Statistica Sinica **6**(1996), 579-588

TOWARDS AN OMNIBUS DISTRIBUTION-FREE GOODNESS-OF-FIT TEST FOR THE COX MODEL

Ian W. McKeague and Yanqing Sun

Florida State University and University of North Carolina

Abstract: A new goodness-of-fit test for Cox's proportional hazards model is introduced. The test is based on a transformation of the difference between nonparametric and Cox model specific estimators of the doubly-cumulative hazard function used by McKeague and Utikal (1991). The transformation is designed to give an asymptotically distribution-free test. The test is shown to be consistent against all alternatives except those in which the baseline hazard is linearly dependent on the covariate.

Key words and phrases: Brownian sheet, martingale methods, nonparametrics, proportional hazards.

1. Introduction

Cox's (1972) proportional hazards model specifies the conditional hazard function of a survival time given a covariate z as $\lambda(t|z) = \lambda_0(t)e^{\beta_0 z}$, where β_0 is an unknown regression parameter and $\lambda_0(t)$ is an unknown baseline hazard function. Numerous goodness-of-fit tests and graphical techniques for checking specific features of this model are available (see, e.g., the survey in Andersen, Borgan, Gill and Keiding (1993), Ch. VII). The key features to check are proportional hazards, i.e., $\lambda(t|z) = \lambda_0(t)r(z)$, where $r(z)$ is some (unknown) relative risk function, and log-linearity of the relative risk function, i.e., $\log r(z) = \beta_0 z$.

Omnibus tests for detecting *arbitrary* departures from the Cox model have been introduced recently, see McKeague and Utikal (1991) and Lin, Wei and Ying (1993). However, the asymptotic null distributions in these tests are model dependent and need to be estimated via Monte Carlo techniques.

In this paper we introduce a new goodness-of-fit test for the Cox model that is asymptotically distribution-free and consistent against a large class of alternatives. Our test is not omnibus, but it comes close to being so. Specifically, it is consistent against all alternatives that are not of the form $\lambda(t|z) = \lambda_1(t,z)e^{\beta_o z}$, where $\lambda_1(t,z)$ is a linear function of z for each fixed t, i.e., $\lambda_1(t,z) = \alpha_0(t)+\alpha_1(t)z$. The asymptotic null distribution is of a standard form and Monte Carlo techniques are not required. This property will be achieved via the transformation method of Khmaladze (1981, 1993). We restrict attention to continuous covariates, but discrete covariates can be handled in an analogous fashion—see the remarks following Theorem 3.1.

The approach is to transform the test statistic process $X(t,z) = \sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}})$ of McKeague and Utikal (1991) in such a way that it converges weakly to Brownian sheet. Here $\tilde{\mathcal{A}}$ is a nonparametric estimator of the doubly cumulative hazard function $\mathcal{A}(t,z) = \int_0^t \int_0^z \lambda(s|x)\,dxds$, and $\hat{\mathcal{A}}$ is a Cox model based estimator of $\mathcal{A}$. Under the Cox model, X converges weakly to a Gaussian random field of the form

$$m(t,z) \quad = \quad \int_0^t \int_0^z \sqrt{h}\,dW - b(z) \int_0^t \int_0^1 g\,dW - c(t,z) \int_0^1 \int_0^1 q\,dW, \qquad (1.1)$$

for $(t,z) \in [0,1]^2$, where W is a Brownian sheet on $[0,1]^2$ and h, b, c, g and q are certain nonrandom functions. The Brownian sheet W is a continuous Gaussian process with mean zero and covariance function $\text{cov}(W(t,z), W(t',z')) = \min(t,t')\min(z,z')$. The integrals with respect to W are defined in the L^2 sense (see Wong and Zakai (1974)).

We shall construct a transformation T that takes m to a Brownian sheet, and provide an estimator $\hat{T}$ of T such that $\hat{T}(X)$ converges weakly to a Brownian sheet. Then the Kolmogorov–Smirnov statistic computed from $\hat{T}(X)$ will converge weakly to $\sup|W(t,z)|$. The exact distribution of the supremum of a Brownian sheet is not known—only approximations are available (see Adler (1990)), but critical values are easily obtained by simulation. McKeague, Nikabadze and Sun (1995), henceforth MNS, found a transformation that takes Gaussian random fields of the form (1.1), but with the last term missing, to a Brownian sheet. They used such a transformation to develop an omnibus test for independence of a survival time from a covariate. Another application, to nonparametric change-point analysis, has been developed by McKeague and Sun (1996).

The transformation T constructed here is a composition $T = J_2 \circ J_1$ of two transformations of the type introduced in MNS. The explicit forms of J_1 and J_2 are given in Section 2. Estimators $\hat{J}_1$, $\hat{J}_2$ of J_1, J_2 are given in Section 3, as well as our main result showing that $\hat{T}(X) = \hat{J}_2(\hat{J}_1(X))$ converges weakly to Brownian sheet. The proof of this result is given in the Appendix.

2. Transformations

In this section we construct the transformation T of m given by (1.1) to a Brownian sheet. First we give the general form of the transformations constructed by MNS that we intend to use. A transformation that takes the sum of the first two terms of (1.1) to a Brownian sheet has the form

$$J(\xi)(t,z) \quad = \quad \int_0^t \int_0^z f_1\,d\xi - \int_0^t \int_0^1 f_2\,d\xi, \qquad (2.2)$$

where f_1, f_2 are nonrandom functions, f_2 depends implicitly on z, and ξ is any function for which the above integrals are defined in the sense of weak net integration (Hildebrandt (1963), Section III.8). A transformation of this type, for other choices of f_1 and f_2, transforms to a Brownian sheet a process of the form $W - \eta K$, where η is any random variable and K is any absolutely continuous nonrandom function of t and z.

Write $m = m_0 - \eta c$, where m_0 is the sum of the first two terms in m, and η is the stochastic integral part of the last term in (1.1). Let J_1 be a transformation of the form (2.2) that takes m_0 to a Brownian sheet B. By linearity of J_1, we can write $J_1(m) = B - \eta K$, where $K = J_1(c)$. Then let J_2 transform $B - \eta K$ to a Brownian sheet, so that $T(m) = J_2(J_1(m))$ is a Brownian sheet, as required.

We can explicitly write the pairs of functions $(f_1^{(j)}, f_2^{(j)})$, $j = 1, 2$ used in J_1 and J_2, respectively, by referring to MNS (Proposition 2.1 and Theorem 2.1). This gives $f_1^{(1)} = h^{-1/2}$,

$$f_2^{(1)}(s,u,z) = h^{-1/2}(s,u) \int_0^{z \wedge u} Q(s,u,x)\, dx,$$

and $f_1^{(2)} = 1$,

$$f_2^{(2)}(s,u,z) \quad = \quad k(s,u) \int_0^{z \wedge u} \frac{k(s,x)}{\int_x^1 k^2(s,v)\, dv}\, dx. \tag{2.3}$$

Here $k(s,x) = \partial^2 K / \partial s \partial x$ and

$$Q(s,u,x) = \frac{h^{-\frac{1}{2}}(s,u) b'(u) h^{-\frac{1}{2}}(s,x) b'(x)}{\int_x^1 h^{-1}(s,v)(b'(v))^2 dv}.$$

An explicit formula for T is given in the Appendix. We assume that h is bounded, positive and bounded away from zero. Also, the various derivatives of b, K and c used above are assumed to exist, and the denominators in $f_2^{(2)}$ and Q are assumed to be finite and positive for $x < 1$.

According to the above discussion we can state the following result:

Theorem 2.1. *The process $T(m)$ is a Brownian sheet on* $[0,1]^2$.

Recall from the Introduction that we are going to apply an estimated version of the transformation T to the process $X = \sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}})$. First we show that $T(X)$ converges weakly to a Brownian sheet, which will follow from the continuous mapping theorem provided that T is continuous on C_2, the space of continuous functions on $[0,1]^2$. We need to modify T slightly to make it continuous. Note that the denominator in Q vanishes at $x = 1$; so to ensure continuity of J_1 we restrict the domain of $J_1(\xi)(\cdot)$ to be $[0,1] \times [0,\rho]$, where $\rho < 1$. This means that the domain of J_2 must be restricted to functions defined on $[0,1]\times[0,\rho]$, which in turn forces us to modify J_2 because the range of integration for the second integral in (2.2) is $[0,t]\times[0,1]$. We replace J_2 by

$$J_2(\xi)(t,z)=\int_0^t\int_0^z f_1^{(2)}\,d\xi-\int_0^t\int_0^\rho f_2^{(2)}\,d\xi,\quad (t,z)\in[0,1]\times[0,\rho], \tag{2.4}$$

where $f_2^{(2)}$ is now defined by replacing the denominator in (2.3) by $\int_x^\rho k^2(s,v)\,dv$. Inspection of the proof of Proposition 2.1 of MNS shows that $J_2(B-\eta K)$ is a Brownian sheet on $[0,1]\times[0,\rho]$. From now on, the transformations J_1 and J_2 in $T=J_2\circ J_1$ are taken to be the modified versions just described. Then T is well-defined as a map from $C_2\cup BV_2$ to $D([0,1]\times[0,\rho])$ and it is continuous on C_2. Here BV_2 is the subspace of $D[0,1]^2$ consisting of functions ξ for which $\xi,\xi(0,\cdot)$ and $\xi(\cdot,0)$ have bounded variation. By the continuous mapping theorem, $T(X)$ converges weakly to a Brownian sheet on $[0,1]\times[0,\rho]$.

3. Cox Model Goodness-of-Fit Test

In this section we construct an estimator $\hat{T}$ of the transformation T and show that $\hat{T}(X)$ converges weakly to a Brownian sheet. This will lead to our distribution-free goodness-of-fit test for the Cox model. We also identify a class of alternatives for which the test is consistent.

Some preliminaries are needed. First we set up our notation for the random censorship model. Suppose that (U_i,C_i,Z_i), $i\geq 1$, are i.i.d. copies of a generic triple (U,C,Z) consisting of a failure time U, a censoring time C, and a covariate Z such that U and C are conditionally independent given Z. The conditional hazard function of U given Z is denoted $\lambda(t|z)$. The observations consist of the (possibly right censored) failure times $\tilde{U}_i=\min(U_i,C_i)$ and the failure indicators $\delta_i=I(U_i\leq C_i)$, for $i=1,\ldots,n$.

Next we define the estimators $\tilde{\mathcal{A}}$ and $\hat{\mathcal{A}}$. As in McKeague and Utikal (1991), partition the interval $[0,1]$ into d_n covariate strata $\mathcal{I}_r=[x_{r-1},x_r)$, $r=1,\ldots,d_n$, where $x_r=rw_n$ and $w_n=1/d_n$ is the width of each stratum. Let $\mathcal{I}_z=\mathcal{I}_r$ for $z\in\mathcal{I}_r$. The nonparametric estimator of the doubly cumulative hazard function $\mathcal{A}$ is

$$\tilde{\mathcal{A}}(t,z)=\int_0^z\int_0^t\frac{N^{(n)}(ds,x)}{Y^{(n)}(s,x)}\,dx,$$

where $N^{(n)}(t,z)=\sum_{i=1}^n I(\tilde{U}_i\leq t,\delta_i=1,Z_i\in\mathcal{I}_z)$ is the number of observed failures and $Y^{(n)}(t,z)=\sum_{i=1}^n I(\tilde{U}_i\geq t,Z_i\in\mathcal{I}_z)$ is the size of the risk set at time t, for the covariate stratum containing z. By convention, $1/0=0$.

The Cox model based estimator of $\mathcal{A}$ is $\hat{\mathcal{A}}(t,z)=\hat{\Lambda}_0(t)\int_0^z e^{\hat{\beta}x}\,dx$, where $\hat{\beta}$ is Cox's maximum partial likelihood estimator of β_0 and $\hat{\Lambda}_0$ is Breslow's estimator of the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$. Suppose that the covariate subdistribution function $F(t,\cdot) = P(Z \le \cdot, \tilde{U} \ge t)$ has sub-density $f(t,\cdot)$ for each fixed t, that f is positive and continuous on $[0,1]^2$, and that the baseline hazard function λ_0 is Lipschitz. Also assume that $nw_n^2 \to 0$ and $nw_n^{1+\delta} \to \infty$ for some $0 < \delta < 1$. Then, under the Cox model, McKeague and Utikal (1991) showed that $X = \sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}})$ converges weakly to m defined by (1.1), where $h = \lambda/f$, $b(\cdot) = \int_0^{\cdot} e^{\beta_0 x}\,dx$, and

$$c(t,z) = \Sigma^{-1}\Big[\Lambda_0(t)\int_0^z xe^{\beta_0 x}\,dx - b(z)\int_0^t \Big(\frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\Big)^2 \lambda_0(u)\,du\Big].$$

Here Σ^{-1} is the asymptotic variance of $\hat{\beta}$, and $s^{(0)}, s^{(1)}$ are standard notations from Andersen and Gill (1982).

