ON NON-ASYMPTOTIC OPTIMAL STOPPING CRITERIA IN MONTE CARLO SIMULATIONS

CHRISTIAN BAYER, HÅKON HOEL, ERIK VON SCHWERIN, AND RAÚL TEMPONE

ABSTRACT. We consider the setting of estimating the mean of a random variable by a sequential stopping rule Monte Carlo (MC) method. The performance of a typical second moment based sequential stopping rule MC method is shown to be unreliable in such settings both by numerical examples and through analysis. By analysis and approximations, we construct a higher moment based stopping rule which is shown in numerical examples to perform more reliably and only slightly less efficiently than the second moment based stopping rule.

1. INTRODUCTION

Given i.i.d. random variables $X_1, X_2, \ldots$ the typical way of approximating their expected value $\mu = \mathrm{E}[X]$ using M samples is the sample average

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

We consider the objective of choosing M sufficiently large so that the error probability satisfies

$$\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) \leq \delta, \tag{1}$$

for some fixed small constants $\mathrm{TOL} > 0$ and $\delta > 0$. Clearly, $\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big)$ decreases as M increases, but at the same time the cost of computing $\overline{X}_M$ increases. From an application and cost point of view it is therefore of interest to derive theory or algorithms that will give upper bounds on M satisfying (1) that are not far too large. When a-priori information about the distribution of X is available, for example if X is a bounded r.v. with an explicitly given bound, it is possible to derive good theoretical upper bounds for M using Hoeffding type inequalities, cf. Hoeffding [5]. In the general case when no or little information of the distribution is given, little theory is however known, and the typical way of estimating $\mathrm{E}[X]$ using a sufficiently large number of samples M is through a sequential stopping rule. Below we give the general structure of the class of sequential stopping rules we have in mind.

(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.

2010 *Mathematics Subject Classification.* Primary 65C05; Secondary 62L12, 62L15.
Key words and phrases. Monte Carlo methods, optimal stopping, sequential stopping rules, non-asymptotic.

Algorithm 1 Sample Variance Based Stopping Rule

Input: Number of samples M_0, accuracy TOL, confidence δ, the cumulative distribution function of the standard normal distributed r.v. $\Phi(x)$.
Output: $\overline{X}_{M_{\overline{n}}}$.

Set $k = 0$, generate M_k samples $\{X_i\}_{i=1}^{M_k}$ and compute the sample variance

$$\overline{\sigma}^2_{M_k} := \frac{1}{M_k - 1} \sum_{i=1}^{M_k} (X_i - \overline{X}_{M_k})^2. \tag{2}$$

while $2\Big(1 - \Phi(\sqrt{M_k}\,\mathrm{TOL}/\overline{\sigma}_{M_k})\Big) > \delta$ **do**
  Set $k = k + 1$ and $M_k = 2M_{k-1}$.
  Generate a batch of M_k i.i.d. samples $\{X_i\}_{i=1}^{M_k}$.
  Compute the sample variance $\overline{\sigma}^2_{M_k}$ as given in (2).
end while
Set $M_{\overline{n}} = M_k$, generate samples $\{X_i\}_{i=1}^{M_{\overline{n}}}$ and compute the output sample mean $\overline{X}_{M_{\overline{n}}}$. (See Section 2 for a motivation of the choice of the stopping criterion in the while loop above.)

Certainly the most natural and important representative of this class of algorithms is given in Algorithm 1. The algorithm estimates the error probability by appealing to the Central Limit Theorem (CLT). Consequently, it only relies on the sample variance in addition to the sample mean. In particular, the algorithm only requires mild additional assumptions on X, namely square integrability.

In the literature, various second moment based sequential stopping rules have been introduced to estimate the steady-state mean of stochastic processes, see for example Law and Kelton [7, 8] for comparisons of the performance of different stopping rules and Bratley, Bennet, and Fox [1] for an overview. Second moment based sequential stopping rules generally tend to perform well in the asymptotic regime when TOL $\to 0$. In fact, Chow and Robbins [2] proved that under very loose restrictions, second moment based sequential stopping rules such as Algorithm 1 are asymptotically consistent, meaning that for a fixed δ,

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

and in Glynn and Whitt [4] the consistency property is proven to hold for such stopping rules applied to more general stochastic processes. The performance for second moment based stopping rules in the non-asymptotic regime – when TOL and δ are fixed values – is however not as well understood. This is unsatisfactory, as in applications this is precisely the interesting regime, in particular since very often we have TOL $\gg \delta$. While consistency is clearly a re-assuring property in any case, in many situations one is in dire need of quantitative estimates of the error probability in the non-asymptotic regime, for instance when one tries to optimize the computational cost needed to meet a certain accuracy target using an adaptive algorithm. We could not find such a quantitative, non-asymptotic analysis of algorithms like Algorithm 1 in the literature.

In this note we demonstrate by numerical examples that second moment based stopping rules can fail convincingly in the non-asymptotic regime, especially when the underlying distribution X is heavy-tailed, see Section 2. We proceed by giving an error analysis of Algorithm 1 specifically in the non-asymptotic regime. We note a-priori that there are two obvious approximation errors in the underlying assumptions of Algorithm 1:

(I) The algorithm appeals to the CLT to approximate the tail probabilities for $\overline{X}_M$ even though M is finite.

(II) In doing so, it uses the sample variance $\overline{\sigma}_M^2$ instead of the true variance σ^2.

To get a hold on the error probability (1) despite the fact that the distribution of the sample mean $\overline{X}_M$ is unknown, we again appeal to the central limit theorem, but we adjust the estimate by adding a Berry-Esseen type term, which extends the validity of the estimate to the non-asymptotic case, thereby dealing with the first approximation error. As the error probability (1) is a tail probability for the distribution of the sample mean and the Berry-Esseen theorem itself is rather aimed at being sharp at the center of the distribution, we appeal to non-uniform versions of the Berry-Esseen theorem, see Theorem 1.1 and Corollary 1.2 below. However, both intuition and numerical tests suggest that the approximation of the tail probabilities by the non-uniform Berry-Esseen theorem is far too pessimistic at least when the second approximation error is small, i.e., when the computed sample variance is actually close to the true variance. In this case, we adjust the normal distribution by a less stringent extra term, which is obtained from an Edgeworth expansion of the distribution function of the sample mean $\overline{X}_M$, c.f. Feller [3].[1]

Having identified possible origins of failure of Algorithm 1, we propose an improvement of Algorithm 1. However, this variant requires third and fourth sample moments, see Section 4. Finally, in Section 5, we test the new algorithm numerically. We find that the new stopping Algorithm 2 indeed satisfies the desired confidence level δ on the error probability (1) even when $\delta \ll \text{TOL}$.

As already discussed above, we need to approximate the unknown distribution of a sample mean in a general, non-asymptotic regime. The uniform and non-uniform Berry-Esseen theorems provide quantitative bounds for the difference between the true distribution of the sample mean and its asymptotic limit, namely the normal distribution. The following classical theorem can be found, for instance, in Petrov [10].

Theorem 1.1 (Uniform and Non-Uniform Berry-Esseen)**.** *Suppose $X_1, X_2, \dots$ are i.i.d. r.v. with $E[X] = 0$, $\sigma^2 = Var(X)$ and $\beta = \frac{E[|X|^3]}{\sigma^3} < \infty$. Then, for a positive constant C_0, the following uniform bound*

$$\left| P\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i < x \right) - \Phi(x) \right| \leq C_0 \frac{\beta}{\sqrt{n}}$$

holds. For another positive constant C_1, the following non-uniform bound holds

$$\left| P\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i < x \right) - \Phi(x) \right| \leq C_1 \frac{\beta}{\sqrt{n}(1+|x|^3)}.$$

Up to our knowledge, the best upper bounds presently known for the Berry-Esseen constants are $C_0 < 0.4785$, cf. Tyurin [11], and $C_1 < C_0 + 8(1+e^1) < 30.2338$,

[1] Note that here we are introducing a gap in the analysis: the estimate based on the non-uniform Berry-Esseen theorem is reliable in the sense that it always leads to an upper bound of the error probability (1). For the Edgeworth expansion, however, there might be situations when the true error probability is underestimated, and, consequently, the accuracy target might still be missed. Numerical evidence, however, suggests that the estimate obtained from solely relying on the non-uniform Berry-Esseen theorem is usually by orders of magnitude too pessimistic. Apart from intrinsic reasons, one reason might be that the constants known in the non-uniform Berry-Esseen theorems might be far from being optimal. In the end, we think that the above compromise between Berry-Esseen type estimations and estimations based on the Edgeworth expansion might be a good compromise which retains the goal of reliably meeting the accuracy target – except maybe for very extreme situations – while keeping a certain level of efficiency. We note, however, that it is also possible to construct even more conservative stopping rules which are only based on the Berry-Esseen theorem.

cf. Michel [9]. For the purpose of this paper, it will be useful to combine the uniform and non-uniform Berry-Esseen bound as follows.

Corollary 1.2 (Berry-Esseen). *Suppose $X_1, X_2, \dots$ are i.i.d. r.v. with $E[X] = 0$, $\sigma^2 = Var(X)$ and $\beta = E\big[|X|^3\big]/\sigma^3 < \infty$. Then*

$$\left| P\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i < x \right) - \Phi(x) \right| \leq C_{\mathrm{BE}}(x) \frac{\beta}{\sqrt{n}}$$

where the bound function $C_{\mathrm{BE}} : \mathbb{R} \to [0, C_0]$ is defined by

$$C_{\mathrm{BE}}(x) := \min\left(C_0, \frac{C_1}{(1+|x|)^3} \right).$$

In the asymptotic regime, the distribution of $\mathrm{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i < x \right)$ can be expressed by so called Edgeworth expansions. Here we present the one-term Edgeworth expansion.

Theorem 1.3 (Edgeworth expansion, cf. Feller [3]). *Suppose $X_1, X_2, \dots$ are i.i.d. r.v. with a distribution which is not a lattice distribution and $E[X] = 0$, $\sigma^2 = Var(X)$ and $E\big[X^3\big] < \infty$. Then*

$$P\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i < x \right) = \Phi(x) + \frac{(x^2-1)e^{-x^2/2}}{6\sqrt{2\pi n}} \frac{E\big[X^3\big]}{\sigma^3} + o\left(n^{-1/2}\right),$$

uniformly for $x \in \mathbb{R}$.

2. Stopping rule failures

Suppose we seek to estimate $\mu = \mathrm{E}[X]$ using Monte Carlo simulation and we actually *do know* the variance $\sigma^2 = \mathrm{Var}(X)$. As before, our objective is to achieve $\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) \leq \delta$, for some fixed, small constants $\mathrm{TOL}, \delta > 0$. The CLT motivates the stopping rule

$$M = \frac{C_{\mathrm{CLT}}^2 \sigma^2}{\mathrm{TOL}^2}, \quad C_{\mathrm{CLT}} := \Phi^{-1}\left(\frac{2-\delta}{2} \right), \tag{3}$$

which would exactly fulfill our objective (1) in the asymptotic regime $M \to \infty$. Of course, this conflicts with our choice (3) for M, since we treat δ and TOL as finite constants. However, we can still estimate the probability in (1) using Corollary 1.2 and obtain

$$\begin{aligned}
\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) &= \mathrm{P}\left(\sqrt{M} \frac{\big|\overline{X}_M - \mu\big|}{\sigma} > \frac{\sqrt{M}\,\mathrm{TOL}}{\sigma} \right) \\
&\leq 2\left(1 - \Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma} \right) \right) + 2C_{\mathrm{BE}}\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma} \right) \frac{\beta}{\sqrt{M}} \\
&= 2(1 - \Phi(C_{\mathrm{CLT}})) + 2C_{\mathrm{BE}}(C_{\mathrm{CLT}}) \frac{\beta}{\sqrt{M}} \mathrm{TOL} \\
&= \delta + 2\frac{C_{\mathrm{BE}}(C_{\mathrm{CLT}})\,\beta}{\sigma C_{\mathrm{CLT}}} \mathrm{TOL}
\end{aligned} \tag{4}$$

This means that in the worst case, the actual error probability could be $\delta + \mathcal{O}(\mathrm{TOL})$ instead of δ [2]. For instance, in situations where the statistical confidence in the result is more stringent than the accuracy so that $\delta \ll \mathrm{TOL}$, the asymptotically motivated choice of M in (3) could, granted the bound (4) is sharp, fail to deliver

[2] Note that C_{CLT} as defined in (3) grows only very slowly as δ decreases, since we have $C_{\mathrm{CLT}} < \sqrt{2\log(\delta^{-1})}$. Thus, the factor in front of TOL in the error probability can almost be neglected.

the expected level of confidence. For most r.v. however, the bound (4) is far too conservative, and one might ask whether it is reasonable to fear overshooting the error probability in the fashion we have described. The following numerical example shows the existence of r.v. for which the stopping rule (3) fails in the non-asymptotic regime

Example 2.1. The heavy-tailed Pareto-distribution has the probability distribution function

$$f(x) = \begin{cases} \alpha x_m^\alpha x^{-(\alpha+1)} & \text{if } x \geq x_m \\ 0 & \text{else,} \end{cases} \tag{5}$$

where $\alpha, x_m \in \mathbb{R}_+$ are respectively the shape and the scale parameter. The moments of $\mathrm{E}[X^n]$ for the Pareto r.v. only exists for $n < \alpha$ and, supposing $\alpha > 2$, its mean and variance are given by

$$\mu = \frac{\alpha x_m}{\alpha - 1} \text{ and } \sigma^2 = \frac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}.$$

