MODELO RECURSIVO CON PUNTOS DE CAMBIO. UNA APLICACIÓN AL TAMAÑO DE CAMADA Y EL NÚMERO DE LECHONES NACIDOS MUERTOS

Ibáñez- Escriche, N[1]., López de Maturana, E[2]., Noguera, J.L.[1] y Varona, L.[3]
[1]Genètica i Millora Animal, IRTA-Lleida, 25198 Lleida
[2]Dpto. de Mejora Genética Animal, INIA, 28040 Madrid
[3]Dpto. de Anatomía, Embriología y Genética Animal, Universidad de Zaragoza, 50013 Zaragoza.
Noelia.ibanez@irta.es

INTRODUCCIÓN

En los últimos años, el interés por el estudio de las relaciones entre caracteres productivos a nivel fenotípico, en el contexto de la genética cuantitativa, ha aumentado gracias a la publicación del trabajo de Gianola y Sorensen (2004). En este trabajo, se describen los procedimientos estadísticos para la resolución de estos modelos de ecuaciones estructurales que tienen una larga tradición en el ámbito de las ciencias sociales.
Recientemente, López de Maturana et al. (2009) han extendido la aplicación de estos modelos recursivos a relaciones no lineales entre fenotipos, a través de una regresión segmentada. Esta aproximación permitió determinar la existencia de correlaciones heterogéneas tanto entre efectos genéticos, como entre permanentes y entre los residuos. Sin embargo, presenta dos desventajas: la predeterminación de los puntos de cambio, que definen la no linealidad, y la discontinuidad en dichos puntos de cambio. El objetivo de este trabajo es proponer un modelo recursivo que permita estimar estos puntos de cambio, y describir una aplicación, en el estudio de la relación no lineal entre el tamaño de camada (número total de lechones nacidos) y el número de lechones nacidos muertos.

MATERIAL Y MÉTODOS

Material Animal

Los individuos de la población analizada pertenecían a un núcleo de selección, de raza *Large White*, seleccionado por tamaño de camada. Los datos analizados corresponden al número total de lechones nacidos (TC) y al número de lechones nacidos muertos (NM) de 4.462 partos producidos por 1070 cerdas y registrados durante el período 1999-2006.

Análisis estadístico

En el análisis se estudiaron los posibles efectos recursivos del tamaño de camada sobre el número de lechones nacidos muertos. Para analizar la posible relación no lineal entre estos caracteres, se utilizó un modelo recursivo con coeficientes estructurales heterogéneos. Tras un análisis exploratorio, se consideró la necesidad de estimar dos puntos de cambio, que distinguen tres grupos de TC.

Asumiendo una distribución conjunta normal multivariante para TC y NM, los datos de un individuo *i* perteneciente al grupo *k* fue modelizado como

$$\mathbf{\Lambda}_k \mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{a} + \mathbf{W}_i\mathbf{p} + \mathbf{e}_i \ , \quad (1)$$

donde

$$\mathbf{y}_i \mid \mathbf{\Lambda}_k, \mathbf{T}_m, \mathbf{b}, \mathbf{a}, \mathbf{p}, \mathbf{R}_0 \sim (\mathbf{\Lambda}^{-1}{}_k(\mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{a} + \mathbf{W}_i\mathbf{p}), \mathbf{\Lambda}^{-1}{}_k\mathbf{R}_0\mathbf{\Lambda}^{-1}{}_k). \quad (2)$$

En este caso, $\mathbf{\Lambda}_k$ (k=1,2,3) es la matriz estructural de coeficientes correspondiente al subgrupo *k* de tamaño de camada, con la forma

$$\mathbf{\Lambda}_k = \begin{bmatrix} 1 & 0 \\ -\lambda_{NM \leftarrow TC} & 1 \end{bmatrix} \quad (3)$$

$\mathbf{T}_m$ es el vector de puntos de cambio y *m* es el número de puntos de cambio. En nuestro caso *m=2* por lo que $\mathbf{\Lambda}_k = \mathbf{\Lambda}_1 \ si \ y_{TC_i} \le T_1$, $\mathbf{\Lambda}_k = \mathbf{\Lambda}_2 \ si \ T_1 < y_{TC_i} \le T_2$ y $\mathbf{\Lambda}_k = \mathbf{\Lambda}_3 \ si \ y_{TC_i} > T_2$.

En la ecuación 2, $\mathbf{y}_i$ es un vector *3 x 1* que contiene los datos de los caracteres TC y NM; **b** es el vector de los efectos sistemáticos (orden de parto, 6 niveles, y la combinación entre el año y la estación, con 31 niveles); **a** es el vector de los efectos genéticos aditivos (1534); **p** es el vector de los efectos permanentes (1070); y **e** es el vector de los efectos residuales. $\mathbf{X_i}$, $\mathbf{Z_i}$, $\mathbf{W_i}$ son matrices de incidencia conocida.

Implementación de MCMC y su análisis posterior.

Para simplificar llamamos $\theta = \mathbf{\Lambda}_k, \mathbf{T}_m, \mathbf{b}, \mathbf{a}, \mathbf{p}, \mathbf{G}_0, \mathbf{P}_0, \mathbf{R}_0$, por lo que la densidad posterior conjunta de θ y de NM viene dada por

$$p(\mathbf{y}_{NM}, \theta \mid \mathbf{y}_{TC}) \propto p(\mathbf{y}_{NM}, \mathbf{y}_{TC}, \theta) = p(\mathbf{y}_{NM}, \mathbf{y}_{TC} \mid \theta) \times p(\theta).$$

Donde, $p(\theta) = p(\mathbf{\Lambda}_k \mid \mathbf{T}_m) p(\mathbf{T}_m) p(\mathbf{b}) p(\mathbf{a} \mid \mathbf{G}_0) p(\mathbf{G}_0) p(\mathbf{p} \mid \mathbf{G}_{0p}) p(\mathbf{P}_0) p(\mathbf{R}_0)$.

Como distribuciones *a priori*, se asignaron normales multivariantes para $\lambda_k \sim N(\mathbf{0}, I\sigma^2_\lambda)$, donde *k=1,2*, y σ^2_λ=100; $\mathbf{b} \sim N(\mathbf{0}, I\sigma^2_\beta)$, con σ^2_β=100; efectos permanentes, $\mathbf{p} \sim N(\mathbf{0}, \mathbf{P}_0 \otimes \mathbf{I})$; y efectos aditivos, $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G}_0 \otimes \mathbf{A})$. Mientras a los puntos de cambio $\mathbf{T}_m$ se les asignó una distribución a priori uniforme $\mathbf{T} \sim Unif[y_{TC\min}, y_{TC\max}]$. Se asumió *a priori* que los parámetros de dispersión se distribuyen siguiendo *Wishart* invertidas con 4 grados de libertad. Para evitar problemas de identificación de parámetros, se asumieron que los errores de los caracteres no estaban correlacionados, por lo que $\mathbf{R_0}$ se asumió diagonal. La inferencia de los parámetros se realizó aplicando el algoritmo de *Gibbs*, excepto en el caso de los puntos de cambio $\mathbf{T}_m$ cuya distribución posterior no es conocida y se muestreó con *Metropolis Hasting*.

Para que los parámetros de los modelos recursivos puedan ser interpretables, como en los modelos mixtos estándar, las matrices de (co)varianza ($\mathbf{G}_0$, $\mathbf{P}_0$ and $\mathbf{R}_0$) necesitan ser ajustadas por la matriz estructural de coeficientes $\mathbf{\Lambda}_k$: $\mathbf{G}^*_k = \mathbf{\Lambda}^{-1}_k \mathbf{G}_0 \mathbf{\Lambda}^{'-1}_k, \mathbf{P}^*_k = \mathbf{\Lambda}^{-1}_k \mathbf{P}_0 \mathbf{\Lambda}^{'-1}_k y\ \mathbf{R}^*_k = \mathbf{\Lambda}^{-1}_k \mathbf{R}_0 \mathbf{\Lambda}^{'-1}_k$ (Sorensen y Gianola 2002).

Después de varios análisis exploratorios se utilizó una cadena de 100.000 iteraciones, con un periodo de "quemado" de 1.000 iteraciones. Cada 10 iteraciones se tomó una muestra, para evitar la alta correlación entre muestras consecutivas.

RESULTADOS Y DISCUSIÓN

Los parámetros de las distribuciones marginales posteriores estimadas de los puntos de cambio, coeficientes estructurales, varianzas, correlaciones y heredabilidades se presentan en las Tablas 1 y 2. Las distribuciones marginales posteriores tuvieron un error de Monte Carlo pequeño y el test de Geweke (Geweke 1992) no detectó falta de convergencia (resultados no mostrados). Como se puede observar en la Tabla 1, los intervalos mínimos de máxima densidad al 95% (IMD95%) de los puntos de cambio y los coeficientes estructurales difieren claramente entre si. Además, los IMD95% de los coeficientes estructurales no contienen el cero. Estos resultados confirmarían la relación no lineal entre TC y NM e indicarían que el número de lechones muertos al parto se incrementa un 0,13, 0,16, y 0,21 en cada subgrupo, respectivamente, por cada lechón nacido adicional.

La media posterior de la heredabilidad para TC fue de 0.09 con un IMD95% de 0.04-0.13, similar a lo publicado por Noguera et al., (2002). A excepción de la correlación del error, la distribución posterior de la heredabilidad para NM y las de las correlaciones genéticas y permanentes entre TC y NM fueron similares para los tres subgrupos (ver tabla 2). Las medias posteriores de las heredabilidades fueron bajas y las correlaciones genéticas y

permanentes cercanas a cero, coincidiendo con los resultados publicados por Ibáñez-Escriche et al. (2009) con un modelo mixto estándar.

Los resultados obtenidos en este análisis mostrarían la utilidad de los modelos recursivos con coeficientes estructurales heterogéneos y estimación de puntos de cambio. No obstante, sería interesante realizar una comparación de modelos con diferente número de puntos de cambio a través de una validación cruzada de los datos.

REFERENCIAS BIBLIOGRÁFICAS

• Geweke, J. 1992. Bayesian Statistics 4, 169-193. • Gianola, D. & Sorensen, D. 2004. Genetics 167: 1407-1424. • Ibáñez- Escriche, N., Varona, L., Quintanilla, R., Casellas, J. & Noguera, J.L. 2009. J. Anim. Sci. 2009.87:80-87. • López de Maturana, E., X.-L. Wu, D. Gianola , K. A., Weigel & G. J. M., Rosa. 2009. Genetics 181:277-287. • Noguera, J.L., Varona, L., Babot, D. & Estany, J. 2002. J. Anim. Sci. 80: 2548-2555.

Tabla 1. *Parámetros de las distribuciones marginales posteriores de los puntos de cambio (T) y los coeficientes estructurales (λ).*

	T_1	T_2	λ_1	λ_2	λ_3
Media	15,30	19,71	0,13	0,16	0,21
IMD95%	14,94 -16,20	17,15 - 21,20	0,13 - 0,14	0,15 - 0,17	0,18 - 0,26

IMD: Intervalo mínimo de máxima densidad al 95%.

Tabla 2. *Parámetros de las distribuciones marginales posteriores de la heredabilidad de los lechones nacidos muertos y de las correlaciones entre tamaño de camada y lechones nacidos muertos para cada subgrupo del tamaño de camada.*

	Grupos de NT					
	(λ_1)		(λ_2)		(λ_3)	
	Media	IMD95%	Media	IMD95%	Media	IMD95%
h^2_{NM}	0,07	0,02 - 0,13	0,07	0,02 -0,12	0,07	0,02 - 012
$\rho_{gTC,\ gNM}$	-0, 02	-0,25 - 0,24	0,01	-0,22 - 0.27	0,07	-0,18 - 0.33
$\rho_{pTC,pNM}$	0,14	-0,08 - 0,37	0,16	-0,07 - 0,39	0,21	-0,02 - 0,46
$\rho_{eTC,eNM}$	1,40	1,23 – 1,54	1,65	1,50-1,79	2,21	1,77 – 2,66

IMD: Intervalo mínimo de máxima densidad al 95%. h^2_{NM}: heredabilidad de los lechones nacidos muertos. $\rho_{g\ TC,gNM}$, $\rho_{pTC,pNM}$ y $\rho_{eTC,eNM}$ correlación genética, permanente y del error entre TC y NM, respectivamente.

A CHANGE-POINT RECURSIVE MODEL. AN APPLICATION TO DATA OF LITTER SIZE AND PIGLETS BORN DEAD.

ABSTRACT: The emergence of statistical methods for the analysis of relationships between phenotypes in a quantitative genetics scenario, stimulated several studies analyzing recursive relationships in the animal breeding and genetics context. Only one of those studies has analyzed nonlinear relationships between phenotypes, through the implementation of a segmented regression. However, this approach assumed fixed change points, with loss of analysis flexibility. A change-point recursive model and its application in litter size (LS) and number of piglets born dead (NBD) is presented in this paper. Field data were provided by a Large White selection nucleus of a commercial breeding company. Field data contained litter size and number of born dead piglets from 4462 farrows. The results revealed the existence of at least 2 change points and different structural coefficients between the change points. These would indicate a non linear relationship between LS and NBD and it would support the change-point recursive model. Nevertheless, a model comparison with different number of change points would be interesting.
Keywords*: Recursive model, non linearity relationship, change-points, prolificacy.*