The Title

The Author

The Date

Table des matières

0.1 Introduction

Historique

L'analyse des données de survie voit le jour au XVII^e siècle, dans le domaine de la démographie. L'objectif des analystes de ce siècle est l'estimation, à partir des registres de décès, de diverses caractéristiques de la population-son effectif, sa longévité, etc. Ces analyses, très générales, ne sont affinées qu'à partir du XIX^e siècle, avec l'apparition de catégorisations suivant des <<variables exogènes>> (sexe, nationalité, catégories socio-professionnelles...). Durant ce siècle, apparaîssent également les premières modélisations consernant la probabilité de mourir à un certain âge, probabilité qui sera par la suite désignée sous le terme de <<fonction de risque>>.

En 1951, Weibull conçoit un modèle paramétrique dans le domaine de la fiabilité ; à cet effet, il fournit une nouvelle distribution de probabilité qui sera par la suite fréquement utilisée en analyse de la survie : la <<loi de Weibull>>.

En 1958, Kaplan et Meier présentent d'importants résultats concernant l'estimation non-paramétrique de la fonction de survie ; de l'estimateur résultant, ils étudient l'espérence, la variance et les propriétés asymptotiques.

L'analyse des données de survie a pour première particularité de ne concerner que des **variables aléatoires positives**.

Une deuxième particularité de cette analyse est l'**incomplétude des données**-différente de la troncature, qui équivaut à une perte d'information. Analysant la survenue d'un certain type d'événement, nous qualifierons de <<donnée complète>> un temps coorespondant à l'observation de la survenue de l'événement, et de <<donnée incomplète> > un temps correspondant à l'absence d'observation de cet événement. Nous emploierons respectivement, par la suite, les termes de **donnée non censurée** et de **donnée censurée.**

Enfin, la troisième particularité de l'analyse des données de survie est la **terminologie** s'y rattachant. Outre les termes conçus à l'origine de cette analyse (tels que *fonction de survie, fonction de risque*).

Constitution et perspectives de mémoire.

Le mémoire se comporte de cinq chapitres.

Au premier chapitre on présente une introduction sur la distribution de survie et les résultats consernant la théorie des données censurées.

Le deuxième chapitre présente les principaux résultats consernant l'estimateur de la fonction de survie tel que l'estimateur de Kaplan Meier (1958) et l'estimateur des quantiles de type noyau (voir, Padjett (1986)).

Le troixième chapitre a été consacré à l'estimateur à noyau des quantiles. Nous présentons ici, les principaux résultat consernant cette classe d'estimateur tels que l'équivalence asymptotique avec leur approximation et la normalité asymptotique.

Nous avons consacrés le quatrième chapitre au loi de logarithme des estimateurs à noyau de la densité des quantiles et de la densité et de taux de hazard pour les données censurées.

Enfin, le dernier chapitre présent la loi limite de l'estimateur de la densité de survie et on définit les accroissements des processus de Kaplan-Meier pour obtenir les lois limites fonctionnelles de ces accroissements.

0.2 Notations

Nous donnons ci-dessous les conventions d'ecriture adoptées :

v.a	variable aléatoire
i.i.d	indépendantes et identiquement distribuées
$p.s$	presque sûrement
$\mathbf{P}$	probabilité
$\mathbf{E}$	espérence
$\mathbf{V}$	variance
$\overset{\pounds}{\rightarrow}$	convergence en loi
$\overset{\mathbf{P}}{\rightarrow}$	convergence en probabilité
$\overset{p.\mathbf{s}}{\rightarrow}$	convergence presque sûre
f_t	filtration à l'instant$t$
$1_{\{A\}}$	fonction indécatrice de l'évenement A
$p_{]s,t]}$	produit-intégrale (ou produit infini) sur $]s,t]$
$P-L$ estimateur	estimateur à limite produit.
$\alpha_n(t)$	processus empirique de Kaplan-Meier
$\beta_n(t)$	Processus de censure de Kaplan-Meier

Chapitre 1

Elements d'analyse des données censurées(DC)

1.1 Distribution de survie

Nous donnons ci-dessous les définitions des principaux outils utilisés en analyse de la survie, nous allons définir les fonctions permettant de décrire une distribution de survie.

Admettons qu'à la date $t = 0$ un élément (un sujet ou un système) commence à fonctionner (à vivre) et qu'à la date t il se produire une panne (la mort, le décès).

Définition 1.1 : La variable **durée de vie** X, délai entre la date d'origine et la date du décès (panne) est une variable aléatoire non négative, $X \in [0, \infty[$. Sa fonction de répartition est :

$$F(t) = \mathbf{P}(X \leq t), \; t \in \mathbb{R}_+. \tag{1.1}$$

Nous ne considérons ici que le cas, où X est continue, c'est-à-dire que la probabilité de décès (de panne) à chaque instant est infiniment petite. Alors :

$$F(t) = \int_0^t f(x)dx,$$

où $f(t)$ est la densité de probabilité de X, définie par :

$$f(t) = F'(t) = \lim_{h \to 0} \frac{\mathbf{P}\{t \leq X \leq t+h\}}{h}, \quad h > 0. \tag{1.2}$$

Donc, $F(t)$ est la probabilité de décéder entre 0 et t, ou la probabilité de défaillance (de panne) au cours de l'intervalle $[0, t]$.

Définition 1.2 : La **fonction de survie** est définie par

$$\begin{aligned} S(t) &= \overline{F}(t) = 1 - F(t) \\ &= \mathbf{P}(X > t), \ t \geq 0. \end{aligned}$$

On remarque que $S(t) = \overline{F}(t)$ est la probabilité de bon fonctionnement continu durant $[0, t]$ ou la probabilité du fonctionnement sans défaillance de l'élément au cours du temps t. La fonction $S(t)$ est monotone décroissante, avec

$$S(0) = 1 \quad \text{et} \quad S(t) \to 0, \quad \text{quand } t \to \infty.$$

La plus importance caractéristique numérique de la durée de survie X est le temps moyen de survie $\mathbf{E}(X)$. Dans ce cas :

$$\begin{aligned} \mathbf{E}(X) &= \int_0^\infty t dF(t) \\ &= -\int_0^\infty t d[1 - F(t)] \\ &= \int_0^\infty S(t) dt, \end{aligned}$$

i.e. , si $\mathbf{E}(X)$ existe, alors :

$$\mathbf{E}(X) = \int_0^\infty S(t) dt.$$

De même, on peut montrer que, si $\mathbf{V}(X)$ existe, alors :

$$\mathbf{V}(X) = 2\int_0^\infty t S(t) dt - (\mathbf{E}(X))^2.$$

1.2 Risque de panne ou taux de défaillance

Considérons tout d'abord le problème suivant :

Supposons que l'élément ait fonctionné sans défaillance jusqu'à linstant u, $u > 0$.

Quelle est la probabilité pour qu'il ne tombe pas en panne dans l'intervalle $]u, u+t]$, $t > 0$? Donc, on s'intéresse à la probabilité

$$S_u(t) = \mathbf{P}\{X > u + t \mid X > u\}, \ u > 0, \ t > 0.$$

La probabilité cherchée est alors la probabilité conditionnelle, et on a :

$$\mathbf{P}\{X > u + t \mid X > u\} = \frac{\mathbf{P}\{X > u + t\}}{\mathbf{P}\{X > u\}} = \frac{S(u+t)}{S(u)} = S_u(t). \tag{1.3}$$

De (1.3) on tire immédiatement que pour tout $\Delta t > 0$:

$$S(t + \Delta t) = \mathbf{P}\{X > t + \Delta t\} = S(t)\Delta t p_t \tag{1.4}$$

où

$$\Delta t p_t = \mathbf{P}\{X > t + \Delta t \mid X > t\}.$$

De (1.3) et (1.4) il suit que la probabilité de panne (de décès) au cours de $]t, t + \Delta t]$, sachant que $X > t$ est :

$$\begin{aligned}
\Delta t q_t &= \mathbf{P}\{t < X \leq t + \Delta t \mid X > t\} \\
&= 1 - \Delta t p_t \\
&= \frac{S(t) - S(t + \Delta t)}{S(t)}.
\end{aligned}$$

Définition 1.3 : La fonction de risque instantané de décès ou taux de déffaillance ou risque de panne est la fonction définie par :

$$\alpha\,(t) = \frac{f(t)}{\overline{F}(t)} = \frac{f(t)}{S(t)}, \quad t \geq 0. \tag{1.5}$$

où f est la densité de probabilité de X.

Notons que cette fonction n'est pas une densité de probabilité.

Le risque instantané est la probabilité que l'événement se produise en x (sachant qu'il ne s'est pas produit auparavant) :

$$\begin{aligned} \alpha\,(x)\,dx &= \frac{\mathbf{P}(x < X \leq x + dx)}{\mathbf{P}(X > x)} \qquad (1.6) \\ &= \mathbf{P}\left(X \in \left]x, x+dx\right] \mid X > x\right). \end{aligned}$$

De la définition (1.2) il suit que :

$$\begin{aligned} \alpha\,(t) &= \lim_{\Delta t \to 0} \frac{\mathbf{P}\left\{t < X \leq t + \Delta t\right\}}{\Delta t S(t)} \\ &= \lim_{\Delta t \to 0} \frac{\mathbf{P}\left\{t < X \leq t + \Delta t\right\}}{\Delta t \mathbf{P}\left\{X > t\right\}} \\ &= \lim_{\Delta t \to 0} \frac{\mathbf{P}\left\{t < X \leq t + \Delta t \mid X > t\right\}}{\Delta t} \\ &= \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{S(\mathbf{t}) - S(t + \Delta t)}{\Delta t} \\ &= -\frac{S'(t)}{S(t)} \qquad (1.7) \end{aligned}$$

Remarque 1.1 : De (1.7) il sensuit que :

$$\alpha\,(t) = -\frac{d \ln S(t)}{dt},$$

d'où

$$\ln S(t) = -\int_0^t \alpha\,(s)\,ds, \qquad t > 0,$$

donc

$$S(t) = \exp\left\{-\int_0^t \alpha\,(s)\,ds\right\}. \tag{1.8}$$

Définition 1.4 : La **fonction de risque cumulé** de $\alpha(s)$ entre 0 et t est donnée par :

$$A(t) = \int_0^t \alpha(s)\, ds.$$

De (1.8) il suit que :

$$A(t) = -\ln S(t),\ t \geq 0,$$

et de (1.5) il vient :

$$f(t) = \alpha(t)\, S(t) = \alpha(t) \exp\{-A(t)\}.$$

Nous pouvons maintenant énoncer les deux résultats suivants :

Propsition 1.1 : La définition de la distribution de probabilité de X repose sur l'une des quatre données suivantes, qui sont équivalentes : $S(x)$, $f(x)$, $\alpha(x)$ et $A(x)$.

Proposition 1.2 : Nous avons :

$$\begin{aligned} S(x) &= P(X > x) \\ &= p_{]0,x]}\left[1 - \alpha(s)\, ds\right] \\ &= \exp\left\{-\textstyle\int_0^x \alpha(s)\, ds\right\} \\ &= \exp(-A(x)) \end{aligned}$$

où $p_{]0,x]}$ désigne le produit infini (ou produit intégrale) (voir, Annexe).

Remarque 1.2 : Dans le cas où X est une variable aléatoire discrète,

$$\mathbf{P}\{X = k\} = p_k, \quad k \in \{1, 2, ...\}.$$

Les fonctions de répartition $F(k)$, de survie $S(k)$ et de risque de défaillance $\alpha(k)$ de X sont données par les formules suivantes :

$$F(k) = \mathbf{P}\{X \leq k\} = \sum_{m \leq k} p_m,$$

$$S(k) = \mathbf{P}\{X > k\} = \mathbf{P}\{X \geq k+1\} = \sum_{m=k+1}^{\infty} p_m,$$

$$\begin{aligned}
\alpha(k) &= \mathbf{P}\{X = k \mid X > k-1\} \\
&= \mathbf{P}\{X = k \mid X \geq k\} \\
&= \frac{p_k}{\sum\limits_{m=k}^{\infty} p_m} \\
&= \frac{p_k}{S(k-1)},
\end{aligned}$$

pour tout $k \in \mathbb{N}$, (on pose ici, que $S(0) = 1$).

Comme

$$1 - \alpha(k) = \frac{S(k-1) - p_k}{S(k-1)} = \frac{S(k)}{S(k-1)}$$

on en tire que :

$$S(k) = [1 - \alpha(k)]\, S(k-1) = \sum_{m=k+1}^{\infty} p_m = \prod_{m=1}^{k} [1 - \alpha(m)], \quad k \in \mathbb{N},$$

puisque :

$$p_k = \alpha(k)\, S(k-1) = \alpha(k) \prod_{m=1}^{k-1} [1 - \alpha(m)], \quad k \in \mathbb{N},$$

en posant $p_1 = \alpha(1)$.

Enfin en remarque que :

$$\mathbf{E}X = \sum_{j=1}^{\infty} j p_j = \sum_{j=1}^{\infty} \sum_{k=1}^{j} p_k = \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} p_j$$

$$= \sum_{k=1}^{\infty} \mathbf{P}\{X \geq k\} = \sum_{k=1}^{\infty} \mathbf{P}\{X > k-1\} = \sum_{k=1}^{\infty} S(k-1).$$

1.3 Modèle de Weibull

Soit :

$F(t) = F(t;\alpha,\lambda) = \mathbf{P}\{X \leq t\} = \left(1 - e^{-\lambda t^{\alpha}}\right) \mathbf{1}_{]0,\infty[}(t), \quad \lambda > 0, \quad \alpha > 0, \quad t \in \mathbb{R},$

i.e., X suit une loi de Weibull $W(\alpha,\lambda)$ de paramètres α et λ. Dans ce modèle :

$$S(t) = S(t;\alpha,\lambda) = e^{-\lambda t^{\alpha}} \mathbf{1}_{]0,\infty[}(t),$$

$$f(t) = f(t;\alpha,\lambda) = \alpha\lambda t^{\alpha-1} e^{-\lambda t^{\alpha}} \mathbf{1}_{]0,\infty[}(t).$$

On peut montrer que :

$$\mathbf{E}(X^k) = \lambda^{-k/\alpha} \Gamma\left(\frac{k}{\alpha} + 1\right),$$

et par conséquent :

$$\mathbf{E}(X) = \frac{1}{\lambda^{1/\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right),$$

et

$$\mathbf{E}(X^2) = \frac{1}{\lambda^{2/\alpha}} \Gamma\left(\frac{2}{\alpha} + 1\right),$$

où

$$\Gamma(\lambda) = \int_0^{\infty} x^{\lambda-1} e^{-x} dx$$

est la fonction gamma,

$$\begin{aligned}\Gamma(\lambda) &= (\lambda-1)\Gamma(\lambda-1),\\ \Gamma(1/2) &= \sqrt{\pi}.\end{aligned}$$

et donc :

$$\mathbf{V}(X) = \frac{1}{\lambda^{2/\alpha}}\Gamma\left(\frac{2}{\alpha}+1\right) - \frac{1}{\lambda^{2/\alpha}}\Gamma^2\left(\frac{1}{\alpha}+1\right).$$

Remarque 1.3 : Soit $X = (X_1, ..., X_n)$ un échantillon tel que :

$$\mathbf{P}\left\{X_i \leq x\right\} = G(x;\alpha,\lambda)\mathbf{1}_{]0,\infty[}(x), \quad \alpha > 0, \quad \lambda > 0, \ \ t \in \mathbb{R},$$

où $G(x;\alpha,\lambda)$ une fonction de répartition vérifiant les conditions :

$$\lim_{x\downarrow 0}\frac{G(x;\alpha,\lambda)}{\lambda x^\alpha} = 1, \qquad G(x;\alpha,\lambda) = 0, \quad x \leq 0,$$

pour tout α et λ fixés.

Soit :

$$X_{(1)} = X_{(n1)} = \min(X_1, ..., X_n).$$

Alors :

$$n^{1/\alpha}X_{(n1)} \xrightarrow{\pounds} W(\alpha,\lambda), \quad \text{quand } n \to \infty,$$

où $W(\alpha,\lambda)$ est la distribution de Weibull de paramètres α et λ, et de densité

$$f(x) = \frac{\alpha}{\lambda}\left(\frac{x}{\lambda}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\lambda}\right)^{\alpha}\right].$$

En effet, pour tout $x > 0$ on a :

$$\mathbf{P}\left\{X_{(n1)} > x\right\} = \left[1 - G(x;\alpha,\lambda)\right]^n$$

et

$$\mathbf{P}\left\{n^{1/\alpha}X_{(n1)} > x\right\} = \left[1 - G(\frac{x}{n^{1/\alpha}}; \alpha, \lambda)\right]^n,$$

d'où on déduit que si $n \to \infty$, alors :

$$\begin{aligned} \ln \mathbf{P}\left\{n^{1/\alpha}X_{(n1)} > x\right\} &= n \ln\left[1 - G(\frac{x}{n^{1/\alpha}}; \alpha, \lambda)\right] \\ &= n\left[-\lambda\left(\frac{x}{n^{1/\alpha}}\right)^\alpha + o\left(\frac{1}{n}\right)\right] \\ &= -\lambda x^\alpha + o\left(1\right), \end{aligned}$$

d'où on tire que pour tout $x > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{n^{1/\alpha}X_{(n1)} > x\right\} = e^{-\lambda x^\alpha} = S(x; \alpha, \lambda),$$

i.e. asymptotiquement $(n \to \infty)$ la statistique $X_{(n1)}$ suit la loi de Weibull $W(\alpha, \lambda)$ de paramètres α et λ.

