

MÉTODO LR EN SIMULACIONES: GANANCIA GENÉTICA Y DISPERSION USANDO HEREDABILIDAD EQUIVOCADA

Macedo¹, F. y Legarra¹, A.

¹GenPhiSE-INRA. 24 chemin de Borde-Rouge, Auzeville, France; fernando.macedo@inra.fr

INTRODUCCIÓN

En la medida que los costos de genotipado disminuyeron, el uso de información genómica en las evaluaciones genéticas se ha hecho más frecuente y hay más estudios de sesgos en este tipo de evaluación. Diversos estudios realizados en bovinos de leche han detectado sesgo en las evaluaciones genéticas tanto en la ganancia genética estimada, como en la dispersión de los valores de cría estimados (EBV) (Spelman et al., 2010; Sargolzaei et al., 2012). Sesgos en evaluaciones genéticas también fueron encontrados en poblaciones de razas Pirenaicas de ovinos lecheros (Baloche et al., 2014).

Una de las posibles causas de sesgos en evaluaciones es el uso de una heredabilidad (h^2) incorrecta, ya sea mayor o menor a la verdadera, principalmente cuando la h^2 no es estimada frecuentemente o cuando cambia a lo largo del tiempo, por ejemplo, por modificaciones en el ambiente.

En la actualidad la herramienta más usada para probar modelos de evaluación y detectar sesgos es la validación cruzada. No obstante, dista de ser una herramienta óptima, acorde a la realidad de los esquemas de selección (Legarra y Reverter, 2017); principalmente en sistemas de evaluación con bajo número de padres y baja cantidad de hijas como pueden ser los sistemas de producción de leche ovina. Recientemente se ha propuesto un método de fácil aplicación y apropiado para usar en cualquier esquema de selección llamado método Linear Regression (LR) (Legarra y Reverter, 2018). Dicho método se basa en la comparación de dos subconjuntos de valores de cría estimados (EBV); uno cuando se usa información parcial, incompleta o antigua (subconjunto parcial) y otro partiendo de información agregada (subconjunto completo). El método asume que el modelo, y la h^2 usada, son correctas. El objetivo de este trabajo ha sido estudiar mediante simulación las propiedades del método LR para estimar el sesgo en evaluaciones genéticas cuando la h^2 usada es incorrecta.

MATERIAL Y MÉTODOS

Simulación. Se simularon 20 réplicas de poblaciones con aproximadamente 450000 individuos para características de h^2 0,10 y 0,30 medidas solamente en las hembras, usando el software QMSim v.1.10 (Sargolzaei y Schenkel, 2009). Para ambas situaciones se plantearon 3 escenarios: (1) los individuos fueron evaluados mediante BLUP con la misma h^2 usada para la simulación (h^2 verdadera); (2) los individuos fueron evaluados usando una h^2 0,05 inferior a la verdadera y (3) los individuos fueron evaluados usando una h^2 0,05 superior a la verdadera. Por lo tanto, evaluamos 6 escenarios: para simulaciones usando h^2 de 0,10: individuos evaluados con h^2 de 0,05, 0,10 y 0,15 y para simulaciones usando h^2 de 0,30: individuos evaluados con h^2 de 0,25, 0,30 y 0,35. La varianza fenotípica siempre fue de 1.

Se aplicó el método LR sobre individuos nacidos en la generación 5 – 9; siendo el subconjunto parcial (\hat{u}_p) los EBV estimados cuando los individuos no poseían hijas (y por lo tanto sin registros) y como subconjunto completo (\hat{u}_w) los EBV estimados cuando los individuos ya poseían registros de sus hijas (en la generación siguiente).

Finalmente se obtuvieron por cada repetición 5 comparaciones entre parcial y completo; totalizando 100 comparaciones, tomando en cuenta todas las réplicas. Para cada par de subconjuntos se estimaron el **sesgo** ($\mu_{wp} = \bar{u}_p - \bar{u}_w$); relacionado con la ganancia genética; y el **coeficiente de regresión** del subconjunto completo en el parcial ($b_{pw} = \text{cov}(\hat{u}_p, \hat{u}_w) / \text{var}(\hat{u}_p)$); relacionado con la medición de la dispersión. Los resultados obtenidos con el método se contrastaron con los valores de obtenidos de comparar los \hat{u}_p y los verdaderos valores de cría (TBV) ($\mu_{TBV,p}$ y $b_{TBV,p}$). Una presentación más profunda del método LR y sus estimadores se puede encontrar en Legarra y Reverter (2018).

RESULTADOS Y DISCUSIÓN

Cuando la h^2 es inferior a la verdadera los EBVs se infraestiman, y cuando la h^2 es superior, se sobreestiman. El método LR intenta estimar este sesgo con los resultados presentados en la Figura 1. Cuando se aplica en el modelo de evaluación una h^2 menor a la usada para simular, se encuentra un sesgo real (\hat{u}_p menor que TBV) mayor al detectado por el estimador

(para h^2 simulada de 0,10: h^2 de evaluación 0,05, μ_{WP} de -0,029 y $\mu_{TBV,p}$ -0,111; para h^2 simulada de 0,30 y h^2 de evaluación de 0,25, μ_{WP} de -0,028 y $\mu_{TBV,p}$ de -0,054). Una situación similar pero inversa sucede cuando la h^2 usada para las evaluaciones es mayor a la usada para simular; se observa una subestimación del sesgo (para h^2 simulada de 0,10: h^2 de evaluación 0,15, μ_{WP} de 0,036 y $\mu_{TBV,p}$ de 0,091; para h^2 simulada de 0,30 y h^2 de evaluación de 0,35, μ_{WP} de 0,032 y $\mu_{TBV,p}$ de 0,050). La diferencia en este caso es que el sesgo es positivo, o sea los \hat{U}_p son mayores que los TBV. Por tanto, incluso cuando la h^2 no es correcta, el método es capaz de detectar la existencia de sesgos (infra- o sobre- estimación sistemática de EBVs) pero no su magnitud.

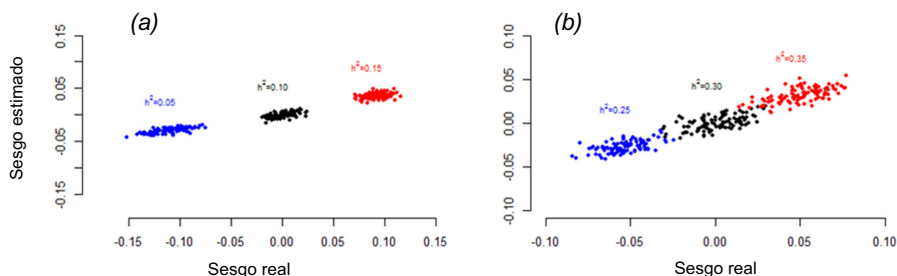


Figura 1. Sesgo observado (en abscisa) y estimado (en ordenada) en todas las simulaciones: (a) simulaciones con h^2 de 0,10 y evaluaciones con h^2 de 0,05, 0,10 y 0,15; y (b) simulaciones realizadas con h^2 de 0,30 y evaluaciones realizadas con h^2 de 0,25, 0,30 y 0,35.

Cuando se usa en la evaluación h^2 mayores a la verdadera, el $b_{TBV,p}$ resulta ser <1 (sobredispersión) para ambos escenarios (h^2 simulada de 0,10 y en modelo de evaluación 0,15: $b_{TBV,p}=0,83$; h^2 simulada de 0,30 y en modelo de evaluación 0,35: $b_{TBV,p}=0,97$), mientras que el uso de h^2 inferiores a la usada para simular resulta en valores medios superiores e inferiores a uno (h^2 simulada de 0,10 y en modelo de evaluación 0,05: $b_{TBV,p}=0,96$; h^2 simulada de 0,30 y en modelo de evaluación 0,25: $b_{TBV,p}=1,06$). De todas formas, en todos los escenarios, este indicador de dispersión se logra estimar de manera satisfactoria (figura 2) tanto en sentido como en magnitud

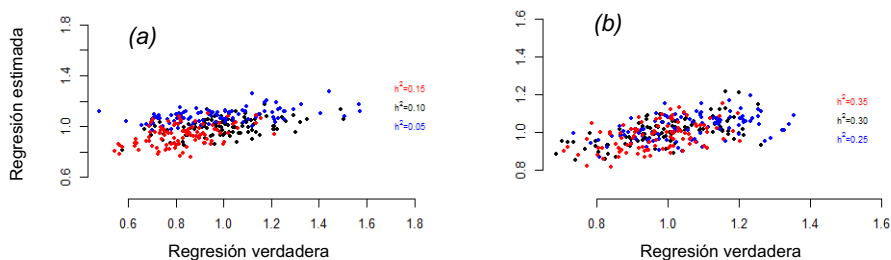


Figura 2. Coeficiente de regresión verdadero (en abscisa) y estimado (en ordenadas) en todas las simulaciones para todos los escenarios probados: (a) simulaciones con h^2 de 0,10 y evaluaciones con h^2 de 0,05, 0,10 y 0,15; y (b) simulaciones realizadas con h^2 de 0,30 y evaluaciones realizadas con h^2 de 0,25, 0,30 y 0,35.

En ese sentido en este trabajo se observó que el uso de h^2 equivocadas en el modelo de evaluación provoca sesgos y valores de dispersión de EBV menores a 1 y que el método LR

fue capaz de estimar estos valores tanto en sentido como en magnitud, de forma más o menos precisa dependiendo de la magnitud de la h^2 usada para la simulación. Así demostramos que el método LR es capaz de detectar la presencia de sesgos y sobredispersiones aun cuando los parámetros usados para realizar las evaluaciones no sean correctos.

REFERENCIAS BIBLIOGRÁFICAS

• Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J.-M., Robert-Granié, C. & Barillet, F. 2014. *J. Dairy Sci.* 97:1107–1116. • Boichard, D., Bonaiti, B., Barbat, A. & Mattalia, S. 1995. *J. Dairy Sci.* 78:431–437. • Legarra, A. & Reverter, A. 2017. Can we frame and understand cross-validation results in animal breeding? *Proc. Assoc. Advmt. Anim. Breed. Genet.* • Legarra, A. & Reverter, A. 2018. *Genet. Sel. Evol.* 50:1–18. • Reverter, A., Golden, B.L., Bourdon, R.M. & Brinks, J.S. 1994. *J. Anim. Sci.* 72:34–37. • Sargolzaei, M., Chesnais, J. & Schenkel, F. 2012. Assessing the bias in top GPA bulls. doi:https://www.cdn.ca/Articles/GEBOCT2012/Jacques_Bias_in_Top_Bulls.pdf. • Sargolzaei, M. & Schenkel, F.S. 2009. *Bioinformatics* 25:680–1. • Spelman, R.J., Keehan, M.D., Obolonkin, V., Winkelman, A.M., Johnson, D.L. & Harris, B. 2010. Application of Genomic Selection 0–3.

Agradecimientos: Los autores quisieran agradecer a las siguientes instituciones financiadoras: Metaprograma GenSel (INRA), La región Occitanie, Proyecto ARDI (Interreg, POCTEFA); así como también a la plataforma de computación Genotoul.

METHOD LR IN SIMULATION: GENETIC GAIN AND DISPERSION USING WRONG HERITABILITIES

ABSTRACT: There is an interest in the study of bias in genetic evaluations, mainly in the context of genomic evaluations. Recently a new method, called Linear Regression, based on statistics obtained from the comparison from estimated breeding values (EBV) obtained from a data set with old records or partial (\hat{u}_p) and from a data set with newer information or whole (\hat{u}_w). In this study, we examine in simulations, the behavior of the method when wrong heritabilities (h^2) are used to evaluate individuals. For simulations we use two h^2 (0.1 and 0.3) and for the evaluation models we use h^2 0.05 points over and under the h^2 used to simulate. Two statistics were analyzed: the bias, related to the genetic gain and the slope of the regression of \hat{u}_w on \hat{u}_p , related to dispersion of EBV. We observed that the use of wrong h^2 results in estimated bias lower than the real one, but in the correct sense. When we observed the slope the use of high h^2 causes a real slope under one, that it could also be detected. In any case, it is possible to estimate by method LR the bias and the slope, with more precision when the true h^2 is high.

Keywords: genetic evaluation, bias, slope