The **sadists** package

Steven E. Pav

March 20, 2017

Abstract

The **sadists** package includes 'dpqr' functions for some obscure distributions, mostly involving sums and ratios of (non-central) chi-squares, chis, and normals.

1 Introduction

The **sadists** package provides density, distribution, quantile and random generation functions (the 'dpqr' functions) for some obscure distributions. For all of these, the 'dpq' functions are approximated via the Edgeworth and Cornish-Fisher expansions. As such, this package is a showcase for the capabilities of the **PDQutils** package, which does the heavy lifting once the cumulants have been computed. [3]

It should be noted that the functions provided by **sadists** do *not* recycle their distribution parameters against the `x`, `p`, `q` or `n` parameters. This is in contrast to the common **R** idiom, and may cause some confusion. This is mostly for reasons of performance, but also because some of the distributions have vector-valued parameters; recycling over these would require the user to provide *lists* of parameters, which would be unpleasant.

First, a function which will evaluate the 'dpq' functions versus random draws of the variable:

```
require(ggplot2)
require(grid)
testf <- function(dpqr, nobs, ...) {
    rv <- sort(dpqr$r(nobs, ...))
    data <- data.frame(draws = rv, pvals = dpqr$p(rv,
        ...))
    text.size <- 6  # sigh

    # http://stackoverflow.com/a/5688125/164611
    p1 <- qplot(rv, geom = "blank") +
        geom_line(aes(y = ..density..,
            colour = "Empirical"), stat = "density") +
        stat_function(fun = function(x) {
            dpqr$d(x, ...)
        }, aes(colour = "Theoretical")) +
        geom_histogram(aes(y = ..density..),
            alpha = 0.3) + scale_colour_manual(name = "Density",
        values = c("red", "blue")) +
```

```
        theme(text = element_text(size = text.size)) +
        labs(title = "Density (tests dfunc)")

    # Q-Q plot
    p2 <- ggplot(data, aes(sample = draws)) +
        stat_qq(distribution = function(p) {
            dpqr$q(p, ...)
        }) + geom_abline(slope = 1, intercept = 0,
        colour = "red") + theme(text = element_text(size = text.size)) +
        labs(title = "Q-Q plot (tests qfunc)")

    # empirical CDF of the p-values;
    # should be uniform
    p3 <- ggplot(data, aes(sample = pvals)) +
        stat_qq(distribution = qunif) +
        geom_abline(slope = 1, intercept = 0,
            colour = "red") + theme(text = element_text(size = text.size)) +
        labs(title = "P-P plot (tests pfunc)")

    # Define grid layout to locate plots
    # and print each graph
    pushViewport(viewport(layout = grid.layout(2,
        2)))
    print(p1, vp = viewport(layout.pos.row = 1,
        layout.pos.col = 1:2))
    print(p2, vp = viewport(layout.pos.row = 2,
        layout.pos.col = 1))
    print(p3, vp = viewport(layout.pos.row = 2,
        layout.pos.col = 2))
}
```

2 Sum of (non-central) chi-squares to a power

Let $X_i \sim \chi^2_{\nu_i}(\delta_i)$ be independent non-central chi-square variates, where δ_i are the non-centrality parameters and ν_i are the degrees of freedom. Let w_i, p_i be given constants. Then

$$Y = \sum_i w_i X_i^{p_i}$$

follows a weighted sum of non-central chi-squares to a power distribution. This is not a common distribution. However, its cumulants can be easily computed, so the 'pdq' functions can be approximated by classical expansions. Moreover, its CDF and quantile functions can be used to compute those of the doubly non-central F, and it is related to the upsilon distribution.

```
require(sadists)
wts <- c(-1, 1, 3, -3)
df <- c(100, 200, 100, 50)
ncp <- c(0, 1, 0.5, 2)
pow <- c(1, 0.5, 2, 1.5)
testf(list(d = dsumchisqpow, p = psumchisqpow,
```

