Prediction of Selection Response for Threshold Dichotomous Traits

Jean L. Foulley

INRA, Unité de Génétique quantitative et appliquée, Equipe de Génétique statistique, Centre de recherches de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex, France

Manuscript received May 11, 1992
Accepted for publication August 14, 1992

ABSTRACT

This paper presents a formula to predict expected response to one generation of truncation selection for a dichotomous trait under polygenic additive inheritance. The derivation relies on the threshold liability concept and on the normality assumption of the joint distribution of additive genetic values and their predictors used as selection criteria. This formula accounts for asymmetry of response when both the prevalence of the trait and the selection rate differ from 0.5 via a bivariate normal integral term. The relationship with the classical formula $R = ipa_c$ is explained with a Taylor expansion about a zero value of the correlation factor. Properties are illustrated with an example of sire selection based on progeny test performance which shows a departure from usual predictions up to 15–20% at low (0.05) or high (0.95) selection rates. Univariate approximations and extensions to several paths of genetic change are also discussed.

QUANTITATIVE genetics as used in animal breeding relies basically on the normality assumption of both genetic and phenotypic values, and consequently, on the hypothesis of linear regression of genetic on phenotypic values. This is especially true for predicting selection response, the theory of which lies, in large populations, on selection index properties and demographical considerations.

Discrete traits cannot be properly analyzed using such assumptions even if described as threshold traits with a normal liability scale (DEMPSTER and LERNER 1950; GIANOLA 1982).

For binary traits such as fertility, calving difficulty and viability whose economic importance is far from negligible, applying the classical formula $R = ipa_c$ (FALCONER 1989) for forecasting genetic response to one generation of selection as the product of selection intensity ($i$), times the square root of selection accuracy ($\rho_{ip}$) and the additive genetic standard deviation ($\sigma_i$), may lead to unreliable results (DANELL and RÖNNINGEN 1981). In such cases, these authors recommended carrying out a simulation study prior to application of selection index.

An alternative would be to employ an analytical formula of prediction. Formulae based on univariate approximations were already proposed: see e.g. FALCONER (1989), MENDEL and ELSTON (1974). The purpose of this paper is to derive a more general expression for such predictions when selection is based on estimated breeding values and to use it for discussing the accuracy of different approximations.

THEORY

For the sake of simplicity, consider a dichotomous trait with a phenotype coded as $y = 0$ or 1, 1 being for individuals showing the attribute of interest. This trait is assumed to be determined by a threshold liability process involving a purely additive model on the underlying scale so that

$$x = \mu_x + a + e$$

with $a \sim N(0,\sigma_a^2)$, $e \sim N(0,\sigma_e^2)$ and $\text{Cov}(a,e) = 0$, $\mu_x, \sigma_a, \sigma_e$ designating the mean, phenotypic, genetic and environmental components on the underlying scale.

Following ROBERTSON and LERNER (1949) and ROBERTSON (1950), the genetic value ($g$) on the observed scale can be defined as the proportion of phenotypes having the attribute among individuals having the same genetic value $a$, i.e.,

$$g = \Pr(y = 1|\mu_x, a) = \Pr(x \geq \tau|\mu_x, a)$$

where $\tau$ is the threshold value. Now, the conditional distribution of $x$ given $\mu_x$ and $a$ say $x|\mu_x, a$ is $N(\mu_x + a, \sigma_x^2)$, so that

$$g = \int_{\tau-\mu_x-a}^{+\infty} \phi(t) \, dt = \Phi[-(\tau - \mu_x - a)/\sigma_x]$$

where $\phi(.)$ and $\Phi(.)$ are the density and the cumulative density functions (CDF) of the standardized normal distribution respectively. Setting the origin to be the threshold, this reduces to:

$$g = \Phi[(\mu_x + a)/\sigma_x].$$

Generally $a$ is not known and one may alternatively condition on $\hat{a}$, the estimated breeding value (EBV). Then, the expression in (2) has to accommodate the uncertainty in $a$ given $\hat{a}$; this is achieved by taking its expectation with respect to the conditional density

$$g = \mathbb{E}[g|\hat{a}] = \int g \, \phi(a|\hat{a}) \, da.$$
function of distribution of sexes and there is no sex dimorphism, it suffices to replace EBVs in male and female candidates (\(\hat{a}_c\) and \(\hat{a}_w\) respectively) by an equivalent variable \(\hat{a}_c = \sqrt{3} (\hat{a}_c + \hat{a}_w)\) such that \(a|\hat{a}\) or \(a|\hat{a}_w\) is \(N_1(1 + (\rho^2/2)|\sigma_a^2),\) i.e., by using \(\theta = 1\) and \(\rho^2 = \rho_{c,w}^2 = \rho_{c,f}^2\).

If use is made of the more commonly tabulated quantity \(L(x,y) = \Pr(X > x, Y > y)\) (JOHNSON and KOTZ 1972), formula (7) is then expressed as \(R^{(+)\dagger}\) as follows:

\[
R^{(+)\dagger} = \left( \frac{\Phi_2(\mu, \eta, \rho \theta) - \Phi(\mu)\Phi(\eta)}{\Phi(\eta)} \right) \frac{\rho}{\theta}
\]

(9) entails the equivalence between upward selection in favor of the trait coded as 1 and downward selection against the trait at \(1 - \Pi\). This can be shown algebraically as follows:

Given \(\Phi^{-1}(1 - \Pi) = -\mu,\) expected response to downward selection can be written as

\[
R^{(-\Pi)} = \left( \frac{\Phi_2(\mu, \eta, -\rho \theta) - \Phi(\mu)\Phi(\eta)}{\Phi(\eta)} \right) \frac{\rho}{\theta}
\]

(10)

RESULTS

The main features of formula (7) are: (i) asymmetry of response thereby generated and consequently, (ii) departure from the usual prediction equation \(R = \rho \theta \sigma_c\).

Asymmetry of response is expected as a consequence of the interaction between fraction selected and incidence of the binary trait as factors determining response to selection for threshold traits (FALCONER 1989).

To illustrate this, let us consider the following simple example dealing with mass selection of \(Q = 0.10\) (e.g., 10 individuals out of 100 candidates) within a monoeocious population for a perfectly heritable trait on the liability scale with a prevalence of \(\Pi = 0.05\). In upward selection among the ten individuals retained, five are displaying the attribute (\(y = 1\)) and the five remaining ones are not, so that the frequency of the trait in the candidates and in their progeny will be \((5 \times 1) + (5 \times 0)/10 = 0.50\) thus providing a response \(R^{(+)}\) equal to 0.50. In selection against the attribute, all the individuals chosen will have the alternative attribute (\(y = 0\)) and the response will be \(R^{-} = -0.05\), thus showing a marked asymmetry \(A = |R^{(+)} - R^{-}| = 0.55\) in favor of upward selec-
tion. More generally, it can be shown that, for traits with incidence $\Pi$ less than one half, the asymmetry $A$ in this example, is given by:

$$A = (1 - Q)/Q$$  \hfil (11a)$$
for a selection rate $Q$ ranging from $\Pi$ to $1 - \Pi$, and

$$A = (1 - \Pi)/\Pi$$  \hfil (11b)$$
for $Q \leq \Pi$, or

$$A = \Pi/(1 - \Pi)$$  \hfil (11c)$$
for $Q > 1 - \Pi$.

Hence, apart from $Q = 1/2$ or $\Pi = 1/2$, asymmetry of response is the rule.

Formulas (7) and (9) account for such a phenomenon. Expected response around $\Pi = \Psi(\mu)$ is asymmetric because the quantity $\Psi(r) = \Phi_2(\mu, \eta; r) - \Phi(\mu)\Phi(\eta)$ is not an odd function, i.e. such as $\Psi(-r) = -\Psi(r)$. In the Taylor expansion of an odd function, all even order derivative terms evaluated about zero must be null. From TALLIS' (1962) result: $d\Psi(r)/dr = \phi_2(\mu, \eta; r)$, $(\phi_2(\mu, \eta; r)$ being the standardized bivariate normal density function with arguments $\mu, \eta$ and correlation $r)$, the Taylor expansion of $\Psi(r)$ about $r = 0$ is

$$\Psi(r) = r\phi(\mu)\phi(\eta) + (r^2/2)\mu\eta\phi(\mu)\phi(\eta) + o(r^3).$$  \hfil (12)$$
Here $\Psi(0) = 0$, but the second order term does not generally vanish except for $Q = 1/2$ or $\Pi = 1/2$. Moreover, keeping the first order term only leads to the following expression of expected response:

$$R(+^*) \equiv \rho h\phi(\mu)\phi(\eta)/\Phi(\eta).$$  \hfil (13)$$
Notice that the ratio $\phi(\eta)/\Phi(\eta)$ is the selection intensity ($i$) and $h\phi(\mu)$ the genetic standard deviation ($\sigma_c$) on the binary scale, so that formula (13) reduces to

$$R(+^*) \equiv i\sigma_c.$$  \hfil (14)$$
which is the very well known formula for predicting expected genetic change after one generation of selection for a continuous trait. The same reasoning applies as well to $R(-^*)$ [see (10)] leading to $R(-^*) \equiv -i\sigma_c$. This argument shows how formula (7) is related to the usual formula and how much it differs from it as a function of $Q$, $\Pi$, $h^2$ and $\rho^2$. In particular, it can be inferred from formula (12) that, for small values of $ph$, the expected response calculated with the usual formula $i\sigma_c$ is underestimated if the product $\mu\eta$ is positive (both $Q$ and $\Pi$ lower than $1/2$ or higher than this value) and overestimated otherwise.

All properties of formulas (7–9) vs. (13–14) are illustrated in an example of selection of progeny-tested sires on their EBVs for liability ($h^2 = 0.20$) of a dichotomous trait with various prevalence levels ($\Pi = 0.05, 0.25, 0.50, 0.75$ and $0.95$), selection rates ($Q = 0.10, 0.50$ and $0.90$) and progeny group sizes ($n = 25, 75$ and $125$).

Estimation of breeding values was assumed to be made according to GILMOUR’S procedure (GILMOUR, ANDERSON and RAE 1985) in which accuracy ($\rho^2$) depends on such population features as $\Pi$ and $h^2$ but not on the breeding value of the candidate as in GIANOLA and FOULLEY (1983). In that case, it can be shown (APPENDIX B) that $\rho^2 = n/(4(n + k))$ where $k = [\Pi(1 - \Pi) + t0^2(\mu)/t0^2(\mu)]$, $t$ being the intra-class coefficient (here $t = h^2/4$) and $1/2$ represents the squared genetic relationship ($\rho^2$) between sire and offspring in which response to selection is measured. It is also worth mentioning that $k$ can also be expressed as usual as $k = (4/h^2) - 1$ with the heritability coefficient calculated on the observed scale according to ROBERTSON (1950) and DEMPSTER and LERNER (1950). It can be seen (APPENDIX B) that this expression for $k$ is consistent with its classical interpretation in formula (14).

Results are given in Table 1 with bivariate normal integrals computed from Dutt’s algorithm as described in DUCROCQ and COLLEAU (1986). As usual in progeny testing, selection intensity has more influence on the rate of response than progeny group size. Also as expected, values of response are strongly dependent on the incidence level with a maximum at $\Pi = 1/2$. Qualitatively, asymmetry of response ($A = |R(+^*)/R(-^*)|$) follows the same pattern as shown in formulae (1 labc). For instance, for a fraction selected $Q$ of 1 in 10, $A > 1$ for $\Pi = 0.05$ and $\Pi = 0.25$ whereas $A < 1$ at the same incidence levels for $Q = 0.9$. The further the

| TABLE 1
<p>| Expected response to one generation of selection of progeny-test ed sires on the estimated breeding value for liability |</p>
<table>
<thead>
<tr>
<th>Progeny group size</th>
<th>Fraction selected</th>
<th>Incidence of the trait (p. 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 U5.95</td>
<td>25 U25.15</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>25</td>
<td>0.1</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td>(1.90)</td>
<td>(7.98)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.86</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(3.63)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.21</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>75</td>
<td>0.1</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>(2.74)</td>
<td>(10.24)</td>
</tr>
<tr>
<td>0.5</td>
<td>1.24</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(4.66)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.26</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(1.14)</td>
</tr>
<tr>
<td>125</td>
<td>0.1</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>(3.10)</td>
<td>(10.98)</td>
</tr>
<tr>
<td>0.5</td>
<td>1.40</td>
<td>5.01</td>
</tr>
<tr>
<td></td>
<td>(1.41)</td>
<td>(4.99)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.29</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(1.22)</td>
</tr>
</tbody>
</table>

Heritability on the liability scale assumed to be equal to 0.20. $U \times U100$: equivalence between upward selection at incidence $x = 100$ and downward selection at $100-x = 100$. Figures within parentheses computes according to the usual formula $ipac$.  


incidence departs from $\frac{1}{2}$, the larger is the asymmetry which also increases with progeny group size. For $n = 125$ and $Q = 0.1$, upward selection exceeds downward selection by 19 and 44 percent at $\Pi = 0.25$ and 0.05, respectively, the same being observed for downward selection at $Q = 0.9$.

As expected from formula (13), predictions based on the usual formula $ipu$ are very close to those proposed here for $\Pi = \frac{1}{2}$ whatever the selection rate, or for any $\Pi$ at $Q = \frac{1}{2}$. Otherwise, the more $\Pi$ departs from $\frac{1}{2}$, the larger is the relative error of prediction by using $ipu$ which amounts up to 15 to 20% for $\Pi = 0.05$ or $\Pi = 0.95$ (Table 2). This is all the more important as the heritability ($h^2 = 0.20$) and the accuracy ($p^2 = 0.15 - 0.20$) are rather small which makes the Taylor expansion a priori better.

**DISCUSSION**

**Approximations:** A simple explicit formula for approximating the bivariate normal integral $L(a,b;r)$ was recently proposed by Cox and Wermuth (1991). Because the expression derived for this approximation is not symmetric in $a$ and $b$ whereas $L(a,b;r)$ is so, three different formulae have to be employed depending on values of the arguments. For instance, for $r$ and $a$ positive, and $a > b$, the formula to apply is

$$L(a,b;r) \approx \Phi(-a)\Phi\left[\frac{r(a) - b}{(1 - r^2)^{1/2}}\right]$$

where $i(a) = \phi(a)/\Phi(-a)$. Here, $a = -\eta$, $h = -\mu$ and $r = \rho h$, so that

$$R^{(s)} \equiv \Phi((\mu + i\rho h)/(1 - \rho^2h^2)^{1/2}) - \Phi(\mu)$$

and it can be easily checked that (15) has a first order Taylor expansion about $\phi h = 0$ equal to $ipu$. Formulas similar to (15) were also proposed by Falconer (1989) and by Mendell and Elston (1974).

Notice also that (15) corresponds to formula (A.3) of Appendix A with $a$ replaced by its expectation $ipu$. Doing so we proceed in the denominator with a phenotypic variance in offspring $u^2(1 - p^2h^2) = u^2 + (1 - p^2\sigma_a^2)$ as if there were no uncertainty about EBVs of selected candidates. This suggests another approximation to adjust this phenotypic variance for a residual variation in $a$ given that truncation selection of the top fraction $Q$ of candidates was performed, i.e. by using $a^2[1 - i(i - \eta^2)h^2]$ (Cochran, 1951). Scaling the new mean in $a$, i.e., $\sigma_a(\mu + i\rho h)$, with this adjusted standard deviation and proceeding as if norm-

---

**Table 2**

**Exact vs. approximate values of response to one generation of selection of progeny-tested sires on the estimated breeding value for liability**

<table>
<thead>
<tr>
<th>Method</th>
<th>Selection rate</th>
<th>5</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivariate$^a$</td>
<td>0.1</td>
<td>3.675</td>
<td>11.904</td>
<td>13.977</td>
<td>10.013</td>
<td>2.556</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.397</td>
<td>5.011</td>
<td>6.427</td>
<td>5.011</td>
<td>1.397</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.284</td>
<td>1.113</td>
<td>1.553</td>
<td>1.323</td>
<td>0.408</td>
</tr>
<tr>
<td>Approx A$^b$</td>
<td>0.1</td>
<td>3.650</td>
<td>11.865</td>
<td>14.029</td>
<td>10.013</td>
<td>2.555</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.400</td>
<td>5.039</td>
<td>6.487</td>
<td>5.013</td>
<td>1.394</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.286</td>
<td>1.137</td>
<td>1.634</td>
<td>1.318</td>
<td>0.406</td>
</tr>
<tr>
<td>Approx B$^c$</td>
<td>0.1</td>
<td>3.674</td>
<td>11.908</td>
<td>13.982</td>
<td>10.013</td>
<td>2.555</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.394</td>
<td>5.019</td>
<td>6.438</td>
<td>5.013</td>
<td>1.394</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.280</td>
<td>1.119</td>
<td>1.569</td>
<td>1.329</td>
<td>0.404</td>
</tr>
<tr>
<td>$ip\sigma_c$</td>
<td>0.1</td>
<td>3.098</td>
<td>10.984</td>
<td>14.040</td>
<td>10.984</td>
<td>3.098</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.409</td>
<td>4.994</td>
<td>6.383</td>
<td>4.994</td>
<td>1.409</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.344</td>
<td>1.220</td>
<td>1.560</td>
<td>1.220</td>
<td>0.344</td>
</tr>
</tbody>
</table>

Heritability on the liability scale assumed to be equal to 0.20 and progeny group size equal to 125; otherwise, same assumptions as in Table 1. $^a$ Bivariate integrals computed via Dutt's algorithm (Duquette and Colleau 1986). $^b$ Approx A: algorithm given by Cox and Wermuth (1991), formulae on page 264 (simplest version). $^c$ Figures within parentheses given percent relative errors, i.e. 100 x (approximate-true values)/true value.
mality holds, one gets
\[ R^{(s)} = \Phi[(\mu + ip\theta)/(1 - i(i - \eta)\theta^2)] - \Phi(\mu). \]  
(16)
As shown by Foulley (1987), this formula can also be derived using a fitness function connecting truncation and stabilizing selection (Chevalet 1988).

An illustration of these procedures is given in Table 2 for the example of progeny test selection on sires assuming progeny group size equal to 125 and heritability to be 0.20 on the liability scale. The approximation based on (16) turns out to be very accurate especially for medium and high selection pressures with, e.g., relative errors lower than 0.05 and 0.25 percent for \( Q = 0.1 \) and 0.5, respectively. Formulas by Cox and Wermuth (1991) without adjustment also provide a rather good approximation to the exact value (i.e., maximum relative error of 1 percent) if both selection rate and prevalence of the attribute are lower than one half. In contrast, calculations based on the usual formula \( ip \sigma_r \) lead in this case to errors up to 15 to 20 percent for extreme values of \( II \) and \( Q \).

Several paths of genetic change: So far, selection was supposed to be carried out within a monogenic population with discrete generations, or in a population with two sexes which can be pooled with respect to selection or with no selection in one sex. In practice this is not always the case, and one would have to adjust formula (4) for differences along those two paths of selection (\( m \), males; \( f \), females) as follows
\[ R^{(s)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \Phi \left( \frac{a + \theta}{\sigma_r} \right) p(a|\hat{a}_{cm},\hat{a}_{cf}) da \right] \Phi(\mu) \]  
(17a)
\[ Q = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\hat{a}_{cm},\hat{a}_{cf}) \Phi(\mu) \]  
(17b)

Before \( \hat{a}_{cm} \) and \( \hat{a}_{cf} \) are the EBVs of candidates \( c \) in sex \( k \) \((k = m,f)\) for males and females respectively and \( \hat{a}_{c} = E(a|\hat{a}_{c}) \) represent the contribution of \((c,k)\) to the EBV \( \hat{a} \) in the offspring (i.e., \( \hat{a} = \Sigma \hat{a}_{c} = \Sigma \theta \hat{a}_{c,x} \); \( \theta_1 = \frac{1}{2} \)). Assuming as previously that \((a,\hat{a}_{cm},\hat{a}_{cf})\) (or equivalently \((a,\hat{a}_{cm},\hat{a}_{cf})\) has a joint normal distribution such that \( \sigma_{a,c}^2 = \rho_{a,c}^2 \sigma_r^2 \), \( \sigma_{a,c}^2 = \rho_{a,c}^2 \sigma_r^2 \), \( Cov(a,\hat{a}_{c}) = \sigma_{a,c} \) \((Cov(a,\hat{a}_{c}) = f_s \sigma_{a,c}^2)\) and \( Corr(\hat{a}_{cm},\hat{a}_{cf}) \) being equal to \( \rho_{m,f} = Corr(\hat{a}_{cm},\hat{a}_{cf}) \), integration of (19a) becomes feasible leading to:
\[ R^{(s)} = \Phi_3(u,\eta_m,\eta_f;\rho_{m,h},\rho_{f,h},\rho_{m,f})/Q - \Phi(\mu) \]  
(18a)
and
\[ Q = \Phi_2(\eta_m,\eta_f;\rho_{m,f}) \]  
(18b)
where arguments are defined as in (8a,b) i.e., \( u = \Phi^{-1}(\Pi) \), \( \eta_k = -\theta \sigma_{a,c}/\sigma_{a,c} = -s_k/\sigma_{a,k} \) \((k = m,f)\) and \( \Phi_2(x_1,x_2,x_3,t_1,t_2,t_3) \) is the trivariate normal CDF with \( x_1,x_2,x_3 \) as arguments and \( r_{12},r_{13},r_{23} \) as corresponding row wise correlations.

Usually, it will be assumed that \( \rho_{m,f} \) the correlation among male and female EBVs in mates is zero, thus simplifying computation of (18a,b). Hence, \( Q = Q_m Q_f \) with \( Q_m, Q_f \) being selection rates in males and females respectively, and \( \eta_k = \Phi^{-1}(\eta_k) \). (\( k = m,f \)).

If different selection procedures are practiced among sexes to produce sires and dams, one has to consider four paths of selection from the grand parent generation, i.e., male to male \((k = 1)\), female to male \((k = 2)\), male to female \((k = 3)\) and female to female \((k = 4)\) as initially described by Rendel and Robertson (1950). Formulas to apply are then straight extensions of (20a,b) with 5 dimensions in the normal CDF instead of 3.

Formulas (18a,b) can be illustrated for the prediction of the expected genetic level of young sires bred in Maine Anjou for a high twin calving rate (Manfredi et al. 1991). For that purpose, the top three service sires were chosen on their EBVs on the underlying scale with an average accuracy of \( p_{m,n}^2 = 0.90 \), and a selection differential of \( \Delta_n = 1.657 \sigma_{a,n} \) corresponding to \( \eta_n = -1.163 \). In the other sex, 58 cows which had twin calvings at least twice each were selected as elite dams among 2536 contemporary recorded cows resulting in \( \eta_n = -2.00 \), \( \Delta_f = 2.373 \sigma_{a,f} \) and a minimum accuracy of \( p_{m,n}^2 = 0.085 \) for an individual evaluation with 3 records per cow (see formula in Appendix B), heritability and repeatability coefficients both equal to 0.13 on the liability scale (Manfredi et al. 1991) and a twin calving frequency of \( \Pi = 0.05 \) in the base population (second and third calves).

