Numerical evaluation of the Gauss hypergeometric function with the hypergeo package

Robin K. S. Hankin
Auckland University of Technology

Abstract

This paper introduces the **hypergeo** package of R routines, for numerical calculation of hypergeometric functions. The package is focussed on efficient and accurate evaluation of the hypergeometric function over the whole of the complex plane within the constraints of fixed-precision arithmetic. The hypergeometric series is convergent only within the unit circle, so analytic continuation must be used to define the function outside the unit circle. This short document outlines the numerical and conceptual methods used in the package; and justifies the package philosophy, which is to maintain transparent and verifiable links between the software and AMS-55. The package is demonstrated in the context of game theory. To cite the package in publications, please use (Hankin 2015).

Keywords: Hypergeometric functions, numerical evaluation, complex plane, R, residue theorem.

1. Introduction

The *geometric* series $\sum_{k=0}^{\infty} t_k$ with $t_k = z^k$ may be characterized by its first term and the constant ratio of successive terms $t_{k+1}/t_k = z$, giving the familiar identity $\sum_{k=0}^{\infty} z^k = (1-z)^{-1}$. Observe that while the series has unit radius of convergence, the right hand side is defined over the whole complex plane except for $z = 1$ where it has a pole. Series of this type may be generalized to a *hypergeometric* series in which the ratio of successive terms is a rational function of k:

$$\frac{t_{k+1}}{t_k} = \frac{P(k)}{Q(k)}$$

where $P(k)$ and $Q(k)$ are polynomials. If both numerator and denominator have been completely factored we would write

$$\frac{t_{k+1}}{t_k} = \frac{(k+a_1)(k+a_2)\cdots(k+a_p)}{(k+b_1)(k+b_2)\cdots(k+b_q)(k+1)} z$$

(the final term in the denominator is due to historical reasons), and if we require $t_0 = 1$ then we write

$$\sum_{k=0}^{\infty} t_k z^k = {}_{\mathrm{a}}\mathrm{F}_{\mathrm{b}}\left[\begin{matrix} a_1, a_2, \ldots, a_p \\ b_1, b_2, \ldots, b_q \end{matrix}; z\right] \qquad (1)$$

when defined. An absent factor is indicated with a dash; thus ${}_0\mathrm{F}_0\left[\begin{matrix} - \\ - \end{matrix};z\right] = e^z$. In most cases of interest one finds that $p = 2$, $q = 1$ suffices. Writing a, b, c for the two upper and one lower argument respectively, the resulting function ${}_2\mathrm{F}_1\left(a, b; c; z\right)$ is known as *the* hypergeometric function. Many functions of elementary analysis are of this form; examples would include logarithmic and trigonometric functions, Bessel functions, etc. For example, ${}_2\mathrm{F}_1\left(\frac{1}{2}, 1; \frac{3}{2}; -z^2\right) = z^{-1}\arctan z$.

Michel and Stoitsov (2008) state that physical applications are "plethora". In addition, naturally-occuring combinatorial series frequently have a sum expressible in terms of hypergeometric functions and an example from the author's work in the field game theory is given below.

1.1. Equivalent forms

The hypergeometric function's series representation, namely

$$ {}_2\mathrm{F}_1\left(a, b; c; z\right) = \sum_{k=0}^{\infty} \frac{(a)_k\,(b)_k}{(c)_k\,k!} z^k, \qquad (a)_k = \Gamma(a+k)/\Gamma(a) \tag{15.1.1} $$

has unit radius of convergence by the ratio test but the integral form

$$ {}_2\mathrm{F}_1\left(a, b; c; z\right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_{t=0}^{1} t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt, \tag{15.3.1} $$

due to Gauss, furnishes analytic continuation; it is usual to follow Riemann and define a cut along the positive real axis from 1 to ∞ and specify continuity from below [NB: equations with three-part numbers, as 15.1.1 and 15.3.1 above, are named for their reference in Abramowitz and Stegun (1965)]. This is implemented as `f15.3.1()` in the package and exhibits surprisingly accurate evaluation.

Gauss also provided a continued fraction form for the hypergeometric function [implemented as `hypergeo_contfrac()` in the package] which has superior convergence rates for parts of the complex plane at the expense of more complicated convergence properties (Cuyt *et al.* 2008).

2. The hypergeo package

The **hypergeo** package provides some functionality for the hypergeometric function; the emphasis is on fast vectorized R-centric code, complex z and moderate real values for the auxiliary parameters a, b, c. The package is released under GPL-2.

Observing the slow convergence of the series representation 15.1.1, the complex behaviour of the continued fraction representation, and the heavy computational expense of the integral representation 15.3.1, it is clear that non-trivial numerical techniques are required for a production package.

The package implements a generalization of the method of Forrey (1997) to the complex case. It utilizes the observation that the ratio of successive terms approaches z, and thus the strategy adopted is to seek a transformation which reduces the modulus of z to a minimum. Abramowitz and Stegun give the following transformations:

$$
\begin{aligned}
{}_2\mathrm{F}_1(a,b;c;z) &= (1-z)^{-a}\,{}_2\mathrm{F}_1\left(a,c-b;c;\frac{z}{z-1}\right) && (15.3.4)\\
&= (1-z)^{-b}\,{}_2\mathrm{F}_1\left(a,c-a;c;\frac{z}{z-1}\right) && (15.3.5)\\
&= \frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}\,{}_2\mathrm{F}_1(a,b;a+b-c+1;1-z)\\
&\quad + (1-z)^{c-a-b}\frac{\Gamma(c)\,\Gamma(a+b-c)}{\Gamma(a)\,\Gamma(b)}\,{}_2\mathrm{F}_1(c-a,c-b;c-a-b+1;1-z) && (15.3.6)\\
&= \frac{\Gamma(c)\,\Gamma(b-a)}{\Gamma(b)\,\Gamma(c-a)}(-z)^{-a}\,{}_2\mathrm{F}_1\left(a,1-c+a;1-b+a;\frac{1}{z}\right)\\
&\quad + \frac{\Gamma(c)\,\Gamma(a-b)}{\Gamma(a)\,\Gamma(c-b)}(-z)^{-b}\,{}_2\mathrm{F}_1\left(b,1-c+b;1-a+b;\frac{1}{z}\right) && (15.3.7)\\
&= (1-z)^{-a}\frac{\Gamma(c)\,\Gamma(b-a)}{\Gamma(b)\,\Gamma(c-a)}\,{}_2\mathrm{F}_1\left(a,c-b;a-b+1;\frac{1}{1-z}\right)\\
&\quad + (1-z)^{-b}\frac{\Gamma(c)\,\Gamma(a-b)}{\Gamma(a)\,\Gamma(c-b)}\,{}_2\mathrm{F}_1\left(b,c-a;b-a+1;\frac{1}{1-z}\right) && (15.3.8)\\
&= \frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}z^{-a}\,{}_2\mathrm{F}_1\left(a,a-c+1;a+b-c+1;1-\frac{1}{z}\right)\\
&\quad + \frac{\Gamma(c)\,\Gamma(a+b-c)}{\Gamma(a)\,\Gamma(b)}(1-z)^{c-a-b}z^{a-c}\,{}_2\mathrm{F}_1\left(c-a,1-a;c-a-b+1;1-\frac{1}{z}\right). && (15.3.9)
\end{aligned}
$$

Observing that the set $\left\{z, \frac{z}{z-1}, 1-z, \frac{1}{z}, \frac{1}{1-z}, 1-\frac{1}{z}\right\}$ forms a group under functional composition[1] we may apply each of the transformations to the primary argument z and choose the one of smallest absolute value to evaluate.

