1986

CRITERIA FOR EVALUATING A SELECTION SCRENE: SOME PROPOSALS

J. M. Elsen
I.N.R.A.

L. Bodin
I.N.R.A.

G. Ricordeau
I.N.R.A.

CRITERIA FOR EVALUATING A SELECTION SCHEME: SOME PROPOSALS

J.M. ELSEN, L. BODIN, G. RICORDEAU, FRANCE

Station d'Amélioration Génétique des Animaux, I.N.R.A.
Toulouse, B.P. 27, 31326 CASTANET-TOLOSAN CEDEX, FRANCE

SUMMARY

It is possible to describe a realized selection by means of an indicator, f(x): probability for an individual of value x to be selected. Various models for this indicator are proposed, in the univariate case, and on the assumption that individuals are ranked on a linear index of measured variables. Estimators are defined on the basis of these models for rating traits controlled during selection. A numeric example with ewes is given.

KEY WORDS: Selection scheme, selection index, selection differential, realized selection.

INTRODUCTION

In order to measure the efficiency of a selection scheme, it is necessary to estimate how breeders rate the various traits when choosing animals. Only those traits that have been controlled can be taken into account, and one must examine not only their relative importance (each in relation to the other), but also the role of a "deviation" representing all the selection criteria that are used without being known. Several studies have been published on this topic concerning dairy cattle (ROBERTSON, 1966; HINKS, 1966, 1975; SCHAEFFER and BURNSIDE, 1974; BERGER et al., 1973; WHITE and NICHOLS, 1965; ROBERTSON and BARKER, 1966; ALLAIRE and HENDERSON, 1966a, 1966b, 1966c). More recently, ELSEN et al. (1985) have extended these methods to an example of a selection scheme for suckling ewes.

I - METHODS

A - Univariate Models

1 - Notation

Pj animal populations: tested (Pt), retained (Pr) and eliminated (Pe)

Nj the number in each category.

Xi random variable: value of the ith trait. Xi can be divided into Cik classes of stock Nijk (Nitk for tested, etc.).

$rik = \frac{Nirk}{Nitk}$: intra-class selection rate. $ri = r = \frac{Nr}{Nt}$

μij and $\sigma^2 ij$: averages and variances of trait i in population Pj.

2 - Studied Models

a) - Generalities

The model, in all cases, is a description of indicator $f(x_i)$: probability for an individual of value x_i to be retained. It is possible to compare the models by evaluating their likelihood (or their logarithms) measured giving parameters the values that maximize it.

Table 1 sketches a few possible hypotheses. Other examples are found in Elsen et al. (1985). Note that selection hypotheses for high (resp. low) values follow from Table 1 assuming: for Model 2, very high (resp. low) λ_i for Models 3 and 4, very small (resp. big) λ_i.

b) - Supplementary Criteria

In addition to their likelihood and the optimal values of the parameters that characterize models, there are other criteria that can give indications on the realized selection.

- Model 4 goes back to those of ROBERTSON (1966) and ALLAIRE and HENDERSON (1966a). These authors use the ρi correlation between X_i and $H_i = X_i + E_i$ as a criterion. This is a non-measurable variable that the breeder uses to select among his animals possibly by a truncation procedure (a tested animal is retained if its h_i value is greater than a γ_i threshold). This correlation can be estimated with $\widehat{\rho i} = ((\widehat{\mu r}i - \widehat{\mu c}i)/\widehat{\sigma i}c)/\delta i = \widehat{D}i/\delta i$

where δi is the maximum value of the difference between retained and tested animals, therefore for $H_i = X_i$. In the case where X_i obeys a normal law, we have $i = i(r).\sigma ic$, where the selection intensity $i(r)$ is $\exp(-\gamma_i^2/2)/\sqrt{2\pi}.r$

- in this context, we define:

* the $\widehat{qi}$ realized selection rate, such that $\widehat{D}i = i(\widehat{qi})$. We will have $r = \widehat{qi}.\widehat{\rho i}$

* a new way of measuring the importance of X_i, $\widehat{c}i = (1-\widehat{qi})/(1-r)$, that varies between 0($X_i$ plays no role in selection) and 1(only X_i plays a role)

* finally, $1-\widehat{\rho i}$ can be interpreted as a lack of benefit due to selecting on E_i.

B - Multivariate Models

1 - Notation and Hypotheses

We will examine here only the case that ALLAIRE and HENDERSON (1966b,c) and BERGER et al. (1973) studied. This is a selection favoring individuals with a high linear index H for the measured variables. It applies to the multivariate case our continuous selection model (Model 4). $H = \Sigma b_i x_i + E$ where E is a non-measured "deviation" variable, that covers all non-controlled traits and randomness in selection. Since b_i are defined to within one proportionality coefficient, it is always possible to assume that E is independent from the X_i, and of unit variance. By hypothesis, the H index of retained reproducers goes beyond a threshold. We also assume multinormality of the X_i, V_c being the matrix of variances and covariances of these variables in Pc.

2 - Weighting Coefficients

a) - Generalities

ALLAIRE and HENDERSON (1966b), restated later on by BERGER et al. (1973) have presented their solution as the result of maximizing the correlation between $H=b'x+E$ and a measurable index $J=\Sigma\beta_i X_i=\beta'X$. Elsen et al. (1985) have shown that it is not necessary to define this J index in order to describe selection. Following our hypotheses, $E(\Delta\mu)=i(r).Vc.b/\sqrt{b'.Vc.b+1}$ and if (as ALLAIRE and HENDERSON do) we replace expectations $E(\Delta\mu)$ with their realizations, the result is

$$\hat{b}=\frac{Vc^{-1}.\widehat{\Delta\mu}}{\sqrt{i(r)^2-\hat{D}^2}} \quad \text{where} \quad \hat{D}^2=\widehat{\Delta\mu}.Vc^{-1}.\widehat{\Delta\mu}$$

b) - Criteria for Rating the Importance of Traits in Selection

This is an extention of the criteria proposed in the univariate case:

- the realized selection rate $\hat{q}$ is the rate that would have given the $\widehat{\Delta\mu_i}$ if selection had been carried out on $\hat{J}$ (or on the $\hat{I}=\hat{b}'.x$ part of $\hat{H}$). The result is $i(\hat{q})=\hat{D}$.
- the selection rate on the other criteria is $\hat{p}=r/\hat{q}$
- the relative importance of controlled traits can be measured with $\hat{c}=(1-\hat{q})/(1-r)$.
- $1-\hat{\rho}$ is the lack of benefit due to selecting on E. The result is $\hat{\rho}=i(\hat{q})/i(r)=\sqrt{var(\hat{I})/var(\hat{H})}$.
- finally, the importance of the ith X variable can be measured with the correlation between X and H (which equals $(\widehat{\Delta\mu_i}/\sigma_{ic})/\hat{D}$), or with the quantity $(\hat{c}_i-\hat{c}_{i-1})/(1-\hat{c}_{i-1})$

c) - Other Approaches

- Elsen et al. (1985) have shown that the previously defined $\hat{b},\hat{\beta}$ coefficients of $\hat{H}$ and $\hat{J}$ indexes are proportional to the coefficients of discriminant analysis between Pr and Pe. On the other hand, the quantity $\hat{D}^2$ is the Mahalanobis distance between tested and retained, and the square of correlation $\hat{\rho}^2$ can be interpreted as the ratio of this distance to the maximum distance.
- we can also use the continuous selection modeling presented above. Here, variable X_i is replaced with $I = b'.X$, H_i with H and E_i with E. The model becomes $f(I)=\Pi(\Lambda-I)$ and it is necessary to adjust for Λ and b.

