MAP 531 : Statistics

26 septembre 2017

Poisson process and exponential distribution

- Poisson process : is the basic model for systems that have discrete outputs (number of people standing in line, break occurences, survival, etc..)
- Between-arrival model : $X_1 \ldots, X_n$ i.i.d. with an exponential distribution of parameter $\theta \in \Theta = \mathbb{R}_+^*$:

$$p(\theta, x) = \theta \mathrm{e}^{-\theta x} \mathbb{1}_{\{x>0\}} \,.$$

- Proprerties : for all $a, b > 0$ and $\theta \in \mathbb{R}_+^*$,

$$\mathbb{P}_\theta \left(X_1 \geq a + b \,|\, X_1 \geq a \right) = \frac{\mathrm{e}^{-(a+b)\theta}}{\mathrm{e}^{-a\theta}} = \mathbb{P}_\theta(X_1 \geq b)$$

Computation of the moments

- Generative function of moments : for $s < \theta$,

$$\mathbb{E}_\theta[\mathrm{e}^{sX}] = \int \mathrm{e}^{sx}\theta\mathrm{e}{-}\theta x \mathrm{d}x$$
$$= \frac{\theta}{s-\theta} = \sum_{k=0}^{\infty} \frac{s^k}{\theta^k}$$

- Development of the exponential

$$\mathbb{E}_\theta[\mathrm{e}^{sX}] = \sum_{k=0}^{\infty} \frac{s^k}{k!}\mathbb{E}_\theta[X^k]$$

- By identification

$$\mathbb{E}_\theta[X^k] = \frac{k!}{\theta^k} .$$

Estimator using moments

Let $T(x) = x$ and $\tilde{T}(x) = x^2$.

$$e(\theta) = \mathbb{E}_\theta \left[T(X_1) \right] = \int_0^{+\infty} x\theta \exp(-\theta x) \mathrm{d}x = \frac{1}{\theta}$$

$$\tilde{e}(\theta) = \mathbb{E}_\theta \left[\tilde{T}(X_1) \right] = \int_0^{+\infty} x^2\theta \exp(-\theta x) \mathrm{d}x = \frac{2}{\theta^2} \; .$$

Estimator using moments

Let $T(x) = x$ and $\tilde{T}(x) = x^2$.

$$e(\theta) = \mathbb{E}_\theta\left[T(X_1)\right] = \int_0^{+\infty} x\theta\exp(-\theta x)\mathrm{d}x = \frac{1}{\theta}$$

$$\tilde{e}(\theta) = \mathbb{E}_\theta\left[\tilde{T}(X_1)\right] = \int_0^{+\infty} x^2\theta\exp(-\theta x)\mathrm{d}x = \frac{2}{\theta^2}\,.$$

- The associated estimators using these moments are solutions of

$$e(\theta) = \frac{1}{\theta} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{et} \quad \tilde{e}(\theta) = \frac{2}{\theta^2} = \frac{1}{n}\sum_{i=1}^n X_i^2\,.$$

Estimator using moments

Let $T(x) = x$ and $\tilde{T}(x) = x^2$.

$$e(\theta) = \mathbb{E}_\theta\left[T(X_1)\right] = \int_0^{+\infty} x\theta\exp(-\theta x)\mathrm{d}x = \frac{1}{\theta}$$

$$\tilde{e}(\theta) = \mathbb{E}_\theta\left[\tilde{T}(X_1)\right] = \int_0^{+\infty} x^2\theta\exp(-\theta x)\mathrm{d}x = \frac{2}{\theta^2}\,.$$

- The associated estimators using these moments are solutions of

$$e(\theta) = \frac{1}{\theta} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{et} \quad \tilde{e}(\theta) = \frac{2}{\theta^2} = \frac{1}{n}\sum_{i=1}^n X_i^2\,.$$

- These eqations have unique solutions :

$$\hat{\theta}_{n,1} = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i} \quad \text{et} \quad \hat{\theta}_{n,2} = \left(\frac{2}{\frac{1}{n}\sum_{i=1}^n X_i^2}\right)^{1/2}\,.$$

