Bootstrap Approximations in Model Checks for Binary Data

Gerhard DIKTA, Marsel KVESIC, and Christian SCHMIDT

Consider a binary regression model in which the conditional expectation of a binary variable given an explanatory variable belongs to a parametric family. To check whether a sequence of independent and identically distributed observations belongs to such a parametric family, we use Kolmogorov–Smirnov and Cramér–von Mises type tests based on a marked empirical process introduced by Stute. We propose and study a new resampling scheme for a bootstrap in this setup to approximate critical values for these tests. We also apply this approach to simulated and real data. In the latter case we check some parametric models that are used to analyze right-censored lifetime data under a semiparametric random censorship model.

KEY WORDS: Binary data; Bootstrap; Goodness of fit; Marked empirical process; Maximum likelihood estimation; Semiparametric random censorship model.

1. INTRODUCTION

Consider a sequence of independent and identically distributed (iid) random variables

$$(\delta_1, Z_1), \dots, (\delta_n, Z_n),$$

where δ is Bernoulli distributed and Z is real-valued with continuous distribution function (df) H. Because our main applications are in survival analysis, we assume throughout that Z is concentrated on the positive real line.

Binary data of this type can be observed in a variety of experiments. Typical examples are given in textbooks focusing on the generalized linear model (GLM) (see, e.g., McCullagh and Nelder 1999). Another application is in survival analysis when the data are generated under the random censorship model (RCM) and δ indicates censoring or not. More precisely, under RCM one has two independent iid sequences: the survival times $X_1, \dots, X_n$ and the censoring times $Y_1, \dots, Y_n$. One observes $Z_i := \min(X_i, Y_i)$ and $\delta_i = \mathbb{1}_{\{X_i \le Y_i\}}$, indicating whether the corresponding Z_i observation is censored or not. Based on these observations, one is interested in estimating the df of the survival time X, denoted here by F. If RCM is the only assumption, then the famous Kaplan–Meier, or product-limit, estimator (see Kaplan and Meier 1958) is an efficient nonparametric estimator of F. But Dikta (1998) and Dikta, Ghorai, and Schmidt (2005) showed that if a parametric regression model can be assumed for the binary data, then the Kaplan–Meier estimator and statistics based on it can be improved by using a semiparametric estimator of F. Note, however, that this improvement is guaranteed only if the assumed regression model for the binary data is true.

Assuming a parametric regression model in this setup means that

$$m(z) = \mathbb{E}(\delta | Z = z),$$

the conditional expectation of δ given $Z = z$, belongs to a known parametric family $\mathcal{M} := \{m(\cdot, \boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$ denotes the parameter space [i.e., $m(\cdot) = m(\cdot, \boldsymbol{\theta}_0)$] for some $\boldsymbol{\theta}_0 \in \Theta$. Logit and probit models are typical examples in this context. Note that for a parameter $\boldsymbol{\theta} \in \Theta$,

$$m(z, \boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(\delta = 1 | Z = z)$$

is the success probability of the Bernoulli variable δ given that $Z = z$.

To fit a regression function from such a parametric family to the data, usually the maximum likelihood estimator (MLE) is applied to estimate the parameter $\boldsymbol{\theta}_0$ by $\boldsymbol{\theta}_n$, say. In particular,

$$\boldsymbol{\theta}_n = \arg\max_{\boldsymbol{\theta} \in \Theta} l_n(\boldsymbol{\theta}), \tag{1}$$

where

$$l_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \big(\delta_i w_1(Z_i, \boldsymbol{\theta}) + (1 - \delta_i) w_2(Z_i, \boldsymbol{\theta})\big)$$

is the normalized log-likelihood function and

$$w_1(x, \boldsymbol{\theta}) = \ln(m(x, \boldsymbol{\theta})), \qquad w_2(x, \boldsymbol{\theta}) = \ln(1 - m(x, \boldsymbol{\theta}))$$

for $x \in \mathbb{R}$.

Needless to say, it is essential for the applicability of the semiparametric estimator in the case of a lifetime study, and also for regression analysis of binary data in a more general context, to check whether the assumed underlying regression model $\mathcal{M}$ is adequate. It is the objective of this article to investigate a universal test to check the null hypothesis

$$H_0 : m(\cdot) \in \mathcal{M} \qquad \text{versus} \qquad H_1 : m(\cdot) \notin \mathcal{M}.$$

Our method here is based on a new bootstrap version of a functional central limit theorem (FCLT) given by Stute (1997) for marked empirical processes R_n^1. In the case of binary data, this process is defined by

$$R_n^1(x) := n^{-1/2} \sum_{i=1}^{n} \big(\delta_i - m(Z_i, \boldsymbol{\theta}_n)\big) \mathbb{1}_{\{Z_i \le x\}}, \qquad 0 \le x \le \infty.$$

Closely related to R_n^1 is the process

$$R_n(x) := n^{-1/2} \sum_{i=1}^{n} \big(\delta_i - m(Z_i, \boldsymbol{\theta}_0)\big) \mathbb{1}_{\{Z_i \le x\}}, \qquad 0 \le x \le \infty.$$

In particular, Stute (1997) provided an iid representation of R_n and R_n^1 from which functional convergence to appropriate centered Gaussian processes follows easily.

Gerhard Dikta is Professor (E-mail: *dikta@fh-aachen.de*), Marsel Kvesic is Student, and Christian Schmidt is Engineer, Department of Applied Sience and Technology, Fachhochschule Aachen, Ginsterweg 1, D-52428 Jülich, Germany.

Journal of the American Statistical Association
June 2006, Vol. 101, No. 474, Theory and Methods
DOI 10.1198/016214505000001032

Based on this result, Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) test statistics can be defined as

$$D_n = \sup_{0\le x\le\infty} |R_n^1(x)| \tag{2}$$

and

$$W_n = \int (R_n^1(x))^2 H_n(dx), \tag{3}$$

where H_n is the empirical distribution function (edf) of the Z-sample. D_n and W_n can then be used to check the null hypothesis.

Note that both processes take their values in the Skorokhod space $D([0,\infty])$. The compactification of $(0,\infty)$ by $[0,\infty]$ is necessary to handle distributional convergence of the underlying processes and the associated test statistics on the whole real line, not just on compact subsets. For example, we have

$$D_n \longrightarrow D_\infty \equiv \sup_{0\le x\le\infty} |R_\infty^1(x)|$$

under H_0.

In principal, this result can be used to obtain asymptotic critical values for D_n from the df of D_∞. However, as pointed out by Stute (1997), the covariance structure of the limiting process R_∞^1 depends not only on the class $\mathcal{M}$, but also on the true but unknown θ_0. Overall, this results in a covariance structure that is not tractable in special cases of interest. In such a situation, a bootstrap may be helpful. To be precise, assume that

$$(\delta_1^*, Z_1^*), \ldots, (\delta_n^*, Z_n^*)$$

is a bootstrap sample and define

$$R_n^{1*}(x) := n^{-1/2}\sum_{i=1}^{n}\big(\delta_i^* - m(Z_i^*, \boldsymbol{\theta}_n^*)\big)\mathbb{1}_{\{Z_i^*\le x\}}, \qquad 0\le x\le\infty,$$

where $\boldsymbol{\theta}_n^*$ is the MLE corresponding to the log-likelihood function based on the bootstrap sample. Now if the process R_n^{1*} tends in distribution to the same limiting process R_∞^1, then we can approximate critical values of the KS or CvM test by the critical values of

$$D_n^* = \sup_{0\le x\le\infty} |R_n^{1*}(x)| \tag{4}$$

and

$$W_n^* = \int (R_n^{1*}(x))^2 H_n^*(dx). \tag{5}$$

Here H_n^* denotes the edf based on the Z^*-sample.

The approach outlined so far is not new. Stute, González Manteiga, and Presedo Quindimil (1998) used it to check for regression models in a more general framework and where the least squares estimator (LSE) was used to estimate the parameter of the model. Four different resampling schemes were considered, and the wild bootstrap was found to outperform the others. But classical bootstrap (CB), in which the sample is drawn from a uniform distribution on the original observations, was found to be inadequate, because the limiting process of R_n^{1*} based on CB does not coincide with R_∞^1.

