AN APPROXIMATION TO THE PROBABILITY INTEGRAL

By J. D. Williams

United States Naval Ordnance Test Station, Inyokern, California

1. Summary. It is shown that

$$\frac{1}{\sqrt{2\pi}}\int_{-x}^{x} e^{-\frac{1}{2}t^2}\,dt \leq [1 - e^{-(2/\pi)x^2}]^{\frac{1}{2}}$$

and that the equality is never in error by as much as three-fourths of one percent. Other approximations are discussed.

2. For use on those occasions when an approximate analytic expression for the integral

$$p(x) = \frac{1}{\sqrt{2\pi}}\int_{-x}^{x} e^{-\frac{1}{2}t^2}\,dt \tag{1}$$

is desired, the approximation

$$p'(x) = [1 - e^{-(2/\pi)x^2}]^{\frac{1}{2}} \tag{2}$$

is simple and reasonably accurate. An approximation equivalent to this is quite commonly used in problems involving a bivariate normal distribution, but its use in the one-dimensional case seems to be less well known.

We shall first show that $p(x) \leq p'(x)$ and then estimate, by calculation, the relative error made when the equality is accepted.

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi}}\int_{-x}^{x} e^{-\frac{1}{2}t^2}\,dt \\ &= \left[\frac{1}{2\pi}\int_{-x}^{x}\int_{-x}^{x} e^{-\frac{1}{2}(t_1^2+t_2^2)}\,dt_1\,dt_2\right]^{\frac{1}{2}} \\ &\leq \left[\frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{(2x/\sqrt{\pi})} re^{-\frac{1}{2}r^2}\,dr\,d\theta\right]^{\frac{1}{2}} \\ &= [1 - e^{-(2/\pi)x^2}]^{\frac{1}{2}} = p'(x), \qquad \text{q.e.d.} \end{aligned} \tag{3}$$

The approximation, introduced at the stage of passage to polar coordinates, comprises replacement of the square region of integration $-x \leq x_i \leq x$ by a circular region, $0 \leq r \leq \frac{2}{\sqrt{\pi}}x$, having the same area. Since we are dealing with a circular normal distribution with zero means, the region of fixed area which covers the greatest density is a circle whose center is at the origin. Therefore our square region of area $4x^2$ must contain less density than the circular region of area $4x^2$ by which we have replaced it.

The maximum value of the relative error,

$$\epsilon_p = \frac{p'(x)}{p(x)} - 1, \tag{4}$$

is found by calculation to be about seven-tenths of one percent, as may be judged from Table 1, column 3.

The question may be asked: Can the relative error be reduced by suitable choice of the parameter c in

$$p'(x) = [1 - e^{-cx^2}]^{\frac{1}{2}}? \tag{5}$$

Calculation indicates that by taking $c = 0.6302$ the relative error is reduced to about one-half of one percent; but this gain is offset, for many purposes, by the loss of the inequality (3).

The density function implied by (2), namely

$$\rho'(x) = \frac{|x|}{\pi} e^{-(2/\pi)x^2}[1 - e^{-(2/\pi)x^2}]^{-\frac{1}{2}}, \tag{6}$$

has the variance

$$\sigma^2 = \pi (1 - \log 2) = 0.964. \tag{7}$$

If c is determined so that the density function will have unit variance, then (5) becomes

$$p'(x) = \left[1 - \left(\frac{e}{2}\right)^{-2x^2}\right]^{\frac{1}{2}}; \tag{8}$$

this approximation to (1) leads to relative errors of almost two percent, which occur when x is small.

The density function (6) may be used to judge the quality of (2) in approximating to an integral of the form

$$p(x_1, x_2) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}t^2} dt, \tag{9}$$

the approximation being

$$p'(x_1, x_2) = \tfrac{1}{2} [p'(x_2) - p'(x_1)] \tag{10}$$

when x_1 and x_2 are positive (which is the severe case). It is evident that the relative error in accepting (10) for (9) cannot exceed the greatest relative discrepancy ϵ_ρ, in the interval $x_1 \leq x \leq x_2$, between density function (6) and the normal density

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \tag{11}$$

The quantity

$$\epsilon_\rho = \frac{\rho'(x)}{\rho(x)} - 1 \tag{12}$$

is tabulated in Table 1, column 6, from which it appears that the relative error committed in using (10) for (9) will surely be less than one-and-a-half percent provided $0 \leq x_i \leq 1.8$; but the relative error may be very great when the interval of integration lies beyond $x = 1.8$.

The approximations described herein were suggested by the following situation, encountered in work done by the Applied Mathematics Panel, NDRC: The probability P of at least one success, defined by $-x \leq x_i \leq x$, in a sample

TABLE 1

x	$p'(x)$	$p(x)$	ϵ_p	$\rho'(x)$	$\rho(x)$	ϵ_ρ
.0	0	0	0	.3989	.3989	0
.1	.0797	.0797	.0002	.3969	.3970	.0005
.2	.1586	.1585	.0005	.3914	.3910	.0010
.3	.2360	.2358	.0008	.3821	.3814	.0018
.4	.3112	.3108	.0013	.3695	.3683	.0033
.5	.3836	.3829	.0018	.3539	.3521	.0051
.6	.4526	.4515	.0024	.3356	.3332	.0072
.7	.5177	.5161	.0031	.3151	.3123	.0089
.8	.5785	.5763	.0038	.2929	.2897	.0111
.9	.6347	.6319	.0044	.2695	.2661	.0128
1.0	.6862	.6827	.0051	.2454	.2420	.0141
1.1	.7329	.7287	.0058	.2211	.2179	.0147
1.2	.7747	.7699	.0063	.1971	.1942	.0149
1.3	.8118	.8064	.0067	.1738	.1714	.0140
1.4	.8443	.8385	.0069	.1516	.1497	.0127
1.5	.8725	.8664	.0070	.1306	.1295	.0085
1.6	.8967	.8904	.0070	.1113	.1109	.0036
1.7	.9171	.9109	.0068	.0937	.0940	−.0032
1.8	.9341	.9281	.0065	.0781	.0790	−.0114
1.9	.9485	.9426	.0063	.0640	.0656	−.0244
2.0	.9600	.9545	.0058	.0520	.0540	−.0370

of n pairs (x_1, x_2) from a population in which the independent component probabilities are $p(x)$, is

$$P = 1 - [1 - p^2(x)]^n. \tag{13}$$

A little numerical exploration, supplemented by examination of the limiting values as $x \to 0$ and $x \to \infty$, revealed that when P is fixed the quantity $\log n$ is very nearly a linear function, of slope minus two, of $\log x$; so nearly, in fact, that one was encouraged to posit the linearity and observe the consequences. This yielded (5), which became (2) by requiring that it go to zero with x in the same manner as (1).