

Liste des Tests du cours

Charles Vin

2022

1 Test de Kolmogorov-Smirnov

Conditions

1. Les X_i semblent provenir d'une loi à fonction de répartition continue. \Rightarrow on n'a pas plusieurs fois la même valeur (sauf si celle-ci on était arrondi).
2. Fonctionne $\forall n$: même si n est petit, ce test est pertinent
3. Si $n \geq 100$, on fait un test asymptotique.

Hypothèse

- H_0 = les X_i ont pour fdr. F_X
- H_1 = les X_i n'ont pas pour fdr. F_X

Statistique de test

$$h(F_n, F) = \max_{1 \leq i \leq n} (\max(\left| \frac{i}{n} - F(X_{(i)}) \right|, \left| \frac{i-1}{n} - F(X_{(i)}) \right|)).$$

Zone de Rejet

$$\mathcal{R} = \{h(F_n, F_X) \leq h_{1-\alpha}\}.$$

avec F_n fonction de répartition empirique, $h_{1-\alpha}$ le quantile à aller chercher dans la table

Méthode

Pour trouver la valeur de $h(F_n, F_X)$: Faire le grand tableau puis trouver le max. Exemple :

i	1	2	3	4	5
$X_{(i)}$	0.3	0.7	0.9	1.2	1.4
$X_{(i)} - 2$	-1.70	-1.30	-1.10	-0.80	-0.60
$F_0(X_{(i)})$	0.04	0.10	0.14	0.21	0.27
$\frac{i}{n}$	0.05	0.1	0.15	0.2	0.25
$\left \frac{i}{n} - F_0(X_{(i)}) \right $	0.01	0.00	0.01	0.01	0.02
$\left \frac{i-1}{n} - F_0(X_{(i)}) \right $	0.04	0.05	0.04	0.06	0.07

TABLE 1 – Ici le max c'est 0.07 à la dernière case

2 Adéquation à une famille d'exponentielle

Conditions

Hypothèse

Statistique de test

Zone de Rejet

Méthode

3 Adéquation à une loi normale

Conditions

Hypothèse

Statistique de test

Zone de Rejet

Méthode

4 Le test du χ^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 remplit pas les conditions, on peut fusionner les classes

Hypothèse

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

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

$$D(\bar{p}_n, p^{ref}) = n \sum_{k=1}^d \frac{(p_{k,n}^- - p_k^{ref})^2}{p_k^{ref}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(d-1).$$

avec

- $N_{k,n} = \sum_{i=1}^n \mathbb{1}_{X_i = x_k}$
- $p_{k,n}^- = \frac{N_{k,n}}{n}$ 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 $\chi^2(d-1)$

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 $\chi^2(d_1)$
5. Etape 4 : Décisions
 - si $D > h_\alpha$, on rejette H_0 (au niveau α).
 - Si $D \leq h_\alpha$ on conserve H_0

4.0.1 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.

4.1 Le χ^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 $\chi^2(d - 1 - M)$ avec M nombre de paramètre.
6. Etape 6 : Décision
 - $D > h_\alpha$ on rejette H_0
 - $D \leq h_\alpha$ on conserve H_0