Modélisation statistique
STT-2902
Automne 2012

Semaine 8: Distribution conjointe de deux variables qualitatives

Emmanuelle Reny-Nolin
Département de mathématiques et de statistique
Université Laval

Semaine 8 : Distribution conjointe de deux variables qualitatives

Distribution théorique de deux variables discrètes ou qualitatives
- Loi conjointe de X et Y
- Loi marginale de X (resp. de Y)
- Loi conditionnelle de X sachant Y (resp. de $Y|X$)
- Indépendance de deux variables qualitatives

La fonction de masse conjointe de X et Y

X est une v.a. pouvant prendre les valeurs $\{x_1, x_2, ..., x_I\}$.

Y est une v.a. pouvant prendre les valeurs $\{y_1, y_2, ..., y_J\}$.

La *loi conjointe* de X et Y est l'ensemble des probabilités p_{ij} telles que

$$\boxed{p_{ij} = \qquad\qquad}$$

Puisque X et Y sont discrètes, il s'agit d'une *fonction de masse conjointe*, qu'on appelle aussi *distribution jointe*.

La fonction de masse conjointe de X et Y

On représente souvent une fonction de masse conjointe dans un tableau :

X \ Y	y_1	...	y_j	...	y_J	Total
x_1	p_{11}	...	p_{1j}	...	p_{1J}	$p_{1\bullet}$
...	...	...	...	...	...	...
x_i	p_{i1}	...	p_{ij}	...	p_{iJ}	$p_{i\bullet}$
...	...	...	...	...	...	...
x_I	p_{I1}	...	p_{Ij}	...	p_{IJ}	$p_{I\bullet}$
Total	$p_{\bullet 1}$	...	$p_{\bullet j}$	...	$p_{\bullet J}$	1

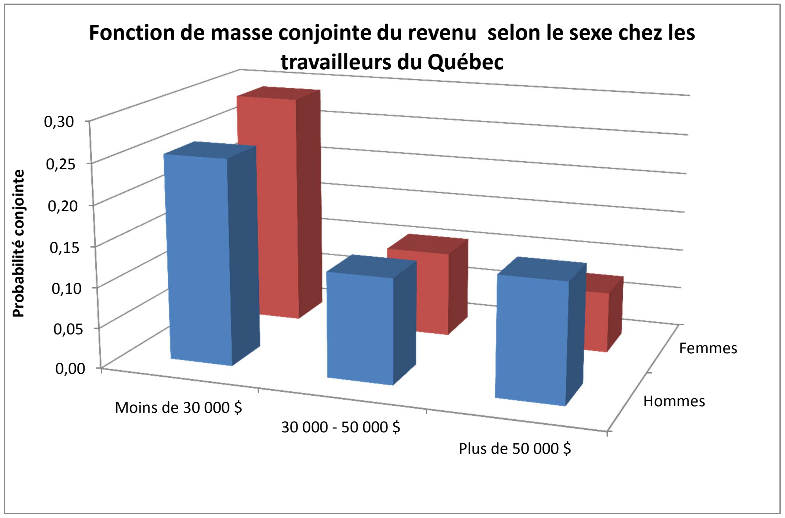
Exemple 1

Sur le site de l'ISQ, on trouve la distribution des travailleurs du Québec selon leur tranche de revenu et leur sexe. Voici la fonction de masse conjointe de ces deux variables discrètes :

Revenu \ Sexe	Hommes	Femmes	Total
Moins de 30 000 \$	0,25	0,29	0,54
30 000 - 50 000 \$	0,13	0,11	0,24
Plus de 50 000 \$	0,15	0,07	0,22
Total	0,53	0,47	1

Source : ISQ (2008), `http://www.stat.gouv.qc.ca/donstat/societe/famls_mengs_niv_vie/revenus_depense/revenus/repart_tranchep08_tab2.htm`

Représentation graphique en diagramme à bâtons



Exemple 2

Une urne contient deux boules rouges numérotées 1 et 2. Elle contient aussi trois boules bleues numérotées 1, 2 et 3.

Vous pigez une boule dans cette urne et vous vous intéressez aux deux variables suivantes :
X : la couleur de la boule tirée
Y : la parité du numéro tiré

Déterminez la distribution conjointe de X et Y.

La loi marginale de X

La fonction de masse *marginale* de X est l'ensemble des probabilités $p_{i\bullet}$ telles que

$$p_{i\bullet} =$$

X	$P(X = x_i)$
x_1	$p_{1\bullet}$
...	...
x_i	$p_{i\bullet}$
...	...
x_I	$p_{I\bullet}$
Total	1

Exemple 1

Déterminer la loi marginale du revenu des travailleurs du Québec.

La loi marginale de Y

La fonction de masse *marginale* de Y est l'ensemble des probabilités $p_{\bullet j}$ telles que

$$p_{\bullet j} =$$

Y	y_1	...	y_j	...	y_J	Total
$P(Y = y_j)$	$p_{\bullet 1}$	...	$p_{\bullet j}$	...	$p_{\bullet J}$	1

Exemple 1

Déterminer la loi marginale du sexe des travailleurs du Québec.

La fonction de masse conditionnelle de $X | Y = y_j$

La fonction de masse *conditionnelle* de X sachant que Y prend la valeur y_j est l'ensemble des probabilités $p_{i|j}$ telles que

$$\boxed{p_{i|j} = \qquad\qquad\qquad\qquad\qquad\qquad}$$

X \ Y	y_1	$\dots$	y_j	$\dots$	y_J
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1J}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$	p_{iJ}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_I	p_{I1}	$\dots$	p_{Ij}	$\dots$	p_{IJ}
Total	$p_{\bullet 1}$	$\dots$	$p_{\bullet j}$	$\dots$	$p_{\bullet J}$

Exemple 1

Déterminer la fonction de masse du revenu des hommes du Québec.

Diagramme en mosaïque

On peut représenter le tableau de fréquences par un diagramme rectangulaire où la surface de chaque cellule est proportionnelle à sa fréquence.

Diagramme en mosaïque (revenu conditionnel au sexe)

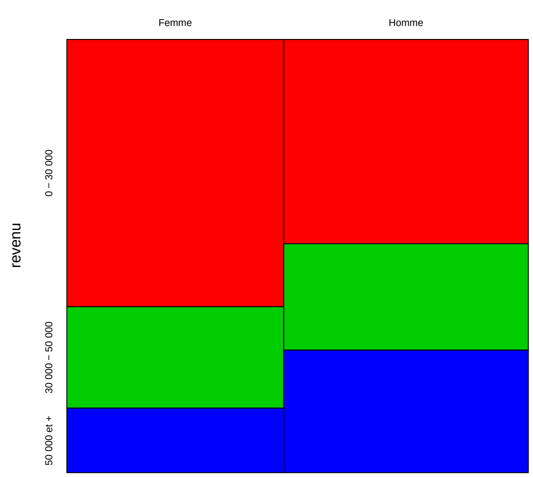
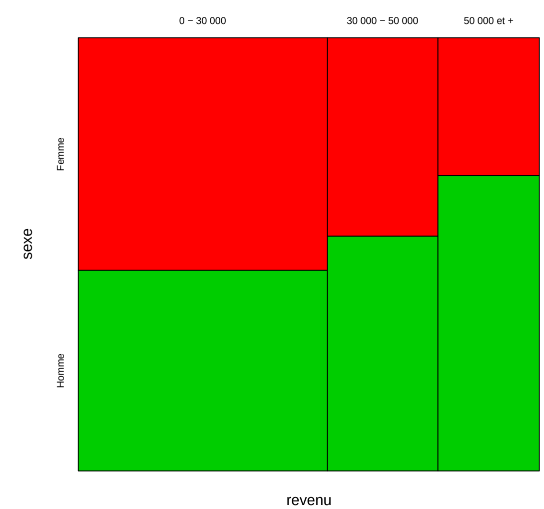


Diagramme en mosaïque

Diagramme en mosaïque (sexe conditionnel au revenu)



L'indépendance de deux variables qualitatives

X et Y sont indépendantes si et seulement si toutes les probabilités conjointes sont égales au produit des probabilités marginales :

$$\boxed{p_{ij} = \qquad\qquad}$$

Exemples 1 et 2

Le revenu est-il indépendant du sexe dans l'exemple 1 ?

La couleur est-elle indépendante de la parité dans l'exemple 2 ?

Si on avait eu 4 boules bleues au lieu de 3, aurait-on obtenu l'indépendance entre X et Y à l'exemple 2 ? Pourquoi ?

Conséquences de l'indépendance de X et Y

Si X est indépendant de Y, alors :

-
-
-

Diagramme en mosaïque sous l'indépendance