

TP Statistiques et Séries Chronologiques

Université de Lorraine

Partiel 2015

Clément Dell'Aiera

1 Estimation de la fonction de répartition

On se donne un n -échantillon X_1, \dots, X_n i.i.d suivant une loi donnée par la même fonction de répartition (f.d.r) F sur \mathbb{R} . \mathcal{F} dénote l'ensemble des fonctions de répartition sur \mathbb{R} .

1. Décrire l'expérience statistique.
2. Le modèle est-il dominé ?
3. On veut estimer $F(x) = \mathbb{P}(X \leq x)$.
 - (a) On pose $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}$. Calculer $\mathbb{E}[\hat{F}_n(x)]$ et $V[\hat{F}_n(x)]$.
 - (b) Montrer que $\hat{F}_n(x)$ converge presque-sûrement vers $F(x)$.
 - (c) Montrer que, si $l(x, y) = (x - y)^2$ est la perte quadratique,

$$\sup_{F \in \mathcal{F}} \mathbb{E}[l(\hat{F}_n(x), F(x))] = \frac{1}{4n}.$$

- (d) En déduire que $\hat{F}_n(x)$ converge uniformément en norme \mathcal{L}^2 vers $F(x)$, et donc en probabilité.
4. (a) Montrer que

$$\mathbb{P}(|\hat{F}_n(x) - F(x)| > t) \leq \frac{1}{t^2} \text{Var}[\hat{F}_n(x)] \leq \frac{1}{4nt^2}$$

- (b) Soit $\alpha \in]0; 1[$. Déterminer

$$t_{\alpha, n} = \inf\{t > 0 : \frac{1}{4nt^2} \leq \alpha\}$$

et en déduire un intervalle de confiance pour $F(x)$ de niveau $1 - \alpha$.

- (c) Comment interpréter $I_{n, \alpha}$? Quelle est sa précision ?
5. On pose $\xi_n = \sqrt{n} \frac{\hat{F}_n(x) - F(x)}{\sqrt{\hat{F}_n(x)(1 - \hat{F}_n(x))}}$.
 - (a) Déterminer la limite en loi de ξ_n .
 - (b) On note $J_{n, \alpha}$ l'intervalle $[-\phi^{-1}(1 - \frac{\alpha}{2}); \phi^{-1}(1 - \frac{\alpha}{2})]$. Calculer la limite de $\mathbb{P}(\xi_n \in J_{n, \alpha})$ lorsque n tend vers ∞ .

- (c) Donner un intervalle de confiance asymptotique pour $J_{n,\alpha}$, ainsi que sa précision asymptotique.
6. Soient Y_j des variables aléatoires réelles indépendantes centrées : $\mathbb{E}Y_j = 0$ et bornées : $a_j \leq Y_j \leq b_j$. On veut démontrer ce que l'on appelle l'*inégalité de Hoeffding* : pour tout $t > 0$,

$$\mathbb{P}(\sum Y_j \geq t) \leq e^{-st} \prod e^{\frac{s^2(b_j - a_j)^2}{8}} \quad \forall s > 0.$$

On pose $\Phi_Y(s) = \log \mathbb{E}[e^{s(Y - \mathbb{E}Y)}]$.

- (a) Montrer que

$$\Phi_Y''(s) = e^{-\Phi_Y(s)} \mathbb{E}[Y^2 e^{sY}] - e^{-2\Phi_Y(s)} (\mathbb{E}[Y e^{sY}])^2.$$

- (b) On définit une nouvelle mesure de probabilité par $\mathbb{Q}(A) = e^{-\Phi_Y(s)} \mathbb{E}[e^{sY} 1_A]$ pour tout borélien A . Comment interpréter $\Phi_Y''(s)$ dans ce cadre ?
- (c) Montrer alors que $\Phi_Y(s) \leq s^2 \frac{(b-a)^2}{8}$.
- (d) En déduire l'inégalité de Hoeffding.
7. (a) Soient X_j des v.a. de Bernoulli de paramètre p et $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$, montrer que

$$\mathbb{P}(|\bar{X}_n - p| \geq t) \leq 2e^{-2nt^2}.$$

- (b) En déduire un intervalle de confiance de niveau $1 - \alpha$ pour $F(x)$.
8. Comparer les différents intervalles de confiance que vous avez obtenu.