

Tests non-paramétriques basés sur la fonction de répartition

Table des matières

1	Test de Kolmogorov-Smirnov d'ajustement à une loi donnée	1
2	Test de Kolmogorov-Smirnov d'ajustement à une famille de loi paramétrique donnée	1
3	Tests de comparaison de Kolmogorov Smirnov	2
4	TP 3	2

1 Test de Kolmogorov-Smirnov d'ajustement à une loi donnée

Exercice 1.

Soit $X = (X_1, \dots, X_n)$ un échantillon de va iid de fonction de répartition F . Soit F_n la fonction de répartition empirique de cet échantillon. Soit F_0 une fonction de répartition donnée. On définit la statistique :

$$h_n(X, F_0) = \|F_n - F_0\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|.$$

1. Montrer que si F_0 est continue alors $h_n(X, F_0) = \max_{1 \leq j \leq n} \max\{j/n - F_0(X_{(j)}), F_0(X_{(j)}) - (j-1)/n\}$.
2. Montrer en utilisant le théorème de Glivenko-Cantelli que si $F \neq F_0$ alors presque sûrement

$$\liminf_{n \rightarrow \infty} \|F_n - F_0\|_\infty > 0.$$

3. Soit $U = (U_1, \dots, U_n)$ un échantillon de loi uniforme sur $[0, 1]$. Montrer que (X_1, \dots, X_n) et $(F^{(-1)}(U_1), \dots, F^{(-1)}(U_n))$ ont la même loi.
4. Montrer qu'on a $(F^{(-1)}(U_i) \leq x) \iff U_i \leq F(x)$.
5. En déduire que si $F = F_0$ alors $h_n(X, F_0)$ ne dépend que de $Im F_0$ (qui est $F_0(\mathbb{R})$).
6. Montrer que si F_0 est continue et si $F = F_0$ alors $h_n(X, F_0)$ a la même loi que $h_n(U, F_U)$ (F_U est la fonction de répartition de la loi uniforme sur $[0, 1]$).
7. On souhaite tester $H_0 : F = F_0$ contre $H_1 : F \neq F_0$.
 - Vérifier qu'il existe un test $\phi_\alpha(X) = I_{h_n(X, F_0) \geq k_\alpha}$ de **niveau** α . Préciser le seuil k_α .
 - Vérifier que si F_0 est continue alors ce test est même de **taille** égale à α . Calculer dans ce cas la p -valeur du test.

2 Test de Kolmogorov-Smirnov d'ajustement à une famille de loi paramétrique donnée

Exercice 2.

Soit X une variable aléatoire de fonction de répartition inconnue F , et soit (X_1, \dots, X_n) un n -échantillon de la loi de X . Pour tout $\theta > 0$, on note F_θ la fonction de répartition définie par

$$F_\theta(x) = (1 - \exp(-x/\theta))I_{x>0}.$$

1. On suppose que $F \in \mathcal{F}$, où $\mathcal{F} = \{F_\theta, \theta \in \mathbb{R}\}$. Déterminer l'estimateur T du maximum de vraisemblance de θ , puis construire un test de $\theta = 100$ contre $\theta \neq 100$ au niveau 5%.

2. On note F_n la fonction de répartition empirique de (X_1, \dots, X_n) , et on définit

$$\Delta_n = \sup_{t \in \mathbb{R}} |F_n(t) - F_T(t)|.$$

Montrer que la loi de Δ_n est libre de θ lorsque $F \in \mathcal{F}$. En déduire un test de l'hypothèse $F \in \mathcal{F}$ contre $F \notin \mathcal{F}$. Si $n = 5$ et que l'on a observé les valeurs 133, 169, 8, 122 et 58, tester de deux manières l'hypothèse $F = F_{100}$ au niveau 10%.

3 Tests de comparaison de Kolmogorov Smirnov

Exercice 3.

Soient (X_1, \dots, X_n) un n -échantillon d'une loi continue μ et (Y_1, \dots, Y_m) un m -échantillon d'une loi ν , indépendant de (X_1, \dots, X_n) . On note F_n la fonction de répartition empirique de (X_1, \dots, X_n) , G_m la fonction de répartition empirique de (Y_1, \dots, Y_m) et on définit

$$D_{n,m} = \sup_{t \in \mathbb{R}} |F_n(t) - G_m(t)|.$$

On s'intéresse aux hypothèses $H_0 : \mu = \nu$ et $H_1 : \mu \neq \nu$.

1. Montrer que si les X_i et les Y_j ont la même loi, alors la loi de $D_{n,m}$ est libre de μ et ν .
2. En utilisant le fait que la loi de $D_{n,m}$ est connue sous H_0 , construire un test de H_0 contre H_1 .
3. On souhaite comparer deux médicaments censés soulager la douleur post-opératoire. On a observé sur 16 patients dont 8 ont pris un médicament A habituel et les 8 autres un médicament B expérimental, les nombres suivants d'heures de soulagement. Y a-t-il une différence significative au niveau 5% entre A et B ?

A	6,8	3,1	5,8	4,5	3,3	4,7	4,2	4,9
B	4,4	2,5	2,8	2,1	6,6	0,0	4,8	2,3

4 TP 3

Exercice 4.

Un test a été étalonné sur une population A de manière que sa distribution suive une loi normale de moyenne 13 et d'écart type 3. sur un échantillon de taille 10 issu d'une population B , on a observé les valeurs suivantes :

8.43 8.70 11.27 12.92 13.05 13.05 13.17 13.44 13.89 18.90

Ces valeurs sont-elles compatibles avec l'hypothèse selon laquelle la variable sous-jacente est distribuée selon une loi normale de moyenne 13 et d'écart type 3?

Exercice 5.

On a testé un échantillon de 5 appareils et noté leurs durées de vie en heures

Appareil	1	2	3	4	5
Durée de vie	133	169	8	122	58

On voudrait savoir si la durée de vie suit une loi de probabilité exponentielle.

1. Estimer le paramètre λ de la loi exponentielle $\lambda e^{\lambda x} I_{x \geq 0}$
2. Formuler les hypothèses (hypothèse nulle H_0 et hypothèse alternative H_1).

3. Comparer la distribution observée à la distribution théorique au moyen d'un test de Kolmogorov-Smirnov. Que peut-on en conclure ?

Exercice 6.

On considère les deux échantillons observés suivants :

$$x = (0.11, 0.20, 0.69, 1.02), \quad y = (0.86, 0.99, 1.24, 1.57, 1.62)$$

On cherche à tester l'hypothèse nulle H_0 : "les deux échantillons proviennent d'une même loi" contre H_1 : "les deux échantillons proviennent de deux lois différentes".

1. Rappeler la forme de la statistique du test de Kolmogorov-Smirnov à deux échantillons.
2. Calculer la p-valeur du test de Kolmogorov-Smirnov.
3. Rappeler la forme de la statistique de Mann-Whitney, et donner son espérance sous H_0 ainsi que ses valeurs minimale et maximale.
4. Calculer la p-valeur du test de Mann-Whitney. Conclure.