1.4 Censure

1.4.1 Définition

Définition 1.5 : la **variable de censure** C est définit par la possible non observation de l'événement. Si l'on observe C, et non X, et que l'on sait que $X > C$ (respectivement $X < C$, $C_1 < X < C_2$), on dit qu'il ya **censure à droite** (respectivement **censure à gauche**, **censure par intervalle**).

Si l'événement se produit, X est ‹‹observée››. S'il ne se produit pas (l'individus étant perdu de vue, ou bien exclu vivant), c'est C qui est ‹‹réalisé› ›.

Par ailleurs, la censure se distingue de la **troncature** : on dit qu'il y a troncature gauche (respectivement droite) lorsque la variable d'intérêt X_i (durée de vie du i^e individu) n'est pas observable quand elle est inferieure à un seuil $c > 0$ fixé (respectivement superieure à un seuil $C > 0$ fixé).

On remarque que ce phénomène de troncature est très différent de celui de la censure, car dans le cas de la censure, on sait que la variable X_i, non observée, est superieure (ou inférieure) à une valeur C qui, elle, a été observée. Donc, la troncature élimine de l'étude une partie des X_i, ce qui a pour conséquence que l'analyse pourra porter seulement sur la loi de X_i conditionnellement à l'événement $(c < X_i \leq C)$, en cas de troncature gauche et droite simultanées.

1.4.2 Caractéristiques

Définition 1.6 : **(Censure de type I).**

Étant donné un échantillon $\mathbb{X} = (X_1, ..., X_n)$ de durées de survie X_i et un nombre positif fixé C, on dit qu'il y a censure à droite de cet échantillon, si au lieu d'observer $X_1, ..., X_n$, on observe n statistiques

$$(T_1, \Delta_1), ..., (T_n, \Delta_n)$$

où

$$T_i = \min(X_i, C),$$

et

$$\Delta_i = \mathbf{1}_{\{T_i = X_i\}} = \begin{cases} 1, & \text{si} \quad X_i \leq C, \\ 0, & \text{si} \quad X_i > C. \end{cases}$$

il est clair que :

$$T_i = X_i \mathbf{1}_{\{X_i \leq C\}} + C \mathbf{1}_{\{X_i > C\}}.$$

Donc, en realité on observe la défaillance (le décès) du sujet i si $X_i \leq C$, et la variable indicatrice Δ_i de l'état aux dernière nouvelles Δ_i du sujet i vaut 1, dans le cas contraire $X_i > C$ et donc l'observation est censurée et l'état aux dernieres nouvelles Δ_i du sujet i vaut 0.

Lorsqu'on ordonne les valeurs de T_i par ordre croissant, obtenant les statistiques d'ordre :

$$T_{(1)} \leq T_{(2)} \leq ... \leq T_{(n)}.$$

C'est par exemple ce qui se passe lorsqu'on observe la durée de fonctionnement de n systèmes au cours d'une expérience de durée C.

Soit $f(x_i;\theta)$ la densité de X_i,

$$X_i \backsim f(x_i;\theta), \quad \theta \in \Theta, \quad x_i \geq 0,$$

et

$$S(x_i;\theta) = 1 - F(x_i;\theta) = \mathbf{P}_\theta \{X_i \geq x_i\}$$

sa fonction de survie où $F(x_i;\theta)$ est la fonction de répartition de X_i.

Dans ce cas, la densité de la statistique (T_i, Δ_i) est donnée par la formule :

$$h(t_i, d_i;\theta) = [f(t_i;\theta)]^{d_i} [S(t_i;\theta)]^{1-d_i}, \quad t_i > 0; \; d_i \in \{0,1\}.$$

Parce que la statistique Δ_i, represente la partie discrète de la statistique (T_i, Δ_i), on a :

$$\begin{aligned}
\mathbf{P}_\theta \{T_i \leq t_i, \Delta_i = 0\} &= \mathbf{P}_\theta \{C \leq t_i, X_i > C\} \\
&= \begin{cases} S(C;\theta), & \text{si } C \leq t_i, \\ 0 & \text{si non,} \end{cases} \\
&= \int_0^{t_i} S(C;\theta) \mathbf{1}_{\{v \geq C\}} dv,
\end{aligned}$$

et donc :

$$h(t_i, 0;\theta) = S(C;\theta) \mathbf{1}_{\{t_i \geq C\}}.$$

De l'autre côté on a :

$$\begin{aligned}
\mathbf{P}_\theta \{T_i \leq t_i, \Delta_i = 1\} &= \mathbf{P}_\theta \{X_i \leq t_i, X_i \leq C\} \\
&= \begin{cases} S(C;\theta), & \text{si } t_i \leq C, \\ 0, & \text{si non,} \end{cases} \\
&= \int_0^{t_i} f(v;\theta) \mathbf{1}_{\{t_i \leq C\}} dv,
\end{aligned}$$

et donc :

$$h(t_i, 1; \theta) = f(t_i; \theta)\mathbf{1}_{\{t_i \leq C\}}.$$

Donc la fonction de vraisemblance, correspondant aux observations

$$(T_1, \Delta_1), ..., (T_n, \Delta_n),$$

est :

$$L(\theta) = \prod_{i=1}^{n} [f(T_i; \theta)]^{\Delta_i} [S(C; \theta)]^{1-\Delta_i}.$$

Définition 1.7 : **(Censure de type II)**.

Étant donné un échantillon $\mathbb{X} = (X_1, ..., X_n)$ de durées de survie X_i et un nombre entier positif r.

On dit qu'il y a censure de type **II**, si au lieu d'observer $X_1, ..., X_n$ on observe n stratistiques

$$(T_1, \Delta_1), ..., (T_n, \Delta_n),$$

où

$$T_i = \min(X_i, X_{(r)}), \quad \Delta_i = \mathbf{1}_{\{T_i = X_i\}},$$

$X_{(r)}$ est la r-ième statistique d'ordre, i.e. $X_{(r)}$ est la r-ième composante du vecteur des statistiques d'ordre $X = (X_{(1)}, ..., X_{(n)})$ associé à l'échantillon $\mathbb{X}$,

$$0 < X_{(1)} < X_{(2)} < ... < X_{(r)} < ... < X_{(r)}.$$

C'est-à-dire que dans la situation considérée, la date de censure est $X_{(r)}$ et les observations sont :

$$\begin{aligned} T_{(i)} &= X_{(i)}, \qquad i = 1, 2, ..., r \\ T_{(i)} &= X_{(r)}, \qquad i = r, r+1, ..., n. \end{aligned}$$

Si :

$$X_i \backsim f(x_i; \theta) \quad \text{et} \quad S(x_i; \theta) = \mathbf{P}\{X_i > x_i\}, \quad x_i > 0, \quad \theta \in \Theta,$$

alors la fonction de vraisemblance associée aux statistiquues

$$(T_1, \Delta_1), ..., (T_n, \Delta_n),$$

est :

$$\begin{aligned} L(\theta) &= \frac{n!}{(n-r)!} \prod_{i=1}^{n} f(T_{(i)}; \theta)^{\Delta_{(i)}} S(T_{(i)}; \theta)^{1-\Delta_{(i)}} \\ &= \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(X_{(i)}; \theta) S(X_{(r)}; \theta)^{n-r}, \end{aligned}$$

puisque $\sum_{i=1}^{n} \Delta_i = r$, où r est donné.

Définition 1.8 : **(Censure de type III) censure aléatoire**

Étant donné un échantillon $\mathbb{X} = (X_1, ..., X_n)$ de durées de survie X_i.

On dit qu'il y a censure aléatoire de cet échantillon s'il existe un autre échantillon $\mathbb{C}$

$$\mathbb{C} = (C_1, ..., C_n)$$

indépendant de $\mathbb{X}$, tel que au lieu d'observer $X_1, ..., X_n$ on observe les statistiques

$$(T_1, \Delta_1), ..., (T_n, \Delta_n),$$

où

$$T_i = \min(X_i, C_i), \qquad \Delta_i = \mathbf{1}_{\{T_i = X_i\}}.$$

1.5 Processus de comptage et l'estimation du risque cumulé

1.5.1 Processus de comtage

Définition 1.9 : Une observation consiste en (T_i, Δ_i), où

$$\begin{cases} T_i & = & \min(X_i, C_i) \\ \Delta_i & = & 1_{\{X_i \le C_i\}}. \end{cases}$$

Nous définissons le **processus de comptage** par :

$$N_i\left(t\right) = 1_{\{T_i \le t \text{ et } \Delta_i = 1\}}.$$

Si l'individu i subit l'événement avant l'instant t, alors $N_i\left(t\right) = 1$; si non, $N_i\left(t\right) = 0$.

Nous définissons également la fonction

$$Y_i(t) = 1_{\{T_i \ge t\}},$$

qui indique le nombre des unités à risque au moment t^-.

On suppose que la variable aléatoire X_i est absolument continue et pour tout t tel que $\mathbf{P}\left\{T_i > t\right\} > 0$ il existe la limite

$$\alpha_{c_i}(t) = \lim_{h \downarrow 0} \frac{\mathbf{P}\left\{T_i \in [t, t+h[\,, \Delta_i = 1 \mid T_i \ge t\right\}}{h}.$$

On dit que la censure est indépendante si :

$$\alpha_{c_i}(t) = \alpha_i(t).$$

Notons que :

$$\begin{aligned} \alpha_{c_i}(t) &= \lim_{h \downarrow 0} \frac{\mathbf{P}\left\{t \le X_i < t+h,\ X_i \le C_i\right\}}{h\mathbf{P}\left\{X_i \ge t,\ C_i \ge t\right\}} \\ &= \lim_{h \downarrow 0} \frac{\mathbf{P}\left\{X_i \le C_i \mid t \le X_i < t+h\right\}}{h\mathbf{P}\left\{X_i \ge t,\ C_i \ge t\right\}} \cdot \mathbf{P}\left\{t \le X_i < t+h\right\} \\ &= \frac{\mathbf{P}\left\{C_i \ge X_i \mid X_i = t\right\} f(t)}{\mathbf{P}\left\{X_i \ge t,\ C_i \ge t\right\}} \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Donc l'égalité $\alpha_{c_i}(t) = \alpha_i(t)$ est équivalente à l'égalité

$$\mathbf{P}\left\{C_i \geq t \mid X_i = t\right\} = \frac{\mathbf{P}\left\{X_i \geq t,\ C_i \geq t\right\}}{S(t)}.$$

Remarque 1.4 : On a :

$$\begin{aligned}
\alpha_c(t) &= \lim_{h\downarrow 0} \frac{\mathbf{P}\left\{t \leq X_i < t+h,\ C_i \geq t\right\}}{h\mathbf{P}\left\{X_i \geq t,\ C_i \geq t\right\}} \\
&= -\frac{1}{\mathbf{P}\left\{X_i \geq t,\ C_i \geq t\right\}} \frac{\partial}{\partial s}\left[\mathbf{P}\left\{X_i \geq s,\ C_i \geq t\right\}\right] |_{s=t} .
\end{aligned}$$

Notons :

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\alpha_i(u)du.$$

Proposition 1.3 : Si la censure est indépendante, alors :

$$\mathbf{E}M_i(t) = 0,$$

pour tout t telque $\mathbf{P}\left\{T_i \geq t\right\} > 0$.

Preuve : L'égalité

$$\mathbf{P}\left\{C_i \geq t \mid X_i = t\right\} = \frac{\mathbf{P}\left\{X_i \geq t,\ C_i \geq t\right\}}{S(t)},$$

implique que :

$$\begin{aligned}
\mathbf{E}M_i(t) &= \mathbf{E}N_i(t) - \textstyle\int_0^t \mathbf{E}Y_i(u)\alpha_i(u)du \\
&= \mathbf{P}\left\{X_i \leq t,\ X_i \leq C_i\right\} - \textstyle\int_0^t \mathbf{P}\left\{X_i \geq u,\ C_i \geq u\right\}\alpha_i(u)du \\
&= \textstyle\int_0^t \mathbf{P}\left\{C_i \geq u \mid X_i = u\right\} f(u)du \\
&\quad - \textstyle\int_0^t \mathbf{P}\left\{C_i \geq u \mid X_i = u\right\} S(u)\alpha_i(u)du = 0.
\end{aligned}$$

La proposition est démontrée.■

Proposition 1.4 : Le processus stochastique définit par :

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\alpha_i(u)du$$

est une martingale par rapport à la filtration $\mathcal{F}_t$, $t \geq 0$ où $\mathcal{F}_t$ est la σ-algèbre engendrée par les prcessus $N_i(t)$ et $Y_i(t)$:

$$f_t = \sigma\left\{N_i(s),\ Y_i(s) :\ 0 \leq s \leq t\right\}.$$

Preuve : Nous avons :

$$\mathbf{E}\left\{N_i(t) - N_i(s) \mid \mathcal{F}_s\right\} = \mathbf{E}\left\{\int_s^t Y_i(u)\alpha_i(u)du \mid \mathcal{F}_s\right\},$$

ce qui implique que :

$$\begin{aligned} \lim_{h\downarrow 0}\frac{1}{h}\mathbf{E}\left\{N_i(t) - N_i(s) \mid \mathcal{F}_s\right\} &= \lim_{h\downarrow 0}\frac{1}{h}\mathbf{E}\left\{\int_s^{s+h} Y_i(u)\alpha_i(u)du \mid \mathcal{F}_s\right\} \\ &= \mathbf{E}\left\{Y_i(s)\alpha_i(s) \mid \mathcal{F}_s\right\} \\ &= Y_i(s)\alpha_i(s), \end{aligned}$$

d'où

$$\mathbf{E}\left\{M_i(t) \mid \mathcal{F}_s\right\} = M_i(s).\ \blacksquare$$

1.5.2 Estimation du risque cumulé

Considérons un échantillon de n individus.

Définissons :

$$\mathbf{N}_+(t) = \sum_{i=1}^{n} N_i(t)$$

et

$$Y_+(t) = \sum_{i=1}^{n} Y_i(t).$$

Définition 1.10 : L'**estimateur de Nelson-Aalen** de la fonction de risque cumulé est définit par(Nelson, (1972) ; Aalen, (1978))

$$\widehat{A}\,(t) = \int_0^t \frac{J\,(u)}{Y_+(u)} d\mathbf{N}_+(u),$$

où $J\,(t) = \mathbf{1}_{\{Y(t)>0\}}$.

Théorème 1.1 : [21]

$\widehat{A}\,(t)$ est un estimateur biaisé de $A(t)$ et, sous l'hypothèse que $F(t) < 1$(i.e. $A(t) < \infty$), nous avons :

$$\sqrt{n}\left(\widehat{A}\,(t) - A(t)\right) \xrightarrow{\pounds} U(t),$$

avec U est une martingale gaussienne telle que :

$$\begin{cases} U\,(0) & = & 0 \\ \mathbf{V}\,(U\,(t)) & = & \displaystyle\int_0^t \frac{\alpha\,(u)}{y\,(u)} du, \end{cases}$$

où

$$y(s) = [1 - F(s)]\,[1 - G(s-)]$$

avec G est la fonction de répartition de C.

Chapitre 2

Estimateur des quantiles de type noyau des données censurées à droite (DCA) et l'estimateur de Kaplan-Meier

2.1 Estimation de la fonction de survie

Si l'on ne peut pas supposer a priori que la loi de la durée de survie X obéit à un modèle paramétrique, on peut estimer la fonction de survie $S(t)$ grâce à plusieurs méthodes non paramétriques dont la plus intéressante est celle de Kaplan-Meier (1958).

Cet estimateur est aussi appelé un estimateur à limite produit (P-L estimateu), car il s'obtient comme un produit.

La probabilité de survivre au delà à l'instant $t_{(n)}$ est éagle au produit suivant :

$$\begin{aligned} S(t_{(n)}) &= \mathbf{P}\left\{X > t_{(n)}\right\} \\ &= \mathbf{P}\left(X > t_{(n)} \mid X > t_{(n-1)}\right) \cdot S(t_{(n-1)}) \\ &= D_n p_{t_{(n-1)}} S(t_{(n-1)}), \end{aligned}$$

où

$$0 = t_{(0)} < t_{(1)} < ... < t_{(n)},$$

et

$$D_n p_{t_{(n-1)}} = S_{t_{(n-1)}}\left(D_n\right), \qquad D_n = t_{(n)} - t_{(n-1)},$$

avec

$$S_s\left(t\right) = \frac{S(s+t)}{S(s)},$$

et $t_{(n-1)}$ est une date antérieur à $t_{(n)}$.

Si on renouvelle l'opération en choisissant une date $t_{(n-2)}$ antérieure à $t_{(n-1)}$, on aura de même

$$\begin{aligned} S_{t_{(n-1)}} &= \mathbf{P}\left\{X > t_{(n-1)}\right\} \\ &= \mathbf{P}\left(X > t_{(n-1)} \mid X > t_{(n-2)}\right) \cdot S(t_{(n-2)}) \end{aligned}$$

et ainsi de suite on obtient la formule :

$$S(t_{(n)}) = \prod_{i=1}^{n} D_i p_{t_{(i-1)}} = \prod_{i=1}^{n}\left(1 - D_i q_{t_{(i-1)}}\right),$$

avec $S(0) = 1$.

Si on désigne par $t_{(i)} = T_{(i)}$, on aura seulement à estimer des quantités de la forme :

$$p_i = \mathbf{P}\left(X > T_{(i)} \mid X > T_{(i-1)}\right) = D_i p_{T(i-1)},$$

où p_i est la probabilité de survivre pendant l'intervalle de temps $D_i = \left]T_{(i-1)}, T_{(i)}\right]$ quand on était vivant au début de cet intervalle.

