Package ‘PDQutils’

March 18, 2017

Maintainer  Steven E. Pav <shabbychef@gmail.com>

Version  0.1.6

Date  2017-03-17

License  LGPL-3

Title  PDQ Functions via Gram Charlier, Edgeworth, and Cornish Fisher Approximations

BugReports  https://github.com/shabbychef/PDQutils/issues

Description  A collection of tools for approximating the ‘PDQ’ functions (respectively, the cumulative distribution, density, and quantile) of probability distributions via classical expansions involving moments and cumulants.

Depends  R (>= 3.0.2)

Imports  orthopolynom, moments

Suggests  ggplot2, reshape2, testthat, knitr

URL  https://github.com/shabbychef/PDQutils

VignetteBuilder  knitr

Collate  'cornish_fisher.r' 'edgeworth.r' 'gram_charlier.r' 'moments.r'
  'PDQutils.r'

RoxygenNote  5.0.1

NeedsCompilation  no

Author  Steven E. Pav [aut, cre]

Repository  CRAN

Date/Publication  2017-03-18 08:37:53 UTC

R topics documented:

AS269 .................................................. 2
cumulant2moment ........................................ 4
dapx_edgeworth ........................................... 5
**Description**

Lee and Lin’s Algorithm AS269 for higher order Cornish Fisher quantile approximation.

**Usage**

\[
\text{AS269}(z, \text{cumul}, \text{order.max}=\text{NULL}, \text{all.ords}=\text{FALSE})
\]

**Arguments**

- **z**
  - the quantiles of the normal distribution. an atomic vector.
- **cumul**
  - the standardized cumulants of order 3, 4, ..., k. an atomic vector.
- **order.max**
  - the maximum order approximation, must be greater than \(\text{length(cumul)}+2\). We assume the cumulants have been adjusted to reflect that the random variable has unit variance (‘standardized cumulants’)
- **all.ords**
  - a logical value. If TRUE, then results are returned as a matrix, with a column for each order of the approximation. Otherwise the results are a matrix with a single column of the highest order approximation.

**Details**

The Cornish Fisher approximation is the Legendre inversion of the Edgeworth expansion of a distribution, but ordered in a way that is convenient when used on the mean of a number of independent draws of a random variable.

Suppose \(x_1, x_2, \ldots, x_n\) are \(n\) independent draws from some probability distribution. Letting

\[
X = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} x_i,
\]

the Central Limit Theorem assures us that, assuming finite variance,

\[
X \to \mathcal{N}(\sqrt{n} \mu, \sigma),
\]

with convergence in \(n\).
The Cornish Fisher approximation gives a more detailed picture of the quantiles of $X$, one that is arranged in decreasing powers of $\sqrt{n}$. The quantile function is the function $q(p)$ such that $P(X \leq q(p)) = q(p)$. The Cornish Fisher expansion is

$$q(p) = \sqrt{n} \mu + \sigma \left( z + \sum_{3 \leq j} c_j f_j(z) \right),$$

where $z = \Phi^{-1}(p)$, and $c_j$ involves standardized cumulants of the distribution of $x_i$ of order up to $j$. Moreover, the $c_j$ include decreasing powers of $\sqrt{n}$, giving some justification for truncation. When $n = 1$, however, the ordering is somewhat arbitrary.

**Value**

A matrix, which is, depending on all. or ds, either with one column per order of the approximation, or a single column giving the maximum order approximation. There is one row per value in $z$. Invalid arguments will result in return value NaN with a warning.

**Note**

A warning will be thrown if any of the $z$ are greater than 3.719017274 in absolute value; the traditional AS269 errors out in this case.

**Author(s)**

Steven E. Pav <shabbychef@gmail.com>

**References**


AS 269. [http://lib.stat.cmu.edu/apstat/269](http://lib.stat.cmu.edu/apstat/269)


**See Also**

qapx_cf

**Examples**

```r
test <- AS269(seq(-2,2,0.01),c(0,2,0,4))
test <- AS269(seq(-2,2,0.01),c(0,2,0,4))
# test with the normal distribution:
s.cumul <- c(0,0,0,0,0,0,0,0,0,0)
pv <- seq(0.001,0.999,0.001)
zv <- qnorm(pv)
```
cumulant2moment

Convert raw cumulants to moments.

Description

Conversion of a vector of raw cumulants to moments.

Usage

cumulant2moment(kappa)

Arguments

kappa  
a vector of the raw cumulants. The first element is the first cumulant, which is also the first moment.

