Original article

Probability statements about the transmitting ability of progeny-tested sires for an all-or-none trait with an application to twinning in cattle

J.L. Foulley[1] and S. Im[2]

[1] *Institut National de la Recherche Agronomique, station de génétique quantitative et appliquée, centre de recherches de Jouy, 78350 Jouy-en-Josas;*
[2] *Institut National de la Recherche Agronomique, laboratoire de biométrie, centre de recherches de Toulouse BP 27, 31326 Castanet-Tolosan Cedex, France*

Summary – This paper compares three statistical procedures for making probability statements about the true transmitting ability of progeny-tested sires for an all-or-none polygenic trait. Method I is based on the beta binomial model whereas methods II and III result from Bayesian approaches to the threshold-liability model of Sewall Wright. An application to lower bounds of the transmitting ability of superior sires with a high twinning rate in their daughter progeny is presented. Results of different methods are in good agreement. The flexibility of these different methods with respect to more complex structures of data is discussed.

genetic evaluation – all-or-none traits – beta binomial model – threshold model – Bayesian methods

Résumé – **Enoncés probabilistes relatifs à la valeur génétique transmise de pères testés sur descendance pour un caractère tout-ou-rien avec une application à la gémellité chez les bovins.** *Cet article compare 3 procédures statistiques en vue de la formulation d'énoncés probabilistes relatifs à la valeur génétique transmise de pères testés sur descendance pour un caractère polygénique tout-ou-rien. La méthode I repose sur le modèle bêta binômial alors que les méthodes II et III découlent d'approches bayésiennes du modèle à seuils de S. Wright. Une application concernant la borne inférieure de la valeur génétique transmise de pères d'élite présentant un taux de gémellité élevé chez leurs filles est présentée. Une bonne concordance des résultats entre méthodes est observée. La flexibilité de ces différentes méthodes vis-à-vis de structures de données plus complexes est abordée en discussion.*

évaluation génétique – caractères tout-ou-rien – modèle bêta binômial – modèle à seuils – méthodes bayésiennes

INTRODUCTION

Genetic evaluation for all-or-none traits is usually carried out via Henderson's mixed model procedures (Henderson, 1973) having optimum properties for the Gaussian linear mixed model. Even though a linear approach taking into account some specific features of binomial or multinomial sampling procedures can be worked out in multi-population analysis (Schaeffer & Wilton, 1976; Berger & Freeman, 1976; Beitler & Landis, 1985; Im *et al.*, 1987), these methods suffer from severe statistical drawbacks (Gianola, 1982; Meijering & Gianola, 1985; Foulley, 1987). Especially as distribution properties of predictors and of prediction errors are unknown for regular or improved Blup procedures applied to all-or-none traits, it would therefore be dangerous to base probability statements on the property of a normal spread of genetic evaluations or of true breeding values given the estimated breeding value.

The aim of this paper is to investigate alternative statistical methods for that purpose. Emphasis will be placed on making probability statements about true transmitting ability (TA) of superior sires progeny-tested for some binary characteristic having a multifactorial mode of inheritance. Numerical applications will be devoted to sires with a high twinning rate in their daughter progeny.

METHODS

The methods presented here are derived from statistical sire evaluation procedures which are based on specific features of the distribution involved in the sampling processes of such binary data. Three methods (referred to as I, II and III) will be described in relation to recent works in this area. The first method is based on the beta binomial model (Im, 1982) and the two other ones on Bayesian approaches (Foulley *et al.*, 1988) to the threshold-liability model due to Wright (1934 a and b). All three methods assume a conditional binomial distribution B (n, π) of binary outcomes (*i.e.* n progeny performance of a sire) given the true value π of a probability parameter (here the sire's true breeding value or transmitting ability). The three methods differ in regard to the modelling of π itself, either directly (beta binomial) or indirectly (threshold-liability model), and consequently in describing the prior distributions of parameters involved, *v.i.z.* π itself or location parameters on an underlying scale.

Method I

Let $y_{ij} = 0$ or 1 be the performance of the jth progeny ($j = 1, 2, ..., n_i$) out of the ith sire ($i = 1, 2, ..., q$) and n_i unrelated dams. Let π_i designate the true transmitting ability (π_i) of sire i.

A priori, the π_i's are assumed to be independently and identically distributed (i.i.d.) as beta random variables with parameters (α, β).

$$f(\pi_1, \ldots, \pi_i, \ldots, \pi_q | \alpha, \beta) \propto \prod_i \pi_i^{\alpha-1} (1 - \pi_i)^{\beta-1} \qquad (1)$$

Conditional on true value π_i, the distribution of binary responses among progeny of a given sire is taken as binomial B (n_i, π_i). These distributions are conditionally independent among sires so that the likelihood is the product binomial

$$P_r(Y_{1o} = y_{1o}, \cdots, Y_{io} = y_{io}, \cdots, Y_{qo} = y_{qo} \mid \pi_1, \cdots, \pi_i, \cdots, \pi_q)$$

$$\propto \prod_i \pi_i^{y_{io}}(1 - \pi_i)^{n_i - y_{io}} \tag{2}$$

where the circle stands for a summation over the corresponding subscript (here $y_{io} = \Sigma_j \, y_{ij}$) and capital letters indicate random variables.

As the prior and the likelihood are conjugate (Cox & Hinkley, 1974, p. 308), the posterior distribution remains in the beta family and can be written as:

$$f(\pi_1, \cdots, \pi_i, \cdots, \pi_q \mid \alpha, \beta; n_i, y_{io}; i = 1, 2, \cdots, q) \propto$$

$$\prod_i \pi_i^{y_{io}+\alpha-1}(1 - \pi_i)^{n_i - y_{io}+\beta-1} \tag{3}$$

with normalizing constant

$$\left[\prod_i B(y_{io} + \alpha; \, n_i - y_{io} + \beta)\right]^{-1}$$

The density in (3) is a product of q independent beta densities. Ignoring subscripts, the posterior density for a given sire is:

$$f(\pi \mid \alpha, \beta, n, y_o) = \pi^{y_o+\alpha-1}(1 - \pi)^{n-y_o+\beta-1}/B\,(y_o + \alpha; \, n - y_o + \beta) \tag{4a}$$