It is further easy to derive that for a Pareto r.v. with $\alpha = 3 + \gamma$ and $0 < \gamma < 1$,

$$\beta = \frac{\mathrm{E}\left[|X-\mu|^3\right]}{\sigma^3} = \mathcal{O}\left(\gamma^{-1}\right).$$

This implies that there exists r.v. for which the second summand of the bound (4) can become arbitrary large. So for such r.v. the stopping rule (3) might fail. Let us investigate by numerical approximations. Considering the distribution with $\alpha = 3.1$ (and $x_m = 1$), yields a heavy-tailed r.v. with known mean, variance and third moment. For a set of accuracies $\mathrm{TOL} \in [0.05, 0.2]$ and confidences $\delta = \mathrm{TOL}^\ell$, $\ell = 0.5, 1, 1.5$, and 2 we have numerically approximated $\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\right) \leq \delta$ using, in accordance with (3), the stopping rule

$$M = \left\lceil \frac{C_{\mathrm{CLT}}^2 \sigma^2}{\mathrm{TOL}^2} \right\rceil$$

The results, illustrated in Figure 1, show that when $\delta \ll \mathrm{TOL}$, the sought confidence is far from met.

Example 2.1 shows that for some r.v. the confidence goal of (1) will not be met by using the stopping rule (3), at least in settings with $\delta \ll \mathrm{TOL}$. Supposing we do not know the variance prior to sampling, yet another type of stopping rule failure is given in Example 2.2; it considers how the MC estimate of Algorithm 1 depends on the initial number of samples M_0.

Example 2.2 (Premature Stopping)**.** In this example we will sample the mean of various r.v. using Algorithm 1 and investigate how the MC estimate $\overline{X}_M$ depends on the initial number of samples M_0. Let $M(M_0)$ denote the number of r.v. used in the stopped estimate as a function of the initial number of samples M_0. Our MC estimate goal then becomes to achieve

$$\mathrm{P}\left(\left|\overline{X}_{M(M_0)} - \mu\right| > \mathrm{TOL}\right) \leq \delta. \tag{6}$$

To investigate whether this goal is fulfilled we plot $\mathrm{P}\left(\left|\overline{X}_{M(M_0)} - \mu\right| > \mathrm{TOL}\right)$ as a function of M_0 for four different r.v. in Figure 2. Figure 2 indicates that the more heavy-tailed or skewed the distribution is, the higher M_0 is needed to ensure that the goal (6) is met.

The demonstrated stopping rule failures motivated us to study and develop ways of constructing more reliable stopping rules. In Section 3, we first analyze the stopping rule of Algorithm 1, and derive an approximate upper bound for the failure probability expressed in terms of M, TOL and δ. In Section 4, we develop

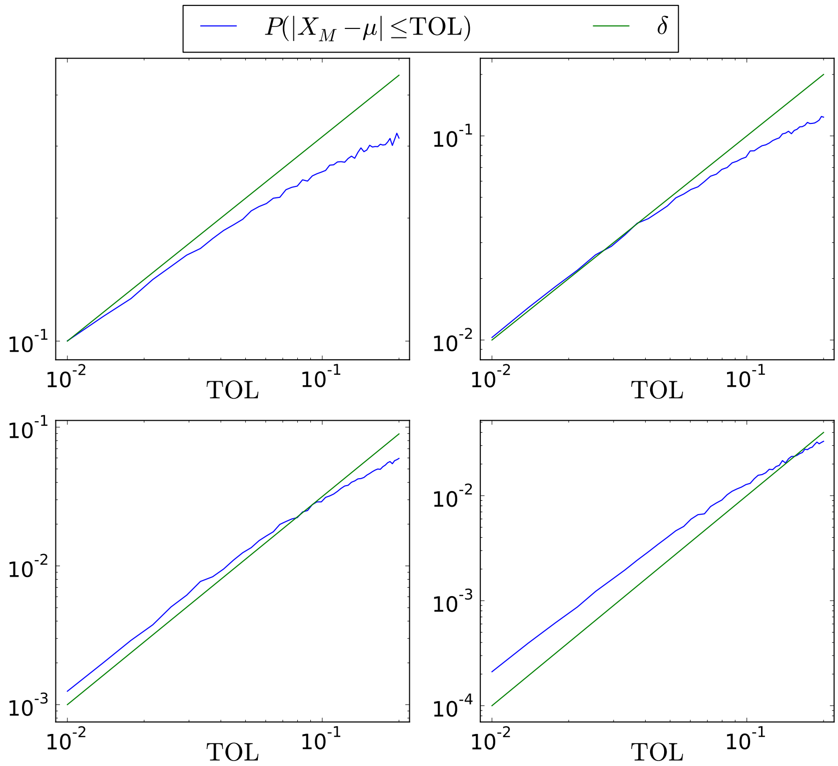

Figure 1. MC estimate using the stopping rule (3) for i.i.d. Pareto r.v. with parameters $\alpha = 3.1$ and $x_m = 1$. The obtained failure probability $\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\right)$ (blue lines) is plotted in comparison to the sought confidence parameter $\delta(\mathrm{TOL}) = \mathrm{TOL}^{\ell}$ (green lines), for $\ell = 0.5$ (upper left), $\ell = 1$ (upper right), $\ell = 1.5$ (lower left), and $\ell = 2$ (lower right). We observe that the smaller δ is relative to TOL, the more apparent does the failure of the stopping rule become.

a more reliable stopping rule algorithm, which in addition to second moment of the r.v. in question, also depends on third and fourth order moments. The paper is concluded with numerical examples comparing the reliability and computational cost of Algorithm 1 with the stopping rule developed in Section 4.

3. Error analysis for Algorithm 1

Examples 2.1 and 2.2 illustrate that for some r.v. the stopping rule in Algorithm 1 does not meet the accuracy-confidence constraint (1). To construct a more reliable stopping rule, penalty terms have to be added to the stopping criterion in Algorithm 1. Some care should be taken to make the penalty terms of right size: if too large penalties are added, the new stopping rule will be reliable but very inefficient, while if too small penalty terms are added, the algorithm will of course be efficient but unreliable.

In this section, we first derive an approximate upper bound for the failure probability

$$\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\,\middle|\, M\right) \tag{7}$$

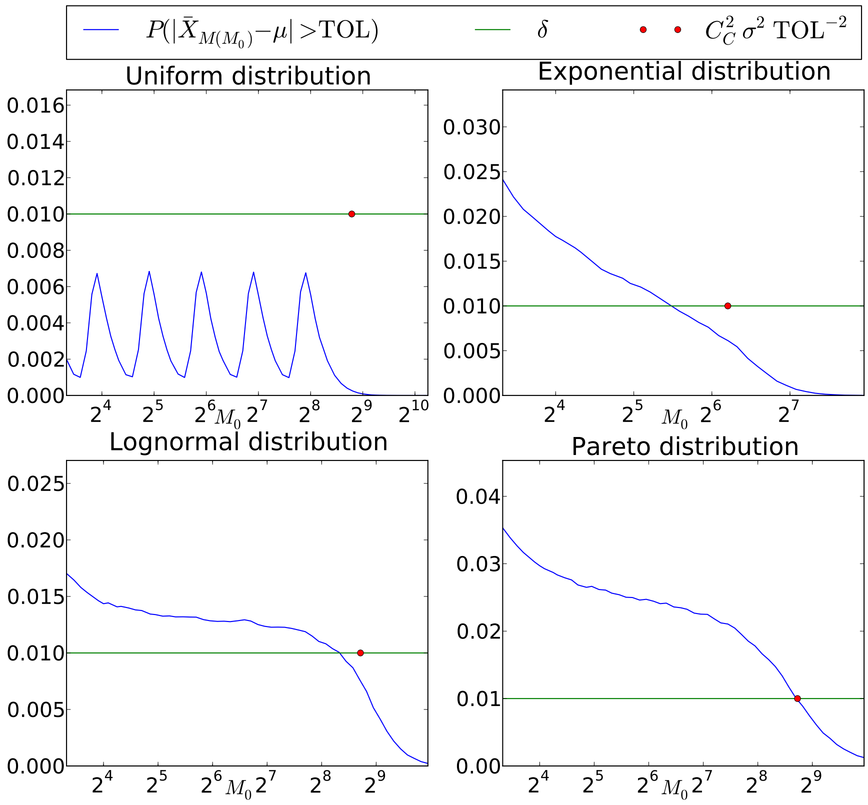

FIGURE 2. Plots of $\mathrm{P}\left(\left|\overline{X}_{M(M_0)} - \mu\right| > \text{TOL}\right)$ as a function of M_0 when using Algorithm 1 to sample $\overline{X}_{M(M_0)}$. The accuracy and confidence is set to $\text{TOL} = 0.1$ and $\delta = \text{TOL}^2$, respectively. **Upper left:** The Uniform distribution $X \sim U(-1, 1)$ with $\mu = 0$ and $\sigma^2 = 2/3$ (light-tailed). **Upper right:** The Exponential distribution with $\mu = 1/3$ and $\sigma^2 = 1/9$ (not heavy-tailed). **Lower left:** The Lognormal distribution $X \sim \log(\mathcal{N}(\mu_{\mathcal{N}}, \sigma^2_{\mathcal{N}}))$ with with $\mu_{\mathcal{N}} = -1$ and $\sigma^2_{\mathcal{N}} = 1$. This gives $\mu = \exp(\mu_{\mathcal{N}} + \sigma^2_{\mathcal{N}}/2)$ and $\sigma^2 = (\exp(\sigma^2_{\mathcal{N}}) - 1)\exp(2\mu_{\mathcal{N}} + \sigma^2_{\mathcal{N}})$ (quite heavy-tailed). **Lower right:** The Pareto distribution with $x_m = 1$ and $\alpha = 3.1$, cf. (5). (heavy-tailed).

corresponding to the stopping rule of Algorithm 1 conditional on the (random) final number of samples M. Clearly, the bound for (7) will also be a r.v. Using the bound for (7), we thereafter construct reasonable penalty terms to be added to the stopping criterion of our new stopping rule.

Let $\overline{\sigma}_M$ denote the sample variance generated from the stopped sample batch, i.e., the samples used to generate the output MC estimate $\overline{X}_M$. Then our first step towards an upper bound for (7) is partitioning the probability (7) into two parts as follows

$$
\begin{aligned}
&\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M\Big) \\
&= \mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M, \big\{|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\big\}\Big)\,\mathrm{P}\Big(|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\Big|M\Big) \\
&\quad + \mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M, \big\{|\overline{\sigma}_M^2 - \sigma^2| < \sigma^2/2\big\}\Big)\,\mathrm{P}\Big(|\overline{\sigma}_M^2 - \sigma^2| < \sigma^2/2\Big|M\Big).
\end{aligned} \tag{8}
$$

The event $|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2$ implies the estimate of the real variance is substantially wrong, and then it is likely that we use far too few samples M to ensure that our MC estimate is reliable. A relatively high penalty term should therefore be added to the stopping criterion to avoid the event $|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2$ from occurring. To derive such a penalty term, we will first bound the factors of the product

$$
\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M, \big\{|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\big\}\Big)\,\mathrm{P}\Big(|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\Big|M\Big) \tag{9}
$$

separately. For the first factor of this product,

$$
\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M, \big\{|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\big\}\Big)\Big),
$$

we recall that in Algorithm 1 the samples used in the output estimate $\overline{X}_M$ and for $\overline{\sigma}_M$ are independent of the samples used to determine M. Keeping this in mind, we derive the following approximate upper bound

$$
\begin{aligned}
&\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M, \big\{|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\big\}\Big) \\
&= \mathrm{P}\Big(\big|\overline{X}_n - \mu\big| > \mathrm{TOL}\,\Big|\big\{|\overline{\sigma}_n^2 - \sigma^2| \geq \sigma^2/2\big\}\Big)\Big|_{n=M} \\
&\lessapprox 2\left(1 - \Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right) + C_{\mathrm{BE}}\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right)\frac{\beta}{\sqrt{M}}\right).
\end{aligned} \tag{10}
$$

Here the Berry-Esseen bound of Corollary 1.2 was used to derive the approximate bound of the last line.

For the second factor of the product (9), we obtain the following equality

$$
\mathrm{P}\Big(|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\Big|M\Big) = \mathrm{P}\big(|\overline{\sigma}_n^2 - \sigma^2| \geq \sigma^2/2\big)\big|_{n=M}.
$$

Furthermore, using Chebycheff's inequality and k-Statistics to bound the variance of the sample variance, cf. Keeping [6], we derive that

$$
\begin{aligned}
\mathrm{P}\Big(|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\Big|M\Big) &= \mathrm{P}\big(|\overline{\sigma}_n^2 - \sigma^2| \geq \sigma^2/2\big)\big|_{n=M} \\
&\leq 4\,\mathrm{E}\left[\frac{|\sigma^2 - \overline{\sigma}_n^2|^2}{\sigma^4}\right]\Bigg|_{n=M} \leq 4\frac{\sigma^4\left(\frac{2}{M-1} + \frac{\kappa}{M}\right)}{\sigma^4} = 4\left(\frac{2}{M-1} + \frac{\kappa}{M}\right).
\end{aligned}
$$

Here κ denotes the *kurtosis*, i.e.

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

We conclude that

$$
\mathrm{P}\Big(|\overline{\sigma}_M^2 - \sigma^2| \geq \sigma^2/2\Big|M\Big) \leq \min\left\{1, 4\left(\frac{2}{M-1} + \frac{\kappa}{M}\right)\right\}. \tag{11}
$$

Next, we want to bound the first factor of the second term of (8),

$$
\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|M, \big\{|\overline{\sigma}_M^2 - \sigma^2| \leq \sigma^2/2\big\}\Big).
$$

The event $|\overline{\sigma}_M^2 - \sigma^2| \leq \sigma^2/2$ indicates that the variance is not substantially wrong-estimated and thereby that it is quite likely that sufficiently many samples are used in our MC estimate. Example 2.1 however illustrated that even in settings with reasonably well-estimated M, failing to meet the confidence is still possible. A relatively weak penalty should thus be added to the stopping criterion to avoid failure in this setting. Applying the Edgeworth expansion with truncated $o\big(n^{-1/2}\big)$ as a weak penalty, cf. Theorem 1.3, we derive the approximate bound

$$\begin{aligned}
&\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|\,M, \big\{|\overline{\sigma}_M^2 - \sigma^2| > \sigma^2/2\big\}\Big)\\
&= \mathrm{P}\Big(\big|\overline{X}_n - \mu\big| > \mathrm{TOL}\,\Big|\,\big\{|\overline{\sigma}_n^2 - \sigma^2| > \sigma^2/2\big\}\Big)\Big|_{n=M}\\
&= \mathrm{P}\left(\sqrt{n}\frac{\big|\overline{X}_n - \mu\big|}{\sigma} > \frac{\sqrt{n}\,\mathrm{TOL}}{\sigma}\,\middle|\,\big\{|\overline{\sigma}_n^2 - \sigma^2| > \sigma^2/2\big\}\right)\Bigg|_{n=M}\\
&\lessapprox 2\left(1 - \Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right) + \frac{\left|\frac{M\,\mathrm{TOL}^2}{\sigma^2} - 1\right| \exp\left(-\frac{M\,\mathrm{TOL}^2}{\sigma^2}\right)}{6\sqrt{2\pi M}}\,\frac{\big|\mathrm{E}\big[(X-\mu)^3\big]\big|}{\sigma^3}\right).
\end{aligned} \tag{12}$$

Combining (10), (11) and (12), and noting that for all $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$\frac{\big|x^2 - 1\big|\, e^{-x^2/2}}{6\sqrt{2\pi n}}\,\frac{\big|\mathrm{E}\big[(X-\mu)^3\big]\big|}{\sigma^3} \leq C_{\mathrm{BE}}(x)\,\frac{\beta}{\sqrt{n}},$$

we obtain the following approximate bound for failing to meet the accuracy of Algorithm 1 conditioned on the stopped number of samples M:

$$\begin{aligned}
&\mathrm{P}\Big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\,\Big|\,M\Big)\\
&\lessapprox 2\left\{1 - \Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right) + C_{\mathrm{BE}}\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right)\frac{\beta}{\sqrt{M}}\right\}\mathrm{P}\big(|\overline{\sigma}_M^2 - \sigma^2| \leq \sigma^2/2\big|\,M\big)\\
&+ 2\left\{1 - \Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right) + \frac{\left|\frac{M\,\mathrm{TOL}^2}{\sigma^2} - 1\right|\,\big|\mathrm{E}\big[(X-\mu)^3\big]\big|}{\exp\left(\frac{M\,\mathrm{TOL}^2}{2\sigma^2}\right) \times 6\sqrt{2\pi M}\sigma^3}\right\}\mathrm{P}\big(|\overline{\sigma}_M^2 - \sigma^2| > \sigma^2/2\big|\,M\big)\\
&\lessapprox 2\left(1 - \Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right)\right) + 2C_{\mathrm{BE}}\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\sigma}\right)\frac{\beta}{\sqrt{M}}\min\left\{1, 4\left(\frac{2}{M-1} + \frac{\kappa}{M}\right)\right\}\\
&+ \frac{\left|\frac{M\,\mathrm{TOL}^2}{\sigma^2} - 1\right|\,\big|\mathrm{E}\big[(X-\mu)^3\big]\big|}{\exp\left(\frac{M\,\mathrm{TOL}^2}{2\sigma^2}\right) \times 3\sqrt{2\pi M}\sigma^3}\max\left\{1 - 4\left(\frac{2}{M-1} + \frac{\kappa}{M}\right), 0\right\}.
\end{aligned} \tag{13}$$

4. A higher moments based stopping rule

From the approximate stochastic error bound (13) we will in this section construct a new, more reliable stopping rule with a stopping criterion based on second, third, and fourth moments of the r.v. that is sampled. The sampled moments our new algorithm will depend on are (here represented in biased form)

$$\overline{\sigma}_M := \sqrt{\sum_{i=1}^{M}\frac{(X_i - \overline{X}_M)^2}{M}}, \qquad \overline{\beta}_M := \sum_{i=1}^{M}\frac{|X_i - \overline{X}_M|^3}{M\overline{\sigma}_M^3}, \tag{14}$$

$$\widehat{\beta}_M := \sum_{i=1}^{M}\frac{(X_i - \overline{X}_M)^3}{M\overline{\sigma}_M^3}, \qquad \text{and} \quad \overline{\kappa}_M := \sum_{i=1}^{M}\frac{(X_i - \overline{X}_M)^4}{M\overline{\sigma}_M^4} - 3.$$

Replacing moments with sample moments in (13), we obtain a computable approximate stochastic error bound

$$\begin{aligned} &\mathrm{P}\Big(\left|\overline{X}_M-\mu\right|>\mathrm{TOL}\,\Big|M\Big) \\ &\lessapprox 2\left(1-\Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\overline{\sigma}_M}\right)\right)+2C_{\mathrm{BE}}\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\overline{\sigma}_M}\right)\frac{\overline{\beta}_M}{\sqrt{M}}\min\left\{1,4\left(\frac{2}{M-1}+\frac{\overline{\kappa}_M}{M}\right)\right\} \\ &\quad+\frac{\left|\frac{M\,\mathrm{TOL}^2}{\overline{\sigma}_M^2}-1\right|\,|\widehat{\beta}_M|}{\exp\left(\frac{M\,\mathrm{TOL}^2}{2\overline{\sigma}_M^2}\right)\times 3\sqrt{2\pi M}\,\overline{\sigma}_M^3}\max\left\{1-4\left(\frac{2}{M-1}+\frac{\overline{\kappa}_M}{M}\right),0\right\}. \end{aligned} \tag{15}$$

The resulting approximate stochastic error bound will be implemented as the following stopping criterion in Algorithm 2:

$$\begin{aligned} &2\left(1-\Phi\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\overline{\sigma}_M}\right)\right)+2C_{\mathrm{BE}}\left(\frac{\sqrt{M}\,\mathrm{TOL}}{\overline{\sigma}_M}\right)\frac{\overline{\beta}_M}{\sqrt{M}}\min\left\{1,4\left(\frac{2}{M-1}+\frac{\overline{\kappa}_M}{M}\right)\right\} \\ &\quad+\frac{\left|\frac{M\,\mathrm{TOL}^2}{\overline{\sigma}_M^2}-1\right|\,|\widehat{\beta}_M|}{\exp\left(\frac{M\,\mathrm{TOL}^2}{2\overline{\sigma}_M^2}\right)\times 3\sqrt{2\pi M}\,\overline{\sigma}_M^3}\max\left\{1-4\left(\frac{2}{M-1}+\frac{\overline{\kappa}_M}{M}\right),0\right\}<\delta \end{aligned} \tag{16}$$

We now present the new stopping rule algorithm.

Algorithm 2 Higher Moments Based Stopping Rule

Input: Accuracy TOL, confidence δ, and initial number of samples M_0.
Output: $\overline{X}_M$.

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 $\overline{\kappa}_{M_n}$ according to (14).
while Inequality (16) is not fulfilled. **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 $\overline{\kappa}_{M_n}$.
end while
Set $M=M_n$, generate i.i.d. samples $\{X_i\}_{i=1}^{M}$ and return the sample mean $\overline{X}_M$.

5. Numerical experiments

In the numerical experiments we will estimate the mean of four differently distributed r.v. by using both the the sample variance based stopping rule in Algorithm 1 and the new higher moments based stopping rule in Algorithm 2. We compare the reliability and complexity of the algorithms, with the complexity measured in terms of average number of r.v. realizations needed to generate the given MC estimate for a given accuracy-confidence pair TOL and δ. The distributions considered here are the light-tailed Uniform distribution, the Exponential distribution, the heavier-tailed Lognormal distribution, and the heavy-tailed Pareto distribution. In all these experiments we have set the algorithm parameter initial number of samples to $M_0=30$. From Figures 3, 4, 5, and 6, which illustrate the results of the numerical experiments, we observe that for the heavy-tailed distributions Algorithm 2 performs reliably and succeeds in meeting the accuracy-confidence constraint while

Algorithm 1 does not. For the light tailed distributions considered, both Algorithms meet the accuracy-confidence constraint. Regarding the complexity of the algorithms, we see that Algorithm 2 is only slightly more costly than Algorithm 1, and, as expected, the complexities of the algorithms seem to become more similar as TOL decreases.

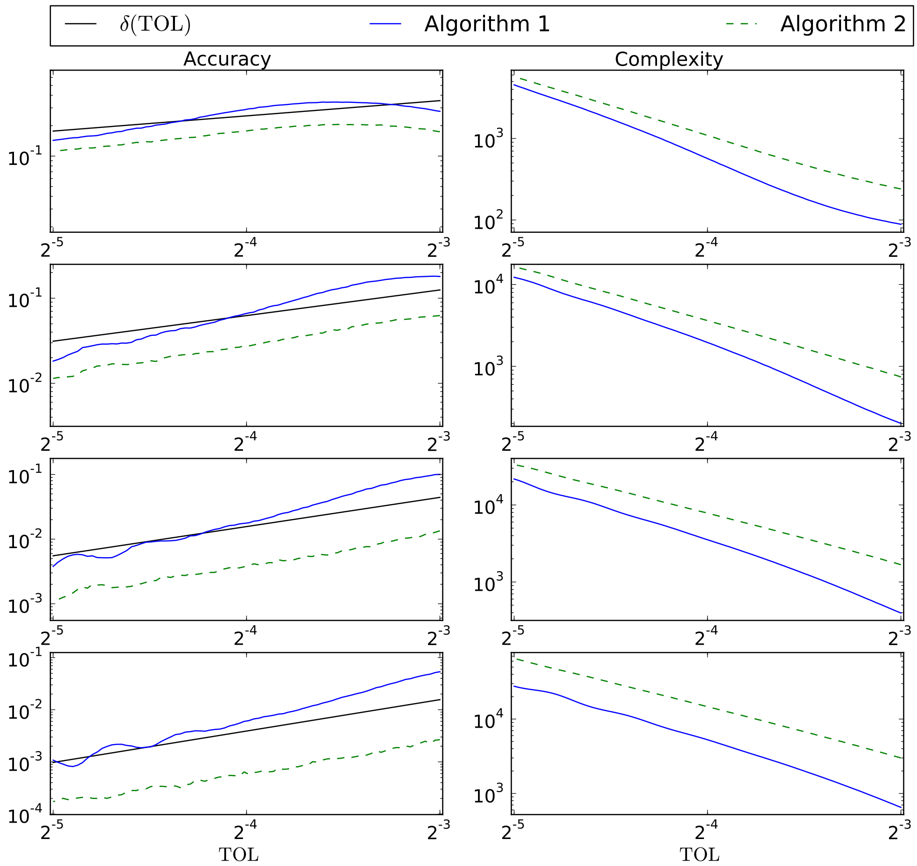

FIGURE 3. (**Pareto Distribution**) Numerical comparison of the accuracy and complexity of reaching the goal $\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) < \delta$ with Algorithm 1 and 2 when sampling Pareto distributed r.v.s with parameters $\alpha = 3.1$ and $x_m = 1$, cf. (5). Row plots from top to bottom is the output for the respective confidences $\delta(\mathrm{TOL}) = \mathrm{TOL}^{1/2}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}^{3/2}$, and $\delta(\mathrm{TOL}) = \mathrm{TOL}^2$.

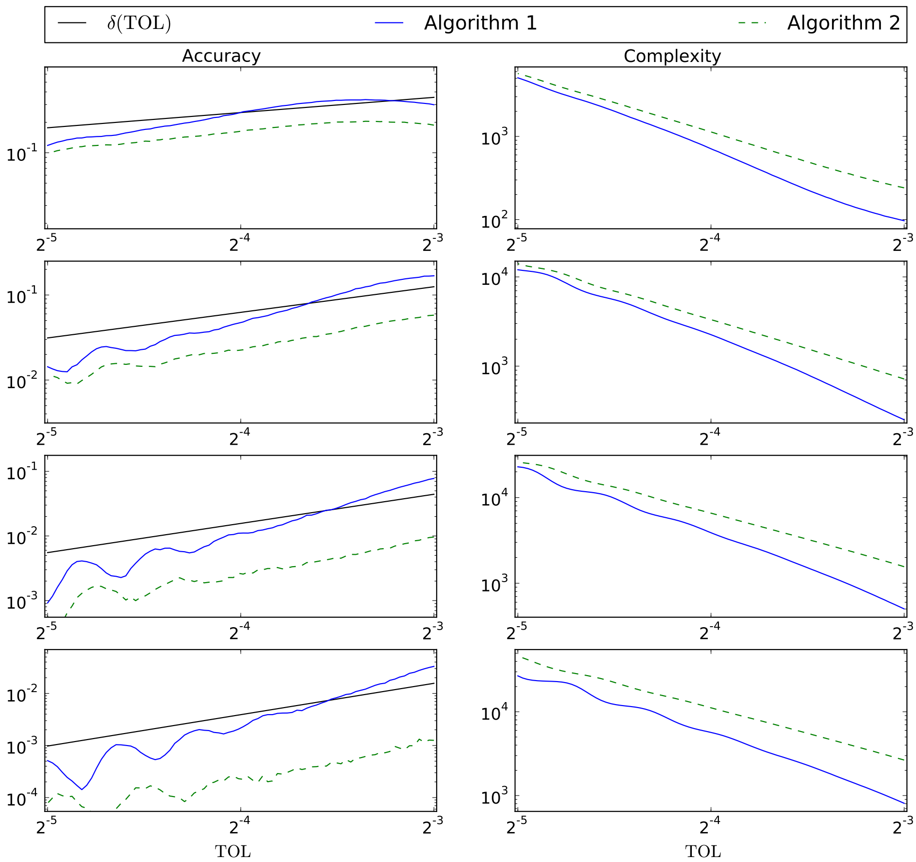

FIGURE 4. (**Lognormal Distribution**) Numerical comparison of the accuracy and complexity of reaching the goal $\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) < \delta$ with Algorithm 1 and 2 when sampling Lognormal distributed r.v. $X \sim \log(\mathcal{N}(\mu_{\mathcal{N}}, \sigma^2_{\mathcal{N}}))$ with with $\mu_{\mathcal{N}} = -1$ and $\sigma^2_{\mathcal{N}} = 1$. Row plots from top to bottom is the output for the respective confidences $\delta(\mathrm{TOL}) = \mathrm{TOL}^{1/2}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}^{3/2}$, and $\delta(\mathrm{TOL}) = \mathrm{TOL}^2$.

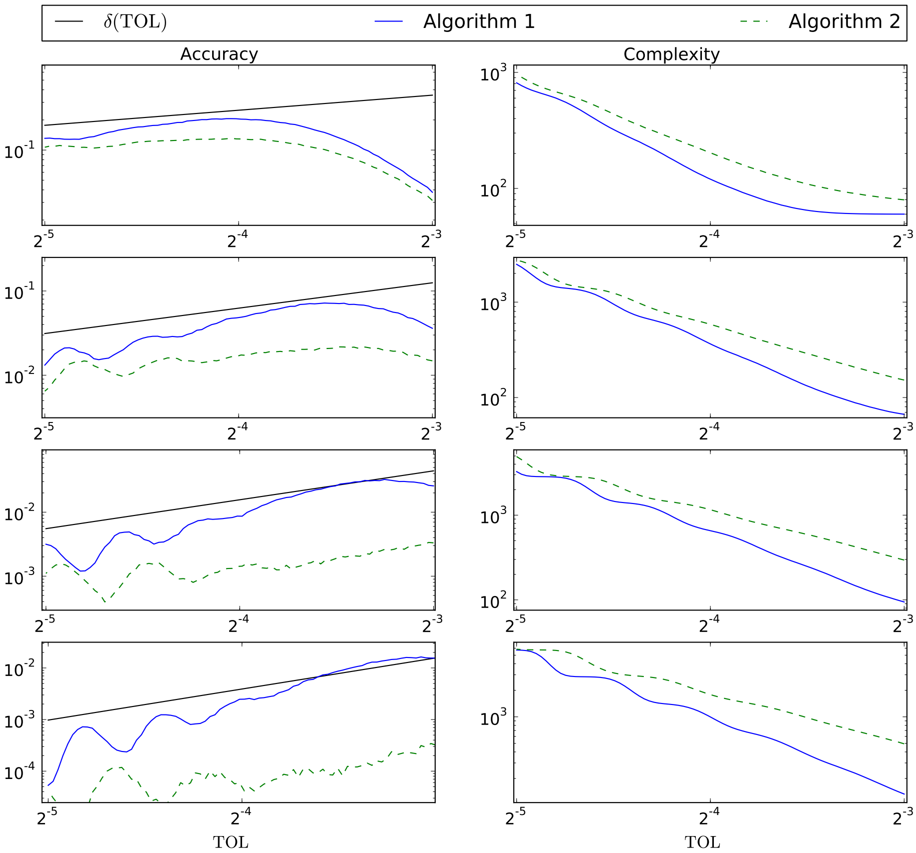

FIGURE 5. (**Exponential Distribution**) Numerical comparison of the accuracy and complexity of reaching the goal $\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) < \delta$ with Algorithm 1 and 2 when sampling exponentially distributed r.v. with $\mu = 1/3$. Row plots from top to bottom is the output for the respective confidences $\delta(\mathrm{TOL}) = \mathrm{TOL}^{1/2}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}^{3/2}$, and $\delta(\mathrm{TOL}) = \mathrm{TOL}^2$.

