

# A Note on Computing Extreme Tail Probabilities of the Noncentral $T$ Distribution with Large Noncentrality Parameter

Viktor Witkovský<sup>a,\*</sup>

<sup>a</sup>*Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia*

## Abstract

The noncentral  $t$ -distribution is a generalization of the Student's  $t$ -distribution. In this paper we suggest an alternative approach for computing the cumulative distribution function (CDF) of the noncentral  $t$ -distribution which is based on a direct numerical integration of a well behaved function. With a double-precision arithmetic, the algorithm provides highly precise and fast evaluation of the extreme tail probabilities of the noncentral  $t$ -distribution, even for large values of the noncentrality parameter  $\delta$  and the degrees of freedom  $\nu$ . The implementation of the algorithm is available at the MATLAB Central, File Exchange: <http://www.mathworks.com/matlabcentral/fileexchange/41790-nctcdfvw>.

*Keywords:* Noncentral  $t$ -distribution, cumulative distribution function (CDF), noncentrality parameter, extreme tail probability, MATLAB algorithm

*2000 MSC:* 62E15, 62-04.

## 1. Introduction

The noncentral  $t$ -distribution, originally derived by R.A. Fisher [2], is a generalization of the Student's  $t$ -distribution [19]. Let  $Z$  be a random variable with a standard normal distribution, i.e.  $Z \sim N(0, 1)$ , and let  $Q$  be an independent chi-square distributed random variable with  $\nu$  degrees of freedom, i.e.  $Q \sim \chi_\nu^2$  with  $\nu > 0$ . Further, let  $\delta$  denote any real constant (the noncentrality parameter), then the probability distribution of the random variable

$$T = \frac{Z + \delta}{\sqrt{Q/\nu}}, \quad (1)$$

is called a noncentral  $t$ -distribution with  $\nu$  degrees of freedom and noncentrality parameter  $\delta$ , or simply written,  $T \sim t_{\nu, \delta}$ . If  $\delta = 0$  the noncentral  $t$ -distribution coincides with the central  $t$ -distribution.

The central as well as the noncentral  $t$ -distributions belong to the most frequently used distributions in statistics. The cumulative distribution function (CDF) of the noncentral  $t$ -distribution is used in the power analysis (as a part of statistical inference based on the normal linear models), see e.g. [11], including  $t$ -test as a special case. That is, the CDF of the noncentral  $t$ -distribution is used to evaluate the probability that a  $t$ -test will correctly reject a false null hypothesis on mean of a normal population  $N(\mu, \sigma^2)$ , i.e. the test of the null hypothesis  $H_0 : \mu \leq \mu_0$  against the alternative  $H_A : \mu > \mu_0$  based on small sample from this population, when the population mean  $\mu$  is actually greater than  $\mu_0$ ; that is, it gives the power of the  $t$ -test.

Broad applicability of the noncentral  $t$ -distribution is also in engineering, measurement science and metrology, quality control applications, as well as in financial mathematics. An interesting problem with important applications is derivation of the exact confidence interval for the noncentrality parameter  $\delta$  based upon a random sample from the normal distribution, or an equivalent problem of derivation of the confidence bounds for the coefficient of variation  $c_V = \frac{\sigma}{\mu}$ , see e.g. [12]. The noncentral  $t$ -distribution is also used for calculating the endpoints of the one-sided tolerance intervals (the tolerance limits) for a normal population. An application of tolerance intervals to manufacturing involves

\*Corresponding author. Tel.: +421 2 59104530; Fax: +421 2 54775943.

Email address: [witkovsky@savba.sk](mailto:witkovsky@savba.sk) (Viktor Witkovský)

comparing specification limits prescribed by the client with tolerance limits that cover a specified proportion of the population. The required values of the tolerance factors can be computed using the quantiles of the noncentral  $t$ -distribution. For more details on statistical tolerance intervals and the one-sided tolerance limits see e.g. [6, 13, 10].

The mathematical expressions for the CDF and the PDF (probability density function) of the noncentral  $t$  are rather complicated. There is not known closed form (analytical) expression of the CDF. So, it is typically given as a series expansion in terms of incomplete beta functions and/or other special functions. A comprehensive list of useful mathematical expressions and alternative representations can be found e.g. in [11].

Algorithms for numerical computation of the CDF, PDF, and the quantiles of the noncentral  $t$ -distribution are available in numerical libraries (e.g. BOOST C++ Libraries [15]), and different implementations are typically available also in standard statistical packages and/or programming environments (e.g. MATLAB [16], R [17], SAS [18], and Mathematica [20]).

However, applicability of the currently used implementations could be limited with respect to the speed and/or precision. This is true especially if the interest is in highly precise evaluation of the extreme tail probabilities of  $t$ -distributions with large values of the noncentrality parameter  $\delta$  and/or the degrees of freedom  $\nu$ .

The aim of this paper is to provide a brief overview of the basic properties of the noncentral  $t$ -distribution, useful for numerical evaluation of the CDF and PDF (consequently also for evaluation of the quantiles of the distribution), and comparisons of the current implementations, especially if used for evaluation of the extreme tail probabilities with large values of the noncentrality parameter.

As an alternative, we have developed a new algorithm based on direct numerical integration (by using standard Gauss-Kronrod quadrature) of a well behaved function which leads to highly precise and fast evaluation of the CDF for any combination of the input parameters  $x$ ,  $\nu$ , and  $\delta$ .

## 2. The noncentral $t$ -distribution

Here we shall briefly summarize some of the known properties of the noncentral  $t$ -distribution useful for numerical evaluation of the CDF, for more details see e.g. [1] and [11].

**Proposition 1.** *Let  $F_{t_{\nu,\delta}}(x) = \Pr(T \leq x)$  be the CDF of the random variable  $T \sim t_{\nu,\delta}$ , defined by (1) for any real  $x$ ,  $\delta$  and  $\nu > 0$ . Then the following properties hold true:*

1. *If  $x = 0$ , then*

$$F_{t_{\nu,\delta}}(x) = \Phi(-\delta), \quad (2)$$

*where  $\Phi(\cdot)$  is the CDF of standard normal distribution.*

2a. *If  $x > 0$ , then*

$$F_{t_{\nu,\delta}}(x) = \int_0^\infty \Phi\left(x \sqrt{\frac{q}{\nu}} - \delta\right) f_{\chi^2_\nu}(q) dq, \quad (3)$$

*where  $f_{\chi^2_\nu}(\cdot)$  is the PDF of chi-square distribution with  $\nu$  degrees of freedom.*

2b. *If  $x > 0$ , then*

$$\begin{aligned} F_{t_{\nu,\delta}}(x) &= \Phi(-\delta) + \int_{-\delta}^\infty \left(1 - F_{\chi^2_\nu}\left(\frac{\nu(z+\delta)^2}{x^2}\right)\right) \phi(z) dz \\ &= \Phi(-\delta) + \int_{-\delta}^\infty \Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) \phi(z) dz \\ &= 1 - \int_{-\delta}^\infty \Gamma_l\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) \phi(z) dz, \end{aligned} \quad (4)$$

*where  $F_{\chi^2_\nu}(\cdot)$  is the CDF of chi-squared distribution with  $\nu$  degrees of freedom,  $\phi(\cdot)$  is the PDF of standard normal distribution,  $\Gamma_u(\cdot, \cdot)$  (resp.  $\Gamma_l(\cdot, \cdot)$ ) denotes the upper (lower) regularized incomplete gamma function, i.e.  $\Gamma_u(a, x) = \frac{1}{\Gamma(a)} \int_x^\infty e^{-t} t^{a-1} dt$ ,  $\Gamma_l(a, x) = 1 - \Gamma_u(a, x)$ , and  $\Gamma(\cdot)$  is the gamma function.*

2c. If  $x > 0$ , then

$$F_{t_{v,\delta}}(x) = \Phi(-\delta) + \frac{1}{2} \sum_{i=0}^{\infty} \left\{ P_i I_y \left( i + \frac{1}{2}, \frac{\nu}{2} \right) + \frac{\delta}{\sqrt{2}} Q_i I_y \left( i + 1, \frac{\nu}{2} \right) \right\}, \quad (5)$$

where

$$P_i = \frac{(\delta^2/2)^i}{i!} e^{-\frac{\delta^2}{2}}, \quad Q_i = \frac{(\delta^2/2)^i}{\Gamma(i+3/2)} e^{-\frac{\delta^2}{2}}, \quad y = \frac{x^2}{\nu + x^2}, \quad (6)$$

and  $I_y(a, b)$  is the incomplete beta function.

3. If  $x < 0$ , then

$$F_{t_{v,\delta}}(x) = 1 - F_{t_{v,-\delta}}(-x). \quad (7)$$

**Proof** The CDF representation given in (5) has been proved by Guenther in [5]. The other properties can be derived directly from the definition of the noncentral  $t$ -variable (1). In particular,

$$F_{t_{v,\delta}}(0) = \Pr(T \leq 0) = \Pr\left(\frac{Z + \delta}{\sqrt{Q/v}} \leq 0\right) = \Pr(Z \leq -\delta) = \Phi(-\delta). \quad (8)$$

If  $x > 0$ , then we get

$$\begin{aligned} F_{t_{v,\delta}}(x) = \Pr(T \leq x) &= \Phi(-\delta) + \Pr(0 < T \leq x) \\ &= \Phi(-\delta) + \Pr\left(-\delta < Z \leq x \sqrt{\frac{Q}{v}} - \delta\right) \\ &= \Phi(-\delta) + E_{(Q)} \left[ \Phi\left(x \sqrt{\frac{Q}{v}} - \delta\right) \right] - \Phi(-\delta) \\ &= E_{(Q)} \left[ \Phi\left(x \sqrt{\frac{Q}{v}} - \delta\right) \right] = \int_0^{\infty} \Phi\left(x \sqrt{\frac{q}{v}} - \delta\right) f_{\chi_v^2}(q) dq, \end{aligned} \quad (9)$$

where  $Q \sim \chi_v^2$ ,  $f_{\chi_v^2}(\cdot)$  is the PDF of  $\chi_v^2$ -distribution, and  $E_{(Q)}(\cdot)$  denotes the expectation operator (with respect to the distribution of a random variable  $Q$ ). Similarly, we get

$$\begin{aligned} F_{t_{v,\delta}}(x) &= \Phi(-\delta) + \Pr\left(-\delta < Z \leq x \sqrt{\frac{Q}{v}} - \delta\right) \\ &= \Phi(-\delta) + \Pr\left(Z \leq x \sqrt{\frac{Q}{v}} - \delta \mid Z > -\delta\right) \\ &= \Phi(-\delta) + \Pr\left(\frac{\nu(Z + \delta)^2}{x^2} \leq Q \mid Z > -\delta\right) \\ &= \Phi(-\delta) + \left(1 - \Pr\left(Q \leq \frac{\nu(Z + \delta)^2}{x^2} \mid Z > -\delta\right)\right) \\ &= \Phi(-\delta) + E_{(Z > -\delta)} \left(1 - F_{\chi_v^2}\left(\frac{\nu(Z + \delta)^2}{x^2}\right)\right) \\ &= \Phi(-\delta) + \int_{-\delta}^{\infty} \Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z + \delta)^2}{2x^2}\right) \phi(z) dz \\ &= 1 - \int_{-\delta}^{\infty} \Gamma_l\left(\frac{\nu}{2}, \frac{\nu(z + \delta)^2}{2x^2}\right) \phi(z) dz \end{aligned} \quad (10)$$

where  $F_{\chi_v^2}(q) = \Gamma_l\left(\frac{\nu}{2}, \frac{q}{2}\right) = 1 - \Gamma_u\left(\frac{\nu}{2}, \frac{q}{2}\right)$  is the CDF of chi-squared distribution with  $\nu$  degrees of freedom, and  $\Gamma_l\left(\frac{\nu}{2}, \frac{q}{2}\right)$  (resp.  $\Gamma_u\left(\frac{\nu}{2}, \frac{q}{2}\right)$ ) denotes the lower (upper) regularized incomplete gamma function, and  $\phi(\cdot)$  is the PDF of standard normal distribution. Note that the representation holds true also for noninteger degrees of freedom,  $\nu > 0$ . Finally, if  $x < 0$ , we get

$$F_{t_{v,\delta}}(x) = \Pr(T \leq x) = \Pr\left(\frac{Z + \delta}{\sqrt{Q/v}} \leq x\right) = \Pr\left(\frac{-Z - \delta}{\sqrt{Q/v}} > -x\right)$$

$$= 1 - \Pr\left(\frac{Z - \delta}{\sqrt{Q/\nu}} \leq -x\right) = 1 - F_{t_{\nu, \delta}}(-x), \quad (11)$$

by using the symmetry of the distribution of the random variable  $Z \sim N(0, 1)$ .  $\square$

The CDF  $F_{t_{\nu, \delta}}(\cdot)$  can be directly used for computing the PDF of the noncentral  $t$  distribution,  $f_{t_{\nu, \delta}}(\cdot)$ , defined by  $f_{t_{\nu, \delta}}(x) = \partial F_{t_{\nu, \delta}}(x)/\partial x$ . In particular, the following holds true:

- If  $x = 0$ , then

$$f_{t_{\nu, \delta}}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{\delta^2}{2}}. \quad (12)$$

- If  $x \neq 0$ , then

$$f_{t_{\nu, \delta}}(x) = \frac{\nu}{x} \left\{ F_{t_{\nu+2, \delta}}\left(x \sqrt{1 + \frac{2}{\nu}}\right) - F_{t_{\nu, \delta}}(x) \right\}. \quad (13)$$

Based on that, the quantiles of the noncentral  $t$ -distribution, say  $x_p$ , can be calculated for any given  $p \in (0, 1)$ . In general, for any fixed combination of the three parameters (from the set  $\nu, \delta, x, p$ ), the remaining parameter can be calculated either directly, or via the usual root-finding techniques, by solving the equation  $F_{t_{\nu, \delta}}(x) = p$ .

### 3. Standard implementations for computing CDF of the noncentral $t$ -distribution

The standard algorithm implementations for computing the CDF of the noncentral  $t$ -distribution are based on its representation (5), which was originally derived by Guenther [5] and later implemented by Lenth [14].

Due to the recurrence properties of the incomplete beta function, the algorithm requires only two evaluations of  $I_y(a, b)$  and the rest is based on simple arithmetic operations. For most typical values of the input arguments  $x, \nu$ , and  $\delta$ , the algorithm is extremely fast and accurate.

The  $R$  implementation (function *pt* in [17]) is based on  $C$  version of the Lenth's algorithm with a restricted range of the noncentrality parameter,  $|\delta| \leq 37.62$ . Otherwise, the result is based on normal approximation,

$$F_{t_{\nu, \delta}}(x) = \Phi(z), \quad \text{where } z = \frac{x(1 - \frac{1}{4\nu}) - \delta}{\sqrt{1 + \frac{x^2}{2\nu}}}, \quad (14)$$

see [1], eqn. (26.7.10), p. 949, which can be rather poor for small  $\nu$ .

Algorithm based on (5) have been implemented also in MATLAB (function *nctcdf* in [16]) and in the BOOST C++ Libraries (function *non\_central\_t* in [15], see also [4]). The BOOST implementation is based on strategies suggested by Benton and Krishnamoorthy [3].

The BOOST function *non\_central\_t* has been tested for wide range of input parameters and compared with test data computed by arbitrary precision interval arithmetic (believed to be accurate to at least 50 decimal places, as declared in [15], and confirmed by a large test data set, kindly provided by J. Maddock [personal communication]). As the complexity of the algorithm based on the series expansion as given in (5) is dependent upon  $\delta^2$ , consequently, the time taken to evaluate the CDF increases rapidly for large noncentrality parameter,  $|\delta| > 500$ , likewise the accuracy decreases rapidly for very large  $\delta$ , see [15].

Moreover, unlike the  $R$  and MATLAB implementations, which compute correctly only the lower tail of the distribution, the BOOST algorithm computes also the upper tail of the distribution (which is important for correct evaluation of the extreme tail probabilities).

As presented in [18], SAS implementation (function *probt*) is based on numerical integration of the representation (3). For most typical values of the input arguments  $x, \nu$ , and  $\delta$ , the algorithm is fast and accurate (for most cases, typically all 14 reported significant digits are correct). However, for more extreme input arguments the algorithm may fail to converge to the prescribed accuracy, and in such case no output is provided by the function *probt*.

Implementation in Mathematica (function *CDF[NoncentralStudentTDistribution[ $\nu, \delta, x]$*  in [20]) is based on numerical integration (computed using Mathematica's high-precision arithmetic) of the noncentral  $t$  PDF function (which is given as an analytical function expressible by using the Hermite polynomials). The computational complexity of this algorithm quickly grows with large  $\nu$  and  $\delta$  and in such cases fail to converge.

#### 4. Algorithm *nctcdfw* based on direct numerical integration

As an alternative, here we suggest a new algorithm based on direct numerical integration (by using standard Gauss-Kronrod quadrature) of a well behaved function, based on expression 2b, which leads to highly precise and fast evaluation of the CDF for any combination of the input parameters  $x$ ,  $\nu$ , and  $\delta$ .

The algorithm has been implemented in MATLAB and its current version is available at the MATLAB Central, File Exchange: <http://www.mathworks.com/matlabcentral/fileexchange/41790-nctcdfw>.

The algorithm computes the lower tail, i.e.  $\Pr(T \leq x) = \Phi(-\delta) + \int_{-\delta}^{\infty} \Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) \phi(z) dz$ , if  $0 < x \leq \delta$ , otherwise, for  $x > 0$  and such that  $x > \delta$ , it computes the upper tail of the distribution,  $\Pr(T > x) = \int_{-\delta}^{\infty} \Gamma_l\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) \phi(z) dz$ .

Notice that in a double-precision arithmetic the integration range  $(-\delta, \infty)$  can be reduced to the limits  $[A_0, B_0]$  given by

$$[A_0, B_0] = [\max(-\delta, \Phi^{-1}(r_{\varepsilon_0})), -\Phi^{-1}(r_{\varepsilon_0})], \quad (15)$$

where  $r_{\varepsilon_0}$  is the minimum real non-zero number (the smallest positive normalized floating point number in IEEE double precision), i.e.  $r_{\varepsilon_0} = 2.2251 \times 10^{-308}$ . So,  $\Phi^{-1}(r_{\varepsilon_0}) = -37.5194$ .

The most important part of the algorithm is the method for subsequent (significant) reduction of the integration range  $[A_0, B_0]$ . For simplicity, we shall illustrate this only for the case of computing the lower tail of the distribution.

Notice that  $\Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right)$  approaches the value 1 for small values of  $z$  ( $z \rightarrow -\delta$ ) and 0 for large values of  $z$  ( $z \rightarrow +\infty$ ). If  $\Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) > 1 - \varepsilon_R$  for  $A_0 \leq z \leq A_1$ , where

$$A_1 = \sqrt{\frac{x^2 q_{\varepsilon_R}}{\nu}} - \delta, \quad (16)$$

with  $q_{\varepsilon_R}$  being the  $\varepsilon_R$ -quantile of the  $\chi_\nu^2$ -distribution (for  $q_{\varepsilon_R} \approx 0$  set  $q_{\varepsilon_R} = 0$ ), where  $\varepsilon_R$  is the required relative tolerance bound (in double-precision arithmetic we set  $\varepsilon_R = 10^{-16}$ ), then the integration range can be further reduced to

$$[A_1, B_1] = [\max(A_0, A_1), B_0], \quad (17)$$

and the CDF is approximated by

$$\Pr(T \leq x) \approx \Phi(A_1) + \int_{A_1}^{B_1} \Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) \phi(z) dz \quad (18)$$

Further reduction of the integration range  $[A_1, B_1]$  is possible since the integrand function

$$g(z) = \Gamma_u\left(\frac{\nu}{2}, \frac{\nu(z+\delta)^2}{2x^2}\right) \phi(z) \quad (19)$$

(typically) quickly fades out from its maximum value. For illustration, Figure 1 presents a typical graph of the integrand function  $g(z)$ , together with the optimally selected integration limits.

Based on that, further reduction of the integration range is given, and the final integration limits are given by

$$[A, B] = [\max(A_1, A_2), \min(B_1, B_2)], \quad (20)$$

where the limits  $A_2$  and  $B_2$  are given as the two possible solutions to the equation

$$g(A_2) = g(B_2) = \varepsilon_A, \quad (21)$$

where  $\varepsilon_A$  denotes the required absolute tolerance bound (which should be properly estimated, see below).

The integrand function  $g(z)$  defined by (19) and its modus can be effectively estimated based on the results and efficient approximations of the CDF and the quantiles of the chi-square distribution, as suggested by Inglot in [8].

In particular, let us denote  $q = \frac{\nu(z+\delta)^2}{x^2}$  and  $h(z) = \log(g(z))$ . Then, by using the lower bound for tails of the  $\chi_\nu^2$ -distribution, as derived in [8] and [9], i.e.  $\frac{1}{2}\mathcal{E}_\nu(q) \leq \Pr(\chi_\nu^2 > q) \leq \frac{1}{\sqrt{\pi}} \frac{q}{q-\nu+2} \mathcal{E}_\nu(q)$ , where

$$\mathcal{E}_\nu(q) = \exp\left\{-\frac{1}{2}\left(q - \nu - (\nu - 2) \log\left(\frac{q}{\nu}\right) + \log(\nu)\right)\right\},$$

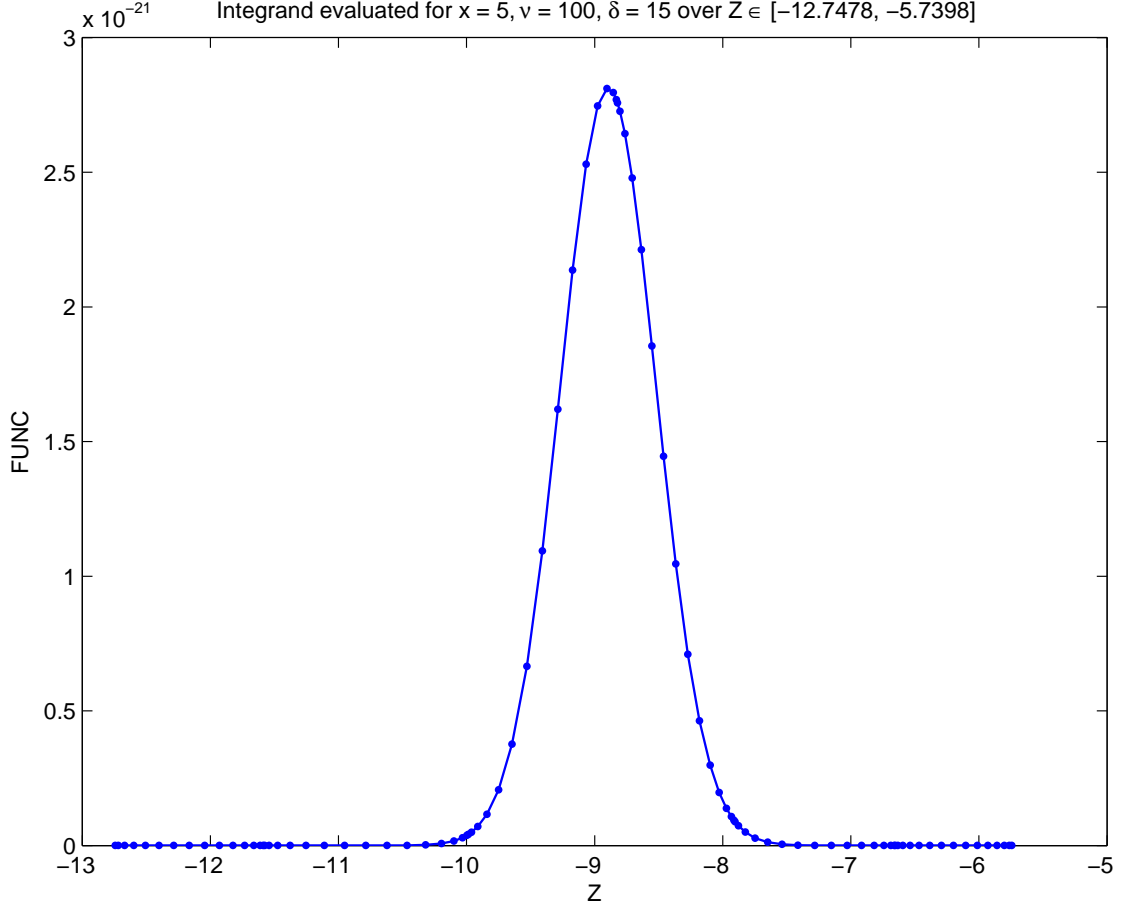


Figure 1: The integrand function  $g(z)$  evaluated at the Gauss-Kronrod nodes ( $90 = 6$  subintervals  $\times$   $15$  nodes) and the selected integration limits  $[A, B]$  used for computing the CDF of the noncentral  $t$ -distribution with the input parameters  $x = 5$ ,  $\nu = 100$  and  $\delta = 15$ . The value of CDF computed by the algorithm *nctcdfvw* is  $\text{CDF} = 2.640405806735035 \times 10^{-21}$ . For comparison, the standard MATLAB function *nctcdf* (Statistics Toolbox) returns  $\text{CDF} = 4.542511227039881 \times 10^{-43}$ , the R function *pt* returns  $\text{CDF} = 2.35515251660662 \times 10^{-21}$ , the SAS function *probt* returns  $\text{CDF} = 2.6404058074408 \times 10^{-21}$ , and the BOOST function *non\_central\_t* returns  $\text{CDF} = 2.64040580673507 \times 10^{-21}$ .

we get

$$\begin{aligned} h(z) &\approx \log\left(\frac{1}{2}\mathcal{E}_\nu(q)\phi(z)\right) \\ &\approx -\log(2) - \frac{1}{2}\left(q - \nu - (\nu - 2)\log\left(\frac{q}{\nu}\right) + \log(\nu) + \log(2\pi) + z^2\right), \end{aligned} \quad (22)$$

which holds true for all  $q$  (and consequently for all  $z$ ) and  $\nu > 2$ . For practical purposes, the algorithm sets  $\nu - 2 \equiv 1$  if  $\nu \leq 2$ .

