M1- Mathématiques Appliquées - ENSTA / Paris-Sud
Modélisation statistique (MAP-STA1), 2016-2017
christine.keribin@math.u-psud.fr

Examen 17 janvier 2017 (9h00-12h00)

L'examen est en deux parties, une partie sur feuille sans ordinateur (Exercices 1 et 2), et une partie TP sur ordinateur. Dans tous les cas, les résultats doivent être justifiés. La notation prendra en compte la clarté et le soin de la rédaction. Tout résultat non démontré d'une question précédente peut être admis pour utilisation dans une question suivante.

Exercice 1 *(environ 7 points)*

On suppose que la durée de vie d'un matériel est une variable aléatoire X qui suit une loi de Weibull $\mathcal{W}(\lambda, 1/2)$ de densité

$$f(x, \lambda) = \frac{1}{2\lambda\sqrt{x}} \exp\left(-\frac{\sqrt{x}}{\lambda}\right) \mathbb{1}_{x>0}$$

1. Montrer que $\sqrt{X}$ suit une loi exponentielle d'espérance à déterminer, et en déduire que si $(X_1, \ldots, X_n)$ est un n-échantillon i.i.d. de loi $\mathcal{W}(\lambda, 1/2)$, alors $2\sum_i \sqrt{X_i}/\lambda \sim \chi^2(2n)$.
2. La loi de Weibull $\mathcal{W}(\lambda, 1/2)$ appartient-elle à la famille exponentielle?

   On considère dans la suite de l'exercice un n-échantillon i.i.d. $(X_1, \ldots, X_n)$ de loi $\mathcal{W}(\lambda, 1/2)$.
3. Déterminer une statistique exhaustive minimale pour estimer λ. Est-elle complète?
4. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\lambda}$:
   (a) L'estimateur est-il UVMB?
   (b) Quelle est sa loi (non asymptotique)?
   (c) L'estimateur est-il efficace?
5. Construire un test UPP pour choisir entre les deux hypothèses $H_0 : \lambda \leq a$ contre $H_1 : \lambda > a$ au niveau α.
6. La p-value du test précédent calculée à partir de l'échantillon vaut 0.5. Rappeler la définition de la p-value. Quelle décision prenez-vous au niveau $\alpha = 5\%$? Avec quel risque?
7. Calculer la fonction puissance et donner son allure pour n=5, 10, 100. La procédure de test est-elle consistante?

Rappel: la densité d'une loi du $\chi^2(k)$ est $f_\chi(x) = \frac{2^{-k/2}}{\Gamma(k/2)} x^{k/2-1} \exp(-x/2)$ pour $x \geq 0$, et celle d'une loi $\Gamma(n, \theta)$ est $f_\Gamma(x) = \frac{\theta^{-n}}{\Gamma(n)} x^{n-1} \exp(-x/\theta)$ pour $x \geq 0$

Exercice 2 *(environ 3 points)*

De façon générale, la loi de Weibull $\mathcal{W}(\lambda,\beta)$ dépend de deux paramètres λ et β, et sa densité s'écrit:

$$f(x;\lambda,\beta) = \frac{\beta}{\lambda} x^{\beta-1} \exp\left(-\frac{x^\beta}{\lambda}\right) 1\!\!1_{x>0}$$

L'exercice 1 a détaillé le cas particulier où $\beta = 1/2$. On considère un n-échantillon i.i.d. $(X_1, \ldots, X_n)$ de loi $\mathcal{W}(\lambda,\beta)$.

1. La loi $\mathcal{W}(\lambda,\beta)$ fait-elle partie de la famille exponentielle?
2. On considère l'estimateur du maximum de vraisemblance $\widehat{\theta} = (\widehat{\lambda}, \widehat{\beta})$. Quelle est sa loi asymptotique ? Vous justifierez la forme de la loi et préciserez comment calculer sa variance (dont on ne demande pas le calcul explicite).
3. La médiane vaut $\nu = (\lambda \log 2)^{1/\beta}$. Soit $b > 0$. Construire le test de Wald de $\nu = b$ contre $\nu \neq b$. Vous indiquerez soigneusement comment les expressions utilisées dépendent des paramètres ou de leur estimation et vous développerez le calcul.
4. Construire une région de confiance de type Wald pour $\theta = (\lambda, \beta)$ de niveau $1 - \alpha$.

Partie TP

Cette partie est à traiter avec le logiciel R. Vous avez accès à vos notes de cours, aide-mémoire R et codes informatiques de TP.
Vous devez remettre votre travail sous la forme d'un fichier `NOM.pdf` *comprenant les figures demandées, et d'un fichier* `NOM.R` *de commandes R commentées, à envoyer sur la messagerie du chargé de TD qui vous sera précisé à l'issue de l'examen .*
Remarque: Pour enregistrer toutes les figures dans un seul pdf, il suffit d'encadrer le code concerné par les commandes suivantes:

```
pdf ("NOM.pdf")
...
dev.off()
```

Exercice 3 *(environ 10 points)*

Des procédés de hautes températures sont souvent utilisés pour traiter des pièces métalliques telles que les engrenages, en modifiant leurs caractéristiques de surface. La profondeur de trempe est une propriété importante de la pièce, et contribue à sa fiabilité. Des essais destructifs sont effectués en découpant la pièce pour mesurer la profondeur de trempe au niveau des dents de l'engrenage. Un ingénieur méthodes et procédés cherche à expliquer la profondeur de trempe `y` en fonction de quatre variables `soaktime`, `soakpct`, `difftime`, `diffpct`[1].

1. Lire les données `ExamSTA1Exam2016-2017.csv` et vérifier que le dataframe obtenu comporte 32 observations et 5 variables quantitatives. Les représenter et commenter.

2. Ajuster une régression linéaire (modèle M) de l'épaisseur `y` en fonction des quatre variables explicatives.

   (a) Préciser les hypothèses sur le modèle.

   (b) La régression est-elle globalement significative au niveau 5%? Justifier en précisant le test (hypothèses, statistique de test et loi sous H_0, la région de rejet et décision)

   (c) Pour la variable `soaktime`, retrouver le calcul de la `t value` et de la p-value en fonction de l'estimation et de l'erreur standard. Cette variable est-elle significative au niveau 5%?

   (d) Commenter la validité du modèle. Vous représenterez l'ensemble des graphes qui vous semblent pertinents à étudier sur une seule fenêtre divisée en sous-régions.

3. On considère le modèle (M2) qui ne contient que les deux variables `soaktime` et `difftime`. Quel modèle retenez-vous?

4. On considère maintenant les variables `x1=soaktime*soakpct` et `x2=difftime*diffpct`.

   (a) Ajouter ces variables au dataframe.

   (b) Ajuster ce nouveau modèle (M3) n'utilisant que ces deux régresseurs (et l'intercept).

   (c) Est-il possible de tester ce modèle contre les autres?

5. Pour les trois modèles, indiquer le R^2, la variance du bruit, et calculer les intervalles de confiance de niveau 95% de la moyenne de l'épaisseur pour `soaktime=1.`, `soakpct=1.1`, `difftime=1.0`, `diffpct=.8`. Ceci permet-il de guider votre choix de modèle?

[1] Generalized Linear Models. R. Myers, D. Montgomery, G. Vining. Wiley 2002