Exercice #1

$X_i \sim Poisson(\lambda_i) \qquad\qquad i = 1, \ldots, k$

$$P\left(X_1 = x \,\middle|\, \sum_{i=1}^{k} X_i = n\right) = \frac{P\left(X_1 = x \bigcap \sum_{i=1}^{k} X_i = n\right)}{P\left(\sum_{i=1}^{k} X_i = n\right)} = \frac{P(X_1 = x)P\left(\sum_{i=2}^{k} X_i = n - x\right)}{P\left(\sum_{i=1}^{k} X_i = n\right)}$$

$$\sum_{i=1}^{k} X_i \sim Poisson\left(\sum_{i=1}^{k} \lambda_i\right)$$
$$\sum_{i=2}^{k} X_i \sim Poisson\left(\sum_{i=2}^{k} \lambda_i\right)$$

$$\begin{aligned}
P\left(X_1 = x \,\middle|\, \sum_{i=1}^{k} X_i = n\right) &= \frac{e^{-\lambda_1}\lambda_1^x}{x!} \times \frac{e^{-\sum_{i=2}^{k}\lambda_i}\left(\sum_{i=2}^{k} \lambda_i\right)^{n-x}}{(n-x)!} \times \frac{n!}{e^{-\sum_{i=1}^{k}\lambda_i}\left(\sum_{i=1}^{k} \lambda_i\right)^{n}} \\
&= \frac{\binom{n}{x}\lambda_1^x\left(\sum_{i=2}^{k} \lambda_i\right)^{n-x}}{\left(\sum_{i=1}^{k} \lambda_i\right)^{n}} \\
&= \binom{n}{x}\left(\frac{\lambda_1}{\sum_{i=2}^{k}\lambda_i}\right)^x \left(\frac{\sum_{i=2}^{k}\lambda_i}{\sum_{i=1}^{k}\lambda_i}\right)^n \\
&= \binom{n}{x}\left(\frac{\lambda_1}{\sum_{i=2}^{k}\lambda_i} \times \frac{\sum_{i=2}^{k}\lambda_i}{\sum_{i=1}^{k}\lambda_i}\right)^x \left(\frac{\sum_{i=2}^{k}\lambda_i}{\sum_{i=1}^{k}\lambda_i}\right)^{n-x} \\
&= \binom{n}{x}\left(\frac{\lambda_1}{\sum_{i=1}^{k}\lambda_i}\right)^x \left(\frac{\sum_{i=2}^{k}\lambda_i}{\sum_{i=1}^{k}\lambda_i}\right)^{n-x}
\end{aligned}$$

$$\frac{\lambda_1}{\sum_{i=1}^{k}\lambda_i} + \frac{\sum_{i=2}^{k}\lambda_i}{\sum_{i=1}^{k}\lambda_i} = 1 \quad\Rightarrow\quad X_1 \,\Big|\, \sum_{i=1}^{k} X_i = n \;\sim\; Bin\left(n, \frac{\lambda_1}{\sum_{i=1}^{k}\lambda_i}\right)$$

Si $k = 2$, alors on a : $X_1 \mid X_1 + X_2 = n \;\sim\; Bin\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$

Exercice #2

$X_1 \sim Bin(n_1, p)$
$X_2 \sim Bin(n_2, p)$
$P(X_1 = x \mid X_1 + X_2 = n) = \dfrac{P(X_1 = x \ \cap \ X_1 + X_2 = n)}{P(X_1 + X_2 = n)} = \dfrac{P(X_1 = x)P(X_2 = n - x)}{P(X_1 + X_2 = n)}$
$X_1 + X_2 \sim Bin(n_1 + n_2, p)$

$$\begin{aligned} P(X_1 = x \mid X_1 + X_2 = n) &= \frac{\binom{n_1}{x} p^x (1-p)^{n_1 - x} \binom{n_2}{n-x} p^{n-x} (1-p)^{n_2 - n + x}}{\binom{n_1 + n_2}{n} p^n (1-p)^{n_1 + n_2 - n}} \\ &= \frac{\binom{n_1}{x}\binom{n_2}{n-x}}{\binom{n_1 + n_2}{n}} \end{aligned}$$

Si on dit que la taille de la population est $n_1 + n_2$, que celle de l'échantillon est n et que n_1 individus dans la population ont un certain caractère, donc n_2 ne l'ont pas, alors $X_1 \mid X_1 + X_2 = n$ est une hypergéométrique ($X_1 \mid X_1 + X_2 = n$ représente le nombre d'individus ayant le caractère dans l'échantillon).

Exercice #3

1) $X^2_{obs} = \dfrac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}} = \dfrac{4526(1198 \times 1278 - 557 \times 1493)^2}{2691 \times 1835 \times 1755 \times 2771} = 92.2$
$\chi^2_{\alpha,(I-1)(J-1)} = \chi^2_{0.05,1} = 3.841$
Au seuil de 5%, il y a dépendance entre X et Y car $X^2_{obs} > \chi^2_{0.05,1}$
$r_{ij} = \dfrac{n_{ij} - \widehat{\mu}_{ij}}{\sqrt{\widehat{\mu}_{ij}}}$ où $\widehat{\mu}_{ij} = \dfrac{n_{i\cdot}n_{\cdot j}}{n}$
$r_{11} = 4.78 \quad r_{12} = -5.79 \quad r_{21} = -3.81 \quad r_{22} = 4.61$
$\widehat{\theta} = \dfrac{n_{11}n_{22}}{n_{12}n_{21}} = \dfrac{1198 \times 1278}{557 \times 1493} = 1.841$

2) $\ln\theta \in \ln\widehat{\theta} \pm z_{\alpha/2}\sqrt{\displaystyle\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{1}{n_{ij}}}$
$\ln\theta \in \ln(1.841) \pm 2.5758\sqrt{\dfrac{1}{1198} + \dfrac{1}{557} + \dfrac{1}{1493} + \dfrac{1}{1278}}$
$\ln\theta \in [0.4457, 0.7749]$
$\theta \in [e^{0.4457}, e^{0.7749}] = [1.562, 2.170]$

3) $\pi_M - \pi_F \in p_M - p_F \pm z_{\alpha/2}\sqrt{\widehat{\sigma}(p_M - p_F)}$
$p_M = \dfrac{n_{11}}{n_{\cdot 1}} = \dfrac{1198}{2691} = 0.4452$
$p_F = \dfrac{n_{12}}{n_{\cdot 2}} = \dfrac{557}{1835} = 0.3035$
$\widehat{\sigma}(p_M - p_F) = \sqrt{\dfrac{p_M(1-p_M)}{n_{\cdot 1}} + \dfrac{p_F(1-p_F)}{n_{\cdot 2}}} = 0.014387$
$\pi_M - \pi_F \in 0.1417 \pm 0.03706 = [0.1046, 0.1788]$

Exercice #4

1)

Z	indépendance entre X et Y	$\widehat{\theta}$
A	non	0.3492
B	oui	0.8025
C	oui	1.1331
D	oui	0.9213
E	oui	1.2216
G	oui	0.8279

Dans le département A, la proportion d'hommes admis est significativement inférieure à celle de femmes admises.

Dans le département B, la proportion d'hommes admis est inférieure à celle de femmes admises, mais pas significativement.

Dans le département C, la proportion d'hommes admis est 13% supérieure à celle de femmes admises, mais cette différence n'est pas significative.

Dans le département D, la proportion d'hommes admis inférieure à celle de femmes admises, mais pas significativement.

Dans le département E, la proportion d'hommes admis est 22% supérieure à celle de femmes admises, mais cette différence n'est pas significative.

Dans le département G, la proportion d'hommes admis est inférieure à celle de femmes admises, mais pas significativement.

2) Sans faire de distinction entre les départements, la proportion d'homme admis est 84% supérieure à celle de femmes admises. Cette différence est significative et va totalement à l'encontre de l'analyse conditionnelle : dans celle-ci, la seule différence significative montre que la proportion d'hommes admis est inférieure. C'est un exemple du paradoxe de Simpson. La conclusion erronée de l'analyse marginale s'explique par le fait que les hommes font beaucoup plus de demandes au département A (où les femmes ne font presque pas de demandes) que dans les autres départements. Et que le département A est celui où il y a beaucoup plus d'admis(es) que dans les autres départements.

Exercice #5

$p(x) = P_{H_0}(X = x) = \binom{x}{7} 0.5^7$

$p(0) = p(7) = 0.0078$
$p(1) = p(6) = 0.0547$
$p(2) = p(5) = 0.1641$
$p(3) = p(4) = 0.2734$

T_i : fréquence théorique de la catégorie i

$T_i = 1334p(i)$

$T_0 = T_7 = 10.42$
$T_1 = T_6 = 72.95$
$T_2 = T_5 = 218.86$
$T_3 = T_4 = 364.77$

Distribution du nombre de familles qui tombent dans chaque catégories :

$P_{obs}(X = 0) = 0.0045$	$P_{obs}(X = 7) = 0.0097$
$P_{obs}(X = 1) = 0.0427$	$P_{obs}(X = 6) = 0.0517$
$P_{obs}(X = 2) = 0.1544$	$P_{obs}(X = 5) = 0.1919$
$P_{obs}(X = 3) = 0.2714$	$P_{obs}(X = 4) = 0.2736$

$X^2 = 13.296$
$\chi^2_{0.01,7} = 18.475$
Je serais plus porté à dire que X^2 provient d'une χ^2_6.
Si X^2 est petit, le modèle est "bien ajusté".
Si on présume une binomiale, une hypothèse pouvant être violée est celle que le deuxième paramètre est 0.5