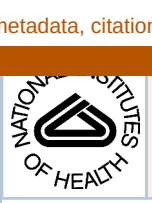

Published in final edited form as:
Biometrics. 2009 March ; 65(1): 116–124. doi:10.1111/j.1541-0420.2008.01037.x.

Bayesian Case Influence Diagnostics for Survival Models

Hyunsoon Cho[1,*], **Joseph G. Ibrahim**[1,**], **Debajyoti Sinha**[2,***], and **Hongtu Zhu**[1,****];

1 *Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, U.S.A*

2 *Department of Statistics, Florida State University, Tallahassee, Florida 32306-4330, U.S.A*

Summary

We propose Bayesian case influence diagnostics for complex survival models. We develop case deletion influence diagnostics for both the joint and marginal posterior distributions based on the Kullback–Leibler divergence (K–L divergence). We present a simplified expression for computing the K–L divergence between the posterior with the full data and the posterior based on single case deletion, as well as investigate its relationships to the conditional predictive ordinate. All the computations for the proposed diagnostic measures can be easily done using Markov chain Monte Carlo samples from the full data posterior distribution. We consider the Cox model with a gamma process prior on the cumulative baseline hazard. We also present a theoretical relationship between our case-deletion diagnostics and diagnostics based on Cox's partial likelihood. A simulated data example and two real data examples are given to demonstrate the methodology.

Keywords

Bayesian diagnostic measure; Cox model; Influential observation; Kullback; Leibler divergence; Marginal influence; Partial likelihood

1. Introduction

The importance of identification of influential observations in a statistical analysis is a well-recognized methodological problem, and the development of diagnostic measures to detect influential observations is of interest to many researchers. Influential observations in a given dataset can have a strong impact on statistical inference and conclusions. In these situations, such influential observations are an important part of the data, and hence require the most careful examination. A common way of assessing the influence of an observation on model fit is through case deletion. In Bayesian analysis, the Kullback–Leibler divergence (K–L divergence) based on case deletion is a measure of discrepancy between the posterior distributions with and without a particular case, and it is a popular Bayesian diagnostic measure. Another popular Bayesian diagnostic measure is the conditional predictive ordinate (CPO) (Gelfand et al., 1992; Geisser, 1993), which is defined as the predictive density of the *i*th case given the data without the *i*th case. A large value of the K–L divergence for the *i*th case implies more influence of the *i*th case on estimation, hypothesis testing, and model fit. A large value

*email: hscho@bios.unc.edu.
**email: ibrahim@bios.unc.edu
***email: sinhad@stat.fsu.edu
****email: hzhu@bios.unc.edu

6. Supplementary Materials

Web Appendices, Table, and Figure referenced in Sections 2, 3, and 4 are available under the Paper Information link at the *Biometrics* website http://www.biometrics.tibs.org.

of CPO for the *i*th case implies better concordance of the *i*th case with the rest of the data, and hence a better model fit.

In Bayesian analysis, considerable research has been done for developing case influence diagnostics using the K–L divergence under various parametric models (Johnson and Geisser, 1985; Pettit, 1986; Carlin and Polson, 1991; Weiss and Cook, 1992; Peng and Dey, 1995; Weiss, 1996; Christensen, 1997; Weiss and Cho, 1998). Pettit (1986) suggested the use of the K–L divergence in detecting influential observations in his review of Bayesian diagnostics. Carlin and Polson (1991) proposed an expected utility approach using the K–L 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 K–L 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 K–L divergence and CPO. Bayesian influence measures for assessing marginal posterior distributions have also been developed for the multivariate linear model and normal random effects models (Johnson and Geisser, 1985; Weiss and Cho, 1998).

Despite the extensive literature on Bayesian diagnostic methods for parametric models, very little has been developed for semiparametric models, including survival models. Due to the potential advantages of fitting a vast array of complex survival models posed by modern survival data, semiparametric Bayesian methodologies in survival analysis have been getting enormous attention in biomedical research. Bayesian case influence diagnostics for survival models pose both theoretical and computational challenges, which are discussed here.

The objective of this article is to propose Bayesian case-deletion influence diagnostics for survival models. First, we develop diagnostic measures to assess the influence of a case on both the joint and marginal posterior distributions based on the directed K–L divergence. In this development, we derive a novel and simplified expression for computing the K–L divergence, which facilitates efficient computation of the proposed diagnostic measures using Markov chain Monte Carlo (MCMC) samples from full data posterior distribution. This avoids the burden of sampling from each of the n posterior distributions, each based on deletion of the ith case, $i = 1, \ldots, n$. Second, we apply the proposed methodology to Bayesian survival models with continuous survival time data. The survival model we consider is the Cox model with a gamma process prior on the cumulative baseline hazard (Sinha, Ibrahim, and Chen, 2003). In addition, we investigate a theoretical connection between the proposed diagnostics based on the K–L divergence and CPO, as well as a connection between diagnostics based on Cox's partial likelihood.

To motivate the proposed methodology, we consider a well known dataset, the Stanford heart transplant data (Miller and Halpern, 1982). The dataset contains 184 transplant cases with the following variables: time measured from the date of the transplant in days; status code (dead or alive); patient age at first transplant in years; and T5 mismatch score (missing for 27 of the cases). This dataset has been analyzed by many, illustrating frequentist diagnostic measures (Pettitt and Daud 1989; Escobar and Meeker, 1992). Here, it is of interest to carry out Bayesian diagnostic methods not only to compare our results with the frequentist results, but also to possibly find other influential (or noninfluential) cases not identified by the previous methods. As shown in Figure 2, our proposed Bayesian diagnostic method identified some cases as influential in this dataset. More details regarding this example are given in Section 4.2. To further illustrate the methodology, we also apply the proposed methods to simulated data and a phase III melanoma clinical trial (E1690) discussed in Sections 4.1 and 4.3, respectively.

The rest of this article is organized as follows. In Section 2, we introduce Bayesian case influence diagnostics based on the K–L divergence. In Section 3, we derive case influence diagnostics for the Cox model with a gamma process prior. In Section 4, we examine the performance of the influence diagnostics using simulated data, the Stanford heart transplant data, and the E1690 trial. We conclude the article with some discussion in Section 5.

2. The Proposed Method

2.1 General Development

Let D be full data and D_{-i} be the data with the ith case deleted. Let $L(\boldsymbol{\beta}|D)$ denote the likelihood based on the full data and $L(\boldsymbol{\beta}|D_{-i})$ denote the likelihood based on the data without the ith case. The posterior distributions for the full data and the ith case deleted can be defined as $p(\boldsymbol{\beta}|D) \propto L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})$ and $p(\boldsymbol{\beta}|D_{-i}) \propto L(\boldsymbol{\beta}|D_{-i})\pi(\boldsymbol{\beta})$, respectively, where $\pi(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$. A typical choice of $\pi(\boldsymbol{\beta})$ is a $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ distribution or a uniform improper prior.

Let $K(P, P_{-i})$ denote the K–L divergence between P and P_{-i}, where P denotes the posterior distribution of $\boldsymbol{\beta}$ for the full data, and P_{-i} denotes the posterior distribution of $\boldsymbol{\beta}$ without the ith case. Specifically,

$$K(P,P_{-i})=\int p(\beta|D)\log\left\{\frac{p(\beta|D)}{p(\beta|D_{-i})}\right\}d\beta. \tag{1}$$

$K(P, P_{-i})$ thus measures the effect of deleting the ith case from the full data on the joint posterior distribution of $\boldsymbol{\beta}$. Note that $K(P, P_{-i}) \neq = K(P_{-i}, P)$ in general. After some algebra, as shown in Web Appendix A, we can derive a simplified expression for $K(P, P_{-i})$ as follows:

$$\begin{aligned}K(P,P_{-i})&=\log E_\beta\left[\frac{L(\beta|D_{-i})}{L(\beta|D)}|D\right]\\&+E_\beta\left[\log\left\{\frac{L(\beta|D)}{L(\beta|D_{-i})}\right\}|D\right],\end{aligned} \tag{2}$$

where $E_{\boldsymbol{\beta}}[\cdot|D]$ represents the expectation with respect to the joint posterior distribution of $\boldsymbol{\beta}$ given D. Equation (2) enables us to compute $K(P, P_{-i})$ for $i = 1 \ldots n$, using only samples from the full data joint posterior distribution of $\boldsymbol{\beta}$. Therefore, equation (2) implies that we completely avoid sampling from $p(\boldsymbol{\beta}|D_{-i})$ for the computation of $K(P, P_{-i})$, and this saves us enormous computational time and effort.

Now suppose that interest lies in assessing the influence of the ith case on the subset $\boldsymbol{\beta}_1$ of the parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Weiss and Cho (1998), Weiss (1996), and Weiss and Cook (1992) pointed out that if the goal of an analysis is to assess the influence of the ith case on the marginal posterior distribution of $\boldsymbol{\beta}_1$, then using the joint posterior of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ to assess this influence may overstate the influence. Hence, in these settings, we need to consider the influence of a case using the marginal posterior distribution of $\boldsymbol{\beta}_1$.

We can express the marginal influence diagnostics of Weiss and Cho (1998) based on the directed K–L divergence as

$$K(P_1,p_{1,-i})=\int p_1(\beta_1|D)\log\left\{\frac{p_1(\beta_1|D)}{p_1(\beta_1|D_{-i})}\right\}d\beta_1, \tag{3}$$

where $p_1(\boldsymbol{\beta}_1|D) = \int p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|D)\, d\boldsymbol{\beta}_2$. The marginal K–L divergence, $K(P_1, P_{1,-i})$, in (3) measures the effect of deleting the ith case from the full data on the marginal posterior distribution of $\boldsymbol{\beta}_1$. Using similar derivations as in (2), we can obtain a simplified expression for $K(P_1, P_{1,-i})$ as follows:

$$\begin{aligned} K(P_1,P_{1,-i}) &= \log E_\beta\left[\frac{L(\beta|D_{-i})}{L(\beta|D)}|D\right] \\ &\quad -E_{\beta_1}\left[\log\int\frac{L(\beta|D_{-i})}{L(\beta|D)}p(\beta_2|\beta_1,D)d\beta_2|D\right], \end{aligned} \tag{4}$$

where $p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, D) = p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|D)/\int p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|D)\, d\boldsymbol{\beta}_2$ and $\int\frac{L(\beta|D_{-i})}{L(\beta|D)}p(\beta_2|\beta_1,D)d\beta_2$ can be evaluated as $E_{\beta_2}[\frac{L(\beta|D_{-i})}{L(\beta|D)}|\beta_1,D]$.

Following McCulloch (1989), calibration of $K(P, P_{-i})$ can be done by solving for p_i such that $K(P, P_{-i}) = K(B(0.5), B(p_i)) = -\log\{4p_i(1 - p_i)\}/2$, where $B(p)$ denotes the Bernoulli distribution with success probability p. This implies that describing outcomes using $p(\boldsymbol{\beta}|D_{-i})$ instead of $p(\boldsymbol{\beta}|D)$ is compatible with describing an unobserved event as having probability p_i when the correct probability is 0.5. After calculating $K(P, P_{-i})$ from (2), we can compute p_i using $p_i=0.5[1+\sqrt{1-\exp\{-2K(P,P_{-i})\}}]$. This equation implies that $0.5 \le p_i \le 1$. $p_i \gg 0.5$ implies that the ith case is influential, because deleting the ith case changes the posterior distribution as much as describing an observed event as having probability p_i when the correct probability is 0.5. In this article, we use p_i as the calibration of $K(P, P_{-i})$ in all of the examples.

2.2 Independence Model

As an illustration, we consider the proposed diagnostic method for the independence model. Suppose that given $\boldsymbol{\beta}$, y_i, $i = 1, 2, \ldots, n$ are independent response variables, not subject to censoring. Then the full data likelihood is $L(\beta|D)=\prod_{k=1}^{n} f(y_k|\beta)$, where $f(y_k|\boldsymbol{\beta})$ is the density of y_k and the likelihood without the ith observation is $L(\beta|D_{-i})=\prod_{k=1,k\neq i}^{n} f(y_k|\beta)$. Therefore, $L(\boldsymbol{\beta}|D)/L(\boldsymbol{\beta}|D_{-i}) = f(y_i|\boldsymbol{\beta})$ and the CPO is given by $CPO_i = [E_\beta[\{f(y_i|\boldsymbol{\beta})\}^{-1}|D]]^{-1}$ (Gelfand et al., 1992).

Using (2) and the above results, we can therefore show that

$$\begin{aligned} K(P,P_{-i}) &= \log E_\beta[\{f(y_i|\beta)\}^{-1}|D] + E_\beta[\log\{f(y_i|\beta)\}|D] \\ &= -\log(CPO_i) + E_\beta[\log\{f(y_i|\beta)\}|D]. \end{aligned} \tag{5}$$

Similarly, using equation (4) we can obtain $K(P_1, P_{1,-i})$ for the influence of the ith case on the marginal posterior distribution of $\boldsymbol{\beta}_1$ and its connection with CPO as follows:

$$\begin{aligned} K(P_1, P_{1,-i}) &= \log E_\beta[\{f(y_i|\beta)\}^{-1}|D] \\ &\quad -E_{\beta_1}\left[\log\int\{f(y_i|\beta)\}^{-1}p(\beta_2|\beta_1,D)d\beta_2|D\right] \\ &= -\log(CPO_i) - E_{\beta_1}\left[\log\int\{f(y_i|\beta)\}^{-1}p(\beta_2|\beta_1,D)d\beta_2|D\right], \end{aligned} \tag{6}$$

where $p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, D) = p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|D)/\int p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|D)\, d\boldsymbol{\beta}_2$ and $\int\{f(y_i|\boldsymbol{\beta})\}^{-1}p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, D)\, d\boldsymbol{\beta}_2$ can be evaluated as $E_{\boldsymbol{\beta}_2}[\{f(y_i|\boldsymbol{\beta})\}^{-1}|\boldsymbol{\beta}_1, D]$. Because equations (2) and (4)–(6) are expressed as posterior expectations with respect to the full data posterior distribution, they can be easily calculated using only MCMC samples from the full data posterior distribution of $\boldsymbol{\beta}$.

