

Aléatoire PC 8 :

Groupes 5 & 17 – Aurélien Alfonsi

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 $(X_1, \dots, X_n)$ 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 $n \rightarrow \infty$.

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. Remarquer que $\sqrt{n}(\hat{\theta}_n - \theta) = -\frac{\theta}{\bar{X}_n} \times \sqrt{n}\left(\bar{X}_n - \frac{1}{\theta}\right)$.

Vérifier que $\left(-\frac{\theta}{\bar{X}_n}, \sqrt{n}\left(\bar{X}_n - \frac{1}{\theta}\right)\right)$ converge en loi vers une limite à préciser et conclure que l'EMV est asymptotiquement normal de variance asymptotique à préciser. Retrouver ce résultat en appliquant la méthode delta.

4. Construire un intervalle de confiance à 95% pour θ en fonction de $\hat{\theta}_n$.

Exercice 2.

Lors de 100 lancers du 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 3.

Soient $X_1, \dots, X_n$ des variables aléatoires I.I.D. suivant la loi $\mathcal{N}_1(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma > 0$. On pose $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ et $S_n = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2}$.

1. Quelle est la loi de $\bar{X}_n$? Quelle est celle de $Y = {}^t(\frac{X_1 - \mu}{\sigma}, \dots, \frac{X_n - \mu}{\sigma})$?
2. Soit $e_1, \dots, e_n$ une base orthonormée de $\mathbb{R}^n$ telle que $e_1 = {}^t(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$ et $O = {}^t(e_1, \dots, e_n)$. Quelle est la loi de $Z = OY$?
3. Vérifier que $\frac{n-1}{\sigma^2} S_n^2 = \sum_{j=2}^n Z_j^2$. En déduire que cette variable est indépendante de $\bar{X}_n$ et suit la loi du χ^2 à $n-1$ degrés de liberté. Quelle est la loi de $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$?

Exercice 4.

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 d'au plus 4 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 5.

On observe un échantillon $(X_1, \dots, X_n)$ de variables aléatoires I.I.D. à valeurs dans un espace fini $A = \{a_1, \dots, a_k\}$. La loi commune est paramétrisée par $p = (p_1, \dots, p_k) : \mathbb{P}_p(X_1 = a_j) = p_j$ pour $j \in \{1, \dots, k\}$. Le paramètre p appartient à $\Theta = \{(p_1, \dots, p_k) \in \mathbb{R}_+^k : p_1 + \dots + p_k = 1\}$. 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 la log-vraisemblance est donnée par $l_n(x, p) = n \sum_{j: \rho_j > 0} \rho_j \ln p_j$ 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 $\frac{1}{n} (l_n(x, p) - l_n(x, p^0)) \leq 0$ 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 $n \rightarrow \infty$.

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 $W_n = \{\zeta_n > \chi_{k-1, 1-\alpha}^2\}$ où $\chi_{k-1, r}^2$ désigne le quantile d'ordre $r \in]0, 1[$ de la loi $\chi_2(k-1)$ est convergent de niveau asymptotique α et donner sa p -valeur asymptotique.