Aléatoire PC 8 :

Benjamin Jourdain

Exercice 1. On observe un échantillon $(X_1, \dots, X_n)$ de n variables aléatoires i.i.d. suivant la loi exponentielle de paramètre $\theta > 0$ de densité $\theta e^{-\theta x} 1_{\{x \geq 0\}}$.

1. Pour $(x_1, \dots, x_n) \in]0, +\infty[^n$, déterminer la valeur $\theta_n(x_1, \dots, x_n)$ qui maximise la densité de $(X_1, \dots, X_n)$ (ou ce qui est équivalent le ln de la densité) au point $(x_1, \dots, x_n)$ comme fonction de $\theta > 0$. En déduire l'estimateur du maximum de vraisemblance $\hat{\theta}_n = \theta_n(X_1, \dots, X_n)$ de θ et montrer qu'il converge p.s. vers θ lorsque $n \to \infty$.
2. Quelle est la loi de $X_1 + \dots + X_n$? En déduire $\mathbb{E}(\hat{\theta}_n)$. L'EMV est-il sans biais ?
3. À l'aide de la méthode delta, montrer que $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une gaussienne centrée de variance à préciser.
4. En déduire que pour $\alpha \in]0, 1[$, $[\hat{\theta}_n(1 - \frac{\phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}), \hat{\theta}_n(1 + \frac{\phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}})]$, où ϕ désigne la fonction de répartition de la loi $\mathcal{N}_1(0, 1)$, est un intervalle de confiance de niveau asymptotique α pour θ.

   A.N. : données de batteries de téléphone portable $\sum_{i=1}^{100} x_i = 3128h$, $\alpha = 5\%$.

Exercice 2.
On désire comparer les moyennes de deux échantillons indépendants $(X_{1,1}, \dots, X_{1,n_1})$ et $(X_{2,1}, \dots, X_{2,n_2})$ distribués respectivement suivant $\mathcal{N}_1(\mu_1, \sigma^2)$ et $\mathcal{N}_1(\mu_2, \sigma^2)$ où $\mu_1, \mu_2 \in \mathbb{R}$ et $\sigma^2 > 0$. Pour $i \in \{1, 2\}$, on pose $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$ et $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2$.

1. Quelle est la loi de $\left(\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma}, \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2}\right)$?

   En déduire que $\sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}}$ suit la loi de Student de paramètre $n_1 + n_2 - 2$.
2. On s'interroge sur les tailles des garçons et des filles de douze ans dans la population. Dans une classe de cinquième, on a observé
   - 16 garçons : moyenne $\bar{x}_G = 148$ cm, écart-type $s_G = 6.9$ cm,
   - 15 filles : moyenne $\bar{x}_F = 153$ cm, écart-type $s_F = 5.8$ cm.

   À l'aide de la question précédente, tester au niveau 5% l'hypothèse nulle “la taille moyenne des filles dépasse celle des garçons de moins de 2 cm”.
3. On suppose maintenant que l'écart-type σ est connu et vaut 6 cm. Reprendre le test précédent en travaillant avec une statistique de test qui tient compte de cette information.

Exercice 3. : étalonnage
On considère que la réponse d'un appareil de mesure à un signal déterministe ξ est égale à $a\xi$ plus un bruit gaussien centré de variance b où $(a, b) \in \mathbb{R} \times \mathbb{R}_+^*$. On se propose d'étalonner l'appareil (c'est à dire estimer les valeurs de a et b) en envoyant une suite $x = (x_1, x_2, \dots, x_n)$ de signaux connus. On note $Y_i = ax_i + \sqrt{b}U_i$ la réponse au i-ième signal où on suppose que les coordonnées du vecteur $U = (U_1, U_2, \dots, U_n)$ sont des variables aléatoires gaussiennes centrées réduites indépendantes. On note

$$\widehat{A} = \frac{\sum_1^n x_i Y_i}{\sum_1^n x_i^2} \text{ et } \widehat{B} = \frac{\sum_1^n (Y_i - x_i \widehat{A})^2}{n-1}$$

1. Donner la loi de $\widehat{A}$. À quelle condition sur la suite $(x_i)_{i\geq 1}$ a-t-on $\mathbb{E}\left((\widehat{A}-a)^2\right) \to_{n\to+\infty} 0$?
2. Vérifier que si on complète $e_1 = \frac{x}{\|x\|}$ en une base orthonormée $(e_1, \dots, e_n)$ de $\mathbb{R}^n$, on a $\widehat{B} = \frac{b}{n-1}\sum_{i=2}^n (e_i.U)^2$. En déduire que $\widehat{A}$ et $\widehat{B}$ sont des variables aléatoires indépendantes. Donner l'espérance et la variance de $\widehat{B}$.
3. Montrer qu'il existe une constante c (dépendant de x) telle que la variable aléatoire $c\dfrac{\widehat{A}-a}{\sqrt{\widehat{B}}}$ suive une loi de Student.
4. Donner un intervalle de confiance à 95% pour le paramètre a, suivant que l'on connaît la valeur de b ou non.

Exercice 4.
On observe un échantillon $X = (X_1, \dots, X_n)$ de variables aléatoires I.I.D. à valeurs dans l'espace fini $\{a_1, \dots, a_k\}$ telles que $\mathbb{P}(X_1 = a_j) = p_j$ pour $j \in \{1, \dots, k\}$. Le paramètre p appartient à $\Theta = \{(p_1, \dots, p_k) \in \mathbb{R}_+^k : p_1 + \ldots + p_k = 1\}$. Pour $p^0 = (p_1^0, \dots, p_k^0) \in \Theta \cap]0,1[^k$, on souhaite tester l'hypothèse nulle $H_0 = \{p = p^0\}$ contre l'hypothèse alternative $H_1 = \{p \neq p^0\}$.

1. Vérifier que pour $x = (x_1, \dots, x_n) \in \{a_1, \dots, a_k\}^n$, $\ln \mathbb{P}(X = x)$ est égal à $l_n(x,p) = n\sum_{j:\rho_j>0} \rho_j \ln p_j$ où $\rho_j = \frac{1}{n}\sum_{i=1}^n 1_{\{x_i = a_j\}}$.
2. En remarquant que pour tout $y > 0$, $\ln(y) \leq y-1$, vérifier que $\frac{1}{n}(l_n(x,p) - l_n(x,\rho)) \leq 0$ et en déduire que l'EMV de p est $\hat{p} = (\hat{p}_1, \dots, \hat{p}_k)$ où $\hat{p}_j = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i = a_j\}}$.

Soit $\zeta_n = n\sum_{j=1}^k \frac{(\hat{p}_j - p_j^0)^2}{p_j^0}$.

3. Montrer que sous H_1, ζ_n tend p.s. vers $+\infty$ avec n.
4. Pour $k = 2$, vérifier que $\zeta_n = Y_n^2$ où $Y_n = \sqrt{\frac{n}{p_1^0(1-p_1^0)}}\left(\frac{1}{n}\sum_{i=1}^n 1_{\{X_i = a_1\}} - p_1^0\right)$ et en déduire que sous H_0, ζ_n converge en loi vers $\zeta \sim \chi^2(1)$ lorsque $n \to \infty$.

On admet plus généralement que pour $k \geq 2$, sous H_0, ζ_n converge en loi vers $\zeta \sim \chi^2(k-1)$.

5. En déduire que le test de région critique $W_n = \{\zeta_n > \chi^2_{k-1,1-\alpha}\}$ où $\chi^2_{k-1,r}$ désigne le quantile d'ordre $r \in]0,1[$ de la loi $\chi^2(k-1)$ est convergent de niveau asymptotique α.

Exercice 5. Lors de 100 lancers d'un dé à 6 faces on observe les résultats suivants

x	1	2	3	4	5	6
$N(x)$	20	13	17	12	23	15

Tester au niveau 5% si le dé n'est pas pipé.

Exercice 6. On souhaite vérifier la qualité du générateur de nombres aléatoires d'une calculatrice scientifique. Pour cela, on procède à 250 tirages dans l'ensemble $\{0, \dots, 9\}$ et on obtient les résultats suivants :

x	0	1	2	3	4	5	6	7	8	9
$N(x)$	28	32	23	26	23	31	18	19	19	31

Tester au niveau $\alpha = 0.1$ si le générateur produit des entiers uniformément répartis sur $\{0, \dots, 9\}$.