Probability Distributions

A probability distribution is a table or an equation that links each outcome (or range of outcomes) of a statistical experiment with its probability of occurrence.

Two Families

- Discrete Distributions (i.e. Count Variables - Integers)
- Continuous Distributions (i.e. Continuous Variables - Real Numbers)

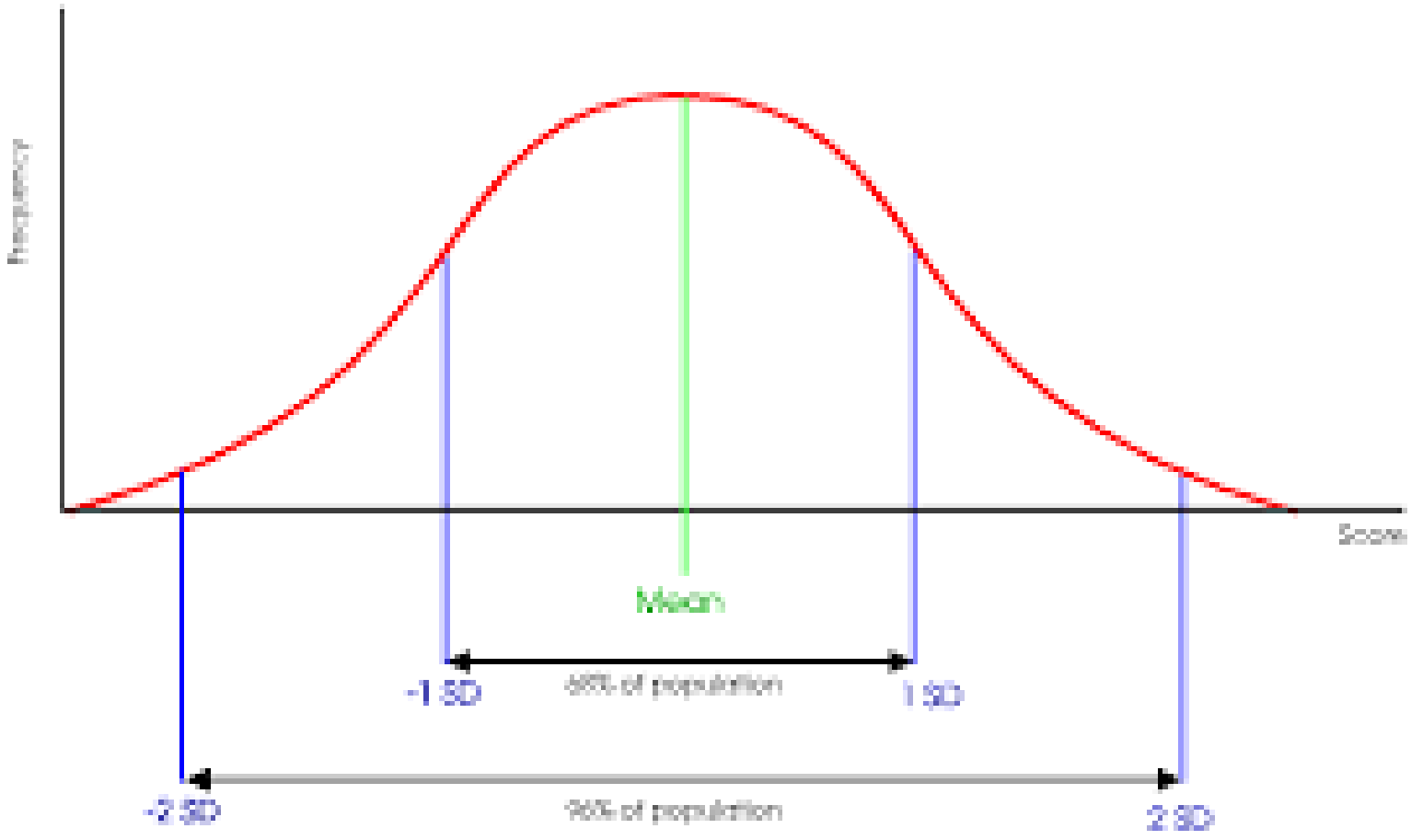
Discrete distributions

- Benford Distribution
- Bernouilli distribution
- Binomial distribution
- Hypergeometric distribution
- Geometric distribution
- Multinomial distribution
- Negative binomial distribution
- Poisson distribution
- Zipf's law

Continuous distributions

- Beta and Dirichlet distributions
- Cauchy distribution
- Chi Square distribution
- Exponential distribution
- Fisher-Snedecor distribution
- Gamma distribution
- Levy distribution
- Log-normal distribution
- Normal and related distributions
- Pareto Distributions
- Student's t distribution
- Uniform distribution
- Weibull distribution
- Extreme values and related distribution
- Distribution in circular statistics

The Normal Distribution Curve



The Normal Distribution

Normal Probability Distribution

- Cornerstone of every undergraduate statistics module.
- Basis of a substantial body of statistical theory .
- Central Limit Theorem - basis of Statistical Inference (i.e. Hypothesis Tesing, Confidence Intervals).

Compendium of Probability Distributions

Compendium of Probability Distributions

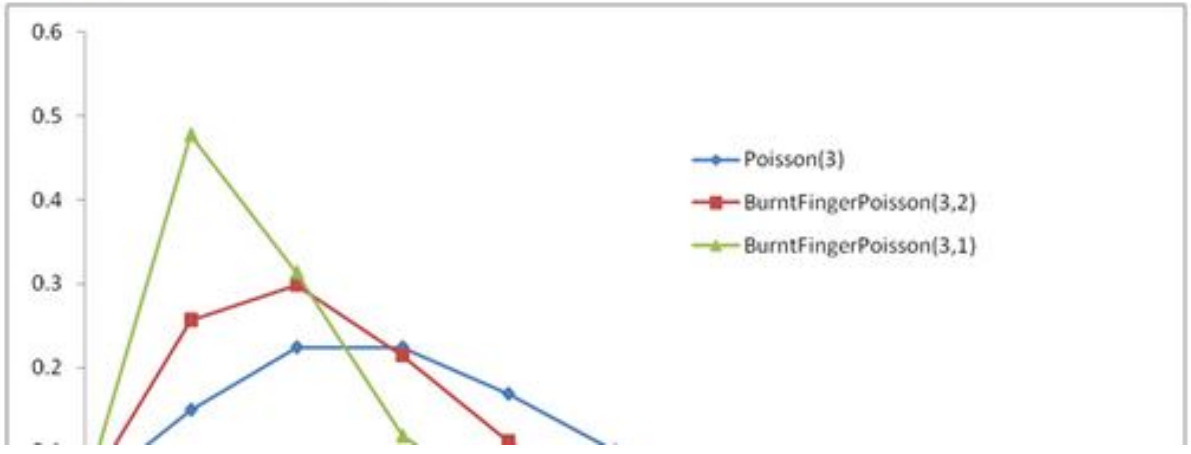
Burnt Finger Poisson Distribution

See also: Poisson distribution

Format: VoseBurntFingerPoisson(λ_1, λ_2, U)

BurntFingerPoisson equations

A BurntFingerPoisson(λ_1, λ_2) distribution returns discrete values from 0 to +infinity:



Compendium of Probability Distributions

Burnt Finger Poisson Disribution

- This type of situation occurs when, for example, an individual has an expected rate of accidents λ_1, but if an accident occurs the individual will become more careful (his/her "fingers got burned") so that for the rest of the modeled time a new, lower expected accident rate λ_1 applies.

Main Functions for R

- `d` - probability density function
- `p` - cumulative distribution function
- `q` - quantile function
- `r` - random number generation function

Distribution	R name	Additional Args
beta	`beta`	`shape1, shape2, ncp`
binomial	`binom`	`size, prob`
Cauchy	`cauchy`	`location, scale`
chi-squared	`chisq`	`df, ncp`
exponential	`exp`	`rate`
F	`f`	`df1, df2, ncp`
...	...	...
normal	`norm`	`mean, sd`
Poisson	`pois`	`lambda`
Student's t	`t`	`df, ncp`
uniform	`unif`	`min, max`
Weibull	`weibull`	`shape, scale`
Wilcoxon	`wilcox`	`m, n`

Figure:

CRAN Task View: Probability Distributions

Maintainer: Christophe Dutang
Contact: Christophe.Dutang at ensimag.fr
Version: 2015-03-27

For most of the classical distributions, base R provides probability distribution functions (p), density functions (d), quantile functions (q), and random number generation (r). Beyond this basic functionality, many CRAN packages provide additional useful distributions. In particular, multivariate distributions as well as copulas are available in contributed packages.

Ultimate bibles on probability distributions are:

- different volumes of N. L. Johnson, S. Kotz and N. Balakrishnan books, e.g. Continuous Univariate Distributions, Vol. 1,
- Thesaurus of univariate discrete probability distributions by G. Wimmer and G. Altmann.
- Statistical Distributions by M. Evans, N. Hastings, B. Peacock.
- Distributional Analysis with L-moment Statistics using the R Environment for Statistical Computing, Asquith (2011).

Random Number Generators:

- *Basic functionality* : R provides several random number generators (RNGs). The random seed can be provided via `set.seed` and the kind of RNG can be specified using `RNGkind`. The default RNG is the Mersenne-Twister algorithm. Other generators include Wichmann-Hill, Marsaglia-Multicarry, Super-Duper, Knuth-TAOCP, Knuth-TAOCP-2002, as well as user-supplied RNGs. For normal random numbers, the following algorithms are available: Kinderman-Ramage, Ahrens-Dieter, Box-Muller, Inversion (default). In addition to the tools above, setRNG provides an easy way to set, retain information about the setting, and reset the RNG.
- *Pseudo-randomness* : RDieHarder offers several dozen new RNGs from the GNU GSL. randtoolbox provides more recent RNGs such as SF Mersenne-Twister and WELL, which are generators of Mersenne Twister type, but with improved quality parameters. rngwell19937 provides one of the WELL generators with 53 bit resolution of the output and allows seeding by a vector of integers of arbitrary length. randaes provides the deterministic part of the Fortuna cryptographic pseudorandom number generator (AES). SuppDists implements two RNGs of G. Marsaglia.

Figure:

Copulas:

- *Unified approaches* : The packages fCopulae, copula, and copBasic provide a lot of general functionality for copulas. Although lacking support for many existing copulas themselves, copBasic is primarily oriented around utility functions for the general mathematics of copulas as described in the well known introduction to copulas by Nelsen.
- *Archimedean copulas* : The Frank bivariate distribution is available in VGAM, RTDE. fCopulae implements the 22 Archimedean copulas of Nelsen (1998, *Introduction to Copulas* , Springer-Verlag) including Gumbel, Frank, Clayton, and Ali-Mikhail-Haq. gumbel is a standalone package for the Gumbel copula and VGAM provides the Ali-Mikhail-Haq bivariate distribution. copula provides Ali-Mikhail-Haq, Clayton, Frank, Gumbel and Joe copulas. CDVine and VineCopula provide Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7 and BB8 copulas. Generalized Archimedean copulas are implemented in the fgac package.
- *Blomqvist copula* : provided in copBasic.
- *Composition of copula* : copBasic provides functions for composition of a single symmetric copula and composition of two copulas.
- *Cubic copula* : Not yet implemented?
- *Dirichlet copula* : Not yet implemented?

Figure: