DEVOIR MAISON

2 NOVEMBRE 2020

Exercice 1 (Inégalité de Bennett)

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes possédant un moment d'ordre 2, et $Z = \sum_{i=1}^n X_i$. On suppose que pour tout $i \in [\![1, n]\!]$, $X_i \leq 1$.

1. Soit $\phi : x \mapsto e^x - x - 1$. Montrer que la fonction $x \mapsto \frac{\phi(x)}{x^2}$, prolongée par continuité en 0, est croissante sur $\mathbb{R}$.
2. En déduire que pour tout $i \in [\![1, n]\!]$, et pour tout $\lambda \geq 0$, on a
$$\mathbb{E}\left[e^{\lambda X_i}\right] \leq 1 + \lambda \mathbb{E}[X_i] + \mathbb{E}[X_i^2]\phi(\lambda)\,,$$
puis que
$$\ln \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}[Z])}\right] \leq v\phi(\lambda)\,,$$
où $v = \sum_{i=1}^n \mathbb{E}[X_i^2]$.
3. Montrer que pour tout $t \geq 0$, on a
$$\mathbb{P}\left(Z - \mathbb{E}Z \geq t\right) \leq \exp\left(-vh\left(\frac{t}{v}\right)\right)\,,$$
où pour tout $x \geq 0$, $h(x) = (1+x)\ln(1+x) - x$.

Exercice 2 (Loi de Pareto)

Soit $\theta > 0$, et X une variable aléatoire de densité f_θ définie pour tout $x \in \mathbb{R}$ par
$$f_\theta(x) = \frac{2\theta^2}{x^3}\mathbb{1}_{x \geq \theta}.$$

Dans ce qui suit on considère $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X.

1. Calculer $\mathbb{E}[X]$. En déduire un estimateur de θ. Est-il consistant ? Le théorème central limite peut-il s'appliquer ?
2. Calculer $\mathbb{E}\left[X^{-1}\right]$. En déduire un nouvel estimateur de θ et montrer qu'il est consistant et asymptotiquement normal.
3. On note $X_{(1)} = \min_{1 \leq i \leq n} X_i$.
   (a) Donner la fonction de répartition de $X_{(1)}$.
   (b) En déduire la loi limite de $n\left(X_{(1)} - \theta\right)$.
   (c) Construire un intervalle de confiance non-asymptotique pour θ de niveau $1 - \alpha$.
   (d) Soit $\theta_0 > 0$. Construire un test de taille $\alpha \in (0, 1)$ pour tester
$$H_0 : \theta \geq \theta_0 \qquad \text{contre} \qquad H_1 : \theta < \theta_0.$$
   (Un test T de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta \setminus \Theta_0$ est de taille α si $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T = 1) = \alpha$.)

Exercice 3 (Régression linéaire polynomiale)

Soient $(z_i, Y_i)_{i=1}^n$ des observations d'un modèle de régression polynomiale, noté (M_r), donné par

$$Y_i = p_r(z_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. avec $\sigma > 0$ inconnu et où p_r est un polynôme inconnu de degré r (r connu). On écrit p_r comme

$$p_r(z) = \sum_{k=0}^{r} a_k z^k, \quad z \in \mathbb{R},$$

avec des coefficients inconnus $(a_k)_{0 \leq k \leq r}$. On suppose $r < n$.

1. Écrire le modèle (M_r) sous forme de modèle linéaire : $\mathbf{Y} = \mathbf{X}_r \boldsymbol{\beta}_r + \varepsilon$. Donner une condition nécessaire et suffisante pour que, pour tout $r < n$, la matrice $\mathbf{X}_r$ soit de rang plein. On supposera dans la suite qu'elle est vérifiée.
2. Donner l'expression générale de l'estimateur des moindres carrés $\widehat{\boldsymbol{\beta}}_r$ de $\boldsymbol{\beta}_r$ et d'un estimateur sans biais de la variance σ^2. Préciser la loi jointe de ces estimateurs.
3. Pour savoir s'il faut utiliser un polynôme de degré r ou si un polynôme de degré $r-1$ est suffisant, on peut comparer les modèles (M_{r-1}) et (M_r). Autrement dit, on considère les hypothèses $H_0 : a_r = 0$ et $H_1 : a_r \neq 0$. Proposer un test de niveau α pour tester H_0 contre H_1.
4. Soit $r_{\max} < n$ fixé. Décrire une procédure qui utilise le test de la question précédente pour choisir un modèle parmi les modèles (M_r) pour $r = 0, \ldots, r_{\max}$.

Supposons maintenant que les $(z_i, Y_i)_{i=1}^n$ sont des observations du modèle $(M_\star)$ défini par

$$(M_\star): \quad Y_i = h(z_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. avec $\sigma > 0$ inconnu et h est une fonction inconnue. Notre but est l'estimation du vecteur $H = (h(z_1), \ldots, h(z_n))^T$. On suppose que toutes les valeurs $z_1, \ldots, z_n$ sont distinctes, et, pour $r < n$ on note

$$p_r = \arg\min_{q \in \mathcal{P}_r} \sum_{i=1}^{n} (h(z_i) - q(z_i))^2,$$

où $\mathcal{P}_r$ est l'ensemble de polynômes de degré au r (au plus).

5. Soit $r < n$ fixé. Une approximation de H est donnée par le vecteur $\widehat{\mathbf{Y}}_r = \mathbf{X}_r \widehat{\boldsymbol{\beta}}_r$ des valeurs ajustées dans le modèle (M_r) avec $p_r \in \mathcal{P}_r$ défini ci-dessus. Montrer que sous $(M_\star)$ la loi de $\widehat{\mathbf{Y}}_r$ est donnée par
$$\widehat{\mathbf{Y}}_r \sim \mathcal{N}((p_r(z_1), \ldots, p_r(z_n))^T, \sigma^2 P_r),$$
où P_r désigne la matrice de projection orthogonale sur l'espace engendré par les colonnes de $\mathbf{X}_r$.
6. Montrer que
$$\mathbb{E}\left[\|\widehat{\mathbf{Y}}_r - H\|^2\right] = \mathbf{b}_r^2 + \mathsf{trace}(\mathbf{V}_r)\,.$$
7. Notons $\mathbf{b}_r = \|\mathbb{E}[\widehat{\mathbf{Y}}_r] - H\|$ la norme du biais de $\widehat{\mathbf{Y}}_r$ et $\mathbf{V}_r$ la matrice de covariance de $\widehat{\mathbf{Y}}_r$. Montrer que, sous $(M_\star)$, on a $\mathbf{b}_r \leq \mathbf{b}_{r-1}$ et $\mathbf{V}_r \succ \mathbf{V}_{r-1}$ (au sens où $\mathbf{V}_r - \mathbf{V}_{r-1}$ est semi-définie positive).

Exercice 4 (Test de Kolmogorov–Smirnov)

Soient $X_1, \dots, X_n$ des variables aléatoires réelles i.i.d. de fonction de répartition F continue sur $\mathbb{R}$. Soit F_0 une fonction de répartition continue sur $\mathbb{R}$. On souhaite tester, au niveau $\alpha \in]0,1[$,

$$H_0 : F = F_0 \quad \text{contre} \quad H_1 : F \neq F_0 .$$

Pour cela, on construit la fonction de répartition empirique $\widehat{F}_n$ définie par

$$\forall x \in \mathbb{R}, \quad \widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} ,$$

et l'on définit la statistique

$$D_n = \|\widehat{F}_n - F_0\|_\infty = \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F_0(x) \right| .$$

Le théorème de Glivenko–Cantelli assure que, sous H_0, on a $D_n \xrightarrow{\text{p.s.}} 0$. Cela conduit à chercher un test de la forme $T = \mathbb{1}_{\{D_n \geq c_\alpha\}}$ pour c_α bien choisi.

1. Pour G une fonction de répartition, l'inverse généralisée de G, notée G^{-1} est définie par

   $$\forall u \in [0,1], \quad G^{-1}(u) = \inf\{x \in \mathbb{R},\, G(x) \geq u\} ,$$

   avec les conventions $\inf \mathbb{R} = -\infty$ et $\inf \emptyset = +\infty$. Montrer que pour tout $x \in \mathbb{R}$ et pour tout $u \in [0,1]$, on a
   $$G(x) \geq u \;\Leftrightarrow\; G^{-1}(u) \leq x .$$
   En déduire que si $U \sim$ Unif$[0,1]$, alors $G^{-1}(U)$ est de fonction de répartition G.
2. Montrer que, sous H_0, la variable D_n a la même loi que

   $$K_n = \sup_{u \in [0,1]} \left| \widehat{G}_n(u) - u \right| ,$$

   où $\widehat{G}_n$ est la fonction de répartition empirique associée à un échantillon $U_1, \dots, U_n$ i.i.d. de loi uniforme sur $[0,1]$.
3. On admet que pour tout $z \in \mathbb{R}_+^*$,

   $$\mathbb{P}\left(\sqrt{n} K_n > z\right) \underset{n \to \infty}{\longrightarrow} \varphi(z) = 2 \sum_{k=1}^{+\infty} (-1)^{k-1} e^{-k^2 z^2} .$$

   La loi de fonction de répartition $1 - \varphi$ est appelée loi de Kolmogorov–Smirnov. En utilisant la fonction φ, construire un test de niveau asymptotique α pour H_0 contre H_1.

Exercice 5 (Méthode de Box–Muller)

Soient X et Y deux variables aléatoires de loi $\mathcal{N}(0,1)$ indépendantes. On note (R, θ) les coordonnées polaires du point (X, Y) (avec R le rayon, et θ l'angle).

1. Déterminer la loi de (R, θ).
2. Déterminer une fonction $\psi : [0,1]^2 \to \mathbb{R}^2$ telle que, si U et V sont deux variables aléatoires de loi uniforme sur $[0,1]$, alors $\psi(U, V)$ a la même loi que (X, Y).