PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Special Issue *FAATNA20>22: Functional Analysis, Approximation Theory and Numerical Analysis* Volume 16 · 2023 · Pages 10–16

Computation of the regularized incomplete Beta function

Vera Egorova[a] · Amparo Gil[a] · Javier Segura[b] · Nico M. Temme[c]

Abstract

An algorithm for the computation of the regularized incomplete Beta function is described. This function has important applications in areas such as Statistics, Physics and Information Theory. The computation of the function can be carried out through a continued function evaluation supplemented with series and asymptotic expansions when both parameters are large. Numerical tests demonstrate the accuracy of the algorithm and show that our algorithm is more accurate than Matlab's built-in function `betainc` for a wide range of parameters.

Keywords: Regularized incomplete Beta function, Asymptotic Expansions, Continued Fractions, Numerical computations.
2010 AMS classification: 33B15, 33C15, 65D20.

1 Introduction

The regularized incomplete Beta function and its complementary function are defined by

$$\begin{aligned} I_x(p,q) &= \frac{1}{B(p,q)}\int_0^x t^{p-1}(1-t)^{q-1}\,dt, \\ J_x(p,q) &= \frac{1}{B(p,q)}\int_x^1 t^{p-1}(1-t)^{q-1}\,dt. \end{aligned} \tag{1}$$

We assume that p and q are positive and $0 \le x \le 1$. $B(p,q)$ is the Beta function

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \tag{2}$$

We notice that from (1) it is easy to check that $J_x(p,q) = I_{1-x}(q,p)$. Also, $I_x(p,q)$ and $J_x(p,q)$ satisfy the relation

$$I_x(p,q) + J_x(p,q) = 1. \tag{3}$$

In the algorithm, it is convenient to compute the smallest of the two. The transition point (indicating when to compute $I_x(p,q)$ or $J_x(p,q)$) is given by

$$x_t = \frac{p}{p+q}. \tag{4}$$

When $x < x_t$ ($x > x_t$), we have (roughly) $I_x(p,q) < J_x(p,q)$ ($J_x(p,q) > I_x(p,q)$).

The inversion problem of the regularized incomplete Beta function was treated in [2]. In this paper, we combine different methods to build an algorithm to compute the regularized incomplete Beta function. A MatLab implementation of the resulting algorithm is provided. Numerical tests demonstrate the accuracy of the algorithm and show that our algorithm is more accurate than MatLab's built-in function `betainc` for a wide range of parameters.

Earlier information on the methods can be found in [7] and [1, §10.5.2]. Related methods are also used in our paper [4] for the incomplete gamma function ratios. For other algorithms applied to the computation of the regularized incomplete Beta function and the incomplete gamma function ratios, see the references given in [6, §8.28].

[a]Departamento de Matemática Aplicada y CC. de la Computación. Universidad de Cantabria. 39005-Santander, Spain.
[b]Departamento de Matemáticas, Estadística y Computación. Universidad de Cantabria. 39005-Santander, Spain.
[c]IAA, 1825 BD 25, Alkmaar, The Netherlands. Former address: Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam, The Netherlands.

2 Methods of computation

2.1 Continued fraction

A continued fraction representation which is useful for computation is given by

$$I_x(p,q) = \frac{x^p(1-x)^q}{pB(p,q)}\left(\frac{1}{1+}\,\frac{d_1}{1+}\,\frac{d_2}{1+}\,\frac{d_3}{1+}\,\ldots\right), \tag{5}$$

where

$$\begin{aligned} d_{2m} &= \frac{m(q-m)x}{(p+2m-1)(p+2m)}, \\ d_{2m+1} &= -\frac{(p+m)(p+q+m)x}{(p+2m)(p+2m+1)}. \end{aligned} \tag{6}$$

The continued fraction converges rapidly for $x < x_t$, where x_t is the transition point given in (4). When $0 \le x \le x_t$ we have $I_x(p,q) \lesssim J_x(p,q)$ we will use the continued fraction. However, for large p and q the computation slows down when x is close to x_t, and we need asymptotic methods. For $x_t \le x \le 1$ (when $1-x \le q/(p+q)$) more rapid convergence is obtained by computing the complementary function $J_x(p,q)$ instead, which is smaller than $I_x(p,q)$.

The computation of the front factor in (5) becomes problematic when p and q are large. Therefore we use the expression given in [8, §11.3.4] to compute $\dfrac{x^p(1-x)^q}{B(p,q)}$

$$\frac{x^p(1-x)^q}{B(p,q)} = \sqrt{\frac{pq}{2\pi(p+q)}}\,\frac{\Gamma^*(p+q)}{\Gamma^*(p)\Gamma^*(q)}e^{p(\log(1+\sigma)-\sigma)+q(\log(1+\tau)-\tau)}, \tag{7}$$

where $\sigma = \dfrac{x-x_t}{x_t}$, $\tau = \dfrac{x_t-x}{1-x_t}$ and $\Gamma^*(x)$ is the scaled gamma function which is defined as

$$\Gamma^*(x) = \frac{\Gamma(x)}{\sqrt{2\pi/x}\,x^x e^{-x}}, \quad x > 0. \tag{8}$$

For an algorithm to compute $\Gamma^*(x)$ see [3]. For small values of σ we use a Taylor expansion to evaluate $\log(1+\sigma)-\sigma$ in (7); similarly to compute $\log(1+\tau)-\tau$ for small values of τ.

2.2 Small or moderate values of p and q, expansions in terms of hypergeometric functions

The regularized incomplete Beta function satisfies the following relations in terms of the Gauss hypergeometric functions (see [6, §8.17(ii)])

$$\begin{aligned} I_x(p,q) &= \frac{x^p}{pB(p,q)}\,{}_2F_1\left(\begin{matrix}p,1-q\\ p+1\end{matrix};x\right) \\ &= \frac{x^p(1-x)^{q-1}}{pB(p,q)}\,{}_2F_1\left(\begin{matrix}1,1-q\\ p+1\end{matrix};\frac{x}{x-1}\right) \\ &= \frac{x^p(1-x)^q}{pB(p,q)}\,{}_2F_1\left(\begin{matrix}p+q,1\\ p+1\end{matrix};x\right). \end{aligned} \tag{9}$$

Using (15.2.1) of [5], the following power series expansions for $I_x(p,q)$ are obtained:

$$I_x(p,q) = \frac{x^p}{pB(p,q)}\sum_{n=0}^{\infty}\frac{(1-q)_n}{p+n}\frac{x^n}{n!}, \quad |x|<1, \tag{10}$$

$$I_x(p,q) = \frac{x^p(1-x)^{q-1}}{pB(p,q)}\sum_{n=0}^{\infty}\frac{(1-q)_n}{(1+p)_n}\left(\frac{x}{x-1}\right)^n, \quad \left|\frac{x}{x-1}\right|<1, \tag{11}$$

$$I_x(p,q) = \frac{x^p(1-x)^q}{pB(p,q)}\sum_{n=0}^{\infty}\frac{(p+q)_n}{(1+p)_n}x^n, \quad |x|<1. \tag{12}$$

Analogous series expansions can be obtained for $J_x(p,q)$.

Denoting the coefficients of the series (10), (11) and (12) by $b_n/(p+n)$, c_n and d_n respectively, we observe that

$$\frac{b_{n+1}}{b_n} = \frac{n+1-q}{n+1}x, \quad \frac{c_{n+1}}{c_n} = \frac{n+1-q}{n+1+p}\,\frac{x}{(x-1)}, \quad \frac{d_{n+1}}{d_n} = \frac{p+q+n}{p+1+n}x. \tag{13}$$

The series (10) and (11) terminate when q is a positive integer.

2.3 Asymptotic expansions

The expansion in (11) can be viewed as an asymptotic expansion for large values of p, with q fixed. The series is convergent for $0 \le x < \frac{1}{2}$, but we can use it for large values of p when $\frac{1}{2} \le x \le 1-\delta$, where δ is a small positive number.

2.3.1 Large p and q, error function approximation

We consider an expansion that is valid around the transition point x_t. The starting point is the representation

$$I_x(p,q) = \tfrac{1}{2}\operatorname{erfc}\left(\eta\sqrt{r/2}\right) - R_r(\eta), \tag{14}$$

where we write $p = r\sin^2\theta$, $q = r\cos^2\theta$ with $0 < \theta < \pi/2$ and η is given by

$$-\tfrac{1}{2}\eta^2 = \sin^2\theta \log\frac{x}{\sin^2\theta} + \cos^2\theta\log\frac{1-x}{\cos^2\theta}. \tag{15}$$

After taking the square root for η we take $\operatorname{sign}(\eta) = \operatorname{sign}(x - \sin^2\theta)$; this means $\operatorname{sign}(\eta) = \operatorname{sign}(x - p/(p+q))$.

