X2015 – MAP 311
PC 9 – 27 juin 2016 – Tests, estimateurs du maximum de vraisemblance et vampires

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

Exercice 1. (Retour sur le théorème de Fubini, Théorème 4.8.1 du poly)

(1) Soit X une variable aléatoire à valeurs dans $\mathbb{N} = \{0, 1, 2, \ldots\}$. Montrer que

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Indication. On pourra écrire que $\mathbb{P}(X \geq i) = \mathbb{E}[\mathbb{1}_{X \geq i}]$ et intervertir somme et espérance.

(2) Soit X une variable aléatoire à valeurs dans $\mathbb{R}_+$. Montrer que

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X \geq u)\, du.$$

Indication. On pourra écrire $\mathbb{P}(X \geq u) = \mathbb{E}[\mathbb{1}_{X \geq u}]$ et intervertir intégrale et espérance.

(3) *Application.* Soient $(X_i)_{1 \leq i \leq n}$ des varables aléatoires réelles indépendantes de même loi exponentielle de paramètre λ. Calculer $\mathbb{E}[\max(X_1, \ldots, X_n)]$.

1 Tests

Exercice 2. Soit $X = (X_1, \ldots, X_n)$ un échantillon de taille n de variables aléatoires indépendantes de même loi exponentielle de paramètre $1/\theta > 0$ avec θ inconnu. Soit $\theta_0 > 0$ et $\alpha \in]0, 1[$. On pose $S_n = X_1 + \cdots + X_n$.

On souhaite tester $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$ au niveau α.

(1) Rappeler pourquoi S_n suit une loi $\Gamma(n, \frac{1}{\theta})$.

(2) Démontrer que sous H_0,

$$Z_n = \frac{2S_n}{\theta_0}$$

suit une loi du χ^2 à un nombre de de degrés de liberté d_n qu'on précisera.

On rappelle qu'une loi du χ^2 à d degrés de liberté est la loi de la somme des carrés de d gaussiennes centrées réduites indépendantes, et qu'une densité de cette loi est

$$\frac{1}{2^{\frac{d}{2}}\Gamma(d/2)} t^{\frac{d}{2}-1} e^{-\frac{t}{2}} \mathbb{1}_{t \geq 0}.$$

On rappelle que la densité de la loi $\Gamma(a,\lambda)$ est $\frac{1}{\Gamma(a)}\lambda^a x^{a-1} e^{-\lambda x} \mathbb{1}_{x>0}$.

(3) En déduire une région de rejet de la forme

$$W_n = \{S_n \geq c\}$$

avec c une constante qu'on exprimera en fonction de θ_0 et $z_{1-\alpha}$, le quantile d'ordre $1-\alpha$ de la loi du χ^2 à d_n degrés de liberté.

Remarque : Il est possible de démontrer que cette région de rejet fournit le test uniformément le plus puissant parmi les tests de niveau α de $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$ (c'est-à-dire que n'importe quelle autre région de rejet au niveau α va donner un test avec une erreur de seconde espèce plus grande que le test qu'on vient de construire).

(4) On suppose que θ est le temps moyen d'attente du RER B à la gare de Lozère. Une association d'usagers et la RATP souhaitent tester si le RER B respecte un temps d'attente moyen d'au plus $\theta_0 = 15\,\text{min}$. Teste-t-on

$$H_0 = \{\theta \leq \theta_0\} \text{ contre } H_1 = \{\theta > \theta_0\} \qquad \text{ou} \qquad H_0 = \{\theta \geq \theta_0\} \text{ contre } H_1 = \{\theta < \theta_0\}?$$

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%.

3 Plus appliqué

Exercice 4. Une colonie de vampires a élu domicile dans un château des Carpates, et le comte Drakul souhaite estimer leur nombre. Pour cela, une nuit de pleine lune, le comte en capture 10, leur mord les oreilles, puis les relâche. La nuit suivante, il en capture 10 au hasard. Trois ont une morsure aux oreilles.

Aidez le comte Drakul en utilisant un estimateur du maximum de vraisemblance.

4 Pour aller plus loin

Exercice 5. (Prolongement de l'exercice 1) Soit X une variable aléatoire réelle et $f : \mathbb{R} \to \mathbb{R}_+$ une fonction croissante C^1 telle que $\lim_{-\infty} f = 0$. Montrer que

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f'(u)\mathbb{P}(X \geq u)\,du.$$

Le résultat reste-t-il vrai si $\lim_{-\infty} f \neq 0$?

Exercice 6. 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 $\mathbb{E}\left[x^S\right] = \mathbb{E}\left[x^U\right] \cdot \mathbb{E}\left[x^V\right]$.

Exercice 7. Le but de cet exercice est d'étudier le nombre de triangles dans un graphe aléatoire d'Erdös-Rényi en utilisant les méthodes du premier et du second moment.

Soit $n \geq 3$ un entier et $p_n \in [0, 1]$. Par définition, une arête est un ensemble de la forme $\{i, j\}$ avec $i, j \in \{1, 2, \ldots, n\}$ et $i \neq j$; un triangle est un ensemble de la forme $\{i, j, k\}$ avec $i, j, k \in \{1, \ldots, n\}$ et où i, j, k sont deux à deux différents.

(1) Combien y a-t-il d'arêtes dans $\{1, \ldots, n\}$? Combien y a-t-il de triangles ?

On note A l'ensemble des arêtes et T l'ensemble des triangles. On considère des variables aléatoires $(X_e)_{e \in A}$ indépendantes de même loi de Bernoulli de paramètre p_n. Si $X_e = 1$, on colorie l'arête e, et si $X_e = 0$, on ne colorie pas l'arête e. On dit qu'un triangle $t = \{i, j, k\}$ est colorié si ses trois arêtes $\{i, j\}$, $\{j, k\}$ et $\{i, k\}$ sont coloriées. On note N_n le nombre de triangles coloriés.

Pour tout triangle t, on considère la variable aléatoire I_t qui vaut 1 si t est colorié et 0 sinon.

(2) Calculer $\mathbb{E}[I_t]$ et $\mathrm{Var}(I_t)$ pour tout triangle t.

(3) En remarquant que $N_n = \sum_{t \in T} I_t$, calculer $\mathbb{E}[N_n]$.

(4) On suppose que $n \cdot p_n \to 0$ lorsque $n \to \infty$. Montrer que $\mathbb{P}(N_n = 0) \to 1$.

(5) Soit X une variable aléatoire positive dont le carré est intégrable. Montrer que $\mathbb{P}(X=0) \le \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2}$.

(6) Quel nombre de manières de choisir un couple de deux triangles (t_1, t_2) tels que t_1 et t_2 ont exactement une arête en commun ?

(7) Montrer que, si t et t' sont deux triangles qui partagent une seule arête commune, alors $\mathrm{Cov}(I_t, I_{t'}) \le p_n^5$.

(8) On suppose que $n \cdot p_n \to \infty$ lorsque $n \to \infty$. Montrer que $\mathbb{P}(N_n = 0) \to 0$.