SOLUTION TP n° 6

Solution 1. Les commandes `sample(c(0, 0, 1), 15, replace = T)` effectuent 15 tirages avec remise d'un chiffre parmi ceux du vecteur $(0, 0, 1)$. La commande `sum` calcule la somme des chiffres obtenus, donc le nombre de 1 obtenus. C'est donc une réalisation de la somme de $n = 15$ *var iid* suivant chacune la loi de Bernoulli $\mathcal{B}(p)$ avec $p = 1/3$ (il y a 1 chance sur 3 d'avoir le 1 à chaque tirage), laquelle suit la loi binomiale $\mathcal{B}(n, p)$. Les commandes équivalentes sont `rbinom(1, 15, 1/3)`.

Solution 2. On rappelle l'inégalité de Bienaymé-Tchebychev: Soit X une *var* admettant un moment d'ordre 2. Alors, pour tout $\epsilon > 0$, on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\mathbb{V}(X)}{\epsilon^2}.$$

L'objectif de cet exercice est d'illustrer cette inégalité avec un exemple.

Soit X une *var* suivant la loi de Poisson $\mathcal{P}(3)$.

1. Donner, sans démonstration, $\mathbb{E}(X)$ et $\mathbb{V}(X)$:

   On a $\mathbb{E}(X) = 3$ et $\mathbb{V}(X) = 3$.

2. Simuler un échantillon de taille 1000 de X :

   ```
   x = rpois(1000, 3)
   ```

3. Une approximation de $\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon)$ est donnée par la proportion $R(\epsilon)$ de valeurs simulées vérifiant $|X - \mathbb{E}(X)| \geq \epsilon$. Évaluer celle-ci avec $\epsilon = 4$:

   ```
   d = abs(x - 3)
   sum(d >= 4) / 1000
   ```

   Puis comparer avec $\mathbb{V}(X)/\epsilon^2$:

   `3 / 4^2` renvoie : `[1] 0.1875`

   On remarque que $\mathbb{V}(X)/\epsilon^2$ est supérieur à la proportion $R(\epsilon)$, ce qui est cohérent avec l'inégalité de Bienaymé-Tchebychev.

4. Reproduire et analyser les commandes suivantes :

   ```
   N = 1000
   x = rpois(N, 3)
   epsilon = 1:50
   proba = NULL
   for (i in 1:50) {
   d = abs(x - 3)
   proba[i] = sum(d >= epsilon[i]) / N
   }
   ```

```
sup = 3 / epsilon^2
plot(epsilon, sup, type = "l", ylim = c(0, 1))
lines(epsilon, proba, col = "red")
```

Le graphique obtenu compare la proportion de valeurs simulées vérifiant $|X - \mathbb{E}(X)| \geq \epsilon$, laquelle approche $\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon)$, et $\mathbb{V}(X)/\epsilon^2$ pour plusieurs valeurs de ϵ. On constate alors que la courbe représentant la première quantitée est toujours inférieure à la deuxième, illustrant ainsi l'inégalité de Bienaymé-Tchebychev.

Écrire une fonction $g(\epsilon) = \mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon)$, $\epsilon \in \mathbb{N}^*$, sans approximation. Comparer graphiquement $g(\epsilon)$ avec $\mathbb{V}(X)/\epsilon^2$ pour $\epsilon \in \{1, \ldots, 50\}$:

On peut écrire $g(\epsilon)$ en fonction de la fonction de répartition de X de commande `ppois` :

$$\begin{aligned}
g(\epsilon) &= \mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) = 1 - \mathbb{P}(|X - 3| < \epsilon) \\
&= 1 - \mathbb{P}(|X - 3| \leq \epsilon - 1) \\
&= 1 - \mathbb{P}(-(\epsilon - 1) \leq X - 3 \leq \epsilon - 1) \\
&= 1 - \mathbb{P}(3 - (\epsilon - 1) \leq X \leq 3 + (\epsilon - 1)) \\
&= 1 - (\mathbb{P}(X \leq 3 + (\epsilon - 1)) - \mathbb{P}(X < 3 - (\epsilon - 1))) \\
&= 1 - (\mathbb{P}(X \leq 3 + (\epsilon - 1)) - \mathbb{P}(X \leq 2 - (\epsilon - 1))) .
\end{aligned}$$

D'où la fonction :

```
g = function(epsilon) {
1 - (ppois(3 + epsilon - 1, 3) - ppois(2 - (epsilon - 1), 3))
}
```

Comparaison :

```
epsilon = 1:50
sup = 3 / epsilon^2
plot(epsilon, sup, type = "l", ylim = c(0, 1))
lines(epsilon, g(epsilon), col = "red")
```

Solution 3. Soit X une *var* suivant la loi binomiale $\mathcal{B}(15, 0.3)$.

1. Simuler un échantillon de taille 500 de X. On note $x_1, \ldots, x_{500}$ les valeurs simulées. Enregistrer les sous forme d'un vecteur de taille 500 :

   ```
   x = rbinom(500, 15, 0.3)
   ```

2. Représenter cet échantillon: tracer le nuage de points $\{(i, x_i);\ i \in \{1, \ldots, 500\}\}$:

   ```
   plot(1:500, x, pch = 19, cex = 0.3)
   ```

3. Calculer la fréquence des différentes observations $\{0, \ldots, 15\}$ en utilisant la fonction table :

   ```
   table(x)
   plot(table(x) / 500)
   ```

   On note $f_0, \ldots, f_{15}$ ces fréquences. Enregistrer les sous forme d'un vecteur :

```
f = NULL
f[1:(dim(table(x)))] = table(x) / 500
```

4. Représenter $f_0, \dots, f_{15}$ à l'aide d'un diagramme à bâtons :

```
barplot(f, names.arg = 0:(dim(table(x)) - 1))
```

5. Comparer la distribution de X et $f_0, \dots, f_{15}$:

```
par(mfrow = c(1,2))
barplot(dbinom(0:15, 15, 0.3), names.arg = 0:15, main =
"Loi binomiale B(15, 0.3)")
barplot(f, names.arg = 0:(dim(table(x)) - 1), main = "Fréquences")
```

6. Calculer et représenter graphiquement les fréquences cumulées (utiliser la fonction `cumsum`) :

```
fcumul = cumsum(f)
barplot(fcumul, names.arg = 0:(dim(table(x)) - 1),
main = "Fréquences cumulées")
```

7. Comparer le graphe de la fonction de répartition de X et la courbe des fréquences cumulées :

```
w = seq(-5, 20, 0.01)
plot(w, pbinom(w, 15, 0.3), pch = 19, cex = 0.3, xlim = c(-5, 20),
main = "Fonction de répartition")
points(0:15, cumsum(dbinom(0:15, 15, 0.3)), pch = 19)
points(0:15, fcumul, pch = 3, col = "red")
```

8. Calculer la moyenne empirique de l'échantillon simulé :

```
mean(x)
```

9. Comparer la moyenne empirique et l'espérance de la loi binomiale $\mathcal{B}(15, 0.3)$:

```
y = 0:15
```

`sum(y * dbinom(y, 15, 0.3))` renvoie : `[1] 4.5` $(= 15 \times 0.3)$, ce qui est proche de l'espérance empirique.

10. Calculer la variance empirique de l'échantillon simulé :

```
var(x)
```

11. Comparer la variance empirique et la variance de la loi binomiale :

`sum(y^2 * dbinom(y, 15, 0.3)) - (sum(y * dbinom(y, 15, 0.3)))^2` renvoie : `[1] 3.15`, ce qui est proche de la variance empirique.

Solution 4. Soient X et Y deux *var* indépendantes telles que X suit la loi de Poisson $\mathcal{P}(5)$ et Y suit la loi de Poisson $\mathcal{P}(3)$. Alors on sait que $Z = X + Y$ suit la loi de Poisson $\mathcal{P}(5 + 3)$.

1. Illustrer ce résultat en simulant des *var* et en utilisant la commande `qqplot` :

```
n = 1000
a = rpois(n, 5)
b = rpois(n, 3)
c = rpois(n, 8)
qqplot(a + b, c)
```

   On constate que le nuage de points peut être ajusté par la droite $y = x$, ce qui illustre le résultat.

   On peut constater que cela n'est pas le cas en faisant les mêmes commandes avec $a = rpois(n, 1)$ (par exemple).

   La commande `qqplot` (signifiant diagramme quantile-quantile) permet de comparer graphiquement les lois de probabilités d'où sont issus 2 échantillons pour savoir si elles sont identiques. Si les points sont alignés sur l'axe $y = x$, les 2 échantillons sont issus de loi de probabilités similaires, sinon elles ne le sont pas. Le terme de quantile-quantile provient du fait que l'on compare la position de certains quantiles dans la population observée avec leur position dans la population théorique.

2. Montrer directement ce résultat en utilisant la commande `convolve` (avec `type = "o"`) :

```
k = 0:25
a = dpois(k, 5)
b = dpois(k, 3)
c = convolve(a, rev(b), type = "o")
c
d = dpois(k, 8)
d
```

   On constate que les valeurs obtenues pour `c` et `d` sont les mêmes, d'où le résultat.

Solution 5. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de *var iid* de loi commune la loi de Bernoulli $\mathcal{B}(p)$. On pose $\overline{X}_n = (1/n)\sum_{i=1}^{n} X_i$. L'objectif de cet exercice est d'illustrer la convergence en probabilité de $(\overline{X}_n)_{n\in\mathbb{N}^*}$ vers p (justifiée par la loi faible des grands nombres). On fixe $p = 0.5$ et $n = 5$.

1. Quelle est la loi de $n\overline{X}_n$?

   Comme $n\overline{X}_n = \sum_{i=1}^{n} X_i$ est la somme de n *var iid* de loi commune la loi de Bernoulli $\mathcal{B}(p)$, $n\overline{X}_n$ suit la loi binomiale $\mathcal{B}(n, p)$.

2. Simuler un échantillon de taille 100 de $n\overline{X}_n$:

```
s = rbinom(100, 5, 0.5)
```

3. Déduire un échantillon de taille 100 de $\overline{X}_n$:

```
w = s / 5
```

4. Évaluer la proportion $R_n(\epsilon)$ de valeurs simulées qui vérifient $|\overline{X}_n - p| > \epsilon$ avec $\epsilon = 0.1$:

```
z = abs(w - 0.5)
sum(z > 0.1) / 100
```

5. Refaire les calculs de $R_n(\epsilon)$ avec $\epsilon = 0.1$ et $n \in \{10, 25, 50, 75, 100, 150, 200, 500\}$:

```
S = matrix(0, ncol = 8, nrow = 100)
n = c(10, 25, 50, 75, 100, 150, 200, 500)
for(i in 1:8) {  S[, i] = rbinom(100, n[i], 0.5) / n[i]  }
colnames(S) = paste("W", as.character(n), sep = " ")
rownames(S) = paste("Simul", as.character(1:100), sep = " ")
Z = abs(S - 0.5)
R_0.1 = c()
for(i in 1:8) { R_0.1[i] = length(Z[Z[, i] >= 0.1, i]) / 100 }
names(R_0.1) = colnames(S)
print(R_0.1)
```

6. Refaire les calculs de $R_n(\epsilon)$ avec $\epsilon = 0.01$ et $n \in \{10, 100, 200, 500, 1000, 5000, 10000, 50000\}$:

```
S = matrix(0, ncol = 8, nrow = 100)
n = c(10, 100, 200, 500, 1000, 5000, 10000, 50000)
for(i in 1:8) {  S[, i] = rbinom(100, n[i], 0.5) / n[i]  }
colnames(S) = paste("W", as.character(n), sep = " ")
rownames(S) = paste("Simul", as.character(1:100), sep = " ")
Z = abs(S - 0.5)
R_0.01 = c()
for(i in 1:8) { R_0.01[i] = length(Z[Z[, i] >= 0.01, i]) / 100 }
names(R_0.01) = colnames(S)
print(R_0.01)
```

7. Commenter les résultats obtenus: On constate que, plus n est grand, plus l'écart entre $\overline{X}_n$ et p est petit. Cela illustre la convergence en probabilité de $(\overline{X}_n)_{n\in\mathbb{N}^*}$ vers p garantie la loi faible des grands nombres.

Solution 6. Reproduire et analyser les commandes suivantes :

```
n = 5000
x = rbinom(n, 10, 0.21)
y = cumsum(x) / (1:n)
plot(y, type = "l")
abline(h = 2.1, col = "red")
```

Quel résultat célèbre est illustré ?

La courbe obtenue est constituée de 5000 réalisations de moyennes de *var iid* $(X_n)_{n\in\mathbb{N}^*}$ suivant chacune la loi binomiale $\mathcal{B}(n,p)$ avec $n=10$ et $p=0.21$. On constate que la courbe se stabilise à 2.1 à mesure que n croit. Cela illustre la loi faible des grands nombres : on pose

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Alors $(\overline{X}_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers $\mathbb{E}(X_1)$. Ici, on a $\mathbb{E}(X_1)=np=2.1$.