Cabras, S., Castellanos, M. E., & Perra, S. (2015). A new minimal training sample scheme for intrinsic Bayes factors in censored data. *Computational Statistics & Data Analysis*, 81, pp. 52-63.

DOI:10.1016/j.csda.2014.07.012

A new minimal training sample scheme for intrinsic Bayes factors in censored data

Stefano Cabras [a,b,*], Maria Eugenia Castellanos [c], Silvia Perra [d]
[a] Department of Statistics, Universidad Carlos III de Madrid, Spain
[b] Department of Mathematics and Informatics, Università di Cagliari, Italy
[c] Department of Informatics and Statistics, Universidad Rey Juan Carlos, Spain
[d] Department of Social Science and Institutions, Università di Cagliari, Italy

A B S T R A C T

The problem of covariate selection for regression models with right censored data is con-sidered. It is approached from a default Bayesian point of view with Bayes factors (BFs) and in particular with Intrinsic BF (IBF) that depends on the minimal training samples (MTSs). In the presence of censored data, the number of possible MTSs increases, due to the fact that uncensored data, relevant for training the improper prior into a proper posterior, must be combined with censored data. For this purpose, the sequential minimal training sam-ple scheme (SMTS) accounts for such requirements but generally leads to IBF correction factors that do not have an analytical form and thus require numerical approximation. In order to obtain an analytical expression of the correction terms, a different TS scheme is introduced based on the Kaplan–Meier (KM) estimator, termed the KM minimal training sample scheme. This new tool works extremely well in the analyzed simulation setting and also in the applications; it produces results which are similar, if not better, than the IBF calculated using MTSs. The resulting new IBF, being based on analytical expressions, is much faster to compute. Evidence of these results comes from a large simulation study, theoretical arguments, and an application to a real data set.

Keywords:
Improper priors
Intrinsic Bayes factor
Kaplan–Meier estimator
Model selection
Survival analysis

1. Introduction

The model selection problem for survival regression analysis is studied with particular attention to covariates selection where the response variable Y is censored. From a Bayesian model selection perspective (see e.g. Nott and Leng, 2010; Wagner and Duller, 2012; and Lee et al., 2014) the most common tools are the Bayes factors (BFs), which unfortunately are undefined when default improper priors are employed. The extensive literature on methods that deal with such a situation is mainly focused on the Intrinsic BF (IBF) (Berger and Pericchi, 1996) that, together with the Fractional BF (FBF) (O'Hagan, 1995), have been defined in order to avoid indeterminacy. Under model regularity conditions, also met in this paper, such approximated BFs converge asymptotically to actual BFs as it is possible to show that there is an intrinsic prior leading to the asymptotic IBF (or asymptotic fractional BF). These two BFs are defined upon a minimal training sample (MTS), drawn, in a suitable manner, from the available data. The MTS is used to train the improper prior into a proper posterior, thus making the indetermination of the BF disappear. In particular, both depend on the size of the MTS, while the IBF also depends on the particular MTS that has been drawn.

Throughout the paper the IBF is mainly used for censored data. The motivation for exploring model selection using this tool instead of other ones such as FBF, or other possibilities like those discussed, for instance, in Celeux et al. (2012) is that IBFs are posterior model consistent, even if the model dimension grows exactly like the sample size. Posterior model consistency here means that the posterior model probability tends to 1 for the "true" sampling model. The necessary and sufficient conditions for this are discussed in Moreno et al. (2014). Such conditions require the priors for parameters and for models that are the same ones used in this paper.

In survival analysis, because of the presence of censored and uncensored observations, the drawing mechanism of the MTS must take into account these different types of observations because each type leads to a different amount of information to estimate the unknown regression parameters. The problem of exploring the space of the MTSs, which includes censored and uncensored observations, has already been formalized in Assumption 0 of Berger and Pericchi (2004), which states that the hypothetical sampling space of possible MTSs must have probability 1 under each model. Satisfying Assumption 0 is a requirement that leads to the introduction of the sequential minimal training sample scheme (SMTS) (Berger and Pericchi, 2004) which consists in drawing samples sequentially, without replacement, until a specified number of uncensored observations has been obtained. This differs from the ordinary resampling without replacement scheme, denoted here by OMTS, used in calculating the IBF in Berger and Pericchi (1996). The OMTS assigns uniform weights to all observations that train the improper prior into a proper density. However, when dealing with censored data only uncensored observations would lead to a proper trained posterior. In this case the OMTS would provide only unweighted uncensored observations, thus not satisfying Assumption 0 and introducing a bias in the estimation of regression parameters and, hence, in the whole model selection procedure.

A different approach to defining MTSs in the presence of censored data is discussed. This new strategy is very useful when it is possible to obtain closed-form expressions for predictive distributions, which is something that occurs mainly when training samples do not contain censored data. In particular, this is true for some models such as the log-normal one, which is a reference model in survival analysis. Calculating the predictive distribution of Y with only uncensored data, i.e., calculating the likelihood involving only the density function of Y, would significantly reduce the computational efforts because the survival function is usually not available in a closed form expression, as in the case of the log-normal model.

As will be discussed later, the theoretical justification behind the KMMTS is that it assures that Assumption 0 is still asymptotically satisfied, and it generates samples that contain only uncensored observations. This simplifies the calculation of IBF and leads to the possibility of exploring a larger number of models with respect to the IBF defined upon the SMTS. Simplification of calculus consists in substituting an MCMC approach for BF approximation with an exact analytical approximation, and this partially explains, as shown later, the better performance of the IBF defined upon the KMMTS with respect to that upon SMTS.

The rest of the paper is organized as follows: Section 2 describes the IBF and the BIC in the presence of censored data. Section 3 contains definitions of different training sample schemes. Section 4 reconsiders the IBF for right censored data using the KMMTS to be used with the IBF. In Section 5 the proposed technique is applied to the log-normal regression model, providing the corresponding expressions for the IBF. Section 6 compares the BIC and the two versions of the IBFs using a simulation study. Section 7 illustrates an application to a well known data set in survival modeling literature. Some remarks and conclusions can be found in Section 8.

2. The regression model and the objective variable selection

It is useful to first recall the regression model considered although the arguments presented here are of general applicability to the problem of model selection. Later on, objective variable selection is introduced, under a perspective mainly focused on the IBF technique.

Let $(t_i, \delta_i, \boldsymbol{x}_i)$ be the survival time, censored indicator and covariates, respectively, for individual $i = 1, \ldots, n$, where $\delta_i = 0$ if right censored and $\delta_i = 1$ otherwise. Without loss of generality, consider a fixed design matrix $\boldsymbol{X}$ with $p + 1$ columns, including the intercept. Let $y_i = \log(t_i)$ be normally distributed according to the following regression model $\mathcal{M}_k$ with a set of covariates denoted by $\boldsymbol{x}_{k,i}$

$$\mathcal{M}_k : y_i = \boldsymbol{\beta}_k^{\top} \boldsymbol{x}_{k,i} + \sigma_k \epsilon_i, \tag{1}$$

where $k \in \{1, 2, \ldots, K = 2^p\}$ is the model index with the corresponding design matrix $\boldsymbol{X}_k = (\boldsymbol{x}_{k,1}, \ldots, \boldsymbol{x}_{k,n})^{\top} \in \mathbb{R}^{n \times p_k}$ and model parameter $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \sigma_k) \in \boldsymbol{\Theta}_k = \mathbb{R}^{p_k} \times \mathbb{R}^+$. Here ϵ_i is assumed to be normally distributed such that the observed times are log-normal. Note that other types of distributions for the time are possible by just fixing that of ϵ_i. For instance, the Weibull model is obtained assuming that ϵ_i is distributed according to a standard Gumbel, or if ϵ_i follows a generalized Gumbel distribution, the observed times would follow a generalized gamma distribution. The proposed approach also fits such models, but the computational advantage is limited, in that when calculating the BF, it is not necessary to integrate the survival function.

The most probable model $\mathcal{M}_k$, given the observed data through BFs and model posterior probabilities (see Kass and Raftery, 1995; Berger, 1999 and Berger and Pericchi, 2001), is selected. In order to calculate BFs a prior distribution $\pi_k(\boldsymbol{\theta}_k)$ needs to be specified separately for each model. This can be complicated because one often initially entertains K models leading to the impossibility of careful subjective prior elicitation. For this purpose, Bayesian model selection is usually done by means of default priors (Berger and Pericchi, 2001) which are often improper. In this context, and because of the specific type of the parametric model, the usual default prior for location–scale models is considered, that is $\pi_k^N(\boldsymbol{\theta}_k) = c_k/\sigma_k$, where c_k is the normalizing pseudo-constant (Yang and Berger, 1997).

Suppose that $\pi_k(\boldsymbol{\theta}_k)$, $k = 1, \dots, K$, is a prior distribution for the unknown parameters of model k, then the predictive density of the observed $\boldsymbol{y}$, under $\mathcal{M}_k$, is

$$m_k(\boldsymbol{y}) = \int_{\Theta_k} f_k(\boldsymbol{y} \mid \boldsymbol{\theta}_k)\pi_k(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k,$$

where f_k is the likelihood of $\boldsymbol{\theta}$ induced by (1). The BF of $\mathcal{M}_i$ against $\mathcal{M}_j$ is

$$B_{ij}(\boldsymbol{y}) = \frac{m_i(\boldsymbol{y})}{m_j(\boldsymbol{y})}, \tag{2}$$

and the posterior probability of $\mathcal{M}_i$ is

$$p(\mathcal{M}_i \mid \boldsymbol{y}) = \frac{m_i(\boldsymbol{y})p_i}{\sum_{j=1}^{K} m_j(\boldsymbol{y})p_j} = \left(1 + \sum_{j \neq i} \frac{p_j}{p_i} B_{ji}(\boldsymbol{y})\right)^{-1}, \tag{3}$$

where p_i are the model prior probabilities. A common choice in default Bayesian analysis is $p_i = 1/K$ and model posterior probabilities only depend on the BFs or equivalently on $\overline{m}_i(\boldsymbol{y}) = m_i(y)/\sum_{j=1}^{K} m_j(\boldsymbol{y})$. Alternative model priors are needed especially when K is large and made up of models with different complexity, a situation that will be discussed later.

Under the improper prior $\pi_k^N(\boldsymbol{\theta}_k)$, the BF in (2)

$$B_{ij}^N(\boldsymbol{y}) = \frac{m_i^N(\boldsymbol{y})}{m_j^N(\boldsymbol{y})} = \frac{c_i}{c_j} \frac{\int f_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)\pi_i^N(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i}{\int f_j(\boldsymbol{y} \mid \boldsymbol{\theta}_j)\pi_j^N(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j},$$

is unscaled because of the presence of the ratio between priors pseudo-constants c_i/c_j. In order to avoid this problem, several proposals can be adopted; in this article intrinsic BFs are considered.

Intrinsic Bayes factors

A solution to avoid the indetermination induced from an improper prior is the IBF where part of the data acts as a training sample (TS). The idea is to use the TS to train the improper prior into a proper posterior, and to use the rest of the data for model discrimination in order to compute the BF, using the updated posterior acting as a data dependent prior distribution (Berger and Pericchi, 1996).

Definition 1 (*Minimal Training Sample*)**.** A TS, indexed by l, is a subset of the data $\boldsymbol{y}$, denoted by $\widetilde{\boldsymbol{y}}(l)$. It is called *proper* if $0 < m_k^N(\widetilde{\boldsymbol{y}}(l)) < \infty$, for all $\mathcal{M}_k$, $k = 1, \dots, K$, and it is called *Minimal* if it is proper and no subset is proper.

Given a MTS, the improper prior $\pi_k^N(\boldsymbol{\theta}_k)$ is trained into a proper posterior as follows

$$\pi_k^N(\boldsymbol{\theta}_k \mid \widetilde{\boldsymbol{y}}(l)) = \frac{f_k(\widetilde{\boldsymbol{y}}(l) \mid \boldsymbol{\theta}_k)\pi_k^N(\boldsymbol{\theta}_k)}{m_k^N(\widetilde{\boldsymbol{y}}(l))}.$$

Using this proper posterior as a prior, the BF is then computed over the rest of the sample. The resulting BF conditioned on $\widetilde{\boldsymbol{y}}(l)$ is

$$B_{ij}(\widetilde{\boldsymbol{y}}(l)) = B_{ij}^N(\boldsymbol{y})B_{ji}^N(\widetilde{\boldsymbol{y}}(l)), \tag{4}$$

where $B_{ji}^N(\widetilde{\boldsymbol{y}}(l)) = m_j^N(\widetilde{\boldsymbol{y}}(l))/m_i^N(\widetilde{\boldsymbol{y}}(l))$. Observe that the conditional BF given in (4) depends on the specific TS $\widetilde{\boldsymbol{y}}(l)$ and thus several proposals have been elaborated to combine a set of L MTSs. In particular, Berger and Pericchi (1996) proposed to average the correction factor $B_{ij}(\widetilde{\boldsymbol{y}}(l))$ over all possible MTSs, leading to the *Arithmetic IBF* (*AIBF*),

$$BF_{ij}^{AI} = B_{ij}^N(\boldsymbol{y})\frac{1}{L}\sum_{l=1}^{L} B_{ji}^N(\widetilde{\boldsymbol{y}}(l)), \tag{5}$$

while another possibility is to consider the median of all the L possible $B_{ij}(\widetilde{\boldsymbol{y}}(l))$ obtaining the *Median IBF* (*MIBF*),

$$BF_{ij}^{MI} = B_{ij}^N(\boldsymbol{y}) \underset{l=1,\dots,L}{\text{Median}}\{B_{ji}^N(\widetilde{\boldsymbol{y}}(l))\}, \tag{6}$$

which leads to a more stable IBF with respect to the MTSs. Further details about *AIBF* and *MIBF* can be found in Berger and Pericchi (1996); Berger and Pericchi (1998, 2001); Moreno et al. (1998, 1999) and, specifically for survival models, in Lingham and Sivaganesan (1999) and Kim and Sun (2000).

For more than two models, possibly non-nested, intrinsic BFs may be incoherent in the sense that $B_{jk} \neq B_{ji}B_{ik}$. To avoid this problem, the *encompassing* approach (Zellner and Siow, 1980; Berger and Pericchi, 2001) is adopted, where each sub-model $\mathcal{M}_i$ is compared with a reference model $\mathcal{M}_0$, e.g., the null model or the one with the intercept only. There are different forms of model encompassing the pairwise comparison from below (Moreno and Girón, 2008) is considered. It consists in comparing model $\mathcal{M}_k$ with the simplest model $\mathcal{M}_0$ always nested in $\mathcal{M}_k$, $k = 1, \ldots, K$. In fact, Moreno and Girón (2008) compared different encompassing methods, in the case of linear regression, and found out that the encompassing from below method has more appealing theoretical properties. For our purposes, the most important feature is that encompassing from below provides coherent model posterior probabilities in the space of all models (Moreno and Girón, 2008). Using the from below procedure, it is possible to obtain pairwise BFs, B_{i1}, $i = 1, \ldots, K$ such that the *BF* of $\mathcal{M}_i$ to $\mathcal{M}_j$ is

$$B_{ij} = \frac{B_{i1}}{B_{j1}}, \quad i, j = 1, \ldots, K, \; i \neq j. \tag{7}$$

Finally, for the sake of comparison, the Bayesian Information Criterion (*BIC*) for censored data discussed in Volinsky and Raftery (2000) is considered. The BIC is based on the Laplace approximation of the posterior distribution and thus of $m(\boldsymbol{y})$. It is a very computationally cheap tool for model selection, although it relies on the effectiveness of the normal approximation of the posterior distribution.

