Using the PDQutils package

Steven E. Pav *

February 26, 2015

Abstract

Example computations via the **PDQutils** package are illustrated.

The **PDQutils** package provides tools for approximating the density, distribution, and quantile functions, and for generation of random variates of distributions whose cumulants and moments can be computed. The PDF and CDF are computed approximately via the Gram Charlier A series, while the quantile is computed via the Cornish Fisher approximation. [2, 5] The random generation function uses the quantile function and draws from the uniform distribution.

1 Gram Charlier Expansion

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 jth probabilist's Hermite polynomial. These polynomials form an orthogonal basis, with respect to the function $w(x) = e^{-x^2/2} = \sqrt{2\pi}\phi(x)$, of the Hilbert space of functions which are square integrable with w-weighting. [1, 22.2.15] The orthogonality relationship is

$$\int_{-\infty}^{\infty} He_i(x) He_j(x) w(x) \mathrm{d}x = \sqrt{2\pi} j! \delta_{ij},$$

where δ_{ij} is the Kronecker delta.

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

$$f(x) = \sum_{0 \le j} \frac{He_j(x)}{j!} \phi(x) \int_{-\infty}^{\infty} f(z) He_j(z) \mathrm{d}z.$$

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

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

*shabbychef@gmail.com

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

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

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.

2 Edgeworth Expansion

Another approximation of the probability density and cumulative distribution functions is the Edgeworth Expansions. These are expressed in terms of the cumulants of the distribution, and also include the Hermite polynomials. However, the derivation of the Edgeworth expansion is rather more complicated than of the Gram Charlier expansion. [2] The Edgeworth series for a zero-mean unit distribution is

$$f(x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)\left[1 + \sum_{1 \le s} \sigma^s \sum_{\{k_m\}} He_{s+2r}\left(x/\sigma\right) \prod_{1 \le m \le s} \frac{1}{k_m!}\left(\frac{S_{m+2}}{(m+2)!}\right)^{k_m}\right],$$

where the second sum is over partitions $\{k_m\}$ such that $k_1 + 2k_2 + \ldots + sk_s = s$, where $r = k_1 + k_2 + \ldots + k_s$, and where $S_n = \frac{\kappa_n}{\sigma^{2n-2}}$ is a semi-normalized cumulant.

3 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

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

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 $P\left(X \le q(p)\right) = q(p)$. The Cornish Fisher expansion is

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

where $z = \Phi^{-1}(p)$ is the normal p-quantile, 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.

4 An Example: Sum of Nakagamis

The Gram Charlier and Cornish Fisher approximations are most convenient when the random variable can be decomposed as the sum of a small number of independent random variables whose cumulants can be computed. For example, suppose $Y = \sum_{1 \le i \le k} \sqrt{X_i/\nu_i}$ where the X_i are independent central chi-square random variables with degrees of freedom $\nu_1, \nu_2, ..., \nu_k$. I will call this a 'snak' distribution, since each summand follows a Nakagami distribution. We can easily write code that generates variates from this distribution given a vector of the degrees of freedom:

```
rsnak <- function(n, dfs) {
    samples <- Reduce("+", lapply(dfs, function(k) {
        sqrt(rchisq(n, df = k)/k)
    }))
}
```

Let's take one hundred thousand draws from this distribution. A q-q plot of this sample against normality is shown in Figure 1. The normal model is fairly decent, although possibly unacceptable in the tails. Using a Cornish Fisher approximation, we can do better.

```
n.samp <- 1e+05
dfs <- c(8, 15, 4000, 10000)
set.seed(18181)
# now draw from the distribution
rvs <- rsnak(n.samp, dfs)
qqnorm(rvs)
qqline(rvs, col = "red")
```

Using the additivity property of cumulants, we can compute the cumulants of Y easily if we have the cumulants of the X_i. These in turn can be computed from the raw moments. The jth moment of a chi distribution with ν degrees of freedom has form

$$2^{j/2} \frac{\Gamma\left((\nu + j)/2\right)}{\Gamma\left(\nu/2\right)}.$$

The following function computes the cumulants of the 'snak' distribution:

```
# for the moment2cumulant function:
library(PDQutils)

# compute the first ord.max raw cumulants of the
# sum of Nakagami variates
snak_cumulants <- function(dfs, ord.max = 10) {
    # first compute the raw moments
    moms <- lapply(dfs, function(nu) {
```

