

Monte Carlo and Empirical Methods for Stochastic Inference (MASM11/FMS091)

Magnus Wiktorsson

Centre for Mathematical Sciences
Lund University, Sweden

Lecture 3

Importance sampling
January 29, 2019

Last time: MC output analysis

- We used the CLT

$$\sqrt{N} (\tau_N - \tau) \xrightarrow{d.} \mathcal{N}(0, \sigma^2(\phi))$$

to target τ by the approximate two-sided confidence interval

$$\left(\tau_N - \lambda_{\alpha/2} \frac{\sigma(\phi)}{\sqrt{N}}, \tau_N + \lambda_{\alpha/2} \frac{\sigma(\phi)}{\sqrt{N}} \right).$$

- In addition, we discussed how to estimate $\varphi(\tau)$ for some function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ having at hand an estimator τ_N of τ . If $\varphi \in \mathcal{C}^1$ one may prove the CLT

$$\sqrt{N} (\varphi(\tau_N) - \varphi(\tau)) \xrightarrow{d.} \mathcal{N}(0, \varphi'(\tau)^2 \sigma^2(\phi)).$$

Consequently, the natural estimator $\varphi(\tau_N)$ works fine, at least asymptotically (but suffers in general from bias for finite N 's).

Example: Buffon's needle

Consider a wooden floor with parallel boards of width d on which we randomly drop a needle of length ℓ , with $\ell \leq d$. Let

$$\begin{cases} X = \text{distance from the lower needlepoint to the upper board edge line,} \\ \theta = \text{angle between the needle and the board edge normal.} \end{cases}$$

Then

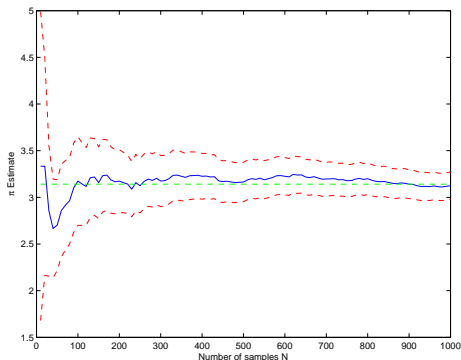
$$\tau = \mathbb{P}(\text{needle intersects board edge}) = \mathbb{P}(X \leq \ell \cos \theta) = \dots = \frac{2\ell}{\pi d}.$$

or, equivalently,

$$\pi = \frac{2\ell}{\tau d}.$$

Example: Buffon's needle (cont.)

Since $\tau = \mathbb{P}(\text{needle intersects board edge}) = \mathbb{E}(\mathbb{1}_{\{X \leq l \cos \theta\}})$ can be easily estimated by means of MC, an approximation of $\pi = \varphi(\tau) = 2l/(\tau d)$ can be obtained via the delta method:



Last time: pseudo-random number generation

- We discussed (briefly) how to generate pseudo-random, uniformly distributed numbers (U_n) using the linear congruential generator

$$U_n = (a \cdot U_{n-1} + c) \pmod{m}.$$

- Having at hand such $\mathcal{U}(0, 1)$ -distributed numbers U , we also looked at how to generate random numbers X from an arbitrary distribution F by means of the inversion method, i.e., by letting

$$X = F^{\leftarrow}(U) = \inf\{x \in \mathbb{R} : F(x) \geq U\}.$$

Conditional methods

Let f be a multivariate density on \mathbb{R}^d . By decomposing f into conditional densities according to

$$f(x_1, \dots, x_d) = f(x_1) \prod_{\ell=2}^d f(x_\ell | x_1, \dots, x_{\ell-1}),$$

the problem of sampling from a multivariate density can be reduced to that of sampling from several univariate densities:

```
draw  $X_1 \sim f(x_1)$   
for  $\ell = 2 \rightarrow d$  do  
  draw  $X_\ell \sim f(x_\ell | X_1, \dots, X_{\ell-1})$   
end for  
return  $X = (X_1, \dots, X_d)$ 
```

Trivially, the resulting draw X has the correct distribution f . This method presumes that the conditional densities are easily obtained, which is not always the case.

Rejection sampling

- In many cases we do not know the inverse of F or not even the normalizing constant of the density f . However, if g is another density such that $f(x) \leq Kg(x)$ for all $x \in \mathbb{R}^d$ and some constant $K < \infty$, we may use rejection sampling:

repeat

 draw $X^* \sim g$

 draw $U \sim \mathcal{U}(0, 1)$

until $U \leq \frac{f(X^*)}{Kg(X^*)}$

$X \leftarrow X^*$

return X

Rejection sampling (cont)

Theorem (Rejection sampling)

The output X of the rejection sampling algorithm is a random variable with density function f . Moreover, the expected number of trials needed before acceptance is K .

Example

We wish to simulate $f(x) = \exp(\cos^2(x))/c$, $x \in (-\pi/2, \pi/2)$, where $c = \int_{-\pi/2}^{\pi/2} \exp(\cos^2(z)) dz = \pi e^{1/2} I_0(1/2)$ is the normalizing constant.

However, since for all $x \in (-\pi/2, \pi/2)$,

$$f(x) = \frac{\exp(\cos^2(x))}{c} \leq \frac{e}{c} = \underbrace{\frac{e\pi}{c}}_K \times \underbrace{\frac{1}{\pi}}_g,$$

where g is the density of $\mathcal{U}(-\pi/2, \pi/2)$, we may use rejection sampling where a candidate $X^* \sim \mathcal{U}(-\pi/2, \pi/2)$ is accepted if

$$U \leq \frac{f(X^*)}{Kg(X^*)} = \frac{\exp(\cos^2(X^*))/c}{e/c} = \exp(\cos^2(X^*) - 1).$$

```

prob = @(x) exp((cos(x))^2 - 1);
trial = 1;
accepted = false;
while ~accepted,
    Xcand = - pi/2 + pi*rand;
    if rand < prob(Xcand),
        accepted = true;
        X = Xcand;
    else
        trial = trial + 1;
    end
end
    
```

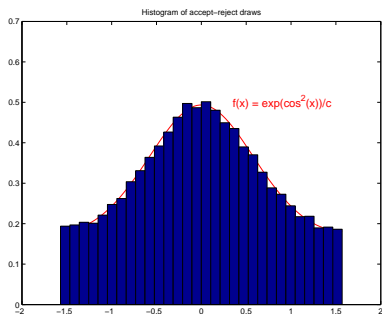


Figure: Plot of a histogram of 20,000 accept-reject draws together with the true density. The average number of trials was $1.5555 (\approx K = e^{1/2}/I_0(1/2) \approx 1.5503)$.

Plan of today's lecture

- 1 Importance sampling (IS)
- 2 Self-normalized IS
- 3 Home assignment 1 (HA1)

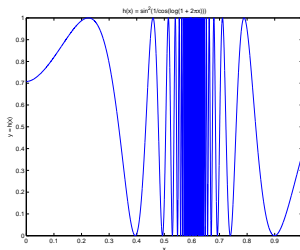
We are here → ●

- 1 Importance sampling (IS)
- 2 Self-normalized IS
- 3 Home assignment 1 (HA1)

Advantages of the MC method

The MC method

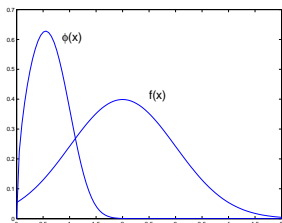
- is more efficient than deterministic methods in high dimensions,
- does in general not require knowledge of the normalizing constant of a density for computing expectations, and
- handles efficiently “strange” integrands that may cause problems for deterministic methods.



Problems with MC integration

OK, MC integration looks promising. We may however run into problems if

- it is hard to sample from f or
- if the integrand ϕ and the density f are dissimilar; in this case we will end up with a lot of draws where the integrand is small, and consequently only a few draws will contribute to the estimate. This gives a large variance.



These problems can often be solved using **importance sampling**.

