Math Tools: Random Variables
(Started: July 19, 2011; Revised: January 18, 2013)

Our first set of math tools concerns randomness: how we think about it, how we quantify it, and so on. It's an essential concept in macroeconomics and finance. You can find more formal versions elsewhere. The idea here is to put the tools to work, and not worry too much about mathematical niceties. At this point you should remind yourself that tools make life easier.

The tools, of course, have uses. *Random variables* (a term we'll define shortly) are described completely by their distribution, the probabilities of the possible outcomes. *Moments* (another term we'll see again) are summary measures that tell us something about the distribution. We focus on two general ideas. The mean and variance reflect the location and scale of the distribution. They depend, for example, on the units of the random variable. Skewness and kurtosis reflect the shape of the distribution and are independent of location and scale. They are particularly helpful in identifying deviations from the normal distribution.

1   Random variables and probability distributions

Random variables are used to describe things that are random, typically in the sense that we don't know their outcomes ahead of time. You might think of the weather (rain or shine?), the economy (boom, bust, or somewhere in between?), the stock market (ditto), or sports (will the Steelers win?).

To make this precise, it's helpful to build the definition from components. This will start off abstract, but turn concrete quickly — and stay that way. The components are:

- *States.* Let's start with what we'll call a *state*, one of the possible outcomes of a random process. You may see the terms *outcome* or *event* used instead, but state serves our purposes. We represent a state with the letter z and the set of all possible states — the *state space* — by $\mathcal{Z}$. Sometimes state $z = 1$ occurs, sometimes state $z = 2$, and so on. Part of the art of applied work is to come up with a useful practical definition of $\mathcal{Z}$. If we're talking about the weather, the states might be $z = $ rain and $z = $ shine. If the stock market, we might assign different states to every possible value of the S&P 500 on (say) the last day of the year.

- *Probabilities.* For each state z, we assign a number $p(z)$ that we refer to as the *probability* of z. We call the complete collection of $p(z)$'s the *probability distribution*. Not every set of numbers works. Legimate probabilities must be nonnegative and sum to one.

- *Random variables.* A *random variable* is a function that assigns a real number to every state: $x(z)$. Note that x inherits randomness from z so it is, precisely, a random variable. Sometimes people distinguish between the random variable and the values the random variable takes, but we'll use x for both.

In practice we often ignore z and define probabilities directly over x. Nevertheless, there are times when the distinction between states and random variables is helpful.

Some common examples of probability distributions of random variables:

- *Bernoulli.* The state space $\mathcal{Z}$ has two elements: $\mathcal{Z} = \{z_1, z_2\}$. If we're flipping a coin, z_1 might represent heads and z_2 tails. A random variable assigns numbers to these two states. The simplest version uses zero and one: $x(z_1) = 0$, $x(z_2) = 1$. The probabilities $p(z_1)$ and $p(z_2)$ are nonnegative numbers that sum to one.

- *Poisson.* This is a little more complicated, but it's a beautiful example that's often used in finance (many other places, too, of course). Suppose $\mathcal{Z}$ consists of the nonnegative integers: $\{0, 1, 2, \ldots\}$. The probability of any particular z is

$$p(z) \;\; = \;\; e^{-\omega} \omega^z / z!,$$

with parameter $\omega > 0$ ("intensity").

Are they legitimate probabilities? Well, they're all positive, so we're good there. Do they sum to one? That's more complicated. The exponential function has the power series expansion

$$e^x \;\; = \;\; 1 + x + x^2/2 + x^3/3! + x^4/4! + \cdots \;\; = \;\; \sum_{j=0}^{\infty} x^j / j!. \tag{1}$$

Our probabilities have a similar form:

$$\sum_{z=0}^{\infty} p(z) \;\; = \;\; e^{-\omega} \sum_{z=0}^{\infty} \omega^z / z! \;\; = \;\; e^{-\omega} e^{\omega} \;\; = \;\; 1.$$

So they are, in fact, legitimate probabilities for the state space $\mathcal{Z}$.

- *Normal (Gaussian).* Here we'll let the state space be the real line and set $x = z$, which allows us to ignore z from now on. We refer to such random variables as continuous, to distinguish them from random variables that based on a discrete set of states, such as the two we just looked at.

For continuous random variables, we describe probabilities with what's called a *probability density function* $p(x)$. Probabilities over an interval $[a, b]$ are integrals,

$$\text{Prob}(a \leq x \leq b) \;\; = \;\; \int_a^b p(x) dx.$$

The function $p(x)$ must be nonnegative for all values of x and integrate to one,

$$\int_{-\infty}^{\infty} p(x) dx \;\; = \;\; 1,$$

the analog of the sum we used earlier.

A *normal* or Gaussian random variable has density function

$$p(x) = (2\pi\sigma^2)^{-1/2} \exp[-(x-\mu)^2/(2\sigma^2)],$$

the well-known "bell-shaped curve." [If you graph this, you'll see why.] It's positive for all x and integrates to one, although we'll take the latter as a fact rather than demonstrate it. The so-called *standard normal* refers to the case $\mu = 0$ and $\sigma = 1$.

There are lots of other common distributions of random variables, many of which are summarized in Wikipedia. (I know Wikipedia has a bad rep, but the math content is often pretty good.)

2 Expectations and moments

The behavior of a random variable is described completely by (i) the set of values it can take and (ii) their probabilities. Sometimes that's too much information: we'd like a smaller number of properties that capture some salient features. I kind of like looking at the whole distribution, or even better a picture of it, but summary numbers can be useful, too.

We start with the concept of an *expectation*. If x is a random variable with probabilities p, its expectation is the weighted average value of x using probabilities as weights:

$$E(x) = \sum_z x(z)p[x(z)].$$

For a continuous random variable, we replace the sum with an integral. The notation is a little cumbersome, which is why we'll just write $E(x)$ most of the time. We can extend this concept to any function of x:

$$E[f(x)] = \sum_z f[x(z)]p[x(z)].$$

There may be cases where the sum (or integral) doesn't converge, but we won't worry about that now — or ever, really.

A *moment* is the expectation of a power of x:

$$\mu'_j = E(x^j)$$

for j a positive integer. The first one (μ'_1) is known as the *mean*. It's a measure of the "location" of the probability distribution. If we change the mean, the probability distribution shifts left and right. Think about graphing the probability distribution of $x+a$ for different values of a. If the mean of x is μ'_1, then the mean of $x+a$ is $\mu'_1 + a$, so when we move the distribution back and forth by changing a, that's reflected in the mean.

We often use *central moments* instead, meaning we look at powers of x minus its mean:

$$\mu_j = E[(x-\mu'_1)^j]$$

3

The idea is to take location out of the calculation. The first central moment is zero by construction:

$$\mu_1 \;=\; E(x - \mu_1') \;=\; E(x) - \mu_1' \;=\; \mu_1' - \mu_1' \;=\; 0.$$

If any of these steps seem mysterious, write out the calculation of the expectation for the discrete case.

The second central moment is called the *variance*, a measure of "dispersion": how spread out the distribution is. The *standard deviation* is the (positive) square root of the variance and is often used the same way. If we write out the definition of the variance, we see it can be expressed in terms of (noncentral or raw) moments:

$$E[(x - \mu_1')^2] \;=\; E[x^2 - 2x\mu_1' + (\mu_1')^2] \;=\; \mu_2' - 2(\mu_1')^2 + (\mu_1')^2 \;=\; \mu_2' - (\mu_1')^2.$$

If we wanted to, we could compute the variance this way. That's pretty common, but we'll see shortly there's a better (by which I mean easier) way. The standard deviation measures the "scale" of a random variable in the following sense: if the standard deviation of x is σ, then the standard deviation of ax is $|a|\sigma$.

Let's try to compute the mean and variance for the distributions we looked at earlier:

- *Bernoulli.* Let the probability that $x = 1$ be ω and the probability that $x = 0$ be $1 - \omega$. How do we find the mean and variance? The easiest way is to look them up in Wikipedia (search "Bernoulli distribution"), but let's see if we can find them on our own. The mean is (apply the definition)

  $$E(x) \;=\; (1 - \omega) \cdot 0 + \omega \cdot 1 \;=\; \omega.$$

  We find the variance from the second moment:

  $$E(x^2) \;=\; (1 - \omega) \cdot 0^2 + \omega \cdot 1^2 \;=\; \omega.$$

  The variance is therefore

  $$\mathrm{Var}(x) \;=\; E(x^2) - [E(x)]^2 \;=\; \omega - \omega^2 \;=\; \omega(1 - \omega).$$

  The standard deviation is the square root of this.

- *Poisson.* The mean is

  $$E(x) \;=\; e^{-\omega} \sum_{j=0}^{\infty} j\omega^j / j! \;=\; e^{-\omega}\omega \sum_{j=1}^{\infty} (j-1)\omega^{j-1} / (j-1)! \;=\; \omega.$$

  [You'll have to think about this a little – or better yet, wait a few minutes and we'll derive this by an easier route.] The second moment is

  $$E(x^2) \;=\; e^{-\omega} \sum_{j=0}^{\infty} j^2\omega^j / j! \;=\; ??.$$

  We could fight our way through this, but since an easier way is just around the corner, I'll surrender now and come back to fight another day.

4

- *Normal.* We'll postpone this one, too.

Our last topic here is *sample moments*: moments computed from data. The idea is to use sample weights rather than probabilities. Given a sample of x_t's for $t = 1, 2, \ldots, T$, the sample mean is

$$\bar{x} = T^{-1} \sum_{t=1}^{T} x_t$$

and the jth sample moment is

$$T^{-1} \sum_{t=1}^{T} x_t^j.$$

Sample central moments are the same, but we subtract the sample mean from x_t first:

$$T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^j.$$

If $j = 2$ we get the sample variance. If the x_t's are produced by a specific distribution, then with enough data we would hope that the sample moments will be "close" to the moments of the distribution.

3 Generating functions

Next up is one of my favorite tools: generating functions. It's a tool with a wide range of uses, but we're interested in one: as a short-cut in computing moments. If you've run across Laplace, Fourier, or z transforms, they're closely related.

The *moment generating function* (mgf) is defined by

$$h(s) = E\left(e^{sx}\right), \tag{2}$$

a function of the real number s. (It's common to use t instead of s, but we need t for time.) Note that $h(0) = 1$. [Ask yourself why.] The mgf is a tool, like a hammer, and we can hammer things with it. If we hammer probability distributions, we get moments as a byproduct.

Recall the power series expansion (1) of the exponential function. If we expand e^{sx} the same way and take expectations, the moments pop out:

$$
\begin{aligned}
h(s) &= E\left(1 + (sx) + (sx)^2/2 + (sx)^3/3! + \cdots\right) \\
&= 1 + \mu_1' s + \mu_3'(s^2/2) + \mu_3'(s^3/3!) + \cdots.
\end{aligned}
$$

With a little more insight, we see that we can recover the moments by differentiating h and setting $s = 0$. The first derivative is the mean:

$$h^{(1)}(0) = \left.\frac{dh(s)}{ds}\right|_{s=0} = \mu_1'.$$

5

Here $h^{(1)}(0)$ means the first derivative of the function $h(s)$ evaluated at $s = 0$. Similarly, high-order moments follow from high-order derivatives:

$$h^{(j)}(s) = \left. \frac{d^j h(s)}{ds^j} \right|_{s=0} = \mu'_j$$

This looks horrible, but it just says that the jth moment is the jth derivative. Bottom line: if we know the mgf, we can find moments by differentiating it. Better yet, we can get Matlab to do the differentiating.

Let's go back to our examples:

- *Bernoulli.* The mgf is

$$h(s) = (1 - \omega)e^{s \cdot 0} + \omega e^{s \cdot 1} = (1 - \omega) + \omega e^s.$$

The first two derivatives give us the first two (noncentral) moments:

$$h^{(1)}(0) = \omega = \mu'_1$$
$$h^{(2)}(0) = \omega = \mu'_2.$$

The variance is therefore $\mu'_2 - (\mu'_1)^2 = \omega(1 - \omega)$, as we saw earlier.

- *Poisson.* The mgf is

$$h(s) = \sum_{z=0}^{\infty} e^{sz} e^{-\omega} \omega^z / z! = e^{-\omega} \sum_{z=0}^{\infty} (e^s \omega)^z / z! = e^{-\omega} e^{e^s \omega} = e^{\omega(e^s - 1)}.$$

The first two derivatives are

$$h^{(1)}(0) = \omega$$
$$h^{(2)}(0) = \omega + \omega^2.$$

The mean and variance are therefore both equal to ω.

- *Normal.* We find the mgf by completing the square:

$$h(s) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} e^{sx} e^{-(x-\mu)^2/2\sigma^2} dx.$$

The exponents are

$$sx - (x - \mu)^2/2\sigma^2 = -(1/2\sigma^2)\left[-2\sigma^2 sx + x^2 - 2\mu x + \mu^2\right]$$
$$= \mu s + \sigma^2 s^2/2 - [x - (\mu + s\sigma^2)]^2/2\sigma^2. \tag{3}$$

This may take you a couple minutes, but try expanding both expressions and lining up terms. It's an important equation, we'll come across it again.

When we plug the result into the integral and rearrange terms, we have

$$h(s) = e^{\mu s + \sigma^2 s^2/2} (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} e^{-[x-(\mu+s\sigma^2)]^2/2\sigma^2} dx.$$

The last term is a normal density function and therefore integrates to one. That leaves us with

$$h(s) \;=\; e^{\mu s + \sigma^2 s^2/2}.$$

If you differentiate, you can show that this implies a mean of μ and a variance of σ^2. [You might try this yourself, to make sure you're following the logic.]

The moment generating function gives us, in these and many other cases, an easy route to finding moments. The *cumulant generating function* (cgf) is even better. It's defined as the logarithm of the moment generating function:

$$k(s) \;=\; \log h(s). \tag{4}$$

[In this class and in Matlab, log means the natural or base-e logarithm. Always.] The cgf has the expansion

$$k(s) \;=\; \left(\kappa_1 s + \kappa_2(s^2/2) + \kappa_3(s^3/3!) + \cdots \right).$$

Its derivatives give us what are called *cumulants* κ_j, defined by

$$\kappa_j \;=\; k^{(j)}(0).$$

We can connect cumulants to moments by linking derivatives of k to those of h using (4). For example, the first two derivatives are

$$\begin{aligned}
k^{(1)}(0) &= h^{(1)}(0)/h(0) \;=\; h^{(1)}(0) \\
k^{(2)}(0) &= h^{(2)}(0) - h^{(1)}(0)^2,
\end{aligned}$$

the mean and variance.

Unlike the mgf, the derivatives of the cgf "centralize" moments automatically. We see that in the second derivative above (we get the variance directly, rather than the second noncentral moment), but that's an example of a more general property. The cgf of $y = a + bx$ is (apply the definition)

$$k(s; y) \;=\; as + k(sb; x).$$

We say a changes the location of the random variable and b the scale. The cumulants of y are connected to those of x by

$$\begin{aligned}
\kappa_1(y) &= a + b\kappa_1(x) \\
\kappa_j(y) &= b^j \kappa_j(x) \quad \text{for } j = 2, 3, \ldots
\end{aligned}$$

That is: after the first one, cumulants aren't affected by location, and scale shows up as a power.

After the mean and variance, the most useful moments/cumulants are the third and fourth measuring, respectively, skewness and kurtosis. Skewness refers to the asymmetry of the

7

distribution: odd cumulants (and central moments) are zero after the first for any symmetric distribution. Kurtosis refers to how much weight is in the tails of the density; holding the mean and variance constant, a distribution with greater kurtosis will have more weight in the tails and, to keep the variance constant, more near the center as well. There's no theorem to that effect, but it's a useful statement nonetheless. [Draw a picture.]

Two standard measures of skewness and kurtosis are based on the third and fourth cumulants:

$$\gamma_1 = \kappa_3/(\kappa_2)^{3/2} \quad \text{(skewness)}$$
$$\gamma_2 = \kappa_4/(\kappa_2)^2 \quad \text{(excess kurtosis)}$$

The denominators take care of scaling. [If you don't see this, note how scaling affects cumulants, and therefore γ_1 and γ_2.] We'll explain the term "excess" shortly.

Our examples again:

- *Bernoulli.* The cgf is

$$k(s) = \log\left[(1-\omega) + \omega e^s\right].$$

  For practice: compute the first four cumulants and the measures of skewness and excess kurtosis.

- *Poisson.* The cgf is

$$k(s) = \omega(e^s - 1).$$

  Its derivatives are all the same, so we have

$$\kappa_j = \omega$$

  for all $j \geq 1$. Skewness is $\gamma_1 = \omega/\omega^{3/2} = \omega^{-1/2} > 0$. Excess kurtosis is $\gamma_1 = \omega/\omega^2 = \omega^{-1} > 0$.

- *Normal.* The cgf is

$$k(s) = \mu s + \sigma^2 s^2/2.$$

  What's wonderful about this is that all cumulants after the first two are zero. Any nonzero cumulants beyond that are signs that the distribution isn't normal. Skewness and excess kurtosis are examples of that.

  One last thing: why do we say "excess" kurtosis? An alternative measure of kurtosis follows from the fourth central moment:

$$\mu_4/(\mu_2)^2.$$

  Since $\mu_2 = \kappa_2$, only the numerator differs from our earlier measure. And since neither measure depends on location or scale, it's enough to look at the standard normal,

which has $\mu = 0$ and $\sigma = 1$. The mgf is therefore $h(s) = e^{s^2/2}$ and its fourth derivative is

$$h^{(4)}(s) = 3e^{s^2/2} + 6s^2 e^{s^2/2} + s^4 e^{s^2/2}.$$

[I did this in Matlab because I'm lazy.] The fourth moment (central because the mean is zero) is therefore $h^{(4)}(0) = 3$. Why 3? That's just the way it is. But it tells us that if the kurtosis of a normal random variable is three, we need to subtract three to detect departures from normality. That's what γ_2 does.

4 Relations between random variables

So far we've looked at single random variables. But economics and finance — and lots of other things as well — are concerned not with the properties of single random variables, but with relations among two or more of them. It's not enough to know the distributions of GDP growth and equity returns, we'd also like to know if they're related, and if so, how. We need a language and tools for talking about that.

We start with *independent* random variables, which we define shortly. The setup for "multivariate" (more than one) random variables is similar to what we've seen. The probability density function for a two-dimensional random variable (x_1, x_2) might be expressed $p_{12}(x_1, x_2)$. We say x_1 and x_2 are independent if this factors into separate functions: $p_{12}(x_1, x_2) = p_1(x_1)p_2(x_2)$. You might think of coin flips. If two flips are independent, then the probability of two heads (say) is just the product of each head separately. If the probability of heads is one-half each time, then the probability of two heads is one-fourth. That's what independence is.

Interdependence — or whatever we call the opposite of independence — comes in many forms. The most direct connection is a linear one, which we can document with *covariances* and *correlations*. The covariance between two random variables x_1 and x_2 is

$$\text{Cov}(x_1, x_2) = E[x_1 - E(x_1)][x_2 - E(x_2)].$$

It's an example of a "cross moment," a moment that involves two or more random variables. If the covariance is positive, high values of x_1 are associated, more often than not, with high values of x_2. If negative, the reverse. Their correlation is a scale-free version:

$$\text{Corr}(x_1, x_2) = \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)^{1/2}\text{Var}(x_2)^{1/2}},$$

a number between minus one and one. You might verify for yourself that if we look instead at the variables $a + bx_1$ and $c + dx_2$, the correlation is the same: a and c drop out when we subtract the means, and b and d cancel when we calculate the ratio.

For now we'll focus on the sample analogs. The sample means and variances, of course, are

$$\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$$

$$\text{Var}(x_i) = T^{-1} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2$$

9

for variables $i = 1, 2$. The sample covariance of x_1 and x_2 is

$$\text{Cov}(x_1, x_2) = T^{-1} \sum_{t=1}^{T} (x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2)$$

and the sample correlation is

$$\text{Corr}(x_1, x_2) = \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)^{1/2}\text{Var}(x_2)^{1/2}}.$$

You can generally get the idea from a scatterplot of the two variables.

This is a measure of linear association. It's entirely possible for two variables to be related in a nonlinear way but have a zero covariance and correlation. Consider this example: Let x_1 take on the values $\{-1, 0, 1\}$ with probability one-third each. Then let $x_2 = (x_1)^2$. The two variables are clearly related, but their covariance is zero.

5 Sums and mixtures

We'll sometimes create random variables from combinations, either sums or mixtures of independent components.

Suppose we start with the independent random variables x_1 and x_2. The distribution of the sum $y = x_1 + x_2$ has a distribution that depends (evidently) on the distributions of x_1 and x_2. Since x_1 and x_2 are independent, the mgf of y is

$$h_y(s) = E\left(e^{sy}\right) = E\left(e^{s(x_1+x_2)}\right) = E\left(e^{sx_1}e^{sx_2}\right) = E\left(e^{sx_1}\right)E\left(e^{sx_2}\right).$$

The last step follows from independence of x_1 and x_2. This implies $h_y(s) = h_1(s)h_2(s)$: the mgf of the sum is the product of the mgf's of the components.

The cgf's of sums are even simpler. The cgf of the sum is

$$k_y(s) = \log h_y(s) = \log h_1(s) + \log h_2(s).$$

[Think about this a minute. Make sure you follow the notation.] In words: the cgf of a sum of independent random variables is the sum of the cgf's of the components. From this it follows that the cumulants of the sum y are the sums of the cumulants of the components x_1 and x_2.

Thinking ahead, sums are a device for introducing nonnormality: if at least one of the components is nonnormal, then so it the sum. [Can you see why that is?]

A mixture is a more complicated object, but also a more interesting one. Suppose we combine x_1 and x_2 with a Bernoulli random variable. With probability $1 - \omega$ we get x_1 and with probability ω we get x_2. Then the mgf is

$$h_y(s) = (1-\omega)E\left(e^{sx_1}\right) + (1-\omega)E\left(e^{sx_2}\right).$$

The cgf is the log of this.

It's less obvious, but this is also a device for introducing nonnormality. That's true here even if x_1 and x_2 are normal. Some of the most popular models in finance start with mixtures like this.

10

Bottom line

We have seen how to formalize the concept of randomness. We also have some summary measures that describe it. The mean and variance describe location and dispersion. Skewness and excess kurtosis describe the shape of the distribution and are independent of scale and location. They are useful in identifying departures from normality, the first focusing on asymmetry, the second on the tails. Covariances and correlations describe relations between random variables. We'll use them all — intensively — from now on.