This Beta distribution B (a, b) can be conveniently expressed with a reparameterization in terms of a prior mean $\pi_o = \alpha/(\alpha + \beta)$, an intra class correlation $\rho_b = (\alpha + \beta + 1)^{-1}$ and the observed frequency $p = y_o/n$. The conditional distribution of π given n, p, π_o and ρ_b is then

$$\pi \mid n, p, \pi_o, \rho_b \sim \text{Beta}\,[np + \pi_o(\rho_b^{-1} - 1); \, n(1 - p) + (1 - \pi_o)\,(\rho_b^{-1} - 1)] \tag{4b}$$

with expectation

$$E(\pi \mid n, p, \pi_o, \rho) = [np + (\rho_b^{-1} - 1)\pi_o]/[n + (\rho_b^{-1} - 1)] \tag{4c}$$

which will be noted $\hat{\pi}$ so as to reflect both its interpretation as a Bayesian estimator of π as well as its equivalence with the best linear predictor or selection index (Henderson, 1973). Using $\hat{\pi}$ and letting $\lambda_b = \rho_b^{-1} - 1$ with λ_b interpretable as a ratio of within to between sire components of variances, the distribution in (4b) can be viewed as a function of n, $\hat{\pi}$ and λ_b, that is, conditional on n, λ_b and $\hat{\pi}$, the distribution of π is:

$$\pi \mid n, \hat{\pi}, \lambda_b \sim \text{Beta}\,[(n + \lambda_b)\,\hat{\pi}, \, (n + \lambda_b)(1 - \hat{\pi})] \tag{5}$$

with expectation

$$E(\pi \mid n, \hat{\pi}, \lambda_b) = \hat{\pi} \tag{6a}$$

and variance

$$\text{Var}\,(\pi \mid n, \hat{\pi}, \lambda_b) = \hat{\pi}(1 - \hat{\pi})/(n + \lambda_b + 1) \tag{6b}$$

Probability statements about true values of TA given the data $(n$ and p or $\hat{\pi})$ and values of the hyperparameters $(\pi_o$ and ρ_b or $\lambda_b)$ can be easily made using

expressions (4b) or (5) of the posterior density of π. Notice that formula (6a) also represents the probability of response $\Pr(Y_{ik} = 1 \mid n_i, \ p_i)$ for a future progeny (k) out of sire (i) with an observed frequency of response p_i in n_i offspring.

These probability statements can be made for specific sires given their progeny test data (n, p or $\hat{\pi}$) and the characteristics of the corresponding population, such as the mean incidence π_o and the intraclass coefficient ρ_b as a parameter of genetic diversity. To allow for comparisons among methods, this ρ_b, or equivalently the ratio λ_b, will be expressed according to Im's (1987) results which relate intraclass coefficients on the binary (ρ_b) and underlying (ρ) scales in a population in which the incidence of the trait is π_o (see next paragraph).

In the case of twinning, interest is usually in superior sires having estimated transmitting ability (ETA) values above the mean π_o. Attention will then be devoted to the lower TA bound π_m which is exceeded with a probability α i.e, to π_m, such that:

$$\Pr\left\{[\pi \mid n, \ p \ (\text{or } \hat{\pi}), \ \rho_b \ (\text{or } \lambda_b)] \geq \pi_m\right] = \alpha \tag{7}$$

This involves computing $x \in [0, 1]$ values of the so-called incomplete beta function defined as, in classical notations

$$\mathrm{I}_x(a, \ b) = \left[\int_0^x t^{a-1}(1 - t)^{b-1}\mathrm{d}t\right] / \mathrm{B}(a, \ b) \tag{8}$$

Details about numerical procedures used to that respect are given in appendix A.

In addition, more general results can be produced for instance in terms of (n, $\hat{\pi}$) values such that formula (7) holds for given values of π_m (TA lower bound) and α (probability level): see appendix A.

Method II

This method is derived from genetic evaluation procedures for discrete traits introduced recently by several authors. All these procedures postulate the Wright threshold liability concept. We restrict our attention here to Bayesian inference approaches proposed independently by Gianola & Foulley (1983), Harville & Mee (1984), Stiratelli *et al.* (1984) and Zellner & Rossi (1984).

Although the methodology is very general vis-à-vis data structures, for the sake of simplicity only its unipopulation version (μ model) will be considered in this paper.

Let l_{ij} be a conceptual underlying variable associated with the binary response y_{ij} of the jth progeny of the ith sire. The variable l_{ij} is modelled as:

$$l_{ij} = \eta_i + e_{ij} \tag{9}$$

where η_i is the location parameter associated with the population of progeny out of sire i and the e_{ij}'s are NID $(0, \sigma_e^2)$ within sire deviations.

Conditional on η_i, the probability that a progeny responds in one of the two exclusive categories coded [0] and [1] respectively is written as:

$$\pi_{i[0]} = \Phi[(\tau - \eta_i)/\sigma_e] \text{ and } \pi_{i[1]} = 1 - \pi_{i[0]}$$

where τ is the value of the threshold, σ_e the within sire standard deviation and $\Phi(\cdot)$ the normal CDF evaluated at $(\tau - \eta_i)/\sigma_e$.

It is convenient to put the origin at the threshold and set σ_e to unity, *i.e.* "standardize" the threshold model (Harville & Mee, 1984)

$$\mu_i = (\eta_i - \tau)/\sigma_e \tag{10}$$

the expression for $\pi_{i[0]}$ can be written as

$$\pi_{i[0]} = \Phi(-\mu_i) = 1 - \Phi(\mu_i)$$

and that for $\pi_{i[1]}$ as:

$$\pi_i = \Phi(\mu_i) \tag{11}$$

In what follows, and to simplify notation, $\pi_{i[1]}$ will be referred to as π_i .

Letting $\mu = \{\mu_i\}$ be an $(q \times 1)$ vector, a natural choice for the prior distribution of $\boldsymbol{\mu}$ under polygenic inheritance, is:

$$\boldsymbol{\mu}|\sigma_u^2 \sim N(\mu_o 1, A\sigma_u^2) \tag{12}$$

where $\mathbf{A}$ is equal to twice Malecot's genetic relationship matrix for the q sires ($\mathbf{A} = \mathbf{I}$ in method I), σ_u^2 is the sire component of variance and μ_o the general phenotypic mean in the underlying scale.

These parameters μ_o and σ_u^2 are linked to the overall incidence π_o via:

$$\pi_o = E_{\mu_i}[\Phi(\mu_i)]$$
$$= \int_{-\infty}^{+\infty} \Phi(\mu)(2\pi\sigma_u^2)^{-1/2} \exp[-(\mu - \mu_o)^2/2\sigma_u^2]d\mu$$
$$= \Phi[\mu_o/(1 + \sigma_u^2)^{1/2}]$$

or, equivalently, defining $\sigma^2 = \sigma_u^2 + \sigma_e^2$ with $\sigma_e^2 = 1$,

$$\mu_o = \Phi^{-1}(\pi_o)\sigma \tag{13a}$$

Similarly, the sire variance σ_{ub}^2 in the binary scale can be related to the underlying distribution *via*

$$\sigma_{ub}^2 = \text{var}_{\mu i}[\Phi(\mu_i)]$$
$$= \Phi_2(\tilde{\mu}, \tilde{\mu}; \rho) - \pi_o^2 \tag{13b}$$

where $\Phi_2(x, y; r)$ is the standardized bivariate cumulative density function with mean 0 and correlation $r, \tilde{\mu} = \mu_o/(1 + \sigma_u^2)^{1/2}$ and ρ in (13b) is the intraclass correlation coefficient $\rho = \sigma_u^2/\sigma^2$. The variance in (13b) can be obtained directly by a probability argument or as the limit of a formula given by Foulley *et al.* (1988) for the variance of the observed frequency p_i when the progeny group size n tends to infinity. Notice also that it differs from the classical expression $\phi^2(\tilde{\mu})\sigma_u^2$ proposed by Dempster & Lerner (1950), $\phi(\cdot)$ designating the standardized normal density function, which is a first order Taylor expansion of (13b) about $\rho = 0$.

