# Travaux Pratiques

stoehr@ceremade.dauphine.fr

## 1 Simulations de variables aléatoires

**Rappel R.** `runif` permet de simuler des réalisations *i.i.d.* de loi $\mathscr{U}([0,1])$. R propose des générateurs aléatoires pour la plupart des lois usuelles. Toutefois, ils ne seront pas utilisés dans cette partie sauf à titre de comparaison.

### 1.1 Fonction inverse et transformations

**Exercice 1** *(Simulation d'une variable aléatoire discrète).* Soit $X$ une variable aléatoire discrète sur l'ensemble $\{5,6,7,8\}$. On définit la loi $\nu$ de $X$ par

$$\mathbb{P}[X=5]=0.4,\quad \mathbb{P}[X=6]=0.2,\quad \mathbb{P}[X=7]=0.3,\quad \mathbb{P}[X=8]=0.1.$$

**1.** Simuler avec la méthode de la fonction inverse un échantillon **x** de 10000 variables aléatoires *i.i.d.* suivant la loi $\nu$.

**2.** Comparer le diagramme en bâtons de l'échantillon **x** à celui de $\nu$.

**Rappel R.** `barplot(height = ...)` trace le diagramme en bâtons de variables catégorielles ou discrètes. `height` est un vecteur contenant la hauteur des barres ou la table de contingence de l'échantillon. Pour un échantillon **x**, la table de contingence s'obtient avec `table(x)`.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Question** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. | 0 | 0 | 0 | 0 | 0 | 0 |

**Exercice 2** *(Loi exponentielle et lois connexes).* Soient $X_1,\ldots,X_n$, variables aléatoires *i.i.d.* de loi exponentielle $\mathscr{E}(\lambda)$, *i.e.* $\mathbb{E}[X_1]=\lambda^{-1}$.

**1.** **(a)** Simuler avec la méthode de la fonction inverse 10000 réalisations de la loi $\mathscr{E}(\lambda)$ pour $\lambda=2$.

**(b)** Vérifier à l'aide d'un histogramme et d'un diagramme Quantile-Quantile que la distribution de cet échantillon est en adéquation avec loi $\mathscr{E}(\lambda)$.

**2.** On rappelle que $S_n = X_1+\ldots+X_n$ suit la loi gamma $\Gamma(n,\lambda)$, *i.e.* $\mathbb{E}[S_n]=n\lambda^{-1}$.

**(a)** Partant de ce résultat, simuler 10000 réalisations de la loi gamma $\Gamma(n,\lambda)$, avec $\lambda=1.5$ et $n=10$.

**(b)** Vérifier graphiquement que la distribution de cet échantillon est en adéquation avec loi $\Gamma(n,\lambda)$.

**3.** Soit $N = \sup\{n \geq 1 : S_n \leq 1\}$ (par convention $N = 0$ si $S_1 > 1$). Alors $N$ suit la loi de Poisson $\mathscr{P}(\lambda)$.

**(a)** Partant de ce résultat, simuler 10000 réalisations de la loi de Poisson $\mathscr{P}(\lambda)$, avec $\lambda = 4$.

**(b)** Vérifier graphiquement que la distribution de cet échantillon est en adéquation avec loi $\mathscr{P}(\lambda)$.

**Rappel R.**

- `hist(x, freq = F)` affiche l'histogramme d'un échantillon `x`. L'option `freq` spécifie si l'histogramme est représenté en densité d'effectifs (`freq = TRUE` par défaut) ou en densité de probabilités (`freq = FALSE`).
- `lines(x,y)` permet d'ajouter une courbe affine par morceaux reliant les points d'abscisse `x` et d'ordonnée `y`.
- `quantile(x, probs)` retourne les quantiles d'un échantillon `x` pour un vecteur de probabilités `probs`.
- `dexp`, `pexp` et `qexp` correspondent respectivement à la densité, à la fonction de répartion et à la fonction quantile d'une loi exponentielle.
- `dgamma`, `pgamma` et `qgamma` correspondent respectivement à la densité, à la fonction de répartion et à la fonction quantile d'une loi gamma.
- `dpois`, `ppois` et `qpois` correspondent respectivement à la densité, à la fonction de répartion et à la fonction quantile d'une loi de Poisson.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Questions** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. | 0 | 0 | 0 | 0 | 0 | 0 |
| 2. | 0 | $\leq 1$ | 0 | 0 | 0 | 0 |
| 3. | 0 | $\leq 1$ | 1 | 0 | 0 | 0 |