Importance sampling (IS, Ch. 6.4.1)

The basis of importance sampling is to take an **instrumental density** g on X such that $g(x) = 0 \Rightarrow f(x) = 0$ and rewrite the integral as

$$\begin{aligned} \tau &= \mathbb{E}_f(\phi(X)) = \int_{\mathcal{X}} \phi(x) f(x) \, dx = \int_{f(x) > 0} \phi(x) f(x) \, dx \\ &= \int_{g(x) > 0} \phi(x) \frac{f(x)}{g(x)} g(x) \, dx = \mathbb{E}_g \left(\phi(X) \frac{f(X)}{g(X)} \right) = \mathbb{E}_g(\phi(X) \omega(X)), \end{aligned}$$

where

$$\omega : \{x \in X : g(x) > 0\} \ni x \mapsto \frac{f(x)}{g(x)}$$

is the so-called **importance weight function**.

Importance sampling (IS, Ch. 6.4.1) (Generalization)

The basis of importance sampling is to take an **instrumental density** g on X such that $g(x) = 0 \Rightarrow \phi(x)f(x) = 0$ and rewrite the integral as

$$\begin{aligned} \tau &= \mathbb{E}_f(\phi(X)) = \int_{\mathcal{X}} \phi(x)f(x) \, dx = \int_{|\phi(x)|f(x)>0} \phi(x)f(x) \, dx \\ &= \int_{g(x)>0} \phi(x) \frac{f(x)}{g(x)} g(x) \, dx = \mathbb{E}_g \left(\phi(X) \frac{f(X)}{g(X)} \right) = \mathbb{E}_g(\phi(X)\omega(X)), \end{aligned}$$

where

$$\omega : \{x \in X : g(x) > 0\} \ni x \mapsto \frac{f(x)}{g(x)}$$

is the so-called **importance weight function**.

Importance sampling (cont.)

We may now estimate $\tau = \mathbb{E}_g(\phi(X)\omega(X))$ using standard MC:

```

for  $i = 1 \rightarrow N$  do
    draw  $X_i \sim g$ 
end for
set  $\tau_N \leftarrow \sum_{i=1}^N \phi(X_i)\omega(X_i)/N$ 
return  $\tau_N$ 
    
```

Here, trivially,

$$\mathbb{V}(\tau_N) = \frac{1}{N} \mathbb{V}_g(\phi(X)\omega(X))$$

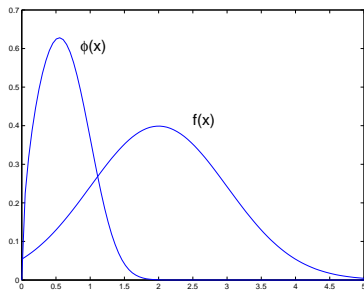
and we should thus aim at choosing g so that the function $x \mapsto \phi(x)\omega(x)$ is close to constant in the support of g . This gives a minimal variance.

Example: A tricky normal expectation

Let X have $\mathcal{N}(2, 1)$ distribution and try to compute

$$\tau = \mathbb{E} \left(\mathbb{1}_{X \geq 0} \sqrt{X} \exp(-X^3) \right) = \int \underbrace{\mathbb{1}_{x \geq 0} \sqrt{x} \exp(-x^3)}_{=\phi(x)} \underbrace{\mathcal{N}(x; 2, 1)}_{=f(x)} dx,$$

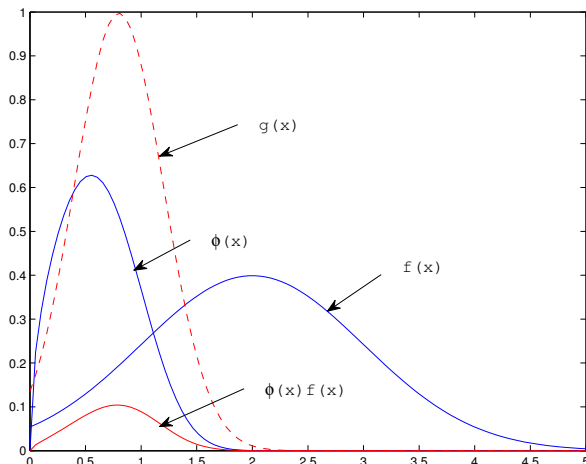
where $\mathcal{N}(x; \mu, \sigma^2)$ denotes the density of the normal distribution.



