# Introduction à la Statistique

## Espérance, variance et la distribution binomiale

Stéphane Guerrier, Mucyo Karemera, Samuel Orso & Lionel Voirol

Data Analytics Lab

# Espérance

$$\mathbb{E}(X) = \sum_{\text{tout } k} k \cdot \mathbb{P}(X = k)$$

Interprétation :

Une moyenne pondérée par les probabilités

La moyenne à long terme de $X$

La valeur juste d'un pari

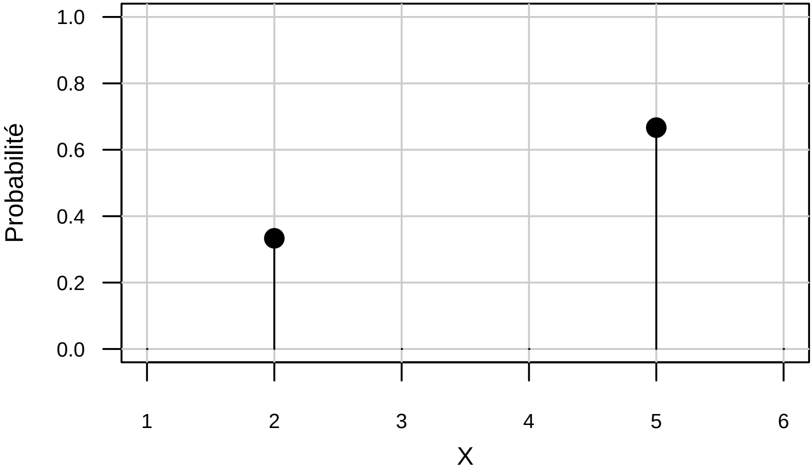
Le point d'équilibre pour un histogramme ou un diagramme en barres de probabilité

# Comment calculer $\mathbb{E}(X)$

| $k$ | 2 | 5 |
|---|---|---|
| $\mathbb{P}(X = k)$ | 1/3 | 2/3 |

$$\begin{aligned}\mathbb{E}(X) &= 2 \cdot \mathbb{P}(X = 2) + 5 \cdot \mathbb{P}(X = 5) \\ &= 2\,(1/3) + 5\,(2/3) \\ &= 2/3 + 10/3 = 12/3 = 4\end{aligned}$$

C'est une *moyenne pondérée* des valeurs de $X$, où les *poids* correspondent aux probabilités associées à ces valeurs.


1.0
0.8
0.6
0.4
0.2
0.0
Probabilité
1
2
3
4
5
6
X

# Un Pari

Vous payez 100 CHF pour un pari.

Si vous gagnez, vous recevez 210 CHF, et si vous perdez, vous recevez 0 CHF.
$X$ = votre gain/perte

$$X = 110 \text{ CHF, si vous gagnez}$$
$$X = -100 \text{ CHF, si vous perdez}$$

Supposons que les probabilités de gagner/perdre sont

$$\mathbb{P}(\text{gagner}) = 0.55$$
$$\mathbb{P}(\text{perdre}) = 0.45$$

Ce pari est-il équitable ?

# Un Pari

La fonction de masse de probabilité (FMP) de $X$ est :

| $k$ | 110 CHF | $-100$ CHF |
|---|---|---|
| $\mathbb{P}(X = k)$ | 0.55 | 0.45 |

Pour vérifier si le pari est équitable ou non, calculez l'espérance de votre gain/perte :

$$\begin{aligned}\mathbb{E}(X) &= 110\ \mathbb{P}(X = 110) - 100\ \mathbb{P}(X = -100)\\ &= 110\ (0.55) - 100\ (0.45)\\ &= 15.5 \text{ CHF}\end{aligned}$$

Est-ce un pari équitable ?

Ce n'est pas un pari équitable, puisque l'espérance n'est pas nulle... mais devriez-vous jouer ?

# Qui veut gagner des millions ?

Quitter = 100 000 CHF
Bonne réponse $X$ = 250 000 CHF
Mauvaise réponse $X$ = 32 000 CHF

Deux cas :

Choisir au hasard parmi les 4 réponses possibles.
$\mathbb{P}(\text{“}\textit{bonne réponse}\text{”}) = 1/4$
$\mathbb{E}(X) = (250\,000)(1/4) + (32\,000)(3/4) = 86\,500$ CHF

Utiliser l'option 50-50, et choisir au hasard entre 2 réponses.
$\mathbb{P}(\text{“}\textit{bonne réponse}\text{”}) = 1/2$
$\mathbb{E}(X) = (250\,000)(1/2) + (32\,000)(1/2) = 141\,000$ CHF

# Ajout d'une constante

$$\mathbb{E}(X + \alpha) = \mathbb{E}(X) + \alpha$$

# Fonctions d'une Variable Aléatoire

| $k$ | 10 | 12 | 15 |
|---|---|---|---|
| $\mathbb{P}(X = k)$ | 0.4 | 0.5 | 0.1 |

$$\mathbb{E}(X) = 10(0.4) + 12(0.5) + 15(0.1) = 11.5$$

Doubler la valeur de $X$

| $k$ | 20 | 24 | 30 |
|---|---|---|---|
| $\mathbb{P}(2\,X = 2\,k)$ | 0.4 | 0.5 | 0.1 |

$$\mathbb{E}(2\,X) = 20(0.4) + 24(0.5) + 30(0.1) = 23$$

$$\mathbb{E}(2\,X) = 2\,\mathbb{E}(X)$$

$$= 2(11.5) = 23$$

# Multiplication par une constante

$$\mathbb{E}(b\,X) = b\,\mathbb{E}(X)$$

# Propriétés de $\mathbb{E}(X)$

1. $\mathbb{E}(X + \alpha) = \mathbb{E}(X) + \alpha$
2. $\mathbb{E}(b\,X) = b\,\mathbb{E}(X)$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. Si $X$ et $Y$ sont indépendants, alors

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$$

# Exemple : *Dé bleu et dé vert*

Deux dés *équilibrés* sont lancés, un bleu et un vert.

$$Y = \text{face sur le dé bleu}, \quad Z = \text{face sur le dé vert}$$

$$X = \text{sommes des deux faces}, \quad W = \text{produit des deux faces}$$

Déterminez l'espérance de $Y$ et $Z$.
L'Univers pour chaque dé bleu ou vert est

$$S = \{1, 2, 3, 4, 5, 6\}$$

La fonction de masse de probabilité (FMP) de $Y$ et $Z$ est

| $y$ | $Y=1$ | $Y=2$ | $Y=3$ | $Y=4$ | $Y=5$ | $Y=6$ |
|---|---|---|---|---|---|---|
| $\mathbb{P}(Y=y)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

# Exemple : *Dé bleu et dé vert*

FMP :

| $y$ | $Y = 1$ | $Y = 2$ | $Y = 3$ | $Y = 4$ | $Y = 5$ | $Y = 6$ |
|---|---|---|---|---|---|---|
| $\mathbb{P}(Y = y)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

Espérance :

$$\mathbb{E}(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6}$$

De manière similaire, $\mathbb{E}(Z) = \frac{21}{6}$

# Exemple : *Dé bleu et dé vert*

Quelle est l'espérance de $X$ ?

$$X = Y + Z$$

$$\mathbb{E}(X) = \mathbb{E}(Y + Z) = \mathbb{E}(Y) + \mathbb{E}(Z) = \frac{21}{6} + \frac{21}{6} = \frac{42}{6} = 7$$

Quelle est l'espérance de $W$ ?

$$W = YZ$$

Puisque le résultat d'un dé ne dépend pas de l'autre, les variables aléatoires $Y$ et $Z$ sont *indépendantes*.

$$\mathbb{E}(W) = \mathbb{E}(YZ) = \mathbb{E}(Y)\mathbb{E}(Z) = \frac{21}{6} \cdot \frac{21}{6} = 12.25$$

# Mesurer la variabilité

La variance = Mesure de l'étendue de la FMP.

La variance = Espérance de la distance au carré de l'espérance :

Espérance : $\mathbb{E}(X)$

Distance de l'espérance : $X - \mathbb{E}(X)$

Distance au carré de l'espérance : $(X - \mathbb{E}(X))^2$

Espérance de la distance au carré de l'espérance :

$$Var(X) = \mathbb{E}\left(\{X - \mathbb{E}[X]\}^2\right)$$

# Une autre façon de calculer la Variance

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}\{X\})^2$$

Remarque :

$\mathbb{E}(X^2)$ n'est pas la même chose que $(\mathbb{E}\{X\})^2$ !

# Comment calculer Var($X$)

| $k$ | 1 | 2 | 3 |
|---|---|---|---|
| $\mathbb{P}(X = k)$ | 0.2 | 0.6 | 0.2 |

$\mathbb{E}(X) = 1(0.2) + 2(0.6) + 3(0.2) = 2$.

