PRACTICE EXAMEN

Modèle de duréee M1 DUAS– Semestre 2
P.-O. Goffard

---

1. Soient U et V deux variables aléatoires continues, positives et indépendantes de fonction de hasards respectives h_U et h_V. On définit $T = \min(U, V)$. Montrer que la fonction de hasard de T s'écrit
$$h_T(t) = h_U(t) + h_V(t), \text{ pour tout } t \geq 0.$$

Solution: La fonction de survie de T s'écrit
$$S_T(t) = \mathbb{P}(T > t) = \mathbb{P}(\min(U, V) > t) = \mathbb{P}(U > t, V > t) = S_U(t)S_V(t).$$
On en déduit que
$$f_T(t) = -S_t'(t) = f_U(t)S_V(t) + S_U(t)f_V(t).$$
On a alors
$$h_T(t) = \frac{f_T(t)}{S_T(t)} = h_U(t) + h_V(t)$$

2. Soit T une variable aléatoire discrète sur un espace d'état $E = \{t_1, \ldots, t_n\}$ tels que $0 = t_0 < t_1 < t_2 < \ldots < t_n = \infty$. On note
$$h(t_k) = \mathbb{P}(T = t_k | T > t_{k-1}) \text{ pour } k \geq 2.$$
Montrer que la fonction de survie de T peut s'écrire
$$S(t) = \prod_{k: t_k \leq t} [1 - h(t_k)]$$

Solution: Pour $t > 0$, il existe $k < n$ tel que $t \in [t_k, t_{k+1})$
$$\begin{aligned}
S(t) &= \mathbb{P}(T > t) \\
&= \mathbb{P}(T > t_k) \\
&= \mathbb{P}(T > t_k | T > t_{k-1})S(t_{k-1}) \\
&= (1 - \mathbb{P}(T \leq t_k | T > t_{k-1}))S(t_{k-1}) \\
&= (1 - h(t_k))S(t_{k-1}) \\
&= (1 - h(t_k))(1 - h(t_{k-1}))S(t_{k-2}) \\
&= \ldots \\
&= \prod_{k: t_k < t} (1 - h(t_k))
\end{aligned}$$

3. On modélise la durée de vie humaine par une variable aléatoire X de fonction de hasard

$$h(x) = a + bc^x,$$

où $a, b, c > 0$. On note $\theta = (a, b, c)$. On note

$$q_x(\theta) = \mathbb{P}(X \leq x + 1 | X > x)$$

la probabilité de décès à l'âge x. Montrer que dans le cadre du modèle, on a

$$q_x(\theta) = 1 - sg^{c^x(c-1)}$$

où on exprimera s et g en fonction de a, b et c.

Solution: On a

$$\begin{aligned} q_x(\theta) &= \mathbb{P}(X \leq x+1|X > x) \\ &= 1 - \mathbb{P}(X > x+1|X > x) \\ &= 1 - \frac{\mathbb{P}(X > x+1, X > x)}{\mathbb{P}(X > x)} \\ &= 1 - \frac{S(x+1)}{S(x)} \\ &= 1 - \exp\left(-\int_x^{x+1} a + bc^x \mathrm{d}x\right) \\ &= 1 - \exp\left(-a - b\int_x^{x+1} e^{x\log(c)} \mathrm{d}x\right) \\ &= 1 - e^a \exp\left(-\frac{b}{\log(c)} c^x(c-1)\right) \end{aligned}$$

4. Soit $T \sim \mathrm{Par}(\theta)$ une variable aléatoire de fonction de survie donnée par

$$S(t) = \frac{1}{(1+t)^\theta}, \text{ pour } t > 0$$

(a) Donner la densité et la fonction de hasard de T.

Solution: On a

$$f(t) = -S'(t) = \theta(1+x)^{-\theta-1}$$

et

$$h(t = \frac{f(t)}{S(t)} = \frac{\theta}{1+x}$$

(b) Calculer l'espérance de vie résiduelle, définie par

$$e(t) = \mathbb{E}(T - t|T > t).$$

Pour quelle valeur de θ cette espérance de vie résiduelle est bien définie.

Solution: L'espérance de vie résiduelle est définie pour $\theta > 1$ et on a

$$\begin{aligned}
e(t) &= \mathbb{E}(T-t|T>t) \\
&= \frac{\mathbb{E}((T-t)\mathbb{I}_{T>t})}{S(t)} \\
&= (1+t)^{\theta}\int_t^{\infty}(s-t)\theta(1+s)^{-\theta-1}\mathrm{d}s \\
&= \frac{1+t}{\theta-1}.
\end{aligned}$$

(c) Soient $t_1, \dots, t_n$ un échantillon iid suivant T. On suppose que l'échantillon est censuré à droite. Donner une écriture de la vraisemblance du modèle prenant en compte la variable aléatoire de censure, notée C, indépendante de T.

Solution: Soient $c_1, ldots, c_n$ des réalisations iid de C. Les données disponible sont

$$\mathcal{D} = (x_i, \delta_i) = (t_i \wedge c_i, \mathbb{I}_{t_i \leq c_i}),$$

et la vraisemblance du modèle s'écrit

$$\mathcal{L}(\mathcal{D}, \theta) = \prod_{i=1}^{n} h(x_i)^{\delta_i} S(x_i)$$

(d) Supposons que $c_i = c$ pour tout $i = 1, \dots, n$. Donner l'expression du maximum de vraisemblance

Solution: On résout

$$\frac{\partial}{\partial\theta}\log\mathcal{L}(\mathcal{D}, \theta) = 0$$

et on obtient

$$\widehat{\theta} = \frac{r}{\sum_i x_i}.$$

où $r = \displaystyle\sum_{i=1}^{n}\delta_i$. On vérifie que

$$\frac{\partial^2}{\partial\theta^2}\log\mathcal{L}(\mathcal{D}, \widehat{\theta}) = -\frac{r}{\widehat{\theta}^2}.$$

(e) Donner un intervalle de confiance pour $\widehat{\theta}$.

Solution: On a

$$\hat{\theta} \sim \mathsf{Normal}\left[\theta, -\left(\frac{\partial^2}{\partial\theta^2}\log\mathcal{L}(\mathcal{D}, \theta)\right)^{-1}\right], \text{ pour } n \to \infty$$

(f) Supposons que la variable aléatoire de censure vérifie $C \sim \mathrm{Par}(\beta\theta)$ pour $\beta > 0$. Calculer

$$\mathbb{P}(T > C),$$

qui correspond à la probabilité qu'une observation soit censuré.

Solution:

$$\frac{\beta}{\beta + 1}$$

(g) Ecrire la vraisemblance du modèle en suppsosant le paramètre β inconnu. Donner l'estimater du maximum de vraisemblance de θ et β.

Solution: La vraisemblance s'écrit

$$\mathcal{L}(\mathcal{D}, \theta, \beta) = \prod_{i=1}^{n} (f(x_i; \theta)^{\delta_i} S(x_i; \beta\theta))^{\delta_i} (f(x_i; \beta\theta)^{\delta_i} S(x_i; \theta))^{1-\delta_i}.$$

On résout les équations du score pour obtenir

$$\widehat{\beta} = \frac{n}{r} - 1, \text{ et } \widehat{\theta} = \frac{r}{\sum_i \log(1 + x_i)}.$$