II - APPLICATION TO A EWE SELECTION SCHEME

A - Presentation of the Analyzed Case

Each year, breeders of Lacaune ewes who practice the Ovitest Cooperative's selection scheme on prolificacy, collectively choose males to be progeny tested. We have analyzed this choice and considered 8 controlled variables (see Table 2). The true selection rate is $r=0.23$ on the average for 1980 and 1981. Table 2 presents some basic statistics on these 8 traits that show:

- few differences between retained and tested for NC and P2
- a higher average among the retained for IL, P1, DT, NA and IP with an increase in variance for the last two variables
- a preference for extremes for MN.

RESULTS

1 - Univariate Analyses

Table 3 shows the values of the various parameters proposed in order to rate selection. The last four pertain to Model 4. Probabilities of sameness of distributions (χ^2 of homogeneity) verify the interest breeders have in maternal traits (MN, NA, IL, IP) and to a lesser extent in DT.

Parameters in Model 4 have the same tendency as the χ^2 of homogeneity for the last seven traits but suggest no particular ram choice on MN. This result can be explained by the aforementioned breeders' preference, regarding this variable, for extreme animals. This preference will be confirmed further on. Comparing likelihoods L1 (no selection) and L2 (general model) shows on one hand, that NC and P1 are of little importance for breeders, and on the other hand, that IP and NA range first. We found a sub-model of the same explanatory value as the general model for only three variables:

- MN: continuous selection for extremes (Model 4). The optimal indicator is $f(MN) = 1 - \Pi(.83\ MN - .522) + \Pi(1.033\ MN - 3.93)$
- NA and IP: selection aimed at greater values. In effect, the optimal value of parameter λi for these two variables is very high.

2 - Multivariate Analysis

2 to 8 variables can be taken into account simultaneously in the linear index. For each i number of variables, Table 3 gives the characteristics of the most explanatory J index. Compared with results of univariate analysis, conclusions regarding trait ratings are similar: maternal aspects (NA, IL, IP) come first, then comes weight P1. The last 4 traits have a negligible role in the description of the realized selection (neither bi nor (X_i , J) correlations are modified between J4 and J8). Finally, Elsen et al. (1985) show that the optimality criterion for linear combinations (discriminant analysis or maximum likelihood of $f(\Lambda - I)$ indicator has little effect on the numeric result). Expressed in standard deviation unit of traits, the optimal weights are about 1 for NA, .67 for IL, .6 for IP, .45 for P1. The index is first correlated with NA and IP, and then with IL.

CONCLUSION

The methods described here should clarify how selection schemes operate and consequently help decide whether to modify them or not. Nevertheless, they have limits: only controlled traits can be analyzed; other models are conceivable; their results regard only one stage in selection and will have to be supplemented by combining analysis of the various stages in each path of genetic improvement.

REFERENCES

ALLAIRE, F.R. and HENDERSON, C.R. 1966a. J. Dairy Sci. 49(11), 1426-1434.
ALLAIRE, F.R. and HENDERSON, C.R. 1966b. J. Dairy Sci. 49(11), 1435-1440.
ALLAIRE, F.R. and HENDERSON, C.R. 1966c. J. Dairy Sci. 50(2), 194-200.
BERGER, P.J., HARVEY, W.R. and RADER, E.R. 1973. J. Dairy Sci. 56, 805-811.
ELSEN, J.M., BODIN, L. and GOUPILLON, C. 1985. Submitted to Genet. Sel. Evol.
HINKS, C.J.M. 1966. Anim. Prod. 8, 467-480.
HINKS, C.J.M. 1975. Anim. Prod. 20, 337-343.
ROBERTSON, A. 1966 Anim. Prod. 8, 95-108.
ROBERTSON, A. and BARKER, J.S.F. 1966. Anim. Prod. 8, 241-252.
SCHAEFFER, L.R. and BURNSIDE E.B. 1974. J. Dairy Sci. 11, 1394-1400.
WHITE, J.M. and NICHOLS, J.R. 1965. J. Dairy Sci. 48, 468-474.