By solving the equation  $\frac{\partial h(z)}{\partial z} = 0$  we get the estimate of the mode (modus), say  $z_{mod}$ , of the integrand function  $g(z)$  as

$$z_{mod} = \frac{-\delta(x^2 + 2\nu) + x\sqrt{4\nu(\nu - 2) + x^2(\delta^2 + 4(\nu - 2))}}{2(x^2 + \nu)}. \quad (23)$$

Based on that, we can estimate the maximum value of the integrand function by

$$g_{max} \approx g(z_{mod}) \approx \exp(h(z_{mod})), \quad (24)$$

and also the required absolute tolerance bound  $\varepsilon_A$ , by solving

$$\log(\varepsilon_A) \approx h(z_{mod}) + \log(\varepsilon_R). \quad (25)$$

Consequently, the limits  $A_2$  and  $B_2$  defined by (21) can be estimated by solving the approximate equation

$$h(A_2) = h(B_2) = \log(\varepsilon_A). \quad (26)$$

The algorithm *nctcdfvw* finds the limits  $A_2$  and  $B_2$  by solving quadratic equation which results from the quadratic approximation (expansion) of the function  $h(\cdot)$  around  $z_{mod}$ .

The quantiles  $q_{\varepsilon_R}$  used in (16) are estimated by using efficient approximation proposed in [8], eqn. (A.3),

$$q_{\varepsilon_R} \approx \nu + 2\varepsilon_R + 1.62 \sqrt{\nu\varepsilon_R} + 0.63012 \sqrt{\nu} \log(\varepsilon_R) - 1.12032 \sqrt{\nu} - 2.48 \sqrt{\varepsilon_R} - 0.65381 \log(\varepsilon_R) - 0.22872. \quad (27)$$

Given the integration limits  $[A, B]$ , the algorithm evaluates the CDF by using the approximation

$$F_{t,\nu,\delta}(x) \approx \Phi(A) + \int_A^B g(z) dz \quad (28)$$

The integral  $\int_A^B g(z) dz$  can be evaluated by using the standard (adaptive) Gauss-Kronrod quadrature which allows to estimate the integration error.

In order to speed-up the computation by using evaluation of the vectorized functions, the MATLAB version of the algorithm *nctcdfvw* uses the non-adaptive version of the (G7,K15)-Gauss-Kronrod quadrature over fixed (prespecified) number of sub-intervals of  $[A, B]$ . The default number of sub-intervals is set to  $n_{subs} = 16$  (a rather conservative choice based on a detailed and extensive testing, in order to ensure that the relative precision to be better than (or equal to)  $10^{-14}$ , in most cases), but for most typical values of the input parameters division to 6 sub-intervals (which requires  $90 = 6 \times 15$  evaluations of the integrand function  $g(z)$ ) is sufficient to achieve the relative error less than  $10^{-12}$ .

## 5. Accuracy comparisons

In order to illustrate and compare the accuracy of the standard algorithms/implementations for computing the noncentral  $t$  distribution (and to compare it with the suggested algorithm based on the expression (4)), Table 1 presents the CDF values of the noncentral  $t$  distribution for several (rather extreme) combinations of input parameters  $x$ ,  $\nu$ , and  $\delta$ , computed by different algorithms/implementations. In particular,

- **MATLAB** function *nctcdfvw* (based on non-adaptive Gauss-Kronrod quadrature, here with integration limits  $[A, B]$  divided into 32 subintervals),
- **MATLAB** function *nctcdf* (Statistics Toolbox),
- **R** function *pt*,
- **SAS** function *probt*,
- **BOOST** C++ Libraries function *non\_central\_t*, as implemented in *DistExplorer*,
- **MATHEMATICA** function *CDF[NoncentralStudentTDistribution]*.

The 'true' values of the CDF have been computed by a version of the MATLAB algorithm *nctcdfvw* (modified for quadruple-precision computation by using the Multiprecision Computing Toolbox for MATLAB [7]).

All computations have been realized on standard PC under 32-bit Windows XP operating system. For detailed comparisons, the results are presented with 18 significant digits. Notice however, that the double-precision arithmetic (used by the presented algorithms, except Mathematica) returns only 16 significant digits.

The differences (with respect to the exact values) are emphasized by underlining the affected digits. The 'NA' value is displayed if the algorithm did not converge. A symbol '\*' is displayed for cases when Mathematica warning message '*NIntegrate* failed to converge to prescribed accuracy' has been issued.

