Monte Carlo

Galin L. Jones

School of Statistics

University of Minnesota

Draft: January 7, 2023

Contents

1 Introduction

The goal of this chapter is to introduce the main themes of the Monte Carlo method. The "Monte Carlo method" means using a computer to simulate data in order estimate fixed unknown quantities (i.e. features or parameters) from a specified distribution. These quantities can be expectations, probabilities, density functions, quantiles, and so on. This likely sounds familiar since it is basically a description of much of classical statistics. The main difference is that it is based on data produced by a computer, rather than collected from an external experiment. There are two fundamental issues in implementing the Monte Carlo method: (1) designing algorithms that produce useful observations and (2) using these observations to estimate features of the given distribution.

The simulated data can be a random sample as in classical Monte Carlo[1] or a realization of a Markov chain as in Markov chain Monte Carlo (MCMC). An independent and identically distributed (iid) sequence is a trivial Markov chain so it is not too surprising that Monte Carlo and MCMC have much in common. However, simulating a Markov chain introduces complications beyond those encountered when iid samples are available. Thus this chapter begins at the beginning and focuses on the simpler setting where simulation of an iid sample is possible.

1.1 Motivating Examples

Suppose distribution F has support X and there is either an associated probability density function (pdf) or probability mass function (pmf), either of which will be denoted by f and referred to as a *probability function* (pf). If, at various points, something more general is needed, it will be carefully stated. The goal is to estimate $\theta \in \mathbb{R}^p$, $p \geq 1$ which is a vector of fixed, unknown features of F. For example, components of θ often include the following.

[1] "Classical Monte Carlo" is a cumbersome phrase so "Monte Carlo" will be used instead. Hopefully, this will not cause confusion as "Monte Carlo" is often used as shorthand for "Monte Carlo method".

1. Expectations. Let $h : \mathsf{X} \to \mathbb{R}$ and

$$\mu_h = E_F[h(X)] = \int_{\mathsf{X}} h(x)F(dx) \; .$$

This notation is used throughout and allows us to avoid having separate formulas for the continuous case where it denotes $\mu_h = \int_{\mathsf{X}} h(x)f(x)dx$ and the discrete case where it denotes $\mu_h = \sum_{x\in\mathsf{X}} h(x)f(x)$.

2. Quantiles. Set $V = h(X)$ and let F_V denote the distribution of V. If $0 < q < 1$, the qth quantile is

$$\xi_q = F_V^{-1}(q) = \inf\{v \; : \; F_V(v) \geq q\} \; .$$

Monte Carlo can be used for more than estimating expectations and quantiles, but these applications are common. A few examples follow which are intended to illustrate more concretely the types of settings where Monte Carlo might be useful. Although it might not be obvious at first glance, this includes settings that have no inherent probabilistic component.

Example 1.1. Consider

$$\int_0^1 \frac{1}{(x+1)^{2.3}[\log(x+3)]^2}\, dx \, ,$$

which does not appear easy to solve, but can be expressed as an expectation. If f is a pdf on $(0,1)$, then

$$\begin{aligned}\int_0^1 \frac{1}{(x+1)^{2.3}[\log(x+3)]^2}\, dx \; &= \int_0^1 \frac{1}{(x+1)^{2.3}[\log(x+3)]^2}\, \frac{f(x)}{f(x)}\, dx \\ &= E_F\left[\frac{1}{f(x)(x+1)^{2.3}[\log(x+3)]^2}\right] \; .\end{aligned}$$

The example also makes the point that any expectation can be converted to an expectation with respect to a different distribution, say G having pf g and support containing X:

$$\mu_h = \int_{\mathsf{X}} h(x)f(x)\, dx = \int_{\mathsf{X}} \frac{h(x)f(x)}{g(x)} g(x)\, dx = \mu_{hf/g} \; .$$

Example 1.2 (Bayesian Logistic Regression). Let X be a known $n \times p$ matrix with rows x_i and let β be a p-vector of parameters. Set

$$h(x_i) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$$

and assume $Y_i \sim \text{Bernoulli}(h(x_i))$, independently for $i = 1, \ldots, m$. Let y denote all of the observed data and assume $\beta \sim \text{N}_p(0, I_p)$ so that the posterior is characterized by

$$q(\beta|y) \propto \left[\prod_{i=1}^{n} \frac{e^{-y_i x_i^T \beta}}{1 + e^{-x_i^T \beta}}\right] e^{-\frac{1}{2}\beta^T \beta} .$$

The normalizing constant or marginal density is

$$m(y) = \int \left[\prod_{i=1}^{n} \frac{e^{-y_i x_i \beta}}{1 + e^{-x_i \beta}}\right] e^{-\frac{1}{2}\beta^T \beta} \, d\beta ,$$

which is analytically intractable. A typical goal is to calculate the posterior mean of β:

$$\mu_\beta = E_q[\beta|y] = \int_{\mathbb{R}^p} \beta \, q(\beta|y) \, d\beta .$$

Posterior inference also often requires expectations of other functions $h(\beta)$, such as second moments, so the general goal is calculation of

$$\mu_h = E_q[h(\beta)|y] = \int_{\mathbb{R}^p} h(\beta) \, q(\beta|y) \, d\beta .$$

Posterior credible intervals can also be based on quantiles. For example, suppose the analysis requires a .95 credible interval $(\xi_{.025}, \xi_{.975})$ for the first component of β, that is β_1. If $q(\beta_1|y)$ is the marginal posterior density of β_1, finding $(\xi_{.025}, \xi_{.975})$ requires solving

$$\int_{-\infty}^{\xi_{.025}} q(\beta_1|y) d\beta_1 = .025 \quad \text{and} \quad \int_{-\infty}^{\xi_{.975}} q(\beta_1|y) d\beta_1 = .975 .$$

Example 1.3 (Bayesian Linear Model). Suppose that, for $i = 1, \ldots, k$ and $j = 1, \ldots, m_i$,

$$\begin{aligned} Y_{ij}|\tau_i, \lambda_e &\sim \text{N}(\tau_i, \lambda_e^{-1}) \\ \tau_i|\mu, \lambda_e &\sim \text{N}(\mu, \lambda_t^{-1}) \\ \mu &\sim \text{N}(m_0, s_0^{-1}) \\ \lambda_e &\sim \text{Gamma}(a_1, b_1) \\ \lambda_t &\sim \text{Gamma}(a_2, b_2) . \end{aligned}$$

Letting y denote all of the observed data and τ denote all of the τ_i, the posterior distribution is characterized by

$$q(\tau, \mu, \lambda_e, \lambda_t|y) \propto f(y|\tau, \lambda_e) f(\tau|\mu, \lambda_t) f(\mu) f(\lambda_e) f(\lambda_t) .$$

The normalizing constant or marginal density of y is

$$m(y) = \int f(y|\tau, \lambda_e) f(\tau|\mu, \lambda_t) f(\mu) f(\lambda_e) f(\lambda_t)\, d\tau\, d\mu\, d\lambda_e\, d\lambda_t \ ,$$

and is analytically intractable. Interest often centers on posterior expectations and quantiles. For example,

$$\mu_\tau = E_q[\tau|y] = \int \tau q(\tau|y)\, d\tau = \int \tau\, q(\tau, \mu, \lambda_e, \lambda_t|y)\, d\mu\, d\lambda_e\, d\lambda_t\, d\tau \ .$$

Because $m(y)$ is unavailable to us, analytical evaluation of posterior expectations or quantiles is unavailable.

While Monte Carlo methods have had a profound impact on the implementation of Bayesian inference, they are also important to the implementation of frequentist inference. Here is a simple example to consider.

