Génét. Sél. Evol., 1984, **16** (2), 177-184

The use of prior information in the estimation of heritability by parent-offspring regression

M.A. TORO and J.V. PRUÑONOSA

Departamento de Genética y Sección de Proceso de Datos
Instituto Nacional de Investigaciones Agrarias
Carretera de La Coruña, Km. 7, Madrid (España)

Summary

The use of prior information in the estimation of the heritability by parent-offspring regression is discussed within a bayesian context. The *a posteriori* distribution is obtained by combining the *a priori* distribution (uniform between 0 and 1), to that obtained from the data. Hence, a bayesian estimator h^{*2} is proposed and its performance compared with those obtained by the least squares and constrained maximum likelihood methods and also with two different bayesian estimators (NIGAM *et al.*, 1979), using Monte Carlo simulation techniques. It is concluded that the estimate h^{*2} should be preferred to the others, particularly for small sets of data.

Key words : heritability, bayesian estimation.

Résumé

Utilisation d'une information a priori *lors de l'estimation de l'héritabilité par la régression parent-descendant*

L'utilisation d'une information *a priori* lors de l'estimation de l'héritabilité par la régression parent-descendant est discutée dans le cadre bayésien. La distribution *a posteriori* résulte de la combinaison de la distribution *a priori* uniforme sur (0, 1) et de celle déduite des données. L'estimateur bayésien correspondant h^{*2} est comparé à ceux des moindres carrés et du maximum de vraisemblance contraint ainsi qu'à deux estimateurs bayésiens proposés par NIGAM *et al.* (1979) en ayant recours à des techniques de simulation. L'étude conclut à la préférence à donner à l'estimateur bayésien h^{*2} notamment dans le cas de petits échantillons.

Mots-clés : héritabilité, estimateur bayésien.

I. Introduction

The heritability of a trait was defined by LUSH (1949) as the ratio of additive variance to phenotypic variance, being the most important genetic parameter in the prediction of selection response. Consequently, the problems involved in its estimation have received considerable attention (HILL, 1974 ; BULMER, 1980 ; FALCONER, 1981).

By definition, the heritability value lies between 0 and 1. Nevertheless, values outside this range can be found in practice and they are usually ascribed to sampling errors. In such cases, the current procedure is to set these anomalous estimates to the nearest valid bound, although the validity of this procedure in unclear (SALES & HILL, 1976 ; HAYES & HILL, 1981).

Theoretically, several authors (THEIL & GOLDBERG, 1961 ; HOERL & KENNARD, 1970 ; MARCQUARDT & SNEE, 1975 ; TOUTENBURG & ROEDER, 1978) have considered the problem of using prior information in the estimation of regression coefficients. NIGAM *et al.* (1979) applied that theory to the estimation of heritability by regression analysis and proposed two new estimators h_1^2 and h_2^2 which were considered to be superior to the traditional ones.

In this paper, a bayesian formulation of h_1^2 and h_2^2 will be given showing that they are not logically sound. A new bayesian estimator h^{*2} is then proposed that seems superior with regard to several statistical criteria.

II. Proposed improved estimators

Consider the linear regression of offspring (y) on single parent (x) expressed in deviations from their means. The statistical model is

$$y = \beta x + \varepsilon$$

where β is the regression coefficient and ε is the associated random error which is normally distributed with zero mean and finite variance (σ^2).

The classical least squares estimator of β is

$$\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

which is unbiased and has sampling variance

$$V(\hat{\beta}) = \frac{\sigma^2}{\Sigma x_i^2}$$

The heritability is estimated as $\hat{h}^2 = 2\hat{\beta}$. When $\hat{h}^2$ does not lie between 0 and 1 the usual practice is to consider the following estimator h_o^2 :

$$h_o^2 = \begin{cases} 0 & \text{if } \hat{h}^2 < 0 \\ h^2 & \text{if } 0 \leq h^2 \leq 1 \\ 1 & \text{if } \hat{h}^2 > 1 \end{cases}$$

It can be shown that this estimate is also the constrained maximum likelihood estimate. The mean $E(h_o^2)$ and the variance $V(h_o^2)$ of the sampling distribution of h_o^2 calculated from the properties of the truncated normal distribution (PEARSON, 1903 ; COCHRAN, 1951) are

$$E(h_o^2) = 2\{\hat{\beta} + [z_o - z_1 + p_o x_o + (1 - p_1)x_1]\sqrt{V(\hat{\beta})}\} \quad (1)$$

$$V(h_o^2) = 4V(\hat{\beta})\{p_1 - p_o + x_o z_o - x_1 z_1 + p_o x_o^2 + (1 - p_1)x_1^2 - [z_o - z_1 + p_o x_o + (1 - p_1)x_1]^2\} \quad (2)$$

where x_o and x_1 are the abscissae of the lower and the upper bounds in standard normal units, respectively, z_o and z_1 the corresponding ordinates on the standard normal curve and p_o and p_1 the values of the standard normal distribution function for x_o and x_1.

$$p_o = \int_{-\infty}^{x_o} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \qquad p_1 = \int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

The general methodology for obtaining bayesian estimators is to combine the prior information and that available from the data in a new posterior distribution. The two estimators (h_1^2 and h_2^2) proposed by NIGAM *et al.* (1979) can be formulated in a bayesian framework assuming and *a priori* distribution of the regression coefficient N (β_o, w^2). Because the regression coefficient obtained from the data is distributed N (β, $\frac{\sigma^2}{\Sigma x_i^2}$), the bayesian estimator will combine the two items of information weighted by the inverses of their variances :

$$h_1^2 = 2\hat{\beta}_1 = 2\,\frac{\hat{\beta}\,\frac{\Sigma x_i^2}{\sigma^2} + \beta_o\,\frac{1}{w^2}}{\frac{\Sigma x_i^2}{\sigma^2} + \frac{1}{w^2}} \tag{3}$$

The variance and the bias of this estimator will be

$$V(h_1^2) = 4V(\hat{\beta}_1) = \frac{1}{\frac{\Sigma x_i^2}{\sigma^2} + \frac{1}{w^2}} \tag{4}$$

$$B(h_1^2) = 2B(\hat{\beta}_1) = 2\,\frac{\frac{1}{w^2}(\beta_o - \beta)}{\frac{\Sigma x_i^2}{\sigma^2} + \frac{1}{w^2}}$$

The estimator h_1^2 proposed by NIGAM *et al.* (1979) by imposing a linear stochastic constraint on the regression equation can be shown to be equivalent to the bayesian estimator for $\beta_o = 1/4$ and $w^2 = 1/64$. On the other hand, the estimator h_2^2 obtained from converting the inequality constraints in the form of a concentration ellipsoid with minimum volume, is equivalent to the bayesian estimator for $\beta_o = 1/4$ and $w^2 = 1/8$. In principle, h_1^2 sould be preferred to h_2^2 because of its lower mean square error.

The arbitrary nature of the h_1^2 and h_2^2 estimators now becomes apparent as there is no empirical or logical reason to assume an *a priori* normal distribution of the regression coefficient with mean 1/4.

As the only initial information available is that the regression coefficient is bounded, it seems reasonable to assume an *a priori* uniform distribution between 0 and 1/2. This type of distribution is justified by the lack of information about the true value of the parameter. The new bayesian estimator proposed here, h^{*2}, is associated with an *a posteriori* distribution which is a combination of the *a priori* distribution (uniform between 0 and 1/2) and that obtained from the data N ($\hat{\beta}$, $\sigma^2/\Sigma x_i^2$), as shown in figure 1. It appears reasonable to choose the mean as measure of the central tendency, given the shape of the

posterior distribution. The use of the mode will result in $h^{*2} = h_0^2$ and, on the other hand, numerical analyses have shown that similar results are obtained by using either the mean or the median. The h^{*2} heritability estimator is then given by the mean of the posterior distribution.

$$h^{*2} = 2\left[\hat{\beta} + \frac{z_o - z_1}{p_1 - p_o}\sqrt{V(\hat{\beta})}\right] \tag{5}$$

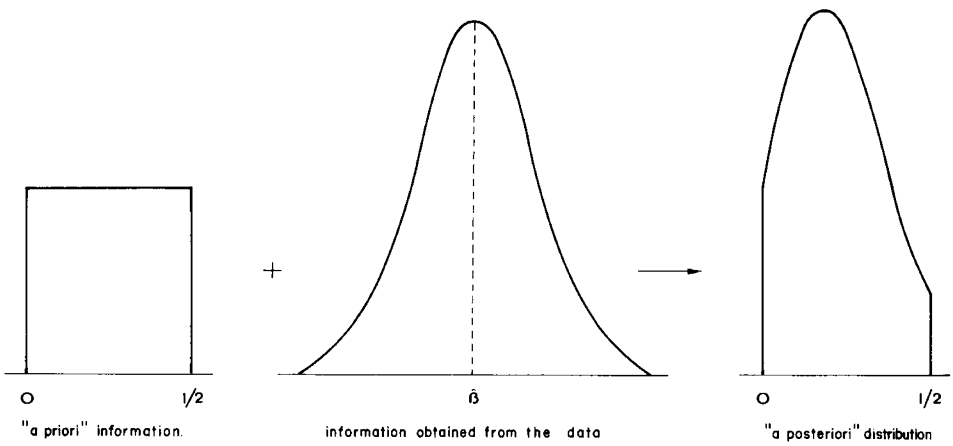

FIG. 1

Posterior distribution obtained by combining a priori *information with that obtained from data.*
Distribution a posteriori *obtenue en combinant l'information* a priori *avec celle obtenue à partir des données.*

Unfortunately, it is not obvious how to derive analytical formulae for the expected value and the variance of the h^{*2} estimator. For this reason its performance will be compared with alternative estimators by simulation methods.

It should be noted that the h^{*2} estimator may also have a non-bayesian and straighforward interpretation. What is meant in essence is that the correct way of truncating the sample distribution of the regression coefficient is not to set the estimate to the nearest valid bound but to reassign the distribution until the probability area between 0 and 1/2 is again unity.

Consider as an example the estimation of the heritability by daughter-dam regression of the number of eggs laid by virgin females of *Tribolium castaneum* scored from the 7th to the 11th day after the emergence of the adult. Data was available from 40 full-sib families of size five. The results are summarized below :

$\Sigma x_i y_i = -160.87$ $\quad$ $\hat{\beta} = -0.05$
$\Sigma x_i^2 = 3\,217.50$ $\quad$ SE $(\hat{\beta}) = 0.11$
$\sigma^2 = 38.93$

The necessary quantities for the estimation of $\hat{h}^2$, h_0^2, h_1^2, h_2^2 and h^{*2} are

$$w^2 = 1/64 \quad \text{or} \quad 1/8, \qquad \beta_o = 1/4$$
$$x_o = (0 - \hat{\beta})/SE\,(\beta) = 0.45, \qquad x_1 = (1/2 - \hat{\beta})/SE\,(\hat{\beta}) = 5.00$$

and, from the standardized normal tables,

$$p_o = 0.67, \qquad p_1 \simeq 1.00$$
$$z_o = 0.36, \qquad z_1 \simeq 0.00$$

The heritability calculated by the least squares method is $\hat{h}^2 = -0.10 \pm 0.22$. Following the usual practice the estimate would be set to the nearest valid bound, $h_o^2 = 0 \pm 0.10$ with a corrected standard error given by formula (2). Estimators h_1^2 and h_2^2 from (3) and (4) are $h_1^2 = 0.16 \pm 0.16$ and $h_2^2 = -0.047 \pm 0.21$, respectively. Finally, from (5), our estimator $h^{*2} = 0.14$.

III. Simulation results

Simulation has been carried out following the methods developed by RØNNINGEN (1974). It must be noted that these methods statistically simulate the genetic model, but do not simulate mendelian sampling.

The study was carried out by generating 1 000 samples, each consisting of 20 half-sib families of size five. The heritability was estimated by twice the parent-offspring regression coefficient using the following methods :

$\hat{h}^2$ = heritability calculated by the least squares method.

h_o^2 = truncated heritability or constrained maximum likelihood estimate.

h^{*2} = new bayesian estimator proposed above.

h_1^2 and h_2^2 = bayesian estimators proposed by NIGAM *et al.* (1975).

Table 1 shows the average values of these estimates over 1 000 runs together with the corresponding empirical standard errors (SE). The true values of the heritability h^2 used to generate the data are also given. The least squares method is the only one resulting in unbiased estimation for each value of the true heritability. The truncated estimator h_o^2 is only biased for extreme values of the true heritability ($h^2 \leqslant 0.20$, $h^2 \geqslant 0.80$) whilst h^{*2} and h_1^2 are biased for almost all values of h^2. The bias of the estimator h_2^2 is very small. It is also apparent that h_o^{*2} and h_1^2 have standard errors which are considerable lower than that of the least squares estimator. On the other hand SE (h_2^2) is similar to SE ($\hat{h}^2$). SE (h_o^2) is appreciably smaller than SE ($\hat{h}^2$) for extreme values of heritability.

Two criteria have been used to compare the different estimators, the mean square error (MSE) and the absolute value of the sum of the deviations from the true value (SAD). Both criteria seem compatible with the practical use of the heritability coefficient in the prediction of response to artificial selection. Traditionally, the bias has been given a greater importance than the magnitude of the variance of the estimators but this procedure has been challenged (HOERL & KENNARD, 1970). For practical purposes, it seems justifiable to prefer the estimator of heritability which is closer to the true value irrespective of all other statistical considerations.

