Institut de Recherche sur l'Enseignement des Mathématiques

Université Montpellier II

Petit lexique de statistique et probabilité

Groupe statistique et probabilité de l'IREM de MONTPELLIER

Michèle ALDEBERT, Noël BASCOU,

Françoise DELATOUR, Jean-Marie SCHADECK

Décembre 2007

Institut de Recherche sur l'Enseignement des Mathématiques

Université Montpellier II

Petit lexique de statistique et probabilité

Groupe statistique et probabilité de l'IREM de MONTPELLIER

Michèle ALDEBERT, Noël BASCOU,

Françoise DELATOUR, Jean-Marie SCHADECK

Décembre 2007

Sommaire

Présentation du groupe et du document

Lors des différents stages de statistique et probabilité que nous avons animés ces dernières années, de nombreuses questions sur des notions de statistique revenaient de façon récurrente. C'est dans le but de répondre à ces questions que dans ce document, nous proposons un « lexique de statistique », qui n'étant pas exhaustif traite les notions qui nous paraissent importantes.

Pour chaque thème abordé, nous proposons :

- Une définition, suivie éventuellement d'une formulation plus simple pour les élèves.
- Les points de référence dans le programme.
- Son champ d'application et son intérêt (par rapport à d'autres notions voisines, ou d'autres qui lui sont liées).
- Un ou plusieurs exemples traités.
- Des compléments éventuels.

Ce travail s'adresse en premier aux professeurs de collège et de lycée, mais peut aussi intéresser des enseignants du supérieur, qui pourraient y trouver les points du programme de statistique et de probabilité enseignés au lycée.

Le document comporte trois chapitres : Statistique descriptive, Probabilité, Statistique inférentielle. Dans chaque chapitre, les thèmes sont classés par ordre alphabétique.

Les membres du groupe statistique de l'IREM de Montpellier ayant participé à la rédaction du document :

Michèle ALDEBERT (Lycée Jean Baptiste DUMAS – ALES)

Noël BASCOU (Lycée Jean MERMOZ – MONTPELLIER)

Françoise DELATOUR (Lycée Jean MERMOZ – MONTPELLIER)

Jean Marie SCHADECK (Lycée Jean MERMOZ – MONTPELLIER)

STATISTIQUE DESCRIPTIVE

Corrélation

Covariance

Diagrammes

Droite de régression

Ecart type

Effectif, effectif cumulé

Echantillon

Etendue

Fréquence, fréquence cumulée

Médiane

Mode

Moyenne arithmétique

Population

Quantiles

Variance

Corrélation linéaire

Tout ce qui suit ne concerne que la méthode des moindres carrés pour l'obtention d'une droite de régression linéaire, voir page 17.

Pour une série double $(x_i, y_i)_{i=1,\dots,n}$

$$s(x) = \sqrt{V(x)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$s(y) = \sqrt{V(y)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Définition
Le **coefficient de corrélation linéaire** est le rapport.

$$r = \frac{Cov(x,y)}{s(x)s(y)}$$ et on a donc $-1 \leq r \leq 1$ (cf covariance)

Remarques

- $r^2 = a \times a'$ où a est la pente de la droite de régression de y en x : $y = ax + b$
  et a' est la pente de la droite de régression de x en y : $x = a'y + b'$
- les droites de régression sont confondues si et seulement si $|r| = 1$
-

$r = -1$	$r < 0$	$r = 0$	$r > 0$	$r = 1$
Liaison fonctionnelle	Corrélation négative	Absence de corrélation linéaire	Corrélation positive	Liaison fonctionnelle

Dans la pratique si $|r|$ est proche de 1, on considère que l'ajustement est « justifié ».

- Selon le problème traité, la notion de forte corrélation est très variable.
- la corrélation est formelle, ce n'est pas un lien de causalité : on peut avoir une forte corrélation entre des caractères totalement étrangers.

Point de référence dans le programme

Terminales, BTS

Champs d'application

- Le coefficient de corrélation permet de valider un ajustement linéaire, lequel permet de faire des prévisions.
- Si un ajustement linéaire n'est pas justifié par la valeur du coefficient de corrélation, il pourra l'être pour la série obtenue après changement de variable. (voir exemple 2)

Exemple 1

x_i	10	12	18	20	25	28	32
y_i	3	5	9	11	15	17	20

$Cov(x,y) = 43,69$

$s(x) = 7,53$

$s(y) = 5,80$ *et* $r = \dfrac{Cov(x,y)}{s(x)s(y)} = 0,999$ (arrondis au 1/100)

L'ajustement linéaire est alors justifié.

Exemple 2

x_i	2	3	5	6	7	8	9
y_i	2	4	8	13	24	44	82

$Cov(x,y) = 54,37$

$s(x) = 2,37$

$s(y) = 26,76$ et $r = \dfrac{Cov(x,y)}{s(x)s(y)} = 0,86$, *qui peut-être jugé insuffisant,* mais

x_i	2	3	5	6	7	8	9
$\ln y_i$	0,69	1,39	2,08	2,56	3,18	3,78	4,41

$Cov(x, \ln y) = 2,88$

$s(x) = 2,37$

$s(\ln y) = 1,22$ *et* $r = \dfrac{Cov(x,\ln y)}{s(x)s(\ln y)} = 0,99$ (arrondis au 1/100)

Les variables x et *ln*y sont fortement corrélées.

On pourra en déduire une relation du type : $\ln y = ax + b$ d'où : $y = e^b e^{ax}$

Compléments

- $R = r^2 = \left(\dfrac{C\,\mathrm{ov}(x,y)}{s(x)s(y)}\right)^2 = \dfrac{V(\hat{y})}{V(y)}$ où *$\hat{y}$ désigne l'estimation affine de y*

  en effet : $V(\hat{y}) = V(ax+b) = a^2V(x)$

  donc $\dfrac{V(\hat{y})}{V(y)} = \dfrac{a^2V(x)}{V(y)} = \left(\dfrac{C\,\mathrm{ov}(x,y)}{V(x)}\right)^2 \times \dfrac{V(x)}{V(y)} = \dfrac{\left(C\,\mathrm{ov}(x,y)\right)^2}{V(x)V(y)} = r^2$

  le résidu quadratique moyen par rapport à la droite de régression de y en x est $V_R(y) = (1-r^2)V(y)$ *, c'est la **variance résiduelle** .*

- *La variance se décompose en* $V(y) = r^2V(y) + (1-r^2)V(y)$

  où $V_E(y) = r^2V(y)$ *est **la variance expliquée** par la droite de régression*

  $V_R(y) = (1-r^2)V(y)$ *est **la variance résiduelle***

  et $V(y) = V_E(y) + V_R(y)$

- *Plus la part de la variance expliquée est important , meilleur est le résumé du nuage de points par la droite de régression et plus forte est la corrélation*

Covariance

Définition

La **covariance Cov(*x*,*y*)** d'une série statistique double $(x_i, y_i)_{i=1,\dots,n}$

est la moyenne des produits des écarts des données aux moyennes.

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

C'est aussi la différence entre « la moyenne des produits » et « le produit des moyennes ».

$$Cov(x, y) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \overline{x}\,\overline{y}$$

.

Remarques

∞ On a aussi, si (par regroupement)

- $x_1, x_2, \dots, x_p$ désignent les p modalités du caractère x.
- $y_1, y_2, \dots, y_q$ désignent les q modalités du caractère y.
- n_{ij} désigne le nombre d'individus présentant à la fois la modalité x_i de X et la modalité y_j de Y.

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(x_i - \overline{x})(y_j - \overline{y}) \qquad \text{où } n = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}$$

Ou encore $$Cov(x, y) = \sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij}(x_i - \overline{x})(y_j - \overline{y}) \qquad \text{où } f_{ij} = \frac{n_{ij}}{n}$$

∞ C'est une forme bilinéaire symétrique positive (dégénérée) définie sur $\mathbb{R}^n$,et c'est la forme polaire associée à la forme quadratique définie sur $\mathbb{R}^n$ par $V(x) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$, c'est à dire la **variance.**

En particulier $Cov(x, y) = \frac{1}{2}\big(V(x+y) - V(x) - V(y)\big)$

L'inégalité de Cauchy-Schwarz donne alors

$$\left|Cov(x, y)\right| \leq \sqrt{V(x)V(y)}$$

Donc $\left|Cov(x, y)\right| \leq s(x)s(y)$

∞ Effet de transformations affines des données :

$$Cov(ax + b, a'y + b') = \frac{1}{n}\sum_{i=1}^{n} aa'(x_i - \overline{x})(y_i - \overline{y}) = aa'Cov(x, y)$$

∞ En particulier la covariance est invariante par translations :

$$Cov(x,y)=\left(\frac{1}{n}\sum_{i=1}^{n}x_iy_i\right)-\overline{x}\,\overline{y}=\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-a)(y_i-b)\right)-(\overline{x}-a)(\overline{y}-b)\quad \forall(a,b)\in\mathbb{R}^2$$
$$Cov(x,y)=Cov(x+a,y+b)$$

Point de référence dans le programme

Terminales, BTS

Champs d'application

Corrélation
Droite des moindres carrés
Indépendance

Exemple 1

x \ y	2	4	6	7	8	9	Dist. Marg. de x
3	*21*	*3*	*0*	*0*	*0*	*0*	*24*
4	*3*	*23*	*3*	*1*	*0*	*0*	*30*
5	*1*	*2*	*25*	*1*	*1*	*1*	*31*
6	*0*	*0*	*2*	*20*	*7*	*2*	*31*
7	*0*	*0*	*0*	*2*	*15*	*27*	*44*
Dist.Marg. de y	*25*	*28*	*30*	*24*	*23*	*30*	*160*

$\overline{x}=5,256$, $\overline{y}=6,025$ et $\sum_{i=1}^{5}\sum_{j=1}^{6}n_{ij}x_iy_j=5569$ donc

$$Cov(x,y)=\frac{5569}{160}-5,256\times 6,025=3,137 \quad \left(\text{ et ici } r=\frac{Cov(x,y)}{s(x)s(y)}\cong\frac{3,137}{1,419\times 2,369}\cong 0,933\right)$$

Complément 1 regroupements de 2 séries doubles de tailles n_1 et n_2

$$(x_i,y_i)_{i=1,\dots,n_1} \text{ et } (x_i,y_i)_{i=n_1+1,\dots,n_1+n_2}$$

la covariance des séries obtenues par regroupement est :

$$Cov(x,y) = \left(\frac{1}{n_1+n_2} \sum_{i=1}^{n_1+n_2} x_i y_i \right) - \overline{x}\,\overline{y} = \frac{1}{n_1+n_2} \left(\sum_{i=1}^{n_1} x_i y_i + \sum_{i=n_1+1}^{n_1+n_2} x_i y_i - (n_1+n_2)\overline{x}\,\overline{y} \right)$$

$$= \frac{1}{n_1+n_2} \left(\sum_{i=1}^{n_1} x_i y_i - n_1 \overline{x}_1 \overline{y}_1 + \sum_{i=n_1+1}^{n_1+n_2} x_i y_i - n_2 \overline{x}_2 \overline{y}_2 + n_1 \overline{x}_1 \overline{y}_1 + n_2 \overline{x}_2 \overline{y}_2 - (n_1+n_2)\overline{x}\,\overline{y} \right)$$

$$= \frac{1}{n_1+n_2} \left(n_1 Cov(x_{(1)}, y_{(1)}) + n_2 Cov(x_{(2)}, y_{(2)}) + n_1 \overline{x}_1 \overline{y}_1 + n_2 \overline{x}_2 \overline{y}_2 - (n_1+n_2)\overline{x}\,\overline{y} \right)$$

$$Cov(x,y) = \frac{n_1 Cov(x_{(1)}, y_{(1)}) + n_2 Cov(x_{(2)}, y_{(2)})}{n_1+n_2} + \left(\frac{1}{n_1+n_2} (n_1 \overline{x}_1 \overline{y}_1 + n_2 \overline{x}_2 \overline{y}_2) - \overline{x}\,\overline{y} \right)$$

c’est donc la somme de la moyenne des covariances et de la covariance des moyennes (formule facile à généraliser)

- la covariance des séries regroupées n’est donc pas la moyenne pondérée des covariances

Complément 2 : preuve de l’inégalité de Cauchy-Schwarz

$Cov(x,y) = \frac{1}{2}\left(V(x+y) - V(x) - V(y)\right)$

donc pour tout $a \in \mathbb{R}$

$Cov(ax,y) = \frac{1}{2}\left(V(ax+y) - V(ax) - V(y)\right)$

$V(ax+y) = V(ax) + 2Cov(ax,y) + V(y)$

$V(ax+y) = a^2 V(x) + 2aCov(x,y) + V(y)$ et donc

$a^2 V(x) + 2aCov(x,y) + V(y) \geq 0$ pour tout $a \in \mathbb{R}$

Donc le discriminant de l'expression, vue comme un trinôme en *a*, est nécessairement négatif :

$Cov^2(x,y) - V(y)V(x) \leq 0$ soit $|Cov(x,y)| \leq s(x)s(y)$

Dans le cas d’égalité :le discriminant est nul si et seulement si le trinôme s’annule pour $a = -\frac{Cov(x,y)}{V(x)}$

Donc pour cette valeur, $V(ax+y) = 0$

$$V(ax+y) = \frac{1}{n} \sum_{i=1}^{n} (ax_i + y_i - a\overline{x} - \overline{y})^2 = 0$$

$$ax_i + y_i - a\overline{x} - \overline{y} = 0 \quad \forall i \in [\![1,n]\!]$$

$$y_i = -a x_i + a\overline{x} + \overline{y} \quad \forall i \in [\![1,n]\!]$$

Donc l'inégalité ne peut être une égalité que s'il existe une relation affine entre y et x.

Diagrammes

Définition

Etant donné une série statistique, une représentation graphique de celle-ci permet d'en avoir une vision synthétique et globale.

Principales représentations :

Dans le cas d'une série statistique correspondant à une variable à caractère qualitatif ou quantitatif discret, les représentations les plus courantes sont :

- les **diagrammes en bâtons**, on porte en abscisse les valeurs du caractère et en ordonnée l'effectif du caractère correspondant.
- les **diagrammes circulaires**, le disque est découpé en secteurs dont l'aire est proportionnelle à l'effectif de la valeur considérée.

Dans le cas d'une série statistique correspondant à une variable à caractère quantitatif où les données sont regroupées en classes, la représentation la plus courante est **l'histogramme**.
On porte en abscisse les classes et on réalise des rectangles dont l'aire est proportionnelle à l'effectif de la classe.

Dans le cas d'une série statistique à deux variables (x_i , y_i) on représentera dans un repère les points de coordonnées (x_i , y_i) le diagramme ainsi obtenu est alors appelé **nuage de points**.

Dans le cas d'une série ordonnée à caractère qualitatif, on peut réaliser un **diagramme en boîte,** on porte sur un axe les valeurs extrêmes de la série ainsi que la médiane et les quartiles.
(voir exemple)

Points de référence dans le programme

Les classes de collège pour les diagrammes en bâtons ou circulaires.
Les classes de lycée pour les histogrammes, les diagrammes en boite ou les nuages de points.

Champs d'application

Permettre des résumés statistiques exploitables rapidement. On en trouve fréquemment dans la presse, les publications scientifiques, les ateliers de production, etc.
Aider au choix d'un modèle statistique pour rendre compte d'une situation.

Exemple 1 Soit la série statistique suivante :

xi	3	4	5	6	7	8
ni	2	5	8	9	5	3

Diagramme en bâton :

Diagramme circulaire:

Exemple 2 Soit la série de données regroupées par intervalle :

Classe	[0,2[	[2,4[	[4,6[	[6,8[	[8,14[
effectif : ni	5	8	12	6	6
effectif corrigé	5	8	12	6	2

Histogramme

Remarque : les effectifs corrigés sont portés en ordonnée.

Exemple 3 Soit la série statistique suivante:

xi	1	2	3	4	5	6	7	8	9	10
ni	4	14	16	20	18	8	8	6	4	2
ni croissant	4	18	34	54	72	80	88	94	98	100

médiane 4

quartile Q1 3

quartile Q2 6

Représentation graphique:
diagramme en boîte

Remarques

- *On peut aussi faire figurer les déciles D1 et D9.*
- *Ce type de représentation graphique permet de mettre en évidence une dissymétrie C'est le cas dans l'exemple ; la médiane et Q1 sont décalés vers les petites valeurs.*

.

Droite de régression
(Ajustement linéaire par la méthode des moindres carrés)

Définition

On étudie simultanément deux caractères quantitatifs, X et Y, des individus d'une population, ou d'un échantillon, dont les valeurs x et y ont été observées.

On a une série double $(x_i, y_i)_{i=1,\dots,n}$, ou regroupée $\left(n_{ij};\left(x_i, y_j\right)\right)_{i=1,\dots,p\ ;\ j=1,\dots,q.}$,

La méthode des moindres carrés permet de déterminer l'équation de la **droite de régression de *y* en *x*** : $y = ax + b$, où :

$$a = \frac{Cov(x, y)}{V(x)}$$

$$\text{et} \quad b = \overline{y} - \frac{Cov(x, y)}{V(x)} \times \overline{x}$$

Les valeurs de a et de b, rendent minimum la somme $S = \sum_{i=1}^{n} \left(y_i - ax_i - b\right)^2$.

Remarques

∞ L'équation de la droite de régression de x en y est

$$x = a'y + b' \quad \text{avec } a' = \frac{Cov(x, y)}{V(y)} \quad \text{et } b' = \overline{x} - \frac{Cov(x, y)}{V(y)} \times \overline{y} = \overline{x} - a'\overline{y}$$

∞ Ces deux droites passent par le point dit moyen $G(\overline{x}, \overline{y})$.

∞ **le coefficient de corrélation linéaire** est $r = \dfrac{\operatorname{cov}(x, y)}{s(x)s(y)}$

et on a donc $-1 \leq r \leq 1$ (cf covariance)

Dans la pratique, on commence par calculer *r*, si $|r|$ est proche de 1, on considère que l'ajustement est « justifié ».

