HOW TO MEASURE HORSE PERFORMANCES PRESENT SITUATION AND PROSPECTS

Anne Tavernier
Station de Génétique Quantitative et Appliquée,INRA
78350 Jouy-En-Josas, France

SUMMARY

Records in competition are often registered as earnings. This measure is useful but leads to difficulties to determine the value allocated to non-earning horses, to standardize by transformation (to a normal distribution) and to account for repeated performances. The major problem is the existence of non optimal rules of endowment, not exactly based on the level of competitors. A new model, assuming an underlying variable on which only the relative position (rank) of the animals is measured, is proposed and an estimate of genetic value is obtained as the mode of a posterior density.

INTRODUCTION

Choosing a good selection criterion is one of the major problems in genetic evaluation of horses. The breeding objective is, in this paper, the ability to be successful in riding competitions (jumping, dressage, three-day-event) or in races (trot and gallop). But how should success be measured ? A "physical" measure of performance is not always available. Such a measure might be racing time for races or number of faults for riding competitions. Such data are not always collected and, furthermore, they may give a poor indication about the real level of the performance. A racing horse must be fast but it must, above all, adapt to particular conditions prevailing in each outing. This may explain the relatively low heritability of time performance of Thoroughbreds (Langlois, 1980a; Hintz, 1980). In Trotter races, time per kilometer in each race or best time in life are often used (Leroy et al., 1989 ; Ojala, 1989 ; Katona & Distl, 1989 ; Petzold et al., 1989 ; Klemetsdal, 1989 ; Arnason et al., 1989) with appropriate heritability but they are only a part of the quality of the winner. In the case of riding horse, it is difficult to assess the technical level of a jumping event. It depends not only on the height of the obstacles but, to a greater extent, on the difficulties encountered when approaching the obstacles and on the distance between the obstacles. None of these variables can be easily quantified. Therefore, information provided by the ranking of horses in each event deserves attention.

A SCALE FOR RANKING ?

Horse's life

A horse may or may not participate to any event. Then he will participate to a variable number of competitions. At each start, he may or not be in the first ones. If he does (for example, within the top 4 in trotting races or in the first quarter in Jumping competitions in France), he will earn money, i.e. he is "placed". As shown by Langlois (1983) money reward in a given race is allocated in an exponential way : for instance, the second horse earns half the amount of the first one, the third half that of the second, etc... The earnings of a horse in a race is then expressed by : $G = a\,x^{(k-1)}\,D$ with a being the proportion of the endowment given to the winner, x the rate of decrease of earnings with rank, k the rank of the horse in the race and D the total amount of money given in the race. For a total number of ranked horses equal to n, the constants a and x are such that : $a(1-x^n)/(1-x) = 1$.

All-or-none variable

The first potential measure is the proportion of starts for which the horse is "placed" (Ojala, 1989 ; Klemetsdal, 1989 ; Arnason et al., 1989, Hedebro et al., 1989), or for a stallion the percentage of horses with earnings divided by the number of horses which have participated to an event or the total number of progeny (Ojala, 1989). This is tantamount to look at the means of a discrete variable with value 1 for a "placed" horse and 0 otherwise. First of all, this type of variable should be analysed as a discrete one (Gianola & Foulley, 1983). But, in any case, such variables do not seem very informative, since they give the same value to any "placed" horse, whether he is a big champion or not, although information about such differences exists.

Earnings

The most frequently used criterion related to ranking is earnings, which has good heritability (0.15-0.30), or points allocated for a given rank (Huizinga H.A. & Van Der Meij G.J.W., 1988).

