Ont the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses

Julian Sampedro

2017 (updated 2025)

Overview

(*i*) Drawing attention on the importatnce of reporting Monte Carlo Error (MCE) in statistical simulation studies.

(*ii*) Define simple and practical tools for estimating MCE.

(*iii*) Providing means for determining the number of replications R required to achieve a prespecified desired level of accuracy.

Notes: *The uncertainty as measured by Monte Carlo Error (MCE) or sampling variability - in published literature is mostly not reported. It variability comes from simulations and number of simulations. This uncertainty comes from the fact that statistical procedures (measures) are repeated on simulated data (finite sample from an infinite population). Thus two different simulation runs will yield differents results. According to a survey (of the same authors) published in 2007 in Biometrics, Biometrika and JASA, out of 223 articles where simulations studies where performed, only 8 reported a formal justification of the number of replications used or a MCE (standard deviation).*

Simulation study: experiment 1 (1/4)

Setting: Quantification of association between Y and X, two binary r.v. We assume the following generating process:

$$\text{logit } P(Y = 1 \mid X) = \beta_0 + \beta_1 X$$

The MCE is defined as:

$$MCE(\hat{\phi}_R) = \sqrt{\text{var}(\hat{\phi}_R)}$$

$N = 100$ (sample size), $R = 100, 500, 1,000, 2,500, 5,000, 10,000$ (number of replicates), $M = 1,000$ (number of simulations)

Notes: *In a context of logistic regression with known values for the parameters $\beta_0 = -1$ and $\beta_1 = ln(2)$, we want to have an estimation of thre operating characteristics of the Maximum Likelihood Estimator (MLE) for the slope parameter β_1 (log-odds ratio). The MCE is the standard deviation of the estimated quantities of interest (operating characteristics).*

Simulation study: experiment 1 (2/4)

The three chosen operating characteristics:

$\hat{\phi}_R^b$, a MC estimate of **percent bias** for MLE of β_X:

$$\hat{\phi}_R^b = \frac{1}{R}\sum_{r=1}^{R}\frac{\hat{\beta}_X^r - \hat{\beta}_X}{\hat{\beta}_X}$$

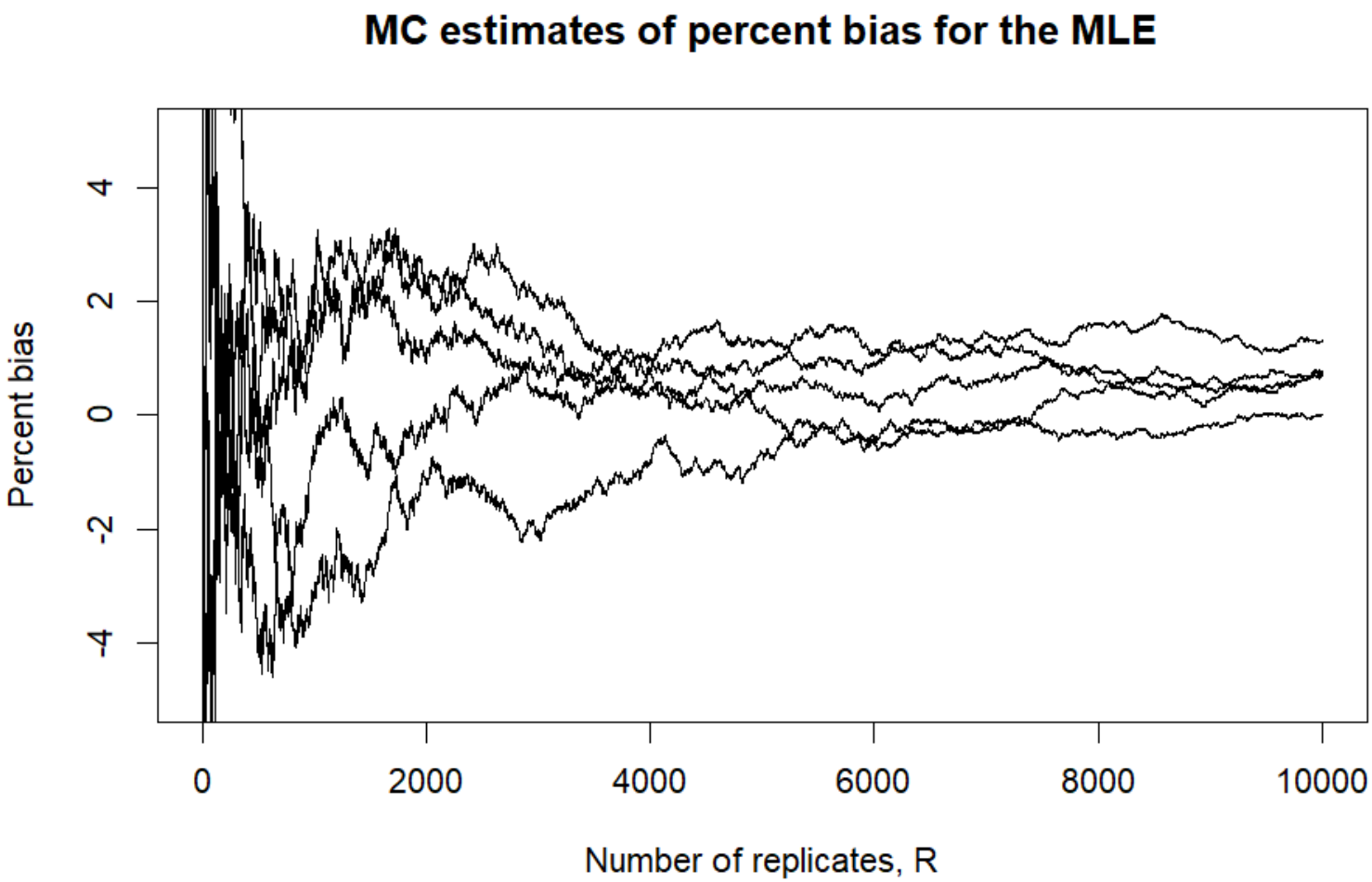
$\hat{\phi}_R^c$, a MC estimate of **coverage probability of the** 95% **CI:**:

$$\hat{\phi}_R^c = \sum_{r=1}^{R} 1\left[\hat{\beta}_X^r - 1.96se(\hat{\beta}_X^r) \leq \beta_X \leq \hat{\beta}_X^r + 1.96se(\hat{\beta}_X^r)\right]$$

$\hat{\phi}_R^p$, a MC estimate of **power** to detect an association:

$$\hat{\phi}_R^b = \sum_{r=1}^{R} 1\left[\frac{\hat{\beta}_X^r}{se(\hat{\beta}_X^r)} < 1.96\right]$$

Simulation study: experiment 1 (3/4)



Notes: *Still a surprising amount of variability for the MC estimate of percent bias for the MLE after 10000 replications. The range of percent bias (after 10000 replicates) is around 2 (this range is also a random variable).*

Simulation study: experiment 1 (4/4))

Operating characteristic	R	min.	max.	mean	MCE
Percent bias	*100*	-22.48	24.42	1.03	6.82
	500	-10.04	11.76	0.92	3.21
	1,000	-6.91	7.74	0.98	2.32
	2,500	-3.71	4.80	0.96	1.36
	5,000	-2.38	3.63	0.90	0.95
	10,000	-1.52	3.09	0.90	0.66
Coverage rate	*100*	85.00	100.00	94.65	2.23
	500	90.80	97.40	94.52	0.99
	1,000	92.30	96.50	94.55	0.71
	2,500	93.04	96.00	94.53	0.48
	5,000	93.22	95.52	94.50	0.36
	10,000	93.57	99.36	94.50	0.32
Power	*100*	20.00	50.00	33.16	4.68
	500	26.20	39.40	32.99	2.16
	1,000	28.40	37.70	33.04	1.55
	2,500	29.32	36.52	33.03	0.96
	5,000	30.70	35.26	32.99	0.66
	10,000	31.44	34.39	32.99	0.46

Note: *I use* $M = 1,000$ *(instead of* $500,000$*). Still, the results are close. Main differences: larger min and smaller max (i.e. for* $R = 100$*.The authors get* -34 *and* 35*). I have about the same MCE. It suggest that more than* $R = 10,000$*to be within one unit* -0.1% *to* 1.9%*,* 95% *on average. Cov. rate means are different (they obtain* 95.3*). Also, less MCE, indicating that* $R = 2,500$ *is optimal (within one unit of the true value* 95% *of the time).*

Simulation study: experiment 2 (1/2)

Setting: Quantification of MCE.

By the strong LLN, we have that: $\hat{\phi}_R \to E[\phi(X)] = \phi$. Also, from the CLT, we know that: $\sqrt{}(\hat{\phi}_R - \phi) \to N(0, \sigma^2_\phi)$

We will consider two measures:

$$\widehat{MCE_{clt}}_{\hat{\phi}_R} = \frac{\hat{\sigma}_\phi}{R} = \frac{1}{R}\sqrt{\sum_{r=1}^{R}\left(\phi(X_r) - \hat{\phi}_R\right)^2}$$

$$\widehat{MCE_{boot}}_{\hat{\phi}_R} = \frac{1}{B}\sqrt{\sum_{b=1}^{B}\left(\hat{\phi}_R(\mathbf{X}^*_b) - \overline{\hat{\phi}_R(\mathbf{X}^*_b)}\right)^2}$$

Note: *A MCE bootstrap estimate is the standard deviation across bootstrap statistics. We use* $B = 100, 200$ *and* 500 *for* $\hat{MCEboot}_{\hat{\phi}_R}$.

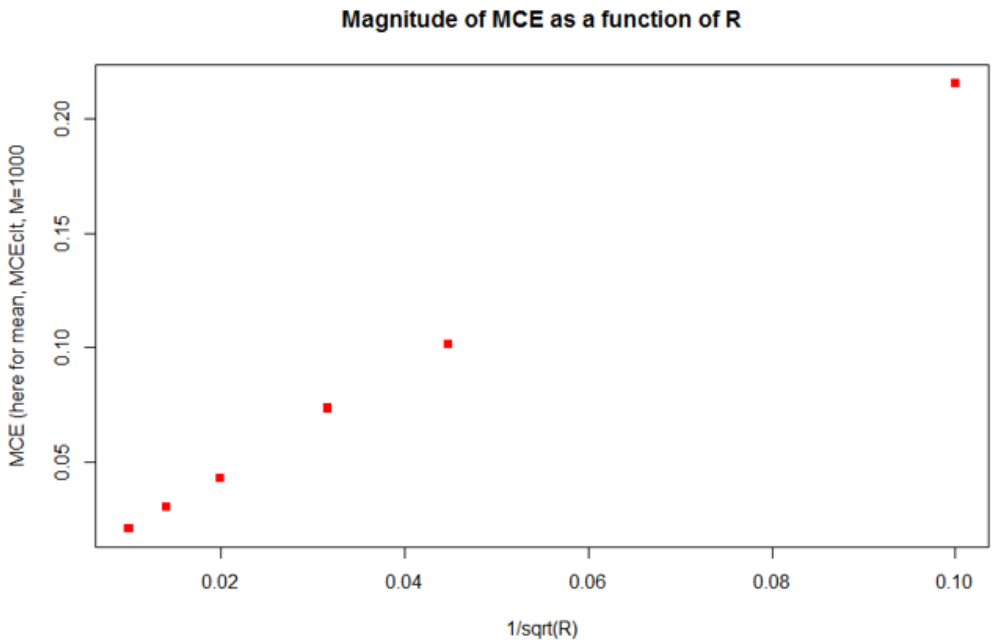
Simulation study: experiment 2 (2/2)

Characteristic	R	value	$\widehat{MCE}_{clt\,\hat{\phi}_R}$	$\widehat{MCE}_{boot\,\hat{\phi}_R}$		
				B = 100	**B** = 200	**B** = 500
Mean	*100*	1.0254	0.2156	0.2210	0.2268	0.2128
	500	0.9248	0.1016	0.0958	0.1027	0.1038
	1,000	0.9810	0.0734	0.0724	0.0782	0.0727
	2,500	0.95572	0.0430	0.0416	0.0397	0.0432
	5,000	0.9045	0.0302	0.0308	0.0307	0.0297
	10,000	0.8984	0.0208	0.0209	0.0224	0.0209
MCE	*100*	6.8191	NA	0.1601	0.1548	0.1600
	500	3.2131	NA	0.0678	0.0660	0.0699
	1,000	2.3205	NA	0.0482	0.0524	0.0501
	2,500	1.3603	NA	0.0291	0.0279	0.0293
	5,000	0.9540	NA	0.0211	0.0191	0.0201
	10,000	0.6571	NA	0.0146	0.0128	0.0143

Note: *Consistent with what the authors obtained. The MCE is reduced in all cases by about* 90% *when we increase R from* 100 *to* $10,000$*, whatever the number of simulations M. The authors use* $M = 500,000$ *but I use* $M = 1,000$ *for computational reasons. We only give point estimation (but could also have reported CI's).*

Remarks

- The magnitude of MCE seems to be linear in $1/\sqrt{R}$
- As $1/\sqrt{R}$ tends to 0, MCE tends to 0
- Further work: Use a constrainde linear regression to estimate R to achieve a desired level of MCE.



Notes: *As $1/\sqrt{R}$ tends to 0, MCE will tend to 0 (useful to assess desired proportion of MCE). The authors proposed an R package called MCE (which has been archived) implementing the BCG plot. The estimated slope helps estimate the number of replications R to achieve an acceptable MCE.*

Conclusions

(*i*) The MCE can be more substantial than traditionally thought.

(*ii*) Even after $500,000$ simulations, there is residual uncertainty.

(*iii*) The magnitude of the MCE depends on various factors, i.e. parameters, operating characteristics, variability in the data.

(*iv*) The MCE can be drastically reduced by increasing R, the number of replicates.

(*iv*) In simulations of most published scientific studies, the MCE is rarely reported though documenting uncertainty (usually by standard errors, p-values and CIs) is commonly insisted on.

Note: *In more complex settings, the MCE is more complicated to estimate.*

References

Elizabeth KOEHLER, Elizabeth BROWN, and Sebastien J.-P. A. HANEUSE, 2009. On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses.

Christian P. ROBERT, George CASELLA, 2005. Monte Carlo Statistical Methods.

Maria L. RIZZO, 2008. Statistical computing with R.

F. H. C. MARRIOTT, 1979. Barnard's Monte Carlo Tests: How Many Simulations?

Full R code of this reproduced study available upon request.