3. Cox Model with Gamma Process Prior

3.1 Model

In the Cox proportional hazards model (Cox, 1972), the gamma process is a very commonly used nonparametric prior process for the cumulative baseline hazard (Kalbfleisch, 1978). The full data are denoted as $D = \{\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{X}\}$, where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ denotes the observed survival times, where y_i may be right censored. We assume that the survival times are all distinct and ordered, i.e., $0 < y_1 < y_2 < \cdots < y_n < \infty$. $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)'$ is an indicator vector with $\delta_i = 1$ if the ith subject failed, and $\delta_i = 0$ if the ith subject was right censored. Also, $\boldsymbol{X}$ is an $n \times p$ matrix of covariates with ith row x_i', and $D_{-i} = \{\boldsymbol{y}_{-i}, \boldsymbol{\delta}_{-i}, \boldsymbol{X}_{-i}\}$ denotes the data with the ith subject (i.e., (y_i, δ_i, x_i')) deleted from D. The hazard function is given by $h(y_i|x_i)=h_0(y_i)\exp(x_i'\beta)$, where $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown regression coefficients, and $h_0(.)$ is an unknown baseline hazard function.

Under the Cox model, the joint probability of the survival of n subjects is given by

$$P(Y>y|\beta,X,H_0)=\exp\left\{-\sum_{k=1}^{n}H_0(y_k)\exp(x_k'\beta)\right\}, \tag{7}$$

where $H_0(y)$ is the cumulative baseline hazard (Ibrahim, Chen, and Sinha, 2001). We take $H_0 \sim GP\,(cH^*(\cdot), c)$, where GP denotes gamma process, $H^*(y)$ is a known differentiable parametric function that represents a parametric guess for the cumulative baseline hazard $H_0(y)$, and $c \geq 0$ is a confidence parameter. $H^*(y)$ is thus the mean of the process. Letting $h_k = H_0(y_k) - H_0(y_{k-1})$, we take $h_k \sim Gamma(ch_{0k}, c)$, the h_k's are independent, where $h_{0k} = H^*(y_k) - H^*(y_{k-1})$ and $Gamma(\alpha, \lambda)$ denotes the gamma distribution with mean $\alpha/\lambda(\alpha > 0$ and $\lambda > 0)$.

The marginal likelihood function of $\boldsymbol{\beta}$ can now be written as follows (Ibrahim et al., 2001; Sinha et al., 2003):

$$\begin{aligned} L(\beta|D)= & \prod_{k=1}^{n} L_k(\beta|D) \\ = & \prod_{k=1}^{n}\exp\left[cH^*(y_k)\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k}\right\}\right] \\ & \times\left[-ch^*(y_k)\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k}\right\}\right]^{\delta_k}, \end{aligned} \tag{8}$$

where $h^*(y)=\frac{d}{dy}H^*(y), A_k=\sum_{l\in\mathcal{R}(y_k)}\exp(x_i'\beta)$, and $\mathcal{R}(y_k) = \{l: y_l \geq y_k\}$ is the set of subjects at risk at time y_k.

We now derive the likelihood function without the ith subject. If $y_k < y_i$ then the risk set at time y_k involves the ith subject, otherwise, the risk set at y_k does not involve the ith subject. Therefore, after deleting the ith subject, the risk set changes to $\mathcal{R}(y_k) = \{l: y_l \geq y_k, l \neq i\}$ for $k < i$. As the risk set changes, the corresponding A_k in the denominators of (8) changes to $A_k - \exp(\boldsymbol{x}'_i\boldsymbol{\beta})$ for $k < i$, whereas for $k > i$, the risk set and A_k remain the same (see Web Appendix A for details). Hence, the likelihood function without the ith subject is given by

$$
\begin{aligned}
&L(\beta|D_{-i})\\
&=\prod_{k=1}^{i-1} L_{k,-i}(\beta|D)\prod_{k=i+1}^{n} L_k(\beta|D)\\
&=\prod_{k=1}^{i-1}\exp\left[cH^*(y_k)\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k-\exp(x_i'\beta)}\right\}\right]\\
&\times\left[-ch^*(y_k)\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k-\exp(x_i'\beta)}\right\}\right]^{\delta_k}\\
&\times\prod_{k=i+1}^{n}\exp\left[cH^*(y_k)\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k}\right\}\right]\\
&\times\left[-ch^*(y_k)\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k}\right\}\right]^{\delta_k}.
\end{aligned}
\tag{9}
$$

The posterior distributions based on the full data and the data without the *i*th subject are thus given by $p(\boldsymbol{\beta}|D) \propto L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})$ and $p(\boldsymbol{\beta}|D_{-i}) \propto L(\boldsymbol{\beta}|D_{-i})\pi(\boldsymbol{\beta})$, respectively.

3.2 Diagnostic Measures

For the Cox model in general, the likelihood function cannot be written as a product of *n* independent terms because the risk set for the *k*th subject involves observations other than the *k*th subject. Because of this dependency, we use (8) for the likelihood function. Another advantage of (8) is its computational feasibility. Because the hazard, h_k, has been integrated out from (8), it is only a function of $\boldsymbol{\beta}$. Therefore, sampling the h_k's is not necessary for Bayesian inference and diagnostics, and thus only samples from the posterior distribution of $\boldsymbol{\beta}$ are needed.

After some algebra, the ratio of likelihoods for the full data and the data without the *i*th subject can be written as $L(\boldsymbol{\beta}|D)/L(\boldsymbol{\beta}|D_{-i}) = g_i(\boldsymbol{\beta})L_i(\boldsymbol{\beta}|D)$. Thus, we can get a simplified expression for computing the influence of the *i*th subject on the joint posterior distribution of $\boldsymbol{\beta}$ as follows:

$$
\begin{aligned}
K(P,P_{-i}) = &\ \log E_\beta[\{g_i(\beta)L_i(\beta|D)\}^{-1}|D]\\
&+E_\beta[\log\{g_i(\beta)L_i(\beta|D)\}|D]\\
= &\ -\log(CPO_i)+E_\beta[\log L_i(\beta|D)|D]\\
&+\log[E_\beta[\{g_i(\beta)\}^{-1}|D]]+E_\beta[\log g_i(\beta)|D],
\end{aligned}
\tag{10}
$$

where

$$
\begin{aligned}
L_i(\beta|D)=&\exp\left[cH^*(y_i)\log\{1-\frac{\exp(x_i'\beta)}{c+A_i}\}\right]\\
&\times\left[-ch^*(y_i)\log\{1-\frac{\exp(x_i'\beta)}{c+A_i}\}\right]^{\delta_i},
\end{aligned}
\tag{11}
$$

$g_i(\beta)=\prod_{k=1}^{i-1}L_k(\beta|D)/\prod_{k=1}^{i-1}L_{k,-i}(\beta|D)$, which can be simplified as

$$
g_i(\beta)=\frac{\prod_{k=1}^{i-1}\left[1-\frac{\exp(x_k'\beta)}{c+A_k}\right]^{cH*(y_k)}\left[-\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k}\right\}\right]^{\delta_k}}{\prod_{k=1}^{i-1}\left[1-\frac{\exp(x_k'\beta)}{c+A_k-\exp(x_i'\beta)}\right]^{cH*(y_k)}\left[-\log\left\{1-\frac{\exp(x_k'\beta)}{c+A_k-\exp(x_i'\beta)}\right\}\right]^{\delta_k}}.
\tag{12}
$$

In addition, CPO_i can be written as,

$$CPO_i=\frac{E_\beta[\{g_i(\beta)\}^{-1}|D]}{E_\beta[\{g_i(\beta)L_i(\beta|D)\}^{-1}|D]}. \tag{13}$$

Because (10) is expressed as a posterior expectation with respect to the full data, computation of (10) can be done using MCMC samples from the full data posterior $p(\boldsymbol{\beta}|D)$. The samples from $p(\boldsymbol{\beta}|D)$ can be easily obtained using adaptive rejection Metropolis sampling (ARMS; Gilks, Best, and Tan et al., 1995) within Gibbs. Specifically, we have

$$\begin{aligned}K(P,P_{-i})&=\log\left[\frac{1}{J}\sum_{j=1}^{J}\{g_i(\beta^{(j)})L_i(\beta^{(j)}|D)\}^{-1}\right]\\&+\frac{1}{J}\sum_{j=1}^{J}\log\{g_i(\beta^{(j)})L_i(\beta^{(j)}|D)\},\end{aligned} \tag{14}$$

and

$$CPO_i=\frac{\frac{1}{J}\sum_{j=1}^{J}\{g_i(\beta^{(j)})\}^{-1}}{\frac{1}{J}\sum_{j=1}^{J}\{g_i(\beta^{(j)})L_i(\beta^{(j)}|D)\}^{-1}} \tag{15}$$

where J is the number of Gibbs samples after burn-in and $\beta^{(j)}=(\beta_1^{(j)},\ldots,\beta_p^{(j)})'$ is the jth Gibbs sample, $j = 1, \ldots, J$.

Similarly, we obtain

$$\begin{aligned}&K(P_1,P_{1,-i})\\&=\log E_\beta[\{g_i(\beta)L_i(\beta|D)\}^{-1}|D]\\&-E_{\beta_1}\left[\log\int\{g_i(\beta)L_i(\beta|D)\}^{-1}p(\beta_2|\beta_1,D)d\beta_2|D\right]\\&=-\log(CPO_i)+\log E_\beta[\{g_i(\beta)\}^{-1}|D]\\&-E_{\beta_1}\left[\log\int\{g_i(\beta)L_i(\beta|D)\}^{-1}p(\beta_2|\beta_1,D)d\beta_2|D\right].\end{aligned} \tag{16}$$

Monte Carlo evaluation of $E_{\boldsymbol{\beta}_1}[\log\int\{g_i(\boldsymbol{\beta})L_i(\boldsymbol{\beta}|D)\}^{-1}\times p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, D)\, d\boldsymbol{\beta}_2|D]$ in (16) can be obtained using the following steps:

Step 1: We use Gibbs sampling to obtain the samples $\beta^{(j)}=(\beta_1^{(j)},\beta_2^{(j)})$ for $j = 1, \ldots, J$ from $p(\boldsymbol{\beta}|D)$ and record $(\beta_1^{(1)},\ldots,\beta_1^{(J)})$ as J Gibbs samples from the marginal posterior of $\boldsymbol{\beta}_1$, $p_1(\boldsymbol{\beta}_1|D)$.

Step 2: We use Gibbs sampling to obtain the samples $\beta^{(r)}=(\beta_1^{(r)},\beta_2^{(r)})$ for $r = 1, \ldots, R$ from $p(\boldsymbol{\beta}|D)$ and record $(\beta_2^{(1)},\ldots,\beta_2^{(R)})$ as R Gibbs samples from the marginal posterior of $\boldsymbol{\beta}_2$ given $\boldsymbol{\beta}_1$, $p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, D)$.

Step 3: For each $\beta_1^{(j)}$, use $\beta_2^{(r)}$ as nested Gibbs samples from $p(\beta_2|\beta_1^{(j)},D)$ to get the Monte Carlo approximation of $E_{\boldsymbol{\beta}_1}[\log\int\{g_i(\boldsymbol{\beta})L_i(\boldsymbol{\beta}|D)\}^{-1}p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, D)\,d\boldsymbol{\beta}_2|D]$ as

$$\frac{1}{J}\sum_{j=1}^{J}\log\left[\frac{1}{R}\sum_{r=1}^{R}\{g_i(\beta_1^{(j)},\beta_2^{(r)})L_i(\beta_1^{(j)},\beta_2^{(r)}|D)\}^{-1}\right].$$

Note that the Gibbs samples in the first and second steps need to be sampled independently. Now, we can get the MCMC approximation of (16) as

$$\begin{aligned}&K(P_1,P_{1,-i})\\&=\log\left[\frac{1}{J}\sum_{j=1}^{J}\{g_i(\beta_1^{(j)},\beta_2^{(j)})L_i(\beta_1^{(j)},\beta_2^{(j)}|D)\}^{-1}\right]\\&-\frac{1}{J}\sum_{j=1}^{J}\log\left[\frac{1}{R}\sum_{r=1}^{R}\{g_i(\beta_1^{(j)},\beta_2^{(r)})L_i(\beta_1^{(j)},\beta_2^{(r)}|D)\}^{-1}\right].\end{aligned}\tag{17}$$

After computing $K(P, P_{-i})$ or $K(P_1, P_{1,-i})$ for all subjects, we can plot $K(P, P_{-i})$ or $K(P_1, P_{1,-i})$ across subjects to identify influential cases.

Because $K(P, P_{-i})$ measures the effect of deleting the ith case on the joint posterior distribution of $\boldsymbol{\beta}$, it can be viewed as a Bayesian analogue of the likelihood displacement (LD), as discussed in Cook and Weisberg (1982), and Cook (1986). Specifically, for the Cox model, $K(P, P_{-i})$ is comparable to the LD based on partial likelihood, which is available in `SAS version 9.1.3`. For more on LD for the Cox model, see Pettitt and Daud (1989). In addition, a limiting expression for $K(P, P_{-i})$ based on model (8) provides a method for computing $K(P, P_{-i})$ under Cox's partial likelihood. The detailed derivations are given in Web Appendix A. Moreover, it can be shown that $K(P_1, P_{1,-i})$ can be approximated by a quadratic form in the weighted score residuals upon a Taylor's series expansion. However, detailed derivations of these results are beyond the scope of this article and will be reported elsewhere.

