Monte Carlo methods with applications

Vivekananda Roy

Iowa State University

June 10, 2019
Chennai Mathematical Institute
Outline

1. Monte Carlo integration
2. Markov chain Monte Carlo
3. Bayesian statistics
1 Monte Carlo integration
2 Markov chain Monte Carlo
3 Bayesian statistics
Monte Carlo methods

We Want to know:

\[ \lambda = \int_S h \, d\pi, \]

which is analytically intractable. Here \( \pi \) is a prob. measure and \( h \) is integrable.

**Ordinary Monte Carlo** is the method of using IID simulations \( X_1, \ldots, X_n \) from \( \pi \) to approximate expectations by sample averages

\[ \bar{h}_n = \frac{1}{n} \sum_{i=1}^n h(X_i). \]

By law of large numbers (LLN), if \( E_\pi |h| < \infty \),

\[ \bar{h}_n \xrightarrow{\text{as}} E_\pi h \equiv \lambda \text{ as } n \to \infty. \]
Monte Carlo error

By SLLN, $\bar{h}_n \xrightarrow{as} E_\pi h$ as $n \to \infty$.

**How do we compute an associated standard error?**

By CLT if $E_\pi h^2 < \infty$,

$$\sqrt{n}(\bar{h}_n - E_\pi h) \xrightarrow{d} N(0, \sigma^2_{\bar{h}}).$$

$$s^2_{\bar{h}} = \frac{1}{n} \sum_{i=1}^{n} (h(X_i) - \bar{h}_n)^2.$$ 

The sample variance $s^2_{\bar{h}}$ is a consistent estimator of $\sigma^2_{\bar{h}}$.

**How large should $n$ be?**

Asymptotic 95% CI for $E_\pi h$: $\bar{h}_n \pm 2s_{\bar{h}}/\sqrt{n}$
Toy Examples

Find

\[ \int_{-\infty}^{\infty} x \exp\left( -\frac{(x - 1)^2}{2} \right) dx \]

\[ \int_{-\infty}^{\infty} x \sin(x) \exp\left( -\frac{(x - 1)^2}{2} \right) dx \]
Toy Examples

Find

\[\int_{-\infty}^{\infty} \sqrt{2\pi} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - 1)^2}{2}\right) \, dx\]

\[\int_{-\infty}^{\infty} \sqrt{2\pi} x \sin(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - 1)^2}{2}\right) \, dx\]
Toy Examples

```r
set.seed(3)

n <- 1000
x <- rnorm(n, me=1)
y <- sqrt(2*pi)*x
est <- mean(y)
est
mcse <- sd(y) / sqrt(n)
interval <- est + c(-1,1)*1.96*mcse
interval

y <- sqrt(2*pi)*x*sin(x)
est <- mean(y)
est
mcse <- sd(y) / sqrt(n)
interval <- est + c(-1,1)*1.96*mcse
interval
```
Find

\[ \int_0^{\infty} \frac{x^2}{2} \exp\left(-\frac{x}{2}\right) dx \]

\[ \int_0^{\infty} \frac{x^2}{2 \log(x + 2)} \exp\left(-\frac{x}{2}\right) dx \]
Toy Examples

```r
n <- 1000
x <- rexp(n, rate=.5)
y <- x^2
est <- mean(y)
est
mcse <- sd(y) / sqrt(n)
interval <- est + c(-1,1)*1.96*mcse
interval

y <- x^2/log(x+2)
est <- mean(y)
est
mcse <- sd(y) / sqrt(n)
interval <- est + c(-1,1)*1.96*mcse
interval
```
1. Monte Carlo integration
2. Markov chain Monte Carlo
3. Bayesian statistics
We Want to know:

\[ \lambda = \int_S h \, d\pi, \]

which is analytically intractable. Here \( \pi \) is a prob. measure and \( h \) is integrable.

