

MAP433 Statistique

PC1: Rappels de probabilités.

Modélisation statistique

29 août 2014

1 Théorème central limite

Soit $(X_n)_n$ une suite de variables aléatoires i.i.d. centrées de variance $\sigma^2 > 0$. Soit

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j.$$

Par le théorème central limite, cette variable converge en loi vers la loi normale centrée réduite, c'est-à-dire, pour tout $t \in \mathbb{R}$, on a $\lim_{n \rightarrow +\infty} \mathbb{E}[e^{itZ_n}] = e^{-\frac{t^2}{2}}$. L'objet de cet exercice est de montrer que la suite Z_n ne peut pas converger en probabilité.

1. Calculer la fonction caractéristique de $Z_{2n} - Z_n$ et montrer que cette différence converge en loi.
2. En étudiant $\mathbb{P}(|Z_{2n} - Z_n| \geq \varepsilon)$, montrer que Z_n ne converge pas en probabilité.

2 Lemme de Slutsky

1. Donner un exemple de suites (X_n) et (Y_n) telles que $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\text{loi}} Y$, mais $X_n + Y_n$ ne converge pas en loi vers $X + Y$.
2. Soient $(X_n), (Y_n)$ deux suites de variables aléatoires réelles, X et Y des variables aléatoires réelles, telles que
 - (i) $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$,
 - (ii) Y est indépendante de (X_n) et X .

Montrer que le couple (X_n, Y_n) converge en loi vers (X, Y) .

3. En déduire que si (X_n) et (Y_n) sont deux suites de variables aléatoires réelles telles que (X_n) converge en loi vers une limite X et (Y_n) converge en probabilité vers une constante c , alors $(X_n + Y_n)$ converge en loi vers $X + c$ et $(X_n Y_n)$ converge en loi vers cX .

3 Estimateur de la variance

Soient X_1, \dots, X_n des variables aléatoires i.i.d., $X_i \sim f(\cdot - \theta)$, où f est une densité de probabilité sur \mathbb{R} symétrique dont on note $\mu_k = \int_{\mathbb{R}} x^k f(x) dx$ les moments d'ordre $k = 2$ et $k = 4$. On note $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Montrer que l'estimateur $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ de la variance des X_i vérifie un théorème central limite.

Indication : on montrera d'abord que l'on peut se ramener au cas où $\theta = 0$, puis on exprimera l'estimateur comme une transformation de $S_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ et de \bar{X}_n .

4 Taux de défaillance

Une chaîne de production doit garantir une qualité minimale de ses produits. En particulier, elle doit garantir que la proportion θ des produits défectueux reste inférieure à un taux fixé par le client. Un échantillon de n produits est prélevé et analysé. On note $\hat{\theta}_n$ la proportion de produits défectueux dans l'échantillon.

1. Proposer un modèle statistique pour ce problème. Quelle est la loi de $n\hat{\theta}_n$?
2. Quelle information donne la loi des grands nombres et le théorème central limite sur le comportement asymptotique de $\hat{\theta}_n$?
3. On donne $\mathbb{P}(N > 1.64) = 5\%$ pour $N \sim \mathcal{N}(0, 1)$. En déduire ε_n (dépendant de n et θ) tel que $\mathbb{P}(\theta \geq \hat{\theta}_n + \varepsilon_n) \xrightarrow{n \rightarrow \infty} 5\%$.
4. La valeur ε_n précédente dépend de θ . A l'aide du lemme de Slutsky, donner ε'_n ne dépendant que de n et $\hat{\theta}_n$ tel que $\mathbb{P}(\theta \geq \hat{\theta}_n + \varepsilon'_n) \xrightarrow{n \rightarrow \infty} 5\%$.