Approximate fixed width confidence intervals in Monte Carlo sampling

Fred J. Hickernell   Lan Jiang   Yuewei Liu   Art Owen

July 9, 2012

Abstract

We present a two stage fixed width confidence interval for the mean, motivated by problems in Monte Carlo sampling. The first stage generates a conservative variance estimate. The second stage uses that variance estimate to make a confidence interval for the unknown mean. It has been known since Bahadur and Savage (1956) that exact non-parametric confidence intervals for the mean do not exist, without some assumptions. Our procedure gives at least the desired coverage level, under a fourth moment (kurtosis) condition on the underlying random variable.

1 Introduction

Monte Carlo sampling is often the method of choice for difficult high dimensional integration problems. One of the strengths of the Monte Carlo method is that the problem data themselves provide a good estimate of the method's accuracy via the central limit theorem (CLT).

When the desired integral is a real value μ, the CLT can be used to construct an approximate 99% confidence interval for μ (described in more detail below). While we get good control of the confidence level (e.g., 99 versus 99.9 percent), what we may really want is control over the length of that interval, either its absolute length, or its length as a proportion of μ. That is, we may want a fixed length confidence interval. Historically such intervals are known as fixed width confidence intervals. There are deterministic analogues. For example, Matlab has a function `quad` designed to approximate $\int_0^1 f(x)\,\mathrm{d}x$ to within precision $\epsilon > 0$ given the function f.

In this paper we present a two stage fixed width confidence interval for the mean of a real random variable, suitable for Monte Carlo sampling. Before presenting the method, we outline the reasons that existing fixed width intervals are not suitable.

The length (equivalently width) of a confidence interval tends to become smaller as the number n of sampled function values increases. In special circumstance, we can choose n to get a confidence interval of at most the desired length and at least the desired coverage level. For instance, if a variance parameter is known then an approach based on Chebychev's inequality is available, though the actual coverage will usually be much higher than the nominal level, meaning that much narrower intervals would have sufficed. Known variance in addition to a Gaussian distribution for the function values supports a fixed width interval construction that is not too conservative. Finally, conservative fixed width intervals for means of bounded random variables, by appealing to exponential inequalities Hoeffding's or Chernoff's inequality.

If the relevant variance or bound is unknown, then approaches based on sequential statistics [**?**] may be available. In sequential methods one keeps increasing n until the interval is narrow enough. Sequential confidence intervals require us to take account of the stopping rule when computing the confidence level. They are available in special circumstances, such as Gaussian or binary data. Similarly, Bayesian methods can support a fixed width interval containing μ with 99% posterior probability, and Bayesian methods famously do not require one to account for stopping rules. They do however require strong distributional assumptions.

The solutions described above require strong assumptions that generally do not hold in Monte Carlo applications. There is no assumption-free way to obtain exact confidence intervals for a mean, as has been known since [**?**]. Some kind of assumption is needed to rule out worst case settings where the desired quantity is the mean of a heavy tailed random variable in which rarely seen large values dominate the mean. The assumption we use is an upper bound on the kurtosis (normalized fourth moment) of the random variable. Under such an assumption we present a two stage algorithm: the first stage generates a conservative variance estimate, and the second stage uses the CLT with the variance from the first stage.

An outline of this paper is as follows:

2 Background probability and statistics

In our Monte Carlo applications, a quantity of interest is written as an expectation: $\mu = \mathbb{E}(Y)$, where Y is a real valued random variable. Very often $Y = f(\boldsymbol{X})$ where $\boldsymbol{X} \in \mathbb{R}^d$ is a random vector with probability density function ρ. Then $\mu = \int_{\mathbb{R}^d} f(\boldsymbol{x})\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$. In other settings the random quantity $\boldsymbol{X}$ might have a discrete distribution or be infinite dimensional (e.g,. a Gaussian process) or both. For Monte Carlo estimation, we can work with the distribution of Y alone. Methods such as quasi-Monte Carlo sampling which exploit smoothness of f and/or ρ require more detailed specification of those functions.

