1 Point estimation

Context

Our engineering team just landed a consulting contract with a company interested in the electricity consumption of its machines. In a first part, we would like to determine how electricity consumption is evenly distributed across the different machines of the same type. To this end, we use the Gini coefficient. In a nutshell, it is an index ranging from 0 to 1 measuring the inequality featured in a distribution. A value of 0 denotes that all our machines use the same amount of electricity while a value of 1 means that all the electricity is used by a single machine. We assume that all of the n machines operate independently and their daily electricity consumption (in MWh) can be modelled as a random variable X with the following density function,

$$f_{\theta_1,\theta_2}(x) = \begin{cases} \frac{\theta_1 \theta_2^{\theta_1}}{x^{\theta_1+1}}, & x \geq \theta_2 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

with $\theta_1 > 2$ and $\theta_2 > 0$.

(a) Derive the quantile function of X

We're looking to solve $P(X \leq x_t) = t$ for x_t.

First let's compute $P(X \leq x_t)$,

$$\begin{aligned} P(X \leq x_t) &= \int_{-\infty}^{x_t} f_{\theta_1,\theta_2}(x)dx \\ &= \int_{\theta_2}^{x_t} \theta_1 \theta_2^{\theta_1} x^{-(\theta_1+1)} dx \\ &= -\frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} \left[x^{-\theta_1}\right]_{x=\theta_2}^{x=x_t} \\ &= -\frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} \left(x_t^{-\theta_1} - \theta_2^{-\theta_1}\right) \end{aligned}$$

Let's solve $P(X \leq x_t) = t$ for x_t,

$$\begin{aligned} -\frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} \left(x_t^{-\theta_1} - \theta_2^{-\theta_1}\right) = t \iff x_t^{\theta_1} &= \frac{t\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \\ \iff x_t &= \left(\frac{t\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1}\right)^{1/\theta_1} \equiv Q_{\theta_1,\theta_2}(t) \end{aligned}$$

(b) Derive the Gini coefficient of X.

The Gini coefficient is defined as,

$$G_{\theta_1,\theta_2} = 2\int_0^1 \left(p - \frac{\int_0^p Q(t)dt}{E(X)}\right) dp \tag{2}$$

Let's first compute the mean of X,

$$\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\
&= \int_{\theta_2}^{+\infty} x \frac{\theta_1 \theta_2^{\theta_1}}{x^{\theta_1+1}} dx \\
&= \theta_1 \theta_2^{\theta_1} \int_{\theta_2}^{+\infty} x^{-\theta_1} dx \\
&= -\frac{\theta_1 \theta_2^{\theta_1}}{(\theta_1 - 1)} \left[x^{-(\theta_1-1)} \right]_{\theta_2}^{+\infty}
\end{aligned}$$

Then the Gini coefficient,

$$G_{\theta_1,\theta_2} = 2 \left(\int_0^1 p dp - \int_0^1 \frac{\int_0^p Q(t)dt}{E(X)} dp \right)$$

We compute each integral separately,

$$\int_0^p Q(t)dt = \int_0^p \left(\frac{t\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \right)^{1/\theta_1} dt$$

We use the change of variable $u = \frac{t\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1}$, $du = \frac{\theta_1}{\theta_1 \theta_2^{\theta_1}} dt$

The boundaries becomes,

$$\begin{cases} t = 0 & \implies u_1 \equiv -\theta_2^{-\theta_1} \\ t = p & \implies u_2 \equiv \frac{p\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \end{cases}$$

Then,

$$\begin{aligned}
\int_0^p Q(t)dt &= \int_{u_1}^{u_2} u^{(1/\theta_1)} \frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} du \\
&= \frac{\theta_1}{\theta_1 \theta_2^{\theta_1}} \left[\frac{u^{(1/\theta_1)+1}}{(1/\theta_1)+1} \right]_{u_1}^{u_2} \\
&= \frac{\theta_1}{\theta_1 \theta_2^{\theta_1}((1/\theta_1)+1)} \left(\left(\frac{p\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} - \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} \right)
\end{aligned}$$

Therefore,

$$\frac{\int_0^p Q(t)dt}{E(X)} = \frac{\theta_1(\theta_1 - 1)}{\theta_2^{(1-\theta_1)}((1/\theta_1)+1)} \left(\left(\frac{p\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} - \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} \right)$$

Then,

$$\int_0^1 \frac{\int_0^p Q(t)dt}{E(X)} dp = \frac{\theta_1(\theta_1 - 1)}{\theta_2^{1-\theta_1}} \left(\underbrace{\int_0^1 \left(\frac{p\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} dp}_{\equiv A} - \underbrace{\int_0^1 \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} dp}_{\equiv B} \right)$$

Computing integral A and B. For A we use the same change of variable as before,

$$\begin{aligned} A &= \int_{u_1}^{u_2} u^{(1/\theta_1)+1} \frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} du \\ &= \frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} \left(\left(\frac{\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \right)^{(1/\theta_1)+2} - \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+2} \right) \end{aligned}$$

$$\begin{aligned} B &= \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} \int_0^1 dp \\ &= \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+1} \end{aligned}$$

Then,

$$\int_0^1 pdp = \frac{1}{2}$$

Eventually,

$$G_{\theta_1,\theta_2} = 2 \left(\frac{1}{2} - \frac{\theta_1(\theta_1 - 1)}{\theta_2^{(1-\theta_1)}((1/\theta_1)+1)} \left[\frac{\theta_1 \theta_2^{\theta_1}}{\theta_1} \frac{1}{(1/\theta_1)+2} \left(\left(\frac{\theta_1}{\theta_1 \theta_2^{\theta_1}} - \theta_2^{-\theta_1} \right)^{(1/\theta_1)+2} - \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+2} \right) - \left(-\theta_2^{-\theta_1} \right)^{(1/\theta_1)+} \right.\right.$$

(c) Derive the maximum likelihood estimator (MLE) of G_{θ_1,θ_2}. Call this estimator $\hat{G}_{\text{MLE}}$

Let's first compute the likelihood function $L(\theta_1, \theta_2)$,

$$\begin{aligned} L(\theta_1, \theta_2) &:= \Pi_{i=1}^n f_{\theta_1,\theta_2}(x) \\ &= \Pi_{i=1}^n \frac{\theta_1 \theta_2^{\theta_1}}{x^{\theta_1+1}} \cdot I(X_i \geq \theta_2 > 0) \\ &= \theta_1 \theta_2^{\theta_1} \frac{1}{\Pi_{i=1}^n X_i^{\theta_1+1}} cdot I(X_{(1)} \geq \theta_2 > 0) \end{aligned}$$

where $X_{(1)} \equiv \min(X_1, ..., X_n)$.

We notice that $L(\theta_1, \theta_2)$ is not continuous along θ_2 and then not differentiable in θ_2. However, we observe that $L(\theta_1, \theta_2)$ increase with θ_2. Therefore, we have to take θ_2 the largest possible in order to maximize $L(\theta_1, \theta_2)$ respecting the condition $X_{(1)} \leq \theta_2 > 0$ otherwise we would have $L(\theta_1, \theta_2) = 0$,

$$\hat{\theta}_2 = X_{(1)}$$

For $\hat{\theta}_1$ we can compute the log-likelihood function $l(\theta_1, \theta_2)$,

$$\begin{aligned} l(\theta_1, \theta_2) &:= \ln(L(\theta_1, \theta_2)) \\ &= \ln(\theta_1) + \ln(\theta_2^{\theta_1}) + \ln(1) - \ln(\pi_{i=1}^{n} X_i^{(\theta_1+1)}) \\ &= \ln(\theta_1) + \theta_1 \ln(\theta_2) - (\sum_{i=1}^{n} \ln(X_i^{(\theta_1+1)})) \\ &= \ln(\theta_1) + \theta_1 \ln(\theta_2) - \sum_{i=1}^{n} (\theta_1 + 1)) \ln(X_i) \end{aligned}$$

We differentiate with respect to θ_1 in order to find the maximum,

$$\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_1} = \frac{1}{\theta_1} + \ln(\theta_2) - \sum_{i=1}^{n} \ln(X_i)$$

Then,

$$\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_1} = 0 \iff \hat{\theta}_1 = \frac{1}{\sum_{i=1}^{n} \ln(X_i) - \ln(\theta_2)}$$

Now we can compute $\hat{G}_{\text{MLE}}$,

$$\begin{aligned} \hat{G}_{\text{MLE}} &:= G_{\hat{\theta}_1, \hat{\theta}_2} \\ &= \end{aligned}$$

(d) Propose a method of moment estimator of G_{θ_1,θ_2}. Call this estimator $\hat{G}_{\text{MME}}$

We already have computed the mean of X,

$$E(X) = -\frac{\theta_1 \theta_2^{\theta_1}}{(\theta_1 - 1)} \left[x^{-(\theta_1 - 1)} \right]_{\theta_2}^{+\infty}$$