Article

Analytic error function and numeric inverse obtained by geometric means

Dmitri Martila † and Stefan Groote *,†

Institute of Physics, University of Tartu, W. Ostwaldi 1, 50411 Tartu, Estonia; eestidima@gmail.com
* Correspondence: groote@ut.ee
† These authors contributed equally to this work.

Abstract: Using geometric considerations, we provide a clear derivation of the integral representation for the error function, known as the Craig formula. We calculate the corresponding power series expansion and prove the convergence. The same geometric means finally help to systematically derive handy formulas that approximate the inverse error function. Our approach can be used for applications in e.g. high-speed Monte Carlo simulations where this function is used extensively.

Keywords: error function; analytic function; inverse error function; approximations

MSC: 62E15; 62E17; 60E15; 26D15

1. Introduction

High-speed Monte Carlo simulations are used for a large spectrum of applications from mathematics to economy. As input for such simulations, the probability distribution are usually generated by pseudo-random number sambling, a method going back to a work of John von Neumann from 1951 [1]. In the era of "big data" such methods have to be fast and reliable, a sign of such neccessity being the release of the very first randomness Quside processing unit in 2023 [2]. Still, these samblings need to be cross-checked by exact methods, and for these the knowledge of analytical functions to describe the stochastic processes, among those the error function, are of tremendous importance.

By definition, a function is called analytic if it locally given by a converging Taylor series expansion. Even if a function itself turns out not to be analytic, its inverse can be analytic. The error function can be given analytically, one of these analytic expressions is the integral representation given by Craig in 1991 [3]. Craig mentioned this representation only in passing and did not give a derivation of it. In the following, there have been a couple of derivations of this formula [4–6]. In Sec. 2 we add a further one which is based on the same geometric considerations as employed in Ref. [7]. In Sec. 3 we give the series expansion for Craig's integral representation and show the fast convergence of this series.

For the inverse error function, handbooks for special functions (cf. e.g. Ref. [8]) do not unveil such an analytic property. Instead, this function have to be approximated. Known approximations are dating back to the late 1960s and early 1970s [9,10]) and reach up to semi-analytical approximations by asymptotic expansion (cf., e.g., Refs. [11–15,Soranzo]. Using the same geometric considerations, in Sec. 4 we develop a couple of handy approximations which can easily be implemented in different computer languages, indicating the deviations from an exact treatment. In Sec. 5 we test the CPU time and give our conclusions.

 doi:10.20944/preprints202302.0476.v1

2. Derivation of Craig's integral representation

Ref. [7] provides an approximation for the Gaussian normal distribution obtained by geometric considerations. The same considerations apply to the error function $\operatorname{erf}(t)$ which is given by the normal distribution $P(t)$ via

$$\operatorname{erf}(t)=\frac{1}{\sqrt{\pi}}\int_{-t}^{t}e^{-x^2}dx=\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{2}t}^{\sqrt{2}t}e^{-x^2/2}dx=P(\sqrt{2}t). \tag{1}$$

Translating the results of Ref. [7] to the error function, one obtains the approximation of order p to be

$$\operatorname{erf}_p(t)^2=1-\frac{1}{N}\sum_{n=1}^{N}e^{-k_n^2t^2}, \tag{2}$$

where the $N=2^p$ values k_n $(n=1,2,\ldots,N)$ are found in the intervals between $1/\cos(\pi(n-1)/(4N))$ and $1/\cos(\pi n/(4N))$ and can be optimised. In Ref. [7] it is shown that

$$\left|\operatorname{erf}(t)-\sqrt{1-e^{-k_{0,1}^2t^2}}\right|<0.0033 \tag{3}$$

for $k_{0,1}=1.116$, and with $14\approx 0.0033/0.00024$ times larger precision

$$\left|\operatorname{erf}(t)-\sqrt{1-\frac{1}{2}(e^{-k_1^2t^2}+e^{-k_2^2t^2})}\right|<0.00024, \tag{4}$$

where $k_{1,1}=1.01$, $k_{1,2}=1.23345$. For the parameters k_n taking the values $k_n=1/\cos(\pi n/(4N))$ of the upper limits of those intervals, it can be shown that the deviation is given by

$$|\operatorname{erf}(t)-\operatorname{erf}_p(t)|<\frac{\exp(-t^2)}{2N}\sqrt{1-\exp(-t^2)}. \tag{5}$$

Given the values $k_n=1/\cos\phi(n)$ with $\phi(n)=\pi n/(4N)$, in the limit $N\to\infty$ the sum over n in Eq. (2) can be replaced by an integral with measure $dn=(4N/\pi)d\phi(n)$ to obtain

$$\operatorname{erf}(t)^2=1-\frac{4}{\pi}\int_0^{\pi/4}\exp\left(\frac{-t^2}{\cos^2\phi}\right)d\phi. \tag{6}$$

3. Power series expansion

The integral in Eq. (6) can be expanded into a power series in t^2,

$$\operatorname{erf}(t)^2=1-\frac{4}{\pi}\sum_{n=0}^{\infty}c_n\frac{(-1)^n}{n!}(t^2)^n \tag{7}$$

with

$$\begin{aligned} c_n &= \int_0^{\pi/4}\frac{d\phi}{\cos^{2n}\phi}=\int_0^{\pi/4}(1+\tan^2\phi)^n d\phi=\int_0^1(1+y^2)^{n-1}dy\\ &= \sum_{k=0}^{n-1}\binom{n-1}{k}\int_0^1 y^{2k}dy=\sum_{k=0}^{n-1}\frac{1}{2k+1}\binom{n-1}{k}, \end{aligned} \tag{8}$$

where $y=\tan\phi$. The coefficients c_n can be expressed by the hypergeometric function, $c_n={}_2F_1(1/2,1-n;3/2;-1)$, also known as Barnes' extended hypergeometric function. On the other hand, we can derive a constraint for the explicit finite series expression for c_n that renders the series in Eq. (7) to be convergent for all values of t. In order to be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2023 doi:10.20944/preprints202302.0476.v1

self-contained, intermediate steps to derive this constraint and to show the convergence are shown in the following. Necessary is Pascal's rule

$$\begin{aligned}\binom{n}{k}+\binom{n}{k-1} &= \frac{n!}{k!(n-k)!}+\frac{n!}{(k-1)!(n-k+1)!}\\ &= \frac{n!(n-k+1+k)}{k!(n-k+1)!} = \frac{(n+1)!}{k!(n+1-k)!} = \binom{n+1}{k}\end{aligned} \tag{9}$$

and the sum over the rows of Pascal's triangle,

$$\sum_{k=0}^{n}\binom{n}{k} = 2^n \tag{10}$$

which can be shown by mathematical induction. The base case $n = 0$ is obvious, as $\binom{0}{0} = 1 = 2^0$. For the induction step from n to $n+1$ we write the first and last elements $\binom{n+1}{0} = 1$ and $\binom{n+1}{n+1} = 1$ separately and use Pascal's rule to obtain

$$\begin{aligned}\sum_{k=0}^{n+1}\binom{n+1}{k} &= 1+\sum_{k=1}^{n}\binom{n+1}{k}+1 =\\ &= 1+\sum_{k=1}^{n}\binom{n}{k}+\sum_{k=1}^{n}\binom{n}{k-1}+1 = 2\sum_{k=0}^{n}\binom{n}{k} = 2^{n+1}.\end{aligned} \tag{11}$$

This proves Eq. (10). Returning to Eq. (8), one has $0 \le k \le n-1$ and, therefore,

$$\frac{1}{2n-1} \le \frac{1}{2k+1} \le 1. \tag{12}$$

For the result in Eq. (8) this means that

$$\frac{1}{2n-1}\sum_{k=0}^{n-1}\binom{n-1}{k} \le c_n \le \sum_{k=0}^{n-1}\binom{n-1}{k} = 2^{n-1}, \tag{13}$$

i.e., the existence of a real number c_n^* between $1/(2n-1)$ and 1 such that $c_n = c_n^* 2^{n-1}$. One has

$$\mathrm{erf}_p(t)^2 = 1-\frac{4}{\pi}\sum_{n=0}^{N}c_n\frac{(-1)^n}{n!}(t^2)^n = 1-\frac{2}{\pi}\sum_{n=0}^{N}c_n^*\frac{(-2t^2)^n}{n!}, \tag{14}$$

and because of $0 \le c_n^* \le 1$ there is again a real number c_N^{**} in the corresponding open interval so that

$$\frac{2}{\pi}\sum_{n=0}^{N}c_n^*\frac{(-2t^2)^n}{n!} = c_N^{**}\frac{2}{\pi}\sum_{n=0}^{N}\frac{(-2t^2)^n}{n!} < \frac{2}{\pi}\sum_{n=0}^{N}\frac{(-2t^2)^n}{n!}. \tag{15}$$

As the latter is the power series expansion of $(2/\pi)e^{-2t^2}$ which is convergent for all values of t, also the original series is convergent and, therefore, $\mathrm{erf}_p(t)^2$ with the limiting value shown in Eq. (7). A more compact form of the power series expansion is given by

$$\mathrm{erf}(t)^2 = \sum_{n=1}^{\infty}c_n\frac{(-1)^{n-1}}{n!}(t^2)^n, \qquad c_n = \sum_{k=0}^{n-1}\frac{1}{2k+1}\binom{n-1}{k}. \tag{16}$$

4. Approximations for the inverse error function

Based on the geometric approach from Ref. [7], in the following we describe how to find simple, handy formulas that, guided by higher and higher orders of the approximation (2) for the error function lead to more and more advanced approximation of the inverse error function. Starting point is the degree $p = 0$, i.e., the approximation in Eq. (3). Inverting

$E = \mathrm{erf}_0(t) = (1 - e^{-k_{0,1}^2 t^2})^{1/2}$ leads to $t^2 = -\ln(1 - E^2)/k_{0,1}^2$, and using the parameter $k_{0,1} = 1.116$ from Eq. (3) gives

$$T_0 = \sqrt{-\ln(1 - E^2)}/k_{0,1}^2. \tag{17}$$

For $0 \leq E \leq 0.92$ the relative deviation $(T_{(0)} - t)/t$ from the exact value t is less than 1.11%, for $0 \leq E < 1$ the deviation is less than 10%. Therefore, for $E > 0.92$ a more precise formula has to be used. As such higher values for E appear only in 8% of the cases, this will not essentially influence the CPU time.

Continuing with $p = 1$, we insert $T_0 = \sqrt{-\ln(1 - E^2)}/k_{0,1}^2$ into Eq. (2) to obtain

$$\mathrm{erf}_1(T_0) = \sqrt{1 - \frac{1}{2}(e^{-k_{1,1}^2 T_0^2} + e^{-k_{1,2}^2 T_0^2})}, \tag{18}$$

where $k_{1,1} = 1.01$ and $k_{1,2} = 1.23345$ are the same as for Eq. (4). Taking the derivative of Eq. (1) and approximating this by the difference quotient, one obtains

$$\frac{\mathrm{erf}(t) - \mathrm{erf}(T_0)}{t - T_0} = \frac{\Delta\,\mathrm{erf}(t)}{\Delta t}\Big|_{t=T_0} \approx \frac{d\,\mathrm{erf}(t)}{dt}\Big|_{t=T_0} = \frac{2}{\sqrt{\pi}} e^{-T_0^2}, \tag{19}$$

leading to $t \approx T_1 = T_0 + \frac{1}{2}\sqrt{\pi} e^{T_0^2}(E - \mathrm{erf}_1(T_0))$. In this case, for in the larger interval $0 \leq E \leq 0.995$ the relative deviation $(T_1 - t)/t$ is less than 0.1%. Using $\mathrm{erf}_2(t)$ instead of $\mathrm{erf}_1(t)$ and inserting T_1 instead of T_0 one obtains T_2 with a relative deviation of maximally 0.01% for the same interval. The results are shown in Fig. 1.

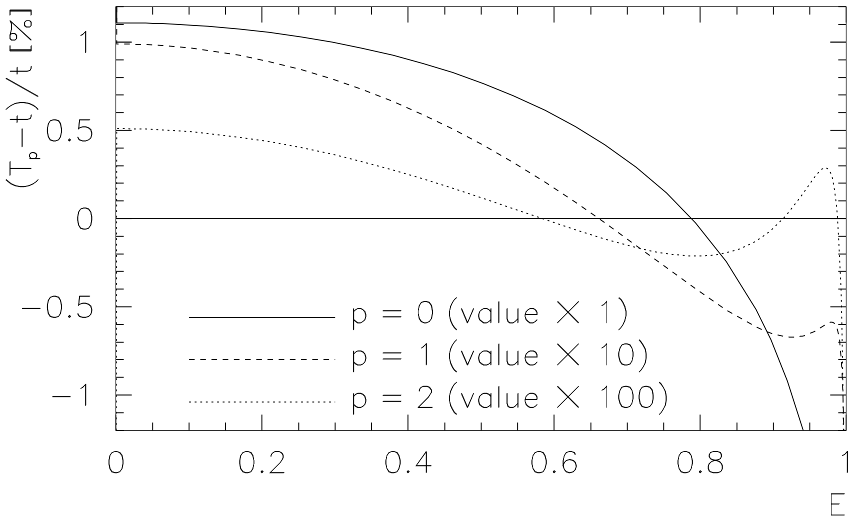


Figure 1. Relative deviations for the statical approximations

The method to can be optimised by a method similar to the shooting method in boundary problems, giving dynamics to the calculation. Suppose that following one of the previous methods, for a particular argument E we have found an approximation t_0 for the value of the inverse error function at this argument. Using $t_1 = 1.01 t_0$, one can adjust the improved result

$$t = t_0 + A(E - \mathrm{erf}(t_0)) \tag{20}$$

by inserting $E = \mathrm{erf}(t)$ and and calculating A for $t = t_1$. In general, this procedure gives a vanishing deviation close to $E = 0$. In this case and for $t_0 = T_1$, in the interval $0 \leq E \leq 0.7$ the maximal deviation is slightly larger than $10^{-6} = 0.0001\%$ while up to $E = 0.92$ the deviation is restricted to $10^{-5} = 0.001\%$. A more general ansatz

$$t = t_0 + A(E - \mathrm{erf}(t_0)) + B(E - \mathrm{erf}(t_0))^2 \tag{21}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2023 doi:10.20944/preprints202302.0476.v1

can be adjusted by inserting $E = \text{erf}(t)$ for $t = 1.01t_0$ and $t = 1.02t_0$, and the system of equations

$$\Delta t = A\Delta E_1 + B\Delta E_1^2, \qquad 2\Delta t = A\Delta E_2 + B\Delta E_2^2 \tag{22}$$

with $\Delta t = 0.01t_0$, $\Delta E_i = \text{erf}(t_i) - \text{erf}(t_0)$ can be solved for A and B to obtain

$$A = -\frac{(2\Delta E_1^2 - \Delta E_2^2)\Delta t}{\Delta E_1\Delta E_2(\Delta E_1 - \Delta E_2)}, \quad B = \frac{(-2\Delta E_1 + \Delta E_2)\Delta t}{\Delta E_1\Delta E_2(\Delta E_1 - \Delta E_2)}. \tag{23}$$

For $0 \leq E \leq 0.70$ one obtains a relative deviation of $1.5 \cdot 10^{-8}$, for $0 \leq E \leq 0.92$ the maximal deviation is $5 \cdot 10^{-7}$. Finally, the adjustment of

$$t = t_0 + A(E - \text{erf}(t_0)) + B(E - \text{erf}(t_0))^2 + C(E - \text{erf}(t_0))^3 \tag{24}$$

leads to

$$\begin{aligned} A &= (3\Delta E_1^2\Delta E_2^2(\Delta E_1 - \Delta E_2) - 2\Delta E_1^2\Delta E_3^2(\Delta E_1 - \Delta E_3) \\ &\quad + \Delta E_2^2\Delta E_3^2(\Delta E_2 - \Delta E_3))\Delta t/D, \\ B &= (-3\Delta E_1\Delta E_2(\Delta E_1^2 - \Delta E_2^2) + 2\Delta E_1\Delta E_3(\Delta E_1^2 - \Delta E_3^2) \\ &\quad - \Delta E_2\Delta E_3(\Delta E_2^2 - \Delta E_3^2))\Delta t/D, \\ C &= (3\Delta E_1\Delta E_2(\Delta E_1 - \Delta E_2) - 2\Delta E_1\Delta E_3(\Delta E_1 - \Delta E_3) \\ &\quad + \Delta E_2\Delta E_3(\Delta E_2 - \Delta E_3))\Delta t/D, \end{aligned} \tag{25}$$

