Simulated switched Z-estimation for accurate finite sample inference

Samuel Orso

Stéphane Guerrier, Mucyo Karamera, Maria-Pia Victoria-Feser

Data Analytics Lab Seminar, Geneva - December 3 2019

Introduction

Motivations:

- For many problems, there are no consistent estimators (readily available); for example, no analytical likelihood (stable distribution, queueing processes, stochastic processes, non-Gaussian factor models,...), robust estimators, ...
- Point-estimators are commonly obtained via simulation-based methods (indirect inference), but they are numerically cumbersome.
- Consequently interval estimation relies on asymptotic theory, bootstrap-like methods cannot be performed.

Simulation-based point-estimation

General setting:

Consider a random sample $X \in \mathbb{R}^n$ generated from model F_{θ_0}. We wish to make inference on θ_0. Pseudo-samples can be generated $X^*(\theta, \omega)$: it depends upon θ and a random variable $\omega \in \mathbb{R}^n$ independent from θ.

Estimator based on the matching:

(inconsistent) estimator based on pseudo-sample X^*

$$\hat{\theta}^*: \quad \hat{\pi}(\theta_0, \omega_0) = \hat{\pi}^*(\theta, \omega)$$

(inconsistent) estimator based on the sample X

Remarks:

- ω_0 is not observed but ω can be simulated.
- we assume there is at least one solution.

Inference based on the matching

A possible strategy for inference is by simulating different ω:

$$\hat{\theta}^*(\omega_1): \quad \boxed{\hat{\pi}(\theta_0, \omega_0)} = \boxed{\hat{\pi}^*(\theta, \omega_1)}$$

$$\vdots$$

$$\hat{\theta}^*(\omega_B): \quad \boxed{\hat{\pi}(\theta_0, \omega_0)} = \boxed{\hat{\pi}^*(\theta, \omega_B)}$$

$$\vdots$$

+ $\hat{\pi}$ can be an inconsistent estimator of θ_0.
+ $\theta_0 = \hat{\theta}^*(\omega_0)$ (under some identifiability conditions) belongs to the set of solutions.
- Computationally intensive (nested optimization).

Proposed approach - SwiZ

Suppose that $\hat{\pi}(\theta_0, \omega_0)$ is a Z-estimator

$$\boxed{\hat{\pi}(\theta_0, \omega_0)}: \quad \Psi\{X(\theta_0, \omega_0), \pi\} = 0$$

We propose to target θ_0 directly in the estimating function:

$$\hat{\theta}^*(\omega): \quad \Psi\{X^*(\theta, \omega), \boxed{\hat{\pi}(\theta_0, \omega_0)}\} = 0$$

We call this estimator the Switched Z-estimator.

Proposition 1: equivalence between SwiZ and matching

If $\hat{\pi}(\theta, \omega)$ is the unique solution of a Z-estimator, then the solutions of the SwiZ and the estimators based on the matching are in the same set of solutions (in finite sample).

+ Computationally simpler (no nested optimization).
- Restricted to Z-estimator.

Summary

The theoretical results presented here aim at answering the following questions:

- **1) Computational issues**: While solving directly the matching may be of interest, it may be extremely computationally demanding. This is a well-known issues of indirect (simulation-based) estimator.
- **2) Properties of $\hat{\theta}^*$**: In addition to being relatively easy to solve, $\hat{\theta}^*$ enjoys the following properties
  - consistency (see Mucyo's presentation),
  - asymptotic normality (not discussed in this presentation),
  - unbiasedness in finite samples (see Mucyo's presentation),
  - accurate inference in finite sample (see next slides).

Confidence distribution - definition

We propose to use $\hat{\theta}^*(\omega_1), \cdots, \hat{\theta}^*(\omega_B), \cdots \sim F_{\hat{\theta}^*}$ to make inference on θ_0.

Definition - confidence distribution (Schweder and Hjort (2002) JSJ)

A data-dependent function $H_n(\cdot)$ is called a confidence distribution for θ_0 if it is a continuous cumulative distribution function and $H_n(\theta_0)$ has a uniform distribution.

Main motivation is a confidence distribution has accurate percentiles:

$$\Pr(\theta_0 \leq H_n^{-1}(\alpha)) = \Pr(H_n(\theta_0) \leq \alpha) = \alpha$$

Therefore, it can be used for:

- Percentile confidence intervals (simple and invariant to 1-1 transformation).
- Point estimator: median is median unbiased.
- Hypothesis testing.

Examples

(From Singh, Xie and Strawderman (2005) AoS)

- Normal mean: $X_1, \cdots, X_n \sim \mathcal{N}(\theta_0, \sigma^2)$, we have $H_n(\theta) = \Phi\left\{\frac{\theta - \bar{X}}{\sigma/\sqrt{n}}\right\}$, where $\bar{X}$ is sample mean, is a confidence distribution.
- Bootstrap distribution: If the bootstrap is consistent $n^{-\alpha}(\hat{\theta}_B - \hat{\theta})$ and $n^{-\alpha}(\hat{\theta} - \theta_0)$ have the same (asymptotic) distribution. An (asymptotic) confidence distribution is $H_n(\theta) = \Pr_B(\hat{\theta}_B - \hat{\theta} \leq \hat{\theta} - \theta)$ where $\Pr_B(\cdot)$ is the probability measure induced by bootstrapping.

Accurate inference in finite sample

Theorem 1 (accurate inference)

If $\Psi\{X(\theta,\omega),\pi\} = \varphi(\theta,\xi,\pi)$ where ξ is a continuous random variable of dimension $\dim(\theta)$ (data reduction) plus some smoothness and identifiability assumptions on φ, then we have that $F_{\hat{\theta}^*}$ is a confidence distribution (componentwise).

Sketch of the proof:

1. From data reduction we have $\hat{\theta}^*(\omega_1),\cdots \sim \hat{\theta}^*(\xi_1),\cdots$
2. Using smoothness of φ, by implicit function theorem plus change of variable we obtain
$$f_\xi(\xi) = f_{\hat{\theta}^*}(\theta)\frac{\det(D_\xi\varphi)}{\det(D_\theta\varphi)}$$
3. Some identifiability condition implies that $f_{\hat{\theta}^*}(\theta_0) \propto f_\xi(\xi_0)$
4. ξ and ξ_0 are independent copies from the same distribution.

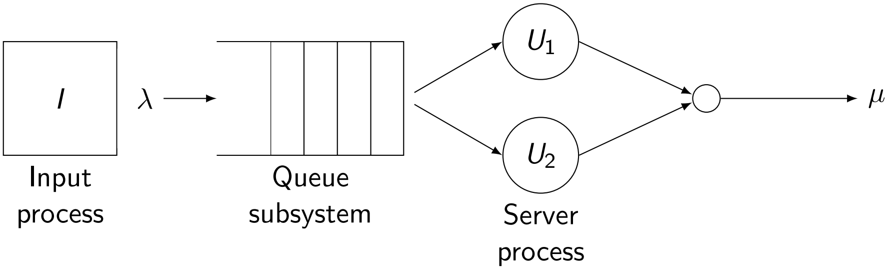
Pareto

Suppose $X_1, \cdots, X_n$ follows a Pareto distribution with parameter minimum η and shape β. The density is $f_X(x) = \frac{\beta\eta^\beta}{x^{\beta+1}}$ if $x \geq \eta$. We use $\hat{\pi} = \{\min(X), \sum_{i=1}^n \log(X_i/\min(X))\}^T$, the sufficient statistics. We can demonstrate that conditions of Theorem 1 hold (not shown here).

n	2	4	8	16
		Coverage (nominal: 95%)		
η	95.02	94.97	94.95	95.04
β	95.10	94.84	95.08	95.02
		Median bias		
η	$1.9 \cdot 10^{-4}$	$-1.8 \cdot 10^{-4}$	$-3.2 \cdot 10^{-4}$	$-7.9 \cdot 10^{-5}$
β	$-2.7 \cdot 10^{-3}$	$-1.1 \cdot 10^{-2}$	$1.2 \cdot 10^{-3}$	$-4.3 \cdot 10^{-3}$

Table: Empirical evidences on 10^5 samples, simulated with parameter values $\eta = 2.5$ and $\beta = 4$. Coverage probabilities target a nominal value of 95%.

Example of queueing process:



Credit to https://tex.stackexchange.com/questions/168113/multi-server-queuing-system-using-tikz

For the M/G/1, we suppose

- Input process follows an exponential distribution (unknown rate).
- Server process follows a uniform distribution (unknown lower and upper bounds).
- We observe only the ouput of the process.

Simulation: M/G/1

This process is expressed, for $i = 1, \cdots, n$, as

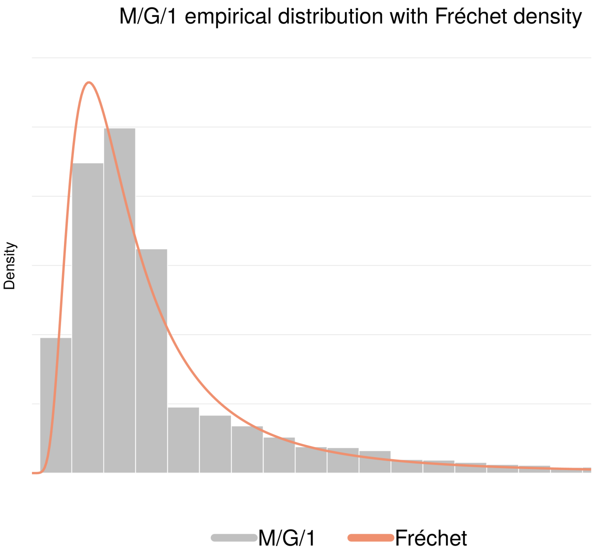
$$x_i = \begin{cases} v_i, & \text{if } \sigma_i^\omega \leq \sigma_{i-1}^x, \\ v_i + \sigma_i^\omega - \sigma_{i-1}^x, & \text{if } \sigma_i^\omega > \sigma_{i-1}^x, \end{cases}$$

where $\sigma_i^\omega = \sum_{j=1}^i \omega_j$, $\sigma_i^x = \sum_{j=1}^i x_j$, $v_i \sim \text{Unif}(\theta_1, \theta_2)$, $0 \leq \theta_1 < \theta_2 < \infty$ and $\omega_i \sim \text{Exp}(\theta_3)$.

For the ith customer: x_i represents interdeparture time, v_i is service time and ω_i corresponds to interarrival time.

There is no tractable likelihood $\Rightarrow$ no easy-to-obtained consistent estimator. (See

Simulation: M/G/1 - choosing Ψ



Simulation: M/G/1 - results

Setup: $\theta_0 = (0.3,\ 0.9,\ 1.0)^T$, number of values: $B = 10,000$.

nominal level:	90%			95%			99%		
n	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
25	91.98	83.96	83.66	93.70	85.41	84.96	94.05	85.42	85.00
50	91.64	99.91	90.79	96.35	99.98	95.39	99.56	100.00	99.32
100	90.52	99.82	89.52	95.37	100.00	94.89	99.09	100.00	99.03
250	89.61	95.59	89.81	94.83	99.04	95.08	99.05	100.00	98.87
500	89.17	91.57	89.44	94.46	96.62	94.62	98.83	99.67	98.96

Numerical evaluation of coverage probabilities (10,000 simulations).

Thank you very much for your attention!

Main references

1. *A simple recipe for making accurate parametric inference in finite sample*, Guerrier, S., Karamera, M., Orso, S. & Victoria-Feser, M.-P., Manuscript available on arXiv: https://arxiv.org/abs/1901.06750
2. *Phase Transition Unbiased Estimation in High Dimensional Settings*, Guerrier, S., Karemera, M., Orso, S. & Victoria-Feser, M.-P., Manuscript available on arXiv: https://arxiv.org/abs/1907.11541

More info...

https://data-analytics-lab.net/
Samuel.Orso@unige.ch

References

Simulation-based matching

C. Gourieroux, A. Monfort, E. Renault, *Indirect inference*, Journal of Applied Econometrics, 1993.

K Heggland, A Frigessi, *Estimating functions in indirect inference*, JRSSB, 2004.

P Fearnhead, D Prangle, *Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation*, JRSSB, 2012.

Confidence distribution

T Schweder, NL Hjort, *Confidence and Likelihood*, Scandinavian Journal of Statistics, 2002.

K Singh, M Xie, WE Strawderman, *Combining information from independent sources through confidence distributions*, The Annals of Statistics, 2005.

M Xie, K Singh, *Confidence distribution, the frequentist distribution estimator of a parameter: A review*, International Statistical Review, 2013.