

# Package ‘smovie’

April 21, 2018

**Type** Package

**Title** Some Movies to Illustrate Concepts in Statistics

**Version** 1.0.1

**Date** 2018-04-21

**Description** Provides movies to help students to understand statistical concepts. The 'rpanel' package <<https://cran.r-project.org/package=rpanel>> is used to create interactive plots that move to illustrate key statistical ideas and methods. There are movies to: visualise probability distributions (including user-supplied ones); illustrate sampling distributions of the sample mean (central limit theorem), the sample maximum (extremal types theorem) and (the Fisher transformation of the) Pearson product moment correlation coefficient; examine the influence of an individual observation in simple linear regression; illustrate key concepts in statistical hypothesis testing. Also provided are dpqr functions for the distribution of the Fisher transformation of the correlation coefficient under sampling from a bivariate normal distribution.

**Depends** R (>= 3.3.0), rpanel (>= 1.1-3)

**License** GPL (>= 2)

**Encoding** UTF-8

**LazyData** TRUE

**RoxygenNote** 6.0.1

**Imports** graphics, methods, revdbayes (>= 1.1.0), stats, SuppDists

**Suggests** knitr, numDeriv, tkrplot, testthat, rmarkdown

**VignetteBuilder** knitr

**URL** <http://github.com/paulnorthrop/smovie>

**BugReports** <http://github.com/paulnorthrop/smovie/issues>

**NeedsCompilation** no

**Author** Paul J. Northrop [aut, cre, cph]

**Maintainer** Paul J. Northrop <p.northrop@uc1.ac.uk>

**Repository** CRAN

**Date/Publication** 2018-04-21 18:10:09 UTC

## R topics documented:

clt	2
continuous	4
correlation	6
discrete	8
ett	10
FPearson	13
lev_inf	14
movies	15
shypo	16
smovie	18
wws	19

<b>Index</b>	<b>22</b>
--------------	-----------

---

clt	<i>Central Limit Theorem (CLT)</i>
-----	------------------------------------

---

### Description

A movie to illustrate the ideas of the sampling distribution of a mean and the central limit theorem.

### Usage

```
clt(n = 20, distn, params = list(), panel_plot = TRUE, hscale = NA,
    vscale = hscale, n_add = 1, delta_n = 1, arrow = TRUE,
    leg_cex = 1.25, ...)
```

### Arguments

n	An integer scalar. The size of the samples drawn from the distribution chosen using distn.
distn	A character scalar specifying the distribution from which observations are sampled. Distributions "beta", "binomial", "chisq", "chi-squared", "exponential", "f", "gamma", "geometric", "gev", "gp", "hypergeometric", "lognormal", "log-normal", "negative binomial", "normal", "poisson", "t", "uniform" and "weibull" are recognised, case being ignored. If distn is not supplied then distn = "exponential" is used. The "gev" and "gp" cases use the <a href="#">gev</a> and <a href="#">gp</a> distributional functions in the <a href="#">revdbayes</a> package. The other cases use the distributional functions in the <a href="#">stats-package</a> . If distn = "gamma" then the (shape, rate) parameterisation is used. If scale is supplied via params then rate is inferred from this. If distn = "negative binomial" then the (size, prob) parameterisation is used. If mu is supplied via params then prob is inferred from this (and size). If distn = "beta" then ncp is forced to be zero.

params	<p>A named list of additional arguments to be passed to the density function associated with distribution <code>distn</code>. The (shape, rate) parameterisation is used for the gamma distribution (see <a href="#">GammaDist</a>) even if the value of the scale parameter is set using <code>params</code>.</p> <p>If a parameter value is not supplied then the default values in the relevant distributional function set using <code>distn</code> are used, except for "beta" (shape1 = 2, shape2 = 2), "chisq" (df = 4), "f" (df1 = 4, df2 = 8), "gev" (shape = 0.2), "gamma" (shape = 2, "gp" (shape = 0.1), "poisson" (lambda = 5) and "t" (df = 4) and "weibull" (shape = 2).</p>
panel_plot	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the <code>tkrplot</code> library is required.
hscale, vscale	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.
n_add	An integer scalar. The number of simulated datasets to add to each new frame of the movie.
delta_n	A numeric scalar. The amount by which <code>n</code> is increased (or decreased) after one click of the + (or -) button in the parameter window.
arrow	A logical scalar. Should an arrow be included to show the simulated sample maximum from the top plot being placed into the bottom plot?
leg_cex	The argument <code>cex</code> to <a href="#">legend</a> . Allows the size of the legend to be controlled manually.
...	Additional arguments to the <code>rpanel</code> functions <a href="#">rp.button</a> and <a href="#">rp.doublebutton</a> , not including <code>panel</code> , <code>variable</code> , <code>title</code> , <code>step</code> , <code>action</code> , <code>initval</code> , <code>range</code> .

## Details

Loosely speaking, a consequence of the [Central Limit Theorem](#) is that the mean of a **large number** of independent and identically distributed random variables, each with mean  $\mu$  and finite standard deviation  $\sigma$  has **approximately** a normal distribution, even if these original variables are not normally distributed.

This movie considers examples where this limiting result holds and illustrates graphically the closeness of the limiting approximation provided by the relevant normal limit to the true finite- $n$  distribution. Of course, when `distn = "normal"` this result is exact.

Samples of size  $n$  are repeatedly simulated from the distribution chosen using `distn`. These samples are summarized using a plot that appears at the top of the movie screen. For each sample the mean of these  $n$  values is calculated, stored and added to another plot, situated below the first plot. This plot is either a histogram or an empirical c.d.f., chosen using a radio button. A [rug](#) is added to a histogram provided that it contains no more than 1000 points.

The p.d.f. (for a continuous variable) or p.m.f. (for a discrete variable) of the original variables is added to the top plot. There is a checkbox to add to the bottom plot the approximate (large  $n$ ) normal p.d.f./c.d.f. (with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$ ), implied by the CLT.

