CS 5135/6035 Learning Probabilistic Models

Lecture 19: Accept-Reject Method; Monte Carlo Integration

Gowtham Atluri

November 4, 2018

Reading Material

- Chapter 2. Random Variable Generation
  Christian Robert and George Casella. Introducing Monte Carlo Methods with R
- Andrieu et al. An introduction to MCMC for machine learning, Machine learning, 2003.

Topics

- Accept-Reject Methods
  - Limitations inverse-transform methods
  - Approach to sampling
  - Example
  - Advantages of Accept-Reject method
- Integrals in Bayesian approaches
  - Challenges
- Monte Carlo Methods
  - History
  - Gentle introduction
  - Integration using sampling

Posterior Estimation: Random sampling

- When the posterior has a familiar functional form (due to conjugacy):
  - we can compute a summary of the distribution analytically
    - mean of a $Beta(a,b)$ is $\frac{a}{a+b}$
  - we can simulate data from the posterior and summarize
    - $\theta \sim Beta(a,b)$
- When posterior does not have a familiar form
  - compute values of the posterior on a grid of points
  - we can approximate the posterior by a discrete posterior
  - How to do point estimation?
  - high-dimensions (i.e., dimensionality of parameters $\boldsymbol{\theta}$)
    - combinatorially evaluating at grid points on all dimensions
    - computationally prohibitive
  - Solution: Draw samples from the posterior and do point-estimation

Random Sampling

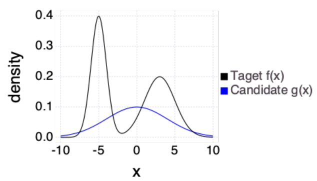
- Uniform random variable is very important
  - many other random variables can be derived and transformed from it
- True vs. Pseudo random numbers
- The inverse transform
  - Compute $F(x) = \int_{-\infty}^{x} f(t)dt$
  - Set $u = F(x)$
  - Determine $x = F^{-1}(u)$
  - Draw $u \sim \mathcal{U}(0,1)$, then compute $x = F^{-1}(u)$
  - $x \sim f(x)$
- Variants
  - General transformation method
  - Discrete distribution
  - Mixture of densities
- Can we sample from a distribution $f(x) \propto x^{\alpha-1}(\frac{1-x}{1+x})^{\beta-1}$?

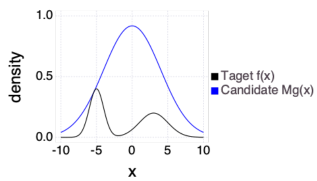
Accept-Reject Methods

- How to generate random variables when inverse transform method and even general transformations fail?
  - E.g., if a normalized version of the pdf $f(x)$ is not available?
    - $f(x) \propto x^{\alpha-1}(\frac{1-x}{1+x})^{\beta-1}$
  - If $F(x) = \int_{-\infty}^{x} f(t)dt$ cannot be expressed in a closed form?
  - If it is not trivial to determine F^{-1}?
- We turn to *indirect* methods
  - We generate a *candidate random variable*
  - *Accept it* subject to passing a test
- This is a powerful class of methods
  - allows us to simulate from virtually any distribution

Accept-Reject Methods

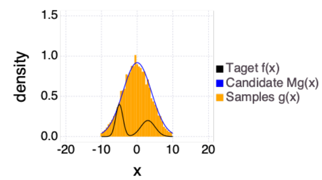
- These *Accept-Reject* methods require us to know the functional form of density f upto a multiplicative constant
  - f is known as *target density*
- We use a simpler density g, called the *candidate density*
  - to generate random variables for which simulation is done
- Constraints:
  1. f and g have compatible supports (i.e., $g(x) > 0$, when $f(x) > 0$)
  2. There is constant M such that $f(x)/g(x) \leq M$ for all x
     - So, $Mg(x)$ *envelopes* $f(x)$

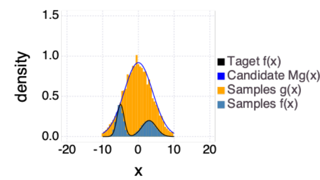




Accept-Reject Methods

- Approach
  1. Generate $y \sim g$
  2. Independently generate $u \sim \mathcal{U}(0, 1)$
  3. If $$u \leq \frac{1}{M}\frac{f(y)}{g(y)},$$ then *accept* y as a sample
  4. else *reject* y, discard u, and start again with step 1.





Accept-Reject Method: Example

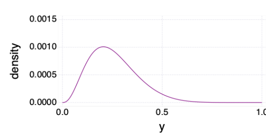
Example: Draw samples from a distribution $f(x) \propto x^{\alpha-1}(\frac{1-x}{1+x})^{\beta-1}$, where $\alpha = 2.7$ and $\beta = 6.3$

- $f(x)$ is our *target density*

Choices involved:

1. We need to select a *candidate density*
   - Let us treat Uniform distribution as a candidate density $g(x) = \mathcal{U}(0, 1)$
   - We can pick any other distribution that we can simulate
2. Next, we need to select M such that $f(x)/g(x) \leq M, \forall x$
   - Here $f(x) \leq M$, so we need to pick M as $\max(f(x))$

```
x = collect(0:0.001:1);
f_x = x.^2.7 .* (((1.-x)./(1.+x)).^6.3);
M = maximum(f_x)
```




```
## 0.001008857762244691
```

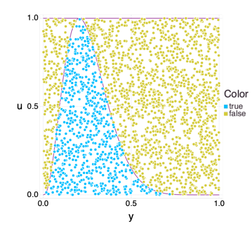

Accept-Reject Method: Example

Approach:

1. We generate a sample y from candidate distribution $y \sim \mathcal{U}(0, 1)$
2. We generate a sample u from $u \sim \mathcal{U}(0, 1)$
3. If $u \leq \frac{1}{M}\frac{f(y)}{g(y)}$, we accept the sample y
4. Otherwise, we *reject* y as a sample from the target distribution.

```
x = collect(0:0.001:1);
f(x) = x.^2.7 .* (((1.-x)./(1.+x)).^6.3);
M = maximum(f(x));

n = 2500;
u = rand(Uniform(0,1),n);
y = rand(n);
x = y[u.<f(y)/M];
plot(x=y,y=u,color = u.<f(y)/M,Geom.point);
```



Properties of Accept-Reject algorithm

- It suffices to know $f(x)$ upto a multiplicative constant
  - The normalizing constant can be absorbed into M
  - $\frac{f(x)}{cg(x)} \leq M \implies \frac{f(x)}{g(x)} \leq M'$
- Efficiency of Accept-Reject algorithm can be measured in terms of its acceptance probability
  - $u \leq \frac{1}{M}\frac{f(y)}{g(y)}$
  - higher the acceptance probability, fewer wasted simulations from g
- If the bound $f(x) \leq Mg(x)$ is not tight (i.e., M is replaced by a larger constant)
  - the algorithm is still valid, but less efficient
- The probability of acceptance is $1/M$
  - M should be as small as possible for computational efficiency.

Monte Carlo Integration

- Integrals in Bayesian approaches
- Challenges with solving integrals
- Monte Carlo methods
  - History
  - A gentle introduction
  - Computing integrals

Integrals in Bayesian approaches

Bayesian approaches require solving integrals in different scenarios:

1. Normalization (e.g., for determining the posterior distribution)

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

2. Marginalization (e.g., for averaging nuisance parameters)

$$p(\theta_1|y) = \int_{\theta_2 \dots \theta_k} p([\theta_1, \theta_2, \dots, \theta_k]|y)d\theta_2 \dots d\theta_k$$

3. Expectation (e.g., to obtain summary statistics of the posterior)

$$\mathbb{E}(f(\theta)) = \int f(\theta)p(\theta|y)d\theta$$

Integrals in Bayesian approaches

Challenges:

- Integrals in large dimensional spaces

$$p(\theta_1|y) = \int_{\theta_2 \dots \theta_k} p([\theta_1, \theta_2, \dots, \theta_k]|y)d\theta_2 \dots d\theta_k$$

- Closed form solutions to integrals are not always possible

Solutions:

- We will rely on numerical approaches
  - Quadrature methods
    - mainly suited for unidimensional integration
    - works with finite intervals
  - Simulation methods: Monte Carlo

