Université de Strasbourg — Cours : Ségolen Geffray
L3 - Mathématique — TD : Myriam Maumy

Examen de Statistique Mathématique – Session de printemps – 17 mai 2011

Les parties I (exercices 1 à 3) et II (exercices 4 à 6) doivent être rédigées sur des copies séparées.
La qualité de la rédaction sera très largement prise en compte.
Les téléphones portables doivent être éteints et rangés.
L'usage des calculatrices n'est pas autorisé.
Les notes de cours et autres documents sont interdits.

Partie I

Exercice 1.
Les questions de cet exercice sont indépendantes.

1. Donner la définition d'un modèle exponentiel général non courbe et d'un modèle exponentiel canonique dans le cas unidimensionnel. Recommencer dans le cas bidimensionnel.
2. Les modèles suivants appartiennent-ils à la famille exponentielle ? Justifier. Préciser, dans chaque cas, l'espace des paramètres et le support de la loi considérée.
   (a)
   $$f_{\alpha,\lambda}(x) = \alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha})I(x>0), \quad \alpha>0,\, \lambda>0$$
   (b)
   $$f_{\mu}(x) = \frac{1}{2}\exp(-|x-\mu|), \quad \mu\in\mathbb{R}$$
3. Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. Exhiber une statistique exhaustive pour (μ, σ^2) dans le modèle à n observations. Justifier.

Exercice 2.
Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de loi admettant une densité f_θ. Déterminer l'estimateur du maximum de vraisemblance de θ, lorsqu'il existe, dans les modèles suivants :

1.
$$f_\theta(x) = \frac{\exp\left(-\frac{|x-\theta|}{2}\right)}{2\sqrt{2\pi|x-\theta|}}, \quad \theta\in\mathbb{R}$$
2.
$$f_\theta(x) = I(-\theta\le x\le\theta), \quad \theta>0.$$

Exercice 3.
Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de variable parente X de loi de Poisson $\mathcal{P}(\lambda)$ dont la fonction de densité par rapport à la mesure de comptage sur $\mathbb{N}$ est rappelée ci-dessous :
$$\mathbb{P}[X=x] = \frac{\lambda^x\exp(-\lambda)}{x!}, \quad \lambda>0.$$

3. Donner un exemple de situation pratique dans laquelle la loi de Poisson peut être utilisée.
4. Déterminer l'estimateur du paramètre λ obtenu par la méthode des moments et étudier ses propriétés.
5. Construire deux intervalles de confiance différents pour λ, tous deux bilatéraux et symétriques et de niveau de confiance $(1-\alpha)$ où α est un réel de $[0; 1]$.

Partie II

Exercice 4.
Un barrage doit être construit dans une région montagneuse de façon à réguler le débit d'une rivière alimentée par la fonte des neiges et par les pluies. Un hydrologue souhaite déterminer la hauteur de la digue de façon à ce que les terres cultivées en aval ne soient pas inondées. Il dispose du relevé des maxima annuels du fleuve à l'endroit de la future construction entre le 1er janvier 1970 et le 31 décembre 2009 exprimés en mètre (m).

1. L'hydrologue trace l'histogramme des données (Figure 1), la boîte à moustaches correspondante (Figure 2) et estime les quantités suivantes :
   - $\widehat{\mu}$=2,55 ;
   - $\widehat{\sigma}^2$=2,67 ;
   - $\widehat{\gamma}_1$=0,94 ;
   - $\widehat{\gamma}_2$=3,32 ;

   en notant $\widehat{\mu}$ l'estimation de la moyenne, $\widehat{\sigma}^2$ l'estimation débiaisée de la variance, $\widehat{\gamma}_1$ l'estimation du coefficient d'asymétrie et $\widehat{\gamma}_2$ l'estimation du coefficient d'aplatissement. Le

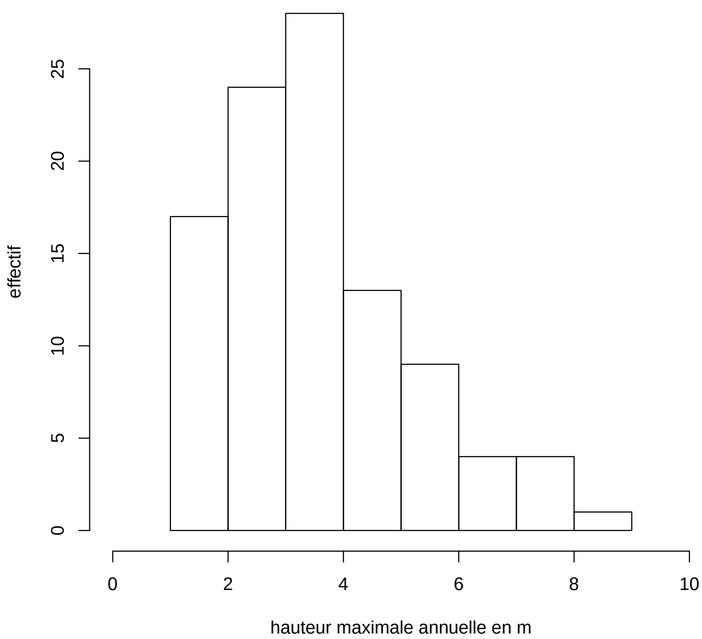

FIGURE 1 – Histogramme des données de maxima annuels

statisticien lui fait remarquer que la normalité des données est peu plausible. Pourquoi ?

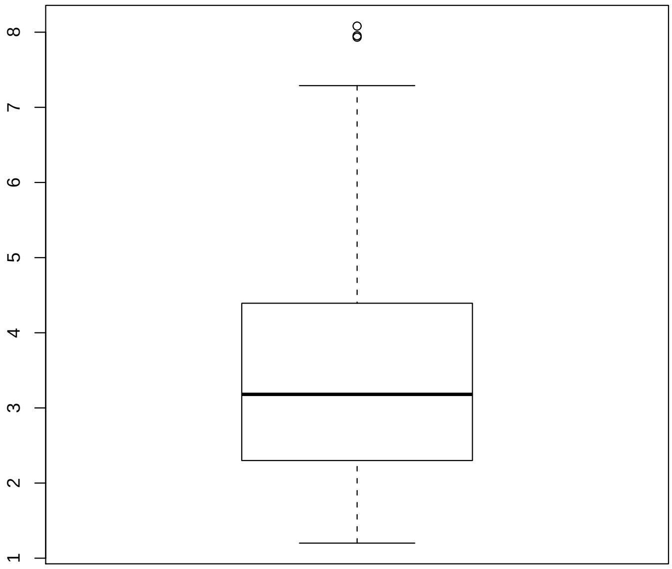

FIGURE 2 – Boîte à moustaches des données de maxima annuels

2. L'hydrologue pense que le niveau maximal de la rivière observé une année donnée n'influe pas sur le niveau maximal de la rivière observé l'année suivante. Il pense également que le fleuve n'a pas subi de changement de régime au cours de la période d'observation (entre le 1er janvier 1970 et le 31 décembre 2009). Par ailleurs, le statisticien effectue un test invalidant la normalité des données. Qu'est-ce que cela implique pour les données ?
3. Le statisticien propose un estimateur respectivement de la moyenne et de la variance du maximum annuel de la rivière en cette région et en détermine la loi jointe asymptotique. Que fait-il ?
4. Le statisticien propose un intervalle de confiance asymptotique au niveau de confiance 95% pour la moyenne du niveau maximal annuel de la rivière en cet endroit. Que fait-il ?
5. Á l'intention de l'hydrologue, le statisticien illustre la convergence de l'intervalle de confiance asymptotique avec des simulations. Pour cela, il génère 10000 échantillons i.i.d. de taille $n = 30; 50; 100; 200$ de loi non gaussienne. Il détermine à partir des 10000 intervalles de confiance obtenus la probabilité de couverture et la largeur moyenne de l'intervalle de confiance en fixant le niveau de confiance à 95%. Il obtient les résultats suivants :

n	probabilité de couverture	largeur moyenne
30	0,933	12,97
50	0,937	10,08
100	0,944	7,18
200	0,949	5,08

Quel commentaire fait-il à l'hydrologue ?
6. L'hydrologue souhaite déterminer la hauteur de la digue ayant une probabilité de 1/10000 d'être dépassée au cours d'un an. Le statisticien explique alors qu'il est nécessaire d'ajuster

une loi paramétrique afin de pouvoir estimer la hauteur de la digue. Le statisticien propose d'ajuster une loi Gamma notée $\Gamma(a, b)$ avec $a, b > 0$ dont la fonction de densité est donnée par :

$$f_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) I(x > 0)$$

et effectue un test validant cet ajustement. Il estime alors les paramètres de la loi Gamma et détermine leur comportement asymptotique. Que fait-il ?

7. Le statisticien en déduit alors une estimation de la hauteur de la digue satisfaisant le critère imposé par l'hydrologue. Que fait-il ?

Exercice 5.

1. Un hydrologue s'intéresse au débit mensuel d'un fleuve donné en un point donné, débit représenté par une variable aléatoire Y observée sur une période de 10 ans ce qui fournit une réalisation d'un échantillon de variables aléatoires $(Y_1, ..., Y_{120})$. Ecrire un modèle de régression linéaire pour expliquer la variable Y par
   (a) la pluviométrie cumulée mensuelle, également observée mensuellement sur une période de 10 ans,
   (b) le débit observé le mois précédent.
2. Un autre hydrologue s'intéresse au débit saisonnier du fleuve (en coupant l'année en quatre saisons : printemps, été, automne, hiver). Il dispose des observations trimestrielles sur une période de 10 ans de la variable Y correspondant à l'échantillon $(Y_1, ..., Y_{40})$. Comment interpréter le modèle suivant :

$$Y_i = \beta_0 + \beta_1 \delta_{1,i} + \beta_2 \delta_{2,i} + \beta_3 \delta_{3,i} + \beta_4 \delta_{4,i} + \varepsilon_i \,, \quad i = 1, ..., 40,$$

   où les ε_i pour $i = 1, ..., n$ sont des variables indépendantes telles que $\mathbb{E}[\varepsilon_i] = 0$ et $\mathrm{Var}(\varepsilon_i) = \sigma^2$ pour $i = 1, ..., n$, avec $\delta_{k,i} = \begin{cases} 1 & \text{si } i \equiv k(4) \\ 0 & \text{sinon} \end{cases}$ pour $k = 1, ..., 4$ et où l'on impose $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$.

Exercice 6.

Soient les couples de variables aléatoires (X_i, Y_i) pour $i = 1, .., n$, indépendants et identiquement distribués. On suppose que la loi conditionnelle de Y_i sachant $X_i = x_i$ est la loi normale de moyenne $\beta_0 + \beta_1 x_i$ et de variance σ^2 pour $i = 1, ..., n$. Déterminer les estimateurs des paramètres β_0, β_1 et σ^2 obtenus par la méthode du maximum de vraisemblance. Sont-ils sans biais ? Sont-ils égaux aux estimateurs obtenus par la méthode des moindres carrés ordinaires ?