Exercices en R, feuille 2

Exercice 1. *Le paradoxe des anniversaires*

On suppose que toutes les années comptent 365 *jours et que toutes les dates de naissances sont équiprobables, c.a.d. qu'il n'existe pas de période dans l'année où il y a plus de naissances. On va calculer la probabilité pour que dans une classe de N élèves au moins deux élèves fêtent leurs anniversaires le même jour.*

1. *Dans une salle contenant deux personnes, quelle est la probabilité que les deux aient leur anniversaire le même jour ?*
2. *Dans une salle contenant trois personnes, quelle est la probabilité qu'au moins deux personnes aient leur anniversaire le même jour ? Généraliser le résultat pour une salle contenant N personnes.*
3. *Créer une fonction* `anniv` *qui à tout entier N tel que $N \geq 2$ associe la probabilité que dans une classe de N élèves au moins deux élèves fêtent leur anniversaire le même jour.*
4. *Trouver la plus petite valeur de N pour laquelle cette probabilité dépasse* 0.5.
5. *Tracer le graphe de la fonction* `anniv` *pour N de* 2 *à* 400.
6. *Martin est un élève de cette classe de N élèves. Quelle est la probabilité qu'il y ait au moins un(e) autre élève ayant son anniversaire le même jour que Martin ?*
7. *Créer une fonction* `anniv2` *qui à tout entier N tel que $N \geq 2$ associe la probabilité trouvée à la question précédente. Superposer le graphe de cette fonction à celui de* `anniv`*. Commenter.*

Exercice 2. *Loi uniforme continue :*

```
dunif(x, min = 0, max = 1)
punif(q, min = 0, max = 1)
qunif(p, min = 0, max = 1)
runif(n, min = 0, max = 1)
```

1. *La fonction* `set.seed()` *permet de spécifier l'amorce du générateur de nombre aléatoire, ce qui se relève utile quand on veut répéter une simulation à l'identique. Exécuter les commandes suivantes plusieurs fois. Qu'observez vous ?*

   ```
   runif(1); set.seed(15); runif(1)
   ```

2. *Pourquoi les commandes suivantes nous retournent toujours* `TRUE` *?*

   ```
   u = runif(20); punif(u) == u; dunif(u) == 1
   ```

3. *Pouvez vous prédire la valeur de* `var(runif(10000))` *?*

Exercice 3. *Loi normale :*

```
dnorm(x, mean = 0, sd = 1)
pnorm(q, mean = 0, sd = 1)
qnorm(p, mean = 0, sd = 1)
rnorm(n, mean = 0, sd = 1)
```

1. *Que font ces lignes de commande ? Indiquer ce que* `R` *vous retourne.*

```
qnorm(0.975);
dnorm(0);
pnorm(1.96);
rnorm(20);
rnorm(10, mean = 5, sd = 0.5);
x = seq(-3, 3, 0.1); pdf = dnorm(x); plot(x, pdf, type = "l");
```

2. *Simuler le tirage de* 1000 *v.a. indépendantes de loi* $\mathcal{N}(0,1)$.
3. *Tracer un histogramme à* 50 *classes correspondant aux résultats de ces tirages.*
4. *Sur le même graphique, tracer la densité de la loi* $\mathcal{N}(0,1)$ *et comparer les deux figures. Commenter.*

Exercice 4. *Loi exponentielle :*

```
dexp(x, rate = 1)
pexp(q, rate = 1)
qexp(p, rate = 1)
rexp(n, rate = 1)
```

1. *Simuler le tirage de* 1000 *v.a. indépendantes de loi* $\mathcal{E}(3)$.
2. *Tracer un histogramme à* 50 *classes correspondant aux résultats de ces tirages.*
3. *Sur le même graphique, tracer la densité de la loi* $\mathcal{E}(3)$ *et comparer les deux figures. Commenter.*

Exercice 5. *Approximation d'une loi de Poisson par une loi normale*

Lorsque λ *est grand, la loi de Poisson de paramètre* λ *peut être approchée par une loi normale de moyenne* λ *et d'écart-type* $\sqrt{\lambda}$.

Par simulation, étudier graphiquement la validité du résultat précédent. Vous pourrez, par exemple, simuler 1000 *réalisations indépendantes d'une loi de Poisson de paramètre* $\lambda = 3, 6, 10, 100$ *et étudier l'évolution de l'histogramme.*

1. *Simuler le tirage de* 1000 *v.a. indépendantes de loi* $\mathcal{P}(3)$.

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

2. *Tracer les histogrammes ayant le titre* `lambda = 3`.

```
hist(x, probability = T,
main = paste("lambda = ", as.character(lambda), sep = ""))
```

3. *Superposer la densité de loi* $\mathcal{N}(3, \sqrt{3})$ *aux histogrammes obtenus.*

```
y = seq(0, max(x), 0.1)
lines(y, dnorm(y, lambda, sqrt(lambda)), col = "red", lwd = 2)
```

4. *Tracer trois autres graphes en répétant les questions 1-3 pour* $\lambda = 6, 10, 100$. *Utiliser* `par(mfrow = c(2,2))` *pour mettre ces quatre graphes dans la même fenêtre.*

Exercice 6. *Illustration de la Loi des grands nombres*

Il s'agît d'illustrer la loi forte des grands nombres : si $X_1, X_2, \ldots, X_n$ *est une suite de v.a. i.i.d. avec* $\mathbb{E}[X_1] < \infty$, *alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} \mathbb{E}[X_1].$$

1. *On tire* n *fois une pièce de monnaie biaisée telle que la probabilité d'obtenir face est* 0.3 *et on calcule ensuite la proportion des faces obtenues.*

```
echantillon = sample(0:1, 1000, replace = T, prob = c(7, 3))

plot(cumsum(echantillon)/(1:1000), type = "l",
main = "Loi des grands nombres", ylim = c(0, 0.7),
xlab = "Taille échantillon")

abline(h = .3, col = 2)

legend(x = "top", lty = 1, col = 1:2,
legend = c("fréquence empirique", "probabilité théorique"), bty="n")
```

2. *Pour obtenir le nombre des faces en fonction de* n *on a utilisé la fonction* `cumsum()`. *Que font les commandes suivantes ?*

```
cumsum(c(0,0,1,0,1,1,0)); cumsum(c(0,0,1,0,1,1,0))/1:7
```

3. *A la place de* `sample()` *pouvez vous utiliser* `rbinom()` *pour simuler les v.a. de loi de Bernoulli ?*
4. *Générer une réalisation* $x_1, \ldots, x_n$ *de la suite* $X_1, \ldots, X_n$ *avec* $n = 1000$ *dans le cas d'une loi normale* $X_1 \sim \mathcal{N}(0, 1)$.
5. *Calculer* $\bar{x}_m = \frac{1}{m}\sum_{i=1}^{m} x_i$ *pour* $m = 1, \ldots, n$.
6. *Tracer* $\bar{x}_m$ *en fonction de* m. *Conclusion ?*
7. *Procéder de même avec la loi de Cauchy. Conclusion ?*

Exercice 7. *Le théorème de la limite centrée*

Il s'agît d'illustrer maintenant le théorème central limite : si $X_1, X_2, \ldots, X_n$ *est une suite de v.a. i.i.d. avec* $\mathbb{E}[X_1] < \infty$ *et* $\mathbb{E}[X_1^2] < \infty$, *alors*

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]}{\sqrt{Var[X_1]}} \xrightarrow{L} \mathcal{N}(0, 1)$$

1. *Générer R réalisations $x_{1,r}, \dots, x_{n,r}$ de la suite $X_1, \dots, X_n$ dans le cas d'une loi gamma $X_1 \sim \mathcal{G}(1,1)$, $(r = 1, \dots, R)$. On a donc $\mathbb{E}[X_1] = 1$ et $Var[X_1] = 1$. On prendra $n = 1000$ et $R = 500$.*
2. *Calculer $\bar{x}_{n,r} = \frac{1}{n} \sum_{i=1}^{n} x_{i,r}$ pour chaque r.*
3. *Tracer (1) un histogramme de $\sqrt{n}(\bar{x}_{n,r} - 1)$ sur le même graphique que la fonction de densité de la loi $\mathcal{N}(0,1)$ et (2) la fonction de répartition empirique de cette suite sur le même graphique que la fonction de répartition de la loi $\mathcal{N}(0,1)$. Essayer avec d'autres valeurs de n. Conclusion ?*
4. *Procéder de même avec la loi binomiale $\mathcal{B}(10, 0.8)$.*