Zhu, Yuen, and Tang (2002) also investigated binary data with the outlined approach, using the MLE to estimate the parameter of the model. This is exactly the same situation that we consider here. Those authors used the CB as the underlying resampling scheme. Because under this resampling scheme R_n^{1*} does not converge to R_∞^1, the authors made some corrections. In particular, they proved under some general assumptions that with probability 1, the process

$$\tilde{R}_n^{1*} = R_n^{1*} - R_n^1 \tag{6}$$

tends in distribution to R_∞^1 in the Skorokhod space $D([0,T])$, where T must be chosen such that $H(T) < 1$. This result shows that the process R_n^{1*} must be corrected by R_n^1 to achieve the correct limiting process when the bootstrap data are generated according to the CB resampling scheme. Of course, this is not a contradiction of the aforementioned result by Stute et al. (1998), because a corrected version of R_n^{1*} is used. Furthermore, Zhu et al. (2002) proposed a random symmetrization (RS) method to approximate the critical values of the CvM test. Overall, they found that both the CB and the RS performed well for moderate sample size whereas RS has better empirical power for small samples.

Generally, bootstrap data are based on the original observations. To ensure that in a bootstrap hypothesis test, the critical values obtained from the bootstrap are meaningful regardless of whether the original observations come from the null hypothesis or the alternative, the bootstrap data should be generated according to the null hypothesis (see Efron and Tibshirani 1993, chap. 16.6, p. 232). The classical bootstrap, as used by Zhu et al. (2002), does not fulfill this requirement!

To reflect the null hypothesis and the heteroscedastic nature of the binary data more closely in the bootstrap dataset, we propose and use the following model-based (MB) resampling scheme.

Definition 1. Let $(\delta_1, Z_1), \ldots, (\delta_n, Z_n)$ be iid observations where the δ_i are binary and the Z_i have a continuous df H. Let $\boldsymbol{\theta}_n$ be the corresponding MLE. The model-based resampling scheme is then defined by the following:

(a) Set $Z_i^* = Z_i$, for $1\le i\le n$.

(b) Generate a sample $\delta_1^*, \ldots, \delta_n^*$ of independent Bernoulli random variables, where δ_i^* has the probability of success given by $m(Z_i, \boldsymbol{\theta}_n)$, for $1\le i\le n$, where

$$m(z, \boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(\delta = 1|Z = z).$$

Note that under this resampling scheme, only the δ's are resampled, whereas the corresponding Z's are taken directly from the original data. Furthermore, the typical heteroscedastic variances of δ_i given Z_i are preserved under this resampling scheme.

The MB resampling scheme can also be generalized to other regression models provided that the parametric family specifies the conditional error distribution completely. In most of the parametric regression models, however, the error distributions are unknown, so that the MB resampling scheme is not applicable. Such models, especially in the case of heteroscedastic errors, should be handled with Wu's (1986) wild bootstrap resampling scheme (see Stute et al. 1998). But the wild bootstrap does not incorporate parametric model assumptions of the error distributions like the MB does; that is, the wild bootstrap is designed mainly to handle a nonparametric error scenario.

This article is organized as follows. In Section 2 we state our main results, which guarantee that the MB resampling scheme can be used to properly approximate the quantities of interest. In Section 3 we present a simulation study to compare the power of the tests under the CB and the MB resampling scheme empirically. In Section 4 we apply the approach to some well-known real lifetime datasets. For the sake of completeness, we list the results obtained for the CvM test based on the RS approach for each simulation and for the real lifetime datasets. Because of its nonparametric nature, the wild bootstrap is not included in this simulation study, in particular because the tests based on this resampling scheme are as costly as those based on the MB. Finally, we provide proofs in the Appendix.

2. MAIN RESULTS

To simplify the notation, we write $D_r m(x, \hat{\boldsymbol\theta})$ for $\frac{\partial m(x,\boldsymbol\theta)}{\partial\theta_r}|_{\boldsymbol\theta=\hat{\boldsymbol\theta}}$ and $Grad(m(x,\hat{\boldsymbol\theta}))$ for the gradient, that is, $Grad(m(x,\hat{\boldsymbol\theta})) = (D_1 m(x,\hat{\boldsymbol\theta}),\dots,D_k m(x,\hat{\boldsymbol\theta}))^t$. Furthermore, we denote partial derivatives of second order by $D_{r,s}(\cdot)$ rather than by $D_r(D_s(\cdot))$.

In this section we assume that the bootstrap data,

$$(\delta_1^*, Z_1), \dots, (\delta_n^*, Z_n),$$

are generated according to the MB resampling scheme given in Definition 1. The bootstrap MLE $\boldsymbol\theta_n^*$ is then defined similar to the MLE of the original data,

$$\boldsymbol\theta_n^* = \arg\max_{\boldsymbol\theta\in\Theta} l_n^*(\boldsymbol\theta), \tag{7}$$

where

$$l_n^*(\boldsymbol\theta) = \frac{1}{n}\sum_{i=1}^{n}\big(\delta_i^* w_1(Z_i,\boldsymbol\theta) + (1-\delta_i^*) w_2(Z_i,\boldsymbol\theta)\big)$$

denotes the corresponding normalized log-likelihood function.

Furthermore, we assume that the MLE $\boldsymbol\theta_n$ is measurable and strongly consistent and that $\boldsymbol\theta_n^*$ is consistent with probability 1; that is, for $\mathbb{P}$ almost all sample sequences and each $\varepsilon > 0$,

$$\mathbb{P}_n(|\boldsymbol\theta_n^* - \boldsymbol\theta_0| > \varepsilon) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty,$$

where $\mathbb{P}_n$ denotes the probability measure carrying the bootstrap sample.

In particular, we use the following assumptions:

(C_1) There exists a measurable solution $\boldsymbol\theta_n \in \Theta$ of the equation $Grad(l_n(\boldsymbol\theta)) = 0$ that tends to $\boldsymbol\theta_0$ with probability 1.

(C_2) For $\mathbb{P}$ almost all sample sequences, there exists a measurable solution $\boldsymbol\theta_n^* \in \Theta$ of the equation $Grad(l_n^*(\boldsymbol\theta)) = 0$ such that $\mathbb{P}_n(|\boldsymbol\theta_n^* - \boldsymbol\theta_0| > \varepsilon) \to 0$ for every $\varepsilon > 0$.

Along with these consistency assumptions, we need some regularity conditions, which were also used by Dikta (1998):

(A_1) For $i = 1, 2$, $w_i(x,\boldsymbol\theta)$ has continuous partial derivatives of second order with respect to $\boldsymbol\theta$ at each $\boldsymbol\theta\in\Theta$ and $x\in\mathbb{R}$. Furthermore, $D_r w_i(\cdot,\boldsymbol\theta)$ and $D_{r,s} w_i(\cdot,\boldsymbol\theta)$ are measurable for each $\boldsymbol\theta\in\Theta$, and there exists a neighborhood $V(\boldsymbol\theta_0)\subset\Theta$ of $\boldsymbol\theta_0$ and a measurable function M such that for all $\boldsymbol\theta\in V(\boldsymbol\theta_0)$, $x \ge 0$, and $1 \le r, s \le k$,

$$|D_{r,s} w_1(x,\boldsymbol\theta)| + |D_{r,s} w_2(x,\boldsymbol\theta)| + |D_r w_1(x,\boldsymbol\theta)| + |D_r w_2(x,\boldsymbol\theta)| \le M(x)$$

and $\mathbb{E}(M^2(Z)) < \infty$.

(A_2) The matrix $\mathbf{I}(\boldsymbol\theta_0) = (\sigma_{r,s})_{1\le r,s\le k}$, where

$$\sigma_{r,s} = \mathbb{E}\left(\frac{D_r(m(Z,\boldsymbol\theta_0))D_s(m(Z,\boldsymbol\theta_0))}{m(Z,\boldsymbol\theta_0)(1-m(Z,\boldsymbol\theta_0))}\right) \tag{8}$$

is positive definite.

(A_3) $m(x,\boldsymbol\theta)$ is continuously differentiable at each $\boldsymbol\theta\in\Theta$. Furthermore, there exists a function N such that for $x \ge 0$ and for all $\boldsymbol\theta\in\Theta$,

$$\big\|Grad(m(x,\boldsymbol\theta))\big\| \le N(x) \qquad \text{and} \qquad \mathbb{E}(N(Z)) < \infty.$$

Conditions (C_1)–(A_2) ensure asymptotic normality of $\boldsymbol\theta_n$ and $\boldsymbol\theta_n^*$. Condition (A_3) corresponds to assumption 2 of Stute (1997). General conditions for the consistency of the MLE [i.e., to guarantee (C_1)], have been discussed by Dikta (1998). Furthermore, consistency of $\boldsymbol\theta_n^*$ can be derived by a proper adaptation of the proof sketched in that article.

Now, set

$$w(\delta,x,\boldsymbol\theta) = \delta w_1(x,\boldsymbol\theta) + (1-\delta) w_2(x,\boldsymbol\theta).$$

Then (A_1) guarantees that

$$\mathbb{E}\big(D_{r,s} w(\delta,Z,\boldsymbol\theta_0)\big) = -\sigma_{r,s};$$

see the proof of theorem 2.3 of Dikta (1998).

If the parameter space Θ is open and connected, then (C_1) together with (A_1) and (A_2) ensure the asymptotic normality of $n^{1/2}(\boldsymbol\theta_n - \boldsymbol\theta_0)$, as was pointed out by Dikta (1998, thm. 2.3). Furthermore, the proof of that theorem shows that

$$n^{1/2}(\boldsymbol\theta_n - \boldsymbol\theta_0) = n^{-1/2}\sum_{i=1}^{n}\mathbf{p}(\delta_i, Z_i, \boldsymbol\theta_0) + o_P(1),$$

where

$$\mathbf{p}(\delta, z, \boldsymbol\theta) := \mathbf{I}^{-1}(\boldsymbol\theta_0)\big(Grad(w(\delta,z,\boldsymbol\theta))\big)$$

and $\mathbf{I}^{-1}(\boldsymbol\theta_0) = (\bar\sigma_{r,s})_{1\le r,s\le k}$ denotes the inverse of $\mathbf{I}(\boldsymbol\theta_0)$. Furthermore, $\mathbb{E}(\mathbf{p}(\delta,Z,\boldsymbol\theta_0)) = \mathbf{0}$. Finally, the proof and some algebra show that

$$\mathbb{E}\big(\mathbf{p}(\delta,Z,\boldsymbol\theta_0)\mathbf{p}^t(\delta,Z,\boldsymbol\theta_0)\big) = \mathbf{I}^{-1}(\boldsymbol\theta_0).$$

Thus, assumption 1 of Stute (1997) is fulfilled. Now (A_3) is tantamount to assumption 2 of Stute (1997), and thus theorem 1.2 of Stute (1997) is applicable. Hence, uniformly in x,

$$R_n^1(x) := R_n(x) - n^{-1/2}\sum_{i=1}^{n}\mathbf{A}^t(x,\boldsymbol\theta_0)\mathbf{p}(\delta_i,Z_i,\boldsymbol\theta_0) + o_P(1), \tag{9}$$

where

$$\mathbf{A}(x,\boldsymbol\theta) = \int_0^x Grad(m(t,\boldsymbol\theta))\, H(dt).$$

Furthermore, corollary 1.3 of Stute (1997) yields

$$R_n^1 \longrightarrow R_\infty^1 \quad \text{in distribution in the space } D([0,\infty]),$$

where R_∞^1 is a centered Gaussian process with covariance function

$$\begin{aligned} K(s,t) = \int_0^{s\wedge t} & m(x,\boldsymbol\theta_0)(1-m(x,\boldsymbol\theta_0))H(dx) \\ &+ \mathbf{A}^t(s,\boldsymbol\theta_0)\mathbf{I}^{-1}(\boldsymbol\theta_0)(\mathbf{A}(t,\boldsymbol\theta_0)) \\ &- \mathbf{A}^t(s,\boldsymbol\theta_0)\mathbb{E}\big(\mathbb{1}_{\{Z\le t\}}\big(\delta - m(Z,\boldsymbol\theta_0)\big)\mathbf{p}(\delta,Z,\boldsymbol\theta_0)\big) \\ &- \mathbf{A}^t(t,\boldsymbol\theta_0)\mathbb{E}\big(\mathbb{1}_{\{Z\le s\}}\big(\delta - m(Z,\boldsymbol\theta_0)\big)\mathbf{p}(\delta,Z,\boldsymbol\theta_0)\big). \end{aligned}$$

Conditioning on Z simplifies the covariance function to the effect that

$$K(s,t) = \int_0^{s\wedge t} m(x,\boldsymbol{\theta}_0)(1-m(x,\boldsymbol{\theta}_0))H(dx) - \mathbf{A}^t(t,\boldsymbol{\theta}_0)\mathbf{I}^{-1}(\boldsymbol{\theta}_0)(\mathbf{A}(s,\boldsymbol{\theta}_0)). \quad (10)$$

In our first theorem, we study the weak convergence of the bootstrap version of the MLE under the MB resampling scheme, that is, the asymptotic distribution of $n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)$.

Theorem 1. Assume that H is continuous and let Θ be a connected, open subset of $\mathbb{R}^k$. Under the null hypothesis, if assumptions (C_1), (C_2), (A_1), and (A_2) are satisfied and the MB resampling scheme is used to generate the bootstrap data, $n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)$ is asymptotically normal, $\mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta}_0))$, with probability 1, where $\mathbf{I}(\boldsymbol{\theta}_0)$ is defined under (8). Furthermore, with probability 1,

$$n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n) = n^{-1/2}\sum_{i=1}^n \mathbf{p}(\delta_i^*, Z_i, \boldsymbol{\theta}_n) + o_{\mathbb{P}_n}(1),$$

where $\mathbf{p}(\delta, z, \boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta}_0)Grad(w(\delta, z, \boldsymbol{\theta}))$.

The next theorem shows that under the null hypothesis, the FCLT holds for the marked empirical process based on the bootstrap data if the MB resampling scheme is used.

Theorem 2. Assume that H is continuous and let Θ be a connected, open subset of $\mathbb{R}^k$. Under the null hypothesis, if assumptions (C_1), (C_2), and (A_1)–(A_3) are satisfied and the MB resampling scheme is used to generate the bootstrap data, then the process R_n^{1*} tends in distribution in $D([0,\infty])$ with probability 1 to the centered Gaussian process R_∞^1.

Remark 1. Unlike in theorem 2 of Zhu et al. (2002), here a restriction of R_n^{1*} to compact intervals $[0,T]$ with $H(T) < 1$ is not necessary.

To analyze the marked empirical process based on the MB bootstrap data under the alternative, we need a proper interpretation of the parameter $\boldsymbol{\theta}_0$. For this we assume the following for the true $m(\cdot)$:

(A_4) There exists a unique parameter $\boldsymbol{\theta}_0 \in \Theta$ that maximizes

$$\mathbb{E}\big(m(Z)\ln(m(Z,\boldsymbol{\theta})) + (1-m(Z))\ln(1-m(Z,\boldsymbol{\theta}))\big).$$

(A_4) is not a strong assumption. It just guarantees that $m(\cdot,\boldsymbol{\theta}_0)$ is the projection of $m(\cdot)$ onto $\mathcal{M}$ with respect to the Kullback–Leibler geometry.

Remark 2. The MB resampling scheme guarantees that the bootstrap data are always generated under the null hypothesis even if the original data come from some alternative. It is this special feature of the MB resampling scheme that allows us to mimic the proofs of the foregoing theorems step by step even under the alternative. Thus the assertions of Theorems 1 and 2 are still correct under alternatives satisfying (A_4).

Remark 3. Remark 2 demonstrates that even under the alternative, the critical values obtained under the MB bootstrap approach are meaningful.

3. SIMULATIONS

In this simulation study we compared the approximations based on the MB, CB, and RS approaches. Generally, we used the following procedure to obtain p values for the KS- and CvM-type tests under the various approaches:

(a) Generate the original iid dataset (ODS),

$$(\delta_1, Z_1), \ldots, (\delta_n, Z_n),$$

and calculate $\boldsymbol{\theta}_n$ and W_n according to (1) and (3).

(b) Use the MB resampling scheme to generate $m = 1{,}000$ independent iid bootstrap datasets,

$$(\delta_{1,j}^*, Z_{1,j}^*), \ldots, (\delta_{n,j}^*, Z_{n,j}^*),$$

and calculate $\theta_{n,j}^*$ and $W_{n,j}^*$ according to (7) and (5), for $1 \le j \le m$.

(c) Obtain the approximated p value of W_n by

$$\frac{1}{m}\sum_{j=1}^m \mathbb{1}_{\{W_{n,j}^* > W_n\}}.$$

A simulation procedure under MB for the KS test is defined similarly. Furthermore, for each ODS, we also obtained approximate p values under the CB resampling scheme for the KS and CvM tests. But in this case we used $\tilde{R}_n^{1*}$ [see (6)] instead of R_n^{1*} to determine D_n^* and W_n^*.

