Non-central Chi-Squared Probabilities – Algorithms in R

Martin Mächler
Seminar für Statistik
ETH Zurich

2019 ff (LaTeX'ed July 20, 2024)

Abstract

1 Introduction

1.1 Definition

Thie following paragraphs are basically verbatim from R 's `? Chisquare` help page:

The chi-squared distribution χ^2_n with `df`$= n \geq 0$ degrees of freedom has density

$$f_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \tag{1}$$

for $x > 0$ and $n \geq 0 \in \mathbb{R}$, i.e., not necessarily integer, where $f_0(x) := \lim_{n \to 0} f_n(x) = \delta_0(x)$, a point mass at zero, is not a density function proper, but a "δ distribution".

The mean and variance are n and $2n$.

The ***non-central*** chi-squared distribution with `df`$=$ n degrees of freedom and non-centrality parameter `ncp`$= \lambda$ has density

$$f_{n,\lambda}(x) = e^{-\lambda/2} \sum_{r=0}^{\infty} \frac{(\lambda/2)^r}{r!} f_{n+2r}(x), \tag{2}$$

for $x \geq 0$ and $f_n()$ defined in (1). For integer n, this is the distribution of the sum of squares of n normals each with variance one, λ being the sum of squares of the normal means; further,
$E(X) = n + \lambda$, $Var(X) = 2(n + 2 * \lambda)$, and $E((X - E(X))^3) = 8(n + 3 * \lambda)$.

Note that the degrees of freedom `df`$= n$, can be non-integer, and also $n = 0$ which is relevant for non-centrality $\lambda > 0$, see Johnson et al. (1994, chapter 29).

In that (noncentral, zero df) case, the distribution is a mixture of a point mass at $x = 0$ (of size `pchisq(0, df=0, ncp=ncp)`) and a continuous part, and `dchisq()` is *not* a density with respect to that mixture measure but rather the limit of the density for $df \to 0$.

χ^2 a special case of Γ

A central chi-squared distribution is special case of a Γ distribution: The chi-squared with n degrees of freedom is the same as a Gamma ('Γ') distribution with `shape` $\alpha = n/2$ and `scale = s`$= \sigma = 2$.

$$\chi^2_n \equiv \Gamma(\alpha = \frac{n}{2}, \sigma = 2) \tag{3}$$

Observed inaccuracies

Since Feb. 2006, the last paragraph of the *Details:* section says

> Note that `ncp` values larger than about 1e5 may give inaccurate results with many warnings for `pchisq` and `qchisq`.

where the '1e5' was extended to '*1e5 (and even smaller)*' in June 2019.

Since April 2010, the help page contains the note

> The code for non-zero `ncp` is principally intended to be used for moderate values of `ncp`: it will not be highly accurate, especially in the tails, for large values.

2 Non-central χ^2 probabilities: History of R 's `pnchisq.c`

The very early versions of R[1] already had R functions for the *non*-central chi-squared (called *"Chi-Square"* there) distribution, at the time functions separate from the central chi-squared, the four non-central functions where called `dnchisq()`, `pnchisq()`, etc, note the extra "n", and had their own separate help page (which was wrongly almost identical to the central chi-squared); e.g. `pnchisq()` with three arguments, already gave the correct result, e.g., for

```
> pnchisq(1,1,1)

[1] 0.4772499
```

In R version 0.50 "Alpha-4" (September 10, 1997), the help page was correct, and the 4 functions all where shown to have 3 arguments, e.g., `pnchisq(q, df, lambda)`.

The source code `R-<ver>/src/math/pnchisq.c` and then, for 0.62 and newer, using directory name `nmath/` had been practically unchanged from the earliest version up to version `0.61.3` (May 3, 1998). The algorithmic implementation in C was just summing up the Poisson weighted central term `term`, " `while (term >= acc)` " with the constant declaration where `acc` $:= 10^{-12}$:

```
double acc = 1.0e-12;

if (x <= 0.0)
 return 0.0;
df = n;
df1 = 0.5 * df;
x = 0.5 * x;
val = pgamma(x, df1, 1.0);
lambda2 = 0.5 * lambda;
c = 1.0;
t = 0.0;
do {
 t = t + 1.0;
 c = c * lambda2 / t;
 df1 = df1 + 1.0;
```

[1] the oldest versions of R still available, the pre-alpha version (before there were version numbers) with source file named `R-unix-src.tar.gz`, dated June 20, 1995, or the oldest still running version 0.16.1, February 11, 1997

```
  term = c * pgamma(x, df1, 1.0);
  val = val + term;
 }
 while (term >= acc);
 return val * exp(-lambda2);
```

Note that this just implements formula (2) replacing the infinite sum with a finite one, declaring convergence after the summand becomes smaller than 10^{-12} (which is not good enough out in the extreme tail). Note that the code does use the equivalence of the central chi-squared to the Gamma distribution with $\alpha = n/2$ and scale 2.

For R version 0.62 (1998-06-14), on the R level, the `*n*` versions of function names became deprecated and the noncentrality parameter was changed from `lambda` to `ncp` and added to the "non-`n`" version of the functions, e.g. R 's `pchisq()` became defined as

```
> pchisq
function (q, df, ncp = 0)
{
    if (missing(ncp))
        .Internal(pchisq(q, df))
    else .Internal(pnchisq(q, df, ncp))
}
```

and the source file `src/nmath/pnchisq.c` (with timestamp `1998-03-17 04:56` and a size of 2669 bytes) now did implement the algorithm AS 275 by Ding (1992). The `NEWS` entry (still in the R sources `doc/NEWS.0`) has been

```
  CHANGES IN R VERSION 0.62

NEW FEATURES

    .........
    .........

    o Some of the t, F, and chisq distribution/probability functions
 now allow a noncentrality parameter `ncp'.
```

But even then, the new `pnchisq.c` with AS 275 contained a comment

> Feb.1, 1998 (R 0.62 alpha); M.Maechler: still have INFINITE loop and/or bad precision in some cases.

At the time I had been pretty confident we'd eliminate these cases pretty quickly. In the meant time, many tweaks have been made, to a large extent by myself, and the code of today works accurately in many more cases than in early 1998. On the other hand, the help page has warned for years now that only moderate values of the noncentrality parameter `ncp` where feasible, and still in R 3.6.1 (July 2019), you can find calls to `pchisq()` which lead to an "infinite loop" (at least on 64-bit Linux and Windows), also for small values of ncp, e.g.,

```
> pchisq(1.00000012e200, df=1e200, ncp=100)
```

References

Ding, C. G. (1992). Algorithm AS 275: Computing the non-central χ^2 distribution function. *Applied Statistics — Journal of the Royal Statistical Society C*, 41(2):478–482.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1 of *Wiley Series in Probability and Mathematical Statistics*. Wiley, N. Y., 2 edition.