ADAPTIVE SIMPLE MONTE CARLO

FRED J. HICKERNELL

ABSTRACT. I attempt a probabilistic analysis of simple Monte Carlo, achieving probabilistic error bounds when the kurtosis is controlled. The algorithm uses a sample size that depends adaptively on the estimated variance of the integrand. Thus, the algorithm is nonlinear (depending essentially on the function). My intention, if what is done here is correct, is to try to extend this to the more sophisticated sampling schemes and infinite dimensional problems.

1. NON-ADAPTIVE MONTE CARLO

Suppose one wishes to compute the following integral or mean, μ, of some function $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mu = \int_{\mathbb{R}^d} g(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x},$$

where $f : \mathbb{R}^d \to [0, \infty)$ is a probability density function. The $\mathcal{L}_{p,f}$ norm of g is defined by

$$\|g\|_{p,f} := \left\{ \int_{\mathbb{R}^d} |g(\boldsymbol{x})|^p f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right\}^{1/p}$$

Note that if $1 \le q < p$, then by Hölder's inequality,

$$
\begin{aligned}
\|g\|_{q,f} &= \left\{ \int_{\mathbb{R}^d} |g(\boldsymbol{x})|^q f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right\}^{1/q} \\
&\le \left\{ \int_{\mathbb{R}^d} |g(\boldsymbol{x})|^p f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right\}^{1/p} \left\{ \int_{\mathbb{R}^d} 1^{p/(p-q)} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right\}^{(p-q)/(pq)} \\
&= \|g\|_{p,f} \, \|1\|_{pq/(p-q),f} = \|g\|_{p,f} \, .
\end{aligned}
\tag{1}
$$

Thus, $\mathcal{L}_q \subseteq \mathcal{L}_p$ for $1 \le q < p$.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be independent and identically distributed random variables with the probability density function f, briefly, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ i.i.d. $\sim f$. Then the simple Monte Carlo estimator of the mean, is

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{X}_i). \tag{2}$$

An important question that springs to mind is how to choose the right sample size, n, to achieve a specified tolerance, ϵ. The random estimator, $\hat{\mu}_n$ has mean μ and variance

$$\mathrm{var}(\hat{\mu}_n) = \frac{\sigma^2}{n},$$

where σ^2 is the variance of the function g:

$$\sigma^2 := \mathrm{var}(g) := \int_{\mathbb{R}^d} [g(\boldsymbol{x}) - \mu]^2 f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{3}$$

1

In practice, one often invokes the Central Limit Theorem to determine sample size. Given a significance level or uncertainty tolerance, α, one has

$$\text{Prob}\left[|\hat{\mu}_n - \mu| \leq \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right] \approx 1 - \alpha.$$

Thus one has the *approximate* probabilistic result

(4) $\text{Prob}\left[|\hat{\mu}_n - \mu| \leq \epsilon\right] \approx 1 - \alpha$ for $n = N_N(\epsilon, \alpha, \sigma^2) := \left\lceil \left(\frac{z_{\alpha/2}\sigma}{\epsilon}\right)^2 \right\rceil.$

The above is exact if $g(\boldsymbol{X}_i)$ are i.i.d. normal, however, in general this result is only approximate and relies on the unknown σ. The above observations are formalized in the proposition below.

Proposition 1. *For a given positive constant, $\sigma_{\max}$, define*

$$\mathcal{G}_{N,\sigma_{\max}} = \{g \in \mathcal{L}_{2,f} : g(\boldsymbol{X}) \sim N(\mu, \sigma^2) \text{ for } \boldsymbol{X} \sim f, \ \sigma^2 \leq \sigma_{\max}^2, \},$$

where the variance of the function is defined in (??). If for a given error tolerance, ϵ, and an uncertainty tolerance, α, then a probabilistic error bound is given by

$$\text{Prob}\left[|\hat{\mu}_n - \mu| \leq \epsilon\right] \geq 1 - \alpha \qquad \text{for } n \geq N_N(\epsilon, \alpha, \sigma_{\max}).$$

Note that the set $\mathcal{G}_{N,\sigma_{\max}}$ assumes that $g(\boldsymbol{X})$ is normal, and that there is a priori knowledge about the variance of g. The first of these drawbacks may be removed by turning to Berry-Esseen inequality.

