Monte Carlo Euler for SDEs

Consider the stochastic differential equation

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t)$$

on $t_0 \le t \le T$, how can one compute the value $E[g(X(T))]$? The Monte-Carlo method is based on the approximation

$$E[g(X(T))] \simeq \sum_{j=1}^{M} \frac{g(\overline{X}(T;\omega_j))}{M},$$

where $\overline{X}$ is an approximation of X, here the Euler method.

The error in the Monte-Carlo method is

$$E[g(X(T))] - \sum_{j=1}^{M} \frac{g(\overline{X}(T;\omega_j))}{M}$$

$$= E[g(X(T)) - g(\overline{X}(T))] \tag{28}$$

$$+ \sum_{j=1}^{M} \frac{E[g(\overline{X}(T))] - g(\overline{X}(T;\omega_j))}{M}. \tag{29}$$

In the right hand side of the error representation (29), the first part is the time discretization error, which we will consider later, and the second part is the statistical error, which we study here.

Monte Carlo Statistical Error

Goal: Approximate the expected value, $E[Y]$, by a sample average of M iid samples

$$\frac{\sum_{j=1}^{M} Y(\omega_j)}{M}$$

and choose M sufficiently large to control the statistical error,

$$E[Y] - \frac{\sum_{j=1}^{M} Y(\omega_j)}{M}.$$

For M independent samples of Y denote sample average $\mathcal{A}(Y;M)$, and sample standard deviation $\mathcal{S}(Y;M)$ of Y by

$$\mathcal{A}(Y;M) \equiv \frac{1}{M}\sum_{j=1}^{M} Y(\omega_j)$$

$$\mathcal{S}(Y;M) \equiv \Big[\mathcal{A}(Y^2;M) - (\mathcal{A}(Y;M))^2\Big]^{1/2}.$$

Let $\sigma_Y \equiv \left\{E\left[\,\left|Y - E[Y]\right|^2\,\right]\right\}^{1/2}$

Exercise 13 *Compute the integral* $I = \int_{[0,1]^d} f(x)dx$ *by the Monte Carlo method, where we assume* $f(x) : [0,1]^d \to \mathbf{R}$.

We have

$$\begin{aligned} I &= \int_{[0,1]^d} f(x)\, dx \\ &= \int_{[0,1]^d} f(x)p(x)\, dx \text{ (where } p \text{ is the uniform pdf)} \\ &= E[f(x)] \text{ (where } x \text{ is uniformly distributed in } [0,1]^d) \\ &\simeq \sum_{j=1}^{M} \frac{f(x(\omega_j))}{M} \\ &\equiv I_M, \end{aligned}$$

The values $\{x(\omega_j)\}$ are sampled uniformly in the cube $[0,1]^d$, by sampling the components $x_i(\omega_n)$ independently and uniformly on the interval $[0,1]$.

Remark 15 (Random number generators) *One can generate approximate random numbers, so called pseudo random numbers, see the lecture notes. By using transformations, one can also generate more complicated distributions in terms of simpler ones.*

Example 8 *Let Y be a given real valued random variable with*

$$P(Y \le x) = F_Y(x).$$

Suppose that we want to sample iid from Y and that we can cheaply compute $F_Y^{-1}(u)$, $u \in [0,1]$.

Then, take U to be uniform distributed in $[0,1]$ and let

$$Y(\omega) = F_Y^{-1}(U(\omega)).$$

We then have

$$P(Y \le x) = P(F_Y^{-1}(U) \le x) = P(U \le F_Y(x)) = F_Y(x)$$

as we wanted!

Acceptance-rejection sampling

It generates sampling values from an arbitrary pdf $\rho_Y(x)$ by using an auxiliary pdf $\rho_X(x)$.

Assumptions:

(i) It is simple to sample from ρ_X,

(ii) There exists $0 < \epsilon \leq 1$ s.t.

$$\epsilon \frac{\rho_Y}{\rho_X}(x) \leq 1, \text{ for all } x.$$

Idea:

Rejection sampling is usually used in cases where the form of ρ_Y makes sampling difficult.

Instead of sampling directly from ρ_Y, we use samples from ρ_X.

These samples from ρ_X are probabilistically accepted or rejected.

Acceptance-rejection sampling

The steps below generate a single realization of Y with pdf ρ_Y.

Step 1 Set $k = 1$

Step 2 Sample two independent random variables:
X_k from ρ_X and $U_k \sim U(0,1)$.

Step 3 **If** $U_k \leq \epsilon \frac{\rho_Y(X_k)}{\rho_X(X_k)}$ **then** *accept* $Y = X_k$ be a sample from ρ_Y.
Otherwise *reject* X_k, increment k by 1 and go to **Step 1**.

Let us see that Y sampled by acceptance-rejection has indeed density ρ_Y

We have the *acceptance probability*

$$
\begin{aligned}
P\left(U_k \le \epsilon \frac{\rho_Y(X_k)}{\rho_X(X_k)}\right) &= \int \int_0^{\epsilon \frac{\rho_Y(x)}{\rho_X(x)}} du \rho_X(x) dx \\
&= \epsilon \int \frac{\rho_Y(x)}{\rho_X(x)} \rho_X(x) dx \\
&= \epsilon \int \rho_Y(x) dx \\
&= \epsilon
\end{aligned}
$$

Let $K(\omega)$ be the first value of k for which X_k is accepted as a realization of Y. We want to show that X_K has the desired density, ρ_Y.

Consider an open set B

$$P(X_K \in B) = \sum_{k \geq 1} P(X_k \in B, K = k)$$

$$= \sum_{k \geq 1} \underbrace{P\left(X_k \in B, U_k \leq \epsilon \frac{\rho_Y(X_k)}{\rho_X(X_k)}\right)}_{\text{does not depend on } k} \prod_{m=1}^{k-1} \underbrace{P\left(U_m > \epsilon \frac{\rho_Y(X_m)}{\rho_X(X_m)}\right)}_{=1-\epsilon}$$

$$= P\left(X_k \in B, U_k \leq \epsilon \frac{\rho_Y(X_k)}{\rho_X(X_k)}\right) \underbrace{\sum_{k \geq 1} (1-\epsilon)^{k-1}}_{=1/\epsilon}$$

To finish compute

$$
\begin{aligned}
P\left(X_k \in B, U_k \le \epsilon \frac{\rho_Y(X_k)}{\rho_X(X_k)}\right) &= \int_B \int_0^{\epsilon \frac{\rho_Y(x)}{\rho_X(x)}} du \rho_X(x) dx \\
&= \epsilon \int_B \frac{\rho_Y(x)}{\rho_X(x)} \rho_X(x) dx \\
&= \epsilon \int_B \rho_Y(x) dx
\end{aligned}
$$

which implies

$$P(X_K \in B) = \int_B \rho_Y(x) dx$$

as we claimed.

Remark 16 (Acceptance-rejection cost) *Compute the expected number of samples per accepted ones:*

$$
\begin{aligned}
E[K] &= \sum_{k\geq 1} k\, P(K=k) \\
&= \sum_{k\geq 1} k\,(1-\epsilon)^k \epsilon \\
&= 1/\epsilon
\end{aligned}
$$

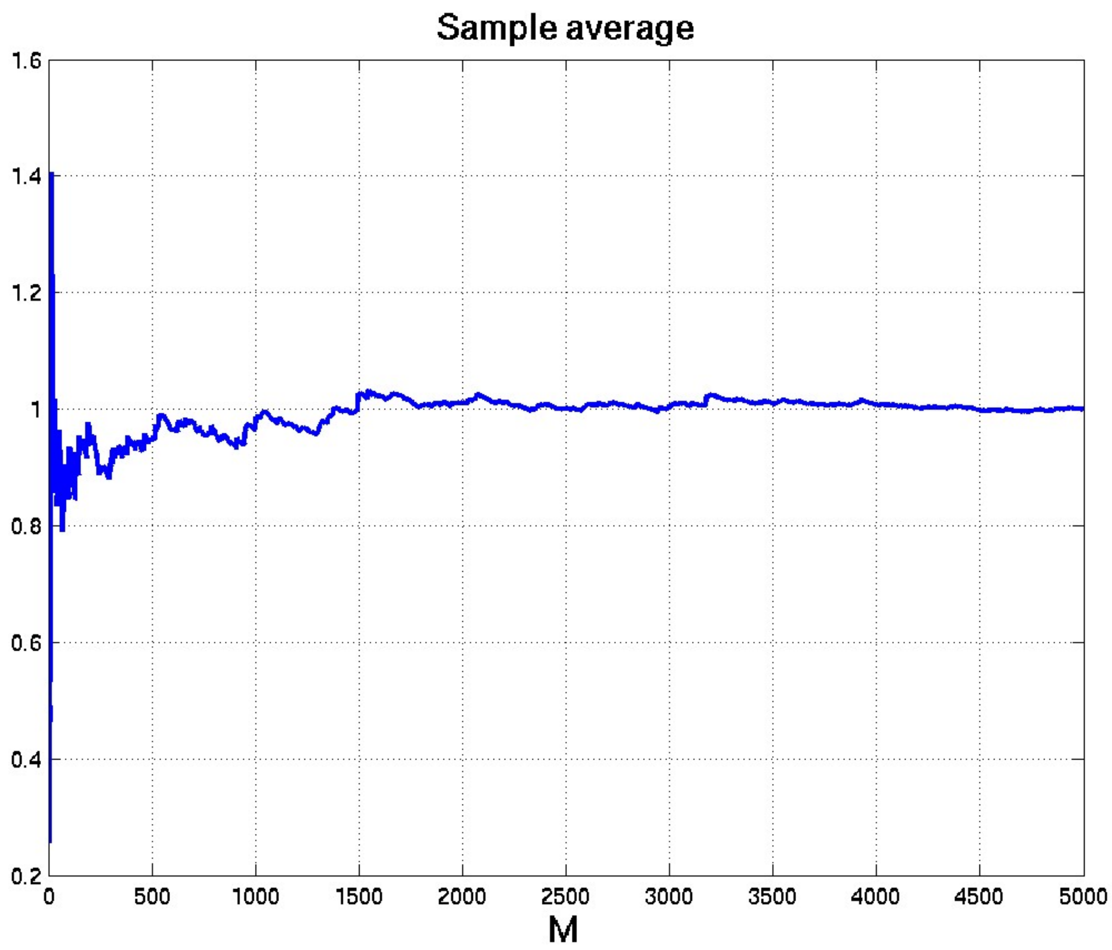
Can you interpret this result?

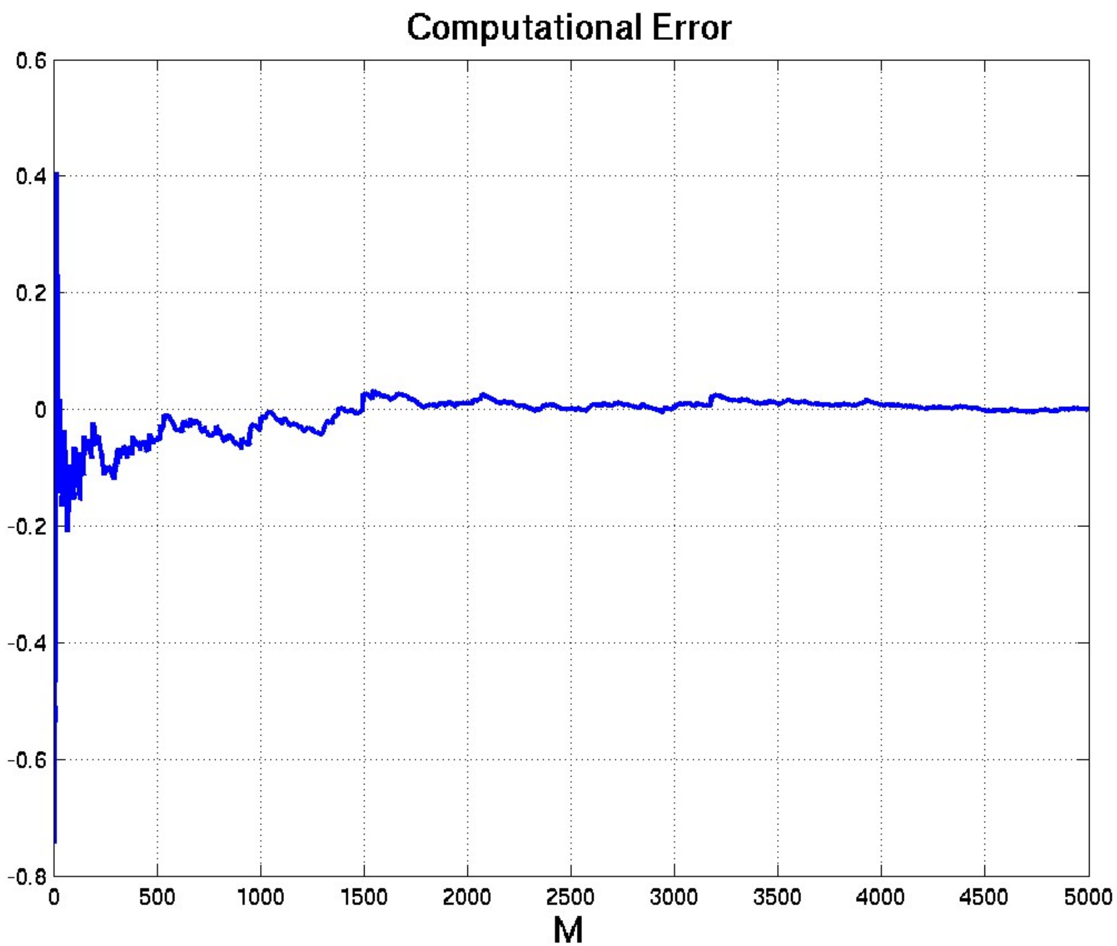
Monte Carlo: Numerical example

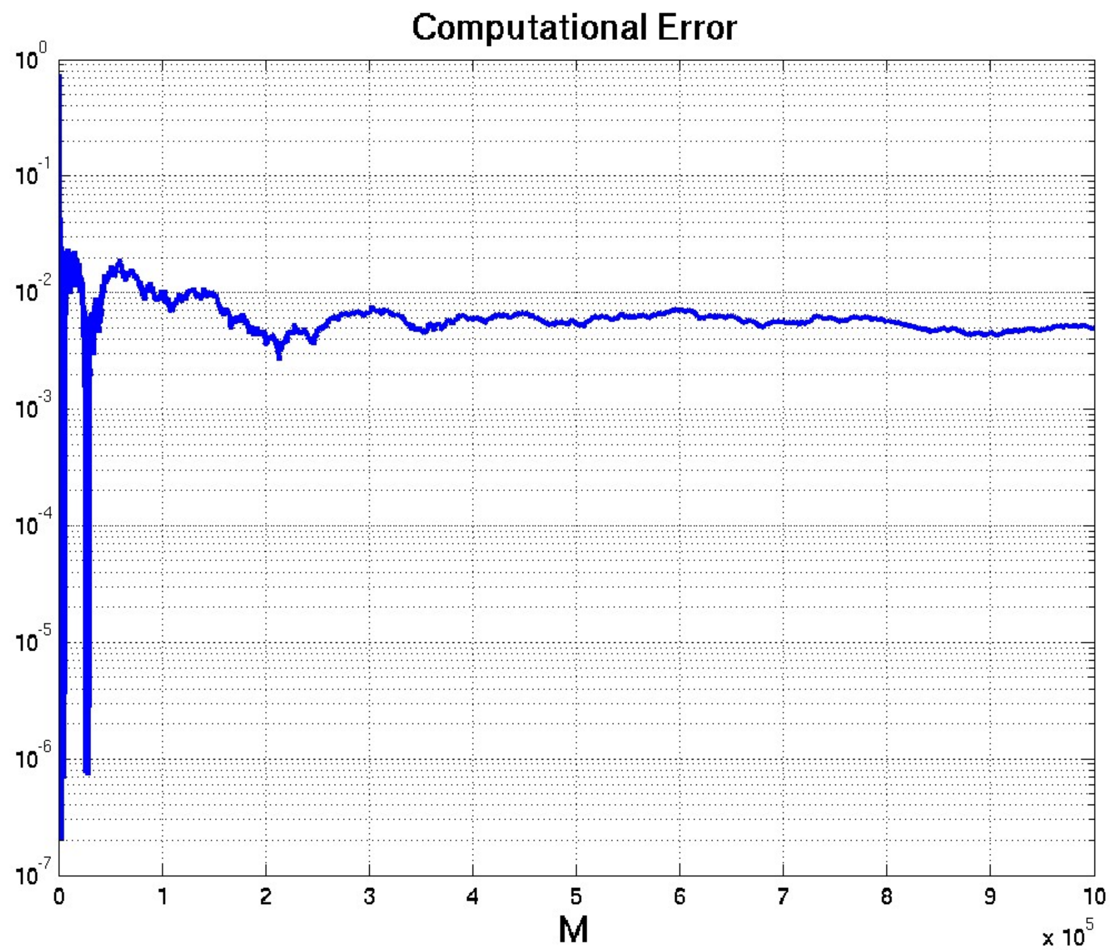
Consider the computation of the integral

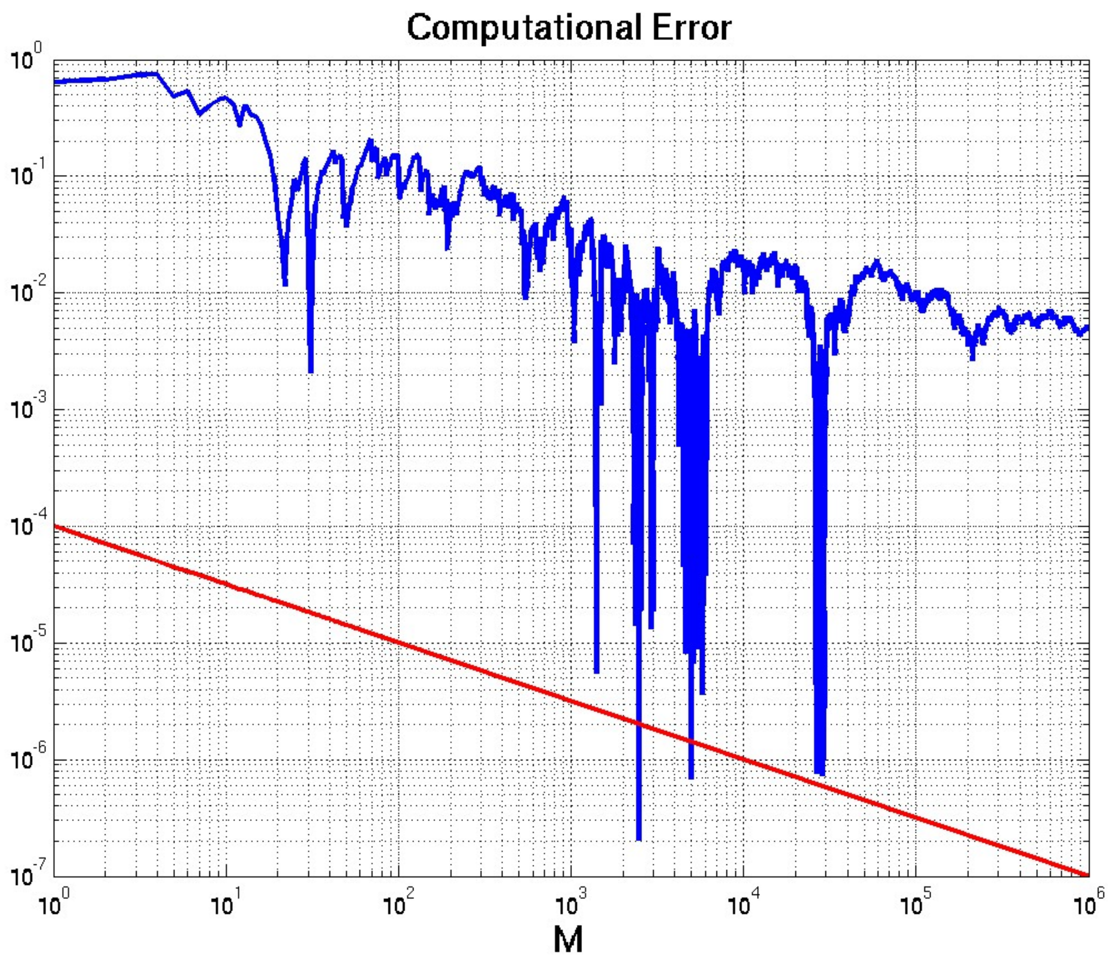
$$1 = \int_{[0,1]^N} \exp(\sum_{n=1}^{N} x_n) dx_1 \ldots dx_N / (e-1)^N$$

```
M = 1e6;  % Max. number of realizations
N = 20;   % Dimension of the problem
u = rand(M,N); f = exp(sum(u')');
run_aver = cumsum(f)./(((1:M)')*(exp(1)-1)^N);
plot(1:M, run_aver),
figure, plot(1:M, run_aver), xlabel 'M'
figure,plot(1:M,(run_aver-1)), xlabel 'M'
figure,semilogy(1:M,abs(run_aver-1)), xlabel 'M',
```


Sample average
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
M


Computational Error
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
M


Computational Error
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
0 1 2 3 4 5 6 7 8 9 10
M
x 10^5


Computational Error
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
10^{6}
M

Monte Carlo error analysis Consider the scaled random variable

$$Z_M \equiv \frac{\sqrt{M}}{\sigma_Y}\big(\mathcal{A}(Y;M) - E[Y]\big)$$

with cumulative distribution function

$$F_{Z_M}(x) \equiv P(Z_M \leq x), \quad \forall\, x \in \mathbb{R}.$$

The Central Limit Theorem is the fundamental result to understand the statistical error of Monte Carlo methods.

Theorem 10 (The Central Limit Theorem) *Assume* $\xi_j,\ j = 1, 2, 3, \ldots$ *are independent, identically distributed (i.i.d) and* $E[\xi_j] = 0$, $E[\xi_j^2] = 1$. *Then*

$$\sum_{j=1}^{M} \frac{\xi_j}{\sqrt{M}} \rightharpoonup \nu, \tag{30}$$

where ν *is* $N(0,1)$ *and* $\rightharpoonup$ *denotes convergence of the distributions, also called weak convergence, i.e. the convergence (36) means* $E[g(\sum_{j=1}^{M} \xi_j/\sqrt{M})] \to E[g(\nu)]$ *for all bounded and continuous functions* g.

Characteristic function Let X be a r.v. then

$$f(t) = E[e^{itX}]$$

is called the *characteristic function* of X. This function identifies completely the distribution of X, namely

Theorem 11 *Two distributions having the same characteristic function are identical*

Example: Consider a standard normal distribution, $X \sim N(0,1)$. Then

$$f(t) = E[e^{itX}] = e^{-\frac{t^2}{2}}$$

In fact, we have inversion formulas closely related to the Fourier transform[a]

Theorem 12 *Let x_1, x_2 be continuity points of F_X. Then*

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t)dt$$

[a] See [Petrov]

Proof. Consider the characteristic function $f(t) = E[e^{it\xi_j}]$. Then its derivatives satisfy

$$f^{(m)}(t) = E[i^m \xi_j^m e^{it\xi_j}]. \tag{31}$$

For the sample average of the ξ_j vars we have

$$\begin{aligned} E[e^{it\sum_{j=1}^M \xi_j/\sqrt{M}}] &= f\left(\frac{t}{\sqrt{M}}\right)^M \\ &= \big(f(0) + \frac{t}{\sqrt{M}} f'(0) + \frac{1}{2}\frac{t^2}{M} f''(0) + o\left(\frac{t^2}{M}\right)\big)^M. \end{aligned}$$

The representation (31) implies

$$\begin{aligned} f(0) &= E[1] = 1, \\ f'(0) &= iE[\xi_n] = 0, \\ f''(0) &= -E[\xi_n^2] = -1. \end{aligned}$$

Therefore

$$\begin{aligned} E[e^{it\sum_{j=1}^M \xi_j/\sqrt{M}}] &= \left(1 - \frac{t^2}{2M} + o\left(\frac{t^2}{M}\right)\right)^M \\ &\to e^{-t^2/2}, \quad \text{as } M \to \infty \\ &= \int_{\mathbb{R}} \frac{e^{itx} e^{-x^2/2}}{\sqrt{2\pi}} dx, \end{aligned} \tag{32}$$

and we conclude that the Fourier transform of the pdf (i.e. the characteristic function) of $\sum_{j=1}^{M} \xi_j/\sqrt{M}$ converges to the Fourier transform of the standard normal distribution. Therefore,

$$
\begin{aligned}
E[g(\sum_{j=1}^{M} \xi_j/\sqrt{M})] &= \int_{\mathbb{R}} g(x)\rho_{\sum_{j=1}^{M} \xi_j/\sqrt{M}}(x)dx \\
&\underbrace{=}_{Parseval} \int_{\mathbb{R}} f(t)F(g)(t)dt \\
&\to \int_{\mathbb{R}} e^{-t^2/2} F(g)(t)dt \\
&\underbrace{=}_{Parseval} E[g(\nu)].
\end{aligned}
$$

Exercise 14 *What is the error of $I_M - I$ in Example 13?*

Let the error ϵ_M be defined by

$$\begin{aligned}\epsilon_M &= \sum_{j=1}^{M} \frac{f(x_j)}{M} - \int_{[0,1]^d} f(x)dx \\ &= \sum_{j=1}^{M} \frac{f(x_j) - E[f(x)]}{M}.\end{aligned}$$

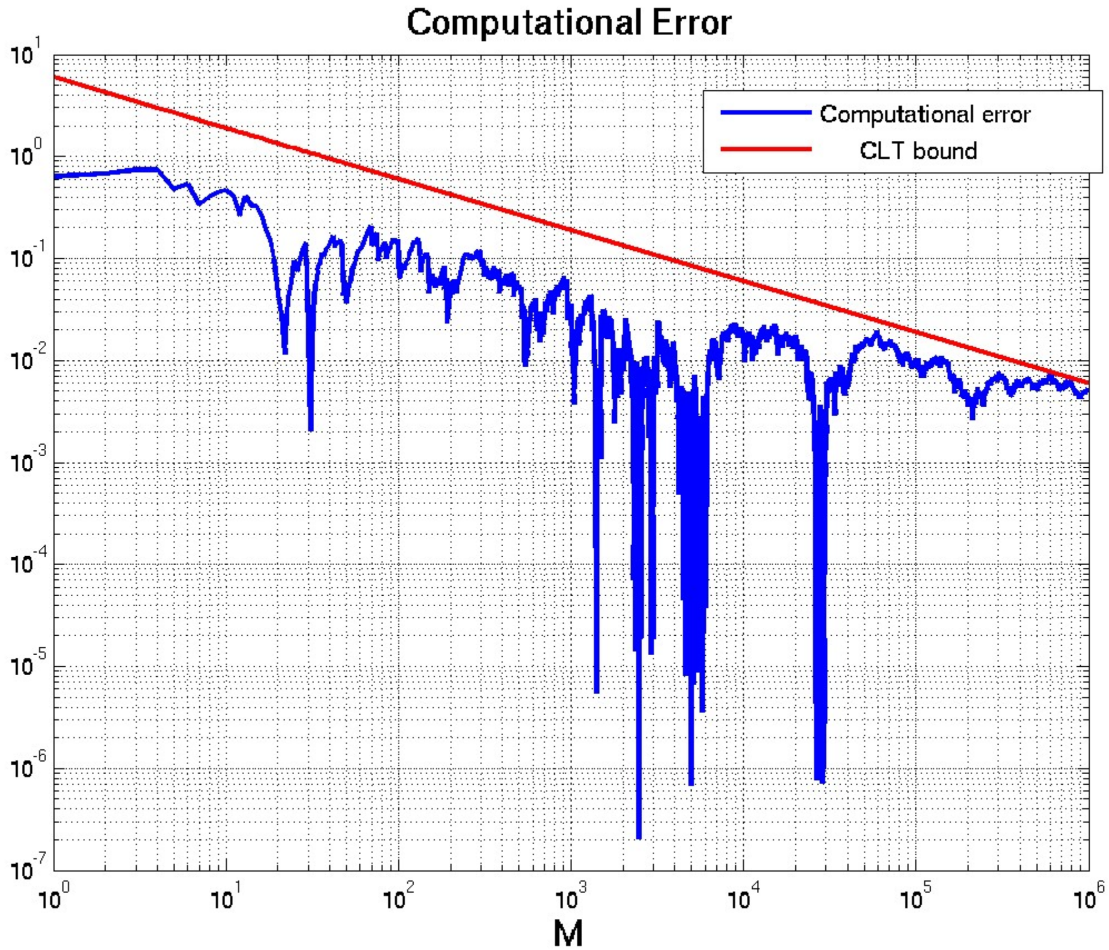
By the Central Limit Theorem, $\sqrt{M}\epsilon_M \rightharpoonup \sigma\nu$, where ν is

$N(0,1)$ and

$$\begin{aligned}\sigma^2 &= \int_{[0,1]^d} f^2(x)dx - \left(\int_{[0,1]^d} f(x)dx\right)^2 \\ &= \int_{[0,1]^d} \left(f(x) - \int_{[0,1]^d} f(x)dx\right)^2 dx.\end{aligned}$$

In practice, σ^2 is approximated by

$$\hat{\sigma}^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left(f(x_j) - \sum_{m=1}^{M}\frac{f(x_m)}{M}\right)^2.$$



Approximate error bound: $C_\alpha \hat{\sigma}/\sqrt{M}$. Here $C_\alpha = 3$.

Theorem 13 (Berry–Esseen) **Assume**

$$\lambda \equiv \frac{\left(E\left[\left|Y - E[Y]\right|^3\right] \right)^{1/3}}{\sigma_Y} < +\infty,$$

then we have a uniform estimate in the central limit theorem

$$|F_{Z_M}(x) - \Phi(x)| \leq \frac{C_{BE}\,\lambda^3}{(1+|x|)^3\sqrt{M}}$$

Here Φ *is the distribution function of* $N(0,1)$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{s^2}{2}\right) ds. \tag{33}$$

and $C_{BE} = 30.51175...$

By the Berry–Esseen thm., the statistical error

$$\mathcal{E}_S(Y;M) \equiv E[Y] - \mathcal{A}(Y;M)$$

satisfies, $\forall\, c_0 > 0$,

$$P\left(\left[|\mathcal{E}_S(Y;M)| \le c_0 \frac{\sigma_Y}{\sqrt{M}}\right]\right) \ge 2\Phi(c_0) - 1 - \frac{C_{BE}\,\lambda^3}{(1+c_0)^3\sqrt{M}}.$$

In practice choose $c_0 \geq 1.65, \Rightarrow 1 > 2\Phi(c_0) - 1 \geq 0.901$ and the event

$$|\mathcal{E}_S(Y;M)| \leq \mathtt{E}_S(Y;M) \equiv c_0 \, \frac{\mathcal{S}(Y;M)}{\sqrt{M}} \tag{34}$$

has probability close to one, which involves the additional step to approximate σ_Y by $\mathcal{S}(Y;M)$.Thus, in the computations $\mathtt{E}_S(Y;M)$ is a good approximation of the statistical error $\mathcal{E}_S(Y;M)$.

Numerical Example:

Taking $c_0 = 3$ yields $2\Phi(c_0) - 1 = 0.9973...$ and

$$P\left(\left[|\mathcal{E}_S(Y;M)| \leq 3\frac{\sigma_Y}{\sqrt{M}}\right]\right) \geq 0.9973 - 0.4766\frac{\lambda^3}{\sqrt{M}}.$$

In particular, if Y is a uniform random variable, then

$$\lambda^3 = \frac{1}{4}3^{1.5} = 1.299...$$

and we have the bound

$$P\left(\left[|\mathcal{E}_S(Y;M)| \leq 3\frac{\sigma_Y}{\sqrt{M}}\right]\right) \geq 0.9973 - \frac{0.6196}{\sqrt{M}}.$$

Obs: the last term on the right will determine the confidence level for $M \leq 5 \times 10^4$

Numerical Example:

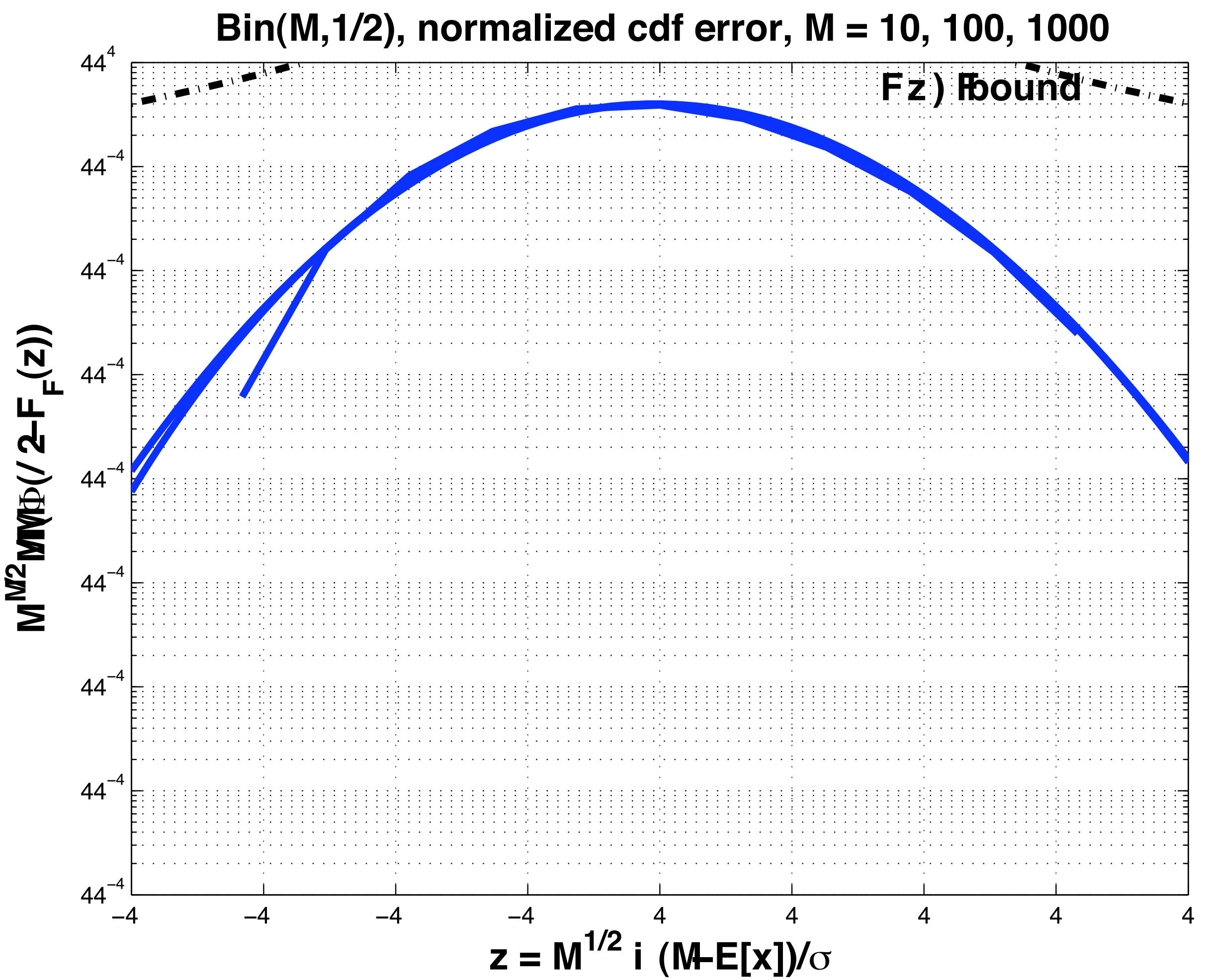
Consider a Binomial r.v. with parameter $p = 1/2$,

