Model checks via bootstrap when there are missing binary data.

Gerhard Dikta[a], Sundarraman Subramanian[b], Thorsten Winkler[a]

[a]Department of Medizintechnik und Technomathematik, Fachhochschule Aachen, Germany

[b]Center for Applied Mathematics and Statistics, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, USA

Abstract

Dikta, Kvesic, and Schmidt proposed a model-based resampling scheme to approximate critical values of tests for model checking involving binary response data. Their approach is inapplicable when the binary response variable is not always observed, however. We propose a missingness adjusted marked empirical process under the framework that the missing binary responses are missing at random. We introduce a resampling scheme for the bootstrap and prove its asymptotic validity. We present some numerical comparisons and illustrate our methodology using a real data set.

KEY WORDS: Covariance function, Functional central limit theorem, Gaussian process, Maximum likelihood estimator, Missing information principle, Model-based resampling.

1 Introduction

Binary data of the type (δ, Z), where δ denotes a dichotomous response variable and Z is an explanatory variable, are common in biomedical studies. For example, δ may represent the death status of a study subject and Z the dose of a particular medication. In survival studies, δ would represent a censoring indicator and Z the possibly censored survival time. For other examples, see Cox and Snell (1989) and Collett (2002), among others. We focus on binary data in which δ may be missing for a subset of study subjects. A case in point is the mice data set analyzed by Dinse (1986), in which out of the 33 mice that died with a certain disease present, 8 died due to the disease ($\delta = 1$), 19 from other known causes ($\delta = 0$), and 6 had unknown cause of death (missing δ). Our goal is to implement a procedure for checking whether any candidate model offers an adequate approximation for $m(t) = E(\delta|Z = t)$. Note that this problem has sound rationale, since a semiparametric survival function estimator, which uses a model-based estimate of $m(t)$, has been shown to be more efficient than the Kaplan–Meier estimator whenever the model for $m(t)$ is correctly specified (Dikta, 1998). Koul and Yi (2006) and Aggarwal and Koul (2008) investigated goodness-of-fit testing for "case 1" interval censored data, where, δ, a "current status" indicator, was fully observed.

Let $\sigma = \xi\delta$, where ξ denotes an indicator variable taking the value 0 when δ is missing and is 1 otherwise. The data consist of n iid observations $(\xi_i, \sigma_i, Z_i), i = 1, \ldots, n$, where each $(\xi_i, \sigma_i, Z_i) \in \{0, 1\}^2 \times [0, \infty]$ is distributed like (ξ, σ, Z). The dichotomous δ is assumed to be missing at random (MAR), which is standard in the literature; see Little and Rubin (1987) or Tsiatis, Davidian, and McNeney, (2002), among others. MAR means that the actual value of δ does not influence whether it is missing or not and that the missingness only depends on Z and not on δ. Formally, $P(\xi = 1|Z = t, \delta = d) = P(\xi = 1|Z = t) \equiv \pi(t)$. Let $\boldsymbol{\Theta} \subset I\!R^k$. We assume that there exists a collection $\mathcal{M} := \{m(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ of parametric functions and a unique $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ such that $m(\cdot) = m(\cdot, \boldsymbol{\theta}_0)$. Checking the validity of the parametric assumption amounts to performing a test of hypothesis for $H_0 : m \in \mathcal{M}$ versus $H_1 : m \notin \mathcal{M}$.

An approach often employed for checking model adequacy for a quantity of interest [$m(t)$ in the present case] is to select a suitable function, which determines the quantity, and compare its model-based estimator with that of a completely data-driven (nonparametric) one, using some measure of discrepancy between the two estimators for the comparison. Stute (1997) employed the difference criterion as the discrepancy measure and proposed a marked empirical process (MEP) for testing the goodness of fit of a parametric regression model. One rejects the null hypothesis if the computed value of any test statistic based on the MEP exceeds a critical value, which is typically calibrated from its asymptotic distribution. Since the limit process has a complicated structure, Stute, González Manteiga, and Presedo Quindimil (1998) approximated the limiting distribution using the wild bootstrap.

In the case of binary regression a reasonable choice is $\tilde{H}_1(t) = P(Z \leq t, \delta = 1)$, because it uniquely determines the conditional probability (Stute et al., 1998), and can be consistently estimated as well. Dikta, Kvesic, and Schmidt (2006) studied the MEP for this case and showed that its asymptotic distribution is intractable. They proposed a model-based resampling scheme for regenerating the binary responses and derived a functional central limit theorem for the bootstrap MEP. The basic and bootstrap MEPs converge to the same limiting distribution, forming the basis for calibrating the asymptotic critical values of tests using the bootstrap counterparts; see Zhu, Yuen, and Tang (2002) for related work. However, when δ is not always observed, their procedure is inapplicable because the MEP is not computable; adjustments that account for the missing binary responses are needed. The adjusted MEP that we propose arises through an application of the well-known missing information principle (MIP) and has also been applied to related censored data settings (e.g., Buckley and James, 1978; McKeague, Subramanian, and Sun 2001).

A goodness-of-fit test directly based on the limit distribution of the adjusted MEP has limitations, however. As pointed out by Stute, Thies, and Zhu (1998), and which applies to our set-up as well, the weak limit of the adjusted MEP is not distribution free, in turn affecting the calibration of critical values. For model checking in regression, Stute, Thies, and Zhu (1998) constructed a transformation which produced a distribution free weak limit. Koul and Yi (2005), based their goodness-of-fit testing for "case 1" interval censored data on the MEP approach combined with the Stute-Thies-Zhu transformation to produce an asymptotically distribution free test. As outlined in Stute, Manteiga, and Quindimil (1998), however, a bootstrap approach would also provide asymptotically correct critical values, which we demonstrate in this paper. We propose a modification of the model-based resampling scheme proposed by Dikta et al. (2006), where we regenerate only the observed binary responses and perform repeated complete cases analysis. We derive a functional central limit theorem, which offers the desired justification for using the bootstrap to approximate the critical values of tests. We also present power studies of the proposed tests. Transformed adjusted MEPs producing distribution free weak limits for our set-up would be a worthwhile direction for future research.

The term model-based resampling (Dikta et al., 2006) derives its rationale from the fact that the responses are always (re)generated according to $H_0 : m \in \mathcal{M}$, irrespective of whether or not the original responses were from the null model. This is also exactly the prescription given by Efron and Tibshirani (1993) and ensures that the bootstrap critical values have the correct size asymptotically whether or not the data follow the null model.

Model-based resampling of the non-missing binary responses has connections with multiple imputations (Lu and Tsiatis, 2001; Tsiatis et al. 2002). Both approaches regenerate binary responses via maximum likelihood based on *complete cases*. Whereas the objective of multiple imputations is to achieve *completion* of data by regenerating the *missing* responses, this is not necessary for our approach. In fact, such completion, followed by regeneration of a fresh set of n binary responses, would not improve power of tests (Subramanian, 2011).

The paper is organized as follows. In Section 2, we introduce the adjusted MEP and derive its large sample properties. In Section 3, we propose our modified model-based resampling and prove its asymptotic validity. In Section 4, we report the results of numerical studies and provide an illustration. Proofs of theorems are detailed in the Appendixes.

2 Adjusted marked empirical processes

The nonparametric and model-based estimators of $\tilde{H}_1(t) = P(Z \le t, \delta = 1)$ are given by

$$H_{1,n}^{(\mathrm{NP})}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i \le t, \delta_i = 1), \qquad H_{1,n}^{(\mathrm{SP})}(t) = \frac{1}{n}\sum_{i=1}^{n} m(Z_i, \boldsymbol{\theta}_n) I(Z_i \le t),$$

where $\boldsymbol{\theta}_n$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ (Dikta, 1998). The MEP is the difference between these two estimators, multiplied by $n^{1/2}$ (Dikta et al., 2006):

$$\mathcal{Q}_n(t) \quad = \quad n^{-1/2} \sum_{i=1}^{n} \{\delta_i - m(Z_i, \boldsymbol{\theta}_n)\} \, I(Z_i \leq t). \tag{2.1}$$

Note that the classical MEP $\mathcal{Q}_n(t)$ bears resemblance to the score for $\boldsymbol{\theta}$ in the following way. If all the binary responses δ_i are observed, the log-likelihood for $\boldsymbol{\theta}$, given by

$$\sum_{i=1}^{n} \left[\delta_i \log(m(Z_i, \boldsymbol{\theta})) + (1 - \delta_i) \log(1 - m(Z_i, \boldsymbol{\theta}))\right],$$

yields the corresponding score for $\boldsymbol{\theta}$ given by (Dikta, 1998):

$$\sum_{i=1}^{n} \frac{\delta_i - m(Z_i, \theta)}{m(Z_i, \theta)(1 - m(Z_i, \theta))} \mathrm{Grad}(m(Z_i, \theta)).$$

The i-th summand of $\mathcal{Q}_n(t)$ is seen to be just the centered part of the score, with the $\boldsymbol{\theta}$ replaced by its MLE, and the resulting quantity normalized and converted to a process in t.

However, $\mathcal{Q}_n(t)$ cannot be computed when δ_i is missing, so has to be adjusted. The proposed adjustment incorporates the above mentioned likeness to handle the case when δ_i are MAR. Note that, when the binary responses are MAR, ξ is independent of δ given Z, see, for example, van der Laan and McKeague (1998) or Subramanian (2004). It follows that

$$m(z) \equiv P(\delta = 1 | Z = z) = P(\delta = 1 | \xi = 1, Z = z),$$

suggesting that $\boldsymbol{\theta}$ can be estimated from the "complete cases", that is, those subjects with $\xi = 1$; see also Lu and Tsiatis (2001) or Tsiatis, Davidian, and McNeney (2002), among others. Indeed, the log-likelihood for $\boldsymbol{\theta}$, given by

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\sigma_i \log(m(Z_i, \boldsymbol{\theta})) + (\xi_i - \sigma_i) \log(1 - m(Z_i, \boldsymbol{\theta}))\right],$$

yields the corresponding score for $\boldsymbol{\theta}$, see Subramanian (2004):

$$\sum_{i=1}^{n} \frac{\xi_i(\delta_i - m(Z_i, \boldsymbol{\theta}))}{m(Z_i, \boldsymbol{\theta})(1 - m(Z_i, \boldsymbol{\theta}))} \mathrm{Grad}(m(Z_i, \boldsymbol{\theta})),$$

which, analogous to the complete data case, results in our proposed adjusted MEP given by

$$\mathcal{R}_n(t) \;=\; n^{-1/2}\sum_{i=1}^{n}\xi_i\left(\delta_i - m(Z_i,\hat{\boldsymbol{\theta}})\right)I(Z_i\le t). \tag{2.2}$$

Here $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l_n(\boldsymbol{\theta})$ is the MLE of $\boldsymbol{\theta}$.

Alternatively, because the binary responses are MAR, the random variables δ and $\sigma+(1-\xi)m(Z)$ have the same conditional mean given Z, namely $m(Z)$:

$$\begin{aligned} E[\sigma+(1-\xi)m(Z)|Z] \;&=\; E(\xi|Z)E(\delta|Z)+(1-E(\xi|Z))m(Z)\\ &=\; \pi(Z)m(Z)+(1-\pi(Z))m(Z) \;=\; m(Z). \end{aligned}$$

This suggests that it is reasonable to substitute $m(Z_i,\hat{\boldsymbol{\theta}})$ for $m(Z_i,\boldsymbol{\theta}_n)$ and $\sigma_i+(1-\xi_i)m(Z_i,\hat{\boldsymbol{\theta}})$ for δ_i in $\mathcal{Q}_n(t)$, resulting in our proposed adjusted MEP.

Henceforth we write $l_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} w_i(Z_i,\boldsymbol{\theta})$, where

$$w_i(t,\boldsymbol{\theta}) \;=\; \xi_i\{\delta_i\log(m(t,\boldsymbol{\theta}))+(1-\delta_i)\log(\bar{m}(t,\boldsymbol{\theta}))\}, \tag{2.3}$$

and $\bar{m}(t,\boldsymbol{\theta}) = 1-m(t,\boldsymbol{\theta})$. We also replace $\boldsymbol{\theta}_n$ in Eq. (2.1) with $\hat{\boldsymbol{\theta}}$, giving the adjusted MEP: Write $\mathrm{Grad}(m(u,\boldsymbol{\theta}_0)) = (D_1(m(u,\boldsymbol{\theta})),\ldots,D_k(m(u,\boldsymbol{\theta})))^T|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, where, for each $1\le r\le k$, $D_r(m(u,\boldsymbol{\theta})) = \partial m(u,\boldsymbol{\theta})/\partial\theta_r$. Let $\mathcal{R}_\infty$ be a zero-mean Gaussian process with covariance

$$K_1(s,t) \;=\; \int_0^{s\wedge t}\pi(u)m(u,\boldsymbol{\theta}_0)\bar{m}(u,\boldsymbol{\theta}_0)dH(u) - \int_0^s\int_0^t \pi(u)\pi(v)\beta(u,v)dH(u)dH(v), \tag{2.4}$$

where $\beta(u,v) = (\mathrm{Grad}(m(u,\boldsymbol{\theta}_0)))^T I^{-1}(\boldsymbol{\theta}_0)\mathrm{Grad}(m(v,\boldsymbol{\theta}_0))$, and $I(\boldsymbol{\theta}_0) = (\sigma_{r,s})_{1\le r,s\le k}$, with

$$\sigma_{r,s} \;=\; E\left(\frac{\pi(Z)D_r(m(Z,\boldsymbol{\theta}_0))D_s(m(Z,\boldsymbol{\theta}_0))}{m(Z,\boldsymbol{\theta}_0)\bar{m}(Z,\boldsymbol{\theta}_0)}\right). \tag{2.5}$$

Theorem 1 gives a functional central limit theorem for $\mathcal{R}_n$; its proof is given in the Appendix.