∞ Si la corrélation est formelle, ce n'est pas nécessairement un lien de causalité : on peut avoir une forte corrélation entre des caractères totalement étrangers.

∞ les deux droites de régression sont d'autant plus proches que $|r|$ est proche de 1.

Point de référence dans le programme

Terminales, BTS.

Champs d'application

- une droite de régression est utilisée pour estimer la valeur d'un caractère connaissant la valeur de l'autre, ou de faire une prévision.
- Dans la pratique, l'un des caractères est observé et l'autre estimé. Ainsi si X est observé on utilise la droite de régression de y en x.
- Certains ajustements, non linéaires, se ramènent par changement de variables à un ajustement linéaire.

Exemple Forces de frottement sur une surface donnée

On a mesuré les forces de frottement, y_i, sur une surface pour des objets de poids différents, x_i.(à surface de contact constante)
Les résultats, en newton, sont portés dans le tableau suivant :

x_i	23	44	65	86	107	128	149	171
y_i	3,4	4,7	5,6	6,8	7,5	8,8	9,6	11,0

Donner une estimation de la force de frottement liée à un objet dont le poids est $55N$.

Réponse :

$\text{cov}(x,y) = 115,978...$ $\qquad V(x) = 2333,734...$ $\qquad V(y) = 5,781...$

$s(x) = 48,309...$

$s(y) = 2,4045...$

donc $r = \dfrac{\text{cov}(x,y)}{s(x)s(y)} = 0,9984...$ *ce qui justifie l'ajustement linéaire.*

$a = \dfrac{Cov(x,y)}{V(x)} = 0,0497$ *(au* $1/1000$*)* $\qquad \overline{x} = 96,625$ $\qquad \overline{y} = 7,175$

$b = \overline{y} - \dfrac{Cov(x,y)}{V(x)} \times \overline{x} = 2,373$ *(au* $1/1000$*)*

l'équation de la droite de régression de y en x est : $y = 0,0497x + 2,373$,

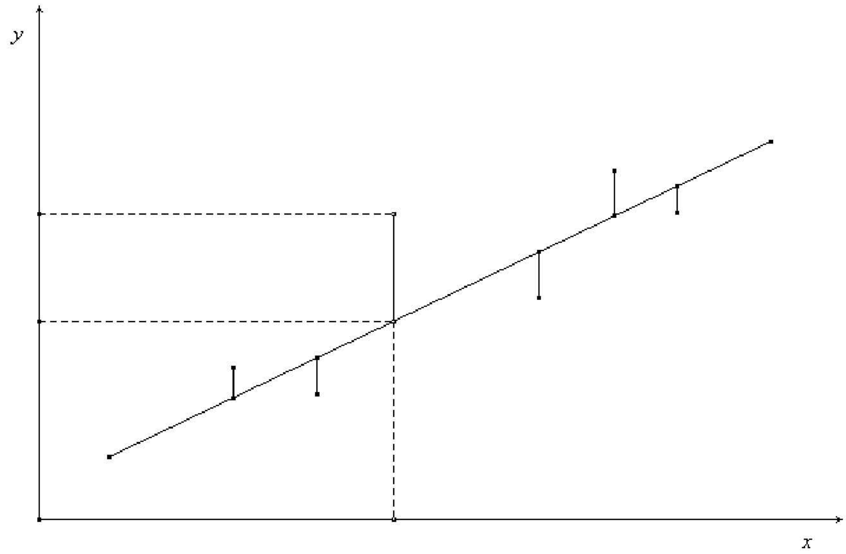
et une estimation la force de frottement liée à un objet dont le poids est 55N

est donc $0,0497 \times 55 + 2,373 = 5,106$, *soit 5,1 N*

Compléments

Cas d'une série double $(x_i, y_i)_{i=1,\dots,n}$ **:**

Ajustement linéaire par la droite des moindres carrés *: (y en fonction de x)*



On associe donc à la droite d, d'équation $y = ax + b$, *la somme* $S = \sum_{i=1}^{n} (y_i - ax_i - b)^2$ **,**

nommée somme quadratique des résidus, et on détermine les valeurs de a et b qui rendent minimum S.

On regarde S comme un polynôme en b.

$$\sum_{i=1}^{n} (y_i - ax_i - b)^2 = nb^2 - 2b\sum_{i=1}^{p} (y_i - ax_i) + \sum_{i=1}^{p} (y_i - ax_i)^2$$

or $\alpha x^2 + \beta x + \gamma$ *, si* $\alpha > 0$ *est minimum pour* $x = -\dfrac{\beta}{2\alpha}$

donc S est minimum pour $b = \dfrac{1}{n}\sum_{i=1}^{n} (y_i - ax_i) = \dfrac{1}{n}\sum_{i=1}^{n} y_i - a\dfrac{1}{n}\sum_{i=1}^{n} x_i$

$$b = \overline{y} - a\overline{x}$$

ainsi , pour tout a : $y = ax + b = ax + \overline{y} - a\overline{x}$ *soit* $y - \overline{y} = a(x - \overline{x})$

la droite passe par le point moyen $G(\overline{x}, \overline{y})$.

On regarde maintenant S comme un polynôme en a.

$$S = \sum_{i=1}^{n} (y_i - \overline{y} + a\overline{x} - ax_i)^2 = \sum_{i=1}^{n} (y_i - \overline{y} - a(x_i - \overline{x}))^2$$

$$S = a^2 \sum_{i=1}^{n} (x_i - \overline{x})^2 - 2a \sum_{i=1}^{n} (x_i - \overline{x})(y_i - b) + \sum_{i=1}^{n} (y_i - b)^2$$

S est minimum pour $\quad a = \dfrac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} =$

Donc $a = \dfrac{Cov(x, y)}{V(x)}$ et $b = \overline{y} - a\overline{x}$

Dans le cas d'une série double donnée sous forme regroupée,un raisonnement analogue conduit au même résultat.

2. Somme quadratique des résidus et résidu quadratique moyen :

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

$$S = \sum_{i=1}^{n} (y_i - (ax_i + \overline{y} - a\overline{x}))^2 = \sum_{i=1}^{n} ((y_i - \overline{y}) - a(x_i - \overline{x}))^2$$

$$S = \sum_{i=1}^{n} (y_i - \overline{y})^2 - 2a \sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x}) + a^2 \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$S = nV(y) \quad - \quad 2naCov(x, y) \quad + \quad na^2 V(x)$$

Comme $a = \dfrac{Cov(x, y)}{V(x)}$*, on a :*

$$S = nV(y) - 2n \frac{[Cov(x, y)]^2}{V(x)} + n \frac{[Cov(x, y)]^2}{V(x)} = nV(y)\left(1 - \frac{[Cov(x, y)]^2}{V(x)V(y)}\right) \quad donc$$

la somme quadratique des résidus par rapport à la droite de régression de y en x est : $S = n(1 - r^2)V(y)$

le résidu quadratique moyen par rapport à la droite de régression de y en x est : $S_m = (1 - r^2)V(y)$ *nommé aussi la variance résiduelle* V_R.

la variance expliquée étant $V_E = r^2 V(y)$

3. Coefficient de détermination : R

C'est $R = r^2$

Et on a $\quad R = r^2 = \dfrac{\text{variance expliquée par la droite de régression}}{\text{variance marginale}}$.

Ecart-type

Définition

L'**écart-type empirique** s d'une série est la racine carrée de la variance :

$$s = \sqrt{V} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

L'unité de l 'écart-type est donc celle des données.

Remarques

- On a aussi pour une série regroupée : $s = \sqrt{\frac{1}{N}\sum_{i=1}^{p} n_i(x_i - \overline{x})^2}$ *où* $N = \sum_{i=1}^{p} n_i$

  Ou encore $s = \sqrt{\sum_{i=1}^{p} f_i(x_i - \overline{x})^2}$ *où* $f_i = \frac{n_i}{N}$

- Comme la moyenne, la variance et l'écart type sont sensibles aux valeurs extrêmes.

Point de référence dans le programme

Premières, BTS.

Champs d'application

C'est un résumé numérique de dispersion.

Le couple (moyenne ; écart-type) permet de faire des calculs sur des regroupements et d'avoir une idée assez précise de la répartition de l'effectif pour une série symétrique.
En pratique ce couple est fondamental dans les nombreux domaines où les grandeurs sont modélisées par des variables gaussiennes (biologie, agronomie, industrie,…)
En théorie ce couple intervient dans de nombreux théorèmes en Statistique Inférentielle.(estimation, tests de validité d'hypothèse, analyse de la variance, théorème de la limite centrée,…)

Exemple 1 : coefficient de variation

Pendant un an on regarde les productions journaliéres de deux chaines de production d'un même véhicule :

Pour la chaine A, on observe $\overline{x}_A = 345$ unités et $s_A = 35,3$
Pour la chaine B, on observe $\overline{x}_B = 445$ unités et $s_B = 43,5$

Quel chaine présente la plus grande dispersion dans sa production journaliére ?

On doit relativiser la valeur de l'écart-type par rapport à la moyenne :

Pour cela on considère le quotient $\dfrac{\text{écart-type}}{\text{moyenne}}$, c'est le **coefficient de variation**

Pour la chaine A, on a : $\dfrac{s_A}{\overline{x}_A} = \dfrac{35,3}{345} \approx 0,1023...$

Pour la chaine B, on a : $\dfrac{s_B}{\overline{x}_B} = \dfrac{43,5}{445} \approx 0,09775...$ c'est donc la chaine A qui répond à la question.

Exemple 2 : distribution des données et écart-type

q un nombre entier naturel

Que peut-on dire de la proportion des données appartenant à $\left[\overline{x} - qs ; \overline{x} + qs\right]$?

On regarde les données extérieures , donc telles que

$|x_i - \overline{x}| > qs \Rightarrow (x_i - \overline{x})^2 > q^2 s^2$

et supposons qu'il y ait p données vérifiant cette inégalité, alors les « *n-p* » autres données vérifient $|x_i - \overline{x}| \leq qs$ (donc $x_i \in \left[\overline{x} - qs ; \overline{x} + qs\right]$) or :

$$ns^2 = \sum_{p \text{ données}} (x_i - \overline{x})^2 + \sum_{n-p \text{ données}} (x_i - \overline{x})^2$$

$$ns^2 > p\, q^2 s^2 + \sum_{n-p \text{ données}} (x_i - \overline{x})^2$$

$$ns^2 - p\, q^2 s^2 > \sum_{n-p \text{ données}} (x_i - \overline{x})^2$$

$$s^2(n - pq^2) > \sum_{n-p \text{ données}} (x_i - \overline{x})^2 > 0$$

donc $n - pq^2 > 0$ et $p < \dfrac{n}{q^2}$

- si $q = 1$: $p <$ n et $p \leq n-1$

  il se peut qu'il y ait une seule donnée dans l'intervalle $\left[\overline{x} - s ; \overline{x} + s\right]$!

- Si $q = 2$: $p < \dfrac{n}{4}$ et $n - p \geq \dfrac{3n}{4}$

on peut affirmer qu'il y a au moins 75% des données dans l 'intervalle $\left[\overline{x} - 2s ; \overline{x} + 2s\right]$

- si $q = 3$: $p < \dfrac{n}{9}$ et $n - p \geq \dfrac{8n}{9}$

on peut affirmer qu'il y a au moins 89% des données dans l 'intervalle $\left[\overline{x} - 3s ; \overline{x} + 3s\right]$...

(Ce sont des pourcentages minimum)

Effectif, effectif cumulé

Définition

∞ L'effectif d'une modalité (classe, valeur,…) est le nombre d'individus correspondant à cette modalité.

Etant donnée une série statistique à caractère quantitatif dans laquelle les valeurs prises par ce caractère ont été classées par ordre croissant,

∞ l'effectif cumulé croissant (resp. décroissant) d'une modalité est la somme de l'effectif de cette modalité et des effectifs de toutes les modalites qui la précèdent (resp. qui la suivent).

Points de référence du programme
Collège, seconde, terminale, BTS.

Champs d'application
Les effectifs cumulés permettent de déterminer une estimation de la médiane et plus généralement les quantiles d'une série.

Remarque : dans le cas où la série a été regroupée on n'aura qu'une estimation de ces paramètres.

Exemple
Soit la série statistique ci-dessous dans laquelle les valeurs prises par le caractère ont été classées par ordre croissant :

valeurs prises par le caractère (ou classes)	$x_1=11$	$x_2=12$	$x_3=13$	$x_4=14$	$x_5=15$	$x_6=16$	Total
effectif	$n_1=2$	$n_2=4$	$n_3=7$	$n_4=10$	$n_5=6$	$n_6=1$	$n=30$
Effectif cumulé croissant	$n_1=2$	$n_1+n_2=6$	$\sum_{j=1}^{3} n_j = 13$	$\sum_{j=1}^{4} n_j = 23$	$\sum_{j=1}^{5} n_j = 29$	$\sum_{j=1}^{6} n_j = 30$	
Effectif cumulé décroissant	$\sum_{j=1}^{6} n_j = 30$	$\sum_{j=2}^{6} n_j = 28$	$\sum_{j=3}^{6} n_j = 24$	$\sum_{j=4}^{6} n_j = 17$	$n_5+n_6=7$	$n_6=1$	

Ainsi il y a :

- *∞10 observations correspondant à la valeur 14 du caractère,*
- *∞23 observations pour lesquelles le caractère prend une valeur au plus égale à14,*
- *∞17 observations pour lesquelles le caractère prend une valeur au moins égale à 14.*
- *∞De ce fait la médiane est égale à 14.*

Echantillon

Définition n°1
Un échantillon issu d'une population est une partie finie d'individus de cette population. Sa taille est le nombre d'individus qui le composent.

Définition n°2
Un échantillon de taille n ($n \in \mathbb{N}^*$) issu d'une population est un n-uplet $(e_1,...,e_n)$ d'individus de la population.

Définition n°3
Un échantillon de taille n ($n \in \mathbb{N}^*$) du caractère mesurable X est un n-uplet $(x_1,...,x_n)$ des valeurs de X sur les individus e_i d'un échantillon $(e_1,...,e_n)$ de la population.

Définition n°4
Un échantillon aléatoire de taille n ($n \in \mathbb{N}^*$) du caractère mesurable X est un n-uplet $(X_1,...,X_n)$ où, pour chaque entier i dans {1 , … , n}, X_i est la variable aléatoire prenant pour valeur la mesure x_i de X sur l'individu n°i de l'échantillon de population choisi.

Points de référence du programme
Définition 1 : classe de seconde.
Définition 2 : classe de terminale.
Définition 3 : classe de terminale, B.T.S.
Définition 4 : B.T.S.

Champ d'application
Définition 3 : un échantillon sert de support à la Statistique Descriptive afin d'appréhender les propriétés de la population.

Définition 4 : un échantillon est utilisé en Statistiques Inférentielles (choix d'un estimateur ponctuel, construction d'un intervalle estimateur, tests d'hypothèses).

Exemples de construction d'un échantillon
Un échantillon peut être choisi :
- De façon déterministe (exemple : on ne s'intéresse qu'aux personnes ayant contracté une certaine maladie, on extrait une pièce fabriquée toutes les 10 sur un tapis roulant, etc.…)
- De façon aléatoire (tirage au hasard avec ou sans remise)
- De façon intermédiaire en tenant compte de la structure de la population (classes d'âge, classes socioprofessionnelles, répartition géographique, etc.…) puis en effectuant un tirage au hasard dans chaque strate ainsi définies (méthode des quotas etc.).
- Un échantillon obtenu de façon aléatoire et à l'aide d'un tirage avec remise (non exhaustif), ou assimilé comme tel, est un échantillon aléatoire indépendant.

Etendue

Définition :
L'étendue d'une série statistique à caractère quantitatif est la différence entre la plus grande et la plus petite des valeurs de la série.

Points de référence du programme:
Seconde, BTS.

Champs d'application:
Son intérêt réside dans la rapidité de sa détermination, dans la mise en évidence de valeurs aberrantes.
Elle est souvent utilisée dans des séries à effectif réduit.
C'est un caractère de dispersion qui ne fait pas intervenir d'autre paramètre.

Exemple
Deux élèves de la même classe ont obtenu les notes suivantes en mathématique :

Elève Antoine : 8- 9- 12- 13- 14- 14- 14.

Elève Bernard : 5- 8- 10- 13- 14- 14- 20.

Ces deux séries ont la même moyenne (12), la même médiane (13) et le même mode (14) pourtant les deux élèves n'ont pas le même profil :
Pour Antoine, l'étendue vaut $14-8=6$, pour Bernard, elle est de $20-5=15$.
Antoine a des notes moins dispersées que celles de Bernard.

Fréquence, fréquence cumulée

Définition

- La fréquence d'une modalité (classe, valeur,…) est la proportion d'individus correspondant à cette modalité.

Etant donnée une série statistique à caractère quantitatif dans laquelle les valeurs prises par ce caractère ont été classées par ordre croissant,

- La fréquence cumulée croissante (resp. décroissante) d'une modalité est la somme de la fréquence de cette modalité et des fréquences de toutes les modalités qui la précèdent (resp. qui la suivent).

Points de référence du programme

Collège, seconde, terminale, BTS.

Champs d'application

Les fréquences cumulées permettent de déterminer une estimation de la médiane et plus généralement les quantiles d'une série.

Remarque : dans le cas où la série a été regroupée on n'aura qu'une estimation de ces paramètres.

