Licence SVT $1^{ème}$ année **Partiel I Statistiques & Probabilités**

Partiel, Mars 2018

Aucun document autorisé, calculatrice autorisée, durée : 02h00

Exercice 1.

La trajectométrie est une des branches des sciences de l'environnement qui s'intéresse à l'étude du déplacement d'individus au sein d'une population étudiée. La miniaturisation des appareils électroniques, l'autonomie croissante des batteries, permettent depuis une dizaine d'années d'équiper de petits animaux et de réaliser des expériences de suivi inimaginables seulement dix ans en arrière. Voici l'exemple du suivi d'un fou du Cap *Morus capensis* réalisé à Port Elysabeth en 2014 à l'aide d'un GPS embarqué et d'un accéléromètre permettant d'avoir respectivement accés à la position de l'animal et à sa vitesse.

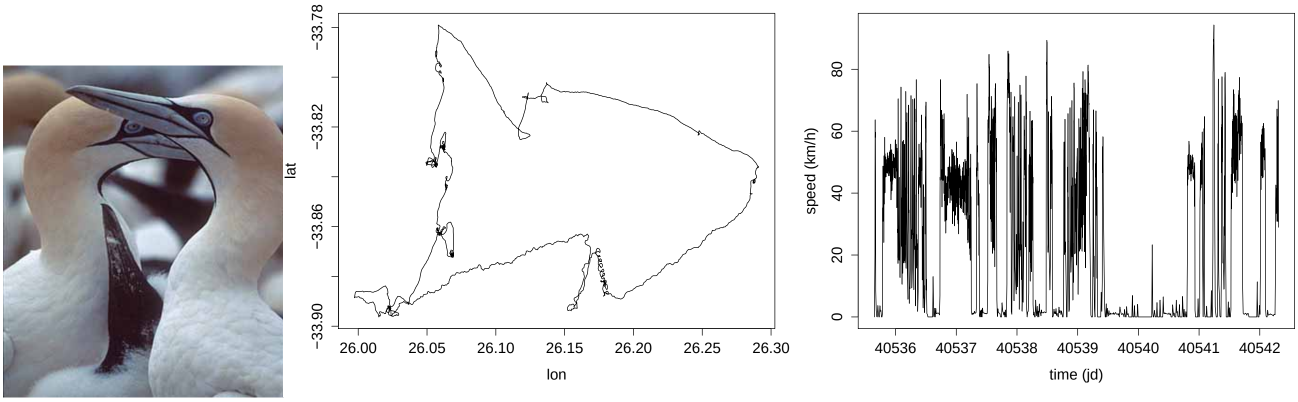

Figure 1: Mouvement d'un oiseau marin équipé en Afrique du Sud (trajectoire et vitesse).

A partir de $n = 200$ trajectoires d'oiseaux, on a compter le nombre de jours qu'un oiseau passe en mer. Cette variable est notée X et voici les observations rangées :

2 2 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 11 11 11 11 11 11
11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 13 13 13 13 13 13 14 14 14 14

1. Quel est le type de la variable X? Donner son domaine de variation et son unité.
2. Donner les x_j, n_j, N_j, f_j, F_j.
3. Construire la distribution observée et la fonction de répartition associée.
4. Donner la médiane en utilisant les données brutes. Calculer la moyenne de cette distribution :
   (a) En utilisant les données brutes (on notera $\overline{x}_1$ le résultat)
   (b) En utilisant les f_j (on notera $\overline{x}_2$ le résultat)
5. En détaillant vos calculs, donner une estimation de la variance empirique s^2 en utilisant les valeurs approchées suivantes :
   $\sum_{i=1}^{n} x_i^2 \simeq 13000$, $\left(\sum_{i=1}^{n} x_i\right)^2 \simeq 2.28 \times 10^6$.
6. Construire la boîte à moustache de cette distribution (on prendra comme valeurs des moustaches les quantiles $x_{0.1}$ et $x_{0.9}$). Tous les quantiles seront calculés avec la distribution observée.