Once it starts, four aspects of this movie are controlled by the user.

- There are buttons to increase (+) or decrease (-) the sample size, that is, the number of values over which a maximum is calculated.

- Each time the button labelled "simulate another  $n_{\text{add}}$  samples of size  $n$ " is clicked  $n_{\text{add}}$  new samples are simulated and their sample mean are added to the bottom histogram.
- There is a button to switch the bottom plot from displaying a histogram of the simulated means and the limiting normal p.d.f. to the empirical c.d.f. of the simulated data, the exact c.d.f. and the limiting normal c.d.f.

### Value

Nothing is returned, only the animation is produced.

### See Also

[movies](#): a user-friendly menu panel.

[smovie](#): general information about smovie.

### Examples

```
# Exponential data
clt()

# Uniform data
clt(distn = "uniform")

# Poisson data
clt(distn = "poisson")
```

---

continuous

*Univariate Continuous Distributions: p.d.f and c.d.f.*

---

### Description

A movie to illustrate how the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of a continuous random variable depend on the values of its parameters.

### Usage

```
continuous(distn, var_range = NULL, params = list(), param_step = list(),
  param_range = list(), p_vec = NULL, smallest = 0.01,
  plot_par = list(), panel_plot = TRUE, hscale = NA, vscale = hscale,
  ...)
```

### Arguments

**distn** Either a character string or a function to choose the continuous random variable. Strings "beta", "cauchy", "chisq" "chi-squared", "exponential", "f", "gamma", "gev", "gp", "lognormal", "log-normal", "normal", "t", "uniform" and "weibull" are recognised, case being ignored. The relevant distributional functions `dxxx` and `pxxx` in the [stats-package](#) are used. The abbreviations

xxx are also recognised. The "gev" and "gp" cases use the [gev](#) and [gp](#) distributional functions in the [revdbayes](#) package. If `distn = "gamma"` then the (shape, rate) parameterisation is used, unless a value for scale is provided via the argument `params` when the (shape, scale) parameterisation is used. Valid functions are set up like a standard distributional function `dxxx`, with first argument `x`, last argument `log` and with arguments to set the parameters of the distribution in between. See the [CRAN task view on distributions](#).

If `distn` is not supplied then `distn = "normal"` is used.

<code>var_range</code>	A numeric vector of length 2. Can be used to set a fixed range of values over which to plot the p.d.f. and c.d.f., in order better to see the effects of changing the parameter values. If <code>var_range</code> is set then it overrides <code>p_vec</code> (see below).
<code>params</code>	A named list of initial parameter values with which to start the movie. If <code>distn</code> is a string and a particular parameter value is not supplied then the following values are used. "beta": shape1 = 2, shape2 = 2, ncp = 0; "cauchy": location = 0, scale = 1; "chi-squared": df = 4, ncp = 0; "exponential": rate = 1; "f": df1 = 4, df2 = 8, ncp = 0; "gamma": shape = 2, rate = 1; "gev": loc = 0, scale = 1, shape = 0.1; "gp": loc = 0, scale = 1, shape = 0.1; "lognormal": meanlog = 0, sdlog = 1; "normal": mean = 0, sd = 1; "t": df = 4, ncp = 0; "uniform": min = 0, max = 1; "weibull": shape = 2, scale = 1. If <code>distn</code> is a function then <code>params</code> must set any required parameters. If parameter value is outside the corresponding range specified by <code>param_range</code> then it is set to the closest limit of the range.
<code>param_step</code>	A named list of the amounts by which the respective parameters in <code>params</code> are increased/decreased after one click of the +/- button. If <code>distn</code> is a function then the default is 0.1 for all parameters. If <code>distn</code> is a string then a sensible distribution-specific default is set internally.
<code>param_range</code>	A named list of the ranges over which the respective parameters in <code>params</code> are allowed to vary. Each element of the list should be a vector of length 2: the first element gives the lower limit of the range, the second element the upper limit. Use NA to impose no limit. If <code>distn</code> is a function then all parameters are unconstrained.
<code>p_vec</code>	A numeric vector of length 2. The p.d.f. and c.d.f. are plotted between the <code>100p_vec[1]%</code> and <code>100p_vec[2]%</code> quantiles of the distribution. If <code>p_vec</code> is not supplied then a sensible distribution-specific default is used. If <code>distn</code> is a function then the default is <code>p_vec = c(0.001, 0.999)</code> .
<code>smallest</code>	A positive numeric scalar. The smallest value to be used for any strictly positive parameters when <code>distn</code> is a string.
<code>plot_par</code>	A named list of graphical parameters (see <code>link[graphics]{par}</code> ) to be passed to <code>plot</code> . This may be used to alter the appearance of the plots of the p.m.f. and c.d.f.
<code>panel_plot</code>	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the <code>tkrplot</code> library is required.
<code>hscale, vscale</code>	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.

... Additional arguments to be passed to `rp.doublebutton`, not including `panel`, `variable`, `title`, `step`, `action`, `initval`, `range`.

### Details

The movie starts with a plot of the p.d.f. of the distribution for the initial values of the parameters. Buttons increase (+) or decrease (-) each parameter. There are radio buttons to switch the plot from the p.d.f. to the c.d.f. and back.

### Value

Nothing is returned, only the animation is produced.

### See Also

[movies](#): a user-friendly menu panel.

[smovie](#): general information about smovie.

### Examples

```
# Normal example
continuous()
# Fix the range of values over which to plot
continuous(var_range = c(-10, 10))

# The same example, but using a user-supplied function and setting manually
# the initial parameters, parameter step size and range
continuous(distn = dnorm, params = list(mean = 0, sd = 1),
           param_step = list(mean = 1, sd = 1),
           param_range = list(sd = c(0, NA)))

# Gamma distribution. Show the use of var_range
continuous(distn = "gamma", var_range = c(0, 15))
```

---

correlation

*Sampling distribution of the Pearson correlation coefficient movie*

---

### Description

A movie to illustrate how the sampling distribution of the Pearson product moment sample correlation coefficient  $r$  depends on the sample size  $n$  and on the true correlation  $\rho$ .

