08.12.2025
lundi, 8 décembre 2025 07:47

Rappel: test d'adéquation du χ^2 :

But: tester H_0 "la loi $\mathbb{P}$ de l'échantillon $X_1, \dots, X_n$ est $\mathbb{Q}$".

Teste basé sur une statistique et une décomposition des réalisations (discrétisation).

$\mathbb{R} = \Omega_1 \cup \dots \cup \Omega_k$ pour $k \geq 1$, $\Omega_i \cap \Omega_j = \emptyset$ si $i \neq j$.

On teste H_0 : " $\mathbb{P}(\Omega_i) = \mathbb{Q}(\Omega_i) \ \forall i = 1, \dots, k$ ".

$$\hookrightarrow \quad Z_n(X_1, \dots, X_n) = n \sum_{i=1}^{k} \frac{\left(\frac{\sum_{j=1}^{n} \mathbb{1}_{\Omega_i}(X_j)}{n} - \mathbb{Q}(\Omega_i) \right)^2}{\mathbb{Q}(\Omega_i)}$$

- $Z_n \xrightarrow{\text{loi}} \chi^2_{k-1}$ si $\mathbb{P}(\Omega_i) = \mathbb{Q}(\Omega_i) \ \forall i$.

- $Z_n \longrightarrow +\infty$ si $\exists i \in \{1, \dots, k\}$, tq. $\mathbb{P}(\Omega_i) \neq \mathbb{Q}(\Omega_i)$.

Exemple d'application: Expérience de Mendel sur la transmission des traits génétiques.

Teste la transmission des charac. de petits pois

Teste la transmission des charac. de petits pois après hybridation.

Théorie : on a un génome constitué de paires de gênes, et on transmet un gêne de chaque paire à nos descendants.

Sur les pois : Gêne couleur : "c" pour couleur verte, récessif.

"C" pour couleur jaune, dominant.

Gêne forme : "f" pour "ridé", récessif

"F" pour "rond / lisse", dominant.

Procédure de teste : on hybride deux plants CcFf et on regarde le génome des "enfants".

↳ on répète ce processus plein de fois et on regarde la fréquence des différentes expressions de gênes

Si Mendel a juste, on doit avoir que tous les 16 génotypes possibles sont équiprobables

↳ on prédis

| Vert | Jaune |

	Vert	Jaune
ridé	$1/16$	$3/16$
lisse	$3/16$	$9/16$

Test du χ^2 pour vérifier si l'échantillon de Mendel respecte ces prédictions.

	J + lisse	J. + ridé	V. + lisse	V + ridé	
Nb. de pois	315	101	108	32	556
Fréq. emp.	$315/556$	$101/556$	- -	. —	
Fréq. théo.	$9/16$	$3/16$	$3/16$	$1/16$	

Notre stat. de test : $n = 556$, $k = 4$

$$Z_{556} = n \sum_{i=1}^{k} \frac{\left(\frac{\sum_{j=1}^{n} \mathbb{1}_{\Omega_i}(X_j)}{n} - Q(\Omega_i) \right)^2}{Q(\Omega_i)}$$

$$= 556 \left(\frac{\left(\frac{315}{556} - 9/16\right)^2}{9/16} + \frac{\left(\frac{101}{556} - 3/16\right)^2}{3/16} + \frac{\left(\frac{108}{556} - 3/16\right)^2}{3/16} + \frac{\left(\frac{32}{556} - 1/16\right)^2}{1/16} \right)$$

≈ 0.47

Région de rejet pour un test à niveau de confiance 95%.

$$\{ Z_{556} \geq C \} \quad \text{avec } C \text{ tq. } \mathbb{P}(\chi^2_3 > C) = 0.05$$

$\hookrightarrow \quad C = 7.82.$

$0.47 \ll 7.82$, on ne rejette pas H_0.

Principe d'appli : on souhaite que $\forall i, i'$

$$\frac{\left(\frac{\sum_{j=1}^{n} \mathbb{1}_{\Omega_i}(X_j)}{n} - Q(\Omega_i) \right)^2}{Q(\Omega_i)}$$

et l'équivalent pour i' ait un ratio au plus ... .

$\Leftrightarrow$ on veut que les $Q(\Omega_i)$ soient toutes du même ordre de grandeur.

Exemple de test d'indépendance du χ^2 :

Yeux \ chev.	Blond	Brun	Noir	Roux	tot.	Freq.
Bleu	25	9	3	7	44	44/124
Gris	13	17	10	7	47	:
Brun	7	13	8	5	33	:

Brun	7	13	8	5	33	:
tot.	45	39	21	19	124	
Freq.	45/124	- -	- -	- -	- -	

<u>But</u>: tester si "couleur de cheveux" est indép. de "couleur d'yeux".

On introduit les quantités:

p_{ij} = $\mathbb{P}$(un individu ait les yeux de couleur i et les cheveux de couleur j)

<u>Yeux</u>: $1 \equiv$ bleu, $2 \equiv$ gris, $3 \equiv$ brun

<u>Cheveux</u>: $1 \equiv$ blond, $2 \equiv$ brun, $3 \equiv$ noir, $4 \equiv$ roux.

$p_{i*} = \sum_{j=1}^{4} p_{ij}$ = $\mathbb{P}$(un individu ait les yeux i)

$p_{*j} = \sum_{i=1}^{3} p_{ij}$ = $\mathbb{P}$(un individu ait les cheveux j)

Si cheveux et yeux indép. on doit avoir

$$p_{ij} = p_{i*} \cdot p_{*j} \quad \forall i,j.$$

On introduit l'hypothèse nulle <u>paramétrique</u>

H_0 : " $p_{ij} = p_{i*} \cdot p_{*j} \quad \forall\ i=1,..,3,$ et $j=1,..,4$ ".

On construit une statistique de test:

On construit une statistique de test :

N_{ij} = nombre d'individus dans l'échantillon avec yeux i et cheveux j.

N_{i*} = nombre — " ———————— avec les yeux i

N_{*j} = nombre — " ———————— avec les cheveux j.

$$Z_n = n \sum_{i=1}^{3} \sum_{j=1}^{4} \frac{\left(\frac{N_{ij}}{n} - q_{ij}\right)^2}{q_{ij}}$$

→ Sous H_0, on a $q_{ij} = p_{ij} = p_{i*} \cdot p_{*j}$

$$\rightsquigarrow Z_n = n \sum_{i=1}^{3} \sum_{j=1}^{4} \frac{\left(\frac{N_{ij}}{n} - p_{i*} \cdot p_{*j}\right)^2}{p_{i*} \cdot p_{*j}}$$

⚠ on ne connait pas p_{i*} et p_{*j}.

<u>Solution</u> : on les approxime

$$\rightsquigarrow Z_n = n \sum_{i=1}^{3} \sum_{j=1}^{4} \frac{\left(\frac{N_{ij}}{n} - \frac{N_{i*}}{n} \cdot \frac{N_{*j}}{n}\right)^2}{\frac{N_{i*}}{n} \cdot \frac{N_{*j}}{n}}$$

$\frac{N_{i*}}{n}$ est l'estimateur "moyenne empirique" de p_{i*}.

Z_n sera notre statistique de test. Il se trouve que Z_n converge vers une loi du χ^2 : degrées de liberté ?

12 paramètre - 1 relation entre eux

- nbre de relations induites par les estimations empirique de p_{i*}, p_{*j}.

$3 + 4 - 1 - 1$ car $p_{1*} + p_{2*} + p_{3*} = 1$

$= 5$ $\quad p_{*1} + p_{*2} + \dots + p_{*4} = 1$

$12 - 1 - 5 = 6$ degrés de liberté.

<u>t - tests</u> : De la même façon que les tests du χ^2 sont basés sur trouver une statistique de test qui suit une loi du χ^2, les t-tests sont basés sur trouver une statistique de test qui suit une loi de Student.

<u>Déf</u> : (loi de Student). Soit $\nu > 0$. On dit qu'une variable aléatoire X suit une <u>loi de Student avec paramètre ν</u> si c'est une v.a. continue avec densité

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\;\Gamma\left(\nu/2\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Si $\nu \in \mathbb{N}^*$, on parlera de loi de Student à

Si $\nu \in \mathbb{N}^*$, on parlera de loi de Student à ν degrés de libertés.

Cette loi apparait en stat à travers le resultat suivant :

Lemme (4.5.1) Soit Z, V deux v.a. indépendantes avec $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_k$ pour $k \in \{2,3,\dots\}$.

Alors,

$$Z\sqrt{\frac{k}{V}} \sim \text{Student}(k).$$

Cette structure va apparaitre sous la forme

$$\frac{(\overline{X}_n - E[X_1])}{\sqrt{\hat{\sigma}_n^2}} = \frac{\sqrt{n}}{\sqrt{n}} \underbrace{\frac{\sqrt{n}\,(\overline{X}_n - E[X_1])}{\sqrt{\mathrm{Var}(X_1)}}}_{\sim \mathcal{N}(0,1)} \cdot \underbrace{\frac{1}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}}}_{\text{"}\sim\text{" } \frac{1}{\sqrt{\chi^2_{n-1}}}}$$