Monte Carlo Sampling

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Motivation for Sampling & Monte Carlo

- Goal: Find the expectation of function, $f(\mathbf{x})$ with respect to a distribution $p(\mathbf{x})$
- The variables, $\mathbf{x}$ may be discrete, continuous, or both
- For the continuous case, we want

$$E[f(\mathbf{x})] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

- For many real problems the evaluation of the integral is intractable
  - Dimensionality of the parameter or latent space
  - Complexity of the distribution
  - Required integrations may not have closed-form solutions and numerical integration is impractical

Basic Sampling Approach

- Draw a set of observations, $\mathbf{x}$, independently from $p(\mathbf{x})$
- Approximate the integral with finite sums over M samples

$$\hat{f} = \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{x}^{(m)})$$

- $E[\hat{f}] = E[f(\mathbf{x})]$ and $\hat{f} \rightarrow E[f(\mathbf{x})]$ as $M \rightarrow \infty$ by LLN as long as independent samples
- The variance of the estimator is given by

$$\text{Var}[\hat{f}] = \frac{1}{N} E[f(\mathbf{x}) - E[f(\mathbf{x})]^2$$

Sampling Issues

- $p(x)$ is too complicated for direct sampling
  - Function transformation only works with a well-described distribution where we can calculate and then invert the integral
- Grid sampling of $p(x)$ is inefficient and inaccurate
- Approaches
  - Rejection sampling
  - Importance sampling
  - Sampling importance resampling
  - Markov chain Monte Carlo (MCMC)

Rejection Sampling Framework

- Rejection sampling allows for sampling from more complex distributions
- Let $p(\mathbf{x})$ be a distribution that is hard to invert, but easy to evaluate at $\mathbf{x}$, possibly up to some normalizing constant Z so

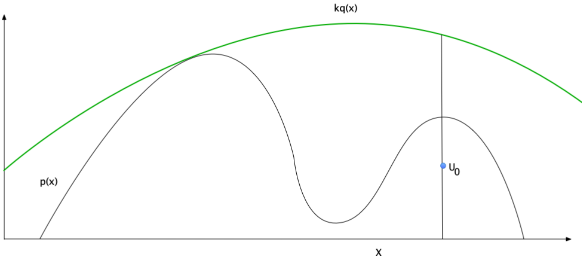
$$p(\mathbf{x}) = \frac{1}{Z_p}\tilde{p}(\mathbf{x})$$

  where $\tilde{p}(\mathbf{x})$ is easy to evaluate but Z_p is unknown

Basis for Rejection Sampling

- Need a simpler distribution, $q(\mathbf{x})$ from which we can easily draw sample, called the *proposal distribution*
- Now choose k such that $kq(\mathbf{x}) \geq \tilde{p}(\mathbf{x})$ for all $\mathbf{x}$, where $kq(\mathbf{x})$ is called the *comparison function*
- Rejection sampling steps for a univariate distribution
  - Generate x_0 from $q(x)$
  - Generate u_0 from $U[0, kq(x_0)]$
  - If $u_0 > \tilde{p}(x_0)$ reject the sample, else retain

Rejection Sampling Example



Selecting k

- The x come from $q(x)$ and are accepted with probability $\tilde{p}(x)/kq(x)$
- So the probability that a sample is accepted is

$$\begin{aligned} \text{Pr(accept)} &= \int \{\tilde{p}(x)/kq(x)\} q(x) dx \\ &= \frac{1}{k} \int \tilde{p}(x) dx \end{aligned}$$

- So the proportion of points rejected depends on k, and we want it as small as possible so that $kq(z) > \tilde{p}(z)$

Comments on Rejection Sampling

- To be effective, the comparison function must be close to $p(x)$
- Multimodal distributions are challenging for envelope functions
- Not effective for high-dimensional distributions, as shown by example
  - Suppose a Gaussian target, $p(\mathbf{x}) \sim N(\mathbf{0}, \sigma_p^2 \mathbf{I})$, and a Gaussian envelope function, $q(\mathbf{x}) \sim N(\mathbf{0}, \sigma_q^2 \mathbf{I})$
  - Clearly $\sigma_q^2 > \sigma_p^2$ so that $kq(\mathbf{x}) > p(\mathbf{x})$
  - In D dimension, the optimal $k = (\sigma_q / \sigma_p)^D$ and the acceptance region is the ratio of volumes under $p(\mathbf{x})$ and $kq(\mathbf{x})$ which is $1/k$
  - This diminishes exponentially with D
- Rejection sampling can be used as a part of a larger approach to sampling in higher-dimensional spaces.