Exercices de Statistiques
Université de Lorraine

Estimation et théorie des tests

Clément Dell'Aiera

1 Principe de Neyman : décision à 2 points

1. Soit f la densité d'une loi de probabilité sur $\mathbb{R}$, et $\mathcal{E}$ l'expérience statistique engendré par un n-échantillon de loi $p_\theta(x) = f(x-\theta)$. On suppose que $\Theta = \{0, \theta_0\}$ avec $\theta_0 \neq 0$. On veut tester $H_0|\theta = 0$ contre $H_1|\theta = \theta_0$.
   (a) Décrire l'expérience statistique et donner la vraisemblance du modèle.
   (b) Donner la zone de rejet du test de Neyman-Pearson de niveau α associé à H_0 et H_1.
2. L'expérimentateur observe une seule réalisation d'une v.a. X de loi de Poisson de paramètre $\theta > 0$. On veut tester $H_0|\theta = \theta_0$ contre $H_1|\theta = \theta_1$, où $\theta_0 \neq \theta_1$.
   (a) Donner la zone de rejet du test de Neyman-Pearson de niveau α associé.
   (b) Sachant que $\mathbb{P}_{\theta_0}(X > 9) = 0.032$ et $\mathbb{P}_{\theta_1}(X > 8) = 0.068$, donner une zone de rejet explicite pour $\alpha = 0.05 = 5\%$. Le test est-il optimal ?

2 Neyman-Pearson : familles à rapport de vraisemblance monotone

1. Soit $\mathcal{E}$ l'expérience statistique engendrée par un n-échantillon de loi normale $\mathcal{N}(\theta, \sigma^2)$, où σ^2 est connu, et $\theta \in \Theta = \mathbb{R}$. On souhaite tester $H_0|\theta = \theta_0$ contre $H_1|\theta = \theta_1$, où $\theta_0 < \theta_1$.
   (a) Décrire le modèle ainsi que la vraisemblance. On choisira la mesure de Lebesgue comme mesure dominante.
   (b) Calculer le rapport de vraisemblance
   $$\frac{f(\theta_1, Z)}{f(\theta_0, Z)}.$$
   (c) Donner la zone de rejet pour le test de Neyman-Pearson associé.
2. Pour la même expérience statistique, on a un test optimal (uniformément plus puissant) de H_0 contre H_1 donné par la région de rejet
   $$\mathcal{R} = \{\overline{X}_n > c\}$$
   où c est solution de $\mathbb{P}_{\theta_0}(\overline{X}_n > c) = \alpha$.
   (a) Calculer explicitement la valeur de la constante $c = c(\theta_0, \alpha)$.
   (b) Calculer la puissance de ce test.

3 Exercice

L'expérimentateur observe 2 échantillons indépendants $X_1, ..., X_n$ et $Y_1, ..., Y_m$ de tailles distinctes $n \neq m$, de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$. Il souhaite tester

$$H_0 \ : \ \mu_1 = \mu_2 \quad \text{contre} \quad H_1 : \ \mu_1 \neq \mu_2.$$

Si $s_{n,1}^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^2$ et $s_{m,2}^2 = \frac{1}{m}\sum_{j=1}^m (Y_j - \overline{Y}_m)^2$, construire un test basé sur la statistique

$$T_{n,m} = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{s_{n,1}^2 + s_{m,2}^2}}$$

et étudier sa consistance.

4 Tests du χ^2

On considère une variable qualitative X, à valeur dans un ensemble fini $E = \{1, .., d\}$. Les lois de probabilité de telles v.a. sont entièrement décrites par le vecteur de probabilité $(p_1, ..., p_d)^T$, où $p_j = \mathbb{P}(X = j)$. On confondera donc les lois de probabilités de E avec

$$\mathfrak{M}_d = \{p = (p_1, ..., p_d)^T : 0 \geq p_j \geq 1 \text{ et } \sum p_j = 1\}.$$

1. **Test d'adéquation du χ^2.** On observe un n-échantillon de loi p et l'on souhaite tester $p = q$ contre $p \neq q$ où $q \in \mathfrak{M}_d$ est une loi fixée.
   (a) Décrire le modèle statistique.
   (b) On définit les fréquences empiriques
   $$\hat{p}_{n,l} = \frac{1}{n} \sum_{j=1}^{n} 1_{X_j = l} \quad \text{pour } l = 1, ..., d.$$
   Donner la limite du vecteur $\hat{p}_n = (\hat{p}_{n,l})^T_{l=1,d}$ pour la topologie de la convergence en probabilité sous $\mathbb{P}_p$.
   (c) On définit
   $$U_n(p) = \sqrt{n}(\frac{\hat{p}_{n,l} - p_l}{\sqrt{p_l}})^T_{l=1,d}.$$
   Donner la limite en loi de chaque composante de $U_n(p)$ sous $\mathbb{P}_p$. Que peut-on dire a priori de la limite en loi de $U_n(p)$? Pourquoi ?
   (d) On définit
   $$Y_l^j = \frac{1}{\sqrt{p_l}}(1_{X_j = l} - p_l).$$
   Si Y_j désigne le vecteur $(Y_1^j, ..., Y_d^j)$, montrer que $\frac{1}{\sqrt{n}} \sum Y_j = U_n(p)$.
   (e) Calculer $E[Y_l^j]$, et $E[Y_l^j Y_{l'}^j]$. Que valent les composantes de la matrice $V(p) = I_d - \sqrt{p}\sqrt{p}^T$, où $\sqrt{p} = (\sqrt{p_1}, ..., \sqrt{p_d})^T$?
   (f) En déduire la limite en loi sous $\mathbb{P}_p$ de $U_n(p)$ et de $||U_n(p)||^2$, le carré de sa norme euclidienne.
   (g) Soient $p, q \in \mathfrak{M}_d$ tels que les coefficients de q soient tous non nuls. On définit :
   $$\chi^2(p, q) = \sum_{l=1}^{d} \frac{(p_l - q_l)^2}{q_l}.$$
   Cette quantité est appelée "distance du χ^2" bien que ce ne soit pas une distance ! Toutefois, $\chi^2(p, q) = 0$ ssi $p = q$.
   Montrer que $n\chi^2(\hat{p}_n, p) = ||U_n(p)||^2$.
   (h) On définit, pour $\alpha \in (0, 1)$, la zone de rejet
   $$\mathcal{R}_{n,\alpha} = \{n\chi^2(\hat{p}_n, p) \geq q^{\chi^2}_{1-\alpha,d-1}\},$$
   où $q^{\chi^2}_{1-\alpha,d-1}$ est le quantile d'ordre $1 - \alpha$ de la loi du χ^2 à $d - 1$ degrés de liberté.
   Montrer que le test associé est asymptotiquement de niveau α et est asymptotiquement consistant.