Inverse transform sampling

Inverse transform sampling (also known as **inversion sampling**, the **inverse probability integral transform**, the **inverse transformation method**, **Smirnov transform**, or the **golden rule**[1]) is a basic method for pseudo-random number sampling, i.e., for generating sample numbers at random from any probability distribution given its cumulative distribution function.

Inverse transformation sampling takes uniform samples of a number u between 0 and 1, interpreted as a probability, and then returns the largest number x from the domain of the distribution $P(X)$ such that $P(-\infty < X < x) \leq u$. For example, imagine that $P(X)$ is the standard normal distribution with mean zero and standard deviation one. The table below shows samples taken from the uniform distribution and their representation on the standard normal distribution.
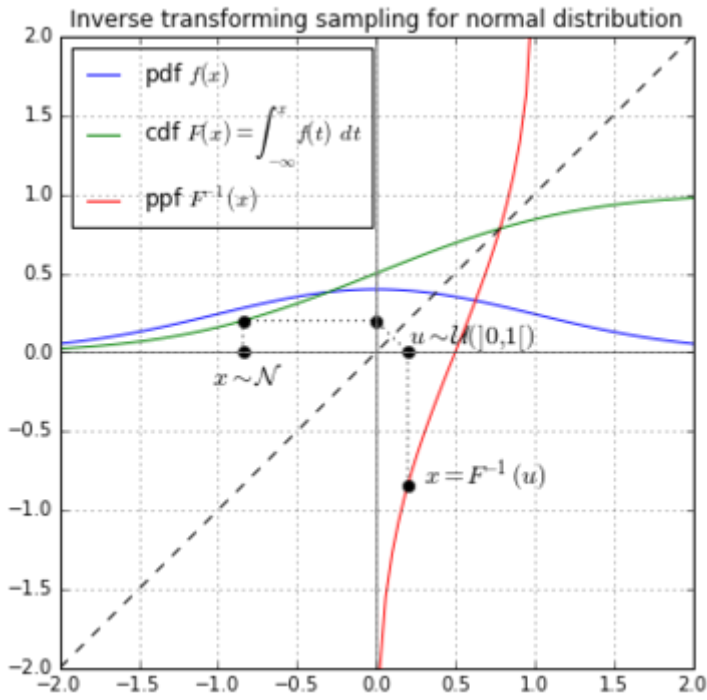
We are randomly choosing a proportion of the area under the curve and returning the number in the domain such that exactly this proportion of the area occurs to the left of that number. Intuitively, we are unlikely to choose a number in the far end of tails because there is very little area in them which would require choosing a number very close to zero or one.

Transformation from uniform sample to normal	
u	$F^{-1}(u)$
.5	0
.975	1.95996
.995	2.5758
.999999	4.75342
1-2^{-52}	8.12589

Computationally, this method involves computing the quantile function of the distribution — in other words, computing the cumulative distribution function (CDF) of the distribution (which maps a number in the domain to a probability between 0 and 1) and then inverting that function. This is the source of the term "inverse" or "inversion" in most of the names for this method. Note that for a discrete distribution, computing the CDF is not in general too difficult: we simply add up the individual probabilities for the various points of the distribution. For a continuous distribution, however, we need to integrate the probability density function (PDF) of the distribution, which is impossible to do analytically for most distributions (including the normal distribution). As a result, this method may be computationally inefficient for many distributions and other methods are preferred; however, it is a useful method for building more generally applicable samplers such as those based on rejection sampling.

Inverse transform sampling for normal distribution

For the normal distribution, the lack of an analytical expression for the corresponding quantile function means that other methods (e.g. the Box–Muller transform) may be preferred computationally. It is often the case that, even for simple distributions, the inverse transform sampling method can be improved on:[2] see, for example, the ziggurat algorithm and rejection sampling. On the other hand, it is possible to approximate the quantile function of

the normal distribution extremely accurately using moderate-degree polynomials, and in fact the method of doing this is fast enough that inversion sampling is now the default method for sampling from a normal distribution in the statistical package R.[3]

Contents

Definition

The probability integral transform states that if X is a continuous random variable with cumulative distribution function F_X, then the random variable $Y = F_X(X)$ has a uniform distribution on [0, 1]. The inverse probability integral transform is just the inverse of this: specifically, if Y has a uniform distribution on [0, 1] and if X has a cumulative distribution F_X, then the random variable $F_X^{-1}(Y)$ has the same distribution as X.

Intuition

From $U \sim \text{Unif}[0, 1]$, we want to generate X with CDF $F_X(x)$. We assume $F_X(x)$ to be a strictly increasing function, which provides good intuition.

We want to see if we can find some strictly monotone transformation $T : [0, 1] \mapsto \mathbb{R}$, such that $T(U) \overset{d}{=} X$. We will have
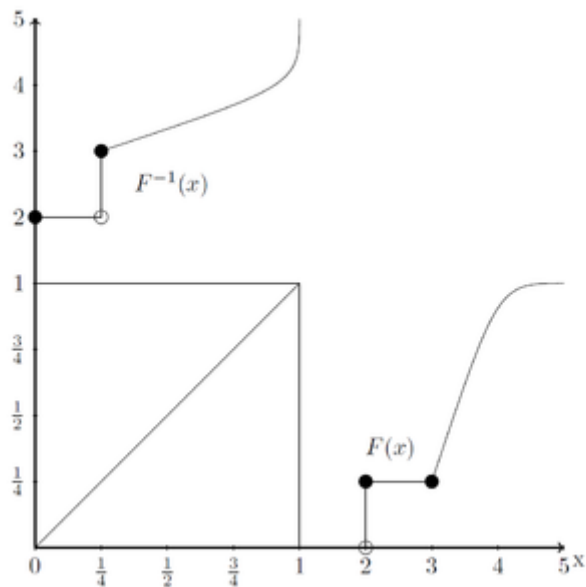
Graph of the inversion technique from x to $F(x)$. On the bottom right we see the regular function and in the top left its inversion.

$$F_X(x) = \Pr(X \leq x) = \Pr(T(U) \leq x) = \Pr(U \leq T^{-1}(x)) = T^{-1}(x), \text{ for } x \in \mathbb{R},$$

where the last step used that $\Pr(U \leq y) = y$ when U is uniform on $(0,1)$.

So we got F_X to be the inverse function of T, or, equivalently $T(u) = F_X^{-1}(u), u \in [0,1]$.

Therefore, we can generate X from $F_X^{-1}(U)$.

The method

The problem that the inverse transform sampling method solves is as follows:

- Let X be a random variable whose distribution can be described by the cumulative distribution function F_X.
- We want to generate values of X which are distributed according to this distribution.

The inverse transform sampling method works as follows:

1. Generate a random number u from the standard uniform distribution in the interval $[0,1]$, e.g. from $U \sim \text{Unif}[0,1]$.
2. Find the inverse of the desired CDF, e.g. $F_X^{-1}(x)$.
3. Compute $X = F_X^{-1}(u)$. The computed random variable X has distribution $F_X(x)$.
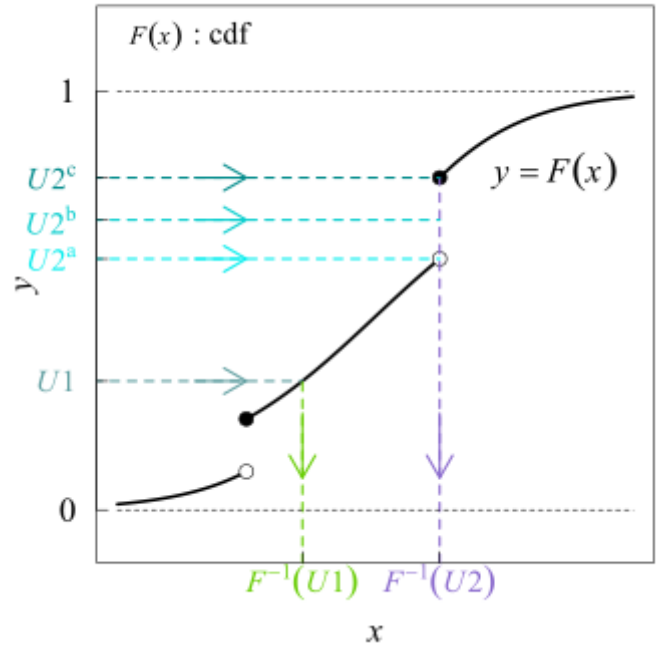
Schematic of the inverse transform sampling. The inverse function of $y = F_X(x)$ can be defined by
$F_X^{-1}(y) = \inf\{x|F_X(x) \geq y\}$.

Expressed differently, given a continuous uniform variable U in $[0,1]$ and an invertible cumulative distribution function F_X, the random variable $X = F_X^{-1}(U)$ has distribution F_X (or, X is distributed F_X).

A treatment of such inverse functions as objects satisfying differential equations can be given.[4] Some such differential equations admit explicit power series solutions, despite their non-linearity.

Examples

- As an example, suppose we have a random variable $U \sim \text{Unif}(0,1)$ and a cumulative distribution function

$$F(x) = 1 - \exp(-\sqrt{x})$$
In order to perform an inversion we want to solve for $F(F^{-1}(u)) = u$
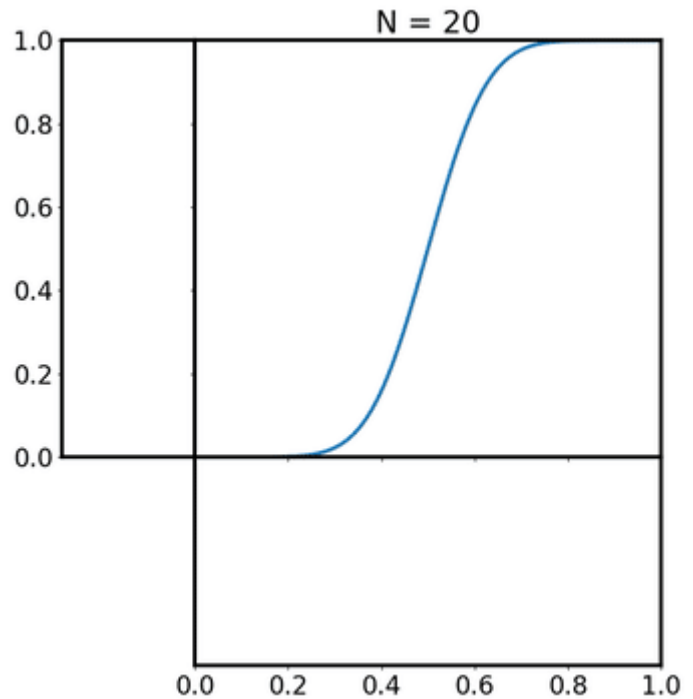
An animation of how inverse transform sampling generates normally distributed random values from uniformly distributed random values

$$F(F^{-1}(u)) = u$$
$$1 - \exp\left(-\sqrt{F^{-1}(u)}\right) = u$$
$$F^{-1}(u) = (-\log(1-u))^2$$
$$= (\log(1-u))^2$$

From here we would perform steps one, two and three.

- As another example, we use the <u>exponential distribution</u> with $F_X(x) = 1 - e^{-\lambda x}$ for x ≥ 0 (and 0 otherwise). By solving y=F(x) we obtain the inverse function
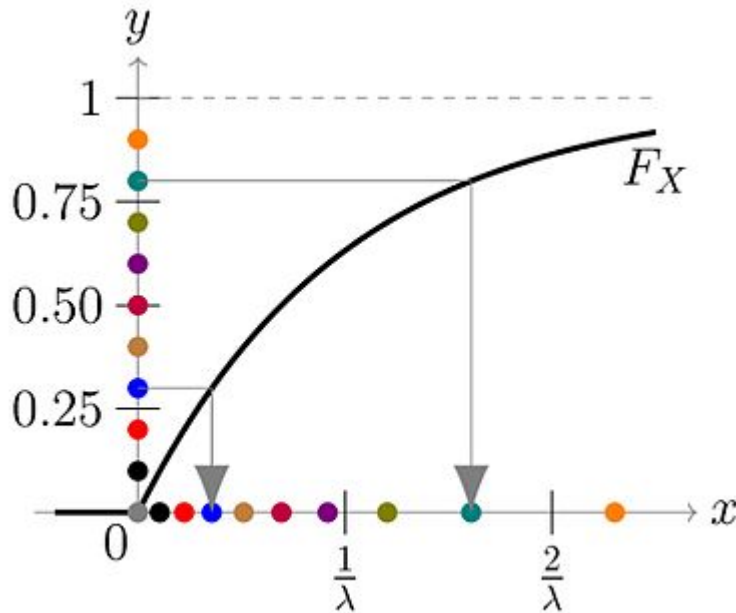
$$x = F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y).$$

It means that if we draw some y_0 from a $U \sim \text{Unif}(0,1)$ and compute $x_0 = F_X^{-1}(y_0) = -\frac{1}{\lambda}\ln(1-y_0)$, This x_0 has exponential distribution.

The idea is illustrated in the following graph:

Random numbers y_i are generated from a uniform distribution between 0 and 1, i.e. Y ~ U(0, 1). They are sketched as colored points on the y-axis. Each of the points is mapped according to $x=F^{-1}(y)$, which is shown with gray arrows for two example points. In this example, we have used an exponential distribution. Hence, for x ≥ 0, the probability density is $\varrho_X(x) = \lambda e^{-\lambda x}$ and the cumulative distribution function is $F(x) = 1 - e^{-\lambda x}$. Therefore, $x = F^{-1}(y) = -\dfrac{\ln(1 - y)}{\lambda}$. We can see that using this method, many points end up close to 0 and only few points end up having high x-values - just as it is expected for an exponential distribution.

Note that the distribution does not change if we start with 1-y instead of y. For computational purposes, it therefore suffices to generate random numbers y in [0, 1] and then simply calculate

$$x = F^{-1}(y) = -\frac{1}{\lambda}\ln(y).$$

Proof of correctness

Let F be a continuous cumulative distribution function, and let F^{-1} be its inverse function (using the infimum because CDFs are weakly monotonic and right-continuous):[5]

$$F^{-1}(u) = \inf\{x \mid F(x) \geq u\} \qquad (0 < u < 1).$$

Claim: If U is a uniform random variable on (0, 1) then $F^{-1}(U)$ has F as its CDF.

Proof:

$$
\begin{aligned}
&\Pr(F^{-1}(U) \leq x)\\
&= \Pr(U \leq F(x)) && \text{(applying } F, \text{ to both sides)}\\
&= F(x) && \text{(because } \Pr(U \leq y) = y, \text{ when U is uniform on}(0,1))
\end{aligned}
$$

Truncated distribution

Inverse transform sampling can be simply extended to cases of <u>truncated distributions</u> on the interval $(a, b]$ without the cost of rejection sampling: the same algorithm can be followed, but instead of generating a random number u uniformly distributed between 0 and 1, generate u uniformly distributed between $F(a)$ and $F(b)$, and then again take $F^{-1}(u)$.

Reduction of the number of inversions

In order to obtain a large number of samples, one needs to perform the same number of inversions of the distribution. One possible way to reduce the number of inversions while obtaining a large number of samples is the application of the so-called Stochastic Collocation Monte Carlo sampler (SCMC sampler) within a <u>polynomial chaos</u> expansion framework. This allows us to generate any number of Monte Carlo samples with only a few inversions of the original distribution with independent samples of a variable for which the inversions are analytically available, for example the standard normal variable.[6]

See also

- <u>Probability integral transform</u>
- <u>Copula</u>, defined by means of probability integral transform.
- <u>Quantile function</u>, for the explicit construction of inverse CDFs.
- <u>Inverse distribution function</u> for a precise mathematical definition for distributions with discrete components.

References

1. Aalto University, N. Hyvönen, Computational methods in inverse problems. Twelfth lecture https://noppa.tkk.fi/noppa/kurssi/mat-1.3626/luennot/Mat-1_3626_lecture12.pdf
2. Luc Devroye (1986). *Non-Uniform Random Variate Generation* (http://www.eirene.de/Devroye.pdf) (PDF). New York: Springer-Verlag.
3. "R: Random Number Generation" (https://stat.ethz.ch/R-manual/R-devel/library/base/html/Random.html).
4. Steinbrecher, G., Shaw, W.T. (2008). Quantile mechanics. *European Journal of Applied Mathematics* 19 (2): 87–112.
5. Luc Devroye (1986). "Section 2.2. Inversion by numerical solution of *F*(*X*) = *U*" (http://luc.devroye.org/chapter_two.pdf) (PDF). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.
6. L.A. Grzelak, J.A.S. Witteveen, M. Suarez, and C.W. Oosterlee. The stochastic collocation Monte Carlo sampler: Highly efficient sampling from "expensive" distributions. https://ssrn.com/abstract=2529691