

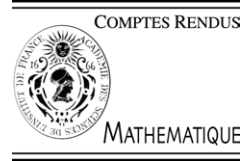


ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

C. R. Acad. Sci. Paris, Ser. I 341 (2005) 49–52



<http://france.elsevier.com/direct/CRASSI/>

Probabilités

# Estimation dans un modèle de Koziol–Green généralisé

Ségolen Geffray, Agathe Guilloux

LSTA, université Paris 6, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 28 juin 2004 ; accepté après révision le 3 mai 2005

Disponible sur Internet le 23 juin 2005

Présenté par Paul Deheuvels

## Résumé

Nous considérons une population dans laquelle la durée de vie  $X$  est censurée à droite de façon informative par une v.a. positive  $Y$ , on obtient alors une v.a.  $Z = X \wedge Y$ . Cette v.a.  $Z$  est à son tour censurée à droite de façon non informative par une v.a. positive  $C$ . Dans le modèle de Koziol–Green généralisé, c'est-à-dire en supposant que  $X$  et  $Y$  ont des fonctions de risques proportionnelles, nous introduisons un estimateur  $\hat{\Lambda}_X$  de la fonction de risque cumulé  $\Lambda_X$  de  $X$  et étudions ses propriétés asymptotiques. *Pour citer cet article : S. Geffray, A. Guilloux, C. R. Acad. Sci. Paris, Ser. I 341 (2005).*

© 2005 Académie des sciences. Publié par Elsevier SAS. Tous droits réservés.

## Abstract

**Estimation in a generalized Koziol–Green model.** We consider a population where the lifetime  $X$  is informatively right-censored by a nonnegative r.v.  $Y$ , thus we obtain a r.v.  $Z = X \wedge Y$ . This r.v.  $Z$  is then noninformatively right-censored by a non-negative r.v.  $C$ . In a generalized Koziol–Green model, i.e. we assume that  $X$  and  $Y$  have proportional hazard functions, we introduce an estimator  $\hat{\Lambda}_X$  of the cumulative hazard function  $\Lambda_X$  of  $X$  and we study its asymptotic properties. *To cite this article: S. Geffray, A. Guilloux, C. R. Acad. Sci. Paris, Ser. I 341 (2005).*

© 2005 Académie des sciences. Publié par Elsevier SAS. Tous droits réservés.

## 1. Introduction

Soit  $X$  une variable aléatoire (v.a.) positive, de fonction de répartition  $F_X(x) = \mathbb{P}(X \leq x)$  supposée continue. Supposons que la v.a.  $X$  est censurée aléatoirement à droite par une v.a. positive  $Y$ , indépendante de  $X$ , de fonction de répartition continue  $F_Y(y) = \mathbb{P}(Y \leq y)$ . On note  $\Lambda_X = -\log(1 - F_X)$  et  $\Lambda_Y = -\log(1 - F_Y)$  les fonctions de risque cumulé de  $X$  et  $Y$  respectivement. Nous supposons que  $X$  et  $Y$  satisfont le modèle des fonctions de risque proportionnelles ou modèle de Koziol–Green, i.e. il existe un  $\beta > 0$  tel que

$$\Lambda_Y = \beta \Lambda_X. \tag{1}$$

Adresses e-mail : [geffray@ccr.jussieu.fr](mailto:geffray@ccr.jussieu.fr) (S. Geffray), [aguilloux@ccr.jussieu.fr](mailto:aguilloux@ccr.jussieu.fr) (A. Guilloux).

Nous supposons de plus que la v.a.  $Z = X \wedge Y$  est censurée à droite par une v.a.  $C$  indépendante de  $Z$ , de fonction de répartition  $F_C$  supposée continue. On observe alors, pour  $i = 1, \dots, n$ , les vecteurs :

$$(T_i = Z_i \wedge C_i, I(Z_i \leq C_i), I(Z_i \leq C_i)I(X_i \leq Y_i)).$$

Les v.a.  $Z_i$  sont i.i.d. de fonction de survie  $1 - F_Z = (1 - F_X)(1 - F_Y)$  et de fonction de risque cumulé  $\Lambda_Z = -\log(1 - F_Z)$ . Les v.a.  $T_i$  sont i.i.d. de fonction de survie  $1 - H = (1 - F_X)(1 - F_Y)(1 - F_C)$ . Sous l'hypothèse de proportionnalité des fonctions de risque de  $X$  et  $Y$ , la v.a.  $Y$  est appelé communément (cf. Andersen et al., [1] et Gather et Pawlitschko, [2]) *censure informative* et la v.a.  $C$  *censure non-informative*. Cette généralisation du modèle de Koziol–Green a été introduite par Gather et Pawlitschko [2], qui ont proposé un estimateur de  $1 - F_X$  et étudié ses propriétés asymptotiques.

Nous proposons un estimateur  $\hat{\Lambda}_X$  de la fonction de risque cumulé  $\Lambda_X$  et nous étudions ses propriétés asymptotiques : consistance uniforme et loi du logarithme itéré de l'estimateur  $\hat{\Lambda}_X$ , convergence faible et approximation forte du processus  $\sqrt{n}(\hat{\Lambda}_X - \Lambda_X)$ .

## 2. Estimation, consistance uniforme et loi du logarithme itéré (LLI)

Par indépendance de  $X$  et  $Y$ , la relation suivante est satisfaite :  $\Lambda_Z = \Lambda_X + \Lambda_Y$ . Avec l'Éq. (1) et en notant  $\alpha = 1/(\beta + 1)$ , on obtient :

$$\Lambda_X = \alpha \Lambda_Z. \tag{2}$$

Notons que l'on peut montrer que  $\alpha = \mathbb{P}(X \leq Y)$ . Nous supposons que  $\alpha \neq 0$  et que  $\mathbb{P}(Z \leq C) \neq 0$ . La relation (2) suggère de proposer un estimateur semi-paramétrique de  $\Lambda_X$  à savoir

$$\hat{\Lambda}_X = \hat{\alpha} \hat{\Lambda}_Z,$$

avec  $\hat{\alpha}$  un estimateur de  $\alpha$  et  $\hat{\Lambda}_Z$  un estimateur de  $\Lambda_Z$ .

On montre que l'égalité suivante est vérifiée :

$$\alpha = \mathbb{P}(X \leq Y, Z \leq C) / \mathbb{P}(Z \leq C).$$

Nous proposons alors d'estimer  $\alpha$  par le rapport des proportions empiriques correspondantes :

$$\hat{\alpha} = \frac{\sum_{i=1}^n I(Z_i \leq C_i)I(X_i \leq Y_i)}{\sum_{i=1}^n I(Z_i \leq C_i)}.$$

Par ailleurs, comme  $Z$  est non-informativement censurée à droite par  $C$ , un estimateur naturel de  $\Lambda_Z$  est celui de Nelson–Aalen défini, pour  $t \geq 0$ , par

$$\hat{\Lambda}_Z(t) = \sum_{i=1}^n \frac{I(T_i \leq t)I(Z_i \leq C_i)}{Y_n(T_i)},$$

où  $Y_n(t) = \sum_{i=1}^n I(T_i \geq t)$ , pour  $t \geq 0$ , est le processus « nombre d'individus à risque en  $t^-$  ». Grâce au résultat de consistance uniforme de l'estimateur de Nelson–Aalen (cf. Andersen et al., [1]), nous établissons le théorème suivant.

**Théorème 2.1.** *Consistence uniforme faible*

Pour tout  $\sigma \in \mathcal{T} = \{t \in [0, \infty[, F_X(t) < 1\}$ , quand  $n$  tend vers  $\infty$ ,

$$\sup_{t \in [0, \sigma]} |\hat{\Lambda}_X(t) - \Lambda_X(t)| \xrightarrow{\mathbb{P}} 0.$$

Considérons maintenant  $T_{1,n} \leq T_{2,n} \leq \dots \leq T_{n,n}$  les statistiques d'ordre de l'échantillon. Soit  $(k_n)$  une suite d'entiers satisfaisant la condition

$$\begin{aligned} \text{(H)} : 1 \leq k_n < n \text{ et } k_n \geq (\log n)^2, \text{ pour } n \text{ assez grand,} \\ \text{la suite } (k_n/n) \text{ est décroissante au sens large et} \\ k_n/n \leq Ak_{2n}/n \text{ pour un } A > 0. \end{aligned}$$

La LLI classique de Hartman et Wintner et la LLI pour l'estimateur de Nelson–Aalen de Giné et Guillou [3] nous permettent d'établir la LLI suivante.

**Théorème 2.2.** *Loi du logarithme itéré*

Soit  $(k_n)$  une suite d'entiers vérifiant la condition (H); on a alors :

$$\sup_{t \leq T_{n-k_n, n}} |\hat{\Lambda}_X(t) - \Lambda_X(t)| = \mathcal{O}\left(\sqrt{\frac{\log \log n}{k_n}}\right) \text{ p.s.}$$

**3. Convergence faible et approximation forte**

Nous obtenons deux résultats asymptotiques pour le processus  $\sqrt{n}(\hat{\Lambda}_X - \Lambda_X)$ .

**Théorème 3.1.** *Convergence faible*

Soit  $\sigma < \tau_H = \inf\{x: H(x) < 1\}$ . Quand  $n$  tend vers l'infini, on a, dans l'espace  $\mathbb{D}[0, \sigma[$  des fonctions càdlàg sur  $[0, \sigma[$ , la convergence faible suivante :

$$\sqrt{n}(\hat{\Lambda}_X(\cdot) - \Lambda_X(\cdot)) \xrightarrow{\mathcal{D}} L(\cdot),$$

où  $L$  est un processus gaussien de moyenne nulle et de fonction de covariance définie, pour tous  $s$  et  $t$  dans  $[0, \sigma[$ , par :

$$\text{Cov}(L(s), L(t)) = \alpha^2 \int_0^{s \wedge t} \frac{d\Lambda_Z}{1-H} + \frac{\alpha(1-\alpha)}{\mathbb{P}(Z \leq C)} \Lambda_Z(s) \Lambda_Z(t).$$

Ce théorème est établi, principalement, à l'aide du résultat de convergence faible de l'estimateur de Nelson–Aalen (cf. Andersen et al., [1]).

**Théorème 3.2.** *Approximation forte*

Soit  $(k_n)$  une suite d'entiers satisfaisant la condition (H). Pour  $n$  assez grand, il existe un espace de probabilité convenablement élargi sur lequel

$$\sup_{t \leq T_{n-k_n, n}} |\sqrt{n}(\hat{\Lambda}_X(t) - \Lambda_X(t)) - L_n(t)| = \mathcal{O}\left(\max\left\{n\sqrt{\log \log n} \left(\frac{\log n}{k_n}\right)^{3/2}, \frac{n^{5/6}(\log n)^{5/2}}{k_n}\right\}\right),$$

où, pour  $n \geq 0$ , le processus  $L_n$  est gaussien de moyenne nulle et de même fonction de covariance que  $L$  (cf. Théorème 3.1).

La preuve de ce théorème nécessite les résultats de Komlos, Major et Tusnady [4] sur l'approximation forte du processus empirique univarié, de Massart [5] sur l'approximation forte du processus empirique multivarié, de Shorack et Wellner [6] sur les grandes déviations du pont brownien et de Giné et Guillou [3] sur la LLI pour l'estimateur de Nelson–Aalen.

Le terme  $n\sqrt{\log \log n}(\log n/k_n)^{3/2}$  provient essentiellement du contrôle des grandes déviations du pont brownien et le terme  $n^{5/6}(\log n)^{5/2}/k_n$  de l'approximation du processus empirique multivarié.

## Références

- [1] P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding, *Statistical Models Based on Counting Processes*, Springer-Verlag, 1993.
- [2] U. Gather, J. Pawlitschko, Estimating the survival function under a generalized Koziol–Green model with partially informative censoring, *Metrika* 48 (1998) 189–207.
- [3] E. Giné, A. Guillou, Laws of the iterated logarithm for censored data, *Ann. Probab.* 27 (1999) 2042–2067.
- [4] J. Komlós, P. Major, G. Tusnady, An approximation of partial sums of independent r.v.'s. and the sample d.f. I, *Z.W.V.G.* 32 (1975) 111–132.
- [5] P. Massart, Strong approximation for multivariate empirical and related processes via KMT constructions, *Ann. Probab.* 17 (1989) 266–291.
- [6] G.R. Shorack, J.A. Wellner, *Empirical Processes with Statistical Application*, Wiley, 1986.