**Ordinary Monte Carlo** is the method of using IID simulations \( X_1, \ldots, X_n \) from \( \pi \) to approximate expectations by sample averages

\[ \bar{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i). \]

**Markov chain Monte Carlo (MCMC)** replaces IID simulations with realizations \( X_1, \ldots, X_n \) of a Markov chain with unique stationary distribution \( \pi \).

By SLLN for Markov chains, under certain conditions,

\[ \bar{h}_n \xrightarrow{\text{as}} E_\pi h \text{ as } n \to \infty. \]
By SLLN for Markov chains, $\overline{h}_n \xrightarrow{as} E_\pi h$ as $n \to \infty$.

**How do we compute an associated standard error?**

An answer to this question requires

$$\sqrt{n}\left(\overline{h}_n - E_\pi h\right) \xrightarrow{d} N(0, \sigma_h^2)$$

and a consistent estimator of $\sigma_h^2$, say, $\hat{\sigma}_h^2$.

**How large should $n$ be?**

Asymptotic 95% CI for $E_\pi h$: $\overline{h}_n \pm 2\hat{\sigma}_h/\sqrt{n}$

**Problem:** $E_\pi h^2 < \infty$ does not guarantee a CLT.

If $\{X_n\}_{n=0}^\infty$ is geometrically ergodic then CLT holds for all $h$ s.t. $E_\pi h^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

**How do we construct a consistent estimator of $\sigma_h^2$?**
Consider a countable state space \( S = \{ s_0, s_1, s_2, \ldots \} \).

**Definition**

A sequence of \( S \) valued random variables \( \{ X_n \}_{n \geq 0} \) defined on a probability space \((\Omega, \mathcal{F}, \mu)\) is called a Markov chain with stationary transition probabilities \( P = ((p_{ij})) \), initial probability distribution \( \nu \), and state space \( S \) if

1. \( X_0 \sim \nu \), and
2. \( P(X_{n+1} = s_j | X_n = s_i, X_{n-1} = s_{i_{n-1}}, \ldots, X_0 = s_{i_0}) = P(X_{n+1} = s_j | X_n = s_i) = p_{ij} \) for all \( s_i, s_j, s_{i_0}, \ldots, s_{i_{n-1}} \) and \( n = 0, 1, \ldots \).
Markov chains

**Definition**

A distribution $\pi$ on $S$ is called stationary (invariant) distribution for $P$ if

$$\pi P = \pi,$$

that is,

$$\sum_{s_i \in S} \pi_i p_{ij} = \pi_j \text{ for all } s_j \in S.$$

Even if a Markov chain has stationary distribution $\pi$, it may still fail to converge to $\pi$.

**Example** Let $S = \{0, 1, 2\}$ with $\pi(i) = 1/3$ for all $i$. Let

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and $X_0 = 0$. 
Markov chains

**Definition**

A Markov chain with stationary transition probability matrix $P = ((p_{ij}))$, and state space $S$ is called irreducible if for all $s_i, s_j \in S$, $P(X_n = s_j \text{ for some } 1 \leq n < \infty \mid X_0 = s_i) > 0$.

For an irreducible Markov chain $\{X_n\}_{n \geq 0}$ with stationary probability distribution $\pi$, SLLN holds, that is, if $E_{\pi} |h| < \infty$, then $\bar{h}_n \xrightarrow{\text{as}} E_{\pi} h$ as $n \to \infty$. 
Let $\pi(x)$ be the target pdf.
Let $x_n$ be the current value of the Markov chain.
The Metropolis-Hastings algorithm performs the following.