$\mathbb{E}(X^2) = \sum k^2\ \mathbb{P}(X = k)$

| $k^2$ | 1 | 4 | 9 |
|---|---|---|---|
| $\mathbb{P}(X = k)$ | 0.2 | 0.6 | 0.2 |

$$\mathbb{E}(X^2) = 1(0.2) + 4(0.6) + 9(0.2) = 4.4$$

Calculer la variance :

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}\{X\})^2 \\ &= 4.4 - 2^2 = 0.4 \end{aligned}$$

# Écart type

L'écart-type (noté $\sigma$) mesure la dispersion des valeurs autour de la moyenne. Il correspond à la racine carrée de la variance, ce qui permet de conserver les mêmes unités que les données.

$$\sigma = \sqrt{Var(X)}$$

# Propriétés

$\text{Var}(\alpha X) = \alpha^2 \text{ Var}(X)$
Par exemple,

$$\text{Var}(3 \cdot X) = 3^2 \cdot \text{Var}(X) = 9 \cdot \text{Var}(X)$$

$\text{Var}(\alpha + X) = \text{Var}(X)$

Par exemple,

$$\text{Var}(3 + X) = \text{Var}(X)$$

Si $X$ et $Y$ sont **indépendants**,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

# Événements indépendants

Si $A$ et $B$ sont *indépendants* :

$$\mathbb{P}(A \text{ et } B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

Si $A_1, A_2, \ldots, A_n$ sont *indépendants* :

$$\mathbb{P}(A_1, A_2, \ldots, \text{ et } A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\ldots\mathbb{P}(A_n)$$

# Lancers de pièces équilibrées

Chaque lancer a une probabilité de $\frac{1}{2}$ de tomber sur Pile

Les issues sont indépendantes
Si l'on lance une pièce $n$ fois :

$$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\dots\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^n$$

$\mathbb{P}(\textit{“16 Piles”}) = \frac{1}{2^{16}} = \frac{1}{65536} = 0.00001526$

$\mathbb{P}(\textit{“10 Piles puis 6 Faces”}) = \frac{1}{2^{16}} = \frac{1}{65536} = 0.00001526$

# Et compter les Piles

Lancer une pièce équilibrée 4 fois

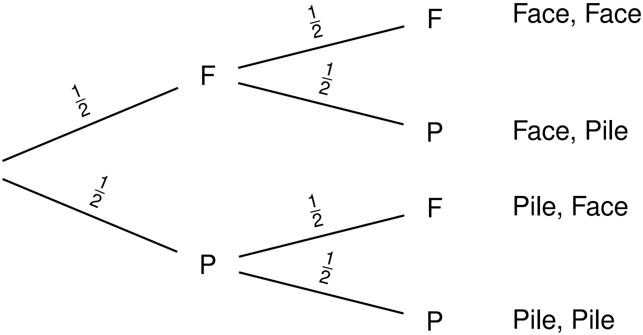
$X$ = nombre de Piles

$\mathbb{P}(X = 4) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{16}$

$\mathbb{P}(X = 0) = \mathbb{P}(4 \text{ Faces}) = \frac{1}{16}$

$\mathbb{P}(X = 1) = ???$

# Univers



F
P
F
P
F
P
1/2
1/2
1/2
1/2
1/2
1/2
Face, Face
Face, Pile
Pile, Face
Pile, Pile

# Univers

| | | | |
|---|---|---|---|
| FFFF | FPFF | PFFF | PPFF |
| FFFP | FPFP | PFFP | PPFP |
| FFPF | FPPF | PFPF | PPPF |
| FFPP | FPPP | PFPP | PPPP |

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\mathbb{P}(X = k)$ | | | | | |

# Univers

| | | | |
|---|---|---|---|
| FFFF | FPFF | PFFF | PPFF |
| FFFP | FPFP | PFFP | PPFP |
| FFPF | FPPF | PFPF | PPPF |
| FFPP | FPPP | PFPP | PPPP |

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\mathbb{P}(X = k)$ | 1/16 | | | | 1/16 |

# Univers

| | | | |
|---|---|---|---|
| FFFF | FPFF | PFFF | PPFF |
| FFFP | FPFP | PFFP | PPFP |
| FFPF | FPPF | PFPF | PPPF |
| FFPP | FPPP | PFPP | PPPP |

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\mathbb{P}(X = k)$ | 1/16 | 1/4 | | | 1/16 |

# Univers

| | | | |
|---|---|---|---|
| FFFF | FPFF | PFFF | PPFF |
| FFFP | FPFP | PFFP | PPFP |
| FFPF | FPPF | PFPF | PPPF |
| FFPP | FPPP | PFPP | PPPP |