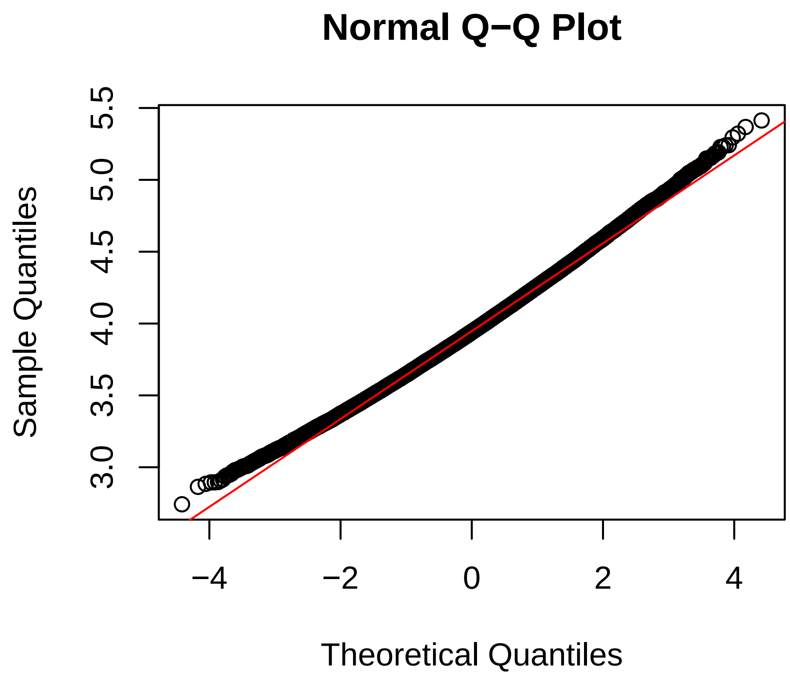

Figure 1: A q-q plot of 1e+05 draws from a sum of Nakagamis distribution is shown against normality.

```
        ords <- 1:ord.max
        moms <- 2^(ords/2) * exp(lgamma((nu + ords)/2) -
            lgamma(nu/2))
        # we are dividing the chi by sqrt the d.f.
        moms <- moms/(nu^(ords/2))
        moms
    })
    # turn moments into cumulants
    cumuls <- lapply(moms, moment2cumulant)
    # sum the cumulants
    tot.cumul <- Reduce("+", cumuls)
    return(tot.cumul)
}
```

We can now trivially implement the 'dpq' functions trivially using the Gram-Charlier and Cornish-Fisher approximations, via **PDQutils**, as follows:

```
library(PDQutils)

dsnak <- function(x, dfs, ord.max = 10, ...) {
    raw.moment <- cumulant2moment(snak_cumulants(dfs,
        ord.max))
    retval <- dapx_gca(x, raw.moment, support = c(0,
        Inf), ...)
    return(retval)
}
psnak <- function(q, dfs, ord.max = 10, ...) {
    raw.moment <- cumulant2moment(snak_cumulants(dfs,
        ord.max))
    retval <- papx_gca(q, raw.moment, support = c(0,
        Inf), ...)
    return(retval)
}
qsnak <- function(p, dfs, ord.max = 10, ...) {
    raw.cumul <- snak_cumulants(dfs, ord.max)
    retval <- qapx_cf(p, raw.cumul, support = c(0,
        Inf), ...)
    return(retval)
}
```

An alternative version of the PDF and CDF functions using the Edeworth expanion would look as follows:

```
dsnak_2 <- function(x, dfs, ord.max = 10, ...) {
    raw.cumul <- snak_cumulants(dfs, ord.max)
    retval <- dapx_edgeworth(x, raw.cumul, support = c(0,
        Inf), ...)
    return(retval)
}
psnak_2 <- function(q, dfs, ord.max = 10, ...) {
    raw.cumul <- snak_cumulants(dfs, ord.max)
    retval <- papx_edgeworth(q, raw.cumul, support = c(0,
        Inf), ...)
```