1. Propose $y \sim q(\cdot|x_n)$.
2. Accept $X_{n+1} = y$ with probability

$$
\alpha(x_n, y) = \min\left\{ \frac{\pi(y)q(x_n|y)}{\pi(x_n)q(y|x_n)}, 1 \right\},
$$

otherwise, set $X_{n+1} = x_n$. 
Metropolis-Hastings algorithm

- Random walk proposal $q(y|x) = f(y - x)$
- Independence proposal $q(y|x) = f(y)$

**Random walk chains**
In the chain is currently at $x$, propose an increment $I$ according to a fixed density $f$. Accept or reject the candidate point $y = x + I$. Thus here $q(y|x) = f(y - x)$ for all $x, y$.
If $f$ is symmetric, that is, $f(-t) = f(t)$ for all $t$, the acceptance probability is

$$\alpha(x_n, y) = \min\left\{ \frac{\pi(y)}{\pi(x_n)}, 1 \right\}.$$

**Independence chains**
Here $q(y|x) = f(y)$ for all $x$.
The acceptance probability is

$$\alpha(x_n, y) = \min\left\{ \frac{\pi(y)f(x_n)}{\pi(x_n)f(y)}, 1 \right\}.$$
Example: Random walk chains

Let

\[ \pi(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) \quad \text{and} \quad f(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-y^2/[2\sigma^2]), \]

that is, the target density is \(N(0, 1)\) and the proposal density is \(N(0, \sigma^2)\) for some known \(\sigma^2\). So

\[ \alpha(x, y) = \min\left\{ \frac{\pi(y)}{\pi(x)}, 1 \right\} = \min\left\{ \exp\left[ -\frac{1}{2} (y^2 - x^2) \right], 1 \right\}. \]
Example

```r
set.seed(3)
library(mcmcse)
n_iterations = 10000
sigma=2.4
log_pi = function(y) {
dnorm(y,log=TRUE)
}
current = 0.5 # Initial value
samps = rep(NA,n_iterations)
for (i in 1:n_iterations) {
  proposed = rnorm(1, current, sigma)
  logr = log_pi(proposed)-log_pi(current)
  if (log(runif(1)) < logr) current = proposed
  samps[i] = current
}
length(unique(samps))/n_iterations # acceptance rate
```
Example

ts.plot(samps[1:1000])
mcse(samps)
x=seq(-3,3,0.01)
fx=sapply(x,function(x) dnorm(x))
plot(x,fx,type='l')
hist(samps[1:10000],prob=T,col='red',add=T)
plot(acf(samps))
Example: Independence chain

Let

\[\pi(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)\]

and

\[f(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{y^2}{2\sigma^2}\right),\]

that is, the target density is \(N(0, 1)\) and the proposal density is \(N(0, \sigma^2)\) for some known \(\sigma^2 > 1\). So

\[\alpha(x, y) = ?\]
Suppose $\pi(x_1, x_2)$ is a joint density. If sampling from the corresponding conditional densities $\pi_{X_1|X_2}$ and $\pi_{X_2|X_1}$ is straightforward, then we can use the Gibbs sampler to explore $\pi$.

Given the current state, $(x_1, n, x_2, n)$, the following two steps are used to move to the new state $(X_1, n+1, X_2, n+1)$.

1. Draw $X_{1, n+1} \sim \pi_{X_1|X_2} (\cdot | x_2, n)$
2. Draw $X_{2, n+1} \sim \pi_{X_2|X_1} (\cdot | x_1, n+1)$

**Example** Let $\pi$ be the bivariate normal density with mean vector $(0, 0)'$ and covariance matrix

\[
\begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix}.
\]

Conditional on $X_1, X_2 \sim N(\rho X_1, 1 - \rho^2)$.
Conditional on $X_2, X_1 \sim N(\rho X_2, 1 - \rho^2)$. 
Example

```r
set.seed(3)
library(mcmcse)
gibbs_bivariate_normal = function(samps_start, n_iterations, rho) {
samps = matrix(samps_start, nrow=n_iterations, ncol=2, byrow=TRUE)
v = sqrt(1-rho^2)
for (i in 2:n_iterations) {
samps[i,1] = rnorm(1, rho*samps[i-1,2], v)
samps[i,2] = rnorm(1, rho*samps[i,1], v)
}
return(samps)
}
samps = gibbs_bivariate_normal(c(-3,3), n_iterations<-1000, rho<-0.9)
```
Example

ts.plot(samps[,1])
ts.plot(samps[,2])
apply(samps, 2, mcse)
apply(samps, 2, quantile, probs=c(0.025,0.975))
cor(samps[,1],samps[,2])
Let $\pi$ be a density on $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$.

