Université d'Angers

Master Data Science – Statistique et science des données pour la biologie

Examen Final (2h)

Le barème est donné à titre indicatif.
Les exercices sont indépendants.

Il sera tenu compte de la rédaction.

Introduction à l'Analyse de Survie

Exercice 1. (3 points)
Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes, continues et non-négatives de fonctions de risque $\lambda_1, \lambda_2, \ldots, \lambda_n$. Montrer que $X = \min\{X_1, X_2, \ldots, X_n\}$ a pour fonction de risque $\sum_{i=1}^n \lambda_i$.

Exercice 2. (5 points)
On suppose que le temps de survie X admet pour densité

$$f(t) := \begin{cases} 0 & \text{si } t > \tau, \\ \alpha\beta(\tau - t)^{\alpha-1} \exp(-\beta(\tau - t)^\alpha) & \text{si } t \leqslant \tau, \end{cases}$$

où α, β et τ sont des paramètres strictement positifs.

On définit par C la censure aléatoire et on suppose que l'on travaille avec un modèle de censure à gauche, c'est-à-dire, que l'on observe les instants $T := \max(X, C)$ muni de l'indicateur de censure $\delta := \mathbb{1}_{\{X \geqslant C\}}$. On note aussi par la suite $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$ un échantillon aléatoire *i.i.d.* issu de (T, δ).

1. Calculer les fonctions de survie S et de risque instantané λ associées à X.
2. Montrer que la variable aléatoire X suit la même loi que $\tau - Y$ où Y est une variable aléatoire dont on précisera la loi.

On suppose à présent que $C \leqslant \tau$ presque sûrement.

3. Proposer modèle de censure à droite à partir de T et δ.
4. Montrer que $\widehat{\tau}_n := \max(T_i,\, i = 1, \ldots, n)$ est un estimateur consistant de τ.
5. Déduire des questions précédentes une procédure d'estimation pour les paramètres τ, α et β à partir de l'échantillon $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$. On ne demande pas de montrer la consistance de chacun des estimateurs.

Exercice 3. (7 points)
On suppose que le temps de survie X admet pour densité

$$f(t) := \begin{cases} \beta \exp(-\beta(t-\tau)) & \text{si } t \geqslant \tau, \\ 0 & \text{si } t < \tau, \end{cases} \tag{1}$$

où $\tau, \beta > 0$.

On définit par C la censure aléatoire et on suppose que l'on travaille avec un modèle de censure à droite, c'est-à-dire, que l'on observe les instants $T := \min(X, C)$ muni de l'indicateur de censure $\delta := \mathbb{1}_{\{X \leqslant C\}}$. On note aussi par la suite $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$ un échantillon aléatoire *i.i.d.* issu de (T, δ). On suppose enfin que $C \geqslant \tau$ presque sûrement.

Partie A. On cherche dans un premier temps à utiliser une approche graphique pour justifier l'adéquation du modèle aux données.

1. Montrer que la fonction de survie associée à X est donnée par $S(t) = \exp(-\beta(t-\tau))$ pour $t \geqslant \tau$ et 1 sinon.

2. Soit $\widehat{S}_n(t)$ l'estimateur de Kaplan-Meier pour la fonction de survie S. Justifier et montrer que

$$\arg\min_{t \in \mathbb{R}} \left\{ \widehat{S}_n(t) < 1 \right\}$$

   définit un estimateur consistant pour τ.

3. Proposer une transformation de $\widehat{S}_n$ pour justifier l'adéquation de X à la distribution décrite par f dans (1).

Partie B. On se propose à présent d'estimer les paramètres par maximum de vraisemblance.

4. Calculer la fonction de log-vraisemblance pour la distribution (1).

5. En déduire les estimateurs de maximum de vraisemblance $\widehat{\tau}_n$ et $\widehat{\beta}_n$.

Plan d'expérience

Exercice 4. (5 points)
Les deux questions sont indépendantes.

1. Montrer que, dans un plan fractionnaire de résolution IV, il n'est pas possible de confondre un effet marginal avec une interaction double mais que deux interactions doubles peuvent être confondues.

2. Reprenons l'exemple traité dans la section 2.2 du cours, en changeant simplement la variable à expliquer : on veut étudier l'influence de la température du mélange (F_1), de la quantité produite (F_2), de la vitesse d'agitation (F_3), de la durée de stockage (G_1) et de la nature du catalyseur (G_2) sur une certaine propriété chimique d'un produit synthétisé en laboratoire (Y), avec comme contrainte que 8 expériences au maximum peuvent être réalisées. Notre plan fractionnaire 2^{5-2} est cette fois muni des clés

$$G_1 = F_1 * F_2 \quad \text{et} \quad G_2 = F_2 * F_3.$$

   a) Écrire les relations complètes. En déduire la résolution et l'aberration du plan.

   b) Le tableau ci-dessous donne les mesures obtenues pour Y.

Expérience	Température (F_1)	Quantité (F_2)	Vitesse (F_3)	Stockage (G_1)	Catalyseur (G_2)	Propriété obtenue (Y)
1	-1	-1	-1			11.2
2	1	-1	-1			16.6
3	-1	1	-1			-1.0
4	1	1	-1			13.2
5	-1	-1	1			12.2
6	1	-1	1			16.8
7	-1	1	1			-2.8
8	1	1	1			14.6

i) Indiquer les niveaux (± 1) à affecter aux facteurs additionnels (on recopiera simplement les deux colonnes sur la copie, pas le tableau complet).

ii) Écrire le modèle liant Y aux facteurs avec ces nouvelles clés (sous la forme $Y \sim e_0 + \ldots$).

iii) Donner l'alias de chaque facteur (de F_1 à G_2) en se limitant aux interactions doubles.

iv) Le vecteur des solutions au système s'écrit

$$\begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_{12} \\ e_{13} \\ e_{23} \\ e_{123} \end{pmatrix} = \begin{pmatrix} 10.1 \\ 5.2 \\ -4.1 \\ 0.1 \\ 2.7 \\ 0.3 \\ -0.2 \\ 0.5 \end{pmatrix}.$$

Identifier les trois facteurs qu'il serait raisonnable de retenir dans le modèle. Peut-on déterminer les effets marginaux de ces trois facteurs ? Si oui, pourquoi ? Et si non, que préconiser au fabricant ?