8

| Example | $x$  | $\nu$ | $\delta$ | TRUE CDF                 | NCTCDFVW                 | MATLAB                    | R                         |
|---------|------|-------|----------|--------------------------|--------------------------|---------------------------|---------------------------|
| 1       | 1    | 1     | 0        | 7.5000000000000000E-001  | 7.5000000000000000E-001  | 7.5000000000000000E-001   | 7.5000000000000000E-001   |
| 2       | -35  | 1     | 0        | 9.09209467564843408E-003 | 9.09209467564843700E-003 | 9.09209467564843500E-003  | 9.09209467564843300E-003  |
| 3       | -35  | 1     | 1        | 1.89903487263458750E-003 | 1.89903487263458800E-003 | 1.89903487263454200E-003  | 1.89903487274889500E-003  |
| 4       | -5   | 1     | 5        | 8.52042451613777143E-009 | 8.52042451613777000E-009 | 8.52042303378652800E-009  | 8.52061943223958500E-009  |
| 5       | -15  | 1     | 15       | 1.29043391190105994E-053 | 1.29043391190106200E-053 | 0                         | 8.16013923099490100E-014  |
| 6       | -35  | 1     | 35       | 7.31501102529248499E-272 | 7.31501102529247900E-272 | 0                         | 5.72875080706580800E-014  |
| 7       | 1    | 10    | 5        | 4.34725285650591657E-005 | 4.34725285650591700E-005 | 4.34725285650592700E-005  | 4.34725284300420300E-005  |
| 8       | 1    | 10    | 10       | 7.95914542988750673E-019 | 7.95914542988750700E-019 | 7.61985302416059300E-024  | 7.94724820780974200E-019  |
| 9       | 1    | 10    | 15       | 1.41346486009205976E-042 | 1.41346486009205900E-042 | 3.67096619931285900E-051  | 9.87042896134168300E-043  |
| 10      | 1    | 10    | 35       | 1.69061467860900429E-237 | 1.69061467860900800E-237 | 1.12491070647255300E-268  | 4.51922957217374300E-250  |
| 11      | 150  | 10    | 200      | 5.88999020094520836E-002 | 5.88999020094520500E-002 | 5.88999020072176300E-002  | 5.45995201321134400E-002  |
| 12      | 150  | 10    | 500      | 3.25241635439258347E-019 | 3.25241635439257900E-019 | 3.25241635469807400E-019  | 2.762470663506697e00-026  |
| 13      | 50   | 100   | 75       | 4.99615060338271916E-011 | 4.99615060338271200E-011 | 4.99615060337484300E-011  | 0                         |
| 14      | 500  | 100   | 510      | 3.71160937464178059E-001 | 3.71160937464177900E-001 | 3.71160937385911300E-001  | 3.75215597825695000E-001  |
| 15      | 1    | 1000  | 10       | 1.14935521338266224E-019 | 1.14935521338266200E-019 | 7.61985302416059300E-024  | 1.14932401536150700E-019  |
| 16      | 100  | 1000  | 105      | 2.05403544901854621E-002 | 2.05403544901854000E-002 | 2.05403544900927000E-002  | 2.01116859042299000E-002  |
| 17      | 1000 | 1000  | 1010     | 3.22438286661716843E-001 | 3.22438286661716400E-001 | 3.22438286530340600E-001  | 3.23499129176421900E-001  |
| Example | $x$  | $\nu$ | $\delta$ | TRUE CDF                 | SAS                      | BOOST                     | MATHEMATICA               |
| 1       | 1    | 1     | 0        | 7.5000000000000000E-001  | 7.5000000000000000E-001  | 7.5000000000000000E-001   | 7.5000000000000000E-001   |
| 2       | -35  | 1     | 0        | 9.09209467564843408E-003 | 9.09209467564840000E-003 | 9.09209467564843000E-003  | 9.09209467564843408E-003  |
| 3       | -35  | 1     | 1        | 1.89903487263458750E-003 | NA                       | 1.89903487263459000E-003  | 1.89903487263458750E-003  |
| 4       | -5   | 1     | 5        | 8.52042451613777143E-009 | 8.52042451613770000E-009 | 8.52042442641561000E-009  | 8.52042451613777143E-009  |
| 5       | -15  | 1     | 15       | 1.29043391190105994E-053 | 1.29043391190100000E-053 | 2.86650837419345000E-016  | 1.29043391190105994E-053  |
| 6       | -35  | 1     | 35       | 7.31501102529248499E-272 | NA                       | -3.77352003162056000E-017 | 7.31501102529248499E-272  |
| 7       | 1    | 10    | 5        | 4.34725285650591657E-005 | 4.34725285650590000E-005 | 4.34725285650591000E-005  | 4.34725285650591657E-005  |
| 8       | 1    | 10    | 10       | 7.95914542988750673E-019 | 7.95914542988750000E-019 | 7.95914542988752000E-019  | 7.95914542988750673E-019  |
| 9       | 1    | 10    | 15       | 1.41346486009205976E-042 | 1.41346486009200000E-042 | 1.41346486009207000E-042  | 1.41346486009205976E-042  |
| 10      | 1    | 10    | 35       | 1.69061467860900429E-237 | 1.69061467860890000E-237 | 1.12491070647253000E-268  | 1.69061467860900429E-237  |
| 11      | 150  | 10    | 200      | 5.88999020094520836E-002 | 5.88999020094520000E-002 | 5.88999020094668000E-002  | NA                        |
| 12      | 150  | 10    | 500      | 3.25241635439258347E-019 | NA                       | 3.25241635439664000E-019  | NA                        |
| 13      | 50   | 100   | 75       | 4.99615060338271916E-011 | 4.99615060338260000E-011 | 4.99615060338272000E-011  | *4.99615060338271916E-011 |
| 14      | 500  | 100   | 510      | 3.71160937464178059E-001 | NA                       | 3.71160937464169000E-001  | NA                        |
| 15      | 1    | 1000  | 10       | 1.14935521338266224E-019 | 1.14935521338260000E-019 | 1.14935521338269000E-019  | *1.14935521338266224E-019 |
| 16      | 100  | 1000  | 105      | 2.05403544901854621E-002 | 2.05403544901850000E-002 | 2.05403544901843000E-002  | *2.05403544901854621E-002 |
| 17      | 1000 | 1000  | 1010     | 3.22438286661716843E-001 | NA                       | 3.22438286661695000E-001  | NA                        |

Table 1: CDF values of the noncentral  $t$ -distribution computed by different algorithms/implementations for selected combinations of the input parameters.



