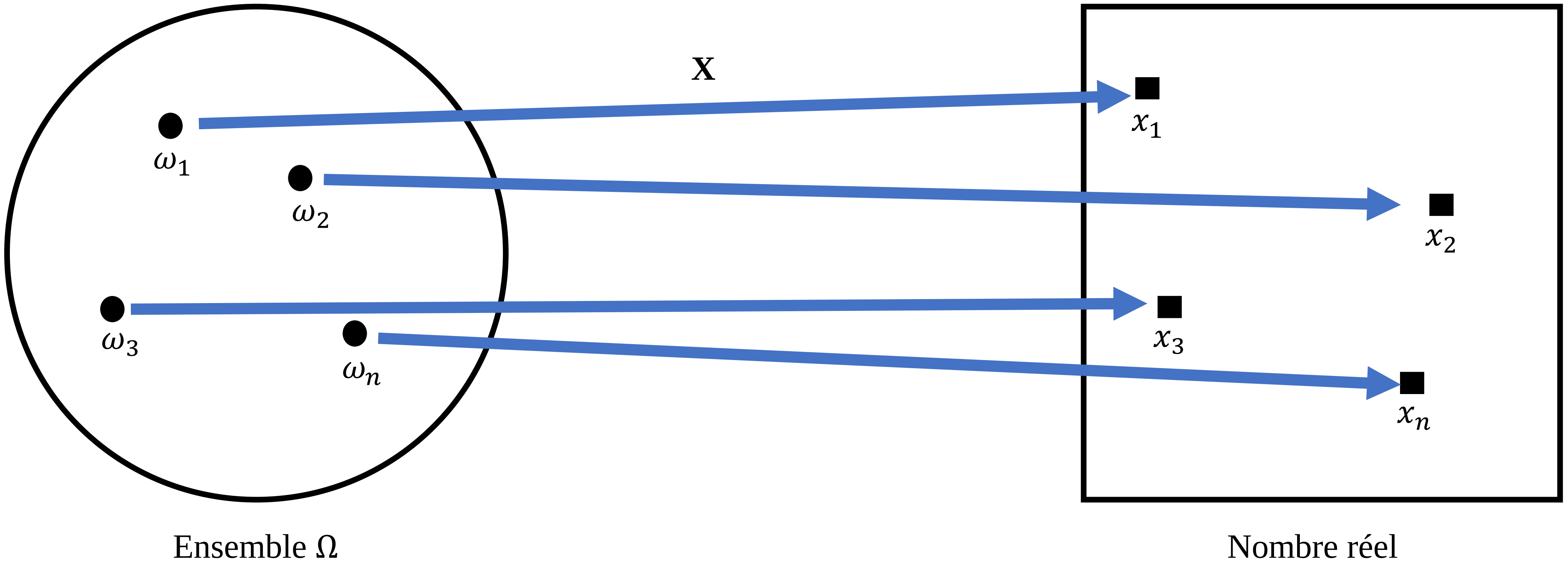
Notion de variables aléatoires

Soit X, une variable aléatoire :



Notion de variables aléatoires

Exemple :

On lance un dé à 6 faces :

- *Si on obtient 6 ou 5, on gagne 8000 Ar*
- *Si on obtient 4, on gagne 2000 Ar*
- *Si on obtient 3, on ne gagne rien*
- *Sinon, on perd 10 000 Ar*

Notons X la variable aléatoire égale au gain algébrique

Les valeurs possibles de X :

$$\left.\begin{array}{lcl} X\,(6) & = & 8\,000 \\ X\,(5) & = & 8\,000 \\ X\,(4) & = & 2\,000 \\ X\,(3) & = & 0 \\ X\,(2) & = & -10\,000 \\ X\,(1) & = & -10\,000 \end{array}\right\} \quad X \in \{-10\,000; 0; 2\,000; 8\,000\}$$

Notion de variables aléatoires

Exemple :

On lance un dé à six faces. Soit X la variable aléatoire égale au double de la valeur indiqué par la face

<u>Les valeurs possibles de X :</u>

$$\left.\begin{array}{lcl} X\ (1) & = & 2 \\ X\ (2) & = & 4 \\ X\ (3) & = & 6 \\ X\ (4) & = & 8 \\ X\ (5) & = & 10 \\ X\ (6) & = & 12 \end{array}\right\} \qquad X \in \{2; 4; 6; 8; 10; 12\}$$

Notion de variables aléatoires

Exemple :

Vous jouez à pile ou face avec votre ami. Si le résultat est pile, vous devez 10 000 Ar à votre ami, sinon, il vous devra 10 000 Ar. Soit X la variable aléatoire indiquant la somme que vous gagnerez au cours de ce jeu.

<u>Les valeurs possibles de X :</u>

$$\left.\begin{array}{l} \text{X (Pile)} = -10\,000 \\ \text{X (face)} = 10\,000 \end{array}\right\} \quad X \in \{-10\,000; 10\,000\}$$

Notion de variables aléatoires

Intérêt des variables aléatoires

1.Permet la création d’une fonction (loi de probabilité) permettant de décrire l’ensemble de l’expérience aléatoire

2. Permet une analyse poussée de l’expérience aléatoire

Variables aléatoires réelles

Définition

Dans un espace probabilité, une **variable aléatoire réelle** est une application de Ω dans $\mathbb{R}$ telle que : pour tout intervalle I de $\mathbb{R}$, l'ensemble $\boldsymbol{X^{-1}(I) = \{\omega \in \Omega | X(\omega) \in I\}}$ est un évènement d'une tribu $\mathcal{A}$.

Notons $\boldsymbol{X(\Omega)}$ **l'ensemble des valeurs prises** par la variable aléatoire X définie sur l'espace probabilisé.

$X^{-1}(\{x\}) = \{\omega \in \Omega | X(\omega) = x\}$ se note $\{X = x\}$ avec $x \in I$

$X^{-1}(\ ]-\infty, a]\ ) = \{\omega \in \Omega | X(\omega) \leq a\}$ se note $\{X \leq a\}$

$X^{-1}(\ ]a, b]\ ) = \{\omega \in \Omega | a < X(\omega) \leq b\}$ se note $\{a < X \leq b\}$

Variables aléatoires réelles

Fonction de répartition

La fonction de répartition d'une variable aléatoire réelle X, notée $F_x(x)$ est une fonction mathématique qui donne la probabilité cumulée que X soit inférieur ou égal à x.

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

Propriétés

1. $\forall x \in \mathbb{R}, \quad F_X(x) \in [0,1]$

2. F_X est une fonction croissante sur $\mathbb{R}$

3. $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$

4. Pour tous réels a et b, $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

Variables aléatoires discrètes

Définition

Une variable aléatoire est dite discrète (v.a.d) X si l'ensemble de ses valeurs $X(\Omega)$, est au plus dénombrable. En d'autres termes, si on a un évènement pour une valeur réelle.

Loi d'une variable aléatoire discrète

La loi de probabilité d'une v.a.d décrit l'attribution des valeurs possibles des probabilités à chacune des valeurs possibles de la variable. Elle se caractérise par une fonction de probabilité qui associe à chaque valeur prise par la variable sa probabilité correspondante:

$$\mathbb{P}(X = k) = p_X(k) \text{ telle que}$$

1. $\forall k \in \mathrm{X}(\Omega), \quad p_X(k) \geq 0$

2. $\sum_{k \in X(\Omega)} p_X(k) = 1$

3. Pour tout réel x, $\mathbb{P}(X \leq x) = \sum_{k \leq x} p_X(k)$ où $\sum_{k \leq x}$ désigne la sommation sur l'ensemble des $k \in X(\Omega)$ inférieurs ou égaux à x.

Variables aléatoires discrètes

Exemple

Soit X une variable aléatoire discrète prenant les valeurs 2, 4, 6, 8. Déterminez la loi de probabilité de X sachant que :

$$\mathbb{P}(X < 6) = \frac{1}{3}, \mathbb{P}(X > 6) = \frac{1}{2}, \mathbb{P}(X = 2) = \mathbb{P}(X = 4)$$

Exercice 1 (durée: 15 min)

Un joueur dispose d'un dé cubique truqué de telle sorte que la probabilité d'apparition d'un numéro est proportionnelle à ce dernier. Nous supposons que les faces sont numérotées de 1 à 6. Soit X la variable aléatoire réelle associée au lancer de ce dé.

1. Déterminez la loi de X.
2. Nous posons $Y = \frac{1}{X}$. Déterminez la loi de Y.

Exercice 2 (durée: 20 min)

Soit $a \in \mathbb{R}$ et X une v.a.d à valeurs dans $\mathbb{N}$, telle que :

$$\forall k \in \mathbb{N}, \mathbb{P}(X = k) = \frac{a}{2^k \times k!}$$

Déterminez a

<u>Indice:</u>

$$e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$$

Moments et fonction génératrice d’une v.a.d

Moments d’une variable aléatoire discrète

1.Espérance mathématique

La variable aléatoire X **admet une espérance mathématique** lorsque I (l’intervalle) est fini ou lorsque la série $\sum_i x_i p_X(x_i)$ est absolument convergente (tend vers une valeur finie).

L’Espérance mathématique, notée $\mathbb{E}(x)$ est la moyenne pondérée définie par :

$$\mathbb{E}(x) = \sum_i x_i p_X(x_i)$$

C'est une mesure centrale qui décrit la "valeur moyenne attendue" de la variable aléatoire après un grand nombre de répétitions de l'expérience aléatoire.

Moments et fonction génératrice d'une v.a.d

Moments d'une variable aléatoire discrète

1.Espérance mathématique

On lance un dé à 6 faces :

- *Si on obtient 6 ou 5, on gagne 8000 Ar*
- *Si on obtient 4, on gagne 2000 Ar*
- *Si on obtient 3, on ne gagne rien*
- *Sinon, on perd 10 000 Ar*

Notons X la variable aléatoire égale au gain algébrique

x_i	-10 000	0	2 000	8 000
$\mathbb{P}(X = x_i)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

Nous avons alors la valeur de l'espérance : $\mathbb{E}(x) = -1\ 666\ Ar$. Donc sur le long terme, jouer au jeu fera perdre au joueur la somme de 1 666 Ar.

Moments et fonction génératrice d’une v.a.d

Moments d’une variable aléatoire discrète

2.Espérance d’une fonction d’une variable aléatoire

Soit g une fonction définie sur $X(\Omega)$ à valeur dans $\mathbb{R}$. Lorsque $g(X)$ ademt une espérance mathématique, c’est-à-dire lorsque $\mathbb{E}\big(g(X)\big)$ est absolument convergente, nous avons :

$$\mathbb{E}(g(X)) = \sum_i g(x_i)p_X(x_i)$$

Exemple

Considérons un jeu de dés. On lance un dé équilibré à 6 faces, et la variable aléatoire X représente le résultat obtenu ($X \in \{1,2,3,4,5,6\}$). Chaque face a la même probabilité d’apparition. Supposons que la fonction d’intérêt soit $g(X) = X^2$, qui représente le résultat obtenu. Calculer l’espérance de la fonction $g(X)$.

$$\mathbb{E}\big(g(X)\big) = \sum_{k=1}^{6} g(X)\mathbb{P}(X = k) = \frac{1}{6}(91) = 15.17$$

Moments et fonction génératrice d'une v.a.d

Moments d'une variable aléatoire discrète

2.Linéarité de l'opérateur $\mathbb{E}$

L'espérance est une forme linéaire : pour tous réels a et b pour toutes variables aléatoires discrètes X et Y admettant une espérance mathématique, nous avons:

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

3. Moment simple et moment centré d'ordre $\boldsymbol{r}$

Soit $r \in \mathbb{N}^*$, sous réserve d'existence**, le moment simple d'ordre $\boldsymbol{r}$** de la variable aléatoire X désigne la valeur :

$$\mu_r(X) = \mathbb{E}(X^r)$$

Soit $r \in \mathbb{N}^*$, sous réserve d'existence**, le moment centré d'ordre $\boldsymbol{r}$** de la variable aléatoire X désigne la valeur :

$$\mu m_r(X) = \mathbb{E}((X - \mu)^r)$$

Moments et fonction génératrice d'une v.a.d

Moments d'une variable aléatoire discrète

Variance et écart-type

La variable aléatoire discrète X admet une variance lorsque X^2 admet une espérance. Nous appelons alors la variance de X la valeur:

$$Var(X) = \mathbb{E}\left((X-\mu)^2\right) = \mu m_2(x)$$

L'écart type de X n'est autre que la valeur $\sigma(x) = \sqrt{Var(x)}$

Formule de Huygens

Si la condition précédente est remplie, nous pouvons appliquer la formule de Huygens pour le calcul de la variance:

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

Vocabulaire

Si $\mathbb{E}(X) = 0$, alors X est une variable aléatoire centrée

Si $\mathrm{Var}(X) = 1$ alors X est une variable aléatoire réduite

Moments et fonction génératrice d'une v.a.d

Moments d'une variable aléatoire discrète

3.Moments d'ordres 3 et 4

Le moment centré d'ordre 3, lorsqu'il existe, de la v.a.d, X dont $\sigma > 0$ fournit **le coefficient d'asymétrie**

$$\frac{\mathbb{E}\left((X-\mu)^3\right)}{\sigma^3}$$

Le moment centré d'ordre 4, lorsqu'il existe, de la v.a.d, X dont $\sigma > 0$ fournit **le coefficient d'aplatissement**

$$\frac{\mathbb{E}\left((X-\mu)^4\right)}{\sigma^4} - 3$$

Variables aléatoires continues

Définitions

Une Variable aléatoire continue est une variable aléatoire qui peut prendre un nombre infini de valeurs dans un intervalle donné. Ces valeurs sont généralement représentées sur une plage continue, comme les nombres réels entre deux bornes.

Fonction de répartition de la variable X

Nous avons la fonction suivante : $F_x(x) = \int_{-\infty}^{x} f_x(t)dt$ avec f_x la densité de probabilité de X. La densité correspond à une plage de probabilité possible.

Remarque

Il faut en particulier que $\int_{-\infty}^{+\infty} f_x(t)dt$ soit convergente et que $\int_{-\infty}^{+\infty} f_x(t)dt = 1$

Propriétés

Soit une variable aléatoire continue admettant une densité de probabilité f_x

- Pour tout $a \in \mathbb{R}, \mathbb{P}([X = a]) = 0$;
- Pour tout (a, b) avec $-\infty \leq a \leq b \leq +\infty$, nous avons :

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f_x(t)dt$$

Statistique descriptive: Statistique univariée

X : un caractère quantitatif simple

$$X(\Omega) = (X_1, X_2, \dots, X_N) \quad \Longrightarrow \quad [x_1, x_2, \dots, x_n]$$

Distribution statistique discrète : valeurs considérées une par une

1. Effectif de la valeur $\boldsymbol{x_i}$: le nombre n_i de fois que la valeur x_i est prise. Il s'agit du cardinal de l'ensemble $X^{-1}(x_i)$

2. Effectif cumulé en $\boldsymbol{x_i}$: la somme $\sum_{j=1}^{i} n_j$

3. Fréquence de la valeur $\boldsymbol{x_i}$: le rapport $f_i = \frac{n_i}{N}$ de l'effectif de x_i à l'effectif total N de la population

4. Fréquence cumulée en $\boldsymbol{x_i}$:la somme $\sum_{j=1}^{i} f_i$

Nous pouvons avoir une distribution statistique discrète : $(\boldsymbol{x_i}, \boldsymbol{n_i})_{\boldsymbol{i=1,\dots,p}}$

Statistique descriptive: Statistique univariée

Exemple

Soit la liste de notes des étudiants : 10, 9, 12, 11, 10, 8, 14, 11, 9, 16, 5, 12, 10, 11, 10, 13

Tri par ordre croissant: 5, 8, 9, 9, 10, 10, 10, 10, 11, 11, 11, 12, 12, 13, 14, 16

Effectif de la valeur

Notes x_i	5	8	9	10	11	12	13	14	16	Total
Effectifs	1	1	2	4	3	2	1	1	1	16

Série statistique : (5,1) ; (8,1) ; (9,2) ; (10,4) ; (11,3) ; (12,2) ; (13,1) ; (14,1) ; (16,1)

Effectif cumulé

Notes x_i	5	8	9	10	11	12	13	14	16	Total
EC	1	2	4	8	11	13	14	15	16	16

Statistique descriptive: Statistique univariée

Exemple

Fréquence de la valeur

Notes x_i	5	8	9	10	11	12	13	14	16	Total
$f_i(\%)$	6.25	6.25	12.5	25	18.75	12.5	6.25	6.25	6.25	100

Fréquence cumulée

Notes x_i	5	8	9	10	11	12	13	14	16	Total
$\sum f_i$	6.25	12.5	25	50	68.75	81.25	87.5	93.75	100	100

Statistique descriptive: Statistique univariée

Distribution statistique groupée : valeurs regroupées en intervalles appelées **classes**

$$]a_i; a_{i+1}] \text{ avec } a_i < a_{i+1}$$

1. Effectif de la valeur $]\boldsymbol{a_i}; \boldsymbol{a_{i+1}}]$: le nombre n_i de fois que la valeur $]a_i; a_{i+1}]$ est prise. Il s'agit du cardinal de l'ensemble $X^{-1}(]a_i; a_{i+1}])$

2. Effectif cumulé **en $\mathbf{a_i}$:** la nombre de valeurs prises dans l'intervalle $]-\infty, a_i]$

3. Fréquence de $]\boldsymbol{a_i}; \boldsymbol{a_{i+1}}]$: : le rapport $f_i = \frac{n_i}{N}$

4. Fréquence cumulée **en $\boldsymbol{a_i}$** :la somme $\sum_{j=1}^{i} f_i$

Nous pouvons avoir une distribution statistique groupée : $(\,]\boldsymbol{a_i}; \boldsymbol{a_{i+1}}], \boldsymbol{n_i})_{\boldsymbol{i=1,\dots,p}}$

Statistique descriptive: Statistique univariée

Distribution statistique groupée : méthode de division en classe

Pour K **le nombre de classe,** et N **le nombre de données totales**

1. **Règle de Sturges ($30 \leq \boldsymbol{N} \leq \boldsymbol{200}$)**

$$\mathrm{K} \approx 1 + \frac{10}{3}\log_{10} N$$

2. **Règle de Rice ($\boldsymbol{N} > \boldsymbol{200}$)**

$$\mathrm{K} \approx 2 \times N^{\frac{1}{3}}$$

3. **Calcul de la racine carrée ($\boldsymbol{N} < \boldsymbol{30}$)**

$$\mathrm{K} = \sqrt{N}$$

Statistique descriptive: Statistique univariée

Distribution statistique groupée : méthode de division en classe

Exemple de calcul de classe

Effectif total	Racine carrée	Règle de Rice	Méthode de Sturges
25	5	5.84	5.65
64	8	7.999	7.02
4000	63.24	31.74	13.00

Statistique descriptive: Statistique univariée

Distribution statistique groupée : méthode de division en classe

Intervalle de classe

1. **Calcul de l'étendue $\boldsymbol{e}$**

$$e = valeur\ maximale\ - valeur\ minimale$$

2. **Calcul de l'amplitude théorique $\boldsymbol{h}$**

$$h = \frac{e}{K}$$

3. **Définition des intervalles**

$$]a_i; a_i + h]$$

Statistique descriptive: Statistique univariée

Distribution statistique discrète : Représentation graphique

1. Diagramme en bâtons : diagramme en bâtons d'effectif (resp. fréquences) d'une D.S.D est une suite de bâtons d'abscisses x_i et de longueur n_i (resp. f_i)

2. Polygone des effectifs (resp. fréquences) : s'obtient en joignant par un segment les sommets des bâtons

3. Polygone des effectifs (resp. fréquences) cumulés: s'obtient en joignant par un segment les sommets du diagramme en bâtons à partir des effectifs (resp. fréquences) cumulés.

Statistique descriptive: Statistique univariée

Distribution statistique discrète : Représentation graphique

Exemple

Soit la distribution statistique discrète : (1,2) ; (2,3) ; (3,4) ; (4,1) ; (5,6) ; (6,5) ; (7,2) ; (8,3) ; (9,1) ; (10,1)

