TP Matlab

Exercice 1. Prise en main des fonctions prédéfinies pour les probabilités.

1. Tester les commandes suivantes :
   (a) `rand(2,4)`
   (b) `normrnd(1,4,5)`
   (c) `exprnd(10,2,3)`
2. Que font les commandes suivantes ?
   (a) `plot(cumsum(rand(1,1000))./[1:1000])`
   (b) `plot([1:1000]./1000,cumsum(rand(1,1000))./[1:1000],'--')`
   (c) `plot(cumsum(-log(rand(1,1000)))./[1:1000],':')`
   (d) `histogram(rand(100))`
   (e) `histogram(normrnd(0,1,1000),25,'Normalization','pdf')`
   `title('Histogramme loi normale centrée réduite')`
3. Écrire une fonction `histonormale(n)` qui génère un échantillon de n variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. Tracer ensuite l'histogramme correspondant et y superposer la densité gaussienne. *On pourra se référer à l'aide pour comprendre le fonctionnement de la commande* `hold`.

Exercice 2. Simulation par fonction de répartition inverse.

Proposition

Soit X une variable aléatoire réelle de fonction de répartition F. Pour $u \in [0,1]$, on désigne par $F^{-1}(u) = \inf\{x \in \mathbb{R} | F(x) \geqslant u\}$ l'inverse généralisée de la fonction F. Si $U \sim \mathcal{U}([0,1])$, alors la variable aléatoire $F^{-1}(U)$ a pour fonction de répartition F.

1. Écrire une fonction `generebernoulli(n,p)` qui génère n variables aléatoires indépendantes de loi $b(p)$, à partir de variables uniformes sur $[0,1]$.
2. Écrire une fonction `simulery(d)` qui génère une variable aléatoire Y de loi uniforme sur les entiers compris entre 0 et $d-1$.
3. On admet que si X suit la loi $\mathcal{E}(2)$, alors l'inverse généralisée de sa fonction de répartition est $F^{-1} : u \mapsto \dfrac{-\ln(1-u)}{2}$. *(On pourra montrer ce résultat après le TP en exercice.)* Écrire une fonction `simulerexpo(n)` qui, à partir de variables aléatoires uniformes sur $[0,1]$, génère n variables aléatoires indépendantes $X_1, \ldots, X_n$ de loi $\mathcal{E}(2)$. Mettre ensuite en évidence la loi des grands nombres en traçant la fonction $k \mapsto \dfrac{X_1 + \cdots + X_k}{k}$ pour k allant de 1 à 1000.

Exercice 3. Méthode de Monte-Carlo. La méthode de Monte-Carlo est basée sur la loi des grands nombres. Elle permet par exemple de calculer des valeurs approchées d'intégrales ou d'espérances, en utilisant des réalisations i.i.d. d'une loi que l'on sait simuler. Par exemple, si $((X_n, Y_n))_{n\geqslant 1}$ est une suite de vecteurs aléatoires i.i.d. de loi uniforme sur $[0,1]^2$, alors la loi des grands nombres entraîne la convergence en probabilité suivante :

$$\frac{f(X_1,Y_1) + \ldots + f(X_n,Y_n)}{n} \to \mathbb{E}\left[f(X_1,Y_1)\right] = \int_{[0,1]^2} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

Un tel résultat reste évidemment vrai pour des vecteurs aléatoires ayant un nombre de composantes quelconque (y compris 1). Si l'on sait majorer la variance de $f(X_1,Y_1)$, on est en plus en mesure de fournir des intervalles de confiance pour contrôler l'erreur commise dans l'approximation. La vitesse de convergence de cette méthode (de l'ordre de $\sqrt{n}$) est lente par rapport à des méthodes déterministes. Cependant, cette vitesse de dépend pas de la régularité de l'intégrande f et dépend plus faiblement du nombre de composantes du vecteur aléatoire que les méthodes déterministes. À l'aide de la méthode de Monte-Carlo, calculer une approximation de l'intégrale suivante :

$$\int_{[-1,1]^2} \mathbb{1}_{\{|y| \leqslant e^{-x^2}\}} \, dx dy.$$

Exercice 4. Estimation statistique de la moyenne d'une loi. On considère une variable aléatoire X, de loi log-normale de paramètres $\mu = 2$ et $\sigma = 1$. Cela signifie que $\ln X \sim \mathcal{N}(2, 1)$. L'objectif de cet exercice est l'obtention d'un intervalle de confiance asymptotique pour $\mathbb{E}[X]$, sans passer par des calculs exacts d'intégrales.

1. Écrire une fonction `generelognormale(n)` qui génère un échantillon de n variables aléatoires de loi log-normale de paramètres $\mu = 2$ et $\sigma = 1$.
2. Appeler `generelognormale(1000)`. Calculer la moyenne et la variance empiriques de l'échantillon obtenu.
3. On admet que pour n grand ($n = 1000$ convenant amplement), on puisse dire en faisant une approximation raisonnable que $\sqrt{n}\frac{\mathbb{E}[X] - \overline{X_n}}{\sqrt{\sigma_n^2}} \sim \mathcal{N}(0, 1)$. On peut montrer alors que

$$\left[\overline{X_n} - 1,96\sqrt{\frac{\sigma_n^2}{n}}, \overline{X_n} + 1,96\sqrt{\frac{\sigma_n^2}{n}}\right]$$

   est un intervalle de confiance asymptotique de niveau de confiance 95% pour $\mathbb{E}[X]$. *(Le faire après le TP en exercice !)* Écrire une fonction `courbeica(n)` qui, à partir d'un échantillon de n variables aléatoires indépendantes de loi log-normale de paramètres $\mu = 2$ et $\sigma = 1$, trace la moyenne empirique des k premières valeurs obtenues (k allant de 1 à n), ainsi que les bornes de l'intervalle de confiance (qui sont censées se resserrer sur le graphique au fur et à mesure que k augmente).