The first alternative when using earnings is to define the occurrence of the measure. Meinardus & Bruns (1987) used only "places" and so dropped all participations without earnings. Such data seem to be selected since it is not fair to penalize a horse with few earnings in an event and not the horse ranked behind him which did not earn any money. This introduces the problem of the value that has to be attributed to non-earning horses. The most usual choice is to give them a 0 value as when using earnings per start or total earnings (in one year or in life) since they are compared to total earnings obtained in all races. This choice is bothersome because it will have different effects according to the average level of endowment of the discipline (the higher it is, the worse the non-placed horses are considered). Thus, the endowment is not independent scale. Furthermore, a non-placed horse in a difficult event with high endowment is as penalized as one in an easier event... It should be less detrimental for the horse in the difficult event.

The second alternative is the choice of a transformation in order to have an optimal distribution of the variable. The most natural choice is the logarithm (Langlois, 1980b, Meinardus & Bruns, 1987) which leads to : $\text{Log}(G) = \text{Log}(D) + (k\text{-}1)\,\text{Log}(x) + \text{Log}(a)$ and therefore represents a linear function of true ranking when D has a suitable distribution. Other transformation to different powers were used with the same objective (square root (Ojala, 1989 ; Katona & Distl, 1989) ; Fourth root (Arnason, 1989) ; or optimal power found after a Box-Cox transformation (Klemetsdal, 1989)).

The third alternative is the choice of the duration of an elementary performance. The natural choice might be each start, with repetition of the performance for each start, but annual earnings are often preferred. Theoretically, earnings per start in one year (Arnason, 1989 ; Langlois, 1989 ; Ojala, 1989) should give the same evaluation as earnings at each start, but in practice it is not the case as horses participate in a different number of events and earn different amounts of money in each one. So, the evaluation will be more influenced by the regularity of the horse with annual criteria than with event criteria (the more regular, the better the horse). Another problem appear with total annual earnings (Langlois, 1980b ; Katona & Distl, 1989 ; Klemetsdal, 1989) as well as life earnings (Minkema, 1989) where earnings is not regressed on the number of starts. The problem is now to introduce the notion of longevity in a criterion of success. It seems preferable to separate the two traits, while taking into account the fact that the better the horse, the more frequently he starts.

In any case, two problems remain. A linear function of rank does not take into account the number of competitors. Moreover, endowment is the only measure of the level of the competition, but may not represent the real level of competition and depends on arbitrary rules. So, ranks must be used but without arbitrary references. Gillespie (1971) first proposed a "performance rate" but treated with a fixed model and without precise statistical predictor. We proposed here a new method assuming the existence of an underlying variable.

A NEW MODEL ASSUMING AN UNDERLYING VARIABLE

Model

The estimation of the animal breeding value using information on ranks is performed without arbitrary scale. In order to analyse such data, the concept of an existing underlying variable will be used as in Gianola & Foulley (1983) for estimating the breeding value with categorical data and in Henery (1981) for likelihood of outcome of a race. The horse's "real" performance (y), which cannot be measured, is viewed as a normal variable with expected value μ. $e=y-\mu$ is a normal $(N(0,\sigma_e^2=1))$ residual variable. These assumptions are reasonable for a trait with polygenic inheritance and under many small environmental influences. μ is modelled as the sum of the additive breeding value of the horse (u), nuisance environmental effects (b), and environmental effects common to all performances of the same horse (p): $y_{ijr} = \mu_{ij} + e_{ijr}$ and: $\mu_{ij} = b_i + u_j + p_j$. The vector of parameters to be estimated is noted $\Theta = (\mathbf{b}', \mathbf{u}', \mathbf{p}')'$. Only the position or rank of the performance relative to the other horses competing in the same event is observed and form the data (Y). Inferences are based on Bayes theorem :

$$f(\Theta/Y) \propto g(Y/\Theta)\, p(\Theta)$$

where $p(\Theta)$ is the prior density of Θ, $g(Y/\Theta)$ is the likelihood function and $f(\Theta/Y)$ is the posterior density of the parameters. Proportionality holds because the marginal density of **Y** does not vary with Θ.

Prior density

The vectors **b**, **u** and **p** are supposed to be mutually independent. Each of them follows a multivariate normal distribution : N(ß,V), N(0,G), N(0,H). Prior information about **b** is vague, so $V^{-1} \rightarrow 0$. Then the prior density of **b** is uniform and the posterior density of Θ does not depend on ß. $G = A\sigma^2_u$ where A is the relationship matrix and σ^2_u the additive genetic variance. **H** is a diagonal matrix with diagonal element equal to the variance of p_j. Then :

$$p(\Theta) \propto \exp(-1/2\mathbf{u}'\mathbf{G}^{-1}\mathbf{u})\, \exp(-1/2\mathbf{p}'\mathbf{H}^{-1}\mathbf{p})$$

Likelihood function

Given μ, the performances y are conditionally independent. The probability of obtaining the observed ranking in an event k can be written as :

$$P_k = \text{Prob}(\, y_1 > y_2 > \ldots > y_{(n-1)} > y_n\,)$$

$$P_k = \int_{-\infty}^{+\infty} \varphi(y_n-\mu_n) \int_{y_n}^{+\infty} \varphi(y_{n-1}-\mu_{n-1}) \ldots \int_{y_{i+1}}^{+\infty} \varphi(y_i-\mu_i) \ldots \int_{y_2}^{+\infty} \varphi(y_1-\mu_1)\, dy_1 \ldots dy_n$$

where n is the number of horses competing in the event k, subscripts for y and μ refer to the rank in the event and φ is the standard normal density. The likelihood function is equal to the product of the likelihood of each event :

$$g(Y/\Theta) = \prod_{k=1}^{m} P_k$$

Estimation of parameters

The posterior density of the parameters is :

$$f(\Theta/Y,G,H) \propto \left(\prod_{k=1}^{m} P_k\right) \exp(-1/2\mathbf{u}'\mathbf{G}^{-1}\mathbf{u})\, \exp(-1/2\mathbf{p}'\mathbf{H}^{-1}\mathbf{p})$$

The parameters are estimated as the mode of this posterior density. It is more convenient to use its logarithm $L(\Theta)$. The system which satisfies the first-order condition for the mode is not linear, so it must be solved iteratively, for example, using a Newton-Raphson (N.R.) algorithm.At each iteration :

$$-\left[\frac{\delta^2 L(\Theta)}{\delta\Theta\delta\Theta'}\right]_{\Theta=\hat{\Theta}^{[q-1]}} \Delta^{[q]} = \left[\frac{\delta L(\Theta)}{\delta\Theta}\right]_{\Theta=\hat{\Theta}^{[q-1]}}$$

Details on practical computation of these derivatives are in Tavernier (1990a).

Estimation of variance components

The estimator described above is the mode of posterior distribution of Θ conditionally on the variance matrices G and H. The structures of theses matrices are known but not the variance components σ_u^2 and σ_p^2. As suggested by Gianola et al (1986) and Foulley et al (1987a, 1987b), when these variances are unknown, inferences on Θ can be made from the distribution

$f(\Theta / Y, \sigma_u^2 = \hat{\sigma}_u^2, \ \sigma_p^2 = \hat{\sigma}_p^2)$, where the variances σ_u^2 and σ_p^2 are replaced by the mode of their marginal posterior distribution. The form of the distribution of $\Theta / Y, \sigma_u^2, \sigma_p^2$ is not known but may be approximated by a normal one (Harville & Mee, 1984). Then, the mode of the marginal posterior distribution of the variances can be found with the iterative scheme involving the expression, at iteration t :

$$\left[\sigma_u^2\right]^{[t+1]} = \left[\left(\mathbf{u}'A^{-1}\mathbf{u} + tr(A^{-1}C_{uu}) \right) / \omega\right]^{[t]}$$

-$u^{[t]}$ is the solution at convergence of the N.R. equations evaluated at $\sigma_u^2 = \sigma_u^{2\,[t]}$,

-$C_{uu}^{[t]}$ is the part of the inverse of the matrix of the N.R. equations at convergence corresponding to u

-ω is the number of elements of u

A similar expression can be found for **p**. These expressions are identical to the EM algorithm for estimation of σ_u^2 by REML under normality.

Application

At present, two applications have been performed. The first one used a small file of two year old French trotters (Tavernier, 1989). The second was on French Jumping horses in 1987 (Tavernier, 1990b). The repeatability was estimated in the latter case to be 0.29. 19798 horses were evaluated from 6513 events and 265077 participations. There were 3 333 668 coefficients (2%). The correlation between these estimations and logarithm of annual earnings and logarithm of annual earnings per start were 0.75 and 0.82 respectively.

CONCLUSION

Earnings remains a good criterion estimator but is only a transformation of rank into a scale suitable for statistical purpose. The new model of analysis of ranks is more difficult to compute but leads to an evaluation which accounts for the place of the horse, the number of horses in the event, the level of each one of the competitors and the total number of events per horse, using multiple comparisons.

REFERENCES

ARNASON, T., BENDROTH, M., PHILIPSSON, J., HENDRIKSSON, K. and DARENIUS, A. 1989. EAAP publication 42 : 106-130.

FOULLEY, J.L., GIANOLA, D., PANCHENAULT, D. 1987a. Génét. Sél. Evol. 19 : 83-102.

FOULLEY, J.L., IM, S., GIANOLA, D., HOSCHELE, I. 1987b. Génét. Sél. Evol. 19 : 197-204.

GIANOLA, D. and FOULLEY, J.L. 1983. Génét. Sél. Evol. 15 : 201-224.

GIANOLA, D., FOULLEY, J.L., FERNANDO, R.L. 1986. Génét. Sél. Evol. 18 : 485-498.

GILLESPIE, R.H. 1971. Thoroughbred Rec. 17 : 961-977.

HARVILLE, D.A. and MEE R.W. 1984. Biometrics, 40 : 393-408.

HEDEBRO-VELANDER, I., GRAAF, K. and PHILIPSSON, J. 1989. 40th Annual Meeting of EAAP. Dublin. 327-328.

HENERY, R.J. 1981. J. R. Statist. Soc. 43 : 86-91.

HINTZ, R.L. 1980. J. Anim. Sci. 51 : 582-594.

HUIZINGA, H.A. and VAN DER MEIJ, G.J.W. 1988. Proc. Coll. Int. CEREOPA, Paris, 9 Mars 1988 .

KATONA, O. and DISTL, O. 1989. EAAP publication 42 : 55-61.

KLEMETSDAL, G. 1989. EAAP publication 42 : 95-105

LANGLOIS, B. 1980a. Livest. Prod. Sci. 7 : 591-605.

LANGLOIS, B. 1980b. Ann. Génét. Sél. Anim. 12 : 15-31.

LANGLOIS, B. 1983. 38th Annual Meeting of EAAP. Madrid.

LEROY, P.L., KAFIDI, N. and BASSLEER, E. 1989. EAAP publication 42 :3-17

MEINARDUS, H. and BRUNS, E. 1987. 38th Annual meeting of EAAP. Lisbon. 44-45.

MINKEMA, D. 1989. EAAP publication 42 : 82-94.

OJALA, M. 1989. EAAP publication 42 : 18-26.

PETZOLD, P., BERGFELD, U. and SCHWARK, H.J. 1989. EAAP publication 42 :62-66.

TAVERNIER, A. 1989. 40th Annual meeting of EAAP. Dublin. 325-326.

TAVERNIER, A. 1990a. Genetic evaluation of horses based on ranks in competitions. Submitted to Genet. Sel. Evol.

TAVERNIER, A. 1990b. Estimation of Breeding Value of Jumping Horses from their Ranks. Submitted to Livest. Prod. Sci.