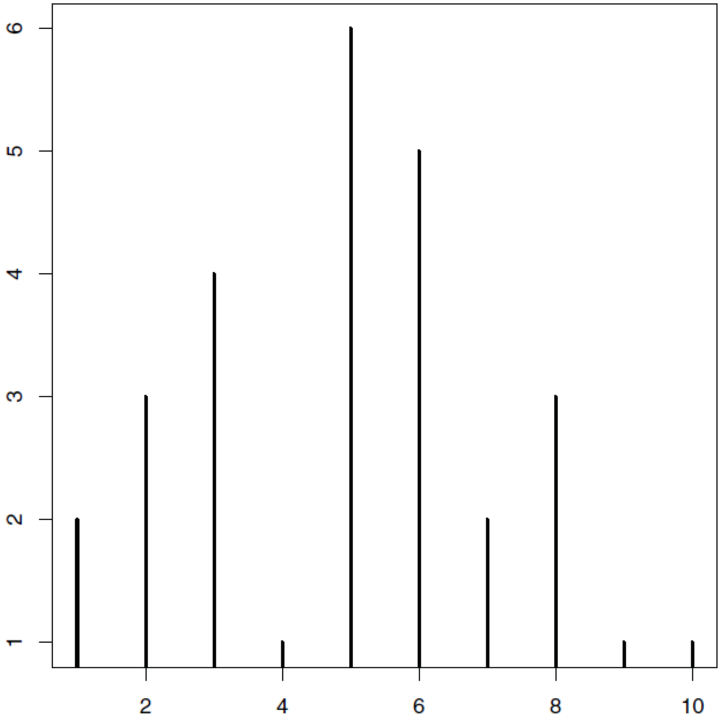
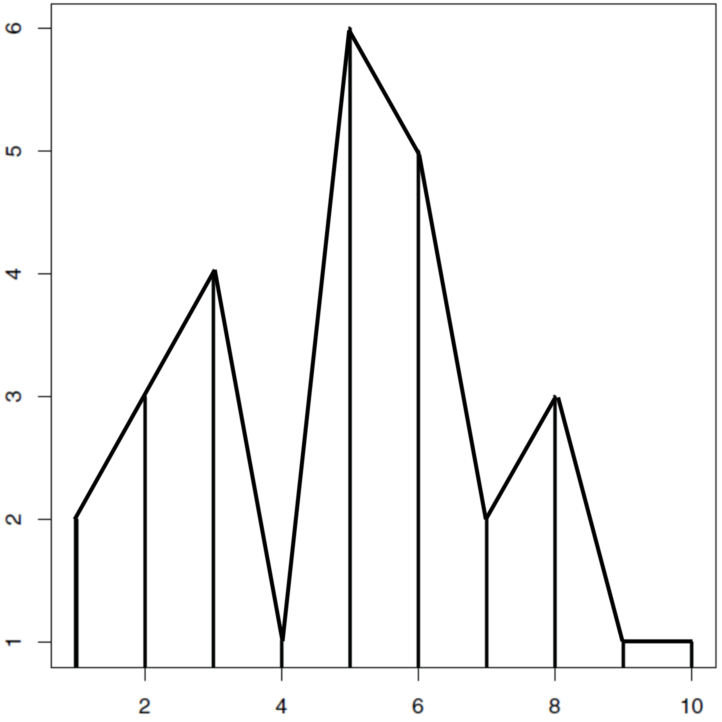
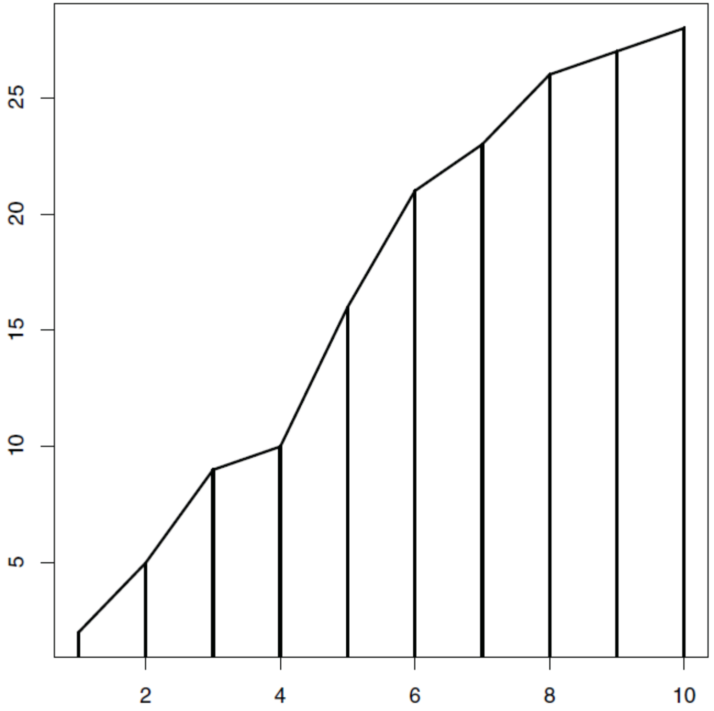


Diagramme en bâtons



Polygone des effectifs



Polygone des effectifs cumulés

Statistique descriptive: Statistique univariée

Distribution statistique groupée : Représentation graphique

1. Histogramme: il s’agit de l’équivalent du diagramme en bâtons d’effectif (resp. fréquences) en fonction des différentes classes définies sur l’axe des abscisses.

2. Polygone des effectifs (resp. fréquences) : s’obtient en joignant les milieux des côtés horizontaux supérieur des rectangles constituant l’histogramme.

3. Polygone des effectifs (resp. fréquences) cumulés: s’obtient en joignant par un segment les sommets des côtés horizontaux supérieur des rectangles constituant l’histogramme des effectifs (resp. fréquences) cumulés.

Statistique descriptive: Statistique univariée

Distribution statistique groupée : Représentation graphique

Exemple

Soit la distribution statistique discrète : (]1,3],4) ; (]3,4],8) ; (]4,5.5],10) ; (]5.5,6],14) ; (]6,8],20) ; (]8,10],12) ; (]10,11],12) ; (]11,12.5],3)

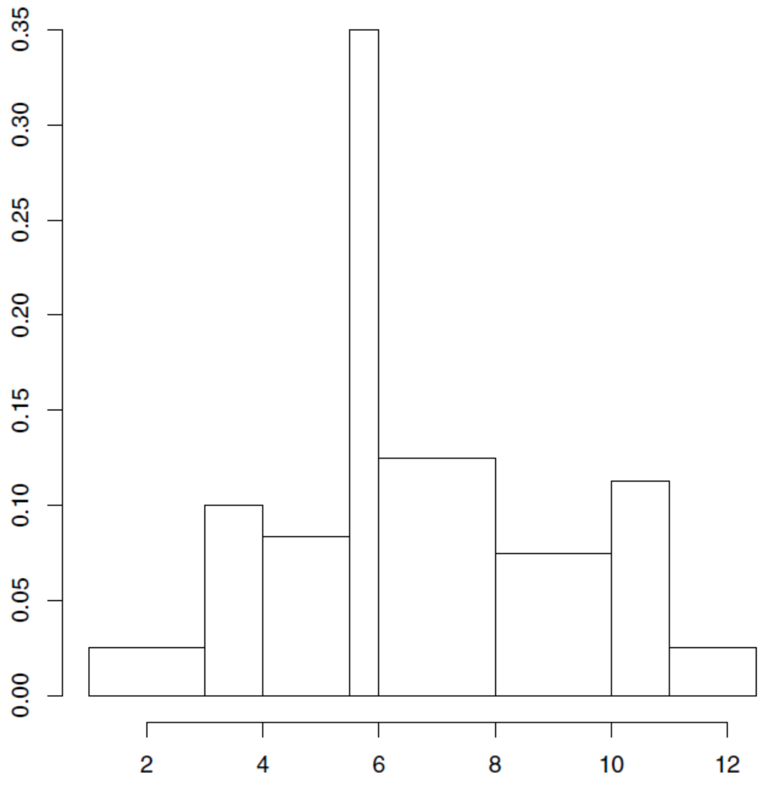
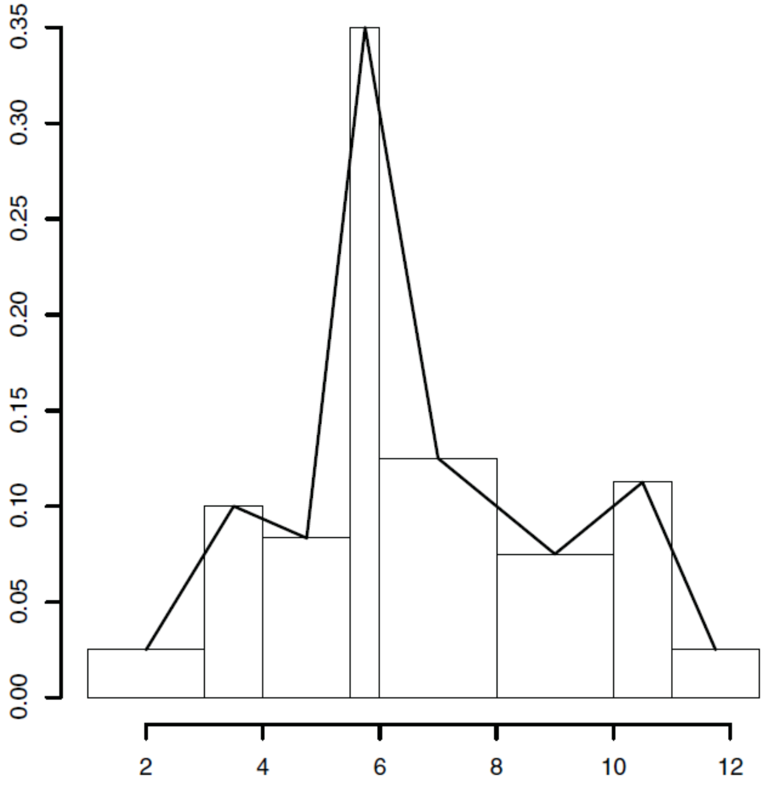
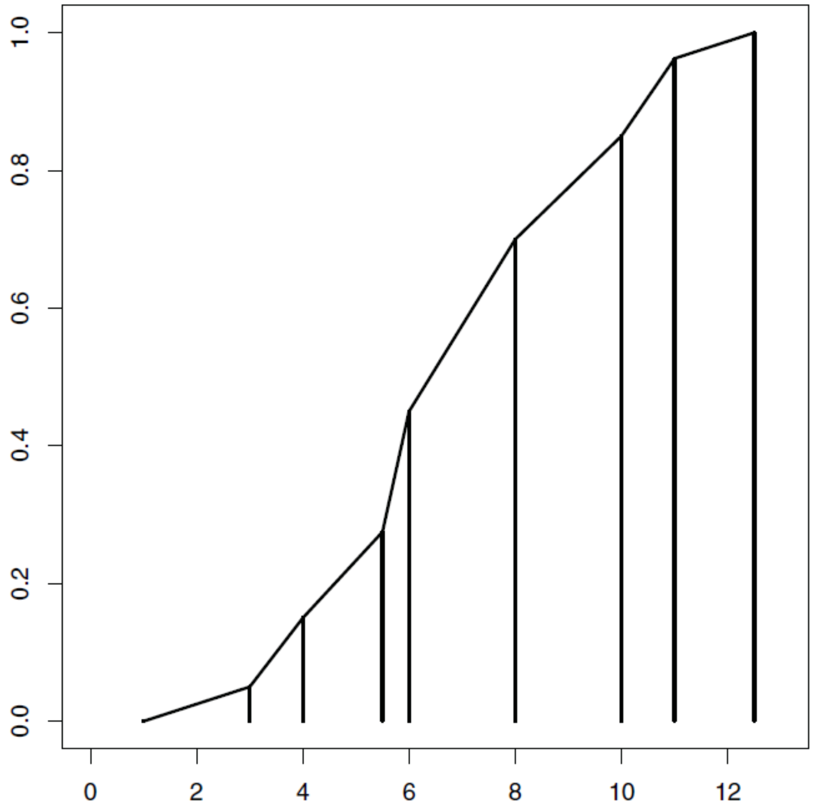


Diagramme en bâtons



Polygone des fréquences



Polygone des fréquences cumulées

Statistique descriptive: Statistique univariée

Mesure de tendance centrale ou caractéristique de position

1. Mode et classe modale :

Distribution statistique discrète : le mode est l'une des valeurs $x_1, x_2, \dots, x_p$ dont la fréquence est maximale

Exemple : 3, 7, 2, 7, 7, 4, 3. Le mode est **7** car cette valeur apparaît 3 fois.

Distribution statistique groupée: la classe modale est la classe $]a_i; a_{i+1}]$ avec l'effectif le plus élevé.

Exemple :

Intervalle	Fréquence
$]0; 10]$	4
$]10; 20]$	7
$]20; 30]$	12
$]30; 40]$	5

$]\mathbf{20}; \mathbf{30}]$ est la classe modale car elle a l'effectif le plus élevé

Statistique descriptive: Statistique univariée

Mesure de tendance centrale ou caractéristique de position

1. Mode et classe modale :

La distribution peut être **unimodale (un seul mode)** ou **plurimodale (plurimodale)**

Exemple : classe bimodale (deux modes)

Intervalle	Fréquence
$]0;10]$	4
$]10;20]$	7
$]20;30]$	12
$]30;40]$	5
$]40;50]$	12

$]\mathbf{20};\mathbf{30}]$ et $]\mathbf{40};\mathbf{50}]$ sont les classes modales

Statistique descriptive: Statistique univariée

Mesure de tendance centrale ou caractéristique de position

1. Médiane:

La médiane d'une série statistique est la valeur qui divise la série en 2 parties égales.

2. Quartiles:

Premier quartile : représente les 25% de la série statistique.

Troisième quartile : représente les 75% de la série statistique.

3. Moyenne arithmétique:

La moyenne $\mu(x)$ ou simplement notée μ est la valeur notée :

$$\mu(x) = \sum_{i=1}^{p} x_i f_i = \left(\frac{1}{N}\right) \sum_{i=1}^{p} x_i n_i$$

$$\mu(x) = \sum_{i=1}^{p} \frac{a_i + a_{i+1}}{2} f_i$$

Statistique descriptive: Statistique univariée

Mesure de dispersion

1. Etendue

l'étendue e est la différence entre la valeur maximale et minimale de la série : $e = \max(x) - \min(x)$

2. Etendue interquartile

L'étendue interquartile, notée $EIQ(x)$ est la différence entre le troisième et le premier quartile : $EIQ(x) = Q_3(x) - Q_1(x)$

3. Variance et écart-type

La variance, notée $\mathrm{Var}(x)$ est une mesure de dispersion qui indique à quel point les données sont éparpillées autour de la moyenne. **L'écart-type**, notée $\sigma(x)$ de même unité que les données de la série, permet d'interpréter la dispersion d'une manière plus intuitive

$$\mathrm{Var}(x) = \sum_{i=1}^{p} \left(x_i - \mu(x)\right)^2 f_i \qquad \sigma(x) = \sqrt{Var(x)}$$

Formule de Huygens: $\mathrm{Var(x)} = \mu(x^2) - \left(\mu(x)\right)^2$ où $\mu(x^2)$ représente la moyenne du carré des valeurs de la série statistique

Statistique descriptive: Statistique univariée

Mesure de dispersion

Exemple

Supposons les notes (/20) suivantes : 10, 12, 14, 16, 18

Calcul de la moyenne : $\mu(x) = \frac{10+12+14+16+18}{5} = 14$

Calcul de la variance : $Var(x) = \frac{(10-14)^2+(12-14)^2+\cdots+(18-14)^2}{5} = 8$

Calcul de l'écart-type :$\sigma(x) = \sqrt{8} \approx 2.83$

Interprétation

Les notes s'écartent en moyenne de 2,84 points autour de la moyenne. La performance des étudiants est plutôt régulière

Statistique descriptive: Statistique univariée

Mesure de dispersion

4. Moment centré d'ordre r

Le moment centré d'ordre r d'une série statistique est une mesure qui décrit la distribution des données autour de la moyenne.

$$\mu m_r(x) = \sum_{i=1}^{p} \left(x_i - \mu(x)\right)^r f_i$$

Pour $r = 2$, il s'agit de la variance

Pour $r = 3$, il s'agit de l'étude de l'asymétrie

Pour $r = 4$, il s'agit de l'étude de l'aplatissement (kurtosis)

Statistique descriptive: Statistique univariée

Mesure de dispersion

Etude de l'asymétrie

Si $\mu m_3(x) = 0$ alors la distribution est symétrique

Si $\mu m_3 < 0$ alors la série est asymétrique à gauche de la moyenne

Si $\mu m_3 > 0$ alors la série est asymétrique à droite de la moyenne

Exemple:

Pour la série statistique suivante : {2,4,6,8}, nous avons $\mu m_3(x) = 0$ donc la série est symétrique (pas de 2 entre chaque valeur)

Pour la série statistique suivante : {2,4,6,10}, nous avons $\mu m_3(x) = 11.25$ donc la série est asymétrique à droite

Pour la série statistique suivante : {2,6,8,10}, nous avons $\mu m_3(x) = 11.25$ donc la série est asymétrique à droite

Statistique descriptive: Statistique univariée

Mesure de dispersion

Etude de l’aplatissement

Si $\mu m_4(x) = 3$ alors la distribution est standardisée

La série est dite platykurtique si elle est aplatie, valeur plus ou moins uniforme

La série est dite leptokurtique si elle est pointue, présence de pics élevés de valeur

Exemple:

Pour la série statistique suivante : {2,4,6,8,20}, nous avons $\mu m_4(x) = 15.6$ donc la série est leptokurtique

Pour la série statistique suivante : {2,4,6,10}, nous avons $\mu m_3(x) = 3.7$ donc la série est platykurtique (plus ou moins aplatie)

Statistique descriptive: Statistique univariée

Mesure de dispersion

5.Coefficient de variation

Le coefficient de variation (ou écart-type relatif) est une mesure de dispersion relative.

$$CV = \frac{\sigma}{\mu}$$

Statistique descriptive: Statistique univariée

Mesure de dispersion

Soit le tableau, représentant le Nombre d’heures travaillées par semaine des personnes ayant un emploi à plein temps dans les États membres de l’UE

Pays	Durée (heures)	Pays	Durée (heures)
Allemagne	41,70	Lettonie	43,00
Autriche	44,10	Lituanie	39,80
Belgique	41,00	Luxembourg	40,90
Chypre	41,80	Malte	41,20
Danemark	40,50	Pays-Bas	40,80
Espagne	42,20	Pologne	42,90
Estonie	41,50	Portugal	41,60
Finlande	40,50	République Tchèque	42,70
France	**41,00**	Royaume-Uni	43,10
Grèce	44,10	Slovaquie	41,60
Hongrie	41,00	Slovénie	42,50
Irlande	40,70	Suède	41,10
Italie	41,30		

Effectuer une étude de la tendance centrale (finir par un diagramme boite à moustache) et une étude de la dispersion des données. Interprétez les résultats obtenus.

Statistique descriptive: Statistique bivariée

Soit (X,Y) une distribution statistique d'un couple de caractères sur une population d'effectif N. Notons $(x_1, x_2, \dots, x_n)$ et $(y_1, y_2, \dots, m)$ les valeurs distinctes pour X et Y ordonnées dans l'ordre croissant.

Exemples

X : « âge d'une personne », Y: « Revenu annuel de cette personne »

X : «Température d'une pièce», Y: « Consommation en énergie»

X : «Taille d'une personne», Y: « Poids de la même personne»

Statistique descriptive: Statistique bivariée

Soit (X,Y) une distribution statistique d'un couple de caractères sur une population d'effectif N. Notons $(x_1, x_2, \dots, x_n)$ et $(y_1, y_2, \dots, y_n)$ les valeurs distinctes pour X et Y ordonnées dans l'ordre croissant.

1.Distribution conjointe

Effectif du couple

L'effectif du couple (x_i, y_i) est égal au nombre $n_{i,j}$ de couple de valeurs égaux à (x_i, y_i)

Fréquence du couple

La fréquence du couple(x_i, y_i) est égal au nombre $f_{i,j} = n_{i,j}/N$

Distributions des effectifs et des fréquences

Elle est notée : $((x_i, y_i), n_{i,j})$ et $((x_i, y_i), f_{i,j})$

Statistique descriptive: Statistique bivariée

2.Distribution marginales

Effectif et fréquences marginaux

L'effectif marginal de x_i(respectivement de y_i) est égal au nombre $n_{i,.} = \sum_{j=1}^{q} n_{i,j}$ (respectivement $n_{.,.j} = \sum_{j=1}^{q} n_{i,j}$). Le même raisonnement est appliqué pour le calcul de la fréquence marginale.

3.Indépendance et distributions conditionnelles

Indépendance

Y est indépendant de X si la répartition des valeurs possibles de Y est la même quelle que soit la valeur de X.

Statistique descriptive: Statistique bivariée

Exemple

Soit X le jour de la semaine, et Y la couleur d’un ballon choisit au hasard

Cas d’indépendance

Jour	Rouge	Bleu	Vert	Total
Lundi	50	30	20	100
Mardi	50	30	20	100
Mercredi	50	30	20	100

Cas de dépendance

Jour	Rouge	Bleu	Vert	Total
Lundi	60	20	20	100
Mardi	30	50	20	100
Mercredi	40	30	30	100

Statistique descriptive: Statistique bivariée

Fréquence conditionnelle

La fraction $\frac{n_{i,j}}{n_{i,.}} = \frac{f_{i,j}}{f_{i,.}}$ représente **la fréquence conditionnelle** de la valeur y_i sachant x_i

Distribution conditionnelle des fréquences

La distribution $\left(y_i, \frac{f_{i,j}}{f_{i,.}}\right)$ est appelée la **distribution conditionnelle des fréquences** de Y sachant que $X = x_i$.

4.Covariance et coefficient de corrélation linéaire

Covariance du couple $(\boldsymbol{X}, \boldsymbol{Y})$

La covariance permet de déterminer dans quelle mesure les deux variables varient ensemble.

$$\mathbb{C}ov(X, Y) = \sum_{i,j} \left(x_i - \mu(X)\right) \times \left(y_j - \mu(Y)\right) f_{i,j}$$

$$\mathbb{C}ov(X, Y) = \mu(XY) - \mu(X)\mu(Y)\ avec\ \mu(XY) =$$

Statistique descriptive: Statistique bivariée

Exemple

Soit le tableau statistique suivant :

X	Y
1	2
2	4
3	6
4	8

Calcul de la fréquence relative : $f_{i,j} = \frac{n_{i,j}}{N} = \frac{1}{4} = 0.25$

Valeur de la covariance: $\mathbb{C}ov(X,Y) = 2.5$

Coefficient de corrélation linéaire

$$r(X,Y) = \frac{\mathbb{C}ov(X,Y)}{\sigma(X)\sigma(Y)}$$