

# Using the **PDQutils** package

Steven E. Pav \*

February 12, 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  $j$ th 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

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

where  $\delta_{ij}$  is the Kronecker delta.

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

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

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

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

---

\*shabbychef@gmail.com

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

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

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.

## 2 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  $x_1, x_2, \dots, x_n$  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,

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

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  $z = \Phi^{-1}(p)$  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.

## 3 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  $Y = \sum_{1 \leq i \leq k} \sqrt{X_i/\nu_i}$  where the  $X_i$  are independent central chi-square random variables with degrees of freedom  $\nu_1, \nu_2, \dots, \nu_k$ . 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:

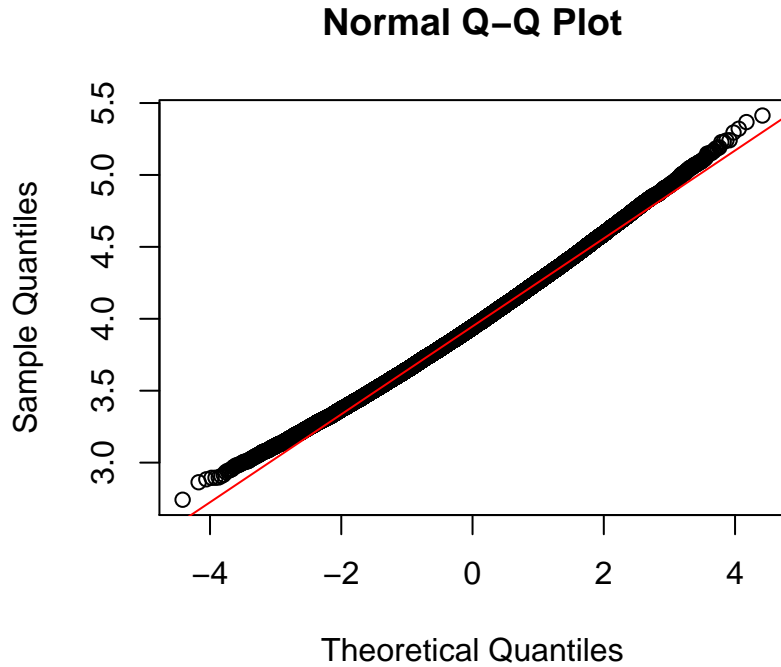


Figure 1: A q-q plot of  $1e+05$  draws from a sum of Nakagamis distribution is shown against normality.

```
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  $j$ th moment of a chi distribution with  $\nu$  degrees of

freedom has form

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

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) {
    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, ...)
  return(retval)
}

psnak <- function(q, dfs, ord.max = 10, ...) {
  raw.moment <- cumulant2moment(snak_cumulants(dfs,
    ord.max))
  retval <- papx_gca(q, raw.moment, ...)
  return(retval)
}

qsnak <- function(p, dfs, ord.max = 10, ...) {
  raw.cumul <- snak_cumulants(dfs, ord.max)
  retval <- qapx_cf(p, raw.cumul, ...)
  return(retval)
}
```

Using this approximate quantile function, the q-q plot looks straighter, as shown in [Figure 2](#).

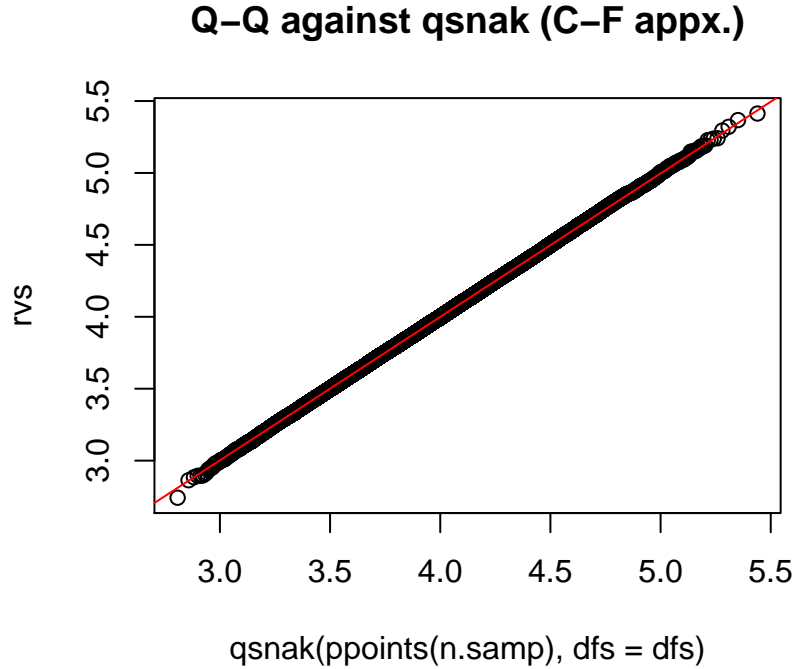


Figure 2: A q-q plot of  $1e+05$  draws from a sum of Nakagamis distribution is shown against quantiles from the ‘qsnak’ function.

```
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")
```

## 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. Rearrang-

- ing 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](http://dx.doi.org/10.1007/978-3-642-04898-2_193).