# Estimation par minimum de contraste de la fonction de risque instantanée dans le modèle de censure multiplicative

Gaëlle Chagny [1] & Fabienne Comte [2] & Angelina Roche [3]

[1] *LMRS, Univ. Rouen; gaelle.chagny@univ-rouen.fr*
[2] *MAP5, Univ. Paris Descartes; fabienne.comte@parisdescartes.fr*
[3] *CEREMADE, Univ. Paris-Dauphine; angelina.roche@dauphine.fr*

**Résumé.** Nous considérons le problème de l'estimation de la fonction de risque instantané d'une variable aléatoire d'intérêt $X$ positive, dans le modèle de censure multiplicative $Y = XU$ (Vardi, 1989), où $U$ est de loi uniforme sur $[0; 1]$, et est indépendant de $X$. Seul un échantillon distribué selon la loi de $Y$ est observé, et retrouver les caractéristiques de la distribution de $X$ peut ainsi être considéré comme un problème inverse. La question de l'estimation non-paramétrique de la densité $f_X$ et de la fonction de survie $\bar{F}_X$ de la variable cible a déjà été étudiée (Brunel *et al.*, 2015, et les références citées), mais nous choisissons de ne pas estimer la fonction de risque instantanée $h = f_X/\bar{F}_X$ par le rapport des deux estimateurs. Dans l'esprit de Plancade (2011), une collection d'estimateurs est définie par minimisation d'une fonction de contraste originale de type régression faisant intervenir un estimateur de la densité de $Y$, sur une collection de modèles linéaires. Nous prouvons un résultat non-asymptotique de contrôle du risque quadratique des estimateurs de la collection. La conséquence est double : des taux de convergence sont obtenus sous des hypothèses de régularité sur la fonction $h$, et la forme du compromis biais-variance devrait conduire à une procédure de sélection de modèle pénalisée - l'estimation adaptative dans ce cadre est encore un travail en cours. La méthode est illustrée par une étude numérique.

**Mots-clés.** Fonction de risque instantanée, estimation par minimum de contraste, modèle de censure multiplicative.

**Abstract.** We address the problem of hazard rate estimation for the nonnegative random variable $X$ of interest in the multiplicative censoring model $Y = XU$ (Vardi, 1989), where $U$ has a uniform distribution on $[0; 1]$, and is independent of $X$. Only a sample distributed like $Y$ is observed, and thus recovering features of the distribution of $X$ can be considered as an inverse problem. Although the problem of estimating the density $f_X$ and the survival function $\bar{F}_X$ of the target variable in a nonparametric way has already been studied by several authors (Brunel *et al.*, 2015, and references therein), we choose not to estimate the hazard $h = f_X/\bar{F}_X$ by the ratio of two estimators. In the spirit of Plancade (2011), a collection of estimators is defined, by minimizing an original regression-type contrast function, involving an estimator for the density of $Y$, over a collection of linear models. Nonasymptotic upper-bounds are proved for the quadratic risks of the estimates. The consequence is twofold: classical convergence rates are derived under smoothness assumptions on the function $h$, and the form of the bias-variance trade off should lead to a penalised model selection procedure - adaptive estimation in this setting is still a work in progress. Simulation experiments illustrate the method.

**Keywords.** Hazard rate; minimum of contrast estimation; multiplicative censoring model.

# 1 Modèle de censure multiplicative

Le modèle de censure multiplicative a été introduit par Vardi (1989) pour unifier différents problèmes statistiques, comme l'estimation sous contrainte de densité décroissante, la déconvolution avec bruit exponentiel ou certains problèmes d'estimation pour des processus de renouvellement (voir aussi Andersen et Hansen, 2001). Il apparaît également naturellement en analyse de survie (van Es *et al.*, 2000).

Ce modèle a été étudié par un certain nombre d'auteurs (voir par exemple Vardi et Zhang, 1992; Asgharian *et al.*, 2012). Vardi (1989) a proposé un estimateur du maximum de vraisemblance de la densité $f_X$.

