Non-asymptotic Optimal Stopping Criteria for Monte Carlo

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Christian Bayer[1], Håkon Hoel[2], Erik von Schwerin[3], and Raúl Tempone[2]

[1] Weierstrass Institute, Berlin, Germany, [2] KAUST, Thuwal, Saudi Arabia, [3] EPF de Lausanne, Switzerland,

Center for Uncertainty Quantification

christian.bayer@wias-berlin.de ,{hakon.hoel,raul.tempone}@kaust.edu.sa, erik.vonschwerin@epfl.ch

Problem Formulation

Monte Carlo sampling: Given i.i.d. random variables $X_1, X_2, \ldots$ approximate their expected value $\mu = \mathrm{E}[X]$ from M samples using the sample average

$$\overline{X}_M := \sum_{i=1}^{M} \frac{X_i}{M}.$$

Objective: Choose M sufficiently large for the error probability to satisfy

$$\mathrm{P}\left(|\overline{X}_M - \mu| > TOL\right) \le \delta, \qquad (1)$$

for some fixed small constants $TOL > 0$ and $\delta > 0$.

Algorithmic Control of the Number of Samples

Without (much) a-priori information on the distribution of X_i not much is known about theoretical upper bounds of M. Typically the number of samples needed is determined through a sequential stopping rule, such as:

(I) Generate a batch of M i.i.d. samples $X_1, X_2, \ldots, X_M$.

(II) Infer distributive properties of $\overline{X}_M$ from the generated batch of samples through higher order sample moments, e.g. the sample mean and the sample variance.

(III) Based on the sample moments, estimate the error probability. When, based on the estimated probability, (1) is violated, increase the number of samples M and return to step (I). Else, break and accept M.

Asymptotically Consistent Stopping Rules

The most common stopping rule is based on the Central Limit Theorem (CLT) and uses sample estimates of $\mathrm{E}[X]$ and $\mathrm{Var}(X)$. It assumes only the existence of the second moment of the X_i.

Algorithm 1 – Sample Variance Based Stopping Rule

Data: Initial number of samples M_0, accuracy TOL, confidence δ, and the cumulative distribution function of the standard normal distribution $\Phi(x)$

Result: The output sample mean $\overline{X}_M$

Initialization: Set $n = 0$, generate M_n samples $\{X_i\}_{i=1}^{M_n}$ and compute the sample variance as follows

$$\overline{\sigma}^2_{M_n} := \frac{1}{M_n - 1} \sum_{i=1}^{M_n} (X_i - \overline{X}_{M_n})^2; \qquad (2)$$

while $2\left(1 - \Phi(\sqrt{M_n} TOL/\overline{\sigma}_{M_n})\right) > \delta$ **do**

Set $n = n + 1$ and $M_n = 2M_{n-1}$;

Generate a batch of M_n i.i.d. samples $\{X_i\}_{i=1}^{M_n}$ and compute the sample variance $\overline{\sigma}^2_{M_n}$ as described in (2);

end

Set $M = M_n$, generate samples $\{X_i\}_{i=1}^{M}$ and compute $\overline{X}_M$;

return $\overline{X}_M$

Second moment based stopping rules perform well in the asymptotic regime when $TOL \ll 0$. In fact, under very loose restrictions they are known to be asymptotically consistent in the sense that for a fixed δ,

$$\lim_{TOL \to 0} \mathrm{P}\left(|\overline{X}_M - \mu| > TOL\right) = \delta.$$

See Chow and Robbins [4] and, for more general stochastic processes, Glynn and Whitt [3].

Non-Asymptotic Failure of a Second Moment Based Stopping Rule

In the non-asymptotic regime $TOL = \mathcal{O}(1)$, second moment based stopping rules may fail to meet the goal (1). Let us illustrate this by sampling a sequence of Pareto distributed i.i.d. r.v. X_i with PDF

$$f(x) = \begin{cases} \alpha x_m^{\alpha} x^{-(\alpha+1)}, & \text{if } x \ge x_m, \\ 0, & \text{otherwise,} \end{cases} \qquad (3)$$

where $\alpha, x_m \in \mathbb{R}_+$. Assume further that the variance $\sigma^2 = \mathrm{Var}(X)$ is known prior to sampling, so that Algorithm 1 takes the simple form

$$M = \frac{C^2_{\mathrm{CLT}} \sigma^2}{TOL^2}, \quad C_{\mathrm{CLT}} := \Phi^{-1}\left(\frac{2-\delta}{2}\right). \qquad (4)$$

Figure 1 gives a performance study of this stopping rule for three different confidence functions $\delta(TOL)$. We observe that the smaller δ is relative to TOL, the larger is the probability of failure for the stopping rule.