Iteration $n+1$ of the Gibbs sampler:

1. Draw $X_{1,n+1} \sim \pi_{X_1\mid\{X_j \neq 1\}}(\cdot, X_{2,n}, \ldots, X_{k,n})$

2. Draw $X_{2,n+1} \sim \pi_{X_2\mid\{X_j \neq 2\}}(X_{1,n+1}, \cdot, X_{3,n}, \ldots, X_{k,n})$

   \[\vdots\]

3. Draw $X_{k,n+1} \sim \pi_{X_k\mid\{X_j \neq k\}}(X_{1,n+1}, \ldots, X_{k-1,n+1}, \cdot)$
1. Monte Carlo integration
2. Markov chain Monte Carlo
3. Bayesian statistics
Bayesian statistics

Suppose $X_i \sim f(x_i|\theta), i = 1, \ldots, m$.

The likelihood function

$$\ell(\theta|x) = \prod_{i=1}^{m} f(x_i|\theta).$$

Before observing the data, the prior density $\pi(\theta)$ is assigned. The posterior density

$$\pi(\theta|x) \propto \ell(\theta|x)\pi(\theta),$$

where $\ell(\theta|x)$ is the likelihood function, and $\pi(\theta)$ is prior. In particular,

$$\pi(\theta|x) = \frac{\ell(\theta|x)\pi(\theta)}{m(x)},$$

where

$$m(x) = \int_{\Theta} \ell(\theta|x)\pi(\theta)d\theta.$$
Example: Normal linear regression

Let \((Y_1, Y_2, \ldots, Y_m)\) denote the vector of normal random variables, \(x_i\) be the \(p \times 1\) vector of known covariates associated with the \(i^{th}\) observation for \(i = 1, \ldots, m\). Let \(\beta \in \mathbb{R}^p\) be the unknown vector of regression coefficients. The multiple linear regression model is

\[
Y_i \overset{\text{ind}}{\sim} N(x_i^T \beta, \sigma^2), \quad i = 1, \ldots, m.
\]

The ordinary least squared estimator

\[
\hat{\beta}_{OLS} = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) = (X^T X)^{-1} X^T Y
\]

For Bayesian analysis, we need to specify priors on \(\beta\) and \(\sigma^2\). For \(\beta\), there are several choices:

- Normal prior
- Improper uniform prior (Jeffreys prior)
- Laplace prior and many others
Example: Normal linear regression

For $\sigma^2$, there are several choices:

- Inverse gamma prior
- Improper power prior (Special case: Jeffreys prior)

The multiple linear regression model is

$$Y \sim N_m(X\beta, \sigma^2I_m).$$

The likelihood function

$$\ell(\beta, \sigma^2|y) \propto (\sigma^2)^{-m/2} \exp \left\{ -\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta) \right\}$$

Priors on $\beta$ and $\sigma^2$:

$$\beta \sim N_p(\mu, C), \quad \sigma^2 \sim IG(\alpha, \gamma)$$

$$\pi(\sigma^2) = \frac{\gamma^\alpha}{\Gamma(\alpha)}(\sigma^2)^{(-\alpha-1)} \exp \left( -\frac{\gamma}{\sigma^2} \right)$$
Example: Normal linear regression

The posterior density

\[ \pi(\beta, \sigma^2 | y) \propto \ell(\beta, \sigma^2 | y) \times \pi(\beta, \sigma^2) \]

\[ \propto (\sigma^2)^{-m/2} \exp \left\{ - \frac{1}{2\sigma^2} (y - X\beta)^T(y - X\beta) \right\} \]

\[ \times \exp \left\{ - \frac{1}{2} (\beta - \mu)^T C^{-1} (\beta - \mu) \right\} (\sigma^2)^{(-\alpha-1)} \exp \left( - \frac{\gamma}{\sigma^2} \right) \]

The posterior density with Jeffreys prior \( \pi_J(\beta, \sigma^2) \propto \frac{1}{\sigma^2} \) is

\[ \pi_J(\beta, \sigma^2 | y) \propto \ell(\beta, \sigma^2 | y) \times \pi_J(\beta, \sigma^2) \]

\[ \propto (\sigma^2)^{-m/2-1} \exp \left\{ - \frac{1}{2\sigma^2} (y - X\beta)^T(y - X\beta) \right\} \]

The posterior densities are intractable, but a Gibbs sampler can be formed as the conditional densities correspond to standard distributions.
Example: Normal linear regression

Indeed,

$$\beta | \sigma^2, y \sim N_p(\mu', C')$$,

where

$$C' = \left( \frac{X^TX}{\sigma^2} + C^{-1} \right)^{-1},$$

and

$$\mu' = C' \left( \frac{X^Ty}{\sigma^2} + C^{-1} \mu \right).$$

Also,

$$\sigma^2 | \beta, y \sim IG(\alpha', \gamma')$$,

where

$$\alpha' = \frac{m}{2} + \alpha,$$

and

$$\gamma' = \left[ (y - X\beta)^T (y - X\beta) + 2\gamma \right]/2.$$
If Jeffreys prior is used,

\[ \beta | \sigma^2, \mathbf{y} \sim N_p((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}) , \]

and

\[ \sigma^2 | \beta, \mathbf{y} \sim IG(m/2, (\mathbf{y} - \mathbf{X} \beta)^T (\mathbf{y} - \mathbf{X} \beta)/2) . \]
library(MCMCpack)
library(mcmcse)
data<-read.csv("stock_treasury.csv")
# Risk Free Rate is in percentage and annualised.
# So the following conversion is required.
Rf<-data$UST_Yr_1/(100*250)
plot(ts(Rf),ylab="US Treasury 1 Year Yield")
n<-nrow(data)
## Compute log-return
ln_rt_snp500<-diff(log(data$SnP500))-Rf[2:n]
ln_rt_ibm<-diff(log(data$IBM_AdjClose))-Rf[2:n]
ln_rt_apple<-diff(log(data$Apple_AdjClose))-Rf[2:n]
ln_rt_msft<-diff(log(data$MSFT_AdjClose))-Rf[2:n]
ln_rt_intel<-diff(log(data$Intel_AdjClose))-Rf[2:n]
```r
y = ln_rt_ibm
n_obs = length(y)
X = cbind(rep(1, n_obs), ln_rt_snp500) # include # an intercept
XtX = t(X) %*% X
n_params = 2
n_obsby2 = n_obs / 2

beta_hat = solve(XtX, t(X) %*% y) # compute # this beforehand
XtX_i = solve(XtX)
beta = c(0, 0) # starting value
# beta = beta_hat
n_iterations = 5000 # number of MCMC iterations
```
beta_out = matrix(data = NA, nrow = n_iterations, ncol = n_params)
sigma_out = matrix(data = NA, nrow = n_iterations, ncol = 1)
for (i in 1:n_iterations) {
    ymxbeta = (y - X %*% beta)
    sigma2 = rinvgamma(1, n_obsby2, t(ymxbeta) %*% ymxbeta * .5)  # draw from sigma2
given beta
    beta = mvrnorm(n = 1, beta_hat, sigma2 * XtXi)  # draw from beta
given sigma2
    sigma_out[i,] = sigma2
    beta_out[i,] = beta
}
ts.plot(sigma_out)
ts.plot(beta_out[,1])
apply(beta_out, 2, mcse)
mcse(sigma_out)
apply(beta_out, 2, quantile, probs=c(0.025,0.975))
quantile(sigma_out, probs=c(0.025,0.975))