3 juin 2022

3 Distribution de Rayleigh

Soit H une variable aléatoire modélisant la hauteur journalière maximale atteinte par un cours d'eau. On suppose que H suit une distribution de Rayleigh de densité

$$f(x;a) = \begin{cases} \dfrac{x}{a} \exp\left(-\dfrac{x^2}{2a}\right) & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

où le paramètre a est inconnu. Supposons que l'on dispose d'une série statistique $\{h_1,h_2,\dots h_n\}$ d'observations indépendantes de H. Proposer une estimation de la probabilité $P(H > \bar{h})$, où $\bar{h} > 0$ est une hauteur donnée arbitraire.

Solution

1. On cherche dans un 1er temps à estimer le paramètre a de la loi de Rayleigh à partir des observations $\{h_1,h_2,\dots h_n\}$. L'expression de la vraisemblance est

$$P(H = h_1,H = h_2,\dots,H = h_n|a) = \prod_{i=1}^{n} P(H = h_i) = \frac{1}{a^n} \prod_{i=1}^{n} h_i \exp\left(-\sum_{i=1}^{n} \frac{h_i^2}{2a}\right).$$

   On en déduit :

$$\ln P(H = h_1,H = h_2,\dots,H = h_n|a) = -n \ln a + \sum_{i=1}^{n} \ln h_i - \sum_{i=1}^{n} \frac{h_i^2}{2a}.$$

   Le maximum de la log-vraisemblance est atteint lorsque

$$a = \frac{1}{2n} \sum_{i=1}^{n} h_i^2.$$

   L'estimateur par maximum de vraisemblance est donc

$$\Theta_n(a) = \frac{1}{2n} \sum_{i=1}^{n} H_i^2.$$

   et conduit à l'estimation :

$$\hat{a} = \frac{1}{2n} \sum_{i=1}^{n} h_i^2.$$

2. On vérifie que

$$P(H \leq \bar{h}) = \int_0^{\bar{h}} \frac{x}{a} \exp\left(-\frac{x^2}{2a}\right) \mathrm{d}x = 1 - \exp\left(-\frac{\bar{h}^2}{2a}\right).$$

   de sorte que

$$P(H > \bar{h}) = \exp\left(-\frac{\bar{h}^2}{2a}\right).$$

On peut donc proposer l'estimation suivante de $P(H > \bar{h})$:

$$P(H > \bar{h}) \simeq \exp\left(-\frac{\bar{h}^2}{2\hat{a}}\right)$$

où

$$\hat{a} = \frac{1}{2n}\sum_{i=1}^{n} h_i^2.$$