Semaine 3

***Exercice 21.** Les hypothèses impliquent que

$$G(t+s) = G(t)G(s), \qquad \forall t, s \geq 0,$$

au moins lorsque $G(t) > 0$. Or, si $G(t) = 0$ l'égalité est évidante, car G est décroissante et nonnegative.
En termes de $g(x) = -\ln G(x)$, cette égalité s'écrit

$$g(t+s) = g(t) + g(s), \qquad \forall t, s \geq 0.$$

(À noter que cette égalité tient, et a un sens, même si $g = \infty$, puisque $g(x) \in [0, \infty]$ pour chaque $x \geq 0$.)
Soit $\lambda = g(1)$, alors $g(2) = 2\lambda$ et par récurrence $g(n) = n\lambda$ pour n entier. Par récurrence encore $g\left(\frac{k}{n}\right) = kg\left(\frac{1}{n}\right)$ pour des entiers n, k. En posant $k = n$ nous obtenons $\lambda = g(1) = ng\left(\frac{1}{n}\right)$, et donc $g\left(\frac{k}{n}\right) = \frac{k}{n}\lambda$, c'est-à-dire que $g(q) = q\lambda$ pour chaque $q > 0$ rationnel. Pour $t > 0$ réel, prenons une suite de rationnels $q_n \searrow t$. En utilisant la continuité à droite de g (qui résulte de celle de G),

$$g(t) = \lim_{n\to\infty} g(q_n) = \lim_{n\to\infty} q_n\lambda = t\lambda.$$

(Nous aurions pu utiliser le fait que G, et par conséquent g, est monotone, sans utiliser la continuité à droite.)
Ainsi $G(t) = \exp(-t\lambda)$ pour chaque t. Puisque $G(t) \to 0$ lorsque $t \to \infty$, forcément $\lambda > 0$ et la fonction qui vaut 0 pour $t < 0$ et $1 - G(t)$ pour $t \geq 0$ est bien la fonction de répartition d'une variable aléatoire exponentielle de paramètre λ. Il est impossible que $\lambda = \infty$, puisque G est continue à droite et $G(0) > 0$.
Remarque 1. Nous n'avons même pas supposé ni que X soit une variable aléatoire continue, ni que $\mathbb{P}(X \geq 0) = 1$!
Remarque 2. Il existe des fonctions « sans mémoire » qui ne sont pas de la forme $G(t) = e^{-\lambda t}$. Ces fonctions, évidemment, ne sont pas continues à droite ni monotones. Leur existence requiert une base de $\mathbb{R}$ sur $\mathbb{Q}$ dont la construction nécessite (une version faible de) l'axiome du choix.

Exercice 22.

(i) La dérivée de f est

$$\frac{d}{d\gamma} f(\gamma) = -2\sum_{i=1}^{n}(x_i - \gamma).$$

En la mettant égale à zéro, on trouve

$$\sum_{i=1}^{n} x_i - n\gamma = 0 \Rightarrow \gamma = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

Puisque $f''(\gamma) = 2n > 0$, $\bar{x}$ est le minimum global de f.

(ii) On peut écrire

$$g(\gamma) = \sum_{i=1}^{n} |x_i - \gamma| = \sum_{i=1}^{n} |x_{(i)} - \gamma|.$$

La fonction g est dérivable pour chaque $\gamma \in \mathbb{R}\backslash\{x_{(1)}, \dots, x_{(n)}\}$.

— Quand $\gamma \in (-\infty, x_{(1)})$, on a $g(\gamma) = \sum_{i=1}^{n}(x_{(i)} - \gamma)$ et donc $g'(\gamma) = \sum_{i=1}^{n} -1 = -n$.
— Quand $\gamma \in (x_{(n)}, \infty)$, on a $g(\gamma) = \sum_{i=1}^{n} -(x_{(i)} - \gamma)$ et donc $g'(\gamma) = \sum_{i=1}^{n} 1 = n$.
— Quand $\gamma \in (x_{(j)}, x_{(j+1)})$, $j = 1, \dots, n-1$, on a

$$g(\gamma) = \sum_{i=1}^{j} -(x_{(i)} - \gamma) + \sum_{i=j+1}^{n} (x_{(i)} - \gamma)$$

et donc $g'(\gamma) = \sum_{i=1}^{j} 1 + \sum_{i=j+1}^{n} -1 = j - (n - j) = 2j - n$.