On note que :

$$0 = T_{(0)} \leq T_{(1)} \leq ... \leq T_{(n)},$$

Notons par :

$R_i =$ **card** $(T_{(i)}^{-})$ le nombre des sujets qui sont vivant juste avant l'instant $T_{(i)}$.

$M_i =$ le nombre de morts à l'instant $T_{(i)}$;

$q_i = 1 - p_i$ la probabilité de mourir pendant l'intervalle D_i sachant que l'on était vivant au début de cet intervalle.

Alors l'estimateur naturel de q_i est :

$$\widehat{q_i} = \frac{M_i}{R_i}.$$

Supposons d'abord qu'il n'y ait pas d'ex-aequo, c'est à dire on suppose que :

$$0 = T_{(0)} < T_{(1)} < ... < T_{(n)}.$$

Dans ce cas,

Si $\Delta_{(i)} = 1$, c'est qu'il y a eu un mort en $T_{(i)}$ et donc $M_i = 1$,

Si $\Delta_{(i)} = 0$, c'est qu'il y a eu une censure en $T_{(i)}$ et donc $M_i = 0$.

Par suite :

$$\widehat{p_i} = 1 - \frac{M_i}{R_i} = \left(1 - \frac{1}{R_i}\right)^{\Delta_{(i)}} = \begin{cases} 1 - \frac{1}{R_i}, & \text{en cas de mort en } T_{(i)}, \\ 1, & \text{en cas de censure en } T_{(i)}, \end{cases}$$

donc $\widehat{p_i}$ n'est différent de 1 qu'aux instants de décès observés.

L'estimateur de Kapan-Meier de la fonction de survie est défini par Kaplan et Meier (1958) :

$$\begin{aligned} \widehat{S}(t) &= \widehat{S}_n(t) = \prod_{T_{(i)} \leq t} \widehat{p_i} = \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{R_i}\right)^{\Delta_{(i)}} \\ &= \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{\Delta_{(i)}}. \end{aligned}$$

Il est évident qu'en absence de la censure, c'est à dire quand $\Delta_i = 1$, $\forall i$, alors :

$$\widehat{S}_n(t) = \begin{cases} 1, & \text{si } t < T_{(1)}, \\ \displaystyle\prod_{T_{(i)} \leq t} \frac{n - i}{n - i + 1}, & \text{si } T_{(i)} \leq t < T_{(i+1)}, \quad i = 1, ..., n - 1, \\ 0, & \text{si } t \geq T_{(n)}. \end{cases}$$

On remarque que $R_i = n - i + 1$.

Il est évident que l'estimateur de Kaplan-Meier $\widehat{F}_n\,(t)$ de $F(t) = 1 - S(t)$ est :

$$\widehat{F}_n\,(t) = 1 - \widehat{S}_n\,(t) = \begin{cases} 0, & \text{si } t < T_{(1)}, \\ 1 - \prod_{T_{(i)} \leq t} \left(\frac{n-i}{n-i+1}\right)^{\Delta_{(i)}}, & \text{si } T_{(i)} \leq t < T_{(i+1)}, \\ 1, & \text{si } t \geq T_{(n)}. \end{cases}$$

Enfin on note que :

$$\widehat{A}_n(t) = -ln\widehat{S}_n\,(t)\,,$$

peut-être considéré comme l'estimateur de Kaplan-Meier de la fonction de risque cumulé $A(t)$.

Définition 2.1 : L'**estimateur de Kaplan-Meier** de la fonction de survie est défini par Kaplan et Meier (1958)

$$\begin{aligned} \widehat{S}\,(t) &= p_{s \leq t} \left(1 - d\widehat{A}\,(s)\right) \\ &= p_{s \leq t} \left(1 - \frac{J(s)}{Y_+(s)} dN_+(s)\right) \end{aligned}$$

ou encore :

$$\begin{aligned} \widehat{S}\,(t) &= \prod\nolimits_{s \leq t} \left(1 - \Delta\widehat{A}\,(s)\right) \\ &= \prod\nolimits_{s \leq t} \left(1 - \frac{J(s)\Delta N_+(s)}{Y_+(s)}\right), \end{aligned}$$

où $\widehat{A}\,(s)$ est l'estimateur de Nelson-Aalen et où, pour un processus $X(s)$ cadlag (continu à droite avec une limite à gauche),

$$\Delta X(s) = X(s) - X(s-).$$

Proposition 2.1 : Un premier estimateur de la variance de $\widehat{S}(t)/S(t)$ est :

$$\widehat{\sigma}^2(t) = \int_0^t \frac{J(s)}{Y_+(s)^2} dN_+(s).$$

L'estimateur de la variance de $\widehat{S}(t)$ correspondant est donné par :

$$\widehat{\mathbf{V}}\left(\widehat{S}(t)\right) = \left[\widehat{S}(t)\right]^2 \widehat{\sigma}^2(t).$$

Propsition 2.2 : Un second estimateur de la variance de $\widehat{S}(t)/S(t)$ est l'estimateur de Greenwood (Greenwood (1926))

$$\widehat{\widehat{\sigma}}^2(t) = \int_0^t \frac{dN_+(u)}{Y_+(s)\left[Y_+(s) - \Delta N_+(s)\right]}.$$

L'estimateur de la variance de $\widehat{S}(t)$ correpondant est donné par :

$$\widehat{\widehat{\mathbf{V}}}\left[\widehat{S}(t)\right] = \left[\widehat{S}(t)\right]^2 \widehat{\widehat{\sigma}}^2(t).$$

2.1.1 Lois asymptotique

Théorème 2.1 : L'estimateur $\widehat{S}(t)$ est biaisé ; de plus, si nous supposons que :

A. pour tout $s \in [0, t]$,

$$n \int_0^s \frac{J(u)}{Y_+(u)} \alpha(u)\, du \xrightarrow{P} \sigma^2(s) \quad (n \to \infty).$$

B. pour tout $\epsilon > 0$,

$$n \int_0^t \frac{J(s)}{Y_+(s)} \alpha(s)\, 1_{\left\{\sqrt{n}|J(s)/Y(s)| > \epsilon\right\}} ds \xrightarrow{P} 0 \quad (n \to \infty).$$

C.
$$\sqrt{n}\int_0^t (1-J(u))\,\alpha(u)\,du \xrightarrow{P} 0 \qquad (n\to\infty),$$

alors cet estimateur vérifié asymptotiquement

$$\sqrt{n}\left(\widehat{S}(t)-S(t)\right) \xrightarrow{\pounds} -U(t)S(t),$$

où U est la martingale gaussienne définie en page (22).

Preuve : Concernant le biais, posons :

$$\begin{aligned} S^*(t) &= p_{s\leq t}\,(1-dA^*(s)) \\ &= \exp(-A^*(t)) \end{aligned}$$

où
$$A^*(t)=\int_0^t J(u)\alpha(u)\,du.$$

D'après le théorème de **Duhamel** (voir. théorème 2, Annexe), on a :

$$\begin{aligned} \frac{\widehat{S}(t)}{S^*(t)}-1 &= -\int_0^t \frac{\widehat{S}(s-)}{S^*(s)}d\left(\widehat{A}-A^*\right)(s) \\ &= -\int_0^\tau \frac{\widehat{S}(s-)\,J(s)}{S^*(s)Y_+(s)}dM_+(s) \qquad (2.1) \end{aligned}$$

pour $t\in[0,\tau)$, et avec $dM_+(t)=dN_+(t)-Y_+(t)\alpha(t)\,dt$.

Comme $\widehat{S}/S^*-1$ une martingale alors :

$$\mathbf{E}\left[\frac{\widehat{S}(t)}{S^*(t)}\right]=1,$$

pour tout $t \in [0, \tau)$.

Enfin,

$$\begin{array}{rccc} & S(t) & \leq & S^*(t) \\ \Longrightarrow & \dfrac{\widehat{S}(t)}{S^*(t)} & \leq & \dfrac{\widehat{S}(t)}{S(t)} \\ \Longrightarrow & \mathbf{E}\left[\dfrac{\widehat{S}(t)}{S^*(t)}\right] & \leq & \mathbf{E}\left[\dfrac{\widehat{S}(t)}{S(t)}\right] \\ \Leftrightarrow & 1 & \leq & \dfrac{\mathbf{E}(\widehat{S}(t))}{S(t)} \\ \Leftrightarrow & \mathbf{E}(\widehat{S}(t)) & \geq & S(t). \end{array}$$

Concernant la convergence asymptotique, d'après (2.2),

$$\sqrt{n}\left\{\frac{\widehat{S}(t)}{S^*(t)} - 1\right\} = -\sqrt{n}\int_0^t \frac{\widehat{S}(s-)\,J(s)}{S^*(s)Y_+(s)}dM_+(s).$$

Concernant la convergence asymptotique, d'après (2.2),

$$\sqrt{n}\left\{\frac{\widehat{S}(t)}{S^*(t)} - 1\right\} = -\sqrt{n}\int_0^t \frac{\widehat{S}(s-)\,J(s)}{S^*(s)Y_+(s)}dM_+(s).$$

D'après les conditions A et B, et d'après le fait que $\widehat{S}(s)/S^*(s) \leq 1/S(t)$ pour tout $s \in [0, t]$, nous déduisons du théorème de **Rebolledo** (théorème 1, Annexe) que :

$$\sqrt{n}\left(\frac{\widehat{S}(t)}{S^*(t)} - 1\right) \xrightarrow{\pounds} -U(t) \text{ quand } n \to \infty.$$

La condition C et le fait que :

$$\begin{aligned}\sqrt{n}\left|\frac{S(s)}{S^*(s)}-1\right| &= \int_0^s \frac{S(u)}{S^*(u)} d\left(A-A^*\right)(u) \\ &\leq \frac{1}{S(t)}\int_0^s \left(1-J(u)\right)\alpha\left(u\right)du\end{aligned}$$

entraînent que :

$$\sup_{s\in[0,t]} \sqrt{n}\left|\frac{S(s)}{S^*(s)}-1\right| \xrightarrow{P} 0 \qquad (n\to\infty)\,.$$

D'où il s'ensuit que :

$$\sqrt{n}\frac{\widehat{S}\left(s\right)-S(s)}{S^*(s)} \xrightarrow{\pounds} -U \qquad (n\to\infty)$$

et nous obtenons le résultat du théorème.■

2.2 Estimateur à noyau des quantiles

Soit $X_1,...,X_n$ une suite des temps de survie de n individus qui sont supposés censurées (à droite) par la suite $C_1,C_2,...,C_n$ qui sont en générale des constantes ou des variables aléatoires i.i.d. ayant une distribution commune G (inconnu) (voir, Page (18)).

On suppose que les X_i, $i\geq 1$, sont des variables aléatoires non négatives ayant une distribution commune F (inconnu).

Les données censurées à droite sont notées par :

$$T_i=\min\left(X_i,C_i\right) \quad , \qquad \Delta_i=\begin{cases}1, & \text{si} \quad X_i\leq C_i \\ 0, & \text{si} \quad X_i>C_i.\end{cases}$$

Notons que la distibution de chaque $(T_i)_{i=1,n}$ est :

$$H=1-(1-F)(1-G).$$

On définit la fonction des quantiles théorique comme suit :

$$Q(p) \equiv \xi_p = \inf\{t : F(t) \geq p\}, \quad 0 \leq p \leq 1.$$

Le problème consiste à estimer $Q(p)$ pour p fixé.

Soit $\hat{F}_n(t)$ l'estimateur de Kaplan-Meier de $F(t)$ (cf. page 26).

L'estimateur classique de $Q(p)$ est :

$$\hat{Q}_n(p) = \inf\left\{t : \hat{F}_n(t) \geq p\right\}.$$

L'estimateur à noyau des quantiles qui à été introduit par Padjett (1986) est défini par :

$$Q_n(p) = h_n^{-1} \int_0^1 \hat{Q}_n(t) K(t - p/h_n) dt,$$

où K est un noyau choisi convenablement, $\{h_n\}$ est une suite de fenêtres.

Soit s_j le saut de $\widehat{S}_n$ (ou $\widehat{F}_n$) en T_j qui est défini par :

$$s_j = \begin{cases} 1 - \widehat{S}_n(T_2), & j = 1 \\ \widehat{S}_n(T_j) - \widehat{S}_n(T_{j+1}), & j = 2, ..., n-1 \\ \widehat{S}_n(T_n), & j = n \end{cases}$$

On a :

$$s_j = 0 \text{ si et seulement si } \Delta_{(j)} = 0, \qquad j < n$$

où les $\Delta_{(j)}$ sont les statistiques d'ordre associées avec Δ_j.

On définit S_i comme suit :

$$S_i \equiv \hat{F}_n(T_{i+1}) = \sum_{j=1}^{i} s_j, \quad (i = 1, 2, ..., n), \; S_n \equiv 1.$$

On suppose que les conditions suivantes sont vérifiées :

Soit $\{h_n\}$ une suite de fenêtres de nombres positifs telles que

1. $h_n \to 0$ quand $n \to \infty$.

Soit K la fonction des valeurs réelles définie sur $]-\infty, +\infty[$ telle que

2- $K(x) \geq 0$,pour tout les nombres réelles x.

3- $\int_{-\infty}^{+\infty} K(x)\, dx = 1$.

4- K à un support fini, c'est-à-dire il existe $c > 0$ telle que $K(x) = 0$ pour $|x| > c$.

5- K est symétrique au voisinage de zéro.

6- K satisfait la condition de lipshitz :

$$\exists \text{ une constante } \Gamma \text{ tel que} : \forall x, y, \ |K(x) - K(y)| \leq \Gamma |x - y| .$$

Notons que les conditions 2 et 3 désignent que K est une densité de probabilité.

De plus F on doit vérifiée :

7- F est continue de densité f.

8- f est continue en ξ_p et $f(\xi_p) > 0, \quad 0 < p < 1$.

9- $H(T_F) \leq 1$, tel que $T_F = \sup \{t : F(t) < 1\}$.

10- F admet une moyenne finie.

L'estimateur à noyau des quantiles est définit pour $0 \leq p \leq 1$ comme :

$$\begin{aligned} Q_n(p) &= h_n^{-1} \int_0^1 \hat{Q}_n(t) K(t - p/h_n)\, dt \\ &= h_n^{-1} \sum_{i=1}^{n} Z_i \int_{S_{i-1}}^{S_i} K(t - p/h_n)\, dt, \end{aligned} \tag{2.2}$$

où $S_0 \equiv 0$.

L'approximation à $Q_n(p)$ est définit par :

$$Q_n^*(p) = h_n^{-1} \sum_{i=1}^{n} Z_i s_i K((S_i - p)/h_n)$$

2.3 Erreur quadratique moyenne(EQM) des estimateurs des quantiles pour les (DCA)

Nous étudions dans cette partie l'erreur quadratique moyenne de l'estimateur à limite produit des quantiles et l'estimateur à noyau des quantiles.

2.3.1 EQM de l'estimateur à limite produit des quantiles

Le théorème suivant donne l'expression asymptotique de l'erreur quadratique moyenne de l'estimateur à limite prioduit des quantiles.

Théorème 2.2 : [18]

Soit p tel que $0 \leq p < \min \left\{1, T_{G(Q)}\right\}$.

Supposons que G est continue, $\mathbf{E}\left(X^2\right) < \infty$, et Q est deux fois différentiable ayant une deuxième dérivée bornée au voisinage de p. Alors :

$$\mathbf{E}\left\{\left[\widehat{Q}_n\left(p\right) - Q\left(p\right)\right]^2\right\},$$

existe, et

$$\begin{aligned}\mathbf{E}\left\{\left[\widehat{Q}_n\left(p\right) - Q\left(p\right)\right]^2\right\} &= n^{-1}\left(Q'\left(p\right)\right)^2\left(1-p\right)^2 \\ &\quad \times \int_0^p \frac{dx}{\left(1-x\right)^2\left(1-H\left(Q\left(p\right)\right)\right)} \\ &\quad + o\left(n^{-3/2}\right) + o\left(n^{-1}\right).\end{aligned}$$

2.3.2 Erreur quadratique moyenne de l'(ENQ)

Ce théorème donne l'expression asymptotique de l'erreur quadratique moyenne de l'estimateur des quantiles à noyau.

Théorème 2.3 : [18]

Soit p tel que $0 \leq p < \min\left\{1, T_{G(Q)}\right\}$.

Supposons que G est continue, $\mathbf{E}\left(X^2\right) < \infty$, Q est deux fois différentiable avec une deuxième dérivée bornée au voisinage de p, et $Q'(p) > 0$, et on suppose que :

$$K \text{ à support } [-c, c], \int K(x)\, dx = 1 \text{ et } \int xK(x)\, dx = 0, \quad c > 0.$$

Alors :

$$\begin{aligned}
\mathbf{E}\left\{\left[Q_n(p) - Q(p)\right]^2\right\} &= n^{-1}(Q'(p))^2 (1-p)^2 \\
&\quad \times \int_0^p \frac{dx}{(1-x)^2 (1 - G(Q(x)))} \\
&\quad + 2n^{-2} (1-p)^2 (Q'(p))^2 \\
&\quad \int_{-c}^{c} K(t) \left[1 - \tilde{K}(t)\right] \\
&\quad \times \int_p^{p+ht} \frac{dx}{(1-x)^2 (1 - G(Q(x)))} dt \\
&\quad + O\left(n^{-3/2}\right) + O\left(h^2\right) + O\left(h^2 n^{-1}\right) \\
&\quad + o(hn^{-1}) + o(n^{-1}),
\end{aligned}$$

où $\tilde{K}(t) = \int_{-c}^{t} K(x)\, dx, \ -c \leq t \leq c.$

Preuve : Avec le théorème (3.1) de Falk (1985), on a : pour $0 \leq p < \min\left\{1, T_{G(Q)}\right\}$,

$$\begin{aligned}
\mathbf{E}(X^2) &\geq \mathbf{E}\left[\sum_{i=1}^{n} X^2 s_i\right] = \mathbf{E}\left[\int_0^{\infty} X^2 d\widehat{F}_n(x)\right] \\
&= \mathbf{E}\left[\int_0^1 \widehat{Q}_n^2(\tau)\, d\tau\right] \geq (1-p)\mathbf{E}[\widehat{Q}_n^2(\tau)].
\end{aligned}$$

Donc, $\mathbf{E}\left\{\left[\widehat{Q}_n(p)-Q(p)\right]^2\right\}<\infty$. De plus,

$$\mathbf{E}\left\{\left[Q_n(p)-Q(p)\right]^2\right\}$$
$$=\mathbf{E}\left\{\left[\int_{-c}^{c}\left(\widehat{Q}_n(p+hu)-Q(p+hu)\right)\cdot K(u)\,du\right]^2\right\}$$
$$+\left\{\int_{-c}^{c}\left[Q(p+hu)-Q(p)\right]K(u)\,du\right\}^2$$
$$+2\mathbf{E}\left\{\int_{-c}^{c}\left[Q(p+hu)-Q(p)\right]K(u)\,du\cdot\right.$$
$$\left.\int_{-c}^{c}\left[\widehat{Q}_n(p+hu)-Q(p+hu)\right]K(u)\,du\right\}.$$

A condition que Q a une deuxième dérivée bornée au voisinage de p, on a :

$$\left\{\int_{-c}^{c}\left[Q(p+hu)-Q(p)\right]K(u)\,du\right\}^2=O\left(h^2\right).$$

Pour $\varepsilon>0$, on définit les évènements

$$A_n=\left\{\left|U_n(p+hu)-(p+hu)\right|>\varepsilon\right\},$$

où $U_n$est la fonction à (p.L) des quantiles basée sur la distribution uniforme sur $(0,1)$.

Avec la propriété de symétrie de Sander (1975), et par Földes et Rejtö (1981), $\varepsilon>0$ est choisi de telle sorte que $P(A_n)\leq d_0\exp(-nd_1)$, où d_0 et d_1 des constantes positives, et on peut choisir h petit tel que $hc<\min\left\{1,T_{G(Q)}\right\}$.

Nous avons,

$$\mathbf{E}\left\{\left[\int_{-c}^{c}\left(\widehat{Q}_n(p+hu)-Q(p+hu)\right)K(u)\,du\right]^2\right\}=E_1+E_2,$$

où

$$E_1 = \mathbf{E}\left\{\left[\int_{-c}^{c}\left(\widehat{Q}_n\left(p+hu\right)-Q\left(p+hu\right)\right)K\left(u\right)du\right]^2 I_{A_n^c}\right\}$$

et

$$E_2 = \mathbf{E}\left\{\left[\int_{-c}^{c}\left(\widehat{Q}_n\left(p+hu\right)-Q\left(p+hu\right)\right)K\left(u\right)du\right]^2 I_{A_n}\right\}.$$

Puisque $p+hu<\min\left\{1,T_{G(Q)}\right\}$, $|E_2| = O\left(\exp\left(-nd_1'\right)\right)$.

En appliquant la formule de Taylor à E_1 et l'inégalité de Sander (1975), on obtient :

$$\begin{aligned} E_1 &= \mathbf{E}\left\{\left[\int_{-c}^{c}K\left(x\right)\left(U_n\left(p+hu\right)-\left(p+hu\right)\right)Q'\left(p+hu\right)du\right]^2\right\} \\ &\quad +O\left(\mathbf{E}\left[\sup_{0\le p\le T^*}(\left|\widehat{U}_n\left(p\right)-p\right|^3\right]\right)+O(\exp(-nd_1')), \end{aligned}$$

où $p<p+hc<T^*<\min\left\{1,T_{G(Q)}\right\}$.

De la preuve de théorème (2) de Lio, Padjett et Yu (1986) pour n grand,

$$O\left(\mathbf{E}\left[\sup_{0\le p\le T^*}(\left|\hat{U}_n\left(p\right)-p\right|^3\right]\right)=O\left(n^{-3/2}\right).$$

On a aussi,

$$\begin{aligned} &\mathbf{E}\left\{\left[\int_{-c}^{c}K\left(x\right)\left(U_n\left(p+hx\right)-\left(p+hx\right)\right)Q'\left(p+hx\right)dx\right]^2\right\} \\ = \ & n^{-1}\mathbf{E}\left\{\left[\int_{-c}^{c}K(x)\left[n^{1/2}\left(U_n\left(p+hx\right)-\left(p+hx\right)\right)\right.\right.\right. \\ & \left.\left.\left.-\,n^{-1/2}K^*\left(p+hx,n\right)\right]dx\right]^2\right\}\cdot\left(Q'\left(p\right)\right)^2 \\ & +2n^{-1}\int_{-c}^{c}K(x)\left[n^{1/2}\left(U_n\left(p+hx\right)-\left(p+hx\right)\right)-n^{-1/2}K^*\left(p+hx,n\right)\right]Q'(p) \end{aligned}$$

$$
\begin{aligned}
& \cdot dx\left[\int_{-c}^{c} K(x) n^{-1/2} K^{*}(p+hx, n)\right] Q'(p) \\
& +n^{-1}\mathbf{E}\left(Q'(p)\right)^{2}\left[\int_{-c}^{c} K(x) n^{-1/2} K^{*}(p+hx, n)\right]^{2} \\
= \; & o\left(n^{-1}\right)+n^{-1}\left(Q'(p)\right)^{2}\mathbf{E}\left[\int_{-c}^{c} K(x) n^{-1/2} K^{*}(p+hx, n)\right]^{2}.
\end{aligned}
$$

On utilisant les résultats de Aly, Csörgő et Horváth (1985), on obtient :

$$
\begin{aligned}
& \mathbf{E}\left\{\left[\int_{-c}^{c} K(x) n^{-1/2} K^{*}(p+hx, n)\right]^{2}\right\} \\
= \; & \mathbf{E}\left\{\int_{-c}^{c} n^{-1}(1-p-hx) W(d(p+hx), n) K(x) dx\right. \\
& \left.\times \int_{-c}^{c}(1-p-ht) W(d(p+ht), n) K(t) dt\right\} \\
= \; & A_1+A_2+A_3,
\end{aligned}
$$

où

$$
\begin{aligned}
A_1 = \; & n^{-1}\mathbf{E}\left\{\int_{-c}^{c}\int_{-c}^{c}(1-p)^{2} W(d(p+hx), n) W(d(p+ht), n)\right. \\
& \left.\times K(x) K(t) dxdt\right\},
\end{aligned}
$$

$$
\begin{aligned}
A_2 = \; & -2n^{-1}h(1-p)\mathbf{E}\left\{\int_{-c}^{c}\int_{-c}^{c} xK(x) K(t) W(d(p+hx), n)\right. \\
& \left.\times W(d(p+ht), n) dxdt\right\},
\end{aligned}
$$

$$
\begin{aligned}
A_3 = \; & n^{-1}h^{2}\mathbf{E}\left\{\int_{-c}^{c}\int_{-c}^{c} xtK(x) K(t) W(d(p+hx), n)\right. \\
& \left.\times W(d(p+ht), n) dxdt\right\},
\end{aligned}
$$

et $W(s,t)$ désigne le processus de Weiner à deux paramètres (voir, Annexe) avec :

$$
\mathbf{E}[W(s,t)]=0,
$$

et

$$\mathbf{E}\left[W(s,t)\,W\left(s',t'\right)\right]=\min\left\{s,s'\right\}\min\left\{t,t'\right\},$$

et avec :

$$dt=\int\limits_{-\infty}^{t}(1-x)^{-2}\left[1-G(Q(x))\right]^{-1}dx,$$

ce qui implique :

$$\begin{aligned}
A_1 &= (1-p)^2\int_{-c}^{c}\int_{-c}^{t}K(t)\,K(u)\int_0^{p+hu}(1-x)^{-2}\left[1-G(Q(x))\right]^{-1}\\
&\quad\cdot dxdudt+(1-p)^2\int_{-c}^{c}\int_{t}^{c}K(t)\,K(u)\\
&\quad\cdot\int_0^{p+ht}(1-x)^{-2}\left[1-G(Q(x))\right]^{-1}dxdudt\\
&= (1-p)^2\left\{\int_{-c}^{c}\int_{-c}^{t}K(t)\,K(u)\right.\\
&\quad\cdot\int_0^{p}(1-x)^{-2}\left[1-H(Q(x))\right]^{-1}dxdudt\\
&\quad+\int_{-c}^{c}\int_{-c}^{t}K(t)\,K(u)\int_p^{p+hu}(1-x)^{-2}\left[1-G(Q(x))\right]^{-1}dxdudt\\
&\quad+\int_{-c}^{c}K(t)\left[1-\tilde{K}(t)\right]\int_0^{p}(1-x)^{-2}\left[1-G(Q(x))\right]^{-1}dxdt\\
&\quad\left.+\int_{-c}^{c}K(t)\left[1-\tilde{K}(t)\right]\int_p^{p+ht}(1-x)^{-2}\left[1-G(Q(x))\right]^{-1}dxdt\right\}.
\end{aligned}$$

On combine le premier et le troisième terme dans la dernière expression de A_1, et avec la combinaison de deuxième et quatrième terme on obtient :

$$
\begin{aligned}
A_1 &= (1-p)^2 \int_0^p (1-x)^{-2} [1-G(Q(x))]^{-1} dx \\
&+2(1-p)^2 \int_{-c}^{c} K(t) \left[1-\tilde{K}(t)\right] \\
&\cdot \int_p^{p+ht} (1-x)^{-2} [1-G(Q(x))]^{-1} dxdt.
\end{aligned}
$$

Avec les même arguments, A_2 et A_3 sont devient :

$$
\begin{aligned}
A_2 &= -2h(1-p) \cdot \\
& \int_{-c}^{c} tK(t) \int_{-c}^{t} K(u) \int_p^{p+hu} (1-x)^{-2} [1-G(Q(x))]^{-1} dxdudt \\
& -4h(1-p) \cdot \\
& \int_{-c}^{c} tK(t) \left[1-\tilde{K}(t)\right] \int_p^{p+ht} (1-x)^{-2} [1-G(Q(x))]^{-1} dxdt,
\end{aligned}
$$

et

$$
A_3 = 2h^2 \int_{-c}^{c} tK(t) \int_p^{p+ht} (1-x)^{-2} [1-G(Q(x))]^{-1} dx \left(\int_t^c sK(s)\, ds \right) dt.
$$

Finalement, avec la combinaison de ces résultats de E_1 et les résultats de E_2, on obtient la résultat.■

La comparaison des erreurs quadratiques moyennes de l'estimateur à $P.L$ des quantiles et de l'estimateur à noyau est donné par le théorème suivant :

Théorème 2.4 : [18]
Si :

$$
\int_{-c}^{c} tK(t)\, \tilde{K}(t)\, dt > 0.
$$

Alors, sous les conditions du théorème (2.3) pour n grand fixé et h tend vers zéro,

$$
\mathbf{E}\left\{[Q_n(p) - Q(p)]^2\right\} - \mathbf{E}\left\{\left[\widehat{Q}_n(p) - Q(p)\right]^2\right\} < 0.
$$

2.4 Selection de la fenêtre optimale

Dans cette partie nous donnons les valeurs des fenêtres optimales pour les calculs de l'estimateur $Q_n(p)$ au sens de minimisation de l'erreur quadratique moyenne.

Puisque cet estimateur est non paramétrique nous utilisons la méthode non paramétrique **bootstrap** (Efron (1981)).

On propose à estimer l'erreur quadratique moyenne de $Q_n(p)$ comme une fonction de h (pour p fixé) et on cherche les valeurs de h qui minimisent l'erreur quadratique moyenne (EqM) de $(Q_n(p))$.

On désigne par $(t_i, \delta_i)_{i=1,n}$ les échantillons censurés observés.

Notons que les échantillons bootstrap pour $(t_i, \delta_i)_{i=1,n}$ sont obtenus en choisissant aléatoirement de n paires (t_i^*, δ_i^*) avec remplaçant pour l'ensemble de n observations (t_i, δ_i).

On note par $\hat{Q}_n^{**}(p)$ l'estimateur à limite produit de $Q_n(p)$ basé sur l'échantillon bootstrap et par $Q_n^{**}(p)$ l'etimateur à noyau des quantiles pour les données bootstrap ; tel que :

$$Q_n^{**}(p) = h^{-1} \int_0^1 \hat{Q}_n^{**}(p) K(t - p/h) \, dt.$$

L'estimateur bootstrap de la variance de $Q_n(p)$ pour grand B (B étant le nombre d'échantillons bootstrap) est :

$$\widehat{V}(h) = (B-1)^{-1} \left\{ \sum_{i=1}^{B} \left[\hat{Q}_{ni}^{**}(p) \right]^2 - \left[\sum_{i=1}^{B} Q_{ni}^{**}(p) \right]^2 / B \right\}, \tag{2.3}$$

et l'estimateur bootstrap de biais de $Q_n(p)$ est :

$$\hat{B}_1(h) = B^{-1} \sum_{i=1}^{B} \hat{Q}_{ni}^{**}(p) - Q_n(p)$$

On désigne l'estimateur bootstrap de l'erreur quadratique moyenne (EqM) de $Q_n(p)$ par $EqM^*_{Q_n(p)}$.

L'estimateur bootstrap de biais est modifié avec utilisation de l'estimateur à limite produit des quantiles qui ne dépend pas de h :

$$\hat{B}_2(h) = B^{-1} \sum_{i=1}^{B} \hat{Q}_{ni}^{**}(p) - \hat{Q}_n(p),$$

où $\hat{Q}_n$ est lestimateur à P.L des quantiles.

Donc :

$$EqM^{*}_{Q_n(p)}(h) = \widehat{\mathbf{V}}(h) + \hat{B}_2^2(h). \tag{2.4}$$

En minimisant (2.4), on obtient la valeur de h cherchée.

Chapitre 3

Lois limites de l'estimateur à noyau des quantiles(ENQ)

3.1 Résultats asymptotiques

On étudie les propriétés asymptotiques de $Q_n(p)$ qui sont données par les théorèmes suivants :

Pour tout distribution H, on note

$$T_H := \sup\{t : \ H(t) < 1\}.$$

Théorème 3.1 : [22]

Supposons que les conditions 1, 2-6, et 7,8 et 10 (voir, Page 32) sont vérifiées, alors :

i]. Si $T_F < T_G \leq \infty$ et $h_n^{-1}(\log\log n/n)^{\frac{3}{4}} \to 0$ quand $n \to \infty$, alors :

$$Q_n(p) \to Q(p) \text{ quand } n \to \infty \text{ avec probabilité } 1,\ 0 \leq p \leq 1.$$

ii]. Si $T_F \leq T_G \leq \infty$. Supposons que G est continue dans un intervalle petit et ouvert (T, T_F).

Soit Φ une fonction définie sur un intervalle $[0, 1 - F(T)]$ tel que:

$$\Phi(x) \geq x, \;\; \Phi(x) \to 0^+ \text{ quand } x \to 0^+,$$

et

$$1 - F(t) \leq \Phi(1 - F(t)) \;\; \text{pour } t \in (T, T_F)\,.$$

Si, $\forall c > 1$,

$$\limsup_n \Phi\left(c\,(\log\log n/(2n))^{1/2}\right) h_n^{-1} < \infty,$$

alors :

$$Q_n(p) \to Q(p) \;\; \text{quand } n \to \infty \text{ avec probabilité 1, } 0 \leq p \leq 1$$

Le théorème suivant indique que les estimateurs Q_n et Q_n^* sont asymptotiquement équivalents presque sûrement (uniformément en $\mathbf{P}$).

On définit :

$$\hat{\mu}(t) = \int_0^t \widehat{S}_n(x)\,dx, \quad t \geq 0,$$

où $\widehat{S}_n$ est l'estimateur de Kaplan-Meier de la fonction de survie (voir, Page (25)).

Théorème 3.2 : [22]

Supposons que les conditions 1, 2, 3, 6 et 10 sont vérifiées.

Sous les conditions du théorème (3.1)(ii), si $\limsup_n \hat{\mu}(T_n) < \infty$ avec probabilité 1, alors :

$$\mathbf{P}\left[\limsup_n |Q_n^*(p) - Q_n(p)| \leq M^* \limsup_n \Phi\left(c\,(\log\log n/2n)^{1/2}\right) h_n^{-2}\right] = 1,$$

où $M^* < \infty$ et $c > 1$ est une constante.

Donc, si $\Phi\left(c\left(\log\log n/2n\right)^{1/2}\right)h_n^{-2} \to 0$, alors sous les conditions de théorème (3.2), Q_n^* et Q_n sont asymptotiquement équivalentes avec probabilité 1 (uniformément en $\mathbf{P}$).

Théorème 3.3 : [17]

Supposons que F et G sont continues et 1, 2, et 6 sont vérifiées et supposons qu'il existe $q > 1$ tel que $E\left(X_1^q\right) < \infty$, où $T_1 = \min\left\{X_1, C_1\right\}$.

Soit η tel que :

$$[1 - F(\eta)][1 - G(\eta)] > 0,$$

et soit $T^* = F(\eta)$, alors pour tout $T \in [0, T^*)$, et $n^{1/2}h_n^4 \to \infty$, quand $n \to \infty$:

$$\lim_{n\to\infty} \mathbf{E}\left[\sup_{0\le p\le T}\left|Q_n^*(p) - Q_n(p)\right|^2\right] = 0,$$

pourvu que $n^{1/2}h_n^4 \to \infty$ quand $n \to \infty$.

3.2 Normalité asymptotique de ENQ

On démontre la normalité asymptotique de $Q_n(p)$ avec le prochain théorème, et de là on utilise les deux lemmes suivants pour la démonstration.

On désigne par $\{K(s,t) : 0 \le s \le t,\ t \ge 0\}$ le processus généralisé de Kiefer (voir, Annexe).

Lemme 3.1 : Quand $n \to \infty$, pour $0 < p < T$, où $T < 1$ et $\delta < \min\{T - p, p\}$, on a:

$$\sup_{|h|<\delta}\left|n^{-1/2}\left[K(p+h,n) - K(p,n)\right]\right| \to 0, \text{ en probabilité.}$$

Lemme 3.2 : Supposons que la dérivée f' est continue en ξ_p et $f\left(\xi_p\right) > 0$.

Sous les conditions 1, 4, 5 et 9, pour $0 < p < 1$, on a :

$$\left|\int_0^1\left[q_n(t) - q_n(p)\right]h_n^{-1}K\left(t - p/h_n\right)dt\right| \to 0 \text{ en probabilité, quand } n \to \infty,$$

où

$$q_n(t) = n^{1/2}\left[\widehat{Q}_n(t) - Q(p)\right],$$

est le processus à limite produit des quantiles.

Preuve de lemme 3.2 : on a : $\forall\delta > 0$, $\exists N$ tel que $\forall n \geq N$,

$$\begin{aligned} & \left|\int_0^1 [q_n(t) - q_n(p)]\, h_n^{-1} K(t - p/h_n)\, dt\right| \\ = & \left|\int_{A(\delta)} [q_n(t) - q_n(p)] \cdot h_n^{-1} K(t - p/h_n)\, dt\right| \\ \leq & \sup_{t\in A(\delta)} |q_n(t) - q_n(p)|, \end{aligned} \tag{3.1}$$

où $A(\delta) = [p-\delta, p+\delta]$.

On a $f(Q(t)) > 0$ pour chaque $t \in A(\delta)$.

Pour δ suffisament petit, on a :

$$\begin{aligned} \sup_{t\in A(\delta)} |q_n(t) - q_n(p)| \leq & \sup_{t\in A(\delta)} \left|\widetilde{Q}_n(t) - \widetilde{Q}_n(p)\right| \cdot \left|\frac{1}{f(Q(t))}\right| \\ & + \sup_{t\in A(\delta)} \left|\widetilde{Q}_n(p)\left[\frac{1}{f(Q(t))} - \frac{1}{f(Q(p))}\right]\right|, \end{aligned}$$

où

$$\widetilde{Q}_n(t) = f(Q(t))\, q_n(t).$$

Soit :

$$a = \sup_{t\in A(\delta)} \left|\frac{1}{f(Q(t))}\right|, \qquad b = \sup_{t\in A(\delta)} \left|\frac{1}{f(Q(t))} - \frac{1}{f(Q(p))}\right|.$$

D'après le corollaire 1 de Cheng (1981), puisque f est continue en ξ_p,

$$\widetilde{Q}_n(p) \to Z \text{ en distribution, quand } n \to \infty,$$

où Z est la variable aléatoire de distribution normale, de moyenne zéro et de variance

$$\sigma^2 = (1-p)^2 \int_0^{\xi_p} [1 - H(u)]^{-2} \, dF^*(u),$$

avec

$$1 - H(u) = [1 - F(u)][1 - G(u)]$$

et

$$F^*(u) = \mathbf{P}(T_i \leq u, \Delta_i = 1).$$

Donc,

$$\sup_{t \in A(\delta)} \left| \widetilde{Q}_n(p) \left[\frac{1}{f(Q(t))} - \frac{1}{f(Q(p))} \right] \right| \leq b \left| \widetilde{Q}_n(p) \right|$$

et pour $\varepsilon > 0$,

$$\lim_{n \to \infty} \sup \mathbf{P} \left(\left| \widetilde{Q}_n(p) \right| > \varepsilon / b \right) \leq \mathbf{P}(|Z| > \varepsilon / b) \leq b\sigma^2 / \varepsilon^2,$$

on a :

$$\begin{aligned} \sup_{t \in A(\delta)} \left| \widetilde{Q}_n(t) - \widetilde{Q}_n(p) \right| \cdot \left| \frac{1}{f(Q(t))} \right| &\leq a \sup_{t \in A(\delta)} \left| \widetilde{Q}_n(t) - \widetilde{Q}_n(p) \right| \\ &\leq a \sup_{t \in A(\delta)} \left\{ \left| \widetilde{Q}_n(t) - n^{-\frac{1}{2}} K(t,n) \right| \right. \\ &\quad + \left| \widetilde{Q}_n(p) - n^{-\frac{1}{2}} K(p,n) \right| \\ &\quad \left. + n^{-1/2} \cdot |[K(p,n) - K(t,n)]| \right\} \end{aligned} \tag{3.2}$$

pour δ assez petit, $p + \delta < T < 1$ et $p - \delta \geq 0$.

Alors, via le corollaire (8.3.3) de Csörgő(1983), ona :

$$\sup_{t \in A(\delta)} \left| \widetilde{Q}_n(t) - n^{-\frac{1}{2}} K(t,n) \right| \to 0 \text{ en probabilité, quand } n \to \infty,$$

et

$$\sup_{t \in A(\delta)} \left| \widetilde{Q}_n(p) - n^{-\frac{1}{2}} K(p,n) \right| \to 0 \text{ en probabilité}$$

En utilisant le lemme 1,

$$\sup_{t\in A(\delta)}\left|n^{-\frac{1}{2}}\left[K\left(p,n\right)-K\left(t,n\right)\right]\right|\rightarrow 0 \text{ en probabilité.}$$

Alors (3.2) converge vers zéro en probabilité quand $n\rightarrow\infty$.

Finalement, puisque b dépend de δ, on peut choisir b de telle sorte que :

$$\sup_{t\in A(\delta)}\left|\widetilde{Q}_n\left(p\right)\left[\frac{1}{f\left(Q\left(t\right)\right)}-\frac{1}{f\left(Q\left(p\right)\right)}\right]\right|\rightarrow 0 \quad \text{en probabilité.}\blacksquare$$

Théorème 3.4 : (**Lio, Padjett et Yu** (1986))

Supposons que les conditions 1, 4, 5, et 7 sont vérifiées, $G\left(T_F\right)\leq 1$, et les f' sont continues en ξ_p et $f\left(\xi_p\right)>0$.

Soit $\{h_n\}$ une suite de fenêtres qui sont vérifiées $n^{1/4}h_n\rightarrow 0$ quand $n\rightarrow\infty$, alors pour $0<p<T$, $(T<1)$,

$$\sqrt{n}\left[Q_n\left(p\right)-Q\left(p\right)\right]\rightarrow Z \text{ en distribution, quand } n\rightarrow\infty,$$

où Z est la variable aléatoire suit la loi normale de moyenne zéro et de variance

$$\sigma_p^2=\left(1-p\right)^2\int_0^{\xi_p}\left[1-H\left(u\right)\right]^{-2}\frac{dF^*\left(u\right)}{f^2\left(\xi_p\right)},$$

où :

$$1-H\left(u\right)=\left[1-F\left(u\right)\right]\left[1-G\left(u\right)\right],$$

et

$$F^*\left(u\right)=\mathbf{P}\left(T_i\leq u,\Delta_i=1\right)$$

est la distribution des observations non censurées.

Un exemple des suites $\{h_n\}$ est $h_n=Dn^{-b}$ avec $b>1/4$.

Preuve : On a :

$$\begin{aligned}\sqrt{n}\left[Q_n\left(p\right)-Q\left(p\right)\right] &= \int_0^1\left[q_n\left(t\right)-q_n\left(p\right)\right]h_n^{-1}K\left(t-p/h_n\right)dt\\ &\quad +n^{\frac{1}{2}}\int_0^1\left[Q\left(t\right)h_n^{-1}K\left(t-p/h_n\right)dt-Q\left(p\right)\right]\\ &\quad +q_n\left(p\right), \end{aligned}\qquad (3.3)$$

où

$$q_n(t) = n^{\frac{1}{2}} \left[\widehat{Q}_n(t) - Q(p) \right].$$

D'après le lemme (3.2), le premier terme de côté de droit de (3.3) est aussi $o_p(1)$.

De même, via l'equation (10) de Yang (1985), on obtient :

$$\begin{aligned} & n^{\frac{1}{2}} \left[\int_0^1 Q(t)\, h_n^{-1} K(t - p/h_n)\, dt - Q(p) \right] \\ = \quad & o\left(n^{\frac{1}{2}} h_n^2 \right) + n^{\frac{1}{2}} h_n^2 Q''(p) \int_{-\infty}^{+\infty} \frac{t^2}{2} K(t)\, dt \end{aligned} \tag{3.4}$$

Pour lasuite de la démonstration, il suffit d'appliquer le corollaire 1 de Cheng (1981).■

Chapitre 4

Lois du logarithme pour des estimateurs de la densité et de la densité des quantiles

On désigne par $f(t)$ et $q(t)$ les densités de $F(x)$ et $Q(t)$ respectivement, où $F(x)$ est la distribution d'une suite de variables aléatoires i.i.d, non négatives et $Q(t)$ leur fonction des quantiles, ce modèle des variables est supposés censurées à droite.

La loi de logarithme des estimatyeurs de densité et de densité des quantiles à été obtenue par Xiang (1992). Ces résultats sont appliquées pour obtenir la fenêtre optimale avec la considération de la convergence presque sûre uniforme.

4.1 Lois de logarithme des estimateurs à noyau de la densité des quantiles

Soit $X_1, ..., X_n$ une suite de variables aléatoires independantes et identiquement distribuées de distribution commune $F(x)$.

Soit $f(x)$ la densité de X_1.

L'estimateur à noyau de $f(x)$ est :

$$f_n(t) = \frac{1}{h_n} \int \left(K(\frac{t-x}{h_n}) \right) dF_n(x), \tag{4.1}$$

où F_n est la distribution empirique de l'échantillon $X_1, ..., X_n$, $\{h_n\}$ est une suite de fenêtres avec $h_n \downarrow 0$ et $K(x)$ un noyau convenable.

Soit :

$$\bar{f}_n(t) = \mathbf{E} f_n(t) = \frac{1}{h_n} \int K\left(\frac{t-x}{h_n}\right) dF(x), \tag{4.2}$$

et soit $I_\varepsilon = (a+\varepsilon, b-\varepsilon)$, $\varepsilon > 0$. On suppose que K à variation borné avec $K(x) = 0$ en dehors de quelque intervalle finit$[r, s)$.

Stute (1982b) a été prouver que pour chaque $\varepsilon > 0$ et $I = (a, b)$ avec $a < b$, si $f(x)$ est uniformément continue sur I avec $0 < \delta \leq f(x) \leq M < \infty$, alors :

$$\forall x \in I, \quad \lim_{n\to\infty} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t \in I_\varepsilon} \frac{\left|f_n(t) - \bar{f}_n(t)\right|}{\sqrt{f(t)}} = \left(2\int_r^s K^2(x)dx\right)^{1/2}.$$

Il a eté trouver la meilleur vitesse de convergence de $f_n(t)$ à $\bar{f}_n(t)$ sur I_ε qui est appliqué pour obtenir les fenêtres optimales avec considération à la convergence presque sûre uniforme de $f_n(t)$ à $f(t)$.

Donc, si $f^{(2)}(t)$ est continue sur I, la fenêtre optimal est :

$$h_n \sim \left(\frac{\int K^2(u)\, du}{10 \sup_{t\in I_\varepsilon}\left[\left|f^{(2)}(t)\right|^2 / f(t)\right]\left(\int K(u)\, u^2 du\right)^2} \frac{\log n}{n}\right)^{1/5},$$

et avec cette fenêtre optimale

$$\frac{f_n(t) - f(t)}{\sqrt{f(t)}} = O\left(\left(\frac{\log n}{n}\right)^{2/5}\right)$$

uniformément dans I_ε.

Soit :

$$Q(t) = \inf\{x : F(x) \geq t\}, \quad 0 < t < 1,$$

la fonction des quantiles associé à $F(x)$, et $q(t)=Q^{'}(t)$ la densité des quantiles.

Parzen (1979) a intoduit un estimateur à noyau de densité des quantiles qui défini par :

$$\widehat{q}_n^*(t) = -\frac{1}{h_n^2}\int_0^1 F_n^{-1}(x)\, K^{'}\left(\frac{x-t}{h_n}\right) dx, \tag{4.3}$$

où

$$F_n^{-1}(x) = \inf\left\{u : F_n(u) \geq x\right\}.$$

Yang (1985) a introduit un nouveau estimateur à noyau des quantiles qui est défini par :

$$\widetilde{Q}_n(t) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{i/n-t}{h_n}\right) X_{(i)},$$

où $X_{(1)}, ..., X_{(n)}$ sont les statistiques d'ordre de l'echantillon $X_1, ..., X_n$.

L'estimateur à noyau de densité des quantiles est défini par :

$$\widehat{q}_n(t) = -\frac{1}{nh_n^2}\sum_{i=1}^{n} K^{'}\left(\frac{i/n-t}{h_n}\right) X_{(i)}. \tag{4.4}$$

Soit, $\widehat{F}_n(t)$ et s_i, $i \geq 1$, l'estimateur de Kaplan Meier et le saux de $\widehat{F}_n(t)$ en $T_{(i)}$, (respectivement), qui sont définis précédemment (cf. page 26, 31).

Soit :

$$\widehat{F}_n^{-1}(x) = \inf\left\{u : \widehat{F}_n(u) \geq x\right\}.$$

Dans le cas de modèle censuré à droite, et correspondant à les estimateurs à noyau de densité des quantiles qui sont définit par (4.3) et (4.4), Xiang a définit les estimateurs à noyau de densité des quantiles comme :

$$\widehat{q}_n^*(t) = -\frac{1}{h_n^2}\int_0^1 \widehat{F}_n^{-1}(x)\, K^{'}\left(\frac{x-t}{h_n}\right) dx,$$

et

$$\widehat{q}_n(t) = -\frac{1}{h_n^2}\sum_{i=1}^{n} T_{(i)} s_i K' \left(\frac{\widehat{F}_n(T_{(i)}) - t}{h_n}\right) X_{(i)}.$$

On suppose que le noyau $K(x)$ est symétrique et, pour l'entier positif l,

$$K(x) \in C^l(-\infty,\infty), \qquad K(x) \text{ à support compact } [-1,1],$$

où

$$C^l(-\infty,\infty) = \left\{f;\ f^{(l)} \text{ est continue sur } (-\infty,\infty)\right\}, \tag{4.5}$$

et, pour quelques entier $m \geq 2$,

$$\begin{aligned}\int_{-1}^{1} K(x)dx &= 1,\\ \int_{-1}^{1} x^j K(x)dx &= 0, \qquad j = 1,...,m-1,\\ \int_{-1}^{1} x^m K(x)dx &= \alpha_m\end{aligned} \tag{4.6}$$

On note que (4.5) implique que $K^{(i)}(x)$, $i = 1,...,n$ à support compact $[-1,1]$.

Pour la suite des fenêtres $\{h_n\}$, on suppose que les conditions suivantes sont vérifiées :

$$\text{(i) } nh_n \uparrow \infty; \qquad \text{(ii) } \frac{\log h_n^{-1}}{nh_n} \qquad \text{(iii) } \frac{\log h_n^{-1}}{\log\log n} \to \infty. \tag{4.7}$$

4.1.1 Résultats intermédiaires

Soit $B(t)$ le pont brownien (voir, Annexe), et $A(x)$ la fonction définie sur l'intervalle $I \subset [0,1]$ avec $0 \leq A(x) \leq 1$ et à dérivée $a(x)$ uniformement continue, $a(x) > \delta > 0$ sur I.

On note (d'après [Shorack et Wellner [(1986), page (532)]] que :

$$\lim_{n\to\infty} \sup_{\substack{\underline{c}h_n \leq t-u \leq \overline{c}h_n \\ t,u\in I}} \frac{|B(t) - B(u)|}{\sqrt{2(t-u)\log h_n^{-1}}} = 1 \quad p.s. \tag{4.8}$$

où $0 < \underline{c} \leq \overline{c} < \infty$ des nombres fixés et,

$$\lim_{n\to\infty} \sup_{\substack{\underline{c}h_n \leq t-u \leq \overline{c}h_n \\ t,u\in I}} \frac{|B(A(t)) - B(A(u))|}{\sqrt{2(t-u)\, a(x_{u,t}) \log h_n^{-1}}} = 1 \quad p.s., \tag{4.9}$$

où $x_{u,t}$ est un point entre u et t.

Soit :

$$L_n(t) = \frac{1}{h_n} \int K\left(\frac{t-x}{h_n}\right) dB(A(x)), \tag{4.10}$$

et soit :

$$J_\varepsilon = [c-d, d+\varepsilon] \subset (0,1) \text{ pour } \varepsilon > 0,$$

et

$$J = [c,d] \text{ avec } c < d.$$

Lemme 4.1 : Supposons que $a(x) = A'(x)$ est continue sur J_ε avec $0<\delta \leq a(x) \leq M < \infty$ pour tout $x \in J_\varepsilon$.

Soit $K(x)$ un noyau à variation borné avec $K(x) = 0$ au dehor de $[-1,1]$, alors avec probabilité 1 :

$$\lim_{n\to\infty} \sqrt{\frac{h_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{|L_n(t)|}{\sqrt{a(t)}} = \left(2\int_{-1}^{1} K^2(x)\, dx\right)^{1/2}.$$

Par la suite, notons que $F, G,$ et H sont les distribution des X_i, C_i et T_i (respectivement).

On suppose que $J_\varepsilon \subset (0, F(T_H))$, et Soit :

$$\overline{q}_n(t) = -\frac{1}{h_n^2} \int_0^1 Q(x) K'\left(\frac{x-t}{h_n}\right) dx.$$

Notre résultats sont les suivante.

Théorème 4.1 : [33]

On suppose que $q(t)$ est continue sur J_ε avec $0 < \delta \leq q(t) \leq M < \infty$ pour tout $t \in J_\varepsilon$.

Soit (4.5) est vérifié pour $l = 1$.

Soit $K'(x)$ est lipshitzienne d'ordre 1 et $G(x)$ est lipshitzienne d'ordre 1/2 dans $[Q(c-\varepsilon), Q(d+\varepsilon)]$, alors si :

$$\frac{\log_2 n}{(nh_n^3 \log h_n^{-1})^{1/2}} \to 0,$$

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\,|\widehat{q}_n^*(t) - \overline{q}_n(t)|}{q(t)} = \left(2\int_{-1}^{1} K^2(x)dx\right)^{1/2}, \tag{4.11}$$

et

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\,|\widehat{q}_n(t) - \overline{q}_n(t)|}{q(t)} = \left(2\int_{-1}^{1} K^2(x)dx\right)^{1/2}. \tag{4.12}$$

Soit :

$$H^u(t) = \mathbf{P}(T_1 \leq t, \Delta_1 = 1) \quad \text{et} \quad H^c(t) = \mathbf{P}(T_1 \leq t, \Delta_1 = 0).$$

Major et Rejtö (1988) ont démontrés que, pour $t < T_H$,

$$\widehat{F}_n(t) - F(t) = \frac{1}{n}\sum_{i=1}^{n} \psi_i(t) + \tau_n(t), \tag{4.13}$$

et

$$\widehat{F}_n(t) - F(t) = \frac{1}{\sqrt{n}} W(t) + \gamma_n(t), \tag{4.14}$$

où

$$\begin{aligned}\psi_i(t) &= (1-F(t))\left\{\int_{-\infty}^{t} \frac{I_{(T_i\leq y,\Delta_i=1)} - H(y)}{(1-H(y))^2} dH^u(y) + \frac{I_{(T_i\leq t,\Delta_i=1)} - H^u(t)}{1-H(t)}\right. \\ &\qquad \left. - \int_{-\infty}^{t} \frac{I_{(T_i\leq y,\Delta_i=1)} - H^u(y)}{(1-H(y))^2} dH(y)\right\},\end{aligned}$$

et $W(t)$ est le processus de Gauss qui est défini par :

$$\begin{aligned} W(t) &= (1-F(t))\left\{\int_{-\infty}^{t} \frac{B(H^u(y)) - B(1-H^c(y))}{(1-H(y))^2} dH^u(y) \right. \\ &\left. + \frac{B(H^u(t))}{1-H(t)} - \int_{-\infty}^{t} \frac{B(H^u(y))}{(1-H(y))^2} dH(y) \right\}. \end{aligned}$$

Le reste des termes (4.13) et (4.14) est satisfait avec probabilité 1,

$$\sup_{t \leq T} |\tau_n(t)| = O\left(\frac{\log n}{n}\right), \quad T < T_H,$$

et, $\forall x > 0$,

$$\mathbf{P}\left(\sup_{t \leq T} |n\gamma_n(t)| > \frac{2c}{\Delta} \log n + x\right) < 2ke^{-\lambda \Delta^2 x},$$

où $0 < \Delta < 1 - H(T)$ et c, k et λ des constants positives.

Remarque 4.1 : Si $Q(t)$ est deux fois différentiable sur J_ε.

Via les résultats de Major et Rejtö (1988) et Lo et Singh (1986) on obtient :

$$\widehat{F}_n^{-1}(t) - Q(t) = -\frac{1}{\sqrt{n}} W(Q(t)) + \beta_n(t),$$

où $W(t)$ est défini par (4.14) et

$$\sup_{t \in J} |\beta_n(t)| = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right).$$

Théorème 4.2 : [33]

Supposons que $Q(t)$ est deux fois différentiable sur J_ε avec $0 < \delta \leq q(t) \leq M < \infty$, pour tout $t \in J_\varepsilon$.

Soit (4.5) est vérifiée pour $l = 1$, et soit $G(x)$est lipchitzienne d'ordre 1/2 sur $[Q(c-\varepsilon), Q(d+\varepsilon)]$, alors si :

$$\frac{(\log n)^3}{nh_n^2 (\log h_n^{-1})^2} \to 0, \quad \text{quand } n \to \infty,$$

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\,|\widehat{q}_n^*(t) - \overline{q}_n(t)|}{q(t)} = \left(2\int_{-1}^{1} K^2(x)\,dx\right)^{1/2},$$

et

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\,|\widehat{q}_n(t) - \overline{q}_n(t)|}{q(t)} = \left(2\int_{-1}^{1} K^2(x)\,dx\right)^{1/2}.$$

On applique le théorème (4.1) pour obtenir les fenêtres optimale, pour cette raison, on suppose que $q^{(m)}(t)$ est continue sur J_ε, avec $m \geq 2$, et que $K(x)$ est satisfait (4.6). De théorème 4.1 et

$$\overline{q}_n(t) - q(t) = \frac{h_n^m}{m!} q^{(m)}(t)\,\alpha_m + o(h_n^m),$$

la fenêtre optimale est obtenu avec minimisation le terme :

$$\frac{h_n^m}{m!} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\,\left|q^{(m)}(t)\right|}{q(t)} \int_{-1}^{1} |K(u)\,u^m|\,du$$
$$+ \left(\frac{2\log h_n^{-1}}{nh_n} \int_{-1}^{1} K^2(u)\,du\right)^{1/2}.$$

Pour $m = 2$, la fenêtre optimal asymptotique est :

$$h_n \sim \left(\frac{\displaystyle\int_{-1}^{1} K^2(u)\,du}{10 \sup_{t\in J} \left[|q^{(2)}(t)|^2 (1 - G(Q(t)))/q^2(t) \right] \left(\displaystyle\int_{-1}^{1} K(u)\,u^2 du \right)^2} \frac{\log n}{n} \right)^{1/5}$$

et, avec probabilité 1,

$$\frac{\widehat{q}_n^*(t) - q(t)}{q(t)} = O\left(\left(\frac{\log n}{n}\right)^{2/5}\right)$$

et

$$\frac{\widehat{q}_n(t) - q(t)}{q(t)} = O\left(\left(\frac{\log n}{n}\right)^{2/5}\right),$$

uniformément sur J.

4.1.2 Estimation des dérivées superieur de la densité des quantiles

Supposons que (4.5) est vérifiée pour $l = r > 1$ et $K^{(r)}$ est lipshitzienne d'ordre 1.

On définit les estimateurs

$$\widehat{q}_n^{*(r)}(t) = \frac{(-1)^r}{h_n^{r+1}} \int_0^1 \widehat{F}_n^{-1}(t) K^{(r)}\left(\frac{x-t}{h_n}\right) dx,$$

et

$$\widehat{q}_n^{(r)}(t) = \frac{(-1)^r}{h_n^{r+1}} \sum_{i=1}^n T_{(i)} s_i K^{(r)}\left(\frac{x-t}{h_n}\right) dx.$$

Soit :

$$\overline{q}_n^{(r)}(t) = \frac{(-1)^r}{h_n^{r+1}} \int_0^1 Q(t) K^{(r)}\left(\frac{x-t}{h_n}\right) dx.$$

Si $q(t) = Q'(t)$ et $G(x)$ est satisfait les suppositions de théorème 4.1, on obtient :

$$\lim_{n\to\infty} h_n^{r-1} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\left|\widehat{q}_n^{*(r)}(t) - \overline{q}_n^{(r)}(t)\right|}{q(t)}$$
$$= \left(2\int_{-1}^1 \left[K^{(r-1)}(x)\right]^2 dx\right)^{1/2},$$

et

$$\lim_{n\to\infty} h_n^{r-1} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(Q(t))}\left|\widehat{q}_n^{(r)}(t) - \overline{q}_n^{(r)}(t)\right|}{q(t)}$$
$$= \left(2\int_{-1}^1 \left[K^{(r-1)}(x)\right]^2 dx\right)^{1/2}.$$

De plus, si, pour $m \geq 2, Q^{(r+m)}(t)$ est continue sur J_ε pour $J_\varepsilon \subset (0, F(T_H))$, la fenêtre optimal est d'ordre $((\log n)/n)^{1/[2(r+m)+1]}$et, avec probabilité 1,

$$\frac{\widehat{q}_n^{*(r)}(t) - q^{(r)}(t)}{q(t)} = O\left(\left(\frac{\log n}{n}\right)^{m/[2(r+m)+1]}\right)$$

et

$$\frac{\widehat{q}_n^{(r)}(t) - q^{(r)}(t)}{q(t)} = O\left(\left(\frac{\log n}{n}\right)^{m/[2(r+m)+1]}\right)$$

uniformément sur J.

4.2 Lois du logarithme pour l'estimateur de la densité et le taux de hazard pour les données censurées

Soit $f(t) = F'(t)$ la densité de X_1.

La fonction de taux de hazard est définie par :

$$h(t) = \frac{f(t)}{1 - F(t)}, \quad \text{pour } \; F(t) < 1. \tag{4.15}$$

L'estimateur à noyau de $f(t)$ est défini par :

$$f_n(t) = \frac{1}{h_n} \int K\left(\frac{t - x}{h_n}\right) d\widehat{F}_n(x), \tag{4.16}$$

où $K(t)$ est un noyau choisi convenablement, $\{h_n\}$ est une suite de fenêtres avec $h_n \downarrow 0$, et $\widehat{F}_n$ est l'estimateur de Kaplan-Meier de F.

De (4.15) et (4.16), l'estimateur de $h(t)$ est :

$$h_n(t) = \frac{f_n(t)}{1 - \widehat{F}_n(t)}, \qquad \text{pour } \; t < T_{(n)}.$$

Soit :

$$\overline{f}_n(t) = \frac{1}{h_n} \int K\left(\frac{t - x}{h_n}\right) dF(x).$$

Sous certain conditions, Diehl et Stute ont démontrés que :

$$\lim_{n\to\infty}\sqrt{\frac{nh_n}{\log h_n^{-1}}}\sup_{t\in J}\sqrt{\frac{1-G(t)}{f(t)}}\left|f_n(t)-\overline{f}_n(t)\right| = \left(2\int K^2(x)dx\right)^{1/2} \qquad p.s.,$$

où $J = [c,d]$.

De plus, si $f^{(r)}(t)$ est continue sur $J_\varepsilon = [c-\varepsilon, d+\varepsilon]$, on a:

$$\forall\varepsilon > 0,\ \sup_{t\in J}|f_n(t) - f(t)| = O\left(\left(\frac{\log n}{n}\right)^{r/(2r+1)}\right) \qquad p.s.,$$

et avec $h_n = (\log n/n)^{1/(2r+1).}$

4.2.1 Résultats intermédiaires

Théorème 4.3 : [34]

Supposons que $f(t)$ est continue sur $J_\varepsilon \subset (0, T_H)$ avec $0 < \delta \leq f(t) \leq M < \infty$ pour chaque $t \in J_\varepsilon$ et (4.5) est vérifiée pour $l = 1$, alors si $nh_n = n^\alpha$, $\alpha > 0$,

$$\lim_{n\to\infty}\sqrt{\frac{nh_n}{\log h_n^{-1}}}\sup_{t\in J}\frac{\sqrt{1-G(t)}\left|f_n(t)-\overline{f}_n(t)\right|}{\sqrt{f(t)}} = \left(2\int_{-1}^{1} K^2(x)dx\right)^{1/2}.$$

Remarque 4.2 : Supposons que $f^{(r)}(t)$ est continue sur J_ε pour $r \geq 2$ et (4.6) est vérifiée, alors de théorème 4.3 et

$$\overline{f}_n(t) - f_n(t) = \frac{h_n^r}{r!}\alpha_r f^{(r)}(t) + o(h_n^r).$$

la fenêtre optimal est obtenu avec minimisation du terme

$$\frac{h_n^r}{r!}\sup_{t\in J}\frac{\sqrt{1-G(t)}\left|f^{(r)}(t)\right|}{\sqrt{f(t)}}\int_{-1}^{1}|K(u)u^r|\,du + \left(\frac{2\log h_n^{-1}}{nh_n}\int_{-1}^{1}K^2(u)du\right)^{1/2}.$$

La fenêtre optimale est d'ordre $O\left((\log n/n)^{1/(2r+1)}\right)$.

Soit :

$$\overline{h}_n(t) = \frac{\overline{f}_n(t)}{1 - F(t)}, \quad \text{pour } F(t) < 1.$$

Le théorème suivante concernant la loi du logarithme de l'estimateur $h_n(t)$.

Théorème 4.4 : [34]

Sous les conditions de théorème (4.3),

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{\log h_n^{-1}}} \sup_{t\in J} \sqrt{\frac{1 - H(t)}{h(t)}} \left|h_n(t) - \overline{h}_n(t)\right| = \left(2\int_{-1}^{1} K^2(u)du\right)^{1/2}.$$

Remarque 4.3 : De théorème (4.4), et sous les suppositions de la remarque (4.2), pour $h_n = ((\log n/n)^{1/(2r+1)})$,

$$\sup_{t\in J} |h_n(t) - h(t)| = O\left(\left(\frac{\log n}{n}\right)^{1/(2r+1)}\right) \quad p.s. \tag{4.17}$$

La vitesse de convergence donné par (4.15) est optimal.

Supposons que (4.5) est vérifiée pour $l = m + 1$, $m \geq 1$.

On considère l'estimateur à noyau :

$$f_n^{(m)}(t) = \frac{1}{h_n^{m+1}} \int K^{(m)}\left(\frac{t - x}{h_n}\right) d\widehat{F}_n(x).$$

On définit :

$$\overline{f}_n^{(m)}(t) = \frac{1}{h_n^{m+1}} \int K^{(m)}\left(\frac{t - x}{h_n}\right) dF_n(x).$$

Théorème 4.5 : [34]

Supposons que (4.5) est vérifiée pour $l = m + 1$ pour $m \geq 1$ et f est continue sur $J\varepsilon$ avec $0 < \delta \leq f(t) \leq M < \infty$ pour tout $t \in J_\varepsilon$, alors si $nh_n = n^\alpha$ pour $\alpha > 0$,

$$\lim_{n\to\infty} \sqrt{\frac{nh_n^{2m+1}}{\log h_n^{-1}}} \sup_{t\in J} \frac{\sqrt{1-G(t)}\left|f_n^{(m)}(t) - \overline{f}_n^{(m)}(t)\right|}{\sqrt{f(t)}} = \left(2\int_{-1}^{1}\left[K^{(m)}(x)\right]^2 dx\right)^{1/2}.$$

En plus si pour $r \geq 2$, $f^{(m+r)}(t)$ est continue sur J_ε, et supposons que (4.6) est vérifiée, alors avec $h_n = ((\log n)/n)^{1/2(r+m)+1}$,

$$\sup_{t\in J}\left|f_n^{(m)}(t) - f^{(m)}(t)\right| = O\left(\left(\frac{\log n}{n}\right)^{r/(2(r+m)+1)}\right) \qquad p.s.$$

Chapitre 5

Lois limite fonctionnelles pour les accroissements des processus à limite produit de Kaplan-Meier.

Soit $X_1, ..., X_n$ une suite de temps de survie de n individus qui sont supposées censurées à droite par la suite $C_1, ..., C_n$ qui sont des constantes ou des variables aléatoires $i.i.d$ définit sur le même espace de probabilité $(\Omega, \mathcal{A}, P)$.

Soit :

$$X = X_1, C = C_1, T = T_1, \Delta = \Delta_1, F(x) = \mathbf{P}(X \leq x), G(x) = \mathbf{P}(C \leq X),$$
$$H(x) = \mathbf{P}(T \leq X) = 1 - (1 - F(x))(1 - G(x)).$$

On désigne par $f(x) = (d/dx)F(x)$ la densité de survie.

Pour chaque entier $n \geq 1$, soit :

$$F_n(x) \quad = \quad 1 - \prod_{\substack{i:\ T_{i,n} \leq x \\ 1 \leq i \leq n}} \left\{ 1 - \frac{\Delta_{i,n}}{n - i + 1} \right\}, \tag{5.1}$$

et

$$G_n(x) \quad = \quad 1 - \prod_{\substack{i:\ T_{i,n} \leq x \\ 1 \leq i \leq n}} \left\{ 1 - \frac{1 - \Delta_{i,n}}{n - i + 1} \right\}, \tag{5.2}$$

les $P.L$ estimateurs de F et G respectivement, où $T_{1,n} \leq ... \leq T_{n,n}$ sont les statistiques d'ordre associées avec $T_1, ..., T_n$, et pour $1 \leq j \leq n$ les $\Delta_{i,n}$ sont les Δ_j correspondantes à $T_{i,n} = T_j$ pour chaque $i = 1, ..., n$, $(\Delta_{i,n})$ représente la valeur des Δ_j correspondant à $T_{i,n} = T_j$, $1 \leq j \leq n$.

On définit le processus empirique de kaplan-Meier comme suit :

$$\alpha_n(x) = n^{1/2} \left(F_n(x) - F(x)\right), \quad n \geq 1, \quad x \in \mathbb{R} \tag{5.3}$$

et le processus de censure de Kaplan-Meier comme suit :

$$\beta_n(x) = n^{1/2}(G_n(x) - G(x)), \quad n \geq 1, \quad x \in \mathbb{R}, \tag{5.4}$$

On définie pour $h \geq 0$ les accroissements $\xi_n(h, t; s)$ et $\eta_n(h, t; s)$ comme suit :

$$\begin{aligned} \xi_n(h, t; s) &= \alpha_n(t + hs) - \alpha_n(t) \\ &= n^{1/2}\eta_n(h, t; s) - n^{1/2}\left(F(t + hs) - F(t)\right), \quad h \geq 0, \quad s, t \in \mathbb{R}, \end{aligned} \tag{5.5}$$

et

$$\begin{aligned} \eta_n(h, t; s) &= F_n(t + hs) - F_n(t) \\ &= n^{-1/2}\xi_n(h, t; s) + \left(F(t + hs) - F(t)\right), \quad h \geq 0, \quad s, t \in \mathbb{R}, \end{aligned} \tag{5.6}$$

où $I(s) = s$ est la fonction identitée et $\{h_n : n \geq 1\}$ est une suite de constantes positives satisfaisants les conditions suivantes :

(H1) (i) $h_n \to 0$; (ii)$h_n \downarrow$; (iii) $nh_n \uparrow$;

(H2) $nh_n / \log_2 n \to \infty$;

(H3) (i) $nh_n / \log n \to \infty$; (ii) $(\log(1/h_n)) / \log_2 n \to \infty$;

(H4) $(\log(1/h_n)) / \log_2 n \to c \in [0, \infty)$;

(H5) $nh_n/\log n \to \gamma \in [0, \infty)$;

(H6) $(\log(1/h_n))/\log n \to d \in [1, \infty)$.

où

$$\log_2 u = \log_+(\log_+ u), \log_+ u = \log(u \vee e).$$

Notations : $u_n = \infty\,(v_n)$ (resp. $u_n \sim v_n$) si $v_n/u_n \to 0$(resp. $u_n/v_n \to 1$).

5.1 Lois limite de l'estimateur de la densité de survie

Soit K le noyau qui vérifie les conditions suivantes :

(K1) K à variation bornée sur $\mathbb{R}$; i.e. $\displaystyle\int_{-\infty}^{+\infty} |dK(u)| < \infty$;

(K2) Pour un certain T; $0 < T < \infty$, $K(u) = 0$ pour tout $|u| \geq \frac{1}{2}T$;

(K3) $\displaystyle\int_{-\infty}^{\infty} K(u)du = 1.$

On note par :

$$f_n(x) = \int_{-\infty}^{\infty} h_n^{-1} K((t-x)/h_n) dF_n(t) \tag{5.7}$$

l'estimateur à noyau de $f(x)$ [voir, eg., Watson et Leadbetter (1964 a, b), Tanner et Wong (1983)].

Soit, pour tout $x \in \mathbb{R}$,

$$\widehat{E} f_n(x) = \int_{-\infty}^{\infty} h_n^{-1} K((t-x)/h_n) dF(t). \tag{5.8}$$

De (5.3), (5.4), (5.5), (5.6), (5.7), et (5.8) et sous les conditions (K1) et (K2), d'aprés l'intégration par partie on obtient :

$$f_n(x) - \widehat{E}f_n(x) = -h_n^{-1}n^{-1/2}\int_{-T}^{T}\xi_n(h_n, x; u)dK(u).$$

Soit :

$$L(x) = \mathbf{P}(V \leq x),$$

la distribution de la variable aléatoire V.

On désigne par :

$$L^{-1}(u) = \inf\{x : L(x) \geq u\}, \quad 0 < u < 1$$

la fonction des quantiles de L, et par :

$$T_L = \sup\{x : L(x) < 1\}$$

le point limite superieur de la distribution de V.

Soit T_F et T_G les points limite superieur des distributions de X et Y tells que $\Theta = \min(T_F, T_G) > 0$, et soit a, a', b, b'des constants tele que $0 < a' < a < b < b' < \Theta$.

On suppose que F et G doivent vérifier les conditions suivantes :

(F1) $F(0) = G(0) = 0$;

(F2) (i) F et G sont continues sur $[a', b']$;

(ii) $f = (d/dx)F$ est définit, continue et stricement positive sur $[a', b']$.

On désigne par Ψ la fonction continue et stictement positive sur $[a', b']$.

On suppose que Ψ_n est un estimateur de Ψ qui vérifiée les conditions suivantes :

(C1) $\sup_{a\leq x\leq b} |\Psi_n/\Psi(x) - 1| \to 0$ en probabilité quand $n \to \infty$;

(C2) $\sup_{a\leq x\leq b} |\Psi_n/\Psi(x) - 1| \to 0$ presque sûrement quand $n \to \infty$.

Sous (H1),(H2), (K1)-(K3), (F1), (F2) et (C2), et les résultats de Deheuvels et Einmahl(1996), pour $x_0 \in [a, b]$, on a :

$$\begin{aligned}
\limsup_{n\to\infty} \quad & \pm \quad \left\{\frac{nh_n}{2\log_2 n}\right\}^{1/2} \left(f_n(x_0) - \widehat{E} f_n(x_0)\right) \left\{\Psi_n(x_0) \times \frac{1-G(x_0)}{f(x_0)}\right\}^{1/2} \\
& = \quad \{\Psi(x_0)\}^{1/2} \left\{\int_{-\infty}^{\infty} K^2(u)du\right\}^{1/2} \qquad p.s.
\end{aligned}$$

Convention : $c/(c+1) = 1$ si (H3) est vérifié, quand $c = \infty$.

Théorème 5.1 : Sous (H1)(i), (H3) ou (H4), (K1)-(K3), (F1)-(F2) et (C1), on a :

$$\begin{aligned}
& \lim_{n\to\infty} \left\{\frac{nh_n}{2\left\{\log\left(1/h_n\right) + \log_2 n\right\}}\right\}^{1/2} \\
& \times \sup_{a\le x\le b} \pm \left(f_n(x) - \widehat{E} f_n(x)\right) \left\{\Psi_n(x_0) \times \frac{1-G(x)}{f(x)}\right\}^{1/2} \\
& = \left(\frac{c}{c+1}\right)^{1/2} \left\{\sup_{a\le x\le b} \Psi(x)\right\}^{1/2} \left\{\int_{-\infty}^{\infty} K^2(u)du\right\}^{1/2}, \quad \text{en probabilité.}
\end{aligned}$$

Si, en plus, (H1)(ii)-(iii) et (C2) sont vérifiées, alors :

$$\begin{aligned}
& \limsup_{n\to\infty} \left\{\frac{nh_n}{2\left\{\log\left(1/h_n\right) + \log_2 n\right\}}\right\}^{1/2} \\
& \times \sup_{a\le x\le b} \pm \left(f_n(x) - \widehat{E} f_n(x)\right) \left\{\Psi_n(x_0) \times \frac{1-G(x)}{f(x)}\right\}^{1/2} \\
& = \left\{\sup_{a\le x\le b} \Psi(x)\right\}^{1/2} \left\{\int_{-\infty}^{\infty} K^2(u)du\right\}^{1/2} \qquad p.s.
\end{aligned}$$

et

$$\liminf_{n\to\infty}\left\{\frac{nh_n}{2\{\log(1/h_n)+\log_2 n\}}\right\}^{1/2}$$

$$\times\sup_{a\le x\le b}\pm\left(f_n(x)-\widehat{E}f_n(x)\right)\left\{\Psi_n(x)\times\frac{1-G(x)}{f(x)}\right\}^{1/2}$$

$$=\left(\frac{c}{c+1}\right)^{1/2}\left\{\sup_{a\le x\le b}\Psi(x)\right\}^{1/2}\left\{\int_{-\infty}^{\infty}K^2(u)du\right\}^{1/2}\quad p.s.$$

5.2 Lois limites fonctionnelles pour les accroissements des processus empirique de Kaplan-Meier

Soit $(B\,[0,1]\,,\mathfrak{U})$ [resp. $(AC\,[0,1]\,,\mathfrak{U})$ l'ensemble $B\,[0,1]$ (resp. $AC\,[0,1]$ de tout les fonctions l bornée (resp. absolument continue) sur $[0,1]$, muni de la topologie uniforme $\mathfrak{U}$ définie par la norme "sup".

Pour chaque $l\in AC\,[0,1]$, on note par $\dot{l}=(d/ds)l$ la dérivée de lebesgue de l et pour chaque $l\in B\,[0,1]$ soit :

$$|l|_H=\begin{cases}\left\{\int_0^1\dot{l}^2\,(s)\,ds\right\}^{1/2}, & \text{si } l\in AC\,[0,1] \text{ et } l(0)=0,\\ \infty, & \text{si non.}\end{cases}$$

Pour tout $\eta\geq 0$, soit :

$$\mathfrak{R}_\eta=\left\{l\in AC\,[0,1]:|l|_H^2\leq\eta\right\}.$$

L'inégalité suivante est une application directe de l'inégalité de Schwarz [voir. eg., (2.36), page 2021 deheuvels (1997)).

Pour chaque $l\in\mathfrak{R}_\eta$

$$\|l\|\leq|l|_H\leq\eta^{1/2}.$$

On définit une suite des sous ensembles des variables de $B\,[0,1]$, pour chaque $n\geq 1$,

$$\mathfrak{K}_n^{\pm}(\Psi_n) = \Big\{\pm\left\{2h_n\left(\log_+(1/h_n)+\log_2 n\right)\right\}^{-1/2}$$
$$\times\,\xi_n(h_n,x;I)\left\{\Psi_n(x)\times\frac{1-G(x)}{f(x)}\right\}^{1/2} : a\le x\le b\Big\}\subseteq B[0,1]$$

Soit $(\mathcal{E},\tau)$ désigne l'ensemble $\mathcal{E}$ muni de la topologie τ introduir par la métrique $d(l,g)$, avec $l,g\in\mathcal{E}$. Pour chaque $\varepsilon>0$ et $A\subseteq\mathcal{E}$, $A\neq\varnothing$, soit :

$$A^{\varepsilon}=\{g\in\mathcal{E}:\exists l\in A, d(l,g)<\varepsilon\}.$$

Soit pour chaque $A,B\subseteq\mathcal{E}$,

$$\mathfrak{D}_\tau(A,B)=\begin{cases}\inf\{\varepsilon>0:A\subseteq B^{\varepsilon}\text{ et }B\subseteq A^{\varepsilon}\}, & \text{ssi } \varepsilon>0 \text{ existes;}\\ \infty, & \text{si non.}\end{cases}$$

On considère la suite :

$$\{\mathfrak{B}_n\subseteq\mathcal{E}:n\ge 1\}\subseteq\mathcal{E}$$

pour lequel il existe un sous-ensemble compact K de ε, tel que :

$$\forall\varepsilon>0,\exists n_\varepsilon<\infty,\ \forall n\ge n_\varepsilon:\mathfrak{B}_n\subseteq K^{\varepsilon}.$$

Définition 5.1 : On dit que $\mathfrak{B}_n$ à un ensemble limite égale à $\mathfrak{B}\subseteq K$, si $\mathfrak{B}$ contient de tout les limites quand $j\to\infty$ des suites convergentes $l_{nj}\in\mathfrak{B}_{nj}$ avec $1\le n_1\le n_2\le ...$ et $n_j\to\infty$.

Définition 5.2 : On dit que $\mathfrak{B}_n$ à couverture minimale $\mathfrak{B}'\subseteq K$ si $\mathfrak{B}'$ contient de tout les limites quand $n\to\infty$ des suites convergente $l_n\in\mathfrak{B}_n$.

Remarque 5.1 : Il est remarquable [voir, eg., Deheuvels (1992)] que la suite $\mathfrak{B}_n\subseteq\mathcal{E}$, à un ensemble limite $\mathfrak{B}\neq\varnothing$ et à couverture minimale $\mathfrak{B}'$, si et seulement si les propriétés suivantes soit vérifiées :

(a) $\mathfrak{B}$ est un sous ensemble compact de $(\mathcal{E},\tau)$;

(b) $\forall\varepsilon>0$, on a pour n suffisament grand

$$\mathfrak{B}'\subseteq\mathfrak{B}_n^{\varepsilon}\ \text{ et }\ \mathfrak{B}_n\subseteq\mathfrak{B}^{\varepsilon}.$$

(c) $\forall \varepsilon > 0$, $l' \notin \mathfrak{B}'$ et $l \in \mathfrak{B}$ on a :

$$l' \notin \mathfrak{B}_n^{\varepsilon} \text{ infiniment souvent (i.s) (en } n\text{) et } l \in \mathfrak{B}_n^{\varepsilon} \text{ i.s (en } n\text{).}$$

Théorème 5.2 : [25]

Supposons que H1(i), (H3) ou (H4), (F1),(F2) et (C1) sont vérifiées.

Soit :

$$M = \sup_{a \leq x \leq b} \Psi(x),$$

alors :

$$\lim_{n \to \infty} \mathfrak{D}_{\mathfrak{U}} \left(\mathfrak{K}_n^{\pm}(\Psi_n), \mathfrak{R}_{Mc/(c+1)} \right) = 0, \quad \text{en probabilité.}$$

Si, en plus de ces conditions, (H1)(ii), (iii) et (C2) sont vérifiées, alors, avec probabilité 1, dans $(B[0,1], \mathfrak{U})$, la suite $\{\mathfrak{K}_n^{\pm}(\Psi_n) : n \geq 1\}$ à un ensemble limite égale à $\mathfrak{R}_M$, et à des couvertures minimales $\mathfrak{R}_{Mc/(c+1)}$.

En particulière, sous (H3), on a si $c/(c+1) = 1$,

$$\lim_{n \to \infty} \mathfrak{D}_{\mathfrak{U}} \left(\mathfrak{K}_n^{\pm}(\Psi_n), \mathfrak{R}_M \right) = 0 \quad p.s.,$$

et $\{\mathfrak{K}_n^{\pm}(\Psi_n) : n \geq 1\}$ à des couvertures minimales $\mathfrak{R}_M$ avec probabilité 1.

Théorème 5.3 : (Deheuvels et Einmahl (1996))

Sous (H1), (H2) et (F1), (F2) pour chaque $x_0 \in [a,b]$, si $M_0 = f(x_0)/(1 - G(x_0))$, la suite

$$\mathcal{P}_n = \left\{ \{2h_n \log_2 n\}^{-1/2} \xi_n(h_n, x_0; I) \right\} \subseteq B[0,1],$$

est presque sûrement relativement compact et a un ensemble limite dans $(B[0,1], \mathfrak{U})$ égale à $\mathfrak{R}_{M_0}$.

5.3 Lois limites fonctionnelles non-standard des processus à $P.L$ de Kaplan Meier

On suppose que :

$$nh_n/\log n \to \gamma \in (0,\infty)\,, \text{ quand } n \to \infty, \tag{5.9}$$

la suite $\{h_n : n \geq 1\}$ qui vérifiée (5.9) est appellé les suites intermédiaires.

il est convenable de travailler avec $\eta_n\,(h,t;I)$ plutôt que avec $\xi_n(h,t;s)$, ce fait est indiqué par la remarque suivante :

Remarque 5.2 : Rappelons les définitions (5.5) et (5.6), sous (5.9) et (F2)(ii), on a quand $n \to \infty$, uniformément sur tout $t \in [a,b]$:

$$\begin{aligned}\frac{n}{\log n}\eta_n\,(h,t;I) &= (1+o(1))(2\gamma)^{1/2}(2h_n\,\{\log(1/h_n)+\log_2 n\})^{-1/2} \\ &\quad \times\xi_n(h_n,t;I)+(1+o(1))\gamma f(t)I. \end{aligned}\tag{5.10}$$

De (5.10), la loi limite fonctionnelle, sous (5.9), avec les fonctios aléatoires

$$\frac{n}{\log n}\eta_n\,(h,t;I) \qquad \text{pour } t \in [a,b]\,,$$

est équivalent, a la loi limite fonctionnelle avec :

$$(2h_n\,\{\log(1/h_n)+\log_2 n\})^{-1/2}\,\xi_n(h_n,t;I) \qquad \text{pour } t \in [a,b]\,.$$

Soit :

$$\{E(x):\ a' \leq x \leq b'\}\,,$$

l'ensemble des fonctions continues et (strictement) positive, et $E_n(x)$, l'estimateur de $E(x)$ pour $a \leq x \leq b$ tel que :

(X.1) $\sup_{a\leq x\leq b}|E(x)/E(x)-1| \to 0$ presque sûrement, quand $n \to \infty$.

Soit :

$$\mathcal{L}_n\,(E_n) = \left\{\frac{n}{\log n}\eta_n\,(h_n,x;I)\,E_n(x) : a \leq x \leq b\right\}.$$

5.3.1 Notations et résultats préliménaires

On note par $I_{RC}[0,1]$ l'ensemble des distributions $l(x) = \mu([0,x])$, $x \in \mathbb{R}$ continue à droite de mésure de Radon μ non négatives bornées et à support dans $[0,1]$, et soit :

$$I_{AC}[0,1] = I_{RC}[0,1] \cap AC[0,1] \cap \{l : l(0) = 0\},$$

muni de la topologie uniforme μ.

Convenablement on définit la métrique de lévy, pour $l, g \in I_{RC}[0,1]$ par :

$$d_L(l,g) = \inf\{\varepsilon > 0 : l(x-\varepsilon) - \varepsilon < g(x) < l(x+\varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}.$$

On considère la fonction $\Phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ tel que, $\exists \mu \in \mathbb{R}$:

(C_μ) (i)$\Phi(\mu) = 0$.

(ii)Φ est convexe et $\Phi(\alpha) \geq 0$, pour tout α.

(iii)$\Phi(\alpha)/\alpha \to \infty$ quand $\alpha \to \infty$.

(iv)$\Phi(\alpha) = \infty$ pour $\alpha < 0$.

Pour tout $l \in B[0,1]$, on définit la fonctionnelle

$$J_\Phi(l) = \begin{cases} \displaystyle\int_0^1 \Phi\left(\dot{l}(u)\right) du, & \text{si } l(0) = 0 \text{ et } l \in AC[0,1] \text{ avec } \dot{l} = (d/du)l; \\ \infty, & \text{si non.} \end{cases}$$

On définit, pour $\mathcal{R} \geq 0$ le sous ensemble de $I_{AC}[0,1]$ par :

$$L_\Phi = \{l \in B[0,1] : J_\Phi(l) < \infty\}$$

et

$$B_\Phi = \{l \in B[0,1] : J_\Phi(l) < \mathcal{R}\}.$$

Lemme 5.1 : Sous (C_μ),

$$l \in B[0,1] \to J_\Phi(l)$$

à loi semi-continue avec considération de la topologie uniforme $\mathfrak{U}$.

Lemme 5.2 : Sous (C_μ), pour chaque $\mathcal{R} > 0$, $B_\Phi(\mathcal{R})$ est un convexe et un sous ensemble de $(B[0,1], \mathfrak{U})$.

Pour chaque $v > 0$ et $x \in \mathbb{R}$, soit :

$$\mathbf{h}_v(x) = v\mathbf{h}(x/v) \quad \text{où} \quad \mathbf{h}(x) = \begin{cases} x\log x - x + 1, & \text{pour } x > 0, \\ 1, & \text{pour } x = 0, \\ \infty, & \text{pour } x < 0. \end{cases}$$

Pour chaque $c > 0$ soit :

$$\delta_c^- = \sup\{x < 1 : \mathbf{h}(x) \geq 1/c\}$$

et

$$\delta_c^+ = \inf\{x > 1 : \mathbf{h}(x) \geq 1/c\}$$

Soit, pour $v > 0$,

$$J_v(l) = J_{\mathbf{h}_v}(l) = \begin{cases} \displaystyle\int_0^1 vh\left(\dot{l}(u)/v\right)dv, & \text{si } l \in I_{AC}[0,1], \\ \infty, & \text{si non.} \end{cases}$$

Soit $\Delta_v(\mathcal{R})$ et Δ_v les ensembles des fonctions définies pour chaque $v > 0$ et $\mathcal{R} > 0$ par :

$$\Delta_v(\mathcal{R}) = B_{hv}(\mathcal{R}) = \{l \in B[0,1] : J_\Phi(l) \leq \mathcal{R}\} \text{ et } \Delta_v = \Delta_v(1).$$

Lemme 5.3 : On a, pour chaque $\omega > 0$ et $\mathcal{R} > 0$

$$\inf_{l \in \Delta_\omega(\mathcal{R})} l(1) = \omega\delta^-_{\omega/\mathcal{R}} \quad \text{et} \quad \sup_{l \in \Delta_\omega(\mathcal{R})} l(1) = \omega\delta^+_{\omega/\mathcal{R}}.$$

Lemme 5.4 : Fixons $0 < A \leq B < \infty$.

Soit $\mathbf{k}(v)$ et $B(v)$ les fonctions continues et positives de $v \in [A, B]$. Alors :

$$\Delta(\mathbf{k}, \beta) = \bigcup_{A \leq v \leq B} \mathbf{k}(v)\,\Delta_{\beta(v)}$$

est un sous ensemble compact de $\left(I_{AC}\left[0,1\right],\mathfrak{U}\right)$.

Théorème 5.4 : [25]

Sous (H5) avec $\gamma \in (0,\infty)$, (F1), (F2) et (X1), on a :

$$\lim_{n\to\infty} \mathfrak{D}_{\mathfrak{U}}\left(\mathcal{L}_n\left(E_n\right),\mathcal{L}\left(E\right)\right)=0 \quad p.s.,$$

où

$$\mathcal{L}\left(E\right)=\bigcup_{a\leq v\leq b}\left\{\frac{E(x)}{1-G(x)}\right\}\Delta_{\gamma f(x)(1-G(x))}.$$

Annexe

Définition 1 : Soit un espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ où $\mathbf{P}$ est la mésure de probabilité sur $(\Omega, \mathcal{A})$. Un processus aléatoire, ou encore une fonction aléatoire réelle (f.a.r.) est une fonction à deux variables : t le temps et ω le hazard, et elle est notée $X(t, \omega)$, avec $t \in [0, \infty[$ et $\omega \in \Omega$.

La trajectoire est : $\omega \mapsto (X(t, \omega),\ t \geq 0)$.

Une f.a.r. à trajectoire continue (f.a.r.c.) est une application

$$X : \begin{array}{ll} [0, \infty[\times \Omega & \longrightarrow \mathbb{R} \\ (t, \omega) & \mapsto X(t, \omega) \end{array}$$

telle que :

a) pour presque tout ω, $t \mapsto X(t, \omega)$ est continue,

b) pour tout $t \geq 0$, $X_t : \omega \mapsto X(t, \omega)$ est une v.a. réelle.

Définition 2 : Soit un espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$, $t \in \mathbb{N}$ ou $\mathbb{R}^+$. Une filtration est une famille $\mathcal{F}_t$ de tribus, telle que :

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{A},$$

$\forall s \leq t$.

Définition 3 : Soient un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et $\mathcal{F}_t$ une filtration. Un processus $X = X(t, \omega)$ est dit $\mathcal{F}_t$-adapté si $\forall t$, X_t est $\mathcal{F}_t$-mésurable.

Définition 4 : Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. Soit $(M_t)_t$, $t \in \mathbb{N}$ ou $\mathbb{R}^+$ un processus réel défini sur Ω. Soit $(\mathcal{F}_t)_t$ une filtration sur Ω.

$(M_t)_t$ est une $\mathcal{F}_t$-**martingale** si :

1) $\forall t$, M_t est $\mathcal{F}_t$-adapté et $M_t \in \mathcal{L}^1$ (i.e. est intégrable),

2) $\forall s$, t, $0 \leq s \leq t$, $E(M_t \mid \mathcal{F}_s) = M_s$ $p.s.$

Définition 5 : Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et une filtration $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}^+}$.

Un $\mathcal{F}$-**processus croissant A** est un processus adapté à $\mathcal{F}$ à valeur réelles satisfaisant la propriété suivante :

pour tout $\omega \in \Omega$, les trajectoires $t \mapsto \mathbf{A}_t(\omega)$ sont croissantes, continues à droite et nulles en zéro.

Définition 6 : Un processus de comtage N est un processus cadlag (continu à droite avec une limite à gauche), adapté, nul en zéro, croissant et ayant des sauts d'amplitude 1.

Proposition 1 : Soit $M(t)$ une martingale. Alors $M^2(t)$ est une sous-martingale et

$$M^2(t) = M_t + \langle M \rangle_t \,,$$

avec $\langle M \rangle_t$ définit par :

$$\langle M \rangle_t = \lim_{|\delta| \to 0} \sum_i E\left[\left(M_{t_{i+1}} - M_{t_i} \right)^2 \mid \mathcal{F}_{t_i} \right]$$

$\langle M \rangle_t$ est appelé le processus prévisible croissant associé à $M(t)$.

Théorème 1 : (Rebolledo)

Si M_n est une suite de martingales, et si

(i) $\langle M_n \rangle_t$ converge en probabilité vers v_t déterministe,

(ii) $\forall \epsilon$, $\exists M_{n,\epsilon}$ suite de martingales telles qu'aucune différence $M_n - M_{n,\epsilon}$ n'ait une amplitude superieure à ϵ,

alors $M_n(t)$ a une limite $M(t)$ de processus croissant v_t, et $M(t)$ est un processus gaussien :

$$\frac{M_n(t)}{v_t} \xrightarrow{\pounds} N\left(0,1\right).$$

Définition 7 : Soit $X(s)$ un processus cadlag, nul en zéro, et à variation bornée. En posant :

$$X(]s,t]) = X(t) - X(s).$$

Soit une partition $t_0 = s < t_1 < ... < t_n = t$. Son **pas** est :

$$|\delta| = \sup_i |t_i - t_{i-1}| \,.$$

On appelle **produit-intégrale** (ou **produit infini**)

$$p_s^t (1 + dX) = \lim_{|\delta| \to 0} \prod_{i=1}^{n} [1 + X (]t_{i-1}, t_i])] ,$$

Propriété 1 : Si $X(t)$ est continue, alors $p_0^t (1 + dX) = e^{X(t)}$.

Théorème 2 : (Duhamel)
Soient :

$$Y = p(1 + dX) \text{ et } Y' = p(1 + dX').$$

Alors :

$$Y(t) - Y'(t) = \int_{s \in [0,t]} p_{[0,s)} (1 + dX) (X(ds) - X'(ds)) \, p_{(s,t]} (1 + dX')$$

Si $Y'(t)$ est régulière, alors :

$$\begin{aligned} \frac{Y(t)}{Y'(t)} - 1 &= \int_{s \in [0,t]} p_{[0,s)} (1 + dX) (X(ds) - X'(ds)) \left[p_{[0,s)} (1 + dX') \right]^{-1} \\ &= \int_0^t \frac{Y(s-)}{Y'(s)} [X(ds) - X'(ds)] . \end{aligned}$$

Définition 8 : (**Processus de Wiener à deux paramètres**)

Soit $X(\tau)$ ($\tau = (x, y) \in \mathbb{R}_+^2$) le processus stochastique à deux paramètres.

On considère le rectangle $R = [x_1, x_2] \times [y_1, y_2] \in \mathbb{R}_+^2$ ($0 \leq x_1 < x_2 < \infty$, $0 \leq y_1 < y_2 < \infty$).

On définit $X(R)$ le mésure de R par :

$$X(R) = X(x_2, y_2) - X(x_1, y_2) - X(x_2, y_1) + X(x1, y_1).$$

Le processus stochastique $\left\{ W(\tau), \ \tau \in \mathbb{R}_+^2 \right\}$ est un processus de Wiener à deux paramètres si et seulement si :

1- $\forall R = [x_1, x_2]\times[y_1, y_2]\,, W(R) \in \mathcal{N}(0, \lambda(R))$, où $\lambda(R) = (x_2 - x_1)(y_{2-}y_1)\,.$

2- $W(0, y) = W(x, 0) = 0 \quad (x \geq 0,\ y < \infty)$.

3- $W(R_2) - W(R_1)$, $W(R_4) - W(R_3)$,..., $W(R_{2i}) - W(R_{2i-1})$ sont des variables aléatioires indépendantes si $R_1, R_2, ..., R_n$ sont des rectangles disjoints, $(i = 2, 3, ...)$.

4- La trajectoire de $W(z;\omega)$ est continue en z avec probabilité 1.

La covariance de $W(\tau)$ est :

$$R(\tau_1, \tau_2) = EW(\tau_1)W(\tau_1) = \min(x_1, x_2) \cdot \min(y_1, y_{2-})$$

où $\tau_1 = (x_1, y_1)$, $\tau_2 = (x_2, y_2)\,.$

Définition 9 : (**Processus de Wiener à un paramètre**)

Le processus stochastique $\{W(t;\omega) = W(t);\quad 0 \leq t < \infty\}$, où $\omega \in \Omega$,et $(\Omega, A, \mathbf{P})$ est un espace de probabilité, est appelée le processus de Wiener si :

1- $W(t) - W(s) \in N(0, t - s)$ pour tout $0 \leq s < t < \infty$ et $W(0) = 0$;

2- $W(t)$ à des accroissement indépendantes, tel que : $W(t_2) - W(t_1)$, $W(t_4) - W(t_3)$,...,$W(t_{2i}) - W(t_{2i-1})$ sont des variables aléatoires indépendantes pour tout $0 \leq t_1 < t_2 \leq t_3 < t_4 \leq ... \leq t_{2i-1} < t_{2i} < \infty$, $(i = 2, 3, ...)$;

3- La trajectoire de $W(t;\omega)$ est continue en t avec probabilité 1.

Définition 10 : (**Processus de Kiefer**)

Soit $W(x, y)$ le processus de Wiener à deux paramètres.

Le processus de Kiefer $\{K(x, y);\ 0 \leq x \leq 1,\ 0 \leq y < \infty\}$ est défini par :

$$K(x, y) = W(x, y) - xW(1, y)\,.$$

Définition 11 : (**Pont Brownien**)

Le processus stochastique $\{B(t);\quad 0 \leq t \leq 1\}$ est un pont brownien si et suelement si :

1- La distribution joint de $B(t_1)$, $B(t_2)$,..., $B(t_n)$ $(0 \leq t_1 < t_2 < ... < t_n \leq 1;\ n = 1,\ 2,...)$ est gaussiene avec $EB(t) = 0$,

2- La covariance de $B(t)$ est :

$$R(s, t) = EB(s)B(t) = \min(s, t) - st.$$

Bibliographie

[1] Aalen O. O., (1978). Non-Parametric inference for a family of counting processus. Annals of statistcs, vol. 6 P. 701-726.

[2] Aly, E.-E.A.A., Csörgő, S., and Horàth, L. (1985). Strong approximation of the quantile process of the product-limit estimator. Journal of multivariate analysis. 16, 185-210.

[3] Cheng, K.F. (1981). on almost sure representations for quantiles of the product limit estiùator with applications. Research report No. 87, SUNY buffalo, dep of statistics.

[4] Csörgő, S., and Horàth, L. (1983), "the rate of strong uniform consistency foe the product limit estimator," Zeitschrift für wahrscheinlichkeitsthecrie and Verwande Gebiete, 62, 411-426.

[5] Csörgő, M. (1983). Quantiles processus with statistical applications. CBMS-NSF regional conference series in applied mathematics, SIAM, philadelphia, PA.

[6] Diehl, S., and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. J. Multivariate Anal. 25 299-310.

[7] Deheuvels, P. (1992). Functional laws of the iterated logarithm for large increments of empirical and quantile processes. Stochastic process. Appl. 43 133-163.

[8] Deheuvels, P. (1996). Functional laws of the iterated logarithm for large increments of empirical processes. Statist. Neerlandica 50 261-280.

[9] Deheuvels, P. and Einmahl, J.H.J. (1996). On the strong limiting behavior of local functionals of empirical processes based upon censored data. Ann. Probab. 18 1693-1722.

[10] Deheuvels, P. (1997).Strong laws for local quantile processes. Ann. Probab. 25 2007-2054.

[11] Deheuvels, P. and Einmahl, J.H.J. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. Ann. Probab 19 1301-1335.

[12] Efron, B. (1981), "censored data and the bootstrap," Journal of the american statistical association, 76, 312-319.

[13] Földes, A. and Rejtö, L. (1981). A LIL type result for the product limit estimator. Z. Wahrsch. Verw. Gebiete 56 75-86.

[14] Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. Annals of statistics, 13, 428-433.

[15] Grenwood. M, (1926). The natural duration of cancer repports on public health and medical subjects, vol. 33 P. 126.

[16] Kaplan, E. L., and Meier, P. (1958), "Non parametric Estimation from incomplete observations," Journal of the American Statistical Association, 53, 457-481.

[17] Lio, Y.L., Padjett, W.J., and Yu , K.F. (1986). On the asymptotic properties of a kernel -type estimator from right censored samples. Journal of statistical planning and inference, 14, 169-177.

[18] Lio, Y.L. and Padjett, W.J. (1987). On the mean squared error of non parametric quantile estimators under random right censorship- Communications in statistics, theory and methods, 16,1617-1628.

[19] Lo, S.H. and SINGH, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. Prob. Theory related. Fields 71 455-465.

[20] Major, P. and Rejtö, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. Ann. Statist. 16 1113-1132.

[21] Nelson, W.B. (1972). Theory and application of hazard plotting for censored data. technometrics, vol. 14 p. 945-965, .

[22] Padgett, W. J. (1986). A kernel-type estimator of a quantile function from right-censored data. Journal of statistical planning and inference, 14, 169-177.

[23] Padjett, W.J., and Thombs, L.A. (1986). Smmoth non parametric quantile estimation under censoring : Simulations and bootstrap methods. comm. statist. Simulation comport., 15, 1003-1025.

[24] Parzen, E. (1979). Nonparametric statistical data modeling. J. Amer. Statist. Assoc. 74 105-131.

[25] Paul Deheuvels and John H.J.Einmahl. Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. Ann.Probab. (to appear).

[26] Sander, J. (1977). The weak convergence of quantiles of the product limit estimator. Technical repports, stanford university, departement of statistics.

[27] Shorack,G. R., and Wellner, J. A. (1986). Empirical processus with applications to statistics. Wiley, New York.

[28] Stute, W. (1982a). the oscillation behavior of empirical processes. Ann. Probab. 10 86-107.

[29] Stute, W. (1982b). A law of the logarithm for kernel density estimators. Ann. Probab. 10 414-422.

[30] Tanner, M. and Wong, W. (1983). The estimation of the hazard function from randomly censored data by the kernel method. Ann. Statist. 11 422-439.

[31] Watson, G. S. and LEADBETTER, M. R. (1964 a). Hazard analysis I. Biometrika 51 175-184.

[32] Watson, G. S. and LEADBETTER, M. R. (1964 b). Hazard analysis II. Sankhyā Ser. A 26 101-116.

[33] Xiang, X. (1994). A law of the logarithm for kernel quantile density estimators. Ann. Probab. 1078-1091.

[34] Xiang, X. (1994). Law of the logarithm for density and hazaed rate estimation for censored data. J. Multivariate Anals. 49 278-286.

[35] yang, S. S. (1985). A smooth nonparametric estimator of a quantile function. J. Amer. Statist. Assoc. 80 1004-1011.