SOLUTION TP n° 4

Solution 1. Certaines lois de *var* peuvent être approchées par d'autres lois plus simples à calculer. Notamment, la loi hypergéométrique $\mathcal{H}(l, m, n)$ peut être approchée par la loi binomiale $\mathcal{B}(n, p)$ avec $p = l/(l + m)$ et $n \leq 0.1(l + m)$, *i.e.*

$$\mathcal{H}(l, m, n) \approx \mathcal{B}(n, p), \qquad p = l/(l + m), \qquad n \leq 0.1(l + m).$$

L'objectif de cet exercice est de visualiser cette approximation avec le logiciel R.

1. Calculer les probabilités $\mathbb{P}(X = k)$ avec $k \in \{0, \ldots, 20\}$ lorsque X est une *var*

   - suivant la loi hypergéométrique $\mathcal{H}(1000, 500, 20)$,
   - suivant la loi binomiale $\mathcal{B}(20, 1000/1500)$.

   ```
   bin = dbinom(0:20, 20, 1000/1500)
   hyp = dhyper(0:20, 1000, 500, 20)
   ```

   Donner un exemple de schéma d'urnes dans lesquels ces 2 lois interviennent :

   Une urne contient 1500 boules, dont 1000 blanches et 500 noires. On extrait au hasard 20 boules de l'urne. Alors

   - la *var* X égale au nombre de boules blanches obtenues en faisant des tirages sans remise suit la loi hypergéométrique $\mathcal{H}(1000, 500, 20)$:

   $$\mathbb{P}(X = k) = \frac{\binom{1000}{k}\binom{500}{20-k}}{\binom{1500}{20}}, \qquad k \in \{0, \ldots, 20\},$$

   - la *var* X égale au nombre de boules blanches obtenues en faisant des tirages avec remise suit la loi binomiale $\mathcal{B}(20, 1000/1500)$:

   $$\mathbb{P}(X = k) = \binom{20}{k} p^k (1 - p)^{20-k}, \qquad k \in \{0, \ldots, 20\},$$

   avec $p = 1000/1500$.

2. Séparer l'écran graphique en 2 avec la commande `par(mfrow = c(2, 1))` : dans la fenêtre 1, représenter le graphe de la loi $\mathcal{H}(1000, 500, 20)$ et dans la fenêtre 2, celui de la loi $\mathcal{B}(20, 1000/1500)$:

   ```
   par(mfrow = c(2, 1))
   plot(k, bin, type = "h", main = "Binomiale", xlab = "k", ylab = "P(X=k)")
   plot(k, hyp, type = "h", main = "Hypergéométrique", xlab = "k",
   ylab = "P(X=k)")
   ```

3. Reproduire et analyser les commandes suivantes:

```
approx.fun = function(k, Ne, Nm, n, p) {
barplot(rbind(dhyper(k, Ne, Nm, n), dbinom(k, n, p)), col = c(1, 2),
legend = c("Hypergéométrique", "Binomiale"), beside = T, names = k,
xlab = "k", ylab = "Proba P(X=k)")
abline(h = 0) }
par(mfrow = c(3 ,1))
approx.fun(0:20, 1000, 500, 20, 2/3)
```

   C'est une très bonne approximation,

```
approx.fun(0:10, 20, 30, 10, 2/5)
```

   C'est une mauvaise approximation,

```
approx.fun(0:10, 10, 15, 10, 2/5)
```

   C'est une très mauvaise approximation; l'approximation est bonne que si $n \leq 0.1(l+m)$, ce qui n'est pas le cas pour les 2 dernières approximations.

Solution 2. La loi binomiale $\mathcal{B}(n,p)$ peut être approchée par la loi de Poisson $\mathcal{P}(np)$ si $n \geq 31$ et $np \leq 10$, *i.e.*

$$\mathcal{B}(n,p) \approx \mathcal{P}(np), \qquad n \geq 31,\ np \leq 10.$$

L'objectif de cet exercice est d'étudier cette approximation avec le logiciel R.