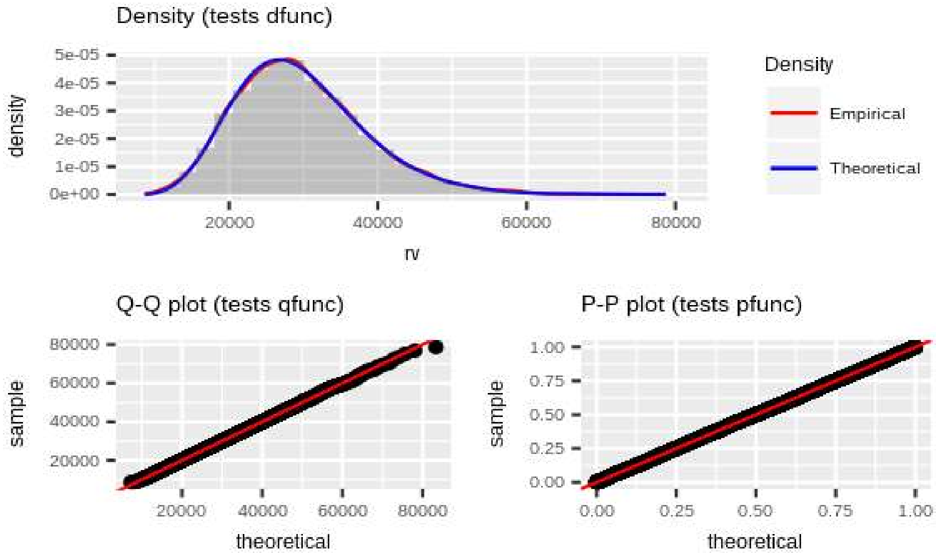

Figure 1: Confirming the dpqr functions of the sum of chi-squares to a power distribution.

```
    q = qsumchisqpow, r = rsumchisqpow),
    nobs = 2^14, wts, df, ncp, pow)
```

3 K-prime distribution

Let $X_i \sim \chi^2_{\nu_i}$ be chi-square random variables with ν_i degrees of freedom for $i = 1, 2$, independent of $Z \sim \mathcal{N}(0, 1)$, a standard normal. Suppose a, b are given constants. Then

$$Y = \frac{bZ + a\sqrt{X_1/\nu_1}}{\sqrt{X_2/\nu_2}}$$

follows a K-prime distribution with degrees of freedom $[\nu_1, \nu_2]$ and parameters a, b. [5] Depending on these four parameters, the K-prime generalizes the following:

- The normal distribution, when $b = 1, a = 0, \nu_2 = \infty$.
- The Lambda-prime distribution (see Section 4), when $b = 1, a \neq 0, \nu_2 = \infty$.
- The (central) t-distribution, when $b = 1, a = 0, \nu_2 < \infty$.
- The square-root of the F-distribution, when $b = 0, a = 1$.
- The (central) chi-distribution, when $b = 0, a = 1, \nu_2 = \infty$.

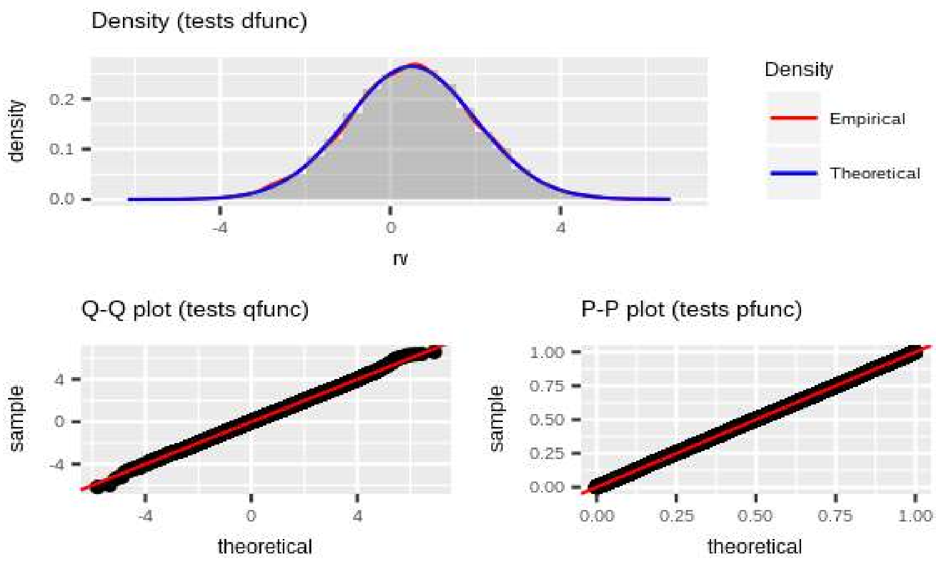

Figure 2: Confirming the dpqr functions of the K-prime distribution.

```r
require(sadists)
v1 <- 50
v2 <- 80
a <- 0.5
b <- 1.5
testf(list(d = dkprime, p = pkprime,
    q = qkprime, r = rkprime), nobs = 2^14,
    v1, v2, a, b)
```

4 Lambda prime distribution

Let $X \sim \chi^2_\nu$ be a chi-square random variable with ν degrees of freedom, independent of $Z \sim \mathcal{N}(0, 1)$, a standard normal. Then

$$Y = Z + t\sqrt{X/\nu}$$

follows a Lambda-prime distribution with parameter t and degrees of freedom ν. [1] It is a special case of the K-prime distribution (Section 3) and of the upsilon distribution (Section 5).

```r
require(sadists)
df <- 50
ts <- 1.5
testf(list(d = dlambdap, p = plambdap,
    q = qlambdap, r = rlambdap), nobs = 2^14,
    df, ts)
```

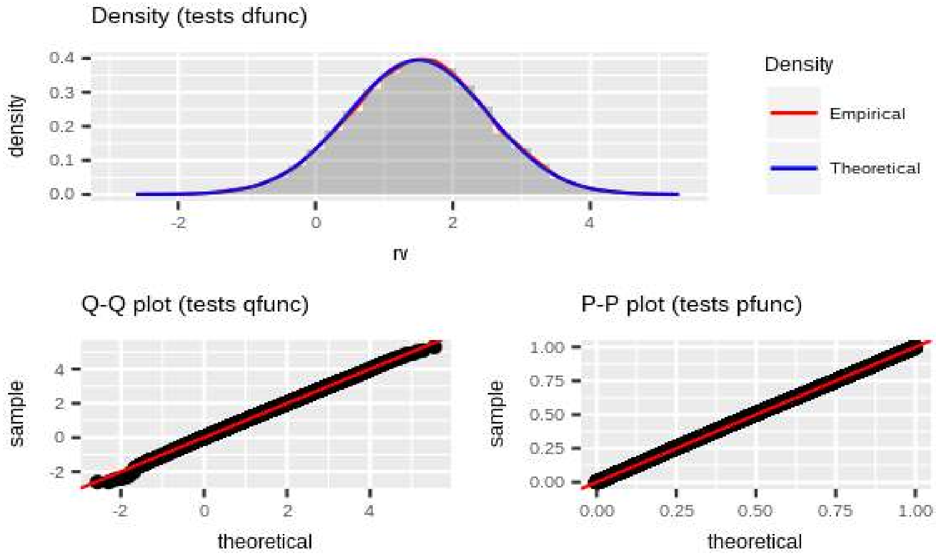

Figure 3: Confirming the dpqr functions of the Lambda-prime distribution.

5 Upsilon distribution

Let $X_i \sim \chi^2_{\nu_i}$ be independent central chi-square variates, where ν_i are the degrees of freedom. Let $Z \sim \mathcal{N}(0,1)$ be a standard normal, independent of the X_i. Let t_i be given constants. Then

$$Y = Z + \sum_i t_i \sqrt{X_i}$$

follows an upsilon distribution with parameter $[t_1, t_2, \ldots, t_k]$ and degrees of freedom $[\nu_1, \nu_2, \ldots, \nu_k]$.

```
require(sadists)
df <- c(30, 50, 100, 20, 10)
ts <- c(-3, 2, 5, -4, 1)
testf(list(d = dupsilon, p = pupsilon,
    q = qupsilon, r = rupsilon), nobs = 2^14,
    df, ts)
```

6 Doubly non-central F distribution

The doubly non-central F distribution generalizes the F distribution to the case where the denominator chi-square is non-central. For $i = 1, 2$, let $X_i \sim \chi^2_{\nu_i}(\delta_i)$ be independent non-central chi-square variates, where δ_i are the non-centrality parameters and ν_i are the degrees of freedom. Then

$$Y = \frac{X_1/\nu_1}{X_2/\nu_2}$$

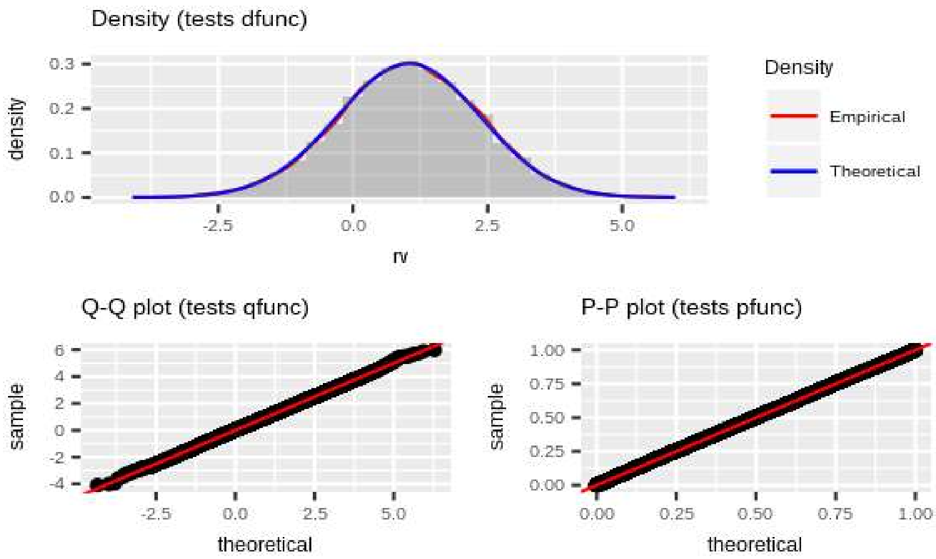

Figure 4: Confirming the dpqr functions of the upsilon distribution.

