Université Paul Valéry
Arts - Lettres - Langues
Sciences humaines & sociales

Cours 2
Distribution conjointe

Distribution conjointe

1 **Rappel : distribution d'une variable** ; soit X une variable observée sur un échantillon E ; l'effectif noté n_i de la modalité m_i [dans E] est le nombre d'individus de E dont la mesure par X est m_i ; on vérifie sans peine que la somme des effectifs vaut n ; la fréquence f_i de m_i [dans E] est le rapport de n_i à n si on l'exprime en proportion ($f_i = \frac{n_i}{n}$), multiplié par 100 si on l'exprime en pourcentage ($f_i = 100 * \frac{n_i}{n}$) ; ce pourcentage indique l'effectif (théorique) qu'on devrait observer dans un échantillon de taille 100 pour que le caractère soit présent dans les mêmes proportions ; on vérifie sans peine que la somme des f_i vaut 1 ou 100 selon que la fréquence est exprimée en proportion ou pourcentage.

La distribution de la variable X [sur E] est la liste des modalités m_i affectée de leur effectif n_i ou à sa fréquence f_i dans l'échantillon ; dans le premier cas la distribution est dite en effectif, et dans le second en fréquence, en proportion ou en pourcentage ; cette liste est généralement présentée sous la forme d'un tableau à deux lignes ou deux colonnes :

modalités	m_1	m_2	...	m_i	...	m_k
effectif ou fréquence	n_1 *ou* f_1	n_2 *ou* f_2	...	n_i *ou* f_i	...	n_k *ou* f_k

modalités	effectif ou fréquence
m_1	n_1 *ou* f_1
m_2	n_2 *ou* f_2
...	...
m_i	n_i *ou* f_i
...	
m_k	n_k *ou* f_k

2 déf **Modalité conjointe.** Lorsqu'on mesure simultanément X et Y sur un individu e de l'échantillon, la mesure de X sur e, notée x_e, est une modalité m_i de X ; de même la mesure de Y sur e, notée y_e, est une modalité m'_j de Y ; le résultat est le couple (x_e,y_e) égal au couple de modalités (m_i,m'_j) ; les différents couples de modalités, formés d'une modalité de X et d'une modalité de Y sont au nombre de k*p et s'appellent les modalités conjointes [de X et Y].

3 déf **Distribution conjointe de X et Y :** c'est la liste des k*p modalités conjointes (m_i,m'_j) associées chacune à son effectif n_{ij} ou à sa fréquence f_{ij} ; comme les modalités sont des couples de valeurs, cette liste est généralement présentée sous la forme d'un tableau à k lignes identifiant les modalités de X et p colonnes identifiant les modalités de , dont chaque cellule représente une modalité conjointe ; pour construire la distribution conjointe, il suffit de remplir chaque cellule c_{ij} de son effectif n_{ij} (le nombre d'individus ayant la mesure m_i pour X et la mesure m'_j pour Y), ou de sa fréquente f_{ij} :

X/Y	m'_1	m'_2	...	m'_j	...	m'_p
m_1	n_{11} ou f_{11}	n_{12} ou f_{12}	...	n_{1j} ou f_{1j}	...	n_{1p} ou f_{1p}
m_2	n_{21} ou f_{21}	n_{22} ou f_{22}	...	n_{2j} ou f_{2j}	...	n_{2p} ou f_{2p}
...	...	...	...	...	...	...
m_i	n_{i1} ou f_{i1}	n_{i2} ou f_{i2}	...	n_{ij} ou f_{ij}	...	n_{ip} ou f_{ip}
...	...	...	...	...	...	...
m_k	n_{k1} ou f_{k1}	n_{k2} ou f_{k2}	...	n_{kj} ou f_{kj}	...	n_{kp} ou f_{kp}

Distribution conjointe de X et Y sous forme de tableau de contingence.

Voilà, par exemple, les distributions en effectif et en proportion de l'exemple 1, « niveau scolaire et absentéisme » :

X / Y	Rare	Moyen	Fréquent
A	7	4	4
B	8	2	2

X / Y	Rare	Moyen	Fréquent	Total X
A	7/27=0,26	0,15	0,15	0,56
B	0,3	0,07	0,07	0,44
Total Y	0,56	0,22	0,22	1

4 ☞ **Construction.** À partir des données brutes et du modèle statistique de la situation, on construit la distribution conjointe de la manière suivante :

1. on construit le tableau des modalités conjointes à k lignes et p colonnes ;
2. dans chaque cellule c_{ij} on place son effectif n_{ij} en comptant le nombre d'individus de l'échantillon ayant simultanément m_i et m'_j pour X et Y, ou la fréquence n_{ij} correspondante.

Exemple de la taille des pères et fils : on choisit arbitrairement comme modalités pour X et Y les 4 intervalles [62 ; 65[, [65 ; 68[, [68 ; 71[et [71 ; 74] ; on trouve 0 individu pour la modalité conjointe (m_1, m'_1) $(n_{11} = 0)$, 3 individus (2, 4 et 6) pour la modalité (m_1, m'_2) $(n_{12} = 3)$, etc. ; on obtient finalement la distribution conjointe :

X/Y	[62 - 65[	[65 - 68[	[68 - 71[	[71 - 74]
[62 - 65[	0	3	0	0
[65 - 68[	0	2	1	1
[68 - 71[	0	0	4	0
[71 - 74]	0	0	1	0

Distributions marginales

5 déf **La distribution marginale de X** est la distribution de X sur l'échantillon (la liste des modalités de X associées à leur effectif ou fréquence dans l'échantillon) ; le nom vient de ce qu'elles sont souvent présentées en marge du tableau de contingence, parallèlement à la liste des modalités. De même pour la distribution marginale de Y.

6 déf **L'effectif marginal de la modalité m_i de X** est le nombre des individus de E dont la mesure par X est m_i ; ces individus sont ceux qui contribuent aux effectifs de la ième ligne du tableau de contingence en effectif ; ce nombre est donc la somme des p effectifs de la ième ligne, $n_{i1}, n_{i2}, ..., n_{ip}$; on le note $n_{i.}$: d'une manière générale, un point comme indice signifie un parcours additif des lignes quand il est en première position (comme dans $n_{.}3$ qui indique un parcours additif des lignes de la troisième colonne), des colonnes quand il est en seconde position (comme dans $n_{2.}$ qui indique un parcours additif des colonnes de la deuxième ligne) ; pour évaluer $n_{i.}$ on

se place ainsi au début de la ligne i et on parcourt les colonnes $c_{i1}, c_{i2}, ..., c_{ij}, ..., c_{ip}$ en cumulant leur contenu $n_{i1}, n_{i2}, ..., n_{ij}, ..., n_{ip}$; ce qu'on résume par la formule :

$$n_{i.} = \sum_{j=1}^{p} n_{ij}$$

Dans une formule de cette nature, la lettre $\sum$ indique un parcours cumulatif, n_{ij} les éléments visitables, $j = 1,p$ la propriété des éléments à visiter (ici, on fait varier j de 1 à p pour passer par les p colonnes, sans faire varier i pour rester sur la ième ligne).

7 déf **La fréquence marginale de la modalité m_i de X** est notée $f_i.$; par définition $\frac{n_{i.}}{n}$.

8 **L'effectif marginal de la modalité m'_j de Y** se calcule de manière similaire : $n.j$ est la somme des k effectifs situés sur la jème colonne, $n_{1j}, n_{2j}, ..., n_{kj}$:

$$n_{.j} = \sum_{i=1}^{k} n_{ij}$$

9 **Le tableau de contingence avec marges** permet de représenter simultanément la distribution conjointe et les deux distributions marginales :

X/Y	m'_1	m'_2	...	m'_j	...	m'_p	$MargeX$
m_1	n_{11} ou f_{11}	n_{12} ou f_{12}		n_{1j} ou f_{1j}		n_{1p} ou f_{1p}	$n_{1.}$ ou $f_{1.}$
m_2	n_{21} ou f_{21}	n_{22} ou f_{22}		n_{2j} ou f_{2j}		n_{2p} ou f_{2p}	$n_{2.}$ ou $f_{2.}$
...							
m_i	n_{i1} ou f_{i1}	n_{i2} ou f_{i2}		n_{ij} ou f_{ij}		n_{ip} ou f_{ip}	$n_{i.}$ ou $f_{i.}$
...							
m_k	n_{k1} ou f_{k1}	n_{k2} ou f_{k2}		n_{kj} ou f_{kj}		n_{kp} ou f_{kp}	$n_{k.}$ ou $f_{k.}$
$MargeY$	$n_{.1}$ ou $f_{.1}$	$n_{.2}$ ou $f_{.2}$		$n_{.j}$ ou $f_{.j}$		$n_{.p}$ ou $f_{.p}$	$n_{..}$ ou $f_{..}$

Ce tableau doit être visuellement décomposé en cinq parties :
- Première colonne : modalités de X
- Dernière colonne : distribution [marginale] de X
- Première ligne : modalités de Y
- Dernière ligne : distribution [marginale] de Y
- L'intérieur : distribution conjointe de X et Y

Il faut remarquer que les sommes des trois distributions sont égales à la taille de l'échantillon.

10 **Exemple : niveau scolaire et absentéisme.** X ayant deux modalités, A et B, construire la distribution marginale de X revient à calculer leur effectif $n_1.$ et $n_2.$; les élèves de niveau A se composent des élèves de niveau A étant peu absents (il y en a 7, n_{11}), des élèves de niveau A étant moyennement absents (il y en a 4, n_{12}) et des élèves de niveau A étant souvent absents (il y en a 4, n_{13}) : au total $n_{1.} = 15$; de la même façon on trouve $n_{2.} = 12$; naturellement, la somme de ces deux effectifs est égale à 27, la taille de l'échantillon.

X / Y	Rare	Moyen	Fréquent	Total X
A	7	4	4	15
B	8	2	2	12
Total Y	15	6	6	27

Voilà le tableau de contingence en proportion avec ses marges :

X / Y	Rare	Moyen	Fréquent	Total X
A	7/27=0,26	0,15	0,15	0,56
B	0,3	0,07	0,07	0,44
Total Y	0,56	0,22	0,22	1

Questions de cours

1. Qu'appelle-t-on modalité conjointe ?
2. Définition d'une distribution conjointe ?
3. Définition d'une distribution marginale ?
4. Que désignent les notations x_1 y_3 k p n m_2 m'_3 c_{44} n_{12} f_{32} $n_2.$ $n_.3$ $f_4.$ $f_.1$ $n_{..}$ dans le modèle d'une situation statistique ?
5. X et Y ont respectivement 4 et 5 modalités : combien y-a-t-il de modalités conjointes ? de distributions marginales ? d'effectifs marginaux ?
6. Quelle est la valeur de $f..$ et $n..$?

Questions sur le cours

1. Montrer que le nombre de modalités conjointes est effectivement égal à $k * p$
2. À quoi est égal la somme des effectifs marginaux des modalités de X ? de Y ?
3. À partir du tableau suivant vérifier les formules suivantes :
4.

X \ Y	m'_1	m'_2	m'_3	m'_4	m'_5
m_1	1	3	6	1	3
m_2	2	4	0	5	1
m_3	3	2	3	1	4
m_4	4	5	3	5	4
m_5	5	2	5	0	1
m_6	6	4	1	2	2

a) $\sum_{i=1}^{3} n_{i5} = 8$;
b) $\sum_{i=1,3} n_{i4} = 7$;
c) $\sum_{j=2}^{4} n_{5j} = 7$;
d) $\sum_{j=1,5} n_{4j} = n_{4.} = 21$;
e) $\sum_{i=4,5} n_{i3}^2 = 34$;
f) $\sum_{i=1,3}(n_{i5} - 2) = 2$;
g) $\sum_{i=2,5}(n_{i2} + n_{i4}) = 22$;
h) $\sum_{j=1,4}(n_{6j} - n_{1j}) = 2$;
i) $\sum_{i=2,5}(n_{i(i-1)} - n_{4i}) = 0$;
j) $\sum_{i=2,4;j=1,4} n_{ij} = 37$;
k) $\sum_{i=4,6;j<4} n_{ij} = 35$;
l) $\sum_{i=3,6;j>=i} n_{ij} = 18$;