STAT 7650 - Computational Statistics Lecture Slides

Simulating Random Variables

Elvan Ceyhan

Updated: March, 2025

AU

- Based on parts of: Chapter 6 in Givens & Hoeting (Computational Statistics), Chapter 22 of Lange (Numerical Analysis for Statisticians), and Chapter 2 in Robert & Casella (Monte Carlo Statistical Methods).

Outline

Motivation

- Simulation is a very powerful tool for statisticians.
- It allows us to investigate the performance of statistical methods before delving deep into difficult theoretical work.
- At a more practical level, integrals themselves are important for statisticians:
    - p-values and expectations are nothing but integrals;
    - Bayesians need to evaluate integrals to produce posterior probabilities, point estimates, and model selection criteria.
- Therefore, there is a need to understand simulation techniques and how they can be used for integral approximations.

Basic Monte Carlo

Suppose we have a function $h(x)$ and we'd like to compute the expected value

$$\mathbf{E}[h(X)] = \int h(x)f(x)\,dx$$

where $f(x)$ is a density. There is no guarantee that the techniques we learn in calculus are sufficient to evaluate this integral analytically.

Thankfully, the law of large numbers (LLN) is here to help: If $X_1, X_2, \ldots, X_n$ are i.i.d samples from $f(x)$, then

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to \int h(x)f(x)\,dx \quad \text{with prob 1 as } n \to \infty.$$

This suggests that a generic approximation of the integral can be obtained by sampling lots of X_i's from $f(x)$ and replacing integration with averaging. *This is the heart of the Monte Carlo method.*

What Follows?

Here, we focus mostly on simulation techniques.

- Some of these will be familiar, others probably not.
- As soon as we know how to produce samples from a distribution, the basic Monte Carlo method described earlier can be used to approximate any expectation.
- However, there are problems where it is not possible to sample from a distribution exactly.

We'll discuss this point more later.

Outline

Generating Uniform RVs

Generating a single U from a uniform distribution on $[0, 1]$ seems simple enough. However, there are a number of concerns to be addressed:

- Is it even possible for a computer, which is precise but ultimately discrete, to produce any number between 0 and 1?
- How can a deterministic computer possibly generate anything that's really random?

While it's important to understand that these questions are out there, we will side-step them and assume that calls of `runif` in R produce bona fide uniform RVs.

Inverse CDF Transform

Suppose we want to simulate X whose distribution has a given CDF F, i.e., for continuous X,

$$\frac{d}{dx}F(x) = f(x).$$

If F is continuous and strictly increasing, then F^{-1} exists.

Procedure
Sampling $U \sim \text{Unif}(0,1)$ and setting $X = F^{-1}(U)$ does the job.
Can you prove it?

This method is (sometimes) called the *inversion method*.

Note:
The assumptions above can be weakened to some extent.

Example: Exponential Distribution

For an exponential distribution with rate λ, we have:

$$f(x) = \lambda e^{-\lambda x} \quad \text{and} \quad F(x) = 1 - e^{-\lambda x}.$$

It is easy to check that the inverse CDF is:

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda}, \quad u \in (0,1).$$

Therefore, to sample X from an Exponential(λ) distribution:

1. Sample $U \sim \text{Unif}(0,1)$.
2. Set $X = -\frac{\log(1-U)}{\lambda}$.

Can be easily "vectorized" to get samples of size n. This is in the `R` function `rexp` — be careful about rate vs. scale parametrization.

Example: Cauchy Distribution

The standard Cauchy distribution has pdf and cdf given by:

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{and} \quad F(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}.$$

This distribution has a shape similar to the normal distribution, but with much heavier tails—Cauchy has no finite moments. However, its CDF can be inverted:

$$F^{-1}(u) = \tan\left[\pi\left(u - \frac{1}{2}\right)\right], \quad u \in (0,1).$$

To generate X from a standard Cauchy distribution:

1. Sample $U \sim \text{Unif}(0,1)$.
2. Set $X = \tan\left[\pi\left(U - \frac{1}{2}\right)\right]$.

To generate a non-standard Cauchy(μ, σ) RV (location μ and scale σ): $\mu + \sigma X$.

Use `rt(n, df=1)` in `R` for simulation from standard Cauchy.

Example: Discrete Uniform Distribution

Suppose we want X to be sampled uniformly from $\{1, \ldots, N\}$. Here is an example where the CDF is neither continuous nor strictly increasing.

The idea is as follows:

1. Divide up the interval $[0, 1]$ into N equal subintervals; i.e., $[0, \frac{1}{N}), [\frac{1}{N}, \frac{2}{N})$, and so forth.
2. Sample $U \sim \text{Unif}(0, 1)$.
3. If $\frac{i}{N} \leq U < \frac{i+1}{N}$, then $X = i + 1$ for $i = 0, 1, \ldots, N - 1$.

