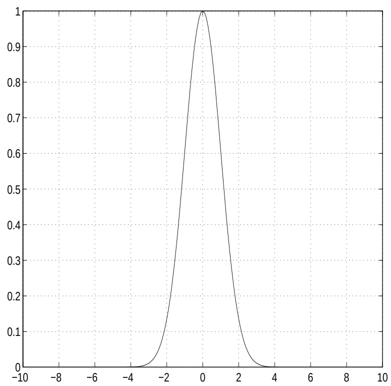
The Normal Distribution: $N(\mu, \sigma^2)$

Introduction The continuous random variable which finds the widest use in applications and theory is the Gaussian or Normal distribution. The probability density function of this random variable is expressed as

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

To begin understanding this distribution consider the simpler function

$$f(x) = e^{-x^2/2}$$



We can determine whether or not this is a valid probability density function by integrating from $-\infty$ to ∞. The standard trick here is to let

$$I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

and consider the quantity I^2.

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-s^2/2} ds$$

We can rewrite the above as

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2/2} e^{-s^2/2} dx ds = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+s^2)/2} dx ds$$

Rewrite the integral in polar coordinates

$$s = rcos(\theta), \qquad x = rsin(\theta)$$

so that the area element $dxds$ becomes $rdrd\theta$ and the limits of integration become $0 \le r < \infty$ and $0 \le \theta \le 2\pi$.

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta$$

This is easily evaluated as

$$I^2 =$$

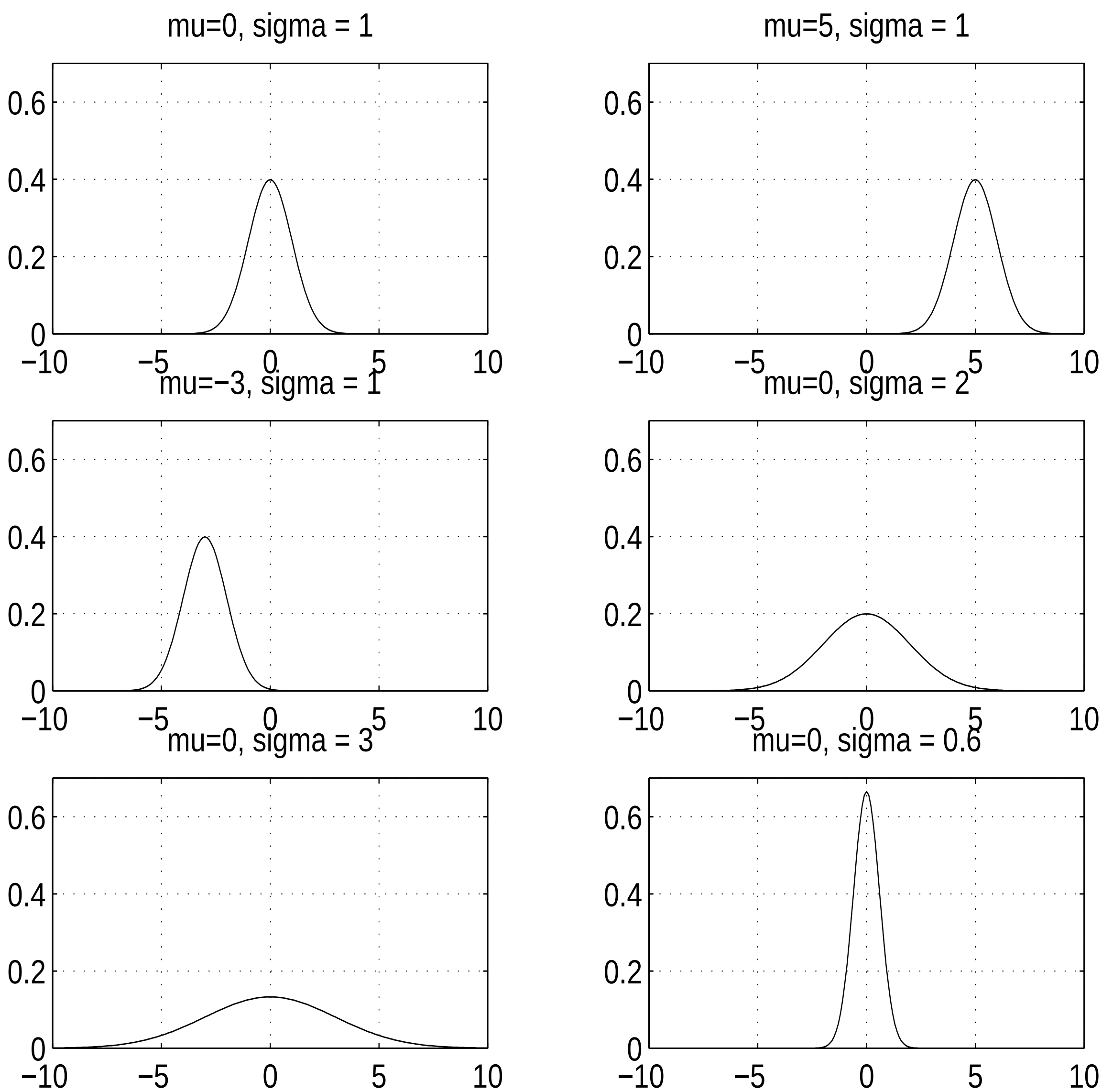
and so

$$I =$$

Using the above result we can define the normal distribution with parameters $\mu = 0$ and $\sigma = 1$ as

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

We can see how the parameters μ and σ alter the appearance of the normal distribution with the following graphs.



Calculating the Mean of a Normal Distribution Recall the definition of the mean of a continuous random variable:

$$E(X) \equiv \int_{-\infty}^{\infty} x f(x) dx$$

Applying this to the normal distribution gives

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

Evaluate this integral with the substitution $s = (x - \mu)/\sigma$. Upon substitution obtain

Break this integral into two parts.

Calculating the Variance of a Normal Distribution Recall the definition of the variance of a continuous random variable:

$$V(X) \equiv E((X-\mu)^2)$$

and that it is often easier to compute

$$V(X) = E(X^2) - E(X)^2$$

$E(X^2)$ may be expressed as

$$E(X) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

Use the same substitution as above $s = (x-\mu)/\sigma$.

Break the integral up into three parts:

The Standard Normal Distribution, $N(0,1)$. It is easy to show that if X is normally distributed with mean μ and standard deviation $= \sigma$, then $Y = aX + b$ is also normally distributed. What is the mean and what is the standard deviation of Y?

We can use this result to our advantage. In the transformation above let $a = 1/\sigma$ and $b = -\mu/\sigma$. Then the distribution of Y is given by

The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *standard normal distribution.* Its values are tabulated and may be readily found.

Consider the following examples from our text.

- The time it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear end collisions. The article ”Fast Rise Break Lamp as a Collision Prevention Device” (*Ergonomics*, 1993:391-395) suggests that reaction time for an in traffic response to a brake signal from standard brake lights can be modelled with a normal distribution having a mean value 1.25 seconds and standard deviation of 0.46 seconds. What is the probability that the reaction time is between 1.00 seconds and 1.75 seconds?

- The breakdown of a randomly chosen diode of a particular type is known to be normally distributed. What is the probability that the diode's breakdown voltage is within k standard deviations of the mean for $k = 1, 2, 3$? Compare with Chebyshev.

The following examples are taken from *Introductory Probability and Statistical Applications* by Paul L. Meyer and from *Probability: An Introduction with Statistical Applications* by John J. Kinney.

1. Suppose X is a continuous random variable with distribution $N(3, 4)$. Find a number c such that

   $$P(X > c) = 2P(X \leq c)$$

2. Suppose that the breaking strength of a cotton fabric (in pounds), say X, has a distribution $N(165, 3)$. A sample of this fabric is considered to be defective if $X < 162$. What is the probability that a sample chosen at random from a large lot will be defective?

3. A bag of cement labelled as 40 pounds actually comes from a distribution which is normal with mean 39.1 pounds and standard deviation $\sqrt{9.4}$ pounds.

   (a) Find the probability that 2 out of 5 randomly selected bags weigh less than 40 pounds.

   (b) Calculate the probability that a sample of $n = 5$ bags will have a sample mean less than 40 pounds.

4. Capacitors from a manufacturer are normally distributed with mean $5\mu f$ and standard deviation $0.4\mu f$. An application requires 4 capacitors between $4.3\mu f$ and $5.9\mu f$. If the manufacturer ships 5 randomly selected capacitors, what is the probability that a sufficient number of capacitors will be within specifications?

One of the more important continuous distributions may be obtained from the normal distribution as follows. Let $Z \sim N(0,1)$, i.e. let Z be a standard normal distribution. We will calculate the distribution of Z^2 and recognize the result as one of our familiar random variables. At this point we just note in passing that it is quite common to obtain "sums of squares" of the sort we are developing in regression and in analysis of variance.

We would like to use one of our previous results: Let X be a continuous random variable with support on (a,b), and let $Y = u(X)$ for a strictly monotone, differentiable (hence invertible) function u. Then the pdf of Y is given as

$$g_Y(y) = f_X(u^{-1}(y))|\frac{dx}{dy}|$$

Since Z^2 is not invertible on the support of Z (why not?) we first let $X = |Z|$. Then, as we have seen, the pdf of X is

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

with support on the nonnegative reals. Now consider $Y = X^2$. Since X may not assume values on the negative reals, $u(X) = X^2$ is now invertible (and differentiable) and so we may use the theorem. We immediately have, since $x = u^{-1}(y) = \sqrt{y}$ and so $dx/dy = (1/2)y^{-1/2}$

$$g_Y(y) = f_X(u^{-1}(x))|\frac{dx}{dy}| = f_X(y^{\frac{1}{2}})|\frac{1}{2}y^{-\frac{1}{2}}|$$

or

$$g_Y(y) = \frac{2}{\sqrt{2\pi}}e^{-\frac{1}{2}(y^{\frac{1}{2}})^2}|\frac{1}{2}y^{-\frac{1}{2}}|$$

so finally

$$g_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}y^{-\frac{1}{2}}$$

What sort of a distribution is this? Since it occurs so often in statistics and probability this particular form of the (   ) distribution is said to be a χ^2 distribution with $\nu = 1$ degree of freedom. We'll see other forms of this distribution (with more degrees of freedom) later.