

# ANÁLISIS BAYESIANO DE LA HETEROGENEIDAD DE LA VARIANZA RESIDUAL DURANTE LA LACTACION UTILIZANDO LA TECNICA DE LOS PUNTOS DE CAMBIO

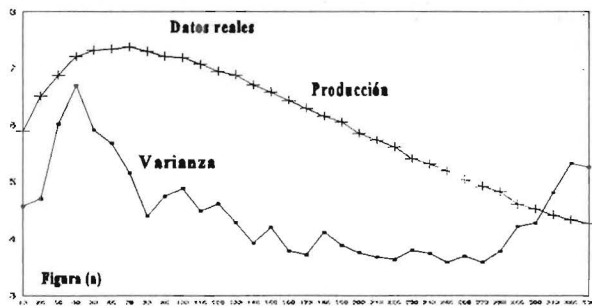
R. Rekaya; M. J. Carabaño y M. A. Toro

Departamento de Producción Animal CIT- INIA

## I/ INTRODUCCION

Problemas en los que interesa identificar puntos de cambio surgen en muchas aplicaciones prácticas especialmente aquellas en los que los datos se han recogido de forma cronológica (modelos multifásicos de crecimiento, producción de leche, etc) y se han desarrollado muchas técnicas estadísticas que permiten identificar los puntos de cambio en una secuencia de datos y como consecuencia los posibles cambios en los parámetros del modelo asumido (Stephens, 1994).

Para modelos no-lineales, con varios puntos de cambio y un gran volumen de datos los problemas computacionales suelen ser prohibitivos y sólo recientemente han podido ser abordados. El análisis bayesiano utilizando el muestreo de Gibbs y algoritmos que permiten muestrear de distribuciones condicionales no estándar es quizás la metodología más prometedora. Esta metodología es la que hemos aplicado al estudio de la heterogeneidad de la varianza residual en las diferentes fases de la curva de lactación, fenómeno que ha sido demostrado en muchos trabajos de análisis de controles diarios ( Schaeffer et al., 1994; Jamrozik et al., 1996). En un estudio de la persistencia con datos españoles (130601 controles diarios) hemos detectado un aumento de la varianza residual a partir del inicio de la lactación hasta los días 40-50, momento en que alcanza un valor máximo y después empieza a disminuir, primero de una manera brusca hasta los días 80 -90 y después de forma más atenuada. A partir de los días 230-240 se observa una aumento importante hasta alcanzar valores parecidos a los encontrados en la primera fase (figura (a)).



## II/ METODOS

### 1. Modelo

Se ha adoptado un modelo bayesiano jerárquico con tres puntos de cambio. En el primer nivel de la jerarquía se asume que la distribución de las observaciones en cada intervalo de la secuencia es

$$Y|\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \sigma_{e1}^2, \sigma_{e2}^2, \sigma_{e3}^2, \sigma_{e4}^2 \sim \begin{matrix} N(X\beta + \eta(\theta), t^{\lambda_1} \sigma_{e1}^2) & i=1, \dots, T_1 \\ N(X\beta + \eta(\theta), t^{\lambda_2} \sigma_{e2}^2) & i=T_1+1, \dots, T_2 \\ N(X\beta + \eta(\theta), \eta(\theta)^{\lambda_3} \sigma_{e3}^2) & i=T_2+1, \dots, T_3 \\ N(X\beta + \eta(\theta), \eta(\theta)^{\lambda_4} \sigma_{e4}^2) & i=T_3+1, \dots, t_{\max} \end{matrix}$$

$\eta(\theta)$  es la parte no lineal del modelo descrita por la fórmula de Wood (1967) y cuya expresión en el día  $t$  es,  $\eta_t(\theta) = a_1 t^{a_2} \exp(-a_3 t)$ .  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  son unos exponentes que relacionan la varianza residual en cada tramo de la lactación con el día del control ( $\lambda_1$  y  $\lambda_2$ ) o con la media de producción calculada a partir de la curva de la lactación ( $\lambda_3$  y  $\lambda_4$ ).

En el segundo estrato de la jerarquía se especifican las distribuciones a priori de los parámetros del modelo definidos en el primer estrato.

$$\beta \sim \beta_{\min} < \beta < \beta_{\max}$$

$$T \sim 1 < T < t_{\max}$$

$$\sigma_{e1}^2 \sim \chi^2(v_e, v_e S_e^2)$$

$$a | \Sigma_0 \sim NMV(m, I \otimes \Sigma_0)$$

$\Sigma_0$  es la matriz de (co)varianzas residuales entre los parámetros de la función de Wood y  $m = Wb + Zu$ , la media de estos parámetros, siendo  $b$  el efecto lactación-edad-época de parto y  $u$  el valor genético aditivo de cada animal para los parámetros de la curva de la lactación.  $W$  y  $Z$  son las correspondientes matrices de incidencias.

