## Lecture 8: More Continuous Random Variables

26 September 2005

Last time: the exponential. Going from saying the density  $\propto e^{-\lambda x}$ , to  $f(x) = \lambda e^{-\lambda x}$ , to the CDF  $F(x) = 1 - e^{-\lambda x}$ . Pictures of the pdf and CDF.

Today: the Gaussian or normal distribution. (I prefer Gaussian.)

Starting point: Make the density proportional to  $e^{-z^2/2}$ . This falls off much faster than the ordinary exponential. Also, we can have it work on both sides of the origin. (Why 1/2? It turns out to make life nicer for us later.) (Why z? Tradition.)

To get the density, we need the normalizing factor, which is  $\int_{-\infty}^{\infty} e^{-z^2/2} dz$ . This is  $\sqrt{2\pi}$ . There are several proofs, but none are really elementary. (There's a nice one involving contour integrals in the complex plane, for instance.) Still, trust me on this one.

We now have a probability density,

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

which looks like this: This is the **standard Gaussian** or **standard normal** density. (I like "Gaussian" better.) It's the famous **bell curve**.

The CDF of the standard Gaussian is an important function, so much so it has its own notation,  $\Phi(z)$  (instead of the usual F(z)). It's also called the **error function**. Unfortunately, it doesn't have any nice analytical form, but there are well-known numerical approximations which are very accurate, at least as accurate at the numerical approximations for sine, cosine, logarithm, etc. From the symmetry of the pdf, though, we know that  $\Phi(0) = 0.5$ . Also, from symmetry, for z > 0,

$$\Pr (Z \ge z) = \Pr (Z \le -z)$$
  
$$1 - \Phi(z) = \Phi(-z)$$

The CDF of the standard Gaussian is an important function, so much so it has its own notation,  $\Phi(z)$  (instead of the usual F(z)). It's also called the **error function**. Unfortunately, it doesn't have any nice analytical form, but there are well-known numerical approximations which are very accurate, at least as accurate at the numerical approximations for sine, cosine, logarithm, etc.



Figure 1: Probability density of the standard Gaussian



Figure 2: CDF  $\Phi$  of the standard Gaussian

At this point we can already find the mean, using the fact that  $e^{-z^2/2}$  is symmetric around the origin.

$$\mathbf{E}[Z] = \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= \int_{-\infty}^{0} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz + \int_{0}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2} dz$$

$$= \int_{0}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz - \int_{0}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= 0$$

As for the variance,  $\operatorname{Var}(Z) = \mathbf{E}[Z^2] - (\mathbf{E}[Z])^2 = \mathbf{E}[Z^2]$ . So let's see that.

$$\mathbf{E}\left[Z^2\right] \quad = \quad \int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}} e^{-z^2/2} dz$$

This doesn't cancel out, because  $z^2 = (-z)^2$ . In fact, one can show that this integral works out to be exactly 1. So the standard Gaussian has mean zero and variance 1.

What about non-standard Gaussians? Well, consider the density of  $X = Z + \mu$ . If  $a \leq X \leq b$ , then  $a - \mu \leq Z \leq b - \mu$ .

$$\Pr(a - \mu \le Z \le b - \mu) = \int_{z=a-\mu}^{z=b-\mu} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$
$$= \int_{x=a}^{x=b} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx$$
$$= \int_{x=a}^{x=b} f_X(x) dx$$
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

Now consider  $X = \sigma Z$ ,  $\sigma > 0$ .

$$\Pr(a \le X \le B) = \Pr\left(\frac{a}{\sigma} \le Z \le \frac{b}{\sigma}\right)$$
$$= \int_{z=\frac{a}{\sigma}}^{z=\frac{b}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$
$$= \int_{x=a}^{x=b} \frac{1}{\sqrt{2\pi}} e^{-z^2/2\sigma^2} \frac{dx}{\sigma}$$
$$= \int_{x=a}^{x=b} f_X(x) dx$$
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Putting these together,  $X = \sigma Z + \mu$  has the density

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

From the rules for expectation,  $\mathbf{E}[X] = \sigma \mathbf{E}[Z] + \mu = \mu$ . From the rules for variance,  $\operatorname{Var}(X) = \sigma^2 \operatorname{Var}(Z) = \sigma^2$ . We say that the equation above gives the distribution of a Gaussian with mean  $\mu$  and variance  $\sigma^2$ ,  $\mathcal{N}(\mu, \sigma^2)$ . An arbitrary Gaussian can be transformed into standard form by  $Z = (X - \mu)/\sigma$ . By standardizing a Gaussian random variable, we keep ourselves from having to think about any distribution function other than  $\Phi$ . Two numbers  $\mu$  and  $\sigma$  completely determine the distribution, and completely specify the transformation back to standard form; they're the parameters.

The sum of two Gaussian variables is always another Gaussian, whether or not they are independent. A Gaussian times a constant is Gaussian. A Gaussian plus a constant is Gaussian.