Here the support of f is significantly larger than that of ϕ .

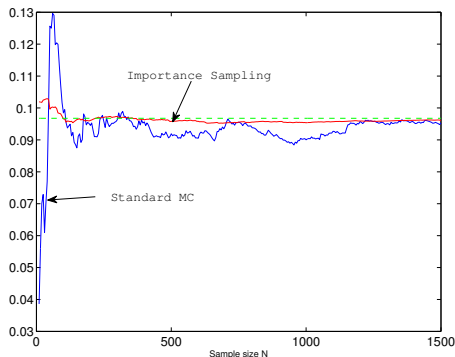
Example: A tricky normal expectation (cont.)

Thus, standard MC will lead to a waste of computational power. Better is to use IS with g being a scale-location-transformed normal-distribution:



Example: A tricky normal expectation (cont.)

```
phi = @(x) (x >= 0) .*sqrt(x) .*exp(- x.^3);  
mu = 0.8;  
sigma = 0.4;  
omega = @(x) normpdf(x,2,1) ./normpdf(x,mu,sigma);  
X = sigma*randn(1,N)+mu;  
tau = mean(phi(X) .*omega(X));
```



We are here → ●

- 1 Importance sampling (IS)
- 2 Self-normalized IS
- 3 Home assignment 1 (HA1)

Self-normalized IS (Ch. 6.4.1)

Often $f(x)$ is known only up to a normalizing constant $c > 0$, i.e. $f(x) = z(x)/c$, where we can evaluate $z(x) = cf(x)$ but not $f(x)$. Then, as before,

$$\begin{aligned} \tau = \mathbb{E}_f(\phi(X)) &= \int_{\mathbf{X}} \phi(x) f(x) \, dx = \frac{c \int_{f(x) > 0} \phi(x) f(x) \, dx}{c \int_{f(x) > 0} f(x) \, dx} \\ &= \frac{\int_{g(x) > 0} \phi(x) \frac{cf(x)}{g(x)} g(x) \, dx}{\int_{g(x) > 0} \frac{cf(x)}{g(x)} g(x) \, dx} = \frac{\int_{g(x) > 0} \phi(x) \frac{z(x)}{g(x)} g(x) \, dx}{\int_{g(x) > 0} \frac{z(x)}{g(x)} g(x) \, dx} \\ &= \frac{\int_{g(x) > 0} \phi(x) \omega(x) g(x) \, dx}{\int_{g(x) > 0} \omega(x) g(x) \, dx} = \frac{\mathbb{E}_g(\phi(X) \omega(X))}{\mathbb{E}_g(\omega(X))}, \end{aligned}$$

where we are able to evaluate

$$\omega : \{x \in \mathbf{X} : g(x) > 0\} \ni x \mapsto \frac{z(x)}{g(x)}.$$

Self-normalized IS (cont.)

Thus, having generated a sample X_1, \dots, X_N from g we may estimate the numerator $\mathbb{E}_g(\phi(X)\omega(X))$ as well as the denominator $\mathbb{E}_g(\omega(X))$ using standard MC:

$$\begin{aligned}\tau &= \frac{\mathbb{E}_g(\phi(X)\omega(X))}{\mathbb{E}_g(\omega(X))} \\ &\approx \frac{\frac{1}{N} \sum_{i=1}^N \phi(X_i)\omega(X_i)}{\frac{1}{N} \sum_{\ell=1}^N \omega(X_\ell)} = \sum_{i=1}^N \underbrace{\frac{\omega(X_i)}{\sum_{\ell=1}^N \omega(X_\ell)}}_{\text{normalized weight}} \phi(X_i).\end{aligned}$$

Note that the denominator yields an estimate of the normalizing constant c :

$$c = \mathbb{E}_g(\omega(X)) \approx \frac{1}{N} \sum_{\ell=1}^N \omega(X_\ell).$$

Example

We reconsider the density

$$f(x) = \exp(\cos^2(x))/c, \quad x \in (-\pi/2, \pi/2),$$

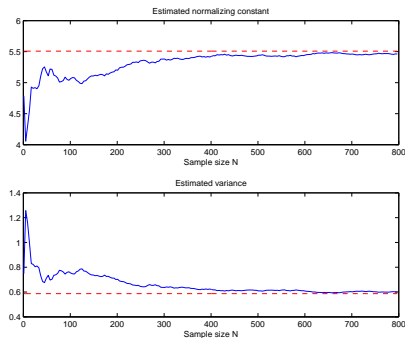
treated last time and estimate its variance as well as the normalizing constant $c > 0$ using self-normalized IS.

Let the instrumental distribution g be the uniform distribution $\mathcal{U}(-\pi/2, \pi/2)$.

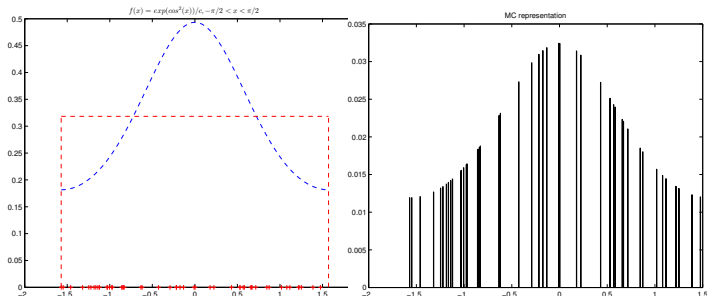
Example (cont.)

```

z = @(x) exp(cos(x).^2);
X = -pi/2 + pi*rand(1,N);
omega = @(x) pi*z(x);
tau = cumsum(X.^2.*omega(X))./cumsum(omega(X));
c = cumsum(omega(X))./(1:N);
subplot(2,1,1);
plot(1:N,c);
subplot(2,1,2);
plot(1:N,tau);
    
```



The weighted sample $(X_i, \omega(X_i))$ can be viewed as an MC representation of the target distribution f .



$$f(x) \xRightarrow{\text{IS}} (X_i, \omega(X_i))$$

We are here → ●

- 1 Importance sampling (IS)
- 2 Self-normalized IS
- 3 Home assignment 1 (HA1)

HA1: Simulation and Monte Carlo integration

HA1 comprises

- one question on random number generation and
- two larger questions on IS (one- and two-dimensional problems) containing
- one sub question (2(c)) on variance reduction (which we will discuss next time).

Submission:

- A written report in PDF format (*No MS Word-files*). The pair Sverker Persson and Lilith Nilsson name their report file `proj1-nlps.pdf`. A printed and stitched copy of the report is given to the lecturer at the very beginning of the lecture on **Tuesday 12 Feb**.
- An email containing the report file as well as *all* your m-files with a file `proj1.m` that runs your analysis. This email has to be sent to `fms091@matstat.lu.se` before **Tuesday 12 Feb, 13:00:00** (that is, 15 minutes before the beginning of the lecture). Use your STIL-IDs in the subject line of the email. Set the subject line to:
“Project 1 by “STILID1” and “STILID2” “
- Late submissions do not qualify for marks higher than 3.

Instructions on report writing

- Explain carefully all introduced notation: $X = ?$.
- Describe/explain the model.
- The text should be readable **without access to the Matlab code**; write plain text instead of including Matlab code in the report.
- Include your solutions in the text; do not write “calculations of ? can be found in the matlab code”, or similar.
- When referring to the lecture notes or the book, be specific (i.e. refer to Chapter/which lecture).
- Refer to your figures in the text. Explain colors etc. in the figure captions (a figure caption is almost never to long).
- Write clear motivations and discussions when it concerns choice of instrumental distributions etc.