In this post I want to talk about one of the applications of the probability integral transform which states that it is possible to transform any given distribution into a standard uniform distribution.

The inverse transform sampling is that application which I got interested in due to a friend who asked me about it, it requires that we are able to compute the quantile function of any probability distribution X, then we can generate data samples of X using a uniformly distributed random variable $U \sim [0, 1]$

Let us set the stage for building intuition, by detailing what is happening in these methods. I will first discuss the case when X takes on continuous real values
(the observation space is $\mathbb{R}$) and we later we can tackle the discrete case.

Infinite possible outcomes

Let X be a random variable defined as a mapping from the sample space to the real number line.

$$X : \Omega \to \mathbb{R}$$

Let this probability distribution have the **probability density function (pdf)** $f_X(x)$, such that the probability that outcomes corresponding to $[a, b]$ happen is,

$$\Pr[a \leq X \leq b] = \int_a^b f_X(t)\, dt$$

and the **cumulative distribution function (cdf)** $F_X(x)$

$$F_X(x) = \int_{-\infty}^x f_X(u)\, du$$

We wish to find a function g such that we can transform the random variable X into U. The following are two proofs that that function g is F_X.

Invariant probability mass

Any function of a random variable is itself a *derived* random variable,

$$Y = g(X)$$

So now we have a mapping $X : \Omega \to x$ from outcomes to real numbers x and further a mapping $g : x \to y$ from each x to $y = g(x)$

So the new derived random variable $Y : \Omega \to y$ is clearly a random variable with pdf h being the composition of f and g^{-1}. It makes sense that the probability of any outcome happening remains invariant under the transformation, since both random variables are simple different ways of talking about the same underlying experiment. So the probability mass is invariant under the change of variables.

$$|f(x)dx| = |f(g^{-1}(y))dy| = |h(y)dy|$$

Since pdf is always positive, we have the following relationship between the two pdfs.

$$h(y)\left|\frac{dy}{dx}\right| = f(x)$$

As a side note, as long as g is reasonably well behaved, i.e. in order for h to also be a pdf, g needs to be monotonic.

So our object is to transform X into U. So we want to find g such that $Y = U$.

> It is very common to start with a distribution which is Uniform(0,1) which is to say that the probability density function $f(x)$ is:
>
> $$f(x) = \begin{cases} 1, & \text{if } 0 \leqslant x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Since the pdf of U is 1, $h(y) = 1$ we find that,

$$f(x) = \left|\frac{dy}{dx}\right|$$

So the cdf of X will be,

$$F_X(x) = \int_{-\infty}^{x}\left|\frac{dy}{dx}\right|dx = y(x) - y(-\infty) = y(x)$$

comparing this with $y = g(x)$, we note that F_X is the function which transforms rv X to U and therefore F_X^{-1} can transform U to X.

CDF of Y

We have $y = g(x)$ and $x = g^{-1}(y)$. Consider the cdf of Y,

We can write probabilities in either in X or Y domain, change of variables does not change probability.

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

if $g = F_X$ then note how $F_Y(y) = y$

We get the $y = x$ that is the cdf of the uniform distribution. If we find the pdf,

$$h(y) = \frac{d}{dy} F_Y(y) = 1$$

It turns out to be 1, the pdf of the $U \sim [0, 1]$

Thus once again we can see how F_X is the function which transforms rv X to U and F_X^{-1} can transform U to X.

The first thing which I personally struggled with is the idea of a derived random variable $Y = F_X(X)$ formed due to the following mappings

$$F_X : X \rightarrow Y$$

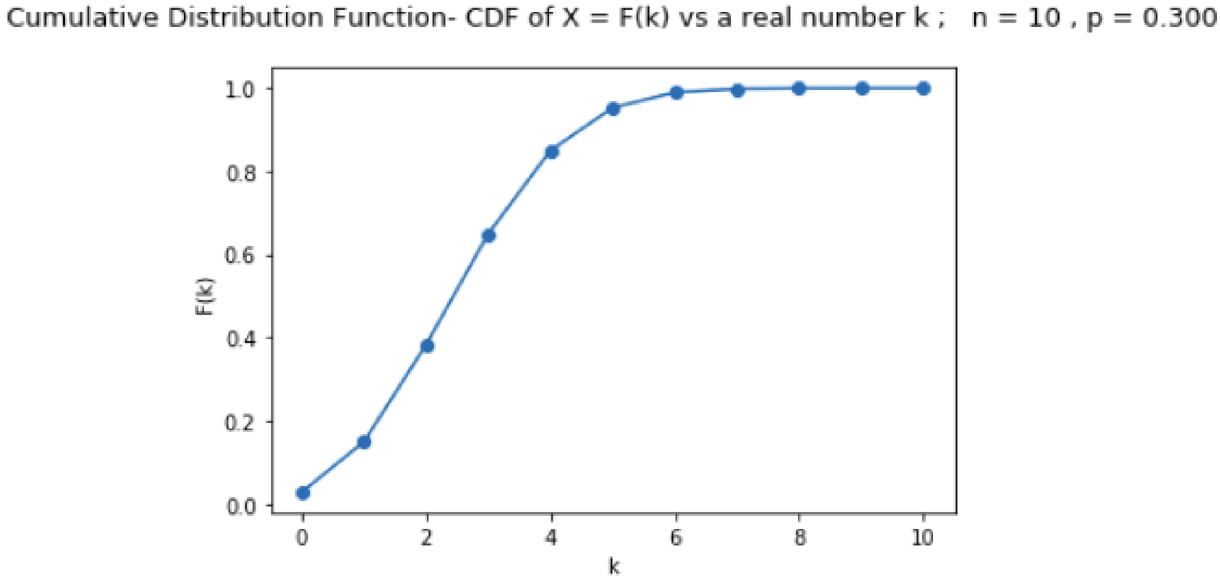
$$X : \Omega \rightarrow \mathbb{R}$$

$$F_X : \mathbb{R} \rightarrow [0, 1]$$

$$Y : \Omega \xrightarrow{\text{X}} \mathbb{R} \xrightarrow{F_X} [0, 1]$$

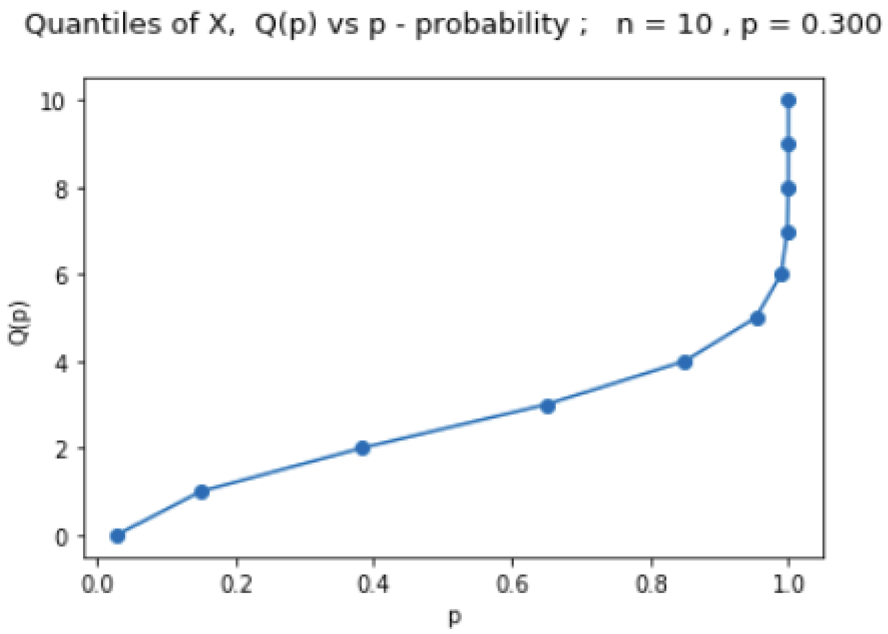
So every $x \in X$ will correspond to some $y \in Y$. So instead of using all real numbers we can use probabilities to index all possible outcomes. This is a change of variable, a different perspective of the same probability distribution.

Say we take n samples drawn from the probability distribution of X. We can visualize $F_X(x) \times n$ to be the number of samples we can expect to find in those n samples corresponding to those outcomes whose random variable values lie below x.



Consider the example of the cdf of the binomial distribution. Here the random variable k is the number of heads when we toss $n = 10$ biased coins. Note how $F(2) \approx 0.4$. This means that if we conduct the 10 coin experiment 100 times, we can expect to see $0, 1$ or 2 heads 40 times. Since $F(6) \approx 0.98$ we can expect to see $7, 8, 9$ or 10 heads only $100 - 98 = 2$ times. This makes sense since the coin is biased with $P(heads) = 0.3$.

So $F_X(x) = y$ is the *fraction* of samples expected to fall below x. With this understanding the quantile function $F_X^{-1}(y)$ takes on new meaning. Taking in a probability y_o and returning a random variable $x_o = F_X^{-1}(y_o)$, means that below x_o lies y_o fraction of the values.




```
[ 0.  1.  2.  3.  4.  5.  6.  7.  8.  9. 10.]
[0.02824752 0.14930835 0.38278279 0.64961072 0.84973167 0.95265101
 0.98940792 0.99840961 0.99985631 0.9999941  1.        ]
```

So asking below what x_o can we expect to see 0.149 fraction of samples will obviously give 1 since we computed 0.149 as that fraction when we computed F_X. Here we are basically composing the two functions.

$$Q(p) = F_X^{-1}(y_o) = F_X^{-1}(F_X(x_o)) = x_o$$

Thus these two transformations are inverses of each other under function composition. While the uniform distribution is the identity element much like in the algebra of numbers.