Highest and median posterior probability model

Once model posterior probabilities have been obtained according to (3), with some choice of the above mentioned BFs, then models are ranked according to their posterior probability and the one that maximizes one of the following two criteria is chosen: (i) the highest posterior probability model (HPPM) or (ii) the median posterior probability model (MPPM) discussed in Barbieri and Berger (2004).

The HPPM is simply the k^*-th model with the highest posterior probability, $\mathcal{M}_{k^*}$, $k^* = \arg\max_{k \in 1,\ldots,K} \; p(\mathcal{M}_k \mid \boldsymbol{y})$, which may still have very low probability.

The MPPM consists in choosing the model containing those regressors which have an overall posterior inclusion probability of being selected greater than 1/2. Let $\boldsymbol{l}_j$ be the set of indexes of all models containing a given variable *j*. The *posterior inclusion probability* for variable *j* is

$$p_j = \sum_{k \in \boldsymbol{l}_j} \Pr(\mathcal{M}_k \mid \boldsymbol{y}), \tag{8}$$

which is the overall posterior probability that the variable *j* is included in one model. If it exists, the MPPM $\mathcal{M}_{\boldsymbol{k}^*}$ is the model consisting of those variables whose posterior inclusion probability is at least 1/2, that is, $\boldsymbol{k}^* = (k_2^*, \ldots, k_{p+1}^*)$, where each k_j^* is

$$k_j^* = \begin{cases} 1 & \text{if } p_j \geq 1/2 \\ 0 & \text{otherwise,} \end{cases}$$

for all $j = 2, \ldots, p + 1$ covariates in $\boldsymbol{X}$ excluding the intercept ($j = 1$). Sometimes it may happen that the MPPM does not exist in the sense that the most probable model is the reference one that only has the intercept. More details can be found in Barbieri and Berger (2004).

3. Training sample schemes

In situations in which the observations in the training sample are differently informative for model parameters, as in the case of censored data, Berger and Pericchi (2004) proposed a general type of training sample scheme. The following definition introduces this concept.

Definition 2 (*Weighted Training Sample*)**.** A weighted TS scheme consists in drawing TS with weights $\boldsymbol{u} = (u_1, \ldots, u_L)$, where $\boldsymbol{u}$ is a probability vector, over the space of all possible TS, $\mathcal{X}^U$ made up of *L* elements.

For example, the OMTS of size *s* drawn from a sample of size *n* has $u_1 = \ldots, u_n = \binom{n}{s}^{-1}$.

As previously noted in the Introduction section, in survival analysis, censored and uncensored observations do not contain the same information. Therefore, the MTS scheme must take this fact into account. In order to gain more insight into this important issue, which is also the motivation of this paper, it is useful to start with a simple example, also used in Berger and Pericchi (2004).

Example 1 (*Right Censored Exponential*)**.** Suppose data $y_1, \ldots, y_n$ are a random sample from the right censored exponential distribution, with censoring time ρ. Thus, if $y_i < \rho$, then the density is $f(y_i \mid \theta) = \theta \exp(-\theta y_i)$, while if the observation is censored, the density is $\Pr(Y_i = \rho \mid \theta) = \exp(-\rho\theta)$.

Consider testing these two hypotheses,

$$M_0 : \theta = \theta_0 \quad \text{vs.} \quad M_1 : \theta \neq \theta_0.$$

Using the usual default prior for the exponential model, $\pi^N(\theta) = 1/\theta$, it can be seen that one single uncensored observation is sufficient to obtain a proper posterior, while no censored observation can achieve this. So the imaginary set of minimal training samples consists of single uncensored observations. As censored observations do not enter in the training sample, the OMTS scheme, consisting in drawing uniformly from all uncensored observations, will result in biased estimators in favor of larger values of $\theta = 1/E(Y_i|\theta)$. In fact, the sampling space of hypothetical training samples is the interval $(0, \rho)$, denoted by $\mathcal{X}^{MT}$. It is clear that Assumption 0 is violated when considering a TS consisting of only one uncensored observation because

$$\Pr^{M_i}_{\boldsymbol{\theta}_i}(\mathcal{X}^{MT}) = \Pr^{M_i}_{\boldsymbol{\theta}_i}(Y < \rho) = 1 - \exp(-\rho\theta_i) < 1, \quad i = 0, 1.$$

In order to solve this problem, Berger and Pericchi (2004) proposed the SMTS scheme, gathered in the following Definition 3.

Definition 3 (*Sequential Minimal Training Sample*)**.** Suppose having s parameters, then the SMTS is constructed by drawing observations, without replacement from $\boldsymbol{y}$, stopping when s uncensored observations are obtained. The SMTS induces a TS of the form

$$\boldsymbol{y}(l) = \{\{s-1 \text{ uncensored and } N_t - s \text{ censored observations}\}, y_s(l)\},$$

with random size $N_t \geq s$.

Note that $\boldsymbol{y}(l)$ is not minimal in the sense of Definition 1 because it contains censored observations that can be removed. This scheme leads to a different set of weights which is induced by the SMTS scheme.

Example 1 (*Cont.*)**.** In the example of the right censored exponential data, it can be easily shown that the SMTS scheme, with $s = 1$, produces a hypothetical training sample space that verifies Assumption 0.

Our proposal consists of another MTS scheme denoted as Kaplan–Meier minimal training sample scheme (KMMTS) in which weights for uncensored observations are positive, but differ from the uniform weights $\boldsymbol{u} > 0$ and are 0 for the censored observations. Thus, excluding censored observations from the MTS, for the log-normal model, a closed form expression for the IBF is obtained. The following definition represents the main proposal throughout the paper.

Definition 4 (*Kaplan–Meier Minimal TS*)**.** A Kaplan–Meier minimal training sample (KMMTS) is an MTS obtained by sampling without replacement s observations from the observed data according to the following probability mass function

$$\begin{aligned}\hat{f}_{KM}(t) &= \hat{F}_{KM}(t_i) - \hat{F}_{KM}(t_{i-1}) \\ &= \begin{cases} \hat{F}_{KM}(t_1) & \text{if } t \leqslant t_1 \\ \hat{F}_{KM}(t_i) - \hat{F}_{KM}(t_{i-1}) & \text{if } t_{i-1} < t \leqslant t_i,\ i = 2, \dots, n-1 \\ 1 - \hat{F}_{KM}(t_{n-1}) & \text{if } t_{n-1} < t \leqslant t_n, \end{cases}\end{aligned} \tag{9}$$

where $\hat{F}_{KM}(t) = 1 - \hat{S}_{KM}(t)$ and $\hat{S}_{KM}(t)$ is the Kaplan–Meier (KM) estimator of the survival function. Details about the calculation of this function can be found in Section 4.2.

Example 1 (*Cont.*)**.** In the example of the right censored exponential model, where censored observations are all equal to ρ, let $y_1 < \cdots < y_{n_u}$ be the ordered uncensored observations, then $\boldsymbol{y} = (y_1, \dots, y_{nu}, \underbrace{\rho, \dots, \rho}_{n_c \text{ times}})$, where n_u and n_c are the number of uncensored and censored observations. The probability mass function $\hat{f}_{KM}(t)$ assigns mass probability of $1/n$ to each $y_1, \dots, y_{n_u-1}$, and $(n_c + 1)/n$ to the last uncensored observation, y_{n_u}. In this case, Assumption 0 is verified asymptotically, when $n \to \infty$, $Y_{n_u} = \max\{Y_1, \dots, Y_{n_u}\}$ converges in probability to ρ, and the mass function in this point is $(n_c + 1)/n$ that converges to $p_{cens} = \exp(-\rho\theta)$. Let $\mathcal{X}^{KMMT}$ be the sampling space of hypothetical training samples using the KMMTS scheme; asymptotically $\mathcal{X}^{KMMT}$ verifies Assumption 0:

$$\begin{aligned}\lim_{n\to\infty} \Pr^{M_i}_{\boldsymbol{\theta}_i}(\mathcal{X}^{KMMT}) &= \lim_{n\to\infty} \left(\Pr^{M_i}_{\boldsymbol{\theta}_i}(Y < \rho) + \Pr^{M_i}_{\boldsymbol{\theta}_i}(Y_{n_u})\right) \\ &= 1 - \exp(-\rho\theta_i) + \exp(-\rho\theta_i) = 1, \quad i = 0, 1.\end{aligned}$$

4. Objective variable selection for regression models with censored data

Assume that observed log-times Y are *right censored*. Although other types of censoring are possible, they only complicate the exposition and calculus.

