Genet. Sel. Evol. 34 (2002) 307–333

DOI: 10.1051/gse:2002010

Original article

Prediction error variance and expected response to selection, when selection is based on the best predictor - for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits

Inge Riis KORSGAARD[a]*, Anders Holst ANDERSEN[b], Just JENSEN[a]

[a] Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, P.O. Box 50, 8830 Tjele, Denmark
[b] Department of Theoretical Statistics, University of Aarhus, 8000 Aarhus-C, Denmark

(Received 9 January 2001; accepted 9 February 2002)

Abstract – In this paper, we consider selection based on the best predictor of animal additive genetic values in Gaussian linear mixed models, threshold models, Poisson mixed models, and log normal frailty models for survival data (including models with time-dependent covariates with associated fixed or random effects). In the different models, expressions are given (when these can be found - otherwise unbiased estimates are given) for prediction error variance, accuracy of selection and expected response to selection on the additive genetic scale and on the observed scale. The expressions given for non Gaussian traits are generalisations of the well-known formulas for Gaussian traits - and reflect, for Poisson mixed models and frailty models for survival data, the hierarchal structure of the models. In general the ratio of the additive genetic variance to the total variance in the Gaussian part of the model (heritability on the normally distributed level of the model) or a generalised version of heritability plays a central role in these formulas.

accuracy of selection / best predictor / expected response to selection / heritability / prediction error variance

* Correspondence and reprints
E-mail: IngeR.Korsgaard@agrsci.dk

1. INTRODUCTION

For binary threshold characters heritability has been defined on the underlying scale (liability scale) and on the observed scale (outward scale) (see [4] and [14]), and the definitions were generalised to ordered categorical traits by Gianola [9]. For Poisson mixed models a definition of heritability can be found in [8], and for survival traits we find several definitions of heritability, see *e.g.* [5,10,11] and [16]. In this paper we consider selection based on the best predictor and the goal is to find out, whether heritability (and which one) plays a central role in formulas for prediction error variance, accuracy of selection and for expected response to selection in mixed models frequently used in animal breeding.

For the Gaussian linear mixed model, the best predictor of individual breeding values, $\hat{a}_i^{bp}$, is linear, *i.e.* a linear function of data, y_i, and under certain conditions given by $\hat{a}_i^{bp} = h^2 (y_i - x_i\beta)$, where h^2 is the heritability of the trait, given by the ratio of the additive genetic variance to the total phenotypic variance, σ_a^2/σ_p^2. In this model accuracy of selection, defined by the correlation between a_i and $\hat{a}_i^{bp}$, is equal to the square root of heritability, *i.e.* $\rho(a_i, \hat{a}_i^{bp}) = h$; and prediction error variance is $\sigma_a^2(1 - h^2)$. The joint distribution of $(a_i, \hat{a}_i^{bp})$ is a bivariate normal distribution that does not depend on fixed effects. Furthermore, if parents of the next generation are chosen based on the best predictor of their breeding values, then the expected response to selection, that can be obtained on the phenotypic scale in the offspring generation (compared to a situation with no selection) is equal to the expected response that can be obtained on the additive genetic scale. The expected response that can be obtained on the additive genetic scale is $\frac{1}{2}h^2 S_{\mathrm{f}} + \frac{1}{2}h^2 S_{\mathrm{m}}$, where S_{f} and S_{m} are expected selection differentials in fathers and mothers, respectively. The expected selection differential does not depend on fixed effects. These results are all very nice properties of the Gaussian linear mixed model with additive genetic effects. We observe (or know) that heritability plays a central role.

In general, if $\mathbf{U}$ and $\mathbf{Y}$ denote vectors of unobservable and observable random variables, then the best predictor of $\mathbf{U}$ is the conditional mean of $\mathbf{U}$ given $\mathbf{Y}$, $\widehat{\mathbf{U}}^{bp} = E(\mathbf{U}|\mathbf{Y})$. The observed value of $\widehat{\mathbf{U}}^{bp}$ is $\hat{\mathbf{u}}^{bp} = E(\mathbf{U}|\mathbf{Y} = \mathbf{y})$ (a predictor is a function of the random vector, $\mathbf{Y}$, associated with observed data). This predictor is best in the sense that it has minimum mean square error of prediction, it is unbiased (in the sense that $E(\widehat{\mathbf{U}}^{bp}) = E(\mathbf{U})$), and it is the predictor of U_i with the highest correlation to U_i. Furthermore, by selecting any upper fraction of the population on the basis of $\widehat{u_i}^{bp}$, then the expected value of U_i (in the selected proportion) is maximised. These properties, which are reasons for considering selection based on the best predictor, and a lot of other results on the best predictor are summarised in [12] (see also references

in [12]). In this paper **U** will be associated with animal additive genetic values and we consider selection based on the best predictor of animal additive genetic values.

The purpose of the paper is to give expressions for the best predictor, prediction error variance, accuracy of selection, expected response to selection on the additive genetic and on the phenotypic scale in a series of models frequently used in animal breeding, namely the Gaussian linear mixed model, threshold models, Poisson mixed models and models for survival traits. The models for survival traits include Weibull and Cox log normal frailty models with time-dependent covariates with associated fixed and random effects. Part of the material in this paper can be found in the literature (mainly results for the Gaussian linear mixed model), and has been included for comparison. Some references (not exhaustive) are given in the discussion. The models we consider are animal models. We will work under the assumptions of the infinitesimal, additive genetic model, and secondly that all parameters of the different models are known.

The structure of the paper is as follows: in Section 2, the various models (four models) we deal with are specified. Expressions for the best predictor, for prediction error variance and accuracy of selection, and for expected response to selection in the different models are given in Sections 3, 4 and 5 respectively. These chapters start with general considerations, next each of the four models are considered and each chapter ends with its own discussion and conclusion. The paper ends with a general conclusion.

2. THE MODELS

Notation 1 *Usually capital letters (*e.g. U_i *and* $\mathbf{U}$*) are used as the notation for a random variable or a random vector; and lower case letters (*e.g. u_i *and* $\mathbf{u}$*) are used as the notation for a specific value of the random variable or the random vector. In this paper we will sometimes use lower case letters (*e.g. a_i *and* $\mathbf{a}$*) for a random variable or a random vector, and sometimes for a specific value of the random variable or the random vector. The interpretation should be clear from the context.*

2.1. Linear mixed model

The animal model is given by

$$Y_i = x_i\beta + a_i + e_i$$

for $i = 1, \ldots, n$, with $\mathbf{a} \sim N_n\left(\mathbf{0}, \mathbf{A}\sigma_a^2\right)$ and $\mathbf{e} \sim N_n\left(\mathbf{0}, \mathbf{I}_n\sigma_e^2\right)$; furthermore $\mathbf{a}$ and $\mathbf{e}$ are assumed to be independent.

2.2. Threshold model

The animal model, for an ordered categorical threshold character with $K \geq 2$ categories, is given by

$$Y_i = \begin{cases} 1 & \text{if } -\infty < U_i \leq \tau_1 \\ 2 & \text{if } \tau_1 < U_i \leq \tau_2 \\ & \vdots \\ K-1 & \text{if } \tau_{K-2} < U_i \leq \tau_{K-1} \\ K & \text{if } \tau_{K-1} < U_i < \infty \end{cases} \tag{1}$$

where $-\infty < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < \infty$, $U_i = x_i\beta + a_i + e_i$, for $i = 1, \ldots, n$ and $\mathbf{a} \sim N_n\left(\mathbf{0}, \mathbf{A}\sigma_a^2\right)$, $\mathbf{e} \sim N_n\left(\mathbf{0}, \mathbf{I}_n\sigma_e^2\right)$, $\mathbf{a}$ and $\mathbf{e}$ are assumed to be independent. Let $\mathbf{X}$ denote the design matrix associated with fixed effects on the underlying scale, the U-scale (or the liability scale). For reasons of identifiability and provided that the vector of ones, $\mathbf{1}$, belongs to the span of the columns of $\mathbf{X}$, then without loss of generality we can assume that $\tau_1 = 0$ and $\sigma_a^2 + \sigma_e^2 = 1$ (or instead of a restriction on $\sigma_a^2 + \sigma_e^2$ we could have put a restriction on only σ_a^2 or σ_e^2 or one of the thresholds, $\tau_2, \ldots, \tau_{K-1}$ (the latter only in case $K \geq 3$)).

2.3. Poisson mixed model

The Poisson animal model is defined by $Y_i|\boldsymbol{\eta} \sim Po\left(\lambda_i\right)$, where $\lambda_i = \exp\left(\eta_i\right)$ with η_i given by

$$\eta_i = \log\left(\lambda_i\right) = x_i\beta + a_i + e_i \tag{2}$$

for $i = 1, \ldots, n$, where $\mathbf{a} \sim N_n\left(\mathbf{0}, \mathbf{A}\sigma_a^2\right)$ and $\mathbf{e} \sim N_n\left(\mathbf{0}, \mathbf{I}_n\sigma_e^2\right)$, furthermore $\mathbf{a}$ and $\mathbf{e}$ are assumed to be independent, and conditional on $\boldsymbol{\eta}$ (the vector of $\eta_i's$) then all of the $Y_i's$ are assumed to be independent.

2.4. Survival model

Consider the Cox log normal animal frailty model with time-dependent covariates for survival times $(T_i)_{i=1,\ldots,n}$. The time-dependent (including time-independent) covariates of animal i are $x_i\left(t\right) = \left(x_{i1}, x_{i2}\left(t\right)\right)$, with associated fixed effects, $\beta = \left(\beta_1, \beta_2\right)$, and $z_i\left(t\right)$, with associated random effects, $\mathbf{u}_2$. The dimension of β_1 (β_2) is p_1 (p_2), and the dimension of $\mathbf{u}_2$ is q_2. The hazard function for survival time T_i is, conditional on random effects, $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{a}, \mathbf{e})$, given by

$$\lambda_i\left(t|\mathbf{u}_1, \mathbf{u}_2, \mathbf{a}, \mathbf{e}\right) = \lambda_0\left(t\right)\exp\left\{x_i\left(t\right)\beta + z_i\left(t\right)\mathbf{u}_2 + u_{1l(i)} + a_i + e_i\right\} \tag{3}$$

for $i = 1, \ldots, n$; $l\left(i\right) \in \{1, \ldots, q_1\}$. The baseline hazard, λ_0: $[0, \infty) \to [0, \infty)$ is assumed to satisfy $\Lambda_0\left(t\right) < \infty$ for all $t \in [0, \infty)$, with

$\lim_{t\to\infty} \Lambda_0(t) = \infty$, where $\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s$ is the integrated baseline hazard function. Besides this, $\lambda_0(\cdot)$ is completely arbitrary. The time-dependent covariates, $x_i(t)$ and $z_i(t)$, are assumed to be left continuous and piecewise constant. Furthermore, the time-dependent covariate, $z_i(t)$, is, for $t \in [0, \infty)$, assumed to be a vector with exactly one element $z_{ik'}(t) = 1$, and $z_{ik}(t) = 0$ for $k \neq k'$. Let $\mathbf{u}_1 = \left(u_{1j}\right)_{j=1,\ldots,q_1}$, $\mathbf{a} = (a_i)_{i=1,\ldots,n}$ and $\mathbf{e} = (e_i)_{i=1,\ldots,n}$, then with regards to the random effects it is assumed that $\mathbf{u}_1 \sim N_{q_1}\left(\mathbf{0}, \mathbf{I}_{q_1}\sigma^2_{u_1}\right)$, $\mathbf{u}_2 \sim N_{q_2}\left(\mathbf{0}, \mathbf{I}_{q_2}\sigma^2_{u_2}\right)$, $\mathbf{a} \sim N_n\left(\mathbf{0}, \mathbf{A}\sigma^2_a\right)$ and $\mathbf{e} \sim N_n\left(\mathbf{0}, \mathbf{I}_n\sigma^2_e\right)$. Furthermore $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{a}$ and $\mathbf{e}$ are assumed to be independent. In this model and conditional on $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{a}, \mathbf{e})$, then all of the $T_i's$ are assumed to be independent. In the following we let $\boldsymbol{\eta} = (\eta_i)_{i=1,\ldots,n}$ with $\eta_i = u_{1l(i)} + a_i + e_i$.

Notation 2 *We introduce the following partitioning of R_+ defined by jumps in the covariate processes $\left(x_i(\cdot), z_i(\cdot)\right)_{i=1,\ldots,n}$: $R_+ = \cup_{m=1}^{P}(l_\mathrm{m}, r_\mathrm{m}]$, with $1 \leq P \leq \infty$; the subsets are disjoint (but not necessarily ordered in the sense that $r_\mathrm{m} = l_{m+1}$ for $m = 1, \ldots, P-1$).*

With $\Lambda_0(\cdot)$ and β_2 known, let the function $h_i^{\mathbf{u}_2}(t)$, conditional on $\mathbf{u}_2$, be defined by

$$h_i^{\mathbf{u}_2}(t) = \int_0^t \lambda_0(s) \exp\left\{x_{i2}(s)\,\beta_2 + z_i(s)\,\mathbf{u}_2\right\} \mathrm{d}s.$$

We note that for $t \in (l_{m'}, r_{m'}]$ with $m' \in \{1, \ldots, P\}$, then

$$h_i^{\mathbf{u}_2}(t) = \sum_{\substack{m=1 \\ m: r_\mathrm{m} < t}}^{P} \exp\left\{x_{i2}(r_\mathrm{m})\,\beta_2 + z_i(r_\mathrm{m})\,\mathbf{u}_2\right\}\left(\Lambda_0(r_\mathrm{m}) - \Lambda_0(l_\mathrm{m})\right) + \exp\left\{x_{i2}(t)\,\beta_2 + z_i(t)\,\mathbf{u}_2\right\}\left(\Lambda_0(t) - \Lambda_0(l_{m'})\right).$$

With $\Lambda_0(\cdot)$ and β_2 known, then, conditional on $\mathbf{u}_2$, it can be shown (see Appendix or a minor generalisation of the proof in Appendix) that the model in (3) is equivalent to a linear model on the $\log\left(h_i^{\mathbf{u}_2}(\cdot)\right)$-scale, *i.e.*

$$\tilde{Y}_i = \log\left(h_i^{\mathbf{u}_2}(T_i)\right) = -x_{i1}\beta_1 - u_{1l(i)} - a_i - e_i + \varepsilon_i$$

where ε_i follows an extreme value distribution, with $E(\varepsilon_i) = -\gamma_\mathrm{E}$, where γ_E is the Euler constant, and $Var(\varepsilon_i) = \pi^2/6$; all of the $\varepsilon_i's$ are independent, and independent of $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{a}$ and $\mathbf{e}$. Note that the scale is specific for each animal (or groups of animals with the same time-dependent covariates). Next let $g_i^{\mathbf{u}_2}$, still conditional on $\mathbf{u}_2$, be an inverse of $\log h_i^{\mathbf{u}_2}$ (*i.e.* $g_i^{\mathbf{u}_2}\left(\log h_i^{\mathbf{u}_2}(T_i)\right) = T_i$ with probability one) then

$$T_i = g_i^{\mathbf{u}_2}\left(-x_{i1}\beta_1 - u_{1l(i)} - a_i - e_i + \varepsilon_i\right).$$

Note the following special cases: Without time-dependent covariates (with associated fixed or random effects) the model in (3) is equivalent to a linear model on the $\log\left(\Lambda_0(\cdot)\right)$-scale, *i.e.* the linear scale is the same for all animals. Furthermore, without time-dependent covariates, and if the baseline hazard is that of a Weibull distribution ($\Lambda_0(t) = (\gamma t)^{\alpha}$), then the model in (3) is a log linear model for T_i given by

$$\tilde{Y}_i = \log(T_i) = -\log(\gamma) - \frac{1}{\alpha}x_{i1}\beta_1 - \frac{1}{\alpha}u_{1l(i)} - \frac{1}{\alpha}a_i - \frac{1}{\alpha}e_i + \frac{1}{\alpha}\varepsilon_i$$

where ε_i follows an extreme value distribution; all of the $\varepsilon_i's$ are independent and independent of $\mathbf{u}_1$, $\mathbf{a}$ and $\mathbf{e}$.

3. BEST PREDICTOR

Assume that we have a population of unrelated and noninbred potential parents, the base population, *i.e.* it is assumed that the vector of breeding values of potential parents is multivariate normally distributed with mean zero and co(variance) matrix $\mathbf{I}_n\sigma_a^2$, where n is the number of animals in the base population. The trait, which we want to improve by selection is either a normally distributed trait, a threshold character, a character following a Poisson mixed model or a survival trait. The models are animal models and assumed to be as described in Section 2, except that $\mathbf{a} \sim N_n\left(\mathbf{0}, \mathbf{I}_n\sigma_a^2\right)$. For each trait, and based on a single record per animal, we will give the best predictor of the breeding values of the potential parents.

First some general considerations which will mainly be used for Poisson mixed models and models for survival data: Let $\mathbf{a} = (a_i)_{i=1,\ldots,n}$ denote the vector of breeding values of animals in the base population (potential parents) then the best predictor of a_i is given by $E(a_i|data)$. If we can find some vector $\mathbf{v} = (v_i)_{i=1,\ldots,N}$, with $(a_i, \mathbf{v})$ following a multivariate normal distribution and with the property that a_i and *data* are conditionally independent given $\mathbf{v}$ (*i.e.* $p(a_i|\mathbf{v},data) = p(a_i|\mathbf{v})$) then the best predictor of a_i is

$$\begin{aligned} E(a_i|data) &= E_{\mathbf{v}|data}\left(E(a_i|\mathbf{v},data)\right) \\ &= E_{\mathbf{v}|data}\left(E(a_i|\mathbf{v})\right) \\ &= E_{\mathbf{v}|data}\left(Cov(a_i, \mathbf{v})Var(\mathbf{v})^{-1}\left(\mathbf{v} - E(\mathbf{v})\right)\right) \end{aligned} \tag{4}$$

The last equation follows because the conditional distribution of a_i given $\mathbf{v}$ is normal. A further simplification can be obtained if the dimension of $\mathbf{v}$ is n, *i.e.* $N = n$, and $p(a_i|\mathbf{v}) = p(a_i|v_i)$, in which case (4) simplifies to

$$\begin{aligned} E(a_i|data) &= E_{\mathbf{v}|data}\left(Cov(a_i, v_i)\,Var(v_i)^{-1}\left(v_i - E(v_i)\right)\right) \\ &= Cov(a_i, v_i)\,Var(v_i)^{-1}\left(E(v_i|data) - E(v_i)\right). \end{aligned} \tag{5}$$

The best predictor of the breeding values of potential parents is given below for each of the four models.

3.1. Linear mixed model

In the linear mixed model we have the well-known formula

$$\hat{a}_i^{bp} = h^2 (y_i - x_i\beta)$$

where y_i is the phenotypic value of animal i, and $h^2 = \sigma_a^2 / \left(\sigma_a^2 + \sigma_e^2\right)$.

3.2. Threshold model

The best predictor of a_i is

$$\begin{aligned}\hat{a}_i^{bp} &= E (a_i | Y_i = y_i) \\ &= \int E (a_i | U_i = u_i)\, p (u_i | Y_i = y_i)\, \mathrm{d}u_i\end{aligned}$$

where $E (a_i | U_i = u_i) = h^2_{\text{nor}} (u_i - x_i\beta)$ and

$$p (u_i | Y_i = y_i) = \frac{p (u_i)}{P (\tau_{k-1} < U_i \leq \tau_k)} \quad \text{if } y_i = k \quad \text{and } \tau_{k-1} < u_i \leq \tau_k$$

for $k = 1, \ldots, K$, with $\tau_0 = -\infty$, $\tau_1 = 0$ and $\tau_K = \infty$. The heritability, h^2_{nor}, on the normally distributed liability scale, the U-scale, is $h^2_{\text{nor}} = \sigma_a^2 / \left(\sigma_a^2 + \sigma_e^2\right) = \sigma_a^2$ because $\sigma_a^2 + \sigma_e^2 = \sigma_u^2 = 1$ and $p (u_i)$ is the density function of U_i. It follows that the best predictor of a_i is, for $Y_i = k$, given by

$$\begin{aligned}\hat{a}_i^{bp} &= h^2_{\text{nor}} \left(E (U_i | Y_i = k) - x_i\beta\right) \\ &= h^2_{\text{nor}} \frac{\varphi (\tau_{k-1} - x_i\beta) - \varphi (\tau_k - x_i\beta)}{P (Y_i = k)} \\ &= h^2_{\text{nor}} \frac{\varphi (\tau_{k-1} - x_i\beta) - \varphi (\tau_k - x_i\beta)}{\Phi (\tau_k - x_i\beta) - \Phi (\tau_{k-1} - x_i\beta)}\end{aligned}$$

where φ (Φ) is the density function (distribution function) of a $N (0, 1)$-distribution.

Note, in particular, for binary threshold characters ($K = 2$) we have

$$\hat{a}_i^{bp} = \frac{h^2_{\text{nor}} \varphi (-x_i\beta)}{\Phi (x_i\beta) (1 - \Phi (x_i\beta))} [Y_i - E (Y_i)]$$

i.e. the best predictor of a_i is linear in y_i; $E (Y_i) = 1 + P (Y_i = 2)$

3.3. Poisson mixed model

In the Poisson mixed model we may use (4) and (5) with $\mathbf{v} = \boldsymbol{\eta} = (\eta_i)_{i=1,\ldots,n}$, where η_i is given by (2). Realising that the conditional density of η_i given *data* is equal to the conditional density of η_i given y_i (*i.e.* $p(\eta_i|data) = p(\eta_i|Y_i = y_i)$) then

$$\hat{a}_i^{bp} = h_{\text{nor}}^2 \left[E(\eta_i|Y_i = y_i) - x_i\beta \right]$$

where $h_{\text{nor}}^2 = \sigma_a^2 / \left(\sigma_a^2 + \sigma_e^2\right)$ and the conditional density of η_i given $Y_i = y_i$ is given by

$$p(\eta_i|Y_i = y_i) = \frac{P(Y_i = y_i|\eta_i)\, p(\eta_i)}{\int_{-\infty}^{\infty} P(Y_i = y_i|\eta_i)\, p(\eta_i)\, d\eta_i}.$$

3.4. Survival model

Notation 3 *Let T_i and C_i denote the random variables associated with the survival time and the censoring time of animal i. We observe $Y_i = \min\{T_i, C_i\}$ and $\delta_i = 1\{T_i \leq C_i\}$. For all of the survival traits we let $data_i = (y_i, \delta_i)$, where y_i is the observed value of the survival time (censoring time) of animal i, depending on the observed value of the censoring indicator. Furthermore we let $data = (data_i)_{i=1,\ldots,n}$ denote data on all animals.*

Assumption 1: For all of the survival traits we will assume that conditional on random effects, then censoring is non-informative of random effects.

For survival traits we will use (4) with $\mathbf{v}$ given as described in the following: For each animal i, we introduce the following m_i random variables: $\{u_{2l} + \eta_i\}_{l \in B_i}$, where B_i consist of those coordinates of the vector $z_i(\cdot)$, which are equal to 1 for some $t \leq y_i$; *i.e.* $m_i = |B_i|$.

Next we let $m = \sum_{i=1}^{n} m_i$, and introduce the random vector $\mathbf{v} = \left(\mathbf{v}_1', \ldots, \mathbf{v}_n'\right)'$ with

$$v_{ij} = u_{2l_{(j)}^i} + \eta_i$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$ where $l_{(1)}^i < \cdots < l_{(m_i)}^i$ are the ordered elements of B_i. The joint distribution of $\mathbf{v}$ is given by

$$\mathbf{v} \sim N_{\text{m}}\left(\mathbf{0}, Var(\mathbf{v})\right)$$

with

$$Var(\mathbf{v}) = \mathbf{Z} Var(\mathbf{u}_2) \mathbf{Z}' + \mathbf{M}$$

where $\mathbf{M}$ is a matrix with blocks $\mathbf{M}_{ik}$, $\mathbf{M} = (\mathbf{M}_{ik})_{i,k=1,\ldots,n}$, and with $\mathbf{M}_{ik}$ given by

$$\mathbf{M}_{ik} = \begin{cases} \mathbf{1}_{m_i \times m_i} \left(\sigma_{u_1}^2 + \sigma_a^2 + \sigma_e^2\right) & \text{for } i = k \\ \mathbf{1}_{m_i \times m_k} \left(1\{l(i) = l(k)\} \sigma_{u_1}^2\right) & \text{for } i \neq k \end{cases}$$

and the (i,j)'th row of the matrix $\mathbf{Z}$, is the vector with all elements equal to zero except for the $l^i_{(j)}$'th coordinate, which is equal to one.

Using (4) with $\mathbf{v}$ given as described above, then the best predictor of a_i is

$$\hat{a}_i^{bp} = Cov\left(a_i, \mathbf{v}\right)\left(Var\left(\mathbf{v}\right)\right)^{-1} E\left(\mathbf{v}|data\right)$$

where $p\left(\mathbf{v}|data\right) = p\left(data|\mathbf{v}\right)p\left(\mathbf{v}\right)\big/p\left(data\right)$ (using the Bayes formula). It follows, under *Assumption 1*, that $p\left(\mathbf{v}|data\right)$ up to proportionality is given by

$$\begin{aligned} p\left(\mathbf{v}|data\right) &\propto \prod_{i=1}^{n}\left[\left(\lambda_i\left(y_i|\mathbf{v}\right)S_i\left(y_i|\mathbf{v}\right)\right)^{\delta_i}\left(S_i\left(y_i|\mathbf{v}\right)\right)^{1-\delta_i}\right] \times p\left(\mathbf{v}\right) \\ &\propto \prod_{i=1}^{n}\left(\lambda_0\left(y_i\right)\exp\left\{x_i\left(y_i\right)\beta + z_i\left(y_i\right)\mathbf{u}_2 + \eta_i\right\}\right)^{\delta_i} \\ &\quad \times \exp\left\{-\left(\sum_{i=1}^{n}\sum_{j=1}^{P} k_{ij}\exp\left\{x_{i2}\left(r_j\right)\beta_2 + z_i\left(r_j\right)\mathbf{u}_2 + \eta_i\right\}\right)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}\mathbf{v}'Var(\mathbf{v})^{-1}\mathbf{v}\right\} \\ &= f\left(\mathbf{v}\right) \end{aligned}$$

where

$$k_{ij} = \exp\left\{x_{i1}\beta_1\right\} \times \begin{cases} \left(\Lambda_0\left(r_j\right) - \Lambda_0\left(l_j\right)\right) & \text{if } r_j < y_i \\ \left(\Lambda_0\left(y_i\right) - \Lambda_0\left(l_j\right)\right) & \text{if } y_i \in \left(l_j, r_j\right] \\ 0 & \text{if } y_i \le l_j \end{cases}.$$

It follows that $p\left(\mathbf{v}|data\right) = f\left(\mathbf{v}\right)\big/\int f\left(\mathbf{v}\right)\mathrm{d}\mathbf{v}$ with $f\left(\mathbf{v}\right)$ as given above.

Note, in the Cox frailty model without time-dependent covariates and with $\mathbf{u}_1$ absent, *i.e.* the special case of (3) with $\lambda_i\left(t|\mathbf{a}, \mathbf{e}\right) = \lambda_0\left(t\right)\exp\left\{x_{i1}\beta_1 + a_i + e_i\right\}$, then we could use (5) with $\mathbf{v} = \boldsymbol{\eta} = \left(a_i + e_i\right)_{i=1,\ldots,n}$. And because, in this model, $p\left(\eta_i|data\right) = p\left(\eta_i|data_i\right)$, then we obtain $\hat{a}_i^{bp} = h^2_{\text{nor}}E\left(\eta_i|data_i\right)$ where $h^2_{\text{nor}} = \sigma_a^2/\left(\sigma_a^2 + \sigma_e^2\right)$ and

$$p\left(\eta_i|data_i\right) \propto \left(\exp\left\{x_{i1}\beta_1 + \eta_i\right\}\right)^{\delta_i}\exp\left\{-\Lambda_0\left(y_i\right)\exp\left\{x_{i1}\beta_1 + \eta_i\right\}\right\} \times \exp\left\{-\frac{1}{2\left(\sigma_a^2 + \sigma_e^2\right)}\eta_i^2\right\}.$$

Example 1. Consider two unrelated and noninbred animals, 1 and 2, and three time periods $(0, r_1]$, $(l_2, r_2]$ and $(l_3, \infty]$, with $r_1 = l_2$ and $r_2 = l_3$ and with

associated random effects u_{21}, u_{22} and u_{23}. Animal 1 is born in period 1 (spent t_{11} units of time in this period) and died or was censored in period 2 (observed to spend $y_1 - t_{11}$ units of time in period 2). Animal 2 is born in period 2 (spent t_{21} units of time in this period) and died or was censored in period 3 (observed to spend $y_2 - t_{21}$ units of time in period 3). Assume that the hazard functions of animal 1 and 2, conditional on random effects are given by

$$\lambda_1(t|\mathbf{u}_2, \boldsymbol{\eta}) = \begin{cases} \lambda_0(t)\exp\{x_1\beta + u_{21} + \eta_1\} & \text{for } t \le t_{11} \\ \lambda_0(t)\exp\{x_1\beta + u_{22} + \eta_1\} & \text{for } t_{11} < t \le t_{11} + (r_2 - l_2) \\ \lambda_0(t)\exp\{x_1\beta + u_{23} + \eta_1\} & \text{for } t_{11} + (r_2 - l_2) < t \end{cases}$$

and

$$\lambda_2(t|\mathbf{u}_2, \boldsymbol{\eta}) = \begin{cases} \lambda_0(t)\exp\{x_2\beta + u_{22} + \eta_2\} & \text{for } t \le t_{21} \\ \lambda_0(t)\exp\{x_2\beta + u_{23} + \eta_2\} & \text{for } t_{21} < t \end{cases}$$

respectively; here $\eta_1 = a_1 + e_1$ and $\eta_2 = a_2 + e_2$. In this example $m_1 = m_2 = 2$ and

$$\begin{aligned} v_{11} &= u_{21} + \eta_1 \\ v_{12} &= u_{22} + \eta_1 \\ v_{21} &= u_{22} + \eta_2 \\ v_{22} &= u_{23} + \eta_2 \end{aligned}$$

with $Var(\mathbf{v})$ given by

$$Var(\mathbf{v}) = \begin{pmatrix} \sigma^2_{u_2} + \sigma^2_a + \sigma^2_e & \sigma^2_a + \sigma^2_e & 0 & 0 \\ \sigma^2_a + \sigma^2_e & \sigma^2_{u_2} + \sigma^2_a + \sigma^2_e & \sigma^2_{u_2} & 0 \\ 0 & \sigma^2_{u_2} & \sigma^2_{u_2} + \sigma^2_a + \sigma^2_e & \sigma^2_a + \sigma^2_e \\ 0 & 0 & \sigma^2_a + \sigma^2_e & \sigma^2_{u_2} + \sigma^2_a + \sigma^2_e \end{pmatrix}.$$

3.5. Discussion and conclusion

For all of the (animal) models considered it was realised or found that heritability, h^2_{nor} (the ratio between the additive genetic variance and the total variance at the normally distributed level of the model) or a generalised version of heritability, $Cov(a_i, \mathbf{v})\left(Var(\mathbf{v})\right)^{-1}$, plays a central role in formulas for the best predictor.

4. PREDICTION ERROR VARIANCE AND ACCURACY OF SELECTION

Having derived the best predictor of breeding values in different models, then we may want to find the prediction error variance, $PEV = E\left(\left(\hat{a}_i^{bp} - a_i\right)^2\right)$.

Remembering that the best predictor, $\hat{a}_i^{bp} = E(a_i|data)$, is an unbiased predictor in the sense that $E(\hat{a}_i^{bp}) = E(a_i)$, then it follows that $Cov(a_i, \hat{a}_i^{bp}) = Var(\hat{a}_i^{bp})$ and that

$$PEV = Var(a_i) - Var\left(\hat{a}_i^{bp}\right).$$

Furthermore, the reliability of $\hat{a}_i^{bp}$, *i.e.* the squared correlation, $\rho^2(a_i, \hat{a}_i^{bp})$, is given by

$$\rho^2\left(a_i, \hat{a}_i^{bp}\right) = \frac{Var\left(\hat{a}_i^{bp}\right)}{Var(a_i)} = 1 - \frac{PEV}{Var(a_i)}.$$

Using the formula $Var(\hat{a}_i^{bp}) = Var(a_i) - E\big(Var(a_i|data)\big)$ (follows from $Var(a_i) = Var\big(E(a_i|data)\big) + E\big(Var(a_i|data)\big)$) and inserting in the expression for PEV, it follows that

$$PEV = E\big(Var(a_i|data)\big)$$

so that an unbiased estimate, PEV_{unbiased}, of PEV is given by

$$PEV_{\text{unbiased}} = Var(a_i|data)$$

(*i.e.* $E(PEV_{\text{unbiased}}) = PEV$) and an unbiased estimate, $\rho^2_{\text{unbiased}}(a_i, \hat{a}_i^{bp})$, of the squared correlation is given by

$$\rho^2_{\text{unbiased}}\left(a_i, \hat{a}_i^{bp}\right) = 1 - \frac{PEV_{\text{unbiased}}}{Var(a_i)}. \tag{6}$$

In both of Poisson mixed models and log normal frailty models for survival data we can find a vector $\mathbf{v} = (v_i)_{i=1,\ldots,N}$ with $(a_i, \mathbf{v})$ following a multivariate normal distribution and with the property that a_i and *data* are conditionally independent given $\mathbf{v}$. Therefore, in the following expression for PEV_{unbiased}

$$\begin{aligned} PEV_{\text{unbiased}} &= Var(a_i|data) \\ &= E_{\mathbf{v}|data}\big(Var(a_i|\mathbf{v}, data)\big) + Var_{\mathbf{v}|data}\big(E(a_i|\mathbf{v}, data)\big) \end{aligned}$$

the first term

$$E_{\mathbf{v}|data}\big(Var(a_i|\mathbf{v}, data)\big) = E_{\mathbf{v}|data}\big(Var(a_i|\mathbf{v})\big)$$

(because $p(a_i|\mathbf{v}, data) = p(a_i|\mathbf{v})$, which follows from the conditional independence of a_i and *data* given $\mathbf{v}$). And because $(a_i, \mathbf{v})$ follows a multivariate normal distribution then $Var(a_i|\mathbf{v})$ $(= \sigma_a^2 - Cov(a_i, \mathbf{v})\, Var(\mathbf{v})^{-1}\, Cov(\mathbf{v}, a_i))$ does not depend on $\mathbf{v}$ and therefore $E_{\mathbf{v}|data}\big(Var(a_i|\mathbf{v})\big) = Var(a_i|\mathbf{v})$. With regards to the second term:

$$E(a_i|\mathbf{v}, data) = E(a_i|\mathbf{v}) = Cov(a_i, \mathbf{v})\, Var(\mathbf{v})^{-1}\left(\mathbf{v} - \mathbf{E}(\mathbf{v})\right)$$

(because $p\left(a_i|\mathbf{v}, data\right) = p\left(a_i|\mathbf{v}\right)$ and $(a_i, \mathbf{v})$ follows a multivariate normal distribution). It follows that the second term

$$Var_{\mathbf{v}|data}\big(E\left(a_i|\mathbf{v}, data\right)\big) = Cov\left(a_i, \mathbf{v}\right) Var\left(\mathbf{v}\right)^{-1} Var\left(\mathbf{v}|data\right) Var\left(\mathbf{v}\right)^{-1} Cov\left(\mathbf{v}, a_i\right).$$

Finally we obtain the following expression for PEV_{unbiased}

$$\begin{aligned} PEV_{\text{unbiased}} &= \sigma_a^2 - Cov\left(a_i, \mathbf{v}\right) Var\left(\mathbf{v}\right)^{-1} Cov\left(\mathbf{v}, a_i\right) \\ &\quad + Cov\left(a_i, \mathbf{v}\right) Var\left(\mathbf{v}\right)^{-1} Var\left(\mathbf{v}|data\right) Var\left(\mathbf{v}\right)^{-1} Cov\left(\mathbf{v}, a_i\right) \\ &= \sigma_a^2 - Cov\left(a_i, \mathbf{v}\right) Var\left(\mathbf{v}\right)^{-1} \left[Var\left(\mathbf{v}\right) - Var\left(\mathbf{v}|data\right)\right] \\ &\quad \times Var\left(\mathbf{v}\right)^{-1} Cov\left(\mathbf{v}, a_i\right). \end{aligned} \tag{7}$$

Again, a further simplification can be obtained if the dimension of $\mathbf{v}$ is n, *i.e.* $N = n$, and $p\left(a_i|\mathbf{v}\right) = p\left(a_i|v_i\right)$, in which case the expression for PEV_{unbiased} simplifies to

$$PEV_{\text{unbiased}} = \sigma_a^2 - Cov\left(a_i, v_i\right) Var\left(v_i\right)^{-1} \left[Var\left(v_i\right) - Var\left(v_i|data\right)\right] Var\left(v_i\right)^{-1} Cov\left(v_i, a_i\right). \tag{8}$$

In the following, either expressions for PEV or PEV_{unbiased} will be given.

Accuracy of selection, $\rho\left(a_i, \hat{a}_i^{bp}\right)$, the correlation between a_i and $\hat{a}_i^{bp}$ is given by

$$\rho\left(a_i, \hat{a}_i^{bp}\right) = \sqrt{1 - \frac{PEV}{Var\left(a_i\right)}}.$$

If we approximate accuracy of selection by $\sqrt{\rho^2_{\text{unbiased}}(a_i, \hat{a}_i^{bp})}$, then we obtain an estimate, which approximately is an unbiased estimate for accuracy.

4.1. Linear mixed model

$$PEV = PEV_{\text{unbiased}} = \sigma_a^2\left(1 - h^2\right)$$

and

$$\rho\left(a_i, \hat{a}_i^{bp}\right) = h$$

see *e.g.* Bulmer [2].

4.2. Threshold model

$$Var\left(\hat{a}_i^{bp}\right) = h_{\text{nor}}^4 \sum_{k=1}^{K} \frac{\left(\varphi\left(\tau_{k-1} - x_i\beta\right) - \varphi\left(\tau_k - x_i\beta\right)\right)^2}{P\left(Y_i = k\right)}$$

$$= h_{\text{nor}}^2 \sigma_a^2 \sum_{k=1}^{K} \frac{\left(\varphi\left(\tau_{k-1} - x_i\beta\right) - \varphi\left(\tau_k - x_i\beta\right)\right)^2}{\Phi\left(\tau_k - x_i\beta\right) - \Phi\left(\tau_{k-1} - x_i\beta\right)}$$

$$PEV = \sigma_a^2 \left(1 - h_{\text{nor}}^2 \sum_{k=1}^{K} \frac{\left(\varphi\left(\tau_{k-1} - x_i\beta\right) - \varphi\left(\tau_k - x_i\beta\right)\right)^2}{\Phi\left(\tau_k - x_i\beta\right) - \Phi\left(\tau_{k-1} - x_i\beta\right)}\right)$$

so that

$$\rho^2\left(a_i, \hat{a}_i^{bp}\right) = h_{\text{nor}}^2 \sum_{k=1}^{K} \frac{\left(\varphi\left(\tau_{k-1} - x_i\beta\right) - \varphi\left(\tau_k - x_i\beta\right)\right)^2}{\Phi\left(\tau_k - x_i\beta\right) - \Phi\left(\tau_{k-1} - x_i\beta\right)}.$$

$$PEV_{\text{unbiased}} = Var\left(a_i|y_i\right)$$

$$= Var\left(a_i|u_i\right) + h_{\text{nor}}^4 Var\left(U_i|y_i\right)$$

$$= \sigma_a^2\left(1 - h_{\text{nor}}^2\right) + h_{\text{nor}}^4 Var\left(U_i|y_i\right)$$

and

$$\rho_{\text{unbiased}}^2\left(a_i, \hat{a}_i^{bp}\right) = h_{\text{nor}}^2\left(1 - h_{\text{nor}}^2 \frac{Var\left(U_i|y_i\right)}{\sigma_a^2}\right)$$

with

$$Var\left(U_i|Y_i = k\right) = \left(1 + \frac{b_{k-1}\varphi\left(b_{k-1}\right) - b_k\varphi\left(b_k\right)}{P\left(Y_i = k\right)} - \left(\frac{\varphi\left(b_k\right) - \varphi\left(b_{k-1}\right)}{P\left(Y_i = k\right)}\right)^2\right),$$

where $b_k = \tau_k - x_i\beta$, $k = 0, ...K$ with $\tau_0 = -\infty$, $\tau_1 = 0$, $\tau_K = \infty$ and $P\left(Y_i = k\right) = \Phi\left(b_k\right) - \Phi\left(b_{k-1}\right)$ for $k = 1, \ldots, K$.

4.3. Poisson mixed model

In the Poisson mixed model we may use (8) with $\mathbf{v} = \boldsymbol{\eta} = (\eta_i)_{i=1,\ldots,n}$, where η_i is given by (2). Furthermore, because $p\left(\eta_i|data\right) = p\left(\eta_i|y_i\right)$, then we obtain

$$PEV_{\text{unbiased}} = \sigma_a^2\left(1 - h_{\text{nor}}^2\right) + h_{\text{nor}}^4 Var\left(\eta_i|y_i\right).$$

It follows that

$$\rho_{\text{unbiased}}^2\left(a_i, \hat{a}_i^{bp}\right) = h_{\text{nor}}^2\left(1 - h_{\text{nor}}^2 \frac{Var\left(\eta_i|y_i\right)}{\sigma_a^2}\right).$$

4.4. Survival model

For survival traits we will use (7) with **v** given in Section 3.4 for calculating PEV_{unbiased}, thereafter $\rho^2_{\text{unbiased}}(a_i, \hat{a}_i^{bp})$ is found from (6).

4.5. Discussion and conclusion

Again, heritability, h^2_{nor}, or a generalised version of heritability, $Cov\,(a_i, \mathbf{v})\,Var\,(\mathbf{v})^{-1}$, plays a central role in the formulas for PEV or PEV_{unbiased}, and therefore also in formulas for reliability (or ρ^2_{unbiased}) and accuracy of selection (which are derived from formulas for PEV or PEV_{unbiased}).

Dempster and Lerner [4] and Robertson [14] gave a formula relating heritability on the observed scale, h^2_{obs}, with heritability on the underlying scale (liability scale) h^2_{nor} for binary threshold characters. In [4] and [14] h^2_{obs} has the interpretation of being reliability, $\rho^2(a_i, \hat{a}_i^{blp})$ of the best linear predictor of a_i, $\hat{a}_i^{blp}$. The work by [4] and [14] was generalised by Gianola [9], who gave a formula relating heritability on the observed scale, h^2_{obs}, with heritability on the liability scale, h^2_{nor}, for threshold characters with $K \geq 2$ categories. Also here h^2_{obs} has the interpretation of being reliability of the best linear predictor of a_i. The best predictor of a_i in the binary threshold model is linear (linear in y_i as we saw in Section 3.2) and therefore equal to the best linear predictor. It follows that reliability of $\hat{a}_i^{bp}$, $\rho^2(a_i, \hat{a}_i^{bp})$, in the binary threshold model is equal to h^2_{obs} found in [4, 14] and [9]. (Note that [4, 14] and [9] used another parameterisation of the threshold model). For threshold characters with $K \geq 3$ categories the best predictor of a_i is no longer linear in y_i and therefore expressions for reliability of the best predictor and the best linear predictor are different.

Foulley and Im [8], in the Poisson mixed model with $\eta_i = \log(\lambda_i) = x_i\beta + a_i$, presented a heritability in the narrow sense,

$$h^2_{\text{narrow}} = E\,(Y_i)^2\,\sigma_a^2 / \left(E\,(Y_i) + E\,(Y_i)^2\left[\exp\left\{\sigma_a^2\right\} - 1\right]\right),$$

which has the interpretation of being reliability, $\rho^2(a_i, \hat{a}_i^{blp})$, of the best linear predictor of a_i, $\hat{a}_i^{blp} = E\,(a_i) + Cov(a_i, Y_i)Var(Y_i)^{-1}\left(Y_i - E\,(Y_i)\right)$.

In order to calculate prediction error variance, PEV, then PEV_{unbiased} should be averaged over the distribution of data. For survival models, we have only partially specified the model of the data, *i.e.* we have only specified the distribution of survival times and not the joint distribution of survival and censoring times. This implies, that we are not able to calculate PEV for survival models, unless a joint distribution for survival and censoring times has been specified (or censoring is absent). In principle we are able to calculate PEV in the remaining models considered — however the calculations may involve integrals without closed form expressions and approximations are required.

For survival traits, a lot of different expressions for heritability have been presented (see *e.g.*[5,10,11] and [16]), some of these do have the interpretation of being reliability of a linear predictor of a random effect (a linear predictor based on survival data or transformed survival data) others are just ratios of variances and others are more difficult to interpret. Most of the heritabilities presented for survival traits have been derived for models without time-dependent covariates.

5. EXPECTED RESPONSE TO SELECTION

For any of the traits under consideration, we will assume that parents of the next generation will be chosen so that the best predictor of breeding values among fathers (mothers) is greater than (or equal to) t_1 (t_2) (or less than (or equal to) t_1 (t_2) for survival traits). Then the expected response to selection on the additive genetic scale, $\Re_a$, is defined by the expected additive genetic value of an offspring, given that the parents of the next generation are selected, and the selected parents are mated at random, minus, the expected additive genetic value obtained without selection (and under the assumption of random mating). Let F and M denote the sets of potential fathers and mothers, respectively, *i.e.* $F = \{i : i \text{ is a male}\}$ and $M = \{i : i \text{ is a female}\}$, then expected response to selection on the additive genetic scale, $\Re_a$, is

$$\Re_a = \sum_{A_1 : A_1 \subseteq F} \sum_{A_2 : A_2 \subseteq M} P(A_1 \times A_2) \times \left[\sum_{(\mathrm{f},\mathrm{m}) \in A_1 \times A_2} \frac{1}{|A_1|\,|A_2|} E\left(a_\mathrm{o} | \hat{a}_\mathrm{f}^{bp} \geq t_1, \hat{a}_\mathrm{m}^{bp} \geq t_2\right) \right] - E(a_\mathrm{o}) \quad (9)$$

where $P(A_1 \times A_2)$ is the probability that exactly those males in A_1 and exactly those females in A_2 are selected, that is, the probability that $\hat{a}_i^{bp} \geq t_1$ for all $i \in A_1$, and $\hat{a}_i^{bp} < t_1$ for all $i \in F \backslash A_1$, and $\hat{a}_i^{bp} \geq t_2$ for all $i \in A_2$ and $\hat{a}_i^{bp} < t_2$ for all $i \in M \backslash A_2$. Let $|A_1|$ ($|A_2|$) denote the number of elements in A_1 (A_2), then conditional on A_1 and A_2 being the sets of selected males and females respectively, the probability of a given mating (assuming random mating among selected parents) is $1/(|A_1|\,|A_2|)$. And $E\left(a_\mathrm{o} | \hat{a}_\mathrm{f}^{bp} \geq t_1, \hat{a}_\mathrm{m}^{bp} \geq t_2\right)$ is the expected value of a_o given that $(\mathrm{f}, \mathrm{m}) \in A_1 \times A_2$ are the parents (subscripts f, m and o are used for the father, mother and offspring). It follows that

$$\sum_{(\mathrm{f},\mathrm{m}) \in A_1 \times A_2} \frac{1}{|A_1|\,|A_2|} E\left(a_\mathrm{o} | \hat{a}_\mathrm{f}^{bp} \geq t_1, \hat{a}_\mathrm{m}^{bp} \geq t_2\right)$$

is the expected additive genetic value of an offspring, conditional on A_1 and A_2 being the sets of selected males and females, and under the assumption of random mating among selected animals.

If we let

$$R_a^{(\mathrm{f,m})} = E\left(a_\mathrm{o} | \hat{a}_\mathrm{f}^{bp} \geq t_1, \hat{a}_\mathrm{m}^{bp} \geq t_2\right) - E\left(a_\mathrm{o}\right) \tag{10}$$

denote the expected response to selection on the additive genetic scale, given that $(\mathrm{f,m}) \in A_1 \times A_2$ are the randomly chosen parents among the selected animals, then it is easily seen that (9) is equal to

$$\Re_a = \sum_{A_1 : A_1 \subseteq F} \sum_{A_2 : A_2 \subseteq M} P\left(A_1 \times A_2\right) \times \left[\sum_{(\mathrm{f,m}) \in A_1 \times A_2} \frac{1}{|A_1|\,|A_2|} R_a^{(\mathrm{f,m})} \right]. \tag{11}$$

If $\left(a_\mathrm{o}, \hat{a}_\mathrm{f}^{bp}, \hat{a}_\mathrm{m}^{bp}\right)$, for all $(\mathrm{f,m}) \in F \times M$, are identically distributed, then (11) simplifies to

$$\Re_a = R_a^{(\mathrm{f,m})}.$$

In general then $R_a^{(\mathrm{f,m})}$ may depend on covariates of both parents, and in general then other mating strategies among selected parents may result in a higher (or lower) expected response to selection on the additive genetic scale, compared to a random mating strategy among selected parents.

Expected response to selection on the phenotypic scale, $\Re_p^\mathrm{o}$, is defined similarly; *i.e.* $\Re_p^\mathrm{o}$ is the expected phenotypic value of an offspring to be raised in a given environment (given covariates of the offspring) given that parents of the next generation are selected, and selected parents are mated at random, minus the expected phenotypic value obtained without selection. *i.e.*

$$\Re_p^\mathrm{o} = \sum_{A_1 : A_1 \subseteq F} \sum_{A_2 : A_2 \subseteq M} P\left(A_1 \times A_2\right) \times \left[\sum_{(\mathrm{f,m}) \in A_1 \times A_2} \frac{1}{|A_1|\,|A_2|} E\left(Y_\mathrm{o} | \hat{a}_\mathrm{f}^{bp} \geq t_1, \hat{a}_\mathrm{m}^{bp} \geq t_2\right) \right] - E\left(Y_\mathrm{o}\right) \tag{12}$$

If we let

$$R_p^{(\mathrm{o}|\mathrm{f,m})} = E\left(Y_\mathrm{o} | \hat{a}_\mathrm{f}^{bp} \geq t_1, \hat{a}_\mathrm{m}^{bp} \geq t_2\right) - E\left(Y_\mathrm{o}\right) \tag{13}$$

denote the expected response to selection on the phenotypic scale, given that $(\mathrm{f,m}) \in A_1 \times A_2$ are the randomly chosen parents among the selected animals, then it is easily seen that (12) is equal to

$$\Re_p^\mathrm{o} = \sum_{A_1 : A_1 \subseteq F} \sum_{A_2 : A_2 \subseteq M} P\left(A_1 \times A_2\right) \times \left[\sum_{(\mathrm{f,m}) \in A_1 \times A_2} \frac{1}{|A_1|\,|A_2|} R_p^{(\mathrm{o}|\mathrm{f,m})} \right] \tag{14}$$

And again, if $\left(Y_{\mathrm{o}}, \hat{a}_{\mathrm{f}}^{bp}, \hat{a}_{\mathrm{m}}^{bp}\right)$, for all $(\mathrm{f}, \mathrm{m}) \in F \times M$, are identically distributed, then (14) simplifies to

$$\Re_p^{\mathrm{o}} = R_p^{(\mathrm{o}|\mathrm{f},\mathrm{m})}.$$

In general, then $R_p^{(\mathrm{o}|\mathrm{f},\mathrm{m})}$ may depend on covariates of the offspring, as well as on covariates of both parents, and in general then other mating strategies among selected parents may result in a higher (or lower) expected response to selection on the phenotypic scale, compared to a random mating strategy among selected parents.

If we want the expected response to selection on the phenotypic scale across all environments, then we must also know the number of offspring to be placed in the different environments.

In the following we give formulas for the expected response to selection on the additive genetic scale conditional on (f, m) being the randomly chosen parents among the selected animals, $R_a^{(\mathrm{f},\mathrm{m})}$, and for the expected response to selection on the phenotypic scale of an offspring to be raised in a given environment, and conditional on (f, m) being the randomly chosen parents among the selected animals, $R_p^{(\mathrm{o}|\mathrm{f},\mathrm{m})}$.

5.1. Linear mixed model

For Gaussian traits we have

$$\begin{aligned} R_p^{(\mathrm{o}|\mathrm{f},\mathrm{m})} &= E\left(x_{\mathrm{o}}\beta + a_{\mathrm{o}} + e_{\mathrm{o}} | \hat{a}_{\mathrm{f}}^{bp} \geq t_1, \hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) - E\left(x_{\mathrm{o}}\beta + a_{\mathrm{o}} + e_{\mathrm{o}}\right) \\ &= E\left(a_{\mathrm{o}} | \hat{a}_{\mathrm{f}}^{bp} \geq t_1, \hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) \\ &= R_a^{(\mathrm{f},\mathrm{m})} \end{aligned}$$

and it follows that

$$\Re_p^{\mathrm{o}} = \Re_a$$

i.e. the expected response to selection on the phenotypic scale is here equal to the expected response on the additive genetic scale. The joint distribution of $(a_{\mathrm{f}}, \hat{a}_{\mathrm{f}}^{bp})'$ is given by:

$$\begin{pmatrix} a_{\mathrm{f}} \\ \hat{a}_{\mathrm{f}}^{bp} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & h^2\sigma_a^2 \\ h^2\sigma_a^2 & h^4 Var(Y_{\mathrm{f}}) \end{pmatrix}\right)$$

so that $a_{\mathrm{f}} | \hat{a}_{\mathrm{f}}^{bp} \sim N\left(\hat{a}_{\mathrm{f}}^{bp}, \sigma_a^2\left(1 - h^2\right)\right)$ (and similarly for the distribution of a_{m} given $\hat{a}_{\mathrm{m}}^{bp}$: $a_{\mathrm{m}} | \hat{a}_{\mathrm{m}}^{bp} \sim N\left(\hat{a}_{\mathrm{m}}^{bp}, \sigma_a^2\left(1 - h^2\right)\right)$), therefore it follows that

$R_p^{(\mathrm{o}|\mathrm{f,m})} = R_a^{(\mathrm{f,m})}$ is given by:

$$\begin{aligned} E\left(a_{\mathrm{o}}|\hat{a}_{\mathrm{f}}^{bp} > t_1, \hat{a}_{\mathrm{m}}^{bp} > t_2\right) &= \frac{1}{2}E\left(a_{\mathrm{f}}|\hat{a}_{\mathrm{f}}^{bp} \geq t_1\right) + \frac{1}{2}E\left(a_{\mathrm{m}}|\hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) \\ &= \frac{1}{2}E\left(\hat{a}_{\mathrm{f}}^{bp}|\hat{a}_{\mathrm{f}}^{bp} \geq t_1\right) + \frac{1}{2}E\left(\hat{a}_{\mathrm{m}}^{bp}|\hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) \\ &= \frac{1}{2}h^2 S_{\mathrm{f}} + \frac{1}{2}h^2 S_{\mathrm{m}} \\ &= \frac{1}{2}h^2\sigma_p i_{\mathrm{f}} + \frac{1}{2}h^2\sigma_p i_{\mathrm{m}} \\ &= \frac{1}{2}h\sigma_a i_{\mathrm{f}} + \frac{1}{2}h\sigma_a i_{\mathrm{m}} \\ &= \frac{1}{2}\rho\left(\hat{a}_{\mathrm{f}}^{bp}, a_{\mathrm{f}}\right)\sigma_a i_{\mathrm{f}} + \frac{1}{2}\rho\left(\hat{a}_{\mathrm{m}}^{bp}, a_{\mathrm{m}}\right)\sigma_a i_{\mathrm{m}} \end{aligned}$$

where $h^2 = \sigma_a^2/\sigma_p^2$, with $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$, and the expected selection differential in fathers, S_{f}, is given by

$$S_{\mathrm{f}} = E\left(Y_{\mathrm{f}} - E\left(Y_{\mathrm{f}}\right)|\hat{a}_{\mathrm{f}}^{bp} \geq t_1\right) = \sigma_p \frac{\varphi\left(\dfrac{t_1}{h^2\sigma_p}\right)}{P\left(\dfrac{Y_{\mathrm{f}} - E\left(Y_{\mathrm{f}}\right)}{\sigma_p} \geq \dfrac{t_1}{h^2\sigma_p}\right)}$$

The intensity of selection in fathers i_{f}, is defined as S_{f}/σ_p, *i.e.* the expected selection differential expressed in phenotypic standard deviations and is here given by $i_{\mathrm{f}} = \varphi\left(\frac{t_1}{h^2\sigma_p}\right)\Big/P\left(\frac{Y_{\mathrm{f}} - E(Y_{\mathrm{f}})}{\sigma_p} \geq \frac{t_1}{h^2\sigma_p}\right)$. The expected selection differential and intensity of selection in mothers, S_{m} and i_{m}, are defined similarly. Note that the accuracy of selection, $\rho(\hat{a}_{\mathrm{f}}^{bp}, a_{\mathrm{f}}) = h$ in this context.

Furthermore, for Gaussian traits, we have

$$R_p^{(\mathrm{o}|\mathrm{f,m})} = R_a^{(\mathrm{f,m})} = \Re_a = \Re_p^{\mathrm{o}}.$$

5.2. Threshold model

Let $\hat{a}_i^{bp}(k)$ denote the best predictor of a_i, conditional on $Y_i = k$, *i.e.*

$$\hat{a}_i^{bp}(k) = h_{\mathrm{nor}}^2\left(E\left(U_i|Y_i = k\right) - x_i\beta\right) = h_{\mathrm{nor}}^2 \frac{\varphi\left(\tau_{k-1} - x_i\beta\right) - \varphi\left(\tau_k - x_i\beta\right)}{P\left(Y_i = k\right)}$$

for $k = 1, \ldots, K$. It is easy to see that $\hat{a}_i^{bp}(1) < \hat{a}_i^{bp}(2) < \cdots < \hat{a}_i^{bp}(K)$ for $i = 1, \ldots, n$. For $t_1 > \hat{a}_{\mathrm{f}}^{bp}(K)$ or $t_2 > \hat{a}_{\mathrm{m}}^{bp}(K)$, then the pair (f, m)

will never be selected as parents of a future offspring; *i.e.* (f, m) will never belong to a $A_1 \times A_2$ with $P(A_1 \times A_2) > 0$. Let $\hat{a}_{\mathrm{f}}^{bp}(0) = \hat{a}_{\mathrm{m}}^{bp}(0) = -\infty$, then for $\hat{a}_{\mathrm{f}}^{bp}(k_1 - 1) < t_1 \leq \hat{a}_{\mathrm{f}}^{bp}(k_1)$ and $\hat{a}_{\mathrm{m}}^{bp}(k_2 - 1) < t_2 \leq \hat{a}_{\mathrm{m}}^{bp}(k_2)$ with $k_1, k_2 = 1, \ldots, K$, the event $\left\{\hat{a}_{\mathrm{f}}^{bp} \geq t_1, \hat{a}_{\mathrm{m}}^{bp} \geq t_2\right\}$ is equivalent to the event $\left\{Y_{\mathrm{f}} \in \{k_1, \ldots, K\}, Y_{\mathrm{m}} \in \{k_2, \ldots, K\}\right\}$. This case corresponds to a situation with possible selection on males if $t_1 > \hat{a}_{\mathrm{f}}^{bp}(1)$ for some $\mathrm{f} \in F$ (and possible selection on females if $t_2 > \hat{a}_{\mathrm{m}}^{bp}(1)$ for some $\mathrm{m} \in M$). It follows that

$$
\begin{aligned}
R_a^{(\mathrm{f},\mathrm{m})} &= \int a_{\mathrm{o}} p\left(a_{\mathrm{o}} | Y_{\mathrm{f}} \in \{k_1, \ldots, K\}, Y_{\mathrm{m}} \in \{k_2, \ldots, K\}\right) \mathrm{d}a_{\mathrm{o}} \\
&= \iiint a_{\mathrm{o}} p\left(a_{\mathrm{o}}, u_{\mathrm{f}}, u_{\mathrm{m}} | U_{\mathrm{f}} > \tau_{k_1-1}, U_{\mathrm{m}} > \tau_{k_2-1}\right) \mathrm{d}a_{\mathrm{o}} \mathrm{d}u_{\mathrm{f}} \mathrm{d}u_{\mathrm{m}} \\
&= \iint E\left(a_{\mathrm{o}} | u_{\mathrm{f}}, u_{\mathrm{m}}\right) p\left(u_{\mathrm{f}}, u_{\mathrm{m}} | U_{\mathrm{f}} > \tau_{k_1-1}, U_{\mathrm{m}} > \tau_{k_2-1}\right) \mathrm{d}u_{\mathrm{f}} \mathrm{d}u_{\mathrm{m}} \\
&= \iint \frac{1}{2} h_{\mathrm{nor}}^2 \big((u_{\mathrm{f}} - x_{\mathrm{f}}\beta) + (u_{\mathrm{m}} - x_{\mathrm{m}}\beta)\big) \\
&\quad \times p\left(u_{\mathrm{f}} | U_{\mathrm{f}} > \tau_{k_1-1}\right) p\left(u_{\mathrm{m}} | U_{\mathrm{m}} > \tau_{k_2-1}\right) \mathrm{d}u_{\mathrm{f}} \mathrm{d}u_{\mathrm{m}} \\
&= \frac{1}{2} h_{\mathrm{nor}}^2 \left[\frac{\varphi\left(\tau_{k_1-1} - x_{\mathrm{f}}\beta\right)}{P\left(U_{\mathrm{f}} > \tau_{k_1-1}\right)} + \frac{\varphi\left(\tau_{k_2-1} - x_{\mathrm{m}}\beta\right)}{P\left(U_{\mathrm{m}} > \tau_{k_2-1}\right)}\right] \\
&= \frac{1}{2} \sigma_a h_{\mathrm{nor}} i_{\mathrm{f}}^{\mathrm{nor}} + \frac{1}{2} \sigma_a h_{\mathrm{nor}} i_{\mathrm{m}}^{\mathrm{nor}} \\
&= \frac{1}{2} h_{\mathrm{nor}}^2 S_{\mathrm{f}}^{\mathrm{nor}} + \frac{1}{2} h_{\mathrm{nor}}^2 S_{\mathrm{m}}^{\mathrm{nor}}
\end{aligned}
$$

where $S_{\mathrm{f}}^{\mathrm{nor}}$ is defined as the expected selection differential on the liability scale obtained by selection on the best predictor, *i.e.*

$$
S_{\mathrm{f}}^{\mathrm{nor}} = E\left(U_{\mathrm{f}} - E\left(U_{\mathrm{f}}\right) | \hat{a}_{\mathrm{f}}^{bp} \geq t_1\right) = \frac{\varphi\left(\tau_{k_1-1} - x_{\mathrm{f}}\beta\right)}{P\left(U_{\mathrm{f}} > \tau_{k_1-1}\right)}
$$

and $i_{\mathrm{f}}^{\mathrm{nor}}$ is defined by $S_{\mathrm{f}}^{\mathrm{nor}}$ divided by σ_u. For categorical threshold characters we have assumed that $\sigma_u^2 = 1$ (for reasons of identifiability), therefore $i_{\mathrm{f}}^{\mathrm{nor}} = S_{\mathrm{f}}^{\mathrm{nor}}$. $S_{\mathrm{m}}^{\mathrm{nor}}$ and $i_{\mathrm{m}}^{\mathrm{nor}}$ are defined similarly. Note, if $t_1 \leq \hat{a}_{\mathrm{f}}^{bp}(1)$ ($t_2 \leq \hat{a}_{\mathrm{m}}^{bp}(1)$) then $S_{\mathrm{f}}^{\mathrm{nor}} = 0$ ($S_{\mathrm{m}}^{\mathrm{nor}} = 0$).

The expected response to selection on the phenotypic scale given that $(\mathrm{f}, \mathrm{m}) \in A_1 \times A_2$ are the randomly chosen parents among the selected animals, is

$$
R_p^{(\mathrm{o}|\mathrm{f},\mathrm{m})} = E\left(Y_{\mathrm{o}} | U_{\mathrm{f}} > \tau_{k_1-1}, U_{\mathrm{m}} > \tau_{k_2-1}\right) - E\left(Y_{\mathrm{o}}\right)
$$

for $k_1, k_2 = 1, \ldots, K$, where $E(Y_o) = \sum_{k=1}^{K} kP(Y_o = k)$ and

$$E\left(Y_o | U_f > \tau_{k_1-1}, U_m > \tau_{k_2-1}\right) = \frac{1}{P\left(U_f > \tau_{k_1-1}, U_m > \tau_{k_2-1}\right)} \sum_{k=1}^{K} kP\left(Y_o = k, U_f > \tau_{k_1-1}, U_m > \tau_{k_2-1}\right)$$

with

$$\begin{aligned}
&P\left(Y_o = k, U_f > \tau_{k_1-1}, U_m > \tau_{k_2-1}\right) \\
&= \iiint P\left(\tau_{k-1} < U_o \leq \tau_k, U_f > \tau_{k_1-1}, U_m > \tau_{k_2-1} | a_o, a_f, a_m\right) \\
&\quad \times p(a_o, a_f, a_m)\, \mathrm{d}a_o \mathrm{d}a_f \mathrm{d}a_m \\
&= \iiint P(\tau_{k-1} < U_o \leq \tau_k | a_o) P\left(U_f > \tau_{k_1-1} | a_f\right) P\left(U_m > \tau_{k_2-1} | a_m\right) \\
&\quad \times p(a_o | a_f, a_m) p(a_f, a_m)\, \mathrm{d}a_o \mathrm{d}a_f \mathrm{d}a_m \\
&= E_{(a_f, a_m)} \left[\left[E_{a_o | (a_f, a_m)} \left(\Phi\left(\frac{\tau_k - x_o\beta - a_o}{\sigma_e} \right) - \Phi\left(\frac{\tau_{k-1} - x_o\beta - a_o}{\sigma_e} \right) \right) \right] \right. \\
&\quad \left. \left(1 - \Phi\left(\frac{\tau_{k_1-1} - x_f\beta - a_f}{\sigma_e} \right) \right) \left(1 - \Phi\left(\frac{\tau_{k_2-1} - x_m\beta - a_m}{\sigma_e} \right) \right) \right] \\
&= E_{(a_f, a_m)} \left[\left[\Phi\left(\frac{\tau_k - x_o\beta - \frac{1}{2}(a_f + a_m)}{\sqrt{\frac{1}{2}\sigma_a^2 + \sigma_e^2}} \right) \right. \right. \\
&\quad \left. - \Phi\left(\frac{\tau_{k-1} - x_o\beta - \frac{1}{2}(a_f + a_m)}{\sqrt{\frac{1}{2}\sigma_a^2 + \sigma_e^2}} \right) \right] \\
&\quad \left. \left(1 - \Phi\left(\frac{\tau_{k_1-1} - x_f\beta - a_f}{\sigma_e} \right) \right) \left(1 - \Phi\left(\frac{\tau_{k_2-1} - x_m\beta - a_m}{\sigma_e} \right) \right) \right]
\end{aligned}$$

where (in obtaining the last equality) we use the formula (from Curnow [3])

$$\int_{-\infty}^{\infty} \varphi(x)\, \Phi(a + bx)\, \mathrm{d}x = \Phi\left(\frac{a}{\sqrt{1 + b^2}} \right)$$

for $a, b \in R$.

Example 2. Consider the trait "diseased within the first month of life" and assume that a binary threshold model can be used for analysing data. Diseased is coded 0, and not diseased is coded 1. To avoid complications we assume a situation where all animals are observed and alive during the first month of life. Assuming that the base population, which we are going to select from, is in two different environments (herds), say 500 males and 500 females in each herd. The model is given by (1) (except that observable values are 0 and 1, instead of 1 and 2) with $h^2_{\text{nor}} = 0.2$, and with $x_i\beta$ for animals in herd 1 (herd 2) determined so that the probability of being diseased is 0.98 (0.5); *i.e.* $x_i\beta \approx -2.054$ ($x_i\beta = 0$) for animals in herd 1 (herd 2). The best predictor of a_i for "not diseased" (diseased) animals in herd 1 is 0.48 (-0.0099). The best predictor of a_i for "not diseased" (diseased) animals in herd 2 is 0.16 (-0.16). For animals in herd 1 (herd 2) accuracy of selection is 0.15 (0.36). Next, selecting all "not diseased" animals (*i.e.* all animals with a best predictor greater than or equal to 0), then

$$R_a^{(\text{f,m})} = \begin{cases} 0.48 & \text{if both of f and m are from herd 1} \\ 0.32 & \text{if f and m are from different herds} \\ 0.16 & \text{if both of f and m are from herd 2} \end{cases}$$

and

$$R_p^{(\text{o}|\text{f,m})} = \begin{cases} 0.187 & \text{if both of f and m are from herd 1, and the offspring, o,} \\ & \text{is going to be raised in herd 2} \\ 0.127 & \text{if f and m are from different herds, and the offspring, o,} \\ & \text{is going to be raised in herd 2} \\ 0.064 & \text{if both of f and m are from herd 2, and the offspring, o,} \\ & \text{is going to be raised in herd 2} \\ 0.037 & \text{if both of f and m are from herd 1, and the offspring, o,} \\ & \text{is going to be raised in herd 1} \\ 0.020 & \text{if f and m are from different herds, and the offspring, o,} \\ & \text{is going to be raised in herd 1} \\ 0.008 & \text{if both of f and m are from herd 2, and the offspring, o,} \\ & \text{is going to be raised in herd 1} \end{cases}$$

In this example we observe that the highest expected response to selection on the additive genetic scale, given that both parents are selected, $R_a^{(\text{f,m})}$, is obtained when both parents are from herd 1. Similarly, the highest expected response to selection on the phenotypic scale, given that both of the parents are selected, and given covariates of the offspring, $R_p^{(\text{o}|\text{f,m})}$, is obtained when both parents are from herd 1.

5.3. Poisson mixed model

$$R_a^{(\mathrm{f,m})} = \frac{1}{2}h^2_{\mathrm{nor}}E\left(\eta_{\mathrm{f}} - E\left(\eta_{\mathrm{f}}\right)|\hat{a}_{\mathrm{f}}^{bp} \geq t_1\right) + \frac{1}{2}h^2_{\mathrm{nor}}E\left(\eta_{\mathrm{m}} - E\left(\eta_{\mathrm{m}}\right)|\hat{a}_{\mathrm{m}}^{bp} \geq t_2\right)$$
$$= \frac{1}{2}h^2_{\mathrm{nor}}S^{\mathrm{nor}}_{\mathrm{f}} + \frac{1}{2}h^2_{\mathrm{nor}}S^{\mathrm{nor}}_{\mathrm{m}}$$

where $S^{\mathrm{nor}}_{\mathrm{f}} = E\left(\eta_{\mathrm{f}} - E\left(\eta_{\mathrm{f}}\right)|\hat{a}_{\mathrm{f}}^{bp} \geq t_1\right)$ ($S^{\mathrm{nor}}_{\mathrm{m}}$ is defined similarly).

$$\begin{aligned}
R_p^{(\mathrm{o|f,m})} &= E\left(\exp\{\eta_{\mathrm{o}}\}|\hat{a}_{\mathrm{f}}^{bp} \geq t_1, \hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) - E\left(\exp\{\eta_{\mathrm{o}}\}\right) \\
&= E\left(\exp\{x_{\mathrm{o}}\beta + a_{\mathrm{o}} + e_{\mathrm{o}}\}|\hat{a}_{\mathrm{f}}^{bp} \geq t_1, \hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) \\
&\quad - E\left(\exp\{x_{\mathrm{o}}\beta + a_{\mathrm{o}} + e_{\mathrm{o}}\}\right) \\
&= \exp\left\{x_{\mathrm{o}}\beta + \frac{1}{2}\sigma_e^2\right\} \\
&\quad \times \left[E\left(\exp\{a_{\mathrm{o}}\}|\hat{a}_{\mathrm{f}}^{bp} \geq t_1, \hat{a}_{\mathrm{m}}^{bp} \geq t_2\right) - \exp\left\{\frac{1}{2}\sigma_a^2\right\}\right].
\end{aligned}$$

5.4. Survival model

With **v** as described in Section 3.4 then

$$\begin{aligned}
R_a^{(\mathrm{f,m})} &= E\left(a_{\mathrm{o}}|\hat{a}_{\mathrm{f}}^{bp} \leq t_1, \hat{a}_{\mathrm{m}}^{bp} \leq t_2\right) - E\left(a_{\mathrm{o}}\right) \\
&= \iint a_{\mathrm{o}} p\left(a_{\mathrm{o}}, \mathbf{v}|\hat{a}_{\mathrm{f}}^{bp} \leq t_1, \hat{a}_{\mathrm{m}}^{bp} \leq t_2\right) \mathrm{d}a_{\mathrm{o}}\mathrm{d}\mathbf{v} \\
&= \int E\left(a_{\mathrm{o}}|\mathbf{v}\right) p\left(\mathbf{v}|\hat{a}_{\mathrm{f}}^{bp} \leq t_1, \hat{a}_{\mathrm{m}}^{bp} \leq t_2\right) \mathrm{d}\mathbf{v} \\
&= Cov\left(a_{\mathrm{o}}, \mathbf{v}\right)\left[Var\left(\mathbf{v}\right)\right]^{-1} E\left(\mathbf{v}|\hat{a}_{\mathrm{f}}^{bp} \leq t_1, \hat{a}_{\mathrm{m}}^{bp} \leq t_2\right).
\end{aligned}$$

Next assume that $\Lambda_0(\cdot)$ and β_2 are known and let $h_i^{\mathbf{u}_2}(t)$ be as described in Section 2.4, then (as we have seen) the model is, conditional on $\mathbf{u}_2$, a linear model for $\tilde{Y}_i = \log\left(h_i^{\mathbf{u}_2}(T_i)\right)$. Then the expected response (given that (f, m) are the selected parents) on this (linear) $\log h_i^{\mathbf{u}_2}(\cdot)$-scale is equal to minus the expected response (given that (f, m) are the selected parents) obtained on the additive genetic scale, *i.e.*

$$R^{(\mathrm{o|f,m})}_{\log\left(h_{\mathrm{o}}^{\mathbf{u}_2}(\cdot)\right)} = -R_a^{(\mathrm{f,m})}.$$

If we want the expected response to selection on the untransformed time scale, then we proceed as follows: Let $g_{\mathrm{o}}^{\mathbf{u}_2}$, still conditional on $\mathbf{u}_2$, denote an inverse

function of $\log h_{\mathrm{o}}^{\mathbf{u}_2}$ (as specified in Sect. 2.4), then

$$T_{\mathrm{o}} = g_{\mathrm{o}}^{\mathbf{u}_2}(\tilde{Y}_{\mathrm{o}}) = g_{\mathrm{o}}^{\mathbf{u}_2}\left(-x_{o1}\beta_1 - a_{\mathrm{o}} - e_{\mathrm{o}} + \varepsilon_{\mathrm{o}}\right).$$

Using a first order Taylor series expansion of $g_{\mathrm{o}}^{\mathbf{u}_2}(\tilde{Y}_{\mathrm{o}})$ around the mean of $\tilde{Y}_{\mathrm{o}}$ ($E(\tilde{Y}_{\mathrm{o}}) = -x_{\mathrm{o}}\beta - \gamma_{\mathrm{E}}$) then we obtain

$$T_0 \approx g_{\mathrm{o}}^{\mathbf{u}_2}\left(E(\tilde{Y}_{\mathrm{o}})\right) + g_{\mathrm{o}}^{\mathbf{u}_2(1)}\left(E(\tilde{Y}_{\mathrm{o}})\right)\left(\tilde{Y}_{\mathrm{o}} - E(\tilde{Y}_{\mathrm{o}})\right).$$

It follows that the expected response (given that (f, m) are the selected parents) on the time scale, $R_T^{(\mathrm{o}|\mathrm{f,m})}$, can be approximated by

$$\begin{aligned} R_T^{(\mathrm{o}|\mathrm{f,m})} &= E\left(T_{\mathrm{o}}|\hat{a}_{\mathrm{f}}^{bp} \le t_1, \hat{a}_{\mathrm{m}}^{bp} \le t_2\right) - E\left(T_{\mathrm{o}}\right) \\ &\approx g_{\mathrm{o}}^{\mathbf{u}_2(1)}\left(E(\tilde{Y}_{\mathrm{o}})\right) R_{\log\left(h_0^{\mathbf{u}_2}(\cdot)\right)}^{(\mathrm{o}|\mathrm{f,m})} \\ &= -g_{\mathrm{o}}^{\mathbf{u}_2(1)}\left(E(\tilde{Y}_{\mathrm{o}})\right) R_a^{(\mathrm{f,m})}. \end{aligned}$$

As pointed out by a reviewer, this formula should be used cautiously, because it is based on a Taylor series expansion of $g_{\mathrm{o}}^{\mathbf{u}_2}(\tilde{Y}_{\mathrm{o}})$ around the mean of $\tilde{Y}_{\mathrm{o}}$, $E(\tilde{Y}_{\mathrm{o}})$. For non-linear functions the Taylor series expansion generally only works well if $\tilde{Y}_{\mathrm{o}}$ is close to $E(\tilde{Y}_{\mathrm{o}})$ - and this is not generally true.

Example 3. In the Weibull frailty model without time-dependent covariates (with associated fixed or random effects), the formulas are even simpler: Let $\tilde{Y}_i = \log(T_i) = -\log(\gamma) - \frac{1}{\alpha}x_i\beta - \frac{1}{\alpha}\eta_i + \frac{1}{\alpha}\varepsilon_i$, with $\eta_i = u_{1l(i)} + a_i + e_i$. It follows that the expected response to selection (given that (f, m) are the selected parents) on the (linear) log time scale is given by

$$R_{\log(\cdot)}^{(\mathrm{o}|\mathrm{f,m})} = -\frac{1}{\alpha}R_a^{(\mathrm{f,m})}.$$

If we want the expected response to selection (given that (f, m) are the selected parents) on the untransformed time scale, then we obtain, using a first order Taylor series expansion of $T_{\mathrm{o}} = \exp(\tilde{Y}_{\mathrm{o}})$ around the mean of $\tilde{Y}_{\mathrm{o}}$ ($E(\tilde{Y}_{\mathrm{o}}) = -\log(\gamma) - \frac{1}{\alpha}x_{\mathrm{o}}\beta - \frac{1}{\alpha}\gamma_{\mathrm{E}}$), that the expected response to selection on the time scale (given that (f, m) are the selected parents) can be approximated by

$$R_T^{(\mathrm{o}|\mathrm{f,m})} \approx -\exp\left\{-\log(\gamma) - \frac{1}{\alpha}x_{\mathrm{o}}\beta - \frac{1}{\alpha}\gamma_{\mathrm{E}}\right\}\frac{1}{\alpha}R_a^{(\mathrm{f,m})}.$$

5.5. Discussion and conclusion

Again heritability (or a generalised version of heritability) is seen to play a central role in the formulas for the expected response to selection.

For Gaussian traits, then the joint distribution of $(a_i, \hat{a}_i^{bp})$ is bivariate normal, this is not the case for any of the other traits studied. Anyhow this assumption has been used (and noted to be critical) in Foulley (1992) and Foulley (1993) for the calculation of response to selection for threshold dichotomous traits and for traits following a Poisson animal mixed models (without a normally distributed error term included), respectively.

For survival traits, note that in order to calculate the expected response to selection, $\Re_a$ in (9) (or $\Re_p^{\mathrm{o}}$ (12)) requires that we either know the joint distribution for survival and censoring times, or censoring is absent.

6. CONCLUSION

All of the models considered are mixed models, where the mixture distribution is the normal distribution. We have observations on the normally distributed scale only in Gaussian mixed linear models. For ordered categorical traits using a threshold model, the observed value is uniquely determined by a grouping on the normally distributed liability scale. In Poisson mixed models we have, conditional on the outcome of the normally distributed random vector, observations from a Poisson distribution. In survival models, and conditional on random effects, then $\log\big(\Lambda_i\,(T_i|\textit{random effects})\big)$ follows an extreme value distribution with mean $-\gamma_{\mathrm{E}}$ and variance $\pi^2/6$.

We have considered selection based on the best predictor of animal additive genetic values. For each trait and based on a single record per animal we have given expressions for the best predictor of breeding values of potential parents (best in the sense that it has minimum mean square error of prediction (PEV), and is the predictor of a_i with the highest correlation to a_i). Furthermore we have given expressions for PEV and/or an unbiased estimate for PEV. We have chosen to select those males (females) with the observed value of the best predictor greater than (or equal to) t_1 (t_2) (or less than (or equal to) t_1 (t_2) for survival traits). Based on this selection criterion we considered the expected response to selection that can be obtained on the additive genetic and the phenotypic scale. Expected response to selection on the additive genetic scale, $\Re_a$, was defined by the expected additive genetic value of an offspring, given that parents of the next generation are selected, and selected parents are mated at random, minus, the expected additive genetic value obtained without selection (and under the assumption of random mating). Expected response to selection on the phenotypic scale, $\Re_p^{\mathrm{o}}$, of an offspring, o, to be raised in a

given environment (given covariates of the offspring) was defined similarly. Note that in general the expected response to selection on the phenotypic scale will depend on covariates of the offspring (in the linear mixed model, this is not the case). In defining the expected response to selection (on both of the additive genetic and the phenotypic scale) note that we have chosen a random mating strategy among selected parents as well as a random mating strategy when there is no selection. Another selection criterion, as well as other mating strategies among selected and/or unselected animals may give other results.

In conclusion, for Gaussian linear mixed models, heritability defined as the ratio between the additive genetic variance and the phenotypic variance plays a central role in formulas for the best predictor, accuracy, reliability, and expected response to selection. Similarly does h^2_{nor}, the ratio between the additive genetic variance and the total variance at the normally distributed level of the model (or a generalised version of heritability, $Cov\,(a_{\text{o}}, \mathbf{v})\,[Var\,(\mathbf{v})]^{-1}$), in all of the other models considered.

Having obtained expressions for the best predictor and related quantities in animal models, then it is relatively easy to generalise and find expressions, in a progeny testing scheme for example. Progeny testing for all-or-none traits was considered by Curnow [3]. In most of the literature for binary traits the mean on the liability scale has been assumed to be the same for all animals. Here, we considered formulas allowing for a more general mean structure.

In this paper we have assumed that all parameters are known. If the parameters are unknown they should be estimated, and for that purpose it is important to ensure the identifiability of the parameters. For all of the models considered in this paper, the theorems concerning identifiability of parameters are given in Andersen *et al.* [1].

In the linear mixed model the best predictor is linear, *i.e.* the best predictor equals the best linear predictor. If the variance components are known, but fixed effects are unknown, then most often BLUP-values for breeding values are presented. These are the expressions for the BLP (equal to the BP in the linear mixed model) with fixed effects substituted by their generalised least square estimates (see *e.g.* [15]). If variance components are unknown as well as fixed effects then "BLUP"-values are presented with estimated variance components inserted for true values. Variance components are often estimated using REML (see [13]). For models other than the linear mixed model the best predictor of breeding values is not necessarily linear and properties of the BP, when estimated values are inserted for true parameter values, are unknown, and will depend on the method of estimation. This topic needs further research.

REFERENCES

[1] Andersen A.H., Korsgaard I.R., Jensen J., Identifiability of parameters in - and equivalence of animal and sire models for Gaussian and threshold characters, traits following a Poisson mixed model, and survival traits, Research Report No. 417 (2000) Department of Theoretical Statistics, University of Aarhus, Denmark.

[2] Bulmer M.G., The Mathematical Theory of Quantitative Genetics, Oxford University Press, 1980.

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

[4] Dempster E.R., Lerner I.M., Heritability of threshold characters, Genetics 35 (1950) 212–236.

[5] Ducrocq V., Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters, in: Proceedings of International workshop on EU concerted action Genetic Improvement of Functional Traits in Cattle (GIFT), Longevity, Interbull Bulletin 21 (1999) 60–67.

[6] Foulley J.L., Prediction of selection response for threshold dichotomous traits, Genetics 132 (1992) 1187–1194.

[7] Foulley J.L., Prediction of selection response for Poisson distributed traits, Genet. Sel. Evol. 25 (1993) 297–303.

[8] Foulley J.L., Im S., A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models, Genet. Sel. Evol. 25 (1993) 101–107.

[9] Gianola D., Heritability of polychotomous characters, Genetics 93 (1979) 1051–1055.

[10] Korsgaard I.R., Andersen A.H., Jensen J., Discussion of heritability of survival traits, in: Proceedings of International workshop on EU concerted action Genetic Improvement of Functional Traits in Cattle (GIFT), Longevity, Interbull Bulletin 21 (1999) 31–35.

[11] Korsgaard I.R., Andersen A.H., Jensen J., On different models, on heritability, reliability and related quantities of survival traits, Book of Abstracts of the 51st Annual Meeting of the European Association for Animal Production, Wageningen Pers, Wageningen, The Netherlands, 2000, page 80.

[12] Searle S.R., Casella G., McCulloch C.E., Variance Components, John Wiley & Sons, 1992.

[13] Patterson H.D., Thompson R., Recovery of inter-block information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[14] Robertson A., Proof that the additive heritability on the P scale is given by the expression $\bar{z}^2 h_x^2/\bar{p}\bar{q}$, Genetics 35 (1950) 234–236.

[15] Robinson G.K., That BLUP is a good thing: The estimation of random effects, Stat. Sci. 6 (1991) 15–51.

[16] Yazdi M.H., Thompson R., Ducrocq V., Visscher P.M., Genetic parameters and response to selection in proportional hazard models, Book of Abstracts of the 51st Annual Meeting of the European Association for Animal Production,Wageningen Pers, Wageningen, The Netherlands, 2000, page 81.

APPENDIX

Let T_i denote the random variable representing survival time of animal i. In the frailty model without time-dependent covariates (with associated fixed or random effects), it is assumed that conditional on the vector of log frailties, $\mathbf{W} = \mathbf{w}$, the hazard function of animal i, is given by

$$\lambda_i(t|\mathbf{w}) = \lambda_0(t) \exp\{x_i\beta + w_i\} \tag{A.1}$$

where $\lambda_0(t)$ is a common baseline hazard function, x_i is a vector of time-independent covariates of animal i, and β is the corresponding vector of regression parameters. Furthermore, conditional on $(W_i)_{i=1,\ldots,n}$, then all of the $T_i's$ are assumed to be independent. In the model specified by (A.1), the conditional integrated hazard function is

$$\Lambda_i(t|\mathbf{w}) = \Lambda_0(t) \exp\{x_i\beta + w_i\}$$

and the conditional survival function is

$$S_i(t|\mathbf{w}) = \exp\{-\Lambda_i(t|\mathbf{w})\}.$$

Because $S_i(T_i|\mathbf{w}) = \exp\{-\Lambda_i(T_i|\mathbf{w})\}$ is uniformly distributed on the interval $(0; 1)$, the transformed random variable, $Y_i = \Lambda_i(T_i|\mathbf{W})$, conditional on $\mathbf{W} = \mathbf{w}$, is exponentially distributed with parameter 1. In turn, ε_i, the logarithm of Y_i, given by

$$\varepsilon_i = \log(Y_i) = \log\big(\Lambda_i(T_i|\mathbf{W})\big) = \log\big(\Lambda_0(T_i)\big) + x_i\beta + W_i \tag{A.2}$$

conditional on $\mathbf{W} = \mathbf{w}$, follows an extreme value distribution. Because the density of ε_i in the conditional distribution given $\mathbf{W} = \mathbf{w}$ does not depend on $\mathbf{w}$, then it follows that ε_i and $\mathbf{W}$ are independent and that the marginal distribution of ε_i is the extreme value distribution. By rearranging terms in (A.2) it follows, that the model in (A.2), is equivalent to a linear model on the $\log\big(\Lambda_0(\cdot)\big)$ scale:

$$\log\big(\Lambda_0(T_i)\big) = -x_i\beta - W_i + \varepsilon_i.$$

The unconditional mean and variance of $\log\big(\Lambda_0(T_i)\big)$ are $-x_i\beta - E(W_i) - \gamma_{\mathrm{E}}$ and $Var(W_i) + \frac{\pi^2}{6}$, respectively, where γ_{E} is the Euler constant.