Empirical Bayes estimation for additive hazards regression models

BY DEBAJYOTI SINHA

Department of Statistics, Florida State University, Tallahassee, Florida 32306, U.S.A.
sinhad@stat.fsu.edu

M. BRENT MCHENRY

Bristol-Myers Squibb, 5 Research Parkway, Wallingford, Connecticut 06492, U.S.A.
brent.mchenry@bms.com

STUART R. LIPSITZ

Harvard Medical School, Boston, Massachusetts 02115, U.S.A.
slipsitz@partners.org

AND MALAY GHOSH

Department of Statistics, University of Florida, Gainesville, Florida 32611, U.S.A.
ghoshm@stat.ufl.edu

SUMMARY

We develop a novel empirical Bayesian framework for the semiparametric additive hazards regression model. The integrated likelihood, obtained by integration over the unknown prior of the nonparametric baseline cumulative hazard, can be maximized using standard statistical software. Unlike the corresponding full Bayes method, our empirical Bayes estimators of regression parameters, survival curves and their corresponding standard errors have easily computed closed-form expressions and require no elicitation of hyperparameters of the prior. The method guarantees a monotone estimator of the survival function and accommodates time-varying regression coefficients and covariates. To facilitate frequentist-type inference based on large-sample approximation, we present the asymptotic properties of the semiparametric empirical Bayes estimates. We illustrate the implementation and advantages of our methodology with a reanalysis of a survival dataset and a simulation study.

Some key words: Gamma process; Integrated likelihood; Posterior process.

1. INTRODUCTION

For univariate survival time T_i subject to noninformative right censoring at C_i , we observe a $k \times 1$ vector x_i of fixed covariates, $y_i = \min(T_i, C_i)$ and the censoring indicator $\delta_i = 1_{T_i \leq C_i}$ for $i = 1, \dots, n$. Since the publication of Cox's (1972) seminal paper, the survival analysis literature has focused mainly on Cox's proportional hazards model $h(t|x) = h_0(t) \exp(\beta^T x)$, where $\text{pr}\{T \in (t, t+dt) | T > t, x\} = h(t|x)dt$ with baseline hazard $h_0(t)$ and $(k \times 1)$ vector β of regression parameters. For $j = 1, \dots, k$, the regression parameter β_j in Cox's model can be

interpreted as the logarithm of the hazard ratio for every unit increment of the corresponding covariate component x_j . The additive hazards model of Aalen (1980) is

$$h(t | x) = h_0(t) + \beta^T x, \quad (1)$$

where, for $j = 1, \dots, k$, β_j is interpreted as the increase in hazard rate for every unit increase in the corresponding component x_j . Aalen's model in (1) is gaining much attention as an alternative to Cox's model. Unlike in Cox's model, the difference in hazard rates $h(t | x_1) - h(t | x_2) = \beta^T(x_1 - x_2)$ and the ratio $S(t | x_1)/S(t | x_2) = \exp\{-t\beta^T(x_1 - x_2)\}$, related to predicted excess mortality, for two subjects can be evaluated without using the unknown $h_0(t)$. However, for Aalen's model, $h_0(t)$ is useful for estimating the survivor function $S(t | x)$ and for predicting the survival behaviour of a particular subject. For a nonparametric version of the model in (1), Aalen (1980, 1989, 1993) presented various frequentist nonparametric estimators of the time-varying regression parameter $\beta(t)$ when the functional form of $h_0(t)$ is unknown. Some authors, notably Lin & Ying (1994) and Lin et al. (1998), have presented semiparametric estimation of β and its standard error using martingale-based method-of-moments estimators. An authoritative review of frequentist estimation under the semiparametric model of (1) is given in Martinussen & Scheike (2006). The estimated $S(t | x)$ resulting from such methods may not be monotone. The post hoc procedures, such as constrained optimization to ensure the nonincreasing property of the estimated survival function, lack necessary theoretical and asymptotic justification and may lead to unknown small-sample bias. Unlike the partial-likelihood method for Cox's model, existing likelihood-based methodology for the semiparametric additive hazards model is not computationally straightforward.

Fully Bayesian methods for Cox's model have been achieved by incorporating the prior information into a prior process for nonparametric $h_0(t)$ and a prior density for β (Kalbfleisch, 1978; Ibrahim et al., 2001). For the Cox model, it has been shown, notably by Kalbfleisch (1978) and Sinha et al. (2003), that the integrated likelihood of β after integrating out $h_0(t)$ with respect to its prior process is similar to the partial likelihood of Cox (1972) and this integrated likelihood can be maximized to obtain an empirical Bayes estimator of β . In this paper, we develop the comparable estimator for the additive hazards regression model by placing a suitable prior process on the baseline cumulative hazard function, $H_0(t) = \int_0^t h_0(u)du$. Beamonte & Bermúdez (2003) presented a full Bayes method based on Markov chain Monte Carlo for $h(t | x) = h_0(t) + h_1(t | x)$, where $h_1(t | x)$ is the hazard function for a gamma density in which the shape parameter is a function of x . In this model, $S(t | x)$ and $h(t | x)$ lack explicit and interpretable expressions as functions of x . Full Bayes methods of Dunson & Herring (2005); Zeng et al. (2006) and Yin & Ibrahim (2005) used different tools based on Markov chain Monte Carlo simulation. All of the full Bayes methods require problem-specific programs, prespecified grids for the piecewise-constant $h_0(t)$, and full specifications and elicitations of the prior process of $h_0(t)$ as well as the conditional prior $\pi\{\beta | h_0(t)\}$. The novelty of our method is that users only need to specify the functional form $\lambda_0(t | \gamma)$ of the prior guess of nonparametric $h_0(t)$, instead of eliciting all the hyperparameters of the prior process.

2. INTEGRATED LIKELIHOOD

Given the observed data vector $D = \{y_i, \delta_i, x_i : i = 1, \dots, n\}$, with time-constant covariates x_i , the likelihood under the additive hazards model in (1) is given by

$$L\{\beta, H_0(\cdot) | D\} \propto \prod_{i=1}^n \{h_0(y_i) + \beta^T x_i\}^{\delta_i} \exp\{-H_0(y_i) - y_i \beta^T x_i\}, \quad H_0(t) = \int_0^t h_0(u)du.$$

For Bayesian analysis, a joint prior density $\pi\{\beta, H_0(\cdot)\}$ can incorporate prior information about the nonparametric $H_0(t)$ and the regression parameter β . For semiparametric Bayesian analysis of Cox's model, the joint prior distribution $\pi\{\beta, H_0(\cdot)\}$ is often simplified by assuming a priori independence of β and $H_0(\cdot)$, and then using a suitable known prior process, such as a gamma process (Kalbfleisch, 1978), for the marginal prior $\pi\{dH_0(\cdot)\}$; see Ibrahim et al. (2001) for a review. In a similar fashion, we use an improper flat prior for β , and the empirical Bayesian inference for the additive hazards model can be based on the integrated likelihood

$$L(\beta, \gamma | D) \propto \int L(\beta, H_0 | D) \pi(dH_0 | \gamma) dH_0, \quad (2)$$

where γ is the vector of hyperparameters associated with the prior process of $H_0(t)$. A natural choice for $\pi\{dH_0(\cdot) | \gamma\}$ is a gamma process for the cumulative hazard $H_0(\cdot)$, denoted by $H_0(t) \sim GP(\Lambda_0, \alpha)$, so that the disjoint increments $H_0(a_j) - H_0(a_{j-1})$ for $0 < a_1 < \dots < a_M < +\infty$ are distributed according to independent gamma distributions with mean $\Lambda_0(a_j) - \Lambda_0(a_{j-1})$ and variance $\{\Lambda_0(a_j) - \Lambda_0(a_{j-1})\}/\alpha$, where α is a known precision parameter and $\Lambda_0(t)$ is an increasing parametric function such that $\Lambda_0(0) = 0$. We assume that the mean function $\Lambda_0(t) = \Lambda(t | \gamma)$ is known except for the unknown hyperparameter vector γ . This mean function is thought of as the prior guess for the unknown nonparametric $H_0(\cdot)$ with α representing the degree of confidence in the prior guess. For example, a simple choice is $\Lambda_0(t | \gamma) = t\gamma$, corresponding to the prior belief that the mean of the prior guess for $h_0(t)$ is a constant. However, the prior belief may not specify the exact value of γ . The known value of α specifies our degree of prior belief in $h_0(t)$ being constant. A small value of α , equal to 1 or less, corresponds to weak confidence in the prior parametric model $\Lambda(t | \gamma)$ of $H_0(t)$, and in that case $\pi\{dH_0(\cdot)\}$ contributes minimally to the estimation of β .

We first derive the integrated likelihood of (2). Without loss of generality, we assume that observed survival data are ordered according to $y_1 < \dots < y_n$ with corresponding censoring indicators $\delta_1, \dots, \delta_n$. The joint probability of survival for n subjects $T = (T_1, \dots, T_n)$, evaluated at $Y = (y_1, \dots, y_n)$ and given the covariate matrix X , is

$$\text{pr}\{T > Y | \beta, H_0(\cdot), X\} = \exp\left[-\sum_{i=1}^n \{H_0(y_i) + \theta_i y_i\}\right] = \exp\left\{-\sum_{i=1}^n (r_i h_i + \theta_i y_i)\right\}, \quad (3)$$

where $\theta_i = \beta^T x_i$, and r_i is the number of subjects in $R_i = \{j : y_j \geq y_i\}$, the risk set at time y_i . As discussed above, the gamma prior process for $H_0(\cdot)$ implies that the disjoint increments, $h_i = H_0(y_i) - H_0(y_{i-1})$, say, are distributed according to independent gamma distributions, $h_i \sim \text{Ga}(\alpha \lambda_i, \alpha)$ with mean $\lambda_i = \Lambda_0(y_i | \gamma) - \Lambda_0(y_{i-1} | \gamma)$ and variance λ_i/α . The joint survival function of (3) given the gamma process is obtained as

$$\begin{aligned} \text{pr}\{T > Y | \beta, \Lambda_0(\cdot), \alpha\} &= E_{GP}[\text{pr}\{T > Y | \beta, X, H_0(\cdot)\} | \Lambda_0(\cdot), \alpha] \\ &= \prod_{i=1}^n \left\{ \left(\frac{\alpha}{\alpha + r_i} \right)^{\alpha \lambda_i} \exp\left(-\sum_{i=1}^n \theta_i y_i\right) \right\} \\ &= \prod_{i=1}^n \exp\left\{-\theta_i y_i - \alpha \lambda_i \log\left(\frac{\alpha + r_i}{\alpha}\right)\right\} \\ &= \prod_{i=1}^n \exp\left\{-\theta_i y_i - \alpha \Lambda_0(y_i) \log\left(\frac{\alpha + r_i}{\alpha + r_{i+1}}\right)\right\}. \end{aligned} \quad (4)$$

The integrated likelihood of β and γ , obtained by differentiating $\text{pr}\{T > Y \mid \beta, \Lambda_0(\cdot), \alpha\}$ in (4) with respect to each y_i whenever the censoring indicator δ_i equals 1, is given by

$$L(\beta, \gamma \mid D, \alpha) \propto \prod_{i=1}^n \exp\{-\theta_i y_i - \Lambda_0(y_i) c_{i\alpha}\} \{\theta_i + \lambda_0(y_i) c_{i\alpha}\}^{\delta_i}, \quad (5)$$

where $c_{i\alpha} = \alpha \log\{(\alpha + r_i)/(\alpha + r_{i+1})\}$ and $\lambda_0(t) = \lambda_0(t \mid \gamma) = (d/dt)\Lambda_0(t \mid \gamma)$. As the precision parameter $\alpha \rightarrow \infty$, the integrated likelihood in (5) converges to the likelihood $L(\beta, \gamma \mid D)$ based on the parametric additive hazards model with $h(t \mid x) = \beta^T x + \lambda_0(t \mid \gamma)$. When the prior mean of the nonparametric baseline hazard $h_0(t)$ is unknown constant γ , i.e. $\lambda_0(t) = \gamma$, the integrated likelihood of (5) further simplifies to

$$L(\beta, \gamma \mid D, \alpha) \propto \prod_{i=1}^n \exp(-\theta_i y_i - \gamma y_i c_{i\alpha}) (\theta_i + \gamma c_{i\alpha})^{\delta_i}. \quad (6)$$

With α known, the integrated likelihood of (5) is proportional to the likelihood of a Poisson linear model with outcome δ_i , parameter vector (γ, β) and covariate vector $(y_i c_{i\alpha}, y_i x_i)$. The empirical Bayes estimator of γ and β can be obtained by treating the integrated likelihood of (5) as a typical Poisson likelihood, that is, by maximizing $L(\gamma, \beta \mid D, \alpha)$ as a function of γ and β . In particular, one can set the score functions, that is, first derivatives of the loglikelihood, equal to zero. When a generalized linear model program is used, the empirical Bayes estimator of (γ, β) is obtained using outcome δ_i , covariates $y_i c_{i\alpha}$ and $y_i x_i$, that is, interactions between y_i and all elements of $(c_{i\alpha}, x_i)$, Poisson error, linear link and no intercept, since, even though γ is constant, the covariate $y_i c_{i\alpha}$ is not constant. The estimated variance-covariance matrix V of estimators $(\hat{\beta}, \hat{\gamma})$ is the inverse of the observed information matrix $I(\hat{\beta}, \hat{\gamma})$. This estimated matrix V is used in § 3 to compute the variance estimators of estimated survival functions. In § 6, it is shown to be a consistent estimator of the asymptotic variance-covariance matrix of $(\hat{\beta}, \hat{\gamma})$. As described in § 3, empirical Bayes estimation of $H_0(t)$ will require additional programming even when $\Lambda_0(t \mid \gamma) = t\gamma$. Furthermore, when the prior mean of $h_0(t)$ is not constant over time, it may not be easy to obtain $(\hat{\beta}, \hat{\gamma})$ through a generalized linear model software program.

3. THE POSTERIOR PROCESS AND ESTIMATOR

To evaluate the empirical Bayes estimator of the cumulative hazard $H_0(t)$ and the survival function $S_0(t) = \exp\{-H_0(t)\}$, we begin with the posterior process $p(H_0 \mid \gamma, \beta, Y)$. We show that the resulting process may not have any standard form; however, it is very closely related to the gamma process of Kalbfleisch (1978). We find a closed-form expression of the Laplace transform of $p\{H_0(t) \mid \lambda, \beta, D\}$, so that posterior moments of $H_0(t)$ and $S_0(t) = \exp\{-H_0(t)\}$ can be evaluated easily from the same Laplace transform.

We first put a grid $t_1 = 0 < t_2 < \cdots < t_{n+1} < \infty$ such that $t_i \in (y_{i-1}, y_i)$, the increments of $H_0(t)$ within intervals generated by this grid, and the ordered survival times $y_1 < \cdots < y_n$ are denoted by $u_{i1} = H_0(y_i) - H_0(t_i)$, $u_{i2} = H_0(t_{i+1}) - H_0(y_i)$ and $u_i = H_0(t_{i+1}) - H_0(t_i) = u_{i1} + u_{i2}$. The increments in prior means of $H_0(t)$ in corresponding intervals are given by $\lambda_{yi1} = \Lambda_0(y_i) - \Lambda_0(t_i)$, $\lambda_{yi2} = \Lambda_0(t_{i+1}) - \Lambda_0(y_i)$, and $\lambda_{ti} = \Lambda_0(t_{i+1}) - \Lambda_0(t_i) = \lambda_{yi1} + \lambda_{yi2}$. The joint survival function given the vector (β, u_1, u_2) is

$$\text{pr}(T > Y \mid u_1, u_2, \beta) = \prod_{i=1}^n \exp(-y_i \theta_i - r_i u_{i1} - r_{i+1} u_{i2}),$$

where $\theta_i = \beta^T x_i$ and $Y = (y_1, \dots, y_n)$. Let $g_1(u_{i1})$ and $g_2(u_{i2})$ be the independent gamma prior densities of u_{i1} and u_{i2} , a consequence of the gamma prior process of $H_0(t)$. The joint distribution of the observed data vector $D = \{y_1, \delta_1, \dots, y_n, \delta_n\}$ and $U = (u_1, \dots, u_{n+1})$ is given by

$$f(D, U | \gamma, \beta) = \prod_{i=1}^n \left(\frac{\partial}{\partial y_i} \right)^{\delta_i} \iint \exp(-y_i \theta_i - r_i u_{i1} - r_{i+1} u_{i2}) g_1(u_{i1}) g_2(u_{i2}) du_{i1} du_{i2},$$

where g_1 and g_2 are gamma densities, the integration is taken over the set $u_i = u_{i1} + u_{i2}$, and the derivative is taken where $\delta_i = 1$, where $\lambda_{yi1} = \Lambda_0(t_i) - \Lambda_0(y_i)$ and $\lambda_{yi2} = \Lambda_0(y_i) - \Lambda_0(t_i)$. Using this joint density, we can show that $E\{\exp(-\sum_{i=1}^n w_i u_i) | D, \beta, \gamma\} = \int \exp(-\sum_{i=1}^n w_i u_i) f(D, U | \gamma, \beta) dU$ is given by

$$\begin{aligned} & E \left\{ \exp \left(- \sum_{i=1}^n w_i u_i \right) | D, \beta, \gamma \right\} \\ &= \prod_{i=1}^n \left(\frac{\partial}{\partial y} \right)^{\delta_i} \left\{ \exp(-y_i \theta_i) \left(\frac{\alpha}{\alpha + r_i + w_i} \right)^{\alpha \lambda_{yi1}} \left(\frac{\alpha}{\alpha + r_{i+1} + w_i} \right)^{\alpha \lambda_{yi2}} \right\} \\ &= \prod_{i=1}^n \left\{ \exp(-y_i \theta_i) \left(\frac{\alpha}{\alpha + r_i + w_i} \right)^{\alpha \lambda_{yi1}} \left(\frac{\alpha}{\alpha + r_{i+1} + w_i} \right)^{\alpha \lambda_{yi2}} \right\} \\ &\quad \times \left\{ \theta_i + \alpha \lambda_0(y_i) \log \left(\frac{\alpha + r_i + w_i}{\alpha + r_{i+1} + w_i} \right) \right\}^{\delta_i}. \end{aligned}$$

The Laplace transform of $p\{H_0(t) | \gamma, \beta, D\}$ is obtained by dividing $\int \exp(-\sum_{i=1}^n w_i u_i) f(D, U | \beta, \gamma) du$ by the marginal density of the observed data, given by $L(\beta, \gamma | D, \alpha)$ in equation (5). After some algebra, we can show that the Laplace transform can be expressed as

$$\begin{aligned} & E \left\{ \exp \left(- \sum_{i=1}^n w_i u_i \right) | D, \beta, \gamma \right\} \\ &= \prod_{i=1}^n \left(\frac{\rho_i}{\rho_i + w_i} \right)^{\alpha \lambda_{ti1}} \left(\frac{\rho_{i+1}}{\rho_{i+1} + w_i} \right)^{\alpha \lambda_{ti2}} \left\{ \frac{\theta_i + \alpha \lambda_0(y_i) \log \left(\frac{\rho_i + w_i}{\rho_{i+1} + w_i} \right)}{\theta_i + \alpha \lambda_0(y_i) \log \left(\frac{\rho_i}{\rho_{i+1}} \right)} \right\}^{\delta_i}, \quad (7) \end{aligned}$$

where $\rho_i = \alpha + r_i$. The first two terms of the right-hand side of (7) correspond to Laplace transforms of independent gamma variables. The third term of (7) corresponds to $B_i + \phi_i(w_i)$, where $\phi_i(w)$ is the Laplace transform of another independent random variable with density proportional to $x^{-1} \{\exp(-\rho_{i+1}x) - \exp(-\rho_i x)\}$ supported on the interval (ρ_{i+1}, ρ_i) and

$$B_i = \frac{\theta_i}{\theta_i + \alpha \lambda_0(y_i) \log(\rho_i / \rho_{i+1})}.$$