As commented on in the previous section, in the case of censored data, the MTS should be defined in order to verify Assumption 0 in Berger and Pericchi (2004), avoiding bias in the estimation of model parameters. This can be accomplished by using the SMTS scheme or asymptotically the KMMTS scheme.

When using the SMTS, N_t, the size of the sample is random and therefore the probability distribution of N_t is useful in order to establish $L_\star$, the number of correction factors, $B_{ij}(\widetilde{\boldsymbol{y}}(l))$, which are actually used to approximate the *AIBF* or *MIBF* when L is large. The following proposition provides the distribution of N_t:

Theorem 1. *Let N_t be the size of the SMTS, then the probability distribution of N_t given n, s and the number of censored observations $n_c = n - \sum_{i=1}^{n} \delta_i$ is*

$$\Pr(N_t = n_t | n, n_c) = \frac{\binom{n_c}{n_t - s}\binom{n-n_c-1}{s-1}(n_t-1)!(n-n_c)}{D_{n,n_t}}, \quad s \le n_t \le n_c + s, \tag{10}$$

where $D_{n,n_t} = n!/(n-n_t)!$.

Proof. The proof is based on straightforward combinatorial calculus. For more details, see Perra et al. (2013).

Theorem 1 allows the probability of obtaining a TS of size $N_t = n_t$ to be calculated. The SMTS can be viewed as a special case of weighted TS, and this distribution can be used to calculate the weights $\boldsymbol{u}$ for each MTS. Given a MTS of size n_t, the weight u_i is just the probability (10) divided by n_t.

4.1. IBF using SMTS

The calculation of the *IBF* can be very computationally demanding, because it is necessary to approximate $2(L+1)$ integrals, where L should be the number of all possible SMTSs. This can be simply unfeasible even for small samples, as L is of order $\sum_{N_t=s}^{n_c+s}\binom{n_c}{N_t - s}\binom{n-n_c-1}{s-1}(N_t-1)!(n-n_c)$. A possible approximation consists in considering a subset of the set of size $L_\star$ of all the possible SMTSs. As mentioned in Section 2 of Berger and Pericchi (2004) and in Varshavsky (1995), it often suffices to randomly choose a subset of cardinality $n \times n_t$ of MTSs, all with fixed size n_t. In order to account for the randomness over the size, consider a subset of cardinality $L_\star$ of SMTSs in which $L_\star$ can be given by the mode or the median of N_t,

$$L_{mo} = n \times \text{mode}\{N_t\} \quad \text{or} \quad L_{me} = n \times [\text{median}\ \{N_t\}],$$

with $[\cdot]$ denoting the integer part and where median and mode are calculated under the probability distribution of N_t given in (10). Distribution (10) can be further employed in implementing a stratified SMTS so that the distribution of sample sizes follows (10). However, this strategy would require enumerating all possible SMTSs, L, which may be unfeasible for large sample sizes. Another way to introduce distribution (10) is to use it in reweighing the SMTSs according to their size. For purposes of comparison with the actual version of the SMTS, this latter strategy will not be pursued further in this paper.

For each SMTS, $\boldsymbol{y}(l)$,

$$B_{i1}^{N}(l) = B_{i1}^{N}(\boldsymbol{y})B_{1i}^{N}(\boldsymbol{y}(l)),$$

is obtained and then the intrinsic BFs are

$$B_{i1,L_\star}^{AI} = B_{i1}^{N}(\boldsymbol{y})\frac{1}{L_\star}\sum_{l=1}^{L_\star} B_{1i}^{N}(\boldsymbol{y}(l)) \quad \text{and} \tag{11}$$
$$B_{i1,L_\star}^{MI} = B_{i1}^{N}(\boldsymbol{y}) \underset{l=1,\ldots,L_\star}{\text{Median}}\ B_{1i}^{N}(\boldsymbol{y}(l)),$$

where $L_\star \in \{L_{mo}, L_{me}\}$. For each $i = 1$ and $j = 2, \ldots, K$, $B_{ij,L_\star}^{AI}$ and $B_{ij,L_\star}^{MI}$ are defined by the equation in (7).

4.2. IBF using KMMTS

In this subsection the new methodology to calculate *IBF* using KMMTS is introduced. First a summary of the Kaplan–Meier estimator and the probability mass function induced by this estimator is presented.

Kaplan–Meier estimator

Let t_i, O_i and d_i be the time, the number of individuals that are still alive at time t_i and the number of subjects that experiment the terminal event at t_i, respectively. The Kaplan–Meier (KM) estimator for the survival function, also known as the *product-limit estimator*, was introduced by Kaplan and Meier (1958) and is expressed as follows:

$$\hat{S}_{KM}(t) = \prod_{t_i \leqslant t}\left(\frac{O_i - d_i}{O_i}\right) = \prod_{t_i \leqslant t}\left(1 - \frac{d_i}{O_i}\right),$$

where $1 - d_i/O_i$ is the conditional probability that an individual survives at the end of a time interval, under the condition that the individual was present at the beginning of the time interval. The KM estimator $\hat{S}_{KM}(t)$ is not well defined for values of t beyond the largest observation; in fact if the largest study time corresponds to a death, then the estimated survival curve is zero after this time. If the largest study time is censored, the survival $\hat{S}_{KM}(t)$ beyond this point will be undetermined because the time to death is unknown. In order to avoid this problem, occurring mainly for very small sample sizes the convention proposed in Efron (1967) is adopted. It consists in fixing the value of the KM estimator, $\hat{S}_{KM}(t)$ to 0 beyond the largest study time. This means that the survivor with the longest time on study has died immediately after the survivor's censoring time. Other possible alternatives can be found in Gill (1980); Efron (1967); Brown et al. (1974); and Moeschberger and Klein (1985).

The KM estimator is based on an assumption of non-informative censoring, that is, the knowledge about a censoring time for an individual does not provide further information about his/her likelihood of survival. This means, for example, that censoring times do not depend on covariates. When this assumption is violated, $\hat{F}_{KM}$ estimates the wrong distribution function and there is a potential conflict between the model selection problem and that of estimating the survival function. When there are indications that censoring could depend on some covariates in the study, KM estimators conditional to these covariates can be considered instead of the marginal KM estimator adopted here. The rest of calculations shown here are not affected by the estimator used to construct the MTS.

Once an estimated survival function, $\hat{S}_{KM}(t)$ is obtained over the whole available sample, the estimation of the cumulative distribution function is obtained as $\hat{F}_{KM}(t) = 1 - \hat{S}_{KM}(t)$ and then the KMMTS scheme in Definition 4 is considered.

The KM estimator of the CDF results in a step function in which the mass is positively defined only at uncensored observations, while the mass function estimated via the $\hat{F}_{KM}$ in a censored observation is 0. In this case, the weights $\boldsymbol{u}$ in Definition 1 are induced by $\hat{f}_{KM}(t_i)$, for $i = 1, \ldots, n - n_c$ uncensored observations and they are 0 for all MTSs containing censored observations, see Eq. (9). This implies that the KMMTS only contains s uncensored observations.

