

Partiel, Février 2017

Aucun document autorisé, calculatrice autorisée, durée : 01h30

L'acoustique est une méthode de mesure utilisée en océanographie pour tenter de quantifier la biomasse d'organismes vivants dans la colonne d'eau. A partir de la marche d'un bateau océanographique effectuée à vitesse constante, une onde acoustique est émise à une certaine fréquence (Fig. 1, 18 kHz). Sur une hauteur d'eau variable, l'écho de certaines espèces animales est détecté, la qualité du signal reçu dépendant, entre autre, de la position des individus dans la colonne d'eau mais également de leur composition (carapace chitineuse des crevettes, par exemple).

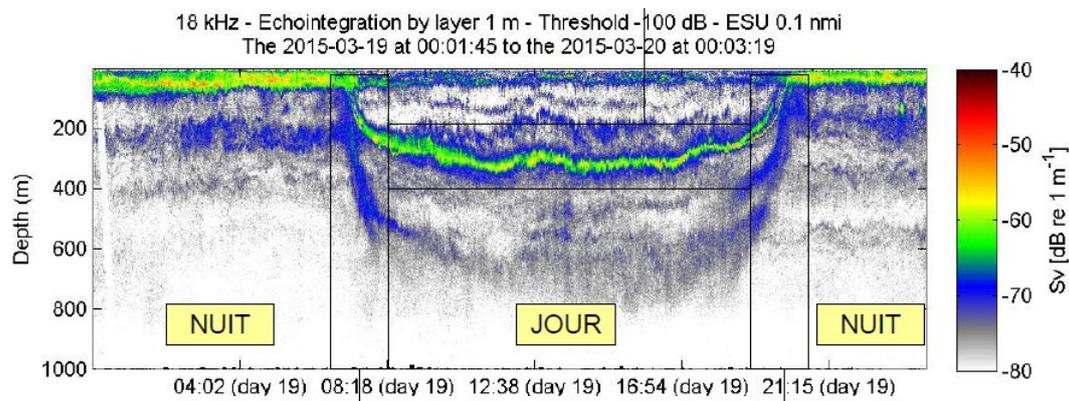


Figure 1: Signal acoustique reçu au cours d'un transect d'une campagne réalisée dans l'Océan Indien.

On s'intéresse à $n = 100$ mesures de la profondeur qui correspond au maximum de la densité de biomasse observée (bande verte sur la fig. 1) sur une journée entre la surface et 400 m de profondeur. Cette variable est notée X , elle est mesurée en m et les valeurs mesurées ont été arrondies au mètre. Voici les relevés rangés de cette variable :

0 2 3 4 5 6 7 7 8 9 10 11 18 19 22 27 31 31 35 41 43 44 52 53 56
 67 70 73 82 84 87 88 91 94 94 98 102 105 106 114 114 134 148 149 149 152 156 158 159 162
 175 175 180 193 207 233 240 249 271 273 282 290 296 305 315 316 318 321 324 324 324 325 325 327 334
 336 339 340 349 350 353 371 372 374 381 386 391 391 392 393 395 397 397 397 399 400 400 400 400 400

1. Quel est le type de la variable X ? Donner son domaine de variation.
2. On fabriquera 8 classes de largeur 50 m, de 0 m à 400 m. Donner les x_{c_j} , n_j , f_j , F_j , h_j .
3. Construire un histogramme 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 \bar{x}_1 le résultat)
 - (b) En utilisant les données mises en classes (on notera \bar{x}_2 le résultat)
5. Commenter les résultats précédents et donner la variance s_1^2 et l'écart type en utilisant les données brutes.
6. Construire la boîte à moustache de cette distribution (on prendra comme valeurs des moustaches les quantiles $x_{0.1}$ et $x_{0.9}$).

On donne : $\sum_{i=1}^n x_i = 19805$, $\sum_{i=1}^n x_i^2 \simeq 6 \times 10^6$, $(\sum_{i=1}^n x_i)^2 \simeq 3.9 \times 10^8$.