Theorem 1 *Suppose **C1** and **A1**–**A3** hold (cf. Appendix). Then* $\mathcal{R}_n \xrightarrow{\mathcal{D}} \mathcal{R}_\infty$ *in* $D[0,\infty]$.

Write $\hat{H}(t)$ for the empirical estimator of $H(t)$, the distribution function of Z. The KS and CvM statistics are given by

$$\mathcal{D}_n = \sup_{0\leq t\leq\infty} |\mathcal{R}_n(t)|; \qquad \mathcal{W}_n = \int \mathcal{R}_n^2(t) d\hat{H}(t). \tag{2.6}$$

We now deduce from the continuous mapping theorem that $\mathcal{D}_n \xrightarrow{\mathcal{D}} \mathcal{D}_\infty := \sup_{0\leq t\leq\infty} |\mathcal{R}_\infty(t)|$, with a corresponding result holding for $\mathcal{W}_n$. Because the covariance of $\mathcal{D}_\infty$ is intractable, we next introduce a model-based resampling scheme to approximate the critical values of $\mathcal{D}_n$.

3 Model-based resampling and asymptotic validity

We obtain $(\xi_1^*, \sigma_1^*, Z_1^*), \ldots, (\xi_n^*, \sigma_n^*, Z_n^*)$ through our model-based resampling as follows:

1. For each $i = 1, \ldots, n$, set $\xi_i^* = \xi_i$ and $Z_i^* = Z_i$.

2. For each $i = 1, \ldots, n$ such that $\xi_i = 1$, regenerate δ_i from the Bernoulli distribution with success probability $m(Z_i, \hat{\boldsymbol{\theta}})$, and call it δ_i^*. Thus $\sigma_i^* = \xi_i^* \delta_i^*$, which is 0 if $\xi_i^* = \xi_i = 0$.

Note that regeneration of δ is carried out only for complete cases, $\xi = 1$. Also, as in Dikta et al. (2006), the Z_i are not resampled. Writing [see (2.3)]

$$w_i^*(t, \boldsymbol{\theta}) = \xi_i^* \{\delta_i^* \log(m(t, \boldsymbol{\theta})) + (1 - \delta_i^*) \log(\bar{m}(t, \boldsymbol{\theta}))\},$$

the bootstrap MLE $\hat{\boldsymbol{\theta}}^*$ solves the equation $\mathrm{Grad}(l_n^*(\boldsymbol{\theta})) = 0$, where $l_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n w_i^*(Z_i^*, \boldsymbol{\theta})$ is the normalized log-likelihood function. Let $I\!P_n, I\!E_n, \mathrm{Var}_n$, and Cov_n denote the probability measure, expectation, variance, and covariance associated with the bootstrap sample. The proof of our first bootstrap result stated in Theorem 2 below is given in the Appendix.

Theorem 2 *Suppose that $\boldsymbol{\Theta}$ is a connected open subset of $I\!R^k$ and that H is continuous. Under H_0, and assumptions **C1**, **C2**, **A1**, and **A2**, given in the Appendix, when the model-based resampling scheme is used to generate the bootstrap data, $n^{1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})$ has asymptotically* *a k-variate normal distribution* $\mathcal{N}_k(\mathbf{0}, I^{-1}(\boldsymbol{\theta}_0))$ *with probability 1. Also, with probability 1,*

$$n^{1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}) \;=\; I^{-1}(\boldsymbol{\theta}_0) n^{-1/2} \sum_{i=1}^{n} \mathrm{Grad}(w_i^*(Z_i, \hat{\boldsymbol{\theta}})) + o_{\mathbb{P}_n}(1). \tag{3.1}$$

Writing $\boldsymbol{a}^{\otimes 2}$ for $\boldsymbol{a}\boldsymbol{a}^T$, and defining $I_n = \mathbb{E}_n \left(n^{-1/2} \sum_{i=1}^n \mathrm{Grad}(w_i^*(Z_i, \hat{\boldsymbol{\theta}})) \right)^{\otimes 2}$, we remark that

$$\begin{aligned} I_n &= \mathbb{E}_n \left(\frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{\xi_i(\delta_i^* - m(Z_i, \hat{\boldsymbol{\theta}}))\mathrm{Grad}(m(Z_i, \hat{\boldsymbol{\theta}}))}{m(Z_i, \hat{\boldsymbol{\theta}})\bar{m}(Z_i, \hat{\boldsymbol{\theta}})} \right)^{\otimes 2} \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{\xi_i \mathrm{Grad}(m(Z_i, \hat{\boldsymbol{\theta}})) \left(\mathrm{Grad}(m(Z_i, \hat{\boldsymbol{\theta}})) \right)^T}{m(Z_i, \hat{\boldsymbol{\theta}})\bar{m}(Z_i, \hat{\boldsymbol{\theta}})} \xrightarrow{\text{a.s.}} I(\boldsymbol{\theta}_0), \end{aligned}$$

by the strong law of large numbers, conditions **C1** and **A1**, and a continuity argument which allows $\hat{\boldsymbol{\theta}}$ to be replaced with $\boldsymbol{\theta}_0$, see Lemma A.1. of Dikta et al., (2006).

The bootstrap MEP is now defined from Eq. (2.2), but using the resampled data, by

$$\mathcal{R}_n^*(t) = n^{-1/2} \sum_{i=1}^{n} \xi_i^* \left(\delta_i^* - m(Z_i^*, \hat{\boldsymbol{\theta}}^*) \right) I(Z_i^* \le t), \qquad 0 \le t \le \infty.$$

Writing $\hat{H}^*(t)$ for the bootstrap version of $\hat{H}(t)$, which, due to our resampling mechanism, is identically equal to $\hat{H}(t)$, the bootstrap KS and CvM statistics are defined as

$$\mathcal{D}_n^* = \sup_{0 \le t \le \infty} |\mathcal{R}_n^*(t)|; \qquad \mathcal{W}_n^* = \int \left(\mathcal{R}_n^*(t) \right)^2 d\hat{H}^*(t).$$

We now state a functional central limit theorem for $\mathcal{R}_n^*$. Proof is given in the Appendix.

Theorem 3 *Suppose that* $\boldsymbol{\Theta}$ *is a connected open subset of* $\mathbb{R}^k$ *and* H *is continuous. Under* H_0 *and assumptions **C1**, **C2**, **A1**–**A3** (cf. Appendix), when model-based resampling is used to generate the bootstrap data, the process* $\mathcal{R}_n^*$ *converges weakly in* $D[0, \infty]$ *with probability 1 to the zero-mean Gaussian process* $\mathcal{R}_\infty$ *with covariance function* $K_1(s, t)$ *given by Eq. (2.4).*

Remark For the analysis of the bootstrap MEPs under the alternative hypothesis we will need a proper interpretation of $\boldsymbol{\theta}_0$. To guarantee that $m(\cdot, \boldsymbol{\theta}_0)$ is the projection of $m(\cdot)$ onto

$\mathcal{M}$ with respect to the Kullback–Leibler geometry, we will need the assumption that there exists a unique parameter $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ that maximizes

$$E\left[\pi(Z)\left\{m(Z)\log(m(Z,\boldsymbol{\theta})) + (1-m(Z))\log(1-m(Z,\boldsymbol{\theta}))\right\}\right].$$

As noted in the Introduction, the resampled data are always generated under the null hypothesis even if the original data follow some alternative. This feature of the model-based resampling scheme allows us to mimic the proof of Theorem 3 even under alternatives.

4 Numerical results

4.1 A power study

A single sample constituted the 100 triplets $(\xi_i, \sigma_i, Z_i)_{1\le i\le 100}$. For each chosen sample of size 100, we calculated $\mathcal{R}_n$ given by Eq. (2.2). We will refer to it henceforth as MIP-MEP. For comparison, we also calculated an augmented inverse probability of nonmissingness weighted MEP, henceforth referred as AIPW-MEP, given by (cf. Subramanian 2010)

$$\mathcal{E}_n(t) \quad = \quad n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i}{\hat{\pi}(Z_i)}(\delta_i - m(Z_i,\hat{\boldsymbol{\theta}}))I(Z_i \le t), \tag{4.1}$$

where $\hat{\pi}(x)$ is a kernel estimator of $P(\xi = 1|Z = x)$, based on bandwidth 0.05. This bandwidth provided the best result when compared with 0.01 and 0.1. We then calculated $\mathcal{D}_n$ and $\mathcal{W}_n$ defined by Eq. (2.6). To approximate the distribution of $\mathcal{D}_n$ and $\mathcal{W}_n$ we employed the model-based resampling described in Section 3 above. We obtained 1000 values (of MIP-MEP and AIPW-MEP based) $\mathcal{D}_n^*$ and $\mathcal{W}_n^*$ from $nboot$ = 1000 bootstrap replications. The proportion exceeding $\mathcal{D}_n$ and $\mathcal{W}_n$ yields *a single bootstrap p-value for a single sample of size 100.* Repeating the entire procedure above over $nsim = 1000$ replications, the empirical power is the proportion of 1000 bootstrap p-values which fell below the nominal 5%.

We considered the generalized proportional hazards model (GPHM) $m(t,\boldsymbol{\theta}) = \theta_1/(\theta_1+t^{\theta_2})$, where $\theta_1 > 0$ and $\theta_2 \in I\!R$, which arises when the failure and censoring distributions are

Weibull: $F(t) = 1 - \exp(-(at)^b)$ and $G(t) = 1 - \exp(-(ct)^d)$, with $\theta_1 = ba^b/(dc^d)$ and $\theta_2 = d - b$. We fixed $(a, b, c) = (2.0, 0.7, 0.9)$ and varied d over a grid of values between 0.2 and 1.35. When $d = 0.7$, the GPHM reduces to the simple proportional hazards model (SPHM) $m(t, \theta_1) = \theta_1/(\theta_1 + 1)$. We introduced misspecification of $m(t)$ by always fitting the SPHM to the generated data. The MLE is $\hat{m}(t) = \sum_{i=1}^n \sigma_i / \sum_{i=1}^n \xi_i$. We used $\pi(t, \boldsymbol{\eta}) = \exp(\eta_1 + \eta_2 t)/(1 + \exp(\eta_1 + \eta_2 t))$, with $(\eta_1, \eta_2) = (0.2, 0.5)$, giving a missingness rate (MR) of about 40%. All tests achieve the nominal level of 5% when there is no misspecification ($d = 0.7$), see Figure 1.

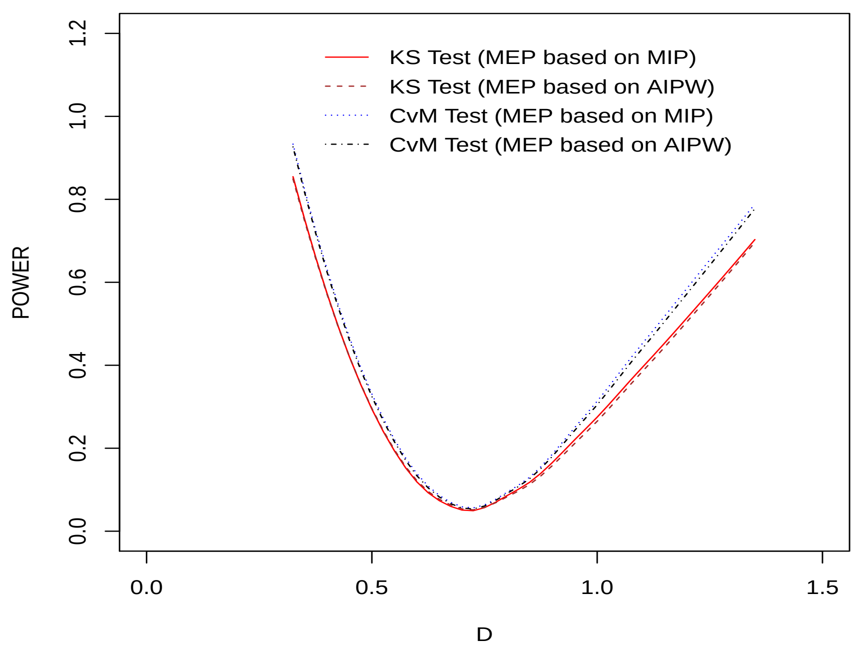

Figure 1: Empirical power of proposed tests.

The empirical distribution function of the 1,000 p-values corresponding to the proposed tests for $(a, b, c, d) = (2.0, 0.7, 0.9, 1.25)$ are displayed in Figure 2. It is seen that both the MIP and AIPW versions of the CvM test perform better than their KS test counterparts.

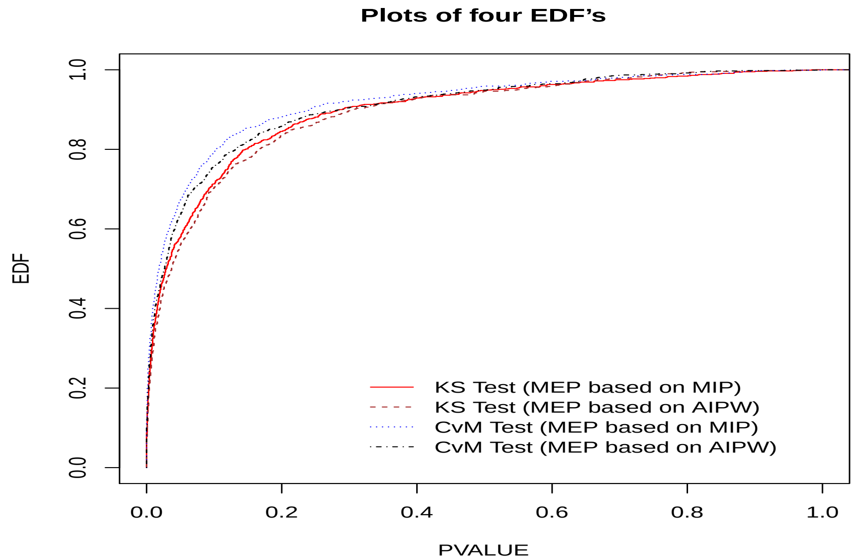

Figure 2: Empirical distribution function of 1,000 bootstrap p-values.

4.2 Illustration using a mice data set

Using our proposed procedures, we tested the adequacy of two models for the mice data reported in the introduction section. They were the three-parameter logit model given by $\operatorname{logit}(m(x,\boldsymbol{\theta})) = \theta_1 + \theta_2 x + \theta_3 x^2$ and the three-parameter probit model given by $m(x,\boldsymbol{\theta}) = \Phi(\theta_1 + \theta_2 x + \theta_3 x^2)$, where $\operatorname{logit}(\varphi) = \log(\varphi/(1-\varphi))$ and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. The fitted three-parameter probit model was $m(x,\hat{\boldsymbol{\theta}}) = \Phi(16.0022 - 17.7897x + 4.844x^2)$. The bootstrap p-value of the KS test based on the MIP-MEP was 0.653. The fitted three-parameter logit model was $\operatorname{logit}(m(x,\hat{\boldsymbol{\theta}})) = -27.9041 + 30.8175x - 8.3415x^2$. The corresponding bootstrap p-value of the KS test based on the MIP-MEP was 0.722. In both cases the null hypothesis cannot be rejected and we conclude that they both offer adequate fits to the mice data, see also Figure 3 below.

Appendix

Write $D_{r,s}(\cdot)$ for the partial derivatives of second order. We need the following conditions:

(**C1**) There exists a measurable solution $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ of $\operatorname{Grad}(l_n(\boldsymbol{\theta})) = 0$, such that $\hat{\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$.

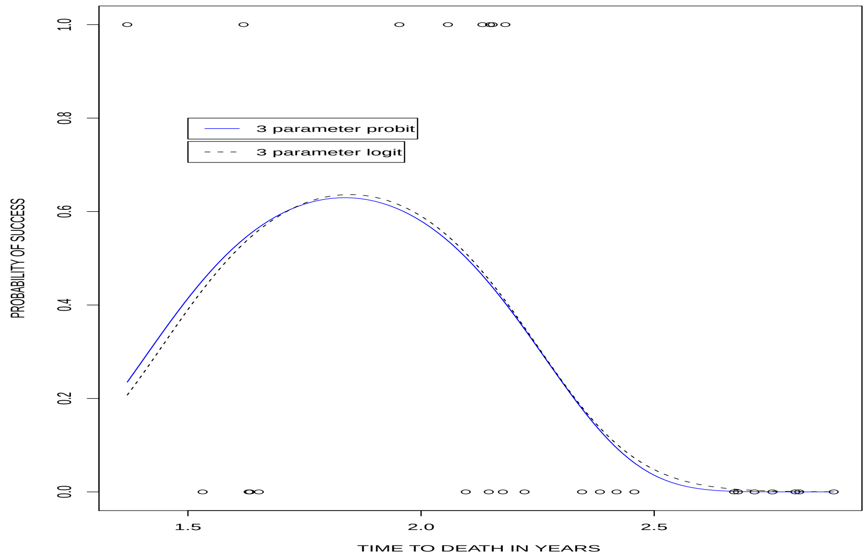

Figure 3: Scatter plot of the nonmissing binary responses and plots of the two fitted models of $m(t)$ for the mice data.

(**C2**) For $I\!\!P$ almost all sample sequences there exists $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, a measurable solution of the equation of $\text{Grad}(l_n^*(\boldsymbol{\theta})) = 0$, such that $I\!\!P_n(|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0| > \epsilon) \longrightarrow 0$ for every $\epsilon > 0$.

(**A1**) The functions $p(t, \boldsymbol{\theta}) = \log(m(t, \boldsymbol{\theta}))$ and $\bar{p}(t, \boldsymbol{\theta}) = \log(\bar{m}(t, \boldsymbol{\theta}))$ have continuous derivatives of second order with respect to $\boldsymbol{\theta}$ at each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $t \geq 0$. Also, the functions $D_r(p(\cdot, \theta))$, $D_r(\bar{p}(\cdot, \theta))$, $D_{r,s}(p(\cdot, \theta))$ and $D_{r,s}(\bar{p}(\cdot, \theta))$ are measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and there exists a neighborhood of $\boldsymbol{\theta}_0$, $V(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ and a measurable square integrable function M [that is, $E(M^2(Z)) < \infty$] such that for all $\theta \in V(\boldsymbol{\theta}_0)$, $t \geq 0$, and $1 \leq r, s \leq k$,

$$|D_{r,s}(p(t, \theta))| + |D_{r,s}(\bar{p}(t, \theta))| + |D_r(p(t, \theta))| + |D_r(\bar{p}(t, \theta))| \leq M(t).$$

(**A2**) The matrix $I(\boldsymbol{\theta}_0)$, whose elements are defined by Eq. (2.5), is positive definite.

(**A3**) The function $m(t, \boldsymbol{\theta})$ is continuously differentiable at each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$; there exists a function N such that for $t \geq 0$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\|\text{Grad}(m(t, \boldsymbol{\theta}))\| \leq N(t)$ and $E(N(Z)) < \infty$.

All the conditions above were given by Dikta (1998) and Dikta et al. (2006). More specifically, conditions **C1** and **C2** can be proved by the methods introduced by Stute (1992).

Sufficient conditions for the strong consistency of the MLE under complete observations are given in Theorem 2.1 and Corollary 2.2 of Dikta (1998). The proof there was based on the ideas presented in Perlman (1972), which can be adapted to our set-up with minor modifications. Moreover, sufficient conditions for **C1** can be derived from proper adaptations of the proofs for strong consistency of the MLE. Since there are different approaches available, we do not wish to insist on a particular sufficient condition for **C1** in our paper.

Since $\hat{\boldsymbol{\theta}}$ is derived via maximum likelihood, we have

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad = \quad I^{-1}(\boldsymbol{\theta}_0) n^{-1/2} \sum_{i=1}^{n} \mathrm{Grad}(w_i(Z_i, \boldsymbol{\theta}_0)) + o_{I\!P}(1), \tag{A.1}$$

see Subramanian (2004) or Dikta (1998). The influence function of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is given by

$$\mathcal{L}(\xi, \sigma, Z, \boldsymbol{\theta}_0) := I^{-1}(\boldsymbol{\theta}_0)\mathrm{Grad}(w(Z, \boldsymbol{\theta}_0)) = \frac{\xi(\delta - m(Z, \boldsymbol{\theta}_0))}{m(Z, \boldsymbol{\theta}_0)\bar{m}(Z, \boldsymbol{\theta}_0)} I^{-1}(\boldsymbol{\theta}_0)\mathrm{Grad}(m(Z, \boldsymbol{\theta}_0)), \tag{A.2}$$

which is centered. Note that $E(\mathcal{L}(\xi, \sigma, Z, \boldsymbol{\theta}_0)\mathcal{L}^T(\xi, \sigma, Z, \boldsymbol{\theta}_0)) = I^{-1}(\boldsymbol{\theta}_0)$ (Subramanian, 2004).

Proof of Theorem 1 By Eqs. (A.1) and (A.2), and by condition **A3**, assumptions 1 and 2 of Stute (1997) are fulfilled. Therefore, although $\mathcal{R}_n(t)$ includes the additional binary random variable ξ_i, we can still follow the steps in the proof of Theorem 1.2 of Stute (1997) exactly and conclude that uniformly in t, $\mathcal{R}_n(t) = \mathcal{R}_{n,1}(t) + \mathcal{R}_{n,2}(t) + o_p(1)$, where

$$\begin{aligned}
\mathcal{R}_{n,1}(t) \quad &= \quad n^{-1/2} \sum_{i=1}^{n} \{\xi_i(\delta_i - m(Z_i, \boldsymbol{\theta}_0))\} \, I(Z_i \le t), \\
\mathcal{R}_{n,2}(t) \quad &= \quad -n^{-1/2} \sum_{i=1}^{n} \int_0^t \pi(s)(\mathrm{Grad}(m(s, \boldsymbol{\theta}_0)))^T I^{-1}(\boldsymbol{\theta}_0) \\
&\qquad\qquad \times \frac{\xi_i(\delta_i - m(Z_i, \boldsymbol{\theta}_0))}{m(Z_i, \boldsymbol{\theta}_0)\bar{m}(Z_i, \theta_0)} \mathrm{Grad}(m(Z_i, \boldsymbol{\theta}_0)) dH(s) \\
&= \quad -n^{-1/2} \sum_{i=1}^{n} \left(\int_0^t \pi(s)(\mathrm{Grad}(m(s, \boldsymbol{\theta}_0)))^T dH(s) \right) \mathcal{L}(\xi_i, \sigma_i, Z_i, \boldsymbol{\theta}_0) \\
&= \quad -n^{-1/2} \alpha^T(t, \boldsymbol{\theta}_0) \sum_{i=1}^{n} \mathcal{L}(\xi_i, \sigma_i, Z_i, \boldsymbol{\theta}_0),
\end{aligned}$$

see Eq. (A.2). Convergence of finite dimensional distributions follows by the central limit theorem. Tightness follows exactly as in the proof of Theorem 1.1 of Stute (1997). Therefore, $\mathcal{R}_n \xrightarrow{\mathcal{D}} \mathcal{R}_\infty$ in $D[0,\infty]$. We can readily obtain $K_1(s,t)$ given by Eq. (2.4) from the following:

$$\begin{aligned}
\mathrm{Cov}(\mathcal{R}_{n,1}(s), \mathcal{R}_{n,1}(t)) &= \int_0^{s\wedge t} \pi(u) m(u, \boldsymbol{\theta}_0)\bar{m}(u,\boldsymbol{\theta}_0)\, dH(u), \\
\mathrm{Cov}(\mathcal{R}_{n,2}(s), \mathcal{R}_{n,2}(t)) &= \int_0^{s}\int_0^{t} \pi(u)\pi(v)\beta(u,v)\, dH(u)dH(v), \\
\mathrm{Cov}(\mathcal{R}_{n,1}(s), \mathcal{R}_{n,2}(t)) &= \mathrm{Cov}(\mathcal{R}_{n,2}(s), \mathcal{R}_{n,1}(t)) \;=\; -\,\mathrm{Cov}(\mathcal{R}_{n,2}(s), \mathcal{R}_{n,2}(t)). \qquad \square
\end{aligned}$$

Proof of Theorem 2 Taylor expansion of $\mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}}^*))$ about $\hat{\boldsymbol{\theta}}$ yields with probability 1

$$0 = \mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}})) + A_n^*(\tilde{\boldsymbol{\theta}}^*)(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}),$$

where $\tilde{\boldsymbol{\theta}}^*$ lies on the line segment joining $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}$, and $A_n^*(\boldsymbol{\theta}) = (a_{r,s}^{n*}(\boldsymbol{\theta}))_{1\le r,s\le k}$ is a $k\times k$ matrix with $a_{r,s}^{n*}(\boldsymbol{\theta}) = D_{r,s}(l_n^*(\boldsymbol{\theta}))$, for $1 \le r, s \le k$. By a straightforward adaptation of Lemma A.1 of Dikta et al. (2006), we have that $A_n^*(\hat{\boldsymbol{\theta}}^*) = -I(\boldsymbol{\theta}_0) + o_{\mathbb{P}_n}(1)$, from which Eq. (3.1) is immediate. To prove the asymptotic normality of $n^{1/2}\mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}})) = n^{-1/2}\sum_{i=1}^n \mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))$, we calculate

$$\mathrm{Var}_n\left(n^{-1/2}\sum_{i=1}^n \boldsymbol{a}^T \mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}_n\left(\sum_{r=1}^k a_r D_r(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))\right) \;\equiv\; \frac{1}{n}\sum_{i=1}^n \mathcal{V}_n.$$

Note that for each $i = 1, \ldots, n$

$$\begin{aligned}
\mathcal{V}_n &= \mathrm{Var}_n\left(\sum_{r=1}^k a_r \frac{\xi_i^*(\delta_i^* - m(Z_i,\hat{\boldsymbol{\theta}}))}{m(Z_i,\hat{\boldsymbol{\theta}})\bar{m}(Z_i,\hat{\boldsymbol{\theta}})} D_r(m(Z_i,\hat{\boldsymbol{\theta}}))\right) \\
&= \sum_{r=1}^k\sum_{s=1}^k a_r a_s \mathbb{E}_n\left(\frac{\xi_i^*(\delta_i^* - m(Z_i,\hat{\boldsymbol{\theta}}))^2 D_r(m(Z_i,\hat{\boldsymbol{\theta}}))D_s(m(Z_i,\hat{\boldsymbol{\theta}}))}{m^2(Z_i,\hat{\boldsymbol{\theta}})\bar{m}^2(Z_i,\hat{\boldsymbol{\theta}})}\right) \\
&= \sum_{r=1}^k\sum_{s=1}^k a_r a_s \frac{\xi_i m(Z_i,\hat{\boldsymbol{\theta}})\bar{m}(Z_i,\hat{\boldsymbol{\theta}}) D_r(m(Z_i,\hat{\boldsymbol{\theta}}))D_s(m(Z_i,\hat{\boldsymbol{\theta}}))}{m^2(Z_i,\hat{\boldsymbol{\theta}})\bar{m}^2(Z_i,\hat{\boldsymbol{\theta}})} \\
&= \sum_{r=1}^k\sum_{s=1}^k a_r a_s \frac{\xi_i D_r(m(Z_i,\hat{\boldsymbol{\theta}}))D_s(m(Z_i,\hat{\boldsymbol{\theta}}))}{m(Z_i,\hat{\boldsymbol{\theta}})\bar{m}(Z_i,\hat{\boldsymbol{\theta}})}.
\end{aligned}$$

From the above calculations, by the strong law of large numbers and a continuity argument (see Lemma A.1 of Dikta et al., 2006), we conclude that, with probability 1,

$$\mathrm{Var}_n\left(n^{-1/2}\sum_{i=1}^{n}\boldsymbol{a}^T\mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))\right)\longrightarrow \boldsymbol{a}^T I(\boldsymbol{\theta}_0)\boldsymbol{a}.$$

To verify Lindeberg's condition, we need to show that

$$T_n(\epsilon)=\sum_{i=1}^{n}\mathbb{E}_n\left[\left(n^{-1/2}\boldsymbol{a}^T\mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))\right)^2 I(|\boldsymbol{a}^T\mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))|>n^{1/2}\epsilon)\right]\rightarrow 0$$

with probability 1 as $n\rightarrow\infty$. Write $U_i(\hat{\boldsymbol{\theta}})=\xi_i\boldsymbol{a}^T\mathrm{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))/(m(Z_i,\hat{\boldsymbol{\theta}})\bar{m}(Z_i,\hat{\boldsymbol{\theta}}))$, and note that $\left(\boldsymbol{a}^T\mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))\right)^2=U_i^2(\hat{\boldsymbol{\theta}})(\delta_i^*-m(Z_i,\hat{\boldsymbol{\theta}}))^2$. Also,

$$\boldsymbol{a}^T\mathrm{Grad}(w_i^*(Z_i,\hat{\boldsymbol{\theta}}))=\begin{cases}\xi_i\boldsymbol{a}^T\mathrm{Grad}(p(Z_i,\hat{\boldsymbol{\theta}})) & \text{with probability } m(Z_i,\hat{\boldsymbol{\theta}})\\ \xi_i\boldsymbol{a}^T\mathrm{Grad}(\bar{p}(Z_i,\hat{\boldsymbol{\theta}})) & \text{with probability } \bar{m}(Z_i,\hat{\boldsymbol{\theta}})\end{cases}$$

It follows that

$$\begin{aligned}T_n(\epsilon) &= \frac{1}{n}\sum_{i=1}^{n}\Big[U_i^2(\hat{\boldsymbol{\theta}})\bar{m}^2(Z_i,\hat{\boldsymbol{\theta}})I\left(|\xi_i\boldsymbol{a}^T\mathrm{Grad}(p(Z_i,\hat{\boldsymbol{\theta}}))|>n^{1/2}\epsilon\right)\cdot m(Z_i,\hat{\boldsymbol{\theta}})\\ &\quad + U_i^2(\hat{\boldsymbol{\theta}})(-m(Z_i,\hat{\boldsymbol{\theta}}))^2 I\left(|\xi_i\boldsymbol{a}^T\mathrm{Grad}(\bar{p}(Z_i,\hat{\boldsymbol{\theta}}))|>n^{1/2}\epsilon\right)\cdot\bar{m}(Z_i,\hat{\boldsymbol{\theta}})\Big].\end{aligned}$$

Follow the steps at the top of page 529 of Dikta et al. (2006) to conclude that

$$T_n(\epsilon)\leq\frac{2}{n}\sum_{i=1}^{n}\sum_{r=1}^{k}\sum_{s=1}^{k}|a_r||a_s|M^2(Z_i)I(\|\boldsymbol{a}\|k^{1/2}M(Z_i)>n^{1/2}\epsilon).$$

Fix A. For properly chosen c, by strong law of large numbers we obtain with probability 1,

$$\limsup_{n\rightarrow\infty}T_n(\epsilon)\leq c\mathbb{E}\left(M^2(Z)I(M(Z)>A)\right).$$

By assumption **A1**, the right hand side term above tends to 0 as $A\rightarrow\infty$. This verifies Lindeberg's condition and we deduce from the Cramér–Wold device that $n^{1/2}\mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}}))$ is asymptotically normal. □

Proof of Theorem 3 Writing $\alpha(t,\boldsymbol{\theta}_0) = \int_0^t \pi(u)\text{Grad}(m(u,\boldsymbol{\theta}_0)dH(u)$, we first note that the second term of Eq. (2.4) is $\alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\alpha(t,\boldsymbol{\theta}_0)$. We write $\mathcal{R}_n^*(t) = \mathcal{R}_{n,1}^*(t) + \mathcal{R}_{n,2}^*(t)$, where

$$\begin{aligned}
\mathcal{R}_{n,1}^*(t) &= n^{-1/2}\sum_{i=1}^n \left\{\xi_i(\delta_i^* - m(Z_i,\hat{\boldsymbol{\theta}}))\right\} I(Z_i \le t),\\
\mathcal{R}_{n,2}^*(t) &= -n^{-1/2}\sum_{i=1}^n \xi_i \left(m(Z_i,\hat{\boldsymbol{\theta}}^*) - m(Z_i,\hat{\boldsymbol{\theta}})\right) I(Z_i \le t).
\end{aligned}$$

Note that $\mathcal{R}_{n,1}^*(t)$ is centered since $I\!\!E_n(\delta_i^*) = m(Z_i,\hat{\boldsymbol{\theta}})$. For $1 \le i \le n$, let $\tilde{\boldsymbol{\theta}}_{n,i}^*$ denote points on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$. A Taylor expansion of $m(Z_i,\hat{\boldsymbol{\theta}}^*)$ about $\hat{\boldsymbol{\theta}}$ yields

$$\mathcal{R}_{n,2}^*(t) = -n^{-1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})^T \frac{1}{n}\sum_{i=1}^n \xi_i \text{Grad}(m(Z_i,\tilde{\boldsymbol{\theta}}_{n,i}^*))I(Z_i \le t).$$

The assumptions and Theorem 2 imply that, with probability 1, $\tilde{\boldsymbol{\theta}}_{n,i}^*$ may each be replaced by $\boldsymbol{\theta}_0$ with remainder term $o_{I\!\!P_n}(1)$. It follows from Theorem 2 of Jennrich (1969) that

$$\mathcal{R}_{n,2}^*(t) = -n^{-1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})^T\alpha(t,\boldsymbol{\theta}_0) + o_{I\!\!P_n}(1).$$

Defining $K_1^*(s,t) = I\!\!E_n\left[\left(\mathcal{R}_{n,1}^*(s) + \mathcal{R}_{n,2}^*(s)\right)\left(\mathcal{R}_{n,1}^*(t) + \mathcal{R}_{n,2}^*(t)\right)\right]$, we can show that

$$\begin{aligned}
K_1^*(s,t) &= \frac{1}{n}\sum_{i=1}^n \xi_i\, m(Z_i,\hat{\boldsymbol{\theta}})\bar{m}(Z_i,\hat{\boldsymbol{\theta}})I(Z_i \le s\wedge t) + \alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)I_n I^{-1}(\boldsymbol{\theta}_0)\alpha(t,\boldsymbol{\theta}_0)\\
&\quad - \alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{n}\sum_{i=1}^n \xi_i\text{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))I(Z_i \le t)\right)\\
&\quad - \alpha^T(t,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{n}\sum_{i=1}^n \xi_i\text{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))I(Z_i \le s)\right).
\end{aligned}$$

By the strong law of large numbers, assumptions **C1** and **A3**, and a continuity argument as in Lemma A.1. of Dikta et al. (2006), we have with probability 1 that, for every $0 \le s \le \infty$,

$$n^{-1}\sum_{i=1}^n \xi_i\text{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))I(Z_i \le s) \longrightarrow \alpha(s,\boldsymbol{\theta}_0).$$

Note that $I_n \to I(\boldsymbol{\theta}_0)$ with probability 1. It follows that, with probability 1,

$$K_1^*(s,t) \quad\longrightarrow\quad \int_0^{s\wedge t} \pi(u)m(u,\boldsymbol{\theta}_0)\bar{m}(u,\boldsymbol{\theta}_0)dH(u) - \alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\alpha(t,\boldsymbol{\theta}_0),$$

which is the same as Eq. (2.4). Therefore $K_1^*(s,t)$ tends with probability 1 to $K_1(s,t)$. Applying the Cramér–Wold device and verifying the corresponding Lindeberg condition, we can show that the finite dimensional distribution of $\mathcal{R}_n^*$ converges to that of $\mathcal{R}_\infty$ with probability 1. Furthermore, $\mathcal{R}_{n,2}^*$ induces a tight sequence of distributions on $D[0,\infty]$. See the proofs of Lemma A.4 and Theorem 2 of Dikta et al. (2006) for the above assertions. To see that $\mathcal{R}_{n,1}^*$ also induces a tight sequence of distribution on $D[0,\infty]$, we follow the method given in the final part of the proof of Theorem 2 of Dikta et al. (2006), concluding that

$$\limsup_{n\to\infty} \mathbb{E}_n \left\{ \left(\mathcal{R}_{n,1}^*(z) - \mathcal{R}_{n,1}^*(z_1)\right)^2 \left(\mathcal{R}_{n,1}^*(z_2) - \mathcal{R}_{n,1}^*(z)\right)^2 \right\} \le \left(H(z_2) - H(z_1)\right)^2,$$

and complete the proof by appealing to Theorem 15.6 of Billingsley (1968). □

References

Aggarwal, D. and Koul, H. L. (2008) Minimum empirical distance goodness-of-fit tests for current status data. *J. Indian Statist. Assoc.* **46** 79–124.

Billingsley, P. (1968). *Convergence of probability measures*, Wiley, New York.

Buckley, J. J. and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66** 429–36.

Collett, D. (2002). *Modelling binary data.* Chapman and Hall/CRC, Boca Raton, Florida.

Cox, D. R. and Snell, E.J. (1989). *Analysis of binary data.* Chapman and Hall, London.

Dikta, G. (1998). On semiparametric random censorship models. *J. Statist. Plann. Inference* **66** 253–279.

Dikta, G., Kvesic, M., and Schmidt, C. (2006). Bootstrap approximations in model checks for binary data. *J. Amer. Statist. Assoc.* **101** 521–530.

Dinse, G.E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. *J. Amer. Statist. Assoc.* **81** 328–336.

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

Jennrich, R. I. (1969). Asymptotic Properties of Non-Linear Least-Squares Estimators. *Ann. Statist.* **40** 633–643.

Koul, H. L. and Yi, T. (2006). Goodness-of-fit testing in interval censoring case 1. *Statist. Probab. Lett.* **76** 709-718.

van der Laan, M. J., and McKeague, I. W. (1998). Efficient estimation from right-censored data when failure indicators are missing at random. *Ann. Statist.* **26** 164–182.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis With Missing Data.* Wiley, New York.

Lu K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in proportional hazards models with missing cause of failure. *Biometrics* **57** 1191–1197.

McKeague, I. W., Subramanian, S., and Sun, Y. (2001). Median regression and the missing information principle. *J. Nonparametr. Statist.* **13** 709–727.

Perlman, M. D. (1972). On the strong consistency of approximate maximum likelihood estimates. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, University of California Press, Berkeley, CA.

Stute, W. (1992). Strong consistency of the MLE under random censoring. *Metrika* **39** 257–267.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641.

Stute, W., González Manteiga, W., and Presedo Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149.

Stute, W., Thies, S., and Zhu, L.-X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26** 1916-1934.

Subramanian, S. (2004). The missing censoring-indicator model of random censorship. *Hand-*

book of Statistics 23: Advances in Survival Analysis. Eds. N. Balakrishnan and C.R.Rao. 123–141

Subramanian, S. (2010). Some adjusted marked empirical processes for model checking with missing binary data. CAMS Technical Report, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey, USA.

Subramanian, S. (2011). Multiple imputations and the missing censoring indicator model. *J. Multivar. Anal.* **102** 105–117.

Tsiatis, A. A., Davidian, M., and McNeney, B. (2002). Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. *Biometrika* **89** 238–244.

Zhu, L. X, Yuen, K. C., and Tang, N. Y. (2002). Resampling methods for testing random censorship models. *Scand. J. Statist.* **29** 111–123.