

Genetics. 2006 Nov; 174(3): 1613–1624.

PMCID: PMC1667069

doi: [10.1534/genetics.106.060673](https://doi.org/10.1534/genetics.106.060673)

## A Thurstonian Model for Quantitative Genetic Analysis of Ranks: A Bayesian Approach

[Daniel Gianola](#)<sup>\*†‡,1</sup> and [Henner Simianer](#)<sup>†</sup>

<sup>\*</sup>Department of Animal Sciences, University of Wisconsin, Madison, Wisconsin 53706, <sup>†</sup>Institute of Animal Breeding and Genetics, Georg-August-University, 37075 Göttingen, Germany and <sup>‡</sup>Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Ås, Norway

<sup>1</sup>*Corresponding author:* Department of Animal Sciences, University of Wisconsin, 1675 Observatory Dr., Madison, WI 53706. E-mail:

[gianola@ansci.wisc.edu](mailto:gianola@ansci.wisc.edu)

Communicating editor: J. B. WALSH

Received 2006 May 12; Accepted 2006 Sep 8.

[Copyright](#) © 2006 by the Genetics Society of America

This article has been [cited by](#) other articles in PMC.

### Abstract

[Go to:](#) ☒ [Go to:](#) ☒

A fully Bayesian method for quantitative genetic analysis of data consisting of ranks of, *e.g.*, genotypes, scored at a series of events or experiments is presented. The model postulates a latent structure, with an underlying variable realized for each genotype or individual involved in the event. The rank observed is assumed to reflect the order of the values of the unobserved variables, *i.e.*, the classical Thurstonian model of psychometrics. Parameters driving the Bayesian hierarchical model include effects of covariates, additive genetic effects, permanent environmental deviations, and components of variance. A Markov chain Monte Carlo implementation based on the Gibbs sampler is described, and procedures for inferring the probability of yet to be observed future rankings are outlined. Part of the model is rendered nonparametric by introducing a Dirichlet process prior for the distribution of permanent environmental effects. This can lead to potential identification of clusters of such effects, which, in some competitions such as horse races, may reflect forms of undeclared preferential treatment.

LATENT variable models for describing mechanisms by which some continuous scale maps into ordered or unordered categories of response have a long tradition in quantitative genetics and psychometry. For example, the threshold-liability model for ordinal categories dates back to [WRIGHT \(1934\)](#), [DEMPSTER and LERNER \(1950\)](#), [GRÜNEBERG \(1952\)](#), and [FALCONER \(1965\)](#); see [GIANOLA \(1982\)](#) for a review. Extremal models were pioneered in psychology by [THURSTONE \(1927\)](#) and were adapted by [BOCK and JONES \(1968\)](#) to explain choices between unordered alternatives.

Suppose that an observation is an assignment into one of  $M$  mutually exclusive and exhaustive categories of response. The classical Thurstonian extremal model postulates the existence of some latent or unobserved continuous valued vector  $\mathbf{y}_{M \times 1} = \{y_i\}$ , such that category  $i$  is observed when  $y_i$  is larger than the other  $M - 1$  elements of the vector. The probability of observing an assignment into category  $M$ , say, is given by

$$\Pr(y_M > y_{M-1}, y_M > y_{M-2}, \dots, y_M > y_1 | \text{parameters}),$$

where “parameters” are typically unknown quantities driving the process. The extremal model can also be applied to data consisting of ranks, presumably reflecting the outcome of some underlying continuous process, *e.g.*, competing ability. For instance, [HENERY \(1981\)](#) and [TAVERNIER \(1991\)](#) used this idea in the context of horse races, and [JOHNSON \*et al.\* \(2002\)](#) for analysis of ranked data collected in primate intelligence experiments. Ranked data also arise in human genetics (*e.g.*, [LAWRENCE \*et al.\* 1994](#)). In [HENERY \(1981\)](#) and [TAVERNIER \(1991\)](#), the setting is one where several individuals compete in a series of events (*e.g.*, horse races). Prior to the event, each individual generates an unobservable random variable, or “liability of winning the event,” and these liabilities are regarded as mutually independent, given some parameters. Posterior to the event, the observed ranking corresponds to the order of the realized values of these liabilities. Clearly, the latent variables are no longer independent, *a posteriori*, because the ranking implies that  $y_{[1]} > y_{[2]} \dots > y_{[M-1]} > y_{[M]}$ , where  $y_{[1]}$  is the liability of the individual ranked first, and so on. In particular, [TAVERNIER \(1991\)](#) described an empirical Bayes procedure for inferring features of the latent distribution, such as breeding value of racing horses or competing ability of each of the members of a list of competitors. The approach of [TAVERNIER \(1991\)](#) is based on calculating, iteratively, a conditional posterior mode, given some values of dispersion parameters; these, in turn, are estimated using an approximation to marginal maximum-likelihood estimation. Her procedure uses statistical and numerical approximations that are not needed in a Bayesian implementation based on Markov chain Monte Carlo (MCMC) sampling.

The objective of this study is to present theory for a fully Bayesian method for quantitative genetic analysis of data consisting of ranks scored at each of a series of runs, experiments, or events. The method uses a Thurstonian representation of the mechanism by which an underlying continuum is translated into observed ranks. The parameters driving the Bayesian hierarchical model include additive genetic and permanent environmental effects, and the MCMC implementation allows easy computation of probabilities of yet to be observed runs. The article is organized as follows. PROBABILITY DISTRIBUTION OF OBSERVED RANKS outlines the Thurstonian concept and gives the sampling model for the observed data. In BAYESIAN STRUCTURE, a hierarchical specification is outlined. CONDITIONAL MAXIMUM *A POSTERIORI* ESTIMATION reviews the empirical Bayes methodology suggested by [TAVERNIER \(1991\)](#) and discusses potential shortcomings of this type of treatment of data. FULLY BAYESIAN ANALYSIS gives details of the MCMC scheme proposed and presents procedures for inference and prediction. Some assumptions are relaxed in NONPARAMETRIC DISTRIBUTION FOR EFFECTS. Specifically, the assumption of a normal distribution of permanent environmental deviates is replaced by a Dirichlet process, using a normal distribution as baseline probability measure; a Gibbs sampling scheme is presented. The article concludes with a discussion.

## PROBABILITY DISTRIBUTION OF OBSERVED RANKS

Go to: ☒ Go to: ☒

Suppose  $n_k$  individuals “compete” in event  $k$  and that the information available for analysis is that furnished by the ranks observed in the event. It is assumed that no ties are possible, but this restriction is easy to lift. As in [HENERY \(1981\)](#) and [TAVERNIER \(1991\)](#), it is supposed that the observed ranking is a manifestation of the magnitudes of latent variables, liability of winning the event, which are specific to each of the individuals.

The model for the unobserved liability of individual  $j$  ( $j = 1, 2, \dots, J$ ) winning event  $k$  ( $k = 1, 2, \dots, K$ ) is the standard specification of quantitative genetics,

$$\begin{aligned} l_{jk} &= \mathbf{x}_{jk}'\beta + a_j + p_j + g_k + e_{jk} \\ &= \mu_{jk} + e_{jk} \end{aligned} \tag{1}$$

(DEMPSTER and LERNER 1950; FALCONER 1965; GIANOLA 1982; GIANOLA and FOULLEY 1983), where  $\beta$  is a set of  $F$  location effects affecting liability and  $\mathbf{x}'_{jk}$  is a known incidence vector peculiar to individual  $j$  in event  $k$ ;  $a_j$  and  $p_j$  are the additive genetic and permanent environmental effects, respectively, of individual  $j$  on liability;  $g_k$  is an effect peculiar to event  $k$  (which may or may not be needed in the model) and  $\mathbf{e}_{jk} \sim N(0, \sigma_e^2)$  is a random residual, distributed independently of all terms in the model, within and across subjects; and  $\sigma_e^2$  is the residual variance. In (1),  $\mu_{jk} = \mathbf{x}'_{jk}\beta + a_j + p_j + g_k$ . The liabilities are not observable, so it is assumed that  $\sigma_e^2 = 1$  for parameter identification purposes. Hence, all effects are measured in units of residual standard deviation.

The liabilities of the “competitors” present in event  $k$  can be packed into the  $n_k \times 1$  vector  $\mathbf{l}_k$ , and the corresponding vectorial representation is

$$\mathbf{l}_k = \mathbf{X}_k\beta + \mathbf{Z}_{ak}\mathbf{a} + \mathbf{Z}_{pk}\mathbf{p} + \mathbf{Z}_{gk}\mathbf{g} + \mathbf{e}_k, \quad (2)$$

where  $\mathbf{a}$  is a  $Q \times 1$  vector of additive genetic effects of all potential participants or genetically related individuals;  $\mathbf{p}$  is a  $C \times 1$  vector of permanent environmental effects (including perhaps nonadditive genetic deviations) of individuals with at least one record of competition;  $\mathbf{g}$  is a  $K \times 1$  vector of event effects;  $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{I}_{n_k})$  is a residual vector, and  $\mathbf{I}_{n_k}$  is an identity matrix of order  $n_k$ ; and  $\mathbf{X}_k$ ,  $\mathbf{Z}_{ak}$ ,  $\mathbf{Z}_{pk}$ , and  $\mathbf{Z}_{gk}$  are incidence matrices of appropriate order. Vertical concatenation of all liabilities into a  $\sum_{k=1}^K n_k \times 1$  vector  $\mathbf{l}$  leads to the representation

$$\begin{aligned} \mathbf{l} &= \mathbf{X}\beta + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_p\mathbf{p} + \mathbf{Z}_g\mathbf{g} + \mathbf{e} \\ &= \mathbf{W}\theta + \mathbf{e}, \end{aligned} \quad (3)$$

where  $\theta = [\beta', \mathbf{a}', \mathbf{p}', \mathbf{g}']'$  and  $\mathbf{W} = [\mathbf{X}, \mathbf{Z}_a, \mathbf{Z}_p, \mathbf{Z}_g]$ .

Prior to the competition, and conditionally on  $\beta$ ,  $\mathbf{a}$ ,  $\mathbf{p}$ , and  $\mathbf{g}$ , the joint density of the yet to be realized liabilities, in view of the mutual independence of the residuals, is

$$\begin{aligned} &p(\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_K | \beta, \mathbf{a}, \mathbf{p}, \mathbf{g}, \delta) \\ &= \prod_{k=1}^K p(\mathbf{l}_k | \theta, \delta) = \prod_{k=1}^K \prod_{j=1}^J [N(l_{jk} | \mu_{jk}, 1)]^{\delta_{jk}}, \end{aligned} \quad (4)$$

where  $N(l_{jk} | \mu_{jk}, 1)$  denotes the density of the normally distributed liability  $l_{jk}$ , with mean  $\mu_{jk}$  and variance 1 and  $\delta = \{\delta_{jk}\}$  is a set of indicator variables; i.e.,  $\delta_{jk} = 1$  if competitor  $j$  is present in event  $k$  and 0 otherwise.

Consider competition  $k$ , such that the “winner” is the individual with the largest liability among the  $n_k$  competitors. An ordered variable is denoted as  $l_{j[\text{rank}]k}$ , which is the realized liability of individual  $j$  attaining rank = 1, 2, ...,  $n_k$  in the event; for instance, if a competitor whose identification is 20 places second in the event, then the corresponding liability is denoted as  $l_{20[2]k}$ . Following HENERY (1981) and TAVERNIER (1991), the probability that a given order or ranking is observed in event  $k$  ( $\mathbf{DATA}_k$ ) is given by

$$\begin{aligned} P_k(\theta) &= P(\mathbf{DATA}_k | \theta, \delta) = P(l_{j[1]k} > l_{j[2]k} \dots > l_{j[n_k]k} | \theta) \\ &= \int_{-\infty}^{\infty} \int_{l_{j[n_k]k}}^{\infty} \dots \int_{l_{j[2]k}}^{\infty} \prod_{j=1}^J [N(l_{jk} | \mu_{j[j]k}, 1)]^{\delta_{jk}} \\ &\quad \times dl_{j[1]k} dl_{j[2]k} \dots dl_{j[n_k]k}. \end{aligned} \quad (5)$$



This probability can also be expressed in terms of nine contrasts between liabilities being  $>0$ . Since the latent variables are not observable, one can “anchor” the liability of the winner to 0, so that all other liabilities are negative; the mean liability of a “perfect” competitor (*i.e.*, one winning all events) would then be null. The probability of the ranks observed in all  $K$  competitions (DATA), assuming that outcomes of different races are conditionally independent, given  $\theta$ , is

$$P(\text{DATA}|\theta) = \prod_{k=1}^K P_k(\theta). \quad (6)$$

## BAYESIAN STRUCTURE

Go to: ☐ Go to: ☐

Consider now the problem of inferring, via the Bayesian approach, unobserved features of interest, *e.g.*, the vector of additive genetics  $\mathbf{a}$ , as it would be the case in horse breeding, or the linear combination  $\mathbf{v} = \mathbf{a} + \mathbf{p}$ , which can be termed “competing ability.” Define  $\Lambda = [\sigma_a^2, \sigma_p^2, \sigma_g^2]'$ , a vector of variance parameters. Assume that the density of the joint prior distribution of  $\theta$  and  $\Lambda$  has the form

$$p(\theta, \Lambda|H) = N_F(\beta|\mathbf{0}, \mathbf{I}\sigma_\beta^2) N_Q(\mathbf{a}|\mathbf{0}, \mathbf{A}\sigma_a^2) N_C(\mathbf{p}|\mathbf{0}, \mathbf{I}\sigma_p^2) \\ \times N_K(\mathbf{g}|\mathbf{0}, \mathbf{I}\sigma_g^2) \times \prod_{i=a,p,g} p(\sigma_i^2|\nu_i, S_i^2). \quad (7)$$

Above,  $N_F(\beta|\mathbf{0}, \mathbf{I}\sigma_\beta^2)$ ,  $N_Q(\mathbf{a}|\mathbf{0}, \mathbf{A}\sigma_a^2)$ ,  $N_C(\mathbf{p}|\mathbf{0}, \mathbf{I}\sigma_p^2)$ , and  $N_K(\mathbf{g}|\mathbf{0}, \mathbf{I}\sigma_g^2)$  denote multivariate normal densities of orders  $F$ ,  $Q$ ,  $C$ , and  $K$ , respectively;  $\sigma_\beta^2$  is a variance reflecting uncertainty about location parameters in  $\beta$  (it is assumed known and large);  $\mathbf{A}$  is the numerator additive relationship matrix ([HENDERSON 1976](#)) and  $\sigma_a^2$  is the additive genetic variance;  $\sigma_p^2$  is the variance of permanent environmental effects and  $\sigma_g^2$  is the variance between events. In addition,  $p(\sigma_i^2|\nu_i, S_i^2)$  is the density of a scaled inverted chi-square distribution on  $\nu_i$  degrees of freedom, with  $S_i^2$  interpretable as a prior guess for  $\sigma_i^2$ , and  $H = [\sigma_\beta^2, \nu_a, S_a^2, \nu_p, S_p^2, \nu_g, S_g^2]'$  is a set of known hyperparameters.

The density of the joint posterior distribution of  $\theta$  and  $\Lambda$  is proportional to the product of (6) and (7):

$$p(\theta, \Lambda|\text{DATA}, H, \delta) \\ \propto \prod_{k=1}^K P_k(\theta) N_F(\beta|\mathbf{0}, \mathbf{I}\sigma_\beta^2) N_Q(\mathbf{a}|\mathbf{0}, \mathbf{A}\sigma_a^2) N_C(\mathbf{p}|\mathbf{0}, \mathbf{I}\sigma_p^2) \\ \times N_K(\mathbf{g}|\mathbf{0}, \mathbf{I}\sigma_g^2) \prod_{i=a,p,g} p(\sigma_i^2|\nu_i, S_i^2). \quad (8)$$

## CONDITIONAL MAXIMUM A POSTERIORI ESTIMATION

Go to: ☐ Go to: ☐

Prior to the advent of MCMC procedures for sampling from posterior distributions, it was standard to use two-stage approaches to carry out an approximate Bayesian analysis. In the first stage, a Gaussian approximation to  $[\Delta|\text{DATA}]$  is used to obtain an approximate maximum marginal-likelihood estimate of  $\Lambda$  ([GIANOLA \*et al.\* 1986](#); [FOULLEY \*et al.\* 1987](#)),  $\hat{\Lambda}$ , say. In the second stage, the joint mode of the conditional posterior distribution  $[\theta|\Lambda = \hat{\Lambda}, \text{DATA}]$  is used to obtain point estimates of  $\theta$  ([GIANOLA and FOULLEY 1983](#)) given  $\Lambda = \hat{\Lambda}$ , or “conditional maximum *a posteriori*” (MAP). If the distribution  $[\theta|\Lambda, \text{DATA}]$  is exactly Gaussian, its posterior mode yields directly the posterior means of each of the elements of  $\theta$  ([GIANOLA and FERNANDO 1986](#)). If, in

addition, a flat prior is used for  $\beta$ , the posterior mean vector of  $[\theta|\Lambda, \text{DATA}]$  yields the best linear unbiased predictor of  $\beta$  and the best linear unbiased predictors of  $\mathbf{a}$ ,  $\mathbf{p}$ , and  $\mathbf{g}$ . Also, if a flat prior is assigned to  $\beta$  and to each of the elements of  $\Lambda$ , under Gaussian assumptions the elements of the mode of the marginal distribution  $[\Delta|\text{DATA}]$  coincide with the restricted maximum-likelihood estimates of the variance components ([HARVILLE 1974](#)).

[TAVERNIER \(1990, 1991\)](#) used this type of approach for finding the  $\theta$ -maximizer of the log of the conditional posterior density  $p(\theta|\Lambda, \text{DATA}, \delta)$ , assuming a flat prior for  $\beta$ . The objective function for the model in (1), assuming a normal prior for  $\beta$ , is

$$L(\theta|\Lambda, \text{DATA}, \delta) = \sum_{k=1}^K \log P_k(\theta) - \frac{1}{2} \left( \frac{\beta' \beta}{\sigma_\beta^2} + \frac{\mathbf{a}' \mathbf{A}^{-1} \mathbf{a}}{\sigma_a^2} + \frac{\mathbf{p}' \mathbf{p}}{\sigma_p^2} + \frac{\mathbf{g}' \mathbf{g}}{\sigma_g^2} \right) \quad (9)$$

[TAVERNIER \(1990, 1991\)](#) employed the Newton–Raphson algorithm for maximizing (9) in the context of horse races. This requires calculation of first and second derivatives of (9) with respect to  $\theta$ . The gradient vector and the Hessian matrix needed for the Newton–Raphson machinery are

$$\frac{\partial L}{\partial \theta} = \sum_{k=1}^K \frac{1}{P_k(\theta)} \frac{\partial P_k(\theta)}{\partial \theta} - \Sigma^{-1} \theta,$$

and

$$\begin{aligned} \frac{\partial^2 L}{\partial \theta \partial \theta'} &= \sum_{k=1}^K \left[ \frac{1}{P_k(\theta)} \frac{\partial^2 P_k(\theta)}{\partial \theta \partial \theta'} - \left( \frac{1}{P_k(\theta)} \frac{\partial P_k(\theta)}{\partial \theta} \right) \right. \\ &\quad \left. \times \left( \frac{1}{P_k(\theta)} \frac{\partial P_k(\theta)}{\partial \theta'} \right) \right] - \Sigma^{-1}, \end{aligned}$$

respectively, where

$$\Sigma^{-1} = \begin{bmatrix} \mathbf{I}_F \sigma_\beta^{-2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \mathbf{A}^{-1} \sigma_a^{-2} & \mathbf{0} & \mathbf{0} \\ & & \mathbf{I}_C \sigma_p^{-2} & \mathbf{0} \\ \text{symmetric} & & & \mathbf{I}_K \sigma_g^{-2} \end{bmatrix}$$

These derivatives are not trivial and involve integration of multivariate normal densities, as well as differentiation under the integral sign. [TAVERNIER \(1991\)](#) gives approximations to first and second derivatives that require calculation of expectations, variances, and covariances of order statistics. Also, numerical calculation of high-order multivariate normal integrals is cumbersome. Since the Newton–Raphson algorithm is iterative and the number of competitors and races in a data set can be very large, the procedure is numerically intensive, as a large number of calculations must be effected.

