Simulations stochastiques - TD 3

Théorèmes limites et méthode de Monte Carlo

Exercice 1. On considère une suite $(X_k)_{k\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées, dont la loi a pour densité $f : \mathbb{R} \to \mathbb{R}_+$ définie par

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ xe^{-x} & \text{si } x \geq 0 \end{cases}$$

1. Expliquer pourquoi f est bien une densité de probabilité sur $\mathbb{R}$.
2. Calculer l'espérance et la variance de X_1.
3. Soit $n \in \mathbb{N}^*$. Calculer l'espérance et la variance de $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Donner une limite presque sûre de $\bar{X}_n$ lorsque n tend vers l'infini.
4. Montrer que la loi de X_1 admet une densité par rapport à la loi exponentielle de paramètre $1/2$. Montrer que cette densité est bornée par $4/e$.
5. ♣ Rappeler comment simuler une variable exponentielle de paramètre $1/2$. En se basant sur la question 4, utiliser la méthode du rejet pour donner une simulation de $(X_k)_{1\leq k\leq 10000}$. Comparer l'histogramme et f et vérifier la convergence de $\bar{X}_n$.
6. Soit $n \in \mathbb{N}^*$ et $t \in \mathbb{R}_+^*$. Appliquer l'inégalité de Bienaymé-Tchebychev pour montrer que
$$\mathbb{P}\left(\left|\bar{X}_n - 2\right| > \frac{t}{\sqrt{n}}\right) \leq \frac{2}{t^2} .$$
7. En déduire que l'intervalle aléatoire
$$I_n = \left[\bar{X}_n - \frac{3}{\sqrt{n}} ; \bar{X}_n + \frac{3}{\sqrt{n}}\right]$$
contient la valeur de $\mathbb{E}(X_1)$ avec probabilité supérieure ou égale à 75%. Pourquoi est-ce naturel que cet intervalle contienne souvent la valeur de $\mathbb{E}(X_1)$? A quelle vitesse décroît sa largeur (en fonction de n) ?
8. ♣ Réaliser $N = 500$ simulations de $\bar{X}_n$ pour $n = 1000$. Donner une approximation de la probabilité l'évènement $\mathbb{E}(X_1) \in I_n$. Faites le lien avec la borne donnée question précédente. Commenter le caractère conservatif de cette borne.
9. ♣ Enoncer le théorème central limite pour $\bar{X}_n$. Utiliser votre dernière simulation pour illustrer graphiquement ce résultat. Proposer une borne moins conservative que celle de la question 7 et compararer au résultat de la question 8.

Exercice 2. On considère l'intégrale suivante :

$$I = \int_0^{\infty} e^{-x^2/2} dx.$$

On prend une suite $(X_k)_{k\geq 1}$ de variables exponentielles indépendantes de paramètre 1, et on pose $Y_k = e^{X_k - X_k^2/2}$, $I_n = (Y_1 + Y_2 + \cdots + Y_n)/n$.

1. Montrer que I_n tend presque sûrement vers I quand n tend vers l'infini.
2. Trouver une expression Σ_n, fonction uniquement de $Y_1, \ldots, Y_n$, qui tende presque sûrement vers l'écart-type de Y_1 quand n tend vers l'infini.
3. Montrer que pour n suffisamment grand, $\mathbb{P}(I \in [I_n - 2\Sigma_n/\sqrt{n}, I_n + 2\Sigma_n/\sqrt{n}])$ est supérieur à 0.95. Quels sont les deux résultats du cours qui sont utiles ici ?
4. Quelle est la valeur exacte de I ?
5. ♣ Simuler 100 fois le calcul de l'intervalle donné dans (3), pour $n = 10000$, et déterminer combien de fois I tombe en dehors de cet intervalle aléatoire.

Exercice 3. On considère l'intégrale suivante sur $[0,1]^6$:

$$I := \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2} dx dx' dy dy' dz dz'.$$