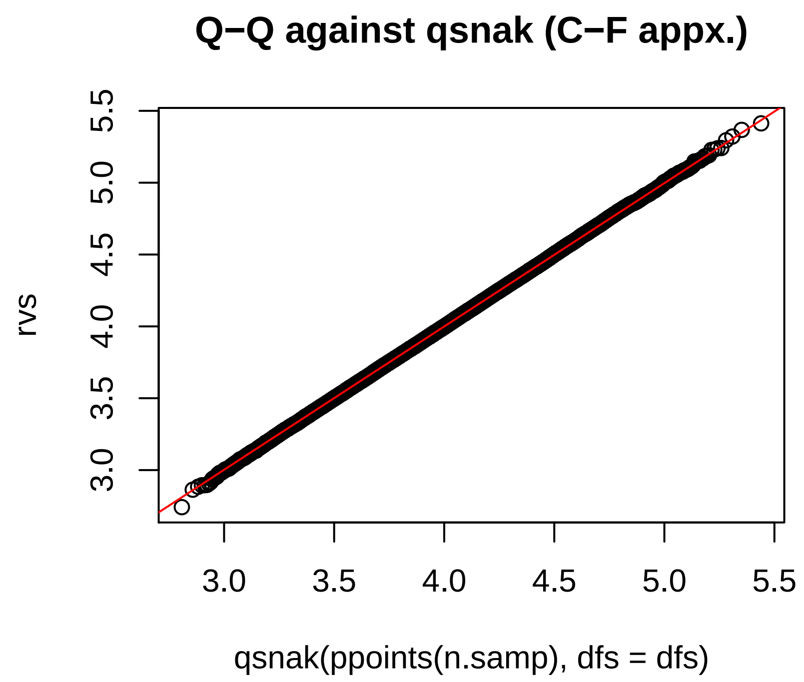

Figure 2: A q-q plot of 1e+05 draws from a sum of Nakagamis distribution is shown against quantiles from the 'qsnak' function.

```
    return(retval)
}
```

Using this approximate quantile function, the q-q plot looks straighter, as shown in Figure 2.

```
qqplot(qsnak(ppoints(n.samp), dfs = dfs), rvs, main = "Q-Q against qsnak (C-F appx.)")
qqline(rvs, distribution = function(p) qsnak(p, dfs),
    col = "red")
```

Note that the q-q plot uses the approximate quantile function, computed via the Cornish-Fisher expansion. We can test the Gram Charlier expansion by computing the approximate CDF of the random draws and checking that it is nearly uniform, as shown in Figure 3.

```
apx.p <- psnak(rvs, dfs = dfs)
require(ggplot2)
ph <- qplot(apx.p, stat = "ecdf", geom = "step")
print(ph)
```

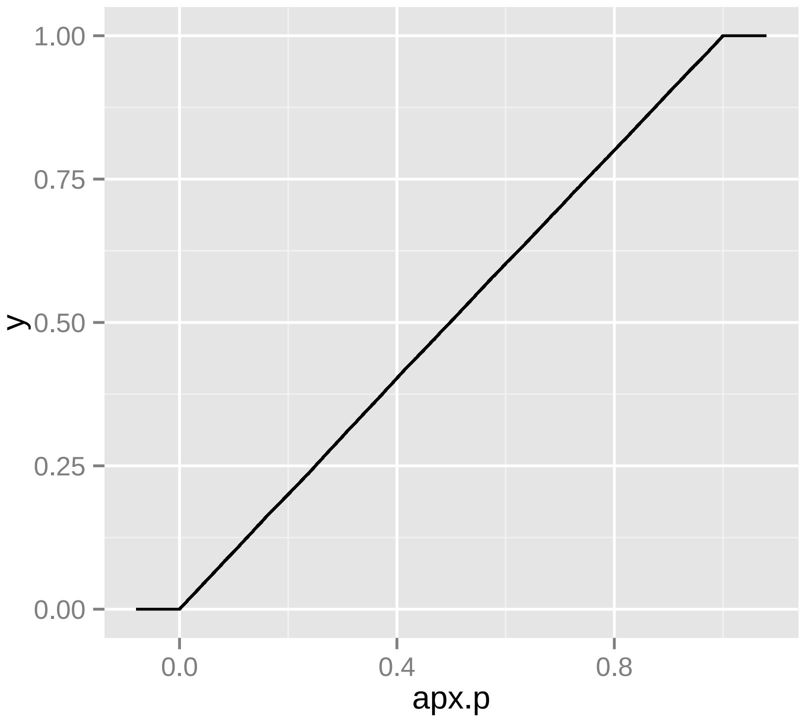

Figure 3: The empirical CDF of the approximate CDF of a sum of Nakagamis distribution on 1e+05 draws is shown.

5 A warning on convergence

Blinnikov and Moessner note that the the Gram Charlier expansion will actually diverge for some distributions when more terms in the expansion are considered, behaviour which is not seen for the Edgeworth expansion. [2] Here, we will replicate their example of the chi-square distribution with 5 degrees of freedom. Blinnikov and Moessner actually transform the chi-square to have zero mean and unit variance. They plot the true PDF of this normalized distribution, along with the 2- and 6-term Gram Charlier approximations, as shown in Figure 4.

```
# compute moments and cumulants:
df <- 5
max.ord <- 20
subords <- 0:(max.ord - 1)
raw.cumulants <- df * (2^subords) * factorial(subords)
raw.moments <- cumulant2moment(raw.cumulants)

# compute the PDF of the rescaled variable:
xvals <- seq(-sqrt(df/2) * 0.99, 6, length.out = 1001)
chivals <- sqrt(2 * df) * xvals + df
pdf.true <- sqrt(2 * df) * dchisq(chivals, df = df)

pdf.gca2 <- sqrt(2 * df) * dapx_gca(chivals, raw.moments = raw.moments[1:2],
    support = c(0, Inf))
pdf.gca6 <- sqrt(2 * df) * dapx_gca(chivals, raw.moments = raw.moments[1:6],
    support = c(0, Inf))

all.pdf <- data.frame(x = xvals, true = pdf.true, gca2 = pdf.gca2,
    gca6 = pdf.gca6)

# plot it by reshaping and ggplot'ing
require(reshape2)
arr.data <- melt(all.pdf, id.vars = "x", variable.name = "pdf",
    value.name = "density")

require(ggplot2)
ph <- ggplot(arr.data, aes(x = x, y = density, group = pdf,
    colour = pdf)) + geom_line()
print(ph)
```

Compare this with the 2 and 4 term Edgeworth expansions, shown in Figure 5.

```
# compute the PDF of the rescaled variable:
xvals <- seq(-sqrt(df/2) * 0.99, 6, length.out = 1001)
chivals <- sqrt(2 * df) * xvals + df
pdf.true <- sqrt(2 * df) * dchisq(chivals, df = df)

pdf.edgeworth2 <- sqrt(2 * df) * dapx_edgeworth(chivals,
    raw.cumulants = raw.cumulants[1:4], support = c(0,
        Inf))
pdf.edgeworth4 <- sqrt(2 * df) * dapx_edgeworth(chivals,
    raw.cumulants = raw.cumulants[1:6], support = c(0,
```

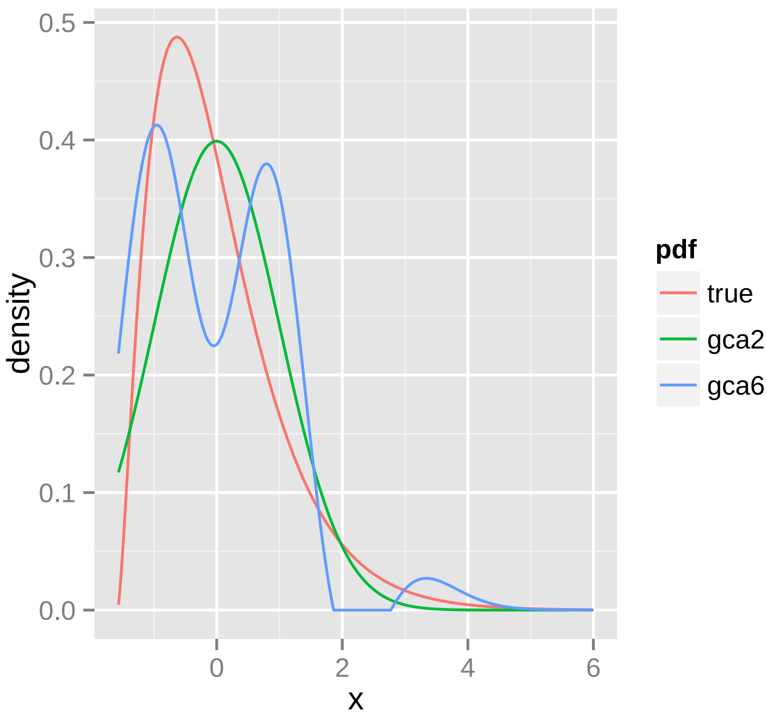

Figure 4: The true PDF of a normalized χ^2_5 distribution is shown, along with the 2- and 6-term Gram Charlier approximations. This replicates Figure 1 of Blinnikov and Moessner. [2]

