Licence STE 1 **$1^{\text{ère}}$ session, Semestre 2, Mai 2005**

Epreuve de Probabilité et Statistiques
Durée 3h, Documents non autorisés, calculatrice autorisée

Il sera tenu compte de la clarté de la rédaction dans la notation finale. Lisez bien l'énoncé recto-verso avant de commencer

Penser à rendre vos graphiques dans la copie

Exercice n° 1

On a mesuré à une station météorologique du Sud de la France des intensités de vent (m/s). La série observée correspond aux intensités mesurées pour des vents de secteurs N-NW, en moyenne horaire. On dispose de $N = 1000$ observations consignées par classes de largeur $h = 1$ dans le tableau ci-dessous:

classe C_i	[0; 1[	[1; 2[	[2; 3[	[3; 4[	[4; 5[	[5; 6[	[6; 7[	[7; 8[	[8; 9[	[9; 10[	[10; 11[
effectif N_i	96	177	203	159	112	77	52	40	31	19	16

classe C_i	[11; 12[	[12; 13[	[13; 14[	[14; 15[	[15; 16[
effectif N_i	6	2	4	5	1

1. Quel est le type de la variable?
2. Tracer l'histogramme de cette distribution groupée sur le graphique réservé à cet effet (2ème feuillet).
3. Tracer l'histogramme en fréquences cumulées. En déduire le tracé de la fonction de répartition empirique.
4. Calculer la moyenne et la variance empirique de la distribution.
5. Tracer le boxplot de cette distribution. Que peut-on en déduire?

Les moustaches du boxplot seront fixées aux quantiles $x_{0.1}$ et $x_{0.9}$.

Correction du premier exercice

1. La variable X est une intensité de vent (m/s). Elle est à valeur dans $\mathbb{R}^+$. C'est donc une variable quantitative continue.
2. Pour tracer l'histogramme, on a besoin de connaître les fréquences $f_j = \frac{n_j}{n}$ de chaque classe j ainsi que la largeur h_j de ces classes. On obtient ainsi la figure 1. Les valeurs qui sont données au-dessus de chaque barre correspondent aux fréquences. La somme des surfaces grises est égale à 1 puisque le $h_j = 1, \ \forall j$.
3. Pour tracer le diagramme en fréquences cumulées on a besoin des $F_j = \frac{N_j}{n}$. Elles sont indiquées sur le graphique, également au-dessus des barres.

   La fonction de répartition empirique (FRE) telle qu'elle a été introduite en cours possède un graphe qui correspond à la courbe bleue sur cette même figure. Comme il s'agit d'une variable continue, la FRE est une fonction continue, linéaire par morceaux. Elle permet de donner la proportion d'observations inférieures à une valeur x fixée et notée :

$$F_n(x) = Prop(X \leq x). \tag{1}$$

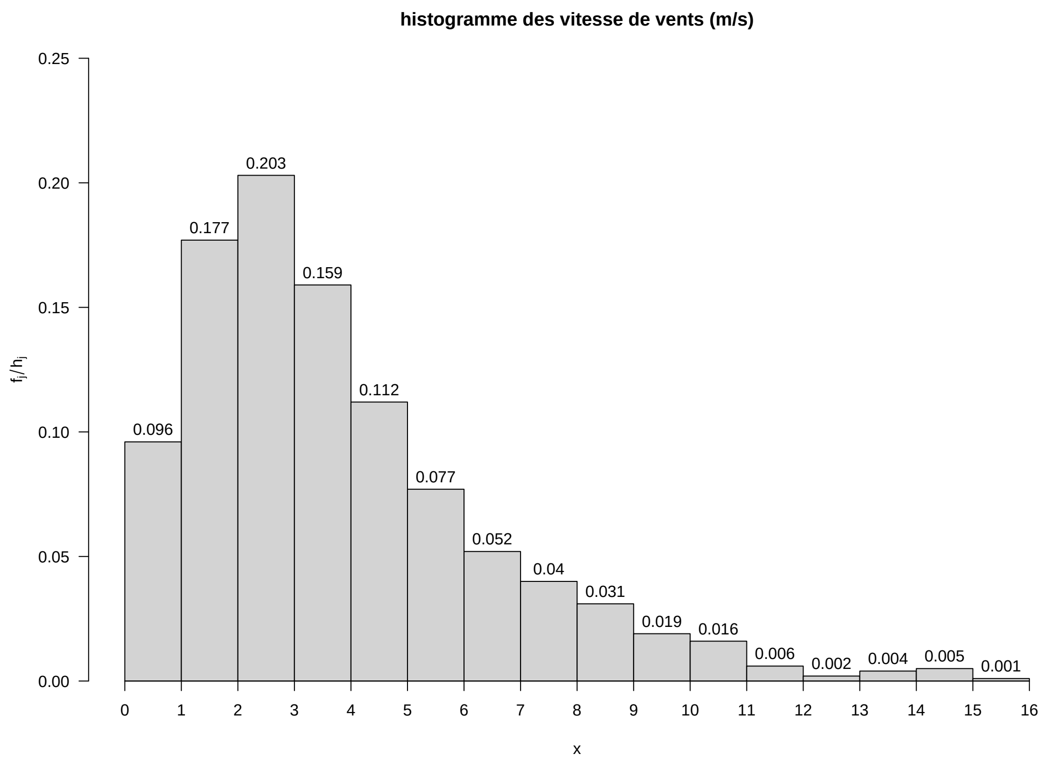

Figure 1:

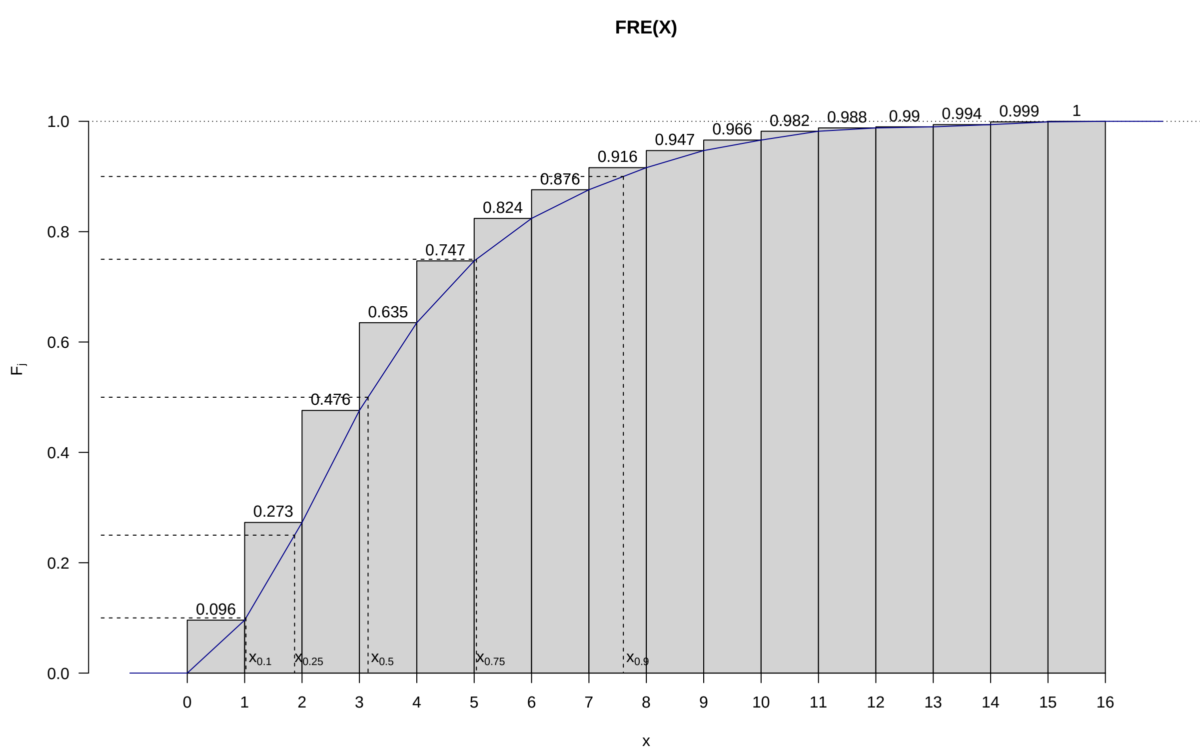

Figure 2:

4. On travaille ici avec une distribution groupée. On n'a donc pas accès à la série observée. Pour calculer la moyenne $\overline{x}$ et la variance s_n^2 empirique, nous allons utiliser les centres de classe notés x_{c_j} ainsi que les fréquences f_j. On obtient ainsi :

$$\overline{x} = \sum_{j=1}^{J}(x_{c_j} \times f_j) = 0.5 \times 0.096 + \cdots + 15.5 \times 0.001 = 3.791 \ m/s, \tag{2}$$

et pour la variance :

$$s_n^2 = \sum_{j=1}^{J}(x_{c_j}^2 \times f_j) - \overline{x}^2 = 0.5^2 \times 0.096 + \cdots + 15.5^2 \times 0.001 - 3.791^2 = 7.156 \ m^2/s^2. \tag{3}$$

5. Le boxplot est tracé en déduisant, avec le graphique 2 de la FRE, les différents quantiles. Il est ici présenté verticalement mais vous pouvez pour des raisons pratiques le présenter horizontalement. Pour le corps du boxplot, on a besoin de $x_{0.25}$ et $x_{0.75}$. Pour les moustaches, on vous indique de prendre $x_{0.1}$ et $x_{0.9}$. La barre centrale est donnéee par la médiane $x_{0.5}$. Pour déterminer ces valeurs, on part de l'axe des ordonnées de la FRE (Figure 2) disons à la valeur $p = 0.5$ et on en déduit, par image inverse, la valeur $x_{p=0.5}$ sur l'axe des abcisses (ici la médiane $x_{0.5} = 3.1$ environ). On fait de même pour les autres quantiles et on peut tracer le graphique 3, sur lequel apparaissent les valeurs des quantiles, déterminées de manière précise. Une estimation visuelle serait suffisante pour avoir un boxplot avec la même allure. On en déduit que la distribution des valeurs de vent est dissymétrique vers les basses valeurs. On aurait pu le constater en remarquant que $\overline{x} > x_{0.5}$. Avec une distribution symétrique, ces deux valeurs seraient identiques.

Exercice n° 2

Soit X une variable aléatoire de densité donnée par :

$$f_X(x) = \begin{cases} kx \exp\left(-\dfrac{x}{\alpha}\right) \text{ si } x \geq 0 \\ 0 \text{ sinon} \end{cases}$$

où α est un paramètre fixé positif et k un paramètre à déterminer.

1. Quel est le type de la variable?
2. Déterminer k en fonction de α pour que $f_X(x)$ définisse une densité de probabilité.
3. Calculer $E(X)$ et $V(X)$.
4. Déterminer la fonction de répartition F_X de X. Que représente les graphiques de la figure ci-dessous?

En fait, on cherche à savoir si la distribution empirique de l'exercice 1 peut être ajustée à l'aide de la densité ci-dessus.

1. Proposer une valeur estimée $\widehat{\alpha}$ de α qui permet d'ajuster la distribution théorique aux données de l'exercice 1.

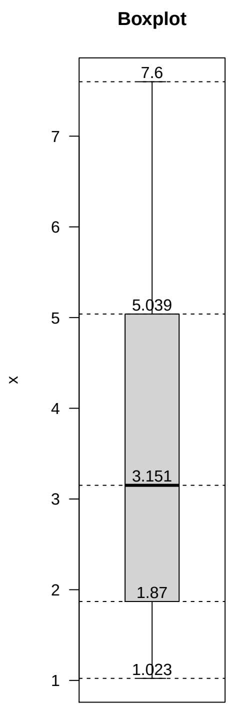

Figure 3:

2. Déterminer graphiquement les déciles d_i^* de la distribution empirique et les déciles d_i de la distribution théorique. Que représente d_5?

3. Tracer les couples $\{(d_i, d_i^*), i = 1, ..., 9\}$ sur le graphique ci-dessous. Que représente ce graphique? Que peut-on en déduire?

On rappelle que $\int_a^b u dv = [uv]_a^b - \int_a^b v du$