X2015 – MAP 311
PC 7 – 13 juin 2016 – TCL et intervalles de confiance

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

Exercice 1. **(Manipulations sur les gaussiennes)** On rappelle qu'une variable aléatoire gaussienne de paramètres (m, σ^2) a pour fonction caractéristique $\phi(t) = e^{imt - \frac{\sigma^2 t^2}{2}}$.

(1) Soit $X = \mathcal{N}(0,1)$ une loi gaussienne centrée réduite. Quelle est la loi de $m + \sigma X$?

(2) Soient $X = \mathcal{N}(m_1, \sigma_1^2)$ et $Y = \mathcal{N}(m_2, \sigma_2^2)$ deux variables aléatoires gaussiennes indépendantes. Quelle est la loi de $X + Y$? Ce résultat reste-t-il vrai si X et Y ne sont pas indépendantes ?

(3) Soit $(X_k)_{k \geq 1}$ une suite de gaussiennes centrées réduites indépendantes. On pose

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} \sqrt{k} X_k.$$

Étudier la convergence en loi de Y_n.

1 Intervalles de confiance

Exercice 2. **(Monte Carlo et Intervalle de confiance pour π)** Soit $((X_i, Y_i))_{i \geq 1}$ une suite de vecteurs aléatoires indépendants tels que pour tout $i \geq 1$, X_i et Y_i sont des variables aléatoires indépdantes de loi uniforme sur $[0,1]$. Posons $Z_i = 1$ si $X_i^2 + Y_i^2 \leq 1$ et $Z_i = 0$ sinon. Finalement, on pose

$$S_n = \frac{4}{n} \cdot \sum_{i=1}^{n} Z_i,$$

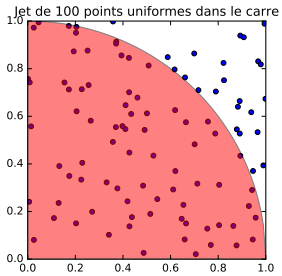

FIGURE 1 – Dans ce exemple, $S_{100} = \frac{4}{100} \cdot 77$ (il y a 77 points dans la région pleine).

(1) Identifier la loi de Z_1.

(2) Montrer que S_n converge presque sûrement vers π.

(3) Montrer que $\mathbb{P}(|S_n - \pi| \geq x) \leq \frac{4}{x^2 n}$ pour tout $x > 0$.
Indication. On pourra utiliser l'inégalité de Bienaymé–Tchebychev.

(4) Expliquer comment simuler un intervalle de confiance de π d'amplitude au plus 10^{-2} à 99%

Exercice 3. **(Intervalles de confiance asymptotiques avec le TCL)** Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi. On suppose que X_1 est de carré intégrable, de moyenne m et de variance $\sigma^2 > 0$. On pose

$$\widehat{m}_n = \frac{X_1 + \cdots + X_n}{n}, \qquad \widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \widehat{m}_n)^2.$$

(1) Justifier que $\sqrt{n}\cdot\frac{\widehat{m}_n - m}{\sigma}$ converge en loi vers $\mathcal{N}(0,1)$, une gaussienne centrée réduite.

(2) En déduire un intervalle de confiance asymptotique pour m au niveau 95% (en supposant σ connu).

(3) Montrer que $\widehat{\sigma}_n$ converge presque sûrement vers σ. L'estimateur $\widehat{\sigma}_n^2$ est-il sans biais ?
Indication. On pourra démontrer que $\frac{n-1}{n}\cdot\widehat{\sigma}_n^2 = \left(\frac{1}{n}\sum_{k=1}^{n} X_k^2\right) - \overline{X}_n^2$.

(4) Montrer que

$$\sqrt{n}\cdot\frac{\widehat{m}_n - m}{\widehat{\sigma}_n} \quad \xrightarrow[n\to\infty]{\text{loi}} \quad \mathcal{N}(0,1).$$

(5) En déduire un intervalle de confiance asymptotique pour m au niveau 95% (en supposant σ inconnu).

Exercice 4. Le but de cet exercice est d'estimer le temps d'attente du RER B, qui suit une loi exponentielle de paramètre $\lambda > 0$ inconnu. Soit $(E_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de même loi exponentielle de paramètre $\lambda > 0$ inconnu. On pose $\widehat{\lambda}_n = \frac{E_1 + \cdots + E_n}{n}$ et

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(E_k - \widehat{\lambda}_n)^2.$$

(1) Donner un intervalle de confiance asymptotique pour $\frac{1}{\lambda}$ au niveau 95%.

(2) Comment obtenir un intervalle de confiance pour λ au niveau 95% ?

2 À chercher pour la prochaine fois

Exercice 5. Avant un référendum, on effectue une enquête pour estimer la proportion p de personnes votant oui. On interroge un échantillon représentatif de n personnes et on note $\widehat{F}_n$ la proportion du nombre de réponses oui dans l'échantillon.

(1) Quelle est la loi de nF_n (en faisant des hypothèses raisonnables) ?

(2) Démontrer que

$$\frac{\sqrt{n}}{\sqrt{\widehat{F}_n(1-\widehat{F}_n)}}\left(\widehat{F}_n - p\right) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

En déduire un intervalle de confiance asymptotique $\widehat{I}_n$ à 95% pour p.

(3) En utilisant l'inégalite de Bienaymé-Tchebychev, obtenir un intervalle de confiance exact $\widehat{J}_n$ à 95%.

(4) On rappelle l'inégalité de concentration (vue en Amphi 3) pour la loi binomiale :

$$\mathbb{P}\left(|\mathsf{Binom}(n,p) - np| > r\right) \leq 2\exp\left(-2\frac{r^2}{n}\right).$$

Utiliser cette inégalité pour obtenir un intervalle de confiance exact $\widehat{K}_n$ à 95%.

(5) Quel intervalle choisiriez-vous ?

3 Plus appliqué

Exercice 6. On effectue une enquête, durant une épidémie de grippe, dans le but de connaître la proportion p de personnes présentant ensuite des complications graves. On observe un échantillon représentatif de 400 personnes et pour un tel échantillon 40 personnes ont présenté des complications.

(1) Donner un intervalle de confiance pour p au risque 5%.

(2) On désire que la valeur estimée $\widehat{p}$ diffère de la proportion inconnue exacte p de moins de 0.005 avec une probabilité égale à 95%. Quel sera l'effectif d'un tel échantillon ?

(3) Quel devrait être le risque pour obtenir le même intervalle qu'à la question précédente en conservant l'effectif $n = 400$? Quelle conclusion peut-on en tirer ?

4 Pour aller plus loin

Exercice 7. **(Stabilisation de la variance)** On dispose d'un échantillon $X_1, \ldots, X_n$ de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $0 < \theta < 1$.

(1) On note $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$ la moyenne empirique des X_i. Que donne la loi forte des grands et let TCL ?

(2) Trouver une fonction g telle que $\sqrt{n}(g(\overline{X}_n) - g(\theta))$ converge en loi vers une loi loi gaussienne centrée réduite.

(3) On note z_α le quantile d'ordre $1-\alpha/2$ de la loi normale ($\mathbb{P}(Z \geq z_\alpha) = \alpha/2$ si Z est une loi normale centrée réduite). En déduire un intervalle de confiance asymptotique $\widehat{I}_{n,\alpha}$ (qui dépend de z_α, n et $\overline{X}_n$) tel que $\lim_{n\to\infty} \mathbb{P}\left(\theta \in \widehat{I}_{n,\alpha}\right) = 1-\alpha$.

Exercice 8. **(Estimateurs linéaires)** Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes de même loi et de carré intégrable. Trouver l'estimateur $\widehat{\theta}_n$ de la moyenne $\theta = \mathbb{E}[X_1]$, qui soit sans biais (c'est-à-dire $\mathbb{E}\left[\widehat{\theta}_n\right] = \theta$ et de variance minimale dans la classe des estimateurs linéaires $\widehat{\theta}_n = \sum_{k=1}^{n} a_k X_k$.

Exercice 9. Soit $g : [0,1] \to [0,1]$ une fonction (mesurable) bornée. On souhaite calculer $m = \int_0^1 g(x)dx$. On pose $\sigma^2 = \int_0^1 g(x)^2 dx - m^2$. Soient X et Y des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose

$$U = \mathbb{1}_{Y \leq g(X)}, \qquad V = g(X), \qquad W = \frac{g(X) + g(1-X)}{2}.$$

(1) Calculer l'espérance et la variance de U, V, W. Comparer les variances de U et V.

(2) Proposer trois méthodes de type Monte-Carlo pour calculer m.

On suppose dans la suite que g est monotone.

(3) Vérifier que $\mathbb{E}[g(X)g(1-W)] \leq m^2$ et comparer les variances de V et W.

Indication. On pourra montrer que $(g(x)-g(y))(g(1-x)-g(1-y)) \leq 0$ pour tout $x, y \in [0,1]$.

Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

(4) On considère les estimateurs suivants de m :

$$A_n = \frac{1}{2n} \sum_{k=1}^{2n} g(X_k) \qquad \text{et} \qquad \frac{1}{2n} \sum_{k=1}^{n} (g(X_k) + g(1-X_k)).$$

Montrer qu'ils sont sans biais. Lequel possède la plus petite variance ?

(5) Dans le cas où $g(x) = x^2$, déterminer le nombre n de simulations nécessaires garantissant une précision relative de 1% sur le calcul de m en erreur quadratique avec A_n et B_n (la précision relative étant $\frac{\mathrm{Var}(A_n)}{m^2}, \frac{\mathrm{Var}(B_n)}{m^2}$).