HAL
open science

A Bayesian analysis of mixed survival models

Vincent Ducrocq, G. Casella

To cite this version:

Vincent Ducrocq, G. Casella. A Bayesian analysis of mixed survival models. Genetics Selection Evolution, 1996, 28 (6), pp.505-529. hal-02686854

HAL Id: hal-02686854

https://hal.inrae.fr/hal-02686854v1

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genet Sel Evol (1996) 28, 505-529
© Elsevier/INRA

Original article

A Bayesian analysis of mixed survival models†

V Ducrocq [1*], G Casella [2]

[1] *Department of Animal Science, Cornell University;*
[2] *Biometrics Unit, Cornell University, Ithaca, NY 14852, USA*

(Received 13 August 1996; accepted 1 October 1996)

Summary – In proportional hazards models, the hazard of an animal $\lambda(t)$, ie, its probability of dying or being culled at time t given it is alive prior to t, is described as $\lambda(t) = \lambda_0(t)e^{\mathbf{w}'\boldsymbol{\theta}}$ where $\lambda_0(t)$ is a 'baseline' hazard function and $e^{\mathbf{w}'\boldsymbol{\theta}}$ represents the effect of covariates $\mathbf{w}$ on culling rate. A distribution can be attached to elements s_q in $\boldsymbol{\theta}$, identifying, for example, genetic effects and leading to mixed survival models, also called 'frailty' models. To estimate the parameters τ of the distribution of frailty terms, a Bayesian analysis is proposed. Inferences are drawn from the marginal posterior density $\pi(\tau)$ which can be derived from the joint posterior density via Laplacian integration, a powerful technique related to saddlepoint approximations. The validity of this technique is shown here on simulated examples by comparing the resulting approximate $\pi(\tau)$ to the one obtained by algebraic integration. This exact calculation is feasible in very specific cases only, whereas the saddlepoint approximation can be applied to situations where $\lambda_0(t)$ is arbitrary (Cox models) or parametric (eg, Weibull), where the frailty terms are correlated through a known relationship matrix, or in more general models with stratification and/or time-dependent covariates. The influence of the censoring rate and the data structure is also illustrated.

survival analysis / mixed model / variance component estimation / Bayesian analysis / proportional hazards model

Résumé – Une analyse bayésienne des modèles de survie mixtes. *Dans le cas des modèles à risques proportionnels, la fonction de risque d'un animal* $\lambda(\mathrm{t})$, *c'est-à-dire sa probabilité de mourir ou d'être réformé au temps* t *sachant qu'il est vivant juste avant* t, *a la forme* $\lambda(\mathrm{t}) = \lambda_0(\mathrm{t})\mathrm{e}^{\mathrm{w}'\theta}$ *où* $\lambda_0(\mathrm{t})$ *est une fonction de risque « de base » et* $\mathrm{e}^{w'\theta}$ *représente l'effet des covariables* w *sur le taux de réforme. Une distribution peut être associée aux termes* $\mathrm{s_q}$ *de* θ, *identifiant, par exemple, des effets génétiques et conduisant à des modèles*

---

* Correspondence and reprints: Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex, France.

[1] On leave from the Station de génétique quantitative et appliquée, Institut national de la recherche agronomique.

† Research supported by NSF Grant No DMS-9625440. This is paper BU-1346-M in the Biometrics Unit, Cornell University, Ithaca, NY 14853, USA.

de survie mixtes, aussi appelés modèles de fragilité. Pour l'estimation des paramètres τ de la distribution des termes aléatoires, une analyse bayésienne est proposée. Les inférences statistiques sont faites à partir de la densité marginale a posteriori $\pi(\tau)$ qui peut être obtenue à partir de la distribution conjointe a posteriori par intégration laplacienne, une technique liée aux approximations point-selles. La validité de cette technique est démontrée ici à partir d'exemples simulés, en comparant les résultats de l'approximation de $\pi(\tau)$ avec ceux obtenus après intégration algébrique. Cette dernière correspond à un calcul exact réalisable uniquement dans des cas très particuliers, alors que l'approximation point-selle peut être appliquée dans des situations où $\lambda_0(t)$ est complètement arbitraire (modèles de Cox) ou paramétrique (par exemple, de type Weibull), où les termes aléatoires sont corrélés à travers une matrice de parenté connue, ou avec des modèles plus généraux avec stratification et/ou covariables dépendantes du temps. L'influence du taux de censure et de la structure des données est aussi illustrée.

analyse de données de survie / modèles mixtes / estimation des composantes de variance / analyse bayésienne / modèle à risques proportionnels

INTRODUCTION

Traits associated with longer productive life of livestock are receiving increasing attention in the animal breeding field: it is recognized that decreasing culling due to the involuntary causes (eg, related to disease, infertility, lameness, etc) by genetic or non-genetic means has a positive effect on economic performance, mainly through decreased replacement costs (van Arendonk, 1986; Strandberg, 1991, Strandberg, 1995, Strandberg and Sölkner, 1996). Huge field data sets are usually available for comprehensive analyses of productive life, for example, as a by-product of the dairy recording schemes in dairy cattle. The obvious methodology of choice for such studies is survival analysis, in which proper techniques to deal with the unavoidable presence of censored data have been developed. However, statistical complexity and computational difficulties related to these methods have delayed the adoption of state-of-the-art methodology and different indirect approaches have been proposed (see Strandberg and Sölkner (1996) for a review). Some large-scale applications (Smith, 1983; Smith and Quaas, 1984; Ducrocq, 1987; Ducrocq et al, 1988a, b; Ruiz, 1991; Fournet, 1992; Egger-Danner, 1993; Ducrocq, 1994) as well as the availability of a software specifically written with animal breeding applications in mind (Ducrocq and Sölkner, 1994) have demonstrated that the use of less appropriate approaches can be avoided.

The most popular class of survival models is the class of proportional hazards models (Cox, 1972; Kalbfleisch and Prentice, 1980; Lawless, 1982; Cox and Oakes, 1984). The hazard of an animal (or in the animal breeding context, its risk of being culled) at time t is described as the product of a baseline hazard function $\lambda_0(t)$, which is either left completely arbitrary (Cox model) or has a parametric form (eg, exponential, Weibull or gamma) and of a positive term which is an exponential function of a vector of covariates $\mathbf{w}'$ multiplied by a vector of regression parameters $\boldsymbol{\theta}$.

Proportional hazard models can be extended to include random (eg, genetic) effects, as in the regular mixed linear models that are used for genetic evaluations worldwide. Mixed survival models are classically referred to as 'frailty' models by statisticians. The 'frailty' term v is defined as an unobserved random quantity which

affects multiplicatively the hazard of individuals or groups of animals. When a term v_m is defined for each animal m ($\lambda_m(t, \mathbf{w}) = v_m \lambda(t, \mathbf{w})$), the frailty component extracts part of the unobserved variation between individuals (Vaupel et al, 1979; Hougaard, 1986a,b; Follmann and Goldberg, 1988; Aalen, 1994) and therefore allows for a correction of the possible discrepancy between the true variance of the observations and the one specified by the model. Such an extra variation is referred to as 'overdispersion' (Louis, 1991; Tempelman and Gianola, 1994). When v_q is defined for a group of individuals, eg, all daughters of a sire q, it describes the shared unobservable (genetic, in this case) characteristics which act on the hazard of each member of the group (Clayton and Cuzick, 1985; Anderson et al, 1992; Klein, 1992; Klein et al, 1992). In all cases, the simple transformation $s = \log v$ allows the inclusion of the frailty term in the linear term $\mathbf{w}'\boldsymbol{\theta}$.

Traditionally, a gamma (Clayton and Cuzick, 1985; Ducrocq, 1987; Klein, 1992) distribution has been attached to the frailty term v because of its flexibility and mathematical convenience. Other distributions have also been proposed, eg, a positive stable distribution or an inverse Gaussian distribution (Hougaard, 1986a,b; Klein et al, 1992). Unfortunately, in all cases, they do not have the theoretical appeal of the (multivariate) normal distribution commonly used in animal breeding when a infinitesimal polygenic model is assumed. However, it has been shown that the estimates obtained for the parameters of the gamma distribution of v were relatively large, at least in dairy cattle, which means that v had an approximate log-normal distribution, ie, s was approximately normally distributed (Ducrocq, 1987; Ducrocq et al, 1988b; Ducrocq, 1994). Therefore, it has been suggested to account for the genetic relationship between animals by assuming a multivariate normal distribution for s, the logarithm of the frailty term v (Ducrocq, 1987; Korsgaard, 1996).

Several approaches have been used to estimate the parameters of the frailty distributions. Klein (1992) and Klein et al (1992) suggested the use of an EM algorithm (Dempster et al, 1977), with iterative estimation of v, $\boldsymbol{\theta}$ and the baseline cumulative hazard distribution for a Cox model, followed by the estimation of the frailty distribution given $\hat{v}$. When a Weibull model is combined with a gamma frailty term, Follmann and Goldberg (1988) showed that the frailty term can be algebraically integrated out from the likelihood function. The same property has been used in a Bayesian context (Ducrocq, 1987; Ducrocq et al, 1988b; Fournet, 1992; Ducrocq, 1994). Monte-Carlo techniques have also been suggested in order to obtain the marginal posterior distributions of the hyperparameters (Clayton, 1991; Dellaportas and Smith, 1993; Korsgaard, 1996) but their use on large data sets with complex models (eg, with time-dependent covariates) may be very tedious.

The objective of this paper is to present a general Bayesian approach to the analysis of mixed survival models, with (but without being restricted to) typical animal breeding situations in mind. The framework will be presented for a simple Weibull model with two types of priors for the frailty term (gamma or log-normal). Straightforward generalization to other models (with stratification and time-dependent covariates, Cox models) will follow. A particular strategy for estimation of the hyperparameters suitable for large applications, complex models and situations where a relationship matrix is used will be presented and its performance will be studied on simulated data.

METHODS

In the Weibull regression case, the baseline hazard function has the Weibull form $\lambda_0(t) = \lambda\rho(\lambda t)^{\rho-1}$. For the time being, we will assume that all covariates are time-independent and that only one baseline is defined (no stratification). The vector $\boldsymbol{\theta}$ includes fixed and random effects. For clarity, and unless specified otherwise, only one random effect in the model, eg, a sire effect $\mathbf{s}$ is considered here. Using the classical linear mixed–model notation:

$$\mathbf{w}'_m = (\mathbf{x}'_m \ \mathbf{z}'_m) \quad \text{and} \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{s} \end{pmatrix}$$

where $\boldsymbol{\beta}$ is the vector of fixed effects.

The hazard function $\lambda(t)$ for animal m is:

$$\begin{aligned} \lambda(t) \mid \boldsymbol{\theta}, \rho) &= \lambda_0(t)\exp\{\mathbf{w}'_m\boldsymbol{\theta}\} \\ &= \lambda_\rho(\lambda t)^{\rho-1}\exp\{\mathbf{w}'_m\boldsymbol{\theta}\} \\ &= \rho t^{\rho-1}\exp\{\rho \log \lambda + \mathbf{w}'_m\boldsymbol{\theta}\} \end{aligned} \qquad [1]$$

and $\rho \log \lambda$ can be incorporated in a grand mean (or any factor) in $\mathbf{w}'_m\boldsymbol{\theta}$.

For simplicity, we will write from now on:

$$\lambda(t \mid \boldsymbol{\theta}, \rho) = \rho t^{\rho-1}\exp\{\mathbf{w}'_m\boldsymbol{\theta}\} \qquad [2]$$

using the same notation but keeping in mind that a component of $\mathbf{w}'_m\boldsymbol{\theta}$ (representing an intercept) now includes $\rho \log \lambda$.

If the record comes from a daughter m of sire q, with observed failure at T_m:

$$\lambda(t \mid \boldsymbol{\theta}) = \rho t^{\rho-1} v_q \exp\{\mathbf{x}'_m\boldsymbol{\beta}\} \qquad \text{for } t \leqslant T_m \qquad [3]$$

Here, $v_q = e^{s_q}$ is the frailty term. The usual relationship $f(t) = \lambda(t)S(t)$ where $S(t) = \int_0^t \lambda(u)\,du$ can be used to show that [3] is a particular case of a log-linear model of the form (Kalbfleisch and Prentice, 1980):

$$Y_m = \log(T_m) = \frac{1}{\rho}\mathbf{x}'_m\boldsymbol{\beta} + \frac{1}{\rho}s_q + \frac{1}{\rho}\omega_m \qquad [4]$$

$$= \mathbf{w}'_m\boldsymbol{\beta}^\star + \frac{1}{\rho}s^\star_q + \frac{1}{\rho}\omega_m \qquad [5]$$

where ω_m follows an extreme value distribution (Kalbfleisch and Prentice, 1980; Lawless, 1982) whose variance is equal to $\pi^2/6$. Note that here ω_m implicitly includes three-quarters of the additive genetic variance. With this presentation, a natural definition of the heritability of the survival trait on the logarithmic scale is:

$$h^2 = \frac{4\ \mathrm{Var}(s^\star)}{\mathrm{Var}(\log T)} = \frac{4\ \mathrm{Var}(s)}{\frac{\pi^2}{6} + \mathrm{Var}(s)} \qquad [6]$$

Formula [6] solves the problem of a proper definition of heritability for survival traits indicated in Ducrocq (1987) and Ducrocq et al (1988b).

Prior distributions

Gamma frailty model

Assume:

$$v_q \sim \text{gamma}(\gamma\,,\gamma)$$

ie,

$$s_q \sim \text{(generalized) log-gamma}(\gamma\,,\gamma)$$

The log-gamma distribution (Bartlett and Kendall, 1946, according to Lawless (1982), p 21) corresponds to the distribution of $\log x$ when x follows a gamma distribution. Note however that the suffix 'log-' (eg, in 'log-normal') is often given to the distribution of x when $\log x$ has a known form (eg, normal). Again, the choice of this prior distribution is mainly related to its flexibility and mathematical convenience (see also Klein, 1992, and Klein et al, 1992). Then:

$$\pi_0(v_q \mid \gamma) = \frac{\gamma^\gamma}{\Gamma(\gamma)}\, v_q^{\gamma-1} \quad \exp\{-\gamma v_q\} \qquad [7]$$

$$\text{and} \qquad \pi_0(\mathbf{s} \mid \gamma) = \prod_{q=1}^{N_q} \left[\frac{\gamma^\gamma}{\Gamma(\gamma)} \exp\left\{\gamma(s_q - e^{s_q})\right\}\right] \qquad [8]$$

Log-normal frailty model

In quantitative genetics, due to the infinitesimal polygenic model usually assumed, it is more natural to consider the following prior distribution for the frailty term:

$$v_q \sim \log - \text{normal}(0\,,\sigma_s^2)$$

ie,

$$s_q \sim N(0\,,\sigma_s^2)$$

and if sires are related:

$$\mathbf{s} \sim \text{MVN}(\mathbf{0}, \mathbf{A}\sigma_s^2)$$

where $\mathbf{A}$ is the relationship matrix between sires, we have

$$\pi_0(\mathbf{s} \mid \sigma_s^2) = \frac{1}{(2\pi\sigma_s^2)^{N_q/2}|\mathbf{A}|^{1/2}} \exp\left(-\frac{1}{2\sigma_s^2}\mathbf{s}'\mathbf{A}^{-1}\mathbf{s}\right) \qquad [9]$$

Hyperparameters

In order to simultaneously consider the two previous cases, we will denote the dispersion parameter of the random effect distribution by τ (with $\tau = \gamma$ or $\tau = \sigma_s^2$) and we will assume a flat prior for τ as well as for $\boldsymbol{\beta}$ and ρ:

$$\pi_0(\boldsymbol{\beta}, \rho) = 1 \quad \text{and} \quad \pi_0(\tau) = 1 \qquad [10]$$

Likelihood construction

Conditionally on $\boldsymbol{\theta}$ and ρ, the contribution to the likelihood of animal m which fails ($\delta_m = 1$) or is censored ($\delta_m = 0$) at time y_m, is:

$$\begin{aligned}\mathcal{L}_m &= \mathcal{L}_m(y_m \mid \boldsymbol{\theta}, \rho) \\ &= [\lambda(y_m \mid \boldsymbol{\theta}, \rho)]^{\delta_m} \times S(y_m \mid \boldsymbol{\theta}, \rho)\end{aligned} \qquad [11]$$

where $S(t)$ is the survivor function at time t. For the Weibull model, these two components are:

$$\begin{aligned}\lambda(y_m \mid \boldsymbol{\theta}, \rho) &= \rho\, y_m^{\rho-1}\, e^{\mathbf{w}'_m \boldsymbol{\theta}} \\ S(y_m \mid \boldsymbol{\theta}, \rho) &= \exp\left(-\int_0^{y_m} \lambda(t \mid \boldsymbol{\theta}, \rho)\, dt\right) \\ &= \exp\left(-y_m^{\rho}\, e^{\mathbf{w}'_m \boldsymbol{\theta}}\right)\end{aligned} \qquad [13]$$

Combining all these contributions (for $m = 1, \ldots, N$) which are conditionally independent, we obtain:

$$\mathcal{L}(\boldsymbol{\theta}, \rho \mid \mathbf{y}) = \prod_m \mathcal{L}_m = \left\{ \rho^N \left[\prod_{\{unc\}} y_m \right]^{\rho-1} \exp\left(\sum_{\{unc\}} \mathbf{w}'_m \boldsymbol{\theta} \right) \right\} \times \exp\left(- \sum_{\{unc, cens\}} y_m^{\rho} e^{\mathbf{w}'_m \boldsymbol{\theta}} \right) \qquad [14]$$

where $\{unc\}$ and $\{cens\}$ represent the sets of indices m corresponding to uncensored and censored records, respectively.

$$\log \mathcal{L}(\boldsymbol{\theta}, \rho \mid \mathbf{y}) = \left(N \log \rho + (\rho - 1) \sum_{\{unc\}} \log y_m + \sum_{\{unc\}} \mathbf{w}'_m \boldsymbol{\theta} \right) - \sum_{\{unc,\, cens\}} y_m^{\rho} e^{\mathbf{w}'_m \boldsymbol{\theta}} \qquad [15]$$

Joint posterior density

Applying Bayes' theorem, we obtain:

$$p(\boldsymbol{\theta}, \rho, \tau \mid \mathbf{y}) \propto \mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta}, \rho) \times \pi_0(\mathbf{s} \mid \tau) \times \pi_0(\boldsymbol{\beta}, \rho) \times \pi_0(\tau) \qquad [16]$$

and taking the logarithm on both sides:

$$\log p(\boldsymbol{\theta}, \rho, \tau \mid \mathbf{y}) = \text{constant} + \sum_m \log \mathcal{L}_m(y_m \mid \boldsymbol{\theta}, \rho) + \log \pi_0(\mathbf{s} \mid \tau) \qquad [17]$$

Inference on θ and ρ

If we assume that τ is known, the logarithm of the joint posterior density of $\Theta = \begin{pmatrix} \boldsymbol{\theta} \\ \rho \end{pmatrix}$ given τ is:

$$\log p(\Theta \mid \mathbf{y}, \tau) = \left(N \log \rho + (\rho - 1) \sum_{\{unc\}} \log y_m + \sum_{\{unc\}} \mathbf{w}'_m \boldsymbol{\theta} \right) - \sum_{\{unc,\, cens\}} y_m^{\rho} e^{\mathbf{w}'_m \boldsymbol{\theta}} + \log \pi_0(\mathbf{s} \mid \tau) \qquad [18]$$