It is difficult to express the posterior process of $H_0(t)$ but we can express its Laplace transform using equation (7). For any $t \in (y_i, y_{i+1})$, the Laplace transform of posterior density

$p\{H_0(t) | \gamma, \beta, Y\}$ is

$$E[\exp\{-wH_0(t)\} | D, \beta, \gamma] = \left(\frac{\rho_i}{\rho_{i+1} + w}\right)^{\alpha\lambda_{0ti}} \prod_{j=1}^i \left(\frac{\rho_j}{\rho_j + w}\right)^{\alpha\lambda_j} \left\{ \frac{\theta_j + \alpha\lambda_0(y_j) \log\left(\frac{\rho_j + w}{\rho_{j+1} + w}\right)}{\theta_j + \alpha\lambda_0(y_j) \log\left(\frac{\rho_j}{\rho_{j+1}}\right)} \right\}^{\delta_j}, \quad (8)$$

where $\lambda_{0ti} = \Lambda_0(t) - \Lambda_0(y_i)$. Formula (8) can be used to find the posterior moments of $H_0(t)$ and $S_0(t) = \exp\{-H_0(t)\}$ as follows,

$$E\{S_0(t) | D, \beta, \gamma\} = \left(\frac{\rho_{i+1}}{\rho_{i+1} + 1}\right)^{\alpha\lambda_{0ti}} \prod_{j=1}^i \left(\frac{\rho_j}{\rho_j + 1}\right)^{\alpha\lambda_j} \left\{ \frac{\theta_j + \alpha\lambda_0(y_j) \log\left(\frac{\rho_j + 1}{\rho_{j+1} + 1}\right)}{\theta_j + \alpha\lambda_0(y_j) \log\left(\frac{\rho_j}{\rho_{j+1}}\right)} \right\}^{\delta_j}, \quad (9)$$

$$E[\{S_0(t)\}^2 | D, \beta, \gamma] = \left(\frac{\rho_{i+1}}{\rho_{i+1} + 2}\right)^{\alpha\lambda_{0ti}} \prod_{j=1}^i \left(\frac{\rho_j}{\rho_j + 2}\right)^{\alpha\lambda_j} \left\{ \frac{\theta_j + \alpha\lambda_0(y_j) \log\left(\frac{\rho_j + 2}{\rho_{j+1} + 2}\right)}{\theta_j + \alpha\lambda_0(y_j) \log\left(\frac{\rho_j}{\rho_{j+1}}\right)} \right\}^{\delta_j}, \quad (10)$$

and empirical Bayes estimators of $S_0(t | x) = \exp\{-H_0(t) - t\beta^T x\} = S_0(t) \exp(-t\beta^T x)$ can be obtained as

$$\hat{S}(t | x) = E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\} = \exp(-t\hat{\beta}^T x) E\{S_0(t) | D, \hat{\beta}, \hat{\gamma}\}, \quad (11)$$

where the second term is obtained by inserting the estimators $\hat{\beta}$ and $\hat{\gamma}$ into equation (9). The $\hat{S}(t | x)$ of (11) is always a nondecreasing function of time. An empirical Bayes estimator of $E[\text{var}\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\}]$ is obtained by using

$$\text{var}\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\} = \exp(-2t\hat{\beta}^T x) (E[\{S_0(t)\}^2 | D, \hat{\beta}, \hat{\gamma}] - [E\{S_0(t) | D, \hat{\beta}, \hat{\gamma}\}]^2), \quad (12)$$

where the last two terms can be computed using equations (9) and (10). To compute the estimate $\hat{V}[E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\}]$ of $\text{var}[E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\} | D]$, one can employ a Monte Carlo approximation using the sampling variance of $\hat{S}(t | x, \hat{\beta}_k, \hat{\gamma}_k)$ for $k = 1, \dots, M$, where $(\hat{\beta}_k, \hat{\gamma}_k)$ are sampled from the joint normal distribution $N\{(\hat{\beta}, \hat{\gamma}), V\}$ defined in § 2. For the special case $\lambda_0(t) = \gamma$, we can obtain a closed-form expression of the estimator of $\text{var}[E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\} | D]$ as

$$\hat{V}[E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\}] = \{\hat{S}(t | x)\}^2 \left\{ v_2 \left(\sum_{j=1}^{i+1} \Delta_j c_{j\alpha} \right)^2 + t^2 v_1 + 2v_{12} \sum_{j=1}^{i+1} \Delta_j c_{j\alpha} \right\}, \quad (13)$$

where $\Delta_j = y_j - y_{j-1}$, for $j = 1, \dots, i$, $\Delta_{i+1} = t - y_i$ and v_1 , v_2 and v_{12} are elements of the variance-covariance matrix V of § 2 derived by inverting the observed information matrix of the integrated likelihood in (5). The expression in equation (13) is derived by first evaluating $\text{var}[\log E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\}]$ from equation (9), and then using the delta method. However, the last term of (9) is close to unity, and its logarithm can be neglected for the purpose of this computation. Our approximation results in some underestimation of $\hat{V}[E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\}]$. Finally, the estimated variance of the empirical Bayes estimator $\hat{S}(t | x, \hat{\beta}, \hat{\gamma})$ is given by

$$\text{var}\{\hat{S}(t | x)\} = \hat{V}[E\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\}] + \text{var}\{S(t | x) | D, \hat{\beta}, \hat{\gamma}\},$$

where the two terms of this estimator are computed from (12) and (13).

Table 1. *Melanoma data. Regression parameter estimates, $\hat{\beta}$, and corresponding 95% interval estimates of gender effect using empirical Bayes estimators for different levels of confidence, α ; using the full Bayes estimator (Dunson & Herring, 2005) under a noninformative prior; and using the method-of-moments estimator*

α	$\hat{\gamma}_0$	$\hat{\beta} \times 10^{-2}$	95% Interval for $\hat{\beta} \times 10^{-2}$
10 000	0.0356	3.38	(0.56, 6.20)
1000	0.0365	3.73	(0.97, 6.50)
100	0.0466	4.04	(0.25, 6.64)
10	0.1189	4.12	(0.26, 8.64)
1	0.5423	4.51	(0.20, 7.02)
0.5	3.2937	5.65	(0.82, 8.16)
0.1	23.3551	5.72	(0.82, 8.23)
FB	—	1.99	(0.24, 5.43)
MME	—	3.34	(0.04, 5.42)

FB, full Bayes; MME, method-of-moments estimator.

4. DATA ANALYSIS AND SIMULATION STUDY

We analyzed a prospective study of 205 patients with malignant melanoma undergoing a radical operation performed during the period 1962–1977 at the Department of Plastic Surgery of the University Hospital of Odense, Denmark (Drzewiecki et al., 1980). There were 57 observed deaths due to melanoma. We used gender as the only covariate for our data analysis to assess its effect on survival. Using the parametric additive hazards model with an underlying exponential distribution, equivalent to $\alpha \rightarrow \infty$ in the empirical Bayes approach, we found the maximum likelihood estimate $\hat{\beta} = 0.0333$ with estimated standard error 0.0144.

We reanalyzed the data using a semiparametric additive hazards regression model with unknown constant γ as the prior mean for $h_0(t)$. To maximize the integrated likelihood of (5) in § 2, we used the Poisson estimating equation in Proc Genmod in SAS. However, our method can be implemented using any generalized linear model program available in standard statistical packages. For different levels of prior confidence, i.e. different values of α , Table 1 gives the estimates $\hat{\beta}$ and $\hat{\gamma}$ with corresponding interval estimates for β based on the asymptotic normal density of the empirical Bayes estimator $\hat{\beta}$.

As α decreases, the numerical values of the estimate of $\hat{\beta}$ and its standard error change initially and then stabilize. However, the value of $\hat{\gamma}$ and its standard error increase steadily. This indicates that, for a finite t , the estimated prior $\text{Ga}(\alpha\hat{\gamma}t, \alpha)$ of $H_0(t)$ approaches a gamma density with a small scale parameter, signifying large variance, but a moderate value of the shape parameter $\alpha\hat{\gamma}t$. This shows that, in the limit as $\alpha \rightarrow \infty$, the observed data suggest a proper unimodal gamma prior for $H_0(t)$. A gamma prior density with scale parameter close to zero and large variance would have represented an unrealistic prior belief that $H_0(t)$ is either very close to zero or is very large. Table 1 shows that, although the prior process is estimated from the data, the influence of the prior mean model on the empirical Bayes estimates of the regression parameter and of the survival curves, under a particular threshold value of α , is practically negligible. Thus, with this method the observed data essentially determine a suitable prior density for the increment in the baseline hazard $H_0(y_j) - H_0(y_{j-1})$ by determining its prior mean $\gamma \Delta_j$ and prior variance $\gamma \Delta_j / \alpha$, where $\Delta_i = y_i - y_{i-1}$. We do not recommend using any value of α below 1: it represents an unrealistic prior model for $dH_0(t)$ because the variability of the prior guess turns out to be more than that of the mean prior guess. For comparison with the full Bayes method of Dunson & Herring (2005)

Fig. 1. Melanoma data. Plots of empirical Bayes survival curve estimates for males, dashed, and females, solid, for (a) $\alpha = 10\,000$, (b) $\alpha = 100$, (c) $\alpha = 10$ and (d) $\alpha = 1$.

and for the method-of-moments estimator (Lin & Ying, 1994), Table 1 also presents the point and interval estimates of β , the gender effect, for these methods. For the full Bayes method, we have used a noninformative prior process with $\alpha = 0.001$ and one event in each grid interval, following the recommendation of Dunson & Herring (2005). We found that, for moderately large α , the empirical Bayes estimates are very similar to the method-of-moments estimates, for this example, and the full Bayes method with a noninformative prior gives a posterior estimate of a somewhat smaller magnitude, approximately 30%, of the gender effect. The full Bayes method is appropriate when one has very useful and precise prior information, but otherwise we recommend the empirical Bayes method.

For different values of α , the empirical Bayes estimates of survival curves for males and females are presented in Fig. 1. The influence of chosen α on the smoothness of the estimated survival curves is clear. When α is below 5 or so, there is hardly any noticeable effect of α on $\hat{\beta}$ and $\hat{S}(t | x)$. For $\alpha > 1000$, the model is for all practical purposes a parametric exponential model with additive hazard $h(t | x_i) = \gamma + \beta^T x_i$.

Using a Rayleigh density $h(t | x) = 0.05t + 1 - 0.5x$ with binary x as the true model, we have conducted a simulation study to compare the empirical Bayes estimator to the following competitors: the maximum likelihood estimator under the correct assumption $h(t | x) = 0.05t + 1 + \beta x$; the maximum likelihood estimator for a wrong exponential model $h(t | x) = \lambda + \beta x$; the full Bayes estimator with a noninformative prior using $\alpha = 0.001$ and the method-of-moments estimator. We simulated 2000 copies of datasets of sample size $n = 100$ from the Rayleigh density. We repeated the study under 0% censoring, 10% censoring and 25% censoring. We computed

Table 2. Results of the simulation study with true hazard $h(t | x) = 1 + 0.05t + \beta x$, with binary x and $\beta = -0.5$, comparing estimates $\hat{\beta}$ obtained from the maximum likelihood method using the correct $h_0(t)$ and using the incorrect $h_0(t) = \lambda$ from the empirical Bayes methods for different α including estimated α , from the full Bayes method using a noninformative prior with $\alpha = 0.0001$ for $h_0(t)$, and from the method of moments method. Three levels of censoring were considered

Censoring percentage	Methods	Simulated mean of $\hat{\beta}$	Relative bias %	Simulated variance
0	MLCORR	-0.513	2.61	0.052
	MLEXPO	-0.413	-17.46	0.035
	EB, $\alpha = 10$	-0.466	-7.02	0.041
	EB, $\alpha = 500$	-0.535	7.07	0.039
	EB, $\alpha = 5000$	-0.448	-10.40	0.037
	EB, $\hat{\alpha}$	-0.503	0.50	0.038
	FB	-0.516	3.10	0.019
	MME	-0.522	4.43	0.059
10	MLCORR	-0.508	1.62	0.054
	MLEXPO	-0.418	-16.44	0.038
	EB, $\alpha = 10$	-0.471	-5.80	0.021
	EB, $\alpha = 500$	-0.530	7.80	0.043
	EB, $\alpha = 5000$	-0.450	-13.80	0.038
	EB, $\hat{\alpha}$	-0.506	1.30	0.041
	FB	-0.561	12.19	0.025
	MME	-0.517	3.34	0.059
25	MLCORR	-0.520	3.30	0.062
	MLEXPO	-0.438	-12.31	0.046
	EB, $\alpha = 10$	-0.481	-3.80	0.032
	EB, $\alpha = 500$	-0.550	11.30	0.052
	EB, $\alpha = 5000$	-0.450	-9.75	0.045
	EB, $\hat{\alpha}$	-0.520	5.01	0.050
	FB	-0.598	19.57	0.031
	MME	-0.521	4.13	0.070