The likelihood function has the same form (*v.i.z.* product binomial) as in method I (formula 2) so that the posterior density reduces to:

$$f(\boldsymbol{\mu}|\mathbf{y},\mu_o,\sigma_u^2) \propto \prod_{i=1}^{m} [\Phi(z'_i\boldsymbol{\mu})]^{y_{io}} [1 - \Phi(z'_i\boldsymbol{\mu})]^{n_i-y_{io}}$$

$$\exp\left[-\boldsymbol{\mu}-\mu_o\mathbf{1})'(\boldsymbol{\mu}-\mu_o\mathbf{1})/2\sigma_u^2\right] \tag{14}$$

where z'_i is a $(1 \times m)$ row vector having 1 in the ith column and 0 elsewhere.

The logposterior density L $(\boldsymbol{\mu};\mathbf{y},\mu_o,\sigma_u^2)$ can be minimized with respect to $\boldsymbol{\mu}$ by a scoring algorithm of the general form

$$-\left\{E_{y|\boldsymbol{\mu}}\left[\partial^2 L(\boldsymbol{\mu};\mathbf{y},\mu_o,\sigma_u^2)/\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}'\right]\right\}_{\boldsymbol{\mu}=\boldsymbol{\mu}^{[t-1]}} \cdot \Delta\boldsymbol{\mu}^{[t]}$$

$$= \partial L[\boldsymbol{\mu};\mathbf{y},\mu_0,\sigma_u^2)/\partial\boldsymbol{\mu}]_{\boldsymbol{\mu}=\boldsymbol{\mu}^{[t-1]}} \tag{15a}$$

The value of $\boldsymbol{\mu}$ in the t-th iteration can be computed by solving the non-linear system

$$(\mathbf{Z}'\,\mathbf{W}^{[t-1]}\mathbf{Z} + I/\sigma_u^2)\Delta\boldsymbol{\mu}^{[t]} = \mathbf{Z}'\,\mathbf{v}^{[t-1]} - (\boldsymbol{\mu}^{[t-1]} - \mu_o\mathbf{1})/\sigma_u^2 \tag{15b}$$

where $\mathbf{W}$ and $\mathbf{v}$ are an $(n \times n)$ diagonal matrix and an $(n \times 1)$ vector respectively having elements

$$v_i = n_i[p_i - \Phi(\mu_i)] + \{\Phi(\mu_i)[1 - \Phi(\mu_i)]\} \tag{16a}$$

$$w_i = n_i\phi^2(\mu_i)/\{\Phi(\mu_i)[1 - \Phi(\mu_i)]\} \tag{16b}$$

Define $\lambda = \sigma_e^2/\sigma_u^2 = 1/\sigma_u^2$ and $\mathbf{u} = \boldsymbol{\mu} - \mu_o\mathbf{1}$ as an $(m \times 1)$ vector of sire deviations. Then the system to be solved $\Delta\boldsymbol{\mu} = \Delta\mathbf{u}$ becomes

$$(w_i^{[t-1]} + \lambda)u_i^{[t]} =$$

$$w_i^{[t-1]}\,u_i^{[t-1]} + n_i\,\phi(\mu_o + u_i^{[t-1]})\,\frac{p_i - \Phi(\mu_o + u_i^{[t]})}{\Phi(\mu_o + u_i^{[t-1]})[1 - \Phi(\mu_o + u_i^{[t-1]})]} \tag{17}$$

An interesting feature of the posterior distribution in (14) is its asymptotic normality

$$n_o^{1/2}\left[(\boldsymbol{\mu} \mid \mathbf{y},\,\mu_o,\,\sigma_u^2) - \boldsymbol{\mu}^*\right] \to N\left\{0, \lim\left[I(\boldsymbol{\mu})/n_o\right]^{-1}\right\} \tag{18a}$$

where $\boldsymbol{\mu}^*$ is the mode of the posterior density of $\boldsymbol{\mu}$ in (14) and $I(\boldsymbol{\mu})$ is defined as

$$I(\boldsymbol{\mu}) = -E_{\mathbf{y}|\boldsymbol{\mu}}[\partial^2 L(\boldsymbol{\mu};\mathbf{y},\mu_o,\,\sigma_u^2)/\partial\boldsymbol{\mu}\,\partial\boldsymbol{\mu}'] \tag{18b}$$

$\lim[I(\boldsymbol{\mu})\,/\,n_o]^{-1}$ can be replaced for test statistics by a consistent estimator such as $-n_o[I(\boldsymbol{\mu}^*)]^{-1}$ where $I(\boldsymbol{\mu})$ is evaluated at $\boldsymbol{\mu} = \boldsymbol{\mu}^*$.

The variance of the limiting normal distribution is usually taken to be (Berger, 1985, p. 224)

$$-n_o[\partial^2 L(\cdot)/\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}' \mid \boldsymbol{\mu} = \boldsymbol{\mu}^*]^{-1}$$

The form shown in (18a and b) applies as well to the asymptotic variance since both of them tend to the same limit as $n_o = \Sigma n_i$ tends to infinity. This involves the use of the following large sample distribution:

$$\mu_i \mid n_i,\, p_i,\, \mu_o,\, \lambda \sim N[\mu_i^*,\, \gamma_i = (w_i^* + \lambda)^{-1}] \tag{19}$$

where w_i^* stands for w_i evaluated at the mode μ_i^*.

Letting $\mu_m = \Phi^{-1}(\pi_m)$ be the parameter value in the underlying scale corresponding to π_m in (7), the probability that the true sire TA, μ_i of sire i (ETA = μ_i^*) exceeds μ (or equivalently $\pi > \pi_m$) can be expressed as

$$\Pr\left(\mu_i > \mu_m \mid n_i,\ p_i,\ \mu_o, \lambda\right) = \Phi[(\mu_i^* - \mu_m)/\gamma_i^{1/2}] \tag{20}$$

For given values of n, p (or $\hat{\pi}$) and $\lambda_b(\pi_o, \sigma_u^2)$, this probability can be computed, given π_m, and compared to the corresponding probability level obtained with the beta binomial model. Alternatively, one can determine the lower bound π_m such that $\Pr(\pi_i > \pi_m) = \alpha$ fixed, by taking $\mu_m = \mu_i^* - \gamma_i^{1/2}\Phi^{-1}(\alpha)$.

Notice that computing the probability in (20), based on the posterior distribution of the true TA, $f(\mu \mid n,\ p,\ \mu_o,\ \lambda)$ is equivalent to computing $\Pr(\pi > \pi_m)$ over the distribution of the probability of response $\pi = \Phi(\mu)$ for a future progeny of sires having an ETA equal to μ^* and a true TA distributed according to (19). This distribution in the observed scale would probably be more appealing for practitioners. This is especially clear as far as ETA's are concerned and one may alternatively to μ^*, consider as a sire evaluation, the expectation of $\pi = \Phi(\mu)$ with respect to the density of μ in (19), say $\hat{\pi}_2$. This expectation is:

$$\hat{\pi}_2 = \Phi[\mu^*/(1 + \gamma_i)^{1/2}] \tag{21}$$

However, the whole distribution of $\pi = \Phi(\mu)$ remains less tractable numerically than that of μ in (19) due to its following form:

$$f(\pi \mid n,\ p,\ \mu_o,\ \lambda) = (2\pi\gamma)^{-1/2} \exp\left\{-[\Phi^{-1}(\pi) - \mu^*(n,\ p,\ \mu_o,\ \lambda)]^2/2\gamma\right\}/\phi[\Phi^{-1}(\pi)] \tag{22}$$

Method III

This method is also derived from the threshold liability model but employs asymptotic properties at an earlier stage.

Let us consider, as previously, the observed frequency of response p_i in n_i progeny of sire i. Conditionally to the true TA, (π_i), p_i has an asymptotic normal distribution, *i.e.*:

$$n_i^{1/2}(p_i - \pi_i) \rightarrow \mathrm{N}\left[0, \pi_i(1 - \pi_i)\right] \tag{23}$$

The normit transformed of p_i, $m_i = \Phi^{-1}(p_i)$ has a conditional distribution which is also asymptotically normal. Following a classical theorem in asymptotic theory (see for instance, formula 6a.23, page 386 in Rao, 1973) and knowing that:

$$\Phi^{-1}(\pi_i) = \mu_i \quad \text{and} \quad \frac{d}{dp}[\Phi^{-1}(p)] = \{\phi[\Phi^{-1}(p)]\}^{-1}$$

one has, given μ_i

$$n_i^{1/2}(m_i - \mu_i)/[\phi(m_i)]^{-1}[p_i(1 - p_i)]^{1/2} \rightarrow \mathrm{N}\,(0, 1) \tag{24}$$

Assuming as in section II B that, *a priori*, the μ_i's are i.i.d $\sim\mathrm{N}\,(\mu_o^2,\ \sigma_u^2)$ leads to a posterior for μ_i with is also normally distributed.

$$\mu_i \mid n_i, \ p_i, \ \mu_o, \ \sigma_u^2 \sim N(\widehat{\mu}_i, \ c_i) \tag{25}$$

The expectation $(\widehat{\mu}_i)$ and variance (c_i) of the distribution in (25) can be easily expressed analytically as (see for instance Cox & Hinkley, 1974, formula 22, page 373).

$$\widehat{\mu}_i/c_i = [n_i\phi^2(m_i)m_i/p_i(1-p_i)] + (\mu_o/\sigma_u^2) \tag{26a}$$

with

$$c_i = \{[p_i(1-p_i)/n_i\phi^2(m_i)]^{-1} + \sigma_u^{-2}\}^{-1}$$
$$= [p_i(1-p_i)\sigma_u^2]/[n_i\phi^2(m_i)\sigma_u^2 + p_i(1-p_i)] \tag{26b}$$

Alternative formulae can be derived so as to mimic usual selection index expressions, which are, ignoring subscripts

$$\widehat{\mu} = \mu_o + CD[\Phi^{-1}(p) - \mu_o] \tag{27a}$$
$$c = (1 - CD)\sigma_u^2 \tag{27b}$$

where CD is given by the usual formula for the coefficient of determination, *i.e.* CD $= n/n + k)$ where the scalar k is defined as:

$$k = p(1-p)/\{\phi^2[\Phi^{-1}(p)]\sigma_u^2 \tag{28a}$$

In method II and with a definition of CD restricted to var $(\mu|y,\sigma_u^2) = (1-CD)\sigma_u^2$, this coefficient is:

$$k = \Phi(\mu_o + u)[1 - \Phi(\mu_o + u)]/\phi^2(\mu_o + u)\sigma_u^2 \tag{28b}$$

Notice that k can be interpreted as in selection index theory as the ratio of a within sire to a sire variance, since $p(1-p)/\phi^2(\cdot)$ is the asymptotic variance of the normit transformation of the frequency of response in progenies of a given sire, conditionally on the sire's true underlying TA.

Substituting $\widehat{\mu}_i$ and c_i for (26a and b) or (27 a and b) for μ_i^* and γ_i respectively in (20) enables us either to determine π_m for a given α probability level knowing μ_o, σ_u^2 (or equivalently π_o and λ), p and n, or to compute the probability level α such that μ_i exceeds a given threshold μ_m. Again, the distribution of $\pi = \Phi(\mu)$ in the observed scale corresponding to the density of μ in (25) can be obtained using the formula (22). Its expectation has the same expression as in (21) with $\widehat{\mu}_i$ and c_i replacing μ_i^* and γ_i respectively.

NUMERICAL APPLICATION

Procedures described in this paper are applied to the problem of screening superior sires with a high twinning rate in their female progeny.

There are relatively large differences among cattle populations with respect to the prevalence π_o of twinning (see for instance Maijala & Syvajarvi, 1977). Two values of π_o, say 2.5 and 4% are considered in this application.

These values of twinning rate are those observed in 2nd and mature (3rd to 7th) calvings respectively in such breeds as Charolais and Maine-Anjou (Frebling *et al.*, 1987). The first value of π_o (0.025) can be viewed as a progeny test of young bulls based on second calvings. Twinning in heifers was not considered because

of its extremely low rate (0.007). The second value (0.04) is an illustration of an evaluation system of service sires based on mature calvings.

Genetic variation for occurrence of multiple births in cattle was assumed to have an heritability coefficient in the underlying scale equal to 0.25, according to estimates published by Syrstad (1974). In this study, h^2 values of the underlying continuous variable are higher than those in the observed scale and, especially more stable over parities as theoretically expected for such binary traits. Using this value in (13a and b) leads to μ_o equal to -2.0242 and -1.8081, and h^2 in the binary scale equal to 0.0394 and 0.053 for π_o equal to 0.025 and 0.04, respectively. Notice, as already pointed out by Im (1987) that h^2 values reported here are slightly higher than those which would be obtained (0.0350 and 0.0483) from the classical formula of Dempster & Lerner (1950).

First application

The first application deals with 5 specific sires known for their high twinning rate in daughters. Table I shows lower bounds (π_m) of the TA in twinning of these 5 sires knowing their progeny test performance (n, p) in mature calvings at 2 different probability levels, $\alpha = 0.90$ and 0.95. In both cases, results are in good agreement across methods with, as expected due to asymptotic approximations, higher values of π_m being obtained with method II and, to a larger extent with III. Differences between I and II, however, are of little importance given the large values of n. Differences among methods are also reflected in ETA values on the observed scale with values for methods II and III very slightly less regressed towards the mean incidence $\pi_o = 0.04$ than in method I. On the other hand, this is a good example of a change of ranking according to criteria used. B and C have close ETA and π_m values although they differ largely in frequency (p). E is lower than D in p, close to D in ETA but larger in π_m due to a greater number of progeny.

Second application

For the two π_o frequencies, tables II and III show the values of progeny group size (n) which provides an $\alpha = 0.9$ probability level for different combinations of π_m, the TA taken as a minimum and $\hat{\pi}$, the ETA value according to formulae (5b) and (6). Minimum TA and ETA values were expressed in percent as well as, for practical purposes, deviations from π_o in σu_b units ($\sigma u_b = 1$ standard deviation in true TA in the $0 - 1$ scale). Results are shown in tables II and III for π_m varying from $+0.25$ to $+2.75$ σu_b and ETA from $+1.00$ to 3.00 σu_b with an elementary increment of 0.25.

The higher the ETA's, the lower are progeny numbers for a given value of π_m. For instance, for $\pi_o = 0.025$ and $\pi_m = 0.048$ (or equivalently $+1.50\sigma u_b$) progeny numbers providing a 0.9 probability level are 5199, 1291, 546, 279, 152 and 81 when the ETA goes from $+ 1.75$ to 3.00 σu_b. Corresponding figures are 3631, 895, 375, 188, 100 and 51 respectively when $\pi_o = 0.04$, i.e. about 60-70% of previous quantities. This special case is illustrated for $\pi_o = 0.025$ in figure 1 with a graph of the beta prior density and posterior distributions corresponding to $\pi_m = 0.048$ and n varying from 81 to 1291.

J.L. Foulley and S. Im
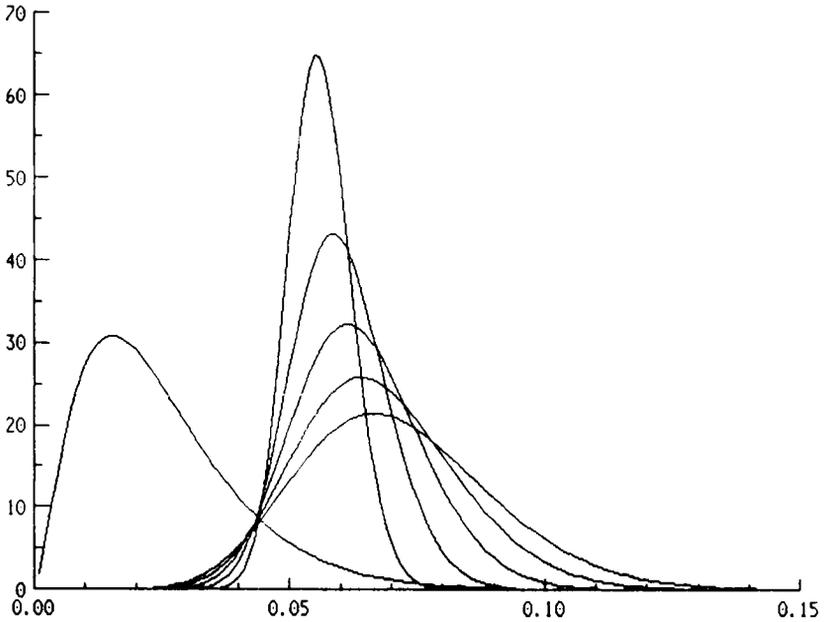
Fig. 1. An illustration of the beta binomial model with a beta prior density ($\pi_o = 0.025$ and $h^2 = 0.25$) and a family of posterior densities corresponding to $\pi_m = 0.048$, $\alpha = 0.90$ and different (n, ETA) combinations, namely, from the bottom to the top, (81, 0.0715); (152, 0.0676); (279, 0.0683); (546, 0.0599) and (1291, 0.0560).

Table I. Lower bounds (π_m) at probability levels of 0.9 and 0.95 for true transmitting abilities in twinning of 5 sires knowing their progeny test performance (n, p) under methods of inference (I: beta binomial; II, III: threshold model).

			(a) ETA according to methods			*(a)* Lower bound (π_m)					
						$\alpha = 0.90$			$\alpha = 0.95$		
Sire	n	p	I	II	III	I	II	III	I	II	III
A	225	0.1156	0.0968	0.0992	0.0996	0.0756	0.0768	0.0779	0.0703	0.0714	0.0727
B	201	0.1144	0.0943	0.0968	0.0972	0.0725	0.0736	0.0748	0.0672	0.0682	0.0695
C	860	0.0977	0.0931	0.0937	0.0938	0.0811	0.0816	0.0817	0.0780	0.0784	0.0786
D	128	0.0938	0.0740	0.0753	0.0758	0.0515	0.0515	0.0531	0.0464	0.0464	0.0481
E	405	0.0815	0.0750	0.0756	0.0757	0.0601	0.0603	0.0607	0.0563	0.0566	0.0570

a) assuming a mean incidence π_o of 0.04 and a coefficient of heritability equal to 0.25 and 0.0531 in the underlying and binary scales respectively.

Table II. Progeny group size and ETA values such as $\Pr(TA > \pi_m) = 0.90$ under the beta binomial model (Method I) and corresponding probability values under the threshold model with methods II (.) and III [.] for $\pi_o = 0.025$ and $h^2 = 0.25$.

$ETA^{(a)}$	Lower bound (π_m) in $TA^{(a)}$										
	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75
	2.89%	3.28%	3.66%	4.05%	4.44%	4.83%	5.21%	5.60%	5.99%	6.38%	6.76%
1.00	330	904	4041								
	(0.899)	(0.900)	(0.901)								
4.05%	[0.909]	[0.905]	[0.902]								
1.25	157	374	1002	4430							
	(0.897)	(0.901)	(0.902)	(0.902)							
4.44%	[0.917]	[0.911]	[0.906]	[0.903]							
1.50	73	182	418	1099	4816						
	(0.894)	(0.900)	(0.903)	(0.904)	(0.903)						
4.83%	[0.930]	[0.920]	[0.912]	[0.907]	[0.903]						
1.75	26	89	206	461	1195	5199					
	(0.890)	(0.899)	(0.903)	(0.905)	(0.905)	(0.903)					
5.21%	[0.945]	[0.932]	[0.921]	[0.914]	[0.909]	[0.904]					
2.00		38	105	231	504	1291	5578				
		(0.897)	(0.903)	(0.907)	(0.908)	(0.907)	(0.904)				
5.60%		[0.947]	[0.934]	[0.924]	[0.916]	[0.910]	[0.905]				
2.25			49	121	255	546	1385	5954			
			(0.903)	(0.908)	(0.910)	(0.910)	(0.908)	(0.905)			
5.99%			[0.949]	[0.936]	[0.926]	[0.918]	[0.911]	[0.906]			
2.50				60	136	279	588	1479	6326		
				(0.910)	(0.913)	(0.914)	(0.913)	(0.910)	(0.906)		
6.38%				[0.951]	[0.938]	[0.928]	[0.920]	[0.913]	[0.906]		
2.75					70	152	302	630	1572	6696	
					(0.915)	(0.917)	(0.917)	(0.915)	(0.911)	(0.906)	
6.76%					[0.952]	[0.941]	[0.930]	[0.922]	[0.914]	[0.907]	
3.00						81	167	325	671	1663	7062
						(0.921)	(0.921)	(0.920)	(0.917)	(0.913)	(0.907)
7.15%						[0.954]	[0.943]	[0.932]	[0.924]	[0.915]	[0.908]

a) expressed in σu_b (standard deviation of TA's) unit and in percentage.

Table III. Progeny group size and ETA values such as $\Pr(TA > \pi_m) = 0.90$ under the beta binomial model (Method I) and corresponding probability values under the threshold model with methods II (.) and III [.] for $\pi_o = 0.04$ and $h^2 = 0.25$.

$ETA^{(a)}$	Lower bound (π_m) in $TA^{(a)}$										
	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75
	4.56%	5.13%	5.69%	6.26%	6.82%	7.39%	7.95%	8.52%	9.08%	9.64%	10.21%
1.00	234	644	2878								
	(0.899)	(0.900)	(0.901)								
6.26%	[0.908]	[0.904]	[0.902]								
1.25	109	263	708	3132							
	(0.897)	(0.900)	(0.902)	(0.901)							
6.82%	[0.915]	[0.910]	[0.906]	[0.902]							
1.50	49	126	292	771	3383						
	(0.894)	(0.900)	(0.903)	(0.903)	(0.902)						
7.39%	[0.926]	[0.918]	[0.908]	[0.907]	[0.903]						
1.75	16	60	142	320	833	3631					
	(0.891)	(0.899)	(0.903)	(0.905)	(0.905)	(0.903)					
7.95%	[0.937]	[0.929]	[0.920]	[0.913]	[0.908]	[0.904]					
2.00		23	70	157	347	895	3875				
		(0.897)	(0.903)	(0.906)	(0.907)	(0.906)	(0.904)				
8.52%		[0.941]	[0.931]	[0.922]	[0.915]	[0.910]	[0.905]				
2.25			30	80	173	375	955	4115			
			(0.902)	(0.908)	(0.910)	(0.910)	(0.908)	(0.905)			
9.08%			[0.944]	[0.934]	[0.925]	[0.917]	[0.911]	[0.905]			
2.50				37	90	188	401	1015	4353		
				(0.908)	(0.912)	(0.913)	(0.912)	(0.909)	(0.905)		
9.64%				[0.947]	[0.936]	[0.927]	[0.919]	[0.912]	[0.906]		
2.75					44	100	203	428	1074	4587	
					(0.914)	(0.916)	(0.916)	(0.914)	(0.911)	(0.907)	
10.21%					[0.949]	[0.939]	[0.929]	[0.921]	[0.914]	[0.907]	
3.00						51	109	218	454	1132	4818
						(0.919)	(0.920)	(0.919)	(0.917)	(0.912)	(0.907)
10.77%						[0.951]	[0.941]	[0.931]	[0.923]	[0.915]	[0.908]

a) expressed in σu_b (standard deviation of TA's) unit and in percentage.

For a given ETA value, the closer this value to π_m, the higher the progeny group size needed to reach a probability level of 0.9. For such a difference (ETA minus π_m) taken as fixed in σu_b units, variations in the progeny number n are rather less pronounced. These variations (Δn) in n are proportional to variations ($\Delta \hat{\pi}$) in ETA within the range of values considered. For instance, for $\hat{\pi} - \pi_m = 0.75 \; \sigma u_b$, $\Delta n = 504 - 330 = 174$ when $\hat{\pi}$ varies from $+ 1.00$ to $+ 2.00 \; \sigma u_b$ at $\pi_o = 0.025$ and $\Delta n = 671 - 504 = 167$ with $\hat{\pi}$ going from $+ 2.00$ to $+ 3.00 \; \sigma u_b$. Clearly, these variations result from the dependency of the posterior variance on ETA as reflected in formula (6b) contrarily to what happened in the Gaussian linear model. For $\pi_o = 0.025$ and $h^2 = 0.25$ and $\alpha = 0.9$, 73 to 167, 157 to 325, 330 to 671, 904 to 1663 and 4041 to 7062 progeny are required to exceed a minimum TA value equal to ETA minus 1.25, 1.00, 0.75, 0.50 and 0.25 σu_b, respectively. Corresponding figures can be found in table III for $\pi_o = 0.04$. Coming back to the case of twinning, practical interest will be for π_m values around $+1.50 \; \sigma u_b$. This corresponds to sires having an ETA equal to 2.25 to 2.50 σu_b (A difference ETA $-\pi_m$ of 1.00 to 1.25 σu_b seems reasonable in practice). A progeny number of 279 to 546 and 188 to 375 is needed depending on $\hat{\pi} - \pi_m$ for $\pi_m = 0.025$ and 0.04, respectively.
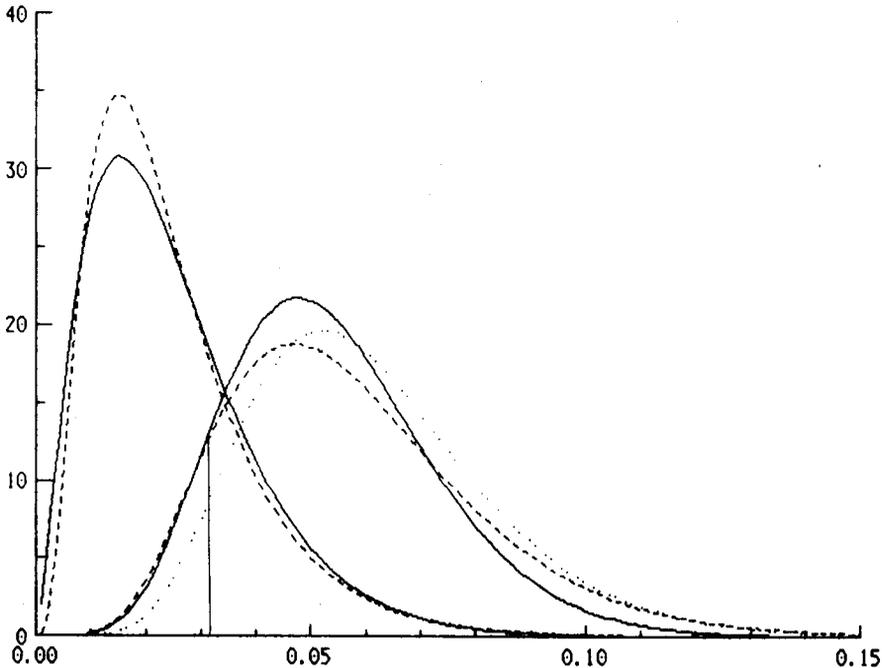

Fig. 2. Prior and posterior distributions of a sire's TA (π) for $\pi_o = 0.025$, $h^2 = 0.25$, $n = 38$ and $p = 0.1316$ (5/38) with probabilities to exceed $\pi_m = 0.03275$ being 0.882, 0.877 and 0.932 for methods I (-); II (- -) and III (. . . .) respectively.

For each n, π_m, ETA, $\lambda_b(\pi_o)$ combination, the corresponding probability that the true TA (μ_i) of sire i exceeds $\mu_m = \Phi^{-1}(\pi_m)$ given n, $p = \hat{\pi} + \lambda_b(\hat{\pi} - \pi_o)/n$ and λ is calculated for methods II and III according to formula (20). As a matter

of fact, methods I, II III can be compared only when using the same amount of information, v.i.z. the same n and p values on the one hand, and the same prior expectations and variances (formulae 13a and b) on the other hand. Probability values for methods II and III are shown in Tables II and III.

The agreement between the 3 methods is generally good, especially between the beta binomial model and the Bayesian approach of the threshold model. For $\pi_m \leq +2.00$ and ETA $\leq +2.25 \, \sigma u_b$, the difference between the 2 probabilities never exceeds 0.01. As pointed out previously (first application), distributions employed to make probability statements in II and III are both asymptotically normal on the underlying scale, and consequently underestimate the real posterior variance and overestimate the probability that the true TA is higher than π_m. Clearly, this drawback is more severe for method III than method II, especially for high values of the ETA as shown in Tables II and III and in figure 2 with the graph of distributions corresponding to $\pi_o = 0.025$, $h^2 = 0.25$, $n = 38$ and $p = 5/38$.

DISCUSSION-CONCLUSION

Adaptation for other practical situations

The situation adressed in this paper was to make probability statements about the true TA of "superior" sires vis-à-vis a minimum (π_m) value for a rare interesting characteristic. The procedure described can also be viewed as making probability statements about TA's of "inferior" sires vis-à-vis an upper TA limit for a frequent unfavourable trait. This is likely to occur in practical animal breeding, especially for the so-called secondary traits (*e.g.* fertility or dystocia in cattle) where selection usually operates against the poorest sires and dams. In some instances for such detrimental traits at a low prevalence, interest might be towards the occurrence of TA's in the lower tail of the distribution. The procedure reported here can be easily accomodated for this case in changing the inequality sign in formula (7) and in taking an opposite argument for $\Phi(\cdot)$ in (20).

Threshold liability concept and beta binomial approaches

In the example studied and within the range of parameters considered, probability statements made by methods I and II were very similar as clearly illustrated by figure 2. This may help reconcile the beta binomial approach praised by statisticians (see, among others Im, 1982) with the threshold liability concept put ahead by quantitative geneticists especially those working in human genetics (see for instance Curnow & Smith, 1975; Falconer, 1965; Fraser, 1980), even under complex segregation patterns (Lalouel *et al.*, 1983). However, the threshold liability model offers more flexibility than the beta binomial. In particular, it can be easily adapted to mixed model structures involving a multipopulation analysis as well as several random sources of variation. Foulley *et al.* (1988) highlighted how inferential issues in such structures can be handled from a unified perspective using the Bayesian paradigm. Thus, taking into account several records per animal (*e.g.* multiple *vs* single births at different parities) can be achieved with this methodology using either a repeatability model or a multiple trait approach (Höschele *et al.*, 1986)

with possible missing data patterns (Foulley & Gianola, 1986) as well as correlated information on normal continuous traits (Foulley *et al.*, 1983). Unfortunately, the beta binomial model in its present state of development remains confined to a single population analysis with one random factor (Williams, 1988).

Gilmour's approach

Although the approach of Gilmour *et al.*, (1985) to the threshold model has its own rationality *via* its connection to the methodology of generalized linear models and quasi-likelihood, its justification for predicting breeding values as compared to Bayesian methods is still questionable (Foulley *et al.*, 1988; Knuiman & Laird, 1988). This is especially true with respect to distribution properties of TA predictors or those of prediction errors for which no formal statistical properties can be claimed for Gilmour's method.

Distribution properties

Distribution properties of the true TA given the data were derived in this study under the implicit assumption that the beta binomial model is true in method I and the threshold concept is also true in methods II and III. Another issue on comparing these methods is to investigate which is the best. Ways of challenging models is a difficult topic which is beyond the scope of this paper (see for instance Smith, 1986). For categorical data and in an animal breeding context, suggestions on how to compare non-nested models were made by Foulley (1987), which are based on the predictive distribution of a future data set given the data observed at present.

Genetic charges

The reasoning followed throughout this paper is a conditional view knowing the progeny test data information on sires. This leads to probability statements about the true TA of specific sires. Discussing progeny group numbers for a planned progeny test programme can involve a different approach (*e.g.* Curnow, 1984) especially when looking *a priori* at this issue with no specific progeny-test results. One could make some statements about a distributional form for ETA's and values of the selection differential and point of truncation. As pointed out by Hill (1977), genetic change due to selecting superior sires on their ETA's under extra constraints with respect to such factors as inbreeding levels, testing facilities and other costs appears to be a natural approach (*e.g.* Curnow, 1984). Response (R) to one generation of upward truncation selection on $\widehat{\pi}(\widehat{\pi} \geq \widehat{\pi}_s)$ is a random variable π having in the beta binomial model the following conditional density given $\widehat{\pi} \geq \pi_o$.

$$f_s(\pi) = \left[\int_{\widehat{\pi}_s}^1 f(\pi, \, \widehat{\pi} \mid n, \pi_o, \rho_b) d\widehat{\pi} \right] \Big/ \left[\int_{\widehat{\pi}_s}^1 f(\widehat{\pi} \mid n, \pi_o, \rho_b) d\widehat{\pi} \right]$$

A similar expression can be written when postulating a threshold model (Foulley, 1987). Clearly, the problem becomes more complex since we have to take into

account, not only the posterior density of π given progeny test results but also the marginal density of the ETA's and integrate $\hat{\pi}$ out in the latter and their product over the selection space.

These are not easy manipulations from an analytical point of view, especially with the threshold model. More research is therefore needed to derive original results in that area.

ACKNOWLEDGEMENTS

REFERENCES

Abramowitz M. & Stegun I.A. (1972) *Handbook of mathematical functions.* 9th ed. Dover publications, New York

Beitler P. & Landis J.R. (1985) A mixed-effects model for categorical data. *Biometrics* 41, 991-1000

Berger J.O. (1985) *Statistical decision theory and bayesian analysis.* 2nd edition. Springer-Verlag, New York

Berger J.P., Freeman A.E. (1978) Prediction of sire merit for calving difficulty. *J. Dairy Sci.* 61, 1146-1150

Cox D.R. & Hinkley D.V. (1974) *Theoretical statistics* Chapman & Hall, London

Curnow R.N. (1984) Progeny-testing for all-or-none traits when a multifactorial model applies. *Biometrics* 40, 375-382

Curnow R.N. & Smith C. (1975) Multifactorial models for familial diseases in man. *J.R. Statist. Soc.* A, 138, 131-169

Dempster E.R. & Lerner I.M. (1950) Heritability of threshold characters. *Genetics* 35, 212-236

Falconer D.S. (1965) The inheritance of liability to certain diseases estimated from the incidence among relatives. *Ann. Hum. Genet.* 29, 51-76

Foulley J.L. (1987) *Méthodes d'évaluation des reproducteurs pour des caractères discrets à déterminisme polygénique en sélection animale.* Thèse d'Etat. Université de Paris-Orsay

Foulley J.L. & Gianola D. (1986) Sire evaluation for multiple binary responses when information is missing on some traits. *J. Dairy Sci.* 69, 2681-2695

Foulley J.L., Gianola D. & Thompson R. (1983) Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet. Sel. Evol.* 15, 401-424

Foulley J.L., Gianola D. & Im S. (1988) Genetic evaluation for discrete polygenic traits in animal breeding. *In: Advances in Statistical Methods for Genetic Improvement of Livestock.* February 16-20, 1987, Armidale. Australia. Springer Verlag (forthcoming)

Fraser F.C. (1980) The William Allan Memorial Award Adress: Evolution of a palatable multifactorial threshold model. *Am. J. Hum. Genet.* 32, 796-813

Frebling J., Gillard P. & Menissier F. (1987) *Résultats préliminaires sur l'aptitude naturelle à la gémellité dans un troupeau de vaches Charolaises et Maine-Anjou et de leur filles.* INRA, mimeo

Gianola D. (1982) Theory and analysis of threshold characters. *J. Anim. Sci.* 54, 1079-1096

Gianola D. & Foulley J.L. (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15, 201-224

Gilmour A.R., Anderson R.D. & Rae A.L. (1985) The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 593-599

Harville D.A. & Mee R.W. (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408

Henderson C.R. (1973) Sire evaluation and genetic trends. *Proceedings of the animal breeding and genetics symposium in honor of Dr. J.L. Lush.* American society of animal science and american dairy science association, Champaign, Illinois, 10-41

Hill W.G. (1977) Comments on statistical efficiency in bull progeny testing for calving difficulty. *Livest. Prod. Sci.* 4, 203-207

Höschele I., Foulley J.L. & Gianola D. (1986) Genetic evaluation for multiple binary responses. *Genet. Sel. Evol.* 18, 299-321

Im S. (1982) *Contribution à l'étude des tables de contingence à paramètres aléatoires: utilisation en biométrie.* Thèse 3e cycle, Université Paul-Sabatier, Toulouse

Im S. (1986) Mutual independence in a mixed effects model letter to the editor. *Biometrics* 42, 997

Im S. (1987) A note on heritability of a dichotomous trait. *Biom. J.* 4, 407-411

Im S., Foulley J.L., & Gianola D. (1987) A linear model for genetic evaluation on categorical traits. 82nd Annual Meeting of the American Dairy Science Association Columbia Missouri, June 21-24. *J. Dairy Sci.* 70, Suppl. 1, 124 abstract

Johnson N.L. & Kotz S. (1972) *Distributions in Statistics Continuous Univariance Distributions,* vol. 3. John Wiley & Sons, New York

Knuiman M.W. & Laird N.M. (1988) Parameter estimation in variance component models for binary response data. *In: Advances in Statistical Methods for Genetic Improvement of Livestock.* February 16-20, 1987, Armidale. Australia. Springer Verlag (forthcoming)

Lalouel M., Rao D.C., Morton N.E. & Elston R.C. (1983) A unified model. *J. Hum. Genet.* 35, 816-826

Maijala K. & Syvajarvi (1977) On the possibility of developing multiparous cattle by selection. *Z. Tierz. Zuechtungsbiol.* 94, 136-150

Meijering A. & Gianola D. (1985) Observations on sire evaluation with categorical data using heteroscedastic mixed linear models. *J. Dairy Sci,* 68, 1226-1232

Pearson K. (1968) *Tables of the incomplete beta function.* 2nd ed. Cambridge University Press, Cambridge

Rao C.R. (1973) *Linear statistical inferences and its applications.* 2nd ed. John Wiley & Sons, New York

Schaeffer L.R. & Wilton J.W. (1976) Methods of sire evaluation for calving ease. *J. Dairy Sci.* 59, 544-551

Smith A.F.M. (1986) Some Bayesian thoughts on modelling and model choice. *The Statistician* 35, 97-102

Stiratelli R., Laird N. & Ware J.H. (1984) Random effects models for serial observations with binary response. *Biometrics* 40, 961-971

Syrstad O. (1974) Inheritance of multiple births in cattle. *Livest. Prod. Sci.* 11, 373-380

Williams D.A. (1988) Extra binomial variation in toxicology. *XIVth International biometric conference*, Namur, July 18-23, 1988, invited papers, 301-316

Wright S. (1934a) An analysis of variability in number of digits in an inbred strain of guinea pig. *Genetics* 19, 506-536

Wright S. (1934b) The results of crosses between inbred strains of guinea pigs differing in number of digits. *Genetics* 19, 537-551

Zellner A. & Rossi P.E. (1984) Bayesian analysis of dichotomous quantal response models. *J. Econ.* 25, 365-393

APPENDIX A: *ALGORITHM FOR SOLVING EQUATION (7)*

Analytical approximations to the inverse function of the incomplete data function are proposed in calculus books: see, for instance, formula 26.5.22, page 945 in Abramowitz & Stegun (1972). To improve the accuracy of these approximations, we employed an algorithm based on finite difference techniques. It consists in iterating from (t) to $(t+1)$ with

$$\Delta x_{t+1} = \Delta x_t [(1 - \alpha) - I_{x_t}(a,b)]/I_{x_t}(a,b) - [I_{x_{t-1}}(a,b)] \tag{A1}$$

Starting values are $x_o = 0$; $I_o(a,b) = 0$ and $x_1 =$ value of the inverse probability function given by *e.g.* Abramowitz & Stegun.

In this equation, the incomplete beta function is calculated from Peizer and Pratt's approximation as recommended by Pearson (1968, formulae 47, 48, 49 page XXX) and Johnson & Kotz (1972, vol. III, pp. 48, 49). For the values of a and b encountered in this study, this approximation is very accurate; the difference |true-approximated values| is less than 10^{-3} according to the previous authors.

Computationally, solving equation (7) in terms of $(n, \hat{\pi})$ requires obtaining the parameter $a = (n + \lambda_b)\hat{\pi}$ of the incomplete beta function $I_x(a,b) = 1 - \alpha$, given x values of α, x and the ratio $a/b = \hat{\pi}/(1 - \hat{\pi})$. This can be worked out with an algorithm similar to [A1] in which a is substituted for x.

For the sake of comparison, n was predicted in the fashion described, and also, using a second approximation to the incomplete beta function (formulae 26.5.21, page 945) in Abramowitz & Stegun (1972). In the examples considered, the agreement between both approximations was excellent; results obtained differed by no more than one progeny.