follows a doubly non-central F distribution.

```
require(sadists)
df1 <- 40
df2 <- 80
ncp1 <- 1.5
ncp2 <- 2.5
testf(list(d = ddnf, p = pdnf, q = qdnf,
    r = rdnf), nobs = 2^14, df1, df2,
    ncp1, ncp2)
```

7 Doubly non-central t distribution

The doubly non-central t distribution generalizes the t distribution to the case where the denominator chi-square is non-central. Let $X_2 \sim \chi^2_{\nu_2}(\delta_2)$ be a non-central chi-square variate, with non-centrality parameter δ_2 and ν_2 degrees of freedom. Let X_2 be independent of Z, a standard normal. Then

$$Y = \frac{Z + \delta_1}{\sqrt{X_2/\nu_2}}$$

follows a doubly non-central t distribution with degrees of freedom ν_2 and non-centrality parameters δ_1, δ_2. The square of a doubly non-central t is, up to scaling, a doubly non-central F, see Section 6.

```
require(sadists)
df <- 75
ncp1 <- 2
```

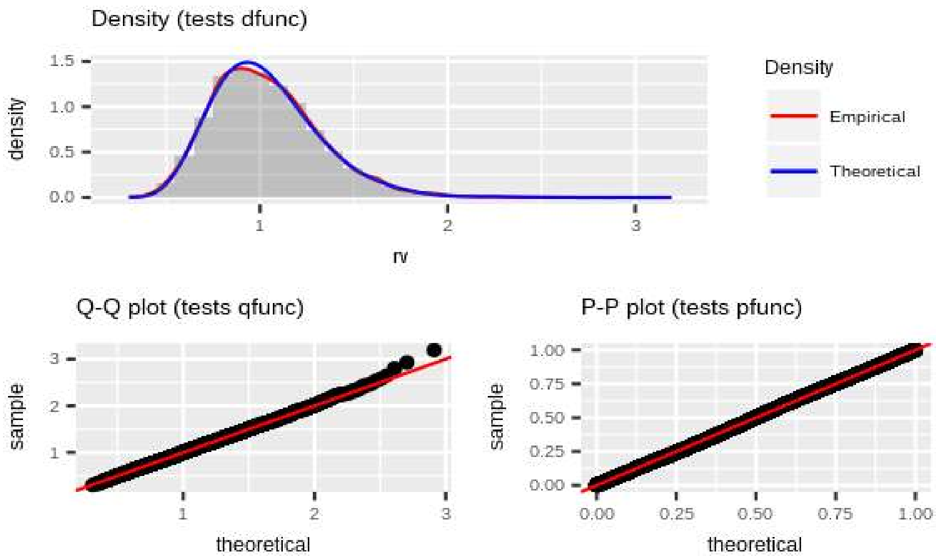

Figure 5: Confirming the dpqr functions of the doubly non-central F distribution.

```
ncp2 <- 3
testf(list(d = ddnt, p = pdnt, q = qdnt,
    r = rdnt), nobs = 2^14, df, ncp1,
    ncp2)
```

8 Doubly non-central Beta distribution

The doubly non-central Beta distribution generalizes the Beta distribution to the case where the denominator chi-square is non-central. For $i = 1, 2$, let $X_i \sim \chi^2_{\nu_i}(\delta_i)$ be independent non-central chi-square variates, where δ_i are the non-centrality parameters and ν_i are the degrees of freedom. Then

$$Y = \frac{X_1}{X_1 + X_2}$$

follows a doubly non-central Beta distribution. Note that

$$F = \frac{\nu_2}{\nu_1}\frac{Y}{1 - Y}$$

follows a doubly non-central F distribution. The 'PDQ' functions use this relationship.

```
require(sadists)
df1 <- 40
df2 <- 80
ncp1 <- 1.5
```

