Key Concepts and Terminology

Piecewise function: Defined by two or more equations with each equation evaluated within a given domain. Two popular ones are the absolute value function and step function.

Discrete function: A function defined by a single equation within a certain domain. A discrete function is specified by a one-to-one mapping or many to one mapping but NEVER a one-to-many mapping.

Probability Density function (PDF): A function (univariate or multivariate) whose value within a given domain or sample space provides the likelihood that the continuous or discrete random variable falls within that range or outside that range. The most common of PDFs is the normal PDF (normal distributions) given by a parametrized multivariate function. The Lognormal and Weibull PDFs are other functions that can be used to model different shapes.

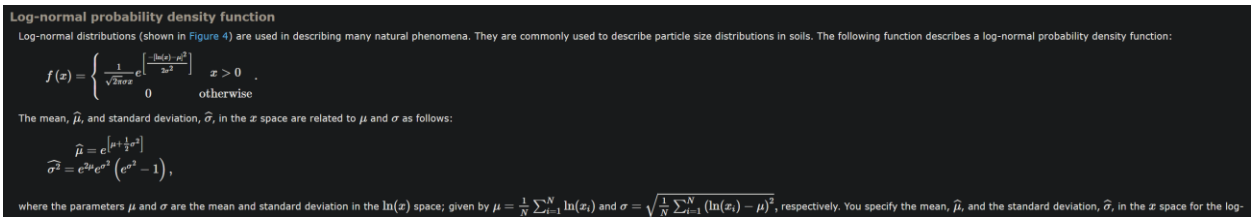
$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} * e^{[-\frac{(x-\mu)^2}{2\sigma^2}]}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

Probability Mass Function (PMF): If the random variable can only take on specific values like rolling a dice, then a PMF is used instead of a PDF.

Lognormal PDF:



Log-normal probability density function

Log-normal distributions (shown in Figure 4) are used in describing many natural phenomena. They are commonly used to describe particle size distributions in soils. The following function describes a log-normal probability density function:

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{\left[\frac{-[\ln(x)-\mu]^2}{2\sigma^2}\right]} & x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

The mean, $\widehat{\mu}$, and standard deviation, $\widehat{\sigma}$, in the x space are related to μ and σ as follows:

$$\widehat{\mu} = e^{\left[\mu + \frac{1}{2}\sigma^2\right]}$$
$$\widehat{\sigma^2} = e^{2\mu} e^{\sigma^2}\left(e^{\sigma^2} - 1\right),$$

where the parameters μ and σ are the mean and standard deviation in the $\ln(x)$ space; given by $\mu = \frac{1}{N}\sum_{i=1}^{N} \ln(x_i)$ and $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\ln(x_i) - \mu)^2}$, respectively. You specify the mean, $\widehat{\mu}$, and the standard deviation, $\widehat{\sigma}$, in the x space for the log-

Cumulative Distribution Function (CDF): A CDF is derived from integrating a PDF of a continuos random variable. Missing constants within a non-parametrized PDF are derived by integrating the PDF equal 1. This is also a part to mention self-starter functions which needs more research.

$$1 = \int_{-\infty}^{\infty} \text{ɣ}.f(x)$$

Key Concepts and Terminology

P-P Plot: A plot of the expected or theoretical vs observed or empirical CDF distributions which is assessed vs a normalized line where both distributions are equal. Data transformations are often capable of inducing normality.

Q-Q Plot: An assessment of the similarity of the quantile function distributions derived for the theoretical vs observed continuous or discrete variable.

Skewness: Measures the asymmetry of a probability distribution; Right or Left skewed.

Kurtosis: A measure that depicts the shape of a distribution's tails (outliers) in comparison to the overall shape. A measure of "tailedness" not "peakness". Leptokurtic (>3/0), mesokurtic (==3/0) and platykurtic (<3/0). A platykurtic distro is an indication of low volatility.

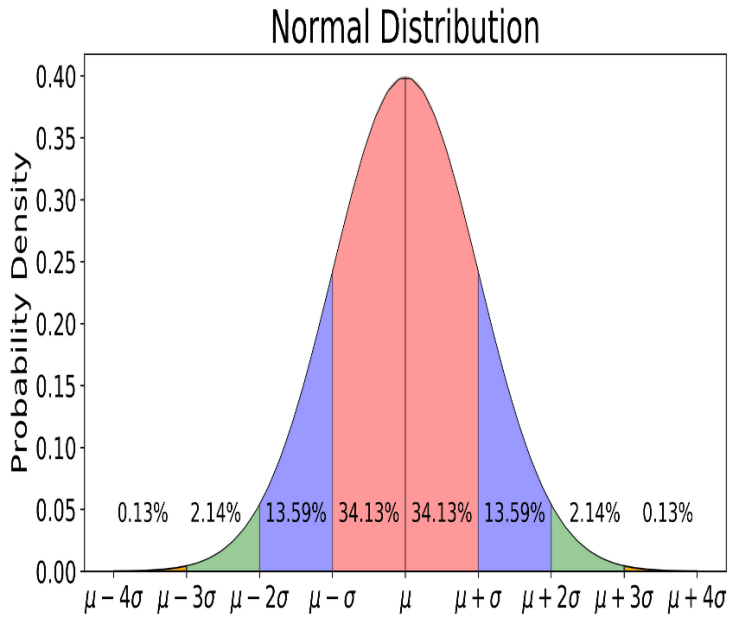
Shapiro-wilk (S-W) Test: A test for assessment of normality for a variable. Is the more powerful test and better for small sample sizes. The Null hypothesis states normality of the data meaning $p < 0.01$ indicates rejection of the null meaning the data is NOT normal.

Kolmogrov-smirnov (K-S) Test: A test for assessment of normality for a variable. Is better for larger sample sizes. Low power and sensitivity to extreme data points.

Paramatric Analysis: Statistical analysis based on the assumption that sampled data can be modelled to a PDF distribution defined by fixed parameters (mean and standard deviation). Examples are modelling data after a normal PDF or Log-normal PDF. Here one comes to the realization that the alpha level represents a percentile and probability simultaneously.

Key Concepts and Terminology

Normal Distribution:



Log-normal Distribution: