Test du chi-carré — Test de conformité (ajustement) / Test d'indépendance

1) Test d'ajustement

i) Problème

On définit sur la population r évènements $E_1, .., E_r$, syst. complet d'évènements.

Modèle théorique : les probabilités des E_i sont $p_1, p_2, ..., p_r$.

Échantillon de taille n : les <u>effectifs observés</u> sont $O_1, O_2, ..., O_r$.

On les compare avec les <u>effectifs théoriques</u> dits <u>effectifs calculés</u> : $\boxed{C_i = n p_i}$
(pas forcém. des entiers)

On compare les O_i et les C_i. Les deux distributions sont-elles conformes ou bien le modèle ne décrit pas bien la réalité ?

ii) <u>Le Test</u> (H_0) : la distribution observée est conforme à la distrib. théorique choisie.

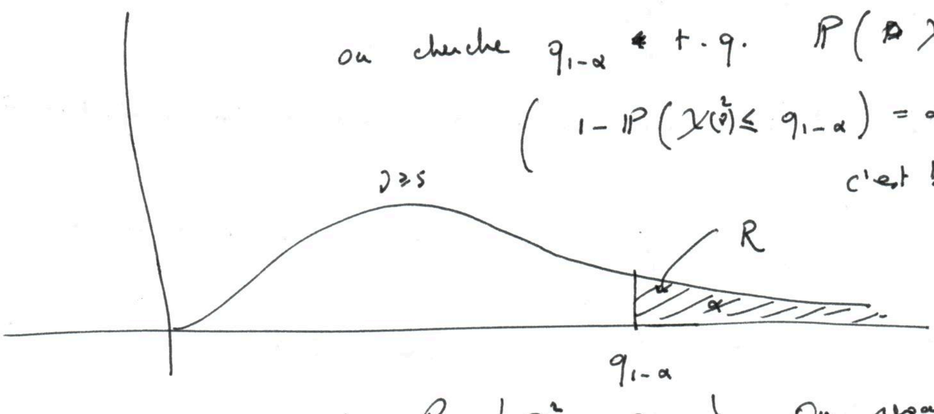
Évènements	E_1	E_2	...	E_r	Totaux
O_i	O_1	O_2	---	O_r	n
p_i	p_1	p_2	...	p_r	1
$C_i = np_i$	$C_1 = np_1$	$C_2 = np_2$	...	$C_r = np_r$	n

<u>Fait</u> Si (H_0) vraie, la variable D_ν^2 qui prend sur chaque échantillon la valeur

$$d_\nu^2 = \sum_{i=1}^{r} \frac{(O_i - C_i)^2}{C_i} = \frac{(O_1 - C_1)^2}{C_1} + \dots + \frac{(O_r - C_r)^2}{C_r}$$

suit la loi du $\chi^2(\nu)$ avec $\nu = r-1$ d.d.l.

iii) <u>Décision, conclusion</u> On fixe α, risque 1ère espèce. On considère la loi $\chi^2(\nu)$



on cherche $q_{1-\alpha}$ t.q. $\mathbb{P}(\chi^2(\nu) \geq q_{1-\alpha}) = \alpha$

$\Big(1 - \mathbb{P}(\chi^2(\nu) \leq q_{1-\alpha}) = \alpha$ donc $\mathbb{P}(\chi^2(\nu) \leq q_{1-\alpha}) = 1-\alpha$
c'est bien le quantile $\Big)$

Alors la zone de rejet est $R = \{ D_\nu^2 \geq q_{1-\alpha} \}$. On regarde la valeur d_ν^2 :

~~Si~~ Si $d_\nu^2 \geq q_{1-\alpha} \Rightarrow H_0$ rejetée avec proba α de se tromper.

si $d_\nu^2 < q_{1-\alpha} \Rightarrow$ on doit accepter H_0.

<u>p-valeur</u> : $\bar{\alpha}$ t.q. $q_{1-\bar{\alpha}} = d_\nu^2$.

2) Test χ^2 d'indépendance de deux caractères

i) Problème

Dans une population P, chaque individu possède deux caractères qualitatifs A et B, ayant les modalités respectives $A_1, \dots, A_r$ et $B_1, \dots, B_l$.

On prend un échantillon de taille n : pour tout $i \in \{1, \dots, r\}$ et tout $j \in \{1, \dots, l\}$ on connaît les effectifs observés O_{ij}, individus qui ont A_i et B_j.

On a évidemment $n = \sum_{i=1}^{r} \sum_{j=1}^{l} O_{ij}$.

Les deux caractères sont-ils indépendants ou il y a une corrélation ?

ii) Le test (H_0) : A et B sont indépendants.

B / A (Obs / Calc)	B_1	B_2	...	B_j	...	B_l	Totaux
A_1	O_{11} / C_{11}	O_{12} / C_{12}		O_{1j} / C_{1j}		O_{1l} / C_{1l}	T_1
A_2	O_{21} / C_{21}	O_{22} / C_{21}		O_{2j} / C_{2j}		O_{2l} / C_{2l}	T_2
$\vdots$							$\vdots$
A_i	O_{i1} / C_{i1}	O_{i2} / C_{i2}		O_{ij} / C_{ij}		O_{il} / C_{il}	T_i
$\vdots$							$\vdots$
A_r	O_{r1} / C_{r1}	O_{r2} / C_{r2}		O_{rj} / C_{rj}		O_{rl} / C_{rl}	T_r
Totaux	S_1	S_2	...	S_j	...	S_l	n

On calcule les effectifs théoriques C_{ij} sous l'hypothèse (H_0) :

sous (H_0), $\forall (i,j)$ les évènements A_i et B_j sont indép., donc

$$\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\,\mathbb{P}(B_j) \Rightarrow \frac{C_{ij}}{n} = \frac{T_i}{n} \cdot \frac{S_j}{n}$$

donc $\boxed{C_{ij} = \frac{T_i \cdot S_j}{n}}$ et on procède comme dans le test d'ajustement :

Fait Sous (H_0), la variable D_ν^2 qui prend sur chaque échantillon la valeur

$$d_\nu^2 = \sum_{i=1}^{r} \sum_{j=1}^{l} \frac{(O_{ij} - C_{ij})^2}{C_{ij}}$$

suit la loi du χ^2 à $\nu = (r-1)(l-1)$ d.d.l.

Décision, conclusion, p-valeur : comme en 1).