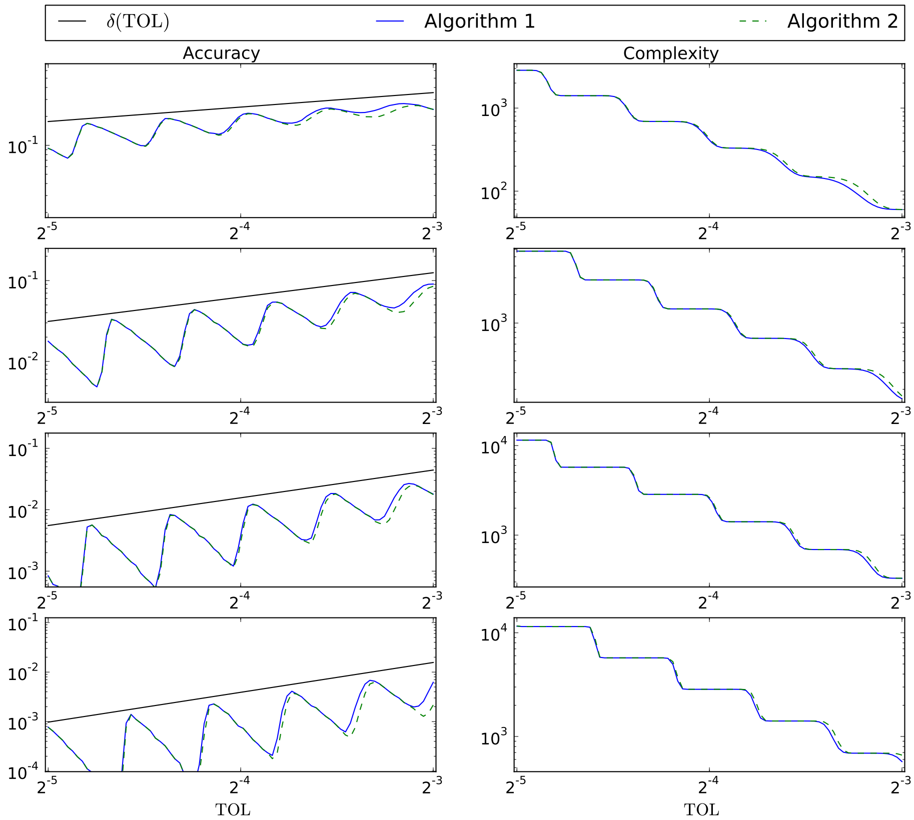

FIGURE 6. (**Uniform Distribution**) Numerical comparison of the accuracy and complexity of reaching the goal $\mathrm{P}\big(\big|\overline{X}_M - \mu\big| > \mathrm{TOL}\big) < \delta$ with Algorithm 1 and 2 when sampling uniformly distributed r.v. $X \sim U(-1,1)$. Row plots from top to bottom is the output for the respective confidences $\delta(\mathrm{TOL}) = \mathrm{TOL}^{1/2}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}$, $\delta(\mathrm{TOL}) = \mathrm{TOL}^{3/2}$, and $\delta(\mathrm{TOL}) = \mathrm{TOL}^2$.

6. Conclusion

We have shown that second moment based sequential stopping rules such as Algorithm 1 run the risk of using too few samples in MC estimates, especially when sampling heavy-tailed r.v. in settings with very stringent confidence requirements, i.e., $\delta \ll \mathrm{TOL}$. Algorithm 2, a higher moment based stopping rule algorithm is proposed in this work, and, according to the numerical examples of Section 5, our new stopping rule performs more much reliable than Algorithm 1 while only slightly increasing the computational cost. In short, we believe that our new stopping rule presented in Algorithm 2 is well worth considering in settings with heavy tailed r.v. and/or $\delta \ll \mathrm{TOL}$.

Note that our analysis of the original Algorithm 1 critically depends on three main ingredients:

(I) a general, non-asymptotic estimate of the tail probabilities for the sample mean $\overline{X}_M$, for which we used either the non-uniform Berry-Esseen theorem given in Corollary 1.2 or the Edgeworth expansion given in Theorem 1.3,

(II) a choice between the more conservative Berry-Esseen bound and the approximate Edgeworth bound made depending on whether the sample variance of the samples used to generate the output MC estimate is close to, or far from the true variance,

(III) an estimate of the conditional distribution function of the sample variance given the output M of the stopping algorithm given in (11).

There is clearly room for improvement in all these steps. First of all, the second ingredient above is dangerous as we do not know how to estimate the correlation between $\overline{X}_M$ and the events $|\overline{\sigma}_M^2 - \sigma^2| > \sigma^2/2$ and $|\overline{\sigma}_M^2 - \sigma^2| \leq \sigma^2/2$. This is problematic, as these approximations can potentially have the wrong sign, i.e., it is possible that the right-hand sides of (10) and (12) are smaller than their respective left-hand sides even though we actually seek upper bounds. It is however our hope that these approximation errors are compensated by the the overly pessimistic non-uniform Berry-Esseen estimate and by using Chebycheff's inequality to bound the conditional distribution function of the sample variance. Even though the numerical evidence obtained in Section 5 seems to confirm that the compensations work well, we would prefer an analysis in which each estimation step can be controlled, at least in the sense that we indeed obtain an upper bound for the error probability.

To a lesser extent, it is not clear that the truncation of the $o(n^{-1/2})$ of the Edgeworth expansion will lead to an upper bound for the error probability, either. In this case, the approximation error is however of higher order, so a stronger case can be made on why the effect will finally be negligible. In fact, when we used truncated Edgeworth expansion also for the estimation of (10) – instead of the non-uniform Berry-Esseen theorem – then the corresponding stopping rule turned out to be not much more reliable than Algorithm 1, indicating that there is a delicate balance between reliability in meeting the accuracy target (1) and maintaining an acceptable efficiency.

Acknowledgments. This work was partially supported by the Center for Predictive Computational Science), the VR project "Effektiva numeriska metoder för stokastiska differentialekvationer med tillämpningar", and King Abdullah University of Science and Technology (KAUST).

References

[1] Paul Bratley, Bennett L. Fox, and Linus E. Schrage. *A guide to simulation (2nd ed.)*. Springer-Verlag New York, Inc., New York, NY, USA, 1987.

[2] Y. S. Chow and Herbert Robbins. On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *The Annals of Mathematical Statistics*, 36(2):pp. 457–462, 1965.

[3] W. Feller. *An introduction to probability theory and its applications. Vol II. 2nd ed.* Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley and Sons, Inc. XXIV, 669 p. , 1971.

[4] Peter W. Glynn and Ward Whitt. The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2(1):pp. 180–198, 1992.

[5] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):pp. 13–30, 1963.

[6] E. S. Keeping. *Introduction to statistical inference.* D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1962.

[7] Averill M. Law and W. David Kelton. Confidence intervals for steady-state simulations, ii: A survey of sequential procedures. *Management Science*, 28(5):pp. 550–562, 1982.

[8] Averill M. Law and W. David Kelton. Confidence intervals for steady-state simulations: I. a survey of fixed sample size procedures. *Operations Research*, 32(6):pp. 1221–1239, 1984.

[9] R. Michel. On the constant in the nonuniform version of the berry-essen theorem. *Probability Theory and Related Fields*, 55:109–117, 1981. 10.1007/BF01013464.

[10] Valentin V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[11] Il'ya S Tyurin. Refinement of the upper bounds of the constants in lyapunov's theorem. *Russian Mathematical Surveys*, 65(3):586, 2010.

Department of Mathematics, University of Vienna, Nordbergstrae 15, 1090 Wien, Austria
E-mail address: christian.bayer@univie.ac.at

Department of Numerical Analysis, Kungl. Tekniska Högskolan, 100 44 Stockholm, Sweden
E-mail address: hhoel@kth.se

Division of Mathematics, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia
E-mail address: erik.vonschwerin@kaust.edu.sa

Division of Mathematics, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia
E-mail address: raul.tempone@kaust.edu.sa