1. Représenter les lois suivantes à l'aide de diagrammes à bâtons :

$$\mathcal{B}(10,0.5), \qquad \mathcal{B}(20,0.25), \qquad \mathcal{B}(50,0.1), \qquad \mathcal{B}(100,0.05), \qquad \mathcal{P}(5):$$

```
par(mfrow = c(2, 3))
barplot(dbinom(0:20, 10, 0.5), main = "Loi binomiale B(10, 0.5)")
barplot(dbinom(0:20, 20, 0.25), main = "Loi binomiale B(20, 0.25)")
barplot(dbinom(0:20, 50, 0.1), main = "Loi binomiale B(50, 0.1)")
barplot(dbinom(0:20, 100, 0.05), main = "Loi binomiale B(100, 0.05)")
barplot(dpois(0:20, 5), main = "Loi de Poisson P(5)")
```

2. Pour chaque $n \in \{5, 10, 20, 50, 100\}$, évaluer le maximum de l'erreur commise lorsque l'on approche la loi binomiale par la loi de Poisson, *i.e.* calculer

$$E_n = \max_{k\in\{0,\ldots,n\}} |\mathbb{P}(X=k) - \mathbb{P}(Y=k)|,$$

   où X est une *var* suivant la loi binomiale $\mathcal{B}(n, 5/n)$ et Y est une *var* suivant la loi de Poisson $\mathcal{P}(5)$:

   Cas $n = 5$:

   `max(abs(dbinom(0:5, 5, 1) - dpois(0:5, 5)))` renvoie : `[1] 0.8245326`

Cas $n = 10$:

`max(abs(dbinom(0:10, 10, 0.5) - dpois(0:10, 5)))` renvoie : `[1] 0.07062638`

Cas $n = 20$:

`max(abs(dbinom(0:20, 20, 0.25) - dpois(0:20, 5)))` renvoie : `[1] 0.02686378`

Cas $n = 50$:

`max(abs(dbinom(0:50, 50, 0.1) - dpois(0:50, 5)))` renvoie : `[1] 0.009457231`

Cas $n = 100$:

`max(abs(dbinom(0:100, 100, 0.05) - dpois(0:100, 5)))` renvoie : `[1] 0.004550458`

Commenter les résultats obtenus: on remarque que l'erreur diminue à mesure que n augmente.

3. *Application 1.* Suite à une campagne de vaccination contre le paludisme, on estime à 2% la proportion de personnes qui seraient pourtant atteintes de la maladie. Soit X la *var* égale au nombre de personnes malades dans un petit village de 100 habitants.

   (a) Déterminer la loi de X :
   X est égale au nombre de réalisations de l'événement A ="la personne est atteinte du paludisme" en $n = 100$ expériences indépendantes. Par conséquent, X suit la loi binomiale $\mathcal{B}(n, p)$, avec $n = 100$ et $p = \mathbb{P}(A) = 2/100 = 0.02$:

$$\mathbb{P}(X = k) = \binom{100}{k} 0.02^k (1 - 0.02)^{100-k}, \qquad k \in \{0, \dots, 100\}.$$

   (b) Calculer la probabilité de constater au moins une personne malade: on veut calculer $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0)$:

   `1 - dbinom(0, 100, 0.02)` renvoie : `[1] 0.8673804`

   (c) Calculer la probabilité de constater au plus 10 personnes malades: on veut calculer $\mathbb{P}(X \leq 10)$:

   `pbinom(10, 100, 0.02)` renvoie : `[1] 0.9999944`

   (d) Recalculer ces probabilités en utilisant une approximation de la loi de X par une loi de Poisson :
   Comme $n = 100 \geq 31$ et $np = 2 \leq 10$, la loi binomiale $\mathcal{B}(n, p)$ que suit X peut être approchée par une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = np = 2$. On obtient ainsi

   - la probabilité de constater plus d'une personne malade: on veut calculer $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0)$:
     `1 - dpois(0, 2)` renvoie : `[1] 0.8646647`
   - la probabilité de constater au plus 10 personnes malades: on veut calculer $\mathbb{P}(X \leq 10)$:
     `ppois(10, 2)` renvoie : `[1] 0.9999917`

