

SOLUTION TP n° 8

Solution 1. Donner une commande R équivalente à `sum(runif(10) < 0.3)` :

La commande `sum(runif(10) < 0.3)` somme le nombre d'événements $\{X_1 < 0.3\}, \dots, \{X_{10} < 0.3\}$ qui se réalisent, où X_1, \dots, X_{10} sont des *var iid* de loi commune la loi uniforme $\mathcal{U}([0, 1])$. C'est donc une réalisation de la somme de $n = 10$ *var iid* de loi commune la loi de Bernoulli $\mathcal{B}(p)$ avec

$$p = \mathbb{P}(X_1 < 0.3) = \int_0^{0.3} dx = 0.3,$$

laquelle suit la loi binomiale $\mathcal{B}(n, p)$. La commande équivalente est `rbinom(1, 10, 0.3)`.

Solution 2. Un échantillon de taille 50 d'une *var* X est présenté dans le tableau suivant :

0.498	2.409	0.577	5.579	2.507	1.695	2.396	1.528	0.365	0.388
3.971	0.984	1.157	12.283	0.052	3.939	9.631	3.727	2.684	0.513
2.540	1.509	1.542	4.771	1.478	4.784	3.266	6.677	3.034	2.426
2.595	0.472	4.074	4.530	2.192	0.418	1.237	1.929	0.729	0.127
6.429	2.072	7.844	0.584	1.769	0.838	1.590	1.554	5.853	4.342

1. Représenter ces données par un histogramme des fréquences :

```
x = c(0.498, ..., 4.342)
```

```
hist(x, prob = T)
```

2. Tracer, sur le graphique précédent, les 3 densités associées à la loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda \in \{0.33, 0.5, 3\}$:

```
curve(dexp(x, 0.33), 0, 15, col = "red", add = T)
```

```
curve(dexp(x, 0.5), 0, 15, col = "blue", add = T)
```

```
curve(dexp(x, 3), 0, 15, col = "green", add = T)
```

Quelle est la densité la plus adaptée ?

La densité la plus adaptée est celle associée à la loi exponentielle $\mathcal{E}(0.33)$.

3. Simuler un échantillon de taille 50 d'une *var* Y suivant la loi exponentielle $\mathcal{E}(0.33)$ et représenter le avec un histogramme des fréquences :

```
z = rexp(50, 0.33)
```

```
hist(z, prob = T)
```

Solution 3. Soient X_1 et X_2 deux *var* indépendantes. Illustrer les résultats ci-dessous avec la commande `qqplot`.

$X_i \sim$	$\mathcal{E}(\lambda)$	$\Gamma(m_i, \lambda)$	$\mathcal{N}(\mu_i, \sigma_i^2)$	$\chi^2(\nu_i)$
$X_1 + X_2 \sim$	$\Gamma(2, \lambda)$	$\Gamma(m_1 + m_2, \lambda)$	$\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$	$\chi^2(\nu_1 + \nu_2)$
Numérique	$\lambda = 3.8$	$m_1 = m_2 = 4.2, \lambda = 2.1$	$\mu_1 = \mu_2 = 1.6, \sigma_1 = \sigma_2 = 1.5$	$\nu_1 = \nu_2 = 3.2$

- $X_1 \sim \mathcal{E}(\lambda)$, $X_2 \sim \mathcal{E}(\lambda)$, X_1 et X_2 indépendantes entraînent $X_1 + X_2 \sim \Gamma(\lambda, 2)$, avec $\lambda = 3.8$:

```
a = rexp(1000, 3.8)
b = rexp(1000, 3.8)
c = rgamma(1000, 2, 3.8)
qqplot(a + b, c)
```

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

- $X_1 \sim \Gamma(m_1, \lambda)$, $X_2 \sim \Gamma(m_2, \lambda)$, X_1 et X_2 indépendantes entraînent $X_1 + X_2 \sim \Gamma(m_1 + m_2, \lambda)$, avec $m_1 = m_2 = 4.2$ et $\lambda = 2.1$:

```
a = rgamma(1000, 4.2, 2.1)
b = rgamma(1000, 4.2, 2.1)
c = rgamma(1000, 8.4, 2.1)
qqplot(a + b, c)
```

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

- $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, X_1 et X_2 indépendantes entraînent $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, avec $\mu_1 = \mu_2 = 1.6$ et $\sigma_1 = \sigma_2 = 1.5$:

```
a = rnorm(1000, 1.6, 1.5)
b = rnorm(1000, 1.6, 1.5)
c = rnorm(1000, 3.2, sqrt(1.5^2 + 1.5^2))
qqplot(a + b, c)
```

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

- $X_1 \sim \chi^2(\nu_1)$, $X_2 \sim \chi^2(\nu_2)$, X_1 et X_2 indépendantes entraînent $X_1 + X_2 \sim \chi^2(\nu_1 + \nu_2)$, avec $\nu_1 = \nu_2 = 3.2$:

```
a = rchisq(1000, 3.2)
b = rchisq(1000, 3.2)
c = rchisq(1000, 6.4)
qqplot(a + b, c)
```

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

Solution 4. Soient X et Y 2 *var* indépendantes. Illustrer les résultats ci-dessous avec la commande `qqplot`.