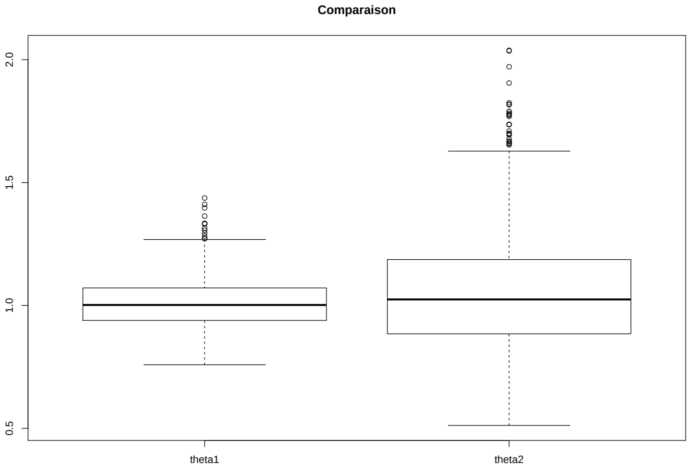

FIGURE – Performance comparison of these estimators $\hat{\theta}_1 = 1/(n^{-1}\sum_{i=1}^n X_i)$ et $\hat{\theta}_2 = 2/(n^{-1}\sum_{i=1}^n X_i^2)$

Translation and scaling model

Let $X_1, \ldots, X_n$ be a n-sample drawn with the following density :

$$p(\theta, x) = \frac{1}{\sigma} p\left(\frac{x-\mu}{\sigma}\right) \ , \quad \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^* \ ,$$

where p is a density on $\mathbb{R}$ satisfying

$$\int xp(x)\mathrm{d}x = 0, \quad \int x^2 p(x)\mathrm{d}x = m_2 > 0 \quad \int x^4 p(x)\mathrm{d}x = m_4 < \infty.$$

Translation and scaling model

Let $X_1, \ldots, X_n$ be a n-sample drawn with the following density :

$$p(\theta, x) = \frac{1}{\sigma} p\left(\frac{x-\mu}{\sigma}\right) \ , \quad \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^* \ ,$$

where p is a density on $\mathbb{R}$ satisfying

$$\int x p(x) \mathrm{d}x = 0, \quad \int x^2 p(x) \mathrm{d}x = m_2 > 0 \quad \int x^4 p(x) \mathrm{d}x = m_4 < \infty.$$

Examples :

- The standard normal distribution, $p(x) = (2\pi)^{-1/2} \exp(-x^2/2)$: in this case, $m_2 = 1$ and $m_4 = 3$.
- The Laplace distribution $p(x) = (1/2) \exp(-|x|)$ such that $m_2 = 2$ and $m_4 = 4!$.

Estimator of moments

Let $T_1(x) = x$ and $T_2(x) = x^2$ and we identify the estimator of moments as the solution of the following system :

$$\begin{cases} e_1(\mu, \sigma^2) & = \mu = \frac{1}{n} \sum_{i=1}^n X_i \ , \\ \\ e_2(\mu, \sigma^2) & = m_2 \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \ . \end{cases}$$

Estimator of moments

$$\begin{cases} \hat{\mu}_n & = \frac{1}{n} \sum_{i=1}^n X_i \ , \\ \\ \hat{\sigma}_n^2 & = \frac{1}{n m_2} \sum_{i=1}^n (X_i - \hat{\mu}_n)^2 \ . \end{cases}$$

Poisson distribution

In this setting $\mathsf{X} = \mathbb{N}$, and the measure is the discrete measure on $\mathbb{N}$. For $\theta \in \Theta = \mathbb{R}_+^*$, the Poisson distribution of parameter θ, is given by the followinf density w.r.t the counding measure

$$f_\theta(x) = \mathrm{e}^{-\theta} \frac{\theta^x}{x!} \ , \quad x \in \mathbb{N}, \theta \in \mathbb{R}_+^* \ .$$

Let $X_1, \ldots, X_n$ be a n-sample from a Poisson distribution of paramter θ. The likelihood writes, for all $\theta > 0$,

$$\mathrm{L}_n(\theta, X_1, \ldots, X_n) = \prod_{i=1}^n \mathrm{e}^{-\theta} \frac{\theta^{X_i}}{X_i!} = \frac{1}{\prod_{i=1}^n X_i!} \exp\left(-n\theta + n\log(\theta)\bar{X}_n\right) \ .$$

This likelihhod has a unique maximum given by

$$\bar{X}_n = n^{-1} \sum_{i=1}^n X_i \ .$$

Translation parameter for a Cauchy distribution

- Let $(X_1, \ldots, X_n)$ be a n-sample from a Cauchy distribution with translation parameter θ (and scaling parameter 1)

$$f(\theta, x) = \frac{1}{\pi\big(1 + (x-\theta)^2\big)} \,, \quad \theta \in \Theta = \mathbb{R}$$

- The log-likelihood is given by

$$\ell_n(\theta, X_1, \ldots, X_n) = -\log \pi - \frac{1}{n} \sum_{i=1}^{n} \log\big(1 + (X_i - \theta)^2\big),$$

The likelihood equation is therefore :

$$\dot{\ell}_n(\theta, X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i - \theta}{1 + (X_i - \theta)^2} = 0.$$

- This equation does not have explicit solutions.

Numerical method to approximate the MLE

One can use a gradient ascent method which is summarised by the following equation : let $\theta^{(k)}$ be the value of the parameter estimate at the k-th iteration of the algorithm. Then,

$$\theta^{(k+1)} = \theta^{(k)} + \gamma \dot{\ell}_n(\theta^{(k)}) \ ,$$

where γ has to be chosen (with care).

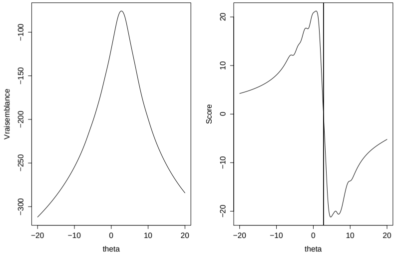

FIGURE – Likelihood and derivative form $n = 50$ observations from a Cauchy model with $\theta = 2$

MLE is not always defined

- Let $(X_1, \ldots, X_n)$ be a n-sample of random variables following $\{f_0(x-\theta), \theta \in \Theta = \mathbb{R}\}$ w.r.t the Lebesgue measure where

$$f_0(x) = \frac{\mathrm{e}^{-\frac{|x|}{2}}}{2\sqrt{2\pi|x|}}, \quad x \in \mathbb{R},$$

- The likelihood writes

$$\mathrm{L}_n(\theta, X_1, \ldots, X_n) = \prod_{i=1}^{n} f_0(X_i - \theta).$$

- For all $i = 1, \ldots, n$, we have

$$\lim_{\theta \to X_i} \mathrm{L}_n(\theta, X_1, \ldots, X_n) = +\infty.$$

- For this statistical experiment, the MLE is not defined.

MLE may not be unique

- Let $(X_1, \ldots, X_n)$ be a n-sample from the Laplace distribution with parameter $\theta \in \Theta = \mathbb{R}$; the pdf w.r.t the Lebesgue measure is

$$f(\theta, x) = \frac{1}{2} \exp\left(- |x - \theta| \right),$$

- The log-likelihood writes

$$\ell_n(\theta, X_1, \ldots, X_n) = -\log(2) - \frac{1}{n} \sum_{i=1}^{n} \left| X_i - \theta \right| .$$

- Maximizing $\theta \mapsto \ell_n(\theta, X_1, \ldots, X_n)$ is equivalent to minimizing the following function

$$\theta \to \sum_{i=1}^{n} \left| X_i - \theta \right| .$$

MLE may not be unique

- This function is differentiable almost everywhere with differential equaled to

$$-\sum_{i=1}^{n} \mathrm{sign}(X_i - \theta) \ .$$

This differential (defined almost everywhere) is piecewise constant.

- If n is odd, there is a unique zero $X_{1:(n+1)/2}$, where $X_{1:n} \leq \ldots \leq X_{n:n}$ is the ordering statistic associated to the sample
- If n is even, there exists an infinite number of solutions : any point in $\left(X_{1:n/2}, X_{1:n/2+1}\right)$ is a MLE.

Translation and sclaing model

- Let $X_1, \dots, X_n$ be a n-sample withthe following pdf

$$p(\theta, x) = \frac{1}{\sigma} p\left(\frac{x-\mu}{\sigma}\right) \ , \quad \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^* \ ,$$

  where p is a pdf on $\mathbb{R}$ such that p is even.
- Let us consider the following functions

$$\psi_1(\mu, \sigma, x) = \mathrm{signe}(x - \mu) \quad \text{et} \quad \psi_2(\mu, x) = -m + \frac{|x-\mu|}{\sigma}$$

  where $m = \int |x| p(x) \mathrm{d}x$. Then :

$$\int_{-\infty}^{\infty} \mathrm{signe}(x-\mu)\sigma^{-1} p(\sigma^{-1}(x-\mu))\mathrm{d}x = 0$$
$$\int_{-\infty}^{\infty} \left\{m - \sigma^{-1}|x-\mu|\sigma^{-1}\right\} p(\sigma^{-1}(x-\mu))\mathrm{d}x = 0$$

Translation and sclaing model

- The *empirical median* of the sample is a Z-estimator of the translation parameter

$$\hat{\mu}_n = \begin{cases} X_{m:n} & \text{si } n = 2m-1 \\ (1/2)(X_{m:n} + X_{m+1:n}) & \text{si } m = 2n \end{cases}$$

where $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ is the ordering statistic associated to the sample $X_1, \ldots, X_n$

- The Z-estimator of the scaling parameter is given by

$$\hat{\sigma}_n = \frac{1}{mn} \sum_{i=1}^{n} |X_i - \hat{\mu}_n| \; .$$