ITEA (1998), Vol. **94A N.° 3,** 305-315

ESTIMA DE PARÁMETROS GENÉTICOS PARA POBLACIONES DE VACUNO LECHERO MEDIANTE METODOLOGÍA BAYESIANA

D. Hernández[1]
M.J. Carabaño[1]
R. Rekaya[2]

1. Departamento de Mejora Genética y Biotecnología
INIA. Apartado 8111
Madrid 28080
España

2. Departament of Animal Sciences
University of Wisconsin
1675 Observatory Dr.
Madison, Wisconsin 53706-1284
U.S.A.

RESUMEN

Dentro de este trabajo se plantea la actualización de estimas de componentes de varianza para caracteres de producción de leche y proteína en la población Frisona Española utilizando procedimientos desarrollados recientemente que permiten el uso de bases de datos de gran tamaño. Con ello se pretende evitar la utilización de información parcial en la estimación de los parámetros genéticos con respecto a la utilizada en las valoraciones genéticas. En el estudio se han usado los datos empleados para la valoración genética nacional de enero de 1998, un total de 1.460.559 datos de producción de leche y 1.084.886 de producción de proteína. La información genealógica incluyó un total de 703.326 y 589.385 animales para leche y proteína, respectivamente. En la estimación de componentes de varianza se empleó el mismo modelo utilizado en la valoración genética nacional, siendo éste un modelo animal con medidas repetidas. Se utilizó el muestreo de Gibbs dentro de un análisis bayesiano para hacer inferencias sobre los parámetros de interés. La media a posteriori de la heredabilidad fue de 0,23 y 0,22 para producción de leche y proteína, respectivamente. La media a posteriori de la repetibilidad fue de 0,46 para ambos caracteres. Las estimas de heredabilidad son ligeramente más bajas que las obtenidas previamente en la misma población con muestras más pequeñas de datos y metodología de máxima verosimilitud e inferiores también al valor actualmente empleado en la valoración nacional. Las causas de estos valores relativamente bajos deben ser investigadas antes de proponer su uso en las valoraciones nacionales.

Palabras claves: Parámetros genéticos, Ganado vacuno lechero, Muestreo de Gibbs, Metodología bayesiana.

SUMMARY

ESTIMATION OF GENETIC PARAMETERS IN DAIRY CATTLE POPULATIONS USING BAYESIAN METHODS

This work aimed at updating the estimates of genetic parameters for milk and protein yield in the Spanish Friesian population by making use of recently developed procedures that allow the utilization of massive data sets. This avoids the use of partial information when estimating variance components with respect to the information used in genetic evaluations. Data used in this study were 1,460,559 and 1,084,886 records from the January, 1998 national genetic evaluation for milk and protein yields. Pedigrees for the milk and protein analyses consisted of 703,326 and 589,385 animals, respectively. Variance components estimation was carried out under the same model used for the genetic evaluation, a repeatability animal model. Gibbs sampling within a Bayesian analysis was used to make inferences about the parameters of interest. The posterior mean for the heritability was 0.23 for milk yield and 0.22 for protein yield. The posterior mean for the repeatability for both traits was 0.46. Heritability estimates were slightly lower than previous estimates obtained in the same population with maximum likelihood procedures and also lower than the value currently used in the genetic evaluations. Additional research is needed to investigate the reasons of the relatively low heritability estimates obtained previous to propose its use in the national evaluations.

Key words: Genetic parameters, Dairy cattle, Gibbs sampling, Bayesian analysis.

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 proce-

dentes 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:

$$\mathbf{y_{ijklmn} = \mu + DA_i + M_j + LE_k + RAE_l + + A_m + P_m + e_{ijklmn}}$$

Donde:

y_{ijklmn} = Producción de leche o proteina expresada en kilogramos.

μ = Media general

DA_i = Días abiertos

M_j = Mes de parto

LE_k = Lactación-Edad

RAE_l = Grupo móvil de comparación Rebaño-Año-Estación.

A_m = Efecto aditivo del animal que produce el dato.

P_m = Efecto permanente ambiental del animal que produce el dato.

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:

$\mathbf{y|b,u,p}, \sigma_e^2 \sim N(\mathbf{Xb + Zu + Wp}, \mathbf{I}\sigma_e^2)$

$\mathbf{b} \sim$ constante

$\mathbf{u}|\ \sigma_a^2 \sim N(0, \mathbf{A}\sigma_a^2)$

$\mathbf{p}|\ \sigma_p^2 \sim N(0, \mathbf{I}\sigma_p^2)$

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 4.000 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 1.000 muestras independientes para cada carácter.

