THE PROBABILITY DENSITY CURVE

The molecules of the atmosphere have many parameters. Some of these parameters—such as velocity—have a very wide distribution of values. With some parameters—again, such as velocity—this distribution of values tends to cluster about the mean value. In these particular cases, this clustering suggests that the farther that a particular value departs from the mean value, the less likely it is to occur.

The Probability Density Function

These distributions of values can be expressed mathematically by a probability density function (pdf) of the form

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \frac{(x-\bar{x})^2}{2\sigma^2} \qquad \text{PDC01}$$

Here, $\phi(x)$ is the probability that the parameter will have the value x, $\bar{x}$ is the mean value of the parameter for the population, and σ is the standard deviation of the distribution. Note that the function is defined by the two variables: the mean ($\bar{x}$) and the standard deviation (σ).

The Mean Value – We can further define the mean value $\bar{x}$ as the sum of all the values divided by the number of values; i. e.,

$$\bar{x} = \tfrac{1}{n}\sum_{1}^{n} x \qquad \text{PDC02}$$

Here, $\boldsymbol{n}$ is the total number of individual values of $\boldsymbol{x}$.

The Standard Deviation – The standard deviation $\boldsymbol{\sigma}$ is defined as the square-root of the mean of the squares of the individual deviations from the mean.

$$\boldsymbol{\sigma} = \left(\frac{1}{n} \sum_{1}^{n} \left(x - \bar{x} \right)^{2} \right)^{\frac{1}{2}} \qquad \text{PDC03}$$

The standard deviation always has a positive value, since negative values of $\boldsymbol{x - \bar{x}}$ are removed in the squaring process.

History of the Probability Density Curve – The curve shown in PDC01 is also known as the bell-shaped curve, the de Moivre-Laplace curve, the Gaussian error curve, and other names when used in particular applications.

The concept was discovered by Abraham de Moivre in 1733 and used in his ***Doctrine of Chances***. The concept was enlarged and extended by Pierre-Simon Laplace in 1812 in his ***Analytical Theory of Probabilities***. At roughly the same time, Karl-Friedrich Gauss claimed to have been using the concept since 1794 and published a dissertation on the normal error distribution in astronomical observations in 1809. Properly speaking, the curve should only be referred to as Gaussian only when it refers to errors in measurements.

De Moivre should retain honors when the curve describes probability density, but this is unlikely to happen. The Gaussian designation is too firmly entrenched in the literature.

The Normal Distribution Curve – When the distribution mean is set to zero ($\bar{x} = 0$) and the standard deviation is set to one ($\sigma = 1$), the probability density curve is said to be "normalized". It becomes the normal distribution curve

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \text{PDC04}$$

It may be written in many forms. In the form shown above, the constant forces the area under the curve to be equal to unity, and x is expressed in multiples of the standard deviation of the distribution. Note that this equation is only valid when the mean is equal to zero. Note also that the equation is independent of the value of the standard deviation, as long as x is always expressed in terms of that standard deviation.

<u>Velocity Distribution Along a Single Cartesian Axis</u> – In kinetic gas theory, the curve identified as PDC01 describes the probability distribution of molecular velocities along any single axis (including our proximity axis) of the standard tri-axial Cartesian reference system for an ideal gas under conditions of equilibrium. For the x-axis, it takes the form

$$\phi\left(v_{\pm x}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \frac{(v_x - \bar{v}_x)^2}{2\sigma^2} \qquad \text{PDC05}$$

Here, $\phi\left(v_{\pm x}\right)$ is the probability that a molecule chosen at random from the total population will have the velocity v_x. The velocity may be in either the positive direction (+x) or in the negative direction (-x).

This expression is valid for any single axis, no matter what the orientation or location of the axis might be within the population—as long as conditions of equilibrium apply. Obviously, then, it is also valid for any combination of such axes, such as make up our proximity axis (v_p) in one-dimensional space.

<u>Velocity Distribution Along the Proximity Axis</u> – When applied to the positive arm of the proximity axis in one-dimensional space, the expression takes the form

$$\phi\left(v_p\right) = \frac{2}{\sigma}\frac{1}{\sqrt{2\pi}}\exp - \frac{v_p^2}{2\sigma^2} \qquad \text{MSV06}$$

The factor of 2 in the constant is necessary because the total probability must remain at unity and we are here only dealing with half the total population (the ***p*** sub-population but not the ***–p*** sub-population). Moreover, the numerator in the exponential fraction has been reduced from $(v_p - \bar{v}_p)^2$ to simply v_p^2 because $\bar{v}_p$ is always zero (0) under conditions of equilibrium.

These equations will tell us the probability that an individual molecule will have some particular speed. However, what if we want to know what proportion of the population has speeds between two specific values? For that, we turn to other members of the family of probability density curves: the error function curve and the complementary error function curve.

The Error Function Curve

In *Statistical Physics*, Brown (p. 298) gives the error function of x as:

$$erf\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}dx \qquad \text{PDC06}$$

The function returns the proportion of the area under the normal curve (the total area being unity) that lies between the value zero and the value x expressed in standard deviations.

Proximity Velocity Error Function – For the range of proximity velocities, Equation PDC06 must be modified to

$$erf\,(v_p) = \frac{2}{\sqrt{\pi}}\int_0^{v_p} e^{-v_p^2}dv_p \qquad \text{PDC07}$$

This expression give the proportion of molecules having proximity velocities lying between zero and v_p. All velocities are expressed in multiples of the standard deviation **σ**.

The Complementary Error Function

The complimentary error function of value ***x***, ***erfc(x),*** is usually given by Brown by the normalized expression:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-x^2} dx \qquad \text{PDC08}$$

The function returns the proportion of the area under the normal curve (the total area being unity) that lies between the value ***x*** and infinity expressed in standard deviations.

Function Relationship – The error function is simply related to the complementary error function by the expression

$$erfc(x) = 1 - erf(x) \qquad \text{PDC09}$$

Proximity Velocity Complementary Error Function – For the range of proximity velocities, Equation PDC08 must be modified to

$$erfc(v_p) = \frac{2}{\sqrt{\pi}} \int_{v_p}^\infty e^{-v_p^2} dv_p \qquad \text{PDC10}$$

This expression gives the proportion of molecules having proximity velocities lying between v_p and infinity. All velocities are expressed in multiples of the standard deviation **σ**.

Selected Values – A selection of values for the error function and the complementary error function are given in Table PDC01. All velocity values are in terms of multiples of the standard deviation.

Table PDC01

v_p	=	$erf(v_p)$	$erfc(v_p)$
$-\infty\ \sigma$		0.0000	1.0000
$-4.0\ \sigma$	$\approx$	0.0000	1.0000
$-3.0\ \sigma$		0.0013	0.9987
$-2.0\ \sigma$		0.0268	0.9732
$-\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\sigma$	$-\bar{v}_i$	0.1050	0.8950
$-(\gamma)^{\frac{1}{2}}\sigma$	$-\bar{v}_s$	0.1180	0.8820
$-1.0\ \sigma$	$-\sigma$	0.1587	0.8413
$-\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\sigma$	$-\bar{v}_p$	0.2128	0.7872
$0.000\ \sigma$		0.5000	0.5000
$\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\sigma$	$\bar{v}_p$	0.7872	0.2128
$1.0\ \sigma$	σ	0.8413	0.1587
$(\gamma)^{\frac{1}{2}}\sigma$	$\bar{v}_s$	0.8820	0.1180
$\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\sigma$	$\bar{v}_i$	0.8950	0.1050
$2.0\ \sigma$		0.9732	0.0268
$3.0\ \sigma$		0.9987	0.0013
$4.0\ \sigma$	$\approx$	1.0000	0.0000
$+\infty\ \sigma$		1.0000	0.0000

Evaluating the Error Function and the Complementary Error Function – These expressions cannot be evaluated explicitly, but can be solved numerically. For those of my readers who wish to construct algorithms for the expressions, Sergei Winitzki proposed (and David Cantrell modified) the following approximation. The approximation is said to be correct to about five significant figures over the entire range of the expression. I offer it here without endorsement, as I have not evaluated it personally.

$$erf(x) \approx \left[1 - \exp\left(-x^2 \frac{\frac{4}{\pi} + ax^2}{1 + ax^2}\right)\right]^{\frac{1}{2}}$$

where the constant a is 0.147 (Cantrell's improvement).

* * * * *

REFERENCES

De Moivre's *Doctrine of Chance*, 1718.
http://www.mathpages.com/home/kmath642/kmath642.htm

Winitzki's Error Function Approximation, 2008.
http://homepages.physik.uni-muenchen.de/~Winitzki/erf-approx.pdf

Samuel Goldberg; **Probability – An Introduction**; Prentice Hall, Englewood, New Jersey; 1960.

Helen M. Walker and Joseph Lev; **Statistical Inference**; Holt, Rinehart, and Winston; New York, NY; 1953.

Murray R. Spiegel; **Mathematical Handbook of Formulas and Tables**, McGraw Hill Book Company, New York, 1968.

Arthur Brown; **Statistical Physics**; Elsevier, New York, 1970.

James Jeans; **An Introduction to the Kinetic Theory of Gases**; Cambridge Library Collection, 1940.

Wolfgang Pauli; **Statistical Mechanics**; Dover Press, Mineola, 1973.