Kaplan and Meier (1958) showed that the Kaplan–Meier estimator, $\hat{S}_{KM}(t)$, is the non-parametric maximum likelihood estimate of $S(t)$, it is asymptotically consistent and the bias of $\hat{S}_{KM}(t)$ declines exponentially with sample size, n. These characteristics assure that assumption 0 is satisfied with the KMMTS, because the sampling space of survival times $\mathcal{X}^U$ is covered when using such a sampling scheme.

Calculation of the IBF using KMMTS

The KMMTS counterparts of AIBF and MIBF are

$$\begin{aligned} BF_{ij}^{AI} &= B_{ij}^{N}(\boldsymbol{y}) \sum_{l=1}^{L'} u_l B_{ji}^{N}(\boldsymbol{y}(l)) \quad \text{and} \\ BF_{ij}^{MI} &= B_{ij}^{N}(\boldsymbol{y}) \ \underset{1,\ldots,L'}{\text{Median}} \ B_{ji}^{N}(\boldsymbol{y}(l)), \end{aligned} \tag{12}$$

where the weights $u_1, \ldots, u_L$ are that obtained with the KMMTS and the median is calculated according to them. The calculation of all possible training samples, jointly with their weights, may still be unviable, but now the size of the MTS is fixed at s and hence it is possible to use a subset of training samples of size $L' = n \times s$.

5. The case of the log-normal regression model

The log-normal model is a well known parametric model with density function $f(t) = 1/(\sqrt{2\pi}\sigma t)\exp\big(-1/(2\sigma^2)(\log(t) - \mu)^2\big)$, survival function $S(t) = 1 - \Phi\left((\log t - \mu)/\sigma\right)$ and hazard rate $h(t) = f(t)/S(t)$, where $\Phi(t)$ is the distribution function of a standard normal variable, $\mu \in \mathbb{R}$, $\sigma > 0$ and $t > 0$. The hazard rate of the log-normal model at 0 is zero; it increases to a maximum and then decreases to 0 as $t \to \infty$. Suppose that the time to the terminal event is log-normal distributed, then $Y_i = \log(T_i)$ follows a normal distribution. Therefore, it is then possible to express the regression model of $Y_i|\boldsymbol{x}_i$ as (1), where $\epsilon_i \sim N(0, 1)$ and suppressing the subscript k to clarify the notation. For right censored data the likelihood of $\boldsymbol{\theta}$ has the form

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}, \boldsymbol{X}) &= \prod_{i=1}^{n} f_Y(y_i)^{\delta_i} \left[S_Y(y_i)\right]^{(1-\delta_i)} \\ &= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y_i - \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{x}_i}{\sigma}\right)^2\right)\right]^{\delta_i} \left[1 - \Phi\left(\frac{y_i - \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{x}_i}{\sigma}\right)\right]^{(1-\delta_i)}, \end{aligned}$$

and using the usual default prior for location–scale models, $\pi(\boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma}$ for $\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma \in \mathbb{R}^+$, the posterior is

$$\pi(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}, \boldsymbol{X}) \propto \pi(\boldsymbol{\beta}, \sigma) L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}, \boldsymbol{X}).$$

Consider the calculation of the marginal distribution $m(\boldsymbol{y})$ under two situations:

(i) only uncensored observations in which $m(\boldsymbol{y})$ has a closed form expression;
(ii) censored and uncensored observations in which $m(\boldsymbol{y})$ must be obtained with numerical approximations such as a Markov Chain Monte Carlo (MCMC).

The marginal distribution $m(\boldsymbol{y})$ in case (i), using well known results (see e.g. Chapter 8 of Ghosh et al., 2006 or Section 2 in Berger and Pericchi, 1997), is

$$m(\boldsymbol{y}) = \frac{\Gamma\left(\frac{1}{2}\right)^r \Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \mid \boldsymbol{X}^\top \boldsymbol{X} \mid^{1/2} \left((\boldsymbol{y}-\hat{\boldsymbol{y}})^\top(\boldsymbol{y}-\hat{\boldsymbol{y}})\right)^{(n-r)/2}}, \tag{13}$$

where r is the rank of matrix $\boldsymbol{X}$, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ and $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$, while for (ii) the kernel of the posterior distribution can be written as

$$\pi(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}, \boldsymbol{X}) \propto \frac{1}{\sigma}\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}\left(\frac{y_i-\boldsymbol{\beta}^\top\boldsymbol{x}_i}{\sigma}\right)^2\right)\right]^{\delta_i}\left[1-\Phi\left(\frac{y_i-\boldsymbol{\beta}^\top\boldsymbol{x}_i}{\sigma}\right)\right]^{(1-\delta_i)}.$$

In this case, the posterior distribution is approximated using a Random Walk Metropolis–Hastings where the proposal is a standard multivariate normal distribution (Albert, 2009) and the algorithm of Chib and Jeliazkov (2001, Subsection 2.1) is used to approximate the BFs.

5.1. IBF calculation

Suppose that to compare the following two log-normal models:

$$M_0 : Y_i = {\boldsymbol{\beta}_0}^\top \boldsymbol{X}_0 + \sigma\epsilon_i \quad \text{and} \quad M_i : Y_i = {\boldsymbol{\beta}_i}^\top \boldsymbol{X}_i + \sigma\epsilon_i.$$

The corresponding KM AIBF (KMAIBF) and KM MIBF (KMMIBF) can be obtained from (12). In particular, the calculation of the correction factors $B_{0i}^N(\boldsymbol{y}(l))$ over the KMMTS, which do not contain censored data, can be obtained in closed-form, while $B_{i0}^N(\boldsymbol{y})$, as calculated over $\boldsymbol{y}$ with uncensored observations, must be approximated using MCMC. Summing up, the computational effort, when using KMAIBF or KMMIBF, is much less than ordinal AIBF or MIBF because the correction factors have closed form expressions and thus the overall computational effort is mainly due to the MCMC necessary to approximate $B_{i0}^N(\boldsymbol{y})$.

The closed form expression for the correction factors can be obtained as follows: let r_0 and r_i be the ranks of the design matrices $\boldsymbol{X}_0$ and $\boldsymbol{X}_i$, respectively, denote by $\boldsymbol{X}_0(l)$ and $\boldsymbol{X}_i(l)$ the matrices of covariates obtained by taking the rows corresponding to a KMMTS for models $\mathcal{M}_0$ and $\mathcal{M}_i$, respectively, then using (13) the correction factor is

$$B_{0i}^N(\boldsymbol{y}(l)) = \frac{\Gamma\left(\frac{1}{2}\right)^{r_0}\Gamma\left(\frac{s-r_0}{2}\right) \mid \boldsymbol{X}_i(l)^\top\boldsymbol{X}_i(l)\mid^{1/2}\left[(\boldsymbol{y}(l)-\hat{\boldsymbol{y}}_i(l))^\top(\boldsymbol{y}(l)-\hat{\boldsymbol{y}}_i(l))\right]^{(s-r_i)/2}}{\Gamma\left(\frac{1}{2}\right)^{r_i}\Gamma\left(\frac{s-r_i}{2}\right) \mid \boldsymbol{X}_0(l)^\top\boldsymbol{X}_0(l)\mid^{1/2}\left[(\boldsymbol{y}(l)-\hat{\boldsymbol{y}}_0(l))^\top(\boldsymbol{y}(l)-\hat{\boldsymbol{y}}_0(l))\right]^{(s-r_0)/2}},$$

where $\hat{\boldsymbol{y}}_i(l) = \boldsymbol{X}_i(l)\hat{\boldsymbol{\beta}}_i$, $\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}_i(l)^\top\boldsymbol{X}_i(l))^{-1}\boldsymbol{X}_i(l)^\top\boldsymbol{y}(l)$, $\hat{\boldsymbol{y}}_0(l) = \boldsymbol{X}_0(l)\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}_0(l)^\top\boldsymbol{X}_0(l))^{-1}\boldsymbol{X}_0(l)^\top\boldsymbol{y}(l)$.

6. Simulation study

A simulation study is presented aimed at comparing the behavior of the different IBFs along with the BIC for censored data. In particular, the comparison concerns the performances of the IBFs calculated over the KMMTS with those obtained using the SMTS scheme and the BIC. The goal is to provide evidence of the fact that IBFs calculated over the KMMTS scheme produce results which are not worse, and are even better, than the IBFs calculated over the SMTS.

The conditional distribution of $y_i|\boldsymbol{x}_i$ has been simulated from the above normal regression model and the censoring indicator has been simulated according to a normal distribution, considering different censoring percentages between 10% and 50%. The right censoring indicators have been simulated by comparing the simulated observed time $y_i|\boldsymbol{x}_i$ with the censoring time. This latter has been drawn from the normal distribution of $y_i|\boldsymbol{x}_i$ shifted to the left in order to guarantee that the probability of having a censoring time smaller than the observed time Y is exactly the percentage of censoring. In any case, the simulated data-set always contains the declared percentages of censored observations.

The generating model for Y has been chosen by fixing the coefficients of the following regression model:

$$Y_i = \log(T_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma\epsilon_i \quad i = 1, \ldots, n,$$

where $\beta_0 = 0$ and $\sigma = 1$ for all generated samples, while four different sets of coefficients are considered:

$\mathcal{M}_0$ = Null model: $(\beta_1, \beta_2, \beta_3) = (0, 0, 0)$,
$\mathcal{M}_1$ = Model with 1 covariate: $(\beta_1, \beta_2, \beta_3) = (1, 0, 0)$,
$\mathcal{M}_2$ = Model with 2 covariates: $(\beta_1, \beta_2, \beta_3) = (1, 1, 0)$ and
$\mathcal{M}_3$ = Model with 3 covariates: $(\beta_1, \beta_2, \beta_3) = (1, 1, 1)$.

Different sample sizes between $n = 10$ and $n = 100$ are coupled with the above percentages of censored observations and the design matrices have been generated from p independent standard normal distributions. Two main scenarios are considered: in the first one $p = 3$ covariates (i.e. $K = 8$ models), and a second one with $p = 6$ covariates leading to a model space made up of $K = 64$ models.

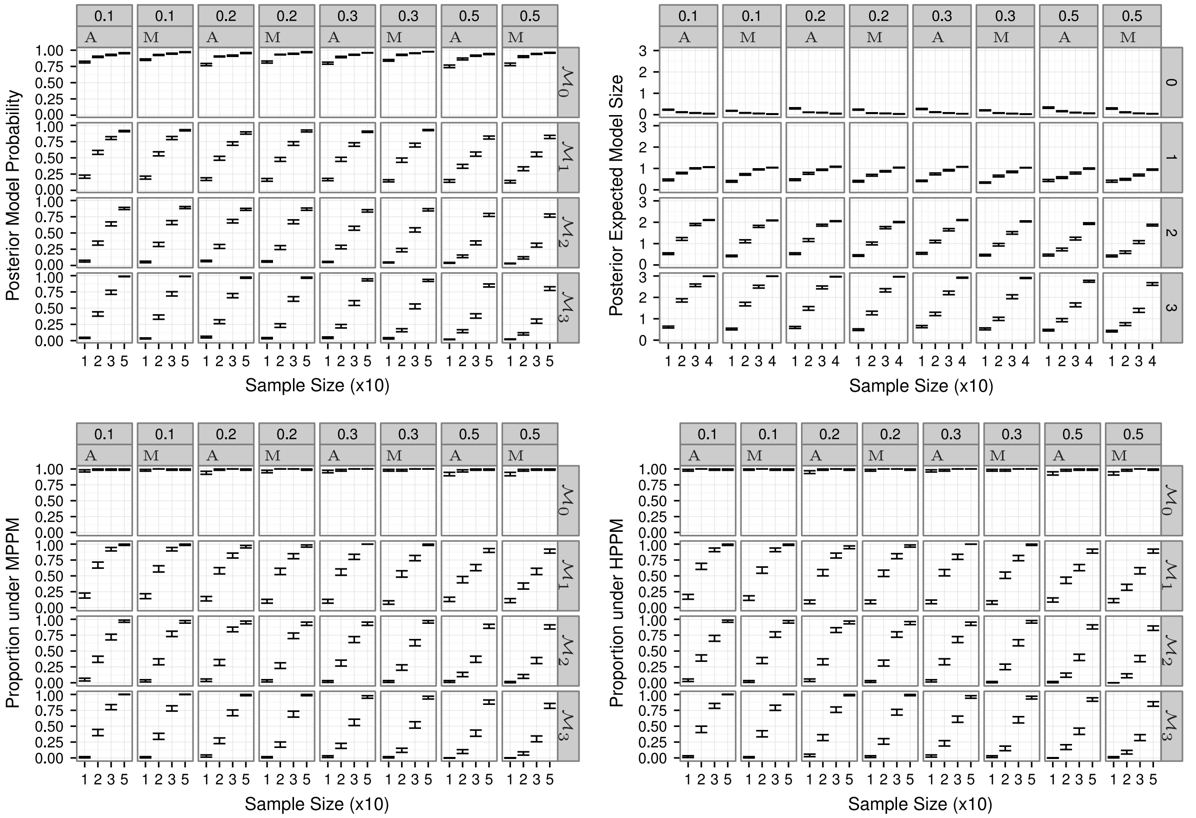

Fig. 1. Results for the IBF under the KMMTS scheme for different models (right scale), percentages of censored observations (top scale: 10%, 20%, 30% and 50%) and different sample sizes (bottom scale: $n = 10$, 20, 30 and 50 samples). Over 100 simulations, the top left panel reports the mean posterior probability for the true model, the top right the PEMS, and the bottom ones show the proportion of times that the true model has been selected according to MPPM and HPPM. All averages are displayed with $\pm$ their standard error.

Results are summarized with the proportion of times that the true model is selected according to HPPM and MPPM criteria along with the posterior expected model size (PEMS), a quantity that provides an idea as to the complexity of the models that are chosen. The PEMS is just the weighted average of the number of covariates in each of the K models, where the weights are the models' posterior probabilities.

All results are based on 100 replications for each simulation scenario. BF^{AI}_{KM} and BF^{MI}_{KM} denote the KMAIBF and the KMMIBF calculated with the KMMTS, respectively.

Fig. 1 reports the results of a wide study for $p = 3$ in order to show the sampling properties of the KMMTS BFs (KMAIBF and KMMIBF) under different percentages of censored data (top scale of all panels), sample sizes (bottom scale of all panels) and true generating model (right scale of all panels). The top-left panel reports the averaged posterior model probability obtained by the true model. For each model selection criteria (MPPM and HPPM), the bottom plots report the average frequency in which the true model has been selected. The top-left panel reports the averaged PEMS. All averages are displayed along with their standard error.

As expected from theoretical results obtained in Moreno et al. (2010) for uncensored observations, simulations suggest that the KMMTS BFs are also consistent as $n \to \infty$ with an impressive learning rate, especially for complex models (e.g. for the full model $\mathcal{M}_3$). Such consistency is reflected with respect to all indicators in the four panels of Fig. 1. The simulation study finally suggests that the increasing proportion of censored observations degrades the performance of the KMMTS BFs, decreasing equally for all models, regardless of their complexity.

Fig. 2 reports results for $p = 6$, a comparison with the SMMTS scheme under a scenario with 30% of censored observations, and two sample sizes: $n = 50$ and $n = 100$. With $p = 6$ there are $K = 64$ models with different complexity. Thus, different from the previous simulation study, it would be convenient to use the model mass prior discussed in Scott and Berger (2010). This is also known as a hierarchical uniform prior because a uniform prior is assumed over the sets of models with the same dimension. Specifically, the prior probability is $1/p$ for the null model $\mathcal{M}_0$ and for the full model with 6 covariates, while for the rest of the models the prior is obtained by dividing $1/p$ by the number of models with 2, 3, 4 and 5 covariates such that the whole probability mass prior for the models with the same complexity is again $1/p$. When all observations are uncensored, the posterior model consistency with this model prior, but not with the uniform prior $1/K$, seems to be guaranteed. Such result applies even if $p = O(n^b)$ and $b = 1$ (Moreno et al., 2014). However, there is not a formal proof of such consistency for

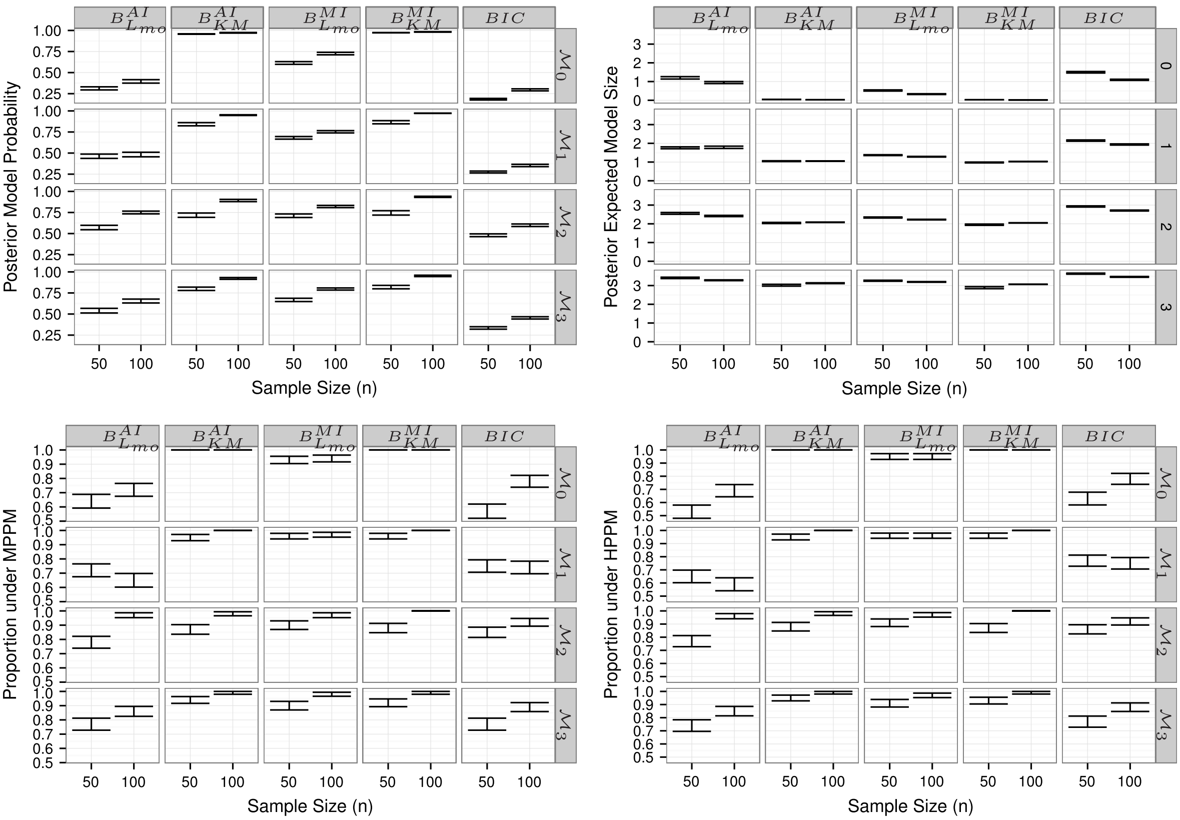

Fig. 2. Simulation results for different simulation scenarios under 30% of censored data and 6 covariates.

censored data, but intuition and the results from simulations suggest that consistency also may be maintained for certain censoring mechanisms when b is chosen in such a way to account for the proportion of censored observations.

From Fig. 2 it can be observed that there is evidence for a significative difference between the SMTS IBFs and those obtained under the KMMTS scheme. Moreover, it seems that the KMMTS scheme also provides more stable BFs with respect to $BF^{AI}_{L_{mo}}$. This result can be ascribed to the more precise computational technique used to approximate KMMTS (all correction factors have closed form expression) with respect to those obtained with the SMTS. Finally, the most appealing property is the computational time, especially relevant in this case with a large number of models. In fact, the results for KMMTS, in this simulation study, have been obtained in one hour using a machine with eight processors, while for the same generated data-sets, two weeks are necessary using the SMTS scheme and the same machine.

Summarizing, the above simulation study suggests that: B^{MI}_{KM} provides satisfactory results along all the true models and is comparable with its far more computationally demanding SMTS counterpart $B^{MI}_{L_{mo}}$; B^{AI}_{KM} performs better than $B^{AI}_{L_{mo}}$, which is less stable than $B^{MI}_{L_{mo}}$. Finally, the *BIC* performs poorly for the most simple models compared to the other BFs.

7. Application to a real data set

As an example, consider the *larynx* data set introduced in Kardaun (1983) and further discussed in Klein and Moeschberger (2003). This data set describes the survival times of $n = 90$ patients suffering from larynx cancer where $n_c = 40$ individuals were alive at the end of the study. There are two possible covariates: the *stage* of the disease, a factor with four levels, and the *age* at diagnosis. In this case, there are $K = 4$ models and results with the model uniform prior are reported, being almost the same with the hierarchical uniform prior.

All BFs agree that survival times are explained by the stage of the disease (see Table 1).

8. Further remarks

This paper provides a new tool to calculate *IBFs* in the case of a regression model with censored data (see Berger and Pericchi, 2004). It is well known that the *AIBF* and *MIBF* are consistent as they satisfy Assumption 0 of Berger and Pericchi (2004), but they can be formidable because they are so computationally demanding for censored data, when the marginal predictive distribution cannot be obtained in a closed form expression. For this purpose, a method which avoids the inclusion

Table 1
The selection criteria is reported in parenthesis when the model is chosen according to HPPM (H) and MPPM (M). The "PEMS" row reports the posterior expected model size.

Model	Model posterior probability (selection criteria)				
	B_{KM}^{AI}	B_{KM}^{MI}	BIC	B_{Lmo}^{AI}	B_{Lmo}^{MI}
stage	0.90 (H, M)	0.97 (H, M)	0.63 (H, M)	0.82 (H, M)	0.82 (H, M)
stage-age	0.09	0.01	0.36	0.14	0.15
NULL	0.01	0.02	0.01	0.04	0.02
age	0.00	0.00	0.00	0.00	0.01
PEMS	1.1	1.0	1.4	1.1	1.1

of censored observations in the TS is developed with a reduction in the computational time for each IBF while preserving, asymptotically, Assumption 0. Applications for this are illustrated in the case of the exponential and log-normal model whose predictive distribution can be obtained in closed form for uncensored data. The simulation study and the example of *larynx* data-set show that KMMTS produces similar and even better BFs than those obtained with the SMTS scheme.

Finally, the problem of deeply exploring the behavior of B_{KM}^{AI} or B_{KM}^{MI}, when the censoring mechanism is stochastically related to the covariates, is left open. It is not considered here because in the KM estimator used, the weights u_i are calculated marginally with respect to $\boldsymbol{x}$. In such a situation other survival estimators are available, such as the one proposed in Cox and Oakes (1984, Section 7.8). Ignoring effects of covariates could result in poor behavior of KMMTS BFs. However, this further complicates the analysis because of the possible intersection between the model selection problem and that of the conditional estimation of the survival function.

Acknowledgments

Maria Eugenia Castellanos was partially supported by Ministerio de Ciencia e Innovación grant MTM2010-19528, MTM2013-42323 and the visiting professor program of the Regione Autonoma della Sardegna. Stefano Cabras has been partially supported by Ministerio de Ciencia e Innovación grant MTM2013-42323, ECO2012-38442, RYC-2012-11455, and Silvia Perra was partially supported by Ministero dell'Istruzione, dell'Univesità e della Ricerca of Italy grant CRP-59903.

References

Albert, J., 2009. Bayesian Computation with R. Springer.
Barbieri, M.M., Berger, J.O., 2004. Optimal predictive model selection. Ann. Statist. 32, 870–897.
Berger, J.O., 1999. Bayes factors. In: Kotz, S., Read, S., Banks, D. (Eds.), Encyclopedia of Statistical Sciences, Vol. 3. Wiley, New York, pp. 20–29.
Berger, J.O., Pericchi, L.R., 1996. The intrinsic Bayes factor for model selection and prediction. J. Amer. Statist. Assoc. 91, 109–122. URL: http://www.jstor.org/stable/2291387.
Berger, J.O., Pericchi, L., 1997. On the justification of default and intrinsic Bayes factors. In: Modeling and Predictions. Springer-Verlag, New York, pp. 276–293.
Berger, J.O., Pericchi, L.R., 1998. Accurate and stable Bayesian model selection: the median intrinsic Bayes factor. Sankhyā: Indian J. Statist. Ser. B 60, 1–18. URL: http://www.jstor.org/stable/25053019.
Berger, J.O., Pericchi, L.R., 2001. Objective Bayesian methods for model selection: introduction and comparison. In: Lahiri, P. (Ed.), Model Selection. In: Institute of Mathematical Statistics, vol. 38. pp. 135–207. URL: http://www.jstor.org/stable/4356165.
Berger, J.O., Pericchi, L.R., 2004. Training samples in objective Bayesian model selection. Ann. Statist. 32, 841–869. URL: http://www.jstor.org/stable/3448577.
Brown, J.B.W., Hollander, M., Korwar, R.M., 1974. Nonparametric tests of independence for censored data, with applications to heart transplant studies. In: Proschan, F., Serfling, R.J. (Eds.), Reliability and Biometry: Statistical Analysis of Lifelength. Society for Industrial and Applied Mathematics, Philadelphia, USA, pp. 327–354.
Celeux, G., El Anbari, M., Marin, J.M., Robert, C.P., 2012. Regularization in regression: comparing Bayesian and frequentist methods in a poorly informative situation. Bayesian Anal. 7, 477–502. http://dx.doi.org/10.1214/12-BA716.
Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the metropolis-hastings output. J. Amer. Statist. Assoc. 96, 270–281. URL: http://pubs.amstat.org/doi/abs/10.1198/016214501750332848, arXiv:http://pubs.amstat.org/doi/pdf/10.1198/016214501750332848.
Cox, D., Oakes, D., 1984. Analysis of Survival Data. Chapman and Hall/CR.
Efron, B., 1967. The two sample problem with censored data. In: The Two Sample Problem with Censored Data. Prentice-Hall, New York, pp. 831–853.
Ghosh, J.K., Delampady, M., Samanta, T., 2006. An Introduction to Bayesian Analysis: Theory and Methods. In: Springer Texts in Statistics, Springer, New York, NY.
Gill, R.D., 1980. The Intrinsic Bayes Factor Explained by Examples. Technical Report, Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam.
Kaplan, E.R., Meier, P., 1958. Non-parametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457–481.
Kardaun, O., 1983. Statistical analysis of male larynx-cancer patients—a case study. Statist. Neerlandica 37, 103–126.
Kass, R., Raftery, A., 1995. Bayes factors. J. Amer. Statist. Assoc. 90, 773–795.
Kim, S.W., Sun, D., 2000. Intrinsic priors for model selection using an encompassing model with applications to censored failure time data. Lifetime Data Anal. 6, 251–269. http://dx.doi.org/10.1023/A:1009641709382.
Klein, J.P., Moeschberger, M.L., 2003. Survival Analysis: Techniques for Censored and Truncated Data, second ed. Springer, New York, URL: http://www.loc.gov/catdir/enhancements/fy0815/2002026667-d.html.
Lee, K.E., Kim, Y., Xu, R., 2014. Bayesian variable selection under the proportional hazards mixed-effects model. Comput. Statist. Data Anal. 75, 53–65. http://dx.doi.org/10.1016/j.csda.2014.02.009. URL: http://www.sciencedirect.com/science/article/pii/S0167947314000498.
Lingham, R.T., Sivaganesan, S., 1999. Intrinsic Bayes factor approach to a test for the power law process. J. Statist. Plann. Inference 77, 195–220. http://dx.doi.org/10.1016/S0378-3758(98)00181-5. URL: http://www.sciencedirect.com/science/article/pii/S0378375898001815.
Moeschberger, M.L., Klein, J.P., 1985. A comparison of several methods of estimating the survival function when there is extreme right censoring. Biometrics 41, 253–259.

Moreno, E., Bertolino, F., Racugno, W., 1998. An intrinsic limiting procedure for model selection and hypotheses testing. J. Amer. Statist. Assoc. 93, 1451–1460. http://www.jstor.org/stable/2670059.
Moreno, E., Bertolino, F., Racugno, W., 1999. Default Bayesian analysis of the Behrens–Fisher problem. J. Statist. Plann. Inference 81, 323–333. http://dx.doi.org/10.1016/S0378-3758(99)00070-1. URL: http://www.sciencedirect.com/science/article/pii/S0378375899000701.
Moreno, E., Girón, F., 2008. Comparison of Bayesian objective procedures for variable selection in linear regression. TEST 3, 472–492.
Moreno, E., Girón, F.J., Casella, G., 2010. Consistency of objective Bayes factors as the model dimension grows. Ann. Statist. 38, 1937–1952. http://dx.doi.org/10.1214/09-AOS754.
Moreno, E., Girón, F.J., Casella, G., 2014. Posterior Model Consistency in Variable Selection as the Model Dimension Grows. Technical Report. Universidad de Granada.
Nott, D.J., Leng, C., 2010. Bayesian projection approaches to variable selection in generalized linear models. Comput. Statist. Data Anal. 54, 3227–3241. http://dx.doi.org/10.1016/j.csda.2010.01.036. URL: http://www.sciencedirect.com/science/article/pii/S016794731000054X.
O'Hagan, A., 1995. Fractional Bayes factors for model comparison. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 99–138. URL: http://www.jstor.org/stable/2346088.
Perra, S., Cabras, S., Castellanos, M.E., 2013. Objective Bayesian Variable Selection for Censored Data. LAP Lambert Academic Publishing, ISBN: 978-3659424519.
Scott, J., Berger, J., 2010. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. Ann. Statist. 38, 2587–2619.
Varshavsky, J., 1995. On the develpment of intrinsic Bayes factors (Ph.D. thesis) Purdue University.
Volinsky, C.T., Raftery, A.E., 2000. Bayesian information criterion for censored survival models. Biometrics 56, 256–262. URL: http://www.jstor.org/stable/2677130.
Wagner, H., Duller, C., 2012. Bayesian model selection for logistic regression models with random intercept. Comput. Statist. Data Anal. 56, 1256–1274.
Yang, R., Berger, J., 1997. A Catalogue of Noninformative Priors, Technical Report, ISDS Discussion Paper 97-42. Duke University.
Zellner, A., Siow, A., 1980. Posterior odds ratios for selected regression hypotheses. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 1. Springer, Berlin, Heidelberg, pp. 585–603. http://dx.doi.org/10.1007/BF02888369.