The function $R_r(\eta)$ in Eq.(14) can be written in the form

$$R_r(\eta) = \frac{1}{F(p,q)}\frac{e^{-\frac{1}{2}r\eta^2}}{\sqrt{2\pi r}}S_r(\eta), \quad F(p,q) = \frac{\Gamma^*(p)\Gamma^*(q)}{\Gamma^*(p+q)}, \tag{16}$$

see also (7). The function $S_r(\eta)$ can be expanded in the form of an asymptotic power series for large values of r, but to avoid the calculation of a number of coefficients in the series we use the power series

$$S_r(\eta) = \sum_{k=0}^{\infty}\tilde{d}_k\eta^k, \quad |\eta| < \eta_c, \quad \eta_c = 2\sqrt{\pi}\min(\sin\theta, \cos\theta). \tag{17}$$

The value η_c follows from the singularities of the relation between η and x. We have from (15)

$$-\eta\frac{d\eta}{dx} = \frac{\sin^2\theta - x}{x(1-x)}. \tag{18}$$

The point $x = \sin^2\theta$ (corresponding to $\eta = 0$) is a regular point, but the derivative $d\eta/dx$ also vanishes for $\sin^2\theta\, e^{\pm 2\pi i}$ and $\cos^2\theta\, e^{\pm 2\pi i}$, which are relevant because of the multivalued logarithms in (15). These singular points determine the domain of convergence of the series in (17).

Using (1), (14) and (15) we derive the differential equation

$$\eta S_r(\eta) - \frac{1}{r}\frac{d}{d\eta}S_r(\eta) = f(\eta) - F(p,q), \tag{19}$$

where

$$f(\eta) = \frac{\eta\sin\theta\cos\theta}{x - \sin^2\theta} = \sum_{k=0}^{\infty}a_k\eta^k, \tag{20}$$

of which the first few a_k are given by

$$a_0 = 1,\ a_1 = -\frac{2}{3}\cot 2\theta,\ a_2 = \frac{\sin^4\theta + \cos^4\theta + 1}{6\sin^2 2\theta}. \tag{21}$$

Substituting the expansion of $f(\eta)$ into (19), we obtain the recurrence relation

$$\tilde{d}_k = a_{k+1} + \frac{1}{r}(k+2)\tilde{d}_{k+2}, \quad k = 0, 1, 2, \ldots, \tag{22}$$

which we use in the backward direction with zero starting values. A special relation is

$$\tilde{d}_1 = r\left(F(p,q) - 1\right) \quad \Longrightarrow \quad F(p,q) = 1 + \frac{\tilde{d}_1}{r}, \tag{23}$$

which can be used in (16) to replace $F(p,q)$. This method can be used when $|\eta|$ is small, say, $|\eta| \le \frac{1}{2}\eta_c$. In [4] a similar method was used for computing the incomplete gamma function ratios.

To compute first $\min\{I_x(p,q), J_x(p,q)\}$: if $\eta < 0$ then

$$I_x(p,q) = \tfrac{1}{2}\operatorname{erfc}\left(-\eta\sqrt{r/2}\right) - R_r(\eta), \quad J_x(p,q) = 1 - I_x(p,q), \tag{24}$$

and if $\eta \ge 0$ then

$$J_x(p,q) = \tfrac{1}{2}\operatorname{erfc}\left(\eta\sqrt{r/2}\right) + R_r(\eta), \quad I_x(p,q) = 1 - J_x(p,q). \tag{25}$$

3 Numerical testing and algorithm

For testing the accuracy of the different methods described in Section 2, we use three-term recurrence relations satisfied by the regularized incomplete Beta function. In particular, we use (8.17.13), (8.17.14) and (8.17.16) of [6] written as

$$
\begin{aligned}
\epsilon_1 &= \left|1-\frac{pI_x(p+1,q)+qI_x(p,q+1)}{(p+q)I_x(p,q)}\right|,\\
\epsilon_2 &= \left|1-\frac{xqI_x(p-1,q+1)+pI_x(p+1,q)}{(p+qx)I_x(p,q)}\right|,\\
\epsilon_3 &= \left|1-\frac{pI_x(p+1,q)+\rho xI_x(p-1,q)}{(p+\rho x)I_x(p,q)}\right|,
\end{aligned}
\tag{26}
$$

where $\rho = p+q-1$.

The tests are applied when $I_x(p,q)$ is greater than the underflow limit in double precision. To compare the accuracy obtained with the different methods, we have first generated a mesh of the values p and q, with 500 nodes for each parameter in the interval (0, 1000) for different (fixed) values of x. For each point, we compare the accuracy obtained when computing (26) with the different methods: ERF, given in (14); S1, given in (10); S2, given in (11); S3, given in (12) and CF, given in (5). Figure 1 shows some of the results obtained for $x = 0.01, 0.25, 0.75, 0.99$. The method for which the error is smaller than the others is shown in the plot with a distinctive colour.

More extensive tests comparing the accuracy of the different methods lead to the algorithm described as Algorithm 1. The MatLab function implementing the algorithm is called `betaincreg` [1]

Algorithm 1 Computation of the regularized incomplete Beta function $I_x(p,q)$

Require: $0 \le x \le 1$, $p \ge 0$, $q \ge 0$
Ensure: $I = I_x(p,q)$
  Compute the transition point, $x_t = p/(p+q)$.
  if $p > 50$ & $q > 50$ & $p+q > 700$ & $|x-x_t| < 0.2$ **then**
    Use the error function approximation in Section 2.3.1.
  else if $p > 100$ & $q < 10$ **then**
    if $x < 0.85$ **then**
      Use the series expansion (10).
    else
      Use the continued fraction (5).
    end if
  else
    if $q > (1-x)p/x$ **then**
      Use the continued fraction (5).
    else
      Use the series expansion (11).
    end if
  end if

The accuracy of our algorithm has been compared against the MatLab built-in function `betainc`. In Figure 2 we show a comparison for fixed values of x ($x = 0.01, 0.25, 0.75, 0.99$) in the (p,q)-plane. The implementation for which the test error is the smallest, is plotted with a distinctive colour: blue (our algorithm) or yellow (MatLab). As can be seen, our algorithm is more accurate than MatLab's built-in function `betainc` at most points of the (p, q)-plane. In addition, in a more extensive test considering a large number of points (10^8) randomly generated in the region $(x, p, q) \in (0, 1) \times (0, 10000) \times (0, 10000)$, the maximum error obtained when computing (26) using our algorithm was 2.8×10^{-12}, which was significantly smaller than the value obtained with the MatLab function (2×10^{-10}).

Acknowledgements

We thank the referees for their helpful remarks. AG, JS and NMT thank financial support from project PGC2018-098279-B-I00 funded by MCIN/ AEI /10.13039/501100011033/ FEDER "Una manera de hacer Europa" and PID2021-127252NB-I00 funded by MCIN/AEI/10.13039/501100011033/ FEDER, UE. VE thanks financial support from project PID2019-107685RB-I00. NMT thanks CWI Amsterdam for scientific support.

[1]The function is available at http://personales.unican.es/gila/betaincreg.m

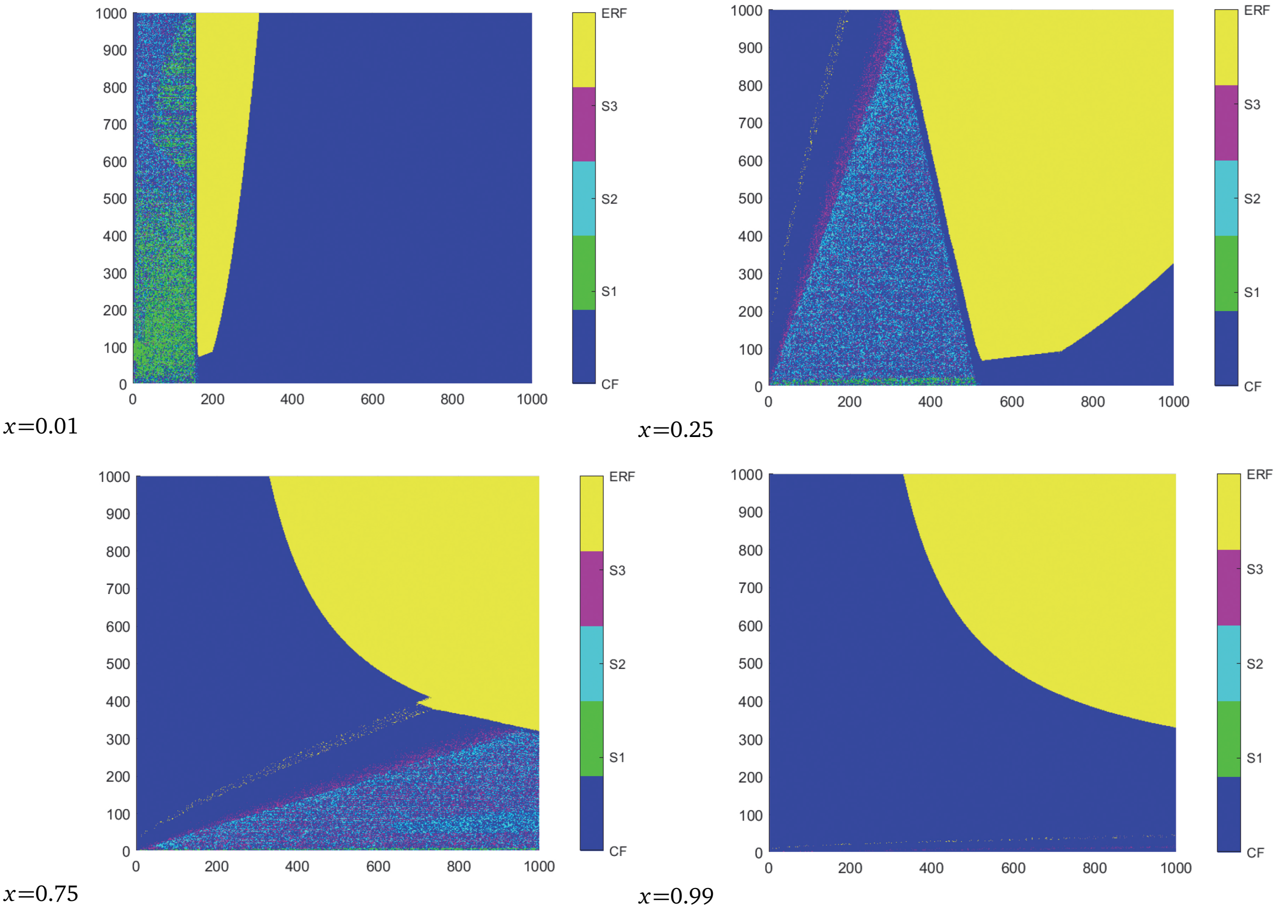

Figure 1: Comparison of methods in the (p,q)-plane to build the algorithm for computing the regularized incomplete Beta function. The p-values (q-values) are represented in the horizontal (vertical) axis. ERF is given in (14), S1 is given in (10), S2 is given in (11), S3 is given in (12) and CF is given in (5). The method where the test error (26) is miminum, is plotted. Four different values of x have been considered in the figures.