## 1.2 Loi normale, vecteurs gaussiens et mouvement brownien

**Exercice 3** *(Algorithme de Box-Muller)*.

**1.** Écrire une fonction `BM(n)` qui retourne $n$ réalisations de la loi normale $\mathcal{N}(0,1)$ par la version cartésienne de la méthode de Box-Muller.

**2.** Valider l'algorithme à l'aide d'un outil graphique.

**Rappel R.** `dnorm`, `pnorm` et `qnorm` correspondent respectivement à la densité, à la fonction de répartion et à la fonction quantile d'une loi normale.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Question** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. | 0 | 0 | 0 | $\leq 1$ | 0 | 0 |

**Exercice 4** *(Simulation de vecteurs gaussiens).* Soit $\mathbf{X} = (X_1, X_2)$ de loi $\mathcal{N}(\mu, \Sigma)$, avec

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{et} \quad \Sigma = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$$

**1.** Simuler une suite de vecteurs $\left(X^{(n)}\right)_{n\geq 1} = \left(X_{1,n}, X_{2,n}\right)_{n\geq 1}$ qui suivent la loi de $\mathbf{X}$.

**2.** Quelle est la loi de $X_1 + X_2$ ? Valider ce résultat graphiquement.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Question** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. | 0 | 0 | 0 | 0 | 0 | 0 |

**Exercice 5** *(Simulation du brownien).* En utilisant les propriétés des accroissements du brownien, simuler une réalisation du brownien aux instants $(t_1, \dots, t_{1110})$ définis par $t_i = i/100$ pour $i \in [\![1, 100]\!]$, $t_i = 1 + (i - 100)/10$ pour $i \in [\![101, 110]\!]$ et $t_i = 2 + (i - 110)/1000$ pour $i \in [\![111, 1110]\!]$.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |

### 1.3 Algorithme du rejet

**Exercice 6** *(Rejet – un premier exemple).* Soit $f$ une fonction de densité définie pour tout réel $x$ par

$$f(x) = \frac{2}{\pi}\sqrt{1 - x^2}\,\mathbb{1}_{\{x \in [-1,1]\}}.$$

**1.** Utiliser la méthode du rejet pour simuler 10000 réalisations suivant la loi de densité $f$.

**2.** Tracer l'histogramme de cet échantillon et le comparer à la densité $f$.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Version** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| ⋆ \| ⋆⋆⋆ | 0 \| 1 | 1 \| 0 | 1 | 0 | 0 | 0 |

**Exercice 7.** Utiliser la méthode du rejet pour simuler 5000 réalisations suivant la loi uniforme sur le disque unité $\mathcal{D} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Version** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| ⋆ \| ⋆⋆⋆ | 0 \| 1 | 1 \| 0 | 1 | 0 | 0 | 0 |

**Exercice 8** *(Loi normale tronquée).* La loi normale $\mathcal{N}(\mu,\sigma^2)$, $\mu>0$, $\sigma>0$, tronquée de support $[b,+\infty[$ admet une densité $f$ définie pour tout réel $x$ par

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}\Phi\left(\frac{\mu-b}{\sigma}\right)}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\mathbb{1}_{\{x\geq b\}},$$

où $\Phi$ est la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$.

**Densité instrumentale n°1 : $\mathcal{N}(\mu,\sigma^2)$.**

1. Écrire une fonction `tr_norm(n, b, mean, sd)` permettant de simuler $n$ réalisations suivant la normale $\mathcal{N}(\texttt{mean},\texttt{sd}^2)$ tronquée de support $[b,+\infty[$.
2. Simuler 10000 réalisations suivant la loi normale $\mathcal{N}(0,2)$ tronquée de support $[2,+\infty[$. Valider votre algorithme graphiquement.

**Densité instrumentale n°2.** La loi exponentielle translatée de $b$, $\tau\mathcal{E}(\lambda,b)$, a pour densité

$$g_\lambda(x)=\lambda e^{-\lambda(x-b)}\mathbb{1}_{\{x\geq b\}},\quad x\in\mathbb{R}.$$

Dans la suite on fixe $\lambda$ à la valeur optimale obtenue dans le TD n°2.

3. Écrire une fonction `tr_norm_2(n, b, mean, sd)` permettant de simuler $n$ réalisations suivant la normale $\mathcal{N}(\texttt{mean},\texttt{sd}^2)$ tronquée de support $[b,+\infty[$.
4. Simuler 10000 réalisations suivant la loi normale $\mathcal{N}(0,2)$ tronquée de support $[2,+\infty[$. Valider votre algorithme graphiquement.
5. Comparer les performances de `tr_norm` et `tr_norm_2`.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Version** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. ⋆ \| ⋆⋆⋆ | 0 \| 1 | 1 \| 0 | 1 | 0 | 0 | 0 |
| 4. ⋆ \| ⋆⋆⋆ | 0 \| 1 | 1 \| 0 | 1 | 0 | 0 | 0 |

## 2 Méthode de Monte Carlo classique

**Exercice 9** *(Estimation de $\pi$).*

1. En utilisant la méthode de Monte Carlo classique, avec $n=150000$ tirages, proposer une estimation de $\pi$ basée sur la génération de
   **(a)** $U_1,\ldots,U_n$ variables aléatoires *i.i.d.* de loi uniforme $\mathcal{U}([0,1])$ ;
   **(b)** $(U_{1,1},U_{2,1}),\ldots,(U_{1,n},U_{2,n})$ couples de variables aléatoires *i.i.d.* suivant la loi $\mathcal{U}([0,1])\otimes\mathcal{U}([0,1])$.
2. Quel est l'estimateur le plus performant ?
3. Avec l'inégalité de Bienaymé-Tchebichev ou l'hypothèse du régime asymptotique au niveau de confiance 95%, pour quelle valeur de $n$ a-t-on une précision de $10^{-2}$ ? Discuter les résultats obtenus.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| Questions | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. | 0 | 0 | 0 | 0 | 0 | 0 |
| 3. | 0 | 0 | 0 | 0 | 0 | 0 |

**Exercice 10** *(Calcul intégral).* On s'intéresse au calcul de l'intégrale suivante

$$\delta = \int_2^{+\infty} \int_0^5 \sqrt{x+y}\sin\left(y^4\right)\exp\left(-\frac{3x}{2} - \frac{y^2}{4}\right)\mathrm{d}x\mathrm{d}y.$$

**1.** Proposer une estimation de $\delta$ par la méthode de Monte Carlo classique en utilisant :

**(a)** un générateur de la loi uniforme et un générateur de la loi normale ;

**(b)** un générateur de la loi exponentielle et un générateur de la loi normale ;

**(c)** un générateur de la loi exponentielle et un générateur de la loi normale tronquée.

**2.** Déterminer le coût de calcul de ces méthodes. En déduire celle qu'il est préférable d'utiliser.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| Questions | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. ⋆ \| ⋆⋆⋆ | 0 \| 1 | 1 \| 0 | 1 | 0 | 0 | 0 |

# 3 Méthodes de réduction de variance

## 3.1 Échantillonnage préférentiel

**Exercice 11** *(Loi de Cauchy).* Soit $X$ une variable aléatoire de loi de Cauchy $\mathscr{C}(0,1)$. On s'intéresse à l'estimation de

$$p := \mathbb{P}[X \geq 50] = \int_{50}^{+\infty} \frac{1}{\pi(1+x^2)}\mathrm{d}x.$$

On se propose de comparer la méthode de Monte Carlo classique et la méthode d'échantillonnage préférentiel basée sur la loi de Pareto dont la fonction de répartition est donnée pour $x \in \mathbb{R}$ par

$$F(x) = \left[1 - \left(\frac{a}{x}\right)^k\right]\mathbb{1}_{\{x \geq a\}}, \quad \text{avec} \quad a > 0 \quad \text{et} \quad k > 0.$$

On considèrera $n = 10000$ tirages.

**1.** Donner une estimation de $p$ à partir de simulations suivant la loi de Cauchy $\mathscr{C}(0,1)$.

**2.** **(a)** Quelles valeurs de $a$ et $k$ doit-on choisir pour la méthode d'échantillonage préférentiel ?

**(b)** En déduire une estimation de $p$ par la méthode d'échantillonnage préférentiel.

**3.** Déterminer l'efficatité relative entre ces deux méthodes.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Questions** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 2. | 0 | 0 | 0 | 0 | 0 | 0 |
| 3. | 0 | 0 | 0 | 0 | 0 | 0 |

**Exercice 12** *(Rejet v.s. échantillonnage préférentiel).* On revient sur le calcul de l'intégrale

$$\delta = \int_2^{+\infty}\int_0^5 \sqrt{x+y}\sin\left(y^4\right)\exp\left(-\frac{3x}{2}-\frac{y^2}{4}\right)\mathrm{d}x\mathrm{d}y.$$

**1.** En écrivant $\delta$ comme une espérance par rapport à la loi exponentielle et à la loi exponentielle translatée, proposer une estimation de $\delta$ par la méthode d'échantillonnage préférentiel.

**2.** Comparer les performances de cette méthode avec la méthode de Monte Carlo classique basée sur un générateur de la loi exponentielle et un générateur de la loi normale tronquée.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Questions** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 2. $\star$ \| $\star\star\star$ | 0 \| 1 | 1 \| 0 | 1 | 0 | 0 | 0 |

### 3.2 Autres méthodes de réductions

**Exercice 13** *(Intégrale de Gauss et réduction de variance).* On souhaite estimer

$$\mathscr{I} = \int_0^2 e^{-x^2}\mathrm{d}x$$

à l'aide des méthodes de Monte-Carlo. Dans cette exercice, on considèrera $n = 10000$ tirages.

**1.** Donner une estimation de $\mathscr{I}$ à l'aide de la méthode de Monte-Carlo classique basée sur deux lois différentes.

**2.** Pour ces estimateurs, proposer une nouvelle estimation de $\mathscr{I}$ basée sur une variable antithétique.

**3.** **(a)** Calculer le moment d'ordre $k \in \mathbb{N}^*$ d'une variable aléatoire uniforme sur $[0,2]$.

**(b)** On utilise le développement en série entière de $x \mapsto \exp(-x^2)$ tronqué à l'ordre $k \in \mathbb{N}^*$ comme variable de contrôle. Proposer une méthode pour trouver la valeur de $k$ à utiliser en pratique.

**(c)** En déduire une estimation de $\mathscr{I}$ par la méthode de la variable de contrôle.

**4.** On considère un estimateur basé sur le générateur de la loi $\mathscr{U}([0,2])$.

**(a)** Choisir une variable de stratification $Z$ et une partition $D_1,\dots,D_K$ telles que $\mathbb{P}[Z \in D_k] = 1/K$.

**(b)** Donner une estimation de $\mathscr{I}$ par la méthode de stratification avec allocation proportionnelle et $K = 10$. Étudier l'influence de $K$ sur la précision de la méthode.