Using Berry-Esseen's inequality (Theorem **??**) yields an exact upper bound rather than approximate one. Choosing $Z = \hat{\mu}_n$ yields

$$\text{Prob}\left[|\hat{\mu}_n - \mu| < \frac{\sigma}{\sqrt{n\alpha}}\right] \geq 1 - \alpha.$$

Thus, a proper choice of sample size guarantees that the estimate is within the tolerance of the true answer with probability $1 - \alpha$:

$$\text{Prob}\left[|\hat{\mu}_n - \mu| < \epsilon\right] \geq 1 - \alpha \qquad \text{for } n \geq N_B(\epsilon, \alpha, \sigma, \varrho)$$

where $N_B(\epsilon, \alpha, \sigma, \varrho)$ is the solution of

(5) $$\Phi\left(-\frac{\sqrt{n}\epsilon}{\sigma}\right) + \frac{A\varrho}{\sqrt{n}} = \frac{\alpha}{2}.$$

By (**??**), we know $n = \frac{\sigma^2}{\epsilon^2}z_{\alpha/2 - A\varrho/\sqrt{n}}^2$, that means, $N_B(\epsilon, \alpha, \sigma, \varrho) \to N_N(\epsilon, \alpha, \sigma)$, when $n \to +\infty$.

Proposition 2. *For a given positive constant, $\sigma_{\max}$, define the set of functions*

$$\mathcal{G}_{C,\sigma_{\max}} = \{g \in \mathcal{L}_{2,f} : \text{var}(g) = \sigma^2 \leq \sigma_{\max}^2, \},$$

where the variance of the function is defined in (??). If for a given error tolerance, ϵ, an uncertainty tolerance, α, and ϱ defined in theorem ??, define $N_B(\epsilon, \alpha, \sigma, \varrho)$ is the solution of (??), then a probabilistic error bound is given by

$$\text{Prob}\left[|\hat{\mu}_n - \mu| \leq \epsilon\right] \geq 1 - \alpha \qquad \text{for } n \geq N_B(\epsilon, \alpha, \sigma_{\max}, \varrho).$$

Proof. Note $Z_n = \hat{\mu}_n - \mu$, by theorem **??**, for $x > 0$ we have

$$\begin{aligned}
\text{Prob}\left[Z_n < \frac{\sigma}{\sqrt{n}}x\right] &= 1 - \text{Prob}\left[Z_n \geq \frac{\sigma}{\sqrt{n}}x\right] \\
&= 1 - \left(\text{Prob}\left[Z_n \geq \frac{\sigma}{\sqrt{n}}x\right] - \Phi(-x) + \Phi(-x)\right) \\
&\geq 1 - A\frac{\varrho}{\sqrt{n}} - \Phi(-x)
\end{aligned}$$

then

$$\begin{aligned}
\text{Prob}\left[|Z_n| < \frac{\sigma}{\sqrt{n}}x\right] &= \text{Prob}\left[Z_n < \frac{\sigma}{\sqrt{n}}x\right] - \text{Prob}\left[Z_n < -\frac{\sigma}{\sqrt{n}}x\right] \\
&\geq 1 - A\frac{\varrho}{\sqrt{n}} - \Phi(-x) - \text{Prob}\left[Z_n < -\frac{\sigma}{\sqrt{n}}x\right] \\
&\geq 1 - A\frac{\varrho}{\sqrt{n}} - \Phi(-x) - \left(A\frac{\varrho}{\sqrt{n}} + \Phi(-x)\right) \\
&= 1 - 2\left(A\frac{\varrho}{\sqrt{n}} + \Phi(-x)\right)
\end{aligned}$$

Let $\epsilon = \sigma x/\sqrt{n}$, we control the probility of

$$\text{Prob}[|Z_n| < \epsilon] \geq 1 - 2\left(A\frac{\varrho}{\sqrt{n}} + \Phi\left(-\frac{\epsilon\sqrt{n}}{\sigma}\right)\right) \geq 1 - \alpha$$

this means we can choose $n := N_B(\epsilon, \alpha, \sigma, \varrho)$, such that

$$\Phi(-\frac{\sqrt{n}\epsilon}{\sigma}) + \frac{A\varrho}{\sqrt{n}} = \frac{\alpha}{2} \tag{6}$$

since $N_B(\epsilon, \alpha, \sigma_{\max}, \varrho) \geq N_B(\epsilon, \alpha, \sigma, \varrho)$, so

$$\text{Prob}\left[|\hat{\mu}_n - \mu| \leq \epsilon\right] \geq 1 - \alpha \qquad \text{for } n \geq N_B(\epsilon, \alpha, \sigma_{\max}, \varrho).$$

$\square$

2. ADAPTIVE MONTE CARLO

Although the Berry-Esseen result is exact for general distributions, it still depends on the typically unknown σ^2. This is why in practice one normally uses observed function values observed to approximate the σ^2 by the sample variance, as follows:

$$\hat{v}_n = \frac{1}{n-1}\sum_{i=1}^{n}[g(\boldsymbol{X}_i) - \hat{\mu}_n]^2. \tag{7}$$

This means that we now have an *adaptive* algorithm. One might choose an initial sample of size n_0, and use it to estimate σ^2 by $\hat{v}_{n_0}$. Then one chooses an *independent* sample of size $n = N_C(\epsilon, \alpha, \hat{v}_{n_0})$ or $n = N_N(\epsilon, \alpha, \hat{v}_{n_0})$ to compute $\hat{\mu}_n$ the final estimate of μ.

Unfortunately, once we approximate σ^2 by $\hat{v}_n$, we again have inexact results. However, they can be made exact by using Cantelli's inequality (Theorem **??**) and the variance of $\hat{v}_n$

in Theorem **??**.

$$1 - \alpha \leq \text{Prob}\left[\hat{v}_n - \sigma^2 \geq -\sigma^2\sqrt{\frac{1}{n}\left(\kappa + \frac{2n}{n-1}\right)\left(\frac{1-\alpha}{\alpha}\right)}\right]$$

$$= \text{Prob}\left[\frac{\hat{v}_n}{1 - \sqrt{\left(\kappa + \frac{2n}{n-1}\right)\left(\frac{1-\alpha}{\alpha n}\right)}} \geq \sigma^2\right]$$

Thus,

(8) $\quad \text{Prob}\left[\hat{\sigma}_{\text{up}}(n, \alpha, \kappa) \geq \sigma\right] \geq 1 - \alpha, \qquad \text{where } \hat{\sigma}_{\text{up}}^2(n, \alpha, \kappa) = \dfrac{\hat{v}_n}{1 - \sqrt{\left(\kappa + \frac{2n}{n-1}\right)\left(\frac{1-\alpha}{n\alpha}\right)}},$

provided that

$$1 > \left(\kappa + \frac{2n}{n-1}\right)\left(\frac{1-\alpha}{n\alpha}\right)$$

$$\frac{n\alpha}{1-\alpha} > \kappa + \frac{2n}{n-1}$$

$$\kappa < \frac{n\alpha}{1-\alpha} - \frac{2n}{n-1} =: \kappa_{\text{poss}}.$$

Theorem 3. *For a given positive constant, $\kappa_{\max}$, define the set of functions with finite fourth momemts:*

$$\mathcal{G}_{C, \sigma_{\max}} = \{g \in \mathcal{L}_{4,f} : \text{kurt}(g) = \kappa \leq \kappa_{\max}, \},$$

where the kurtosis of the function is defined in (**??**). *Suppose that one has an error tolerance, ϵ, and an uncertainty tolerance, α. Let $\alpha_1 = 1 - \sqrt{1-\alpha}$. Pick any $n_0 > 1$ satisfying*

$$\kappa_{\max} < \frac{n_0 \alpha_1}{1 - \alpha_1} - \frac{2n_0}{n_0 - 1},$$

and compute the sample variance, $\hat{v}_{n_0}$ of a simple random sample of size n_0. Use this to compute $\hat{\sigma}_{\text{up}}^2(n_0, \alpha_1, \kappa_{\max})$ by (**??**). *ϱ is defined in theorem* **??**. *Define $N_B(\epsilon, \alpha, \sigma, \varrho)$ is the solution of* (**??**). *Next choose an independent random sample of size n and compute $\hat{\mu}_n$, the simple Monte Carlo estimator of μ. A probabilistic error bound is given by*

$$\text{Prob}\left[|\hat{\mu}_n - \mu| \leq \epsilon\right] \geq 1 - \alpha \qquad \text{for } n := \max\{n_0, N_B(\epsilon, \alpha_1, \sigma_{\max}, \varrho)\}$$

Proof. By (**??**) it follows that $\hat{\sigma}_{\text{up}}(n, \alpha_1, \kappa) \geq \sigma$ with probability $1 - \alpha_1$. $\qquad \square$

Consider a spacial case in (**??**), let $\Phi(-\frac{\sqrt{n}\epsilon}{\sigma}) = \frac{A\varrho}{\sqrt{n}} = \frac{\alpha}{4}$, since $\varrho \leq (\kappa_{\max} + 3)^{3/4}$, we can get the upper bound of $N_B(\epsilon, \alpha_1, \sigma_{\max}, \varrho)$ by

$$N_B(\epsilon, \alpha_1, \sigma_{\max}, \varrho) \leq \max\{\frac{Z_{\alpha_1/4}^2 \sigma_{\max}^2}{\epsilon^2}, \frac{16 A^2 (\kappa_{\max} + 3)^{3/2}}{\epsilon^2}\}$$

3. Example

Consider the case of the uniform probability distribution on $[0, 1]$, i.e., $f = 1$. Define

$$g(x) = \begin{cases} 1 + \sigma\sqrt{\frac{1-p}{p}}, & 0 \leq x \leq p, \\ 1 - \sigma\sqrt{\frac{p}{1-p}}, & p < x \leq 1, \end{cases}$$

where p and σ are parameters, with $0 < p < 1$. Note that

$$\mu = \int_0^1 g(x)\,dx = 1$$

$$\text{var}(g) = \int_0^1 [g(x) - \mu]^2\,dx = \sigma^2 \frac{1-p}{p}p + \sigma^2 \frac{p}{1-p}(1-p) = \sigma^2,$$

$$\kappa = \text{kurt}(g) = \frac{1}{\sigma^4}\int_0^1 [g(x) - \mu]^4\,dx - 3 = \left(\frac{1-p}{p}\right)^2 p + \left(\frac{p}{1-p}\right)^2 (1-p) - 3$$

$$= \frac{(1-p)^3 + p^3}{p(1-p)} - 3 = \frac{1}{p(1-p)} - 6.$$

Note that κ ranges from a minimum of -2, when $p = 1/2$ to a maximum of ∞ when $p = 0, 1$. When

4. QUESTIONS

Here are some questions that suggest themeselves:

- Is this analysis above known already? Is this the typical probabilistic setting? Is it better to look at a randomized setting where one considers the expected value of the error?
- Can this type of analysis be extended to randomized quasi-Monte Carlo for finited dimension, d? Infinite dimension? In this latter case one needs some multilevel algorithm, but the specification of the levels perhaps could be deduced from the data.
- Are there inequalities like Chebyshev's inequality that apply when Z is the sum of i.i.d. random variables? Some of the better known ones, like Hoeffding's inequality assume boundedness, which we cannot presume here.
- How do we prove Cantelli's inequality?

APPENDIX

Theorem 4 (Berry-Esseen Inequality). *Let $X_1, \ldots, X_n$ be independent random variables haveing a common distribution.*
Suppose

$$EX_1 = 0, \quad EX_1^2 = \sigma^2 > 0, \quad E\,|X_1|^3 < \infty, \quad \varrho = \frac{E\,|X_1|^3}{\sigma^3}.$$

Then

$$\sup_x \left| \text{Prob}\left[\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n X_j < x\right] - \Phi(x) \right| \le A\frac{\varrho}{\sqrt{n}}.$$

where A satisfied $0.4097 \le A \le 0.5600$.

Theorem 5. *Let $\hat{v}_n$ be the sample variance as defined in* (??). *It's variance is*

$$\text{var}(\hat{v}_n^2) = \frac{\sigma^4}{n}\left(\kappa + \frac{2n}{n-1}\right),$$

where κ denotes the kurtosis:

(9) $$\kappa := \mathrm{kurt}(g) := \frac{\gamma}{\sigma^4} - 3, \qquad \gamma := \int_{\mathbb{R}^d} [g(\boldsymbol{x}) - \mu]^4 f(\boldsymbol{x}) \, d\boldsymbol{x}.$$

Proof. The sample variance has mean σ^2/n. To facilitate the derviation, let $Y_i = g(X_i) - \mu$.

$$\hat{v}_n = \frac{1}{n-1} \sum_{i=1}^{n} \left[Y_i - \left(\frac{1}{n} \sum_{j=1}^{n} Y_j \right) \right]^2 = \frac{1}{n(n-1)} \left[n \sum_{i=1}^{n} Y_i^2 - \sum_{j,k=1}^{n} Y_j Y_k \right]$$

$$\hat{v}_n^2 = \frac{1}{n^2(n-1)^2} \left[n^2 \sum_{i,j=1}^{n} Y_i^2 Y_j^2 - 2n \sum_{i,j,k=1}^{n} Y_i^2 Y_j Y_k + \sum_{i,j,k,l=1}^{n} Y_i Y_j Y_k Y_l \right]$$

$$E[Y_i^2 Y_j^2] = \begin{cases} \gamma, & i = j, \\ \sigma^4, & i \neq j, \end{cases}$$

$$\sum_{i,j=1}^{n} E[Y_i^2 Y_j^2] = n\gamma + n(n-1)\sigma^4,$$

$$E[Y_i^2 Y_j Y_k] = \begin{cases} \gamma, & i = j = k, \\ \sigma^4, & i \neq j, j = k, \\ 0, & j \neq k, \end{cases}$$

$$\sum_{i,j,k=1}^{n} E[Y_i^2 Y_j Y_k] = n\gamma + n(n-1)\sigma^4$$

$$E[Y_i Y_j Y_k Y_l] = \begin{cases} \gamma, & i = j = k = l, \\ \sigma^4, & i, j, k, l \text{ have 2 distinct values}, \\ 0, & \text{otherwise}, \end{cases}$$

$$\sum_{i,j,k,l=1}^{n} E[Y_i Y_j Y_k Y_l] = n\gamma + 3n(n-1)\sigma^4$$

$$E[\hat{v}_n^2] = \frac{n^3[\gamma + (n-1)\sigma^4] - 2n^2[\gamma + (n-1)\sigma^4] + n[\gamma + 3(n-1)\sigma^4]}{n^2(n-1)^2}$$

$$= \frac{(n-1)\gamma + (n^2 - 2n + 3)\sigma^4}{n(n-1)}$$

$$\mathrm{var}(\hat{v}_n^2) = E[\hat{v}_n^2] - [E(\hat{v}_n)]^2 = \frac{(n-1)\gamma + (n^2 - 2n + 3)\sigma^4}{n(n-1)} - \sigma^4$$

$$= \frac{(n-1)\gamma + (-n+3)\sigma^4}{n(n-1)} = \frac{1}{n} \left(\gamma - \frac{n-3}{n-1}\sigma^4 \right)$$

$$= \frac{1}{n} \left(\sigma^4(\kappa+3) - \frac{n-3}{n-1}\sigma^4 \right) = \frac{\sigma^4}{n} \left(\kappa + \frac{2n}{n-1} \right)$$

and then divide both sides by σ_Z^2/α.　　　　　　　　　　　　　　　　□

Theorem 6 (Single tailed Cantelli's inequality). *ξ is a random variable with bounded variance σ^2 and for any $a \geq 0$, it follows that:*

$$\text{Prob}[\xi - \text{E}\xi \geq a] \leq \frac{\sigma^2}{a^2 + \sigma^2}$$

Proof. Let random variable $X = \xi - \text{E}\xi$, S is like that:

$$S = \left\{ \begin{array}{ll} 1, & X \geq a \\ -1, & X < a \end{array} \right.$$

we know

$$\begin{aligned}
\text{var}(X) &= \text{E}\{X^2\} - (\text{E}\{X\})^2 \\
&= \text{E}\{\text{E}\{X^2|S\}\} - (\text{E}\{\text{E}\{X|S\}\})^2 \\
&= \text{E}\{\text{var}(X|S) + (\text{E}\{X|S\})^2\} - (\text{E}\{\text{E}\{X|S\}\})^2 \\
&= \text{E}\{\text{var}(X|S)\} + \text{E}\{(\text{E}\{X|S\})^2\} - (\text{E}\{\text{E}\{X|S\}\})^2 \\
&= \text{E}\{\text{var}(X|S)\} + \text{var}(\text{E}\{X|S\})
\end{aligned}$$

(10)

Let's look at $\text{E}\{X|S\}$, obviously, $\text{E}\{X|S = 1\} \geq a$, and $\text{E}\{\text{E}\{X|S\}\} = \text{E}X = 0$, e.g.

(11) $$\text{E}\{\text{E}\{X|S\}\} = \text{E}\{X|S = 1\}P_g + \text{E}\{X|S = -1\}P_l = 0$$

where $P_g = \text{Prob}[X \geq a]$, $P_l = \text{Prob}[X < a]$. From (??) we have

(12) $$(\text{E}\{X|S = -1\})^2 = \left(\frac{\text{E}\{X|S = 1\}P_g}{P_l}\right)^2 \geq \left(\frac{aP_g}{P_l}\right)^2$$

then by (??),

$$\begin{aligned}
\text{var}(X) &= \text{E}\{\text{var}(X|S)\} + \text{var}(\text{E}\{X|S\}) \\
&\geq \text{var}(\text{E}\{X|S\}) \\
&= \text{E}\{\text{E}\{X|S\}^2\} - (\text{E}\{\text{E}\{X|S\}\})^2 \\
&= \text{E}\{\text{E}\{X|S\}^2\} \\
&\geq a^2 P_g + P_l(\frac{aP_g}{P_l})^2 = a^2 P_g/(1 - P_g)
\end{aligned}$$

Note $\text{var}(X) = \text{var}(\xi) = \sigma^2$, it says

$$\begin{aligned}
\sigma^2 \geq a^2 P_g/(1 - P_g) &\leftharpoondown \sigma^2(1 - P_g) \geq a^2 P_g \\
&\rightleftharpoons \sigma^2 \geq (a^2 + \sigma^2)P_g \\
&\rightleftharpoons P_g \leq \frac{\sigma^2}{a^2 + \sigma^2}
\end{aligned}$$

also is $P[\xi - \text{E}\xi \geq a] \leq \frac{\sigma^2}{a^2 + \sigma^2}$

□

Theorem 7. *Proof inequality(??)*

Proof. From theorem **??** we know for $a > 0$

$$\text{Prob}[\hat{v}_n - \sigma^2 \leq -a] = \text{Prob}[\sigma^2 - \hat{v}_n \geq a] \leq \frac{\text{var}(\hat{v}_n)}{a^2 + \text{var}(\hat{v}_n)}$$

Let

$$a = \sigma^2 \sqrt{\frac{1}{n}\left(\kappa + \frac{2n}{n-1}\right)\left(\frac{1-\alpha}{\alpha}\right)}$$

by theorem **??**

$$\text{Prob}[\hat{v}_n - \sigma^2 \leq -a] \leq \frac{\frac{\sigma^4}{n}\left(\kappa + \frac{2n}{n-1}\right)}{\frac{\sigma^4}{n}\left(\kappa + \frac{2n}{n-1}\right)\left(\frac{1-\alpha}{\alpha}\right) + \frac{\sigma^4}{n}\left(\kappa + \frac{2n}{n-1}\right)}$$

$$(13) \qquad\qquad = \frac{1}{\left(\frac{1-\alpha}{\alpha}\right)+1} = \alpha$$

by (**??**) we get equation (**??**) easily.   □