Inverse CDF Sampling

Troy Stribling

June 30, 2018

1 Introduction

Inverse CDF sampling is a method for obtaining samples from both discrete and continuous probability distributions that requires the CDF to be invertable. The method proposes a CDF value from a Uniform random variable on $[0, 1]$ which is then used as input into the inverted CDF to generate a sample with the desired discrete or continuous distribution. Here examples for both cases are discussed. For the continuous case a proof is given that demonstrates the samples produced have the expected distribution.

2 Sampling Discrete Distributions

A discrete probability distribution consisting of a finite set of N probability values is defined by,

$$\{p_1, p_2, \ldots, p_N\} \tag{1}$$

with,

$$\sum_{i=1}^{N} p_i = 1.$$

The CDF specifies the probability that $i \leq n$ and is given by,

$$P(n) = \sum_{i=1}^{n} p_i, \tag{2}$$

where $P(N) = 1$.

For a given CDF propsal, P^*, equation (2) can always be inverted by evaluating it for each n and searching for the value of n that satisfies, $P(n) \geq P^*$. A sampler can be implemented in Python with the following,

```
import numpy

nsamples = 100000
cdf_proposals = numpy.random.rand(nsamples)
samples = [numpy.flatnonzero(cdf >= cdf_proposals[i])[0] for i in range(nsamples)]
```

Consider the following disrete distribution,

$$\left\{\frac{1}{12}, \frac{1}{12}, \frac{1}{6}, \frac{1}{6}, \frac{1}{12}, \frac{5}{12}\right\} \tag{3}$$