**5.** Déterminer l'efficacité relative de ces différentes méthodes.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Questions** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 3.b. | 0 | 2 | 0 | 0 | 0 | 0 |
| 3.c. | 0 | 1 | 0 | 0 | 0 | 0 |
| 4.b. ($\star$) | 0 | 1 | 0 | 0 | 0 | 0 |
| 4.b. ($\star\star\star$) | 0 | 0 | 0 | 0 | 0 | 1 |
| 4.c. | 0 | 1 | 0 | 0 | 0 | 0 |

**Exercice 14** *(« Mouvement brownien » et finance).* Soit $\mathbf{X} = (X_1, X_2)$ un vecteur gaussien de loi $\mathcal{N}(0, \Sigma)$, avec

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

On souhaite estimer

$$\mathscr{I} = \mathbb{E}\left[\max\left\{0, \frac{1}{2}e^{-\sigma^2/2+\sigma X_1} + \frac{1}{2}e^{-\sigma^2/2+\sigma X_2} - 1\right\}\right].$$

Pour les applications numériques, on prendra $\sigma = 2$, $n = 5000$ tirages et on fournira un intervalle de confiance asymptotique au niveau 95%.

**1.** Donner une estimation de $\mathscr{I}$ par la méthode Monte-Carlo classique.

On s'intéresse dans un premier temps à la méthode de la variable antithétique. Soient

$$\mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \quad \text{et} \quad A = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

**2.** **(a)** Montrer que le vecteur $A\mathbf{Z}$ suit la même loi que $\mathbf{Z}$

**(b)** Donner une estimation de $\mathscr{I}$ par la méthode de la variable antithétique.

**(c)** La méthode de la variable antithétique permet-elle de réduire la variance ?

On s'intéresse enfin à la méthode de la variable de contrôle.

**3.** **(a)** Calculer $\mathbb{E}\left[\exp(\sigma X_1) + \exp(\sigma X_2)\right]$.

**(b)** En déduire une estimation de $\mathscr{I}$ par la méthode de la variable de contrôle.

**4.** Déterminer l'efficacité relative de ces trois méthodes.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Questions** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. | 0 | 0 | 0 | 0 | 0 | 0 |
| 2. | 0 | 0 | 0 | 0 | 0 | 0 |
| 3. | 0 | 1 | 0 | 0 | 0 | 0 |

**Exercice 15**. Le nombre de précipitation $S$ sur un mois est modélisé par une loi de Poisson de paramètre $\lambda = 3.7$. La quantité d'eau $Q_s$ tombant lors d'une précipitation $s$ est modélisé par un loi de Weibul de paramètre de forme $k = 0.5$ et de paramètre d'échelle $\lambda = 2$ (on supposera que les précipitations sont indépendantes). La quantité de pluie tombant en 1 mois est donc

$$X = \begin{cases} 0 & \text{, si } S = 0, \\ \sum_{s=1}^{S} Q_s & \text{, sinon.} \end{cases}$$

On s'intéresse aux mois présentant de faibles précipitations et on cherche à estimer $p = \mathbb{P}[X < 3]$ (*i.e.*, il y a moins de 3 cm de pluie par mois).

**1.** Donner une estimation de $p$ par la méthode de Monte Carlo classique. On donnera l'intervalle de confiance au niveau 95%.

**2.** Donner une estimation de $p$ par la méthode de stratification avec allocation proportionnelle, en précisant les strates utilisées. On donnera l'intervalle de confiance au niveau 95%.

**3.** Proposer une méthode d'estimation de $p$ par la méthode de stratification avec allocation optimale. Quelles difficultés rencontrez-vous ?

**4.** Comparer les variances et l'efficacité relative de ces trois méthodes d'estimations. Discuter de façon concise les résultats obtenus.

**Auto-évaluation.** Nombre de fois où les éléments suivants interviennent dans la solution.

| **Questions** | `c()` | `for` | `while` | `if-else` | `Vectorize` | `apply` |
|---|---|---|---|---|---|---|
| 1. ⋆ \| ⋆⋆⋆ | 0 | 1 \| 0 | 0 | 0 | 0 | 0 \| 1 |
| 2. ⋆ \| ⋆⋆⋆ | 0 | 5 \| 2 | 0 | 0 | 0 | 0 \| 4 |