**Modèles de durée / Examen / Janvier 2023**

**Durée 2h – aucun document n'est autorisé**

## Corrigé

La qualité de la rédaction, des justifications apportées et de la présentation de la copie seront prises en compte dans la notation. Chaque réponse doit être correctement justifiée : une réponse juste sans justification sera considérée comme fausse.

## Quelques propriétés du modèle de Weibull

On 'intéresse au modèle de Weibull ; dont la fonction de hasard est de la forme :

$$\mu(t) = \lambda\alpha t^{\alpha-1},\ \alpha,\lambda > 0$$

**Question n°1 (1 point) :** Calculez la fonction de survie.

> Comme $S(t) = \mathbf{exp}\left(-\int_0^t \mu(s)\,ds\right)$, on trouve $S(t) = e^{-\lambda t^{\alpha}}$

**Question n°2 (3 point) :** Calculez les moments non centrés d'ordre *k*, $E\left(X^k\right)$ à l'aide de la fonction $\Gamma(x) = \int_0^{+\infty} u^{x-1}e^{-u}du$. En déduire l'espérance, la variance et le coefficient de variation. Que peut-on dire de ce dernier ?

> $E\left(T^k\right) = \lambda^{-k/\alpha}\Gamma\left(\frac{k}{\alpha}+1\right)$
>
> $E(T) = \lambda^{-1/\alpha}\Gamma\left(\frac{\alpha+1}{\alpha}\right)$ et $V(T) = \lambda^{-2/\alpha}\left(\Gamma\left(\frac{2}{\alpha}+1\right) - \Gamma^2\left(\frac{1}{\alpha}+1\right)\right)$, ce qui montre que le coefficient de variation $\frac{\sigma(T)}{E(T)}$ ne dépend pas du facteur d'échelle $\lambda$.
>
> Cf. la section 3.2 du support de cours pour les calculs.

**Question n°2 (1 point) :** Démontrez que si la variable *T* est distribuée selon une loi exponentielle de paramètre $\lambda > 0$ alors $T^{1/\alpha}$ suit $W(\alpha,\lambda)$.

> $P\left(T^{1/\alpha} > x\right) = P\left(T > x^{\alpha}\right) = e^{-\lambda x^{\alpha}}$.

**Question n°3 (4 points) :** Prouver que, si $(X_1,\ldots,X_n)$ est un échantillon d'une loi de fonction de répartition G sur $]0,+\infty[$ qui vérifie :

$$\lim_{x\to 0^+}\frac{G(x)}{\lambda x^{\alpha}}=1,$$

alors $n^{1/\alpha}X_{(1)}$ converge en loi lorsque *n* tend vers l'infini vers une distribution $W(\alpha,\lambda)$. On rappelle que $X_{(1)}=\min(X_i, 1\le i\le n)$.

> Cf. la section 3.2 du support de cours.

On s'intéresse maintenant à l'estimation des paramètres $(\alpha,\lambda)$.

**Question n°4 (2 points) :** Trouver une relation entre $(\alpha,\lambda)$ et $\ln(-\ln S(t))$ et en déduire un estimateur graphique de $(\alpha,\lambda)$.

> Comme $S(t)=e^{-\lambda t^{\alpha}}$, on a $\ln(-\ln(S(t)))=\ln(\lambda)+\alpha\ln(t)$ ; si $\hat{S}(t)$ désigne un estimateur empirique de la fonction de survie, les points $\left(\ln(t),\ln\left(-\ln\left(\hat{S}(t)\right)\right)\right)$ doivent donc être approximativement alignés sur une droite d'ordonnée à l'origine $\ln(\lambda)$ et de pente $\alpha$.

**Question n°5 (3 points) :** Rappelez la forme générale de la log-vraisemblance d'un modèle avec une censure aléatoire droite non informative et une troncature gauche. Vous redonnerez la définition de « censure aléatoire droite non informative » et « troncature gauche ».

> $$\ln L(\theta)=\sum_{i=1}^{n}d_i\ln\left(\mu_\theta(t_i)\right)+\ln S_\theta(t_i)-\ln S_\theta(e_i),$$
>
> Cf. les sections 1.2 et 1.5.1 du support de cours.

**Question n°6 (4 points) :** Écrire les conditions du premier ordre du programme de maximisation de la log-vraisemblance. Que peut-on en déduire dans le cas particulier $\alpha=1$ ? Dans le cas général ?

> On trouve, en notant $(e_i,t_i,d_i)$ l'instant d'entrée dans l'observation (troncature gauche), le dernier instant observé et l'indicatrice de non censure :
>
> $$\frac{\partial}{\partial\lambda}\ln L=\frac{d}{\lambda}-\sum_{i=1}^{n}\left(t_i^{\alpha}-e_i^{\alpha}\right)$$
>
> $$\frac{\partial}{\partial\alpha}\ln L=\frac{d}{\alpha}+\sum_{i=1}^{n}d_i\ln(t_i)-\alpha\lambda\sum_{i=1}^{n}\left(t_i^{\alpha-1}-e_i^{\alpha-1}\right)$$

avec $d = \sum_{i=1}^{n} d_i$.

Dans le cas particulier $\alpha = 1$, on retrouve l'estimateur du paramètre d'une loi exponentielle, $\hat{\lambda} = \frac{d}{\sum_{i=1}^{n}\left(t_i - e_i\right)}$. Dans le cas général, les équations de vraisemblance n'admettent pas de solution explicite, il faut se tourner vers une approximation numérique.

**Question n°7 (2 points) :** Décrire un algorithme permettant de résoudre le système d'équations ci-dessus.

La méthode la plus simple est d'utiliser l'algorithme de Newton-Raphson, basé sur la récurrence suivante ;

$$\theta_{k+1} = \theta_k - \left[\frac{\partial^2}{\partial\theta\partial\theta'}\mathbf{ln}\,L\left(\theta_k\right)\right]^{-1}\frac{\partial\,\mathbf{ln}\,L\left(\theta_k\right)}{\partial\theta}$$, avec $\theta = \left(\alpha, \lambda\right)$

Dans le cas particulier de la distribution de Weibull, on a

$$\frac{\partial^2}{\partial\theta\partial\theta'}\mathbf{ln}\,L\left(\theta_k\right) = \begin{bmatrix} -\frac{d}{\alpha^2} - \lambda\sum_{i=1}^{n}\left(t_i^{\alpha-1} - e_i^{\alpha-1}\right) - \alpha\lambda\sum_{i=1}^{n}\left(t_i^{\alpha-1}\,\mathbf{ln}\left(t_i\right) - e_i^{\alpha-1}\,\mathbf{ln}\left(e_i\right)\right) & \alpha\sum_{i=1}^{n}\left(t_i^{\alpha-1} - e_i^{\alpha-1}\right) \\ \alpha\sum_{i=1}^{n}\left(t_i^{\alpha-1} - e_i^{\alpha-1}\right) & -\frac{d}{\lambda^2} \end{bmatrix}$$

pour les valeurs initiales de l'algorithme, on peut par exemple prendre l'estimateur graphique de la question n°4.

Cf. la section 2.4.1 du support de cours.