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)

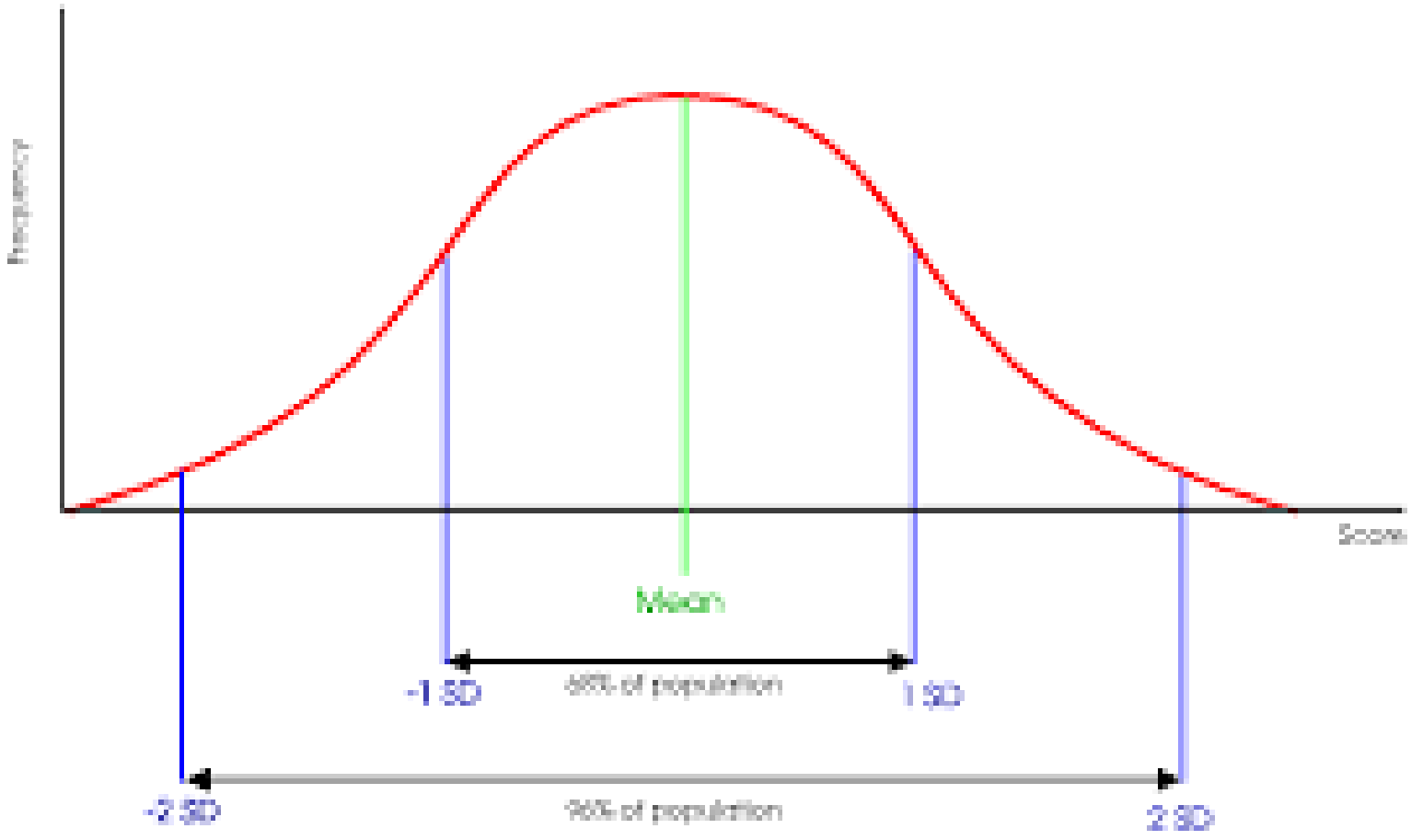
Discrete distributions

- 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

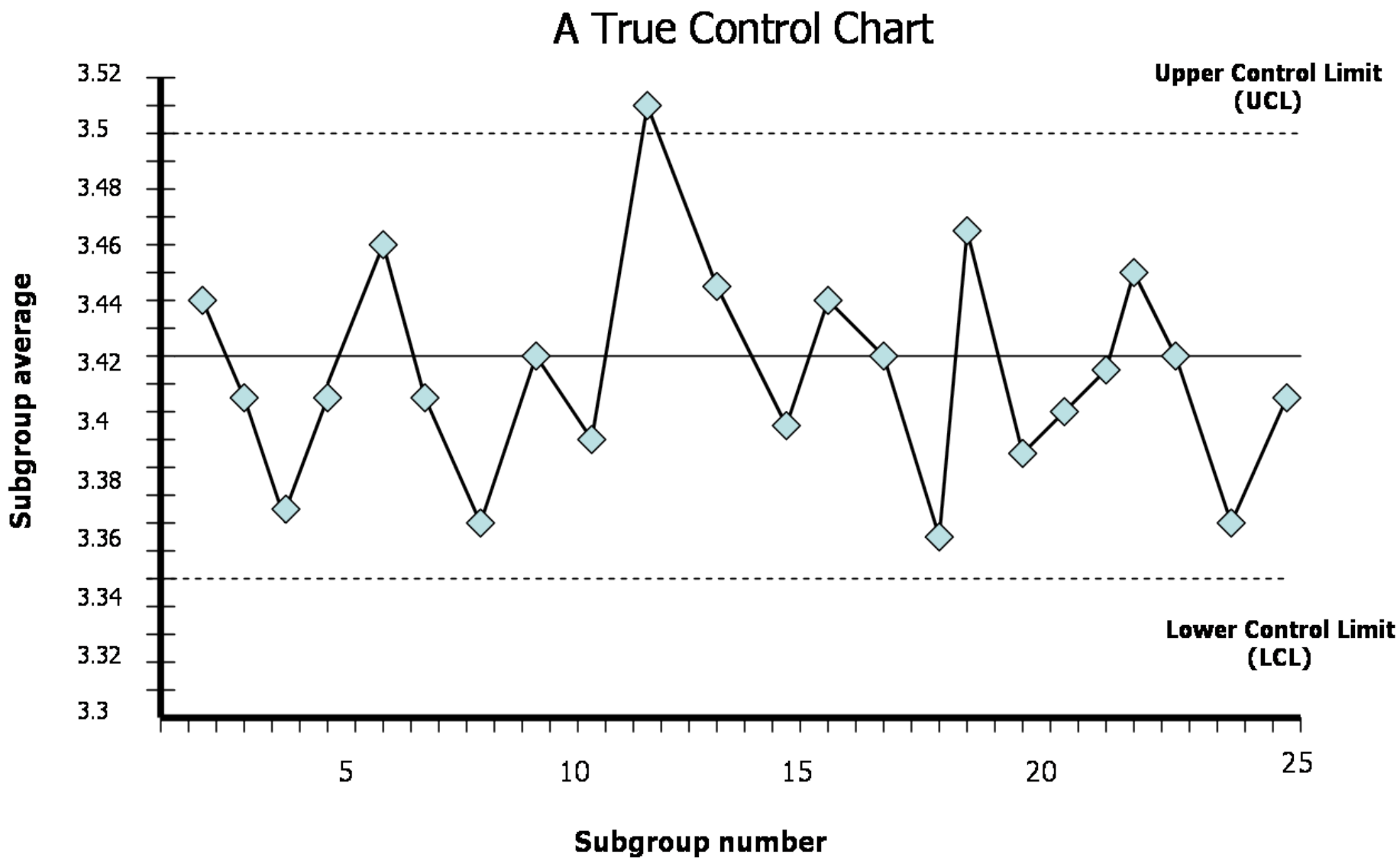
The Normal Distribution Curve



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).


A True Control Chart
Upper Control Limit
(UCL)
Lower Control Limit
(LCL)
Subgroup average
3.52
3.5
3.48
3.46
3.44
3.42
3.4
3.38
3.36
3.34
3.32
3.3
5
10
15
20
25
Subgroup number

Compendium of Probability Distributions

Compendium of Probability Distributions

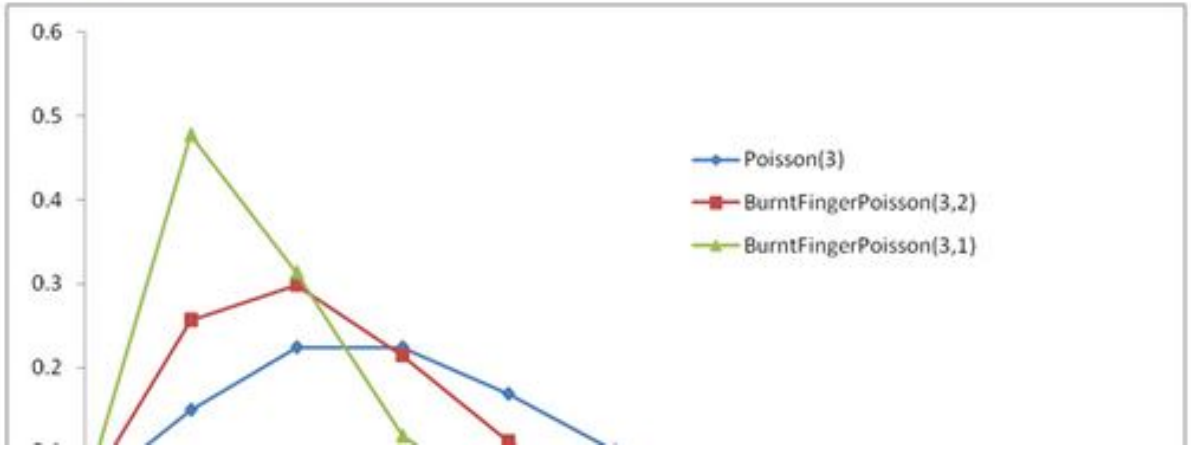
Burnt Finger Poisson Distribution

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.

Compendium of Probability Distributions

Burr Disribution

- The Burr distribution is a right-skewed distribution bounded at the minimum value of a. b is a scale parameter while c and d control its shape.
- It is frequently used to model insurance claim sizes, and is sometimes considered as an alternative to a Normal distribution when data show slight positive skewness.

Compendium of Probability Distributions

Burr Disribution

Burr(0,1,c,d) is a unit Burr distribution. Examples of the Burr distribution are given below:

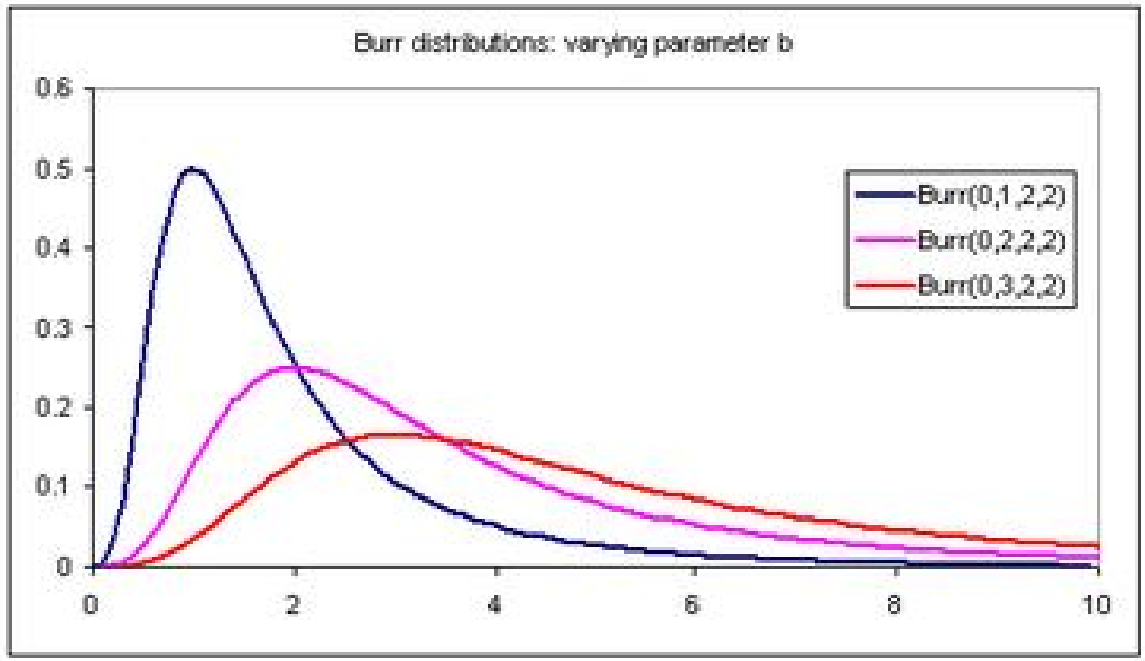


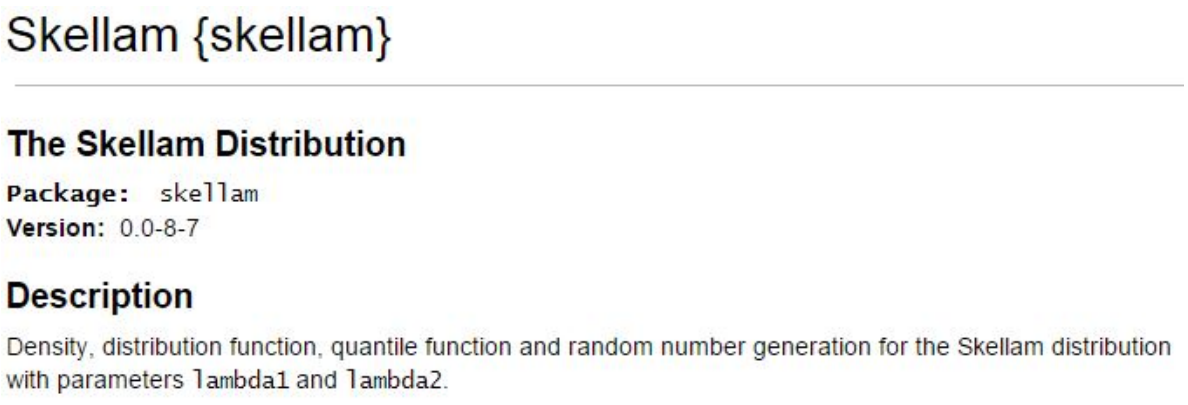
Figure:

Skellam distribution

- The **Skellam distribution** is the discrete probability distribution of the difference $n_1 - n_2$ of two statistically independent random variables N_1 and N_2 each having Poisson distributions with different expected values μ_1 and μ_2.
- It is useful in describing the statistics of the difference of two images with simple photon noise.
- It is also useful in describing the **point spread distribution** in sports where all scored points are equal, such as baseball, hockey and soccer.

Skellam distribution

Package no longer on CRAN - seemingly implemented on some other pacakges though.



Skellam {skellam}

The Skellam Distribution

Package: skellam
Version: 0.0-8-7

Description

Density, distribution function, quantile function and random number generation for the Skellam distribution with parameters lambda1 and lambda2.

Figure:

Compendium of Probability Distributions

Braford Distribution

The theory has a lot of implications in researching and investment in periodicals: for example, how many journals an institute should subscribe to, or one should review in a study. It also gives a guide for advertising, by identifying the first third of journals that have the highest impact, helps determine whether journals on a new(ish) topic (or arena like e-journals) have reached a stabilised population, and test the efficiency of Web browsers.

Main Functions for R

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

Some Probability Distributions in R

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`

Distributions in the stats package

Description

Density, cumulative distribution function, quantile function and random variate generation for many standard probability distributions are available in the **stats** package.

Details

The functions for the density/mass function, cumulative distribution function, quantile function and random variate generation are named in the form `dxxx`, `pxxx`, `qxxx` and `rxxx` respectively.

For the beta distribution see `dbeta`.

For the binomial (including Bernoulli) distribution see `dbinom`

- How many people have to be in a room for there to be a better than 50/50 chance of two people sharing a birthday?

```
> pbirthday(2)
[1] 0.002739726
> 1/365
[1] 0.002739726
> pbirthday(23)
[1] 0.5072972
> 
```

CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics

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).

- *Zipf law* : Package zipfR provides tools for the Zipf and the Zipf-Mandelbrot distributions. VGAM also implements the Zipf distribution.
- *Further distributions* : The VGAM package provides several additional distributions, namely: Skellam, Yule-Simon, Zeta and Haight's Zeta, Borel-Tanner and Felix distribution.

The VGAM package

VGAM: Vector Generalized Linear and Additive Models

An implementation of about 6 major classes of statistical regression models. At the heart of it are the vector generalized linear and additive model (VGLM/VGAM) classes. Many (150+) models and distributions are estimated by maximum likelihood estimation (MLE) or penalized MLE, using Fisher scoring. VGLMs can be loosely thought of as multivariate GLMs. VGAMs are data-driven VGLMs (i.e., with smoothing). The other classes are RR-VGLMs (reduced-rank VGLMs), quadratic RR-VGLMs, reduced-rank VGAMs, RCIMs (row-column interaction models)—these classes perform constrained and unconstrained quadratic ordination (CQO/UQO) models in ecology, as well as constrained additive ordination (CAO). Note that these functions are subject to change, especially before version 1.0.0 is released; see the NEWS file for latest changes.

Version:	0.9-7
Depends:	R ($\geq$ 3.0.0), methods, stats, stats4, splines
Suggests:	VGAMdata, MASS

zipfR: Statistical models for word frequency distributions

Statistical models and utilities for the analysis of word frequency distributions. The utilities include functions for loading, manipulating and visualizing word frequency data and vocabulary growth curves. The package also implements several statistical models for the distribution of word frequencies in a population. (The name of this library derives from the most famous word frequency distribution, Zipf's law.)

Version:	0.6-6
Depends:	R (≥ 2.10.1)
Published:	2012-04-03
Author:	Stefan Evert, Marco Baroni
Maintainer:	Stefan Evert <stefan.evert at uos.de>

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.

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?