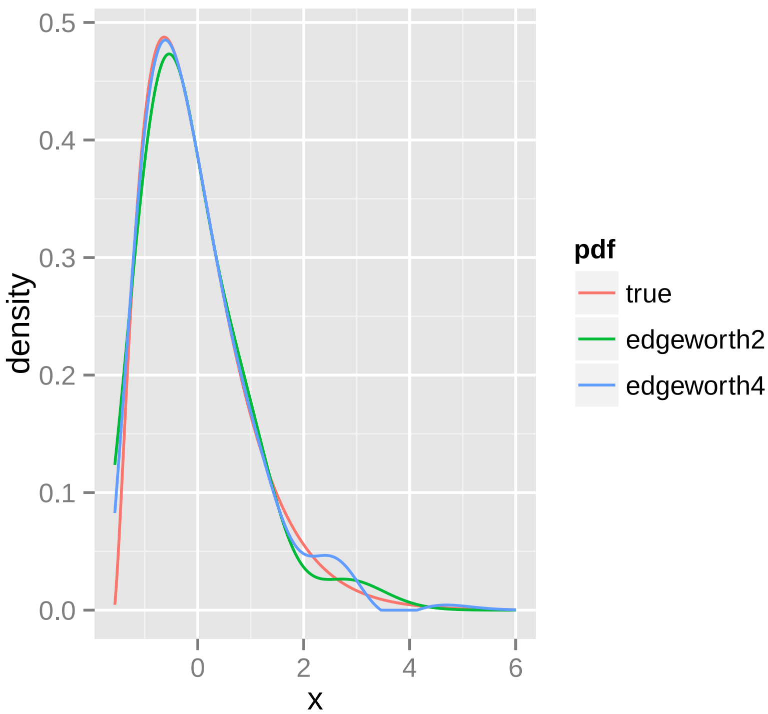

Figure 5: The true PDF of a normalized χ^2_5 distribution is shown, along with the 2- and 4-term Edgeworth expansions. This replicates Figure 6 of Blinnikov and Moessner. [2]

```
        Inf))

all.pdf <- data.frame(x = xvals, true = pdf.true, edgeworth2 = pdf.edgeworth2,
    edgeworth4 = pdf.edgeworth4)

# plot it by reshaping and ggplot'ing
require(reshape2)
arr.data <- melt(all.pdf, id.vars = "x", variable.name = "pdf",
    value.name = "density")

require(ggplot2)
ph <- ggplot(arr.data, aes(x = x, y = density, group = pdf,
    colour = pdf)) + geom_line()
print(ph)
```

References

[1] Milton Abramowitz and Irene˜A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York,

ninth dover printing, tenth gpo printing edition, 1964. URL http://people.math.sfu.ca/~cbm/aands/toc.htm.

[2] S.~Blinnikov and R.~Moessner. Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement*, 130:193–205, May 1998. doi: 10.1051/aas:1998221. URL http://arxiv.org/abs/astro-ph/9711239.

[3] P.~K. Cheah, D.~A.~S. Fraser, and N.~Reid. Some alternatives to Edgeworth. *Canadian Journal of Statistics*, 21(2):131–138, 1993. URL http://fisher.utstat.toronto.edu/dfraser/documents/174.pdf.

[4] Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Rearranging Edgeworth-Cornish-Fisher expansions. Privately Published, 2007. URL http://arxiv.org/abs/0708.1627.

[5] Stefan~R. Jaschke. The Cornish-Fisher-expansion in the context of Delta - Gamma - Normal approximations. Technical Report 2001,54, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 2001. URL http://www.jaschke-net.de/papers/CoFi.pdf.

[6] Yoong-Sin Lee and Ting-Kwong Lin. Algorithm AS 269: High order Cornish-Fisher expansion. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):pp. 233–240, 1992. ISSN 00359254. URL http://www.jstor.org/stable/2347649.

[7] Vladimir~V. Ulyanov. Cornish Fisher expansions. In Miodrag Lovric, editor, *International Encyclopedia of Statistical Science*, pages 312–315. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-04897-5. doi: 10.1007/978-3-642-04898-2_193. URL http://dx.doi.org/10.1007/978-3-642-04898-2_193.