where $D = \Delta E_1\Delta E_2\Delta E_3(\Delta E_1 - \Delta E_2)(\Delta E_1 - \Delta E_3)(\Delta E_2 - \Delta E_3)$. For $0 \leq E \leq 0.70$ the relative deviation is restricted to $5 \cdot 10^{-10}$ while up to $E = 0.92$ the maximal relative deviation is $4 \cdot 10^{-8}$. The results for the deviations of $T_{(n)}$ $(n = 1, 2, 3)$ for linear, quadratic and cubic dynamical approximation are shown in Fig. 2.

5. Conclusions

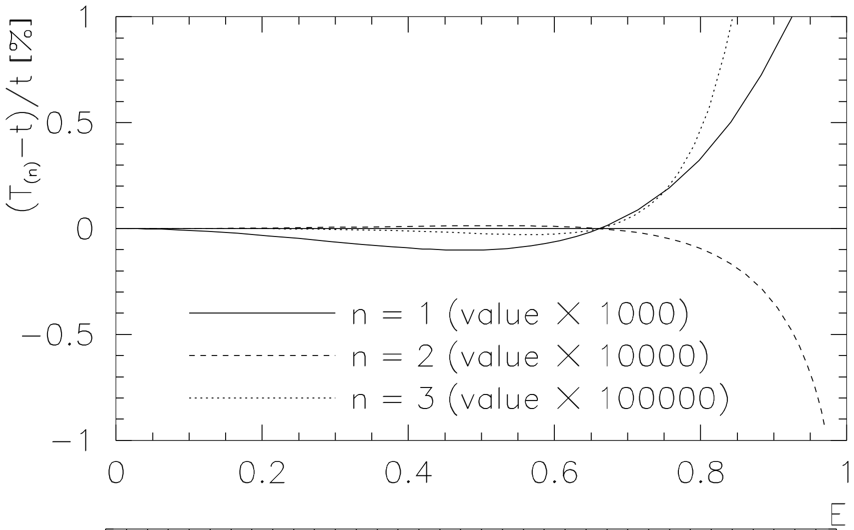
In order to test the feasibility and speed, we have coded our algorithm in the computer language C under `Slackware 15.0` on an ordinary hp laptop. The dependence of the CPU time for the calculation is estimated by calculating the value 10^6 times in sequence. The speed of the calculation of course turns does not depend on the value for E, as the precision is not optimised. This would have to be neccessary for a practical application. For an arbitrary starting value $E = 0.8$ we perform this test, and the results are given in Table 1. An analysis of the table shows that a further step in the degree p doubles the run time while the dynamics for increasing n adds a constant value of approximately 0.06 seconds to the result. Despite the fact the increase of the dynamics needs the solution of a linear system of equations and the coding of the result, this endeavour is justified, as by using the dynamics one can increase the precision of the result without loosing calculational speed.

Table 1. Run time experiment for our algorithm under C for $E = 0.8$ and different values of n and p (CPU time in seconds). As indicated, the errors are in the last displayed digit, i.e., ± 0.01 seconds.

	$n = 0$	$n = 1$	$n = 2$	$n = 3$
$p = 0$	0.06(1)	0.12(1)	0.15(1)	0.18(1)
$p = 1$	0.13(1)	0.19(1)	0.22(1)	0.25(1)
$p = 2$	0.24(1)	0.29(1)	0.33(1)	0.36(1)

Author Contributions: Conceptualization, D.M.; methodology, D.M. and S.G.; writing—original draft preparation, D.M.; writing—review and editing, S.G. and D.M.; visualization, S.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Regional Development Fund under Grant No. TK133.



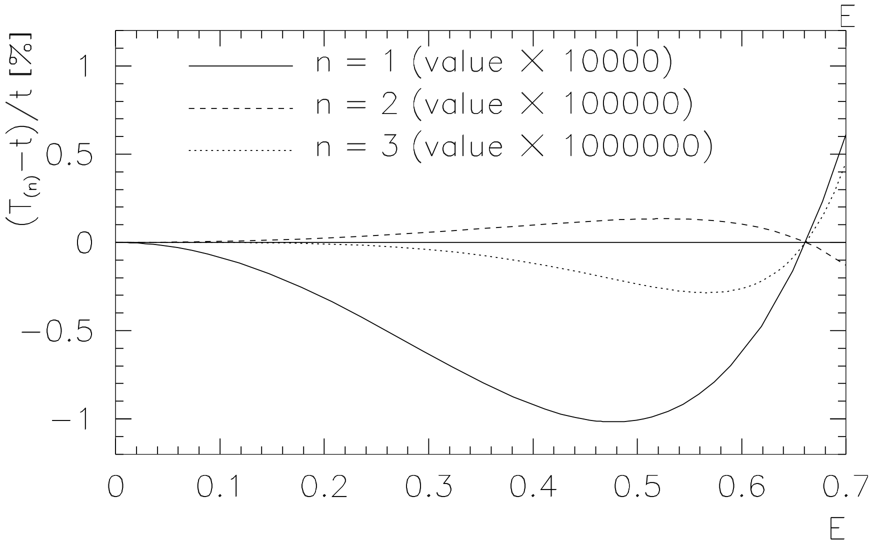


Figure 2. Relative deviation for the dynamical approximations (the degree is chosen to be $p = 1$)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. von Neumann, J. Various Techniques Used in Connection with Random Digits. in Householder, A.S.; Forsythe, G.E.;and Germond, H.H. (eds.). Monte Carlo Methods. *National Bureau of Standards Applied Mathematics Series* **1951** *12*, 36–38.
2. https://quside.com/quside-unveils-the-worlds-first-randomness-processing-unit
3. Craig, J.W. A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations. *Proceedings of the 1991 IEEE Military Communication Conference* **1991**, *2*, 571–575.
4. Lever, K.V. New derivation of Craig's formula for the Gaussian probability function. *Electronics Letters* **1998** *34*, 1821–1822.
5. Tellambura, C; Annamalai, A. Derivation of Craig's formula for Gaussian probability function. *Electronics Letters* **1999** *35*, 1424–1425.
6. Stewart, S.M. Some alternative derivations of Craig's formula. *The Mathematical Gazette* **2017** *101*, 268–279.
7. Martila, D; Groote, S. Evaluation of the Gauss Integral. *Stats* **2022** *5*, 538–545.
8. Andrews, L.C. Special functions of mathematics for engineers. *SPIE Press* **1998**, 110.
9. Strecok, A.J. On the Calculation of the Inverse of the Error Function. *Math. Comp.* **1968** *22*, 144–158.
10. Blair, J.M; Edwards, C.A.; Johnson, J.H. Rational Chebyshev Approximations for the Inverse of the Error Function. *Math. Comp.* **1976** *30*, 827–830.

doi:10.20944/preprints202302.0476.v1

11. Bergsma, W.P. A new correlation coefficient, its orthogonal decomposition and associated tests of independence. arXiv:math/0604627 [math.ST].
12. Dominici, D. Asymptotic analysis of the derivatives of the inverse error function. arXiv:math/0607230 [math.CA].
13. Dominici, D; Knessl, C. Asymptotic analysis of a family of polynomials associated with the inverse error function. arXiv:0811.2243 [math.CA].
14. Winitzki, S. A handy approximation for the error function and its inverse. `https://www.academia.edu/9730974/`
15. Giles, M. Approximating the erfinv function. in *GPU Computing Gems Jade Edition* **2011** 109–116.

Soranzo. Soranzo, A; Epure, E. Simply Explicitly Invertible Approximations to 4 Decimals of Error Function and Normal Cumulative Distribution Function. arXiv:1201.1320 [stat.CO].