Details

The 'raw' cumulants $\kappa_i$ are connected to the 'raw' (uncentered) moments, $\mu'_i$ via the equation

$$\mu'_n = \kappa_n + \sum_{m=1}^{n-1} \binom{n-1}{m-1} \kappa_m \mu'_{n-m}$$

Value

a vector of the raw moments. The first element of the input shall be the same as the first element of the output.
**dapx_edgeworth**

**Author(s)**

Steven E. Pav &lt;shabbychef@gmail.com&gt;

**See Also**

`moment2cumulant`

**Examples**

```r
# normal distribution, mean 0, variance 1
n.mom <- cumulant2moment(c(0,1,0,0,0,0))
# normal distribution, mean 1, variance 1
n.mom <- cumulant2moment(c(1,1,0,0,0,0))
```

**Description**

Approximate the probability density or cumulative distribution function of a distribution via its raw cumulants.

**Usage**

```r
dapx_edgeworth(x, raw.cumulants, support=c(-Inf,Inf), log=FALSE)

papx_edgeworth(q, raw.cumulants, support=c(-Inf,Inf), lower.tail=TRUE, log.p=FALSE)
```

**Arguments**

- **x**: where to evaluate the approximate density.
- **raw.cumulants**: an atomic array of the 1st through kth raw cumulants of the probability distribution. The first cumulant is the mean, the second is the variance. The third is *not* the typical unitless skew.
- **support**: the support of the density function. It is assumed that the density is zero on the complement of this open interval.
- **log**: logical; if TRUE, densities f are given as log(f).
- **q**: where to evaluate the approximate distribution.
- **log.p**: logical; if TRUE, probabilities p are given as log(p).
- **lower.tail**: whether to compute the lower tail. If false, we approximate the survival function.
Details

Given the raw cumulants of a probability distribution, we can approximate the probability density function, or the cumulative distribution function, via an Edgeworth expansion on the standardized distribution. The derivation of the Edgeworth expansion is rather more complicated than that of the Gram Charlier approximation, involving the characteristic function and an expression of the higher order derivatives of the composition of functions; see Blinnikov and Moessner for more details. The Edgeworth expansion can be expressed succinctly as

\[ \sigma f(\sigma x) = \phi(x) + \phi(x) \sum_{1 \leq s} \sigma^s \sum_{\{k_m\}} H_{s+2r}(x)c_{k_m}, \]

where the second sum is over some partitions, and the constant \( c \) involves cumulants up to order \( s+2 \). Unlike the Gram Charlier expansion, of which it is a rearrangement, the Edgeworth expansion is arranged in increasing powers of the standard deviation \( \sigma \).

Value

The approximate density at \( x \), or the approximate CDF at \( q \).

Note

Monotonicity of the CDF is not guaranteed.

Author(s)

Steven E. Pav <shabbychef@gmail.com>

References


See Also

the Gram Charlier expansions, `dapx_gca`, `papx_gca`

Examples

```r
# normal distribution, for which this is silly
xvals <- seq(-2,2,length.out=501)
d1 <- dapx_edgeworth(xvals, c(0,1,0,0,0))
d2 <- dnorm(xvals)
d1 - d2

qvals <- seq(-2,2,length.out=501)
p1 <- papx_edgeworth(qvals, c(0,1,0,0,0))
p2 <- pnorm(qvals)
p1 - p2
```
Approximate density and distribution via Gram-Charlier A expansion.

Description

Approximate the probability density or cumulative distribution function of a distribution via its raw moments.

Usage

dapx_gca(x, raw.moments, support=NULL, basis=c('normal', 'gamma', 'beta', 'arcsine', 'wigner'), basepar=NULL, log=FALSE)
papx_gca(q, raw.moments, support=NULL, basis=c('normal', 'gamma', 'beta', 'arcsine', 'wigner'), basepar=NULL, lower.tail=TRUE, log.p=FALSE)

Arguments

x where to evaluate the approximate density.
raw.moments an atomic array of the 1st through kth raw moments of the probability distribution.
support the support of the density function. It is assumed that the density is zero on the complement of this open interval. This defaults to c(-Inf, Inf) for the normal basis, c(0, Inf) for the gamma basis, and c(0, 1) for the Beta, and c(-1, 1) for the arcsine and wigner.
basis the basis under which to perform the approximation. 'normal' gives the classical 'A' series expansion around the PDF and CDF of the normal distribution via Hermite polynomials. 'gamma' expands around a gamma distribution with parameters basepar$shape and basepar$scale. 'beta' expands around a beta distribution with parameters basepar$shape1 and basepar$shape2.
basepar the parameters for the base distribution approximation. If NULL, the shape and rate are inferred from the first two moments and/or from the support as appropriate.
log logical; if TRUE, densities f are given as log(f).
q where to evaluate the approximate distribution.
log.p logical; if TRUE, probabilities p are given as log(p).
lower.tail whether to compute the lower tail. If false, we approximate the survival function.
Details

Given the raw moments of a probability distribution, we can approximate the probability density function, or the cumulative distribution function, via a Gram-Charlier expansion on the standardized distribution. This expansion uses some weighting function, \( w \), typically the density of some 'parent' probability distribution, and polynomials, \( p_n \) which are orthogonal with respect to that weighting:

\[
\int_{-\infty}^{\infty} p_n(x)p_m(x)w(x)dx = h_n\delta_{mn}.
\]

Let \( f(x) \) be the probability density of some random variable, with cumulative distribution function \( F(x) \). We express

\[
f(x) = \sum_{n \geq 0} c_n p_n(x)w(x)
\]

The constants \( c_n \) can be computed from the known moments of the distribution.

For the Gram Charlier 'A' series, the weighting function is the PDF of the normal distribution, and the polynomials are the (probabilist’s) Hermite polynomials. As a weighting function, one can also use the PDF of the gamma distribution (resulting in generalized Laguerre polynomials), or the PDF of the Beta distribution (resulting in Jacobi polynomials).

Value

The approximate density at \( x \), or the approximate CDF at

Note

Monotonicity of the CDF is not guaranteed.

Author(s)

Steven E. Pav <shabbychef@gmail.com>

References


See Also

qapx_cf
Examples

# normal distribution:
xvals <- seq(-2,2,length.out=501)
d1 <- dapx_gca(xvals, c(0,1,0,3,0), basis='normal')
d2 <- dnorm(xvals)
# they should match:
d1 - d2

qvals <- seq(-2,2,length.out=501)
p1 <- papx_gca(qvals, c(0,1,0,3,0))
p2 <- pnorm(qvals)
p1 - p2

xvals <- seq(-6,6,length.out=501)
mu <- 2
sigma <- 3
raw.moments <- c(2,13,62,475,3182)
d1 <- dapx_gca(xvals, raw.moments, basis='normal')
d2 <- dnorm(xvals, mean=mu, sd=sigma)
## Not run:
plot(xvals,d1)
lines(xvals,d2,col='red')

## End(Not run)
p1 <- papx_gca(xvals, raw.moments, basis='normal')
p2 <- pnorm(xvals, mean=mu, sd=sigma)
## Not run:
plot(xvals,p1)
lines(xvals,p2,col='red')

## End(Not run)

# for a one-sided distribution, like the chi-square
chidf <- 30
ords <- seq(1,9)
raw.moments <- exp(ords * log(2) + lgamma((chidf/2) + ords) - lgamma(chidf/2))
xvals <- seq(0.3,10,length.out=501)
dlg <- dapx_gca(xvals, raw.moments, support=c(0,Inf), basis='gamma')
d2 <- dchisq(xvals,df=chidf)
## Not run:
plot(xvals,dlg)
lines(xvals,d2,col='red')

## End(Not run)
p1g <- papx_gca(xvals, raw.moments, support=c(0,Inf), basis='gamma')
p2 <- pchisq(xvals,df=chidf)
## Not run:
plot(xvals,p1g)
lines(xvals,p2,col='red')

## End(Not run)
# for a one-sided distribution, like the log-normal
mu <- 2
sigma <- 1
ords <- seq(1,8)
raw.moments <- exp(ords * mu + 0.5 * (sigma*ords)^2)
xvals <- seq(0.5,10,length.out=501)
dig <- dapx_gca(xvals, raw.moments, support=c(0,Inf), basis='gamma')
d2 <- dnorm(log(xvals),mean=mu,se=sigma) / xvals
## Not run:
plot(xvals,dig)
lines(xvals,d2,col='red')
## End(Not run)

moment2cumulant  

Convert moments to raw cumulants.

Description

Conversion of a vector of moments to raw cumulants.

Usage

moment2cumulant(moms)

Arguments

moms  
a vector of the moments. The first element is the first moment (the mean). If centered moments are given, the first moment must be zero. If raw moments are given, the first moment must be the mean.

Details

The 'raw' cumulants $\kappa_i$ are connected to the 'raw' (uncentered) moments, $\mu'_i$ via the equation

$$\kappa_n = \mu'_n - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \kappa_m \mu'_{n-m}$$

Note that this formula also works for central moments, assuming the distribution has been normalized to zero mean.

Value

a vector of the cumulants. The first element of the input shall be the same as the first element of the output.
Note

The presence of a NA or infinite value in the input will propagate to the output.

Author(s)

Steven E. Pav <shabbychef@gmail.com>

See Also

cumulant2moment

Examples

# normal distribution, mean 0, variance 1
n.cum <- moment2cumulant(c(0,1,0,3,0,15))
# normal distribution, mean 1, variance 1
n.cum <- moment2cumulant(c(1,2,4,10,26))
# exponential distribution
lambda <- 0.7
n <- 1:6
e.cum <- moment2cumulant(factorial(n) / (lambda^n))

PDQ Functions via Gram Charlier, Edgeworth, and Cornish Fisher Approximations

Description

PDQ Functions via Gram-Charlier, Edgeworth, and Cornish Fisher Approximations

Gram Charlier and Edgeworth Expansions

Given the raw moments of a probability distribution, we can approximate the probability density function, or the cumulative distribution function, via a Gram-Charlier 'A' expansion on the standardized distribution.

Suppose \( f(x) \) is the probability density of some random variable, and let \( F(x) \) be the cumulative distribution function. Let \( He_j(x) \) be the \( j \)th probabilist’s Hermite polynomial. These polynomials form an orthogonal basis, with respect to the function \( w(x) \) of the Hilbert space of functions which are square \( w \)-weighted integrable. The weighting function is \( w(x) = e^{-x^2/2} = \sqrt{2\pi}\phi(x) \). The orthogonality relationship is

\[
\int_{-\infty}^{\infty} He_i(x)He_j(x)w(x)dx = \sqrt{2\pi}j!\delta_{ij}.
\]

Expanding the density \( f(x) \) in terms of these polynomials in the usual way (abusing orthogonality) one has

\[
f(x) \approx \sum_{0 \leq j} \frac{He_j(x)}{j!}\phi(x) \int_{-\infty}^{\infty} f(z)He_j(z)dz.
\]
The cumulative distribution function is 'simply' the integral of this expansion. Abusing certain facts regarding the PDF and CDF of the normal distribution and the probabilist’s Hermite polynomials, the CDF has the representation

$$F(x) = \Phi(x) - \sum_{1 \leq j} \frac{He_{j-1}(x)}{j!} \phi(x) \int_{-\infty}^{\infty} f(z)He_j(z)dz.$$ 

These series contain coefficients defined by the probability distribution under consideration. They take the form

$$c_j = \frac{1}{j!} \int_{-\infty}^{\infty} f(z)He_j(z)dz.$$ 

Using linearity of the integral, these coefficients are easily computed in terms of the coefficients of the Hermite polynomials and the raw, uncentered moments of the probability distribution under consideration. Note that it may be the case that the computation of these coefficients suffers from bad numerical cancellation for some distributions, and that an alternative formulation may be more numerically robust.

**Generalized Gram Charlier Expansions**

The Gram Charlier 'A' expansion is most appropriate for random variables which are vaguely like the normal distribution. For those which are like another distribution, the same general approach can be pursued. One needs only define a weighting function, $w$, which is the density of the 'parent' probability distribution, then find polynomials, $p_n(x)$ which are orthogonal with respect to $w$ over its support. One has

$$f(x) = \sum_{0 \leq j} p_j(x)w(x) \frac{1}{h_j} \int_{-\infty}^{\infty} f(z)p_j(z)dz.$$ 

Here $h_j$ is the normalizing constant:

$$h_j = \int w(z)p_j^2(z)dz.$$ 

One must then use facts about the orthogonal polynomials to approximate the CDF. Another approach to arrive at the same computation is described by Berberan-Santos.

**Cornish Fisher Approximation**

The Cornish Fisher approximation is the Legendre inversion of the Edgeworth expansion of a distribution, but ordered in a way that is convenient when used on the mean of a number of independent draws of a random variable.

Suppose $x_1, x_2, \ldots, x_n$ are $n$ independent draws from some probability distribution. Letting

$$X = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} x_i,$$ 

the Central Limit Theorem assures us that, assuming finite variance,

$$X \to N(\sqrt{n}\mu, \sigma),$$

with convergence in $n$. 
The Cornish Fisher approximation gives a more detailed picture of the quantiles of $X$, one that is arranged in decreasing powers of $\sqrt{n}$. The quantile function is the function $q(p)$ such that $P(X \leq q(p)) = q(p)$. The Cornish Fisher expansion is

$$q(p) = \sqrt{n \mu} + \sigma \left( z + \sum_{3 \leq j} c_j f_j(z) \right),$$

where $z = \Phi^{-1}(p)$, and $c_j$ involves standardized cumulants of the distribution of $x_i$ of order $j$ and higher. Moreover, the $c_j$ feature decreasing powers of $\sqrt{n}$, giving some justification for truncation. When $n = 1$, however, the ordering is somewhat arbitrary.