Example 1.4 (Logit-Normal Generalized Linear Mixed Model). Let

$$p(\beta, u) = \frac{e^{\beta+u}}{1+e^{\beta+u}} \ .$$

Suppose $Y|u, \beta \sim \text{Bernoulli}(p(\beta, u))$ and $U|\lambda \sim \text{N}(0, \lambda)$.

Then the likelihood is

$$L(\beta, \lambda) = \int_{-\infty}^{\infty} f(y|u, \beta) f(u|\lambda)\, du = \frac{1}{\sqrt{2\pi\lambda}} \int_{-\infty}^{\infty} \frac{e^{\beta+u}}{(1+e^{\beta+u})^2} e^{-\frac{1}{2\lambda}u^2}\, du \ .$$

Now $L(\beta, \lambda)$ can be written as an expectation: let G be a distribution having density g on $\mathbb{R}$ so that

$$\begin{aligned} L(\beta, \lambda) &= \int_{-\infty}^{\infty} \frac{f(y|u, \beta) f(u|\lambda)}{g(u)} g(u)\, du \\ &= E_G \left[\frac{f(y|u, \beta) f(u|\lambda)}{g(u)} \right] \ . \end{aligned}$$

Since the likelihood can be expressed as an expectation Monte Carlo can be used to approximate the function [5].

2 Monte Carlo

Settings where Monte Carlo is appropriate often begin with a given distribution F and the goal is to estimate θ, a vector of features of F. In Monte Carlo experiments, observations $X_1, \ldots, X_m$ are simulated and are used to construct an estimator θ_m in such a way that $\theta_m \approx \theta$ for large m. Calculation of the estimator θ_m alone is an incomplete solution to the problem. No matter how large m is, there will be an unknown *Monte Carlo error*, $\theta_m - \theta$ and hence θ_m will be more valuable if a measure of the Monte Carlo error is included.

Monte Carlo sample size is used to mean the size of the simulation effort and it will be denoted m while n will be used to denote the sample size associated with the original statistical setting. A couple of examples may help illuminate the difference.

Example 2.1. Suppose Y_1, Y_2, Y_3 are iid Poisson(λ) and $\lambda \sim$ Gamma$(2, 3)$. Then the posterior is $\lambda | y_1, y_2, y_3 \sim$ Gamma$(3\bar{y} + 2, 6)$, where $\bar{y}$ is the sample mean of the three observations. There is no simple closed form for the median of the posterior, but simulation can be used to estimate it. Simulate a large number, 10000 say, observations from the posterior distribution. The sample median is an estimate of the true posterior median. Here $n = 3$ and $m = 10000$.

Example 2.2. Suppose $Y_1 \ldots, Y_{100}$ and $x_i = i/5$ for $i = 1, \ldots, 50$. Lets use linear regression to model the observations as $Y_i = \beta x_i + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$. The goal is to test the hypotheses $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ with a type 1 error rate of $\alpha = .05$. Then β is estimated using least squares and a standard t-test used for testing the hypotheses.

How robust is this procedure to departures from the assumption $\varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$? If, in fact, $\varepsilon_i \overset{iid}{\sim}$ Cauchy$(0, \sigma)$, but the t-test is used, what happens to the type 1 error rate? One way to find out is via simulation. Do the following 1000 times: fix $\beta = 0$, simulate $\varepsilon_i \overset{iid}{\sim}$ Cauchy$(0, \sigma)$ for $i = 1, \ldots, 50$ and conduct the t-test. The proportion out of 1000 that reject is an estimate of the type 1 error. Here $n = 50$ and $m = 1000$.

While the theory of Monte Carlo is large sample frequentist theory, the asymptotics are as the Monte Carlo sample size m increases. Typically, the observed data sample size n is treated as fixed and known, but there are situations where it makes more sense to let n increase to infinity while

fixing m or let m and n increase simultaneously. These last two settings will not be addressed further.

2.1 Producing a random sample

Monte Carlo methods are based on the ability to have the computer generate independent Uniform$(0, 1)$ observations. Of course, the observations are not random since they are produced by deterministic methods. However, good pseudorandom number generators produce sequences that effectively mimic independent Uniform$(0, 1)$ observations and hence are known as pseudorandom sequences. It is not clear that truly random sequences would be desirable since repeatability would be problematic, making debugging much more challenging. For the most part, this issue will not be considered further since the distinction between pseudorandom and random often will not be useful here.

This section considers some basic ways of obtaining a random sample from (non-uniform) F, including inversion, ratio of uniforms, the accept-reject algorithm, and linchpin variable sampling. This presentation is not in any way intended to be comprehensive.

2.1.1 Inversion

Define $F^{-1} : (0, 1) \to \mathbb{R}$ by

$$F^{-1}(y) = \inf\{x \,:\, F(x) \geq y\} \,.$$

The *quantile function theorem* is the foundation for simulating random variates from an arbitrary distribution given the ability to simulate from a Uniform distribution.

Theorem 2.1. *If $U \sim$ Uniform$(0, 1)$ and $X = F^{-1}(U)$, then $X \sim F$.*

Example 2.3. Suppose $\beta > 0$. Inversion can be used to construct a draw from an Exp(β) distribution. Then $F^{-1}(y) = -\beta \log(1 - y)$ so if $U \sim$ Uniform$(0, 1)$, then $F^{-1}(U) = -\beta \log(1 - U) \sim$ Exp(β).

When F^{-1} is explicitly available and calculation of it is fast, inversion is practical, but these limitations often prevent its use.

2.1.2 Accept-Reject

The accept-reject algorithm uses draws from a convenient distribution G, having pf g, say, and converts them into draws from F. Suppose the support of G contains X and that

$$M = \sup_{x \in \mathsf{X}} \frac{f(x)}{g(x)} < \infty \, .$$

Algorithm 1: Accept-Reject

1: Draw $Y \sim G$

2: Draw $U \sim \text{Uniform}(0, 1)$

3: If

$$u \leq \frac{f(y)}{Mg(y)}$$

accept y as a draw form F; otherwise return to step 1.

The accept-reject algorithm is a stochastic algorithm in that the accept-reject step is random. The probability of acceptance on a given step is

$$\begin{aligned} P(U \leq f(y)/Mg(y)) &= E\left[P(U \leq f(y)/Mg(y))|Y\right] \\ &= E\left[\frac{f(Y)}{Mg(Y)}\right] \\ &= \frac{1}{M} \, . \end{aligned}$$

Theorem 2.2. *Algorithm 1 produces $X \sim F$.*

Proof. Notice that

$$\begin{aligned} P(X \leq x) &= P(Y \leq x | U \leq f(y)/Mg(y)) \\ &= \frac{P(Y \leq x, U \leq f(y)/Mg(y))}{P(U \leq f(y)/Mg(y))} \, . \end{aligned}$$

Now consider numerator and denominator separately:

$$\begin{aligned} P(Y \le x, U \le f(y)/Mg(y)) &= E\left[P(Y \le x, U \le f(y)/Mg(y))|Y\right] \\ &= E\left[I(Y \le x)\frac{f(Y)}{Mg(Y)}\right] \\ &= \frac{1}{M}F(x) \end{aligned}$$

and

$$\begin{aligned} P(U \le f(y)/Mg(y)) &= E\left[P(U \le f(y)/Mg(y))|Y\right] \\ &= E\left[\frac{f(Y)}{Mg(Y)}\right] \\ &= \frac{1}{M}\,. \end{aligned}$$

Putting these together yields $P(X \le x) = F(x)$, which proves the claim. □

Clearly, the choice of proposal density g is crucial to the success of the algorithm. Note that M is the expected number of proposals required before a draw is obtained. To make M smaller and the algorithm more efficient requires a proposal g that mimics f in its tails.

Also notice that accept-reject can be used when the normalizing constant for f is unknown. Let $f(x) = c_1 h(x)$, $g(x) = c_2 l(x)$, and

$$\sup_{x \in \mathsf{X}} \frac{h(x)}{l(x)} = K\,.$$

If $U \sim \text{Uniform}(0,1)$ and $Y \sim G$, then

$$U \le \frac{h(y)}{Kl(y)} = \frac{f(y)}{\frac{c_1}{c_2}Kg(y)} = \frac{f(y)}{Mg(y)}$$

and hence yields an equivalent accept-reject algorithm.

2.1.3 Ratio of Uniforms

Let h be a positive integrable function on (a,b) where a and b are not necessarily finite. Define

$$\mathsf{A}_h = \{(x,y) \ : \ 0 \le x \le h^{1/2}(y/x), \quad a < y/x < b\}\,. \tag{1}$$

Example 2.4. Suppose $X \sim \text{Cauchy}(0,1)$, then $h(x) = [1+x^2]^{-1}$ and

$$\mathsf{A}_h = \{(x,y) \ : \ x > 0 \text{ and } x^2 + y^2 \leq 1\}\,.$$

Theorem 2.3. *If (U,V) is uniformly distributed on A_h, then $X = V/U$ has pdf $f(x) \propto h(x)$.*

Theorem 2.3 suggests a simple algorithm for generating from a non-uniform distribution having pf f.

Algorithm 2: Ratio of Uniforms

1: Draw (U,V) uniformly on A_h

2: Set $X = V/U$

Algorithm 2 can be efficient, but generating uniformly on A_h can be challenging. Fortunately, there is a special case of the accept-reject algorithm for avoiding this bottleneck. Set $a = 0$,

$$b = \sup_x \sqrt{h(x)} < \infty\,, \qquad c = \sup_x x\sqrt{h(x)} < \infty\,, \text{ and } \ d = \inf_x x\sqrt{h(x)} > -\infty\,.$$

Then $\mathsf{A}_h \subseteq A = [a,b] \times [c,d]$ which suggests the following ratio of uniforms algorithm using accept-reject.

Algorithm 3: Ratio of Uniforms using Accept-Reject

1: Draw $U \sim \text{Uniform}(a,b)$

2: Draw $V \sim \text{Uniform}(c,d)$

3: If $U \leq h^{1/2}(V/U)$, set $X = V/U$; otherwise, repeat step 1.

Now it is easy to see that the probability of acceptance is

$$\frac{\text{area}(\mathsf{A}_h)}{\text{area(A)}} = \frac{\text{area}(\mathsf{A}_h)}{b(d-c)} \tag{2}$$

and hence that the mean number of proposals until success is finite.

Example 2.5. This is a continuation of example 2.4. Notice that

$$A = [0,1] \times [-1,1]$$

and the acceptance probability (2) is $\pi/4$.

2.1.4 Linchpin Variables

The accept-reject and ratio of uniforms algorithms can be difficult to apply in multivariate settings, that is, when $d > 1$. However, surprisingly often a complicated multivariate simulation setting can be converted to simulating from a simpler distribution. Suppose f can be expressed as a product of a conditional pf $f_{X|Y}$ and a marginal pf f_Y so that

$$f(x, y) = f_{X|Y}(x|y) f_Y(y)$$

If sampling from $f_{X|Y}$ is straightforward, then say Y is a *linchpin variable*.

Algorithm 4: Linchpin Variable Algorithm

1: Draw $Y \sim F_Y$

2: Draw $X \sim F_{X|Y}(\cdot \mid Y)$

It should be obvious that the linchpin variable algorithm will be useful only when sampling from the marginal F_Y is easier than sampling from the joint F.

Example 2.6. For $i = 1, \ldots, n$ assume $t_i > 0$ is known and let $Y_i|\lambda_i \overset{ind}{\sim} \text{Poisson}(t_i\lambda_i)$. Assume priors $\lambda_i \overset{ind}{\sim} \text{Gamma}(a, \beta)$ and $\beta \sim \text{Gamma}(c, d)$ with a, c, and d known positive constants. The posterior is characterized by

$$f(\beta, \lambda|y) \propto \left(\prod_{i=1}^{n} \lambda_i^{a+y_i-1} e^{-(\beta+t_i)\lambda_i} \right) \beta^{an+c-1} e^{-\beta d} \, .$$

By inspection the conditionals $\lambda_i|\beta \overset{ind}{\sim} \text{Gamma}(a + y_i, \beta + t_i)$ and the marginal pf for β is characterized by

$$f(\beta|y) \propto \beta^{an+c-1} e^{-\beta d} \left(\beta + t_i\right)^{-(a+y_i)} \, .$$

Thus β is a linchpin variable. See exercise 2.12 for an accept-reject algorithm to sample from the posterior marginal.

Example 2.7. For $i = 1, \ldots, K$ suppose that for known $a, b, c > 0$,

$$\begin{aligned} Y_i|\theta_i &\sim \text{N}(\theta_i, a) \qquad & \theta_i|\mu, \lambda &\sim \text{N}(\mu, \lambda) \\ \lambda &\sim \text{IG}(b, c) \qquad & f(\mu) &\propto 1 \, . \end{aligned}$$

Then the hierarchy yields a proper posterior $f(\theta, \mu, \lambda|y)$ with $\theta = (\theta_1, \ldots, \theta_K)^T$ and $y = (y_1, \ldots, y_K)^T$. Consider the factorization [see 6]

$$f(\theta, \mu, \lambda|y) = f(\theta|\mu, \lambda, y) f(\mu|\lambda, y) f(\lambda|y) \; .$$

Then $f(\theta|\mu, \lambda, y)$ is the product of univariate normal densities $\theta_i|\mu, \lambda, y \sim \mathrm{N}((\lambda y_i + a\mu)/(a + \lambda), a\lambda/(a + \lambda))$. Now $f(\mu|\lambda, y)$ is also a normal density $\mathrm{N}(\bar{y}, (a + \lambda)/K)$. Finally,

$$f(\lambda|y) \propto \frac{1}{\lambda^{b+1}(a + \lambda)^{(K-1)/2}} \exp\left\{ -\frac{c}{\lambda} - \frac{1}{2(a + \lambda)} \sum_{i=1}^{K} (y_i - \bar{y})^2 \right\} \; .$$

Thus λ is a linchpin variable. See exercise 2.14 for an accept-reject algorithm to sample from the posterior marginal.

Linchpin samplers will typically not be useful when the dimension of the linchpin variable is too large. The following example illustrates this.

Example 2.8. Consider a version of the so-called Bayesian lasso. Let X be a known $m \times p$ design matrix and assume $\lambda > 0$ is known. Assume a, $b > 0$ are known and

$$\begin{aligned} Y|\beta, \gamma &\sim \mathrm{N}_m(X\beta, \gamma^{-1} I_m) \\ \nu(\beta|\gamma) &= \left(\frac{\lambda\gamma}{4}\right)^p \exp\left\{ -\frac{\lambda\gamma}{2} \|\beta\|_1 \right\} \\ \gamma &\sim \mathrm{Gamma}(a, b) \; . \end{aligned}$$

This hierarchy gives rise to a posterior density which has conditional

$$\gamma|\beta, y \sim \mathrm{Gamma}\left(p + a + \frac{n}{2}, \; b + \frac{\lambda\|\beta\|_1 + \|y - X\beta\|_2^2}{2} \right)$$

and marginal

$$f(\beta|y) \propto \left(1 + \frac{\lambda\|\beta\|_1 + \|y - X\beta\|_2^2}{2b} \right)^{-(a+p+n/2)} \; .$$

In principle, one can construct an accept-reject sampler for sampling form the marginal of $\beta|y$ when X is full rank. However, the method is so inefficient as to be useless. Moreover, in situations where the lasso may be useful X is often not of full rank or p is large so that this linchpin variable sampler is not useful.

2.2 Estimation Theory for Monte Carlo

The estimation theory of Monte Carlo largely coincides with classical large-sample frequentist statistics. As such this will not be a comprehensive review, instead it is focused on a few key ideas.

2.2.1 Monte Carlo estimation

Suppose there is a Monte Carlo sample $X_1, \ldots, X_m \overset{iid}{\sim} F$ and we want to evaluate an expectation with respect to F

$$\mu_h := E_F[h(X)] = \int_{\mathsf{X}} h(x) F(dx) \; .$$

If $E_F|h(X)| < \infty$, then the strong law of large numbers (SLLN) obtains so that, with probability 1, as $m \to \infty$,

$$\mu_m := \frac{1}{m} \sum_{i=0}^{m-1} h(X_i) \to \mu_h \; . \tag{3}$$

Thus the sample mean is an asymptotically valid estimator of μ_h.

Estimation of quantiles follows much the same pattern as estimation of expectations. Set $V = h(x)$ and let F_V denote the distribution of V. If $0 < q < 1$, the qth quantile is

$$\xi_q = F_V^{-1}(q) = \inf\{v \, : \, F_V(v) \geq q\} \; .$$

If F_V is absolutely continuous and has continuous density function f_V satisfying $0 < f_V(\xi_q) < \infty$, then ξ_q is the unique solution of $F_V(y-) \leq q \leq F_V(y)$.

If $X_1, \ldots, X_m$ are iid F, set $Y_i = h(X_i)$ for $i = 1, \ldots, m$. Let $Y_{m(j)}$ be the jth order statistic of Monte Carlo sample. Then an estimator of ξ_q is

$$\xi_{m,q} = Y_{m(j)} \quad \text{where} \quad j - 1 < mq \leq j \tag{4}$$

and, with probability 1,

$$\xi_{n,q} \to \xi_q \quad \text{as} \quad m \to \infty \, . \tag{5}$$

Exercises

Exercise 2.1. Let $R = \{(x_1, x_2, x_3) \ : \ 0 \leq x_1 \leq x_2, \ 0 \leq x_1 \leq \sqrt{x_3}, \text{ and } x_1^2 + x_2^2 + x_3^2 \leq 1\}$.

1. Show that estimating the volume of R fits the framework of Section 1.1.
2. Given the ability to simulate independent copies of $U \sim \text{Uniform}(0, 1)$ devise an algorithm for estimating the volume of R.

Exercise 2.2. Prove Theorem 2.1.

Exercise 2.3. Given $U_1, \ldots, U_m \overset{iid}{\sim} \text{Uniform}(0, 1)$ how how to produce $X \sim \text{Gamma}(m, \beta)$ for any $\beta > 0$.

Exercise 2.4. Suppose F is continuous and that F^{-1} is not available in explicit form. Show how to use a root finding algorithm (such as bisection or Newton-Raphson) to use inversion to produce $X \sim F$.

Exercise 2.5. Let F have support $\{1, 2, \ldots, K\}$. Use inversion to construct an algorithm for simulating $X \sim F$ based on the ability to simulate $U \sim \text{Uniform}(0, 1)$.

Exercise 2.6. Suppose inversion can be used to simulate $X \sim F$ and let $F_{[a,b]}$ denote the F restricted to the interval $[a, b]$ where $[a, b]$ is contained in the support of F. Show that inversion can also be used to simulate $X \sim F_{[a,b]}$.

Exercise 2.7. Show that if $U_1, U_2 \overset{iid}{\sim} \text{Uniform}(0, 1)$ and

$$X_1 = \cos(2\pi U_1)\sqrt{-2\log U_2} \qquad X_2 = \sin(2\pi U_1)\sqrt{-2\log U_2}\,,$$

then $X_1, X_2 \overset{iid}{\sim} \text{N}(0, 1)$.

Exercise 2.8. Prove Theorem 2.3.

Exercise 2.9. Suppose $A \subseteq \mathsf{X}$ and that F_A is the distribution F constrained to have support A. Notice that if f is the pf of F and $c = P(X \in A)$, then the pf of F_A is

$$f_A(x) = c^{-1} f(x) I_A(x)\ .$$

Devise a simple algorithm for simulating from F_A given the ability to simulate from F. Establish that the algorithm is valid.

Exercise 2.10. Implement the algorithm in the previous exercise to sample from a Beta(2,8) distribution constrained to $A = [.25, .45]$. Compare the theoretical and observed acceptance rates. Estimate the mean of the constrained distribution and report the Monte Carlo standard error.

Exercise 2.11. Let

$$f_{X_1,X_2}(x_1, x_2) \propto \exp\left\{-5(x_2 - x_1^2)^2 - \frac{1}{20}(x_1 - 1)^2\right\} .$$

Show that $X_1 \sim \mathrm{N}(1, 10)$ and $X_2|x_1 \sim \mathrm{N}(x_1^2, 1/10)$. Use a linchpin variable approach to sample from the joint distribution and plot the results.

Exercise 2.12. Recall Example 2.6.

1. Verify the expressions given for the posterior conditional of $\lambda_i|\beta, y$ and the posterior marginal $\beta|y$.

2. If the observed data are $0, 0, 1, 0, 0, 2, 1, 1, 0, 0$, and $t_i = 1$ for all i, implement an accept-reject sampler with a Gamma$(an+c, d)$ proposal distribution to sample from the $\beta|y$ marginal. How does the acceptance rate change as d changes?

3. Implement a linchpin variable sampler to estimate the posterior mean of $\beta, \lambda_1, \ldots, \lambda_{10}$. What is the Monte Carlo error of estimation?

Exercise 2.13. Suppose $Y_1, \ldots, Y_n \sim \mathrm{N}(\mu, \theta)$ and suppose the prior on (μ, θ) is proportional to $1/\sqrt{\theta}$.

1. Show that the posterior is proper if $n \geq 3$.

2. Show that $\mu|\theta, y \sim \mathrm{N}(\bar{y}, \theta/n)$ and $\theta|y \sim \mathrm{IG}(n/2, ns^2/2)$.

3. Suppose $n = 5$, $s^2 = 2$ and $\bar{y} = 1$. Implement a linchpin variable algorithm to estimate the posterior mean and 85% credible region for μ.

Exercise 2.14. Recall Example 2.7.

0.9981504	-0.5370935	0.4000770	-0.4832548	-1.1237313
0.9712100	2.1511597	1.0207962	-0.9021560	-1.6078151
-2.2382052	0.5014717	1.3792220	0.3905185	-0.8079672
0.5799192	2.4626378	-1.0220088	1.5830721	-0.7893418

1. Verify the expressions given for $f(\theta|\mu, \lambda, y)$, $f(\mu|\lambda, y)$, and $f(\lambda|y)$.

2. Consider the data in the table 2. Implement an accept-reject sampler with an $\text{IG}(b, c)$ proposal distribution to sample from the $\lambda|y$ marginal.

Appendix

More on pseudorandom number generation: Devroye [3] and Fishman [4]

More on the quantile function theorem: Angus [1].

More on accept-reject: Caffo et al. [2] and Martino et al. [7] and squeeze principle and ARS

References

[1] Angus, J. E. (1994). The probability integral transform and related results. *SIAM Review*, 36:652–654.

[2] Caffo, B. S., Booth, J. G., and Davison, A. C. (2002). Empirical sup rejection sampling. *Biometrika*, 89:745–754.

[3] Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer-Verlag Inc.

[4] Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York.

[5] Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society,* Series B, 56:261–274.

[6] Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547.

[7] Martino, L., Luengo, D., and Míguez, J. (2012). On the generalized ratio of uniforms as a combination of transformed rejection and extended inverse of density sampling. *arXiv:1205.0482*.