TECNICAS ESTADISTICAS DE VALORACION DE REPRODUCTORES

J Altarriba, LA García Cortés, C Moreno, L Varona y G Yagüe
Unidad de mejora genética. Facultad de Veterinaria
Miguel Servet 177. 50013-Zaragoza
e-mail: altarrib@posta.unizar.es

Modelos lineales

La estimación del valor genético de los candidatos a la selección se ha venido realizando durante la última decada a partir de las ecuaciones de modelo mixto de Henderson (BLUP). Es bien conocido que, al igual que el índice clásico de selección, este procedimiento predice los valores genéticos a partir de una transformación lineal de los datos. Sin embargo, el método de Henderson se ha adoptado debido a que las asunciones que el mejorador debe hacer sobre la distribución de los datos afectan solamente a su varianza, en tanto que los procedimientos anteriores debían realizar también asunciones sobre las medias de los datos.

Si bien el método de Henderson supone un avance respecto al índice clásico de selección, ambos proporcionan ordenaciones de los animales que son susceptibles de modificación según los componentes de varianza asumidos por el mejorador. Los componentes de varianza verdaderos pueden ser reemplazados por estimadores con buenas propiedades estadísticas (REML), y los predictores BLUP así obtenidos, aunque se carezca de un análisis que garantice sus propiedades estadísticas, son útiles a efectos prácticos. No debe olvidarse que los estimadores REML han de ser suficientemente precisos para que las ordenaciones de los animales no se vean sustancialmente alteradas.

Así, pueden cometerse errores importantes si se utiliza esta técnica en doble etapa REML-BLUP para analizar experimentos de selección de tamaño reducido, poblaciones autóctonas de tamaño pequeño, etc. Sin embargo, para analizar datos de poblaciones grandes este procedimiento dará buenos resultados.

Por otro lado, asumir como verdaderos los componentes de varianza REML lleva a una reducción aparente de los errores de predicción de los valores genéticos, lo que puede traducirse en una falsa y peligrosa sensación de seguridad cuando se aborden planes de selección sobre poblaciones pequeñas. No olvidemos que lo mismo ocurre cuando se comparan los errores de predicción del índice clásico frente al índice BLUP. El error de predicción del índice clásico es menor que el de BLUP debido a que asume más información conocida.

Siguiendo el mismo argumento, predecir los valores genéticos por procedimientos que no asumen conocidos los componentes de varianza proporcionará errores de predicción mayores que BLUP, pero ello sólo será debido al mayor número de asunciones en BLUP y no a una mayor eficiencia estadística del método. Por estos motivos, la inferencia bayesiana, al permitir análisis en los que los componentes de varianza son estimados conjuntamente con los efectos fijos y aleatorios, presentará mejores propiedades que REML-BLUP, y será tanto mejor cuanto menor sea la información con la que se estimen los componentes de varianza.

Los estimadores de Monte Carlo

Debido a la dificultad de cálculo que presenta la integración multidimensional en la inferencia bayesiana. Las distribuciones marginales posteriores no se calculan explícitamente sino que se obtienen sus estimadores de Monte Carlo, generalmente a partir de la técnica conocida como *Gibbs sampler*. Está técnica, ampliamente descrita en la bibliografía (Gelfand y Smith, 1990; Wang et al, 1994), se basa en cadenas de Markov que convergen hacia las distribuciones marginales posteriores de cada variable del análisis.

Esta estimación por Monte Carlo es computacionalmente más eficiente que la integración explícita, más aún, es la única posibilidad viable hasta el momento. Sin embargo, presenta el problema de que su convergencia es lenta en general y en algunos casos tan díficil de determinar que puede mover a tomar por buenos los estimadores de Monte Carlo calculados a partir de cadenas que no han convergido todavía (Gelman y Rubin, 1992). Una aportación interesante que resuelve este problema para el caso de los modelos lineales es el procedimiento conocido como cadenas emparejadas (Johnson, 1996). Valga como ejemplo el siguiente: usando este procedimiento, se demuestra que la convergencia de un modelo lineal que asuma los componentes de varianza conocidos (el equivalente de un BLUP), el Gibbs sampler converge a la misma velocidad que un Gauss-Seidel al resolver el sistema de las ecuaciones de modelo mixto (García Cortés et al, 1997).

Modelos no lineales

A nuestro juicio, la inferencia bayesiana aventaja a la inferencia frecuentista en sencillez y flexibilidad, lo que ha permitido desarrollar en pocos años la estimación de valores genéticos en modelos no lineales con carácter general. Nuestro grupo ha centrado su actividad durante los últimos años en modelos relacionados con funciones de producción (Varona et al, 1996) y en modelos categóricos (Moreno et al, 1997).

Funciones de producción.

En producción animal existen caracteres que se manifiestan a lo largo del tiempo como una función de producción. En ocasiones, alguno de los

parámetros corresponde a una característica, por ejemplo, el parámetro c de la función de Wood indica la persistencia en la curva de lactación. La importancia de la evaluación genética de los animales para parámetros de estas funciones radica principalmente en que permite orientar la selección hacia el ritmo de producción en lugar de hacia el total de producción. Por ejemplo, puede permitir la selección de animales que tengan un pico bajo en su curva de lactación, una mayor persistencia, o pesos al nacimiento bajos con altos crecimientos.

