S6 STATISTIQUES DESCRIPTIVES A DEUX DIMENSIONS

Cadre : on dispose désormais d'une série statistique double de n couples de valeurs $(u_i ; v_i)$

on considère 2 caractères (2 variables) X et Y

ex. poids — X $\quad u_1 \quad u_2 \cdots u_k \cdots u_n$

taille — Y $\quad v_1 \quad v_2 \cdots v_k \cdots v_n$

même valeur (effectif de 2, au moins pour la valeur en question)

Expression de la série par effectifs

La série (1D) $(u_1, \ldots, u_n)$ atteint p valeurs différentes notées $(x_1, \ldots, x_p)$ $(p \leq n)$

La série (1D) $(v_1 \cdots v_n)$ atteint q valeurs différentes notées $(y_1 \cdots y_q)$ $(q \leq n)$

On peut déterminer les effectifs $n_{i,j}$ $i \in [1..p]$, $j \in [1..q]$ des couples d'observations (x_i, y_j)

On peut, si on le souhaite, regrouper les valeurs en classes et alors $n_{i,j}$ désigne l'effectif du couple de valeurs de la série double tq la première valeur est contenue dans la classe x_i et la seconde valeur est contenue dans la classe y_j

Tableau d'effectifs

$x_i \backslash y_j$	y_1	y_2	...	y_q	Total
x_1	$n_{1,1}$	$n_{1,2}$	...	$n_{1,q}$	$n_{1\bullet}$
x_2	$n_{2,1}$	$n_{2,2}$	...	$n_{2,q}$	$n_{2\bullet}$
$\vdots$					
x_p	$n_{p,1}$	$n_{p,2}$	...	$n_{p,q}$	$n_{p\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$		$n_{\bullet q}$	n

$$n_{i\bullet} = \sum_{j=1}^{q} n_{ij} \qquad i \in [1,p]$$

$$n_{\bullet j} = \sum_{i=1}^{p} n_{ij} \qquad j \in [1..q]$$

$$\sum_{i=1}^{p} n_{i\bullet} = \sum_{j=1}^{q} n_{\bullet j} = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij} = n$$

Exemple avec $n=12$ couples (Poids, Taille)

3 classes de Taille

7 classes de Poids

Poids \ Taille	]160;170]	]170;180]	]180;190]	Totaux
]55;60]		1		1
]60;65]	1	1	1	3
]65;70]		2		2
]70;75]			1	1
]75;80]				0
]80;85]	1	2	1	4
]85;90]			1	1
Totaux	2	6	4	12

En fréquences relatives (approximation / estimation de probabilités)

Poids \ Taille	]160;170]	]170;180]	]180;190]	Totaux

Poids	[illegible]	]…;180]	[illegible]	Totaux
]55;60]		1/12		1/12
]60;65]	1/12	1/12	1/12	3/12
]65;70]		2/12		2/12
]70;75]			1/12	1/12
]75;80]				0/12
]80;85]	1/12	2/12	1/12	4/12
]85;90]			1/12	1
Totaux	2/12	6/12	4/12	1

$$\mathbb{P}\left[(\text{Poids}, \text{Taille}) \in \left( ]80;85] \times]170;180] \right)\right] =$$

$$\mathbb{P}[\underbrace{(X,Y)}_{\text{couple de v.a.r (discrète)}} = \underbrace{(x_i, y_i)}_{\text{réalisation possible pour le couple}}] = \underbrace{p_{ij}}_{\text{loi du couple}} \approx \underbrace{\frac{n_{ij}}{n}}_{\text{estimation empirique } \hat{p}_{ij}} = \underbrace{f_{ij}}_{\text{fréq. relative}}$$

$n_{ij} = 2$, $n = 12$

STATS à partir d'une série

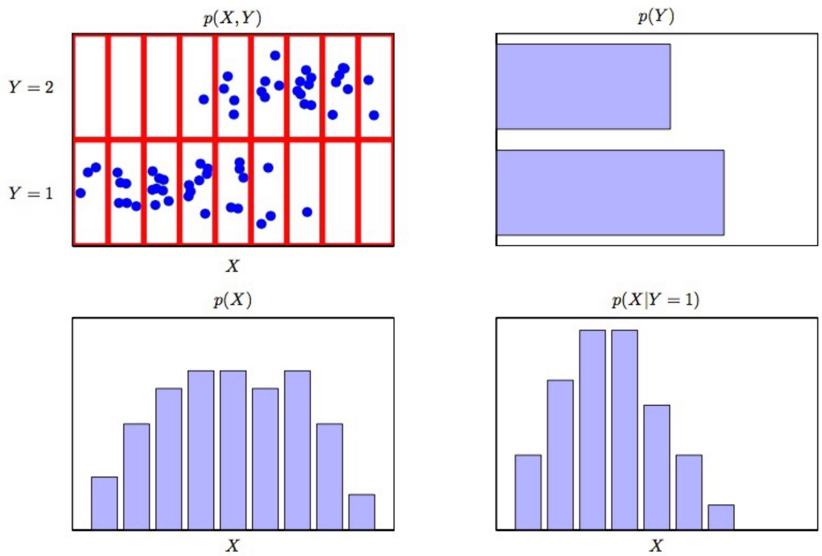
PAS

PROBA…

(réalisation d'un n-échantillon)

$\theta = \mu_{1,2}, \sigma_{1,2}$

x_i, y_j



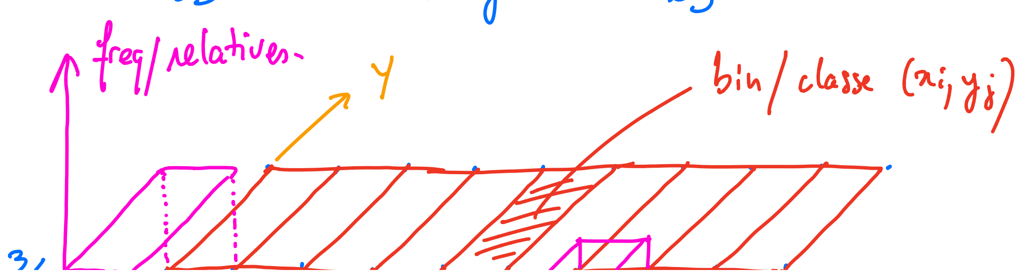
Idée nouvelle

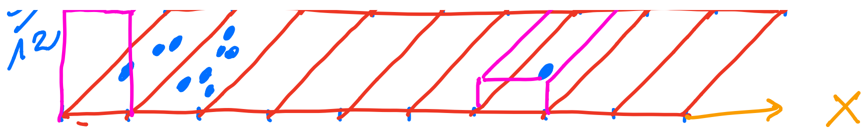
loi conditionnelle ex. $\mathbb{P}[X \mid Y=1]$
déduite des fréquences conditionnelles relatives

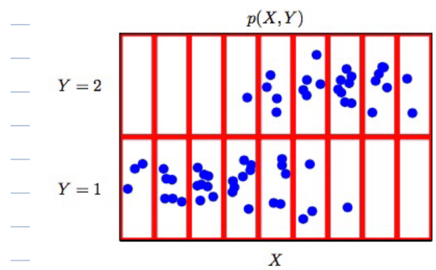
$$f_{i|j} = \frac{n_{ij}}{n_{\bullet j}}$$

$$f_{j|i} = \frac{n_{ij}}{n_{\bullet i}}$$

Représentation 3D de l'histogramme 2D







(S7) Notion de dépendance entre deux caractères, (dépendance entre deux v.a.r.)

Idée intuitive

Lorsque les observations (2D) présentes dans une série statistique double ont tendance
- à s'aligner
- à se structurer le long d'une droite (linéaire)

alors on dit qu'il y a une dépendance linéaire entre les caractères/les variables

on parle également d'une corrélation linéaire