More simply, set $X = \lfloor NU \rfloor + 1$.

This is equivalent to `sample(N, size=1)` in `R`.

Example: Triangular Distribution

The (symmetric) pdf of X is given by:

$$f(x) = \begin{cases} 1 + x & \text{if } -1 \leq x < 0; \\ 1 - x & \text{if } 0 \leq x \leq 1. \end{cases}$$

and the corresponding cdf is

$$F(x) = \begin{cases} \frac{x^2}{2} + x + \frac{1}{2} & \text{if } -1 \leq x < 0; \\ -\frac{x^2}{2} + x + \frac{1}{2} = 1 - \frac{1}{2}(1 - x)^2 & \text{if } 0 \leq x \leq 1. \end{cases}$$

If we restrict X to $[0, 1]$, then the pdf becomes:

$$\widetilde{f}(x) = 2\,(1 - x) \quad \text{for } 0 \leq x \leq 1$$

and the CDF is simply:

$$\widetilde{F}(x) = 2x - x^2 = 1 - (1 - x)^2, \quad x \in [0, 1].$$

Example: Triangular Distribution

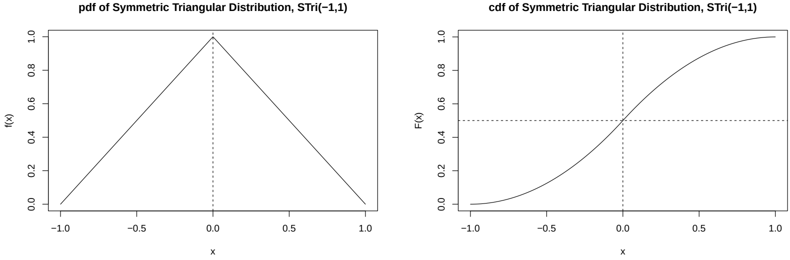

Figure 1: The pdf (left) and cdf (right) of the symmetric triangular distribution.

Example: Triangular Distribution

For this (restricted) "sub-problem", the inverse is:

$$\widetilde{F}^{-1}(u) = 1 - \sqrt{1-u}, \quad u \in [0,1].$$

To sample X from the symmetric triangular distribution:

1. Sample $U \sim \text{Unif}(0,1)$.
2. Set $\widetilde{X} = 1 - \sqrt{1-U}$.
3. Take $X = \pm\widetilde{X}$ based on a flip of a fair coin.

Sampling Normal RVs

While normal RVs can, in principle, be generated using the CDF transform method, this requires evaluation of the standard normal inverse CDF, which is a non-trivial calculation.

- There are a number of fast and efficient alternatives for generating normal RVs.
- The method below, due to Box and Muller, is based on some trigonometric transformations.

Box-Muller Method

This method generates a pair of (independent) standard normal RVs X and Y based on the following facts:

For $X, Y \overset{iid}{\sim} N(0,1)$, the joint pdf in the Cartesian coordinates (X, Y) becomes

$$(2\pi)^{-1} r e^{-\frac{r^2}{2}}, \quad (\theta, r) \in [0, 2\pi) \times [0, \infty)$$

when transformed to to the polar coordinates (θ, R). Then $\theta \sim \text{Unif}(0, 2\pi)$ and $R^2 \sim \text{Exp}(1/2)$ (with the rate) are independent.

So, to generate independent normal X and Y:

1. Sample $U, V \sim \text{Unif}(0, 1)$.
2. Set $R = \sqrt{-2 \log V}$ and $\theta = 2\pi U$.
3. Finally, take $X = R\cos\theta$ and $Y = R\sin\theta$.

Take a linear function to get different mean and variance.

Bernoulli RVs

Perhaps the simplest RVs are Bernoulli RVs — ones that take only values 0 or 1.

To generate $X \sim \text{Ber}(p)$:

1. Sample $U \sim \text{Unif}(0, 1)$.
2. If $U \leq p$, then set $X = 1$; otherwise, set $X = 0$.

In R, use `rbinom(n=1, size=1, prob=p)`.

Binomial RVs

Since $X \sim \text{Bin}(n, p)$ is equal in distribution to $X_1 + \ldots + X_n$, where the X_i's are independent $\text{Ber}(p)$ RVs, the previous slide gives a natural strategy to sample X.

That is, to sample $X \sim \text{Bin}(n, p)$, generate $X_1, \ldots, X_n$ independently from $\text{Ber}(p)$ and set X equal to their sum.

