Université Claude Bernard - Lyon 1
ISFA - M2SAF
Frédéric Planchet et Maxime Ben-Brik

Année 2021-2022

Travaux dirigés : modèles de durée
Séance n°4

Exercise 1 Manipulation de l'estimateur de Kaplan-Meier.

Soit un échantillon de n individus i.i.d. de durée de vie $T_1, \ldots, T_n$. Chaque observation i est soumise à censure à droite C_i, supposée indépendante et non-informative. On observe (Y_i, D_i) pour chaque i, où $Y_i = T_i \wedge C_i$ et $D_i = \mathbb{1}_{\{T_i \leq C_i\}}$. On note S la fonction de survie de T et G celle de C.

1. En notant $H_1(t) = \mathbb{P}(Y > t, D = 1)$ et $H(t) = \mathbb{P}(Y > t)$, montrer que
$$\frac{dS(t)}{S(t-)} = \frac{dH_1(t)}{H(t-)}.$$
2. Retrouver à partir de cette expression, l'estimateur non-paramétrique de Nelson-Aalen pour la fonction de hasard cumulée $\Lambda(t)$.
3. Rappeler l'expression de la fonction de hasard $h(t)$, de la fonction de hasard cumulée $\Lambda(t)$ et de la fonction de survie d'une loi discrète. En déduire l'expression de l'estimateur de Kaplan-Meier de la fonction de survie.
4. En suivant un raisonnement similaire avec $H_1(t) = \mathbb{P}(Y > t, D = 1 | Y > L)$ et $H(t) = \mathbb{P}(L \leq t < Y | Y > L)$, donner l'expression de l'estimateur de Kaplan-Meier avec troncature à gauche L indépendante (T indépendant de (L, C) et C indépendant de L).
5. En utilisant directement l'expression de la première question, exprimer l'estimateur de Kaplan-Meier sous la forme d'une somme faisant intervenir l'estimateur de G.

Exercise 2 Variance de l'estimateur de Kaplan-Meier.

Soit un échantillon de n individus i.i.d. de durée de vie $T_1, \ldots, T_n$. Chaque observation i est soumise à censure à droite C_i, supposée indépendante et non-informative. On observe (Y_i, D_i) pour chaque i, où $Y_i = T_i \wedge C_i$ et $D_i = \mathbb{1}_{\{T_i \leq C_i\}}$. On note S la fonction de survie de T et G celle de C.

On rappelle que l'estimateur de Kaplan-Meier vérifie
$$\sqrt{n}\left(\widehat{S} - S\right) \xrightarrow[n \to \infty]{\mathcal{L}} \boldsymbol{U} ,$$
avec $\boldsymbol{U}$ un processus gaussien centrée de variance-covariance
$$\rho(s,t) = -S(s)\, S(t) \int_0^{s \wedge t} \frac{dS(u)}{S(u)^2\, G(u)}.$$

1. On note $H_1(t) = \mathbb{P}(Y > t, D = 1)$ et $H(t) = \mathbb{P}(Y > t)$. Puisque

$$\frac{dS(t)}{S(t-)} = \frac{dH_1(t)}{H(t-)} \, ,$$

donner un estimateur de la variance de $\widehat{S}(t)$.

2. En notant $Y_{(i)}$ la statistique d'ordre i des Y et $D_{[i]}$ la valeur de D associée, réécrire cet estimateur de la variance et vérifier qu'il correspond à l'estimateur de Greenwood.
3. Proposer un intervalle de confiance de niveau α pour l'estimateur de Kaplan-Meier.
4. Déterminer la distribution asymptotique de $\widehat{p}_t = \dfrac{\widehat{S}(t+1)}{\widehat{S}(t)}$, puis en déduire l'estimateur de sa variance (de type Greenwood).