

Nous souhaitons désormais estimer les quantiles extrêmes de la distribution de températures, notre objectif étant de déterminer le "risque de canicule" entre le 15 juillet et le 15 août à Bordeaux.

## 0.1 Méthodologie

Nous conservons la même base de données que dans la partie précédente, mais nous considérons dans cette partie les mesures de températures maximales journalières prises entre le 15 juillet et le 15 août de chaque année, afin d'effacer le caractère saisonnier mis en évidence dans la première partie.

Nous supposons ainsi que ces données correspondent à  $n$  observations i.i.d.  $X_1, \dots, X_n$  d'une loi  $\mathbb{P}$  inconnue. L'objectif est d'estimer, pour  $\alpha \in [0, 1]$  proche de 1, le quantile d'ordre  $\alpha$  de cette loi. Cela nous donnera la valeur de température qui ne sera pas dépassée - avec un niveau de confiance  $\alpha$ . Notre mesure de "risque de canicule" sera alors la valeur de  $\alpha$  pour laquelle cette température maximale est  $30^\circ\text{C}$ .

Pour cela, comme mis en évidence dans le cours, nous ne considérons pas de méthodes de type paramétriques ou de quantiles empiriques, mais préférons une approche par domaine d'attraction. Nous supposons ainsi que  $X_1, \dots, X_n$  sont dans le domaine d'attraction d'une certaine loi max-stable, ce qui d'après le cours implique l'existence de  $\xi \in \mathbb{R}$  caractérisant cette loi max-stable sous la forme

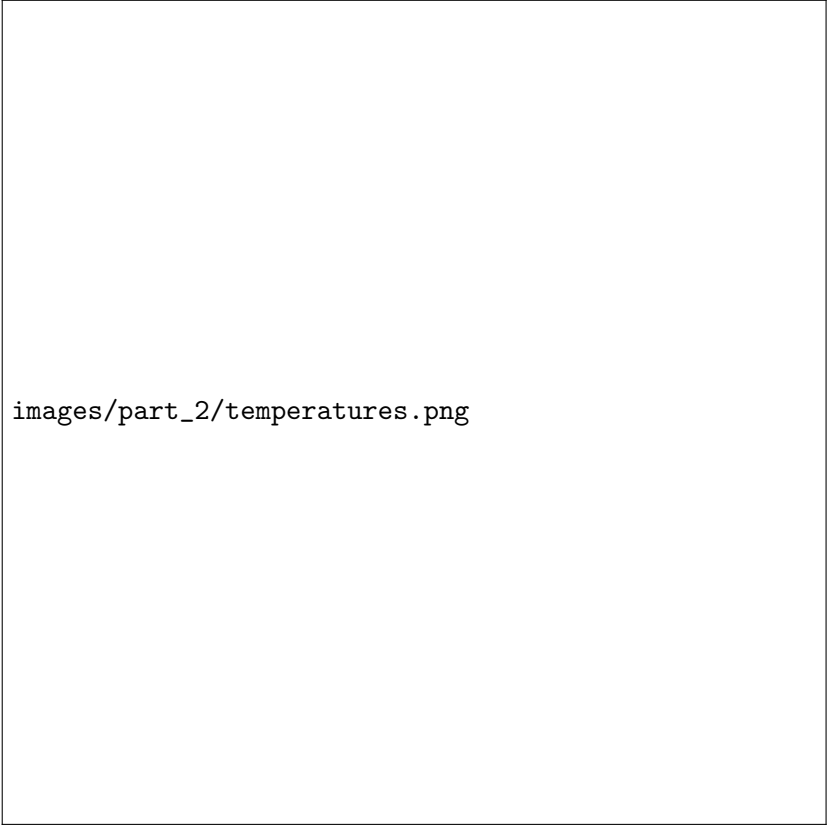
$$H_\xi = \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}} & \text{si } \xi \neq 0 \\ e^{-e^{-x}} & \text{si } \xi = 0 \end{cases}$$

La première chose est de déterminer une estimation du paramètre  $\xi$ , ce que nous faisons dans la section suivante. Nous en déduirons ensuite une estimation du quantile en adaptant une méthode du cours à nos données.

En revanche, il est difficile de chercher à vérifier les résultats, sachant que nous ne connaissons pas la loi exacte de  $X_1$ . Il est donc difficile de proposer un test statistique de vérification, et les résultats que nous obtenons restent spéculatifs.

## 0.2 Estimation de $\xi$

Au lieu d'appliquer aveuglément des calculs d'estimateurs à nos données, étudions-les. Nous pouvons représenter graphiquement les valeurs de températures :



images/part\_2/temperatures.png

FIGURE 1 – Données

### 0.3 Estimateurs de Hill et Pickands

Nous estimons le paramètre  $\xi$  selon les formules suivantes :

$$\hat{\xi}_{n,k(n)}^H = \frac{\sum_{i=n-k(n)+1}^n (\log(X_{(i,n)}) - \log(X_{(n-k(n)+1,n})))}{k(n)} \quad (\text{Estimateur de Hill})$$

et

$$\hat{\xi}_{n,k(n)}^P = \frac{1}{\log(2)} \log \left( \frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}} \right) \quad (\text{Estimateur de Pickands})$$

où  $X_{(i,n)}$  correspond à la  $i$ -ème plus grande valeur parmi tous les  $X_j$ .

Dans les deux cas, des résultats théoriques montrent que l'estimateur converge en probabilités vers la vraie valeur sous les conditions

$$\begin{cases} k(n) & \xrightarrow{n \rightarrow +\infty} +\infty \\ \frac{k(n)}{n} & \xrightarrow{n \rightarrow +\infty} 0 \end{cases}$$

Il s'agit donc de trouver un compromis entre la valeur de  $k$  et celle de  $n$ . En pratique, nous traçons le graphe de  $\hat{\xi}_{n,k(n)}$  en fonction de  $k$  (la valeur de  $n$  est fixée et correspond au nombre d'observations), et nous cherchons une "zone de stabilité" correspondant à la valeur estimée de  $\xi$ .

#### 0.4 Méthode *peak over threshold*