Correction

1. La variable X est une variable réelle. Il s'agit d'une profondeur donc $X \in [0; 400] \subset \mathbb{R}$, même si les valeurs échantillonnées ont été discrétisées.

c_j	x_{c_j}	n_j	$f_j = \frac{n_j}{n}$	F_j	h_j
[0;50[25	22	0.22	0.22	50
[50; 100[75	14	0.14	0.36	50
[100; 150[125	9	0.09	0.45	50
2. [150; 200[175	9	0.09	0.54	50
[200; 250[225	4	0.04	0.59	50
[250; 300[275	5	0.05	0.63	50
[300; 350[325	16	0.16	0.79	50
[350; 400[375	21	0.21	1	50

3. Les barres de l'histogramme ont une hauteur de $\frac{f_j}{h_j}$. De cette manière, il est un estimateur de la densité et la somme des surfaces de tous les rectangles est égale à 1.

4. La médiane $x_{0.5}$, une des caractéristiques centrales d'une distribution, correspond à la moyenne entre la 50^{ème} et la 51^{ème} valeur de la série observée, soit $x_{0.5} = \frac{162+175}{2} = 168.5$. Avec les données brutes, la moyenne $\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{100} \times 19805 = 198.05 m$. Avec les données en J classes, $\bar{x}_2 = \sum_{j=1}^J f_j x_{c_j} = (0.22 \times 25 + \dots + 0.21 \times 375) = 196.5 m$.

5. On s'aperçoit que les deux valeurs sont différentes. Ceci est dû au fait que, quand les données sont groupées en classes, on perd une partie de l'information qui permet d'estimer la moyenne empirique. Mieux vaut en fait travailler directement sur les données brutes. Une estimation de la variance est donnée grâce aux valeurs numériques fournies dans l'énoncé et $s_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_1)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_1^2 \simeq \frac{1}{100} \times 6 \times 10^6 - \frac{1}{10000} \times 3.9 \times 10^8 = 21000 m^2$.

6. Pour construire le boxplot on a besoin d'estimer les quantiles empiriques $x_{0.1}$, $x_{0.25}$, $x_{0.5}$, $x_{0.75}$ et $x_{0.9}$. Pour cela, on peut, soit utiliser les données brutes, soit la fonction de répartition empirique. C'est ce qui est fait ci-dessous.

En utilisant les données brutes, on obtient : $x_{0.1} = 9.5$, $x_{0.25} = 61.5$, $x_{0.5} = 168.5$, $x_{0.75} = 335$ et $x_{0.9} = 394$. On se rend compte que les estimations peuvent varier, d'une méthode à l'autre. Ci-dessous, le boxplot construit à partir des quantiles sur données brutes :

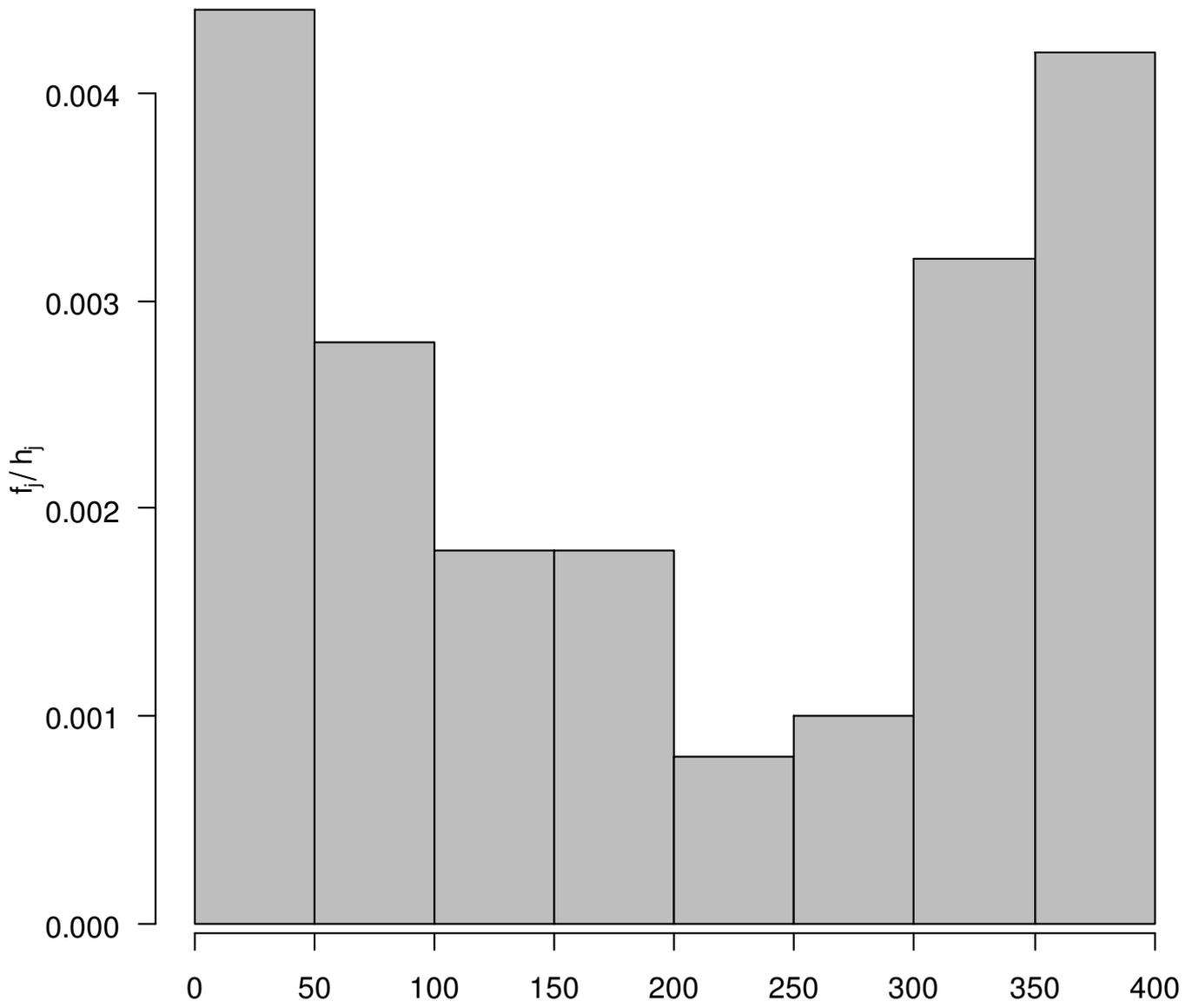


Figure 2:

