Corrigé - Série 4
Lois conjointes et tableaux de fréquences à double entrée

Exercice 1

a) Loi conjointe de X et Y :

X \ Y	2	3	4	Total
1	1/6	1/6	1/6	3/6
2	0	1/6	1/6	2/6
3	0	0	1/6	1/6
Total	1/6	2/6	3/6	1

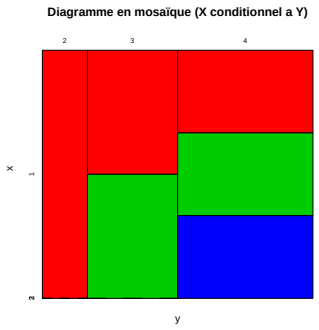
Loi marginale de X :

x_i	1	2	3	Total
$p_{i\bullet}$	3/6	2/6	1/6	1

Loi marginale de Y :

y_j	2	3	4	Total
$p_{\bullet j}$	1/6	2/6	3/6	1

b)



c) X et Y ne sont pas indépendantes, car il y a plusieurs cas où $p_{ij} \neq p_{i\bullet}p_{\bullet j}$.

d) Loi conditionnelle de Y lorsque le plus petit numéro tiré vaut 3 :

y_j	2	3	4	Total
$p_{j\|X=3}$	0	0	1	1

e) $E(Y|X=3) = 4$; $\sqrt{Var(Y|X=3)} = 0$.

f) Loi conditionnelle de X lorsque le plus grand numéro tiré est pair :

x_i	1	2	3	Total
$p_{i\|Y=2\,ou\,4}$	1/2	1/4	1/4	1

g) $Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{6}[1(2)+1(3)+1(4)+2(3)+2(4)+3(4)] - \left(\frac{3}{6}(1)+\frac{2}{6}(2)+\frac{1}{6}(3)\right)\left(\frac{1}{6}(2)+\frac{2}{6}(3)+\frac{3}{6}(4)\right) = \frac{35}{6} - \left(\frac{10}{6}\right)\left(\frac{20}{6}\right) = \frac{5}{18}$.

Exercice 2

a) Loi conjointe de X et Y :

X \ Y	-2	5	8	Total
1	0,21	0,35	0,14	0,7
2	0,09	0,15	0,06	0,3
Total	0,3	0,5	0,2	1

b) $P(X \text{ et } Y \text{ pairs}) = 0,09 + 0,06 = 0,15$.

c) $P(X = 1|Y = 5 \text{ ou } 8) = \frac{P(X = 1 \text{ et } Y = 5) + P(X = 1 \text{ et } Y = 8)}{P(Y = 5) + P(Y = 8)} = \frac{0,35 + 0,14}{0,5 + 0,2} = 0,7$

d) Il n'est pas nécessaire d'effectuer le calcul, car les variables sont indépendantes. Ainsi, leur covariance est nulle. Si vous avez besoin de vous convaincre :
$Cov(X,Y) = 4,55 - (1,3)(3,5) = 0$.

Exercice 3

$$\begin{aligned} X &= \text{nombre de cartes de pique } (\spadesuit) \text{ pigées} \\ Y &= \text{nombre de rois pigés} \end{aligned}$$

a) Loi conjointe de X et Y :

X ($\spadesuit$) \ Y (K)	0	1	2	Total
0	$\frac{\binom{36}{2}}{\binom{52}{2}}$	$\frac{\binom{3}{1}\binom{36}{1}}{\binom{52}{2}}$	$\frac{\binom{3}{2}}{\binom{52}{2}}$	$\frac{\binom{39}{2}}{\binom{52}{2}}$
1	$\frac{\binom{12}{1}\binom{36}{1}}{\binom{52}{2}}$	$\frac{\left[\binom{36}{1}+\binom{3}{1}\binom{12}{1}\right]}{\binom{52}{2}}$	$\frac{\binom{3}{1}}{\binom{52}{2}}$	$\frac{\binom{13}{1}\binom{39}{1}}{\binom{52}{2}}$
2	$\frac{\binom{12}{2}}{\binom{52}{2}}$	$\frac{\binom{12}{1}}{\binom{52}{2}}$	0	$\frac{\binom{13}{2}}{\binom{52}{2}}$
Total	$\frac{\binom{48}{2}}{\binom{52}{2}}$	$\frac{\binom{4}{1}\binom{48}{1}}{\binom{52}{2}}$	$\frac{\binom{4}{2}}{\binom{52}{2}}$	1

Loi conjointe de X et Y en version fractionnaire :

X ($\spadesuit$) \ Y (K)	0	1	2	Total
0	$\frac{630}{1326}$	$\frac{108}{1326}$	$\frac{3}{1326}$	$\frac{741}{1326}$
1	$\frac{432}{1326}$	$\frac{72}{1326}$	$\frac{3}{1326}$	$\frac{507}{1326}$
2	$\frac{66}{1326}$	$\frac{12}{1326}$	0	$\frac{78}{1326}$
Total	$\frac{1128}{1326}$	$\frac{192}{1326}$	$\frac{6}{1326}$	1

b) X et Y ne sont pas des variables indépendantes, car le produit des probabilités marginales n'est pas toujours égal à la probabilité conjointe correspondante.