Monte Carlo Methods: some history

- Stanislow Ulam invented Monte Carlo Methods
- In 1946, he was recovering from illness and was playing solitaire
- He wondered: *what are the chances that a particular solitaire laid out with 52 cards would come out successfully?*
- Exhaustive combinatorial calculations were not feasible
- He came up with a practical approach
  - simply record the outcomes of 100 games and calculate a crude percentage that way
- He reached out to John von Neumann
  - the pair began to work on the problem together
- Metropolis, a colleague of Ulam and von Neumann, gave a code name 'Monte Carlo'
  - after the European gambling capital

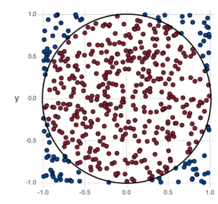
Monte Carlo Methods: a general introduction

- Monte Carlo methods are a broad class of computational algorithms
  - that rely on repeated random sampling to estimate a desired quantity

Example: Can we determine the value of π using MC method?

Approach:

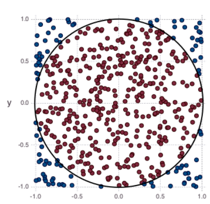
1. Draw a square, and inscribe a circle in it
2. Uniformly scatter points over the square
3. Count the number of points inside the circle
4. Compute fraction of points inside the circle
   - Area of Circle/Square $= \pi r^2/(2r)^2 = \pi/4$
5. $\hat{\pi} = 4 \times$ fraction of points in circle



Monte Carlo Methods: a general introduction

```
n=10000;
x = -1+rand(n)*2;
y = -1+rand(n)*2;
in_or_out = Int.(x.^2 + y.^2 .<= 1);
pi = 4*sum(in_or_out.==1)/length(in_or_out)
```

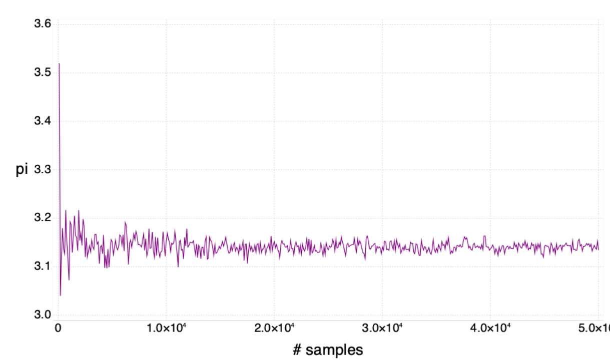
```
## 3.1516
```



- When estimating a value is not possible to compute in traditional ways
  - Monte Carlo methods can be effective
- Monte Carlo Methods rely on an endless flow of random numbers
  - for well-known or new distributions

Monte Carlo methods: Convergence

Estimates converge as the number of samples increase.

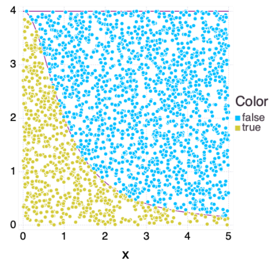


Monte Carlo Integration: Introduction

- Computing a definite integral $\int_a^b f(x)dx$ is equivalent to computing the area under the curve

$$\text{Example: compute } \int_0^5 \frac{4}{1+x^2} dx$$

- The same Monte Carlo approach for computing π applies here too!
  - We know value of integral $A_1 = \int_0^5 1dx = 5; A = 4A_1 = 20$
  - Scatter n points uniformly in the range $[0, 5]$
  - Compute proportion of points p in region of interest
  - Area under the curve is the area Ap



Summary

- Accept-reject methods
  - for sampling from arbitrary distribution
  - when inverse-transform methods are not suited
  - only needs to know $f(x)$ upto a normalizing constant
  - a tighter envelope $g(x)$ results in higher efficiency
- Integrals in Bayesian approaches
  - normalization of posterior
  - point-estimation
  - averaging nuisance parameters
- Integrals can be intractable
  - high-dimensional spaces
  - no closed form solution
- Monte Carlo methods (gentle introduction)
  - relies on random sampling to compute a desired quantity
  - compute integrals by drawing random samples