Accuracy of asymptotic approximations to the log-Gamma and Riemann-Siegel theta functions

Richard P. Brent

Australian National University
and University of Newcastle

In memory of Jon Borwein 1951–2016

26 September 2016

Abstract

This talk will describe some new bounds on the error in the asymptotic approximation of the log-Gamma function $\ln \Gamma(z)$ for complex z in the right half-plane. These improve on bounds by Hare (1997) and Spira (1971).

I will show how to deduce similar bounds for asymptotic approximation of the Riemann-Siegel theta function $\vartheta(t)$, and show that the attainable accuracy of a well-known approximation to $\vartheta(t)$ can be improved by including an exponentially small term in the approximation.

This improves the attainable accuracy for real positive t from $O(e^{-\pi t})$ to $O(e^{-2\pi t})$.

For further details, see the preprint at arXiv:1609.03682.

The Riemann-Siegel theta function

The *Riemann-Siegel theta function* $\vartheta(t)$ is defined for real t by

$$\vartheta(t) := \arg \Gamma\left(\frac{it}{2} + \frac{1}{4}\right) - \frac{t}{2} \ln \pi.$$

The argument is defined so that $\vartheta(t)$ is continuous on $\mathbb{R}$, and $\vartheta(0) = 0$. Since $\vartheta(t)$ is an odd function, i.e. $\vartheta(-t) = -\vartheta(t)$, we can assume that $t > 0$.

The significance of $\vartheta(t)$

It follows from the functional equation for the ζ function that

$$Z(t) := e^{i\vartheta(t)}\,\zeta(\tfrac{1}{2} + it)$$

is a real-valued function.

In a sense, $\vartheta(t)$ encodes half the information contained in $\zeta(\frac{1}{2} + it)$ (albeit the less interesting half), while $Z(t)$ encodes the other (more interesting) half.

Zeros of $\zeta(s)$ on the critical line $\Re(s) = \frac{1}{2}$ can be isolated by finding sign changes of $Z(t)$.

If $a < b$ and $Z(a)Z(b) < 0$, then there is an odd number of zeros (counted by their multiplicities) in (a, b).

The Riemann-Siegel formula

The Riemann-Siegel formula is

$$Z(t) = \sum_{k=1}^{\lfloor (t/2\pi)^{1/2} \rfloor} 2k^{-1/2} \cos[\vartheta(t) - t \ln k] + R(t),$$

where $t^{1/4} R(t)$ has a rather complicated asymptotic expansion in descending powers of $t^{1/2}$.

This gives a way of computing accurate approximations to $Z(t)$ in time roughly $O(t^{1/2})$.

The "easy" part is the computation of $\vartheta(t)$.

However, $\vartheta(t)$ is an interesting function in its own right. Today I will consider the computation of $\vartheta(t)$, not $R(t)$.

A different representation of $\vartheta(t)$

Recall the definition

$$\vartheta(t) := \arg\Gamma\left(\tfrac{it}{2} + \tfrac{1}{4}\right) - \tfrac{1}{2}t\ln\pi.$$

The following equivalent representation of $\vartheta(t)$ is more convenient for our purposes:

$$\vartheta(t) = \tfrac{1}{2}\arg\Gamma\left(it + \tfrac{1}{2}\right) - \tfrac{1}{2}t\ln(2\pi) - \tfrac{\pi}{8} + \tfrac{1}{2}\arctan\left(e^{-\pi t}\right).$$

The red terms: having $\Re(s) = \frac{1}{2}$ rather than $\Re(s) = \frac{1}{4}$ makes it easier to derive an asymptotic expansion with only odd powers of t and with rigorous error bounds. The $\arctan\left(e^{-\pi t}\right)$ term will be important later, when we consider numerical approximation of $\vartheta(t)$.

Sketch of proof

Use the reflection formula

$$\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$$

and the duplication formula

$$\Gamma(s)\Gamma(s+\tfrac{1}{2}) = 2^{1-2s}\pi^{1/2}\Gamma(2s)$$

with $s = \frac{it}{2} + \frac{1}{4}$. Multiplying gives

$$\Gamma(\tfrac{it}{2}+\tfrac{1}{4})^2\,|\Gamma(\tfrac{it}{2}+\tfrac{3}{4})|^2 = \frac{2^{1/2-it}\pi^{3/2}\Gamma(it+\frac{1}{2})}{\sin\pi\left(\frac{it}{2}+\frac{1}{4}\right)}\,.$$

The result follows on taking the argument of each side and simplifying, using the fact that

$$\arctan\left(\frac{1-e^{-\pi t}}{1+e^{-\pi t}}\right) = \frac{\pi}{4} - \arctan\left(e^{-\pi t}\right).$$

$\vartheta(t)$ and $\ln\Gamma(z)$

To compute $\vartheta(t)$ we need $\arg\Gamma(z)$ where $z = it + \frac{1}{2}$ (or $\frac{it}{2} + \frac{1}{4}$). We use

$$\arg\Gamma(z) = \Im(\ln\Gamma(z))$$

(modulo a multiple of 2π which can be handled with care, so I won't worry about it in this talk).

Taking logarithms, the duplication formula

$$\Gamma(z)\Gamma(z+\tfrac{1}{2}) = 2^{1-2z}\pi^{1/2}\Gamma(2z)$$

gives

$$\ln\Gamma(z+\tfrac{1}{2}) = \ln\Gamma(2z) - \ln\Gamma(z) + \text{ known terms.}$$

Thus, the problem of finding an asymptotic expansion for $\vartheta(t)$ can be solved via Stirling's asymptotic expansion for $\ln\Gamma(z)$ in the case $z = it$ (i.e. on the imaginary axis in the z-plane).

Stirling's formula for $\ln \Gamma(z)$

Recall Stirling's asymptotic expansion (for $z \in \mathbb{C} \backslash (-\infty, 0]$):

$$\ln \Gamma(z) = (z - \tfrac{1}{2}) \log z - z + \tfrac{1}{2} \log(2\pi) + \sum_{j=1}^{k} T_j(z) + R_{k+1}(z),$$

where

$$T_j(z) = \frac{B_{2j}}{2j(2j-1)z^{2j-1}}$$

is the j-th term in the sum, and the "error" or "remainder" after summing k terms may be written as

$$R_{k+1}(z) = -\int_0^\infty \frac{B_{2k}(\{u\})}{2k\,(u+z)^{2k}}\,\mathrm{d}u.$$

Here $\{u\} := u - \lfloor u \rfloor$ denotes the fractional part of u, and $B_{2k}(x)$ is the $2k$-th Bernoulli polynomial.

The remainder term in Stirling's formula

If z is real and positive, life is easy: the asymptotic series is *strictly enveloping* in the sense of Pólya and Szegö, so $R_k(z)$ has the same sign as $T_k(z)$ and is smaller in absolute value, i.e. $|R_k(z)| < |T_k(z)|$. Note that $R_k(z)$ is the remainder after summing $k-1$ terms, so $T_k(z)$ is the first term omitted.

For complex z, life is not so simple.
If $|\arg(z)| \le \pi/4$, the inequality $|R_k(z)| < |T_k(z)|$ still holds [Whittaker and Watson].

If $\Re(z) < 0$, we can (and probably should) use the reflection formula $\Gamma(z)\Gamma(-z) = -\dfrac{\pi}{z\sin(\pi z)}$, so assume that $\Re(z) > 0$.

If $\Im(z) < 0$, we can take complex conjugates, so assume that $\Im(z) \ge 0$.

Thus, we are left with the case $\theta := \arg(z) \in (\pi/4, \pi/2]$.

New error bounds

We have two (related) bounds, valid for $\Re(z) \geq 0$, $z \neq 0$:

$$\left| \frac{R_{k+1}(z)}{T_k(z)} \right| < \sqrt{\pi k}$$

and

$$\left| \frac{R_k(z)}{T_k(z)} \right| < 1 + \sqrt{\pi k}.$$

The first bound is useful if we want to bound the error as a multiple of the last term included in the sum; the second bound applies if we want to bound the error as a multiple of the first term omitted from the sum.

The second bound follows from the first by the triangle inequality, since $R_k(z) = T_k(z) + R_{k+1}(z)$.

Proof (sketch)

Since $|B_{2k}(v)| \le |B_{2k}|$ for all $v \in [0,1]$, we have

$$|R_{k+1}(z)| = \left| \int_0^\infty \frac{B_{2k}(\{u\})}{2k(u+z)^{2k}}\,\mathrm{d}u \right| \le \frac{|B_{2k}|}{2k} \int_0^\infty |u+z|^{-2k}\,\mathrm{d}u.$$

Let $x := \Re(z) \ge 0$ and $y := \Im(z)$. Inside the last integral,

$$|u+z|^2 = (u+x)^2 + y^2 \ge u^2 + x^2 + y^2 = u^2 + |z|^2,$$

so

$$\int_0^\infty |u+z|^{-2k}\,\mathrm{d}u \le \int_0^\infty (u^2+|z|^2)^{-k}\,\mathrm{d}u.$$

Now a change of variables $u \mapsto |z| \tan\psi$ allows us to evaluate the integral on the right in closed form (obtaining a ratio of Gamma functions) via "Wallis's formula". Finally, we use the inequality $\Gamma(k+\frac{1}{2})/\Gamma(k) < \sqrt{k}$ to simplify the result.

Hare's bound

Kevin Hare[1] (1997) gave a bound (in our notation)

$$\left|\frac{R_k(z)}{T_k(z)}\right| \le \frac{4\pi^{1/2}\,\Gamma(k+\frac{1}{2})}{\Gamma(k)\sin^{2k-1}\theta},$$

where $\theta := \arg(z) \in (0,\pi)$. Note that $\sin\theta = y/|z|$.

If $\sin\theta < 1$, our bound is much better than Hare's, because we do not have a $\sin^{2k-1}\theta$ factor in the denominator.

If $\theta = \pi/2$ then $\sin\theta = 1$ (the best case for Hare's bound), and Hare's upper bound on $|R_k/T_k|$ is about $4\sqrt{\pi k}$. This is between 2.74 and 4 times larger than our bound $1+\sqrt{\pi k}$.

[1] Hare was a student of Jon Borwein at Simon Fraser University, 2002.

Improvements on Hare's bound

We made three improvements on Hare's bound.

- By using a form of the remainder with numerator (inside the integral) $B_{2k}(\{u\})$ instead of $B_{2k} - B_{2k}(\{u\})$, we save (almost) a factor of two, since we can use $|B_{2k}(\{u\})| \le |B_{2k}|$, but Hare has to use $|B_{2k} - B_{2k}(\{u\})| \le 2|B_{2k}|$.

  This trick reduces $2\sqrt{\pi k}$ to $1 + \sqrt{\pi k}$.
- By assuming that $x = \Re(z) \ge 0$, we save another factor of two because the integral that we have to bound is over $[0, \infty)$, but Hare's is over $[x, \infty) \subset (-\infty, \infty)$.
- We can use $|u + z|^2 \ge u^2 + |z|^2$ in our proof, whereas Hare has to use $|u + z|^2 = (u + x)^2 + y^2$, since he does not assume that $x \ge 0$. This gives us an improvement by a factor $(|z|/y)^{2k-1} = 1/\sin^{2k-1}\theta$.

Some other bounds

Spira (1971) proved a bound that is similar to Hare's, but with a larger constant factor. He stated his bound without the $\sin^{2k-1}\theta$ factor in the denominator. However, the bound that he actually proved *did* have the $\sin^{2k-1}\theta$ factor in the denominator.