Exemple

Soit la série statistique ci-dessous dans laquelle les valeurs prises par le caractère ont été classées par ordre croissant :

valeurs prises par le caractère (ou classes)	$x_1=11$	$x_2=12$	$x_3=13$	$x_4=14$	$x_5=15$	$x_6=16$	Total
effectif	$n_1=2$	$n_2=4$	$n_3=7$	$n_4=10$	$n_5=6$	$n_6=1$	$n=30$
fréquence	$f_1=0.07$	$f_2=0.14$	$f_3=0.23$	$f_4=0.33$	$f_5=0.2$	$f_6=0.03$	1
fréquence cumulée croissante	$f_1=0.07$	$\sum_{j=1}^{2} f_j = 0.21$	$\sum_{j=1}^{3} f_j = 0.44$	$\sum_{j=1}^{4} f_j = 0.77$	$\sum_{j=1}^{5} f_j = 0.97$	$\sum_{j=1}^{6} f_j = 1$	
fréquence cumulée décroissante	$\sum_{j=1}^{6} f_j = 1$	$\sum_{j=2}^{6} f_j = 0.93$	$\sum_{j=3}^{6} f_j = 0.79$	$\sum_{j=4}^{6} f_j = 0.56$	$\sum_{j=5}^{6} f_j = 0.23$	$f_6 = 0.03$	

Ainsi il y a :
- 33% des observations qui correspondent à la valeur 14 du caractère,
- pour 77% des observations le caractère prend une valeur au plus égale à14,
- pour 56% des observations le caractère prend une valeur au moins égale à 14.
- De ce fait la médiane est égale à 14.

Médiane

Définition
Pour une série statistique à caractère quantitatif, la médiane est le plus petit élément q de la série ordonnée tel que au moins 50% des données soient inférieures ou égales à q.

Détermination

- Si on possède n observations ($n \in \mathbb{N}^*$) :
Dans le cas n impair, la médiane est la valeur centrale de la série ordonnée.
Dans le cas n pair, on choisit (par convention) en classe de seconde et de première, la demi somme des deux valeurs centrales.
- Si les données sont regroupées, on parlera de classe médiane.

Remarque
Si la médiane est sensible aux fluctuations d'échantillonnage, elle l'est moins aux valeurs aberrantes. Toutefois elle se prête difficilement à des calculs algébriques.

Points de référence du programme
Classes de seconde, première, terminale et B.T.S.

Champ d'application
La médiane donne une information plus significative que la moyenne dans les séries où les valeurs extrêmes sont importantes.
Par exemple : salaires en économie, médiane des notes d'une classe etc. .

Exemples

Exemple °1 : : notes d'une classe.

Les notes d'une classe A sont : 2 – 4 – 5 – 5 – 8 – 10 – 10 – 11 - 13
Les notes d'une classe B sont : 2 – 4 – 5 – 5 – 8 – 10 – 10 – 11
Pour la classe A, la médiane est égale à 8 (valeur centrale)
Pour la classe B, la médiane est égale par définition à 5, et, si on applique la convention ci-dessus, elle est égale à $\frac{5+8}{2} = 6.5$

Exemple °2 :

La distribution de 57 mesures de longueur, en mètres, a fourni les résultats suivants :

Longueur	[0 ;2[	[2 ;4[	[4 ;6[	[6 ;8[	[8 ;10[
fréquence	0.14	0.26	0.32	0.23	0.05

On ne peut pas calculer la médiane car on ne possède pas les données observées avant regroupement. Cependant on peut en donner une estimation par interpolation linéaire :

Longueur	[0 ;2[	[2 ;4[	[4 ;6[	[6 ;8[	[8 ;10[
fréquence	0.14	0.26	0.32	0.23	0.05
Fréq. Cum.	0.14	0.40	0.72	0.95	1.00

La classe médiane est la classe [4 ;6[.
40% des longueurs observées sont inférieures à 4m et 72% inférieures à 6m.

On estime qu'il y a 50% des longueurs inférieures à $4+\dfrac{(50-40)(6-4)}{(72-40)}=4.625$.

La médiane est estimée à 4.625 m.

Exemple °3 la demi durée de vie d'un atome.

La durée de vie d'un atome est une variable aléatoire X qui suit une loi exponentielle de paramètre λ.

La médiane m_e vérifie : $\displaystyle\int_0^{m_e}\lambda e^{-\lambda t}\,dt=\frac{1}{2} \quad\Leftrightarrow\quad 1-e^{-\lambda m_e}=\frac{1}{2}$

$$\Leftrightarrow \quad m_e=\frac{\ln 2}{\lambda}$$

On remarque que $E(X)=\dfrac{1}{\lambda}$ donc la moyenne et la médiane sont différentes.

Mode

Définition :
Pour une série statistique à caractère quantitatif discret X, le mode est la valeur du caractère correspondant au plus grand effectif.

Pour une série statistique à caractère quantitatif continu X la classe modale est la classe du caractère correspondant au plus grand effectif.

- Si les classes sont de même amplitude, on peut estimer le mode en choisissant le centre de la classe modale, ou en affinant le calcul comme il est fait à l'exemple 1.
- Si les classes ne sont pas de même amplitude, on peut estimer le mode en choisissant la classe correspondant à la plus grande valeur de l'effectif corrigé (voir exemple 2).

Remarque :
Le mode n'est pas nécessairement unique. S'il y a plusieurs modes, on parle de série plurimodale.

Points de référence du programme :
Seconde, B.T.S.

Champ d'application
Le mode est représentatif de la valeur la plus courante.
Il est facile à déterminer et graphiquement il représente d'emblée une valeur centrale.
Il peut cependant présenter une certaine ambiguïté (pas de mode, plusieurs modes)
Il peut s'imposer comme paramètre de position dans certaines situations

Exemples

1. Un marchand de chaussures d'enfants a fait une étude statistique pour connaître la taille de chaussures qu'il vend le plus :

taille	[21 – 25[	[25 -29[	[29 -33[	[33 -37[
effectif	100	120	110	90

Les classes sont d'étendue égales à $\alpha = 4$
La classe modale est la classe [25 – 29[
Il peut estimer que le mode (c'est-à-dire la taille la plus vendue) est le 27.
Il peut aussi, si c'est possible, calculer les différences d'effectifs entre la classe modale $[x_k\,;x_{k+1}[$ *et les deux classes limitrophes* $[x_{k-1}\,;x_k[$ *et* $[x_{k+1}\,;x_{k+2}[$: $d_1 = 120 - 100 = 20$ *et* $d_2 = 120 - 110 = 10$ *Alors le mode peut être estimé grâce à la formule* $m_o = x_k + \dfrac{\alpha d_1}{d_1 + d_2} = 25 + \dfrac{4 \times 20}{20 + 10} \oplus 27.67$

2. Un publiciste veut cibler la classe modale pour promouvoir son produit. Une étude statistique lui a fournit les résultats suivants portant le nombre de ventes (en milliers) en fonction du poids en gammes d'un paquet de lessive:

poids	[250 – 500[	[500 – 575[	[575-1000[	[1000 – 2000[
effectif n_k	10	50	70	80

Les classes ne sont pas d'étendue égales. La plus petite amplitude est $\alpha = 250$. *Pour chaque classe* $[x_k ; x_{k+1}[$ *l'effectif* n_k *est remplacé par l'effectif corrigé* $\tilde{n}_k$ *de telle sorte que* $\tilde{n}_k \times (x_{k+1} - x_k) = n_k$ *(les aires des rectangles de l'histogramme sont proportionnelles aux effectifs observés)*
On obtient les résultats suivants :

poids	[250 – 500[	[500 – 575[	[575-1000[	[1000 – 2000[
effectif $\tilde{n}_k$	10	50	70	20

La classe modale est la classe [575- 1000[.
Le mode peut être estimé par le centre de la classe modale : $m_O = 787.5$ *g*

Moyenne arithmétique

Définition
Pour une série statistique à caractère quantitatif discret X, la moyenne arithmétique est :

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ où $n \in IN^*$ et $x_1, x_2, \ldots, x_n$ sont les n valeurs observées distinctes ou non.
- $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$ où $n \in IN^*$ et $x_1, x_2, \ldots, x_k$ sont les k valeurs distinctes observées avec pour effectif correspondant $n_1, n_2, \ldots, n_k$ et $n = \sum_{i=1}^{k} n_i$

Pour une série statistique à caractère quantitatif continu X, la moyenne arithmétique peut être estimée en remplaçant les valeurs numériques x_i par les centres des classes c_i.

Remarques

- La moyenne arithmétique est peu sensible aux fluctuations d'échantillonnage.

Elle se prête bien aux comparaisons.
Des valeurs aberrantes peuvent toutefois la modifier sensiblement.
Sa détermination est aisée car algébrique.

- $\bar{x}$ est le barycentre de la famille (n_i ; x_i) pour i entier entre 1 et k.
- On définit aussi la moyenne géométrique $m_g = (x_1 x_2 \ldots x_n)^{1/n}$ ou $(x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k})^{1/n}$

  la moyenne quadratique $m_q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$ ou $\frac{1}{n}\sqrt{\sum_{i=1}^{k} n_i x_i^2}$

  la moyenne harmonique $m_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$ ou $\frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$

  (Selon que les valeurs des xi sont regroupées ou non.)

Points de référence du programme
Seconde, première, terminale,B.T.S.

Champ d'application
C'est le paramètre de position le plus utilisé en statistique.
Elle sert d'outil dans des calculs algébriques en statistiques inférentielles.

Exemple
Moyenne de la série : 11.6 - 11.7 - 11.8 - 12 - 12.2 - 12.3 - 12.4 - 12.8

$\bar{x} = \frac{1}{8}(11.6 + 11.7 + 11.8 + 12 + 12.2 + 12.3 + 12.4 + 12.8) = 12.1$

On peut aussi utiliser la linéarité de la moyenne en remplaçant chaque valeur x_i par $10x_i - 120$. On obtient la série : $-4 \quad -3 \quad -2 \quad 0 \quad 1 \quad 3 \quad 4 \quad 8$ de moyenne $10\bar{x} - 120 = 1$ d'où $\bar{x} = 12.1$

Population

Définition
Ensemble dont les éléments sont appelés individus et sur lequel porte l'étude statistique.

Remarque
En statistique, le point de départ est la problématique qu'il faut préciser, à savoir :

- la population sur laquelle on va travailler.
- le ou les caractère(s) étudié(s)

Une étude statistique induit donc une population, des individus, un ou des caractère(s).

Points de référence du programme
Classes de seconde, première, terminale, BTS.

Champ d'application
Les études statistiques portent sur un ou des caractères d'une population, ou d'un échantillon de cette population.

Quantiles

Définitions

On considère une série statistique discrète $\left(x_i\right)_{i=1,\dots,n}$ ordonnée, c'est à dire la liste de toutes les valeurs numériques prises par le caractère de tous les individus de la population ou de l'échantillon étudié rangées par ordre croissant, la même valeur pouvant être répétée.

1. les quartiles
Intuitivement, l'idée est de partager la série en quatre parties « égales»

- le **premier quartile** : $\boldsymbol{q_1}$

On considère l'ensemble des valeurs q des termes de la série ordonnée, pour lesquelles au moins 25 % des données soient inférieures (ou égales) à q.
$\boldsymbol{q_1}$ est le plus petit élément de cet ensemble.

- le **troisième quartile :** $\boldsymbol{q_3}$

On considère l'ensemble des valeurs q des termes de la série ordonnée, pour lesquelles au moins 75 % des données soient inférieures (ou égales) à q.
$\boldsymbol{q_3}$ est le plus petit élément de cet ensemble.

- le **second quartile** est alors la médiane m_e

- **L'intervalle interquartile :** c'**est** l'intervalle dont les extrémités sont le premier et le troisième quartiles : $[\boldsymbol{q_1} ; \boldsymbol{q_3}]$.

Au moins 50% des termes de la série ont une valeur qui appartient à cet intervalle :

• Le pourcentage des termes dont la valeur est strictement supérieur à q_3 est inférieur à 25%, sinon q_3 ne serait pas le plus petit élément q' des valeurs des termes de la série ordonnée, tel qu'au moins 75 % des données soient inférieures à q'.
• Le pourcentage des termes dont la valeur est strictement inférieur à q_1 est inférieur à 25%, sinon q_1 ne serait pas le plus petit élément q des valeurs des termes de la série ordonnée, tel qu'au moins 25 % des données soient inférieures ou égales à q.
• Ainsi le pourcentage des termes dont la valeur appartient à cet intervalle est supérieur à 50%.

- **L'écart interquartile :** est la différence entre le troisième et le premier quartile $\boldsymbol{q_3} - \boldsymbol{q_1}$

  Il donne une « mesure » de la dispersion de la série.

- **le couple (médiane ; intervalle interquartile)** pour résumer une série statistique.

On peut construire une « **boite à moustaches** » ou diagramme de <u>Tuckey</u> pour visualiser une série (et les comparer d'un seul coup d'œil entres elles)

Dans l'exemple ci dessous on a préféré laisser les valeurs en abscisses comme dans la majorité des représentations graphiques.

min q_1 $\mathrm{m_e}$ q_3 max

2. les déciles

On partage cette fois la série en 10 parties égales.

- **Premier décile** : d_1

On considère l'ensemble des valeurs q des termes de la série ordonnée, pour lesquelles au moins 10 % des données sont inférieures (ou égales) à q.
d_1 est le plus petit élément de cet ensemble.

- **Neuvième décile** : d_9

On considère l'ensemble des valeurs q des termes de la série ordonnée, pour lesquelles au moins 90 % des données sont inférieures (ou égales) à q.
d_9 est le plus petit élément de cet ensemble.

- **L'intervalle inter décile** : $[d_9\ ;\ d_1]$ qui « contient au moins » 80% des termes de la suite.

- **Et l'écart inter décile :** $d_9 - d_1$

Enfin, on peut construire une autre « boite à moustaches » d'où sont exclues, en principe les valeurs extrêmes, aberrantes ou erronées.

d_1 q_1 m_e q_3 d_9

- Enfin on peut définir de la même façon les **vingtiles**, les **centiles**,…

Point de référence dans le programme

Premières, Terminales et BTS.

Champs d'application

- diagramme de <u>Tuckey</u>

- ils permettent d'apprécier la symétrie ou la dissymétrie d'une série statistique, indication importante dans le choix d'un modèle.
- adéquation à une loi
- test

Remarques

∞ Il est immédiat qu'une transformation affine conserve les quantiles :

Par exemple, pour les quartiles

Si $(x'_i ; n_i) = (ax_i + b ; n_i)$ $\forall i \in \{1;\ldots;p\}$ où $a \in \mathbb{R}_+$ et $b \in \mathbb{R}$,

Alors : $q'_1 = aq_1 + b$ *et* $q'_3 = aq_3 + b$

et pour l'écart interquartile :

$q'_3 - q'_1 = (aq_3 + b) - (aq_1 + b) = a(q_3 - q_1)$

∞ par construction,

Ils donnent une bonne idée de la répartition des données autour de la valeur centrale.

Les données extrêmes (ou aberrantes) n'influent pas sur la valeur des quantiles. L'inconvénient de ces résumés est qu'ils ne se prêtent pas aux calculs. Par exemple on ne peut calculer les résumés du regroupement de plusieurs séries à partir des résumés de chacune des séries initiales.

Exemple

On mesure la masse, en grammes, des lettres postales d'un échantillon de taille 500 :

x_i (en gr)	15	16	17	18	19	20	21	22	25	30
n_i	12	10	40	78	110	120	70	30	20	10
$\sum n_i$	12	22	62	140	250	370	440	470	490	500

a. $q_1 = 18$, $q_3 = 21$ et $m_e = 19$ (on peut réaliser la « boite à moustache »)

$\frac{440-62}{500} = \frac{378}{500} = 0,756$ donc 75,6% des individus sont dans l'intervalle interquartile.

b. $d_1 = 17$ et $d_9 = 22$

$\frac{470-22}{500} = \frac{448}{500} = 0,896$ donc 89,6% des individus sont dans l'intervalle interdécile.

c. le $5^{\text{ième}}$ centile est $c_5 = 17$, le $95^{\text{ième}}$ centile est $c_{95} = 25$.

Variance

Définition

La **variance V** d'une série statistique $(x_i)_{i=1,\ldots,n}$ est la moyenne quadratique des écarts des données à la moyenne.

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

C'est aussi la différence entre « la moyenne des carrés » et « le carré de la moyenne ».

$$V = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - \overline{x}^2 \quad \text{formule dite de } \textit{Huyghens-König}$$

C'est un résumé numérique de dispersion.

Remarques

- Expression nommée **variance** par *Fisher* en 1918.
- On a aussi, pour les séries statistiques regroupées $(n_i, x_i)_{i=1,\ldots,n}$

$$V = \frac{1}{N}\sum_{i=1}^{p}n_i(x_i - \overline{x})^2 \quad où \ N = \sum_{i=1}^{p}n_i$$

Ou encore $$V = \sum_{i=1}^{p}f_i(x_i - \overline{x})^2 \quad où \quad f_i = \frac{n_i}{N}$$

- $\sum_{i=1}^{n}(x_i - \overline{x})^2$ mesure le carrée de la distance usuelle de la série à la série non dispersée (dont l'étendue est nulle) la plus proche :

La fonction $\alpha \longrightarrow \sum_{i=1}^{n}(x_i - \alpha)^2$ est minimale pour $\alpha = \overline{x}$

- C'est la forme quadratique associée à la forme bilinéaire symétrique (dégénérée) sur R^n définie par $f(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ c'est à dire la **covariance** $\mathrm{cov}(x,y)$

En particulier $$\mathrm{cov}(x,y) = \frac{1}{2}\big(V(x+y) - V(x) - V(y)\big)$$

- Effet d'une transformation affine des données :

Si on applique une fonction affine à toutes les données :

$$\begin{cases} x_i \text{ a } x'_i = ax_i + b & où \ a \in \mathrm{R} \ et \ b \in \mathrm{R} \\ \forall i \text{ Î } \{1;\ldots;n\} \end{cases}$$

Comme $\overline{x'} = \overline{ax+b} = a\overline{x}+b$, $V' = \frac{1}{n}\sum_{i=1}^{p}(x_i' - \overline{x}')^2 = \frac{1}{n}\sum_{i=1}^{p}(ax_i - a\overline{x})^2 = a^2 V$

∞ En particulier la variance est invariante par translation :

$$V = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\ 2}\right) - \overline{x}^2 = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2\right) - (\overline{x} - a)^2 \quad \forall a \in \mathrm{R}$$

∞ la formule dite de Huyghens-König fournit une inégalité intéressante en elle même :

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \ldots 0$$

donc $\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 „ \frac{1}{n}\sum_{i=1}^{n} x_i^2$ ou bien $\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right)^2 „ \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}$

le carré de la moyenne est inférieure à la moyenne des carrés

Point de référence dans le programme

Premières, BTS.

Champs d'application

C'est un résumé numérique de dispersion.

Exemple

On mesure la masse, en grammes, des lettres postales d'un échantillon de taille 500 :

x_i (en grammes)	15	16	17	18	19	20	21	22	25	30
n_i	12	10	40	78	110	120	70	30	20	10

Calculer la variance

$$On\ a \quad \sum_{k=1}^{10} n_i x_k = 9844 \quad et \quad \sum_{k=1}^{10} n_i x_k^{\ 2} = 196692$$

$\overline{x} = \frac{9844}{500} = 19,7$ d'où $V = \frac{196692}{500} - \left(\frac{9844}{500}\right)^2 = 5,294$

Compléments :

1 Regroupements de 2 séries de tailles n_1 et n_2 ; $(x_i)_{i=1,\ldots,n_1}$ et $(x_i)_{i=n_1+1,\ldots,n_2}$

La variance empirique de la série obtenue par regroupement est :

$$V = \frac{1}{n_1 + n_2} \sum_{i=1}^{n_1+n_2} (x_i - \overline{x})^2 = \frac{1}{n_1 + n_2} \sum_{i=1}^{n_1+n_2} x_i^{\ 2} - \overline{x}^2$$

$$V = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1+n_2} x_i^{\ 2} - (n_1 + n_2)\overline{x}^2 \right)$$

$$V = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} x_i^{\ 2} - n_1\overline{x}_1^2 + \sum_{i=n_1+1}^{n_1+n_2} x_i^{\ 2} - n_2\overline{x}_2^2 + n_1\overline{x}_1^2 + n_2\overline{x}_2^2 - (n_1 + n_2)\overline{x}^2 \right)$$

$$V = \frac{1}{n_1 + n_2} \left(n_1V_1 + n_2V_2 + n_1\overline{x}_1^2 + n_2\overline{x}_2^2 - (n_1 + n_2)\overline{x}^2 \right)$$

$$V = \frac{n_1V_1 + n_2V_2}{n_1 + n_2} + \frac{n_1(\overline{x}_1^2 - \overline{x}^2) + n_2(\overline{x}_2^2 - \overline{x}^2)}{n_1 + n_2}$$

On peut faire apparaître outre la moyenne des variances , la variance des moyennes :

$$\frac{n_1(\overline{x}_1^2 - \overline{x}^2) + n_2(\overline{x}_2^2 - \overline{x}^2)}{n_1 + n_2} = \frac{n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 + 2n_1\overline{x}_1\overline{x} + 2n_2\overline{x}_2\overline{x} - 2(n_1 + n_2)\overline{x}^2}{n_1 + n_2}$$

$$= \frac{n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 + 2(n_1\overline{x}_1 + n_2\overline{x}_2)\overline{x} - 2(n_1 + n_2)\overline{x}^2}{n_1 + n_2}$$

$$= \frac{n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2}{n_1 + n_2} \quad \text{puisque } (n_1\overline{x}_1 + n_2\overline{x}_2) = (n_1 + n_2)\overline{x}$$

D'où la formule standard $V = \dfrac{n_1V_1 + n_2V_2}{n_1 + n_2} + \dfrac{n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2}{n_1 + n_2}$ (formule facile à généraliser)

Le premier terme est donc la moyenne des variances :

$\overline{V(x_j)}$ nommée **variance intra population**

Le second terme est donc la variance des moyennes :

$V(\overline{x}_j)$ nommée **variance inter population**

∞ la variance de la série regroupée n'est donc pas la moyenne pondérée des variances

2 fluctuations d'échantillonnage

Si on regroupe ***p*** séries de taille ***n***, on sait que la moyenne des ***p*** moyennes est la moyenne des ***np*** valeurs

Il n'en va pas de même pour la variance des moyennes :

La variance V des np valeurs est

$$V = \frac{1}{np}\sum_{j=1}^{p}\sum_{i=1}^{n}(x_{i;j}-\overline{x})^2 = \frac{1}{p}\sum_{j=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i;j}-\overline{x})^2\right) \qquad \left(\text{donc plus grande que: } \frac{1}{p}\sum_{j=1}^{p}\frac{1}{n}\sum_{i=1}^{n}(x_{i;j}-\overline{x}_j)^2\right)$$

(On voit ici, de plus que la variance des *np* valeurs est supérieure à la moyenne des variances des p échantillons)

La variance des moyennes V_m est

$$V_m = \frac{1}{p}\sum_{j=1}^{p}(\overline{x}_j-\overline{x})^2 = \frac{1}{p}\sum_{j=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n}x_{i;j}-\overline{x}\right)^2$$

$$Or \quad \forall j \in [\![1;p]\!], \qquad \left(\frac{1}{n}\sum_{i=1}^{n}x_{i;j}-\overline{x}\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n}(x_{i;j}-\overline{x})^2$$

(Le carré de la moyenne est inférieure à la moyenne des carrés)

Donc $$V_m = \frac{1}{p}\sum_{j=1}^{p}(\overline{x}_j-\overline{x})^2 \leq \frac{1}{p}\sum_{j=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i;j}-\overline{x})^2\right) \qquad \text{et} \qquad V_m < V$$

C'est à dire que la variance de la série des *p* moyennes est plus petite que la variance des *np* données.

PROBABILITE

Dénombrement

Espérance

Fonction de répartition

Indépendance

Loi binomiale

Loi de probabilité

Loi de Poisson

Loi exponentielle

Loi normale

Loi uniforme

Probabilité conditionnelle

Variable aléatoire continue

Variable aléatoire discrète

Variance, écart type

Dénombrement

Définition
Dénombrer c'est déterminer le nombre d'éléments d'un ensemble fini E.
Si le cardinal de E est petit, on peut utiliser un arbre sinon on aura recours à des formules.

Les formules les plus usitées sont :

- n^p qui permet de dénombrer le nombre d'applications d'un ensemble à « p » éléments vers un ensemble à « n » éléments ; il s'agit aussi d'arrangements avec répétition de « p » éléments parmi « n ».

- A_n^p qui permet de dénombrer le nombre d'applications injectives d'un ensemble à « p » éléments vers un ensemble à « n » éléments ; il s'agit aussi d'arrangements sans répétition de « p » éléments parmi « n ». $A_n^p = n(n-1)...(n-p+1) = \dfrac{n!}{(n-p)!}$

- $n!$ qui permet de dénombrer le nombre d'applications bijectives d'un ensemble à « n » éléments dans un ensemble à « n » éléments ; il s'agit aussi des permutations de « n » éléments.

- C_n^p *ou* $\begin{pmatrix} n \\ p \end{pmatrix}$ qui permet de dénombrer le nombre de parties ou de combinaisons à « p » éléments dans un ensemble à « n » éléments. $C_n^p = \dfrac{A_n^p}{p!} = \dfrac{n!}{p!(n-p)!}$

Complément

- Γ_n^p qui permet de dénombrer le nombre de parties ou de combinaisons avec répétition à « p » éléments dans un ensemble à « n » éléments $\Gamma_n^p = C_{n+p-1}^p$

- Pour déterminer le nombre de permutations de « n » éléments dont « p » groupes sont formés d'éléments identiques, on divisera « $n\,!$ » par la factorielle du cardinal de chaque groupe d'éléments identiques ; il s'agit de permutations avec répétition.

Points de référence dans le programme
Classes de terminale et de BTS.

Champs d'application

En calcul des probabilités dans le cas où l'univers $|$ est formé d'un nombre fini d'évènements élémentaires équiprobables, la probabilité d'un évènement A est : $p(A) = \dfrac{Card\ A}{Card\ |}$, il est donc nécessaire de calculer $Card\ A$ et $Card\ |$

Exemples

Exemple 1

En informatique, on code les différents symboles par des « octets » (suite de 8 chiffres prenant les valeurs 0 ou 1). Combien y a-t-il d'octets possibles ?

Chaque octet correspond à une application de l'ensemble {1,2,3,4,5,6,7,8} dans l'ensemble {0,1}. Il y a donc $2^8 = 256$ *possibilités.*

Exemple 2

Avec les chiffres {1, 2, 3, 4, 5, 6, 7, 8} pris une seule fois, combien de nombres de trois chiffres peut on faire ?

Chaque nombre de trois chiffres est un arrangement de trois chiffres parmi les huit.

Il y a donc $A_8^3 = 8 \times 7 \times 6 = 336$ *possibilités.*

Exemple 3

Combien peut on faire d'anagrammes du mot : « nombre »

Chaque anagramme est une permutation des 8 lettres il y en a donc 6 ! = 720.

Exemple 4

Combien y a-t-il de tirages distincts de 4 cartes d'un jeu de 32 ?

Chaque tirage correspond à une partie de 4 cartes parmi 32, il y a donc $C_{32}^4 = \dfrac{32 \times 31 \times 30 \times 29}{4!} = 35960$

Exemple 5

Combien y a-t-il de pièces différentes dans un jeu de dominos ?

Pour réaliser un domino, il faut choisir deux nombres de 0 à 6 avec possibilité de répétition car il peut y avoir des doubles.

Il y a donc autant de pièces que de combinaisons avec répétition de 2 nombres parmi 7 soit :

$$\Gamma_7^2 = C_8^2 = 28$$

Exemple 6

Combien peut on faire d'anagrammes du mot : « probabilité »

Chaque anagramme est une permutation avec répétition des deux « i » et des deux « b » des 11 lettres il y en a donc $\dfrac{11!}{2! \times 2!}$ *= 9979200 permutations.*

Espérance mathématique

Définition

∞ Si X est une variable aléatoire discrète finie prenant les valeurs x_1, x_2,……x_n $n \in$ ¥ , l'espérance mathématique de x, notée E(X) est définie par :

$$E(X) = \sum_{i=1}^{i=n} x_i p(X=x_i)$$

∞ Si l'ensemble des valeurs x_i n'est pas fini, et si E(X) existe alors :

$$E(X) = \sum_{i=1}^{+\infty} x_i \, p(X = x_i)$$

∞ Si X est une variable aléatoire continue de densité f définie sur ¡ (voir fonction de répartition) et si E(X) existe alors :

: $E(X) = \int_{¡} x\, f(x)dx$.

Propriétés

a et b sont des réels, X et Y des variables aléatoires

∞ Propriété de linéarité : E(aX) = aE(X) et : E(X + Y) = E(X) + E(Y)

∞ E(aX +b) = aE(X) + b

Point de référence dans le programme

Terminales, BTS.

Champs d'application

L'espérance mathématique permet de calculer le gain moyen espéré dans un jeu, la durée de vie moyenne d'un produit, etc……

Exemple 1

On lance trois dés simultanément. On gagne dix euros si on obtient un triple six, cinq euros pour un double six, un euro pour un seul six, et on perd 2 euros si on n'obtient aucun 6 .

$\Omega = \{ (i, j, k) / i \in \{ 1 ;2 ;3 ;4 ;5 ;6\}, j \in \{ 1 ;2 ;3 ;4 ;5 ;6\} \text{ et } k \in \{ 1 ;2 ;3 ;4 ;5 ;6\} \}$

Card Ω = 216

X est la variable aléatoire « gain obtenu ». X suit la loi de probabilité donnée par le tableau :

x_i	-2	1	5	10

$P(X=x_i)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

$E(X) = -2 \times \frac{125}{216} + 1 \times \frac{75}{216} + 5 \times \frac{15}{216} + 10 \times \frac{1}{216} = -\ 0,4$ arrondie au dixième.

L'espérance mathématique est donc de -0,4€.

Exemple 2 :

On lance trois dés simultanément. On gagne cent euros si on obtient un triple six, cinq euros pour un double six, un euro pour un seul six, et on perd treize euros si on n'obtient aucun 6 .

X est le gain obtenu.

La loi de probabilité devient :

x_i	- 13	1	5	100
$P(X=x_i)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

$$E(X) = \frac{-13 \times 125 + 1 \times 15 + 5 \times 15 + 1 \times 100}{216} = -\ 6,4$$

Donc l'espérance mathématique est de -6,4€.

L'espérance mathématique est inférieure à celle de l'exemple 1, mais le jeu sera plus attractif.

Exemple 3

Calcul de l'espérance mathématique de la loi exponentielle.

La fonction de densité est $\begin{cases} f(x) = \lambda\, e^{-\lambda x} & \text{si } x \geq 0 \\ f(x) = 0 & \text{sinon} \end{cases}$ $\lambda \in]0, +\infty[$

Donc $E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx$

On pose $u(x) = x$ $v'(x) = \lambda\ e^{-\lambda x}$ $u'(x) = 1$ $v(x) = -\ e^{-\lambda x}$

La formule d'intégration par parties donne :

$$\begin{aligned} \int_0^a \lambda x e^{-\lambda x} dx &= \left[-xe^{-\lambda x}\right]_0^a + \int_0^a e^{-\lambda x} dx \\ &= -a\, e^{-\lambda a} + \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^a \\ &= -\, a\, e^{-\lambda a} + \frac{1}{\lambda}(-\, e^{-\lambda a} + 1) \end{aligned}$$

$E(X)$ est la limite de cette intégrale lorsque a tend vers $+\infty$

Donc $E(X) = \frac{1}{\lambda}$

En fiabilité,ce nombre est appelé MTBF (Moyenne des Temps de Bon Fonctionnement).

Fonction de répartition

Définition
On appelle fonction de répartition d'une variable aléatoire X la fonction de la variable réelle x définie par : $F(x) = P(X \leq x) \quad où \; x \in \mathbb{R}$

Remarques :

- La fonction de répartition « F» peut parfois être définie par :
$F(x) = P(X < x) \quad où \; x \in \mathbb{R}$

- La fonction de répartition est croissante, à valeurs dans $[0,1]$.

- Dans le cas d'une variable aléatoire continue X, une fonction f , est densité de probabilité de X si elle est positive, intégrable sur $\mathbb{R}$ et si son intégrale sur $\mathbb{R}$ est égale à 1.
  La fonction de répartition est alors définie par : $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$. Pour tout x réel. On a alors : F' = f .

Points de référence dans le programme
Classes de BTS

Champs d'application
Dans le cas d'une variable aléatoire discrète elle peut intervenir, si on s'intéresse aux variables aléatoire Max(X) ou min (X).

Exemple 1

Une urne contient 6 boules vertes et 4 boules rouges.
On tire 3 boules de l'urne, soit X le nombre de boules vertes.
Déterminer la fonction de répartition de X.

X prend les valeurs : {0,1,2,3}. Card(Ω) = $C_{10}^{3} = 120$.

$$P(X=0) = \frac{C_4^3}{120} = \frac{1}{30}$$
$$P(X=1) = \frac{C_6^1 C_4^2}{120} = \frac{9}{30}$$
$$P(X=2) = \frac{C_6^2 C_4^1}{120} = \frac{15}{30}$$
$$P(X=3) = \frac{C_6^3}{120} = \frac{5}{30}$$

$F(x) = P(X \leq x)$

$pour\ x \in]-\infty, 0[\quad F(x) = 0$

$pour\ x \in [0,1[\quad F(x) = \dfrac{1}{30}$

$pour\ x \in [1,2[\quad F(x) = \dfrac{10}{30}$

$pour\ x \in [2,3[\quad F(x) = \dfrac{25}{30}$

$pour\ x \in [3,+\infty[\quad F(x) = 1$

Exemple 2

Soit X une variable aléatoire continue de densité *f* définie par :

$$\begin{cases} f(x) = 2e^{-2x} & pour\ x \in [0,+\infty[\\ f(x) = 0 & pour\ x \in]-\infty, 0[\end{cases}$$

Déterminer la fonction de répartition de X.

$F(x) = P(X \leq x)$

$Pour\ x \in]-\infty, 0[\quad F(x) = 0$

$Pour\ x \in [0,+\infty[, \ F(x) = \int_0^x 2e^{-2t}\, dt = [-e^{-2t}]_0^x = 1 - e^{-2x}$

Indépendance

Définition 1: événements indépendants

P une probabilité sur $(\Omega, \mathcal{P}(\Omega))$

Deux événements A et B ($A \subset \Omega$ et $B \subset \Omega$) sont **indépendants** si et seulement si pour la probabilité P : $P(A \cap B) = P(A) \times P(B)$.

Cela équivaut à $P(A/B) = P(A)$ si $P(B) \neq 0$ (ou encore $P(B/A) = P(B)$ si $P(A) \neq 0$).

L'emploi de ce qualificatif est justifié par le fait que : la réalisation de A ne dépend pas de celle de B *au sens de la probabilité P*.

Remarque

- L'indépendance n'est pas intrinsèque à A et B, elle dépend de P : A et B peuvent être indépendants pour P et non pour une autre probabilité P'.
- L'indépendance n'est pas « intuitive ».

Cependant :

- Deux événements *indépendants* au « sens du langage commun» sont indépendants au « sens des probabilités ». L'inverse est faux !
- Il faut signaler le risque de confusions entre **indépendants** et **incompatibles.**

Définition 2 : variables aléatoires indépendantes

P une probabilité sur $(\Omega, \mathcal{P}(\Omega))$

Deux variables aléatoires X_1 et X_2 sont **indépendantes** pour *P si :*

$\forall I_1 \subset \mathbb{R}$ et $\forall I_2 \subset \mathbb{R}$, les événements $A_1 = X_1^{-1}(I_1)$ et $A_2 = X_2^{-1}(I_2)$ sont **indépendants** pour *P*, soit. : $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$.

Point de référence dans le programme

Terminales, BTS.

Champs d'application

Tirage avec remise.
Schéma de Bernouilli.
Echantillonnage.