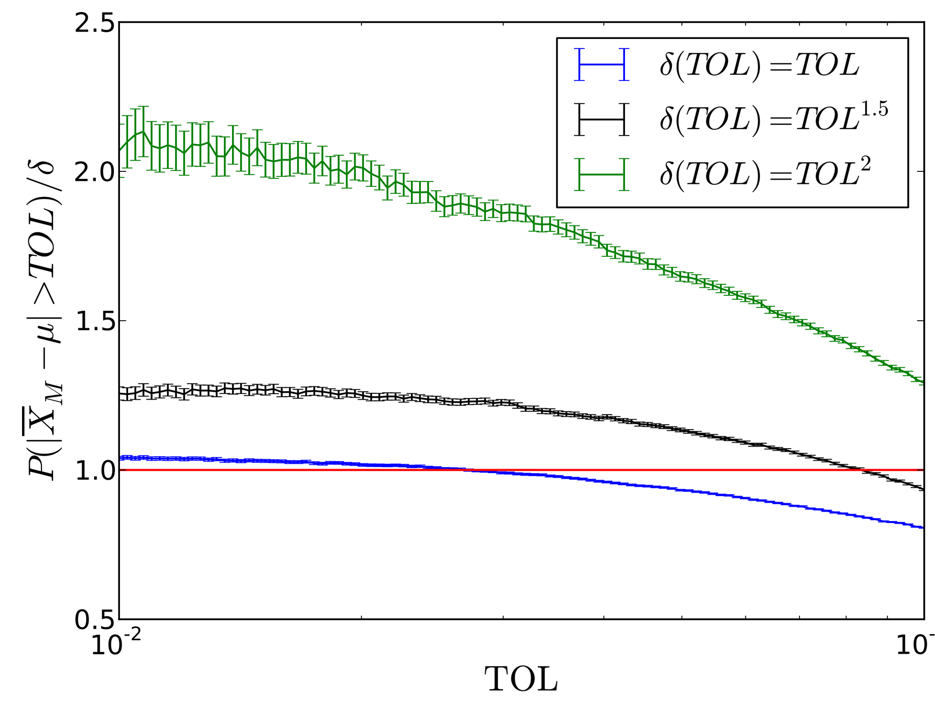


Figure 1: *The scaled probability of failure $\overline{p}_N(TOL,\delta)/\delta$, cf. (5), is plotted with error bars for three different confidence functions $\delta(TOL)$ (blue, black and green lines) in the setting of sampling i.i.d. Pareto r.v. with parameters $\alpha = 3.1$ and $x_m = 1$ and using the stopping rule (4). When $\overline{p}_N(TOL,\delta)/\delta > 1$, the stopping rule is unable to meet the goal (1).*

In the above figure the probability of failure is estimated by

$$\overline{p}_N(TOL,\delta) = N^{-1} \sum_{i=1}^{N} \mathbf{1}_{|\overline{X}_M(\omega_i) - \mu| > TOL}, \qquad (5)$$

using $N = 10^7$ Monte Carlo outer loop samples of $\overline{X}_M(\omega_i)$, and the error bars for the estimate of $\overline{p}_N$ is approximated by $1.96\sqrt{\overline{p}_N(1-\overline{p}_N)/N}$.

Second Moment Based Algorithm for the Non-Asymptotic Regime

Hickernell et al. [2] recently proposed an algorithm that is guaranteed to meet condition (1) provided that an upper bound for the *kurtosis*

$$\kappa = \frac{\mathrm{E}\left[|X - \mu|^4\right]}{\sigma^4} - 3$$

is given prior to sampling. In applications the demand for an a priori bound on κ can however be impractical.

Stopping Rules Based on Higher Moments

While the guaranteed accuracy of the algorithm by Hickernell et al. is highly desirable, the algorithm can sometimes be overly pessimistic leading to excessive computational work. The aim of this work [1] is to construct a sequential stopping rule that uses higher order sample moments to become more reliable than Algorithm 1, but not at a prohibiting increase in the computational cost.