Stieltjes (c. 1900) showed that, for $|\theta| < \pi$,

$$\left|\frac{R_k(z)}{T_k(z)}\right| \le \sec^{2k}(\theta/2).$$

If $\theta = \pi/2$ (the case that is of interest for $\vartheta(t)$), this is larger than our bound by a factor $2^k/(1+\sqrt{\pi k})$.

Gauss's asymptotic expansion for $\ln \Gamma(z + \frac{1}{2})$

Taking logarithms in the duplication formula

$$\Gamma(z + \tfrac{1}{2}) = 2^{1-2z} \pi^{1/2} \Gamma(2z)/\Gamma(z),$$

it is easy to deduce Gauss's (1813) asymptotic expansion

$$\ln \Gamma(z + \tfrac{1}{2}) \sim z \log z - z + \tfrac{1}{2} \log(2\pi) + \sum_{j \geq 1} \widehat{T}_j(z),$$

where

$$\widehat{T}_j(z) = -(1 - 2^{1-2j}) T_j(z) = \frac{B_{2j}(\frac{1}{2})}{2j(2j-1)z^{2j-1}}.$$

The result is well-known, but the point is that we also inherit bounds on the error when the sum in Gauss's formula is truncated.

The Riemann-Siegel theta function (again)

Returning to the Riemann-Siegel theta function $\vartheta(t)$, recall that

$$\vartheta(t) = \tfrac{1}{2}\arg\Gamma\left(it + \tfrac{1}{2}\right) - \tfrac{1}{2}t\ln(2\pi) - \tfrac{\pi}{8} + \tfrac{1}{2}\arctan\left(e^{-\pi t}\right).$$

If we put $z = it$ in Gauss's asymptotic expansion for $\ln\Gamma(z + \frac{1}{2})$, we quickly get an asymptotic expansion for $\vartheta(t)$, along with error bounds. The result is

$$\vartheta(t) = \frac{t}{2}\log\left(\frac{t}{2\pi e}\right) - \frac{\pi}{8} + \frac{\arctan\left(e^{-\pi t}\right)}{2} + \sum_{j=1}^{k} \widetilde{T}_j(t) + \widetilde{R}_{k+1}(t),$$

where $\widetilde{T}_j(t) = \dfrac{|B_{2j}(\frac{1}{2})|}{4j(2j-1)t^{2j-1}} > 0$, and $\widetilde{R}_{k+1}(t)$ is the remainder after taking k terms in the sum.

Error bounds

Using our error bounds on Gauss's asymptotic expansion for $\ln \Gamma(z + \frac{1}{2})$ in the case $z = it$, we get several bounds for the error $\widetilde{R}_{k+1}(t)$ in the asymptotic expansion of $\vartheta(t)$:

$$|\widetilde{R}_{k+1}(t)| \le \frac{\pi^{1/2}\,\Gamma(k - \frac{1}{2})\,|B_{2k}|}{8\,k!\,t^{2k-1}}\,,$$

and if $k \ge 3$ or $t \ge 1$ then

$$\left|\frac{\widetilde{R}_{k+1}(t)}{\widetilde{T}_k(t)}\right| < \sqrt{\pi k}$$

and

$$\left|\frac{\widetilde{R}_k(t)}{\widetilde{T}_k(t)}\right| < 1 + \sqrt{\pi k}.$$

The arctan term

What if we omit the term $\frac{1}{2}\arctan\left(e^{-\pi t}\right)$ in the asymptotic approximation to $\vartheta(t)$? This always seems to be done in the literature (Lehmer, Edwards, Gabcke, ...).

We still get a valid asymptotic expansion in the sense of Poincaré. However, numerically the approximation can be much worse.

The error bound has to be increased to compensate, e.g. we could replace our second error bound by

$$|\widetilde{R}_{k+1}(t)| < \widetilde{T}_k(t)\sqrt{\pi k} + \tfrac{1}{2}e^{-\pi t}.$$

The term $\frac{1}{2}e^{-\pi t}$ is not always negligible, because

$$\min_{k\geq 1} \widetilde{T}_k(t)\sqrt{\pi k} \approx \tfrac{1}{2}e^{-2\pi t} \ll \tfrac{1}{2}e^{-\pi t}.$$

The smallest term(s)

Define

$$\widetilde{T}_{\min}(t) := \min_{k \geq 1} \widetilde{T}_k(t).$$

Let $k_{\min}(t)$ be the corresponding index, so $\widetilde{T}_{\min}(t) = \widetilde{T}_{k_{\min}}(t)$.

Using $|B_{2k}| = \dfrac{2(2k)!\,\zeta(2k)}{(2\pi)^{2k}}$, it is straightforward to show that

$$k_{\min}(t) = \left\lfloor \pi t + \tfrac{5}{4} + O\left(t^{-1}\right) \right\rfloor$$

and

$$\widetilde{T}_{\min}(t) = \frac{e^{-2\pi t}}{2\pi\sqrt{t}} \left(1 + O\left(t^{-1}\right)\right).$$

Attainable accuracy

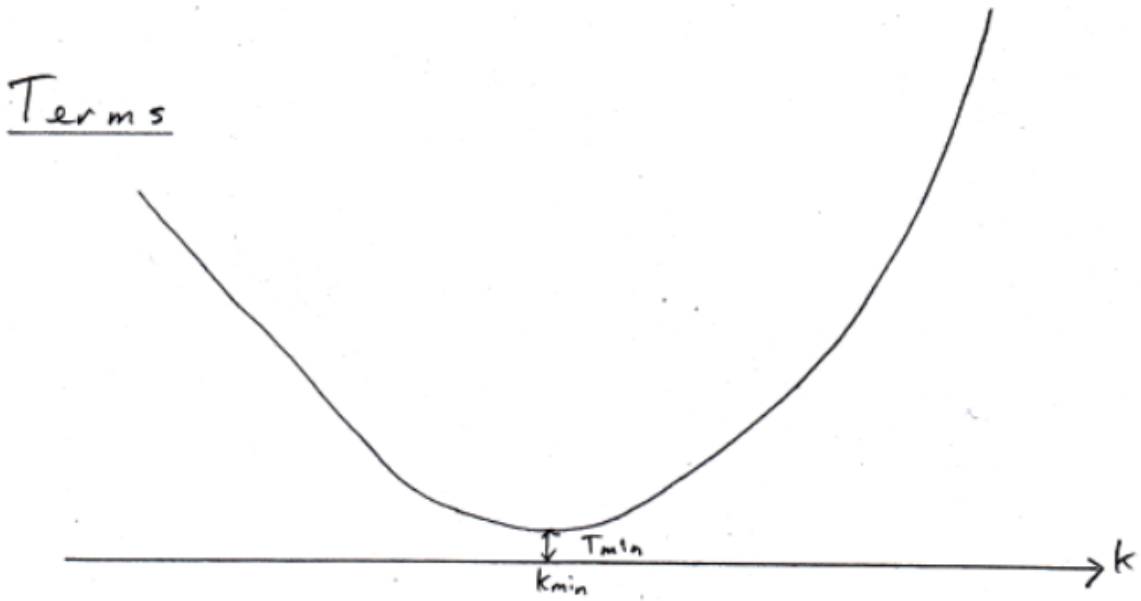
Using our error bound $|\widetilde{R}_{k+1}(t)| < \widetilde{T}_k(t)\sqrt{\pi k}$ with $k = k_{\min}$,
it is clear that we can guarantee an error of at most

$$\tfrac{1}{2} e^{-2\pi t}(1 + O(1/t))$$

if we truncate the asymptotic series for $\vartheta(t)$ after $k_{\min}(t)$ terms,
provided that we include the arctan term in the approximation.

Graphical interpretation: the terms

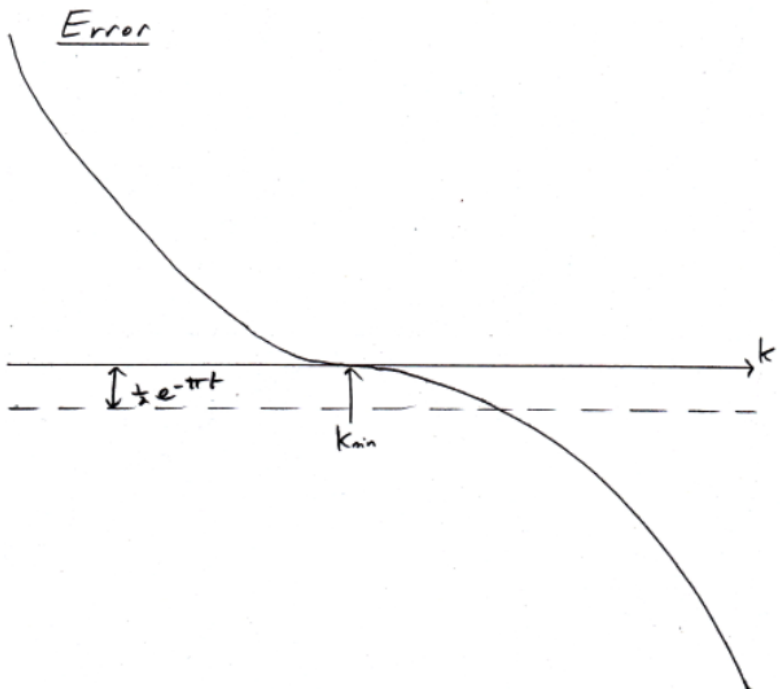
The terms $\widetilde{T}_k(t)$ are all positive. They decrease for $k \le k_{\min}$, then increase. Pictorially:



Close to $k_{\min}$, the terms approximate a (discretised) parabola $y = \widetilde{T}_{\min} + c(x - k_{\min})^2$.

Graphical interpretation: the error

The error curve $\widetilde{R}_{k+1}(t)$ approximates a cubic: $-$(integral of the previous curve).



If the $\frac{1}{2}\arctan(e^{-\pi t})$ term is included in the approximation, the curve crosses the k-axis close to $k = k_{\min}$.
If the $\frac{1}{2}\arctan(e^{-\pi t})$ term is omitted, the error curve is displaced upwards (drawn with the same curve but a dashed axis). The zero-crossing is near some $k \geq 2k_{\min}$.

Numerical results

t	$k_{\min}$	A	B	C
1	4	7.2×10^{1}	-0.79	-1.1×10^{-2}
2	7	2.4×10^{3}	-0.63	$+2.4\times10^{-4}$
5	16	4.6×10^{7}	-0.21	$+2.8\times10^{-3}$
10	32	4.4×10^{14}	-0.50	$+8.3\times10^{-4}$
20	64	2.7×10^{28}	-1.08	$+8.3\times10^{-5}$
50	158	3.7×10^{69}	-0.84	-1.5×10^{-4}
100	315	8.6×10^{137}	-0.76	-5.2×10^{-5}

The table gives

A : the error in the standard asymptotic approximation (no arctan term) after taking $k_{\min}(t)$ terms in the sum, normalised by the smallest term $\widetilde{T}_{\min}(t)$;

B : the same but including the $\frac{1}{2}\arctan(e^{-\pi t})$ term;

C : the error in an empirically improved approximation (next slide), again normalised by $\widetilde{T}_{\min}(t)$.

Observations

- The normalised value A is approximately $\pi t^{1/2} \exp(\pi t)$, which is large because $\widetilde{T}_{\min}(t)$ is much smaller than the error, which is about $\frac{1}{2} \exp(-\pi t)$.
- The entries in column B are negative. We would be better off truncating the sum after $k_{\min} - 1$ terms instead of $k_{\min}$ terms (which would have the effect of adding one to the entries in column B). However, a much better approximation is obtained by adding a "correction term"

$$\left(\pi t - k_{\min}(t) + \tfrac{1}{12}\right) \widetilde{T}_{\min}(t).$$

The motivation for the correction term is to smooth out the sawtooth nature of approximation B, which has jumps at the values of t where $k_{\min}(t)$ changes. This explains the addition of $(\pi t - k_{\min}(t) + c)\, \widetilde{T}_{\min}(t)$, where c is an arbitrary constant. Column C assumes that $c = \frac{1}{12}$.

The mysterious constant $\frac{1}{12}$

We do not have a theoretical explanation for the value of the constant $c \approx \frac{1}{12}$, although it is clearly related to the asymptotic location of the positive zero of the function $\widetilde{R}_{k+1}(t)$. (There should be a unique positive zero.)

By a theorem of Karl Dilcher (1987), for $u \in [-\frac{1}{2}, \frac{1}{2}]$,

$$\frac{B_{2k}(u+\frac{1}{2})}{B_{2k}} = \cos(2\pi u) + O(4^{-k}) \text{ uniformly as } k \to \infty.$$

Thus, in the expression

$$\widetilde{R}_{k+1}(t) = \Im\left(-\frac{1}{4k}\int_0^\infty \frac{B_{2k}(\{u+\frac{1}{2}\})}{(u+it)^{2k}}\,\mathrm{d}u\right),$$

we may be able to approximate $B_{2k}(\{u+\frac{1}{2}\})$ by $B_{2k}\cos(2\pi u)$, which could make the problem more tractable.

Olver's example

We have seen that an exponentially small term $\approx \frac{1}{2}e^{-\pi t}$ can be significant in the numerical approximation of $\vartheta(t)$.

This is not an isolated example. Olver (1964) gives a simpler example, which is discussed by Meyer (1989) in a more accessible paper. Briefly,

$$F(n) := \int_0^\pi \frac{\cos(nt)}{t^2+1}\,\mathrm{d}t$$

has (for large integer $n > 0$) an asymptotic expansion

$$F(n) \sim (-1)^{n-1} \sum_{k\geq 1} \lambda_k n^{-2k},$$

where $\lambda_1 \approx 0.05318, \lambda_2 \approx 0.04791, \ldots$ However, this gives a poor approximation (16% relative error) for $n = 10$ because it does not allow for a term $\frac{\pi}{2}e^{-n}$ that arises because the integrand has poles at $t = \pm i$.

References

M. V. Berry, The Riemann-Siegel expansion for the zeta function: high orders and remainders, *Proc. R. Soc. Lond. A* **450** (1995), 439–462.

R. P. Brent, On the zeros of the Riemann zeta function in the critical strip, *Math. Comp.* **33** (1979), 1361–1372.

R. P. Brent, *On asymptotic approximations to the log-Gamma and Riemann-Siegel theta functions*, arXiv:1609.03682, 13 Sept. 2016.

K. Dilcher, Asymptotic behaviour of Bernoulli, Euler, and generalized Bernoulli polynomials, *J. Approximation Theory* **49** (1987), 321–330.

H. M. Edwards, *Riemann's Zeta Function*, Academic Press, New York, 1974; reprinted by Dover Publications, 2001.

W. Gabcke, *Neue Herleitung und Explizite Restabschätzung der Riemann-Siegel-Formel*, Ph.D. thesis, Göttingen, 1979.

References

J.-P. Gram, Note sur les zéros de la fonction $\zeta(s)$ de Riemann, *Acta Mathematica* **27** (1908), 289–304.

D. E. G. Hare, Computing the principal branch of log-Gamma, *J. of Algorithms* **25** (1997), 221–236.

R. E. Meyer, A simple explanation of the Stokes phenomenon, *SIAM Review* **31** (1989), 435–445.

F. W. J. Olver, Chapter in *Asymptotic Solutions of Differential Equations and their Applications*, C. H. Wilcox (ed.), Wiley, NY, 1964, 163–183. (See also Meyer (1989).)

F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, New York, 1974.

R. Spira, Calculation of the Gamma function by Stirling's formula, *Math. Comp.* **25** (1971), 317–322.

E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 3rd ed., Cambridge Univ. Press, 1920.