3.1. The estimator $\hat{T}$

Note that the transformation J_1 has unknown components λ_0, β_0 and f. A simple calculation shows that $K = J_1(c_1)$, where c_1 is first term in the definition of c. Also note that Σ^{-1} appears as a factor in K due to its presence in c_1, but, being constant, it disappears from J_2. Thus the only unknown components of J_2 are λ_0, β_0 and f. Hence, to estimate T, it suffices estimate these three components of the model. Naturally, $\hat{\beta}$ is used to estimate β_0.

We shall need a uniformly consistent estimator of λ_0. Moreover, methods from stochastic calculus are to be applied to martingale integrals involving this estimator, so it needs to be predictable. Following Ramlau-Hansen (1983), we use a kernel estimator of the form

$$\hat{\lambda}_0(t) = \frac{1}{b_n}\int_0^1 \kappa\Big(\frac{t-s}{b_n}\Big)\,d\hat{\Lambda}_0(s),$$

where b_n is a bandwidth parameter, κ is a Lipschitz kernel function having support $[0,1]$ and integral 1. This estimator is predictable and can be shown to be uniformly $O_P(b_n)$-consistent over intervals strictly contained within $[0,1]$ if $b_n = n^{-\delta}$, where $0 < \delta < 1/4$. Thus, to be able to use $\hat{\lambda}$ we need to restrict the entire analysis to a time interval strictly contained within $[0,1]$. However, to preserve the earlier notation we assume that $[0,1]$ is strictly contained within a larger follow-up interval, say $[-\epsilon, 1+\epsilon]$, for some $\epsilon > 0$, so that $\hat{\lambda}$ is uniformly consistent over the whole of $[0,1]$.

We estimate f in a similar fashion, using the kernel estimator

$$\hat{f}(t,z) = \frac{1}{b_n}\int_0^1 \kappa\Big(\frac{z-x}{b_n}\Big)\,d\hat{F}(t,x),$$

where $\hat{F}$ is the empirical estimator of F. It is readily shown, using the tightness criterion in Bickel and Wichura (1971), that $\sqrt{n}(\hat{F} - F)$ converges weakly to a Gaussian random field on $[0,1]^2$. This implies, using the technique of Ramlau-Hansen (1983), Proof of Theorem 4.1.2, that $\hat{f}$ is uniformly $O_P(b_n)$-consistent over regions that are contained within $[-\epsilon, 1+\epsilon] \times [0,1]$ and bounded away from its edges. To use $\hat{f}$ we need to restrict the analysis to such a region. But again, to preserve our earlier notation, we assume that the covariate data are collected over the interval $[-\epsilon', 1+\epsilon']$, for some $\epsilon' > 0$, so that $\hat{f}$ is uniformly consistent over the whole of $[0,1]^2$. Note that this consistency holds for general λ, not only for the Cox model.

Our estimator of $T = T(f, \lambda_0, \beta_0)$ is $\hat{T} = T(\hat{f}, \hat{\lambda}_0, \hat{\beta}_0)$.

3.2. The test statistic

Our goodness-of-fit test statistic is taken to be the Kolmogorov-Smirnov statistic $S = \sup_{t,z} |\hat{T}(X)(t,z)|$, where the supremum is over $[0,1] \times [0,\rho_0]$. Here $\rho_0 < \rho < 1$. Note that we have enlarged the original unit square over which data is collected in order to estimate T consistently, as discussed above. The further restriction of z to $[0,\rho_0]$ is required to ensure that J_2 is continuous, for which the denominator in $f_2^{(2)}$, namely $\int_x^\rho k^2(s,v)\,dv$, must be bounded away from zero. The latter is guaranteed by restricting z to a smaller interval $[0,\rho_0]$, where $0 < \rho_0 < \rho$, and assuming that f and λ_0 are bounded away from 0. To show that the transformation T is a continuous map from $C_2 \cup BV_2$ to $D([0,1] \times [0,\rho_0])$ we further assume that f and λ_0 are Lipschitz continuous and $f \in BV_2$.

We now state the main result of the paper. The proof is given in the Appendix.

Theorem 3.1. *Under the Cox model, $\hat{T}(X)$ converges weakly to a Brownian sheet in $D([0,1] \times [0,\rho_0])$.*

To calculate P-values based on S, refer to the distribution of $S^* = \sup_{t,z} |W(t,z)|$, where the supremum is over $[0,1] \times [0,\rho_0]$. This distribution can be found accurately by simulation of Brownian sheet. For $\rho_0 = .9$, the 5% critical level is 2.2811.

Can we extend our approach to discrete covariates? A *direct* extension is not possible because the assumption of a continuous covariate is crucial for obtaining the Brownian sheet limit of the "test process" $\hat{T}(X)$. This is comparable to the situation in standard survival analysis where it is necessary to have a continuous survival time to to be able to express the asymptotic distribution of the Nelson–Aalen estimator in terms of a continuous time-changed Brownian motion. We briefly indicate how discrete covariates can be handled using a modification of our approach.

For discrete covariates it is unnatural to integrate over z (as in $\tilde{\mathcal{A}}$), but rather one should consider each covariate level separately. Suppose that there are p covariate levels $z_1, \ldots, z_p$. The idea is to compare general and Cox-model based estimators of the p-vector of cumulative conditional hazard functions

$$\mathbf{A}(t) = \left(\int_0^t \lambda(s|z_1)\, ds, \ldots, \int_0^t \lambda(s|z_p)\, ds \right).$$

The general estimator of the jth component of $\mathbf{A}(t)$ is of Nelson–Aalen type:

$$\tilde{\mathbf{A}}_j(t) = \int_0^t \frac{N^{(n)}(ds, z_j)}{Y^{(n)}(s, z_j)}$$

and the Cox-model based estimator is $\hat{\mathbf{A}}_j(t) = \hat{\Lambda}_0(t) e^{\hat{\beta} z_j}$. By adapting the proof of Theorem 3.1 of McKeague and Utikal (1991), it can be shown that $\mathbf{X} \equiv \sqrt{n}(\tilde{\mathbf{A}} - \hat{\mathbf{A}})$ converges weakly to a p-variate Gaussian process $\mathbf{m}$ that can be expressed as a sum of stochastic integrals with respect to a p-dimensional Wiener process $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_p)$. The form of $\mathbf{m}$ is analogous to (1.1) except that the Brownian sheet W is replaced by $\mathbf{W}$ and there is no integration over the covariate. Next, use an innovation martingale approach (cf. MNS, Section 2) to construct a transformation $\mathbf{T}$ such that $\mathbf{T}(\mathbf{m})$ is a p-dimensional Wiener process. Let $\hat{\mathbf{T}}_j$ be a consistent estimator of the jth component of $\mathbf{T}$. Then the test statistic to be used in the discrete covariate case is $\sup_{t,j} |\hat{\mathbf{T}}_j(\mathbf{X})(t)|$, which converges in distribution to $\sup_{t,j} |\mathbf{W}_j(t)|$.

3.3. Consistency

First note that under general alternatives, $\hat{\beta}$ and $\hat{\Lambda}_0(t)$ converge in probability to some nonrandom β^* and $\Lambda_0^*(t) = \int_0^t \lambda_0^*(s)\, ds$, respectively (see Lin and Wei (1989)). Thus $\hat{\mathcal{A}}(t,z)$ converges in probability to $\mathcal{A}^*(t,z) = \Lambda_0^*(t) \int_0^z e^{\beta^* x}\, dx$. Moreover, it can be seen that $\hat{T} = \hat{T}(\hat{f}, \hat{\lambda}_0, \hat{\beta})$ estimates $T^* = T(f, \lambda_0^*, \beta^*)$ under general alternatives. It follows (cf., MNS, Proof of Theorem 3.3) that our test is consistent provided

$$T^*(\mathcal{A} - \mathcal{A}^*)(t,z) \neq 0$$

for some $(t,z) \in [0,1] \times [0,\rho_0]$. By solving $T^*(\mathcal{A} - \mathcal{A}^*) = 0$ for λ we can identify alternatives for which our test may fail to be consistent. This leads to the following result.

Theorem 3.2. *The test based on S is consistent against all alternatives λ that are not of the form $\lambda(t|z) = (\alpha_0(t) + \alpha_1(t)z)e^{\beta z}$ over the rectangle $[0,1] \times [0,\rho_0]$, where α_0 and α_1 are arbitrary functions of t.*

Proof. Let J_1^* denote J_1 with λ_0 and β_0 replaced by λ_0^* and β^*, similarly for J_2^*. Slightly abusing our earlier notation, h, k, Q and b are now considered to have λ_0 and β_0 replaced by λ_0^* and β^* .

We first show that if $J_2^*(\xi) = 0$ then $\psi(t,z) \equiv \partial^2\xi/\partial t\partial z = \gamma(t)k(t,z)$ for some function γ. Suppose $J_2^*(\xi) = 0$. Then, differentiating $J_2^*(\xi)(t,z)$ with respect to t and z, we obtain

$$\psi(t,z)\int_z^\rho k^2(t,v)\,dv = k(t,z)\int_z^\rho k(t,v)\psi(t,v)\,dv \tag{3.5}$$

for $z < \rho$.

Taking partial derivatives w.r.t. z on both sides, multiplying both sides by the derivatives, and solving the resulting differential equation, we find that ψ has the stated form.

Now set $\xi = J_1^*(\mathcal{A} - \mathcal{A}^*)$. Then $T^*(\mathcal{A} - \mathcal{A}^*) = 0$ implies that $\psi = \gamma k$, which can be written explicitly as

$$\begin{aligned} & h^{-1/2}(t,z)\eta(t,z) - \int_z^1 h^{-1/2}(t,u)Q(t,u,z)\eta(t,u)\,du \\ = \ & \gamma(t)\left\{h^{-1/2}(t,z)\lambda_0^*(t)ze^{\beta^* z} - \int_z^1 h^{-1/2}(t,u)Q(t,u,z)\lambda_0^*(t)ue^{\beta^* u}\,du\right\}, \end{aligned}$$

where $\eta(t,z) = \lambda(t|z) - \lambda_0^*(t)e^{\beta^* z}$. Rearranging this equation gives

$$\zeta(t,z)\int_z^1 h^{-1}(t,v)(b'(v))^2\,dv = b'(z)\int_z^1 h^{-1}(t,u)b'(u)\zeta(t,u)\,du,$$

where $\zeta(t,z) = \lambda(t|z) - \gamma(t)\lambda_0^*(t)ze^{\beta^* z}$. The above equation can be solved for ζ in the same way that (3.5) was solved for ψ, to give $\zeta(t,z) = \alpha_0(t)e^{\beta^* z}$. This completes the proof.

Remark. The alternatives of the form

$$\lambda(t|z) = (\alpha_0(t) + \alpha_1(t)z)e^{\beta_0 z} \tag{3.6}$$

represent perturbations of the Cox model in which the baseline hazard function λ_0 is linearly dependent on the covariate. By a Taylor series argument, linear perturbations provide a satisfactory approximation to general hazards of the form $\lambda(t|z) = \lambda_0(t,z)e^{\beta_0 z}$, provided the dependence of $\lambda_0(t,z)$ on z is relatively mild. It would be of interest to investigate the model (3.6) in its own right. A test of $\alpha_1 \equiv 0$ would provide a means of detecting simple linear violations of the proportional hazards assumption. Aalen's (1980) additive risk model is contained in (3.6) (take $\beta = 0$). Thus, the new model gives a joint generalization of both the Cox and Aalen models.

Acknowledgement

Ian McKeague's research was partially supported by NSF Grant ATM-9417528. Yanqing Sun's research was partially supported by NIH Training Grant T32CA09667.

Appendix. Proof of Theorem 3.1

As discussed in Section 2, we first need to show that $T: C_2 \cup BV_2 \to D([0,1] \times [0, \rho_0])$ is continuous on C_2, where $\rho_0 < \rho < 1$. To do this we write $T(\xi)$, for $\xi \in C_2$, explicitly as

$$\begin{aligned} T(\xi) \;=\; & \int_0^t \int_0^z \phi_1 \, d\xi - \int_0^z \int_0^t \int_u^1 \phi_2, d\xi(s,x) du \\ & - \int_0^z \int_0^t \int_u^\rho \phi_3 \, d\xi(s,x) du + \int_0^z \int_v^\rho \int_0^t \int_u^1 \phi_4 \, d\xi(s,x) du dv, \end{aligned}$$