Using the same notation as in Tempelman and Gianola (1993), let $\tilde{\Theta}_\tau$ be the mode of this joint posterior density:

$$\tilde{\Theta}_\tau = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_\tau \\ \tilde{\rho}_\tau \end{pmatrix} = \mathrm{Arg}_\Theta \max p(\Theta \mid \mathbf{y}, \tau) \qquad [19]$$

At the mode, the gradient vector is null:

$$\nabla p_\tau(\tilde{\Theta}) = \left(\frac{\partial \log p(\Theta \mid \mathbf{y}, \tau)}{\partial \Theta} \right)_{\Theta = \tilde{\Theta}_\tau} = \mathbf{0} \qquad [20]$$

For latter use, we also need to define the negative Hessian matrix:

$$\tilde{\mathbf{H}}_\tau = \left(-\frac{\partial^2 \log p(\Theta \mid \mathbf{y}, \tau)}{\partial \Theta \, \partial \Theta'} \right) \qquad [21]$$

Joint inference on β, ρ and τ

Consider here the particular case of the gamma frailty model, where the random effect $\mathbf{s}$ has a log-gamma distribution ($\tau = \gamma$; this implies that the genetic relationship between sires is ignored). Then the marginal posterior density of $\boldsymbol{\beta}$, ρ and τ is obtained by integrating out $\mathbf{s}$ from the joint posterior density $p(\boldsymbol{\theta}, \rho, \tau \mid \mathbf{y}) = p(\boldsymbol{\theta}, \rho, \gamma \mid \mathbf{y})$:

$$p(\boldsymbol{\beta}, \rho, \tau \mid \mathbf{y}) = \int_{-\infty}^{\infty} p(\boldsymbol{\theta}, \rho, \tau \mid \mathbf{y}) \, ds \propto \int_{-\infty}^{\infty} \mathcal{L}(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{s}, \rho) \times \pi_0(\mathbf{s} \mid \rho) \times \pi_0(\tau) \, ds \qquad [22]$$

Grouping the contributions to the likelihood of all daughters of each sire q:

$$p(\boldsymbol{\beta}, \rho, \gamma \mid \mathbf{y}) = \int_{-\infty}^{\infty} \prod_{q=1}^{N_q} \rho^{n_q} \left[\prod_{\{unc,q\}} y_m \right]^{\rho - 1} \exp\left(\sum_{\{unc,q\}} \mathbf{w}'_m \boldsymbol{\theta} \right) \exp\left(- \sum_{\{unc,cens,q\}} y_m^{\rho} \, e^{\mathbf{w}'_m \boldsymbol{\theta}} \right) \pi_0(s_q \mid \gamma) \, ds_q \qquad [23]$$

where now $\{unc, q\}$ and $\{cens, q\}$ are the sets of indices m of the n_q uncensored and the censored daughters of sire q, respectively.

Writing $e^{\mathbf{w}'_m \boldsymbol{\theta}} = e^{\mathbf{x}'_m \boldsymbol{\beta}} e^{s_q}$ for all daughters of sire q, one can factor out the terms which do not depend on s_q, which leads to:

$$p(\boldsymbol{\beta}, \rho, \gamma \mid \mathbf{y}) = \prod_{q=1}^{N_q} \alpha_q \int_{-\infty}^{\infty} \{\exp(n_q\, s_q - \Omega_q\, e^{s_q})\} \left\{ \frac{\gamma^\gamma}{\Gamma(\gamma)} \exp(\gamma(s_q - e^{s_q})) \right\} ds_q \qquad [24]$$

with:

$$\alpha_q = \rho^{n_q} \left[\prod_{\{unc,q\}} y_m \right]^{\rho-1} \exp\left(\sum_{\{unc,q\}} \mathbf{x}'_m \boldsymbol{\beta} \right) \qquad [25]$$

and:

$$\Omega_q = \sum_{\{unc,cens,q\}} y_m^\rho e^{\mathbf{x}'_m \boldsymbol{\beta}} \qquad [26]$$

Each of these products, for $q = 1, \ldots N_q$, is of the form:

$$\alpha_q \frac{\gamma^\gamma}{\Gamma(\gamma)} \int_{-\infty}^{\infty} \exp((n_q + \gamma) s_q - (\Omega_q + \gamma) e^{s_q})\, ds_q \qquad [27]$$

The term under the integral can be recognized as the kernel of a log-gamma distribution with parameters $(n_q + \gamma)$ and $(\Omega_q + \gamma)$. Therefore,

$$\int_{-\infty}^{\infty} \frac{(\Omega_q + \gamma)^{n_q + \gamma}}{\Gamma(n_q + \gamma)} \exp((n_q + \gamma) s_q - (\Omega_q + \gamma) e^{s_q}) ds_q = 1$$

Hence, the integration of the random effects s_q out of the joint posterior density can be done algebraically:

$$p(\boldsymbol{\beta}, \rho, \gamma \mid \mathbf{y}) = \left[\frac{\gamma^\gamma}{\Gamma(\gamma)} \right]^{N_q} \prod_{q=1}^{N_q} \left\{ \frac{\alpha_q\, \Gamma(n_q + \gamma)}{(\Omega_q + \gamma)^{n_q + \gamma}} \right\} \qquad [28]$$

or:

$$\log p(\boldsymbol{\beta}, \rho, \gamma \mid \mathbf{y}) = N_q[\gamma \log \gamma - \log(\Gamma(\gamma))] + \sum_{q=1}^{N_q} \{\log \alpha_q + \log \Gamma(n_q + \gamma) - (n_q + \gamma) \log(\Omega_q + \gamma)\} \qquad [29]$$

Expressions [28] and [29] are essentially those used in Ducrocq (1987), Ducrocq et al (1988b) and Ducrocq (1994) for the estimation of the sire variance of the length of productive life of dairy cows. Follmann and Goldberg (1988) referred to the distribution in [28] as a multivariate Burr distribution. Again, $\boldsymbol{\beta}, \rho$ and γ can

be estimated as the mode of this posterior distribution:

$$\begin{pmatrix} \breve{\boldsymbol{\beta}} \\ \breve{\rho} \\ \breve{\gamma} \end{pmatrix} = \mathrm{Arg}_{(\boldsymbol{\beta},\rho,\gamma)} \max \log p(\boldsymbol{\beta}, \rho, \gamma \mid \mathbf{y}) \qquad [30]$$

with associated negative Hessian matrix $\breve{\mathbf{H}}$.

Inference on τ

Inferences on the dispersion parameter τ should be based on its marginal posterior distribution, after integrating out the nuisance parameters $\boldsymbol{\theta}$ and ρ (Berger, 1985; Robert, 1992):

$$p(\tau \mid \mathbf{y}) = \int\int p(\boldsymbol{\theta}, \rho, \tau \mid \mathbf{y})\, d\boldsymbol{\theta}\, d\rho \qquad [31]$$

or:

$$p(\tau \mid \mathbf{y}) = \int\int \exp\{\log p(\boldsymbol{\theta}, \rho, \tau \mid \mathbf{y})\}\, d\boldsymbol{\theta}\, d\rho \qquad [32]$$

Except in trivial cases, this integration cannot be performed algebraically. To obtain the marginal posterior distribution of the dispersion parameter τ, one can either simulate random samples from it (Clayton, 1991; Dellaportas and Smith, 1993; Korsgaard, 1996), compute the integral numerically (Smith et al, 1985) or find an approximation. We will choose the third alternative, using a technique known as Laplacian integration (Tierney and Kardane, 1986; Achcar and Bolfarine, 1986; Tierney and Kardane, 1986; Tierney et al, 1989; Tempelman and Gianola, 1993; Goutis and Casella, 1996). For any given value $\tau^\star$ of τ, we want to approximate:

$$p(\tau^\star \mid \mathbf{y}) = \int\int \exp\{\log p(\boldsymbol{\theta}, \rho \mid \mathbf{y}, \tau^\star)\}\, d\boldsymbol{\theta}\, d\rho \qquad [33]$$

Intuitively, if $p(\boldsymbol{\theta}, \rho \mid \mathbf{y}, \tau) = p(\Theta_{\tau^\star})$ is unimodal, the value of the integral will heavily depend on the value of the density at its mode $\tilde{\Theta}_{\tau^\star}$. Then, using the first terms of a Taylor series expansion of $\log p(\Theta_{\tau^\star})$ around this mode and noticing that $\nabla p_{\tau^\star}(\tilde{\Theta}_{\tau^\star}) = \mathbf{0}$, we have:

$$\begin{aligned} p(\tau^\star \mid \mathbf{y}) &\propto \int \exp\{\log p(\Theta \mid \mathbf{y}, \tau^\star)\}\, d\Theta \\ &\overset{\text{approx}}{\propto} \int \exp\left\{\log p(\tilde{\Theta}_{\tau^\star} \mid \mathbf{y}, \tau^\star) - \frac{1}{2}(\Theta - \tilde{\Theta}_{\tau^\star})'\tilde{\mathbf{H}}_{\tau^\star}(\Theta - \tilde{\Theta}_{\tau^\star})\right\} d\Theta \\ &= \frac{p(\tilde{\Theta}_{\tau^\star} \mid \mathbf{y}, \tau^\star)}{(2\pi)^{n/2} \mid \tilde{\mathbf{H}}_{\tau^\star} \mid^{\frac{1}{2}}} \end{aligned} \qquad [34]$$

The determinant part in the last equation is obtained by recognizing the kernel of a multivariate normal density of mean $\tilde{\Theta}_{\tau^\star}$ and variance $\tilde{\mathbf{H}}_{\tau^\star}$ under the integral sign.

