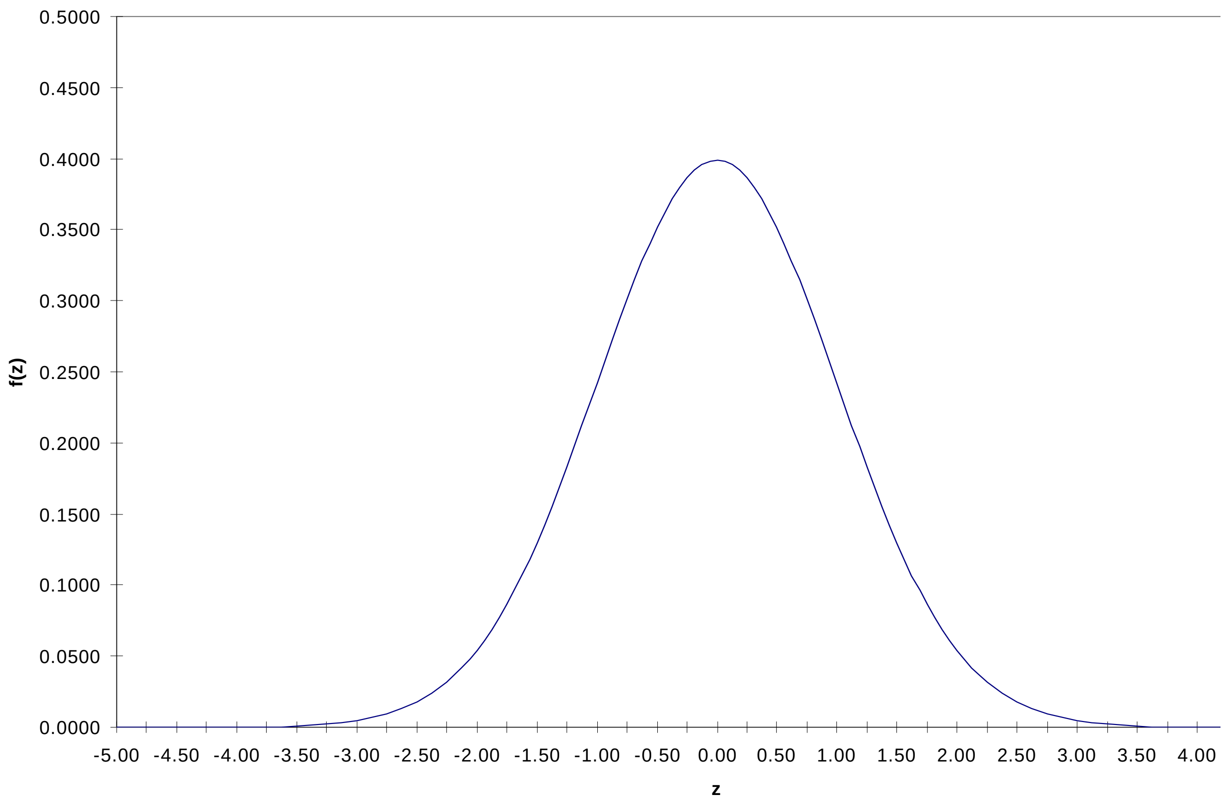
TEACHING NOTE 97-01:
THE NORMAL PROBABILITY DISTRIBUTION

Version date: July 18, 2008 C:\Classes\Teaching Notes\TN97-01.doc

The normal probability distribution, known commonly to the layman as the bell-shaped curve, was first identified in the 17th century by Abraham de Moivre (1667-1754). Its mathematical structure was developed by Carl Frederich Gauss (1777-1855) and the curve is often referred to as the Gaussian distribution. A modified version of it is



depicted below.

The mathematical function that plots the normal curve, called the density function, is

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

where x is the value of the random variable, μ is its expected value and σ is its standard deviation. Any normally distributed random variable can be expressed as a standard normal random variable by subtracting its expected value and dividing by its standard deviation. This standardized normal variable is often referred to with the letter z and its density is written as

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

The standard normal variable has expected value of zero, a variance of 1 and is symmetric. Symmetry gives it a skewness factor of zero. Other distributions that are more (less) peaked than the normal are said to be leptokurtic (platykurtic). The only information required by the normal distribution is the expected value and variance. It is the standard normal that is plotted above.

The density function gives only the height of the curve. The actual probability of a random variable lying within a particular finite or infinite range is provided by the distribution function, which is the cumulative density. Its mathematical specification for a standard normal variable is

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx .$$

This is interpreted as the probability that a standard normal variable will be less than z. It is sometimes written as Prob($z \le Z$) or Pr($z \le Z$) or simply N(z). The probability is the area under the curve. The total area under the curve, i.e. the integral from $-\infty$ to $+\infty$, is 1.0. Thus $N(\infty) = 1$ and $N(-\infty) = 0$. The probability that a standard normal variable will lie between a and b is given as N(a)-N(b). As with any continuous variable, it is impossible to speak of the probability of obtaining a specific value such as the probability that $z = a$. The probability of observing any one value is zero. Only finite or infinite ranges can be determined. Thus, the expressions Prob($z \le a$) and Prob($z < a$) are equivalent.

Calculation of the probability for a particular range involves the evaluation of the above integral. It is well known that the distribution function for the normal probability cannot be integrated by standard mathematical means. Instead numerical techniques must be used. Tables for the function are widely available and found in nearly every statistics book. Fortunately, there are several other excellent and simple means of computing the normal probability.

It is a general rule that any well-behaved mathematical function with defined derivatives up to a given order can be approximated by a polynomial function. There are many such approximations of the normal probability, most of which are provided in Abramowitz and Stegun (1972). Probably the most widely used is the following.

$$N(z) = 1 - f'(z)(0.31938153k - 0.356563782k^2 + 1.781477937k^3 - 1.821255978k^4 + 1.330274429k^5)$$

where k = 1/(1 + 0.2316419z) and $f'(z) = \left(1/\sqrt{2\pi}\right)\exp\left(-z^2/2\right)$. If z < 0, the fact that the curve is symmetric enables us to obtain N(z) as 1 - N(-z). This function is known to be accurate to at least four digits. The table shown after the references was created using it.

For example, suppose we wish to find the probability of observing a value less than 1.35. This is found by looking up 1.3 in the left column and moving over to the 0.05 column. We see that N(1.35) = .9115. Suppose we wish to find N(-0.52). Since z is negative, we simply find N(0.52), which is .6985 and subtract this value from 1.0, giving us 0.3015. Note that N(0.0) = .5, meaning that half of the area under the curve is to the left of zero. It is well known that approximately two-thirds of the time, a normally distributed random variable will lie within one standard deviation of the expected value. This is verified by noting that N(1) - N(-1) = N(1) - (1 - N(1)) = 2N(1) - 1 = 2*0.8413 - 1 = 0.6826. Similarly we would find that approximately 95 percent of the time, the random variable will be within two standard deviations of the expected value and 99 percent of the time, it will be within three standard deviations.

If you are working in Excel, the normal probability can be obtained with the function =normsdist(.) where the value of z is inserted in the parentheses either directly or through a cell reference.

A useful result in involving the normal distribution is the so-called *Central Limit Theorem*. It states that the distribution of the sum or mean of a sample of n independent random variables will converge to the normal distribution as $n \to \infty$. A general rule of thumb is that n must be at least 30. Note that this refers to the distribution of the sample mean or sum. The distribution of each observation in a sample does not converge to the normal distribution, except for certain distributions such as the binomial and the poisson.

The moment generating function of any probability distribution is $M(t) = E(e^{tx})$. For the normal distribution, the moment generating function

$$M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

Successive differentiation of the moment generating function and setting t = 0 provides the moments, E(x), $E(x)^2$, $E(x^3)$, etc. Note then that dM(t)/dt = M(t)(μ. + $(1/2)\sigma^2 t$). Evaluated at t = 0, we have dM(t)/dt|t = 0 = μ. The second derivative, $d^2M(t)/dt^2$ = $M(t)(1/2)2\sigma^2$. Evaluated at t = 0, we have $d^2M(0)/dt^2|t=0 = \sigma^2$. All odd number moments of the standard normal distribution are zero and the even numbered moments are simplified to $\sigma^m m!/(2^{m/2}(m/2)!)$, i.e., m = 2,4,6, etc.

References

Any good book on statistical theory will give extensive coverage of the normal probability distribution.

The polynomial approximations are in

Abramowitz, M. and I. Stegun. *Handbook of Mathematical Functions* (New York: Cover Publications, 1972), Ch. 26.

Normal Probability Table

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990