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\mathbb{P}(X = k)$ | 1/16 | 1/4 | | 1/4 | 1/16 |

# Univers

| | | | |
|---|---|---|---|
| FFFF | FPFF | PFFF | PPFF |
| FFFP | FPFP | PFFP | PPFP |
| FFPF | FPPF | PFPF | PPPF |
| FFPP | FPPP | PFPP | PPPP |

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\mathbb{P}(X = k)$ | 1/16 | 1/4 | 3/8 | 1/4 | 1/16 |

# En général

- Lancez une pièce $n$ fois
- $2^n$ résultats dans l'espace des échantillons
- $\mathbb{P}(\text{une séquence}) = (1/2)^n$
- $\mathbb{P}(X = k) = (1/2)^n$ (# de séquences avec $k$ Faces)

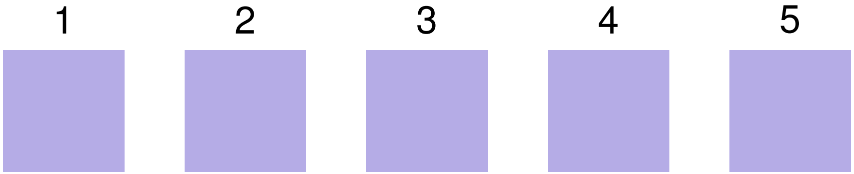
Comment comptons-nous ces combinaisons ? ? ?

# Principe fondamental du comptage

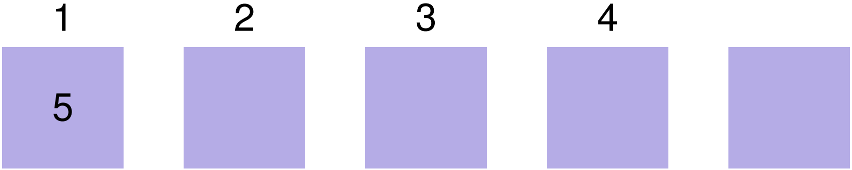
Supposons que $r$ actions (choix, expériences) doivent être effectuées dans un certain ordre :

| ACTION | NOMBRE DE RÉALISATIONS POSSIBLES |
|---|---|
| Action 1<br>Action 2<br>...<br>Action $r$ | $m_1$<br>$m_2$<br>...<br>$m_r$ |
| TÂCHE ENTIÈRE | $m_1 \cdot \ldots \cdot m_r$ |

# Choisir un ordre

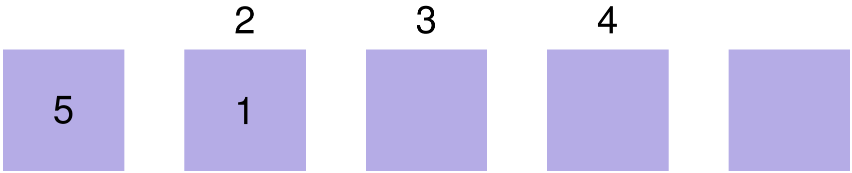


# Choisir un ordre




5

# Choisir un ordre




$$5 \times 4$$

# Choisir un ordre

| | | | | |
|---|---|---|---|---|
| | 2 | | 4 | |
| 5 | 1 | 3 | | |

$$5 \times 4 \times 3$$

# Choisir un ordre

| | 2 | | | |
|---|---|---|---|---|
| 5 | 1 | 3 | 4 | |

$$5 \times 4 \times 3 \times 2$$

# Choisir un ordre

| | | | | |
|---|---|---|---|---|
| 5 | 1 | 3 | 4 | 2 |

$$5 \times 4 \times 3 \times 2 \times 1$$

# Choisir un ordre

| | | | | |
|---|---|---|---|---|
| 5 | 1 | 3 | 4 | 2 |

$$5 \times 4 \times 3 \times 2 \times 1 = 120$$

# Comptage des séquences

Dans l'exemple ci-dessus :

$$(5)\ (4)\ (3)\ (2)\ (1) \ = \ 5!$$

Plus généralement, n factoriel est noté par :

$$n! = n\,(n-1)\,(n-2)\ \ldots\ (3)(2)1$$

Par exemple, $30! = 30\ (29)\ (28)\ \ldots\ (2)1 = 2.65 \times 10^{32}$

Par définition

$$0! = 1$$

# 2 parmi 5


Joueur 1

Joueur 2

# 2 parmi 5


Joueur 2

5

# 2 parmi 5


$$(5 \times 4)/2 = 10$$

# $k$ parmi $n$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

- $n = 5,\ k = 2$

$$C_5^2 = \frac{5!}{2!\,3!} = \frac{5(4)(3)(2)1}{2(1) \times (3)(2)1} = \frac{120}{2 \times 6} = 10$$

- $n = 6,\ k = 4$

$$C_6^4 = \frac{6!}{4!\,2!} = \frac{6(5)(4)(3)(2)1}{(4)(3)(2)(1) \times (2)1} = \frac{720}{48} = 15$$

- $n = 52,\ k = 3$

$$C_{52}^3 = \frac{52(51)50}{(3)(2)1} = \frac{132\,600}{6} = 22\,100$$

# Probabilité d'obtenir 6 Faces sur 10 lancers

$$\mathbb{P}(\text{“}\textit{6 Faces en 10 Lancers}\text{”}) = C_{10}^{6}\left(\frac{1}{2}\right)^{10}$$

# de combinaisons

probabilité de chaque combinaison

# Probabilité d'obtenir 6 Faces sur 10 lancers

$$\mathbb{P}(\text{“}\textit{6 Faces en 10 Lancers}\text{”}) = C_{10}^{6}\left(\frac{1}{2}\right)^{10}$$

\# de combinaisons

probabilité de chaque combinaison

$$\frac{10!}{6!\,4!}\left(\frac{1}{1024}\right) = \frac{10(9)(8)(7)}{4(3)(2)1}\left(\frac{1}{1024}\right) = \frac{210}{1024} = 0.2051$$

# Suite d'événements indépendants

$A_1, A_2, \ldots, A_n$ sont indépendants avec

$$\mathbb{P}(A_i) = p$$
$$\mathbb{P}(A_i^c) = 1 - p$$

Alors,

$$\mathbb{P}(A_1, A_2^c, \text{ et } A_3) = p\,(1-p)\,p = p^2\,(1-p)$$

# Autres probabilités

Lancé d'une pièce déséquilibrée 5 fois

$$\mathbb{P}(F) = 0.2 \ \text{ and } \ \mathbb{P}(P) = 0.8$$

$$\text{Alors, } \ \mathbb{P}(FP) = 0.2\ (0.8) = 0.16$$

$$\mathbb{P}(FPPFP) = 0.2(0.8)(0.8)(0.2)(0.8) = (0.2)^2(0.8)^3 = 0.02048$$
$$\mathbb{P}(PPFPF) = 0.8(0.8)(0.2)(0.8)(0.2) = (0.2)^2(0.8)^3 = 0.02048$$
$$\mathbb{P}(PPPFF) = 0.8(0.8)(0.8)(0.2)(0.2) = (0.2)^2(0.8)^3 = 0.02048$$

$$\mathbb{P}(\text{ “2 Faces” }) = C_5^2(0.2)^2(0.8)^3 = 10(0.02048) = 0.2048$$

# Expérience binomiale

$n$ essais

Les essais sont indépendants

Deux issues possibles : “Succès” (encodé par 1) ou “Échec” (encodé par 0)

$\mathbb{P}($ “Succès” $) = p$ et $\mathbb{P}($ “Échec” $) = 1 - p$

# Exemple : Votation d'une loi

Échantillon de 100 votants ($n = 100$)

Hypothèse : les individus sont indépendants

"Soutien au projet de loi" correspond à un "Succès"

$p = 0.44$

$X$ = # de partisans en faveur de l'initiative

# Probabilité binomiale

$$\mathbb{P}(X = k) = C_n^k \, p^k \, (1 - p)^{n-k}$$

$n$ = nombre total d'essais, $p$ = probabilité de succès

$$\begin{aligned}
&\mathbb{P}(\text{ 40 partisans }) = \mathbb{P}(X = 40) \\
&= C_{100}^{40}(0.44)^{40}(0.56)^{60} \\
&= (1.3746 \times 10^{28})(5.47151 \times 10^{-15})(7.78541 \times 10^{-16}) \\
&= 0.0586
\end{aligned}$$

# Distribution Binomiale

Expérience binomiale :

$n$ essais

Les essais sont indépendants

Deux issues possibles $\mathbb{P}($ "Succès" $) = p$

Probabilité binomiale

$$\begin{aligned}\mathbb{P}(X = k) &= C_n^k \, p^k \, (1-p)^{n-k} \\ &= \frac{n!}{k!\,(n-k)!} \, p^k \, (1-p)^{n-k}\end{aligned}$$