We can think of the Y vs X axes switching by rotation around $y = x$ line in the cdf vs rv graph (F_X vs x) and thus forming the quantile function vs probability ($F_X^{-1}(y) = x$ vs y = F_X)

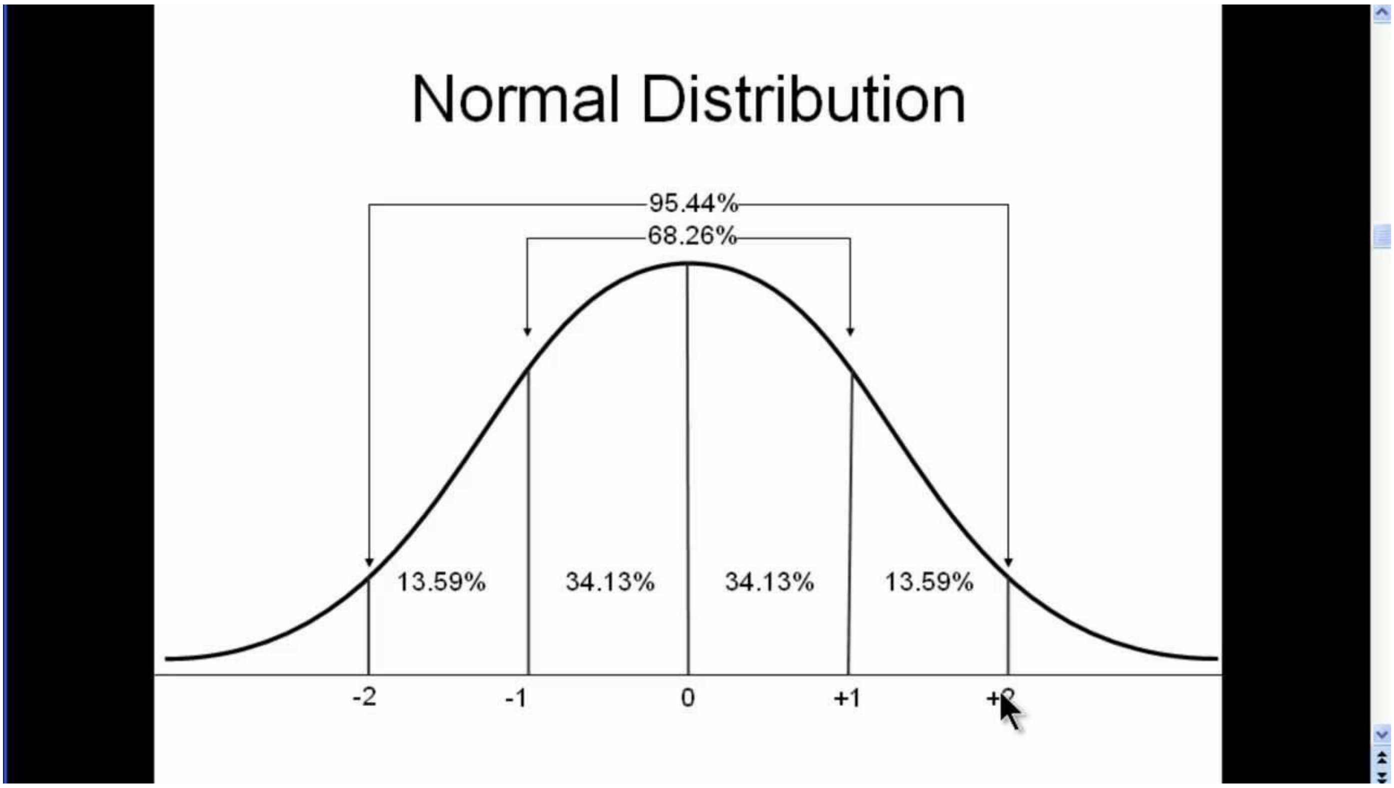
Let us take an example,

Uniform to Normal

Consider the example of turning uniform samples into normally distributed samples using the Quantile function of the Normal distribution - the probit function.

$$X = F_X^{-1}(Y)$$

So given a random probability y, the probit function F_X^{-1} returns a real number x such that the fraction of the area lying to the left of x is y.



Now since X is normally distributed, the fraction of area to the left changes rapidly near the mean. The numbers at both ends of the real line correspond to nearly 0 and 1 probability and there is negligible change as you go further outward.

So even if Y is uniformly distributed, by asking this specific question, we see that by using the probit function, we mostly get values near the mean. Thus by sampling the quantile function uniformly we get data samples distributed as per X.

This is because the Quantile function just like the CDF or PDF or the characteristic function fully and uniquely defines any probability distribution.

Normal to Uniform

Say we have a collection of data samples from from the normal distribution. Most of the x lie near the mean and a few scattered out in both directions. We create a collection of y using,

$$y = F_X(x)$$

The normal distribution CDF Φ is almost a straight line near the mean while it is mostly 0 and 1 towards either ends of the real line. Due to how x is distributed, y rarely takes extreme probabilities and is evenly spread across $[0, 1]$.

Concerns

The lack of an analytical expression for the quantile functions of many continuous probability distributions and the fact it is hard to generalize this method to higher dimensions are reasons for the lack of popularity for this method.

> Written with StackEdit.