Lecture 15: Basic Sampling (Beyond Random Sampling)

Lecturer: Anshumali Shrivastava *Scribe By: Mingyi Xu (mx13), Jiaxin Fan (jf68)*

1 Intro: Estimation and Sampling

1.1 Estimation

Assume that there is a function $f(x) = -x^2 + 2 * x$ as Figure 1 shows. How can we get the area between $f(x)$ and the x-axis over the interval $x \in (0, 1)$?

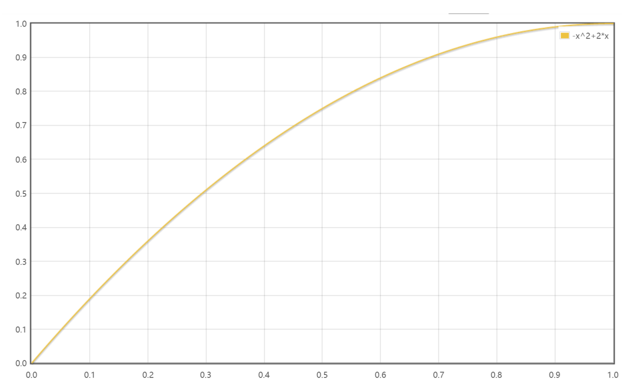

Figure 1: $f(x) = -x^2 + 2 * x$

One way to compute the area is using integral:$\int_0^1 f(x)dx$. However, the task can be challenging when the function has multiple variables, say x_1, x_2, x_3,..., x_{10}, which leads to a 10-dimensional integral.

To solve this problem, we can use an estimation method derived from the Monte Carlo methods to approximate the area between $f(x)$ and the x-axis for $x \in (0, 1)$:

1. Select a rectangular region called U spanning from $[0, 1]$ in both x, y directions. Denote its area as S_U.
2. Implement a function $Random(x, y)$ to generate a random point (x, y). This function selects x and y independently and uniformly from the interval $(0, 1)$. If $f(x) \geq y$, set the value of the point as 1; otherwise, set the value as 0.
3. Repeat Step 2 for T times (which T is large enough), summing up the values of every generated point and setting the total as t.
4. The estimated area under $f(x)$ can be computed as follows: $\frac{t}{T} * S_U$.

This method gives us an easy approach to compute the estimated area under $f(x)$, even if the function has n variables. By generating random values for x_1, x_2, x_3,..., x_n, and a corresponding y value multiple times within a chosen region U, the method avoids considering high-dimension integrals.

Besides, the choice of region U would affect the result. Since a large area of U would increase the variance of the result by reducing the proportion of the expected area, the closer the region U is to the desired area, the better the results will be.

1.2 Sampling

Sampling is a widely used technique in many disciplines. However, generating samples from complex distributions can be challenging. Since many of the available algorithms are designed for well-known distributions like the uniform, Gaussian, or Gamma distributions, how can we sample based on more complex ones?

Thankfully, we can transform the known distributions into the ones we want via several tools. In this section, we are introduced to three key techniques: the Inversion Method, Rejection Sampling, and Importance Sampling.

Before we proceed, let's recap two key concepts:

1. Probability Density Function (PDF): Represented typically by $f(x)$, the PDF traces the likelihood of a continuous random variable assuming a specific value.

2. Cumulative Distribution Function (CDF): Often symbolized by $F(x)$, the CDF quantifies the probability that the random variable X will not exceed the value x.

With these foundational concepts in mind, let's get a closer look at the three techniques.

2 Inversion Method

The Inversion Method is a technique used to generate random numbers that conform to a specific, even complex, probability distribution. It effects when the cumulative distribution function (CDF) of the desired distribution is known.

Assume X is a continuous random variable with its cumulative distribution functionß F(x). In this case, the random variable $y = F(x)$ follows a uniform distribution on the interval [0, 1]. Inverse transform sampling reverses this process: first, for the random variable y, which is drawn from the interval $U[0, 1]$, since the random variable $F^{-1}(y)$ has the same distribution as X, $u = F^{(-1)}(y)$ can be regarded as a random sample generated from the distribution $F(x)$.

To prove the Inversion Method works, we assume that $F(x)$ is a continuous, strictly increasing cumulative distribution function. Then generate a uniform random variable U in the interval $[0, 1]$. We can define a new random variable X as follows:

$$X = F^{-1}(U) \tag{1}$$

To show that X has the desired distribution, we prove that the cumulative distribution function of X matches $F(x)$:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(F(F^{-1}(U)) \leq F(x)) = P(U \leq F(x)) = F(x) \tag{2}$$

Therefore, X has the cumulative distribution function $F(x)$, indicating that it follows the desired distribution.

3 Rejection Sampling

Rejection sampling allows us to produce samples from a target distribution by leveraging a different, more easily sampled distribution. Here's the key:

1. Find a Known Distribution: Find an available distribution, denote the PDF as $g(x)$.
2. Determine a Constant: Find a constant M to satisfy the inequality $f(x) \leq M * g(x)$ for all x.

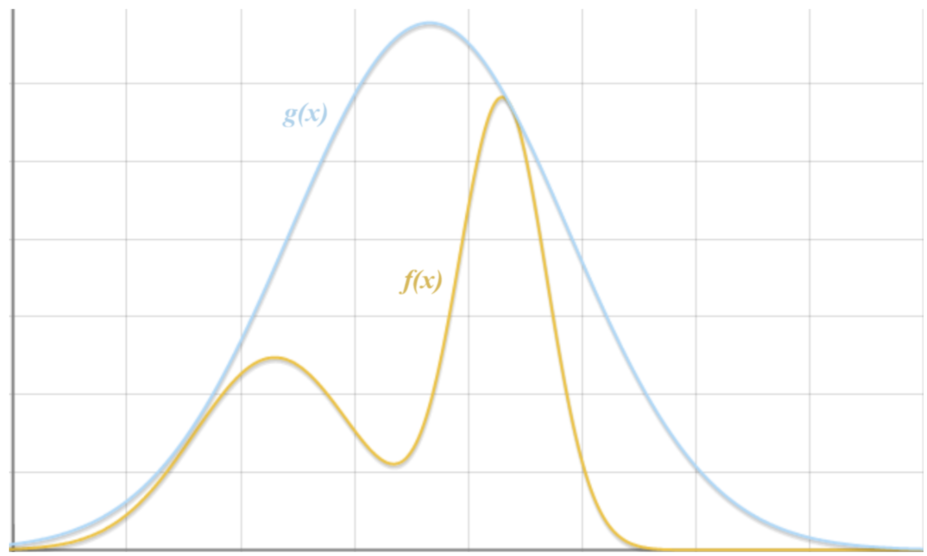

Figure 2: An example of rejection sampling.

The algorithm is shown as follow:

Algorithm 1 Rejection Sampling

Require: Generate samples from desired distribution $f(x)$

```
Draw x ~ g, u ~ Uniform[0,1]
if u ≤ f(x) / (M·g(x)) then
    return x
else
    continue
end if
```

For any specific value of x, the probability of acceptance is given by $P_accept(x) = \frac{f(x)}{M \cdot g(x)}$. Considering all values of x, the overall acceptance rate for the samples that fit the target distribution $f(x)$ is:

$P(x|accepted) = \int \frac{f(x)}{M \cdot g(x)} * g(x)dx = \frac{1}{M} \int f(x)dx.$

A high value of M results in a lower acceptance rate would decreases the performance of the sampling process. To keep it effective, it's important to find an M that's as small as possible, which often means choosing a $g(x)$ that is close enough to the target distribution. This is similar to what we have discussed in section 1.1, wherein a large domain U can lead to increased variance in estimation.

4 Importance Sampling

Importance Sampling is a Monte Carlo method used for estimating properties instead of sampling. Suppose we are trying to estimate the expected value of a function f(x) under a certain probability distribution $w(x)$. We can estimate the expectation as :

$$E[f(x)] = \int f(x)w(x)dx \approx \frac{1}{n}\sum_i f(x_i) \tag{3}$$

Using this equation, we can simply sample from the distribution $w(x)$ and take the advantage of all samples to estimate the expected value $E[f(x)]$. But if sampling from $w(x)$ is hard, how can we still obtain the estimation? The answer is to find another distribution function $g(x)$, which is easier to draw samples. So we can get the following transformation:

$$E[f(x)] = \int f(x)w(x)dx = \int f(x)\frac{w(x)}{g(x)}g(x)dx \approx \frac{1}{n}\sum_i f(x_i)\frac{w(x_i)}{g(x_i)} \tag{4}$$

where x is sampled from $g(x)$ and g(x) is non-zero. And $\frac{w(x_i)}{g(x_i)}$ is called importance weight.