

## Modèles de durée / Examen du 20 janvier 2012

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

### Problème : mesure de l'incertitude sur des taux de mortalité

On considère une variable aléatoire  $X$  décrivant une durée de vie et on note  $S$  et  $h$  respectivement la fonction de survie et la fonction de hasard associées. On utilise les notations usuelles, *ie*  $T_i = X_i \wedge C_i$  et  $D_i = 1_{\{X_i \leq C_i\}}$  avec  $C_i$  la variable de censure. On rappelle que l'estimateur de Greenwood de la variance de l'estimateur Kaplan-Meier de la fonction de survie  $S$  est :

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \gamma(t)^2$$

où  $\gamma(t) = \sqrt{\sum_{T_i \leq t} \frac{d_i}{r_i(r_i - d_i)}}$  avec  $d_i$  le nombre de sorties non censurées observées à l'instant  $T_i$  et  $r_i$  l'exposition au risque. On rappelle également l'approximation de la variance de  $f(X)$  obtenue par la méthode Delta :

$$V(f(X)) \approx \left( \frac{df}{dx}(E(X)) \right)^2 V(X).$$

**Question n°1 (2 points)** : Donner une justification de l'approximation ci-dessus et préciser dans quel cadre elle est valable.

**Question n°2 (3 points)** : En partant de l'expression suivante de l'estimateur de Kaplan-Meier,  $\hat{S}(t) = \prod_{T_i \leq t} \left(1 - \frac{d_i}{r_i}\right)$ , donner une justification de la formule de Greenwood (on pourra considérer d'abord  $\ln(\hat{S}(t))$ )

**Question n°3 (2 points)** : Rappeler le lien entre la probabilité conditionnelle de sortie  $q_x$  et la fonction de survie  $S(x)$ . En déduire un estimateur Kaplan-Meier de la probabilité conditionnelle de sortie,  $\hat{q}(x)$ . Quel est le lien entre  $q_x$  et la fonction de survie de la variable  $X_x = (X - x) | X > x$  ?

**Question n°4 (3 points)** : Déduire des questions précédentes un estimateur de  $\hat{V}(\hat{q}(x))$

**Question n°5 (3 points)** : Comment déduire de la question précédente un intervalle de confiance (asymptotique) au niveau  $\alpha$  pour  $q_x$ ? Donner l'expression explicite d'un tel intervalle.

On considère maintenant le nombre de sorties par décès sur  $[x, x+1[$ , que l'on note  $D_x$  et l'exposition au risque  $E_x$ .

**Question n°6 (2 points)** : Rappelez la définition (mathématique) de  $E_x$  et le lien qui existe entre  $E_x$  et  $D_x$ .

**Question n°7 (3 points)** : Proposer dans ce cadre un estimateur non paramétrique de  $q_x$ , dont on calculera l'espérance et la variance. Quelle est la loi asymptotique de cet estimateur lorsque  $E_x$  est grand ?

**Question n°8 (2 points)** : Déduire de la question précédente un intervalle de confiance (asymptotique) au niveau  $\alpha$  pour  $q_x$ .