4. Illustrative Examples

In this section, we illustrate our methodology with simulated data and two real datasets.

4.1 Simulated Data

To examine the performance of the proposed diagnostics measures, we considered simulated datasets with one or more of the generated cases perturbed. The covariate x_{i1}, $i = 1, \ldots, n$, was generated from a $N(30, 4)$ distribution and standardized for numerical stability. An additional covariate, x_{i2}, was independently generated from a *Bernoulli*(0.5) distribution. The failure time T_i was generated from an exponential distribution with hazard rate λ_i, where $\lambda_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$ with $\beta_0 = 1$, $\beta_1 = -0.5$, and $\beta_2 = 2$, and the censoring time C_i was generated from an exponential distribution with $\lambda_c = 2.56$, where T_i and C_i were assumed independent. The survival times y_i, $i = 1, \ldots, 150$, were taken as $y_i = \min(T_i, C_i)$, δ_i was the censoring indicator equal to 1, if $T_i \leq C_i$, and 0, if $T_i > C_i$. In the simulated data, y_i ranged from 0.000008 to 0.8269 with median = 0.0553, mean = 0.1045, and standard deviation = 0.1321, whereas λ_i ranged from 1.11 to 58.79 with median = 5.97, mean = 12.89, and standard deviation = 13.28. The observed censoring rate was 32%.

We selected cases 10, 59, and 62 for perturbation. To create influential observations in the dataset, we choose one or two of those selected cases and perturbed the survival time (y_i), the covariate (x_{i1}), or both the survival time and the covariate of the chosen case(s). Detailed descriptions regarding the perturbations are given in Table 1 and Web Appendix A. In Table

1, dataset (a) denotes the original simulated dataset with no perturbation and datasets (b)–(o) denote datasets with perturbed case(s) added by the perturbation schemes (I)–(VI).

We fit the gamma process model of Section 3 with an exponential $H^*(y) = 2.7y$. We chose a noninformative prior distribution for $\boldsymbol{\beta}$ as $N_2(\mathbf{0}, 10^6\mathbf{I})$. We used ARMS within Gibbs to obtain posterior samples. After burn-in, 40,000 MCMC posterior samples were used in the analysis. The proposed joint and marginal K–L divergences, $K(P, P_{-i})$ in (10), $K(P_1, P_{1,-i})$, $K(P_2, P_{2,-i})$ in (16), and calibrations of those divergences were computed for the simulated data with and without perturbation of the cases. We used p_i in Section 2.1 to compute the calibrations of $K(P, P_{-i})$, $K(P_1, P_{1,-i})$, and $K(P_2, P_{2,-i})$. We monitored convergence of the Gibbs chain using the method proposed by Geweke (1992), as well as trace plots. We conducted sensitivity analyses using $c = 0.01$, 0.1, 1, 10, and 100. For brevity, we present results for only the low confidence value of $c = 0.01$. For the computation of $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$, we used every fifth sample from the 40,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results.

Table 1 shows that the posterior inferences are sensitive to the perturbation of the selected case (s). Overall, the inferences are most sensitive to the perturbation of both the survival time and the covariate. Because we used noninformative priors on $\boldsymbol{\beta}$ and $c = 0.01$, the posterior estimates were similar to the maximum likelihood estimates based on partial likelihood. The results regarding the diagnostics showed that $K(P, P_{-i})$, as well as $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$, changed very little for the nonperturbed cases, whereas they changed a lot for the perturbed case(s).

The results in Table 2 show that before perturbation (dataset [a]), all of the selected cases are not influential, each providing a small $K(P, P_{-i})$ with its calibration close to 0.5. However, after perturbation (datasets [b–o]), $K(P, P_{-i})$ for those perturbed cases increased a lot and the corresponding calibrations become much larger than 0.5, indicating those cases are influential. Specifically, perturbing both the survival time and the covariate of a case increases $K(P, P_{-i})$ a lot. For example, $K(P, P_{-i})$ (and its calibration) for case 10 in dataset (h) is increased from 0.0006 (0.5168) to 5.8040 (1). We also note that the perturbed cases are similarly identified as influential using the LD based on partial likelihood. Moreover, Figure 1 clearly shows that $K(P, P_{-i})$ performed well for identifying influential case(s) in each dataset providing larger $K(P, P_{-i})$ for the perturbed case(s) compared to the other cases.

Moreover, in Table 1, we observe that perturbing the survival time of a case had influence on the posterior estimates of both β_1 and β_2, whereas perturbing the covariate (x_1) of a case alone had more influence on the estimates of β_1, corresponding to the perturbed covariate. We see that $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$ in Table 2 describe these marginal influences well. Specifically, both $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$ increase for the perturbation of the survival time, whereas $K(P_1, P_{1,-i})$ increases relative to $K(P_2, P_{2,-i})$ for the perturbation of the covariate (x_1). For example, perturbing the survival time of case 62 in dataset (d) increases $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$ from 0.0107 to 1.3214, and 0.0036 to 1.7394, respectively, whereas perturbing the covariate (x_1) of case 62 in dataset (g) increases $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$ from 0.0107 to 2.4073, and 0.0036 to 0.0618, respectively.

Although there may be masking effects when there is more than one perturbed case (datasets [k–o]), $K(P, P_{-i})$ identifies the influential cases by providing a larger $K(P, P_{-i})$ and its calibration compared to the other cases. In addition, $K(P_1, P_{1,-i})$ and $K(P_2, P_{2,-i})$ also describe the influence of the cases on posterior inference regarding β_1 and β_2, respectively. However, the magnitude of the measures become much smaller and the existence of an extremely influential case may mask the influence of other cases. This is not surprising because the proposed diagnostics are based on single case deletion.

Overall, the proposed joint and marginal influence diagnostic measures, $K(P, P_{-i})$, $K(P_1, P_{1,-i})$, and $K(P_2, P_{2,-i})$ performed well for identifying influential cases as well as describing the influence of a case on posterior inference.

4.2 Stanford Heart Transplant Data

To further illustrate the proposed methodology, we revisit the Stanford heart transplant data discussed in Section 1. Escobar and Meeker (1992) used 184 transplant cases to identify influential cases using an accelerated failure time lognormal regression model. We used the same dataset here with some minor modifications and identified influential cases using the proposed methodology. Of the 184 cases, 71 cases were right censored. The covariate included in this analysis was age (x_1) (mean = 41.09 and standard deviation = 11.036) as well as a quadratic term of age (x_2). Similar to Miller and Halpern (1982) and Escobar and Meeker (1992), the T5 mismatch score covariate was not used in this analysis due to its non-significance. For numerical stability in MCMC sampling, we standardized age and divided survival time by 365 to make time in years instead of days.

