07.Advanced DIstributions

The Gamma distribution is an extension of the Erlang distribution where the number of stages can be any positive real number. It is called in this way because it uses the Gamma function in its PDF.

$$f_{Gamma<\theta,k>}(t) = \frac{t^{k-1}e^{-t/\theta}}{\theta^k \Gamma(k)}$$

The Gamma distribution has the following mean and variance:

$$\mu_{Gamma<\theta,k>} = k\theta, \sigma^2_{Gamma<\theta,k>} = k\theta^2$$

Although very complex, it still has analytical results available: for this reason it is sometimes used to obtain random variables with a given mean and c.v:

$$k = \frac{1}{\sqrt{c_v}}, \theta = \frac{E[X]}{k}$$

Weibull

The Weibull distribution is characterized by two parameters: the shape k and the scale l. It similar to an exponential distribution, but its exponent is raised to the power of k

$$f_{Weibull<\lambda,k>}(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k}$$
$$F_{Weibull<\lambda,k>}(t) = 1 - e^{-\left(\frac{t}{\lambda}\right)^k}$$

It is has the following mean, variance and moments:

$$\mu_{Weibull<\lambda,k>} = \lambda\Gamma\left(1 + \frac{1}{k}\right)$$
$${\sigma^2}_{Weibull<\lambda,k>} = \lambda^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right]$$
$$E[X^n] = \lambda^n\Gamma\left(1 + \frac{n}{k}\right)$$

It is used for modeling the lifetime of components subject to aging.

Pareto

It is a time raised by a given value

The Pareto distribution has a simple monomial expression that is characterized by a shape parameter α and a scale parameter m:

$$f_{Pareto<\alpha,m>}(t) = \begin{cases} \dfrac{\alpha m^\alpha}{t^{\alpha+1}} & t \geq m \\ 0 & t < m \end{cases}$$

$$F_{Pareto<\alpha,m>}(t) = \begin{cases} 1 - \left(\dfrac{m}{t}\right)^\alpha & t \geq m \\ 0 & t < m \end{cases}$$

Since the integral of a PDF must equal to 1, the Pareto is a proper distribution only for α > 0, and it generates values in the range [m, ∞].

$$\int_m^\infty \frac{\alpha m^\alpha}{x^{\alpha+1}} dx = \alpha m^\alpha \int_m^\infty x^{-\alpha-1} dx = -\frac{\alpha m^\alpha}{\alpha} \left(x^{-\alpha} \right|_m^\infty$$

$$= \begin{cases} \infty & \alpha \leq 0 \\ \dfrac{\alpha m^\alpha}{\alpha} m^{-\alpha} = 1 & \alpha > 0 \end{cases}$$

The distribution has a finite mean only for α > 1.

$$E\left[X_{Pareto<\alpha,m>}\right] = \int_m^\infty x \frac{\alpha m^\alpha}{x^{\alpha+1}} dx = \alpha m^\alpha \int_m^\infty x^{-\alpha} dx = \frac{\alpha m^\alpha}{1-\alpha} \left(x^{1-\alpha} \right|_m^\infty$$

$$\mu_{Pareto<\alpha,m>} = \begin{cases} \infty & \alpha \leq 1 \\ \dfrac{\alpha m}{\alpha - 1} & \alpha > 1 \end{cases}$$

t has finite mean, but infinite variance for 1 < α <= 2. The distribution has finite variance only for α > 2.

$$E[X^2{}_{Pareto<\alpha,m>}] = \int_m^{\infty} x^2 \frac{\alpha m^\alpha}{x^{\alpha+1}} dx = \alpha m^\alpha \int_m^{\infty} x^{1-\alpha} dx = \frac{\alpha m^\alpha}{2-\alpha} (x^{2-\alpha}|_m^{\infty}$$

$$E[X^2{}_{Pareto<\alpha,m>}] = \begin{cases} \infty & \alpha \leq 2 \\ \frac{\alpha m^2}{\alpha-2} & \alpha > 2 \end{cases} \qquad \sigma^2{}_{Pareto<\alpha,m>} = \begin{cases} \infty & \alpha \leq 2 \\ \frac{\alpha m^2}{(\alpha-1)^2(\alpha-2)} & \alpha > 2 \end{cases}$$

We are interested in it as its expression is very simple, but in many case the length of the request follow a Pareto distribution. It cannot generate values smaller than a certain values called shape of the distribution.
It is used to model systems where very large values can have a non-negligible probability to occur. This phenomenon is called Heavy tail, and it can shown plotting 1-F(x) in log-log-scale.

