Projet à rendre Probabilités

Université de Thiès, UFR SES
Département de Management des Organisations
Master 1 Sciences de Données et Applications

Année universitaire 2019-2020
Fatou Néné Diop

1 Loi faible des grands nombres et simulations sous Python

Théorème 1.1 (Loi faible des grands nombres ou théorème de Khintchine) *Soient des variables aléatoires* $Y_1, \cdots, Y_n$ *indépendantes telles que* $\mathbf{E}(Y_i) = m$ *et* $\mathbf{V}(Y_i) = \sigma^2$, $\forall i = 1, \ldots, n$. *La moyenne empirique de ces variables converge en probabilité vers l'espérance* m *:*

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{\mathbf{P}} \mathbf{E}(Y_i) = m. \tag{1}$$

Exercice 1:

On considère n variables aléatoires discrètes $Z_1, \cdots, Z_n$ indépendantes et identiquement distribuées (i.i.d.) telles que $Z_i \sim \mathcal{P}(\lambda)$. D'après les propriétés de la loi de Poisson, on sait que $\mathbf{E}(Z_i) = \lambda$. Par conséquent, d'après le théorème de Khintchine, on a :

$$\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{\mathbf{P}} \lambda \tag{2}$$

Afin d'illustrer cette propriété, menons l'expérience suivante. On considère des variables aléatoires indépendantes et identiquement distribuées selon une loi uniforme $Y_i \sim U_{[0,10]}$, $\forall 1 = 1, \ldots, n$, avec $\mathbf{E}(Y_i) = 5$.

On applique la procédure suivante :

1. Grâce au logiciel Python, on tire des réalisations $\{y_1, ..., y_n\}$ des n variables $\{Y_1, \cdots, Y_n\}$.
2. On calcule une réalisation de la moyenne empirique $\overline{Y}_n$. Cette réalisation est notée $\bar{y}_n = n^{-1}\sum_{i=1}^{n} y_i$.
3. On répète 5 000 fois les étapes 1 et 2. On obtient ainsi 5 000 réalisations de la variable $\overline{Y}_n$.
4. On construit l'histogramme de ces 5 000 réalisations.
5. On répète l'expérience pour différentes valeurs de la dimension n (taille d'échantillon). Prendre les valeurs $n = 10$, $n = 100$, $n = 1000$ et $n = 10000$. Il convient de ne pas confondre ici la taille d'échantillon n (par exemple 3) et le nombre de réplications (5 000).

Tracer les histogrammes des 10000 réalisations $\bar{y}_n$ obtenues pour quatre valeurs de n, à savoir $n = 10$, $n = 100$, $n = 1000$ et $n = 10000$.

Commenter les résultats obtenus.

2 Théorème central limite et simulations sous Python

Théorème 2.1 (Théorème central limite, Lindeberg-Levy) *Soit $Y_1, \cdots, Y_n$ une séquence de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) avec une espérance finie $E(Y_i) = m$ et une variance finie $\mathbf{V}(Y_i) = \sigma^2$, $\forall i = 1, \ldots, n$. Alors la moyenne empirique $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ vérifie :*

$$\sqrt{n}\left(\frac{\overline{Y}_n - m}{\sigma}\right) \xrightarrow{Loi} \mathcal{N}(0,1). \tag{3}$$

Exercice 2:

On considère des variables aléatoires indépendantes et identiquement distribuées selon une loi du khi-deux $Y_i \sim \chi^2(2)$, $\forall i = 1, \ldots, n$, avec $\mathbf{E}(Y_i) = 2$ et $\mathbf{V}(Y_i) = 4$. On applique la procédure suivante :

1. Grâce au logiciel Python, on tire des réalisations $\{y_1, \ldots, y_n\}$ des n variables $\{Y_1, \ldots, Y_n\}$.
2. On calcule une réalisation de la moyenne empirique $\overline{Y}_n$. Cette réalisation est notée $\bar{y}_n = n^{-1}\sum_{i=1}^{n} y_i$.
3. On considère la variable aléatoire transformée :

$$Z_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mathbf{E}(Y_i)}{\sqrt{\mathbf{V}(Y_i)}}\right) = \sqrt{n}\left(\frac{\overline{Y}_n - 2}{2}\right) \tag{4}$$

   À partir de la réalisation de la moyenne empirique $\bar{y}_n$, on calcule une réalisation decette variable transformée comme $z_n = \sqrt{n}\left(\overline{y}_n - 2\right)/2$.
4. On répéte cette procédure 5 000 fois (étapes 1 à 3). On obtient alors 5 000 réalisations de la variable z_n.
5. On construit un histogramme de ces 5 000 réalisations et l'on compare cet histogramme à la densité d'une loi normale centrée réduite $\mathcal{N}(0,1)$.

Exercice 3: *Théorème cental limite appliqué au traitement de texte*

Le texte "LesMiserables1" est disponible sur le site (web www.gutenberg.org).

- Télécharger https ://r-stat-sc-donnees.github.io/LesMiserables1.txt.
- Supprimer la ponctuation, les majuscules / minuscules et divisez le texte en jetons individuels (mots).
- Pour les mots (*un*, *et*, *le*, *il*, *est*) calcule leurs nombres respectifs au fur et à mesure que le livre progresse, exemple

$$n_{le}[i] = \sum_{j=1}^{i} \{w_j = le\}.$$

- Tracer les proportions $n_{mot}[i]/i$ sur le document dans un seul graphique.
- Pourquoi ne pouvons-nous pas appliquer le théorème central limite (TCL) directement ?
- Que devons-nous faire sur le texte pour appliquer le TCL ?

```
import string
import urllib.request
x = urllib.request.urlopen('https://r-stat-sc-donnees.github.io/LesMiserables1.txt')
livre=str(x.read())
print(livre)

#Ajouter vos codes ici
```

Définition 2.1 *Soit X un vecteur aléatoire sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ à valeurs dans $\mathbb{R}^d$. La loi image $\mathbf{P}X^{-1}$ est appellée distribution de la v.a. X, et est notée $\mathbf{P}_X$.*

On appelle fonction caractéristique de X la fonction $\Phi_X : \mathbb{R}^n \to \mathbb{C}$ définie par

$$\Phi_X(u) = \mathbf{E}\left[e^{i\langle u,X\rangle}\right] \quad \textit{pour tout}\, u \in \mathbb{R}^d.$$

La fonction caractéristique dépend uniquement de la loi de X :

$$\Phi_X(u) = \int_{\mathbb{R}^n} e^{i\langle u,X\rangle} d\mathbf{P}_X(x),$$

et n'est rien d'autre que la transformée de Fourier de $\mathbf{P}_X$ au point $-u/2\pi$.

Exercice 4: *Fonction caractéristique*

1. Pour un vecteur gaussien X de moyenne b et de matrice de variance V, montrer que

$$\Phi_X(u) = e^{\langle u,b\rangle - \frac{1}{2}\langle u,Vu\rangle}.$$

2. Si $\mathbf{P}_X$ est symétrique par rapport à l'origine, i.e. $\mathbf{P}_X = \mathbf{P}_{-X}$, montrer que Φ_X est à valeurs réelles.
3. Pour une v.a. réelle, supposons que $\mathbf{E}\left[|X|^p\right] < \infty$ pour un certain entier $p \geqslant 1$. Montrer que Φ_X est p fois dérivable et

$$\Phi_X^{(k)}(0) = i^k \mathbf{E}\left[X^k\right], \quad \text{pour}\, k = 1, \ldots, p.$$