UNIVERSITÉ DE GENÈVE

Introduction à la Statistique

La distribution binomiale et les variables aléatoires continues

Stéphane Guerrier, Mucyo Karemera, Samuel Orso & Lionel Voirol

Data Analytics Lab

n factoriel

n factoriel

$$n! = n\,(n-1)\,(n-2)\ \ldots\ (3)(2)(1)$$
$$0! = 1$$

Exemples :

$$4! = (4)\,(3)\,(2)\,(1) = 24$$
$$5! = (5)\,(4)\,(3)\,(2)\,(1) = (5)\,4! = 120$$
$$10! = (10)\,(9)\,(8)\,(7)\,(6)\,5! = 3'628'800$$

Combinaisons

k parmi n

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

$n = 6,\ k = 4$

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

$n = 20,\ k = 4$

$$C_{20}^4 = \frac{20!}{4!\,(20-4)!} = \frac{20\,(19)\,(18)\,(17)\,16!}{4(3)(2)1\times 16!}$$
$$= \frac{116\,280}{24} = 4'845$$

Exemple : création d'un examen

Nous avons un ensemble de :

15 questions vrai/faux (V/F) et
20 questions à choix multiple.

Question : Nous voulons créer un examen avec 10 questions dont exactement 4 sont V/F. Combien d'examens avec exactement 4 questions V/F peuvent être construits ?

Remarque : Ignorez l'ordre des questions.

Exemple : création d'un examen

Le nombre total de façons de choisir 4 questions V/F est

$$C_{15}^{4} = \frac{15!}{4!\,11!} = 1'365$$

Le nombre total de façons de choisir 6 questions à choix multiple est

$$C_{20}^{6} = \frac{20!}{6!\,14!} = 38'760$$

Le nombre total d'examens avec 4 questions V/F et 6 questions à choix multiple est

$$C_{15}^{4} \times C_{20}^{6} = 1'365 \times 38'760 = 52'907'400$$

Exemple : création d'un examen

Nous avons un ensemble de 15 questions vrai/faux et 20 questions à choix multiple.

Parmi les examens avec 10 questions, quelle est la probabilité de choisir un examen avec 4 questions V/F (et 6 questions à choix multiple) ?

Rappel de la définition classique de la probabilité :

$$\begin{aligned}
&\mathbb{P}(\text{“\textit{Examen avec 4 V/F parmi 10 questions}”})\\
&= \frac{\#\text{ d'examens avec 4 questions V/F parmi 10 questions}}{\#\text{ d'examens avec 10 questions}}\\
&= \frac{C_{15}^{4} \times C_{20}^{6}}{C_{35}^{10}}\\
&= \frac{52'907'400}{183'579'396} = 0.2882
\end{aligned}$$

Exemple : lancer d'une pièce équilibrée

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

\# de combinaisons

probabilité de chaque combinaison

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

Exemple : lancer d'une pièce déséquilibrée

Lancé d'une pièce déséquilibrée 5 fois

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

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

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

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

Exemple : lancer d'une pièce déséquilibrée

Vous lancez une pièce déséquilibrée 100 fois

$n = 100$

Les lancers sont indépendants

"Face" = "Succès"
"Pile" = "Échec"

$p = \mathbb{P}(\text{ "Face" }) = 0.44$
$1 - p = \mathbb{P}(\text{ "Pile" }) = 0.56$

$$X = \# \text{ de Faces}$$

Exemple : lancer d'une pièce déséquilibrée

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

- Quelle est la probabilité d'obtenir 40 Faces parmi les 100 lancers ?

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

Distribution binomiale

Expérience binomiale :

n essais

Les essais sont indépendants

Deux issues possibles $\mathbb{P}(\text{ “Succès” }) = p$

Probabilité binomiale

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

Exemple : Le tiroir à chaussettes

Un tiroir contient 6 chaussettes rouges et 4 chaussettes noires. Quelle est la probabilité de sélectionner 2 chaussettes rouges lorsque vous en tirez 5 du tiroir (avec remise) ?

Il y a 5 essais.

Chaque tirage est indépendant des autres, puisque nous choisissons une chaussette à chaque fois avec remise.

Comme nous sommes intéressés à sélectionner des chaussettes *rouges*, nous appelons cela un succès et sélectionner une chaussette *noire* est un échec.

Pour chaque essai, la probabilité de succès est $p = 6/10 = 0,6$, elle est constante.

Ceci est une *expérience binomiale*.

Exemple : Le tiroir à chaussettes

Un tiroir contient 6 chaussettes rouges et 4 chaussettes noires. Quelle est la probabilité de sélectionner 2 chaussettes rouges lorsque vous en tirez 5 du tiroir (avec remise) ?

X : # de chaussettes rouges (c'est-à-dire succès) parmi les 5 essais.

$$\begin{aligned}\mathbb{P}(X=2) &= C_5^2 \cdot \left(\frac{6}{10}\right)^2 \cdot \left(1-\frac{6}{10}\right)^{5-2} \\ &= 10 \cdot (0.6)^2 \cdot (0.4)^3 \\ &= 0.2304.\end{aligned}$$

Exemple : Le tiroir à chaussettes

Un tiroir contient 6 chaussettes rouges et 4 chaussettes noires. Quelle est la probabilité de sélectionner *au moins 4 chaussettes rouges* lorsque vous en tirez 5 du tiroir (avec remise) ?

$$\begin{aligned}
\mathbb{P}(X \geq 4) &= \mathbb{P}(X = 4) + \mathbb{P}(X = 5) \\
&= C_5^4 \cdot (0.6)^4 \cdot (0.4)^1 + C_5^5 \cdot (0.6)^5 \cdot (0.4)^0 \\
&= 5 \cdot (0.6)^4 \cdot (0.4)^1 + (0.6)^5 \cdot (0.4)^0 \\
&= 0.2592 + 0.0778 = 0.3369
\end{aligned}$$

Prenez la route la plus courte

450 étudiants suivent le cours “Introduction à la Statistique”, mais la probabilité qu'un étudiant ne soit pas endormi à la fin d'un cours est de 0.01.

Quelle est la probabilité qu'au moins un étudiant ne soit pas endormi à la fin du cours ?

$$\mathbb{P}(X \geq 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \cdots + \mathbb{P}(X = 450)$$

$$\begin{aligned}
\mathbb{P}(X \geq 1) &= 1 - \mathbb{P}(X < 1) \\
&= 1 - \mathbb{P}(X = 0) \\
&= 1 \ - \ C_{450}^{0} \ (0.01)^{0}(0.99)^{450} \\
&= 0.9891
\end{aligned}$$

Distribution binomiale ($n = 1$)

$$X = \begin{cases} 1, & \text{avec probabilité } p \\ 0, & \text{avec probabilité } (1-p) \end{cases}$$

$X \sim Bin(1,\ p)$

k	0	1
$\mathbb{P}(X = k)$	$1 - p$	p

$$\mathbb{E}(X) = 0\ (1-p) + 1\ p = p$$
$$\mathbb{E}(X^2) = 0^2\ (1-p) + 1^2\ p = p$$
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 = p - p^2 = p(1-p)$$

Espérance & variance de W

Soit $W = X_1 + X_2 + \ldots + X_n$

Si $X_1, X_2, \ldots, X_n$ sont indépendantes avec $X \sim Bin(1, p)$, alors

$$W \sim Bin(n, p)$$

Espérance de W

$$\begin{aligned} \mathbb{E}(W) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n) \\ &= n\, \mathbb{E}(X) = n\, p \end{aligned}$$

Variance de W

$$\begin{aligned} \mathrm{Var}(W) &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n) \\ &= n\, \mathrm{Var}(X) = n\, p\, (1 - p) \end{aligned}$$

Loi binomiale : espérance et variance

$X \sim Bin(n,\ p)$

Fonction de Masse de Probabilité (FMP)

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

Espérance

$$\mathbb{E}(X) = n\ p$$

Variance

$$\mathrm{Var}(X) = n\ p\ (1-p)$$

Écart-type

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{n\ p\ (1-p)}$$

Variable aléatoire discrète

X prend des valeurs de manière aléatoire

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

Exemple de variables aléatoires continues

$X =$ taille

$Y =$ temps d'attente

$U =$ nombre aléatoire compris entre 0 et 1

Variables aléatoires discrètes vs continues

Variables aléatoires discrètes

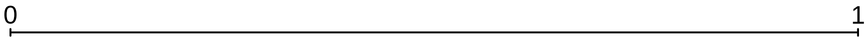
- Prennent des valeurs dans une liste de nombres distincts
- Correspondent à des dénombrements

Variables aléatoires continues

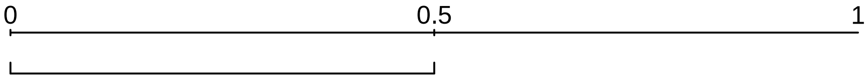
- Prennent des valeurs sur un intervalle
- Correspondent par exemple à des mesures physiques

Une variable aléatoire U

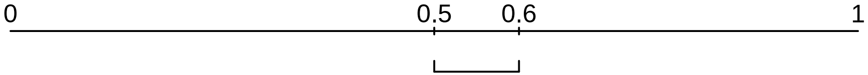
U est un nombre entre 0 et 1. On supose une distribution uniforme.



- $\mathbb{P}(U < 1/2) = 1/2$

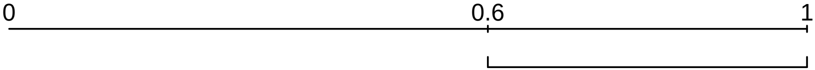


- $\mathbb{P}(0.5 < U < 0.6) = 0.1$



Longueur d'un intervalle

- $\mathbb{P}(U \text{ dans } A) =$ Longueur de A
- $\mathbb{P}(U > 0.6) = 1 - 0.6 = 0.4$

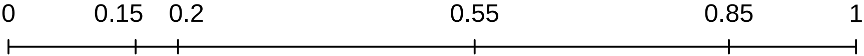


- $\mathbb{P}(U = 0.3) = ?$

$$\mathbb{P}(\text{“}U = \textit{n'importe quel nombre}\text{”}) = 0$$

Pourquoi ?

Exemples



- $A = \{0.15 < U < 0.55\}$

$$\mathbb{P}(A) = \mathbb{P}(0.15 < U < 0.55) = 0.55 - 0.15 = 0.40$$

- $B = \{0.2 < U < 0.85\}$

$$\mathbb{P}(B) = \mathbb{P}(0.2 < U < 0.85) = 0.85 - 0.2 = 0.65$$

- A ou B

$$\mathbb{P}(A \;\; ou \;\; B) = \mathbb{P}(0.15 < U < 0.85) = 0.85 - 0.15 = 0.7$$

- A et B

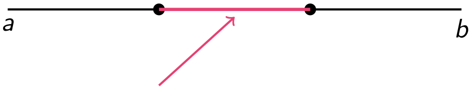
$$\mathbb{P}(A \;\; et \;\; B) = \mathbb{P}(0.2 < U < 0.55) = 0.55 - 0.2 = 0.35$$

- $B|A$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \;\; et \;\; B)}{\mathbb{P}(B)} = \frac{0.35}{0.4} = 0.875$$

Généralement

Soit X une variable aléatoire uniforme sur l'intervalle de a à b



$$\mathbb{P}(\text{“}X \in \textit{intervalle}\text{”}) = \frac{\text{longueur de l'intervalle de l'événement}}{\text{longueur de l'intervalle total}}$$

Variable aléatoire continue

- $\mathbb{P}(X = 6) = 0$
  La probabilité que la variable aléatoire prenne une certaine valeur unique est nulle.

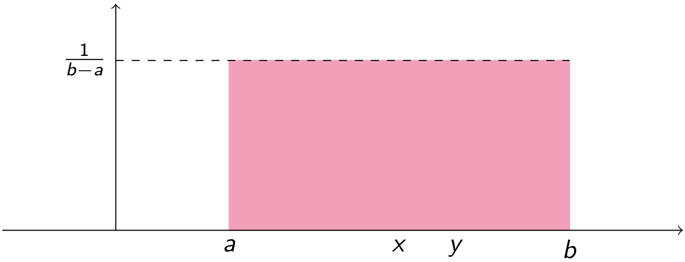
- Les intervalles ont une probabilité > 0

- Pour une variable aléatoire Uniforme U :

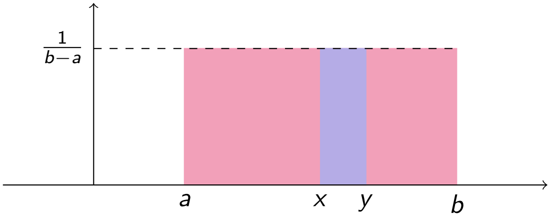
$$\mathbb{P}(x < U < y) = \frac{y - x}{\text{longueur de l'intervalle total}}$$

Loi uniforme : fonction de densité de probabilité

$$f(x) = \frac{1}{b-a}$$



Loi uniforme : fonction de densité de probabilité



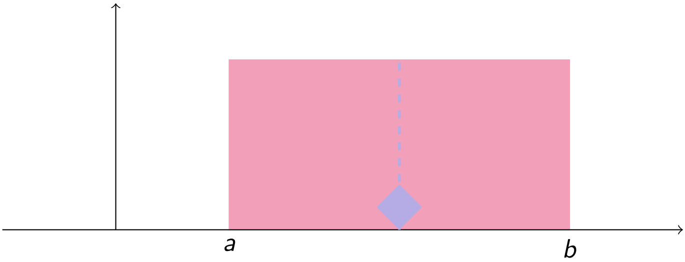
Calcul des probabilités

$$\begin{aligned}\mathbb{P}(x < U < y) &= \text{ Aire} \\ &= \text{ largeur} \times \text{ hauteur} \\ &= (y-x)\cdot\left(\frac{1}{b-a}\right)\end{aligned}$$

Aire totale sous la courbe $= 1$

Loi uniforme : espérance et variance

Centre de gravité



$$\mathbb{E}(U) = \frac{b+a}{2}$$

$$\mathrm{Var}(U) = \frac{(b-a)^2}{12}$$

Exemple : attendre le bus

Jean prend le bus tous les jours pour aller en cours. Le bus arrive à l'arrêt ponctuellement toutes les 30 minutes. *Supposons que Jean arrive à l'arrêt de bus à un moment quelconque de la journée, avec une probabilité uniforme.*

Soit

$$X = \text{le temps (en minutes) que Jean doit attendre pour le bus}$$

Quelle est la probabilité que Jean attende moins de 10 minutes ?

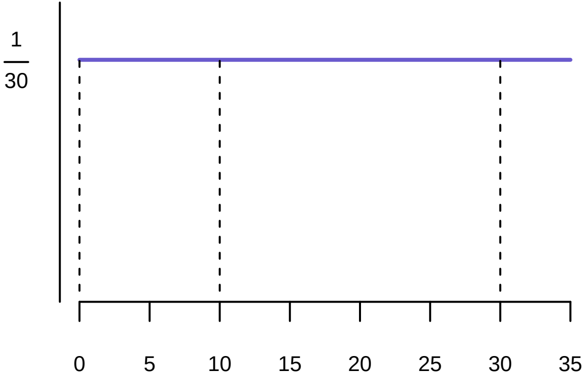
Exemple : attendre le bus

X est une variable aléatoire *continue* qui prend ses valeurs dans $(0, 30)$

Comme tous les temps d'attente sont équiprobables, la variable aléatoire suit la loi uniforme :

$$X \sim U(0, 30)$$

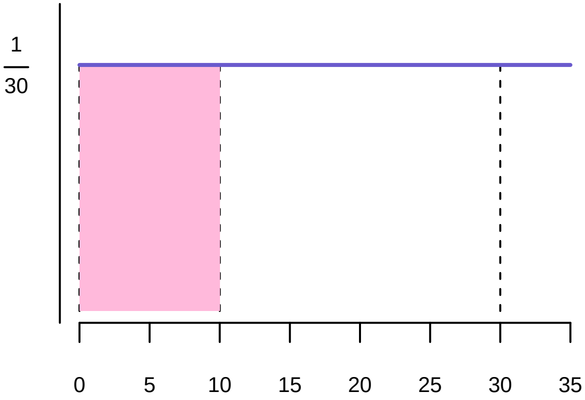
Fonction de densité de probabilité de X :



Exemple : attendre le bus

L'aire sous la courbe de densité entre 0 et 10 représente la probabilité $\mathbb{P}(X \leq 10)$, c'est-à-dire la probabilité que John attende moins de 10 minutes.

$$\mathbb{P}(X \leq 10) = \text{Base} \times \text{Hauteur} = 10 \times \frac{1}{30} = \frac{1}{3}.$$



Fonction de répartition cumulative

Une autre manière de décrire la distribution de X

$F(x) = \mathbb{P}(U \leq x)$

$F(x)$ correspond à l'aire sous la courbe jusqu'à x

Pour $U \sim U(a, b)$:

$$F(x) = \mathbb{P}(U \leq x) = \frac{x - a}{b - a}$$

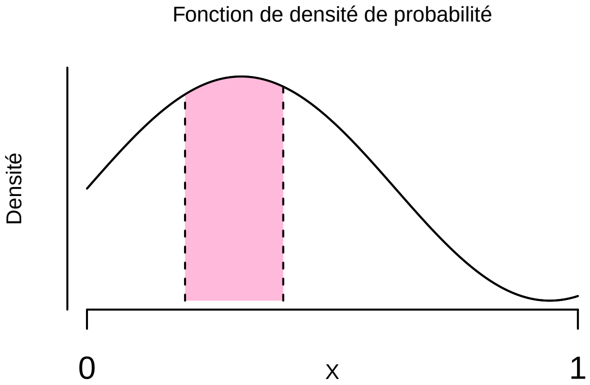
Fonction de répartition cumulative

X est une variable aléatoire continue

$F(x) = \mathbb{P}(X \leq x)$ (définition)

Probabilité	Calcul à l'aide de la fonction de répartition cumulative
$\mathbb{P}(X \leq x)$	$= F(x)$
$\mathbb{P}(x \leq X \leq y)$	$= F(y) - F(x)$
$\mathbb{P}(X \geq x)$	$= 1 - F(x)$

Fonction de densité de probabilité



$P(\text{“}X \in \ \textit{Intervalle}\text{''}\ ) =$ Aire sous la fonction de densité de probabilité