NIH Public Access

Author Manuscript

Lifetime Data Anal. Author manuscript; available in PMC 2012 April 9.

Published in final edited form as:

Lifetime Data Anal. 2011 January ; 17(1): 43–70. doi:10.1007/s10985-010-9170-0.

Bayesian local influence for survival models

Joseph G. Ibrahim,
Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA

Hongtu Zhu, and
Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA

Niansheng Tang
Department of Statistics, Yunnan University, Kunming, People's Republic of China

Joseph G. Ibrahim: ibrahim@bios.unc.edu

Abstract

The aim of this paper is to develop a Bayesian local influence method (Zhu et al. 2009, submitted) for assessing minor perturbations to the prior, the sampling distribution, and individual observations in survival analysis. We introduce a perturbation model to characterize simultaneous (or individual) perturbations to the data, the prior distribution, and the sampling distribution. We construct a Bayesian perturbation manifold to the perturbation model and calculate its associated geometric quantities including the metric tensor to characterize the intrinsic structure of the perturbation model (or perturbation scheme). We develop local influence measures based on several objective functions to quantify the degree of various perturbations to statistical models. We carry out several simulation studies and analyze two real data sets to illustrate our Bayesian local influence method in detecting influential observations, and for characterizing the sensitivity to the prior distribution and hazard function.

Keywords

Bayesian local influence; Bayesian perturbation manifold; Perturbed model; Posterior distribution; Prior; Survival model

1 Introduction

Survival data are common in various settings, including medicine, biology, engineering, public health, epidemiology, and econometrics. There is a large literature on developing various statistical models including parametric, semiparametric and nonparametric models for analyzing survival data. See the references in Anderson et al. (1993), Fleming and Harrington (1991), Kalbfleisch and Prentice (2002), Lawless (2003) and Ibrahim et al. (2001). For instance, frailty models are extensions of the proportional hazards model that allow us to model the association between individual survival times within clusters and for modeling multivariate survival data. In addition, cure rate models have been developed to model time-to-event data for various types of cancers, including breast cancer, non-

Correspondence to: Joseph G. Ibrahim, ibrahim@bios.unc.edu.

Hodgkins lymphoma, leukemia, prostate cancer, and some other diseases, where for these diseases, a significant proportion of patients are "cured" (Chen et al. 1999).

Recent advances in computation and prior elicitation have made Bayesian analysis of complex survival models feasible. For instance, nonparametric prior processes including the gamma process, the beta process, the Dirichlet process, mixtures of Dirichlet processes, nested Dirchlet processes, stick-breaking processes, and Polya tree processes have been developed in semiparametric Bayesian inference (Ibrahim et al. 2001; Sinha et al. 2003; Hanson and Johnson 2002; Hanson et al. 2009; Dunson and Park 2008; Rodriguez et al. 2008; De Iorio et al. 2009).

The literature on Bayesian survival analysis is enormous, and too numerous to list here. Ibrahim et al. (2001) give a comprehensive review of Bayesian survival analysis methods up to 2001 and we refer the reader to their book for more details. More recent work includes methods for multivariate survival data (Dunson and Dinse 2002; Yin and Ibrahim 2005a), mixtures of Polya tree priors for accelerated failure time models (Hanson and Johnson 2002), order restricted Bayesian survival analysis (Dunson and Herring 2003; Chen and Dunson 2004), Bayesian model selection and model averaging in survival analysis (Dunson and Herring 2005), methods for missing data in survival models (Chen et al. 2002b, 2006; Ibrahim et al. 2008), Bayesian transformation survival models (Yin and Ibrahim 2005b), cure rate models (Kim et al. 2009; Yin and Ibrahim 2005c,d; Cooner et al. 2007; Chen et al. 2002a,b,c), methods for recurrent events and panel count data (Sinha et al. 2008; Sinha and Maiti 2004), additive hazards models (Sinha et al. 2009), dynamic frailty models for multivariate survival data (Pennell and Dunson 2006), and dependent Dirichlet process models for survival data (De Iorio et al. 2009). Although not specifically mentioned in their papers, applications of kernel stick-breaking processes (Dunson and Park 2008) and nested Dirichlet processes (Rodriguez et al. 2008) to models for survival data are also potentially promising. Moreover, there has been a substantial literature on Bayesian methods for joint modeling of longitudinal and survival data. Some recent papers include Brown and Ibrahim (2003a,b), Ibrahim et al. (2004), Chen et al. (2004), Brown et al. (2005), and Chi and Ibrahim (2006, 2007).

The literature on Bayesian diagnostics mainly addresses methods based on case deletion using the Conditional Predictive Ordinate (CPO) (Geisser 1993; Gelfand et al. 1992; Gelfand and Dey 1994) and the Kullback-Leibler (KL) divergence (Peng and Dey 1995). Considerable research has been done for developing case influence diagnostics using the KL divergence under various parametric models (Johnson and Geisser 1983, 1985; Pettit 1986; Carlin and Polson 1991; Weiss and Cook 1992; Peng and Dey 1995; Weiss 1996; Christensen 1997; Sinha and Dey 1997; Weiss and Cho 1998). Pettit (1986) suggested the use of the KL divergence in detecting influential observations in his review of Bayesian diagnostics. Carlin and Polson (1991) proposed an expected utility approach using the KL divergence as a utility function to define the influence of a set of observations in a parametric modeling framework, considering the normal linear model and mixed models. Weiss and Cook (1992) introduced the KL divergence to assess the divergence between posteriors in the context of case deletion in generalized linear models. Weiss (1996) and Weiss and Cho (1998) proposed assessing the influence of case deletion using model perturbations as well as establishing its relationship to the KL divergence and the CPO. Sinha and Dey (1997) give a review paper on Bayesian methods for survival analysis and discuss Bayesian residuals and goodness of fit. Bayesian influence measures for assessing marginal posterior distributions have also been developed for the multivariate linear model and normal random effects models in Johnson and Geisser (1985) and Weiss and Cho (1998).

Although the above-mentioned methods have not been directly applied to survival models, they are general enough so that their extensions to censored survival data appear straightforward. More recently, Bayesian diagnostic methods for semiparametric survival models using the KL divergence have been examined by Cho et al. (2009).

The extensive literature on Bayesian diagnostic methods for parametric or semiparametric models has primarily focused on case influence diagnostic procedures, and there is essentially no literature in Bayesian survival analysis examining more general diagnostic procedures such as local influence methods, for example (Cook 1986; Zhu et al. 2007; Zhu and Lee 2001; Zhu and Zhang 2004). In Bayesian analysis of survival data, posterior quantities such as the Bayes factor or posterior mean for a given dataset may be sensitive to a small perturbation to any of the three key elements of a Bayesian analysis: the data, the prior, or the sampling distribution. For this reason, sensitivity analyses should be done to check the degree of sensitivity of the parameters of interest with respect to these three key elements of a Bayesian analysis.

In the Bayesian literature, local influence has been developed as a class of sensitivity analysis methods to perturb each of these three key elements of a Bayesian analysis and assess the influence of these perturbations on the posterior distribution and its associated posterior quantities (Berger 1990, 1994; McCulloch 1989; Gustafson 2000; Sivaganesan 2000; Kass et al. 1989; Weiss 1996; Oakley and O'Hagan 2004). The key idea of the local influence approach primarily computes the derivatives of posterior quantities with respect to a small perturbation to the prior and the sampling distribution. McCulloch (1989) generalizes Cook's (1986) local influence approach to assess the effects of perturbing the prior in a Bayesian analysis. Moreover, a nonparametric analogue of Cook's (1986) local influence approach is to calculate the Frechét derivative of the posterior with respect to the prior (Gustafson 1996a,b; Gustafson and Wasserman 1995; Berger 1994; Berger et al. 2000). Zhu et al. (2009) develop a general Bayesian influence method for assessing various perturbations to the prior and the sampling distribution of a class of parametric models while allowing for incomplete data in a Bayesian analysis. According to the best of our knowledge, however, very little has been done on developing a general Bayesian local influence approach for simultaneously perturbing the three components of a Bayesian model, assessing their effects, and examining their applications in Bayesian survival analysis.

The aim of this paper is to develop a Bayesian local influence method to perturb the three components of a Bayesian model and to assess minor perturbations in Bayesian survival analysis. We propose a perturbation model to individually or simultaneously perturb the three components of the Bayesian model. We construct a Bayesian perturbation manifold to characterize the intrinsic structure of the perturbation model and quantify the degree of each perturbation in the perturbation model. We develop local influence measures for selecting the most influential perturbation based on various objective functions including the posterior mean distance and Bayes factor, and examine their statistical properties in Bayesian survival analysis.

This rest of this paper is organized as follows. In Sect. 2, we introduce the general survival model and the perturbation model to characterize various perturbations to the initial model. We construct a perturbation manifold for the perturbation model and derive its associated geometric quantities. We also develop local influence measures to quantify the effects of perturbing the data, the prior and the sampling distribution on the posterior quantities. We present simulation studies and two real data analyses in Sect. 3. We conclude the article with some final discussion in Sect. 4.

2 The general method

2.1 Bayesian survival model

Consider data from n independent clusters, with n_i observations in the i th cluster for $i = 1, \ldots, n$. For the jth observation in the ith cluster ($j = 1, \ldots, n_i$), we observe a possibly right censored event time y_{ij}, censoring indicator ν_{ij}, where $\nu_{ij} = 1$ is y_{ij} is a failure and $\nu_{ij} = 0$ if y_{ij} is right censored, and a $p \times 1$ vector of covariates $\mathbf{x}_{ij}$. There is a total of $N=\sum_{i=1}^{n} n_i$ observations. Within a cluster, the observations may be dependent, but conditional on the cluster-specific latent vector $\mathbf{b}_i = (b_{i1}, \ldots, b_{iq})^T$, (y_{ij}, ν_{ij}) for different j are independent. Let D_{obs} and $D_{mis} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ denote the observed data and missing data, respectively, and $D_{com} = (D_{mis}, D_{obs})$ denotes the complete data.

A formal Bayesian analysis of D_{com} involves the specification of a sampling distribution $p(D_{com} | \boldsymbol{\theta})$ and a prior distribution $p(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_F, \boldsymbol{\theta}_I)$ includes a finite-dimensional parameter vector $\boldsymbol{\theta}_F$ and a vector of infinite-dimensional parameters $\boldsymbol{\theta}_I$, such as the baseline hazard function. For the sampling distribution, we consider a statistical model $p(D_{com} | \boldsymbol{\theta})$ such that

$$p(D_{com}|\theta)=\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}, \nu_{ij}|\mathbf{b}_i, \theta)p(\mathbf{b}_i|\theta_F) \quad (1)$$

and $p(D_{obs} | \boldsymbol{\theta}) = \int p(D_{com} | \boldsymbol{\theta}) d\, D_{mis}$. Without the presence of $\boldsymbol{\theta}_I$, model (1) includes various parametric models for survival data (Ibrahim et al. 2001; Lawless 2003; Kalbfleisch and Prentice 2002). This class of survival models also includes proportional hazard models, frailty models, cure rate models, and many other parametric and semiparametric models. We mention here that for ease of exposition, we have assumed that the distribution of $\mathbf{b}_i$ is parametric and hence its parameters are finite dimensional. However, our methodology is quite general and can still be applied when the distribution of $\mathbf{b}_i$ is nonparametric such as a Dirichlet or Polya tree process, where in this case, the parameters of the distribution of $\mathbf{b}_i$ are infinite, and therefore we would write $p(\mathbf{b}_i | \boldsymbol{\theta}_I)$ in (1). A typical joint prior specification is to assume that $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_F)\, p(\boldsymbol{\theta}_I)$. We usually assume parametric prior distributions for the components of $\boldsymbol{\theta}_F$ and nonparametric prior distributions for the components of $\boldsymbol{\theta}_I$. To carry out Bayesian inference, we usually use Markov chain Monte Carlo (MCMC) methods to obtain samples from the posterior distribution of the observed data, $p(\boldsymbol{\theta}|D_{obs})$, given by

$$p(\theta|D_{obs}) \propto \int \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}, \nu_{ij}|\mathbf{b}_i, \theta)p(\mathbf{b}_i|\theta_F)p(\theta_I)p(\theta_F)d\left(\prod_{i=1}^{n}\mathbf{b}_i\right). \quad (2)$$

Subsequently, we can calculate posterior quantities of $\boldsymbol{\theta}$, such as the posterior mean of $d(\boldsymbol{\theta}_F)$, where $d(\cdot)$ is an arbitrary function.

For the purposes of illustration, we consider the following examples.

Example 1 (Proportional hazard model)—We consider Bayesian analysis of the proportional hazards model with right censored data (Cox 1972; Ibrahim et al. 2001). This model is often referred to as the Cox model. In this case, $n_i = 1$ for all i, $b_i = 0$, $y_{i1} = T_{i1} \wedge C_{i1}$ is the minimum of the censoring time C_{i1} and the survival time T_{i1} and $\nu_{i1} = \mathbf{1}(y_{i1} = T_{i1})$, where $\mathbf{1}(\cdot)$ is an indicator function. The Cox model assumes that the conditional hazard function of y_{i1} given $\mathbf{x}_{i1}$ is given by

$$h(y|\mathbf{x}_{i1})=h_0(y)\exp\left(\mathbf{x}_{i1}^T\beta\right), \quad (3)$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and $h_0(\cdot)$ is an unknown baseline hazard function. In this case, $\boldsymbol{\theta}_F = \boldsymbol{\beta}$ and $\boldsymbol{\theta}_I = H_0(\cdot)$, where $H_0(y)=\int_0^y h_0(u)$ is the baseline cumulative hazard function. The full likelihood for the observed data is given by

$$p(D_{obs}|\theta)=\prod_{i=1}^{n}\left[\exp\left(\mathbf{x}_{i1}^T\beta\right)dH_0(y_{i1})\right]^{\nu_{i1}}\exp\left\{-\exp\left(\mathbf{x}_{i1}^T\beta\right)H_0(y_{i1})\right\}. \quad (4)$$

Example 2 (Shared-frailty model)—We consider the commonly used shared-frailty model with right censored data (Vaupel et al. 1979). Here we have $q = 1$ and ν_{ij} takes the value 1 if y_{ij} is a failure time or zero if y_{ij} is right censored. It also assumes that the conditional hazard function of y_{ij} given the latent frailty random variable b_i for the ith cluster and $\mathbf{x}_{ij}$ is given by

$$h(y|\mathbf{x}_{ij}, b_i)=h_0(y)b_i\exp\left(\mathbf{x}_{ij}^T\beta\right), \quad (5)$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and $h_0(\cdot)$ is an unknown baseline hazard function. It is common to assume a gamma distribution for the latent frailty, that is $b_i \sim \mathcal{G}(\kappa^{-1}, \kappa^{-1})$. In this case, $\boldsymbol{\theta}_F = (\boldsymbol{\beta}, \kappa)$ and $\boldsymbol{\theta}_I = H_0(\cdot)$. The complete-data likelihood function for all subjects is given by

$$p(D_{com}|\theta)=\prod_{i=1}^{n}\left\{b_i^{\kappa^{-1}-1}\exp(-b_i/\kappa)\times\prod_{j=1}^{n_i}\left[b_i\exp\left(\mathbf{x}_{ij}^T\beta\right)dH_0\left(y_{ij}\right)\right]^{\nu_{ij}}\exp\left(-\exp\left(\mathbf{x}_{ij}^T\beta\right)b_iH_0\left(y_{ij}\right)\right)\right\}. \quad (6)$$

Example 3 (Cure rate model)—We consider a Bayesian analysis of the cure rate model with right censored survival data (Chen et al. 1999; Yakovlev 1994; Ibrahim et al. 2001). We consider here the promotion time cure rate model as discussed in Chen et al. (1999). Suppose we have n subjects, and let N_i denote the number of carcinogenic cells for the ith subject, where the N_i 's are assumed to be *i.i.d.* Poisson random variables with mean $\theta_i=\exp\left(\mathbf{x}_i^T\beta\right)$, $i = 1, \ldots, n$. Further, suppose $Z_{i1}, \ldots, Z_{i,N_i}$ are the *i.i.d* incubation times for the N_i carcinogenic cells for the i th subject, which are unobserved, and all have baseline survival function $S_0(\cdot)$ with baseline hazard function $h_0(\cdot)$, $i = 1, \ldots, n$. The survival time for subject i is $y_{i1} = \min\{Z_{ij}, j = 1, \ldots, N_i\}$, and the indicator $\nu_{i1} = 1$ if y_{i1} is a failure time and 0 if it is right censored. Following Chen et al. (1999), the complete-data likelihood $p(D_{com}|\boldsymbol{\theta})$ is given by

$$\left(\prod_{i=1}^{n}S_0(y_{i1})^{N_i-\nu_{i1}}\left[N_ih_0(y_{i1})\right]^{\nu_{i1}}\right)\exp\left\{\sum_{i=1}^{n}\left[N_i\mathbf{x}_{i1}'\beta-\log(N_i!)-\exp(\mathbf{x}_{i1}'\beta)\right]\right\}. \quad (7)$$

In this case, $\boldsymbol{\theta}_F = \boldsymbol{\beta}$ and $\boldsymbol{\theta}_I = H_0(\cdot)$. Alternative parametric, semiparametric, and multivariate cure rate models have been developed for modeling time-to-event data for various type of cancers, such as melanoma and breast cancer (Ibrahim et al. 2001).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Example 4 (Prior distributions)—To carry out a Bayesian analysis of the models in Examples 1–3, we take a joint prior distribution for $\boldsymbol{\theta}$ as follows. First we take independent priors for $\boldsymbol{\beta}$, κ, H_0, so that $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})\, p(\kappa)\, p(H_0)$. For κ and $\boldsymbol{\beta}$, we may assume $\kappa \sim \mathcal{G}(\varphi_1, \varphi_2)$ and $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where $\mathcal{G}(\varphi_1, \varphi_2)$ denotes the gamma distribution with shape parameter $\varphi_1 > 0$ and scale parameter $\varphi_2 > 0$, and $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denotes the multivariate normal distribution with $p \times 1$ mean vector $\boldsymbol{\mu}_0$ and $p \times p$ covariance matrix $\boldsymbol{\Sigma}_0$. If the smallest eigenvalue $\lambda_{min}(\boldsymbol{\Sigma}_0)$ converges to ∞, then $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ tends to an improper prior. In contrast, if the largest eigenvalue $\lambda_{max}(\boldsymbol{\Sigma}_0)$ is very small, then $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ tends to a strongly informative prior.

We can take different prior distributions including a piecewise constant hazards model, Gamma process model, Beta process model, or a Dirichlet process model for the baseline hazard $h_0(\cdot)$ or cumulative baseline hazard $H_0(\cdot)$. One of the most convenient and popular specifications for $h_0(\cdot)$ is the piecewise constant hazards model. To construct this model, we first construct a finite partition of the time axis, $0 < s_1 < s_2 < \cdots < s_J$, with $s_J > y_{ij}$ for all i, j, which leads to J intervals $(0, s_1], \ldots, (s_{J-1}, s_J]$. In the jth interval, we assume $h_0(y) = \lambda_j$ for $y \in I_j = (s_{j-1}, s_j]$. A common prior for $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)^T$ is the independent gamma prior $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \alpha_{1j})$ for $j = 1, \ldots, J$, where α_{0j} and α_{1j} are prior hyperparameters. Another approach is to build a priori correlation among the λ_j's (Leonard 1978; Sinha 1993; Arjas and Gasbarra 1994; Ibrahim et al. 2001) using correlated priors for $\boldsymbol{\lambda}$. For instance, Arjas and Gasbarra (1994) proposed a first-order autoregressive model on the λ_j's by taking $\lambda_k | \lambda_{k-1} \sim \mathcal{G}(\alpha_k, \alpha_k/\lambda_{k-1})$ for $k > 1$. Another common approach to building correlation in the hazard is to define $\psi_i = \log(\lambda_j)$, $j = 1, \ldots, J$, and then specify a multivariate normal prior for $\boldsymbol{\psi}$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_J)^T$.

We may also consider a gamma process prior for $H_0(\cdot)$, that is, $H_0 \sim \mathcal{G}P(c_0 H^*, c_0)$ (Kalbfleisch 1978), where c_0 is a fixed constant and $H^*(\cdot)$ is a known increasing function with $H^*(0) = 0$. $H^*(\cdot)$ can be viewed as a parametric guess for the unknown cumulative baseline hazard function $H_0(\cdot)$. For example if $H^*(\cdot)$ is a Weibull distribution, then $H^*(y) = \gamma_0 y^{k_0}$, in which (γ_0, k_0) are specified hyperparameters. Thus, $H^*(\cdot)$ is the mean of the process and $H_0(\cdot)$ is a stochastic process with the properties: $H_0(0) = 0$; $H_0(t)$, $t > 0$, has independent increments in disjoint intervals; and for $t > s$, $H_0(t) - H_0(s) \sim \mathcal{G}(c_0(H^*(t) - H^*(s)), c_0)$. An alternative approach is to specify a gamma process prior on the baseline hazard function rate (Dykstra and Laud 1981; Ibrahim et al. 2001).

Our main interest is to make valid Bayesian inferences about $\boldsymbol{\theta}$, and this requires a reasonably robust prior $p(\boldsymbol{\theta})$ and the correct specification of the sampling distribution. A non-robust prior for $p(\boldsymbol{\theta})$, the presence of outliers, and misspecifying some of those modeling assumptions may introduce serious bias in the estimation and inference regarding $\boldsymbol{\beta}$. Thus, it is crucial to assess the robustness of both the prior and the sampling distribution as well as the identification of outliers. According to the best of our knowledge, no sensitivity analyses involving all three of the components of a Bayesian model have been carried out for the survival models considered here.

2.2 Perturbation model

We develop a class of perturbation models to characterize various perturbation schemes to perturb the data, the prior, and the sampling distribution as follows:

$$p(D_{com}, \theta | \omega) = p(\theta, \omega_P(\theta)) \prod_{i=1}^{n} \times \left\{ \left[\prod_{j=1}^{n_i} p(y_{ij}, \nu_{ij} | \mathbf{b}_i, \theta, \omega_D, \omega_S) \right] p(\mathbf{b}_i, \theta_F, \omega_D, \omega_S) \right\} \quad (8)$$

and $\int p(D_{com}, \boldsymbol{\theta}|\boldsymbol{\omega})d\, D_{com}\, d\boldsymbol{\theta} = 1$, where $\boldsymbol{\omega}_P \in R^{mP}$, $\boldsymbol{\omega}_S \in R^{mS}$, and $\boldsymbol{\omega}_D \in R^{mD}$ represent perturbations to the prior, the sampling distribution, and the data, respectively. Moreover, let $m = m_P + m_S + m_D$, and we assume that $\omega^0 = \left(\omega_P^0, \omega_S^0, \omega_D^0\right) \in R^m$ represents no perturbation. For instance, we may only perturb individual observations (or clusters) in order to identify influential observations (or clusters). Specifically, the likelihood function $p(D_{com}\,|\boldsymbol{\theta}, \boldsymbol{\omega}_D)$ for perturbing the data is defined by

$$\prod_{i=1}^{n}\prod_{j=1}^{n_i} p\left(y_{ij}, \nu_{ij}|\mathbf{b}_i, \theta, \omega_{D,i}\right)\; p\;\left(\mathbf{b}_i, \theta_F|\omega_{D,i}\right), \tag{9}$$

where $\boldsymbol{\omega}_{d,i}$ represents the perturbation vector to the observations in the ith cluster. To assess the sensitivity of model assumptions to a small perturbation, we usually surround $p(D_{com}\,|\boldsymbol{\theta})$ by a class of distributions $p(D_{com}\,|\boldsymbol{\theta}, \boldsymbol{\omega}_S)$ such that $p(D_{obs}\,|\boldsymbol{\theta}, \boldsymbol{\omega}_S) = \int p(D_{com}\,|\boldsymbol{\theta}, \boldsymbol{\omega}_S)d\, D_{mis}$, $p\left(D_{com}|\theta, \omega_S^0\right) = p(D_{com}|\theta)$, and $p\left(D_{obs}|\theta, \omega_S^0\right) = p(D_{obs}|\theta)$. We may statistically assess whether $\omega_S = \omega_S^0$ is valid.

As an illustration, we examine various perturbations to the proportional hazards and shared-frailty models.

Example 1 (continued)—To carry out the Bayesian analysis, we specify independent priors given by $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})\, p(H_0)$, where $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $H_0 \sim \mathscr{G}P(c_0 H^*, c_0)$ as discussed in Example 4. We consider a prior perturbation to H_0 by assuming $H_0 \sim \mathscr{G}P\left(c_0 H_k^*(\omega_P), c_0\right)$, in which $H_k^*(y, \omega_P) = \int_0^y k_*(t, \omega_P) h_0(t) dt$, and a local perturbation to $h(y|\mathbf{x}_{i1})$ by assuming that $h(y|\mathbf{x}_{i1}, \omega_D) = h_0(y)\exp\left(\mu_2(\mathbf{x}_{i1}^T\beta, \omega_{D,i})\right)$, where $\mu_2\left(\mathbf{x}_{i1}^T\beta, \omega_{D,i}\right)$ are differentiable functions of $\boldsymbol{\omega}_{D,i}$ and $\mu_2\left(\mathbf{x}_{i1}^T\beta, 0\right) = \mathbf{x}_{i1}^T\beta$. Without loss of generality, we assume that the y_{i1}'s are distinct and $y_{(1)1} < \cdots < y_{(n)1}$ denote the ordered failure or censoring times. Following the derivation in Sinha et al. (2003), the perturbed posterior distribution of $\boldsymbol{\beta}$ is given by

$$p(\beta|D_{com}, \omega) \propto \varphi(\beta|\mu_0,$$
$$\Sigma_0)\prod_{j=1}^{n}\left[\left(\frac{c_0}{c_0 + A_j(\omega_D;\beta)}\right)^{c_0\left[H_k^*(y_{(j)1},\omega_P) - H_k^*(y_{(j-1)1},\omega_P)\right]} \times \left(-c_0 h_0(y_{(j)1}) k_*(y_{(j)1}, \omega_P)\log\left(\frac{c_0 + A_{j+1}(\omega_D;\beta)}{c_0 + A_j(\omega_D;\beta)}\right)\right)^{\nu_{(j)1}}\right], \tag{10}$$

where φ $(\boldsymbol{\beta}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denotes the multivariate normal density with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, $A_j(\omega_D;\beta) = \sum_{y_{l1} \geq y_{(j)1}} \exp\left(\mu_2(\mathbf{x}_{l1}^T\beta, \omega_{D,l})\right)$, and $\nu_{(j)1}$ is the censoring indicator for the jth ordered survival time $y_{(j)1}$.

Example 2 (continued)—We consider a data perturbation to $h(y|\mathbf{x}_{ij}, b_i)$ by assuming

$$h(y|\mathbf{x}_{ij}, b_i) = h_0(y) b_i \exp\left(\mu_2(\mathbf{x}_{ij}^T\beta, \omega_{D,i})\right). \tag{11}$$

If $\mu_2\left(\mathbf{x}_{ij}^T\beta, \mathbf{0}\right) = \mathbf{x}_{ij}^T\beta$, then $\omega_{D,1}^0 = \cdots = \omega_{D,n}^0 = 0$ represents no perturbation. Thus, the sampling distribution for the data perturbation, $p(D_{com}\,|\boldsymbol{\theta}, \boldsymbol{\omega}_D)$, is given by

$$p(D_{com}|\theta, \omega_D) \propto \prod_{i=1}^{n} \left(b_i^{\kappa^{-1}-1} \exp(-b_i/\kappa) \prod_{j=1}^{n_i} \left\{ \left[b_i \exp\left(\mu_2\left(\mathbf{x}_{ij}^T \beta, \omega_{D,i}\right)\right) h_0(y_{ij}) \right]^{\nu_{ij}} \times \exp\left\{ -b_i \exp\left(\mu_2\left(\mathbf{x}_{ij}^T \beta, \omega_{D,i}\right)\right) H_0(y_{ij}) \right\} \right\} \right). \quad (12)$$

The perturbation to the prior $\boldsymbol{\omega}_P$ ($\boldsymbol{\theta}$) includes the additive ε–contamination class and the geometric contamination class as special cases (Berger 1990, 1994; Gustafson and Wasserman 1995; Moreno 2000). We examine perturbations to the prior distribution of $\boldsymbol{\theta}$ in Example 4.

