



Una aplicación del enfoque bayesiano para estimar componentes de varianza y heredabilidad

An application of Bayesian analysis to estimate components of variance and heritability

Ana Vargas Paredes¹, Víctor Maehara Oyata¹

¹ Universidad Nacional Agraria La Molina, Lima, Perú. Email: anavargas@lamolina.edu.pe; vmaehara@lamolina.edu.pe

Recepción: 5/08/2018; Aceptación: 05/06/2019

Resumen

Este trabajo tiene por objetivo estimar los componentes de variancia del efecto genético aditivo del animal y del rebaño, así como la heredabilidad aditiva y la proporción de la variancia ambiental rebaño, a partir de un conjunto de datos de lactaciones de ganado lechero de raza Holsteins con información genealógica. Para obtener estos estimados se utilizó un modelo lineal mixto para la producción de leche, en función de los efectos fijos del factor número de lactación y de la covariable logaritmo del número de días en leche. La parte aleatoria del modelo estuvo en función del efecto genético aditivo del animal, del efecto del rebaño y del error. El modelo fue ajustado vía inferencia bayesiana utilizando el muestreo de Gibbs.

Palabras clave: Análisis bayesiano; modelo lineal mixto; heredabilidad.

Abstract

The aim of this work is to estimate components of variance of the additive genetic effects of the animal and the herd. Also we estimate the heritability and the proportion of the herd environmental variance using a data set of lactation of Holsteins dairy cattle with genealogical information. To obtain these estimates, a linear mixed model was used for milk production, based on the fixed effects of the lactation number factor and the covariate logarithm of the number of days in milk. The random part of the model was a function of the additive genetic effect of the animal, the effect of the herd and the error. The model was adjusted via Bayesian inference based on Gibbs sampling.

Keywords: Bayesian analysis; linear mixed models; heritability.

Forma de citar el artículo: Vargas, A.; Maehara, V. 2019. Una aplicación del enfoque bayesiano para estimar componentes de varianza y heredabilidad. *Anales Científicos* 80 (2):391-396 (2019).

DOI: <http://dx.doi.org/10.21704/ac.v80i2.1455>

Autor de correspondencia (*): Ana Vargas Paredes. Email: anavargas@lamolina.edu.pe

© Universidad Nacional Agraria La Molina, Lima, Perú.

1. Introducción

En el campo de mejoramiento genético animal, la estimación de parámetros genéticos como la heredabilidad (proporción de la varianza fenotípica atribuida a factores genéticos aditivos) se usa para predecir los valores de cría, los cuales son utilizados para realizar la selección de los animales. Los procedimientos estadísticos utilizados para realizar estas estimaciones se basan principalmente en dos grandes metodologías, una basada en máxima verosimilitud restringida conocida como REML y otra basada en procedimientos bayesianos (Sorensen y Gianola, 2002).

En la aproximación bayesiana se combina lo que se conoce de los parámetros (distribuciones *a priori*) con la información que proporciona los datos para obtener la distribución *a posteriori*, la cual representa la incertidumbre sobre los parámetros después de que se ha tomado en cuenta la información de los datos (Blasco, 2001). El procedimiento computacional estándar, en la estimación bayesiana de parámetros genéticos, es el método de Montecarlo y cadenas de Markov (MCMC) para obtener muestras de la distribución *a posteriori*, a partir de diferentes algoritmos, entre ellos el algoritmo de muestreo de Gibbs y el algoritmo de Metropolis-Hastings, los cuales son muy populares y han sido implementados en diferentes softwares. Esta metodología ha sido aplicada en muchas de las áreas de interés en el mejoramiento genético animal. Uno de los primeros estudios fue realizado por Wang *et al.* (1993, 1994), quienes aplicaron el análisis bayesiano vía muestreo de Gibbs, para estimar parámetros genéticos relacionados con el tamaño de la camada de cerdos ibéricos a partir de un modelo univariado. Este trabajo tiene como objetivo principal estimar el parámetro genético heredabilidad de la producción de leche mediante un modelo animal unicarácter a partir de registros de 3397 lactaciones

de ganado lechero de raza Holsteins con información genealógica que comprende 6547 animales descargados desde United State Department of Agriculture, USDA, utilizando inferencia bayesiana.

2. Materiales y Métodos

Modelo animal formulado para estimar componentes de varianza

Siguiendo la formulación de Henderson del modelo animal, Sorensen & Gianola (2002) y Mrode (2014) resumen matricialmente el modelo animal univariado a utilizar para datos de una variable fenotípica con distribución normal, como:

$$y = x\beta + Z_1u_1 + Z_2u_2 + \varepsilon \quad (1)$$

donde y es un vector que corresponde a la producción de leche estandarizada, β es un vector que corresponde a los efectos del factor número de lactación o parto; y la covariable logaritmo del número de días en leche; el vector u_1 corresponden al efecto genético aditivo del animal y u_2 efecto del rebaño, X , Z_1 y Z_2 son matrices de incidencia relacionadas con β , u_1 y u_2 respectivamente y ε es un vector aleatorio de residuales.

El modelo (1) escrito de forma individual para cada observación por Vazquez et al. 2010 es:

$$y_{ijk} = \beta_0 + L_i + \beta_1 \log(dim)_{ij} + c_j + h_k + e_{ijk} \quad (2)$$

donde:

y_{ijk} es la producción de leche estandarizada sobre el parto i , j ésima vaca, y del rebaño k ; β_0 es la media general; L_i es el efecto fijo del número de lactación o parto ($i = 1, 2, \dots, 5$); dim_{ij} es el número de días en leche de la vaca j en la i -ésima lactación (covariable); β_1 es el coeficiente de regresión de dim ; c_j es el efecto aleatorio aditivo de la vaca j ($j = 1, 2, \dots, 1359$); h_k es el efecto aleatorio del rebaño k ($k = 1, 2, \dots, 57$); y e_{ijk} es el efecto aleatorio residual.

Modelo lineal mixto para estimar componentes de varianza por el método bayesiano

La distribución condicional que generan los datos del modelo descrito en (1) es:

$$y/\beta, \mathbf{u}_1, \mathbf{u}_2, \sigma_e^2 \sim N(\mathbf{X}\beta + \sum_{i=1}^2 \mathbf{Z}_i \mathbf{u}_i, \mathbf{I}\sigma_e^2)$$

Respecto a los supuestos de las distribuciones que son: $\mathbf{u}_1/\mathbf{A} \sim N(\mathbf{0}, \mathbf{A}\sigma_1^2), \mathbf{u}_2 \sim N(\mathbf{0}, \mathbf{I}\sigma_2^2)$ y $\varepsilon \sim N(\mathbf{0}, \mathbf{I}\sigma^2\varepsilon$, donde \mathbf{A} es la matriz de covarianzas aditivas entre los individuos, además $\mathbf{u}_1, \mathbf{u}_2$ y ε son asumidas por ser independientes entre sí.

Respecto a los a prioris para β como es usual se asumió una distribución uniforme, es decir $p(\beta) \propto$ constante, además de independencia entre $\beta, \mathbf{u}_1, \mathbf{u}_2$ y ε .

La densidad posterior conjunta de todos los parámetros desconocidos es proporcional a:

$$p(\beta, \mathbf{u}_1, \mathbf{u}_2, \sigma_1^2, \sigma_2^2, \sigma_e^2/y) \propto p(\beta) p(\mathbf{u}_1/\sigma_1^2) p(\sigma_1^2) p(\mathbf{u}_2/\sigma_2^2) p(\sigma_2^2) p(\sigma_e^2) p(y/\beta, \mathbf{u}_1, \mathbf{u}_2, \sigma_e^2)$$

a partir de esta se deduce la distribución posterior completa de cada parámetro a estimar.

Para conseguir muestras de la distribución posterior conjunta se aplicó el muestreo de Gibbs con una sola cadena de 100000 iteraciones y almacenadas cada 10 iteraciones, descartándose las 10000 primeras (burn-in). Para ello se utilizó la librería MCMCglmm implementado en el paquete R (Hadfield, 2010).

Para chequear el comportamiento del algoritmo MCMC se observó la convergencia a través de los gráficos de traza, que muestran la evolución de los valores muestreados a lo largo de las iteraciones y las autocorrelaciones de la cadena de las muestras. Asimismo, se obtuvo el error

de Montecarlo para cada estimación con la finalidad de medir la variabilidad de la estimación a través de la simulación.

Descripción de los datos

Los datos son registros de 3397 lactaciones del primer al quinto parto de 1359 vacas Holsteins, hijas de 38 toros en 57 rebaños. Todos los registros corresponden a vacas con al menos 100 días de leche. La información genealógica, pedigrí, de estas vacas comprende 5 generaciones con un total de 6547 animales. Toda esta información ha sido descargada desde USDA (United State Department of Agriculture) <http://www.aipl.arsusda.gov/>, 2010 y están disponibles en el conjunto de datos *milk* y *pedCows* de la librería *pedigreemm* en R. (Vazquez et al., 2010)

3 Resultados y discusión

Los estimados de las medidas de centralidad (media, mediana y moda) de las distribuciones posteriores para los componentes de variancia y heredabilidad fueron estimados utilizando la librería *MCMCglmm* del paquete estadístico R, y son presentados en el **Tabla 1**. Los resultados muestran una heredabilidad moderada de este carácter.

Tabla 1. Estimados de los componentes de varianza y heredabilidad para la producción de leche

	σ_c^2	σ_h^2	σ_e^2	h^2	c^2
Media	0,3157	0,204	0,4828	0,3154	0,20162
Mediana	0,3154	0,19076	0,4824	0,3158	0,1979
Moda	0,3199	0,1876	0,482	0,3172	0,1925

$\sigma_c^2, \sigma_h^2, \sigma_e^2, c^2$ y h^2 : varianza aditiva, varianza respecto al rebaño, varianza residual, heredabilidad aditiva y proporción de la varianza ambiental rebaño respectivamente

En la **Figura 1** se muestra las densidades

marginales para los componentes del modelo, las cuales tienden a ser simétricas (media, mediana y moda son cercanas), y tienen un comportamiento prácticamente normal. Las trazas (convergencia) muestran un comportamiento aparentemente aleatorio para estos estimados de la distribución a posteriori.

La **Figura 1** permite verificar el comportamiento del algoritmo MCMC. En las gráficas de las trazas (evolución de las muestras a través de las iteraciones), lado izquierdo, no se observa ninguna tendencia en los nueve componentes.

Las correlaciones entre sucesivas muestras son bajas en casi todos los componentes del modelo, **Tablas 3 y 4**, lo que podría indicar una fuerte convergencia de la cadena. Esto se refleja en los tamaños efectivos de la muestra no correlacionada que son altos para los componentes de interés, mayores a 1000 según lo mostrado en el **Tabla 2** que es lo mínimamente recomendado por **Hadfield (2010)**. El componente que corresponde al error presenta una autocorrelación no baja, sin

embargo, el error de Monte Carlo descrito es bajo, **Tabla 2**, pero como el error de Monte Carlo está directamente relacionado con la inversa de la longitud de la cadena (o número de iteraciones del algoritmo), en definitiva, este disminuirá cuando se aumente dicha longitud. En este estudio no se corrió los modelos con mayor longitud de cadena, puesto que los resultados fueron similares a las distribuciones posteriores de las características de interés mostrando ser bastante cercanas a la normal.

Tabla 2. Tamaño efectivo muestral (TE) y error de Monte Carlo (EMC) de las distribuciones posteriores de la varianza genética y heredabilidad para los caracteres analizados

	σ_c^2	σ_h^2	σ_e^2	h^2	c^2
TE	2136	9000	4712	3125	9000
EMC	0,000576	0,00055	0,000244	0,000491	0,000432

σ_c^2 , σ_h^2 , σ_e^2 , c^2 y h^2 : varianza aditiva, varianza respecto al rebaño, varianza residual, heredabilidad aditiva y proporción de la varianza ambiental rebaño respectivamente

Tabla 3. Autocorrelaciones de los componentes: número de lactación y logaritmo del número de días en leche

	Intercept	lact2	lact3	lact4	lact5	log(dim)
Lag 0	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Lag 10	0,01198	-0,02787	-0,00425	-0,00056	-0,00991	0,00894
Lag 50	-0,00183	0,02923	0,00780	0,02572	-0,00552	0,00601
Lag 100	-0,00058	-0,01004	-0,01298	0,00434	-0,00859	0,00202
Lag 500	-0,00336	-0,00314	-0,00982	0,00200	-0,00122	0,01191

Tabla 4. Autocorrelaciones de los componentes: animal, rebaño y error

	animal	herd	units
Lag 0	1,00000000	1,00000000	1,00000000
Lag 10	0,616322115	0,001095029	0,11517117
Lag 50	0,090100800	-0,013593962	0,03285455
Lag 100	0,022891093	-0,009120291	0,01713553
Lag 500	0,006656649	-0,004943959	0,01954783

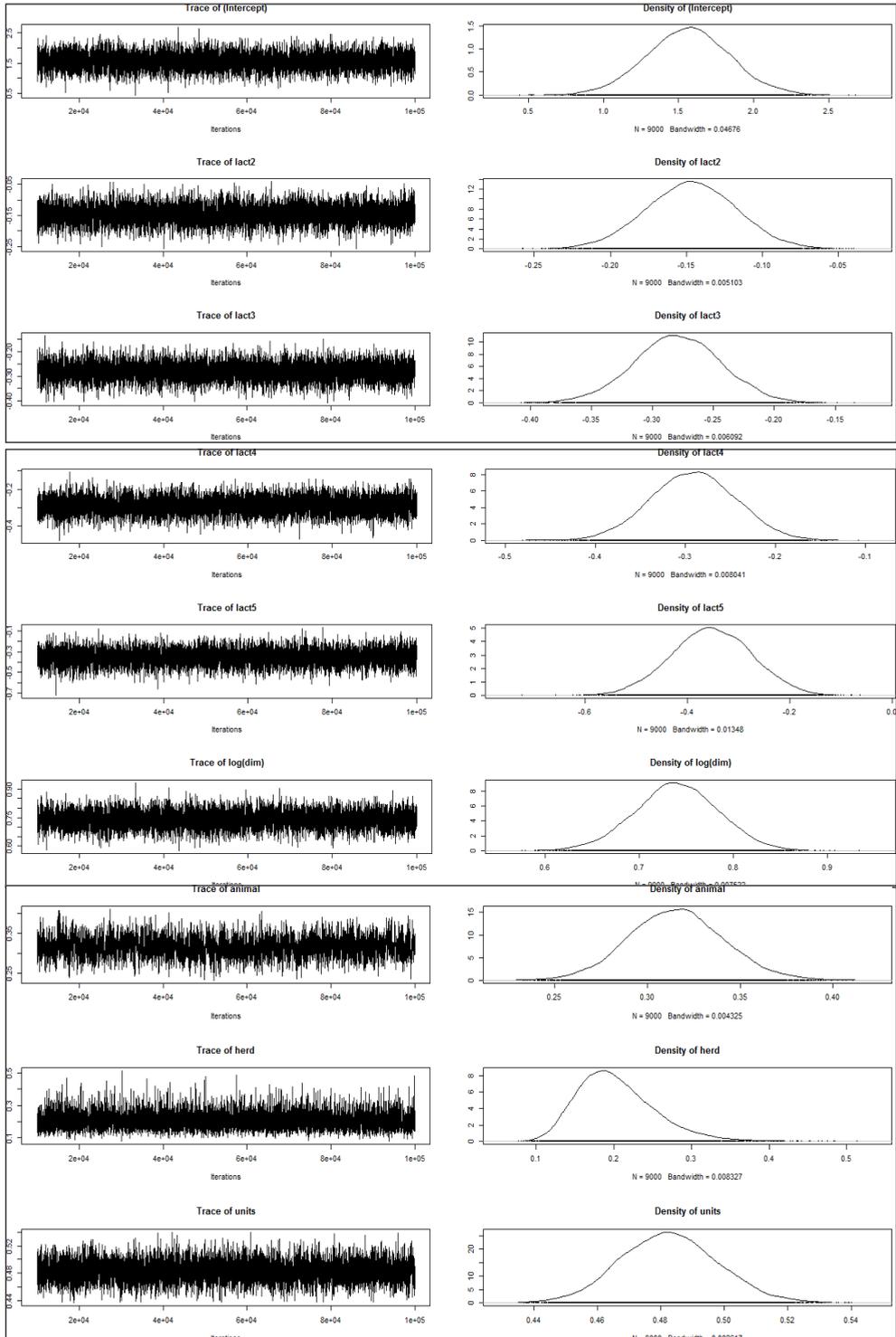


Figura 1. Evolución de los valores estreados a lo largo de las iteraciones y limaciones de las funciones de densidades a posteriori para cada componente.

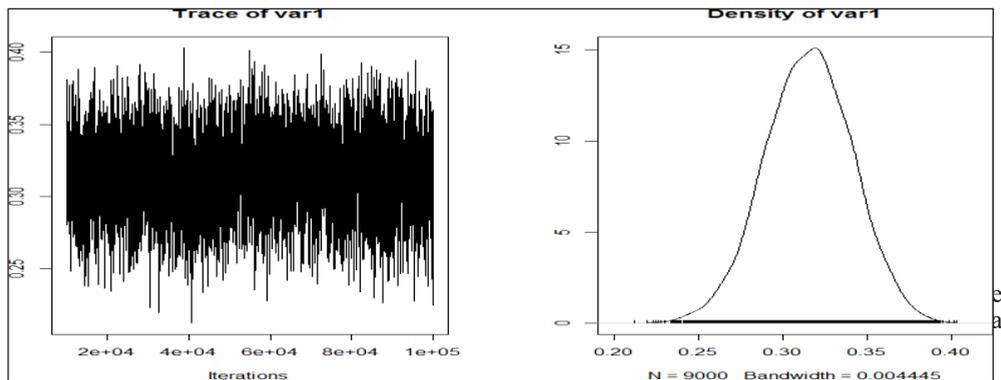


Figura 2. Evolución de los valores muestreados a lo largo de las iteraciones y las estimaciones de la función de densidad a posteriori para la heredabilidad

4. Conclusiones

En el diagnóstico del modelo vía estimación bayesiana, no se encontraron problemas de convergencia de la cadena. Se obtuvieron errores de Montecarlo bajos y tamaños efectivos de muestra mayores a 1000 para cada componente del modelo. La media estimada de la heredabilidad vía muestreo de Gibbs fue de 0.3154 para la