Statistics and Probability Letters 83 (2013) 219–226

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bootstrap based model checks with missing binary response 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

ARTICLE INFO

Article history:
Received 18 July 2011
Received in revised form 11 September 2012
Accepted 11 September 2012
Available online 17 September 2012

Keywords:
Covariance function
Functional central limit theorem
Maximum likelihood estimator
Continuous mapping theorem
Wild bootstrap

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.

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), 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).

Let $\sigma = \xi\delta$, where ξ denotes an indicator variable taking the value 0 when δ is missing and 1 otherwise. The data consist of n i.i.d. 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 van der Laan and McKeague (1998), 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 $\boldsymbol{\Theta} \subset \mathbb{R}^k$. We assume that there exist 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 $H_1 : m \notin \mathcal{M}$.

Stute (1997) proposed a marked empirical process (MEP) for testing the goodness of fit of a parametric regression model. Stute et al. (1998a) approximated the MEP's limiting distribution using wild bootstrap. Stute et al. (1998b), on the other

* Corresponding author. Tel.: +1 973 642 4496.
E-mail address: sundars@njit.edu (S. Subramanian).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2012.09.014

hand, constructed a transformation which produced a distribution free weak limit. Koul and Yi (2006) and Aggarwal and Koul (2008) applied the latter approach for goodness-of-fit testing for "case 1" interval censored data, where, δ, a "current status" indicator, was fully observed. Dikta et al. (2006) specialized Stute's (1997) approach for binary regression when the outcomes are always observed and proposed a model-based resampling procedure; see also Zhu et al. (2002) for related work.

We propose a modification of Dikta et al.'s (2006) model-based resampling, where we regenerate only the observed binary responses and perform repeated complete case 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 a power study 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 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 some theorems are detailed in the Appendix.

2. Adjusted marked empirical processes

Our proposed adjusted MEP is a special case of Stute (1997). We start with the process

$$t \longrightarrow n^{-1/2}\sum_{i=1}^{n}\{\delta_i - m(Z_i,\boldsymbol{\theta}_0)\}\, I(Z_i \le t). \tag{2.1}$$

Because the binary responses are MAR, ξ is independent of δ given Z. Then the random variables δ and $\sigma + (1-\xi)m(Z)$ have the same conditional mean given Z, namely $m(Z)$:

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

By MAR, $\boldsymbol{\theta}$ can be estimated from the "complete cases": $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbb{R}^k} l_n(\boldsymbol{\theta})$, where

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

see Subramanian (2004). Substituting $\sigma_i + (1-\xi_i)m(Z_i,\boldsymbol{\theta}_0)$ for δ_i and the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}$ for θ_0 in Eq. (2.1) we obtain the adjusted MEP

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

We will need some basic results. 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 $\bar{m}(t,\boldsymbol{\theta}) = 1 - m(t,\boldsymbol{\theta})$. 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 $I(\boldsymbol{\theta}_0) = (\sigma_{r,s})_{1\le r,s\le k}$, where

$$\sigma_{r,s} = E\left(\frac{\pi(Z)D_r(m(Z,\boldsymbol{\theta}_0))D_s(m(Z,\boldsymbol{\theta}_0))}{m(Z,\boldsymbol{\theta}_0)\bar{m}(Z,\boldsymbol{\theta}_0)}\right). \tag{2.4}$$

Since $\hat{\boldsymbol{\theta}}$ is derived via maximum likelihood, we have (cf. Subramanian, 2004; Dikta, 1998)

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = I^{-1}(\boldsymbol{\theta}_0)\, n^{-1/2}\sum_{i=1}^{n}\mathrm{Grad}(w_i(Z_i,\boldsymbol{\theta}_0)) + o_{\mathbb{P}}(1). \tag{2.5}$$

The influence function of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is readily seen to be

$$\mathcal{L}(\xi,\sigma,Z,\boldsymbol{\theta}_0) := I^{-1}(\boldsymbol{\theta}_0)\mathrm{Grad}(w(Z,\boldsymbol{\theta}_0)) = \frac{\xi(\delta - m(Z,\boldsymbol{\theta}_0))}{m(Z,\boldsymbol{\theta}_0)\bar{m}(Z,\boldsymbol{\theta}_0)} I^{-1}(\boldsymbol{\theta}_0)\mathrm{Grad}(m(Z,\boldsymbol{\theta}_0)). \tag{2.6}$$