where

$$\begin{aligned} \phi_1(s,x,\beta) \;&=\; \Big[\lambda_0(s) e^{\beta x} / f(s,x)\Big]^{-1/2} \\ \phi_2(s,x,u,\beta) \;&=\; \frac{f(s,x) f^{1/2}(s,u) e^{\frac{1}{2}\beta u}}{\lambda_0^{1/2}(s) \int_u^1 f(s,v) e^{\beta_0 v} \, dv} I(u \le x \le 1) \\ \phi_3(s,x,u,\beta) \;&=\; \phi_1(s,x,\beta) \frac{k(s,x) k(s,u)}{\int_u^\rho k^2(s,v) \, dv} I(u \le x \le \rho) \\ \phi_4(s,x,u,v,\beta) \;&=\; \phi_2(s,x,u,\beta) \frac{k(s,u) k(s,v)}{\int_v^\rho k^2(s,w) \, dw} \\ k(s,x) \;&=\; \Big[f(s,x) \lambda_0(s) e^{\beta x}\Big]^{1/2} \frac{\int_x^1 f(s,y)(x-y) e^{\beta y} \, dy}{\int_x^1 f(s,v) e^{\beta v} \, dv} \end{aligned}$$

and β is set to β_0. Under our assumptions that f and λ_0 are Lipschitz and bounded away from zero, and $f \in BV_2$, it follows that all the integrands ϕ_i are continuous and have bounded variation in (s,x), uniformly in (u,v). The desired continuity of T then follows from the integration by parts formula for weak net integrals (see, e.g., Lemma 2 of MNS).

The proof can be completed by showing that $\|(T - \hat{T})(X)\| \xrightarrow{P} 0$, where $\|\cdot\|$ is the supremum norm on $D([0,1] \times [0, \rho_0])$. There is a term corresponding to each of the four terms in the above explicit representation of T. Taking the first term, we need to show that

$$\left\| \int_0^\cdot \int_0^\cdot (\hat{\phi}_1 - \phi_1) \, dX \right\| \xrightarrow{P} 0,$$

where $\hat{\phi}_1$ is the estimated version of ϕ_1. The decomposition of X given in McKeague and Utikal (1991, (8.8)) can be used to split the above integral into a sum of martingale integrals and integrals with respect to t and z. Martingale techniques cannot be used directly to bound the martingale integrals since $\hat{\phi}_1$ involves $\hat{\beta}$, so the integrand is not predictable. It is necessary to Taylor expand $\hat{\phi}_1(\beta)$ about $\beta = \beta_0$, and bound the remainder term of order $(\hat{\beta} - \beta_0)^2 = O_P(n^{-1})$. Details are given in Sun and McKeague (1995), Proof of Theorem 3.2. The integrals involving $\hat{\phi}_2$, $\hat{\phi}_3$ and $\hat{\phi}_4$ can be treated in a similar fashion.

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In: *Mathematical Statistics and Probability Theory*, Lecture Notes in Statistics **2**, (Edited by W. Kloneki, A. Kozek and J. Rosinski), pp. 1-25. Springer-Verlag, New York.

Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes.* Hayward, Calif.: Inst. Math. Statist.

Andersen, P. K. , Borgan, Ø. , Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656-1670.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser.B* **34**, 187-220.

Hildebrandt, T. H. (1963). *Introduction to the Theory of Integration.* Academic Press, New York.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26**, 240-257.

Khmaladze, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21**, 798-829.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84**, 1074-1078.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-572.

McKeague, I. W., Nikabadze, A. M. and Sun, Y. (1995). An omnibus test for independence of a survival time from a covariate. *Ann. Statist.* **23**, 450-475.

McKeague, I. W. and Sun, Y. (1996). Transformations of Gaussian random fields to Brownian sheet and nonparametric change-point tests. *Statist. Probab. Lett.*, in press.

McKeague, I. W. and Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand. J. Statist.* **18**, 177-195.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.

Sun, Y. and McKeague, I. W. (1995). A distribution-free goodness-of-fit test for the Cox proportional hazards model. *Technical Report.* University of North Carolina at Charlotte.

Wong, E. and Zakai, M. (1974). Martingales and stochastic integrals for processes with a multi-dimensional parameter. *Z. Wahrsch. verw. Gebiete* **51**, 109-122.

Department of Statistics, Florida State University, Tallahassee, FL 32306. U.S.A.

Department of Mathematics, University of North Carolina, Charlotte, NC 28223, U.S.A.

(Received December 1994; accepted July 1995)