Los planes de selección que incluyen caracteres medidos a lo largo del tiempo y que están sujetos a funciones de producción son normalmente difíciles de diseñar ya que animales en diferentes fases de producción que conviven en la explotación deben ser controlados a la vez. Por ejemplo, en el caso del crecimiento, los animales han de ser pesados cuando el controlador visita la explotación, independientemente de la edad que tengan. Un retraso o adelanto de pocos días en el control del rendimiento de un animal puede favorecer o perjudicar su valor genético predicho. Tradicionalmente esto se ha resuelto ajustando los datos antes de ser analizados, aunque esto puede provocar sesgos sensibles en algunos parámetros, como es el caso de las correlaciones genéticas y principalmente residuales entre caracteres (Varona et al, 1997).

La inferencia bayesiana permite abordar este problema haciendo una estimación conjunta de los datos observados sin corregir, si se incluye información en el modelo acerca del momento en el que ha sido hecho el control de cada animal. El análisis bayesiano que permite incluir toda esta información conjuntamente corresponde a lo que técnicamente se conoce como modelo bayesiano jerárquico (Berger, 1985; Wakefield et al, 1994). Se denomina así porque incluye dos modelos jerarquizados, el primero el que ajusta los datos de un animal a su función de producción, y el segundo el que considera la función de producción de un animal como un carácter heredable.

El análisis bayesiano jerárquico se ha mostrado como una herramienta flexible para analizar diferentes funciones que afectan a diversos tipos de producción (leche, carne) y sencilla de aplicar, aunque cuando se aplique a funciones de producción no lineales, las distribuciones condicionales pueden tomar formas no conocidas. En general, se ha detectado una convergencia lenta de las cadenas.

Otra caracteristica que conviene resaltar es que atempera el efecto de controles de rendimientos mal registrados (*outliers*). Esto ocurre porque el ajuste de las funciones se realiza dentro del análisis genético y por tanto incluye información de parientes o de otros datos que compartan los mismos efectos fijos.

Caracteres categóricos.

Una técnica tan interesante como útil cuando se combina con el Gibbs sampler es la llamada *data augmentation*. Consiste en introducir en el modelo

unas variables latentes adicionales que corresponden a variables no observadas. Por ejemplo, siguiendo esta técnica los datos pueden ser aumentados para equilibrar los modelos, rellenando las casillas vacías del diseño para que las distribuciones posteriores sean más sencillas de manejar; también puede ser utilizada en un modelo categórico para incluir las variables subyacentes y aligerar así el análisis.

Si se utiliza el Gibbs sampler conjuntamente con data augmentation permite predecir los valores genéticos en caracteres categóricos utilizando modelos animales (Sorensen et al, 1995). El data augmentation aporta al modelo unas variables adicionales que corresponden a los valores genéticos de los animales en la escala subyacente. De este forma los valores genéticos de los animales y los componentes de varianza se estiman en la escala subyacente de una forma más cómoda.

El análisis proporciona resultados sesgados cuando el número de observaciones por efecto fijo es pequeño (Moreno et al, 1997), si bien puede utilizarse cuando el diseño incluye efectos fijos con suficientes observaciones (Altarriba et al, 1997). Moreno et al (1997) proponen también un procedimiento basado en Bootstrap para corregir el sesgo de las estimaciones que ha dado buenos resultados en modelos animales en los que los efectos fijos tengan pocas observaciones por nivel.

Citas Bibliográficas

ALTARRIBA, MORENO, VARONA y GARCIA CORTES, 1997, J. Anim. Sci (enviado)

BERGER, 1985. Statistical Decision and Bayesian Analysis. 2nd Ed. Springer Verlag.

GARCIA CORTES, RICO y GROENEVELD. 1997. J.Anim.Sci (enviado)

GELFAND y SMITH, 1990. J. Am. Stat. Assoc. 85:398-409.

GELMAN y RUBIN. 1992. (Bernardo, J. M., J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds.), Oxford Univ. Press, UK, 169

JOHNSON. 1996. J. Amer. Stat. Assoc. 91:154

MORENO, SORENSEN, VARONA, GARCIA y ALTARRIBA, 1997. Genet. Sel. Evol. (en prensa)

SORENSEN, ANDERSEN, GIANOLA y KORSGAARD, 1995. Genet. Sel. Evol. 27:229.

TANNER y WONG. 1987. J. Amer. Statis. Assoc. 82:528-550

VARONA, MORENO, GARCIA CORTES y ALTARRIBA, 1997. Lives. Proc. Sci. (en prensa).

VARONA, MORENO, GARCIA CORTES y ALTARRIBA. J. Dairy Sci. (enviado)

WANG, RUTLEDGE Y GIANOLA, 1994. Genet. Sel. Evol. 26:91-115.

WAKEFIELD, SMITH, RACINE-POON y GELFAND, 1994. Appl. Statist. 43:201-221.