MLCORR, maximum likelihood with correct baseline hazard; MLEXPO, maximum likelihood with incorrect baseline hazard; EB, empirical Bayes; FB, full Bayes; MME, method of moments estimator.

empirical Bayes estimates of β for different values of α including α estimated by maximizing the integrated likelihood of (6) for different values of α over a grid. For the empirical Bayes estimators, we purposely used a wrong prior guess, namely constant $\lambda_0(t)$ for $h_0(t)$; in practice, after all, we may not correctly guess a priori the true functional form of $h_0(t)$. The performance of the estimate $\hat{\beta}$ of $\beta = -0.5$ for each method has been evaluated in terms of the relative bias, as a percentage and for the simulated variance of the estimate.

The results of the simulation study are summarized in Table 2. The main feature is that the empirical Bayes estimator performs almost as well as the correct-model maximum likelihood estimator when α is either estimated or of moderate size, i.e. about 10. It is clearly preferable to use empirical Bayes methods whenever we are in doubt about the true form of $h_0(t)$, since an incorrectly specified full likelihood can lead to severely biased estimators of β compared to empirical Bayes estimators, particularly when α is estimated. The full Bayes estimators under noninformative priors appear to have a very high level of relative bias compared to empirical Bayes and moment estimators, especially if there is a moderate level of censoring and a noninformative prior. When we cannot elicit effectively a prior opinion, it is safe to use an empirical

Bayes estimator with either a moderately large value of α or even an estimated α . The moment estimators have reasonable relative biases, but the method produces larger variances of the estimates compared to the full Bayes and empirical Bayes methods.

5. TIME-VARYING COVARIATES AND COEFFICIENTS

The empirical Bayesian methodology described here can be easily extended to the case with a vector of time-dependent covariates $x(t)$. We need a simple assumption that each component of $x(t)$ is either continuous or piecewise-continuous with points of discontinuity different from observed survival times $(y_1, \dots, y_n)$. For an additive hazards regression model with time-varying covariates, the joint survival function is given by

$$\begin{aligned} \text{pr}\{T > Y \mid \beta, H_0(\cdot), X\} &= \exp \left[- \sum_{i=1}^n \{H_0(y_i) + \beta^T Z_i(y_i)\} \right] \\ &= \exp \left\{ - \sum_{i=1}^n \left(r_i h_i + \sum_{j \in R} [\beta^T \{Z_j(y_i) - Z_j(y_{i-1})\}] \right) \right\}, \end{aligned}$$

where r_i is the number of subjects in $R_i = \{j : y_j \geq y_i\}$, the risk set at time y_i , and $Z_i(t) = \int_0^t x_i(u) du$ is the vector of integrated covariates. Following calculations similar to those in § 2, we can show that the integrated likelihood of (β, γ) from $\text{pr}(T > Y \mid \beta, \Lambda_0, \alpha)$ is

$$L(\beta, \gamma \mid D, \alpha) \propto \prod_{i=1}^n \exp \{ -\beta^T Z_i(y_i) - \Lambda_0(y_i) c_{i\alpha} \} \{ \beta^T x_i(y_i) + \lambda_0(y_i) c_{i\alpha} \}^{\delta_i}, \quad (14)$$

where $c_{i\alpha} = \alpha \log(\rho_i / \rho_{i+1})$. The integrated likelihood of (14) is very similar to the integrated likelihood of the fixed-covariates model given in equation (5). Following mathematical steps given in § 3 with minor changes in notation, we can also derive the Laplace transform of the posterior density of $p\{H_0(t) \mid \gamma, \beta, D\}$ for the time-varying covariate case as

$$\begin{aligned} E[\exp\{-wH_0(t)\} \mid D, \beta, \gamma] \\ = \left(\frac{\rho_i}{\rho_{i+1} + w} \right)^{\alpha \lambda_{0ti}} \prod_{j=1}^i \left(\frac{\rho_j}{\rho_j + w} \right)^{\alpha \lambda_j} \left\{ \frac{\beta^T x_j(y_j) + \alpha \lambda_0(y_j) \log \left(\frac{\rho_j + w}{\rho_{j+1} + w} \right)}{\beta^T x_j(y_j) + \alpha \lambda_0(y_j) \log \left(\frac{\rho_j}{\rho_{j+1}} \right)} \right\}^{\delta_j} \end{aligned}$$

and the posterior moments of $S_0(t) = \exp\{-H_0(t)\}$ as

$$\begin{aligned} E[\{S_0(t)\}^k \mid D, \beta, \gamma] \\ = \left(\frac{\rho_{i+1}}{\rho_{i+1} + k} \right)^{\alpha \lambda_{0ti}} \prod_{j=1}^i \left(\frac{\rho_j}{\rho_j + k} \right)^{\alpha \lambda_j} \left\{ \frac{\beta^T x_j(y_j) + \alpha \lambda_0(y_j) \log \left(\frac{\rho_j + k}{\rho_{j+1} + k} \right)}{\beta^T x_j(y_j) + \alpha \lambda_0(y_j) \log \left(\frac{\rho_j}{\rho_{j+1}} \right)} \right\}^{\delta_j}. \end{aligned}$$

For $X^*(t) = \{x(u) : u < t\}$ and $Z(t) = \int_0^t x(u) du$, the empirical Bayes estimators of $S\{t \mid X^*(t)\} = \exp\{-H_0(t) - \beta^T Z(t)\} = S_0(t) \exp\{-\beta^T Z(t)\}$ and their variances can be obtained using equations similar to § 3. For brevity, we omit the details. To find the estimators and their standard errors, we need the values of $\{x_i(t), Z_i(t)\}$ only at observed failure times no greater than y_i .