For each ODS obtained under step (a), p values of the CvM test under the RS approach were calculated based on the following procedure:

(b′) Generate $m = 1{,}000$ independent iid datasets

$$\varepsilon_{1,j}, \ldots, \varepsilon_{n,j},$$

where each $\varepsilon_{i,j}$ takes values ± 1 with probability .5. Calculate, for $1 \le j \le m$,

$$W_{n,j}^{rs} = n\int_0^\infty (V_{n,j}^{rs}(x))^2 H_n(dx),$$

$$V_{n,j}^{rs}(x) = n^{-1}\sum_{i=1}^n \varepsilon_{i,j} f(Z_i, \delta_i, x),$$

$$f(Z,\delta,x) = \big(\delta - m(Z,\boldsymbol{\theta}_n)\big) \times \left(\mathbb{1}_{\{Z\le x\}} - \frac{\int_0^x \alpha(u,Z,\boldsymbol{\theta}_n)H_n(du)}{m(Z,\boldsymbol{\theta}_n)(1-m(Z,\boldsymbol{\theta}_n))}\right),$$

$$\alpha(x,Z,\boldsymbol{\theta}_n) = Grad(m(x,\boldsymbol{\theta}_n))\mathbf{I}_n^{-1}Grad(m(Z,\boldsymbol{\theta}_n)),$$

and $\mathbf{I}_n = (\sigma_{n;r,s})_{1\le r,s\le k}$, with

$$\sigma_{n;r,s} = \int_0^\infty \frac{D_r(m(x,\boldsymbol{\theta}_n))D_s(m(x,\boldsymbol{\theta}_n))}{m(x,\boldsymbol{\theta}_n)(1-m(x,\boldsymbol{\theta}_n))}H_n(dx).$$

(c′) Obtain the approximated p value of W_n by

$$\frac{1}{m}\sum_{j=1}^m \mathbb{1}_{\{W_{n,j}^{rs} > W_n\}}.$$

In our first two simulations, which focused on some models of the Aranda-Ordaz (1981) family, we repeated steps (a)–(c′) 1,000 times, where we used $\alpha = 5\%$ as a significance level. In each case $Z \sim \mathcal{N}(0,4)$; that is, Z has a normal df with expectation 0 and variance 4. Because the Z-data are distributed on $\mathbb{R}$, we must consider the processes over the whole real line.

Simulation 1. In the first simulation we used a three-parameter logit model,

$$m(z, \boldsymbol{\theta}) = \frac{\exp(\theta_1 + \theta_2 z + \theta_3 z^2)}{1 + \exp(\theta_1 + \theta_2 z + \theta_3 z^2)},$$

to generate the ODS. For the null hypothesis, we took a two-parameter logit model,

$$\mathcal{M} = \left\{ m(z, \boldsymbol{\theta}) = \frac{\exp(\theta_1 + \theta_2 z)}{1 + \exp(\theta_1 + \theta_2 z)} : \theta_1, \theta_2 \in \mathbb{R} \right\}.$$

All ODS were generated with $\theta_1 = .2$ and $\theta_2 = 1.0$. To observe the change of the power when the ODS departs from the null hypothesis, the third parameter, θ_3, was varied in steps of .1 from 0 to .3. Note that the null hypothesis is correct if $\theta_3 = 0$. The results are reported in Table 1.

In both bootstrap approaches the nominal level of 5% was reasonably well attained for all sample sizes; however, the RS-based CvM test performed far too sensitively (13.6%) for $n = 20$. Generally, the empirical power of all of the tests increased substantially with increasing sample size, at least in those cases where the ODS is not too close to H_0. All tests performed poorly under the alternative in the case $n = 20$. The best results were derived by the RS-based CvM approach here. However, because this test did not hold the nominal level of 5% under the null hypothesis for this small sample size, it is useless to discuss its empirical power for $n = 20$. In most cases the bootstrap-based CvM tests seemed to be more efficient than the corresponding KS tests. Comparing the two different resampling schemes for the bootstrap, we always observe better results under MB. For moderate sample sizes of $n = 50$ and $n = 100$, the MB-based CvM tests outperformed the other two tests, whereas the CvM tests based on the RS approach showed better results than the CB-based CvM tests.

Simulation 2. In our next simulation we generated the ODS according to a three-parameter complementary log–log model; that is, we used

$$m(z, \boldsymbol{\theta}) = 1 - \exp\bigl(-\exp(\theta_1 + \theta_2 z + \theta_3 z^2)\bigr),$$

with different parameter sets $\boldsymbol{\theta}^t = (\theta_1, \theta_2, \theta_3)$, whereas the null hypothesis was a two-parameter complementary log–log model, that is,

$$\mathcal{M} = \bigl\{ m(z, \boldsymbol{\theta}) = 1 - \exp(-\exp(\theta_1 + \theta_2 z)) : \theta_1, \theta_2 \in \mathbb{R} \bigr\}.$$

Table 1. Percentages of Rejecting H_0 in Simulation 1, $\alpha = 5\%$

	Sample	RS	MB		CB	
$\boldsymbol{\theta}^t$	size	CvM	CvM	KS	CvM	KS
(.2, 1.0, 0)	20	13.6	4.4	5.3	2.0	4.9
	50	7.2	5.0	6.8	3.6	6.5
	100	4.3	4.0	5.0	3.9	5.6
(.2, 1.0, .1)	20	10.6	4.2	4.2	2.2	3.3
	50	10.7	11.0	9.2	6.0	7.7
	100	11.7	12.4	10.0	10.1	9.3
(.2, 1.0, .2)	20	15.2	11.6	9.6	6.5	9.1
	50	22.6	25.0	19.3	18.9	15.5
	100	46.2	53.1	39.2	43.3	31.6
(.2, 1.0, .3)	20	24.1	22.8	18.7	16.1	15.7
	50	52.3	56.4	41.2	50.1	37.4
	100	83.9	86.9	72.9	83.3	70.4

Table 2. Percentages of Rejecting H_0 in Simulation 2, $\alpha = 5\%$

	Sample	RS	MB		CB	
$\boldsymbol{\theta}^t$	size	CvM	CvM	KS	CvM	KS
(−.2, −1, 0)	20	22.1	3.5	4.5	3.9	9.1
	50	10.1	4.6	4.7	2.5	4.6
	100	7.9	4.3	5.0	4.9	6.4
(−.2, −1, −.175)	20	24.6	14.6	14.9	2.3	5.0
	50	32.4	33.9	30.6	5.0	4.9
	100	52.9	60.5	52.8	13.0	11.7
(−.2, −1, −.200)	20	31.3	25.9	22.8	2.6	4.4
	50	48.8	53.9	49.0	6.4	8.3
	100	69.9	80.7	71.8	14.8	13.5
(−.2, −1, −.225)	20	34.0	34.6	31.3	2.3	4.5
	50	56.6	67.1	60.5	8.3	8.9
	100	80.1	91.8	86.3	14.7	13.9

As in the first simulation, we left the first two parameters constant, $\theta_1 = -.2$ and $\theta_2 = -1.0$, and varied the third parameter, $\theta_3 = 0, -.175, -.200, -.225$. Again, we generated the ODS under the null hypothesis in the case of $\theta_3 = 0$. The results are given in Table 2.

In the case of the complementary log–log model, the results are similar to those observed in the first simulation under the logit model. The nominal level of 5% was attained by the MB-based tests quite well and more accurately than by the CB-based tests, whereas the KS test, at 9.1%, was also too sensitive in the case of sample size $n = 20$. The RS-based CvM tests were far too sensitive under the null hypothesis for $n = 20$ (22.1%). In addition, this test also failed to attain the nominal level for $n = 50$ (10.1%). As for the logit model, the empirical power increased with increasing distance of the ODS from the null hypothesis and with increasing sample size, when the RS- or MB-based tests are used. The CB-based tests perform poorly under all of the alternatives considered. All results under the considered alternatives, provided that the test attains the nominal level under the null hypothesis, indicate that the MB-based CvM tests were more efficient than the RS- and CB-based tests, whereas the RS-based tests outperformed the CB-based tests. For the MB-based tests, CvM provided better results than KS.

To compare the empirical power of CvM under the two resampling schemes and the RS approach more illustratively, we plot the empirical df of the obtained 1,000 p values for all tests in Figure 1, where $\theta_1 = -.2$, $\theta_2 = -1$, $\theta_3 = -.225$, and sample size $n = 100$. Obviously, the MB-based CvM test outperformed the other two tests for all reasonable nominal levels.