Contre-exemple : $p_{22} = 0$, ce qui n'égale pas $p_{2\bullet}p_{\bullet 2} = \frac{6\times 78}{1326^2}$

c)

$$\begin{aligned} P(Y \geq 1 | X \geq 1) &= \frac{P(Y \geq 1 \cap X \geq 1)}{P(X \geq 1)} \\ &= \frac{(72+3+12+0)/1326}{(507+78)/1326} \\ &= 0,1487 \end{aligned}$$

d) • Vous payez 1$ pour chaque carte de pique pigée.

• Vous recevez 2$ pour chaque roi pigé.

Un jeu est équitable si l'espérance de gain est nulle. Pour calculer l'espérance du gain, on peut procéder de deux façons :

1) On détermine la valeur du gain pour chaque couple de valeurs (x_i, y_j), que l'on notera $g(x_i, y_j)$. On calcule l'espérance du gain comme suit :

$$\begin{aligned} E(Gain) &= \sum_{i=1}^{I} \sum_{j=1}^{J} g(x_i\, y_j)\; P(X = x_i \text{ et } Y = y_j) \\ &= 0 \left(\frac{630}{1326}\right) + 2 \left(\frac{108}{1326}\right) + \; ... = \; -0,19\$ \end{aligned}$$

2) On définit la variable Gain comme une combinaison linéaire des variables X et Y :

$$G \; = (-1)\, X + 2\, Y$$

On calcule l'espérance du gain comme suit :

$$\begin{aligned} E(G) &= (-1)\, E(X) + 2\, E(Y) \\ &= (-1) \left(\frac{663}{1326}\right) \; + \; 2 \left(\frac{204}{1326}\right) \\ &= -0,19 \; \$ \end{aligned}$$

Le jeu n'est donc pas équitable, car en moyenne, le joueur perd de l'argent.

Quel montant un roi devrait-il vous faire gagner pour le jeu devienne équitable ? Supposons qu'un roi vous donne k dollars. La valeur de k sera déterminée d'après l'équation :

$$E(G) \; = (-1)\, E(X) + k\, E(Y) \; = \; 0 \qquad \Rightarrow \qquad k = \frac{663}{204} = 3,25 \; \$$$

Exercice 4

Le tabac est-il plus associé aux décès par cancer du poumon ou aux décès par maladies coronariennes ?

On veut savoir si $P(Cancer|Fum)$ est supérieure ou inférieure à $P(Mal.coron.|Fum)$.

L'énoncé nous dit que

$$\frac{P(Cancer|Fum)}{P(Cancer|Non-Fum)} = 10$$

et que

$$\frac{P(Mal.coron.|Fum)}{P(Mal.coron.|Non-Fum)} = 1,7$$

On sait également que
$P(Cancer|Non-Fum) = 5/100\,000$ et que $P(Mal.coron.|Non-Fum) = 170/100\,000$.

Il suit que

$$P(Cancer|Fum) = 10 \times \frac{5}{100\,000} = \frac{50}{100\,000}$$

et que

$$P(Mal.coron.|Fum) = 1,7 \times \frac{170}{100\,000} = \frac{289}{100\,000}$$

Ainsi, puisque les maladies coronariennes sont beaucoup plus présentes dans la population que le cancer du poumon, il est normal qu'elles soient associées à plus de décès de fumeurs. Cette analyse ne permet toutefois pas de déterminer si le tabac a causé ces décès.

Exercice 5

a) Taux de mortalité des mères avant 1847 :

Médecins accoucheurs : $p_M = \frac{1\,989}{20\,024} = 0,098$

Sages-femmes : $p_{SF} = \frac{691}{17\,791} = 0,039$

b) Y a-t-il un lien statistique entre le type d'accoucheur et la survie ? (La différence entre les deux taux de mortalité est-elle significative ou fortuite ?)

On peut conduire un test d'indépendance et tester les hypothèses suivantes à l'aide de la distribution du khi-carré.

H_0 : La survie et le métier de l'accompagnant sont indépendants

H_1 : Il existe une relation entre la survie et le métier de l'accompagnant

On calcule les fréquences espérées, puis la distance observée entre le modèle d'indépendance (le tableau des fréquences espérées) et les observations.

Fréq. obs. O_{ij}	Survie	Décès	Total
Médecins	18 215	1 989	20 204
Sages-femmes	17 100	691	17 791
Total	35 315	2 680	37 995

Fréq. esp. E_{ij}	Survie	Décès	Total
Médecins	18 778,9	1 425,1	20 204
Sages-femmes	16 536,1	1 254,9	17 791
Total	35 315	2 680	37 995

Valeur observée de la statistique du test : $D_{obs} = 512,68$.

Puisque notre tableau de fréquences a les dimensions 2×2, l'espérance de la distance sous H_0 est $(2-1) \times (2-1) = 1$. La valeur observée est beaucoup plus grande que n'importe quelle valeur critique, et le seuil observé ($P(D > 512,68)$ sous H_0) est presque 0.