Given the appropriate transformation, the right hand side is evaluated using direct summation. If $|z| < 1$, the series is convergent by the ratio test, but may require a large number of terms to achieve acceptable numerical precision. Summation is dispatched to `genhypergeo_series()` which evaluates the generalized hypergeometric function 1; the R implementation uses multiplication by repeatedly incremented upper and lower indices a_i, b_i. Thus for example if $(1-z)^{-1}$ is small in absolute value we would use function `f13.3.8()`:

```
> require("hypergeo")
> f15.3.8
```

```
function(A,B,C,z,tol=0,maxiter=2000){
 jj <- i15.3.8(A,B,C)
```

[1]It is the anharmonic subgroup of the Möbius transformations, generated by $z \longrightarrow 1/z$ and $z \longrightarrow 1-z$. It is isomorphic to S_3, the symmetric group on 3 elements.

```
  jj[1]*(1-z)^(-A)*genhypergeo(U=c(A,C-B),L=A-B+1,z=1/(1-z),tol=tol,maxiter=maxiter) +
  jj[2]*(1-z)^(-B)*genhypergeo(U=c(B,C-A),L=B-A+1,z=1/(1-z),tol=tol,maxiter=maxiter)
}
```

(slightly edited in the interests of visual clarity). This is a typical internal function of the package and like all similar functions is named for its equation number in Abramowitz and Stegun (1965). Note the helper function `i15.3.9()`, which calculates the Gamma coefficients of the two hypergeometric terms in the identity. This structure allows transparent checking of the code.

2.1. Special cases

The methods detailed above are not applicable for all values of the parameters a, b, c. If, for example, $c = a + b \pm m$, $m \in \mathbb{N}$ (a not uncommon case), then equation 15.3.6 is not useful because each term has a pole; and it is numerically difficult to approach the limit. In this case the package dispatches to `hypergeo_cover1()` which uses 15.3.4 through 15.3.9 but with 15.3.6 replaced with suitable limiting forms such as

$$ {}_2\mathrm{F}_1\left(a, b; a+b+m; z\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(n!)^2} \left[2\psi(n+1) - \psi(a+n) - \psi(b+n) - \log(1-z)\right](1-z)^n, \qquad \pi \tag{15.3.11} $$

(Abramowitz and Stegun give a similar expression for negative m). This equation is comparable to 15.3.6 in terms of computational complexity but requires evaluation of the digamma function ψ. Equation 15.3.11 is evaluated in the package using an algorithm similar to that for `genhypergeo_series()` but includes a runtime option which specifies whether to evaluate $\psi(\cdot)$ *ab initio* each time it is needed, or to use the recurrence relation $\psi(z+1) = \psi(z) + 1/z$ at each iteration after the first. These two options appear to be comparable in terms of both numerical accuracy and speed of execution, but further work would be needed to specify which is preferable in this context.

A similar methodology is used for the case $b = a \pm m$, $m = 0, 1, 2, \ldots$ in which case the package dispatches to `hypergeo_cover2()`.

However, the case $c - a = 0, 1, 2, \ldots$ is not covered by Abramowitz and Stegun (1965) and the package dispatches to `hypergeo_cover3()` which uses formulae taken from the Wolfram functions site (Wolfram 2014). For example `w07.23.06.0026.01()` gives a straightforwardly implementable numerical expression for ${}_2\mathrm{F}_1$ as a sum of two *finite* series and a generalized hypergeometric function ${}_3\mathrm{F}_2$ with primary argument z^{-1}.

In all these cases, the limiting behaviour is problematic. For example, if $a+b-c$ is close to, but not exactly equal to, an integer then equation 15.3.11 is not applicable. The analytic value of the hypergeometric function in these circumstances is typically of moderate modulus, but both terms of equation 15.3.6 have large amplitude and numerics are susceptible to cancellation errors.

2.2. Critical points

All the above methods fail when $z = \frac{1}{2} \pm \frac{i\sqrt{3}}{2}$, because none of the transformations 15.3.6-15.3.9

change the modulus of z from 1. The function is convergent at these points but numerical evaluation is difficult. This issue does not arise in the real case considered by Forrey (1997).

These points were considered by Buhring (1987) who presented a computational method for these values; however, his method is not suitable for finite-precision arithmetic (a brief discussion is presented at `?buhring`) and the package employs either an iterative scheme due to Gosper (Johansson *et al.* 2010), or the residue theorem if z is close to either of these points.

3. Package testing suite

The package comes with an extensive test suite in the `tests/` directory. The tests fall into two main categories, firstly comparison with either Maple or Mathematica output (although Becken and Schmelcher (2000) caution that Mathematica routines cannot be used as reference values); and secondly, verification of identities which appear in AMS-55 as named equations.

4. The package in use

The **hypergeo** package offers direct numerical functionality to the R user on the command line. One example from the author's current work is in game theory (Hankin 2020). Consider a game in which a player is given n counters each of which she must allocate into one of two boxes, A or B. At times $t = 1, 2, 3\ldots$ a box is identified at random and, if it is not empty, a counter removed from it; box A is chosen with probability p and box B with probability $1-p$. The object of the game is to remove all counters as quickly as possible. If the player places a counters in box A and b in B, then the probability mass function of removing the final counter at time $t = a + b + r$ is

$$p^a(1-p)^b\left[\binom{a+b+r-1}{a-1,b+r}(1-p)^r+\binom{a+b+r-1}{a+r,b-1}p^r\right],\qquad r=0,1,2,\ldots. \tag{2}$$

The two terms correspond to the final counter being removed from box A or B respectively. This PMF has expectation

$$\begin{aligned} p^a(1-p)^b\Bigg[p\binom{a+b}{a+1,b-1}\ {}_2\mathrm{F}_1\left(a+b+1,2;a+2;p\right)+\\ (1-p)\binom{a+b}{a-1,b+1}\ {}_2\mathrm{F}_1\left(a+b+1,2;b+2;1-p\right)\Bigg] \end{aligned} \tag{3}$$

with R idiom:

```
> expected <- function(a,b,p){
+   Re(
+   choose(a+b,b) * p^a * (1-p)^b * (
+      p *b/(1+a) * hypergeo(a+b+1,2,a+2,  p) +
+   (1-p)*a/(1+b) * hypergeo(a+b+1,2,b+2,1-p) ))
+ }
```

Thus if $p = 0.8$ and given $n = 10$ counters we might wonder whether it is preferable to allocate them $(8, 2)$ or $(9, 1)$:

```
> c(expected(8,2,0.8),expected(9,1,0.8))

[1] 3.019899 1.921089
```

showing that the latter allocation is preferable in expectation.

The package is designed for use with R and Figure 1 shows the package being used to visualize ${}_2\mathrm{F}_1\left(2, \frac{1}{2}; \frac{2}{3}; z\right)$ over a region of the complex plane.