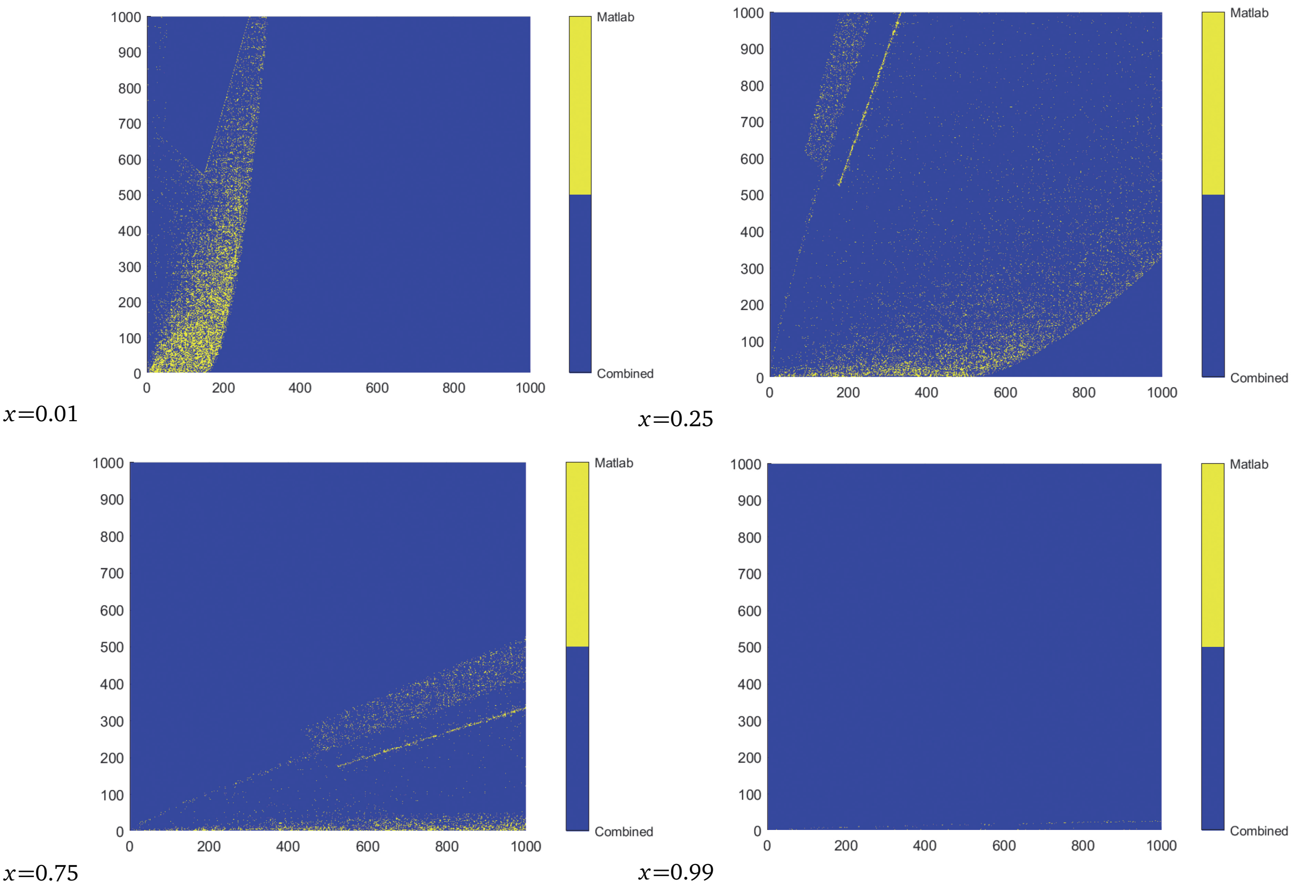

Figure 2: Comparison of accuracy between our algorithm and the MatLab built-in function **betainc** in the (p,q)-plane. The p-values (q-values) are represented in the horizontal (vertical) axis. Four different values of x have been considered in the figures. The implementation for which the test error is the smallest, is plotted with a distinctive colour: blue (our algorithm) or yellow (MatLab).

References

[1] A. Gil, J. Segura, and N. M. Temme. *Numerical Methods for Special Functions*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007.

[2] A. Gil, J. Segura, and N. M. Temme. Efficient algorithms for the inversion of the cumulative central beta distribution. *Numer Algo*, 74(1):77–91, 2017.

[3] A. Gil, J. Segura, and N. M. Temme. Gammachi: A package for the inversion and computation of the gamma and chi-square cumulative distribution functions (central and noncentral). new version announcement. *Comput Phys Commun*, 267:108083, 2021.

[4] A. Gil, N. M. Temme, and J. Segura. Efficient and accurate algorithms for the computation and inversion of the incomplete gamma function ratios. *SIAM J. Sci. Comput.*, 34(6):A2965–A2981, 2012.

[5] A.B. Olde-Daalhuis. Chapter 15, Hypergeometric Function. In *NIST Handbook of Mathematical Functions*, pages 383–402. Cambridge University Press, Cambridge, 2010a. http://dlmf.nist.gov/15.

[6] R. B. Paris. Chapter 8, Incomplete Gamma and Related Functions. In *NIST Handbook of Mathematical Functions*, pages 321–349. Cambridge University Press, Cambridge, 2010a. http://dlmf.nist.gov/8.

[7] N. M. Temme. Asymptotic inversion of the incomplete beta function. *J. Comput. Appl. Math.*, 41(1-2):145–157, 1992.

[8] N. M. Temme. *Special functions: An introduction to the classical functions of mathematical physics*. Wiley-Interscience, New York, 1996.