Simulations 3–5 concentrate on survival data generated under semiparametric random censorship models (SRCMs). In general, we have an iid sequence $X_1, \dots, X_n$ of survival times with df F and, independently of this sequence, another iid sequence $Y_1, \dots, Y_n$ of censoring times with df G. The observations are denoted by $(\delta_1, Z_1), \dots, (\delta_n, Z_n)$, where $Z_i = \min(X_i, Y_i)$ and $\delta_i = \mathbb{1}_{\{X_i \le Y_i\}}$. Furthermore, we denote the df of Z by H. So far, this is the RCM assumption. In addition, in SRCM, a parametric model for $m(z) = \mathbb{E}(\delta | Z = z)$ is assumed. As pointed out by Dikta (1998, p. 255), m is linked to the hazard functions λ_f, λ_g, and λ_h through

$$m(z) = \frac{\lambda_f(z)}{\lambda_h(z)} = \frac{\lambda_f(z)}{\lambda_f(z) + \lambda_g(z)}, \tag{11}$$

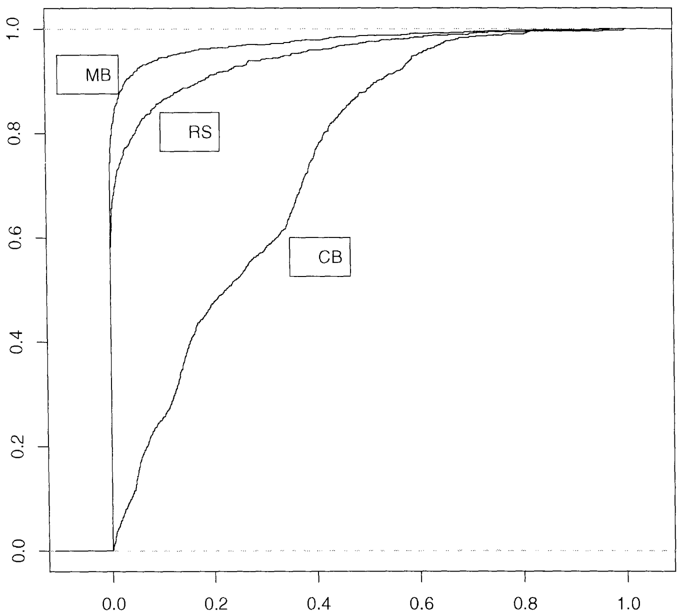

Figure 1. Empirical df of 1,000 p Values of the CvM Tests.

where f, g, and h denote the density functions corresponding to F, G, and H.

Among other things, Zhu et al. (2002) investigated the simple proportional hazards model (PHM), which is a special example of a SRCM (see Dikta 1998, ex. 2.8), in a simulation study. To be precise, under PHM it is assumed that

$$m \in \mathcal{M}_1 = \{m(z, \theta) = \theta : 0 < \theta < 1\}. \tag{12}$$

In the following simulations, we compared the MB-based CvM test with the CB-based and RS-based tests under conditions similar to those of Zhu et al. (2002). After this, we repeated the steps (a)–(c′) 1,000 times to obtain the estimated p values.

Simulation 3. In this simulation we used for the censoring distribution the exponential distribution with parameter 1. The survival time was also exponentially distributed and adjusted to get $c = 25\%$ and 35% of censored observations. Note that the ODS was generated under the null hypothesis, $m \in \mathcal{M}_1$. The results are reported in Table 3. Overall, we found that both procedures attain the nominal levels quite well (see also Zhu et al. 2002). However, as also reported by Zhu et al. (2002, p. 115), the CB-based CvM test may be a bit too sensitive for a small sample size like $n = 30$.

Simulation 4. A generalization of PHM was given by Dikta (1998, ex. 2.9) and is denoted here by GPHM. This model is adequate when X and Y are Weibull distributed, that is, $F(x) = 1 - \exp(-(ax)^b)$ and $G(x) = 1 - \exp(-(cx)^d)$. In this case,

$$m \in \mathcal{M}_2 = \left\{ m(z, \boldsymbol{\theta}) = \frac{\theta_1}{\theta_1 + x^{\theta_2}} : \theta_1 > 0, \theta_2 \in \mathbb{R} \right\}, \tag{13}$$

where $\theta_1 = (ba^b)/(dc^d)$ and $\theta_2 = d - b$. To generate the ODS, we used (as in Zhu et al. 2002, table 2), the parameters $(a, b, c, d) = (1.5, .8, .1, .5)$ and $(a, b, c, d) = (1.7, .8, .3, .5)$, which led to $c = 20\%$ and $c = 30\%$ of censoring. Note that the ODS is now generated under the alternative and $m \in \mathcal{M}_1$ is no longer correct. The results are reported in Table 4.

All three approaches showed better results under $c = 30\%$ censoring than under $c = 20\%$ censoring. In all cases, the MB-based CvM tests and (except under the nominal level of 5% for the case of $n = 200$ and $c = 30\%$) the CB-based tests were slightly better than the RS-based ones. The MB-based CvM tests were slightly better than the CB-based tests for sample sizes $n = 200$, $n = 100$, and $n = 50$ (under $c = 20\%$ of censoring). Only for sample sizes $n = 30$ and $n = 50$ (under $c = 30\%$ of censoring) did we observe slightly higher empirical power for the CB-based tests than for the MB-based tests. But, as we observed in Simulation 3, the CB-based test might be too sensitive for small sample sizes.

Simulation 5. In our last simulation, we used for the censoring distribution G the mean 1 exponential and for F Weibull distributions of the type $F(x) = 1 - \exp(-(ax)^2)$ (cf. Zhu et al. 2002, table 3). The parameter a was chosen to get again $c = 20\%$ and $c = 30\%$ of censoring. According to (11), we then had $m(z) = 2az/(1 + 2az)$, which showed that the ODS are not generated according to the null hypothesis $m \in \mathcal{M}_1$. The results of this simulation, reported in Table 5, reflect the results observed in Simulation 4.

4. REAL DATA APPLICATIONS

PHM, as a model for survival data, was introduced by Koziol and Green (1976). They analyzed CvM-type goodness-of-fit tests for the survival function based on the KM estimator under

Table 3. Percentages of Rejecting H_0 in Simulation 3

Sample size	Nominal level	RS $c = 25\%$	RS $c = 35\%$	MB $c = 25\%$	MB $c = 35\%$	CB $c = 25\%$	CB $c = 35\%$
30	1%	.3	1.0	.6	.8	.6	1.6
	5%	4.8	5.0	4.6	4.0	5.5	5.7
	10%	11.3	10.1	11.0	9.7	12.7	10.9
50	1%	1.0	1.0	1.1	1.1	1.7	1.4
	5%	4.9	5.1	5.1	4.8	5.6	5.8
	10%	10.1	11.1	9.5	10.6	11.4	11.6
100	1%	1.0	.7	1.0	.5	1.2	.8
	5%	3.6	5.3	3.9	5.2	4.1	5.7
	10%	8.5	10.8	9.3	10.8	9.0	11.5
200	1%	1.0	.9	1.0	.8	1.3	1.0
	5%	5.2	4.3	4.7	4.2	5.2	4.4
	10%	10.2	9.5	10.2	9.8	10.3	9.6

Table 4. Percentages of Rejecting H_0 in Simulation 4

Sample size	Nominal level	RS $c = 20\%$	RS $c = 30\%$	MB $c = 20\%$	MB $c = 30\%$	CB $c = 20\%$	CB $c = 30\%$
30	1%	5.6	9.1	5.8	10.1	6.1	12.1
	5%	16.0	22.6	17.3	23.5	17.0	25.6
	10%	26.1	36.1	28.1	37.7	27.2	38.6
50	1%	9.2	17.5	10.7	19.0	9.8	20.0
	5%	25.2	37.3	27.2	39.1	26.6	39.9
	10%	37.5	49.6	39.1	51.4	38.4	51.4
100	1%	24.6	42.4	27.8	45.4	26.3	45.1
	5%	49.4	67.8	51.5	69.8	50.1	68.8
	10%	63.6	79.4	65.4	80.4	64.0	80.0
200	1%	61.4	78.2	64.0	80.3	61.8	79.3
	5%	80.1	91.4	81.9	92.2	80.6	90.9
	10%	87.4	95.0	88.4	95.5	87.5	95.4

PHM. Later, Abdushukurov (1987) and Cheng and Lin (1987) independently introduced an estimator of the survival function that improves the Kaplan–Meier estimator in terms of asymptotic variance if the assumption of PHM is correct. (See Csörgő 1988b for a review of fundamental properties of estimation under PHM.)

Special tests to check for PHM have been given by Csörgő (1988a) and Henze (1993). These tests have been applied to some well-known real datasets by both authors, with the result that PHM is a model assumption that can hardly be found in practice.

In what follows, we apply the MB-based and CB-based tests to three of the datasets investigated by Csörgő (1988a) and Henze (1993) to check for PHM and for GPHM. In particular, we take $H_{0,1}$ given by (12) and $H_{0,2}$ given by (13) as null hypotheses.

The first dataset, the Channing House data (CHD), was published by Hyde (1977). The dataset comprises survival times of $n = 97$ men who lived in Channing House, a Palo Alto retirement center, over the period 1964 to July 1, 1975. A total of 51 observations are censored, and we ignore (as in Csörgő 1988a and Henze 1993) the entry dates.

The second dataset, the Stanford heart transplant data (HTD), was provided by Miller and Halpern (1982). This dataset reports the lifetimes of $n = 184$ patients who received a heart transplant within the heart transplantation program from October 1967 to February 1980. In this dataset, 71 observations are censored.

Our third dataset, the Oestrogen treatment data (OTD), were given by Hollander and Proschan (1979). These data show the survival times of $n = 211$ stage IV prostate cancer patients treated with estrogen. A total of 121 observations in this dataset are censored.

Table 6 lists the results of these tests. Because GPHM includes PHM, we observe higher p values in testing $H_{0,2}$ than in testing $H_{0,1}$. Like the tests proposed by Csörgő (1988a) and Henze (1993), the p values obtained under all three approaches reject PHM for the HTD and OTD datasets, whereas PHM cannot be rejected for the CHD data. In the case of GPHM, we obtained p values that also reject this model for the OTD data (at least for a significance level of 1.5%), whereas GPHM fit to CHD and HTD. Thus we can analyze the CHD and the HTD with the semiparametric estimator introduced by Dikta (1998), with GPHM as an underlying parametric model for m.

5. CONCLUSIONS

In this article we have considered the class of marked empirical processes where the dependent variables are binary. We have seen that these processes can be approximated in distribution by corresponding bootstrap processes if the new MB resampling scheme is used. Based on this result, universal goodness-of-fit tests, like KS or CvM, can be used for model checks. In the simulation study we observed a significant gain in empirical power of CvM- and KS-type tests under the MB resampling scheme compared with the corresponding tests, which are based on a

Table 5. Percentages of Rejecting H_0 in Simulation 5

Sample size	Nominal level	RS $c = 20\%$	RS $c = 30\%$	MB $c = 20\%$	MB $c = 30\%$	CB $c = 20\%$	CB $c = 30\%$
30	1%	12.2	21.0	14.7	20.9	15.2	24.8
	5%	30.9	42.5	33.1	45.0	33.3	46.2
	10%	42.8	55.3	46.7	58.0	45.6	57.4
50	1%	24.4	37.9	29.2	40.6	26.2	42.4
	5%	48.8	63.8	52.9	67.9	50.7	66.6
	10%	61.1	75.6	65.3	77.7	61.7	76.3
100	1%	58.1	78.0	63.9	80.3	59.6	80.5
	5%	81.1	92.1	84.7	93.3	81.3	92.7
	10%	88.4	96.4	90.8	96.8	88.7	96.4
200	1%	91.4	98.5	93.3	98.8	91.6	99.0
	5%	97.8	99.7	98.3	99.9	97.9	99.6
	10%	99.1	100.0	99.3	100.0	99.2	99.9

Table 6. p Values of PHM and GPHM in RS-, MB-, and CB-Based Tests

Data	Null hypothesis	RS CvM	MB CvM	MB KS	CB CvM	CB KS
CHD	$H_{0,1}$	.685	.641	.434	.663	.429
	$H_{0,2}$	.826	.702	.563	.781	.611
HTD	$H_{0,1}$	0	0	0	0	0
	$H_{0,2}$	.662	.449	.756	.656	.822
OTD	$H_{0,1}$	0	0	0	0	0
	$H_{0,2}$	.004	.006	.010	.005	.014

corrected bootstrap version of the marked empirical process under the CB resampling scheme. As a consequence, bootstrapping in this scenario should be done under the MB resampling scheme.

Furthermore, in our simulation study the CvM tests based on the random symmetrization approach were outperformed by the MB-based bootstrap tests. However, the RS-based test had the advantage of lower computational cost than the bootstrap-based tests, because for the second one must calculate the maximizer of the likelihood function for every bootstrap dataset, which is numerically expensive.

Finally, we have tested PHM for some real survival data with the RS-based and the bootstrap based tests and obtained p values that led to the same decisions as some specialized tests, when they are applied to check for PHM.

APPENDIX: PROOFS

Throughout this section we let $\mathbb{P}_n, \mathbb{E}_n$, var_n, and cov_n denote the probability measure, expectation, variance, and covariance associated with the bootstrap sample.

Lemma A.1. Under the assumptions of Theorem 1, we get, with probability 1,

$$\mathbf{A}_n^*(\tilde{\boldsymbol\theta}_n^*) = -\mathbf{I}(\boldsymbol\theta_0) + o_{\mathbb{P}_n}(1),$$

where $\mathbf{A}_n^*(\boldsymbol\theta) = (a_{r,s}^{n*}(\boldsymbol\theta))_{1\le r,s\le k}$ is a $k\times k$ matrix with $a_{r,s}^{n*}(\boldsymbol\theta) = D_{r,s}l_n^*(\boldsymbol\theta)$, for $1\le r,s\le k$, and where $\tilde{\boldsymbol\theta}_n^*$ lies on the line segment joining $\boldsymbol\theta_n^*$ and $\boldsymbol\theta_n$, for every $n\in\mathbb{N}$.

Proof. Fix $1\le r,s\le k$, $\varepsilon>0$, $\gamma>0$, and denote by V_γ the γ-neighborhood of $\boldsymbol\theta_0$. (C_1) and (C_2) then yield that, with probability 1, $\mathbb{P}_n(|\tilde{\boldsymbol\theta}_n^* - \boldsymbol\theta_0| > \gamma) \longrightarrow 0$. Apply Markov's inequality to get, with probability 1,

$$\mathbb{P}_n\big(|a_{r,s}^{n*}(\tilde{\boldsymbol\theta}_n^*) - a_{r,s}^{n*}(\boldsymbol\theta_0)| > \varepsilon\big) \le \varepsilon^{-1}\mathbb{E}_n\Big(\sup_{\boldsymbol\theta\in V_\gamma}|a_{r,s}^{n*}(\boldsymbol\theta) - a_{r,s}^{n*}(\boldsymbol\theta_0)|\Big) + o(1).$$

Now we use a continuity argument that occurs in similar form throughout this proof section but that we carry out in detail only once, in the next step, to handle the expectation on the right side of the foregoing inequality. This expectation is bounded from above by

$$n^{-1}\sum_{i=1}^n\Big(\sup_{\boldsymbol\theta\in V_\gamma}|D_{r,s}w_1(Z_i,\boldsymbol\theta) - D_{r,s}w_1(Z_i,\boldsymbol\theta_0)| + \sup_{\boldsymbol\theta\in V_\gamma}|D_{r,s}w_2(Z_i,\boldsymbol\theta) - D_{r,s}w_2(Z_i,\boldsymbol\theta_0)|\Big),$$

which tends to

$$\mathbb{E}\Big(\sup_{\boldsymbol\theta\in V_\gamma}|D_{r,s}w_1(Z,\boldsymbol\theta) - D_{r,s}w_1(Z,\boldsymbol\theta_0)| + \sup_{\boldsymbol\theta\in V_\gamma}|D_{r,s}w_2(Z,\boldsymbol\theta) - D_{r,s}w_2(Z,\boldsymbol\theta_0)|\Big)$$

with probability 1, according to (A_1) and the strong law of large numbers (SLLN). (A_1) and Lebesgue's theorem then guarantee that this expectation tends to 0 when $\gamma\to 0$. Because γ was chosen arbitrarily, we therefore get, with probability 1, that

$$\mathbf{A}_n^*(\tilde{\boldsymbol\theta}_n^*) = \mathbf{A}_n^*(\boldsymbol\theta_0) + o_{\mathbb{P}_n}(1).$$

Furthermore,

$$\mathbb{E}_n(a_{r,s}^{n*}(\boldsymbol\theta_0)) = n^{-1}\sum_{i=1}^n m(Z_i,\boldsymbol\theta_n)D_{r,s}w_1(Z_i,\boldsymbol\theta_0) + (1-m(Z_i,\boldsymbol\theta_n))D_{r,s}w_2(Z_i,\boldsymbol\theta_0),$$

which tends to $\mathbb{E}(D_{r,s}w(\delta,Z,\boldsymbol\theta_0)) = -\sigma_{r,s}$ with probability 1 according to the SLLN and a continuity argument similar to the one given earlier. Finally, Chebyshev's inequality, a continuity argument, and (A_1) yield that, with probability 1,

$$\mathbf{A}_n^*(\tilde{\boldsymbol\theta}_n^*) = -\mathbf{I}(\boldsymbol\theta_0) + o_{\mathbb{P}_n}(1).$$

This completes the proof of the lemma.