Le lien est très clair : le taux de mortalité est plus élevé chez les médecins.

c) Taux de mortalité des mères après 1847 :

Médecins accoucheurs : $p_M = \dfrac{1\,712}{47\,938} = 0,036$

Sages-femmes : $p_{SF} = \dfrac{1\,248}{40\,770} = 0,031$

(Entre vous et moi, c'est encore très élevé, dans les deux cas !)

d) La différence entre les deux taux de mortalité est-elle encore significative ? On fait le test d'indépendance de la même façon qu'en b).

Valeur observée de la statistique du test : $D_{obs} = 17,78$.

Le seuil observé est $P(D > 17,78) = 0,00002478$.

Le lien est encore significatif : le taux de mortalité est plus élevé chez les médecins, mais la différence est moins grande que précédemment. Les médecins n'ont pas tous

accepté de changer instantanément leur pratique : cela les aurait obligés à admettre qu'ils étaient responsables de tant de morts...

Exercice 6

On veut savoir si la distribution de la variable d'intérêt (ici : années vécues après le décès) est la même pour toutes les populations considérées (ici les hommes et les femmes).

• Puisque lorsque deux variables sont indépendantes leurs lois conditionnelles sont toutes égales à leur loi marginale, il est équivalent de dire "Les $I = 2$ distributions conditionnelles sont les mêmes" et "La variable d'intérêt (durée de vie) et la variable qui distingue les populations (sexe) sont indépendantes". Dans notre exemple, la question revient à se demander s'il existe un lien entre les variables "sexe" et "durée de vie", et la statistique du test d'indépendance nous permet de répondre à la question.

• Voici le tableau des fréquences observées et espérées :

Population	Années vécues après le décès < 5 ans	5 à 10 ans	> 10 ans	Total
Hommes (veufs)	25	42	33	100
	$29,3$	$41,3$	$29,3$	
Femmes (veuves)	19	20	11	50
	$14,7$	$20,7$	$14,7$	
Total	44	62	44	150

• Calcul de la valeur observée de la statistique du test :

$D_{obs} = 0.64 + 0.01 + 0.46 + 1.28 + 0.02 + 0.92 = 3,328$

• Décision et conclusion :

Puisque $\chi^2_{2,0.05} = 5,99$, on ne rejette pas H_0 au seuil de 5%, et on conclut que la distribution de la durée de vie ne diffère pas selon le sexe.

On peut aussi calculer le seuil observé du test : $P(D > 3,328) = 0.1894$ (où $D \sim \chi^2_2$ sous H_0), ce qui indique qu'on ne rejetterait pas l'indépendance même en utilisant un seuil de 10% ou 15%.

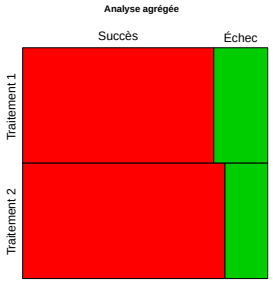
Exercice 7

a) Construire les tableaux de fréquences associés à ces trois situations.

Analyse agrégée

	Succès	Échec	Total
Traitement 1	273	77	350
Traitement 2	289	61	350
Total	562	138	700

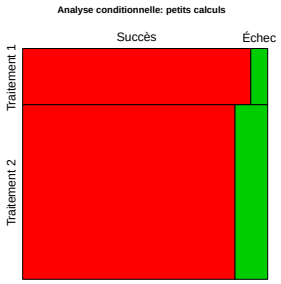
Valeur-P = 0,1285



Analyse conditionnelle, calculs < 2 cm

	Succès	Échec	Total
Traitement 1	81	6	87
Traitement 2	234	36	270
Total	315	42	357

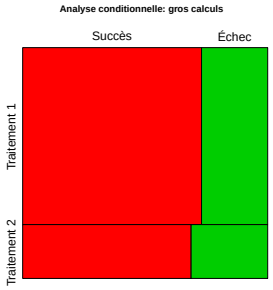
Valeur-P = 0,1051



Analyse conditionnelle, calculs > 2 cm

	Succès	Échec	Total
Traitement 1	192	71	263
Traitement 2	55	25	80
Total	247	96	343

Valeur-P = 0,4580



b) Quand on considère les résultats des traitements sans tenir compte de la taille des calculs rénaux, on conclut que le traitement 2 a un plus grand taux de succès que le traitement 1. (Cette différence est non significative statistiquement).

Quand on considère les résultats des traitements en tenant compte de la taille des calculs rénaux, i.e. en faisant l'analyse séparément pour les petites pierres et les grosses pierres, on conclut l'inverse. (Encore non significatif).

c) Il s'agit d'une réalisation du paradoxe de Simpson. Une troisième variable influence la relation entre le traitement et le succès : la taille des calculs (et il faut en tenir compte). Cette apparente contradiction est due au fait que peu de calculs inférieurs à 2 cm sont traités avec les chirurgies ouvertes (qui ont beaucoup de succès), et que beaucoup de petits calculs sont traités par chirurgie percutanée (qui semble avoir moins de succès).

Bien sûr, dans le choix d'un traitement, il faut aussi tenir compte des risques collatéraux (anesthésie générale, grande incision, etc.), mais c'est une autre histoire...