- Caractérisation de la loi du chi-deux $\chi^2(2)$: Si $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$, alors

$$X^2 + Y^2 \sim \chi^2(2) :$$

```
a = rnorm(1000)
b = rnorm(1000)
c = rchisq(1000, 2)
qqplot(a^2 + b^2, c)
```

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

- Caractérisation de la loi de Student $\mathcal{T}(\nu)$: Si $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(\nu)$, alors

$$\frac{X}{\sqrt{\frac{Y}{\nu}}} \sim \mathcal{T}(\nu).$$

Prendre $\nu = 3.9$:

```
a = rnorm(1000)
b = rchisq(1000, 3.9)
c = rt(1000, 3.9)
qqplot(a / sqrt(b / 3.9), c)
```

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

- Caractérisation de la loi de Fisher $\mathcal{F}(\nu_1, \nu_2)$: Si $X \sim \chi^2(\nu_1)$ et $Y \sim \chi^2(\nu_2)$, alors

$$\frac{\frac{X}{\nu_1}}{\frac{Y}{\nu_2}} = \frac{\nu_2 X}{\nu_1 Y} \sim \mathcal{F}(\nu_1, \nu_2).$$

Prendre $(\nu_1, \nu_2) = (2.1, 8.3)$:

```
a = rchisq(1000, 2.1)
b = rchisq(1000, 8.3)
c = rf(1000, 2.1, 8.3)
qqplot((a / 2.1) / (b / 8.3), c)
```

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

Solution 5.

1. Reproduire et comprendre l'enjeu des commandes suivantes : $n = 5000$

```
par(mfrow = c(2, 1))
x = rnorm(n)
```

```

y = cumsum(x) / (1:n)
plot(y, type = "l")
abline(h = 0, col = "red")
xc = rcauchy(n, 0, 1)
yc = cumsum(xc) / (1:n)
plot(yc, type = "l")
abline(h = 0, col = "red")

```

2. Quel résultat célèbre est illustré dans le premier graphique ?

Par le biais de données simulées, on obtient une courbe constituée de 5000 réalisations de moyennes de *var iid* $(X_n)_{n \in \mathbb{N}^*}$ suivant chacune la loi normale $\mathcal{N}(0, 1)$. On constate que la courbe se stabilise à 0 à mesure que n croît. Cela illustre la loi faible des grands nombres qui nous assure que $(\bar{X}_n)_{n \in \mathbb{N}^*}$, avec

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

converge en probabilité vers $\mathbb{E}(X_1) = 0$.

3. Comment expliquer le résultat du deuxième graphique ?

La seule différence réside dans l'utilisation de valeurs simulées associées à la loi de Cauchy $\mathcal{C}(0, 1)$. Contrairement au premier graphique, la courbe ne stabilise pas à mesure que n augmente. Cela s'explique par le fait qu'une *var* X suivant la loi de Cauchy $\mathcal{C}(0, 1)$, *i.e.* de densité

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

n'admet pas d'espérance. Donc la loi faible des grands nombres ne s'applique pas et la suite de *var* $(\bar{X}_n)_{n \in \mathbb{N}^*}$ diverge.

Solution 6. Soient $(X_n)_{n \in \mathbb{N}^*}$ une suite de *var iid* de loi commune la loi uniforme $\mathcal{U}([0, 1])$, *i.e.* de densité

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

1. Pour tout $n \in \mathbb{N}^*$, on pose $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$. Expliquer pourquoi on a l'approximation suivante :

$$\bar{X}_n \approx \mathcal{N}\left(0.5, \frac{1}{12n}\right),$$

pour n suffisamment grand:

C'est une conséquence du théorème de central limite : Pour tout $n \in \mathbb{N}^*$, on pose

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad Z_n = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sigma(\bar{X}_n)} = \sqrt{n} \left(\frac{\bar{X}_n - \mathbb{E}(X_1)}{\sigma(X_1)} \right).$$

Alors la suite de $\text{var}(Z_n)_{n \in \mathbb{N}^*}$ converge en loi vers une $\text{var } Z \sim \mathcal{N}(0, 1)$. Comme

$$\mathbb{E}(X_1) = \frac{1}{2}, \quad \mathbb{V}(X_1) = \frac{1}{12},$$

pour n grand, cela justifie l'approximation:

$$\bar{X}_n \approx \mathcal{N}\left(0.5, \frac{1}{12n}\right).$$

2. Reproduire et analyser les commandes suivantes :

```
X = matrix(runif(500 * 10), ncol = 10, nrow = 500)
x = apply(X, 1, mean)
par(mfrow = c(1, 2))
hist(x, main = "Taille d'échantillon 10", prob = T)
curve(dnorm(x, 0.5, sqrt(1 / 120)), add = T)
Y = matrix(runif(500 * 50), ncol = 50, nrow = 500)
x = apply(Y, 1, mean)
hist(x, main = "Taille d'échantillon 50", prob = T)
curve(dnorm(x, 0.5, sqrt(1 / 600)), add = T)
```

Par le biais de valeurs simulées, on montre que l'histogramme des fréquences de réalisations de \bar{X}_n épouse la densité de la loi normale $\mathcal{N}\left(0.5, \frac{1}{12n}\right)$ avec $n \in \{10, 50\}$. Le résultat visuel est meilleur avec $n = 50$. Cela illustre l'approximation $\bar{X}_n \approx \mathcal{N}\left(0.5, \frac{1}{12n}\right)$.