Définitions

Introduction

L'objectif de cette séance est d'utiliser les définitions de probabilités et de probabilités conditionnelles.

Probabilités

- Un **événement** est un ensemble d'éventualités dont on peut mesurer la probabilité comprise entre 0 et 1.
- Soit A_1, A_2, …, A_n sont n événements **disjoints**. Nous avons

$$\mathrm{Prob}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathrm{Prob}(A_i)\,.$$

- Par définition, cette propriété est aussi vérifiée pour une suite infinie d'événements disjoints.
- La probabilité de l'événement dit **certain**, constitué de toutes les éventualités, est égale à un.
- La probabilité conditionnelle de l'événement A sachant B, un événement de probabilité non nulle est définie par

$$\mathrm{Prob}(A|B) = \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(B)}\,.$$

- Les événements A et B sont indépendants si

$$\mathrm{Prob}(A|B) = \mathrm{Prob}(A)\,.$$

- De manière équivalente, A et B sont indépendants si

$$\mathrm{Prob}(A \cap B) = \mathrm{Prob}(A)\mathrm{Prob}(B)\,.$$

  (Se convaincre mathématiquement de cette dernière affirmation).

Variables aléatoires

Une variable aléatoire à valeurs discrètes est un nombre aléatoire, X, dont la valeur appartient à un ensemble fini ou dénombrable, par exemple les valeurs observées à la suite d'un lancer de dé.

- On peut alors numéroter les valeurs de ce nombre.
- La loi de probabilité de la variable X est définie de la manière suivante

$$\mathrm{Prob}(X = i) = p_i\,,$$

  où

$$p_i \geq 0 \quad \text{et} \quad \sum_{i=1}^{\infty} p_i = 1\,.$$

  En d'autres termes, une loi de probabilité correspond à une suite, peut-être finie, de nombre positifs dont la somme est égale à un.
- La formule des probabilités totales est particulièrement importante. Pour tout événement A, nous avons

$$\mathrm{Prob}(A) = \sum_{i=1}^{\infty} \mathrm{Prob}(A|X = i)\mathrm{Prob}(X = i)\,.$$

Valeur moyenne

La valeur moyenne, ou espérance, de la variable aléatoire X est égale à

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i \, \mathrm{Prob}(X = i) \,.$$

Pour bien comprendre cette formule, supposons que X peut prendre les valeurs $1, 2, \ldots, N$. Supposons maintenant que l'on dispose de n observations de la variable X, disons $(x_1, x_2, \ldots, x_n)$.

Il nous est plus familier de calculer la moyenne à l'aide de la formule suivante

$$\mathrm{moyenne}(X) = \frac{1}{n} \sum_{k=1}^{n} x_k \,.$$

Pour retrouver la formule de l'espérance, il faut parcourir l'ensemble des N résultat et calculer la fréquence de chacune des observations

$$\mathrm{moyenne}(X) = \sum_{i=1}^{N} i \, \frac{\#\{k : x_k = i\}}{n} \,.$$

en acceptant le fait que la fréquence de l'observation i est proche de sa probabilité (loi des grands nombres),

$$\frac{\#\{k : x_k = i\}}{n} \approx \mathrm{Prob}(X = i),$$

nous retrouvons bien la formule donnée plus haut.

Exercice

On lance dé à 6 faces et on note X le résultat du lancer. Puis on lance un dé à X faces (plus exactement, on simule son lancer numériquement). On note Y le résultat final.

- Ecrire un programme R pour simuler $n = 100000$ observations, $(y_1, y_2, \ldots, y_n)$, de cette experience (commandes utiles : `sample` et `sapply`). Commenter les opérations effectuées (deux lignes).

```r
# comment 1
programme_chatgpt <- NULL #à modifier
```

- Donner une table des valeurs de probabilités empiriques issues de l'expérience simulées arrondies à 3 décimales (commandes utiles : `table` et `round`).

```r
# comment 1
programme_chatgpt <- NULL #à modifier
```

- Calculer la valeur moyenne de l'echantillon simulé $(y_1, y_2, \ldots, y_n)$.

```r
# comment 1
```

- A l'aide de la formule des probabilités totales, donner une formule mathématique décrivant la probabilité théorique $\mathrm{Prob}(Y = i)$, pour tout $i = 1, \ldots, 6$.

- Ecrire une fonction en R calculant la probabilité $\mathrm{Prob}(Y = i)$, pour tout $i = 1, \ldots, 6$.

```r
# comment 1
prob <- function(i) {
  if (!(i %in% 1:6)) stop("Valeur attendue entre 1 et 6")
  return(NULL) #à modifier
  }
```

- Donner une table des valeurs de probabilités théoriques arrondies à 3 décimales (commandes utiles : `sapply` et `round`).
- Comparer graphiquement les probabilités empiriques et théoriques.

```
# NULL a modifier
#plot(NULL,
#     lwd = 6, col ="red3",ylab = "frequency", las = 1)

# NULL a modifier
#points(1:6+0.1, sapply(1:6, NULL) ,
#       type = "h", col = "green3", lwd = 6)

#legend(x = 4,y = 0.3, col = c("red3","green3"),
#       pch =19,legend = c("empirique", "théorique"))
```

- Calculer numériquement l'espérance de la variable Y (commandes utiles : `sapply` et `sum`). Comparer sa valeur à la valeur moyenne de l'echantillon simulé $(y_1, y_2, \ldots, y_n)$.

```
# comment 1
```