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

Let $\mathcal{R}_\infty$ denote a zero-mean Gaussian process with covariance function given by

$$K_1(s,t) = \int_0^{s\wedge t}\pi(u)m(u,\boldsymbol{\theta}_0)\bar{m}(u,\boldsymbol{\theta}_0)dH(u) - \int_0^s\int_0^t \pi(u)\pi(v)\beta(u,v)dH(u)dH(v), \tag{2.7}$$

where $\beta(u, v) = (\mathrm{Grad}(m(u, \boldsymbol{\theta}_0)))^T I^{-1}(\boldsymbol{\theta}_0)\mathrm{Grad}(m(v, \boldsymbol{\theta}_0))$. We can write

$$\mathcal{R}_n(t) = n^{-1/2}\sum_{i=1}^{n} \xi_i\,[\delta_i - m(Z_i, \boldsymbol{\theta}_0)]\,I(Z_i \le t) + n^{-1/2}\sum_{i=1}^{n} \xi_i\left[m(Z_i, \boldsymbol{\theta}_0) - m(Z_i, \hat{\boldsymbol{\theta}})\right] I(Z_i \le t).$$

Denote the first sum by $\mathcal{R}_{n,1}(t)$. By differentiability and Eq. (2.5), the second sum is seen to be $\mathcal{R}_{n,2}(t) + o_p(1)$ uniformly in t (cf. Dikta et al., 2012), where

$$\mathcal{R}_{n,2}(t) = -n^{-1/2}\sum_{i=1}^{n}\left(\int_0^t \pi(s)\,(\mathrm{Grad}(m(s, \boldsymbol{\theta}_0)))^T\,dH(s)\right)\mathcal{L}(\xi_i, \sigma_i, Z_i, \boldsymbol{\theta}_0),$$

and H is the distribution function of Z. By Eqs. (2.5), (2.6), and condition (A3) given in the Appendix, assumptions 1 and 2 of Stute (1997) are fulfilled, whose Corollary 1.3 then implies that $\mathcal{R}_n \xrightarrow{\mathcal{D}} \mathcal{R}_\infty$ in $D[0, \infty]$. The following expressions yield $K_1(s, t)$ given by Eq. (2.7):

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

Under conditions (C1) and (A1)–(A3) (cf. Appendix), we have proved the following result.

Theorem 1. *Suppose that conditions* (C1) *and* (A1)–(A3) *hold. 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 KS and CvM statistics are given by

$$\mathcal{D}_n = \sup_{0\le t\le\infty} |\mathcal{R}_n(t)|; \qquad \mathcal{W}_n = \int \mathcal{R}_n^2(t)d\hat{H}(t). \tag{2.8}$$

Theorem 1 and the continuous mapping theorem imply that $\mathcal{D}_n \xrightarrow{\mathcal{D}} \mathcal{D}_\infty := \sup_{0\le t\le\infty} |\mathcal{R}_\infty(t)|$, 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 $\xi_i^* = \xi_i = 0$.

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 $\hat{\boldsymbol{\theta}}^*$ is a measurable solution of $\mathrm{Grad}(l_n^*(\boldsymbol{\theta})) = 0$, 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 $\mathbb{P}_n$, $\mathbb{E}_n$, Var_n, and Cov_n denote the probability measure, expectation, variance, and covariance associated with the bootstrap sample. We will need the condition (C2) given in the Appendix, which can be shown, for example, using the methods of Stute (1992). Proof of Theorem 2 is given in the Appendix.

Theorem 2. *Suppose that* $\boldsymbol{\Theta}$ *is a connected open subset of* $\mathbb{R}^k$ *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

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

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

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

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 conditions* (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. (2.7).*

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) and then $\mathcal{D}_n$ and $\mathcal{W}_n$ given by Eq. (2.8). We obtained 1000 values of $\mathcal{D}_n^*$ and $\mathcal{W}_n^*$ from $n$boot = 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 $n$sim = 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) given by $m(t,\boldsymbol{\theta}) = \theta_1/(\theta_1 + t^{\theta_2})$, $\theta_1 > 0$, $\theta_2 \in \mathbb{R}$, which arises when the failure and censoring are each 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 of about 40%. Both tests achieve the nominal 5% when there is no misspecification ($d = 0.7$); see Fig. 1.

The empirical distribution function of the 1000 p-values corresponding to the two tests for $(a,b,c,d) = (2.0, 0.7, 0.9, 1.25)$ is displayed in Fig. 2. The CvM test performed best.

For comparison, we also investigated with an augmented inverse probability of nonmissingness weighted MEP, given by (cf. Subramanian, 2010)

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

where $\hat{\pi}(x)$ is a kernel estimator of $P(\xi = 1|Z = x)$. Related comparison plots can be found in Subramanian (2010) and Dikta et al. (2012).

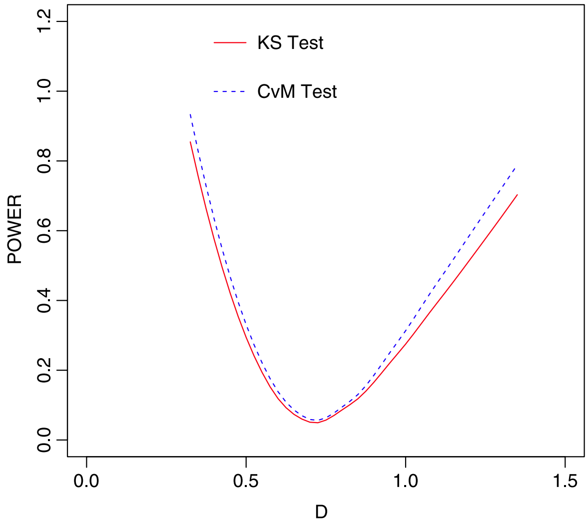

Fig. 1. Empirical power of proposed tests.

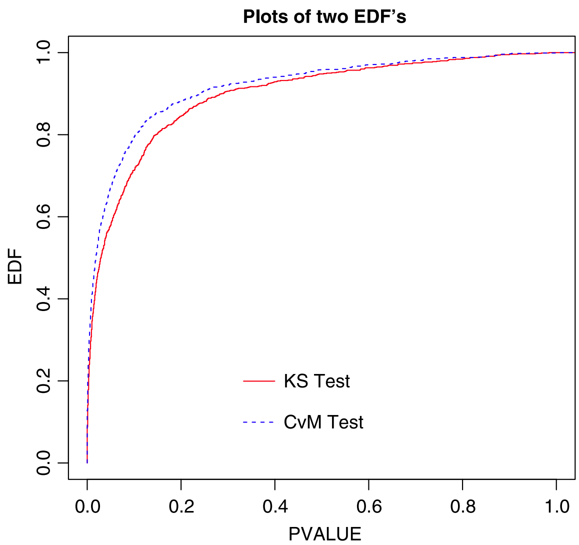

Fig. 2. Empirical distribution function of 1000 bootstrap p-values.

4.2. Illustration using a mice data set

Using our proposed procedure, 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 $\text{logit}(m(x, \boldsymbol{\theta})) = \theta_1 + \theta_2 x + \theta_3 x^2$ and the three-parameter probit model given by $m(x, \boldsymbol{\theta}) = \Phi(\theta_1 + \theta_2 x + \theta_3 x^2)$, where $\text{logit}(\varphi) = \log(\varphi/(1-\varphi))$ 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 was 0.653. The fitted three-parameter logit model was $\text{logit}(m(x, \hat{\boldsymbol{\theta}})) = -27.9041 + 30.8175x - 8.3415x^2$. The corresponding bootstrap p-value of the KS test 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 Fig. 3.

Acknowledgments

The authors thank the Editor Professor Koul and a reviewer for their comments and suggestions, which led to improvements in the paper.

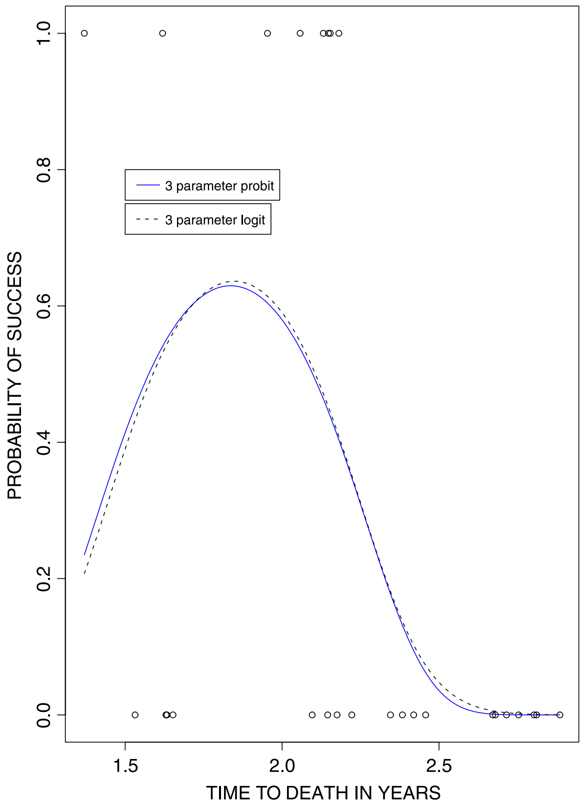

Fig. 3. Scatter plot of the nonmissing binary responses and plots of the two fitted models of $m(t)$ for the mice data.

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 $\mathrm{Grad}(l_n(\boldsymbol{\theta})) = 0$, such that $\hat{\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$.

(C2) For $\mathbb{P}$ almost all sample sequences there exists $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, a measurable solution of the equation of $\mathrm{Grad}(l_n^*(\boldsymbol{\theta})) = 0$, such that $\mathbb{P}_n(|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0| > \epsilon) \longrightarrow 0$ 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. (2.4), 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 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\|\mathrm{Grad}(m(t, \boldsymbol{\theta}))\| \leq N(t)$ and $E(N(Z)) < \infty$.

