République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université Mohammed Seddik Ben Yahai - Jijel

Faculté des Sciences Exacte et Informatique
Département de Mathématique

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques.

Option : Probabilités et Statistique.

Thème

Propriétés asymptotiques d'estimateurs non paramétriques de la distribution des données censurées

Présenté par :

Bounefikha Yasmina

Devant le jury :

Président	**Khaoula Boudjerda**	**M.C.B** Université Mohammed Seddik Ben Yahia, Jijel
Encadreur	**Merieme Madi**	**M.A.A** Université Mohammed Seddik Ben Yahia, Jijel
Examinateur	**Zeyneb Abdi**	**M.A.A** Université Mohammed Seddik Ben Yahia, Jijel

Promotion **2020/2021**

♡ *Remerciements* ♡

Avant de présenter ce travail, mes remerciements vont tout d'abord à :

Dieu le tout puissant qui m'a donné la volonté et la santé pour accomplir ce travail et qui m'a aidé à franchir un pas vers le chemin du savoir.

je remercie beaucoup mon encadreur **Mme.Madi Merieme** qui m'a encadré pendant la période de la réalisation de ce travail, sa disponibilité, malgré ses responsabilités, ses orientations permis de mener à merveille ce travail.

je tient aussi a remercier mon enseignant **Mr.GHERDA Mebrouk** pour ses sacrifice et aide précieuse jusqu'à l'accomplissement de ce travail.

mes remerciements vont également aux membres du jury pour l'intéret qu'ils ont porté à mon recherche en acceptant d'examiner mon mémoire et de l'enrichir par leurs propositions.

Je remercie beaucoup l'enseignant **Boukeloua Mohamed** pour leur conseils et pour l'aide moi.

Mes remerciements les plus vives et chaleureuses a tous les membres de ma famille.

Enfin, que tous ceux et celles qui, de loin ou de prés ils m'ont apporté leur aide et soutien trouveront ici, mon reconnaissance et sympatie.

♡ *Dédicases* ♡

Je voudrais dédier ce travail :

A ma mère, a mon cher père qui m'a soutenu et encouragé, a mes frères et sœur, a tous ma famille et particulièrement ma tante.

A mes amies **Halima, Nour Elhouda, Imen, Aziza** et particulièrement **Nadjet** pour leurs conseils.

A mes collèges de promo 2020/2021.

Table des matières

Table des figures

Liste des tableaux

Notations

$\mathbb{P}$: convergence en probabilité.

P.s : convergence presque sur.

m.q : convergence en moyenne quadratique.

p.c : convergence presque complète.

D : convergence en loi.

O : l'ordre de vitesse de la convergence.

0(h) : une fonction qui tend vers 0 quand h tend vers 0.

Introduction

On rencontre les données censurées dans différent domaines de recherche, en médecine, en biologie, en économie, en fiabilité,... Des observations sont dites censurées lorsque la variables étudiée représente la durées à un événement terminal ; et que l'étude est limitée dans le temps. Cette variable est dite de survie, variable aléatoire positive. Dite censurée si elle n'est pas intégralement observée. On s'intéresse dans ce mémoire à l'estimation non paramétrique des fonctions de distribution des données censurées. En fixant l'étude sur les estimateurs des fonctions de survie, de densité et du taux de hasard, on démontre le comportement asymptotique et la convergence presque complète de chaque estimateur. Le premier chapitre est consacré aux rappels sur la terminologie utilisée dans le domaine des données de survie, et les estimateurs non paramétriques, les plus utilisés, des fonctions de distributions des données censurées. En particulier, l'estimateur de Kaplan-Meyer de la fonction de survie, et les estimateurs à noyaux des fonctions de densité et du taux de hasard. Dans le deuxième chapitre, on s'intéresse aux propriétés asymptotiques des estimateurs cités précédemment, en se basant sur les travaux de Kaplan-Meyer et Boukeloua. On se concentre le comportement asymptotique et la convergences des estimateurs On conclu ce mémoire, par une étude de simulation dont le but est l'évaluation des qualités des estimateurs à noyau étudiés dans le deuxième chapitre.

CHAPITRE 1

Estimation non paramétrique de fonctions de distribution des données censurées

Introduction

L'objectif de ce chapitre est de faire un rappel sur les estimateurs non paramétriques, les plus utilisés, des fonctions de distribution des données censurées, notamment de la fonction de survie, de densité de probabilité et celle du taux de hasard. On introduit ce chapitre, par un rappel sur la terminologie des données censurées. Pour ces derniers on calcul la fonction de densité et de répartition dans les cas de censure à droite et double. Puis on donne quelque estimateurs des fonctions de distribution.

1.1 Données censurées (données de survie)

Dans nombreuses des applications statistiques on a mené à étudier une variable de durée,c'est la durée d'apparition d'un évènement au cour du temps ; ou bien le temps qui s'écoule entre deux évènements.Comme par exemple :

En médecine : durée de guérison d'un patient, durée de rémission d'un malade,...

En fiabilité : durée de vie d'une lampe, durée de vie d'un matériel,..

En biologie : durée d'apparition de parasite en culture de cellules,...

En économie : durée de chômage.

En éducation : durée d'apprentissage d'une langue.

En assurance : durée de vie d'un contrat (durée entre la date de résiliation et la date de création du contrat).

Les études de survie nécessitent la connaissance d'un certain nombres de données qu' on doit les connaitre pour chaque sujet.

Date d'origine (DO) : C'est la date d'entrée dans l'étude, pour les essais cliniques ,c'est à dire l'origine de l'analyse de survie. Mais généralement pour un patient c'est la date d'atteindre une maladie.cette date se diffère d'un individu à un autre.
Date de point (DP) : C'est la date au-delà de laquelle on arrêttera l'étude et on ne tiendra plus compte des informations sur les sujets après cet instant. Commune à tous les individus.
Date des dernières nouvelles (DDN) : C'est la date la plus récente ou des informations sur un sujet ont été recueillies.

A partir de la date des dernières nouvelles, de la date d'origine et de l'état du sujet, il est possible de définir :
Durée de surveillance T_i : C'est la durée entre la date d'origine et la date des dernières nouvelles, appelé aussi période de surveillance de l'instant de la survenue de l'évènement d'un sujet i.
Durée de recul L_i : appelée aussi délai de censure ; c'est la période entre la date d'origine et la date de point. C'est la période maximale d'observation. Elle peut être connue ou inconnue ; déterministe ou aléatoire, dépendante de l'individu ou non.
Durée de participation t_i : C'est le délai réellement observé pour chaque individu. C'est la durée de suivi. On distingue deux cas :
$DDN < DP$: Ce qui est équivalent à $t_i \in [DDN, DP]$, la durée de participation est identique à la durée de surveillance $t_i = T_i$. Si l'individu n'a pas subi l'évènement à la DDN, le temps de participation nous donne une observation incomplète, il sera considéré comme perdu de vue. Mais si l'individu a subi l'évènement à la DNN, le temps de participation nous donne une observation complète de son état, .
$DDN > DP$: Ce qui est équivalant à $t \in [DO, DP]$, dans ce cas $t_i = L_i$.
Le sujet est considéré comme exclu vivant , car s'il n'a pas encore subi l'évènement avant la DP, c'est sont état initial qui sera pris en considération, même s'il y a un changement d'état après la DP.

1.1.1 Cas et types de censure

On dit que les données sont censurées lorsque il y a des informations partielles sur la valeur de la variable d'intérêt (temps de survie), elle se présentent quand on ne dispose pas d'assez de temps pour attendre que toutes les observations atteignent l'évènement d'intérêt.

Définition 1.1. [11] *Le temps (la durée) de survie T; c'est le temps écoulé entre la date d'origine et la date de l'évènement. C'est une variable aléatoire positive.*

Définition 1.2. [11] *La durée de survie T est dite censurée si elle n'est pas intégralement observée.*

Cas de censure

Pour l'individu i considérons :
T_i : le temps de survie.
C_i : le temps de censure, c'est à dire la durée d'observation de l'individu entre le début de l'étude et la date de visite.
X_i : la durée réellement observée

Cas de censure à droite

Il y a censure à droite lorsque la durée de vie T est supérieure à la durée observée X .
Par exemple la durée de vie d'un genre de machines précis mais que ces dernières tombent en panne s'il se produit une surtension d'électricité. Ici la durée de vie de la machine est censurée à droite par instant auquel se produit la surtension.

Cas de censure à gauche

Il ya censure à gauche lorsque la durée de survie T est inférieure à la durée observée X. Par exemple en fiabilité un composant électronique monté en parallèle avec un ou plusieurs autres composants. Une panne de ce composant n'entraine pas nécessairement l'arrêt du système : le système peut continuer à fonctionner jusqu'à ce que cette panne soit détectée(par exemple lors d'un contrôle ou en cas de l'arrêt du système). La durée observée pour ce composant est alors censurée à gauche.

Cas de censure double

On dit qu'on a une censure double si on a des données censurées à droite et des données à gauche dans le même échantillon. Par exemple, l'étude qui s'intéresse à l'age auquel les enfants savaient déjà effectuer les taches étudiées, on sait seulement alors que l'âge où ils ont appris est inférieur a leur âge à la date du début de l'étude. A la fin de l'étude, certains enfants ne savaient pas encore accomplir ces tache et on sait alors seulement que l'âge auquel ils apprendront éventuellement ont appris est supérieur à leur âge à la fin de l'étude. L'âge au début de l'étude est évidement inférieur à l'âge à la fin de l'étude, l'âge d'intérêt est observé ssi il se trouve dans la période d'étude.

Cas de censure par intervalle

Dans le cas de la censure par intervalle, on observe à la fois une borne inférieure et une borne supérieure de la durée d'intérêt. Ceci arrive dans les études de suivi médical où les patients sont contrôlé périodiquement. Ceci se produit notamment si un patient se rend à l'hôpital à des dates régulières : s'il ne se présente pas à un rendez-vous, on sait seulement que son décès s'est produit dans l'intervalle entre la dernière visite et le rendez-vous.

Cas de censure mixte

On dit qu'il ya censure mixte lorsque deux phénomènes de censure l'un à droite et l'autre à gauche peuvent empêcher l'observation du phénomène d'intérêt sans qu'on puisse nécessairement déterminer un intervalle auquel appartient, au lieu d'observer un échantillon de la variable d'intérêt T, on observe un échantillon du couple (Z,A) avec

$$Z = max(min(T, C), L),$$

et

$$A = \begin{cases} 0, \ si \ L < T < C; \\ 1, \ si \ L < C < R; \\ 2, \ si \ min(T, C). \leq L \end{cases}$$

Où R,L sont des variables de censure et A est l'indicateur de censure

Types de censure

Les catégories de censure peuvent se décliner en fonction du mode ou mécanisme de censure . On obtient les types suivants :

Censure de type I : L' expérimentateur fixe une date (non aléatoire) de fin d'expérience. La durée de participation maximale est alors fixée et vaut, pour chaque observation, la différence entre la date de fin d'expérience et la date d'entrée du patient dans l'étude. Le nombre d'évènements observés est quant à lui aléatoire. Ce modèle est souvent utilisée dans les études épidémiologiques. Pour une censure à date fixe $C_i = C$, i=1,...,n. On peut avoir

• **une censure à droite de type** I : La durée de survie n'est pas observable au delà d'une durée maximale C fixée. C'est à dire au lieu d'observer les variables aléatoires $T_1, ..., T_n$ on observe T_i uniquement lorsque $T_i \leq C$ sinon on sait que $T_i > C$ on utilise

$$Z_i = min(T_i, C), i = 1, ..., n$$

Par exemple, dans l'apprentissage d'une langue par un groupe d'étudiant durant un stage de période fixée, on note T la durée d'apprentissage de cette langue, pour certains étudiants nous allons observer leurs durées T_i d'apprentissage de la langue. Par contre pour d'autres leurs T_i ne seront pas observées car le stage est limité dans le temps.

Censure de type II :(Attente) L'expérimentateur fixe à priori le nombre d'évènements d'intérêt à observer. La date de fin d'expérience devient alors aléatoire, le nombre d'évènements d'intérêt étant quand à lui non aléatoire. Au lieu d'observer $T_1, ..., T_n$ on observe $T_1 \leq T_2 \leq ... \leq T_n$.

Ce modèle est souvent utilisé dans les études de fiabilité.

• **une censure à droite de type II** :On décide d'observer les durées de survie de n patients jusqu'à ce que k d'entre eux soient décédés et d'arrêté l'étude à ce moment là, $T_{(i)}$ et $Z_{(i)}$les statistiques d'ordres des variables aléatoires T_i et Z_i, la data de censure est Z_k et on observe les variables $Z_1 = T_1, Z_2 = T_2, ..., Z_k = T_k, Z_k = T_{k+1}, ..., Z_k = T_n$

Censure de type III (aléatoire de type I)

C'est typiquement ce modèle qui utilisé pour les essais thérapeutique. Dans ce type d'expérience, la date d'inclusion du patient dans l'étude est fixée, mais la date de fin d'observation est inconnue. Le nombre d'évènement observés et la durée totale de l'expérience sont aléatoire.

• **Censure à droite de type III** : soit D un variable aléatoire au lieu d'observer la variable aléatoire T qui on intéresse , on observe le couple des variables aléatoires (Z, δ) avec

$$Z = min(T, D)$$

et

$$\delta = \mathbb{1}_{T \leqslant D} = \begin{cases} 1 \text{ si } T \leqslant D \text{ (pas de censure, on observe les données complètes)} \\ 0 \text{ si } T > D \text{ (il y a une censure à droite)} \end{cases}$$

• **Censure à gauche de type III** : soit G une variable aléatoire au lieu d'observer la variable aléatoire T qu'on intéresse, on observe le couple des variables aléatoires (Z, δ) avec

$$Z = max(T, G),$$

et

$$\delta = \mathbb{1}_{\{T \geq G\}} = \begin{cases} 1, \text{ si } T \geq G \text{ (pas de censure, on observe les données complètes);} \\ 0, \text{ si } T < G \text{ (il y a une censure à gauche).} \end{cases}$$

• **censure par intervalle de type** III

La durée T est dite censurée par intervalle si au lieu d'observer $T_1, ..., T_n$ on observe aléatoirement $(X_i, \delta_i), i = 1, ..., n$ où

$$X_i = \max[\min(T_i, C_i^{(R)}), C_i^{(L)}].$$

Dans ce type d'expérience, la date d'inclusion du patient dans l'étude est fixée, mais la date de fin d'observation est inconnu ; le nombre d'évènements observés et la durée totale de l'expérience sont aléatoires. Par exemple, notons T l'âge à laquelle une certaine maladie apparait pour la première fois chez un individu
Après un examen médical on a reçu deux types de réponses :

i) l'individu a déjà été malade mais l'age exact de la première apparition n'a pas été retenu,

ii) l'individu n'a jamais eu la maladie.

Dans le premier cas on n'a pas observé T mais est inférieur à l'âge de l'individu lors de l'examen il s'agit une observation à gauche.
Dans le $2^{\text{ème}}$ cas on sait seulement que T est supérieur à l'age de l'individu donc on a une observation censurée à droite.

Un autre exemple, pour détecter les composants défectueux d'un processus de production industriel, on effectue des contrôles selon des dates aléatoires, lorsqu'on constate qu'un composant est à changer, on sait seulement qu'il est tombé en panne entre les dates de deux contrôle successifs ; c'est un exemple de censure par intervalle.

1.2 Distributions de la durée de survie

On s'intéresse dans cette section aux définitions et interprétation dans différentes fonctions de distribution de la durée de survie, ainsi que les relations existant entres ces fonctions.

1.2.1 Durée de survie absolument continue

On suppose que la durée de survie T est une variable aléatoire positive et absolument continue. Alors sa loi de probabilité peut être définie par l'une des cinq équivalentes fonctions :

Définition 1.3. [16]
textbf(fonction de répartition F)
La fonction de répartition de T, notée F, est définie comme suit :

$$F(t) = \mathbb{P}(T \leq t),\ t \geq 0,$$

elle désigne la probabilité que l'événement d'intérêt ait lieu avant t.

Définition 1.4. [16]*(densité de probabilité f)*

Si F admet une dérivée au point t, notons cette dérivée f, alors f est appelée densité de probabilité de T, définie sur$[0, +\infty[$ *par*

$$\begin{aligned} f(t) &= \lim_{dt \longrightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + dt)}{dt} \\ &= \lim_{dt \longrightarrow 0} \frac{F(t + dt) - F(t)}{dt}. \end{aligned}$$

Elle désigne que l'évènement d'intérêt ait lieu après, dans un petit intervalle de temps.

Remarque 1. *Puisque T est une variable aléatoire absolument continue, les notations*

$$F(t) = \mathbb{P}(T \leq t) \text{ et } F(t) = \mathbb{P}(T < t)$$

sont identiques.

Définition 1.5. [16] *(fonction de survie S)*
La fonction de survie S représente la probabilité de survivre au moins jusqu'au temps t. Autrement dit, la probabilité de ne pas avoir fait l'évènement d'intérêt jusqu'à l'instant t. Elle est définie comme suit :

$$S(t) = \mathbb{P}(T \geq t) = 1 - F(t), t \geq 0.$$

Remarque 2. $S(t)$ *est une fonction monotone décroissante et continue telle que* $S(0) = 1$ *et* $\lim_{t \to +\infty} S(t) = 0$.

Définition 1.6. *[16] (taux de hasard ou risque instantané* λ*)*
Le taux de risque instantané λ *est la probabilité qu'un évènement survienne dans un petit*

intervalle de temps après t, sachant qu'il n'a pas eu lieu avant t.

$$\lambda(t) = \lim_{dt \longrightarrow 0} \frac{\mathbb{P}(t \leq T < t + dt | T \geq t)}{dt}.$$

Définition 1.7. [16]*(taux de risque cumulé ou taux de hasard cumulé Λ)*
Le taux de risque cumulé est défini sur l'intervalle $[0, t]$ *par :*

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

Il représente la probabilité de passage à l'évènement d'intérêt dans [t,t+dt] en prenant en considération les informations précédentes.

Relations entre les fonctions de distribution de survie

Les fonctions de distribution de probabilité de la durée de survie T sont liée par les relations suivantes :

1. $\lambda(t) = \frac{f(t)}{s(t)} = \frac{f(t)}{1-F(t)} = \frac{d\Lambda(t)}{dt}$
2. $\Lambda(t) = -\ln(S(t))$
3. $S(t) = -\exp(\Lambda(t))$
4. $F(t) = 1 + \exp(\Lambda(t))$
5. $f(t) = -\frac{dS(t)}{dt}$

Preuve.

$$\begin{aligned}
\lambda(t) &= \lim_{dt \longrightarrow 0} \frac{\mathbb{P}(t \leqslant T < t + dt \; T \geqslant t)}{dt} \\
&= -\frac{d\ln(S(t))}{dt} \\
&= \frac{f(t)}{S(t)} \\
&= \frac{f(t)}{1 - F(t)} \\
&= \frac{d\Lambda(t)}{dt} \\
\Lambda(t) &= \int_0^t \lambda(u)du \\
&= -\ln(S(t))
\end{aligned}$$

$$
\begin{aligned}
S(t) &= \mathbb{P}(T \geq t) \\
&= 1 - F(t) \\
&= 1 - \int_0^t f(u)du \\
&= \exp\{-\int_0^t \lambda(u)du\} \\
&= -\exp(\Lambda(t)) \\
F(t) &= \mathbb{P}(T > t) \\
&= 1 - S(t) \\
&= \int_0^t f(u)du \\
&= 1 + \exp\int_0^t \lambda(u)du \\
&= 1 + \exp\Lambda(t) \\
f(t) &= \lim_{dt \longrightarrow 0} \frac{\mathbb{P}(t \leq T < t + dt)}{dt} \\
&= \frac{dF(t)}{dt} \\
&= -\frac{dS(t)}{dt}
\end{aligned}
$$

■

Exemple 1.8. *Supposons que le temps de survie d'une population est de densité de probabilité* $f(t) = e^{-t};\ t \geqslant 0$, *donc la fonction de répartition de* T *est* $F(t) = \int_0^t f(x)dx = 1 - e^{-t}$ *et la fonction de survie est* $S(t) = e^{-t}$ *et la fonction de taux de hasard est* $\lambda(t) = \frac{e^{-t}}{e^{-t}} = 1$.

1.2.2 Paramètres de position et de dispersion associés à la distribution de survie

Paramètre de position

On calcule l'espérance et la variance de la durée de survie T en utilisant n'importe quelle des ces fonctions de distribution.

Définition 1.9. [16] *(temps moyen de survie)*
L'espérance de la durée de survie T, appelé aussi temps moyen de survie E est donnée par :

$$\begin{aligned} E(T) &= \int_0^{+\infty} tf(t)dt \\ &= \int_0^{+\infty} S(t)dt. \end{aligned}$$

Paramètre de dispersion

Définition 1.10. [16] *(variance de survie)*
La variance de la durée de survie est donnée par :

$$\begin{aligned} var(T) &= E(T^2) - [E(T)]^2 \\ &= 2\int_0^{+\infty} tS(t)\,dt - [\mathbb{E}(T)]^2. \end{aligned}$$

Définition 1.11. [16] *(quantiles)*
Les quantiles de la durée de survie pour $0 \leq p \leq 1$, notés Q_p, sont définis par :

$$\begin{aligned} Q_p &= \inf(t, F(t) \geqslant p) \\ &= \inf(t, S(t \leqslant 1-p)) \end{aligned}$$

Q_p représente le temps où la proportion p d'une population a subi l'évènement d'intérêt.

Remarque 3. *La médiane $Q_{\frac{1}{2}}$ est le quantile particulier satisfaisant $S(Q_{\frac{1}{2}}) = 0.5$. Distingue que la moitie de la population a subi l'évènement d'intérêt.*

1.2.3 Durée de survie discrète

Soit T une variable aléatoire discrète , ses valeurs $(t_i)_{i=0,\dots,n}$ sont des valeurs ordonnées en ordre croissant.

Définition 1.12. *[16] (fonction de survie)*

$$S_T(t) = \mathbb{P}(T > t) = \sum_{i,t_i>t} \mathbb{P}(T = t_i)$$

Définition 1.13. [16] *Le taux de hasard λ_T, de la variable aléatoire discrète est donnée par :*

$$\begin{aligned}\lambda_T(t_i) &= \mathbb{P}(T = t_i \; T \geq t) \\ &= \frac{\mathbb{P}(T = t_i)}{S_T(t_{i-1})} \\ &= 1 - \frac{S_T(t_i)}{S_T(t_{i-1})}\end{aligned}$$

Exemple 1.14. *Soit T une variable aléatoire prenant les valeurs 1,2 et 3, avec même probabilité qui est $\frac{1}{3}$, la fonction de survie de T est*

$$S_T(t) = \begin{cases} 1, \; si \, 0 \leqslant t < 1 \\ \frac{2}{3}, \; si \, 1 \leqslant t < 2 \\ \frac{1}{3}, \; si \, 2 \leqslant t < 3 \\ 0, \; si \, t \geqslant 3 \end{cases}$$

et le taux de hasard est

$$\lambda_T(t) = \begin{cases} \frac{1}{3}, \; si \, t = 1 \\ \frac{1}{2}, \; si \, t = 2 \\ 1, \; si \, t = 3 \\ 0, \; sinon \end{cases}$$

1.3 Fonctions de loi de probabilité pour les données censurées

On représente les fonction de répartition et de densité pour les modèles de censure à droite (à gauche) et double

1.3.1 Le modèle à censure à droite

Soit T et C deux variables aléatoires positives et absolument continue de densité de probabilités respective f_T, f_C, de fonctions de survies S_T, S_C et de fonctions de répartitions F_T, F_C. Sous l'hypothèse que T et C sont indépendantes,on considère le cas de la censure aléatoire à

droite, les observations sont les couples $(Z_1;\delta_1),...,(Z_n;\delta_n)$ où $Z_i = min(T_i;C_i)$ et $\delta_i = \mathbb{1}_{T_i \leq C_i}$

$$\delta_i = \begin{cases} 1, \text{ si } Z_i = T_i, \text{on observe la durée de survie(pas de censure)} \\ 0, \text{ si } Z_i = C_i, \text{ on observe } C_i, \text{ il ya une censure} \end{cases}$$

La loi de probabilité des données observées est définie comme suite :
La fonction de répartition :

$$\begin{aligned} F_{(Z;\delta)}(z,0) &= \mathbb{P}(Z \leq z, \delta = 0) \\ &= \mathbb{P}(min(T,C) \leq z, T > C) \\ &= \int_0^z \int_t^\infty dF_T(u) dF_C(t) \\ &= \int_0^z S_T(t) dF_C(t) \end{aligned}$$

donc la densité est

$$f_{Z;\delta}(z,0) = S_T(z) f_C(z)$$

Aussi

$$\begin{aligned} F_{(Z;\delta)}(z,1) &= \mathbb{P}(Z \leq z, \delta = 1) \\ &= \mathbb{P}(min(T,C) \leq z, T \leq C) \\ &= \mathbb{P}(T \leq z, T \leq C) \\ &= \int_0^z \int_t^\infty dF_C(u) dF_T(t) \\ &= \int_0^z S_C(t) dF_T(t) \end{aligned}$$

et

$$f_{Z;\delta}(z,1) = S_C(z) f_T(z)$$

Par conséquence : la densité pour $\delta_i = 0, 1, \forall i = 1, ..., n$ est

$$f_{(Z;\delta_i)}(z_i,\delta_i) = (S_{C_i}(z) f_{T_i}(z))^{\delta_i} (S_{T_i}(z) f_{C_i}(z))^{1-\delta_i}$$

Remarque 4. *Pour les fonctions de loi des censurées à gauche, on trouve la même fonction de densité déjà calculer pour la censure à droite.*

1.3.2 Pour le modèle de censure double

En et plus de T C, soit L une variable aléatoire positive et absolument continue de densité f_L, de fonction de survie $S_L$et fonction de répartition F_L.
Sous l'hypothèse que L est indépendante que T est C, on considère le cas de la censure aléatoire double, les observations sont les couples $(Z_1; A_1), ..., (Z_n; A_n)$ où $Z_i = max(min(T_i; C_i), L_i)$

$$A_i = \begin{cases} 0,\ si\ L_i \leqslant T_i \leqslant C_i, alors\ Z_i = T_i,\ \text{pas de censure} \\ 1,\ si\ C_i \leqslant T_i,\ alors\ Z_i = C_i;\ \text{il y a censure à droite} \\ 2,\ si\ T_i \leqslant L_i,\ alors\ Z_i = L_i;\ \text{il y a censure à gauche} \end{cases}$$

La loi de probabilité des données observées est définie comme suite :
La fonction de répartition :

$$\begin{aligned} F_{Z;A}(z,0) &= \mathbb{P}(Z \leqslant z, A = 0) \\ &= \mathbb{P}(max(min(T;C), L) \leqslant z, L \leqslant T \leqslant C) \\ &= \mathbb{P}(T \leq z, L \leq T \leq C) \\ &= \int_0^z \int_0^t dF_L(u) \int_t^\infty dF_C(u) dF_T(t) \\ &= \int_0^z F_L(t) S_C(t) dF_T(t) \end{aligned}$$

donc la densité est

$$f_{Z,A}(z,0) = F_L(t) S_C(z) f_T(z)$$

Aussi

$$\begin{aligned} F_{Z;A}(z,1) &= \mathbb{P}(Z \leq z, A = 1) \\ &= \mathbb{P}(max(min(T;C), L) \leq z, C < T) \\ &= \mathbb{P}(C \leq z, C < T) \\ &= \int_0^z \int_t^\infty dF_T(u) dF_C(t) \\ &= \int_0^z S_T(t)\, dF_C(t) \end{aligned}$$

d'ou

$$f_{(Z;A)}(z,1) = S_T(z)f_C(z).$$

et

$$\begin{aligned} F_{(Z,A)}(z,2) &= \mathbb{P}(Z \leq z, A = 2) \\ &= \mathbb{P}(max(min(T;C),L) \leq z, T < L) \\ &= \mathbb{P}(L \leq z, T < L) \\ &= \int_0^z \int_t^\infty dF_T(u)dF_L(t) \\ &= \int_0^z S_T(t)dF_L(t) \end{aligned}$$

donc

$$f_{(Z;A)}(z,2) = S_T(z)f_l(z)$$

On obtient la densité pour i= $0, 1, 2$

$$f_{(Z;\alpha)}(z,i) = (F_L(t)S_C(z)f_T(z))^{I_{A=0}(i)}(S_T(z)f_T(z))^{I_{A=1}(i)}(S_T(z)f_L(z))^{I_{A=2}(i)}$$

Dans ce modèle T est observée ssi $T \in [L;C]$ est une donnée censurée soit à droite soit à gauche mais pas les deux à la fois.

1.4 Estimation non paramétrique

L'approche non paramétrique est une estimation fonctionnelle, elle ne nécessite aucune hypothèse sur la loi de probabilité des observations ordonnées par ordre croissant des temps de participation. Cette approche est très efficace quand nous considérons un échantillon important d'observations. On intéresse à l'estimation de la fonction de survie ,la fonction du risque cumulé et celle de densité de probabilité.

1.4.1 L'estimation de la la fonction de survie

Estimateur de Kaplan-Meier

On s'intéresse à l'estimation de la fonction de survie de la v.a T censurée à droite par une v.a.C positive et indépendante de T. On observe un échantillon $T_1, ..., T_n$ de v.a indépendants représentant les durées de survie de même loi que T et C, $C_1, ..., C_n$ un échantillon représentant les temps de censure que l' on suppose indépendante des durées de survie.

Dans le modèle de censure à droite, on observe pas la durée de survie T ; mais on observe l'échantillon $\Big(Z_i = (X_i = T_i \bigwedge C_i, \delta_i = \mathbb{1}_{\{T_i \leqslant C_i\}})_{i=1,...,n}\Big)$ de n v.a. i.i.d. de même loi que $Z = \Big(X = T \bigwedge C, \delta = \mathbb{1}_{\{T \leqslant C\}}\Big)$.

Kaplan et Meier ont proposés un estimateur de la fonction de survie S. Cet estimateur vient de l'idée que, survivre après un temps t c'est être en vie juste avant t et ne pas mourir à l'instant t c'est à dire.

Pour $0 < t_1 < ... < t_i < t_{i+1} < ... < t_n = t$

$$\begin{aligned} S(t) &= \mathbb{P}(T > t) = \mathbb{P}(T > t, T > t_{n-1}) \\ &= \mathbb{P}(T > t | T > t_{n-1}) \mathbb{P}(T > t_{n-1}) \\ &\vdots \\ &= \mathbb{P}(T > t | T > t_{n-1}) ... \mathbb{P}(T > 1 | T > 0) \mathbb{P}(T > 0). \end{aligned}$$

Si on choisit les instants de conditionnement où il se produit un évènement t_i (panne,mort,...) on aura estimé des quantités de la forme :

$$\mathbb{P}(T > t_{(i)} | T > t_{(i-1)}) = p_i,$$

qui est la probabilité de survivre pendant l'intervalle de temps $I_i = [t_{(i-1)}, t_{(i)}]$ sachant qu'on était "vivant" au début de cet intervalle.

Notons r_i le nombre d'individus à risque de subir l'événement juste avant le temps $t_{(i)}$et m_i le nombre de décès à l'instant t_i. Alors la probabilité $1 - p_i$ de mourir dans l'intervalle $]t_{i-1}, t_i[$ sachant que l'on état vivant en t_{i-1}, notée

$$q_i = 1 - p_i$$

peut être estimée par q_i est la fréquence

$$\hat{q}_i = \frac{m_i}{r_i}$$

Soit (Z_i) avec $(1 \leqslant i \leqslant n)$, s'il n'y a pas ex-æquo, et dans le cas contraire $(1 \leqslant i \leqslant M)$, si :

$$\delta_i = \begin{cases} 1, \text{ c'est qu'il ya eu un évènement en } ; t_i, m_i = 1, \\ 0, \text{ c'est q'u'il ya eu une censure en } t_i, m_i = 0. \end{cases}$$

Par suite, on a $r_i = n - (i-1)$, on obtient l'estimateur de Kaplan-Meier est :

$$\hat{S}_{KM}(Z_i) = \prod_{j \leqslant i} \mathbb{P}(T > Z_j | T > Z_{j-1}) = \prod_{j \leqslant i} (1 - p_j),$$

alors

$$\hat{S}_{KM}(Z_i) = \prod_{j \leqslant i} (1 - \frac{m_j}{r_j})$$

Remarque 5. *L'estimateur de Kaplan-meier est une fonction en escaliers qui fait des sauts à chaque instant t_i. La valeur du saut dépend du nombre d'évènements au temps t_i et aussi du nombre de censures à ce temps là.*

Traitement des ex-aequo : Dans le d'existence des ex-aequo on n'a plus *mi* égale à *ti* mais le nombre d'évènements en t_i et aussi on n'a plus $r_i = n - (i-1)$. Dans ce cas on doit garder r_i et m_i et l'estimateur de Kaplan-Meier devient :

$$\hat{S}_{KM}(t) = \prod_{x_i \leqslant t} (1 - \frac{m_i}{r_i})^{\delta_i}$$

Exemple 1.15. *[9] On observe la durée de vie de 10 diodes exprimées en mois*
1 3 4$^+$ *5* 7^+ *8 9* 10^+ *11* 13^+

Temps t_i	r_i	mi	$\hat{S}_{KM}(t_i)$	intervalle
0	10	0	1	[0,1[
1	10	1	(1-1/10)$\hat{S}(0) = 0.900$	[1,3[
3	9	1	(1-1/9)$\hat{S}(1) = 0.800$	[3,5[
5	7	1	(1-1/7)$\hat{S}(3) = 0.686$	[5,8[
8	5	1	(1-1/5)$\hat{S}(5) = 0589$	[8,9[
9	4	1	(1-1/4)$\hat{S}(8) = 0.411$	[9,11[
11	2	1	(1-1/2)$\hat{S}(9) = 0.206$	[11,∞[

TABLE 1.1 – Exemple d'estimateur de Kaplan-Meier

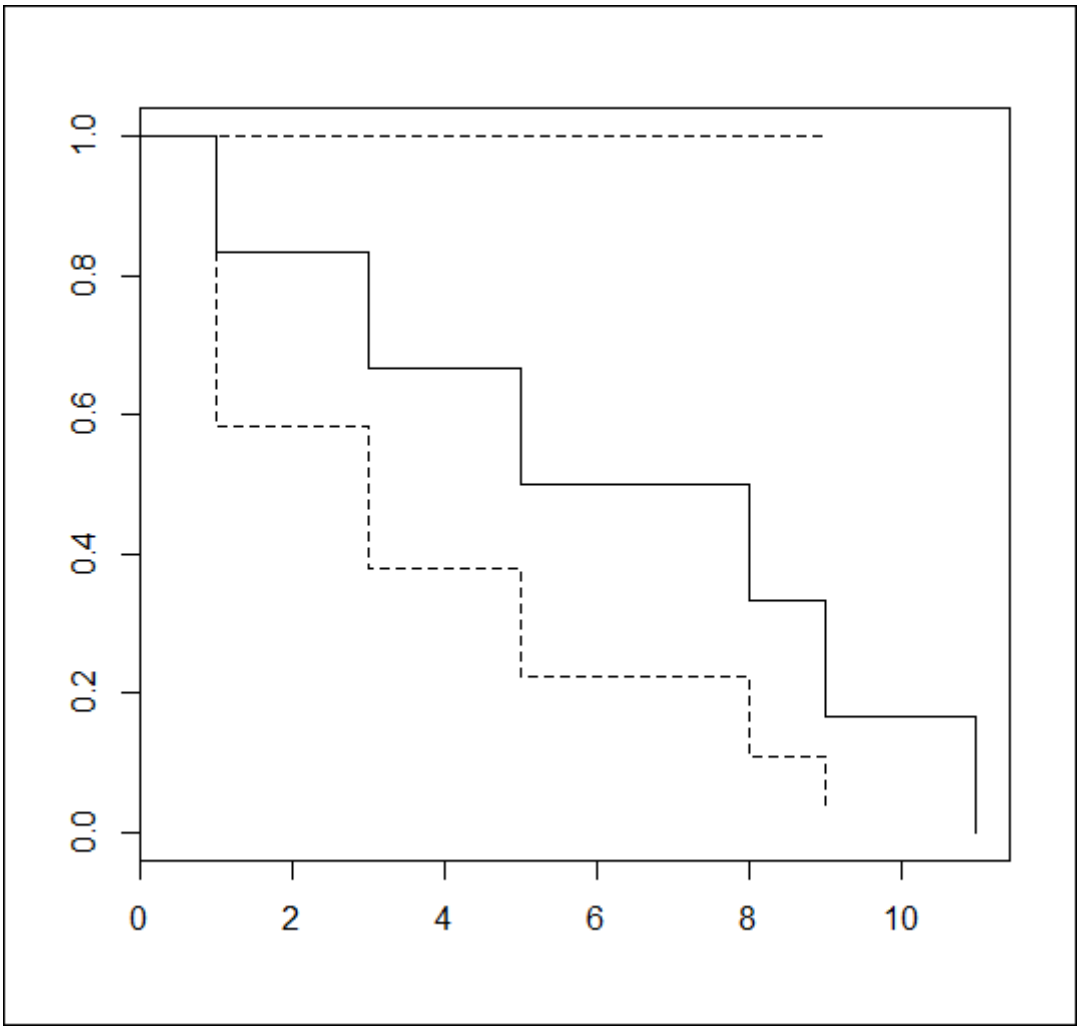

FIGURE 1.1 – Le graphe de la fonction de survie

1.4.2 Estimation à noyau de la densité et de taux de hasard

On a plusieurs méthodes classiques pour faire l'estimation non paramétrique, on lisse la fonction de hasard ; comme les estimateurs à noyaux, les estimateurs isotoniques, les estimateurs splines ou l'estimation par les k points les plus proches.

On s'intéresse dans ce paragraphe à l'estimation non paramétrique à noyau, des fonctions de densité f=(F') et du taux de hasard $\lambda = \frac{f}{1-F}$.

Un noyau K est une fonction borélienne, positive et intégrable, telle que $\int_{\mathbb{R}} K(x)dx = 1$. Ce type d'estimateur à été introduit, pour la densité, par Rosenblat(1956) puis amelioré par Parsen (1962).

Exemples de noyaux

Les estimateurs à noyaux d'une densité sont :

Noyau gaussien(normal) : $\forall u \in \mathbb{R}, K(u) = \frac{1}{\sqrt{2\Pi}} \exp -\frac{u^2}{2}$.

Noyau quartic : $\forall u \in \mathbb{R}, K(u) = \frac{15}{16}(1-u^2)\mathbb{1}_{|u|\leqslant 1}$.

Noyau d'Epanechnikov : $\forall u \in \mathbb{R}, K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{|u|\leqslant 1}$.

Noyau uniforme : $\forall u \in \mathbb{R}, K(u) = \frac{1}{2}\mathbb{1}_{|u|\leqslant 1}$.

Noyau triangulaire : $\forall u \in \mathbb{R}, K(u) = (1-|u|)\mathbb{1}_{|u|\leqslant 1}$.

Soit $X_1, X_2, .., X_n$ un échantillon de n v.a.i.i.d centrée à droite. C_1, C_2, C_n n v.a.i.i.d indépendante des v.a $X_1, X_2, ..., X_n$, on observe les couples $(Z_i, \delta_i)_{i=1,n}$ tels que pour tout i, $Z_i = min(X_i, C_i)$ et $\delta_i = \mathbb{1}_{\{X_i \leqslant C_i\}}$ l'estimateur à noyau de la fonction de densité, noté $\hat{f}_n$ est définie comme suit :

$$\hat{f}_n(t) = \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{t-x}{h_n}\right) d\hat{F}_n(x)$$

où $(h_n)_{n\geqslant 0}$ est une suite de nombres réels positifs, rappelé paramètre de lissage, ou fenêtre de l'estimateur avec h_n lorsque $n \to \infty$, K est un noyau. soit K : $\mathbb{R} \longrightarrow \mathbb{R}$ un noyau réel qui est une fonction de densité intégrable, symétrique , borné et intégrable 1, soit $\hat{S}_{KM}$ l'estimateur empirique naturel de la fonction de survie. Ce qui donne l'estimateur à noyau.

$$\begin{aligned}\hat{f}_n(t) &= -\frac{1}{h_n}\sum_{i=1}^{n} K(\frac{t-u}{h_n})\hat{S_{KM}}d(u) \\ &= \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{t_i - X_i}{h_n})\frac{\delta_i}{n-i+1}\hat{S}_{KM}(X_i^-)\end{aligned}$$

CHAPITRE

2

Propriétés asymptotiques

Ce chapitre est consacré à l'étude théorique des propriétés asymptotiques des estimateurs de Kaplan-Meier et ceux à noyau de la fonction de densité et de taux de hasard, on démontre la normalité asymptotique et la convergence complète dans les cas général de censure et la censure à droite.

2.1 Rappel sur la consistance et la convergence des estimateurs

Un estimateur $\hat{\theta}_n$ est de θ est dit

1. Faiblement consistant si $\forall x \in \mathbb{R}, \hat{\theta}_n(x) \xrightarrow[n\to\infty]{\mathbb{P}} \theta(x)$.
2. Faiblement et uniformément consistant si

$$\sup_{x\in\mathbb{R}} |\hat{\theta}_n(x) - \theta(x)| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

3. Fortement consistant si

$$\forall x \in \mathbb{R}, \hat{\theta}_n(x) \xrightarrow[n\to\infty]{p.s} \theta(x).$$

4. Fortement et uniformément consistant si

$$\sup_{x\in\mathbb{R}} |\hat{\theta}_n(x) - \theta(x)| \xrightarrow[n\to\infty]{p.s} 0.$$

5. Asymptotiquement sans biais si

$$\forall x \in \mathbb{R}, \lim_{n\to\infty} E(\hat{\theta}_n(x)) = \theta(x).$$

6. Asymptotiquement et uniformément sans biais si

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} E\left[\hat{\theta}_n(x) - \theta(x)\right] = 0.$$

7. Convergence en moyenne quadratique(m.q) si

$$\lim_{n\to\infty} E\left[\hat{\theta}_n(x) - \theta(x)\right]^2 = 0.$$

8. La convergence presque complète (p.c)
   La suite des variables aléatoires X lorsque $n \longrightarrow \infty$ si

$$\forall \varepsilon > 0, \sum \mathbb{P}(|X_n - X| > \varepsilon) < \infty.$$

   On dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ vers X est d'ordre (u_n) si

$$\forall \varepsilon_0 > 0, \sum \mathbb{P}(|X_n - X| > \varepsilon_0 u_n) < \infty.$$

2.2 Propriétés asymptotiques de l'estimateur Kaplan-Meier

L'estimateur de Kaplan-Meier possède un certain nombre de bonnes propriétés qui en font la généralisation naturelle de l'estimateur empirique de la fonction de répartition en présence de censure, l'estimateur de Kaplan-Meier est l'unique estimateur cohérent de la fonction de survie.

Propriétés

1) Si aucune donnée n'est censurée,i.e $T_i = X_i$ pour $i = 1, ..., n$ alors

$$\hat{S}_{KM}(t) = \hat{S}_n(t) = 1 - F_n(t),$$

où

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{T_i \leq t},$$

est la fonction de répartition empirique.

2) Soit $S_C(t) = \mathbb{P}(C > t)$ et $\tau_x = \inf\{x \geq 0 | S(x)S_C(x) = 0\}$:

— Si S et S_C n'ont pas de points de discontinuités en commun, on a pour tout $\tau < \tau_x$

$$\sup_{0 \leq t \leq \tau} |\hat{S}_{KM}(t) - S(t)| \xrightarrow[n \longrightarrow \infty]{ps} 0.$$

— En tout point $t \in [0, \tau]$

$$\sqrt{n}(\hat{S}_{KM}(t) - S(t) \xrightarrow[n \longrightarrow \infty]{} N(0, V^2(t)),$$

avec

$$V^2(t) = -S^2(t) \int_0^t \frac{S(u)}{S^2(u)S_c(u)} du$$

C'est a dire l'estimateur de Kaplan-Meier est un estimateur fortement consistant et asymptotiquement gaussien de $S(t)$.

3) Variance de l'estimateur de Kaplan-Meier est

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \gamma(t)^2,$$

avec

$$\gamma(t) = \sqrt{\sum_{T_i \leq t} \frac{m_i}{r_i(r_i - m_i)}}.$$

preuve (voir [4])

2.3 Propriétés asymptotiques de l'estimateur à noyau de la densité (cas général)

soit X une durée de vie positive avec la fonction de distribution F et la fonction de densité f inconnues. Nous supposons que X peut être censuré donc au lieu d'observer un échantillon de X,

nous avons à disposition un échantillon $(Z_1, \delta_1), ..., (Z_n, \delta_n)$ de copies indépendantes du couple de variables aléatoires réelles, où Z = X si et seulement si l'indicateur de la censure prend la valeur 0. En empruntant l'idée à l'estimation de type à noyau, nous estimons f par

$$\hat{f}_n(t) = \frac{1}{h_n} \int K\left(\frac{t-y}{h_n}\right) d\hat{F}_n(y), \tag{2.1}$$

où $\hat{F}_n$ est un estimateur de F, continus droits à variation bornée et satisfaisant certaines conditions supplémentaires à préciser plus tard (voir H_4), K est une fonction du noyau et h_n est la fenêtre de l'estimateur.

Posons $g(t) = \mathbb{P}(\delta = 0|X = t)$, les hypothèses suivantes seront nécessaires pour énoncer la normalité asymptotique de $\hat{f}_n$ pour un x fixe :

H_1 : f est différentiable en x ;

H_2 : g est continue en x ;

H_3 : $\exists a > 0; \inf_{t\in[x-a,x+a]} g(t) > 0;$;

H_4 : F_n est une fonction étagée avec seulement des sauts possibles à $(Z_i)_{1\leqslant i\leqslant n}$ et telle que

$$\sqrt{n} \max_{i, Z_i \in]x-a, x+a[} |n\Delta F_n(Z_i) 1_{\delta_i=0} - 1_{\delta_i=0}(g(Z_i))| = O_p(1);$$

et

$$\sum_{i=1}^{n} \Delta F_n(Z_i) 1_{\delta_i \neq 0} = o_p\left(\sqrt{\frac{h_n}{n}}\right) \ que\ n \to \infty.$$

Si nous avons à disposition un échantillon complet $(X_i)_{1\leqslant i\leqslant n}$ pour X ensuite

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right),$$

comme

$$E\left(\left(\frac{x-Z_i}{h_n}\right)\frac{1_{\delta_i=0}}{g(Z_i)}\right) = E\left(K\left(\frac{x-X_i}{h_n}\right)\right),$$

il semble naturel de remplacer

$$K\left(\frac{x-X_i}{h_n}\right) \text{ par } K\left(\frac{x-Z_i}{h_n}\frac{1_{\delta_i=0}}{g(Z_i)}\right).$$

Chaque fois que X est censuré, nous introduisons :

$$\tilde{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\frac{1_{\{\delta_i=0\}}}{g(Z_i)},$$

qui sera asymptotiquement normale (comme le montre la preuve du théorème(2.1)).

Les hypothèses H_3 et H_4 : sont satisfaites pour des données censurées à droites et doublement censurées sous approprié. Nous supposons en plus une hypothèse standard, dans le cadre non paramétrique, concernant K et h_n

H_5 : K est la fonction de densité bornée par un réel $M > 0$ et à un support compact.

H_6 : $nh_n \to \infty$ et $nh_n^3 \to 0$, comme $n \to \infty$.

Théorème 2.1. *Sous H_1 et H_6 on a*

$$\sqrt{nh_n}(f_n(x) - f(x)) \xrightarrow{D} N\left(0, \frac{f(x)}{g(x)\int K^2(z)dz}\right) \text{ lorsque } n \to \infty,$$

Preuve. Rappelons que

$$\tilde{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}.$$

■

remarquons que sous H_3, H_5 et H_6 on peut supposer que $g(Z_i) \neq 0$ pour n suffisamment grand nous nous faisons de la décomposition suivante

$$f_n(x) - f(x) = (f_n(x) - \tilde{f}_n(x)) + (\tilde{f}_n(x) - E\tilde{f}_n(x)) + (E\tilde{f}_n(x) - f(x))$$

nous traitons chaque terme à tour de rôle

étape 1 : étude de terme $\tilde{f}_n(x) - E\tilde{f}_n(x)$

On a $\tilde{f}_n(x) - E\tilde{f}_n(x) = \sum_{i=1}^{n} \xi_{n,i}$ où

$$\xi_{n,i} = \frac{1}{nh_n}K\left(\frac{x-Z_i}{h_n}\right)\frac{1_{\{\delta_i=0\}}}{g(Z_i)} - \frac{1}{nh_n}E\left(K\left(\frac{x-Z_i}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right)$$

pour appliquer le théorème de Lindberg (voir Billingskly 1995) à $\xi_{n,i}$ nous devons montrer que pour tout $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{1}{Var(\tilde{f}_n(x))}\sum_{i=1}^{n} E\left(\xi_{n,i}^2 \mathbb{1}_{|\xi_{n,i}|>\varepsilon\sqrt{Var(\tilde{f}_n(x))}}\right) = 0. \tag{2.2}$$

Sous H_5 et H_6 on considère que $Z_i \in [x-a, x+a]$ pour n est grand, on obtient

$$|\xi_{n,i}| \leqslant \frac{2M}{\inf_{t\in[x-a,x+a]} g(t)} \times \frac{1}{nh_n},$$

où

$$\begin{aligned}\frac{1}{Var(\tilde{f}_n(x))}\sum_{i=1}^{n} E\left(\xi_{n,i}^2 \mathbb{1}_{\{|\xi_{n,i}|>\varepsilon\sqrt{Var(\tilde{f}_n(x))}\}}\right) &\leqslant \frac{1}{Var(\tilde{f}_n(x))} \times \frac{4M^2}{(\inf_{t\in[x-a,x+a]} g(x))^2} \times \frac{1}{n^2h_n^2} \\ &\times n \times \mathbb{P}\left(|\xi_{n,1}| > \varepsilon\sqrt{Var(\tilde{f}_n(x))}\right) \\ &\leqslant \frac{1}{Var(\tilde{f}_n(x))} \times \frac{4M^2}{(\inf_{t\in[x-a,x+a]} g(t))^2} \\ &\times \frac{1}{nh_n^2} \times \frac{E\xi_{n,1}^2}{\varepsilon^2(Var(\tilde{f}_n(x)))}\end{aligned}$$

A partir de l'inégalité de Chebyshev-Bienaymé (voir Billingsley 1995), on a

$$\begin{aligned}\frac{1}{Var(\tilde{f}_n(x))}\sum_{i=1}^{n} E\left(\xi_{n,i}^2 \mathbb{1}_{\{|\xi_{n,i}|\}>\varepsilon\sqrt{Var(\tilde{f}_n(x))}}\right) &\leqslant \frac{4M^2}{\varepsilon^2(\inf_{t\in[x-a,x+a]} g(t))^2} \times \frac{1}{n^2h_n^2 Var(\tilde{f}_n(x))} \\ &= \frac{4M^2}{\varepsilon^2(\inf_{t\in[x-a,x+a]} g(t))^2} \times \frac{1}{nh_n \times \frac{1}{h_n} var\left(K(\frac{x-Z_i}{h_n})\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right)}.\end{aligned} \tag{2.3}$$

Il reste à calculer

$$\lim_{n\to\infty} \frac{1}{h_n} Var\left(K(\frac{x-Z_1}{h_n})\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_1)}\right).$$

On calcule la limite suivante

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{h_n} E\left(K(\frac{x-Z_1}{h_n})\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_1)}\right)^2 &= \lim_{n\to\infty} \frac{1}{h_n} E\left(K^2\left(\frac{x-X_1}{h_n}\right)\frac{1}{g^2(X_1)} E(\mathbb{1}_{\{\delta_i=0\}}|X_1)\right) \\ &= \lim_{n\to\infty} \frac{1}{h_n}\int K^2\left(\frac{x-u}{h_n}\right)\varphi(u)du \text{ où } \varphi(u) = \frac{f(u)}{g(u)}\mathbb{1}_{[x-a,x+a]}(u) \\ &= \frac{f(x)}{g(x)}\int K^2(z)dz\end{aligned} \tag{2.4}$$

d'après le théorème de Bochner, grâce à H_1, H_3, H_5 et H_6

par un argument similaire, on obtient de même que

$$\lim_{n\to\infty} \frac{1}{h_n} E\left(K\left(\frac{x-Z_1}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_1)}\right) = f(x)$$

On trouve

$$\lim_{n\to\infty}\frac{1}{h_n}\left(E\left(K\left(\frac{x-Z_1}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_1)}\right)\right)^2=0$$

Donc

$$\lim_{n\to\infty}\frac{1}{h_n}Var\left(K\left(\frac{x-Z_1}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_1)}\right)=\frac{f(x)}{g(x)}\int K^2(z)dz \tag{2.5}$$

d'après l'équation

$$\frac{\tilde{f}_n(x)-E\tilde{f}_n(x)}{\sqrt{var(\tilde{f}_n(x))}}\xrightarrow{D}N(0,1)$$

et

$$\sqrt{nh_n}\left(\tilde{f}_n(x)-E\tilde{f}_n(x)\right)\xrightarrow{D}N\left(0,\frac{f(x)}{g(x)}\int K^2(z)dz\right) \tag{2.6}$$

étape 2 : étude de terme $E\tilde{f}_n(x)-f(x)$, on trouve

$$\begin{aligned}E\tilde{f}_n(x)&=\frac{1}{h_n}E\left(K\left(\frac{x-Z_i}{h_n}\right)\frac{\mathbb{1}_{\delta_i=0}}{g(Z_1)}\right)\\&=\frac{1}{h_n}E\left(K\left(\frac{x-u}{h_n}\right)f(u)du\right)\\&=\int K(z)f(x-h_nz)dz\end{aligned}$$

Donc

$$E\tilde{f}_n(x)-f(x)=\int K(z)(f(x-h_n)-f(x))dz$$

Sous H_1, H_5 et H_6 et l'utilisation de la formule de Taylor-Young, on trouve

$$E\tilde{f}_n(x)-f(x)=O(h_n), n\to\infty$$

en suite, on utilise H_6 qui donne

$$\lim_{n\to\infty}\sqrt{nh_n}(E\tilde{f}_n(x)-f(x))=0.$$

étape 3 : étude de terme $f_n(x) - \tilde{f}_n(x)$

$$\begin{aligned}
\sqrt{nh_n}\left|f_n(x) - \tilde{f}_n(x)\right| &= \sqrt{nh_n}\left|\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\Delta F_n(Z_i) - \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right| \\
&\leqslant \sqrt{nh_n}\left|\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i=0\}} - \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right| \\
&+ \frac{\sqrt{nh_n}}{h_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i\neq 0\}} \\
&\leqslant \frac{\sqrt{nh_n}}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\left|n\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i=0\}} - \frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right| \\
&+ \sqrt{\frac{n}{h_n}}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i\neq 0\}} \\
&\leqslant \sqrt{nh_n}\max_{i,Z_i\in[x-a,x+a]}\left|n\Delta F_n(Z_i)\mathbb{1}\{\delta_i=0\} - \frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right| \\
&\times \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) + M\sqrt{\frac{n}{h_n}}\sum_{i=1}^{n}\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i\neq 0\}}
\end{aligned}$$

Sous H_4 et H_5, si n $\to \infty$:

$$\sqrt{nh_n}\max_{i,Z_i\in[x-a,x+a]}\left|n\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i=0\}} - \frac{\mathbb{1}_{\{\delta_i=0\}}}{g(Z_i)}\right| \xrightarrow{p} 0$$

et

$$\sqrt{\frac{n}{h_n}}\sum_{i=1}^{n}\Delta F_n(Z_i)\mathbb{1}_{\{\delta_i\neq 0\}} \xrightarrow{p} O$$

de plus il est prouvé dans Parzen (1962) que lorsque n tend vers l'infini

$$\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \xrightarrow{p} f(x)$$

D'où

$$\sqrt{nh_n}\left|f_n(x) - \tilde{f}_n(x)\right| \xrightarrow{p} 0 \text{ comme } n \longrightarrow \infty.$$

En combinant cela avec les équations (2.6) (2.7) nous obtenons le résultat que nous cherchons.

2.4 Les propriétés asymptotiques dans le cas de la censure à droite

2.4.1 Estimateur de Kaplan-Meier

Soit T une variable aléatoire positive, censurée à droite par une variable aléatoire C positive et indépendante de T. Nous observons l'échantillon $(X_i = T_i \bigwedge C_i, \delta_i = \mathbb{1}_{\{T_i \leqslant C_i\}})_{1 \leqslant i \leqslant n}$ de n couples de variables aléatoires i.i.d et de même loi que $(X = T \bigwedge C, \delta = \mathbb{1}_{T \leqslant C})$, et nous notons F,G et H les fonctions de répartition respectives de T, C et X, S,$\bar{G}, \bar{H}$ leur fonctions de survie respectives, et $(Z_j)_{1 \leqslant j \leqslant n}$ les valeurs distinctes des $(X_j)_{1 \leqslant i \leqslant n}$ rangées dans l'ordre croissant. L'estimateur de Kaplan-Meier de S est donné pour tout $t \in \mathbb{R}$ par :

$$S_n(t) = \prod_{j | Z_j \leqslant t} \left(1 - \frac{M(Z_j)}{R(Z_j)}\right),$$

où $M(Z_j) = \sum_{i=1}^{n} \delta_i \mathbb{1}_{\{X_i = Z_j\}}$ c'est le nombre de morts exactes au j$^{\text{ème}}$ instant
et $R(Z_j) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \geqslant Z_j\}}$ est le nombre d'individus à risque juste avant le j$^{\text{ème}}$ instant.
FÖldes A.et al. ont trouvé en 1980 un taux de convergence presque complète uniforme de S_n de l'ordre de $\sqrt{\frac{\log n}{\sqrt{n}}}$, mais la convergence n'a lieu qu'avant le plus petit des temps terminaux des supports de F et de G (voir[Foldes et al. 1980] théorème 2.2 page 237).

Puis en imposant que F et G sont continues, Foldes A. et Rejto L. ont amélioré le taux de convergence qui est passé à l'ordre de $\sqrt{\frac{\log n}{n}}$ (voir[Foldes et Rejto 1981] preuve du théorème 3.2). Dans Boukeloua[], il ont retrouvé ce même taux sans exiger la continuité de F ni celle de G. Et pour cela, nous avons besoin du lemme suivant dont la preuve est donnée dans [Shorak et Wellner 1986](lemme 1 page 302).

Lemme 2.2. *Si A et B sont deux fonctions croissantes et continues sur* $[0, \infty[$ *avec A(t)=B(t) pour* $t < 0$ *et* $\Delta A \leqslant 1$ *et* $\Delta B \leqslant 1$ *sur* $[0, +\infty[$ *et si* $\theta_\beta = \inf t \in \mathbb{R} | B(t) = +\infty$ *alors la seule solution local bornée Z de l'équation*

$$Z(t) = Z(0) - \int \frac{Z(x^-)}{1 - \Delta B(x)} d(A(x) - B(x))$$

sur $[0, \theta_\beta]$ *est donnée par*

$$Z(t) = Z(0) \exp(B^c - A^c(t)) \frac{\prod_{0 \leqslant x \leqslant t}(1 - \Delta A(x))}{\prod_{0 \leqslant x \leqslant t}(1 - \Delta B(x))}$$

Théorème 2.3. *pour tout* $\theta \in [0, T_H[$*, on a :*

$$\sup_{t\leqslant\theta} |S_n(t) - S(t)| \xrightarrow{p.co} 0$$

et

$$\sup_{t\leqslant\theta} |S_n(t) - S(t)| = O_{p.co}\left(\sqrt{\frac{\log n}{n}}\right)$$

Preuve. Soit $\theta \in]0, T_H[$, pour tout t¡ 0, on a : $S_n(t) - S(t) = 0$ donc $\sup_{t\leqslant\theta} |S_n(t) - S(t)| = \sup_{0\leqslant\leqslant t\theta\leqslant} |S_n(t) - S(t)|$

Alors soit $t \in [0, \theta]$, on pose $N_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\leqslant t, \delta_i=1\}}$ et $Y_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\geqslant t\}}$, $N_n(t)$ et $Y_n(t)$ sont les lois empiriques associées respectivement à $H_1(t) = \mathbb{P}(X \leqslant t, \delta = 1)$ et à $\bar{H}(t^-)$

par ailleurs, la fonction de hasard cumulé de T est donnée par $\Lambda(t) = \int_{]0,t[} \frac{dF(x)}{S(x^-)} = \int_{]0,t[} \frac{dH_1}{H(x^-)}$ qui estimée par l'estimation de Nelson-Aalen suivant : $\Lambda_n(t) = \int_{]0,t[} \frac{dN_n(x)}{Y_n(x)}$

De plus, nous avons ; $S(t) = 1 - \mathbb{P}(T \leqslant t) = 1 - \int_{]0,t[} dF(x) = 1 - \int_{]0,t[} S(x^-)d\Lambda(x)$

donc le lemme (1) donne

$$S(t) = \prod_{x\leqslant t}(1 - \Delta\Lambda(x)) \exp(-\Lambda^c(t)) \tag{2.7}$$

D'autre part, nous avons : $\Lambda_n(t) = \int_{]0,t[} \sum_{j|Z_j\leqslant t} \frac{\Delta N_n Z_j}{Y_n(Z_j)} = \sum \frac{M(Z_j)}{R(Z_j)} \Rightarrow \Delta\Lambda_n(Z_j) = \frac{M(Z_j)}{R(Z_j)}$ d'où

$$S_n(t) = \prod_{x\leqslant t}(1 - \Delta\Lambda_n(Z_j)) \tag{2.8}$$

Les relations (2.7) et (2.8) montrent, d'après le lemme 1, que $\frac{S_n(t)}{S(t)}$ vérifie

$$\begin{aligned}
\frac{S_n(t)}{S(t)} &= 1 - \int_{]0,t[} \frac{S_n(x^-)d(\Lambda_n(x) - \Lambda(x))}{S(x^-)(1 - \Delta\Lambda(x))} \\
&\Rightarrow S_n(t) - S(t) = -S(t) \int_{]0,t[} \frac{S_n(x^-)d(\Lambda_n(x) - \Lambda(x))}{S(x^-)(1 - \Delta\Lambda(x))} \\
&\Rightarrow |S_n(t) - S(t)| \leqslant \left| \int_{]0,t[} \frac{S_n(x^-)}{S(x^-)} dK_n(x) \right| \text{ où } K_n(t) = \int_{]0,t[} \frac{d(\Lambda_n(x) - \Lambda(x))}{1 - \Delta\Lambda(x)}
\end{aligned}$$

Nous en déduisons, en appliquant la formule d'intégration par partie que

$$
\begin{aligned}
|S_n(t)-S(t)| &\leqslant \frac{S_n(t)}{S(t)}\,|K_n(t)| + \left|\int_{]0,t[} K_n(x)d\left(\frac{S_n(x)}{S(x)}\right)\right| \\
&\leqslant \frac{1}{S(\theta)}\sup_{0\leqslant u\leqslant\theta}|K_n(u)| + \left|\int_{]0,t]} K_n(x)S_n(x^-)d(\frac{1}{S(x)})\right| + \left|\int_{]0,t]}\frac{K_n(x)}{S(x)}dS_n(x)\right| \\
&\leqslant \frac{1}{S(\theta)}\sup_{0\leqslant u\leqslant\theta}|K_n(u)| + \sup_{0\leqslant u\leqslant\theta}|K_n(u)|\left(\frac{1}{S(t)}-1\right) + \frac{1}{S(\theta)}\sup_{0\leqslant u\leqslant\theta}|K_n(u)||S_n(t)-1| \\
&\leqslant \left(\frac{3}{S(\theta)-1}\right)\sup_{0\leqslant u\leqslant\theta}|K_n(u)| \\
\text{or} S(x) = S(x^-)(1-\Delta\Lambda(x)) &\Rightarrow \frac{1}{1-\Delta\Lambda(x)} = \frac{S(x^-)}{S(x)} = 1-\frac{\Delta S(x)}{S(x)} \\
&\Rightarrow K_n(u) = \int_{]0,u]}(1-\frac{\Delta S(x)}{S(x)}d(\Lambda_n(x)-\Lambda(x)) \\
&\Rightarrow |K_n(u)| \leqslant \left|\int_{]0,u]} d\Lambda_n(x)-\Lambda(x)\right| + \left|\int_{]0,u]}\frac{\Delta S(x)}{S(x)}d(\Delta\Lambda_n(x)-\Delta\Lambda(x))\right| \\
&\leqslant |\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u)-\Lambda(u)| + \sum_{\substack{x\in]0,u]\\ \Delta S(x)>0}}\left|\frac{\Delta S(x)}{S(x)}\right||\Delta\Lambda_n(u)-\Delta\Lambda(u)| \\
&\leqslant |\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u)-\Lambda(u)| + \frac{1}{S(\theta)}\sup_{0\leqslant u\leqslant\theta}|\Delta\Lambda_n(u)-\Delta\Lambda(u)|\sum_{\substack{x\in]0,u]\\ \Delta S(x)>0}}|\Delta S(x)| \\
&\leqslant \left(1+\frac{1}{S(\theta)}\right)\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u)-\Lambda(u)| + \frac{1}{S(\theta)}\sup_{0\leqslant u\leqslant\theta}\left|\Lambda_n(u^-)-\Lambda(u^-)\right|
\end{aligned}
$$

et ceci pour tout $u\in[0,\theta]$, d'où

$\sup_{0\leqslant u\leqslant\theta}|K_n(u)| \leqslant (1+\frac{1}{S(\theta)})\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u)-\Lambda(u)| + \frac{1}{S(\theta)}\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u^-)-\Lambda(u^-)|$

$\Rightarrow |S_n(t)-S(t)| \leqslant \frac{(3-S(\theta))(1+S(\theta))}{S(\theta)^2}\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u)-\Lambda(u)| + \frac{3-S(\theta)}{S(\theta)^2}\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(u^-)-\Lambda(u^-)|, \forall t\in$ $[0,\theta]$ donc :

$$
\sup_{0\leqslant u\leqslant\theta}|S_n(t)-S(t)| \leqslant \frac{(3-S(\theta))(1+S(\theta))}{S(\theta)^2}\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(t)-\Lambda(t)| + \frac{3-S(\theta)}{S(\theta)^2}\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(t^-)\Lambda(t^-)| \tag{2.9}
$$

De plus nous avons :

$$
\begin{aligned}
|\Lambda_n(t)-\Lambda(t)| &= \left|\int_{]0,t]}\frac{dN_n(x)}{Y_n(x)} - \int_{]0,t]}\frac{dH_1}{\bar{H}(x^-)}\right| \\
&= \left|\int_{]0,t]}\frac{dN_n(x)}{Y_n(x)} + \int_{]0,t]}\frac{dH_1}{\bar{H}(x^-)} + \int_{]0,t]}\frac{dN_n(x)}{\bar{H}(x^-)} - \int_{]0,t]}\frac{dN_n(x)}{\bar{H}(x^-)}\right|
\end{aligned}
$$

$$\leqslant \left|\int_{]0,t]}\left(\frac{1}{Y_n(x)}-\frac{1}{\bar{H}(x^-)}\right)dN_n(x)\right|+\left|\int_{]0,t]}\frac{1}{\bar{H}(x^-)}d\left(N_n(x)-H_1(x)\right)\right|$$
$$=\left|\int_{]0,t]}\frac{\hat{H}(x^-)-Y_n(x)}{Y_n(x)\times\bar{H}(x^-)}dN_n(x)\right|+\left|\frac{1}{\bar{H}(x^-)}d\left(N_n(x)-H_1(x)\right)\right|$$
$$\leqslant \sup_{0\leqslant u\leqslant\theta}\left|\frac{\bar{H}(u^-)-Y_n(u)}{Y_n(u)\times\bar{H}(u^-)}\right|N_n(t)+\left|\frac{N_n(t)-H_1(t)}{\bar{H}(x)}\right|$$
$$\leqslant \frac{1}{Y_n(\theta)\times\bar{H}(\theta^-)}\sup_{0\leqslant u\leqslant\theta}\left|Y_n(u)-\bar{H}(u^-)\right|+\frac{1}{\bar{H}(\theta)}\sup_{0\leqslant u\leqslant\theta}|N_n(u)-H_1(u)|+\sup_{0\leqslant u\leqslant\theta}\left|N_n(u)-\bar{H}_1(u)\right|\left(\frac{1}{\bar{H}(t)}\right.$$
$$\leqslant \frac{1}{Y_n(\theta)\times\bar{H}(\theta^-)}\sup|Y_n(u)-\bar{H}(u^-)|+\left(\frac{2}{\bar{H}(\theta)}-1\right)\sup_{0\leqslant u\leqslant\theta}|N_n(u)-\bar{H}_1(u)|$$

et comme $Y_n(\theta)\xrightarrow{p.s}\bar{H}(\theta^-)\neq 0$ on a $\frac{1}{Y_n(\theta)}\xrightarrow{p.s}\frac{1}{\bar{H}(\theta^-)}\Rightarrow\exists C(\theta)>0/\frac{1}{Y_n(\theta)}\leqslant C(\theta)$ p.s. et par conséquent :

$|\Lambda_n(t)-\Lambda(t)|\leqslant\frac{C(\theta)}{\bar{H}(\theta^-)}\sup_{0\leqslant u\leqslant\theta}\left|Y_n(u)-\bar{H}(u^-)\right|+\left(\frac{2}{\bar{H}(\theta^-)}-1\right)\sup_{0\leqslant u\leqslant\theta}|N_n(u)-H_1(u)|$ p.s.

et ceci pour tout $t\in[0,\theta]$, donc :

$\sup_{0\leqslant u\leqslant\theta}|\Lambda_n(t)-\Lambda(t)|\leqslant\frac{C(\theta)}{\bar{H}(\theta^-)}\sup_{0\leqslant u\leqslant\theta}\left|Y_n(t)-\bar{H}(t^-)\right|+\frac{1}{\bar{H}(\theta^-)}\sup_{0\leqslant u\leqslant\theta}\left|N_n(t^-)-\bar{H}_1(t^-)\right|+\left(\frac{1}{\bar{H}(\theta^-)}-1\right)\sup_{0\leqslant u\leqslant}$

Nous pouvons montrer de la même façon que

$$\sup_{0\leqslant t\leqslant\theta}|\Lambda_n(t^-)-\Lambda(t^-)|\leqslant\frac{C(\theta)}{\bar{H}(\theta^-)}\sup_{0\leqslant t\leqslant\theta}|Y_n(t)-\bar{H}(t^-)|+\frac{1}{\bar{H}(\theta^-)}\sup_{0\leqslant t\leqslant\theta}|N_n(t^-)-H_1(t^-)|$$
$$+\left(\frac{1}{\bar{H}(\theta^-)}-1\right)\sup_{0\leqslant t\leqslant\theta}|N_n(t)-H_1(t)|$$

Il est s'ensuite alors :

$\sup_{0\leqslant u\leqslant\theta}|S_n(t)-S(t)|<\alpha(\theta)\sup_{0\leqslant u\leqslant\theta}|Y_n(t)-\bar{H}(t^-)|+\beta(\theta)\sup_{0\leqslant u\leqslant\theta}|N_n(t)-H_1(t)|$
$+\gamma(\theta)\sup_{0\leqslant u\leqslant\theta}|N_n(t^-)-H_1(t^-)|$ p.s.

où : $\alpha(\theta)=\frac{C(\theta)(3-S(\theta))(2+S(\theta))}{\bar{H}(\theta^-)S(\theta)^2},\beta(\theta)=\left(\frac{2}{\bar{H}(\theta)}-1\right)\left(\frac{3-S(\theta)(1+S(\theta))}{S(\theta)^2}\right)+\left(\frac{1}{\bar{H}(\theta^-)}-1\right)\frac{3-S(\theta)}{S(\theta)^2}$

et $\gamma(\theta)=\frac{3-S(\theta)}{\bar{H}(\theta^-)S(\theta)^2}$

Donc

$$\mathbb{P}\left(\sup_{0\leqslant u\leqslant\theta}|S_n(t)-S(t)|>5(\alpha(\theta)+\beta(\theta)+\gamma(\theta)\sqrt{\frac{\log n}{n}}\right)\leqslant\mathbb{P}\left(\sup_{0\leqslant u\leqslant\theta}\left|Y_n(t)-\bar{H}(t^-)\right|>5\sqrt{\frac{\log n}{n}}\right)$$
$$+\mathbb{P}\left(\sup_{0\leqslant u\leqslant\theta}|N_n(t)-H_1(t)|>5\sqrt{\frac{\log n}{n}}\right)$$
$$+\mathbb{P}\left(\sup_{0\leqslant u\leqslant\theta}\left|N_n(t^-)-H_1(t^-\right|>5\sqrt{\frac{\log n}{n}}\right)\tag{2.10}$$

et ceci tout $n \geqslant 2$, d'où :

$$\sum_{n\geqslant 2}\mathbb{P}\left(\sup_{0\leqslant u\leqslant\theta}|S_n(t)-S(t)|>5(\alpha(\theta)+\beta(\theta)+\gamma(\theta))\sqrt{\frac{\log n}{n}}\right)\leqslant\sum_{n\geqslant 2}\mathbb{P}\left(\sup_{0\leqslant u\leqslant\theta}|Y_n(t)-\bar{H}(t^-)|>5\sqrt{\frac{\log n}{n}}\right)$$
$$+\sum_{n\geqslant 2}\mathbb{P}\left(\sup|N_n(t)-H_1(t)|>5\sqrt{\frac{\log n}{n}}\right)$$
$$+\sum_{n\geqslant 2}\left(\sup_{0\leqslant u\leqslant\theta}\left|N_n(t^-)-H_1(t^-)\right|>5\sqrt{\frac{\log n}{n}}\right)$$

La relation (), la remarque la suivant et le fait que $\sup_{0\leqslant u\leqslant\theta}|S_n(t)-S(t)|=\sup_{t\leqslant\theta}|S_n(t)-S(t)|$ entrainent que :

$$\sum_{n\geqslant 2}\mathbb{P}\left(\sup_{t\leqslant\theta}|S_n(t)-S(t)|>5\left(\alpha(\theta)+\beta(\theta)+\gamma(\theta)\sqrt{\frac{\log n}{n}}\right)\right)<\infty$$

Avec $\alpha(\theta)+\beta(\theta)+\gamma(\theta)>0$ ce qui achève la démonstration. ■

2.4.2 Estimateur à noyau de la densité

Soit T une variable aléatoire positive censuré à droite par une variable aléatoire C positive et indépendante de T .

Notons F,G,H les fonctions de répartition respective de T,C et X

Supposons f=F' est la de probabilité de T. Étendant le cas des données complètes. Foldes et al(1981) ont proposé d'estimé f comme suit

$$\hat{f}_n(x)=\frac{1}{h_n}\int k(\frac{x-y}{h_n})dF_n(y) \tag{2.11}$$

où $F_n=1-S_n$.

Il est clair que nous retrouvons l'estimateur de Parzen-Rosenblatt si les données sont complètes Introduisons les quantités suivantes

$$\bar{f}_n(x)=\frac{1}{h_n}\int k(\frac{x-y}{h_n})dF_n(y)=\frac{1}{h_n}\int k(\frac{x-y}{h_n})f(y)dy$$

et

$$\tilde{K}(y)=K(\frac{x-y}{h_n})$$

Remarquons que $\bar{f}_n(x) = Ef_n(x)$ dans le cas des données complètes mais ceci n'est pas dans les données censurées. Cependant $\bar{f}_n(x)$ tend vers $f(x)$ (presque surement),sauf l'indication contraire,nous utilisons les même notations que dans le premier chapitre.En particulier l'estimation se base sur l'échantillon $(X_i = T_i \bigwedge C_i, \delta_i = 1_{T_i \leqslant C_i})_{1 \leqslant i \leqslant n}$,G est la fonction de répartition de la variable de censure T_F et le point terminal de la variable d'intérêt.

Convergence presque complète

La convergence presque complète implique la convergence presque sûre et se prête bien aux calculs faisant intervenir des sommes de variable aléatoire.
Elle est surtout utilisée en statistique non paramétrique .

Cette définition du taux a été introduite par Ferraty et Vien(2006). Elle a l'avantage théorique d'impliquer les deux vitesses de convergence classiques en probabilité et presque sûre ,et l'avantage pratique d'être souvent plus facile à démontrer.

Sous l'hypothèse de continuité de f au point x et sous les conditions assez faibles sur le noyau et la fenêtre, Foldes A et. al, ont montré en 1981 la convergence presque sur de $f_n(x)$ vers f(x). Quand à nous, nous allons montrer la convergence presque complète de f_n en point $x < T_H$, en précisant éventuellement le taux de convergence, sous des conditions un peu plus forte sur le noyau, de plus nous précisons le taux de convergence pour cela considérons les hypothèses suivantes :

H_1 : f est continue au point x

H_2 : f est de classe C^2 au voisinage de x

$H_3 : \exists k, p, \varepsilon_0 \in \mathbb{R}_+^*, \forall y \in]x - \varepsilon_0, x + \varepsilon_0[, |f(x) - f(y)| \leq k|x - y|^p$

$H_4 : h_n \longrightarrow 0$ et $nh_n^2/logn \longrightarrow \infty$

H_5 : K est une densité continue à droite à variation bornée sur $\mathbb{R}$ et telle que $\exists M > 0, \forall U \in \mathbb{R}, |u| \geq M \Longrightarrow K(u) = 0$

H_6 : K est bornée

$H_7 : \int uK(u)du = 0$ et $\int u^2K(u)du < \infty$

Théorème 2.4.

i) Sous $(H_1), (H_2), (H_5)$ et (H_6) nous avons :

$$f_n(x) \xrightarrow{p.co} f(x)$$

ii) Sous $(H_2), (H_4), (H_5)$ et (H_7) nous avons :

$$f_n(x) - f(x) = O_{p.co}\left(h_n^2 + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

iii) Sous $(H_3), (H_4), (H_5)$ *nous avons :*

$$f_n(x) - f(x) = O_{p.co}\left(h_n^p + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

Pour montrer ce théorème on utilise les deux lemmes suivantes :

Lemme 2.5. *Sous* (H_4) *et* (H_5)*, on a pour tout* $\theta < T_H$ *:*

$$\sup_{x\leq\theta} |f_n(x) - Ef_n(x)| = O_{p.co}\left(\frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

Où $E(f_n(x)) = \frac{1}{h_n}\int K\left(\frac{x-y}{h_n}\right) dF(y)$

Preuve. Soient $\theta < T_H$ et $x \leq \theta$, nous avons

$$|f_n(x) - Ef_n(x)| = \frac{1}{h_n}|\int K\left(\frac{x-y}{h_n}\right) d\left(F_n(y) - F(y)\right)|$$

On posant $u = \frac{x-y}{h_n}$ nous obtenons

$$|f_n(x) - Ef_n(x)| = \frac{1}{h_n}\int_{-M}^{+M} K(u) d\left(\tilde{F}_n(u) - \tilde{F}(u)\right)|$$

$$\tilde{F}_n(u) = F_n(x - uh_n) \text{ et } \tilde{F}(u) = F(x - uh)$$

En intégrant par parties,il vient

$$\begin{aligned}
|\int K(u)d(\tilde{F}_n(u) - \tilde{F}(u))| &= |\tilde{F}_n(u) - \tilde{F}(u)K(u)|_{-M}^{+M} - \int_{-M}^{+M}(\tilde{F}_n(u) - \tilde{F}(u))dK(u)| \\
&\leqslant |\tilde{F}_n(u) - \tilde{F}(u)K(u)|_{-M}^{+M} + |\int_{-M}^{+M}(\tilde{F}_n(u) - \tilde{F}(u))dK(u)| \\
&\leqslant |\int_{-M}^{+M}(\tilde{F}_n(u) - \tilde{F}(u))dK(u)| \\
&= \frac{1}{h_n}\int_{-M}^{+M} K(u) d\left(\tilde{F}_n(u) - \tilde{F}(u)\right)\Big| \leq \frac{1}{h_n}\int_{-M}^{+M}(\tilde{F}_n(u) - \tilde{F}(u)dK(u)\Big|
\end{aligned}$$

De plus, K étant à variation bornée et continue à droite s'écrit $K = K_1 - K_2$ ou K_1, K_2 deux

fonctions croissantes et continues à droite, d'où

$$\begin{aligned}|f_n(x) - Ef_n(x)| &\leq \frac{1}{h_n}\int_{-M}^{+M}\left(\tilde{F}_n(u) - \tilde{F}(u)\right)dK_1(u)| + \frac{1}{h_n}\int_{-M}^{+M}\left(\tilde{F}_n(u) - \tilde{F}(u)\right)dK_2(u)| \\ &\leq \frac{1}{h_n}\sup_{u>-M}\left|\tilde{F}_n(u) - \tilde{F}(u)\right|\int_{-M}^{+M} dK_1(u)| + \frac{1}{h_n}\sup_{u>-M}\left|\tilde{F}_n(u) - \tilde{F}(u)\right|\int_{-M}^{+M} dK_2(u)| \\ &\leq \frac{V_K}{h_n}\sup_{u>-M}\left|\tilde{F}_n(u) - \tilde{F}(u)\right| = \frac{V_K}{h_n}\sup_{u>-M}|F_n(x-uh_n) - F(x-uh_n)|\end{aligned}$$

ou V_K est la variation totale de K sur $\mathbb{R}$

Soit $\theta^* \in]\theta, T_H[$, puisque $h_n \longrightarrow 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0 : Mh_n < \theta_* - \theta$, ce qui montre que $x - uh_n < \theta^*$ du fait que $u > -M$ et $x \leq \theta$, par conséquent

$$\sup_{u>-M}|F_n(x-uh_n) - F(x-uh_n)| \leq \sup_{t<\theta^*}|F_n(t) - F(t)|$$

Donc pour tout $x \leq \theta$ nous avons

$$|f_n(x) - Ef_n(x)| \leq \frac{V_K}{h_n}\sup_{t<\theta^*}|F_n(t) - F(t)|$$

d'où

$$\sup_{x\leq\theta}|f_n(x) - Ef_n(x)| \leq \frac{V_K}{h_n}\sup_{t<\theta^*}|F_n(t) - F(t)| = O_{p.co}\left(\frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right) \tag{2.12}$$

■

Lemme 2.6. *Sous* $(H_1), (H_4), (H_5)$ *et* (H_6) *nous avons :*

$$Ef_n(x) \xrightarrow[n\to\infty]{} f(x)$$

Sous $(H_2), (H_4), (H_5)$ *et* (H_7) *nous avons :*

$$Ef_n(x) - f(x) = O(h_n^2)$$

Sous $(H_3), (H_4)$ *et* (H_5) *nous avons :*

$$Ef_n(x) - f(x) = O(h_n^p)$$

Preuve. En utilisons le changement de variable z=x-y nous pouvons écrire :

$$\begin{aligned} Ef_n(x) &= \frac{1}{h_n}\int K\left(\frac{x-y}{h_n}\right) f(y)dy \\ &= \frac{1}{h_n}\int \left(\frac{z}{h_n}\right) f(x-z)dz \xrightarrow[|z|\to\infty]{} f(x) \end{aligned}$$

D'après le théorème de Bochner($zK(z) \xrightarrow[|x|\to\infty]{}$ car K est à support compact) En utilisant le changement de variable $u = \frac{x-y}{h_n}$,il est ensuite que

$$\begin{aligned} |Ef_n(x) - f(x)| &= \left|\frac{1}{h_n}\int K\left(\frac{x-y}{h_n}\right) f(y)dy - f(x)\right| \\ &= |K(u)f(x-uh_n)du - f(x)| \\ &= \left|\int_{-M}^{+M} K(u)f(x-uh_n)du - f(x)\int_{-M}^{+M} K(u)du\right| \end{aligned}$$

$$= \left|\int_{-M}^{+M} K(u)(f(x-uh_n)du - f(x))du\right| \tag{2.13}$$

Comme f est de classe C^2 au voisinage de x , on peut lui appliquer le développement de Taylor à l'ordre 2, ce qui donne :

$$\begin{aligned} |Ef_n(x) - f(x)| &= \left|\int_{-M}^{+M} K(u)\left[f(x) - uh_n f'(x) + \frac{u^2h_n^2}{2}f''(\eta_n) - f(x).\right]du\right| \\ &= \frac{h_n^2}{2}\left|\int_{-M}^{+M} f''(\eta_n)u^2K(u)du\right| \end{aligned}$$

où η_n est entre x et $x - uh_n$
f'' étant continu au point x, nous avons pour $\varepsilon > 0$ quelconque :

$$\forall\delta > 0, \forall y, |y-x| < \delta \Rightarrow |f''(y) - f''(x)| < \varepsilon$$

et comme $h_n \longrightarrow 0, \exists n_0 \in \mathbb{N}, \forall n > n_0, h_n < \frac{\delta}{M}$, donc pour tout $n \geq n_0$ nous avons :

$$|\eta_n - x| \leq |uh_n < \delta \Rightarrow |f''(\eta_n) - f''(x)| < \varepsilon$$

d'où :

$$|Ef_n(x) - f(x)| \leq \frac{h_n^2}{2}\int_{-M}^{+M} |f''(\eta_n) - f''(x)|\, u^2K(u)du + \frac{|f''(x)h_n^2}{2}\int_{-M}^{+M} u^2K(u)du$$

$$\leq \left[\frac{\varepsilon + |f''(x)|}{2} \int_{-M}^{+M} u^2 K(u) du\right] h_n^2 = O(h_n^2) \tag{2.14}$$

On a

$$|Ef_n(x) - f(x)| \leq \int_{-M}^{+M} K(u) \, |f(x - uh_n - f(x))| \, du$$

et comme $h_n \to 0, n_1 \in \mathbb{N}, \forall n \geq n_1; h_n < \frac{\varepsilon_0}{M}$, donc pour tout $n \geq n_1$ nous avons :

$$|x - uh_n - x| = |u| h_n < \varepsilon_0 \Rightarrow |f(x - uh_n) - f(x)| \leq K|u|^p h_n^p < KM^p h_n^p$$

d'ou

$$|Ef_n(x) - f(x)| \leq KM^p h_n^p \int_{-M}^{+M} K(u) du = KM^p h_n^p = O(h_n^p)$$

■

H_8 : f est continue sur C

H_9 : f est de classe C^2 sur C

$H_{10} : \exists k, p, \varepsilon_0 \in \mathbb{R}_+^*, \forall x \in C, \forall y \in]x - \varepsilon_0, x + \varepsilon_0[, |f(x) - f(y)| \leq k|x - y|^p$

Théorème 2.7.

i) Sous $(H_8), (H_4), (H_5)$ *et* (H_6)*, nous avons :*

$$\sup_{x \in C} |f_n(x) - f(x)| \xrightarrow{p.co} 0$$

ii) Sous $(H_9), (H_4), (H_5)$ *et* (H_7)*, nous avons :*

$$\sup_{x \in C} |f_n(x) - f(x)| = O_{p.co}\left(h_n^2 + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

iii) Sous $(H_{10}), (H_4)$ *et* (H_5) *, nous avons :*

$$\sup_{x \in C} |f_n(x) - f(x)| = O_{p.co}\left(h_n^p + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

Ce théorème résulte les lemmes suivants :

Lemme 2.8. *i) Sous* $(H_8), (H_4), (H_5)$ *et* (H_6) *, nous avons :*

$$\sup_{x\in C} |Ef_n(x) - f(x)| \xrightarrow[n\to\infty]{} 0$$

ii) Sous $(H_9), (H_4), (H_5)$ *et* (H_7) *, nous avons :*

$$\sup_{x\in C} |f_n(x) - f(x)| = O(h_n^2)$$

ii) Sous $(H_9), (H_4), (H_5)$ *et* (H_7) *, nous avons :*

$$\sup_{x\in C} |f_n(x) - f(x)| = O(h_n^p)$$

Preuve.

i) Comme pour tout points i)lemme(4) ce point découle le théorème de Bochner du fait que f est uniformément continue car elle est continue sur le compact C

ii) D'après la relation

$$|Ef_n(x) - f(x)| \leq \left[\frac{\varepsilon + |f''(x)|}{2}\int_{-M}^{+M} u^2 K(u)du\right] h_n^2 = O(h_n^2)$$

et comme f" est continue sur le compact C :

$$\forall A > 0/\forall x \in C, |f''(x)| \leq A, d'ou : \sup_{x\in C} |Ef_n(x) - f(x)| \leq \left[\frac{\varepsilon + A}{2}\int_{-M}^{+M} u^2 K(u)du\right] h_n^2 = O(h_n^2)$$

iii) D'après la relation , nous avons pour tout $x \in C$

$|Ef_n(x) - f(x)| \leq \int_{-M}^{+M} K(u)\,|f(x - uh_n) - f(x)|\,du$ et comme $h_n \to 0/\exists n_0 \in \mathbb{N}/\forall n \geq n_0, h_n <$

donc pour tout $n \geqslant n_0$ nous avons : $|x - uh_n - x| = |u|h_n < \varepsilon_0 \Rightarrow k|u|^p h_n^p \leqslant kM^p h_n^p$ d'où

$$|Ef_n(x) - f(x)| \leq kM^p h_n^p \int_{-M}^{+M} K(u)du = kM^p h_n^p$$
$$\Rightarrow \sup_{x\in C} |Ef_n(x) - f(x)| \leq kM^p h_n p = O(h_n^p)$$

■

2.4.3 Estimateur à de taux de hasard

Le taux de hasard de T est définie par $\lambda(x) = \frac{f(x)}{S(x)}$ si $S(x) \neq 0$ et 0 sinon. L'estimateur de taux de hasard définie par

$$h_n(x) = \frac{f_n(x)}{S_n(x) + u_n}$$

où $(u_n)_{n\in\mathbb{N}}$ est une suite de nombres réels strictement positifs convergeant vers 0, et servant à éviter la division par 0.(pour $u_n = \frac{1}{n}$ nous retrouvons l'estimateur proposé par *Földes* A et al. dans [Földes et al. 1981].
Nous allons montre des résultats de convergence de λ_n similaires à ceux données à la section précédente pour f_n.

Théorème 2.9. *Soit $x < T_H$*

i) Sous $(H_1), (H_4), (H_5)$ et (H_6), nous avons :

$$\lambda_n(x) \xrightarrow{p.co} \lambda(x)$$

ii) Sous $(H_2), (H_4), (H_5)$ et (H_7) et pour un choix de $u_n = O\left(h_n^2 + \frac{1}{h_n}\sqrt{\frac{logn}{n}}\right)$, nous avons :

$$\lambda_n(x) - \lambda(x) = O\left(h_n^2 + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

iii) Sous $(H_3), (H_4)$ et (H_5) et pour un choix de $u_n = O\left(h_n^p + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$, nous avons :

Preuve. On utilise la décomposition suivante :

$$\lambda_n(x) - \lambda(x) = \frac{f_n(x) - f(x)}{S_n(x)} + u_n + (S(x) - S_n(x) - u_n)\frac{f(x)}{S(x)(S_n(x) + u_n)}$$

d'où $|\lambda_n(x) - \lambda(x)| \leq \frac{1}{S_n+u_n}|f_n(x) - f(x)| + (|S_n(x) - S(x)| + u_n)\frac{f(x)}{S(x)(S_n(x)+u_n)}$

et comme $(S_n(x) + u_n) \xrightarrow{p.s} S(x) \neq 0$,on a $\frac{1}{S_n(x)+u_n} \xrightarrow{p.s} \frac{1}{S(x)} \Rightarrow \exists c(x) > 0/\frac{1}{S_n(x)+u_n} \leq c(x)/\frac{1}{S_n(x)+u_n} \leq c(x)$ p.s donc

$$|\lambda_n(x) - \lambda(x)| \leqslant c(x)|f_n(x) - f(x)| + \frac{f(x)c(x)}{S(x)}\left(|S_n(x) - S(x)| + u_n\right) \text{ p.s}$$

et les résultats visés en découlent grâce aux théorèmes , en tenant compte à chaque fois du choix de u_n et du fait que $\sqrt{\frac{\log n}{n}} = O\left(\frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$

Théorème 2.10.

i) Sous $(H_8), (H_4), (H_5)$ *et* (H_6)*, nous avons :*

$$\sup_{x\in C} |\lambda_n(x) - \lambda(x)| \xrightarrow{p.co} 0$$

ii) Sous $(H_9), (H_4), (H_5)$ *et* (H_7) *et pour un choix de* $u_n = O\left(h_n^2 + \frac{1}{h_n}\sqrt{\frac{logn}{n}}\right)$*, nous avons :*

$$\sup_{x\in C} |\lambda_n(x) - \lambda(x)| = O\left(h_n^2 + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

iii) Sous $(H_{10}), (H_4)$ *et* (H_5) *pour un choix de* $u_n = O\left(h_n^p + \frac{1}{h_n}\sqrt{\frac{logn}{n}}\right)$*, nous avons :*

$$\sup_{x\in C} |\lambda_n(x) - \lambda(x)| = O\left(h_n^p + \frac{1}{h_n}\sqrt{\frac{\log n}{n}}\right)$$

Preuve.

$|\lambda_n(x) - \lambda(x)| \leq \frac{1}{S_n+u_n}|f_n(x) - f(x)| + (|S_n(x) - S(x)| + u_n)\frac{f(x)}{S(x)(S_n(x)+u_n)}$

et comme f est continue sur le compact C :$\exists A > 0/\forall x \in C : f(x) \leqslant A$, nous en déduisons alors , en notant $\theta = max(C)$ que :

$$\begin{aligned}\sup_{x\in C} |\lambda_n(x) - \lambda(x)| \leqslant & \frac{1}{\inf_(S_n(x) + u_n)} \sup_{x\in C} |f_n(x) - f(x)| \\ & + \frac{A \sup_{x\in C} (|S_n(x) - S(x)| + u_n)}{S(\theta) \inf(S_n(x) + u_n)} \end{aligned} \tag{2.15}$$

■

De plus , nous avons pour $\eta \in]0, S(\theta)[$ quelconque

$$\begin{aligned}\inf(S_n(x) + u_n(x)) \leqslant \frac{\eta}{2} &\Rightarrow \sup_{x\in C} (-S_n(x) - u_n) \geqslant -\frac{\eta}{2} \\ \Rightarrow \sup_{x\in C} |S_n(x) + u_n - S(x)| &\geqslant \sup_{x\in C} (-S_n(x) - u_n + S(x)) \\ &\geqslant \sup_{x\in C} (-S_n(x) - u_n) + S(\theta) > \frac{\eta}{2}\end{aligned}$$

Donc

$$\mathbb{P}\left(\inf(S_n(x)+u_n) \leqslant \frac{\eta}{2}\right) \leqslant \mathbb{P}\left(\sup_{x\in C}|S_n(x)+u_n-S(x) > \frac{\eta}{2}\right) \tag{2.16}$$

d'où

$$\sum_{n\geqslant 0}\mathbb{P}\left(\inf(S_n(x)+u_n) < \frac{\eta}{2}\right) < \infty \tag{2.17}$$

Les résultats visés découlent de (2.5) en tenant compte de (2.7) et des théorèmes (2.3)

Remarque 6. *La comparaison des résultats que nous venons de montrer pour λ_n avec ceux de Foldes A.et.al est identique à la comparaison que nous avons fait pour f_n.*

■

Dans cette section, nous appliquons le résultat du théorème afin d'obtenir la normalité asymptotique de l'estimateur de densité du noyau basé sur des données censurées doublement ou deux fois puis nous dérivons le résultat pour des données censurées à droite.

Annex

Théorème de Lindberge(théorème centrale limite) :
Soit X_1, X_1,...une suite de variables aléatoires réelles définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant la même loi. Supposons que l'espérance μ et l'écart σ de loi existent et soient finis avec $\sigma \neq 0$. Considérons la somme $S_n = X_1 + X_2 + ... + X_n$ alors l'espérance de S_n est $n\mu$ et son écart-type vaut $\sigma\sqrt{n}$, de plus quand n est assez grand, la loi normale $N(n\mu, n\sigma^2)$ est une bonne approximation de la loi de S_n

$$\bar{X}_n = \frac{S_n}{n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

et

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

Le théorème centrale limite énonce alors que la suite de variables aléatoires $Z_1, Z_2, ..., Z_n$ converge en loi vers une variable aléatoire z, définie sur le meme espace probabilité et de loi normale centrée et réduite N(0, 1) lorsque n tends vers l'infini. Cela signifie que si ϕ est la fonction de répartition de n(0, 1) alors pour tout réel Z :

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leqslant z) = \phi(z)$$

Ou équivalente

$$\lim_{n\to\infty} \mathbb{P}(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant z) = \phi(z)$$

Théorème de Chebyshev-Bienaymé :
Soit Y une variable aléatoire positive définie sur un espace probabilité.
On note E(Y) l'espérance de Y. Alors, pour tout $a > 0$ on a :

$$\mathbb{P}(Y \geqslant a) \leqslant \frac{E(Y)}{a}.$$

Théorème de Bochner : est un theorème d'analyse harmonique caracterisant la transformée de fourier d'une mesure positive sur un groupe localement compact.
Une fonction φ d'une variable aléatoire réelle si et seulement si elle vérifie les conditions suivantes :

- $\varphi(0) = 1$.
- φ est continue.
- φ est définie positive, c'est à dire que quels que soient les familles de réels $(u_1, ..., u_n)$ et de complexes $(z_1, ..., z_n)$, on a $\sum_{i,j=1}^{n} \varphi(u_i - u_j) z_i \bar{z}_j \geqslant 0$.

Bibliographie

[1] **Bertholon, H(2001)** *Une modelisation du vieillissement', Ph.D.thesis, Université Joseph Fourier, Grenoble.*

[2] **Billingsley, P(1995)** *Probability and Mesure, The University of Chigago :Willy.*

[3] **Boukeloua Mouhamed** : *Étude des estimateurs de la fonction de répartition et de densité dans un modèle de censure* , mémoire de Master, Département de Mathématiques, Université de Constantine 1, 2012-2013.

[4] **Boukeloua Mouhamed et Messaci Farida** : *Asymptotic normality of Kernel estimators based upan incomplete data*, Journal of non parametric statics, 24 mars 2016.

[5] **Elisa T-Lee, John Wenyu Wang** : *Statiscal methods for survival - data amalyse*, Wiley series in probability and statistics, Third edition.2003.

[6] **Kaplan, E. L. Meier, P.(1958)**. Non parametric estimation from incomplete observations. JASA, vol 53, pp.457-481.

[7] **Messaci Farida** : *Statistique non paramétrique pour les données censurées*, 2016/2017.

[8] **Philip Saint Pierre** : *cour introduction à l'analyse des durées de vie*, Pierre et Marie Curie, 25 février 2015.

[9] **Jian - Jian Ren and Minaggaogu** : *Regression M-estimators with doubly censored data*, Tulan University and MCGill University.

[10] **K. Dietz, M. Gail, K. Kricke bery, J.A. Tsiatic** : *Statistics for Biology and Health series editors.*

[11] **Lyasmine Harrouche** : *Analyse statistique des modèles de survie*, mémoire de Master, Département de Mathématiques, Université de Moulod Mammeri Tizi-Ouzou, 2017-2018.

[12] **Parsen, E., 1962**. On estimation of a probability density function and mode.*Annals of Mathematical.* , vol-33, pp.1065-1076.

[13] **Pollard, D.(1984)**. *Convergence of Stochastic Processes, New Haven : Yale University.*

[14] **Ren, J-J.(1997)**. *On self-Consistent Estimators and Kernel Density Estimators With Doubly Censored Data, Journal of statistical Planning and Inference, 64, 27-43.*

[15] **Rosnblatt, M., 1956.** : *Remarks on some non parametric estimates of a density function.* Annals of Mathematical statistics, vol-27, pp.832-837.

[16] **Souad Abdlaziz et Amira Sissaoui** : *Estimation pour les données censurées*, mémoire de Master, Département de Mathématiques, Université de Jijel, 2019-2020.

[17] **Turnbull, B.W.(1974)** : *Non Parametric Estimation Of Survivorship Function With Doubly Censored Data*, Journal of the American Statistical Association, 69-169-173.

Résume

Cette mémoire porte sur l'étude des propriétés asymptotiques des estimateurs non paramétriques des données censurées, comme l'estimateur de Kaplan-Meier, l'estimateur à noyau de la densité et de taux.

Nous établissons la convergence presque complète, la consistance et la normalité asymptotique de ces estimateurs dans le cas de la censure générale et la censure à droite.

Abstract

This thesis focuses on the study of the asymptotic properties of nonparametric estimators of censored data, such as the Kaplan-Meier estimator, the kernel estimator of density and rate. hspace * 0.5cm We establish the almost complete convergence, consistency and asymptotic normality of these estimators.