The Monte Carlo estimate of μ is

$$\hat{\mu} = \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{1}$$

where Y_i are independent random variables with the same distribution as Y.

2.1 Moments

Our methods require conditions on higher moments of Y as described here. The variance of Y is

$$\sigma^2 = \mathbb{E}((Y-\mu)^2).$$

The standard deviation σ of Y is the non-negative square root of σ^2. Some of our expressions assume without stating it that $\sigma > 0$, and all will require $\sigma < \infty$. The skewness of Y is $\gamma = \mathbb{E}((Y-\mu)^3)/\sigma^3$, and the kurtosis of Y is $\kappa = \mathbb{E}((Y-\mu)^4)/\sigma^4 - 3$. The mysterious 3 in κ is there to make it zero for Gaussian random variables. Also, $\mu, \sigma^2, \gamma, \kappa$ are the first four cumulants [**?**] of the distribution of Y. We will make use of

$$\tilde{\kappa} = \kappa + 3$$

in some of our derivations.

In addition to the moments above we will also use some centered absolute moments of the form $M_k = M_k(Y) = \mathbb{E}(|Y-\mu|^k)$. Normalized versions of these are $\widetilde{M}_k = M_k(Y)/\sigma^k$. In particular, $\widetilde{M}_4 = \tilde{\kappa}$ and $\widetilde{M}_3$ governs the convergence rate of the CLT.

It is a standard result that $1 \le q \le p < \infty$ implies $M_q(Y) \le M_p(Y)^{q/p}$ and similarly $\widetilde{M}_q(Y) \le \widetilde{M}_p(Y)^{q/p}$. The special case

$$\widetilde{M}_3(Y) \le \widetilde{M}_4(Y)^{3/4} \tag{2}$$

will be important for us.

2.2 CLT intervals

A random variable Z has the standard normal distribution, denoted by $\mathcal{N}(0,1)$, if

$$\Pr(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)\,\mathrm{d}t.$$

We use $\Phi(z)$ for this cumulative distribution function.

Under the central limit theorem, the distribution of $\sqrt{n}(\hat{\mu}_n - \mu)/\sigma$ approaches $\mathcal{N}(0,1)$ as $n \to \infty$. As a result

$$\Pr\bigl(\hat{\mu}_n - 2.58\sigma\sqrt{n} \le \mu \le \hat{\mu}_n + 2.58\sigma/\sqrt{n}\bigr) \to 0.99 \tag{3}$$

as $n \to \infty$. We write the interval in (3) as $\hat{\mu}_n \pm 2.58\sigma/\sqrt{n}$. Equation (3) is not usable when σ^2 is unknown, but the usual estimate

$$\hat{v}_n = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{\mu}_n)^2 \tag{4}$$

may be substituted, yielding the interval $\hat{\mu}_n \pm 2.58s/\sqrt{n}$ which also satisfies the limit in (3) by Slutsky's theorem. For an arbitrary confidence level $1-\alpha \in (0,1)$, we replace the constant 2.58 by $Z^{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$. The width of this interval is $2Z^{1-\alpha/2}s/\sqrt{n}$, and when μ is in the interval then the absolute error $|\mu - \hat{\mu}_n| \le \varepsilon \equiv Z^{1-\alpha/2}s/\sqrt{n}$.

The coverage level of the CLT interval is only asymptotic. In more detail, [**?**] shows that

$$\Pr\bigl(\mu \in \hat{\mu}_n \pm 2.58s\sqrt{n}\bigr) = 0.99 + \frac{1}{n}(A + B\gamma^2 + C\kappa) + O\Bigl(\frac{1}{n^2}\Bigr) \tag{5}$$

for constants A, B, and C that depend on the desired coverage level (here 99%). Hall's theorem requires only that the random variable Y has sufficiently many finite moments and is not supported solely on a lattice (such as the integers). It is interesting to note that the $O(1/n)$ coverage error in (5) is better than the $O(1/\sqrt{n})$ root mean squared error for the estimate $\hat{\mu}_n$ itself.

2.3 Standard inequalities

Here we present some well known inequalities that we will make use of. First, Chebychev's inequality ensures that a random variable (such as $\hat{\mu}_n$) is seldom too far from its mean.

Theorem 1 (Chebyshev's Inequality)**.** *Let Z be a random variable with mean μ and variance $\sigma^2 > 0$. Then for all $\alpha > 0$,*

$$\Pr\left[|Z-\mu| \ge \frac{\sigma}{\sqrt{\alpha}}\right] \le \alpha, \qquad \Pr\left[|Z-\mu| < \frac{\sigma}{\sqrt{\alpha}}\right] \ge 1-\alpha.$$

Proof. This is a slight restatement of the version in [**?**][page 52]. ∎

In some settings we need a one sided inequality like Chebychev's. We will use this one due to Cantelli.

Theorem 2 (Cantelli's inequality)**.** *Let Z be any random variable with mean μ and finite variance σ^2. For any $a \ge 0$, it follows that:*

$$\Pr[Z - \mu \ge a] \le \frac{\sigma^2}{a^2 + \sigma^2}.$$

Proof. [**?**][page 53] ∎

Berry-Esseen type theorems govern the rate at which a CLT takes hold. We will use the following one.

Theorem 3 (Non-uniform Berry-Esseen Inequality)**.** *Let $Y_1, \dots, Y_n$ be i.i.d. random variables. Suppose that $\mathbb{E}(Y_i) = \mu$, $\mathrm{Var}(Y_i) = \sigma^2 > 0$, and $\widetilde{M}_3 = \mathbb{E}(|Y_i - \mu|^3)/\sigma^3 < \infty$. Then for any $x \in \mathbb{R}$,*

$$\left|\Pr\left[\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (Y_i - \mu) < x\right] - \Phi(x)\right| \le \frac{A\widetilde{M}_3}{\sqrt{n}}(1+|x|)^{-3},$$

where A is some number satisfying $0.4097 \le A \le 0.5600$.

Proof. [**?**, Theorem 5.16, p. 168] $\square$

Our method requires us to bound the probability of the sample variance being too large. For that, we will use some moments of the variance estimate.

Theorem 4. *Let $Y_1, \dots, Y_n$ be i.i.d. random variables with finite kurtosis κ and modified kurtosis $\tilde{\kappa} = \kappa + 3$. Let $\hat{\mu}_n$ be the sample mean as defined in* (1) *and $\hat{v}_n$ be the sample variance as defined in* (4)*. Then the variance of the sample variance and the kurtosis of the sample mean are given by*

$$\mathrm{Var}(\hat{v}_n) = \frac{\sigma^4}{n}\left(\tilde{\kappa} - \frac{n-3}{n-1}\right), \qquad \kappa(\hat{\mu}_n) = \frac{\kappa}{n}.$$

Proof. The first claim is proved in [**?**, Chapter 7]. The second claim is a standard result on cumulants. $\square$

3 Two stage confidence interval

Our two stage procedure works as follows. In the first stage, we take a sample of independent values $Y_1, \dots, Y_{n_\sigma}$ from the distribution of Y. From this sample we compute $\hat{v}_{n_\sigma}$ according to (4) and estimate the variance of Y_i by $\hat{\sigma}^2 = \mathfrak{C}^2 \hat{v}_{n_\sigma}$, where $\mathfrak{C}^2 > 1$ is a "variance inflation factor" that will reduce the probability that we have underestimated $\mathrm{Var}(Y)$. For the second stage, we use the estimate $\hat{\sigma}^2$ as if it were the true variance of Y_i and use Berry-Esseen theorem to obtain a suitable sample size.

The next two subsections give details of these two steps that will let us bound their error probabilities. Then we give a theorem on the method as a whole.

3.1 Conservative variance estimates

We need to ensure that our first stage estimate of the variance σ^2 is not too small. The following result bounds the probability of such an underestimate.

Proposition 1. *Let $Y_1, \dots, Y_n$ be IID random variables with variance $\sigma^2 > 0$ and kurtosis κ. Let $\hat{v}_n$ be the sample variance defined at* (4)*, and let $\tilde{\kappa} = \kappa + 3$. Then*

$$\Pr\left[\hat{v}_n < \sigma^2\left\{1 + \sqrt{\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)\left(\frac{1-\alpha}{\alpha n}\right)}\right\}\right] \geq 1 - \alpha, \tag{6a}$$

$$\Pr\left[\hat{v}_n > \sigma^2\left\{1 - \sqrt{\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)\left(\frac{1-\alpha}{\alpha n}\right)}\right\}\right] \geq 1 - \alpha. \tag{6b}$$

Proof. Choosing

$$a = \sigma^2\sqrt{\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)\left(\frac{1-\alpha}{\alpha n}\right)} > 0,$$

(a) (b)

Figure 1: (a) The maximum kurtosis, $\kappa_{\max}(\alpha, n_\sigma, 1.5)$, as defined in (8); (b) comparison of sample sizes $N_G(0.01, \alpha)$, $N_C(0.01, \alpha)$, and $N_B(0.01, \alpha, \kappa_{\max}^{3/4}(\alpha, 1000, 1.5))$.

it follows from Cantelli's inequality (Theorem 2) that

$$\Pr\left[\hat{v}_n - \sigma^2 \geq \sigma^2 \sqrt{\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)\left(\frac{1-\alpha}{\alpha n}\right)}\right] = \Pr\left[\hat{v}_n - \sigma^2 \geq a\right]$$

$$\leq \frac{\operatorname{Var}(\hat{v}_n)}{a^2 + \operatorname{Var}(\hat{v}_n)} = \frac{\frac{\sigma^4}{n}\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)}{\frac{\sigma^4}{n}\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)\left(\frac{1-\alpha}{\alpha}\right) + \frac{\sigma^4}{n}\left(\tilde{\kappa} - \frac{n-3}{n-1}\right)} = \frac{1}{\left(\frac{1-\alpha}{\alpha}\right) + 1} = \alpha.$$

Then (6a) follows directly. By a similar argument, applying Cantelli's inequality to the expression $\Pr\left[-\hat{v}_n + \sigma^2 \geq a\right]$ implies (6b). $\square$ $\square$

Using Proposition 1 we can bound the probability that $\hat{\sigma}^2 = \mathfrak{C}^2 \hat{v}_{n_\sigma} < \sigma^2$. Equation (6a) implies that

$$\Pr\left[\frac{\hat{v}_{n_\sigma}}{1 - \sqrt{\left(\tilde{\kappa} - \frac{n_\sigma - 3}{n_\sigma - 1}\right)\left(\frac{1-\alpha}{\alpha n_\sigma}\right)}} > \sigma^2\right] \geq 1 - \alpha, \tag{7}$$

where $\tilde{\kappa}$ is $\widetilde{M}_4(Y)$. It follows that we require the kurtosis of the integrand to be small enough, relative to n_σ, α, and $\mathfrak{C}$, in order to ensure that (7) holds. Specifically, we require

$$\frac{1}{1 - \sqrt{\left(\tilde{\kappa} - \frac{n_\sigma - 3}{n_\sigma - 1}\right)\left(\frac{1-\alpha}{\alpha n_\sigma}\right)}} \leq \mathfrak{C}^2,$$

or equivalently,

$$\tilde{\kappa} \leq \frac{n_\sigma - 3}{n_\sigma - 1} + \left(\frac{\alpha n_\sigma}{1-\alpha}\right)\left(1 - \frac{1}{\mathfrak{C}^2}\right)^2 =: \tilde{\kappa}_{\max}(\alpha, n_\sigma, \mathfrak{C}). \tag{8}$$

Figure 1a shows how large a kurtosis can be accommodated for a given n_σ, α, and $\mathfrak{C} = 1.5$. Note that for $n = 30$, a common rule of thumb for applying the central limit theorem, even the modest value $\alpha = 0.1$ yields $\tilde{\kappa}_{\max}$ of only about 2, corresponding to a kurtosis of about 5. While that kurtosis is reasonably large for many observational data settings, Monte Carlo applications often involve a larger kurtosis than this.

3.2 Conservative interval widths

Here we consider how to choose the sample size n to get the desired coverage level from an interval with half-length at most ε. We suppose here that σ is known. In practice we will use a conservative (biased high) estimate for σ.

First, if the CLT held exactly and not just asymptotically, then we could use a CLT sample size, of

$$N_{\mathrm{CLT}}(\varepsilon,\sigma,\alpha) = \left\lceil \Bigl(\frac{Z^{1-\alpha/2}\sigma}{\varepsilon}\Bigr)^2 \right\rceil$$

independent values of Y_i in an interval like the one in (3).

Given knowledge of σ, but no assurance of a Gaussian distribution for $\hat\mu_n$, we could instead select a sample size based on Chebychev's inequality, taking

$$N_{\mathrm{Cheb}}(\varepsilon,\sigma,\alpha) = \left\lceil \frac{\sigma^2}{\alpha\varepsilon^2} \right\rceil \tag{9}$$

observations Y_i and using an interval like (3) gives

$$\Pr(|\hat\mu - \mu| \le \varepsilon) \ge 1-\alpha. \tag{10}$$

Naturally $N_{\mathrm{Cheb}} \ge N_{\mathrm{CLT}}$.

Finally, we could use the non-uniform Berry-Esseen inequality from Theorem 3. This inequality requires a finite third moment $M_3(Y) = \mathbb{E}(|Y-\mathbb{E}(Y)|^3)$. Let $\widetilde{M}_3 = M_3(Y)/\sigma^3$. The non-uniform Berry-Esseen inequality implies that

$$\begin{aligned}
\Pr\left[|\hat\mu_n - \mu| < \frac{\sigma}{\sqrt{n}}x\right] &= \Pr\left[\hat\mu_n - \mu < \frac{\sigma}{\sqrt{n}}x\right] - \Pr\left[\hat\mu_n - \mu \le -\frac{\sigma}{\sqrt{n}}x\right] \\
&\ge \left[\Phi(x) - \frac{0.56\widetilde{M}_3}{\sqrt{n}}(1+|x|)^{-3}\right] - \left[\Phi(-x) + \frac{0.56\widetilde{M}_3}{\sqrt{n}}(1+|x|)^{-3}\right] \\
&= 1 - 2\left(\frac{0.56\widetilde{M}_3}{\sqrt{n}}(1+|x|)^{-3} + \Phi(-x)\right), \qquad (11)
\end{aligned}$$

for all $x \in \mathbb{R}$. Choosing $x = \varepsilon\sqrt{n}/\sigma$, we find that the probability of making an error less than ε is bounded below by $1-\alpha$, i.e.,

$$\Pr[|\hat\mu_n - \mu| < \varepsilon] \ge 1-\alpha, \quad \text{provided } n \ge N_{\mathrm{BE}}(\varepsilon,\sigma,\alpha,\widetilde{M}_3), \tag{12a}$$

where the Berry-Esseen sample size is

$$N_{\mathrm{BE}}(\varepsilon,\sigma,\alpha,M) := \min\left\{ n \in \mathbb{N} : \Phi\left(-\sqrt{n}\varepsilon/\sigma\right) + \frac{0.56M}{\sqrt{n}\left(1+\sqrt{n}\epsilon/\sigma\right)^3} \le \frac{\alpha}{2} \right\}. \tag{12b}$$

To compute this value, we need to know $\widetilde{M}_3$. In practice, substituting an upper bound on $\widetilde{M}_3$ yields an upper bound on the necessary sample size.

It is possible that in some situations $N_{\mathrm{BE}} > N_{\mathrm{Cheb}}$ might hold. In such cases we could use N_{Cheb} instead.

3.3 Algorithm and theorem

In detail, the two stage algorithm works as follows. First, the user specifies four quantities:

- an initial sample size $n_\sigma \geq 2$ for variance estimation,
- a variance inflation factor $\mathfrak{C}^2 \in (1, \infty)$,
- an uncertainty tolerance $\alpha \in (0, 1)$, and,
- an error tolerance, $\varepsilon > 0$.

Though n_σ can be as small as 2 it should ordinarily be larger.

At the first stage of the algorithm, $Y_1, \dots, Y_{n_\sigma}$ are sampled independently from the same distribution as Y. Then the variance estimate $\hat{\sigma}^2 = \mathfrak{C}^2 \hat{v}_{n_\sigma}$ is computed using (4) to compute $\hat{v}_{n_\sigma}$.

To prepare for the second stage of the algorithm we compute $\tilde{\alpha} = 1 - \sqrt{1 - \alpha}$ and then $\tilde{\kappa}_{\max} = \tilde{\kappa}_{\max}(\tilde{\alpha}, n_\sigma, \mathfrak{C})$ using equation (8). The sample size for the second stage is

$$n = N_\mu(\varepsilon, \hat{\sigma}, \tilde{\alpha}, \tilde{\kappa}_{\max}^{3/4}),$$

where

$$N_\mu(\varepsilon, \sigma, \alpha, M) := \min\bigl(N_{\mathrm{Cheb}}(\varepsilon, \sigma, \alpha), N_{\mathrm{BE}}(\varepsilon, \sigma, \alpha, M)\bigr). \tag{13}$$

Recall that N_{Cheb} is defined in (9) and N_{BE} is defined in (12b).

After this preparation, the second stage is to sample $Y_{n_\sigma+1}, \dots, Y_{n_\sigma+N_\mu}$ independently from the distribution of Y and compute

$$\hat{\mu} = \frac{1}{N_\mu} \sum_{i=n_\sigma+1}^{n_\sigma+N_\mu} Y_i. \tag{14}$$

Theorem 5. *Let Y be a random variable with mean μ. If $\widetilde{M}_4(Y) \leq \tilde{\kappa}_{\max}(\tilde{\alpha}, n_\sigma, \mathfrak{C})$ then the two stage algorithm above yields an estimate $\hat{\mu}$ given by* (14) *which satisfies*

$$\Pr(|\hat{\mu} - \mu| \leq \varepsilon) \geq 1 - \alpha.$$

Proof. The first stage yields a variance estimate satisfying $\Pr(\hat{\sigma}^2 > \sigma^2) \geq 1 - \tilde{\alpha}$ by (8) applied with error tolerance $\tilde{\alpha}$. The second stage yields $\Pr(|\hat{\mu}_n - \mu| \leq \varepsilon) \geq 1 - \tilde{\alpha}$ by the Berry-Esseen result (11), so long as $\hat{\sigma} \geq \sigma$ and $\widetilde{M}_3(Y) \leq \tilde{\kappa}_{\max}(\tilde{\alpha}, n_\sigma, \mathfrak{C})$. The second condition holds under the theorem's hypothesis because $\widetilde{M}_3 \leq \tilde{\kappa}_{\max}^{3/4}$; see (2). Thus, in the two stage algorithm we have

$$\begin{aligned} \Pr\left(|\hat{\mu}_n - \mu| \leq \varepsilon\right) &= \mathbb{E}\bigl(\Pr\left(|\hat{\mu}_n - \mu| \leq \varepsilon \mid \hat{\sigma}\right)\bigr) \\ &\geq \mathbb{E}\left((1 - \tilde{\alpha}) 1_{\sigma \leq \hat{\sigma}}\right) \\ &\geq (1 - \tilde{\alpha})(1 - \tilde{\alpha}) = 1 - \alpha. \end{aligned}$$

□

Table 1: Probability of meeting the error tolerance for test function (15) using the adaptive algorithm in Theorem 5.

p	0.0001	0.0002	0.0005	0.001	0.002	0.005
$\Pr(\lvert\mu - \hat{\mu}_n\rvert \le 0.01)$	8.90%	21.30%	39.80%	63.20%	85.80%	99.50%

Figure 2: Empirical distribution function of $\left|\mu - \hat{\mu}_n\right|/0.01$ for example (15) and various values of p using the adaptive algorithm in Theorem 5.

4 Illustrative examples

Here we illustrate the difficulties of fixed precision computation. Our first example uses the two stage algorithm from Section 3. Our second example illustrates difficulties with the deterministic `quad` function from Matlab.

4.1 Two stage algorithm

To illustrate our two stage algorithm, consider the following univariate step-function integrated against the uniform probability distribution on $[0, 1]$, i.e., $\rho = 1$:

$$f(x) = \begin{cases} \mu + \sigma\sqrt{\frac{1-p}{p}}, & 0 \le x \le p, \\ \mu - \sigma\sqrt{\frac{p}{1-p}}, & p < x \le 1, \end{cases} \tag{15}$$

where $p \in (0, 1)$ is a parameter. The test function parameters are $\mu = \sigma = 1$. The algorithm parameters are an absolute error tolerance of 0.01, an uncertainty of $\alpha = 5\%$, a sample size for estimating the variance of $n_\sigma = 1000$, and a variance inflation factor $\mathfrak{C} = 1.5$. Table 1 shows the percentage of times that the error tolerance is met for various values of p. For $p = 0.005$ the algorithm exceeds the required 95% success, but for $p = 0.0001$ the algorithm only gets the correct answer less than 10% of the time. Figure 2 shows the empirical distribution of the normalized error. A normalized error no greater than one means that the error tolerance has been met.

In this case, the initial sample used to estimate the variance of the integrand may miss a very narrow spike. Without a good estimate of this variance, there is no reliable determination of the sample size needed for computing a sample mean that is close enough to the mean of the function. Theorem 5 describes under what conditions this adaptive algorithm will not be fooled. In the present example, small values of p lead to large values of $\tilde{\kappa}$ and for small enough p the kurtosis condition in Theorem 5 is violated.

Figure 3: Integrand that fools MATLAB's `quad` function.

4.2 MATLAB's `quad`

The difficulties of error estimation are not unique to Monte Carlo methods. For the step function f defined in (15), MATLAB's `quad` function approximates $\int_0^1 f(x - 1/\sqrt{2} \pmod 1)\, \mathrm{d}x$ with an error tolerance of 10^{-14} to be ≈ 0.92911, instead of the true answer of 1. Thus, the step function can fool automatic quadrature routines. Figure 3 displays the integrand

$$f(x) = 1+\cos\left(8\pi \min\left(\max\left(\frac{x-0.27158}{0.45684}, 0\right), 1\right)\right), \qquad \int_0^1 f(x)\, \mathrm{d}x = 1.54316.$$

Here the constants 0.27158 and 0.45684 are chosen to fool MATLAB's `quad` function. Applying `quad` to approximate the above integral with an error tolerance of 10^{-14} gives the answer 2, since all of the integrand values sampled by `quad` are 2. Thus, `quad` is fooled again, even though this integrand has continuous first derivative.

5 Algorithm cost

6 Relative error