Corrigé de l'évaluation de travaux pratiques

Exercice 1 Des dés

Partie A : Un seul dé

On considère une expérience aléatoire consistant à lancer un dé bien équilibré. On note D la variable aléatoire égale au résultat obtenu.

1. ✐ Quelle est la loi de D ? Quelle est son espérance μ, sa variance σ^2 ?
   Solution : La loi de D est uniforme sur l'ensemble fini $\{1, 2, 3, 4, 5, 6\}$. On peut également la décrire par le tableau suivant :

i	1	2	3	4	5	6
$\mathbb{P}(D = i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

   Son espérance et sa variance se calculent très facilement :

$$\mu = \mathbb{E}[D] = \frac{1}{6} \times 1 + \cdots + \frac{1}{6} \times 6 = \frac{7}{2}$$
$$\mathbb{E}\left[D^2\right] = \frac{1}{6} \times 1^2 + \frac{1}{6} \times 2^2 + \cdots + \frac{1}{6} \times 6^2 = \frac{91}{6}$$
$$\sigma^2 = \mathbb{V}\mathrm{ar}[D] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

2. Créer une fonction `de6faces(N)` qui renvoie un vecteur ligne de taille `N` dont les coefficients sont des variables aléatoires indépendantes de même loi que D. (On pourra utiliser la fonction `floor` qui renvoie la partie entière d'un nombre ou d'une matrice, élément par élément.)
   Solution :

```
function D = de6faces(N)
  D = floor(rand(1,N)*6)+1;
end
```

3. Tracer l'histogramme d'un échantillon de $N = 1000$ valeurs suivant la loi de D.
   Solution : L'histogramme est tracé avec les instructions :

```
N = 10000;
D = de6faces(N);
hist(D); % hist(D, 1:6) est plus joli
```

   Voir la figure 1 pour un exemple de ce qu'on pouvait obtenir.

4. Écrire une fonction `moy_emp(X)` qui prend en entrée un vecteur (ligne ou colonne) `X` et qui renvoie sa moyenne empirique.
   Solution :

```
function m = moy_emp(X)
  % X est un vecteur
  m = sum(X)/numel(X);
end
```

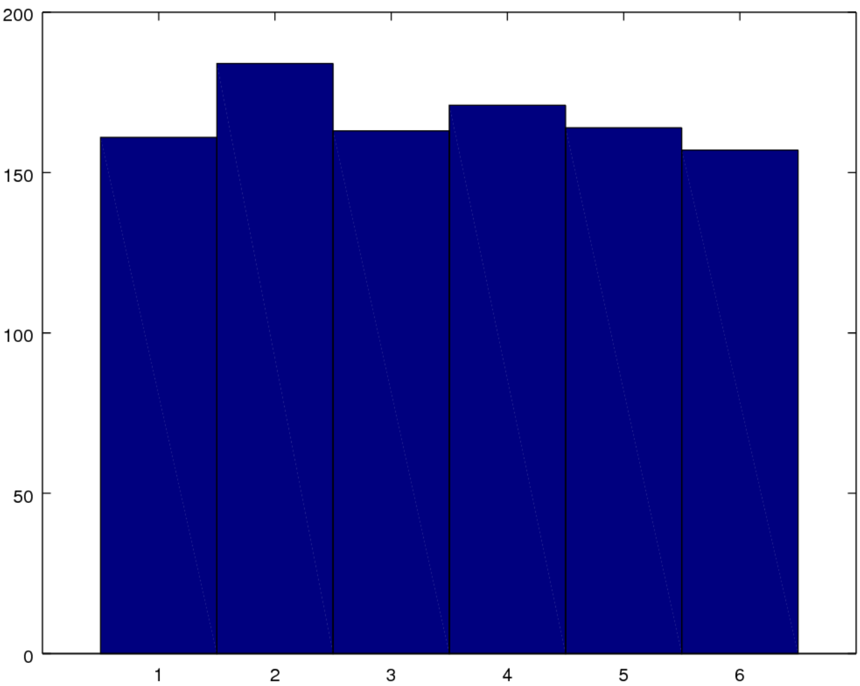

Figure 1 – Histogramme de 1000 lancers de dé

5. Calculer la moyenne empirique d'un échantillon de $N = 10000$ valeurs suivant la loi de D. ✐ Pouvait-on s'attendre à ce résultat ? Justifier la réponse.

   Solution : Pour afficher une moyenne empirique, on peut par exemple exécuter le script suivant :

```
N = 10000;
D = de6faces(N);
m = moy_emp(D);
printf("moyenne empirique de %d valeurs : %g\n", N, m);
```

   Cela donne un nombre « proche » de $3,5$ qui est l'espérance de D. Ceci est tout à fait cohérent avec la loi des grands nombres qui dit précisément que, si $D_1, D_2, \dots, D_n$ sont n tirages indépendants ayant la même loi que D, on a

$$\frac{1}{N}\left(D_1 + D_2 + \cdots + D_n\right) \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}\left[D\right] = \frac{7}{2}$$

Partie B : un seul dé, beaucoup de moyennes

Soit n un entier naturel non nul. On note D_n la variable aléatoire égale à la moyenne empirique de n copies indépendantes de D.

1. Écrire un programme `de6moyennes(n, N)` qui renvoie un vecteur ligne de taille N, dont les composantes sont des variables aléatoires indépendantes de même loi que D_n.

   Solution :

```
function moys = de6moyennes(n,N)
% renvoie une ligne de taille N
```

```
% en faisant la moyenne de n valeurs
  D = floor(6*rand(n,N)) + 1;
  moys = sum(D) / n;
end
```

2. Tracer l'histogramme d'un échantillon de $N = 1000$ valeurs indépendantes suivant la loi de D_{100}.
   Solution : Pour tracer un histogramme avec $K = 30$ classes par exemple, on peut utiliser le script suivant :

```
N = 1000;
n = 100;
moyennes = de6moyennes(n,N);
K = 30;
hist(moyennes, K);
```

3. Quelle est l'allure de cet histogramme ? ✐ Quelle est la « loi approchée » de D_{100} ? Justifier.
   Solution : La figure 2 montre un exemple de ce qu'on pouvait obtenir. Cela illustre le théorème central limite qui nous dit que la loi de D_{100} est approchée par la loi normale $\mathcal{N}\left(\mu, \left(\frac{\sigma}{\sqrt{100}}\right)^2\right)$, où μ et σ sont l'espérance et l'écart-type de D, calculés à la première question.

Partie C : un seul dé, qu'on relance parfois

On considère le jeu (récursif) suivant :

— le joueur commence par lancer un dé ;

— si le résultat d'un lancer est 6, le joueur a un nouveau lancer, sinon il doit arrêter ;

— le score du joueur est la somme de tous les résultats des dés qu'il a lancés.

1. ✐ Est-il possible d'avoir un score de 13 avec ce jeu ? Si oui, avec quelle probabilité ?
   Solution : Notons S la variable aléatoire égale au score du joueur et D_1, D_2, ..., les résultats des lancers de dé successifs. Alors,

$$\mathbb{P}(S = 13) = \mathbb{P}(D_1 = 6, D_2 = 6, D_3 = 1) = \left(\frac{1}{6}\right)^3.$$

2. Écrire une fonction `jeu_de_des(N)` qui renvoie un échantillon de N valeurs indépendantes qui simulent le score d'un joueur de ce jeu de dés.
   Solution : Il n'y avait pas tellement d'autre solution que d'utiliser une boucle `while`.

```
function S = jeu_de_des(N)
  S = zeros(1,N);
  for ii = 1:N
    de = de6faces(1);
    score = de;
    while de == 6
      de = de6faces(1);
      score = score + de;
```

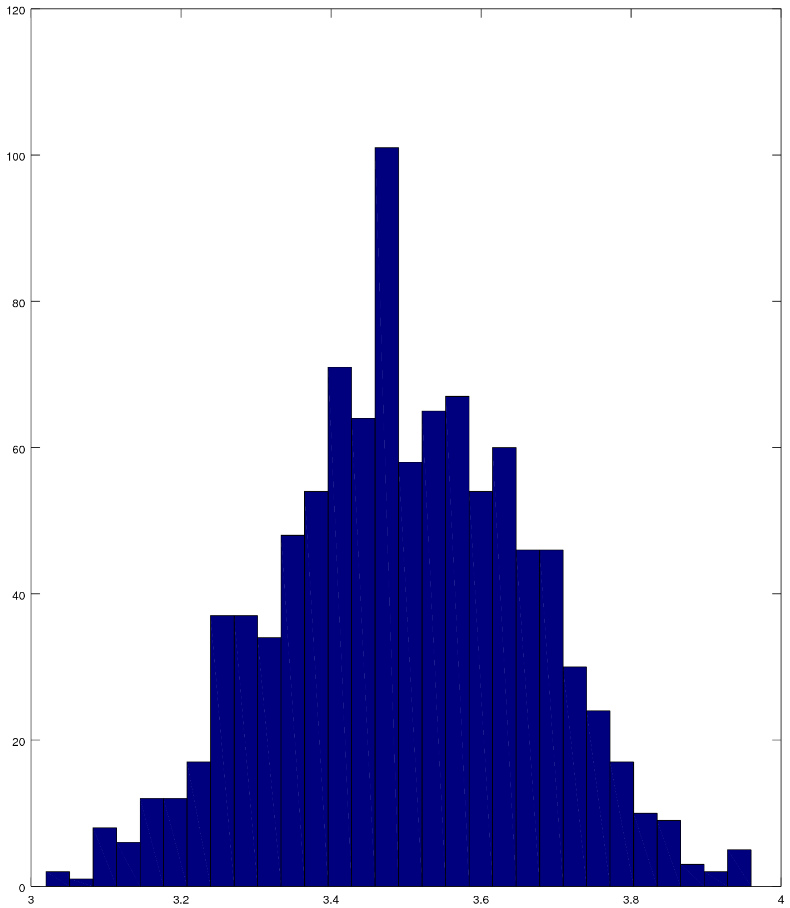

FIGURE 2 – Illustration du théorème central limite

```
        end
        S(ii) = score;
    end
end
```

3. Tracer l'histogramme d'un échantillon pour $N = 10000$. Reproduire l'allure de cet histogramme sur la copie.
   Solution : Pour tracer l'histogramme, on peut utiliser le script suivant :

```
N = 10000;
S = jeu_de_des(N);
% pour rendre l'histogramme plus joli :
m = max(S);
hist(S, 1:m);
```

Suivant la machine utilisée, le tirage de S peut être assez long car on ne peut pas utiliser ici de solution vectorisée. L'histogramme obtenu est visible à la figure 3.

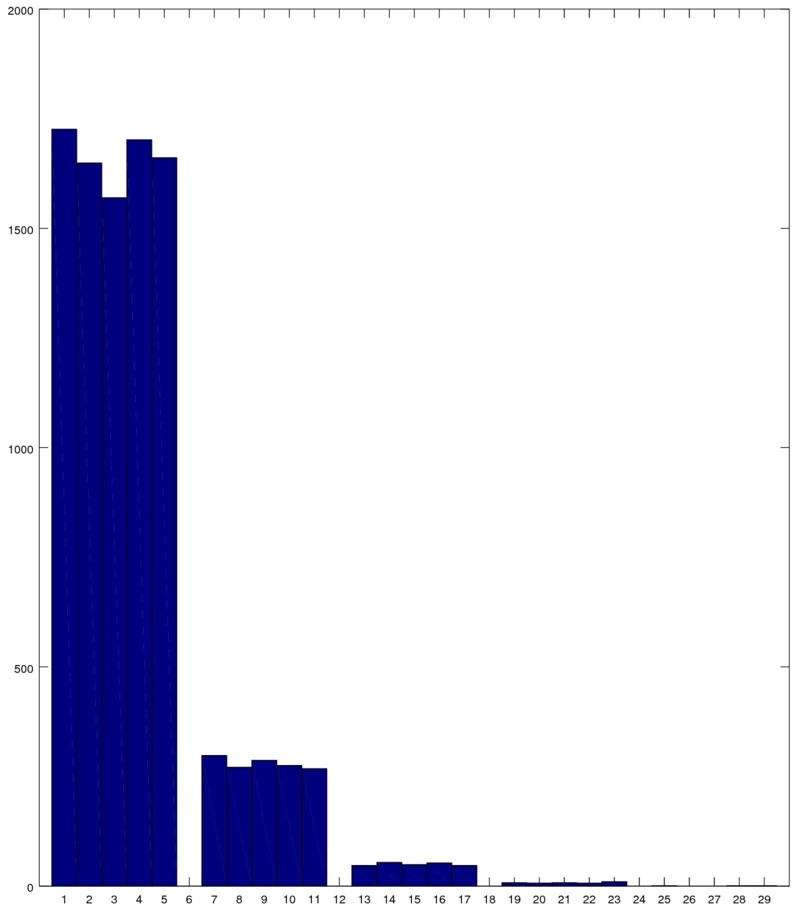

FIGURE 3 – Histogramme correspondant à 10000 réalisations du jeu de dés

4. Calculer la fréquence empirique de la valeur 13, toujours pour $N = 10000$ et comparer avec la réponse donnée à la question 1.

   Solution : On calcule la fréquence de la valeur 13 avec la commande `freq13 =sum( S == 13 )/ N`. Le résultat est proche de $(1/6)^3$.

5. Estimer, à l'aide de la simulation, l'espérance du score d'un joueur et la probabilité que le score soit supérieur à 10.

   Solution : Pour estimer l'espérance du score, on utilise la moyenne empirique (`sum(S)/ N`) et pour estimer la probabilité que le score soit supérieur ou égal à 10, on utilise la commande `sum((S >=10)/ N`.

Exercice 2

On considère la fonction f définie par :

$$f(x) = \begin{cases} 3x^2 & \text{si } x \in [0\,;1], \\ 0 & \text{sinon.} \end{cases}$$

1. ✐ Justifier que f est une fonction de densité de probabilité.
   Solution : Il est clair que la fonction f est positive. Il reste à montrer que son intégrale vaut 1 :
   $$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_0^1 3x^2\,\mathrm{d}x = \left[x^3\right]_0^1 = 1^3 - 0^3 = 1.$$

   On note X une variable aléatoire de densité f.

2. ✐ Justifier que la fonction F définie par
   $$F(x) = \begin{cases} x^3 & \text{si } x \in [0\,;1], \\ 0 & \text{si } x < 0, \\ 1 & \text{si } x > 1, \end{cases}$$
   est la fonction de répartition de X.
   Solution : On a bien $\lim_{x\to-\infty} F(x) = 0$. Il reste à montrer que F est une primitive de f, ce qui est le cas car $F'(x) = 3x^2$ pour tout x dans $[0\,;1]$ et $F'(x) = 0$ si x est en-dehors de $[0\,;1]$.

3. Simulation par la méthode de l'inverse de la fonction de répartition.
   (a) ✐ Soit U une variable aléatoire de loi uniforme sur $[0\,;1]$. Justifier que la variable aléatoire $Y = \sqrt[3]{U}$ a la même loi que X.
   Solution : Il suffit de montrer que Y et X ont la même fonction de répartition. Soit $y \in [0\,;1]$.
   $$\mathbb{P}\left(Y \le y\right) = \mathbb{P}\left(\sqrt[3]{U} \le y\right) = \mathbb{P}\left(U \le y^3\right) = y^3 = F(y).$$

   (b) À l'aide de la question précédente, écrire une fonction `inv_rep(N)` qui renvoie un vecteur ligne de taille N dont les composantes sont des copies indépendantes de X.
   Solution :

```
function X = inv_rep(N)
  X = rand(1, N) .^ (1 / 3);
end
```

   (c) Tracer l'histogramme d'un échantillon de taille $N = 1000$.
   Solution : On trace l'histogramme avec, par exemple, le script suivant.

```
N = 1000;
X = inv_rep(N);
K = 50;
hist(X, K);
```

   Un exemple de ce qu'on pouvait obtenir est visible sur la figure 4.

4. Simulation par la méthode du rejet.
   (a) ✐ Expliquer brièvement comment simuler la loi de X avec la méthode du rejet.
   Solution : On doit trouver une fonction de densité g d'une loi connue et facilement simulable, et une constante c la moins grande possible telle que pour tout $x \in [0\,;1]$, $f(x) \le cg(x)$. On va utiliser ici la loi uniforme sur $[0\,;1]$ (donc ici g est constante et égale à 1) et $c = 3$.

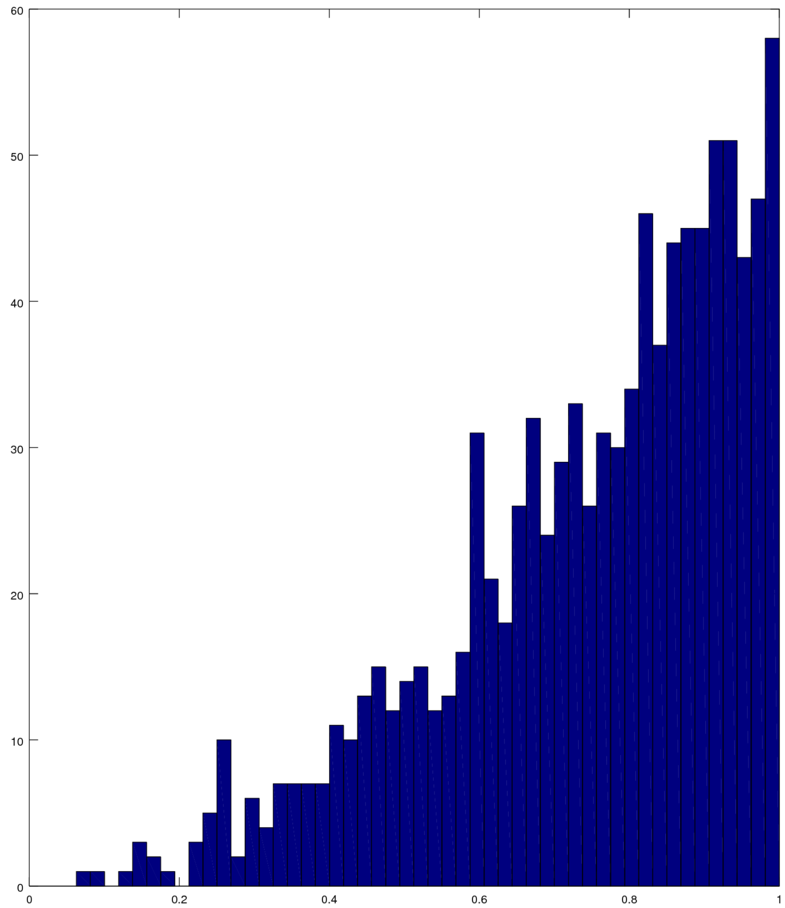

FIGURE 4 – Histogramme correspondant à un échantillon de taille 1000 obtenu par la méthode de l'inverse de la fonction de répartition

Pour simuler X, on commence par tirer V suivant la loi uniforme sur $[0\,;1]$ avec `rand`, puis U une autre variable uniforme sur $[0\,;1]$ indépendante de V. On accepte V comme un tirage de X si

$$U \leq \frac{f\left(V\right)}{cg\left(V\right)} = \frac{3V^2}{3 \times 1} = V^2.$$

(b) Écrire un programme `rejet(N)` qui simule par la méthode du rejet un échantillon de N copies indépendantes de X.

Solution :

```
function X = rejet(N)
  X = zeros(1, N);
  for ii = 1:N
    while true
      essai = rand();
      valeur_test = rand();
      if valeur_test <= essai^2
```

```
        break;
      end
    end
    X(ii) = essai;
  end
end
```

(c) ✐ Pour N grand, combien d'appels à la fonction `rand` votre programme fait-il en moyenne ?

Solution : Le nombre d'itération dans la boucle `while` est aléatoire, de loi géométrique de paramètre $\frac{1}{c}$ donc son espérance est $c = 3$. Si l'on demande un nombre N important de valeurs, d'après la loi des grands nombres, le nombre moyen d'itérations dans la boucle sera proche de 3. Comme chaque boucle `while` fait appel deux fois à la fonction `rand`, cela fait en moyenne 6 appels à rand par valeur renvoyée.

Exercice 3 Méthode de Monte-Carlo **6 points**

À l'aide de scripts ou de fonctions, en utilisant la méthode de Monte-Carlo, donner une valeur approchée des intégrales suivantes :

$$I = \int_0^2 \sqrt{x^4+3}\,\mathrm{d}x,$$
$$J = \int_{-\infty}^{+\infty} \sqrt{x^4+3}e^{-\frac{x^2}{2}}\,\mathrm{d}x.$$

Solution : La première intégrale est égale à

$$I = \int_0^2 \sqrt{x^4+3}\,\mathrm{d}x = \int_0^2 \sqrt{x^4+3} \times 2 \times \frac{1}{2}\,\mathrm{d}x = \mathbb{E}\left[f(U)\right],$$

où U suit la loi uniforme sur $[0\,;2]$ et $f : x \mapsto 2\sqrt{x^4+3}$. On approche cette espérance en utilisant la loi des grands nombres, ainsi

$$I = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(U_i),$$

où U_1, U_2, ... sont des copies indépendantes de la variable U. Cela donne lieu à la fonction suivante :

```
function I = monte_carlo_1(N)
  U = 2 * rand(1,N);
  I = sum( 2 * sqrt( U .^ 4 + 3 ) ) / N;
end
```

On peut comparer le résultat à la valeur numérique calculée par Octave en utilisant le script suivant :

```
f = @(x) sqrt( x.^4 + 3);
quad(f, 0, 2)
```

Pour la deuxième intégrale, on utilise cette fois la loi normale :

$$J = \int_{-\infty}^{+\infty} \sqrt{2\pi}\sqrt{x^4+3}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x = \mathbb{E}\left[g(N)\right],$$

où N suit la loi normale centrée réduite et $g : x \mapsto \sqrt{2\pi\,(x^4+3)}$. La fonction suivante permet d'approcher l'intégrale J :

```
function J = monte_carlo_2(N)
  U = randn(1,N);
  J = sum( sqrt( 2 * pi * ( U.^4 + 3 ) ) ) / N;
end
```

Le résultat devrait être proche de celui donné par le script suivant :

```
g = @(x) sqrt( x.^4 + 3) * exp(- (x.^2 / 2) )
quad(g, -Inf , Inf)
```