inversetransformsampling

May 8, 2023

1 Inverse Transform Sampling

Inverse Transform Sampling (ITS) is one of the most fundamental techniques for generating samples from a specified probability distribution based on samples from a uniform distribution. The cumulative distribution function of a given probability distribution is a function $F : R \to R$ that is non-decreasing and right-continuous, expressed as:

$$\lim_{x \to -\infty} F(x) = 0 \quad \lim_{x \to \infty} F(x) = 1.$$

The cumulative distribution function uniquely defines the probability distribution. The relationship between the cumulative distribution function and the probability density function $f(x)$ is given by:

$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

When it is possible to find the inverse of the cumulative distribution function, denoted as:

$$U = F(x) \to x = F^{-1}(U),$$

then the random variable x has a distribution with cumulative distribution function F, and U is a random variable with a uniform distribution on the interval $(0, 1)$. The new random variable is defined as $X = F^{-1}(U)$. We have:

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

A sequence of pseudorandom numbers is generated: $U_1, U_2, ..., U_n \in (0, 1)$, which is then transformed into a sequence $X_1, X_2, ..., X_n \in (-\infty, \infty)$. The numbers X_i have a probability distribution with the cumulative distribution function F. The inversion of the cumulative distribution function can also be used in the case of discrete probability distributions with probabilities $p_k = \mathbb{P}(X = k)$, where $k = 0, 1, 2, ...$, which can be generated using the sequence U_i according to the formula:

$$X_n = min\{k : U_n \leq \sum_{i=0}^{k} p_i\}, \quad n = 1, 2, 3, ...$$

Below code demonstrates an example of using the Inverse Transform Sampling (ITS) method to generate samples from a continuous distribution, specifically the Weibull $\mathcal{W}e(2, 1)$ distribution with shape parameter 2 and scale parameter 1. The generator takes a function as an argument, which in this case is an anonymous function that applies the inverse transformation formula for the Weibull distribution: $\sqrt{-\log(1 - x)}$. This function maps values from a uniform distribution

(generated internally) to values from the Weibull distribution. The following plots represent the following: the cumulative distribution function, where the Empirical CDF is depicted in blue and the Theoretical CDF in red. It can be observed that both curves largely overlap, indicating a correctly implemented algorithm. The next plot is the Probability Density Function, which shows a histogram and the theoretical density. The last of the plots is the QQ-plot, also known as the Quantile-Quantile plot. The empirical quantiles are plotted on the Y-axis, while the theoretical quantiles are described on the X-axis. Below, in the table, basic descriptive statistics such as mean, variance, skewness, and kurtosis are presented for both absolute error and theoretical error. Next code demonstrates an example of using the inverse transform sampling method to generate samples from a discrete distribution, specifically the Bernoulli $\mathcal{B}(\frac{1}{2})$ distribution with a success probability of $\frac{1}{2}$. This distribution represents a binary random variable that takes either the value 0 or 1. If the input value x is less than or equal to 1/2, the function returns 0; otherwise, it returns 1. This function serves as the inverse of the cumulative distribution function for the Bernoulli distribution. This generator applies the inverse transformation defined by bernoulli_quantile to map values from a uniform distribution to values from the Bernoulli distribution. Just as with the presentation of results for the continuous probability distribution, the results for the discrete probability distribution version are also presented in the form of three plots. The Cumulative Distribution Function plot shows that the empirical CDF and the theoretical CDF are very close to each other. The only significant difference is that the theoretical CDF starts from the point (0, 0.5), while the empirical CDF starts from the initial point (0, 0). As the curves progress, we observe slight differences. The next plot represents the Probability Mass Function, comparing the empirical PMF with the theoretical PMF. It can be noticed that the probability values are close to each other, with differences not exceeding 0.1. The last plot is the Quantile-Quantile Plot, which exhibits a linear character. It behaves almost like the line y=x. Below the plots, a table is presented, showing basic descriptive statistics such as mean, variance, skewness, and kurtosis. It can be observed that the absolute error for the theoretical value in each row of the table is significantly small, indicating an acceptable level of error.

[ ]: