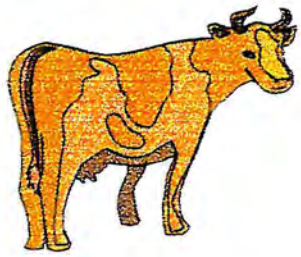
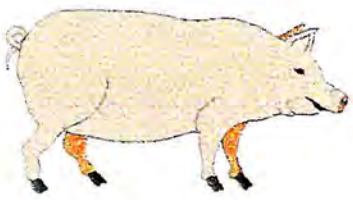
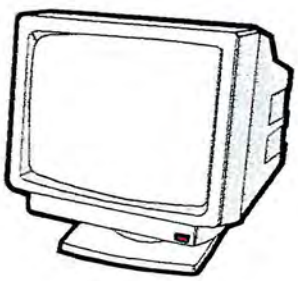


IX REUNION NACIONAL DE MEJORA GENETICA ANIMAL





Arkaute, 25-26 junio 1998

INDICE

PONENCIAS

Ponencia 1. ***Conceptos económicos en un programa de mejora genética en vacuno de leche.*** N. CHARFFEDDINE y R. ALENDA
Ponencia 2. ***Selección para un óptimo: aplicación al tamaño de camada en ovino.*** LOYS BODIN Y COL.
Ponencia 3. ***Mejora genética de la producción de fibra de Cachemira.*** C. DIAZ y A. MORENO

COMUNICACIONES LIBRES

.- ***Selección para morfología mamaria en ovino de leche.*** Seldas, M. E. y col.
.- ***Concentración de células somáticas en leche de oveja y resistencia a la mamitis: nuevos objetivos de selección.*** San Primitivo, F. y col.
.- ***Nuevos caracteres en los objetivos de selección de los programas de mejora.*** Campo, J. L. y García M.
.- ***Selección de ganado caprino utilizando el gen de la caseina α_{s1}.*** Sanchez-Palma A. y col.
.- ***Detección de QTLs mediante la partición de la varianza genética en función del parentesco atribuible a segmentos del genoma.*** Varona, L. y Perez-Enciso, M.
.- ***Teoría de los juegos y optimización de los programas de mejora: fábula del mejorador egoísta y del mejorador altruista.*** Yague, G. y García, L.A.
.- ***Estimación bayesiana de parámetros genéticos y efectos de procedencia de caracteres de crecimiento en Eucalyptus Nitens.*** Fernandez, A. y col.
.- ***Mejora genética del pino insignis***. Espinel, S. y Aragonés, A

PONENCIAS

Ponencia 4. ***La mejora genética frente a los retos medioambientales.*** JOAN TIBAU
Ponencia 5. ***La controversia bayesiana y la metodología de Gibbs.*** AGUSTIN BLASCO

COMUNICACIONES LIBRES

.- ***Estima de parámetros genéticos para prolificidad en ganado ovino mediante muestreo de Gibbs: resultados preliminares.*** Espinosa, M. J. y Jurado J. J.
.- ***Estima de parámetros genéticos para poblaciones de vacuno lechero mediante metodología bayesiana.*** Hernandez, D. y col.
.- ***La huella de Ruperta.*** Rodrigañez, J y col.
.- ***Análisis y comparación de curvas de crecimiento.*** Blasco. A.

POSTERS

.- ***Modelización y estima de componentes de varianza para la producción de leche y proteína en el día del control en la raza ovina Manchega.*** Serrano. M. y col.
.- ***Estrategias de selección para controlar la consanguinidad en programas de mejora***. Fernandez, B. y col

.- ***Selección asistida por marcadores para caracteres no aditivos.*** Toro, M. y Fernandez, J.

.- ***Aplicación de la inseminación artificial en la optimización del esquema de selección en la agrupación caprina canaria: resultados preliminares.*** Lorenzo, M. y col.

.- ***Caracterización de variedades minoritarias del cerdo ibérico***. Delgado J.M. y col.

.- ***Asociación de una lordosis/escoliosis/sifosis a la estructura familiar en la cría intensiva de dorada (sparus aurata).*** Afonso J.M. y col.

.- ***Informatividad de marcadores moleculares tipo microsatélite en dos poblaciones de cerdos de raza ibérica y landrace.*** Clop, A. y col

.-***Mejora y conservación genética de las variedades del podenco andaluz.*** Barba, C. J. y col.

.- ***Caracterización morfométrica de los recursos genéticos ovinos canarios.*** Alvarez, S. Y col.

Conceptos económicos en un programa de mejora genetica en vacuno de leche

*Charffeddine, N. y R. Alenda**
Dpto. Producción Animal, E.T.S.I. Agrónomos, Madrid
Tel.: +34 91 549.18.80 Fax.: +34 91 549.97.63 Email: ralenda@pan.etsia.upm.es*

Resumen:

En una primera parte de este trabajo, se hizo una revisión bibliográfica de los conceptos económicos más relevantes y de los estudios económicos más recientes en vacuno de leche. Las etapas que componen un estudio económico dentro de un programa de mejora genetica son: la definición de un genotipo agregado representativo del objetivo del programa, el establecimiento de una función de beneficio apropiada de los aspectos productivos, económicos y sociales del sistema de producción, la estimación de los pesos económicos de cada uno de los caracteres que forman parte del agregado y él calculo de la expresión acumulativa de descuento correspondiente a cada uno.

En la segunda parte, se desarrolló un ejemplo practico poniendo en manifiesto dichas etapas, haciendo uso de datos económicos y productivos procedentes de ganaderías de la Comunidad Autónoma del País Vasco y de la Cooperativa de Alta Moraña en Avila. Los resultados muestran que los pesos económicos de los caracteres productivos y de la longevidad varían según las perspectivas de producción (libre mercado o existencia de cuota de producción), los niveles de producción y los precios a nivel del mercado. La introducción del sistema de cuota reduce la importancia económica de los caracteres de producción de leche, sobre todo los caracteres bajo restricción, producción de grasa y volumen de leche. Los niveles productivos asi que el sistema de pago de la leche afecta también los pesos económicos de los caracteres productivos. La importancia económica de la longevidad varia especialmente en función de las tasas de desecho y los gastos de recría.

Dado dichas variaciones, el ajuste del genotipo agregado de un programa de mejora a los niveles de producción y las circunstancias de mercado de cada zona es necesario.

Introducción

Un programa de mejora genética busca contribuir a lograr lo que parece lógico en toda empresa: maximizar sus beneficios mediante el incremento de la eficiencia productiva de los animales que la componen. Los animales de la siguiente generación han de producir de forma más eficiente y, por lo tanto, más rentable que la generación actual, bajo condiciones económicas y sociales futuras (Pieters *et al.* 1997). Llevar a la práctica este objetivo está cada vez más influenciado por conceptos económicos que por consideraciones puramente empíricas.

En general, los programas de mejora realizados hasta la actualidad han buscado una simplificación tanto en la organización como en los caracteres a seleccionar. En porcino, avicultura y cunicultura, por ejemplo, los programas de mejora explotan la heterosis y/o la complementariedad de los caracteres y seleccionan las líneas paternas por crecimiento y las maternas por caracteres reproductivos y maternos. En vacuno de carne, la mejora genética se basa casi en exclusiva en la selección por el peso al destete, tanto por las influencias directas como maternas. Las especies que se seleccionan en pureza, como son el vacuno y el ovino de leche, suelen basar su mejora, principalmente, en la selección de los caracteres de producción de leche.

En la ultima década, los programas de mejora en vacuno de leche de varios países del mundo han incorporado otros criterios, algo más complejos, combinando la calidad y la cantidad de leche con caracteres funcionales en índices de selección llamados índices combinados o índices de mérito total, para buscar animales de alta producción y a su vez más longevos. La elaboración de estos índices de selección ha sido posible con métodos objetivos y métodos no objetivos (Groen, 1992). Los métodos no objetivos definen la ponderación de cada

carácter según unas respuestas genéticas deseadas y/o unas restricciones sobre la mejora de algunos caracteres (Kempthone y Nordskog, 1959; Brascamp, 1984, Charffeddine et al., 1995). Los métodos objetivos calculan los valores económicos modelizando el beneficio en función del mérito genético de los caracteres elegidos (Groen, 1988; Van Arendonk, 1985, Bekman y Van Arendonk 1993, Charffeddine, 1998).

En cuanto a los métodos objetivos, la integración de los caracteres elegidos en el genotipo agregado depende de la contribución de cada uno de ellos al aumento del beneficio global (Hazel, 1943). La metodología utilizada para valorar y estimar tal contribución económica es bastante variada y distinta (Groen, 1992). Según Groen *et al.* (1997b) es difícil dar con la metodología ideal para realizar un estudio económico, debido a que esta dependerá, en cada caso, de los caracteres considerados y de las circunstancias de producción y de mercado asumidas. No obstante, el método más apropiado es aquel que refleja fielmente los aspectos productivos y socioeconómicos del beneficio, tratando la mejora genética como cualquier desarrollo tecnológico. Groen y Ruyter (1990) distinguen dos perspectivas o puntos de vista para evaluar esa importancia económica: el punto de vista positivo (Positive Approach) y el punto de vista normativo (Normative Approach) (James y Ellis, 1979). El punto de vista positivo llamado también "Data evaluation" consiste en utilizar datos económicos y técnicos reales para hacer balances económicos, de los cuales se deduce la importancia económica de cada uno de los caracteres estudiados (Andrus y McGilliard, 1975; Dijkhuizen, 1983, France y Thornley, 1984). El punto de vista normativo, llamado también "Data Simulation", consiste en desarrollar un modelo de simulación con parámetros técnicos, económicos y niveles de producción representativos de la población estudiada (Tess *et al.* 1983; Van Arendonk, 1985; Groen 1988).

Según Groen (1992), el modelo desarrollado desde el punto de vista normativo requiere el establecimiento de una función de beneficio (Profit function). Aunque una función de eficiencia sería la terminología más adecuada.

Los modelos normativos ofrecen más ventajas que los modelos positivos a la hora de considerar un gran número de elementos y factores del sistema de producción y sus relaciones y reacciones ante un cambio en el mérito genético de los animales. Groen *et al.* (1997b), afirman que el mayor inconveniente de los modelos positivos consiste en el uso de precios históricos cuando la mejora genética es una orientación hacia el futuro, mientras que, con los modelos normativos, hay mas posibilidades de usar distintos precios, distintos niveles de producción y distintos parámetros técnicos. La mayoría de los estudios económicos hoy en día están basados en una función de beneficios establecida desde el punto de vista normativo (Bekman y Van Arendonk, 1993; Colleau *et al.* 1994; Visscher *et al.* 1994; Pieters *et al.* 1997, Charffeddine, 1998).

El objetivo de este trabajo es hacer una pequeña revisión sobre los conceptos claves de los estudios económicos en mejora genetica y desarrollar un ejemplo de una función de beneficio desde una perspectiva normativa que describa de forma cuantitativa las relaciones que existen entre el nivel genético, las rentas (Output) y los gastos (Input) de una ganadería de vacuno de leche.

Él articulo ha sido estructurado de la siguiente manera:

En la primera parte, se presenta una revisión bibliográfica de los conceptos claves en

los cuales se basa la obtención de los pesos económicos desde un punto de vista normativo, así como los resultados de los estudios económicos más recientes en vacuno de leche.

En la segunda parte, se desarrolla un ejemplo practico en lo cual se establece una función de beneficio desde una perspectiva normativa que describe de forma cuantitativa las relaciones que existen entre el nivel genético, los ingresos y los gastos a partir de datos reales de dos regiones, una en el centro y la otra en el norte de España. Se calculan los pesos económicos de los caracteres de producción, kg. de leche, grasa y proteína a 305 días en situación de cuota y de libre mercado y la duración de la vida productiva en cada zona. El fin de calcular los pesos económicos usando datos económicos y productivos de distintas zonas es estudiar la sensibilidad de estos parámetros a las variaciones en los niveles de producción y las condiciones de mercado.

I- Conceptos claves en un estudio económico dentro de un programa de mejora genética

Para definir la importancia económica de los caracteres a seleccionar en un programa de mejora genetica se requiere:

- La definición de un genotipo agregado
- El establecimiento de una función de beneficio
- La estimación de los pesos economicos
- El calculo de las expresiones acumulativas de descuento.

I.1- Genotipo agregado

Un programa de mejora genética definido desde una perspectiva económica se inicia con la definición de un genotipo agregado (traducción de la terminología inglesa "Aggregate genotype") llamado también valor aditivo económico o valor genético económico (Blasco y Muñoz, 1994). El genotipo agregado define el objetivo del programa de mejora, expresando la eficiencia económica del mérito genético global del animal en función de caracteres que controlan o determinan dicha eficiencia (Groen y Van Arendonk, 1997). Es importante distinguir, en la teoría de los índices de selección, entre el genotipo agregado y el índice que combina las distintas fuentes de información para predecir el valor genético total del animal, llamado comúnmente índice de selección. El genotipo agregado involucra a todos los caracteres que afectan o controlan la eficiencia económica del sistema de producción, mientras que el índice de selección combina todos los caracteres que han sido elegidos como criterio de selección, caracteres cuya información es disponible a nivel de cada animal y cuyos parámetros genéticos permiten alcanzar un progreso genético razonable (Gibson, 1992).

El genotipo agregado (frecuentemente indicado en la bibliografía con la notación H) se define como una ecuación que combina los valores genéticos verdaderos de los caracteres de interés ponderados según su importancia económica relativa (Hazel, 1943) y se escribe como:

$$H = \sum_{i=1}^{n} a_i \cdot g_i$$

Donde :

g_i : Valor genético aditivo verdadero para el carácter i

a_i : Valor económico de descuento del carácter i

El valor económico de descuento del carácter i es el peso económico de dicho carácter multiplicado por su expresión acumulativa de descuento (Groen *et al.*, 1997b). El peso

económico o valor económico de un carácter *i* es, sencillamente, el incremento generado en el beneficio global, debido a un incremento genético de una unidad en el carácter elegido. Se define, generalmente, en la unidad monetaria en la cual ha sido expresado el beneficio por unidad de variación del carácter, por año y por animal (Groen *et al.* 1997b). La expresión acumulativa de descuento es un coeficiente que refleja el tiempo y la frecuencia con los cuales se manifiesta en el futuro la superioridad genetica de cada carácter en la población seleccionada (Groen, 1990).

El índice de selección (frecuentemente indicado en la bibliografia con la notación I) trata de predecir H a partir de **m** caracteres medibles y se escribe como :

$$I = \sum_{i=1}^{m} b_i \cdot x_i$$

Donde,

x_i : Observación o valor genético predicho para el carácter **i**

b_i : Coeficiente de ponderación del carácter **i** a nivel del índice de selección.

m y n pueden ser iguales o no. No todos los caracteres incluidos en el índice de selección son necesariamente incluidos en el genotipo agregado y viceversa.

Los coeficientes de ponderación b_i se calculan maximizando la correlación entre el genotipo agregado y el índice de selección (lo que resulta igual a maximizar el logaritmo de dicha correlación) o minimizando la varianza del error de predicción (estadisticamente equivalente a la minimización de la suma de los cuadrados de las desviaciones entre H e I) (Weller, 1994).

I.2- Función de beneficio

Una función de beneficio es una ecuación que describe los cambios en las rentas económicas netas en función de una serie de parámetros biológicos y económicos (Gibson, 1992). El uso de la función de beneficio en mejora genética animal se limita, principalmente, a describir en un modelo el beneficio en función del mérito genético para los caracteres incluidos en el genotipo agregado, con el fin de calcular sus pesos económicos. Realmente, el beneficio es una función de un conjunto de caracteres x_i , un conjunto de precios de los productos p_j, un conjunto de costes variables cv_k y costes fijos cf_m :

$$B = f(x_1, x_2, ..., x_i, p_1, p_2, ..., p_j, cv_1, cv_2, ..., cv_k, cf_1, cf_2, ..., cf_m)$$

Sin embargo, para simplificar los procedimientos de cálculo de los pesos económicos, se considera habitualmente que los precios y los gastos (tanto fijos como variables) por unidad de producto son constantes, sea cual sea el nivel de producción (Gibson, 1992). Por lo tanto, la función de beneficio se expresa solamente de los valores genéticos aditivos de los caracteres que forman parte del genotipo agregado.

$$B = f(x_1, x_2, ..., x_i)$$

La función de beneficio, para que sea apropiada, ha de reunir según Gibson (1992) las siguientes tres condiciones:

- El cambio en el beneficio debe responder únicamente a una modificación en el mérito genético y no a otro cambio ambiental como podría ser la alimentación, la sanidad, etc.

- Las condiciones de manejo han de ser representativas al momento y el lugar donde se desarrolló el cambio genético.
- Los parámetros económicos han de reflejar la situación del mercado y el sistema de producción del momento y del lugar donde se desarrolla el cambio genético.

La importancia de definir bien la función de beneficio ha sido resaltada en varios trabajos (Groen, 1989b,c; Pieters *et al.*, 1997; Groen *et al.*, 1997b). La función de beneficio es decisiva en dos de las fases más importantes de un programa de mejora genética: el cálculo de los pesos económicos y la elección de los criterios de selección. Cualquier error o simplificación inadecuada podría acarrear consecuencias negativas sobre las respuestas de un programa de mejora genética.

Habitualmente, en los estudios económicos de la mejora genética animal se asume que la relación entre el beneficio y los caracteres considerados es lineal, cuando es obvio que en muchas situaciones no es así (Gibson, 1992). La mejora genética tiene un progreso relativamente lento, sobre todo en vacuno de leche, por lo que los términos de orden superior son, normalmente, despreciables. La aproximación lineal es razonable en el momento concreto en el que se aplica y en ausencia de una situación de saturación de mercado.

La función de beneficio, como se ha dicho anteriormente, se define atendiendo a la relación entre el beneficio y el cambio del nivel genético de los caracteres elegidos. Sin embargo, en la mayoría de los estudios, este concepto básico se olvida y las ecuaciones de beneficio se establecen describiendo las relaciones entre las condiciones de manejo y alimentación y los cambios ocasionados en los rendimientos de los animales. El ejemplo utilizado por Gibson (1992) explica claramente esta situación. Éste ha considerado una función de beneficio que depende del valor genético para la producción de leche, donde la capacidad de ingestión determina los costes de alimentación, de forma que al aumentar la producción de leche por vaca aumentará su capacidad de ingestión. Para estimar los costes de alimentación asociados con el aumento en la producción de leche, se ha de saber la relación entre el incremento del nivel genético de producción y el incremento de la capacidad de ingestión. Sin embargo, lo que se suele utilizar en estos casos para definir dicha relación son fórmulas entre el nivel de producción a escala fenotípica y la capacidad de ingestión recomendados por NRC (National Research Council), ARC (Agricultural Research Council) y INRA (Institut National de Recherche Agricole). La mayoría de estos manuales han sido confeccionados a partir de experimentos en los cuales se asume que todos los animales manifiestan el mismo mérito genético a la hora de relacionar el nivel productivo con la capacidad de ingestión.

Por otra parte, la función de beneficio se define habitualmente igualando los caracteres considerados a la media fenotípica de la población y asumiendo que las posibles relaciones entre los distintos factores de producción y el nivel productivo son extrapolables a nivel genético (Gibson, 1992).

La función de beneficio desde un punto de vista económico, se compone de dos elementos principales: los ingresos (Revenues) y los gastos (Costs), confundidos frecuentemente en producción animal con los productos (Output) y los factores de producción (Input), respectivamente (Gibson, 1992). Para definir la perspectiva según la cual se combinan los gastos y los ingresos en la función de beneficio, Groen (1989a) distinguen dos importantes aspectos a fijar. Primero el objetivo de selección, y segundo, la base de evaluación, es decir, las características del sistema de producción.

Harris (1970) diferencia tres puntos o intereses de selección: primero, maximizar el

beneficio, definido como ingresos menos gastos; Segundo, minimizar los gastos por unidad de producto, y tercero, maximizar los ingresos por unidad de gasto. En mejora genética animal, se considera principalmente el primer o el segundo punto (Groen y Ruyter, 1990). La elección del interés de selección depende del beneficiario de tal proceso de selección o el "decision-maker" como le denomina Groen (1988). Habitualmente, los programas de mejora genética buscan mejorar la rentabilidad económica del propio ganadero, por lo que se busca preferentemente el primer punto, dado que es el que más beneficio global genera para el ganadero. Sin embargo, se ha de utilizar con bastante precaución cuando se trata de una situación de saturación de mercado o de limitación del nivel productivo, como es el caso de la cuota aplicada sobre la producción de leche por parte de la Unión Europea. El segundo punto tiene una importancia relevante únicamente cuando el nivel productivo llega a su nivel óptimo.

Groen (1989a) definió tres opciones para definir la situación base de evaluación: un número fijo de animales dentro del sistema de producción; un nivel fijo de productos que salen fuera del sistema y un nivel fijo de factores que entran dentro del sistema de producción. La elección de una de las tres opciones depende de las condiciones socioeconómicas en las cuales se desenvuelve el sistema de producción. Si el sistema se encuentra en una situación de libre mercado y con un tamaño que se considera suficientemente óptimo, la base de evaluación sería la primera opción, pero si se trata de un sistema de producción en un mercado con restricciones sobre los productos (p.e. cuota de leche) o sobre los factores de producción (p.e. falta de forraje, etc.) sería la segunda o la tercera opción, respectivamente.

Finalmente, hay que evitar dos errores importantes que frecuentemente se cometen a la hora de definir la función de beneficio. Primero, ocultar caracteres bajo otros caracteres, es decir, que en ocasiones un carácter es función de otros y, al no hacerlo explícito, conduce a errores en el cálculo de las derivadas y a confusiones a la hora de determinar los objetivos (Blasco y Muñoz, 1994). Segundo, el "double counting", es decir, considerar un gasto o un ingreso dos veces de formas distintas en la misma función. Por ejemplo, al considerar la mamitis y la producción de leche en un genotipo agregado, la pérdida de producción debida a una infección mamaria no se debe contabilizar a la hora de calcular el peso económico de la mamitis, dado que al incluir el nivel de producción, esta pérdida será doblemente contabilizada (Groen *et al.*, 1997b).

I.3- Los pesos económicos

El peso económico de un carácter determinado es el cambio en el beneficio debido al incremento de una unidad en el mérito genético de tal carácter. Esto se suele calcular comparando la diferencia de beneficios entre la situación actual y la situación en la que un carácter aumenta una unidad, manteniendo los demás caracteres constantes.

Los pesos económicos, definidos desde una perspectiva normativa, dependen tanto de la base de evaluación según la cual ha sido establecida la función de beneficio, como del interés de selección elegido (Moav, 1973).

Brascamp *et al.* (1985) demuestran que al definir la función de beneficio según el concepto "Beneficio Cero" (traducción de la terminología inglesa "Zero Profit"), sea cual sea la base de evaluación, el peso económico de un carácter determinado es constante.

El "Beneficio Cero" es un concepto económico que hace uso de un beneficio llamado "Beneficio Normal" (traducción de la terminología inglesa "Normal Profit") y que se contabiliza en la función de beneficio como un gasto de producción. El "Beneficio Normal" se define como un beneficio necesario para hacer frente a unas inversiones de expansión que

permitan al productor permanecer largo plazo en el sector. El hecho de situarse en el momento en el cual el beneficio adquirido es igual al beneficio normal y, por lo tanto, el beneficio neto es cero, permite equiparar los distintos sistemas de producción, bajo distintas circunstancias y condiciones de mercado y, por lo tanto, hacer factible una valoración comparable de lo que podría generar un cambio genético.

Groen y Ruyter (1990) definen el peso económico según dos tipos de interés de selección y tres bases de evaluación. En la tabla 1 se resumen estas definiciones expresadas en función de conceptos económicos de producción.

Tabla 1. Definición de pesos económicos expresados en función de conceptos económicos de producción según dos tipos de interés de selección y tres situaciones de producción (Groen y Ruyter, 1990).

Base de evaluación	Interés de selección		
	Maximizar el beneficio $\beta = (I - G)$		*Minimizar los gastos por unidad de producto* $\beta = (G/I)$
Libre mercado	Ingresos marginales por unidad de producto menos gastos marginales de factores de producción por unidad de producto.		Media de los gastos totales por unidad de producto menos gastos variables marginales por unidad de producto.
Restricciones sobre la producción (Output)	Ingresos marginales por unidad de producto menos la media por unidad de factores de producción de la sustracción (Ingresos - gastos fijos por animal)		Media de los gastos totales por unidad de producto menos la media de los gastos fijos totales de la explotación por unidad de factores de producción.
Restricciones sobre los factores de producción (Input)	Media de los gastos variables por unidad de producto menos una unidad de gastos marginales.		Media de los gastos variables por unidad de producto menos una unidad de gastos marginales.

I.4- Las expresiones acumulativas de descuento

La expresión acumulativa de descuento (traducción de la terminología inglesa "Cumulative Discounted Expression") define el momento y la frecuencia con los cuales se manifiesta en el futuro la superioridad genética originada por el uso de un animal seleccionado en un programa de mejora genética (Brascamp, 1978).

Multiplicando la expresión acumulativa de descuento por el valor económico se obtiene el valor económico de descuento (traducción de la terminología inglesa "Discounted Economic Value"), que es el valor económico a considerar cuando se quiere comparar la importancia económica relativa de caracteres que se manifiestan de distintas formas y en diferentes momentos a lo largo de la vida del animal (Groen *et al.*, 1997b).

La expresión acumulativa de descuento en un tiempo **t** para un carácter determinado (E_t) se define como :

$$E_t = E_{t-1} + h' \cdot m_t \left(\frac{1}{1+r}\right)^t$$

Donde:

E_t, E_{t-1} : Expresión acumulativa de descuento en el tiempo *t* y *t-1*, respectivamente.

m_t : Vector de las proporciones de genes transmitidos a partir de la generación base en el tiempo t.

r : tasa de descuento.

h : Vector de incidencia con $(n+q)$ elementos, describe la frecuencia relativa de los animales en la población que expresan el carácter dentro de cada clase de edad y de sexo.

El calculo de la expresión acumulativa de descuento de un carácter determinado requiere la definición de dos conceptos claves:

- El factor de descuento ligado al carácter
- El flujo de genes según el cual se transmite el carácter de una generación a la siguiente.

Factor de descuento

El concepto de descuento parte del principio económico de inversión, según el cual el valor verdadero del dinero sigue un proceso de depreciación continuo a lo largo de los años, es decir, que una suma de dinero hoy vale más que mañana y así sucesivamente a medida que transcurre el tiempo. Si un programa de mejora genética o una decisión determinada de selección (como es el caso de elegir un semental) cuesta una suma de dinero hoy y sus resultados no se pueden cosechar hasta varios años después, la evaluación de los costes o beneficios se debería hacer teniendo en cuenta el valor verdadero del dinero en cada año y, referir todos los gastos e ingresos a una época determinada para que sean comparables. Ocurre lo mismo a la hora de valorar el beneficio marginal que se podría obtener al mejorar genéticamente caracteres que se expresan de distintas formas y en diferentes momentos de la vida del animal.

Económicamente, el concepto de descuento consiste en equiparar los gastos y los ingresos producidos en años distintos y expresarlos como gastos e ingresos en un año **t**. De forma que un gasto $\mathbf{X_0}$ en el año **0** vale $\mathbf{X_0(1 + r)^t}$ en el año **t** y un gasto $\mathbf{X_i}$ en el año **i** vale $\mathbf{X_i(1 + r)^{t-i}}$ en el año **t**. Siendo **r** la tasa de descuento.

Dado que el hecho de referir los gastos y los ingresos a un año **t** podría ser algo arbitrario a la hora de fijar **t** , Gibson (1992) ve más conveniente referirlos al momento en el cual se inició la inversión y expresarlos en función del valor inicial del citado momento. El valor inicial (VI) de un gasto o ingreso en el año **t** es igual a :

$$VI = \frac{X_t}{(1+r)^t} = d \cdot X_t$$

$$\text{con} \quad d = \frac{1}{(1+r)^t}$$

donde:

r : tasa de descuento

d : factor de descuento para un valor financiero del año **t**

La tasa de descuento r, llamada en términos económicos "la tasa atractiva mínima de retribución o ganancias" (traducción de la terminología inglesa "Minimum attractive rate of return "), define la tasa anual de ganancia mínima que hay que generar al invertir en un negocio para equipararla a la depreciación del dinero (Gibson, 1992). No obstante, Gibson (1992) recomienda utilizar el tipo de interés real o actual (es decir, el tipo de interés corregido

por la tasa de inflación) como tasa de descuento, siendo ésta la mejor opción de las utilizadas. Las tasas de descuento utilizadas en los estudios económicos de mejora genética varían del 3 al 5% (Groen, 1990; Dekkers, 1994; Pieters *et al.*, 1997).

Gibson (1992) sugiere utilizar una tasa de descuento ligeramente superior para los ingresos que para los gastos, para favorecer los caracteres que generan ingresos a edades más tempranas y que producen gastos de una forma más escalonada.

Flujo de genes

El flujo de genes estudia el proceso de transmisión de los genes a través de la población y estima en cada periodo la proporción de genes transmitida a cada grupo de animales proveniente de la generación base, sobre la cual se ha tomado la decisión de selección (Gibson, 1992).

La transmisión de genes se puede definir en términos de una proporción de genes en un grupo de animales procedentes de la población base (Hill, 1974) o en términos de porcentaje del mérito genético aditivo de los animales de la población base transmitido a su descendencia (Elsen, 1980).

Estimar la proporción según la cual se transmiten los genes a lo largo de los años en los distintos grupos de animales, precisa definir la población base, conocer la estructura de la población, es decir, definir los distintos grupos de animales, el periodo de tiempo necesario para que un animal pase de un grupo a otro y la forma según la cual se transfieren los genes entre grupos de animales (reproducción o envejecimiento) (Groen y Van Arendonk, 1997).

La elección de la población base depende del objetivo de selección y del beneficiario de la decisión de selección tomada, por lo que puede estar formada únicamente por sementales, por hembras o por los dos juntos. El número de grupos de animales que estructura la población se determina según los cambios del estado fisiológico de cada sexo. El periodo de tiempo según el cual se rige el cambio de un grupo a otro depende de la especie animal considerada y del sistema de producción. Por ejemplo, en porcino los intervalos de tiempo considerados son de 6 meses mientras que en vacuno de leche son de un año (Groen y Van Arendonk, 1997).

La transmisión de genes se sigue y se estima durante la vida de una generación, por lo que la estimación de la proporción de genes en cada clase o grupo de animales se hace periódicamente, siendo el intervalo de tiempo de cada periodo igual al periodo de tiempo de cada clase de edad.

Sea $\mathbf{m}_t$ un vector de elementos $m_{lt}(i)$ que define la proporción de genes de interés en la clase de sexo **l** con edad **i**, en el tiempo **t**.

Existen dos procesos que influyen en el tránsito de genes entre un grupo de animales y otro: la reproducción y el envejecimiento.

$$m_t' = \left[\underbrace{m_{1t}(1) \quad m_{1t}(2) \quad \ldots \quad m_{1t}(n)}_{\text{n clases de edad de los machos}} \quad \underbrace{m_{2t}(1) \quad m_{2t}(2) \quad \ldots \quad m_{2t}(q)}_{\text{q clases de edad de las hembras}} \right]$$

m_t se calcula como :

$$m_t = [R + Q]m_{t-1} = P \cdot m_{t-1}$$ (Groen y Van Arendonk, 1997)

Donde:

R : matriz que describe la transmisión de genes por reproducción entre las distintas clases de edad (las **n** clases de machos y **q** clases de hembras).

Q : matriz que describe la transmisión de genes por envejecimiento entre animales de distintas clases de edad.

m_{t-1} : es la proporción de genes que hay en cada clase de edad y sexo en el momento ***t-1***.

P : matriz de probabilidad de transición de genes por las dos vías (reproducción y envejecimiento) entre animales de distintas clases.

La forma general de P es la siguiente :

$$P = \underbrace{\text{clases de edad - sexo a donde llegan los genes}}_{\text{clases de edad-sexo de donde proceden los genes}} \left\{ \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1,n+q} \\ \cdot & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & \cdot \\ p_{n+q,1} & \cdots & \cdots & p_{n+q,n+q} \end{bmatrix} \right.$$

Donde:

p_{ij} : proporción de genes en la clase de edad-sexo i en el tiempo t que procede de la clase edad-sexo j en el tiempo ***t-1***.

Los elementos de m_t en cada periodo se calculan de la siguiente forma:

$$m_1 = P \cdot m_0$$
$$m_2 = P \cdot m_1$$
$$\cdots \quad \cdots \quad \cdots$$
$$m_t = P \cdot m_{t-1}$$

m_t se define en función de la proporción de genes en la generación de base (m_0) de la siguiente forma:

$$m_t = P^t \cdot m_0$$

II- Resumen de los estudios económicos más recientes en vacuno de leche

En este apartado se recapitulan los trabajos más recientes que han estimado los pesos económicos de los caracteres de producción y de los caracteres funcionales en vacuno de leche.

Los pesos económicos dependen de los parámetros económicos del mercado, del sistema y nivel de producción y de la metodología utilizados en cada estudio, por lo que no es adecuado comparar los valores obtenidos entre un estudio y otro (Groen *et al.* 1997b). Por este motivo, se va a hacer más énfasis en describir la metodología y las condiciones de cada

estudio, que en comparar e intentar dar explicaciones a las divergencias entre los valores económicos de un carácter determinado en un estudio y otro.

II.1. Pesos económicos de los caracteres productivos

Numerosos estudios han estimado recientemente los pesos económicos de la producción de leche y de sus componentes en distintos sistemas de producción, condiciones de mercado y bajo diferentes perspectivas (Van Arendonk *et al.*, 1985; Wilmink, 1988; Gibson, 1989; Groen, 1989b y c; Gibson, 1992; Bekman y Van Arendonk, 1993; Harris y Freeman, 1993; Colleau *et al.*, 1994; Steverink *et al.*, 1994; Visscher *et al.*, 1994; Veerkamp, 1996; Pieters *et al.*, 1997).

El primer trabajo que tuvo en cuenta la incidencias de la cuota sobre los objetivos de selección fue el de Van Arendonk *et al.* (1985). En este trabajo se ha realizado una aplicación de las consideraciones teóricas, desarrolladas por Brascamp *et al.*, (1985) y Smith *et al.*, (1986), estudiando el efecto del redimensionamiento de las explotaciones lecheras en los márgenes económicos. Van Arendonk *et al.* (1985) consideran dos situaciones extremas: una en la cual hay una cantidad fija de factores de producción (forraje) (si el número de animales disminuye, cada animal tendrá que consumir más factores de producción) y la otra donde la cantidad de factores de producción consumida por animal es fija y la disminución del número de animales, como consecuencia de la cuota, implica una sobreproducción de factores de producción que será vendida. La aplicación de la cuota sobre la producción en la primera situación, considerada como más verosímil, ocasionó una revalorización de la selección sobre los dos porcentajes (grasa y proteína). La pérdida económica debida a la aplicación de la cuota fue evaluada en un 18% del progreso económico ligado a la mejora genética antes de la instauración de la cuota.

Wilmink (1988) estudió la posibilidad de modificar la composición de la leche a través de un programa de mejora genética. Se estimaron los pesos económicos de la leche y de sus componentes en una situación de libre mercado y en presencia de cuota. La desviación típica del índice de selección establecido disminuyó en un 40% de su valor inicial al imponer el sistema de cuota, lo que genera una reducción en la rentabilidad de la selección por caracteres productivos y, por lo tanto, aumenta el interés relativo en los caracteres llamados secundarios.

Gibson (1989), con una metodología similar a la de Van Arendonk *et al.* (1985), estima los pesos económicos del volumen de leche, de grasa y de proteína en ausencia y presencia de restricciones sobre la producción de leche en las condiciones económicas inglesas. Los gastos fijos fueron considerados como proporcionales al número de vacas y los gastos de alimentación suplementarios de los animales seleccionados fueron cubiertos por un suplemento de pienso. En este trabajo, el peso económico de la grasa disminuye en un 60% de su valor inicial al instaurar la cuota.

Groen (1988, 1989b,c, 1990) estudia la incidencia de la cuota, en las condiciones holandesas, sobre el comportamiento general del beneficio de un programa de mejora genética. El peso económico de la grasa disminuye en un 62% de su valor inicial bajo un sistema de cuotas.

Harris y Freeman (1993), mediante una programación lineal, han estudiado el efecto de

diferentes niveles de progreso genético para distintos caracteres (especialmente la grasa) en un rebaño de 80 vacas durante 10 años. El peso económico de los caracteres considerados ha sido calculado variando el beneficio total y cuando el progreso genético de un carácter determinado varía en una unidad, manteniendo los demás caracteres en unos niveles constantes y teniendo en cuenta una perspectiva de un número fijo de animales en el rebaño. En situación de cuota, el peso económico de la grasa disminuye en un 105% de su valor inicial y el de proteína aumenta en un 25%. Este resultado ha sido posible debido a una hipótesis poco verosímil que supone que el volumen de leche y de grasa que excede de la cuota no serán pagados, mientras que la proteína sí, por lo que el peso económico de la proteína sale favorecido en situación de cuota.

Bekman y Van Arendonk (1993) han establecido una función de beneficio para estimar la importancia económica de la producción de leche, la producción de carne, la dificultad al parto y la mortalidad de los terneros en los rebaños de vacuno de leche dentro de las circunstancias de producción y de mercado holandeses. Bajo una situación de cuota, los pesos económicos del volumen de leche, de la grasa y de la proteína disminuyen en un 300, 81,7 y 1,7% de sus valores iniciales, respectivamente.

Colleau *et al.* (1994) han estudiado las modificaciones introducidas en el índice de selección (INEL) en Francia en 1993 para ajustar la importancia relativa de cada carácter a las condiciones de la cuota. Han sido consideradas distintas alternativas y condiciones de mercado para estudiar la eficiencia y la adaptabilidad del nuevo INEL a diferentes circunstancias de producción. El nuevo INEL limita drásticamente la producción de grasa, estimula el aumento del porcentaje de proteína y favorece animales con mayor cociente % proteína / % de grasa . La aplicación de la cuota reduce el peso económico de la grasa en más del 100% de su valor inicial, por lo que un incremento en el nivel genético para la producción de grasa, pasa de ser premiado a ser penalizado.

Visscher *et al.*, (1994) han estimado los pesos económicos de los caracteres de producción de leche (Volumen de leche, de grasa y de proteína) en un sistema de producción basado en el pastoreo en Australia. Los factores de producción (alimentos) han sido considerados como constantes por lo que un aumento en las necesidades energéticas de los animales seleccionados genera una disminución del número de animales. Han sido estudiadas varias situaciones de nivel de producción, de parámetros económicos, etc. La producción de leche y de grasa tienen la misma importancia económica, que es equivalente al 40% de la importancia económica de la proteína.

Veerkamp (1996), utilizando un modelo bioeconómico, ha estimado los pesos económicos de los caracteres de producción, de la capacidad de ingestión y del peso adulto de los animales considerando dos niveles de producción. Los pesos económicos de los caracteres productivos no varían al pasar de una producción base de 5.500 kg./vaca/año a una producción de 7.700 kg./vaca/año, dado que no se ha considerado ninguna restricción sobre el nivel productivo.

Pieters *et al.* (1997), utilizando el mismo modelo desarrollado por Bekman y Van Arendonk (1993), han estudiado tres sistemas de pago de la leche en distintas regiones de Italia y la incidencia de la introducción de la cuota sobre cada uno de ellos. Los valores económicos de los caracteres productivos disminuyen entre un 47 y un 52 % de sus valores iniciales con la instauración de la cuota. La diferencia de la renta económica entre un sistema de pago y otro,

ambos sin cuota, varía de un 4,2 a un 16,7%. Sin embargo, al aplicar la cuota láctea, la diferencia entre los sistemas de pago asciende a un 57% en algunos casos.

II.2. Peso económico de los caracteres funcionales

II.2.1. Longevidad:

La longevidad es un componente muy importante en la rentabilidad económica de las ganaderías de vacuno de leche. El incremento en longevidad de los animales aumenta el beneficio global de una ganadería de cuatro formas: reduce los gastos de reemplazo por vaca y por rebaño, cambia la estructura de edades en el rebaño, aumentando la proporción de las vacas adultas en máximo de producción, incrementa la intensidad de selección voluntaria generando así más progreso genético y aumenta la proporción de vacas en producción con respecto a las novillas en todo el rebaño. Para valorar la importancia económica de la longevidad, se necesita un modelo bastante extenso y técnicas de programación matemática sumamente complejas (Groen *et al.*, 1997a).

La optimización del momento y la intensidad del reemplazo en un rebaño, así como la importancia económica de la longevidad, están estrechamente ligados con la magnitud relativa del coste de criar o comprar una novilla con respecto al coste de recuperar, y por lo tanto, no desechar una vaca (Van Arendonk, 1985).

Se han considerado dos enfoques para valorar la importancia económica de la longevidad. El primero valora el beneficio marginal generado por un aumento en la vida productiva de las vacas (Van Arendonk, 1991; Allaire y Gibson, 1992; Harris y Freeman; 1993; Weigel *et al.*, 1995 y Pieters y Bagnato 1996) y el segundo estima el beneficio económico provocado por una reducción en el porcentaje de desecho involuntario (Van Arendonk, 1985; Rogers *et al.*, 1988 y Brandts *et al.*, 1996).

La importancia económica de la longevidad respecto a la producción de leche en un programa de mejora genética ha sido estimada en varios trabajos. Los resultados son bastante contradictorios. Allaire y Gibson (1992), Dekkers (1993) y Van Raden y Wiggans (1995) concluyen que el beneficio generado por la producción de leche es de 2,5 a 2,7 veces más importante que el beneficio debido a una vida productiva más larga. Sin embargo, Rogers *et al.*, (1989) y Van Arendonk (1991) afirman que el peso económico de la producción lechera, respecto a la vida productiva, varía solamente de 0,8 a 1,4 veces. Los distintos planteamientos seguidos en cada uno de los estudios (considerar o no el coste de oportunidad, ajustar o no la vida productiva por la producción de leche, etc.), así como la sensibilidad de estas evaluaciones económicas a los sistemas de producción y las condiciones de mercado podrían explicar dichas diferencias (Groen *et al.*, 1997b).

II.2.2. Recuento de células somáticas

La estimación de la importancia económica del recuento de células somáticas suscita dos puntos de discusión, la no linealidad del beneficio respecto a los niveles del recuento

celular y la separación entre los costes debidos exclusivamente a un alto recuento celular y los costes generados por una infección mamaria.

El recuento de células somáticas a determinados niveles no tiene ninguna consecuencia sobre el beneficio global de una ganadería. Sin embargo, como consecuencia de una infección mamaria, el recuento celular supera rápidamente esos niveles y estará acompañado de una reducción en la cantidad y la calidad de producción de leche. Por ella, el recuento de células somáticas no influye de forma lineal sobre el beneficio. Weller *et al.*, (1996) han definido una función de beneficio no lineal para el recuento de células somáticas según las condiciones de producción y de mercado en Israel. Sin embargo, concluyeron que cuando la función de beneficio no es lineal, no hay una solución uniforme y única a la hora de establecer un índice de selección que maximice el beneficio. El índice que maximiza el objetivo de selección en las siguientes generaciones depende de la intensidad de selección, sobre todo si la media de la población está cerca del óptimo económico (Weller *et al.*, 1996).

Los costes de un alto recuento celular se confunden frecuentemente con las pérdidas y los costes de una infección subclínica mamaria (Reneau y Packard, 1991, Schepers y Dijkhuizen, 1991). El recuento de células somáticas influye negativamente en el beneficio, principalmente a través de un deterioro de la composición química y la calidad higiénica de la leche. La mayoría de los países del entorno económico español han incluido en los criterios de pago de la leche una penalización a partir de determinados niveles de recuento celular a nivel del tanque (Reents, 1997).

Kolstad y Dekkers (1994) han estimado la importancia económica del recuento celular en la primera y segunda lactación, contabilizando como único coste el deterioro de la calidad y la cantidad de los componentes de la leche relacionados con la producción de queso.

Weller *et al.*, (1996) estimaron la importancia económica del recuento de células somáticas considerando únicamente la penalización en el sistema de pago de la leche a partir de determinados niveles de recuento celular.

II.2.3. El peso adulto de las vacas en producción

El peso adulto de las vacas en vacuno de leche muestra generalmente un peso económico negativo (Groen *et al.*,1997a). Los gastos marginales asociados al incremento de las necesidades energéticas tanto para el crecimiento de las novillas de recría como para el mantenimiento de las vacas en producción, superan los ingresos marginales generados por un aumento en el peso de los animales desechados (Groen, 1989b).

El peso económico del peso adulto se estima generalmente sin tener en cuenta los cambios que podrían ocurrir en la composición corporal de los animales. La magnitud de los valores económicos estimados en la bibliografía depende principalmente de los costes de alimentación y del precios de la carne asumido en cada estudio (Groen, 1989b).

Visscher *et al.* (1994) han estimado el peso económico del peso adulto en un sistema de producción basado en pastoreo en Australia concluyendo que, al limitar la cantidad de forraje disponible en cada explotación, el peso económico del peso adulto disminuye drásticamente. Morris y Wilton, (1977) y Groen (1989b) dedujeron que al ignorar el periodo de recría, el peso económico del peso adulto se veía poco influenciado por el sistema de

producción y las condiciones de mercado.

III- Ejemplo practico de un estudio económico en vacuno de leche

III.1. Datos Utilizados

En este estudio se han utilizado datos del programa de Gestión Técnico-Económica de los 3 centros provinciales (Lurgintza, Lorra y Asgafal) de la Comunidad Autónoma del País Vasco correspondientes a 239 explotaciones durante año 1995 y datos económicos y productivos correspondientes al año 1997, de 22 explotaciones pertenecientes a la Cooperativa Alta Moraña, en Ávila.

La tabla 2 expone las medias de los parámetros productivos y económicos más relevantes correspondientes a las ganaderías del País Vasco y la cooperativa de Alta Moraña.

Tabla 2. Medias de los parámetros productivos y económicos más relevantes calculadas sobre los datos de las ganaderías de la Cooperativa Alta Moraña y del País Vasco.

		País Vasco	**Alta Moraña**
Datos productivos			
Producción de leche	(kg.vaca^{-1}.año^{-1})	7850	8630
Producción de grasa	(kg.vaca^{-1}.año^{-1})	298	379
Producción de proteína	(kg.vaca^{-1}.año^{-1})	235	309
Longitud de la vida productiva	(años)	4.02	3,64
Intervalo entre partos	(dias)	402	415
Precios y costes del mercado			
Leche	(ptas.kg^{-1})	42,4	49.01
Volumen de leche ($i_{Volumen}$)	(ptas.kg^{-1})	10.49	10.54
Grasa (i_{Grasa})	(ptas.kg^{-1})	420.49	510,54
Proteína ($i_{Proteina}$)	(ptas.kg^{-1})	550,49	710.54
Terneros	(ptas.kg de peso vivo^{-1})	850	337
Vacas de desecho	(ptas.kg de peso vivo^{-1})	194	145
Gastos globales de recría	(ptas.novilla^{-1})	218.941	256.341
Coste de una unidad forrajera			
ración de las vacas	(ptas.UFL^{-1})	27.40	31.67
ración de las novillas	(ptas.UFL^{-1})	25.90	38,40

III.2. Descripción de la función de beneficio utilizada

El modelo utilizado es un modelo estático, en el cual el beneficio anual de cada rebaño ha sido descrito en función del número de animales (N) y de los caracteres productivos y funcionales siguientes:

- kg. de Volumen de Leche a 305 días (V)
  (Volumen de leche = kg. de leche total – kg de Grasa – kg de Proteína)
- kg. de Grasa a 305 días (G)
- kg. de Proteína a 305 días (P)
- Longitud de Vida Productiva (L)

El beneficio anual se contabilizó como ingresos menos gastos. Se asume que el cambio en el mérito genético de un carácter determinado no genera cambios en los ingresos ni en los gastos que pudieran derivarse del cambio asociado en los demás caracteres y tampoco en los parámetros técnicos y económicos iniciales. El único cambio posible se refiere al número de animales. Una restricción sobre la producción, como es el caso de la cuota, se traducirá en una

reducción del efectivo de cada rebaño.

La ecuación general del beneficio anual dentro de cada rebaño es similar a las desarrolladas en los trabajos de Groen (1988, 1989b, c), Gibson (1989), Van Arendonk y Brascamp (1990), Bekman y Van Arendonk (1993) y Pieters *et al.* (1997).

$$ß = N (I_a - G_a) - G_f$$

donde:

N : Número de vacas presentes en el rebaño

I_a : Ingresos de una vaca durante un año

G_a : Gastos de una vaca durante un año

G_f : Gastos fijos del rebaño

El beneficio anual de cada animal $ß_a$ es igual a : $ß_a = I_a - G_a$

Los ingresos relacionados con la actividad de vacuno de leche considerados son de dos tipos:

- Ingresos por venta de leche.
- Ingresos por venta de carne.

Los ingresos de cada animal debidos a la venta de leche (I_{VL}) dependen de su cantidad (l) y de su calidad que es el contenido en grasa y proteina.

$$I_{VL} = l\left[p_L + p_G \cdot (G - G_0) + p_P \cdot (P - P_0)\right]$$

Donde:

l : Producción media de leche (kg.).

G : Contenido en grasa de un kg. de leche (g·kg. de leche^{-1}).

P : Contenido en proteína de un kg. de leche (g·kg. de leche^{-1}).

G_0 : Grasa de referencia según el criterio de pago (g·kg. de leche^{-1}).

P_0 : Proteína de referencia según el criterio de pago (g·kg. de leche^{-1}).

p_L : Precio de la leche (ptas.kg^{-1}).

p_G : Prima de calidad de la grasa (ptas.g^{-1}.kg de leche^{-1}).

p_P : Prima de calidad de la proteina (ptas.g^{-1}.kg de leche^{-1}).

Los ingresos anuales por venta de carne (I_{VC}) proceden de la venta de vacas de desecho y terneros recién nacidos.

$$I_{VC} = \left[\left(p_{VD} \cdot PV_{VD}\right)\left(1/L - Mortd_V\right) + \left(p_{tr}\right)\left(NP\left(1 - Mortd_t\right)\right)\right]$$

Donde:

p_{VD} : Precio de un kg. de peso vivo de las vacas de desecho (ptas.kg^{-1}).

PV_{VD} : Peso medio de las vacas desechadas (kg.).

p_{tr} : Precio medio de un ternero recién nacido (ptas.ternero^{-1})

Número de vacas de desecho = $N_{VD} = N * (1/L - Mortd_v)$

L : Longitud de la vida productiva media de las vacas del rebaño (años).

N : Número de vacas presentes en el rebaño.

$Mortd_v$: Mortalidad de las vacas en tanto por uno.

Número de terneros vendidos = $N_{TV} = N\ (NP)\ (1 - Mortd_t)$

NP : Número de partos por año = 365 / IP

IP : Intervalo entre partos (días).

$Mortd_t$: Mortalidad de terneros en tanto por uno.

Los gastos de cada ganadería han sido detallados tanto para vacas en producción como para novillas de reemplazo. Para cada grupo de animales se han considerado dos tipos de gastos : gastos de alimentación y gastos fijos por animal.

Los gastos de alimentación han sido definidos en términos de necesidades energéticas. Las necesidades energéticas de las vacas en producción se han calculado en energía neta de leche y se han expresado en Unidades Forrajeras Leche (UFL), distinguiendo los diferentes estados fisiológicos, conservación o mantenimiento, gestación, crecimiento y lactación.

La ecuación global de los gastos de cada animal (G_a), dentro de una ganadería durante un año, es la siguiente :

$$G_a = c_g \cdot G + c_p \cdot P + c_V \cdot V + C\left(E_{Mant_v}\right) + C\left(E_{Crec_v}\right) + C\left(E_{gest_v}\right) + GF_V$$

$$+ \left(1/L + Mortd._N\right)\left(\underbrace{C(E_{Mant_N}) + C(E_{Crec_N}) + C(E_{gest_N}) + GF_N}_{G_{recria}}\right)$$

Donde,

c_g, c_p y c_v : Costes marginales de producir un kilo de grasa, proteína y volumen de leche, respectivamente (ptas.kg^{-1}).
G, P y V : Producción media de grasa, proteína y volumen de leche (kg.vaca^{-1}.año^{-1}).
$C(E_{Mant_V})$: Coste de la energía de mantenimiento de las vacas (ptas.vaca^{-1}.año^{-1}).
$C(E_{Crec_V})$: Coste de la energía de crecimiento de las vacas (ptas.vaca^{-1}.año^{-1}).
$C(E_{gest_V})$: Coste de la energía de gestación de las vacas (ptas.vaca^{-1}.año^{-1}).
Gf_v : Gastos fijos por vaca (ptas.vaca^{-1}.año^{-1}).
L : Longitud de la vida productiva (años).
$Mortd._N$: Mortalidad entre las novillas en el periodo de recría en tanto por uno.
$C(E_{Mant_N})$: Coste de la energía de mantenimiento de las novillas (ptas.novilla^{-1}).
$C(E_{Crec_N})$: Coste de la energía de crecimiento de las novillas (ptas.novilla^{-1}).
$C(E_{gest_N})$: Coste de la energía de la primera gestación de las novillas (ptas.novilla^{-1}).
GF_N : Gastos fijos por novilla en el periodo de recría (ptas.novilla^{-1}).

Los costes marginales c_i (i = grasa, proteína, volumen) se calculan multiplicando los gastos energéticos necesarios (GE_i) por el coste de una unidad forrajeara (p_{UFL}):

$$c_i = p_{UFL} \cdot GE_i$$

III.3. Cálculo de los pesos económicos

Para calcular el peso económico de un carácter determinado se deriva parcialmente la función de beneficio (β) con respecto a dicho carácter manteniendo los demás caracteres en sus niveles medios de producción en los rebaños del estudio.

Los pesos económicos de los caracteres de producción de leche han sido calculados considerando dos situaciones: una de libre mercado y otra en la cual opera la cuota sobre productos lácteos impuesta por la Unión Europea.

El peso económico a_{x_i} de un carácter x_i en situación de libre mercado se calcula mediante la siguiente ecuación:

$$a_{x_i} = \frac{1}{N} \cdot \frac{\partial \beta}{\partial x_i} = \frac{\partial I_a}{\partial x_i} - \frac{\partial G_a}{\partial x_i} \Big|_{x=\mu} \qquad (3)$$

Donde:

x_i : Valor del carácter considerado.
x : Vector de todos los caracteres.
μ : Valores medios de todos los caracteres en la población del estudio.

En el caso de una restricción sobre la producción, como es el caso de la cuota, el número de animales en el rebaño no es constante. El peso económico a_{x_i} de un carácter x_i se calcula mediante la siguiente ecuación:

$$a_{x_i} = \frac{1}{N} \cdot \frac{\partial \beta}{\partial x_i} = \frac{\partial I_a}{\partial x_i} - \frac{\partial G_a}{\partial x_i} + \frac{1}{N}\left(I_a - G_a\right) \cdot \frac{\partial N}{\partial x_i} \Big|_{x=\mu} \qquad (4)$$

El régimen de tasa suplementaria en el sector lácteo, más conocido como régimen de cuotas, fue impuesto por la Comunidad Europea en 1984 con el fin de mantener el equilibrio del mercado de los productos lácteos en la Unión Europea, permitiendo al mismo tiempo la evolución estructural del sector. La cuota asignada a cada ganadero está compuesta por dos elementos: una cantidad de referencia de leche a producir y un porcentaje de grasa de referencia que tiene que contener como máximo la leche producida. Según este sistema, los productores de leche deben pagar una multa de 115% del precio de la leche, llamada tasa suplementaria, por cada kg. de leche que venden a las centrales lecheras y que supere la cuota de leche equivalente asignada. La cantidad de leche equivalente, sobre la cual se aplica la tasa suplementaria, es la cantidad de leche producida después de su eventual corrección por la diferencia entre su contenido en grasa y la grasa de referencia. El volumen de leche equivalente al nivel de grasa de referencia en un rebaño se calcula mediante la siguiente ecuación:

$$Q = N\left[leche\left(1 + k_g \cdot \left(pg - pg_{ref}\right)\right)\right]$$

Donde:

Q : Volumen de leche con un nivel de grasa igual a la grasa de referencia (kg.).
N : Número de vacas en producción.
$Leche$: Producción de leche media en el rebaño (kg.vaca^{-1}).
k_g : Factor de escala que traduce las décimas de diferencia entre los niveles de grasa en kg. de leche (=18).
pg : Porcentaje de grasa medio de la leche producida en el rebaño en tanto por uno.
pg_{ref} : Porcentaje de grasa de referencia en tanto por uno (= 0,037).

Teniendo en cuenta que:

$$leche = V + G + P \quad \text{y} \quad leche \cdot pg = G$$

la expresión de la cuota queda:

$$Q = N\left[leche\left(1 + k_g \cdot \left(pg - pg_{ref}\right)\right)\right]$$

$$= N\left[\underbrace{(1 - k_g \cdot pg_{ref})}_{q_v} \cdot V + \underbrace{(1 - k_g \cdot pg_{ref} + k_g)}_{q_g} \cdot G + \underbrace{(1 - k_g \cdot pg_{ref})}_{q_p} \cdot P\right]$$

$$= N\left[q_v \cdot V + q_g \cdot G + q_p \cdot P\right]$$

$$= N\sum_{i=1}^{3} q_i \cdot x_i$$

La derivada de la cantidad de leche equivalente en situación de cuota con respecto a la producción de leche y sus componentes es por definición igual a cero:

$$\frac{\partial Q}{\partial x_i} = \frac{\partial N}{\partial x_i} \cdot \sum_{i=1}^{3} q_i \cdot x_i + N \cdot q_i = 0$$

Por lo que ,

$$\frac{\partial N}{\partial x_i} = -N \cdot \frac{q_i}{\sum_{i=1}^{3} q_i \cdot x_i}$$

Sustituyendo la derivada de N con respecto a x_i en (4), obtendremos la expresión del peso económico de la producción de leche y de cada uno de sus constituyentes bajo una situación de cuota:

Peso económico del volumen de leche:

$$a_{Volumen} = i_{Volumen} - c_{Volumen} - \frac{q_v}{\sum q_i x_i} \cdot \beta_a$$

Donde:

$i_{Volumen}$: Ingreso marginal de un kg. de volumen de leche (ptas.kg^{-1}).

$c_{Volumen}$: Coste marginal de un kg. de volumen de leche (ptas.kg^{-1}).

Peso económico de la grasa:

$$a_{Grasa} = i_{Grasa} - c_{Grsasa} - \frac{q_g}{\sum q_i x_i} \cdot \beta_a$$

Donde:

i_{Grasa} : Ingreso marginal de un kg. de grasa (ptas.kg^{-1}).

c_{Grasa} : Coste marginal de un kg. de grasa (ptas.kg^{-1}).

Peso económico de la proteina:

$$a_{Proteina} = i_{Proteina} - c_{Proteina} - \frac{q_p}{\sum q_i x_i} \cdot \beta_a$$

Donde:

$i_{Proteina}$: Ingreso marginal de un kg. de proteína (ptas.kg^{-1}).

$c_{Proteina}$: Coste marginal de un kg. de Proteína (ptas.kg^{-1}).

Al derivar la función de beneficio β con respecto a L, la expresión del peso económico de la longevidad es la siguiente:

$$a_L = \frac{-p_{VD} \cdot PV_{VD} + G_{recria}}{L^2}$$

III.3. Estimación de las expresiones acumulativas de descuento para cada carácter

Las expresiones acumulativas de descuento para los caracteres kg. de leche, volumen de leche, de grasa, de proteína y longitud de la vida productiva han sido calculados mediante el programa informático denominado "geneflow" (Brascamp, 1978) basado en los conceptos descritos por Hill (1974) y Brascamp (1975).

Descomponiendo las vías de progreso genético, la población base de la cual parten los genes ha sido dividida en dos grupos:

- Sementales de inseminación artificial subdivididos en 9 clases donde cada una representa un año de edad. Se asume que los toros se ponen en prueba en el transcurso del tercer año de edad y obtienen las pruebas oficiales a los 6 años, es decir, entre el tercer y sexto año los toros no se utilizan, aunque esta última consideración no se ha cumplido en los últimos años en España.
- Vacas y novillas de recría subdivididas en 9 clases que representan un año de edad. La edad al primer parto es de 24 meses, por lo que la producción de leche se inicia en el tercer año de edad.

La matriz que describe la transición de genes entre una clase de animales a otra, tanto por envejecimiento como por reproducción, ha sido definida siguiendo los pasos de Groen (1990), utilizando una estructura de edades acorde con la distribución de las vacas en control lechero en España. La tasa de descuento utilizada ha sido del 5% acumulada sobre un periodo de 25 años.

Los vectores de incidencia de cada carácter (tabla 3) han sido definidos teniendo en cuenta el número proporcional de animales y la expresión del carácter en cada clase de edad dentro de cada sexo. Las expresiones de descuento han sido calculadas considerando como en Groen (1990) que las cuatro vías de progreso genético, padre-hijo, padre-hija, madre-hijo y madre-hija aportan el 43, 18, 33 y el 6% del progreso total, respectivamente.

Tabla 3 : Vectores de incidencia utilizados en los cálculos de las expresiones de descuento para los caracteres kg. de leche, volumen de leche, kg. de grasa, kg. de proteina, recuento de células somáticas en primera lactación, longitud de la vida productiva y peso adulto.

	Clases de edad y sexo																
	Sementales									Vacas y novillas de reemplazo							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Leche	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0,23	0.19	0.13	0,09	0.08	0.04
Volumen de leche	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0,24	0,23	0.19	0,13	0.09	0,08	0,04
Grasa	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0,23	0.19	0.13	0,09	0.08	0.04
Proteina	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0,24	0,23	0.19	0.13	0.09	0.08	0.04
Longitud de vida productiva	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0,21	0,20	0.19	0.12	0,11	0.10	0.07

III.5. Resultados y discusión

La tabla 4 muestra los pesos económicos de los caracteres: volumen de leche, grasa y proteina según los parámetros económicos y niveles de producción en las dos zonas del estudio, considerando dos condiciones de producción, una de libre mercado y otra con las limitaciones de la cuota láctea.

La tabla 5 muestra el peso económico de la longevidad según los parámetros

productivos y de mercado en los cuales se envuelven las ganaderías de Alta Moraña y del País Vasco.

Tabla 4 Pesos económicos de los componentes de la producción de leche en el País Vasco y Alta Moraña, considerando dos condiciones de producción: una de libre mercado y otra bajo las limitaciones de la cuota (ptas.kg^{-1}.vaca^{-1}.año^{-1})

	Situación de libre mercado			Situación de cuota		
	Alta Moraña	País Vasco	Variación (%)	Alta Moraña	País Vasco	Variación (%)
Volumen Leche	7,05	7,5	- 6,0	2,02	3,2	- 36,9
Grasa	323,19	258,4	+ 25,1	46,75	25,5	+ 83,3
Proteína	618,92	471,2	+ 31,4	613,89	467,0	+ 31,5

Tabla 5: Peso económico de la longevidad en las condiciones de producción y de mercado de Alta Moraña y el País Vasco (ptas.dia^{-1}.vaca^{-1}.año^{-1}).

	Alta Moraña	País Vasco	Variación (%)
Longevidad	36,46	16,2	+ 125,1

En situación de libre mercado, la mejora del potencial genético medio en un kg. de leche con un 3,8% de grasa y un 3% de proteína en las ganaderías del País Vasco y de Alta Moraña genera un beneficio marginal de 30,9 y 37,4 ptas.·vaca^{-1}·año^{-1}, respectivamente.

Los principales componentes de la leche (grasa y proteína) muestran pesos económicos bastante distintos que reflejan el sistema de pago de la leche en cada zona, tanto por calidad como por cantidad. El volumen de leche muestra un peso económico positivo en las dos zonas en contra de lo que se puede observar en la mayoria de los trabajos realizados en países de nuestro entorno (Wilmink, 1988, Groen, 1989c, Bekman y Van Arendonk, 1993, Colleau *et al.*,1994 y Pieters *et al.*, 1997).

A pesar de que el ingreso marginal por producir un kilo de volumen de leche en el caso de Alta Moraña es superior, los altos costes de alimentación registrados en dicha zona reducen el peso económico del volumen de leche un 6%. En situación de cuota, el mayor nivel de producción con respecto a la situación del Pais Vasco, eleva esta reducción hasta el 36,9%.

La mayor bonificación, dada por los porcentajes de grasa y proteína por parte de la central lechera de la Cooperativa de Alta Moraña, aumenta los pesos económicos de los caracteres producción de grasa y proteina, a pesar de los altos costes de alimentación y los mayores niveles de producción en los rebaños pertenecientes a la Cooperativa.

El alto coste de alimentación y, por lo tanto el mayor coste de recría, así como una vida productiva más corta de las vacas en los rebaños de Alta Moraña con respecto a los rebaños considerados en el estudio del País Vasco, explican el considerable incremento en el peso económico de la longevidad.

La tabla 6 muestra las expresiones acumulativas de descuento relativas a cada carácter y en cada vía de selección. Estas estimas se realizaron tendiendo en cuenta las diferencias tanto en formas como en tiempo de expresión entre los caracteres de producción y la longevidad.

Las expresiones de descuento son relativas a los animales de la población de base y expresan el efecto relativo de la mejora genética por una unidad en cada uno de los distintos caracteres con respecto al mismo hecho en la población base. Se expresan en animal·año por animal de la población base.

Las diferencias entre caracteres son debidas exclusivamente a las diferencias en las incidencias de expresión de las distintas clases de edad de cada sexo en la población. Mientras que las diferencias entre vías de selección son originadas en parte por los distintos momentos de introducción de genes, y por otra parte, por el tiempo que tarden en expresarse por primera vez los genes en la población.

Tabla 6. Expresiones acumulativas de descuento de los distintos caracteres aplicadas a cada vía de selección (animal·año por animal de base).

	Las distintas vías de progreso genético			
Caracteres	padre-hijo	padre-hija	madre-hijo	madre-hija
Volumen de leche	0,196	0,597	0,642	0,661
Grasa	0,196	0,597	0,642	0,661
Proteína	0,196	0,597	0,642	0,661
Longevidad	0,188	0,580	0,624	0,643

Teniendo en cuenta la participación de cada vía de selección en el progreso genético total indicadas anteriormente, las expresiones acumulativas de descuento aplicadas a los caracteres de producción de leche y a la longevidad son de 0,443 y 0,429, respectivamente.

Por lo que si queremos seleccionar por los tres componentes de producción de leche y por longevidad en las dos regiones en situación de cuota, el genotipo agregado en cada una de ellas quedaría definido como:

País Vasco:

$$H = 1{,}41 \cdot V.\ leche + 11{,}29 \cdot Grasa + 206{,}88 \cdot Proteína + 6{,}95\ Longevidad$$

Alta Moraña:

$$H = 0{,}89 \cdot V.\ leche + 20{,}71 \cdot Grasa + 271{,}95 \cdot Proteína + 15{,}64\ Longevidad$$

Donde,

H : El valor genetico global verdadero de un animal expresado en pesetas
V. leche : El valor genetico verdadero de un animal por volumen de leche (kg)
Grasa : El valor genetico verdadero de un animal por producción de grasa (kg)
Proteína : El valor genetico verdadero de un animal por producción de proteina (kg)
Proteína : El valor genetico verdadero de un animal por longevidad (días)

III.5. Conclusión

Los pesos económicos de los tres componentes de la producción de leche en las dos regiones han sido distintos tanto en situación de cuota como en un mercado libre. Las diferencias existentes en los costes marginales de producción (principalmente los costes de alimentación) y en el sistema de pago de leche en cada zona son en gran medida las causas de dichas diferencias.

El beneficio marginal debido a la mejora genetica de la longevidad de los animales es netamente superior en las ganaderías de Alta Moraña en comparación con el País Vasco dado

los altos costes de recría y los bajos ingresos por venta de animales de desecho por un lado por otro.

La introducción del sistema de cuota láctea reduce el peso económico de los tres componentes de la producción de leche, sobre todo el de grasa y del volumen de leche.

Las variaciones en las perspectivas de producción, en el nivel productivo y en las condiciones de mercado afectan sensiblemente los pesos económicos de los caracteres estudiados. Por lo que se genera distintas definiciones del genotipo agregado correspondiente a cada situación.

IV. Conclusión

Definir correctamente el objetivo de un programa de mejora genetica requiere una estimación exacta y real de la importancia económica de cada uno de los caracteres a seleccionar. Dicha estimación nos orienta, en acuerdo con las circunstancias de producción y de mercado, sobre los niveles óptimos de mejora a realizar en cada carácter con el fin de maximizar la rentabilidad económica de los sistemas de producción.

Una estimación correcta y fiable de los pesos económicos requiere una base teórica sólida, una función de beneficio apropiada de los aspectos productivos, económicos y sociales del sistema de producción, una situación base definida a partir de datos reales y unas suposiciones cercanas a las circunstancias futuras de producción. Sin embargo, a nivel practico, cumplir todas estas exigencias queda todavía como un desafio a pesar de que la definición de los objetivos de mejora, sobre todo en vacuno de leche que es el sector que más de cerca hemos estudiado, está cada vez de mas interés.

Agradecimientos

Los autores agradecen a IKT y a la Cooperativa de Alta Moraña su colaboración en la realización de este estudio.

Referencias bibliográficas

Allaire, F. R., and J. P. Gibson. 1992. Genetic value of herd life adjusted for milk production. *J. Dairy Sci. 75: 1349.*

Andrus, D. F. and L. D. McGilliard. 1975. Selection of dairy cattle for Overall excellence. *J. Dairy Sci. 58: 1876.*

Bekman, H. and J. A. M. Van Arendonk, 1993. Derivation of economic values for veal, beef and milk production traits using profit equations. *Livest. Prod. Sci.*, 34: 35.

Blasco, A. y Muñoz, A., 1994. Algunos aspectos económicos de los programas de mejora genética. *VII Reunión sobre Mejora Genetica Animal. Lleida.*

Brandts, A., Canavesi, F., and Cassandro, M., 1996. Economic weights for longevity in the Italian Holstein breed. *In Proc, of the Open Session of the INTERBULL Ann. Meeting, Veldhoven, The Netherlands. Bulletin Nº 14, pp.76.*

Brascamp, E. W., 1975. Model Calculations concerning economic optimization of A.I.-breeding with cattle. *PhD Thesis, Dept. Animal Breeding, Wageningen Agricultural University, Wageningen.*

Brascamp, E.W., 1978. Methods on economic optimization of animal brreding plans. *Report B- 134, Research Institute for animal Husbandry "Schoonoord", Zeit.*

Brascamp, E.W., 1984. Selection indices with constraints. *Anim. Breed. Abstr., 52:645.*

Brascamp, E. W., Smith, C., Guy, D. R., 1985. Derivation of economic weights from profit equations. *Anim. Prod. 40, 175.*

Colleau, J.J.; Regaldo, D. and P.L. Gastinel. 1994. Adaptation de l'index français de sélection laitière (INEL) au contexte des quotas. INRA. *Prod. Anim. 7(3); 151.*

Charffeddine, N. 1998. Selección por mérito económico global en el ganado vacuno frisón en España. *Tesis doctoral, dpto. Producción animal, E.T.S.I. Agrónomos Madrid.*

Charffeddine, N.; Alenda, R.; Carabaño, M.J.; Béjar, F. 1995. La selección por el mérito total en el ganado vacuno de leche. *Frisona Española Nº 89 p.43*

Dekkers, J.C.M. 1993. Theoretical basis for genetic parameters of Herd life and effects on

response to selection. *J. Dairy. Sci. 76:1433.*

Dekkers, J. C. M., 1994. Optimal Breeding strategies for calving ease. *J. Dairy Sci. 77: 3441.*

Dijkhuizen, A.A., 1983. Economische aspecten van ziekten en ziektesbestrijding bij melkvee (with English Summary). *Ph.D. Thesis, Veterinaire Faculteit, Utrecht.*

Elsen, J. M., 1980. Diffusion du progrès génétique dan les populations avec générations imbriquées: quelques propriétés d'un modèle de prévision. *Ann. Génét. Sél. Anim., 12: 49.*

France, J. and Thornley, J. H. M. 1984. A quantitative approach to problems in agriculture and related sciences. *Butterworths, London.*

Gibson, J.P., 1989. Economic weights and index selection of milk production traits when multiple production quotas apply. *Anim. Prod., 49: 171.*

Gibson, J.P., 1992. The design and economics of animal breeding strategies. *First Presented as Nordic Graduate course, Greve, Denmark.*

Groen, Ab. F., 1988. Derivation of economic values in cattle breeding. A model at farm level . *Agric. Systems 27: 195.*

Groen, Ab. F., 1989a. Cattle breeding goals and production circumstances. *PhD Thesis, Department of Farm Management and Department of Animal Breeding, Wageningen Agricultural University, Wageningen, The Netherlands.*

Groen, Ab. F., 1989b. Economic values in cattle breeding. I. Influence of production circumstances in situation without output limitations. *Livest. Prod. Sci., 22: 1.*

Groen, Ab. F., 1989c. Economic values in cattle Breeding. II. Influence of production circumstances in situations with output limitations. *Livest. Prod. Sci., 22: 17.*

Groen, Ab. F. 1990. Influences of production circumstances on the economic revenue of cattle breeding programmes. *Anim. Prod., 469.*

Groen, Ab. F. 1992. The derivation of economic values. *Workshop on advanced biometrical methods in animal breeding. October 9th - 15th, Flawill, Switzerland.*

Groen, Ab. F. and Ruyter, T. P. L. 1990. Derivation of economic values of milk production traits: a literature review. *Proceeding 4th world congress on genetics applied to livestock production, Edimburgh. XIV: 191.*

Groen, Ab. F. and J. A. M. Van Arendonk, 1997. Breeding Programmes. *Lecture notes for E250-210, Department of Animal Breeding and Genetics, Wageningen Agricultural University.*

Groen, Ab. F., Van Arendonk, J. A. M. and Ant. R. Vollema. 1997a. On the use of opportunity costs in deriving the economic value of herd life. *Animal Science (Submitted).*

Groen, Ab. F.; Steine, T.; Colleau, J.J.; Pedersen, J., Pribyl, J. and N. Reinsch, 1997b. Economic values in dairy cattle breeding, with special reference to functional traits. *Report of an EAAP-working grou. Livest. Prod. Sci. 49; p. 1.*

Harris, D. L., 1970. Breeding for efficiency in livestock production: defining the economic obectives. *J. Anim. Sci. 30, 860.*

Harris, B.L.; Freeman, A.E. 1993. Economic weights for milk yield traits and herd life under various economic conditions and production quotas. *J. Dairy Sci. 76:868.*

Hazel, L. N., 1943. The genetic basis for constructing selection indexes. *Genetics, 28: 476.*

Hill, W.G., 1974. Prediction and evaluation of reponse to selection with overlapping generations. *Anim. Pord. 18: 117.*

James, A. D. and Ellis, P.R., 1979. The evaluation of production and economic effects of disease. *Proceedings 2nd International Sympo. on Veterinary Epidemiology and Economics. Canberra, pp. 363.*

Kempthorne, O. and A.W. Nordskog, 1959. Restricted selection indices. *Biometrics, 15: 10.*

Kolstad, B. and Dekkers, J., 1994. Incorporation of Mastitis into the breeding goal. *Report to the Genetic Evaluation Board of Canada, 9pp.*

Moav, R., 1973. Economic evaluation of genetic differences. *In, Agricultural Genetics. (R. Moav, ed.) wiley, New York. pp. 319.*

Morris, C. A., and Wilton, J. W., 1977. The influence of body size on the economic efficiency of cows: a review. *Anim. Breed. Abstr., 45: 139.*

Pieters, T. and A. Bagnato. 1996. Genetic parameters and economic values for longevity in the Italian Holstein breed. *Int. Workshop on genetic improvement of functional traits in cattle. Gembloux, Belguim. Interbull, Nº 12.*

Pieters, T., Canavesi, F., Cassandro, M, Dadati, E., and J. A. M. Van Arendonk, 1997. Consequences of differences in pricing systems between regions on economic values and revenues, of a natonal dairy cattle breeding scheme in Italy. *Livest. Prod. Sci. 49:23.*

Reents, R. 1997. Somatic cell count as an indicator trait for genetic selection against mastitis susceptibility. *48th Annual Meeting of EAAP, August 25-28, 1997, Vienna, Austria.*

Reneau, J. K., and Packard, V.S., 1991. Monitoring Mastitis, Milk quality and economic losses in dairy fields. *Dairy, Food and Environmental Sanitation, Vol. 11(1), p. 4.*

Rogers, G.W., Van Arendonk, J.A.M.; McDaniel, B.T. 1988. Influence of involutary culling on optimum culling rates and annualized net revenues. *J. Dairy Sci. 71:3463.*

Rogers, G. W., McDaniel, B. T., Dentine, M. R., Funk, D. A. 1989. Genetic correlation between survival and linear type traits measured in first lactation. J. Dairy Sci. 72:523.

Schepers, J. A. and Dijkhuizen, A. A., 1991. The economics of mastitis and mastitis control in dairy cattle: a critical analysis of estimates published since 1970. *Prev. Vet. Med. 10, 213..*

Smith, C.; James, J. W. and E.W. Brascamp. 1986. On the derivation of economic weight in livestock improvement. *Anim. Prod. 43:545.*

Steverink, M. H.A., Groen, Ab. F., and Berentsen, P.B.M., 1994. The influence of environmental policies for dairy farms on dairy cattle breeding goals. *Livest. Prod. Sci., 40:251.*

Tess, M. W.; Bennett, G.L. and Dickerson, G. E., 1983. Simulation of genetic changes in life cycle efficiency of pork production. I.A. bio-economic model. *J. Anim. Sci. 56: 336.*

Van Arendonk, J. A. M., 1985. Studies on the replacement policies in dairy cattle. II Optimum policy and influence of changes in production and prices. *Livest. Prod. Sci., 14:101.*

Van Arendonk, J. A. M. 1991. Use of profit equations to determine relative econmic value od dairy cattle herd life and production from yield field data. *J. Dairy Sci. 74:1101.*

Van Arendonk, J. A. M. and Brascamp, E.W. 1990. Economic considerations in dairy cattle Breeding. *4th World congress on genetics applied to livestock production, edimburgh. XIV: p. 78.*

Van Arendonk, J. A.M., Wilmink, J.B.M., and Dijkhuizen, A.A., 1985. Consequences of a restriction of the herd milk production for the selection on milk, fat and protein. *Adapting EEC cattle breeding programmes to market realities. Krausslich H., and A. Lutterback, Editors. Commission of the European Communities, Brussels, pp. 211.*

Van Raden, P. M. and G. R. Wiggans. 1995. Productive life evaluations: calculation, Accuracy, and Economic value. *J. Dairy Sci. 78:631.*

Veerkamp, R.F., 1996. Liveweight and feed intake in dairy cattle breeding. *Proceedings of International Workshop on Genetic Improvement of Functional Traits in Cattle. Gembloux, Belgium. January. Interbul Bullet. 12:173.*

Visscher, P. M., Bowman, P.J. and Goddard, M.E., 1994. Breeding objectives for pasture based dairy production systems. *Livest. Prod. Sci., 40: 123.*

Weigel, D.J., Cassell, B.G., Hoeschele, I.; Pearson, R.E. 1995. Multiple-trait prediction of

transmitting abilities for herd life and estimation of economic weights using relative net income adjusted for opportunity cost. *J. Dairy Sci. 78:639.*

Weller, J. I., 1994. Economic Aspects of Animal Breeding. *Chapman &Hall, London, Uk. Pp. 244.*

Weller, J. I., Pasternak, H. and A. F. Groen, 1996. Selection indices for non-linear breeding objectives, selection for optima. *Proc. Int. Workshop on Genetic Improvement of functional traits in cattle. Gembloux, Belgium, January 1996. p206.*

Wilmink, J. B. M., 1988. Selection on fat and protein to maximize profit in dairy herds. *Livest. Prod. Sci., 20: 299.*

SELECCIÓN PARA UN ÓPTIMO APLICACIÓN AL TAMAÑO DE LA CAMADA EN OVINO

Magali SAN CRISTOBAL-GAUDY[1], Loys BODIN[2],
Jean-Michel ELSEN[2], Claude CHEVALET[1]

[1] INRA - Laboratoire de Génétique Cellulaire, BP 27, 31326 Castanet-Tolosan France.
[2]INRA - Station Génétique des Animaux, BP 27, 31326 Castanet-Tolosan France.

Introducción

El propósito principal de la mejora genética en producción animal es aumentar (o reducir) la expresión de los caracteres para incrementar directamente o indirectamente el rendimiento económico de las ganaderías o de otros segmentos de producción. Por eso, en producción animal, la mayoría de los esquemas de selección son direccionales (aumento de la cantidad y la calidad de la leche, incremento del números de huevos; aumento de la resistencia a enfermedades etc...). Sin embargo, la homogeneidad de la producción es un factor importante de la eficiencia económica (por ejemplo, el beneficio de un ganadero depende en parte de su capacidad de mandar lotes de animales homogéneos al matadero); además, muchos caracteres presentan un óptimo que proporciona una rentabilidad máxima. Así, existe un peso óptimo para los canales de pollo, corderos o cerdos, existe también un ph óptimo de la carne de cerdo, y reducir la variabilidad del diámetro de la fibra de lana presenta un verdadero interés económico. En producción de ovino, cuando los recursos alimenticios no son demasiado escasos, las ovejas que proporcionan mayor beneficio son las que paren dos corderos puesto que un solo cordero reduce mucho el margen por oveja y criar corderos nacidos triples o más supone demasiado incremento de gastos debido al exceso de mortalidad de los corderos excesivamente ligeros al nacer, a la necesidad de recurrir a la lactancia artificial para criar los corderos supernumerarios y a los efectos negativos del modo de nacimiento sobre el crecimiento y la duración del engorde.

La estabilización de los fenotipos alrededor de una expresión dominante es conocida desde hace tiempo como decisiva para la evolución de las especies y varios autores han propuesto modelos para explicar esa estabilización natural. Algunos como Falconer (1989) hacen intervenir directa o indirectamente genes de "fitness", otros como Rendel (1979) recurren a genes que actúan directamente sobre la variabilidad del carácter. Cualquiera que sea su origen, la estabilización depende siempre del medio, haciendo imprescindible discriminar por una parte, la estabilización de un carácter en un medio cpncreto (normalising selection) y, por otra parte, la capacidad de mantener un fenotipo constante, o por lo menos con baja variabilidad, en medios variables (canalising selection).

Númerossos experimentos han sido realizados para entender los mecanismos subyacentes a la canalización natural o para medir las posibilidades de canalizar un carácter alrededor de un cierto valor. Todos se refieren a animales de laboratorio como drosophila, tribolium o incluso ratones. Casi siempre la selección era del tipo normalizante (eliminación de los extremos observados en un solo medio) y era realizada globalmente (Peinero, 1992; Campo y Gill, 1993), intra familia (Kaufman *et al.*, 1977) o entre familias (Cardin y Minvielle, 1986).

La selección canalizante ha sido experimentada por Waddington (1960), por Sheiner y Lyman (1991) y Gibson y Bradley (1974) quienes eliminaban animales extremos de una población criada en un medio variable (temperatura variable entre 20 y 29°C).

Con esos experimentos se llega a las siguientes conclusiones: 1) la selección estabilizante es eficaz y permite una fuerte reducción de la varianza fenotípica; 2) las estimaciones de heredabilidad hechas durante y al final de los experimentos de selección indican que generalmente la varianza genética del carácter seleccionado disminuye; 3) En muchos casos se ha podido mostrar que la varianza residual o la sensibilidad de los individuos a fluctuaciones del medio ha sido reducida con la selección. Sin embargo, en esos diversos experimentos no hubo análisis estadísticos de parámetros genéticos que necesitan utilizar modelos en los cuales los términos de error no siguen una distribución con varianza homogénea.

Foulley *et al.* (1990, 1992, 1994), San Cristobal *et al.* (1993), Foulley y Quaas (1995), Robert *et al.* (1995a,b), han desarrollado métodos estadísticos para tener en cuenta la heterogeneidad de varianzas en mejora animal. Recientemente, San Cristóbal *et al.* (1998a) han extendido esos modelos para introducir efectos genéticos como factores de heterogeneidad, para estimar nuevos valores genéticos y para estudiar la eficacia de una selección canalizante basada en esos modelos. San Cristobal *et al.* (1998b) han adaptados esos modelos al análisis de parámetros genéticos de caracteres discretos. Aquí se presenta una síntesis corta de esos dos últimos artículos.

El modelo propuesto

El modelo propuesto por SanCristobal *et al.* (1998a) corresponde a la modelización de las observaciones de Rendel y Sheldon (1960). Ellos demuestran que la selección canalizanteorienta efectivamente el promedio de la población hacia un óptimo y que la varianza residual es reducida. Algunos individuos aparecen menos sensibles que otros a las variaciones del medio, estando esa particularidad bajo control genético puesto que responde a la selección. Como en Wagner *et al.* (1997), se puede expresar el efecto de un locus sobre la varianza residual con un factor multiplicativo de esa varianza ligado a ese locus y se puede extender ese modelo a un locus de un modelo poligénico o infinitesimal. Esos modelos se justifican también perfectamente para caracteres discretos como el tamaño de la camada si consideramos previamente como Gianola y Foulley (1983) una variable subyacente de distribución normal y un conjunto de umbrales que definen las clases de observación.

Modelo de base

Para un carácter continuo, el fenotipo y_j de un animal con medidas repetidas ($y=\{y_1, \ldots, y_j, \ldots, y_n\}$) es función de dos valores genéticos u y v que contribuyen a la media y al logaritmo de la varianza ambiental según el modelo siguiente :

$$y_j = \mu + u + e^{\left(\frac{\eta + v}{2}\right)} \varepsilon_j \qquad j = 1, \ldots, n \qquad (1)$$

μ media de la población,

η media del logaritmo de la varianza,

ε_j variables normales independientes con distribución $\sim \mathcal{N}(0, 1)$

De acuerdo al modelo infinitesimal de la genética cuantitativa las variables u y v son

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}_2 \left[0, \begin{pmatrix} \sigma_u^2 & r\sigma_u\sigma_v \\ r\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \right]$$

σ_u^2 y σ_v^2 son las varianzas genéticas aditivas y r es el coeficiente de correlación entre u y v.

El promedio y la varianza (varianza fenotípica) de la variable Y son :

$$E(Y) = \mu \tag{2}$$

$$Var(Y) = \sigma_u^2 + e^{\left(\frac{\eta + \sigma_v^2}{2}\right)} = \sigma_p^2 \tag{3}$$

Si los animales están medidos en distintos ambientes, podemos escribir un modelo más general

$$y = x'\beta + z'u + \exp\left[\, 1/2\, (p'\beta^* + q'v) \,\right] \varepsilon \tag{4}$$

β y β^* efectos fijos,
u y v efectos genéticos aleatorios, y
los parámetros de posición son entonces β y u
y los parámetros de dispersión son β^* y v
(x', z') y (p', q') matrices de incidencia respectivas

en ese caso las variables u y v son

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}_{2q} \left[0, \begin{pmatrix} \sigma_u^2 & r\sigma_u\sigma_v \\ r\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \otimes A \right] \tag{5}$$

con A la matriz de parentesco de los animales incluidos en el análisis.

Adaptación para caracteres discontinuos

Se supone que los datos de K categorías discretas que se observan en una clase i (nivel de un efecto fijo o aleatorio) provienen de poblaciones de variables aleatorias

$$N_i = (N_{i1}, \ldots, N_{ik}, \ldots, N_{iK}) \sim \mathcal{M}\ (n_{i+}\ ;\ (\Pi_{i1}, \ldots, \Pi_{ik}, \ldots, \Pi_{iK}))$$

donde n_{i+} es el número total de observaciones de la clase i ; $n_{i+} = \Sigma_k\, n_{ik}$

y Π_{ik} es la probabilidad de que una variable inobservable $Y_i \sim \mathcal{N}\ (\mu_i\ ,\ \sigma_i^2)$ esté entre dos umbrales τ_{k-1} y τ_k :

$$\Pi_{ik} = \Phi\left((\tau_k - \mu_i)/\sigma_i\right) - \Phi\left((\tau_{k-1} - \mu_i)/\sigma_i\right) \tag{6}$$

Objetivos de selección

En el caso continuo, si queremos que la media de la siguiente generación de una población Y_d se acerque a un cierto óptimo y_0 con una variabilidad mínima alrededor de ese óptimo, podemos fijar a priori el objetivo siguiente :

$$\min\ \{\ E[(Y_d - y_0)^2]\ \}$$

Para un individuo con diferentes observaciones y hechas sobre él o su descendencia, un criterio de selección sencillo puede escribirse :

$$I(y) = (\bar{y} - y_0)^2 + S^2 y \qquad (7)$$

En el caso discontinuo, la canalización se efectúa sobre la escala subyacente, y el objetivo es función de un cierto óptimo fijado a priori

$$\Pi_o = (\Pi_{1,o}, ..., \Pi_{k,o}, ..., \Pi_{K,o})$$

Por ejemplo, pesos económicos relativos a cada categoría. En el caso del tamaño de la camada en ovino, si queremos que las ovejas paran exclusivamente dos corderos, el óptimo será : $\Pi_o = (0, 1, 0, ..., 0)$.

Hemos estudiado así, un índice muy elemental y que parece natural sobre la escala observada : la proporción de partos dobles

$$I(N) = N_2/n \qquad (8)$$

Análisis del tamaño de la camada en raza Lacaune

Los datos provienen del control oficial de producción realizado en 57 ganaderías de la raza Lacaune que participan en el esquema de selección para la prolificidad. Son 11 723 primeros partos realizados en 2 épocas y a 4 edades distintas de las hijas de 157 padres. Como son padres en testaje por descendencia, cada padre tiene por lo menos 30 hijas.

La prolificidad media es 1.72 y la distribución de los partos es la siguiente :

Tamaño de la camada	1	2	3	4	5+
%	41.1	47.5	9.8	1.5	0.1

El análisis que hemos hecho consistió en modificar poco a poco un modelo clásico con umbrales hasta poder estimar los componentes de varianza y los valores genéticos de un modelo mixto heteroscedástico con umbrales.

A) Primero hemos ajustado un modelo clásico considerando los efectos año, rebaño, época y edad como efectos fijos y el efecto padre (u) como aleatorio (incluyendo la matriz de parentesco entre los padres).

$$\mu_{ijklm} = año_i + rebaño_j + estación_k + edad_l + u_m$$
$$\sigma^2_{ijklm} = \sigma^2_e$$

El proceso de estimación de Fouley y Manfredi (199?), da una estimación de la componente padre de $\sigma^2_u = 0.109$ y una heredabilidad del tamaño de la camada de $h^2_u = 0.39$

B) Después hemos ajustado un modelo heteroscedástico considerando todos los efectos como fijos para ver cuáles eran los efectos significativos sobre la variabilidad. Por eso se analizó el cociente entre los logaritmos de la verosimilitud del modelo de base

$$\mu_{ijklm} = año_i + rebaño_j + estación_k + edad_l + u_m$$
$$\log \sigma^2_{ijklm} = \eta$$

y el modelo que incluye el factor considerado (por ejemplo factor año)

$$\mu_{ijklm} = año_i + rebaño_j + estación_k + edad_l + u_m$$
$$\log \sigma^2_{ijklm} = \eta + año_i$$

Ese cociente sigue asintóticamente una distribución de $\aleph^2$ con grados de libertad iguales a la diferencia entre el números de parámetros estimados en cada modelo. El cuadro siguiente presenta la probabilidad de ese test para los distintos efectos y indica que solo el factor padre (considerado como efecto fijo) tiene un efecto significativo sobre el logaritmo de la varianza.

Efecto	Valor del test	gdl	Prob.
Año	19.82	10	0.031
Rebaño	58.52	56	0.38
Edad	3.8	2	0.15
Padre	235.06	156	0.00004

C) Finalmente, se ha ajustado un modelo mixto heteroscedástico con umbrales para estimar los valores genéticos y los componentes de varianza. En ese ultimo modelo, se ha añadido la matriz de parentesco entre los padres y, los efectos fijos fueron los que previamente habían sido detectados significativos.

$$\mu_{ijklm} = año_i + rebaño_j + estación_k + edad_l + u_m$$
$$\log \sigma^2_{ijklm} = \eta + v_m$$

donde u_m y v_m representan los efectos padre respectivamente sobre la media y el logaritmo de la varianza del carácter.

La estimación de los componentes $\sigma^2_u = 0.64$ y $\sigma^2_v = 0.25$ permite calcular una heredabilidad de $h^2_u = 0.36$. Ese valor es ligeramente inferior a lo que da el modelo homoscedástico con umbrales e informa sobre el sesgado que puede existir en la estimación de la heredabilidad de la media si no se considera el efecto genético sobre la heterogeneidad de las varianzas. Las estimaciones son relativamente altas lo que indica una buena posibilidad para canalizar ese carácter alrededor de un óptimo. En la figura 1, se muestra una representación de $\hat{u}$ y $\hat{v}$ para los 157 machos que nos permite apreciar su capacidad de producir hijas con alta o baja prolificidad y con alta o baja variabilidad.

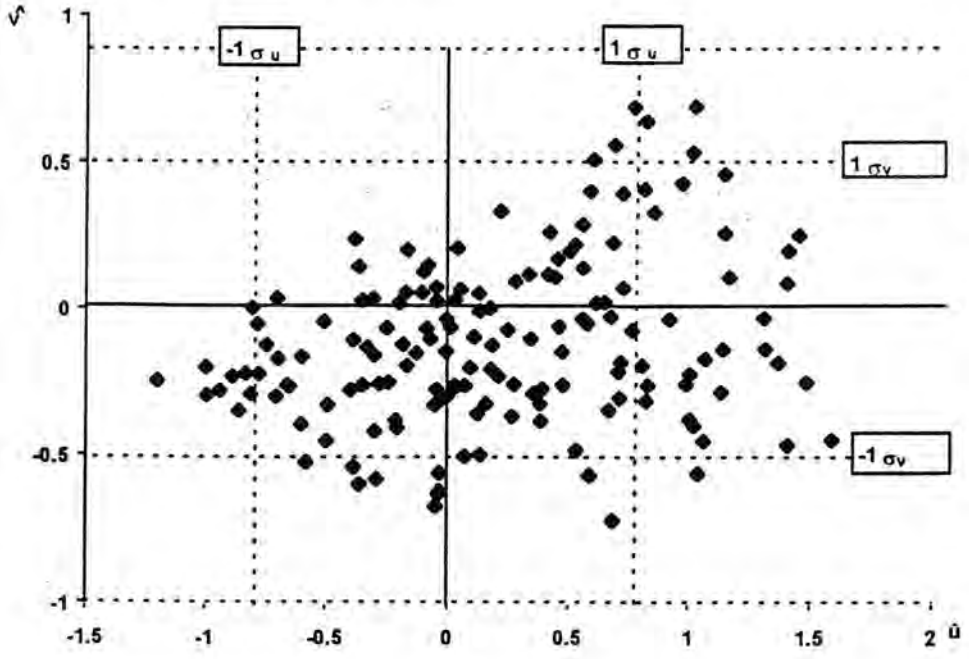


Figure 1 : Plot de los valores geneticos $\hat{u}$ y $\hat{v}$ de los 157 machos *Lacaune*.

Simulaciones

Hemos hecho varias simulaciones para estudiar la repuesta de la media, de la varianza y del porcentaje de cada tipo de partos, cuando se realiza durante varias generaciones una selección basada sobre un índice para una variable continua (7) y cuando se considera un índice muy sencillo pero adaptado a la variable discreta (8). Esas simulaciones nos han permitido también, ver cómo reaccionaba la selección según varias combinaciones de los parámetros h^2_u, η y σ^2_v.

Modelización de la población

Los valores genéticos u y v de p machos y de m hembras fueron representadas por variables aleatorias que siguen una distribución conjunta (5), suponiendo A : matriz de identidad (ausencia de relaciones de parentesco) y r igual a 0.

a) Cada uno de los p machos está apareado al azar con n hembras

b) Cada pareja (i, j) produce una hija cuyo resultado en su primer parto está sobre la escala subyacente :

$$y_{ij} = \mu + 1/2\ (u_i + v_j) + \alpha_{ij} + \exp\ [(\eta + 1/2\ (u_i + v_j) + \alpha^*_{ij})/2]\ \varepsilon_{ij}$$

donde el término de error ε_{ij}, y los términos α_{ij} y α^*_{ij} que representan los efectos de meiosis siguen distribuciones normales $\mathcal{N}$ (0, 1), $\mathcal{N}$ (0, $\sigma^2_u/2$) y $\mathcal{N}$ (0, $\sigma^2_v/2$) respectivamente.

Para las simulaciones de la variable discreta se pasa de esa escala continua subyacente a la escala observada mediante 4 umbrales que definen 5 categorías. Esos parámetros corresponden a la distribución del tamaño de la camada observada en Lacaune.

c) esos resultados de descendencia permiten estimar un índice (7 o 8) para el macho

d) los mejores s machos se seleccionan. Esos padres 'elite' apareados a una muestra de las hembras de la etapa (a) producen p hijos

e) iterativamente, esos p machos se aparean al azar con las n hembras nacidas en la etapa (b) para producir la nueva generación de hijas de testaje.

Las etapas b, c, d y e se repiten así hasta la décima generación.

Parámetros del modelo

Hemos examinado un esquema que considera en cada generación 500 machos en testaje con 100 hijas cada uno. De esos machos, 100 padres 'elite' se seleccionan por índice para producir cada uno 5 hijos para la generación siguiente. Hemos considerado también dos valores para σ^2_u (0.033 y 0.114) que corresponden a una baja (h^2_u=0.10) o alta (h^2_u=0.35) heredabilidad ; así como dos valores para σ^2_v (0.03 y 0.25) y tres valores para la media inicial (1, 1.8 y 2), teniendo un óptimo μ_0 fijado a 2. Para cada conjunto de parámetros hemos hecho 100 simulaciones y calculado el promedio y la varianza de los resultados.

Resultados de las simulaciones

La figura 2 indica, en el caso de una selección sobre una variable continua, la evolución de la media y de la varianza fenotípica en el transcurso de las generaciones. SE han considerado varias combinaciones de la media de la población al principio de la selección y de las varianzas de los efectos genéticos sobre la media (σ^2_u) y el logaritmo de la varianza ambiental (σ^2_v).

Esas simulaciones muestran que la selección es eficaz : la media se acerca al óptimo definido y la varianza fenotipica se va reduciendo poco a poco. Comparando las distintas combinaciones de parámetros se observa que ;

La media de la población varía más rápidamente cuando σ^2_u es alta. Por otra parte, aunque esa media varía poco con los cambios de σ^2_v , los mayores cambios se observan cuando σ^2_v es baja y σ^2_u alta.

La varianza de la población disminuye en el transcurso de la selección y más rápidamente cuando σ^2_v es alta y σ^2_u baja. Existe así una relación entre el progreso sobre una variable y la variabilidad sobre la otra que traduce el hecho que cuando la presión de selección sobre una variable es mínima, la selección opera más sobre la otra.

Para una combinación de varianzas (σ^2_u y σ^2_v), se observa que la disminución de la varianza fenotípica es mayor cuando la media inicial de la población esta próxima al óptimo μ_o. Eso se debe al hecho de que, cerca del óptimo, los índices varían más a causa de v que de u.

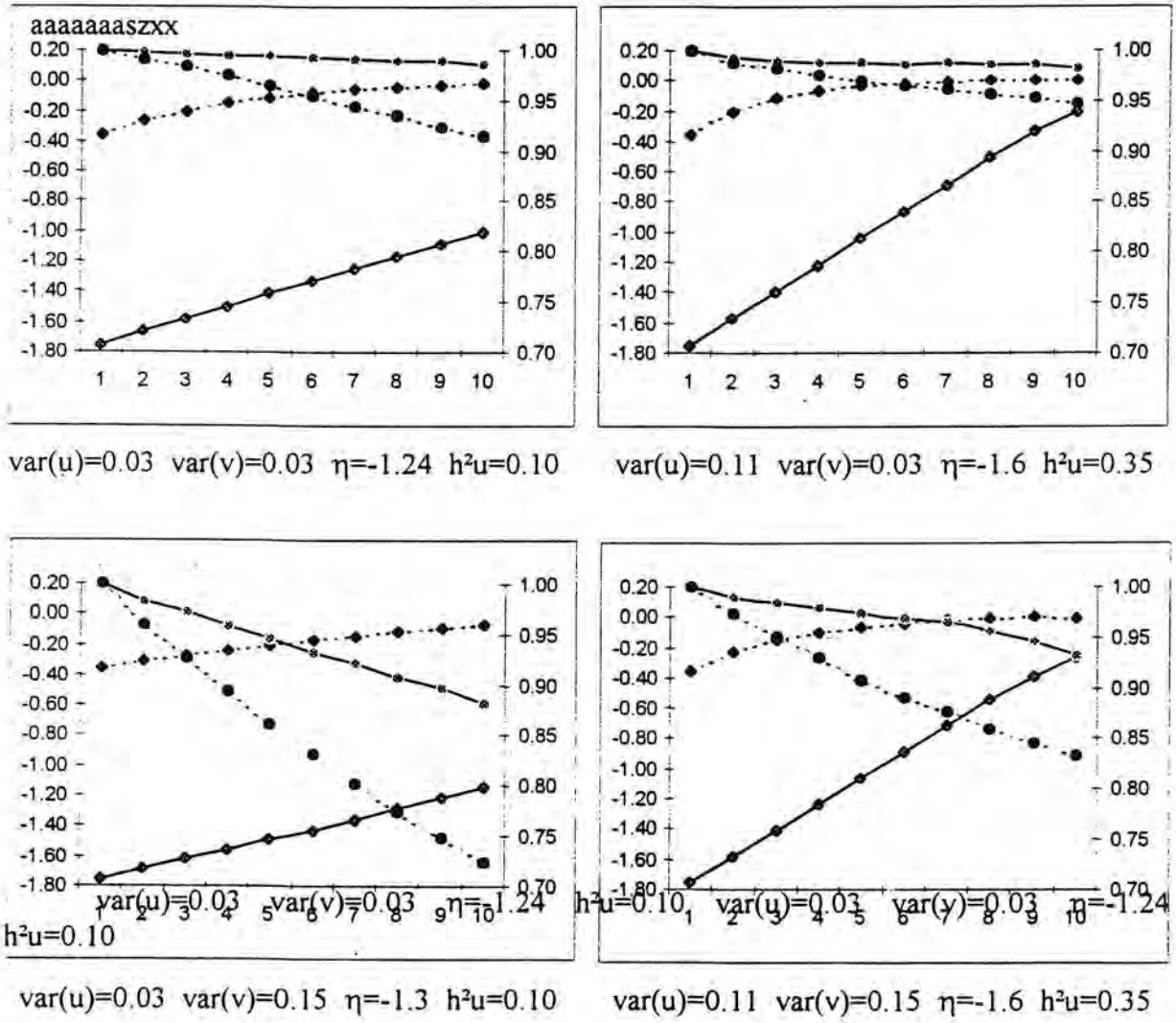


Figura 2 : Evolución de la media (rombos) y de la desviación tiípica (círculos) durante 10 generaciones de selección en dos poblaciones (cerca (linea) o lejos (puntos) del óptimo) y varias combinaciones de σ^2_u y σ^2_v

Las simulaciones para la canalización de una variable discreta muestran que es posible modificar la distribución de los partos acercándose al óptimo. En la figura 3 se puede comparar la evolución del porcentaje de partos simples, dobles, triples y cuádruples o superior cuando se selecciona con un índice sencillo y cuando se selecciona con un índice clásico que no cambia la varianza subyacente. Diez generaciones de selección para un óptimo producen una distribución con más partos dobles que lo que da la selección clásica (a prolificidad media comparable), y con menos partos triples, cuádruples y superiores. Sin embargo, maximizar el porcentaje de partos dobles no impide que el porcentaje de partos triples aumente también durante la primeras generaciones de selección lo que no cuadra con el propósito real de los ganaderos. El porcentaje de dobles es sin duda un criterio demasiado sencillo aunque se observa una disminución de los partos superior a 4 después de la generación 5. Si a largo plazo el objetivo parece adecuado, su definición no es suficiente a corto plazo.

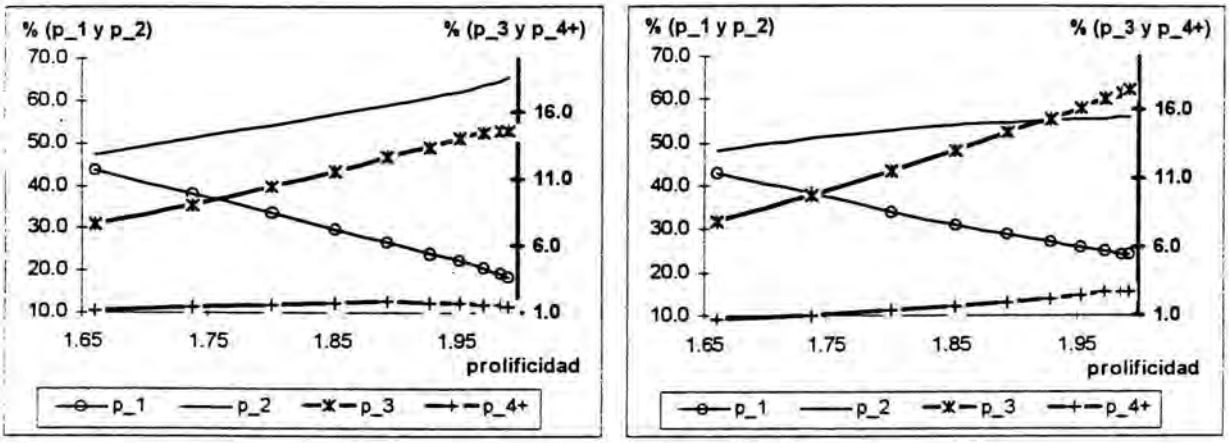


Figura 3 : Evolución de los porcentajes de partos simples, dobles, triples y más cuando se selecciona con un índice sencillo que reduce la varianza y con un índice clásico.

Conclusiones

Las varianzas del tamaño de la camada en raza Lacaune presentan una heterogeneidad debida parcialmente a factores genéticos. La introducción de la matriz de parentesco en los análisis permite estimar los valores de varianza genética aditiva para la media y el logaritmo de la varianza. Con las simulaciones, se puede apreciar que buscar un óptimo reduciendo la varianza fenotípica es posible. Sin embargo, el peso relativo de cada variable (media y varianza) para conseguir el objetivo final depende de la situación en cada generación. Para un objetivo final único, existe un gran número de caminos (combinaciones de los pesos sobre u y v en cada generación) para alcanzarlo. Existe en particular un camino más rápido, y un camino más rentable; desde luego fijarse un objetivo constante (igual al objetivo final) a cada generación no dará el progreso más rápido, ni el más rentable. Estamos desarrollando otros estudios para optimizar la canalización.

Referencias

Campo J.L., Garcia Gill M. 1993 Assortative mating and directional or stabilizing selection for a non-linear function of traits in *Tribolium*. *J. Anim. Breed. Genet.* **110**, 74-80

Cardin S., Minvielle F. 1986 Selection on phenotype variation of pupa weight in *Tribolium Castaneum*. *Can. J. Genet. Cytol.* **28**, 856-861

Falconer D.S. 1989 The problem of environment and selection. *Am Natur* **86**, 293-298

Foulley J.L., Gianola D., San Cristobal M., Im S. 1990 A method for assessing extended and sources of heterogeneity of residual variances in mixed linear models. *J Dairy Sci.* **73**, 1612-1624

Foulley J.L., San Cristobal M., Gianola D., Im S. 1992 Marginal likelihood and Bayesian approaches to the analysis of heterogenous residual variances in mixed linear Gaussian models. *Comput Statis Data Anal.* **13**, 291-305

Foulley J.L., Hébert D., Quass R.L. 1994 Inferences on homogeneity of between family components of variances and covariance among environments in balanced cross-classified designs. *Genet. Sel. Evol.* **26,** 117-136.

Foulley J.L., Quaas R.L. 1995 Heterogenous variances in Gaussian linear mixed models. *Genet. Sel. Evol.* **27**, 211-228.

Gianola y Foulley 1983

Gibson J.B., Bradley B.P. 1974 Stabilising selection in constant and fluctuating environments. *Heredity* **33**, 293-302.

Kaufman P.K., Enfield F.D., Comstock R.E. 1977 Stabilising selection for pupa weight in *Tribolium Castaneum*. *Genetics* **87**, 327-341.

Piniero R. 1992 Selection for canalization at two dorsocentral bristles in *Drosophila Melanogaster*. *J. Heredity* **83**, 49-55.

Rendel J.M. 1979 Canalisation and selection . In: *Quantitative genetic variation.* Academic press. Inc., J.N. Thompson Jr. and J.M. Thoday ed. 139-156.

Rendel J.M., Sheldon B.L. 1960 Selection for canalization of the scute phenotype in *Drosophilia Melanogaster*. *Austr. J. Biol. Sci.* **13**, 36-47.

Robert C., Foulley J.L., Ducrocq V. 1995a Genetic variation of trait measured in several environments. I. Estimation and testing of homogenous genetic and intra-class correlations between environments. *Genet. Sel. Evol.* **27**, 111-123.

Robert C., Foulley J.L., Ducrocq V. 1995b Genetic variation of trait measured in several environments. II. Inference on between-environments homogeneity of intra-class correlations. *Genet. Sel. Evol.* **27**, 125-134.

San Cristobal M., Foulley J.L., Manfredi E. 1993 Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet. Sel. Evol.* **25**, 3-30.

San Cristóbal-Gaudy M., Elsen J.M., Bodin L., Chevalet C. 1998a Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genet. Sel. Evol.* (Submitted).

San Cristóbal-Gaudy M., Elsen J.M., Bodin L., Chevalet C. 1998b Genetic components of litter size variability in sheep. *Genet. Sel. Evol.* (in preparation).

Sheiner S.M., Lyman R.F. 1991 The genetics of phenitypic plasticity. II : Response to selection. *J. Evol. Biol.* **4**, 23-50.

Waddington C.H. 1960 Experiments on canalizing selection. *Genet. Res.* **1**, 140-150.

Wagner G.P., Booth G., Bagheri-Chaichian H. 1997 A population genetic theory of canalization. *Evolution* **51**, 329-347.

Mejora Genética de la Producción de fibras de Cachemira

C. Díaz y A. Moreno

Dpto. de Mejora Genética y Biotecnología. Ctra. de la Coruña km. 7. 28040 Madrid.

Introducción

El sector agropecuario español, como el del resto de los países de la Unión Europea, está bajo la creciente necesidad de diversificar sus actividades dado que las producciones tradicionales están cerca de satisfacer las demandas del mercado. Una alternativa clara, son los productos no alimentarios de alto valor añadido, alternativa que coincide con la política agrícola de la UE.

La producción de fibras de origen animal para uso en la industria textil, presenta una situación muy diversa en función del origen de la fibra (Tabla 1). De aquellos tipos de fibra cuya producción está estabilizada, la fibra de Cachemira es la de mayor valor económico.

Tabla 1. Situación de la Producción Mundial de fibras de origen animal

Tipo de Fibra	Animal Productor	Producción (Tm)	Tendencia
Alpaca y Llama	Camélidos	4.000	Estable
Angora	Conejo	7.000	Estable
Pelo Camello	Camélidos	1.500	Decrec.
Cashgora	Cabra Angora x Cabra salvaje	50	Decrec.
Cachemira	Cabra cachemira	5.000	Estable
Mohair	Cabra Angora	20.800	Estable
Vicuña	Camélido	No disp.	Inaccesible
Lana de Yak	Bóvido	1.000	Decrec.

Fuente: Ho Phan. 1997. European Fine Fibre Network Workshop Report 1.

Tradicionalmente el término de cachemira hacía referencia a las cabras procedentes de Asia. Actualmente, la definición de cachemira viene impuesta por razones de mercado. Se define como cabras cachemira aquellas cabras con doble capa de abrigo que producen fibra con diámetro inferior a 18,5 µm. En las cabras de cachemira existe una gran diversidad de genotipos con un rango muy amplio de variación tanto en diámetro como en producción de fibra. Se describen algunos tipos, tales como, Siberia (producción >500g y diámetro >20 µm), China y Mongolia (diámetro <16 µm. y producción entre 100-300g.). Existen algunas regiones chinas donde se dan animales de excelente calidad (diámetro entre 15-17 µm) y alto nivel de producción (500 g). Los países tradicionalmente productores de fibra cachemira son asiáticos (China, Mongolia, Irán, Afganistán, Siberia, Pakistán e India) aunque a partir de la década de los ochenta se incorporaron Australia, Nueva Zelanda y Reino Unido, que actualmente junto con Estados Unidos, son los principales países procesadores.

En Europa existe un creciente interés por la producción de fibras textiles de origen animal, tales como cachemira, mohair, angora y lanas de calidad (Milne, 1997). Existen en realidad distintas iniciativas por parte de cooperativas de productores que aúnan esfuerzos para la producción de la materia prima, el procesado y la comercialización del producto elaborado. Así al amparo de una red temática financiada por la UE se han de constituir bases de datos que englobarán la información generada en cada país para las distintas especies.

Osoro y col., 1995, han identificado la producción de fibra de cachemira en España como una complemento a los sistemas de producción tradicionales por el valor añadido que se puede obtener de la fibra. La cachemira es la fibra de origen animal que presenta una mayor asociación con el lujo y la calidad. El valor de la fibra de cachemira es una función de varias características, fundamentalmente de calidad y cantidad.

La introducción de cabras de Cachemira en España se hizo en el marco de un proyecto INIA de carácter multidisciplinar, que integró aspectos relacionados con la genética, los sistemas de manejo, reproducción, enfermedades parasitarias y economía.

Fisiología y Medida de la fibra

Las cabras de cachemira son animales de doble capa de abrigo. Sobre la piel se distinguen dos tipos de folículos pilosos; los folículos primarios que producen pelo con diámetro > 50 µm y longitud de 6-20cm y cuyo papel biológico es la protección física y, folículos secundarios productores de fibra con diámetro de 4 -30 µm y longitud de 4-8 cm, cuya función es la de aislante térmico (Lupton y col., 1995 citado por Herrmann y Wortmann, 1995).

La población de folículos, aparentemente, queda fijada entre los 3-6 meses de edad (Lambert y col., 1984). El folículo desarrolla un ciclo continuo de actividad de crecimiento, maduración, permanencia, y muda (Chase, 1954). En el Hemisferio Norte la máxima actividad folicular tiene lugar entre el solsticio de verano y el solsticio de invierno. Las cabras de cachemira presentan un patrón de crecimiento de la fibra y muda estacional. Durante el invierno presentan doble capa de abrigo. En primavera, bajo la influencia del aumento de horas de luz, se produce la muda de la fibra que se lleva a cabo durante algunas semanas. Se observa variación en el inicio de la muda entre animales aunque dentro de animales, los ciclos de actividad folicular están más o menos sincronizados.

Protocolo para determinar las características de fibra cachemira

Los animales, machos y hembras, son pesados y controlados a los cinco meses de edad tomando muestra de vellón de 10 cm^2 en el punto medio del lateral del animal. Estas muestras se almacenan en una habitación en seco. Antes de iniciar el proceso de medida, las citadas muestras, se llevan a una cámara de acondicionamiento con temperatura y humedad relativa controladas (20°C ± 2°C y 65% ± 5%) En primer lugar se pesa la muestra obtenida. A continuación se procede a la separación manual de pelo y fibra y se pesan por separado. El siguiente paso es el lavado de la fibra con éter de petróleo. Posteriormente se mide la longitud de la fibra (con una regla graduada), extendiendo ésta sobre un tablero de terciopelo. La determinación del diámetro de la fibra se realiza por microscopía de proyección y análisis de imagen (CUE.2.2) midiéndose 200 fibras por muestra.

En la actualidad se está imponiendo el analizador óptico del diámetro de la fibra (OFDA), por su rapidez y por que permite trabajar con un mayor volumen de muestra. Sin embargo, aún no está resuelta la falta de consistencia entre las determinaciones del diámetro de la fibra obtenida por diferentes métodos y/o en diferentes laboratorios (Herrmann and Wortmann, 1994-95). Lo cual obviamente puede presentar un problema a la hora de establecer criterios de selección en un programa de Mejora que implique la intervención de distintos países y/o laboratorios.

Definición de los caracteres de posible interés en el Programa de Mejora Genética

En el Programa de Mejora Genética que se está desarrollando actualmente los caracteres de posible interés son medidos sobre el animal joven, a los cinco meses. Esto se hace con el fin de no dilatar las decisiones a la hora de seleccionar los reproductores y así disminuir el intervalo generacional. Sobre el cabrito se definen:

- Diámetro medio de la fibra (**D.M**): media en µm de doscientas fibras de la muestra.
- Desviación estándar del diámetro de la fibra (DE_d).
- Peso vivo (Kg.): peso del animal el día en que se toma la muestra del vellón.
- Rendimiento: definido como la razón del peso de la fibra al peso total (pelo más fibra), en porcentaje.
- Peso en gramos de fibra de cachemira en la muestra (P_CACH): se obtiene como el producto del rendimiento por el peso total de la muestra.
- Longitud de la fibra (mm): se toman tres medidas máxima, media y mínima.
- Medulación: número de fibras meduladas sobre el total de fibras medidas. Expresado en porcentaje.
- Producción anual estimada (PAE) en gramos: obtenida utilizando la fórmula de Bishop y Russel (1994)

$$PAE = 53{,}6\ (PV)^{0{,}703}\ P_CACH / 0{,}4$$

  donde, PV representa el peso vivo del animal el día que se toma la muestra y, P_CACH el peso de cachemira en la muestra. Mediante esta fórmula se extrapola la fibra producida en la superficie muestreada a la superficie corporal total.

Existen algunos trabajos que están determinando caracteres histológicos que podrían tener algún interés para ser utilizados en la selección, tales como densidad de folículos secundarios y razón de folículos secundarios a primarios (Pattie y Restall, 1989). La definición de los caracteres es un tema aún abierto. Se ha visto que el PAE sobrestima la producción real y por otro lado existe la ya citada falta de consistencia entre laboratorios. A estas dificultades podemos añadir el desconocimiento que existe de los caracteres de fibra en el animal adulto ya que el muestreo siempre de hace sobre el animal joven.

Información para la venta de fibra en el animal adulto

Hay dos procedimientos para la recogida de la fibra cachemira que son peinado y esquilado. En España, por razones de manejo y costes de mantenimiento de los animales, se ha optado por el peinado.

El valor de la fibra en la industria textil viene determinado por el peso total del esquilado o peinado junto a una escala de apreciación visual relativa al contenido en fibra (1 < 20%; 2 = 20-40%; 3 = 40-60%;4 = 60-80%; 5 > 80%.) y la calidad de fibra. Para ésta última se establecen tres categorías según la apreciación visual del diámetro, hilado con diámetro de fibra inferior a 16,5 micras; tejido para diámetro de fibra entre 16,5 y 18,5 micras. Las fibras

con diámetro superior a 18,5 micras no se considera fibra de cachemira. Otros criterios de calidad son el color (blanco, no blanco, coloreado) y la apreciación visual de la longitud de la fibra. Fibras inferiores a 3 cm o superiores a10 cm son rechazadas.

Caracterización productiva de la población de Cachemira española

La población de cachemira española procede de Escocia. La cachemira escocesa representa un 'pool' de genes de distintas procedencias (Islandia, Siberia, Tasmania, Nueva Zelanda y cabra autóctona) que han sido tratadas como una población única y sometida a un proceso de selección (Bishop y Russel, 1994).

En la Tabla 2 se presentan las medias fenotípicas de los caracteres que han sido medidos en la población española a los cinco meses de edad. Como puede apreciarse la fibra española es de alta calidad (comercialmente, fibras por debajo de 16,5 micras son así consideradas) con una producción de cachemira (PAE) muy próxima a las referidas en otras poblaciones de cachemira a éstas edades.

Tabla 2. Medias fenotípicas de los caracteres controlados en cabras de cachemira en España.

Carácter	N	Media	D.E.
D.M (μm)	411	13,95	1,30
DE_d (μm)	411	2,64	0,43
PAE (g)	347	228	146
PDM (kg)	401	19,1	6,1
PCACH (g)	361	0,216	0,132
LM (mm)	362	31,9	12,7
Rendimiento %	362	35,4	14,2
Medulación %	362	1,2	1,4

De acuerdo a las características productivas medias de la población de cachemira española, se ha determinado cual sería el ingreso bruto medio por animal de un productor únicamente obtenido en función del valor de la fibra. Para ello hemos recogido el sistema de precios del Mercado Europeo de 1997 (Russel, 1998). En éste mercado se consideran tres categorías de precios de acuerdo al diámetro medio. Fibras por debajo de 16,5 μm (17.908 pts./kg); fibras con un diámetro entre 16,5 μm y 18,5 μm (13.585 pts./kg), y fibras por encima de 18,5 μm sin precio establecido ni valor comercial en el mercado de la Cachemira. El valor comercial que se podría asignar a la fibra producida por la cachemira española se establece siguiendo el procedimiento de Ponzoni y Guifford (1990). A precios de mercado de 1997 el precio medio de la Cachemira española sería de 17.799 pts. / kg. o 4.058 pts por animal.

Parámetros genéticos

Existe un número muy limitado de referencias bibliográficas relativas a los caracteres productivos en poblaciones de Cachemira. Los caracteres involucrados en la producción de fibra presentan, en general, heredabilidades entre moderadas y altas. El rango de valores encontrados para diámetro medio de la fibra varía entre 0,47 y 0,99 (Bishop, 1994; Bishop y Russel, 1996) para peso de la cachemira entre 0,36 y 0,62 (Bishop, 1994; Bishop y Russel, 1996) mientras que el peso vivo presenta heredabilidades más consistentes y similares a las encontradas en otras especies (Bishop, 1994). Este rango de valores representa una mezcla de poblaciones y de edades.

Es a su vez de especial importancia la correlación genética existentes entre todos los caracteres relacionados con la fibra. En este sentido, las estimas encontradas en la literatura son altas y positivas; siendo especialmente altas las estimas de correlaciones genéticas entre diámetro medio y cantidad de fibra que oscilan entre 0,62 y 0,83 (Pattie y Restall, 1989; Bigham et al., 1993; Bishop y Russel, 1996). Por el contrario, no parece existir una asociación clara entre las características de la fibra y peso vivo de los animales. Las correlaciones genéticas con el peso vivo, tanto para diámetro como para cantidad, citadas en la literatura no parecen ser distintas de cero (Pattie y Restall, 1989; Bighman y col., 1993; Bishop y Russel, 1996).

Los parámetros genéticos que han sido utilizados en la población española son los estimados por Bishop y Russel (1996), dado que la población española procede de la escocesa y la cantidad de información originada en la primera hasta la fecha es limitada. Por otra parte dada la relativa consistencia encontrada en la literatura de las estimas de heredabilidad (moderadas o altas) de los caracteres relacionados con la fibra, se esperaría una respuesta rápida a la selección. Sin embargo, la existencia de correlaciones genéticas (igualmente consistentes en la literatura) positivas y altas entre los distintos caracteres hacen crítico el planteamiento de los objetivos y las estrategias de selección por las posibles respuestas correlacionas indeseables. Por ejemplo, en un estudio de simulación de un núcleo realizado por Díaz y col. (1998) se observó que seleccionar únicamente para aumentar la producción de fibra conllevaría un incremento anual en el diámetro de 0,5 micras. Por otra parte existen evidencias experimentales que así lo indican; de la línea de selección desarrollada en el MLURI para disminuir el diámetro se obtuvo como respuesta correlacionada una reducción del 38% de la Producción Anual Estimada en cuatro años de selección (Russel, 1998).

Según las estimas obtenidas por Bishop y Russel (1996), la heredabilidad del diámetro medio de la fibra (DM) es de 0,68 y 0,55 la heredabilidad asociada a la PAE. La estima de correlación genética es de 0,83.

Objetivos de selección

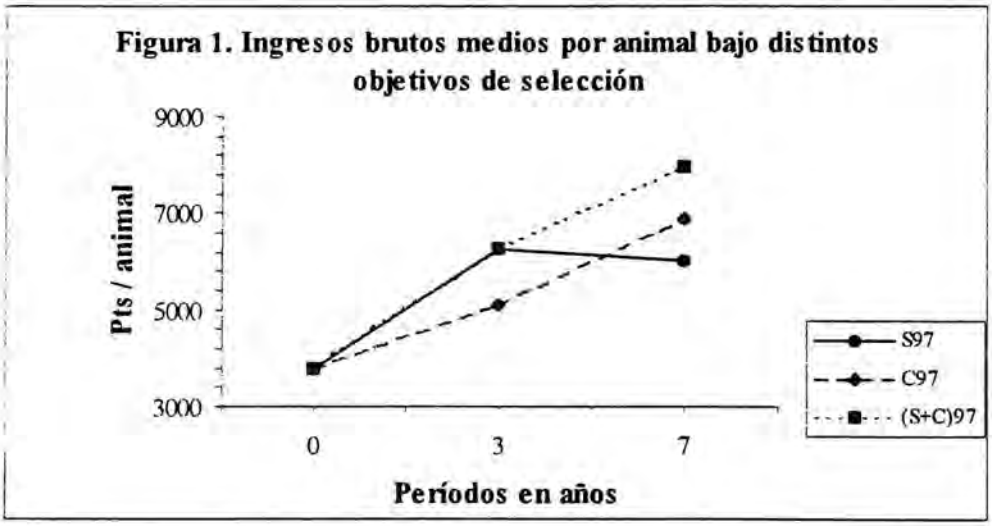
Como ya se ha dicho anteriormente, desde un punto de vista comercial calidad y cantidad de fibra son esenciales. Por tanto, la producción de fibra y por consiguiente, la mejora genética de la fibra de Cachemira, deben abordar esos dos grupos de caracteres. Los caracteres de calidad vienen determinados por la industria textil y su principal componente es el diámetro medio de la fibra; aunque existen otras características adicionales tales como el color, la longitud, el lustre, el rizo etc. Desde el punto de vista de la selección centraremos nuestro interés en el

diámetro medio de la fibra y la Producción Anual Estimada (PAE) como caracteres de calidad y cantidad, respectivamente.

En función de los precios medios de mercado de 1997 que han sido indicados anteriormente, si la selección se orientara a la búsqueda de calidad, un decremento de una unidad en el diámetro medio produciría un incremento de 95 pts / kg en el precio de la fibra producida. Por el contrario, si el proceso de selección resultara en el incremento de un micrómetro en el diámetro medio supondría una pérdida de valor de 397 pts. / kg a precios del mercado 1997 y con las medias de la población actual de cachemira española (Tabla 2). Como puede observarse la relación calidad precio no es lineal, y por lo tanto la consecuencia, en términos de ingresos percibidos por el productor, dependerá en gran medida del efecto que, incrementar o disminuir una micra en el diámetro medio de la fibra, tenga sobre la cantidad de fibra producida.

Desde esta perspectiva económica discutiremos a continuación distintos objetivos de selección. Para ello, se utilizan los resultados del estudio de simulación realizado por Diaz y col. (1998) anteriormente mencionado. En dicho trabajo, se compararon distintas estrategias de selección con restricciones con una aplicación a un núcleo de cabras de cachemira. Las distintas estrategias de selección se orientaron a mantener el diámetro de la fibra constante. Los resultados del citado estudio mostraron que la selección para incrementar la producción sin restricción en el diámetro se traduciría en un aumento de producción de 52 g anuales con un incremento de diámetro de 0,5 micras. Sin embargo, seleccionando para aumentar cantidad con restricción en el diámetro el aumento de producción obtenido era de 25 g mientras que el diámetro se mantenía constante.

Figura 1. Ingresos brutos medios por animal bajo distintos objetivos de selección



En la Figura 1 se presentan los efectos medidos en términos de ingresos brutos por animal de tres objetivos de selección distintos (S = selección para cantidad sin restricciones en diámetro; C = selección para producción con restricción en diámetro y S+C = selección S en los tres primeros años y C en los siguientes). Se observa claramente que la elección entre los distintos objetivos depende de la perspectiva en el tiempo. Al cabo de tres años de selección, S implicaría un ingreso adicional de 1.166 pts. por animal respecto a C. Por el contrario, si consideramos un periódo de siete años de selección, S supondría una pérdida de 875 pts. por

animal respecto a C. Una combinación de ambos objetivos parecería ser la mejor opción, puesto que a corto plazo implicaría los mismos ingresos que S y a más largo plazo supondría una ventaja con respecto a C (1.090 pts./animal). Las diferencias que serían esperadas entre los tres escenarios podrían explicarse por el efecto que el incremento del diámetro tiene sobre el precio de la fibra y que no parece ser compensado por una mayor producción. Así por ejemplo, bajo la estrategia S, tendríamos animales de mayor producción, pero en su mayor parte estarían en el segundo tramo de precios (16,5 – 18,5 μm) y la fibra del 28 % de los animales de nuestra población no tendría valor económico. En cambio, en la combinación de S+C, el 76% de la población se mantendría en el primer nivel de precios (< 16,5 μm) y el 24 % restante en el segundo tramo.

Por lo tanto, del conocimiento del mercado internacional se deduce que seleccionar para disminuir diámetro no tendría sentido dado los niveles de calidad de nuestra población en la actualidad; mientras que apostar por cantidad únicamente tendría sentido a corto plazo. Si además tenemos en cuenta que existen ciertas evidencias de que el diámetro medio de la fibra incrementa con la edad (Pattie y Restall, 1989) se espera que restringir el diámetro medio de la fibra nos permitiría asegurar un valor comercial para todo el rebaño.

Actuaciones en la población de cachemira española

En la población de cachemira española, de entre todos los caracteres controlados en los animales, hemos centrado nuestro interés sobre el diámetro medio de la fibra y la Producción Anual Estimada como caracteres representativos de la calidad y cantidad de la fibra.

En un primer período, se testaron los machos importados de Escocia y los apareamientos se organizaron para evitar consanguinidad. En el año 1995 se pudieron evaluar por primera vez los animales de cachemira. Dado que en ese momento se disponía únicamente de información sobre el diámetro medio de las muestras tomadas en los cabritos nacidos en el año 1994, se realizó una evaluación genética mediante la aplicación de un BLUP-Modelo Animal. La heredabilidad considerada fue 0,68 (Bishop y Russel, 1996). El modelo de evaluación contempló, la localización, el sexo y la edad del animal el día de la toma de la muestra como covariable. Puesto que sólo disponíamos de información para el diámetro de la fibra los machos fueron seleccionados para disminuir dicho diámetro. Para comprobar si los caracteres respondían a la selección como era esperado, se estimó la respuesta esperada para diámetro en función de los valores medios predichos para los machos padres de los individuos nacidos en el año 1996 (única información de la que se dispone actualmente en este sentido). Para obtener la respuesta esperada, se utilizaron las valoraciones genética de esos machos obtenidas a partir de su propia información. Asumiendo que todos los machos tenían la misma contribución a la descendencia se hubiera esperado un decremento en el diámetro de 0,16 micras, pero tomando en cuenta las contribuciones diferenciales de los macho el decremento esperado sería de 0,26 micras, coincidiendo éste último valor con la media de los individuos nacidos en 1996.

A partir de 1995, los animales fueron controlados para los otros caracteres que se citan en la Tabla 2. Para la evaluación genética se implementó un BLUP-Modelo Animal Multicarácter para diámetro medio y Producción Anual Estimada. Los animales han estado ubicados en Asturias (dos fincas) y Guadalajara. De los 17 machos que figuran como padres con hijos con datos, 10 tienen descendencia en al menos dos de las localizaciones para provocar la conexión. En la evaluación realizada en el año 1997, se valoraron un total de 856 animales, 367 de los

cuales poseían información propia. Vía hembras no se ha practicado selección hasta la fecha dado que uno de los objetivo era incrementar el tamaño de rebaño. Los apareamientos se verificaron para minimizar el incremento de consanguinidad.

Con posterioridad y en base a los resultados de la simulación para el manejo del núcleo de selección se ha implementado un BLUP con Restricciones (Quaas y Henderson, 1976) para la selección de hembras en el núcleo, y técnicas de programación lineal aplicadas a las valoraciones BLUP multicarácter, para la selección de los machos (Díaz y col, 1998). La programación lineal nos permite maximizar la respuesta en cantidad de fibra sujeta a una serie de restricciones tanto en el diámetro como en el número mínimo de machos y número máximo de hembras a cubrir por macho. Las soluciones obtenidas nos indican el número de machos a seleccionar y la proporción de hembras a cubrir por cada uno de los machos para obtener el máximo de la función sometido a las restricciones impuestas. La programación lineal es un instrumento de gran flexibilidad para imponer restricciones (Díaz, 1998, resultados no publicados), de manera tal que nosotros podríamos establecer el incremento en diámetro deseado (esperado) y en función de esta restricción buscar la combinación machos que maximizan el incremento de producción.

Referencias citadas.

Bishop, S. 1994. Strain comparisons and genetic parameters for cashmere goats. In: Proc. 5th WCGALP. Vol 19:401-408.

Bishop, S. C. y Russel, F.J. .1994. Cashmere production from feral and imported cashmere goat kids. Anim. Prod. 58: 135-144.

Bishop, S. y Russel, A. 1996. The inheritance of fibre traits in a crossbred population of cashmere. Anim. Sci. 63:429-436.

Chase,H. B.1954. Growth of hair. Physiological Reviews, 34: 632-633.

Díaz, C.; Toro, M. y Rekaya, R.,1998. Comparison of restricted selection strategies: an application to selection of Cashmere goats. Livest. Prod. (En prensa)

Herrmann, S. y Wortmann F.J. 1994 - 1995. First results of the round trial of fibre diameter testing facilities. Fine Fibre News. 4: 6-16.

Herrmann, S. y Wortmann F.J. 1995. Developments of models for the simultaneus estimation of fibre quality and yield in raw cashmere fleeces. Fine Fibre News 5: 6-10.

Lambert A., Restall B. J., Norton B W. y Winter J. D. 1984. The post-natal development of hair follicles groups in the skin of the Australian feral goat. Proc. Aust. Soc. Anim. Prod., 15. 420-423.

Mandakh, B. and Zagdsuren, Y., 1998. Selection index for improvement on Mongolian Cashmere goat performance. In: Proc. 6thWCGALP. Vol 24:96-98.

Milne, J., 1997. European Fine Fibre Network Workshop Report 1. Pgs. 2-3.

Osoro, K.; Martínez, A. y Ormaizabal, J.J. 1995. Posibilidades de la cabra de Cachemira. ITEA. Vol. Extra. 16 (I):204-206.

Pattie, W.A. and Restall, B.J. 1989. The inheritance of Cashmere in Australian Goats. 2. Genetic Parameters and breeding values. Livest. Prod. Sci. 21: 251-261.

Pattie, W. A. and. Restall B.J. 1990. Breeding for cashmere. In. Proc. Of The Scottish Cashmere. The variable alternative. Battleby, 21st July, 1990, pp 13-31.

Ponzoni, R.W. and Gifford, D.R. 1990. Developing breeding objectives for Australian Cashmere goats. J. Anim. Breed. Genet. 107: 351 – 370.

Quaas, R.L. y Henderson, C.R. 1976. Selection criteria for altering groth curve. Dept. Anim. Sci., Cornell University. Ithaca, N.Y. Mimeo.

Restall, B. J. and Pattie, W.A. 1989. The inheritance of Cashmere in Australian Goats. 1. Characteristics of the base population and the effects of environmental factors. Livest. Prod. Sci. 21: 157-172.

Russel, A. 1998. Breeding programme achieves significant gains in Cashmere production. Scotish Cashnere News. Num. 36: 1–2.

PRIMEROS RESULTADOS DEL ESTUDIO DE LA MORFOLOGÍA MAMARIA EN GANADO OVINO MANCHEGO

SELDAS,MªE.[1]; JURADO,J.J.[2]; GARCÍA,M.[1]; MONTORO,V.[1]; PÉREZ-GUZMÁN,MªD.[1]
[1]CERSYRA. Avda. del Vino, 6. 13300-Valdepeñas. Ciudad Real
[2]INIA. Ctra. de la Coruña, Km. 7. 28040-Madrid

1. RESUMEN.

Se han analizado 4.706 puntuaciones morfológicas de ubres correspondientes a 3.122 ovejas que pertenecen al núcleo de selección del programa de mejora genética de la raza Manchega. La finalidad de esta evaluación es mejorar por selección la aptitud morfológica de la ubre para el ordeño mecánico.

En cada puntuación morfológica se tomaron en cuenta los siguientes caracteres mamarios básicos: **Profundidad e Inserción** de la ubre, **Verticalidad** y **Tamaño** de los pezones. Se consideró además un quinto carácter global denominado **Conformación** de la ubre. Dichos parámetros se valoraron siguiendo una escala lineal de 9 puntos (del 1 al 9).

Las medias de las puntuaciones para los caracteres fueron, para la Profundidad 5.27; para la Inserción 5.07; para la Verticalidad 4.63; para el Tamaño 4.92 y para la Conformación 4.56.

En el modelo de análisis de varianza realizado, el efecto Oveja(rebaño) es el que explica el mayor porcentaje de varianza para los cinco caracteres evaluados. A continuación, los factores más significativos e influyentes son el Rebaño, el Intervalo desde el Parto hasta el Control y el Calificador.

Finalmente, se ha obtenido una correlación fenotípica positiva entre la Conformación de la ubre y la Producción Láctea. En función de los resultados que se obtengan del estudio genético, podemos confirmar si se puede mejorar la aptitud morfológica para el ordeño mecánico y al mismo tiempo incrementarse los niveles productivos.

2. INTRODUCCIÓN.

El futuro de la producción de leche de oveja en nuestro país depende en gran medida de la mecanización de la operación del ordeño. Esto es debido a que, sobre dicho tipo de ordeño se atribuye una realización del trabajo en condiciones más adecuadas, una mejora de la eficacia de la mano de obra al aumentar los rendimientos horarios, la posibilidad de obtener leche en mejores condiciones higiénicas, la facilidad para la organización del trabajo y el control de resultados e, incluso, que permite la versatilidad de los operarios.

Las características de la ubre apta para el ordeño mecánico en las ovejas son conocidas y fueron descritas por Mikus (1969, 1972, 1978) quien denominó "udder machine" a la ubre óptima, desde el punto de vista morfológico, para la utilización de la máquina de ordeño.

Así pues, sabiendo que los caracteres morfológicos de la ubre en el ganado ovino presentan heredabilidades entre medias a altas (Mavrogenis *et al.*, 1988; Charon, 1990); y conociendo las características que debe tener la ubre apta para el ordeño mecánico "udder machine", sería posible incluir dentro de los programas de selección de la raza ovina Manchega los caracteres morfológicos mamarios, con el fin de disponer lo más pronto posible de sementales valorados para dichos caracteres.

3. MATERIALES Y MÉTODOS.

3.1. Material Animal.

Para la realización del presente trabajo se han utilizado ovejas de 12 rebaños pertenecientes al núcleo de selección del programa de mejora genética de la raza Manchega, llevado a cabo en el Centro Regional de Selección y Reproducción Animal (CERSYRA) de Valdepeñas.

Las ovejas de estas ganaderías están sometidas al Control Lechero Oficial. En los 12 rebaños se utiliza el ordeño mecánico.

Se han analizado y estudiado un total de 4.706 puntuaciones realizadas a 3.122 ovejas durante el año 1997.

3.2. Metodología de Valoración.

Para la evaluación morfológica de la ubre se ha seguido un sistema de puntuación lineal con valores comprendidos entre 1 y 9, diseñado y desarrollado por primera vez en ganado ovino por el grupo de Genética de la Universidad de León (Fernández de Sierra y cols., 1995). Dicho método está basado en la valoración de 5 caracteres mamarios que se describen a continuación:

"Profundidad" (PROF), este carácter se define como la distancia entre la inserción posterior de la ubre y el piso de la misma tomando como punto de referencia el corvejón de la oveja. Se considera como puntuación óptima el valor 5, correspondiente al punto medio de la escala.

"Inserción" (INS), se define como la sujeción que presenta la glándula mamaria a la pared abdominal de la oveja. Este carácter está determinado por el perímetro de la base de inserción. El valor óptimo corresponde al máximo perímetro de inserción, al que se le adjudica la máxima puntuación (9 puntos).

"Verticalidad" (VERT), este carácter está determinado por el ángulo de inserción de los pezones. La mofología considerada óptima (9 puntos) corresponde a la máxima verticalidad; es decir, pezones dirigidos hacia el suelo y muy poco adelantados.

"Tamaño" (TAM), definido por la longitud de los pezones, dada la alta correlación existente entre la longitud y la anchura de los mismos. La puntuación óptima coincide con el valor medio de la escala (5 puntos).

"Conformación" (CONF), considerado como un valor global de la aptitud morfológica de la ubre para el ordeño mecánico. El valor máximo de la escala (9 puntos) se corresponde con el tipo de ubre ideal para el ordeño mecánico (Mikus, 1978).

3.3. Metodología Estadística.

Los análisis estadísticos se han realizado mediante el paquete estadístico SAS (SAS Institute INC., 1988). Para ello, la información contenida en la base de datos se ha transformado a ficheros formato SAS, mediante ficheros puente DBF.

Para el estudio de las variables dependientes se han realizado análisis de varianza mediante el procedimiento GLM (Modelo Lineal General). Así mismo, se han utilizado los procedimientos MEANS, FREQ y CORR para el estudio de las distribuciones de los datos y correlaciones entre las diferentes variables.

El modelo lineal de análisis de varianza utilizado fue el siguiente:

$$Y_{ijklmn} = \mu + R_i + O_{j(i)} + P_k + NP_l + CA_m + IN_n + \varepsilon_{ijklmn}$$

Siendo:

$Y_{ijklmno}$ = Valor de cada observación (Variable Dependiente).

μ = Valor Medio

R_i = Efecto del Rebaño (Factor Fijo). Contiene 12 clases.

$O_{j(i)}$ = Efecto de la oveja j dentro del rebaño i (Factor Jerarquizado).

P_k = Efecto de la Producción (Covariable).

NP_l = Efecto del Número de Parto (Factor Fijo). Agrupa 3 clases.

CA_m = Efecto del Calificador (Factor Fijo). Posee 5 clases.

IN_n= Efecto de Intervalo de tiempo comprendido entre la fecha de parto y la realización de la calificación (Factor Fijo). Se han establecido 3 clases.

ε_{ijklmn} = Error residual.

4. RESULTADOS Y DISCUSIÓN.

Las medias aritméticas, los errores estándar y los coeficientes de variación de las puntuaciones, para cada uno de los cinco parámetros mamarios calificados, figuran en la Tabla 1.

Tabla 1. Valores estadísticos para los parámetros mamarios calificados.

	PROF	INS	VERT	TAM	CONF
MEDIA	5.27	5.07	4.63	4.92	4.56
ERROR	±0.81	±1.00	±1.05	±0.82	±1.04
COEF. VAR.	15.39	19.80	22.55	16.59	22.73

Las medias de las puntuaciones oscilan entre un mínimo de 4.56 (CONF), y un máximo de 5.27 (PROF), siendo la mayoría muy próximas a 5 puntos y presentando una distribución de puntuaciones casi normal.

La CONF y la VERT destacan como los caracteres más dispersos, coincidiendo con los resultados obtenidos por Fernández de Sierra y cols. en 1995 para la raza Churra.

4.1. Factores de Variación sobre las puntuaciones de los caracteres mamarios.

Tal y como se muestra en la Tabla 2, el modelo de análisis de varianza realizado es muy significativo ($P \leq 0.001$) para los cinco parámetros mamarios calificados.

El efecto Oveja(rebaño) es el que explica mayor porcentaje de varianza para los cinco caracteres, lo que indica la gran variabilidad que existe en la morfología de la ubre del ganado ovino de raza Manchega. Estos porcentajes se sitúan entre un máximo del 79.95% para la VERT y un mínimo del 66.94% para la INS, resultando estos valores muy interesantes para la realización de selección sobre la morfología mamaria.

En un 2° lugar (en cuanto a porcentaje de varianza explicada) se sitúa el factor Rebaño, seguido por el factor Intervalo Parto-Control para los caracteres PROF, INS y CONF y por el factor Calificador para la VERT y el TAM.

El efecto del N° de Parto es el menos significativo y el que menor porcentaje de varianza explica para los cinco caracteres mamarios.

Tabla 2. Porcentaje de Varianza Explicada por los distintos Factores de Variación sobre los Caracteres Mamarios y Significación del Efecto.

	VARIABLES DEPENDIENTES				
Fte. de Variación	**PROF**	**INS**	**VERT**	**TAM**	**CONF**
Rebaño	8.92 ***	3.92 ***	2.80 ***	2.80 ***	2.41 ***
Oveja(rebaño)	71.40 ***	66.94 ***	79.95 ***	76.36 ***	72.16 ***
N° Parto	0.14 ***	0.02 NS	0.06 NS	0.01 NS	0.04 NS
Calificador	1.27 ***	0.45 ***	0.34 ***	0.25 ***	0.31 ***
Producción	0.19 ***	0.54 ***	0.08 **	0.10 **	0.26 ***
Intervalo P-C	1.50 ***	2.45 ***	0.21 ***	0.14 **	1.04 ***
Total	**83.42** ***	**74.32** ***	**83.44** ***	**79.66** ***	**76.22** ***

*** = P≤0.001 ** = 0.001<P≤0.01 * = 0.01<P≤0.05 NS = P>0.05

A continuación se expone el efecto de los principales factores sobre los parámetros mamarios calificados.

Rebaño

El efecto Rebaño resulta altamente significativo sobre la puntuación de los 5 caracteres de la ubre evaluados (Tabla 2). El porcentaje de varianza que explica oscila entre un mínimo de 2.41 para la CONF y un máximo de 8.92 para la PROF.

El rebaño es el factor fijo que aporta el mayor porcentaje de varianza explicada para los cinco caracteres morfológicos. La importancia de este efecto se basa en las variaciones genéticas existentes entre las distintas ganaderías, debidas a la selección intrarebaño que vienen realizando los ganaderos de forma tradicional.

Las máximas diferencias encontradas en las puntuaciones medias entre rebaños son: 1.43 para la Profundidad; 1.77 para la Inserción; 1.68 para la Verticalidad; 0.75 para el Tamaño y 1.57 para la Conformación.

Número de Parto

Aunque este efecto no aparece significativo en el modelo (solamente resulta significativo para el carácter PROF) (Tabla 2), si se puede observar una tendencia en las medias de las puntuaciones alcanzadas por los caracteres INS y CONF a lo largo de los

sucesivos partos. Dicha tendencia que nos parece importante resaltar, queda reflejada en la Figura 1.

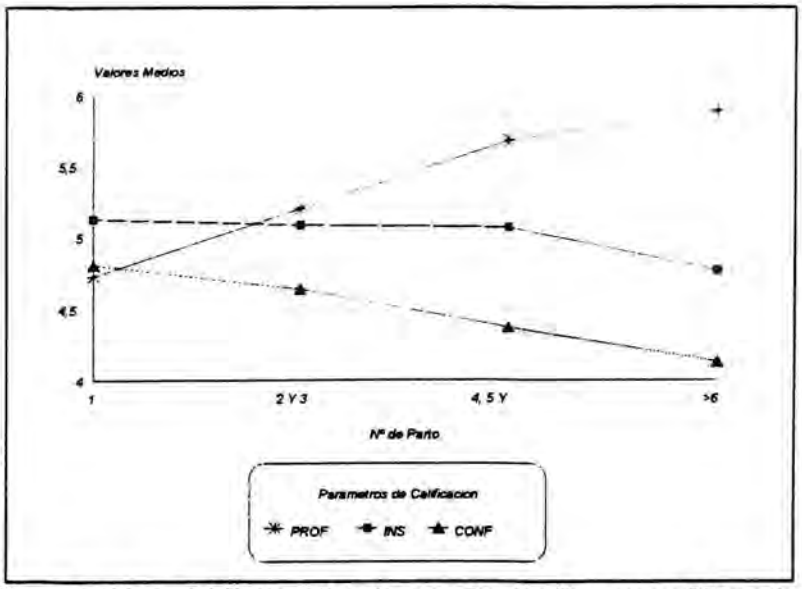


Figura 1. Efecto del Factor N° de Parto sobre los Caracteres Mamarios

Se muestra en esta figura cómo la profundidad aumenta considerablemente según se incrementa el N° de Lactaciones. Este aumento entre los primeros partos se corresponde con el incremento de la producción de leche. A partir del 6° parto las mayores puntuaciones se deben a una pérdida en la capacidad de retracción del ligamento suspensor medio (Ubres descolgadas).

La Inserción de la ubre decrece lentamente hasta el 6° parto, disminuyendo más bruscamente a partir de este momento. Esta disminución de la INS junto con el aumento de la PROF, muestran una evolución con la edad hacia una ubre pendulosa o descolgada.

La Conformación general de la ubre empeora cuantos más partos hayan sufrido las ovejas, como consecuencia del aumento de Profundidad, y de la disminución de la Inserción que experimenta la ubre.

Calificador

El efecto Calificador es estadísticamente muy significativo para todos los parámetros mamarios calificados. Ejerce su máxima influencia sobre los caracteres VERT y TAM.

Este hecho pone de manifiesto diferencias en la forma de evaluar de los calificadores. Por ello resultaría conveniente tratar de unificar algo más los distintos criterios de evaluación.

No obstante, la Inserción, la Profundidad y la Conformación general son los parámetros más fácilmente evaluados por los calificadores. Esto resulta interesante porque la facilidad en la puntuación de la Conformación permite valorar de forma objetiva la morfología de la ubre con respecto a su aptitud al ordeño mecánico.

Producción

La Producción es un efecto significativo para todos los caracteres mamarios evaluados. Dicho efecto ejerce mayor influencia sobre la INS (siendo el 3er efecto fijo más importante a continuación del Rebaño y del Intervalo Parto-Control) que sobre el resto de los parámetros mamarios.

En la Tabla 3 se exponen los coeficientes de regresión obtenidos entre el efecto Producción y cada uno de los parámetros calificados.

Conviene destacar como muy interesante el coeficiente de regresión positivo existente entre la Producción y la Conformación (0.39***), el cual pone de manifiesto que a las ubres más productivas les corresponde mayor puntuación del carácter CONF. Esto repercute en la posibilidad de incluir este carácter en los programas de selección, ya que estaremos seleccionando indirectamente a favor de mayores producciones lecheras.

Tabla 3. Coeficientes de Regresión (Puntos/Litro) entre la Producción Láctea y los Parámetros Mamarios.

	PROF	INS	VERT	TAM	CONF
Produc. Láctea	0.33***	0.57***	0.26**	0.20**	0.39***

*** = P≤0.001 ** = 0.001<P≤0.01

Intervalo Parto-Control

El efecto del intervalo de tiempo transcurrido desde el parto hasta la realización del control ejerce una influencia muy significativa sobre los caracteres morfológicos mamarios.

Es el 2° factor fijo más determinante sobre la Profundidad, la Inserción y la Conformación de la ubre y el 3° sobre la Verticalidad y el Tamaño de los pezones (Tabla 2).

Tal y como se muestra en la Figura 2, las puntuaciones de todos los parámetros evaluados (a excepción del carácter TAM que no presenta diferencias muy significativas) decrecen a medida que va transcurriendo la lactación.

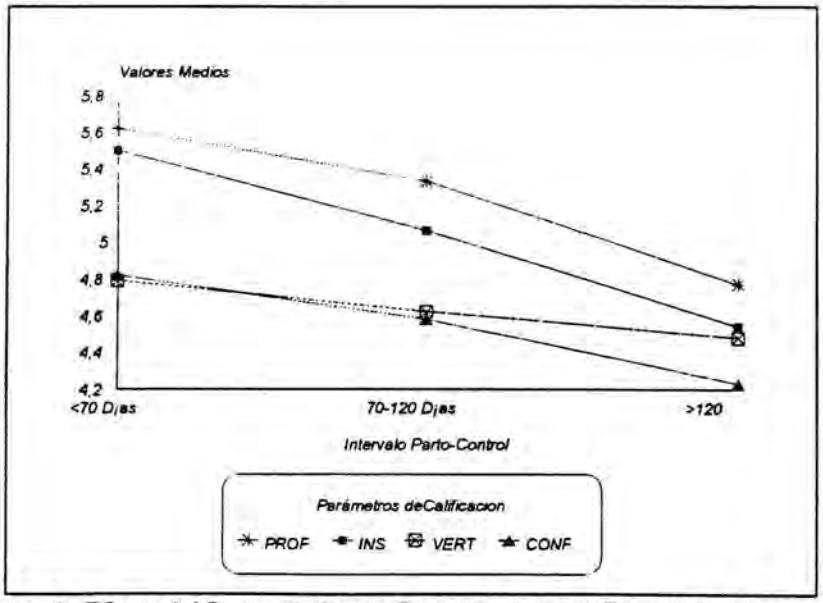


Figura 2. Efecto del Intervalo Parto-Control sobre los Caracteres Mamarios

La disminución de la Profundidad y de la Inserción está relacionada con la pérdida de productividad que se va produciendo conforme avanza la lactación.

El ligero descenso del carácter VERT indica que al final de la lactación los pezones tienden hacia la horizontalidad.

Las puntuaciones de Conformación también van disminuyendo a medida que se incrementa el periodo de tiempo transcurrido entre el parto y el control. Este empeoramiento de la Conformación es debido a la evolución negativa del resto de parámetros mamarios.

4.2. Correlaciones fenotípicas entre los caracteres puntuados linealmente y la producción láctea.

Las correlaciones de las puntuaciones obtenidas entre los 5 caracteres mamarios y la producción láctea se resumen en la Tabla 4.

Tabla 4. Correlaciones fenotípicas entre los Caracteres de la Ubre puntuados linealmente y la Producción Láctea

	INS	VERT	TAM	CONF	PRODUC.
PROF	0.27***	-0.07***	0.15***	0.03*	0.31***
INS		0.2***	0.11***	0.65***	0.32***
VERT			0.35***	0.49***	0.07***
TAM				0.11***	0.13***
CONF					0.21***

*** = P≤0.001 * = 0.01<P≤0.05

Podemos observar en esta tabla, que en general, todas las correlaciones son positivas y significativas, a excepción de la existente entre la PROF y la VERT (-0.07***).

Con respecto a las correlaciones entre los caracteres básicos (PROF, INS, VERT y TAM), las más importantes son las existentes entre los parámetros PROF e INS (0.27***) y entre la VERT y el TAM (0.35***). La primera de ellas (PROF e INS), se debe a que generalmente un aumento en la Inserción de la ubre va acompañado de un aumento de la Profundidad (Incremento del volumen total de ubre). A su vez, las ubres más voluminosas se corresponden con las ubres más lecheras, justificándose así la correlación que se produce entre la Productividad y los caracteres PROF e INS (0.31*** y 0.32*** respectivamente)

La correlación entre los parámetros básicos VERT y TAM es debida a que los pezones ubicados verticalmente actúan como "cisternas" y aumentan por tanto su tamaño (Fernández de Sierra y cols., 1995).

En cuanto a las correlaciones entre los caracteres básicos y la Conformación, la más importante es la correlación entre CONF e INS (0.65***), seguida de la correlación entre CONF y VERT (0.49***), siendo por tanto estos dos caracteres básicos los que más peso poseen sobre la puntuación final de Conformación.

Es interesante analizar la baja y poco significativa correlación entre la CONF y la PROF (a pesar de que la Profundidad está correlacionada con la Inserción y ésta última posee una importante repercusión sobre la Conformación final), debido a la evolución que sufre dicha correlación con el transcurso de los partos. Así en el 1er parto la correlación entre dichos parámetros es 0.36 (se puntúan con mayor Conformación las ubres más desarrolladas), descendiendo hasta hacerse negativa (-0.21***) a partir del 6° parto (una elevada puntuación

de Profundidad equivale a una baja puntuación de Conformación, puesto que se corresponden en la mayoría de los casos, con ubres descolgadas).

Por último destacar la importancia de las correlaciones positivas obtenidas entre la Producción Láctea y los parámetros mamarios. Dichas correlaciones pondrían de manifiesto que, siempre y cuando las correlaciones genéticas de posteriores estudios resulten también positivas, una selección a favor de cualquiera de estos caracteres (sobre todo PROF, INS y CONF) contribuiría a un aumento de la productividad. Estos resultados contrastan con los estudios realizados en la raza Churra por Fernández de Sierra y cols. en 1995. Dichos autores llegaron a la conclusión de que la selección para producción de leche hace que la morfología de la ubre evolucione en contra de la "Udder Machine".

5. CONCLUSIONES

- El valor medio de las puntuaciones en Conformación mamaria de las ovejas manchegas evaluadas es 4.56. Este reducido valor indica una morfología mamaria poco apta para el ordeño mecánico.

- En el modelo estadístico realizado destacan los elevados porcentajes de varianza explicada alcanzada por el efecto Oveja(rebaño). Se pone así de manifiesto la gran variabilidad existente en la morfología de la ubre del ganado ovino Manchego.

- La Inserción de la ubre y la Verticalidad de los pezones son los caracteres que más altas correlaciones poseen con la Conformación global. Dichos caracteres deben considerarse como prioritarios en los programas de selección.

- Las correlaciones fenotípicas positivas que se han obtenido entre los caracteres morfológicos de la ubre y la Producción Láctea, harían posible la selección simultánea para ambas características si el correspondiente estudio genético resultase favorable.

6. REFERENCIAS BIBLIOGRÁFICAS

CHARON, K. M. (1990). Genetic parameters of the morphological traits of sheep udder. World Review of Animal Productions. Vol XXV nº 1: 73.

FERNÁNDEZ DE SIERRA, G. y Cols. (1995). Valoración genética de la morfología de la ubre en ganado ovino de aptitud láctea. Aplicación al programa de selección de la raza Churra. Tesis Doctoral. Facultad de Veterinaria de la Universidad de León.

MAVROGENIS, A. P.; PAPACHRISTOFOROU, C.; LYSANDRIDES, P. y ROUSHIAS, A. (1988). Enviromental and genetics factors affecting udder characters and milk production in Chios sheep. Génét. Sel. Evol. 20 (4): 477.

MIKUS, M. (1969). Manual stripping in the case of machine milking of sheep in relation to the measurements of udder. to the lactation days and certain further indices. Zyvocisna Vyroba, 14(12): 827.

MIKUS, M. (1972). Héritabilité des propriétes morphologiques du pis chez les brebis Cigaja en considérations avec la traite mécanique. In: Vedecke Prace VUO Trecin, 6: 89.

MIKUS, M. (1978). Study of the mutual relationships between dimensions of the udder with regard to improvement of sheep for machine milking. 2me Symposium International sur le traite mécanique des petits ruminants. Alghero (Italia). INRA-ITOVIC: 102.

CONCENTRACIÓN DE CÉLULAS SOMÁTICAS EN LECHE DE OVEJA Y RESISTENCIA A LA MAMITIS: NUEVOS OBJETIVOS DE SELECCIÓN

San Primitivo, F., El-Saied, U.M. y Carriedo, J.A.

Dpto. Producción Animal I. Universidad de León. 24071. León.

Los procesos mamíticos constituyen uno de los problemas patológicos de mayor importancia en las explotaciones lecheras. De su control depende en gran medida el rendimiento económico de las ganaderías. En consecuencia, los métodos que permiten un diagnóstico precoz y que pueden detectar incluso las mamitis subclínicas, han adquirido una importancia relevante. Entre ellos, el recuento de células somáticas en la leche se ha mostrado muy efectivo, de forma que se ha incluido entre los análisis rutinarios que se realizan para determinar los componentes bromatológicos de la leche, hace ya varios años en el ganado vacuno y más recientemente en el ovino y el caprino.

La concentración de células somáticas (RCS) está influenciada, fundamentalmente, por la higiene y las prácticas de manejo. No obstante, en el ganado vacuno se ha identificado un componente heredable apreciable para el RCS, lo que permite establecer una predicción significativa de la respuesta a la selección. En el ganado ovino este tipo de estudios son muy escasos.

El recuento de células somáticas (RCS) en la leche de animales sanos, puede considerarse como una variable más de composición de la leche, mientras que el RCS procedente de ubres enfermas, puede ser considerado como una variable diferente, indicadora de la presencia de mamitis.

Esta dualidad conceptual del RCS debe ser considerada, desde el punto de vista genético, de forma que sería deseable realizar estimaciones de los parámetros genéticos con datos procedentes de animales sanos, independientemente de las estimaciones realizadas sobre animales enfermos o sobre la población completa.

Con objeto de diferenciar de forma fácil y rápida las ovejas sanas de las enfermas, puede utilizarse un umbral de discriminación mediante los valores del RCS. Por debajo del umbral la oveja se admite como sana y por encima se considera enferma. González y col. (1995) han establecido este umbral entre 250 y 300 células/(ml x 10^{-3}).

Las estimaciones de la heredabilidad y la repetibilidad del RCS permiten tomar decisiones sobre los programas de selección y los métodos de valoración genética. Teniendo en cuenta ambos parámetros, en el ganado vacuno Da y cols. (1992) y Zhang y cols. (1994), muestran la adecuación de un Modelo Animal con repetibilidad para la valoración genética, considerando la precisión de las predicciones de los valores genéticos.

Los programas de selección aplicados en la actualidad al ganado ovino de aptitud láctea, consideran fundamentalmente la producción de leche y de proteína. Es importante conocer si estos programas pueden provocar una respuesta correlacionada que afecte al RCS, para lo cual deben estimarse las correlaciones genéticas.

CARACTERÍSTICAS ESTRUCTURALES DE LA VARIABLE

La variable RCS, incluida ya dentro del control lechero cualitativo oficial, como un parámetro más de composición, está siendo utilizada por los ganaderos y técnicos de ganado ovino, para identificar las hembras que padecen procesos mamíticos y establecer las medidas de control que consideren más adecuadas. Por lo tanto, esta variable está cumpliendo una misión importante, por cuanto produce una mejora sanitaria y, en consecuencia, productiva.

Desde el punto de vista de su utilización en programas de selección, la variable RCS debe cumplir una serie de características, que la cualifiquen como objetivo de selección.

Desde el punto de vista de su valoración, la variable puede considerarse muy adecuada. Es de fácil determinación por métodos automáticos, incluso puede incorporarse al proceso de cuantificación de las variables bromatológicas clásicas; no sufre alteración con la adición de las sustancias conservantes normalmente utilizadas y no se modifica durante los primeros días de conservación a temperatura ambiente o, preferentemente, refrigerada.

El RCS es una variable muy influenciada por el estado sanitario de la ubre, lo que hace que presente una enorme variabilidad, moviéndose entre valores muy extremos. Sus propiedades distributivas se encuentran muy alejadas de la normalidad, con una moda de 88, una mediana de 126 y una media de 734 y con un sesgo de 4,28. Con objeto de que adquiera una distribución más cercana a la normal, es necesario realizar su transformación logarítmica (Log RCS). La variable Log RCS presenta una distribución próxima a la normal, con una moda de 5,28, una mediana de 5,11, una media de 4,79 y un sesgo de 0,63, muy próximo al valor de la distribución normal.

CARACTERÍSTICAS GENÉTICAS

Heredabilidad y Repetibilidad

En el ganado ovino, las estimaciones de la heredabilidad del RCS que se han publicado hasta el momento han sido realizadas en la raza Churra. En la primera estimación (Baro y cols., 1994), se obtuvo un valor muy bajo, 0,04. Al ser efectuada para una población con valores muy altos de RCS, media aritmética de 2.254 células/(ml x 10^{-3}) y escasa conexión entre los rebaños, no creemos que pueda ser considerada como representativa del ganado ovino, en la actualidad. Posteriormente, El-Saied y cols. (1998a y b), utilizando datos procedentes de 10 rebaños incluidos en el núcleo de selección de la oveja Churra, con una media de 734 células/(ml x 10^{-3}), obtienen como estimaciones de la heredabilidad 0,09, para controles y 0,12 para medidas lactacionales. (Tabla 1). Las estimaciones de la heredabilidad del RCS obtenidas en el ganado vacuno, han estado comprendidas en el rango 0,09 -0,13 (Banos y Shook, 1990; Schutz y cols., 1990; Da y cols., 1992; Zhang y cols., 1994), siendo por lo tanto semejantes a las obtenidas en el ganado ovino.

Tabla 1. Heredabilidades y repetibilidades (r) y sus errores típicos aproximados (ET) para las medidas del RCS.

Variable	h^2	ET	r	ET
RCS como indicador de mamitis:				
Log RCS (medida lactacional)	0,12	0,03	0,35	0,03
Log RCS (controles)	0,09	0,02	0,38	0,01
RCS de ubres sanas:				
Log (RCS ≤250)	0,03	0,02	0,10	0,02

Cuando para la estimación de la heredabilidad se consideraron únicamente controles inferiores al umbral, considerado éste en 250 células/(ml x 10^{-3}), es decir datos procedentes de ovejas sanas, la heredabilidad que se obtiene es de 0,03, estadísticamente no distinta de cero (Tabla 1). Por lo tanto, en los animales sanos el componente genético heredable parece ser nulo o muy pequeño, en todo caso menor que el obtenido cuando se utilizan todos los datos (0,09 a 0,12), es decir cuando el RCS se considera como indicador de la mamitis.

En consecuencia, en las ovejas enfermas, y por extensión en la población en general, el

RCS puede ser considerado como un indicador aceptable para realizar selección a favor de la resistencia genética a procesos mamíticos, variable que puede ofrecer respuestas relativamente bajas pero significativas, de acuerdo con los valores de su heredabilidad. La selección para esta misma variable no sería eficaz si se consideran únicamente ovejas sanas, debido al bajo valor de la heredabilidad.

El Saied y cols. (1998a y b) han obtenido estimaciones de 0,38 y 0,36, para la repetibilidad del RCS, utilizando controles y medidas lactacionales, respectivamente. A partir de controles efectuados cada 20 días, en la raza Lacaune, Lagriffoul y cols.(1993), estimaron como valor de la repetibilidad del RCS 0,39, semejante al estimado en la raza Churra.

Considerando los valores obtenidos para las repetibilidades, podemos deducir que el sistema de control lechero que se utiliza en los rebaños comerciales de ovejas de raza Churra (un control mensual simplificado alternativo a.m.-p.m), es adecuado para el RCS. A esta misma conclusión han llegado también Fuertes y cols. (1998).

Correlaciones genéticas y fenotípicas

En el ganado ovino, los estudios genéticos del recuento de células somáticas en la leche son muy escasos, debido al reducido número de poblaciones que cuentan con registros genealógicos y productivos adecuados. Por otra parte, en el ganado vacuno, especie en la que estos estudios están más avanzados, los resultados obtenidos para las correlaciones genéticas pueden considerarse contradictorios, habiéndose estimado valores muy diversos (Banos y Shook, 1990; Schutz y cols., 1990).

En la Tabla 2, figuran las estimaciones de las correlaciones genéticas y fenotípicas para las variables RCS, producción láctea y porcentaje de proteína, obtenidas en la oveja Churra (El-Saied y cols., 1.998c).

Tabla 2. Correlaciones genéticas (debajo de la diagonal) y fenotípicas (encima) para los controles de RCS, producción láctea y porcentaje de proteína.

	Pr.láctea	Log RCS	Proteína
Pr.láctea		-0,15	-0,31
Log RCS	-0,10		0,10
Proteína	-0,27	0,03	

Considerando el bajo valor de la correlación genética entre RCS y leche (-0,10), y el

valor nulo entre RCS y porcentaje de proteína (0,03), no cabe esperarse una respuesta correlacionada en los niveles de RCS, cuando se seleccione para incrementar la producción láctea y proteica (El-Saied y cols., 1998c).

Correlaciones fenotípicas inversas entre RCS y producción láctea han sido también encontradas en el ganado vacuno (Schutz, 1977; Ng-Kawi-Hang y cols., 1984; Emanuelson y cols., 1988) y ovino (Gonzalo y cols. 1994; Lagriffoul y cols., 1994; Fuertes y cols., 1998).

Factores ambientales

Sobre el Log RCS, se muestran significativos los efectos del rebaño y el número de lactación. Sin embargo, el tipo de parto no detecta diferencias significativas.

Para el estudio de la concentración de células somáticas en la leche, pueden utilizarse datos procedentes directamente de los controles o medidas lactacionales (media de los controles de cada lactación). En cada caso, como indicaremos más adelante, los factores de variación que suelen utilizarse en los modelos matemáticos destinados a la valoración genética, pueden ser diferentes.

Cuando la variable es el Log del RCS obtenido en los controles, los factores de variación utilizados más frecuentemente son el Rebaño-fecha de control, el número de lactación, el tipo de parto y el estado de la lactación. En nuestros estudios, salvo el tipo de parto, el resto de los factores se han mostrado altamente significativos ($P< 0,001$) y con una influencia muy importante, cuando se utilizan todos los datos. Sin embargo, cuando únicamente se incluyen datos del Log ($RCS \leq 250$), los resultados cambian. El momento de la lactación y el número de parto (al menos en los 6 primeros), parecen no influir sobre el Log ($RCS \leq 250$). Estos resultados permiten deducir que el Log ($RCS \leq 250$), es decir aplicado a ubres sanas y el Log RCS, es decir tomado como indicador de mamitis, se comportan como variables fisiológicamente diferentes.

Cuando la variable es la medida lactacional del RCS, los factores de variación corrientemente utilizados son el Rebaño-año-estación, el número de lactación y el tipo de parto. En nuestros análisis, solo el Rebaño-año-estación y el número de lactación mostraron diferencias altamente significativas ($P < 0,001$).

APLICACIÓN PRÁCTICA

Considerando los valores moderados de la repetibilidad y los relativamente bajos de la heredabilidad de la variable lactacional del RCS, así como de los controles, resulta adecuado

utilizar mediciones repetidas, con la máxima información posible, para aumentar la precisión de la predicción de los valores genéticos, de la misma forma que se ha aplicado en el ganado vacuno (Da y cols., 1992 y Zhang y cols., 1994). Además, con el mismo objetivo, puede utilizarse un Modelo Animal con repetibilidad, especialmente interesante en el ganado ovino, debido a que el número de hijas de los moruecos es más bajo que el de los toros.

En los programas de selección aplicados al ganado ovino, para la predicción de los valores genéticos considerando las variables lactacionales, puede considerarse adecuado el siguiente Modelo Animal multicarácter con repetibilidad:

$$y_{ijkmn} = A_i + PE_i + RAE_j + P_k + T_m + e_{ijkmn}$$

donde:
y_{ijkmn} = medida lactacional de RCS, producción láctea en 120 d, porcentaje de proteína por lactación,
A_i = efecto aleatorio genético aditivo del animal i,
PE_i = efecto aleatorio ambiental permanente del animal i,
RAE_j = efecto fijo del rebaño-año-estación j,
P_k = efecto fijo del numero de parto k,
T_m = efecto fijo del tipo de parto m, y
e_{ijkmn} = efecto residual aleatorio.

Una alternativa en el procedimiento de predicción de los valores genéticos podría ser considerar los controles de RCS en lugar de las medidas lactacionales. Para la producción láctea del ganado vacuno Ptak y Schaeffer (1993) y Swelve (1995) se refieren a las posibles ventajas de este tipo de modelos.

Para la predicción de los valores genéticos en el ganado ovino, siguiendo este tipo de modelo de controles, sería adecuada la utilización del siguiente Modelo Animal multicarácter con repetibilidad:

$$y_{ijklmn} = A_i + PE_i + RDC_j + P_k + L_l + T_m + e_{ijklmn}$$

donde:
y_{ijklmn} = producción láctea, Log (RCS), porcentaje de proteína, por control.
A_i = efecto aleatorio genético aditivo del animal i,
PE_i = efecto aleatorio ambiental permanente del animal i,
RDC_j = efecto fijo del rebaño-día de control j,
P_k = efecto fijo del numero de parto k,
L_l = efecto fijo del periodo de la curva de lactación l,
T_m = efecto fijo del tipo de parto m, y
e_{ijklmn} = efecto residual aleatorio.

Ambos tipos de modelos matemáticos han sido utilizados (El-Saied y cols., 1998a, b y c), para la estimación de los parámetros genéticos, siguiendo los procedimientos AIREML de

Gilmur y cols. (1995) y DFREML de Thompson y cols. (1990).

La adecuación del Modelo Animal con repetibilidad se basa, por lo tanto, en una concepción puramente estadística, pero posiblemente puede tener, también, una justificación en términos biológicos. En este sentido, el Modelo Animal con repetibilidad, supone la utilización de medidas repetidas de RCS, tanto relativas a los diferentes controles de una misma lactación, como a la información procedente de lactaciones diferentes, de una misma oveja.

Biológicamente, la utilización de medidas repetidas permite valorar el historial de cada oveja respecto al recuento. La persistencia de valores altos del RCS, podría indicar una baja capacidad de la oveja para superar los procesos mamíticos y por lo tanto del nivel de respuesta inmunitaria a la infección, es decir un bajo nivel de resistencia. También podría interpretarse como un efecto debido a la patogenicidad del microorganismo causante del proceso mamítico, con resistencia variable frente al sistema inmunitario ovino.

En la situación actual del ganado ovino en Castilla y León, con unas condiciones sanitarias deficientes, aunque en clara evolución positiva, el RCS es una variable muy influenciada por agentes patógenos. Aunque las últimas estimaciones de heredabilidad permiten deducir la posibilidad de obtener respuestas positivas, si se ejerce una presión de selección en contra de altos RCS, se obtendrían respuestas más rápidas y adecuadas mejorando las condiciones sanitarias de los rebaños. Alcanzado un nivel sanitario aceptable, la selección en contra de RCS altos y persistentes y, en consecuencia, hacia mayor resistencia genética, estaría justificada, utilizando para la valoración genética de los reproductores un Modelo Animal con repetibilidad.

BIBLIOGRAFÍA

Banos, G. y Shook, G.E. (1990). Genotype by environment interaction and genetic correlations among parities for somatic cell count and milk yield. *J. Dairy Sci.*, 73: 2563-2573.

Baro, J.A., Carriedo, J.A. y San Primitivo, F. (1994). Genetic parameters of test day measures for somatic cell count, milk yield and protein percentage of milking ewes. *J. Dairy Sci.*, 77: 2658-2662.

Da, Y., Grossman, M., Misztal, I. y Wigans, R.R. (1992). Estimation of genetic parameters for somatic cell score in Holsteins. *J. Dairy Sci.*, 75: 2265-2271.

El-Saied, U.M., Carriedo, J.A. y San Primitivo, F. (1998a). Heritability of test day somatic cell counts and its relationship with milk yield and protein percentage in dairy ewes. *J. Dairy Sci.* Aceptado para su publicación.

El-Saied, U.M., Carriedo, J.A., de la Fuente, L.F. y San Primitivo, F. (1998b). Genetic parameters of lactation cell counts and milk and protein yields in dairy ewes. Enviado.para

su publicación.

El-Saied, U.M., Carriedo, J.A., de la Fuente, L.F. y San Primitivo, F. (1998c). Genetic and environmental correlations among somatic cell counts, milk yield, and protein percentage in milking ewes. 6° International Symposium on the Milking of Small Ruminants. Aceptado para su publicación.

Emanuelson, U., Danell, B. y Philipsson, J. (1988). Genetic parameters for clinical mastitis, somatic cell counts and milk production estimated by multitrait restricted maximum likelihood. *J. Dairy Sci.*, 71: 467-476.

Fuertes, J.A., Gonzalo, C., Carriedo, J.A. y San Primitivo, F. (1998) Parameters of test day milk yield and milk componentes for dairy milk and milk components for dairy ewes. *J. Dairy Sci.*, 81: 1300-1307.

Gilmur, A.R., Thomson, R. y Cullis, R.R. (1995). AIREML, an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics, 51: 1440-1450.

Gonzalez, M.C., Gonzalo, C., San Primitivo, F. y Carmenes, P. (1995). Relationship between somatic cell count and intramammary infection of the half udder in dairy ewes. *J. Dairy Sci.*, 78: 2753-2759.

Gonzalo, C., Carriedo, J.A., Baro, J.A. y San Primitivo, F. (1994). Factors influencing variation of test day milk yield, somatic cell count, fat, and protein in dairy sheep. *J. Dairy Sci.*, 77: 1537-1542.

Lagriffoul, G., Aurel, M. A., Barillet, F. Bergonier, D., Bernard, J. y Berthelot, X. (1993). Evaluation des coptages de cellules somatiques de brebis de race Lacaune: Resultats preliminaires. Acts del V Int. Symp. Machine Milking Small Ruminants. S. Kukovics Publ., Budapest, Hungary: 110-119.

Lagriffoul, G., Bergonier, D., Berthelot, X., Jacquin, M., Guillouet, P. y Barillet, F. (1994). Facteurs de variation génétiques et non génétiques des comptages de cellules somatiques du lait de brebis en relation avec les caractères laitiers et les mesures portant sur le lait du tank. Int. Symp. "Somatic cells and milk of small ruminants", Bella, Italy.10-15.

Ng-Kawi-Hang, K.F., Hayes, J.F., Moxley, J.E. y Monardes, H.G. (1984). Variability of test-day milk production and composition and relation of somatic cell counts with yield and compositional changes of Bovine milk. *J. Dairy Sci.*, 67: 361-366.

Ptak, E. y Schaeffer, L.R. (1993). Use of test day yields for genetic evaluation of dairy sires and cows. *Lives. Prod. Sci.*, 34: 23-34.

Schutz, M.M., Hansen, L.B., Steuernagel, G.R., Reneau, J.K. y Kuck, A.L. (1990). Genetic parameters for somatic cells, protein, and fat in milk of Holsteins. *J. Dairy Sci*, 73: 494-502.

Swelve, H.H. (1995). The effect of test day models on estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.*, 78: 929-938.

Thompson, R. y Hill, W.G. (1990). Univariate REML analysis for multivariate data with the animal models. Proc. 4th World Cong. Genet. Appl. Livest. Prod., Edinburgh, Scotland XIII: 484-487.

Zhang, W.C., Dekkers, C.M., Banos; G. y Burnside, E.B. (1994). Adjustment factors and genetic evaluation for somatic cell score and relatioships with other traits of Canadian Holsteins. *J. Dairy Sci.*, 77: 659-665.

NUEVOS CARACTERES EN LOS OBJETIVOS DE SELECCIÓN DE LOS PROGRAMAS DE MEJORA

J. L. Campo y M. García Gil

Departamento de Genética y Biotecnología, Instituto Nacional de Investigación Agraria y Alimentaria, Apartado 8111, 28080 Madrid

Introducción

Junto a los caracteres relativos a la cantidad de producto (carne, leche, huevos, fibra), que han mantenido su importancia relativa a lo largo de los años, dos grupos adicionales de caracteres han tenido importancia relativa creciente en los programas de mejora genética. Estos dos grupos se refieren a la calidad del producto y al bienestar de los animales, cuya importancia relativa ha llegado a igualar e incluso a superar a la de los caracteres de producción en algunas especies (EWART, 1993), en las que los primeros han triplicado su importancia relativa en los últimos 30 años, mientras que los segundos la han cuadriplicado. Tanto los caracteres de un grupo como los de otro tienen correlación negativa con los caracteres de producción, existiendo además un marcado antagonismo genético entre calidad y bienestar, por lo que es difícil incorporar todos ellos en el objetivo de un programa de selección. Como ejemplos citaremos la selección en contra de lípidos en el primer caso y la selección para adaptación a ambientes especiales en el segundo caso.

Junto a estos caracteres, se han incorporado nuevos criterios de selección en el grupo de caracteres relacionados con la cantidad de producto, y más concretamente con la eficiencia de transformación de alimento. La selección directa para eficiencia, definida como el cociente entre cantidad de producto (peso corporal por ejemplo) y consumo de alimento no es el mejor procedimiento para mejorarla. Una alternativa a considerar es la selección contra el consumo residual, definido como la diferencia entre el consumo real y el predicho en función de las necesidades de mantenimiento y producción. Este procedimiento ha sido estudiado en gallinas

(SCHULMAN y col., 1994), cerdos (MRODE y KENNEDY, 1993), vacuno de carne (JENSEN y col., 1992), y vacuno lechero (KORVER y col., 1991).

Dentro de los caracteres relacionados con el bienestar animal suelen utilizarse indicadores del nivel de estrés, ya sean medidas fisiológicas o medidas relativas al comportamiento. Las medidas fisiológicas más utilizadas son el nivel de corticosterona (FREEMAN, 1976, en gallinas; BOANDLE y col., 1989, en vacuno lechero; ZAVY y col., 1992, en vacuno de carne), el ritmo de latidos del corazón (JONES y col., 1981, en gallinas; HARGREAVES y HUTSON, 1990, en ovino), y el cociente entre heterofilos y linfocitos en aves (GROSS y SIEGEL, 1983, en gallinas) o entre neutrofilos y linfocitos en mamíferos (ARAVE y col., 1974, en vacuno lechero; PUPPE y col., 1997, en porcino).

Las medidas de comportamiento suelen indicar niveles de estrés psicológico (miedo) y los procedimientos pueden agruparse en tres grupos, basados en impedir el movimiento del animal, en la exposición a un objeto nuevo, y en el contacto con la gente (manejo), respectivamente. Los primeros se basan en la aversión que presentan aves (prueba de inmovilidad muscular; GALLUP, 1979) y mamíferos (DUNN, 1990) a posiciones invertidas sobre el dorso. Los métodos más utilizados basados en la novedad son la prueba del campo abierto (colocación del animal en un medio desconocido), utilizada en cerdos (FRASER, 1974), gallinas (FAURE, 1975), vacas (KILGOUR, 1975), y ovejas (MOBERG y WOOD, 1982), incluyendo la prueba de emergencia en este grupo (tiempo de salida de un corral o jaula; MURPHY y WOOD-GUSH, 1978), y la prueba del objeto desconocido (colocación de un cono multicolor por ejemplo en el corral o en la jaula; JONES, 1987), cuantificando los niveles de aproximación-alejamiento. Finalmente, el estrés producido por el contacto del animal con las personas se cuantifica en función de los niveles de ataque-huída (MILLS y col., 1991). Todos estos caracteres suelen desviarse de la distribución normal más o menos intensamente, y exigen tiempo y paciencia por parte del experimentador. Aunque suele considerarse que el comportamiento está determinado principalmente por el medio ambiente,

la importancia de los factores genéticos es similar a la existente en otros caracteres, siendo de magnitud moderada la mayoría de las estimas de heredabilidad.

En este trabajo, revisaremos los experimentos efectuados en nuestro departamento relacionados con la calidad de huevo y el bienestar animal, así como los referentes a la selección para consumo residual de alimento. Los dos primeros se han llevado a cabo en razas autóctonas de gallinas, mientras que el último consta de dos experimentos distintos realizados en *Tribolium castaneum*. Los caracteres de calidad de huevo se refieren a la dureza y al color de cáscara, y al contenido de colesterol en la yema. Los estimadores del nivel de estrés han sido la prueba de inmovilidad muscular y el cociente entre heterofilos y linfocitos.

Material y métodos

Las razas de gallinas utilizadas se mantienen en la Estación Experimental de "El Encín" (Madrid), dentro de un programa de conservación y estudio de recursos genéticos iniciado en 1975 (CAMPO y OROZCO, 1982). Las razas son ponedoras de huevo blanco (Castellana, Andaluza), crema (Prat), o marrón oscuro (Villafranquina), o se usan para producción de carne y puesta de huevo marrón (Vasca). Se incluye en el programa una población sintética (Castellana Codorniz), obtenida a partir del cruzamiento F_2 entre la Castellana y la Prat, seguido de cuatro generaciones de selección para eumelaninas (CAMPO, 1991). El control es una población de Leghorn Blanca, originada por cruzamiento de tres estirpes seleccionadas para puesta (CAMPO y JURADO, 1982).

La inmovilidad muscular se induce fácilmente colocando la gallina sobre su espalda en una cuna de madera con forma de U (JONES y FAURE, 1981). El ave es sujetada durante 15 s, y si permanece inmóvil durante 10 s después de que el experimentador haya retirado su mano, se pone en marcha un cronómetro para medir el tiempo que tarda el animal en levantarse. Si se levanta antes de 10 s, se considera que la inmovilidad muscular no ha sido inducida y se repite el procedimiento. Suelen imponerse restricciones al número total de intentos de inducción (3 ó 4) y al tiempo máximo de inmovilidad (10 ó 15 min).

La sangre necesaria para realizar el conteo de leucocitos se recoge a partir de un pequeño pinchazo en la cresta, extendiendo una gota en un porta de cristal. La tinción se hace usando los colorantes May-Grünwald y Giemsa (LUCAS y JAMROZ, 1961) después de la fijación con alcohol metílico. Se cuentan 100 leucocitos, incluyendo los granulares (heterofilos, eosinofilos y basofilos) y los no granulares (linfocitos y monocitos), en un microscopio con objetivo de inmersión de 100 aumentos, calculando el cociente entre heterofilos y linfocitos. Dentro de los leucocitos granulares, los heterofilos se distinguen por tener las granulaciones alargadas en lugar de redondeadas, mientras que dentro de los agranulares los linfocitos se distinguen por tener el núcleo redondeado en lugar de tenerlo dentado.

La dureza de cáscara se estima por medio de la gravedad específica, la porosidad o la densidad de cáscara. La gravedad específica se mide por flotación (WELLS, 1968), utilizando nueve soluciones de cloruro sódico de densidades 1,068 a 1,100 g/cm^3 en incrementos de 0,004 g/cm^3. La porosidad de cáscara se estima como la pérdida de peso de huevo hasta los 17 días de incubación. La densidad de cáscara expresa el peso por unidad de superficie, estimando ésta en función del peso, la longitud y la anchura del huevo (CARTER, 1975). El color de la cáscara se mide por medio de un reflectómetro. El contenido de colesterol en yema se determina por el método enzimático (SALE y col., 1984).

En cada uno de los dos experimentos con *Tribolium castaneum* (población Consejo), se hicieron cuatro repeticiones con cuatro generaciones de selección por línea. En cada repetición había una línea seleccionada para consumo residual, una línea seleccionada para conversión de alimento, y una línea control sin seleccionar. En el experimento 1, la selección trataba de disminuir el cociente entre el consumo de alimento y la ganancia en peso. El consumo residual de alimento se definió como el consumo menos la regresión genética del consumo sobre la ganancia en peso desde los 7 a los 14 días, y el peso inicial a los 7 días. Esta regresión genética es una alternativa lógica a la regresión fenotípica, y es equivalente a un

índice de selección con restricción total para los caracteres incluidos en la regresión (KENNEDY y col., 1993). En el experimento 2, la selección trataba de disminuir el cociente entre el consumo de alimento y la masa de huevos. El consumo residual se basaba en la masa de huevos puestos entre los días 28 y 30, y el peso adulto a los 28 dias.

Resultados y discusión

No había diferencias significativas entre las razas Villafranquina y Vasca en términos de duración de la inmovilidad muscular (484 y 475 s, respectivamente), mientras que ambas mostraban mayor duración que las otras razas. La duración de la inmovilidad muscular era significativamente menor en la Castellana Codorniz (189 s) comparada con las otras razas. No se encontraron diferencias entre Leghorn (319 s), Prat (302 s), Castellana (261 s), y Andaluza (256 s). Las razas de huevo blanco y crema son de peso ligero y de temperamento nervioso, similares a la Leghorn, mientras que las razas de huevo marrón son de peso medio y de temperamento tranquilo. El resultado encontrado en la población sintética de Castellana Codorniz, sugiere importantes efectos heteróticos para este carácter. Un porcentaje de heterosis mayor todavía se encontró en el cruzamiento F_1 de dos variedades diferentes de la raza Prat. La heredabilidad realizada en la población de Leghorn era moderada (0,32 ± 0,10), lo que unido a lo anterior indica selección direccional intensa actuando sobre el carácter.

El cociente entre heterofilos y linfocitos era significativamente mayor en la Castellana Codorniz (0,56). Las razas Villafranquina y Vasca tenían cocientes significativamente menores (0,32 y 0,27 respectivamente), aunque el valor encontrado en la Villafranquina no difería significativamente de los de la Castellana y la Prat (0,39 en ambas), y estas dos razas no diferían ni del control de Leghorn (0,42) ni de la Andaluza (0,41). El mayor valor encontrado en la Castellana Codorniz, en comparación con las dos poblaciones parentales, indica la presencia de importantes efectos heteróticos, y sugiere un valor bajo para la heredabilidad de este carácter. Las razas con mayor valor del cociente tenían valores significativamente más altos para el número de heterofilos y más bajos para el número de linfocitos y vice versa. La

correlación entre el cociente de heterofilos y linfocitos y la inmovilidad muscular fue negativa y significativamente distinta de cero (-0,64).

La gravedad especifica del huevo era significativamente mayor en la Castellana Codorniz y en la Prat (1,089 y 1,088 g/cm^3, respectivamente), aunque estos valores no diferian significativamente de la estima calculada en la Villafranquina (1,087 g/cm^3). La gravedad específica de la Villafranquina no difería de la de las razas Castellana, Andaluza y Vasca (1,084 g/cm^3 en todas ellas). El control de Leghorn tenía menor dureza de cáscara que todas las razas (1,080 g/cm^3). La porosidad de cáscara era significativamente menor en la Vasca y en la Castellana Codorniz (6,3 y 6,4 g, respectivamente). La raza Prat presentaba mayor valor (7,9 g), siendo intermedios los de las razas Villafranquina (7 g), Castellana (7,2 g) y Andaluza (7,3 g), que no diferían significativamente de la Leghorn (7,4 g). La densidad de cáscara era significativamente mayor en la Prat (86 mg/cm^2), la Villafranquina (85 mg/cm^2), la Vasca (84 mg/cm^2), y la Castellana Codorniz (81 mg/cm^2), mientras que la Castellana (79 mg/cm^2) y la Andaluza (78 mg/cm^2) no diferian del control de Leghorn (77 mg/cm^2). El porcentaje de heterosis encontrado en los tres estimadores de la dureza de cáscara era sólo apreciable en el caso de la porosidad, sugiriendo distinta acción genètica para este componente de la dureza en relación con los otros dos. El color de cáscara era mucho más oscuro en la Villafranquina (19 unidades de reflectómetro) que en la Vasca, la otra raza de huevo marrón (32 unidades). Los valores encontrados en la Prat y en la Castellana Codorniz eran intermedios (57 unidades en ambas), mientras que Andaluza (66 unidades) y Castellana (70 unidades) presentaban valores algo inferiores a los de la Leghorn (78 unidades).

El contenido de colesterol en la yema era significativamente menor en la Castellana Codorniz (220 mg/huevo), presentando un porcentaje de heterosis apreciable con respecto a las razas parentales. Una de estas razas, la Prat, y la Villafranquina tenían significativamente menor contenido en colesterol (251 y 252 mg/huevo, respectivamente) que el control de Leghorn (276 mg/huevo). La Castellana (263 mg/huevo) no difería significativamente de la

Leghorn, mientras que el contenido de colesterol en la Vasca (304 mg/huevo) era el mayor de todos.

En ambos experimentos con *Tribolium* el uso del consumo residual como criterio de selección para disminuir la conversión de alimento, era más eficiente que la selección directa basada en el cociente. En el experimento 1, la respuesta obtenida para el cociente entre consumo y ganancia en peso era significativamente mayor en la línea seleccionada para consumo residual de alimento (-16,01 ± 3,56), en comparación con la obtenida en la línea seleccionada directamente para el cociente (-4,75 ± 1,97). En el experimento 2, las dos líneas diferían significativamente para el cociente entre consumo y masa de huevos, siendo el valor observado en la línea seleccionada para consumo residual casi el doble que el correspondiente a la línea seleccionada directamente para el cociente (-113,49 ± 16,39 y -64,75 ± 16,93, respectivamente). La superioridad del consumo residual utilizado como criterio de selección se mantenía al considerar este carácter como objetivo de selección. En ambos experimentos, las respuestas obtenidas en los caracteres utilizados para estimar el consumo residual (peso inicial y ganancia en peso en el experimento 1, y masa de huevos y peso adulto en el experimento 2), no eran significativamente distintas de cero.

Referencias

ARAVE, C. W., J. L. ALBRIGHT, C. L. SINCLAIR. 1974. *J. Dairy Sci.* 59, 1497-1503.

BOANDLE, K. E., J. E. WOHLT, R. V. CARSIA. 1989. *J. Dairy Sci.* 72, 2193-2199.

CAMPO, J. L. 1991. *Poultry Sci.* 70, 1469-1473.

CAMPO, J. L., J. J. JURADO. 1982. *Proc. 2nd WCGALP* 8, 869-874.

CAMPO, J. L., F. OROZCO. 1982. *Proc. 2nd WCGALP* 6, 88-93.

CARTER, T. C. 1975. *Br. Poultry Sci.* 16, 541-543.

DUNN, C. S. 1990. *Vet. Rec.* 126,522-527.

EWART, J. 1993. *Br. Poultry Sci.* 34, 3-10.

FAURE, J. M. 1975. *Ann. Genet. Sel. Anim.* 7, 197-204.

FRASER, D. 1974. *Appl. Anim. Ethol.* 1, 3-16.

FREEMAN, B. M. 1976. *W. Poultry Sci. J.* 32, 249-256.

GALLUP, G. G. 1979. *Anim. Behav.* 27, 316-317.

GROSS, W. B., H. S. SIEGEL. 1983. *Avian Dis.* 27, 972-979.

HARGREAVES, A. L., G. D. HUTSON. 1990. *Appl. Anim. Behav. Sci.* 26, 91-96.

JENSEN, J., I. L. MAO, B. B. ANDERSEN, P. MADSEN. 1992. *J. Anim. Sci.* 70, 386-395.

JONES, R. B. 1987. *Br. Poultry Sci.* 28, 319-326.

JONES, R. B., J. M. FAURE. 1981. *Behav. Proc.* 6, 47-55.

JONES, R. B., I. J. DUNCAN, B. O. HUGHES. 1981. *Behav. Proc.* 6, 121-133.

KENNEDY, B. W., J. H. WERF, T. H. MEUWISSEN. 1993. *J. Anim. Sci.* 71, 3239-3250.

KILGOUR, R. 1975. *Anim. Behav.* 23, 615-624.

KORVER S., E. A. EEKELEN, H. VOS, G. J. NIEUWHOF. 1991. *Liv. Prod. Sci.* 29, 49-59.

LUCAS, A. M., C. JAMROZ. 1961. *Atlas of avian hematology*. USDA, Washington, DC.

MILLS, A. D., Y. NYS, J. GAUTRON, J. ZAWADSKI. 1991. *Br. Poultry Sci.* 32, 117-129.

MOBERG, G. P., V. A. WOOD. 1982. *Appl. Anim. Ethol.* 8, 269-276.

MRODE, R. A., B. W. KENNEDY. 1993. *Anim. Prod.* 56, 225-232.

MURPHY, L. B., D. G. WOOD-GUSH. 1978. *Biol. Behav.* 3, 39-61.

PUPPE, B., M. TUCHSCHERER, A. TUCHSCHERER. 1997. *Liv. Prod. Sci.* 48, 157-164.

SALE, F., S. MARCHESINI, P. FISHMAN, B. BERRA. 1984. *Anal. Biochem.* 142, 347-350

SCHULMAN, N., M. TUISKULA, L. SIITONEN. 1994. *Poultry Sci.* 73, 1479-1484.

WELLS, R. G. 1968. *Egg quality*. Oliver and Boyd, Edinburgh.

ZAVY, M. T., P. E. JUNIEWICZ, W. A. PHILLIPS. 1992. *Am. J. Vet. Res.* 53, 551-558.

Resumen

Se resumen tres grupos de experimentos realizados con caracteres de comportamiento, calidad de producto, y eficiencia alimenticia. Los dos primeros se han efectuado en gallinas de puesta: duración de la inmovilidad muscular, cociente heterofilos:linfocitos, dureza y color de cáscara, y contenido de colesterol en yema. La eficiencia alimenticia se ha estudiado en *Tribolium*, considerando selección para consumo residual de alimento.

Selección de ganado caprino utilizando la información del gen de la caseína α_{s1}.

Sánchez-Palma, A.[1] Manfredi E.[2] y Serradilla, J.M.[1]

[1] Departamento de Producción Animal. ETSIAM. Universidad de Córdoba. España.

[2] SAGA - INRA Castanet Tolosan-Toulouse. Francia.

Resumen.

Mediante simulación de Montecarlo, se analiza por un lado las posibles utilizaciones de la información del polimorfismo de la caseína αs1 para la selección de los machos para ser testados por IA y por otro el modelo de indexación a utilizar cuando se incorpora esta información en el modelo de valoración.

Introducción.

Si observamos el contenido proteico de la leche de cabra, podemos distinguir entre las proteinas del suero y las coagulables o caseínas. Dentro de las caseínas, proteínas de la leche que precipitan a pH 4,6 podemos diferenciar 4 tipos: αs1, αs2, β y κ, según su diferente estructura primaria (Angulo, 1997). La caseína αs1 es la fracción más variable en el ganado caprino y presenta un polimorfismo con al menos 13 alelos, 5 de los cuales con una frecuencia apreciable, que se asocian con diferentes niveles de caseína producida. De esta manera los alelos A, B y C, se asocian con una alta producción de caseína y, por tanto, de proteína, el E con una producción media, el F con una producción baja y el O, o nulo, con una producción nula (Grosclaude y col., 1987). Además de esta relación, los polimorfismos de la caseína αs1 también se asocian con a diferencias en las características tecnológicas (pH, tiempo de coagulación, firmeza de la cuajada) en la fabricación de queso e, incluso, con la intensidad de "sabor a cabra" de los quesos madurados (Angulo, 1997).

Serradilla y col. (1992) apuntan tres posibles vías de utilización de estos polimorfismos con el fin de mejorar el contenido de proteína de la leche y su rendimiento quesero: selección poligénica por el contenido de determinadas fracciones de las caseínas ligadas a un mayor rendimiento y calidad del queso (la fracción αs1 y posiblemente la κ), utilización de las variantes alélicas de las distintas fracciones caseínicas como marcadores asociados a una mayor producción de caseína y el empleo de la ingeniería genética.

Larzul (1993) y Barbieri (1995) recomiendan la utilización de la información correspondiente al polimorfismo de la caseína αs1 en los esquemas de selección del ganado caprino. Esta práctica ya se está dando en Francia, al menos para limitar la entrada en el centro de inseminación de aquellos machos con genotipos más desfavorables (Manfredi y col. 1995).

El presente trabajo de simulación se plantea con el fin de comparar, por un lado, los resultados de la indexación de los animales, obtenida con la utilización de un modelo animal incluyendo el efecto del gen mayor como efecto fijo (BCC) (Kennedy y col. 1992) y sin incluirlo (BSC). Por otro lado, se comparan las posibles formas de utilización de la información del polimorfismo de la caseína αs1 para la selección de los machos de IA (machos en prueba de descendencia), de entre los que presentan un mejor índice obtenido con los modelos anteriores. En el se analizan 3 grados de utilización de la información de los polimorfismos de la caseína αs1 que de mayor a menor intensidad de utilización son:

* No se tiene en cuenta para nada. (SS)

* Se utiliza para elegir entre hermanos, candidatos a ser sometidos a la prueba de descendencia por el índice de sus padres en la valoración genética, aquel o aquellos que tienen los mejores genotipos del gen de la caseína αs1 (SH).

* Se utiliza para seleccionar de entre todos los candidatos a ser sometidos a prueba de descendencia, aquellos con genotipo más favorable. (SG).

Material y métodos.

Se realizan simulaciones para generar los valores de los caracteres tasa de proteína (TP, g/kg.), sobre el que se supone un efecto fuerte del gen mayor, y cantidad de materia proteica (MP, kg.) asociado con un efecto débil del gen mayor (Larzul, 1993; Barbieri, 1995).

El modelo de simulación es el siguiente:

$$y_{ijklmn} = \mu + A_j + R_k + E_l + Tp_m + La_n + G_i + g_i + ep_i + \varepsilon_{ijklmn}$$

Donde, y representa el valor del carácter para un individuo (cantidad o tasa de proteína); μ representa la media poblacional; **A** es el efecto del año (con 20 niveles); R es el efecto de la ganadería (con 100 niveles); **E** es el efecto de la estación de parto (con 4 niveles: primavera, verano, otoño e invierno); **Tp** es el efecto del tipo de parto (con 3 niveles: 1, 2 o 3 cabritos); **La** es el efecto del número de lactación (con 10 niveles); **G** representa el efecto del gen mayor de la caseína αs1 (con tres niveles: AA, AF y FF); **g** es valor poligénico del animal para el carácter y se distribuye como N (0, σ^2_g); **ep** es el efecto ambiental permanente que se distribuye como N (0, σ^2_{ep}) y ε es el efecto residual, que se distribuye como N (0, σ^2_ε).

Los efectos de los distintos niveles de los factores fijos (A, R, E, Tp y La) se estimaron con los de los datos correspondientes a cinco años del control de rendimientos de la raza Malagueña, mediante el procedimiento "GLM" del paquete estadístico SAS.

Para simular el efecto del genotipo de la caseína αs1, se supuso que el efecto del gen mayor es un factor fijo más (G) con tres niveles, ya que para simplificar la simulación se consideraron solo tres genotipos (AA genotipo fuerte, FF genotipo débil y AF genotipo intermedio) (Larzul, 1993; Barbieri, 1995). La parte no controlada por el gen mayor se consideró aleatoria (parte poligénica). Además de esta parte aleatoria, se consideraron como factores aleatorios el efecto ambiental permanente (ep) y el efecto residual (ε). Estos valores se generan en el proceso de simulación mediante el procedimiento de Box y Muller (Jurado e Ibañez, 1982)

En la Tabla 1, se muestran los valores de los efectos del gen mayor sobre cada uno de los caracteres, así como la frecuencia de partida de los distintos genotipos:

Tabla 1. Efectos del gen mayor para cada uno de los genotipos y caracteres.*

Genotipos	TP (g/kg.)	MP (kg.)	Frecuencias* (%)
AA	2,39	0,97	4
AF	0	0	32
FF	-2,39	-0,97	64

**Efectos y frecuencias estimados por Barbieri (1995) en la raza Alpina francesa*

Las medias poblacionales de los caracteres MP (25,26 kg.) y TP (28,30 g/kg.) y los diferentes parámetros genéticos que se utilizan para la simulación (Tablas 2 y 3) son estimaciones obtenidas por Barbieri (1995) en la raza Alpina francesa.

Tabla 2. Parámetros genéticos utilizados para la simulación.

	Valores poligénicos		Valores genéticos	
	MP	TP	MP	TP
σ^2	3,55	0,96	3,85	2,79
σ^2_{ep}	1,71	1,23	1,71	1,23
σ^2_{ε}	4,01	0,65	4,01	0,65
h^2	0,38	0,34	0,40	0,60
r	0,57	0,77	0,58	0,86

σ^2 = Varianza genética
σ^2_{ep} = Varianza ambiental permanente
σ^2_{ε} = Varianza residual
r = Repetibilidad.

Tabla 3. Covarianzas utilizadas en la simulación.

	Covarianza poligénica	Covarianza genética
$\sigma_{mp\text{-}tp}$	-0,41 (-0,22) *	0,30 (0,09) *
$\sigma_{ep1\text{-}ep2}$	0,08 (-0,19) *	0,01 (-0,05) *
$\sigma_{\varepsilon 1\text{-}\varepsilon 2}$	-0,64	-0,65

$\sigma_{mp\text{-}tp}$ = Covarianza genética.
$\sigma_{ep1\text{-}ep2}$ = Covarianza ambiental permanente.
$\sigma_{\varepsilon 1\text{-}\varepsilon 2}$ = Covarianza residual.
** Entre paréntesis aparecen las correlaciones.*

El esquema de selección simulado es un esquema basado únicamente en la IA. Se considera un núcleo de selección constituido por 100 ganaderías de distinto tamaño y un centro de inseminación constituido por 40 machos de diferentes edades (10 de 1 año, 10 de 2 años, 10 de 3 años, 5 de 4 años y 5 de 5 años) considerando los machos probados a partir del cuarto año, utilizándose como tales un año más. Por tanto la tasa de selección de los machos en prueba en el centro de IA es del 50%. En cada ganadería se insemina con un 30% de semen de machos en prueba y un 70% de machos probados. Toda la reposición, tanto de las hembras en las ganaderías (20 %) como de los machos para el centro de inseminación, se considera que procede de la IA. Los machos que se ponen en prueba provienen de las inseminaciones de las mejores hembras de cada ganadería con semen de machos probados. Se considera una efectividad del IA del 50 %.

Los modelos utilizados para la valoración son: en un caso el mismo que para la simulación, indexando los animales por la suma del efecto estimado mediante BLUP para el genotipo del gen mayor de la caseína αs1 (G) y la estima BLUP de la parte restante (g, parte poligénica); en el segundo caso, un modelo que no considera el efecto del gen mayor, los animales se indexan entonces por la estima BLUP del valor genético:

$$y_{ijklmn} = \mu + A_j + R_k + E_l + Tp_m + La_n + vg_i + ep_i + \varepsilon_{ijklmn}$$

Donde, vg_i representa el valor genético del animal *i* para el carácter considerado.

La metodología de simulación es similar a la utilizada por Analla en 1996, primero se genera una población base y, posteriormente, se elige en cada ciclo de selección (cada año) la reposición, utilizando el índice que corresponda y teniendo en cuenta la consanguinidad y la correlación poligénica entre ambos caracteres. Por tanto en la valoración genética se tiene en cuenta también la depresión por consanguinidad (Analla, 1996). Se simularon un total de 20 años. Cada simulación se repitió 10 veces modificando la semilla de génesis de valores aleatorios.

Combinando, por una parte, las 3 formas de utilización de la información del polimorfismo de la caseína αs1 para la selección de los machos en prueba (SS, SH y SG) y, por otra, los dos modelos de valoración genética a utilizar (BSC y BCC), tenemos 6 combinaciones posibles. Como se realizaron simulaciones independientes para la selección por cantidad de materia proteica (MP) y por de proteína (TP), tenemos un total de 12 casos para simular con sus correspondientes réplicas.

Resultados y discusión.

En la tabla 4 se recogen los valores genéticos medios de los dos caracteres, después de 20 años de selección para cada uno de ellos, en los diferentes casos simulados. Podemos observar que cuando el criterio de selección es la materia proteica (MP), el que el modelo de valoración incluya o no la información del gen mayor no supone diferencias significativas en cuanto a la ganancia genética obtenida. Mientras que si el efecto es la tasa de proteína (TP), estas diferencias si son significativas y aparecen en favor del modelo que incluye el efecto del gen mayor. En cualquier caso esta claro que cuando el efecto del gen mayor es fuerte (TP), es más favorable para la ganancia genética la selección genotípica de los candidatos para ser testados. Esto no esta tan claro cuando el efecto es débil (MP).

Tabla 4. Valor genético medio de las hembras de reposición para el año 20. Se acompaña del error de Montecarlo (error típico entre replicas).

		MP (kg.)	TP (g/kg.)
BSC	SS	10,11 ± 0,19	*6,41 ± 0,10*
	SH	10,25 ± 0,15	6,69 ± 0,08
	SG	10,20 ± 0,12	**6,93 ± 0,09**
BCC	SS	10,34 ± 0,16	6,76 ± 0,13
	SH	**10,35 ± 0,21**	6,74 ± 0,14
	SG	*9,87 ± 0,23*	6,82 ± 0,08

En las tablas 5 y 6 se recogen los valores del incremento o reducción del valor algunos otros estadísticos de interés, entre el año 0 y el 20, correspondientes a los distintos criterios de selección, los diferentes modelos de valoración y las distintas formas de utilización de la información del genotipo de la caseína αs1.

Tabla 5. Diferencia de valores entre la población base y la existente a los 20 años para el carácter cantidad de materia proteica (MP) de las hembras de reposición.

MP		Valor Poligénico medio (kg.)	Reducción de varianza poligénica (%)	Reducción de varianza genética (%).
BSC	SS	9,62	26,11	*26,34*
	SH	**9,75**	**26,73**	28,04
	SG	9,23	24,84	30,36
BCC	SS	9,52	*23,66*	29,04
	SH	9,46	25,17	29,97
	SG	*8,91*	25,16	**30,84**

En cuanto a la evolución de los valores poligénicos medios para cada año, en general coinciden con la evolución de los valores genéticos medios. Observamos como estos valores son mucho menores a los valores genéticos medios en el caso de la selección para TP (tabla 6 y 4) que cuando se selecciona para MP (tabla 5 y 4).

Tabla 6. Diferencia de valores entre la población base y la existente a los 20 años para el carácter tasa de proteína (TP) de las hembras de reposición.

TP		Valor Poligénico medio (kg.)	Reducción de varianza poligénica (%)	Reducción de varianza genética (%).
BSC	SS	*4,03*	22,86	*66,13*
	SH	4,30	24,82	**75,45**
	SG	**4,53**	*21,39*	73,29
BCC	SS	4,41	**27,81**	75,25
	SH	4,36	24,22	73,83
	SG	4,45	23.94	73,29

Las reducciones de varianza, tanto poligénica como genética, se muestran en las Tablas 5 y 6. Se observa la gran reducción de la varianza genética que se produce cuando el efecto del gen mayor es fuerte (TP) sobre el carácter a seleccionar, mientras que esta reducción es menor si el efecto del gen mayor es más débil. Por contra la reducción en cuanto a la varianza poligénica es del mismo orden. No obstante las reducciones de la varianza observadas son consecuencias de las tendencias genéticas positivas que se observan en todos los caso simulados.

A continuación podemos observar una serie de figuras en las que se representa la evolución durante los 20 años simulados para las hembras de reposición de las frecuencias de los diferentes genotipos AA, AF y FF (figuras 1, 2, 3 y 4) y las varianzas poligénicas y genéticas (figuras 5 y 6). En estas figuras se puede observar la fijación de alelos favorables y la reducción tanto de varianza poligénica como genética.

Figura 1. *Evolución de las frecuencias para los tres genotipos (AA, AF y FF) a lo largo de los 20 años simulados para las hembras de reposición. Caso de la simulación en la que se selecciona para la materia proteica (MP) se valora con un modelo que incorpora la información del genotipo del gen mayor (BCC) y se utiliza esta información para elegir entre hermanos candidatos (SH).*

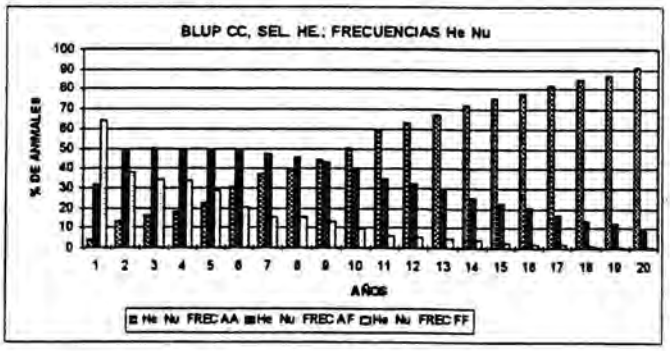


Figura 2. *Evolución de las frecuencias para los tres genotipos (AA, AF y FF) a lo largo de los 20 años simulados para las hembras de reposición. Caso de la simulación en la que se selecciona para la materia proteica (MP) se valora con un modelo que incorpora la información del genotipo del gen mayor (BCC) y se utiliza esta información para elegir entre los candidatos para ser puestos en testaje (SG).*

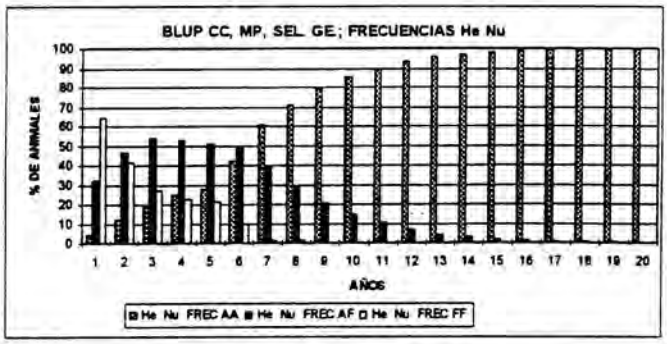


Figura 3. *Evolución de las frecuencias para los tres genotipos (AA, AF y FF) a lo largo de los 20 años simulados para las hembras de reposición. Caso de la simulación en la que se selecciona para la tasa de proteína (TP) se valora con un modelo que no tiene en cuenta la información del genotipo del gen mayor (BSC) y esta información no se utiliza para elegir entre los machos (SS).*

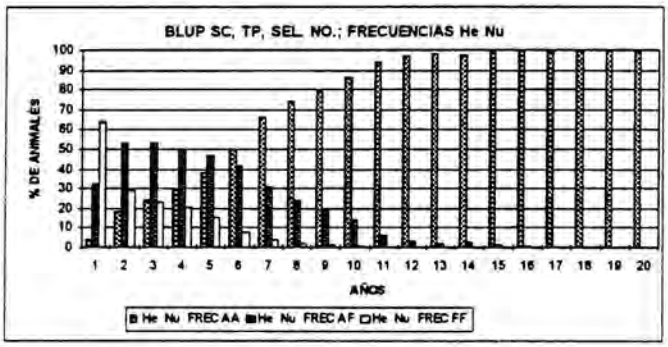


Figura 4. *Evolución de las frecuencias para los tres genotipos (AA, AF y FF) a lo largo de los 20 años simulados para las hembras de reposición. Caso de la simulación en la que se selecciona para la tasa de proteína (TP) se valora con un modelo que no tiene en cuenta la información del genotipo del gen mayor (BSC) y utiliza la información del genotipo de las caseínas para seleccionar los candidatos para entrar en el centro de IA (SG).*

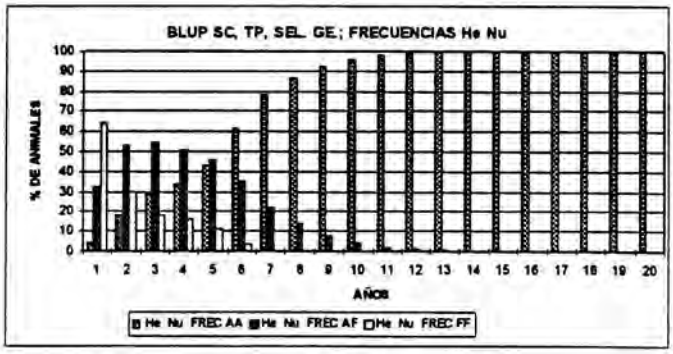


Figura 5. *Evolución de las varianzas poligénicas (VRPMP) y varianzas genéticas (VRGMP) para las hembras de reposición ,durante los 20 años simulados, para la simulación en la que se selecciona para materia proteica (MP), se valora con un modelo en el que se incluye el efecto del gen mayor (BCC) y utiliza la información del genotipo de las caseínas para seleccionar entre hermanos para ser puestos en testaje (SH).*

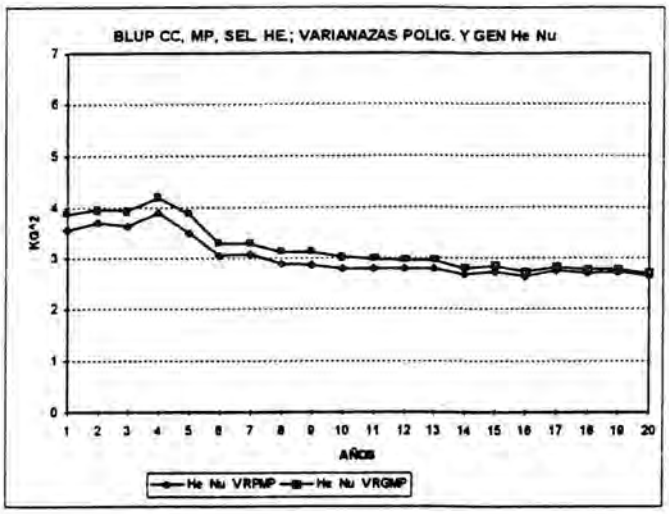
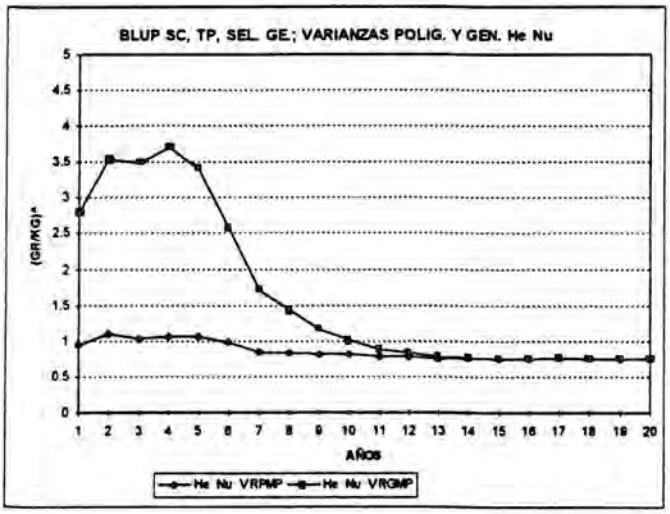


Figura 6. *Evolución de las varianzas poligénicas (VRPMP) y varianzas genéticas (VRGMP) para las hembras de reposición, durante los 20 años simulados, para la simulación en la que se selecciona para la tasa de proteína (TP), se valora con un modelo en el que no se incluye el efecto del gen mayor (BSC) y utiliza la información del genotipo de las caseínas para seleccionar por este los candidatos para ser testados (SG).*



Antes de sacar conclusiones debemos tener en cuenta que para poder utilizar un modelo de valoración en el que se incluya el genotipo de la caseína αs1 como un factor fijo más, es indispensable conocer los genotipos de toda la población cuya información se tiene en cuenta para la valoración de los reproductores. Por el contrario si el modelo que se utiliza para valorar no incorpora la información del gen mayor, solo tendríamos que conocer los genotipos de aquella parte de la población con los mejores índices de valoración genética. A otro nivel podemos asociar la selección para la cantidad de materia proteica, con la selección de un carácter que esta poco afectado (1 vez la desviación típica poligénica) por el gen mayor

(genotipo de la caseína αs1) mientras que la selección para el carácter tasa de proteína esta asociado a un efecto fuerte del gen mayor (3 veces la desviación típica poligénica). Finalmente tenemos que tener en cuenta que partimos de una frecuencia alta del alelo desfavorable, f(F)=0,8. No obstante si tuviésemos unas frecuencias menores únicamente se trataría de desplazarse en el tiempo hasta que con la fijación de las frecuencias nos encontrásemos en el caso correspondiente.

Podemos concluir de acuerdo con Larzul y col. (1997) que por un lado a medida que cuando el efecto del gen mayor es fuerte, existe una gran variabilidad y la heredabilidad del carácter es baja, se debe utilizar la selección genotípica para los polimorfismos de la caseína αs1. Por otro lado teniendo en cuenta el coste de análisis de todos los animales y los beneficios que se obtienen con un modelo de valoración en el que se incluye el efecto del gen mayor (lo cual exige conocer el genotipo de toda la población objeto de valoración) de momento, con los efectos del gen considerados en este trabajo no parece justificado incluir la información genotípica en el modelo de valoración.

Referencias bibliográficas.

Analla, M. 1996. Valoración genética de reproductores y selección de ovino Segureño: Un estudio de simulación. Tesis Doctoral. ETSIAM. Universidad de Córdoba.

Angulo,C. 1997. Análisis del efecto del genotipo de las caseínas sobre las características tecnológicas de la leche de cabra mediante la técnica NIRS. Tesis Doctoral. ETSIAM. Universidad de Córdoba.

Barbieri, M.E. 1995. Polymorphisme de la caseine αs1. Effets des genotype sur des performances zootechniques et utilization en selection caprinejuanma

. Tesis doctoral. INA Paris-Grignon, 161 pp.

Grosclaude, F., Mahé, M.F., Brignon, G. Di Stasio, L. y Jeunet, R., 1987. A mendelian polymorphisme underlying quantitaive variations of goats αs1-casein. Genet. Sel. Evol. 4:399-412.

Jurado, J.J. y Ibañez, V. 1989. Simulación genética en ordenador. Vol III: Teoría de la simulación genética en computador. Monografias de la ETSIA. Num 83. Universidad Politécnica de Madrid.

Kennedy, B.W., Quinton, M. y Van Arendonk, J.A.M., 1992. Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70:2000-12.

Larzul, C. 1993. Propositions pour lámelioration de la selection caprine: definition des nouvelles grilles de qualifications d'index combinés, prise en compte du génotype de la caséine αs1. Rappor INA Paris-Grignon, 39 pp.

Larzul, C., Manfredi, E. y Elsen J.M. 1995. Potential gain including major gene infomation in breeding value estimation. *Gen. Sel. Evol.* 29: 161-84.

Manfredi, E., Ricordeau, G., Barbieri, M.E., Amigues, Y. y Bibé, B. 1995. Génotipe caséine αs1 et index des boucs de testage en race Alpine et Saanen. Gen. Sel. Evol. 27: 451-8.

Serradilla, J.M., Diaz, E., Muñoz, A. y Alonso, A. 1992. Variacion cuantitativa y polimorfismos de las caseinas de la leche de cabra. Posinbilidades de selcción para aumentar el rendimiento en queso. *ITEA* 1:13-22.

DETECCION DE QTLS MEDIANTE LA PARTICION DE LA VARIANZA GENETICA EN FUNCION DEL PARENTESCO ATRIBUIBLE A SEGMENTOS DEL GENOMA.

L. VARONA, M. PEREZ-ENCISO
Centre UDL-IRTA. Area de Producción Animal. 25198. Lleida.

INTRODUCCION

A pesar del gran número de métodos estadísticos disponibles para la identificación de genes que afectan a la expresión de caracteres cuantitativos (QTLS), todos ellos se basan, explícita o implícitamente, en asumir que existe uno o dos genes más un residuo genético infinitesimal. Este enfoque presenta una serie de problemas, a) la distribución de efectos génicos asumida no es realista, el tránsito entre genes de efecto grande y de efecto muy pequeño es seguramente gradual , b) la elección del nivel de significación no es obvia y c) puede no ser un método robusto si las frecuencias alélicas no son las asumidas. Por ejemplo, en una F2, si no hay alelos fijados para el QTL en cada una de las poblaciones origen.

Un enfoque alternativo es identificar regiones que expliquen más variación genética que la esperada bajo un modelo infinitesimal. Esta posibilidad ha sido planteada por Visscher y Haley (1996) y Liu y Dekkers (1998), pero no desarrollaron la metodología necesaria. En este trabajo definimos la teoría y la ilustramos con un ejemplo sencillo.

MATERIAL Y METODOS.

1. TEORIA.

Bajo un modelo estrictamente infinitesimal, la covarianza genética aditiva entre 2 individuos es proporcional al porcentaje de genoma compartido:

$$a(i,i') = \frac{1}{2G}\sum_{H=1}^{2}\sum_{H'=1}^{2}\int_{0}^{G}\delta_{H,H'}(x)dx = \frac{1}{2\sum_{1}^{n_c}G_i}\sum_{i=1}^{n_c}\sum_{H=1}^{2}\sum_{H'=1}^{2}\int_{0}^{G_i}\delta_{H,H'}(x)dx$$

donde H y H' son los haplotipos de cada individuo, G es la longitud del genoma, con

$$G = \sum_{i=1}^{n_c} G_i$$

n_c es el número de cromosomas y δ es una variable indicadora que toma el valor 1 si los haplotipos comparados son idénticos por descendencia en la posición x y 0 en caso contrario. En ausencia de marcadores, a(i,i') es el coeficiente de parentesco de Wright. Por el contrario, si se dispusiese de un numero infinito de marcadores completamente informativos, a(i,i') podría ser calculado con total precisión. En la practica la situación es intermedia.

- a) MARCADORES COMPLETAMENTE INFORMATIVOS.

Si los marcadores son completamente informativos, se conoce con certeza el origen gamético de cada uno de los marcadores y, por lo tanto, la probabilidad de origen de cualquier posición del genoma es sencillamente calculable.

Si dos haplotipos presentan los mismos alelos en ambos marcadores y en la misma fase que sus padres, el coeficiente de parentesco para el segmento de longitud λ del haplotipo j entre el individuo i y el haplotipo k del individuo i' es:

$$\delta(H_{ij},H_{i'k}) = \frac{1}{\lambda}\int_{0}^{\lambda}\frac{\left[(1-r_x)^2(1-r_{\lambda-x})^2+(r_x r_{\lambda-x})^2\right]}{(1-r_\lambda)^2}dx$$

donde r_i es la fracción de recombinación correspondiente a la distancia i.
Si comparten ambos marcadores en fase distinta a sus padres:

$$\delta(H_{ij}, H_{i'k}) = \frac{1}{\lambda}\int_0^{\lambda} \frac{\left[(1-r_x)^2 r_{\lambda-x}{}^2 + (1-r_{\lambda-x})^2 r_x{}^2\right]}{r_\lambda{}^2} dx$$

Si no comparten ninguno de los alelos en los marcadores y los haplotipos se encuentran en la misma fase que sus padres, el parentesco es:

$$\delta(H_{ij}, H_{i'k}) = \frac{1}{\lambda}\int_0^{\lambda} \frac{\left[2(1-r_x)(1-r_{\lambda-x})r_x r_{\lambda-x}\right]}{(1-r_\lambda)^2} dx$$

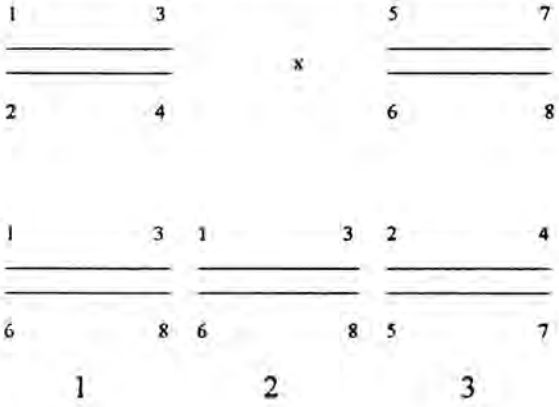
Si no comparten ninguno de los marcadores y los haplotipos se encuentran en distinta fase que los padres:

$$\delta(H_{ij}, H_{i'k}) = \frac{1}{\lambda}\int_0^{\lambda} \frac{\left[2(1-r_x)r_{\lambda-x}(1-r_{\lambda-x})r_x\right]}{r_\lambda{}^2} dx$$

Si los haplotipos sólo comparten uno de los marcadores:

$$\delta(H_{ij}, H_{i'k}) = 0.5$$

Supongamos dos marcadores polimórficos separados por una distancia de 20cM y el siguiente pedigrí:

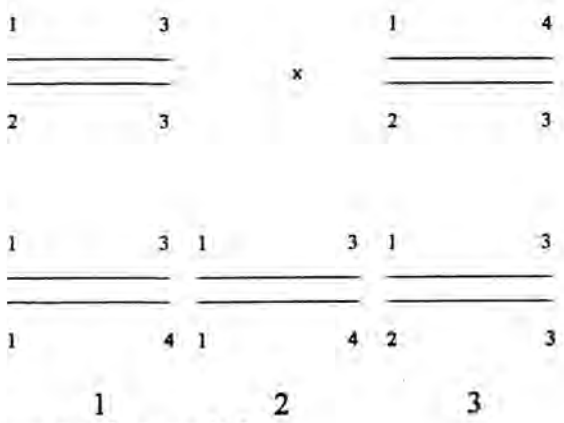


$$a(1,2) = \frac{1}{2} \times \left(\sum_{i=1}^{2} \sum_{j=1}^{2} \delta\left(H_{ij}, H_{i'j}\right) \right) = \frac{0.987 + 0.000 + 0.000 + 0.987}{2} = 0.987$$

$$a(1,3) = \frac{1}{2} \times \left(\sum_{i=1}^{2} \sum_{j=1}^{2} \delta\left(H_{ij}, H_{i'j}\right) \right) = \frac{0.013 + 0.000 + 0.000 + 0.013}{2} = 0.013$$

- b) MARCADORES PARCIALMENTE INFORMATIVOS Y/O FASE DESCONOCIDA.

Cuando los marcadores son parcialmente informativos y se desconoce la fase de los mismos, el origen genético no puede ser definido completamente y es necesario ponderar todas las posibles combinaciones. Supongamos la misma genealogía con marcadores bialélicos:



En este caso, y si de nuevo λ=0.2

a(1,2)=0.86

a(1,3)=0.38

para el segmento considerado.

2. MODELO.

Supongamos:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

donde **y** es el vector de datos fenotípicos, **X** y **Z** son las matrices de incidencia, β es el vector de efectos sistemáticos, **u** es el vector de efectos genéticos aditivos y **e** es el vector de residuos.

$$Var(\mathbf{u}) = \mathbf{A}^{M}\sigma_a^2$$

donde $\mathbf{A}^{\mathbf{M}}$ es la matriz de relaciones aditivas calculadas utilizando la información molecular tal y como se ha descrito. En principio, la estimación de h^2 y de los valores genéticos aditivos será más precisa que utilizando **A** (calculada exclusivamente con la información aportada por el pedigrí). Sin embargo, el incremento de precisión no es muy elevado si la longitud del genoma es suficientemente grande.

Más interesante es discernir cuál es la contribución de cada segmento del genoma a la variación genética total. Por ejemplo, se puede describir un modelo que atribuya una parte de la variación de origen genético a cada cromosoma.

$$\mathbf{u} = \sum_{i=1}^{n_c} \mathbf{u}_i \qquad Var(\mathbf{u}_i) = Diag\left\{\mathbf{A}_i\sigma_{a_i}^2\right\}$$

3. EJEMPLO. ESTIMACION DEL PORCENTAJE DE VARIACION ATRIBUIDO A CADA CROMOSOMA.

Simulación.

Se ha simulado una población consistente en 100 machos y 500 hembras, en un diseño jerárquico con 5 hembras cruzadas con cada macho y 5 descendientes por hembra, generándose familias de 25 medio hermanos en grupos de 5 hermanos completos. Para cada individuo se han generado 4 cromosomas con 20, 15, 10 y 5 loci infinitamente polimórficos repartidos homogéneamente en cada cromosoma. Cada

loci explica el 1% de la variación, por lo tanto la heredabilidad simulada es 0.50. En cada cromosoma se han simulado 2 marcadores infinitamente polimórficos (Caso II) y Bialélicos (Caso III).

Estimación.

Se ha estimado la heredabilidad utilizando un modelo animal que no hace uso de la información de marcadores moleculares (Caso I), además se han estimado la heredabilidad atribuida a cada uno de los cromosomas utilizando 2 marcadores infinitamente polimórficos por cromosoma (Caso II) y dos marcadores bialélicos (Caso III), cuya frecuencia génica era 0.5. Los estimadores se han obtenido mediante inferencia bayesiana con información a priori vaga.

Tabla I. Heredabilidad estimada (Media ± Desviación Típica Posteriores).

Caso	Infinitesimal	Crom. I	Crom. II	Crom. III	Crom. IV
I	0.475 ± 0.035	-	-	-	-
II	0.023 ± 0.017	0.184 ± 0.032	0.175 ± 0.030	0.110 ± 0.027	0.035 ± 0.020
III	0.045 ± 0.034	0.213 ± 0.066	0.096 ± 0.060	0.092 ± 0.043	0.058 ± 0.045

La heredabilidad total es estimada con precisión por los tres procedimientos. Sin la utilización de marcadores moleculares no es posible discriminar entre el origen cromosómico de la variación. Cuando se utilizan marcadores moleculares sí es posible discriminar entre los orígenes de la variación, aunque la precisión que se obtiene depende de la cantidad de información disponible en los marcadores. En el Caso II, la cantidad de información polimórfica (PIC) es 1, mientras que en el Caso III, su PIC es 0.375. Este hecho se refleja en la desviación de la distribución posterior marginal del porcentaje de variación explicada entre los casos II y III (Por ejemplo, 0.032 vs. 0.066 para el primer cromosoma).

La información molecular permite, además de atribuir las causas de la variación de origen genético a distintos segmentos de ADN, refinar la matriz de parentesco y, por lo tanto, mejorar la capacidad predictiva del mérito genético global, que puede ser usado como criterio de selección. En la Tabla II, se presenta la correlación entre el mérito genético simulado y estimado para los tres casos considerados. Como se observa, con un marcador molecular con información polimórfica baja (Caso III), el aumento de la precisión es despreciable (1.6%), por el contrario, con un marcador completamente informativo se consigue un incremento del 6.8% en la precisión (Caso II).

Tabla II. Correlación entre valores genéticos simulados y predichos.

Caso	r^2
I	0.617
II	0.659
III	0.627

CONCLUSIONES Y PERSPECTIVAS FUTURAS.

Los marcadores moleculares permiten refinar el parecido genético entre individuos para cada segmento de ADN, lo que permite calcular los porcentajes de variación atribuibles a dicho segmento, así como mejorar la capacidad predictiva de los modelos de evaluación. Incluso cuando no existen loci de gran efecto, pero los loci de efecto pequeño están repartidos desigualmente por el genoma, la información molecular se manifiesta de utilidad. Pese a todo, es necesario disponer de marcadores muy informativos

para que estas ventajas se hagan evidentes. El procedimiento ha resultado eficiente en la estimación de variación genética a segmentos de ADN, por lo que puede ser utilizado para la detección de loci de variación cuantitativa (QTL). Con respecto a otras aproximaciones tiene la ventaja de que es fácilmente generalizable al análisis de varios caracteres simultáneamente y que, en principio, será más robusto que el enfoque clásico si existe pleiotropía y ligamiento.

Referencias

- Visscher, P.M. y C. Haley. 1996. Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. Theoretical Applied Genetics. 93:691-702.

- Lui, Z y J. C. M. Dekkers. 1998. Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. Genetics. 148:495-505.

Agradecimientos

Trabajo financiado por los proyectos UE PL96-2243 y CICYT AGF96-2510.

TEORÍA DE LOS JUEGOS Y OPTIMIZACIÓN DE LOS PROGRAMAS DE MEJORA: FÁBULA DEL MEJORADOR EGOÍSTA Y DEL MEJORADOR ALTRUISTA"

G Yagüe, L A García-Cortés, C Moreno y J Altarriba
Unidad de Genética Cuantitativa y Mejora Animal
Facultad de Veterinaria. Universidad de Zaragoza
c/ Miguel Servet, 177. Zaragoza-50013

Resumen: Se analiza el comportamiento de varias empresas de genética que, diseñando simultáneamente sus pruebas de descendencia, tratan de competir en un mismo mercado. Este trabajo muestra que el equilibrio comercial, es decir, el tamaño de familia que maximiza los resultados económicos para cada empresa, no coincide con el tamaño de familia que maximiza el progreso genético. El equilibrio comercial, en general, lleva a tamaños de familia menores. Se muestra también cómo las empresas inocentes, que no hacen uso de la teoría de juegos, verán sus cuotas de mercado arrebatadas por empresas que, desoyendo los consejos de la genética cuantitativa, tratan de maximizar sus resultados particulares a costa del progreso genético común.

Introducción

En este trabajo se estudia una prueba de progenie con los recursos de testaje limitados. Habitualmente en este tipo de problemas, el tamaño de familia se determina atendiendo al criterio de maximizar el progreso genético (Hill, 1971).
Cuando varias empresas compiten por el mismo mercado, su interés se centra en conseguir animales mejores que los de la competencia. Estas empresas pueden pensar que optimizar unilateralmente el progreso genético las hará más competitivas, sin embargo, disponen de una herramienta más formal para analizar el problema: La teoria de juegos (von Neumann y Morgestern, 1944). Una revisión interesante sobre el tema puede encontrarse en Rasmusen (1994).

Las reglas del juego

Presentaremos esta sección directamente sobre un ejemplo. Dos empresas, la primera con 1000 plazas y la segunda con 500 plazas ($T=1500$, $T_1=1000$ y $T_2=500$), realizan una prueba de progenie (una sola generación). El mercado escogerá los 10 mejores sementales tras la prueba ($N=10$). La heredabilidad es 0,25. Ambas empresas, independientemente de la superioridad genética de los sementales seleccionados, buscan que sus sementales queden escogidos por delante de los de la competencia.

Ambas empresas conocen todas las reglas del juego excepto el tamaño de familia que emplearán (tendrán que decidirlo) y el tamaño de familia de la competencia (tendrán que adivinarlo). En este juego, el tamaño de familia que cada empresa utiliza se denomina su "estrategia".
Llamaremos n_1 y n_2 a los tamaños de familia o estrategias, S_1 y S_2 a la cantidad de sementales probados ($S_i=T_i/n_i$).

Cómo jugar

Primero, el juego propuesto es un juego se suma cero. Es decir, lo que gana una empresa lo pierde la otra y viceversa. En estos juegos no hay ninguna razón para cooperar. La mejor manera de plantear el juego es, pues, competir.

En el juego propuesto, es de esperar que dos terceras partes de los sementales seleccionados hayan sido probados por la empresa número 1 y una tercera parte hayan sido seleccionados por la empresa número 2. Definiremos los pagos del juego como el número de sementales que consigue colocar en el mercado una empresa por encima o por debajo de su número esperado. O sea:

$$P_i = N_i - \frac{NT_i}{T}$$

Definimos también la matriz de pagos como una tabla que representa los pagos para ambas empresas en cualquier combinación de estrategias. Esta matriz de pagos puede ser obtenida por todos los jugadores antes de determinar su estrategia. Esta matriz se presenta en la Tabla 1.

Estrategia de la empresa 1

Estrategia de la empresa 2	2	4	6	**8**	10	12	14	16	18	20
2	0	1.591	1.942	**1.980**	1.884	1.713	1.494	1.246	.981	.708
4	-1.745	0	.494	**.609**	.563	.434	.260	.061	-.150	-.365
6	-2.087	-.487	0	**.132**	.110	.007	-.139	-.308	-.491	-.678
8	**-2.014**	**-.578**	**-.128**	**0**	**-.016**	**-.106**	**-.237**	**-.391**	**-.557**	**-.729**
10	-1.796	-.509	-.101	**.015**	0	-.084	-.206	-.351	-.507	-.669
12	-1.529	-.375	-.006	**.097**	.080	0	-.117	-.254	-.403	-.558
14	-1.255	-.215	.118	**.209**	.190	.112	0	-.132	-.275	-.424
16	-.990	-.049	.253	**.333**	.312	.235	.127	0	-.138	-.282
18	-.743	.114	.388	**.459**	.436	.361	.256	.133	0	-.139
20	-.515	.270	.520	**.582**	.557	.484	.382	.264	.135	0

Tabla 1. Matriz de pagos para el juego del duopolio. Esta matriz representa los pagos para la empresa 1 solamente, ya que los pagos para la empresa 2 son iguales pero de signo contrario.

Nótese que para obtener los pagos, es necesario en cada caso buscar el punto de truncamiento adecuado que deje 10 sementales en el lado derecho. Determinar la posición de este punto de truncamiento puede hacerse por tanteo usando un método numérico (Burden y Faires, 1985).

El punto de truncamiento ha de cortar las distribuciones de los índices de selección obtenidos por los sementales de ambas empresas. Estos sementales son en cada caso una muestra de S_i individuos con media nula y varianza igual a

$$\sigma_i^2 = \frac{0.25 n_i h^2}{1 + 0.25(n_i - 1)h^2}$$

El juego consiste en que, atendiendo a la tabla 1, cada empresa debe adivinar cuál será la estrategia que la empresa contraria elegirá, y decidir entonces la estrategia propia que maximice sus pagos.

Resultado del juego del duopolio

Aunque cada empresa tiene libertad para elegir su estrategia, un análisis razonable del juego lleva a una elección unívoca de las estrategias óptimas de ambas empresas. La empresa 1 elegirá como estrategia un tamaño de familia de 8. Si nos fijamos en la tabla 1, la columna que da mejores pagos es la columna 8. Así, la empresa 1 no tendrá dudas en elegir esta estrategia. La empresa 2, a su vez, sabe que la empresa 1 elegirá 8, luego el mejor pago para ella está en la celda (8,8), ya que cualquier otro valor de esta columna da pagos positivos para la empresa 1 y negativos para ella. El resultado del juego si ambas empresas quieren competir es (8,8) y los pagos son cero para ambas empresas, es decir, cada empresa consigue colocar en el mercado una proporción de sementales que corresponde a la proporción de sus recursos de testaje.

Podemos representar la tabla 1 en forma de gráfica (Figura 1). Esta figura tiene forma de silla de montar, y el equilibrio comercial (8,8) corresponde a maximizar esa superficie en una dirección pero minimizarla en la dirección perpendicular.

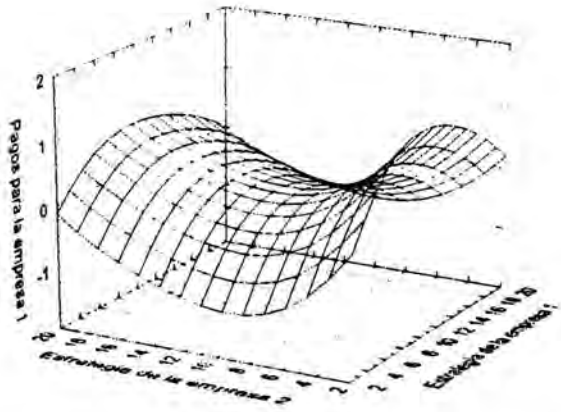


Figura 1. *Pagos para la empresa 1 en el juego del duopolio.*

Vamos a ver el resultado del juego un poco más en detalle. Ambas empresas pueden calcular, usando métodos numéricos, que con 1500 plazas y 10 sementales, el tamaño de familia que maximiza el progreso genético es 14. Entonces, ambas empresas pueden escribir la Tabla 2, que se presenta a continuación:

n_1	n_2	N_1	N_2	P_1	P_2	$E(\Delta G)$
8	14	6.876	3.124	0.209	-0.209	1.215
14	8	6.430	3.570	-0.237	0.237	1.229
8	8	6.667	3.333	0	0	1.201
14	14	6.667	3.333	0	0	1.242

Tabla 2. Estrategias simultáneas de ambas empresas, sus pagos y el progreso genético obtenido de esas combinaciones.

Razonemos como la empresa 1: Si la empresa 2 escoge 14, ya que quiere maximizar el progreso genético, mi estrategia puede ser 8 o 14. Si escojo 8, conseguiré colocar entre los seleccionados 6,8 sementales y si escojo 14, sólo colocaré 6,6. Por tanto, si la empresa 2 escoge 14, a mi me interesa escoger 8. Por otro lado, si la empresa 2 escoge 8, ya que quiere maximizar sus resultados comerciales, mi estrategia puede ser 8 o 14. Si escojo 8, conseguiré colocar entre los seleccionados 6,6 sementales y si escojo 14, sólo colocaré 6,4. Por tanto, si la empresa 2 escoge 8, a mi me interesa escoger 8. Por tanto, haga lo que haga la empresa 2, a mi me interesa escoger 8.

Al razonar como la empresa 2, el argumento se repite de igual forma. Por tanto, aunque ambas empresas obtienen los mismos pagos en (8,8) que en (14,14), dos empresas egoistas forzarán al competir una situación tal como (8,8), reduciendo el progreso genético un 3% sin que ninguna de ellas saque beneficio de la competencia. En cualquier caso, una determinada empresa puede jugar 8 esperando que la contraria, bien por altruismo o bien por desconocimiento de la teoría de juegos, juegue 14.

Para la teoría de juegos, este tipo de equilibrios estables se denominan equilibrios de Nash (Nash, 1950). En el caso de juegos en los que las estrategias deban ser elegidas dentro de un rango de valores numéricos, el equilibrio de Nash se denomina más específicamente equilibrio de Cournot-Nash (Binmore, 1992).

Un resultado inesperado de este juego es que el tamaño de la empresa no influye en su decisión, y empresas grandes no tienen motivos para comportarse de forma distinta a las empresas pequeñas, ni pueden forzar jugadas ganadoras.

Más de dos empresas

En el juego anterior, ambas empresas jugaban tras haber calculado todas las posibles combinaciones de estrategias y atendiendo a la superficie de la Figura 1. Cuando hay más de dos empresas, la tabla de pagos es multidimensional y el número de combinaciones crece de forma galopante. Por este motivo, cuando hay varias empresas, éstas tendrán que emplear otros procedimientos para determinar el equilibrio de Nash y jugar de forma óptima.

Planteamos aquí un procedimiento para encontrar ese equilibrio teniendo en cuenta que la superficie multidimensional ha de ser optimizada para cada empresa en su eje de coordenadas. Puede utilizarse procedimientos similares a los de maximización de funciones como los presentados en Press (1986), pero teniendo en cuenta que la función de pagos es particular para cada empresa.

Presentaremos el algoritmo sobre un ejemplo, supongamos el mismo caso del juego anterior, pero ahora juegan 5 empresas cuyos recursos de testaje son 700 plazas, 400 plazas, 200 plazas, 100 plazas y 100 plazas respectivamente.

El algoritmo que hemos empleado es el siguiente (ver Tabla 3): Todas las empresas comienzan jugando 5. La empresa 1 se comporta de forma egoista y calcula que jugando 8,29 obtendrá el máximo beneficio. La empresa 2, que desconfia de la empresa 1, sabe que va jugar 8,29 y calcula que debe jugar 8,46 para maximizar sus beneficios, y así sucesivamente. Cuando todas las empresas se hayan comportado de forma egoista, las estrategias serán 8,29, 8,46, 8,55, 8,59 y 8,64. La empresa 1 todavía puede ir más lejos y utilizar esa información para optimizar

de nuevo su estrategia y obtener 8,64, pero su decisión puede ser anticipada por las demás. Cuando todas las empresas hayan repetido esta operación varias veces, ninguna de ellas podrá mejorar sus pagos con un cambio unilateral de su estrategia, entonces esta combinación de estrategias corresponde al equilibrio de Nash. En la tabla 3 se presenta el resultado de este algoritmo para varios casos con diferente número de empresas.

En cada paso del algoritmo, la empresa correspondiente debe optimizar su estrategia asumiendo que las otras empresas no cambian la suya. Buscar el máximo de esa función puede hacerse con un método numérico como el algoritmo de acotamientos sucesivos (Press et al, 1986).

caso	T_1	T_2	T_3	T_4	T_5	iter	n_1	n_2	n_3	n_4	n_5
1	700	400	200	100	100	1	8.299	8.461	8.551	8.598	8.647
						2	8.649	8.649	8.649	8.649	8.649
						3	8.649	8.649	8.649	8.649	8.649
2	600	300	300	150	150	1	8.261	8.377	8.505	8.574	8.649
						2	8.649	8.649	8.649	8.649	8.649
						3	8.649	8.649	8.649	8.649	8.649
3	850	350	250	50		1	8.357	8.506	8.623	8.648	
						2	8.649	8.649	8.649	8.649	
						3	8.649	8.649	8.649	8.649	
4	900	500	100			1	8.378	8.599	8.648		
						2	8.649	8.649	8.649		
						3	8.649	8.649	8.649		
5	900	600				1	8.378	8.648			
						2	8.649	8.649			
						3	8.649	8.649			

Tabla 3. Resultados de la convergencia del algoritmo para encontrar el equilibrio de Nash en varios casos con diferente número de empresas. La heredabilidad fue 0.25 y el número de sementales seleccionados fue 10.

La tabla 3 presenta también otros 4 casos con diferente número y tamaño de empresas. En los 5 casos, los recursos de testaje totales fueron 1500, y el equilibrio

de Nash fue igual para todos los casos. Esto parece sugerir que el punto de equilibrio depende sólo del cociente entre los recursos de testaje y el número de sementales a seleccionar (T/N), y del coeficiente de heredabilidad, conclusión ésta que recuerda el trabajo de Robertson (1957) acerca del tamaño de familia que optimiza el progreso genético.

Papel del cociente T/N y de la heredabilidad en el equilibrio comercial: En las figuras 2 y 3 presentamos los efectos del cociente T/N y de la heredabilidad en el tamaño de familia escogido por empresas altruistas o egoístas y en el progreso genético obtenido. Se muestra también la aproximación de Robertson para el tamaño de familia óptimo.

$$n = 0.56\sqrt{\frac{T}{Nh^2}}$$

Estas figuras muestran que el equilibrio comercial siempre da un tamaño de familia menor que el que maximiza el progreso genético, especialmente cuando la heredabilidad es baja. El progreso genético obtenido por las 5 empresas egoístas es alrededor de un 3% menor que el obtenido por 5 empresas altruistas.

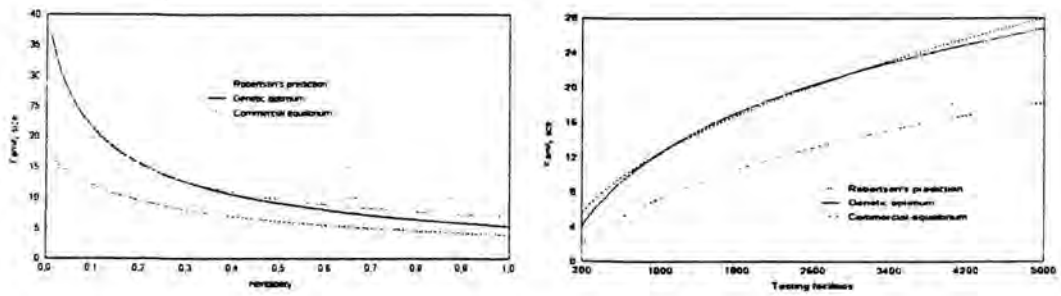


Figura 2. *Óptimo genético, equilibrio comercial y predicción de Robertson en función de la heredabilidad y el cociente T/N.*

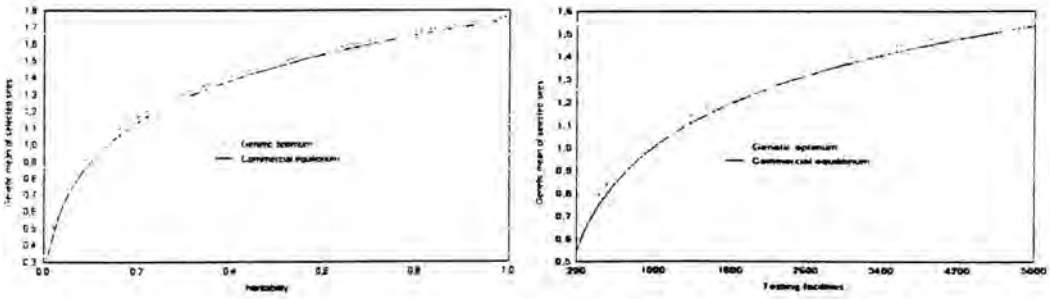


Figura 3. *Progreso genético obtenido cuando las empresas respetan el óptimo genético o el equilibrio comercial*.

BIBLIOGRAFÍA

Binmore K (1992) Fund and games: A text on game theory. D.C. Heath, Lexington. Mass.

Burden RL and JD Faires (1985) Numerical analysis. PWS, Boston, US

Hill WG (1971) Investment appraisal for national breeding programs. Anim. Prod. 71:37-50

Nash J (1951) Non-cooperative games. Ann. Math. 54:286-295

Press WH, BP Flannery, SA Teukolski and WT Vetterling (1986). Numerical recipes: The art of scientific computing. Cambridge Univ. Press. Cambridge, UK

Rasmusen E (1994) Games and information. An introduction to game theory. Blackwell, Oxford, UK.

Robertson A (1957) Optimum group size in progeny testing and family selection. Biometrics. 13:442-450

von Neumann J and O Morgestern (1944) The theory of games in economic behavior. Wiley. New York

ESTIMACIÓN BAYESIANA DE PARÁMETROS GENÉTICOS Y EFECTOS DE PROCEDENCIA DE CARACTERES DE CRECIMIENTO EN *EUCALYPTUS NITENS*

A. Fernández[1], M.C. Rodríguez[1] y R. Ipinza[2]
[1]Dep. de Mejora Genética y Biotecnología, SGIT-INIA, Madrid. [2]Universidad Austral de Chile.

INTRODUCCIÓN

En los últimos años se ha propuesto el uso de la metodología REML/BLUP en mejora genética forestal (WHITE y HODGE, 1989; BORRALHO, 1995) como alternativa a métodos tradicionales como la selección fenotípica o la basada en índices de selección. Sin embargo, este enfoque aún presenta algunos problemas que pueden ser minimizados con el uso de aproximaciones Bayesianas, especialmente cuando la información disponible es escasa, o la estructura genealógica no está bien definida.

En ciencia forestal existen ya algunos trabajos realizados con este enfoque (GREEN y STRAWDERMAN, 1996) pero su uso en programas de mejora no esta todavía muy extendido (SORIA *y col.*, 1998). El presente artículo tiene como objetivo la aplicación de estas técnicas para el análisis genético de caracteres de crecimiento en cuatro ensayos de progenie y procedencias de *Eucalyptus nitens* localizados en la región VIII de Chile, usando un modelo bivariante individual.

MATERIAL Y MÉTODOS

Los datos del presente estudio fueron recogidos por el Instituto Forestal de Chile (INFOR) para evaluar el crecimiento de 162 familias de polinización abierta de *Eucalyptus nitens*, procedentes del sudeste australiano, en distintas regiones de Chile. La colección de semillas involucradas en los ensayos fue adquirida al CSIRO, Australia, en los años 1987 y 1988 (TABLA 1) y asignadas a 13 procedencias distintas.

TABLA 1. Antecedentes de la colección de semillas de *Eucalyptus nitens*.

LUGAR DE ORIGEN		Nº FAMILIAS	LATITUD (SUR)	LONGITUD (OESTE)	ALTITUD (NSNM)
1. Major Point Ebor	NSW	10	30°25′	152°25′	1.450
2. Barrington Tops	NSW	9	31°55′	151°30′	1.450
3. Tallaganda SF	NSW	5	35°48′	149°31′	1.280
4. Tallaganda SF	NSW	7	35°48′	149°31′	1.250
5. Tallaganda SF	NSW	11	35°52′	149°30′	1.105
6. Tallaganda SF	NSW	7	35°52′	149°28′	1.417
7. Anembo Tallaganda	NSW	5	35°54′	149°30′	1.300
8. Tallaganda	NSW	13	36°00′	149°54′	1.050
9. Brown Mt. Nimmitabel	NSW	11	36°32′	149°28′	1.000
10. Glenborg SF	NSW	9	36°38′	149°24′	1.050
11. Toorongo Plateau	VIC.	39	37°47′	146°16′	900
12. Toorongo Plateau	VIC.	16	37°54′	146°00′	1.000
13. Rubicon	VIC.	20	37°18′	145°49′	1.000

Los ensayos se establecieron en cuatro localidades del país, tres de ellas en la región VIII de Chile y la restante en la región X situada más al sur y con unas condiciones climáticas muy diferentes (TABLA 2).

TABLA 2. Identificación de los ensayos considerados en los análisis

ENSAYO	LOCALIZACIÓN REGIÓN, PROVINCIA, COMUNA	ATITUD SU	LONGITUD OESTE	FECHA DE ESTABLECIMIENTO	Nº PROGENIE
1. El Durazno	VIII R, Ñuble, San Carlos	36°26′	71°35′	Ago. 91	1957
2. San Lorenzo	VIII R, Bio Bio, Sta Barbara	37°30′	71°45′	Jul. 90	3259
3. El Morro	VIII R, Bio Bio, Mulchén	37°55′	72°00′	Jul. 90	4485

4. Vista Alegre	X R. Valdivia, Mafil	39°60′	72°50′	Jul. 90	4889

Se empleó el diseño de "*bloques de familias compactas*" equivalente al de parcelas divididas ("*split splot*") en su distribución en el campo. En este esquema, la parcela principal estuvo compuesta por la procedencia, y dentro de ésta se distribuyeron al azar las progenies, representadas por parcelas de cuatro plantas en línea. Cada ensayo estuvo formado por 8 - 10 bloques rodeados por dos hileras perimetrales de plantas de aislación para minimizar el efecto de borde. La altura y el diámetro a la altura del pecho (1.3 m) fueron medidos en todos los árboles supervivientes a los 5 años de edad. Los datos disponibles se muestran en la TABLA 3.

TABLA 3. Registros disponibles en la base de datos.

	TODOS LOS SITIOS	EXCLUIDO SITIO 4
N° Árboles con dato (*n*)	14590	9701
N° Familias	162	162
N° Árboles en pedigrí (n_u)	14752	9863
N° Procedencias	13	13
Media Altura (s.d.), m	8.213 (2.643)	7.995 (2.696)
Media Diámetro (s.d.), cm	8.602 (3.358)	8.361 (3.386)

Interacción Genotipo x Ambiente

La verificación de la no existencia de interacción genotipo x medio es un asunto de gran importancia en mejora forestal debido a la peculiar estructura de los ensayos forestales. En nuestro estudio se emplearon dos métodos de análisis preliminares para examinar la importancia de la interacción procedencia por sitio.

Componentes de Varianza

Para cada uno de los caracteres registrados se realizó un análisis de varianza y se calculó la proporción de la varianza total representada por cada fuente de variación. El modelo empleado fue:

$$Y_{ijklm} = P_i + F_{j(i)} + S_k + B_{l(k)} + PS_{ik} + PB_{il(k)} + e_{ijklm}$$

donde Y_{ijklm} es la altura o diámetro del árbol m^{th} de la familia j^{th} de la i^{th} procedencia creciendo en el bloque l^{th} del sitio k^{th}; P_i es el efecto de la ith procedencia; $F_{j(i)}$ es el efecto de la familia j^{th} dentro de la procedencia i^{th}; S_k es el efecto del k^{th} sitio; $B_{l(k)}$ es el efecto del bloque l^{th} dentro del sitio k^{th}; PS_{ik} es el efecto de la interacción de la procedencia i^{th} y el sitio k^{th}; $PB_{il(k)}$ es el efecto de la interacción de la procedencia i^{th} del bloque l^{th} dentro del sitio k^{th} y e_{ijklm} es el error aleatorio asociado al dato $ijklm^{th}$. La esperanza de los cuadrados medios (EMS) fue calculada utilizando el procedimiento GLM del SAS (SAS Institute Inc., 1988).

En los programas de mejora, uno de los criterios que se siguen para despreciar el efecto de la interacción, cuando esta ha sido significativa, es el indicado por SHELBOURNE (1972). Este autor sugirió que cuando el componente de varianza debido a la interacción es menor a la mitad del debido al genotipo ($\sigma^2_{GXE} / \sigma^2_G < 0.5$), la interacción genotipo x ambiente carece de significación práctica.

Correlaciones genéticas

Otro de los criterios seguidos para estudiar la importancia de la interacción es el valor de las correlaciones genéticas entre un mismo carácter medido en varios sitios. Estas fueron estimadas con el programa VCE (GROENEVELD, 1998). JOHNSON y BURDON (1990) y WOOLASTON *y col.*. (1990) dan como valor umbral 0.8, por encima del cual la interacción genotipo x medio puede despreciarse.

Estandarización de los datos

La heterogeneidad de varianza entre bloques fue corregida para los dos caracteres mediante una estima de la desviación estándar dentro de grupos:

$$\dot{y_{ij}} = y_i + \sigma_B (y_{ij} - y_i) / s_i$$

donde y_{ij} es la observación individual del árbol j^{th} en el bloque i^{th}, y_i es la media del carácter en el bloque i^{th}, s_i es una estima de la desviación estándar dentro de bloques obtenida usando un procedimiento empírico Bayesiano que combina la información de los datos de cada bloque con la información *a priori* del resto de los bloques (WEIGEL y GIANOLA, 1993) y σ_b es la desviación estándar base correspondiente a cada carácter y región Esta se definió como la desviación estándar media del bloque, es decir, valores de 2.07m (altura) y 2.88 cm (diámetro).

Las desviaciones estándar medias de los bloques después de la transformación de los datos fueron de 2.02 m y 2.83 cm para la altura y el diámetro, respectivamente. Aunque la heterogeneidad de varianzas de ambos caracteres se redujo considerablemente después de la transformación, el test de Barlett siguió siendo significativo para ambos caracteres. Actualmente se están estudiando nuevos métodos de corrección de la heterogeneidad de varianzas.

Modelo

La evaluación genética de los arboles se realizó siguiendo el modelo mixto bivariante:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

donde $\mathbf{y}$ es la matriz de observaciones (altura y diámetro) de orden n x 2, $\mathbf{X}$ y $\mathbf{Z}$ son matrices de incidencia conocidas que relacionan $\mathbf{y}$ con los efectos fijos β y aleatorios $\mathbf{u}$, respectivamente; β es una matriz p x 2 de los efectos fijos que incluyen la media y el efecto de bloque; $\mathbf{u}=[\mathbf{u}_1,\mathbf{u}_2]$ es una matriz n_u x 2 de efectos genéticos aditivos; $\mathbf{e}=[\mathbf{e}_1,\mathbf{e}_2]$ es una matriz n x 2 de efectos residuales. Las distribuciones asumidas en este modelo que consideran las procedencias como grupos genéticos fueron:

$$\mathbf{u}|\Sigma_u, A, Q, g \sim N(Qg, A \otimes \Sigma_u), \qquad e|\Sigma_e \sim N(0, I \otimes \Sigma_e)$$

donde A es la matriz de relaciones aditivas; Q es una matriz n_u x 13 cuyos elementos q_{ij} representan la proporción de genes del árbol i derivados de cada procedencia j; g es una matriz 13 x 2 de los efectos de las procedencias para ambos caracteres (QUAAS, 1988) y Σ_u y Σ_e son las matrices de (co)varianzas de los efectos genéticos aditivos y residuales.

Un aspecto fundamental de los programas de mejora genética vegetal que implican polinización abierta es el desconocimiento de los padres. En Eucalyptus existe una complicación adicional debido a la existencia de autopolinización parcial. En las evaluaciones genéticas clásicas (REML/BLUP), las matrices de relaciones aditivas A se construyen asumiendo una estructura familiar de medios hermanos. Posteriormente las estimas de heredabilidad se ajustan para considerar la tasa de autopolinización (s) (BORRALHO, 1995)

$$h^2 = 0.25 h_0^2 / (a - 0.25 F h_0^2)$$

siendo $F = 0.5s$, h_0^2 la heredabilidad obtenida en el análisis REML y a las relaciones aditivas medias entre los árboles

$$a = s^2 + 2s(1-s)0.5 + (1-s)^2 0.25$$

En el presente análisis, las matrices aditivas se construyeron teniendo en cuenta una tasa de autopolinización del 40%, según PÉREZ-ENCISO y FERNANDO (1992), de forma que fue considerada tanto en la estimación de los parámetros como en la evaluación de los individuos.

Para el análisis bayesiano se siguió la metodología propuesta por SORENSEN *y col.* (1994) y extendida a modelos bivariantes usando una generalización multivariante de la distribución χ^2 (distribución Whishart) por VARONA *y col.* (1994). Tanto para los efectos fijos como para los componentes de varianza se emplearon distribuciones *a priori* planas.

En el muestreo de Gibbs se realizaron 506.000 iteraciones, descartándose las 6000 primeras y un intervalo de muestreo de 1/50 salvándose, por tanto, 10.000 muestras de cada parámetro de interés. Para el análisis post-Gibbs se estimaron los perfiles de las distribuciones marginales con estimadores tipo *kernel* (SILVERMAN, 1986). Los análisis se efectuaron sobre una submuestra de 1000 puntos (uno de cada 10) de la muestra de Gibbs.

Los análisis se realizaron con los datos estandarizados y sin estandarizar para evaluar el impacto de la transformación en la predicción de los valores mejorantes y sus implicaciones en la posible selección. Éstos, a su vez, se compararon con los resultados obtenidos con la metodología REML/BLUP para los que se utilizaron los programas de estimas de componentes de varianza VCE (GROENEVELD, 1994) y de evaluación genética PEST (GROENEVELD *y col.*, 1990)

RESULTADOS Y DISCUSIÓN

Interacción Genotipo x Ambiente

Las varianzas de la interacción procedencia por sitio fueron muy significativas para ambos caracteres estandarizados, siendo la proporción de dicha varianza respecto a la varianza total del 1.5% y 1.4% para altura y diámetro, respectivamente. De acuerdo con el criterio propuesto por SHELBOURNE (1972) la interacción no puede descartarse en este estudio. Si bien para la altura el valor de $\sigma^2_{PXS} / \sigma^2_P$ fue 0.312, para el diámetro este valor fue 0.837, superior al umbral de 0.5. En los datos sin estandarizar, el efecto de la interacción fue todavía mayor (0.474 para altura y 1.017 para diámetro).

Las correlaciones genéticas entre los distintos sitios para ambos caracteres se presentan en las Tablas 4 y 5.

TABLA 4. Correlaciones genéticas aditivas (errores típicos) entre altura en diferentes sitios.

	EL DURAZNO	SAN LORENZO	EL MORRO
SAN LORENZO	0.958 (0.056)		
EL MORRO	0.800 (0.087)	0.872 (0.059)	
VISTA ALEGRE	0.672 (0.100)	0.778 (0.068)	0.691 (0.077)

TABLA 5. Correlaciones genéticas aditivas (errores típicos) entre diámetro en diferentes sitios.

	EL DURAZNO	SAN LORENZO	EL MORRO
SAN LORENZO	0.795 (0.089)		
EL MORRO	0.846 (0.066)	0.926 (0.049)	
VISTA ALEGRE	0.766 (0.085)	0.756 (0.074)	0.708 (0.062)

Las correlaciones entre los sitios El Durazno, San Lorenzo y El Morro (situados en la VIII Región de Chile) son mayores al valor umbral definido por JOHNSON y BURDON (1990) y WOOLASTON *y col.* (1990), mientras que entre estos sitios y Vista Alegre (situado en la X Región) son menores.

Según estos resultados procedimos a realizar el análisis de varianza descrito anteriormente excluyendo el sitio Vista Alegre del mismo y el efecto de la interacción perdió importancia ($\sigma^2_{PXS} / \sigma^2_P$ = 0.093 y 0.334 para altura y diámetro respectivamente).

A efectos prácticos habría que utilizar un modelo multicarácter considerando los datos de este sitio como caracteres distintos a los del resto de los sitios, sin embargo debido a la poca información de la que disponiamos se optó por descartar los datos de Vista Alegre en los análisis subsiguientes.

Estandarización de los datos

La influencia de la estandarización de los datos en la evaluación genética se muestra en la TABLA 6. Las estimas de los coeficientes de correlación entre las medias posteriores de los valores genéticos obtenidas con los datos sin tipificar y tipificados para las procedencias familias y árboles fueron muy altas (1.000, 0.997 y 0.997 para altura y 0.984, 0.995 y 0.996 para diámetro, respectivamente). Se observaron cambios muy pequeños en el ranking de las procedencias, familias y en el conjunto de los árboles. IBÁÑEZ *y col.* (1996) demostraron que las predicciones de los valores mejorantes de los datos individuales en bloques con mayor / menor varianzas se espera que disminuyan / incrementen cuando se realiza la estandarización de las varianzas fenotípicas.

TABLA 6. Principales estadísticos del cambio en ranking genético cuando los valores mejorantes predichos son estimados con datos sin estandarizar y estandarizados.(Q^1, Q^2 ,Q^3 = Percentiles 25, 50 y 75, r = correlación de Spearman).

	Media	Moda	Q^1	Q^2	Q^3	Min. , Max.	r
Altura							
Procedencias	0.00	0.0	0.00	0.00	0.00	0,0	1.000
Familias	0.00	1.0	-2.25	0.00	2.00	-10,10	0.997
Arboles	0.00	-9.0	-120.00	4.00	122.00	-1162,1161	0.997
100 mejores	34.07	N.U	-3.00	32.50	74.00	-85,204	0.340
Diámetro							
Procedencias	0.00	0.0	-0.50	0.00	0.50	-1,1	0.984
Familias	0.00	1.0	-2.00	0.00	2.25	-16,17	0.995
Arboles	0.00	-24.0	-140.00	6.00	154.00	-1354,1046	0.996
100 mejores	25.02	N.U	-12.00	9.50	63.75	-66,171	0.553

N.U.: NO ÚNICA

El efecto más importante se observó al analizar el mérito genético de los mejores árboles y, por tanto, de los candidatos a la selección (correlaciones de 0.340 y 0.553 para altura y diámetro respectivamente). La comparación de las dos listas de los valores mejorantes de los 100 mejores predichos por ambos métodos, muestra que solo el 63% (Altura) y el 71% (Diámetro) de éstos están presentes en ambos catálogos. Para un programa de mejora, los cambios en el ranking genético son de particular relevancia si coinciden con diferencias en la superioridad genética de los individuos seleccionados.

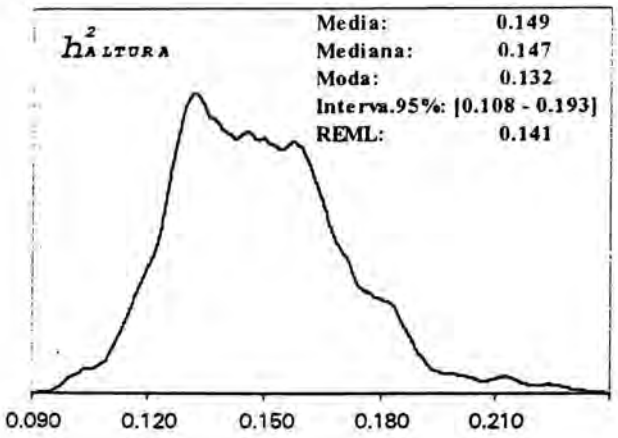
Parámetros Genéticos

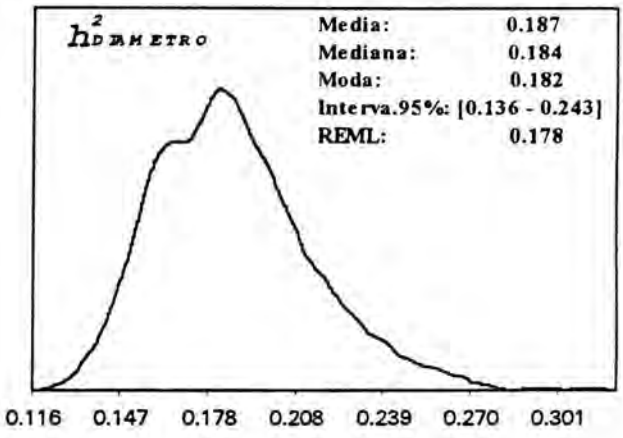
En la FIGURA 1 se muestran las distribuciones marginales posteriores de los parámetros de interés (heredabilidades y correlaciones) obtenidas con los datos estandarizados. También se incluye un resumen de los principales estadísticos de dichas distribuciones, así como las estimas puntuales obtenidas con REML y ajustadas a una tasa de polinización del 40%.

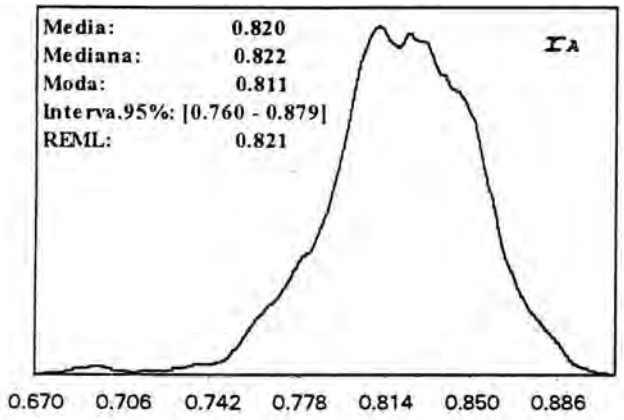
Una de las principales limitaciones de las estimas REML de (co)varianzas es que su distribución no es conocida. Sin embargo con los métodos bayesianos el conocimiento de las distribuciones posteriores de los parámetros permite la construcción de intervalos de confianza. Los intervalos del 95% de mayor densidad posterior son los intervalos de confianza más usuales y fueron calculados para todos los parámetros analizados. Para casi todos ellos las

distribuciones fueron asimétricas.

Como cabría esperar al emplear distribuciones *a priori* planas, las estimas REML fueron muy parecidas a las modas posteriores obtenidas usando técnicas Bayesianas. Además fueron semejantes a las encontradas en la bibliografía (IPINZA *y col.*, 1997).







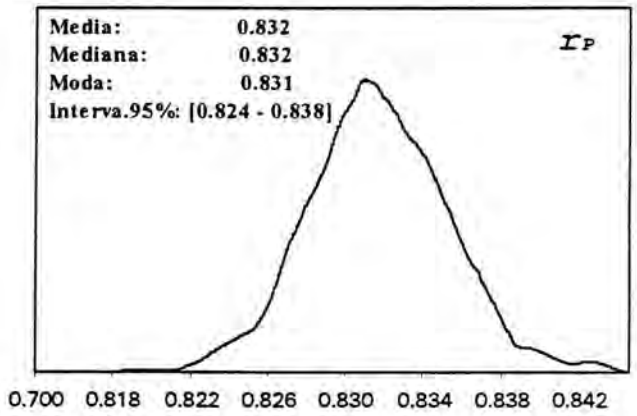


FIGURA 1: Distribuciones marginales posteriores de las heredabilidades de la altura (h^2_{ALTURA}) y diámetro ($h^2_{DIAMETRO}$) y de la correlaciones genéticas (r_A) y fenotípicas (r_p).

Efectos de las procedencias y valores mejorantes de las familias

En el análisis bayesiano las procedencias fueron incluidas en el pedigrí como padres fantasmas en vez de incluirlas como efectos fijos en el modelo de evaluación. Los valores de la media y desviaciones estandar de las correspondientes distribuciones marginales de los efectos de las procedencias Australianas se presentan en la TABLA 7.

TABLA 7. Media y desviación estándar de las distribuciones posteriores de los efectos de las procedencias Australianas y estimas BLUP para altura y diámetro

	Altura		Diámetro	
	Media (s.d.)	BLUP	Media (s.d.)	BLUP
Procedencia				
1.Major Point Ebor	-1.712 (0.188)	-1.665	1.593 (0.302)	-1.542
2.Barrigton Tops	-0.826 (0.199)	-0.803	0.038 (0.312)	-0.010
3.Tallaganda SF	-1.233 (0.261)	-1.210	0.976 (0.401)	-0.957
4.Tallaganda SF	-1.647 (0.226)	-1.586	1.194 (0.364)	-1.110
5.Tallaganda SF	-1.422 (0.179)	-1.376	1.412 (0.296)	-1.361
6.Tallaganda SF	-1.297 (0.207)	-1.270	1.082 (0.327)	-1.049
7.Anembo Tallaganda	-1.311 (0.255)	-1.277	1.008 (0.410)	-0.959
8.Tallaganda	-1.262 (0.176)	-1.214	0.907 (0.279)	-0.910
9.Brown Mt. Nimmitabel	-1.594 (0.185)	-1.546	1.285 (0.292)	-1.230
10.Glenborg SF	-1.386 (0.201)	-1.319	1.056 (0.314)	-0.951
11.Toorongo Plateau	0.000 -	0.000	0.000 -	0.000
12.Toorongo Plateau	-0.300 (0.165)	-0.276	0.410 (0.253)	-0.385
13.Rubicon	0.602 (0.148)	0.653	0.980 (0.225)	1.049

Los efectos de las procedencias en ambos caracteres se calcularon como la diferencia en valores genéticos en relación a la procedencia 11 (Toorongo Plateau) de Australia, considerada la población base al ser la más representada por el mayor número de familias.

No se observaron grandes diferencias entre las estimas puntuales de los efectos de las procedencias obtenidas por REML/BLUP y aquellas obtenidas por el método Bayesiano.

Las medias posteriores de los valores mejorantes de las familias oscilaron entre los valores extremos: -2.974 a 1.424 m (altura) y -3.532 a 2.374 cm (diámetro) y sus desviaciones estandar posteriores variaron de 0.254 a 0.448 m (altura) y de 0.381 a 0.672 cm (diámetro). Estos valores indican la diferente información disponible de las familias de cada procedencia, de manera que los efectos de las procedencia 3 (Tallaganda SF) y 7 (Anembo Tallaganda), con solo 5 familias, y 200 y 325 datos de árboles respectivamente, son las estimadas con mayor error, mientras que la procedencia 13 (Rubicon) con 20 familias y 1047 registros es la estimada con mayor precisión (excluida la procedencia 11 (Toorongo Plateau) con 39 familias y 2493 registros considerada como referencia por este motivo,).

Evaluación BLUP vs. Muestreo de Gibbs

En la comparación entre los ranking de las procedencias y valores mejorantes de las familias y árboles obtenidos por BLUP y muestreo de Gibbs, las diferencias fueron insignificantes (TABLA 8).

TABLA 8. Principales estadísticos del cambio en ranking genético cuando los valores mejorantes predichos son estimados con BLUP o técnicas Bayesianas.(Q^1, Q^2 ,Q^3 = Percentiles 25, 50 y 75, r = correlación de Spearman)

	Media	Moda	Q^1	Q^2	Q^3	Min., Max.	r
Altura							
Procedencias	0.00	0.00	0.00	0.00	0.00	0,0	1.000
Familias	0.00	0.00	-2.25	0.00	1.00	-9,17	0.995
Arboles	0.00	0.00	-84.00	9.00	103.00	-919,2973	0.998
100 mejores	2.48	0.00	-7.00	1.50	10.00	-31,55	0.878
Diámetro							
Procedencias	0.00	0.00	0.00	0.00	0.00	-2,1	0.984
Familias	0.00	-1.00	-3.00	-1.00	2.00	-11,16	0.996
Arboles	0.00	1.73	-95.00	11.00	112.00	-3676,1369	0.998
100 mejores	2.09	0.00	-5.75	0.00	7.00	-26,80	0.934

En consecuencia, el impacto de la técnica de evaluación genética es muy poco importante tanto para las familias (para la mitad de las familias el cambio en posición fue menor a 3) como para el caso del ranking de los 100 mejores árboles (solo 8 para la altura y 5 para el diámetro serían seleccionados por técnicas bayesianas y no estarían presentes si se seleccionara por el mérito genético estimado vía BLUP). La respuesta a la selección, por tanto, no estaría muy condicionada por el método de estimación siendo trivial la diferencia de las medias de los valores mejorantes de los 100 árboles seleccionados estimados vía Gibbs y REML/BLUP.

Los resultados obtenidos confirman la aplicabilidad de las técnicas de modelos mixtos al análisis de datos forestales y evidencian, en nuestra opinión, la riqueza de información que proporcionan las técnicas bayesianas y las ventajas que su futura aplicación en programas de mejora de Eucalyptus podrían conllevar.

REFERENCIAS BIBLIOGRÁFICAS

BORRALHO, N.M.G. 1995. Prooc. of the CRC-IUFRO Conference on Eucalypt Plantations: Improving Fibre yield and Quality, 19-24 Feb. 1995. Australia.

GREEN, E.J. Y STRAWDERMAN, W.E., 1996. For. Sci. 42: 456-464.

GROENEVELD, E. 1990. Programa PEST.Prooc. 4th World Congress of Genetic Applied to Livestock Production, Edinburgh, vol. XII, pp. 488-491.

GROENEVELD, E. 1998. Programa VCE versión 4.2 para SunOS.

IBAÑEZ, M., CARABAÑO, M.J., FOULLEY, J.L. Y ALLENDA, R. 1996. Livest. Prod. Sci. 45: 137-147.

IPINZA R., GUTIÉRREZ, B. Y MOLINA, M. 1997. IUFRO Conference Modelling Growth of Fast_Grown Tree Specie. 3-5 Sep. 1997. Valdivia-Chile.

JOHNSON, G.R., Y BURDON, R.D. 1990. Silvae Genetica 39: 55-62.

PÉREZ-ENCISO, M., Y FERNANDO, R.L. 1992. Theor. Appl. Genet. 84: 173-179.

QUAAS, R.L. 1988. Journal of Dairy Science 71: 1338-1345.

SAS, 1990. SAS/ETS versión 6. Sas Institute Inc. Cary, NC. USA.

SHELBOURNE, C.J.A. 1972. Prooc. IUFRO Genetics-SABRAO Joint Symposia. Japan, Tokyo. pp. 59-65.

SILVERMAN, B.W. 1986. Density estimation for statistics and data analysis. Chapman and Hall.

SORENSEN, D., WANG, C.S., JENSEN, J. Y GIANOLA, D. 1994. Genet. Sel. Evol. 26: 333-360.

SORIA, F., BASURCO, F., TOVAL, G., SILIÓ, L., RODRÍGUEZ, M.C., Y TORO, T. 1998. Canadian Journal of Forest Research (en prensa).

VARONA, L., MORENO, C., GARCIA-CORTÉS, L.A. Y ALTARRIBA, J. 1994. Rev. Portug.

Zootec. 1: 185-195.
WEIGEL, K.A. Y GIANOLA, D. 1993. J. Dairy Sci. 76: 1445-1465.
WHITE, T.L. Y HOGGE, G. 1989. Predicting breeding values with applications in forest tree improvement. Kluwer Academic Publishers, Dordrecht, The Netherlands.
WOOLASTON, R.R., KANOWSKY, P.J. Y NIKLES, D.G. 1990. Silvae Genetica 39: 21-28.

LA CONTROVERSIA BAYESIANA EN MEJORA ANIMAL

A. Blasco
Departamento de Ciencia Animal. Universidad Politécnica de Valencia.
Apartado 22012. Valencia 46071.
ablasco@dca.upv.es

"Un tema constante en el desarrollo de la estadística ha sido la búsqueda de una justificación a lo que los estadísticos hacen. Leyendo los libros de texto uno puede hacerse la idea equivocada de que 'Student' propuso su test porque era el test insesgado uniformemente más potente para la media de una distribución Normal, pero sería más preciso decir que el concepto de test insesgado uniformemente más potente ha ganado mucho de su predicamento porque produce el test t, y como todo el mundo sabe el test t es algo que está muy bien".

DAWID (Comentario en una discusión del artículo de O. Bandorff-Nielsen "Plausibility inference", J. Royal Stat. Soc. Serie B 38: 123-125, 1976)

1. INTRODUCCIÓN

La reciente irrupción de métodos bayesianos en la mejora genética animal ha sumido en una cierta perplejidad a los mejoradores. Los partidarios del uso de estas técnicas aseguran que dan respuesta a problemas mal resueltos en mejora animal, por ejemplo, citando a Gianola et al. (1994) ([1]):

"Supongamos que nos encontramos en el caso donde se aplican análisis de máxima verosimilitud (ML) y BLUP sobre un modelo animal. Con respecto a la primera cuestión [¿Qué puede decirse acerca de la varianza genética en la población base?] obtendríamos una estimación puntual de la varianza genética y una simple medida de incertidumbre que, técnicamente hablando, sólo tendría sentido en muestras grandes y si los datos se distribuyen de forma normal. Respecto a la segunda cuestión [¿Cómo estimar la incertidumbre de los valores de mejora cuando no se conocen las medias y varianzas?] tendríamos que utilizar las estimas máximo verosímiles de las varianzas como si fueran los parámetros verdaderos, ignorando su error. La cuestión relativa

[1] Las citas textuales no figuran en el idioma original. La traducción no es literal

> a la respuesta a la selección [¿Cómo estimar la eficacia de la selección, añadiendo a esa estima una medida de su incertidumbre?] incluso es más complicada. Se pueden derivar estimas de la respuesta a partir de un modelo animal, pero sus propiedades son desconocidas...Una alternativa es adoptar la posición bayesiana. Para cada una de las cuestiones expuestas la respuesta Bayesiana reside en calcular la densidad marginal posterior del parámetro de interés. Esta distribución cuantifica exactamente la incertidumbre acerca del parámetro desconocido."

Pero también conocidos estadísticos de la mejora animal muestran su escepticismo, cuando no su abierta oposición al uso de estos métodos:

> "He indicado por qué algunos estimadores bayesianos tienen propiedades poco afortunadas, pues suelen conducir a estimas de cero para las componentes de varianza entre grupos, en experimentos bien diseñados. Sin embargo aún se advoca por este tipo de estimador". (Thompson, 1987).

> "La controversia entre las escuelas de inferencia frecuentista y bayesiana tiene unas implicaciones enormes. Desde el punto de vista del análisis de datos la consecuencia es que todos los procedimientos expuestos en libros como el de Snedecor y Cochran 'Métodos estadísticos' no es sólo que son malos, sino que además están mal fundamentados. Por otra parte, cada informe sobre un trabajo de investigación debería contener afirmaciones del investigador del tipo '*mi probabilidad* de que un parámetro θ caiga entre 2.3 y 5.7 es 0.90'". (Kempthorne, 1984).

Aunque hasta el momento me he referido a dos escuelas de inferencia (frecuentista y bayesiana), puesto que son las únicas que tienen relevancia hoy en día (al menos en el campo de la mejora genética animal), la inferencia que proviene del examen de la verosimilitud está situada de forma un tanto intermedia entre ambas. Por un lado la verosimilitud juega un papel central en la inferencia bayesiana como portadora de la información que proviene de los datos, y por otro el método de máxima verosimilitud goza de propiedades atractivas desde el punto de vista frecuentista ([2]). Normalmente el mejorador está interesado en soluciones eficaces para estimar el valor de mejora de sus animales, y no necesariamente interesado en la filosofía subyacente, por lo que resulta tranquilizador saber que cuando el número de datos es grande los resultados son los mismos en los tres casos, o que los métodos de mejora animal

---

[2] No he incluido la inferencia fiducial de Fisher o la estructural de Fraser porque no han tenido incidencia en los métodos utilizados en mejora animal y por otra parte están casi abandonadas. El lector interesado puede consultar el primero de los dos artículos sobre inferencia de la Enciclopedia of Statistical Sciences (Kotz y Johnson, 1982), Vol 3, pp.:89-104. Una visión frecuentista de la inferencia se encuentra en Stuart y Ord (1991), mientras que para la visión bayesiana se puede consultar Bernardo y Smith (1994). La visión de verosimilitud se puede encontrar en Edwards (1992). El excelente libro de Barnett (1982) no recoge los últimos avances en la materia, por lo que la comparación de sistemas de inferencia queda un poco anticuada desde el punto de vista práctico.

corrientemente utilizados pueden derivarse de una u otra forma según los presupuestos de una u otra escuela. Robinson (1991) llega a sugerir que en el caso de la estima de valores de mejora la distancia entre escuelas es mínima, y que la apreciación de que no es así se debe esencialmente a la insistencia de los mejoradores adscritos a una u otra escuela en subrayar las diferencias entre ellas más que las similitudes. Creo que estos planteamientos andan lejos de resolver el problema. La estadística tiene como objetivo proponer cuánto conocemos sobre algo y con qué grado de certeza o de duda, por lo que el caso de muestras grandes y representativas es generalmente el caso en el que menos se nota el auxilio de esta rama de las matemáticas ([3]). Además, en la mayor parte de experimentos se intenta maximizar la eficacia respecto a los medios empleados, por lo que no se pretende utilizar grandes muestras sino muestras de un tamaño suficiente como para poder extraer conclusiones relevantes. Aunque es reconfortante saber que el BLUP y el REML tienen una interpretación adecuada en ambas escuelas, las diferencias entre ellas son demasiado notorias, desde el punto de vista práctico, como para ignorarlas: por ejemplo,

1. La forma de presentar los resultados es completamente diferente. La escuela bayesiana trabaja con probabilidades, y ante un problema como el de estimar la heredabilidad su resultado es una distribución de densidad de probabilidad de este parámetro, dada una muestra determinada. La escuela frecuentista aporta ante el mismo problema un valor estimado y un intervalo de confianza. Por supuesto, la escuela frecuentista puede en ocasiones ofrecer las distribuciones de la muestra en el caso de que se repitiera infinitas veces el experimento, o una aproximación -por ejemplo, utilizando técnicas de bootstrap, que últimamente han alcanzado cierta relevancia en algunos problemas de genética molecular- pero la interpretación es completamente diferente, pues la distribución que se obtiene es la del estimador, no la del parámetro sobre el que se intentan hacer inferencias. Además, algunos conceptos de uso frecuente en mejora animal, como el de estimadores insesgados o la diferencia entre efectos fijos y aleatorios, no tienen cabida en la escuela bayesiana ([4]).

---

[3] Esto no siempre es cierto, hay ocasiones en que con una muestra grande el resultado de un test frecuentista es opuesto al de uno bayesiano (paradoja de Lindley, ver p. ej. Press 1989, pp.36-37), pero aún siendo un asunto importante, el tema excede los límites de este artículo.

[4] Esta última afirmación puede resultar chocante al mejorador habituado a trabajar con los métodos BLUP (Best Linear Unbiased Predicition, Henderson 1973) o REML (máxima verosimilitud residual, también llamada máxima verosimilitud modificada o restringida, Patterson y Thompson, 1971). El propio Henderson ha comentado muchas veces que métodos sesgados pueden ser más eficaces que el BLUP (ver Henderson, 1984, por ejemplo). Fisher y Yates consideraban que la distinción entre efectos fijos y aleatorios aportaba más confusión que claridad al análisis de la varianza (ver, p. ej. el prólogo de Yates a la reedición de las obras de Fisher, 1990), y Fisher consideraba la propiedad de insesgamiento un tanto irrelevante (Fisher, 1956, pp. 146-147), esencialmente debido a que no era invariante funcional; (por ejemplo, si bien la esperanza de la media muestral es la media poblacional, la esperanza del cuadrado de la media muestral no es el cuadrado de la media poblacional).

2. Los resultados aportados por estas escuelas no siempre coinciden, particularmente en muestras pequeñas y cuando la escuela bayesiana hace uso de información *a priori*, (ver, por ejemplo, Wang et al., 1994).

3. Algunos problemas mal resueltos en la escuela frecuentista o en la que utiliza la verosimilitud tienen solución en la escuela bayesiana, como hemos comentado al principio al exponer la cita de Gianola et al. (1994).

Dadas las ventajas claras de la escuela bayesiana en los puntos 1 y 3, surgen dos cuestiones obvias: ¿porqué aparecen ahora los métodos bayesianos en mejora animal si la metodología bayesiana es incluso más antigua que los análisis frecuentistas? Y si es una metodología que aventaja sustancialmente a las que hasta ahora se utilizan, ¿porqué no las ha desplazado? ([5]). La respuesta a la primera pregunta es más simple que la repuesta a la segunda. Operativamente hablando, la metodología bayesiana requiere habitualmente, y particularmente en el caso de la mejora animal, la resolución de integrales complicadas frecuentemente sólo accesibles por métodos aproximados (ver, por ejemplo, Cantet et al., 1992). La aparición de métodos nuevos para obtener las distribuciones de probabilidad bayesianas (singularmente el muestreo de Gibbs; ver Sorensen et al., 1994 y Gianola et al., 1994) ha facilitado enormemente el abordaje de estos problemas, haciéndolos operativamente factibles. La respuesta a la segunda pregunta requiere un examen detallado de las características de las dos escuelas en su forma de abordar la inferencia y los problemas de decisión ([6]). El objetivo de este artículo es exponer los puntos de vista de estas dos escuelas de inferencia en lo que atañe a los métodos utilizados hoy en mejora genética animal, con los pros y los contras de las distintas aproximaciones.

2. LA ESCUELA FRECUENTISTA

La escuela frecuentista desarrolló sus métodos en los años 30 y 40. Los nombres más notorios de esta escuela son el polaco Jerzy Neyman, el británico Egon Pearson -hijo de Karl Pearson, uno de los fundadores de la estadística moderna-, quienes trataron los problemas de inferencia, y Abraham de Wald que trató los problemas de teoría de la decisión.

2.1. Inferencia y precisión

Aunque no es en realidad necesario exponer sus fundamentos, dado que sus métodos se explican de forma estándar hoy en día en la mayor parte de cursos de estadística, comentaré

---

[5] O, como se preguntaba Brad Efron en un breve artículo del que lo más interesante es la discusión que lo acompaña: ¿por qué no somos todos bayesianos? (Efron, 1986).

[6] Hay una respuesta más obvia: los mejoradores son educados en Universidades donde sólo se les explica el método frecuentista, presentándose generalmente una caricatura del método bayesiano. De todas formas el fenómeno no es ajeno al "revival" que vive la escuela Bayesiana desde mediados de los 50.

los dos procedimientos de inferencia usados en Mejora Animal: la estimación([7]) y el test de hipótesis (aunque algunos artículos recientes están introduciendo la teoría de la decisión en problemas de Mejora, la producción científica en este campo es aún muy reducida; ver por ejemplo, Wooliams y Meuwissen, 1993). Se considera que el valor del parámetro que se quiere estimar (la heredabilidad de un carácter en una población, por ejemplo) es una cantidad fija determinada, y la inferencia se basa en observar cómo se distribuirían hipotéticas estimas del parámetro de interés si la experiencia se repitiera infinitas veces.

Intervalos de confianza: Si el experimento se repitiera infinitas veces aparecerían infinitos intervalos de confianza que incluirían al valor del parámetro en un 95% de los casos. Obsérvese que cuando se da un intervalo de confianza, por ejemplo $h^2 = 0.20 \pm 0.05$, no se está indicando que la probabilidad de que la heredabilidad se encuentre entre 0.15 y 0.25 es del 95%, sino que si se repitiera el experimento infinitas veces aparecerían infinitos intervalos de confianza (de los que el intervalo [0.15,0.25] es un ejemplo) que contendrían al valor verdadero en un 95% de los casos.

Test de hipótesis: Asumiendo unos riesgos determinados de considerar una hipótesis verdadera cuando no lo es, o de rechazarla cuando es verdadera, se crean unas regiones de aceptación y de rechazo sobre la distribución de los resultados de hipotéticas infinitas repeticiones del experimento, bajo la hipótesis que se quiere contrastar y bajo su alternativa, y se observa si la muestra cae en la región de aceptación o de rechazo . Por ejemplo, cuando la hipótesis es que dos razas tienen la misma media para un determinado carácter, si la diferencia que se observa es lo suficientemente pequeña como para no caer en la región de rechazo, bien se acepta que efectivamente las dos razas tienen la misma media o bien simplemente no se afirma que las razas sean diferentes (dependiendo de la potencia ([8]) del test). Obsérvese que aumentando el tamaño de la muestra se acabaría por rechazar siempre esta hipótesis, por lo que en experimento bien diseñados la cuestión de si dos razas tienen la misma media para un carácter o no, es una cuestión de tipo práctico: antes de empezar el experimento se debe decidir qué valor es lo suficientemente relevante como para considerar esas medias diferentes. El problema surge en experimentos no bien diseñados, en datos de campo o en los caracteres para los cuales no se diseñó el experimento, en esos casos la respuesta del test (aceptar o rechazar si dos valores son diferentes) es insatisfactoria .

2.2. Los índices de selección y el BLUP como estimadores frecuentistas

Los índices de selección hacen su aparición en un artículo de Karl Pearson (1903) sobre selección natural, aunque tardarían mucho tiempo en ser aplicados a la mejora genética

---

[7] No distinguiré entre estimación puntual y por intervalos porque esta distinción no es necesaria para los propósitos del artículo. Fisher, con la aspereza que le caracterizaba, acusó a Neyman y Pearson de inventar el concepto de estimación puntual sólo para, confiando en la ignorancia de los estudiantes, pretender haber traído algo original a la estadística con el concepto de intervalos de confianza (Fisher, 1956, p. 143).

[8] El test que considera dos riesgos es tal vez el más frecuente, aunque experimentadores hay que diseñan con potencias del 50%, lo que equivale a ignorar el de segunda especie.

animal, y expresados de forma simple en los trabajos de Lush sobre selección combinada (Lush, 1947). El BLUP aparece, sin este nombre, en 1949 como un método máximo verosímil, lo que resultó no ser cierto (Henderson 1949, 1950) y condujo a que no se le prestara una importancia especial hasta años más tarde. La historia de su concepción y desarrollo está narrada por Henderson (1973). Examinaremos aquí las propiedades frecuentistas de estos métodos.

Consideremos el modelo lineal

$$\mathbf{y} = \mathbf{X}\,\mathbf{b} + \mathbf{Z}\,\mathbf{u} + \mathbf{e} \qquad (1)$$

donde **y** es el vector de datos, **b** el de efectos fijos (por ejemplo, rebaño-año-estación, covariables, etc), **u** el de efectos aleatorios (valores de mejora de los individuos), **e** el de residuos y **X, Z** son matrices de incidencia. El estimador

$$\hat{\mathbf{u}} = \mathbf{A}'\,\mathbf{V}^{-1}\,(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

donde **A** y **V** son las matrices de varianzas-covarianzas de los efectos aleatorios y de los datos respectivamente, y $\hat{\mathbf{b}}$ es la estima por mínimos cuadrados de **b**, da la *mejor predicción* ([9]) *lineal insesgada (Best Linear Unbiased Prediction)* de **u**. Caso de que no existan efectos fijos, suprimiendo $\mathbf{X}\hat{\mathbf{b}}$ de la fórmula se obtiene la expresión de los índices de selección, por lo que estos pueden considerarse un caso particular.

Invertir **V** no es sencillo cuando se disponen de miles de datos, como es el caso habitual de la mejora genética animal. Henderson (1963) demostró que la solución de las ecuaciones

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

coincidía con las estimas de $\hat{\mathbf{b}}$ y $\hat{\mathbf{u}}$ anteriores, y posteriormente (Henderson, 1976) mostró cómo calcular $\mathbf{A}^{-1}$ sin necesidad de invertir **A**, lo que convirtió en factible al método, que hoy ha alcanzado una difusión universal entre los mejoradores.

El BLUP es, sin embargo, un estimador insesgado en un sentido diferente al que se le da en la estadística frecuentista. Tomemos uno de los componentes de $\hat{\mathbf{b}}$ cualquiera, al que llamaremos $\hat{b}$ (sin subíndice, por comodidad en la notación). Si tomásemos infinitas muestras obtendríamos infinitos $\hat{b}$ que cumplirían

$$E(\hat{b}) = b$$

[9] Henderson introdujo la costumbre de llamar *predicciones* a las estimas de los valores aleatorios. Parece innecesario.

es decir; en cada repetición hipotética del experimento los valores de $\hat{b}$ que se obtienen se agrupan en torno al valor verdadero b. Sin embargo esta condición es imposible de cumplir para los valores aleatorios u, puesto que, por ser aleatorios, cambian en cada repetición hipotética del experimento. Cuando los valores son aleatorios se dice que su estimador es insesgado si

$$E(\hat{u}) = E(u)$$

Esta condición es mucho menos sugestiva que la anterior. Ahora el estimador no se distribuye alrededor del valor verdadero, puesto que en cada repetición del experimento no sólo tengo un $\hat{u}$ diferente sino que también u ha cambiado. Esto afecta también a la interpretación del riesgo de estimación, que es lo que en definitiva se trata de minimizar. En el caso anterior

$$\text{Riesgo}(\hat{b}) = [\text{sesgo}(\hat{b})]^2 + \text{varianza}(\hat{b})$$

ahora

$$\text{Riesgo}(\hat{u}) = [\text{sesgo}(\hat{u})]^2 + \text{varianza}(\hat{u}) + \text{varianza}(u)$$

([10]) La cantidad varianza (u) no se puede modificar, puesto que depende de la naturaleza (habitualmente será la varianza aditiva, dominante o de efectos maternos). El problema está, por otra parte, indeterminado, puesto que hay infinitos estimadores que, según su sesgo y varianza, tienen el mismo valor de riesgo. Para determinar el problema se elige el de varianza mínima de entre los que tienen sesgo nulo (en el sentido que hemos dado a la palabra sesgo cuando hablamos de factores aleatorios) .

Para poder utilizar el método BLUP hay que conocer las matrices **A** y **V**, de lo contrario el método ya no es BLUP. Como su valor verdadero se desconoce, se han hecho varias propuestas para que el método conserve buenas propiedades, más adelante volveré sobre este asunto([11]).

---

[10] Aquí sesgo significa $E(\hat{\mathbf{u}}) - E(\mathbf{u})$, que es una expresión diferente a la que habitualmente utiliza la estadística frecuentista. En el caso de efectos aleatorios el sesgo debería ser $E(\hat{\mathbf{u}}|\mathbf{u}) - \mathbf{u}$, valor distinto de cero, con lo que el método BLUP sería un método *sesgado* . Llamar al BLUP insesgado porque se llama sesgo a otra cosa, parece un uso un tanto liberal del lenguaje.

[11] Es tan habitual considerar al BLUP como "el mejor estimador", que se olvida a veces que sólo es el mejor de entre los de una clase muy concreta, y sólo cuando se conocen las componentes de varianza de los efectos aleatorios sin error, por lo que el uso de la palabra "Best" resulta un tanto equívoco.

2.3. Los métodos basados en esperanzas de formas cuadráticas

El estudio de los componentes de la variación (ANOVA) fue formalizado por Fisher en 1922 sobre bases muy intuitivas, como le gustaba subrayar al propio Fisher. Su formalización se hizo para modelos equilibrados con factores aleatorios únicamente. Tras varios intentos de generalización a modelos desequilibrados por varios autores, Henderson publica en 1953 tres métodos de análisis de varianza de dificultad de computación creciente: el primero trata del análisis de modelos desequilibrados con factores aleatorios y los otros dos ya tratan de modelos en los que hay efectos fijos y aleatorios simultáneamente -modelos mixtos-. El método 2 no admite interacciones entre efectos fijos y aleatorios, pero el método 3 resuelve cualquier modelo. Estos métodos son implementados por Harvey (1960) para estimar heredabilidades y correlaciones genéticas en un programa que fue muy popular entre los genetistas ([12])

Posteriormente se demostró que el método de Fisher era óptimo para diseños equilibrados, pero de los métodos de Henderson, aparte de ser insesgados no se conoce que tengan alguna otra propiedad favorable. Otro intento de encontrar estimas con propiedades adecuadas es el desarrollado por Rao en 1971 con sus estimas insesgadas de varianza mínima (MIVQUE). Este método no es iterativo, pero es función de unos valores previos. El método produce estimas insesgadas de error mínimo sólo cuando los valores previos que se le suministran son los verdaderos componentes de varianza, lo que -al menos para el profano- resulta un tanto peculiar ([13]). El método se basa, además, en la normalidad de los caracteres que se analizan. Cuando no es asi, el método se llama MINQUE porque tiene la propiedad de que minimiza la norma (módulo al cuadrado) del vector que define la distancia entre el valor estimado y el verdadero. Esta última forma de proceder, un tanto oscura, no ha gozado de popularidad no ya entre los mejoradores sino ni siquiera entre los estadísticos profesionales.

Para resolver dificultades de cómputo, Henderson (1984) y Schaeffer (1986) presentaron sendos métodos basados también en esperanzas de formas cuadráticas. Todos estos métodos

---

[12] Una advertencia respecto a esta breve nota histórica y a las que siguen: precedentes hay de casi todo. Alguien aducirá que el MIVQUE fue propuesto por Lamotte en 1970 en notas internas de su Universidad, pero se puede aducir que en 1968 un estudiante de Searle terminó su tesis doctoral sobre MIVQUE para modelos sencillos. De hecho, rastreando el método, en 1906, en un tratado de geodesia, aparece la primera descripción de lo que hoy podemos llamar MIVQUE. En 1861 hay indicios del uso del ANOVA para un factor, y su autor dice que "no es procedimiento que sea nuevo" (ver Schéeffe, 1959). Máxima verosimilitud restringida -con este nombre- se encuentra para un caso simple en Thompson -otro Thompson- 1962, y el título del artículo de Patterson y Thompson (1971) es casi el mismo que el del capítulo 24 del libro de Anderson y Bancroft (1952), quienes ya sugirieron estos procedimientos. Por su parte el método de máxima verosimilitud, aunque propuesto por Fisher y examinadas sus propiedades en 1922, se encuentra antes en el propio Fisher (1912) y en varios autores desde Bernouilli (1782, en Kendall, 1961).

[13] Además pronto se descubrió que sustituyendo en la fórmula como valores iniciales los que se deberían obtener en la primera iteración, se obtenía una fórmula idéntica a una de las soluciones de algoritmos que dan lugar a estimas REML, por lo que -con la notable excepción del propio Rao, quien en 1988 produjo un libro sobre estimación de componentes de varianza prácticamente dedicado al MIVQUE (Rao y Kleffe, 1988), el método ha caído en desuso en Mejora Animal.

gozan de una propiedad típicamente frecuentista: el insesgamiento, pero no son necesariamente los estimadores más eficientes (de mínimo riesgo) y, sobretodo, tienen un inconveniente que los hace particularmente inadecuados para su uso en mejora genética animal: son métodos sesgados cuando los datos están seleccionados (ver, p.ej., Gianola et al., 1989). Este inconveniente es importante, puesto que la mayor parte de bases de datos que se usa en mejora genética animal consiste en datos seleccionados. Escoger sólo a las generaciones base para hacer el análisis no resuelve el problema, puesto que en ese caso es probable que el número de datos sea insuficiente para obtener una precisión aceptable.

2.4. El método de máxima verosimilitud

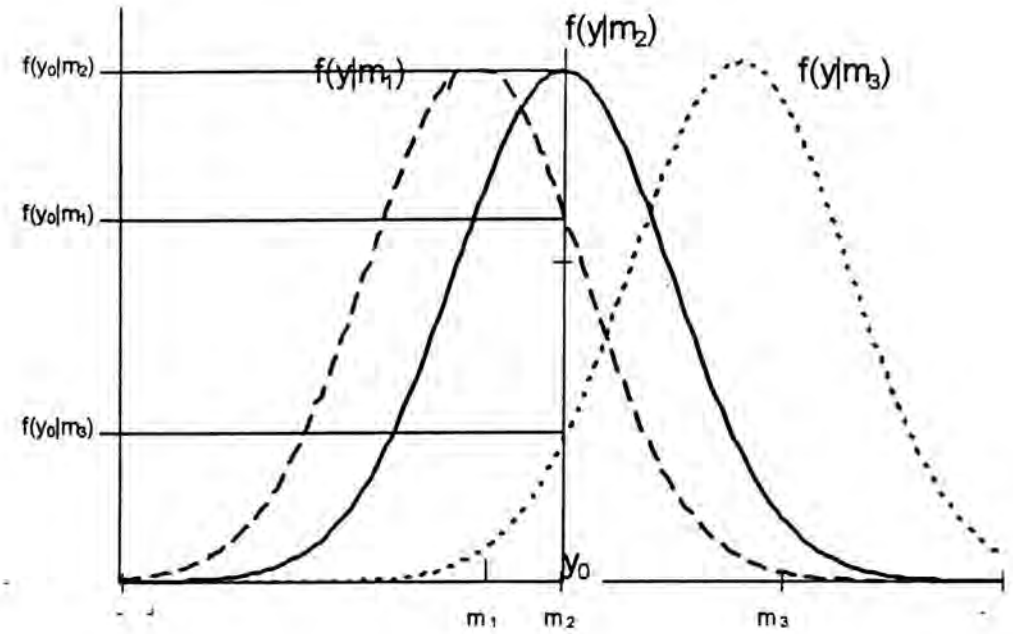
El concepto de verosimilitud y el método de máxima verosimilitud fueron formalizados por Fisher entre 1912 y 1922. En 1912 la teoría de la estimación aún no estaba desarrollada, por lo que no se le podían atribuir al método propiedades particularmente interesantes y el artículo quedó prácticamente ignorado. En 1922 Fisher publica un artículo en el que se definen las propiedades que debe reunir un buen estimador, y se encuentra que el método de máxima verosimilitud produce estimadores con buenas propiedades, por lo que es entonces cuando se produce su aceptación en la comunidad científica (Fisher 1912, 1922).

Supongamos que conocemos cómo se distribuyen las muestras en el espacio muestral (esto es, cómo se agruparían distintas muestras si repitiéramos infinitas veces el experimento). Si llamamos $f(\mathbf{y}|u)$ a la función de densidad de la muestra para un valor de u determinado, el método de máxima verosimilitud consiste en maximizar $f(\mathbf{y}|u)$. Se supone que $f(\mathbf{y}|u)$ es función de u para cada muestra **y**. Para explicarlo recurriremos a un ejemplo sencillo.

Supongamos que quiero averiguar la media del peso de conejos de una determinada raza a las ocho semanas de edad, y tomo una muestra. Para hacer el ejemplo más sencillo supongamos que la muestra es de un sólo conejo que pesa y_0=1.60 Kg. En la figura 1 se representan varias de las posibles poblaciones de las que puede haber sido extraído el conejo, con medias m_1 = 1.50, m_2= 1.60, m_3= 1.80.

Obsérvese que las densidades de probabilidad de la primera o la tercera población $f(y_0|m_1)$ y $f(y_0|m_3)$, son inferiores a la de la segunda población, $f(y_0|m_2)$. Debo reconocer que de todas las poblaciones de conejos posible, la que tiene una media como el peso de mi conejo es la que hace más verosímil que al muestrear haya sacado yo a ese conejo en concreto. Obsérvese que la función de verosimilitud está compuesta por valores $f(y_0|m_1)$, $f(y_0|m_2)$, etc. que forman una curva que no tiene por qué tener ni siquiera el aspecto de una función conocida.

Figura 1. Valores de verosimilitud para distintas medias



Obsérvese también que si represento los valores $f(y_0|m_1)$, $f(y_0|m_2)$,... obtengo una función en la que lo que varía son los valores de m mientras que lo que está fijado es el valor de la muestra, y_0 . En general esto es equivalente a suponer que disponemos de un número muy grande de valores $f(y|m_1)$, $f(y|m_2)$, $f(y|m_3)$,.... Cada uno representa una "probabilidad instantánea"([14]), en el sentido de que $f(y|m_1)$ es, por supuesto, una función de densidad de la muestra y *en unas circunstancias determinadas* (esto es, si y proviene de una población cuya media es m_1), y la probabilidad de que y esté entre dos valores y_0 e y_1 *en esas circunstancias concretas* viene dada por el área entre $f(y_0|m_1)$ y $f(y_1|m_1)$, pero al mismo tiempo es obvio que no se pueden sumar las probabilidades definidas por áreas en el entorno de $f(y_0|m_1)$, $f(y_0|m_2)$, $f(y_0|m_3)$, ..., porque están tomadas de poblaciones diferentes, y el conjunto de estos valores no obedece a las leyes de la probabilidad -no suman 1, por ejemplo- ([15]).

---

[14] Con este nombre es como en el artículo de 1912 aparece la verosimilitud. Fisher no utilizará el término *verosimilitud* hasta mucho después.

[15] Se genera, además, un problema de notación, puesto que hablar de varias funciones de densidad $f(y|m_1)$, $f(y|m_2)$, $f(y|m_3)$, ... para un y dado, es idéntico matemáticamente a hablar de una cierta función $L(m|y)$, con la diferencia de que mientras f es una función de densidad, que representa probabilidades, L ni es de densidad ni representa probabilidades. Problema al que contribuyen textos clásicos como el de de Kendall, utilizando para la verosimilitud la notación $L(x|\theta)$ (Stuart and Ord, 1991), o la de miembros de la escuela frecuentista distinguiendo entre "dado un parámetro" (que siempre está fijado) y "dada la muestra" (que es la variable aleatoria); para el primer caso usan la notación $(x|\theta)$ y para el segundo $(\theta;x)$, con lo que se

Fisher (1912) propone tomar el valor m_i que dé lugar a un valor mayor de $f(y|m_i)$, porque le parece intuitivo que de todas las poblaciones a las que da lugar la serie $f(y|m_1)$, $f(y|m_2)$, $f(y|m_3)$, ... la que tiene mayor valor de $f(y|m_i)$ es la que hace que la muestra que se ha tomado parezca más *probable*. Como aquí la palabra *probable* puede inducir a confusión, puesto que ya hemos dicho que esas "probabilidades instantáneas" no pueden ser tomadas como probabilidades en conjunto, posteriormente dirá que hay que tomar el valor m_i que hace más *verosímil* el que cuando hemos muestreado haya salido esa muestra. Obsérvese que el método de máxima verosimilitud no es, por tanto, el que procura el estimador *más probable* dada la muestra. Literalmente, el método de máxima verosimilitud provee el valor del parámetro que, *de ser verdadero*, haría más probable a la muestra observada ([16]).

Cuando Fisher propuso el método no quedaba claro que este estimador fuera particularmente bueno. Fisher lo propuso porque le parecía que la verosimilitud suministraba un *grado de creencia racional* que, aunque no gozara de las propiedades de una probabilidad, le permitía expresar la incertidumbre de una forma no muy diferente. Fisher proponía en realidad usar toda la curva de verosimilitud y no sólo su máximo, lo que hoy en día es frecuente en algunos problemas de marcadores genéticos. En esos casos se corre el riesgo de acabar interpretando la verosimilitud como si fuera una probabilidad ([17]). En realidad el estadístico frecuentista no usa la curva de verosimilitud sino sólo su máximo. El método ha sido aceptado por sus buenas propiedades frecuentistas -es asintóticamente insesgado, suficiente cuando hay estimadores suficientes, eficiente, óptimo asintóticamente normal, etc.- , pero estas propiedades son asintóticas, y si las muestras son pequeñas no tienen por qué conducir a un estimador particularmente bueno. Por otra parte no es necesariamente el estimador que minimiza el Riesgo([18]). El método presenta, sin embargo, una ventaja aparte de sus propiedades como estimador frecuentista: Cualquier reparametrización conduce a la misma

---

puede encontrar la verosimilitud con cuatro notaciones: $L(\theta|x)$, $L(x|\theta)$, $f(x|\theta)$ y $L(\theta;x)$. La primera de ellas es la que menos confusión ocasiona.

[16] Suponiendo, claro está, que el sistema del que se muestrea genere las muestras con probabilidad máxima. Esta crítica la hizo Euler cuando Bernouilli presentó su trabajo, y es razonable, puesto que suponer esto puede conducir a sesgos en muestras pequeñas. Por ejemplo, en el caso de la varianza, si tomamos una muestra de dos individuos el estimador máximo verosímil tenderá a estimar varianzas pequeñas, puesto que en las distribuciones con varianzas más pequeñas los valores de $f(y|\sigma^2)$ son más elevados (de hecho esto es exactamente lo que ocurre: el estimador máximo verosímil de la varianza el la varianza muestral, s^2, que es sesgado y tiende a infraestimarla).

[17] De hecho Von Mises (1957, pp. 157-158) acusa a Fisher de exponer con gran cuidado las diferencias entre probabilidad y verosimilitud, para acto seguido usar la verosimilitud como se usa la palabra probabilidad en el sentido cotidiano. No le falta razón.

[18] Un ejemplo conocido es la paradoja de James-Stein: supongamos que disponemos de K poblaciones $N(m_i, \sigma^2)$ independientes y se intenta estimar las m_i para que el riesgo $E[\Sigma(m_i - m_i)^2]$ sea mínimo. El estimador máximo verosímil está compuesto por las medias muestrales de cada población: $\mathbf{m}=[x_1, \ldots, x_K]$. James y Stein (1961) encontraron un estimador de menor riesgo que el máximo verosímil para $K>2$.

estima; por ejemplo, la estima máximo verosímil de la varianza es el cuadrado de la estima máximo verosímil de la desviación típica, y un estimador máximo verosímil función de otros que también lo sean, es a su vez máximo verosímil.

Desde un punto de vista práctico el método de máxima verosimilitud es una herramienta importante para el investigador aplicado. La escuela frecuentista ha determinado una serie de propiedades interesantes que deben reunir los estimadores, pero no indica cómo hacerse con estimadores que las posean. El método de máxima verosimilitud da al investigador y al técnico una forma de obtener estimadores razonables, si bien de forma asintótica.

Se puede obtener, además, una medida de la precisión de la estimación derivada de la propia función de verosimilitud. Una medida de la precisión se obtiene examinando la forma de la curva de verosimilitud; si esta es apuntada, hay razones para considerar a su máximo como un valor fiable; si la función es más bien aplanada, su valor máximo es tan *verosímil* como otros valores que difieren notablemente de él. Para medir la precisión Fisher introdujo el concepto de *cantidad de información* ([19]),

$$E[\delta \ell(u|y)/\delta u]^2$$

La escuela frecuentista ha reducido el problema a encontrar el estimador máximo verosímil debido a que éste tiene buenas propiedades desde el punto de vista frecuentista, la principal de las cuales es que converge a una distribución Normal cuya media es el valor verdadero del parámetro que se quiere estimar, y su varianza la inversa de la cantidad de información

---

[19] "Cantidad de información" es un término con un contenido emocional que intentaré expresar de forma intuitiva. Llamando l al logaritmo de la verosimilitud, dl(u|y)/du es la pendiente de la curva del logaritmo de la verosimilitud. Este valor, positivo o negativo, será elevado si la función de verosimilitud es apuntada, y será pequeño si es plana. Elevando al cuadrado para evitar problemas de signo y hallando la media de todos estos valores, esta media será un valor elevado si la función es apuntada y bajo en caso de que sea plana. En el entorno del máximo, si la función es apuntada ello quiere decir que el grado de credibilidad de los puntos alejados del máximo es pequeño; si por el contrario la función es muy plana, puntos distantes tendrán un grado de credibilidad similar. Esta cantidad fue llamada por Fisher cantidad de información intrínseca a los datos, probablemente porque cuanto mayor sea la cantidad de información, más apuntada es la curva de verosimilitud y menos verosímiles son las estimas alejadas del máximo. En muestras pequeñas, además, la función de verosimilitud puede presentar máximos locales o ser asimétrica en torno al máximo, con lo que el "grado de creencia racional" que la verosimilitud proporciona no es el mismo a un lado del máximo que al otro lado. El motivo de usar logaritmos es facilitar el que la cantidad de información sea aditiva. Si los n individuos de una muestra son independientes, la verosimilitud

$L(q|x_1,...x_n) = L(q|x_1) \cdot L(q|x_2) \cdot ... \cdot L(q|x_n)$,

por tanto

$\log L(q|x_1,...x_n) = \log L(q|x_1) + \log L(q|x_2) + ... + \log L(q|x_n)$

aplicada en el punto máximo (ver, p. ej., Stuart y Ord, 1991, pp. 659 y 660) ([20]). El principal problema de esta medida de la precisión reside en que depende del tamaño de la muestra, puesto que se basa en una aplicación del teorema central del límite, por lo que no está claro qué sucede en muestras pequeñas, ni tampoco cuál es el tamaño muestral a partir del cual estos problemas desaparecen.

Si la verosimilitud indica un grado de creencia racional en una estimación, lo lógico sería utilizar el cociente entre dos verosimilitudes para decidir sobre la importancia de un efecto en un modelo. Lamentablemente ya hemos visto que las verosimilitudes no son cantidades que se puedan comparar como se comparan las probabilidades; una verosimilitud cuatro veces mayor que otra no proporciona un "grado de creencia racional" cuatro veces mayor. Sin embargo el cociente de verosimilitudes tiene propiedades frecuentistas interesantes y es habitual el uso del test de razones de verosimilitud. Si en un modelo se ignora un factor aleatorio, los cocientes de verosimilitudes se distribuirían, si se repitiera infinitas veces la experiencia, como una χ^2_1 , lo que permite utilizarlo de forma frecuentista examinando las regiones de aceptación y rechazo para su uso ([21]).

2.5. El REML como estimador frecuentista

El método de máxima verosimilitud ha encontrado su aplicación más frecuente en mejora genética animal en la estimación de componentes de varianza, singularmente debido a una variante del procedimiento, la máxima verosimilitud residual o restringida (REML).

Hartley y Rao (1967) propusieron obtener estimas de componentes de varianza utilizando el método de máxima verosimilitud (ML). El procedimiento da lugar a ecuaciones bastante complicadas que deben resolverse de forma iterativa, por lo que dificultades de cómputo impidieron su utilización rutinaria en programas de mejora. Surgió, además, alguna dificultad conceptual: al resolver modelos mixtos, ML estima componentes de varianza como si la estima de los efectos fijos se hubiera realizado sin error; esto es, sin tener en cuenta los grados de libertad perdidos al estimar los efectos fijos, lo que podría ser preocupante cuando se estiman muchos efectos fijos, como es el caso del vacuno de leche. Por estas razón se propuso realizar un cambio de sistema de coordenadas y proyectar los datos en un subespacio en el que no hubieran efectos fijos, maximizando la verosimilitud en este subespacio. A este procedimiento se le llamó máxima verosimilitud restringida (REML) y fue generalizado por Patterson y Thompson en 1971. El REML presenta, además, la ventaja -más bien de tipo psicológico- de que sus estimas coinciden con las del ANOVA cuando los datos estaban equilibrados. Las estimas ML en diseños equilibrados dan resultados diferentes a las del ANOVA -que es óptimo para esos diseños-, lo que resulta un tanto inquietante.

---

[20] Recuerdo que esta cantidad es equivalente matemáticamente a $- E[dL^2(u|y)/du^2]$, que es como suele encontrarse en la literatura (ver Stuart y Ord, 1991, pp. 659).

[21] Edwards (1992) recomienda, de todas formas, la comparación de verosimilitudes, pero basado en argumentos ajenos a la escuela frecuentista.

Para examinar mejor las diferencias entre máxima verosimilitud y REML, estimaremos la varianza por ambos métodos utilizando un modelo sencillo.

$$\mathbf{y} = \mathbf{X}\mu + \mathbf{e} = \mathbf{1}\mu + \mathbf{e}$$

donde $\mathbf{X} = \mathbf{1}$ es un vector de unos. La matriz de varianzas-covarianzas de los errores es $\mathbf{V} = \mathbf{I}\sigma^2$, y su determinante $|\mathbf{V}| = (\sigma^2)^n$. La función de verosimilitud, en el caso de que el carácter se distribuya de forma normal $N(\mu,\sigma^2)$ y la muestra tenga n datos, es

$$L(\sigma^2|\mathbf{y}) = cte \cdot \sigma^{2(-n/2)} \exp[-(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) / 2\sigma^2]$$

El valor que hace máximo $L(\sigma^2|\mathbf{y})$ es

$$\hat{\sigma}^2 = (1/n)(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = (1/n)\Sigma(y_i - \mu)^2$$

Obsérvese que debemos conocer μ para obtener la estima de la varianza. Como no conocemos μ lo sustituimos por la estima máximo verosímil de μ,

$$\hat{\mu} = (1/n)\Sigma y_i$$

con lo que la estima ML de la varianza es

$$\hat{\sigma}^2_{ML} = (1/n)(\mathbf{y}-\mathbf{1}\hat{\mu})'(\mathbf{y}-\mathbf{1}\hat{\mu})$$

Esta estima, a pesar de ser función de otra, sigue teniendo las mismas buenas propiedades asintóticas que todas las estimas de máxima verosimilitud, y no hay razón formal para rechazarla. Para calcular las estimas REML se proyectan los datos en un subespacio sin efectos fijos. Si la matriz de proyección es **K**, el método consiste en hacer

$$\mathbf{K}'\mathbf{y} = \mathbf{K}'\mathbf{1}\mu + \mathbf{K}'\mathbf{e} = \mathbf{K}'\mathbf{e}$$

de forma que $\mathbf{K}'\mathbf{1} = \mathbf{0}$. Por ejemplo, la matriz

$$\mathbf{K}' = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & \dots & 0 \\ 1 & 0 & -1 & 0 & \dots & \dots & 0 \\ 1 & 0 & 0 & -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & \dots & \dots & -1 \end{bmatrix}$$

cumple la condición.

La varianza de $\mathbf{K'y}$ es $\mathbf{K'VK} = \mathbf{K'K}\sigma^2$. La verosimilitud es, ahora

$$L(\sigma^2|\mathbf{K'y}) = cte \cdot |\mathbf{K'K}\sigma^2|^{1/2} + \exp\left[(\mathbf{K'y})' (\mathbf{K'K})^{-1} (\mathbf{K'y}) / 2\sigma^2\right]$$

Si se hace la multiplicación $\mathbf{K'K}$ se deduce inmediatamente que

$$|\mathbf{K'K}\sigma^2| = n (\sigma^2)^{n-1}$$

con lo que se puede obtener el valor que maximiza la verosimilitud, y que ahora es

$$\hat{\sigma}^2 = [1/(n-1)]\, \mathbf{y'K(K'K)^{-1}K'y}$$

y aunque se puede demostrar formalmente para el caso general ([22]), es sencillo ver en nuestro ejemplo que

$$\mathbf{y'K(K'K)^{-1}K'y} = [\mathbf{y} - \mathbf{1}(1/n)\Sigma y_i\,]\,'\,[\mathbf{y} - \mathbf{1}(1/n)\Sigma y_i\,] = (\mathbf{y} - \mathbf{1}\hat{\mu})'(\mathbf{y} - \mathbf{1}\hat{\mu})$$

y la estima REML de la varianza es

$$\hat{\sigma}^2_{REML} = \left[1 / (n-1)\right](\mathbf{y} - \mathbf{1}\hat{\mu})'(\mathbf{y} - \mathbf{1}\hat{\mu})$$

que es idéntica a la estima ML, pero dividiendo por n-1 en vez de por n. A pesar de la similitud de las fórmulas hay que hacer notar que, al contrario que en la estima ML, no se sustituye el valor verdadero de μ por el estimado, sino que al deducir la fórmula del estimador de la varianza aparece una expresión, $(1/n)\Sigma y_i$, que coincide con la estima máximo verosímil de μ. El hecho de que se ha tenido en cuenta el grado de libertad perdido al estimar μ, se refleja en que se divide por (n-1) en lugar de por n. Cuando hay muchos efectos fijos esta distinción es notable, puesto que la estima máximo verosimil es

$$\hat{\sigma}^2_{ML} = (1 / n)(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

mientras que la estima REML es

$$\hat{\sigma}^2_{REML} = \left[1 / (n - rg\mathbf{X})\right](\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

---

[22] Se puede demostrar que, en general,

$$\mathbf{y'K(K'VK)^{-1}K'y} = [\mathbf{y} - \mathbf{X}\,(\mathbf{X'V^{-1}X})^{-}\,\mathbf{X'V^{-1}y}\,]'\,\mathbf{V^{-1}}\,[\mathbf{y} - \mathbf{X}\,(\mathbf{X'V^{-1}X})^{-}\,\mathbf{X'V^{-1}y}\,]$$

En nuestro caso, $\mathbf{X} = \mathbf{1}$, $\mathbf{V} = \mathbf{I}\sigma^2$, con lo que

$$\mathbf{X}(\mathbf{X'V^{-1}X})^{-1}\,\mathbf{X'V^{-1}y} = \mathbf{1}\,[(1/\sigma^2)\mathbf{1'1}]^{-1}\,(1/\sigma^2)\mathbf{1'y} = \mathbf{1}[(1/\sigma^2)\,n\,]^{-1}\,(1/\sigma^2)\Sigma y_i = \mathbf{1}\,(1/n)\Sigma y_i$$

donde rg **X** es el rango de **X** y $\hat{\mathbf{b}}$ la estima máximo verosímil de **b**.

Con la matriz **K** propuesta, el análisis de la varianza se hace sobre un vector

$$(\mathbf{K'y})' = [\, y_1 - y_2 ,\ y_1 - y_3 , \ldots , y_1 - y_n \,]$$

la idea es que una forma de hacer desaparecer la media es hacer el análisis sobre diferencias de datos en lugar de sobre los propios datos, puesto que

$$y_i - y_j = (\mu + e_i) - (\mu + e_j) = e_i - e_j$$

donde μ ha desaparecido. Hay varias matrices que cumplen la condición $\mathbf{K'1} = \mathbf{0}$. El análisis se podría hacer también sobre

$$(\mathbf{K'y})' = [\, y_1 - y_2 ,\ y_2 - y_3 , \ldots , y_{n-1} - y_n \,]$$

con el mismo resultado. Esto no quiere decir que cualquier **K** valga. Una matriz con la mitad de sus filas compuestas por ceros también cumple $\mathbf{K'1} = 0$, y también lo cumple la matriz $\mathbf{K} = \mathbf{0}$. Se trata de encontrar matrices que no hagan perder información relativa a la dispersión, lo que se consigue introduciendo en **K** el máximo número de contrastes lineales independientes. De hecho no importa la **K** concreta, puesto que en la fórmula final no aparece. Obsérvese que **K** es de dimensiones (n-1) x n puesto que sólo hay n-1 parejas de diferencias: perdemos un grado de libertad, o lo que es lo mismo, utilizamos la información algo resumida: el nuevo vector de datos **K'y** es un vector de n-1 elementos. Si tuviéramos que hacer una representación geométrica de nuestra muestra, necesitaríamos un espacio de n dimensiones, sin embargo, al utilizar REML deberíamos representar nuestros datos en un espacio de n-1 dimensiones, nos movemos en un espacio más restringido, hemos perdido un *grado de libertad*. Cuando se estiman muchos efectos fijos, esta pérdida es más notable.

No hay un argumento claro para preferir REML a ML. En el ejemplo anterior la estima REML es insesgada pero tiene un riesgo mayor que la ML para todos los valores posibles de la varianza, pero esto puede no ocurrir en casos más complejos. En otras situaciones el riesgo de la estimación REML será mayor o menor que la de ML dependiendo de los valores verdaderos de las componentes de varianza y de la estructura de los datos. La elección que en mejora animal se ha hecho hacia el método REML está relacionada más con argumentos indirectos como los expuestos antes que con una razón basada en el riesgo del estimador o sus propiedades. El argumento de que las estimas REML coinciden con las de ANOVA en diseños equilibrados es escasamente convincente en mejora animal, en donde los diseños están siempre sometidos a fuertes desequilibrios. Sin embargo el REML se ha impuesto como el "método a escoger", tal vez, parafraseando a Dawid en la cita que da comienzo a este artículo, porque "todo el mundo sabe que es algo que está muy bien".

3. LA ESCUELA BAYESIANA

La escuela bayesiana fue fundada por Laplace por medio de varios trabajos publicados de 1774 a 1812, y durante el siglo XIX ocupó un papel preponderante en la inferencia científica (Stigler, 1986). Antes que Laplace, y sin que al parecer éste tuviera conocimiento, se había presentado en la Royal Society de Londres un trabajo póstumo atribuido a un oscuro clérigo, el reverendo Thomas Bayes, formalizando el mismo principio de inferencia ([23]). Los trabajos sobre verosimilitud de Fisher en los años 20 y los de la escuela frecuentista en los 30 y 40 hicieron casi desaparecer a la escuela bayesiana, hasta que comenzó un "revival" a mediados de los 50 que dura *in crescendo* hasta nuestros días. En mejora genética animal el bayesianismo fue introducido por Daniel Gianola, primero en trabajos sobre caracteres umbral en colaboración con J.L. Foulley, y posteriormente en artículos en los que se desarrollan aplicaciones a prácticamente todos los campos de la mejora genética animal (ver Gianola et al., 1996 para una revisión sobre los métodos bayesianos en mejora genética animal).

3.1. Inferencia y precisión

La forma esencial de trabajar de la escuela bayesiana consiste en, dados los datos observados en el experimento, describir toda la incertidumbre que puede existir en torno a un parámetro, usando como medida natural de la incertidumbre la probabilidad de que el parámetro tome determinados valores. Por ejemplo, en el caso de la heredabilidad se obtendría la función de densidad de probabilidad $f(h^2|\mathbf{y})$ siendo $\mathbf{y}$ el vector de valores observados (figura 2). Una vez obtenida esa distribución se pueden hacer inferencias de múltiples maneras: por ejemplo, se puede desear averiguar entre qué valores se encuentra h^2 con una probabilidad del 95%, o qué probabilidad tiene el que h^2 esté entre tal y tal valor. En la figura 2 el área entre los valores 0.1 y 0.3 es la probabilidad de que h^2 se encuentre entre esos dos valores. En los casos en los que es necesaria una estimación puntual de h^2, por ejemplo para un índice de selección, hay varios parámetros de la función de densidad $f(h^2|\mathbf{y})$ que pueden ser usados como estimación puntual, y cuyo uso depende de la preferencia del investigador. Por ejemplo,

- *La moda*: es el valor de h^2 que maximiza $f(h^2|\mathbf{y})$; esto es, el valor más probable de h^2 dada la muestra $\mathbf{y}$. Puede demostrarse que es el valor que maximiza el riesgo de estimación cuando la función de pérdidas es del tipo todo-nada; esto es, la función de pérdidas toma valor 0 cuando $h^2 = \hat{h}^2$ y valor 1 en caso contrario, lo que equivale en realidad a no asignar una función de pérdidas definida.

---

[23] Bayes no llegó a publicar ningún trabajo matemático en vida y era miembro de la Royal Society por unos estudios sobre metafísica. El manuscrito original no se conserva y debemos confiar en la adscripción a Bayes que su amigo R. Price hizo al presentar la publicación en la Royal Society. Aparentemente el principio había sido formulado con anterioridad, y Stigler (1983) se lo atribuye a Saunderson, un profesor de óptica ciego autor de un gran número de publicaciones sobre varias ramas de las matemáticas.

- *La mediana*: Es tan probable que el valor verdadero sea superior como inferior a esta estima. Minimiza el riesgo de estimación cuando la función de pérdidas es $|h^2 - \hat{h}^2|$.

- *La media*: Es el estimador que minimiza el riesgo mínimo cuadrático, que es por otra parte el de uso más corriente en mejora animal: $E(h^2 - \hat{h}^2)^2$.

Para poder hacer todas estas inferencias es menester disponer de la función de densidad de probabilidad $f(h^2|\mathbf{y})$. De acuerdo con las leyes de la probabilidad, la probabilidad P(A,B)de que se presenten dos sucesos simultáneamente es

$$P(A,B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

con lo que

$$P(A|B) = P(B|A) \cdot P(A) / P(B)$$

En nuestro caso,

$$f(h^2|\mathbf{y}) = f(\mathbf{y}|h^2)\, f(h^2) \;/\; f(\mathbf{y}) = \text{cte} \cdot f(\mathbf{y}|h^2)\, f(h^2)$$

Obsérvese que $f(h^2|\mathbf{y})$ es una función de h^2, pero no de **y**, que está fijada; por tanto $f(\mathbf{y}|h^2)$ es aquí función de h^2, pero no de **y**, que es exactamente la definición que antes vimos de verosimilitud. Por la misma razón $f(\mathbf{y})$ es una constante, ya que no depende de h^2 e **y** está fijado. Finalmente, $f(h^2)$ es la densidad de probabilidad de h^2 al margen de nuestro experimento.

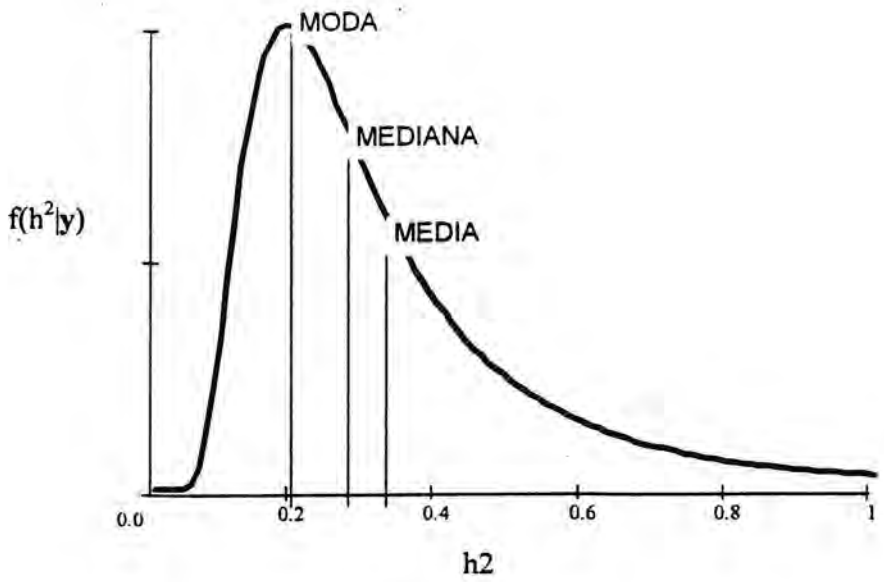
Las críticas al bayesianismo tienen que ver con esta última probabilidad ([24]), llamada *a priori* ([25]) porque no depende de los datos, es previa al experimento. En ocasiones esta información está claramente determinada; por ejemplo, la probabilidad *a priori* de obtener un individuo recesivo en el cruce de dos heterocigotos es 1/4, al margen del experimento, pero en el caso de la heredabilidad no está claro qué se quiere decir con esta probabilidad previa.

---

[24] Obviamente no hay probabilidades "absolutas" al margen de cualquier experimento. Los valores de la heredabilidad serán más o menos probable dadas unas circunstancias determinadas. Está mal escrito, por tanto, $f(h^2)$, y debiera escribirse siempre $f(h^2|H)$, donde H es el conjunto de hipótesis bajo las cuales los valores de h^2 son más o menos probables. El teorema de Bayes se reescribiría, pues, conteniendo en cada probabilidad un "dado H", pero no se hace por simplificar la escritura.

[25] Jeffreys (1931) distingue entre información previa ("prior") e información *a priori*. Para esta exposición la distinción no me parece relevante.

Figura 2. Densidad posterior de la heredabilidad



Obsérvese que:

1) Es discutible qué valor de estimación puntual de h^2 se debe ofrecer. En casos extremos, en los que la función de densidad posterior sea bimodal, podría ocurrir que el valor medio, que es el que minimiza el riesgo cuadrático, fuera mucho menos probable que los valores de las modas (figura 3). En este caso una estimación puntual no describe bien la incertidumbre en torno al verdadero valor de h^2.

2) Aunque la función de densidad posterior sea asimétrica, se puede calcular siempre un intervalo de confianza ([26]). Si se desea que el valor más probable esté en el centro del intervalo, en ese caso el intervalo es asimétrico ([27]).

---

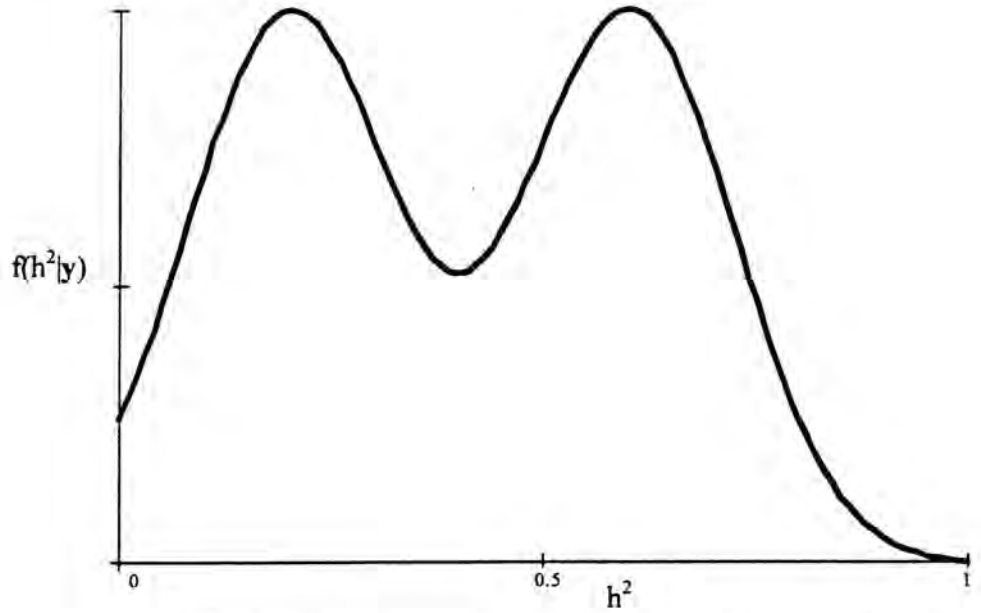
[26] Los bayesianos suelen hablar de intervalos o regiones 'de credibilidad', más que 'de confianza', puesto que son regiones en las que el verdadero valor del parámetro se encuentra con una probabilidad determinada.

[27] Una dificultad adicional, si se quiere usar el sistema de intervalos de confianza para la inferencia científica, es que hay que conocer exactamente la densidad posterior, por lo que hay que conocer el valor de f(**y**). Este valor es difícil de computar porque

$$f(\mathbf{y}) = \int f(\mathbf{y}, h^2) f(h^2) d(h^2)$$

como **y** es un vector, esta integral es multidimensional. Aunque los elementos de **y** fueran independientes, esta integral pasaría a ser una integral múltiple de tantas dimensiones como datos, y en ambos casos el problema es irresoluble incluso por métodos numéricos. Hoy en día las técnicas de muestreo de Gibbs que se mencionan en el apéndice I han resuelto este problema.

Figura 3. Densidad posterior bimodal de una heredabilidad.



3) Al no haber repeticiones hipotéticas del experimento, tampoco hay nada que pueda ser considerado como "sesgo". Tampoco hay factores fijos, dado que a todos los factores desconocidos se les asigna una distribución aleatoria *a priori* basada en las opiniones del investigador.

Los tests de hipótesis tienen una interpretación más clara en la escuela bayesiana. Por ejemplo, se desea probar si las medias de dos razas son iguales. Para ello se calcula el cociente entre la probabilidad de que sean iguales y la probabilidad de que difieran:

$$P(m_1 - m_2 = 0 \mid \mathbf{y}) \;/\; P(m_1 - m_2 \neq 0 \mid \mathbf{y})$$

si este cociente es igual a 7, por ejemplo, esto quiere decir que la probabilidad de que ambas medias sean iguales es 7 veces mayor que la probabilidad de que sean diferentes.

3.2. El uso de la información previa

Cuando claramente hay información a priori

En ocasiones existe información a priori sobre los parámetros a estimar. En esos casos el uso de esta información previa no representa ningún problema, como veremos a continuación con un ejemplo de Fisher (para quien, dicho sea de paso, el bayesianismo era una auténtica "bestia

negra") ([28]). Supongamos un carácter controlado por un gen con dos alelos y con dominancia, como el caso del color negro que domina sobre el marrón de la piel de ciertos ratones. Se desea averiguar si un ratón negro proveniente de un apareamiento entre heterocigotos es homocigoto o heterocigoto. Para ello se cruza ese ratón negro con uno marrón, y se obtienen siete descendientes, todos ellos negros.

La probabilidad a priori de que un ratón negro proveniente de un apareamiento entre heterocigotos sea homocigoto es 1/3, puesto que un apareamiento Aa x Aa da tres tipos de animales negros: AA, Aa, aA.

$$f(AA) = 1/3$$

A partir de ahora representaremos por Aa a los animales heterocigotos, sean estos Aa ó aA. La probabilidad a priori de que un ratón negro proveniente de un apareamiento entre heterocigotos sea heterocigoto es, por la misma razón, 2/3.

$$f(Aa) = 2/3$$

Una vez realizado el experimento, la verosimilitud de que el ratón negro sea homocigoto es 1; es decir, dado que se observan siete descendientes negros, en el caso de que el padre negro fuera homocigoto (AA), la probabilidad de ese suceso sería 1, puesto que los descendientes del cruce con el ratón marrón (aa) cabe esperar que sean negros (Aa).

$$L(AA|\ y=7 \text{ descendientes, todos negros}) \equiv \text{Prob}(y=7 \text{ descendientes, todos negros} \mid AA) = 1$$

Una vez realizado el experimento, la verosimilitud de que el ratón negro sea heterocigoto es $(1/2)^7 = 1/128$; es decir, dado que se observan siete descendientes negros, en el caso de que el padre negro fuera heterocigoto (Aa), la probabilidad de ese suceso sería 1/2, puesto que los descendientes del cruce con el ratón marrón (aa) cabe esperar que sean negros (Aa) y marrones (aa) en la misma proporción.

$$L(Aa|\ y=7 \text{ descendientes negros}) = \text{Prob}\ (y=7 \text{ descendientes todos negros} \mid Aa) = \frac{1}{128}$$

Por tanto, la probabilidad de que el padre sea AA debiera ser

$$L(AA \mid y) \cdot f(AA) = 1 \cdot 1/3 = 1/3$$

y de que sea heterocigoto

$$L(Aa \mid y) \cdot f(Aa) = (1/128)\ (2/3) = 1/192$$

---

[28] El lector puede creer que exagero. Lane (1980) usa exactamente esta expresión al revisar las opiniones de Fisher sobre el bayesianismo.

pero como estas cantidades no suman 1 y ambos son sucesos excluyentes que no admiten otra posibilidad, hay que dividir por la suma de ambas:

f(y) = Prob (7 descendientes todos negros) =
= Prob (padre AA) + Prob (padre Aa) = (1/3) + (1/192)

Probabilidad de padre homocigoto, dado que se obtuvieron siete descendientes negros al cruzar con un ratón marrón

$$f(AA \mid y) = L(AA \mid y) \cdot f(AA) / f(y) = (1/3) / (1/3 + 1/192)$$

Probabilidad de padre heterocigoto, dado que se obtuvieron siete descendientes negros al cruzar con un ratón marrón

$$f(Aa \mid y) = L(Aa \mid y) \cdot f(Aa) = (1/192) / (1/3 + 1/192)$$

que es la respuesta buscada.

Cuando la información a priori se admite como conocida

Es frecuente en mejora animal admitir que se conocen los parámetros genéticos, por ejemplo para usarlos en estimaciones BLUP o en índices de selección, aunque sólo se tengan estimas de los mismos. En el caso de los índices se requiere además el conocimiento de las medias (es decir, de los efectos fijos). Este uso de la información previa está admitido en mejora genética animal, y no genera ninguna polémica. Consideremos el modelo

$$\mathbf{y} = \mathbf{Z}\,\mathbf{u} + \mathbf{e}$$

donde **y** es el vector de datos, **u** el valor aditivo de un individuo, **e** la componente ambiental y **Z** la matriz de incidencia. Hallaremos la función de densidad posterior f(**u**|**y**).

Se considera que **u** se distribuye de forma Normal con media cero y matriz de varianzas-covarianzas genéticas conocida **A**

$$f(\mathbf{u}) = cte \cdot \exp\{(-1/2\sigma_u^2)\,\mathbf{u}'\,\mathbf{A}^{-1}\,\mathbf{u}\}$$

$$f(\mathbf{y}|\mathbf{u})_{\text{como función de } \mathbf{u}} = cte \cdot \exp\{(-1/2\sigma_e^2)\,(\mathbf{y} - \mathbf{Zu})'\,(\mathbf{y} - \mathbf{Zu})\}$$

$$f(\mathbf{u}|\mathbf{y}) = f(\mathbf{y}|\mathbf{u})_{\text{como función de } \mathbf{u}} \cdot f(\mathbf{u}) = cte \cdot \exp\{(-1/2\sigma_e^2)\,[(\mathbf{y} - \mathbf{Zu})'\,(\mathbf{y} - \mathbf{Zu}) + \mathbf{u}'\,\mathbf{A}^{-1}\,\mathbf{u}\,\alpha]\}$$

donde $\alpha = \sigma_e^2 / \sigma_u^2$

Esta función de densidad posterior f(**u**|**y**) tiene el máximo en

$$\hat{\mathbf{u}} = (\mathbf{Z'Z} + \mathbf{A}^{-1}\,\alpha)^{-1}\,\mathbf{Z'}\,\mathbf{y}$$

lo que corresponde a la mejor predicción lineal de **u**, y es un resultado conocido de los índices de selección. Como la distribución es Normal, el máximo corresponde a la moda, la media y la mediana simultáneamente. A partir de f(**u**|**y**) se pueden hacer predicciones como hemos visto anteriormente.

Cuando la información a priori es vaga

En muchas ocasiones es difícil cuantificar la información a priori de una forma tan objetiva como la de los ejemplos que acabo de citar. En esos casos los estadísticos no bayesianos consideran que no es posible aplicar el teorema de Bayes y el problema no tiene solución por la vía de las probabilidades. Dentro del campo bayesiano se ha intentado dar respuesta a esta dificultad de varias formas:

1) *Definiendo la probabilidad como un estado de creencias del investigador*

La probabilidad definida como la frecuencia con la que se puede presentar un suceso es un concepto demasiado restrictivo como para dar solución a nuestro problema, por lo que se considera que la probabilidad es un valor subjetivo que expresa la opinión del investigador acerca del grado de conocimiento existente sobre el parámetro que se desea estimar. Probabilidad subjetiva no quiere decir probabilidad arbitraria. Como hace notar Keynes (1921), esta opinión subjetiva debe estar fundada de forma tal que el investigador pueda persuadir a los científicos de su entorno de que esa opinión es razonable.

Una forma de imaginar cómo puede establecerse esta probabilidad es pensar, de momento, en un problema distinto. Supongamos que intentamos evaluar el precio de un producto -p.ej. el kg de conejo-. Al margen de cualquier muestra que se tome, podemos afirmar que será menos de 10.000 pts el kg y más de 0 pts. Podemos afinar más diciendo que es poco probable que sea de 100 pts/kg y también que es poco probable que esté a 1000 pts./kg. Continuando, podemos poco a poco establecer una serie de sucesos que nos parecen igualmente probables, o unos más que otros, y llegar a producir una distribución más o menos escalonada que será considerada como aceptable por muchos interlocutores. Esto, que es sencillo para problemas univariantes, se vuelve prácticamente imposible para problemas multivariantes. Por ejemplo, si hablamos de tres caracteres con varios efectos ambientales (lo que se llaman efectos fijos en el contexto frecuentista), hace falta determinar la probabilidad a priori que opinamos sobre el caso "heredabilidad del primer carácter 0.1, del segundo 0.2, del tercero 0.2, correlación genética del primero con el segundo -0.7, del segundo con el tercero 0.1, del primero con el tercero de 0.1, para un efecto de estación de -0.5 para el primer carácter, de 7.2 para el segundo, etc. ¡Y esto debe hacerse para cada caso posible! Determinar todas estas opiniones puede conducir al investigador a una crisis psicológica, por lo que en la mayor parte de problemas de mejora genética animal esta determinación de la opinión a priori debe ser abandonada.

2) *Eliminando en la práctica la influencia de la probabilidad a priori a base de aumentar el tamaño muestral*

Si se dispusiera de suficientes datos, la probabilidad a priori no influiría en la distribución de la densidad posterior de probabilidades, por tanto se deben hacer experimentos con un número de datos suficiente como para que la función a priori carezca de relevancia. En ese caso la función de densidad de probabilidades a priori se busca de forma arbitraria, y habitualmente se procura que facilite los cálculos de la función posterior y que no conduzca a paradojas o a resultados inadmisibles. Este argumento fue introducido en la literatura bayesiana por Jeffreys (1931) y ha tomado popularidad en los últimos años ante las dificultades que presenta el uso de la información previa cuando ésta es vaga. Su dificultad principal reside en la determinación del tamaño de la muestra en experimentos con pocos datos, en los que esa densidad previa puede influir el resultado final de forma relevante.

Cuando no hay información a priori

Hasta aquí, si bien discutible el procedimiento sobre cómo representar las creencias, no se presenta ninguna dificultad de tipo teórico. Habrá investigadores que prefieran no considerar a la probabilidad como un fenómeno subjetivo, pero si se da este paso adelante, la inferencia que se produce no da lugar a paradojas notables ni a contradicciones. El problema aparece al intentar representar la ignorancia *a priori*. Los primeros bayesianos (Bayes incluido) usaron una técnica a la que Keynes denominó *Principio de Indiferencia*([29]), y que consiste en admitir el siguiente postulado: *Cuando no hay información sobre los distintos sucesos posibles, todos tienen la misma probabilidad a priori*. Admitir este postulado equivale a decir que en el viejo saco que contiene bolas blancas y negras usado para enseñar probabilidades, si no dispongo de ninguna información sobre en que proporción están, mi creencia a priori es que cualquier combinación posible tiene la misma probabilidad. En el caso de variables continuas esto implica representarlas como una recta paralela al eje de las X en un intervalo concreto, por ejemplo al intervalo [0,1] para el caso de la heredabilidad, por lo que se les conoce también como *a prioris planos*. Este terrible postulado es origen no sólo de problemas ontológicos, sino matemáticos. Por ejemplo, si quiero representar ignorancia total sobre la heredabilidad, puedo decir que puede tomar cualquier valor en el intervalo [0,1] con la misma probabilidad. Entonces,

$$P(h^2 < 0.5) = 1/2$$

Pero el suceso $h^2 < 0.5$ es el mismo que $h^4 < 0.25$, por lo que su probabilidad es la misma:

$$P(h^4 < 0.25) = 1/2$$

---

[29] Keynes (1921) ilustra su capítulo sobre el principio de indiferencia con una cita interesante que transcribo:
ABSOLUTO: - Seguro, Señor, que no es muy razonable entregar mi amor a una dama de la que no conozco nada.
SIR ANTHONY: - Estoy seguro, Señor, que es menos razonable poner objeciones a una dama de la que usted nada conoce.

Esto indica que es más probable que h^4 se encuentre cerca de 0 que cerca de 1, luego si admitimos este resultado queremos decir ni más ni menos que sobre h^2 no hay información a priori, pero sobre h^4 sí, lo que es absurdo. No hay una respuesta sencilla a este problema. Las soluciones propuestas son:

1) *Admitir la validez de los aprioris planos*

Es razonable representar la ausencia de información de forma plana sobre h^2 y no sobre h^4. Si admitiéramos que h^{16} tiene una densidad uniforme entre 0 y 1 ello implicaría que $P(h^{16} > 0.5) = 1/2$, por lo que consecuentemente $P(h^2>0.5^{1/8}) = P(h^2>0.92) = 1/2$, algo que no estaríamos dispuestos a admitir. Esta respuesta, propuesta por Box y Tiao (1973), resulta poco convincente. Como Edwards (1992) señala, parece lógico admitir que la ausencia de información sobre un parámetro implica también la ausencia de información sobre cualquier transformación. Decir que ignoramos qué probabilidades tienen los valores de un parámetro no parece lo mismo que decir que todos ellos tienen la misma probabilidad.

2) *La ignorancia se representa respecto a la verosimilitud*

Hay simplemente un problema de escala. Cambiando la escala es posible encontrar funciones de densidad a priori inmunes a reparametrizaciones. Para ello en primer lugar hay que definir bien qué quiere decir ausencia de información. De la expresión del teorema de Bayes vemos que la información de los datos está contenida en la verosimilitud $f(y|h^2)$, mientras que la densidad a priori $f(h^2)$ es independiente de los datos. Una distribución de densidad de probabilidades que exprese ignorancia se debe expresar en una escala en la cual la verosimilitud no cambie. Esta solución fue propuesta inicialmente por Jeffreys (1961) y está desarrollada con detalle en Box y Tiao (1973, cap. 1). En principio no siempre es posible encontrar una escala así, por lo que se recurre a aproximaciones que son buenas cuando el tamaño de la muestra es grande (Box y Tiao dicen de tamaño moderado, sin duda para evitar las críticas consecuentes). En el caso de que se intente expresar ignorancia de dos parámetros simultáneamente, por ejemplo, la media y la varianza, estas funciones no existen o conducen a paradojas. Las funciones a priori de Jeffreys no han tenido éxito en Mejora animal, a pesar de ser ampliamente usadas en otros campos de la estadística, precisamente porque los problemas de mejora animal suelen ser multivariantes de una u otra forma (hay efectos ambientales además de los genéticos, en muchas ocasiones los problemas son multicarácter, etc.).

Sucede también que muchas de las funciones de densidad a priori propuestas no son verdaderas probabilidades, puesto que su integral no es 1; por ejemplo, la función plana que da el mismo valor de probabilidad a todos los puntos entre $-\infty$ y $+\infty$. Estas funciones pueden en ocasiones dar lugar a densidades posteriores que no integren a 1 y que no sean, por tanto, verdaderas probabilidades, con lo que la inferencia se vuelve imposible. En muchas ocasiones es difícil de detectar que la función posterior encontrada no es una auténctica función de densidad, con lo que se corre el riesgo de realizar afirmaciones probabilísticas falsas. En un artículo reciente, Hobert y Casella (1997) recomiendan utilizar densidades a priori propias para evitar estos problemas. En cualquier caso, si utilizamos densidades a priori impropias

persiste siempre el problema ontológico de considerar una incertidumbre a priori mediante una función que no es una verdadera probabilidad.

3) *Hallar una densidad posterior en la que la información a priori sea mínima*

Para casos en los que haya objetivamente una ignorancia a priori, Bernardo (1979) propuso el cálculo de funciones de densidad posterior de forma que las funciones a priori debían dar el máximo peso posible a los datos, lo que tiene una justificación intuitiva aparentemente razonable. La ventaja de este planteamiento es que se calcula la densidad posterior directamente, lo que asegura que se trata de una verdadera probabilidad. En cualquier caso, para quienes propugnan esta solución,

> "el problema de caracterizar la ignorancia a priori o la información vaga a priori es mucho más complejo de lo que intuitivamente parece... 'vago' es un concepto demasiado vago para que pueda ser útil. No hay un a priori objetivo que represente la ignorancia... el apriori de referencia componente del análisis es simplemente una herramienta matemática" (Bernardo y Smith, 1994, pp. 298).

El principal problema de esta solución es operativo: estos aprioris son difíciles de determinar, y hay que calcularlos para cada modelo ([30]).

3.3. El BLUP y los índices de selección como estimadores bayesianos

Ronningen (1971) primero y Dempfle (1977) después, observaron que el BLUP puede considerarse como un estimador bayesiano particular. Dado el modelo mixto (1), la forma bayesiana de plantear la estimación del valor de mejora en animales consiste en estimar la función de densidad $f(\mathbf{u}|\mathbf{y})$ para poder hacer inferencias. En un contexto bayesiano no hay "efectos fijos", todos los efectos son aleatorios, por lo que el caso de índices de selección vuelve a ser un caso particular en el que no existe el vector **b**.

Trataremos, pues de encontrar un estimador para el vector

$\mathbf{t}' = [\,\mathbf{b}'\ \ \mathbf{u}'\,]$

[30] En el caso multivariante debe distinguirse entre el parámetro de interés y los de ruido. Los aprioris de referencia de Bernardo han sido criticados porque cambian según se dispone el orden de parámetros en el que se realiza el análisis (ver, p. ej. Robert, 1992, pp. 109-111). Naturalmente, si se admite que son meras herramientas matemáticas sin valor ontológico, este es un problema menor. Una dificultad seria reside en que aparentemente el cálculo de estas densidades posteriores es extremadamente complejo, particularmente cuando los datos están seleccionados y, como es habitual, el proceso de selección no se puede describir de forma sencilla (nunca se aplica la misma barrera de selección, los individuos tienen información variable, etc.). Aparecen en el proceso integrales múltiples difíciles de resolver.

a partir de los datos **y**. Para ello determinaremos primero f(**t** | **y**). Sabemos que

$$f(\mathbf{t} \mid \mathbf{y}) = \text{cte. } f(\mathbf{y} \mid \mathbf{t})\, f(\mathbf{t})$$

Si llamamos $\mathbf{W} = [\mathbf{X}\ \mathbf{Z}]$ y suponemos que $f(\mathbf{y} \mid \mathbf{t}) \sim N([\mathbf{Xb}+\mathbf{Zu}], \mathbf{I}\sigma_e^2) = N(\mathbf{Wt}, \mathbf{I}\sigma_e^2)$ puesto que al estar **t** "dado" esto implica que es un valor fijado y constante,

$$f(\mathbf{y} \mid \mathbf{t}) = \text{cte. } \exp[(\mathbf{y} - \mathbf{Wt})'\ (\mathbf{y} - \mathbf{Wt})]$$

La densidad a priori de **t** podemos suponer que es también Normal. Esto es razonable para los valores aditivos, pero para los efectos ambientales discutible, y luego lo discutiremos. La media y la varianza *a priori* de **t** son

$$E(\mathbf{t}') = E[\mathbf{b}'\ \mathbf{u}'] = [\mathbf{m}_b'\ \mathbf{0}] = \mathbf{m}$$

$$\mathbf{V} = \text{Var}(\mathbf{t}') = \text{Var}[\mathbf{b}'\ \mathbf{u}'] = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix}$$

donde **m** y S son la media y la varianza *a priori* de los efectos ambientales. Suponemos que los efectos ambientales expresados en **b** y los aditivos de **u** están incorrelacionados ([31]).

$$f(\mathbf{t}) = \text{cte. } \exp[(\mathbf{t} - \mathbf{m})'\ \mathbf{V}^{-1}\ (\mathbf{t} - \mathbf{m})]$$

$$f(\mathbf{t} \mid \mathbf{y}) = \text{cte. } f(\mathbf{y} \mid \mathbf{t})\, f(\mathbf{t}) = \text{cte. } \exp[(\mathbf{y} - \mathbf{Wt})'\ (\mathbf{y} - \mathbf{Wt}) + (\mathbf{t} - \mathbf{m})'\ \mathbf{V}^{-1}\ (\mathbf{t} - \mathbf{m})]$$

Hemos visto tres estimadores bayesianos: media, mediana y moda de f(**t** | **y**). Como la distribución es Normal, coinciden. Hallaremos, por ejemplo, la moda, que es el valor que maximiza f(**t** | **y**). Derivando respecto a **t** e igualando a cero se obtiene

$$[\mathbf{W}'\mathbf{W} + \mathbf{V}^{-1}]\ \mathbf{t} = \mathbf{W}'\mathbf{y} + \mathbf{V}^{-1}\ \mathbf{m}$$

que, en forma expandida, es

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} + \mathbf{S}^{-1} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} + \mathbf{S}^{-1}\mathbf{m}_b \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

---

[31] Esto no es así en ciertas ocasiones. Se sabe, por ejemplo, que las mejores granjas de vacuno de leche gozan de un ambiente mejor y además usan semen de mejor calidad. Deshacerse de estas correlaciones fue una de las razones que impulsaron a Henderson (1973) a proponer el efecto granja como "fijo". Cuando el tamaño de las granjas es pequeño estos efectos están peor estimados que los aleatorios, lo que repercute en la estima de los efectos genéticos y genera un problema de difícil solución.

Estas ecuaciones son muy parecidas a las del modelo mixto, pero al producto matricial de los efectos antes considerados como "fijos", **X'X**, se le añade la inversa de la matriz de varianzas covarianzas de estos efectos, justamente como ocurre en los efectos aditivos con la matriz **A**. En el caso de una ignorancia total sobre los valores *a priori* de los efectos fijos, estos podrían variar en el intervalo $]-\infty\ ,\ +\infty[$, con lo que su varianza tendería a infinito y $\mathbf{S}^{-1}$ tendería a cero. Si $\mathbf{S}^{-1}$ se anula tenemos entonces las ecuaciones del modelo mixto.

El BLUP no es, pues, sino un estimador bayesiano que considera que la densidad *a priori* de los efectos aditivos es normal de media cero y varianza **A**, y la de los efectos ambientales es plana (todos los posibles valores tienen la misma probabilidad *a priori* ([32])) y varía a lo largo de toda la recta real (pueden tomar cualquier valor). Este último punto es discutible, puesto que cualquier efecto imaginable tiene en ganadería si no una distribución de densidad de probabilidades asociada, al menos unos límites que claramente no puede sobrepasar. Su uso queda justificado sólo como el de una densidad de referencia y por motivos operativos a la hora de calcular densidades posteriores.

3.4. Interpretación bayesiana de los métodos máximo verosímiles

Suponiendo que cualquier valor de h^2 es igualmente probable, $f(h^2)$=cte. (lo que sólo es en realidad admisible en el intervalo [0 1])

$$f(h^2|\mathbf{y}) = cte \cdot f(\mathbf{y}|h^2) \qquad (1)$$

con lo que se puede obtener $f(h^2|\mathbf{y})$ simplemente obteniendo $f(\mathbf{y}|h^2)$. En principio tampoco tenemos por qué saber cómo se distribuye $f(\mathbf{y}|h^2)$, pero esto es más fácil de resolver porque $f(\mathbf{y}|h^2)$ es una función de las muestras **y** que se toman, lo que facilita en muchas ocasiones el problema. Por ejemplo, en determinadas circunstancias, basándonos en experimentos pasados, podemos decir que **y** sigue una distribución Normal. Obsérvese, y es importante observarlo, que aunque la forma de la verosimilitud coincida con $f(h^2|\mathbf{y})$, la interpretación de la verosimilitud y de la densidad posterior es completamente diferente, puesto que $f(h^2|\mathbf{y})$ es una función de densidad de probabilidad y podemos hacer inferencias probabilísticas como hemos indicado antes.

En el caso particular de las estimas REML, fue Harville (1974) quien primero se percató de que admitían una interpretación bayesiana. La moda de la densidad posterior $f(\mathbf{b}, \sigma_u^2, \sigma_e^2)$ representa el máximo de esa distribución trivariante y tiene tres componentes, uno para **b** y otros dos para σ_u^2 y σ_e^2. La densidad posterior $f(\sigma_u^2) = \iint f(\mathbf{b}, \sigma_u^2, \sigma_e^2)\ d\mathbf{b}\ d\sigma_e^2$ tiene un máximo para un valor de σ_u^2 que no tiene por qué coincidir con el componente de la moda de

[32] Para ser estricto debiera decir que todos los intervalos de valores de la misma longitud tienen la misma probabilidad, puesto que la probabilidad de un punto concreto es cero ya que hay infinitos puntos en cualquier intervalo de la recta real.

$f(\mathbf{b}, \sigma_u^2, \sigma_e^2)$. La estima máximo verosímil (ML) de σ_u^2 coincide con la componente correspondiente de la moda de $f(\mathbf{b}, \sigma_u^2, \sigma_e^2)$, mientras que la estima REML coincide con la moda de $f(\sigma_u^2, \sigma_e^2)$. Cuando hay varios efectos aleatorios, como por ejemplo en el modelo animal con repetibilidad:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e}$$

donde **p** son los efectos aleatorios de ambiente común entre dos mediciones del carácter, la estima REML de σ_u^2 coincide con la componente correspondiente de la moda de la distribución $f(\sigma_u^2, \sigma_p^2, \sigma_e^2)$ mientras que la estima ML coincide con la componente correspondiente de la moda de la distribución $f(\mathbf{b}, \sigma_u^2, \sigma_p^2, \sigma_e^2)$.

Obsérvese que el REML, cuya justificación desde el punto de vista frecuentista es oscura, aparece aquí justificado de forma bayesiana (con las limitaciones de considerar la densidad previa de los efectos "fijos" como plana).

Obsérvese también que desde un punto de vista bayesiano puede irse más lejos y, en el caso del modelo con varios efectos aleatorios, proponer la moda de la densidad

$$f(\sigma_u^2) = \iint f(\sigma_u^2, \sigma_p^2, \sigma_e^2)\, d\sigma_p^2\, d\sigma_e^2$$

como estimador de la varianza σ_u^2. Esto es exactamente lo que proponen Gianola y Foulley (1990) con su estimador *VEIL*. La razón para proponer la moda, aparte de argumentos en torno a la función de pérdidas o a que sea el valor más probable, es de tipo práctico: se obtiene maximizando una función. Actualmente el desarrollo de técnicas con cadenas de Markov por simulación de Monte-Carlo (muestreo de Gibbs, etc.) hace innecesaria la obligatoriedad de recurrir a la moda, puesto que se estima la totalidad de la densidad posterior.

3.6. Los estimadores bayesianos del valor genético y los parámetros genéticos

Al margen de la interpretación bayesiana que tengan ciertos métodos de mejora, ¿cómo aborda un bayesiano la estima de los valores genéticos? En primer lugar, no existen "efectos fijos" en un contexto bayesiano, por lo que no sería de aplicación el modelo mixto expuesto en (1), sin embargo mantendremos la nomenclatura con el objeto de separar los efectos genéticos de los ambientales. La forma más amplia de abordar el problema consistiría en estimar la densidad posterior conjunta $f(\mathbf{b},\mathbf{u}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y})$, y sus correspondientes densidades marginales, aunque pronto aparecen las mismas dificultades que en el caso de los estimadores frecuentistas ([33]).

---

[33] En realidad la forma más amplia incluiría las distribuciones de posibles matrices **Z** y **X**. Como no se sabe bien cómo abordar este problema, en realidad lo que se estima es la

En primer lugar, se desconoce la densidad a priori conjunta de **b** y **u**, por lo que se admite que son independientes: f(**b**,**u**) = f(**b**)f(**u**). Como esto no es así en algunos casos importantes (ver nota 31), es frecuente tomar como valor a priori f(**b**)=constante, con lo que podemos obtener la solución BLUP con sus ventajas e inconvenientes ([34]). En el caso de **u** se admite que se distribuye de forma Normal con media cero y varianza σ_u^2 . Una forma alternativa de proceder sería dar información previa a f(**b**), con lo que la solución final puede verse afectada si hay pocos datos. Ya hemos discutido antes el uso de la información a priori, y habitualmente se espera que sea irrelevante debido a la información que proporcionen los datos, lo que es de esperar que ocurra frecuentemente, dada la gran cantidad de datos disponibles en cualquier programa de mejora.

En segundo lugar, hay que estimar las componentes de varianza a partir de los datos. Esto también afecta a la forma de la densidad a priori, puesto que también se suele suponer la independencia de las distribuciones y esto no es necesariamente cierto. Se suele suponer que la densidad a priori de las componentes de varianza es una χ^2 invertida para calcular más fácilmente la función de densidad posterior. Esto, que parece enormemente arbitrario, no lo es tanto, puesto que estas funciones χ^2 pueden tener formas muy diversas en función de dos parámetros ([35]), y puede expresarse un amplio margen de opiniones previas en función de sus valores. En cualquier caso, desde la aparición de las técnicas de muestreo de Gibbs y de muestreo en "caja negra", ya no es estrictamente necesario usar funciones que acomoden particularmente bien a los cálculos, aunque se sigue haciendo para facilitar el cómputo de las densidades posteriores.

Una vez determinadas las densidades previas y la verosimilitud, se calcula la densidad posterior multiplicando ambas densidades. Si la función de densidad resultante es conocida, por ejemplo una Normal, se pueden aprovechar sus propiedades para producir inferencias (por ejemplo, que el 95% de probabilidad cae entre 1.96 y -1.96 veces su desviación típica). Cuando no es así se puede intentar realizar una estimación puntual, por ejemplo hallando la moda de la función -con lo que basta hallar el máximo-, o bien se pueden extraer valores de esa densidad posterior y trabajar con esos valores para representar aproximadamente esta función de densidad. Las inferencias se hacen a partir de los puntos muestreados de la densidad posterior multivariante. Por ejemplo, supongamos que hemos muestreado 5.000 puntos de una densidad posterior conjunta $f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y})$. Cada uno de ellos tiene un valor para σ_u^2 y σ_e^2, por lo que realizando el cociente se obtienen 5.000 puntos muestreados

---

densidad posterior de $f(\mathbf{b},\mathbf{u},\sigma_u^2,\sigma_e^2 \mid \mathbf{y}, \mathbf{X}, \mathbf{Z})$, aunque la forma habitual de expresarla sea la que indicamos en el texto. Se pueden incluir más efectos y más jerarquías, pero para el ejemplo es más que suficiente el modelo propuesto.

[34] Nótese que aunque a priori sean independientes, en la densidad posterior típicamente no lo serán.

[35] Parámetros a los que se les llama "grado de credibilidad" ν y "parámetro de posición" S de forma un tanto equívoca, puesto que ambos parámetros afectan a la forma de la curva.

al azar de la densidad posterior de la heredabilidad. Hallando la media de valores de heredabilidad obtenidos, tengo una estima de la media de la densidad posterior; creando un histograma o dibujando la función de densidad a partir de esos 5.000 valores puedo obtener una estima de la moda, y ordenándolos puedo obtener la mediana con facilidad, así como cualquier intervalo de confianza.

El problema se centra ahora en cómo obtener esos valores tomados al azar de las funciones de densidad posterior. Hay un conjunto de funciones de los que se sabe cómo extraer valores al azar, y si la densidad posterior es una de ellas, no hay problemas para hacerlo. Cuando no es así, existen técnicas de muestreo "en caja negra" que permiten, con más o menos coste computacional, realizar ese muestreo ([36]). En el caso multivariante, o que involucra estima de muchos parámetros, como es el caso habitual en problemas de mejora genética (piénsese que además de las estimas de componentes de varianza hay que estimar los valores genéticos), no es posible computacionalmente -al menos de momento- muestrear directamente de la función de densidad posterior, y hay que transformar el problema en univariante, o en problema de menos dimensiones, mediante técnicas de muestreo de Gibbs que se explican en el apéndice I. Las técnicas de muestreo de Gibbs permiten, además muestrear de la función de densidad posterior y no de funciones proporcionales, eludiendo el problema expuesto en la nota 27.

4. DISCUSIÓN

Al comparar las ventajas e inconvenientes de las dos escuelas es menester distinguir los de tipo ontológico (relativas a la inferencia científica) de las de tipo práctico. Aunque las primeras han dado lugar a una copiosa y entretenida bibliografía, el mejorador genético animal estará a menudo más interesado en las segundas.

4.1. Ventajas e inconvenientes respecto a la inferencia científica

Hasta hace algunos años por lo menos, se ha venido considerando que la diferencia esencial entre ambas escuelas residía en el uso que hacen de la información previa al experimento. La escuela frecuentista produce inferencias a partir de los datos y de su conocimiento previo de la distribución del espacio muestral. Aunque se han hecho muchas bromas respecto a la validez de realizar inferencias a partir de algo más que los datos ([37]), el problema del

---

[36] No podemos detallarlas aquí. Esencialmente consisten en muestrear de una función conocida que "cubre" a la desconocida, y decidir por métodos simples si el punto muestreado pertenece a una o a otra función. Para poder "cubrir" a la función desconocida es necesario conocer su máximo, lo que en el caso multivariante representa un problema y hay que acudir a otras técnicas como el muestreo de Metrópolis-Hastings. Tanner (1993) trata con detalle las recetas de esta "cocina bayesiana".

[37] Berger y Wolpert (1984) recogen una buena colección de estas bromas. Sin embargo no deja de ser razonable hacer inferencias a partir no sólo de los datos sino del conocimiento previo de cómo se distribuyen.

frecuentismo reside en que para poder hablar de probabilidades sin usar información previa, como querían sus fundadores (ver Pearson, 1962), no hay más remedio que referirse a la muestra y no al parámetro que se desea estimar, lo que introduce una forma un tanto antinatural de expresar la incertidumbre asociada al resultado de un experimento. Fisher era perfectamente consciente de la pérdida que ello suponía:

> "Bayes percibió la importancia fundamental de este problema y propuso un axioma que, de ser cierto, bastaría para traer al campo de la probabilidad un gran número de inferencias inductivas; de forma que después de observar una muestra se podrían hacer inferencias acerca de la población, inferencias inciertas, por supuesto, pero con la bien conocida clase de incertidumbre que proviene de las leyes de la probabilidad". (Fisher, 1936).

La escuela bayesiana realiza inferencias a partir de probabilidades asociadas a valores del parámetro sobre el que se desea establecer el grado de conocimiento existente, lo que representa un medida mucho más natural de la incertidumbre. La inquietud al usar este procedimiento surge de que todo pueda ser una ilusión; es decir, de que en realidad no sean probabilidades las densidades posteriores debido a que no se ha podido precisar el grado de información previo o a que ese grado es inexistente o vago.

Las diferencias de fondo residen en que las dos escuelas dan respuestas diferentes al problema de la inducción([38]). En términos esquemáticos, este principio niega la posibilidad de la inferencia; por ejemplo, niega la posibilidad no ya de estimar con certeza parámetros poblacionales a partir de muestras, sino de estimarlos con una probabilidad determinada (Popper y Millers, 1983). La solución de Hume (1739) a este problema era de tipo psicológico: cuando observo una repetición de acontecimientos (por ejemplo, que las vacas Jersey producen menos que las Frisonas) espero que en futuro el comportamiento del fenómeno será el mismo. Obviamente esta solución no es satisfactoria y desde los primeros intentos de Kant (1787) hasta los más recientes de Russell (1948) y Popper (1972) se ha intentado dar una justificación racional de por qué a pesar de todo creemos que la repetición de un fenómeno tiene una causa y podemos estimarla, aunque las soluciones han sido poco satisfactorias. Russell (1948, pp.491) propone cinco postulados que, de admitirse (y esto es voluntario, por lo que el problema de la inducción sigue sin resolverse) justificarían la

---

[38] El problema de la inducción fue presentado en su forma moderna por Hume aunque como señala Copleston (1959, Vol. V, pp. 286) la esencia de ese problema -la dificultad de asociar efectos a causas- había sido ya planteada mucho antes. Kant buscó la justificación racional en el conocimiento innato (*a priori*). El a priori de Kant sólo tiene en común con el de Bayes el nombre. Kant dice: "Es por tanto una de las primeras y más necesarias cuestiones, y que no puede resolverse a simple vista, la de saber si hay algún conocimiento independiente de la experiencia y también de toda impresión sensible. Llámese a este conocimiento a priori, y distíngase del empírico en que las fuentes del último son a posteriori, es decir, que las tiene en la experiencia". Kant (1787, pp.98). Naturalmente ningún científico utiliza la información a priori como algo ajeno a la experiencia. Pese a ello, Kant es aún citado por algún matemático bayesiano (Robert, 1992, pp. 336) que no parece haberlo leído bien. La exposición más detallada que conozco del problema de la inducción se encuentra en Stove (1995, pp. 101-143).

inferencia, y Popper (1972) se limita a demarcar mejor el problema de la inducción sin aportar nada nuevo ([39]). En la actualidad se admite que el problema de la inducción no tiene una solución del tipo lógico-deductivo, y la investigación actual se centra en el examen de las "buenas razones" que hay para creer en este principio, y de la naturaleza de estas "buenas razones" ([40]).

Entre los estadísticos frecuentistas, Neyman prefería hablar de "comportamiento" de un científico ante los resultados del experimento antes que de inferencia científica. Fisher defendía, sin embargo, que era posible asociar una probabilidad a un suceso, y en una larga discusión con el estadístico bayesiano Jeffreys (recogida en Lane, 1980) este problema marcó las principales diferencias entre los dos planteamientos. En realidad la estadística frecuentista tampoco resuelve el problema de la inducción, por un lado porque no hay justificaciones claras para afirmar que se conoce con precisión la distribución del espacio muestral, y por otro porque cada medida de incertidumbre acaba por ir asociada a un parámetro poblacional que se desconoce (por ejemplo, el error estándar de la media depende de la varianza poblacional).

Una forma clara de comprender el problema de la inducción es examinar el teorema de Bayes. Para poder asociar una probabilidad a un acontecimiento debo saber su densidad de probabilidades *a priori*. Si tomo una muestra de vacas frisonas y observo que producen 400 litros de leche al año por término medio, debo saber que ese valor es muy improbable *a priori* para poder establecer una conclusión razonable. Sin embargo esta probabilidad a priori nos es desconocida, por lo que estoy basándome en mi grado de creencia para establecer la inferencia final. Por supuesto, en el caso de las vacas, y si soy una persona relacionada con la ganadería, tengo un convencimiento psicológicamente fuerte de que las vacas frisonas producen más de 400 litros de leche al año porque creo que son correctas mis lecturas a través de los años, los comentarios de los colegas en los Congresos, etc., pero estoy hablando de un estado de creencias, que será mucho más vago si el campo en el que estoy experimentando me es poco conocido.

Obsérvese que el uso del estado de creencias es perfectamente razonable, y no implica que no puedan ser compartidas por otros colegas de mi profesión. Un asunto diferente, y que coloca al paradigma bayesiano en una situación incómoda desde el punto de vista ontológico, es la imposibilidad psicológica que se puede dar cuando hay que establecer un grado de creencias multivariante, como es el caso habitual de la mejora genética animal. Debido a la dificultad de precisar bien el estado de creencias a priori, o el de ignorancia, el bayesianismo moderno tiende a considerar la función de densidad a priori como un artificio matemático, como hemos

---

[39] Popper sólo es considerado como un gran filósofo en ambientes científicos, probablemente por ser un filósofo fácil de leer y por ocuparse de temas relacionados con la ciencia. Stove (1995), realiza una crítica demoledora de su filosofía.

[40] El conocido filósofo americano Max Black ha hecho una revisión del estado actual de la investigación en torno al principio de inducción, y también de su relación con la probabilidad (Black, 1984).

dicho antes ([41]). Por otro lado, se ha producido una cierta aproximación al modo de trabajar bayesiano por parte de la estadística frecuentista mediante el uso de técnicas de boot-strap, que permiten tener una representación de la distribución muestral; aunque el problema de fondo sigue siendo el mismo, puesto que sigue sin poderse establecer la incertidumbre respecto al parámetro desconocido en términos de probabilidad. Así las cosas, la discusión sobre inferencia promete continuar produciendo una cantidad considerable de literatura, no necesariamente útil a la hora de ir a seleccionar conejos a las ocho de la mañana de cada Martes.

4.2. Ventajas e inconvenientes respecto a la aplicación

Hasta hace muy pocos años, la crítica fundamental que desde un punto de vista práctico se hacía a los estadísticos bayesianos era su esterilidad a la hora de producir resultados útiles. La metodología implicaba la resolución de integrales complicadas o el uso de aproximaciones más o menos afortunadas. La aparición de técnicas de Monte-Carlo en el muestreo de cadenas de Markov, de los que el popular "Gibbs sampling" es sólo un ejemplo([42]), ha resuelto la mayor parte de problemas operativos, y aunque ha generado otros nuevos, particularmente cuando los modelos no son Normales, la entidad de estos problemas operativos es ahora mucho menor. Las ventajas de disponer con facilidad de muestras realizadas de la densidad posterior vuelve a la estadística bayesiana particularmente atractiva, puesto que no es menester conocer la forma de la función para realizar inferencias: la media se estima directamente de los individuos muestreados, así como los intervalos de confianza u otros estimadores puntuales. El principal inconveniente se centra en que no se puede tener, al menos por el momento, la certeza matemática de que las cadenas de Markov utilizadas tengan las propiedades requeridas para garantizar que los puntos muestreados pertenecen, efectivamente, a la densidad posterior. Este tema se discute con más detalle en el Apéndice.

Las dificultades de las técnicas frecuentistas en mejora animal son de dos tipos: por un lado las derivadas del uso de grandes bases de datos y por otro las que derivan de la dificultad del frecuentismo para precisar el error estándar de un parámetro en situaciones medianamente complejas. Un ejemplo de lo primero es la dificultad de obtener estimas REML multivariantes cuando se dispone de algunas decenas de miles de datos, y un ejemplo de lo segundo es la dificultad de estimar con precisión el error de la respuesta a la selección cuando ésta se calcula mediante medias genéticas estimadas por BLUP usando parámetros genéticos

---

[41] Aunque esto es relativamente moderno. En un trabajo clásico, Lindley y Smith (1972) arguyen "que es típicamente cierto que hay información a priori disponible acerca de los parámetros, y que puede ser explotada para encontrar mejores y a veces sustancialmente mejores estimas [respecto a las de mínimos cuadrados]". Compárese esta afirmación con la de Bernardo y Smith, 1994, pp. 298 que figura al final del apartado 3.2.

[42] Tan popular, que hay quien lo confunde con un método de estimación. Josiah Willard Gibbs (1839-1903), célebre físico americano fundador de la termodinámica estadística, tiene poco que ver con estas técnicas. El nombre proviene de que estas cadenas de Markov generadas por muestreo aleatorio se aplicaron por primera vez a distribuciones inventadas por Gibbs, aunque más de 80 años después de la muerte de éste.

estimados mediante REML. Sin embargo hay casos en los que los métodos frecuentistas son rápidos y precisos (por ejemplo en el uso habitual de los índices de selección o el BLUP para estimar los valores aditivos de los individuos con objeto de realizar selección truncada) y no es fácil que técnicas más sofisticadas los desplacen ([43]). El abordaje bayesiano presenta, sin embargo ciertas ventajas en algunos casos:

1. *En el abordaje de nuevos problemas*: La forma de trabajar es clara y siempre la misma: obtener densidades posteriores de las variables a estimar. En esto se parece en cierto modo al método de máxima verosimilitud.

2. *Cuando se dispone de pocos datos*: Aunque la información usada en programas de mejora animal es habitualmente masiva, en ciertos experimentos la cantidad de datos está limitada por el coste o las facilidades disponibles. Incluso si el total de datos es grande pero la cantidad de datos por "efecto fijo" es pequeña -como es frecuente con el uso de animales en bandas, o en datos de campo tomados en granjas pequeñas- las estimas de los efectos fijos no son buenas y esto repercute en las estimas de los efectos aleatorios. Los métodos máximo verosímiles tienen propiedades asintóticas y podrían no ser útiles en estos casos, por lo que las densidades posteriores bayesianas sistematizan mejor el grado de ignorancia. Por otra parte, encontrar un máximo multivariante cuando hay pocos datos es poco seguro, la función de verosimilitud es muy plana y hay riesgos de caer en máximos locales; sin embargo las densidades bayesianas son habitualmente densidades marginales, por lo que su representación ilustra mucho mejor el grado de imprecisión del resultado que el dato del máximo y el error típico asociado (los perfiles de verosimilitud no son fiables con pocos datos, debido a las propiedades asintóticas de la verosimilitud que acabamos de comentar). Además, si no se desea confiar mucho en la información producida por los datos para las componentes de varianza, puede usarse información subjetiva (un ejemplo de esta forma de proceder se encuentra en Blasco et al., 1997); de una forma similar, los usuarios frecuentistas de BLUP usan a menudo componentes de varianza basadas en la bibliografía, sólo que el estadístico bayesiano puede precisar mejor el grado de ignorancia a priori.

3. *Cuando se desconoce la distribución en el espacio muestral del parámetro que se desea estimar*: Este es un caso frecuente, particularmente cuando el carácter analizado no tiene una distribución Normal. Por ejemplo, si se comparan dos curvas de Gompertz -lo que es habitual en estudios de crecimiento con mamíferos- no hay fórmulas que precisen el error típico de las diferencias entre parámetros de las curvas. La ventaja de la aproximación bayesiana, unida a las nuevas técnicas de muestreo de Monte-Carlo en cadenas de Markov, es que se pueden obtener muestras de una densidad posterior, y establecer inferencias bayesianas sin necesidad de realizar las operaciones matemáticas, frecuentemente irresolubles incluso por métodos numéricos, que implica la utilización de la fórmula matemática de la densidad

---

[43] Aunque la afición de los mejoradores animales al cómputo es tan notable, que ya aparecen BLUPs calculados mediante cadenas de muestreo de Gibbs (ver, por ejemplo, las últimas Jornadas de Producción Animal de Zaragoza (1997). Esto haría válida la ironía de uno de los autores del REML, Robin Thompson, cuando en el 6º Congreso mundial de genética aplicada a la mejora animal (Guelph, 1994) comentaba: "¿Qué es el Gibbs sampling?: ¡El BLUP con ruido!

posterior para hallar medias, modas o intervalos de confianza. Una solución de tipo general aplicable a cualquier curva ha sido propuesta por Varona et al. (1997), y dos aplicaciones, aún sin publicar en el momento de escribir estas líneas, han sido propuestas por Rekaya et al. (1998) y Blasco y Varona (1998) para curvas de lactación y de crecimiento respectivamente. Hay ocasiones en que el problema planteado tiene una solución bayesiana trivial, por ejemplo cuando el parámetro a estimar es combinación de otros parámetros estimados (por ejemplo un coeficiente de correlación). En estos casos, como se dispone de individuos muestreados de la densidad posterior conjunta, basta con tomar el valor de la covarianza y de las desviaciones típicas de cada individuo muestreado y calcularle un coeficiente de correlación. Todos ellos serán una muestra de la densidad posterior conjunta del coeficiente de correlación, con lo que se puede ya hacer inferencias sobre este coeficiente y calcular su precisión.

4. *En problemas de estimación jerarquizados*: Cuando se necesita un parámetro para realizar una estimación de otro parámetro es complicado o imposible hallar un error estándar que incluya ambos procesos de estimación. Por ejemplo, cuando se desea estimar el valor genético de los animales mediante un BLUP es necesario estimar las componentes de varianza de los efectos aleatorios. Es frecuente estimarlas mediante REML a partir de los mismos datos, pero cuando se calculan los errores estándar de los valores genéticos no se tiene en cuenta el error cometido al estimar las componentes de varianza, sino que se toman éstas como si fueran valores verdaderos (o estimados sin error). Una forma de abordar el problema desde la perspectiva bayesiana consiste, como hemos visto, en calcular la densidad posterior conjunta $f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2 | \mathbf{y})$ e integrar respecto a las componentes de varianza y al vector **b** para obtener la densidad posterior $f(\mathbf{u}|\mathbf{y})$. Esto equivale a dar a las componentes integradas todos los valores posibles y ponderarlos por sus probabilidades respectivas. Sorensen et al. (1994) han mostrado cómo hacerlo de forma sencilla utilizando técnicas de muestreo de Gibbs y ha sido aplicado primero por Wang et al. (1994) en porcino y posteriormente por un número creciente de autores.

5. *Cuando se imponen condiciones a los datos o a los parámetros a estimar*: Por ejemplo, en el caso de la durabilidad de las vacas en la explotación, se dispone de individuos con distinto número de lactaciones pero se desea estimar el número máximo posible. O por ejemplo en el caso de establecer el orden correcto de los sementales de un catálogo, y en general en cualquier problema de estadísticos ordenados. En ambos casos la solución bayesiana es trivial, en el primero usando técnicas de datos faltantes ([44]) y en el segundo la solución es inmediata a partir de las densidades posteriores obtenidas.

En cualquier caso, el mejorador que no quiera entrar en problemas de la filosofía de la inferencia, sino más bien usar útiles establecidos para resolver sus problemas, debe saber que las escuelas mencionadas están perfectamente establecidas y que no tiene sentido el tener que justificar el uso de un tipo de inferencia u otro al presentar sus resultados. Ninguna de las escuelas presenta hoy dificultades operativas salvo para casos especialmente complejos, y hay software disponible para analizar una gran variedad de problemas genéticos desde el punto de

[44] No puedo aquí entrar en el detalle de la técnica. Consiste esencialmente en hallar la densidad posterior suponiendo que se conocen los datos faltantes e integrarla luego poniendo los límites que describen la situación real.

vista de ambas. Escoger una u otra es un acto que para un mejorador aplicado depende esencialmente de qué problemas resuelve una que la otra no haya resuelto.

Ambas escuelas presentan problemas formales y paradojas. Suele ocurrir que se admiten con más facilidad las paradojas a las que el investigador se encuentra acostumbrado. También -por qué no decirlo- las revistas científicas encuentran "socialmente admisibles" los problemas mal resueltos de la escuela frecuentista (por ejemplo, aplicar un método máximo verosimil con pocos datos) antes que los de la bayesiana, aunque esto está cambiando ya. Cualquier refinamiento estadístico no debe hacer olvidar al investigador que el problema esencial de su trabajo reside en disponer de suficiente cantidad de información experimental para que sus conclusiones puedan expresarse de forma lo bastante precisa. La angustia del técnico o del científico que desearía expresar con certeza sus resultados, y no sólo con un grado de probabilidad, no es nueva; ya Virgilio en su Egloga IX, y en un contexto agrícola, hace exclamar a su protagonista: "¡Feliz aquél que puede conocer las causas de las cosas!".

AGRADECIMIENTOS

Este artículo fue revisado por un número considerable de colegas antes de su envio para publicación. Deseo expresar mi agradecimiento a todos ellos, y especialmente a mis amigos Daniel Gianola, Daniel Sorensen y Susie Bayarri por sus detallados comentarios.

REFERENCIAS

ANDERSON R.L., BANCROFT, T.A., 1952. Statistical theory in Research. Mc-Graw Hill.

BLASCO A., SORENSEN D., BIDANEL J.P., 1997. A bayesian analysis of genetic parameters and selection response for litter size components in pigs. Genetics: En prensa.

BLASCO A., VARONA L. 1998., Ajuste y comparacion de curvas de crecimiento. ITEA. Remitido.

BARNETT V., 1973/1982. Comparative Statistical Inference. Second edition in 1982. Chichester: Wiley.

BERGER J.O., WOLPERT R.L. 1984., The likelihood principle. Institute of Mathematical Statistics. Lecture Notes - Monograph Series. Series Editor, Shanti S. Gupta, Purdue University

BERNARDO J.M., 1979. Reference posterior distributions for bayesian inference. J. Roy. Statist. Soc. B 41:113-147.

BERNARDO J.M., SMITH A.F.M., 1994. Bayesian theory. Wiley

BLACK M., 1984. Inducción y probabilidad. Cátedra .

BOX G.E.P., TIAO G.C., 1973. Bayesian inference in statistical analysis. Reading, MA: Addison-Wesley.

CANTET R.J.C., FERNANDO R.L., GIANOLA D., 1992. Bayesian inference about dispersion parameters of univariate mixed models with maternal effects: theoretical considerations. Genet. Sel. Evol. 24:107-135.

COPLESTON F., 1963/1985. A History of Philosophy. Vol V: Hobbes to Hume. Bantam Doubleday Dell Publ. Inc.

DEMPFLE L., 1977. Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs bayesiens. Ann. Genet. Sel. Anim. 9:27-32.

EDWARDS A.W.F., 1992. Likelihood. Cambridge:University Press.

EFRON B., 1986. Why isn't everyone a bayesian?. The American Statistician 40 :1-11.

FISHER R.A. 1912. On an absolute criterion for fitting frequency curves. Messeng. Math. 41:155-160.

FISHER R.A. 1922. On the mathematical foundations of theoretical statistics. Phil. Trans., A. 222:309-368.

FISHER R.A. 1935. The logic of inductive inference. J. Roy. Stat. Soc. 98:39-54.

FISHER R.A., 1936. Uncertain inference. Proc. Am. Acad. Arts Sci. 71:245-258.

FISHER R.A. 1956. Statistical methods and scientific inference. Hafner Press.

GIANOLA D., FOULLEY J.L., 1990. Variance estimation from integrated likelihoods. Genet. Sel. Evol. 22: 403-417.

GIANOLA D., RODRIGUEZ-ZAS S., SHOOK G.E., 1994. The Gibbs sampler in the animal model: a primer. Séminaire Modele Animal. INRA Departament de Genetique Animale. Ed. J.L. Foulley y M. Molenat. pp. 47-56.

HARTLEY H.O., RAO J.N.K., 1967. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54:93-108.

HARVEY W.R., 1960. Least squares of data with unequal subclass numbers. U.S. Department of Agriculture, A.R.S. 20-28:1-157.

HARVILLE D., 1974. Bayesian inference for variance components using only error contrasts. Biometrika 61: 383-385.

HENDERSON C.R., 1949. Estimation of changes in herd environment. J. Dairy Sci. 32:706

HENDERSO C.R., 1950. Estimation of genetic parameters. Ann Math. Stat. 21:309-310.

HENDERSON C.R., 1953. Estimation of variance and covariance components. Biom. 9:226-252.

HENDERSON C.R., 1963. Selection index and expected genetic advance. En: Statistical Genetics and Plant Breeding, W.D. Hanson y H.F. Robinson (Ed.). National Academy of Sciences-National Research Council, Washington. pp., 141-153.

HENDERSON C.R., 1973. Sire evaluation and genetic trends. En: Proceedings of the animal breeding and genetics symposium in honor of Dr. L. Lush. Blacksburg, Virginia. pp.10-41.

HENDERSON C.R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biom. 32:69-84.

HENDERSON C.R., 1984. Applications of linear models in animal breeding. University of Guelph.

HUME D., 1739/1983. Tratado de la naturaleza humana. Orbis S.A.

JAMES W., STEIN C., 1961. Estimation with quadratic loss. Proc. Fourth Berkeley Symp., 1 (J. Neyman and E.L. Scott, eds.) Berkeley. Univ. California Press, pp. 361-380.

JEFFREYS H., 1931. Scientific inference. Cambridge University Press.

JEFFREYS H., 1961. Theory of probability. Oxford University Press.

KANT E. 1787/1983. Crítica de la razón pura. Orbis.

KEMPTHORNE O., 1984. Revisiting the past and anticipating the future. En: Statistics: an appraisal. Proceedings 50th Anniversary Iowa State Statistical Laboratory (ed: H.A. David and H.T. David). The Iowa State University Press. pp.31-52.

KENDALL M.G., 1961. Daniel Bernoulli on maximum likelihood. Biometrika 48:1-8.

KEYNES J.M., 1921. A treatise on probability. London. Macmillan.

KOTZ S., JOHNSON N.L., 1982. Encyclopedia of statistical sciences. Wiley .

LAMOTTE L.R., 1970. A class of estimators of variance components. Tech. Report 10, Univ. of Kentucky.

LANE D.A., 1980. Fisher, Jeffreys, and the nature of probability. En: Lecture Notes in Statistics (ed. S.E. Fienberg and D.V. Hinkley), Springer-Verlag Berlin Heidelberg. pp., 148-160.

LINDLEY D.V., SMITH A.F.M., 1972. Bayes estimates for the linear model. J. Royal Statist. Soc. Ser. B 34:1-41.

LUSH J.L., 1947. Family merit and individual merit as bases for selection. Amer. Nat. 81:241-261.

PATTERSON H.D., THOMPSON R., 1971. Recovery of interblock information when block sizes are unequal. Biometrika 58:545-554.

PEARSON E.S., 1962 Some thoughts on statistical inference. :275-283.

PEARSON K., 1903. Mathematical contributions to the theory of evolution. XI on the influence of natural selection on the variability and correlation of organs. Phil. Trans. of the Royal Soc. of London A, 200:1-66.

POPPER K.R., 1972. Conocimiento objetivo. Tecnos.

POPPER K., MILLER D., 1983. A proof of the impossibility of inductive probability. Nature 302: 687-688.

PRESS S.J., 1989. Bayesian statistics. Wiley.

RAO C.R., 1971. Minimum variance quadratic unbiased estimation of variance components. J. Mult. Anal., 1:445-456.

RAO C.R., KLEFFE J., 1988. Estimation of variance components and applications. North-Holland.

REKAYA R., CARABAÑO M.J., TORO M.A., 1998. Bayesian analysis of lactation curves with a non-linear model in lactation curves. J. of Dairy Sci. Remitido.

ROBINSON G.K., 1991. That BLUP is a good thing: the estimation of random effects. Statistical Science 6:15-51.

ROBERT C.P., 1992. L'Analyse statistique bayesienne. Economica.

RONNINGEN K., 1971. Some properties of the selection index derived by "Henderson's Mixed Model Method". Z. Tierz. Züchtbiol. 83: 186-193.

RUSSELL B., 1948/1983. El conocimiento humano. Orbis S.A.

SCHAEFFER L.R., 1986. Pseudo expectation approach to variance component estimation. J. Dairy Sci. 69: 2884-2889.

SCHEEFFE H., 1967. The analysis of variance. Wiley .

SORENSEN D.A., WANG C.S., JENSEN J., GIANOLA D., 1994. Bayesian analysis of genetic change due to selection using Gibbs sampling. Genet. Sel. Evol. 26:333-360.

SORENSEN D., 1997. Gibbs sampling in quantitative genetics. *National Institute of Animal Sciences. Internal rapport* No. 82. Tjele, Dinamarca, 188pp.

STIGLER S.M., 1986. The history of statistics. The measurement of uncertainty before 1900. Harvard University Press.

STIGLER S.M., 1983. Who discovered Bayes's theorem?. The American Statiscian 37:290-296.

STOVE D.C., 1995. Popper y después. Cuatro irracionalistas contemporáneos. Tecnos.

STUART A., ORD J.K., 1991. Kendall's avanced theory of statistics Vol.2. Arnold

TANNER M.A., 1993. Tools for Statistical Inference. Springer-Verlag.

THOMPSON W.A., 1962. The problem of negative estimates of variance components. Ann. Math. Stat. 33: 273-289.

THOMPSON R., 1987. Statistical methods and applications to animal breeding. Thesis presented for degree of Doctor of Science. University of Edinburgh.

Von MISES R., 1928/1957. Probability statistics and truth. London. Macmillan.

VARONA L., MORENO C., GARCIA-CORTES L.A. ALTARRIBA J., 1994. Multiple trait analysis of underlying biological variables of production functions. Livest. Prod. Sci. 47:201-209.

WANG C.S., GIANOLA D.,SORENSEN D.A., JENSEN J., CHRISTENSEN A., RUTLEDGE, J.J., 1994. Response to selection for litter size in Danish Landrace pigs: a bayesian analysis. Theor. Appl. Genet. 88:220-230.

WOOLLIAMS J.A., MEUWISSEN T.H.E., 1993. Decision rules and variance of response in breeding schemes. Anim. Prod. 56:179-186.

YATES F., 1990. Foreword. En: R.A. Fisher, Statistical methods, experimental design and scientific inference. Oxford Univ. Press.

APENDICE I. Muestreo de Gibbs

Se pretenden muestrear valores *al azar* de la función de densidad posterior $f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y})$ Para ello se procede de la siguiente forma:

Se toman valores arbitrarios de inicio para $\mathbf{u}, \sigma_u^2, \sigma_e^2$, y se muestrea un valor al azar de la función condicional

$$f(\mathbf{b} \mid \mathbf{u}, \sigma_u^2, \sigma_e^2, \mathbf{y})$$

Con el valor de **b** así muestreado, y con los valores anteriores de σ_u^2, σ_e^2, se muestrea un valor al azar de la función condicional

$$f(\mathbf{u} \mid \mathbf{b}, \sigma_u^2, \sigma_e^2, \mathbf{y})$$

Con el valor de **u** así muestreado, con el de **b** muestreado antes y con el de σ_e^2 inicial, se muestrea un valor al azar de la función condicional

$$f(\sigma_u^2 \mid \mathbf{b}, \mathbf{u}, \sigma_e^2, \mathbf{y})$$

Con el valor de σ_u^2 así muestreado y con los de **b** y **u** muestreados antes, se muestrea un valor al azar de la función condicional

$$f(\sigma_e^2 \mid \mathbf{b}, \mathbf{u}, \sigma_u^2, \mathbf{y})$$

con los valores muestreados $\mathbf{u}, \sigma_u^2, \sigma_e^2$, se reinicia el proceso. Finalmente, al cabo de un cierto número de iteraciones, los valores que se muestrean pertenecen a la densidad conjunta posterior $f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y})$. Cada valor es, además, miembro de cada una de las distribuciones marginales $f(\mathbf{b}|\mathbf{y})$, $f(\mathbf{u}|\mathbf{y})$, $f(\sigma_u^2 \mid \mathbf{y})$, $f(\sigma_e^2 \mid \mathbf{y})$, que permiten hacer inferencias por separado en cada parámetro.

Lo mismo se haría para realizar inferencias sobre combinaciones lineales de efectos "fijos" y aleatorios. Si no se sabe cómo extraer muestras de una función directamente, puesto que la función condicional puede tener una forma complicada, se utilizan muestreos en "caja negra",

por lo que casi siempre es posible realizar un proceso de muestreo de Gibbs. Los problemas asociados a este procedimiento son tres:

1) *No se puede demostrar que se está muestreando de la función de densidad posterior.* Para que una cadena de Markov converja a una densidad posterior debe cumplir ciertas propiedades y ser una cadena *irreducible*. Aunque se puede demostrar que ciertas cadenas *no* son irreducibles, no hay una demostración general que asegure que la cadena que estamos usando tiene esta propiedad (los detalles pueden consultarse en los apuntes del curso que impartió Daniel Sorensen en la Escuela de Agrónomos de Valencia; Sorensen, 1997).

2) *No se puede saber con certeza a partir de qué iteración los puntos extraídos pertenecen a la función de densidad posterior.* Es decir, aunque supiéramos que nuestra cadena es irreducible, no sabríamos a partir de qué iteración hemos llegado a la convergencia.

3) *No se puede asegurar que la cadena de Markov haya convergido a una distribución estacionaria.* A veces se encuentran zonas de tránsito que dan la impresión de que pertenecen ya a la distribución de convergencia, pero continuando el muestreo un número suficiente de veces la cadena evoluciona hacia otra distribución.

Estas dificultades se podrían resumir diciendo que no sabemos cuándo ni a qué hemos convergido, en el caso de que hayamos convergido a algo. Como se ve, el problema no es trivial y es en la actualidad objeto de una investigación intensiva por matemáticos especializados en cadenas de Markov. Los científicos aplicados se conforman con iniciar varias cadenas desde puntos de partida diferentes y comprobar que aparentemente se converge a la misma distribución. Este convencimiento es de tipo psicológico, pero suficiente para la mayor parte de usuarios. No es diferente del convencimiento psicológico que tiene quien aplica el método de máxima verosimilitud a un problema multivariante y cree que ha obtenido un máximo global en lugar de uno local. O del convencimiento de quien estima una tendencia genética y cree que los parámetros genéticos que ha usado (y que van a determinar sus resultados) son próximos a los verdaderos. Ningún método operativo de cualquiera de las dos escuelas, frecuentista o bayesiana, está hoy por hoy libre de problemas, aunque hay esperanzas razonables de que en ambos casos se vayan encontrando soluciones aceptables para todos.

ESTIMA DE PARAMETROS GENETICOS PARA PROLIFICIDAD EN GANADO OVINO MEDIANTE MUESTREO DE GIBBS. RESULTADOS PRELIMINARES

Mª. José Espinosa Pascual
Juan José Jurado García
Area de Mejora Genética y Biotecnología.
CIT-INIA
Crta de La Coruña, Km. 7,00
28040 MADRID

INTRODUCCION

La cooperativa Carne-Aragón S.C.L. esta llevando a cabo un programa de selección genética en ganado ovino de raza Rasa Aragonesa en colaboración con la DGA de Aragón y el INIA de Madrid, cuyo objetivo de selección es incrementar la productividad numérica, utilizando como criterio de selección la prolificidad de las ovejas. Este programa se puso en marcha a finales de 1995.

La metodología BLUP es generalmente, aceptada como la forma más eficiente y versátil de obtener una clasificación correcta de los animales en función de su valor genético. En condiciones adecuadas las predicciones genéticas son insesgadas y de mínima varianza. No obstante hay considerables reservas de tipo teórico acerca de su posible aplicación a caracteres de tipo discreto.

La prolificidad es un carácter que presenta distribución fenotípica discontínua, con un reducido número de clases y claramente asimétrica. Estas características ponen en cuestión que su análisis pueda llevarse a cabo mediante la técnica de los modelos lineales mixtos. No obstante dicha naturaleza discontínua, en ocasiones, se suele ignorar aduciendo que los modelos mixtos es una metodología robusta que soporta anomalías en la distribución de los carácteres y que permite una aceptable ordenación de los animales en función de su valor genético.

La definición del modelo umbral por Wright (1934) y su posterior desarrollo por Demspter y Lerner(1950), en el caso de un solo umbral, puso en evidencia que las heredabilidades en la escala subyacente y en la escala visible pueden ser diferentes y de mayor magnitud en la primera que en la última, estando ambas relacionadas en función de la incidencia del carácter y del valor de la ordenada en el valor umbral. Gianola (1982) generaliza estos resultados a un mayor
número de umbrales.

Gianola y Foulley (1983) y Harville y Mee (1984) proponen métodos con modelos no lineales con el propósito de obtener predicciones genéticas, pero presentan grandes problemas computacionales cuando se pretende aplicar un modelo animal con medidas repetidas sobre los datos de una población de gran tamaño. Foulley y col (1987) proponen la utilizacion de un modelo Poisson con estos mismos propósitos, no encontrando ventajas respecto a otros modelos.

Desde un punto de vista teórico, y puesto que la escala subyacente es más heredable, sería deseable que la elección de los reproductores se basara en los valores geneticos predichos sobre la escala subyacente mas que en los predichos sobre la visible, lo que implica acceder de algún modo a dicha escala. Sorensen y col. (1995) proponen un método para efectuar inferencias bayesianas en modelos umbrales mediante muestreo de Gibbs, basado en el muestreo de funciones de densidad truncadas que conducen a distribuciones posteriores de los parámetros genéticos y de las predicciones genéticas de los animales en la escala subyacente.

El propósito de este trabajo es presentar los primeros resultados de la aplicación de la metodología descrita por Sorensen y col (1995) mediante un modelo animal con medidas repetidas, al esquema de selección genética de Carne-Aragón, en especial a la obtencion de estimas de heredabilidad y repetibilidad de la prolificidad, obtenidas de formas diferentes y en las dos escalas.

MATERIAL Y METODOS

Los datos utilizados fueron parte de los acumulados en la base de datos del programa de selección y que se emplearon en la ultima valoración de reproductores en Abril de 1998. El número total de animales en dicha base de datos fue de 36159 ovejas en 44 rebaños que sumaban un total de 107921 partos. Hasta la fecha hay datos de 27 machos con 366 hijas controladas

Los datos utilizados fueron los correspondientes a los rebaños conectados, entendiéndose por tales los que tenían al menos 8 hijas de machos de referencia. El número de ovejas en estas circunstancias fue de 21605 con un total de 71130 partos y 24 machos. El modelo utilizado fue el siguiente:

$$Y_{ijkl} = \mu + RA_i + Est_j + Np_k + Itp_l + Mc_m + U_n + Ep_n + \varepsilon_{p(ijklmn)}$$

En donde:

Y_{ijkl} el la prolificidad de la oveja en el parto

μ es la media general de la población

RA_i es el grupo de comparación rebaño- año de parto (108 niveles).

Est_j es el efecto de la estación de parto (12 niveles).

Np_k es el efecto del numero de parto (10 niveles)

Itp_l es el efecto del intervalo parto y siguiente cubrición (6 niveles: ovejas de primer parto, <90 días, 90-160 dias, >160, anomalías: intervalo entre parto < 150 dias, >600 dias)

Mc_m es el efecto del modo de cubrición (3 niveles, sincronización de celos sin IA, sincronización mas IA, monta natural)

U_n es el efecto genético del animal que produce el dato

Ep_n es el efecto ambiental permanente

$\varepsilon_{p(ijklmn)}$ es el residuo del modelo

Para estimar los parámetros genéticos se utilizaron diversos métodos con el fin de comparar resultados.

Sobre la escala visible se utilizó el programa VCE de Groeneveld (1998) que implementa la metología REML usando un método de gradiente analítico para maximizar la función de verosimilitud.

También se utilizó sobre la escala visible el muestreo de Gibbs según Wang(1994), realizandose un muestreo de los efectos fijos, efectos aleatorios y las varianzas genéticas, del efecto permanente y residual. Para los efectos fijos se asumieron "priors" uniformes de modo que todo los valores tuvieran la misma probabilidad y para los componentes de varianza se asumieron distribuciónes χ^2 invertidas con bajos grados de credibilidad (ν_e=4). Se limitó el campo de variación de las varianzas entre limites amplios.

Para trabajar sobre la escala subyacente se utilizó la metodología propuesta por Sorensen (1995), mediante la técnica conocida como de "aumento de datos" (data augmentation) citada por el mismo autor. De acuerdo con el modelo umbral se asume la existencia de una escala subyacente sobre la que se definen los valores fenotípicos de las ovejas (y_s) correspondientes a los de la escala visible. Se asume asimismo un modelo lineal igual al usado en la escala visible en el que la variable es y_s. Se asume también la existencia de unos umbrales (t_1, t_2, y t_3) que determinan unos intervalos en la distribución subyacente. En cada iteración el valor y_s de cada parto se obtiene mediante la suma de los correspondientes efectos

fijos, genéticos y permanentes calculados en la iteración anterior (y_{sc}), mas el efecto del residuo (ε_s). Este ultimo se deduce a partir de una distribución normal de media y_{sc} y varianza la obtenida como residual en la iteración anterior. Si la oveja es de parto simple el valor ε_s se toma al azar de la distribución entre $-\infty$ y t_1, si es de parto doble se toma entre t_1 y t_2, si es de parto triple se toma entre t_2 y t_3 y si es cuádruple o mas se toma entre t_3 y $+\infty$. Una vez muestreados los y_s, se muestrean los efectos fijos, los efectos aleatorios y las componentes de varianza de forma semejante a la citada anteriormente. La estructura de "priors" utilizada fue la misma que la usada para el caso de la escala visible.

Los umbrales elegidos fueron los deducidos a partir de las frecuencias fenotípicas en la escala visible. La aplicación del método conduce a obtener estimas de los parámetros genéticos y de los efectos fijos y predicciones de los valores genéticos de todos los animales como es propio del modelo animal. Puesto que los umbrales se asumen fijos, el proceso de muestreo conduce a que las varianzas tomen valores según los umbrales elegidos, pero los parámetros genéticos obtenidos a partir de las mismas son constantes con independencia de los umbrales fijados. Las cadenas de muestreo utilizadas tuvieron longitudes variables. Se descartaron las primeras iteraciones y a partir de una determinada se tomaron periódicamente algunas de ellas.

También se utilizó la técnica de cadenas pareadas (coupling method) aplicada por Garcia-Cortés y col (1997) al modelo animal, con el fin de estimar el número de iteraciones que deberían ser descartadas previamente para conseguir buenas inferencias marginales. El periodo de calentamiento se consideró terminado cuando las diferencia entre los valores de los parámetros en ambas cadenas pareadas fue inferior a un valor prefijado.

Con el fin de estudiar la autocorrelación entre los puntos que integran la distribución marginal se calcularon las estimas de parámetros usando una única cadena de gran longitud con un periodo de calentamiento variable, así como también fue variable el intervalo entre las iteraciones elegidas. En todos estos casos se usan cadenas pareadas para vigilar la convergencia del proceso. La autocorrelación y el tamaño efectivo del conjunto de puntos utilizados para formar la distribución marginal se calculó por la técnica descrita por Cox y Miller (1965) y utilizada por Sorensen (1995).

RESULTADOS

En la tabla 1 se presenta una síntesis de los resultados obtenidos hasta este momento, por lo que éstos deben ser considerados como preliminares.

Los valores obtenidos en la escala visible para la heredabilidad (h^2) de 0.059± 0.006 y para la repetibilidad (r_{ep}) de 0.102±0.006 mediante REML están dentro del orden de valores obtenidos por otros autores para esta raza. Occon(1988), obtiene valores de h^2 para el primer parto que oscilan entre 0.07 y 0.21. La r_{ep} oscila entre 0.11 y 0.16.

Cuando se utiliza el muestreo de Gibss en la escala visible se usó una única cadena con 300000 iteraciones, un periodo previo de 10000 y se eligió una de cada 70 iteraciones. En total se obtuvieron 4142 puntos que integran las distribuciones marginales. Los valores de los parámetros son muy semejantes a los obtenidos mediante REML. (h^2=0.049±0.004 y r_{ep}=0.106±0.002). En este caso se observa en la h^2 una alta autocorrelación (0.77) lo que supone un tamaño efectivo de 491.5. En cambio en la r_{ep} esta autocorrelación es de 0.105 con un tamaño efectivo de 2890,18, lo cual indicaría que los puntos que integran la distribución marginal de la r_{ep} son más independientes que los de la h^2, debido seguramente a que la r_{ep} implica dos varianzas muestreadas frente a solo una de la h^2.

TABLA 1.- Estimas de la heredabilidad y repetibilidad obtenidas por REML y muestreo de Gibss en las escala visible y subyacente, asi como autocorrelación y tamaño efectivo.

METOD.	ESCALA	LOGITUD CADENA	PERIODO PREVIO	INT	PARAMETROS. GENETICOS	AUTOCORR. TAM. EFEC.
REML	VISIBLE				h^2=0.059±0.006 r_{ep}=0.102±0.006	
GIBBS	VISIBLE	300000	10000	70	h^2=0.049±0.004 r_{ep}=0.106±0.002 n=4142	0.77 491.5 0.106 2890.2
GIBBS	SUBYAC	30100	3000	50	h^2 =0.101±0.008 r_{ep}=0.164±0.004 n=542	0.74 40.02 0.136 425.5
GIBBS	SUBYAC	50000	6000	200	h^2=0.094±0.008 r_{ep}=0.168±0.004 n=220	0.541 52.13 -0.015 190.7

En la aplicación del muestreo de Gibbs a la escala subyacente, las cadenas pareadas tuvieron 30100 iteraciones con un periodo previo de 3000 y se tomó una de cada 50 iteraciones. En cada una de las cadenas los valores iniciales de los parámetros genéticos fueron diferentes, pero no así los efectos fijos. Se observó que en la iteración 3000, la diferencia máxima entre ambas cadenas para las soluciones del modelo mixto era del orden de 0.01. En el caso de las varianzas genéticas y permanentes dicha diferencia entre cadenas fue de 0.0004 lo que supone 1% de la diferencia inicial y en el caso de la varianza residual fue de 0.0001, un 0.005% de la inicial. Esto implica que el periodo de calentamiento fue suficiente para asegurar que las cadenas habían convergido y se estaba ya muestreando de la distribución conjunta (Figura 1). El valor obtenido de h^2 fue de 0.101 ±0.008 y el de la r_{ep} de 0.164±0.004, que son claramente superiores a los obtenidos trabajando sobre la escala visible. Este mayor valor de h^2 en la escala subyacente que en la visible era esperable según la teoría y demuestra la bondad del método propuesto por Sorensen (1985). Las predicciones genéticas obtenidas de esta forma tienen la ventaja de estar corregidas por los efectos ambientales según un modelo animal con medidas repetidas(como es propio de la metodología BLUP) y, además, están basadas en unas medidas del carácter mucho más heredables como es la escala subyacente. El número de puntos que en este caso integran las distribuciones marginales fue de 542. La autocorrelación para h^2 fue de 0.747 lo que supone un tamaño efectivo de 40.02. Para la r_{ep} dicha autocorrelación fue de 0.136 con un tamaño efectivo de 425.5. Nuevamente se presenta la misma circunstancia que anteriormente, los puntos que integran la distribución marginal de la h^2 son menos independientes que los que integran la r_{ep}.

Con el propósito de estudiar si la alta autocorrelación afectaba de alguna forma a las estimas de los parámetros genéticos, se corrió otra cadena semejante a la anterior de 50000 iteraciones, pero con un periodo previo más largo (6000) y tomando una de cada 200 iteraciones. En total se obtuvieron 220 puntos En este caso los valores alcanzados para h^2 (0.094±0.008) y r_{ep} (0.168±0.004) fueron del mismo orden que en el caso anterior con un autocorrelación mas baja (0.53) y un tamaño efectivo de 52.13 para la h^2 y de –0.051 y 190.69 para r_{ep} respectivamente. Esto podría ser un indicio de que la autocorrelación no afecta de forma notable a las estimas de los parámetros genéticos

CONCLUSIONES

Las conclusiones a las que podemos llegar en función de los resultados obtenidos

serian las siguientes.

a) Los valores estimados de la heredabilidad y repetibilidad en la escala visible en esta población de ovino de raza Rasa-Aragonesa fueron de 0.049±0.004 y 0.107±0.002 respectivamente.

b) En la escala subyacente el valor de dichos parámetros fueron: para la heredabilidad de 0.101±0.008 y para la repetibilidad de 0.164±0.004, considerablemente mayores que en la escala visible.

c) Las predicciones genéticas de los animales de la población para el carácter prolificidad en la escala subyacente es posible obtenerlas mediante la técnica del aumento de datos y muestreo de Gibbs.

d) Aunque se presenta autocorrelación entre los puntos integrante de las distribuciones marginales, esto no parece afectar de forma notable a las estimas de los parámetros genéticos.

e) Parece razonable que las predicciones genéticas de los animales en la escala subyacente deberían ser el criterio de selección de reproductores en razón de su mayor heredabilidad.

BIBLIOGRAFÍA

COX, D.R., MILLER, H.D. 1965.- The theory of Stochastic Process. Chapman and Hall, London U.K.DEMPSTER, .R., LERNER, I.M. 1950. Heritability of threshold characters.Genetics,35:212.FOULLEY, J.L., GIANOLA, D., IM, S. 1987.- Genetic evaluation of traits distributted as Poisson-binomial with reference to reproductive traits. Theor. Appl. Genet. 73:870

GARCIA-CORTES, L.A., RICO, M., GROENEVELD,E. 1997. Coupling the gibbs Sampler to asses convergence in animal Models. Fotocopia

GIANOLA, D. 1980. A method of sire evaluation for dichotomies. J. Anim. Sci. 51: 1266-1271.

GIANOLA, D. 1982. Theory and analysis of treshold characters. J. Anim. Sci. 54: 1079-1096.

GIANOLA, D., FOULLEY, J.L. 1983. Sire evaluation for ordered categorical data with a threshold model. Gen. Sel. Evol. 15(2): 201-224.

GROENEVELD, E. 1998.- Programa VCE 4, version 4.2.

HARVILLE, D.A., MEE, R.W. 1984. A mixed-model procedure for analising ordered categorical data. Biometrics, 40: 393-408.

HOESCHELE, I., TIER, B. 1995. Estimation of variance components of treshold characters by marginal posterior modes and means via Gibbs sampling. Genet. Sel. Evol. 27: 519-540.

OCCON, A. 1988. Repetibilidad y heredabilidad de la prolificidad en la raza *rasa Aragonesa*. Optimización de los progresos genéticos en los procesos de selección. Tesis doctoral.

SORENSEN, D., ANDERSEN, J., GIANOLA, D., KORSGAARD, I. 1995. Bayesian inference in treshold models using Gibbs sampling. Genet. Sel. Evol. 27: 229-249.

WANG, C.S., RUTLEDGE,J.J., GIANOLA,D., 1994.- Bayesian analysis of mixed models via gibbs sampling with application to litter size in Iberian pigs. Genet. Sel. Evol. 26, 91-115.

WRIGHT. 1934. An analysis in number of digits in an inbred strain of guinea pigs. Genetics, 19: 506.

La huella de Ruperta

J. Rodrigáñez, M. Toro, C. Rodríguez y L. Silió

Departamento de Mejora Genética y Biotecnología, I.N.I.A.
Ctra. de La Coruña km 7, 28040 Madrid

Introducción

El análisis de la información genealógica de una población permite describir el proceso de cambios en variabilidad genética a lo largo de su historia. Recientemente, se han aplicado en algunas poblaciones de animales domésticos diversos métodos desarrollados por especialistas en conservación de poblaciones de animales salvajes mantenidas en cautividad (Lacy y col., 1995). Estos trabajos abordan el balance de la gestión de programas de conservación (Rodrigáñez y col., 1997a), el efecto sobre la variación genética de la estructura de grandes poblaciones (Söllkner y Filipcic, 1997), la contribución de grupos genéticos bien definidos (Gandini y col., 1997) o la identificación del número efectivo de ascendientes y de cuellos de botella en poblaciones con diverso uso de inseminación artificial (Boichard y col., 1997). A partir de estas técnicas, se ha implementado una partición del coeficiente de consanguinidad en componentes atribuibles a los diferentes fundadores o a agrupaciones de éstos, que ha permitido analizar la variación de depresión consanguínea entre linajes de fundadores (Lacy y col., 1996; Rodrigáñez y col., 1998).

El presente trabajo es una aplicación de estos métodos al análisis de la variación genética y de la depresión consanguínea en tres poblaciones de cerdo ibérico emparentadas, que comparten un destacado animal fundador, la cerda *Ruperta*, cuyas huellas intentaremos rastrear por diversas vías en estas poblaciones.

Tabla 1. *Resumen de información genealógica y de prolificidad en las tres estirpes*

	Guadyerbas	*Gamito*	*Torbiscal*
Nº de fundadores	23	14	71
Nº de animales en genealogia	1020	793	3489
Nº de verracos	243	196	761
Nº de madres con datos	670	458	1453
Nº de camadas	2020	1308	5562
Lechones nacidos por camada	7.36 (2.29)	7.81 (1.99)	8.38 (2.30)
Lechones nacidos vivos por camada	7.02 (2.18)	7.48 (1.94)	8.00 (2.22)

Material y Métodos

Tres poblaciones cerradas

Las poblaciones consideradas en este estudio (*Guadyerbas*, *Gamito* y *Torbiscal*) forman parte de la piara experimental de cerdos ibéricos establecida desde 1944-45 en la finca *Dehesón del Encinar* (Oropesa, Toledo). La formación de esta piara se realizó a partir de animales de cuatro estirpes representativas de los principales tipos de cerdo ibérico existentes en la amplia

y heterogénea población de la época: dos portuguesas coloradas: *Ervideira* y *Caldeira* y dos extremeñas negro lampiñas: *Campanario* y *Puebla* (Odriozola, 1976).

Guadyerbas es la actual denominación de la estirpe negro lampiña *Puebla*, que aun se conserva como población cerrada (Rodrigáñez y col., 1997b). Entre las cerdas fundadoras de ésta, se identificó una de ellas heterocigota para el par de alelos *Ee* (Negro/colorado). A partir de esta cerda, de nombre *Ruperta*, se formó una línea colorada (*ee*), derivada exclusivamente de *Puebla* mediante tácticas de infiltración progresiva del gen recesivo en cuestión. La línea *Gamito*, colorada lampiña, fue el resultado de este proceso. Por último, *Torbiscal* es una línea compuesta, que procede de la fusión de las cuatro estirpes fundadoras realizada lentamente entre 1961 y 1963 para que pudieran contribuir a la misma el mayor número posible de reproductores. Se dispone de la genealogía completa y registros reproductivos de todas estas poblaciones. La información utilizada en el presente trabajo se presenta en la Tabla 1.

Análisis genealógico

Para el análisis genealógico, los reproductores de cada población se agruparon, de acuerdo con su fecha de nacimiento, en cohortes trianuales. Un fundador se define como un animal sin relación genética conocida con ningún otro de la genealogía, exceptuados sus descendientes. La contribución genética esperada de cada fundador a un animal determinado se mide mediante los coeficientes de parentesco entre el animal y la población de fundadores (James, 1972). La contribución genética proporcional de cada fundador a la población (cohorte) analizada se calcula mediante la expresión $p_i = \sum_{j=1}^{N} a_{ij} / N$, donde a_{ij} es el coeficiente de relación aditiva entre el fundador i y el animal j, siendo N el número de animales de la cohorte considerada. Los valores de p_i se denominan también probabilidades de origen genético y permiten calcular el número efectivo de fundadores $N_{fe} = 1 / \sum_{i=1}^{n} p_i^2$, siemdo n el número de fundadores (Lacy, 1989). Otro parámetro es el número efectivo de ascendientes N_{fa}, basado en las contribuciones marginales de los antepasados necesarios para explicar la diversidad genética de la población, y que identifica los posibles cuellos de botella en la genealogía (Boichard y col., 1997).

Debido al número finito de descendientes de cada reproductor, existe el riesgo de extinción de genes en la transmisión de padres a hijos. Mediante la técnica de goteo de genes (MacCluer y col., 1986), que simula la segregación mendeliana a través de la genealogía, es posible estimar la proporción de alelos de cada fundador (r_i) retenida en el censo de cada cohorte, así como el número esperado de alelos de cada fundador ($2r_i$) y de todos los fundadores ($\Sigma 2r_i$). Estos análisis se realizaron con 250.000 iteraciones por estirpe. El número equivalente de genomas fundadores $N_{ge} = 1 / \sum_{i=1}^{n} p_i^2 / r_i$ es otro parámetro, originalmente propuesto por Lacy (1989), que mide la diversidad genética preservada en la población. Con el mismo objetivo Boichard y col. (1997) han propuesto el número efectivo de genomas fundadores.

El coeficiente de consanguinidad (F) puede descomponerse en coeficientes parciales (F'), que miden las probabilidades de que un animal sea homocigoto para un alelo descendiente de cada fundador específico. Estos coeficientes se han calculado usando una modificación del método clásico de cálculo de coeficientes de consanguinidad (Lacy y col., 1996; Lacy, 1997). Dado el número de fundadores, la descomposición de la consanguinidad realizada en cada estirpe fue

algo más simple, reflejando las características de su origen. Así en *Guadyerbas* y *Gamito* se consideraron dos componentes: F^R, consanguinidad debida a la fundadora *Ruperta* y F^{NR}, contribución del resto de los fundadores de la estirpe *Puebla* a la consanguinidad del individuo. En *Torbiscal*, junto a estos dos componentes de la consanguinidad se consideraron separadamente las sumas de los coeficientes parciales debidos a los fundadores de las otras tres estirpes *Ervideira* (F^A), *Campanario* (F^B) y *Caldeira* (F^C). A efectos del análisis de depresión consanguínea, se realizó el cálculo de estos coeficientes tanto para las cerdas con registros de prolificidad (F_m) como para las camadas nacidas (F_c).

Análisis de depresión consanguínea

El análisis de los registros de tamaño de camada fue realizado, separadamente en cada población, mediante un modelo animal con repetibilidad:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{p} + \mathbf{e}$$

siendo como es habitual **y**, el vector de observaciones; **b**, el vector de efectos sistemáticos estación (4 niveles), ordinal de parto (5 niveles) y los coeficientes de regresión de los correspondientes coeficientes de consanguinidad (b_{Fm} y b_{Fc}, en *Torbiscal* y b_{Fc}^R y b_{Fc}^{NR}, en las otras dos estirpes); **u**, **p** y **e**, son los vectores de efectos genéticos aditivos, de ambiente permanente y de residuos; **X**, $\mathbf{Z}_1$ y $\mathbf{Z}_2$ son matrices de incidencia que relacionan **u**, **p** y **e** con las observaciones en **y**.

La implementación del análisis bayesiano de este modelo mediante muestreo de Gibbs ha sido descrita por Wang y col. (1994), precisamente con una aplicación a registros procedentes de la estirpe *Gamito*. Otros aspectos relativos a las características del muestreo y análisis post-Gibbs son similares a los descritos por Rodrigáñez y col. (1998).

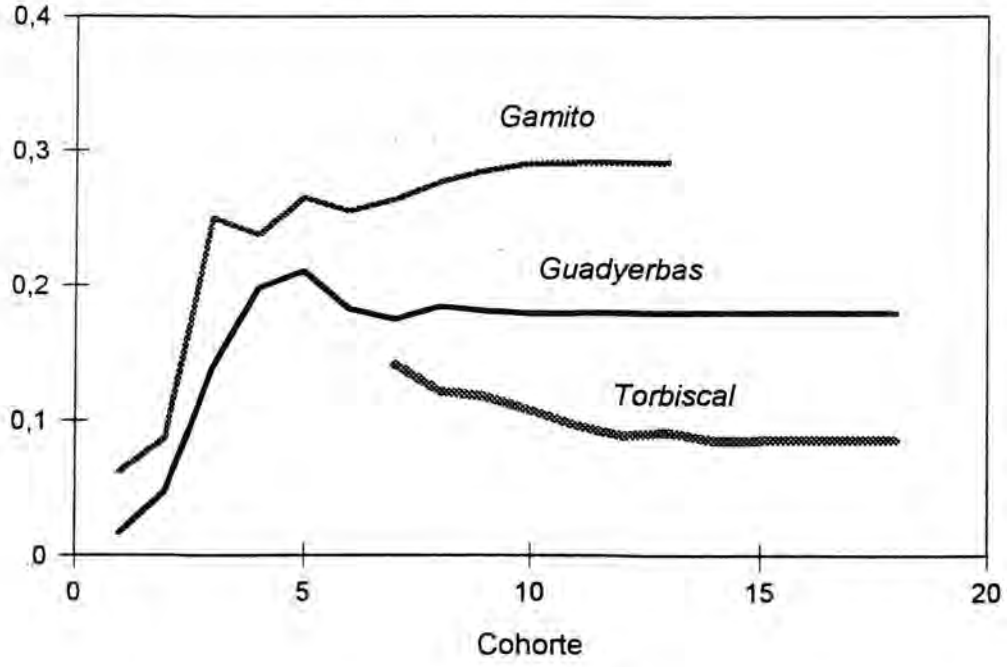


Figura 1. *Contribución genética proporcional (p) de la fundadora Ruperta a las estirpes Guadyerbas, Gamito y Torbiscal.*

Resultados y discusión

El número de cohortes trianuales analizadas ha sido 18 en *Guadyerbas* (1945-1997), 13 en *Gamito* (1945-1983) y 12 en *Torbiscal* (1963-1997). Entre todos los animales fundadores, la cerda *Ruperta* es la de mayor contribución a las tres poblaciones. En la Figura 1 se representa su contribución proporcional a las tres estirpes, que tras fluctuaciones iniciales se estabilizó con valores respectivos de 0.18, 0.29 y 0.09. A lo largo de los periodos analizados, el número efectivo de fundadores N_{fe} pasó de valores iniciales de 11.54, 8.00 y 22.29, en *Guadyerbas*, *Gamito* y *Torbiscal* a valores finales de 10.35, 6.81 y 35.53.

La importancia relativa de *Ruperta* se observa, aunque de forma algo más borrosa, si se utiliza un parámetro alternativo: el número efectivo de ancestros (fundadores o no). Pese a su elevada contribución a *Guadyerbas*, *Ruperta* no figura entre los 13 ancestros que explican la diversidad genética de la población actual (N_{fa} = 6.81). Ello puede atribuirse a un cuello de botella existente en la población hacia 1965. En *Gamito* son también 13 los ancestros (N_{fa} = 5.37) que explican la diversidad de la última cohorte. Aunque la contribución marginal de *Ruperta* es la décima en importancia, dos de sus hijos y uno de sus nietos ocupan los tres primeros puestos. Por último, 54 ancestros explican la diversidad actual de *Torbiscal* (N_{fa} = 14.45), siendo *Ruperta* la única fundadora, la principal hembra y el sexto ancestro en cuanto a contribución marginal.

En cuanto a los resultados obtenidos por goteo de genes, en el que se asignan dos alelos distintos a cada fundador, el número equivalente de genomas fundadores N_{ge} descendió de valores iniciales de 10.76, 5.85 y 11.59, en *Guadyerbas*, *Gamito* y *Torbiscal* a valores finales de 2.02, 2.04 y 4.92, reflejando el efecto de la deriva genética (Lacy, 1989). Los respectivos valores del número efectivo de genomas fundadores (Boichard y col., 1997) fueron, en las cohortes iniciales de cada estirpe: 9.71, 5.40 y 6.67, y en las finales: 1.38, 1.40 y 3.03.

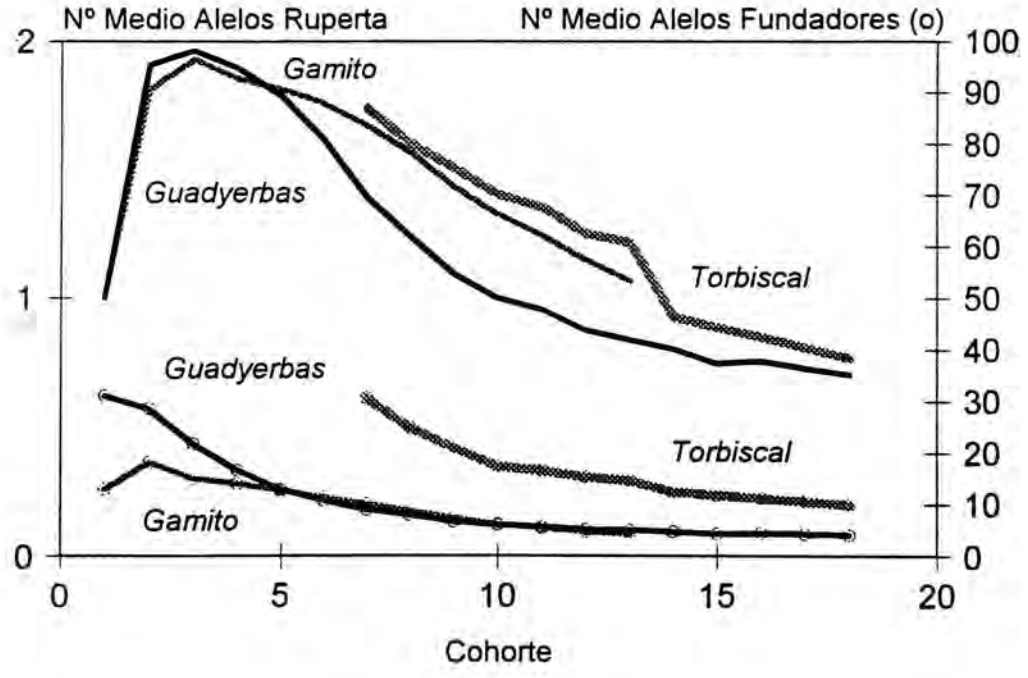


Figura 2. *Número medio de alelos de Ruperta y Número medio de alelos fundadores ($\Sigma 2r_j$) en las estirpes Guadyerbas, Gamito y Torbiscal.*

Los cambios registrados en el número medio de alelos fundadores representados y en el número de alelos con origen en *Ruperta* se describen en la Figura 2. El máximo de algunos de estos parámetros se observa en las cohortes 2ª o 3ª, debido a que hasta ellas no se obtuvieron descendientes reproductores de todos los fundadores. Los valores finales respectivos son, en cuanto a alelos fundadores: 4.04 (d.t. 1.06), 4.21 (d.t. 1.12) y 9.90 (d.t. 1.73) y, en cuanto a alelos procedentes de *Ruperta*: 0.70, 1.07 y 0.77.

El tercero de los aspectos respecto al que podemos examinar el rastro de los fundadores, y de *Ruperta* entre ellos, es el de su contribución a la consanguinidad de las estirpes. En la Tabla 2 se sintetizan los valores de los componentes parciales de la consanguinidad , de acuerdo con las agrupaciones previamente descritas, de las 8890 camadas nacidas en las tres estirpes. La proporción de la consanguinidad media que puede atribuirse a *Ruperta* es 0.193 en *Guadyerbas*, 0.329 en *Gamito* y 0.101 en *Torbiscal*.

Tabla 2. *Componentes parciales de la consanguinidad de las camadas en las tres estirpes: Consanguinidad debida a la fundadora Ruperta (F_c^R) y al resto de los fundadores en Guadyerbas y Gamito (F_c^{NR}) y en Torbiscal (F_c^A, F_c^B, F_c^C y F_c^{NR})*

		Componentes parciales de consanguinidad de las camadas		
Estirpe		Media	d.t.	Mínimo, Máximo
Guadyerbas	F_c^R	0.043	0.021	0.000, 0.125
	F_c^{NR}	0.180	0.087	0.000, 0.424
Gamito	F_c^R	0.097	0.022	0.042, 0.183
	F_c^{NR}	0.198	0.055	0.060, 0.420
Torbiscal	F_c^R	0.011	0.004	0.000, 0.055
	F_c^{NR}	0.029	0.012	0.000, 0.055
	F_c^A	0.027	0.014	0.000, 0.205
	F_c^B	0.011	0.007	0.000, 0.087
	F_c^C	0.025	0.012	0.000, 0.113

En el análisis de la depresión consanguínea del tamaño de camada se obtuvieron resultados relativos a parámetros genéticos de este carácter en las estirpes estudiadas que se resumen en la Tabla 3. Los resultados de *Gamito* concuerdan con los obtenidos por Wang y col. (1994) y los de *Torbiscal* con los obtenidos mediante DFREML en un análisis anterior a partir de una muestra más reducida de madres y camadas (Rodriguez y col., 1994).

Tabla 3. *Media y desviacion tipica posterior de heredabilidad (h^2), coeficiente de ambiente permanente (p^2) y repetibilidad del tamaño de camada en las tres estirpes.*

	Guadyerbas	*Gamito*	*Torbiscal*
Nacidos			
h^2	0.10 (0.04)	0.08 (0.04)	0.07 (0.02)
p^2	0.07 (0.03)	0.04 (0.03)	0.10 (0.02)
Repetibilidad	0.17 (0.03)	0.12 (0.03)	0.17 (0.02)
Nacidos vivos			
h^2	0.06 (0.03)	0.08 (0.04)	0.06 (0.02)
p^2	0.08 (0.03)	0.03 (0.02)	0.08 (0.02)
Repetibilidad	0.13 (0.03)	0.11 (0.03)	0.14 (0.01)

Se realizó un primer análisis de la depresión consanguinea en *Torbiscal* para determinar la importancia relativa del efecto de la consanguinidad de la madre (F_m) y de la camada (F_c) sobre el tamaño de ésta. Los resultados obtenidos (Tabla 4) indican un efecto negativo de la consanguinidad de la camada, con una disminución de -0.13 lechones vivos/camada por cada 10% de incremento de F_c, mientras que el efecto de F_m no es significativo. Un resultado similar se ha obtenido en el análisis de la depresión consanguínea del tamaño de camada realizado en otra piara cerrada (Rodrigáñez y col., 1998). Por ello, en el modelo de análisis de la depresión consanguínea mediante componentes parciales de consanguinidad, se ha contemplado únicamente la consanguinidad de las camadas.

Tabla 4. *Media y desviacion tipica posterior de los efectos de los coeficicientes de consanguinidad de la madre (F_m) y de la camada (F_c) y de los componentes parciales de ésta (F_c^i, i= R, NR) sobre el número de lechones nacidos y nacidos vivos en las tres estirpes.*

Estirpe	Efecto coeficiente consanguinidad	Nacidos por camada	Nacidos vivos por camada
Guadyerbas	F_c^R	-3.70 (5.35)	-1.64 (5.07)
	F_c^{NR}	-1.49 (1.61)	-1.94 (1.48)
Gamito	F_c^R	8.36 (13.20)	3.97 (12.93)
	F_c^{NR}	-4.70 (5.20)	-3.38 (5.09)
Torbiscal	F_m	-0.98 (1.00)	-0.51 (0.92)
	F_c	-0.72 (0.82)	-1.31 (0.80)

Los resultados correspondientes a *Guadyerbas* y *Gamito*, se incluyen asimismo en la Tabla 4. Para el rastreo de la huella de *Ruperta*, la partición realizada de la depresión consanguínea distingue un componente atribuible a esta cerda (F_c^R) y otro al resto de los fundadores (F_c^{NR}). En *Guadyerbas*, las medias posteriores de los coeficientes de regresión de ambos componentes (b_{Fc}^R y b_{Fc}^{NR}) sobre el número total de nacidos y el de nacidos vivos presentan valores negativos, indicativos de depresión consanguínea. Los valores de las desviaciones típicas posteriores reflejan las limitaciones de la información disponible, en particular la escasa dispersión de F_c^R (Tabla 2). El problema se acentúa en *Gamito*, donde son todavía menores tanto el número de animales y de registros como los coeficientes de variación de los componentes parciales de consanguinidad. Aunque los valores positivos de las medias posteriores de b_{Fc}^R para ambos caracteres resultan sugestivos, los amplios valores de las desviaciones típicas posteriores imposibilitan cualquier inferencia sobre estos parámetros. En este punto del análisis, la imagen que podemos obtener de la huella de *Ruperta* aparece muy difuminada. En *Torbiscal*, la huella de *Ruperta* en cuanto a su posible contribución a la depresión consanguínea se pierde por completo. Al reducido valor de F_c^R y de su dispersión en esta estirpe hay que añadir que, agrupado por estirpes el elevado número de fundadores, los cinco coeficientes resultantes F_c^A, F_c^B, F_c^C F_c^R y F_c^{NR} están fuertemente correlacionados lo que hace ilegítimo el análisis de sus efectos.

Las pesquisas en torno al rastro de *Ruperta* nos han proporcionado hallazgos muy dispares, aunque ninguno de gran mérito. Nuestro propósito, más modesto, era ensayar nuevas herramientas de investigación, y en este sentido, creemos que la búsqueda realizada ha sido un ejercicio saludable.

Referencias

Boichard D., Maignel L. y Verrier E. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics, Selection, Evolution* **29**: 5-23.

Gandini G.C., Samore A. y Pagnacco G. 1997. Genetic contribution of the Arabian to the Italian Haflinger horse. *J. Anim. Breed. Genet.* **114**: 457-464.

James J.W. 1972. Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.* **42**: 272-273.

Lacy R.C. 1989. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. *Zoo Biology* **8**: 111-123.

Lacy R.C., Ballou J.D., Princée F., Starfield A. y Thompson E. A. 1995. Pedigree analysis for population management. En *Population management for survival and recovery. Analytical methods and strategies in small population conservation* (ed. J.D. Ballou, M. Gilpin y T.J.R. Foose) pp. 57-75. Columbia University Press, New York.

Lacy R.C., Alaks G. y Walsh A. 1996. Hierarchical analysis of inbreeding depression in Peromyscus polionotus. *Evolution* **50**: 2187-2200.

Lacy R.C. 1997. Errata. *Evolution* **51**: 1025.

MacCluer J.W., Van der Berg J.L., Read B. y Ryder O.A. 1986. Pedigree analysis by computer simulation. *Zoo Biology* **5**: 147-160.

Odriozola M. 1976. *Investigación sobre los datos acumulados en dos piaras experimentales*, IRYDA, Madrid.

Rodrigáñez J., Toro M.A., Silió L. y Rodríguez M.C., 1997a. Supervivencia de alelos en la piara Large White de la Misión Biológica de Galicia. *ITEA* Vol. Extra **18 (I):** 351-353.

Rodrigáñez J., Toro M.A., Silió L. y Rodríguez M.C., 1997b. Fifty years of conservation of a black hairless strain of Iberian pigs. *International Symposium on Mediterranean Animal Germplasm and Animal Challenges*. Benevento, 1995., E.A.A.P. Publication No. **85**, pp. 183-186.

Rodrigáñez J., Toro M.A., Rodríguez M.C. y Silió L. 1998. Founder allele survival and inbreeding depression on litter size in a closed line of Large White pigs. *Animal Science* (aceptado).

Rodriguez M.C., Rodrigáñez J. y Silió L. 1994. Genetic analysis of maternal ability in Iberian pigs. *J. Anim. Breed. Genet.* **111**: 220-227.

Söllkner J. y Filipcic L. 1997. Genetic variability of population and similarity of subpopulations in cattle breeding evaluated by analysis of pedigrees. *48th Annual Meeting of the E.A.A.P.*, Viena.

Wang C.S., Rutledge J.J. y Gianola D. 1994. Bayesian analysis in mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics, Selection, Evolution* **26**: 91-115.

Estima de parámetros genéticos para poblaciones de vacuno lechero mediante metodología bayesiana

Hernández, D.[1]; Carabaño, M.J. [1]; Rekaya, R[2].

[1]Departamento de Mejora y Biotecnología. INIA. Apartado 8111. Madrid 28080.
[2]Departament of Animal Sciences. University of Wisconsin. 1675 Observatory Dr. Madison, Wisconsin 53706-1284. U.S.A.

Introducción

Las estimas de la heredabilidad o repetibilidad para los caracteres de producción lechera se han venido obteniendo con muestras de datos seleccionadas según distintos criterios de acuerdo con la finalidad del estudio y con las limitaciones computacionales ligadas a la utilización de los métodos de máxima verosimilitud.

Las estimas más recientes obtenidas en nuestra población de Frisón para la heredabilidad de la producción de leche y proteína son cercanas a 0.26 (Ibáñez, 1996, Rekaya, 1997) o ligeramente superiores (Charffeddine, 1998). En cuanto a la repetibilidad, las estimas más recientes son cercanas a 0.50 para la producción de leche y algo inferiores para la producción de proteína (Rekaya, 1997).

La tendencia en otros países es la de obtener nuevas estimas de heredabilidad superiores a las encontradas en estudios más antiguos (ver, p.ej., Van Tassell et al., 1997). Dentro de los países que participan en la valoración internacional que lleva a cabo Interbull, dos tercios de éstos informan que las heredabilidades que utilizan en su valoración nacional están por encima de 0.25 y el resto de países (incluido el nuestro) utiliza este valor o valores ligeramente inferiores (Noruega con 0.20 representa el valor mínimo) (Interbull, 1998. Datos no publicados). El valor de la heredabilidad empleado en las valoraciones genéticas tiene cierta importancia en las comparaciones internacionales. Aquellos países con heredabilidades más altas producen valoraciones genéticas con un rango de variación más amplio y con una fiabilidad más alta. Ello hace que se espere que, dentro de la valoración internacional, los toros procedentes de estos países tengan cierta ventaja con respecto a los toros valorados en países donde la heredabilidad es más baja.

La utilización de métodos Monte Carlo ligados a distintos métodos de estimación de componentes de varianza ha permitido en los últimos años el uso de bases de datos de gran tamaño. Dentro de este trabajo se plantea la obtención de estimas de parámetros genéticos para los caracteres kg de leche y proteína basadas en la utilización del método bayesiano que utiliza el muestreo de Gibbs sobre el total de datos disponibles para las valoraciones genéticas de estos caracteres.

Material y Métodos

Se analizó la totalidad de los datos disponibles para la valoración genética de vacuno frisón español correspondientes a producción de leche y de proteína. El número de lactaciones utilizadas en este estudio fue de 1.460.559 para producción de leche y 1.084.886 para la proteína, correspondiendo a 605.902 y 472.494 animales respectivamente. La información genealógica abarcó a 703.326 y 589.385 animales

para cada carácter estudiado. Tanto los datos como la genealogía fueron aportados por CONAFE.

Para la estima de los parámetros de interés, los componentes de varianza, y más concretamente, la heredabilidad y repetibilidad, se empleó un procedimiento Bayesiano. El algoritmo de muestreo de Gibbs fue utilizado para obtener las distribuciones marginales a posteriori de los mencionados parámetros.

El modelo aplicado fue el que actualmente se emplea en la valoración nacional de Frisón (Pena, 1998, comunicación personal), que es un modelo animal con efecto permanente:

$$y_{ijklmn} = \mu + DA_i + M_j + LE_k + RAE_l + A_m + P_m + e_{ijklm}$$

Donde:
- $y_{ijkl\,mn}$ = Producción de leche o proteina expresada en kilogramos.
- μ = Media general
- $\mathbf{DA_i}$ = Días abiertos
- $\mathbf{M_j}$ = Mes de parto
- $\mathbf{LE_k}$ = Lactación-Edad
- $\mathbf{RAE_l}$ = Grupo móvil de comparación Rebaño-Año-Estación.
- $\mathbf{A_m}$ = Efecto aditivo del animal que produce el dato.
- $\mathbf{P_m}$ = Efecto permanente ambiental del animal que produce el dato.
- $\mathbf{e_{ijklmn}}$ = Efecto residual

En notación matricial,

$$\mathbf{y = Xb + Zu + Wp + e}$$

, donde **y** es el vector de observaciones (kg de leche o proteína), **b** contiene los efectos μ, DA, M, LE y RAE , **u** los efectos genéticos aditivos, **p** los efectos ambientales permanentes y **e** los efectos residuales.

La verosimilitud y distribuciones a priori asumidas fueron:

$$\mathbf{y}|\mathbf{b},\mathbf{u},\mathbf{p},\ \sigma_e^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$$
$$\mathbf{b} \sim \text{constante}$$
$$\mathbf{u}|\sigma_a^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$$
$$\mathbf{p}|\sigma_p^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$$

dado que σ_a^2, σ_p^2 y σ_e^2 son desconocidos, se asumieron en un segundo estrato de la jerarquía las siguientes distribuciones a priori para estos parámetros:

$\sigma_a^2|\nu_a, S_a^2 \sim \chi^{-2}(\nu_a, S_a^2)$, donde, $\nu_a = 4$ y $S_a^2 = 250.000$ y 237 para kg de leche y proteína respectivamente.

$\sigma_p^2|\nu_p, S_p^2 \sim \chi^{-2}(\nu_p, S_p^2)$, donde, $\nu_p = 4$ y $S_p^2 = 200.000$ y 207 para kg de leche y proteína respectivamente.

$\sigma_e^2|\nu_e, S_e^2 \sim \chi^{-2}(\nu_e, S_e^2)$, donde, $\nu_e = 4$ y $S_e^2 = 550.000$ y 501 para kg de leche y proteína respectivamente.

Los valores de los parámetros s_a^2, s_p^2, s_e^2 han sido obtenidos a partir de estimas de los correspondientes componentes de varianza en estudios previos sobre esta misma población.

A partir de la verosimilitud y distribuciones a priori de los parámetros desconocidos se obtuvieron las correspondientes distribuciones condicionales a posteriori para llevar a cabo el muestreo de Gibbs.

La implementación del muestreo de Gibbs se hizo mediante el uso de una única cadena larga, excepto en las primeras 10.000 iteraciones. Durante esta fase inicial se utilizaron dos cadenas que partían de distintos valores iniciales pero usaban la misma semilla para generar el muestreo aleatorio, de acuerdo con el método de cadenas pareadas propuesto por Johnson (1996), según describen García Cortés et al. (1998). Mediante este procedimiento se determinó el periodo mínimo de calentamiento como aquel en que la diferencia de las dos cadenas era inferior a 5/10000 partes del valor absoluto de cada uno de los parámetros. Por otro lado, el periodo de calentamiento fue también determinado mediante el procedimiento propuesto por Raftery y Lewis (1992).

Una vez determinado el periodo mínimo de calentamiento se obtuvo el número efectivo de muestras y la autocorrelación entre las mismas a partir de 4000 puntos obtenidos tras el periodo de calentamiento y con un intervalo entre muestras determinado por el procedimiento de Raftery y Lewis (1992). A partir del número de tamaño efectivo más bajo se calculó el número de iteraciones necesarias para obtener un total aproximado de 1000 muestras independientes para cada carácter.

Resultados y Discusión

Las Gráficas 1 a 4 muestran el resultado de aplicar el método de las cadenas pareadas para la heredabilidad y repetibilidad de kg de leche y kg de proteína. De acuerdo con este procedimiento, la convergencia de las dos cadenas se consigue tanto para los mencionados parámetros como para los componentes de varianza en unas 700 iteraciones para el carácter kg de leche y 500 para kg de proteína . Según el procedimiento de Raftery y Lewis (1992), la longitud del periodo de calentamiento debería ser de 16 para ambos caracteres, un valor sensiblemente inferior al obtenido con el método de cadenas pareadas. La subestimación del periodo de calentamiento por el procedimiento de Raftery y Lewis (1992) frente al de cadenas pareadas ha sido encontrada también en el trabajo de García Cortés et al. (1998).

En la Tabla 1 se presenta la autocorrelación entre muestras y el número de muestras efectivas de cada 100 para los componentes de varianza y la heredabilidad y repetibilidad de los caracteres kg de leche y kg de proteína, respectivamente. De acuerdo con estos resultados, la varianza del efecto permanente es el parámetro para el que se obtienen muestras con menor grado de autocorrelación, siendo la autocorrelación menor para el carácter kg de proteína que para el de kg de leche. De acuerdo con los resultados de esta Tabla y asumiendo que la proporción de muestras independientes se mantiene constante, el número total de muestras requerido para obtener 1000 muestras independientes en todos los parámetros sería de 166667 y 52631 para kg de leche y proteína, respectivamente. Sin embargo, el número de puntos totales a tomar de acuerdo con el procedimiento de Raftery y Lewis (1992), resultó ser de 45.096 y 45.068 para leche y proteína respectivamente.

Por otra parte, por motivos computacionales, el número máximo de muestras que fue posible obtener fue de 115.000 para ambos caracteres, que representaría un total de 690 muestras independientes para leche y 2185 para proteína.

Las densidades marginales posteriores para los parámetros de interés fueron estimadas a partir de las últimas 30.000 muestras (10.000 puntos) para producción de leche y 90.000 muestras (30.000 puntos) para kg de proteína, utilizando estimadores tipo kernel no paramétricos (Silverman, 1986). La idea inicial era la de obtener estas densidades marginales con 90.000 muestras para ambos caracteres, pero problemas de estimabilidad en kg de leche hicieron que se realizara sobre las 30.000 últimas.

En las Tablas 2 y 3 se muestran los estadísticos de tendencia central, así como la varianza correspondientes a las distribuciones marginales posteriores de las varianzas (σ_a^2, σ_p^2 y σ_e^2), heredabilidad y repetibilidad de los caracteres en estudio. En general, los valores de las varianzas son inferiores a los obtenidos en estudios previos recientes sobre la población Frisona española (Ibáñez, 1996, Rekaya, 1997, Charffeddine, 1998). Lo mismo sucede con la heredabilidad, lo que parece ser debido especialmente al menor valor obtenido en este estudio para la varianza aditiva. Estos resultados van en contra de la tendencia observada en otras poblaciones en cuanto a la obtención de estimas de heredabilidad por encima de 0.25. En un trabajo reciente sobre la población Holstein de EE.UU., Van Tassell et al., (1997) estiman mediante el método R (Reverter et al., 1994) la heredabilidad de los caracteres kg de leche, grasa y proteína a partir de todos los datos disponibles de vacas en primera lactación nacidas a partir de 1980. Como conclusión del trabajo se recomienda usar en las valoraciones nacionales de Holstein un valor medio de la heredabilidad de 0.30 (variando entre 0.25 y 0.35) frente al valor medio de 0.25 (variando entre 0.20 y 0.30) que se empleaba con anterioridad.

Las razones del bajo valor de la heredabilidad están siendo investigadas. Una posible razón podría ser la incorporación al control lechero en una alta proporción de animales provenientes de zonas en las que se ha observado una heredabilidad más baja (Ibáñez, 1996). Por otra parte, se está planteando la utilización del método R (Reverter et al., 1994), único procedimiento que permitiría, además del ya empleado, la estimación con un volumen de datos tan elevado, para contrastar los resultados obtenidos.

Agradecimientos

Este trabajo ha sido elaborado dentro del proyecto SC96-046 del Programa Sectorial de I+D del MAPA. La contribución de CONAFE en la preparación y cesión de los archivos de datos de producción y genealógicos es agradecida.

Bibliografía

Charffeddine, N. 1998. Selección por mérito económico global en el ganado vacuno Frisón en España. Tesis Doctoral. ETSI Agrónomos. Universidad Politécnica de Madrid.

García-Cortés, L.A.; Rico, M.; Groeneveld, E. 1998. Coupling the Gibbs Sampler to Assess Convergence in Animal Models. Journal Animal Science (enviado).

Ibáñez, M.A. 1996. Heterogeneidad de varianzas en la población Frisona española. Tesis Doctoral. ETSI Agrónomos. Universidad Politécnica de Madrid.

Johnson, V.E. 1996. Studying convergence of Markov Chain Monte Carlo algoritms using coupled sample paths. J. Amer. Statist. Assoc. 91:154-166.

Raftery, A.E.; Lewis, S.M. 1992. How many iterations in the Gibbs sampler. Bayesian Statistic 4, 763-773. Edit. Bernardo, J.M. Oxford University Press.

Rekaya. R. 1997. Análisis bayesiano de datos de producción en los días del control para la selección de caracteres lecheros. Tesis Doctoral. ETSI Agrónomos. Universidad Politécnica de Madrid

Reverter, A., B.L. Golden, R.M. Bourdon. 1994. Method R variance component procedure: application on the simple breeding value model. J. Anim. Sci. 72:2247-2253.

Silverman, B.W. 1986. Density estimation for statistic and data analysis. Chapman and Hall Ed.

Van Tassell, C.P. , G.R. Wiggans, H.D. Norman y R.L. Powell. Estimation of heritability for yield of U.S. dairy cattle. Proc. of the 1997 Interbull Meeting. Vienna, Austria. Bulletin 16:104-107.

Tabla 1.- Autocorrelación (r) y número de muestras efectivas de cada 100 (M.E.) para kg de leche y proteína respectivamente.

Parámetro	Kg de leche		Kg de proteína	
	r	M.E.	r	M.E.
Varianza Aditiva	0.982	0.62	0.947	1.94
Varianza Permanente	0.966	14.88	0.887	30.5
Varianza Residual	0.503	0.68	0.210	2.70
Heredabilidad	0.984	0.59	0.952	1.90
Repetibilidad	0.873	4.46	0.670	18.6

Tabla 2.- Estadísticos de tendencia central y varianza de la distribución marginal posterior para el carácter kg de leche.

Parámetro	Media	Moda	Mediana	Varianza
Varianza Aditiva	202123.0	201551.8	202030.2	5530378
Varianza Permanente	201420.5	201673.5	201455.7	3502246
Varianza Residual	472413.9	472479.9	472415.2	211905
Heredabilidad	0.23074	0.23038	0.23064	6.055E-6
Repetibilidad	0.46069	0.46081	0.46069	1.012 E-6

Tabla 3.- Estadísticos de tendencia central y varianza de la distribución marginal posterior para el carácter kg de proteína.

Parámetro	Media	Moda	Mediana	Varianza
Varianza Aditiva	174.35	173.90	174.32	5.39308
Varianza Permanente	191.47	191.25	191.44	4.57930
Varianza Residual	421.29	421.40	421.29	0.62053
Heredabilidad	0.22149	0.22158	0.22148	1.040E-5
Repetibilidad	0.46476	0.46471	0.46475	2.521E-6

Figura 5.- Densidad marginal posterior de la heredabilidad de kg de leche

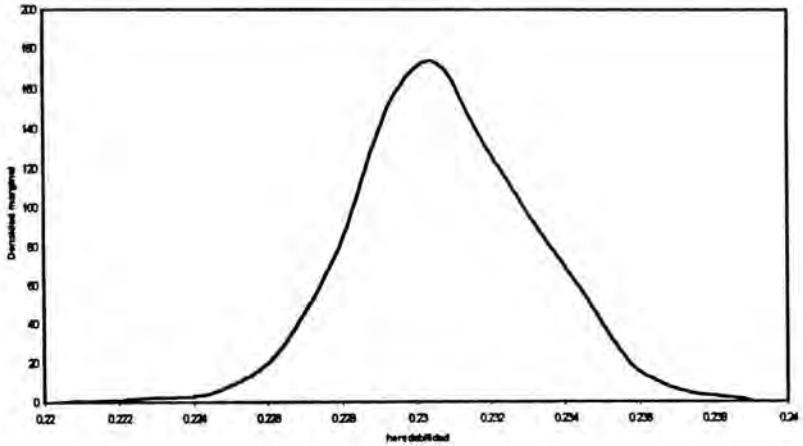


Figura 6.- Densidad marginal posterior de la repetibilidad de kg de leche

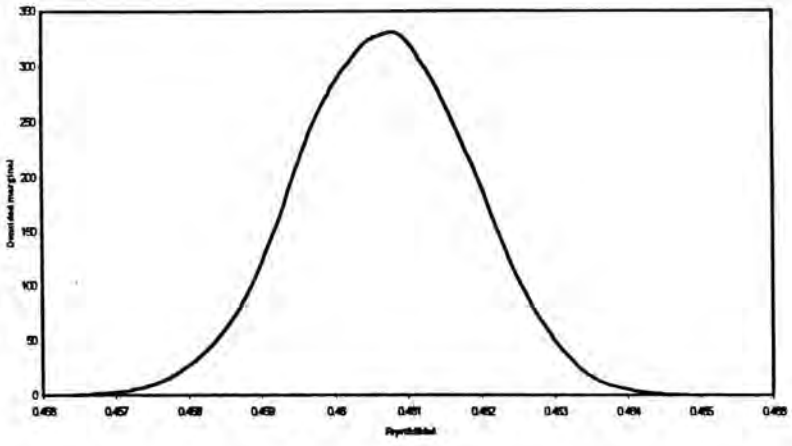


Figura 7.- Densidad marginal posterior de la Heredabilidad de Kg de proteina

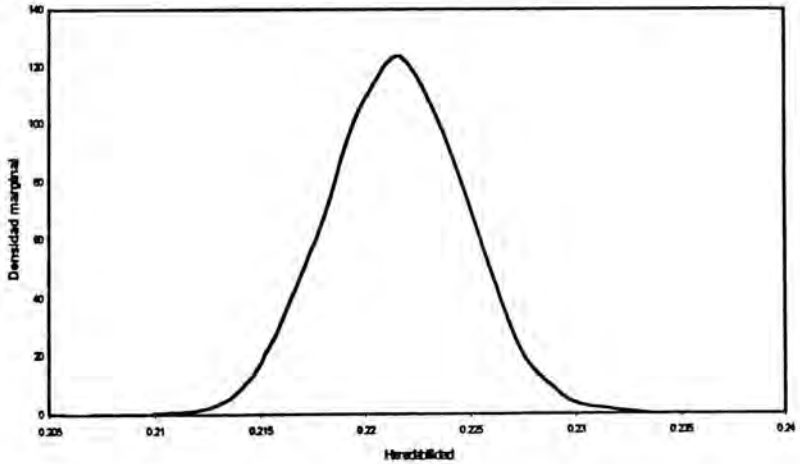


Figura 8.- Densidad marginal posterior de la repetibilidad de kg de proteina

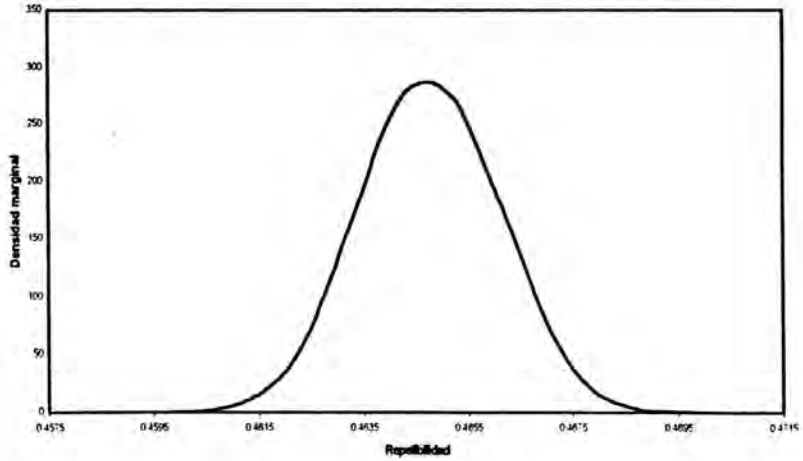


Figura 1.- Valores estimados de la Heredabilidad de kg de leche mediante cadenas pareadas

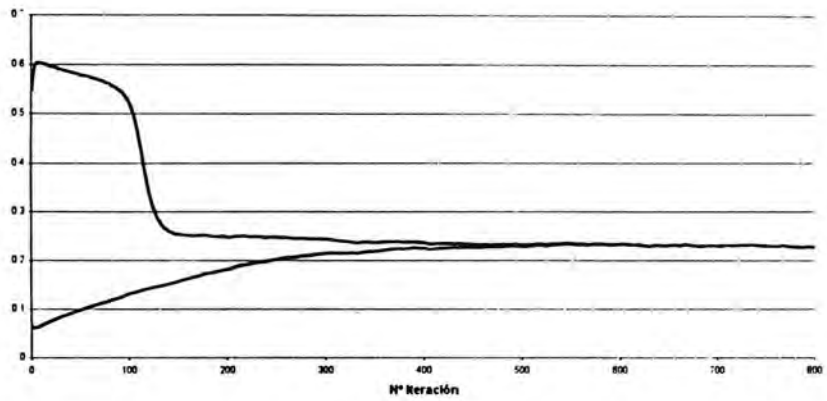


Figura 2.- Valores estimados de la Repetibilidad de kg de leche mediante cadenas pareadas

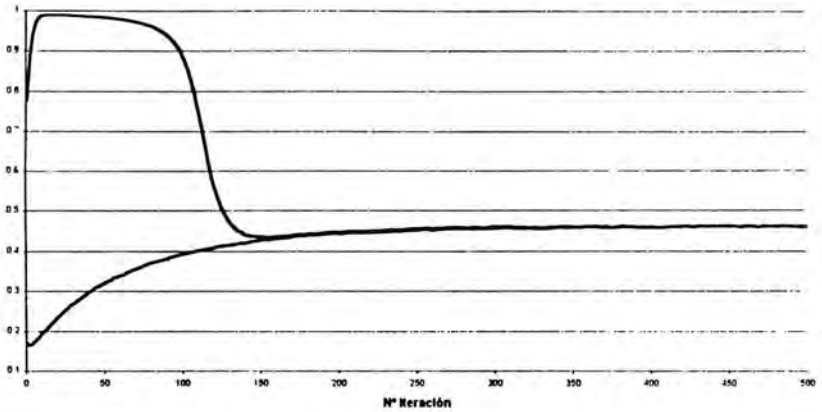


Figura 3.- Valores estimados de la Heredabilidad de kg de proteina mediante cadenas pareadas

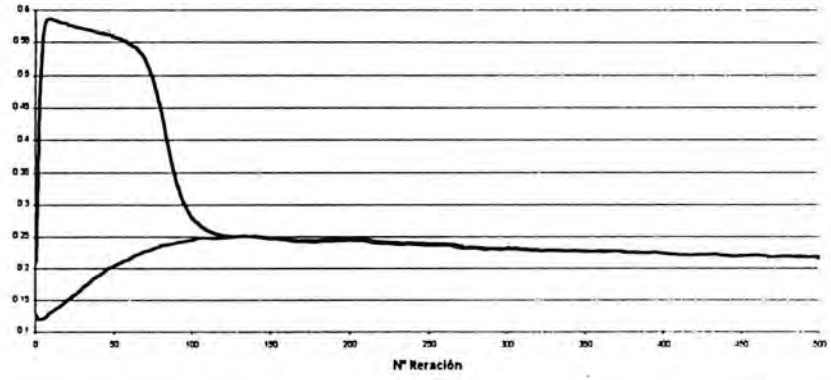
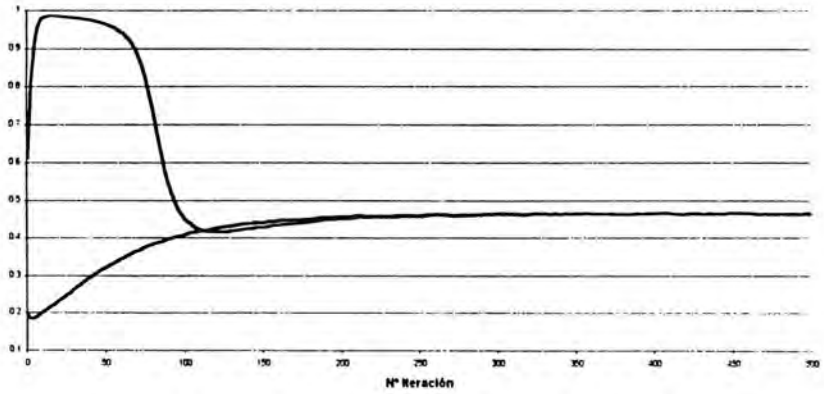


Figura 4.- Valores estimados de la Repetibilidad de kg de proteina mediante cadenas pareadas



AJUSTE Y COMPARACION DE CURVAS DE CRECIMIENTO

A. Blasco
Departamento de Ciencia Animal. Universidad Politécnica de Valencia.
Apartado 22012. Valencia 46071

Luis Varona
Ente UdL-IRTA. Alcalde Rovira Roure, 177. Lleida

EL MODELO

Tomemos la función de Gompertz, probablemente la más usada en la descripción del crecimiento de aves, cerdos y conejos. En el modelo, cada individuo i tiene una curva de crecimiento distinta. Cada individuo dispone de varios datos, tomados cada cierto tiempo j (por ejemplo, cada semana).

$$y_{ij} = a_i \cdot \exp\left[-b_i \cdot \exp\left(-k_i \cdot t_j\right)\right] + \varepsilon_{ij} \qquad (1)$$

Sobre los parámetros a_i, b_i, k_i, actúan efectos ambientales (estación, temperatura, tamaño de la camada en la que nació el individuo, sexo, etc.) que pueden estar correlacionados. Cada individuo tiene un valor genético para cada parámetro, que lógicamente está correlacionado con el valor genético de sus parientes. Si representamos por **a**, **b**, **k**, los vectores con los parámetros de todos los individuos,

$$\mathbf{a} = \mathbf{X}\beta_a + \mathbf{Z}\mathbf{u}_a + \mathbf{e}_a$$

$$\mathbf{b} = \mathbf{X}\beta_b + \mathbf{Z}\mathbf{u}_b + \mathbf{e}_b$$

$$\mathbf{k} = \mathbf{X}\beta_k + \mathbf{Z}\mathbf{u}_k + \mathbf{e}_k$$

Para evitar innecesarias complicaciones, asumimos que las matrices de diseño **X**, **Z** son las mismas (esto es, suponemos que si hay un efecto ambiental como el de estación, éste afecta a los tres parámetros, aunque sea de forma diferente), lo que permite simplificar la notación:

$$\mathbf{p} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

donde

$\mathbf{p}' = [\, \mathbf{a}', \mathbf{b}', \mathbf{k}' \,]$ es el vector de parámetros

$\beta' = [\, \beta'_a, \beta'_b, \beta'_k \,]$ es el vector de efectos ambientales sistemáticos

$\mathbf{u}' = [\, \mathbf{u}'_a, \mathbf{u}'_b, \mathbf{u}'_k \,]$ es el vector de efectos genéticos

$\mathbf{e}' = [\, \mathbf{e}'_a, \mathbf{e}'_b, \mathbf{e}'_k \,]$ es el vector de efectos aleatorios

el vector **e** no incluye los errores de ajuste, que son los e_{ij} de la ecuación (1), sino que pretende reflejar el hecho de que las curvas de crecimiento de los animales pueden estar afectadas por factores aleatorios no considerados en el modelo. Esto es distinto a que un dato concreto tomado sobre un animal en un momento dado se aleje de la curva de crecimiento de este animal debido a factores aleatorios relacionados con el muestreo (por ejemplo, se toma el dato justo cuando acaba de comer el animal).

Desde un punto de vista frecuentista se considera **b** como un conjunto de efectos fijos, pero desde un punto de vista bayesiano se debe intentar encontrar una función que represente la opinión que se tiene a priori en función de la información previa disponible. Una forma sencilla de hacer esto es suponer que

$\mathbf{b} \sim N(\mathbf{m_b}, \mathbf{V_b})$

donde $\mathbf{m_b}$, $\mathbf{V_b}$ son subjetivas e intentan dar lugar a una función Normal que represente el estado de creencias previas. Dada la dificultad de hacerlo en el caso multivariante, habitualmente $\mathbf{V_b}$ será una matriz diagonal.

Se considera que $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, donde

$$\mathbf{G} = \begin{bmatrix} \mathbf{G_a} & \mathbf{G_{ab}} & \mathbf{G_{ak}} \\ & \mathbf{G_b} & \mathbf{G_{bk}} \\ & & \mathbf{G_k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma^2_{Ga} & \mathbf{A}\sigma_{Gab} & \mathbf{A}\sigma_{Gak} \\ & \mathbf{A}\sigma^2_{Gb} & \mathbf{A}\sigma_{Gbk} \\ & & \mathbf{A}\sigma^2_{Gk} \end{bmatrix} = \mathbf{A} \otimes \mathbf{G_p}$$

donde **A** es la matriz de parentesco de los individuos, y $\mathbf{G_p}$ la matriz de varianzas-covarianzas genéticas de los parámetros a, b, k

$$\mathbf{G_p} = \begin{bmatrix} \sigma^2_{Ga} & \sigma_{Gab} & \sigma_{Gak} \\ & \sigma^2_{Gb} & \sigma_{Gbk} \\ & & \sigma^2_{Gk} \end{bmatrix}$$

Se considera que $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, donde

$$\mathbf{R} = \begin{bmatrix} \mathbf{R_a} & \mathbf{R_{ab}} & \mathbf{R_{ak}} \\ & \mathbf{R_b} & \mathbf{R_{bk}} \\ & & \mathbf{R_k} \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma^2_{Ra} & \mathbf{I}\sigma_{Rab} & \mathbf{I}\sigma_{Rak} \\ & \mathbf{I}\sigma^2_{Rb} & \mathbf{I}\sigma_{Rbk} \\ & & \mathbf{I}\sigma^2_{Rk} \end{bmatrix} = \mathbf{I} \otimes \mathbf{R_p}$$

donde **I** es la matriz de identidad, y $\mathbf{R_p}$ la matriz de varianzas-covarianzas de los efectos aleatorios no sistemáticos de los parámetros a, b, k.

$$\mathbf{R_p} = \begin{bmatrix} \sigma^2_{RGa} & \sigma_{Rab} & \sigma_{Rak} \\ & \sigma^2_{Rb} & \sigma_{Rbk} \\ & & \sigma^2_{Rk} \end{bmatrix}$$

Esto es, se considera que los efectos aleatorios no sistemáticos que actúan sobre a, b, k, pueden estar relacionados entre sí para cada individuo, pero que no hay correlaciones entre efectos de distintos individuos; es decir, que para un individuo i

$r(e_{ai}, e_{bi}) \neq 0$; $r(e_{ai}, e_{ki}) \neq 0$; $r(e_{ki}, e_{bi}) \neq 0$

mientras que para dos individuos i, j

$r(e_{ai}, e_{aj}) = 0$; $r(e_{ai}, e_{bj}) = 0$; etc.

Es razonable suponer que habrá una cierta correlación entre parámetros, por ejemplo, si uno representa la tasa de crecimiento (k) y otro el peso adulto (A) es difícil suponer que no estén relacionados. Si se quiere suponer que hay correlaciones entre individuos, es menester incluir un nuevo efecto aleatorio que las considere, aunque no hay motivos para pensar que deban existir.

Por tanto, se considera que

$$(\mathbf{p} \mid \mathbf{b}) \sim N(\mathbf{Xb}, \mathbf{Z'GZ} + \mathbf{R})$$

$$(\mathbf{p} \mid \mathbf{b},\mathbf{u}) \sim N(\mathbf{Xb} + \mathbf{Zu}, \mathbf{R}) \qquad (2)$$

Se considera, finalmente, que los errores de estimación e_{ij} se distribuyen de forma Normal con media cero y varianza que va aumentando con el tiempo. La razón es que la varianza está ligada a la media del carácter: cuando los animales son jóvenes, todos los datos que se recogen se agrupan en torno a valores pequeños, y conforme los animales crecen la dispersión aumenta por un efecto de escala. Al llegar al estado adulto, las medidas sucesivas que se toman tienen la misma dispersión. Si hemos considerado que la varianza aumenta con la media y ésta última lo hace con arreglo a una curva de crecimiento, podemos considerar que la varianza o que la desviación típica también aumentan con arreglo a una curva de crecimiento. De hecho el modelo es muy flexible, la mayor o menor linearidad de la pendiente que describe el crecimiento de la varianza estará determinada por los parámetros de la curva. Por tanto,

$\varepsilon \sim N(\mathbf{0}, \mathbf{V})$, donde

$$\mathbf{V} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & & & . \\ . & & . & & . \\ . & & & . & . \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \mathbf{D} \end{bmatrix} \qquad (3)$$

donde **D** es una matriz diagonal correspondiente a los datos de un individuo

$$\mathbf{D} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & . \\ . & & . & . \\ 0 & \dots & \dots & \sigma_T^2 \end{bmatrix}$$

donde

$$\sigma_j = a_\varepsilon \cdot \exp\left[-b_\varepsilon \cdot \exp\left(-k_\varepsilon \cdot t_j\right)\right]$$

Es decir, se considera que los errores de estimación no están relacionados ni entre individuos ni entre dos medidas consecutivas de un individuo. Aunque es obvio que hay autocorrelación entre dos medidas sucesivas de un individuo, ésta ha sido tenida en cuenta al incluir efectos ambientales y genéticos en los parámetros, por lo que los errores no deben estar autocorrelacionados. Si hubiera otras fuentes de autocorrelación, deberían ser tenidas en cuenta en el modelo de los parámetros a, b, k. Si no es posible, porque se considera que hay efectos ambientales similares entre dos medidas sucesivas que pueden no estar recogidos en el modelo y que afectan al error de estimación, puede modificarse **D** para incluir covarianzas entre medidas sucesivas. Cómo hacerlo es un tanto arbitrario, pero suele suponerse que dos medidas sucesivas están correlacionadas con una correlación r, por lo que la primera y la tercera medida lo está con una correlación r^2, la primera y la cuarta con r^3, etc. Si las medidas no son tomadas de forma regular en el tiempo, hay que tenerlo en cuenta al definir la estructura de los errores.

Se considera también que las desviaciones típicas del error aumentan con el tiempo con arreglo a una curva de Gompertz de parámetros a_ε, b_ε, k_ε, y que son las mismas para todos los individuos. Para simplificar la notación, llamaremos

$\mathbf{p'}_\varepsilon = [a_\varepsilon, b_\varepsilon, k_\varepsilon]$

no consideraremos que sobre estos parámetros ejerzan su influencia factores genéticos ni ambientales sistemáticos.

Se desea estimar:

1) Los parámetros a_i, b_i, k_i, de cada individuo.
2) Los parámetros a, b, k, medios de la población.
3) Los parámetros genéticos de a, b, k (esto es, sus heredabilidades y correlaciones genéticas).
4) La diferencia entre parámetros a_i, b_i, k_i de dos individuos o entre los parámetros a, b, k medios de dos poblaciones.
5) Los efectos ambientales que se ejercen sobre los parámetros a, b, k.
6) Los parámetros a_ε, b_ε, k_ε que describen la evolución de la desviación típica del error de ajuste con el tiempo.

PROBLEMAS DE ESTIMACIÓN

1) ***No existe una representación lineal de la curva***

Aunque se tomen logaritmos para hacer desaparecer un exponente, queda siempre una parte exponencial. Esto significa que o bien debe ajustarse mediante regresión no lineal o bien debe utilizarse alguna aproximación lineal, normalmente basada en una serie de Taylor. En cualquier caso, los ajustes no lineales se resuelven también por aproximaciones lineales usando habitualmente series de Taylor.

2) ***En al caso de linealizar tomando logaritmos, el tratamiento de error de ajuste obliga a hipótesis violentas***

Pongamos un ejemplo más simple, ajustar la curva exponencial

$$y = a \cdot \exp(-b \cdot t) + \varepsilon$$

pero esto impide linealizar la ecuación tomando logaritmos, por lo que se suele proponer el modelo

$$y = a \cdot \exp(-b \cdot t) \cdot \exp(\varepsilon) = a \cdot \exp(-b \cdot t + \varepsilon)$$

que es fácilmente linealizable y se ajusta sin problemas por regresión simple

$$\ln y = \ln a - b \cdot t + e$$

pero que obliga a considerar que los errores son multiplicativos y exponenciales.

3) ***Se desconocen las distribuciones de las estimas de los parámetros. No es posible comparar curvas de crecimiento***

Se sabe que los estimadores de los coeficientes de regresión se distribuyen como t-Student, pero en una regresión no lineal no se sabe qué distribución tienen los estimadores de los coeficientes a,b,k de la curva de crecimiento, lo que tiene la fatal consecuencia de *impedir la comparación de curvas*. Si se hacen transformaciones nos encontramos con el problema de que al deshacerlas, el antilogaritmo del error estándar de un coeficiente no es el error estándar del coeficiente. Por ejemplo, en la ecuación de Brody, una vez estimados b y [ln a], se desea saber el valor de a y para ello se toma el antilogaritmo de [ln a], pero el error estándar de a no se obtiene hallando el antilogaritmo del error estándar de [ln a].

4) ***La corrección por efectos sistemáticos es oscura o imposible***

Se supone que el hecho de nacer en una camada numerosa, o de pertenecer a uno u otro sexo, afectan al peso adulto y a los parámetros b y k de forma distinta. Cómo hacer estas correcciones es oscuro, y habitualmente se ha preferido trabajar con datos precorregidos que no han tenido en cuenta que el efecto sistemático puede afectar de forma diferente a uno u otro parámetro de la curva, aparte de no considerar los errores cometidos en la precorrección en el resultado final.

5) ***No se tiene en cuenta el parentesco, lo que altera la estructura de los errores***

Por métodos clásicos no está claro cómo incluir el efecto del parentesco en el ajuste a cada parámetro. Los modelos de regresión aleatoria, en la que cada parámetro es (o lleva asociada) una variable aleatoria pretenden resolver parcialmente el problema, pero se encuentran la dificultad de que en ese caso los parámetros cambian de valor en función de la información disponible (i.e.: si hay pocos datos su valor se aproxima a cero). Es frecuente que en el ajuste de curvas de crecimiento no todos los individuos tengan datos hasta el final de la toma de medidas, por lo que hay parámetros mejor estimados que otros. Esto no es un problema para la estimación de una curva media pero sí para la estimación de curvas individuales.

6) ***No se tiene en cuenta la estructura de los errores debida a causas no genéticas***

No se suele tener en cuenta, por ejemplo, la autocorrelación ocasionada por medir a un individuo en periodos de tiempo consecutivos o el hecho de que la varianza del error aumenta con la edad hasta estabilizarse al llegar al estado adulto. Para abordar este último problema se puede utilizar una regresión dando un peso proporcional a al inversa de la varianza en cada momento de medida, pero no se incluye en el modelo el error cometido al estimar los pesos.

7) ***La estimación de los parámetros genéticos de los coeficientes se realiza por métodos que no optimizan la información***

La mayor parte de autores estiman las curvas individuales y realizan un análisis genético de las estimaciones, con lo que pierden la información debida al parentesco y no incluyen el error de estimación de estos parámetros a, b, k en el error de los parámetros genéticos.

SOLUCIÓN BAYESIANA

1. La solución analítica

1.1) ***El problema***

Desde un punto de vista bayesiano, el problema consiste en encontrar la función de densidad posterior de los parámetros que se desea estimar, dados los datos. Esto es, encontrar

$$f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e \mid \mathbf{y})$$

Aplicando el teorema de Bayes,

$$f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e \mid \mathbf{y}) = f(\mathbf{y} \mid \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e) \cdot f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e) / f(\mathbf{y})$$

1.2) ***Cálculo de la verosimilitud***

En el modelo (ecuación 1), si se fija $\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e$ lo único que queda variable es el error de estimación e_{ij}, por lo que

$$\mathbf{y} \sim N(\mathbf{m}, \mathbf{V})$$

donde **V** es la matriz de varianzas covarianzas de los errores, expuesta en (2), y en **m** se encuentran los valores m_{ij} que se obtendrían para cada momento de medida t_j en la ecuación (1) (sin el error e_{ij}) al sustituir los valores de los parámetros a_i, b_i, k_i de cada animal, puesto que estos parámetros están fijados.

Como los errores e_{ij} son independientes, y como la densidad conjunta de variables independientes es el producto de sus densidades, la expresión que queda es bastante simple:

$$f(\mathbf{y} \mid \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e) = \prod_i \prod_j f(y_{ij} | a_i, b_i, k_i, \sigma^2_{eij}) = \prod_i \prod_j \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left[\frac{(y_{ij} - m_{ij})^2}{\sigma^2_{ij}}\right]$$

1.3) ***Cálculo de las funciones de densidad a priori***

$f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e) = f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R})\, f(\mathbf{p}_e)$

ya que $\mathbf{p}_e$ no depende del resto de parámetros. Utilizando reglas de probabilidad,

$P(A,B) = P(A|B)\, P(B)$

En nuestro caso,

$f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}) = f(\mathbf{p} \mid \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R})\, f(\mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R})$

$f(\mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}) = f(\mathbf{b}, \mathbf{u} \mid \mathbf{G}, \mathbf{R})\, f(\mathbf{G}, \mathbf{R}) = f(\mathbf{b}, \mathbf{u} \mid \mathbf{G}, \mathbf{R})\, f(\mathbf{G})\, f(\mathbf{R})$

ya que las varianzas genéticas no están relacionadas con las ambientales. Es posible considerar dependencia entre ambos componentes de varianza en la distribucion a priori (Weiss et al., 1997), aunque un modelo así nunca ha sido utilizado en mejora genética animal. En principio no hay razones por las que los dos tipos de componentes de varianza deban estar relacionados.

Supondremos también que los efectos sistemáticos son independientes de los aleatorios. En ocasiones se sabe con certeza que esto no es así (por ejemplo, cuando se evalúan toros lecheros se sabe que las mejores granjas, las que mejor ambiente procuran a las vacas, son las que más invierten en genética y traen semen de mejores animales), pero aquí no tenemos motivos para suponer que no puedan ser independientes, salvo errores en el diseño del experimento. Así pues,

$f(\mathbf{b}, \mathbf{u} \mid \mathbf{G}, \mathbf{R}) = f(\mathbf{b})\, f(\mathbf{u} \mid \mathbf{G})$

Con lo que finalmente queda

$f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{p}_e) = f(\mathbf{p} \mid \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R})\, f(\mathbf{b})\, f(\mathbf{u} \mid \mathbf{G})\, f(\mathbf{G})\, f(\mathbf{R})\, f(\mathbf{p}_e)$

donde

$f(\mathbf{p} \mid \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}) \sim N(\mathbf{Xb} + \mathbf{Zu}, \mathbf{R}) = N(\mathbf{Xb} + \mathbf{Zu}, \mathbf{I} \otimes \mathbf{R}_p)$

$f(\mathbf{b}) \sim N(\mathbf{m}_b, \mathbf{V}_b)$ según discutimos al hablar del modelo

$f(\mathbf{u} \mid \mathbf{G}) \sim N(\mathbf{0}, \mathbf{G}) = N(\mathbf{0}, \mathbf{I} \otimes \mathbf{G}_p)$

$f(\mathbf{G})$ y $f(\mathbf{R})$ deberían construirse en función de las creencias previas proporcionadas por la información previa disponible. Como no es posible establecer estrictamente estas creencias, suele recurrirse a alguna función que con pocos parámetros pueda cambiar de forma adaptándose al estado más o menos vago de creencias previas. Por conveniencias matemáticas suele usarse un función conjugada, en este caso una distribucion Whishart invertida, aunque hay muchas otras soluciones posibles. Esta función depende de dos

parámetros, una matriz de escala y otro parámetro al que equivocadamente se le denomina "grados de libertad" o "grados de credibilidad"; ambos parámetros modifican la forma de la función.

Es prácticamente obligado renunciar a describir el estado de creencias previo multivariante puesto que es difícil de definir, ya que habría que describir la opinión previa al experimento sobre cada conjunto de valores posible; esto es, en cada punto del espacio (ver Blasco (1997) para una discusión sobre este punto). Se puede, sin embargo, definir el estado de creencias para cada variable por separado utilizando una matriz de escala diagonal, puesto que en ese caso se obtienen chi-cuadrados invertidas que son fácilmente representables. En cualquier caso se debe intentar realizar los análisis variando los valores de estos parámetros para examinar hasta qué punto la opinión previa es importante en el resultado final.

$f(p_e) = c$, un vector de constantes dentro de límites admisibles del espacio paramétrico para asegurar la propiedad de la distribución posterior conjunta. Es decir, los posibles valores de a_e tienen todos la misma probabilidad a priori, con ciertos límites para evitar que la función sea impropia, y lo mismo podemos decir de b_e y de k_e. Con esto pretendemos reflejar un estado de incertidumbre sobre los valores de estos parámetros.

1.4) ***Cálculo de*** **f(y)**

El cálculo de f(y) no es estrictamente necesario cuando lo que se desea es hallar la moda (el valor más probable) de la densidad posterior, puesto que como la densidad posterior no depende de y (está condicionada a y), 1/f(y) es una constante de proporcionalidad que se puede ignorar al buscar el máximo de la densidad posterior. Sin embargo si se desea hallar la media o usar el sistema de intervalos de confianza para la inferencia científica, hay que conocer exactamente la densidad posterior, por lo que hay que conocer el valor de f(y). Este valor es difícil de computar porque

$$f(y) = \int f(y, p, b, u, G, R, p_e)\, f(p, b, u, G, R, p_e)\, d(p, b, u, G, R, p_e)$$

como y es un vector, esta integral es multidimensional. Aunque los elementos de y fueran independientes, esta integral pasaría a ser una integral múltiple de tantas dimensiones como datos, y en ambos casos el problema es irresoluble incluso por métodos numéricos. Si sólo se está interesado en las distribuciones marginales, la constante de integración es unidimensional y el problema es resoluble mediante integración numérica. De todas formas, hoy en día las técnicas de muestreo de Gibbs, que se mencionan en el apartado siguiente, han resuelto ambos problemas.

2. Modus operandi

Finalmente tenemos que

$$f(p, b, u, G, R, p_e \mid y) =$$

$$= f(y \mid p, b, u, G, R, p_e)\, f(p \mid b, u, G, R)\, f(b)\, f(u \mid G)\, f(G)\, f(R)\, f(p_e) \;/\; f(y) \qquad (4)$$

que es un producto de funciones Normales o Whishart invertida. Se puede intentar hallar el máximo de esa función, lo que dará la solución más probable, o hallar la media, lo que dará la solución que minimiza el riesgo cuadrático. En el primer caso nos encontramos con un problema de la misma envergadura que el de hallar soluciones máximo verosímiles: el número de parámetros y la complejidad de las funciones hacen de ésta una tarea muy compleja. En el segundo caso habría que integrar la función y hallar la constante de proporcionalidad, lo que tampoco parece viable.

Una solución propuesta recientemente es la de extraer muestras al azar de la función de densidad posterior para representar aproximadamente esta función de densidad. Las inferencias se hacen a partir de los puntos muestreados de la densidad posterior multivariante. Creando histogramas o dibujando las función de densidad marginales a partir de esos puntos, se puede obtener una estima de la moda; la media de esos valores es una estima

de la media de la densidad posterior; y finalmente, ordenándolos, se puede obtener la mediana con facilidad, así como cualquier intervalo de confianza.

El problema se centra ahora en cómo obtener esos valores *tomados al azar* de la función de densidad posterior. En el caso multivariante no es posible computacionalmente -al menos de momento- muestrear directamente de la función de densidad posterior, y hay que transformar el problema en univariante, o en problemas de menos dimensiones, mediante técnicas de muestreo de Gibbs. Las técnicas de muestreo de Gibbs, basadas en el muestreo de las funciones condicionales de la densidad posterior, permiten, además, eludir el cálculo de f(y) (ver Sorensen (1997) para una amplia revisión). Se trata, pues, de extraer muestras al azar de las funciones

$f(\mathbf{p}|\ \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$, $f(\mathbf{b}|\ \mathbf{p}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$, $f(\mathbf{u}|\ \mathbf{b}, \mathbf{p}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$,

$f(\mathbf{G}|\ \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{R}, p_e, \mathbf{y})$, $f(\mathbf{R}|\ \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, p_e, \mathbf{y})$, $f(p_e|\ \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{y})$

Para iniciar el proceso se toman valores arbitrarios de $\mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e$, con ellos:

1) se muestrea al azar un valor de **p** en la primera función $f(\mathbf{p}|\ \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$,

2) con ese valor y los $\mathbf{u}, \mathbf{G}, \mathbf{R}, p_e$ arbitrarios de antes se muestrea al azar un valor de **b** en la siguiente función $f(\mathbf{b}|\ \mathbf{p}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$.

3) con esos dos valores de **p** y **b** se muestrea un valor al azar de **u** en la tercera función $f(\mathbf{u}|\ \mathbf{b}, \mathbf{p}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$.

y así sucesivamente hasta que se tienen valores al azar de $\mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e$, momento en el que se reinicia el ciclo muestreando un valor al azar de **p** en la primera función. Al cabo de varios ciclos, los valores muestreados pertenecen a la densidad posterior $f(\mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e | \mathbf{y})$.

Para poder aplicar estas técnicas es menester saber cómo muestrear al azar de las funciones condicionales. Es sencillo escribirlas, puesto que de (4) se trata de tomar como constante todo aquello que está condicionado. Por ejemplo, para el caso de $f(\mathbf{p}|\ \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$ se escribiría de forma explícita (4), se tomaría como constantes $\mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e$ y como variable **p**, y se intentaría ver si esa función es de alguna forma conocida (Normal, Wishart u otra) de la que hayan algoritmos para extraer valores al azar. Haciendo esto se obtiene que

$f(\mathbf{b}|\ \mathbf{p}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$, $f(\mathbf{u}|\ \mathbf{b}, \mathbf{p}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$ son Normales de parámetros conocidos,

$f(\mathbf{G}|\ \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{R}, p_e, \mathbf{y})$ y $f(\mathbf{R}|\ \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, p_e, \mathbf{y})$ son Wishart invertidas,

$f(p_e|\ \mathbf{p}, \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \mathbf{y})$ y $f(\mathbf{p}|\ \mathbf{b}, \mathbf{u}, \mathbf{G}, \mathbf{R}, p_e, \mathbf{y})$ no pertenecen a ninguna familia conocida, por lo que para extraer valores al azar de ellas es necesario utilizar técnicas de aceptacion-rechazo (Ripley, 1987) o introducir muestreos mediante un algoritmo Metropolis-Hastings de estas variables (ver Tanner,1993, para una revisión).

Supongamos que se quieren comparar las curvas de crecimiento medias de dos grupos de animales. Para ello basta con realizar el análisis de los datos de los dos grupos simultáneamente introduciendo como efecto fijo el efecto de grupo, con lo que tendremos dos niveles para cada parámetro a, b, k. Luego se calcula en cada iteración del proceso Gibbs las diferencias entre los dos niveles de este efecto fijo, y con ello obtenemos puntos de la función de densidad posterior de las diferencias entre grupos para a, b, k. A partir de ahí se actúa como antes: se calcula un estimador o los intervalos de confianza para esa diferencia. También se pueden calcular las medias de los valores aditivos de cada parámetro, y en cada iteración hallar la media de los de cada grupo y restar esas dos medias. Con ello se obtendrían puntos de la función de densidad posterior de las diferencias genéticas entre grupos para a, b, k.

Modelización y estima de componentes de varianza para la producción de leche y proteína en el día de control en la raza ovina Manchega.

Serrano, M.*; Jurado, J.J.*, Pérez-Guzmán, M.D.** y Montoro, V.**
*Departamento de Mejora Genética y Biotecnología. INIA. Apdo. 8111, 28080 Madrid.
**CERSYRA. Avda. del Vino 2, 13300 Valdepeñas. Ciudad Real.

Introducción.

Es norma generalizada que la producción de leche de las ovejas en una lactación se estime de forma aproximada mediante el método del día centrado o Fleischman, tipificando la lactación de los animales a una longitud prefijada de la misma. La bondad de dicho método depende principalmente de la calidad del control lechero en cuanto al aspecto temporal de la recogida de las muestras o controles. Es muy frecuente, dado el tipo de manejo de los animales de esta raza, que los controles productivos se realicen con intervalos de tiempo muy variables lo cual tiene como consecuencia la desincronización entre el número cronológico del control y el estado de lactación del animal controlado. Por ello se plantea la necesidad de buscar métodos alternativos que permitan superar estas dificultades.

En los últimos años, en ganado vacuno de leche se han realizado estudios sobre un carácter alternativo a la producción estandarizada de leche, que es la producción diaria de leche o producción en el control, que presenta la ventaja de ser un carácter que depende de una sola medida y no del conjunto de varias y que permite la eliminación de controles anómalos, sin por ello tener que desechar una lactación completa.

Siguiendo por tanto las experiencias realizadas en vacuno de leche, en este trabajo se han llevado a cabo los primeros estudios encaminados a la definición de un modelo de valoración genética para la producción de leche y porcentaje de proteína en el día de control y estimas de parámetros genéticos para dichos caracteres, en ovejas de raza Manchega.

Material y Métodos.

El primero de los estudios tuvo como fin la definición del modelo de efectos fijos más adecuado para el tratamiento de la producción leche y porcentaje de proteína en el día de control. Este estudio se basó en aquellos ya realizados en vacuno de leche (Strabel y Szwaczkowski, 1995; Rekaya, 1997) y algunos existentes en ovino de leche (Baro, 1994; Barrillet, y Boichard, 1994).

Posteriormente se realizaron estimas de componentes de varianza de estos caracteres. Existen referencias en vacuno de leche donde los controles de una misma lactación son tratados como medidas repetidas del mismo carácter, y las distintas lactaciones como caracteres distintos (Ptak y Schaeffer, 1993 y Reents y col. 1995). Por otro lado, escasos estudios abordan modelos multicarácter de los controles dentro de lactación (Rekaya, 1997), debido a las grandes exigencias computacionales que requieren. Sin embargo dado que en ovino de leche el número de controles es más reducido (4 controles en ovino frente a 10 en vacuno) se decidió comenzar el estudio realizando estimas de componentes de varianza con modelos multicarácter intralactación, considerando cada control productivo como un carácter independiente.

Para el estudio se han empleado los datos productivos del control lechero oficial de la raza ovina Manchega acumulados hasta Noviembre de 1997 y los registros genealógicos correspondientes a dichos datos. Esto supone 106.515 lactaciones pertenecientes a 67.250 ovejas, y un registro genealógico de 75.759 animales. Del total de los datos se seleccionaron aquellos que cumplían los siguientes requisitos:

- Intervalo parto-primer control entre 30 – 72 días después del parto.

- Intervalo entre controles 26 –34 días.
- Producción de leche en el control mayor o igual a 200 ml. y menor o igual a 3074 ml.
- Porcentaje de proteína distinto de cero en cada control.
- Cuatro controles registrados como mínimo en la lactación.
- Tener al menos registro genealógico de madre.
- Tener número de lactación y número de crías al parto conocido.

Con estas condiciones se obtuvo una base de datos de 22.954 registros productivos de 7.996 ovejas de 1ª lactación, 7.882 ovejas de 2ª lactación y 7.076 ovejas de 3ª o más lactaciones. En la tabla 1 se presenta un resumen de las producciones medias para ml. de leche y % de proteína de cada uno de los controles productivos correspondientes a dichos datos.

Los modelos de efectos fijos para definir la producción en el día de control para leche y % de proteína fueron analizados mediante el procedimiento GLM del paquete estadístico SAS (SAS, 1990). Estos modelos fueron estudiados para cada lactación, 1ª, 2ª y 3ª o más, y control productivo, 1°, 2° 3° y 4°.

TABLA 1: Medias y desviaciones típicas para los días en lactación, producción de leche en ml. y % de proteína registrados en el día de control, para cada uno de los controles productivos, en el total de los datos y en 1ª, 2ª y 3ª o más lactaciones.

	CONTROL	TOTAL LACTACIONES N = 22.954	1ª LACTACIÓN N = 7.996	2ª LACTACIÓN N = 7.882	3ª o más LACTACIÓNES N = 7.076
Días en Lactac.	1° C.	47.29±9.49	47.33±9.25	47.20±9.60	47.34±9.68
	2° C.	77.19±9.78	77.23±9.57	77.16±9.83	77.18±10.04
	3° C.	107.05±10.07	107.17±9.86	107.05±10.05	106.88±10.40
	4° C.	136.72±10.43	136.95±10.15	136.67±10.42	136.45±10.83
LECHE (ml.)	1° C.	1327.12±544.53.	1225.02±517.66	1370.63±558.47	1414.75±538.98
	2° C.	1108.91±482.69	1113.37±488.37	1117.94±493.64	1089.15±456.86
	3° C.	875.96±409.67	908.84±431.51	866.53±407.64	841.04±374.38
	4° C.	667.96±341.38	691.41±356.43	667.09±343.13	634.49±312.00
PROTE. (%)	1° C.	5.33±0.67	5.21±0.64	5.37±0.66	5.43±0.69
	2° C.	5.65±0.68	5.49±0.65	5.71±0.66	5.79±0.70
	3° C.	6.02±0.76	5.86±0.74	6.08±0.75	6.16±0.78
	4° C.	6.31±0.89	6.22±0.86	6.33±0.88	6.41±0.92

Los modelos analizados fueron los siguientes:

Modelo 1: $y = ncr + edp + rba + est + del + \varepsilon$

Modelo 2: $y = ncr + edp + rba + est + del/cov + \varepsilon$

Modelo 3: $y = ncr + edp/cov + rba + est + del + \varepsilon$

Modelo 4: $y = ncr + edp + rdc + + del + \varepsilon$

Modelo 5: $y = ncr + \quad edp + rdc + \quad + del/cov + \varepsilon$

Modelo 6: $y = ncr + edp/cov + rdc + \quad + del \quad + \varepsilon$

Donde:

y = registro de producción (ml. de leche o % de proteína) en el control.
ncr = número de crías al parto (3 niveles).
edp = edad al parto en meses como variable categórica (12 niveles en 1ª lactación, 29 niveles en 2ª lactación y 61 niveles en 3ª lactación).
edp/cov = edad al parto en meses como covariable.
rba = efecto rebaño-año del control (196 niveles en 1ª y 2ª lactación y 201 en 3ª lactación).
rdc = efecto rebaño-año-mes de control (nº de niveles variable según control y lactación, mínimo 658 niveles y máximo 688).
est = estación del control (12 niveles en cada control).
del = días en lactación como variable categórica (43 niveles en el 1º control, 50 niveles en el 2º control, 55 niveles en el 3º control y 58 niveles en el 4º control).
del = días en lactación como covariable.
ε = residuo.

Los componentes de varianza se estimaron mediante el programa VCE 4.2 (Groeneveld, 1998) con un modelo multicarácter dentro de lactación (1ª, 2ª y 3ª o más lactaciones) basado en el modelo1 de efectos fijos, que incluyó cada uno de los cuatro controles productivos de la lactación de un animal, y cuya ecuación fue la siguiente:

$$y_{ijklmnp} = ncr_j + edp_k + rba_l + est_{m(i)} + del_{(n)i} + g_{(p)i} + \varepsilon_{ijklmnp}$$

Donde:

$y_{ijklmnp}$ = produción de leche (ml.) o porcentaje de proteína en el control i (i=1,4) de la oveja p.
ncr_j = número de crías al parto j.
edp_k = edad al parto k.
rba_l = grupo de comparación rebaño-año de control l.
$est_{(m)i}$ = mes m del control i.
$del_{(n)i}$ = días en lactación n hasta el control i.
$g_{(p)i}$ = genotipo del animal p para el control i.
$\varepsilon_{ijklmnp}$ = residuo.

En este análisis intervinieron 7.996 ovejas de 1ª lactación con un registro genealógico de 14.311 animales, 7.882 ovejas de 2ª lactación con un registro genealógico de 13.689 animales y 7.076 ovejas de 3ª o más lactaciones con un registro genealógico de 10.357 animales.

Posteriormente se realizó una estima de componentes de varianza de los caracteres leche y porcentaje de proteína tipificados a 120 días de lactación, con el modelo utilizado en el esquema de selección de esta raza (Serrano y col, 1996) sobre el total de los datos (22.954 registros productivos y 23.051 registros genealógicos).

Resultados y discusión.

Los resultados de los modelos de efectos fijos probados para todas las lactaciones y cada uno de los controles, se presentan en la Tabla 2. En general para todos los controles, lactaciones y modelos la mayoría de los efectos fueron muy significativos ($p<0.001$). Solo en algunos casos los efectos número de crías al parto y la edad al parto fueron significativas a más del 6% de probabilidad.

En general todos los modelos presentaron un coeficiente de determinación decreciente desde el primer control al cuarto control para ml. de leche en todas las lactaciones y una tendencia contraria, de aumento, para el porcentaje de proteína. El modelo que presentó un mayor coeficiente de determinación, para todas las lactaciones y cada uno de los controles dentro de lactación, fue el modelo 4. Estos resultados coinciden con gran cantidad de estudios en vacuno y ovino de leche (Rekaya, 1997; Strabel y Szwaczkowski, 1995; Barrillet y Boichard, 1994, etc.) en los que el grupo de comparación rebaño-fecha de control es el efecto sistemático más significativo en la determinación de la producción en el día de control. Sin embargo los modelos que incluyen dicho efecto, presentan problemas, ya que los grupos de comparación así constituidos son de pequeño tamaño, y como indican algunos autores en la revisión realizada por Swalve en 1995, esto puede ser el obstáculo más importante a la hora de abordar este tipo de modelos.

En nuestro caso para la estima de componentes de varianza, dada la imposibilidad de crear grupos de comparación rebaño-fecha de control para cada control dentro de lactación, con un suficiente número de observaciones por nivel, se recurrió a la utilización de los efectos sistemáticos del modelo 1, en el cual los efectos rebaño-año

TABLA 2: Coeficientes de determinación (R^2) en porcentaje de cada uno de los modelos de efectos fijos obtenidos para cada lactación y control productivo mediante el procedimiento GLM del paquete estadístico SAS, para ml. de leche y % de proteína.

		CONTROL 1		CONTROL 2		CONTROL 3		CONTROL 4	
LACTA	MODELO	Leche	Prot.	Leche	Prot.	Leche	Prot.	Leche	Prot.
1ª LACTA.	Modelo 1	48.81	27.40	47.80	32.74	47.33	31.07	42.46	33.87
	Modelo 2	48.49	26.90	47.55	32.27	46,96	30.42	41.81	33.06
	Modelo 3	48.74	27.32	47.69	32.67	47.25	30.87	42.37	33.68
	Modelo 4	55.97	38.76	55.54	43.23	55.14	40.70	50.37	43.71
	Modelo 5	55.70	38.37	55.30	42.82	54.85	43.71	49.80	43.02
	Modelo 6	55.92	38.66	55.45	43.17	55.02	40.58	50.33	43.57
2ª LACTA.	Modelo 1	43.55	26.47	40.85	30.35	35.16	26.63	31.93	31.80
	Modelo 2	43.19	26.10	40.52	29.84	34.82	25.98	31.48	31.25
	Modelo 3	43.34	26.02	40.60	30.05	34.93	26.29	31.64	31.60
	Modelo 4	50.91	36.87	49.68	41.33	43.67	37.45	42.70	43.20
	Modelo 5	50.63	36.53	49.39	40.88	43.34	36.98	42.27	42.77
	Modelo 6	50.71	36.50	49.51	41.19	43.50	37.19	42.50	43.07
3ª LACTA.	Modelo 1	35.60	21.02	29.61	24.72	29.02	25.91	26.99	31.80
	Modelo 2	36.14	21.67	30.00	25.30	29.81	26.15	27.52	32.12
	Modelo 3	35.61	21.22	29.70	24.95	29.26	26.15	27.24	31.84

	Modelo 4	43.64	31.38	38.25	36.05	39.15	37.19	38.10	42.99
	Modelo 5	43.47	31.12	31.12	35.71	38.96	36.77	37.69	42.56
	Modelo 6	42.97	30.70	30.70	35.48	38.48	36.80	37.44	42.44

de control, edad al parto y número de crías al parto son comunes a todos los controles de un mismo animal, y el mes de control y los días en lactación son distintos para cada uno de los controles productivos del animal.

Los parámetros genéticos estimados con el modelo multicarácter intra-lactación antes descrito para la producción de leche y % de proteína en el día de control se presentan en las tablas 3 y 4, respectivamente. En el caso de 3ª o más lactaciones para leche, las estimas se presentan sin error debido a limitaciones en el software.

Los resultados coinciden en líneas generales con aquellos obtenidos por algunos autores en vacuno de leche. El primer control presenta heredabilidades inferiores tanto para leche como % de proteína frente al resto de los controles, en todas las lactaciones estudiadas, como ocurre en Pander y col., 1992 y en Rekaya, 1997 con datos de 1ª lactación en vacuno de leche. Este hecho podría deberse a que es precisamente el primer control aquel que tiene mayor variabilidad en los días en lactación por ser mayor el intervalo de tiempo en el que se puede tomar la muestra (30 a 72 días después del parto según la normativa oficial).

Las correlaciones genéticas entre controles de una misma lactación presentaron valores altos y positivos, y generalmente de mayor magnitud entre controles sucesivos, tanto para leche como para proteína, coincidiendo con la tendencia general observada en las estimas obtenidas en vacuno de leche (Rekaya, 1997, Vargas y col, 1998). Las correlaciones fenótipicas en todos los casos presentaron valores muy inferiores a las genéticas, coincidiendo de nuevo con los resultados obtenidos en vacuno de leche (Pander y col. 1992, Rekaya, 1997 y Vargas y col, 1998). Las estimas de heredabilidad para la producción de leche y % de proteína tipificadas a 120 días de lactación tomaron los valores de 0.15±0.01 y 0.28±0.01, respectivamente, que resultan muy similares a los encontradas para los distintos controles y lactaciones.

TABLA 3: Parámetros genéticos y sus errores típicos, (heredabilidades en la diagonal, correlaciones genéticas sobre la diagonal y correlaciones fenotípicas bajo la diagonal) para cada uno de los cuatro controles productivos de leche (ml.), estimados para cada una de las lactaciones, primera, segunda y tercera o más.

		1° CONTROL	2ª CONTROL	3° CONROL	4° CONTROL
1ª LACTAC. 7996 reg. 14311 ani.	1° CONTROL	**0.15±0.02**	0.80±0.05	0.82±0.05	0.60±0.08
	2° CONTROL	0.52	**0.18±0.01**	0.95±0.02	0.77±0.05
	3° CONTROL	0.44	0.61	**0.17±0.02**	0.91±0.03
	4° CONTROL	0.33	0.50	0.61	**0.18±0.02**
2ª LACTAC. 7882 reg. 13689 ani.	1° CONTROL	**0.13±0.01**	0.89±0.05	0.70±0.08	0.60±0.09
	2° CONTROL	0.53	**0.13±0.02**	0.94±0.03	0.80±0.07
	3° CONTROL	0.43	0.64	**0.16±0.02**	0.79±0.06
	4° CONTROL	0.35	0.52	0.64	**0.14±0.02**

3ª o más LACTAC. 7076 reg. 10357 ani.	1° CONTROL	**0.10**	0.95	0.71	0.84
	2° CONTROL	0.56	**0.19**	0.89	0.96
	3° CONTROL	0.45	0.66	**0.14**	0.97
	4° CONTROL	0.34	0.53	0.63	**0.25**

TABLA 4: Parámetros genéticos y sus errores típicos, (heredabilidades en la diagonal, correlaciones genéticas sobre la diagonal y correlaciones fenotípicas bajo la diagonal) para cada uno de los cuatro controles productivos de % de proteína, estimados para cada una de las lactaciones, primera, segunda y tercera o más.

		1° CONTROL	2ª CONTROL	3° CONTROL	4° CONTROL
1ª LACTAC. 7996 reg. 14311 ani.	1° CONTROL	**0.17±0.02**	0.68±0.06	0.84±0.06	0.67±0.09
	2° CONTROL	0.38	**0.23±0.02**	0.93±0.03	0.60±0.07
	3° CONTROL	0.25	0.45	**0.19±0.02**	0.82±0.05
	4° CONTROL	0.17	0.28	0.43	**0.23±0.02**
2ª LACTAC. 7882 reg. 13689 ani.	1° CONTROL	**0.13±0.02**	0.85±0.06	0.84±0.07	0.50±0.08
	2° CONTROL	0.41	**0.25±0.02**	0.96±0.02	0.55±0.06
	3° CONTROL	0.26	0.44	**0.21±0.02**	0.75±0.05
	4° CONTROL	0.18	0.30	0.44	**0.24±0.02**
3ª o más LACTAC. 7076 reg. 10357 ani.	1° CONTROL	**0.17±0.03**	0.83±0.07	0.81±0.07	0.55±0.10
	2° CONTROL	0.48	**0.26±0.03**	0.99±0.01	0.92±0.05
	3° CONTROL	0.32	0.49	**0.21±0.02**	0.93±0.04
	4° CONTROL	0.20	0.32	0.48	**0.28±0.03**

Como conclusión a los estudios preliminares llevados a cabo en este trabajo podríamos apuntar que:

- Las heredabilidades de los distintos controles productivos tanto para leche como % de proteína son similares dentro y entre lactaciones y al valor de ésta para los caracteres tipificados a 120 días de lactación, a excepción del primer control cuyos valores de heredabilidad fueron claramente inferiores a los del resto de los controles en todas las lactaciones.
- Las correlaciones genéticas entre los controles de una misma lactación fueron distintas a 1, y mayores entre controles consecutivos, lo cual indica que los controles podrían no responder a la misma base genética.
- La producción de leche y porcentaje de proteína en el día de control podría ser una alternativa como criterio de selección al actual de producción tipificada a 120 días de lactación, ya que las heredabilidades son similares y el primero evitaría los errores cometidos en la propia toma de datos del control lechero.

Son necesarios más estudios, en el sentido de definir el tratamiento de estos

caracteres, dentro y entre lactaciones, y de profundizar en el sentido de las correlaciones entre la cantidad de leche y los parámetros de calidad de ésta (porcentaje de grasa, proteína y extracto seco) en los controles productivos. Por otro lado, también será objeto de futuros estudios, la definición del grupo de comparación más adecuado al modelo de valoración de estos nuevos caracteres, y la posible incorporación de otros efectos ambientales que afecten a estos caracteres productivos.

Referencias.

- **Baro, J.A.; Carriedo, J.A. y San Primitivo, F. 1994**. Genetic parameters of test day measures for somatic cell count, milk yield and protein percentage of milking ewes. J. Dairy Sci. 77: 2658-2662.
- **Barrillet, F. y Boichard, D. 1994**. Use of first lactation test-day data for genetic evaluation of the Lacaune dairy sheep. Proc. Of the 5th WCGALP, Guelph. 18: 111-114.
- **Groeneveld, E. 1998**. Programa VCE versión 4.2 para SunOS.
- **Jurado, J.J.; Serrano, M.; Pérez-Guzmán, M.D. y Montoro, V. 1997**. Estimas de componentes de varianza y respuesta a la selección en la raza ovina Manchega mediante Muestreo de Gibbs. Investigación Agraria, producción y sanidad animales. Vol. 12 (1, 2 y 3). Pags: 27-40.
- **Pander, N.L.; Hil, W.G. y Thompson, R. 1992**. Genetic parameters of test day records of British Holstein-Friesian heifers, Anim. Prod. 55:11-21.
- **Ptak, E. Y Schaeffer, L.R. 1993**. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34: 23-34.
- **Reents, R.; Dekkers, J.M.C. y Schaeffer, L.R. 1995**. Genetic evaluation with a multiple lactation test day model for SCS and production traits. Interbull Open Meeting. Praga.
- **Rekaya, R. 1997**. Análisis Bayesiano de datos de producción en los días del control para la selección de caracteres lecheros. Tesis Doctoral. E.T.S.I.A. Univ. Politécnica de Madrid.
- **SAS, 1990**. SAS/ETS versión 6. Sas Institute Inc. Cary, NC. USA.
- **Serrano, M.; Pérez-Guzmán, M.D.; Montoro, V. Y Jurado, J.J. 1996**. Genetic parameters estimation and selection progress for milk yield in Manchega sheep. Small Ruminant Research. 23:51-57.
- **Strabel, T. y Szwaczkowski, T. 1995**. The modelling of some nongenetic effects on test day milk yields in dairy cattle. 46th Annual Meeting of the EAAP. Praga.
- **Swalve, H.H. 1995**. Test day models in the analysis of dairy production data –a rewiev. Arch. Tierz., Dummerstorf. 38: 11-30.
- **Vargas, B.; Pérez. E. Y Van Arendonk, J.A.M. 1998**. Analysis of test day yield data of Costa Rica dairy cattle. J. Dairy Sci. 81: 225-263.

Estrategias de selección para controlar la consanguinidad en programas de mejora

B. Fernández†, M. Toro y A. Caballero†*

† Dep. Bioquímica, Genética e Inmunología, Universidad de Vigo

* Instituto Nacional de Investigaciones Agrarias

La selección con los métodos actuales de evaluación conlleva una mayor precisión y, con ello, una mayor respuesta, pero también conlleva un incremento paralelo en la consanguinidad. Es de todos conocido que la consanguinidad produce efectos indeseables, tales como la reducción de variabilidad genética, la depresión consanguínea y la reducción en la probabilidad de fijación de genes favorables. Por ello, en los últimos años se ha llevado a cabo un esfuerzo investigador considerable en la búsqueda de métodos para controlar la consanguinidad en programas de mejora de animales domésticos. Las estrategias investigadas están basadas en métodos de selección (*e.g.*, Wray y Goddard, 1994; Brisbane y Gibson, 1995; Meuwissen, 1997), métodos de apareamiento (*e.g.*, Caballero y col, 1996) o combinaciones de ambos (*e.g.*, Toro y Pérez-Enciso, 1990). Todos los métodos sugeridos están basados, de una manera u otra, en el mismo principio. La elección de los parentales debe llevarse a cabo penalizando a aquéllos individuos que tienen un mayor grado de parentesco con sus congéneres. Los métodos más recientes y sofisticados (por ejemplo el de Meuwissen, 1997), aunque muy eficientes, pueden resultar difíciles de ejecutar por mejoradores sin una base matemática sólida. En este trabajo, investigamos uno de los métodos más simples y clásicos de control de la consanguinidad, la minimización de la varianza de las contribuciones de ascendientes a descendientes.

La tasa de incremento en consanguinidad viene reflejada por el inverso del doble del censo efectivo de población (N_e). Sin selección y suponiendo apareamiento aleatorio, la expresión del censo efectivo es

$$\frac{1}{Ne}=\frac{1}{16M}\left[2+S^2_{mm}+2\left(\frac{M}{F}\right)S_{mm,mf}+\left(\frac{M}{F}\right)^2 S^2_{mf}\right]+\frac{1}{16F}\left[2+S^2_{ff}+2\left(\frac{F}{M}\right)S_{ff,fm}+\left(\frac{F}{M}\right)^2 S^2_{fm}\right] \quad (1)$$

(Hill, 1979), donde M (F) es el número de reproductores machos (hembras), S^2_{sm} (S^2_{sf}) es la varianza del número de hijos (hijas) con los que contribuyen los parentales de sexo s (m o f), y $S_{sm,sf}$ es la correspondiente covarianza. Es bien conocido que minimizando la varianza de las contribuciones de padres a hijos se produce un aumento del censo efectivo y, por tanto, un descenso de la tasa de consanguinidad. Por ejemplo, si cada padre contribuye con un hijo y F/M hijas, y cada madre contribuye con una hija, y con un hijo con probabilidad F/M, el censo efectivo se obtiene mediante la expresión $1/Ne = 3/16M + 1/16F$ (Gowe y col., 1959) y, si $M=F$, $Ne=2N$, aproximadamente. Se pueden obtener, sin embargo, censos efectivos aún mayores si, además de minimizar la varianza de las contribuciones, también se lleva a cabo aparemiento no aleatorio y subdivisión de la población (ver Wang, 1997).

En el contexto de la selección artificial, la mínima varianza de las contribuciones de padres a hijos se consigue con la selección intrafamiliar, pero la precisión de la selección y la respuesta se ven muy reducidas con este método (ver Hill y col., 1996). Se puede, sin

embargo, reducir la varianza restringiendo la pérdida en respuesta a un valor aceptable, por ejemplo el 5% de la respuesta que se obtendría sin ningún control en la consanguinidad (Toro y Pérez-Enciso, 1990). Este es el método que llamaremos de aquí en adelante, método MV.

No obstante, la fórmula (1) no es una predicción válida de la tasa asintótica de consanguinidad en programas de selección, ya que no sólo depende de las varianzas de las contribuciones en una generación, sino de las contribuciones a largo plazo. Esto ocurre porque con selección sobre un carácter heredable la ventaja selectiva se transmite a generaciones sucesivas exacerbando las asociaciones fortuitas entre genes neutros y seleccionados (ver Woolliams y col., 1993, Santiago y Caballero, 1995). Las contribuciones de ascendientes a descendientes se hacen constantes al cabo de unas pocas generaciones (cuatro o cinco sin ligamiento) y la varianza de estas contribuciones determina el censo efectivo. De esta forma, la ecuación (1) puede validarse si se sustituyen las varianzas y covarianzas de una generación por las correspondientes varianzas y covarianzas de las contribuciones de ascendientes en una generación a descendientes en futuras generaciones. Estas varianzas a largo plazo se pueden predecir en función de la varianza de una generación, la intensidad de selección, y la correlación genética y ambiental entre parientes.

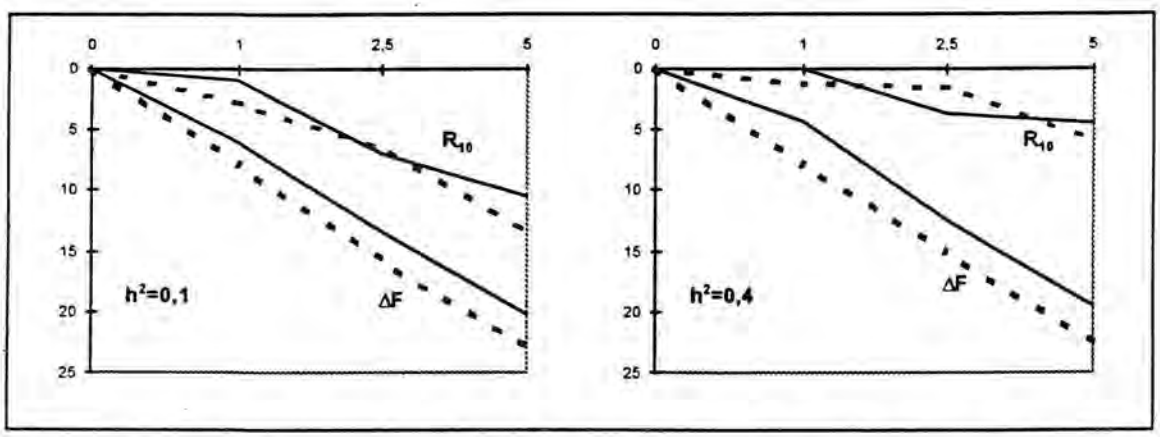
De la discusión anterior parece evidente que la minimización de la consanguinidad pasa por la minimización de las varianzas de las contribuciones de ascendientes a descendientes. Esto implicaría la minimización simultánea de la varianza de las contribuciones de todas las generaciones anteriores a la presente generación t, es decir, de la t-1 a la t, de la t-2 a la t, de la t-3 a la t, etc. Un resultado equivalente se obtendría seleccionando aquéllos individuos que presentan un menor parentesco entre ellos, en razón del siguiente argumento. Seleccionar individuos minimizando la varianza de las contribuciones de la generación parental, t-1, a la actual, t (o sea, maximizar el censo efectivo expresado en la ecuación 1) equivale a seleccionar aquéllos individuos que tienen un parentesco menor en relación a sus padres, es decir, equivale a minimizar el número de hermanos y medios hermanos. Análogamente, seleccionar individuos minimizando la varianza de las contribuciones de la generación t-2 a la t equivale a seleccionar aquéllos individuos que tienen un parentesco menor en relación a sus abuelos, es decir, equivale a minimizar el número de primos; y así sucesivamente. Por tanto, la selección de individuos con mínimo parentesco equivale a la selección con mínima varianza de las contribuciones a largo plazo y, con ello, a la forma teóricamente más efectiva de controlar la consanguinidad.

La selección de individuos con el mínimo parentesco (que llamaremos método MP) ha sido estudiada mediante métodos aproximados por Brisbane y Gibson (1995) y otros autores, y por Ballou y Lacy (1995) en el contexto de poblaciones no seleccionadas. Cuando la población base es una población de individuos no emparentados (o con igual parentesco entre ellos) y no hay selección, los métodos MV y MP deben ser equivalentes. Cuando la población base presenta parentescos desiguales entre los individuos, sin embargo, el método MP es más efectivo hasta que el parentesco medio se uniformice en la población (ver Ballou y Lacy, 1995). Con selección, sin embargo, se espera que la ventaja de MP sea general, ya que la propia selección producirá continuamente parentescos desiguales en la población.

Mediante simulación estocástica por ordenador llevamos a cabo una comparación entre los métodos MV y MP en una población con 16 machos y 16 hembras, de los que se seleccionan durante 10 generaciones los mejores 4 machos y 8 hembras de acuerdo con su fenotipo, y cada macho seleccionado se aparea con dos hembras para obtener de nuevo los 32 individuos evaluados en cada generación. El carácter seleccionado esta controlado por un modelo infinitesimal con heredabilidad 0,1 o 0,4. En cada generación, de entre los 32 individuos evaluados se buscan todas las posibles combinaciones de machos y hembras seleccionados cuya media fenotípica (desviada de la media de la población) no sea inferior en

una cierta proporción (la máxima respuesta que estamos dispuestos a perder) a la media obtenida seleccionando a los individuos con el mejor fenotipo. A continuación se calcula la varianza de las contribuciones de padres a hijos y el parentesco promedio para la combinación correspondiente. La combinación con mínima varianza (MV) o parentesco (MP) constituye la población parental de la siguiente generación. Puesto que se utilizan todas las posibles combinaciones de individuos, ambos métodos son exactos, a diferencia de estudios anteriores, en los que se usan procedimientos aproximados.

La siguiente figura muestra la reducción porcentual en la tasa de consanguinidad, así como la reducción porcentual en respuesta en la generación 10 para poblaciones con heredabilidad 0,1 y 0,4 y una restricción en el diferencial de selección del 1%, 2,5% y 5%. Los resultados con MV se representan con trazo continuo y los de MP con trazo discontinuo. Mientras que la pérdida en respuesta fue similar para los dos métodos (no hay diferencias significativas), la reducción en la tasa de consanguinidad fue significativamente mayor con el método MP que con el método MV. Estos resultados indican que en poblaciones seleccionadas la minimización del parentesco es un método más eficaz que la minimización de la varianza en una generación, incluso aunque la población base esté constituida por individuos no emparentados. Con heredabilidad alta, la pérdida en respuesta fue similar a la pretendida, pero con heredabilidad baja la pérdida fue mayor, probablemente por una reducción en la precisión de la selección.



Los resultados anteriores refieren a la situación en la que los genes seleccionados y neutros son independientes. Con ligamiento y selección se espera que los métodos anteriores sean ineficientes, ya que ambos utilizan la información genealógica, y el coeficiente de consanguinidad obtenido a partir de las genealogías no es un fiel reflejo de la probabilidad de homocigosis por descendencia (ver Santiago y Caballero, 1998 y Caballero y Santiago 1998).

Para contrastar los métodos en presencia de ligamiento, se corrieron simulaciones similares a las anteriores (con h^2=0,4) en las que el carácter seleccionado estaba controlado por un numero elevado de genes aditivos (300) con igual efecto, y 300 genes neutros intercalados entre los primeros. Se investigaron distintos niveles de ligamiento, con un único cromosoma sin recombinación (L=0), y uno (L=1) o cinco (L=5) cromosomas de un Morgan de longitud cada uno. La tasa de consanguinidad de evaluó a partir del incremento observado en la varianza de las frecuencias génicas de los genes neutros.

La siguiente tabla indica la respuesta en las generaciones 5 y 10, y la tasa de consanguinidad obtenidas para cada modelo de genoma cuando se aplica selección fenotípica sin ningún método de control de la consanguinidad (primera columna en cada caso), así como las correspondientes reducciones porcentuales cuando la selección se lleva a cabo utilizando el método MV y MP. Como es de esperar, la respuesta a la selección es mucho menor, y la tasa

de consanguinidad es mucho mayor cuando hay ligamiento estrecho que cuando hay recombinación libre ($L=\infty$).

	$L=\infty$			$L=5$			$L=1$			$L=0$		
		MV	MP		MV	MP		MV	MP		MV	MP
R_5	1.628	-4	-1	1.618	0	-4	1.564	-5	-3	1.446	-5	-6
R_{10}	2.92	-4	-1	2.827	-1	-3	2.500	-3	-3	1.816	-2	-3
ΔF	0.046	-12	-15	0.053	-11	-15	0.073	-14	-12	0.134	-5	-5

Los resultados indican que tanto MV como MP son ineficaces en reducir la tasa de consanguinidad cuando existe ligamiento total (L=0,) pero, sorprendentemente, con una mínima recombinación (L=1) el efecto desaparece, y las eficiencias de los métodos son similares a las que tendrían en un genoma con recombinación libre ($L=\infty$).

Este trabajo fue financiado por el Ministerio de Educación y Ciencia (proyecto PB96-0304).

REFERENCIAS

BALLOU, J.D., LACY, R.C. (1995): Identifying genetically important individuals for management of genetic variation in pedigreed populations. En *Population management for survival and recovery*. Ballou, J.D., Gilpin, M., Foose, T.J. (eds.) Columbia University Press. pp. 76-111.

BRISBANE, J.R., GIBSON, J.P. (1995): Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. Theoretical and Applied Genetics, 91: 421-431.

CABALLERO, A., SANTIAGO, E. (1998): Survival rates of major genes in selection programmes. Proc. 6th World Congress on Genetics Applied to Livestock Production, 26: 5-12.

CABALLERO, A., SANTIAGO, E., TORO, M.A. (1996): Systems of mating to reduce inbreeding in selected populations. Animal Science, 62: 431-442.

GOWE, R.S., ROBERTSON, A., LATTER, B.D.H. (1959): Environment and poultry breeding problems. 5. The design of poultry control strains. Poultry Science, 38 :462-471.

HILL, W.G. (1979): A note on effective population size with overlapping generations. Genetics, 92: 317-322.

HILL, W.G., CABALLERO, A., DEMPFLE, L. (1996): Prediction of response to selection within families. Genetics, Selection, Evolution, 28: 379-383.

MEUWISSEN, T.H.E. (1997). Maximising the response of selection with a predefined rate of inbreeding. Journal of Animal Science, 75: 934-940.

SANTIAGO, E., CABALLERO, A. (1995): Effective size of populations under selection. Genetics, 139: 1013-1030.

SANTIAGO, E., CABALLERO, A. (1998): Effective size and polymorphism of linked neutral loci in populations under selection. Genetics, en prensa.

TORO, M.A., PÉREZ-ENCISO, M. (1990). Optimization of selection response under restricted inbreeding. Genetics, Selection, Evolution, 22: 93-107.

WANG, J. (1997): More efficient breeding systems for controlling inbreeding and effective size in animal populations. Heredity, 79: 591-599.

WOOLLIAMS, J.A., WRAY, N.R., THOMPSON, R. (1993). Prediction of long-term contributions and inbreeding in populations undergoing mass selection. Genetical Research, 62: 231-242.

WRAY, N.R., GODDARD, M.E. (1994). Increasing long-term responses to selection. Genetics, Selection, Evolution, 26: 431-451.

SELECCIÓN ASISTIDA POR MARCADORES PARA CARACTERES NO ADITIVOS

M. A. Toro Ibáñez y J. Fernández Martín

Departamento de Mejora Genética y Biotecnología
S. G. I. T. - I. N. I. A. Crta. Coruña Km 7
28040. Madrid

Introducción

En un programa de mejora se obtendrían mayores respuestas si se aprovechasen los efectos dominantes del carácter sobre que el que se ejerce selección. Para ello se necesitan esquemas en los que se consideren dos poblaciones: una con los individuos que se comercializarán y otra con los individuos del núcleo que mantendrán la población (Toro, 1993). En este contexto las diferentes estrategias que pueden seguirse dependerán del criterio elegido para seleccionar los individuos que generarán la población comercial y los reproductores del núcleo y del sistema de apareamiento que se implemente con éstos últimos.

Por otra parte la disponibilidad creciente de marcadores que cubren todo el genoma de las especies de interés hace que la selección asistida por marcadores pueda convertirse en una alternativa real a los métodos tradicionales de selección.

En este trabajo se analizan los resultados de simular selección asistida por marcadores según esquemas con poblaciones comercial y reproductora para aprovechar la dominancia de un carácter.

Métodos

El carácter sobre el que se ejerce selección está determinado por la acción de 25 QTL dialélicos. La diferencia entre los efectos de los dos homocigotos para cada QTL es $2a$, siendo a valores en progresión geométrica (primer término igual a 10) de manera que la suma de todos ellos es 100 (Lande y Thompson, 1990; Gimelfarb y Lande, 1994). Los QTL están aleatoriamente distribuidos a lo largo de 20 cromosomas de 100 cM de longitud. La heredabilidad se fijó en 0,2.

En cada generación, compuesta por 64 individuos, se seleccionaron 16 individuos por sexo para obtener la siguiente generación permitiendo un solo apareamiento por individuo y una contribución igual de cada pareja (cuatro hijos y cuatro hijas). También se eligieron 16 individuos por sexo (los mismos que para el núcleo o diferentes) para generar la población comercial mediante apareamiento al azar y contribución igual por parejas.

Cada simulación constó de 10 generaciones de selección y los resultados fueron el promedio de 500 repeticiones.

Para cubrir un amplio rango de acciones génicas el valor del heterocigoto se calculó como el valor del homocigoto para el alelo favorable (a) por un factor que abarcaba desde cero hasta 1,125. En una simulación alternativa el valor de los heterocigotos se hizo proporcional a la magnitud del efecto de manera que al mayor de ellos correspondiese dominancia completa y al menor recesividad completa.

La selección se realizó sobre los efectos aparentes de 11 marcadores bialélicos por cromosoma regularmente espaciados. Para que la asociación con los QTL fuera máxima la población de partida fue una F1 heterocigota tanto para los QTL como para los marcadores; los QTL estaban en fase de repulsión total. Los efectos de cada marcador fueron calculados a través de una regresión múltiple del genotipo de cada marcador en el fenotipo de 1000 individuos provenientes de la F2. Sólo aquellos marcadores con un efecto significativo en la citada regresión fueron considerados en el proceso de selección.

Los esquemas de selección testados consistían en combinaciones diferentes de las siguientes opciones:

Selección

- Aptitud Combinatoria General (G).
- Aptitud Combinatoria Específica (S).
- Valor genotípico individual (I).

Apareamiento

- Al azar (R).
- Mínima Aptitud Combinatoria Específica (s).
- Mínima (Aptitud Combinatoria Específica - Aptitud Combinatoria General) (s^*).

Tanto la selección como el apareamiento (cuando no era al azar) se realizó mediante programación lineal entera (en concreto se implementó el *algoritmo húngaro* para problemas de asignación).

Como medida de la eficiencia de la selección asistida por marcadores se realizó selección sobre el valor fenotípico de los individuos.

Resultados y conclusiones

En las Tablas 1 y 2 aparecen las estrategias que rindieron mejores respuestas para cada una de las situaciones consideradas a corto, medio y largo plazo (2, 5 y 10 generaciones de selección). Las casillas sombreadas se refieren a la respuesta en la población comercial y las otras a la reproductora. La primera posición de cada terna indica el criterio de elección de los individuos reproductores, la segunda el sistema de apareamiento seguido con éstos y la tercera el modo de elección de los individuos que generan la población comercial. Cuando aparece un guión en alguna de las posiciones significa que varias estrategias obtuvieron el mismo resultado.

Tabla 1. Estrategias que generan la mejor respuesta cuando todos los QTL presentan el mismo grado de dominancia

		Generación 2						Generación 5						Generación 10					
d = 0	p = 0.5	*G*	-	-	*G*	*s*	*S*	*G*	*s*	*S*	-	*s*	*S*	*G*	*s*	*S*	*G*	*s*	*S*
d = 0.5	p = 0.5	*G*	*R*	-	*I*	*R*	*S*	*G*	*R*	-	-	*s*	*S*	*G*	*R*	-	*S*	*s*	*S*
d = 1	p = 0.5	*G*	*R*	-	*S*	s^*	*S*	*I*	*R*	*S*	*S*	s^*	*S*	*I*	*R*	*S*	*S*	*s*	*S*
d = 1.125	p = 0.5	*G*	*R*	-	*S*	s^*	*S*	*I*	*R*	*S*	*S*	s^*	*S*	*I*	*R*	*S*	*S*	s^*	*S*

Tabla 2. Estrategias que generan la mejor respuesta cuando el grado de dominancia de los QTL es proporcional a su efecto

	Generación 2						Generación 5						Generación 10					
p = 0.5	*G*	*R*	-	*G*	*s*	*S*	*G*	*R*	-	*G*	*s*	*S*	*G*	*R*	-	*G*	*s*	*S*

Población reproductora

Población comercial

G.- Aptitud Combinatoria General.
S.- Aptitud Combinatoria Específica.
I.- Valor individual.

R.- Apareamiento al azar.
s.- Apareamiento de mínima Aptitud Combinatoria Específica.
s^*.- Mínima (Aptitud Combinatoria Específica – Aptitud Combinatoria General).

Debemos hacer la consideración de que la eficacia de la selección asistida por marcadores se reduce a lo largo de las generaciones; de hecho a diez generaciones es menos efectiva que la selección fenotípica (resultados no mostrados). La causa es que la

recombinación va deshaciendo el desequilibrio de ligamiento entre QTL y marcadores y por tanto debilitando su asociación. Sería recomendable por tanto reevaluar los efectos aparentes cada pocas generaciones a efecto de eliminar los marcadores que ya no son significativos.

A la vista de los resultados algunas de las conclusiones a las que se puede llegar son:

- Respecto de la respuesta en la población comercial:
  * Para aditividad y, a corto plazo, dominancia parcial la mejor estrategia es elegir los individuos reproductores siguiendo el criterio G.
  * Para dominancia completa o sobredominancia es mejor la selección por S.
  * El sistema de apareamiento idóneo es s o s^*
  * El mejor criterio de selección de los individuos que van a producir la población comercial es S.
- Respecto de la respuesta en el núcleo:
  * Con aditividad o dominancia parcial el criterio de selección óptimo es G.
  * El apareamiento aleatorio resulta más adecuado (excepto en aditividad a largo plazo).

Cuando la dominancia es proporcional las combinaciones óptimas a cualquier tiempo son $G\,s\,S$ para la comercial y $G\,R-$ para el núcleo.

Obviamente el interés del mejorador será obtener la máxima respuesta en la población comercial aunque se debe tener en cuenta la consanguinidad que se genera en el núcleo ya que puede dificultar el mantenimiento de la población por los efectos deletéreos sobre eficacia biológica.

Bibliografía

- Gimelfarb, A. y Lande, R. 1994. Simulation of marker-assisted selection in hybrid populations. Genet. Res., Camb 63: 39-47.
- Lande, R. y Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756.
- Toro, M. A. 1993. A new method aimed at using the dominance variance in closed breeding populations. Genet. Sel. Evol. 25: 63-74.

INFORMATIVIDAD DE MARCADORES MOLECULARES TIPO MICROSATÉLITE EN DOS POBLACIONES DE CERDOS DE RAZA IBÉRICA Y LANDRACE

A. Clop [1], C. Ovilo [2], A. Coll [1], C. Barragán [2], J. M. Folch [1].

(1) Unitat de Genètica i Millora, Universitat Autònoma de Barcelona, 08193 Bellaterra;

(2) Area de Mejora Genética Animal, CIT-INIA, 28040 Madrid.

INTRODUCCIÓN

Este trabajo forma parte de un proyecto pluridisciplinar más amplio, cuyo objetivo es caracterizar regiones del genoma responsables de las características relacionadas con la calidad de la carne en porcino (the IBMAP consortium, 1998).

Gracias al trabajo realizado en el desarrollo del mapa genético del cerdo (Sus scrofa), en el que actualmente se dispone de más de 1000 microsatélites posicionados a lo largo del genoma (Rohrer et al., 1996), se pueden llevar a cabo experimentos para localizar QTL's ligados a estos marcadores.

El diseño experimental realizado se basa en un cruzamiento de tipo F2 entre ejemplares de características bien diferenciadas en cuanto a la calidad de la carne, como son las razas Ibérica y Landrace. En los animales de este cruzamiento se determinará el genotipo para algunos marcadores moleculares de tipo microsatélite y se medirá el valor fenotípico de los caracteres de calidad cárnica. La existencia en la F2 de asociaciones entre los alelos de un marçador con distintos niveles de expresión de alguno de los caracteres estudiados, nos indicará la presencia de factores genéticos, cercanos al microsatélite, responsables de las diferencias observadas entre las razas parentales en dichos caracteres de calidad.

En una primera fase de desarrollo del proyecto el objetivo más inmediato es la elección de los microsatélites, en función de su posición en el genoma y su informatividad en los cruzamientos realizados, para proceder al genotipado de los individuos de este cruzamiento.

MATERIAL Y MÉTODOS

Los animales usados en este trabajo son los que componen la F0 del cruzamiento del proyecto IBMAP Consortium.

Las extracciones del ADN genómico se han realizado mediante un protocolo de precipitación salina con Cloruro Sódico en saturación (5.5 M), (Miller et al. 1988).

Se han probado 139 microsatélites marcados fluorescentemente con las condiciones de amplificación descritas por Rohrer et al. (1996), con algunas modificaciones. El análisis de los productos de PCR se ha realizado mediante electroforesis capilar en un equipo automatizado de detección fluorescente (ABI PRISM 310 Genetic analyzer).

Para valorar la informatividad de cada uno de los microsatélites se ha utilizado el Índice de Ron (Ie) (Ron et al., 1995). Este cálculo ha sido adaptado a nuestro diseño experimental para poder aplicarlo a la población parental. Así, siendo Ai,Aj el genotipo para cada macho Ibérico, calculamos:

$$Ie = 1 - 1/2(Pi + Pj)$$

donde: Pi = Frecuencia del alelo Ai en la muestra de hembras Landrace.

Pj = Frecuencia del alelo Aj en la muestra de hembras Landrace.

Una vez calculado el Ie para cada macho, se calcula la media de los tres resultados. Así, cuanto menor sea la frecuencia de los alelos de los individuos Ibéricos en las hembras Landrace, más alto será el valor Ie para aquel microsatélite.

RESULTADOS Y DISCUSIÓN

Los microsatélites son repeticiones en tándem de 2 a 5 nucleótidos uniformemente distribuidos por todo el genoma, con una elevada tasa de variabilidad poblacional en el número de repeticiones. Una vez determinada su posición en el genoma, se pueden usar como marcadores para detectar la presencia de QTL'sguedant@club-internet.fr ligados a ellos y asociados a los fenotipos observados.

Para este tipo de experimentos nos interesan solamente los marcadores que sean suficientemente informativos (reflejado en un Ie alto), es decir, aquéllos en los que haya una alta divergencia en los alelos presentes en cada población y en las frecuencias de los mismos, ya que esto permite seguir la herencia de los alelos al analizar la segregación de la F2.

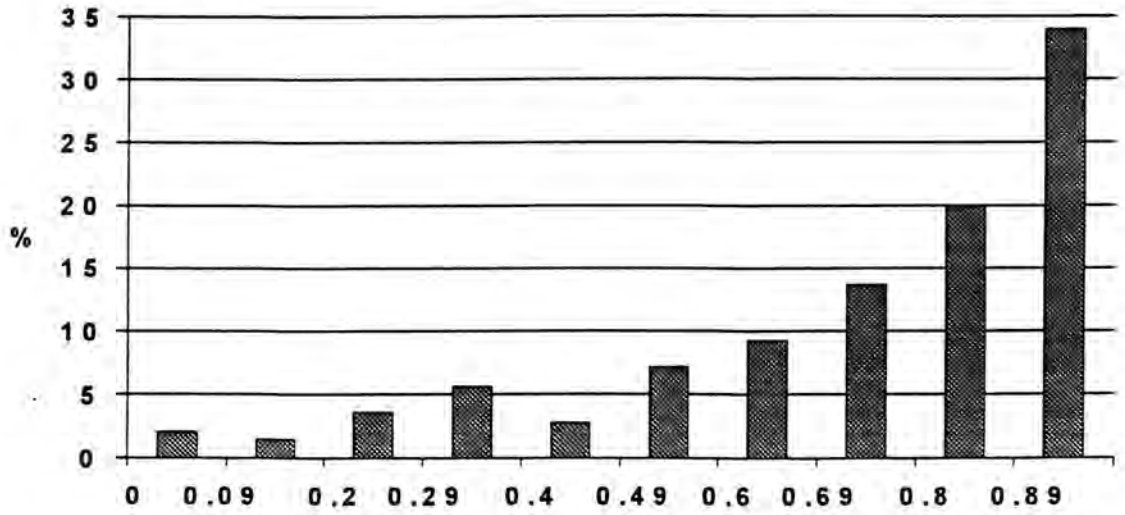
Inicialmente se analizaron los 139 microsatélites para los 3 animales Ibéricos y 5 Landrace, obteniendo los siguientes resultados : Ie = 0 en 3 microsatélites, Ie < 0.5 en 19, Ie =[0.50 -0.69] en 23 casos, Ie = [0.70-0.99] en 69 y Ie = 1 en 25 marcadores (Figura 1). La mayoria de estos marcadores presentan un nivel elevado de informatividad que refleja la diferencia genética existente entre las dos poblaciones del experimento. De estos 139 microsatélites analizados, se rechazaron 27 por estar en zonas del genoma suficientemente cubiertas por otros microsatélites más informativos, y 22 debido a su baja informatividad.

Finalmente, los marcadores seleccionados se genotiparon en el resto de hembras de la población F0 (26), y se volvió a calcular el Ie de cada microsatélite. Cuando se comparan los valores de Ie inferidos a partir de una muestra de la población parental con los calculados para la generación parental completa no se encuentran diferencias significativas, concluyéndose que en nuestro caso, y con la muestra de animales utilizada, el cálculo del Índice de Ron ha requerido pocos datos para obtener valores suficientemente fiables.

Los marcadores moleculares finalmente seleccionados, según su posición y su informatividad, son 90 microsatélites, los cuales están siendo usados actualmente para analizar

la descendencia del cruzamiento del proyecto IBMAP Consortium. Asimismo se han probado 80 microsatélites mediante tinción con nitrato de plata, con el fin de disponer de una batería adicional de microsatélites polimórficos que permitiría cubrir zonas del genoma en las que no se disponga de suficientes microsatélites fluorescentes informativos. El objetivo siguiente será relacionar los genotipos obtenidos con los caracteres fenotípicos de calidad estudiados.

figura 1.Porcentage de microsatélites (ordenadas) que presentan un determinado valor de Ie (abcisas), en intervalos de 0.1 unidades



AGRADECIMIENTOS

-a Nova Genètica, en especial a P. Borrás y E. Ramells, y al Area de Producción Animal del UdL-IRTA (M. Pérez Enciso, J. L. Noguera) por la cesión del material animal. La Junta de Castilla la Mancha cedió muy amablemente los verracos Guadyerbas a Nova Genètica.

-al *"THE usda SUPPORTED U.S. Pig Genome Coordination Project"* por la contribución de los primers que hemos usado para la elaboración del trabajo molecular del proyecto.

-Proyecto financiado por la CICYT AGF96-2510.

BIBLIOGRAFIA

-IBMAP Consortium, (1998), *An experiment to detect QTL's affecting meat quality in Iberian and Landrace pigs.* Abstract presentado en el *"6th world Congress on genetics applied to livestock production"*, celebrado en Sidney en 1998.

-Miller, S.A., Dykes, D.D. & Polensky, H. F. (1998), *A simple salting out procedure for extracting DNA from human nucleated cells*, Nucleic Acid Research, 16, n. 3, 1215.

-Rohrer G., Alexander L., Hu Z., Smith T., Keele J., Beattie C., (1996). *A comprehensive Map of the Porcine Genome*, Genome Research, 6:371-391.

-Ron. M., Lewin H., Da Y., Band. M., Yanai A., Blank Y., Feldmesser E. & Weller J. I, (1995), *Prediction of informativeness for microsatellite markers among progeny of sires used for detection of economic trait loci in dairy cattle,* Animal Genetics, 26, 439-441.

ASOCIACIÓN DE UNA LORDOSIS/ESCOLIOSIS/SIFOSIS A LA ESTRUCTURA FAMILIAR EN LA CRIA INTENSIVA DE DORADA (*Sparus aurata*)

Afonso, J.M[1]., Montero, D.[2], Robaina, L.[2], Izquierdo, M.[3], Ginés, R.[1]

[1] Departamento de Producción Animal, Facultad de Veterinaria, Universidad de las Palmas de Gran Canaria.
[2] Instituto Canario de Ciencias Marinas.
[3] Departamento de Biología, Facultad de Ciencias del Mar, Universidad de las Palmas de Gran Canaria.

INTRODUCCIÓN

Dentro de la piscicultura marina en nuestro país, la dorada ocupa un lugar preponderante, habiendo multiplicado por cinco su producción en los últimos siete años (datos de la Subdirección General del Caladero Nacional, Acuicultura y Recursos Marinos, M.A.P.A.), y con perspectivas de seguir creciendo. Uno de los problemas que aparecen con la cría intensiva es la presencia de proporciones importantes de individuos con algún tipo de malformación, fundamentalmente en la pigmentación, en las escamas, en la vejiga natatoria y en el sistema esquelético (Divanach *et al.*, 1996), motivos por los cuales hay una notable depreciación del valor final del producto. De las anomalías esqueléticas, las de la columna vertebral son las que tienen mayor importancia porque pueden representar hasta el 5% de la producción total de una empresa de engorde (Andrades *et al.*, 1996), todo ello a pesar de la criba a la que son sometidos los peces durante su periodo de alevinaje en las empresas multiplicadoras (*hatchery*). Los tipos más frecuentes que afectan a la columna son la lordosis (deformidad dorsal de la columna vertebral, forma de V), la escolisis (deformidad lateral de la columna vertebral, forma de zig-zag) y la sifosis (deformidad ventral de la columna vertebral, forma de Λ).

Se han identificado diferentes factores no genéticos (*post-hatching*) como causantes de las incidencias de malformaciones esqueléticas, especialmente la escoliosis y la lordosis, en las especies cultivadas más relevantes como el salmón atlántico (*Salmo salar*), la trucha arcoiris (*Oncorhynchus mykiss*), la tilapia (*Tilapia nilotica*), la dorada (*Sparus aurata*), la lubina (*Dicentrarchus labrax*), el pez gato(*Ictalurus punctatus*), el pez limón (*Seriola sp.*) y la dorada japonesa (*Pagrus major*). Entre ellos se encuentran los relacionados con los nutrientes en dieta (deficiencias en triptófano o vitamina C, o excesos en vitamina D o tirosina), las condiciones hidrodinámicas y de cultivo, la calidad del agua (presencia de metales pesados, pesticidas y otros agentes químicos), y el parasitismo (Divanach *et al.*, 1996). Por el contrario, son escasos los trabajos que indican alguna naturaleza genética de éstas malformaciones, ya sea mediante herencia mendeliana simple de un gen recesivo o dominante, o por segregación poligénica (Purdom, 1993; Divanach *et al.*, 1996). En dorada, sólo han sido descritas algunas evidencias por Andrades *et al.* (1996) en dorada europea y por Taniguchi *et al.* (1984) en dorada japonesa, quienes aducen razones genéticas, asícomo de composición del huevo y factores ambientales en los reproductores o embriones, para explicar el 27% de larvas deformes al día 0 de vida *post-hatching* y las diferencias en incidencias de lordosis entre grupos familiares, respectivamente

El objetivo de la presente comunicación es poner de manifiesto la asociación encontrada entre una lordosis/escoliosis/sifosis a la estructura familiar de dorada bajo condiciones de cría intensiva.

MATERIALES Y METODOS

Las doradas utilizadas en éste estudio fueron crecidas bajo las condiciones naturales de luz propias de la latitud 28° 10'N y en régimen de cría intensiva, en el Instituto Canario de Ciencias Marinas, y pertenecen a la primera generación de un programa de selección divergente para los niveles de cortisol plasmático post-estrés de confinamiento (Afonso *et al.*, 1998). Se utilizaron los 11640 descendientes pertenecientes a familias de hermanos carnales (algunos a partir de cruces recíprocos), distribuidos en 2 ó 3 réplicas por familia a lo largo del crecimiento, asícomo 3 réplicas de una puesta masal de toda la población. Las réplicas de cada familia se establecieron en diferentes momentos de su puesta, la cual se prolongó desde diciembre de 1996 hasta marzo de 1997.

Desde el inicio del experimento, las larvas de dorada fueron crecidas en tanques de fibra de vidrio de 500 litros con circulación de agua, y alimentadas con rotíferos (*Brachionus plicatilis*) enriquecidos con SELCO, *Nannochloropsis sp.* e *Isochrypsis sp.*, a partir del 3-4° día de vida y hasta el día 21. Entonces se les suministró *Artemia* hasta el día 36 y posteriormente piensos comerciales de la casa comercial E.W.O.S., S.A.. La densidad de los alevines fue de 12 kgs/m^3, el oxígeno, suministrado mediante aireación, estuvo comprendido entre 6 y 8 mg/l y la temperatura del agua osciló entre 18,3 °C y 22,8 °C.

Entre el 12 de noviembre y el 12 diciembre de 1997, y con aproximadamente un año de edad, todos los peces de cada tanque fueron analizados morfológicamente, clasificando las diferentes malformaciones por réplica y familia. A una muestra de peces con las deformidades morfológicas más patentes se les hicieron radiografías mediante rayos X sobre placas de mamografía a 150 Kv y 6.8 mA por segundo, a fin de corroborar la valoración macroscópica de las deformidades.

Para el análisis estadístico se utilizó el siguiente modelo loglineal, sobre el paquete estadístico SPSS:

$$\ln \hat{f}_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

$\hat{f}_{ij}$ = es la frecuencia esperada de la celda *ij*.

μ = es la media de los logaritmos de las frecuencias esperadas.

α_i = es el efecto de la categoría *i* del factor A.

β_j = es el efecto de la categoría *j* del factor B.

$\alpha\beta_{ij}$ = es el efecto debido a la interacción de la categoría *i* del factor A y la categoría *j* del factor B.

RESULTADOS

Se determinaron 17 deformidades individuales de columna (lordosis, fusiones de vértebras, etc.), prognatismos, opérculos abiertos, mandíbulas dobladas y deformaciones dorso-ventrales. Las 22 clases de deformidades restantes respondieron a combinaciones de dos o más de las individuales. Así, en la familia denominada 2H, y en cada una de sus tres réplicas, apareció una Lordosis/Escoliosis/Sifosis (LES) (ver figuras 1a y 1b), consistente en una repetición sucesiva de las tres deformidades desde la cabeza hasta la cola, y que en comparación a un pez normal (figura 1c) da al pez el aspecto de estar comprimido, permitiéndole sólo mover el pedúnculo terminal de la aleta caudal.

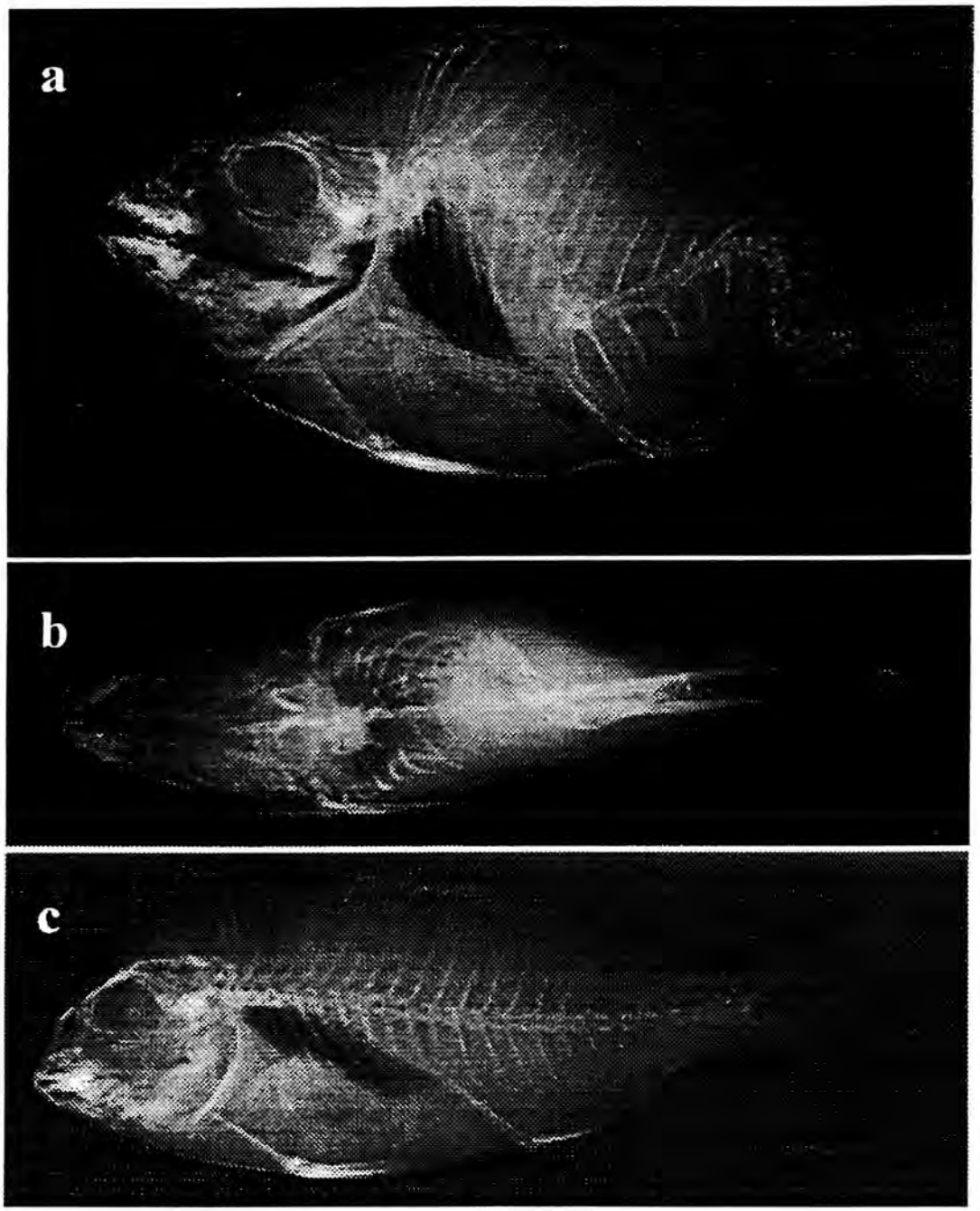


Figura 1. a, visión lateral de la deformidad LES. b, visión dorso-ventral de la deformidad LES. c, visión lateral de un individuo normal.

En la tabla I se muestran las frecuencias observadas de las diferentes deformidades encontradas para las familias 2H, 4H, 2mh y 4mh; donde 2mh y 4mh corresponden con las descendencias de los cruces recíprocos de los padres de las familias 2H y 4H. No hubo interacción estadísticamente significativa dentro de éstas familias entre el tipo de deformidad y la réplica, ni tampoco entre el tipo de deformidad y las familias 4H, 2mh y

4mh ($0,10<P<0,99$) . Sin embargo, si hubo una asociación estadísticamente significativa entre la familia 2H con la deformidad ($P<0.001$), de modo que aparecieron más LES de los que cabría esperar si hubiera independencia entre la familia y la deformidad ($Z_{2H.LES}=3,18$; $P<0,05$). Esta asociación también surgió considerando toda la población (Tabla II) ($P<0,001$), con un valor de $Z_{2H.LES}=4,28$ ($P<0,05$).

Tabla I. Frecuencias observadas de las diferentes deformidades en las familias 2H, 4H, 2mh y 4mh.

DEFORMIDAD	FAMILIA			
	2H	4H	2mh	4mh
Normal	257	280	651	69
Lordosis	8	0	12	0
prognatismo Superior	0	3	2	0
prognatismo Inferior	1	3	14	0
Mandíbula doblada	2	0	6	0
Opérculo I	0	1	6	0
FRente prominente	2	25	39	1
Opérculo II	1	0	9	0
Ojos	0	0	2	0
Dorso cóncavo	1	0	0	0
L+FR	0	0	2	0
L+I	1	0	0	0
FR+I	0	5	14	0
FR+B	1	0	1	0
LES	19	0	0	0

Tabla II. Frecuencias observadas de las 39 malformaciones para toda la población, excepto la familia 2H, agrupadas según el tipo de deformidad de columna.

DEFORMIDAD	POBLACION
Normal	11005
Lordosis	321
Fusión de vértebras	21

DISCUSIÓN

Las deformidades de columna representaron aproximadamente el 3,2 % de todos los peces, y la deformidad LES sólo el 0,2 %, lo cual está dentro de los rangos normales de las empresas donde se lleva un control de los requerimientos (Andrades *et al.*, 1996), y algo por encima de los valores que se han detectado sobre poblaciones naturales de otras especies, 0,7% en *Morone saxatilis* (Hickev & Young, 1984).

Para desmarcar algo más los condicionantes ambientales, hicimos una comparación de las réplicas de la familia 2H con aquellas réplicas pertenecientes a otras familias de la población, las cuales fueron establecidas el mismo dia y por tanto con las condiciones de cría más parecidas posibles, y nuevamente encontramos la asociación estadísticamente significativa entre la familia 2H y la deformidad LES ($P<0,001$) ($Z_{2H,LES}=2,77$; $P<0,05$).

Los factores nutricionales juegan un papel importante en el normal crecimiento de los peces (Koeniq, 1979). Así, deficiencias o excesos de la vitaminas C o D, respectivamente, generan altas incidencias de deformidades vertebrales en distintas especies (Hilton *et al.*,1978; Sato *et al.*, 1978; John *et al.*, 1979; Mahajan *et al.*, 1979; Wunder, 1981; Chávez de Martínez, 1990; Lall *et al.*, 1990; Mgbenka, 1991; Kanazawa *et al.*,1992; Dabrowski *et al.*, 1996). De igual modo, las deficiencias en triptófano o excesos en tirosina deforman irreversiblemente a los peces (Nacario, 1983; Walton, *et al.*,1984; Akiyama, *et al.*,1985; Akiyama, *et al.*,1986a; 1986b), y cuyos grados de deformación se pueden acentuar con la velocidad del agua (Chatain, 1994). Sin embargo, los efectos nutricionales debieron ejercer una influencia similar sobre las diferentes familias debido a la homogeneidad del cultivo, ya que se emplearon los mismos piensos para todas las descendencias.

A fin de contrastar la posible determinación genética de la deformidad LES, y el modo de segregación, se harán cruces con las descendencias de las familias 2H, 2mh, 4mh y 4H, con control cuantitativo de parámetros de cultivo que afectan a las incidencias, previo y posterior al apareamiento.

AGRADECIMIENTOS

Agradecer la ayuda prestada al Dr. Grau-Brassas y al Dr. Pestano en la realización de las placas radiográficas y su adecuación al texto. Este trabajo ha sido financiado por la Unión Europea con el proyecto FAIR-CT95-0152.

BIBLIOGRAFIA

Afonso, J.M., Montero, D., Robaina, L., Fernández, H., Izquierdo, M. and Ginés, R., 1998. Selection programmes for stress tolerance in fish (Review). *Options Méditerranéennes*, 34: 235-245.

Akiyama, T., Arai, S., Murai, T. and Nose, T., 1985. Tryptophan requirement of chum salmon fry. *Bull. Jap. Soc. Sci. Fish. Nissuishi.*,51(6):1005-1008.

Akiyama, T., Murai, T. and Mori, K., 1986. Role of tryptophan metabolites in inhibition of spinal deformity of chum salmon fry caused by tryptophan deficiency. *Bull. Jap. Soc. Sci. Fish. Nissuishi.*,52(7):1255-1259.

Akiyama, T., Murai, T. and Nose, T., 1986b. Oral administration of serotonin against spinal deformity of chum salmon fry induced by tryptophan deficiency. *Bull. Jap. Soc. Sci. Fish. Nissuishi.*,52(7):1249-1254.

Andrades, J.A., Becerra, J. and Fernández-Llebrez, P., 1996. Skeletal deformities in larval, juvenile and adult stages of cultured gilthead sea bream (*Sparus aurata* L.). *Aquaculture*, 141:1-11.

Chatain, B., 1994. Abnormal swimbladder development and lordosis in sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus aurata*). *Aquaculture*, 119: 371-379.

Chávez de Martínez, M.C., 1990. Vitamin C requierement of the mexican native cichlid *Cichasoma urophthalamus. Aquaculture*, 86(4): 409-416.

Dabrowski, K., Moerau, R. and El Saidy, D., 1996. Ontogenetic sensitivity of channel catfish to ascorbic acid deficiency. *J. Aquat. Anim. Health.*, 8(1):22-27.

Divanach, P., Boglione, C., Menu, B., Koumoundouros, G., Kentouri, M. and Cataudella, S., 1996. Abnormalities in finfish mariculture: an overview of the problem, causes and solutions. SEABASS AND SEABREAM CULTURE: PROBLEMS AND PROSPECTS INTERNATIONAL WORKSHOP; 45-66.

Hickey, C.R. and Young, B.H., 1984. Incidence of morphological abnormalities in striped bass from the Houston river and coastal Long Island, New York. *N.Y. Fish Game J.*, 31(1):104-108.

Hilton, J.W., Cho, C.Y. and Slinger, S.J., 1978. effect of graded levels of supplemental ascorbic acid in practical diets fed to rainbow trout (*Salmo gairdneri*).*J. Fish. Res. Board Can.*, 35(4):131-136.

John, T.M., George, J.C., hilton, J.W. and Slinger, S.J., 1979. Influence of dietary ascorbic acid on plasm lipid levels in the rainbow trout. *Int. J. vitam. Nutr. Res.*, 49(4):400-405.

Kanazawa, A., Teshima, S., Koshio, S., Higashi, M. and Itoh, S., 1992. Effect of L-ascorbyl-2-phosphate-Mg on the yellowtail *Seriola quinqueradiata* as a vitamin C source. *Nippon Suisan Gakkishi - Bull. Jap. Soc. Sci. Fish.*, 1992, 58(2):337-341.

Koeniq, J., 1979. Vitamins, macro and micro-mineral trace-elements in fish nutrition. *Oceanis Doc. Oceanogr.*, 5(2):131-144.

Lall, S.P., Olivier, G., Weerakoon, D.E.M. and Hines, J.A., 1990. The effect of vitamin C deficiency and excess on immune response in Atlantic salmon (*Salmo salar*). *3 Int. Symp. on Feeding and Nutrition in Fish.*:427-441.

Mahajan, C.L. and Agrawal, N.K., 1979. Vitamin C deficiency in *Channa punctatus*. *J. Fish Biol.*, 15(5):613-622.

Mgbenka, B.O., 1991. Digestibility of some agricultural by products by African catfish (*Clarias gariepinus* Burchell 1822) and preliminary investigation on the effects of L-ascorbic acid (vitamin C) in an agricultural by product based diet on the African catfish. *Aquaculture Research and Development in rural Africa*. Edited by Costa, P.B.A.,

Lightfoot, C., Ruddle, K. and Pullin, R.S.V., 27: 30.

Nacario, J.F., 1983. The effect of thyroxine on the larvae and fry of *Sarotherodon niloticus* (Tilapia nilotica). *Aquaculture*, 34(12):73-83.

Purdom, C.E., 1993. Genetics and Fish Breeding. Chapman & Hall.

Sato, M., Yoshinaka, R. and Ikeda, S., 1978. dietary ascorbic acid requirement of rainbow trout for growth and collagen formation. *Bull. Jap. Soc. Sci. Fish.*, 44(9): 1029-1035.

Taniguchi, N., Azuma, K. And Umeda, S., 1984. Differences due to parents in incidence of vertebral malformation in artificially bred red seabream. *Bull. Jap. Soc. Sci. Fish Nissuishi*, 50(5):787-792.

Walton, M.J., Coloso, M., Cowey, C.B., Adron, J.W. and Knox, D., 1984. The effects of dietary tryptophan levels on growth and metabolism of rainbow trout (*Salmo gairdneri*). *Br. J. Nutr.*, 51(2)279-287.

Wunder, W., 1981. Lateral Curvature of the Vertebral Column (Scoliosis) of Carp (*Cyprinus carpio*). A Photopgraphic study. *Fisch Teichwirt*, 32(10):293-296.

APLICACIÓN DE LA INSEMINACIÓN ARTIFICIAL EN LA OPTIMIZACIÓN DEL ESQUEMA DE SELECCIÓN EN LA AGRUPACIÓN CAPRINA CANARIA : RESULTADOS PRELIMINARES

Lorenzo, M1 ; Fresno, M1 ; Darmanin, N 1; Delgado, J.V2; Barba, C. J.2 y Afonso, J.M.3

1 Unidad de Producción Animal Pastos y Forrajes. ICIA. Apdo. 60. La Laguna. Tenerife
2 Unidad de Veterinaria. Departamento de Genética. UCO. Medina Azahara, 9. Córdoba.
3 Departamento de Producción Animal. F. Veterinaria. Universidad de Las Palmas.

El desarrollo del esquema de selección de la agrupación caprina canaria y ha atravesado por las mismas dificultades que el resto de razas pertenecientes a esta funcionalidad en nuestro País, y que se centran en el un bajo desarrollo estructural y metodológico de las ganaderías que ocasionan por un lado grandes dificultades en el control de rendimientos y por otro lado, una información genealógica de escasa credibilidad.

Nuestro equipo está tratando de resolver el segundo aspecto por la vía del uso de planes de apareamientos alternativos en las explotaciones, la exclusión de paternidad con la aplicación restringida de marcadores del ADN (microsatélites) y con el desarrollo de las técnicas de inseminación artificial con semen fresco y refrigerado que además nos ofrecerán unas importantes aportaciones para la conexión genética de los rebaños y posteriormente, en fases de mayor desarrollo del esquema, para la difusión de la mejora.

Estamos aplicando las técnicas tradicionales tanto para la sincronización de celos como para la inseminación con semen fresco y refrigerado. Nuestro propósito inicial ha contemplado tres aspectos :
desarrollo metodológico de las técnicas de inseminación
sensibilización de los ganaderos
integración de esta técnicas en el esquema

Por el momento nuestras experiencias con semen fresco nos han ofrecido un 75 % de fertilidad y una prolificidad total de 192 %, en las 88 cabras inseminadas en las cuatro estaciones del año 1.997. Además hemos realizado inseminaciones de este tipo en dos ganaderías del núcleo de control lechero de Tenerife con un total de 50 animales. Estas inseminaciones se desarrollaron en primavera y su contrastación preliminar está pendiente de realizarse en estos días por ecografía.

La inseminación refrigerada nos ofreció unos resultados del 80,95 % de fertilidad en las 19 cabras inseminadas en la estación experimental.

CARACTERIZACIÓN DE LAS VARIEDADES MINORITARIAS DEL CERDO IBÉRICO.

Delgado, J.V.1; Barba, C.J.1; Rodero, E.2; Pardo, J.1; Mata, C.1; Cañuelo, P.3; Dieguez, E.3 y Afonso, J. M.4

1 Unidad de Veterinaria. Departamento de Genética. UCO. Medina Azahara, 9. Córdoba.
2 Departamento de Producción Animal. F. Veterinaria. UCO. Medina Azahara, 9. Córdoba.
3 A.E.C.E.R.I.B.E.R. Apdo. 40. 06300 Zafra. Badajoz.
4 Departamento de Producción Animal. F. Veterinaria. Universidad de Las Palmas.

La agrupación porcina Ibérica tiene una estructura poblacional especialmente compleja manifestándose diversas estructuras genéticas en su interior. Los tipos, estirpes y líneas del cerdo Ibérico constituyen una riqueza genética que le ofrece una versatilidad adaptativa y productiva para la explotación en los ambientes más diversos de la España seca.

En 1.996 se establece una relación científica entre AECERIBER, MAPA y UCO con objeto de definir y caracterizar desde los puntos de vista morfológico, productivo y genético las diversas formas existentes dentro de esta agrupación.

En esta comunicación describimos los planteamientos generales de este trabajo de caracterización en sus tres facetas.

En el plano morfológico se han medido 24 variables morfológicas, 16 cuantitativas y 8 cualitativas. Con esta información se ha desarrollado un estudio en tres fases :

Análisis estadísticos descriptivos

Estudio pormenorizado de las diferencias existentes para cada variable con el uso de GLM y test de proporciones.

Análisis discriminante canónico y valoración de las distancias de Mahalanovis.

Desde el punto de vista productivo se han analizado 16 variables de pesos y crecimientos así como del peso de piezas nobles. El análisis estadístico realizado fue similar al descrito anteriormente.

En cuanto a la caracterización morfológica nos hemos basado en el estudio de los 27 microsatélites recomendados por la ISAG y la FAO para estudios de este tipo. En este aspecto, se han tipificado los animales, se ha realizado un análisis de frecuencias alélicas, se han calculados los niveles de heterocigosidad, se ha testado el equilibrio de cada uno de los caracteres estudiados y finalmente se han calculado las distancias de Nei.

MEJORA Y CONSERVACIÓN GENÉTICA DE LAS VARIEDADES DEL PODENCO ANDALUZ

Barba, C.J.1; Delgado, J.V.1; Sierra, A.C.1 y Afonso, J.M.2

1 Unidad de Veterinaria. Departamento de Genética. UCO. Medina Azahara, 9. Córdoba.
2 Departamento de Producción Animal. F. Veterinaria. Universidad de Las Palmas.

El Podenco Andaluz es quizás la raza canina española cuyos criadores cuentan con el mayor nivel organizativo todo ello a pesar de ser volumétricamente la población canina más importante de nuestro País., al menos en número de animales.

El Club Nacional del Podenco Andaluz tenía conciencia de la necesidad de aplicar criterios científicos a la gestión genética de sus poblaciones, pero en nuestro País se carecía casi por completo de experiencias en esta materia. En 1995 se inician los contactos entre esta asociación de criadores y nuestro equipo con el fin de iniciar trabajos en colaboración conducentes a resolver los siguientes problemas :

Gestión genética de las pequeñas poblaciones que conforman las variedades minoritarias del Podenco Andaluz

Desarrollo de criterios objetivos de selección para la evaluación genética tanto morfológica como funcional (aptitud de caza) de estos animales

Optimización de la metodología de evaluación fenotípica y genética de los candidatos a reproductores

Diseño y optimización de esquemas de selección y conservación

En esta comunicación presentamos en primer lugar la estructura poblacional de esta raza, en la que se encuentran representadas y reconocidas en el libro de registros nueve variedades completamente aisladas reproductivamente. Con esta información realizamos una clasificación de las mismas en dos niveles : variedades en conservación y susceptibles de mejora. En una segunda fase describimos las acciones en desarrollo para cada uno de estos niveles y finalmente profundizamos en la metodología de control y de evaluación genética de los candidatos a reproductores.

CARACTERIZACIÓN MORFOMÉTRICA DE LOS RECUSOS GENÉTICOS OVINOS CANARIOS.

Alvarez, S1 ; Capote, J.1; Fresno, M1 ; Delgado, J.V2; Barba, C. J.2 ; Afonso, J.M.3 y Rodero, E.4

1 Unidad de Producción Animal Pastos y Forrajes. ICIA. Apdo. 60. La Laguna. Tenerife
2 Unidad de Veterinaria. Departamento de Genética. UCO. Medina Azahara, 9. Córdoba.
3 Departamento de Producción Animal. F. Veterinaria. Universidad de Las Palmas.
4 Departamento de Producción Animal. F. Veterinaria. UCO. Medina Azahara, 9. 14005 Córdoba.

En la actualidad existen en las Islas Canarias dos razas ovinas autóctonas ; la raza Canaria y la raza Palmera. La primera de ellas es de gran implantación en algunas islas del Archipiélago, disponiendo de unos censos entre 19000 y 21000 cabezas que manifiestan una aptitud láctea muy marcada e interesante, con cifras de producción que pueden alcanzar los 180 kg. en lactaciones de 120 días. Por el contrario, la raza Palmera dispone de unos efectivos por debajo del centenar de ejemplares y se encuentra por tanto al borde de la extinción, situación a la que ha llegado por haber quedado desubicada zootécnicamente en la actual estructura de sector agrario canario.

Estas razas mantienen una importancia genética adicional por sus relaciones ancestrales con ovinos ibéricos (Churra, Gallega y Lacha) y norteafricanos (ovinos magrebíes) de lana basta, así como con múltiples agrupaciones raciales criollas de Iberoamérica como puedan ser el Churro Chiapas, el Navajo Churro Sheep y los ovinos criollos argentinos y uruguayos, entre otros.

En el presente trabajo se ha realizado un estudio estadístico concienzudo para la caracterización morfológica de ambas razas y para el establecimiento de diferencias y similaridades entre ellas, sentando las bases para futuros estudios filogenéticos de la línea de emigración de los ovinos de lana basta desde Europa hacia América a partir del siglo XV. Estos estudios morfológicos se podrán ampliar en el futuro a una contrastación genética con aplicación de marcadores del ADN, fundamentalmente mitocondrial.

En este estudio se han utilizado 12 variables morfológicas medidas en una muestra de 139 animales para la raza Canaria y de 64 animales para la raza Palmera, a las que se sometió a un análisis estadístico descriptivo, una comparación directa mediante pruebas de homogeneidad de medias y un análisis discriminante canónico.