

Série de TD n°2 : Tests de Kolmogorov-Smirnov et Cramer-von Mises

Master 2 Probabilités et Statistiques

Département de mathématiques - BISKRA

Objectifs de la série

Cette série de travaux dirigés vise à maîtriser les tests d'adéquation non paramétriques, en particulier les tests de Kolmogorov-Smirnov et Cramer-von Mises. Les exercices couvrent à la fois les aspects théoriques et les applications pratiques.

Exercice 1 : Application Directe

On considère l'échantillon suivant de 12 observations, supposées indépendantes et identiquement distribuées :

0.15	0.22	0.31	0.42	0.48	0.53	0.61	0.67	0.72	0.79	0.85	0.93
------	------	------	------	------	------	------	------	------	------	------	------

On souhaite tester si ces données proviennent d'une loi uniforme sur $[0, 1]$.

- Formuler les hypothèses H_0 et H_1 du test.
- Calculer la fonction de répartition empirique $F_n(t)$ pour chaque point de l'échantillon.
- Calculer la statistique de Kolmogorov-Smirnov D_n .
- Calculer la statistique de Cramer-von Mises W_n^2 .
- Pour un niveau $\alpha = 5\%$, déterminer les valeurs critiques des deux tests à l'aide des tables appropriées.
- Conclure pour chaque test et comparer les résultats.

Exercice 2 : Test de Normalité

Les mesures suivantes représentent la taille (en cm) de 15 étudiants :

168.2	172.5	175.1	169.8	178.3	171.6	174.2	167.9
173.4	170.7	176.1	169.3	172.8	175.9	171.1	

On souhaite tester l'adéquation de ces données à une loi normale.

- Estimer les paramètres μ et σ^2 de la loi normale.
- Calculer la statistique de Kolmogorov-Smirnov D_n pour le test de normalité.
- En utilisant la table de Lilliefors (ou une approximation appropriée), déterminer la p-value du test.
- Effectuer le test de normalité au niveau $\alpha = 5\%$.
- Discuter des limitations du test KS lorsque les paramètres sont estimés à partir des données.

Exercice 3 : Test à Deux Échantillons

Deux méthodes d'apprentissage ont été testées sur deux groupes d'étudiants. Les résultats aux examens sont les suivants :

Groupe A (Méthode traditionnelle)

12.3	14.7	11.8	13.5	15.2	12.9	14.1	13.8
------	------	------	------	------	------	------	------

Groupe B (Nouvelle méthode)

13.8	15.6	14.3	16.1	12.7	15.9	14.8	16.3	15.1
------	------	------	------	------	------	------	------	------

On souhaite tester si les deux méthodes donnent des résultats provenant de la même distribution.

- Formuler les hypothèses du test d'homogénéité.
- Calculer les fonctions de répartition empiriques $F_n(x)$ et $G_m(x)$.
- Calculer la statistique de Kolmogorov-Smirnov à deux échantillons :

$$D_{n,m} = \sup_x |F_n(x) - G_m(x)|$$

- Déterminer la valeur critique pour un niveau $\alpha = 5\%$.
- Effectuer le test et conclure.
- Proposer une interprétation pédagogique des résultats.

Exercice 4 : Test d'adéquation à la loi uniforme

On observe l'échantillon de 20 observations suivant :

0.278	0.452	0.464	0.494	0.496	0.505	0.576	0.592	0.602	0.608
0.661	0.683	0.690	0.696	0.704	0.728	0.754	0.850	0.902	0.949

On souhaite tester si ces données proviennent d'une loi uniforme sur $[0, 1]$.

- Calculer la fonction de répartition théorique $F_U(t)$ de la loi uniforme sur $[0, 1]$.
- Construire un tableau contenant pour chaque observation :
 - La valeur observée x_i
 - La fonction de répartition empirique $F_n(x_i)$
 - La fonction de répartition théorique $F_U(x_i)$
 - L'écart $|F_n(x_i) - F_U(x_i)|$
- Déterminer la statistique $D_n = \sup |F_n(t) - F_U(t)|$.
- Calculer la statistique de Cramer-von Mises W_n^2 .
- Pour $\alpha = 5\%$, utiliser les valeurs critiques :

$$c_{KS} \approx \frac{1.36}{\sqrt{n}} \quad \text{et} \quad c_{CvM} \approx 0.461$$

pour conclure sur les deux tests.

- Comparer les sensibilités des deux tests pour cet échantillon.

Exercice 5 : Synthèse Théorique

a) Propriétés d'invariance

- i) Démontrer que la statistique de Kolmogorov-Smirnov est invariante par transformation monotone croissante.
- ii) Montrer que cette propriété s'applique également au test de Cramer-von Mises.
- iii) Expliquer l'importance pratique de cette invariance.

b) Consistance des tests

- i) Démontrer que le test de Kolmogorov-Smirnov est consistant contre toute alternative fixe.
- ii) Montrer que le test de Cramer-von Mises est également consistant.
- iii) Comparer la puissance asymptotique des deux tests.

c) Distribution asymptotique

- i) Établir que sous H_0 :

$$\sqrt{n}D_n \xrightarrow{\mathcal{L}} \sup_{0 \leq t \leq 1} |B(t)|$$

où $B(t)$ est un pont brownien.

- ii) Donner l'expression de la fonction de répartition limite de $\sqrt{n}D_n$.
- iii) Expliquer comment cette distribution est utilisée pour construire les tables du test KS.

d) Extensions et généralisations

- i) Proposer une extension du test KS pour des données censurées.
- ii) Discuter des adaptations nécessaires pour le test CvM dans le cas de données groupées.
- iii) Mentionner d'autres tests d'adéquation non paramétriques et situer KS et CvM dans ce paysage.