ISFA
GRANDE ÉCOLE D'ACTUARIAT ET DE GESTION DES RISQUES
INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

MODÈLES DE DURÉE

Exercices – Modèles paramétriques

Corrigés

Exercice n°1 (loi de Pareto)

Considérons un groupe de *n* individus d'âges $(\theta_1, \ldots, \theta_n)$ et de durées de maintien résiduelles $(X_1, \ldots, X_n)$ indépendantes et marginalement distribuées selon une loi de Pareto de 2^e espèce :

$$S_i(t) = \theta_i^{\alpha} (\theta_i + t)^{-\alpha} \text{ pour } t \geq 0.$$

1. Soit un individu d'âge θ dans ce groupe. Expliciter sa fonction de risque *h* et sa fonction espérance de vie résiduelle *e*. Que peut-on en déduire quant à la pertinence du modèle proposé ?

$$f(t) = \alpha\theta^{\alpha}(\theta+t)^{-(\alpha+1)} \Rightarrow h(t) = \frac{\alpha}{\theta+t} \text{ pour } t \geq 0.$$

$$e(t) = \frac{1}{S(t)}\int_t^{\infty} S(x)dx = \theta^{-\alpha}(\theta+t)^{\alpha}\theta^{\alpha}\int_t^{\infty}(\theta+x)^{-\alpha}dx.$$

$$\int_t^{\infty}(\theta+x)^{-\alpha}dx < +\infty \text{ pour } \alpha > 1 \text{ donc}$$

$$e(t) = \frac{\theta+t}{\alpha-1} \text{ pour } t \geq 0 \text{ si } \alpha > 1.$$

Le modèle n'est pas pertinent pour modéliser la durée de vie humaine car la fonction de risque est décroissante avec le vieillissement. Par contre il pourrait être envisagé pour le temps passé au chômage, en modélisant l'ancienneté dans l'état.

2. Expliciter la fonction de survie de cet individu conditionnée par le fait qu'il sera en vie dans *x* années. Que peut-on en déduire ?

Soient $0 \leq x \leq t$, on a :

$$P\left[X>t \mid X>x\right]=\frac{P\left[X>t\right]}{P\left[X>x\right]}=\left(\frac{\theta+x}{\theta+x+(t-x)}\right)^{\alpha}=P\left[Y>t-x\right],$$

où $Y \sim Par(\theta+x,\alpha)$.

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.

3. Supposons que ce groupe de *n* individus soit observé jusqu'à son extinction et notons $x=(x_1,\ldots,x_n)$ les durées de survie observées. Donnez l'estimateur du maximum de vraisemblance de α.

$$\mathbf{L}(\alpha,\underline{x})=\prod_{i=1}^{n}\alpha\theta_i^{\alpha}(\theta_i+x_i)^{-(\alpha+1)}$$

$$\ln\mathbf{L}(\alpha,\underline{x})=n\ln\alpha+\alpha\sum_{i=1}^{n}\ln\theta_i-(\alpha+1)\sum_{i=1}^{n}\ln(\theta_i+x_i)$$

$$\frac{\partial}{\partial\alpha}\ln\mathbf{L}(\alpha,\underline{x})=\frac{n}{\alpha}+\sum_{i=1}^{n}\ln\theta_i-\sum_{i=1}^{n}\ln(\theta_i+x_i)$$

$$\frac{\partial}{\partial\alpha}\ln\mathbf{L}(\hat{\alpha},\underline{x})=0\Leftrightarrow\hat{\alpha}=\frac{n}{\sum_{i=1}^{n}\ln\left(1+\frac{x_i}{\theta_i}\right)}$$

4. Supposons à présent que le groupe n'est plus observé que pendant *c* années. Donnez l'estimateur du maximum de vraisemblance de α à l'aide des observations $t=(t_1,\ldots,t_n)$ où $t_i=x_i\wedge c$.

Notons $r=\sum_{i=1}^{n}1_{\{x_i\le c\}}$ et considérons le vecteur ordonné $y=(y_1,\ldots,y_n)$ où $y_i=x_{(i)}$ pour $i\in\{1,\ldots,r\}$ et $y_i=c$ pour $i\in\{r+1,\ldots,n\}$. Notons θ_i l'âge de début d'observation de l'individu ayant produit y_i.

$$\mathbf{L}(\alpha,\underline{x})=\prod_{i=1}^{r}\alpha\theta_i^{\alpha}(\theta_i+y_i)^{-(\alpha+1)}\prod_{i=r+1}^{n}\theta_i^{\alpha}(\theta_i+y_i)^{-\alpha}$$

$$\ln\mathbf{L}(\alpha,\underline{x})=r\ln\alpha+\alpha\sum_{i=1}^{n}\ln\theta_i-(\alpha+1)\sum_{i=1}^{r}\ln(\theta_i+y_i)-\alpha\sum_{i=r+1}^{n}\ln(\theta_i+y_i)$$

$$\frac{\partial}{\partial\alpha}\ln\mathbf{L}(\alpha,\underline{x})=\frac{r}{\alpha}-\sum_{i=1}^{n}\ln\left(1+\frac{y_i}{\theta_i}\right)$$

$$\frac{\partial}{\partial \alpha} \mathbf{ln\,L}\left(\hat{\alpha}, \underline{x}\right) = 0 \Leftrightarrow \hat{\alpha} = \frac{r}{\sum_{i=1}^{r} \mathbf{ln}\left(1 + \frac{x_{(i)}}{\theta_i}\right) + \sum_{i=r+1}^{n} \mathbf{ln}\left(1 + \frac{c}{\theta_i}\right)}$$

On constate en particulier que si on choisissait de travailler sur échantillon réduit (*ie* en éliminant les données censurées) on surestimerait le paramètre α.

5. Considérons à présent la situation où tous les individus du groupe ont la même ancienneté inconnue θ au début de la période s'observation et que cette-dernière se prolonge jusqu'à la dernière sortie. Ecrivez le programme d'optimisation dont est solution l'estimateur du maximum de vraisemblance $\left(\hat{\theta}, \hat{\alpha}\right)$ de $\left(\theta, \alpha\right)$ et proposez une méthode pour fournir une valeur initiale à l'algorithme de résolution.

$$\mathbf{ln\,L}\left(\theta, \alpha, \underline{x}\right) = n\,\mathbf{ln}\,\alpha + n\alpha\,\mathbf{ln}\,\theta - \left(\alpha + 1\right)\sum_{i=1}^{n} \mathbf{ln}\left(\theta + x_i\right)$$

Considérons le système aux dérivés partielles :

$$\begin{cases} \frac{\partial}{\partial \theta} \mathbf{ln\,L}\left(\theta, \alpha, \underline{x}\right) = \frac{n\alpha}{\theta} - \left(\alpha + 1\right)\sum_{i=1}^{n} \frac{1}{\theta + x_i} = 0 \\ \frac{\partial}{\partial \alpha} \mathbf{ln\,L}\left(\theta, \alpha, \underline{x}\right) = \frac{n}{\alpha} + n\,\mathbf{ln}\,\theta - \sum_{i=1}^{n} \mathbf{ln}\left(\theta + x_i\right) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + \frac{x_i}{\theta}} \left\{1 + \frac{1}{n}\sum_{i=1}^{n} \mathbf{ln}\left(1 + \frac{x_i}{\theta}\right)\right\} - 1 = 0 \\ \alpha = \frac{n}{\sum_{i=1}^{n} \mathbf{ln}\left(1 + \frac{x_i}{\theta}\right)} \end{cases}$$

Les moments d'une loi de Pareto n'existent que jusqu'à l'ordre $\left[\alpha\right]$; α étant inconnu, on n'utilisera pas la méthode des moments pour fournir une valeur initiale à l'algorithme d'obtention des estimateurs du maximum de vraisemblance. On pourra néanmoins se tourner vers la méthode des quantiles qui consiste pour $0 < p_1 < p_2 < 1$, à résoudre l'équation en α :

$$\frac{Q\left(p_1\right)}{Q\left(p_2\right)} = \frac{F^{-1}\left(p_1\right)}{F^{-1}\left(p_2\right)} = \frac{\left(1 - p_1\right)^{-1/\alpha} - 1}{\left(1 - p_2\right)^{-1/\alpha} - 1},$$

où $Q\left(p\right)$ désigne le quantile empirique d'ordre *p*.

Exercice n°2 (Modèles à espérance de vie affine)

Soit *X* une variable aléatoire « durée de maintien » dont la loi admet une densité. On suppose de plus que *X* admet une espérance. On note *S* la fonction de survie, *h*(*x*) le taux de hasard calculés en *x* et *e*(*x*) l'espérance de maintien au delà de *x*.

1. Montrer que, pour tout *x* où *e*(*x*) est non nul, on a la relation :

$$h(x) = \frac{1+e'(x)}{e(x)}.$$

$$e(x) = \frac{1}{S(x)}\int_x^{\infty} S(t)\,dt \Rightarrow e'(x) = -\frac{S'(x)}{S^2(x)}\int_x^{\infty} S(t)\,dt + \frac{0-S(x)}{S(x)} = h(x)e(x) - 1.$$

2. On cherche les lois telles que pour tout *x* réel positif on ait : $e(x) = ax + b$ (avec $a \geq 0$ et $b > 0$).

- Calculer *h*(*x*).
- Vérifier que pour $x \geq 0$, $S(x) = \left(1 + \frac{a}{b}x\right)^{-(1+1/a)}$.
- Donner l'espérance de *X* (sans calculs !).
- Montrer que si l'on suppose $a < 1$, la variance de *X* existe et est égale à $b^2 \frac{1+a}{1-a}$.

Remarque : Il n'est pas indispensable, pour calculer $\mathbf{E}\left[X^2\right]$ *de calculer la densité f*(*x*)*.*

- $h(x) = \dfrac{1+a}{ax+b}$.
- Soit $\psi : x \mapsto \left(1 + \frac{a}{b}x\right)^{-(1+1/a)}$ pour $x \geq 0$. Sur son ensemble de définition $\mathbf{R}_+$, ψ est à valeurs dans $\mathbf{R}_+^*$.

$$\psi'(x) = -\frac{a+1}{b}\left(1 + \frac{a}{b}x\right)^{-(2+1/a)}$$

$$\Rightarrow \frac{\psi'(x)}{\psi(x)} = -\frac{a+1}{b}\left(1+\frac{a}{b}x\right)^{-1} = -\frac{a+1}{ax+b} = -h(x) \Rightarrow S = \psi.$$

- $\mathbf{E}[X] = e(0) = b$
- $\mathbf{Var}[X] = 2\int_0^{\infty} t\,S(t)\,dt - \{\mathbf{E}[X]\}^2$

$$\begin{aligned}\int_0^\infty t\,S(t)\,dt &= \int_0^\infty t\left(1+\frac{a}{b}t\right)^{-(1+1/a)} dt = \int_1^\infty \frac{b}{a}(x-1)\,x^{-(1+1/a)}\frac{b}{a}\,dx \\ &= \frac{b^2}{a^2}\int_1^\infty x^{-1/a} - x^{-1-1/a}dx < +\infty \text{ si } a>1 \\ &= \frac{b^2}{a^2}\left(\frac{a}{1-a}-a\right) = \frac{b^2}{1-a}\end{aligned}$$

$$\Rightarrow \mathbf{Var}[X] = 2\frac{b^2}{1-a} - b^2 = b^2\frac{1+a}{1-a}.$$

3. On observe 100 durées de maintien correspondant à l'observation d'un échantillon i.i.d de loi commune une loi prise dans la famille définie ci-dessus. La moyenne empirique calculée sur l'échantillon est égale à 15 et la variance empirique égale à 900.

- Donner une estimation ponctuelle du paramètre *b* puis du paramètre *a* obtenus par la méthode des moments.
- Montrer que si on suppose $a=0$ l'estimation de *b* est la même celle que l'on obtiendrait par la méthode du maximum de vraisemblance.

Notons $x=(x_1,\dots,x_{100})$ les durées observées.

- $\begin{cases}\tilde{b} = \dfrac{1}{100}\displaystyle\sum_{i=1}^{100} x_i = 15 \\ \tilde{b}^2\dfrac{1+\tilde{a}}{1-\tilde{a}} = \dfrac{1}{100}\displaystyle\sum_{i=1}^{100}(x_i-\bar{x})^2 = 900\end{cases} \Leftrightarrow \begin{cases}\tilde{b} = 15 \\ \tilde{a} = 0{,}6\end{cases}$

- Si $a=0$, la fonction de risque est constante : $h(x)=\frac{1}{b}$. Il s'agit du modèle exponentiel de paramètre $\lambda=\frac{1}{b}$ dont l'estimateur du maximum de vraisemblance en l'absence de données incomplètes est donné par :

$$\hat{\lambda} = \frac{100}{\sum_{i=1}^{100} x_i} \Rightarrow \hat{b} = \frac{1}{100}\sum_{i=1}^{100} x_i.$$

Exercice n°3 (Propriétés des modèles de durée de vie humaine)

En considérant que l'espérance de vie résiduelle est décroissante avec l'âge et que l'espérance de vie est croissante avec l'âge, quelles sont les contraintes que l'on doit imposer à un modèle de survie caractérisé par *e* ?

L'espérance de vie résiduelle est décroissante avec l'âge : $e'(x) \leq 0$.

L'espérance de vie est croissante avec l'âge : $\frac{d}{dx}\left(e(x)+x\right) \geq 0 \Leftrightarrow e'(x) \geq -1$.

Au final, pour tout $x > 0$, $-1 \leq e'(x) \leq 0$.

Dans le même contexte, quel doit être le sens de variation de *h* par rapport à l'âge ? Quelles conditions peut-on en déduire sur les paramètres d'un modèle de Weibull, d'un modèle Gamma et d'un modèle Makeham ?

$$h(x) = \frac{1+e'(x)}{e(x)} \Rightarrow h(x) \geq 0 .$$

De plus le phénomène de vieillissement implique que le risque de décéder augmente avec l'âge soit : $h'(x) \geq 0$.

Modèle de Weibull : $h(x) = \lambda \alpha x^{\alpha-1}$ avec $\lambda, \alpha > 0$.

$h'(x) \geq 0$ pour $\alpha \geq 1$.

Modèle Gamma : $h(x) = \dfrac{x^{r-1}e^{-\lambda x}}{\int_x^\infty t^{r-1}e^{-\lambda t}dt}$ avec $r, \lambda > 0$.

$h'(x) \geq 0$ pour $r \geq 1$.

Modèle Makeham : $h(x) = a + bc^x$ avec $a, b, c \geq 0$.

$h'(x) = b \ln c \, c^x \geq 0$ pour $c \geq 1$.

Exercice n°4 (Fraction polynomiale pour espérance de vie résiduelle)

Soit $x \mapsto e(x)$ une fonction de la forme :

$$x \mapsto e(x) = \frac{a(x+b)}{(x+c)^2},$$

où *a* et *c* sont strictement positifs et *b* positif.

1. Déterminer le graphe de $e(x)$ pour $c > 2b$.

$e'(x) = a\dfrac{c-2b-x}{(x+c)^3}$ donc *e* est croissante sur $[0, c-2b]$ puis décroissante sur $[c-2b, +\infty]$ avec $\lim\limits_{x\to\infty} e(x) = 0$.

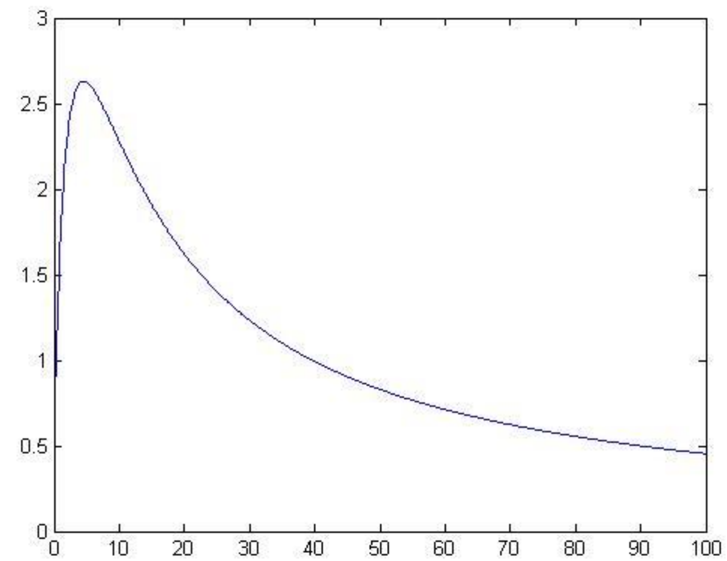

Fig : Graphe de e pour $a = 50 ; b = 0{,}25 ; c = 5$

2. Déterminer les conditions supplémentaires que doivent vérifier *a*, *b* et *c* pour que $e(x)$ soit associée à une fonction de survie ayant bien $e(x)$ comme espérance résiduelle.

