The Normal Probability Distribution Function: An Alternate Derivation

Ishtiaq A. Qazi
Institute of Environmental Science and Engineering, School of Civil and Environmental Engineering,
National University of Sciences and Technology, Rawalpindi, Pakistan
ishtiaq_qazi@iese.edu.pk

Abstract

Considering the manner in which the hits along an axis are distributed on a common target board, a frequency distribution function is developed. This function has two arbitrary constants which are determined using simple mathematical techniques. This leads to an alternate derivation of the normal distribution function which does not make use of the polar coordinates and double integration over two variables.

Introduction

The standard method of derivation of the normal probability distribution function involves changing form Cartesian to Polar coordinates followed by double integration over 'θ' and 'r' [1, 2]. To most students the sudden switch of coordinates seems odd and they feel uncomfortable with the derivation process. This becomes particularly difficult for those not thoroughly familiar with the polar coordinate system.

Following other authors [3, 4, 5], an alternate derivation of the function has been developed below which it is hoped would be conceptually more palatable to the reader.

The Target Board

Let us consider a circular target board commonly used in fire-arm (pistol, rifle) or archery contests (Figure 1).

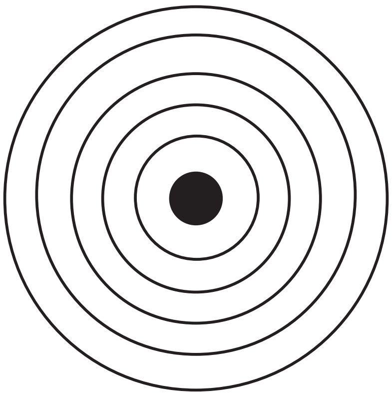

Fig. 1. Target board commonly used in fire-arm and archery contests

Considering the 'Bull's Eye' as the origin, let us try to develop the frequency distribution function f(x) for the hits scored, along the x-axis (Figure 2).

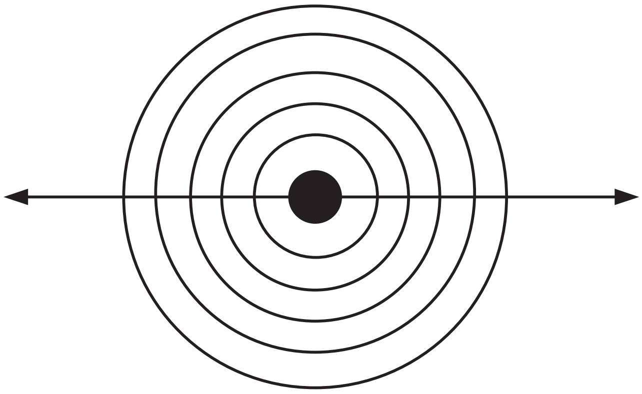

Fig. 2. The horizontal axis of the target board

Our intuition tells us that f(x) has a general shape as given in Figure 3 and must satisfy the following conditions:

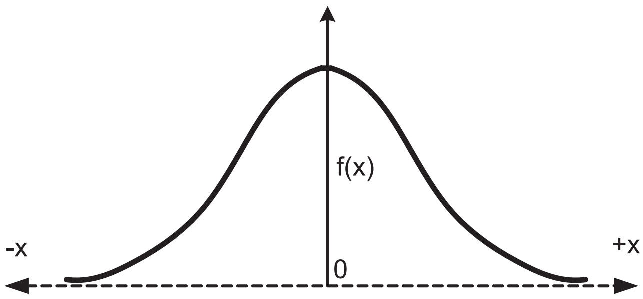

Fig. 3. Frequency distribution function for target hits along the horizontal axis

a) (x) at infinity, along both sides should be zero
b) Even for a novice, the value of f(x) at x = 0, is maximum
c) Because of the condition 'b' the derivative of f(x) at origin is zero, i.e. $f'(0) = 0$
d) The mean value of 'x' lies at the 'Bull's Eye'

The General Form of the Distribution Function

It is easy to see that the function $f(x) = Ae^{-Bx^2}$ satisfies all the above conditions. For example:

$$f(\pm\infty) = 0 \tag{1}$$

$$f(0) = A \tag{2}$$

$$f'(x) = -2ABxe^{-x^2} \quad \Rightarrow \quad f'(0) = 0 \tag{3}$$

We therefore consider f(x) as a true probability distribution function (pdf) and take the following condition, associated with such functions, to be necessarily true:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} Ae^{-Bx^2} dx = \qquad (4)$$

More importantly, due to symmetry conditions (for a normal shooter no bias is expected, for either the left or the right of the origin):

$$\int_{0}^{+\infty} f(x)dx = \frac{1}{2}\int_{-\infty}^{+\infty} Ae^{-Bx^2} dx = \frac{1}{2} \qquad (5)$$

Mean of f(x)

Let us now test f(x) once more and try to find the mean value of 'x' as follows:

$$\text{Mean} = \int_{-\infty}^{+\infty} x.f(x)dx = \int_{-\infty}^{+\infty} xAe^{-Bx^2} dx = -\frac{A}{2B} e^{-Bx^2}\Big|_{-\infty}^{+\infty} = 0 \qquad (6)$$

This is exactly what we expected and our function passes another 'pdf' test.

