INTEGRATING THE NORMAL CURVE

IVAN CHRISTOV

1. Introduction

Let us denote the normal distribution function as $N(s) = \int_0^s e^{-x^2}\mathrm{d}x$. Evaluating $N(s)$ arises often in the study of probability and statistics, since the area under the normal curve represents the probability of a certain event occurring. In this paper we will attempt to integrate e^{-x^2}. However the integral cannot be evaluated by traditional means (i.e. the fundamental theorem of calculus) because an analytical expression for the anti-derivative of e^{-x^2} does not exist. Thus we will endeavor to find a different way of approaching the problem. Eventually a method for finding the value of N for all s will be proposed.

2. Solution to the Problem

In order to evaluate $N(s)$ we will first evaluate $N^2(s)$ and then simply take the square root to find the desired values.

2.1. **Solution on an Infinite Interval.**

Before attempting to evaluate the normal distribution for all s, let us first consider the case when we are finding out all the area under the curve, i.e. $N(\infty)$. Thus let us consider $N^2(\infty)$,

$$N^2(\infty) = \left(\int_0^\infty e^{-x^2}\mathrm{d}x\right)^2 = \int_0^\infty e^{-x^2}\mathrm{d}x \int_0^\infty e^{-x^2}\mathrm{d}x.$$

Since the two integrals are independent of each other, we can change the integrating variable of the second integral from x to, say, y, thus

$$N^2(\infty) = \int_0^\infty e^{-x^2}\mathrm{d}x \int_0^\infty e^{-y^2}\mathrm{d}y.$$

The limits of integration do not depend on x or y, thus we can combine the product of the two integrals into a single double integral since the area of integration will not change. Consequently,

$$N^2(\infty) = \int_0^\infty\int_0^\infty e^{-x^2}e^{-y^2}\mathrm{d}x\mathrm{d}y = \iint_R e^{-(x^2+y^2)}\mathrm{d}A, \tag{1}$$

where $R = \{(x,y) \mid x, y \geq 0\}$ and $\mathrm{d}A$=$\mathrm{d}x\mathrm{d}y$.

Although it is not obvious, a change to polar coordinates at this point will let us evaluate the integral very easily using the fundamental theorem of calculus. By way of the following theorem (see p.853, [2])

$$\iint_S f(x,y)\mathrm{d}A = \int_\alpha^\beta\int_a^b f(r\cos\theta, r\sin\theta)r\mathrm{d}r\mathrm{d}\theta,$$

we change (1) from rectangular to polar coordinates. Hence,

$$N^2(\infty) = \iint_D re^{-r^2}\mathrm{d}A, \tag{2}$$

where $D = \{(r,\theta) \mid r \geq 0,\ 0 \leq \theta \leq \frac{\pi}{2}\}$ and dA=drdθ. An important note must go here: in order to achieve the same result, we must make sure that we are integrating over the same region, or that at least we achieve the same result when integrating over a different region. In general $R \neq D$, but since $\lim_{x\to\infty} e^{-x^2} = 0$ (we are considering the case of $x = \infty$ at this moment) and e^{-x^2} is monotonous, decreasing and $\left(e^{-x^2}\right) < \left(x^2 + y^2 = \infty\right)$ on $(0,\infty)$, then we will not be loosing any area when integrating over the sector D rather than over the square R.

Substituting the limits into the integral we achieve

$$\int_0^{\frac{\pi}{2}}\int_0^{\infty} re^{-r^2}\mathrm{d}r\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \lim_{a\to\infty}\left(-\frac{1}{2}e^{-a^2} + \frac{1}{2}\right)\mathrm{d}\theta = \frac{1}{2}\int_0^{\frac{\pi}{2}}\mathrm{d}\theta = \frac{\pi}{4}. \tag{3}$$

From (3) we see that $N^2(\infty) = \frac{\pi}{4}$, consequently $N(\infty) = \frac{\sqrt{\pi}}{2}$. Thus we have found at least one exact value of N. From the results of this section we can conclude that the area of the normal curve on $(-\infty,\infty)$ is **exactly** $\sqrt{\pi}$.

2.2. **Solution on a Finite Interval.** Now we find the value of $N(s)$ from $N^2(s)$ for any s. Let us rewrite (1) as,

$$N^2(s) = \iint_\Sigma e^{-\left(x^2+y^2\right)}\mathrm{d}A, \tag{4}$$

where $\Sigma = \{(x,y) \mid 0 \leq x, y \leq s\}$ and dA=dxdy. Now let us define $\Omega_1 = \{(x,y) \mid 0 \leq r \leq s, 0 \leq \theta \leq \frac{\pi}{2}\}$ and $\Omega_2 = \{(x,y) \mid 0 \leq r \leq s\sqrt{2}, 0 \leq \theta \leq \frac{\pi}{2}\}$. Simply put, the region Ω_1 defines the circle inscribed in Σ while the region Ω_2 describes the circle circumscribed over Σ. As we discussed earlier, e^{-x^2} is monotonous and decreasing on $(0,\infty)$ and thus (see p.306, [1])

$$\iint_{\Omega_1} < \iint_\Sigma < \iint_{\Omega_2}. \tag{5}$$

We can easily evaluate the double integrals over Ω_1 and Ω_2, because the regions are polar. Using the same method as when we evaluated (1) we transform the inequality in (5) to

$$\frac{\pi}{4}\left(1 - e^{-s^2}\right) < \iint_\Sigma < \frac{\pi}{4}\left(1 - e^{-2s^2}\right),$$

which in turn is

$$\frac{\sqrt{\pi}}{2}\left(\sqrt{1 - e^{-s^2}}\right) < N(s) < \frac{\sqrt{\pi}}{2}\left(\sqrt{1 - e^{-2s^2}}\right). \tag{6}$$

The last inequality tells us precisely where the value of $N(s)$ lies, however there does not exist a way of finding the value exactly. Thus the best approximate that can be offered at this point is

$$N(s) = \frac{\sqrt{\pi}}{4}\left(\sqrt{1-e^{-s^2}} + \sqrt{1-e^{-2s^2}}\right). \tag{7}$$

In other words the midpoint of the interval in (6).

3. Analysis and Conclusion

3.1. **Speed of Convergence of the Proposed Method.** Table 1 below summarizes the results of three methods of integrating the normal curve. In column one are shown the values of N accurate to nine decimal places [3]. In the second column we present the result of integrating the McLauren Series for e^{-x^2} with ten terms. In the third column are the results from using Simpson's $\frac{1}{3}$ Rule for numerically evaluating an integral, using 100 subdivisions. In the fourth column we have the results of the method derived in the previous section.

x	Exact	McLauren Series	Simpson's Rule	Method from (7)
10.0	0.886226925	$-1.31672402\text{x}10^{12}$	0.852893592	0.886226925
6.0	0.886226925	-98744195	0.866266925	0.886226925
4.0	0.886226912	-24195.970	0.872893573	0.886226901
3.0	0.886207348	-76.48332393	0.876203293	0.886199579
2.5	0.885866274	-1.05976082	0.877481543	0.885798188
2.0	0.882081391	0.861525336	0.875033411	0.882075889
1.5	0.856188394	0.856133349	0.849571416	0.859756818
1.0	0.746824133	0.746824121	0.739775095	0.764341297
0.5	0.461281006	0.461281006	0.455710617	0.486356708

Table 1

From the data in Table 1 we see that method proposed here is actually quite accurate for large s and is quickly convergent – seven digit accuracy at $s = 4$. We see that Simpson's Rule is consistently at one digit accuracy, and the McLauren Series diverges rapidly for moderate and large s, although it is very accurate around $s = 0$.

3.2. **Closing Remarks.** In conclusion, we have managed to find an accurate and computationally inexpensive method for evaluating $N(s)$. In the process we have also found one exact value of the function, and that is $N(\infty)$. From that value we inferred that the area under the entire normal curve is $\sqrt{\pi}$.

References

[1] Widder, D. V., *Advanced Calculus*, Prentice-Hall, New York, 1947.
[2] Stewart, *Calculus 3 ed.*, Brooks/Cole, Pacific Grove 1995.
[3] McClave, J. T. and Sincich, T., *A First Course in Statistics 6 ed.*, Prentice-Hall, New York 1997.