We fit the gamma process model of Section 3 with $H^*(y) = 0.35y$, $c = 0.01$ and $c = 100$. We chose a noninformative prior distribution for $\boldsymbol{\beta} = (\beta_1, \beta_2)$ as $N_2(\mathbf{0}, 10^6\mathbf{I})$. MCMC computations were done similarly as described in Section 4.1, and 14,000 MCMC posterior samples were used in this analysis after burn-in. The posterior means (standard deviations) and 95% highest posterior density (HPD) intervals for $\boldsymbol{\beta}$ were: For $c = 0.01$, 0.4588 (0.1134), and (0.2404, 0.6830) for β_1, and 0.2323 (0.0841) and (0.0650, 0.3949) for β_2; For $c = 100$, they were 0.3793 (0.1068) and (0.1746, 0.5916) for β_1, and 0.1117 (0.0766) and (−0.0385, 0.2606) for β_2.

Table 3 shows subjects having large $K(P, P_{-i})$ and calibration values compared to the other subjects in the dataset. For both small and large c, case 74 was identified as the most influential, having $K(P, P_{-i})$ (calibration) = 0.1539 (0.7573) for $c = 0.01$ and $K(P, P_{-i})$ (calibration) = 0.1818 (0.7761) for $c = 100$. Cases 159, 119, and 139 were also identified as influential. In addition, we identified cases 160, 108, and 133 as somewhat influential compared to other cases in the dataset for both small and large c. Figure 2 shows a plot of $K(P, P_{-i})$ for all the cases using $c = 0.01$. Upon examination of these cases, it appears that these cases are influential due to low values of the covariate age, and because there were not many low age cases. Specifically, cases 159, 139, 160, 108, and 133 had small failure times in spite of their low age values. An analysis using the LD based on partial likelihood showed that cases 74, 159, 119, and 139 were also identified as influential. In addition, our analysis identified similar cases as being influential as in Escobar and Meeker (1992), in which they identified influential cases using either case weight perturbations (patient number: 21, 74, 119, 133, 159, 160) or response perturbations (patient number: 18, 21, 133, 139, 159) based on an accelerated failure time lognormal regression model. Although a different model than ours was being fit, we used the results in Escobar and Meeker (1992) as a benchmark for the proposed Bayesian methodology to examine whether the proposed Bayesian methodology was at least consistent and yielding results in the same direction as commonly used frequentist methodology. We note that we used patient number as case number whereas Escobar and Meeker (1992) used case number sorted by age.

4.3 Melanoma Data

As a further demonstration of the proposed methodology, we considered a phase III clinical trial, labeled E1690 (Kirkwood et al., 2000). The trial evaluated the efficacy of interferon alfa-2b therapy on melanoma patients. The dataset consisted of 427 patients. The response variable was relapse-free survival time in years. The covariates included in this analysis were age, treatment, sex, and performance status. For details, see Web Appendix B. We fit the gamma process model of Section 3 with $H^*(y) = 0.26y$ and $c = 0.01$. We chose a noninformative

prior distribution for $\boldsymbol{\beta}$ as $N_4(\mathbf{0}, 10^6\mathbf{I})$. MCMC computations were done similarly as described in Section 4.1.

For the E1690 data, we did not find any highly influential cases. The $K(P, P_{-i})$ was smaller than 0.034 for all cases and the corresponding calibrations were not much larger than 0.5 (Web Figure 1). However, cases 16784, 16074, 16179, 16109, 16221, and 16504 had larger $K(P, P_{-i})$ compared to the other cases (Web Table 1). Specifically, case 16784 ($K(P, P_{-i}) = 0.0338$, calibration = 0.6279) and case 16074 ($K(P, P_{-i}) = 0.0303$, calibration = 0.6213) were identified as the most and the second-most influential cases compared to the other cases. After an investigation as to the reason why these identified cases were more influential than others, we found that the identified cases had longer relapse-free survival time (although they were censored) in spite of their large ages compared to other cases having moderate performance status. The marginal influence for the individual β_j's showed that the identified observations were more influential on posterior inference of β_4, which corresponds to the performance status covariate, compared to the other covariates (Web Table 1).

5. Discussion

We have proposed Bayesian case influence diagnostics using the Kullback–Leibler divergence for survival models. We have provided simple computational formulas for computing case influence on both the joint and marginal posterior distributions using MCMC techniques. We have only considered diagnostics based on single case deletion. This can be easily expanded to deletion of more than a single case or subsets of cases. In principle, this methodology can also be applied to any regression model by specifying the ratio of likelihoods with full data and data with a single case (or subset of cases) deleted. We have presented the full development for survival models here for focus and clarity of exposition.

The issue of what to do in a statistical analysis once an influential observation has been detected is a huge issue with no easy answer. Most researchers in this area recommend that (i) analyses with and without the influential case should be clearly reported, indicating differences in point and interval estimates, as well as variance estimates, and (ii) if one seeks remedies to the problem, three strategies are typically mentioned: one can transform the data, reparameterize the model, or fit a new model all together. Remedies for influential observations is a very large research area on its own.

Supplemental Materials

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank the editor, associate editor, and referee for very helpful comments and suggestions, which have led to an improvement of this article. This research was partially supported by NIH grants GM 70335, CA 74015, NSF grant SES-0643663, and NCI grant R01CA69222.

References

Carlin BP, Polson NG. An expected utility approach to influence diagnostics. Journal of the American Statistical Association 1991;86:1013–1021.

Christensen, R. Log-linear Models and Logistic Regression. New York: Springer-Verlag; 1997.

Cook DR. Assessment of local influence. Journal of the Royal Statistical Society, Series B: Methodological 1986;48:133–169.

Cook, RD.; Weisberg, S. Residuals and Influence in Regression. New York: Chapman & Hall; 1982.

Cox DR. Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, Series B: Methodological 1972;34:187–220.

Escobar LA, Meeker WQ. Assessing influence in regression analysis with censored data. Biometrics 1992;48:507–528. [PubMed: 1637975]

Geisser, S. Predictive Inference: An Introduction. London: Chapman & Hall; 1993.

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: Oxford University Press; 1992. p. 147-159.

Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Disc: P189–193). In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. Bayesian Statistics. Vol. 4. Oxford: Oxford University Press; 1992. p. 169-188.

Gilks WR, Best NG, Tan KKC. Adaptive rejection Metropolis sampling within Gibbs sampling. Applied Statistics 1995;44:455–472.

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

Johnson W, Geisser S. Estimative influence measures for the multivariate general linear model. Journal of Statistical Planning and Inference 1985;11:33–56.

Kalbfleisch JD. Nonparametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society, Series B: Methodological 1978;40:214–221.

Kirkwood J, Ibrahim J, Sondak V, Richards J, Flaherty L, Ernstoff M, Smith T, Rao U, Steele M, Blum R. High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial e1690/s9111/c9190. Journal of Clinical Oncology 2000;18:2444–2458. [PubMed: 10856105]

McCulloch RE. Local model influence. Journal of the American Statistical Association 1989;84:473–478.

Miller R, Halpern J. Regression with censored data. Biometrika 1982;69:521–531.