Resultados y Discusión

Las Figuras 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 el cuadro 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

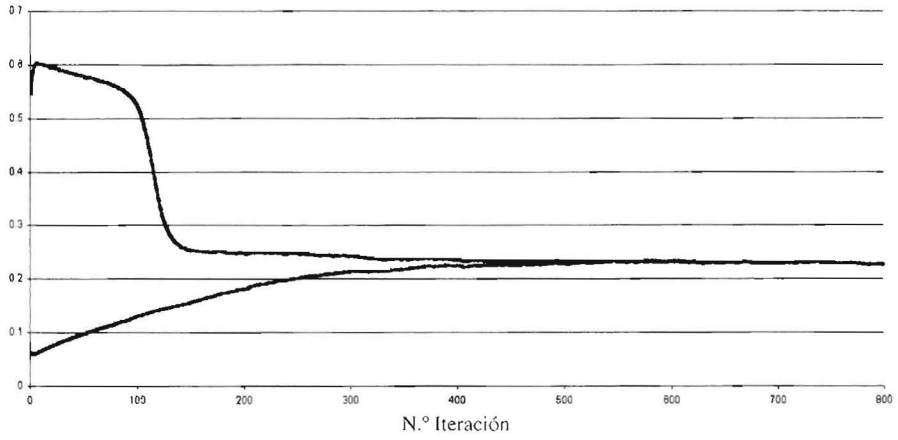


Figura 1. Valores estimados de la Heredabilidad de kg de leche mediante cadenas pareadas

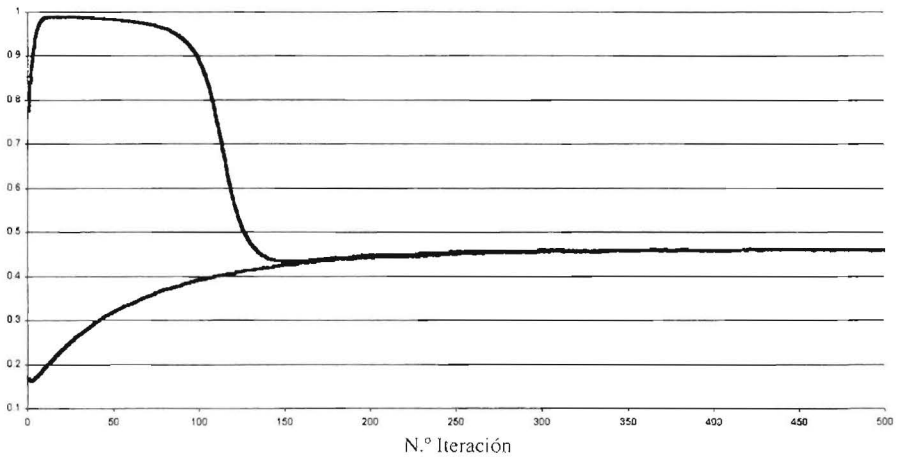


Figura 2. Valores estimados de la Repetibilidad de kg de leche mediante cadenas pareadas

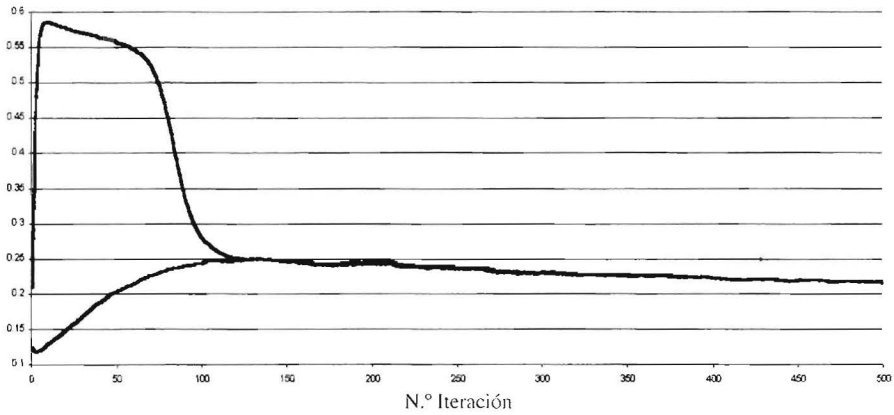


Figura 3. Valores estimados de la Heredabilidad de kg de proteína mediante cadenas pareadas

