

# TD 1: APPROCHE PARAMÉTRIQUE DE LA SURVIE.

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

---

1. L'objectif est de déterminer l'estimateur du maximum de vraisemblance du maximum de vraisemblance de la loi exponentielle  $\text{Exp}(\beta)$  de densité

$$f(t) = \frac{e^{-t/\beta}}{\beta} \mathbb{I}_{(0,\infty)}(t).$$

Soit  $\mathcal{D} = \{t_1, \dots, t_n\}$  un échantillon i.i.d. de réalisation d'une v.a. exponentielle.

- (a) Donner l'estimateur du maximum de vraisemblance  $\hat{\theta}$ .

**Solution:** La log vraisemblance s'écrit

$$l(\mathcal{D}; \theta) = -n \log(\beta) - \frac{1}{\beta} \sum t_i.$$

On a

$$\frac{\partial l}{\partial \theta}(\mathcal{D}; \theta) = -n \frac{n}{\beta} + \frac{1}{\beta^2} \sum t_i.$$

On en déduit que

$$\frac{\partial l}{\partial \theta}(\mathcal{D}; \theta) = 0 \Leftrightarrow \hat{\beta} = \frac{\sum t_i}{n},$$

après avoir vérifié que

$$\frac{\partial l}{\partial \theta}(\mathcal{D}; \hat{\theta}) < 0.$$

- (b) Supposons que nous soyons en présence de données censurées, avec une censure à droite, non informative, de niveau  $c_1, \dots, c_n$ . Donner l'estimateur du maximum de vraisemblance de  $\beta$ .

**Solution:** Les observations sont

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

et la log vraisemblance s'écrit

$$-\sum \delta_i \log(\beta) - \frac{\sum x_i}{\beta} + \text{Cste.}$$

On en déduit que

$$\frac{\partial l}{\partial \theta}(\mathcal{D}; \theta) = 0 \Leftrightarrow \hat{\beta} = \frac{\sum x_i}{\sum \delta_i}.$$

On peut vérifier la cohérence de l'estimation. En prenant  $C \rightarrow \infty$ , on retombe sur l'estimateur obtenu à la question 1.

2. Nous allons construire un test d'adéquation à la loi exponentielle  $\text{Exp}(\beta)$  de densité

$$f(t) = \frac{e^{-t/\beta}}{\beta} \mathbb{I}_{(0,\infty)}(t),$$

basé sur la transformée de Laplace. La transformée de Laplace d'une v.a.  $T$  est donnée par

$$\psi(\theta) = \mathbb{E}(e^{-\theta T}).$$

- (a) Soit  $t_1, \dots, t_n$  un échantillon i.i.d. de réalisation de  $T \sim \text{Exp}(\beta)$ . L'estimateur du maximum de vraisemblance de  $\beta$  est donnée par

$$\hat{\beta}_n = \frac{1}{n} \sum_{i=1}^n t_i.$$

Quelle est la distribution de  $T/\hat{\beta}_n$  lorsque  $n \rightarrow \infty$ ?

**Solution:**  $T/\hat{\beta}_n \xrightarrow{D} \text{Exp}(1)$

- (b) Donner l'expression de la transformée de Laplace  $\psi(\theta)$  de  $Y \sim \text{Exp}(\beta = 1)$  et montrer qu'elle vérifie l'équation différentielle suivante

$$\psi'(\theta)(1 + \theta) + \psi(\theta) = 0$$

**Solution:** On a

$$\psi(\theta) = \frac{1}{1 + \theta}$$

et

$$\psi'(\theta) = -\frac{1}{(1 + \theta)^2} = -\frac{1}{1 + \theta} \psi(\theta)$$

- (c) On définit  $y_i = t_i/\hat{\beta}_n$  et

$$\psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n e^{-\theta y_i}$$

A quoi correspond cette quantité?

**Solution:** Il s'agit d'un estimateur asymptotiquement sans biais de la transformée de Laplace d'une loi  $\text{Exp}(1)$ .

- (d) On définit la statistique de test suivante

$$S_n = n \int_0^\infty [\psi'_n(\theta)(1 + \theta) + \psi_n(\theta)]^2 e^{-\theta} d\theta,$$

Quel est votre interprétation de cette statistique de test?

**Solution:** On retrouve dans le carré sous l'intégrale, l'équation différentielle satisfaite par  $\psi$  dans le cas où les données suivent une loi exponentielle. On intègre pour prendre en compte l'ensemble du domaine de définition de la transformée de Laplace. On ajoute la fonction exponentielle pour rendre la statistique de test intégrable. L'avantage est d'obtenir un test explicite, voir la question d'après.

(e) Montrer que l'on peut estimer  $S_n$  par

$$S_n = \frac{1}{n} \sum_{j,k=1}^n \left[ \frac{(1-y_j)(1-y_k)}{y_j + y_k + 1} - \frac{y_j + y_k}{(y_j + y_k + 1)^2} + \frac{2y_j y_k}{(y_j + y_k + 1)^2} + \frac{2y_j y_k}{(y_j + y_k + 1)^3} \right].$$

**Indication:** Il faut développer le carré sous l'intégrale et calculer chaque terme séparément. C'est un peu fastidieux mais on y arrive.

**Solution:** Il s'agit d'un petit calcul intégral, je vous fait confiance.

(f) Illustrer ce test avec R (comparer son efficacité à celui de Kolmogorov-Smirnov).

On fixe le niveau du test à 0.05. Calculer la probabilité de rejeter l'hypothèse

$$(H_0) : T \sim \text{Exp}(\beta)$$

lorsque les données sont des réalisations i.i.d. de loi  $\text{Gamma}(\alpha, 1)$ . Il s'agit de la puissance du test. Evaluer la puissance du test pour  $\alpha = 1/4, 1/2, 3/4, 1, 5/4, 3/2, 7/4$ . et  $n = 50$ . On fera le graphique de la puissance en fonction de  $\alpha$  avec une courbe pour le test présenté dans l'exercice (baptisé LT test) et le test de Kolmogorov.

**Indications:** Voici les étapes de l'algorithme

1. Simuler  $t_i \sim \text{Gamma}(\alpha, 1)$  pour  $i = 1, \dots, n$
2. Estimer  $\beta$  par  $\hat{\beta}$
3. Définir  $y_i = t_i / \hat{\beta}$  et calculer  $S_n$
4. Simuler  $\tilde{t}_i \sim \text{Exp}(\hat{\beta})$
5. Estimer  $\beta$  par  $\tilde{\beta}$
6. Définir  $\tilde{y}_i = \tilde{t}_i / \tilde{\beta}$  et calculer  $\tilde{S}_n$  sur la base des  $\tilde{y}_i$

Répéter ces étapes  $J = 1000$  fois. On a une suite de valeur de test statistiques  $S_n^j$  et  $\tilde{S}_n^j$  pour  $j = 1, \dots, J$ . La valeur critique du test est donnée par

$$S_{0.95} = \text{Quantile}(\tilde{S}_n^j, j = 1, \dots, J; 0, 95),$$

qui correspond au quantile empirique d'ordre 95% des  $\tilde{S}_n^j$ . La puissance du test est alors

$$\frac{1}{J} \sum_{j=1}^J \mathbb{I}_{S_n^j > S_{0.95}}.$$

Vous devriez obtenir un résultat proche de celui de la Figure 1.

(g) Commenter le résultat de la Figure 1.

- La forme de la courbe est elle celle attendue?
- Quel est le meilleur test?

Pour plus d'information sur ce test, on pourra se référer au travail de Henze et Meintanis [1]

**Solution:**

- La forme de la courbe est celle attendue, pour  $\alpha = 1$  les données provienne du modèle exponentielle, la puissance des test doit donc atteindre le niveau fixé à 0.05 et augmenté de part et d'autre de  $\alpha = 1$

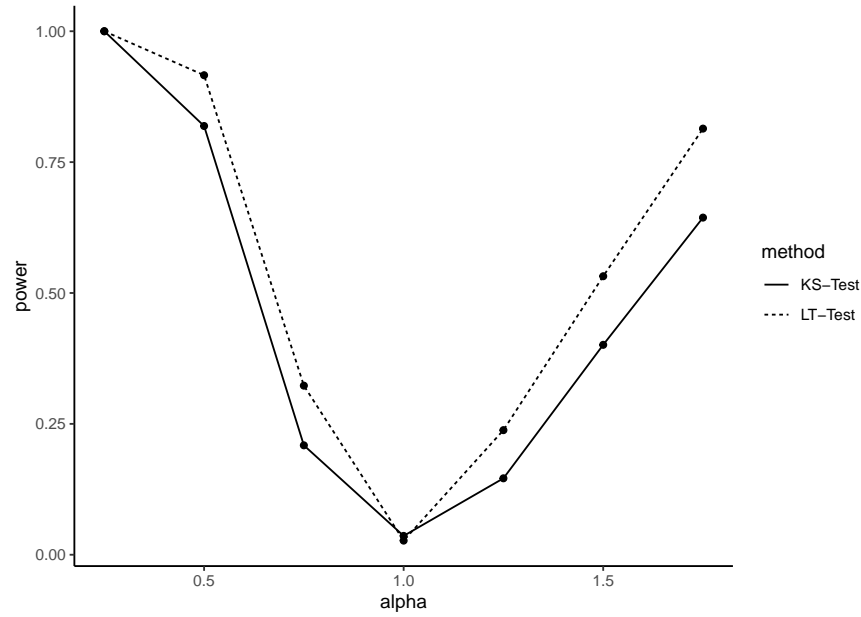


Figure 1: Puissance du LT test et du test de K-S.

- Le LT test semble supérieur. Il est tout à fait possible qu'un résultat inverse soit obtenu si les données proviennent d'un modèle de Weibull par exemple. On peut conclure que le LT test semble meilleur pour discriminer des données qui proviennent de la loi gamma.

## References

- [1] Norbert Henze and Simos G. Meintanis. Tests of fit for exponentiality based on the empirical laplace transform. *Statistics*, 36(2):147–161, jan 2002.