- TABLE 1 -

Compared univariate models

Tested hypothesis	Corresponding model	Parameters to be estimated	Method used
I : no selection on Xi	$f(x_i) = q \quad \forall x_i$	q	$\hat{q} = \frac{N_r}{N_c}$
II : selection for extremes : general case	$f(x'_i) \leqslant f(x''_i)$ if $\lambda_i \geqslant x'_i \geqslant x''_i$ $f(x'_i) \geqslant f(x''_i)$ if $\lambda_i \leqslant x''_i \leqslant x'_i$	set of the $f(x_i)$	Dynamic programming
III : selection for extremes : choosing by truncation	$f(x_i) = 1 - q_i$ if $x_i \in [\lambda_{i1}, \lambda_{i2}]$ $= q_i$ elsewhere	λ_{i1}, λ_{i2} and q_i	Trial and error
IV : selection for extremes : continuous choice	$f(x_i) = 1 - \Pi\left(\frac{\lambda_{i2} - x_i}{\sigma\varepsilon_i}\right) + \Pi\left(\frac{\lambda_{i1} - x_i}{\sigma\varepsilon_i}\right)$	λ_{i1}, λ_{i2} and σ_{Ei}	Newton - Raphson

Π is an ascending function of x_i. We choose for Π the distribution function of the standardized normal distribution

- TABLE 2 -

Analysed variables

Variables regarding the dam	Averages tested		Averages selected	Standard deviation tested		Standard deviation selected
MN : birth type	2.044	=	2.021	.65	< (1%)	.87
NC : lambing sequence number of the dam	4.79	=	4.78	1.57	=	1.57
NA : mean number of lambs per year of the dam	2.45	< (1%)	2.70	.39	< (1%)	.49
IL : milk production index of the dam	.35	< (1%)	.62	.91	=	.92
IP : prolificacy index of the dam	1.45	< (1%)	1.49	.08	< (1%)	.10
Variables regarding the individual						
P1 : weight on day 40	224.54	<	230.13	29.92	=	30.45
P2 : weight on day 140	480.24	=	486.44	49.11	=	55.03
DT : testis diameter	59.17	< (5%)	60.40	6.29	=	6.42

- TABLE 3 -

Univariable selection parameters

Trait	MN	NC	NA	IL	IP	P1	P2	DT
Sameness probability of distributions	0	.67	0	0	0	.42	.86	.07
Models likelihood								
- L1	227	227	227	227	227	227	227	227
- L2	215	223	196	212	206	221	215	215
- L3	224	226	215	229	219	228	225	228
- L4	215	227	202	221	212	224	226	224
Realized selection rate $\hat{q}$	.99	.995	.59	.83	.68	.90	.94	.90
Rate $\beta = r/\hat{q}$	.22	.22	.37	.26	.32	.24	.23	.24
Coefficient $\hat{c} = \frac{1 - \hat{q}}{1 - r}$	.02	.01	.53	.22	.40	.12	.08	.13
Lack of gain $100(1 - \hat{D}/g)$	97	100	51	78	63	86	91	86

- TABLE 4 -

Multivariate analysis : parameter values according to the number of traits

- T A B L E 4 -

Multivariate analysis : parameter values according to the number of traits

Number of variables studied	Discriminant variables MN	NC	NA	IL	IP	P1	P2	DT	Realized selection rate	Relative importance $\hat{c}$	Good ranking probability	Possible gain $1 - \hat{e}$	Maximum Likelihood -L
1			X						.586	.529	.66	.51	196.4
2			X	X					.541	.585	.68	.46	194.8
3			X	X	X				.516	.618	.69	.43	191.4
4			X	X	X	X			.489	.652	.70	.40	186.7
8	X	X	X	X	X	X	X	X	.487	.655	.70	.35	186.1

- T A B L E 5 -

Characteristics of the 4 variable selection indexes

Variables	MN	NC	NA	IL	IP	P1	P2	DT
J4 coefficients :								
discriminant analysis			1.	.26	3.01	.006		
maximum likelihood			1.	.30	3.06	.007		
Correlations (X, J4) :	-.026	-.035	.817	.371	.632	.229	.093	.174

λ parameter of the optimal indicator is -10.07