Exemple 1

L'expérience consiste à jeter 1 dé. S'il y a équiprobabilité, et si A est l'évènement *« impair »* et B est l'évènement *« supérieur à 5 », alors :*

$$P(A) = \frac{3}{6} = \frac{1}{2}, \quad P(B) = \frac{2}{6} = \frac{1}{3} \quad \text{et} \quad P(A \cap B) = \frac{1}{6} = P(A) \times P(B)$$

donc A et B sont indépendants pour P , ce qui n'était pas intuitif…

Mais il se peut que la probabilité (dans le cas d'un dé truqué) soit définie par :

	1	2	3	4	5	6
$p_i\,'$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{15}$	$\frac{1}{6}$

Alors $P'(A)=\dfrac{7}{15}$, $P'(B)=\dfrac{3}{10}$ mais $P'(A\cap B)=\dfrac{2}{15}$:
A et B ne sont plus indépendants pour P'.

Exemple 2

L'expérience consiste à jeter 2 dés cubiques discernables (un rouge et un vert par exemple)
On a 2 v.a. : X_1 la valeur de la face du premier dé, X_2 la valeur de la face du second. Puis deux autres : X_1+X_2 et $X_1\times X_2$.

$\Omega=\{(i,j)/i\in\{1;2;3;4;5;6\}\ ,j\in\{1;2;3;4;5;6\}\}$,

Ω est formé de 36 couples ordonnés correspondant à des événements que l'on peut supposer « équiprobables »

X_1 et X_2 sont « clairement » indépendantes et :

$P(X_1=2)=\dfrac{6}{36}=\dfrac{1}{6}$ $\quad P(X_2=3)=\dfrac{6}{36}=\dfrac{1}{6}$ $\quad$ (*selon l'usage* : $P\left(X_1^{-1}(\{x_i\})\right)=P(X_1=x_i)$)

$P\left((X_1=2)\cap(X_2=3)\right)=\dfrac{1}{36}=\dfrac{1}{6}\times\dfrac{1}{6}=P(X_1=2)\times P(X_2=3)$ etc…

$\boxed{X_1 \text{ et } X_2 \text{ sont indépendantes}}$

En revanche, X_1 et $X_1\times X_2$ ne sont pas « naïvement » indépendantes et :

$P(X_1=4)=\dfrac{6}{36}=\dfrac{1}{6}$ $\quad P(X_1\times X_2=4)=\dfrac{1}{36}$

et $\quad P\left((X_1=4)\cap(X_1\times X_2=4)\right)=\dfrac{1}{36}\neq\dfrac{1}{6}\times\dfrac{1}{12}=P(X_1=4)\times P(X_1\times X_2=4)$

$\boxed{X_1 \text{ et } X_1\times X_2 \text{ ne sont pas indépendantes}}$

De plus :

$P(X_1+X_2=4)=\dfrac{3}{36}=\dfrac{1}{12}$ $\quad P(X_1\times X_2=4)=\dfrac{3}{36}$

et $\quad P\left((X_1+X_2=4)\cap(X_1\times X_2=4)\right)=\dfrac{1}{36}\neq\dfrac{1}{12}\times\dfrac{3}{36}=P(X_1+X_2=4)\times P(X_1\times X_2=4)$

$\boxed{X_1+X_2 \text{ et } X_1\times X_2 \text{ ne sont pas indépendantes}}$

Loi Binomiale

Définition

Dans le cas où une épreuve a deux issues « S » et « E » de probabilités complémentaires p et q (q=1-p), si cette épreuve est répétée n fois de façon indépendante, la variable aléatoire X : nombre de réalisations de « S » suit la loi binomiale B(n,p).

Dans ce cas $P(X=k)=C_n^k p^k q^{n-k}$ l'espérance : E(X)= n p , et l'écart type : $\sigma_X = \sqrt{npq}$

Remarque

Dans la pratique, si $n \ldots 50$, $p < 0{,}1$, et np „ 10, on approche la loi binomiale par la loi de Poisson de paramètre $\lambda = n\,p$

Points de référence dans le programme

Elle est introduite en classe de terminale, et se retrouve dans les différents programmes de la plupart des BTS

Champs d'application

Elle permet de modéliser les situations d'événements à deux issues qui se répètent de façons indépendantes (succès, échec ; tirages avec remises…..).

Dans le cas de « tirages sans remise » dans une population très grande par rapport au nombre de tirages elle est utilisée en approximation de la loi Hypergéométrique.

Exemples

1. Si on joue à la roulette dans un casino il y a 30 résultats possibles pour chaque lancers de la boule ; soit X le nombre d'apparitions du « 1 » pour 20 parties. Quelle est la loi de X ? Si on joue le « 1 » aux 20 parties, quelle est la probabilité de gagner trois fois exactement ?
   Si on s'intéresse uniquement à l'apparition du « 1 » il n'y a alors que deux issues : soit le 1 apparaît la probabilité est alors 1/30 soit un autre nombre apparaît la probabilité est alors 29/30. Les 20 lancers sont indépendants.
   Donc X suit la loi binomiale B(20,1/30).
   La probabilité de gagner trois fois est alors : $P(X=3)=C_{20}^3\left(\frac{1}{30}\right)^3\left(\frac{29}{30}\right)^{17}$; $0{,}085$
2. Considérons que la fréquence des votes en France en faveur d'un candidat A est de 30%. Si on effectue un sondage sur 20 personnes et si on appelle X le nombre de réponses en faveur de A, quelle est la loi suivie par X ? Quelle est la probabilité que sur l'échantillon, plus de deux personnes aient voté pour A ?
   Le tirage est effectué sans remise donc X suit une loi Hypergéométrique,. Cependant ici le nombre de personnes votant en France étant très grand par rapport à l'échantillon, on assimile la loi de X à la loi Binomiale B(20 ; 0,3). Les calculs se feront alors avec la loi binomiale.

$$
\begin{aligned}
P(X>2) &= 1-P(X \text{ " } 2) \\
&= 1-P(X=0)-P(X=1)-P(X=2) \\
&= 1-0{,}7^{20}-C_{20}^1 0{,}3.\,0{,}7^{19}-C_{20}^2 0{,}3^2.\,0{,}7^{18} \\
&;\ 0{,}963
\end{aligned}
$$

Loi de probabilité

Définition

Dans le cas d'un univers fini : $\Omega = \{\omega_1, \omega_2 .... \omega_n\}$ une loi de probabilité discrète sur Ω est définie par la donnée des probabilités élémentaires : $p_i = P(\omega_i)$ (la somme des probabilités étant égale à 1).

Remarque : l'univers peut être aussi dénombrable.

Une loi de probabilité est continue si tous les évènements élémentaires sont de probabilité nulle ; l'univers est alors infini, un intervalle de $\mathbb{R}$ par exemple, la loi est alors définie par une fonction appelée densité de probabilité.

Points de référence dans le programme

Elle est introduite dès la seconde avec l'utilisation de la loi équirépartie dans les modélisations permettant les simulations, puis en terminale avec la loi uniforme ou exponentielle. En classe de BTS on introduira la loi de Poisson et la loi normale.

Champs d'application

Elles sont les supports théoriques dans la modélisation.

Exemples

1. Dans le cas du lancer d'un dé équilibré, l'univers étant l'ensemble des 6 numéros portés par les faces, quelle est la loi de probabilité ?

Chaque résultat a la même probabilité 1/6 d'apparaître il s'agit de la loi discrète uniforme ou équirépartie, définie par :

$p_i = 1/6$ pour tout $i \in \{1,2,..6\}$.

2. Soit *f* la fonction définie par : $\begin{cases} f(x) = ke^{-\lambda x} & pour \quad x \geq 0 \\ f(x) = 0 & pour \quad x < 0 \end{cases}$ $\quad \lambda \in]0, +\infty[$

Pour quelle valeur de « k » f est elle une loi de probabilité ?

Il faut que f remplisse deux conditions :

- $f(x) \geq 0 \quad soit \quad k \geq 0$
- $\int_0^{+\infty} ke^{-\lambda x} dx = 1$

$$\int_0^{+\infty} ke^{-\lambda x} dx = 1$$

$$\left[\frac{k}{-\lambda} e^{-\lambda x} \right]_0^{+\infty} = 1$$

$$\lim_{x \to +\infty} \frac{k}{-\lambda} e^{-\lambda x} + \frac{k}{\lambda} = 1$$

$$Soit : \frac{k}{\lambda} = 1 \quad d'où \quad k = \lambda$$

La loi ainsi définie est appelée la loi exponentielle définie par sa densité : $f(x) = \lambda e^{-\lambda x} \quad pour \ x \geq 0 \quad (\lambda > 0)$

Loi de Poisson

Définition

La loi de Poisson de paramètre $\lambda \in]0,+\infty[$, notée $P(\lambda)$, est la probabilité définie sur $\mathbb{N}$ par : $p_k = e^{-\lambda}\dfrac{\lambda^k}{k!}$ $k \in \mathbb{N}$.

Une variable aléatoire discrète X suit une loi de Poisson
Si - l'ensemble des valeurs de X est $\mathbb{N}$

- $P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$

On a : $E(X) = \lambda$ $\quad$ $\mathrm{Var}(X) = \lambda$ $\quad$ donc $\sigma(X) = \sqrt{\lambda}$

La fonction de répartition est définie sur $\mathbb{R}$ par :

pour $x \in [n, n+1[$, $n \in \mathbb{N}$, $F(x) = P(X \leq x) = e^{-\lambda}\sum_{k=0}^{n}\dfrac{\lambda^k}{k!}$

Remarque

C'est aussi la loi « limite » de la loi binomiale $B(n, p)$:

$B(n, p)$ converge en loi vers $P(\lambda)$ si np tend vers λ quand n tend vers $+\infty$.

Brièvement :

$$C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!}(1-p)^{-k} \times p^k \times (1-p)^n \qquad \text{et si } p = \frac{\lambda}{n}$$

$$= \frac{n!}{(n-k)!} \times \frac{1}{k!}\left(1-\frac{\lambda}{n}\right)^{-k} \times \frac{\lambda^k}{n^k} \times \left(1-\frac{\lambda}{n}\right)^n = \frac{n!}{(n-k)!} \times \frac{1}{n^k}\left(1-\frac{\lambda}{n}\right)^{-k} \times \frac{\lambda^k}{k!} \times \left(1-\frac{\lambda}{n}\right)^n$$

Si k est fixé, et si n tend vers $+\infty$, la limite est $e^{-\lambda}\dfrac{\lambda^k}{k!}$.

En effet $\displaystyle \lim_{n\to\infty}\frac{n!}{(n-k)!} \times \frac{1}{n^k} = \lim_{n\to\infty}\frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k} = 1$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1 \quad (k \text{ est fixé})$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

Dans la pratique :

Si $p < 0,1$, $n \geq 50$ et $np \leq 10$, on approche la loi binomiale par la loi de Poisson de paramètre $\lambda = np$

Point de référence dans le programme

Classes de BTS

Champs d'application
La loi de Poisson intervient dans la modélisation de certaines variables aléatoires liées à des phénomènes rares.
-nombre de pannes.
-nombre d'accidents.
-nombre de clients qui se présentent à un guichet.

Plus généralement elle intervient dans les Processus de Poisson homogènes :
Soit un événement instantané et rare pouvant se produire à tout instant et un nombre illimité de fois.
On considère le nombre d'observations de l'événement *N(t)* dans l'intervalle [0 ;t]

- Les accroissements de *N(t)* sont indépendants :
  Pour $0 = t_0 < t_1 < \ldots < t_k$, les variables aléatoires $N(t_1), N(t_2) - N(t_1), \ldots, N(t_k) - N(t_{k-1})$ sont indépendantes,
  En particulier : les variables aléatoires $N(t)$ *et* $N(t+\Delta t) - N(t)$ sont indépendantes,
- Le processus est homogène dans le temps :
  la loi de $N(t+\Delta t) - N(t)$ ne dépend que de la durée Δt de l'observation, son espérance est proportionnelle à Δt :
  $E\big(N(t+\Delta t) - N(t)\big) = \lambda \Delta t$
- Les événements sont isolés et rares :
  $P\big(N(t+\Delta t) - N(t) > 1\big) = \lambda \Delta t + \varepsilon$ $\qquad \lambda > 0$ et $\varepsilon > 0$
  ε négligeable devant Δt

Ainsi
Si T_1 est le temps d'attente du premier événement, $P(T_1 > t) = P\big(N(t) = 0\big)$
$P(T_1 > t + \Delta t) = P\big(N(t) = 0\big) \times P(N(t+\Delta t) - N(t) = 0)$,
On montre alors que T_1 suit la loi exponentielle de paramètre λ : $\gamma_1(\lambda)$,
et on montre que pour *t* fixé $P(N(t) = n) = \dfrac{(\lambda t)^k}{k!} e^{-\lambda t}$.

Exemple

Si X suit une loi binomiale $B(200, 10^{-2})$, pour calculer $P(X \leq k)$, on utilise la table de la fonction de répartition de Poisson de paramètre $\lambda = 2$
On lit par exemple : $P(X \leq 5) \approx 0,9834$.

le calcul avec la loi binomiale donne
$$\binom{200}{0} \times 0,01^0 \times 0,99^{200} + \ldots + \binom{200}{0} \times 0,01^5 \times 0,99^{195} = 0,9839770$$

On a donc ici une approximation de $P(X \leq 5)$ à 0,006 prés.

Loi exponentielle

Définition

La **loi exponentielle** de paramètre $\lambda \in \mathbb{R}_+^*$, notée $\gamma_1(\lambda)$, est la loi continue dont la **fonction de densité** est $\begin{cases} f(x) = \lambda e^{-\lambda x} & si\ x \geq 0 \\ f(x) = 0 & \text{sinon} \end{cases}$

Sa fonction de répartition est définie sur $\mathbb{R}$ par :

$$\begin{cases} F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} & x > 0 \\ F(x) = P(X \leq x) = 0 & \text{sinon} \end{cases}$$

Si une variable aléatoire continue X suit la loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$

On a : $\mathrm{E(X)} = \dfrac{1}{\lambda}$ et $\mathrm{Var(X)} = \dfrac{1}{\lambda^2}$ donc $\sigma(\mathrm{X}) = \dfrac{1}{\lambda}$

(la médiane est $T = \dfrac{1}{\lambda} \ln 2$, on résout $P(X \leq T) = \int_0^T \lambda e^{-\lambda t} dt = 1 - e^{-\lambda T} \leq \dfrac{1}{2}$)

Remarque

C est la loi continue qui correspond à la loi géométrique dans le cas discret.

Point de référence dans le programme

Terminale S (désintégration des noyaux) et BTS.

Champs d'application

- C'est la loi de durée de vie sans vieillissement, après rodage et avant usure.
  (c'est à dire à taux de survie constant).
- fiabilité.
- Temps d'attente d'un événement.
- les Processus de Poisson homogènes, voir « loi de Poisson ».

Exemple 1

la durée de vie moyenne d'un type d'ampoule est 3000 heures.
Quelle est la probabilité que la durée de vie d'une ampoule dépasse 4000 heures ?

La durée de vie d'une ampoule est une variable aléatoire qui suit la loi exponentielle de paramètre $\lambda = \dfrac{1}{3000}$

et $P(t \geq 4000) = 1 - P(t < 4000) = 1 - \left(1 - e^{-\frac{4}{3}}\right) \approx 0{,}2636$.

Exemple 2

On appelle X_t, la variable aléatoire réelle égale au nombre d'appels reçus, par un standard téléphonique pendant une période de durée *t* exprimée en heure.
On admet que X_t suit la loi de Poisson de paramètre λt, $\lambda \in \mathbb{R}_+^*$.

1. Déterminer la probabilité pour que X_t soit nulle.
2. Ayant choisi une origine des temps, on appelle Y la variable aléatoire réelle désignant le temps d'attente du premier appel.
   a. Exprimer en fonction de λ et t la probabilité: $P(Y > t)$, puis la probabilité $P(Y \leq t)$.
   b. En déduire que Y suit une loi exponentielle.

Solution :

1. *On a simplement* $P(X_t = 0) = e^{-\lambda t}\dfrac{(\lambda t)^0}{0!} = e^{-\lambda t}$.
2. *C'est la probabilité qu'il n'y a eu aucun appel durant le temps t,*
   donc $P(Y > t) = P(X_t = 0) = e^{-\lambda t}$.
   Et $P(Y \leq t) = 1 - P(Y > t) = 1 - e^{-\lambda t}$ *, pour* $t \geq 0$
   On reconnaît la fonction de répartition de la loi exponentielle de paramètre λ.

Loi normale

Définition (**loi normale** ou **loi de Laplace-Gauss**)

1. la loi normale centrée réduite : $N(0,1)$

C'est la loi **continue** dont la **fonction de densité** est définie par **:** $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ pour $x \in \mathbb{R}$**.**

Si une variable aléatoire continue X suit la loi normale centrée réduite, on a : $E(X)=0$ et $\mathrm{Var(X)}=1$ donc $\sigma(\mathrm{X})=1$.
Sa fonction de répartition est définie sur $\mathbb{R}$ par :
$F(x)=P(X \leq x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$. (Aussi notée $\Pi(x)$).

2. la loi normale de paramètres m et σ : $N(m,\sigma)$ est la loi **continue** dont la **fonction de densité** est définie par : $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$ pour $x \in \mathbb{R}$.

Si une variable aléatoire continue X suit la loi normale de paramètres m et σ , on a : $E(X)=m$ et $\mathrm{Var(X)}=\sigma^2$ donc $\sigma(\mathrm{X})=\sigma$
Sa fonction de répartition est définie sur $\mathbb{R}$ **par** :
$F(x)=P(X \leq x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2}dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{1}{2}t^2}dt=\Pi\left(\frac{x-m}{\sigma}\right)$.

(On se ramène à $N(0,1)$ par changement de variable en posant : $T=\frac{X-m}{\sigma}$).

Remarque

C'est la loi des erreurs sur les mesures physiques (« *seconde loi des erreurs* » de Laplace obtenue par un *théorème central limite*) et en particulier en astronomie (problème « *d'ajustement* » dans l'étude des orbites planétaires , la loi normale y est introduite par Gauss en recourant à la *méthode des moindres carrés*)

Point de référence dans le programme

Classes de BTS.

Champs d'application

Elle intervient notamment quand un phénomène dépend d'un grand nombre de paramètres, connus ou non, sans que l'un d'eux soit prédominant et dont les effets s'ajoutent.

-Poids, tailles, ou autres caractéristiques des êtres vivants.
-Mesures des grandeurs.

- Approximation d'une loi binomiale $B(n, p)$ (théorème de Moivre-Laplace) par la loi normale $N\left(np, \sqrt{npq}\right)$

Pratiquement : quand si $n \geq 30$, $np \geq 15$ et $npq \geq 5$ avec une **correction de continuité** (cf. exemples)

Exemples

1. Le taux de cholestérol est dit « normal » s'il est compris entre 1,60 g/l et 2,2 g/l, car on suppose la distribution de ces taux « normale » , c'est à dire que le taux de cholestérol est une variable suivant la loi normale d'espérance (ou moyenne) $m = 1,9$ et d'écart type $\sigma = 0,15$. D'autre part on considère « normal » le taux appartenant à l'intervalle $[m - 2\sigma; m + 2\sigma]$, lequel contient 95 % de la population.

Non normal signifie seulement extérieur à cet intervalle.

2. Approximation d'une loi binomiale par une loi normale

X suit la loi binomiale $B(50 ; 0,5)$ et on cherche $P(24 \leq X \leq 26)$

On approche la loi $B(50; 0,5)$ par la loi normale $N\left(25; \sqrt{12,5}\right)$

On considère X' qui elle suit $N\left(25; \sqrt{12,5}\right)$

On pose $T = \dfrac{X' - 25}{\sqrt{12,5}}$:

$$P(24 \leq X' \leq 26) = P\left(-\frac{1}{\sqrt{12,5}} \leq T \leq \frac{1}{\sqrt{12,5}}\right) = 2\Pi\left(\frac{1}{\sqrt{12,5}}\right) - 1 \approx 0,228$$

(en faisant une interpolation affine entre $\Pi(0,28)$ et $\Pi(0,29)$)

Or ce résultat n'est pas satisfaisant car, avec la loi binomiale,

$P(X = 24) = 0,107956897$, $P(X = 25) = 0,112275173$ et $P(X = 26) = 0,107956897$

donc $P(24 \leq X \leq 26) = 0,328188966$

Il faut affiner la façon dont on passe d'une v.a. discrète à une v.a. continue, c'est la **« correction de continuité »** :

On remplace 24 par l'intervalle [23.5 , 24.5], 25 par l'intervalle [24.5 , 25.5], 26 par l'intervalle [25.5 , 26.5].

Ainsi on remplace l'ensemble {24 , 25 , 26} des valeurs de la variable discrète par l'intervalle [23.5 , 26.5] pour la variable continue.

$$P(24 \leq X \leq 26) \approx P(23,5 \leq X' \leq 26,5) = 2\Pi\left(\frac{1,5}{\sqrt{12,5}}\right) - 1 \approx 0,334$$

De même $P(X \leq 25) \approx 0,556137586$

$$P(X \leq 25) \approx P(0 \leq X' \leq 25) = P\left(-\frac{25}{\sqrt{12,5}} \leq T \leq \mathbf{0}\right) = \Pi(0) - \left(1 - \Pi(5\sqrt{2})\right) \approx 0,5 \quad \text{et}$$

$$P(-0,5 \leq X' \leq 25,5) = P\left(-\frac{25,5}{\sqrt{12,5}} \leq T \leq \frac{0,5}{\sqrt{12,5}}\right) = \Pi(0,1\sqrt{2}) - \left(1 - \Pi(5,1\sqrt{2})\right) \approx 0,556$$

Loi uniforme

Définition

La loi uniforme sur $[a,b]$, notée $\mathcal{U}([a,b])$,est la loi continue dont la fonction de densité est $\begin{cases} f(x) = \dfrac{1}{b-a} & x \in [a,b] \\ f(x) = 0 & \text{sinon} \end{cases}$

Si une variable aléatoire continue X suit la loi uniforme $\mathcal{U}([a,b])$

$$P(\alpha \leq X \leq \beta) = \frac{1}{b-a}\int_{\alpha}^{\beta} dt = \frac{\beta-\alpha}{b-a} \qquad \text{si} \quad a \leq \alpha \leq \beta \leq b$$

On a : $E(X) = \dfrac{a+b}{2}$ et $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$ donc $\sigma(X) = \dfrac{b-a}{2\sqrt{3}}$.

Sa fonction de répartition est définie sur R par :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \qquad \text{donc} \quad F(x) = \begin{cases} 0 & si\ x \leq a \\ \dfrac{x-a}{b-a} & si\ a < x \leq b \\ 1 & si\ x > \mathrm{b} \end{cases}$$

Cas particulier

La loi uniforme sur $[0,1]$, notée $\mathcal{U}([0,1])$,est la loi continue dont la fonction de densité est

$$\begin{cases} f(x) = 1 & x \in [0,1] \\ f(x) = 0 & \text{sinon} \end{cases}$$

Si une variable aléatoire continue X suit la loi uniforme $\mathcal{U}([0,1])$

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} dt = \beta - \alpha \qquad \text{si} \quad 0 \leq \alpha \leq \beta \leq 1$$

On a : $E(X) = \dfrac{1}{2}$ et $\mathrm{Var}(X) = \dfrac{1}{12}$ donc $\sigma(X) = \dfrac{1}{2\sqrt{3}}$.

Sa fonction de répartition est définie sur R par :

$$F(x)=P(X \leq x)=\int_{-\infty}^{x} f(t)dt \qquad \text{donc} \qquad F(x)=\begin{cases} 0 & si\ x \leq 0 \\ x & si\ \ 0 < x \leq 1 \\ 1 & si\ x > 1 \end{cases}$$

Remarques

-Pour une variable aléatoire suivant une loi uniforme, la probabilité ne dépend que de la longueur de l'intervalle.

- Elle est au cas continu ce qu'est l'équiprobabilité au cas discret fini.

Point de référence dans le programme
Terminales S et classes de BTS.

Champs d'application
Calculs de probabilité, simulation.

Exemples

- Si l'on a rendez-vous avec une personne entre 10 h et 11 h (sans autres information) alors la probabilité que ce rendez-vous ait lieu entre 10 h 5 min et 10 h 10 min est $\frac{5}{60}=\frac{1}{12}$: On utilise la loi $\mathcal{U}([0,60])$.
- la touche Ramdon des calculatrices simule une variable suivant $\mathcal{U}([0,1])$

Compléments

La méthode d'inversion qui est une méthode générale, consistant à simuler une variable aléatoire de loi quelconque, en composant un appel de Ramdom avec sa **fonction quantile**.

La fonction quantile est « une fonction **inverse** » de la fonction de répartition :

Définition
Si X est une variable aléatoire à valeurs dans R et F sa fonction de répartition, la fonction quantile de X est la fonction :

$$\begin{cases}]0;1[\xrightarrow{Q} \mathbb{R} \\ u \longrightarrow Q(u)=\inf\{x / F(x) \geq u\} \end{cases}$$

Par convention, $Q(0)$ et $Q(1)$ sont respectivement la plus petite et la plus grande valeur de X *(éventuellement infinies).*

Dans le **cas continu** : par exemple

si f est une densité de X continue et strictement positive sur $[a;b]$ et nulle ailleurs, alors $F(x)=\int_{-\infty}^{x} f(t)dt$ pour $x \in \mathbb{R}$, est continue et strictement croissante de $[a;b]$ sur $[0;1]$ est donc bijective, et donc $Q(u)=\inf\{x / F(x) \geq u\}$ est l'unique x tel que $F(x)=u$

alors $Q=F^{-1}$ si on note ainsi la fonction réciproque de la restriction de F *à* $[a;b]$.

(si a et/ou b sont infinis, il faut ouvrir les crochets en conséquence)

Dans le **cas discret** :

La fonction de répartition et la fonction quantile sont des fonctions en escalier .

Si X prend les valeurs $x_1,...,x_n$, $n \in \mathbb{N}$, rangées dans l'ordre croissant :

pour , $x \in [x_k, x_{k+1}[$, $F(x)=P(X \leq x_k)=P(X=x_1)+...+P(X=x_k)$

$$Q(0)=x_1 \quad \text{et} \quad Q(u)=\begin{cases} x_1 & \text{si } u \in]0;F(x_1)] \\ & \dots \\ x_k & \text{si } u \in]F(x_{k-1});F(x_k)] \\ & \dots \end{cases}$$

On obtient la proposition suivante

Proposition :

Soit F_X la fonction de répartition sur $\mathbb{R}$ d'une variable aléatoire X,

Q la fonction quantile correspondante.

U une variable aléatoire suivant la loi uniforme $\mathcal{U}([0,1])$.

Alors la variable aléatoire $Y=Q \circ U$ a pour fonction de répartition F et donc $Y=Q \circ U$ suit la même loi que X.

Preuve :

$\forall x \in \mathbb{R}$,

$F_Y(x)=P(Y \leq x)=P(Q \circ U \leq x)$

$\{u \in [0;1] / Q(u) \leq x\}=\{u \in [0;1] / \inf\{y / F_X(y) \geq u\} \leq x\}=\{u \in [0;1] / \inf\{y / u \leq F_X(y)\} \leq x\}$

puisque F est croissante sur $\mathbb{R}$

On a $\{u \in [0;1] / Q(u) \leq x\}=\{u \in [0;1] / u \leq F_X(x)\}$

Donc

$P(Y \leq x)=P(U \leq F_X(x))=F_X(x)$.

Un exemple dans le cas continu :

Pour la loi exponentielle de paramètre λ, la fonction de répartition est

$F(x)=1-e^{-\lambda x}$ pour $x>0$ et 0 sinon.

$\forall u \in]0;1[\qquad Q(u)=\inf\{x / F(x) \geq u\}=x$ tel que $1-e^{-\lambda x}=u$

Donc $Q(u) = -\frac{1}{\lambda}\ln(1-u)$ $\qquad$ ($Q(0)=0$ et $\lim_{u\to 1} Q(u) = +°$)

D'où l'algorithme de simulation :
Puisque (Ramdon) et (1 – Ramdon) suivent la même loi

$$X \longleftarrow -\frac{\ln(Ramdon)}{\lambda}$$

Un exemple dans le cas discret
X prend les valeurs $\{x_1, x_2, ...\}$ rangées dans l'ordre croissant.

Notons F_k la valeur de la fonction de répartition sur l'intervalle $[x_k, x_{k+1}[$

Alors $Q(u) = x_k$ sur $]F_{k-1}, F_k]$
d'où l'algorithme de simulation :

$k \longleftarrow 1$

$U \longleftarrow Ramdon$
tant que $(U > F_k)$ *faire*

$k \longleftarrow k+1$
fin tant que $\qquad$ (alors $U \in]F_{k-1}, F_k]$)

$X \longleftarrow x_k$

Remarque
La méthode d'inversion n'est exacte qu'à condition de connaître l'expression explicite de F^{-1}, C'est rarement le cas. Si on veut appliquer la méthode à la loi normale par exemple, il faudra utiliser un algorithme d'approximation. En plus de l'imprécision, la méthode d'inversion sera alors relativement lente.

Même quand on connaît explicitement F^{-1}, la méthode d'inversion est rarement la plus efficace pour les variables continues.
En revanche, elle est efficace pour de nombreuses loi discrètes.

Probabilité conditionnelle

Définition

Soit *P* une probabilité quelconque sur un espace probabilisé $(\Omega, \mathcal{P}(\Omega))$.et soit $B \subset \Omega$ tel que $P(B) > 0$.

La **probabilité conditionnelle** P_B est la probabilité définie sur $(\Omega, \mathcal{P}(\Omega))$

par quelque soit $A \subset \Omega$, $P_B(A) = \dfrac{P(A \cap B)}{P(B)}$.

Remarque

L'événement B étant réalisé, on note $P(A/B)$ la « probabilité de A sachant B ».

C'est le nombre $P(A/B) = P_B(A) = \dfrac{P(A \cap B)}{P(B)}$.

Cas particulier

Si Ω est fini, et P est la probabilité uniforme,

	A	$\overline{A}$
B	$A \cap B$	$\overline{A} \cap B$
$\overline{B}$	$A \cap \overline{B}$	$\overline{A} \cap \overline{B}$

on remarque que, pour A et B non vides

$$P(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}, \quad P(A \cap B) = \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(\Omega)}, \quad P(B) = \frac{\mathrm{Card}(B)}{\mathrm{Card}(\Omega)}$$

Si B est réalisé, l'univers est réduit à B.

$$P(A/B) = \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(B)} \quad \text{et} \quad P(A/B) = \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(\Omega)} \cdot \frac{\mathrm{Card}(\Omega)}{\mathrm{Card}(B)} = \frac{P(A \cap B)}{P(B)}$$

Remarque : Si A et B sont indépendantes, $P(A/B)=P(A)$.

Point de référence dans le programme
Terminales, BTS.

Champs d'application
Tirages sans remise
Epreuves successives non indépendantes
Test d'hypothèses

Exemple

∞ Quelle est la probabilité de tirer un cœur dans un jeu de 32 cartes sachant que ce n'est pas un as ? (On admet qu'il y équiprobabilité)

A est l'évènement : « la carte est un cœur » , B est l'évènement : « la carte n'est pas un as »

$$P(A) = \frac{8}{32} = \frac{1}{4} \qquad P(B) = 1 - P(\overline{B}) = 1 - \frac{4}{32} = \frac{28}{32} = \frac{7}{8} \qquad P(A \cap B) = \frac{7}{32}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{7}{32}}{\frac{7}{8}} = \frac{1}{4} \qquad \textit{(ou } \frac{7}{28} \textit{ comme on pouvait s'y attendre).}$$

∞ Quelle est la probabilité d'avoir tiré un roi dans un jeu de 32 cartes sachant que l'on a tiré une figure ?

R est l'évènement : « la carte est un roi » , F est l'évènement : « la carte es t une figure »

$$P(R) = \frac{4}{32} = \frac{1}{8}$$

$$P(R/F) = \frac{P(R \cap F)}{P(F)} = \frac{\frac{4}{32}}{\frac{12}{32}} = \frac{1}{3}$$

Ici les évènements R et F ne sont pas indépendants.

Variable aléatoire continue

Définition :
Une variable aléatoire réelle continue est une variable X telle que pour tout x réel, la probabilité P(X=x) = 0. L'ensemble des valeurs prises par la variable aléatoire est un (ou plusieurs) intervalle (s) de IR.

Elle est définie par la donnée de sa fonction de répartition F définie par : $F(x) = P(X \leq x)$. ou par la densité de probabilité f de X ; f est la dérivée de F.

Remarques :
Soit X une variable aléatoire prenant ses valeurs sur l'ensemble D inclus dans $\mathbb{R}$

- $P(a \leq X \leq b) = F(b) - F(a)$.
- L'espérance de X (ou moyenne espérée) est définie par $E(X) = \int_D x\, f(x)dx$
- La variance de X est définie par $V(X) = E[X-E(X)]^2 = E(X^2)-[E(X)]^2$ où
  $E(X^2) = \int_D x^2\, f(x)dx$
- L'écart type est défini par $\sigma_X = \sqrt{V(X)}$

Points de référence dans le programme :
Loi uniforme, loi exponentielle en terminale et en BTS.
Loi normale en BTS.

Champs d'application :
Modélisation de phénomènes continus (mesures physiques, temps d'attente, . . .)

Remarque : Cela peut être l'occasion de montrer l'intérêt de la fonction de répartition.

Exemple :

Une personne doit se rendre en un lieu entre 12h et 13h. Calculer la probabilité qu'elle arrive entre 12h30mn et 12h45mn

L'heure d'arrivée de la personne est une variable aléatoire X qui suit la loi uniforme sur l'intervalle [0,1] (si on prend l'origine des temps à 12h), la densité de probabilité est alors définie par f(x)=1

On obtient : $P(0{,}5<X<0{,}75) = \int_{0,5}^{0,75} 1\, dx = [x]_{0,5}^{0,75} = 0{,}25$

Variable aléatoire discrète

Définition :
Soit un espace probabilisé (Ω,P(Ω),p), une variable aléatoire est une application de Ω dans $\mathbb{R}$. On dit quelle est discrète si l'ensemble de ses valeurs est fini ou dénombrable

Remarques :

Soit X une variable aléatoire prenant ses valeurs sur l'ensemble I = X(Ω)

- La loi de probabilité de X est définie par P(X=k) pour k élément de I.
- La fonction de répartition F(x) pour x réel est définie par : F(x) = P(X $\leq$ x)
- L'espérance de X (ou moyenne espérée) est définie par E(X) = $\sum_{x \in I} x\, P(X = x)$
- La variance de X est définie par V(X) = E[X-E(X)]² = E(X²)-[E(X)]²
- L'écart type est défini par $\sigma_X = \sqrt{V(X)}$

Points de référence dans le programme :
Terminales et BTS.

Champs d'application :
Gains aux jeux de hasard, . . .
Elle est associée aux lois de probabilités discrètes, loi binomiale, loi de Poisson . . .

Exemple 1

Un sac contient 2 jetons de 10 €, 3 jetons de 5 €, et 5 jetons de 0 €.
On tire deux jetons du sac, soit X la somme obtenue.

Quelle est la loi de X, son espérance, son écart type, sa fonction de répartition ?

- *L'ensemble des valeurs prises par X est :* X(Ω)= *{0,5,10,15,20}*
- *Card(*Ω) $= C_{10}^{2} = 45$
- La loi de X est définie par :

$$P(X=0)=\frac{C_5^2}{45}=\frac{10}{45}$$

$$P(X=5)=\frac{15}{45}$$

$$P(X=10)=\frac{C_3^2+2\times 5}{45}=\frac{13}{45}$$

$$P(X=15)=\frac{6}{45}$$

$$P(X=20)=\frac{1}{45}$$

- *E(X)= 0 P(X=0) +5 P(X=5) + . . . +20 P(X=20)* $= \dfrac{315}{45}=7$
- *V(X)= 0² P(X=0) +5² P(X=5) + . . . +20² P(X=20) -7²*; $27{,}11 \quad \sigma_X \; ; \; 5{,}21$

- *Fonction de répartition :*

$$Si\ x\in]-°\ ,0[\quad F(x)=0$$

$$Si\ x\in [0,5[\quad F(x)=\frac{10}{45}$$

$$Si\ x\in [5,10[\quad F(x)=\frac{25}{45}$$

$$Si\ x\in [10,15[\quad F(x)=\frac{38}{45}$$

$$Si\ x\in [15,20[\quad F(x)=\frac{44}{45}$$

$$Si\ x\in [20,+°\ [\quad F(x)=1$$

Exemple 2 :
On lance 4 fois un dé. Soit X le nombre de « 6 » obtenus. Quelle est la loi de X, l'espérance la variance ?

