Statistiques décisionnelles : TD 5

Pour effectuer le test d'homogénéité de Kolmogorov-Smirnov, on utilisera ks.test(X,Y) où X et Y sont des vecteurs contenant les observations. Attention, si vous faites ks.test(X,"F") où X est un vecteur et F est une fonction, vous effectuerez un test d'ajustement de Kolmogorov-Smirnov.

Exercice 1. *On mesure les temps de course au* 80 *mètres de deux groupes d'athlètes français et belges. Les temps mesurés sont donnés ci-dessous :*

français : 9.07 9.06 10.7 9.56 10.46 9.3 9.08 11.94 12.28 9.84

belges : 10.52 9.6 11.0 10.67 12.24 10.86 9.74.

Ces deux groupes sont-ils homogènes, au niveau 5% *?*

Exercice 2. *On a mesure les températures moyenne en Janvier tous les ans à Lille entre* 2000 *et* 2010 *et à Zurich entre* 2011 *et* 2020*. Toutes les données sont réelles.*

Lille : 4.2 3.6 4.1 4 5.3 2 7.4 6.5 0.7 0.2

Zurich : 1.2 2 0.6 3.2 1.9 2.8 -3 4.7 0.1 2.6.

Ces deux groupes sont-ils homogènes, au niveau 0.05 *? Quand vous aurez traité l'exercice* 4*, vous pourrez refaire cet exercice en vous demandant s'il fait meilleur en janvier à Lille ou à Zurich.*

Exercice 3. *On a mesuré la vitesse moyenne du vent toutes les semaines pendant les deux derniers mois au cap blanc nez et au cap gris nez, exprimé en* $km.h^{-1}$

Blanc-nez : 34 49 2.7 53 81 61 41 150

Gris-nez : 40 52 5 47 90 64 39 173

Peut-on dire que ces relevés ont la même loi ?

Exercice 4 (Test d'homogénéité Kolmogorov-Smirnov unilatéral ou « one sided »)**.** *Le but de cet exercice est d'introduire une variante du test d'homogénéité de Kolmogorov-Smirnov avec une hypothèse alternative différente de celle vue en cours. Pour deux variables aléatoires* X *et* Y*, on dit que* Y *domine* X *si* $\forall t \in \mathbb{R},\ F_Y(t) \leq F_X(t)$*. On le note* $Y \succ X$*. Cela signifie que* Y *a tendance à prendre des valeurs plus grandes que* X*. Cela ne veut pas dire que* Y *prend toujours des valeurs plus grandes que* X *! Les question ci-dessous ont pour but de vous familiariser avec cette notion.*

1. *Montrer que si* X *est uniforme sur* $[0,1]$ *et* Y *uniforme sur* $[2,3]$*, alors* $Y \succ X$*.*
2. *Montrer que si* X *est uniforme sur* $[0,1]$ *et* Y *est uniforme sur* $[0,2]$ *alors* $Y \succ X$*.*
3. *Montrer que si* X *est uniforme sur* $[0,1[$ *et* $Y \sim \mathcal{N}(0,1)$ *alors* Y *ne domine pas* X*, et* X *ne domine pas* Y*.*

Cette notion de domination est utile pour comparer deux variables aléatoires en statistiques. Par exemple, si on teste l'efficacité d'un médicament, on peut vouloir prendre H_0 "le médicament ne fait pas d'effet" contre H_1 "le médicament a un effet positif" (l'hypothèse usuelle serait "le médicament a un effet").

On dispose de deux échantillons indépendants des variables i.i.d. $(X_1, \ldots, X_n)$ *et* $(Y_1, \ldots, Y_m)$. *On cherche à construire un test pour :*

— H_0 *:* X_1 *et* Y_1 *ont même loi, c'est à dire* $F_{X_1} = F_{Y_1}$.

— H_1 *:* X_1 *et* Y_1 *n'ont pas la même loi, et* $Y_1 \succ X_1$.

On introduit la statistique de test

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

où F_n *est la fonction de répartition empirique des* X_i *et* G_m *est la fonction de répartition empirique des* Y_j.

Les questions suivantes ont pour but de vous aider à vous poser quelques questions sur cette statistique de test, j'attends de vous que vous compariez cette statistique de test au test d'homogénéité de Kolmogorov-Smirnov fait en cours pour vous familiariser avec ces deux objets.

4. *Sous* H_0, *la loi de* $T_{n,m}$ *dépend-elle de la loi des* X_i *?*
5. *Sous* H_1, *si* n *et* m *tendent vers l'infini, quelle est la limite de* $T_{n,m}$ *? Quel est son signe ?*
6. *Construire un test de niveau* α *pour le test de* H_0 *contre* H_1.

Application : *On teste l'effet d'un somnifère sur deux groupes indépendants de* 11 *personnes. Le groupe contrôle a reçu un produit inactif (I), alors que le groupe test a reçu le somifère (S). Les temps de sommeil moyens pour ces individus, exprimés en minutes, sont présentés dans le tableau ci-dessous.*

S	560	470	580	570	550	485	460	540	620	550	620
I	590	530	435	360	430	570	490	480	380	400	350

Si on note X_i *les durées de sommeil du groupe contrôle et* Y_i *les durées de sommeil du groupe test, traduire l'efficacité du somnifère en terme de domination de variables aléatoires. Déterminer si, au niveau* 5%, *ce produit est efficace.*

Super Bonus : *Construisez une variante du test d'ajustement de Kolmogorov-Smirnov pour un échantillon i.i.d.* $X_1, \ldots, X_n$ *et une fonction de répartition cible* F *et les hypothèses* $H_0 : F_{X_1} = F$ *et* $H_1 : F_{X_1} \neq F$ *et* $F_{X_1} \leq F$.

Exercice 5. *Neufs parcelles ont été cultivées dans une région A et sept parcelles dans une région B. On vous donne les rendements constatés :*

Région A : 48,0 48,2 50,3 53,5 54,6 56,4 57,8 58,5 60,5.
Région B : 44,2 46,3 48,3 48,5 50,5 51,2 55,4.

1. *Les deux régions sont-elles aussi fertiles ?*
2. *La région A est-elle plus fertile que la B ?*