Young bulls born out of those sires and dams are expected to have daughters with a twin calving rate of \( q_{m,n}^2 = 0.05 \), \( q_{m,n}^2 = 0.25 \), \( p_{m,n}^2 = 0.25 \), \( \Phi(\eta_n) \) of \( \Phi(\eta_n) \), i.e., 7.41%. It can easily be checked that the classical approach will lead in that case to an expected value in the daughters of \( 0.05 + 0.25 (\Delta_n + \Delta_f) = 0.05 + 0.566 \sigma_{a,f} \), i.e., 7.10% for \( \sigma_{a,c} = \Phi^{-1}(0.05) = 0.7572 \). Had the selection on elite dams being more severe, with only the 19 cows having at least 3 twin calvings, the expected twinning rate in daughters of young bulls would have been 7.55% using (20a,b), or 7.21 percent with the ipo formula.

Other issues: This study is an attempt to address the general problem of predicting selection response for threshold dichotomous traits. Curnow (1984) derived formulas for genetic progress based on progeny testing in the case where none of the progeny exhibit the unfavorable attribute which represents a very particular form of selection. As mentioned previously, some work was already made in this field by Falconer (1989) and Mendell and Elston (1974) using univariate approximations. This study developed a more general framework and allows results to be related to other approximations; Moreover, formulae derived illustrate very well why the expected response to se-
lection is asymmetric and results clearly show that prediction based on the classical formula \( ip \sigma \) are reliable approximations only for intermediate values of the selection rate \((Q \approx \frac{1}{2})\) and the frequency of the trait \((\Pi \approx \frac{1}{2})\). For instance, in this study the error of prediction using \( ip \sigma \) may amount up to 15 to 20\% for extreme values of both the incidence level \((\Pi = 0.05; 0.95)\), and the selection rate \((Q = 0.1; 0.9)\). In such cases, asymmetry should be taken into account via, e.g., the threshold liability model. Such predictions might be a means to test the relevance of this model experimentally.

One critical assumption made here consists in the normality assumption for the joint distribution of \((a, \tilde{a})\), and especially for the marginal distribution of \(\tilde{a}\). This assumption may be tenable for Maximum a Posteriori (GIANOLA and FOULLEY 1988) and quasi-likelihood procedures (GILMOUR, ANDERSON and RAE 1985) if there is enough information provided by the data to each candidate for selection. This usually occurs in sire evaluation based on many progeny born out of artificial insemination: see, e.g., histograms of sire proofs for twinning rate in daughter groups shown in MANFREDI et al. (1991). One practical provision to apply in such a case would be at first to adjust the expression \((\Delta \tilde{a})\) of the selection differential in \(\tilde{a}\) by replacing \(ip \tilde{a}\) in (15) with its effective value calculated from the real distribution of EBVs. For instance, for phenotypic selection on single records, EBVs of candidates are \(\tilde{a}(1) = h^2 \phi(\mu)/\Pi\) and \(\tilde{a}(0) = -h^2 \phi(\mu)/(1 - \Pi)\) for \(y = 1\) and \(y = 0\) respectively (see formulas 11 and 12a with \(m = 1\) in APPENDIX B). For \(\Pi < Q\) such values lead to \(\Delta \tilde{a} = \tilde{a}(1)(\Pi/Q) + \tilde{a}(0)(Q - \Pi)/Q\) which, for \(Q = \frac{1}{2}\), reduces exactly to FALCONER’s (1989) formula \(\Delta \tilde{a} = h^2 S\) with \(S = 2\Pi i_\mu + (1 - 2\Pi) i_\nu\) (see p. 407) using his notations \(i_\mu = \phi(\mu)/\Pi\) and \(i_\nu = -i_\mu(1 - \Pi)/\Pi\). IM and GIANOLA (1988) also treated a similar case using a direct approach to parent offspring regression with either one or both parents measured. However, in general, the problem remains extremely intricate since one has to integrate out in (4) not with respect to \(\tilde{a}\), but to the discrete phenotypic variable used as selection criterion.

Other pending tasks are to extend this procedure to more than one generation taking into account the effects of both genetic drift and linkage disequilibrium on genetic variance: see, e.g., formulas by KEIGHTLEY and HILL (1987), CHEVALET (1988), WRAY and HILL (1989) and VERRIER, COLLEAU and FOUILLY (1990) for continuous traits. The problem of overlapping generations also deserves attention especially as for mid range or long term predictions which are of great interest in practical animal breeding programs for livestock production.

We would like to thank J. J. COLLEAU, V. DUCROCQ and M. POUTOUS for reading this manuscript as well as two anonymous reviewers for their helpful comments and criticisms. Special thanks are expressed to one reviewer who suggested presenting the second univariate approximation in the form given in (16).

