Probabilités
Université Clermont Auvergne
IUT Clermont Auvergne
2A BUT Info, 2024-2025

A. El Oudi
A. Khaldi
C. Samir
A. Wohrer

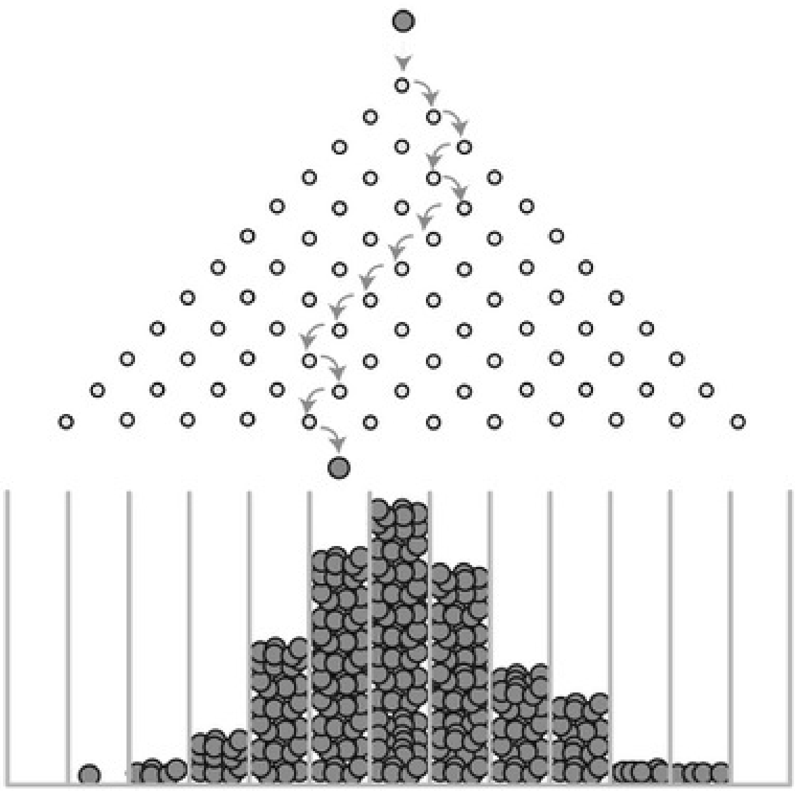
TP2 : Planche de Galton

Objectifs

1. **Simuler** une loi de probabilités discrète.
2. Représenter une loi discrète par son **histogramme**.
3. Explorer certaines lois de probabilités classiques.

Introduction

La planche de Galton est un dispositif imaginé par Sir Francis Galton en 1889, afin d'observer concrètement le phénomène de convergence de certaines lois de probabilités.



Il s'agit d'un empilement régulier de clous sur une planche, suivant le schéma triangulaire illustré ci-dessus. Des billes sont lâchées depuis le haut. À chaque étage, une bille tapant un clou peut tomber à gauche ou à droite du clou, et ce de manière *probabiliste*.

Dans la planche de Galton "classique", la bille a autant de chances de tomber à gauche qu'à droite du clou. Cependant, on peut aussi imaginer des planches "biaisées" dans lesquelles, à chaque clou, la bille a plus de chances de tomber d'un côté que de l'autre.

Le but de ce TP est de simuler des planches de Galton, et d'étudier les distributions de probabilité résultantes.

Modélisation

On modélise la planche de Galton de la manière suivante :

- Il existe un nombre $0 \leq p \leq 1$, tel que à chaque clou, la bille a
  - une probabilité p de passer à droite du clou,
  - une probabilité $1 - p$ de passer à gauche du clou.

  Cette probabilité p est la même pour tous les clous, il s'agit d'une caractéristique « globale » de la planche. Dans une planche de Galton classique, $p = 0.5$, mais vous devez pouvoir simuler une planche avec n'importe quelle valeur de p.
- On note $N + 1$ le nombre d'étages de la planche. Les étages sont numérotés depuis le haut : l'étage tout en haut possède le numéro 0, et l'étage tout en bas possède le numéro N. Dans le TP, vous pourrez simuler des planches avec différentes valeur de N.
- À l'étage 0, il y a une unique position de départ pour la bille. À l'étage 1, il y a 2 positions de passage possibles pour la bille (suivant qu'elle est tombée à gauche ou à droite du premier clou). À l'étage 2, il y a 3 positions de passage possibles pour la bille, etc. De manière générale, si on considère l'étage numéro n, il y a $n + 1$ positions de passage possibles, indexées de 0 (position la plus à gauche) à n (position la plus à droite).

Rappels de Python

Rappels en vrac. À vous d'aller y trouver ce dont vous avez besoin !

- Documentation officielle de `numpy.random` : par ici.
  Vous pouvez par exemple instantier un générateur de nombres aléatoires :

```
rng = np.random.default_rng()
```

  puis l'utiliser pour tirer des nombres suivant la loi qui vous intéresse.
- Le module `scipy.special` contient des fonctions qui peuvent vous être utiles :

```
from scipy.special import factorial,binom
print(factorial(5))
```

```
print(binom(7,3))
```

- La librairie `matplotlib` permet d'afficher des figures :

```
import matplotlib.pyplot as plt
k = np.array(range(10))
plt.figure("La fonction x^2 sur les entiers de 0 à9")
plt.plot(k, k**2, 'r')
plt.show()
```

Exercices

Exercice 1 (Simulation d'une planche).

1. On considère une bille dans la planche, à un certain étage n donné. Notons k la position actuelle de la bille ($k = 0$ si elle est complètement à gauche, $k = n$ si elle est complètement à droite). Quelles positions la bille peut-elle atteindre à l'étage suivant ?

2. On définit la variable aléatoire Y telle que

$$\text{position suivante} = \text{position actuelle} + Y$$

Quelle loi de probabilité suit la variable Y ? Trouvez comment générer en Python des échantillons de cette loi de probabilité.

3. Déduisez-en un moyen simple de simuler la planche de Galton. Écrivez une fonction

```
def simul_planche(N,p):
```

qui simule le passage d'une bille à travers une planche de Galton, et renvoie la position d'arrivée de la bille. Cette fonction prend en entrée deux paramètres définissant la nature de la planche (sa hauteur N et son paramètre gauche/droite p), et doit renvoyer un nombre entier aléatoire entre 0 et N (position d'arrivée de la bille).

4. Simulez un grand nombre de billes (par exemple 10000) passant à travers une planche de Galton de paramètres $N = 12$ et $p = 0.5$. Stockez les 10000 positions d'arrivée obtenues.

5. Tracez l'histogramme des 10000 positions d'arrivée précédentes. Vous pouvez utiliser la fonction `hist` de la librairie `matplotlib`, ou encore la fonction `histplot` de la librairie `seaborn`.

Question en passant : avec une planche de Galton réelle, et de vraies billes, comment pourrait-on obtenir concrètement cet histogramme ?

6. Reproduisez l'expérience avec différentes autres valeurs possibles pour le paramètre p. Quelles sont les conséquences au niveau de l'histogramme ?

Question en passant : comment pourrait-on fabriquer une planche de Galton réelle dont le paramètre p soit différent de 0.5 ?

Exercice 2 (Identification de la loi sous-jacente).

1. Soit X la variable aléatoire donnant la position d'arrivée d'une bille passant à travers une planche de Galton de hauteur N et de paramètre p. Justifiez que X suit une loi de probabilités connue.

2. Rappelez la valeur théorique de la loi pour X, c'est-à-dire l'expression des nombres $P(X = k)$, pour toutes les valeurs possibles de k. Dans les histogrammes simulés à l'exercice précédent, rajoutez le tracé de cette loi "théorique", et vérifiez qu'elle se superpose bien à l'histogramme.

Exercice 3.

On fixe les paramètres suivants pour une planche de Galton :

$$N = 1000, \quad p = 0.004$$

On note X la variable aléatoire donnant la position d'arrivée d'une bille dans cette planche.

1. Décrivez, avec des mots, comment fonctionnerait une telle planche de Galton.
2. Quelle est l'espérance (théorique) de X ?
3. Justifiez, avec un résultat de votre cours, que la variable aléatoire X se comporte comme une loi "simplifiée" dont vous donnerez le paramètre.
4. Vérification : tracez l'histogramme empirique d'une telle planche (par exemple avec 10000 billes) ainsi que la loi de probabilité "simplifiée" proposée à la question précédente. Vérifiez que les deux se superposent bien.

Exercice 4 (théorème central limite).

On fixe les paramètres suivants pour une planche de Galton :

$$N = 1000, \quad p = 0.5$$

On note X la variable aléatoire donnant la position d'arrivée d'une bille dans cette planche.

1. Étant donnée la loi pour la variable X, donnez la valeur des nombres $m = \mathrm{E}(X)$ et $V = \mathrm{Var}(X)$.
2. Justifiez qu'on peut écrire $X = \sum_{i=1}^{N} Y_i$, pour des variables Y_i dont vous préciserez la loi de probabilité. Les variables Y_i sont-elles indépendantes entre elles ?

3. Un célèbre résultat, intitulé **théorème central limite**, dit en gros la chose suivante : *Soit X une variable aléatoire d'espérance m et de variance V. Si X est construite comme la somme d'un grand nombre de variables aléatoires indépendantes entre elles, alors elle suit approximativement une **loi normale** :*

$$P(X = k) \simeq C.\exp\left(-\frac{(k-m)^2}{2V}\right)$$

Ici, C est juste une constante qui assure que la loi de probabilité est *normalisée* (i.e., que sa somme fait 1). Sa valeur s'estime facilement à l'aide d'un ordinateur.

Vérifiez ce résultat, en simulant l'histogramme de cette planche de Galton, et en y superposant la loi normale donnée par la formule ci-dessus.

Exercice 5 (pour les curieux).

Si vous êtes parvenu jusqu'ici et qu'il vous reste quelques neurones, vous vous poserez peut-être la question existentielle suivante : *dans l'exercice 3 aussi, la variable X est construite comme la somme d'un grand nombre de variables aléatoires indépendantes.* Or sa loi n'est pas bien approchée par une loi normale !

La raison est que l'énoncé du *théorème central limite* donné à l'exercice 4 est un peu approximatif. Un énoncé un peu plus précis est le suivant : si on se donne une famille $(Y_i)_{i\in\mathbb{N}}$ de variables aléatoires, toutes de même loi, et indépendantes entre elles, alors la variable

$$X = \sum_{i=1}^{N} Y_i$$

tend vers une loi normale lorsque N tend vers l'infini.

Justifiez qu'on est bien dans une telle situation à l'exercice 4, mais qu'on n'est pas dans une telle situation à l'exercice 3.