The general form of the additive hazards regression model for the time-varying regression parameter $\beta(t)$ can be expressed as

$$H(t | x) = H_0(t) + B^T(t)x,$$

where $B(t) = \int_0^t \beta(u)du$ is the vector of cumulative regression functions, and $H_0(t) = \int_0^t h_0(u)du$ is the cumulative baseline hazard function. Earlier, we discussed the special case, in which the regression parameter β is constant over time, that is $B(t) = t\beta$. The joint probability of survival for n subjects given the covariates is

$$\begin{aligned} \text{pr}\{T > Y | \beta, H_0(\cdot), X\} &= \exp \left[- \sum_{i=1}^n \{H_0(y_i) + B^T(y_i)x_i\} \right] \\ &= \exp \left(- \sum_{i=1}^n \left[r_i h_i + \sum_{j \in R_i} \{B^T(y_i) - B^T(y_{i-1})\}x_i \right] \right). \end{aligned}$$

Similarly to the argument used in equation (5), the joint survival function of T given the gamma process can be obtained as

$$\begin{aligned} \text{pr}\{T > Y | \beta, \Lambda_0(\cdot), \alpha\} &= E_{GP}[\text{pr}\{T > Y | B(\cdot), X, H_0(\cdot)\} | \Lambda_0(\cdot), \alpha] \\ &= \prod_{i=1}^n [\exp\{-(A_i - A_{i+1})B(y_i) - \Lambda_0(y_i)c_{i\alpha}\}], \end{aligned}$$

where $c_{i\alpha} = \alpha \log(\rho_i/\rho_{i+1})$ and $A_i = \sum_{j \in R_i} x_j$. By differentiating $\text{pr}(T > Y | B, \Lambda_0, \alpha)$ with respect to each y_i with $\delta_i=1$, we obtain the integrated likelihood

$$\begin{aligned} L(B, \gamma | D, \alpha) &\propto \prod_{i=1}^n [\exp\{-(A_i - A_{i+1})B(y_i) - \alpha \Lambda_0(y_i)c_{i\alpha}\} \\ &\quad \times \{(A_i - A_{i+1})dB(y_i) + \lambda_0(y_i)c_{i\alpha}\}^{\delta_i}], \end{aligned} \quad (15)$$

and when there are no ties in the data and $\lambda_0(t) = \gamma$, i.e. constant, maximizing the likelihood in (15) can be simplified as maximizing the likelihood

$$L(B, \gamma | D, \alpha) \propto \prod_{i=1}^n [\exp\{-\delta_i A_i dB(y_i) - \gamma y_i c_{i\alpha}\} \{x_i^T dB(y_i) + \gamma c_{i\alpha}\}^{\delta_i}],$$

where $dB(y_i)$ is the increment in $B(t)$ at y_i . For any $\gamma > 0$, the increments of $B(t)$ are estimated only at y_i values with $\delta_i = 1$.

6. ASYMPTOTIC PROPERTIES

Asymptotic properties of parametric empirical Bayes procedures are well documented in the statistical literature (Robbins, 1956, 1980; Morris, 1983). To present interval estimates of β and other quantities, these properties need to be applicable to semiparametric empirical Bayes estimators. However, a rigorous proof of the asymptotic properties of semiparametric empirical Bayes estimators for a general set-up is not available yet. As a first step, we prove the asymptotic normality of the score equation, with respect to the limiting marginal density of the data vector given (γ, β, α) , for the case of $\lambda_0(t) = \gamma$ and under the assumption that the random censoring variable is distributed as exponential with unknown rate ϕ , independent of survival times and covariates. The prior assumption that $h_0(t)$ is constant puts no restriction on the actual forms

and shapes of the baseline hazard $h_0(t)$ and its empirical Bayes estimator. We have obtained a similar proof of our asymptotic theory for Type-II censoring. However, here we present the proof only for exponentially distributed informative censoring, because in practice this assumption is often made for design purposes and theoretical convenience. Extensions to more general forms of $\lambda_0(t)$ and censoring mechanisms will be topics for future study. The integrated loglikelihood from (6) based on random censoring is given by $l(\beta, \gamma | D, \alpha) = \sum_{i=1}^n -(y_i q_i - \delta_i \log q_i)$, where $q = x_i \beta^\top + \gamma c_{i\alpha}$ and $y_1 < \dots < y_n$ are observed ordered survival/censoring times. The corresponding score functions are given by

$$-\frac{\partial l(\gamma, \beta)}{\partial \beta} = \sum_{i=1}^n (y_i x_i - \delta_i q_i^{-1} x_i), \quad -\frac{\partial l(\gamma, \beta)}{\partial \gamma} = \sum_{i=1}^n (y_i c_{i\alpha} - \delta_i q_i^{-1} c_{i\alpha}).$$

Let $z_i = y_i - y_{i-1}$ for $i = 1, \dots, n$, with $y_0 = 0$. We can show that the pairs (z_i, δ_i) for $i = 1, \dots, n$ are independently distributed with $\delta_i \sim \text{Ber}(q_i/Q_i)$ and $z_i \sim \text{Ex}(Q_i)$ with $Q_i = \phi(n - i + 1) + \sum_{j=i}^n q_j$. We can rewrite the score function in terms of $(z_1, \delta_1, \dots, z_n, \delta_n)$, as

$$\sum_{i=1}^n (y_i - \delta_i q_i^{-1}) W_i = \sum_{i=1}^n \{(n - i + 1)z_i - \delta_i q_i^{-1}\} W_i, \quad (16)$$

where $W_i = (x_i^\top c_{i\alpha})$. To show the asymptotic normality of the standardized score vector of (16) using the Cramér–Wold device, it is sufficient to show the asymptotic normality of $\sum_{i=1}^n \sum_{i=1}^n \{(n - i + 1)z_i - \delta_i q_i^{-1}\} W_i^\top a$ for every nonzero vector $a = (a_1, a_2)^\top$. To this end, using the C_Δ -inequality, we can first show that

$$\begin{aligned} & \sum_{i=1}^n E \{ |(n - i + 1)z_i - \delta_i q_i^{-1}\} W_i^\top a |^{2+\Delta} \\ & \leq \sum_{i=1}^n 2^{1+\Delta} |W_i^\top a| [E \{ |(n - i + 1)z_i|^{2+\Delta} \} + q_i^{-2-\Delta} E \{ |\delta_i|^{2+\Delta} \}] \end{aligned}$$

for any $\Delta > 0$. Using some mild regularity conditions, such as uniform boundedness of $|x_{ik}|$, and some inequality arguments, we can show that the above expression is $O_e(n)$, where O_e denotes the exact order. Again, we can further show that

$$\begin{aligned} & \text{var} \left[\sum_{i=1}^n \{(n - i + 1)z_i - \delta_i q_i^{-1}\} W_i^\top a \right] \\ & = \sum_{i=1}^n (W_i^\top a)^2 \left\{ Q_i^{-2} (n - i + 1)^2 + \frac{(n - i + 1)\phi}{Q_i q_i} \right\} = O_e(n). \end{aligned}$$