This results in an approximation of the marginal posterior density which is similar to what is described in the statistical literature as a saddlepoint approximation of this density (Daniels, 1954; Reid, 1988; Kolassa, 1994; Goutis and Casella, 1996). Taking the logarithm on both sides, we get the following approximation:

$$\log p(\tau^\star \mid \mathbf{y}) \approx \text{constant} + \log p(\tilde{\Theta}_{\tau^\star} \mid \mathbf{y}, \tau^\star) - \frac{1}{2} \log |\tilde{\mathbf{H}}_{\tau^\star}| \qquad [35]$$

$$= \text{constant} + f(\tau^\star) \qquad [36]$$

An obvious point estimate of τ is $\hat{\tau}$ at the mode of this approximate marginal posterior density:

$$\hat{\tau} = \text{Arg}_\tau \max \log p(\tau \mid \mathbf{y}) \qquad [37]$$

However, the use of [34] is not limited to the computation of its mode. Other point estimates or other types of inferences (credible sets or hypothesis testing, etc (Berger, 1985; Robert, 1992)) can be derived from the knowledge of the full marginal posterior density. Repeated computations of [34], and in particular of the negative Hessian matrix $\tilde{\mathbf{H}}$, for many different values of τ may quickly become too heavy, though. We propose to summarize the general characteristics of the distribution [34] through the computation of its first three moments by unidimensional numerical integration based on Gauss–Hermite quadrature. To obtain a more precise estimate of these moments after quadrature, the iterative strategy proposed by Smith et al (1985) is implemented. Using initial values of the mean and the variance of the distribution of $\log \tau$ (to force the integration domain to be $(-\infty, +\infty)$), the integration variable is standardized. New estimates are obtained by quadrature and the standardization is repeated. After a few iterations, this strategy ensures that the quadrature rules are applied in an appropriate region of the function to integrate. Details are given in the *Appendix*. The results can be used to obtain a second approximation of the marginal posterior density based on its first three moments. Using an expression known as the Gram–Charlier series expansion of a function $f(x)$ of a variable x with moments μ, σ^2 and κ, we have (McCullagh, 1987):

$$f(x) \approx \phi(x)(1 + \kappa(z^3 - 3z)/6) \qquad [38]$$

where $\phi(x)$ is the density of a normal distribution with mean μ and variance σ^2 and $z = (x - \mu)/\sigma$.

Other situations

Cox model

The application of the saddlepoint approximation to obtain the marginal posterior density of the dispersion parameter of the random effect is not restricted to the Weibull regression model. It can be applied, at least in theory, to any joint posterior density. For example, in the case of a Cox mixed model, for which the baseline hazard function $\lambda_0(t)$ is assumed to be completely arbitrary, $p(\tilde{\Theta}_{\tau^\star} \mid, \mathbf{y}, \tau^\star)$ and the corresponding negative Hessian matrix $\tilde{\mathbf{H}}_{\tau^\star}$ in [34] can be derived replacing the

likelihood function in [16] by the partial likelihood function initially proposed by Cox (1972):

$$\mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta}, \tau) = \prod_{T_{[i]}} \left(\frac{e^{w_i' \boldsymbol{\theta}}}{\sum_{m \in \mathrm{Risk}(T_{[i]})} e^{w_m' \boldsymbol{\theta}}} \right) \qquad [39]$$

where the $T_{[i]}$'s are the distinct observed failure times and $\mathrm{Risk}(T_{[i]})$ is the set of individuals at risk at time $T_{[i]}$, ie, alive just prior to $T_{[i]}$. Then, assuming that τ is known, the estimate of $\boldsymbol{\theta}$ to be used in [34] is obtained from the joint posterior density as:

$$\boldsymbol{\theta}_\tau = \mathrm{Arg}_{\boldsymbol{\theta}} \max p(\boldsymbol{\theta} \mid \mathbf{y}, \tau) \qquad [40]$$

Stratification. Time-dependent covariates

Stratification and the use of time-dependent covariates are common approaches to accommodate situations for which the proportional hazards is not valid for all effects or throughout the whole time range. As for the Cox model, the main changes with respect to the situation described so far occur in the computation of the likelihood and its derivatives and do not interfere with the validity of the saddlepoint approximation. For example, if the covariates in $bfm\mathbf{w}_m$ are step-functions of time with changes at times $\varphi_{m,i}, i = 0, ...I$ with $\varphi_{m,0} = 0$ and $\varphi_{m,I} = y_m$, then $\mathbf{w}_m$ is piecewise constant on intervals $[\varphi_{m,i}, \varphi_{m,i+1}[$ and the expressions to use in [12] are:

$$\lambda(y_m \mid \boldsymbol{\theta}, \rho) = \rho\, y_m^{\rho-1} e^{\mathbf{w}_m'(y_m)\boldsymbol{\theta}} \qquad [41]$$

$$S(y_m \mid \boldsymbol{\theta}, \rho) = \exp\left[- \sum_{i=0}^{i=I-1} e^{\mathbf{w}_m'(\varphi_{m,i})\boldsymbol{\theta}} (\varphi_{m,i+1}^{\rho} - \varphi_{m,i}^{\rho}) \right] \qquad [42]$$

In the case of stratification, the hazard function $\lambda(y_m)$ and the survivor function $S(y_m)$ include parameters ρ and $\rho \log \lambda$ (the 'intercept' in $\mathbf{w}_m'\boldsymbol{\theta}$ in [1]) specific to the relevant stratum.

ILLUSTRATION

In order to illustrate the approach described above for the estimation of dispersion parameters of the random effects in frailty models, simulated data were generated based on a Weibull model with a random effect (that will be referred to as a sire effect) and mimicking the data structure that is often encountered in animal breeding situations. The objective was to assess the quality of the saddlepoint approximation by comparing the exact marginal posterior distribution of the variance parameter of the sire effect ([28] obtained via algebraic integration) with its approximation ([34] after Laplacian integration). This comparison was done under the following conditions: a log-gamma distribution [8] was considered as a prior for the sire effect (which is a prerequisite for possible algebraic integration); only one

fixed effect ($\boldsymbol{\beta} = \mu$ the grand mean) was included; and it was assumed that in [28], we have:

$$p(\mu, \rho, \gamma \mid \mathbf{y}) \approx p(\gamma \mid \mathbf{y}, \mu = \hat{\mu}, \rho = \hat{\rho}) \qquad [43]$$

Preliminary examination of [43] showed that in all cases studied, the density [43] was virtually identical to the approximate density $p(\gamma \mid \mathbf{y})$ after integrating out μ and ρ by Laplacian integration. In other words, what was actually compared here are two approximate densities obtained after Laplacian integration of μ, ρ and N_q sire effects s_q in one case, of μ and ρ (with algebraic integration of the s_q's) in the other case.

The general behavior of the saddlepoint approximation of the marginal posterior density of the sire variance was also examined under a variety of situations (different types of censoring, of unbalanced structure, with a multivariate normal prior, with relationships between sires, using a Cox model, etc).

Simulation strategy

In all situations (unless specified otherwise), 5 000 records were generated using the following Weibull hazard function:

$$\lambda_{jkq}(t \mid \boldsymbol{\theta}) = \rho t^{\rho-1} \exp\{\mu + \beta_k + s_q\} \qquad [44]$$

where $\lambda_{jkq}(t)$ represents the hazard at time t of the jth animal ($j = 1, \cdots 5\,000/N_q$) under the influence of the kth level of a fixed effect, hereafter referred to as the 'herd' effect ($k = 1, \cdots K$) and daughter of the qth sire ($q = 1, \cdots N_q$). Values $\mu = -11$ and $\rho = 1.5$ were used in all cases described here, corresponding to an average failure time of about 1 800. For the comparison between Laplacian and algebraic integrations, it was assumed that $K = 0$, ie, $\beta_k = 0$ and the sire effects s_q were generated from a log-gamma distribution with parameter $\gamma = 50$. This corresponds to a variance of s_q equal to $\Psi^{(1)}(\gamma) \approx 0.02$, where

$$\Psi^{(1)}(\gamma) = \frac{\partial^2 \log \Gamma(x)}{\partial x^2}$$

is the trigamma function evaluated at γ. Using expression [6], we get:

$$h^2 = \frac{4\ \Psi^{(1)}(\gamma)}{\frac{\pi^2}{6} + \Psi^{(1)}(\gamma)} \approx 0.05 \qquad [45]$$

which is in the typical range of heritability values encountered for this kind of trait.

When a normal distribution was assumed, a sire variance of $\sigma = 0.02$ was retained to generate the sire effects. When herd effects were used in model [44] ($K > 0$), these were arbitrarily generated from a uniform $[-2, 2]$ distribution.

Two different censoring schemes were simulated. In censoring type A, all generated records greater than a given value C_A were considered as censored at C_A. The value of C_A was chosen by trial and error in order to obtain a given proportion of censored records. Censoring type B tried to mimic an overlapping generations scheme. The daughters of a first batch (10%) of sires had a censored record equal

to C_B when their simulated failure time was greater than C_B. The daughters of the following batch (also 10%) of sires were considered as censored when their failure time was greater that $2C_B$, and so on. The censoring time for the last 10% was $10C_B$. Therefore, the daughters of the first group of sires were heavily censored ('young daughters of young sires') while the proportion of censored records for the last group was small ('daughters of old sires'). Again, C_B was determined by trial and error.

Different unbalanced situations were also simulated. In scheme $U1$, the daughters of 100 sires (with 50 daughters each) were distributed over 505 herds, five with 500 animals and 500 with five daughters. In scheme $U2$, half of the animals (2 500) were assumed to be daughters of five sires with 500 daughters each while the other half were daughters of 500 sires with five daughters each. These animals were randomly distributed over 100 herds. Finally, in scheme $U3$, the daughters of the 50 'best' sires (with 50 daughters each) were raised in the 'best' 50 herds (where 'best' means lowest relative culling rate) while the daughters of the 'worst' 50 sires were raised in the 'worst' herds.

To study the impact of the existence of genetic relationships between individuals, data were generated according to a model slightly different from [44]. First, the effects s_{gs} of ten grandsires ('sires of sires') were generated from a normal distribution with mean 0 and variance $\sigma_s^2/4$ (with $\sigma_s^2 = 0.02$). For each of them, ten sire effects s_q were obtained by adding to s_{gs} a normally distributed random effect with variance $3\sigma_s^2/4$. Finally, 50 records of daughters of each of these sires were simulated according to the model:

$$\lambda_{jkq}(t \mid \boldsymbol{\theta}) = \rho t^{\rho-1}\exp\{\mu + \beta_k + s_q + r_j\} \qquad [46]$$

where r_j represents the remaining additive genetic effect for the jth animal and was generated from a normal distribution with mean 0 and variance $3\sigma_s^2$, leading to records with a global additive genetic variance equal to $\sigma_a^2 = 4\,\sigma_s^2$. These data were analyzed and the marginal posterior density of the sire variance component was obtained under three different genetic models: two sire models identical to [44] assuming no relationships between sires (case $S1$) or including the relationship matrix between sires (case $S2$), and an 'animal' model (case An), describing the individual additive genetic effect a_j of each animal j and including the complete relationship matrix between the 5 110 animals (5 000 with records + 100 sires + 10 grand-sires):

$$\lambda_{jk}(t \mid \boldsymbol{\theta}) = \rho t^{\rho-1}\exp\{\mu + \beta_k + a_j\} \qquad [47]$$

All computations were done using the 'Survival Kit', a set of Fortran programs developed by Ducrocq and Sölkner (1994). The 'Survival Kit' was specifically written to efficiently analyze the very large field data sets encountered by animal breeders and implements all the features described in this paper with Weibull and Cox models, possibly with strata, time-dependent covariates and random effects. In particular, the maximization of the expressions [18] or [29] is based on a limited memory quasi-Newton method (Liu and Nocedal, 1989) which only requires the computation of the vector of first derivatives of [18] or [29]. If required (for example, in [36] or when computing asymptotic standard errors), the negative Hessian is computed but only at convergence. Sparse matrix subroutines (Perez-Enciso et al,

1994) are used to compute the determinant or the inverse of this negative Hessian in the Weibull case.

Results

Laplacian integration vs Algebraic integration

Figure 1 represents the marginal posterior distribution obtained after integrating out the sire effects s_q from the joint posterior distribution, either algebraically or using the Laplacian approximation. All records were uncensored. In the three samples presented here, the true value $\gamma = 50$ is obviously included in any reasonable HPD credible set. When there were few sires with many daughters each, the two computed forms of the marginal posterior distribution were virtually indistinguishable. When little information was available for each sire effect (ten daughters each in the 500 sires case), the marginal posterior distributions were rather flat, with a long tail towards large values of γ (ie, small sire variances). The agreement between Laplacian and algebraic integration was not as good, although the modes of the two distributions were close. With even less information per sire (five daughters or less per sire), neither of the two marginalization techniques worked in most of the cases: the mode of the distribution or its first moments could not be computed.

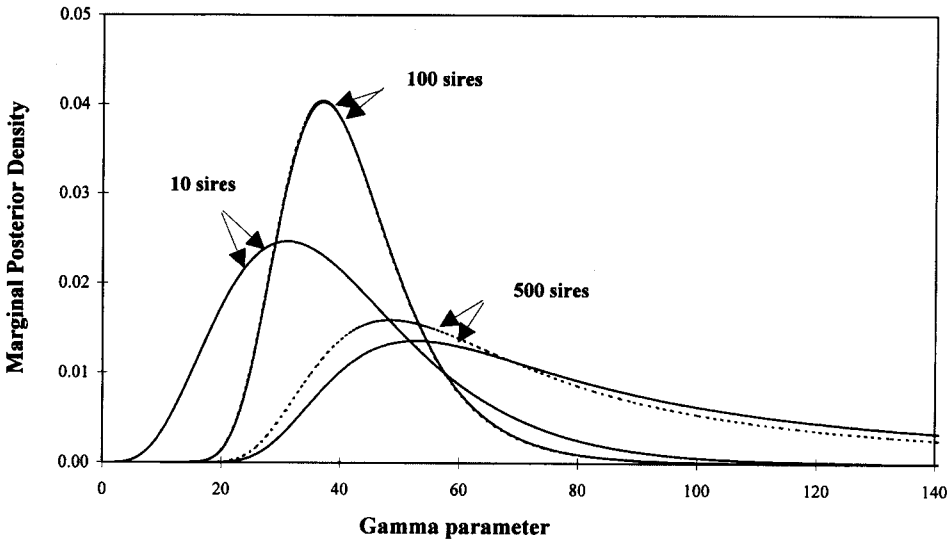

Fig 1. Marginal posterior density of the variance parameter γ obtained after Laplacian integration (- - - - -) or algebraic integration (———); 5 000 records; 10, 100 or 500 sires with 500, 50 or 10 daughters each. No censoring. $\gamma = 50$ = true value of the gamma parameter.

Effect of censoring

Figure 2 presents again the result of the same two marginalization approaches, for 100 sires with 50 daughters each but under censoring schemes A and B, with in both cases a proportion of 50% censored records ($C_A = 1\,200$ and $C_B = 270$). Clearly, censoring had little effect on the quality of the approximation when the Laplacian integration was used. However, because the amount of information available to estimate a rather small sire variance was drastically reduced, it was not always possible to obtain a well-defined posterior density (see Breslow and Clayton (1993) for similar results in the context of generalized linear mixed models). For example, in figure 2, the posterior density in the case of censoring scheme A does not integrate to 1. The same phenomenon also occurred for some samples with censoring scheme B. Interestingly, when sire effects with a larger variance $\gamma = 10$ were simulated, which corresponds to an heritability of 0.24, even extreme situations with more than 80% censored records (with $C_A = 520$) led to well-defined, very peaked posterior densities.

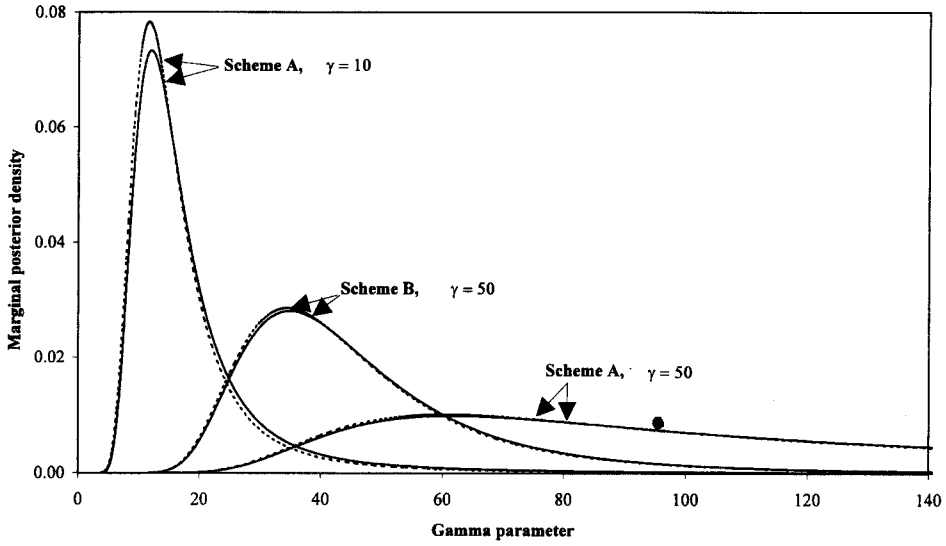

Fig 2. Marginal posterior density of the variance parameter γ obtained after Laplacian integration (- - - - -) or algebraic integration (———); 5 000 records; 100 sires with 50 daughters each. γ = true value of the gamma parameter. Censoring scheme A: all records above 540 days are censored at 540. Censoring scheme B: the sires are distributed over ten batches of ten sires, mimicking overlapping generations schemes. Censoring occurs after 270 days for batch 1, 540 days for batch 2, etc.

Normally distributed random effects

Having shown the validity of the saddlepoint approximation of the marginal posterior density, other samples were generated with normally distributed sire

effects and with 100 (fixed) 'herd' effects. Figure 3 displays the marginal posterior density for ten such samples, with 100 sires and no censoring. The obtained distributions were not as skewed as in the case of a log-gamma distribution. At least in the examples studied, the true value 0.02 was always in any HPD credible set. Note however that the variance of these densities were quite large (standard deviations between 0.0049 and 0.0079 for a true parameter value of 0.02).

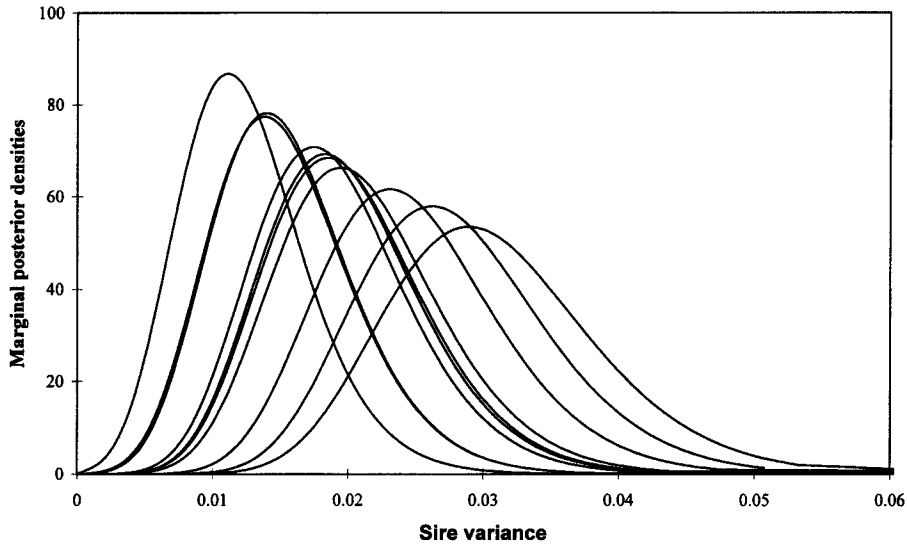

Fig 3. Marginal posterior density of the sire variance σ_s^2 obtained after Laplacian integration. Ten samples with 5 000 records; 100 sires with 50 daughters each; 100 'herd' effects; no censoring; $\sigma_s^2 = 0.02 =$ true value.

Effect of unbalancedness

When unbalancedness was induced by simulating both very large and small herds (case $U1$), the effect on the marginal posterior density appeared to be minimal (fig 4). When a large heterogeneity was created in the number of daughters per sire (case $U2$), the main consequence was a less precise estimation of the sire variance. The most negative impact was observed when the animals were not randomly distributed across herds (case $U3$). It seems that a part of the favorable influence of the best sires on the survival of their daughters was attributed to the herd effects, resulting in a sire variance strongly biased downwards.

Including a relationship matrix

The two marginal posterior densities obtained under a sire model with or without inclusion of the true relationship matrix between sires were very similar (fig 5). As may have been expected, the inclusion of the relationship matrix slightly increased the variance of this posterior density, because it accounts for the fact that

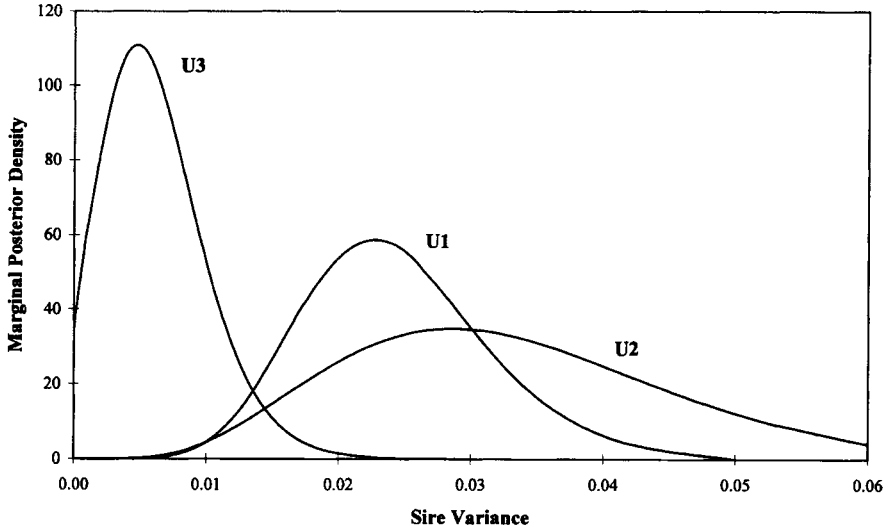

Fig 4. Marginal posterior density of the sire variance σ_s^2 obtained after Laplacian integration in three unbalanced cases; 5 000 records. Case $U1$: 500 herds with five animals; five herds with 500 animals. Case $U2$: 500 sires with five daughters each; five sires with 500 daughters each. Case $U3$: the daughters of the best sires are located in the best herds. No censoring; $\sigma_s^2 = 0.02 =$ true value.

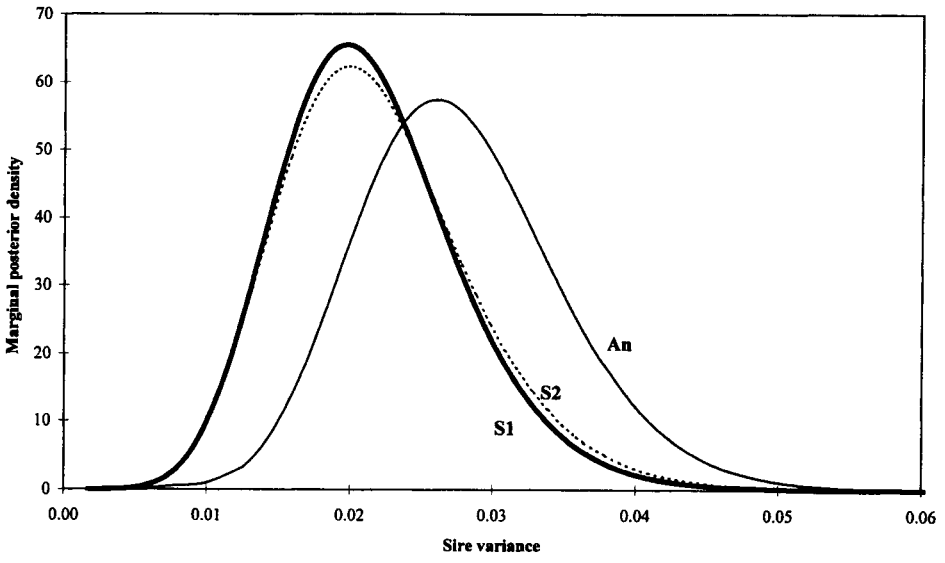

Fig 5. Marginal posterior density of the sire variance σ_s^2 obtained after Laplacian integration under two sire models (without a relationship matrix ($S1$) or with the true relationship matrix between sires ($S2$)) and under an animal model (An); 5 000 records; no censoring; $\sigma_s^2 = 0.02 =$ true value.

the records of related animals are more similar, hence globally less variable. In all the samples simulated, the animal model consistently led to a slight overestimation of the sire variance: the marginal posterior density in the case of the animal model was systematically to the right of those for the two sire models. This may be attributed, at least in part, to the fact that a much larger number of parameters have to be integrated out with an animal model than with a sire model. Such a problem has been pointed out for example by Mayer (1995) in the context of a threshold model. The Laplacian integration probably does not perform as well in such a case. Note, however, that this may be worsened by the fact that only a very simple pedigree structure was simulated here. In particular, no information at all was assumed to be available on the female side. The sire model used does not account for the overdispersion implicitly created by the effect r_j, which represents three-quarters of the total additive genetic variance. An attempt to fit a model similar to [46] assuming a log-gamma prior distribution for r_j and performing the algebraic integration of r_j led to a marginal posterior density of the sire variance similar to that obtained with the two sire models and a very large estimate ($\gamma > 400$ at the mode) for the gamma parameter, synonymous of a very small variance for the r_j's. This is likely the result of the lack of information available for the estimation for γ that was already illustrated in figure 1.

Cox model vs Weibull model

When a parametric (Weibull) or semi-parametric (Cox) model was used in the construction of the likelihood function, it was repeatedly observed that the resulting marginal posterior densities of σ_s^2 were very similar (fig 6), with often a slightly larger variance in the case of the Cox model. It is not known if similar results would have been obtained had the data been generated assuming a baseline hazard function different from the Weibull hazard.

Approximation of the marginal posterior density of τ based on its first three moments

The first three moments of the marginal posterior density of the parameter τ were computed by numerical integration of [34] using a five-point Gauss–Hermite quadrature formula and after standardization of the function to integrate. New standardization factors were obtained and the procedure was repeated until the computed moments stabilized, which usually occurred after only three iterations. Figure 7 illustrates the fact that the knowledge of these moments leads to a reasonable approximation of the marginal posterior density of τ

DISCUSSION AND CONCLUSION

Bayesian analysis offers a coherent framework for the otherwise unclear problem of variance components estimation in mixed nonlinear models (Ducrocq, 1990): all the elements for inferences on dispersion parameters are contained in the marginal posterior distribution of these parameters and the construction of the latter is based on general principles. Particular applications to animal breeding situations were

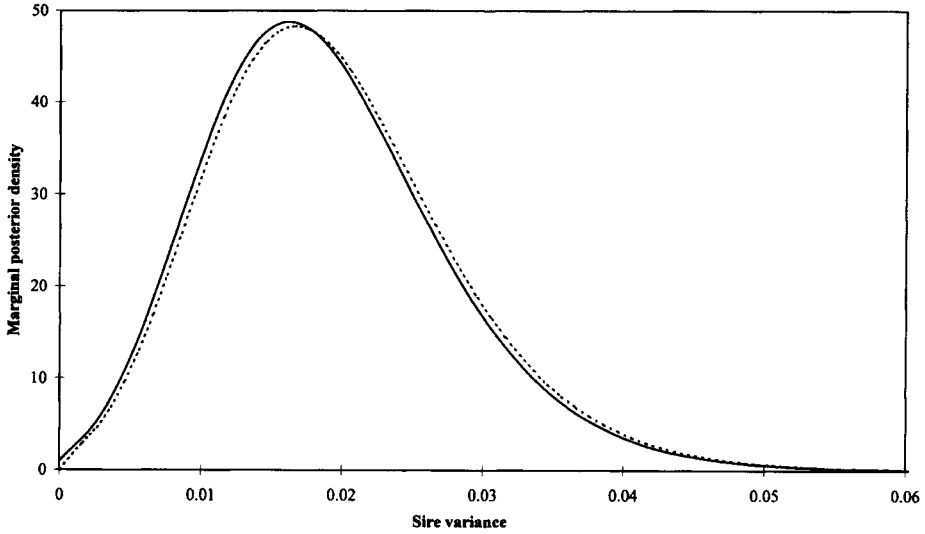

Fig 6. Marginal posterior density of the sire variance σ_s^2 obtained after Laplacian integration when a Weibull (——) or a Cox (- - - - -) model is used; 5 000 records generated from a Weibull distribution; 100 sires with 50 daughters each; 100 'herd' effects; no censoring; $\sigma_s^2 = 0.02 =$ true value.

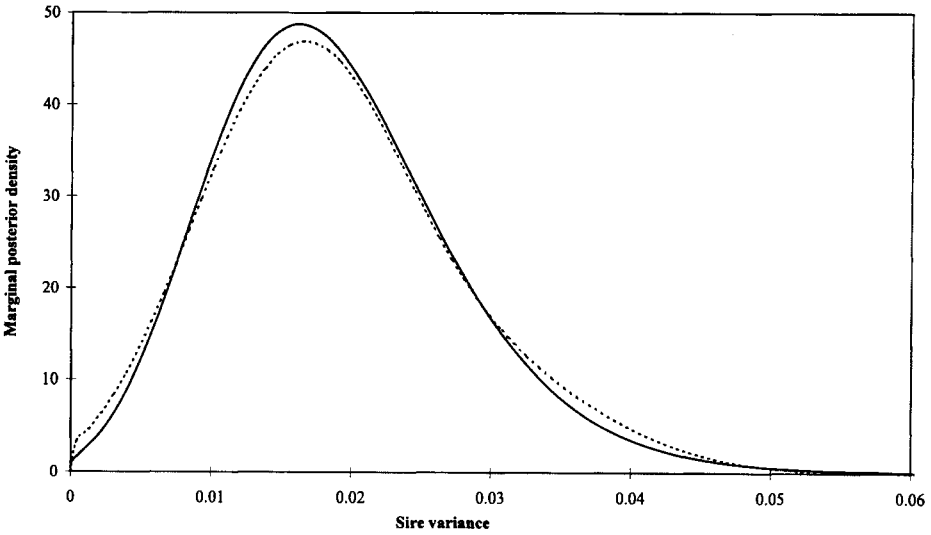

Fig 7. Marginal posterior density of the sire variance σ_s^2 obtained after Laplacian integration computed at 80 different points (——) or approximated using a Gram–Charlier series expansion based on its first three moments (- - - - -); 5 000 records generated from a Weibull distribution; 100 sires with 50 daughters each; 100 'herd' effects; no censoring; $\sigma_s^2 = 0.02 =$ true value, mode at $\tau = 0.0197$. Final estimates of the mean, standard deviation and coefficient of skewness: 0.0215, 0.0064, 0.588.

proposed for categorical data (Foulley et al, 1987; Höschele et al, 1987; Foulley et al, 1989) and for Poisson mixed models (Tempelman and Gianola, 1993). In this paper, a general approach for genetic evaluation and estimation of dispersion parameters for Weibull and Cox mixed models was described. Its main attractive features are its generality and its computational feasability, even for very large applications. As an example of the latter, the largest analysis that we have carried out involved the estimation of the mode and the first three moments of the marginal posterior distribution of the sire variance component for the length of productive life of 633 516 Holstein cows, daughters of 3 613 related sires. The Weibull mixed model used was quite complex and included time-dependent effects such as a herd–year–season effect (with 82 713 levels, assumed to be randomly distributed with a log-gamma distribution), a lactation number × stage of lactation effect, a herd size effect and a year-to-year variation in herd size effect as well as continuous linear and quadratic effects of covariates such age at first calving, milk, fat and protein yield.

Popular extensions of proportional hazards models such as stratification or the use of time-dependent covariates complicate the actual likelihood computations but do not interfere with the marginalization procedures described here. The inclusion of genetic relationships between individuals is straightforward through the use of an appropriate prior distribution. Other prior distributions (including informative priors) or other parametric baseline hazard functions could have been incorporated. More complex genetic structures (eg, with maternal effects) can be fitted. When more than one random effect is considered in the model, the approximation described here leads to the joint marginal posterior of all the dispersion parameters for all random effects. Further marginalization can be performed numerically along the lines described in the *Appendix* for the calculation of the moments of the marginal posterior distribution but this may be considered too costly. In the case of a Weibull mixed model with two random effects, one of them having a log-gamma distribution, the possibility of integrating out the latter algebraically avoids this difficulty.

Laplacian integration can be applied to other situations too. For example, Tierney and Kadane (1986) and Tierney et al (1989) suggested the direct computation of the mean of the marginal posterior density using second-order approximation formulae. These formulae were derived applying Laplacian integration to both the numerator and the denominator of a ratio of integrals. However, this requires the maximization of the joint posterior density for the dispersion parameters, the fixed effects and the random effects. This approach failed when we attempted it as the maximization procedure led to dispersion parameters estimates corresponding to random effects with null variance. The same phenomenon had been described previously in similar situations (Tempelman and Gianola, 1993).

At least in theory, Laplacian integration could have been used to obtain the marginal posterior distribution of parameters other than the dispersion parameters. However, this may be considered far too demanding, because each application of the Laplace expansion requires the maximization of one particular function involving all parameters except the one of interest. This is in contrast with some Monte-Carlo methods, such as Gibbs sampling, where the marginal distributions for all parameters can be obtained simultaneously. However, in practical animal breeding

situations, the separate consideration of all marginal densities is often not required, because estimated breeding values are point estimates mainly used to rank animals: when little information is available for the genetic evaluation, an accurate ranking of the candidates to selection is unrealistic. In the opposite case (precise estimation), the rankings based on, say, the mode or the mean of either the marginal or the joint posterior distribution are likely to be very similar. Marginal posterior densities of nonlinear functions of parameters can also be calculated (Wong and Li, 1992).

Marginalization based on Laplacian integration has been shown to give excellent results in standard situations. For many nonlinear applications, the quality of the saddlepoint approximation would have to rely on the comparison of the approximate marginal distribution of the dispersion parameters with the actual distribution obtained via Monte-Carlo simulations. The exceptional situation studied here where an exact algebraic integration of a log-gamma random effect is possible permits a more straightforward comparison. It was found that the designs for which the two marginal posterior distributions (exact and approximate) depart from each other correspond to situations where the quantity of information available for the estimation of genetic parameters is quite limited. This means, in particular, that the saddlepoint approximation is likely to be unsuccessful for the estimation of the parameters of a frailty term used to describe an extra variation (overdispersion). However, one can still use algebraic integration of the random effects in the case of a gamma frailty component in a Weibull model.

ACKNOWLEDGMENTS

The first author thanks the INRA for making possible his stay at Cornell University and expresses his appreciation to Genex Cooperative, Inc, Ithaca, New York for its support.

REFERENCES

Aalen OO (1994) Effects of frailty in survival analysis. *Statist Meth in Med Res* 3, 227–243

Abramowitz M, Stegun IA (1964) *Handbook of Mathematical Functions.* US Department of Commerce, National Bureau of Standards, 1046 p

Achcar JA, Bolfarine H (1986) Use of accurate approximations for posterior densities in regression models with censored data. *Rev Soc Chil Estad* 3, 84–104

Anderson JE, Louis TA, Holm NV, Harvald B (1992) Time-dependent association measures for bivariate survival analysis. *J Am Stat Ass* 87, 641–650

Berger JO (1985) *Statistical Decision Theory and Bayesian Analysis.* Springler-Verlag, New York, NY

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *J Am Stat Ass* 88, 9-25

Clayton DG (1991) A Monte-Carlo method for Bayesian inference in frailty models. *Biometrics* 47, 467–485

Clayton DG, Cuzick J (1985) Multivariate generalizations of the proportional hazards model. *J R Stat Soc, Series A* 148, 82–117

Cox D (1972) Regression models and life table (with discussion). *J R Stat Soc, Series B* 34, 187–220

Cox DR, Oakes D (1984) *Analysis of Survival Data.* Chapman and Hall, London, UK

Daniels, HE (1954) Saddlepoint approximations in statistics. *Ann Math Statist* 25, 631–650

Dellaportas P, Smith AFM (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Appl Stat* 42, 443–459

Dempster AP, Laird NM, Rubin DR (1977) Maximum likelihood estimation for incomplete data via the EM algorithm (with discussion). *J R Stat Soc, Series B* 39, 1–38

Ducrocq V (1987) An analysis of length of productive life in dairy cattle. PhD dissertation, Cornell University, Ithaca, NY, USA

Ducrocq V (1990) Estimation of genetic parameters arising in nonlinear models. In: *4th World Cong Genet Appl Livest Prod, Edinburgh, UK, July 23–27* 13, 419–428

Ducrocq V (1994) Statistical analysis of length of productive life for dairy cows of the Normande breed. *J Dairy Sci*, 77, 855–866

Ducrocq V, Quaas RL, Pollak EJ, Casella G (1988a) Length of productive life of dairy cows. I. Justification of a Weibull model. *J Dairy Sci* 71, 3061–3070

Ducrocq V, Quaas RL, Pollak EJ, Casella G (1988b) Length of productive life of dairy cows. II. Variance component estimation and sire evaluation. *J Dairy Sci* 71, 3071–3079

Ducrocq V, Sölkner J (1994) 'The Survival Kit', a Fortran package for the analysis of survival data. In: *5th World Cong Genet Appl Livest Prod*, Dep Anim Poult Sci, Univ of Guelph, Guelph, Ontario, Canada 22, 51–52

Egger-Danner C (1993) Zuchtwerschätzung für Merkmale der Langlebigkeit beim Rind mit Methoden des Lebensdaueranalyse. PhD dissertation, University of Agriculture, Vienna, Austria

Follmann DA, Goldberg MS (1988) Distinguishing heterogeneity from decreasing hazard rates. *Technometrics* 30, 389–396

Foulley JL, Gianola D, Im S, Misztal I (1989) Une approche bayésienne de l'analyse de caractères discrets. In: *Biométrie et données discrètes* (B Asselain, C Duby, JP Masson, J Tranchefort, eds). École nationale supérieure agronomique, Rennes, France, 6–35

Foulley JL, Im S, Gianola D, Höschele I (1987) Empirical Bayes estimation of parameters for n polygenic binary traits. *Gen Sel Evol* 19, 197–224

Fournet F (1992). *Étude de la durée de carrière sportive du cheval de concours hippique.* Mémoire de Diplôme d'Agronomie Approfondie, Inst Natl Agron, Paris-Grignon, Paris, France

Goutis C, Casella G (1996) Explaining the saddlepoint approximation. Technical report, Biometrics Unit, Cornell University, Ithaca, NY, USA, BU-1311-M

Hoeschele I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334–349

Hougaard P (1986a) A class of multivariate failure time distributions. *Biometrika* 73, 671–678

Hougaard P (1986b) Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 73, 387–396

Kalbfleisch JD, Prentice RL (1980) *The Statistical Analysis of Failure Time Data.* John Wiley and sons, New York, USA

Klein JP (1992) Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* 48, 795–806

Klein JP, Moeschberger M, Li, YH, Wang ST (1992) Estimation random effects in the Framingham heart study. In: *Survival Analysis: State of the Art* (J Klein, P Goel, eds), Kluwer Academic Publishers, 99–120

Kolassa, JE (1994) *Series Approximation Methods in Statistics.* Springer-Verlag, New York, USA

Korsgaard, IR (1996) Genetic analysis of survival data – a Gibbs sampling approach. International Workshop on Genetic Improvement of Functional Traits in Cattle, Faculté universitaire des sciences agronomiques, Gembloux, Belgium, Jan 21-23, 1996

Lawless J (1982) *Statistical Models and Methods for Lifetime Data.* John Wiley and sons, New York, NY

Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. *Math Program* 45, 503–528

Louis T (1991) Assessing, accommodating and interpreting the influences of heterogeneity. *Environ Health Perspect* 80, 215–222

Mayer M (1995) Inequality of maximum a posteriori estimators with equivalent sire and animal models for threshold traits. *Gen Sel Evol* 27, 423-435

McCullagh P (1987) *Tensor Methods in Statistics.* Chapman and Hall, London, UK

Perez-Enciso M, Mizstal I, Elzo MA (1994) Fspak: an interface for public domain sparse matrix subroutine. In: *5th World Cong Genet Appl Livest Prod,* Dept Anim Poultry Sci, Univ of Guelph, Guelph, Ontario, Canada, 22, 87–88

Reid N (1988) Saddlepoint methods and statistical inference. *Stat Sci* 3, 213–238

Robert C (1992) *L'analyse statistique Bayesienne.* Economica, Paris, France

Ruiz F (1991) Relationship among length of productive life, milk yield and profitability of US, Canadian and Mexican Holstein sires in Mexico. PhD dissertation, Cornell University, Ithaca, NY, USA

Smith A, Skene AM, Shaw J, Naylor J, Dransfield M (1985) The implementation of the Bayesian paradigm. *Commun Stat, Theor Meth* 14, 1079–1102

Smith S (1983) The extension of failure time analysis to problems of animal breeding. PhD dissertation, Cornell University, Ithaca, NY, USA

Smith SP, Quaas RL (1984) Productive life span of bull progeny groups: failure time analysis. *J Dairy Sci,* 67, 2999–3007

Strandberg E (1991) Breeding for lifetime performance in dairy cattle. PhD dissertation, Swedish Univ of Agricultural Sciences, Uppsala, Sweden

Strandberg E (1995) Breeding for longevity in dairy cows. In: *Progress in Dairy Science,* (C Philipps, ed), CAB International, Wallingford, UK, 125–144

Strandberg E, Sölkner J (1996) Breeding for longevity and survival in dairy cattle. International Workshop on Genetic Improvement of Functional Traits in Cattle, Faculté universitaire des sciences agronomiques, Gembloux, Belgium, Jan 21-23, 1996

Tempelman RJ, Gianola D (1993) Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration. *Gen Sel Evol* 25, 305–319

Tempelman RJ, Gianola D (1996) A mixed effects model for overdispersed count data in animal breeding. *Biometrics* 52, 265–279

Tierney L, Kardane JB (1986) Accurate approximations for posterior moments and marginal densities. *J Am Stat Ass* 81, 82–86.

Tierney L, Kass RE, Kardane JB (1989) Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J Am Stat Ass* 84, 710–716

van Arendonk J (1986) Economic importance and possibilities for improvement of dairy cow herd life. In: *3rd World Cong Genet Appl Livest Prod, July 16–22,* Lincoln, Nebraska, USA 9, 95–100

Vaupel J, Manton KG, Stallard E (1979) The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography* 16, 439–454.

Wong WH, Li B (1992) Laplace expansion for posterior densities of nonlinear functions of parameters. *Biometrika* 79, 393–398

APPENDIX: MOMENTS OF THE MARGINAL POSTERIOR DENSITY OF τ

Define $g(\tau) = \exp\{f(\tau) - f(\hat{\tau})\}$. Expressions [36] or [37] imply:

$$p(\tau \mid \mathbf{y}) = k\, g(\tau) \qquad \text{[A1]}$$

for some integration constant k $\left(\int_0^\infty g(\tau)d\tau = \frac{1}{k}\right)$. Let:

$$h_n(\tau) = \int_0^\infty \tau^n\, g(\tau)d\tau \qquad \text{[A2]}$$

Knowing $h_n(\tau), n = 0, \ldots, 3$, one can compute k and the first three moments of the approximate marginal posterior density of τ:

$$k = \frac{1}{h_0(\tau)} \qquad \text{[A3]}$$

$$\mu(\tau) = \mu_1 = \frac{h_1(\tau)}{h_0(\tau)} \qquad \text{[A4]}$$

$$\mathrm{Var}(\tau) = \mu_2 - \mu_1^2 = \frac{h_2(\tau)}{h_0(\tau)} - \mu_1^2 \qquad \text{[A5]}$$

$$\kappa = \frac{(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3)}{\mathrm{Var}(\tau)^{3/2}} \qquad \text{[A6]}$$

with $\mu_3 = \frac{h_3(\tau)}{h_0(\tau)}$. Adapting the approach of Smith et al (1985) to our particular case, the expressions [A3], [A4], [A5] and [A6] are computed iteratively using the following algorithm:

– Reparameterize τ in such a way that the new variables take values between $-\infty$ and $+\infty$. Here, this can be done with the change of variable $\xi = \log \tau$

$$h_n(\tau) = s_n(\xi) = \int_{-\infty}^{+\infty} e^{(n+1)\xi}\, g(e^\xi)\, d\xi \qquad \text{[A7]}$$

– Let μ_ξ and σ_ξ^2 be the (approximate) marginal posterior mean and variance of ξ. By definition:

$$\mu_\xi = \int_{-\infty}^{+\infty} \xi\, g(e^\xi)\, d\xi \qquad \text{[A8]}$$

$$\sigma_\xi^2 = \int_{-\infty}^{+\infty} \xi^2\, g(e^\xi)\, d\xi - \mu_\xi^2 \qquad \text{[A9]}$$

Let $\mu_\xi^{(0)}$ and $\sigma_\xi^{(0)}$ be initial estimates of these moments. Standardize ξ using the transformation $\nu = \dfrac{\xi - \mu_\xi^{(0)}}{\sigma_\xi^{(0)}}$. Then, we get a first estimate of the moments in [A2]:

$$h_n(\tau) = \int_{-\infty}^{+\infty} \sigma_\xi^{(0)}\, e^{(n+1)(\mu_\xi^{(0)}+\sigma_\xi^{(0)}\nu)} \quad g(e^{(\mu_\xi^{(0)}+\sigma_\xi^{(0)}\nu)})\, d\nu \qquad \text{[A10]}$$

and new estimates for μ_ξ and σ_ξ^2 by computing:

$$\mu_\xi^{(1)} = \int_{-\infty}^{+\infty} \sigma_\xi^{(0)}\, (\mu_\xi^{(0)} + \sigma_\xi^{(0)}\nu) \quad g(e^{(\mu_\xi^{(0)}+\sigma_\xi^{(0)}\nu)})\, d\nu \qquad \text{[A11]}$$

$$\sigma_\xi^{2(1)} = \int_{-\infty}^{+\infty} \sigma_\xi^{(0)}\, (\mu_\xi^{(0)} + \sigma_\xi^{(0)}\nu)^2 \quad g(e^{(\mu_\xi^{(0)}+\sigma_\xi^{(0)}\nu)})\, d\nu - \mu_\xi^{(1)^2} \qquad \text{[A12]}$$

Finally, factoring out the expression $e^{-\frac{\nu^2}{2}}$ in the integrand, we get:

$$h_n(\tau) = \int_{-\infty}^{+\infty} e^{-\frac{\nu^2}{2}} \left[\sigma_\xi^{(0)}\, e^{(n+1)(\mu_\xi^{(0)}+\sigma_\xi^{(0)}\nu)+\frac{\nu^2}{2}}\, g(e^{(\mu_\xi^{(0)}+\sigma_\xi^{(0)}\nu)})\right] d\nu \qquad \text{[A13]}$$

$$= \int_{-\infty}^{+\infty} e^{-\frac{\nu^2}{2}} \quad r(\nu, \mu_\xi^{(0)}; \sigma_\xi^{(0)})\, d\nu \qquad \text{[A14]}$$

Similar expressions exist for $\mu_\xi^{(1)}$ and $\sigma_\xi^{2(1)}$. They are of the form required for the application of the Gauss–Hermite quadrature rules. For example, $h_n(\tau)$ will be evaluated as:

$$h_n(\tau) \approx \sum_{i=1}^{I} \omega_i \quad r(\nu_i, \mu_\xi^{(0)}, \sigma_\xi^{(0)}) \qquad \text{[A15]}$$

where ν_i and ω_i, for $i = 1, ...I$, are the roots and the associated weights of the Hermite polynomial of order I (Abramowitz and Stegun, 1964). Again, similar formulae apply to $\mu_\xi^{(1)}$ and $\sigma_\xi^{2(1)}$. Once those new values of μ_ξ and σ_ξ^2 have been computed, they can replace the initial values $\mu_\xi^{(0)}$ and $\sigma_\xi^{(0)}$ and the procedure can be iterated until convergence.

It is important to note that the main work involved is the computation of $g(e^{(\nu)})$ at I points ν_i and that the resulting values are used repeatedly in the computation of $h_n(\tau), \mu_\xi^{(1)}$ and $\sigma_\xi^{2(1)}$.