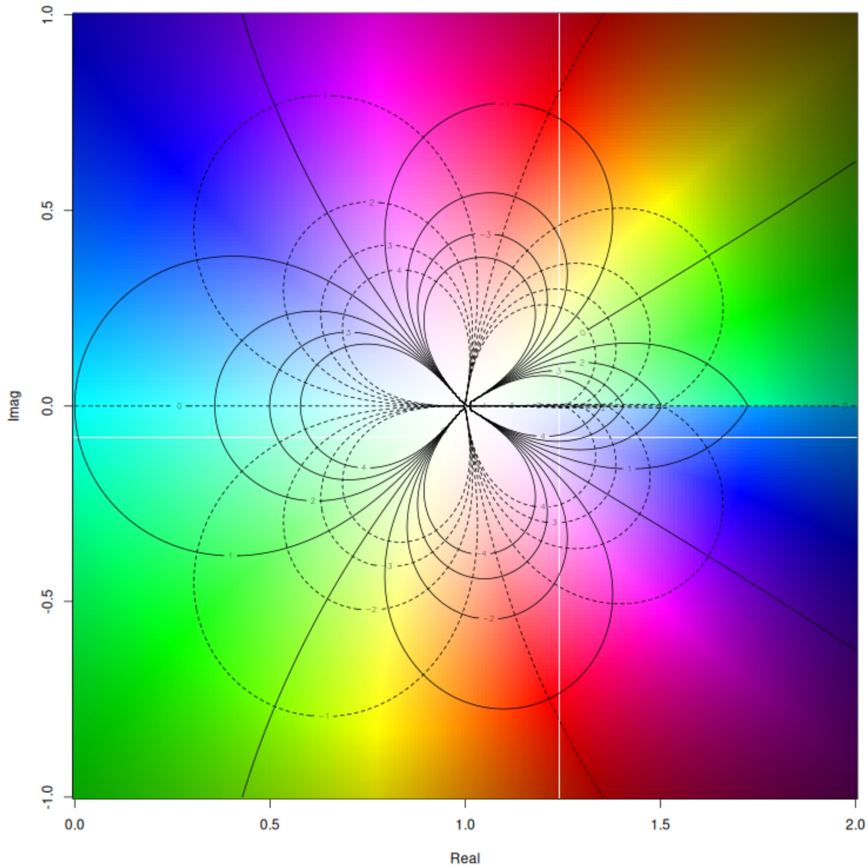

Figure 1: View of the function ${}_2\mathrm{F}_1\left(2, \frac{1}{2}; \frac{2}{3}; z\right)$ evaluated over a part of the complex plane using the **hypergeo** package. Function visualization following Thaller (1998) and the **elliptic** package (Hankin 2006); hue corresponds to argument and saturation to modulus. Solid contour lines correspond to real function values and dotted to imaginary function values. Note the cut line along the real axis starting at $(1, 0)$, made visible by an abrupt change in hue

4.1. Conclusions and further work

Evaluation of the hypergeometric function is hard, as evidenced by the extensive literature concerning its numerical evaluation (Becken and Schmelcher 2000; Michel and Stoitsov 2008; Forrey 1997; Buhring 1987). The **hypergeo** package is presented as a partial implementation, providing reasonably accurate evaluation over a large portion of the complex plane and covering moderate real values of the auxiliary parameters a, b, c. Difficulties arise when $b - a$ or $c - b - a$ become close to, but not exactly, integers because the terms in equations 15.3.6 through 15.3.9 become large and cancellation errors become important.

Further work might include extension to complex auxiliary parameters but Michel and Stoitsov caution that this is not straightforward.

References

Abramowitz M, Stegun IA (1965). *Handbook of Mathematical Functions*. New York: Dover.

Becken W, Schmelcher P (2000). "The analytic continuation of the Gaussian hypergeometric function ${}_2F_1(a, b; c; z)$ for arbitrary parameters." *Journal of Computational and Applied Mathematics*, **126**, 449–478.

Buhring W (1987). "An analytic continuation of the hypergeometric series." *Siam J. Math. Anal.*, **18**(3), 884–889.

Cuyt A, *et al.* (2008). *Handbook of continued fractions for special functions*. Springer.

Forrey RC (1997). "Computing the hypergeometric function." *Journal of computational physics*, **137**, 79–100.

Hankin RKS (2006). "Introducing **elliptic**, an R package for elliptic and modular functions." *Journal of Statistical Software*, **15**.

Hankin RKS (2015). "Numerical evaluation of the Gauss hypergeometric function with the **hypergeo** package." *The R journal*, **7**, 81–88.

Hankin RKS (2020). "A liability allocation game." *J. Comput. Methods Sci. Eng*, **20**(1), 65–79. doi:https://doi.org/10.3233/JCM-193658.

Johansson F, *et al.* (2010). *mpmath: A Python library for arbitrary-precision floating-point arithmetic (version 0.14)*. http://code.google.com/p/mpmath/.

Michel N, Stoitsov MV (2008). "Fast computation of the Gauss hypergeometric function with all its parameters complex with application to the Pöschl-Teller-Ginocchio potential wave functions." *Computer Physics Communictions*, **178**, 535–551.

Thaller B (1998). "Visualization of Complex Functions." *The Mathematica Journal*, **7**(2), 163–180.

Wolfram S (2014). "The Hypergeometric function." File `Hypergeometric2F1.pdf`, downloaded from http:/functions.wolfram.com/HypergeometricFunctions/.

Affiliation:

Robin K. S. Hankin
Auckland University of Technology
New Zealand
E-mail: hankin.robin@gmail.com