

TD 2 : INTÉGRATION MONTE-CARLO

EXERCICE 1 (Échantillonnage préférentiel)

Soit X une variable aléatoire suivant une loi de Pareto de paramètres $(1, 3)$ (dont une densité est $f : x \in \mathbb{R} \mapsto 3x^{-4}\mathbf{1}_{x \geq 1}$). On s'intéresse à l'estimation de $p = \mathbb{P}(X \geq s)$ pour $s > 1$ fixé.

1. Déterminer la valeur de p .
2. Proposer un estimateur Monte-Carlo et déterminer sa loi asymptotique.
3. Proposer un estimateur Monte-Carlo par échantillonnage préférentiel construit sur une loi exponentielle translatée. Est-il satisfaisant ?

EXERCICE 2 (Somme d'exponentielles et conditionnement)

Soient X et Y deux variables aléatoires indépendantes de lois respectives $\text{Exp}(\lambda)$ et $\text{Exp}(\mu)$, pour $\lambda > 0$ et $\mu > 0$. On cherche à estimer $p = \mathbb{P}(S > s)$, où $S = X + Y$ et $s > 0$ est fixé.

1. Proposer un estimateur Monte-Carlo $\hat{p}_n$ de p .
2. Déterminer $\mathbb{P}(S > s | Y = y)$ pour tout $y > 0$. En déduire un estimateur Monte-Carlo par conditionnement $\tilde{p}_n$.

EXERCICE 3 (Estimation d'une p -valeur)

On considère un vecteur gaussien à deux dimensions $X = (U, V) \sim \mathcal{N}(\theta, I_2)$, où $\theta \in \mathbb{R}^2$ et I_2 est la matrice identité de dimension 2. On souhaite tester $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$. Pour ce faire, on considère la statistique de test $S(X) = \max(|U|, |V|)$.

1. On note Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$ et F celle de la variable aléatoire $S(X)$ sous H_0 . Exprimer F en fonction de Φ puis proposer un test $T(X)$ de taille $\alpha \in]0, 1[$ pour décider entre les hypothèses H_0 et H_1 .
2. Soit $x \in \mathbb{R}^2$ une réalisation de X . Déterminer la p -valeur $\alpha_0(x)$ du test T pour la réalisation x .
3. On rappelle que, puisqu'ici le test est de la forme $T(X) = \mathbf{1}_{S(X) > F^{-1}(1-\alpha)}$, avec F continue et strictement croissante sur son support, la p -valeur s'écrit aussi $\alpha_0(x) = \mathbb{P}_{\theta_0}(S(X) \geq S(x))$. En déduire un estimateur Monte-Carlo de la p -valeur.
4. On suppose maintenant que $X \sim \mathcal{N}(\theta, \Sigma)$, où Σ est une matrice symétrique et définie positive, et on souhaite toujours tester $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$, avec la même statistique de test $S(X) = \max(|U|, |V|)$. La formule théorique de la p -valeur n'est plus aussi simple et on ne cherchera pas à l'expliciter. Proposer cependant une manière de l'approcher (on admettra que F est dans ce cas continue et strictement croissante).

EXERCICE 4 ((*) Test du rapport de vraisemblance)

Soient $N \in \mathbb{N}^*$, $\theta \in]0, 1[$ et $X \sim \text{Bin}(N, \theta)$. Pour $\theta_0 \in]0, 1[$, on souhaite tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau $\alpha \in]0, 1[$. Pour ce faire, on considère la statistique

$$S(X) = -2 \log \frac{\theta_0^X (1 - \theta_0)^{N-X}}{\hat{\theta}_N^X (1 - \hat{\theta}_N)^{N-X}},$$

où $\hat{\theta}_N$ est l'estimateur du maximum de vraisemblance (on prend comme d'habitude la convention usuelle $0^0 = 1$), et le test du rapport de vraisemblance : $T(X) = \mathbf{1}_{S(X) > F^{-1}(1-\alpha)}$, où F est la fonction de répartition de $S(X)$ sous H_0 .

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_N$.
2. De manière générale, considérons un modèle statistique paramétrique $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, un échantillon observé $\mathbf{X}$ de loi P_θ (pour un certain $\theta \in \Theta$) et une alternative de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$, pour $\theta_0 \in \Theta$ fixé. Soient une statistique de test $S(\mathbf{X})$, de fonction de répartition F_θ et, pour $\alpha \in]0, 1[$, le test : $T(\mathbf{X}) = \mathbf{1}_{S(\mathbf{X}) > F_{\theta_0}^{-1}(1-\alpha)}$. Montrer que $T(\mathbf{X})$ est un test de taille $1 - F_{\theta_0}(F_{\theta_0}^{-1}(1-\alpha))$ et donc de niveau α .
3. Supposons que la statistique $S(X)$ n'ait qu'un nombre fini d'issues, notées $\{v_1, \dots, v_m\}$. Soit alors $\delta = \min_{1 \leq i, j \leq m} |v_i - v_j|$. En remarquant que $F_{\theta_0}^{-1}(1-\alpha) \in \{v_1, \dots, v_m\}$ et que $S(\mathbf{X}) > F_{\theta_0}^{-1}(1-\alpha) \iff S(\mathbf{X}) - \delta \geq F_{\theta_0}^{-1}(1-\alpha)$, montrer que, pour une réalisation x de $\mathbf{X}$, la p -valeur $\alpha_0(x)$ vaut :

$$\alpha_0(x) = \mathbb{P}_{\theta_0}(S(X) \geq S(x)).$$

4. Revenons au cadre particulier qui nous intéresse et considérons une réalisation $x \in \{0, \dots, N\}$ de X . Proposer un estimateur Monte-Carlo visant à approcher la p -valeur $\alpha_0(x)$.
5. Soit la fonction

$$\varphi : x \in]0, 1[\mapsto -2 \log \frac{\theta_0^x (1 - \theta_0)^{1-x}}{x^x (1-x)^{1-x}},$$

de sorte que $S(X) = N\varphi(\hat{\theta}_N)$. Montrer qu'un développement limité de φ en θ_0 donne :

$$\forall x \in]0, 1[, \quad \varphi(x) = \frac{1}{\theta_0(1-\theta_0)}(x - \theta_0)^2 + (x - \theta_0)^2 r(x),$$

où $r :]0, 1[\rightarrow \mathbb{R}$ est une fonction telle que $\lim_{x \rightarrow \theta_0} r(x) = 0$.

6. En remarquant que l'on peut prolonger r par continuité de sorte que $r(\theta_0) = 0$, conclure que, sous H_0 , $S(X) \xrightarrow[N \rightarrow \infty]{\text{Loi}} \chi_1^2$.
7. Donner une autre expression de la p -valeur $\alpha_0(x)$ et en déduire une manière de l'approcher lorsque N est grand.

EXERCICE 5 (Variables antithétiques)

Nous ne présentons pas cette méthode dans le cadre le plus général, mais plutôt par l'intermédiaire d'une application classique. On souhaite estimer $I = \mathbb{E}[\varphi(X)]$ avec X une variable aléatoire réelle telle que $\mathbb{E}[\varphi(X)^2] < \infty$, F sa fonction de répartition et on suppose disposer d'une formule pour la fonction quantile F^{-1} . Dans ce cas, on sait qu'une façon de simuler X est de considérer U uniforme sur $[0, 1]$ et de poser $X := F^{-1}(U)$. L'estimateur Monte-Carlo standard s'écrit alors

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \varphi(X_i) = \frac{1}{n} \sum_{i=1}^n \varphi(F^{-1}(U_i)),$$

où (U_n) est une suite i.i.d de loi $\text{Unif}([0, 1])$ et pour tout $n \geq 1$, $X_n = F^{-1}(U_n)$. L'estimateur de I par variables antithétiques est défini par

$$\tilde{I}_n := \frac{1}{n} \sum_{i=1}^n \frac{\varphi(F^{-1}(U_i)) + \varphi(F^{-1}(1 - U_i))}{2}.$$

1. Montrer que $\tilde{I}_n$ est sans biais, fortement consistant et asymptotiquement normal, c'est-à-dire

$$\sqrt{n} \left(\tilde{I}_n - I \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, s^2),$$

avec

$$s^2 := \frac{1}{2} [\text{Var}(\varphi(F^{-1}(U))) + \text{Cov}(\varphi(F^{-1}(U)), \varphi(F^{-1}(1-U)))] .$$

2. Montrer que $s^2 \leq \sigma^2 := \text{Var}(\varphi(X))$.
3. Soit $\alpha \in]0, 1[$. Déterminer un estimateur fortement consistant $\tilde{s}_n^2$ de s^2 et en déduire un intervalle de confiance de niveau asymptotique $(1 - \alpha)$ pour I .
4. Application. Considérons l'exemple jouet suivant : soit $q > \log 2$ fixé connu, $X \sim \text{Exp}(1)$ et $p := \mathbb{P}(X > q) = e^{-q}$. Montrer que $s^2 = (\sigma^2 - p^2)/2$ où σ^2 est la variance asymptotique du Monte-Carlo standard et s^2 celle obtenue par variables antithétiques.

EXERCICE 6 ((*) Preuve par couplage de l'inégalité de covariance de Tchebychev)

Soit X une variable aléatoire, φ une fonction croissante et ψ une fonction décroissante telles que $\varphi(X)$ et $\psi(X)$ soient de carré intégrable.

1. Soit X' une variable aléatoire de même loi que X et indépendante de celle-ci. Montrer que

$$\text{Cov}(\varphi(X) - \varphi(X'), \psi(X) - \psi(X')) = \mathbb{E}[(\varphi(X) - \varphi(X'))(\psi(X) - \psi(X'))] \leq 0.$$

2. Montrer par ailleurs que : $\text{Cov}(\varphi(X) - \varphi(X'), \psi(X) - \psi(X')) = 2 \text{Cov}(\varphi(X), \psi(X))$, et en déduire l'inégalité de covariance de Tchebychev, à savoir que : $\text{Cov}(\varphi(X), \psi(X)) \leq 0$.
3. En déduire le principe des variables antithétiques en Monte-Carlo, à savoir que si φ est monotone et h décroissante telle que $h(X)$ ait même loi que X alors, sous réserve d'existence des variances,

$$\text{Var} \left(\frac{\varphi(X) + \varphi(h(X))}{2} \right) \leq \frac{1}{2} \text{Var}(\varphi(X)).$$

4. Exemple jouet : proposer une méthode de variables antithétiques pour estimer $\mathbb{E}[\exp X]$ lorsque $X \sim \mathcal{N}(0, 1)$.

EXERCICE 7 ((†) Marche aléatoire avec dérive et échantillonnage préférentiel)

Soit $(X_n)_{n \geq 1}$ une suite de variables gaussiennes indépendantes $\mathcal{N}(-m, 1)$ avec $m > 0$, et $S_n = X_1 + \dots + X_n$. Les réels $a < 0$ et $b > 0$ étant fixés, on considère le temps d'arrêt $N = \min\{n : S_n < a \text{ ou } S_n > b\}$ et on s'intéresse à la probabilité $p = \mathbb{P}(S_N > b)$ que la marche aléatoire (S_n) sorte de l'intervalle $[a, b]$ par le haut.

1. Montrer que le temps d'arrêt N est presque sûrement fini.
2. Proposer un estimateur $\hat{p}_k$ de p basé sur k simulations successives de trajectoires. Intuitivement, pourquoi cet estimateur n'est-il pas satisfaisant ?
3. Plutôt que de simuler X suivant la densité d'une loi $\mathcal{N}(-m, 1)$, on simule X' suivant la loi $\mathcal{N}(m, 1)$. On note (S'_n) la marche aléatoire obtenue et N' le temps de sortie de l'intervalle $[a, b]$. Notons le rapport de vraisemblance $w(s) := f(s)/g(s)$, où f et g sont les densités respectives des variables aléatoires S_N et $S'_{N'}$. Le but est de montrer que $w(s) = \exp(-2ms)$.

(a) Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction test, montrer que

$$\mathbb{E}[\varphi(S_N)] = \sum_{n=1}^{\infty} \mathbb{E}[\varphi(S_n) \mathbf{1}_{N=n}].$$

(b) Notant $I = [a, b]$ et I^c son complémentaire, montrer que

$$\mathbb{E}[\varphi(S_n) \mathbf{1}_{N=n}] = \int_{\mathbb{R}^n} \varphi(s_n) \mathbf{1}_{I^{n-1}}(s_1, \dots, s_{n-1}) \mathbf{1}_{I^c}(s_n) f(s_1, \dots, s_n) ds_1 \dots ds_n,$$

où, avec la convention $s_0 = 0$,

$$f(s_1, \dots, s_n) = \exp(-2ms_n) \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{((s_i - s_{i-1}) - m)^2}{2}\right\}.$$

(c) En déduire que

$$\mathbb{E}[\varphi(S_n) \mathbf{1}_{N=n}] = \mathbb{E}[\varphi(S'_n) \exp(-2mS'_n) \mathbf{1}_{N'=n}],$$

et conclure.

4. Déduire de la question précédente un nouvel estimateur $\tilde{p}_k$ de p basé sur k simulations successives.

5. Déduire de la forme de $\tilde{p}_k$ que $p \leq \exp(-2mb)$.