Université Abou Bekr Belkaid – Tlemcen
Faculté des Sciences
Département de Mathématiques

Année Universitaire 2020/2021
$1^{ère}$ année Master
Semestre 2

Examen Final "Analyse de Survie"

Aucun document n'est autorisé
Durée : 1h30mn
03 Juin 2021

* La qualité de la rédaction, des justification apportées et de la présentation de la copie seront prises en compte dans la notation.

Exercice 1. (4pts)

Considérons un individu d'age θ et de durée de maintien résiduelle T distribuée selon une loi de "**Pareto**" ($T \curvearrowright Par(\theta, \alpha)$) :

$$S_T(t) = \theta^\alpha (\theta + t)^{-\alpha} \qquad pour\ \ t \geq 0\ \ et\ \ \alpha > 0.$$

1. Expliciter la fonction de risque h et la fonction de vie résiduelle e.
2. Expliciter la fonction de survie de cet individu conditionnée par le fait qu'il sera en vie dans x années. Que peut-on déduire.

Exercice 2. (8pts)

Soit un groupe de 10 patients (Groupe A) suivi pour un cancer de type "cancer pancréatique exocrine". En parallèle, l'on suit un autre groupe de 12 patients (Groupe B) atteints d'un "cancer pancréatique endocrine". Les durées de vie sont ci-après données :

Groupe A : 4,$4^+, 4^+, 5, 6^+, 6^+, 7, 7, 9, 9^+$.
Groupe B : $1^+, 2, 2^+, 5, 5, 5^+, 6^+, 7, 7^+, 8^+, 8, 9$.

1. Calculer l'estimateur de **KAPLAN-MEIER** de la fonction de survie du groupe A, puis tracer le graphe correspondant.
2. On souhaite comparer la survie de chacun des deux groupes A et B dans un délai de 9 mois. Tester l'égalité des deux survies (test de **LOG-RANK**) à un risque $\alpha = 5$ %, conclure. ($\mathbb{P}(\chi_1^2 > 3.84) = 0.05$).

Tourner la page.

Exercice 3. (8pts)

On considère une situation de censure aléatoire à droite :

$$\mathsf{X}_i = T_i \wedge C_i \quad et \quad D_i = \begin{cases} 1 & si \;\; T_i \leq C_i \\ 0 & si \;\; T_i > C_i \end{cases}$$

dans laquelle on suppose les censures C_i indépendantes des durées T_i. On fait l'hypothèse qu'il existe $\beta > 0$ tel que $S_C(t) = S_T(t)^\beta$ pour tout $t \geq 0$.

1. Calculer la densité de C en fonction de celle de f_T et de S_T, on déduire la fonction de risque de C en fonction de celle de T.
2. Commenter les cas particulier du modèle $\beta = 1$ et $\beta \to 0$.
3. Déterminer la loi de D_i.

   A partir de maintenant, on suppose que T suit la loi exponentielle de paramètre θ i.e. $S_T(t) = exp(-\theta t)$. La vraisemblance de l'échantillon $(X_1, D_1), ..., (X_n, D_n)$ s'écrit :

   $$L(\theta) = \Pi_{i=1}^n \left[f_T(X_i, \theta) S_C(X_i, \theta)\right]^{D_i} \left[f_C(X_i, \theta) S_T(X_i, \theta)\right]^{1-D_i}$$

4. Écrire la log-vraisemblance de (θ, β) en fonction de n, $\bar{\mathsf{D}} = \dfrac{1}{n}\Sigma_{i=1}^n D_i$ et $\bar{\mathsf{X}} = \dfrac{1}{n}\Sigma_{i=1}^n X_i$. En déduire les estimateurs de maximum de vraisemblance de β lorsque θ est connu et de θ lorsque β est connu.

Bonne courage.

Correction Epreuve finale "Analyse de survie"

2020/2021.

Exercice n°1 :

$T \hookrightarrow Par(\theta, \alpha)$: $S_T(t) = \theta^\alpha (\theta + t)^{-\alpha}$, $t \geq 0$ et $\alpha > 0$

1°/. Fonction de Risque h :

(0,5) $$h(t) = \frac{-S'(t)}{S(t)} = \frac{-(-\alpha)\theta^\alpha(\theta+t)^{-\alpha-1}}{\theta^\alpha(\theta+t)^{-\alpha}} = \alpha(\theta+t)^{-1} = \frac{\alpha}{(\theta+t)} \cdot \quad t \geq 0$$

- Fonction de vie Residuelle e :

$$e(t) = \frac{\int_t^{+\infty} S_T(s)\,ds}{S_T(t)} \quad \text{or :}$$

(01,5) $$\int_t^{+\infty} S_T(s)\,ds = \int_t^{+\infty} \theta^\alpha (\theta+s)^{-\alpha} ds = \theta^\alpha \left[\frac{(\theta+s)^{-\alpha+1}}{-\alpha+1}\right]_t^{+\infty} \quad (\text{pour } \alpha \neq 1)$$

cette intégrale converge pour $-\alpha + 1 < 0$ c.à.d $\alpha > 1$:

Donc ; $$\int_t^{+\infty} S_T(s)\,ds = \theta^\alpha \left(0 - \frac{(\theta+t)^{-\alpha+1}}{-\alpha+1}\right) = \frac{\theta^\alpha (\theta+t)^{-\alpha+1}}{\alpha - 1}$$

Alors, $$e(t) = \frac{\theta^\alpha (\theta+t)^{-\alpha+1}}{\alpha-1} \cdot \frac{1}{\theta^\alpha (\theta+t)^{-\alpha}} = \frac{\theta+t}{\alpha-1} \quad \text{pour } t \geq 0 \text{ et } \alpha > 1$$

Conclusion :
$$\begin{cases} h(t) = \dfrac{\alpha}{\theta+t} & \text{si } t \geq 0 \text{ et } \alpha > 0. \\[2ex] e(t) = \dfrac{\theta+t}{\alpha-1} & \text{si } t \geq 0 \text{ et } \alpha > 1 \end{cases}$$

2°/ $\mathbb{P}(T > t \,/\, T > x) = ?$

$$\mathbb{P}(T > t \,/\, T > x) = \frac{\mathbb{P}(T > t, T > x)}{\mathbb{P}(T > x)} \qquad 0 \leq x \leq t$$

(01,5) $$= \frac{\mathbb{P}(T > t)}{\mathbb{P}(T > x)}$$

$$= \frac{\theta^\alpha (\theta+t)^{-\alpha}}{\theta^\alpha (\theta+x)^{-\alpha}}$$

$$= \frac{(\theta+x+(t-x))^{\alpha}}{(\theta+x)^{-\alpha}}$$

Alors, $\mathbb{P}(T>t \mid T>x) = (\theta+x)^{\alpha}\,(\theta+x+(t-x))^{-\alpha}$.

$= \mathbb{P}(Y > t-x)$ avec $Y \hookrightarrow Par(\theta+x, \alpha)$.

Déduction : Dans ce modèle, il n'y a pas de modification générationnelle de la durée de vie. Un individu qui a 40 ans et qui survit 10 ans aura la même loi de survie dans 10 ans qu'un individu qui a aujourd'hui 50 ans. (0,5)

Exercice n° 2 :

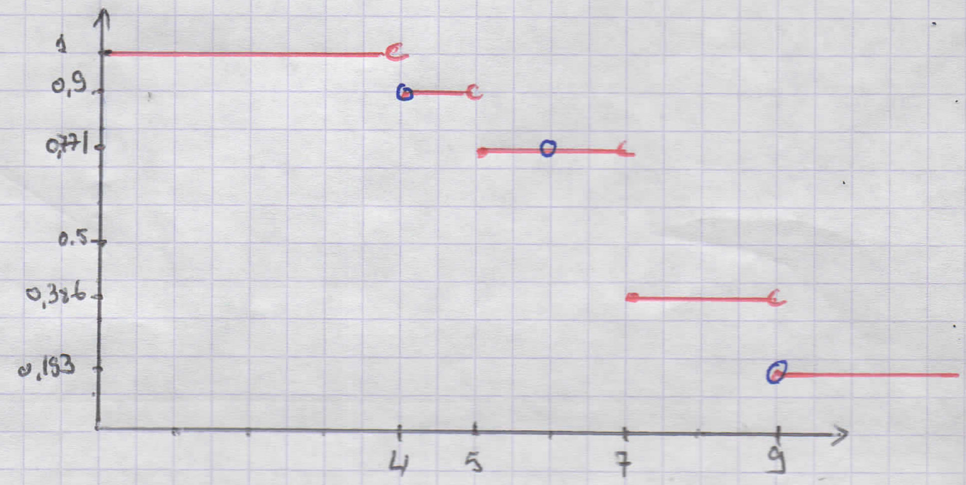
(0.5) 1°/ $\hat{S}_{KM}(t) = \prod_{i / t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$: Estimateur de Kaplan-Meier.

(0,75) 0,25x7

t_i	d_i	c_i	n_i	d_i/n_i	$1-\frac{d_i}{n_i}$	$\hat{S}_{KM}(t_i)$
0	0	0	10	0	1	1
4	1	2	10	0.1	0.9	0.9
5	1	2	07	0.143	0.857	0.771
7	2	0	04	0.5	0.5	0.386
9	1	1	02	0.5	0.5	0.193

Graphe :

(0.5)



2°/ Test de Log-Rank :

$H_0 : S_A = S_B$

t_i	c_{Ai}	d_{Ai}	n_{Ai}	c_{Bi}	d_{Bi}	n_{Bi}	n_i	d_i	d_i/n_i	$e_{Ai}=n_{Ai}\cdot\frac{d_i}{n_i}$	$e_{Bi}=n_{Bi}\cdot\frac{d_i}{n_i}$
0	0	0	10	1	0	12	22	0	0	0	0
2	0	0	10	1	1	11	21	1	0.048	0,480	0,528
4	2	1	10	0	0	09	19	1	0.053	0.530	0,477
5	2	1	07	2	2	09	16	3	0,188	1.316	1,692
7	0	2	04	1	1	05	09	3	0,333	1.332	1,665
8	0	0	02	1	1	03	05	1	0,200	0,400	0,600
9	1	1	02	0	1	01	03	2	0,667	1.334	0,667
Σ		$O_A=5$			$O_B=6$					$E_A=5,392$	$E_B=5,629$

03 0,25x12

* Sur notre période de survie, on a 5 décédé pour l'échantillon A et 6 dans l'échantillon B.
* "$H_0 : S_A = S_B$" : la survie des patients atteints d'un cancer pancréatique exocrine ne diffère pas de la survie des patients atteints d'un cancer pancréatique endocrine.

Alors pour un risque $\alpha = 5\%$, l'indice calculer :

0,5 $$T_n = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} = 0,053 < 3,841.$$

(car $T_n \hookrightarrow \chi^2_1$ et : $\mathbb{P}(\chi^2_1 > 3,841) = 0.05$)

Ainsi T_n est inférieur au seuil, (Région de Rejet $D_n = \mathbb{P}(T_n > 3,84)$)

0,5 donc, on ne rejet pas H_0. c.à.d la différence est non significative : Il n'y a pas de différence de survie entre les deux groupe de cancers pancréatiques.

Exercice n° 3:

$$X_i = T_i \wedge C_i \quad \text{et} \quad D_i = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{si } T_i > C_i \end{cases}$$

$$C_i \perp\!\!\!\perp T_i \quad \text{et} \quad \exists\, \beta > 0 \,/\, S_C(t) = [S_T(t)]^{\beta}$$

1°/ • $f_C = ?$: Soit $t \geq 0$.

$$f_C(t) = -d\, S_C(t)$$

(01)

$$= -\beta\, S_T(t)^{\beta-1} \cdot S'_T(t)$$

$$= -\beta\, S_T(t)^{\beta-1} \cdot S'_T(t)$$

donc, $\boxed{f_C(t) = \beta\, S_T^{\beta-1}(t) \cdot f_T(t)}$ car $f_T(t) = -S'_T(t)$

(01) • $h_C = ?$ $h_C(t) = \dfrac{f_C(t)}{S_C(t)} = \dfrac{\beta\, S_T^{\beta-1}(t)\, f_T(t)}{S_T^{\beta}(t)} = \beta\, \dfrac{f_T(t)}{S_T(t)} = \beta\, h_T(t)$.

(05) 2°/ $\beta = 1$: $f_C(t) = f_T(t)$ donc C et T suit la même loi (Identiquement distribuée).

$\beta \to \infty$: $S_C(t) \longrightarrow 1$ donc $F_C(t) \longrightarrow 0$ c.à.d la censure disparaît.

3°/ Loi de D_i : $D_i = \mathbb{1}_{\{T_i \leq C_i\}}$ donc $D_i \hookrightarrow \mathcal{B}(p)$ avec.

$p = \mathbb{P}(D_i = 1) = \mathbb{P}(T_i \leq C_i) = \mathbb{P}(X_i \in I)$ avec.

$$I = \{(x,y) \in \mathbb{R}^2 \,/\, x \geq 0 \text{ et } x \leq y\}.$$

donc, $p = \displaystyle\int_0^{+\infty}\int_x^{+\infty} f_T(x)\, f_C(y)\, dy\, dx$ ($T_i \perp\!\!\!\perp C_i$).

(01,5)

$$= \int_0^{+\infty} f_T(x)\, S_C(x)\, dx.$$

$$= \int_0^{+\infty} f_T(x) \cdot S_T^{\beta}(x)\, dx$$

$$= \int_0^{+\infty} -S'_T(x)\, S_T^{\beta}(x)\, dx.$$

$$= \frac{-1}{\beta+1}\left[S_T^{\beta+1}(x)\right]_0^{+\infty}$$

$= \dfrac{-1}{\beta+1}(0-1)$ donc $\boxed{p = \dfrac{1}{\beta+1}}$.

Conclusion : $D_i \hookrightarrow \mathcal{B}\left(\frac{1}{\beta+1}\right)$.

4°/ $S_T(x) = \exp(-\theta t) \Rightarrow f_T(x) = +\theta \exp(-\theta t)$.

$S_C(x) = \exp(-\theta\beta t) \Rightarrow f_C(x) = +\theta\beta \exp(-\theta\beta t)$. (0,75)

$$L(\theta) = \prod_{i=1}^{n} \left[f_T(X_i,\theta)\, S_C(X_i,\theta) \right]^{D_i} \left[f_C(X_i,\theta)\, S(X_i,\theta) \right]^{1-D_i}$$

Donc, $\log[L(\theta)] = \sum_{i=1}^{n} \Big(D_i \log\left[\theta \exp\left[-\theta X_i(1+\beta)\right]\right] + (1-D_i) \log\left[\theta\beta \left(\exp\left[-\theta X_i(1+\beta)\right]\right]\right) \Big)$

(01.25)

$$= \sum_{i=1}^{n} \Big(\cancel{D_i \log\theta} + \cancel{D_i(-\theta X_i(1+\beta))} + \log\theta + \log\beta - \theta X_i(1+\beta) - \cancel{D_i \log\theta} - D_i \log\beta - \cancel{D_i(-\theta X_i(1+\beta))} \Big)$$

$$= \sum_{i=1}^{n} \left(\log\theta + (1-D_i)\log\beta - \theta X_i(1+\beta) \right)$$

$$= n\log\theta + \left(n - \sum_{i=1}^{n} D_i\right)\log\beta - \theta(1+\beta) \quad X_i .$$

$$\log(L(\theta)) = n\log\theta + n(1-\bar{D})\log\beta - n\theta(1+\beta)\bar{X} .$$

- E.M.V(β) :

$$\frac{\partial L(\theta,\beta)}{\partial\beta} = \frac{n(1-\bar{D})}{\beta} - n\theta\bar{X}$$

(01)

$$\frac{\partial L(\theta,\beta)}{\partial\beta} = 0 \iff \beta = \frac{1-\bar{D}}{\theta\bar{X}} \qquad \text{Alors}$$

$$\hat{\beta} = \frac{1-\bar{D}}{\hat{\theta}\bar{X}}$$

- E.M.V(θ) ;

$$\frac{\partial L(\theta,\beta)}{\partial\theta} = \frac{n}{\theta} - n(1+\beta)\bar{X}$$

(01)

$$\frac{\partial L(\theta,\beta)}{\partial\theta} = 0 \iff \theta = \frac{1}{(1+\beta)\bar{X}}$$

$$\hat{\theta} = \frac{1}{(1+\hat{\beta})\bar{X}}$$