Peng F, Dey DK. Bayesian analysis of outlier problems using divergence measures. The Canadian Journal of Statistics 1995;23:199–213.

Pettitt AN, Daud IB. Case-weighted measures of influence for proportional hazards regression. Applied Statistics 1989;38:51–67.

Pettit LI. Diagnostics in Bayesian model choice. The Statistician: Journal of the Institute of Statisticians 1986;35:183–190.

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

Weiss R. An approach to Bayesian sensitivity analysis. Journal of the Royal Statistical Society, Series B: Methodological 1996;58:739–750.

Weiss RE, Cho M. Bayesian marginal influence assessment. Journal of Statistical Planning and Inference 1998;71:163–177.

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

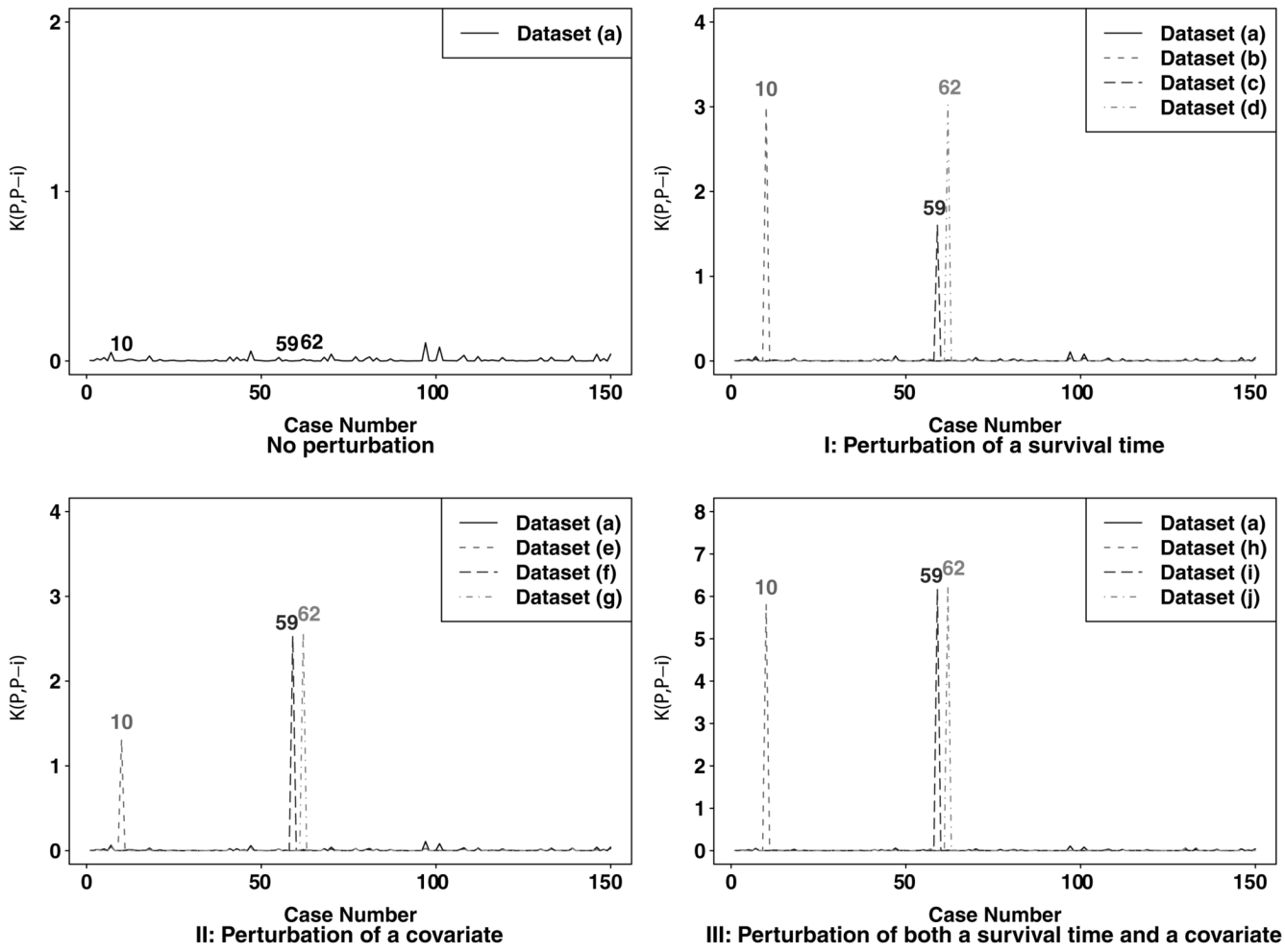

Figure 1.
$K(P, P_{-i})$ for the simulated data with $c = 0.01$.

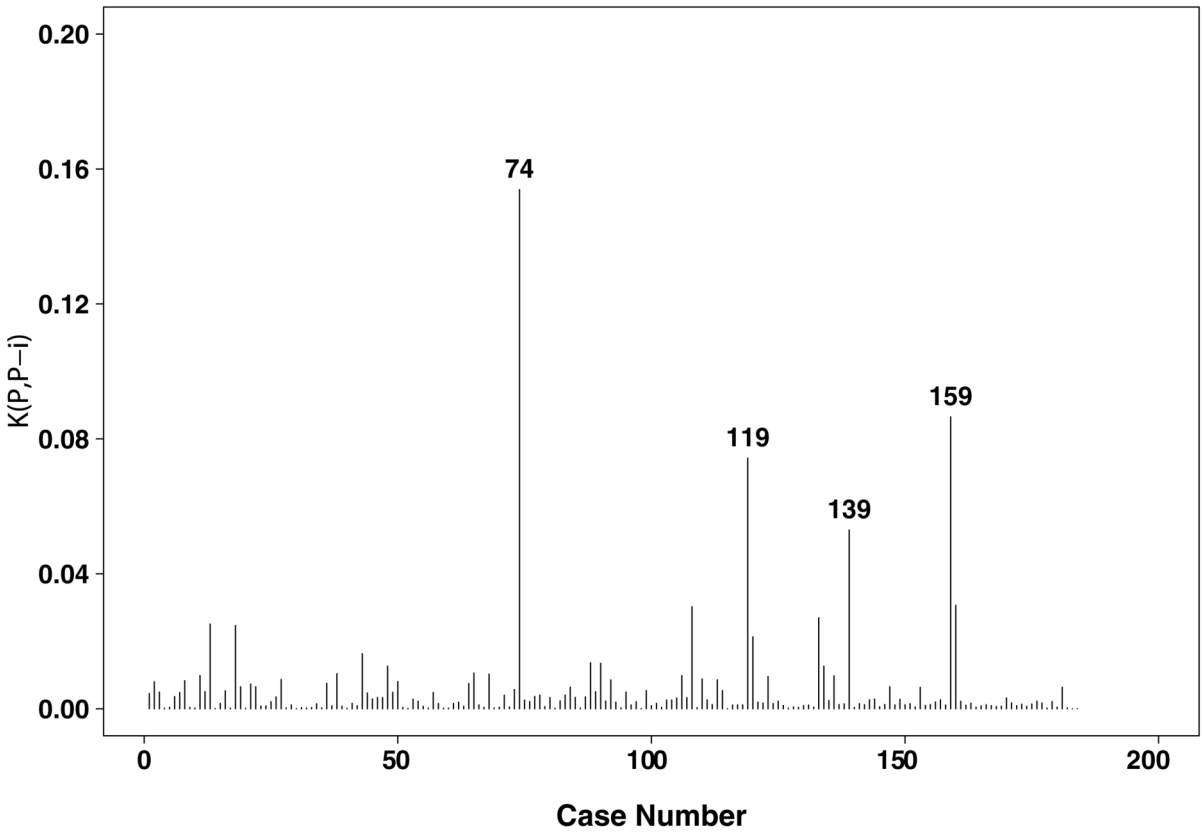

Figure 2.
$K(P, P_{-i})$ for the heart transplant data with $c = 0.01$.

Table 1

Posterior means and standard deviations for the simulated data with c = 0.01

					Maximum likelihood estimates				Posterior estimates			
					β_1		β_2		β_1		β_2	
Number of perturbed case	**Perturbation scheme**	**Dataset names**	**Description of perturbation**	**Perturbed case number**	**MLE**	**SE**	**MLE**	**SE**	**Mean**	**SD**	**Mean**	**SD**
None	No perturbation	a	Original data	None	−0.5292	0.1082	1.9945	0.2545	−0.5411	0.1077	2.0348	0.2567
One case	I	b	Survival time	10	−0.3355	0.1019	1.5003	0.2191	−0.3424	0.1020	1.5256	0.2197
		c	Survival time	59	−0.4412	0.1091	1.5765	0.2219	−0.4483	0.1090	1.6038	0.2229
		d	Survival time	62	−0.3563	0.1036	1.5126	0.2191	−0.3619	0.1037	1.5380	0.2194
	II	e	Covariate	10	−0.3480	0.0703	1.8916	0.2490	−0.3482	0.0703	1.9275	0.2508
		f	Covariate	59	−0.2630	0.0654	1.8329	0.2472	−0.2621	0.0655	1.8663	0.2483
		g	Covariate	62	−0.2413	0.0610	1.8280	0.2468	−0.2390	0.0613	1.8620	0.2475
	III	h	Survival time and covariate	10	−0.0086	0.0566	1.4968	0.2259	−0.0061	0.0564	1.5199	0.2274
		i	Survival time and covariate	59	−0.0279	0.0609	1.4819	0.2247	−0.0251	0.0607	1.5055	0.2263
		j	Survival time and covariate	62	−0.0099	0.0574	1.4943	0.2255	−0.0071	0.0571	1.5198	0.2264
Two cases	IV	k	Survival time	10	−0.3196	0.1046	1.3616	0.2155	−0.3257	0.1051	1.3870	0.2162
			Survival time	59								
		l	Survival time	10	−0.2677	0.1021	1.3416	0.2154	−0.2727	0.1026	1.3646	0.2149
			Survival time	62								
	V	m	Covariate	10	−0.2401	0.0601	1.8187	0.2471	−0.2385	0.0606	1.8529	0.2479
			Covariate	59								
		n	Covariate	59	−0.1911	0.0568	1.7958	0.2463	−0.1899	0.0572	1.8290	0.2479
			Covariate	62								
	VI	o	Survival time	10	−0.1896	0.0676	1.4792	0.2209	−0.1858	0.0678	1.5034	0.2206
			Covariate	62								

Table 2

Case influence diagnostics for the simulated data

Perturbation scheme	Dataset names	Case number	LD	Joint influence		Marginal influence			
				$K(P, P_{-i})$	Cal.	$K(P_1, P_{1,-i})$	Cal.	$K(P_2, P_{2,-i})$	Cal.
No perturbation	a	10	0.0009	0.0006	0.5168	0.0014	0.5266	0.0008	0.5205
		59	0.0001	0.0001	0.5067	0.0002	0.5090	0.0003	0.5118
		62	0.0204	0.0109	0.5736	0.0107	0.5727	0.0036	0.5421
I	b	10	1.9946	3.0065	0.9994	1.8175	0.9934	2.1358	0.9965
	c	59	1.1502	1.6045	0.9898	0.4879	0.8947	1.4537	0.9862
	d	62	1.9791	3.0257	0.9994	1.3214	0.9819	1.7394	0.9922
II	e	10	0.9141	1.3144	0.9816	1.4099	0.9849	0.0160	0.5901
	f	59	1.9342	2.5355	0.9984	2.5159	0.9984	0.0771	0.7042
	g	62	2.0644	2.5896	0.9986	2.4073	0.9980	0.0618	0.6814
III	h	10	3.6459	5.8040	1.0000	5.8891	1.0000	0.5759	0.9135
	i	59	3.8906	6.1829	1.0000	5.1433	1.0000	0.6485	0.9262
	j	62	3.8218	6.2595	1.0000	5.4944	1.0000	0.3322	0.8484
IV	k	10	1.1233	0.9574	0.9617	0.5498	0.9084	0.4243	0.8782
		59	0.3366	0.2051	0.7900	0.0066	0.5572	0.1890	0.7805
	l	10	0.8173	0.5998	0.9179	0.2924	0.8327	0.2824	0.8284
		62	0.7663	0.5439	0.9071	0.2448	0.8111	0.2681	0.8221
V	m	10	0.0744	0.0627	0.6717	0.0630	0.6720	0.0041	0.5452
		59	1.1271	1.0910	0.9710	1.1114	0.9722	0.0249	0.6102
	n	59	0.4191	0.2971	0.8347	0.2833	0.8288	0.0010	0.5219
		62	0.7247	0.6499	0.9264	0.6118	0.9201	0.0309	0.6262
VI	o	10	1.0668	1.2878	0.9806	0.3136	0.8413	0.9959	0.9646
		62	0.8762	1.1853	0.9761	1.1941	0.9765	0.0685	0.6789

Note that Cal. represents calibration.

Table 3

Case influence diagnostics for the heart transplant data

	Case identification			c = 0.01		c = 100	
Patient no.	**Time (days)**	**Status**	**Age**	$K(P, P_{-i})$	**Cal.**	$K(P, P_{-i})$	**Cal.**
74	2006	Alive	15	0.1539	0.7573	0.1818	0.7761
159	10	Dead	13	0.0865	0.6993	0.0973	0.7102
119	1116	Alive	14	0.0743	0.6858	0.0628	0.6718
139	86	Dead	12	0.0530	0.6585	0.0871	0.6999
160	5	Dead	20	0.0307	0.6219	0.0337	0.6277
108	42	Dead	19	0.0303	0.6211	0.0359	0.6316
133	1	Dead	21	0.0270	0.6145	0.0289	0.6185

Note that Cal. represents calibration.