Math 217 Probability and Statistics

Prof. D. Joyce, Clark University

Wednesday, 24 Oct 2007

Quiz Today.

Due Friday. From Chapter 5.1: 7, 9, 16, 18, 25.

Last time. Functions of random variables. The Cauchy distribution.

Today. Survey of distributions: Gamma, Beta, and the Gamma and Beta functions (see the notes on distributions). The normal distribution.

The normal distribution. The central limit theorem says that the sum or average of a large number of independent random variables is approximately a normal distribution. We'll look at the exact statement of it later in the course. Now, we'll just look at what the normal distribution is.

There are various normal distributions. One of them is the *standard normal distribution.* A random variable with a standard normal distribution is often denoted Z. It's density function is

$$f_Z(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

It's graph is a bell shaped curve, but not nearly as flat as the Cauchy distribution we looked at last time. We haven't yet defined the mean μ and the standard deviation σ of a random variable, but we will in the next chapter. (Actually, the mean is introduced in one of the exercises in this chapter.) The mean will be what you know as the mean, or arithmetic average, and the standard deviation will be an indication of the spread of the distribution. The mean of the standard normal distribution is 0, and its standard deviation is 1.

If we translate a standard normal random variable, the mean will change by the translation. That means that if we add a constant μ to a standard normal random variable to get a random variable $X = Z + \mu$, then the mean of X will be μ. But translations don't change the standard deviation, so the standard deviation of X will still be 1.

On the other hand if we scale a standard normal random variable by multiplying it by a constant, then its mean won't change, but its standard deviation will be multiplied by that same constant. That means that if multiply a standard normal random variable by a constant σ to get a random variable $X = \sigma Z$, then the mean of X will still be 0, but its standard deviation will be σ.

If we both scale and translate to get $X = \sigma Z + \mu$, then the mean of X will be μ and its standard deviation will be σ. Such an X is called a *normal* random variable, and it's distribution is called a *normal distribution*, or a *Gaussian distribution.* We can determine the density f_X for X from the density f_Z of a standard normal distribution defined above by means of the formula we developed last time

$$f_X(x) = f_Z(\phi^{-1}(z))\, \frac{d}{dx}\phi^{-1}(x).$$

Now $X = \phi(Z)$, so $\phi(z) = \sigma z + \mu$. Therefore, $\phi^{-1}(x) = (x-\mu)/\sigma$, and so $\frac{d}{dx}\phi^{-1}(x) = 1/\sigma$. Thus, the normal distribution with mean μ and standard deviation σ has the probability density function

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)} \\ &= \frac{1}{\sigma\sqrt{2\pi}}\, \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

(Note that $\exp x$ is another notation used for e^x when the exponent is complicated.)

The cumulative distribution function $F_X(x)$ for a normal distribution can be found by integrating its probability density function $f_X(x)$. Unfortunately, this is not an elementary integral, and that means it can not be found in terms of the standard elementary functions which include all the algebraic functions, trig functions, inverse trig functions, logs, and exponential functions. In fact, a new function has to be named to integrate it called the *error function*, $\operatorname{erf}(x)$:

$$\operatorname{erf}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt$$

which is just the cumulative distribution function $F_Z(x)$ for the standard normal distribution. There are tables for this function in text books, but calculators that can do statistics can calculate the function directly.