Poisson RVs

Poisson RVs can be constructed from a Poisson process, an integer-valued continuous-time stochastic process.

By definition, the number of events of a Poisson process in a fixed interval of time is a Poisson RV with mean proportional to the length of the interval.

But the time between events are independent exponentials. Therefore, if $Y_1, Y_2, \ldots$ are independent Exp(1) RVs, then

$$X = \max\left\{k : \sum_{i=1}^{k} Y_i \leq \lambda\right\}$$

then $X \sim \text{Poi}(\lambda)$.

In `R`, use `rpois(n, lambda)`.

Chi-square RVs

The chi-square RV X (with n degrees of freedom) is defined as follows:

- Generate n independent $N(0, 1)$ RVs: $Z_1, \ldots, Z_n$
- Take $X = Z_1^2 + \cdots + Z_n^2$.

Therefore, to sample $X \sim \text{ChiSq}(n)$ (or $X \sim \chi_n^2$), take the sum of squares of n independent standard normal RVs.

Independent normals can be sampled using the Box-Muller method.

Student-t RVs

A Student-t RV X (with ν degrees of freedom) is defined as the ratio of a standard normal and the square root of an independent (normalized) chi-square RV with ν df.

More formally, let $Z \sim N(0, 1)$ and $Y \sim \chi^2_\nu$; then

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

is a t_ν RV.

Remember the scale mixture of normals representation?

In `R`, use `rt(n, df=nu)`.

Representation of Student-t as a Scale Mixture of Normals

The scale mixture of normals representation elucidates that the Student-t distribution is effectively a normal distribution with a random variance.

Formally, we define:

- An auxiliary variable $W = 1/\sqrt{Y/\nu}$, with W following an inverse-gamma distribution due to its derivation from a chi-square distribution $Y \sim \chi^2_\nu$.
- The Student-t RV X as $X = Z \cdot W$, where W serves as the scaling factor, adjusting Z's variance.

Thus, the Student-t distribution can be viewed as an infinite mixture of normal distributions with mean 0 and variable variances. It's particularly adept at modeling data with outliers or heavy tails, thus, offers a significant advantage over the normal distribution for such scenarios.

Multivariate Normal RVs

The p-dimensional normal distribution has a mean vector $\boldsymbol{\mu}$ and a $p \times p$ variance-covariance matrix $\boldsymbol{\Sigma}$.

The techniques above can be used to sample a vector $\mathbf{Z} = (Z_1, \ldots, Z_p)^\top$ of independent normal RVs. But how to incorporate the dependence contained in $\boldsymbol{\Sigma}$?

Let $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top$ be the Cholesky decomposition of $\boldsymbol{\Sigma}$. It can be shown that $\mathbf{X} = \boldsymbol{\mu} + \mathbf{L}\mathbf{Z}$ is the desired p-dimensional normal distribution.

Can you prove it?

Outline

Intuition Behind Simulating from a Density Function

Let f be a density function on an arbitrary space $\mathcal{X}$; the goal is to simulate from f.

Note the trivial identity:

$$f(x) = \int_0^{f(x)} du$$

This identity implicitly introduces an *auxiliary variable* U with a conditionally uniform distribution.

The intuition behind this viewpoint is that simulating from the joint distribution of (X, U) might be easy, and then we can just throw away U to get a sample of $X \sim f$...

The "Theorem" on Simulation

Theorem: Simulating $X \sim f$ is equivalent to simulating $(X, U) \sim \text{Unif}\{(x, u) : 0 < u < f(x)\}$ and then throwing away U.

Proof: Write the density of (X, U) and integrate out U.

How to implement this?

- One idea: $X \sim f$ and $U|\{X = x\} \sim \text{Unif}(0, f(x))$.
- A better idea: "conditioning preserves uniformity," i.e., for $A_0 \subset A$,

$$Z \sim \text{Unif}(A) \Rightarrow Z|\{Z \in A_0\} \sim \text{Unif}(A_0).$$

Note: The first approach is used in the accept-reject method.

More on Implementation

The "conditioning preserves uniformity" point can be interpreted as follows:

Procedure: Suppose that A is the set that contains $C_f := \{(x, u) : 0 < u < f(x)\}$. Simulate (X, U) uniformly on A, and keep (X, U) only if $U < f(X)$. Such a kept pair (X, U) is uniformly distributed on the constraint set, C_f, so X has the desired distribution f.

Efficiency of Sampling: The efficiency of sampling depends on how tightly A fits the desired constraint set. For a one-dimensional X, with bounded support and bounded f, a reasonable choice for A is a rectangle.

Extension: This idea extends—this is what the next sections are about!

Example: Beta Simulation

Objective: Simulate from a Beta$(2.7, 6.3)$ distribution.

Procedure

- Start with uniforms in a box containing the Beta$(2.7, 6.3)$ pdf, but keep only those that satisfy the constraint.
- Simulated 2000 uniforms, kept only 744 betas.

This approach demonstrates the use of uniform distributions as a starting point and applying constraints to achieve the desired distribution.

Example: Beta Simulation

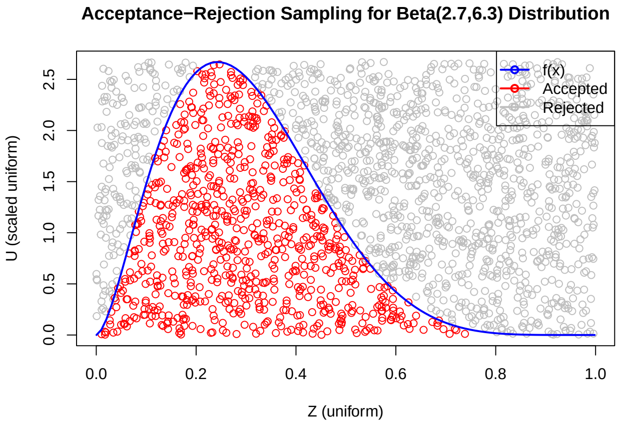

Figure 2: Simulation from Beta$(2.7, 6.3)$ distribution. Initial uniform points are shown as dots. Rejected points are in gray, and red dots represent accepted samples matching the Beta distribution, with its theoretical density curve overlaid.

Example: Beta Simulation

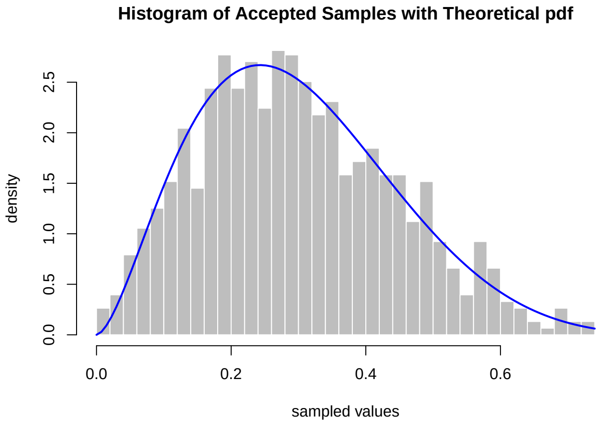

Figure 3: Histogram of simulated Beta$(2.7, 6.3)$ points with theoretical pdf overlaid, showing the close match between empirical pdf (i.e. the density histogram) and the theoretical pdf curve overlaid.

Outline

Outline

Majorization

Concept: Sampling from a distribution with pdf $f(x)$.

Suppose $f(x) \leq h(x) := Mg(x)$ where $g(x)$ is a (simple) pdf and $M > 1$ is a constant. We say $h(x)$ *majorizes* $f(x)$.

Goal: Use samples from the (easy-to-sample) pdf $g(x)$ as "candidate samples" from $f(x)$.

- It's clear that there will be samples from g that are not representative of f, since $g(x) \neq f(x)$.
- The idea is to throw away those "bad" samples.

This approach provides a mechanism for simplifying the sampling process from complex distributions by utilizing a simpler distribution as a proxy.

Acceptance-Rejection Method

The rule that determines when a sample is "thrown away" is called the *acceptance-rejection method*.

To sample X from $f(x)$:

1. Sample $U \sim \text{Unif}(0, 1)$.
2. Sample $\widetilde{X} \sim g(x)$.
3. Keep $X = \widetilde{X}$ if $U \leq \frac{f(\widetilde{X})}{h(\widetilde{X})}$.

Try to prove the following:

- Accept-reject returns a sample from f (see "theorem").
- Acceptance rate or probability is $1/M$.

Goal: To make acceptance rate high, i.e., $M \gtrsim 1$... (why?)

Note: It is not necessary to know f exactly—it's enough to know f only up to a proportionality constant.

Example: Sampling from a Trigonometric Density

Context: The true density is a complex trigonometric function, used in the Robert & Casella book cover (see the code and the figure below).

Approach:

- The majorant used for the acceptance-rejection method is a normal distribution.
- We simulate $n = 2000$ samples from the trigonometric density.

Outcome:

- A histogram of the samples showcases the distribution across the range $[-4, 4]$.
- The density plot highlights the complex behavior of the trig function, contrasting it with the majorant normal distribution.

Insight: So, the acceptance-rejection method can be very effective in simulating with non-standard densities, facilitating the exploration of their properties through simulation.

Example: Sampling from a Complex Trigonometric Density

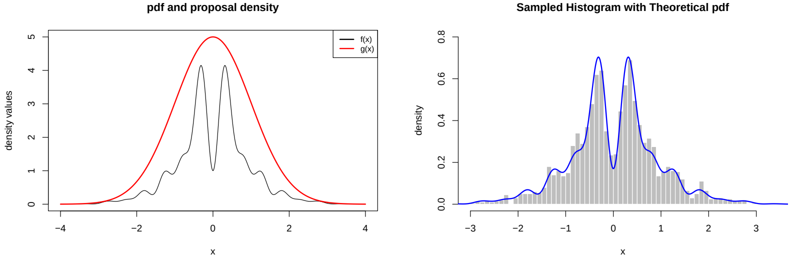

Figure 4: Left: Complex trigonometric density function $f(x)$ and a envelope function $h(x)$, with $f(x)$ in black and $h(x)$ in red (both non-normalized). Right: Normalized histogram of simulated values with theoretical pdf $f(x)$ overlaid for comparison.

Example: von Mises Distribution

Definition and Characteristics: The von Mises distribution is defined by its probability density function (PDF):

$$f(x) = \frac{\exp(-\kappa \cos x)}{2\pi I_0(\kappa)}, \quad x \in [0, 2\pi],$$

where $I_0(\kappa)$ denotes the modified Bessel function of the first kind and order zero, and κ represents the concentration parameter, analogous to the inverse of variance in the normal distribution. This distribution is pivotal for modeling angles and directions, hence its domain $x \in [0, 2\pi]$.

Significance and Applications:

- Often referred to as a *circular normal distribution* due to its bell-shaped curve when plotted in polar coordinates, illustrating its utility in capturing the essence of circular data.

von Mises Distribution (Cont'd)

Significance and Applications:

- Popular in fields such as geology, biology, and environmental science for modeling directional data, showcasing its versatility and applicability across diverse scientific disciplines.

Efficient Sampling Techniques: Sampling from the von Mises distribution, with a specific concentration parameter κ, can be done with an acceptance-rejection method which is optimized by designing a "good" envelope or majorant, typically composed of exponential PDFs, to approximate the target distribution closely. The challenge is constructing an envelope that minimizes the rejection rate.

Implementation Insight: For practical implementation (including constructing the optimal majorant), refer to the `R` code.

von Mises Distribution (Cont'd)

Example with $\kappa = 2$:[1]

- Uses two oppositely oriented exponentials for the envelope.
- $n = 5000$ samples; acceptance rate (for $\kappa = 2$) is approximately 0.81.

Observations:

- The (un-normalized) density plot would show the true density versus the majorant.
- True density and majorant plots illustrate the efficiency and fit of the majorant to the true distribution.

[1] For some reason, majorant construction in the `R` code only works for $\kappa > .66\ldots$

von Mises Distribution (Cont'd)

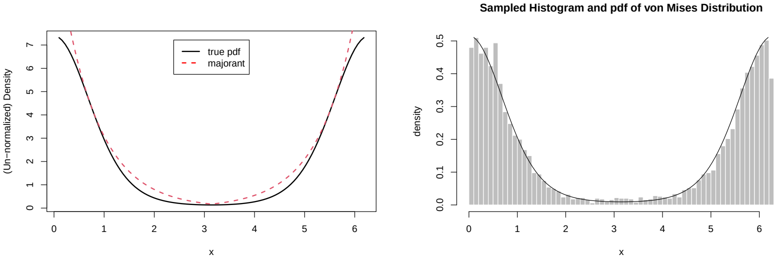

Figure 5: Left: von Mises pdf with parameter $\kappa = 2$ (black) and its majorant function (red) (both un-normalized). Right: Normalized histogram of simulated points from the von Mises distribution using the accept-reject method, with the theoretical pdf curve overlaid.

Example: Small-Shape Gamma Distribution

Problem: Suppose we want $X \sim \text{Gamma}(\nu, 1)$, with $\nu \leq 0.01$.

- Using `rgamma` for this will result in many exact zeros!

Question: Can we develop a better/more efficient method?

Towards an Acceptance-Rejection Method:

- As $\nu \to 0$, $-\nu \log X \to \text{Exp}(1)$ in distribution.
- Suggests good samples of $\log X$ can be obtained using acceptance-rejection with an exponential envelope.

Work in Literature:

- A version of the paper by R. Martin is available at arXiv:1302.1884.
- `R` code is on his research website.

Example: Small-Shape Gamma (Cont'd)

Efficiency Analysis:

- Left panel shows (un-normalized) density of $Z = -\nu \log X$, for $\nu = 0.005$, along with the proposed envelope.
- This approach is more efficient than other accept-reject methods for this problem, based on acceptance rate (as a function of ν).

Acceptance Rate Comparison:

- The plot would typically show acceptance rates for Ahrens-Dieter, Best, Kundu-Gupta, and `rgamss` methods.
- Indicates superiority of the proposed method in terms of acceptance rate.

Note: This advanced method provides a significant improvement in sampling from distributions with very small shape parameters.

Example: Small-Shape Gamma (Cont'd)

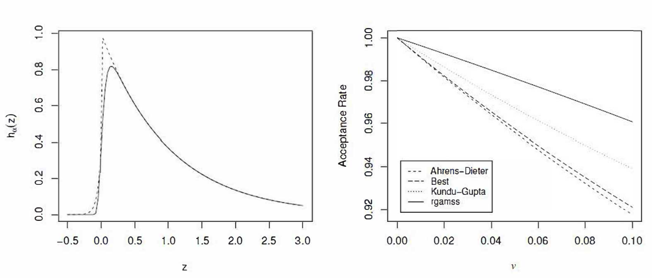

Figure 6: Left: The un-normalized density function (solid) for $Z = -\nu \log X$ with $\nu = 0.005$, compared to a carefully tailored envelope (dashed). Right: Acceptance rates of various methods—Ahrens-Dieter, Best, Kundu-Gupta, and `rgamss`—for sampling from distributions with small shape parameters.

Outline

Bounding Set

Suppose we want to sample from a pdf $f(x)$.

- All that matters is the shape of $f(x)$, so we can remove any constants and consider $h(x) = cf(x)$, for some $c > 0$.
- Define the set in $\mathbb{R}^2$:

$$\mathcal{S}_h = \left\{(u, v) : 0 < u \leq (h(v/u))^{1/2}\right\}.$$

- If $\mathcal{S}_h$ is bounded, then we can find a bounding set that encloses it.
- Ideally, the bounding set should be simple, e.g., a rectangle.

This strategy helps in simplifying the sampling process by focusing on the shape of the distribution, allowing for efficient sampling methods.

Ratio Method

Given: $\mathcal{S}_h$ is bounded, allowing for a rectangle that encloses it.

Sampling Process:

1. Sample (U, V) uniformly from the bounding rectangle.
2. If $(U, V) \in \mathcal{S}_h$, then $X = V/U$ is a sample from $f(x)$.

Proof: Requires some Jacobian-type calculations (see Lange).

Efficiency:

- Some draws from the rectangle will be rejected.
- The efficiency of the sampling algorithm depends on how closely the bounding set matches $\mathcal{S}_h$.

Why Ratio Method Works?

Proposition (Lange): Suppose $k_u = \sup_x\{\sqrt{h(x)}\}$ and $k_v = \sup_x\{|x|\sqrt{h(x)}\}$ are finite. Then the rectangle $[0, k_u] \times [-k_v, k_v]$ encloses $\mathcal{S}_h$. If $h(x) = 0$ for $x < 0$, then the rectangle $R_{u,v} = [0, k_u] \times [0, k_v]$ encloses $\mathcal{S}_h$. Finally, if the point (U, V) sampled uniformly from the enclosing set (i.e. the rectangle $R_{u,v}$) falls within $\mathcal{S}_h$, then the ratio $X = V/U$ is distributed according to $f(x)$.

Proof (modified from Lange):

- From the definition of $\mathcal{S}_h$ it is clear that the permitted u lie in $[0, k_u]$.
- Multiplying the inequality $u \le \sqrt{h(v/u)}$ by $|v|/u$ implies that $|v| \le k_v$.
- If $h(x) = 0$ for $x < 0$, then no $v < 0$ yields a pair (u, v) in $\mathcal{S}_h$.

Proof (continued)

Finally, note that the transformation $(u, v) \to (v/u, u)$ has Jacobian $-1/u$. Hence,

$$\begin{aligned}
F_X(x_0) &= P(X = V/U \leq x_0) \;\; \text{(since } X = V/U \text{ are accepted)} \\
&\propto \int\int \mathbf{1}_{\{\frac{v}{u} \leq x_0\}} \mathbf{1}_{\left\{0 < u \leq \sqrt{h\left(\frac{v}{u}\right)}\right\}} \, du \, dv \;\; \text{(unif. integration on } \mathcal{S}_h) \\
&= \int\int \mathbf{1}_{\{x \leq x_0\}} \mathbf{1}_{\left\{0 < u \leq \sqrt{h(x)}\right\}} \, u \, du \, dx \;\; \text{(change of var. } x = v/u) \\
&= \int_{-\infty}^{x_0} \frac{1}{2} h(x) dx = \int_{-\infty}^{x_0} c' f(x) dx
\end{aligned}$$

is the distribution function of the accepted X up to a normalizing constant. □

Example: Gamma Distribution

Objective: Sample $X \sim \text{Gamma}(\nu, 1)$ for non-integer ν.

