

Derive MLE for parameter of Pareto distribution given by following PDF

$$f(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} & x > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Discuss existence of moments (depending on values  $\alpha$ ) for this probability distribution.

$$1) L(\alpha) = \prod_{i=1}^n \frac{\alpha}{x_i^{\alpha+1}} \rightarrow \ell(\alpha) = \sum_{i=1}^n \ln\left(\frac{\alpha}{x_i^{\alpha+1}}\right) = n \cdot \ln(\alpha) - \sum_{i=1}^n (\alpha+1) \cdot \ln(x_i)$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^n \ln(x_i) \rightarrow \hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln(x_i)}$$

$$2) m^k \rightarrow \int_1^{\infty} x^k \cdot \frac{\alpha}{x^{\alpha+1}} dx \rightarrow \text{distribution pro jake } \alpha \text{ } \exists \text{ jake momenty}$$

Consider MLE for a parameter  $\sigma^2$  of Normal distribution (with unknown  $\mu$ ).

- Find out whether the MLE is biased or not.
- If the MLE is biased, find an unbiased estimator.
- Discuss bias of any estimators for standard deviation.

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i \bar{x} + n \bar{x}^2 \right) - (\bar{x}^2 - 2 \bar{x} \bar{x} + \bar{x}^2) =$$

$$= \frac{1}{n} \left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) - (\bar{x} - \bar{x})^2 \Rightarrow E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right) = E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2\right)$$

$$E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right) \rightarrow x_i - \mu \sim N(0; \sigma^2) \Rightarrow E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right) = \frac{1}{n} \cdot n \sigma^2 = \sigma^2$$

$$E((\bar{x} - \mu)^2) \rightarrow \text{rozptyl } \bar{X}; \bar{X} \sim N(\mu; \frac{\sigma^2}{n}) \Rightarrow E((\bar{x} - \mu)^2) = \frac{\sigma^2}{n}$$

$$\text{dokonady: } E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n\sigma^2 - \sigma^2}{n} = \frac{n-1}{n} \cdot \sigma^2 \neq \sigma^2$$

$$\text{Nesluší odhad: } \frac{1}{n} \rightarrow \frac{1}{n-1} \Rightarrow \frac{n-1}{n-1} \sigma^2 = \sigma^2$$

$$\text{Směrodatná odchylka: } \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \rightarrow \sqrt{E(x)} \neq E(\sqrt{x})$$

$$\sqrt{x} \text{ konvexní} \Rightarrow E\left(\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}\right) < \sqrt{E\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2\right)} = \sigma$$

$\Rightarrow$  odhad je vždýlens a podhodnocuje

Derive MLE for a parameter of geometric distribution (number of Bernoulli trials before first success) given by following PMF

$$p(x) = (1-p)^{x-1} p$$

Discuss a bias of this estimator (if the MLE is biased try to use Jensen's inequality). Show that geometric distribution is a part of exponential family. Use factorization criterion to show that the resulting MLE is using sufficient statistic.

$$L(p) = \prod_{i=1}^n (1-p)^{x_i-1} \cdot p$$

$$\ell(p) = \sum_{i=1}^n \ln[(1-p)^{x_i-1} \cdot p] = \sum_{i=1}^n [(x_i-1) \cdot \ln(1-p) + \ln p] =$$

$$= \sum_{i=1}^n x_i \cdot \ln(1-p) - n \ln(1-p) + n \ln(p)$$

$$\frac{\partial \ell}{\partial p} = - \frac{\sum_{i=1}^n x_i}{1-p} + \frac{n}{1-p} + \frac{n}{p} = 0$$

$$\frac{\sum_{i=1}^n x_i}{n} = \frac{1-p}{1-p} + \frac{1-p}{p} = \frac{1}{p} \Rightarrow \hat{p} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$$

$$E\left(\frac{1}{\bar{x}}\right) > \frac{1}{E(\bar{x})} \text{ protože } \frac{1}{x} \text{ je konvexní} \Rightarrow$$

$\Rightarrow \hat{p}$  nadhodnocuje

$$p(x) = (1-p)^{x-1} p \rightarrow e^{\ln((1-p)^{x-1} p)} = e^{(x-1) \ln(1-p) + \ln p}$$

$$\Rightarrow h(x) = 1; T(x) = x-1; \eta(\theta) = \ln(1-p); A(\theta) = \ln p$$

$$\prod_{i=1}^n [(1-p)^{x_i-1} \cdot p] = (1-p)^{\sum_{i=1}^n (x_i-1)} \cdot p^n = (1-p)^{\sum_{i=1}^n x_i} \cdot \left(\frac{p}{1-p}\right)^n \Rightarrow$$

$$T(x) = \sum_{i=1}^n x_i \text{ is sufficient}$$

$$1) p(x_i, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \cdot e^{-n\lambda} = \prod_{i=1}^n \left( \frac{\lambda}{x_i!} \right) \cdot \left[ \lambda^{\sum_{i=1}^n x_i} \cdot e^{-n\lambda} \right] \rightarrow$$

$$\mu(x) = \prod_{i=1}^n \left( \frac{1}{x_i!} \right); \quad \nu(x_i, \lambda) = \lambda^{\sum_{i=1}^n x_i} \cdot e^{-n\lambda} \rightarrow T_1(x) = \sum_{i=1}^n x_i$$

$$2) f(x_i, \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = \left( \frac{1}{\sqrt{2\pi} \cdot \sigma} \right)^n \cdot e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} =$$

$$= \left( \frac{1}{\sqrt{2\pi} \cdot \sigma} \right)^n \cdot e^{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \cdot \sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}} \Rightarrow \begin{matrix} T_1 = \sum_{i=1}^n x_i \\ T_2 = \sum_{i=1}^n x_i^2 \end{matrix}$$

$$3) f(x_i, \alpha, \beta) = \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \cdot e^{-\beta \sum_{i=1}^n x_i} \cdot \sum_{i=1}^n \ln(x_i) \cdot e^{-\beta \sum_{i=1}^n x_i}$$

$$T_1(x) = \sum_{i=1}^n x_i \quad T_2(x) = \sum_{i=1}^n \ln(x_i) \quad \text{protože } \sum_{i=1}^n \ln(x_i) = \ln\left(\prod_{i=1}^n x_i\right)$$

nebo

$$T_2^*(x) = \prod_{i=1}^n (x_i)$$

Find a (jointly) sufficient statistics for following distributions:

- Poisson distribution  $p(x) = \frac{\lambda^x}{x!} e^{-\lambda}; x \in \{0, 1, \dots\}$
  - Normal distribution
  - Gamma distribution  $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}; x \geq 0$
- [Gamma function wiki](#)