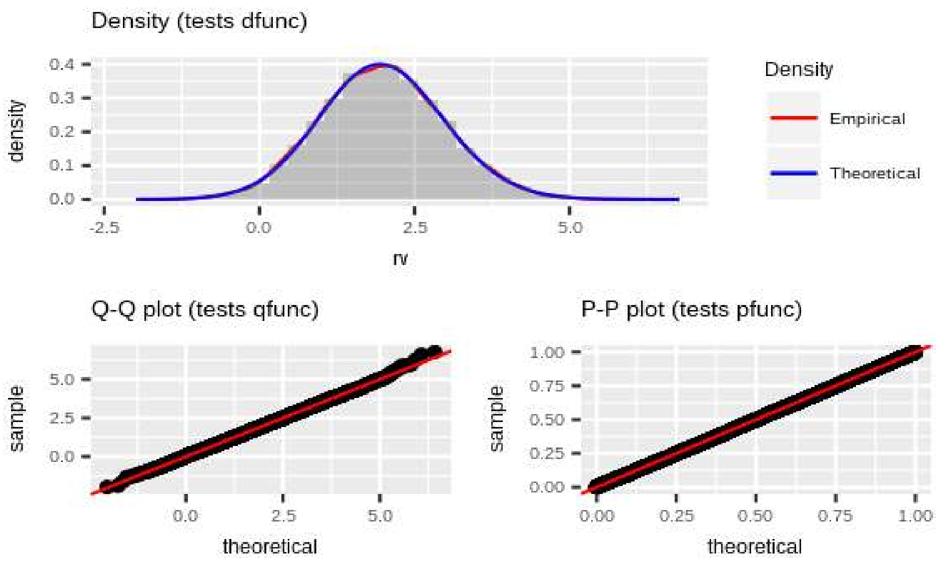

Figure 6: Confirming the dpqr functions of the doubly non-central t distribution.

```
ncp2 <- 2.5
testf(list(d = ddnbeta, p = pdnbeta,
    q = qdnbeta, r = rdnbeta), nobs = 2^14,
    df1, df2, ncp1, ncp2)
```

9 Doubly non-central Eta distribution

The doubly non-central Eta distribution is to the doubly non-central Beta what the doubly non-central t is to the doubly non-central F. Let $X_2 \sim \chi^2_{\nu_2}(\delta_2)$ be a non-central chi-square variate, with non-centrality parameter δ_2 and ν_2 degrees of freedom. Let X_2 be independent of Z, a standard normal. Then

$$Y = \frac{Z}{\sqrt{Z^2 + X_2}}$$

follows a doubly non-central Eta distribution with degrees of freedom ν_2 and non-centrality parameters δ_1, δ_2. The square of a doubly non-central Eta is a doubly non-central Beta, see Section 8. Note that

$$t = \sqrt{\nu_2}\frac{Y}{\sqrt{1 - Y^2}}$$

follows a doubly non-central t distribution. The 'PDQ' functions use this relationship.

```
require(sadists)
df <- 100
ncp1 <- 0.5
```

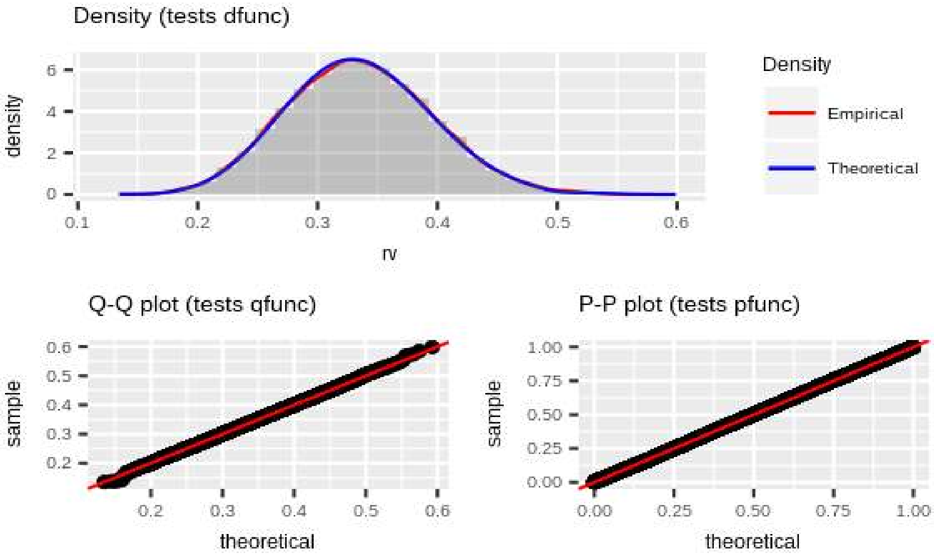

Figure 7: Confirming the dpqr functions of the doubly non-central Beta distribution.

```
ncp2 <- 2.5
testf(list(d = ddneta, p = pdneta, q = qdneta,
    r = rdneta), nobs = 2^14, df, ncp1,
    ncp2)
```

10 Sum of logs of (non-central) chi-squares

Let $X_i \sim \chi^2_{\nu_i}(\delta_i)$ be independent non-central chi-square variates, where δ_i are the non-centrality parameters and ν_i are the degrees of freedom. Let w_i be given constants. Then

$$Y = \sum_i w_i \log X_i$$

follows a weighted sum of log of non-central chi-squares distribution. This is not a common distribution. However, its cumulants can easily be computed. [4]

```
require(sadists)
wts <- c(5, -4, 10, -15)
df <- c(100, 200, 100, 50)
ncp <- c(0, 1, 0.5, 2)
testf(list(d = dsumlogchisq, p = psumlogchisq,
    q = qsumlogchisq, r = rsumlogchisq),
    nobs = 2^14, wts, df, ncp)
```

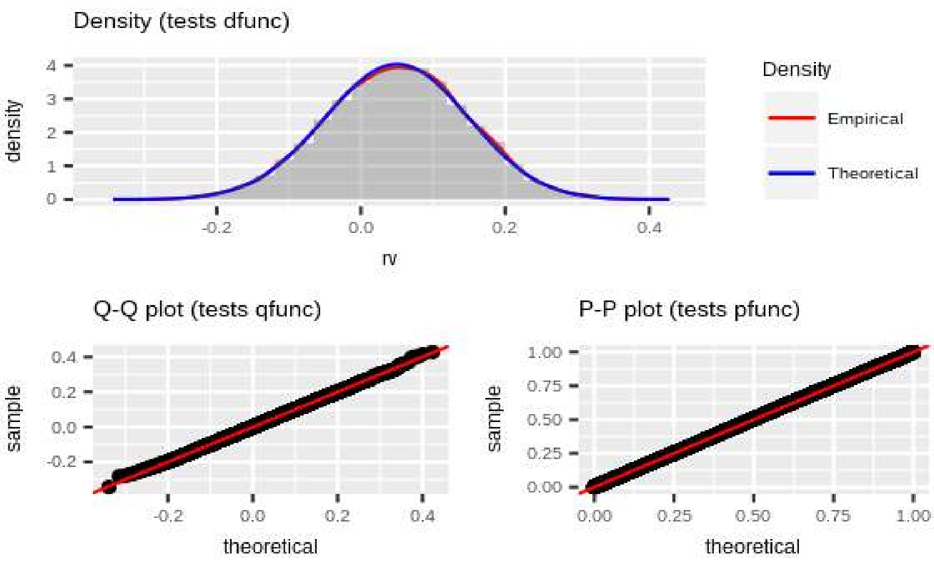

Figure 8: Confirming the dpqr functions of the doubly non-central Eta distribution.

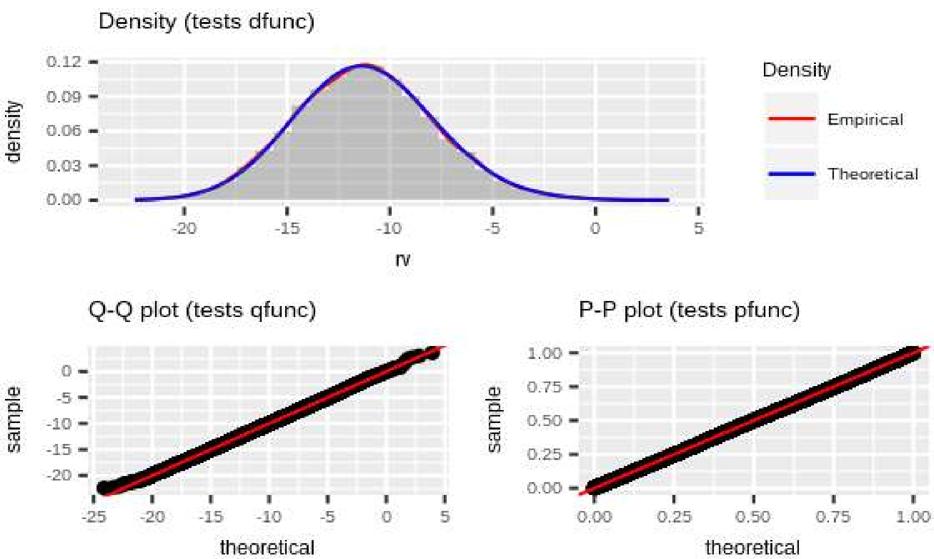

Figure 9: Confirming the dpqr functions of the sum of log of chi-squares distribution.

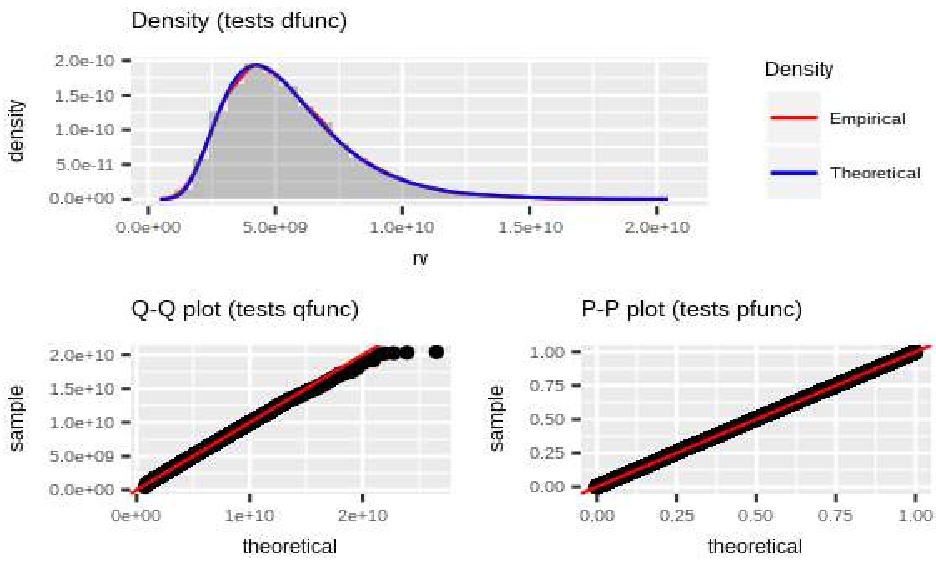

Figure 10: Confirming the dpqr functions of the product of chi-squares to a power distribution.

11 Product of (non-central) chi-squares to a power

Let $X_i \sim \chi^2_{\nu_i}(\delta_i)$ be independent non-central chi-square variates, where δ_i are the non-centrality parameters and ν_i are the degrees of freedom. Let p_i be given constants. Then

$$Y = \prod_i X_i^{p_i}$$

follows a product of non-central chi-squares to a power distribution. This is not a common distribution. The 'PDQ' functions are computed using a transform of the logs of chi-squares distribution, see Section 10.

```
require(sadists)
df <- c(100, 200, 100, 50)
ncp <- c(0, 1, 0.5, 2)
pow <- c(1, 0.5, 2, 1.5)
testf(list(d = dprodchisqpow, p = pprodchisqpow,
    q = qprodchisqpow, r = rprodchisqpow),
    nobs = 2^14, df, ncp, pow)
```

12 Product of doubly non-central F variates

Let $X_j \sim F_{\nu_{1,j},\nu_{2,j}}(\delta_{1,j}, \delta_{2,j})$ be independent doubly non-central F variates, where $\delta_{i,j}$ are the non-centrality parameters and $\nu_{i,j}$ are the degrees of freedom.

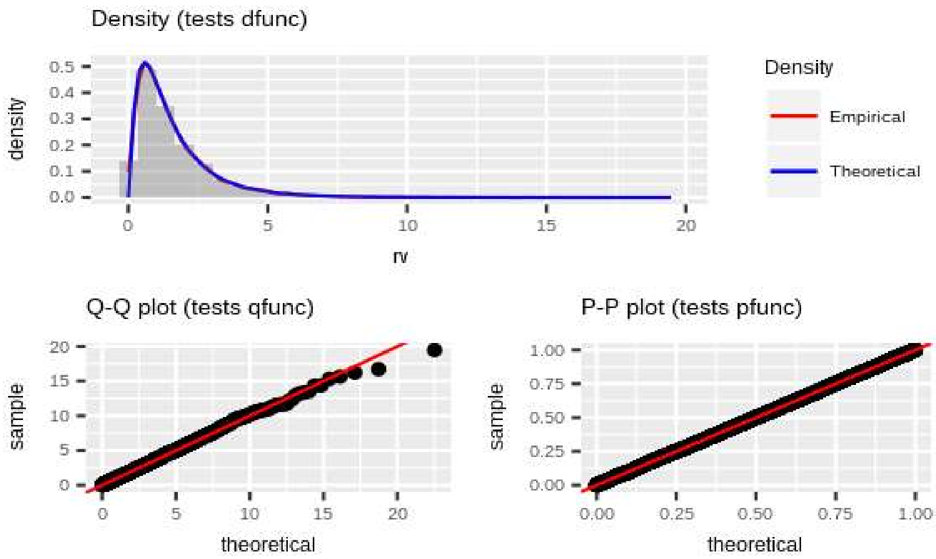

Figure 11: Confirming the dpqr functions of the product of doubly non-central Fs distribution.

Then

$$Y = \prod_j X_j$$

follows a product of doubly non-central Fs distribution. This is not a common distribution. The 'PDQ' functions are computed using a transform of the logs of chi-squares distribution, see Section 10.

```
require(sadists)
df1 <- c(10, 20, 5)
df2 <- c(1000, 500, 150)
ncp1 <- c(1, 0, 2.5)
ncp2 <- c(0, 1.5, 5)
testf(list(d = dproddnf, p = pproddnf,
    q = qproddnf, r = rproddnf), nobs = 2^14,
    df1, df2, ncp1, ncp2)
```

13 Product of normal variates

Let $Z_j \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right)$ be independent normal variates with means μ_j and variances σ_j^2. Then

$$Y = \prod_j X_j$$

follows a product of normals distribution. This is not a common distribution. When the coefficients of variation, σ_j/μ_j are large for some of the variates, the approximations given in this package tend to break down.

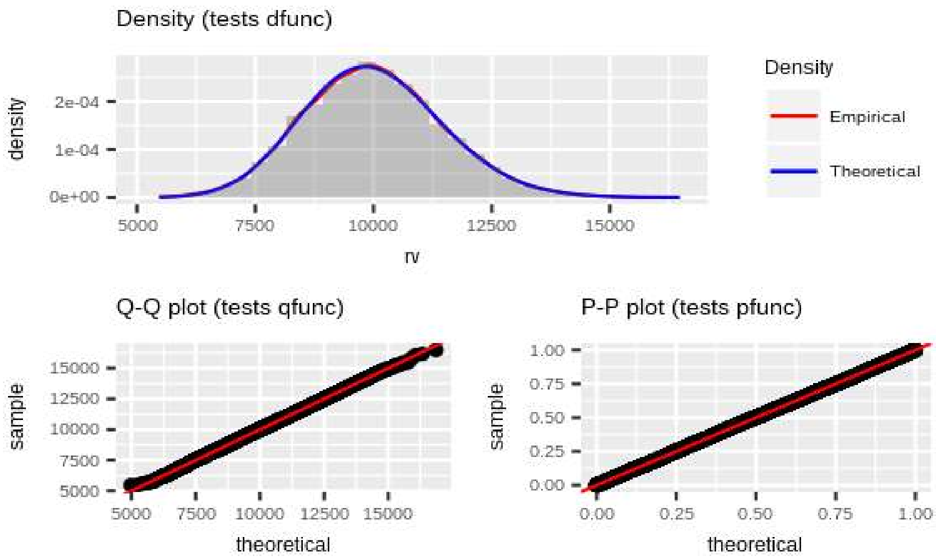

Figure 12: Confirming the dpqr functions of the product of normals distribution.

```
require(sadists)
mu <- c(100, 20, 5)
sigma <- c(10, 2, 0.2)
testf(list(d = dprodnormal, p = pprodnormal,
    q = qprodnormal, r = rprodnormal),
    nobs = 2^14, mu, sigma)
```

References

[1] Bruno Lecoutre. Two useful distributions for Bayesian predictive procedures under normal models. *Journal of Statistical Planning and Inference*, 79:93–105, 1999. URL http://www.researchgate.net/publication/242997315_Two_useful_distributions_for_Bayesian_predictive_procedures_under_normal_models/file/5046352b9cba661c83.pdf.

[2] M. Paolella. *Intermediate Probability: A Computational Approach*. Wiley, 2007. ISBN 9780470035054. URL http://books.google.com/books?id=9SHARfvyiR4C.

[3] Steven E. Pav. *PDQutils: PDQ Functions via Gram Charlier, Edgeworth, and Cornish Fisher Approximations*, 2015. URL https://github.com/shabbychef/PDQutils. R package version 0.1.1.

[4] Steven E. Pav. Moments of the log non-central chi-square distribution. Privately Published, 2015. URL http://arxiv.org/abs/1503.06266.

[5] Jacques Poitevineau and Bruno Lecoutre. Implementing Bayesian predictive procedures: The K-prime and K-square distributions. *Compu-*

tational Statistics and Data Analysis, 54(3):724–731, 2010. URL http://arxiv.org/abs/1003.4890v1.