Parameter-Varying Support for Maximum Likelihood on Data with Unknown Left Endpoint

Matheus H J Saldanha, University of São Paulo & Federal University of São Carlos
Adriano K Suzuki, University of São Paulo

Abstract

- Lower-bounded continuous data are ubiquitous in many fields. But the left endpoint is not always known.
- How to perform inference, then? At present, the options are: **i)** guess the left endpoint, **ii)** add it as an extra parameter to the distribution, **iii)** estimate it with endpoint estimators.
- **Our proposal:** an "iterative method" where the likelihood function $L(\cdot \mid \boldsymbol{\theta})$ modifies the sample on-the-fly, using an estimate of what the sample minimum would be if $\boldsymbol{\theta}$ was the true parameter.
- **Evaluation:** Simulation experiments (with the gamma, Weibull and generalized gamma distributions) shows that the proposed method outperforms all the other methods, achieving better performance in $\sim 65\%$ of the trials in the worst case scenario.
  - + It can be proved to share some properties of usual MLE estimates,
  - + it works better on smaller samples.

ICMC USP SÃO CARLOS ufscar CAPES

Context and challenges

- The problem of inference with unknown population minimum has not been tackled directly.
- How have researchers dealt with this so far?

1. Guess the population minimum.
2. Add the population minimum as a parameter:

$$g(x \mid \boldsymbol{\theta}, m) := f(x - m \mid \boldsymbol{\theta}), \qquad (1)$$

3. Use a left endpoint estimator.

- e.g. Estimator of Alves and Neves (2014):

$$\hat{\mathcal{M}} = X_{(n-k)} + X_{(n)} + \frac{1}{\log 2} \sum_{i=0}^{k-1} \log\left(\frac{k+i}{k+i+1}\right) X_{(n-k-i)}$$

- Our objectives are twofold:
  - ↪ Estimate the population minimum m
  - ↪ Estimate the parameters $\boldsymbol{\theta}$

Proposed method

- Let X have cdf $F(x - m \mid \boldsymbol{\theta})$
- The cdf of the sample minimum is:

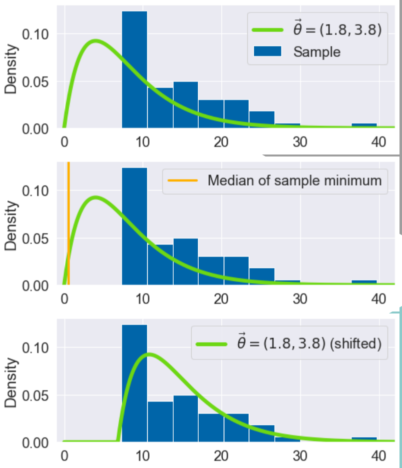
$$F_{\text{min}}(x \mid \boldsymbol{\theta}) = 1 - [1 - F(x \mid \boldsymbol{\theta})]^n$$

- The corresponding quantile function:

$$F_{\text{min}}^{-1}(q \mid \boldsymbol{\theta}) = F^{-1}\left(1 - (1-q)^{1/n} \mid \boldsymbol{\theta}\right)$$

- The median $F_{\text{min}}^{-1}(0.5 \mid \boldsymbol{\theta})$ is a **good guess** for what the **sample minimum** would be if $\boldsymbol{\theta}$ was the true parameter.
- Thus, we use the modified family of pdfs:

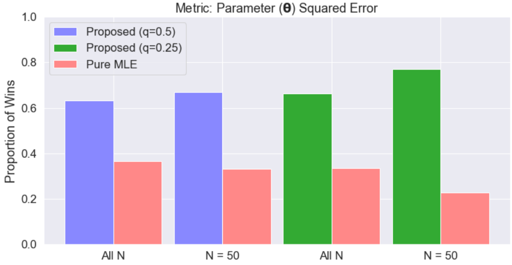
$$\left\{f\left(x - x_{(1)} + F_{\text{min}}^{-1}(q \mid \boldsymbol{\theta}) \mid \boldsymbol{\theta}\right) \; : \; \boldsymbol{\theta} \in \Omega\right\}$$



- Preserves properties of usual MLE estimator
- Iterative procedure for more stability:

$$\boldsymbol{\theta}^{(n+1)} = \underset{\tilde{\boldsymbol{\theta}}}{\arg\max} \sum_{j=1}^{n} \log f\left(x_j + x_{q,\boldsymbol{\theta}^{(n)}} \mid \tilde{\boldsymbol{\theta}}\right)$$

Simulation experiments



- Generalized Gamma distribution
- 8 combinations of parameters
- Sample sizes 50, 100, 500

- For this particular case, $q = 0.25$ is better
- $66.4\%/77.1\%$ for parameter squared error
- $70.4\%/79.8\%$ for left endpoint squared error

- In Gamma and Weibull simulations, $q = 0.5$ performs better