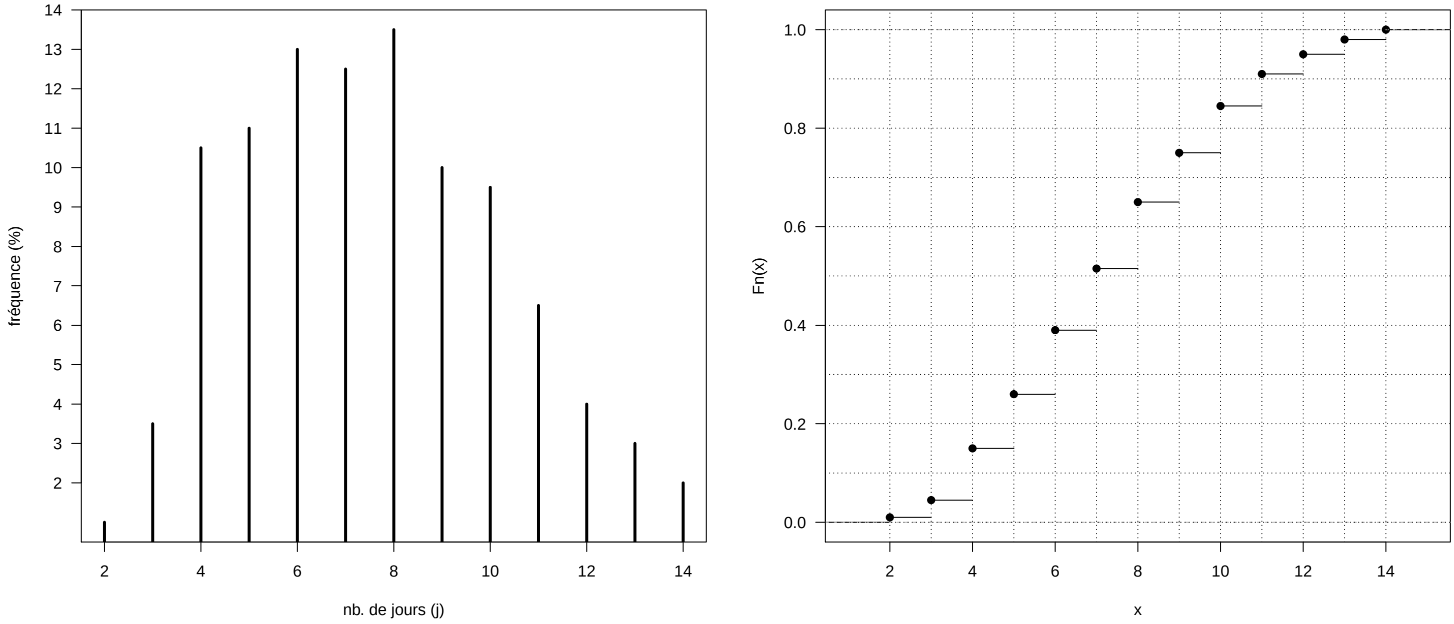

Figure 2: Distribution et fonction de répartition de la variable X

Correction

Exercice 1.

1. La variable est un nombre de jour : il s'agit donc d'une variable discrète. Son unité est le jour (j) et $X \in \mathbb{N}$.

2. Le tableau ci-dessous résume les quantités demandées soit, le nombre de modalités $J = 12$, le nombre d'observation $n = 200$, les modalités x_j, les effectifs n_j, les effectifs cumulés N_j, les fréquences f_j et les fréquences cumulées F_j :

x_j	n_j	N_j	f_j	F_j
2	2	2	0.010	0.010
3	7	9	0.035	0.045
4	21	30	0.105	0.150
5	22	52	0.110	0.260
6	26	78	0.130	0.390
7	25	103	0.125	0.515
8	27	130	0.135	0.650
9	20	150	0.100	0.750
10	19	169	0.095	0.845
11	13	182	0.065	0.910
12	8	190	0.040	0.950
13	6	196	0.030	0.980
14	4	200	0.020	1

3. La courbe gauche de la figure 2. représente la distribution des valeurs de X. Il s'agit d'un diagramme en bâtons car la variable est discrète. Le graphique de droite est la fonction de répartition empirique, en marches d'escalier, caractéristique d'une variable discrète.

4. Soit $x_{0.5}$, la médiane de X. En utilisant les données brutes, elle est telle que :

$$x_{0.5} = \frac{x_{(n/2)} + x_{(n/2-1)}}{2} = 7.$$

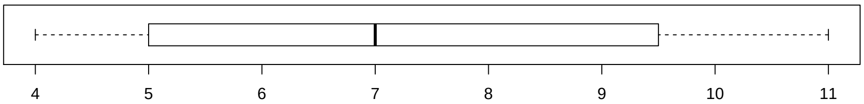

Figure 3: Boxplot des réalisations de X

Elle est calculée de cette manière car n est pair.
Le calcul de la moyenne sur les données brutes est donnée par

$$\overline{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1509}{200} = 7.545.$$

En utilisant les fréquences, on aura :

$$\overline{x}_2 = \sum_{j=1}^{J} f_j x_j = 7.545.$$

Evidemment, on obtient le même résultat car les données ne sont pas mises en classe (variable discrète).

5. Le calcul approximatif de la variance est donné par :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \simeq \frac{13000}{200} - \frac{2.28 \times 10^6}{200^2} = 65 - 57 = 8.$$

6. Pour tracer le boxplot, on a besoin de l'intervalle interquartile $I = x_{0.75} - x_{0.25} = 9.5 - 5 = 4.5$, de la médiane $x_{0.5} = 7$, des quantiles $x_{0.1} = 4$ et $x_{0.9} = 11$. Tous les quantiles sont estimés graphiquement. Le quantile $x_{0.75}$ tombant sur la marche 0.75 de la fonction de répartition, il n'est pas unique : on prend le centre de l'intervalle quantile $[9; 10]$, soit $x_{0.75} = 9.5$. La figure 3. représente le boxplot associé à ces valeurs.