Soit $f_Y$ la densité de $Y$, et $f_X$ celle de $X$. Nous pouvons voir facilement que

$$f_Y(y) = \int_y^{+\infty} \frac{f_X(x)}{x} dx, \quad y \geq 0$$

ce qui implique que l'obtention de la densité de $X$ à partir de la densité de $Y$ est un problème inverse. Partant de ce constat, Andersen et Hansen (2001) ont proposé une procédure d'estimation s'appuyant sur la décomposition en valeurs singulières de l'opérateur à inverser.

Plus récemment, Asgharian *et al.* (2012) ont considéré un estimateur de la densité $f_X$ pour lequel des résultats de consistance uniforme sont prouvés.

Cependant, toutes les procédures citées ci-dessus reposent sur l'observation directe de $X_1, ..., X_m$ avec $m$ grand, en plus de $Y_1, ..., Y_n$, les résultats théoriques n'étant plus valables quand $m = 0$.

Brunel *et al.* (2015) ont proposé des estimateurs adaptatifs à noyaux de la densité $f_X$ et de la fonction de survie $\bar{F}_X$ de la variable $X$ lorsque $m = 0$ ($X$ pouvant éventuellement prendre des valeurs négatives). La fenêtre des estimateurs proposés est sélectionnée par un critère inspiré des travaux de Goldenshluger et Lepski (2011) permettant d'atteindre automatiquement le meilleur compromis biais-variance.

Nous considérons ici le problème de l'estimation de la fonction de risque instantanée $h = f_X/\bar{F}_X$. Un estimateur possible dans ce cadre est un estimateur quotient de la forme $\widehat{h} = \widehat{f}_X/\widehat{\bar{F}}_X$ avec $\widehat{f}_X$ (resp. $\widehat{\bar{F}}_X$) un estimateur de $f_X$ (resp. $\bar{F}_X$) mais ce type d'estimateur souffre de problèmes d'instabilité lorsque $\widehat{\bar{\bar{F}}}_X$ est proche de 0. Nous proposons une approche alternative, inspirée des travaux de Comte *et al.* (2011) et Plancade (2011).

# 2 Estimation par projection

Nous considérons l'estimation de la fonction $h$ sur un compact déterministe $A = [0, \mathbf{a}] \subset [0, +\infty[$. Soit $(\mathbb{L}^2(A), \|.\|, \langle ., . \rangle)$ l'espace des fonctions de carré intégrable sur $A$, muni de son produit scalaire usuel $\langle f, g \rangle = \int_A f(t)g(t)dt$ et de la norme associée $\|f\| = \sqrt{\langle f, f \rangle}$, pour tous $f,\ g \in \mathbb{L}^2(A)$.

Nous considérons la collection $(S_m)_{m \geq 1}$ de sous-espaces vectoriels engendrés par les fonctions de la base de Fourier sur $A$ : $S_m = \mathrm{Span}\{\varphi_1, \ldots, \varphi_{D_m}\}$, avec $\varphi_1(x) = \mathbf{a}^{-1/2}\mathbf{1}_A(x)$,

$$\varphi_{2j}(x) = \mathbf{a}^{-1/2}\mathbf{1}_A(x)\sqrt{2}\cos\left(2\pi jx/\mathbf{a}\right) \text{ et } \varphi_{2j+1}(x) = \mathbf{a}^{-1/2}\mathbf{1}_A(x)\sqrt{2}\sin\left(2\pi jx/\mathbf{a}\right),$$

et $D_m = 2m + 1$, $m \geq 1$. Nous choisissons une dimension maximale $D_{N_n}$ telle que $D_{N_n} \leq K\sqrt{n/\ln(n)^3}$ avec $K$ une constante positive. La stratégie d'estimation que nous adoptons nécessite la différentiabilité de la base $(\varphi_j)_{j \geq 1}$, d'où le choix de la base trigonométrique. D'autres bases qui vérifient cette condition, comme une base de splines, pourraient être considérées.

Nous notons également $\| \cdot \|_{\bar{F}_X}$ la semi-norme qui apparaît naturellement dans notre problème d'estimation, définie par $\|t\|^2_{\bar{F}_X} := \langle t, t \rangle_{\bar{F}_X}$, où pour tous $s, t \in \mathbb{L}^2(A)$, $\langle s, t \rangle_{\bar{F}_X} := \int_A s(x)t(x)\bar{F}_X(x)dx$. Nous pouvons voir aisément que $\|t\|_{\bar{F}_X} \leq \|t\|$.

Dans la suite, nous faisons l'hypothèse suivante :

$(H_A)$ Il existe $\bar{F}_0 > 0$ tel que $\inf_{x \in A} \bar{F}_X(x) \geq \bar{F}_0 > 0$ et $f_X$ est majorée sur $A$.

Cette hypothèse justifie en particulier le choix de l'ensemble $A$ d'estimation comme étant strictement inclu dans $\mathbb{R}_+$ (sans quoi $(H_A)$ ne pourrait être vérifiée). L'hypothèse $(H_A)$ permet d'établir que $h$ est de carré intégrable sur $A$: en effet, sous $(H_A)$, $\|h\|^2_A \leq \sup_{x \in A} |f_X(x)|/(\bar{F}_0)^2$. De plus, si $(H_A)$ est vérifiée, alors les normes $\|.\|$ et $\|.\|_{\bar{F}_X}$ sont équivalentes sur $\mathbb{L}^2(A)$ :

$$\forall t \in \mathbb{L}^2(A), \bar{F}_0^{-1}\|t\|^2 \leq \|t\|^2_{\bar{F}_X} \leq \|t\|^2.$$

## 2.1 Définition du contraste

Pour tout $m = 1, ..., N_n$, nous cherchons à définir un estimateur de la fonction de risque instantanée par minimisation d'un contraste.

La clef pour la définition du contraste repose sur les égalités suivantes, qui découlent de la définition du modèle

$$yf'_Y(y) = -f_X(y),\ y \geq 0 \tag{1}$$

et
$$\bar{F}_Y(y) + y f_Y(y) = \bar{F}_X(y),\ y \geq 0. \tag{2}$$

À partir de ces relations, nous construisons le contraste suivant, qui est défini pour toute fonction dérivable $t \in \mathbb{L}^2(A)$

$$\gamma_n(t) = \|t\|_n^2 - \frac{2}{n}\sum_{i=1}^{n}\left[t(Y_i) + Y_i t'(Y_i)\right]\mathbf{1}_{\{Y_i \in A\}} - 2\mathbf{a}t(\mathbf{a})\widehat{f}_Y(\mathbf{a}), \tag{3}$$

$$\text{avec} \quad \|t\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\int_A t^2(x)\mathbf{1}_{\{Y_i \geq x\}}dx + Y_i t^2(Y_i)\mathbf{1}_{\{Y_i \in A\}}\right],$$

la semi-norme empirique et $\widehat{f}_Y(\mathbf{a})$ un estimateur adaptatif de la densité $f_Y$ de $Y$ au point $\mathbf{a}$, par exemple $\widehat{f}_Y(\mathbf{a}) = (n\widehat{h})^{-1}\sum_{i=1}^{n} K((Y_i - \mathbf{a})/\widehat{h})$ avec $K$ un noyau et $\widehat{h}$ sélectionné par un critère imitant au mieux la décomposition biais-variance du risque ponctuel en $\mathbf{a}$ (voir Rebelles, 2015).

Les relations (1) et (2) impliquent la propriété suivante :

**Lemme 2.1** *Supposons ($H_A$) vérifiée. Soit $t \in \mathbb{L}^2(A)$ une fonction dérivable alors*

$$\mathbb{E}(\gamma_n(t)) = \|t - h\|_{\bar{F}_X}^2 - \|h\|_{\bar{F}_X}^2 + \mathbf{a}t(\mathbf{a})\left(\mathbb{E}\left[\widehat{f}_Y(\mathbf{a})\right] - f_Y(\mathbf{a})\right). \tag{4}$$

Par conséquent, minimiser le contraste $\gamma_n$ revient, pour $n$ suffisamment grand, à trouver la fonction $t \in S_m$ pour laquelle $\|t - h\|_{\bar{F}_X}$ est minimal, dès que l'estimateur $\widehat{f}_Y$ est un bon estimateur de $f_Y$, au point $\mathbf{a}$. Sous $(H_A)$ l'équivalence des normes $\|\cdot\|$ et $\|\cdot\|_{\bar{F}_X}$, nous indique que l'estimateur défini par minimisation du contraste $\gamma_n$ est une approximation de la projection $h_m$ de $h$ sur $S_m$ et, lorsque $m$ est suffisamment grand, de la fonction à estimer $h$.

## 2.2 Définition des estimateurs

L'élément $t = \sum_{j=1}^{D_m}\alpha_j\varphi_j$ qui minimise $\gamma_n$ sur $S_m$ vérifie

$$\frac{\partial\gamma_n(\sum_{j=1}^{D_m}\alpha_j\varphi_j)}{\partial\alpha_{j_0}} = 0, \quad \text{pour tout } j_0 \in \{1, \ldots, D_m\}.$$

Soit $\alpha = {}^t(\alpha_j)_{j=1,\ldots,D_m} \in \mathbb{R}^{D_m}$, le vecteur des coefficients de $t$. La dernière condition équivaut à

$$\widehat{\Psi}_m\widehat{\alpha} = \widehat{b}, \ \text{ avec } \ \widehat{b} = {}^t\left(\frac{1}{n}\sum_{i=1}^{n}(\varphi_j(Y_i) + Y_i\varphi_j'(Y_i))\mathbf{1}_{\{Y_i \in A\}} - \mathbf{a}\widehat{f}_Y(\mathbf{a})\varphi_j(\mathbf{a})\right)_{j=1,\ldots,D_m},$$

et $\widehat{\Psi}_m := (\Psi_{j,k})_{j,k\in\{1,...,D_m\}}$, où

$$\Psi_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\left\{\int_A \varphi_j(x)\varphi_k(x)\mathbf{1}_{Y_i\geq x}dx + Y_i\varphi_j(Y_i)\varphi_k(Y_i)\mathbf{1}_{\{Y_i\in A\}}\right\}.$$

Il existe donc une unique fonction de $S_m$ minimisant $\gamma_n$ si et seulement si $\widehat{\Psi}_m$ est inversible. Par conséquent, étant donné un réel $0<\rho_1<1$, nous définissons, pour tout $m=1,...,N_n$, l'ensemble $\widehat{\Delta}^m_{\rho_1} = \left\{\min \mathrm{Sp}(\widehat{\Psi}_m) \geq (1-\rho_1)\widehat{\bar{F}}_0\right\}$, avec $\widehat{\bar{F}}_0$ un estimateur de $\bar{F}_0$ vérifiant la propriété suivante

**$(H_F)$** $\mathbb{P}(|\widehat{\bar{F}}_0 - \bar{F}_0| \geq \rho_0\bar{F}_0) \leq c/n^4$ pour un $\rho_0 \in (0,1)$.

Deux définitions au moins sont envisageables pour $\widehat{\bar{F}}_0$ : si l'on fixe $\bar{F}_0 = \inf_{x\in A}\bar{F}_Y(x) = \bar{F}_Y(\mathbf{a})$ (qui vérifie bien $(H_A)$ car $\bar{F}_X \leq \bar{F}_Y$ par Eq. (2)), alors $\widehat{\bar{F}}_Y(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Y_i\geq \mathbf{a}\}}$ vérifie $(H_F)$. Une autre possibilité, qui donne un estimateur plus précis mais plus difficile à calculer, est de considérer que $\bar{F}_0 = \bar{F}_X(\mathbf{a})$ et d'estimer la survie $\bar{F}_X$ de $X$ au point $\mathbf{a}$, par exemple par une procédure d'estimation à noyau (Brunel *et al.*, 2015).

Finalement, notre estimateur est défini par :

$$\widehat{h}_m = \sum_{j=1}^{D_m}\widehat{\alpha}_j\varphi_j, \text{ avec } \widehat{\alpha} = {}^t(\widehat{\alpha}_j)_{j=1,...,D_m} = \begin{cases} \widehat{\Psi}_m^{-1}\widehat{b} & \text{sur } \widehat{\Delta}^m_{\rho_1} \\ 0 & \text{sur } (\widehat{\Delta}^m_{\rho_1})^c \end{cases}.$$

# 3 Résultats théoriques

Nous montrons la propriété suivante.

**Proposition 3.1** *Si $(H_A)$ et $(H_F)$ sont vérifiées et $f_Y \in \mathbb{L}^2(A)$, $\mathbb{E}(Y_1^2) < +\infty$ alors l'estimateur $\widehat{h}_m$ satisfait*

$$\mathbb{E}(\|\widehat{h}_m - h\|^2) \leq C_1(\|h_m - h\|^2 + V(m)) + C_2 D_m \mathbb{E}\left[\left(\widehat{f}_Y(\mathbf{a}) - f_Y(\mathbf{a})\right)^2\right] + \frac{C_3}{n}$$

*avec $C_1$, $C_2$ et $C_3$ des constantes indépendantes de $m$ et $n$, et*

$$V(m) = \frac{1}{\mathbf{a}\bar{F}_0 n}\left((1+\mathbf{a}\|h\|^2)D_m + \pi^2 D_m^3\right).$$

La borne supérieure de la Proposition 3.1 fait apparaître un terme de biais $\|h-h_m\|^2$ qui décroît lorsque $m$ croît, et un terme de variance $V(m)$ qui, à l'inverse, croît avec $m$. Le dernier terme $C_3/n$ est un terme résiduel négligeable. L'avant dernier terme est de l'ordre du risque quadratique ponctuel de l'estimateur $\widehat{f}_Y$ au point $\mathbf{a}$.

Pour obtenir la vitesse de convergence de l'estimateur $\widehat{h}_m$, nous supposons de plus que $h$ est dans une boule de Hölder $\Sigma(\beta, L, [0, \mathbf{a} + \varepsilon])$ avec $\varepsilon > 0$ où

$$\begin{aligned}\Sigma(\beta, L, I) &= \{f : I \to \mathbb{R}, \text{ t.q. } f \text{ est } \ell = \lfloor \beta \rfloor - \text{fois dérivable} \\ &\qquad \text{et } |f^{(\ell)}(x) - f^{(\ell)}(y)| \leq L|x - y|^{\beta - \ell} \text{ pour tous } x, y \in I\big\},\end{aligned}$$

où $\lfloor \beta \rfloor$ désigne le plus grand entier strictement plus petit que $\beta$.

**Corollaire 3.1** *Nous supposons que les hypothèses de la Proposition 3.1 sont vérifiées. Alors, pour tout* $\varepsilon > 0$,

$$\sup_{h \in \Sigma(\beta, L, [0, \mathbf{a}+\varepsilon])} \mathbb{E}(\|\widehat{h}_m - h\|^2) \leq C n^{-2\beta/(2\beta+3)},$$

*pour une constante* $C$ *indépendante de* $m$ *et* $n$.

La vitesse $n^{-2\beta/(2\beta+3)}$ est la vitesse obtenue pour l'estimation de la densité de $X$ par Brunel *et al.* (2015) qui ont prouvé également des résultats d'optimalité (vitesse minimax). Comme le montre le corollaire précédent, le terme additionnel ne dégrade pas l'ordre du risque obtenu en choisissant le modèle qui réalise le compromis biais-variance. Cependant, la prise en compte de ce terme pour définir une pénalité adéquate permettant de sélectionner le bon modèle uniquement sur la base des observations reste pour l'instant un travail en cours.

# Bibliographie


[1] K. E. Andersen et M. B. Hansen (2001), Multiplicative censoring: density estimation by a series expansion approach. *J. Statist. Plann. Inference*, 98(1-2):137–155.

[2] M. Asgharian, M. Carone, et V. Fakoor (2012), Large-sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.*, 40(1):159–187.

[3] E. Brunel, F. Comte, et V. Genon-Catalot (2015), Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST*. to be published.

[4] F. Comte, S. Gaïffas, et A. Guilloux (2011), Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1171–1196.

[5] A. Goldenshluger et O. Lepski (2011), Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3): 1608–1632.

[6] S. Plancade (2011), Model selection for hazard rate estimation in presence of censoring. *Metrika*, 74(3):313–347.

[7] B. van Es, C. A. J. Klaassen, et K. Oudshoorn (2000), Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. *J. Statist. Plann. Inference*, 91(2):295–312.

[8] Y. Vardi (1989), Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika*, 76(4):751–761.
[9] G. Rebelles (2015), Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli*, 21(4), 1984–2023.
[10] Y. Vardi et C.-H. Zhang (1992), Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.*, 20(2):1022–1039.