

## SOME SIMPLE FULL-RANGE INVERSE-NORMAL APPROXIMATIONS

RAYMOND F. KOOPMAN\*

**Abstract.** Two approximations are given for numerically inverting the normal distribution function. Both approximations minimize the maximum absolute error in the approximate  $z$  over the full range of the distribution. The first approximation has two parameters and only modest accuracy but is very simple; the second has four parameters and is substantially more accurate. Both approximations have smaller maximum absolute errors than others of equivalent complexity. Alternate versions of the approximations are also given, with their parameter values chosen to minimize the maximum relative error in the value of  $q$  implied by the approximate  $z$ , rather than the maximum absolute error in the approximate  $z$ . This “back-translation” approach to evaluating the quality of an approximation seems not to have been used before. Considering the two sets of approximations together draws attention to the issue of the relative seriousness of errors at different points in the range, which is a subject-field question.

**MSC.** 65D15.

**Keywords.** numerical approximation, inverse normal distribution function, inverse error function.

### 1. INTRODUCTION

Let  $f$  and  $\bar{F}$  denote the standard normal density and complementary distribution functions, respectively, and let  $q = \bar{F}(z)$  denote the upper-tail area. The intent is to approximate  $z$  numerically as a simple function of  $q$  with errors that are small over the full range of the distribution. Almost all the functions that have been proposed are accurate for either the central part of the distribution or the tail, but not both; such approximations must be either truncated, making no attempt to cover the full range, or concatenated with another function to extend the range, possibly creating a discontinuity (see, *e.g.*, Odeh and Evans, 1974 [13]; Beasley and Springer, 1977 [3]; Derenzo, 1977 [4]; Page, 1977 [14]; Hamaker, 1978 [5]; Schmeiser, 1979; Bailey, 1981 [2]; Lin, 1989, 1990 [9], [10]; Vedder, 1993 [16]; Moro, 1995 [12]; Acklam, 2000 [1]; Voutier, 2010 [17]; Soranzo and Epure, 2014 [15], Lipoth *et al.* 2022 [11]). Only Hastings (1955, sheets 67 and 68) [6], Hill and Davis (1973) [7],

---

\*Psychology Department, Simon Fraser University, Burnaby, BC, Canada,  
e-mail: [koopman@sfu.ca](mailto:koopman@sfu.ca).

and Winitzki (2008) [18] limit the errors over the full range and are relatively simple.

## 2. TWO NEW APPROXIMATIONS

The method used here takes advantage of symmetry: it requires  $q \leq 1/2$  and gives an estimate  $\hat{z} \geq 0$ ; if  $q > 1/2$ , then  $1 - q$  should be used in place of  $q$ , and  $\hat{z}$  should be given a negative sign. Let  $t = -2 \ln(2q)$ . Then  $t = z^2 + \ln r$ , where  $r = u^2/u_0^2$ ,  $u = f(z)/\bar{F}(z)$  is the upper-tail mean, and  $u_0 = f(0)/\bar{F}(0) = (2/\pi)^{1/2}$ . The general form of the approximation is  $\hat{z} = (t - \ln R(t))^{1/2}$ , where  $R$  is a rational function of  $t$  that approximates  $r$ , subject to  $R(0) = R'(0) = 1$ . The error is asymptotically zero:  $\hat{z} - z = \frac{\ln(r/R)}{(z^2 + \ln(r/R))^{1/2} + z}$ ; as  $z \rightarrow \infty$ ,  $u^2 \rightarrow z^2 + 2$ , so  $t \rightarrow u^2 - 2 + \ln R$ ,  $r/R \rightarrow$  a constant, and  $\hat{z} - z \rightarrow 0$ . Parameter values that optimize the fits were found by hand.

*Approximation 1:*  $R(t) = 1 + t + t^2/(at + b)$ ,  $a = 1.991162$ ,  $b = 10.05113$ ,  $|\hat{z} - z| < 1.19 \times 10^{-3}$ . Rounding  $a$  to 2 and  $b$  to 10 increases the maximum absolute error to  $1.25 \times 10^{-3}$ , but also reduces the absolute errors over a substantial interval. The maximum absolute error of Winitzki's approximation (one parameter) is  $2.76 \times 10^{-2}$ , after converting it from erf to normal and optimizing the parameter. The maximum absolute error of Hastings' sheet 67 (four parameters) is  $2.84 \times 10^{-3}$ .

