X2016 – MAP 311

PC 9 – 26 juin 2017 – Tests

Corrigé des questions non abordées en PC

Igor Kortchemski – igor.kortchemski@cmap.polytechnique.fr

Corrigé des exercices non traités sur http://www.normalesup.org/~kortchem/MAP311 *un peu après la PC.*

2 Estimateurs du maximum de vraisemblance

Exercice 3. La hauteur maximale en mètres de la crue annuelle d'un fleuve est modélisée par une variable aléatoire de Rayleigh de paramètre a qui a pour densité :

$$f(x) = \frac{x}{a} e^{-\frac{x^2}{2a}} \mathbb{1}_{x \geq 0}$$

où a est un paramètre inconnu.

(1) Déterminer l'esimateur du maximum de vraisemblance $\widehat{a}_n$ de a.

(2) Si X suit une loi de Rayleigh de paramètre a, démontrer que X^2 suit une loi exponentielle de paramètre $1/(2a)$.

(3) L'estimateur $\widehat{a}_n$ est-il sans biais ?

(4) Trouver un intervalle de confiance asymptotique pour a au niveau 95%.
On rappelle que la variance d'une loi exponentielle de paramètre λ est $1/\lambda^2$.

(5) Une compagnie d'assurance estime qu'une catastrophe correspond à une crue de plus de 6m. Trouver un intervalle de confiance asymptotique pour la probabilité p qu'une catastrophe se produise durant une année au niveau 95%.

Corrigé : Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes de loi de Rayleigh.

(1) Une densité de $(X_1, \ldots, X_n)$ au point $(x_1, \ldots, x_n)$ est

$$\frac{\prod_{i=1}^n x_i}{a^n} e^{-\frac{\sum_{i=1}^n x_i^2}{2a}}.$$

Pour chercher quelle valeur de a maximise cette quantité, on calcule sa dérivée par rapport à a, qui vaut

$$-n \frac{\prod_{i=1}^n x_i}{a^{n+1}} e^{-\frac{\sum_{i=1}^n x_i^2}{2a}} + \frac{\prod_{i=1}^n x_i}{a^n} \cdot \frac{\sum_{i=1}^n x_i^2}{2a^2} \cdot e^{-\frac{\sum_{i=1}^n x_i^2}{2a}} = \frac{\prod_{i=1}^n x_i}{a^{n+2}} e^{-\frac{\sum_{i=1}^n x_i^2}{2a}} \left(-an + \frac{\sum_{i=1}^n x_i^2}{2} \right).$$

Ainsi,

$$\widehat{a}_n = \frac{\sum_{i=1}^n X_i^2}{2n}.$$

(2) On utilise la méthode de la fonction muette. Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Alors, avec le changement de variable $u = x^2$ (et donc $x = \sqrt{u}$ et $dx = \frac{du}{2\sqrt{u}}$) :

$$\mathbb{E}[F(X)] = \int_0^\infty F(x^2) \frac{x}{a} e^{-\frac{x^2}{2a}} dx = \int_0^\infty F(u) \frac{1}{2a} e^{-\frac{u}{2a}} du,$$

d'où le résultat.

(3) Comme X_1^2 suit une loi exponentielle de paramètre $1/(2a)$, on a $\mathbb{E}[X_1^2] = 2a$, et par linéarité de l'espérance $\widehat{a}_n$ est sans biais.

Pour des questions, demande d'explications etc., n'hésitez pas à m'envoyer un mail.

(4) D'après le théorème central limite,

$$\sqrt{n} \cdot \left(\frac{\sum_{i=1}^n X_i^2}{n} - 2a \right) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \mathrm{Var}(X_1^2)) = \mathcal{N}(0, 4a^2)).$$

On en déduit en utilisant le lemme de Slutsky que

$$\frac{\sqrt{n}}{\widehat{a}_n} (\widehat{a}_n - a) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1).$$

En notant $z_{1-\alpha/2} \simeq 1.96$ le quantile de niveau $1-\alpha/2$ de la loi gaussienne centrée réduite, on en déduit que

$$\widehat{I}_n = \left[\widehat{a}_n \left(1 - \frac{z_{1-\alpha/2}}{\sqrt{n}} \right), \widehat{a}_n \left(1 + \frac{z_{1-\alpha/2}}{\sqrt{n}} \right) \right]$$

eſt un intervalle de confiance asymptotique pour a au niveau 95%.

(5) La probabilité d'une cataſtrophe eſt

$$p = \int_6^\infty \frac{x}{a} e^{-\frac{x^2}{2a}} dx = \left[e^{-\frac{x^2}{2a}} \right]_6^\infty = e^{-18/a}.$$

Ainsi, un interalle de confiance asymptotique pour p au niveau 95% eſt :

$$\widehat{J}_n = \left[\exp\left(-\frac{18}{\widehat{a}_n \left(1 + \frac{z_{1-\alpha/2}}{\sqrt{n}} \right)} \right), \exp\left(-\frac{18}{\widehat{a}_n \left(1 - \frac{z_{1-\alpha/2}}{\sqrt{n}} \right)} \right) \right].$$

□

3 Plus appliqué

Exercice 4. On considère les décimales du nombre π eſt on souhaite savoir si elles sont uniformément réparties sur $\mathcal{X} = \{0, \ldots, n\}$. On calcule π à 10^{-n} près avec $n = 10000$, ce qui nous donne les n premières décimales de π. On trouve $N_0 = 968$, $N_1 = 1026$, $N_2 = 1021$, $N_3 = 975$, $N_4 = 1012$, $N_5 = 1046$, $N_6 = 1021$, $N_7 = 969$, $N_8 = 948$ et $N_9 = 1014$.

Teſter l'hypothèse que les décimales de π sont uniformément réparties au niveau 5%.

Corrigé : On effectue un teſt du χ^2 d'adéquation à une loi uniforme sur $\{0, 1, \ldots, 9\}$. En notant $\widehat{\theta}_i$ la proportion du nombre de i et en posant $\theta_i = 1/10$, on forme

$$\zeta_n = n \sum_{j=0}^{9} \frac{\left(\widehat{\theta}_j - \theta_j\right)^2}{\theta_j} \simeq 9.3.$$

Sous l'hypothèse H_0 "les décimales sont uniformément réparties", ζ_n converge en loi cers une loi du χ_2 à 9 degrés de liberté. Si $Z \sim \chi_9^2$, on a $\mathbb{P}(Z \geq 16.9) = 0.05$. Comme $\zeta_n \geq 16.9$, on accepte au niveau (asymptotique) 5%. □

4 Quelques exercices pour finir...

Exercice 5. Démontrer qu'il n'eſt pas possible de truquer deux dés indépendants de façon à ce que la somme des points obtenus en les lançant indépendamment suive la loi uniforme sur l'ensemble $\{2, 3, 4, \ldots, 12\}$.

Indication. On pourra noter U et V deux variables aléatoires indépendantes à valeurs dans $\{1,2,\ldots,6\}$, supposer que $S = U + V$ suit la loi uniforme sur $\{2,3,\ldots,12\}$ et exploiter le fait que $\mathbb{E}\left[x^S\right] = \mathbb{E}\left[x^U\right] \cdot \mathbb{E}\left[x^V\right]$.

Corrigé : Notons $u_i = \mathbb{P}(U = i)$ et $v_i = \mathbb{P}(V = i)$ pour $1 \leq i \leq 6$. Tout d'abord, on doit avoir $u_6 > 0$ et $v_6 > 0$ (car sinon $\frac{1}{11} = \mathbb{P}(S = 12) = \mathbb{P}(U = 6 \text{ et } V = 6) = 0$. D'autre part,

$$\frac{1}{11}(x^2 + \cdots + x^{12}) = (u_1 x + \cdots + u_6 x^6)(v_1 x + \cdots + v_6 x^6).$$

Donc

$$\frac{1}{11} \cdots \frac{x^{12} - 1}{x - 1} = (u_1 + \cdots + u_6 x^5)(v_1 + \cdots + v_6 x^5).$$

Comme un polynôme à coefficients réels de degré impair admet au moins une racine réelle, on en déduit que $\frac{x^{12}-1}{x-1}$ possède au moins deux racines réelles, ce qui est absurde. $\square$

Exercice 6. Selon Shakespeare, César dit à Brutus au moment de mourir : “Et tu, Brute ? Then fall, Cæsar !”. Estimez la probabilité que vous inhaliez à cet instant l'une des molécules d'air exhalées par César prononçant cette ultime phrase.

On supposera qu'il y a 10^{44} molécules d'air en tout, qu'il y a environ $6 \cdot 10^{23}$ molécules d'air dans 22.5 litres d'air et qu'une inhalation ou une exhalation représentent 0.5 litres d'air.

Corrigé : On note $N = 10^{44}$ le nombre de molécules d'air (supposé constant). Soit a le nombre de molécules d'air inhalées/exhalées pendant une respiration. Notons p la probabilité qu'on inhale à cet instant l'une des molécules d'air exhalées par César prononçant cette ultime phrase.

La probabilité qu'une moléculare d'air donnée ait été exhalée par César vaut a/N. Donc

$$1 - p = \left(1 - \frac{a}{N}\right)^a$$

(en supposant que les molécules d'air sont uniformément réparties dans l'atmosphère et en utilisant l'approximation tirage sans remise = tirage avec remise car a est beaucoup plus petit que N).

On a $a = 0.5 \cdot \frac{6 \cdot 10^{23}}{22.5} = \frac{4}{3} \cdot 10^{22}$, de sorte

$$1 - p = \exp\left(\frac{4}{3} \cdot 10^{22} \cdot \ln\left(1 - \frac{4}{3} \cdot 10^{-22}\right)\right) \simeq \exp\left(-\frac{16}{9}\right) \simeq 0.17,$$

en utilisant l'approximation $\ln(1 - x) \simeq -x$ pour x petit. Ainsi, $p \simeq 0.83$. $\square$