

Lemme: Pour tout $\beta \in [0, 1]$, **Myrtille** est (γ, p) -accurate pour l'estimation d'un décile sur la distribution uniforme sur $[0, 1]$ avec

$$\alpha = 8 (\log k + \log 2/\beta) / \varepsilon$$

$$p = \sum_{k=0}^{\alpha} \binom{n/10}{k} (\gamma)^k (1-\gamma)^{n-k} \times \sum_{k=0}^{\alpha} \binom{9n/10}{k} (\gamma)^k (1-\gamma)^{n-k} + \beta$$

Démonstration: Soit $x \in \mathbb{R}^n$ et $\gamma, \beta \in [0, 1]^2$. On suppose x triée par ordre croissant pour faciliter les notations.

- On pose $\alpha = 8 (\log k + \log 2/\beta) / \varepsilon$. On sait que **AboveThreshold** est alors (α, β) -accurate. On en déduit que,

$$\mathbb{P}(\mathcal{M}(x)_i \in [x_{in/10-\alpha}, x_{in/10+\alpha}]) \geq 1 - \beta$$

- Posons alors $(d_i)_i$ les déciles de x . On note que

$$\begin{aligned} & \mathbb{P}(\mathcal{M}(x)_i \in [d_i - \gamma, d_i + \gamma]) \\ & \geq \mathbb{P}([x_{in/10-\alpha}, x_{in/10+\alpha}] \subset [d_i - \gamma, d_i + \gamma] \wedge \mathcal{M}(x)_i \in [x_{in/10-\alpha}, x_{in/10+\alpha}]) \end{aligned}$$

Or,

$$\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c) = \mathbb{P}(A) + \mathbb{P}(B) - 1$$

Ainsi,

$$\begin{aligned} & \mathbb{P}(\mathcal{M}(x)_i \in [d_i - \gamma, d_i + \gamma]) \\ & \geq \mathbb{P}([x_{in/10-\alpha}, x_{in/10+\alpha}] \subset [d_i - \gamma, d_i + \gamma]) + \mathbb{P}(\mathcal{M}(x)_i \in [x_{in/10-\alpha}, x_{in/10+\alpha}]) - 1 \\ & \geq \mathbb{P}([x_{in/10-\alpha}, x_{in/10+\alpha}] \subset [d_i - \gamma, d_i + \gamma]) - \beta \end{aligned}$$

- Minorons enfin $\mathbb{P}([x_{in/10-\alpha}, x_{in/10+\alpha}] \subset [d_i - \gamma, d_i + \gamma])$ pour conclure!

L'événement dont nous essayons de minorer la probabilité est l'événement $A =$ “au moins α valeurs, parmi celles plus petites que d_i , sont dans $[d_i - \gamma, d_i]$ de longueur γ et au moins α valeurs, parmi celles plus grandes que d_i , sont dans $[d_i, d_i + \gamma]$ ”. Ainsi,

$$\begin{aligned} & \mathbb{P}([x_{i \times n/10-\alpha}, x_{i \times n/10+\alpha}] \subset [d_i - \gamma, d_i + \gamma]) \\ & = 1 - \sum_{k=0}^{\alpha} \binom{n/10}{k} (\gamma)^k (1-\gamma)^{n-k} \times \sum_{k=0}^{\alpha} \binom{9n/10}{k} (\gamma)^k (1-\gamma)^{n-k} \end{aligned}$$

Finalement,

$$\mathbb{P}(\mathcal{M}(x)_i \in [d_i - \gamma, d_i + \gamma]) \geq 1 - \sum_{k=0}^{\alpha} \binom{n/10}{k} (\gamma)^k (1-\gamma)^{n-k} \times \sum_{k=0}^{\alpha} \binom{9n/10}{k} (\gamma)^k (1-\gamma)^{n-k} - \beta$$

Théorème: Pour tout $\beta \in [0, 1]$ et tout $\gamma \in \mathbb{R}_+$, **Myrtille** a une erreur quadratique moyenne inférieur à $\sqrt{9}\gamma$ sur la distribution uniforme sur $[0, 1]$ avec une probabilité au moins $(1-p)^9$ pour

$$\alpha = 8 (\log k + \log 2/\beta) / \varepsilon$$

$$p = \sum_{k=0}^{\alpha} \binom{n/10}{k} (\gamma)^k (1-\gamma)^{n-k} \times \sum_{k=0}^{\alpha} \binom{9n/10}{k} (\gamma)^k (1-\gamma)^{n-k} + \beta$$

Démonstration: C'est un corollaire direct du lemme précédent.