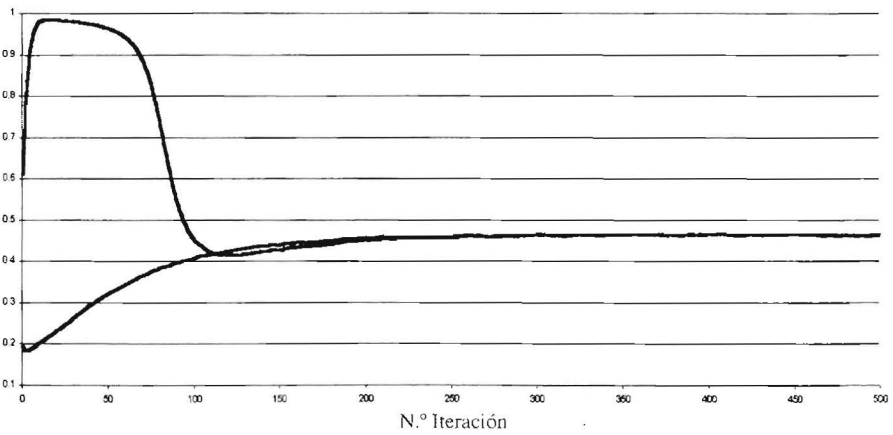


Figura 4. Valores estimados de la repetibilidad de kg de proteína mediante cadenas pareadas

CUADRO 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	30,50
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	18,60

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 este cuadro y asumiendo que la proporción de muestras independientes se mantiene constante, el número total de muestras requerido para obtener 1.000 muestras independientes en todos los parámetros sería de 166.667 y 52.631 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 figuras 5,6,7 y 8 se muestran las densidades marginales posteriores de las heredabilidades y repetibilidades para leche y proteina respectivamente.

En los cuadros 2 y 3 se muestran los estadísticos de tendencia central, así como la varianza correspondiente a las distribuciones marginales posteriores de las varianzas (σ^2_a, σ^2_p y σ^2_e), 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 es-

CUADRO 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

CUADRO 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

pecialmente 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.

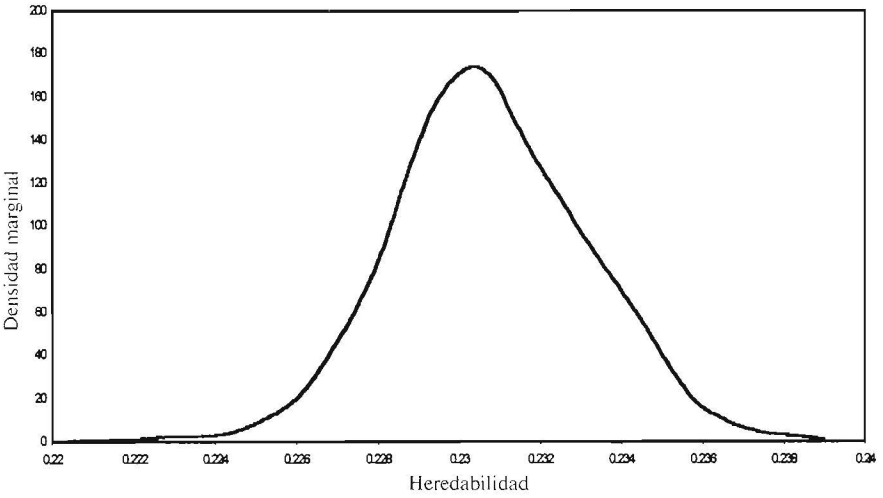


Figura 5. Densidad marginal posterior de la heredabilidad de kg de leche

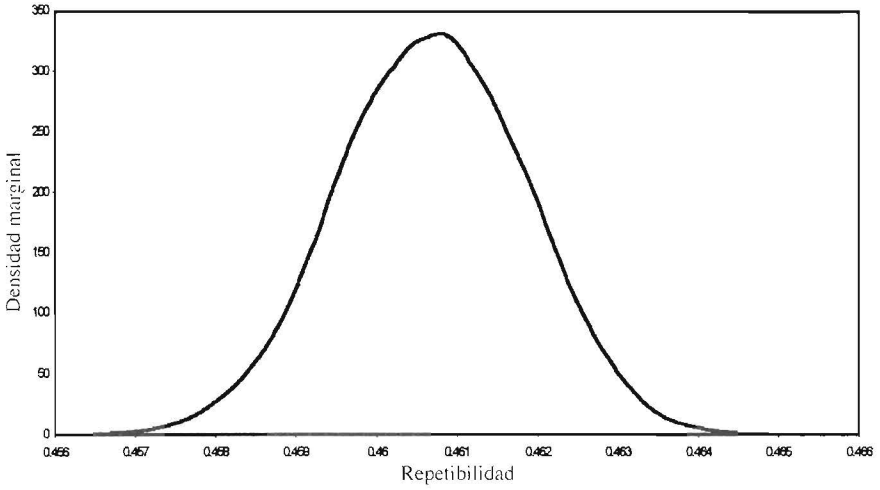


Figura 6. Densidad marginal posterior de la repetibilidad de kg de leche

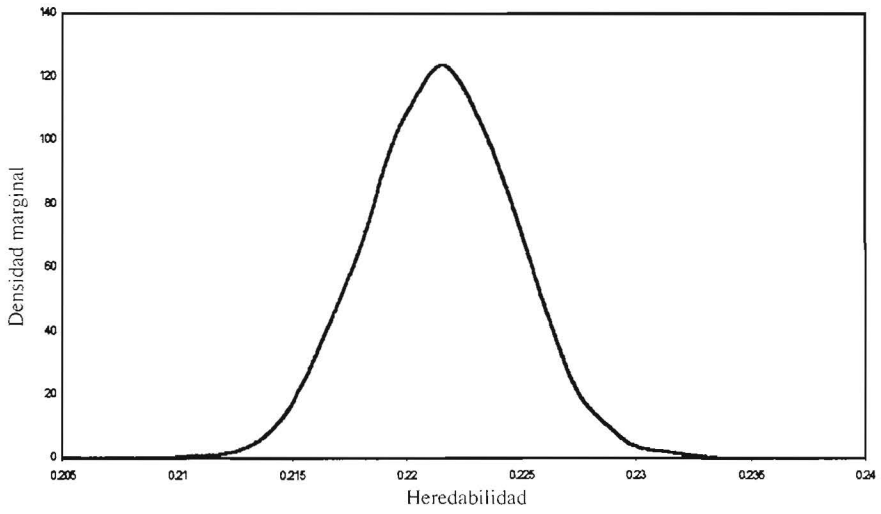


Figura 7. Densidad marginal posterior de la Heredabilidad de kg de proteína

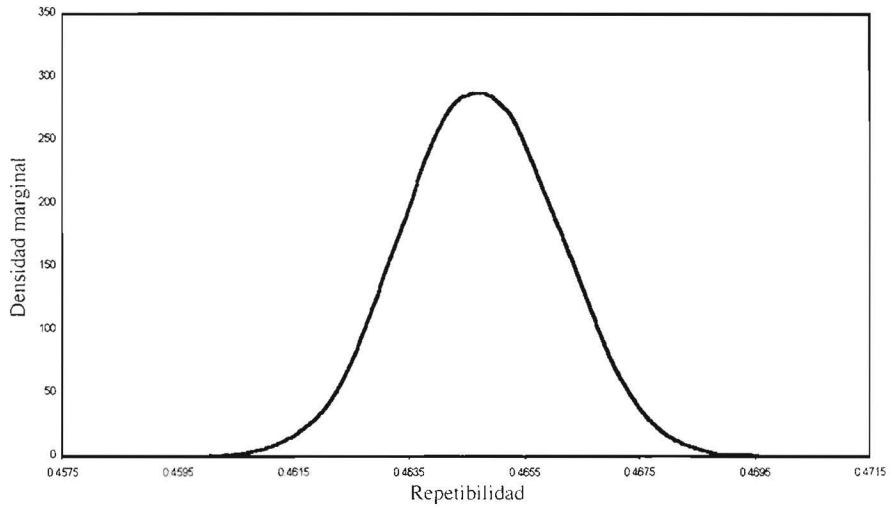


Figura 8. Densidad marginal posterior de la repetibilidad de kg de proteína

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

CHARFEDDINE 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., GOLDEN B.L., BOURDON R.M., 1994. Method R variance component procedure: application on the simple breeding value model. J. Anim. Sci. 72, 2.247-2.253.

SILVERMAN B.W., 1986. Density estimation for statistic and data analysis. Chapman and Hall Ed.

VAN TASSELL C.P., WIGGANS G.R., NORMAN H.D. y POWELL R.L., Estimation of heritability for yield of U.S. dairy cattle. Proc. of the 1997 Interbull Meeting. Vienna, Austria. Bulletin 16,104-107.