Méthodes de Monte Carlo et processus stochastiques

Aurélien Alfonsi – Stefano De Marco – Emmanuel Gobet – Clément Rey

MAP 556 - TP 1 - 19 et 21 septembre 2018

Simulation de variables aléatoires et thèoremes limites fondamentaux

Quelques lignes guide pour l'usage de Python:

- Un script `python` est un fichier texte avec l'extension `.py`. Un tel script peut être modifié avec n'importe quel editeur de texte, et executé à partir d'une console Python. Si Python n'est pas déjà installé sur votre machine, nous conseillons de télécharger la distribution *Anaconda*

  `https://www.anaconda.com/download/`

  qui installe: une version de Python, les librairies de base (`numpy`, `scipy`, `matplotlib`) et l'environnement de développement `spyder`.

- Rappel: dans Python 2.x, il faudra s'interdir l'utilisation de caractères accentués dans un script (même dans les commentaires!).

- On écrira un fichier `.py` par question, en donnant au fichier un nom explicite, comme par exemple `MAP556_TP1_Exo1_1.py`

- Nous utiliserons la programmation matricielle, en travaillant avec les variables de la librairie `numpy`, en evitant autant que possible les boucles. Cela permet d'avoir un gain en temps de calcul dans la plus part de nos applications.

EXERCICE 1 - Simulation de variables aléatoires dans Python. Loi des grands nombres.
Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires i.i.d. telles que $\mathbb{E}[|X_1|] < \infty$. La loi forte des grands nombres affirme que la suite des *moyennes empiriques*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converge presque sûrement vers $\mathbb{E}[X_1]$ lorsque $n \to \infty$.

1. En completant le fichier `MAP556_PC1_Exo1_1.py`, simuler un grand nombre N de variables aléatoires uniformes sur $[0, 1]$ indépendantes et répresenter graphiquement la suite des moyennes empiriques $\overline{X}_n$ pour n allant de 1 à N. On pourra utiliser la fonction `numpy.random.rand` pour obtenir les tirages de la loi uniforme et la fonction `numpy.cumsum` pour calculer la moyenne empirique cumulée jusqu'à n.

2. La loi de Cauchy de paramètre $a > 0$ est la loi sur $\mathbb{R}$ de densité

$$f(x) = \frac{a}{\pi(a^2 + x^2)}.$$

   (a) Soit $F(x) = \int_{-\infty}^{x} f(y)dy$ sa fonction de répartition. Soit U une v.a. de loi uniforme sur $[0, 1]$. Quelle est la loi de $F^{-1}(U)$?

(b) Fichier `MAP556_PC1_Exo1_2.py`. En remarquant que $F(x) = \frac{1}{\pi}\arctan(\frac{x}{a}) + \frac{1}{2}$, générer un grand nombre N de variables aléatoires indépendantes de loi de Cauchy de paramètre $a = 1$, toujours en utilisant la fonction `numpy.random.rand`. Répresenter graphiquement la suite des moyennes empiriques $\overline{X}_n$ pour n allant de 1 à N. Qu'observez-vous?

EXERCICE 2 - Théoreme central limite. Intervalles de confiance. Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires i.i.d. telles que $\mathbb{E}[X_1^2] < \infty$, et $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ la suite de leurs moyennes empiriques. Le théoreme central limite affirme que la suite des erreurs renormalisées

$$e_n = \frac{\sqrt{n}}{\sqrt{Var(X_1)}}\left(\overline{X}_n - \mathbb{E}[X_1]\right)$$

converge en loi vers une gaussienne $\mathcal{N}(0,1)$ lorsque $n \to \infty$.

1. (a) Fichier `MAP556_PC1_Exo2_1_a.py`. On considère une suite iid de v.a. exponentielles de paramètre 2. Pour N fixé, obtenir un échantillon de M valeurs de la moyenne empirique $(\overline{X}_N^j)_{j=1,\dots,M}$. Tracer l'histogramme des valeurs correspondantes de l'erreur normalisée

$$e_N^j = \sqrt{\frac{N}{Var(X_1)}}(\overline{X}_N^j - \mathbb{E}[X_1]), \qquad j = 1,\dots,M$$

et le comparer avec la densité gaussienne centrée réduite.

(b) Fichier `MAP556_PC1_Exo2_1_b.py`. Afin d'illuster la différence entre convergence en loi et convergence presque sûre, tracer une réalisation de la suite e_n pour n variant de 1 à N. Qu'observez-vous?

2. *Intervalles de confiance.* A différence de l'exemple précédent, en pratique on ne connait pas la valeur $Var(X_1)$. On peut cependant l'estimer à partir des mêmes simulations, en utilisant l'estimateur de la variance $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - (\overline{X}_n)^2$.

(a) Quel resultat permet d'affirmer que la suite

$$\frac{\sqrt{n}}{\sigma_n}\left(\overline{X}_n - \mathbb{E}[X_1]\right)$$

converge en loi vers $\mathcal{N}(0,1)$?

(b) En définissant l'intervalle (aléatoire) $I_n^\delta = \left[\overline{X}_n - \delta\frac{\sigma_n}{\sqrt{n}}, \overline{X}_n + \delta\frac{\sigma_n}{\sqrt{n}}\right]$, $\delta > 0$, la question précédente entraîne

$$\mathbb{P}\Big(\mathbb{E}[X_1] \in I_n^\delta\Big) = \mathbb{P}\Big(\frac{\sqrt{n}}{\sigma_n}|\overline{X}_n - \mathbb{E}[X_1]| \leq \delta\Big) \longrightarrow \mathbb{P}(|\mathcal{N}(0,1)| \leq \delta) = 2\int_0^\delta \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx$$

lorsque $n \to \infty$. Pour $\delta = 1.96$, le dernier terme à droite vaut approximativement 0.95, ce qui permet d'idéntifier $I_n^{1.96}$ comme étant l'intervalle de confiance *asymptotique* à 95% pour l'espérance $\mathbb{E}[X_1]$.

Fichier `MAP556_PC1_Exo2_2.py`. Fournir une estimation de l'espérance de v.a. uniformes sur $[0,1]$ avec son intervalle de confiance à 95% donné par le TCL.