LITERATURE CITED


ROBERTSON, A., 1950 Proof that the additive heritability on the p scale is given by the expression \(z^2h^2/pq\). Genetics 35: 234–236.


VERRIER, E., J. J. COLLEAU and J. L. FOULLEY, 1990 Predicting
APPENDIX A

Analytical integration of (5): One can solve the first integral within brackets in (5) say
\[
\int_{-\infty}^{+\infty} \Phi(\mu + a)/\sigma_\alpha p(a/\hat{a}) \, da
\]
(A.1)

using the following general result about normal probability integrals by Curnow (1984)
\[
\int_{-\infty}^{+\infty} \phi(u)\Phi(\alpha + \beta u) \, du = \Phi(\alpha + \beta^2)^{-1/2}
\]
where \( \phi(u) \) is the standardized normal density. Putting \( \alpha^* = (a - \hat{a})/(1 - \rho_i^2)^{1/2} / \sigma_\alpha \), (A.1) can be written as
\[
\int_{-\infty}^{+\infty} \phi(\alpha^*) \Phi(\mu_a + \hat{a} + \alpha^*(1 - \rho_i^2) / \sigma_\alpha \sigma_i) \, da^*
\]
which, after setting \( \alpha = (\mu_a + \hat{a}) / \sigma_a \), and \( \beta = (1 - \rho_i^2)^{1/2} / \sigma_a \sigma_i \), reduces to
\[
\Phi(\mu_a + \hat{a}) / [(1 - \rho_i^2) \sigma_a^2 + \sigma_i^2]^{1/2}
\]
From (A.3) and (6b), formula (5) can be rewritten as
\[
R^{(+)} = \left\{ \int_{\theta_k}^{\infty} \Phi\left( \frac{\mu_a + \hat{a}}{[(1 - \rho_i^2) \sigma_a^2 + \sigma_i^2]^{1/2}} \right) \cdot \hat{a}^* \sigma_i^2 \, d\hat{a}^* \right\} - \Pi
\]
(A.4)

Letting \( \hat{a}^* = -\hat{a}/\sigma_2 \) and noting that \( (1 - \rho_i^2) \sigma_a^2 + \sigma_i^2 \) is equal to \( (1 - \rho_i^2)^2 / \sigma_a^2 \sigma_i^2 \), the argument of \( \Phi(\cdot) \) in (A.4) can be expressed as \( (\mu - \hat{a}^* \sigma_i^2)/(1 - \rho_i^2)^{1/2} / \sigma_2 \) with \( \mu = \mu_a / \sigma_a \) and the integral to be computed in (A.4) is eventually
\[
\int_{-\infty}^{0} \Phi\left( \frac{\mu - \hat{a}^* \sigma_i^2}{(1 - \rho_i^2)^{1/2} \sigma_2} \right) \phi(\hat{a}^*) \, d\hat{a}^*
\]
It is well known that
\[
\int_{-\infty}^{0} \Phi\left( \frac{u - rt}{(1 - r^2)^{1/2}} \right) \phi(t) \, dt
\]
is an alternative form for the bivariate normal CDF, \( \Phi_2(u,v;r) \) with arguments \( u,v \) and correlation \( r \), so that formula (A.4) can be simplified into
\[
R^{(+)} = [\Phi_2(\mu, \eta; \sigma_2) - \Phi(\mu)\Phi(\eta)] / \Phi(\eta)
\]
(A.5)

where
\[
\mu = \Phi^{-1}(\Pi)
\]
(6a)

and
\[
\eta = -\theta s/\sigma_a = -s/\sigma_a = \Phi^{-1}(Q)
\]
(A.6b)

\(( \sigma_a \), being the standard deviation of EBVs in candidates).

APPENDIX B

Expression for the accuracy. Sire evaluation on progeny performance: It is usual in such a case to work on the sire transmitting ability \( s \) which represents half the additive genetic value \( a_i \) in our notations. The Fisher scoring algorithm to implement in Gilmour's procedure (Gilmour, Anderson and Rae 1985) can be written as, from round \( I \) to \( I + 1 \) (Foulley, Gianola and Im 1990)
\[
(q^{[I]} + \lambda) s^{[I+1]} = q^{[I]}(f_i - \Pi)/\phi(\mu)
\]
(B.1)

where \( \Pi = \Phi(\mu) \) is the true frequency of the trait as before; \( n_u f_i \) are the progeny group size and observed frequency in progeny of sire \( i \), respectively; \( \lambda = s_a^2 / s_i^2 \) is the ratio of phenotypic \( (s_a^2) \) to sire \( (s_i^2) \) variances; \( \gamma = \phi^2(\mu)/[I(1 - \Pi) - t\phi^2(\mu)] \), \( t = s_i^2 / s_a^2 \) being the intraclass correlation, and \( q_i = n_i \gamma \).

From the left hand side of (B.1), one gets the expression for the asymptotic variance of prediction errors, say \( V_i = \text{Var}(s_i - s) \)
\[
V_i = (q_i + \lambda) s_i^2
\]
(B.2)

The accuracy \( (\rho_s^2) \) in predicting \( s_i \) by \( \hat{s}_i \) or equivalently \( a_i \) by \( \hat{a}_i \), is related to \( V_i \) by \( \rho_s^2 = \varphi(q_i + \lambda) \), which can also be written in the classical form
\[
\rho_s^2 = n_i / (n_i + k)
\]
(B.3)

with
\[
k = \lambda / \gamma = [I(1 - \Pi) - t\phi^2(\mu)] / \phi(\mu)
\]
(B.4)

Here \( t = h^2/4 \) and \( k \) can be expressed as for continuous traits as
\[
k = (4/h_i^2) - 1
\]
(B.5)

using the formula of Dempster and Lerner (1950) and Robertson (1950) for the heritability in the observed scale \( h_i^2 = h^2\phi^2(\mu)/[I(1 - \Pi)] \).

Sire evaluation on repeated records: In that situation, the model can be formulated as
\[
x_{ij} = \mu_a + a_i + p_i + e_{ij}
\]
(6b)

where \( x_{ij} \) is the \( j \)th performance \( (j = 1, \ldots, m) \) of the \( i \)th dam \( (i = 1, \ldots, J) \); \( a_i \sim (0, \sigma_a^2) \) is the additive genetic value, \( p_i \sim (0, \sigma_p^2) \) is the permanent environmental effect, and \( e_{ij} \sim (0, \sigma_e^2) \) is the residual effect.

The system to be solved is then (ignoring iteration round for the sake of simplicity)
\[
\begin{cases}
(q_i + \lambda) \hat{a}_i + q^* \hat{p}_i = q^*(f_i - \Pi)/\phi(\mu) \\
q^* \hat{a}_i + (q_i + \lambda) \hat{p}_i = q^*(f_i - \Pi)/\phi(\mu)
\end{cases}
\]
(7)

for \( i = 1, \ldots, I \)

where \( f_i \) is the observed frequency of the attribute.
among the \( m_i \) repeated records of dam \( i \); \( \lambda_a = \sigma_a^2 / \sigma_x^2 \), \( \lambda_p = \sigma_p^2 / \sigma_x^2 \). \( \gamma \) is as before with \( t \) here being equal to the repeatability coefficient \( r = (\lambda_a + \lambda_p)^{-1} \) and \( q^* = m_r \gamma \).

After eliminating the \( p_i \) unknowns in (B.7), the system becomes:

\[
\frac{q^*(\lambda_a + \lambda_p) + \lambda_a \lambda_p}{q^* + \lambda_p} \hat{a}_i = \frac{q^* \lambda_p}{(q^* + \lambda_p) \phi(\mu)} (f_i - \Pi) 
\]

i.e.

\[
\hat{a}_i = \frac{q^* \lambda_p}{q^*(\lambda_a + \lambda_p) + \lambda_a \lambda_p \phi(\mu)} (f_i - \Pi). \quad (B.8)
\]

Then,

\[
V_i = \text{Var}(\hat{a}_i - a_i) = \frac{(q^* + \lambda_p) \sigma_x^2}{q^*(\lambda_a + \lambda_p) + \lambda_a \lambda_p} 
\]

and

\[
\rho_{p_i}^2 = 1 - (V_i / \sigma_x^2) 
\]

\[
= \frac{q^* \lambda_p}{q^*(\lambda_a + \lambda_p) + \lambda_a \lambda_p}.
\]

\[
\hat{a}_i = \rho_{p_i}^2 (f_i - \Pi) / \phi(\mu) \quad \text{(B.11)}
\]

Noting that \( \lambda_a^{-1} = h_a^2 \), and \( (\lambda_a + \lambda_p) / \lambda_a \lambda_p = r \), and after dividing the numerator and denominator in (B.10) by \( \lambda_a \lambda_p \), one gets

\[
\rho_{p_i}^2 = \frac{q^* h_a^2}{q^* r + 1} \frac{m_r \phi^2(\mu)}{\Pi(1 - \Pi) + (m_i - 1) r \phi^2(\mu)} \quad \text{(B.12a)}
\]

or, alternatively, under Lush's (1945) classical form

\[
\rho_{p_i}^2 = \frac{m_r h_a^2}{1 + (m_i - 1) r \phi^2(\mu)} \quad \text{(B.12b)}
\]

where

\[
r \phi^2(\mu) = \frac{r \phi^2(\mu)}{\Pi(1 - \Pi)}
\]

is defined in the same way as \( h_a^2 \), as a repeatability coefficient on the observed scale.