1. Donner une interprétation géométrique de cette intégrale.
2. Soient $(X_k)_{k\geq 1}$, $(Y_k)_{k\geq 1}$, $(Z_k)_{k\geq 1}$, $(X'_k)_{k\geq 1}$, $(Y'_k)_{k\geq 1}$, $(Z'_k)_{k\geq 1}$ des variables indépendantes uniformes sur $[0,1]$. Déterminer une suite I_n de variables aléatoires, uniquement fonctions des variables $X_k, Y_k, Z_k, X'_k, Y'_k, Z'_k$ pour $1 \leq k \leq n$, qui tende presque sûrement vers I quand n tend vers l'infini.
3. Montrer que la variance de toute variable aléatoire à valeurs dans un intervalle $[a,b]$ est inférieure ou égale à $(b-a)^2/4$ (on pourra d'abord supposer $a = -b$).
4. Montrer que
$$\mathrm{Var}\left[\sqrt{(X_1 - X'_1)^2 + (Y_1 - Y'_1)^2 + (Z_1 - Z'_1)^2}\right] \leq 3/4.$$
5. Montrer que pour n assez grand
$$\mathbb{P}\left(I \in [I_n - \sqrt{3/n}, I_n + \sqrt{3/n}]\right) > 0.95.$$
6. Montrer que pour n assez grand, la borne supérieure I_n^+ de l'intervalle aléatoire précédent est inférieure ou égale à $I + 2\sqrt{3/n}$ avec probabilité supérieure à 0.95.
7. La valeur exacte de I est (!)
$$I = \frac{1}{105}(4 + 17\sqrt{2} - 6\sqrt{3} + 21\log(1+\sqrt{2}) + 84\log(1+\sqrt{3}) - 42\log(2) - 7\pi) \simeq 0.6617.$$
Trouver l'ordre de grandeur minimum que n doit avoir pour qu'avec probabilité au moins 0.95, l'intervalle de confiance donné en 5) ne contienne pas 2/3.
8. ♣ Simuler l'intervalle donné en (5) pour $n = 10000$, puis pour $n = 1000000$, et regarder dans chacun de ces deux cas si 2/3 est compris dedans. Cela confirme-t-il le calcul fait dans 7) ?

Exercice 4. ♣

1. Rappeler le résultat de convergence de la loi géométrique (renormalisée) vers une loi exponentielle.
2. Écrire une fonction qui, pour une valeur de p donnée, génère un vecteur de 1000 tirages indépendants d'une variable aléatoire X de loi géométrique de paramètre p.

3. Vérifier à l'aide d'histogrammes que, lorsque p tend vers 0, p fois la loi géométrique de paramètre p tend vers la loi exponentielle de paramètre 1.

Exercice 5. ♣

1. Rappeler le résultat de convergence de la loi binomiale vers la loi de Poisson. Quelles sont les conditions à vérifier ? Donner un exemple de la vie de tous les jours qui correspond à ces conditions.
2. Écrire une fonction qui génère un vecteur de 1000 tirages indépendants d'une variable aléatoire X de loi binomiale $\mathcal{B}(n, p)$.
3. Vérifier à l'aide de graphiques que, lorsque n tend vers l'infini, la loi binomiale $\mathcal{B}\big(n, 3/n\big)$ tend vers la loi de Poisson de paramètre 3.

Exercice 6. ♣
On donne les matrices suivantes

$$L := \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix} \qquad \Sigma := LL^T = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}.$$

Combiner cette observation avec l'algorithme de Box-Müller vu au TP 1 pour générer une suite de vecteurs gaussiens indépendants, $(X_k, Y_k)_{k\geq 1}$, de moyenne $(1, 1)$ et de covariance Σ. Pour $k = 10000$ vérifiez votre simulation en calculant les statistiques empiriques.

Exercice 7. Soit $(X_k)_{k\geq 1}$, $(Y_k)_{k\geq 1}$ des variables aléatoires indépendantes, uniformes sur $[-1, 1]$.

1. Montrer que les vecteurs aléatoires $A_k = (X_k, Y_k)$, $k \geq 1$, sont indépendants et uniformes sur le carré $[-1, 1]^2$.
2. ♣ Vérifier ce résultat graphiquement à l'aide de 1000 tirages indépendants de la suite $(A_k)_{k\geq 1}$.
3. On note $\mathcal{D} = \big\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\big\}$ le disque unité dans $\mathbb{R}^2$. Quelle est la nature de la suite de variables aléatoires $(Z_k)_{k\geq 1}$ définie par $Z_k = \mathbb{1}_{\{A_k \in \mathcal{D}\}}$?
4. ♣ Donner une simulation de $(Z_k)_{1\leq k\leq 500000}$. Approximer $\mathbb{P}(Z_1 = 1)$. Retrouver numériquement l'aire du disque.
5. Pour $r \geq 1$, on définit K_r comme étant le r-ième indice k tel que $A_k \in \mathcal{D}$. Quelle est la loi de K_1 ? Quelle est la loi de A_{K_1} ?
6. ♣ Vérifier ce résultat graphiquement à l'aide de 1000 tirages indépendants de la suite $(A_k)_{k\geq 1}$ (on ne simulera la suite que jusqu'à $k = K_1$).
7. Montrer que les variables aléatoires K_1 et $K_2 - K_1$ sont indépendantes et de même loi. Comment pourrait-on généraliser ce résultat ? Pouvez-vous donner un exemple de la vie de tous les jours qui correspond à cette situation ?
8. ♣ A l'aide de la méthode de Monte Carlo, estimer les probabilités

$$\mathrm{P}(K_2 - K_1 = j \,|\, K_1 = i) \qquad \text{et} \qquad \mathrm{P}(K_2 - K_1 = j)$$

pour les premières valeurs de i et j. (Attention à prendre suffisamment de tirages pour la méthode de Monte Carlo.) Que remarquez-vous ?