Computing the Beta Function for Large Arguments

Martin Mächler
Seminar für Statistik
ETH Zurich
1997, 2002, 2019 ff

Abstract

I was excited about having derived nice asymptotic formulas enabling to accurately compute \( \log B(a, b) \) for very large \( b \) etc, but then realized there were other existing solutions, partly applied already e.g., in TOMS 708 algdiv() (which I now also provide as R function in package DPQ).

1 Introduction

The beta distribution function and its inverse are widely used in statistical software, since, e.g., the critical values of the \( F \) and \( t \) distributions can be expressed using the inverse beta distribution, see, e.g., ?, sec. 5.5.

Whereas sophisticated algorithms are available for computing the beta distribution function and its inverse (Majumder and Bhattacharjee (1973a) and 1973b, Cran et al. (1977); Berry et al. (1990), (Johnson et al., 1995, ch. 25)), these algorithms rely on the computation of the beta function itself which is not a problem in most cases. However, for large arguments \( p \), the usual formula of the beta which uses the gamma function can suffer severely from cancellation when two almost identical numbers are subtracted or divided.

For \( x > 0 \), \( \Gamma(x) \) is positive and analytical, i.e. infinitely many times continuously differentiable. From (2), integrating by parts gives \( \Gamma(x + 1) = x\Gamma(x) \), and hence the recursion formula

\[
\Gamma(x + n) = \Gamma(x) \cdot x \cdot (x + 1) \cdots (x + n - 1). \tag{3}
\]

This entails the best-known property of the gamma function, i.e., the fact that it generalizes the factorial \( n! \). Namely, for \textit{integer} arguments \( n \in \mathbb{N} \), one has \( \Gamma(n + 1) = n! \). For this and many more properties, see, e.g., ?, ch. 6.

For the beta function \( B(p, q) \), it is well known that for larger values of \( p, q \) the corresponding \( \Gamma \) values may become larger than the maximal (floating point) number on the computer, even though \( B(p, q) \) itself may remain relatively small. For this and other
numerical reasons, one usually works with the (natural) logarithms of beta and gamma functions, i.e.,
\[ \log B(p, q) = \log \Gamma(p) + \log \Gamma(q) - \log \Gamma(p + q). \] (4)

For the beta function \( B(p, q) \) which is symmetric in \( p, q \) we assume without loss of generality that \( p < q \), and now consider the situation where \( q \) is very large, or more generally \( q \) is large compared to \( p \),\( p \ll q. \) (5)

For convenience, we write \( B(p, q) = \frac{\Gamma(p)}{Q_{pq}} \) where \( Q_{pq} := \frac{\Gamma(p+q)}{\Gamma(q)} \).

The beta function is closely related to the binomial binomial coefficient \( \binom{N}{n} \),
\[ \binom{N}{n} = \frac{N!}{n! (N-n)!} = \frac{\Gamma(N+1)}{n! \Gamma(N-n+1)} = \frac{Q_{n,N-n+1}}{n!}. \] (7)

where we need \( Q_{pq} \) for integers \( p = n \) and \( q = N - n + 1 \).

Note that for \( p \ll q \), or \( q/p \to \infty \), the ratio in (6) will become more and more imprecise, since \( \log Q_{pq} = \log \Gamma(p+q) - \log \Gamma(q) \) tends to the difference of two almost identical numbers which extinguishes most significant digits. The goal of this paper can be restated as finding numerically useful asymptotic formula for \( Q_{pq} \) when \( q \to \infty \).

For the problem of the binomial coefficient when \( N \to \infty \) and because \( Q_{pq} \) is a smooth (infinitely continuous) function in both arguments, we will consider the special case of \( p = n \in \mathbb{N} \). Using the recursion (3) for the numerator of \( Q_{nq} \), we get
\[ Q_{n,q} = q \cdot (q+1) \cdots (q+n-1) = q^n \cdot \left( 1 + \frac{1}{q} \right) \left( 1 + \frac{2}{q} \right) \cdots \left( 1 + \frac{n-1}{q} \right) \]
\[ = q^n \prod_{k=1}^{n-1} (1 + k/q) = q^n \cdot f_n(1/q), \] (8)
where
\[ f_n(x) = \prod_{k=1}^{n-1} (1 + kx) = \sum_{k=0}^{n-1} a_{kn} x^k, \quad \text{where } a_{0n} \equiv 1, \] (9)
and from (8),
\[ Q_{n,q} = q^n \cdot \left( 1 + \frac{a_{1n}}{q} + \frac{a_{2n}}{q^2} + \ldots + \frac{a_{n-1,n}}{q^{n-1}} \right). \] (10)

In the following section, I will derive closed formulas (in \( n \)) for \( a_{kn} \).

2 Series Expansions

If we apply (9) for \( n+1 \), we get
\[ f_{n+1}(x) = \sum_{k=0}^{n} a_{k,n+1} x^k = \prod_{k=1}^{n} (1 + kx) = (1 + nx) \prod_{k=1}^{n} (1 + kx) = (1 + nx) \cdot f_n(x) \]
\[ = (1 + nx) \sum_{k=0}^{n-1} a_{kn} x^k = 1 + \sum_{k=1}^{n-1} (a_{kn} + na_{k-1,n}) x^k + n a_{n-1,n} x^n. \]
Comparison of coefficients gives

\begin{equation}
    a_{k,n+1} = a_{kn} + na_{k-1,n} \quad (k = 1, \ldots, n),
\end{equation}

where \( a_{n,n} := 0 \).

If we set \( n = k \), and let \( \tilde{a}_n := a_{n,n+1} \), we get \( \tilde{a}_n = a_{n,n} + n\tilde{a}_{n-1} \) from which we conclude that \( \tilde{a}_n = n! \), since \( a_{n,n} = 0 \) and \( \tilde{a}_0 = a_{0,1} = 1 \) by definition (9). Hence,

\begin{equation}
    a_{k,k+1} = k!,
\end{equation}

and applying (11) successively for \( n, n-1, \ldots \) yields

\begin{equation}
    a_{k,n} = \sum_{m=1}^{n-1} m \cdot a_{k-1,m} \quad (k = 1, \ldots, n-1).
\end{equation}

Hence, we can compute \( a_{k,n} \) if \( a_{k-1,n} \) are known and therefore may compute all \( a_{k,n} \) starting with \( k = 0 \) where \( a_{0,n} \equiv 1 \). To derive useful direct formulae, we consider \( a_{kn} \) for given \( k \) as a polynomial in \( n \). It is now useful, to apply (13) for all \( n = 1, 2, \ldots \) instead of only for \( n > k \), i.e., \( k \leq n - 1 \).

References