Unlike many approximations, there is little to be gained by optimizing approximation 1 over a domain that excludes very small values of  $q$ . On the other hand, increasing the degree of  $R$  provides a substantial gain in accuracy without restricting  $q$ .

*Approximation 2:*  $R(t) = 1 + t + t^2(at + b)/(t^2 + ct + d)$ ,  $a = .5583971$ ,  $b = 6.268748$ ,  $c = 20.36848$ ,  $d = 58.95569$ ,  $|\hat{z} - z| < 8.43 \times 10^{-5}$ .

The maximum absolute error of Hastings' sheet 68 (six parameters) is  $4.47 \times 10^{-4}$ . The maximum absolute error of the Hill–Davis approximation (six parameters) is  $3.58 \times 10^{-4}$  if  $|z| \leq 304$ ; otherwise the limit is  $7.11 \times 10^{-4}$ .

Fig. 1 and Fig. 2 show the absolute error curves for approximations 1 and 2. The dashed curve in Fig. 1 shows the result of rounding  $a$  and  $b$  to integers. To display the asymptotic behavior of the approximations, the  $z$  axis in all the figures extends far beyond the usual single-digit range and is plotted in arcsinh units. The computations used Mathematica's arbitrary precision capability.

Simple approximations are no longer as widely needed as they once were, but where simplicity is still important, approximation 1 is so simple, especially with the rounded parameters, that it should be used if its error bound is acceptable; if it is not, then approximation 2 should be considered. No other approximations of comparable simplicity have such small absolute errors over the full range of  $z$ .

Aside from any immediate practical uses these approximations may have, the fact that they are so simple, yet still manage to cover the full range with

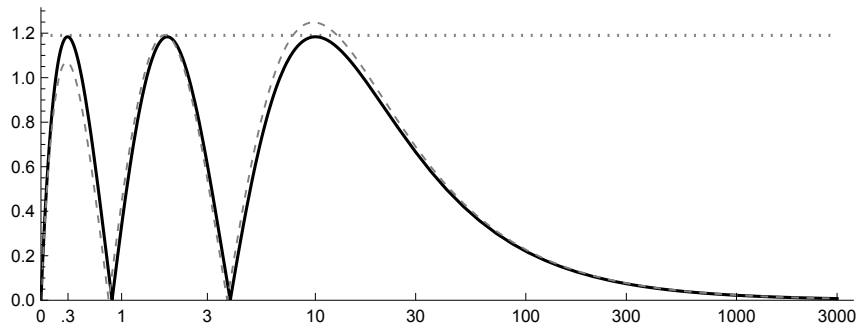


Fig. 1. Approximation 1:  $10^3|\hat{z} - z|$  as a function of  $z$ . Optimized parameters give the solid curve. Rounded parameters give the dashed curve.

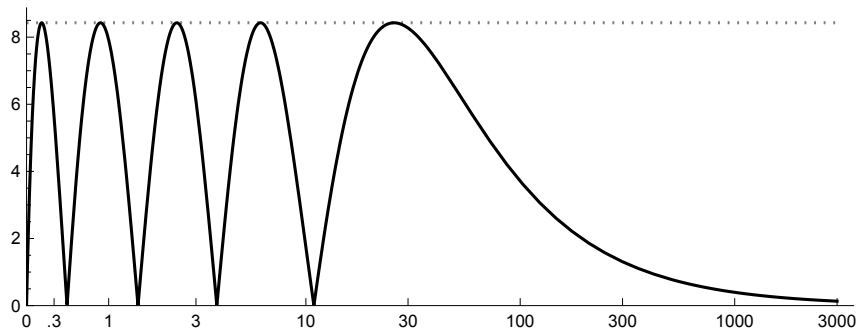


Fig. 2. Approximation 2:  $10^5|\hat{z} - z|$  as a function of  $z$ .

smaller errors than many more complex approximations, suggests that further study is warranted to see why they are so much better than the others and how far that advantage extends. Investigating the general form of the approximation  $(t - \ln R(t))^{1/2}$ , with  $R$  not necessarily a rational function, would accordingly be a logical next step, consonant with Howard's (2022) [8] suggestion "that finding approximations to the inverse error function is best treated directly and as a separate problem, rather than approaching it via finding the inverse of an approximation to the error function."

### 3. A DIFFERENT APPROACH

It may not always be appropriate to minimize the absolute error in  $\hat{z}$ . In general, the choice of a minimand ought to be based on the intended use of the approximate value and the effects that errors in it may have. For instance, if we were constructing a table of approximate critical values  $\hat{z}$  for given values of  $q$ , and if each implied  $\tilde{q} = \bar{F}(\hat{z})$  would be taken as if it were the given  $q$ , then it might be better to minimize the relative error<sup>1</sup> in  $\tilde{q}$  rather than the absolute

<sup>1</sup>We use  $\ln(\tilde{q}/q)$  instead of the more common  $(\tilde{q} - q)/q$  because relative error is a multiplicative concept, not an additive one, and thinking additively leads to the fallacy that  $x\%$  over and  $x\%$  under are equally serious and cancel one another. Some may say that an even

error in  $\hat{z}$ . This approach — comparing  $\tilde{q}$  to the original  $q$  — is analogous to evaluating a translation from one language to another by comparing its back-translation to the original. It seems to have not been used before to evaluate numeric approximations.

*Alternate approximation 1:*  $a = 1.80977$ ,  $b = 11.627$ ,  $|\ln(\tilde{q}/q)| < 5.84 \times 10^{-3}$ .

*Alternate approximation 2:*  $a = .5688184$ ,  $b = 14.97873$ ,  $c = 38.46443$ ,  $d = 145.668$ ,  $\ln(\tilde{q}/q) < 6.30 \times 10^{-4}$ .

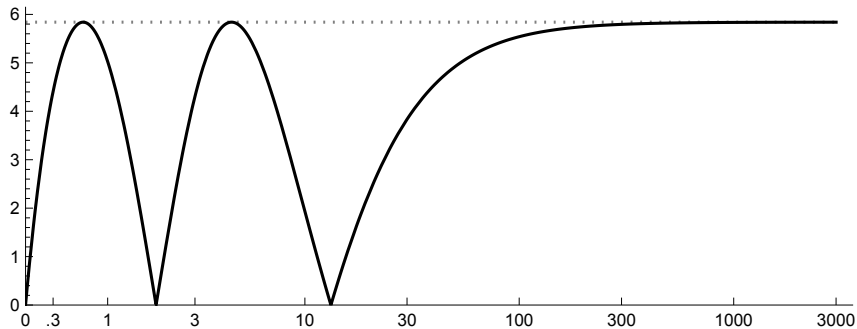


Fig. 3. Alternate approximation 1:  $10^3 |\ln(\tilde{q}/q)|$  as a function of  $z$ .

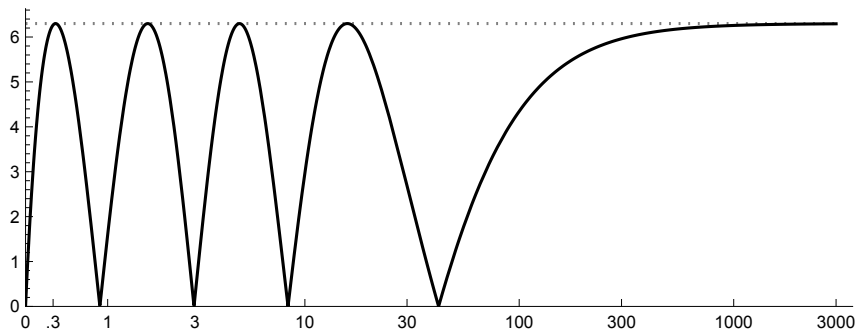


Fig. 4. Alternate approximation 2:  $10^4 |\ln(\tilde{q}/q)|$  as a function of  $z$ .

Fig. 3 and Fig. 4 show the relative error curves for the alternate approximations. The asymptotic relative error can be made zero by setting  $a$  to  $2/(\pi-2)$  in the first approximation or  $\pi/2-1$  in the second, but that would increase the maximum relative error even after adjusting the other parameters to equate the magnitudes of the extrema.
















The alternate approximations, too, may have some immediate practical uses, but there is a larger point to be made here. Minimizing the relative error in  $\tilde{q}$  is equivalent to minimizing the weighted error in  $\hat{z}$  using weights that increase with  $z$ , which is opposite to the weighting needed to minimize

---

better measure of relative error for probabilities would be the logit difference, the log of the odds ratio, because it treats the domain symmetrically.

the relative error in  $\hat{z}$ . The relative seriousness of errors in  $\hat{z}$  at different values of  $z$  will depend on what  $z$  represents and the effects that errors in  $\hat{z}$  will have, which is a subject-field question; there is unlikely to be one approximation that is best for all purposes. This is true for approximations in general, not just the inverse-normal.

#### REFERENCES

- [1] P. J. ACKLAM, *An algorithm for computing the inverse normal cumulative distribution function*, 2000. Online at <https://gist.github.com/roguetrainer/8188638> 
- [2] B. J. R. BAILEY, *Alternatives to Hastings' approximation to the inverse of the normal cumulative distribution function*, *Appl. Stat.*, **30** (1981), pp. 275–276. <https://doi.org/10.2307/2346351> 
- [3] J. D. BEASLEY, S. G. SPRINGER, *Algorithm AS 111: The percentage points of the normal distribution*, *Appl. Stat.*, **26** (1977) no. 1, 118–121. <https://doi.org/10.2307/2346889> 
- [4] S. E. DERENZO, *Approximations for hand calculators using small integer coefficients*, *Math. Comput.*, **31** (1977), pp. 214–222. <https://doi.org/10.1090/S0025-5718-1977-0423761-X> 
- [5] H. C. HAMAKER, *Approximating the cumulative normal distribution and its inverse*, *Appl. Stat.*, **27** (1978), pp. 76–77. <https://doi.org/10.2307/2346231> 
- [6] C. HASTINGS, *Approximations For Digital Computers*, Princeton, NJ: Princeton University Press, 1955.
- [7] G. W. HILL, A. W. DAVIS, *Algorithm 442: Normal deviate*, *Commun. ACM*, **16** (1973), pp. 51–52. <https://doi.org/10.1145/361932.361949> 
- [8] R. M. HOWARD, *Arbitrarily accurate analytical approximations for the error function*, *Math. Comput. Appl.*, **27** (2022). <https://doi.org/10.3390/mca27010014> 
- [9] J. T. LIN, *Approximating the normal tail probability and its inverse for use on a pocket calculator*, *Appl. Stat.*, **28** (1989), pp. 69–70. <https://doi.org/10.2307/2347681> 
- [10] J. T. LIN, *A simpler logistic approximation to the normal tail probability and its inverse*, *Appl. Stat.*, **29** (1990), pp. 255–257. <https://doi.org/10.2307/2347764> 
- [11] J. LIPOTH, Y. TEREDA, S. M. PAPALEXIOU, R. J. SPITERI, *A new very simply explicitly invertible approximation fo the standard normal cumulative distribution function*, *AIMS Mathematics*, **7** (2022), pp. 11635–11646. <https://doi.org/10.3934/math.2022648> 
- [12] B. MORO, *The full Monte*, *Risk*, **8** (1995), pp. 57–58.
- [13] R. E. ODEH, J. O. EVANS, *Algorithm AS 70: The percentage points of the normal distribution*, *Appl. Stat.*, **23** (1974), pp. 96–97. <https://doi.org/10.2307/2347061> 
- [14] E. PAGE, *Approximations to the cumulative normal function and its inverse for use on a pocket calculator*, *Appl. Stat.*, **26** (1977), pp. 75–76. <https://doi.org/10.2307/2346872> 
- [15] A. SORANZO, E. EPURE, *Very simply explicitly invertible approximations of normal cumulative and normal quantile function*, *Appl. Math. Sci.*, **8** (2014), pp. 4323–4341. <https://doi.org/10.12988/ams.2014.45338> 
- [16] J. D. VEDDER, *An invertible approximation to the normal distribution function*, *Comput. Stat. Data Anal.*, **16** (1993), pp. 119–123. [https://doi.org/10.1016/0167-9473\(93\)90248-R](https://doi.org/10.1016/0167-9473(93)90248-R) 
- [17] P. M. VOUTIER, *A new approximation to the normal distribution quantile function*, 2010. <https://doi.org/10.48550/arXiv.1002.0567> 
- [18] S. WINITZKI, *A handy approximation for the error function and its inverse*, 2008. Online at <https://issuu.com/julianprice/docs/erf-approx> 