Phase Transition Unbiased Estimation in High Dimensional Settings

Mucyo Karemera

University of Geneva

joint work with
S. Guerrier, S. Orso & M.-P. Victoria-Feser

December 16, 2019

Introduction

Motivation:

- The need for statistical methods that control the finite sample and asymptotic biases of estimators is growing.
- Finite sample biases can arise in numerous situations but they become increasingly common in situations where the ratio p/n is "large" (e.g. logistic regression).
- Asymptotic biases are also becoming increasingly common due to the ever increasing data size and model complexity. Indeed, an important challenge encountered in this case are the numerical aspects of the estimation procedure.
- To carry out estimation, approximate methods are increasingly used in practice. These methods are typically easier to implement although they can lead to asymptotically biased estimators.

Our Goal

Find a general method that correct the bias while being simple to implement.

Proposed Approach

General setting:

Consider a random sample $\mathbf{Y}(\theta_0, n, \omega_0) \in \mathbb{R}^n$ being generated under model F_{θ_0} (possibly conditional on a set of fixed covariates). Our goal is to estimate the unknown parameter θ_0 using a **biased** (asymptotically and/or in finite samples) estimator that is either readily available or can easily be computed.

"Observed" (random) sample

True parameter value, $\theta_0 \in \Theta \subset \mathbb{R}^p$

$$\mathbf{Y}\left(\theta_0, n, \omega_0\right) \sim F_{\theta_0}$$

Sample size, $\mathbf{Y}(\theta_0, n, \omega_0) \in \mathbb{R}^n$

True seed $\omega_0 \in \mathbb{R}^m$

EXAMPLE: Let $Y_i \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $i = 1, \ldots, n$. Then we could write:

$$Y_i(\theta_0, n, \omega_{0,i}) = \mu_0 + \sigma_0 \omega_{0,i}, \quad \omega_{0,i} \sim \mathcal{N}(0,1).$$

Proposed Approach

We can also consider **simulated samples** denoted as $\mathbf{Y}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h) \in \mathbb{R}^n$, $h \in \mathbb{N}^*$ (where $\mathbb{N}^* \equiv \mathbb{N} \setminus \{0\}$):

Simulated sample

Generic parameter value, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$

$$\mathbf{Y}\left(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h\right) \sim F_{\boldsymbol{\theta}}$$

Sample size, $\mathbf{Y}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h) \in \mathbb{R}^n$

Generic seed $\boldsymbol{\omega}_h \in \mathbb{R}^m$

Remarks:

- The use of seeds $\boldsymbol{\omega}$ allows to explicitly display the randomness of a sample.
- The difference (or equivalence) between two simulated samples, say $\mathbf{Y}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_{h_1})$ and $\mathbf{Y}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_{h_2})$, clearly depends on the seed.

Proposed Approach - JINI Estimator

Suppose that $\hat{\boldsymbol{\pi}}\left(\mathbf{Y}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0)\right) \in \boldsymbol{\Theta}$ is an available estimator of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. For simplicity, we'll write $\hat{\boldsymbol{\pi}}\left(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0\right) := \hat{\boldsymbol{\pi}}\left(\mathbf{Y}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0)\right)$. **This estimator may be biased (asymptotically and/or in finite sample).** We propose to target $\boldsymbol{\theta}_0$ indirectly through $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0)$, by means of $\hat{\boldsymbol{\theta}}_H$ defined as

$$\hat{\boldsymbol{\theta}}_H \in \hat{\boldsymbol{\Theta}}_H := \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argzero}} \; \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0) - \frac{1}{H} \sum_{h=1}^{H} \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h), \tag{1}$$

where $H \in \mathbb{N}^*$ and $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h)$ is computed on the simulated sample $\mathbf{Y}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h)$. We call this estimator a **Just Identified iNdirect Inference Estimator or simply JINI Estimator**.

Remarks:

- We are assuming that the solution set $\hat{\boldsymbol{\Theta}}_H$ is not empty (and that no solution is outside $\boldsymbol{\Theta}$).
- We also assume that the sequence $\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_H$ is fixed.

Summary

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

- **1) Computational issues**: While the approach based on $\hat{\boldsymbol{\theta}}_H$ may be of interest, it may be extremely computationally demanding. This is a well-known issues of indirect (simulation-based) estimator. Theorem 1 shows that **computationally efficient algorithm**, i.e. the Iterative Bootstrap (IB) algorithm proposed by Kuk (1995), can be used to compute $\hat{\boldsymbol{\theta}}_H$.

- **2) Properties of $\hat{\boldsymbol{\theta}}_H$**: In addition to being computable, $\hat{\boldsymbol{\theta}}_H$ enjoys the following properties
  - **Consistency** by Theorem 1,
  - **Phase Transition (PT) unbiasedness** by Theorem 4,
  - asymptotic normality (not discussed in this presentation).

Algorithm for $\hat{\boldsymbol{\theta}}_H$ & Consistency - Main Result

Theorem 1

Under certain assumptions and for $n \in \mathbb{N}$ large enough, the ***Iterative Bootstrap (IB) algorithm*** *defined as*

$$\hat{\boldsymbol{\theta}}_H^{(k)} := \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0) - \left[\frac{1}{H}\sum_{h=1}^{H} \hat{\boldsymbol{\pi}}\left(\hat{\boldsymbol{\theta}}_H^{(k-1)}, n, \boldsymbol{\omega}_h\right) - \hat{\boldsymbol{\theta}}_H^{(k-1)}\right],$$

where $\hat{\boldsymbol{\theta}}_H^{(0)} := \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0)$ is such that

1. $\lim_{k\to\infty} \hat{\boldsymbol{\theta}}_H^{(k)} = \hat{\boldsymbol{\theta}}_H = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argzero}}\ \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n, \boldsymbol{\omega}_0) - \frac{1}{H}\sum_{h=1}^{H} \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, n, \boldsymbol{\omega}_h),$
2. *there exists an $\varepsilon \in (0,1)$ such that for all $k \in \mathbb{N}$ we have*

$$\left\|\hat{\boldsymbol{\theta}}_H^{(k)} - \hat{\boldsymbol{\theta}}_H\right\|_2 = \mathcal{O}_{\mathrm{p}}(p^{1/2}\varepsilon^k),$$

3. $\hat{\boldsymbol{\theta}}_H$ *is a consistent estimator of $\boldsymbol{\theta}_0$, i.e., $\left\|\hat{\boldsymbol{\theta}}_H - \boldsymbol{\theta}_0\right\|_2 = o_{\mathrm{p}}(1)$.*

PT-unbiasedness of $\hat{\boldsymbol{\theta}}_H$ - Definition

It is in general difficult to establish whether an estimator is **unbiased** for all n and therefore the **asymptotic order** of the bias is used to "quantify" its magnitude. The **Phase Transition (PT) unbiasedness** can serve as a middle-ground between unbiasedness and the asymptotic order of the bias.

Definition 2 (Phase Transition unbiasedness)

An estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is said to be **PT-unbiased** if $\left\|\mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}_0\right\|_2 = \mathcal{O}(0)$.

In other words, $\hat{\boldsymbol{\theta}}$ is PT-unbiased if there exists a $n^* \in \mathbb{N}^*$ such that for all $n \in \mathbb{N}^*$ with $n \geq n^*$, we have $\left\|\mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}_0\right\|_2 = 0$.
In particular, it implies that for *all* $\delta \in \mathbb{N}$, $\left\|\mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}_0\right\|_2 = \mathcal{O}\left(n^{-\delta}\right)$.

PT-unbiasedness of $\hat{\theta}_H$ - Insufficent condition

Let $f(n)$ be a real-valued function. Do we have

$$f(n) = \mathcal{O}(n^{-\delta}),\ \ \forall \delta \in \mathbb{N} \ \ \overset{???}{\Longrightarrow} \ \ \exists n^* \in \mathbb{N}^*,\ \text{ s. t. }\ f(n) = 0,\ \ \forall n \geq n^*.$$

In general, the answer is **no**. Indeed, by definition $f(n) = \mathcal{O}(n^{-\delta})$, $\forall \delta \in \mathbb{N}$ if and only if

$$\forall \delta \in \mathbb{N},\ \ \exists n_\delta \geq 0,\ \ \exists M_\delta > 0\ \text{ s. t. }\ \forall n > n_\delta\ \ |f(n)| \leq M_\delta n^{-\delta}. \qquad (2)$$

From this, one can easily construct a step function $f(n)$ such that $f(n) = \mathcal{O}(n^{-\delta}),\ \ \forall \delta \in \mathbb{N}$ but $f(n) \neq 0\ \ \forall n \in \mathbb{N}^*$.
An other example is simply

$$\exp(-n) = \mathcal{O}(n^{-\delta}),\ \ \forall \delta \in \mathbb{N} \text{ but } \exp(-n) \neq 0\ \ \forall n \in \mathbb{N}^*.$$

PT-unbiasedness of $\hat{\boldsymbol{\theta}}_H$ - Sufficient condition

However, by **switching the first two argument** in (2), we get to a sufficient condition to prove the PT-unbiasdness. More precisely, defining $f(n) = \mathcal{O}_{\delta\in\mathbb{N}}(n^{-\delta})$ if and only if

$$\exists n^* \geq 0, \ \ \forall \delta \in \mathbb{N}, \ \ \exists M_\delta > 0 \ \text{ s. t. } \forall n > n^* \ \ |f(n)| \leq M_\delta n^{-\delta}, \tag{3}$$

we have

$$f(n) = \mathcal{O}_{\delta\in\mathbb{N}}(n^{-\delta}) \implies f(n) = \mathcal{O}(0).$$

More generally, we have the following result:

Lemma 3

Let $g(n)$ be a strictly positive real-valued function such that $\lim_{n\to\infty} g(n) = 0$. If $f(n) = \mathcal{O}_{\delta\in\mathbb{N}}\left(g(n)^\delta\right)$ then $f(n) = \mathcal{O}(0)$.

PT-unbiasedness of $\hat{\boldsymbol{\theta}}_H$ - Main Result

Theorem 4

Under certain assumptions, $\left\|\mathbb{E}\left[\hat{\boldsymbol{\theta}}_H\right] - \boldsymbol{\theta}_0\right\|_2 = \mathcal{O}(0)$.

- One of the step in proving Theorem 4 is to show that

$$\left\|\mathbb{E}\left[\hat{\boldsymbol{\theta}}_H\right] - \boldsymbol{\theta}_0\right\|_2 \; = \; \mathcal{O}_{\delta\in\mathbb{N}}\left((p/n)^{\delta}\right).$$

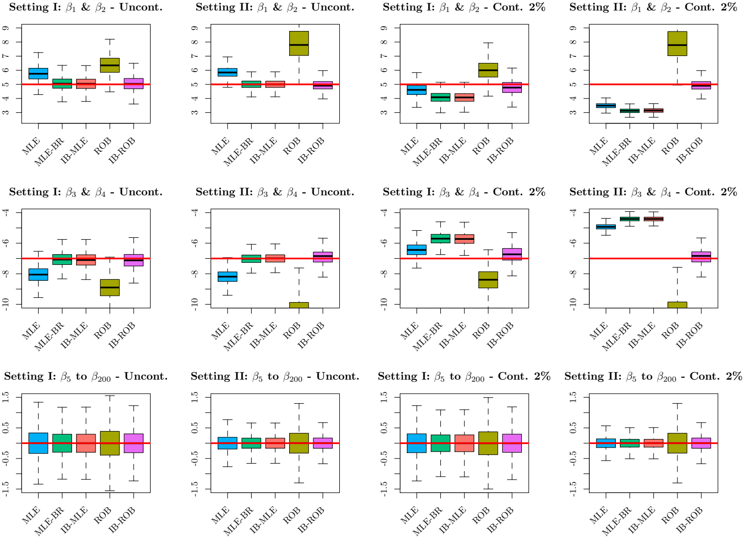
Note that by assumption we have that $p/n \to 0$.

Simulation Study - Logistic Regression

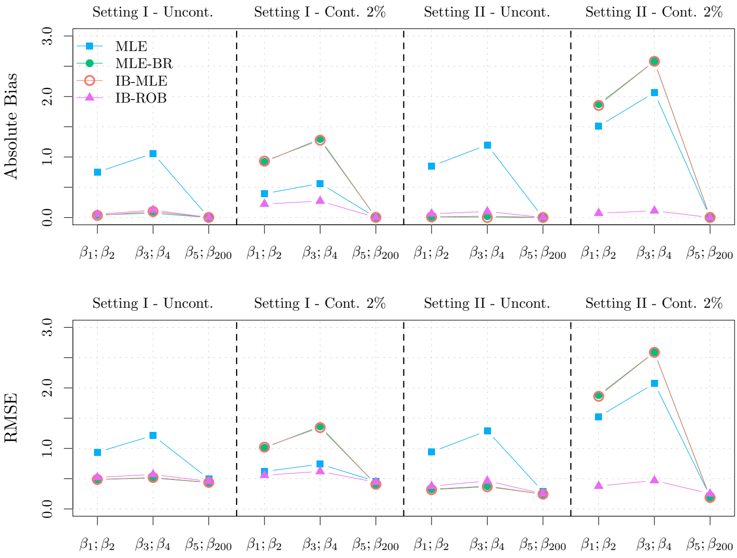
- We consider 5 estimators:
  - **MLE**
  - The bias reduced MLE (**MLE-BR**) using the R *brglm* function (with default parameters) of the *brglm* package (see Kosmidis et al., 2017)
  - JINI based on the MLE computed on the transformed $\tilde{\mathbf{Y}}$ (**IB-MLE**).
  - Robust estimator (**ROB**) using the function *glmrob*.
  - JINI based on ROB computed on the transformed $\tilde{\mathbf{Y}}$ (**IB-ROB**).
- Two simulation settings (with and without contamination) which are presented in the table below. Note that the covariates were simulated independently from a $\mathcal{N}(\mu, 4/\sqrt{n})$ in order to ensure that $\mathbb{E}[\mathbf{Y}_i]$ is well "distributed" between 0 or 1.

Parameters	Setting I	Setting II
$p =$	200	200
$n =$	2000	3000
μ	0	0.6
EPV $\approx$	5	3.75
$\beta_1 = \beta_2 =$	5	5
$\beta_3 = \beta_4 =$	-7	-7
$\beta_5 = \ldots = \beta_{200} =$	0	0
$\delta =$	0.01	0.01

Simulation Study - Logistic Regression



Simulation Study - Logistic Regression

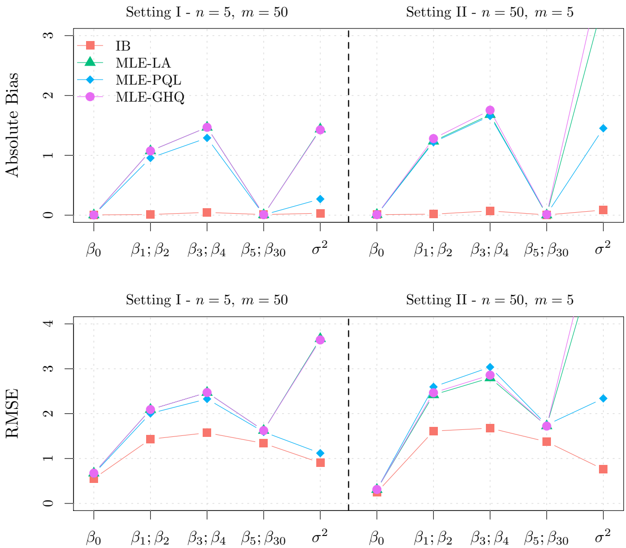


Simulation Study - GLMM

- We consider a **random intercept logistic regression model** and four estimators:
  - **MLE-LA**: MLE based LAplace approximations.
  - **MLE-PQL**: MLE based on the pseudo- and Penalized Quasi-Likelihood.
  - **MLE-GHQ**: MLE based adaptive Gauss-Hermite Quadrature.
  - **IB**: IB estimator based on the penalized iteratively reweighted least squares estimator (a less accurate approximation of the MLE) and computed on the transformed $\tilde{\mathbf{Y}}$.

Parameters	Setting I	Setting II
$p = q + 2$	32	32
$n =$	5	50
$\forall i,\ m_i = m =$	50	5
$N =$	250	250
EPV $\approx$	4	4
$\beta_0 =$	0	0
$\beta_1 = \beta_2 =$	5	5
$\beta_3 = \beta_4 =$	-7	-7
$\beta_5 = \ldots = \beta_{30} =$	0	0
$\sigma^2 =$	1.5	1.5
$\delta =$	0.01	0.01

Simulation Study - GLMM



Thank you very much for your attention!

Main references

1. *Simulation-Based Bias Correction Methods for Complex Models*, Guerrier, S., Dupuis-Lozeron, E., Ma, Y. & Victoria-Feser, M.-P., JASA, 2019
2. *Phase Transition Unbiased Estimation in High Dimensional Settings*, Guerrier, S., Karemera, M., Orso, S. & Victoria-Feser, M.-P., Submitted manuscript available on arXiv: https://arxiv.org/abs/1907.11541

Any questions?

More info...

https://data-analytics-lab.net/

Mucyo.Karemera@unige.ch

References

I. Kosmidis, A. Guolo, and C. Varin. Improving the accuracy of likelihood-based inference in meta-analysis and meta-regression. *Biometrika*, 104:489–496, 2017.

A. Y. C. Kuk. Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 395–407, 1995.