

Rappel: $N_i(N_1, \dots, N_k)$ suit la loi multinomiale de paramètre (n, p) ($n \in \mathbb{N}^*$, $p \in [0, 1]^k$ avec $\sum_{i=1}^k p_i = 1$)
 ou $\forall (n_1, \dots, n_k) \in \mathbb{N}^k$ t.q. $\sum_{i=1}^k n_i = n$ $P(N_1=n_1, \dots, N_k=n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$
 (on laisse de côté le test d'appartenance à un modèle paramétrique)

Exercice 1

Test du χ^2 ...

d'ajustement/adéquation

d'indépendance

d'homogénéité

formalisation: On observe $N \sim \text{Mult}(n, p)$, p inconnu

$H_0: p = p_0$

où p_0 est connu
(vecteur de proba de référence)

$H_1: p \neq p_0$

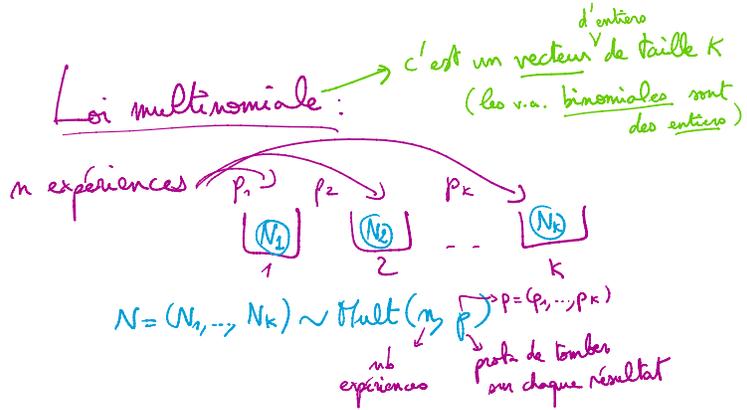
Statistique de test:

$T = \sum_{i=1}^k \frac{(N_i - n p_{0i})^2}{n p_{0i}}$

nb observé \rightarrow nb attendu
somme sur toutes les issues possibles
 χ^2 , $np_i \geq 5$

Sous H_0 , dans la limite $n \rightarrow +\infty$,

$T \sim \chi^2(K-1)$



Exercice 2

$N?$ $m?$ $p?$ $p_0?$

2 issues: anticorps et pas anticorps

$I \text{ à } K=2$
 $p_0 = (p_{01}, p_{02})$

$m=50$
 $p = (p_1, p_2) \leftarrow$ INCONNU

on a tiré 50 personnes blondes

$N = (N_1, N_2)$

Si $K=2$ et $N = (N_1, N_2) \sim \text{Mult}(n, p)$
alors $N_1 \sim \text{Bin}(n, p_1)$
et $N_2 = m - N_1$
($N_2 \sim \text{Bin}(n, p_2)$)

(Suite du rappel de cours, test d'ajustement)

Sous H_1 , $T \xrightarrow{n \rightarrow +\infty} +\infty$

On prend donc une région de rejet $R = [t_\alpha, +\infty[$

Niveau: $P_{H_0}(T \in R) \leq \alpha : P_{T \sim \chi^2(K-1)}(T \geq t_\alpha) \leq \alpha$
 $\hookrightarrow t_\alpha =$ quantile d'ordre $1-\alpha$ de $\chi^2(K-1)$

- 1 = produire des anticorps
- 2 = ne pas en produire

Alors $p_0 = (p_{01}; p_{02})$
 $N = (10; 40) \sim \text{Mult}(n, p)$
 $n = 50$

p inconnu.

* H_0 : la probabilité de produire des anticorps est la même entre les blonds et la population globale (blonds inclus)
 H_1 : différente

* $H_0: p = p_0$ $H_1: p \neq p_0$

* $T = \frac{(N_1 - n p_{01})^2}{n p_{01}} + \frac{(N_2 - n p_{02})^2}{n p_{02}} = \frac{(10-5)^2}{5} + \frac{(40-45)^2}{45} = 5,56 = T_{obs}$

Sous H_0 , dans la limite $n \rightarrow +\infty$, $T \sim \chi^2(2-1) = \chi^2(1)$.

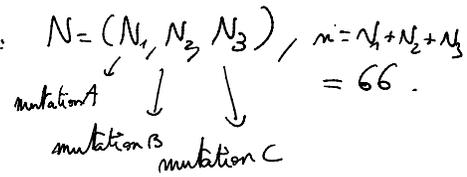
* Région de rejet: $R = [t_\alpha, +\infty[$ avec $t_\alpha =$ quantile d'ordre $1-\alpha = 95\%$ de $\chi^2(1)$
 $[t_\alpha = 3,84]$

Sous H_0 , dans la norme χ^2 ,
 * Région de rejet: $R = [t_{\alpha}, +\infty[$ avec $t_{\alpha} =$ quantile d'ordre $1-\alpha = 95\%$ de $\chi^2_{(1)}$
 $t_{\alpha} = 3,84$
 * $T^{obs} \geq 3,84$ donc on rejette H_0 : la probabilité de produire des anticorps est différente entre la population blonde et la population totale.

⚠ On n'a pas montré une relation de causalité, seulement une corrélation.

Exercice 3

On observe les mutations exhibées par des bactéries cultivées dans l'espace: $N = (N_1, N_2, N_3)$, $n = N_1 + N_2 + N_3 = 66$.



On suppose $N \sim \text{Mult}(n, p)$, p inconnue.

On veut comparer p à sa valeur sur Terre $p_0 = (0,25; 0,40; 0,35)$.

$$N = (12, 33, 21)$$

H_0 : la répartition des mutations est la même sur Terre ou dans l'espace
 H_1 : _____ n'est pas _____

$H_0: p = p_0$ $H_1: p \neq p_0$

Statistique de test: $T = \sum_{i=1}^3 \frac{(N_i - n(p_0)_i)^2}{n(p_0)_i} = 12 - \text{ (interruption pour cause de panne de stylet) }$