TABLE 1

Average values of heritability estimates (x10²) (± SE) obtained by the least-squares ($\hat{h}^2$) and constrained maximum likelihood (h_o^2) methods, the bayesian estimators proposed by NIGAM et al. (1979) *(h_1^2, h_2^2) and that proposed in this paper (h_o^{*2}), compared with its true value (h^2).*

Valeurs moyennes des estimations de l'héritabilité (x10²) (± écart-type) obtenues avec la méthode des moindres carrés ($\hat{h}^2$), le maximum de vraisemblance avec contraintes (h_o^2), les estimateurs bayésiens proposés par NIGAM et al. (1979) *(h_1^2, h_2^2) et l'estimateur présenté dans ce travail (h^{*2}), en comparaison avec sa vraie valeur (h^2).*

h^2	$\hat{h}^2$	h_o^2	h_1^2	h_2^2	h^{*2}
0	− 1 ± 21	8 ± 12	20 ± 13	3 ± 19	18 ± 9
10	9 ± 22	14 ± 16	27 ± 13	13 ± 20	24 ± 12
20	21 ± 23	23 ± 20	34 ± 13	23 ± 21	30 ± 14
30	29 ± 23	30 ± 21	38 ± 13	31 ± 21	36 ± 15
40	40 ± 23	40 ± 22	44 ± 12	40 ± 22	43 ± 16
50	50 ± 24	50 ± 23	50 ± 13	50 ± 21	50 ± 16
60	61 ± 24	60 ± 23	56 ± 13	60 ± 21	57 ± 16
70	70 ± 23	69 ± 21	61 ± 13	68 ± 21	64 ± 15
80	81 ± 23	78 ± 18	68 ± 14	78 ± 21	71 ± 13
90	90 ± 20	86 ± 15	74 ± 12	87 ± 18	77 ± 11
100	100 ± 19	93 ± 11	83 ± 13	97 ± 18	83 ± 9

TABLE 2

Mean square error (x10⁴) (MSE) of heritability estimates obtained by the least-squares ($\hat{h}^2$) and constrained maximum likelihood (h_o^2) methods and the bayesian estimators proposed by NIGAM et al. (1979) *(h_1^2, h_2^2) and that proposed in this paper (h^{*2}), for different true values of the heritability (h^2).*

Valeurs de l'erreur quadratique moyenne (x10⁴) (MSE) des estimations de l'héritabilité obtenues avec la méthode des moindres carrés ($\hat{h}^2$), le maximum de vraisemblance avec contraintes (h_o^2), les estimateurs bayésiens proposés par NIGAM et al. (1979) *(h_1^2, h_2^2) et l'estimateur présenté dans ce travail (h^{*2}), pour différentes valeurs vraies de l'héritabilité (h^2).*

h^2	MSE ($\hat{h}^2$)	MSE (h_o^2)	MSE (h_1^2)	MSE (h_2^2)	MSE (h^{*2})
0	451	195	564	383	424
10	507	285	457	425	335
20	544	403	358	450	303
30	531	436	234	426	253
40	540	486	175	428	257
50	578	519	163	452	266
60	571	513	180	453	264
70	531	440	249	435	270
80	516	346	317	426	256
90	394	236	397	342	279
100	372	182	471	333	363

Both criteria lead to similar conclusions and therefore, only the MSE criterium will be discussed in detail. The MSE values of the different heritability estimates are shown in table 2. The two estimators h_o^2 and h^{*2} are always better than $\hat{h}^2$ for all values of the true heritability. Preference should be given to h^{*2} over $\hat{h}^2$ because of the lower MSE implied (almost half of that corresponding to the least squares estimate). The truncated estimator h_o^2 is accompained by a reduction in the MSE only for extreme values of the true heritability. Although the use of the h_1^2 estimator leads to a considerable reduction of the MSE there are several reasons for avoiding its use : (1) this estimator is not better than that obtained by least squares for extreme heritability values ; (2) the assumption of an *a priori* distribution N (1/4, 1/64) of the heritability is not logically or empirically sound ; (3) h_1^2 is clearly worse than h^{*2} whatever criteria are applied.

Table 3 shows the MSE values for different number of families and family sizes. It appears clear that the h^{*2} estimator is more efficient if the number of families and/or the family size is small (50 families of 10 half-sibs or less).

TABLE 3

Mean square error (x10⁴) of heritability estimates obtained by the least-squares ($\hat{h}^2$) and constrained maximum likelihood (h_o^2) methods and the bayesian estimator proposed in this paper (h^{*2}) calculated from different number of families and family sizes for different true values of the heritability (h^2).

Valeurs de l'erreur quadratique moyenne (x10⁴) des estimations de l'héritabilité obtenues avec la méthode des moindres carrés ($\hat{h}^2$), le maximum de vraisemblance avec contraintes (h_o^2) et l'estimateur bayésien proposé dans ce travail (h^{*2}) calculées pour des nombres et tailles de familles différents ainsi que diverses valeurs de l'héritabilité vraie (h^2).

No. families	Family size	True heritability	Mean square error		
			$\hat{h}^2$	h_o^2	h^{*2}
100	10	20	54	54	46
		40	60	60	60
		60	59	59	59
100	5	20	94	90	68
		40	102	102	99
		60	94	94	92
50	10	20	117	107	77
		40	128	128	120
		60	128	128	120
50	5	20	188	160	116
		40	175	175	151
		60	191	191	165
20	10	20	324	256	195
		40	371	362	240
		60	356	335	232
20	5	20	545	403	303
		40	540	486	257
		60	571	513	264
20	1	20	2 323	1 128	645
		40	2 297	1 190	273
		60	2 044	1 133	246

In some situations, prior information would allow to bound the true heritability value to lie within narrower bounds.

$$L_1 \leq h^2 \leq L_2 \quad \text{where} \quad L_1 \geq 0 \quad \text{and} \quad L_2 \leq 1$$

In this case, the reduction in the resulting MSE values arising is quite considerable. It is not obvious how the estimators h_1^2 and h_2^2 can now be used and this therefore results in further disadvantage.

It can thus be concluded that the new bayesian estimator proposed here h^{*2} is superior to the usual ones and must be preferred specially if the sample sizes are small. Furthermore, the principles involved in its derivation can be generalized to more complex situations and this work is now in progress.

Received October 10, 1982.
Accepted November 26, 1983.

Acknowledgement

The authors thank L. SILVELA, C. LÓPEZ-FANJUL, E. CARBONELL and M.T. DOBAO fir his valuable assistance.

References

BULMER M.G., 1980. *The Mathematical Theory of Quantitative Genetics.* 254 pp., Oxford University Press, Oxford.

COCHRAN W.G., 1951. Improvement by means of selection. *Proceedings of the second Berkeley symposium on mathematical statistics and probability.* Neyman J. (ed.), 449-470, University of California Press, Berkeley.

FALCONER D.S., 1981. *Introduction to Quantitative Genetics.* 2nd edition, 340 pp., Longman, London.

HAYES J.F., HILL W.G., 1981. Modification of estimates of parameters in the construction of genetic selection indices (« bending »). *Biometrics,* **37,** 483-493.

HILL W.G., 1974. Heritabilities : estimation problems and the present state of information. *1st World Congress on Genetics Applied to Livestock Production,* Madrid, October 7-11 1974, **1,** 343-351, Garsi, Madrid.

HOERL A.E., KENNARD R.W., 1970. Ridge regression : application to non-orthogonal problems. *Technometrics,* **12,** 69-82.

LUSH J.L., 1949. *Animal Breeding Plans.* 3rd edition, 443 pp., Iowa State University Press, Ames.

MARCQUARDT D.W., SNEE R.D., 1975. Ridge regression in practice. *Am. Stat.,* **29,** 3-20.

NIGAM A.K., SRIVASTAVA V.K., JAIN J.P., GOPALAN R., 1979. A note on estimation of heritability by regression analysis. *Biom. J.,* **21,** 667-673.

PEARSON K., 1903. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. R. Soc. Lond.,* A **200,** 1-66.

RØNNINGEN K., 1974. Monte Carlo simulation of statistical-biological models which are of interest in animal breeding. *Acta Agric. Scand.,* **24,** 135-142.

SALES J., HILL W.G., 1976. Effect of sampling errors on efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. *Anim. Prod.,* **23,** 1-14.

THEIL H., GOLDBERGER A.S., 1961. On pure and mixed statistical estimation in economics. *Int. Econ. Rev.,* **2,** 65-78.

TOUTENBURGH H., ROEDER B., 1978. Mini max-linear and Theil estimator for restricted regression coefficients. *Math. Operationsforsch. Stat. (Ser. Stat.),* **9,** 499-505.