
TP 3: Simulations de variables aléatoires discrètes

Exercice 3.1 — Autour des lois discrètes usuelles

la fonction préimplémentée `sample` prélève successivement un nombre donné d'éléments d'un ensemble de cardinal fini, avec ou sans remise.

Plus précisément : `sample(x, size, replace = FALSE, prob = NULL)` où :

- `x` : Décrit l'ensemble dans lequel on va échantillonner. Il s'agit soit d'un vecteur, soit d'un entier `k`. Dans ce dernier cas, l'ensemble considéré sera "tous les entiers allant de 1 à `k`".
- `size` : Un entier positif ou nul donnant le nombre d'élément à tirer. Par défaut, il s'agit du cardinal de l'ensemble décrit par `x`.
- `replace` : Un booléen : Avec (T) ou sans remise (F).
- `prob` : Un vecteur de poids de probabilités d'obtenir les éléments de l'ensemble à partir duquel on échantillonne.

1) Taper les commandes suivantes, que signifient-elles et qu'observe-t-on ?

```
> X <- 1:10
> sample(X, 7)
> sample(X, 7, replace=TRUE)
> sample(X) # Rearangement: permutation (aléatoire)
> sample(5) # Rearangement: permutation (aléatoire)
> sample(X[X > 8], 7)
> sample(X[X > 8], 7, replace=TRUE)
> sample(X[X > 9], 7)
> sample(X[X > 10], 7)
> example(sample)
```

Exemple d'application : l'épreuve du lancé de dé

Considérons dans un premier temps un dé équilibré. On effectue 500 lancers, puis on regarde l'effectif des résultats

```
> res_DE = sample(1:6, 500, replace = T)
> table(res_DE)
```

On réitère l'expérience, mais cette fois avec un dé pipé (une chance sur deux de tomber sur la valeur 1)

```
> res_DE_pipo = sample(1:6, 500, replace = T, prob=c(0.5,0.1,0.1,0.1,0.1,0.1))
> table(res_DE_pipo)
```

Pour comparer les deux résultats, on peut réaliser le graphique suivant :

```
> titre1 = "Fréquences obtenues pour \n 500 lancers de dé équilibré"
> titre2 = "Fréquences obtenues pour \n 500 lancers de dé pipé"
> par(mfrow=c(1,2))
> barplot(table(res_DE)/500, main=titre1)
> barplot(table(res_DE_pipo)/500, main=titre2)
```

Autre application :

Une urne contient 5 boules blanches et 10 boules noires. On considère les deux épreuves suivantes :

E1 - On tire successivement 10 boules dans l'urne, avec remise.

E2 - On tire successivement 10 boules dans l'urne, sans remise.

2) Simuler une réalisation de chacune des deux épreuves avec la fonction **sample**. On représentera une boule blanche par le chiffre 1 et une boule noire par le chiffre 0.

3) On s'intéresse à la variable X (resp. Y) comptant le nombre de boules blanches tirées lors d'une réalisation de l'épreuve E1 (resp. de l'épreuve E2).

Simuler 500 réalisations de chacune de ces variables en utilisant la fonction **sample**. Comparer suivant le type d'épreuve, le diagramme en barre des fréquences observées avec la distribution des lois binomiales et hypergéométriques correspondantes (on utilisera les fonctions **dbinom** et **dhyper**).

Exercice 3.2 — Loi faible des grands nombres

Soit X_1, \dots, X_n n variables aléatoires indépendantes et distribuées suivant la loi de Bernoulli de paramètre p. On veut illustrer la convergence en probabilité de :

$$W_n = \frac{X_1 + \dots + X_n}{n} = \frac{S_n}{n}, \text{ vers } p.$$

On fixe $p = 0.5$ et $n = 5$

1) On commence par simuler un échantillon de taille 100 distribué suivant la loi de W_n
Indications :

- Quelle est la loi de $S_n = X_1 + \dots + X_n$?
 - Simuler un échantillon de taille $N = 100$ distribué suivant la loi de S_n . On le note s_1, \dots, s_N
 - On déduit un échantillon distribué suivant la loi de W_n en prenant $\frac{s_1}{n}, \dots, \frac{s_N}{n}$
- On fixe $\varepsilon = 0.1$ Evaluer la proportion $R_n(\varepsilon)$ de valeurs simulées qui vérifient $|W_n - p| > \varepsilon$
 - Refaire le calcul de $R_n(0.1)$ pour différentes valeurs de $n = 10, 25, 50, 75, 100, 150, 200, 500$
 - Refaire les calculs pour $\varepsilon = 0.01$ et $n = 10, 100, 200, 500, 1000, 5000, 10000, 50000$
 - Commenter les résultats obtenus.

Exercice 3.3 — Approximation binomiale-Poisson

1) Comparer à l'aide de diagrammes bâton la loi de Poisson de paramètre 5 avec les lois binomiales suivantes :

- $\mathcal{B}(10, \frac{1}{2})$
- $\mathcal{B}(20, \frac{1}{4})$
- $\mathcal{B}(50, \frac{1}{10})$

d) $\mathcal{B}(100, \frac{1}{20})$

2) Pour $n=10, 20, 50$ et 100 , évaluer le maximum de l'erreur commise lorsque l'on approxime la loi binomiale par la loi de Poisson, c'est-à-dire calculer :

$$Erreur_n = \max \{|P(X_n = k) - P(Y = k)|, k \in \mathbb{N}\},$$

où X_n suit une loi binomiale $\mathcal{B}(n, \frac{5}{n})$ et Y suit une loi de Poisson $\mathcal{P}(5)$ (regarder la fonction `dpois`)

3) Commenter les résultats obtenus.

Exercice 3.4 — Application

Suite à une campagne de vaccination contre le paludisme, on estime à 2%, la proportion de personnes qui seront pourtant atteintes de la maladie.

On note X la variable aléatoire égale au nombre de personnes malades dans un petit village de 100 habitants.

- 1) Quelle est la loi de X (préciser les paramètres)?
- 2) Calculer la probabilité de constater plus d'une personne malade
- 3) Calculer la probabilité de constater au plus dix personnes malades
- 4) Recalculer ces probabilités en utilisant une approximation de la loi de X par une loi de Poisson (préciser le paramètre)
- 5) Quelle est la qualité de l'approximation ?

Exercice 3.5 — Simulation : la loi binomiale

Un générateur de nombres aléatoires est une fonction qui reproduit les résultats d'une expérience aléatoire. On obtient ainsi une suite de nombres simulés suivant une loi de probabilité donnée. Par exemple, le générateur de la loi de Bernoulli de paramètre $\frac{1}{2}$ reproduit le lancer d'une pièce équilibrée.

- 1) Simuler un échantillon de taille 500 distribué suivant la loi binomiale de paramètres $n = 15$ et $p = 0.3$. On note x_1, \dots, x_{500} les valeurs simulées (les enregistrer sous forme d'un vecteur de taille 500)
- 2) Représenter cet échantillon : tracer le nuage de points $\{(i, x_i) \mid i = 1, \dots, 500\}$
- 3) Calculer la fréquence des différentes observations $\{0, \dots, 15\}$ en utilisant la fonction `table`. On note f_0, \dots, f_{15} ces fréquences (empiriques). (les enregistrer sous forme d'un vecteur)
- 4) Représenter les fréquences f_0, \dots, f_{15} à l'aide d'un diagramme "bâton"
- 5) Comparer la distribution de la loi binomiale de paramètres $(15, 0.3)$ et les fréquences empiriques f_0, \dots, f_{15}
- 6) Calculer et représenter graphiquement les fréquences cumulées (utiliser la fonction `cumsum`).
- 7) Comparer la fonction de répartition de la loi binomiale de paramètres $(15, 0.3)$ et la courbe des fréquences cumulées.
- 8) Calculer la moyenne empirique de l'échantillon simulé.
- 9) Comparer la moyenne empirique et l'espérance de la loi binomiale.
- 10) Calculer la variance empirique de l'échantillon simulé.

- 11) Comparer la variance empirique et la variance de la loi binomiale.
- 12) Simuler un second échantillon distribué suivant la loi binomiale de paramètres $n=15$ et $p=0.3$ et refaire les questions précédentes.
- 13) Comparer et commenter les résultats obtenus.