Distinguons les deux cas suivants :

1. n pair :
   — $g'(\gamma) < 0$ quand $\gamma \in (-\infty, x_{(1)})$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = 1, \dots, \frac{n}{2} - 1$.
   — $g'(\gamma) = 0$ quand $\gamma \in (x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)})$.
   — $g'(\gamma) > 0$ quand $\gamma \in (x_{(n)}, \infty)$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = \frac{n}{2} + 1, \dots, n-1$.
   Puisque g est continue, chaque point en $[x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}]$ est un minimum de g et en particulier $M = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$ est un minimum.
2. n impair :
   — $g'(\gamma) < 0$ quand $\gamma \in (-\infty, x_{(1)})$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = 1, \dots, \frac{n+1}{2} - 1$.
   — $g'(\gamma) > 0$ quand $\gamma \in (x_{(n)}, \infty)$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = \frac{n+1}{2}, \dots, n-1$.
   Puisque g est continue, $M = x_{(\frac{n+1}{2})}$ est l'unique minimum de g.

Remarque : il est possible que $x_{(k)} = x_{(k+1)}$ pour un certain k (c'est-à-dire qu'on observe la même valeur plusieurs fois), mais la preuve reste valide même dans ce cas.

Exercice 23.

(i) Nous obtenons $\bar{x} = 10.24$ et $M = 10.05$.

(ii) Maintenant nous obtenons $\bar{x} = 13.89$ et $M = 10.05$.

(iii) On observe que dans la partie (i) les valeurs de $\bar{x}$ et de M sont similaires, tandis que dans la partie (ii) la valeur de $\bar{x}$ a beaucoup changé à cause de la valeur atypique 48.6. En même temps, la valeur de M n'a pas changé. On note que la moyenne $\bar{x}$ est plus susceptible aux valeurs aberrantes que la médiane M. En fait, dans la partie (ii), $\bar{x}$ est plus grande que chaque observation sauf la valeur extrême 48.6. À cause de cette valeur, la moyenne n'est pas un très bon résumé de la position de cet échantillon. En revanche, la médiane n'est pas affectée par cette valeur extrême.

Exercice 24. Considerons l'échantillon :

$$-2, -2, 0, 1, 3$$

La moyenne et la médiane sont égal 0, mais l'échantillon n'est pas symétrique autour de 0.
Remarque pour ceux qui ont besoin d'une définition mathématique formelle de la symétrie. L'échantillon $x_1, \dots, x_n$ s'appelle symétrique autour de $a \in \mathbb{R}$, si

$$\{x_1, \dots, x_n\} = \{-(x_1 - a) + a, \dots, -(x_n - a) + a\}.$$

L'égalité est comprise comme l'égalité des ensembles.

Exercice 25. Nous écrivons :

$$\begin{aligned} n\widehat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2\sum_{i=1}^n x_i\bar{x} \\ &= \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2\bar{x}\sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2. \end{aligned}$$

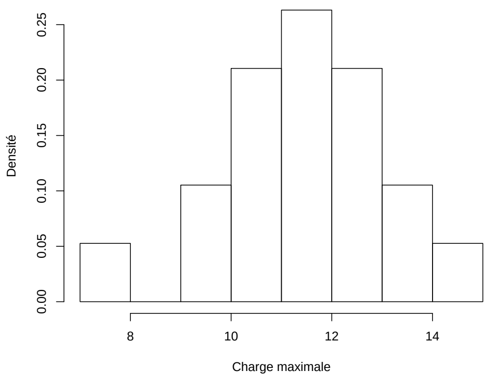
Cette formule est plus pratique, car elle demande de calculer les carrés de $n+1$ nombres et une différence, au lieu de devoir calculer n différences, et puis n carrés, comme dans la formule originale.

