



HAL
open science

Estimation De La Fonction De Hasard Pour Un Processus Fortement Mélangeant Avec Censure

Jean Lecoutre, Elias Ould-Said

► **To cite this version:**

Jean Lecoutre, Elias Ould-Said. Estimation De La Fonction De Hasard Pour Un Processus Fortement Mélangeant Avec Censure. Annales de l'ISUP, 1993, XXXVII (1-2), pp.71-99. hal-03664830

HAL Id: hal-03664830

<https://hal.science/hal-03664830v1>

Submitted on 11 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pub. Inst. Stat. Univ. Paris
XXXVII, fasc. 1-2, 1993, 71 à 99

ESTIMATION DE LA FONCTION DE HASARD
POUR UN PROCESSUS FORTEMENT MELANGEANT AVEC CENSURE.

PAR

Jean Pierre LECOUTRE, Elias OULD – SAID

*Laboratoire de Statistique Theorique et Appliquée,
Universite Paris VI, 4 place Jussieu, 75252 Paris Cedex 05.*

Résumé: On étudie les propriétés de convergence simple et uniforme d'un estimateur du noyau de la fonction de hasard, dans le cas de variables aléatoires multidimensionnelles, issues d'un processus fortement mélangeant, avec censure aléatoire.

Mots clés: fonction de hasard, censure, forte mélangeance.

Abstract. It is shown that a kernel-type estimate for the hazard rate is pointwise, as well as uniformly, strongly consistent, from a strongly mixing process and randomly right-censored samples.

A.M.S. Subject Classification: 62G05, 62N05.

1. Introduction

Dans l'analyse statistique des durées de vie, on considère généralement des variables aléatoires (v.a.) réelles positives et indépendantes (cf J.J. Droesbeke et al, 1989). Cependant, dans certaines applications, la durée passée dans un état donné, ou la durée séparant deux événements est relative à un système ou à un groupe d'individus, et sera représentée par une v.a. X à valeurs dans \mathbf{R}_+^s , $s \in \mathbf{N}^*$. Dans ce cas, la fonction de hasard h n'admet pas une définition unique (cf D. Shanbhag, S. Kotz, 1987). Par analogie avec le cas unidimensionnel ($s=1$), nous définirons h à partir de la fonction de répartition (f.r.) F , et de la densité de probabilité f , supposée bornée, de X par:

$$h(t) = \frac{f(t)}{1 - F(t)},$$

pour toutes les valeurs de t du sous-ensemble \mathbf{E} de \mathbf{R}^s où $F(t) < 1$.

La relation d'ordre utilisée sur \mathbf{R}^s est définie à partir des coordonnées par :

$$\forall x \in \mathbf{R}^s, \forall y \in \mathbf{R}^s, x < y \iff x_i < y_i \text{ pour } 1 \leq i \leq s.$$

De nombreuses méthodes d'estimation de h , à partir d'observations X_1, X_2, \dots, X_n ont été proposées et étudiées, dans le cas $s = 1$ et pour des observations indépendantes (voir par exemple *N.Singpurwalla, M.Wong (1983), S.Hassani et al (1986)*). Plus récemment, certains auteurs (*G.Roussas (1989), L.Györfi et al (1989)*) ont établi des résultats de convergence forte d'estimateurs à noyau dans le cas $s > 1$ et pour des observations dépendantes. En effet, les durées observées peu avant peuvent influencer la durée actuelle, ce qui correspond à une hypothèse de mélangeance.

Par ailleurs, certaines durées pouvant ne pas être observées, nous nous proposons ici d'obtenir les mêmes résultats sous une hypothèse plus faible de mélangeance, et dans le cas où la variable X , est censurée à droite par une v.a. positive C de f.r. G et de densité g .

On supposera que $G < 1$ sur \mathbf{E} et, pour des raisons d'identifiabilité, que les variables X et C sont indépendantes.

Les variables observables sont donc les couples

$$T_i = \min(X_i, C_i) \quad \text{et} \quad \delta_i = I_{X_i \leq C_i}, \quad 1 \leq i \leq n$$

On fera enfin l'hypothèse suivante :

H1. Le processus $(T_i)_{i \in \mathbf{N}}$ est strictement stationnaire et α -mélangeant (ou fortement mélangeant).

On rappelle que cela signifie que :

$$\alpha_n = \sup\{|P(A \cap B) - P(A)P(B)| \mid A \in \mathcal{M}_0^t, B \in \mathcal{M}_{t+n}^\infty\} \longrightarrow 0, \quad n \rightarrow \infty$$

où \mathcal{M}_0^t (resp. \mathcal{M}_{t+n}^∞) désigne la tribu engendrée par $(T_s; s \leq t)$ (resp. $(T_s; s \geq t+n)$).

Si on note $f^* = f(1 - G)$ la densité commune des v.a. $(T_i, \delta_i = 1)_{i \in \mathbf{N}}$ et L la f.r. de $T = \min(X, C)$, la fonction de hasard peut s'écrire sous la forme :

$$\forall t \in \mathbf{E}, \quad h(t) = \frac{f^*(t)}{1 - L(t)}.$$

Un estimateur naturel de h sera alors :

$$\forall t \in \mathbf{E}, \quad h_n(t) = \frac{f_n^*(t)}{1 - L_n(t)},$$

où L_n est la fonction de répartition empirique de (T_1, T_2, \dots, T_n) et f_n^* un estimateur quelconque de la densité f^* . Pour des estimateurs de ce type, dans le cas de v.a. indépendantes, voir par exemple *J.J. Dreesbeke et al. (1989)*.

Nous retiendrons ici comme estimateur de f^* :

$$\forall t \in \mathbf{E}, \quad f_n^*(t) = \frac{1}{nb_n^s} \sum_{i=1}^n \delta_i K\left(\frac{t - T_i}{b_n}\right),$$

où K est un noyau de *Parzen-Rosenblatt* (cf *D. Bosq, J.P. Lecoutre, 1987*) et $(b_n)_{n \in \mathbf{N}}$ une suite de nombres réels strictements positifs, tels que

$$\lim_{n \rightarrow \infty} b_n = 0.$$

2. Principaux résultats

Nous pouvons écrire :

$$\forall t \in \mathbf{E}, \quad h_n(t) - h(t) = \frac{1}{D_n(t)} [(f_n^*(t) - f^*(t))(1 - L(t)) + f^*(t)(L_n(t) - L(t))] \quad (2.1)$$

en ayant posé

$$D_n(t) = [1 - L(t)][1 - L_n(t)].$$

Le processus $(T_i)_{i \in \mathbf{N}}$ étant strictement stationnaire, et l'ergodicité se déduisant de la mélangeance, on obtient:

$$\forall t \in \mathbf{E}, \quad L_n(t) = \frac{1}{n} \sum_{i=1}^n I_{]-\infty, t](T_i) \xrightarrow{p.co.} L(t), \quad n \rightarrow \infty.$$

Ainsi, pour tout t dans \mathbf{E} et pour n assez grand, on aura :

$$D_n(t) > 0 \quad p.s.$$

L'expression (2.1) est alors définie p.s et permet de déduire la convergence de h_n vers h de celle de f_n^* vers f^* .

Théorème 1.

Sous l'hypothèse H1, si on choisit la suite (b_n) telle que :

$$\frac{nb_n^2}{m_n} = \infty(\text{Log}n),$$

où m_n est une suite croissante, non majorée, avec $1 \leq m_n \leq n/2$ et telle qu'il existe une constante finie A vérifiant

$$\frac{n}{m_n} \left(\frac{1}{m_n} \sum_{j=1}^{m_n} \alpha_j + 2\sqrt{e} \alpha_{m_n}^{2m_n/3n} \right) < A,$$

alors, en tout point de \mathbf{E} où f^* est continue :

$$h_n(t) \xrightarrow{p.co.} h(t) \quad , \quad n \rightarrow \infty.$$

La suite m_n vérifiera dans la suite les conditions du théorème 1.

Pour obtenir la convergence uniforme, on introduit l'hypothèse suivante:

H2. La densité f^* est uniformément continue sur un certain compact C.

Théorème 2.

