Liste des Tests du cours

Charles Vin

2022

Table des matières

3 Le test du $\mathcal{X}^2$ d'ajustement

Conditions

1. Les X_i sont à valeur dans un ensemble fini (loi discrète). Si a valeur dans $\mathbb{N}$, on fusionne les classes à partir d'un certain rang choisis
2. Test asymptotique : $\forall k \in \{1, \dots, d\}, np_k^{ref}(1 - p_k^{ref}) \geq 5 \Leftrightarrow n \geq 20$

Si on ne remplis pas les conditions, on peut fusionner les classes

Hypothèse

$$H_0 = p = p^{ref} \text{ i.e. } \forall k \in \{1, \dots, d\}, p_k = p_k^{ref}$$
$$H_1 = p \neq p^{ref} \text{ i.e. } \exists k \in \{1, \dots, d\} : p_k \neq p_k^{ref}$$

Avec p^{ref} un vecteur fixé à tester (par exemple pour un lancé de dé $(\frac{1}{6}, \dots, \frac{1}{6})$)

Statistique de test

$$\begin{aligned} D(\bar{p_n}, p^{ref}) &= n \sum_{k=1}^{d} \frac{(\bar{p_{k,n}} - p_k^{ref})^2}{p_k^{ref}} \rightarrow_{n \to \infty}^{\mathcal{L}} \mathcal{X}^2(d-1) \\ &= \sum_{k=1}^{d} \frac{(N_{k,n} - np_k^{ref})^2}{np_k^{ref}} \end{aligned}$$

avec

- $N_{k,n} = \sum_{i=1}^{n} \mathbb{1}_{X_i x_k}$ (ce qu'il y a dans le tableau de la consigne)
- $\bar{p_{k,n}} = \frac{N_{k,n}}{n}$ les proportions observés

Zone de Rejet

$$\mathcal{R} = \{D(\bar{p_n}, p^{ref}) \geq h_\alpha\}.$$

avec h_α le quantile d'ordre $1-\alpha$ de la loi $\mathcal{X}^2(d-1)$

Méthode

1. Etape 0 : On vérifie les conditions
$$\forall k \in \{1, \dots, d\}, n * p_k \geq 5.$$
C'est la condition de Cochran (1954), il avait testé cas possible en observant l'approximation faites.
2. Etape 1 : On calcule les effectifs et proportions observées : $N_{k,n}$ et $\hat{p}_{k,n}$
3. Etape 2 : Calcul de la statistique de test
$$D = n \sum_{d}^{k=1} \frac{(\hat{p}_{k,n} - p_k)^2}{p_k}.$$
4. Etape 3 : Détermination de la zone de rejet au niveau α. On lit h_α le quantile d'ordre $1-\alpha$ de la loi $\mathcal{X}^2(d_1)$
5. Etape 4 : Décisions
   — si $D > h_\alpha$, on rejette H_0 (au niveau α).
   — Si $D \leq h_\alpha$ on conserve H_0

Bilan de la méthode

Aspects positifs :
— **Fonctionne pour toutes les lois**
— Facile à faire
Aspects négatifs :
— Problème de consistance. Regrouper les variables par intervalle ruiner l'erreur de seconde espèce.
— Asymptotique
— Dépendant du choix des intervalles. Ce qui n'est pas canonique.

3.1 Le $\mathcal{X}^2$ d'ajustement à une famille paramétrique de loi

Pratiquement comme avant, pas encore fait en TD, mais copier collé du cours quand même

1. Etape 1 : Soit $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ (pour P_θ). On estime **tous** les paramètres de la loi $(p_1^{\hat{\theta}_n}, \dots, p_d^{\hat{\theta}_n})$
2. Etape 2 : On vas tester l'ajustement de $X_1, \dots, X_n$ à $P_{\hat{\theta}_n}$ On calcule les fréquences observées $\hat{p}_{k,n}$.
3. Etape 3 : Vérification des conditions $np_k^{\hat{\theta}_n}$ et possible regroupement en classes
4. Etape 4 : Calcul de la stat de test D
5. Etape 5 : Zone de rejet : lecture de H_α le quantile d'ordre $1-\alpha$ d'une $\mathcal{X}^2(d-1-M)$ avec M nombre de paramètre.
6. Etape 6 : Décision
   — $D > h_\alpha$ on rejette H_0
   — $D \leq h_\alpha$ on conserve H_0

5 Test du $\mathcal{X}^2$ d'indépendance

Donnée

$(X_1, Y_1), \ldots, (X_T, Y_T)$ iid appariés.

— X_1 à valeur dans $A_1, \ldots, A_M$

— Y_1 à valeur dans $B_1, \ldots, B_N$

Conditions

— Loi discrète

— n ou T plutôt grand

— $\forall i < M, j < N$: $T * \hat{p}_m \hat{q}_m \geq 5$ ou avec la notation en TD : $E_{i,j} \geq 5$

Hypothèse

— H_0 : $X_1 \perp Y_1$

— H_1 : $X_1 \not\perp Y_1$

Statistique de test

$$
\begin{aligned}
D &= T * \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{(\hat{p}_{m,n} - \hat{p}_m \hat{q}_n)^2}{\hat{p}_m \hat{q}_n} \\
&= \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{(N_{m,n} - \frac{N_{m,\bullet} N_{\bullet,n}}{T})^2}{\frac{N_{m,\bullet} N_{\bullet,n}}{T}}
\end{aligned}
$$

On utilise la deuxième en TD, la fraction est équivalent à $E_{i,j}$ aka le produit en croix à l'intérieur du tableau durant les TD (groupe 2)

Zone de Rejet

- Sous H_0, $D \to \mathcal{X}^2((M-1)(N-1))$
- Sous H_1, $D \to +\infty$

$$\mathcal{R} = \{D \geq h_\alpha\}.$$

Méthode

yeux \ cheveux	Noirs (A_1)	Bruns	Blonds	Roux	Total
Marrons (B_1)	$N_{1,1} = 152$	$N_{2,1} = 247$	83	11	$N_{\bullet,1} = 152$
Vert ou Gris	73	114	37	8	232
Bleus	36	167	127	10	275
Total	$N_{1,\bullet} = 261$	463	247	29	1000

$$E_{1,1} * 1000 = 152 * 261$$

$$\Leftrightarrow E_{1,1} = \frac{152 * 261}{1000} = 128.67$$

yeux \ cheveux	Noirs (A_1)	Bruns	Blonds	Roux	Total
Marrons (B_1)	128.67	228.26	121.77	14.3	$N_{\bullet,1} = 152$
Vert ou Gris	60.55	107.42	57.3	6.73	232
Bleus	71.78	127.32	67.93	7.98	275
Total	$N_{1,\bullet} = 261$	463	247	29	1000

Puis calculer la stat de test

$$D = \sum_{\text{chaque case du tableau}} \frac{(N_{1,1} - E_{1,1})^2}{E_{1,1}}.$$

6 Test du $\mathcal{X}^2$ d'homogénéité

Donnée

- $X_1, \ldots, X_{n_1}$ échantillons iid
- $Y_1, \ldots, Y_{n_2}$ échantillons iid
- Échantillons indépendant entre eux

Les variables sont toutes à valeurs dans les mêmes classes $A_1, \ldots, A_M$.

Conditions

Hypothèse

On veut tester l'homogénéité

- $H_0 = X_1$ et Y_1 ont la même loi $\Leftrightarrow \forall m \in \{1, \ldots, M\}, P(X_1 \in A_m) = P(Y_1 \in A_m)$
- $H_1 = X_1$ et Y_1 n'ont pas la même loi $\Leftrightarrow \exists m \in \{1, \ldots, M\}$ tel que $P(X_1 \in A_m) \neq P(Y_1 \in A_m)$