Here are some important facts:

$$\begin{aligned} &\Pr\left(|Z| \le 1\right) = \Phi(1) - \Phi(-1) &= 0.6826895\\ &\Pr\left(|Z| \le 2\right) = \Phi(2) - \Phi(-2) &= 0.9544997\\ &\Pr\left(|Z| \le 3\right) = \Phi(3) - \Phi(-3) &= 0.9973002 \end{aligned}$$

By applying the standardization above, we can see that a Gaussian falls within one standard deviation of its mean about two-thirds of the time, within two standard deviations about ninety-five percent of the time, and within three standard deviations about 99.7% percent of the time.

Remember the Chebyshev inequality from before:

$$\Pr\left(|X - \mu| \ge k\sigma\right) \le \frac{\operatorname{Var}\left(X\right)}{\sigma^2}$$

If X is Gaussian, then

$$\Pr(|X - \mu| \ge k\sigma) = 1 - (\Phi(k) - \Phi(-k))$$

Let's compare the probability of large deviations from the mean given by Chebyshev to that obtained by the Gaussian. Clearly, the Gaussian distribution is much more tightly concentrated around its mean than arbitrary distributions. This will be very useful later.

We will see starting next week that the sum of independent random variables tends towards a Gaussian; this is why it is so important and so ubiquitous.

In particular, let's look at a binomial distribution with contant p and growing n. I claim that X Bin(n, p) approaches in distribution  $\mathcal{N}(np, np(1-p))$ . Let's look at some plots, keeping p = 0.5 for simplicity.

There are a number of other distributions which are closely related to the Gaussian.



Figure 3: Comparison of the probability of large deviations for the Gaussian distribution (solid line) against the Chebyshev bounds, which apply to all distributions (dashed line). The lower figure shows the probability on a log scale.



Figure 4: Binomial distribution (solid line) versus a Gaussian density with the same mean and variance (dotted line), with p = 0.5 and n = 4, 10 and 30, respectively.

 $\chi^2$  distribution: if Z is a standard Gaussian, then the distribution of  $Z^2$  is said to be " $\chi^2$  with one degree of freedom":

$$f_{\chi^{2},1}(x) = \frac{x^{-1/2}e^{-x/2}}{\sqrt{2\pi}}$$
  
**E**[X] = 1  
Var(X) = 2

This comes up a lot in statistical testing, especially in the following variant form. If  $Z_1, Z_2 \dots Z_n$  are all independent standard normals, then  $\sum_{i=1}^n Z_i^2$  has a  $\chi^2$  distribution with *n* degrees of freedom:

$$f_{\chi^2,n} = \frac{x^{n/2-1}e^{-x/2}}{\Gamma(n/2)2^{n/2}}$$
$$\mathbf{E}[X] = n$$
$$\operatorname{Var}(X) = 2n$$

where  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ , the **gamma function**. If x is an integer, then  $\Gamma(x) = (x-1)!$ . For any x,  $\Gamma(x) = (x-1)\Gamma(x-1)$ . Finally,  $\Gamma(1/2) = \sqrt{\pi}$ . The  $\chi^2$  distribution is a special case of a broader family of distributions

called the  $\gamma$  distributions, which have the pdf

$$f_{\gamma(\alpha,\theta)}(x) = \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^{\alpha}}$$

They're defined over non-negative x. If  $\alpha > 1$ , then for small x this looks like  $x^{\alpha-1}$ , and rises, but for large x it always looks like an exponential. In fact, the exponential is the space case where  $\alpha = 1$ . For more, see the book.

If  $Y = \log X$  is Gaussian, we say that X is **log-Gaussian** or **lognormal**. Then  $X \ge 0$ , and the distribution is specified by the mean and variance of Y, i.e. of  $\log X$ , call them M and  $S^2$ . In explicit form,

$$f_{lognormal}(x) = \frac{e^{-\frac{(\log x - M)^2}{2S^2}}}{x\sqrt{2\pi S^2}}$$
  

$$\mathbf{E}[X] = e^{M + \frac{S^2}{2}}$$
  

$$\operatorname{Var}(X) = e^{2M + S^2} \left(e^{S^2} - 1\right)$$

The lognormal is the theoretical distribution that best fits the observed distribution of income and wealth. The mean and variance vary from country to country, but the general form doesn't. One consequence of this is that rich people are really *a lot* richer than ordinary people, and that the mean isn't so representative as the median. The median income is after all  $e^M$ , but the mean is much higher! For instance, median household income in the United States is \$42,228 (as of 2001), and the standard deviation of the log income is about



Figure 5: PDF and CDF of a gamma distribution with  $\alpha=3$  and  $\theta=1.$ 

0.879. This gives a mean income of about \$62 thousand, which is very close to the true value of \$58 thousand.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>The main problem with the log-normal model of income distribution is that there are slightly more really poor people than it allows for, and the rich people are much, much richer.



Figure 6: Density and CDF of a lognormal distribution with M = 42,228 and S = 0.879, based on US income data.