These two results show that

$$\frac{\sum_{i=1}^n E \{ |(n - i + 1)z_i - \delta_i q_i^{-1}|^{2+\Delta} \} |W_i^\top a|^{2+\Delta}}{(\text{var} [\sum_{i=1}^n \{(n - i + 1)z_i - \delta_i q_i^{-1}\} W_i^\top a])^{1+\Delta/2}} = O_e(n^{-\Delta/2}) \rightarrow 0, \quad (17)$$

for every $\Delta > 0$ as $n \rightarrow \infty$. The result of (17) verifies the Liapounov condition for the asymptotic normality of $\sum_{i=1}^n \{(n - i + 1)z_i - \delta_i q_i^{-1}\} W_i^\top a$. For brevity, we again omit the details of the proof of (17). Following the proof of the asymptotic normality of the score function, a straightforward application of a Taylor series expansion gives us the proof of the asymptotic normality of the empirical Bayes estimator $(\hat{\beta}, \hat{\gamma})$ with asymptotic variances consistently estimated using the observed information matrix derived in § 2.

7. DISCUSSION

Various authors, notably Samaniego & Reneau (1994) and Samaniego & Neath (1996), have argued for the need for proper centring, i.e. selection, and assessment of the prior to ensure superior performance of a Bayes procedure compared to the frequentist approach with respect to expected mean squared error. They have advocated the use of the empirical Bayes procedure as a practical and sensible tool for selecting such priors. We have presented a very convenient procedure for performing the inference under Aalen's model when the class of priors of $\{H_0(t), \beta\}$ belongs to a class Π_α characterized by the user-specified confidence parameter α . The restriction we put on this class Π_α is that, for any prior $\pi \in \Pi_\alpha$, the marginal posterior of $H_0(t)$ is a gamma process with known precision parameter α and the prior for β is constant with support defined appropriately to make the joint prior π valid. Like the similarities between the parametric empirical Bayes procedure of Morris (1983) and a fully Bayesian parametric method, our approach has a close relationship with a full semiparametric Bayesian analysis because we can regard the latter as a special case of empirical Bayes when the class Π is assumed to have only one member π , fully specified by the user. Following Berger (1984), we can call our method a robust semiparametric Bayes procedure. However, the rigorous theoretical justification for using such a term to describe our estimators is beyond the scope of this paper. Another topic for future investigation is the asymptotic optimality properties of the expected mean square $r_0(\pi_0, \hat{\beta}) = E[\int_y \{\hat{\beta}(y) - \beta\}^2 dy | \pi_0]$ of the empirical Bayes estimator (Samaniego & Reneau, 1994; Carlin & Louis, 2000a,b), where the expectation is taken with respect to the true but unknown joint prior $\pi_0(H_0, \beta)$ and the integration is over the sampling distribution of observed data vector y given (H_0, β) . An important special case is when the prior $\pi_0(H_0, \beta)$ is equivalent to the prior density $\pi_0(\beta | H_0^*)$, where H_0^* is the unknown fixed cumulative baseline hazard.

ACKNOWLEDGEMENT

The authors are grateful for the support provided by grants from the U.S. National Institutes of Health and the U.S. National Science Foundation.

REFERENCES

- AALLEN, O. O. (1980). *A Model for Non-parametric Regression Analysis of Counting Processes. Lecture Notes in Statistics*, **2**, pp. 1–25. New York: Springer.
- AALLEN, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **8**, 907–25.
- AALLEN, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statist. Med.* **12**, 1569–88.
- BEAMONTE, E. & BERMÚDEZ, J. D. (2003). A Bayesian semiparametric analysis for additive hazards models with censored observations. *Test* **12**, 347–63.
- BERGER, J. O. (1984). The robust Bayesian viewpoint. In *Robustness in Bayesian Statistics*, Ed. J. Kadane, pp. 63–124. Amsterdam: North Holland.
- CARLIN, B. P. & LOUIS, T. A. (2000a). *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall.
- CARLIN, B. P. & LOUIS, T. A. (2000b). Empirical Bayes: past, present and future. *J. Am. Statist. Assoc.* **95**, 1286–9.
- COX, D. R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc. B* **34**, 187–220.
- DRZEWIECKI, K. T., LADEFOGED, C. & CHRISTENSEN, H. E. (1980). Biopsy and prognosis for cutaneous malignant melanomas in clinical stage 1. *Scand. J. Plast. Reconstr. Surg.* **14**, 141–4.
- DUNSON, D. B. & HERRING, A. H. (2005). Bayesian model selection and averaging in additive and proportional hazards models. *Lifetime Data Anal.* **11**, 213–32.
- IBRAHIM, J. G., CHEN, M.-H. & SINHA, D. (2001). *Bayesian Survival Analysis*. New York: Springer.
- KALBFLEISCH, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *J. R. Statist. Soc. B* **40**, 214–21.
- LIN D. Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.
- LIN, D. Y., OAKES, D. & YING, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85**, 289–98.

- MARTINUSSEN, T. & SCHEIKE, T. H. (2006). *Dynamic Regression Models for Survival Data*. New York: Springer.
- MORRIS, C. N. (1983). Parametric empirical Bayes inference: theory and applications (with Discussion). *J. Am. Statist. Assoc.* **78**, 47–55.
- ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proc. Third Berkeley Symp. Math. Statist. Prob.*, vol. 1, Ed. J. Neyman, pp. 157–63. Berkeley, CA: University of California Press.
- ROBBINS, H. (1980). An empirical Bayes estimation problem. *Proc. Nat. Acad. Sci.* **77**, 6988–9.
- SAMANIEGO, F. J. & RENEAU, D. M. (1994). Toward a reconciliation of the Bayesian and frequentist approaches to point estimation. *J. Am. Statist. Assoc.* **89**, 947–57.
- SAMANIEGO, F. J. & NEATH, A. A. (1996). How to be a better Bayesian. *J. Am. Statist. Assoc.* **91**, 733–42.
- SINHA, D., IBRAHIM, J. G. & CHEN, M.-H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika* **90**, 629–41.
- YIN, G. & IBRAHIM, J. G. (2005). A class of Bayesian shared Gamma frailty models with multivariate failure time data. *Biometrics* **61**, 208–16.
- ZENG, D., YIN, G. & IBRAHIM, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *J. Am. Statist. Assoc.* **101**, 670–84.

[Received January 2006. Revised November 2008]