3.1.3 Distributions conditionnelles

Les distributions conditionnelles s'obtiennent en fixant la valeur d'une des deux variables (où la modalité d'une des deux variables).

Exemple 3.1.2

On considère le tableau suivant, relatif à une population de 100 ménages, tel que : $X =$ « le nombre d'enfants du ménage » et $Y =$ « le nombre de pièces du logement »

$X \backslash Y$	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	*Total*
$x_1 = 2$	15	10	05	30
$x_2 = 3$	30	5	10	45
$x_3 = 4$	10	5	0	15
$x_4 = 5$	10	0	0	10
Total	65	20	15	100

1. La distribution conditionnelle de X sachant $Y = 3$ est donnée par la première colonne du tableau.
2. La distribution conditionnelle de X sachant $Y = 4$ est donnée par la deuxième colonne du tableau.
3. La distribution conditionnelle de Y sachant $X = 2$ est donnée par la première ligne du tableau.
4. De même, la distribution conditionnelle de Y sachant $X = 5$ est donnée par la quatrième ligne du tableau.

Ces quatre distributions se présentent dans les tableaux suivants :

Distribution conditionnelle de X sachant $Y = 3$

$X / Y = 3$	$n_{i/1}$
$x_1 = 2$	15
$x_2 = 3$	30
$x_3 = 4$	10
$x_4 = 5$	10
Total	65

Distribution conditionnelle de X sachant $Y = 4$

$X / Y = 4$	$n_{i/2}$
$x_1 = 2$	10
$x_2 = 3$	5
$x_3 = 4$	5
$x_4 = 5$	0
Total	20

Distribution conditionnelle de Y sachant $X=2$

$Y/X=2$	$n_{j/1}$
$y_1=3$	15
$y_2=4$	10
$y_3=5$	05
Total	30

et

Distribution conditionnelle de Y sachant $X=5$

$Y/X=5$	$n_{j/4}$
$y_1=3$	10
$y_2=4$	0
$y_3=5$	0
Total	10

Remarque 3.1.2 $n_{i/1}=n_{i1}$; $n_{i/2}=n_{i2}$; $n_{j/1}=n_{1j}$; $n_{j/4}=n_{4j}$.

En général si on prend le tableau des contingents des effectifs suivant :

Distribution conjointe en effectif de X et Y

$X \backslash Y$	y_1	y_2	...	y_j	...	y_k	Total
x_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2.}$
$\vdots$							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i.}$
$\vdots$							
x_r	n_{r1}	n_{r2}		n_{rj}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.k}$	N

$\Rightarrow$

Distribution conditionnelle de X sachant $Y=y_j$

$X/Y=y_j$	$n_{i/j}$
x_1	n_{1j}
x_2	n_{2j}
$\vdots$	
x_i	n_{ij}
$\vdots$	
x_r	n_{rj}
Total	$n_{.j}$

et

Distribution conditionnelle de Y sachant $X=x_i$

$Y/X=x_i$	$n_{j/i}$
y_1	n_{i1}
y_2	n_{i2}
$\vdots$	
y_j	n_{ij}
$\vdots$	
y_k	n_{ik}
Total	$n_{i.}$

Remarque 3.1.3 La distribution conditionnelle de chacune des variables X et Y peut être définie à partir des fréquences.

1. Dans le cas de la distribution conditionnelle de X sachant $Y=y_j$, on a :

$$f_{i/j}=\frac{n_{ij}}{n_{.j}}=\frac{n_{ij}\times N}{n_{.j}\times N}=\frac{f_{ij}}{f_{.j}}; \quad \text{avec } \sum_{i=1}^{r} f_{i/j}=1$$

2. Dans le cas de la distribution conditionnelle de Y sachant $X = x_i$, on a :

$$f_{j/i} = \frac{n_{ij}}{n_{i.}} = \frac{n_{ij} \times N}{n_{i.} \times N} = \frac{f_{ij}}{f_{i.}}; \text{ avec } \sum_{j=1}^{k} f_{j/i} = 1$$

Exemple 3.1.3

On considère le tableau suivant, relatif à une population de 100 ménages, tel que :
X = « le nombre d'enfants du ménage » et Y = « le nombre de pièces du logement ».
Pour calculer les **distributions conditionnelles en fréquences** de X sachant $Y = 4$ et de Y sachant $X = 2$, on a deux façons pour le faire. En effet ;

$X \backslash Y$	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	*Total*
$x_1 = 2$	15	10	05	30
$x_2 = 3$	30	5	10	45
$x_3 = 4$	10	5	0	15
$x_4 = 5$	10	0	0	10
Total	65	20	15	100

1ere méthode : En passant par les distributions conditionnelles des effectifs.

Distribution conditionnelle de X sachant Y = 4

$X / Y = 4$	$n_{i/2}$
$x_1 = 2$	10
$x_2 = 3$	5
$x_3 = 4$	5
$x_4 = 5$	0
Total	20

Distribution conditionnelle de Y sachant X = 2

$Y / X = 2$	$n_{j/1}$
$y_1 = 3$	15
$y_2 = 4$	10
$y_3 = 5$	05
Total	30

$\Rightarrow$

Distribution conditionnelle de X sachant Y = 4

$X / Y = 4$	$f_{i/2}$
$x_1 = 2$	0,5
$x_2 = 3$	0,25
$x_3 = 4$	0,25
$x_4 = 5$	0
Total	1,00

Distribution conditionnelle de Y sachant X = 2

$Y / X = 2$	$f_{j/1}$
$y_1 = 3$	0,5
$y_2 = 4$	0,33
$y_3 = 5$	0,17
Total	1,00

2^eme^ méthode : En passant par la distribution conjointe en fréquences.

Donc

Distribution conjointe en fréquences

$X \backslash Y$	$y_1=3$	$y_2=4$	$y_3=5$	*Total*
$x_1=2$	0,15	0,10	0,05	0,30
$x_2=3$	0,30	0,05	0,10	0,45
$x_3=4$	0,10	0,05	0	0,15
$x_4=5$	0,10	0	0	0,10
Total	0,65	0,20	0,15	1,00

En utilisant les formules $f_{i/j} = \dfrac{f_{ij}}{f_{.j}}$ et $f_{j/i} = \dfrac{f_{ij}}{f_{i.}}$ on obtient

Distribution conditionnelle de X sachant $Y=4$

$X/Y=4$	$f_{i/2}$
$x_1=2$	0,5
$x_2=3$	0,25
$x_3=4$	0,25
$x_4=5$	0
Total	1,00

Distribution conditionnelle de Y sachant $X=2$

$Y/X=2$	$f_{j/1}$
$y_1=3$	0,5
$y_2=4$	0,33
$y_3=5$	0,17
Total	1,00

Remarque 3.1.4 $f_{i/1} \neq f_{i1}$; $f_{i/2} \neq f_{i2}$; $f_{j/1} \neq f_{1j}$; $f_{j/4} \neq f_{4j}$.

3.1.4 Variables indépendantes

Les variables X et Y sont indépendantes si et seulement si

$$\forall i, j \Rightarrow f_{ij} = f_{i.} \times f_{.j}$$

Remarque 3.1.5 On a l'équivalence suivante :

$$\forall i, j : f_{ij} = f_{i.} \times f_{.j} \Leftrightarrow n_{ij} = \frac{n_{i.} \times n_{.j}}{N}$$

Exemple 3.1.4 On considère le tableau suivant, relatif à une population de 54 ménages, tel que :

$X \backslash Y$	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	*Total*
$x_1 = 2$	2	4	12	18
$x_2 = 3$	4	8	24	36
Total	6	12	36	54

Les deux variables X et Y sont indépendantes car :

1	$\frac{n_{1.} \times n_{.1}}{N} = \frac{18 \times 6}{54} = 2 = n_{11}$
2	$\frac{n_{1.} \times n_{.2}}{N} = \frac{18 \times 12}{54} = 4 = n_{12}$
3	$\frac{n_{1.} \times n_{.3}}{N} = \frac{18 \times 36}{54} = 12 = n_{13}$

4	$\frac{n_{2.} \times n_{.1}}{N} = \frac{36 \times 6}{54} = 4 = n_{21}$
5	$\frac{n_{2.} \times n_{.2}}{N} = \frac{36 \times 12}{54} = 8 = n_{22}$
6	$\frac{n_{2.} \times n_{.3}}{N} = \frac{36 \times 36}{54} = 24 = n_{23}$

3.2 Paramètres de liaison

Dans toute la suite on considère N observations sur les deux variables X et Y.

3.2.1 Covariance entre X et Y

La covariance est égale à la moyenne des écarts des couples les (x_i, y_i) de X et **Y** par rapport au point $(\bar{x}, \bar{y})$.

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

- <u>**Le rôle de la covariance**</u>

La covariance indique le sens de la relation entre les variables X et Y. Ainsi, On peut distinguer les cas suivants :

<u>1^er^ *Cas*</u> : Si $Cov(x, y) > 0$, alors on peut dire que **la relation entre les deux variables est positive**. Dans ce cas, **ces deux variables varient dans le même sens.**

<u>2^eme^ *Cas*</u> : Si $Cov(x, y) < 0$, alors on peut dire que **la relation entre les deux variables est négative**. Dans ce cas, **ces deux variables varient en sens inverse.**

<u>3^eme^ *Cas*</u> : Si $Cov(x, y) = 0$, alors on peut dire qu'**il n'y a pas de relation entre les deux variables.** Dans ce cas, **les variations de l'une n'entraînent pas la variation de l'autre.**