Proof of Theorem 1

Because of assumptions (C_1) and (C_2), Taylor's expansion yields that, with probability 1,

$$n^{1/2}\mathbf{A}_n^*(\tilde{\boldsymbol\theta}_n^*)(\boldsymbol\theta_n^* - \boldsymbol\theta_n) = -n^{-1/2}\sum_{i=1}^n Grad(w(\delta_i^*, Z_i, \boldsymbol\theta_n)), \quad (A.1)$$

where $\tilde{\boldsymbol\theta}_n^*$ lies on the line segment joining $\boldsymbol\theta_n^*$ and $\boldsymbol\theta_n$. Thus we can apply Lemma A.1 to get

$$\mathbf{A}_n^*(\tilde{\boldsymbol\theta}_n^*) = -\mathbf{I}(\boldsymbol\theta_0) + o_{\mathbb{P}_n}(1).$$

Consider now the right side of (A.1) and note that each term of the sum is centered. To show the asymptotic normality, we apply the Cramér–Wold device and fix $\mathbf{a}\in\mathbb{R}^k$ to get

$$\text{var}_n\left(n^{-1/2}\sum_{i=1}^n \mathbf{a}^t Grad(w(\delta_i^*, Z_i, \boldsymbol\theta_n))\right) = n^{-1}\sum_{i=1}^n\sum_{1\le r,s\le k} a_r a_s \frac{D_r(m(Z_i,\boldsymbol\theta_n))D_s(m(Z_i,\boldsymbol\theta_n))}{m(Z_i,\boldsymbol\theta_n)(1-m(Z_i,\boldsymbol\theta_n))}.$$

The SLLN, (A_1), and a continuity argument then yield that, with probability 1,

$$\text{var}_n\left(n^{-1/2}\sum_{i=1}^n \mathbf{a}^t Grad(w(\delta_i^*, Z_i, \boldsymbol\theta_n))\right) \longrightarrow \mathbf{a}^t\mathbf{I}(\boldsymbol\theta_0)(\mathbf{a}).$$

To verify Lindeberg's condition, we need to prove, because of the result just mentioned, that for every $\varepsilon>0$,

$$L_n(\varepsilon) = n^{-1}\sum_{i=1}^n W_i^2(\boldsymbol\theta_n)\mathbb{E}_n\Big(\big(\delta_i^* - m(Z_i,\boldsymbol\theta_n)\big)^2 \times \mathbb{1}_{\{|W_i(\boldsymbol\theta_n)(\delta_i^* - m(Z_i,\boldsymbol\theta_n))| > n^{1/2}\varepsilon\}}\Big) \longrightarrow 0$$

with probability 1, where

$$W_i(\boldsymbol\theta_n) = \frac{\mathbf{a}^t Grad(m(Z_i,\boldsymbol\theta_n))}{m(Z_i,\boldsymbol\theta_n)(1-m(Z_i,\boldsymbol\theta_n))}.$$

Calculating the expectation yields

$$L_n(\varepsilon) = n^{-1}\sum_{i=1}^{n}\sum_{1\le r,s\le k} a_r a_s \frac{D_r(m(Z_i,\boldsymbol{\theta}_n)D_s(m(Z_i,\boldsymbol{\theta}_n)))}{m(Z_i,\boldsymbol{\theta}_n)(1-m(Z_i,\boldsymbol{\theta}_n))}$$
$$\times\Big((1-m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{|\mathbf{a}^t Grad(w_1(Z_i,\boldsymbol{\theta}_n))>n^{1/2}\varepsilon\}}$$
$$+ m(Z_i,\boldsymbol{\theta}_n)\mathbb{1}_{\{|\mathbf{a}^t Grad(w_2(Z_i,\boldsymbol{\theta}_n))>n^{1/2}\varepsilon\}}\Big).$$

According to (C_1), $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$ with probability 1. Therefore, we can apply (A_1) to bound

$$L_n(\varepsilon) \le 2n^{-1}\sum_{i=1}^{n}\sum_{1\le r,s\le k}|a_r||a_s|M^2(Z_i)\mathbb{1}_{\{\|\mathbf{a}\|\sqrt{k}M(Z_i)>n^{1/2}\varepsilon\}},$$

where we also used the Cauchy–Schwarz inequality to simplify the indicator functions.

Now fix $T>0$ and use SLLN to get, with probability 1,

$$\limsup_{n\to\infty} L_n(\varepsilon) \le c\mathbb{E}\big(M^2(Z)\mathbb{1}_{\{M(Z)>T\}}\big),$$

for $c<\infty$ properly chosen. But the right side tends to 0 as $T\to\infty$, due to assumption (A_1), which finally verifies Lindeberg's condition.

Thus, with probability 1, the right side of (A.1) is asymptotically normal with covariance matrix $\mathbf{I}(\boldsymbol{\theta}_0)$. Together with a Cramér–Slutsky argument and the regularity of $\mathbf{I}(\boldsymbol{\theta}_0)$, this completes the proof of Theorem 1.

The following lemma handles the bootstrap counterpart of the decomposition given under (9).

Lemma A.2. Under the assumptions of Theorem 2, we get with probability 1 uniformly in x,

$$R_n^{1*}(x) = R_n^*(x) - n^{-1/2}\sum_{i=1}^{n}\mathbf{A}^t(x,\boldsymbol{\theta}_0)\mathbf{p}(\delta_i^*, Z_i,\boldsymbol{\theta}_n) + o_{\mathbb{P}_n}(1)$$
$$\equiv R_n^{2*}(x) + o_{\mathbb{P}_n}(1),$$

where

$$R_n^*(x) = n^{-1/2}\sum_{i=1}^{n}\big(\delta_i^* - m(Z_i,\boldsymbol{\theta}_n)\big)\mathbb{1}_{\{Z_i\le x\}}.$$

Proof. Consider the decomposition

$$R_n^{1*}(x) = n^{-1/2}\sum_{i=1}^{n}\big(\delta_i^* - m(Z_i,\boldsymbol{\theta}_n^*)\big)\mathbb{1}_{\{Z_i\le x\}}$$
$$= n^{-1/2}\sum_{i=1}^{n}\big(\delta_i^* - m(Z_i,\boldsymbol{\theta}_n)\big)\mathbb{1}_{\{Z_i\le x\}}$$
$$+ n^{-1/2}\sum_{i=1}^{n}\big(m(Z_i,\boldsymbol{\theta}_n) - m(Z_i,\boldsymbol{\theta}_n^*)\big)\mathbb{1}_{\{Z_i\le x\}}$$
$$\equiv R_n^*(x) + S_n^*(x),$$

and apply Taylor's expansion to get

$$S_n^*(x) = -n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)^t n^{-1}\sum_{i=1}^{n} Grad(m(Z_i,\tilde{\boldsymbol{\theta}}_{n,i}^*))\mathbb{1}_{\{Z_i\le x\}},$$

where $\tilde{\boldsymbol{\theta}}_{n,i}^*$ lies on the line segment joining $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}_n^*$, for $1\le i\le n$.

Theorem 1, assumptions (C_1), (C_2), and (A_3), the SLLN, and a continuity argument then yield that, with probability 1,

$$S_n^*(x) = -n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)^t n^{-1}\sum_{i=1}^{n} Grad(m(Z_i,\boldsymbol{\theta}_0))\mathbb{1}_{\{Z_i\le x\}} + o_{\mathbb{P}_n}(1),$$

uniformly in x. Note that our processes are defined on the compact set $[0,\infty]$. Therefore, we can apply, due to (A_3), theorem 2 of Jennrich (1969) to obtain that, with probability 1,

$$S_n^*(x) = -n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)^t\mathbf{A}(x,\boldsymbol{\theta}_0) + o_{\mathbb{P}_n}(1),$$

which completes the proof of the lemma in view of Theorem 1.

In the next lemma we analyze the covariance function of R_n^{2*}.

Lemma A.3. Under the assumptions of Theorem 2, the covariance function

$$K_n^*(s,t) = \mathrm{cov}_n\big(R_n^{2*}(s), R_n^{2*}(t)\big)$$

tends, with probability 1, to the covariance function $K(s,t)$ given under (10), for $0\le s,t\le\infty$.

Proof. Fix $0\le s\le t\le\infty$ and set

$$\mathbf{I}_n = \left(n^{-1}\sum_{i=1}^{n}\frac{D_r(m(Z_i,\boldsymbol{\theta}_n))D_s(m(Z_i,\boldsymbol{\theta}_n))}{m(Z_i,\boldsymbol{\theta}_n)(1-m(Z_i,\boldsymbol{\theta}_n))}\right)_{1\le r,s\le k}.$$

A direct calculation of $K_n^*(s,t)$ yields

$$K_n^*(s,t) = n^{-1}\sum_{i=1}^{n} m(Z_i,\boldsymbol{\theta}_n)(1-m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{Z_i\le s\}}$$
$$+ \mathbf{A}^t(t,\boldsymbol{\theta}_0)\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{I}_n\mathbf{I}^{-1}(\boldsymbol{\theta}_0)(\mathbf{A}(s,\boldsymbol{\theta}_0))$$
$$- \mathbf{A}^t(t,\boldsymbol{\theta}_0)\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\left(n^{-1}\sum_{i=1}^{n} Grad(m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{Z_i\le s\}}\right)$$
$$- \mathbf{A}^t(s,\boldsymbol{\theta}_0)\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\left(n^{-1}\sum_{i=1}^{n} Grad(m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{Z_i\le t\}}\right).$$

Note that $\mathbf{I}_n \to \mathbf{I}(\boldsymbol{\theta}_0)$ with probability 1 according to the SLLN, assumptions (C_1) and (A_1), and a continuity argument. Furthermore, the SLLN, assumptions (C_1) and (A_3), and a continuity argument guarantee that, with probability 1,

$$n^{-1}\sum_{i=1}^{n} Grad(m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{Z_i\le x\}} \longrightarrow \mathbf{A}(x,\boldsymbol{\theta}_0),$$

for every $0\le x\le\infty$, which finally completes the proof of the lemma.

Lemma A.4. Under the assumptions of Theorem 2, the finite-dimensional distribution of R_n^{2*} converges to those of R_∞^1 with probability 1.

Proof. A straightforward calculation shows that

$$R_n^{2*}(x)$$
$$= n^{-1/2}\sum_{i=1}^{n}\left(\mathbb{1}_{\{Z_i\le x\}} - \frac{\mathbf{A}^t(x,\boldsymbol{\theta}_0)\mathbf{I}^{-1}(\boldsymbol{\theta}_0)(Grad(m(Z_i,\boldsymbol{\theta}_n)))}{m(Z_i,\boldsymbol{\theta}_n)(1-m(Z_i,\boldsymbol{\theta}_n))}\right)$$
$$\times\big(\delta_i^* - m(Z_i,\boldsymbol{\theta}_n)\big).$$

In view of Lemma A.3, it suffices to show asymptotic normality for the finite-dimensional distributions of R_n^{2*}. This can be done by applying the Cramér–Wold device and verifying the corresponding Lindeberg condition. We omit the proof here, because the arguments are similar to those given in the proof of Theorem 1.

Proof of Theorem 2

Because of the foregoing lemmas, it remains to show that, with probability 1, $(R_n^{2*})_{n\geq 1}$ is tight. Because the sequence defined by the second summands, that is, by

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{A}^t(\cdot,\boldsymbol{\theta}_0)\mathbf{p}(\delta_i^*,Z_i,\boldsymbol{\theta}_n)$$

is C-tight according to Theorem 1 and the continuity of $\mathbf{A}(\cdot,\boldsymbol{\theta}_0)$, it remains to analyze $(R_n^*)_{n\geq 1}$. In particular, we apply theorem 15.6 of Billingsley (1968) to prove that, with probability 1,

$$R_n^* \longrightarrow B\circ T \tag{.2}$$

in distribution in $D([0,\infty])$, where B denotes the Wiener process and

$$T:[0,\infty]\ni x\longrightarrow T(x)=\int_0^x \operatorname{var}(\delta|Z=t)H(dt)$$

a nondecreasing transformation (see also Stute 1997, thm. 1.1; Stute et al. 1998, lemma A.3).

For the covariance function, we get, with $0\leq s\leq t\leq\infty$,

$$\begin{aligned}\operatorname{cov}_n\big(R_n^*(s),R_n^*(t)\big) &= n^{-1}\sum_{i=1}^{n}m(Z_i,\boldsymbol{\theta}_n)(1-m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{Z_i\leq s\}}\\ &\longrightarrow \int_0^s m(z,\boldsymbol{\theta}_0)(1-m(z,\boldsymbol{\theta}_0))H(dz)\end{aligned}$$

with probability 1, by applying the SLLN and a continuity argument. Thus the covariance function coincides with the one of $B\circ T$.

That the finite-dimensional distributions of R_n^* tend to those of $B\circ T$ with probability 1 can be shown using the same technique as sketched under Lemma A.4. Thus we omit the proof here.

For the final part of the proof, we assume without loss of generality that R_n^* is defined on the unit interval (cf. Stute 1997, p. 637). Now fix $0\leq x_1\leq x\leq x_2\leq 1$ to obtain

$$\begin{aligned}&\mathbb{E}_n\Big(\big(R_n^*(x)-R_n^*(x_1)\big)^2\big(R_n^*(x_2)-R_n^*(x)\big)^2\Big)\\ &\quad= n^{-2}\sum_{i=1}^{n}m(Z_i,\boldsymbol{\theta}_n)(1-m(Z_i,\boldsymbol{\theta}_n))\mathbb{1}_{\{x_1<Z_i\leq x\}}\\ &\qquad\times\left(\sum_{j=1,j\neq i}^{n}m(Z_j,\boldsymbol{\theta}_n)(1-m(Z_j,\boldsymbol{\theta}_n))\mathbb{1}_{\{x<Z_j\leq x_2\}}\right)\\ &\quad\leq 4^{-2}\big(H_n(x)-H_n(x_1)\big)\big(H_n(x_2)-H_n(x)\big).\end{aligned}$$

The Glivenko–Cantelli theorem then yields that, with probability 1,

$$\limsup_{n\to\infty}\mathbb{E}_n\Big(\big(R_n^*(x)-R_n^*(x_1)\big)^2\big(R_n^*(x_2)-R_n^*(x)\big)^2\Big)\leq\big(H(x_2)-H(x_1)\big)^2,$$

which finally completes the proof of Theorem 2.

[Received November 2003. Revised July 2005.]

REFERENCES

Abdushukurov, A. A. (1987), "Nonparametric Estimation in the Proportional Hazards Model of Random Censorship," in *Conference of Young Scientists*, Mathematical Institute Academy of Sciences Uzbek SSR, Tashkent, VINITI No. 8756-V [in Russian].

Aranda-Ordaz, F. J. (1981), "On Two Families of Transformations to Additivity for Binary Response Data," *Biometrika*, 68, 357–363.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Cheng, P. E., and Lin, G. D. (1987), "Maximum Likelihood Estimation of a Survival Function Under the Koziol–Green Proportional Hazards Model," *Statistics and Probability Letters*, 5, 75–80.

Csörgő, S. (1988a), "Testing for the Proportional Hazards Model of Random Censorship," in *Proceedings of the 4th Prague Symposium on Asymptotic Statistics*, eds. P. Mandl and M. Hušková, Prague: Charles University, pp. 41–53.

——— (1988b), "Estimation in the Proportional Hazards Model of Random Censorship," *Statistics*, 19, 437–463.

Dikta, G. (1998), "On Semiparametric Random Censorship Models," *Journal of Statistical Planning and Inference*, 66, 253–279.

Dikta, G., Ghorai, J., and Schmidt, C. (2005), "The Central Limit Theorem Under Semiparametric Random Censorship Models," *Journal of Statistical Planning and Inference*, 127, 23–51.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Henze, N. (1993), "A Quick Omnibus Test for the Proportional Hazards Model of Random Censorship," *Statistics*, 24, 253–263.

Hollander, M., and Proschan, F. (1979), "Testing to Determine the Underlying Distribution Using Randomly Censored Data," *Biometrics*, 35, 393–401.

Hyde, J. (1977), "Testing Survival Under Right Censoring and Left Truncation," *Biometrika*, 64, 225–230.

Jennrich, R. I. (1969), "Asymptotic Properties of Non-Linear Least-Squares Estimators," *The Annals of Mathematical Statistics*, 40, 633–643.

Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation From Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.

Koziol, J. A., and Green, S. B. (1976), "A Cramér–von Mises Statistic for Randomly Censored Data," *Biometrika*, 63, 465–474.

McCullagh, P., and Nelder, J. A. (1999), *Generalized Linear Models* (2nd ed.), London: Chapman & Hall/CRC.

Miller, R., and Halpern, J. (1982), "Regression With Censored Data," *Biometrika*, 69, 521–531.

Stute, W. (1997), "Nonparametric Model Checks for Regression," *The Annals of Statistics*, 25, 613–641.

Stute, W., González Manteiga, W., and Presedo Quindimil, M. (1998), "Bootstrap Approximations in Model Checks for Regression," *Journal of the American Statistical Association*, 93, 141–149.

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *The Annals of Statistics*, 14, 1261–1295.

Zhu, L. X., Yuen, K. C., and Tang, N. Y. (2002), "Resampling Methods for Testing Semiparametric Random Censorship Models," *Scandinavian Journal of Statistics*, 29, 111–123.