producción de leche, la cual es moderada, lo que respalda la idea de que la variabilidad fenotípica de esta característica (producción de leche) está explicada en aproximadamente el 32% por la acción genética aditiva y el resto por otros factores.

5. Literatura citada

Blasco, A. 2001. The Bayesian controversy in animal inbreeding. *Journal of Animal Science* 79(8): 2023-2046.

Hadfield, JD. 2010. MCMC Methods for Multi-response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, 33(2): 1-22. Consultado 13 julio 2014. Disponible en: <https://www.jstatsoft.org/article/view/v033i02/v33i02.pdf>

Mrode, R. 2014. *Linear Model for the Prediction of Animal Breeding Value*. 3 ed. Edinburgh, UK, CABI. 343 p.

Sorensen, D.; Gianola, D. 2002. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. New York. Estados Unidos, Springer-Verlag. 740 p

Vazquez, AI.; Bates, D.; Rosa GJ., Gianola, D.; Weigel, KA. 2010. Technical note: An R package for fitting generalized linear mixed models in animal breeding1. *Journal of Animal Science* 88(2): 497-504.

Wang, D.; Rutledge, J.; Gianola, D. 1993. Marginal inferences about variance components in a mixed linear models using Gibbs sampling. *Genetics Selection Evolution*, 25(1):41-62.

Wang, D.; Rutledge, J.; Gianola, D. 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics Selection Evolution*, BioMed Central 26(2): 91-115.