We propose an algorithm based on splitting the probability of the error being greater than TOL conditioned on M

$$\mathrm{P}\left(|\overline{X}_M - \mu| > TOL \,\middle|\, M\right) = \mathrm{P}\left(|\overline{X}_M - \mu| > TOL \,\middle|\, M\right) \mathrm{P}\left(|\overline{\sigma}^2_M - \sigma^2| \ge \sigma^2/2 \,\middle|\, M\right) + \mathrm{P}\left(|\overline{X}_M - \mu| > TOL \,\middle|\, M\right) \mathrm{P}\left(|\overline{\sigma}^2_M - \sigma^2| < \sigma^2/2 \,\middle|\, M\right), \qquad (6)$$

to combine a conservative estimate of $\mathrm{P}\left(|\overline{X}_M - \mu| > TOL \,\middle|\, M\right)$ based on the uniform and non-uniform Berry-Esseen Theorems with a more optimistic bound based on an Edgeworth expansion up to 3rd moment. The estimates are combined by weighting them with probability estimates of the error when estimating the variance by the sample variance, $\overline{\sigma}^2_M$. Using a Chebycheff inequality for k-statistics and Markov's inequality, we bound

$$\mathrm{P}\left(|\overline{\sigma}^2_M - \sigma^2| \ge \sigma^2/2 \,\middle|\, M\right) \le C_p(M) := \min\left(1,\ 4\left(\frac{2}{M-1} + \frac{\kappa}{M}\right), \frac{2^p}{\sigma^{2p}} \mathrm{E}\left[|\overline{\sigma}^2_M - \sigma^2|^p \mid M\right]\right), \qquad (7)$$

using $p = 2$ in the numerical examples. One might take the minimum over several values of $p = p_1, \ldots, p_n > 0$. This leads to an approximate stochastic error bound and to replace the stopping criterion in the **while** loop of Algorithm 1 by

$$2\left(1 - \Phi\left(\frac{\sqrt{M}TOL}{\overline{\sigma}_M}\right)\right) + 2C_{\mathrm{BE}}\left(\frac{\sqrt{M}TOL}{\overline{\sigma}_M}, \overline{\beta}_M\right) \frac{1}{\sqrt{M}} \overline{C}_p(M) + \frac{\left|\frac{MTOL^2}{\overline{\sigma}^2_M} - 1\right| |\widehat{\beta}_M|}{\exp\left(\frac{MTOL^2}{2\overline{\sigma}^2_M}\right) \times 3\sqrt{2\pi M \overline{\sigma}^3_M}} (1 - \overline{C}_p(M)) > \delta, \qquad (8)$$

where $C_{\mathrm{BE}}(\cdot,\cdot)$ represents the constants in the Berry-Esseen theorems and $\overline{C}_p(\cdot)$ is obtained from (7) by replacing all moments of $\overline{\sigma}$ by their empirical counterparts. This leads to the following algorithm based on sample moments of orders 2, 3 and possibly 4 (if the 4th moment is bounded).

Algorithm 2 – Higher Moments Based Stopping Rule

Data: Initial number of samples M_0, accuracy TOL, and confidence δ

Result: $\overline{X}_M$

Initialization: Set $n = 0$, generate i.i.d. samples $\{X_i\}_{i=1}^{M_n}$ and compute the sample moments $\overline{\sigma}_{M_n}$, $\overline{\beta}_{M_n}$, $\widehat{\beta}_{M_n}$ and $\overline{\kappa}_{M_n}$ and all (other) moments needed for $\overline{C}_P$;

while *Inequality* (8) *holds* **do**

Set $n = n + 1$ and $M_n = 2M_{n-1}$;

Generate M_n i.i.d. samples $\{X_i\}_{i=1}^{M_n}$ and compute the sample moments $\overline{\sigma}_{M_n}$, $\overline{\beta}_{M_n}$, and $\overline{\kappa}_{M_n}$ and all (other) moments needed for $\overline{C}_P$;

end

Set $M = M_n$, generate i.i.d. samples $\{X_i\}_{i=1}^{M}$ and compute the sample mean $\overline{X}_M$;

return $\overline{X}_M$

Numerical Examples

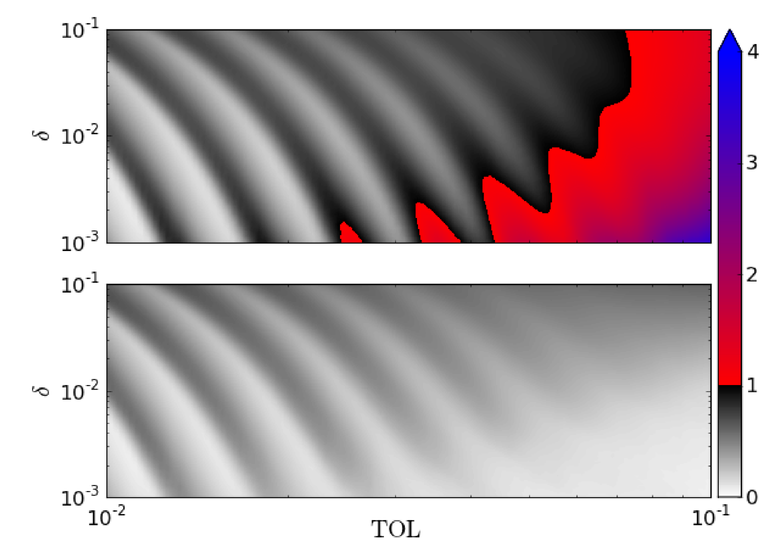
We compare the three presented stopping rules for two heavy tailed distributed r.v. with standard deviation $\sigma = 1$. The first, a Pareto distribution, has a finite 3rd moment, but lacks 4th moment; the second, a Normal-Inverse Gaussian distribution, has finite kurtosis $\kappa = 123$.

The color in the top plots in the examples below show the scaled probability of failure $\overline{p}_N(TOL,\delta)/\delta$ for the respective algorithms, cf. (5), and the bottom plots show the respective algorithms' complexity in terms of the average number of r.v. samples required to generate $\overline{X}_M$ at the given accuracy and confidence pair (TOL,δ). An algorithm is considered unreliable at points (TOL,δ) where $\overline{p}_N(TOL,\delta)/\delta > 1$.

Pareto with $\alpha = 3.1$ and $x_m = (\alpha - 1)\sqrt{(\alpha-2)/\alpha}$

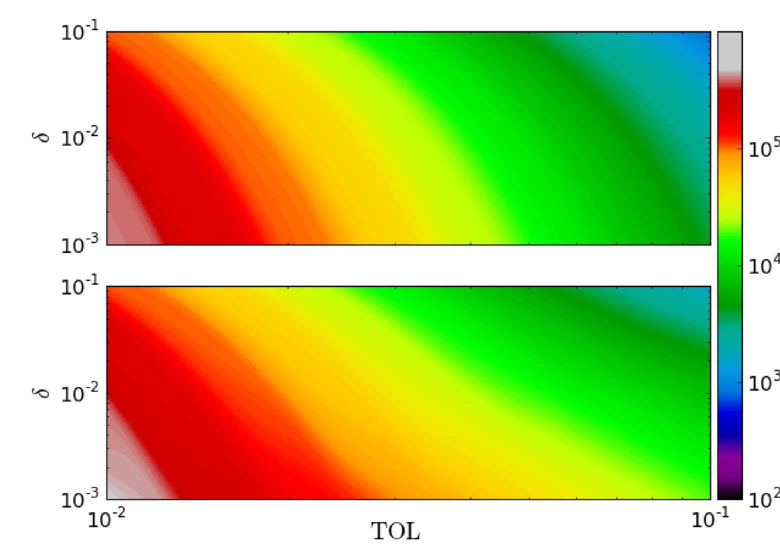
Algorithm 1 $\overline{p}_N(TOL,\delta)/\delta$

Algorithm 2 $\overline{p}_N(TOL,\delta)/\delta$



Algorithm 1 Complexity

Algorithm 2 Complexity



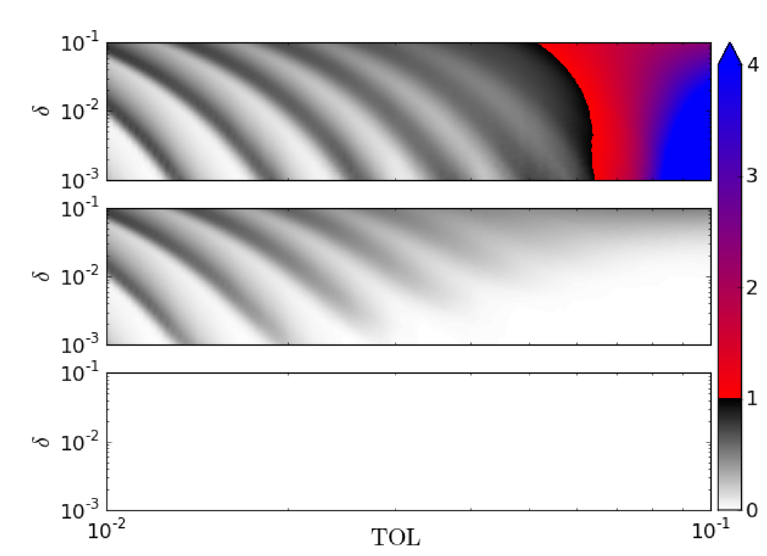
Normal-inverse Gaussian Distribution

Normal-inverse Gaussian distributed r.v. with parameters $\alpha = 3$, $\beta = \sqrt{\alpha^2 - 1}$, $\gamma = 1$, $\delta = \alpha^{-2}$, and $mu = -\beta/\gamma$. This yields r.v. with standard deviation $\sigma = 1$ and kurtosis $\kappa = 123$.

Algorithm 1 $\overline{p}_N(TOL,\delta)/\delta$

Algorithm 2 $\overline{p}_N(TOL,\delta)/\delta$

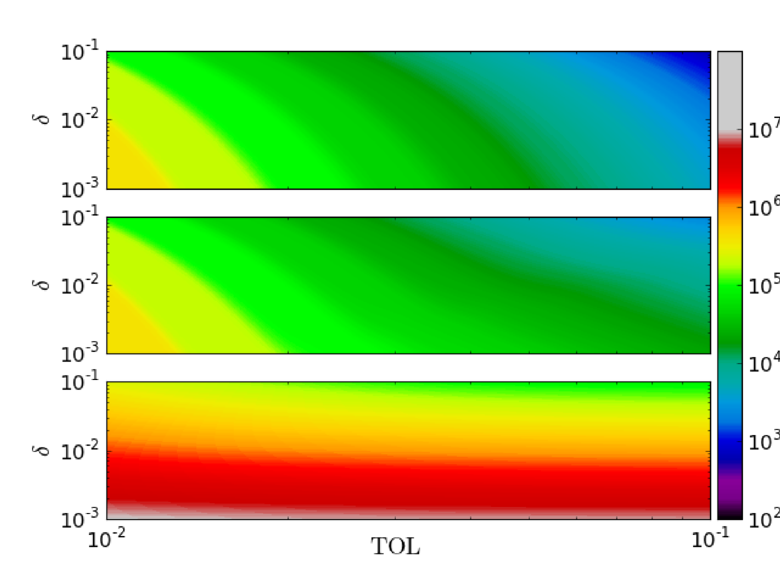
Algorithm 3 Hickernell et al. $\overline{p}_N(TOL,\delta)/\delta$



Algorithm 1 Complexity

Algorithm 2 Complexity

Algorithm 3 Hickernell et al. Complexity



Summary of several distributions

Considered r.v.	Algorithm	Algorithm performance			
		N	$\max \overline{p}_N/\delta$	$\max(\overline{p}_N(1-\overline{p}_N)/N)^{1/2}/\delta$	runtime$/N$
Pareto	Alg 1	$5 \cdot 10^6$	3.529600	0.026522	0.124712 s
($\sigma = 1$ and $\kappa = \infty$)	Alg 2	$5 \cdot 10^6$	0.686309	0.009506	0.189077 s
Normal-inv. Gaussian	Alg 1	$5 \cdot 10^5$	12.014000	0.154076	3.019442 s
	Alg 2	$5 \cdot 10^5$	0.755712	0.028699	3.048026 s
($\sigma = 1$, $\kappa = 123$)	Hick. et al.'s	$5 \cdot 10^4$	0.000419	0.000296	77.322941 s
Exponential $\lambda = 1$,	Alg 1	$5 \cdot 10^6$	0.919659	0.012642	0.033680 s
	Alg 2	$5 \cdot 10^6$	0.912216	0.012040	0.076942 s
($\sigma = 1$ and $\kappa = 6$)	Hick. et al.'s	$5 \cdot 10^6$	0.206530	0.000676	0.229495 s

References

[1] C. Bayer, H. Hoel, E. von Schwerin, and R. Tempone. On non-asymptotic optimal stopping criteria in MC simulations. MATHICSE Technical Report 07.2013, preprint 2013

[2] F. J. Hickernell, L. Jiang, Y. Liu, and A. Owen. Guaranteed Conservative Fixed Width Confidence Intervals Via Monte Carlo Sampling. arXiv 1208.4318, preprint 2012

[3] P. Glynn and W. Whitt. The Asymptotic Validity of Sequential Stopping Rules for Stochastic Simulations. *Ann. Appl. Probab.*, 2(1):180–198, 1992

[4] Y. S. Chow and H. Robbins. On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean. *Ann. Math. Statist.* 36(2), 457–462, 1965.