Statistiques décisionnelles : TD 1

Exercice 1.

1. *Vous vous promenez le long d'une voie ferrée, en pleine rêverie. Vous prenez soudain conscience d'un grand bruit, droit derrière vous. En quoi consiste l'erreur de première espèce ? L'erreur de deuxième espèce ? Laquelle est la plus grave ? Que faites-vous ?*
2. *Vrai ou faux : si on compare un test d'hypothèses à un jugement, on peut dire que l'erreur de première espèce est de condamner un homme coupable.*
3. *Dans l'analogie du jugement, que se passerait-il, si on souhaitait rendre l'erreur de première espèce nulle ?*
4. *Dans l'analogie du jugement, comment interpréter la puissance du test ?*

Définition 1. *Si $X_1, \ldots, X_n$ sont des variables aléatoires, on note $X_{(1)}, \ldots, X_{(n)}$ l'échantillon ordonné croissant des X_i. Ainsi, $X_{(1)}$ est la plus petite valeur prise par les X_i, $X_{(2)}$ la seconde plus petite valeur et ainsi de suite jusqu'à $X_{(n)}$ qui est la plus grande valeur des X_i. Cette notation avec des parenthèses est standard en probabilités et statistiques. Attention, les X_i sont i.i.d. mais les $X_{(i)}$ ne sont ni indépendants, ni identiquement distribués. Gardez en tête que $X_{(i)}$ dépend de toutes les variables $X_1, \ldots, X_n$.*

Exercice 2 (Rappels sur les fonctions de répartitions)**.** *Soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires i.i.d. de loi inconnue.*

1. *Rappeler la définition de la fonction de répartition des X_i, notée F_X, ainsi que la fonction de répartition empirique, notée F_n.*
2. *Représenter graphiquement F_X et F_n lorsque les X_i suivent une loi uniforme sur $[0, 1]$ et où on a obtenu les valeurs*
$$X_1 = 0.17 \quad X_2 = 0.66 \quad X_3 = 0.42 \quad X_4 = 0.62, \quad X_5 = 0.41.$$
Sur ce graphe, que vaut $\sup |F_X(t) - F_n(t)|$ *?*
3. *Représenter graphiquement F_X et F_n lorsque les X_i suivent une loi $\mathcal{E}(1)$ et où on a obtenu les valeurs*
$$X_1 = 0.33 \quad X_2 = 0.98 \quad X_3 = 0.53 \quad X_4 = 0.22, \quad X_5 = 0.39.$$
Sur ce graphe, que vaut $\sup |F_X(t) - F_n(t)|$ *?*
4. *Montrer que, quel que soit $t \in \mathbb{R}$, la suite $(F_n(t))_{n\geq 1}$ converge presque sûrement vers $F_X(t)$.*
5. *Donner un intervalle de confiance asymptotique pour $F_X(t)$ de niveau* 95%.
6. *Rappeler le théorème de Glivenko-Cantelli. Quelle est la différence avec ce qui a été montré précédemment ? Comment calculez-vous en pratique* $\sup_{t\in\mathbb{R}} |F_X(t)- F_n(t)|$ *? On pourra ordonner les variables X_i de manière croissante en notant $X_{(i)}$ la i-ème plus grande valeur dans l'échantillon $X_1, \ldots, X_n$.*

Exercice 3. *Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de loi $\mathcal{E}(\lambda)$.*

1. *Si U est une variable aléatoire de loi uniforme sur $[0,1]$, calculer la fonction de répartition de $-\ln(1-U)$ et en déduire la loi de cette variable. Quel théorème du cours est illustré ici ?*
2. *Si X est une variable aléatoire de loi $\mathcal{E}(1)$, quelle est la loi de $\frac{1}{\lambda}X$? En déduire la loi de $\frac{-1}{\lambda}\ln(1-U)$ si U est une variable aléatoire de loi uniforme sur $[0,1]$.*
3. *Si X est une variable aléatoire de loi $\mathcal{E}(1)$, quelle est la loi de $1-e^{-X}$? Bonus : Si F_X est inversible, quelle est la loi de $F_X(X)$?*
4. *Soient $U_1,\dots,U_n$ des variables aléatoires i.i.d. uniformes sur $[0,1]$. On définit les Y_i par $Y_i = -\frac{1}{\lambda}\ln(1-U_i)$. Quelle est la relation entre les variables ordonnées $Y_{(1)},\dots,Y_{(n)}$ et $U_{(1)},\dots,U_{(n)}$?*
5. *Pour différentes valeurs de n, on cherche à comprendre la loi de la variable aléatoire $\sup_{t\in\mathbb{R}}|F_X(t)-F_n(t)|$. En vous inspirant de l'exercice précédent, écrivez cette quantité comme un maximum d'un nombre fini de quantités mettant en jeu les $X_{(i)}$*
6. *Si X suit une lui exponentielle de paramètre 1 et $n=1$, explicitez la variable aléatoire $\sup_{t\in\mathbb{R}}|F_X(t)-F_n(t)|$. Même question si X suit une loi exponentielle de paramètre λ.*
7. *Même question si U suit une loi uniforme sur $[0,1]$.*
8. *Même question si $Y = -\frac{1}{\lambda}\ln(1-U)$ où U suit une loi uniforme sur $[0,1]$. Que remarquez-vous ? Reliez cette question aux deux précédentes.*
9. *Recommencez la question précédente pour $n=2$ puis n quelconque. Après réflexion, de quoi dépend la variable $\sup_{t\in\mathbb{R}}|F_X(t)-F_n(t)|$?*
10. *Comment feriez-vous pour connaître approximativement les propriétés de cette variable pour $n=1$, $n=5$? (à l'aide d'un ordinateur et de R).*