Example 4 (continued)—We perturb $p(\boldsymbol{\beta})$ by assuming that $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_0 + \boldsymbol{\omega}_{P,1}, \omega_{P,2} \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\omega}_{P,1} \in R^p$ and $\omega_{P,2} \geq 0$ is a positive scalar. Thus, $\omega_P(\theta) = \left(\omega_{P,1}^{'}, \omega_{P,2}\right)^{'} \in R^p \times [0, \infty)$ is independent of $\boldsymbol{\theta}$, and thus $\omega_P^0(\theta) = \left(\mathbf{0}_p^{'}, 1\right)^{'}$ represents no perturbation. Consider the additive ε–contamination class given by $p(\boldsymbol{\theta}; \boldsymbol{\omega}_P(\boldsymbol{\theta})) = p(\boldsymbol{\theta}) + \boldsymbol{\omega}_P [g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})]$, where $\boldsymbol{\omega}_P \in [0, 1]$ and $g(\boldsymbol{\theta})$ is a contamination distribution (Berger 1994). We may assume that the density of $p(\boldsymbol{\beta})$ can be approximated by

$$p(\beta) = P(\beta; \omega_P, k)^2 \varphi(\beta; \mu_0, \sum\nolimits_0),$$

where $P(\boldsymbol{\beta}; \boldsymbol{\omega}_P, k)$ is a multivariate polynomial of order k and $\boldsymbol{\omega}_P$ are coefficients of $P(\boldsymbol{\beta}; \boldsymbol{\omega}_P, k)$ such that $P(\boldsymbol{\beta}; \mathbf{0}, k) = 1$ (Gallant and Nychka 1987). Thus, $\omega_P^0 = \mathbf{0}$ represents no perturbation.

For the piecewise exponential model, we have $h_0(y) = \lambda_j$, where $s_{j-1} < y \leq s_j$, $j = 1, \ldots, J$. We assume that $\lambda_j \sim \mathcal{G}(\omega_{P0,j}\, \alpha_{0\,j}, \omega_{P1,j}\, \alpha_{1j})$ for $j = 1, \ldots, J$, where $\boldsymbol{\omega}_{P,j} = (\omega_{P0,\,j}, \omega_{P1,j})^T$ can be regarded as perturbation vectors and $\omega_{P,1} = \cdots = \omega_{P,n} = 1$ represents no perturbation. Let $\log(\lambda_j) = \psi_j$, $j = 1, \ldots, J$, and let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_J)^T$. We consider a perturbation to $\boldsymbol{\psi}$ using a discretized integrated Ornstein-Uhlenbeck process (Sinha and Dey 1997), given by

$$\psi_{j+1} - \psi_{0,j+1} = \omega_{P,j}[\psi_j - \psi_{0,j}] + \varepsilon_{j+1}, \quad (13)$$

where $\psi_{0,j}$ is the prior mean of ψ_j, the ε_j are *i.i.d*, $\varepsilon_j \sim N\left(0, \sigma_\psi^2\right)$, and $\omega_{P,j}$ can be regarded as a perturbation parameter, $j = 1, \ldots, J - 1$. Thus, $\omega_{P,1} = \cdots = \omega_{P,\,J-1} = 0$ represents no perturbation. A simple perturbation to the gamma process prior for $H_0(\cdot)$ is to use a positive scale function $k_*(t, \boldsymbol{\omega}_P)$ to perturb $H^*(\cdot)$ such that $H_k^*(y, \omega_P) = \int_0^y k_*(t, \omega_P) h_0(t) dt$, where $k_*(t, \mathbf{0}) = 1$ represents no perturbation.

2.3 Bayesian perturbation manifold

We propose a Bayesian perturbation manifold to quantify each perturbation $\boldsymbol{\omega}$ in the perturbation model to the Bayesian survival model (1). Since Ω is a subset of R^m, the perturbed model $\mathcal{M} = \{ p(D_{com}, \boldsymbol{\theta}| \boldsymbol{\omega}): \boldsymbol{\omega} \in R^m\}$ can be regarded as an m-dimensional manifold under some conditions (Amari 1990). The geometric structure of $\mathcal{M}$ is mainly characterized by an $m \times m$ generalized Fisher information matrix within the Bayesian framework, denoted by $G(\boldsymbol{\omega}) = (g_{jk}(\boldsymbol{\omega}))$. Let ω_k be the kth component of $\boldsymbol{\omega}$, $\ell_c(\boldsymbol{\omega}) = \log p(D_{com}, \boldsymbol{\theta}|\boldsymbol{\omega})$, and $\partial_{\omega_k} = \partial/\partial_{\omega_k}$. The m^2 quantities

$$g_{jk}(\omega)=\mathrm{E}_{\omega}\left\{[\partial_{\omega_j}\ell_c(\omega)-\mathrm{E}_{\omega}\partial_{\omega_j}\ell_c(\omega)][\partial_{\omega_k}\ell_c(\omega)-\mathrm{E}_{\omega}\partial_{\omega_k}\ell_c(\omega)]\right\} \quad (14)$$

for $j, k = 1, \ldots, m$, in which $\mathrm{E}_{\boldsymbol{\omega}}$ denotes the expectation taken with respect to $p(D_{com}, \boldsymbol{\theta}| \boldsymbol{\omega})$, form the *metric* tensor of $\mathcal{M}$, denoted by $G(\boldsymbol{\omega})$. Geometrically, the m functions $\partial_{\omega_j} \ell_c(\boldsymbol{\omega})$ can be regarded as the basis functions in the tangent space $T\boldsymbol{\omega}$ of $\mathcal{M}$ at each $\boldsymbol{\omega} \in \mathcal{M}$. If there are no random effects or missing data, $p(D_{com}, \boldsymbol{\theta}| \boldsymbol{\omega})$ and $\ell_c(\boldsymbol{\omega})$ reduce to $p(D_{obs}, \boldsymbol{\theta}| \boldsymbol{\omega})$ and $\ell_o(\boldsymbol{\omega}) = \log p(D_{obs}, \boldsymbol{\theta}| \boldsymbol{\omega})$, respectively. $\mathcal{M} = \{ p(D_{obs}, \boldsymbol{\theta}| \boldsymbol{\omega}): \boldsymbol{\omega} \in R^m\}$ can be regarded as a generalization of the perturbation manifold in Zhu et al. (2007, 2009).

An *appropriate perturbation* to the survival model (1) requires that $G(\boldsymbol{\omega}^0)$ be a diagonal matrix. Since $G(\boldsymbol{\omega})$ is essentially a Fisher information matrix, the (i, i)th element $g_{ii}(\boldsymbol{\omega})$ can be interpreted as the amount of perturbation introduced by ω_i, whereas $g_{ij}(\boldsymbol{\omega})$ represents the association between ω_i and ω_j. A diagonal $G(\boldsymbol{\omega})$ indicates that all components of $\boldsymbol{\omega}$ may be regarded as being orthogonal to each other in the perturbed model (Cox and Reid 1987). A large value of $|g_{ij}(\omega)/\sqrt{g_{ii}(\omega)g_{jj}(\omega)}|$ indicates that ω_i and ω_j play similar roles in the perturbation model $\mathcal{M}$. An extreme scenario is that ω_i and ω_j are linearly dependent, that is $|g_{ij}(\omega)|=\sqrt{g_{ii}(\omega)g_{jj}(\omega)}$, and thus one of them can be dropped. For interpretation purposes, it is important to introduce an appropriate perturbation to the survival model (1) in order to make it easier to identify the source of a large perturbation.

Although $G(\boldsymbol{\omega}^0)$ may not be diagonal for an arbitrary perturbation $\boldsymbol{\omega}$, we can always choose a new perturbation vector $\tilde{\boldsymbol{\omega}}$, defined by

$$\tilde{\omega}(\omega)=\omega^0+G\left(\omega^0\right)^{1/2}\left(\omega-\omega^0\right), \quad (15)$$

such that $G(\tilde{\boldsymbol{\omega}})$ evaluated at $\boldsymbol{\omega}^0$ equals $\mathbf{I}_m$ (Zhu et al. 2007, 2009). Given the geometric structure of $\mathcal{M}$, we can further carry out influence analysis of $\boldsymbol{\omega}$.

We can compute several additional geometric quantities of the manifold $\mathcal{M}$, which are useful for characterizing the geometric structure of $\mathcal{M}$ (Amari 1990). To connect two tangent spaces at two neighboring points $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$, we introduce the Levi-Civita connection of $G(\boldsymbol{\omega})$, given as follows:

$$\Gamma_{jkl}(\omega)=0.5\left[\partial_{\omega_l}g_{jk}(\omega)+\partial_{\omega_j}g_{lk}(\omega)-\partial_{\omega_k}g_{lj}(\omega)\right]. \quad (16)$$

Furthermore, we can determine a *geodesic* $\boldsymbol{\omega}(t) = (\omega_1(t), \ldots, \omega_m(t))$ with respect to the affine connection $\Gamma_{jkl}(\omega)$ on $\mathcal{M}$. Specifically, the geodesic $\boldsymbol{\omega}(t)$ satisfies the equation

$$\frac{d^2\omega_l(t)}{dt^2}+\sum_{s,j,k}g^{ls}(\omega(t))\Gamma_{jks}(\omega(t))\frac{d\omega_j(t)}{dt}\frac{d\omega_k(t)}{dt}=0,$$

where $g^{ls}(\boldsymbol{\omega})$ is the (l, s)th element of $G(\boldsymbol{\omega})^{-1}$. As we move along a geodesic, the tangent vector of the geodesic does not change in length and direction. The geodesic is an extension of the straight line $\boldsymbol{\omega}(t) = \boldsymbol{\omega}^0 + t\mathbf{h}$ in Euclidean space (Amari 1990).

Finally, we can obtain the *Bayesian perturbation manifold* ($\mathcal{M}$, $G(\boldsymbol{\omega})$) with metric tensor $G(\boldsymbol{\omega})$. As an illustration, we examine the geometric structure of the proportional hazards and shared-frailty models in Bayesian analysis.

Example 1 (continued)—The perturbation model $\mathcal{M} = \{ p(D_{obs}, \boldsymbol{\beta} | \boldsymbol{\omega}_P, \boldsymbol{\omega}_D)\ (\boldsymbol{\omega}_P, \boldsymbol{\omega}_D) \in \Omega\}$ can be regarded as a Riemannian manifold. The tangent space $T\boldsymbol{\omega}$ of $\mathcal{M}$ is spanned by

$$\begin{aligned}\partial_{\omega_P}\ell_o(\omega)&=\sum_{j=1}^{n}c_0\log\left(\frac{c_0}{c_0+A_j(\omega_D;\beta)}\right)\int_{y_{(j-1)1}}^{y_{(j)1}}\partial_{\omega_P}k_*(t,\omega_P)h_0(t)dt+\sum_{j=1}^{n}\nu_{(j)1}k_*(y_{(j)1},\omega_P)^{-1}\partial_{\omega_P}k_*(y_{(j)1},\omega_P),\\ \partial_{\omega_D}\ell_o(\omega)&=\sum_{j=1}^{n}\left[\nu_{(j)1}-\nu_{(j-1)1}-c_0\int_{y_{(j-1)1}}^{y_{(j)1}}k_*(t,\omega_P)h_0(t)dt\right]\frac{\partial_{\omega_D}A_j(\omega_D;\beta)}{c_0+A_j(\omega_D;\beta)},\end{aligned} \tag{17}$$

where $\nu_{(0)1} = 0$ and $y_{(0)1} = 0$. The submatrices of $G(\boldsymbol{\omega}^0)$ are given by

$$\begin{aligned}E_{\omega^0}\left[-\partial^2_{\omega_P}\ell_o(\omega^0)\right]&=-\sum_{j=1}^{n}c_0\log\left(\frac{c_0}{c_0+A_j(\omega_D^0;\beta)}\right)\\ E&\left[\int_{y_{(j-1)1}}^{y_{(j)1}}\partial^2_{\omega_P}k_*(t,\omega_P^0)h_0(t)dt\right]-\sum_{j=1}^{n}E\left\{\nu_{(j)1}\left[\partial_{\omega_P}2\log k_*(y_{(j)},\omega_P^0)\right]\right\},\\ E_{\omega^0}\left[-\partial^2_{\omega_D}\ell_o(\omega^0)\right]&=\sum_{j=1}^{n}\left[\nu_{(j)1}-\nu_{(j-1)1}-c_0\int_{y_{(j-1)1}}^{y_{(j)1}}k_*\left(t,\omega_P^0\right)h_0(t)dt\right]\times\partial_{\omega_D}\left[\frac{\partial_{\omega_D}A_j(\omega_D^0;\beta)}{c_0+A_j\left(\omega_D^0;\beta\right)}\right],\\ E_{\omega^0}\left[-\partial^2_{\omega_D\omega_P}\ell_o(\omega^0)\right]&=\sum_{j=1}^{n}c_0\int_{y_{(j-1)1}}^{y_{(j)1}}\partial_{\omega_P}k_*\left(t,\omega_P^0\right)h_0(t)dt\frac{\partial_{\omega_D}A_j\left(\omega_D^0;\beta\right)^T}{c_0+A_j\left(\omega_D^0;\beta\right)}.\end{aligned} \tag{18}$$

Clearly, $G(\boldsymbol{\omega}^0)$ is not a diagonal matrix, and thus, $\boldsymbol{\omega}$ is not an appropriate perturbation. However, we can always choose a new perturbation vector $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}^0)$, which gives an appropriate perturbation.

Example 2 (continued)—We consider the frailty model using the piecewise constant hazards model and assume the data perturbation (11) to $h(y|\mathbf{x}_{ij}, b_i)$. Thus, $p(D_{com}, \boldsymbol{\theta} | \boldsymbol{\omega}_D)$ is given by

$$\begin{aligned}&p(D_{com},\theta|\omega_D)\\ &\propto\varphi(\beta;\mu_0,\textstyle\sum_0)p(\kappa)\prod_{i=1}^{n}\left(b_i^{\kappa^{-1}-1}\exp(-b_i/\kappa)\prod_{j=1}^{n_i}\prod_{l=1}^{J}\left[b_i\exp\left(\mu_2\left(\mathbf{x}_{ij}^T\beta,\omega_{D,i}\right)\right)\lambda_l\right]^{\delta_{ijl}\nu_{ij}}\right.\\ &\left.\times\exp\left\{-b_i\exp\left(\mu_2\left(\mathbf{x}_{ij}^T\beta,\omega_{D,i}\right)\right)\delta_{ijl}\left[\lambda_l\left(y_{ij}-s_{l-1}\right)+\sum_{g=1}^{l-1}\lambda_g(s_g-s_{g-1})\right]\right\}\right).\end{aligned} \tag{19}$$

The tangent space $T\boldsymbol{\omega}_D$ of $\mathcal{M} = \{p(D_{com}, \boldsymbol{\theta} | \boldsymbol{\omega}_D): \boldsymbol{\omega}_D \in \Omega\}$ is spanned by

$$\partial_{\omega_{D,i}}\ell_c(\omega)=\sum_{j=1}^{n_i}\sum_{l=1}^{J}\delta_{ijl}\left[\partial_{\omega_{D,i}}\mu_2\left(\mathbf{x}_{ij}^T\beta,\omega_{D,i}\right)\right]\times\left\{\nu_{ij}-b_i\exp\left(\mu_2\left(\mathbf{x}_{ij}^T\beta,\omega_{D,i}\right)\right)\Delta_l(y_{ij})\right\}$$

for $i = 1, \ldots, n$, where $\Delta_l(y_{ij})=\lambda_l(y_{ij}-s_{l-1})+\sum_{g=1}^{l-1}\lambda_g(s_g-s_{g-1})$. It can be shown that $g_{D,ii}(\omega_D^0)=E_{\omega_D^0}\left[-\partial^2_{\omega_{Di}}\ell_c\left(\omega_D^0\right)\right]$ is given by

$$\sum_{j=1}^{n_i}\sum_{l=1}^{J} E_{\omega_D^0}\left\{\delta_{ijl} b_i \exp\left(\mathbf{x}_{ij}^T\beta\right)\left[\partial_{\omega_{D,i}}\mu_2\left(\mathbf{x}_{ij}^T\beta, \omega_{D,i}^0\right)\right]^{\otimes 2}\Delta_l(y_{ij})\right\}$$
$$-\sum_{j=1}^{n_i}\sum_{l=1}^{J}\times E_{\omega_D^0}\left\{\delta_{ijl}\left[\partial^2_{\omega_{D,i}}\mu_2\left(\mathbf{x}_{ij}^T\beta, \omega_{D,i}\right)\right]\left[\nu_{ij} - b_i \exp\left(\mu_2\left(\mathbf{x}_{ij}^T\beta, \omega_{D,i}\right)\right)\Delta_l(y_{ij})\right]\right\} \quad (20)$$

for $i = 1, \ldots, n$ and $G(\omega^0)=\mathrm{diag}(g_{D,11}(\omega_D^0), \ldots, g_{D,nn}(\omega_D^0))$. Here, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for a vector $\mathbf{a}$. When $\boldsymbol{\omega}_{D,i}$ is a scalar for each i, $G(\boldsymbol{\omega}^0)$ is a diagonal matrix and thus, $\boldsymbol{\omega}$ is an appropriate perturbation. In general, for multivariate $\boldsymbol{\omega}_{D,i}$, we can choose a new perturbation vector $\tilde{\omega}_{D,i}=\omega_{D,i}^0+g_{D,ii}\left(\omega_D^0\right)^{1/2}\left(\omega_{D,i}-\omega_{D,i}^0\right)$ for each i to obtain an appropriate perturbation $\tilde{\boldsymbol{\omega}}_D = (\boldsymbol{\omega}_{D,1}, \ldots, \boldsymbol{\omega}_{D,n})^T$.

2.4 Local influence measures

We consider an $l \times 1$ objective function $f(\boldsymbol{\omega})$: $\mathcal{M} \to R^l$, such as the posterior mean distance, φ-divergence ($l = 1$, see Example 6), or Bayes factor ($l = 1$), which defines the aspect of inference of interest for sensitivity analysis. We first consider the finite-dimensional manifold $\mathcal{M}$. Let $\boldsymbol{\omega}(t)$ be a geodesic on $\mathcal{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}^0$. It follows from a Taylor's series expansion that $f(\boldsymbol{\omega}(t)) = f(\boldsymbol{\omega}(0)) + \dot{f}_{\mathbf{h}}(0)t + O(t^2)$, where $\partial_t \boldsymbol{\omega}(t)|_{t=0} = \mathbf{h} \in R^m$ and $\dot{f}_{\mathbf{h}}(0)=\sum_j \partial_{\omega_j} f(\omega^0)h_j=\nabla_f^T\mathbf{h}$, in which $\nabla_f = \partial_{\boldsymbol{\omega}} f(\boldsymbol{\omega}^0)$.

First, we consider the case with $\nabla_f \neq 0$. We define a *first-order influence measure* (FI) in the direction $\mathbf{h} \in R^m$ as follows:

$$\mathrm{FI}_{f,\mathbf{h}}=\mathrm{FI}_{f(\omega^0),\mathbf{h}}=\frac{\mathbf{h}^T\nabla_f \boldsymbol{W}_f \nabla_f^T\mathbf{h}}{\mathbf{h}^T\boldsymbol{G}\mathbf{h}}, \quad (21)$$

where $\boldsymbol{G} = \boldsymbol{G}(\tilde{\boldsymbol{\omega}}^0)$ and $\boldsymbol{W}_f$ is a positive semi-definite matrix. Particularly, for the appropriate perturbation $\tilde{\boldsymbol{\omega}}(\boldsymbol{\omega})$ in (15), FI $_{f,\mathbf{h}}$ reduces to

$$\mathrm{FI}_{f(\tilde{\omega}),\mathbf{h}}\Big|_{\tilde{\omega}=\omega^0}=\frac{\mathbf{h}^T\boldsymbol{G}^{-1/2}\nabla_f \boldsymbol{W}_f \nabla_f^T\boldsymbol{G}^{-1/2}\mathbf{h}}{\mathbf{h}^T\mathbf{h}}. \quad (22)$$

The maximum value of FI $_{f,\mathbf{h}}$ equals the principal eigenvalue of $\boldsymbol{G}^{-1/2}\nabla_f \boldsymbol{W}_f \nabla_f^T\boldsymbol{G}^{-1/2}$, which quantifies the largest degree of local influence of $\tilde{\boldsymbol{\omega}}$ to a statistical model, while the corresponding eigenvector of $\boldsymbol{G}^{-1/2}\nabla_f \boldsymbol{W}_f \nabla_f^T\boldsymbol{G}^{-1/2}$, denoted by $\mathbf{h}_{\max}$, can be used either for identifying robustness of priors, influential observations, or an inadequate sampling distribution. The quantity FI $_{f,\mathbf{h}}$ has a strong connection with McCulloch's (1989) Bayesian local influence measure for assessing the prior. The quantity $\mathbf{h}_{\max}$ indicates the worst perturbation direction for $f(\tilde{\boldsymbol{\omega}})$, that is, the direction that gives maximum change to the objective function. We also suggest inspection of $\mathrm{FI}_{f,\mathbf{e}_i}$, where $\mathbf{e}_i$ is an $m \times 1$ vector with ith component 1 and 0 otherwise (Zhu and Lee 2001; Zhu et al. 2007). The use of the FI $_{f,\mathbf{e}_i}$'s can identify the most significant components of $\tilde{\boldsymbol{\omega}}$.

Example 5 (Influence measures)—We consider the logarithm of the Bayes factor for comparing $\boldsymbol{\omega}$ with $\boldsymbol{\omega}^0$ as follows:

$$B(\omega) = \log p(D_{obs}|\omega) - \log p(D_{obs}|\omega^0), \tag{23}$$

where $p(D_{obs}|\boldsymbol{\omega}) = \int p(D_{obs}, \boldsymbol{\theta}|\boldsymbol{\omega})d\boldsymbol{\theta} = \int p(D_{com}, \boldsymbol{\theta}|\boldsymbol{\omega})d\, D_{mis}d\boldsymbol{\theta}$. Under some smoothness conditions, $B(\boldsymbol{\omega})$ is a continuous map from $\mathcal{M}$ to R. If we set $f(\boldsymbol{\omega}) = B(\boldsymbol{\omega})$, it can be shown that

$$\nabla_B = E_{\omega^0}\left[\partial_\omega \log p\left(D_{com}, \theta; \omega^0\right)|D_{obs}\right] \quad \text{and} \quad \mathbf{h}_{\max} = \frac{\boldsymbol{G}\left(\omega^0\right)^{-1/2}\nabla_B}{\sqrt{\nabla_B^T \boldsymbol{G}(\omega^0)^{-1}\nabla_B}}. \tag{24}$$

To calculate the local influence measures associated with $B(\boldsymbol{\omega})$, we just need to compute ∇_B and $\boldsymbol{G}(\boldsymbol{\omega}^0)$ and choose an appropriate $\boldsymbol{W}_f$, such as $\boldsymbol{W}_f = I$. For instance, for Examples 1 and 2, we can easily calculate them given the formulas given in Sect 2.3 along with MCMC methods for computing ∇_B.

We consider the posterior mean of a function of $\boldsymbol{\theta}$, denoted $d(\boldsymbol{\theta}) \in R^l$, after introducing $\boldsymbol{\omega}$ as follows:

$$M_d(\omega) = \int d(\theta) p(D_{mis}, \theta|D_{obs}, \omega) dD_{mis} d\theta. \tag{25}$$

We can set $f(\boldsymbol{\omega}) = M_d(\boldsymbol{\omega})$ and $\boldsymbol{W}_f \equiv W_{M_d} = [\mathrm{Cov}(d(\boldsymbol{\theta})|D_{obs})]^{-1}$. It can be shown that

$$\nabla_{M_d} = \mathrm{Cov}_{\omega^0}\left\{\partial_\omega \log p(D_{com}, \theta; \omega^0), d(\theta)|D_{obs}\right\} \tag{26}$$

and these local influence measures are associated with Gustafson's (1996a; 1996b) local sensitivity quantities. To calculate the local influence measures associated with $M_d(\boldsymbol{\omega})$, we just need to compute ∇_{M_d} and $\boldsymbol{G}(\boldsymbol{\omega}^0)$. For instance, for Examples 1 and 2, we can easily approximate them using the formulas given in Sect. 2.3 along with MCMC methods for computing ∇_{M_d}.

We can also carry out Bayesian local influence when $\nabla_f = 0$. For notational simplicity, we assume that the dimension of $f(\boldsymbol{\omega})$ equals 1. It follows from a Taylor's series expansion that $f(\boldsymbol{\omega}(t)) = f(\boldsymbol{\omega}(0)) + 0.5\, \ddot{f}_{\mathbf{h}}(0)t^2 + O(t^3)$, where $\ddot{f}_{\mathbf{h}}(0) = \mathbf{h}^T \boldsymbol{H}_f\, \mathbf{h}$ and $\boldsymbol{H}_f = \partial_\omega^2 f(\omega^0)$. We define a *second-order influence measure* (SI) in the direction $\mathbf{h} \in R^m$ as follows:

$$\mathrm{SI}_{f,\mathbf{h}} = \mathrm{SI}_{f(\omega^0),\mathbf{h}} = \frac{\mathbf{h}^T \boldsymbol{H}_f \mathbf{h}}{\mathbf{h}^T \boldsymbol{G}\mathbf{h}}. \tag{27}$$

Particularly, for the appropriate perturbation $\tilde{\boldsymbol{\omega}}(\boldsymbol{\omega})$ in (15), $\mathrm{SI}_{f,\mathbf{h}}$ reduces to

$$\mathrm{SI}_{f(\tilde{\omega}),\mathbf{h}}\Big|_{\tilde{\omega}=\omega^0} = \frac{\mathbf{h}^T \boldsymbol{G}^{-1/2}\boldsymbol{H}_f \boldsymbol{G}^{-1/2}\mathbf{h}}{\mathbf{h}^T\mathbf{h}}. \tag{28}$$

Similar to the first-order influence measure, we only consider the eigenvalue-eigenvector pairs of $\boldsymbol{G}^{-1/2}\, \boldsymbol{H}_f\, \boldsymbol{G}^{-1/2}$, which can be used either for identifying non-robust priors, influential observations, or inappropriate and non-robust sampling distributions. We also examine $\mathrm{SI}_{f,\mathbf{e}_i}$ (Zhu and Lee 2001; Zhu et al. 2007, 2009).

Example 6 (φ-divergence)—We consider the φ-divergence between two posterior distributions before and after introducing the perturbation $\boldsymbol{\omega}$ as follows:

$$\mathrm{D}_{\varphi}(\omega)=\int \varphi(R(D_{mis}, \theta;\omega))p(D_{mis}, \theta|D_{obs})dD_{mis}d\theta, \quad (29)$$

where $R(D_{mis}, \boldsymbol{\theta}; \boldsymbol{\omega}) = p(D_{mis}, \boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega})/p(D_{mis}, \boldsymbol{\theta}|D_{obs})$ and $\varphi(\cdot)$ is a convex function with $\varphi(1) = 0$, such as the Kullback-Leibler divergence or the χ^2-divergence (Kass et al. 1989; Weiss 1996). We set $f(\boldsymbol{\omega}) = D_{\varphi}(\boldsymbol{\omega})$. It can be shown that $\dot{f}_{\mathbf{0}} = 0$ and

$$\boldsymbol{H}_f=\ddot{\varphi}(1)\int \{\partial_{\omega}\log p(D_{mis}, \theta|D_{obs}, \omega_0)\}^2 p(D_{mis}, \theta|D_{obs})dD_{mis}d\theta. \quad (30)$$

In practice, we use MCMC methods to draw samples $\left\{(\theta^{(s)}, D_{mis}^{(s)}:s=1, \ldots, S_0)\right\}$ from $p(\boldsymbol{\theta}, D_{mis}|D_{obs})$ to approximate $\boldsymbol{H}_f$.

3 Examples

3.1 Simulation studies

Survival times y_{i1} ($i = 1, \ldots, 100$) are generated from an exponential distribution with hazard $h_i=\exp\left(\mathbf{x}_{i1}^T\beta\right)$, where $\mathbf{x}_{i1} = (x_{i11}, x_{i12}, x_{i13})^T$ in which x_{i11}, x_{i12} and x_{i13} are generated from the normal distributions $N(1.8, 0.20)$, $N(1.7, 0.30)$ and $N(2.0, 0.25)$, respectively, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = (0.8, 0.8, 0.8)^T$. To introduce some outliers, we generated new survival times $\{y_{i1}: i = 99, 100\}$ from the exponential distribution with hazard $h_i=\exp\left(\mathbf{x}_{i1}^T\beta+5x_{i13}^2\right)\varepsilon_i^{\exp(5x_{i13}^2)-1}$, where ε_i is generated from a uniform distribution $U(0, 1)$. The survival times $\{y_{i1}: i = 1, \ldots, 98\}$ are randomly right censored with probability 0.10. The censoring proportion of the survival times is about 7.2%.

We fit the piecewise constant hazards model to the simulated data in which we chose subintervals $(s_{j-1}, s_j]$ with equal lengths for $J = 200$ and used MCMC (Metroplis-Hastings) sampling to carry out Bayesian influence analysis (Chen et al. 2000). We specify the following prior distributions for $\boldsymbol{\beta}$ and λ_j ($j = 1, \ldots, J$):

$$p(\beta)=N(\mu_0, \textstyle\sum_0), \qquad p(\lambda_j)=\mathscr{G}(\alpha_{0j}, \alpha_{1j}), \quad (31)$$

where $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, α_{0j}, α_{1j} are specified hyperparameters. We set $\boldsymbol{\mu}_0 = (0.8, 0.8, 0.8)^T$, $\alpha_{0j} = 8.0$, $\alpha_{1j} = 10.0$, and $\boldsymbol{\Sigma}_0 = \mathrm{diag}(0.25, 0.25, 0.25)$. In the MCMC sampler, we choose a normal proposal density which yields an average acceptance rate of 32.5%.

We simultaneously perturbed the piecewise constant hazards model and the prior distributions of $\boldsymbol{\beta}$, whose perturbed log-posterior is given by

$$\log p(\theta|D_{obs}, \omega)=\sum_{i=1}^{n}\sum_{j=1}^{J}\left\{\delta_{ij}v_i\left(\log\lambda_j+\mathbf{x}_{i1}^T\beta+\omega_i\right)-\delta_{ij}\left[\lambda_j(y_i-s_{j-1})+\sum_{g=1}^{j-1}\lambda_g(s_g-s_{g-1})\right]\times\exp\left(\mathbf{x}_{i1}^T\beta+\omega_i\right)\right\}$$
$$-\left[\log|\textstyle\sum_0/\omega_{\beta 2}|+\omega_{\beta 2}(\beta-\omega_{\beta 1}\mu_0)^T\sum_0^{-1}(\beta-\omega_{\beta 1}\mu_0)\right]/2,$$

where $\delta_{ij} = 1$ if subject i either failed or was right censored in the jth interval and 0 otherwise, $v_i = 1$ if subject i failed and $v_i = 0$ if subject i was right censored, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n,$

$\omega_{\beta1}, \omega_{\beta2})^T$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$. In this case, $\boldsymbol{\omega}^0 = (0, \ldots, 0, 1, 1)^T$ represents no perturbation. After some calculations, we have

$$G\left(\omega^0\right)=\text{diag}\left(g_{11},\ldots,g_{nn},\mu_0^T\sum\nolimits_0^{-1}\mu_0, 3/2\right),$$

where $g_{ii}=E\left\{\sum\nolimits_{j=1}^{J}\delta_{ij}[\lambda_j(y_i - s_{j-1})+\sum\nolimits_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})]\exp\left(\mathbf{x}_{i1}^T\beta\right)\right\}\ (i=1,\ldots,n)$, and the expectation is with respect to the joint distribution of $(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\lambda})$. Then, we chose a new perturbation scheme $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}^0)$ and calculated the associated local influence measures $\mathbf{h}_{\max}$ corresponding to (24) which we denote by $\mathbf{h}_{\max}^B$, $\text{SI}_{\text{D}_{\varphi}\mathbf{e}_j}$ and $\text{SI}_{M_d\mathbf{e}_j}$, in which $\varphi(\cdot)$ was chosen to be the Kullback-Leibler divergence and $d(\boldsymbol{\theta}) = \boldsymbol{\theta}$. Cases 99 and 100 were detected to be influential by our local influence measures (Fig. 1a, c, e).

We used the same setup except that we employed a perturbed prior distribution for $\boldsymbol{\beta}$, namely $p(\boldsymbol{\beta}) = N(\boldsymbol{\mu}_0, 0.4\boldsymbol{\Sigma}_0)$, and then we applied the same MCMC method, perturbation scheme, and local influence measures. Cases 99 and 100 and the perturbed prior distribution of $\boldsymbol{\beta}$ were identified to have a big effect (Fig. 1b, d, f).

To detect the misspecified survival model, we generated new survival times $\{y_{i1}: i = 1, \ldots, 100\}$ from an exponential distribution with hazard $h_i=\exp\left(\mathbf{x}_{i1}^T\beta+5x_{i13}^2\right)$, where $\mathbf{x}_{i1} = (x_{i11}, x_{i12}, x_{i13})^T$ is generated from the multivariate normal distribution $N_3((1.8, 1.7, 2.0)^T, \text{diag}(0.2, 0.3, 0.25))$. The survival times $\{y_{i1}: i = 1, \ldots, 100\}$ are randomly right censored with probability 0.10. The actual censoring proportion of the survival times is about 13.0%. Similarly, the piecewise constant hazards model with $h_i=\exp\left(\mathbf{x}_{i1}^T\beta\right)$ is used to fit the simulated data using the priors for $\boldsymbol{\beta}$ and λ_j given in equation (31), where we chose subintervals $(s_{j-1}, s_j]$ with equal lengths using $J = 120$. In this simulation study, we consider the following perturbation to the sampling distribution $p(D_{obs}|\boldsymbol{\theta})$, given by

$$p(D_{obs}|\theta,\omega)=\prod_{i=1}^{n}\prod_{j=1}^{J}\{\lambda_j h_i^{\omega}\}^{\delta_{ij}\nu_i}\exp\left\{-h_i^{\omega}\delta_{ij}\left[\lambda_j(y_i - s_{j-1})+\sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\right]\right\},$$

where $h_i^{\omega}=\exp\left(\mathbf{x}_{i1}^T\beta+\omega\right)$. In this case, $\omega = 0$ represents no perturbation. The local influence measures including the Bayes factor, KL-divergence and posterior mean for the above generated dataset were calculated under a $N(0.8\mathbf{1}_3, 0.5\boldsymbol{I}_3)$ prior for $\boldsymbol{\beta}$ and are, respectively, denoted as t_B, t_{φ} and t_M. Let $\tilde{\boldsymbol{\beta}}$ denote the posterior mean of $\boldsymbol{\beta}$ for the above generated dataset. To formally assess the sizes of t_B, t_{φ} and t_M, we computed a "p-value" as follows. We used the parametric bootstrap and simulated 100 data sets according to the fitted model with $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$. Then, for each simulated data set, three corresponding local influence measures, denoted as $t_{bB}^k, t_{b\varphi}^k$ and $t_{bM}^k\ (k=1,\ldots,100)$, were calculated. The p-values, which were calculated using the formula $p-\text{value}=\frac{1}{100}\sum\nolimits_{k=1}^{100}I(t_{bl}^k \geq t_l)\ (l=B,\varphi,M)$, are given by 0.0, 0.04 and 0.0, respectively. These results show that the survival model is misspecified at "significance level" $\alpha = 0.05$. In the MCMC scheme, we used a normal proposal density which yielded an average acceptance rate of 26% in calculating t_l, and used a normal proposal to yield an average acceptance rate of 32% for calculating $t_{bl}^k\ (l=B,\varphi,M)$.

In the second simulation study, we considered the shared-frailty model. For the shared-frailty model (5), we set $n = 100$, and we chose varying values of n_i in order to create a scenario with different cluster sizes. In particular, we set $n_1 = \dots = n_{10} = 3$, $n_{91} = \dots = n_{100} = 20$, and $n_j \in \{5, 7, 8, 10, 12, 14, 16, 18\}$ for $i = 11, \dots, 90$. For each cluster i ($i = 1, \dots, n$), the survival times y_{ij} ($j = 1, \dots, n_i$) are generated from an exponential distribution with hazard $h_{ij}=b_i\exp\left(\mathbf{x}_{ij}^T\beta\right)$, where $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})^T$ in which $x_{ij1} = 1.0$, x_{ij2} and x_{ij3} are, respectively, generated from the $\mathcal{G}(5.0, 8.0)$ distribution and the $N(2.0, 0.25)$ distribution, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = (0.8, 0.8, 0.8)^T$. To introduce some outliers, we generated new survival times $\{y_{ij}: i = 99, 100, j = 1, \dots, 20\}$ from the exponential distribution with hazard $h_{ij}=b_i\exp\left(\mathbf{x}_{ij}^T\beta+5x_{ij3}^2\right)$. The survival times $\{y_{ij}: i = 1, 2, \dots, 98, j = 1, \dots, n_i\}$ are randomly right censored with probability 0.10. The censoring proportion of the survival times is about 15.8% in the dataset.

We fit the piecewise constant hazards shared-frailty model in which we chose subintervals $(s_{j-1}, s_j]$ with equal lengths for $J = 120$ and used MCMC sampling to carry out Bayesian influence analysis. Here, we adopted the same prior distributions for $\boldsymbol{\beta}$ and λ_l ($l = 1, \dots, J$) as those given in Eq. 31. Also, the same values of the hyperparameters for $\boldsymbol{\mu}_0$, α_{0j}, α_{1j} and $\boldsymbol{\Sigma}_0$ (see the first simulation study) were used in the analysis, and we chose a normal proposal density in the MCMC scheme to give us an average acceptance rate of 33%.

We simultaneously perturbed the piecewise constant hazards shared-frailty model and the prior distribution of $\boldsymbol{\beta}$. The perturbed complete-data log-posterior of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$ is given by

$$\begin{aligned}\log p(\theta|D_{com},\omega)=&\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=1}^{J}\Big\{\delta_{ijl}\nu_{ij}\Big[\log(b_i\lambda_l)+\mathbf{x}_{ij}^T\beta+w_i\Big]\\&-\delta_{ijl}\,b_i\exp\left(\mathbf{x}_{ij}^T\beta+\omega_i\right)\Big[\lambda_l\left(y_{ij}-s_{l-1}\right)+\sum_{g=1}^{l-1}\lambda_g\left(s_g-s_{g-1}\right)\Big]\Big\}\\&-\Big[\log|\Sigma_0/\omega_{\beta2}|+\omega_{\beta2}\left(\beta-\omega_{\beta1}\mu_0\right)^T\Sigma_0^{-1}(\beta-\omega_{\beta1}\mu_0)\Big]/2,\end{aligned}$$

where $\delta_{ijl} = 1$ if the jth observation in the ith cluster either failed or was right censored in the lth interval and 0 otherwise, and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n, \omega_{\beta1}, \omega_{\beta2})^T$. In this case, $\boldsymbol{\omega}^0 = (0, \dots, 0, 1, 1)^T$ represents no perturbation. After some calculations, we have

$$G\left(\omega^0\right)=\text{diag}\left(g_{11},\dots,g_{nn},\mu_0^T\sum\nolimits_0^{-1}\mu_0,3/2\right),$$

where $g_{ii}=E\left\{\sum_{j=1}^{n_i}\sum_{l=1}^{J}\delta_{ijl}b_i\left[\lambda_l(y_{ij}-s_{l-1})+\sum_{g=1}^{l-1}\lambda_g(s_g-s_{g-1})\right]\exp\left(\mathbf{x}_{ij}^T\beta\right)\right\}(i=1,\dots,n)$, and the expectation is with respect to the joint distribution of $(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\lambda})$. Then, we chose a new perturbation scheme $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}^0)$ and calculated the associated local influence measures $\mathbf{h}_{\max}^B$, $\text{SID}_{D_\varphi \mathbf{e}_j}$ and $\text{SI}_{CM_h\mathbf{e}_j}$, in which $\varphi(\cdot)$ was chosen to be the Kullback divergence and $d(\boldsymbol{\theta}) = \boldsymbol{\theta}$. Cases 99 and 100 were detected to be influential by our local influence measures (Fig. 2a, c, e).

We used the same setup except that we employed a perturbed prior distribution for $\boldsymbol{\beta}$, given by $p(\boldsymbol{\beta}) = N(4.0\boldsymbol{\mu}_0, 0.6\boldsymbol{\Sigma}_0)$, and then we applied the same MCMC method, perturbation scheme, and local influence measures. Cases 99 and 100 and the perturbed prior distribution of $\boldsymbol{\beta}$ were identified to have a big effect (Fig. 2b, d, f).

Similarly, to detect a misspecified model, we generated new survival times $\{y_{ij}: i = 1, \ldots, 100, j = 1, \ldots, n_i\}$ from the exponential distribution with hazard $h_{ij}=b_i\exp\left(\mathbf{x}_{ij}^T\beta+5x_{ij3}^2\right)$. The survival times $\{y_{ij}: i = 1, \ldots, 100, j = 1, \ldots, 7\}$ are randomly right censored with probability 0.10. The actual censoring proportion of the survival times is about 10.3%. The piecewise constant hazards shared-frailty model with $h_{ij}=b_i\exp\left(\mathbf{x}_{ij}^T\beta\right)$ is used to fit the simulated data using the priors of $\boldsymbol{\beta}$ and λ_j given in Eq. 31, where the subintervals $(s_{j-1}, s_j]$ have equal lengths using $J = 120$. Here, we consider the following perturbation to $p(D_{obs}|\boldsymbol{\theta})$:

$$p(D_{obs}|\theta,\omega)=\prod_{i=1}^{100}\prod_{j=1}^{n_i}\prod_{l=1}^{J}\left\{\lambda_l h_{ij}^{\omega}\right\}^{\delta_{ijl}\nu_{ij}}\times\exp\left\{-h_{ij}^{\omega}\delta_{ijl}\left[\lambda_l(y_{ij}-s_{l-1})+\sum_{g=1}^{l-1}\lambda_g(s_g-s_{g-1})\right]\right\},$$

where $h_{ij}^{w}=b_i\exp\left(\mathbf{x}_{ij}^T\beta+\omega\right)$. In this case, $\omega = \omega^0 = 0$ represents no perturbation. The local influence measures including the φ-divergence and posterior mean for the above generated dataset were calculated using the prior $p(\boldsymbol{\beta})=N(0.8\mathbf{1}_3, 0.25\boldsymbol{I}_3)$ for $\boldsymbol{\beta}$. In the MCMC scheme, we take a normal proposal density such that the average acceptance rate is 34%. To compute a p-value, again we used the parametric bootstrap method and generated 100 bootstrap datasets from the fitted model with $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$. Two corresponding local influence measures, denoted as $t_{b\varphi}^k$ and t_{bM}^k $(k=1,\ldots,100)$, obtained from the 100 simulated data sets were calculated and used to estimate the p-values as 0.04 and 0.01, respectively. These results show that these influence measures are significant.

3.2 Multiple myeloma data

Multiple myeloma is a hematologic cancer characterized by an overproduction of antibodies. The Eastern Cooperative Oncology Group (ECOG) carried out a clinical trial (E2479) examining a chemotherapy to treat this disease and to also identify important prognostic factors that are predictive of survival. Our primary goal here is to illustrate the proposed methodology for carrying out Bayesian influence analysis. The response variable y is time to death, which is subject to right censoring. There are a total of $n = 339$ observations with eight observations being right censored. We examine eight covariates, which are blood urea nitrogen (x_1), hemoglobin (x_2), platelet count (x_3) (1 if normal, 0 if abnormal), age (x_4), white blood cell count (x_5), bone fractures (x_6), percentage of the plasma cells in bone marrow (x_7), and serum calcium (x_8). To ease the computational burden, we standardized all the covariates.

We fit the piecewise constant hazards model to the E2479 dataset in which $J = 28$ is used with the intervals chosen so that at least one failure or censored observation falls in each interval. The prior distributions for $\boldsymbol{\beta}$ and λ_j $(j = 1, \ldots, J)$ given in Eq. 31 are adopted, but $\boldsymbol{\mu}_0$ is taken to be that given in Ibrahim et al. (2001, Table 3.2 with $a_0 = 0$ on p. 63), $\alpha_{0j} = 8.0$, $\alpha_{1j} = 10.0$, and $\boldsymbol{\Sigma}_0 = \text{diag}(1.0, \ldots, 1.0)$. In the MCMC sampling, we chose the proposal density so that it gave an average acceptance rate of 34%.

We simultaneously perturbed the piecewise constant hazards model and the prior distributions of $\boldsymbol{\beta}$ and λ_j $(j = 1, \ldots, J)$, whose perturbed log-posterior is given by

$$\log p(\theta|D_{obs},\omega)=\sum_{i=1}^{n}\sum_{j=1}^{J}\left\{\delta_{ij}\nu_i\left(\log\lambda_j+\mathbf{x}_{i1}^T\beta+\omega_i\right)-\delta_{ij}[\lambda_j(y_i-s_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(s_g-s_{g-1})]\exp(\mathbf{x}_{i1}\beta+\omega_i)\right\}-[\log|\textstyle\sum_0/\omega_{\beta2}| + \omega_{\beta2}(\beta-\omega_{\beta1}\boldsymbol{u}_0)^T\textstyle\sum_0^{-1}(\beta-\omega_{\beta1}\boldsymbol{u}_0)]/2 + \sum_{j=1}^{J}\{(\alpha_{0j}-1)\log\lambda_j-\omega_{\alpha j}\lambda_j\alpha_{1j} + \alpha_{0j}\log(\omega_{\alpha j}\alpha_{1j})-\log\Gamma(\alpha_{0j})\}, \quad (32)$$

where $\boldsymbol{\omega}$ = (ω_1, …, ω_n, $\omega_{\beta1}$, $\omega_{\beta2}$, $\omega_{\alpha1}$, …, $\omega_{\alpha J}$)^T and $\boldsymbol{\theta}$ = ($\boldsymbol{\beta}$, $\boldsymbol{\lambda}$). In this case, $\boldsymbol{\omega}^0$ = (0, …, 0, 1, 1, 1, …, 1)^T represents no perturbation. After some calculations, we have

$$G(\omega^0)=\text{diag}\left(g_{11},\ldots,g_{nn},\mu_0^T\textstyle\sum_0^{-1}\mu_0,8/2,\alpha_{01},\ldots,\alpha_{0j}\right),$$

where $g_{ii}=E\left\{\sum_{j=1}^{J}\delta_{ij}\left[\lambda_j(y_i-s_{j-1})+\sum_{g=1}^{j-1}\lambda_g(s_g-s_{g-1})\right]\exp\left(\mathbf{x}_{i1}^T\beta\right)\right\}\ (i=1,\ldots,n)$ and the expectation is taken with respect with the joint distribution of ($\boldsymbol{y}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$). Then, we chose a new perturbation scheme $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega}-\boldsymbol{\omega}^0)$ and calculated the associated local influence measures $\mathbf{h}_{\max}^B$, $\mathrm{SID}_{\mathrm{D}_\varphi \mathbf{e}_j}$ and $\mathrm{SI}_{C\,M_h\mathbf{e}_j}$, in which $\varphi(\cdot)$ was chosen to be the Kullback divergence and $d(\boldsymbol{\theta}) = \boldsymbol{\theta}$. Examination of Fig. 3a indicates that cases 188, 261, 271, 294, 301, 307, 309, 321 and 339 were detected to be influential by our local influence measures; Fig. 3b shows that cases 110, 185, 261, 271, 304, 307, 321 and 339 were detected to be influential by our local influence measures; whilst cases 110, 185, 234, 261, 271, 304, 306, 307 and 339 were detected to be influential by Fig. 3c. Cases 261, 271, 307 and 339 were identified to be influential by all three figures. All these figures show that the priors for $\boldsymbol{\beta}$ and λ_j (j = 1, …, 28) do not impact the analysis very much.

To examine the robustness of the sampling model, we consider the following perturbation to $p(D_{obs}|\boldsymbol{\theta})$:

$$p(D_{obs}|\theta,\omega)=\prod_{i=1}^{n}\prod_{j=1}^{J}\left\{\lambda_j h_i^{\omega}\right\}^{\delta_{ij}\nu_i}\exp\left\{-h_i^{\omega}\delta_{ij}\left[\lambda_j(y_i-s_{j-1})+\sum_{g=1}^{j-1}\lambda_g(s_g-s_{g-1})\right]\right\},$$

where $h_i^{w}=\exp\left(\mathbf{x}_{i1}^T\beta+\omega\right)$. In this case, $\omega = \omega^0 = 0$ represents no perturbation. The local influence measures including the Bayes factor and φ-divergence were calculated with $p(\boldsymbol{\beta})=N(\boldsymbol{\mu}_0, \boldsymbol{I}_8)$, where $\boldsymbol{\mu}_0$ is taken to be that given in Ibrahim et al. (2001, Table 3.2 with a_0 = 0 on p. 63) and are, respectively, denoted by t_B and t_φ. Let $\tilde{\boldsymbol{\beta}}$ denote the posterior mean of $\boldsymbol{\beta}$ for the E2479 dataset. To calculate a p-value, we simulated 100 datasets from the fitted model with $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, and calculated the two corresponding local influence measures, denoted by t_{bB}^k and $t_{b\varphi}^k$ $(k=1,\ldots,100)$, which yielded the p-values 0.0 and 0.0, respectively. These results show that these influence measures are highly significant and therefore the sampling model may be misspecified. In the MCMC algorithm, we chose the normal proposal density to yield an average acceptance rate of 34%.

3.3 Melanoma data

Melanoma incidence is increasing at a rate that exceeds all solid tumors. Here, we examine a phase III clinical trial (E1690) carried out by the Eastern Cooperative Oncology Group (ECOG) involving post-operative chemotherapy for melanoma patients. The two main treatment arms on the E1690 trial are high-dose interferon (IFN) and observation (OBS), and one of the main aims in this study was to compare these two treatment arms with respect to relapse-free survival (RFS). There are a total of $n = 427$ observations with relapse-free survival (y) subject to right censoring. To illustrate our proposed methodology, we consider the three covariates treatment (x_1: IFN, OBS), age (x_2), and sex (x_3).

We fit a semiparametric cure rate model to the E1690 dataset using the likelihood in (7) along with a piecewise constant hazards model for $h_0(\cdot)$ in which $J = 10$ is used with s_j being the $((1 - e^{-j/J})/(1 - e^{-1}))$th quantile of the y_j's. The prior distributions for $\boldsymbol{\beta}$ and λ_j ($j = 1, \ldots, J$) given in Eq. 31 are adopted with $\boldsymbol{\mu}_0 = 0$, $\alpha_{0j} = 8.0$, $\alpha_{1j} = 10.0$, and $\boldsymbol{\Sigma}_0 = \text{diag}(0.25, \ldots, 0.25)$. In the MCMC algorithm, we choose a normal proposal density so that the acceptance rate is approximately 30%.

We simultaneously perturbed the cure rate model (7), a subset of the λ_j's, and the prior distributions of $\boldsymbol{\beta}$ and λ_j ($j = 1, \ldots, J$), whose perturbed log-posterior is given by

$$\begin{aligned}\log p(\theta|D_{com},\omega)=&\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K}\Big\{\delta_{ij}I(j\in T_k)\Big[\nu_i\log(N_i\lambda_j\omega_{\lambda k})-N_i\lambda_j\omega_{\lambda k}(y_i-s_{j-1})\Big]\\&-\delta_{ij}N_i\sum_{l=1}^{K}\sum_{g=1}^{j-1}\omega_{\lambda l}\lambda_g(s_g-s_{g-1})I(g\in T_l)\Big\}+\sum_{i=1}^{n}\Big\{N_i\left(\mathbf{x}_i^T\beta+\omega_i\right)\\&-\log(N_i!)-\exp\left(\mathbf{x}_i^T\beta+\omega_i\right)\Big\}-[\log|\textstyle\sum_0/\omega_{\beta 2}|\\&+\omega_{\beta 2}(\beta-\omega_{\beta 1}\boldsymbol{u}_0)^T\textstyle\sum_0^{-1}(\beta-\omega_{\beta 1}\boldsymbol{u}_0)]/2+\displaystyle\sum_{j=1}^{J}\{(\alpha_{0j}-1)\log\lambda_j\\&-\omega_{\alpha j}\lambda_j\alpha_{1j}+\alpha_{0j}\log(\omega_{\alpha j}\alpha_{1j})-\log\Gamma(\alpha_{0j})\},\end{aligned}\tag{33}$$

where δ_{ij} and ν_i are defined as before, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n, \omega_{\beta 1}, \omega_{\beta 2}, \omega_{\alpha 1}, \ldots, \omega_{\alpha J}, \omega_{\lambda 1}, \ldots, \omega_{\lambda K})^{\mathrm{T}}$, $T_k \in \{1, \ldots, J\}$ ($k = 1, \ldots, K$) is an index set and satisfies $T_{k_1} \cap T_{k_2} = \varphi$ for every $k_1 \neq k_2 \in \{1, \ldots, K\}$ and $T_1 \cup \ldots \cup T_K = \{1, \ldots, J\}$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$. In this case, $\boldsymbol{\omega}^0 = (0, \ldots, 0, 1, 1, 1, \ldots, 1)^T$ represents no perturbation. Here, we take K to be 5 and $T_1 = \{1, 2\}$, $T_2 = \{3, 4\}$, $T_3 = \{5, 6\}$, $T_4 = \{7, 8\}$, $T_5 = \{9, 10\}$. After some calculations, we have

$$\boldsymbol{G}(\omega^0)=\text{diag}\left(\boldsymbol{G}_N,\mu_0^T\textstyle\sum_0^{-1}\mu_0,4/2,\alpha_{01},\ldots,\alpha_{0j},\boldsymbol{H}_\lambda\right),$$

where $\boldsymbol{G}_N = E\{\xi\xi^T\} - E(\xi)E(\xi)^T$ with $\boldsymbol{\xi}=\left(N_1-\exp\left(\boldsymbol{x}_1^T\beta\right),\ldots,N_n-\exp\left(\boldsymbol{x}_n^T\beta\right)\right)^T$, $\boldsymbol{H}_\lambda = E\{\boldsymbol{\eta}\boldsymbol{\eta}^T\} - E(\boldsymbol{\eta})\mathrm{E}(\boldsymbol{\eta})^T$ with $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)^T$ in which

$$\eta_k=\sum_{i=1}^{n}\sum_{j\in T_k}\delta_{ij}\left\{\nu_i-N_i\left[\lambda_j(y_i-s_{j-1})+\sum_{g\in T_k\cap\{1,2,\ldots,j-1\}}\lambda_g(s_g-s_{g-1})\right]\right\}-\left(\sum_{i=1}^{n}N_i\sum_{j\in\{T_{k+1},\ldots,T_K\}}\delta_{ij}\right)\left(\sum_{g\in T_k}\lambda_g(s_g-s\right.$$

for $k = 1, \ldots, K$, and the expectation is taken with respect with the joint distribution of $(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\lambda})$. Then, we chose a new perturbation scheme $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}^0)$ and calculated the associated local influence measures $\mathrm{SI}_{\mathrm{D}_\varphi \mathbf{e}_j}$ and $\mathrm{SI}_{C\,M_h \mathbf{e}_j}$ in which $\varphi(\cdot)$ was chosen to be the Kullback divergence and $d(\boldsymbol{\theta}) = \boldsymbol{\theta}$. It is important to note here that cure rate models have improper survival functions (i.e., $\lim_{t\to\infty} S(t) > 0$), and therefore the survival density of the cure rate model integrates to a constant less than 1. In order for the metric tensor to be meaningful here, we must therefore normalize the survival density for the cure rate model so that it is a bonafide density (i.e., integrates to 1). After doing such a normalization, the metric tensor $\boldsymbol{G}$ is well defined for this model and has a theoretical justification. Such a theoretical justification, however, is beyond the scope of this paper.

Results for the Bayesian analysis are reported in Fig. 4. Figure 4a shows that cases 110, 132, 205, 221, 230, 237, 257, 264, 296, 326, 388 and 405 were detected to be influential by our local influence measures; whilst cases 26, 35, 77, 267, 296, 297, 341, 397 and 405 were detected to be influential by Fig. 4b. Cases 296 and 405 were identified to be influential by Fig. 4a, b.

4 Discussion

We have developed a Bayesian local influence method to perturb D_{com}, $p(\boldsymbol{\theta})$, or $p(D_{com}|\boldsymbol{\theta})$ in assessing minor perturbations to the prior and/or the sampling distribution in Bayesian survival analysis. We have introduced a perturbation model to characterize simultaneous (or individual) perturbations to the data, the prior distribution and the sampling distribution. We have constructed a Bayesian perturbation manifold to the perturbation model and calculated its associated geometric quantities including its metric tensor. We have developed first-order and second-order local influence measures based on several objective functions to quantify the degree of various perturbations to the statistical model. Finally, we have also examined a number of examples to highlight the broad spectrum of applications of this local influence method in Bayesian survival analysis.

Finally, we mention that in order for the metric tensor $\boldsymbol{G}$ to be well defined for the proposed methodology, the joint density of $(\boldsymbol{y}, \boldsymbol{\theta})$ needs to be proper, that is, it needs to integrate to 1. Therefore, proper priors for $\boldsymbol{\theta}$ must be used as well as proper sampling densities. The cure rate model, as noted earlier, does not have a proper sampling (survival) density and therefore, the survival density must first be normalized before the metric tensor and other geometric quantities can be calculated. The normalized survival density in this case is given by

$$f(\boldsymbol{y}|\theta)=\frac{f^{*}(\boldsymbol{y}|\theta)}{\int f^{*}(\boldsymbol{y}|\theta)d\boldsymbol{y}}$$

where $f^{*}(\boldsymbol{y}|\boldsymbol{\theta})$ is the (improper) survival density for the cure rate model. Theoretical justifications for these types of models as well as extensions of this methodology to Bayesian models with improper priors will be pursued elsewhere.

Acknowledgments

This work was supported in part by NSF grant SES-06-43663 to Dr. Zhu, and NIH grants GM 70335 and CA 74015 to Dr. Ibrahim, NSFC grant 10561008 and NCET grant NCET-07-0737 to Dr. Tang.

References

Amari, S. Lecture notes in statistics. 2. Springer-Verlag; Berlin: 1990. Differential-geometrical methods in statistics.

Anderson, PK.; Borgan, O.; Gill, RD.; Keiding, N. Statistical models based on counting processes. Springer-Verlag; New York: 1993.

Arjas E, Gasbarra D. Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. Stat Sin. 1994; 4:505–524.

Berger JO. Robust Bayesian analysis: sensitivity to the prior. J Stat Plan Inference. 1990; 25:303–328.

Berger JO. An overview of robust Bayesian analysis. TEST. 1994; 3:5–58.

Berger, JO.; Rios Insua, D.; Ruggeri, F. Bayesian robustness. In: Rios Insua, D.; Ruggeri, F., editors. Robust Bayesian analysis. Springer-Verlag; New York: 2000.

Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics. 2003; 59:221–228. [PubMed: 12926706]

Brown ER, Ibrahim JG. Bayesian approaches to joint cure rate and longitudinal models with applications to cancer vaccine trials. Biometrics. 2003; 59:686–693. [PubMed: 14601770]

Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomark-ers and survival. Biometrics. 2005; 61:64–73. [PubMed: 15737079]

Carlin BP, Polson NG. An expected utility approach to influence diagnostics. J Am Stat Assoc. 1991; 86:1013–1021.

Chen Z, Dunson DB. Bayesian estimation of survival functions under stochastic precedence. Lifetime Data Anal. 2004; 10:159–173. [PubMed: 15293630]

Chen M-H, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. J Am Stat Assoc. 1999; 94:909–919.

Chen, M-H.; Shao, Q-M.; Ibrahim, JG. Monte Carlo methods in Bayesian computation. Springer-Verlag; New York: 2000.

Chen MH, Harrington DP, Ibrahim JG. Bayesian cure rate models for malignant melanoma: a case study of eastern cooperative oncology group trial E1690. Appl Stat. 2002a; 51:135–150.

Chen MH, Ibrahim JG, Lipsitz SR. Bayesian methods for missing covariates in cure rate models. Lifetime Data Anal. 2002b; 8:117–146. [PubMed: 12048863]

Chen MH, Ibrahim JG, Sinha D. Bayesian inference for multivariate survival data with a surviving fraction. J Multivar Anal. 2002c; 80:101–126.

Chen MH, Ibrahim JG, Sinha D. A new joint model for longitudinal and survival data with a cure fraction. J Multivar Anal. 2004; 91:18–34.

Chen MH, Ibrahim JG, Shao QM. Posterior propriety and computation for the Cox regression model with applications to missing covariates. Biometrika. 2006; 93:791–807.

Chi Y, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. Biometrics. 2006; 62:432–445. [PubMed: 16918907]

Chi Y, Ibrahim JG. A new class of joint models for longitudinal and survival data accomodating zero and non-zero cure fractions: a case study of an international breast cancer study group trial. Stat Sin. 2007; 17:445–462.

Cho H, Ibrahim JG, Sinha D, Zhu H. Bayesian case influence diagnostics for survival models. Biometrics. 2009; 65:116–124. [PubMed: 18422991]

Christensen, R. Log-linear models and logistic regression. Springer-Verlag, Inc; New York: 1997.

Cook RD. Assessment of local influence (with discussion). J R Stat Soc B. 1986; 48:133–169.

Cooner F, Banerjee S, Carlin BP, Sinha D. Flexible cure rate modelling under latent activation schemes. J Am Stat Assoc. 2007; 102:560–572. [PubMed: 21031152]

Cox DR. Regression models and life tables (with discussion). J R Stat Soc B. 1972; 34:187–220.

Cox DR, Reid N. Parameter orthogonality and approximate conditional influence (with discussion). J R Stat Soc B. 1987; 49:1–39.

De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric nonproportional hazards survival modeling. Biometrics. 2009; 65:762–771. [PubMed: 19210742]

Dunson DB, Dinse GE. Bayesian models for multivariate current status data with informative censoring. Biometrika. 2002; 58:79–88.

Dunson DB, Herring AH. Bayesian inferences in the Cox model for order restricted alternatives. Biometrics. 2003; 59:918–925.

Dunson DB, Herring AH. Bayesian model selection and averaging in additive and proportional hazards. Lifetime Data Anal. 2005; 11:213–232. [PubMed: 15938547]

Dunson DB, Park H-H. Kernel stick breaking processes. Biometrika. 2008; 95:307–323. [PubMed: 18800173]

Dykstra RL, Laud PW. A Bayesian nonparametric approach to reliability. Ann Stat. 1981; 9:356–367.

Fleming, TR.; Harrington, DP. Counting processes and survival analysis. Wiley; New York: 1991.

Gallant AR, Nychka DW. Semiparametric maximum likelihood estimation. Economatrica. 1987; 55:363–390.

Geisser, S. Predictive inference: an introduction. Chapman & Hall; London: 1993.

Gelfand AE, Dey DK. Bayesian model choice: asymptotics and exact calculations. J R Stat Soc B. 1994; 56:501–514.

Gelfand, AE.; Dey, DK.; Chang, H. Model determination using predictive distributions, with implementation via sampling-based methods (Disc: P160-167). In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. Bayesian statistics. Vol. 4. Oxford University Press; Oxford: 1992. p. 147-159.

Gustafson P. Local sensitivity of inferences to prior marginals. J Am Stat Assoc. 1996a; 91:774–781.

Gustafson P. Local sensitivity of posterior expectations. Ann Stat. 1996b; 24:174–195.

Gustafson, P. Local robustness in Bayesian analysis. In: Rios Insua, D.; Ruggeri, F., editors. Robust Bayesian analysis. Springer-Verlag; New York: 2000. p. 71-88.

Gustafson P, Wasserman L. Local sensitivity diagnostics for Bayesian inference. Ann Stat. 1995; 23:2153–2167.

Hanson T, Johnson WO. Modeling regression error with a mixture of Polya trees. J Am Stat Assoc. 2002; 97:1020–1033.

Hanson T, Johnson W, Laud P. Semiparametric inference for survival models with step process covariates. Can J Stat. 2009; 37:60–79.

Ibrahim, JG.; Chen, M-H.; Sinha, D. Bayesian survival analysis. Springer-Verlag, Inc; New York: 2001.

Ibrahim JG, Chen M-H, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine studies. Stat Sin. 2004; 14:863–883.

Ibrahim JG, Chen MH, Kim S. Bayesian variable selection for the Cox regression model with missing covariates. Lifetime Data Anal. 2008; 14:496–520. [PubMed: 18836829]

Johnson W, Geisser S. A predictive view of the detection and characterization of influential observations in regression analysis. J Am Stat Assoc. 1983; 78:137–144.

Johnson W, Geisser S. Estimative influence measures for the multivariate general linear model. J Stat Plan Inference. 1985; 11:33–56.

Kalbfleisch JD. Nonparametric Bayesian analysis of survival time data. J R Stat Soc B. 1978; 40:214–221.

Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. 2. Wiley; New York: 2002.

Kass RE, Tierney L, Kadane JB. Approximate methods for assessing influence and sensitivity in Bayesian analysis. Biometrika. 1989; 76:663–674.

Kim S, Xi Y, Chen M-H. A new latent cure marker model for survival data. Ann Appl Stat. 2009; 3:1124–1146.

Lawless, JF. Statistical models and methods for life time data. 2. Wiley; New York: 2003.

Leonard T. Density estimation, stochastic processes and prior information. J R Stat Soc B. 1978; 40:113–146.

McCulloch RE. Local model influence. J Am Stat Assoc. 1989; 84:473–478.

Moreno, E. Global Bayesian robustness for some classes of prior distributions. In: Rios Insua, D.; Ruggeri, F., editors. Robust Bayesian analysis. Springer-Verlag; New York: 2000.

Oakley J, O’Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc B. 2004; 66:751–769.

Peng F, Dey DK. Bayesian analysis of outlier problems using divergences measures. Can J Stat. 1995; 23:199–213.

Pennell ML, Dunson DB. Bayesian semiparametric dynamic frailty models for multiple event time data. Biometrics. 2006; 62:1044–1052. [PubMed: 17156278]

Pettit LI. Diagnostics in Bayesian model choice. Statistician. 1986; 35:183–190.

Rodriguez A, Dunson DB, Gelfand AE. The nested Dirichlet process (with discussion). J Am Stat Assoc. 2008; 103:1131–1144.

Sinha D. Semiparametric Bayesian analysis of multiple event time data. J Am Stat Assoc. 1993; 88:979–983.

Sinha D, Dey DK. Semiparametric Bayesian analysis of survival data. J Am Stat Assoc. 1997; 92:1195–1212.

Sinha D, Maiti T. A Bayesian approach for the analysis of panel-count data with dependent termination. Biometrics. 2004; 60:34–40. [PubMed: 15032771]

Sinha D, Ibrahim JG, Chen M-H. A Bayesian justification of Cox's partial likelihood. Biometrika. 2003; 90:629–641.

Sinha D, Maiti T, Ibrahim JG, Ouyang B. Current methods for recurrent events data with dependent termination: a Bayesian perspective. J Am Stat Assoc. 2008; 103:866–878. [PubMed: 19169433]

Sinha D, McHenry BM, Lipsitz SR, Ghosh M. Empirical Bayes estimation for additive hazards regression models. Biometrika. 2009; 96:529–544.

Sivaganesan, S. Global and local robustness approaches: uses and limitations. In: Rios Insua, D.; Ruggeri, F., editors. Robust Bayesian analysis. Springer-Verlag; New York: 2000. p. 89-108.

Vaupel JM, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 1979; 16:439–454. [PubMed: 510638]

Weiss R. An approach to Bayesian sensitivity analysis. J R Stat Soc B. 1996; 58:739–750.

Weiss RE, Cho M. Bayesian marginal influence assessment. J Stat Plan Inference. 1998; 71:163–177.

Weiss RE, Cook RD. A graphical case statistic for assessing posterior influence. Biometrika. 1992; 79:51–55.

Yakovlev AY. Letter to the editor. Stat Med. 1994; 13:983–986. [PubMed: 8080583]

Yin G, Ibrahim JG. A class of Bayesian shared gamma frailty models with multivariate failure time data. Biometrics. 2005a; 61:208–216. [PubMed: 15737095]

Yin G, Ibrahim JG. Bayesian frailty models based on Box-Cox transformed hazards. Stat Sin. 2005b; 15:781–794.

Yin G, Ibrahim JG. A general class of Bayesian survival models with zero and non-zero cure fractions. Biometrics. 2005c; 61:403–412. [PubMed: 16011686]

Yin G, Ibrahim JG. Cure rate models: a unified approach. Can J Stat. 2005d; 33:559–570.

Zhu HT, Lee SY. Local influence for incomplete data models. J R Stat Soc B. 2001; 63:111–126.

Zhu HT, Zhang HP. A diagnostic procedure based on local influence. Biometrika. 2004; 91:579–589.

Zhu HT, Ibrahim JG, Lee SY, Zhang HP. Perturbation selection and influence measures in local influence analysis. Ann Stat. 2007; 35:2565–2588.

Zhu, HT.; Ibrahim, JG.; Tang, NS. Bayesian influence analysis: a geometric approach. 2009. Submitted

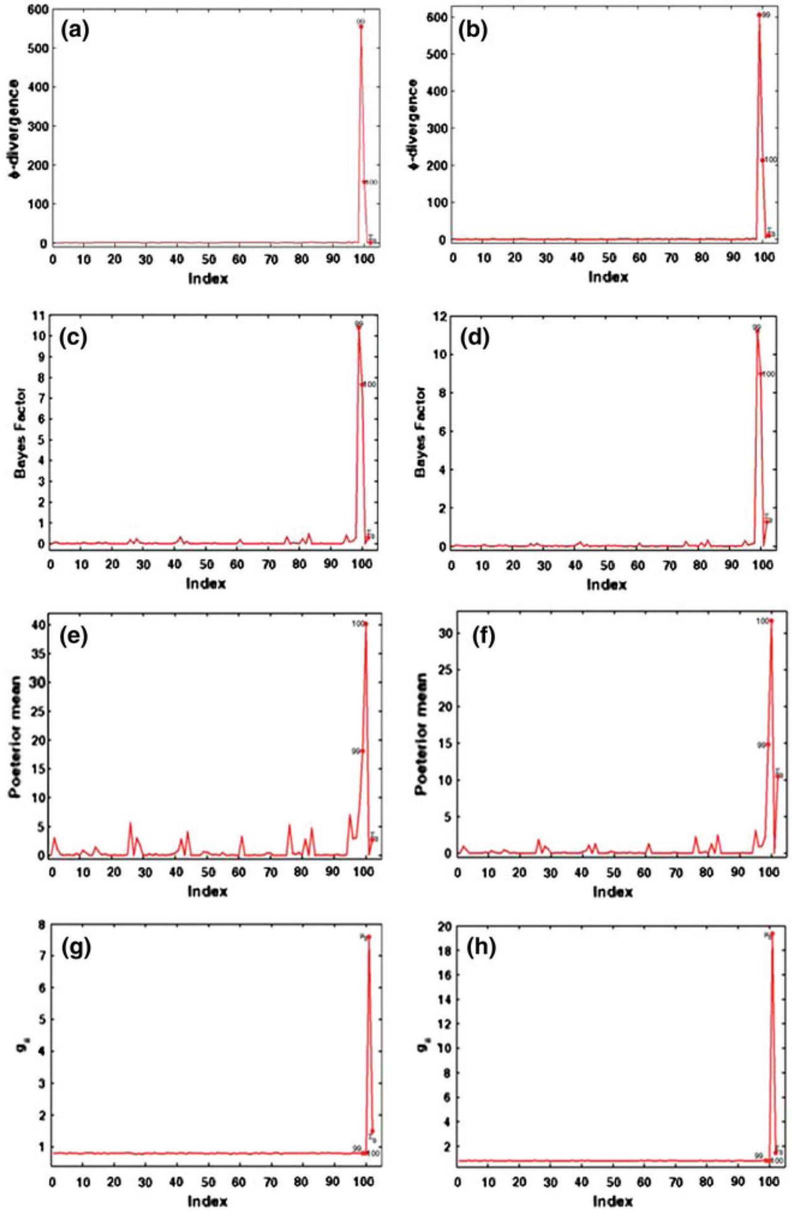

Fig. 1.

Index plots of local influence measures for simultaneous perturbation. In the first column, three local influence measures including (**a**) $SI_{D_{\varphi}\mathbf{e}_j}$, (**c**) $\mathbf{h}^B_{\max}$, and (**e**) $SI_{M_d\mathbf{e}_j}$ can detect the two influential cases: 99 and 100, and (**g**) g_{ii}. In the second column, three local influence measures including (**b**) $SI_{D_{\varphi}\mathbf{e}_j}$, (**d**) $\mathbf{h}^B_{\max}$, and (**f**) $SI_{M_d\mathbf{e}_j}$ can detect both the two influential cases (99 and 100) and the impact of the perturbed prior distribution for $\boldsymbol{\beta}$, and (**h**) g_{ii}

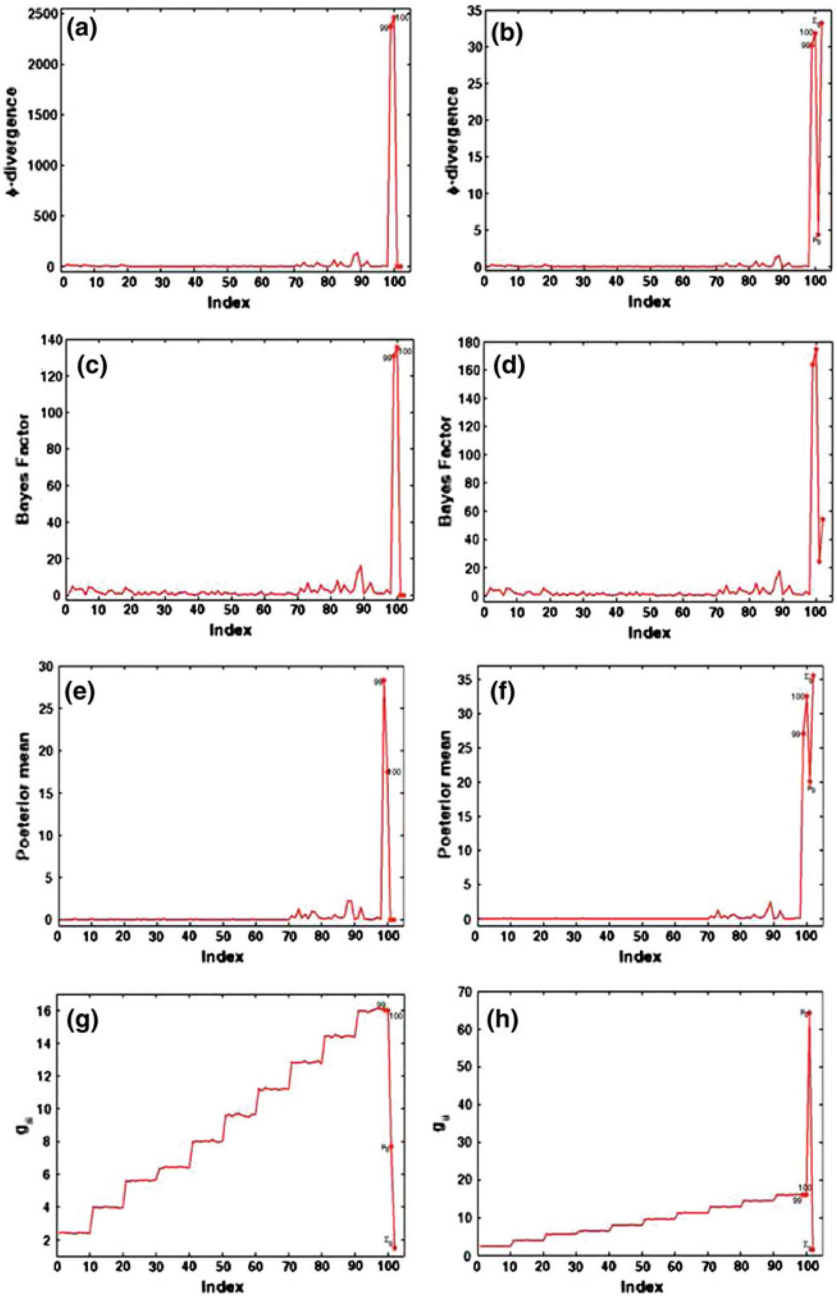

Fig. 2.

Index plots of local influence measures for simultaneous perturbation. In the first column, three local influence measures including (**a**) $\mathbf{h}^B_{\max}$, (**c**) $\mathrm{SI}_{\mathrm{D}_{\varphi}\mathbf{e}_j}$, and (**e**) $\mathrm{SI}_{C\,M_h\mathbf{e}_j}$ can detect the two influential cases: 99 and 100, and (**g**) g_{ii}. In the second column, three local influence measures including (**b**) $\mathbf{h}^B_{\max}$, (**d**) $\mathrm{SI}_{\mathrm{D}_{\varphi}\mathbf{e}_j}$, and (**f**) $\mathrm{SI}_{C\,M_h\mathbf{e}_j}$ can detect both the two influential cases (99 and 100) and the impact of the perturbed prior distribution for $\boldsymbol{\beta}$, and (**h**) g_{ii}

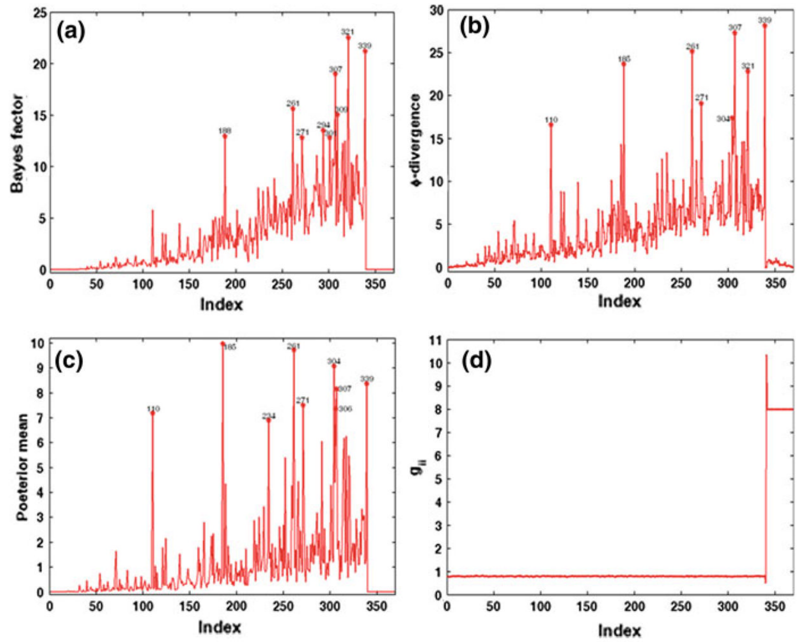

Fig. 3.

Index plots of local influence measures (**a**) $\mathbf{h}^B_{\max}$, (**b**) $SI_{D_\varphi \mathbf{e}_j}$, (**c**) $SI_{C\,M_h \mathbf{e}_j}$, and (**d**) g_{ii} for simultaneous perturbation (30)

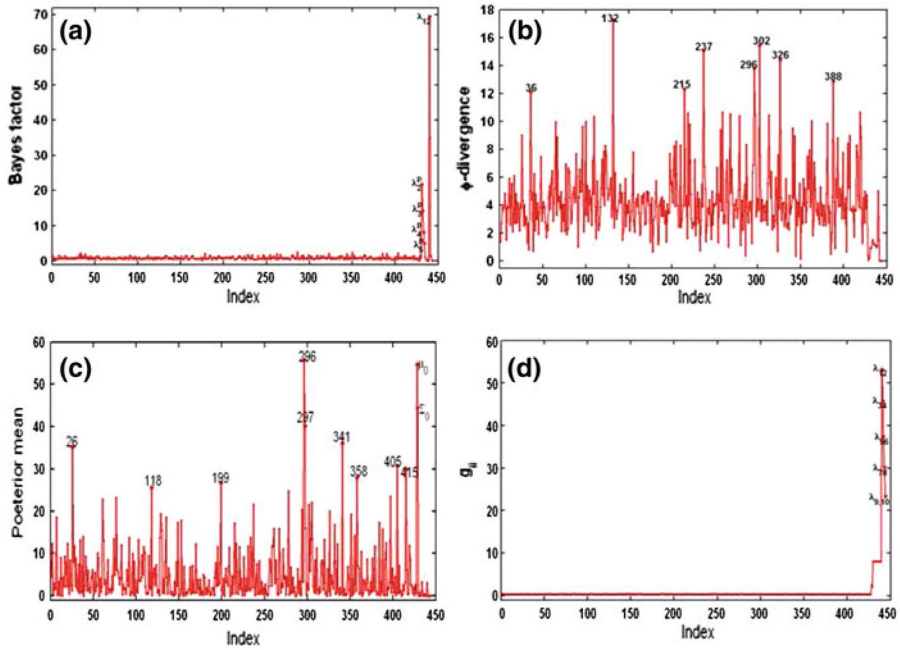

Fig. 4.
Index plots of local influence measures (**a**) $SI_{D_\varphi \mathbf{e}_j}$, (**b**) $SI_{C M_h \mathbf{e}_j}$, and (**c**) g_{ii} for simultaneous perturbation (31)