STT 1000 - STATISTIQVES

ARTHVR CHARPENTIER

Loi de Pareto

$$F(x) = \begin{cases} 1 - \left(\dfrac{a}{x}\right)^{\theta} & \text{si } x \geq a \\ 0 & \text{si } x < a \end{cases} \quad \text{et } f(x) = \begin{cases} \dfrac{\theta}{a}\left(\dfrac{a}{x}\right)^{\theta+1} & \text{si } x \geq a \\ 0 & \text{si } x < a \end{cases}$$

$$\begin{cases} E(X) = \frac{\theta a}{\theta - 1} \text{ (n'existe que si } \theta > 1\text{)} \\ V(X) = \frac{\theta a^2}{(\theta-1)^2(\theta-2)} \text{ (n'existe que si } \theta > 2\text{)} \end{cases}$$

La fonction de moyenne en excédant est

$$\mu(t) = \mathbb{E}[X - t | X > t] = \frac{1}{\overline{F}(t)} \int_t^{\infty} \overline{F}(x)dx = xxxxxxxxx$$

Loi de Pareto

Soit $x_1, \cdots, x_n$ un échantillon i.i.d. de densité $f_\alpha(x) = \alpha x^{-(1+\alpha)}$ pour $x \geq 1$, et $\alpha > 0$.
La vraisemblance s'écrit

$$\mathcal{L}(\alpha) = \alpha^n \left(\prod_{i=1}^n x_i\right)^{-\alpha-1}$$

$$\frac{\partial \log \mathcal{L}(\alpha)}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^n \log(x_i)$$

$$\frac{\partial^2 \log \mathcal{L}(\alpha)}{\partial \alpha^2} = -\frac{n}{\alpha^2} < 0$$

$$\widehat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^n \log(x_i)\right)^{-1}$$

Posons $y = \log(x)$ alors la densité de Y est $g(y) = \alpha \exp[-\alpha y]$ for $y \in \mathbb{R}_+$, i.e. $Y \sim \mathcal{E}(\alpha)$.

$$Z = \sum_{i=1}^n Y_i \sim \mathcal{G}(n, \alpha),\ \mathbb{E}[Z] = \frac{n}{\alpha}$$

Loi de Pareto

$$\mathbb{E}[\widehat{\alpha}^2] = \mathbb{E}\left[\frac{n^2}{Z^2}\right] = \frac{n^2 p^2}{(n-1)(n-2)}\int_0^\infty \frac{\alpha^{n-2}z^{n-3}e^{-\alpha z}}{(n-3)!}dz = \frac{n^2\alpha^2}{(n-1)(n-}$$

$$\text{Var}[\widehat{\alpha}^2] = \frac{n^2\alpha^2}{(n-1)(n-2} - \left(\frac{n\alpha}{n-1}\right)^2 = \frac{n^2\alpha^2}{(n-1)^2(n-2)}$$

Notons que

$$\mathbb{E}\left(-\frac{\partial^2 \log \mathcal{L}(\alpha)}{\partial\alpha^2}\right) = \frac{n}{\alpha^2}$$

Loi de Pareto

Pour la méthode des moments,

$$\mathbb{E}[X] = \int_1^\infty x\alpha x^{-\alpha-1}dx = \left[\frac{\alpha x^{1-\alpha}}{1-\alpha}\right]_1^\infty = \frac{\alpha}{\alpha-1}$$

dès lors que $\alpha > 1$ (sinon l'espérance n'existe pas).

$$\overline{x} = \frac{\widehat{\alpha}}{\widehat{\alpha}-1} \text{ ou } \widehat{\alpha} = \frac{\overline{x}}{\overline{x}-1}$$

Loi de Pareto

Soit $x_1, \cdots, x_n$ un échantillon i.i.d. de densité $f_\alpha(x) = \dfrac{3\alpha^3}{x^4}$ pour $x > \alpha$, et $\alpha > 0$.

$$f(x_1, x_2, \cdots, x_n) = \left(3^n \alpha^{3n} \mathbf{1}_{\min\{x_i\} > \alpha}\right) \prod_{i=1}^{n} x_i^{-4},$$

Posons $y = \min\{x_i\}$. La densité de Y est

$$g(y) = n f_\alpha(y) \big[1 - F_\alpha(y)(\big]^{n-1} = \frac{3n\alpha^{3n}}{y^{3n+1}}, \text{ poury} > \alpha.$$

puisque

$$1 - G_\alpha(y) = 3\alpha^3 \int_\alpha^\infty u^{-4} du = \left(\frac{\alpha}{y}\right)^3$$

$$\mathbb{E}[Y] = \int_\alpha^\infty y g(y) dy = \int_\alpha^\infty \frac{3n\alpha^{3n}}{y^{3n}} dy = \frac{3n}{-3n+1} \left[\frac{\alpha^{3n}}{y^{3n-1}}\right]_\alpha^\infty = \frac{3n}{3n-1}\alpha$$

De plus

$$\mathbb{P}[n(Y - \alpha) \leq u] = \mathbb{P}\left[Y \leq \alpha + \frac{u}{n}\right] = 1 - \left(\frac{\alpha}{\alpha + u/n}\right)^{3n}$$

Loi de Pareto

Si on suppose que Y suit une loi de Pareto de paramètres (y_0, θ) avec y_0 connu.

L'estimateur du maxumum de vraisemblance est

$$\widehat{\theta} = \left(\frac{1}{n}\log(y_i) - \log(y_0)\right)^{-1}$$

et l'inverse de l'information de Fisher est θ^2/n, donc

$$\sqrt{n}\frac{\widehat{\theta}-\theta}{\theta} \approx \mathcal{N}(0,1)$$

Aussi, si on veut un intervalle de confiance à 95%, notons que

$$\mathbb{P}\left(-1.96 \leq \sqrt{n}\frac{\widehat{\theta}-\theta}{\theta} \leq 1.96\right) \simeq 0.95$$

soit, de manière équivalente

$$\mathbb{P}\left(\frac{\widehat{\theta}}{1+1.96/\sqrt{n}} \leq \theta \leq \frac{\widehat{\theta}}{1-1.96/\sqrt{n}}\right) \simeq 0.95$$

or $(1+u)^{-1} \simeq 1-u$ donc