Determining the value of 'B'

We are now fairly confident that the function developed above is correct and our problem reduces to finding the values of the two arbitrary constant 'A' and 'B'; sinceit is much simpler, we shall tackle the latter one first. For this, after mean, we use the other important idea in statistics which is the 'variance'. By definition

$$\text{Variance} = \sigma^2 = \int_{-\infty}^{+\infty} x^2.f(x)dx = \int_{-\infty}^{+\infty} x^2.Ae^{-Bx^2} dx \qquad (7)$$

The integral can be solved 'by parts' by considering

$$\begin{aligned} u = x \quad &\Rightarrow \quad u' = 1 \\ v' = Axe^{-Bx^2} \quad &\Rightarrow \quad v = \frac{-A}{2B} e^{-Bx^2} \end{aligned} \qquad (8)$$

Thus

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2.Ae^{-Bx^2} dx = -\frac{Ax}{2B} e^{-Bx^2}\Big|_{-\infty}^{+\infty} + \frac{1}{2B}\int_{-\infty}^{+\infty} Ae^{-Bx^2} dx \qquad (9)$$

Since, as discussed above, $\int_{-\infty}^{+\infty} Ae^{-Bx^2} dx = 1$, we have

$$\sigma^2 = \frac{1}{2B} \quad \Rightarrow \quad B = \frac{1}{2\sigma^2} \qquad (10)$$

Determining the value of 'A'

Having done the easier part we shall now determine the value of the other constant 'A'. For this let us lay down our target board horizontally and try to find the volume of the solid 'bell' generated by rotating the f(x) curve, for positive x values [Fig 4], through a full circle. The value of

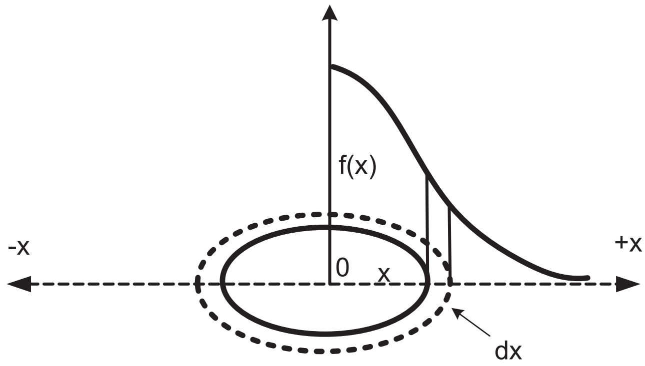

Fig. 4. Thin cylindrica shell generated by on half of the frequency distribution curve

f(x) along a circular path with radius, says 'x' will be constant. Considering a thin strip of thickness 'dx', the volume 'dΦ' of the thin shell, of the cylinder generated by rotating half of the f(x) curve through a full circle, would thus be given by

$$d\Phi = 2\pi x.f(x)dx = 2\pi x.Ae^{-Bx^2} dx \qquad (11)$$

Integration then gives us the value of the volume 'Φ' as

$$\Phi = 2\pi\int_{0}^{\infty} xf(x)dx = 2\pi\int_{-\infty}^{\infty} Axe^{-Bx^2} dx = -\frac{\pi A}{B} e^{-Bx^2}\Big|_{0}^{\infty} = \frac{\pi A}{B} \qquad (12)$$

Although we know the value of 'B', and we have found a relationship between 'A' and 'B', we have generated a third variable 'Φ'. Hence, in order to determine 'A' we have to find an additional relationship containing 'Φ'. We can do this by considering the target board to lie in the x-y plane [Fig. 5]. Now, if in addition to the x-axis, we consider the pdf alvong the y-axis [Fig. 6]. Again due to symmetry considerations, it is safe to assume that f(y) would be given by

$$f(x) = Ae^{-By^2} \qquad (13)$$

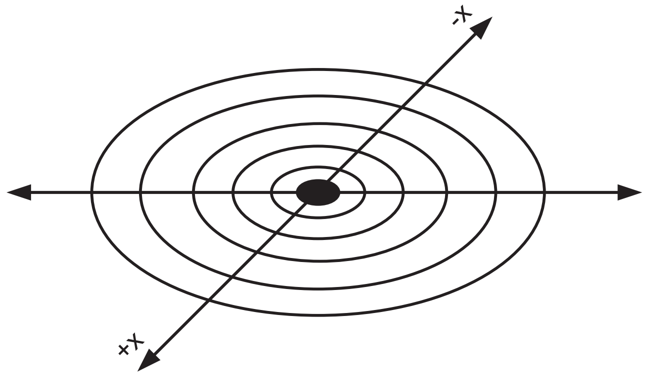

Fig. 5. Target board laid flat with the two axes

Let us consider a rectangle generated by the interception of two strips as shown in Fig. 7, to lie at a distance 'r' from the origin. We can now consider our pdf to be a simultaneous function of x and y or a function of 'r' so that

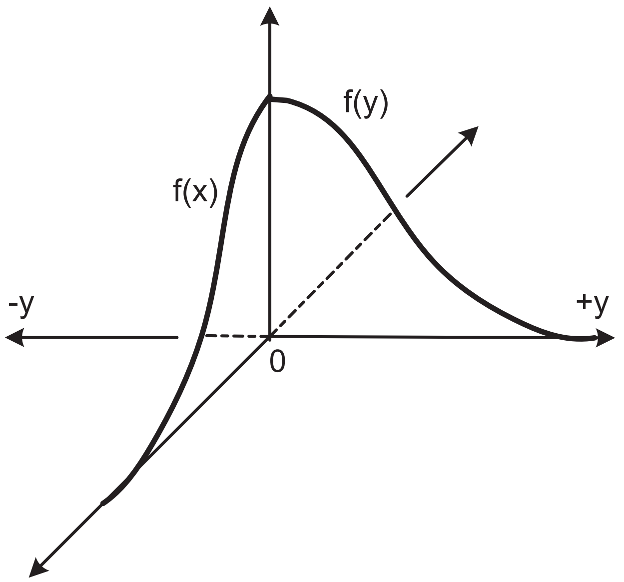

Fig. 6. One fourth of the soled bell generated gy the frequency distribution function of target hits in the positive 'x' and 'y' quadrant

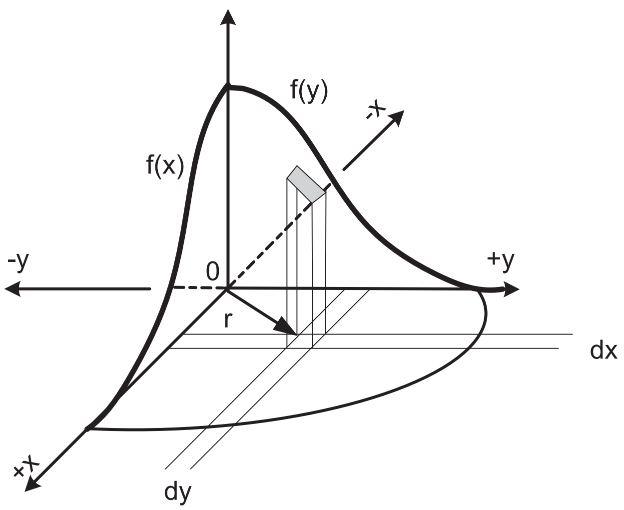

Fig. 7. ???

$$f(x, y) = f(r) = Ae^{-Br^2} \tag{14}$$

Keeping in view that $r = x^2 + y^2$ and considering the area of the rectangle as 'dxdy' the volume of the rectangular prism under the 'bell' is obviously

$$d\Phi = f(x, y)dxdy = Ae^{-Br^2}dxdy = Ae^{-B(x^2+y^2)}dxdy = Ae^{-Bx^2}e^{-By^2}dxdy \tag{15}$$

Volume of the quarter of the bell, lying in the positive x and y plane, is thus given by

$$\frac{\Phi}{4} = \int_0^\infty\int_0^\infty f(x, y)dxdy = \int_0^\infty\int_0^\infty Ae^{-Bx^2}.e^{-By^2}dxdy = \int_0^\infty\left[\int_0^\infty Ae^{-Bx^2}dx\right].e^{-By^2}dy \tag{16}$$

We know that

$$\int_0^\infty Ae^{-Bx^2}dx = \int_0^\infty Ae^{-By^2}dy = \frac{1}{2} \tag{17}$$

Thus

$$\frac{\Phi}{4} = \frac{1}{2}\int_0^\infty e^{-By^2}dy = \frac{1}{2A}\int_0^\infty Ae^{-By^2}dy = \frac{1}{4A} \tag{18}$$

This gives us the other relationship containing 'Φ' as

$$\Phi = \frac{1}{A} \tag{19}$$

Equating it with the first relationship, we get

$$\Phi = \frac{1}{A} = \frac{\pi A}{B} \quad \Rightarrow \quad A^2 = \frac{B}{\pi} \tag{20}$$

But

$$B = \frac{1}{2\sigma^2} \tag{21}$$

Thus

$$A = \frac{1}{\sqrt{2\pi}\,\sigma} \tag{22}$$

We finally have the normal probability density function, with mean zero, and standard deviation 'б' as

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-x^2/2\sigma^2} \tag{23}$$

For a distribution with a nonzero mean the function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{\frac{-(x-\bar{x})^2}{2\sigma^2}} \tag{24}$$

REFERENCES

1. J. Branson, 'Integral of Gaussian', available at: http://quantummechanics.ucsd.edu/ph130a/130_notes/node87.html
2. Unknown, 'Gaussian Integrals', http://www.umich.edu/~chem461/Gaussian%20Integrals.pdf
3. W. D. Cairns, 'A Derivation of the Equation of the Normal Probability Curve', Bull. Amer. *Math. Soc*. 26 (3) (1919), pp 105-108.
4. C. P. Nicholas and R. C. Yates, 'The Probability Integral', this MONTHLY 57 (1950) 412-413
5. D. Teague, 'The Normal Distribution: A derivation from basic principles, available at: moodle.cs.huji.ac.il/mod/resource/view.php?id=3313