1 Overview

This library implements a number of functions for mathematical operations, it has dependencies upon the scala-nlp-breeze matrix library, and its purpose is to explore the implementation and useage of a variety of tools and techniques starting with some statistical tools. This document is structured with the main headings corresponding to the subject areas implemented in the library, and minor headings corresponding to the major package names in the project. Each subsection below the minor headings is also decomposed into individual subject areas, and the lowest level heading represents a class.

2 Methods of Counting

2.1 package au.id.cxd.math.count

The count package contains a series of modules dedicated to methods of counting.

2.1.1 Factorial

The factorial operation is provided as $n!$ implementing:

$$\prod_{i=1}^{n-1}(n-i)$$

The Factorial implementation will memoize results, allowing for efficient reuse during runtime.

2.1.2 Choose

The choose module implements $\binom{n}{m}$, how many ways can m items be selected with replacement from a set of n items.
Determined as:

$$\frac{n!}{m!n!}$$

2.1.3 Permutation

The method of selecting m ordered items from a set of n ordered items $P\binom{n}{m}$.

$$\frac{n!}{(n-m)!}$$

3 Probability

3.1 package au.id.cxd.math.probability

This package provides a series of modules that support operations for inference via probability, and for estimation of distributions.

3.1.1 Inequalities

The class TchebysheffInequality implements a simple estimation of a pdf using the inequality rule:

$$P(\mu - k\sigma < Y < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Which can be restated as:

$$P(lower < Y < upper) \geq 1 - \frac{1}{k^2}$$

The value of k is derived from either upper and lower bounds since:

$$k = \frac{upper - \mu}{\sigma}$$

After determining the value of k the probability can be estimated by substituting

$$p = 1 - \frac{1}{k^2}$$

3.1.2 Discrete Distributions

The discrete distributions packages contains the following.

Binomial

The binomial distribution (class name Binomial) has the parameters p for the prior proportion of successes and n for the total number of trials and calculates the probability of y successes

$$P(y;n;p) = \sum_{i=1}^{n} \binom{n}{y_i} p_i^y (1-p)^{n-y_i}$$

The properties of the binomial are:

Mean: $\mu = np$
Variance: $\sigma^2 = np(1-p) = npq$

Geometric Distribution

The geometric distribution (class name Geometric) with 1 parameter for probability p and $y \geq 1$. The variable represents the nth trial where the success occurs (for instance if y=2 then the trial was successful on the 2nd attempt). The parameter p represents the probability of success. The probability function calculates the probability of success at the nth trial.

$$P(y;p) = p(1-p)^{y-1}$$

The simple properties of the distribution are:

Mean: $\mu = \frac{1}{p}$
Variance: $\sigma^2 = \frac{1-p}{p^2}$

Hyper Geometric Distribution

The Hypergeometric (class name HyperGeometric) distribution represents the probability of choosing y number of events of the same kind from a subset of r like items within a population of all N possible items (of different kinds) for the sample of size n containing the mixed items. The constraints are such that $r \leq n \leq N$ and $y \leq r \leq n$. The parameters are y, r, n, N. The probability distribution is defined as follows.

$$P(y; r, n, N) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

The simple properties of the distribution are:

Mean: $\mu = \frac{nr}{N}$
Variance: $\sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Negative Binomial Distribution

The Negative Binomial Distribution (class name NegativeBinomial) provides the probability of the nth success or potentially nth failure of a bernoulli trial. The parameters are r representing the (r -1) initial trials that where the successful and y the total number of trials before the next success r occurs. The distribution is calculated as follows:

$$P(y; r) = \binom{y-1}{r-1} p^r q^{y-r}$$

where $y = r, r+1, \ldots$

The simple properties of the distribution are:

Mean: $\mu = \frac{r}{p}$
Variance: $\sigma^2 = \frac{r(1-p)}{p^2}$

The Poisson Distribution

The Poisson Distribution (class name Poisson) provides the probability of an event occurring a certain number of times within an interval. It is commonly used to model a number of events occuring in a certain period of time. We can use the two parameters lambda and y to represent the number of events, and period of time respectively. The distribution is defined as:

$$P(y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where $y \geq 0$ and $\lambda > 0$
The simple properties of the distribution are:

Mean: $\mu = \lambda$
Variance: $\sigma^2 = \lambda$

3.1.3 Continuous Distributions

3.1.4 Inference

3.1.5 Regression