Computational and Applied Mathematics, I
C. Brezinski and U. Kulish (Editors)
Elsevier Science Publishers B.V. (North-Holland)

Asymptotic inversion of a class of cumulative distribution functions

N. M. Temme

CWI, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
e-mail: nicot@cwi.nl

Abstract

The cumulative distribution functions are written in terms of the Gaussian distribution, and for large values of a parameter, represented as the normal distribution function plus remainder. First the normal distribution function is inverted, and next an asymptotic expansion is constructed for the solution of the inversion problem. For the symmetric incomplete beta function the transformation to the Gaussian standard form is given, together with an indication of the numerical accuracy of the inversion method.

1. INTRODUCTION

We consider cumulative distribution functions of the Gaussian form

$$F_a(\eta) = \sqrt{\frac{a}{2\pi}}\,\frac{1}{\Phi(a)}\int_{-\infty}^{\eta} e^{-\frac{1}{2}a\zeta^2} f(\zeta)\,d\zeta, \tag{1.1}$$

where $a > 0$ and $\eta \in \mathbb{R}$. We assume that f is an analytic function in a domain containing the real axis, and that f is positive on $\mathbb{R}$ with the normalization $f(0) = 1$. The function $\Phi(a)$ is a normalization, such that $F_a(\infty) = 1$. That is,

$$\Phi(a) = \sqrt{\frac{a}{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}a\zeta^2} f(\zeta)\,d\zeta. \tag{1.2}$$

In [1] it is shown that several well-known distribution functions can be written in this form, including the incomplete gamma and beta functions. It is also shown that the following representation holds

$$F_a(\eta) = \tfrac{1}{2}\mathrm{erfc}\left(-\eta\sqrt{a/2}\right) + R_a(\eta), \tag{1.3}$$

where erfc is the error function (the normal distribution function) defined by

$$\mathrm{erfc}\, z = \frac{2}{\sqrt{\pi}}\int_z^{\infty} e^{-t^2}\,dt$$

and $R_a(\eta)$ is written as

$$R_a(\eta) = \frac{e^{-\frac{1}{2}a\eta^2}}{\sqrt{2\pi a}} S_a(\eta). \tag{1.4}$$

The function $S_a(\eta)$ is expanded in the form

$$S_a(\eta) \sim \sum_{n=0}^{\infty} \frac{C_n(\eta)}{a^n}, \quad \text{as} \quad a \to \infty, \tag{1.5}$$

$\eta \in \mathbb{R}$. For the class considered in [1] no restrictions on η are needed. In fact, (1.5) holds uniformly with respect to $\eta \in \mathbb{R}$ (and in a larger domain of the complex plane).

The normalizing function $\Phi(a)$ can be expanded in the form

$$\Phi(a) \sim \sum_{n=0}^{\infty} \frac{A_n}{a^n}, \quad \text{as} \quad a \to \infty, \quad A_0 = 1. \tag{1.6}$$

This expansion can be obtained by substituting the expansion

$$f(\zeta) = \sum_{n=0}^{\infty} f_n \zeta^n$$

(we assume that f is analytic at the origin). It follows that

$$A_n = \frac{2^n \Gamma(n+\frac{1}{2})}{\Gamma(\frac{1}{2})} f_{2n}, \quad n = 0, 1, 2, . \tag{1.7}$$

We are concerned with solving the equation

$$F_a(\eta) = q, \tag{1.8}$$

that is, finding η for a given value positive number a and a given number $q \in (0, 1)$. We are especially interested in solving (1.8) when a is large.

This problem is of importance in probability theory and mathematical statistics. Several approaches are available in the (statistical) literature, where often a first approximation of η is constructed, based on asymptotic expansions, but this first approximation is not always reliable. Higher approximations may be obtained by numerical inversion techniques, which require evaluation of the function $F_a(\eta)$. This may be rather time consuming, especially when a is large.

In the present method we also use an asymptotic result. The approximation is quite accurate, especially when a is large. It follows from numerical results for the incomplete gamma function, however, that a three term asymptotic expansion already gives an accuracy of 4 significant digits for $a \geq 2$, uniformly with respect to $q \in (0, 1)$.

2. METHOD OF INVERSION

Let $\eta_0 = \eta_0(q, a)$ be the real number satisfying the equation

$$\tfrac{1}{2}\mathrm{erfc}\left(-\eta_0\sqrt{a/2}\right) = q. \tag{2.1}$$

Then the requested value η is written in the form

$$\eta(q, a) = \eta_0(q, a) + \varepsilon(\eta_0, a), \tag{2.2}$$

and we try to determine the function ε. It appears that we can expand this quantity in the form

$$\varepsilon(\eta_0, a) \sim \frac{\varepsilon_1}{a} + \frac{\varepsilon_2}{a^2} + \frac{\varepsilon_3}{a^3} + \ldots, \tag{2.3}$$

as $a \to \infty$. The coefficients ε_i can obtained explicitly in terms of the function f (and derivatives) evaluated at η_0.

We first remark that (1.1) and (1.8) yield the relation

$$\frac{dq}{d\eta} = \sqrt{\frac{a}{2\pi}}\,\frac{1}{\Phi(a)}\,e^{-\frac{1}{2}a\eta^2} f(\eta)$$

From (2.1) we obtain

$$\frac{dq}{d\eta_0} = \sqrt{\frac{a}{2\pi}}\,e^{-\frac{1}{2}a\eta_0^2}.$$

Upon dividing these two differential equations, we eliminate q, although it is still present in η_0. So we obtain

$$\frac{d\eta}{d\eta_0} = \frac{\Phi(a)}{f(\eta)} e^{\frac{1}{2}a(\eta^2-\eta_0^2)}, \quad -\infty < \eta_0 < \infty. \tag{2.4}$$

Substitution of (2.2) gives the differential equation

$$f(\eta_0 + \varepsilon)\Phi(a)\left[1 + \frac{d\varepsilon}{d\eta_0}\right] = \Phi(a)\,e^{a\varepsilon(\eta_0 + \frac{1}{2}\varepsilon)},$$

a relation between ε and η_0, with a as (large) parameter.

It is convenient to write η in place of η_0. That is, we try to find the function $\varepsilon = \varepsilon(\eta, a)$ that satisfies the equation

$$f(\eta + \varepsilon)\left[1 + \frac{d\varepsilon}{d\eta}\right] = \Phi(a)\,e^{a\varepsilon(\eta + \frac{1}{2}\varepsilon)}. \tag{2.5}$$

When we have obtained the solution $\varepsilon(\eta, a)$ (or an approximation), we write it as $\varepsilon(\eta_0, a)$ and the final value of η follows from (2.2).

For large values of a we have $\Phi(a) = 1 + \mathcal{O}(a^{-1})$, see (1.6). Comparing dominant terms in (2.5), we infer that the first coefficient ε_1 in (2.3) is defined by

$$f(\eta) = e^{\eta\varepsilon_1},$$

giving

$$\varepsilon_1 = \frac{1}{\eta}\ln f(\eta). \tag{2.6}$$

We have assumed = that f is positive on $\mathbb{R}$, $f(0) = 1$, and that f is analytic in a neighbourhood of $\mathbb{R}$. Hence, it follows that $\varepsilon_1 = \varepsilon_1(\eta)$ is also analytic on $\mathbb{R}$. Further coefficients in (2.3) are obtained by using standard perturbation methods. We need the expansion of $\Phi(a)$ given in (1.6), and

$$f(\eta + \varepsilon) = f(\eta) + \varepsilon f'(\eta) + \tfrac{1}{2}\varepsilon^2 f''(\eta) + \ldots,$$

in which (2.3) is substituted to obtain an expansion in powers of a^{-1}. Putting all this in (2.5), we find by comparing terms with equal powers of a^{-1}

$$\varepsilon_2 = \frac{1}{2\eta f}(2f\varepsilon_1' + 2f'\varepsilon_1 - 2A_1 f - f\varepsilon_1^2); \tag{2.7}$$

$$\begin{aligned}\varepsilon_3 = &\frac{1}{8\eta f}(8f\varepsilon_2' + 8f'\varepsilon_1\varepsilon_1' - 4A_1 f\varepsilon_1^2 + 8f'\varepsilon_2 + 4f''\varepsilon_1^2 + \\ &- 8A_1 f\varepsilon_2\eta - 8A_2 f - 8f\varepsilon_1\varepsilon_2 - 4f\varepsilon_2^2\eta^2 - 4f\varepsilon_2\eta\varepsilon_1^2 - f\varepsilon_1^4).\end{aligned} \tag{2.8}$$

The derivatives f', ε', etc., are with respect to η, and all functions are evaluated at η.

The above representations of ε_i are not suitable for computing when $|\eta|$ is small. We give the first part of the power series expansion of these functions. All coefficients can be expressed in terms of the coefficients f_n occurring in the expansion $f(\eta) = \sum_{n=0}^{\infty} f_n\eta^n$. We also need the relations $A_1 = f_2$, $A_2 = 3f_4$ (see (1.7)).

$$\begin{aligned}\varepsilon_1 = &f_1 + (f_2 - \tfrac{1}{2}f_1^2)\eta + (f_3 - f_1 f_2 + \tfrac{1}{3}f_1^3)\eta^2 + (f_4 - f_1 f_3 - \tfrac{1}{2}f_2^2 + f_2 f_1^2 - \tfrac{1}{4}f_1^4)\eta^3 \\ &+ (-f_3 f_2 + f_1 f_2^2 - f_2 f_1^3 + f_3 f_1^2 - f_1 f_4 + f_5 + \tfrac{1}{5}f_1^5)\eta^4 \\ &+ (-\tfrac{1}{2}f_3^2 + 2f_3 f_1 f_2 - f_3 f_1^3 - f_4 f_2 + f_4 f_1^2 + f_6 - f_1 f_5 + \\ &\tfrac{1}{3}f_2^3 - \tfrac{3}{2}f_2^2 f_1^2 + f_2 f_1^4 - \tfrac{1}{6}f_1^6)\eta^5 \\ &+ (2f_4 f_1 f_2 + f_7 - 3f_3 f_2 f_1^2 + \tfrac{1}{7}f_1^7 + 2f_2^2 f_1^3 - f_2 f_1^5 - f_1 f_2^3 + f_3 f_1^4 + f_1 f_3^2 + f_3 f_2^2 \\ &- f_4 f_1^3 - f_5 f_2 + f_5 f_1^2 - f_4 f_3 - f_1 f_6)\eta^6 + O(\eta^7);\end{aligned}$$

$$\begin{aligned}\varepsilon_2 &= (2f_3 - \tfrac{1}{3}f_1^3) + (-\tfrac{3}{2}f_2f_1^2 + \tfrac{5}{8}f_1^4 + 3f_4)\eta \\ &+ (-\tfrac{13}{15}f_1^5 + \tfrac{10}{3}f_2f_1^3 - 2f_1f_2^2 - 2f_3f_1^2 + 4f_5)\eta^2 \\ &+ (\tfrac{25}{6}f_3f_1^3 + \tfrac{25}{4}f_2^2f_1^2 - \tfrac{65}{12}f_2f_1^4 - \tfrac{5}{6}f_2^3 - \tfrac{5}{2}f_4f_1^2 - 5f_3f_1f_2 + \tfrac{77}{72}f_1^6 + 5f_6)\eta^3 \\ &+ (-6f_4f_1f_2 + 6f_7 + 15f_3f_2f_1^2 - \tfrac{87}{70}f_1^7 - 13f_2^2f_1^3 + \tfrac{77}{10}f_2f_1^5 + 5f_1f_2^3 \\ &- \tfrac{13}{2}f_3f_1^4 - 3f_1f_3^2 - 3f_3f_2^2 + 5f_4f_1^3 - 3f_5f_1^2)\eta^4 + O(\eta^5);\end{aligned}$$

$$\begin{aligned}\varepsilon_3 &= (\tfrac{4}{15}f_1^5 - \tfrac{1}{3}f_2f_1^3 - f_3f_1^2 + 8f_5) \\ &+ (-f_2^2f_1^2 + \tfrac{67}{24}f_2f_1^4 - \tfrac{3}{2}f_4f_1^2 + \tfrac{1}{3}f_3f_1^3 + 2f_3^2 - 4f_3f_1f_2 - \tfrac{127}{144}f_1^6 + 15f_6)\eta \\ &+ (-6f_4f_1f_2 + 24f_7 + 2f_3f_2f_1^2 + \tfrac{391}{210}f_1^7 + 9f_2^2f_1^3 \\ &- \tfrac{87}{10}f_2f_1^5 - f_1f_2^3 + \tfrac{7}{2}f_3f_1^4 - 6f_1f_3^2 - 3f_3f_2^2 - 2f_5f_1^2 + 6f_4f_3)\eta^2 + O(\eta^3);\end{aligned}$$

3. SIMPLE EXAMPLES

We consider a few simple examples in which η can be written explicitly in terms of η_0. Apart from the inversion in (2.1) an exact solution of the inversion problem can then be given.

First let $f(\zeta) = \exp(-b\zeta)$. Then (1.1) can be written as

$$\begin{aligned}F_a(\eta) &= \sqrt{\frac{a}{2\pi}}\,\frac{1}{\Phi(a)}\int_{-\infty}^{\eta} e^{-\frac{1}{2}a\zeta^2} f(\zeta)\,d\zeta, \\ &= \sqrt{\frac{a}{2\pi}}\,\frac{e^{\frac{1}{2}b^2/a}}{\Phi(a)}\int_{-\infty}^{\eta+b/a} e^{-\frac{1}{2}a\zeta^2}\,d\zeta \\ &= \frac{1}{2}\mathrm{erfc}\left[-(\eta+b/a)\sqrt{a/2}\right],\end{aligned}$$

where we have taken $\Phi(a) = \exp[b^2/(2a)]$, to obtain the normalization $F_a(\infty) = 1$. It follows that the solution η of (1.8) can be expressed in terms of the solution of (2.1) by writing $\eta = \eta_0 - b/a$, and (2.2) gives $\varepsilon(\eta_0, a) = -b/a$. The quantities ε_i occurring in (2.3) are given by

$$\varepsilon_1 = -b, \quad \varepsilon_i = 0, \quad i \geq 2.$$

It easily verified that (2.6) and (2.7) indeed reduce to these simple values.

Next consider $f(\zeta) = \exp(-\frac{1}{2}b\eta^2)$, giving

$$\begin{aligned}F_a(\eta) &= \sqrt{\frac{a}{2\pi}}\,\frac{1}{\Phi(a)}\int_{-\infty}^{\eta} e^{-\frac{1}{2}a\zeta^2} f(\zeta)\,d\zeta, \\ &= \frac{1}{2}\mathrm{erfc}\left[-\eta\sqrt{(a+b)/2}\right],\end{aligned}$$

where we have taken $\Phi(a) = \sqrt{a/(a+b)}$. The solution of (1.8) now reads $\eta = \eta_0\sqrt{a/(a+b)}$ and we have

$$\varepsilon(\eta_0, a) = \eta_0\left[\sqrt{a/(a+b)} - 1\right].$$

Expanding we obtain

$$\varepsilon(\eta_0, a) \sim \eta_0\left(-\frac{b}{2a} + \frac{3b^2}{8a^2} - \frac{5b^3}{16a^3} + \ldots\right), \quad a \to \infty$$

giving

$$\varepsilon_1 = -\tfrac{1}{2}b\eta_0, \quad \varepsilon_2 = \tfrac{3}{8}b^2\eta_0, \quad \varepsilon_3 = -\tfrac{5}{16}b^3\eta_0.$$

To verify (2.8), (2.9) we need also the first coefficients A_n of (1.6), that is, $A_1 = -b/2,\ A_2 = 3b^2/8$.

4. TRANSFORMATION TO STANDARD FORM

We give a simple example of a transformation to standard form (1.1). Consider the incomplete beta function

$$I_x(a,b) = \frac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}\,dt, \quad a > 0, \quad b > 0,$$

where $B(a,b)$ is the complete beta function, given by $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. In the symmetric case $a = b$, which gives Student's distribution, we can write

$$I_x(a,a) = \frac{4^{-a}}{B(a,a)}\int_0^x [4t(1-t)]^a\,\frac{dt}{t(1-t)}.$$

We bring this into a standard form with Gaussian character by writing

$$-\frac{1}{2}\zeta^2 = \ln[4t(1-t)], \quad 0 < t < 1, \quad \operatorname{sign}(\zeta) = \operatorname{sign}(t - \frac{1}{2}),$$

and a similar relation between η and x. After straightforward manipulations we obtain

$$I_x(a,a) = \sqrt{\frac{a}{2\pi}}\,\frac{1}{\Phi(a)}\int_{-\infty}^{\eta} e^{-\frac{1}{2}a\zeta^2} f(\zeta)\,d\zeta,$$

with

$$\Phi(a) = \frac{\sqrt{a}\,\Gamma(a)}{\Gamma(a+\frac{1}{2})}, \quad f(\zeta) = \sqrt{\frac{\frac{1}{2}\zeta^2}{1 - \exp(-\frac{1}{2}\zeta^2)}}.$$

Observe that f is singular at the points where $\zeta^2 = \pm 4 \pm i$, and that, hence, f is analytic in the strip $|\Im\zeta| < \sqrt{2\pi}$. The first coefficients A_n of expansion (1.6) are $A_0 = 1,\ A_1 = \frac{1}{8},\ A_2 = \frac{1}{128}$.

We give a numerical verification of the asymptotic method. When $a = 2$, equation $I_x(a, a) = q$ is equivalent with $2x^3 - 3x^2 + q = 0$, which for $q = 5/32$ has a solution $x = 1/4$. The method described in this paper gives with three terms in expansion (2.3) the approximation 0.25000202. It follows that in this case the value $a = 2$ gives 5 digits accuracy. For the general incomplete beta functions comparable results can be obtained. In [2], [3] we give more details, with similar numerical results, on the inversion of the incomplete gamma and beta functions. It appears that the expansion (2.3) leads to a powerful and efficient expansion for the solution of the inversion problem, in which the parameter a is assumed to be large, although already for $a \geq 2$ quite accurate results can be obtained.

5. REFERENCES

[1] N.M. TEMME (1982), The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions, SIAM J. Math. Anal. **13**, 239-253.

[2] N.M. TEMME (1991), Asymptotic inversion of incomplete gamma functions, to appear in Mathematics of Computation.

[3] N.M. TEMME (1991), Asymptotic inversion of the incomplete beta function, to appear in J. Comput. and Appl. Math.