### Usage

```
correlation(n = 30, rho = 0, panel_plot = TRUE, hscale = NA,
           vscale = hscale, delta_n = 1, delta_rho = 0.1, ...)
```

**Arguments**

<code>n</code>	An integer scalar. The initial value of the sample size. Must not be less than 2.
<code>rho</code>	A numeric scalar. The initial value of the true correlation $\rho$ . Must be in $[-1, 1]$ .
<code>panel_plot</code>	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the <code>tkrplot</code> library is required.
<code>hscale, vscale</code>	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.
<code>delta_n</code>	An integer scalar. The amount by which the value of the sample size is increased/decreased after one click of the +/- button.
<code>delta_rho</code>	A numeric scalar. The amount by which the value of <code>rho</code> is increased/decreased after one click of the +/- button.
<code>...</code>	Additional arguments to the <code>rpanel</code> functions <code>rp.button</code> and <code>rp.doublebutton</code> , not including <code>panel</code> , <code>variable</code> , <code>title</code> , <code>step</code> , <code>action</code> , <code>initval</code> , <code>range</code> .

**Details**

Random samples of size  $n$  are simulated from a bivariate normal distribution **bivariate normal distribution** in which each of the variables has a mean of 0 and a variance of 1 and the correlation  $\rho$  between the variables is chosen by the user.

The movie contains two plots. On the top is a scatter plot of the simulated sample, illustrating the strength of the association between the simulated values of the variables. A new sample is produced by clicking "simulate another sample. For each simulated sample the sample (Pearson product moment) correlation coefficient  $r$  is calculated and displayed in the title of the top plot.

The values of the sample correlation coefficients are stored and are plotted in a histogram in the bottom plot. A rug displays the individual values, with the most recent value coloured red. As we accumulate a large number of values in this histogram the shape of the sampling distribution of  $r$  emerges. The exact p.d.f. of  $r$  is superimposed on this histogram, as is the value of  $\rho$ .

The bottom plot can be changed in two ways: (i) a radio button can be pressed to replace the histogram and pdf with a plot of the empirical c.d.f. and exact cdf; (ii) the variable can be changed from  $\rho$  to Fisher's z-transformation  $F(\rho) = \operatorname{arctanh}(\rho) = [\ln(1+\rho) - \ln(1-\rho)]/2$ . For sufficiently large values of  $n$ ,  $F(\rho)$  has approximately a normal distribution with mean  $\rho$  and variance  $1/(n-3)$ .

The values of the sample size  $n$  or true correlation coefficient  $\rho$  can be changed using the respective +/- buttons. If one of these is changed then the bottom plot is reset using the sample correlation coefficient of the first sample simulated using the new combination of  $n$  and  $\rho$ .

**Value**

Nothing is returned, only the animation is produced.

**See Also**

[movies](#): a user-friendly menu panel.

[smovie](#): general information about smovie.

**Examples**

```
correlation(rho = 0.8)
correlation(n = 10)
```

discrete

*Univariate Discrete Distributions: p.m.f and c.d.f.***Description**

A movie to illustrate how the probability mass function (p.m.f.) and cumulative distribution function (c.d.f.) of a discrete random variable depend on the values of its parameters.

**Usage**

```
discrete(distn, var_support = NULL, params = list(), param_step = list(),
  param_range = list(), p_vec = NULL, smallest = 0.01,
  plot_par = list(), panel_plot = TRUE, hscale = NA, vscale = hscale,
  observed_value = NA, ...)
```

**Arguments**

- |             |   |
|-------------|---|
| distn       | <p>Either a character string or a function to choose the discrete random variable. Strings "binomial", "geometric", "hypergeometric", "negative binomial" and "poisson" are recognised, case being ignored. The relevant distributional functions <code>dxxx</code> and <code>pxxx</code> in the <a href="#">stats-package</a> are used. The abbreviations <code>xxx</code> are also recognised. If <code>distn = "hypergeometric"</code> then the (size, prob) parameterisation is used, unless a value for <code>mu</code> is provided via the argument <code>params</code> when the (size, mu) parameterisation is used.</p> <p>Valid functions are set up like a standard distributional function <code>dxxx</code>, with first argument <code>x</code>, last argument <code>log</code> and with arguments to set the parameters of the distribution in between. See the <a href="#">CRAN task view on distributions</a>. It is assumed that the support of the random variable is a subset of the integers, unless <code>var_support</code> is set to the contrary.</p> <p>If <code>distn</code> is not supplied then <code>distn = "binomial"</code> is used.</p> |
| var_support | <p>A numeric vector. Can be used to set a fixed set of values for which to plot the p.m.f. and c.d.f., in order better to see the effects of changing the parameter values or to set a support that isn't a subset of the integers. If <code>var_support</code> is set then it overrides <code>p_vec</code> (see below).</p>  |
| params      | <p>A named list of initial parameter values with which to start the movie. If <code>distn</code> is a string and a particular parameter value is not supplied then the following values are used. "binomial": <code>size = 10</code>, <code>prob = 0.5</code>; "geometric": <code>prob = 0.5</code>; "hypergeometric": <code>m = 10</code>, <code>n = 7</code>, <code>k = 8</code>; "negative binomial": <code>size = 10</code>, <code>prob = 0.5</code>; "poisson": <code>lambda = 5</code>.</p> <p>If <code>distn</code> is a function then <code>params</code> must set any required parameters.</p> <p>If parameter value is outside the corresponding range specified by <code>param_range</code> then it is set to the closest limit of the range.</p>  |

<code>param_step</code>	A named list of the amounts by which the respective parameters in <code>params</code> are increased/decreased after one click of the +/- button. If <code>distn</code> is a function then the default is 0.1 for all parameters. If <code>distn</code> is a string then a sensible distribution-specific default is set internally.
<code>param_range</code>	A named list of the ranges over which the respective parameters in <code>params</code> are allowed to vary. Each element of the list should be a vector of length 2: the first element gives the lower limit of the range, the second element the upper limit. Use <code>NA</code> to impose no limit. If <code>distn</code> is a function then all parameters are unconstrained.
<code>p_vec</code>	A numeric vector of length 2. The p.d.f. and c.d.f. are plotted between the $100p\_vec[1]\%$ and $100p\_vec[2]\%$ quantiles of the distribution. If <code>p_vec</code> is not supplied then a sensible distribution-specific default is used. If <code>distn</code> is a function then the default is <code>p_vec = c(0.001, 0.999)</code> .
<code>smallest</code>	A positive numeric scalar. The smallest value to be used for any strictly positive parameters when <code>distn</code> is a string.
<code>plot_par</code>	A named list of graphical parameters (see <code>link[graphics]{par}</code> ) to be passed to <code>plot</code> . This may be used to alter the appearance of the plots of the p.m.f. and c.d.f.
<code>panel_plot</code>	A logical parameter that determines whether the plot is placed inside the panel ( <code>TRUE</code> ) or in the standard graphics window ( <code>FALSE</code> ). If the plot is to be placed inside the panel then the <code>tkrplot</code> library is required.
<code>hscale, vscale</code>	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.
<code>observed_value</code>	A non-negative integer. If <code>observed_value</code> is supplied then the corresponding line in the plot of the p.m.f. is coloured in red.
<code>...</code>	Additional arguments to be passed to <code>rp.doublebutton</code> , not including <code>panel</code> , <code>variable</code> , <code>title</code> , <code>step</code> , <code>action</code> , <code>initval</code> , <code>range</code> .

### Details

The movie starts with a plot of the p.m.f. of the distribution for the initial values of the parameters. Buttons increase (+) or decrease (-) each parameter. There are radio buttons to switch the plot from the p.m.f. to the c.d.f. and back.

Owing to a conflict with the argument size of the function `rp.control` the parameter size of, for example, the binomial and negative binomial distributions, is labelled as `n`.

### Value

Nothing is returned, only the animation is produced.

### See Also

[movies](#): a user-friendly menu panel.

[smovie](#): general information about smovie.

## Examples

```
# Binomial example
discrete()

# The same example, but using a user-supplied function and setting manually
# the initial parameters, parameter step size and range
discrete(distn = dbinom, params = list(size = 10, prob = 0.5),
         param_step = list(size = 1),
         param_range = list(size = c(1, NA), prob = c(0, 1)))

# Poisson distribution. Show the use of var_support
discrete(distn = "poisson", var_support = 0:20)
```

---

 ett

*Extremal Types Theorem (ETT)*


---

## Description

A movie to illustrate the extremal types theorem, that is, convergence of the distribution of the maximum of a random sample of size  $n$  from certain distributions to a member of the Generalized Extreme Value (GEV) family, as  $n$  tends to infinity. Samples of size  $n$  are simulated repeatedly from the chosen distribution. The distributions (simulated empirical and true) of the sample maxima are compared to the relevant GEV limit.

## Usage

```
ett(n = 20, distn, params = list(), panel_plot = TRUE, hscale = NA,
    vscale = hscale, n_add = 1, delta_n = 1, arrow = TRUE,
    leg_cex = 1.25, ...)
```

## Arguments

n	An integer scalar. The size of the samples drawn from the distribution chosen using <code>distn</code> . <code>n</code> must be no smaller than 2.
distn	<p>A character scalar specifying the distribution from which observations are sampled. Distributions "beta", "cauchy", "chisq", "chi-squared", "exponential", "f", "gamma", "gp", lognormal, log-normal, "ngev", "normal", "t", "uniform" and "weibull" are recognised, case being ignored.</p> <p>If <code>distn</code> is not supplied then <code>distn = "exponential"</code> is used.</p> <p>The "gp" case uses the <a href="#">gp</a> distributional functions in the <a href="#">revdbayes</a> package.</p> <p>The "ngev" case is a negated GEV(<math>1/\xi</math>, 1, <math>\xi</math>) distribution, for <math>\xi &gt; 0</math>, and uses the <a href="#">gev</a> distributional functions in the <a href="#">revdbayes</a> package. If <math>\xi = 1</math> then this coincides with Example 1.7.5 in Leadbetter, Lindgren and Rootzen (1983).</p> <p>The other cases use the distributional functions in the <a href="#">stats-package</a>. If <code>distn = "gamma"</code> then the (shape, rate) parameterisation is used. If scale is supplied via <code>params</code> then rate is inferred from this. If <code>distn = "beta"</code> then <code>ncp</code> is forced to be zero.</p>

params	<p>A named list of additional arguments to be passed to the density function associated with distribution <code>distn</code>. The (shape, rate) parameterisation is used for the gamma distribution (see <a href="#">GammaDist</a>) even if the value of the scale parameter is set using <code>params</code>.</p> <p>If a parameter value is not supplied then the default values in the relevant distributional function set using <code>distn</code> are used, except for "beta" (shape1 = 2, shape2 = 2), "chisq" (df = 4), "f" (df1 = 4, df2 = 8), "ngev" (shape = 0.2), "gamma" (shape = 2, "gp" (shape = 0.1), "t" (df = 4) and "weibull" (shape = 2).</p>
panel_plot	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the <code>tkrplot</code> library is required.
hscale, vscale	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.
n_add	An integer scalar. The number of simulated datasets to add to each new frame of the movie.
delta_n	A numeric scalar. The amount by which <code>n</code> is increased (or decreased) after one click of the + (or -) button in the parameter window.
arrow	A logical scalar. Should an arrow be included to show the simulated sample maximum from the top plot being placed into the bottom plot?
leg_cex	The argument <code>cex</code> to <a href="#">legend</a> . Allows the size of the legend to be controlled manually.
...	Additional arguments to the <code>rpanel</code> functions <a href="#">rp.button</a> and <a href="#">rp.doublebutton</a> , not including <code>panel</code> , <code>variable</code> , <code>title</code> , <code>step</code> , <code>action</code> , <code>initval</code> , <code>range</code> .

## Details

Loosely speaking, a consequence of the **Extremal Types Theorem** is that, in many situations, the maximum of a *large number*  $n$  of independent random variables has *approximately* a  $GEV(\mu, \sigma, \xi)$  distribution, where  $\mu$  is a location parameter,  $\sigma$  is a scale parameter and  $\xi$  is a shape parameter. See Coles (2001) for an introductory account and Leadbetter et al (1983) for greater detail and more examples. The Extremal Types Theorem is an asymptotic result that considers the possible limiting distribution of linearly normalised maxima as  $n$  tends to infinity. This movie considers examples where this limiting result holds and illustrates graphically the closeness of the limiting approximation provided by the relevant GEV limit to the true finite- $n$  distribution.

Samples of size  $n$  are repeatedly simulated from the distribution chosen using `distn`. These samples are summarized using a histogram that appears at the top of the movie screen. For each sample the maximum of these  $n$  values is calculated, stored and added to another plot, situated below the first plot. A [rug](#) is added to a histograms provided that it contains no more than 1000 points. This plot is either a histogram or an empirical c.d.f., chosen using a radio button.

The probability density function (p.d.f.) of the original variables is superimposed on the top histogram. There is a checkbox to add to the bottom plot the exact p.d.f./c.d.f. of the sample maxima and an approximate (large  $n$ ) GEV p.d.f./c.d.f. implied by the ETT. The GEV shape parameter  $\xi$  that applies in the limiting case is used. The GEV location  $\mu$  and scale  $\sigma$  are set based on constants used to normalise the maxima to achieve the GEV limit. Specifically,  $\mu$  is set at the  $100(1-1/n)\%$  quantile of the distribution `distn` and  $\sigma$  at  $(1/n) / f(\mu)$ , where  $f$  is the density function of the distribution `distn`.

Once it starts, four aspects of this movie are controlled by the user.

- There are buttons to increase (+) or decrease (-) the sample size, that is, the number of values over which a maximum is calculated.
- Each time the button labelled "simulate another  $n_{\text{add}}$  samples of size  $n$ " is clicked  $n_{\text{add}}$  new samples are simulated and their sample maxima are added to the bottom histogram.
- There is a button to switch the bottom plot from displaying a histogram of the simulated maxima, the exact p.d.f. and the limiting GEV p.d.f. to the empirical c.d.f. of the simulated data, the exact c.d.f. and the limiting GEV c.d.f.
- There is a box that can be used to display only the bottom plot. This option is selected automatically if the sample size  $n$  exceeds 100000.
- There is a box that can be used to display only the bottom plot. This option is selected automatically if the sample size  $n$  exceeds 100000.

For further detail about the examples specified by `distn` see Chapter 1 of Leadbetter et al. (1983) and Chapter 3 of Coles (2001). In many of these examples ("exponential", "normal", "gamma", "lognormal", "chi-s", the limiting GEV distribution has a shape parameter that is equal to 0. In the "uniform" case the limiting shape parameter is -1 and in the "beta" case it is  $-1 / \text{shape2}$ , where `shape2` is the second parameter of the `Beta` distribution. In the other cases the limiting shape parameter is positive, with respective values `shape` ("gp", see `gp`),  $1 / \text{df}$  ("t", see `TDist`), 1 ("cauchy", see `Cauchy`),  $2 / \text{df}^2$  ("f", see `FDist`).

## Value

Nothing is returned, only the animation is produced.

## References

Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London. [http://dx.doi.org/10.1007/978-1-4471-3675-0\\_3](http://dx.doi.org/10.1007/978-1-4471-3675-0_3)

Leadbetter, M., Lindgren, G. and Rootzen, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York. <http://dx.doi.org/10.1007/978-1-4612-5449-2>

## See Also

`movies`: a user-friendly menu panel.

`smovie`: general information about smovie.

## Examples

```
# Exponential data: xi = 0
ett()

# Uniform data: xi = -1
ett(distn = "uniform")

# Student t data: xi = 1 / df
ett(distn = "t", params = list(df = 5))
```

---

FPearson	<i>Fisher's transformation of the Pearson product moment correlation coefficient</i>
----------	--

---

### Description

Density, distribution function, quantile function and random generator for the distribution of Fisher's transformation of Pearson's product moment correlation, based on a random sample from a bivariate normal distribution

### Usage

```
dFPearson(x, N, rho = 0, log = FALSE)

pFPearson(q, N, rho = 0, lower.tail = TRUE, log.p = FALSE)

qFPearson(p, N, rho = 0, lower.tail = TRUE, log.p = FALSE)

rFPearson(n, N, rho = 0, lower.tail = TRUE, log.p = FALSE)
```

### Arguments

x, q	Numeric vectors of quantiles.
N	Numeric vector. Number of observations, (N > 3).
rho	Numeric vector. Population correlations, (-1 < rho < 1).
log, log.p	A logical scalar; if TRUE, probabilities p are given as log(p).
lower.tail	A logical scalar. If TRUE (default), probabilities are P[X <= x], otherwise, P[X > x].
p	A numeric vector of probabilities in [0,1].
n	Numeric scalar. The number of observations to be simulated. If length(n) > 1 then length(n) is taken to be the number required.

### References

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, **10**(4), 507-521. <http://dx.doi.org/10.2307/2331838>

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3-32. <https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15169/1/14.pdf>

### See Also

[Pearson](#) in the SuppDists package for dpqr functions for the untransformed Pearson produce moment correlation coefficient.

[correlation](#): correlation sampling distribution movie.

**Examples**

```
dFPearson(-1:1, N = 10)
dFPearson(0, N = 11:20)

pFPearson(0.5, N = 10)
pFPearson(0.5, N = 10, rho = c(0, 0.3))

qFPearson((1:9)/10, N = 10, rho = 0.2)
qFPearson(0.5, N = c(10, 20), rho = c(0, 0.3))

rFPearson(6, N = 10, rho = 0.6)
```

lev\_inf

*Leverage and influence in simple linear regression movie***Description**

A movie to examine the influence of a single outlying observation on a least squares regression line.

**Usage**

```
lev_inf(association = c("positive", "negative", "none"), n = 25,
        panel_plot = TRUE, hscale = NA, vscale = hscale)
```

**Arguments**

association	A character scalar. Determines the type of association between (not-outlying) observations: "positive" for positive linear association; "negative" negative linear association; "none" for no association.
n	An integer scalar. The size of the sample of (non-outlying) observations.
panel_plot	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the tkrplot library is required.
hscale, vscale	Numeric scalars. Scaling parameters for the size of the plot when panel_plot = TRUE. The default values are 1.4 on Unix platforms and 2 on Windows platforms.

**Details**

$n$  pairs of observations are simulated with the property that the mean of response variable  $y$  is a linear function of the values of the explanatory variable  $x$ . These pairs of observations are plotted using filled black circles. An extra observation is plotted using a filled red circle. Initially this observation is placed in the middle of the plot.

Superimposed on the plot are two least squares regression lines: one based on all the data ('with observation') and one in which the 'red' observation has been removed ('without observation'). Initially these lines coincide.

The location of the ‘red’ observation can be changed using the +/- buttons so that the effect of the position of this observation on the ‘with observation’ line can be seen.

We see that if the red observation is outlying, that is, it is far from the least squares line fitted to the other observations, then its **influence** on the least squares regression line depends on its x-coordinate. If its x-coordinate is much larger or smaller than the x-coordinate of the other observations (**high leverage**) then the influence is higher than if it has a similar x-coordinate to the other observations (**low leverage**). An observation with high leverage does not necessarily have high influence: if its y-coordinate falls very close to the regression line fitted to the other observations then its influence will be low.

### Value

Nothing is returned, only the animation is produced.

### See Also

[movies](#): a user-friendly menu panel.

[smovie](#): general information about smovie.

### Examples

```
# Positive association
lev_inf()

# No association
lev_inf(association = "none")
```

---

`movies`

*Main menu for smovie movies*

---

### Description

Uses the template [rp.cartoons](#) function to produce a menu panel from which any of the movies in [smovie](#) package can be launched. For greater control of an individual example call the relevant function directly.

### Usage

```
movies(fixed_range = TRUE, hscale = NA, vscale = hscale)
```

### Arguments

`fixed_range` A logical scalar. Only relevant to the **Discrete** and **Continuous** menus. If TRUE then in the call to [discrete](#) or [continuous](#) the argument `var_support` (`discrete`) or `var_range` (`continuous`) is set so that the values on the horizontal axes are fixed at values that enable the movie to show the effects of changing the parameters of the distribution, at least locally to the default initial values for the parameters. For greater control call [discrete](#) or [continuous](#) directly.

hscale, vscale Numeric scalars. Scaling parameters for the size of the plot when panel\_plot = TRUE. The default values are 1.4 on Unix platforms and 2 on Windows platforms.

### See Also

[discrete](#), [continuous](#), [clt](#), [ett](#), [correlation](#), [lev\\_inf](#), [wvs](#), [shypo](#).

[smovie](#): general information about smovie.

### Examples

```
movies()
```

---

shypo

*Testing simple hypotheses*

---

### Description

A movie to illustrate statistical concepts involved in the testing of one simple hypothesis against another. The example used is a random sample from a normal distribution whose variance is assumed to be known. The simple hypotheses relate to the value of the mean  $\mu$ .

### Usage

```
shypo(mu0 = 0, sd = 6, eff = sd, n = 10, a = mu0 + eff/2,
      target_alpha = 0.05, target_beta = 0.1, panel_plot = TRUE,
      hscale = NA, vscale = hscale, delta_n = 1, delta_a = sd/(10 *
      sqrt(n)), delta_eff = sd, delta_mu0 = 1, delta_sd = 1)
```

### Arguments

mu0	A numeric scalar. The value of $\mu$ under the null hypothesis $H_0$ with which to start the movie.
sd	A positive numeric scalar. The (common) standard deviation $\sigma$ of the normal distributions of the data under the two hypotheses.
eff	A numeric scalar. The <i>effect size</i> . The amount by which the value of $\mu$ under the alternative hypothesis is greater than the value mu0 under the null hypothesis. That is, $\mu_1 = \text{eff} + \text{mu0}$ . eff must be non-negative.
n	A positive integer scalar. The sample size with which to start the movie.
a	A numeric scalar. The critical value of the test with which to start the movie. $H_0$ is rejected if the sample mean is greater than a.
target_alpha	A numeric scalar in (0,1). The target value of the type I error to be achieved by setting a and/or n if the user asks for this using a radio button.
target_beta	A numeric scalar in (0,1). The target value of the type II error to be achieved by setting a and/or n if the user asks for this using a radio button.

<code>panel_plot</code>	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the tkrplot library is required.
<code>hscale</code> , <code>vscale</code>	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.
<code>delta_mu0</code> , <code>delta_eff</code> , <code>delta_a</code> , <code>delta_n</code> , <code>delta_sd</code>	Numeric scalars. The respective amounts by which the values of <code>mu0</code> , <code>eff</code> , <code>a</code> , <code>n</code> and <code>sd</code> are increased (or decreased) after one click of the + (or -) button in the parameter window.

## Details

The movie is based on two plots.

The top plot shows the (normal) probability density functions of the sample mean under the null hypothesis  $H_0$  (mean  $\mu_0$ ) and the alternative hypothesis  $H_1$  (mean  $\mu_1$ , where  $\mu_1 > \mu_0$ ), with the values of  $\mu_0$  and  $\mu_1$  indicated by vertical dashed lines.  $H_0$  is rejected if the sample mean exceeds the critical value  $a$ , which is indicated by a vertical black line.

The bottom plot shows how the probabilities of making a type I or type II error (alpha and beta respectively) depend on the value of  $a$ , by plotting these probabilities against  $a$ .

A parameter window enables the user to change the values of  $n$ ,  $a$ ,  $\mu_0$ ,  $\text{eff} = \mu_1 - \mu_0$  or  $sd$  by clicking the +/- buttons.

Radio buttons can be used either to:

- set  $a$  to achieve the target type I error probability `target_alpha`, based on the current value of  $n$ ;
- set  $a$  and (integer)  $n$  to achieve (or better) the respective target type I and type II error probabilities of `target_alpha` and `target_beta`, based on the current value of  $n$ .

## Value

Nothing is returned, only the animation is produced.

## See Also

[movies](#): a user-friendly menu panel.

[smovie](#): general information about smovie.

## Examples

```
# 1. Change a (for fixed n) to achieve alpha = 0.05
# 2. Change a and n to achieve alpha <= 0.05 and beta <= 0.1
shypo(mu0 = 0, eff = 5, n = 16, a = 2.3, delta_a = 0.01)
```

---

smovie

*smovie: some movies to illustrate concepts in statistics*

---

## Description

These movies are animations used to illustrate key statistical ideas. They are produced using the [rpanel-package](#), which has [BWidget](#) as a system requirement. [BWidget](#) is included in the R installers for Windows and macOS. For other platforms please see [Section 1.1.7 of Writing R Extensions](#) for installation advice.

## Details

When one of these functions is called R opens up a small *parameter window* containing clickable buttons that can be used to change parameters underlying the plot. For the effects of these buttons see the documentation of the individual functions.

See `vignette("smovie-vignette", package = "smovie")` for an overview of the package and the [user-friendly menu panel](#).

There are movies on the following topics.

## Probability distributions

- [Discrete distributions](#)
- [Continuous distributions](#)

## Sampling distributions

- [Central Limit Theorem: sampling distribution of a sample mean](#)
- [Extremal Types Theorem: sampling distribution of a sample maximum](#)
- [Pearson product moment correlation coefficient](#)

## Regression

- [Leverage and influence in simple linear regression](#)

## Hypothesis testing

- [Wald, Wilks and Score tests](#)
- [Testing simple hypotheses](#)

## References

Bowman, A., Crawford, E., Alexander, G. and Bowman, R. W. (2007). rpanel: Simple Interactive Controls for R Functions Using the tcltk Package. *Journal of Statistical Software*, **17(9)**, 1-18. <http://www.jstatsoft.org/v17/i09/>.

**Description**

A movie to illustrate the nature of the Wald, Wilks and score likelihood-based test statistics, for a model with a scalar unknown parameter  $\theta$ . The user can change the value of the parameter under a simple null hypothesis and observe the effect on the test statistics and (approximate) p-values associated with the tests of this hypothesis against the general alternative. The user can specify their own log-likelihood or use one of two in-built examples.

**Usage**

```
wvs(model = c("norm", "binom"), theta_range = NULL, mult = 3,
    theta0 = if (!is.null(theta_range)) sum(c(0.25, 0.75) * theta_range) else
    NULL, panel_plot = TRUE, hscale = NA, vscale = hscale,
    delta_theta0 = if (!is.null(theta_range)) abs(diff(theta_range))/20 else
    NULL, theta_mle = NULL, loglik = NULL, alg_score = NULL,
    alg_obs_info = NULL, digits = 3, ...)
```

**Arguments**

model	<p>A character scalar. Name of the the distribution on which one of two in-built examples are based.</p> <p>If model = "norm" then the setting is a random sample of size n from a normal distribution with unknown mean <math>\mu = \theta</math> and known standard deviation sigma.</p> <p>If model = "binom" then the setting is a random sample from a Bernoulli distribution with unknown success probability <math>\theta</math>.</p> <p>The behaviour of these examples can be controlled using arguments supplied via . . . . In particular, the data can be supplied using data. If model = "norm" then n, mu, and sigma can also be chosen. The default cases for these examples are:</p> <ul style="list-style-type: none"> <li>• model = "norm": n = 10, mu = 0, sigma = 1 and data contains a sample of a sample of size n simulated, using <a href="#">Normal</a>, from a normal distribution with mean mu and standard deviation sigma.</li> <li>• model = "binom": data = c(7, 13), that is, 7 successes and 13 failures observed in 20 trials. For the purposes of this movie there must be at least one success and at least one failure.</li> </ul>
theta_range	A numeric vector of length 2. The range of values of $\theta$ over which to plot the log-likelihood. If theta_range is not supplied then the argument mult is used to set the range automatically.
mult	A positive numeric scalar. If theta_range is not supplied then an interval of width 2 x mult standard errors centred on theta_mle is used. If model = "binom" then theta_range is truncated to (0,1) if necessary.
theta0	A numeric scalar. The value of $\theta$ under the null hypothesis to use at the start of the movie.

<code>panel_plot</code>	A logical parameter that determines whether the plot is placed inside the panel (TRUE) or in the standard graphics window (FALSE). If the plot is to be placed inside the panel then the <code>tkrplot</code> library is required.
<code>hscale, vscale</code>	Numeric scalars. Scaling parameters for the size of the plot when <code>panel_plot = TRUE</code> . The default values are 1.4 on Unix platforms and 2 on Windows platforms.
<code>delta_theta0</code>	A numeric scalar. The amount by which the value of <code>theta0</code> is increased (or decreased) after one click of the + (or -) button in the parameter window.
<code>theta_mle</code>	A numeric scalar. The user may use this to supply the value of the maximum likelihood estimate (MLE) of $\theta$ . Otherwise, <code>optim</code> is used to search for the MLE, using <code>theta0</code> as the initial value and <code>theta_range</code> as bounds within which to search.
<code>loglik</code>	An R function, vectorised with respect to its first argument, that returns the value of the log-likelihood (up to an additive constant). The movie will not work if the observed information is not finite at the maximum likelihood estimate.
<code>alg_score</code>	A R function that returns the score function, that is, the derivative of <code>loglik</code> with respect to $\theta$ .
<code>alg_obs_info</code>	A R function that returns the observed information that is, the negated second derivative of <code>loglik</code> with respect to $\theta$ .
<code>digits</code>	An integer indicating the number of significant digits to be used in the displayed values of the test statistics and p-values. See <code>signif</code> .
<code>...</code>	Additional arguments to be passed to <code>loglik</code> , <code>alg_score</code> and <code>alg_obs_info</code> if <code>loglik</code> is supplied, or to functions relating to the in-built examples otherwise. See the description of <code>model</code> above for details.

## Details

The **Wald**, **Wilks** (or likelihood ratio) and **Score** tests are asymptotically equivalent tests of a simple hypothesis that a parameter of interest  $\theta$  is equal to a particular value  $\theta_0$ . The test statistics are all based on the log-likelihood  $l(\theta)$  for  $\theta$  but they differ in the way that they measure the distance between the maximum likelihood estimate (MLE) of  $\theta$  and  $\theta_0$ . The Wilks statistic is the amount by which the log-likelihood evaluated  $\theta_0$  is smaller than the log-likelihood evaluated at the MLE. The Wald statistic is based on the absolute difference between the MLE and  $\theta_0$ . The score test is based on the gradient of the log-likelihood (the score function) at  $\theta_0$ . For details see Azzalini (1996).

This movie illustrates the differences between the test statistics for simple models with a single scalar parameter. In the (default) normal example the three test statistics coincide. This is not true in general, as shown by the other in-built example (`distn = "binom"`).

A user-supplied log-likelihood can be provided via `loglik`.

## Value

Nothing is returned, only the animation is produced.

## References

Azzalini, A. (1996) *Statistical Inference Based on the Likelihood*, Chapman & Hall / CRC, London.

**See Also**

[movies](#): a user-friendly menu panel.  
[smovie](#): general information about smovie.

**Examples**

```
# N(theta, 1) example, test statistics equivalent
wvs(theta0 = 0.8)

# binomial(20, theta) example, test statistics similar
wvs(theta0 = 0.5, model = "binom")

# binomial(20, theta) example, test statistic rather different
# for theta0 distant from theta_mle
wvs(theta0 = 0.9, model = "binom", data = c(19, 1), theta_range = c(0.1, 0.99))

# binomial(2000, theta) example, test statistics very similar
wvs(theta0 = 0.5, model = "binom", data = c(1000, 1000))

set.seed(47)
x <- rnorm(10)
wvs(theta0 = 0.2, model = "norm", theta_range = c(-1, 1))

# Log-likelihood for a binomial experiment (up to an additive constant)
bin_loglik <- function(p, n_success, n_failure) {
  return(n_success * log(p) + n_failure * log(1 - p))
}

wvs(loglik = bin_loglik, theta0 = 0.5, theta_range = c(0.1, 0.7),
     theta_mle = 7 / 20, n_success = 7, n_failure = 13)

bin_alg_score <- function(p, n_success, n_failure) {
  return(n_success / p - n_failure / (1 - p))
}
bin_alg_obs_info <- function(p, n_success, n_failure) {
  return(n_success / p ^ 2 + n_failure / (1 - p) ^ 2)
}
wvs(loglik = bin_loglik, theta0 = 0.5, theta_range = c(0.1, 0.7),
     theta_mle = 7 / 20, n_success = 7, n_failure = 13,
     alg_score = bin_alg_score, alg_obs_info = bin_alg_obs_info)
```

# Index

- Beta, [12](#)
- Cauchy, [12](#)
- Central Limit Theorem: sampling distribution of a sample mean, [18](#)
- clt, [2](#), [16](#)
- continuous, [4](#), [15](#), [16](#)
- Continuous distributions, [18](#)
- correlation, [6](#), [13](#), [16](#)
  
- dFPearson (FPearson), [13](#)
- discrete, [8](#), [15](#), [16](#)
- Discrete distributions, [18](#)
  
- ett, [10](#), [16](#)
- Extremal Types Theorem: sampling distribution of a sample maximum, [18](#)
  
- FDist, [12](#)
- FPearson, [13](#)
  
- GammaDist, [3](#), [11](#)
- gev, [2](#), [5](#), [10](#)
- gp, [2](#), [5](#), [10](#), [12](#)
  
- legend, [3](#), [11](#)
- lev\_inf, [14](#), [16](#)
- Leverage and influence in simple linear regression, [18](#)
  
- movies, [4](#), [6](#), [7](#), [9](#), [12](#), [15](#), [15](#), [17](#), [21](#)
  
- Normal, [19](#)
  
- optim, [20](#)
  
- Pearson, [13](#)
- Pearson product moment correlation coefficient, [18](#)
  
- pFPearson (FPearson), [13](#)
- plot, [5](#), [9](#)
  
- qFPearson (FPearson), [13](#)
  
- revdbayes, [2](#), [5](#), [10](#)
- rFPearson (FPearson), [13](#)
- rp.button, [3](#), [7](#), [11](#)
- rp.cartoons, [15](#)
- rp.control, [9](#)
- rp.doublebutton, [3](#), [6](#), [7](#), [9](#), [11](#)
- rug, [3](#), [11](#)
  
- shypo, [16](#), [16](#)
- signif, [20](#)
- smovie, [4](#), [6](#), [7](#), [9](#), [12](#), [15–17](#), [18](#), [21](#)
- smovie-package (smovie), [18](#)
  
- TDist, [12](#)
- Testing simple hypotheses, [18](#)
  
- user-friendly menu panel, [18](#)
  
- Wald, Wilks and Score tests, [18](#)
- wws, [16](#), [19](#)