All the conditions above were given by Dikta (1998) and Dikta et al. (2006). To a remark by a reviewer whether there are sufficient conditions for (C1), we have the following observations: 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. Different approaches are available and we do not wish to insist on a particular sufficient condition for (C1).

Proof of Theorem 2. For details see our technical report (Dikta et al., 2012). 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 $A_n^*(\boldsymbol{\theta}) = (a_{r,s}^{n*}(\boldsymbol{\theta}))_{1 \leq r,s \leq k}$ is a $k \times k$ matrix with $a_{r,s}^{n*}(\boldsymbol{\theta}) = D_{r,s}(l_n^*(\boldsymbol{\theta}))$, for $1 \leq r, s \leq k$. 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 can show 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 the Lindeberg condition, we can also show that

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

with probability 1 as $n\to\infty$. Deduce from the Cramér–Wold device that $n^{1/2}\mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}}))$ is asymptotically normal. □

Proof of Theorem 3. For details see our technical report (Dikta et al., 2012). Writing $\alpha(t,\boldsymbol{\theta}_0) = \int_0^t \pi(u)\mathrm{Grad}(m(u,\boldsymbol{\theta}_0))dH(u)$, we first note that the second term of Eq. (2.7) 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

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

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

Note that $\mathcal{R}_{n,1}^*(t)$ is centered since $\mathbb{E}_n(\delta_i^*) = m(Z_i,\hat{\boldsymbol{\theta}})$. 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

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

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 $o_{\mathbb{P}_n}(1)$. It follows from Theorem 2 of Jennrich (1969) that

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

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

$$\begin{aligned}
K_1^*(s,t) = \quad & \frac{1}{n}\sum_{i=1}^n \xi_i\, m(Z_i,\hat{\boldsymbol{\theta}})\bar{m}(Z_i,\hat{\boldsymbol{\theta}})I(Z_i\le s\wedge t) + \alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)I_nI^{-1}(\boldsymbol{\theta}_0)\alpha(t,\boldsymbol{\theta}_0) \\
& - \alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{n}\sum_{i=1}^n \xi_i\mathrm{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))I(Z_i\le t)\right) \\
& - \alpha^T(t,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{n}\sum_{i=1}^n \xi_i\mathrm{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))I(Z_i\le s)\right) \\
\longrightarrow & \int_0^{s\wedge t}\pi(u)m(u,\boldsymbol{\theta}_0)\bar{m}(u,\boldsymbol{\theta}_0)dH(u) - \alpha^T(s,\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)\alpha(t,\boldsymbol{\theta}_0),
\end{aligned}$$

which is the same as Eq. (2.7). 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, we can show that $\mathcal{R}_{n,2}^*$ and $\mathcal{R}_{n,1}^*$ each induce a tight sequence of distributions on $D[0,\infty]$. □

References

Aggarwal, D., Koul, H.L., 2008. Minimum empirical distance goodness-of-fit tests for current status data. J. Indian Statist. Assoc. 46, 79–124.
Cox, D.R., 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., Schmidt, C., 2006. Bootstrap approximations in model checks for binary data. J. Amer. Statist. Assoc. 101, 521–530.
Dikta, G., Subramanian, S., Winkler, T., 2012. Model checks via bootstrap when there are missing binary data. CAMS Technical Report. Department of Mathematical Sciences, New Jersey Institute of Technology. Newark, New Jersey, USA.
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.
Jennrich, R.I., 1969. Asymptotic properties of non-linear least-squares estimators. Ann. Statist. 40, 633–643.

Koul, H.L., Yi, T., 2006. Goodness-of-fit testing in interval censoring case 1. Statist. Probab. Lett. 76, 709–718.
Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis With Missing Data. Wiley, New York.
Perlman, M.D., 1972. On the strong consistency of approximate maximum likelihood estimates. In: 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., Presedo Quindimil, M., 1998a. Bootstrap approximations in model checks for regression. J. Amer. Statist. Assoc. 93, 141–149.
Stute, W., Thies, S., Zhu, L.-X., 1998b. Model checks for regression: an innovation process approach. Ann. Statist. 26, 1916–1934.
Subramanian, S., 2004. The missing censoring-indicator model of random censorship. In: Balakrishnan, N., Rao, C.R. (Eds.), Handbook of Statistics 23: Advances in Survival Analysis. pp. 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.
van der Laan, M.J., McKeague, I.W., 1998. Efficient estimation from right-censored data when failure indicators are missing at random. Ann. Statist. 26, 164–182.
Zhu, L.X., Yuen, K.C., Tang, N.Y., 2002. Resampling methods for testing random censorship models. Scand. J. Stat. 29, 111–123.