Aside from these technical issues, the procedure relies on statistical approximations that may not be always adequate. For instance, in a high-dimensional, nonnormal, multivariate distribution, the components of the mode

of the joint posterior density may be far from the corresponding posterior expectations. Also, computing a measure of precision of estimates of  $\theta$  is a task by itself: a Gaussian approximation to the conditional (given  $\Lambda$ ) posterior covariance matrix is obtained by inverting the negative of the Hessian matrix and by evaluating this inverse at the converged values of  $\theta$ . The diagonal elements of this inverse give a large sample approximation to the conditional posterior variances of the elements of  $\theta$ . Last but not least, the procedure for estimating parameters in  $\Lambda$  is also based on Gaussian approximations and, furthermore, the approximate modal inferences about  $\theta$  do not take into account the uncertainty of the approximate estimates of  $\Lambda$ . All these difficulties can be circumvented via MCMC procedures, as discussed in the following sections.

## FULLY BAYESIAN ANALYSIS

Go to: ☐ Go to: ☐

**Joint posterior distribution:** The Bayesian structure given above is maintained, but the parameter vector is augmented with the unobserved liabilities as latent variables ([TANNER and WONG 1987](#); [SORENSEN \*et al.\* 1995](#)). A main difference with a standard latent variable problem is that, in our setting, the sample space of the liabilities is order constrained, within events.

After augmentation with the liabilities  $\mathbf{l}$ , the joint posterior density is expressible as

$$\begin{aligned}
 & p(\mathbf{l}, \theta, \Delta | \text{DATA}, H, \delta) \\
 & \propto \prod_{k=1}^K \left\{ \prod_{j=1}^J [N(l_{jk} | \mu_{jk}, 1)]^{\delta_{jk}} I_k \right\} N_F(\beta | \mathbf{0}, \mathbf{I}\sigma_\beta^2) \\
 & \quad \times N_Q(\mathbf{a} | \mathbf{0}, \mathbf{A}\sigma_a^2) N_C(\mathbf{p} | \mathbf{0}, \mathbf{I}\sigma_p^2) \\
 & \quad \times N_K(\mathbf{g} | \mathbf{0}, \mathbf{I}\sigma_g^2) \prod_{i=a,p,g} p(\sigma_i^2 | \nu_i, S_i^2).
 \end{aligned} \tag{10}$$

Above, the indicator set  $I_k$  ( $k = 1, 2, \dots, K$ ) denotes the order of liabilities, or ranking attained, observed in competition  $k$ . For example, if competitors 1, 5, 7, and 14 participate in event 20, and the ranking observed is  $5 > 14 > 1 > 7$ , then, following the notation employed in (10),  $I_{20} = \{l_{5[1]20}, l_{7[4]20}, l_{14[2]20}, l_{1[3]20}\}$ .

**Markov chain Monte Carlo sampling:** A Gibbs sampler is described for effecting draws from the joint posterior distribution (10). This is a well-known procedure for sampling from joint distributions; see, *e.g.*, [GILKS \*et al.\* \(1996\)](#) and [ROBERT and CASELLA \(2005\)](#). Briefly, the Gibbs sampler loops through all conditional posterior distributions, with uncertain quantities sampled either blockwise or piecewise. All samples from the fully conditionals are accepted with probability 1. Samples from early iterations are discarded as “burn in”; subsequently, at the end of each cycle of sampling, the coordinates of sample  $(\mathbf{l}^{[s]}, \theta^{[s]}, \Delta^{[s]}; s = 1, 2, \dots, S)$  are regarded as draws from the corresponding marginal posterior distributions. The chain can be thinned, such that samples are as lowly correlated (serially) as desired. If storage capacity is not a limiting factor, then it is better to use all samples in the post-Gibbs analysis. Convergence diagnostics and processing issues are discussed in [COWLES and CARLIN \(1996\)](#), [SORENSEN and GIANOLA \(2002\)](#), and [GELMAN \*et al.\* \(2003\)](#).

**Sampling of liabilities:** The form of (10) indicates that the conditional posterior distribution of the liabilities in event  $k$  has the form

$$p(\mathbf{l}_k | \text{ELSE}) \propto \prod_{j=1}^J [N(l_{jk} | \mu_{jk}, 1)]^{\delta_{jk}} I_k; \quad k = 1, 2, \dots, K, \tag{11}$$

where ELSE denotes the unknown liabilities at all other events,  $\theta$ ,  $\Delta$ ,  $\delta$ , and DATA. Note that, conditionally on ELSE, the liabilities at different events are mutually independent. However, within event  $k$ , liabilities are not independent, because of the sample space constraint imposed by knowledge of its outcome, represented by  $I_k$ . The dependence arises through the sample space, but the kernel of the conditional distribution of the  $n_k$  liabilities in event  $k$ , given ELSE, is proportional to the product of the kernels under conditional independence. Sampling from the space-constrained distribution is straightforward:

Draw “independently” the  $n_k$  liabilities of competitors in event  $k$  (whose indexes are in  $I_k$ ) from the distributions  $N(l_{jk}|\mu_{jk}, 1), j \in I_k$ .

If the liabilities sampled satisfy the order constraint in  $I_k$ , then accept the vector  $\mathbf{l}_k$  as a draw from the posterior distribution. Otherwise, reject the proposal and repeat the sampling until a valid draw is effected.

This procedure, although simple, is typically inefficient and it may have a large rejection rate. A more efficient sampling technique, described by [DEVROYE \(1986\)](#), is given in more detail in [SORENSEN and GIANOLA \(2002\)](#). To illustrate, let the outcome of competition 20 be as in the example above; that is,  $I_{20} = \{l_{5[1]20}, l_{14[2]20}, l_{1[3]20}, l_{7[4]20}\}$ . Hence, in event 20, the sample space of the liabilities is  $l_{7,20} < l_{1,20} < l_{14,20} < l_{5,20}$ . To sample the four liabilities, one proceeds from either first to last or last to first (the liability of the winner can be set equal to zero, as stated above). For instance:

Sample  $l_{7,20}$  from  $N(l|\mu_{7,20}, 1)$  without any constraint. This is the realized value of the liability of the last competitor.

Sample  $l_{1,20}$  from the truncated normal distribution  $\text{TN}_{l_{7,20};\infty}(l|\mu_{1,20}, 1)$ , where the parameters are as in the absence of truncation. The realized value is formed as

$$\begin{aligned} l_{1,20} &= \mu_{1,20} + \Phi^{-1}\{\Phi(l_{7,20} - \mu_{1,20}) + U[1 - \Phi(l_{7,20} - \mu_{1,20})]\} \\ &= \mu_{1,20} + \Phi^{-1}\{U + (1 - U)\Phi(l_{7,20} - \mu_{1,20})\}, \end{aligned}$$

where  $\Phi(\cdot)$  is the standard normal distribution function and  $U$  is a uniform random number in  $(0,1)$ .

Sample  $l_{14,20}$  from the truncated normal process  $\text{TN}_{l_{1,20};\infty}(l|\mu_{14,20}, 1)$ , forming the realized value as

$$l_{14,20} = \mu_{14,20} + \Phi^{-1}\{U + (1 - U)\Phi(l_{1,20} - \mu_{14,20})\}.$$

Sample the liability of the winner from  $\text{TN}_{l_{14,20};\infty}(l|\mu_{5,20}, 1)$  as

$$l_{5,20} = \mu_{5,20} + \Phi^{-1}\{U + (1 - U)\Phi(l_{14,20} - \mu_{5,20})\}.$$

Once the  $K$  events are processed independently, this produces a realization of the  $\sum_{k=1}^K n_k \times 1$  vector of liabilities at iteration  $s$ ,  $\mathbf{l}(s)$ , which becomes the “data” in model (3) at iteration  $s = 1, 2, \dots, S$  of the Gibbs sampler.

**Sampling of parameters:** Given the liabilities, the conditional posterior density of all parameters can be deduced from (10) after fixing the vector  $\mathbf{l}$ . This yields



$$\begin{aligned}
& p(\boldsymbol{\theta}, \boldsymbol{\Delta} | \mathbf{l}, H, \delta) \\
& \propto \prod_{k=1}^K \left\{ \prod_{j=1}^J [N(l_{jk} | \mu_{jk}, 1)]^{\delta_{jk}} \right\} N_F(\boldsymbol{\beta} | \mathbf{0}, \mathbf{I}\sigma_{\boldsymbol{\beta}}^2) \\
& \quad \times N_Q(\mathbf{a} | \mathbf{0}, \mathbf{A}\sigma_{\mathbf{a}}^2) N_C(\mathbf{p} | \mathbf{0}, \mathbf{I}\sigma_{\mathbf{p}}^2) \\
& \quad \times N_K(\mathbf{g} | \mathbf{0}, \mathbf{I}\sigma_{\mathbf{g}}^2) \prod_{i=a,p,g} p(\sigma_i^2 | \nu_i, S_i^2).
\end{aligned} \tag{12}$$

Note that, given the liabilities, the information provided by the sets  $I_k$  is redundant. The density above is that of a Bayesian mixed-effects model under Gaussian assumptions ([GIANOLA and FERNANDO 1986](#); [WANG \*et al.\* 1993, 1994](#)) with the sampled liabilities playing the role of data. The conditional posterior distributions of  $\boldsymbol{\theta}$  and  $\boldsymbol{\Delta}$  are in closed form ([WANG \*et al.\* 1993, 1994](#)), as follows. The location parameters  $\boldsymbol{\theta}$  can be sampled from the multivariate normal process

$$\boldsymbol{\theta} | \text{ELSE} \sim N(\mathbf{C}^{-1}\mathbf{r}, \mathbf{C}^{-1}), \tag{13}$$

where

$$\mathbf{C} = \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1},$$

and

$$\mathbf{r} = \mathbf{W}'\mathbf{l}$$

The draw from distribution (13) can be done either in a single pass ([GARCIA-CORTÉS and SORESENSEN 1996](#)) or element by element ([WANG \*et al.\* 1994](#)); the rate of convergence to the equilibrium distribution is lower and the autocorrelation between samples higher when sampling is piecewise.

All conditional posterior distributions of the dispersion parameters are scaled inverted chi square and are mutually independent. The distributions are

$$\sigma_{\mathbf{a}}^2 | \text{ELSE} \sim (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + \nu_{\mathbf{a}}S_{\mathbf{a}}^2) \chi_{\nu_{\mathbf{a}}+Q}^{-2}, \tag{14}$$

$$\sigma_{\mathbf{p}}^2 | \text{ELSE} \sim (\mathbf{p}'\mathbf{p} + \nu_{\mathbf{p}}S_{\mathbf{p}}^2) \chi_{\nu_{\mathbf{p}}+C}^{-2}, \tag{15}$$

and

$$\sigma_{\mathbf{g}}^2 | \text{ELSE} \sim (\mathbf{g}'\mathbf{g} + \nu_{\mathbf{g}}S_{\mathbf{g}}^2) \chi_{\nu_{\mathbf{g}}+K}^{-2}. \tag{16}$$

The Gibbs sampling algorithm consists of drawing liabilities from distributions (11), such that the order constraints are satisfied, followed by draws from (13) and (14)–(16) to obtain parameter updates. The sampling process is repeated as needed, after convergence diagnostics are satisfied, and such that the Monte Carlo error of estimation of features of the posterior distribution is small enough.

**Inference:** The procedure of [Tavernier \(1991\)](#) produces conditional posterior modes of location parameters, as already noted, but can also be used to estimate the probability of observing a certain outcome, even in a



hypothetical competition, although ignoring uncertainty about estimates. The fully Bayesian Markov chain Monte Carlo procedure is more flexible. First, it is possible to produce a complete posterior distribution for any uncertain quantity, *e.g.*, the competing ability  $a_j + p_j$  of individual  $j$  or the additive genetic variance  $\sigma_a^2$ . In addition, it is possible to estimate the probability of any event related to a future competition. Let  $I_f$  be an order associated with a yet to be realized event referred to as “future,” with  $n_f$  participants. The predictive probability distribution of  $I_f$ , posterior to historical data (**DATA**), is given by

$$\begin{aligned} & P(I_f | \mathbf{DATA}, H, \delta) \\ &= \int P(l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f} | \theta) p(\theta | \mathbf{DATA}, H, \delta) d\theta \\ &= E_{\theta | \mathbf{DATA}} [P(l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f} | \theta)], \end{aligned} \quad (17)$$

where  $E_{(.)}$  denotes expectation with respect to the distribution  $(.)$ . The integral in (17) does not have a closed form, since the posterior distribution  $[\theta | \mathbf{DATA}, H, \delta]$  is unknown. However, it can be estimated from the Markov chain Monte Carlo samples via ergodic averaging.

Let  $\theta^{[s]}$ ,  $s = 1, 2, \dots, S$  be the  $\theta$ -coordinate of a sample from the posterior distribution  $[\theta, \Delta | \mathbf{DATA}, H, \delta]$ , with density (10), and let

$$P(l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f} | \theta^{(s)})$$

be the probability of observing the ranking in  $I_f$  in the future liabilities when  $\theta = \theta^{[s]}$ . Note that, since event effects are assumed to be independent, **DATA** will not contain information about a future event effect  $g_f$ . This does not pose a problem, as the posterior distribution can be augmented with this effect; however, **DATA** will convey information about  $\sigma_g^2$ . A simulation-consistent estimator of (17) is

$$\hat{P}(I_f | \mathbf{DATA}, H, \delta) = \frac{1}{S} \sum_{s=1}^S J^{(s)} \left( l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f} \right), \quad (18)$$

where  $J^{(s)} \left( l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f} \right)$  is an indicator variable taking the value 1 if the ranking in  $I_f$  is attained in sample  $s$  and 0 otherwise. In practice, one samples liabilities for future event  $f$  from

$$p(l_f | \theta^{(s)}) \propto \prod_{j=1}^J [N(l_{jf} | \mu_{jf}^{(s)}, 1)]^{\delta_{jf}}$$

and checks whether or not the realized values satisfy  $l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f}$ ; if so,

$J^{(s)} \left( l_{j[1]f} > l_{j[2]f} \dots > l_{j[n_f]f} \right) = 1$ . This process is repeated for every Monte Carlo sample, leading directly to (18).

A similar procedure can be used to estimate the probability of any future event  $A$ , such as whether “participant”  $j$  is first, second, or third in a competition involving  $n_f$  competitors; let such an event be denoted as  $A$ . Then, the probability of observing  $A$  in a future event can be estimated as

$$\hat{P}(A|\text{DATA}, H, \delta) = \frac{1}{S} \sum_{s=1}^S J^{(s)}(j \text{ first, second, or third}).$$

Operationally, one samples liabilities  $\mathbf{l}_f$  from  $[\mathbf{l}_f|\theta^{(s)}]$  as given above and checks whether or not  $A$  is realized, in which case  $J^{(s)}(j \text{ first, second, or third}) = 1$  and 0 otherwise.

## NONPARAMETRIC DISTRIBUTION FOR EFFECTS

Go to: ☐ Go to: ☐

Above, it was assumed that prior uncertainty about permanent environmental or event effects could be described reasonably by the normal distributions  $N_C(\mathbf{p}|\mathbf{0}, \mathbf{I}\sigma_p^2)$  and  $N_K(\mathbf{g}|\mathbf{0}, \mathbf{I}\sigma_g^2)$ , respectively. However, it might be sensible to make less strong assumptions about these distributions. For instance, there may be clusters of events that are more similar to each other, *i.e.*, clusters representing different, unknown, levels of competition. Likewise, there may be some competitors that receive similar “preferential” treatment. A suitable representation of uncertainty may be that furnished by the Dirichlet process prior ([ESCOBAR 1994](#); [MACEachern 1994](#); [KLEINMAN and IBRAHIM 1998](#); [VAN DER MERWE and PRETORIUS 2003](#)). The case of the permanent environmental effects is discussed here. *A priori*, let the permanent effects  $p_i$  be independently distributed as

$$p_i \sim G,$$

where  $G$  is some general distribution. In turn, assume that  $G$  follows a Dirichlet process (DP) prior

$$G \sim \text{DP}(M, G_0(\sigma_p^2))$$

([FERGUSON 1973](#); [ANTONIAK 1974](#)), meaning that  $G$  is a random member of a space of distributions, with  $M$  and  $G_0(\sigma_p^2)$  being parameters of these process. For example, it could be that  $G_0(\sigma_p^2) \equiv N_C(\mathbf{p}|\mathbf{0}, \mathbf{I}\sigma_p^2)$  is taken as “base prior,” a distribution approximating the nonparametric shape of  $G$ , with  $M$  interpretable as a degree of belief on how close  $N_C(\mathbf{p}|\mathbf{0}, \mathbf{I}\sigma_p^2)$  is to  $G$ . The parameter  $M$  is such that, when  $M \rightarrow \infty$ , then  $G \rightarrow G_0$ ; the same is true if all random effects are identical ([KLEINMAN and IBRAHIM 1998](#)). When  $G$  is mixed over the Dirichlet process, the Polya urn representation of the prior distribution of the random effects ([KLEINMAN and IBRAHIM 1998](#)) is, for  $N(0, \sigma_p^2)$  as base prior,

$$\begin{aligned} & p_1 \sim N(0, \sigma_p^2) \\ p_2 & \begin{cases} = p_1 \text{ with } \Pr(p_2 = p_1) = P_1 \\ \sim N(0, \sigma_p^2) \text{ with probability } 1 - P_1 \end{cases} \\ p_3 & \begin{cases} = p_1 \text{ with } \Pr(p_3 = p_1) = P_1 \\ = p_2 \text{ with } \Pr(p_3 = p_2) = P_2 \\ \sim N(0, \sigma_p^2) \text{ with probability } 1 - P_1 - P_2 \end{cases} \\ & \vdots \\ & \vdots \\ p_C & \begin{cases} = p_1 \text{ with } \Pr(p_C = p_1) = P_1 \\ = p_2 \text{ with } \Pr(p_C = p_2) = P_2 \\ \dots \\ = p_{C-1} \text{ with } \Pr(p_C = p_{C-1}) = P_{C-1} \\ \sim N(0, \sigma_p^2) \text{ with probability } 1 - \sum_{i=1}^{C-1} P_i, \end{cases} \end{aligned}$$

where the probabilities  $P_1, P_2, \dots, P_{C-1}$  follow from the parameters of the Dirichlet process. Using results in [ESCOBAR \(1994\)](#), [KLEINMAN and IBRAHIM \(1998\)](#), and [VAN DER MERWE and PRETORIUS \(2003\)](#), the prior distribution of the random effects generates the sequence

$$p_C | p_1, p_2, \dots, p_{C-1} \begin{cases} p_1 \sim N(0, \sigma_p^2) \\ = p_j \text{ with } \Pr(p_C = p_j) = \frac{1}{M+C-1}; \\ j = 1, 2, \dots, C-1 \\ \sim N(0, \sigma_p^2) \text{ with probability } \frac{M}{M+C-1}. \end{cases}$$

Note that as  $M \rightarrow \infty$ , the prior distribution tends toward the parametric process  $N(0, \sigma_p^2)$ , as noted earlier. The preceding implies that the conditional prior distribution of a permanent environmental effect ( $i$ , say), given all other such effects, is

$$p(p_i | p_{-i}) = \frac{1}{M+C-1} \sum_{j \neq i}^C \delta(p_j, B) + \frac{M}{M+C-1} N(0, \sigma_p^2), \quad (19)$$

where

$$\delta(p_j, B) = \begin{cases} p_j & \text{when } p_j \in B \\ 0 & \text{when } p_j \notin B, \end{cases}$$

and  $B$  is the set over which the sum is taken ([ESCOBAR 1994](#); [VAN DER MERWE and PRETORIUS 2003](#)). Note that  $p_i$  can be equal to the permanent environmental effect of other subjects with nonnull prior probability; this has an implication in the construction of conditional posterior distributions. Representation (19) illustrates that the conditional prior distribution of any of the permanent effects, given all other  $p$ 's, is a mixture of  $C-1$  degenerate distributions ( $\delta(p_j, B)$ ) with point masses at  $p_j$  ( $j \neq i$ ) and of the parametric base distribution  $N(0, \sigma_p^2)$ . The mixing probabilities of states  $j \neq i$  are  $1/(M+C-1)$  for each of the  $C-1$  “conditioning” states, and  $M/(M+C-1)$  for the probability that  $p_i$  is a draw from the normal process.

Given the liabilities, and assuming temporarily that  $M$  is assigned a fixed value, the joint posterior density of all uncertain quantities is obtained by modifying (12) into

$$\begin{aligned} & p(\theta, \Delta | l, H, \delta, M) \\ & \propto \prod_{k=1}^K \left\{ \prod_{j=1}^J [N(l_{jk} | \mu_{jk}, 1)]^{\delta_{jk}} \right\} N_F(\beta | \mathbf{0}, \mathbf{I}\sigma_\beta^2) N_Q(\mathbf{a} | \mathbf{0}, \mathbf{A}\sigma_a^2) \\ & \quad \times \prod_{i=1}^C [p_i \sim G][G \sim \text{DP}(M, G_0(\sigma_p^2))] \\ & \quad \times N_K(\mathbf{g} | \mathbf{0}, \mathbf{I}\sigma_g^2) \prod_{i=a, p, g} p(\sigma_i^2 | \nu_i, S_i^2). \end{aligned} \quad (20)$$

Above,  $p_i \sim G$  denotes that  $p_i$  follows the unknown distribution  $G$  drawn from a Dirichlet process with

$G_0(\sigma_p^2) \equiv N(p_i|0, \sigma_p^2)$  as base measure and with known parameter  $M$ . Since the Dirichlet process allows  $p_i$  to take any of the other  $p$ 's as a current value, data from all individuals with records of performance contribute to the conditional posterior distribution of  $p_i$ . One has as conditional posterior distribution

$$\begin{aligned}
 & p(p_i|\text{ELSE}) \\
 & \propto \prod_{k=1}^K \left\{ \prod_{i=1}^J [N(l_{ik}|\mu_{ik}, 1)]^{\delta_{ik}} \right\} p(p_i|p_{-i}) \\
 & \propto \prod_{k=1}^K \left\{ \prod_{i=1}^J [N(l_{ik}|\mu_{ik}, 1)]^{\delta_{ik}} \right\} \\
 & \times \left[ \frac{1}{M+C-1} \sum_{j \neq i}^C \delta(p_j, B) + \frac{M}{M+C-1} N(p_i|0, \sigma_p^2) \right] \\
 & \propto \sum_{j \neq i}^C \left\{ \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik}|\mu_{ik}, 1)]^{\delta_{ik}} \right\} \delta(p_j, B) \\
 & + M \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik}|\mu_{ik}, 1)]^{\delta_{ik}} N(p_i|0, \sigma_p^2).
 \end{aligned} \tag{21}$$

Under standard parametric assumptions, the fully conditional posterior distribution of  $p_i$  is given by

$$\begin{aligned}
 & p(p_i|\text{ELSE, parametric}) \\
 & = \frac{\prod_{k=1}^K \prod_{i=1}^J [N(l_{ik}|\mu_{ik}, 1)]^{\delta_{ik}} N(p_i|0, \sigma_p^2)}{\int_{-\infty}^{\infty} \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik}|\mu_{ik}, 1)]^{\delta_{ik}} N(p_i|0, \sigma_p^2) dp_i}.
 \end{aligned} \tag{22}$$

Using well-known results (e.g., [WANG et al. 1993, 1994](#); [SORENSEN and GIANOLA 2002](#)) it can be shown that

$$p_i|\text{ELSE, parametric} \sim N(\tilde{p}_i, \tilde{v}_{p_i}), \tag{23}$$

where

$$\tilde{p}_i = \frac{\mathbf{z}_{p_i}' (1 - \mathbf{X}\beta - \mathbf{Z}_a\mathbf{a} - \sum_{j \neq i}^C \mathbf{z}_{p_j} p_j - \mathbf{Z}_g\mathbf{g})}{\mathbf{z}_{p_i}' \mathbf{z}_{p_i} + (1/\sigma_p^2)},$$

$\mathbf{z}_{p_j}$  is the column of  $\mathbf{Z}_p$  in (3) pertaining to individual  $j$ , and

$$\tilde{v}_{p_i} = \left( \mathbf{z}_{p_i}' \mathbf{z}_{p_i} + \frac{1}{\sigma_p^2} \right)^{-1}.$$

In addition, let



$$\int_{-\infty}^{\infty} \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}} N(p_i | 0, \sigma_p^2) dp_i = K_i, \quad (24)$$

with this integral expressible in closed form. Note that the integrand is the product of the densities of all liabilities of individual  $i$ , with  $\beta$ ,  $\mathbf{a}$ , and  $\mathbf{g}$  as fixed parameters (the  $p$ 's of other individuals do not enter into the model for the liabilities of individual  $i$ ) times the  $N(p_i | 0, \sigma_p^2)$  density. Hence, with  $\mathbf{l}_i$  being the  $r_i \times 1$  vector of liabilities of the individual, and  $\bar{\mathbf{X}}_i$ ,  $\bar{\mathbf{Z}}_{ai}$ , and  $\bar{\mathbf{Z}}_{gi}$  being incidence matrices relating  $\mathbf{l}_i$  to  $\beta$ ,  $\mathbf{a}$ , and  $\mathbf{g}$ , respectively, a standard integration yields

$$\begin{aligned} K_i &= N_{r_i}(\mathbf{l}_i | \bar{\mathbf{X}}_i \beta + \bar{\mathbf{Z}}_{ai} \mathbf{a} + \bar{\mathbf{Z}}_{gi} \mathbf{g}, \mathbf{J}_{r_i} \sigma_p^2 + \mathbf{I}_{r_i}) \\ &= \frac{1}{(2\pi)^{r_i/2} |\mathbf{J}_{r_i} \sigma_p^2 + \mathbf{I}_{r_i}|^{1/2}} \exp \left[ -\frac{T_i}{2} \right], \end{aligned} \quad (25)$$

where

$$\begin{aligned} T_i &= (\mathbf{l}_i - \bar{\mathbf{X}}_i \beta - \bar{\mathbf{Z}}_{ai} \mathbf{a} - \bar{\mathbf{Z}}_{gi} \mathbf{g})' (\mathbf{J}_{r_i} \sigma_p^2 + \mathbf{I}_{r_i})^{-1} \\ &\quad \times (\mathbf{l}_i - \bar{\mathbf{X}}_i \beta - \bar{\mathbf{Z}}_{ai} \mathbf{a} - \bar{\mathbf{Z}}_{gi} \mathbf{g}), \end{aligned}$$

and  $\mathbf{J}_{r_i}$  is a matrix of 1's of order  $r_i$ . Employing (23) and (24) in (22) leads to

$$\prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}} N(p_i | 0, \sigma_p^2) dp_i = K_i N(\tilde{p}_i, \tilde{v}_{p_i}),$$

so that

$$\begin{aligned} p(p_i | \text{ELSE}) &\propto \sum_{j \neq i}^C \left\{ \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}} \right\} \\ &\quad \times \delta(p_j, B) + M K_i N(\tilde{p}_i, \tilde{v}_{p_i}). \end{aligned} \quad (26)$$

The form of (26) indicates that the fully conditional posterior distribution of  $p_i$  is a mixture of the  $C - 1$  degenerate distributions  $\delta(p_j, B)$  with point masses at  $p_j$  ( $j \neq i$ ) and of the parametric distribution  $N(\tilde{p}_i, \tilde{v}_{p_i})$ . Let now

$$\begin{aligned} Q_j &= \frac{\prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}}}{\sum_{j \neq i}^C \{ \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}} \} + M K_i}; \\ &\quad j = 1, \dots, i-1, i+1, \dots, C \\ 1 - \sum_{j \neq i}^C Q_j &= \frac{M K_i}{\sum_{j \neq i}^C \{ \prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}} \} + M K_i}. \end{aligned}$$

Similar to that in [VAN DER MERWE and PRETORIUS \(2003\)](#), the rule for drawing samples from the conditional posterior distribution  $p_i | \text{ELSE}$  is to take as a current value for  $p_i$

$$p_i \begin{cases} = p_j \text{ with } \Pr(p_i = p_j | \text{ELSE}); \\ j = 1, \dots, i-1, i+1, \dots, C \\ \sim N(\tilde{p}_i, \tilde{v}_{p_i}) \text{ with probability } 1 - \sum_{j \neq i}^C Q_j \end{cases}, \quad (27)$$

where  $p_j$  is the current state in the Markov chain of the permanent environmental effect of individual  $j$ , and  $Q_j$  is evaluated at the current values of all parameters entering into this probability. In the calculation of  $Q_j$ , the

permanent effect  $p_i$  is replaced by  $p_j$  when computing  $\prod_{k=1}^K \prod_{i=1}^J [N(l_{ik} | \mu_{ik}, 1)]^{\delta_{ik}}$ . This means that values of  $p_j$  that generate liabilities  $\mathbf{l}_i$  with higher plausibility are assigned a higher probability of selection. Eventually, this leads to clustering of permanent environmental effects into “plausibility” groups.

The fully conditional posterior distributions of all other parameters can be readily deduced from (20) by fixing  $\mathbf{p}$  as well as the liabilities

$$\begin{aligned} & p(\beta, \mathbf{a}, \mathbf{g}, \Delta | \mathbf{l}, \mathbf{p}, H, \delta, M) \\ & \propto \prod_{k=1}^K \left\{ \prod_{j=1}^J [N(l_{jk} | \mu_{jk}, 1)]^{\delta_{jk}} \right\} \\ & \times N_F(\beta | \mathbf{0}, \mathbf{I}\sigma_\beta^2) N_Q(\mathbf{a} | \mathbf{0}, \mathbf{A}\sigma_a^2) \\ & \times N_K(\mathbf{g} | \mathbf{0}, \mathbf{I}\sigma_g^2) \prod_{i=a, pg} p(\sigma_i^2 | \nu_i, S_i^2). \end{aligned} \quad (28)$$

The conditional posterior distribution of  $\beta$ ,  $\mathbf{a}$ , and  $\mathbf{g}$  given everything else is the multivariate normal process

$$\beta, \mathbf{a}, \mathbf{g} | \mathbf{l}, \mathbf{p}, \Delta, H, \delta, M \sim N(\mathbf{C}_*^{-1} \mathbf{r}_*, \mathbf{C}_*^{-1}), \quad (29)$$

where

$$\begin{aligned} \mathbf{C}_* &= \begin{bmatrix} \mathbf{X}'\mathbf{X} + \mathbf{I}_F\sigma_\beta^{-2} & \mathbf{X}'\mathbf{Z}_a & \mathbf{X}'\mathbf{Z}_g \\ \mathbf{Z}_a'\mathbf{X} & \mathbf{Z}_a'\mathbf{Z}_a + \mathbf{A}^{-1}\sigma_a^{-2} & \mathbf{Z}_a'\mathbf{Z}_g \\ \mathbf{Z}_g'\mathbf{X} & \mathbf{Z}_g'\mathbf{Z}_a & \mathbf{Z}_g'\mathbf{Z}_g + \mathbf{I}_K\sigma_g^{-2} \end{bmatrix}, \\ \mathbf{r}_* &= \begin{bmatrix} \mathbf{X}'(\mathbf{l} - \mathbf{Z}_p\mathbf{p}) \\ \mathbf{Z}_a'(\mathbf{l} - \mathbf{Z}_p\mathbf{p}) \\ \mathbf{Z}_g'(\mathbf{l} - \mathbf{Z}_p\mathbf{p}) \end{bmatrix} \end{aligned}$$

The conditional posterior distributions of the variance components are as in the parametric model, that is,

$$\sigma_a^2 | \text{ELSE} \sim (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + \nu_a S_a^2) \chi_{\nu_a + Q}^{-2}, \quad (30)$$

$$\sigma_p^2 | \text{ELSE} \sim (\mathbf{p}'\mathbf{p} + \nu_p S_p^2) \chi_{\nu_p + C}^{-2}, \quad (31)$$

and

$$\sigma_g^2 | \text{ELSE} \sim (g'g + \nu_g S_g^2) \chi_{\nu_g + K}^{-2}. \quad (32)$$

The Gibbs sampler draws liabilities as before, permanent environmental effects from (27),  $\beta$ ,  $\mathbf{a}$ , and  $\mathbf{g}$  from (29), and the three variance components from the preceding expressions.

At each iteration of the Gibbs sampler, the procedure induces a clustering structure of permanent environmental effects, with each cluster consisting of individuals sharing the same value of  $p_i$ . Following [VAN DER MERWE and PRETORIUS \(2003\)](#), let  $t$  be the number of unique permanent environmental effects ( $0 < t \leq C$ ). The posterior distribution of  $t$  can be obtained readily from the sampling procedure. [BUSH and MACEachern \(1996\)](#) suggest embedding an optional additional step in the Gibbs sampling procedure: let the number of clusters identified at iteration  $m$  be  $t^{(m)}$ ; this means that there are  $t^{(m)}$  distinct values of the  $p$ 's, which are labeled as  $\boldsymbol{\eta}^{(m)} = [\eta_{1^{(m)}}, \eta_{2^{(m)}}, \dots, \eta_{t^{(m)}}]'$ . The model for the liabilities in (3) is then rewritten as

$$l = \mathbf{X}\beta + \mathbf{Z}_a \mathbf{a} + \mathbf{Z}_{p^{(m)}} \boldsymbol{\eta} + \mathbf{Z}_g \mathbf{g} + \mathbf{e},$$

where  $\mathbf{Z}_{p^{(m)}}$  is a condensation of  $\mathbf{Z}_p$  of order  $\sum_{k=1}^K n_k \times t^{(m)}$  obtained by assigning liabilities to the appropriate cluster of permanent environmental effects to which they belong. At that point of the scheme, [KLEINMAN and IBRAHIM \(1998\)](#) and [VAN DER MERWE and PRETORIUS \(2003\)](#) suggest drawing a sample of  $\boldsymbol{\eta}$  from the process

$$\boldsymbol{\eta} | \text{ELSE} \sim N(\hat{\boldsymbol{\eta}}, \hat{\mathbf{V}}_{\boldsymbol{\eta}}),$$

where

$$\hat{\boldsymbol{\eta}} = \left( \mathbf{Z}_{p^{(m)}}' \mathbf{Z}_{p^{(m)}} + \mathbf{I}_{t^{(m)}} \frac{1}{\sigma_p^2} \right)^{-1} \mathbf{Z}_{p^{(m)}}' (l - \mathbf{X}\beta - \mathbf{Z}_a \mathbf{a} - \mathbf{Z}_g \mathbf{g}),$$

and

$$\hat{\mathbf{V}}_{\boldsymbol{\eta}} = \left( \mathbf{Z}_{p^{(m)}}' \mathbf{Z}_{p^{(m)}} + \mathbf{I}_{t^{(m)}} \frac{1}{\sigma_p^2} \right)^{-1}.$$

Once a sample of  $\boldsymbol{\eta}$  is obtained, say  $\boldsymbol{\eta}^{(m)}$ , the permanent environmental effects are reconstituted as  $\mathbf{p}^{(m+1)} = \mathbf{L}^{(m)} \boldsymbol{\eta}^{(m)}$ , where  $\mathbf{L}^{(m)}$  is a  $C \times t^{(m)}$  matrix that allocates cluster values to permanent environmental effects. Then, the MCMC process continues for other parameters as usual.

Turn attention now to inferring  $M$ , the degree of belief parameter of the Dirichlet process. Using results in [ANTONIAK \(1974\)](#) and [WEST \(1992\)](#), but giving more details, as their developments are compact, the prior distribution of the number of clusters  $t$  may be written as

$$p(t|M) = w_C(t) M^t \frac{\Gamma(M)}{\Gamma(M+C)}; \quad t = 1, 2, \dots, C,$$

where the factor  $w_C(t)$  does not involve  $M$ ;  $\Gamma(\cdot)$  is the Gamma function. As pointed out by [WEST \(1992\)](#), if one

has samples of  $t$  (see above), the conditional posterior distribution of  $M$  is expressible as

$$p(M|t, \text{ELSE}) \propto (M|t) = \frac{p(t|M)p(M)}{\sum_{t=1}^C p(t|M)p(M)},$$

where  $p(M)$  is the prior density of  $M$ . This result follows, because, if one knows  $\mathbf{p}$ , this gives immediately the number of clusters  $t$  as well as the number of permanent environmental effects per cluster. Then, with  $B(\cdot, \cdot)$  being the Beta function

$$\begin{aligned} p(M|t, \text{ELSE}) &\propto M^t \frac{\Gamma(M)}{\Gamma(M+C)} p(M) \\ &= M^t \frac{(M+C)B(M+1, C)}{M\Gamma(C)} p(M) \\ &\propto M^{t-1} (M+C) p(M) \int_0^1 x^M (1-x)^{C-1} dx. \end{aligned}$$

[WEST \(1992\)](#) notes that this density results from marginalizing the joint distribution  $[M, x|t, \text{ELSE}]$ , where  $x$  is a continuous random variable taking values between 0 and 1. Hence, it follows that

$$p(M, x|t, \text{ELSE}) \propto M^{t-1} (M+C) p(M) x^M (1-x)^{C-1} dx.$$

Consequently, the conditional posterior density of the auxiliary variable  $x$  is

$$p(x|t, M, \text{ELSE}) \propto x^M (1-x)^{C-1} dx, \quad (33)$$

so that  $x|t, M, \text{ELSE} \sim \text{Beta}(M+1, C)$ . In addition, the conditional posterior distribution of  $M$  is

$$p(M|x, t, \text{ELSE}) \propto M^{t-1} (M+C) p(M) x^M,$$

and its form depends on the prior adopted for  $M$ . Since  $M$  must be positive, [WEST \(1992\)](#) suggests a Gamma ( $a, b$ ) distribution as prior, so that

$$\begin{aligned} p(M|x, t, \text{ELSE}) &\propto M^{t-1} (M+C) M^{a-1} \exp(-bM) x^M \\ &\propto M^{t+a-1} \exp[-M(b - \log x)] \\ &\quad + C M^{t+a-2} \exp[-M(b - \log x)]. \end{aligned}$$

The normalized density is

$$\begin{aligned} &p(M|x, t, \text{ELSE}) \\ &= \frac{M^{t+a-1} \exp[-M(b - \log x)] + C M^{t+a-2} \exp[-M(b - \log x)]}{\int_0^\infty \{ M^{t+a-1} \exp[-M(b - \log x)] + C M^{t+a-2} \exp[-M(b - \log x)] \} dM}. \end{aligned}$$

The integrals in the denominator yield



$$\int_0^\infty M^{t+a-1} \exp[-M(b - \log x)] dM = \frac{\Gamma(t+a)}{(b - \log x)^{t+a}}$$

$$\int_0^\infty CM^{t+a-2} \exp[-M(b - \log x)] dM = \frac{C\Gamma(t+a-1)}{(b - \log x)^{t+a-1}}.$$

Hence, letting  $a^* = (t + a)$  and  $b^* = (b - \log x)$  the conditional posterior distribution becomes

$$\begin{aligned} & p(M|x, t, \text{ELSE}) \\ &= \frac{M^{a^*-1} \exp(-Mb^*) + CM^{a^*-1-1} \exp(-Mb^*)}{(\Gamma(a^*)/b^{*a^*}) + (C\Gamma(a^*-1)/b^{*a^*-1})} \\ &= \frac{(\Gamma(a^*)/b^{*a^*})(b^{*a^*}/\Gamma(a^*))M^{a^*-1} \exp(-Mb^*) + C(\Gamma(a^*-1)/b^{*a^*-1})(b^{*a^*-1}/\Gamma(a^*-1))M^{a^*-1-1} \exp(-Mb^*)}{(\Gamma(a^*)/b^{*a^*}) + (C\Gamma(a^*-1)/b^{*a^*-1})} \\ &= \pi_x \text{Gamma}(a^*, b^*) + (1 - \pi_x) \text{Gamma}(a^* - 1, b^*). \end{aligned} \quad (34)$$

This is a mixture of the two Gamma distributions indicated, with mixing probabilities  $\pi_x$  and  $1 - \pi_x$ . Note that

$$\begin{aligned} \pi_x &= \frac{\Gamma(a^*)/b^{*a^*}}{(\Gamma(a^*)/b^{*a^*}) + (C\Gamma(a^*-1)/b^{*a^*-1})} \\ &= \frac{\Gamma(a^*)}{\Gamma(a^*) + Cb^*\Gamma(a^*-1)}. \end{aligned}$$

Since  $\Gamma(a^*) = (a^* - 1)\Gamma(a^* - 1)$ ,

$$\begin{aligned} \pi_x &= \frac{a^* - 1}{a^* - 1 + Cb^*} \\ 1 - \pi_x &= \frac{Cb^*}{a^* - 1 + Cb^*}. \end{aligned}$$

At the end of the MCMC sampling process there will be  $S$  samples of the number of clusters  $t$  and of the auxiliary variables  $x$ . The density of the marginal posterior distribution of  $M$  can be estimated ([WEST 1992](#)) using the Rao–Blackwell estimator

$$\begin{aligned} & p(M|\text{ELSE}) \\ &= \frac{1}{S} \sum_{s=1}^S p(M|x^{(s)}, t^{(s)}, \text{ELSE}) \\ &= \frac{1}{S} \sum_{s=1}^S [\pi_{x^{(s)}} \text{Gamma}(a^{*(s)}, b^{*(s)}) \\ &+ (1 - \pi_{x^{(s)}}) \text{Gamma}(a^{*(s)} - 1, b^{*(s)})], \end{aligned} \quad (35)$$

where  $a^{*(s)} = t^{(s)} + a$ , and  $b^{*(s)} = b - \log x^{(s)}$ . If  $a = 0$  and  $b = 0$ , the Gamma prior degenerates to  $p(M) \propto M^{-1}$  or, equivalently, to  $p[\log(M)] \propto \text{constant}$ . If such an improper prior is adopted, the Rao–Blackwell estimator reduces to

$$p(M|\text{ELSE}) \\ = \frac{1}{S} \sum_{s=1}^S [\pi_{x^{(s)}} \text{Gamma}(t^{(s)}, -\log x^{(s)}) \\ + (1 - \pi_{x^{(s)}}) \text{Gamma}(t^{(s)} - 1, -\log x^{(s)})],$$

with  $a^* = t$  and  $b^* = -\log x$  used in the calculation of  $\pi_{x^{(s)}}$ . While the conditional posterior distribution of  $M$  is well defined for  $a = b = 0$ , this does not guarantee that its marginal posterior distribution will be always proper. Hence, the uniform prior on  $\log(M)$  should be used with caution.

## CONCLUSION

Go to: ☒ Go to: ☒

This study presents parametric and semiparametric procedures for analysis of rank data. Observations presented as ranks arise often in behavioral or cognitive sciences and in the context of competitions such as horse jumping events or greyhound races. For instance, [JOHNSON \*et al.\* \(2002\)](#) carried out a metaanalysis of 30 primate (24 genera represented) intelligence experiments grouped into “paradigms” and “procedures within paradigm.” Paradigms are types of intelligence tests (*e.g.*, “reversal learning,” where individuals learn to pick one object rather than another, and then the value of the objects changes, so that a previously unrewarded object is now rewarded); “procedures” refer to different methodologies used to investigate the paradigm. In one of six reversal studies the observed ranking was Cebus > Hylobates > Aotus > Saimiri. In the analysis of the ranks, they used Thurstonian models similar to those described in this article. Their Bayesian formulation allowed a clear assessment of uncertainty, which was viewed as especially important because, in these experiments, raters usually do not rank >15 items, so asymptotic approximations may be inadequate.

In a study of competitions of jumping horses, [TAVERNIER \(1990\)](#) argued that a “physical” measure of performance (*e.g.*, time, number of obstacles broken or refused) is not always available, and that it is unclear how to measure the technical difficulty of an event. Also, measures such as yearly earnings are available only for horses that earn money, so that the information comes from a selected sample, which may distort inference considerably. As discussed earlier, [TAVERNIER \(1990, 1991\)](#) proposed an approximate Bayesian method in which ranks are viewed as being the manifestation of latent variables, as in the present study. In the analysis, involving ~19,000 horses participating in all jumping competitions that took place in France in 1987, ~14% of the horses had an accuracy <0.14 (“accuracy” ranges between 0 and 1), and 28% did not reach 0.45. These accuracies are approximations and probably understate uncertainty, because the error of estimates of parameters is not taken into account in the calculations. Important numerical problems were encountered by [TAVERNIER \(1990\)](#). For each Newton–Raphson iteration, and for a race with  $n$  horses, she had to calculate the following multivariate normal integrals: (1) an integral of order  $n - 1$ , (2)  $n - 1$  integrals of order  $n - 2$ , and (3)  $(n - 2)!$  integrals of order  $n - 3$ . For this reason, she used approximations to integrals based on Taylor series expansions. None of these integrals need to be computed explicitly or approximated in our procedure, as all calculations rely on Monte Carlo sampling. Another difficulty in the conditional maximum *a posteriori* method of [TAVERNIER \(1990, 1991\)](#) is the computation of the joint posterior mode, which, in the context of horse races in France may involve solving (reiteratively) a system exceeding 100,000 unknowns. In our procedure, solving of equations is replaced by effecting draws from conditional posterior distributions and, under the assumptions of this article, all these are recognizable. A potential difficulty may be slow mixing of the MCMC algorithm, but this can often be accelerated by either changing the algorithm or adopting a different parameterization. Introducing a Dirichlet process prior enhances the level of difficulty of the calculations, but it confers robustness with respect to possible departures from the normality assumption for the distributions of events or of permanent environmental effects. Further, the nonparametric treatment of permanent environmental effects is a flexible alternative to a thick-tailed

distribution ([STRANDÉN and GIANOLA 1998, 1999](#); [ROSA et al. 2001, 2003, 2004](#)), in the sense that it allows for asymmetry, as well as for allocation of effects into clusters with some biological meaning.

Here, it was assumed that event effects were equally distributed *a priori*, but there may be heterogeneity with respect to level of difficulty. This can be accommodated via hierarchical modeling. For instance, the prior distribution  $N_K(\mathbf{g}|\mathbf{0}, \mathbf{I}\sigma_g^2)$  could be modified into  $N_K(\mathbf{g}|\mathbf{M}\boldsymbol{\alpha}, \mathbf{I}\sigma_g^2)$ , where  $\boldsymbol{\alpha}$  is a vector of unknown “levels of difficulty” effects, and  $\mathbf{M}$  is a known matrix relating  $\mathbf{g}$  to  $\boldsymbol{\alpha}$ ; in turn,  $\boldsymbol{\alpha}$  could be assigned a vague prior. In their primate studies, [JOHNSON et al. \(2002\)](#) fitted a latent variable model including a genus effect and a “paradigm-genus” deviation; they found that the latter did not contribute to variability appreciably and concluded that taxa differed in domain-general ability. A nonparametric alternative would be to assign a Dirichlet process prior to the event effects  $\mathbf{g}$  and let the data drive to clusters representing level of difficulty.

In the context of horse breeding, it may be that there are individuals whose genes predispose them to perform well in certain types of competition, but not in others; *i.e.*, there may be different sets of genes for different types of competition. Our model can be extended by assuming that each individual has a liability peculiar to each type of competition. Suppose there are three types of competition; here, there would be three liabilities per individual with a within-individual correlation matrix  $\mathbf{R}_0$ , say. All three liabilities would be realized only if the individual in question participates in the three types of events. A simpler model could postulate that liabilities for different events are independent, but that each individual possesses an additive genetic effect peculiar to the type of event. An additive genetic variance–covariance matrix  $\mathbf{G}_0$  would be introduced and inferred as in a multivariate Bayesian mixed-effects model ([SORENSEN and GIANOLA 2002](#)).

In many sports events, *e.g.*, races or marathons, there are no ties. However, ties can be handled in our procedure via a simple modification of the limits the liability variates can take. For example, consider a competition with four participants (5, 7, 14, and 20), such that the observed ranking is 5<sub>[1]</sub>, 7<sub>[2.5]</sub>, 14<sub>[2.5]</sub>, 20<sub>[4]</sub>; here, [2.5] denotes a tie in second place between individuals 7 and 14. In the process of sampling the liabilities, it suffices to accept samples such that 5 and 20 take the largest and smallest values, respectively, with the liabilities 7 and 14 taking any order between those of 5 and 20; then, the ranking of 7 and 14 would vary at random over MCMC samples. Alternatively, one could assign the average of the two liabilities sampled for 7 and 14 in each of the MCMC iterations. [JOHNSON et al. \(2002\)](#) handled ties via a parametric model for the probability that two genera are tied, with one additional parameter (with a corresponding prior distribution) introduced in the implementation.

A useful extension of this model, at least in the context of evaluation of horse races, would be one for a joint analysis of ranks and of some “continuous” variate, such as time in the event or earnings. As mentioned above, there are situations in which the rank information is available for all competitors, but earnings are available only for the placed competitors. Clearly, earnings for the other competitors are not missing completely at random. However, if the conditional probability of missing earning information depends on the ranks, then missingness would be ignorable in a joint analysis of ranks and of earnings ([RUBIN 1976](#); [IM et al. 1989](#); [SORENSEN et al. 2001](#)). For example, such a model could postulate a joint normal distribution of the liabilities and of the continuous trait measured on some competitors.

## Acknowledgments

Go to: ☒ Go to: ☒

Yu-Mei Chang and Xiao-Lin Wu are thanked for helpful comments. Research was completed while the senior author was Mercator Professor (Deutsche Forschungsgemeinschaft) at Georg-August-University, Göttingen, Germany. Support by the Wisconsin Agriculture Experiment Station and by grants National Research Initiatives Competitive Grants Program/U.S. Department of Agriculture 2003-35205-12833, National Science Foundation



## References

- Antoniak, C. E., 1974. Mixtures of Dirichlet processes with applications to nonparametric problems. *Ann. Stat.* 2: 1152–1174.
- Bock, R. D., and L. V. Jones, 1968. *The Measurement and Prediction of Judgment and Choice*. Holden–Day, San Francisco.
- Bush, C. A., and S. N. MacEachern, 1996. A semiparametric Bayesian model for randomized block designs. *Biometrika* 83: 275–285.
- Cowles, M. K., and B. P. Carlin, 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Stat. Assoc.* 91: 883–904.
- Dempster, E. R., and I. M. Lerner, 1950. Heritability of threshold characters. *Genetics* 35: 212–236.  
[\[PMC free article\]](#) [\[PubMed\]](#)
- Devroye, L., 1986. *Non-Uniform Random Variate Generation*. Springer, New York.
- Escobar, M. D., 1994. Estimating normal means with a Dirichlet process prior. *J. Am. Stat. Assoc.* 89: 268–275.
- Falconer, D. S., 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* 29: 51–76.
- Ferguson, T. S., 1973. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* 1: 209–230.
- Foulley, J. L., S. Im, D. Gianola and I. Hoeschele, 1987. Empirical Bayes estimation of genetic value for n binary traits. *Genet. Sel. Evol.* 19: 197–224. [\[PMC free article\]](#) [\[PubMed\]](#)
- Garcia-Cortés, L. A., and D. Sorensen, 1996. On a multivariate implementation of the Gibbs sampler. *Genet. Sel. Evol.* 28: 121–126.
- Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin, 2003. *Bayesian Data Analysis*, Ed. 2. Chapman & Hall, London.
- Gianola, D., 1982. Theory and analysis of threshold characters. *J. Anim. Sci.* 54: 1079–1096.
- Gianola, D., and R. L. Fernando, 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63: 217–244.
- Gianola, D., and J. L. Foulley, 1983. Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15: 201–223. [\[PMC free article\]](#) [\[PubMed\]](#)
- Gianola, D., J. L. Foulley and R. L. Fernando, 1986. Prediction of breeding values when variances are not known. *Genet. Sel. Evol.* 18: 485–498. [\[PMC free article\]](#) [\[PubMed\]](#)
- Gilks, W. R., S. Richardson and D. J. Spiegelhalter, 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Grüneberg, H., 1952. Genetical studies on the skeleton of the mouse. IV. Quasi-continuous variations. *J. Genet.* 51: 95–114.
- Harville, D. A., 1974. Bayesian inference of variance components using only error contrasts. *Biometrika* 61: 383–385.
- Henderson, C. R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32: 69–83.
- Henery, R. J., 1981. Permutation probabilities as models for horse races. *J. R. Stat. Soc. B* 43: 86–91.
- Im, S., R. L. Fernando and D. Gianola, 1989. Likelihood inferences in animal breeding under selection: a missing data theory viewpoint. *Genet. Sel. Evol.* 21: 399–414.
- Johnson, V. E., R. O. Deaner and C. P. Van Schaik, 2002. Bayesian analysis of multi-study rank data with application to primate intelligence ratings. *J. Am. Stat. Assoc.* 97: 8–17.
- Kleinman, K. P., and J. G. Ibrahim, 1998. A semiparametric Bayesian approach to the random effects model.



- Biometrics 54: 921–938. [[PubMed](#)]
- Lawrence, S., R. Beasley, I. Doull, B. Begishvili, F. Lampe *et al.*, 1994. Genetic analysis of atopy and asthma as quantitative traits and ordered polychotomies. *Ann. Hum. Genet.* 58: 359–368. [[PubMed](#)]
- MacEachern, S. N., 1994. Estimation of normal means with a conjugate style Dirichlet process prior. *Comm. Stat. Sim.* 23: 727–741.
- Robert, C. P., and G. Casella, 2005. *Monte Carlo Statistical Methods*, Ed. 2. Springer, New York.
- Rosa, G. J. M., D. Gianola and J. I. Urioste, 2001. Assessing relationships between genetic evaluations using robust regression with an application to Holsteins in Uruguay. *Acta Agric. Scand. Sect. A* 51: 21–34.
- Rosa, G. J. M., C. R. Padovani and D. Gianola, 2003. Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biom. J.* 45: 573–590.
- Rosa, G. J. M., D. Gianola and C. R. Padovani, 2004. Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *J. Appl. Stat.* 31: 855–873.
- Rubin, D. B., 1976. Inference and missing data. *Biometrika* 63: 581–582.
- Sorensen, D., and D. Gianola, 2002. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer–Verlag, New York.
- Sorensen, D., S. Andersen, D. Gianola and I. Korsgaard, 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27: 229–249.
- Sorensen, D. A., R. L. Fernando and D. Gianola, 2001. Inferring the trajectory of genetic variance in the course of artificial selection. *Genet. Res.* 77: 83–94. [[PubMed](#)]
- Strandén, I., and D. Gianola, 1998. Attenuating effects of preferential treatment with Student-t mixed linear models: a simulation study. *Genet. Sel. Evol.* 30: 565–583.
- Strandén, I., and D. Gianola, 1999. Mixed effects linear models with t-distributions for quantitative genetic analysis: a Bayesian approach. *Genet. Sel. Evol.* 31: 25–42.
- Tanner, M. A., and W. Wong, 1987. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* 82: 528–550.
- Tavernier, A., 1990. Estimation of breeding value of jumping horses from their ranks. *Livest. Prod. Sci.* 26: 277–290.
- Tavernier, A., 1991. Genetic evaluation of horses based on ranks in competitions. *Genet. Sel. Evol.* 23: 159–173.
- Thurstone, L. L., 1927. A law of comparative judgement. *Psych. Rev.* 34: 278–286.
- Van Der Merwe, A. J., and A. L. Pretorius, 2003. Bayesian estimation in animal breeding using the Dirichlet process prior for correlated random effects. *Genet. Sel. Evol.* 35: 137–158. [[PMC free article](#)] [[PubMed](#)]
- Wang, C. S., J. J. Rutledge and D. Gianola, 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.* 25: 41–62.
- Wang, C. S., J. J. Rutledge and D. Gianola, 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* 26: 91–115.
- West, M., 1992. Hyperparameter estimation in Dirichlet process mixture models. Technical Report 92-A03. ISDS, Duke University, Durham, NC.
- Wright, S., 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19: 506–536. [[PMC free article](#)] [[PubMed](#)]