Normal

The Normal (or Gaussian) distribution is defined by its mean and its variance.

$$f_{N<\mu,\sigma^2>}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

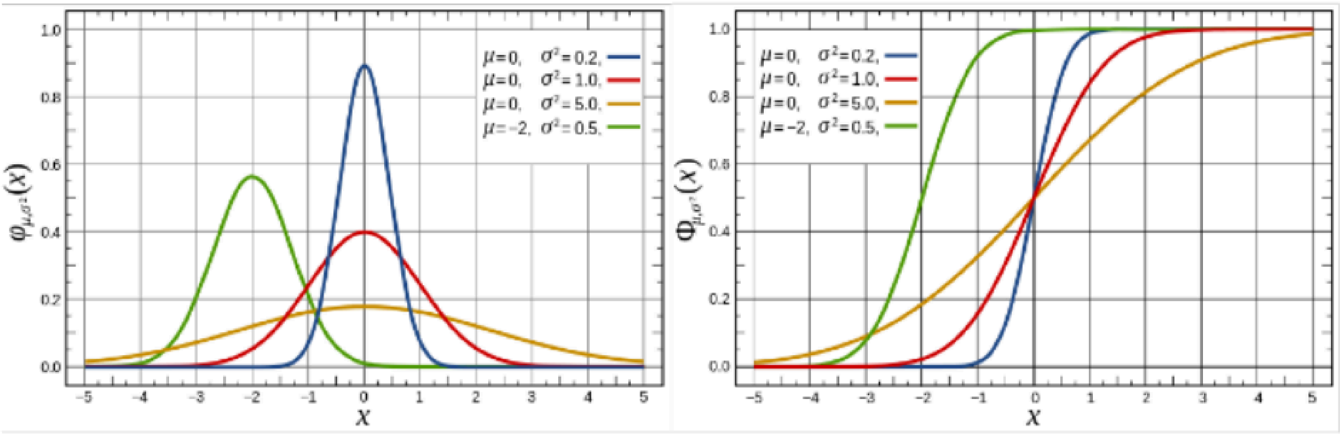
$$F_{N<\mu,\sigma^2>}(t) = \frac{1}{2}\left[1 + erf\left(\frac{t-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$\mu_{N<\mu,\sigma^2>} = \mu \qquad \sigma^2{}_{N<\mu,\sigma^2>} = \sigma^2$$

It is very important since, for the central limit theorem, the sum of a large number of independent random variables tends to a Normal distribution

$$\sum_i X_i \cong X_{N<\sum_i \mu_i, \sum_i \sigma^2{}_i>}$$

It has the following PDF and CDF:



When $\mu = 0$ and σ^2=1, the distribution is called Standard Normal. If $N_{<0,1>}$ is a Standard Normal distribution, then we can obtain normal distributions with different average and variance with a simple linear transform:

$$X_{N<\mu,\sigma^2>} = \mu + \sigma X_{N_<0,1>}$$

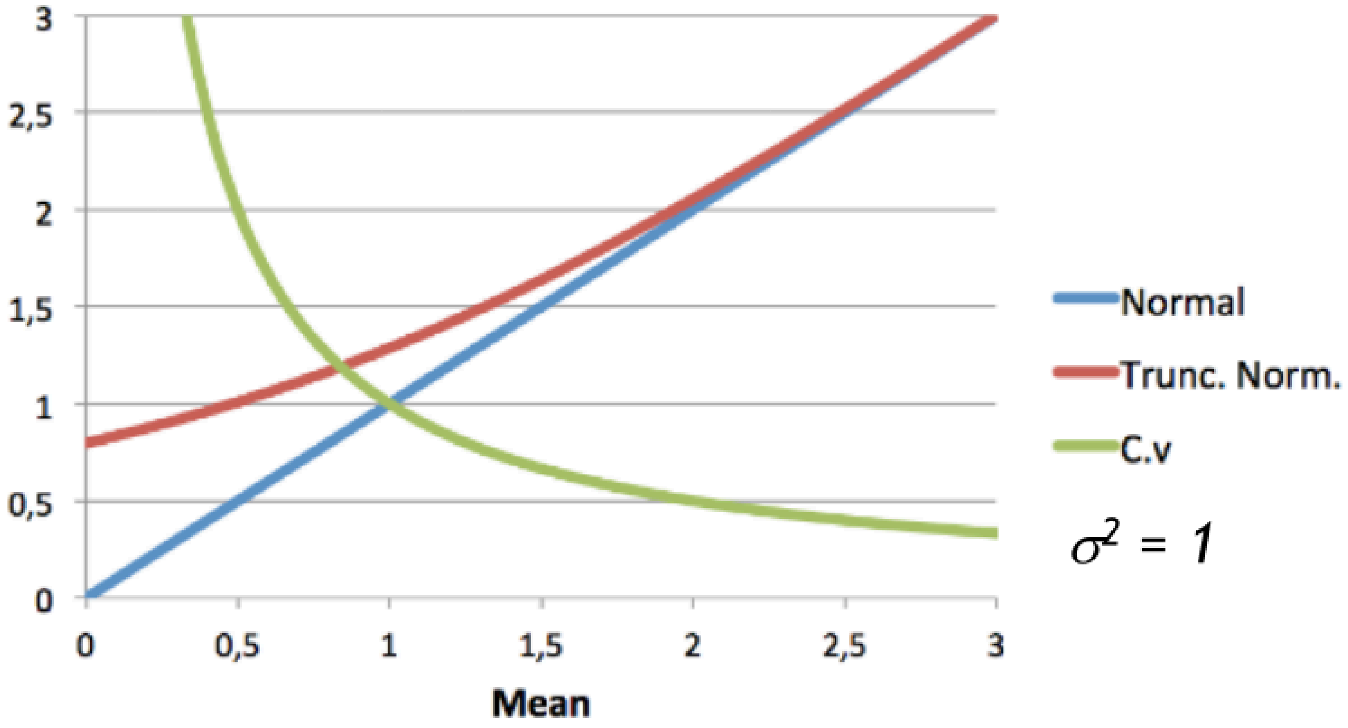
Normal distribution can also generate negative number so we need to truncate all of this negative values as we cannot have them in performance evaluation.

Truncated normal

The Truncated Normal distribution discards all the values that are negative when sampling a conventional Normal distribution:

$$f_{TruncN<\mu,\sigma''>}(t) = \frac{f_{N_{<\mu\sigma^2>}}(t)}{1 - F_{N_{\mu,\sigma^2}}(0)}$$

The truncated distribution, however, has different mean and variance with respect to the original Normal distribution. It can still be a good approximation if the c.v. is not too large, since the difference between the two values can be negligible.



Log Normal

The Log Normal distribution, is computed by using a Normal distribution sample (with a given average and standard deviation) as the exponent for the exponential function

$$X_{LogNormal<\mu,\sigma^2} = e^{X_{N<\mu,\sigma^2}}$$

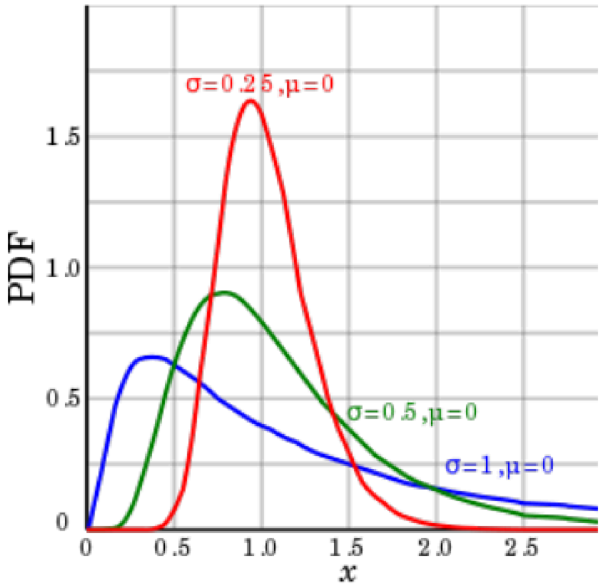
Thanks to the properties of the exponential, it produces only positive numbers. Its PDF, CDF, average and variance are:

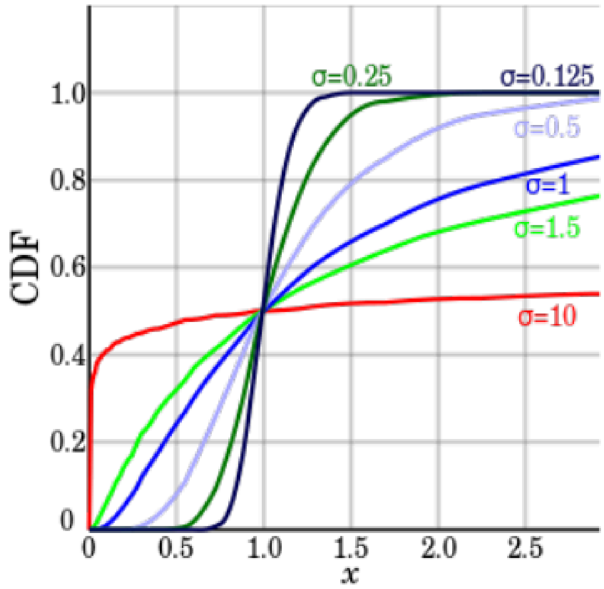
$$f_{LogNormal<\mu,\sigma^2>}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}$$

$$F_{LogNormal<\mu,\sigma^2>}(t) = \frac{1}{2}\left[1 + erf\left(\frac{\ln(t) - \mu}{\sigma\sqrt{2}}\right)\right]$$

$$\mu_{LogN<\mu,\sigma^2>} = e^{\mu+\frac{\sigma^2}{2}} \quad {\sigma^2}_{LogN<\mu,\sigma^2>} = \left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2}$$

It has the following shape:





Given a target average *E[X]* and coefficient of variation c_v, its parameters can be computed in the following way:

$$\mu = ln\left(\frac{E[X]}{\sqrt{1+c_v^2}}\right) \qquad \sigma^2 = \sqrt{ln(1+c_v^2)}$$

Random numbers can then be generated by simply applying its definition, starting from a normally distributed random value (generated for example with the Box-Muller technique).

x_i = exp(*GenNormalRandomSample*(μ, σ^2))