Legal Mumbo Jumbo

PDQutils is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Lesser General Public License for more details.

Note

This package is maintained as a hobby.

Author(s)

Steven E. Pav <shabbychef@gmail.com>

References


Description

News for package ‘PDQutils’.

**PDQutils Version 0.1.6 (2017-03-18)**

- Package maintenance–no new features.
- move github figures to location CRAN understands.

**PDQutils Version 0.1.5 (2016-09-18)**

- Package maintenance–no new features.
- Remove errant files from test directory.

**PDQutils Version 0.1.4 (2016-03-03)**

- Package maintenance–no new features.
- Incompatibilities in vignette with ggplot2 release.

**PDQutils Version 0.1.3 (2016-01-04)**

- Package maintenance–no new features.

**PDQutils Version 0.1.2 (2015-06-15)**

- Generalized Gram Charlier expansions.
- bugfixes.

**PDQutils Version 0.1.1 (2015-02-26)**

- Edgeworth expansions.

**PDQutils Initial Version 0.1.0 (2015-02-14)**

- first CRAN release.
Description

Approximate the quantile function of a distribution via its cumulants.

Usage

`qapx_cf(p, raw.cumulants, support=c(-Inf,Inf), lower.tail = TRUE, log.p = FALSE)`

Arguments

- **p**: where to evaluate the approximate distribution.
- **raw.cumulants**: an atomic array of the 1st through kth raw cumulants. The first value is the mean of the distribution, the second should be the variance of the distribution, the remainder are raw cumulants.
- **support**: the support of the density function. It is assumed that the density is zero on the complement of this open interval. This defaults to `c(-Inf,Inf)` for the normal basis, `c(0,Inf)` for the gamma basis, and `c(0,1)` for the Beta, and `c(-1,1)` for the arcsine and wigner.
- **lower.tail**: whether to compute the lower tail. If false, we approximate the survival function.
- **log.p**: logical; if TRUE, probabilities p are given as log(p).

Details

Given the cumulants of a probability distribution, we approximate the quantile function via a Cornish-Fisher expansion.

Value

The approximate quantile at p.

Note

Monotonicity of the quantile function is not guaranteed.

Author(s)

Steven E. Pav <shabbychef@gmail.com>
References


AS 269. http://lib.stat.cmu.edu/apstat/269


See Also
dapx_gca, papx_gca, AS269, rapx_cf

Examples

```r
# normal distribution:
pvals <- seq(0.001, 0.999, length.out=501)
q1 <- rapx_cf(pvals, c(0,1,0,0,0,0))
q2 <- qnorm(pvals)
q1 - q2
```

**rapx_cf**  
Approximate random generation via Cornish-Fisher expansion.

**Description**

Approximate random generation via approximate quantile function.

**Usage**

`rapx_cf(n, raw.cumulants, support=c(-Inf,Inf))`

**Arguments**

- `n`: number of observations. If `length(n) > 1`, the length is taken to be the number required.
- `raw.cumulants`: an atomic array of the 1st through kth raw cumulants. The first value is the mean of the distribution, the second should be the variance of the distribution, the remainder are raw cumulants.
- `support`: the support of the density function. It is assumed that the density is zero on the complement of this open interval. This defaults to `c(-Inf,Inf)` for the normal basis, `c(0,Inf)` for the gamma basis, and `c(0,1)` for the Beta, and `c(-1,1)` for the arcsine and wigner.
Details

Given the cumulants of a probability distribution, we approximate the quantile function via a Cornish-Fisher expansion.

Value

A vector of approximate draws.

Author(s)

Steven E. Pav <shabbychef@gmail.com>

See Also

qapx_cf

Examples

# normal distribution:
rl <- rapx_cf(1000, c(0,1,0,0,0,0))
Index

*Topic distribution
   AS269, 2
   cumulant2moment, 4
   dapx_edgeworth, 5
   dapx_gca, 7
   moment2cumulant, 10
   qapx Cf, 15
   rapx Cf, 16
*Topic package
   PDQutils, 11

AS269, 2, 16

cumulant2moment, 4, 11

dapx_edgeworth, 5
dapx_gca, 6, 7, 16

moment2cumulant, 5, 10

papx_edgeworth (dapx_edgeworth), 5
papx_gca, 6, 16
papx_gca (dapx_gca), 7
PDQutils, 11
PDQutils-NEWS, 14
PDQutils-package (PDQutils), 11

qapx Cf, 3, 8, 15, 17

rapx Cf, 16, 16