Université Paris-Dauphine
Année 2011-2012
Département de Mathématique

Exercices d'entrainement

(Warning : Les exercices tirés au sort pour le partiel de Novembre dans cette liste n'auront pas les numéros utilisés dans cette série.)

Exercise 1
On considère la v.a. X de densité

$$f_X(x) = C\left[\left(x-\frac{1}{2}\right)^2 - x + \frac{3}{4}\right]\mathbb{I}_{[0,1]}(x)\,.$$

1. Calculer la constante de normalisation C ainsi que l'espérance et la variance de $X \sim f_X$.
2. Ecrire une fonction simulant n réalisations de $X \sim f_X$ par inversion générique.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$.
4. Donner une estimation de l'ésperance et de la variance de X ainsi que l'intervalle de confiance de niveau $1-\alpha$ avec $\alpha = 0.05$.
5. Donner une estimation numérique de C.

Exercise 2
Considérons une variable aléatoire X de densité

$$f(x) = 3x^2\exp(-x^3),\ x \in \mathbb{R}_+.$$

a. Donner la fonction de répartition F de X.

b. Simuler $n = 10^4$ variables indépendantes de F (iid de F).

c. Calculer les quantiles $q_1 = F^{-1}(0.025)$ et $q_2 = F^{-1}(0.975)$.

d. Calculer les valeurs d'estimateurs $\hat{q}_{n,1}$ de q_1 et $\hat{q}_{n,2}$ de q_2 à partir de l'échantillon de taille n simulé à la question b. Donner une évaluation de la précision de ces estimations.

Exercise 3

a. Soient X_1 et X_2 deux variables indépendantes $X_1, X_2 \sim Exp(\lambda)$. Quelle est la densité de la distribution de la transformation $Z = X_1/X_2$?

b. Simuler $n = 10^4$ réalisations de cette loi. Tracer l'histogramme associé et vérifier l'adéquation à la densité trouvée à la question a.

c. Quels sont les deux premiers moments de cette loi ? *(Attention piège !)* Etudier le problème à partir des simulations de la question b.

Exercise 4
Soit $X_1, \ldots, X_n$ un n-échantillon distribué sous la loi caractérisée par la fonction de répartition F (et de densité correspondante f). Soit $\alpha \in]0,1[$ et $\hat{q}_n$ est l'estimateur empirique du quantile $F^{-1}(\alpha)$

On voudrait vérifier le théorème limite suivant :

$$\sqrt{n}(\hat{q}_n - F^{-1}(\alpha)) \to_{\mathcal{D}} \mathcal{N}(0, \sigma_q^2) \tag{1}$$

où la variance de la loi normale asymptotique est égale à

$$\sigma_q^2 = \frac{\alpha(1-\alpha)}{[f(F^{-1}(\alpha))]^2}.$$

Dans la suite du problème on supposera que F est la fonction de répartition de $\mathcal{N}(0, 100)$, c'est-à-dire une normale de moyenne 0 et d'écart type 10, et donc de densité

$$f(x) = \frac{1}{10\sqrt{2\pi}} \exp(-\frac{x^2}{200}).$$

On prendra $\alpha = 0.18$.

1. Simuler $N = 10^3$ fois des échantillons de taille $n = 10^2$ de la loi $\mathcal{N}(0, 100)$, et calculer $\hat{q}_n$ pour chacun de ces échantillons. On pourra écrire une fonction R prenant comme argument (N, n) et fournissant les $\hat{q}_n$.
2. Tracer l'histogramme des $\hat{q}_n$ ainsi obtenus et tester l'adéquation à la loi asymptotique donnée en (1).
3. Donner un intervalle de confiance à 95% sur l'espérance de $\hat{q}_n$.

Exercise 5
On cherche à approcher l'intégrale

$$\int_{\mathbf{R}} \frac{3 + \sin^2(x)}{(1 + x^2)^2} \mathrm{d}x$$

à l'aide de la méthode de Monte-Carlo.

a. Donner une méthode d'acceptation rejet permettant de simuler une variable aléatoire dont la densité est proportionnelle à $\frac{3+\sin^2(x)}{(1+x^2)^2}$.

b. Illustrer à l'aide d'un histogramme sur 10^4 simulations la pertinence de votre algorithme de génération.

c. Déduire du taux d'acceptation de l'algorithme d'acceptation rejet la valeur de l'intégrale, en précisant l'erreur statistique de cette approximation (pour une confiance de 95%).

Exercise 6
Soit l'intégrale

$$I = \int_0^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, \mathrm{d}y$$

a. Donner la valeur exacte de I (à 10^{-10} près) par une commande R.

b. Proposer une méthode de Monte-Carlo d'évaluation de I reposant sur la génération d'un n-échantillon de variables aléatoires gaussiennes. Fournir un intervalle de confiance à 95% sur I pour $n = 10^4$.

c. Proposer une autre méthode de Monte-Carlo reposant sur un n-échantillon de variables aléatoires de loi uniforme. Fournir un intervalle de confiance à 95% sur I pour $n = 10^4$. Parmi ces deux méthodes, laquelle est la meilleure ?

Exercise 7
Soit une densité de probabilité f sur $\mathbb{R}$ telle que

$$f(x) \propto 0.7\exp(-x^2/2) + 0.3\exp(-x^2/2 + 7(x - 7/2)), x \in \mathbb{R}.$$

C est la constante de proportionalité définie par l'identité

$$C\int_{\mathbb{R}} \left\{0.7\exp(-x^2/2) + 0.3\exp(-x^2/2 + 7(x - 7/2))\right\} \mathrm{d}x = 1\,.$$

On génére suivant f par acceptation-rejet en partant d'une densité g, avec $f/g \leq MC$.

a. Entourer les couples (g, M) valides. (On pourra utiliser la fonction `optimise`.)

1. $\mathcal{E}xp(3)$ et $M = 345$
2. $N(2.1, 3^2)$ et $M = 7$
3. $Student_3$ et $M = 300$
4. $U_{[-10,10]}$ et $M = 24$
5. $N(2.1, 3^2)$ et $M = 11$
6. Cauchy et $M = 40$

b. Entourer le couple (g, M) le plus efficace en termes de nombres de rejets.

1. $N(2.1, 3^2)$ et $M = 11$
2. Cauchy et $M = 40$
3. $\mathcal{E}xp(3)$ et $M = 345$
4. $U_{[-10,10]}$ et $M = 24$
5. $N(2.1, 3^2)$ et $M = 7$
6. $Student_3$ et $M = 300$

On choisit finalement d'utiliser pour g une loi de Cauchy et $M = 50$.

c. Simuler un échantillon de départ de taille $n = 10000$ réalisations de $Y \sim g$ en utilisant le couple (g, M) trouvé dans la question précédent. Parmi les propositions suivantes, entourer le taux d'acceptation le plus proche du résultat obtenu.

1. 0.05
2. 0.89
3. 0.01
4. 0.34
5. 0.82

d. Déduire la valeur de la constante de normalisation C du taux d'acceptation obtenu précédemment.

Exercise 8
Etant donnée la densité $f(x) \propto \exp\{-x^2\sqrt{x}\}[\sin(x)]^2$, $0 < x < \infty$, de la variable aléatoire X, on considère les lois de densité

$$g_1(x) = \frac{1}{2}e^{-|x|}, \quad g_2(x) = \frac{1}{2\pi}\frac{1}{1 + x^2/4}, \quad g_3(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,.$$

1. Pour des échantillons produits suivant chacune de ces lois, estimer par une expérience de Monte-Carlo le nombre M de simulations nécessaire pour obtenir une précision de trois décimales sur $\mathbb{E}_f[X]$. Fournir le code R utilisé.
2. En utilisant g_1 comme proposition, générer un échantillon suivant f par acceptation-rejet et en déduire une estimation de la constante de normalisation de f. Fournir le code R utilisé.
3. Comparer aux estimations obtenues en utilisant g_2 et g_3. Fournir le code R utilisé.

Exercise 9

On considère la densité

$$f_X(x;\theta) \;=\; \theta\, x^{\theta-1}\mathbb{I}_{[0,1]}(x) \;\;,\;\; \theta > 0\,.$$

1. Donner l'expression de l'espérance et de la variance de $X \sim f_X$ en fonction de θ.
2. En utilisant la méthode d'inversion génerique, écrire une fonction qui permette de simuler un n-échantillon $(X_1, X_2, \ldots, X_n)$ de réalisations de variables aleatoires de densité f_X.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$ et $\theta = 3$.
4. Montrer graphiquement la convergence des estimateurs Monte-Carlo de l'espérance et de la variance de f_X vers leurs valeurs théoriques dans le cas $\theta = 0.9$. $\theta = 3$. Tracer les courbes relatives à cette étude.

Exercise 10

Afin d'étudier le propagation des noms de familles d'une génération à l'autre dans l'Angleterre Victorienne, Sir Galton a proposé en 1873 le modèle stochastique suivant. Chaque adulte mâle transmet son nom à ses fils. Supposons que le nombre de fils de chaque homme soit une variable aléatoire entière (de même loi pour tous les hommes). Alors on peut montrer que

- si le nombre moyen de fils par homme (noté m) est inférieur ou égal à 1, le nom de famille va disparaître presque surement.
- Si $m > 1$ la probabilité de survie du nom est non-nulle et le nombre de porteurs du nom connait une croissance exponentielle.

Notons X_n le nombre de porteurs du nom de famille à la génération n. Soit $Y_{n,k}$ le nombre de fils de l'individu k de la n-ième génération. Alors

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$$

avec $X_1 = 1$.

On suppose que $Y_{n,k}$ est une variable aléatoire dans $\{0, 1, 2\}$ telle que

$$P(Y_{n,k} = 0) = p_0, \quad P(Y_{n,k} = 1) = p_1 \quad P(Y_{n,k} = 2) = 1 - (p_0 + p_1)$$

1. Calculer m le nombre moyen de fils par homme en fonction de (p_0, p_1). Calculer la variance σ^2 de $Y_{n,k}$ en fonction de (p_0, p_1).

2. Ecrire une fonction galton prenant en argument n_{gene} le nombre de générations et le vecteur des probabilités (p_0, p_1) et donnant en sortie une réalisation du processus $(X_n)_{n=1\ldots n_{gene}}$ décrit précédemment.
3. On pose $(p_0, p_1, p_2) = (0.2, 0.6, 0.2)$ et $n_{gene} = 100$
   (a) Que vaut m ?
   (b) Représenter sur une même figure 5 réalisations simulées avec la fonction galton.
   (c) Proposer une méthode de Monte-Carlo permettant de calculer la probabilité d'extinction du nom de famille

Exercise 11
On sait que si i_1, i_2, i_3, i_4 sont quatre entiers tirés aléatoirement entre 1 et n, la probabilité que ces quatre nombres soient premiers entre eux vaut approximativement $90/\pi^4$ (cette probabilité tend vers $90/\pi^4$ quand n tend vers l'infini).
Utiliser cette propriété pour calculer en R une approximation de π. Observer l'influence de la valeur de n.

On pourra utiliser les instructions

```
install.packages("schoolmath")
require(schoolmath)
```

pour installer le package `schoolmath`, puis utiliser la fonction `gcd(a,b)` pour calculer le plus grand diviseur commun des nombres a et b.

Exercise 12
Lorsque f correspond à un mélange de lois normales

$$p\,\mathcal{N}(0,1) + (1-p)\mathcal{N}(3,1),$$

on peut simuler un échantillon Z de distribution f de taille `nech` de la manière suivante :

```
p=0.6
nech=1000
Z=replicate(nech, rnorm(1, sample(c(0,3),prob=c(p,1-p)),1))
```

On s'intéresse à la quantité $P = \mathbb{P}(X > 4)$.

a. Déterminer la valeur exacte de P :

1. par une intégration numérique ;
2. en utilisant le fait que f est un mélange de lois normales.

On estime maintenant la quantité $P = \mathbb{P}(X > 4)$ en utilisant plusieurs méthodes de Monte-Carlo.

b. Simuler $N = 500$ échantillons, chacun de taille `nech`, simulés suivant f. En déduire un intervalle de confiance à 95% sur P en fournissant le code R utilisé.

c. A partir d'un unique échantillon de taille `nech` simulé suivant f, et en remarquant que P correspond à l'estimation de la fonction de répartition de f, donner un intervalle de confiance sur P à 95%.

d. On souhaite utiliser un échantillon W généré suivant une loi g autre que f. Quelle distribution proposez-vous pour estimer P ? Donner le nom de la méthode utilisée et produire un intervalle de confiance à 95% sur P pour un échantillon W de taille $n = 1000$, ainsi que le code R utilisé.

Exercise 13
Soit $X_1, \ldots, X_n$ un n-échantillon distribué sous la loi caractérisée par la fonction de répartition F (et de densité correspondante f). Soit $\alpha \in]0,1[$ et $\hat{q}_n$ est l'estimateur empirique du quantile $F^{-1}(\alpha)$
On voudrait vérifier le théorème limite suivant :

$$\sqrt{n}(\hat{q}_n - F^{-1}(\alpha)) \to_{\mathcal{D}} \mathcal{N}(0, \sigma_q^2) \tag{2}$$

où la variance de la loi normale asymptotique est égale à

$$\sigma_q^2 = \frac{\alpha(1-\alpha)}{[f(F^{-1}(\alpha))]^2}.$$

Dans la suite du problème, on supposera que F est la fonction de répartition de $\mathcal{N}(0, 300)$, c'est-à-dire une normale de moyenne 0 et de variance 300, et donc de densité

$$f(x) = \frac{1}{\sqrt{600\pi}} \exp(-\frac{x^2}{600}).$$

On prendra $\alpha = 0.11$.

1. Simuler $N = 10^3$ fois des échantillons de taille $n = 10^2$ de la loi $\mathcal{N}(0, 100)$, et calculer $\hat{q}_n$ pour chacun de ces échantillons. On pourra écrire une fonction R prenant comme argument (N, n) et fournissant les $\hat{q}_n$.
2. Tracer l'histogramme des $\hat{q}_n$ ainsi obtenus et tester l'adéquation à la loi asymptotique donnée en (2).
3. Donner un intervalle de confiance à 95% sur l'espérance de $\hat{q}_n$.

Exercise 14
Considérons une variable aléatoire X de densité

$$f(x) = 4x^3 \exp(-x^4), \; x \in \mathbb{R}_+.$$

a. Donner la fonction de répartition F de X.

b. Simuler $n = 10^3$ variables indépendantes de F (iid de F).

c. Calculer les quantiles $q_1 = F^{-1}(0.075)$ et $q_2 = F^{-1}(0.9)$.

d. Calculer les valeurs d'estimateurs $\hat{q}_{n,1}$ de q_1 et $\hat{q}_{n,2}$ de q_2 à partir de l'échantillon de taille n simulé à la question b. Donner une évaluation de la précision de ces estimations.

Exercise 15
Soit une densité sur $\mathbb{R}$

$$f(x) \propto x^2 \, \frac{3 + \sin^2(x)}{(1 + \cos(x)^2)^2} \, \exp\{-x^2\}.$$

On pose $g(x) = x^2 \, \exp\{-x^2\}/\Gamma(3/2)$.

1.. Montrer par un changement de variable que g est bien une densité, ce que R confirme par

```
> integrate(function(x) x^2*exp(-x^2)/gamma(1.5),-10,10)
1 with absolute error < 3.2e-05
```

2.. Décrire un algorithme d'acceptation-rejet permettant simuler une variable aléatoire de densité f. On pourra utiliser la fonction `rgamma()`.

3.. Déduire de la simulation ci-dessus un code R donnant une approximation de Monte-Carlo de la constante

$$\int_{\mathbf{R}} x^2 \, \frac{3+\sin^2(x)}{(1+\cos(x)^2)^2} \, \exp\{-x^2\} \mathrm{d}x \, .$$

4.. On obtient

```
> integrate(function(x) x^2*(3+sin(x)^2)*exp(-x^2)/
+ ((1+cos(x)^2)*gamma(1.5)),-10,10)
2.955234 with absolute error < 0.00018
```

Pouvez vous en déduire la probabilité d'acceptation de l'algorithme d'acceptation-rejet ?