$$X = \sum_{i=1}^{M} Y_i$$

and Y_i iid Bernoulli r.vars., $\sigma^2 = p(1-p)$. Let

$$Z = \frac{(X - Mp)}{\sigma\sqrt{M}},$$

then we compare its cdf (computed exactly) vs. the CLT approximation, $\Phi(z)$. We do it for several values of M ...


Bin(M,1/2), normalized cdf error, M = 10, 100, 1000
Fz) Fbound
z = M^1/2 i (M-E[x])/σ

Adaptive Monte Carlo

For a given $TOL_S > 0$, the goal is to find M such that $\mathtt{E}_S(Y;M) \leq TOL_S$. The following algorithm adaptively finds the number of realizations M to compute the sample average $\mathcal{A}(Y;M)$ as an approximation to $E[Y]$. With probability close to one, depending on c_0, the statistical error in the approximation is then bounded by TOL_S.

`routine Monte-Carlo`$(TOL_S, Y, M_0; EY)$

Set the batch counter $m = 1$, $M[1] = M_0$ and

$\mathtt{E}_S[1] = 2\,TOL_S$.

Do while $(\mathtt{E}_S[m] > TOL_S)$

Compute $M[m]$ <u>new</u> samples of Y, along with the sample

average $EY \equiv \mathcal{A}(Y; M[m])$, the sample variance

$\mathcal{S}[m] \equiv \mathcal{S}(Y; M[m])$ and

the deviation $\mathtt{E}_S[m+1] \equiv \mathtt{E}_S(Y; M[m])$.

Compute $M[m+1]$ by

`change_`M $(M[m], \mathcal{S}[m], TOL_S; M[m+1])$.

Increase m by 1.

end-do

`end of Monte-Carlo`

routine change_M $(M_{\text{in}}, \mathcal{S}_{\text{in}}, TOL_S; M_{\text{out}})$

$$\begin{aligned} M^* &= \min\left\{\text{integer part } \left(\frac{c_0\, \mathcal{S}_{\text{in}}}{TOL_S}\right)^2,\ \text{MCH} \times M_{\text{in}}\right\} \\ n &= \text{integer part } (\log_2 M^*) + 1 \\ M_{\text{out}} &= 2^n. \end{aligned} \tag{35}$$

end of change_M

Remark 17 (Parameters for change M) *Here, M_0 is a given initial value for M, and $MCH > 1$ is a positive integer parameter introduced to avoid a large new number of realizations in the next batch due to a possibly inaccurate sample standard deviation $\mathcal{S}[m]$. Indeed, $M[m+1]$ cannot be greater than $MCH \times M[m]$.*

We will use $MCH = 2$ in the next example:

Numerical Example: Adaptive MC, $TOL = 1e - 2$ **for** $1 = \int_{[0,1]^N} \exp(\sum_{n=1}^N x_n) dx_1 \dots dx_N/(e-1)^N$, $N = 20$.

M	Sample E	Sample std	Error est.	Comp. Error
200	1.2526	4.4003e+00	9.3e-01	2.5e-01
400	1.0411	2.1068e+00	3.1e-01	4.1e-02
800	0.9889	1.8730e+00	2.0e-01	-1.1e-02
1600	1.0699	2.3410e+00	1.7e-01	7.0e-02
3200	1.0324	1.9087e+00	1.0e-01	3.2e-02
6400	1.0764	2.1659e+00	8.1e-02	7.6e-02
12800	1.0212	2.0788e+00	5.5e-02	2.1e-02
25600	9.9549	1.9036e+00	3.6e-02	-4.5e-03
51200	1.0104	1.8946e+00	2.5e-02	1.0e-02
102400	0.98721	1.8248e+00	1.7e-02	-1.3e-02
204800	0.99375	1.9103e+00	1.3e-02	-6.6e-03

```
409600  0.99611   1.9320e+00   9.1e-03  -3.9e-03
```

Question: Can you compute the confidence level corresponding to the above computations as a function of M using the BE Theorem?

Large Deviations

Theory for rare events, deep in the distribution tails.

Remember CLT and BET: Assume $\xi_j,\ j = 1, 2, 3, \ldots$ are independent, identically distributed (i.i.d) and $E[\xi_j] = 0$, $E[\xi_j^2] = 1$. Then

$$\sum_{j=1}^{M} \frac{\xi_j}{\sqrt{M}} \rightharpoonup \nu, \tag{36}$$