El tercer estrato de la jerarquía consiste en asignar distribuciones a  $b$ ,  $u$  y  $\Sigma_0$  definidos en el estrato anterior.

$$b \sim b_{\min} < b < b_{\max}$$

$$u | \Sigma_g \sim NMV(0, A \otimes \Sigma_g)$$

$$\Sigma_0 | v_0 S_0^2 \sim W^{-1}(v_0, v_0 S_0^2)$$

La estructura jerarquizada termina asignando una distribución a priori para la matriz de (co)varianzas genéticas  $\Sigma_g$ .

$$\Sigma_g | v_g S_g^2 \sim W^{-1}(v_g, v_g S_g^2)$$

$S_0^2$  y  $S_g^2$  son matrices  $3 \times 3$  de hiperparámetros conocidos para  $\Sigma_0$  y  $\Sigma_g$ , respectivamente.

## 2. Inferencia bayesiana

Para implementar el algoritmo de Gibbs sampling se requieren las distribuciones condicionales de todos los parámetros desconocidos del modelo. Estas distribuciones condicionales son normales para  $\beta$ ,  $b$  y  $u$ , chi-cuadrado invertida para  $\sigma_{e1}^2$  y Wishart invertida para las matrices  $\Sigma_g$ ,  $\Sigma_0$ . Para los demás parámetros del modelo ( $T_1, T_2, T_3, \lambda_1, \lambda_2, \lambda_3, \lambda_4, a_1, a_2, a_3$ ) las distribuciones condicionales no son formas estándar por lo que el muestreo se ha realizado utilizando el algoritmo "adaptive rejection metropolis sampling" (Gilks et al., 1995)

## 3. Simulación

Para contrastar esta metodología, se ha aplicado inicialmente a datos simulados con la misma estructura de los datos reales y con valores conocidos de los parámetros. La relación entre la media

de producción y la varianza residual para los datos simulados se indica en la figura (b).

### III/ RESULTADOS Y DISCUSION

En la tabla 1 se presenta un sumario de las distribuciones marginales de los puntos de cambio (fechas de cambio) y de los exponentes que relaciona la producción o los días en lactación y la varianza residual. Se observa un ajuste razonablemente satisfactorio entre las modas a posteriori y los valores verdaderos y éstos están en todos los casos en el intervalo del 95% de HPD. La figura (c) que representa la relación entre la media de producción o los días en lactación y la varianza residual confirma la adecuación de la metodología utilizada.

Queda por aplicar esta metodología a los datos reales de vacuno lechero así como analizar su posible impacto en el análisis de los controles diarios y la evaluación genética de la producción de leche y de la persistencia de la curva de lactación.

#### REFERENCIAS BIBLIOGRAFICAS.

Gilks et al., 1995. Appl. Statist. 4:455-472.

Jamrozik et al., 1996. Interbull Open Meeting, Veldhoven, The Netherlands, June 23-24. 1996.

Schaeffer et al, 1994. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Vol. XVIII:443

Stephens, 1994. Appl Statist. 1:159-178.

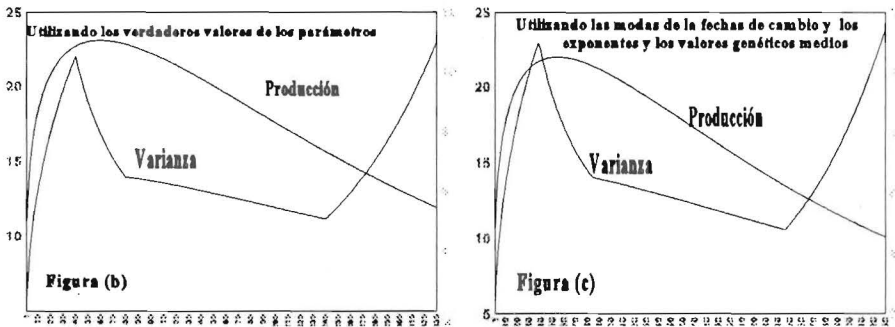


Tabla 1: Sumario de las distribuciones marginales de las fechas de cambio y de las exponentes

Parámetro	Verdadero	Media	Moda	Mediana	HPD
$\sigma_{e1}^2$	2.40	2.351	2.372	2.346	2.17 - 2.572
<b>Fecha de cambio</b>					
$T_1$	40	38.54	39.13	38.92	29.87 - 48.44
$T_2$	80	84.76	81.27	82.80	67.29 - 103.19
$T_3$	240	246.42	248.25	248.97	226.74 - 258.48
<b>Exponentes</b>					
$\lambda_1$	0.40	0.423	0.413	0.418	0.298 - 0.544
$\lambda_2$	-0.70	-0.671	-0.694	-0.683	-0.792 --0.492
$\lambda_3$	0.60	0.624	0.618	0.634	0.508 - 0.767
$\lambda_4$	-2.50	-2.675	-2.651	-2.686	-2.934 - -2.312