- To apply the ratio method, take $h(x) = x^{\nu-1}e^{-x}$ for $x > 0$.
- It can be shown that, in general, if $h(x) = 0$ for $x < 0$, then the rectangle $[0, k_u] \times [0, k_v]$, with:

$$k_u = \sup_x \{h(x)^{\frac{1}{2}}\} \quad \text{and} \quad k_v = \sup_x \left\{|x|h(x)^{\frac{1}{2}}\right\}$$

  encloses $\mathcal{S}_h$.
- For the Gamma$(\nu, 1)$ case:

$$k_u = \left(\frac{\nu - 1}{e}\right)^{\frac{\nu-1}{2}} \quad \text{and} \quad k_v = \left(\frac{\nu + 1}{e}\right)^{\frac{\nu+1}{2}}$$

Example: Gamma Distribution (Cont'd)

Ratio Method Implementation: Sampling $X \sim \text{Gamma}(\nu, 1)$

1. Sample $U', V' \sim \text{Unif}(0,1)$, set $U = k_u U'$ and $V = k_v V'$.
2. Set $X = V/U$.
3. If $U \leq \sqrt{h(X)} = X^{(\nu-1)/2} e^{-X/2}$, then accept X.

Example: $\nu = 7.7$.

- The ratio method has an acceptance rate ≈ 0.44.
- This example illustrates the density of accepted X values across the range $[5, 20]$.

Insight: The acceptance rate indicates the efficiency of the sampling algorithm. Higher acceptance rates suggest a closer match between the sampling method and the target distribution's characteristics.

Example: Gamma Distribution (Cont'd)

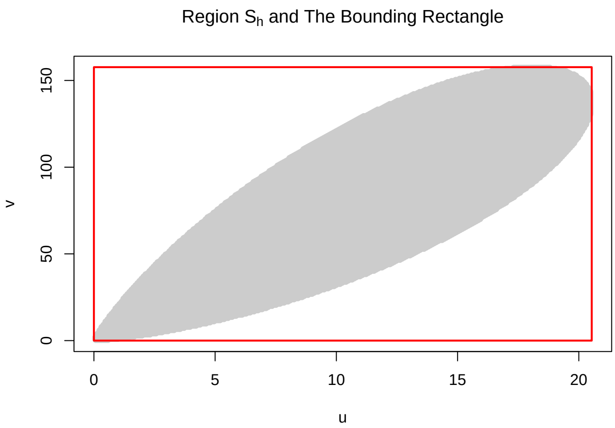

Figure 7: The region $\mathcal{S}_h$ in the uv-plane (shaded region) for Gamma($\nu = 7.7, 1$) distribution and the bounding rectangle (red) encompassing $\mathcal{S}_h$.

Example: Gamma Distribution (Cont'd)

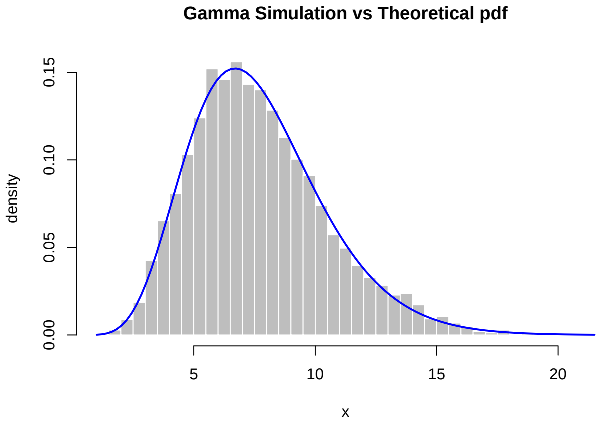

Figure 8: Histogram of 5000 simulated points from Gamma($\nu = 7.7, 1$) distribution, generated using the ratio method. The theoretical Gamma pdf curve is also overlaid (solid line).

Outline

Motivation for Sampling Importance Resampling (SIR)

Exact Sampling:

- Previous methods are "exact" in that the distribution of the draw X (given that it's accepted) matches the target distribution f.

Question:

- Is it necessary for the sampling to be exact?

An Approximate Sampling Approach:

- An interesting idea is to sample from a different distribution $g(x)$, similar to $f(x)$, and weight these samples in such a way that a resample according to the given weights resembles a sample from $f(x)$.
- This is the core idea behind Sampling Importance Resampling (SIR) (note the similarity to acceptance-rejection sampling).