## 6. Conclusions

It seems that (currently) there is no implementation of the algorithm for computing CDF of the noncentral  $t$ -distribution which is uniformly efficient (reasonably fast) and accurate for all input parameters  $x$ ,  $\nu$ , and  $\delta$  in double-precision arithmetic.

According to our present study, this goal is best satisfied by the SAS and BOOST implementations, if we restrict to the typical (most frequently used) values of the input parameters. However, when the output of such algorithm is supposed to be used subsequently for further computations, as e.g. computing the PDF or quantiles of the distribution (and/or the noncentrality parameter  $\delta$ , or the degrees of freedom  $\nu$ , for given values  $x$  and the CDF/PDF), the possible inaccuracy, slow evaluation or failure to converge, can be critical.

Here we have suggested a new algorithm based on numerical quadrature of a well behaved function which is reasonably precise and fast in double-precision arithmetic for all input parameters  $x$ ,  $\nu$  and  $\delta$ . Precision of the MATLAB version of the algorithm was tested for wide range of input parameters (not presented here). In most of the tested cases the relative error was below  $10^{-14}$ .

## 7. Acknowledgements

The author wishes to thank two anonymous referees for identifying the errors in original version of the manuscript and for their constructive comments which helped to improve presentation of the paper. The work was supported by the Slovak Research and Development Agency, grant APVV-0096-10, and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, grants VEGA 2/0038/12, 2/0019/10.

## References

- [1] Abramowitz M., Stegun I. A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. *National Bureau of Standards*, Tenth Edition, 1972.
- [2] Airey, J. R., Irwin, J. O., Fisher, R. A.: Introduction to Tables of  $H_h$  Functions. *British Association for the Advancement of Science, Mathematical Tables I*, XXIV–XXXV, 1931.
- [3] Benton D., Krishnamoorthy K.: *Computing discrete mixtures of continuous distributions: noncentral chisquare, noncentral t and the distribution of the square of the sample multiple correlation coefficient*. *Computational Statistics & Data Analysis* **43** (2003), 249–267.
- [4] Bristow, P. A., Maddock, J.: DistExplorer: Statistical Distribution Explorer. *Boost Software License*. Edition: Version 1.0., 2012. <http://sourceforge.net/projects/distexplorer/>.
- [5] Guenther, W. C.: *Evaluation of probabilities for the noncentral distributions and the difference of two t variables with a desk calculator*. *Journal of Statistical Computation and Simulation* **6** (1978), 199–206.
- [6] Hahn, G. J., Meeker, G. J.: *Statistical Intervals: A Guide for Practitioners*. *John Wiley & Sons*, New York, 1991.
- [7] Holoborodko P.: Multiprecision Computing Toolbox for MATLAB. *Advpix*, Yokohama. Edition: Version 3.4.3, 2013. <http://www.advpix.com>.
- [8] Inglot, T.: *Inequalities for quantiles of the chi-square distribution*. *Probability and Mathematical Statistics* **30** (2010), 339–351.
- [9] Inglot, T., Ledwina, T.: *Asymptotic optimality of a new adaptive test in regression model*. *Annales de l'Institut Henri Poincaré* **42** (2006), 579–590.
- [10] Janiga, I., Garaj, I.: *OnE-sided tolerance factors of normal distributions with unknown mean and variability*. *Measurement Science Review* **8** (2006), 12–16.
- [11] Johnson, N. L., Kotz, S., Balakrishnan, N.: *Continuous Univariate Distributions, Volume 2*. *John Wiley & Sons*, New York, Second Edition, 1995.
- [12] Kim, J.: *Efficient Confidence Interval Methodologies for the Noncentrality Parameters of the Noncentral T-Distributions*. *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology*, PhD Thesis, May 2007.
- [13] Krishnamoorthy, K., Mathew, T.: *Statistical Tolerance Regions: Theory, Applications, and Computation*, *John Wiley & Sons*, New York, 2009.
- [14] Lenth, R. V.: *Algorithm AS 243 Cumulative distribution function of the non-central t distribution*. *Applied Statistics* **38** (1989), 185–189.
- [15] Maddock, J., Bristow, P. A., Holin, H., Zhang, X., Lalande, B., Rade, J., Sewani, G., van den Berg, T., Sobotta, B.: *Noncentral T Distribution. Boost C++ Libraries*, Edition: Version 1.53.0, 2012. <http://www.boost.org>.
- [16] The MathWorks Inc.: *MATLAB Edition: Version 8.0.0.783 (R2012b)*. Natick, Massachusetts, 2012. <http://www.mathworks.com>.
- [17] R Development Core Team.: *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Edition: Version 3.0.0, Vienna, Austria, 2013. <http://www.R-project.org>.
- [18] SAS Institute Inc.: *PROBT Function. SAS(R) 9.3 Functions and CALL Routines: Reference*, 2013. <http://support.sas.com/>.
- [19] Student: *The probable error of a mean*. *Biometrika* **6** (1908), 1–25.
- [20] Wolfram Research, Inc.: *Mathematica Edition: Version 9.0*. *Wolfram Research, Inc.*, Champaign, Illinois, 2013. <http://www.wolfram.com/mathematica/>.