Exercise 16
On cherche à approcher l'intégrale

$$\int_{\mathbf{R}^+} \frac{5+\sin^5(x/5)}{(1+x^2)^2} \mathrm{d}x$$

à l'aide de la méthode de Monte-Carlo.

a. Donner une méthode d'acceptation rejet permettant de simuler une variable aléatoire dont la densité est proportionnelle à $\frac{5+\sin^5(x/5)}{(1+x^2)^2}$.

b. Illustrer à l'aide d'un histogramme sur 10^4 simulations la pertinence de votre algorithme de génération.

c. Déduire du taux d'acceptation de l'algorithme d'acceptation rejet la valeur de l'intégrale, en précisant l'erreur statistique de cette approximation (pour une confiance de 95%).

Exercise 17
Déterminer par un code R si, parmi les nombres de la forme $(8+17n)$, $n \in \mathbb{N}$, il existe une puissance de 6, de 7, et de 8.

Exercise 18
Voici une carte d'un jeu de loto

3	15			40	50	62		
	17	21		43			72	85
8			39			65	76	90

Elle est construite sur la base des règles suivantes :
– chaque carte contient 15 nombres choisis entre 1 et 90 disposés sur une grille 9×3 ;
– chaque nombre ne peut figurer qu'une fois ;
– la première colonne ne peut contenir que les nombres de 1 à 9 (9 nombres), la dernière les nombres de 80 à 90 (11 nombres) et les autres colonnes `i*10+0:9` avec `i=1,...,7` (10 nombres) ;
– chaque ligne doit contenir exactement 5 nombres et chaque colonne doit contenir au moins 1 nombre ;
– les nombres d'une même colonne doivent être rangés en ordre croissant.

1. Ecrire une fonction qui génère une carte selon les mêmes règles.
2. A l'aide des fonctions `data.frame(...)`, `plot(...)` et `lines(...)` et avec le bon choix de leurs paramètres, visualiser la carte.

Exercise 19
Soit $X_1, \dots, X_n$ un n-échantillon distribué sous la loi caractérisée par la fonction de répartition F (et de densité correspondante f). Soit $\alpha \in]0,1[$ et $\hat{q}_n$ est l'estimateur empirique du quantile $F^{-1}(\alpha)$
On voudrait vérifier le théorème limite suivant :

$$\sqrt{n}(\hat{q}_n - F^{-1}(\alpha)) \to_{\mathcal{D}} \mathcal{N}(0, \sigma_q^2) \tag{3}$$

où la variance de la loi normale asymptotique est égale à

$$\sigma_q^2 = \frac{\alpha(1-\alpha)}{[f(F^{-1}(\alpha))]^2}.$$

On supposera ici que F est la fonction de répartition de $\mathcal{N}(0,16)$, c'est-à-dire une normale de moyenne 0 et d'écart type 4, et donc de densité

$$f(x) = \frac{1}{4\sqrt{2\pi}} \exp(-\frac{x^2}{32}).$$

On prendra $\alpha = 0.28$.

1. Simuler $N = 10^3$ fois des échantillons de taille $n = 10^2$ de la loi $\mathcal{N}(0,100)$, et calculer $\hat{q}_n$ pour chacun de ces échantillons. On pourra écrire une fonction R prenant comme argument (N, n) et fournissant les $\hat{q}_n$.
2. Tracer l'histogramme des $\hat{q}_n$ ainsi obtenus et tester l'adéquation à la loi asymptotique donnée en (3).
3. Donner un intervalle de confiance à 95% sur l'espérance de $\hat{q}_n$.

Exercise 20
Trouver, par un programme R, toutes les paires de “carrés amis”, c'est-à-dire les paires de carrés parfaits $x^2 > 10$ et $y^2 > 10$ ayant le même nombre de chiffres et tels qu'une translation de tous les chiffres de x^2 par un même chiffre a (modulo 10) redonne y^2. Par exemple, 121 et 676 sont amis.

Exercise 21
On considère la densité

$$f_X(x;\theta) \;=\; \theta\, x^{\theta-1} \mathbb{I}_{(a,b)}(x) \;\;,\;\; 0 < \theta < 1, \;\; 0 < a < b\,.$$

1. Donner l'expression de l'espérance et de la variance de $X \sim f_X$ en fonction de θ.
2. Trouver λ^* et la constante M telles que

$$f_X(x) \leq M f_{\mathcal{E}(\lambda^*)}(x) \quad \forall x \in [a, b]\,,$$

   où $f_{\mathcal{E}(\lambda^*)}$ est la densitè de probabilité exponentielle.
3. En utilisant l'algorithme d'acceptation-rejet, écrire une fonction qui permette de simuler un échantillon $(X_1, X_2, \dots, X_n)$ de n réalisations de variables aléatoires de densité f_X à partir des réalisations d'une variable $\mathcal{E}(\lambda^*)$ pour différentes valeurs des paramètres (a, b, θ) et qui retourne X et le taux d'acceptation τ.
4. A l'aide de cette fonction générer $n = 10^4$ réalisations de $X \sim f_X$ dans les cas $\underline{\theta = 0.4,\ a = 0.5 \text{ et } b = 8}$ et $\underline{\theta = 0.4,\ a = 0.5 \text{ et } b = 1}$ et vérifier graphiquement la pertinence de cet algorithme de génération dans les deux cas : dans quel cas l'algorithme est le moins efficace ? Comment pourrait-on l'améliorer ?

Exercise 22
Considérons une variable aléatoire X de densité

$$f(x) = 6x^5 \exp(-x^6),\ x \in \mathbb{R}_+.$$

a. Donner la fonction de répartition F de X.
b. Simuler $n = 10^4$ variables indépendantes de F (iid de F).
c. Calculer les quantiles $q_1 = F^{-1}(0.125)$ et $q_2 = F^{-1}(0.775)$.
d. Calculer les valeurs d'estimateurs $\hat{q}_{n,1}$ de q_1 et $\hat{q}_{n,2}$ de q_2 à partir de l'échantillon de taille n simulés à la question b.
d. Donner une évaluation de la précision de ces estimations.

Exercise 23
Soit Σ un cercle de rayon unitaire centré en l'origine des axes. Soit X et Y un couple de variables aléatoires distribuées uniformément dans le cercle Σ. On note

$$X = R\cos\Theta\,,$$

$$Y = R\sin\Theta\,,$$

les coordonnées d'un point quelconque $P \in \Sigma$ avec $R \in [0, 1]$ et $\Theta \in [0, 2\pi)$.

1. Déterminer la densité de probabilité jointe de $f_{R,\Theta}(r, \theta)$. Ces deux variables sont-elles indépendantes ?
2. En se servant de $f_{R,\Theta}$ construire un algorithme de génération de points dans Σ.
3. Générer $n = 10^4$ points et vérifier graphiquement la pertinence de votre algorithme.

Exercise 24
On considère la variable aléatoire X de fonction de répartition

$$F(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}$$

1. Écrire du code R pour simuler des réalisations de X par inversion générique.
2. Calculer la densité de X, et la tracer par-dessus un histogramme de $n = 1000$ réalisations de X.

3. On sait que $Var(X) = \pi^2/6$. Construire une expérience de Monte-Carlo pour obtenir un estimateur de la variance. Quelle taille d'échantillon est nécessaire pour obtenir une approximation de π^2 à trois chiffres après la virgule ?

Exercise 25
Soit Y une variable de loi $\mathcal{G}(m, \lambda)$ de densité

$$f_{\mathcal{G}(m,\lambda)}(x) = \frac{1}{(m-1)!}\,\lambda e^{-\lambda x}\,(\lambda x)^{m-1}\,. \quad x \in \mathbb{R}^+$$

1. Générer $n = 10^4$ réalisations de la variable aléatoire Y par inversion générique, en sachant que Y peut être obtenue comme la somme de m variables $\mathcal{E}(\lambda)$ indépendantes.
2. Tracer, sur un même graphique, l'histogramme des réalisations et la densité $f_{\mathcal{G}(m,\lambda)}$.
3. Montrer graphiquement la convergence des estimateurs Monte-Carlo de l'espérance et de la variance de Y vers leur valeurs théoriques m/λ et m/λ^2.

Exercise 26
Ecrire un programme R permettant de résoudre le problème suivant : Soit une loterie comprenant N tickets numérotés de 1 à N. On suppose les N tickets vendus. Les tickets gagnants sont ceux comportant un 1 et un 3 à droite du 1, comme par exemple 123 et 8135. Écrire un programme R permettant de trouver l'unique valeur de $999 < N < 9999$ telle que la proportion de billets gagnants soit exactement 10%.

Exercise 27
On veut simuler $n = 10^4$ réalisation d'une variable aléatoire $X \sim \mathcal{P}(\lambda)$ de densité

$$f(k;\lambda) \;=\; P(X=k) \;=\; P_k \;=\; \frac{\lambda^k e^{-\lambda}}{k!}\,.$$

1. Vérifier la relation de récurrence $P_{k+1} = \frac{\lambda}{k+1}\,P_k$, avec $P_0 \;=\; e^{-\lambda}$.
2. Ecrire une fonction permettant de générer n réalisations de $X \sim \mathcal{P}(\lambda)$ par inversion générique. On rappelle que

$$F_X^{-1}(u) = \inf\left\{x; F_X(x) \geq u\right\}\,.$$

3. Donner une estimation de l'espérance et de la variance et les comparer à la valeur théorique λ.
4. Tracer, sur un même graphique, les fonctions de répartition empirique et théorique (par `ppois()`).

Exercise 28
On considère la variable aléatoire X de densité

$$f(x) = Ce^{-x^2/2}(x - \lfloor x \rfloor) \quad x \in \mathbb{R}$$

où $\lfloor x \rfloor$ représente la partie entière de x, calculée en R par l'instruction `floor(x)`.

1. Écrire un algorithme d'acceptation-rejet permettant de simuler des réalisations de X.
2. En déduire une expérience de Monte-Carlo permettant d'estimer la valeur de la constante C.

3. Quelle est la probabilité d'acceptation de votre algorithme d'acceptation-rejet ?

Exercise 29
Soit l'intégrale
$$I = \int_{-\infty}^{\infty} x^2 \, e^{x-\frac{x^2}{2}} \, dx \, .$$

1. Calculer l'estimation $\hat{I}$ de I en utilisant la méthode Monte-Carlo avec $n = 10^4$ réalisations.
2. Donner une estimation de l'intervalle de confiance de niveau $1-\alpha$ avec $\alpha = 0.05$.
3. Montrer graphiquement la convergence de $\hat{I}$ vers la valeur théorique I (la fonction `integrate()` pourra être utilisée). Tracer sur le meme graphique l'évolution de l'intervalle de confiance en fonction de la taille de l'échantillon.

Exercise 30
On souhaite estimer la constante
$$e = \sum_{n=0}^{\infty} \frac{1}{n!} \, .$$
On considère une suite infinie de variables aléatoires $U_1, U_2, \ldots, U_i \ldots$ de loi $\mathcal{U}_{[0,1]}$ et la variable aléatoire K qui indique la première réalisation qui dépasse la réalisation qui la précède, c'est à dire
$$K = \min_{\{k \geq 2\}} \{k : \; U_k \geq U_{k-1}\} \, .$$
On observe que $P(K = k) = \frac{k-1}{k!}$ car $P(K > k) = P(U_1 \geq U_2 \geq U_3 \geq \cdots \geq U_k) = \frac{1}{k!}$ et donc
$$E[K] = \sum_{k=2}^{\infty} k \, \frac{k-1}{k!} = \sum_{n=0}^{\infty} \frac{1}{n!} = e \, .$$

1. Ecrire une fonction qui simule n réalisations de la variable k.
2. Pour $n = 10^3$ donner une estimation par Monte-Carlo de e et la comparer à sa valeur théorique.
3. Donner une estimation de l'intervalle de confiance de niveau $1-\alpha$ avec $\alpha = 0.05$.
4. Vérifier graphiquement la convergence de l'estimateur vers sa valeur théorique. Tracer sur le même graphique l'évolution de l'intervalle de confiance.

Exercise 31
On considère la v.a. X de densité
$$f_X(x) = C \left[\left(x - \frac{1}{2} \right)^2 - x + \frac{3}{4} \right] \mathbb{I}_{[0,1]}(x) \, .$$

1. Calculer la constante de normalisation C, l'espérance et la variance de $X \sim f_X$.
2. Proposer une méthode de simulation de $X \sim f_X$, motiver son choix et décrire dans les détails son application à ce problème.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$.
4. Donner une estimation de la constante de normalisation C.

Exercise 32
Considérons une variable aléatoire X de densité

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0, \sigma > 0$$

1. Donner la fonction de répartition F de X.
2. Simuler un échantillon de taille $n = 10^4$ tel que les X_i sont iid de F en utilisant la méthode d'inversion générique.
3. Illustrer graphiquement la pertinence de votre algorithme.
4. Calculer les quantiles $q_1 = F^{-1}(0.025)$ et $q_2 = F^{-1}(0.975)$.

Exercise 33
A l'aide de la méthode Acceptation-Rejet,

1. Générer une réalisation d'une variable aléatoire de densité

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

   On pourra utiliser la densité uniforme sur l'intervalle $[-1, 1]$. Illustrer graphiquement la pertinence de votre algorithme de génération.
2. On se place maintenant dans $\mathbb{R}^2$. Générer une réalisation d'une variable aléatoire de densité

$$g(x_1, x_2) = \frac{1}{2} \mathbb{I}_{\{(x_1,x_2) \in R^2 |\ |x_1| + |x_2| \leq 1\}}(x_1, x_2).$$

   On pourra utiliser la densité uniforme sur le carré $[-1, 1]^2$. Illustrer graphiquement la pertinence de votre algorithme de génération.

Exercise 34
Générer une réalisation d'une variable aléatoire de densité

$$f(x) = \frac{45}{16} x (1 - x)^{\frac{1}{4}}, 0 < x < 1$$

On pourra utiliser la densité $g(x) = 2x, 0 < x < 1$.Illustrer graphiquement la pertinence de votre algorithme de génération.

Exercise 35
Nous considérons une variable aléatoire X de densité de probabilité
$f(x) = \frac{2}{\pi}\sqrt{1 - x^2}, -1 \leqslant x \leqslant 1$.

1. Déterminer par la méthode de Monte-Carlo des approximations de l'espérance et de la variance de X.
2. Fournir un intervalle de confiance à 95% pour $n = 10^4$.

Exercise 36
Calculer par une méthode de simulation des approximations de la fonction de répartition de $N(0, 1)$. Donner la variance de l'estimateur.

Exercise 37
Soient $a \in \mathbb{R}$ et $k > 1$. La fonction de répartition de la loi de Pareto de paramètres (a, k) s'écrit de la façon suivante :

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ 1 - \left(\frac{x}{a}\right)^{-k} & \text{sinon} \end{cases}$$

1. Ecrire une fonction rpareto2 prenant pour arguments n un entier et les paramètres a et k, permettant de simuler n réalisations indépendantes de loi de Pareto(a, k) à partir de n réalisations indépendantes de loi $\mathcal{U}_{[0,1]}$.
2. Choisir un couple de paramètres (a, k) et illustrer graphiquement la pertinence de votre méthode de simulation.
3. A partir d'un 1000-échantillon que vous aurez simulé, donner un intervalle de confiance de niveau asymptotique 5% pour l'espérance d'une v.a. de Pareto(a, k).

Exercise 38
Soient $\lambda > 0$ et $k > 0$. On considère la fonction de répartition suivante :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-(x/\lambda)^k} & \text{sinon} \end{cases}$$

1. Ecrire une fonction rweibull2 prenant pour arguments n un entier ainsi que les paramètres a et k et permettant de simuler n réalisations indépendantes de loi de Weibull(λ, k) à partir de n réalisations indépendantes de loi $\mathcal{U}_{[0,1]}$.
2. Choisir un couple de paramètres (λ, k) et illustrer graphiquement la pertinence de votre méthode de simulation.
3. A partir d'un 1000-échantillon que vous aurez simulé, donner un intervalle de confiance de niveau asymptotique 5% pour l'espérance d'une v.a. de Weibull(λ, k).

Exercise 39
On considère la densité

$$f_X(x; \theta) \;=\; \theta\, x^{\theta-1} \mathbb{I}_{(a,b)}(x) \;\;,\;\; 0 < \theta < 1, \;\; 0 < a < b\,.$$

1. En utilisant l'algorithme d'acceptation-rejet, écrire une fonction qui permette de simuler un échantillon $(X_1, X_2, \ldots, X_n)$ de n réalisations de variables aléatoires de densité f_X à partir des réalisations d'une variable aléatoire uniforme pour différentes valeurs des paramètres (a, b, θ) et qui retourne X et le taux d'acceptation τ.
2. A l'aide de cette fonction générer $n = 10^4$ réalisations de $X \sim f_X$ dans les cas $\underline{\theta = 0.9,\ a = 0.5 \text{ et } b = 8}$ et $\underline{\theta = 0.4,\ a = 0.5 \text{ et } b = 8}$ et vérifier graphiquement la pertinence de cet algorithme de génération dans les deux cas : dans quel cas l'algorithme est le moins efficace ? Comment pourrait-on l'améliorer ?

Exercise 40
On considère la fonction de répartition de la loi logistique définie de la façon suivante :

$$F(x) = \frac{1}{1 + e^{-x}} \quad x \in \mathbb{R}$$

1. Ecrire une fonction rlogis2 prenant pour argument n un entier et permettant de simuler n réalisations indépendantes de loi logistique à partir de n réalisations indépendantes de loi $\mathcal{U}_{[0,1]}$.
2. Illustrer graphiquement la pertinence de votre méthode de simulation.
3. A partir d'un 1000-échantillon que vous aurez simulé, donner un intervalle de confiance de niveau asymptotique 5% pour l'espérance d'une v.a. de loi logistique.

Exercise 41

On cherche à simuler des points uniformément répartis dans le domaine contenu dans un cercle de centre $(0,0)$ et de rayon ρ. On rappelle que la densité de probabilité de cette loi est :

$$f_{X,Y}(x,y) = \frac{1}{\mathcal{A}ire(\mathcal{C})}\mathbb{I}_{(x,y)\in\mathcal{C}} = \frac{1}{\pi\rho^2}\mathbb{I}_{x^2+y^2\leq\rho^2}$$

Soient (R,θ) tels que $(X,Y) = (R\cos\theta, R\sin\theta)$.

1. Justifier la forme de la densité $f_{X,Y}$.
2. En utilisant en changement en coordonnées polaires, calculer la loi du couple (R,θ).
3. Après avoir isolé la densité de R, proposer une méthode de simulation pour R reposant sur une v.a $\mathcal{U}_{[0,1]}$.
4. En déduire une méthode de simulation de points uniformément répartis dans le cercle de rayon ρ.
5. Illustrer graphiquement votre méthode : tracer le cercle de rayon ρ et afficher les points simulés par la méthode précédente.

Exercise 42

1. Soient (X,Y) deux variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[-1,1]$.
   (a) Montrer que le couple (X,Y) est de loi uniforme sur un domaine $\mathcal{D}$ de $\mathbb{R}^2$ à préciser.
   (b) Donner la probabilité théorique que $P(X^2+Y^2\leq 1)$ (un dessin et un raisonnement géométrique pourront suffire).
2. Proposer une méthode de Monte-Carlo pour calculer la probabilité $P(X^2+Y^2\leq 1)$. En déduire une approximation de π.

Exercise 43

On se place dans $\mathbb{R}^2$ et on considère le domaine ellipsoidal $\mathcal{D} = \{(x,y)\in\mathbb{R}^2, \left(\frac{x}{a}\right)^2+\left(\frac{y}{b}\right)^2\leq 1\}$.

Choisir (a,b) différents de $(1,1)$.

1. Tracer l'ellipse en utilisant le fait que l'on peut décrire cette ellipse de façon paramétrique : l'ellipse est l'ensemble des points $\mathcal{E} = \{(x(t),y(t)), t\in[0,2\pi]\}$ tel que

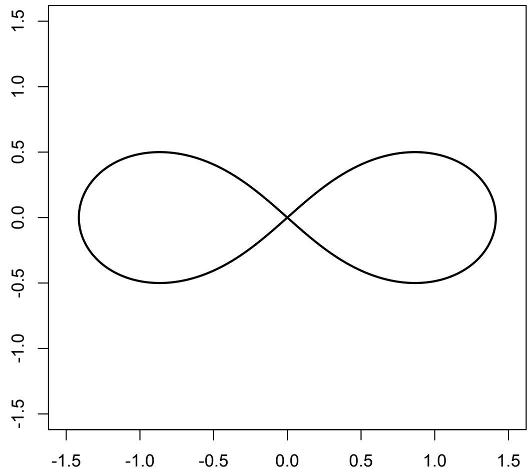
$$\begin{cases} x(t) &= a\cos(t) \\ y(t) &= b\sin(t) \end{cases}$$

2. Tracer le plus petit rectangle possible contenant cette ellipse. On notera $\mathcal{R}$ ce rectangle

3. Proposer une méthode permettant de générer des points (X, Y) de loi uniforme sur ce rectangle.
4. En déduire une approximation par Monte-Carlo de la surface du domaine $\mathcal{D}$.

Exercise 44
On se place dans $\mathbb{R}^2$ et on considère le domaine inclus dans la courbe du Lemniscate de Bernoulli

$$\mathcal{D} = \{(r\cos(\theta), r\sin(\theta)) \in \mathbb{R}^2, r^2 \leq 2\cos(2\theta)\}$$



1. On peut réécrire la courbe de Lemniscate par une équation paramétrique de la forme :

$$x(t) = \frac{\sqrt{2}\sin(t)}{\cos^2(t)+1} \quad y(t) = \frac{\sqrt{2}\cos(t)\sin(t)}{\cos^2(t)+1}, \quad t \in [0, 2\pi]$$

   Tracer cette courbe. *On fera varier t entre 0 et 2π. Pour chaque valeur de t, on peut tracer les points de coordonnées $(x(t), y(t))$ et relier ces points.*
2. Proposer un carré le plus petit possible contenant ce domaine. Le tracer sur le même graphique
3. Simuler des points (X, Y) uniformément répartis dans ce carré. Les afficher en rouge.
4. En posant

$$R^2 = X^2 + Y^2 \quad \text{et} \quad \theta = tan^{-1}(\frac{Y}{X})$$

   afficher en vert les points contenu dans le domaine du Lemniscate. *Utiliser la fonction* `atan` *de R.*
5. En déduire une estimation de l'aire du domaine contenu dans le Lemniscate
6. Peut-on améliorer la méthode ?

Exercise 45
On considère la v.a. X de densité

$$f_X(x) = C\left(x - \frac{1}{2}\right)^2 \mathbb{I}_{[0,1]}(x).$$

1. Calculer la constante de normalisation C, l'espérance et la variance de $X \sim f_X$.
2. Proposer une méthode de simulation de $X \sim f_X$, motiver son choix et décrire dans les détails son application à ce problème.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$.
4. Donner une estimation de la constante de normalisation C.

Exercise 46
Afin d'étudier le propagation des noms de familles d'une génération à l'autre dans l'Angleterre Victorienne, Sir Galton a proposé en 1873 le modèle stochastique suivant. Chaque adulte mâle transmet son nom à ses fils. Supposons que le nombre de fils de chaque homme soit une variable aléatoire entière (de même loi pour tous les hommes). Alors on peut montrer que

- si le nombre moyen de fils par homme (noté m) est inférieur ou égal à 1, le nom de famille va disparaître presque surement.
- Si $m > 1$ la probabilité de survie du nom est non-nulle et le nombre de porteurs du nom connait une croissance exponentielle.

Notons X_n le nombre de porteurs du nom de famille à la génération n. Soit $Y_{n,k}$ le nombre de fils de l'individu k de la n-ième génération. Alors

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$$

avec $X_1 = 1$.
On suppose que $Y_{n,k}$ est une variable aléatoire dans $\{0, 1, 2\}$ telle que

$$P(Y_{n,k} = 0) = p_0, \quad P(Y_{n,k} = 1) = p_1 \quad P(Y_{n,k} = 2) = 1 - (p_0 + p_1)$$

1. Calculer m le nombre moyen de fils par homme en fonction de (p_0, p_1). Calculer la variance σ^2 de $Y_{n,k}$ en fonction de (p_0, p_1).
2. Ecrire une fonction `galton` prenant en argument n_{gene} le nombre de générations et le vecteur des probabilités (p_0, p_1) et donnant en sortie une réalisation du processus $(X_n)_{n=1 \ldots n_{gene}}$ décrit précédemment.
3. On pose $(p_0, p_1, p_2) = (0.25, 0.5, 0.25)$
   (a) Que valent m et σ^2 ?
   (b) Représenter sur une même figure 5 réalisations simulées avec la fonction `galton` et $n_{gene} = 100$.
   (c) Proposer une méthode de Monte-Carlo permettant de calculer la $P(X_{100} = 0)$
   (d) Après avoir récupéré 100 trajectoires telles que $X_{n_{gene}} > 0$, donner une approximation de $E[X_{gene}|(X_{gene} > 0)]$ avec $n_{gene} = 10, 20, 50$. Vérifier que cette moyenne est proche de $n_{gene}\sigma^2/2$.

Exercise 47
Nous cherchons à évaluer l'intégrale suivante :

$$I = \int_0^1 e^{-\frac{x^2}{2}}\,dx$$

par des méthodes de Monte-Carlo.

1. Donner la commande R permettant d'obtenir la valeur exacte de I sous la forme

   `0.8556244 with absolute error < 9.5e-15`

2. Proposer une méthode de Monte-Carlo reposant sur la génération d'un n-échantillon de variables aléatoires gaussiennes. Donner le code permettant de construire un intervalle de confiance à 95% sur I pour $n = 1000$.
3. Proposer une méthode de Monte-Carlo reposant sur la génération d'un n-échantillon de variables aléatoires de loi uniforme.
4. Comment comparer ces 2 méthodes ?
5. Montrer que
$$I = \int_0^1 e^{\frac{-(1-x)^2}{2}}\mathrm{d}x = \frac{1}{2}\int_0^1 \left[e^{\frac{-x^2}{2}} + e^{\frac{-(1-x)^2}{2}}\right]\mathrm{d}x$$
6. En déduire une nouvelle méthode d'estimation de I reposant sur des variables aléatoires uniformes.
7. En utilisant des simulations d'une loi Béta $B(a,a)$, par `rbeta(n,a,a)`, détailler l'algorithme d'optimisation du choix de a pour l'estimation. Fournir le code R correspondant.

Exercise 48
On souhaite simuler des réalisations de la loi sur $\mathbb{R}$ dont la densité est proportionelle à :

$$f(x) \propto \begin{cases} \frac{\sin^4(x/2)}{x^2} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$$

On souhaite utiliser pour cela la loi de Cauchy $\mathcal{C}(0,1)$, de densité $g(x) = \dfrac{1}{\pi(1+x^2)}$.

1. Soit C la constante de normalisation de la densité f, telle que : $f(x) = \frac{\sin^4(x/2)}{Cx^2}$ pour $x \neq 0$. Montrer que :
$$\sup_x \frac{f(x)}{g(x)} \quad \leq \quad \frac{2\pi}{C}$$
2. Proposer un algorithme d'acceptation-rejet permettant de simuler des réalisations de la densité f (on pourra utiliser la fonction `rcauchy`).
3. Déduire des sorties de l'algorithme une estimation par Monte-Carlo de la constante de normalisation C. Utiliser cette estimation pour illustrer graphiquement la pertinence de l'algorithme.

Exercise 49
On cherche à simuler des réalisations d'une loi de Poisson de paramètre $\lambda > 0$ à partir de réalisations d'une loi uniforme $\mathcal{U}_{[0,1]}$.

1. Construire une fonction rexp2 prenant pour arguments un entier n et λ et permettant de simuler n réalisations indépendantes de loi exponentielle de paramètre λ à partir de n réalisations de loi $\mathcal{U}_{[0,1]}$.
2. Soient $Y_1, Y_2, \ldots$ des variables aléatoires indépendantes et identiquement distribuées de loi exponentielle de paramètre λ. Posons

$$Z_k = Y_1 + Y_2 + \cdots Y_k$$

et définissons la variable aléatoire X telle que :

$$X = k \text{ si } Z_k < 1 < Z_{k+1}$$

Autrement dit on simule des variables aléatoires exponentielles de paramètre λ et on compte le nombre de simulations nécéssaires pour que leur somme cumulée dépasse 1.
Ecrire une fonction rpoisson2 prenant pour argument λ et permettant de simuler 1 réalisation de la variable aléatoire X.
3. Utiliser la fonction rpoisson2 pour simuler un échantillon de taille 1000. Proposer une méthode graphique pour vérifier que X suit une loi de poisson de paramètre λ.

Exercise 50
On considère la loi $\Gamma(5,1)$ tronquée en a, de densité

$$g(x) \propto x^4 e^{-x} \mathbb{I}_{[a,\infty[}(x).$$

On souhaite générer des réalisations de cette loi.

1. Proposer un algorithme de rejet pour générer selon cette loi. Quel est son taux d'acceptation ?
2. Soit $U \sim \mathcal{U}([0,1])$ et soit F la fonction de répartition de la loi $\Gamma(5,1)$ (non tronquée). On pose $Z = F^{-1}(F(a) + (1 - F(a))U)$. Montrer que la densité de Z est g.
3. Utiliser cette seconde méthode pour générer selon la densité g. Comparer les deux méthodes.
4. Donner une formule R pour calculer l'intégrale $\int_a^\infty x^4 e^{-x} dx$. Proposer une méthode de Monte-Carlo pour calculer une valeur approchée de cette intégrale.

Exercise 51
On considère la variable aléatoire X de densité

$$f(x) = \frac{1}{x^2} \mathbb{I}_{[1,\infty[}(x)$$

1. Écrire une fonction R pour générer n réalisations de X par inversion générique.
2. Observer le comportement de la moyenne empirique quand n augmente. Comment expliquez-vous ce phénomène ?

Exercise 52
On considère les variables aléatoires X et Y de densités définies sur $\mathbb{R}$

$$f_X(x) = C\, e^{-|x|}$$

$$f_Y(x) = e^{-x} e^{-e^{-x}}$$

1. Calculer C de manière à ce que f_X soit bien une densité.
2. Écrire une fonction R qui génère des réalisations de X. On pourra utiliser la fonction `rexp()`.
3. Trouver une constante M telle que $\forall x \in \mathbb{R}, f_Y(x) \leq M f_X(x)$.
4. Écrire une fonction R pour générer des réalisations de Y par acceptation-rejet.

Exercise 53
On considère la variable aléatoire X de densité

$$f(x) = C\, e^{-x(\log(x))^2} \mathbb{I}_{\mathbb{R}_+}(x).$$

1. Écrire un algorithme d'acceptation-rejet permettant de simuler des réalisations de X.
2. En déduire une expérience de Monte-Carlo permettant d'estimer la valeur de la constante C.
3. Quelle est la probabilité d'acceptation de votre algorithme d'acceptation-rejet ?

Exercise 54
Soient X et Y deux variables aléatoires indépendantes de loi exponentielle de paramètre 1.

1. Calculer la loi conditionnelle de X sachant que $\{X > \sqrt{1-2Y}\}$
2. En déduire un algorithme de simulation de la loi normale $\mathcal{N}(0,1)$.
3. Écrire une fonction R suivant cet algorithme. Calculer le taux d'acceptation.

Exercise 55
On sait que si i et j sont deux entiers tirés aléatoirement entre 1 et n, la probabilité que i et j soient premiers entre eux vaut approximativement $6/\pi^2$ (cette probabilité tend vers $6/\pi^2$ quand n tend vers l'infini).
Utiliser cette propriété pour calculer en R une approximation de π. Observer l'influence de la valeur de n.

On pourra utiliser les instructions

```
install.packages("schoolmath")
require(schoolmath)
```

pour installer le package `schoolmath`, puis utiliser la fonction `gcd(a,b)` pour calculer le plus grand diviseur commun des nombres a et b.

Exercise 56
Soient $\alpha > 0$ et $\beta > 0$. Soit X une variable aléatoire de loi admettant comme fonction de répartition :

$$F(x) = \begin{cases} 1 - \exp\{-\alpha x^\beta\} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

1. Expliciter la densité de la loi de X
2. Proposer une méthode de simulation de cette loi. Ecrire une fonction `rexo` prenant pour arguments un entier n, α et β permettant de simuler n réalisations indépendantes de loi caractérisée par F à partir de n réalisations de loi $\mathcal{U}_{[0,1]}$.

3. Choisir un couple de paramètres (α, β) et simuler un échantillon de taille 1000.
4. Proposer une méthode graphique pour illustrer la pertinence de votre méthode de simulation.

Exercise 57
On cherche à simuler des réalisations d'une loi de Poisson de paramètre $\lambda > 0$ à partir de réalisations d'une loi uniforme $\mathcal{U}_{[0,1]}$.
Rappelons d'abord que si X suit une loi de Poisson de paramètre λ alors pour tout entier naturel k,

$$p_k = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Ceci implique la relation suivante entre p_k et p_{k+1}.

$$p_{k+1} = e^{-\lambda}\frac{\lambda^{k+1}}{(k+1)!} = p_k\frac{\lambda}{k+1}$$

Notons P_k les probabilités cumulées

$$P_k = P(X \leq k) = \sum_{j=1}^{k} p_j$$

On a alors :

$$P_{k+1} = P_k + p_{k+1} = P_k + p_k\frac{\lambda}{k+1}$$

Soit $U \sim \mathcal{U}_{[0,1]}$. On pose

$$X = k \text{ si } P_{k-1} < U < P_k$$

1. En utilisant les remarques faites sur les probabilités cumulées P_k, écrire un programme permettant de simuler une réalisation de la variable aléatoire X définie ci-dessus. Ecrire une fonction `rpoiss2` prenant pour argument λ.
2. Utiliser la fonction `rpoiss2` pour simuler un échantillon de taille 1000. Proposer une méthode graphique pour vérifier que X suit une loi de Poisson de paramètre λ.

Exercise 58
Une calculatrice est abimée : elle démarre en affichant 0 ; si on appuie sur 4, 6 ou 0, ce chiffre est ajouté à droite du nombre affiché (à l'exception du 0 affiché qui est remplacé par le chiffre) ; si on appuie sur 2, le nombre affiché est divisé par 2. Les autres touches ne fonctionnent pas. Un exemple de suite est

```
0
4 [press 4]
46 [press 6]
23 [press 2]
230 [press 0]
```

Est-il possible d'afficher n'importe quel entier sur cette calculatrice ? Ecrire un code R permettant d'aborder ce problème.

Exercise 59
On considère la v.a. X de densité

$$f_X(x) \,=\, C\,\left[1-3\left(x-\frac{1}{2}\right)^2\right]\mathbb{I}_{[0,1]}(x)\,.$$

1. Caculer l'espérance et la variance de $X \sim f_X$.
2. Proposer une méthode de simulation de $X \sim f_X$ basée sur l'algorithme d'acceptation-rejet.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$.
4. Donner une estimation de l'ésperance et de la variance de X ainsi que l'intervalle de confiance de niveau $1-\alpha$ avec $\alpha = 0.05$.
5. Calculer la constante de normalisation C

$$C \quad : \quad C\int_0^1 f_X(x)\,dx \,=\, 1\,,$$

et en donner une estimation numérique.

Exercise 60
Considérons une expérience de Bernoulli : soient $X_1, X_2, \cdots \sim_{i.i.d}\ Bern(p)$. Posons Z le premier instant de succès : Z est le premier entier k tel que $X_k = 1$ et $X_1 = X_2 = \cdots = X_{k-1} = 0$.

1. Proposer une méthode permettant de simuler une réalisation d'une v.a. de loi de Bernoulli de paramètre p à partir de la réalisation d'une v.a. de loi uniforme $\mathcal{U}_{[0,1]}$.
2. Ecrire un programme permettant de simuler UNE réalisation de la variable aléatoire Z définie ci-dessus. Ecrire une fonction rgeom2 prenant pour argument p.
3. Utiliser la fonction rgeom2 pour simuler un échantillon de taille 1000. Proposer une méthode graphique pour vérifier que Z suit une loi géomtrique de paramètre p (on pourra utiliser les fonctions pgeom, dgeom ou qgeom existant sous R).

Exercise 61
On cherche à approcher l'intégrale

$$\int_{\mathbf{R}^-} \frac{2+\sin^4(x/6)}{(1+x^2)^{3/2}}\mathrm{d}x$$

à l'aide de la méthode de Monte-Carlo.

a. Donner une méthode d'acceptation rejet permettant de simuler une variable aléatoire dont la densité est proportionnelle à $\frac{2+\sin^4(x/6)}{(1+x^2)^{3/2}}$.

b. Illustrer à l'aide d'un histogramme sur 10^4 simulations la pertinence de votre algorithme de génération.

c. Déduire du taux d'acceptation de l'algorithme d'acceptation rejet la valeur de l'intégrale, en précisant l'erreur statistique de cette approximation (pour une confiance de 95%).

Exercise 62
On considère la loi de Pareto $\mathcal{P}(\alpha, x_m)$ de fonction de répartition :

$$F(x) \quad = \quad [1 - (x_m/x)^\alpha]\mathbf{1}_{\{x \geq x_m\}}$$

On souhaite générer des réalisations de cette loi.

1. Proposer un algorithme d'inversion générique pour simuler n réalisations de la loi de Pareto de paramètres α et x_m donnés.
2. Montrer que la densité de la loi de Pareto s'écrit :

$$f(x) \quad = \quad \frac{\alpha x_m^\alpha}{x^{\alpha+1}}\mathbf{1}_{\{x \geq x_m\}}$$

et illustrer la pertinence de l'algorithme de simulation (on prendra $\alpha = 2, x_m = 1, n = 10\,000$)
3. *Autre méthode :* Montrer que si Y suit la loi exponentielle $\mathcal{E}(\alpha)$, de fonction de répartition $F_Y(x; \alpha) = 1 - e^{\alpha x}$, et si $X = x_m e^Y$, alors X suit la la loi de Pareto $\mathcal{P}(\alpha, x_m)$. En déduire un deuxième algorithme permettant de simuler n réalisations de la loi de Pareto de paramètres α et x_m donnés.
4. Illustrer graphiquement la pertinence de ce deuxième algorithme.

Exercise 63
Afin d'étudier le propagation des noms de familles d'une génération à l'autre dans l'Angleterre Victorienne, Sir Galton a proposé en 1873 le modèle stochastique suivant. Chaque adulte mâle transmet son nom à ses fils. Supposons que le nombre de fils de chaque homme soit une variable aléatoire entière (de même loi pour tous les hommes). Alors on peut montrer que

- si le nombre moyen de fils par homme (noté m) est inférieur ou égal à 1, le nom de famille va disparaître presque surement.
- Si $m > 1$ la probabilité de survie du nom est non-nulle et le nombre de porteurs du nom connait une croissance exponentielle.

Notons X_n le nombre de porteurs du nom de famille à la génération n. Soit $Y_{n,k}$ le nombre de fils de l'individu k de la n-ième génération. Alors

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$$

avec $X_1 = 1$.
On suppose que $Y_{n,k}$ est une variable aléatoire dans $\{0, 1, 2\}$ telle que

$$P(Y_{n,k} = 0) = p_0, \quad P(Y_{n,k} = 1) = p_1 \quad P(Y_{n,k} = 2) = 1 - (p_0 + p_1)$$

1. Calculer m le nombre moyen de fils par homme en fonction de (p_0, p_1). Calculer la variance σ^2 de $Y_{n,k}$ en fonction de (p_0, p_1).
2. Ecrire une fonction `galton` prenant en argument n_{gene} le nombre de générations et le vecteur des probabilités (p_0, p_1) et donnant en sortie une réalisation du processus $(X_n)_{n=1\ldots n_{gene}}$ décrit précédemment.

3. On pose $(p_0, p_1, p_2) = (0.1, 0.6, 0.3)$ et $n_{gene} = 80$.
   (a) Que vaut m ?
   (b) Représenter sur une figure 5 réalisations simulées avec la fonction `galton`.
   (c) Proposer une méthode de Monte-Carlo permettant de calculer la probabilité d'extinction.
   (d) Pour 5 trajectoires telles que $X_{n_{gene}} > 0$, représenter l'évolution de X_n/m^n. Que remarquez-vous ?

Exercise 64
On souhaite simuler $X_1, X_2 \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$

1. Ecrire la densité jointe $f_{(X_1,X_2)}(x,y)$ du couple (X_1, X_2)
2. *Passage en coordonnées polaires :* Soit (Θ, R) le couple de variables aléatoires telles que :

$$\begin{aligned} X_1 &= R\cos(\Theta); \\ X_1 &= R\sin(\Theta). \end{aligned}$$

   Ecrire la densité jointe $f_{(\Theta,R)}(\theta, r)$ du couple (Θ, R).
3. Montrer que R et Θ sont indépendantes, que $\Theta \sim \mathcal{U}([0, 2\pi])$ et que $R^2 \sim \mathcal{E}(1/2)$, où $\mathcal{E}(1/2)$ est la loi exponentielle de paramètre $1/2$, de fonction de répartition $F(x) = 1 - e^{-x/2}$.
4. En déduire une transformation permettant d'obtenir X_1, X_2 à partir de deux réalisations indépendantes U_1, U_2 de la loi $\mathcal{U}([0,1])$.

Exercise 65
Considérons une variable aléatoire X de densité

$$f(x) = 2x\exp(-x^2),\ x \in \mathbb{R}_+.$$

a. Donner la fonction de répartition F de X.
b. Simuler $n = 10^4$ variables indépendantes de F (iid de F).
c. Calculer les quantiles $q_1 = F^{-1}(0.225)$ et $q_2 = F^{-1}(0.925)$.
d. Calculer les valeurs d'estimateurs $\hat{q}_{n,1}$ de q_1 et $\hat{q}_{n,2}$ de q_2 à partir de l'échantillon de taille n simulés à la question b. Donner une évaluation de la précision de ces estimations.

Exercise 66
On considère la v.a. X de densité

$$f_X(x) = C\left(x - \frac{1}{2}\right)^2 \mathbb{I}_{[0,1]}(x)\,.$$

1. Calculer la constante de normalisation C, l'espérance et la variance de $X \sim f_X$.
2. Proposer une méthode de simulation pour $X \sim f_X$ basée sur l'algorithme d'acceptation-rejet.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$.

4. Donner une estimation de l'espérance et de la variance de X ainsi que l'intervalle de confiance de niveau $1-\alpha$ avec $\alpha = 0.05$.
5. Donner une estimation numérique de la constante de normalisation C.

Exercise 67

On souhaite simuler des réalisations de la loi sur $\mathbb{R}$ dont la densité est proportionelle à :

$$f(x) \propto \begin{cases} \frac{\sin^2(x)}{x^2} & \text{si } x \neq 0 \\ 1 & \text{si } x = 0 \end{cases}$$

On souhaite utiliser pour cela la loi de Cauchy $\mathcal{C}(0,1)$, de densité $g(x) = \dfrac{1}{\pi(1+x^2)}$.

1. Soit C la constante de normalisation de la densité f, telle que : $f(x) = \frac{\sin^2(x)}{Cx^2}$ pour $x \neq 0$. Montrer que :

$$\sup_x \frac{f(x)}{g(x)} \leq \frac{2\pi}{C}$$

2. Proposer un algorithme d'acceptation-rejet permettant de simuler des réalisations de la densité f (on pourra utiliser la fonction `rcauchy`)
3. Déduire des sorties de l'algorithme une estimation par Monte-Carlo de la constante de normalisation C. Utiliser cette estimation pour illustrer graphiquement la pertinence de l'algorithme.

Exercise 68

On cherche à approcher l'intégrale

$$\int_{\mathbf{R}} \frac{4+\sin^3(x/2)}{(1+x^2)^{5/2}} \mathrm{d}x$$

à l'aide de la méthode de Monte-Carlo.

a. Donner une méthode d'acceptation rejet permettant de simuler une variable aléatoire dont la densité est proportionnelle à $\frac{4+\sin^3(x/2)}{(1+x^2)^{5/2}}$

b. Illustrer à l'aide d'un histogramme sur 10^4 simulations la pertinence de votre algorithme de génération.

c. Déduire du taux d'acceptation de l'algorithme d'acceptation rejet la valeur de l'intégrale, en précisant l'erreur statistique de cette approximation (pour une confiance de 95%).

Exercise 69

On tire un entier k au hasard, uniformément entre 1 et n. On sait que la probabilité de l'événement "k est divisible par un carré parfait" vaut approximativement $6/\pi^2$ (cette probabilité tend vers $6/\pi^2$ quand n tend vers l'infini).

Utiliser cette propriété pour calculer en R une approximation de π. Observer l'influence de la valeur de n.

On rappelle que l'instruction `a%%b` calcule $a \bmod b$, le reste dans la division de a par b.

Exercise 70
On considère la loi Gamma $\mathcal{G}(k,\theta)$, à valeurs dans $\mathbb{R}_+$ et de densité : $f(x;k,\theta) = \frac{\theta^k}{(k-1)!}x^{k-1}e^{-\theta x}$, pour $k \in \mathbb{N}^*$ et $\theta > 0$.

1. On admet la propriété suivante : si $X_1, \ldots, X_k$ sont des réalisations indépendantes de la loi exponentielle $\mathcal{E}(\theta)$, de fonction de répartition $F(x) = 1 - e^{\theta x}$, et si $Z = X_1 + \ldots + X_k$, alors $Z \sim \mathcal{G}(k,\theta)$. En déduire un algorithme permettant de simuler n réalisations de la loi $\mathcal{G}(k,\theta)$, pour k entier naturel non nul (on pourra utiliser la fonction `rexp`).
2. Illustrer graphiquement la pertinence de l'algorithme (on utilisera $k = 3, \theta = 0.1, n = 10\,000$.)
3. À partir des sorties de l'algorithme, donner un estimateur de Monte-Carlo de l'espérance et de la variance de la loi $\mathcal{G}(k,\theta)$. Comparer aux valeurs théoriques (k/θ et k/θ^2, respectivement)
4. Donner un intervalle de confiance à 95% de l'espérance de la loi $\mathcal{G}(k,\theta)$.

Exercise 71
On considère la v.a. X de densité

$$f_X(x) = C\left(x - \frac{1}{2}\right)^2 \mathbb{I}_{[0,1]}(x)\,.$$

1. Calculer la constante de normalisation C, l'espérance et la variance de $X \sim f_X$.
2. Proposer une méthode de simulation pour $X \sim f_X$ basée sur la méthode d'inversion générique.
3. Tracer l'histogramme des réalisations et le confronter à la densité f_X pour $n = 10^4$.
4. Donner une estimation de la constante de normalisation C.