Probability Theory/Statistics

Estimation of the survival probabilities by adjusting a Cox model to the tail

Estimation des probabilités de survie par l'ajustement sur la queue d'un modèle de Cox

Ion Grama, Jean-Marie Tricot, Jean-François Petiot

Université de Bretagne Sud, centre de recherche Yves-Coppens, BP 573, 56017 Vannes, France

ARTICLE INFO

ABSTRACT

Within the framework of a survival analysis model with censored life time data and an explanatory covariate, our goal is to predict the survival probability beyond the largest observed time. A Cox model with a constant underlying hazard function is proposed to adjust the tail of the life time distribution. Under some regularity conditions, we prove that the parameter estimators are convergent.

© 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

RÉSUMÉ

Dans le contexte d'un modèle d'analyse de survie avec données censurées et en présence d'une covariable explicative, on veut prédire la probabilité de survie au-delà de la plus grande observation. Un modèle de Cox dont la fonction de risque de base est constante, est ajusté sur la queue de la distribution de la durée de vie. Sous certaines conditions de régularité, on démontre la convergence des estimateurs des paramètres du modèle.

© 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Version française abrégée

On utilise le modèle de Cox à hasard proportionnel ([7,6]) pour prédire la durée de vie à partir des temps d'observation censurés $t_1, \ldots, t_n$, lorsqu'elle est proche ou même dépasse $\max\{t_1, \ldots, t_n\}$. Pour cela on propose d'ajuster la queue de la distribution de la durée de vie par un modèle de Cox. Cette approche est similaire à celle des références [2] et [3].

On considère une variable aléatoire (v.a.) X à valeur sur $[x_0, \infty)$, $x_0 \geqslant 0$, de fonction de répartition (f.r.) F, de densité f_F, de fonction de survie $S_F(x) = 1 - F(x)$, $x \geqslant x_0$ et de fonction de risque $h_F = f_F / S_F$. Les temps de censure à droite, à valeur sur $[x_0, \infty)$, sont notés C. Soient f_C et F_C respectivement, les densité et f.r. de C. Soient $S_C(x) = 1 - F_C(x)$, $x \geqslant x_0$ et $h_C = f_C / S_C$. On note $P_F(x, \delta)$, $x \geqslant 0$, $\delta \in \{0, 1\}$ la loi du vecteur (T, Δ), où $T = \min\{X, C\}$ et $\Delta = 1$ $(X \leqslant C)$. Soient $X_1, \ldots, X_n$ et $C_1, \ldots, C_n$ deux échantillons indépendants de X et C et soient $T_i = \min\{X_i, C_i\}$ et $\Delta_i = 1$ $(X_i \leqslant C_i)$. On note par $\mathbb{P}_F$ la loi des $Y_i = (T_i, \Delta_i)$, $i = 1, \ldots, n$ et par $\mathbb{E}_F$ l'espérance par rapport à $\mathbb{P}_F$. On suppose que $h_F(\cdot) = h(\cdot|z)$ dépend de la valeur z d'une covariable Z dans $\{0, 1\}$. Par la suite on considérera une suite de réalisations $z_1, \ldots, z_n$ de Z et les calculs se feront conditionnellement aux $Z_i = z_i$.

E-mail addresses: ion.grama@univ-ubs.fr (I. Grama), jean-marie.tricot@univ-ubs.fr (J.-M. Tricot), jean-francois.petiot@univ-ubs.fr (J.-F. Petiot).

Notons $\lambda_{\beta,z} \equiv \lambda e^{\beta z} > 0$. Pour $\lambda > 0$, $\beta \in \mathbb{R}$ et $t > x_0$, notons $F_{\lambda,\beta,t}(x|z)$ la f.r. telle que :

$$1 - F_{\lambda,\beta,t}(x|z) = \begin{cases} S_F(x), & x \in [x_0, t], \\ S_F(t)\exp(-\lambda e^{\beta z}(x-t)), & x > t. \end{cases}$$

Soit $\hat{n}_{z,t} = \sum_{T_i \geqslant t, Z_i = z} \Delta_i$, $z = 0, 1$. Les estimateurs locaux du maximum de la quasi-vraisemblance pour $\lambda_{\beta,0} = \lambda$ et $\lambda_{\beta,1} = \lambda e^{\beta}$ sont : $\hat{\lambda}_{z,t} = \hat{n}_{z,t} / \sum_{T_i \geqslant t, Z_i = z}(T_i - t)$, $z = 0, 1$ (avec $0/0 = 0$).

Soit $\mathcal{K}(\lambda_1, \lambda)$ la divergence de Kullback–Leibler entre deux lois exponentielles de paramètres $\lambda_1 > 0$ et $\lambda > 0$. Soit $\chi^2(P, P_0) = \int_{\mathbb{Y}} \frac{dP}{dP_0} dP - 1$ l'entropie du χ^2 entre les deux mesures de probabilité équivalentes P et P_0. On dit que la queue de F appartient au domaine d'attraction du modèle de Cox si il existe $\lambda > 0$ et $\beta \in \mathbb{R}$ tels que $\lim_{t \to \infty} \chi^2(P_F, P_{F_{\lambda,\beta,t}}) = 0$. Alors il existe une suite $\tau_n \in [x_0, \infty)$ telle que :

$$\chi^2(P_F, P_{F_{\lambda,\beta,\tau_n}}) = O\left(\frac{\ln n}{n}\right). \tag{0.1}$$

Si ξ_n et η_n sont des v.a. non négatives, la notation $\xi_n = O_{\mathbb{P}_F}(\eta_n)$ signifie qu'il existe $c > 0$ tel que $\mathbb{P}_F(\xi_n/\eta_n \leqslant c) \to 1$ lorsque $n \to \infty$ (avec $\frac{+\infty}{+\infty} = 0$).

Théorème 0.1. *Si F appartient au domaine d'attraction du modèle de Cox et τ_n est une suite satisfaisant* (0.1)*, alors $\mathcal{K}(\hat{\lambda}_{z,\tau_n}, \lambda_{\beta,z}) = O_{\mathbb{P}_F}(\frac{\ln n}{\hat{n}_{z,\tau_n}})$, pour $z = 0, 1$.*

Notons $q_{C,F}(t) = \int_t^\infty \frac{S_C(x)}{S_C(t)} \frac{f_F(x)}{S_F(t)} dx$ et $n_z = \sum_{i=1}^n 1 \ (Z_i = z)$.

Théorème 0.2. *Supposons qu'il existe deux constantes $q_0 > 0$ et $\kappa > 0$ telles que $q_{C,F}(t) \geqslant q_0$, pour $t \geqslant x_0$, et $\kappa \leqslant n_z/n \leqslant 1 - \kappa$. Si F appartient au domaine d'attraction du modèle de Cox et τ_n est une suite satisfaisant (0.1) et $nS_C(\tau_n)S_F(\tau_n) \to \infty$ lorsque $n \to \infty$, alors $\mathbb{E}_F \hat{n}_{z,\tau_n} \geqslant c_0 \kappa n S_C(\tau_n) S_F(\tau_n)$, pour une constante $c_0 > 0$, $\mathbb{P}_F(\hat{n}_{z,\tau_n} \geqslant \frac{1}{2} \mathbb{E}_F \hat{n}_{z,\tau_n}) \to 1$ lorsque $n \to \infty$, et $\mathcal{K}(\hat{\lambda}_{z,\tau_n}, \lambda_{\beta,z}) = O_{\mathbb{P}_F}(\frac{\ln n}{nS_C(\tau_n)S_F(\tau_n)})$.*

Soit $A > 0$, $\lambda_{\max} > \lambda_{\min} > 0$, $\beta_{\max} > \beta_{\min}$ des constantes. Dans le cas particulier où F appartient à la classe $\mathcal{H}(\alpha)$ des fonctions telles que $h(x|z) = \lambda_{\beta,z} + r(x|z)$, $|r(x|z)| \leqslant A\exp(-\alpha x)$, avec $\alpha > 0$, $\lambda_{\min} \leqslant \lambda \leqslant \lambda_{\max}$, $\beta_{\min} \leqslant \beta \leqslant \beta_{\max}$, on peut donner la vitesse explicite en n. En effet, supposons la condition suivante concernant les censures :

$$h_C(x) = \gamma + r_C(x), \quad |r_C(x)| \leqslant Mx^{-\mu}, \quad \mu > 0, \ \gamma_{\min} \leqslant \gamma \leqslant \gamma_{\max}, \tag{0.2}$$

où $M > 0$, $\gamma_{\max} > \gamma_{\min} > 0$ sont des constantes. Notons $\lambda_\beta = \lambda \max\{1, e^\beta\}$.

Théorème 0.3. *Soit $F \in \mathcal{H}(\alpha)$ et la condition* (0.2) *réalisée. Supposons qu'il existe deux constantes $q_0 > 0$ et $\kappa > 0$ telles que $q_{C,F}(t) \geqslant q_0$, pour $t \geqslant x_0$, et $\kappa \leqslant n_z/n \leqslant 1 - \kappa$. Alors, lorsque $z = 0, 1$, on a $\mathcal{K}(\hat{\lambda}_{z,\tau_n}, \lambda_{\beta,z}) = O_{\mathbb{P}_F}(n^{-\frac{2\alpha}{\lambda_\beta + \gamma + 2\alpha}} \ln n)$, où $\tau_n = \frac{\ln n}{\lambda_\beta + \gamma + 2\alpha} + O(1)$.*

1. Introduction and main results

The non-parametric Cox proportional hazard model has become a common tool for predicting the time to event data in various applied fields (see for example [7,6,1] and the references therein, among many others). However it is well known that the estimated survival probabilities become very unstable when the life time is close to the right end of the observed data range. In the present paper we propose a new approach for estimating the survival probabilities from the censored failure time data $t_1, \ldots, t_n$, when the life time is close to or even exceeds $\max\{t_1, \ldots, t_n\}$. Our idea consists in adjusting a Cox model to the tail of the unknown lifetime distribution and is similar to that used in [2] and [3].

Assume that a random variable (r.v.) X with values in $[x_0, \infty)$, $x_0 \geqslant 0$, has the distribution function (d.f.) F, the density function f_F and the survival function $S_F(x) = 1 - F(x)$, $x \geqslant x_0$. By $h_F = f_F/S_F$ we denote its hazard function. Consider a right random censoring time C with values in $[x_0, \infty)$ admitting the density f_C. Denote by F_C its d.f. and let $S_C(x) = 1 - F_C(x)$, $x \geqslant x_0$, and $h_C = f_C/S_C$. Let $P_F(x, \delta)$, $x \geqslant x_0$, $\delta \in \{0, 1\}$ be the law of the random vector (T, Δ), where $T = \min\{X, C\}$ and $\Delta = 1 \ (X \leqslant C)$. The density of P_F is given by

$$p_F(x, \delta) = f_F(x)^\delta S_F(x)^{1-\delta} f_C(x)^{1-\delta} S_C(x)^\delta, \quad x \geqslant x_0, \ \delta \in \{0, 1\}.$$

Let $X_1, \ldots, X_n$ and $C_1, \ldots, C_n$ be two independent samples from X and C and let $T_i = \min\{X_i, C_i\}$ and $\Delta_i = 1 \ (X_i \leqslant C_i)$. Denote by $\mathbb{P}_F$ the joint law of the sample $Y_i = (T_i, \Delta_i)$, $i = 1, \ldots, n$ and by $\mathbb{E}_F$ the expectation w.r.t. $\mathbb{P}_F$ on $\mathbb{Y}^n$, where $\mathbb{Y} = [x_0, \infty) \times \{0, 1\}$.

Assume that

$$S_F(x) = \exp\left(-\int_{x_0}^{x} h_F(v)\,dv\right),$$

where $h_F(v) = h(v|z)$ is supposed to depend on the values z of a covariate Z with values in the set $\{0, 1\}$. In the sequel we consider that we are given a sequence of realizations $z_1, \ldots, z_n$ of the covariate Z and that all the reasonings are made conditionally w.r.t. these realizations.

Since the hazard function $h(\cdot|z)$ is not supposed to have a particular structure, the non-parametric estimator of the proposed model does not provide reliable estimation for the survival probabilities $S_F(x)$ for large values of x. In particular, when $x \geqslant \max\{t_1, \ldots, t_n\}$ for a given sample $t_1, \ldots, t_n$ of $T_1, \ldots, T_n$, these predictions are not meaningful. To ensure better predictive properties, we shall assume the Cox model (see for instance [7]) $h(x|z) = \lambda_{\beta,z} \equiv \lambda e^{\beta z} > 0$, $x \geqslant t$, for some unknown threshold level $t \geqslant x_0$. For any $\lambda > 0$, $\beta \in \mathbb{R}$ and $t \geqslant x_0$, denote by $F_{\lambda,\beta,t}(x|z)$ the distribution function defined by

$$1 - F_{\lambda,\beta,t}(x|z) = \begin{cases} S_F(x), & x \in [x_0, t], \\ S_F(t)\exp(-\lambda e^{\beta z}(x - t)), & x > t. \end{cases}$$

For any $t \geqslant x_0$ define quasi-likelihood by $L_n(F_{\lambda,\beta,t}) = \sum_{i=1}^{n} \ln p_{F_{\lambda,\beta,t}}(T_i, \Delta_i)$. For $z = 0, 1$, denote $\hat{n}_{z,t} = \sum_{T_i \geqslant t, Z_i = z} \Delta_i$. Maximizing $L_n(F_{\lambda,\beta,t})$ in $\lambda_{\beta,0} = \lambda$ and in $\lambda_{\beta,1} = \lambda e^{\beta}$ gives the local quasi-likelihood estimators (with $0/0 = 0$)

$$\hat{\lambda}_{z,t} = \frac{\hat{n}_{z,t}}{\sum_{T_i \geqslant t, Z_i = z}(T_i - t)}, \quad z = 0, 1.$$

Let $\mathcal{K}(\lambda_1, \lambda)$ be Kullback–Leibler divergence between two exponential laws with parameters $\lambda_1 > 0$ and $\lambda > 0$ respectively. It is easy to verify that $\mathcal{K}(\lambda_1, \lambda) = G(\frac{\lambda}{\lambda_1} - 1)$, where $G(x) = x - \ln(x + 1)$ for $x > -1$. By continuity we set $\mathcal{K}(0, \lambda) = +\infty$. Consider the χ^2 entropy between two equivalent probability measures P and P_0:

$$\chi^2(P, P_0) = \int_{\mathbb{Y}} \frac{dP}{dP_0}\,dP - 1.$$

We say that the tail of the distribution function F belongs to the domain of attraction of Cox model if there exist $\lambda > 0$ and $\beta \in \mathbb{R}$ such that $\lim_{t \to \infty} \chi^2(P_F, P_{F_{\lambda,\beta,t}}) = 0$. It is easy to see that if the tail of the distribution function F belongs to the domain of attraction of Cox model then there is a sequence τ_n such that the following condition is satisfied

$$\chi^2(P_F, P_{F_{\lambda,\beta,\tau_n}}) = O\left(\frac{\ln n}{n}\right). \tag{1.1}$$

Now we are in position to formulate our main results. For non-negative r.v.'s ξ_n and η_n the notation $\xi_n = O_{\mathbb{P}_F}(\eta_n)$ means that there exists $c > 0$ such that $\mathbb{P}_F(\xi_n/\eta_n \leqslant c) \to 1$ as $n \to \infty$ (with $\frac{+\infty}{+\infty} = 0$).

Theorem 1.1. *If the distribution function F belongs to the domain of attraction of Cox model and τ_n is a sequence satisfying* (1.1)*, then, for $z = 0, 1$,*

$$\mathcal{K}(\hat{\lambda}_{z,\tau_n}, \lambda_{\beta,z}) = O_{\mathbb{P}_F}\left(\frac{\ln n}{\hat{n}_{z,\tau_n}}\right).$$

Denote $q_{C,F}(t) = \int_t^{\infty} \frac{S_C(x)}{S_C(t)} \frac{f_F(x)}{S_F(t)}\,dx$. Remark that $q_{C,F}(t)$ is the proportion of the non-censored survival times among all survival times exceeding the threshold t. Denote $n_z = \sum_{i=1}^{n} 1\,(Z_i = z)$. Then n_z/n is the proportion of observed times associated to the value z of the covariate Z. The following theorem is a consequence of Theorem 1.1.

Theorem 1.2. *Assume there exist constants $q_0 > 0$ and $\kappa > 0$ such that $q_{C,F}(t) \geqslant q_0$, for $t \geqslant x_0$, and $\kappa \leqslant n_z/n \leqslant 1 - \kappa$. If the distribution function F belongs to the domain of attraction of Cox model and the sequence τ_n is satisfying* (1.1) *and $nS_C(\tau_n)S_F(\tau_n) \to \infty$ as $n \to \infty$, then $\mathbb{E}_F \hat{n}_{z,\tau_n} \geqslant c_0 \kappa n S_C(\tau_n)S_F(\tau_n)$, for some constant $c_0 > 0$, $\mathbb{P}_F(\hat{n}_{z,\tau_n} \geqslant \frac{1}{2}\mathbb{E}_F \hat{n}_{z,\tau_n}) \to 1$ as $n \to \infty$, and*

$$\mathcal{K}(\hat{\lambda}_{z,\tau_n}, \lambda_{\beta,z}) = O_{\mathbb{P}_F}\left(\frac{\ln n}{nS_C(\tau_n)S_F(\tau_n)}\right).$$

Remark that the condition $q_{C,F}(t) \geqslant q_0$, for $t \geqslant x_0$, implies the identifiability of the model, for any fixed $t \geqslant x_0$.

2. Explicit rate of convergence

Let $A > 0$, $\lambda_{\max} > \lambda_{\min} > 0$, $\beta_{\max} > \beta_{\min}$ be some constants. We shall give the rate of convergence in the case when F belongs to the class of distribution functions $\mathcal{H}(\alpha)$ with hazard function of the form $h(x|z) = \lambda_{\beta,z} + r(x|z)$, $|r(x|z)| \leqslant A \exp(-\alpha x)$, where $\alpha > 0$, $\lambda_{\min} \leqslant \lambda \leqslant \lambda_{\max}$, $\beta_{\min} \leqslant \beta \leqslant \beta_{\max}$. This class is similar to the class of functions studied in [4] and [5] for the extreme values models. To do so we shall impose the following condition on the censoring variable C:

$$h_C(x) = \gamma + r_C(x), \quad |r_C(x)| \leqslant M x^{-\mu}, \quad \mu > 0, \ \gamma_{\min} \leqslant \gamma \leqslant \gamma_{\max}, \tag{2.1}$$

where $M > 0$, $\gamma_{\max} > \gamma_{\min} > 0$ are some given constants. Denote $\lambda_\beta = \lambda \max\{1, e^\beta\}$.

Theorem 2.1. *Assume $F \in \mathcal{H}(\alpha)$ and condition* (2.1). *Suppose that there exist constants $q_0 > 0$ and $\kappa > 0$ such that $q_{C,F}(t) \geqslant q_0$, for $t \geqslant x_0$, and $\kappa \leqslant n_z/n \leqslant 1 - \kappa$. Then, for $z = 0, 1$,*

$$\mathcal{K}(\hat{\lambda}_{z,\tau_n}, \lambda_{\beta,z}) = O_{\mathbb{P}_F}\left(n^{-\frac{2\alpha}{\lambda_\beta + \gamma + 2\alpha}} \ln n\right),$$

where $\tau_n = \frac{\ln n}{\lambda_\beta + \gamma + 2\alpha} + O(1)$.

A consistent estimator of the parameter $\beta = \ln \frac{\lambda_{\beta,1}}{\lambda_{\beta,0}}$ is given by $\hat{\beta}_{\tau_n} = \ln \frac{\hat{\lambda}_{1,\tau_n}}{\hat{\lambda}_{0,\tau_n}}$. A rate of convergence result for $\hat{\beta}_{\tau_n}$ is straightforward.

3. Proofs of the main results

Proposition 3.1. *For any $\lambda, \lambda_1 > 0$, $\beta, \beta_1 \in \mathbb{R}$ and any $x \geqslant 0$,*

$$\mathbb{P}_F\left(\sum_{i=1}^n \ln \frac{p_{F_{\lambda_1,\beta_1,t}}}{p_{F_{\lambda,\beta,t}}}(Y_i) \geqslant x + n\chi^2(P_F, P_{F_{\lambda,\beta,t}})\right) \leqslant e^{-\frac{x}{2}}.$$

The proof of the proposition is similar to that of Proposition 7.1 in [3] and therefore will be not detailed here.

Proposition 3.2. *For any $\lambda > 0$, $\beta \in \mathbb{R}$, $t \geqslant x_0$ and any $x \geqslant 0$ it holds*

$$\mathbb{P}_F\left(\hat{n}_{z,t} \mathcal{K}(\hat{\lambda}_{z,t}, \lambda_{\beta,z}) \geqslant x + n\chi^2(P_F, P_{F_{\lambda,\beta,t}}) + 2\ln n\right) \leqslant 2e^{-\frac{x}{2}},$$

where $z = 0, 1$.

The assertion of Theorem 1.1 follows immediately from Proposition 3.2 if we set $x = 2\ln n$.

3.1. Proof of Proposition 3.2

Note that

$$\sum_{i=1}^n \ln \frac{p_{F_{\lambda_1,\beta_1,t}}}{p_{F_{\lambda,\beta,t}}}(Y_i) = \sum_{T_i \geqslant t} \Delta_i \ln \frac{\lambda_1 e^{\beta_1 z_i}}{\lambda e^{\beta z_i}} - \sum_{T_i \geqslant t} (\lambda_1 e^{\beta_1 z_i} - \lambda e^{\beta z_i})(T_i - t).$$

Stratifying w.r.t. values of Z_i we get

$$\sum_{i=1}^n \ln \frac{p_{F_{\lambda_1,\beta_1,t}}}{p_{F_{\lambda,\beta,t}}}(Y_i) = \hat{n}_{0,t} \Lambda_0(\lambda_1, \lambda_{\beta,0}) + \hat{n}_{1,t} \Lambda_1(\lambda_1 e^{\beta_1}, \lambda_{\beta,1}),$$

where $\Lambda_z(\lambda', \lambda'') = \ln \frac{\lambda'}{\lambda''} - (\lambda' - \lambda'')\frac{1}{\lambda_{z,t}}$, $z = 0, 1$. Denote for brevity $y = x + n\chi^2(P_F, P_{F_{\lambda,\beta,t}}) \geqslant 0$ and $g(u, k) = \frac{\ln \frac{u}{\lambda_{\beta,z}} - \frac{y}{k}}{u - \lambda_{\beta,z}}$. Let $\lambda^+(k) = \arg\max_{u \geqslant \lambda_{\beta,z}} g(u, k)$ and $\lambda^-(k) = \arg\min_{u \leqslant \lambda_{\beta,z}} g(u, k)$. It is easy to see that

$$\begin{aligned}
\{\hat{n}_{z,t} \Lambda_z(\hat{\lambda}_{z,t}, \lambda_{\beta,z}) \geqslant y, \ \hat{\lambda}_{z,t} \geqslant \lambda_{\beta,z}\} &= \{g(\hat{\lambda}_{z,t}, \hat{n}_{z,t}) \geqslant \hat{\lambda}_{z,t}, \ \hat{\lambda}_{z,t} \geqslant \lambda_{\beta,z}\} \\
&\subset \{g(\lambda^+(\hat{n}_{z,t}), \hat{n}_{z,t}) \geqslant \hat{\lambda}_{z,t}, \ \hat{\lambda}_{z,t} \geqslant \lambda_{\beta,z}\} \\
&= \{\hat{n}_{z,t} \Lambda_z(\lambda^+(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y, \ \hat{\lambda}_{z,t} \geqslant \lambda_{\beta,z}\} \\
&\subset \{\hat{n}_{z,t} \Lambda_z(\lambda^+(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y\}.
\end{aligned}$$

In the same way we get

$$\{\hat{n}_{z,t}\Lambda_z(\hat{\lambda}_{z,t}, \lambda_{\beta,z}) \geqslant y, \ \hat{\lambda}_{z,t} < \lambda_{\beta,z}\} \subset \{\hat{n}_{z,t}\Lambda_z(\lambda^-(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y\}.$$

These inclusions imply

$$\{\hat{n}_{z,t}\mathcal{K}(\hat{\lambda}_{z,t}, \lambda_{\beta,z}) \geqslant y\} = \{\hat{n}_{z,t}\Lambda_z(\hat{\lambda}_{z,t}, \lambda_{\beta,z}) \geqslant y\}$$
$$\subset \{\hat{n}_{z,t}\Lambda_z(\lambda^+(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y\} \cup \{\hat{n}_{z,t}\Lambda_z(\lambda^-(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y\}. \tag{3.1}$$

From (3.1) we get

$$\mathbb{P}_F(\hat{n}_{z,t}\mathcal{K}(\hat{\lambda}_{z,t}, \lambda_{\beta,z}) \geqslant y) \leqslant \mathbb{P}_F(\hat{n}_{z,t}\Lambda_z(\lambda^+(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y) + \mathbb{P}_F(\hat{n}_{z,t}\Lambda_z(\lambda^-(\hat{n}_{z,t}), \lambda_{\beta,z}) \geqslant y)$$
$$\leqslant \sum_{k=1}^n \mathbb{P}_F(\hat{n}_{z,t}\Lambda_z(\lambda^+(k), \lambda_{\beta,z}) \geqslant y) + \sum_{k=1}^n \mathbb{P}_F(\hat{n}_{z,t}\Lambda_z(\lambda^-(k), \lambda_{\beta,z}) \geqslant y). \tag{3.2}$$

Suppose $z = 0$. Choosing β_1 such that $\lambda_1 e^{\beta_1} = \lambda e^\beta$, from Proposition 3.1, it follows

$$\mathbb{P}_F(\hat{n}_{0,t}\Lambda_0(\lambda_1, \lambda_{\beta,0}) \geqslant y) \leqslant e^{-\frac{x}{2}}, \tag{3.3}$$

for any $\lambda_1 > 0$. Then, by (3.2) and (3.3),

$$\mathbb{P}_F(\hat{n}_{z,t}\mathcal{K}(\hat{\lambda}_{0,t}, \lambda_{\beta,0}) \geqslant y) \leqslant 2ne^{-x/2} = 2e^{-x/2 + \ln n}.$$

The proposition is proved for $z = 0$. The proof in the case $z = 1$ is similar.

3.2. Sketch of the proof of Theorem 2.1

Theorem 2.1 is a consequence of Theorem 1.1 if we show condition (1.1). Consider the distance $\rho_t(h_1, h_2) = \sup_{x \geqslant t} |h_1(x) - h_2(x)|$, where h_1, h_2 are two positive functions on $[x_0, \infty)$ and $t \geqslant x_0$. By Lemma 8.5 in [3],

$$\chi^2(P_F, P_{F_{\lambda,\beta,\tau_n}}) \leqslant P_F\left(\ln \frac{dP_F}{dP_{F_{\lambda,\beta,\tau_n}}}\right)^2 \exp\left(\left|\ln \frac{dP_F}{dP_{F_{\lambda,\beta,\tau_n}}}\right|\right).$$

Using the fact that $F \in \mathcal{H}(\alpha)$ and condition (2.1) we obtain

$$\chi^2(P_F, P_{F_{\lambda,\beta,\tau_n}}) = O(S_C(\tau_n)S_F(\tau_n)\rho_{\tau_n}^2)$$
$$= O\left(\exp\left(-(\gamma + \lambda_{\beta,z} + 2\alpha)\tau_n + \frac{A}{\alpha}e^{-\alpha\tau_n} + M\tau_n^{-\mu}\right)\right),$$

where $\rho_{\tau_n} = \max_{z \in \{0,1\}} \rho_{\tau_n}(h(\cdot|z), \lambda e^{\beta z})$. This implies that (1.1) holds with $\tau_n = \frac{\ln n}{\lambda_\beta + \gamma + 2\alpha} + O(1)$.

References

[1] J.-F. Dupuy, I. Grama, M. Mesbah, Asymptotic theory for the Cox model with missing time dependent covariate, Ann. Statist. 34 (2006) 903–924.
[2] I. Grama, V. Spokoiny, Pareto approximation of the tail by local exponential modeling, Bull. Acad. Sci. Moldova 1 (53) (2007) 1–22.
[3] I. Grama, V. Spokoiny, Statistics of extremes by oracle estimation, Ann. Statist. 36 (2008) 1619–1648.
[4] P. Hall, A.H. Welsh, Best attainable rates of convergence for estimates of parameters of regular variation, Ann. Statist. 12 (1984) 1079–1084.
[5] P. Hall, A.H. Welsh, Adaptive estimates of regular variation, Ann. Statist. 13 (1985) 331–341.
[6] J.D. Kalbfleisch, R.L. Prentice, The Statistical Analysis of Failure Time Data, Wiley, 2002.
[7] J.P. Klein, M.L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, Springer, 2003.