Pour chaque lancé deux possibilités :
- *On obtient le 6, p = 1/6*
- *On n'obtient pas le 6, q = 5/6*

Les lancés sont indépendants, donc X suit la loi binomiale B(4 ; 1/6)

- $P(X=k)=C_4^k\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{4-k}$

- E(X) = 4/6 = 2/3 .

- V(X)= 20/36 = 5/9.

Variance, écart type

I Variance

Compte tenu des difficultés théoriques pour donner une définition générale, on donnera une définition pour les cas rencontrés au lycée.

1. cas d'une variable aléatoire discrète finie

X prend les valeurs $x_1,\dots,x_n$ $\quad n \in \mathbb{N}$

On pose $p_i = P(X = x_i)$ pour tout *i*.

Définition

X une variable aléatoire discrète finie d'espérance $E(X) = \sum_{i=1}^{n} x_i p_i$

La **variance de X** est le réel positif $V(X) = \sum_{i=1}^{n} \left(x_i - E(X)\right)^2 p_i$

Remarque

- Par définition, $V(X) = E\left((X-\overline{X})^2\right)$, si on pose $E(X) = \overline{X}$,

  d'où $\quad V(X) = E(X^2 - 2X\overline{X} + \overline{X}^2) = E(X^2) - 2\overline{X}E(X) + \overline{X}^2) = E(X^2) - \overline{X}^2$

  et enfin $\quad V(X) = E(X^2) - \left(E(X)\right)^2$ (formule dite de *Huyghens-König*).

2. cas d'une variable aléatoire discrète infinie (mais dénombrable)

X prend les valeurs $x_i \in \mathbb{R}$, $i \in \mathbb{N}$

On pose $p_i = P(X = x_i)$ pour tout *i*.

On suppose que les séries $\sum_{i=1}^{+\infty} x_i\, p_i$ et $\sum_{i=1}^{+\infty} x_i^2\, p_i$ sont convergentes

Définition

X une variable aléatoire discrète d'espérance $E(X) = \sum_{i=1}^{+\infty} x_i p_i$

La **variance de X** est le réel positif $V(X) = \sum_{i=1}^{+\infty} \left(x_i - E(X)\right)^2 p_i$

ou encore $\quad V(X) = E(X^2) - \left(E(X)\right)^2 = \sum_{i=1}^{+\infty} x_i^2 p_i - \left(\sum_{i=1}^{+\infty} x_i\, p_i\right)^2$

3. Cas d'une variable aléatoire continue.

X *une variable aléatoire de densité* f *à support dans* $\mathbb{R}$: $P(a \leq X < b) = \int_a^b f(t)dt$

Si les intégrales $\int_{\mathbb{R}} tf(t)dt$ et $\int_{\mathbb{R}} t^2 f(t)dt$ existent.

Définition

X une variable aléatoire de densité f d'espérance $E(X) = \int_{\mathbb{R}} tf(t)dt$

La **variance de X** est le réel positif $V(X) = \int_{\mathbb{R}} \left(t - E(X)\right)^2 f(t)dt$

ou encore $\quad V(X) = E(X^2) - \left(E(X)\right)^2 = \int_{\mathbb{R}} t^2 f(t)dt - \left(\int_{\mathbb{R}} tf(t)dt\right)^2$

Remarques

- La variance mesure la dispersion des valeurs prises par la variable aléatoire autour de la sa valeur centrale qui est l'espérance.
- La fonction $u \longrightarrow E\left((X-u)^2\right)$ est minimale pour $u = E(X)$
- $\forall (a,b) \in \mathbb{R}^2 \quad V(aX+b) = a^2 V(X)$

II. Ecart type

Définition

X une variable aléatoire admettant une variance $V(X)$

L'écart type de X est le nombre $\sigma = \sqrt{V(X)}$

Remarques

- L'écart-type mesure la dispersion des valeurs prises par la variable aléatoire autour de la valeur centrale qui est l'espérance, mais dans l'unité de la variable.
- La lettre s est réservée à la valeur numérique de l'écart type d'un échantillon.

Point de référence dans le programme

Terminales, BTS.

Champs d'application

La variance et l'écart type permettent en probabilité de mesurer le « risque », ce sont des paramètres importants pour différentes lois.

Exemple 1

On lance deux dés

X est la somme des faces, sa loi est alors

x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$E(X) = \sum_{i=1}^{11} x_i p_i = 7$

$$V(X)=\sum_{i=1}^{11}\left(x_i-7\right)^2 p_i=\frac{35}{6} \qquad \text{ou} \qquad V(X)=\sum_{i=1}^{11} x_i^2 p_i-7^2=\frac{329}{6}-49=\frac{35}{6} \quad \text{et} \quad \sigma=\sqrt{\frac{35}{6}}$$

Compléments

- **Cas de la loi de Poisson**

X suit la loi de Poisson de paramètre $\lambda\in]0,+\infty[$

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!} \qquad k\in \mathbb{N}$$

- $\forall n\in\mathbb{N}$, $\displaystyle\sum_{k=0}^{n} k^2\times e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=1}^{n}\frac{k\lambda^k}{(k-1)!}$

$$\sum_{k=1}^{n}\frac{k\lambda^k}{(k-1)!}=\lambda+\lambda^2\sum_{k=2}^{n}\frac{\lambda^{k-2}}{(k-2)!}\frac{k}{k-1}\leq \lambda+2\lambda^2\sum_{k=2}^{n}\frac{\lambda^{k-2}}{(k-2)!}\leq \lambda+2\lambda^2 e^{\lambda}$$

Donc $\displaystyle\sum_{k=0}^{+\infty} k^2\times e^{-\lambda}\frac{\lambda^k}{k!}$ est convergente

- $\displaystyle E(X)=\sum_{k=0}^{+\infty} k\times e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\times\lambda\times e^{\lambda}=\lambda$ car $\displaystyle\sum_{k=0}^{n} k\times\frac{\lambda^k}{k!}=\sum_{k=1}^{n}\frac{\lambda^k}{(k-1)!}=\lambda\sum_{k=1}^{n}\frac{\lambda^{k-1}}{(k-1)!}$

- $\displaystyle V(X)=E(X^2)-\left(E(X)\right)^2=\sum_{i=1}^{+\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!}-\lambda^2$

$$\sum_{i=1}^{+\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{i=1}^{+\infty} k^2\frac{\lambda^k}{k!}$$

$$\sum_{k=1}^{n}\frac{k^2\lambda^k}{k!}=\sum_{k=1}^{n}\frac{k\lambda^k}{(k-1)!}=\lambda\sum_{k=1}^{n}\frac{d}{d\lambda}\left(\frac{\lambda^k}{(k-1)!}\right)=\lambda\frac{d}{d\lambda}\left(\sum_{k=1}^{n}\frac{\lambda^k}{(k-1)!}\right)=\lambda\frac{d}{d\lambda}\left(\lambda\sum_{k=1}^{n}\frac{\lambda^{k-1}}{(k-1)!}\right)$$

On peut en déduire (séries de fonctions réelles de la variable réelle λ)

La convergence normale sur tout intervalle borné de $\mathbb{R}_+$, donc uniforme de $\displaystyle\sum_{k=1}^{+\infty}\frac{d}{d\lambda}\left(\frac{\lambda^k}{(k-1)!}\right)$

et $\displaystyle\sum_{i=1}^{+\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{i=1}^{+\infty} k^2\frac{\lambda^k}{k!}=\lambda e^{-\lambda}\frac{d}{d\lambda}\left(\lambda\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}\right)=\lambda e^{-\lambda}\frac{d}{d\lambda}\left(\lambda e^{\lambda}\right)=\lambda(\lambda+1)$

D'où enfin $V(X)=\lambda(\lambda+1)-\lambda^2=\lambda$ **et** $\sigma=\sqrt{\lambda}$

- **Cas de la loi exponentielle**

Si X suit la loi exponentielle de paramètre $\lambda\in\mathbb{R}_+$, la fonction de densité est définie par

$$\begin{cases} f(x)=\lambda e^{-\lambda x} & si\ x\geq 0 \\ f(x)=0 & \text{sinon} \end{cases}$$

On vérifie que $\displaystyle\int_{\mathbb{R}} t^2 f(t)dt$ existe :

$$\int_0^x t^2\lambda e^{-\lambda t}dt = \left[-t^2e^{-\lambda t}\right]_0^x + 2\int_0^x te^{-\lambda t}dt$$

$$= -x^2e^{-\lambda x} + 2\left(\left[-t\frac{e^{-\lambda t}}{\lambda}\right]_0^x + \int_0^x \frac{e^{-\lambda t}}{\lambda}dt\right)$$

$$= -x^2e^{-\lambda x} + 2\left(-\frac{xe^{-\lambda x}}{\lambda} + \left[-\frac{e^{-\lambda t}}{\lambda^2}\right]_0^x\right)$$

$$= -x^2e^{-\lambda x} + 2\left(-\frac{xe^{-\lambda x}}{\lambda} + \frac{1}{\lambda^2} - \frac{e^{-\lambda x}}{\lambda^2}\right)$$

Donc $\lim\limits_{x\to°} \int_0^x t^2\lambda e^{-\lambda t}dt = \dfrac{2}{\lambda^2}$ et on sait que $E(X) = \int_0^{°} \lambda te^{-\lambda t}dt = \dfrac{1}{\lambda}$

Ainsi $V(X) = \int_0^{°} t^2 f(t)dt - \left(\dfrac{1}{\lambda}\right)^2 = \dfrac{1}{\lambda^2}$ et $\sigma = \dfrac{1}{\lambda}$

∞ **Iinégalité de Bienaymé-Tchébychev**

Pour toute variable aléatoire X admettant une espérance et une variance, et pour tout $\varepsilon > 0$

$$P\left(\left|X - E(X)\right| \ldots \varepsilon\right) „ \frac{\sigma^2}{\varepsilon^2}$$

Preuve *(pour le cas d'une variable continue)*

On utilise l'inégalité de Markov :

Pour toute variable aléatoire X ***positive*** *de densité f admettant une espérance, et pour tout* $\delta > 0$

$$P\left(X \ldots \delta E(X)\right) „ \frac{1}{\delta}$$

En effet, $E(X) = \int_{R} tf(t)dt \ldots \int_D tf(t)dt$ si $D = \left\{x \in R / x \ldots \delta E(X)\right\}$

$$E(X) \ldots \int_D tf(t)dt \ldots \delta E(X)\int_D f(t)dt = \delta E(X)P(X \ldots \delta E(X))$$

D'où l'inégalité.

On considère alors la variable $Y = (X - E(X))^2$ *qui est positive, et admet une espérance* $E(Y) = E(X - E(X))^2 = V(X) = \sigma^2$

L'inégalité de Markov appliquée à Y et $\lambda = \dfrac{\varepsilon^2}{\sigma^2}$ *donne alors* $P\left(Y \geq \dfrac{\varepsilon^2}{\sigma^2}\sigma^2\right) \leq \dfrac{\sigma^2}{\varepsilon^2}$

Soit $P\left(\left(X - E(X)\right)^2 \geq \varepsilon^2\right) \leq \dfrac{\sigma^2}{\varepsilon^2}$,

et comme la fonction carrée est croissante sur $\mathbb{R}_+$: $\left(X - E(X)\right)^2 \geq \varepsilon^2 \Leftrightarrow \left|X - E(X)\right| \geq \varepsilon$

$$P\left(\left|X - E(X)\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

(majoration grossière, elle ne dépend pas de la loi)

STATISTIQUES INFERENTIELLES

Adéquation à une loi

Distribution d'échantillonnage

Estimation ponctuelle

Estimation par intervalles

Fluctuation d'échantillonnage

Modélisation

Simulation

Tests d'hypothèses

Adéquation à une loi

Définition
Etant donné une série statistique et un modèle théorique, un test d'adéquation conduit à comparer les effectifs observés et les effectifs théoriques et à décider si les différences entre ces effectifs sont significatives.

Point du programme
Terminale, BTS.

Champs d'application
Si l'adéquation à une loi n'a pas été rejetée, le caractère étudié pourra être modélisé à l'aide de cette loi.
Dans le cadre des programmes, on se limitera à un modèle d'équirépartition.

Exemple
On souhaite savoir si un dé à six faces est bien équilibré. On le jette 60 fois dans les mêmes conditions et on observe le nombre n_i d'apparitions de chaque face ($i \in \{1,2,3,4,5,6\}$). On obtient :

N° de la face	1	2	3	4	5	6
Effectif observé n_i	15	7	4	11	6	17
Fréquence observée f_i	15/60	7/60	4/60	11/60	6/60	17/60

Le modèle théorique retenu sous l'hypothèse H_0 : « le dé est parfaitement équilibré » est la loi uniforme discrète. Les fréquences théoriques sont donc $p_i = 1/6$ pour chaque face.

On calcule la quantité $d^2 = \sum_{i=1}^{k} (f_i - p_i)^2$ *=0.038*

Si d^2 est « suffisamment » petit (c'est-à-dire inférieur à une valeur critique) on pourra considérer que les écarts entre les fréquences observées et les fréquences théoriques sont dues à des fluctuations d'échantillonnage, et qu'il n'y a pas d'objection à adopter le modèle théorique

Afin de déterminer cette valeur critique, on se fixe un seuil α (dit risque de première espèce ici α = 10%).

On simule sous l'hypothèse H_0 d'équirépartition, un « grand » nombre de fois une série de 60 lancers : par exemple 1000 fois une série de 60 entiers au hasard entre 1 et 6. Pour chaque série, on calcule la valeur de d^2. On obtient ainsi une série de 1000 valeurs de d^2 dont on calcule le 9° décile λ : on obtient λ =0.04
Pour 1000 échantillons de 60 lancers d'un dé équilibré, 90% donnent d^2<0.04
Dans le cas du dé que l'on veut tester, d^2 =0.038<0.04 : il n'y a pas de raison de refuser l'hypothèse de l'équirépartition, au seuil α = 10%.

Remarques

- rien ne permet de dire s'il y a véritablement équirépartition.

- si on avait trouvé $d^2 > 0.04$, on aurait refusé l'hypothèse d'équirépartition avec un risque de de se tromper de10%.

- Le choix du seuil α impose le choix du quantile calculé.

- Plus généralement le test du Khi-deux permet de tester l'adéquation à une loi. Le critère ci-dessus (qui est une adaptation du test du Khi-deux) n'est utilisable que dans le cas de l'équirépartition et dans le cas où les simulations (sur lesquelles le quantile est calculé) sont faites en refaisant un grand nombre de fois (ici 1000 fois) le **même** nombre n d'observations (ici n= 60).

- *La quantité utilisée dans le test du khi-deux est* $\sum_{i=1}^{k} \frac{(o_i - n_i)^2}{o_i}$ *où les* o_i *et* n_i *sont respectivement les effectifs observés et théoriques. Dans le cas d'un test d'adéquation à une loi uniforme:*

  Soit θ_i *la variable aléatoire : nombre d'observations de la classe i.*

  Soit $F_i = \frac{\theta_i}{n}$ *la fréquence empirique de la classe i.*

  Soit $D^2 = \sum_{i=1}^{k} (F_i - \frac{1}{k})^2$

  $$\sum_{i=1}^{k} \frac{(\theta_i - n_i)^2}{n_i} = \sum_{i=1}^{k} \frac{(nF_i - np_i)^2}{np_i} = \sum_{i=1}^{k} n\frac{(F_i - p_i)^2}{p_i} = n\sum_{i=1}^{k} \frac{(F_i - \frac{1}{k})^2}{\frac{1}{k}} = k\,n\,D^2.$$

  La loi de nD^2 *dépend du nombre k de classes et non de n.*

Distribution d'échantillonnage

Définition
Soit une population sur laquelle est étudié un caractère X, considérons tous les échantillons de même taille n ($n \in \mathbb{N}^*$) issus de cette population. Pour chacun de ces échantillons, on peut calculer un paramètre statistique (par exemple une moyenne, une variance, un écart type, etc…) dont les valeurs observées sont liées aux fluctuations d'échantillonnage.
La distribution de ces paramètres est appelée distribution d'échantillonnage.

Tant qu'un tel échantillon n'a pas été choisi, les paramètres statistiques étudiés sont des variables aléatoires suivant un ensemble de lois appelés distribution des variables d'échantillon.

Dans le cas d'une moyenne, la loi de distribution d'échantillonnage est donnée par le théorème suivant.

Théorème limite central :
Si X_1 , X_2 , X_3... , X_n sont des V.A. indépendantes de même loi, d'espérance E(X) = m et d'écart type σ, alors, si $n \to +\infty$:
∞ *La loi de $S = X_1 + X_2 + ... + X_n$ converge vers la loi normale de moyenne nm et d'écart type* $\sigma\sqrt{n}$.
∞ *La loi de probabilité de la moyenne $\overline{X} = \frac{S}{n}$ tend vers la loi normale de moyenne m, d'écart type* $\frac{\sigma}{\sqrt{n}}$.

Points de référence du programme
Classes de BTS

Champs d'application
La connaissance des distributions des variables d'échantillon permet de faire de la statistique inférentielle (estimation, tests d'hypothèse…)

Exemple
La population est l'ensemble des ampoules fabriquées en un mois par une entreprise.
X est la variable aléatoire durée de vie d'une ampoule. On suppose que X suit une loi normale N(1200,60).
On extrait de la fabrication un échantillon aléatoire de taille n=100 avec remise. Quelle est la probabilité que la durée de vie d'un échantillon de taille 100 soit supérieure à 1215 h.

La variable aléatoire $\overline{X}$ moyenne d'un échantillon quelconque suit la loi normale
$N\left(1200, \frac{60}{\sqrt{100}}\right)$, soit $N(1200\,,\,6)$

$$
\begin{aligned}
P(\overline{X} > 1215) &= 1 - P(\overline{X} \leq 1215) \\
&= 1 - P\left(\frac{\overline{X} - 1200}{6} \leq \frac{1215 - 1200}{6}\right) \\
&= 1 - F(2,5) \\
&\approx 0,007
\end{aligned}
$$

Remarques

- Si la population est grande par rapport à la taille de l'échantillon il est fréquent d'assimiler un échantillon issu d'un tirage sans remise à un échantillon issu d'un tirage avec remise.

- Si la variable aléatoire X n'est pas gaussienne et n grand (dans la pratique n>30), le théorème central limite conduit aux mêmes conclusions.

Estimation ponctuelle

Définition
Estimation d'un paramètre
Etant donné un caractère X, étudié sur une population, dont le modèle dépend d'un paramètre θ inconnu, une estimation de θ est une valeur numérique attribuée à θ.
Dans la pratique, cette estimation est fournie par un échantillon.

Remarque
L'estimation ponctuelle a pour intérêt de fournir une valeur numérique pour un paramètre inconnu. Par contre cette valeur numérique est soumise aux aléas de l'échantillon.

Points de référence du programme
Classes de B.T.S.

Champ d'application:

- une estimation d'une moyenne de la population est la moyenne observée $\overline{x}$ sur un échantillon.
- une estimation d'une proportion de la population est la fréquence observée f sur l'échantillon.
- une estimation de la variance de la population est la « variance corrigée » observée sur un échantillon :
  Pour un échantillon de taille n ($n \in \mathbb{N}^* \setminus \{1\}$) dont les valeurs observées sont $(x_1, x_2, ..., x_n)$ la « variance corrigée » est $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ alors que la variance est $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$.
  Si s^2 est la variance de l'échantillon, la variance corrigée est : $\frac{n}{n-1}s^2$

Exemple :
On mesure la résistance d'un ciment en moulant des éprouvettes cylindriques et en observant la pression (en kg/cm^2) à partir de laquelle elles se brisent (essai de compression simple). Pour 10 cylindres utilisés on a observé les pressions suivantes :
19.6 , 19.9 , 20.4 , 19.8 , 20.5 , 21.0 , 18.5 , 19.7 , 18.4 , 19.4
On suppose que la pression X est une variable aléatoire normale dont la moyenne est m et l'écart type σ.
Donner une estimation ponctuelle de m et σ^2.

Solution : *La moyenne de l'échantillon des 10 cylindres est* $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 19.72$ *et la variance corrigée est* $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2$ *qui donne s = 0.823*

On propose donc pour estimation de m : $m = 19.72$ kg/cm^2
et pour estimation de : $\sigma = 0.823$ kg/cm^2.

Estimation par intervalle

Définitions :

- **Intervalle estimateur d'un paramètre.**

Soit X un caractère étudié sur une population, dont le modèle dépend d'un paramètre θ inconnu.

Etant donné : un échantillon aléatoire ($X_1, X_2, \ldots, X_n$) issu de la population
un réel α dans [0 ; 1],
deux fonctions A et B des n variables aléatoires X_i ,

un intervalle aléatoire estimateur de θ, au niveau de confiance (ou au seuil de confiance) $1-\alpha$, est un intervalle]A ; B[ayant la probabilité $1-\alpha$ de contenir θ.

- **Intervalle de confiance :**

Etant donné un caractère X, étudié sur une population, dont le modèle dépend d'un paramètre θ inconnu, une estimation par intervalle de θ, au niveau de confiance (ou au seuil de confiance) $1-\alpha$, est la réalisation d'un intervalle estimateur de θ au niveau de confiance $1-\alpha$, au vu d'un échantillon.

Remarques :

On ne peut pas savoir si θ est ou non dans l'intervalle fourni par le seul échantillon de taille n choisi.

Si θ appartient à cet intervalle, rien ne permet de savoir s'il est ou non proche de son centre.

La méthode proposée, si par exemple $\alpha = 0.05$, donnera dans 95% des cas un intervalle qui contient θ.

Une estimation par intervalle prend en compte la taille de l'échantillon et le niveau de confiance choisi.

Points de référence du programme :

Classes de seconde (fourchettes de sondage) et B.T.S. (statistiques inférentielles).

Champ d'application :

Sondages, etc.

Exemple 1

Une estimation par intervalle de la moyenne de la population, dans le cas où X est normale ou dans le cas où n est grand ($n \geq 30$) et dans le cas où l'écart-type σ est connu est $]\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} ; \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}[$ (où $\overline{x}$ est la moyenne de l'échantillon).

Démonstration

Soit $\overline{X}$ la variable aléatoire moyenne d'échantillon. On sait que $\frac{\overline{X} - m}{\sigma / \sqrt{n}} \to N(0,1)$

Comme la loi normale N(0 ;1) est symétrique, on choisit de centrer l'intervalle d'estimation .

Or prob($-u < \frac{\overline{X}-m}{\sigma/\sqrt{n}} < u$*)=1-*$\alpha$ *est équivalent à* $2F(u)-1=1-\alpha$ *(F fonction de répartition de la loi normale centrée réduite).*

Donc prob($\overline{X}-u\frac{\sigma}{\sqrt{n}} < m < \overline{X}+u\frac{\sigma}{\sqrt{n}}$*)=1-*$\alpha$

L'intervalle]$\overline{X}-u\frac{\sigma}{\sqrt{n}};\overline{X}+u\frac{\sigma}{\sqrt{n}}$*[est l' intervalle estimateur de m au seuil de confiance* $1-\alpha$.

En donnant à $\overline{X}$ la valeur $\overline{x}$ fournie par l'échantillon, on obtient un intervalle de confiance pour m au niveau de confiance $1-\alpha$, qui est : $]\overline{x}-u\frac{\sigma}{\sqrt{n}};\overline{x}+u\frac{\sigma}{\sqrt{n}}[$

Le risque α le plus utilisé est 5% . dans ce cas $u=1.96$ d'où le résultat.

Application numérique

La distribution des tailles, en centimètres, des individus d'une population est supposée suivre une loi normale dont la moyenne m est inconnue et dont l'écart-type est $\sigma=10$ cm. On choisit un échantillon aléatoire et indépendant de $n=49$ individus de cette population et on observe sur cet échantillon une taille moyenne $\overline{x}=178$ cm.

Une estimation par intervalle de m au niveau de confiance 95% est $]178-1.96\frac{10}{7};178+1.96\frac{10}{7}[$

soit : $]175.2;180.8[$

Exemple 2

Une estimation par intervalle d'une proportion de population au risque 5% est $]f-1.96\sqrt{\frac{f(1-f)}{n}}\ ;\ f+1.96\sqrt{\frac{f(1-f)}{n}}[$ (f fréquence observée sur un échantillon de taille n et $n\geq 30$) parfois remplacé par $]f-\sqrt{\frac{1}{n}}\ ;\ f+\sqrt{\frac{1}{n}}[$

Démonstration

Soit A un évènement de probabilité p à estimer. *Les variables d'échantillon X_i sont des variables de Bernouilli (X_i = 1 si l'observation n°i a fourni la réalisation de A, X_i = 0 sinon) de même moyenne p et de même variance p(1-p) où p désigne la proportion de la population à estimer. Or si n est grand et p pas trop petit (en pratique n> 30 et* $0.3 \leq p \leq 0.7$*), d'après le théorème central limite, la variable aléatoire fréquence d'échantillon* $F=\frac{1}{n}\sum_{i=1}^{n}X_i$ *suit approximativement la loi normale N(p,* $\sqrt{\frac{p(1-p)}{n}}$*).*

On procède alors comme dans l'exemple précédent.

- *Soit on remplace* $\sigma=\sqrt{\frac{pq}{n}}$ *par* $\sqrt{\frac{f(1-f)}{n}}$
- *Soit on considère que p(1-p) a pour maximum ¼, donc pour tout p*$\in[0;1]$,

En arrondissant 1.96 à 2, on obtient l' intervalle]$f-\sqrt{\frac{1}{n}}\ ;\ f+\sqrt{\frac{1}{n}}$*[qui contient]*$f-1.96\sqrt{\frac{p(1-p)}{n}}\ ;\ f+1.96\sqrt{\frac{p(1-p)}{n}}$ *[(Cf. programme de seconde)*

Fluctuation d'échantillonnage

Définition
Si on choisit des échantillons de taille « n » dans une population et si on observe dans cet échantillon la moyenne, la médiane, la variance ou toute autre fonction de la variable statistique ces valeurs fluctuent d'un échantillon à l'autre.

Points de référence dans le programme :
En classe de seconde les élèves doivent être sensibilisés à cette notion.
Elle est liée à la notion de test statistique introduite en terminale.
En BTS elle est liée aux statistiques inférentielles.

Champs d'application
Cette prise de conscience permet par la suite d'éclairer la notion de » décision statistique » dans les tests.
Elle permet d'avoir un regard critique sur les résultats de sondages.

Exemples

On dispose d'une urne contenant 40 boules vertes et 60 boules rouges. On désire observer la fréquence d'apparition des boules vertes sur un tirage de 30 boules avec remise.

On tire 30 fois une boule avec remise et on note le nombre de boules vertes obtenues, on obtient ainsi leur fréquence d'apparition sur un tirage.
Si on répète 100 fois le tirage de 30 boules on observera des fréquences différentes, la fluctuation d'échantillonnage sera alors mise en évidence.
Si on répète 50 fois l'expérience précédente, on peut aussi observer la fluctuation de la fréquence la plus haute (ou la plus basse) ou la fréquence médiane ou toute autre fonction de ces fréquences sur les 100 tirages.

L'utilisation d'un tableur ou d'une calculatrice permet de simuler ces expériences ; il suffit par exemple avec la fonction « aléa » de choisir un nombre aléatoire entre 1 et 10. Le tirage d'une boule verte correspondra à un nombre de 1 à 4, celui d'une boule rouge à un nombre de 5 à 10.

Modélisation

Définition

En statistique, modéliser, c'est trouver une loi de probabilité qui rende compte d'un phénomène aléatoire.

Points de référence dans le programme :
Elle intervient dès la classe de seconde, en effet pour mettre en place des simulations on est amené à utiliser un modèle, qui sera au lycée une loi équirépartie.
En terminale les modélisations utilisent la loi uniforme, la loi binomiale ou la loi exponentielle.

Champs d'application
Au lycée elle permet de faire des simulations sur calculatrice ou sur tableur dans le but d'observer des fluctuations d'échantillonnage lorsque l'on veut des résultats sur des échantillons assez grands et que l'expérimentation devient trop lourde.

Par la suite, pour modéliser une situation, de nombreuses notions mises en place au lycée (histogramme, boites à moustache, résumés statistiques,…..) permettront de faire le choix d'un modèle adapté.

Exemple

Pour modéliser le lancé du dé, on utilisera la loi équirépartie (probabilité 1/6 pour chaque face).
On peut alors simuler le lancer du dé avec une calculatrice ou bien un tableur, voir « simulation, exemple 1 »

Simulation

Définition
Effectuer une simulation, c'est remplacer une expérience réelle par une expérience virtuelle donnant à priori les mêmes résultats.

Remarque : Etant donné un phénomène, il faut déterminer un modèle, puis le faire fonctionner sur tableur ou calculatrice. On peut ainsi obtenir un grand nombre de réalisations de l'expérience par simulation.

Points de référence dans le programme :
En classe de seconde est introduite la notion de simulation pour remplacer le lancé d'une pièce ou d'un dé. En terminale, elle intervient aussi pour mettre en place la notion de test d'adéquation à la loi uniforme.

Champs d'application
Si on veut observer un phénomène en le répétant plusieurs fois, l'expérimentation peut devenir vite fastidieuse, et si elle est nécessaire au départ elle pourra si on dispose d'un modèle, être remplacée par une simulation qui donnera rapidement de nombreux résultats.
Elle devient nécessaire dans le cas ou ne connaît pas de modèle théorique du phénomène, par exemple le lancer de deux pièces qui pourra être simulé à partir du lancé d'une pièce qui lui peut être modélisé avec la loi équirépartie.
La simulation d'un très grand nombre d'expériences permettra éventuellement de choisir un modèle du phénomène, ici le résultat du lancer de deux pièces.

Exemples

1. Si on veut observer le lancé d'un dé 500 fois on peut simuler cette expérience sur calculatrice ou sur tableur en utilisant la loi équirépartie

La fonction « aléa » du tableur ou la touche « random »de la calculatrice donne un nombre aléatoire entre 0 et 1. En prenant 6 fois la partie entière de ce nombre plus 1 on obtient un nombre aléatoire prenant une valeur entière entre 1 et 6.En répétant ceci 500 fois on obtient par simulation la réalisation de 500 lancers du dé.

2. Si on veut observer la somme de deux dés on ne dispose pas en seconde de modèle. Cependant à partir de la simulation précédente du lancé d'un dé, on pourra simuler le lancé de deux dès.

Il suffit pour cela d'utiliser deux listes obtenues comme précédemment et de faire la somme des termes deux par deux. On dispose alors des résultats obtenus par simulation du lancer de deux dés 500 fois.

Tests d'hypothèses

Définition

Les tests permettent de rejeter ou à défaut d'accepter une hypothèse, portant généralement sur un paramètre de la population (moyenne, proportion etc..) ou sur le choix d'un modèle.
Dans la pratique, on fixe un seuil critique α , puis à partir d'un échantillon donné, on refusera (ou on acceptera) l'hypothèse.

Lorsqu'on effectue un test par rapport à une hypothèse H_0, contre l'hypothèse alternative H_1 quatre décisions sont possibles :

- On accepte H_0 alors que H_0 est vraie.
- On rejette H_0 alors que H_0 est vraie, l'erreur commise est α, appelée erreur de première espèce.
- On accepte H_0 alors que H_0 est fausse, l'erreur commise β, est appelée erreur de deuxième espèce.
- On refuse H_0 alors que H_0 est fausse.

Construction du test voir exemple 1

Points de référence dans les programmes
En classe de BTS, les tests de conformité pour une moyenne ou une proportion.
En terminale les tests d'adéquation à une loi équirépartie.

Champs d'application
Dans l'industrie, la conformité d'une partie de la production . . .
Le choix de modèles probabilistes pour décrire des phénomènes.

Exemple 1 :
Test de conformité.
Au niveau national, le pourcentage d'étudiants qui réussissent à un examen donné est :
$p_0 = 62\%$.
Dans une faculté où 400 étudiants ont présenté cet examen, le taux de réussite est : $f = 65\%$.
La différence entre les deux taux de réussite est elle significative ?

Les quatre points suivants constituent la construction du test :

1. *Supposons l'échantillon issu d'une population où la fréquence de réussite est « p ».*

Soit F la variable aléatoire, fréquence d'un échantillon quelconque de taille n (n>30) issu de cette population. Alors F suit la loi : $N\left(p, \sqrt{\frac{pq}{n}}\right)$ *(Théorème central limite)*

2. *Hypothèse nulle :* $H_0 : p = p_0$
   Hypothèse alternative : $H_1 : p \neq p_0$*, il s'agit d'un test bilatéral.*

Sous l'hypothèse H_0, F suit la loi : $N\left(p_0, \sqrt{\frac{p_0 q_0}{n}}\right)$ *ou* $\frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \equiv N(0,1)$

3. *On choisit un seuil α par exemple 0,05 on lit sur la table N(0,1) la valeur critique : u=1,96 (lue à 0,975)*

On obtient : $0,95 = P\left(-1,96 < \frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < 1,96 \right)$

4. *Règle de décision :*

Posons : $T = \frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Si pour un échantillon donné, F prend la valeur « f » et T la valeur « t »

- *Si :* $t \notin]-1,96\,;1,96[$ *on rejette l'hypothèse H_0 au risque 5%.*
- *Si :* $t \in]-1,96\,;1,96[$ *on n'a pas de raison de refuser l'hypothèse H_0, donc on l'accepte au risque β (indéterminé).*

5. *Utilisation du test :*
Pour l'échantillon donné, f = 0,63 et n = 400 donc t = 1,24 on ne peut donc refuser H_0. On conclura donc au seuil de 5% que la fréquence de réussite de l'échantillon est conforme à celle de la population (la différence est due à la fluctuation d'échantillonnage).

Remarque :

Dans le cas où l'on veut tester si l'augmentation de la fréquence de reçus est significative, on effectue un test unilatéral.
H_1 devient : $p > p_0$
Pour le seuil α de 0,05 on lit sur la table N(0,1) la valeur critique : u=1,65 (lu à 0,95)

On obtient : $0,95 = P\left(\frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < 1,65 \right)$

La règle de décision est alors :

- *Si t> 1,65 on refuse H_0.*
- *Si t<1,65 on accepte H_0, ce qui est le cas pour notre échantillon.*

Exemple 2 :
Test d'adéquation à une loi, voir page 74

Bibliographie

Nous avons utilisé la bibliographie suivante :

MURRAY.R. SPIEGEL, 1983, *Probabilités et Statistique.*
Série Schaum.

M BERTAUD,B CHARLES, 1989, *Initiation à la statistique et aux probabilités*
Editions Eyroles Paris.

G. SAPORTA, 1990, *Probabilités, Analyse des données et Statistique*.
Edition Technip.

Paul G HOEL, 1991, S*tatistique mathématique.*
Edition Armand Collin.

P. DAGNELIE, 1992, *Statistique Théorique et Appliquée*, Tome 1.
Edition Les Presses Agronomiques de Grembloux.

Bernard PY, 1994, *Statistique descriptive.*
Edition Economica.

Pascal KAUFFMAN, 1994, *Statistique information, estimation test.*
Edition Dunod

J J DAUDIN, S ROBIN, C VUILLET, 2001,*Statistiques inférentielles, idées, démarche.*
Edition Presse universitaire de Rennes.

Brochure C I I Paris Nord., 2000.
SIMULATION et STATISTIQUE en SECONDE, IREM Paris Nord.