   (e) Quelle est la qualité de l'approximation ?
   La qualité de l'approximation dépend de l'erreur qui est obtenue par :

   `max(abs(dbinom(0:100, 100, 0.02)-dpois(0:100, 2)))` renvoie :

   `[1] 0.002743345`

4. *Application 2.* Une entreprise fabrique, en grande quantité, des tiges métalliques cylindriques pour l'industrie. Leur longueur et leur diamètre sont exprimés en millimètres. Dans un lot, 3% des tiges ne sont pas conformes pour la longueur. On prélève au hasard 50 tiges de ce lot pour vérification. Le lot est suffisamment important pour que l'on puisse assimiler ce prélèvement à des tirages avec remise. Soit la *var* X qui, à tout prélèvement de 50 tiges, associe le nombre de tiges non conformes pour la longueur.

   (a) Déterminer la loi de X :
   On a $X(\Omega) = \{0, \ldots, 50\}$. Par l'énoncé, on assimile le prélèvement de 50 tiges à des tirages avec remise. Donc X est égale au nombre de réalisations de l'événement
   $A =$ "la tige est non conforme" en 50 expériences indépendantes. On en déduit que X suit la loi binomiale $\mathcal{B}(n, p)$ avec $n = 50$ et $p = \mathbb{P}(A) = 0.03$:

$$\mathbb{P}(X = k) = \binom{50}{k}(0.03)^k(1 - 0.03)^{50-k}, \qquad k \in \{0, \ldots, 50\}.$$

   (b) Calculer la probabilité qu'au plus deux tiges ne soient pas conformes pour la longueur : `pbinom(2, 50, 0.03)` renvoie : `[1] 0.8107981`

   (c) Refaire (b) en utilisant une approximation de la loi de X :
   Comme $n = 50 \geq 31$ et $np = 1.5 \leq 10$, la loi binomiale $\mathcal{B}(n, p)$ que suit X peut être approchée par une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = np = 1.5$. Donc

$$\mathbb{P}(X = k) \simeq e^{-1.5}\frac{(1.5)^k}{k!}, \qquad k \in \{0, \ldots, 50\}.$$

   Ainsi, $\mathbb{P}(X \leq 2)$ est presque égale à : `ppois(2, 1.5)` renvoie : `[1] 0.8088468`

   (d) Quelle est la qualité de l'approximation ?
   La qualité de l'approximation dépend de l'erreur qui est obtenue par :
   `max(abs(dbinom(0:50, 50, 0.03) - dpois(0:50, 1.5)))` renvoie :
   `[1] 0.005064785`

Solution 3. On pose

$$p_{i,j} = \frac{1}{(i+1)^{j+1}}, \qquad (i, j) \in (\mathbb{N}^*)^2.$$

On admet que $(p_{i,j})_{(i,j)\in(\mathbb{N}^*)^2}$ caractérise la loi de probabilité d'un couple de *var* discrètes (X, Y); pour tout $(i, j) \in (\mathbb{N}^*)^2$, on a $p_{i,j} \geq 0$ et $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{i,j} = 1$.

1. Montrer que

$$\sum_{i=1}^{1000}\sum_{j=1}^{1000} p_{i,j} \approx 0.999.$$

   On va créer une matrice nulle de dimension 1000×1000, y mettre les composantes $(p_{i,j})_{(i,j)\in\{1,\ldots,1000\}^2}$, puis on va sommer toutes ces composantes.

```
m = matrix(rep(0, 1000^2), 1000, 1000)
for (i in 1:1000){
for (j in 1:1000){
m[i, j] = 1 / (i + 1)^(j + 1) } }
sum(m)
```

Cela renvoie : `[1] 0.999001`

2. Représenter graphiquement $(p_{i,j})_{(i,j)\in\{1,\dots,5\}^2}$:

```
persp(1:5, 1:5, m[1:5, 1:5], phi = 20, expand = 0.45, col = rainbow(7))
```