SOBRE LA CONVERGENCIA DEL MUESTREO DE GIBBS EN LOS MODELOS BAYESIANOS JERÁRQUICOS

L.A. García Cortés y C. Cabrillo

Resumen

Desde la introducción del muestreo de Gibbs en mejora genética se ha extendido su uso en combinación con los modelos bayesianos jerárquicos. Aquí mostramos cómo el modelo aumentado con los valores aditivos de los animales y resuelto con un muestreo de Gibbs ofrece resultados que numéricamente no concuerdan con los obtenidos a partir de un modelo bayesiano en el que no se haya realizado el aumento. La aparente paradoja se resuelve en favor de la técnica de aumento de datos y en contra del muestreo de Gibbs cuya convergencia presenta problemas cerca del origen de varianzas.

Introducción

Durante los últimos años, el muestreo de Gibbs (Gelfand y Smith, 1989) se ha popularizado en mejora genética debido a su sencillez de cálculo y su capacidad para proporcionar las distribuciones marginales de los parámetros de interés. El modelo bayesiano jerárquico se adapta perfectamente a este algoritmo, debido a que las distribuciones condicionales necesarias presentan formas conocidas y sencillas de implementar (Wang et al, 1993).

En el modelo bayesiano jerárquico, es bien conocido que las distribuciones condicionales de la varianza entre machos presentan una forma proporcional a una ji-cuadrado inversa. No parece extraño entonces que las distribuciones marginales de la varianza entre machos obtenidas usando el muestreo de Gibbs, es decir, promediando condicionales a partir de una cadena de Markov, presenten una ordenada nula en el cero. Tal es el caso de las figuras presentadas en (Korsgaard et al, 1998; Sorensen et al, 1994; Stranden y Gianola, 1999; Wanget al, 1993) entre otras.

Por contra, en los resultados que presentamos en García Cortés et al (1999), la distribución marginal de un componente de varianza toma un valor mayor que cero en el origen. La diferencia entre este trabajo y los anteriormente mencionados es únicamente la manera de implementar el muestreo de Gibbs. Concretamente en este último trabajo se utiliza un modelo donde se integran los efectos aditivos de los animales, es decir, un análisis en el que no se utiliza un modelo aumentado.

En este último trabajo se construye la verosimilitud a partir de una única distribución, esto es, asumiendo que los datos observados siguen una distribución normal multivariante cuyos parámetros se desean inferir. En el modelo que vamos a usar aquí (un modelo macho por simplicidad) esta distribución se expresa como

$$\mathbf{y} \sim N\left(\mathbf{Xb}, \mathbf{ZZ}'\sigma_u^2 + \mathbf{I}\sigma_e^2\right) \qquad (1)$$

Como se expone en lo que sigue, las distribuciones de la componente de varianza presentan diferencias sustanciales en un entorno del cero según que en el análisis se utilize o no la técnica de aumento de datos en clara contradicción con la esencia misma de dicha técnica, a saber, que la inferencia no se ve modificada en el proceso

(Tanner y Wong, 1987). Demostraremos aquí como la aparente paradoja se debe a un problema de convergencia del muestreo de Gibbs cuando se aplica al modelo con aumento de datos. Como cabría esperar, la inferencia en sí es independiente del modelo elegido y, en contra de la intuición, las verdaderas distribuciones marginales no son nulas en el origen aunque de una forma singular que impide la convergencia correcta del muestreo de Gibbs.

El modelo

En aras de una mayor simplicidad y, puesto que no afecta a la generalidad de nuestras conclusiones, nos restringiremos a un modelo macho descrito por

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

donde $\mathbf{y} \sim N\left(\mathbf{Xb}, \mathbf{ZZ}'\sigma_u^2 + \mathbf{I}\sigma_e^2\right)$ son los datos observados, **b** son efectos fijos desde el punto de vista frecuentista, $\mathbf{u} \sim N\left(0, \mathbf{I}\sigma_u^2\right)$ son los efectos genéticos de los padres, $\mathbf{e} \sim N\left(0, \mathbf{I}\sigma_e^2\right)$ son los residuos del modelo y **X** y **Z** son las matrices de incidencias que relacionan los niveles de cada efecto con las observaciones.

La distribución posterior de σ_u^2 puede obtenerse por marginalización con respecto a **b** y σ_e^2 de la posterior conjunta:

$$f\left(\mathbf{b}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y}\right) \qquad (2)$$

Llamaremos a éste el análisis 1 (o modelo 1). Se corresponde con el utilizado en Garcia Cortés et al (1999).

Es bien conocido que puede simplificarse el problema aumentando el modelo anterior con las variables incluidas en **u**, para después obtener la marginal posterior de σ_u^2 marginalizando con respecto a **u**, **b** y σ_e^2 en:

$$f\left(\mathbf{u}, \mathbf{b}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y}\right) \qquad (3)$$

Este segundo análisis (o modelo) está basado en el aumento de datos y está garantizado que las marginales de σ_u^2 en (2) y (3) son iguales (Tanner y Wong, 1987). Este último será el análisis 2 y consiste en el modelo bayesiano jerárquico habitual en mejora genética (Wang et al, 1993).

Implementación

Las marginalizaciones en (2) y (3) van a ser realizadas usando un muestreo de Gibbs. Dicha marginalización se llevará a cabo de la forma habitual en tres fases: definir la posterior conjunta en función de la verosimilitud y distribuciones a priori, tomar la condicional de σ_u^2 en cada ciclo de la cadena y promediar las condicionales. Es decir, no usamos los puntos muestreados sino que acumulamos las densidades condicionales, método conocido como de Rao-Blackwell (Gelfand y Smith, 1989). En ambos análisis las posteriores son (suponemos todas las priori planas):

Análisis 1.

$$f\left(\mathbf{b},\sigma_u^2,\sigma_e^2\,|\,\mathbf{y}\right)\propto f\left(\mathbf{y}\,|\,\mathbf{b},\sigma_u^2,\sigma_e^2\right)$$

$$f\left(\mathbf{b},\sigma_u^2,\sigma_e^2\,|\,\mathbf{y}\right)\propto \left|\mathbf{ZZ}'\sigma_u^2+\mathbf{I}\sigma_e^2\right|^{-0.5}\exp\left\{-0.5(\mathbf{y}-\mathbf{Xb})'\left(\mathbf{ZZ}'\sigma_u^2+\mathbf{I}\sigma_e^2\right)^{-1}(\mathbf{y}-\mathbf{Xb})\right\}$$

Análisis 2.

$$f\left(\mathbf{u},\mathbf{b},\sigma_u^2,\sigma_e^2\,|\,\mathbf{y}\right)\propto f\left(\mathbf{y}\,|\,\mathbf{u},\mathbf{b},\sigma_u^2,\sigma_e^2\right)f\left(\mathbf{u}\,|\,\sigma_u^2\right)$$

$$f\left(\mathbf{u},\mathbf{b},\sigma_u^2,\sigma_e^2\,|\,\mathbf{y}\right)\propto \sigma_e^{-n}\sigma_u^{-q}\exp\left\{-\frac{(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})}{2\sigma_e^2}-\frac{\mathbf{u}'\mathbf{u}}{2\sigma_u^2}\right\}$$

donde n es el número de datos y q es el número de animales.

Tomando las condicionales completas para σ_u^2 en ambos modelos se obtiene

Análisis 1.

$$f\left(\sigma_u^2\,|\,\mathbf{b},\sigma_e^2,\mathbf{y}\right)\propto \left|\mathbf{ZZ}'\sigma_u^2+\mathbf{I}\sigma_e^2\right|^{-0.5}\exp\left\{-0.5(\mathbf{y}-\mathbf{Xb})'\left(\mathbf{ZZ}'\sigma_u^2+\mathbf{I}\sigma_e^2\right)^{-1}(\mathbf{y}-\mathbf{Xb})\right\}$$

Análisis 2.

$$f\left(\sigma_u^2\,|\,\mathbf{u},\mathbf{b},\sigma_e^2,\mathbf{y}\right)\propto \sigma_u^{-q}\exp\left\{-\frac{\mathbf{u}'\mathbf{u}}{2\sigma_u^2}\right\} \qquad (4)$$

en donde las correspondientes constantes de proporcionalidad son independientes de **u**. Una vez obtenidas las condicionales completas, las marginales se obtienen promediando las de cada ciclo del muestreo de Gibbs. Centrándonos en el punto $\sigma_u^2=0$, las densidades a promediar resultan en

Análisis 1.

$$f\left(\sigma_u^2=0\,|\,\mathbf{b},\sigma_e^2,\mathbf{y}\right)\propto \left|\mathbf{I}\sigma_e^2\right|^{-0.5}\exp\left\{-0.5(\mathbf{y}-\mathbf{Xb})'\left(\mathbf{I}\sigma_e^2\right)^{-1}(\mathbf{y}-\mathbf{Xb})\right\} \qquad (5)$$

Análisis 2.

$$\lim_{\sigma_u^2\to 0} f\left(\sigma_u^2\,|\,\mathbf{u},\mathbf{b},\sigma_e^2,\mathbf{y}\right)\propto \delta(\mathbf{u}'\mathbf{u}) \qquad (6)$$

donde $\delta(\ )$ denota delta de Dirac. La variable **u'u** representa una densidad localizada en cero. Es fácil convencencerse de que este comportamiento singular es el correcto inspeccionando (4). En el caso de que **u'u** valga cero el factor exponencial en el lado

derecho de (4) se reduce a 1 de tal forma que la divergencia en $\sigma_u^2 = 0$ del prefactor σ_u^{-q} no se ve compensada en este caso por dicho factor exponencial. Al mismo tiempo, puesto que (4) representa una distribución de probabilidad es fácil convencerse de que el límite $\sigma_u^2 \to 0$ aunque divergente, se aproxima manteniendo el area finita con respecto a todas las variables y en particular con respecto a **u'u**. Matemáticamente este tipo de divergencia define la delta de Dirac que, si bien no es una función en el sentido habitual, adquiere un sentido riguroso bien definido dentro de la llamada teoría de las funciones generalizadas. En particular, la integral de la delta para este caso en el que está centrada en el origen del intervalo de integración ha de identificarse con 1/2. Este tipo de comportamiento singular conduce a una probabilidad finita en $\sigma_u^2 \to 0$ garantizando la equivalencia con el modelo 1 cuya probabilidad finita en $\sigma_u^2 = 0$ es evidente por simple inspección de (5). Sin embargo, al mismo tiempo este comportamiento singular impide una convergencia correcta del muestreo de Gibbs puesto que la probabilidad de alcanzar el valor **u'u**=0, el único punto que contribuye al valor no nulo de la marginal en el origen, es infinitesimal.

Ejemplo numérico

Dejaremos un análisis matemático más completo para una publicación posterior. Aquí nos limitaremos a ilustrar el problema con un ejemplo numérico. La muestra corresponde a un modelo con 50 machos, 25 efectos fijos en un solo factor y 500 datos. Para la simulación se usó una varianza residual de 95 y una varianza entre machos de 5. La figura 1 presenta las marginales en ambos modelos y corrobora lo anteriormente expuesto: Ambas condicionales son distintas en el punto $\sigma_u^2 = 0$, forzando a que las marginales sean distintas en el entorno de cero.

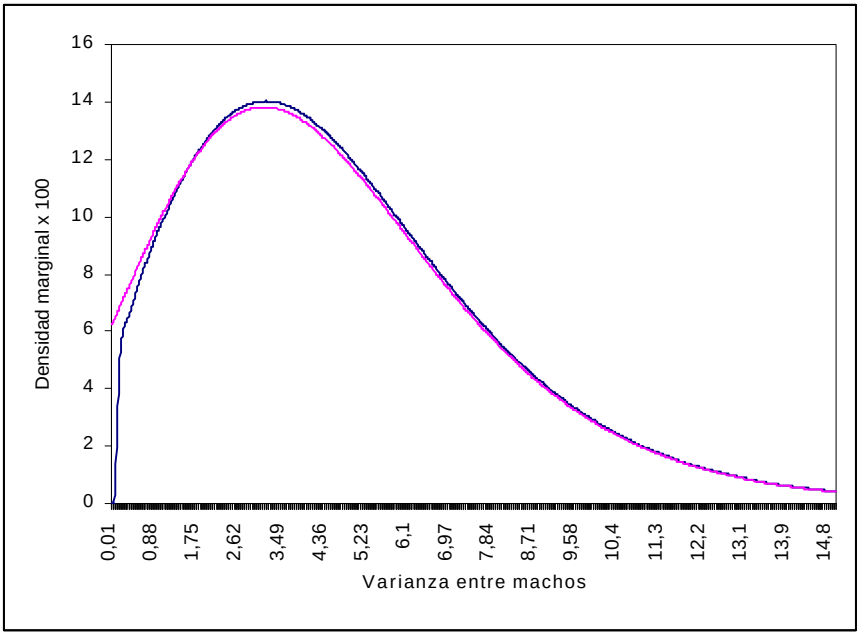

***Figura 1**. Distribuciones marginales posteriores para la varianza entre machos en los modelos 1 y 2.*

Un simple experimento numérico puede convencernos de que el resultado correcto es el del análisis 1. Para ello utilizaremos una a priori discreta para σ_u^2, con el 50% de la

densidad en el punto 0.1 y el 50% restante en el punto 0.2 (dos puntos escogidos por estar en el entorno de cero) y comparemos los resultados de los siguientes casos:

Caso 1. Densidad marginal posterior para el análisis 1 con una a priori plana.
Caso 2. Densidad marginal posterior para el análisis 2 con una a priori plana.

Estos dos primeros casos se obtienen directamente de la figura 1, tomando las ordenadas en los puntos 0.1 y 0.2.

Caso 3. Probabilidad marginal posterior para el análisis 1 con una a priori discreta
Caso 4. Probabilidad marginal posterior para el análisis 2 con una a priori discreta

Estos dos últimos casos se obtienen repitiendo el análisis con dicha a priori incluida explícitamente en el modelo. La diferencia con los dos análisis anteriores está en el muestreo de S_u^2: esta condicional solamente tiene probabilidad en los puntos 0.1 y 0.2 y la cadena muestrea solamente entre estas dos posibilidades. Nótese que en estos dos casos hablamos de probabilidades en lugar de densidades, ya que la distribución a priori discreta da lugar a una marginal posterior discreta. La idea subyacente es que a pesar de que las a priori discretas usadas están dentro de la zona en la que ambos modelos discrepan el carácter localizado de éstas debería evitar, o al menos disminuir fuertemente, la influencia perniciosa de la mala convergencia en el origen. Por otro lado, como ambas a priori, la plana y la discreta, presentan el mismo cociente entre sus valores en los dos puntos estudiados (esto es, 1 en ambos casos), los correspondientes cocientes de las posteriores deberían también coincidir (aunque, evidentemente, ya no serían igual a 1)

Caso	Análisis	A priori	Densidad en 0.1	Densidad en 0.2	cociente
1	1	Plana	0.06572	0.06948	1.057
2	2	Plana	0.01905	0.05441	2.856
3	1	Discreta	0.4861	0.5139	1.057
4	2	Discreta	0.4822	0.5178	1.073

Tabla 1. *Densidades o probabilidades marginales posteriores en los cuatro casos*

La Tabla 1 confirma estas hipótesis dentro de una precisión numérica más que razonable. Efectivamente, los casos 1 y 3 correspondientes al modelo 1 dan el mismo cociente de densidades. Así mismo, los casos 3 y 4 correspondientes a modelos diferentes pero con la a priori discreta coinciden razonablemente bien tanto en las densidades como en el cociente. El cociente en el caso 2, por contra, difiere fuertemente del obtenido en los otros casos, confirmando así, que los dos modelos son equivalentes pero que el muestreo de Gibbs es inapropiado cuando las a priori toman valores finitos en el origen.

La figura 2 permite visualizar el comportamiento patológico que invalida el muestreo de Gibbs. En ella se muestra una serie de 5 condicionales de las que se promedian para obtener la curva del análisis 2 de la figura 1 correspondientes de derecha a izquierda a valores decrecientes de **u'u**. Según **u'u** se va aproximando a cero, se observa claramente como al mismo tiempo que se van desplazando hacia el origen se van estrechando manteniendo un área finita. En el límite **u'u**=0 colapsan en una condicional infinitamente estrecha, centrada en el origen y de área finita, esto es, en algo proporcional a una delta de Dirac. A la vista de la figura 2 es fácil entender que según nos acercamos al origen la densidad marginal se hace progresivamente más difícil de estimar puesto que en el promedio van pesando cada vez más, condicionales

más estrechas y más infrecuentes, hasta que en el propio origen la estimación fracasa completamente.

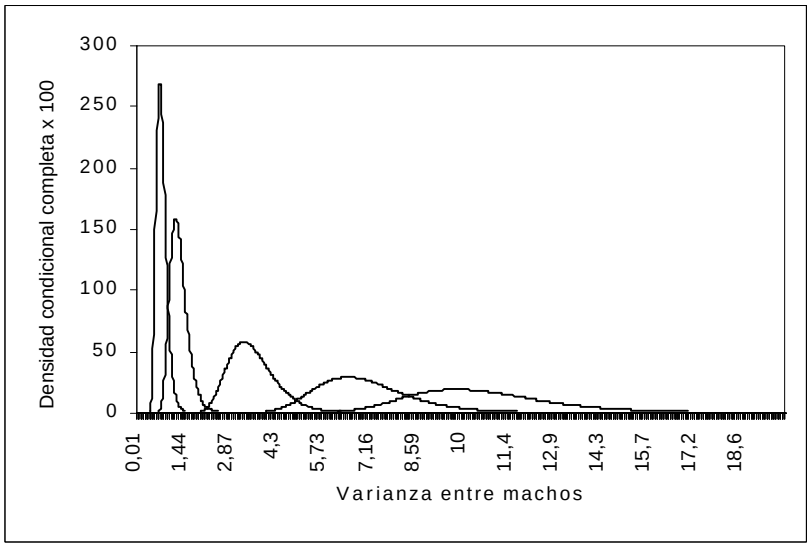

***Figura 2**. 5 condicionales de las que se promedian para obtener la gráfica anterior*

Por último, nótese que el haber utilizado distribuciones a priori no informativas (planas) y por tanto impropias, no tiene relación alguna con el mal comportamiento del muestreo. De hecho, en las simulaciones numéricas las a priori planas están perfectamente definidas puesto que el rango de la variable aleatoria es evidentemente finito. Se han utilizado nada más que por su simplicidad de implementación. Cualquier otra a priori con valor no nulo y finito en el origen presentaría el mismo problema de convergencia.

Implicaciones

La implicación de lo aquí expuesto es clara: En caso de necesitar probabilidades de intervalos en la cola izquierda de la distribución marginal posterior, se debería o bien prescindir del muestreo de Gibbs junto con aumento de datos o bien corroborarlo con alguna simulación similar a la expuesta en la sección anterior. En el caso extremo en el que se necesite la ordenada en el eje, el modelo jerárquico es del todo inadecuado.

Así mismo, habría que ser cuidadoso con los análisis que proporcionan marginales cuya densidad disminuye rápidamente al acercarse al cero. Estos deberían ser revisados, bien utilizando un experimento similar al empleado en la sección anterior (usando una a priori discreta) o bien implementando un modelo sin aumento de datos.

Citas bibliográficas

García Cortés et al (1999) ITEA, Vol Extra
Gelfand y Smith (1990) J. Amer. Stat. Assoc.,85:398-409
Korsgaard et al (1998) Genet. Sel. Evol. 30:241-256
Sorensen et al (1994) genet. Sel. Evol. 26:333-360
Stranden y Gianola (1999) Genet. Sel. Evol. 31:25-42
Tanner y Wong (1987) Tools for statistical inference. Springer
Wang et al (1993) Genet. Sel. Evol. 25:41-62