Université Paris Dauphine - PSL **M1 Math**
copyright 2019-20 **cours G. Turinici**

Introduction à la statistique non-paramétrique

TD1 : Rappels

Exercice 1 : Soit X une variable aléatoire réelle, absolument continue de densité continue f, de fonction de répartition F. On observe un n-échantillon iid $(X_1, \dots, X_n)$ de même loi que X. On considère la statistique T qui ordonne l'échantillon dans le sens croissant :

$$T(X_1, \dots, X_n) = (X_{(1)}, \dots, X_{(n)}),$$

avec $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. $(X_{(1)}, \dots, X_{(n)})$ s'appelle la statistique d'ordre.

1. On suppose pour cette question uniquement que les X_i sont seulement indépendants et de lois continues (c'est-à-dire que les X_i sont indépendants et ont tous une fonction de répartition F_i continue, mais pas forcément absolument continue) . Montrer que
$$\mathbb{P}\left(\exists\, i \neq j : \quad X_i = X_j\right) = 0,$$
et que dans la définition de la statistique d'ordre, on peut donc se limiter à des inégalités strictes : $X_{(1)} < X_{(2)} < \dots < X_{(n)}$.
2. Déterminer la densité de la loi du n-uplet $(X_{(1)}, \dots, X_{(n)})$.
3. Déterminer la fonction de répartition F_k et la densité f_k de $X_{(k)}$.
4. Montrer que si $\mathbb{E}[|X|]$ est finie, alors il en est de même de $\mathbb{E}[|X_{(k)}|]$.
5. Rappeler les densités des lois de $X_{(1)}$ et $X_{(n)}$ et déterminer la densité du couple $(X_{(1)}, X_{(n)})$. Quelle est la loi de $W_n = X_{(n)} - X_{(1)}$?
6. On considère une suite $(U_i)_{i \in \mathbb{N}}$ de variables i.i.d. selon la loi uniforme sur $[0, 1]$, et on pose
$$Y_n = \min_{1 \leq i \leq n} U_i \qquad Z_n = \max_{1 \leq i \leq n} U_i - \min_{1 \leq i \leq n} U_i$$
   (a) Montrer que nY_n converge en loi vers une loi exponentielle.
   (b) Étudier la convergence en loi de Z_n, puis sa convergence en probabilité et $\mathbb{L}^1$.
   (c) Soit $\epsilon > 0$. Calculer $\mathbb{P}[|Z_n - 1| > \epsilon]$. En déduire que Z_n converge presque sûrement.
   (d) Rappeler les implications logiques entre les modes de convergence étudiés : en loi, en probabilité, en norme $\mathbb{L}^1$, en norme $\mathbb{L}^2$, presque sûre.

Exercice 2 : On reprend un exemple du cours. La limite légale d'un polluant contenu dans les déchets d'une usine est de 6mg/kg. On effectue un dosage sur 20 prélèvements sur lesquels on observe une moyenne empirique de 7mg/kg avec un écart-type empirique de 2.4mg/kg. On admet que la loi de dosage est gaussienne.

On observe donc $X_1, \dots, X_{20} \overset{iid}{\sim} N(\mu, \sigma^2)$ avec μ et σ^2 inconnus.

1. Faire un test de niveau α pour le problème de test suivant :
$$H_0 : \mu \leq 6 \quad \text{contre} \quad H_1 : \mu > 6.$$
2. On calcule à partir de ces données $\bar{x} = 7$ et $\hat{\sigma}^2 = \frac{1}{19}\sum_{i=1}^{20}(x_i - \bar{x})^2 = 2.4^2$. Calculer la p-valeur observée et conclure si on choisit le niveau $\alpha = 5\%$.

Exercice 3 : On dispose d'un échantillon de loi Bernoulli de paramètre p : $X_1, \dots, X_n \overset{iid}{\sim} Be(p)$.

1. Proposer une procédure de test pour le problème suivant
$$H_0 : p = 1/2 \quad \text{contre} \quad H_1 : p > 1/2.$$
2. Proposer une procédure de test pour le problème suivant
$$H_0 : p = 1/2 \quad \text{contre} \quad H_1 : p \neq 1/2$$
3. Proposer un test asymptotique pour le problème de la question précédente.
4. Calculer la puissance du test asymptotique de la question 3. La puissance tend-elle simplement vers 1 quand n tend vers l'infini ?
5. Application numérique. On calcule à l'aide des données, $n = 100$, $\bar{x} = 0.59$, $q_{0.95} = 58$, $q_{0.975} = 60$ où on note q_α le quantile d'ordre α de la loi binomiale $B(100, 1/2)$. Quelle est la conclusion des deux premiers tests ci-dessus, au niveau $\alpha = 0.05$, pour ces données ?