3. Montrer que si on se donne comme paramètres

- $\alpha = e(0)$
- $\beta =$ durée pour laquelle $e(x)$ est maximale,
- $\gamma =$ valeur maximale de $e(x)$,

on peut définir les paramètres *a*, *b* et *c*.

Application numérique : $\alpha = 1{,}5 ; \beta = 2{,}5 ; \gamma = 5$
Réponse : $a = 54{,}88 ; b = 0{,}244 ; c = 2{,}988$

$$\begin{cases} \alpha = e(0) = \dfrac{ab}{c^2} \\ \beta = \mathbf{arg\,max}\, e(x) = c - 2b \\ \gamma = \mathbf{max}\, e(x) = e(c-2b) = \dfrac{a}{4(c-b)} \end{cases} \Leftrightarrow \begin{cases} \alpha(\beta+2b)^2 = ab \\ c = \beta + 2b \\ 4\gamma(\beta+b) = a \end{cases}$$

$$4\gamma(\beta + b) = \frac{\alpha(\beta+2b)^2}{b} \Leftrightarrow 4\gamma\beta b + 4\gamma b^2 = \alpha\beta^2 + 4\alpha\beta b + 4\alpha b^2$$

$$\Leftrightarrow 4(\gamma-\alpha)b^2 + 4(\gamma-\alpha)\beta b - \alpha\beta^2 = 0 .$$

Comme $b > 0$, $b = \dfrac{-2(\gamma-\alpha)\beta + 2\beta\sqrt{\gamma(\gamma-\alpha)}}{4(\gamma-\alpha)} = \dfrac{\beta}{2}\left(\sqrt{\dfrac{\gamma}{\gamma-\alpha}} - 1\right)$.

On en déduit $c = \beta\sqrt{\dfrac{\gamma}{\gamma-\alpha}}$ et $a = 2\beta\gamma\left(\sqrt{\dfrac{\gamma}{\gamma-\alpha}} + 1\right)$.

4. Donner l'expression de la fonction de survie *S*.

$$S(x) = \frac{e(0)}{e(x)} \mathbf{exp}\left\{-\int_0^x \frac{dt}{e(t)}\right\}$$

$$\int_0^x \frac{dt}{e(t)} = \int_0^x \frac{(t+c)^2}{a(t+b)}dt = \int_b^{x+b} \frac{(t+c-b)^2}{at}dt = \int_b^{x+b} \frac{t}{a} + 2\frac{c-b}{a} + \frac{(c-b)^2}{at}dt$$

$$= \frac{(x+b)^2 - b^2}{2a} + 2\frac{(c-b)x}{a} + \frac{(c-b)^2}{a}\mathbf{ln}\frac{x+b}{b}$$

$$S(x) = \frac{ab}{c^2}\frac{(x+c)^2}{a(x+b)}\mathbf{exp}\left\{-\frac{x^2+2bx}{2a} - 2\frac{(c-b)x}{a} - \frac{(c-b)^2}{a}\mathbf{ln}\frac{x+b}{b}\right\}$$

$$S(x) = \left(\frac{x+c}{c}\right)^2 \left(\frac{b}{x+b}\right)^{1+\frac{(c-b)^2}{a}} \mathbf{exp}\left\{-\frac{x^2}{2a} + \frac{b-2c}{a}x\right\}.$$

5. Donner l'expression du taux de hasard *h*.

$$h(x) = \frac{1+e'(x)}{e(x)} = \frac{(x+c)^3 + a(c-2b-x)}{(x+c)^3}\frac{(x+c)^2}{a(x+b)} = \frac{(x+c)^3 + a(c-2b-x)}{a(x+c)(x+b)}$$