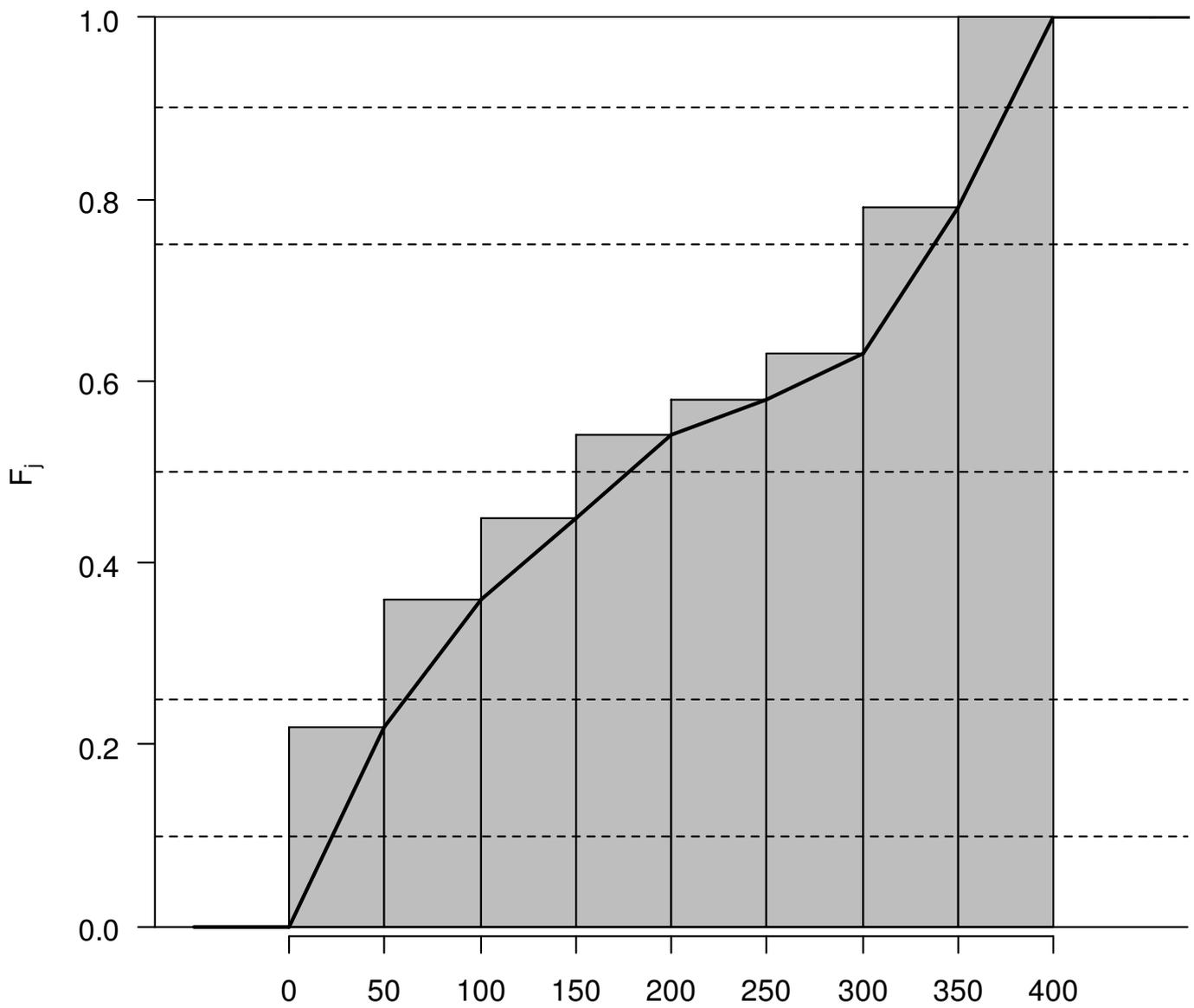


Figure 3: Les quantiles sont estimés en prenant l'intersection de chaque ligne pointillée avec le graphe de la FRE (ligne épaisse).

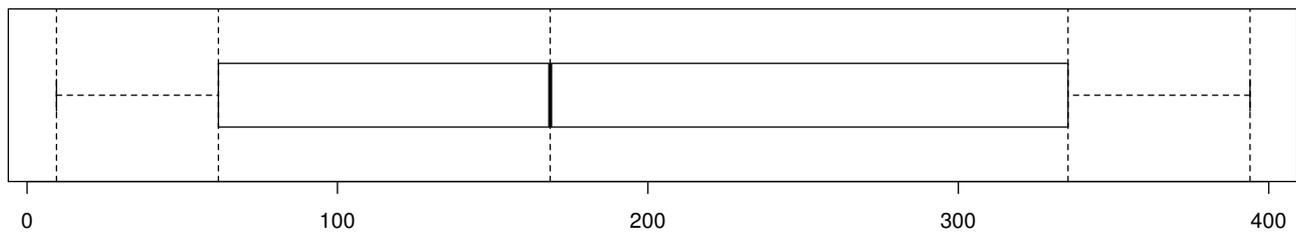


Figure 4: