$\boxed{\text{TD n°1}}$ (Analyse de Survie)

Des indications sur la résolutions des exercices

Ex 1 1°) Facile

2°) $Y_i(t) = \mathbb{1}_{(X_i > t)}$ (au lieu de $Y_i(t) = \mathbb{1}_{(X_i \geq t)}$) pour $t \geq 0$.

$\mathbb{P}(Y_i = 1) = \mathbb{P}(X_i > t) = (1-F(t))(1-G(t))$ cf (1re question)

Donc $Y_i \sim \mathcal{B}((1-F(t))(1-G(t)))$. (loi de Bernoulli de para. $(1-F(t))(1-G(t))$)

$Y_n(t) = \sum_{i=1}^{n} Y_i(t)$ où les Y_i $i=1\dots,n$ sont indépendantes de même loi $\mathcal{B}(p)$

où $p = (1-F(t))(1-G(t))$.

Ainsi $Y_n(t)$ suit une loi Binomiale $\mathcal{B}(n,p)$

3°) Soit $J_n(t) = \mathbb{1}_{(Y_n(t) > 0)} = 1 - \mathbb{1}_{(Y_n(t) \leq 0)} = 1 - \mathbb{1}_{(Y_n(t) = 0)}$

Montrons que $\mathbb{1}_{(Y_n(t)=0)} \xrightarrow[n\to\infty]{\mathbb{P}} 0$? et ainsi $J_n(t) \xrightarrow{\mathbb{P}} 1$.

Soit $\varepsilon > 0$. On a $\mathbb{P}(\mathbb{1}_{(Y_n(t)=0)} > \varepsilon) = \mathbb{P}(\mathbb{1}_{(Y_n(t)=0)} = 1) = \mathbb{P}(Y_n(t) = 0)$

Or $Y_n(t) \sim \mathcal{B}(n,p)$ donc $\mathbb{P}(Y_n(t) = k) = C_n^k p^k (1-p)^{n-k}$, $k = 0, \dots, n$

et donc $\mathbb{P}(Y_n(t) = 0) = (1-p)^n = \left(1 - (1-F(t))(1-G(t))\right)^n$.

Avec $(1-F(t))(1-G(t)) < 1$ on a $\mathbb{P}(Y_n(t) = 0) \longrightarrow 0$.

Pour la v.a. $Y_n(t)$ on a par la loi faible des Grands nombres :

$$\frac{1}{n} Y_n(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t) \xrightarrow{\mathbb{P}} \mathbb{E}(Y_1(t)) = (1-F(t))(1-G(t))$$

4°) On a d'après 3°:

$$n \cdot \frac{J_n(t)}{Y_n(t)} = \frac{J_n(t)}{\left(\frac{Y_n(t)}{n}\right)} \xrightarrow[n\to\infty]{\mathbb{P}} \frac{1}{(1-F(t))(1-G(t))}$$ Ssi $(1-F(t))(1-G(t)) > 0$

Par suite

$$\frac{J_n(t)}{Y_n(t)} = \left(\frac{1}{n}\right) \cdot \left(\frac{n J_n(t)}{Y_n(t)}\right) \xrightarrow{\mathbb{P}} 0 \cdot \frac{1}{(1-F(t))(1-G(t))} = 0$$

5°). De $\widetilde{H}_n(t) = \int_0^t J_n(s)\, h(s)\, ds$ pour $t>0$, on a :

$$\mathbb{E}(\widetilde{H}_n(t)) = \int_0^t \mathbb{E}(J_n(s))\, h(s)\, ds \quad \text{(par Fubini)}$$

Or $\mathbb{E}(J_n(s)) = \mathbb{E}(\mathbb{1}_{(Y_n(s)>0)}) = \mathbb{P}(Y_n(s)>0) \xrightarrow[n\to\infty]{} 1$ (par 3°).)

et $|\mathbb{E}(J_n(s))\, h(s)| \le h(s)$. Or $\int_0^t h(s)\, ds = H(t) < \infty$, par th. T.C.D

$$\mathbb{E}(\widetilde{H}_n(t)) \xrightarrow[n\to\infty]{} \int_0^t h(s)\, ds = H(t) \text{, ie } \mathbb{E}(\widetilde{H}_n(t) - H(t)) \xrightarrow[n\to\infty]{} 0.$$

Ex 2. 1°) déjà vu en TD !

2°) voir cours !

3°) Dans le cas de v.a. censurées ~~soit~~ l'estimateur $\widetilde{S}_n(t)$;

$$\widetilde{S}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(T_i > t,\, D_i = 1)}$$

ici on introduit la v.a D_i qui exprime l'existence de v. censurées

- Rappelons que les (T_i) sont iid, (C_i) iid et $(T_i) \perp\!\!\!\perp (C_i)$

On note f_T la densité de T et f_C celle de C.

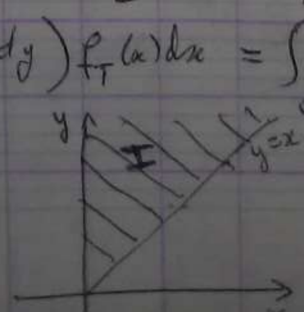
Rappel : si $Z_i := \mathbb{1}_{(T_i \le C_i)}$ alors $Z_i \sim \mathcal{B}(p)$ (sont iid)

où $p = \int_0^\infty S_C(t)\, f_T(t)\, dt$ où $S_C(t) = \mathbb{P}(C>t) = 1 - F_C(t)$

En effet : $\mathbb{P}(Z_i = 1) = \mathbb{P}(T_i \le C_i) = \mathbb{P}((T_i, C_i) \in I)$

$$= \iint_I f_{(T,C)}(x,y)\, dx\, dy = \iint_I f_T(x)\, f_C(y)\, dx\, dy$$

$$= \int_0^\infty \left(\int_x^{+\infty} f_C(y)\, dy \right) f_T(x)\, dx = \int_0^\infty \underbrace{\mathbb{P}(C > x)}_{S_C(x)} \cdot f_T(x)\, dx =: p.$$



③

Pour l'estimateur $\tilde{S}_n(t)$:

De même on pose $U_i := \mathbb{1}_{(T_i > t,\, D_i = 1)}$ (des va. de Bernoulli iid !)

et on a : $\mathbb{P}(U_i = 1) = \mathbb{P}(T_i > t, D_i = 1) = \mathbb{P}(T_i > t, T_i \leq C_i)$

$$= \mathbb{P}(t < T_i \leq C_i) = \int_t^{\infty} \left(\int_x^{+\infty} f_C(y)\,dy \right) f_T(x)\,dx$$

$$= \int_t^{\infty} S_C(x)\, f_T(x)\,dx$$

Donc $U_i \sim \mathcal{B}(q)$ avec $q = \int_t^{\infty} S_C(x) f_T(x)\,dx$.

Par suite en appliquant la LFGN on arrive à :

$$\tilde{S}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(T_i > t,\, D_i = 1)} = \frac{1}{n} \sum_{i=1}^{n} U_i \xrightarrow[\mathbb{P}.p.s]{} \mathbb{E}(U_1) = q.$$

ie. $\tilde{S}_n(t) \xrightarrow{\mathbb{P}.p.s} \int_t^{\infty} S_C(x) f_T(x)\,dx$.

Mais $\int_t^{\infty} S_C(x) f_T(x)\,dx = \int_t^{\infty} (1 - F_C(x)) f_T(x)\,dx = \int_t^{\infty} f_T(x)\,dx - \int_t^{\infty} F_C(x) f_T(x)\,dx$

$$= S_T(t) - \int_t^{\infty} F_C(x) f_T(x)\,dx < S_T(t)$$

si il y a censure
ie $F_C(x) \neq 0$

Donc $\tilde{S}_n(t) \not\to S_T(t)$ dans <u>le cas d'existence de censure</u>

En fait $\tilde{S}_n(t) \xrightarrow{\mathbb{P}.p.s} \ell$ avec $\ell < S_T(t)$.

TD n°2. Analyse de survie

Ex I et **Ex II** déjà faits !

Ex III. Avec $f_T(t) = \dfrac{\lambda \alpha t^{\alpha-1}}{(1+\lambda t^{\alpha})^2}$, $t \geq 0$, $\lambda > 0$, $\alpha \geq 1$.

On a $S_T(t) = \displaystyle\int_t^{\infty} f_T(t)\,dt = \dots$

à calculer !

et $h(t) = \dfrac{f_T(t)}{S_T(t)} = \dots$

Ex IV Ici $T \sim \mathcal{E}(\theta)$ loi exponentielle de paramètre $\theta > 0$

$C \sim \mathcal{E}(\beta)$ " " $\beta > 0$

et refaire les calculs de l'Ex II cette fois avec ces deux lois exponentielles..

Ex V. Avec la présence dans les données de l'exercice de v.a. censurées (5^+ etc...) on utilise l'estimateur $\hat{S}_{KM}$.

On calcule l'estimateur de Kaplan-Meier $\hat{S}_{KM}(t)$ par :

(*) $\boxed{\hat{S}_{KM}(t) = \displaystyle\prod_{t_{(i)} \leq t} \left(1 - \frac{m_i}{n_i}\right)^{D_i}}$ où $\underset{\substack{\parallel \\ 0}}{t_{(0)}} < t_{(1)} < t_{(2)} < \dots < t_{(n)}$

où m_i = nbre d'événements E (observés) ~~avant~~ à $t_{(i)}$

n_i = nbre "d'individus à risque" (en vie) à $t_{(i-1)}$

$$n_i = n - \sum_{j=1}^{n-1} m_j - \sum_{j=1}^{i-1} c_j$$

avec c_j = nbre de censures dans l'intervalle $[t_{(j)}, t_{(j+1)}[$

ou encore $\quad n_i = n_{i-1} - m_{i-1} - c_{i-1}$

et m_{i-1} = nbre d'évènements (obs) à $t_{(i-1)}$

c_{i-1} = nbre de v.a censurées ds $[t_{(i-1)}, t_{(i)}[$, $i > 1$.

et on calcule $\hat{S}_{KM}(t)$ ~~pour~~ chaque intervalle $[t_{(i-1)}, t_{(i)}[$

et $\hat{S}_{KM}(t)$ est constant ~~pour~~ dans $t \in [t_{(i-1)}, t_{(i)}[$.

(voir l'exemple du cours!)

Pour le taux de service $h(t)$ (ou fonction de risque) :

un estimateur est $\hat{h}_n(t_{(i)}) := \dfrac{m_i}{n_i}$, $i = 1, 2, \ldots, n$.

puis on fait une interpolation linéaire entre les points $(t_{(i)}, \hat{h}_n(t_{(i)}))$.

Tracer les graphes correspondants à $\hat{S}_{KM}(t)$ et $\hat{h}_n(t)$.

Remarque : les instants $t_{(i)}$ correspondent soit à l'observation de l'évènement E ou une censure.