Sampling Importance Resampling (SIR) Algorithm

The target pdf $f(x)$, which may be known only up to a proportionality constant.

Preparation: Let $g(x)$ be another pdf with the same support as $f(x)$.

SIR Algorithm Steps:

1. Take an independent sample $Y_1, \ldots, Y_m$ from g.
2. Calculate the standardized importance weights
$$\widetilde{w}(Y_j) = \frac{w(Y_j)}{\sum_{i=1}^{m} w(Y_i)}, \text{ where } w(Y_j) = \frac{f(Y_j)}{g(Y_j)}, \quad j = 1, \ldots, m.$$
3. Resample $X_1, \ldots, X_n$ with replacement from $\{Y_1, \ldots, Y_m\}$ with probabilities $\widetilde{w}(Y_1), \ldots, \widetilde{w}(Y_m)$.

Outcome: The resulting sample $X_1, \ldots, X_n$ is approximately distributed as $f(x)$.

Remarks on Sampling Importance Resampling (SIR)

Approximation Quality: As $m \to \infty$, $P(X_i \in A) \to \int_A f(x)\,dx$, indicating the approximate nature of sampling.

Choice of Envelope g:

- The selection of g is critical; $f(x)/g(x)$ should not be too large to avoid dominance by a single weight.
- Violation may result in X-samples being almost identically a single value.

Sample Size Considerations: Theory suggests m should be large, where "large" depends on the desired sample size n from the target.

Comparison with Other Methods: Monte Carlo estimates based on the SIR sample typically have variances larger than those obtained with direct sampling or importance sampling.

Example: Bayesian Inference via SIR

Problem Context: Want to sample from the posterior for the following setup.

- Consider the likelihood function
$$L(\theta) \propto \prod_{i=1}^{n} (1 - \cos(x_i - \theta)), \quad 0 \leq \theta \leq 2\pi.$$
- Observed data $(x_1, \ldots, x_n)$ is provided in the `R` code.

Prior Distribution: Assume θ is given a Unif$(0, 2\pi)$ prior distribution, $\pi(\theta) = \frac{1}{2\pi}$.

Posterior Distribution: Then posterior $\propto$ prior $\times$ likelihood $\propto$ likelihood.

SIR Algorithm Application: Use the SIR algorithm with the prior as the envelope; $N = 1000$.

Visualization: See the plots below.

Bayesian Inference via SIR (Cont'd)

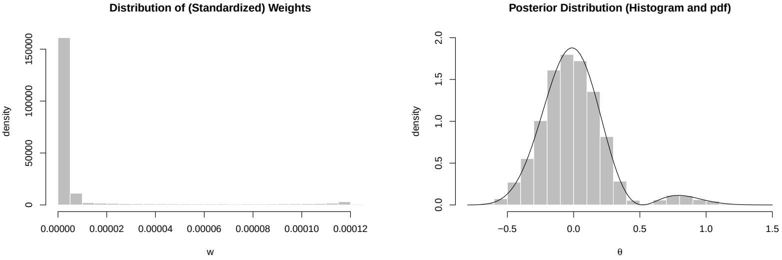

Figure 9: Left: Histogram of the importance weights, showing the distribution of weights assigned to each sampled point. Right: Histogram of the SIR sample with a density plot overlaid, illustrating the distribution of θ after resampling according to importance weights.

Bayesian Inference via SIR (Cont'd)

Interpretation:

- The left histogram provides insights into the variability and distribution of importance weights across the sampled points.
- The right histogram, along with the density overlay, demonstrates how the SIR process adjusts the prior distribution to better reflect the observed data, resulting in a posterior distribution of θ.

Note: This example showcases the practical application of SIR in Bayesian inference, emphasizing the transformation from prior to posterior distribution through the weighting and resampling mechanism.

Outline

Remarks on Simulating Random Variables

Importance of Simulation:

- Simulating random variables is crucial for various applications of the Monte Carlo method.

Methods of Simulation:

- Some distributions can be easily simulated via the inversion method, while others require more intricate approaches.

Strategies and Implementations:

- The "Fundamental Theorem of Simulation" provides a general strategy for simulating from non-standard distributions, though its implementation may be complex.
- The accept-reject method is a practical implementation with efficiency depending on how closely the envelope function matches the target density.

Remarks on Simulating Random Variables

Challenges and Solutions:

- Identifying an effective envelope function is challenging.
- Various automatic/adaptive methods have been developed to address this.

Recurring Concepts:

- The accept-reject idea not only plays a crucial role in Rejection Sampling context, but will recur in future discussions and applications.