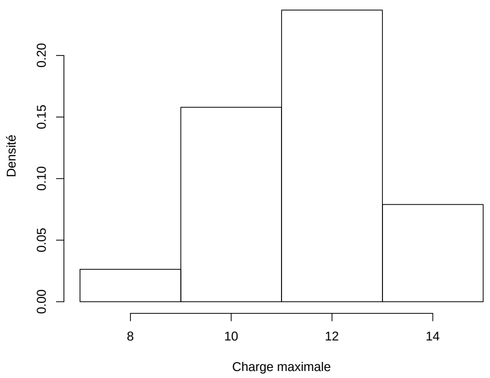
Exercice 26. Si $n = 12$ alors $M = (x_{(6)} + x_{(7)})/2$, $Q_1 = x_{(4)}$ et $Q_3 = x_{(9)}$.
Si $n = 13$ alors $M = x_{(7)}$, $Q_1 = x_{(4)}$ et $Q_3 = x_{(10)}$.
Si $n = 14$ alors $M = (x_{(7)} + x_{(8)})/2$, $Q_1 = (x_{(4)} + x_{(5)})/2$ et $Q_3 = (x_{(10)} + x_{(11)})/2$.
Si $n = 15$ alors $M = x_{(8)}$, $Q_1 = (x_{(4)} + x_{(5)})/2$ et $Q_3 = (x_{(11)} + x_{(12)})/2$.
Pour n quelconque, on obtient les formules

$$Q_1 = \begin{cases} x_{\left(\frac{n}{4}+1\right)} & n \equiv 0 \mod 4 \\ x_{\left(\frac{n-1}{4}+1\right)} & n \equiv 1 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{n-2}{4}+1\right)} + x_{\left(\frac{n-2}{4}+2\right)}\right) & n \equiv 2 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{n-3}{4}+1\right)} + x_{\left(\frac{n-3}{4}+2\right)}\right) & n \equiv 3 \mod 4, \end{cases} \qquad Q_3 = \begin{cases} x_{\left(\frac{3n}{4}\right)} & n \equiv 0 \mod 4 \\ x_{\left(\frac{3(n-1)}{4}+1\right)} & n \equiv 1 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{3(n-2)}{4}+1\right)} + x_{\left(\frac{3(n-2)}{4}+2\right)}\right) & n \equiv 2 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{3(n-3)}{4}+2\right)} + x_{\left(\frac{3(n-3)}{4}+3\right)}\right) & n \equiv 3 \mod 4. \end{cases}$$

Exercice 27.

(i) À gauche : $h = 1$, $\kappa = 10$; à droite : $h = 2$, $\kappa = 11$.

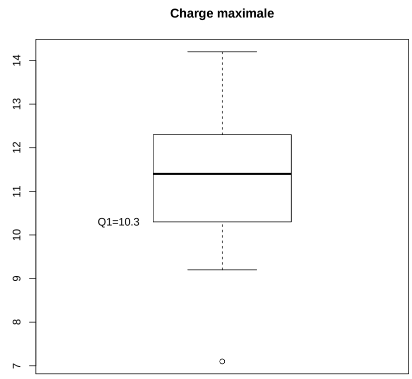




Les deux histogrammes donnent plus ou moins le même message : la distribution est unimodale et légèrement asymétrique à gauche. Le premier histogramme a une plus grande “résolution”, mais avec plus de variabilité. Par exemple, on peux déduire la location du mode plus précisément avec le premier histogramme, mais il y a un intervalle vide entre 8 et 9.

(ii) Il s'agit du premier quartile de l'échantillon, Q_1. Ici $n = 19$ et donc la médiane est $M = x_{(10)}$. Le premier quartile est donc défini comme étant la médiane du sous-échantillon $x_{(1)}, \dots, x_{(10)}$, il est donc donné par $(x_{(5)} + x_{(6)})/2 = 10.3$.

(iii) Le troisième quartile est défini comme étant la médiane du sous-échantillon $x_{(10)}, \dots, x_{(19)}$, il est donc donné par $(x_{(14)} + x_{(15)})/2 = 12.3$.



(iv) Voir le graphique ci-dessous. La valeur 7.1 est une valeur aberrante et le premier quartile Q_1 détermine la borne inférieure de la boîte.

Exercice 28.

(i) La fonction de répartition de la loi exponentielle de paramètre λ est donnée par

$$F_X(x) = 1 - \exp(-\lambda x), \qquad x \geq 0.$$

Puisque cette fonction est continue et strictement croissante sur son support $[0, \infty)$, nous obtenons que $q_\alpha = F_X^-(\alpha) = F_X^{-1}(\alpha)$ et donc

$$\alpha = F_X(q_\alpha) = 1 - \exp(-\lambda q_\alpha) \Longrightarrow q_\alpha = \frac{-\ln(1-\alpha)}{\lambda}.$$

(ii) Supposons par l'absurde que $F_Y(t) < F_X(t)$ pour un certain $t \in \mathbb{R}$. Il existe un $\varepsilon > 0$ tel que $F_Y(t+\varepsilon) < F_X(t)$, car F_Y est continue à droite. Il existe un α tel que $F_Y(t+\varepsilon) < \alpha < F_X(t)$. Visiblement $\alpha \in]0,1[$ et par les définitions de F_X^- et F_Y^- nous avons

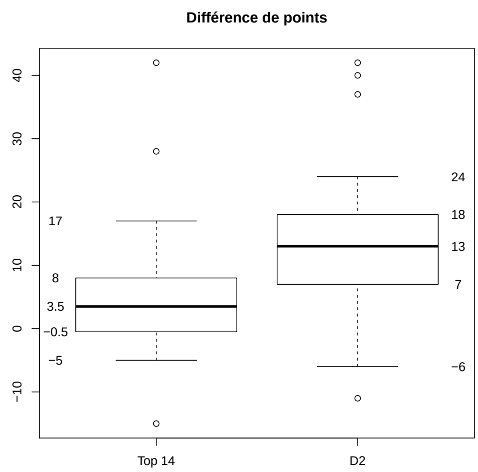
$$F_X^-(\alpha) \leq t < t + \varepsilon \leq F_Y^-(\alpha),$$

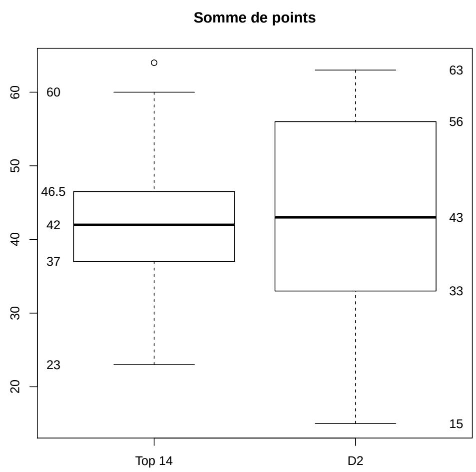
ce qui contredit l'hypothèse $F_X^- = F_Y^-$ sur $]0,1[$. En supposant qu'il existe un $t \in \mathbb{R}$ tel que $F_X(t) < F_Y(t)$, on arrive à une contradiction semblable.

Exercice 29. (i) Voici les tables :

Différence de points			Nombre total de points		
	Top 14	D2		Top 14	D2
Moyenne	6.7	14.2	Moyenne	43.1	43.1
Médiane	3.5	13	Médiane	42	43
Q_1	-0.5	7	Q_1	37	33
Q_3	8	18	Q_3	46.5	56
EIQ	8.5	11	EIQ	9.5	23
W_1	-5	-6	W_1	23	15
W_2	17	24	W_2	60	63

(ii) Voici les graphiques :





En regardant le premier graphique ci-dessus, il semble que dans les deux ligues l'équipe jouant à domicile gagne plus souvent. En plus, l'avantage du terrain est nettement plus prononcé en D2. Il y a une proportion importante de valeurs aberrantes (4 sur 16, 3 sur 14), ce qui pourrait suggérer que les ligues ne sont pas équilibrées.

En regardant le second graphique, on ne peut pas dire qu'une certaine ligue est plus défensive que l'autre. En revanche, la variation entre les matchs semble être plus grande en D2. Il est intéressent de noter que la valeur aberrante correspond au match Grenoble–Lyon, un classique du championnat de France, d'autant plus que la plupart des équipes de rugby à XV viennent du sud de la France.