Sous les hypothèses H1 et H2, si le noyau K est lipschitzien, en choisissant une suite (b_n) vérifiant la condition du théorème 1, on obtient

$$\sup_{t \in C} |h_n(t) - h(t)| \xrightarrow{p.co.} 0, n \rightarrow \infty.$$

Pour préciser la vitesse de convergence, nous aurons besoin d'une hypothèse complémentaire.

H3. Toutes les dérivées partielles secondes de f^* sont uniformément bornées dans C.

Théorème 3.

Sous les hypothèses H1, H2 et H3, si le noyau K est pair, lipschitzien, avec $\int u_i^2 |K(u)| du < \infty$ pour $1 \leq i \leq s$, avec $u = (u_1, u_2, \dots, u_s)$, alors :

$$\theta_n^{-1} \sup_{t \in C} |h_n(t) - h(t)| = O(1) \text{ p.co.}, n \rightarrow \infty$$

si on choisit la suite b_n de telle sorte que

$$\theta_n = \frac{m_n \text{Log} n}{n b_n^s} \rightarrow 0, \theta_n m_n \rightarrow \infty, n \rightarrow \infty$$

et il existe une constante finie, positive D telle que

$$\forall n \in \mathbb{N}, b_n^2 \theta_n^{-1} \leq D$$

Remarque: les résultats précédents sont identiques à ceux obtenus par L.Györfi et al (1989) dans le cas non censuré, sous des hypothèses un peu différentes, relatives ici à f^* au lieu de f .

3. Démonstrations

3.1 Démonstration du théorème 1.

Il suffit de montrer que

$$f_n^*(t) \xrightarrow{p.co.} f^*(t), \quad n \rightarrow \infty.$$

On peut écrire :

$$Ef_n^*(t) - f^*(t) = \frac{1}{b_n^s} \int [f^*(u) - f^*(t)] K\left(\frac{u-t}{b_n}\right) du ; \quad (3.1)$$

donc, par application du lemme de Bochner (*D.Bosq, J.P. Lecoutre, 1987*):

$$Ef_n^*(t) - f^*(t) \rightarrow 0, \quad n \rightarrow \infty.$$

Il reste seulement à montrer que

$$f_n^*(t) - Ef_n^*(t) \xrightarrow{p.co.} 0.$$

Pour cela on écrit

$$f_n^*(t) - Ef_n^*(t) = \sum_{i=1}^n Z_i,$$

où

$$Z_i = \frac{1}{nb_n^s} \delta_i K\left(\frac{t-T_i}{b_n}\right) - \frac{1}{nb_n^s} E[\delta_i K\left(\frac{t-T_i}{b_n}\right)] \quad 1 \leq i \leq n \quad (3.2)$$

Par construction : $EZ_i=0$.

Nous allons alors appliquer un lemme dû à *M.Carbon (1988)*

Lemme 1.

Soit Z_i un processus réel fortement mélangeant, tel que

$$\forall i \in \mathbf{N}, E Z_i = 0 ;$$

$$\forall n \in \mathbf{N}, \forall i \in \mathbf{N}, 1 \leq i \leq n, |Z_i| \leq d_n, E Z_i^2 \leq D_n.$$

Alors : $\forall \epsilon > 0, \forall n \in \mathbf{N}, n \geq 3$

$$P\left\{\left|\sum_{i=1}^n Z_i\right| > \epsilon\right\} \leq 2 \exp\{-\gamma \epsilon + 4\gamma^2 n e [D_n + 8d_n^2 \sum_{j=1}^k \alpha_j] + 2\sqrt{e} \alpha_k^{2k/3n} \frac{n}{k}\},$$

où k est un entier inférieur à $n/2$, γ vérifiant

$$0 < \gamma \leq \frac{1}{4kd_n e}.$$

Si \bar{K} et B sont des majorants respectifs de K et de f :

$$|Z_i| \leq \frac{2\bar{K}}{nb_n^s}, \quad E(Z_i^2) \leq \frac{2B\bar{K}}{n^2 b_n^s} = D_n.$$

On applique alors le lemme 1 en prenant $k=m_n$ et $\gamma = 1/4m_n d_n e$:

$$P\left\{\left|\sum_{i=1}^n Z_i\right| > \epsilon\right\} \leq 2 \exp\left\{-\frac{nb_n^s}{8e\bar{K}m_n} \left(\epsilon - \frac{B}{m_n}\right) + \frac{2n}{em_n} \left(\frac{1}{m_n} \sum_{j=1}^k \alpha_j + 2\sqrt{e}(\alpha(m_n))^{2m_n/3n}\right)\right\}$$

Pour n assez grand, on aura :

$$\frac{B}{m_n} < \frac{\epsilon}{2};$$

d'après les hypothèses du théorème 1, on peut donc trouver deux constantes C_1 et C_2 telles que

$$P\left\{\left|\sum_{i=1}^n Z_i\right| > \epsilon\right\} \leq C_1 \exp(-C_2 \epsilon \frac{nb_n^s}{m_n}) \quad (3.3)$$

le choix de b_n montre que

$$\sum_n P\left\{\left|\sum_{i=1}^n Z_i\right| > \epsilon\right\} < \infty.$$

3.2 Démonstration du théorème 2

Le lemme précédent permet d'obtenir une vitesse de convergence dans le théorème de *Glivenko-Cantelli*.

Lemme 2.

Soit $\theta_n = \sqrt{\frac{m_n \text{Log} n}{n}}$; si $\theta_n = o(1)$, alors :

$$\frac{1}{\theta_n} \sup_{t \in C} |L_n(t) - L(t)| = O(1) \quad \text{p.co } n \rightarrow \infty.$$

D'après la décomposition (2.1), il suffit alors de montrer que

$$\sup_{t \in C} |f_n^*(t) - f^*(t)| \xrightarrow{\text{p.co.}} 0, \quad n \rightarrow \infty.$$

L'application du lemme de *Bochner* dans (3.1) permet d'obtenir :

$$\sup_{t \in C} |E f_n^*(t) - f^*(t)| \rightarrow 0, \quad n \rightarrow \infty.$$

Il ne reste donc plus qu'à établir que :

$$\sup_{t \in C} |f_n^*(t) - E f_n^*(t)| \xrightarrow{\text{p.co.}} 0, \quad n \rightarrow \infty.$$

Pour cela, considérons un recouvrement de C en l_n boules B_j de centre v_j et de rayon inférieur à b_n^η avec $l_n \leq \bar{l} b_n^{-\eta s}$, où η est un nombre fixé tel que $\eta > s + 1$, et \bar{l} une constante positive.

Soit t fixé dans C et B_j la boule contenant t ; en notant $Z_i(t)$ les v.a. définies en (3.2), on obtient;

$$Z_i(t) = Z_i(v_j) + \tilde{Z}_i(t),$$

avec $\tilde{Z}_i(t) = Z_i(t) - Z_i(v_j)$.

Le noyau K étant lipschitzien, il existe une constante M finie, positive, tel que

$$\sup_{t \in C} \left| \sum_{i=1}^n \tilde{Z}_i(t) \right| \leq 2M b_n^{-s} b_n^\eta.$$

Il reste seulement à montrer que :

$$\max_{1 \leq j \leq l_n} \left| \sum_{i=1}^n Z_i(v_j) \right| \xrightarrow{p.co.} 0, \quad n \rightarrow \infty.$$

Pour tout $\epsilon > 0$, on a :

$$P \left\{ \max_{1 \leq j \leq l_n} \left| \sum_{i=1}^n Z_i(v_j) \right| > \epsilon \right\} \leq l_n \sup_{t \in C} P \left\{ \left| \sum_{i=1}^n Z_i(t) \right| > \epsilon \right\}.$$

De la majoration (3.3) on déduit l'existence d'une constante C_3 telle que

$$P \left\{ \max_{1 \leq j \leq l_n} \left| \sum_{i=1}^n Z_i(v_j) \right| > \epsilon \right\} \leq C_3 b_n^{-\eta s} \exp(-C_2 \epsilon \frac{nb_n^s}{m_n}).$$

Le majorant précédent peut s'écrire :

$$\frac{C_3}{(nb_n^s)^\eta} n^{\eta - C_2 \epsilon \nu_n},$$

où $\nu_n = nb_n^s / m_n \text{Log} n \rightarrow \infty$; c'est le terme général d'une série convergente, ce qui achève la démonstration.

3.3 Démonstration du théorème 3.

La suite θ_n vérifie les conditions du lemme 2 et par conséquent :

$$\theta_n^{-1} \sup_{t \in C} |L_n(t) - L(t)| = O(1) \text{ p.co. }, n \rightarrow \infty.$$

Par ailleurs, en reprenant la démonstration du théorème 2 avec $\epsilon_0 \theta_n$ au lieu de ϵ , pour un certain $\epsilon_0 > 0$, on établit que

$$\theta_n^{-1} \sup_{t \in C} |f_n^*(t) - E f_n^*(t)| = O(1) \text{ p.co. }, n \rightarrow \infty.$$

Enfin, en effectuant dans (3.1) un développement de *Taylor* à l'ordre deux, compte tenu des hypothèses sur le noyau K on montre qu'il existe une constante $C_4 > 0$, telle que :

$$\sup_{t \in C} |Ef_n^*(t) - f^*(t)| \leq C_4 b_n^2.$$

Ce dernier résultat permet de conclure grâce à la décomposition (2.1).

BIBLIOGRAPHIE

D.Bosq, J.P.Lecoutre(1987). Théorie de l'estimation fonctionnelle. *Economica*.

M.Carbon(1988). Inégalités de grandes déviations dans les processus. Applications à l'estimation fonctionnelle. Thèse de Doctorat de l'Univ. Paris VI.

J.J.Droesbeke, B.Fichet, P.Tassi(1989),éditeurs. Analyse statistique des durées de vie. Modélisation des données censurées. *Economica*.

L.Györfi, W.Härdle, P.Sarda, P.Vieu(1989). Nonparametric curve estimation from time series. *Lecturs Notes in Statistics*, 60 Springer-Verlag.

S.Hassani, P.Sarda, P.Vieu(1986). Approche non paramétrique en théorie de la fiabilité : revue bibliographique. *Rev. Statist. Appl.*, 35, 27-41.

G.Roussas (1989). Hazard rate estimation under dependence conditions. *J. Statist. Plann. Inference*, 22, 81-93.

D. Shanbhag, S. Kotz (1987). Some new approaches to multivariate probability distributions. *J. Multivariate. Anal.*, 22, 189-211.

N. Singpurwalla, M. Wong (1983). Estimation of the failure rate. A survey of nonparametric methods. Part I : non-bayesian methods. *Comm. Statist.*, 12, 559-588.

VERGENCES STOCHASTIQUES DE SUITES DE
MESURES ALÉATOIRES SIGNÉES CONSIDÉRÉES
COMME VARIABLES ALÉATOIRES HILBERTIENNES

Christine SUQUET

Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées
Laboratoire de Statistique et Probabilité - Bât. M2
59663 - Villeneuve d'Ascq Cedex (France)

Résumé. Soit B la borne bornée d'un espace métrique X . On étudie le plongement de l'espace \mathcal{M} des mesures signées bornées sur B dans un espace de Hilbert autoréplémentaire séparé H_X afin de construire les mesures aléatoires signées comme variables aléatoires hilbertiennes. Cette construction est effectivement possible dans les trois cas où X est compact, localement compact, séparable. Les problèmes de mesurabilité des H_X progressivement mesurés dans ces 3 cas, les variables aléatoires naturellement associées à une mesure aléatoire ainsi définie sont les intégrales des éléments de H_X par rapport à cette mesure. On étudie la convergence progressive en probabilité et en loi des suites de mesures aléatoires et de leurs variables associées.

Summary. Let X be a metric space and B its bounded subset. Let \mathcal{M} the space of bounded signed measures on B . In order to construct the signed random measures as hilbertian random variables we use the embedding of \mathcal{M} in a separable reproducing kernel space H_X . This construction is effectively possible when X is compact or locally compact or separable. Measurability and integrable embedding are studied in these three cases. Real random variables naturally associated with a so defined random measure μ^* are integrals of H_X elements with respect to μ^* . For random sequences and their associated random variables, three modes of convergence are studied (almost surely, in probability, in law).

CLASSIFICATION J. M.S. (M.O.S.) : 60B35 - 60E27

KEYS WORDS : Espace à produit scalaire — Mesure signée — Mesure aléatoire — Variable hilbertienne — Convergence faible — Convergence stochastique