Notes on Random Variables, Expectations, Probability Densities, and Martingales

Includes Exercise Due Tuesday, April 4.

For many or most of you, parts of these notes will be review. If you have had multivariable calculus and econometrics or statistics, it should all be review until the section on martingales. If the math used in these notes and in the exercise at the end seems very unfamiliar or difficult to you, be sure to let me know and to raise questions in class when the math starts to go by too fast. From the first class meeting it appears that everyone has at least univariate calculus plus some course, probably econometrics, in which they studied probability and the normal distribution. These notes use integrals with respect to several variables, which may be new if you've not had multivariable calculus.

The first part of these notes is material that should be in most introductory undergraduate probability or econometrics textbooks. The sections on the law of iterated expectations and on martingales overlap the assigned material in Campbell, Lo and MacKinlay.

Parts of the exercise and the notes assume that you can work with matrix multiplication, inversion, determinants, etc. If this is not true for you, let me know and we will go over these concepts in class Thursday, 3/30.

1. Random Variables

A **random variable** is a mathematical model for something we do not know but which has a range of possible values, possibly some more likely than others. Examples include the sum of the dots on rolled dice, the value of a share of Microsoft stock 6 months from now, and the name of the prime minister of Russia 3 weeks from now. For most of us the value of a share of Microsoft stock 6 months ago is also a random variable, though once we have looked it up somewhere we could treat it as a simple number. If dice are fair, each number of dots, from 1 through 6, is equally likely on each die. We formalize this by saying that each of the numbers from 1 to 6 has probability $\frac{1}{6}$. The set of numbers from 1 to 6, the range of possible values, is what we call the **probability space** of the random variable, and the set of numbers, adding up to 1, that we attach to the elements of the probability space (in this case $\frac{1}{6}$ on each of the six numbers in the space) is the **probability distribution** on the space. When we roll two dice together and consider their sum, the probability space becomes the integers from 2 to 12, and it no longer makes sense to give them all equal probability.

A random variable like the value of Microsoft shares in 6 months does not have a finite list of possible values. It could lie anywhere between zero and some very large positive number. (Actually, stock prices are always quoted in discrete increments, so that there really are only finitely many positive values, but the number of possible values is very large, so we are going to pretend, as does most financial modeling, that the price could actually be any real number.) For a random variable like this, every individual real number has zero probability, but intervals of real numbers can have positive probability. We usually specify the probability of such intervals by specifying a **probability density function** or pdf. A pdf for a single random variable X taking on real values is a function $f(\cdot)$ defined on the real line that is everywhere non-negative and satisfies

$$\int_{-\infty}^{\infty} f(x)\,dx = 1\,. \tag{1}$$

The probability of an interval (a,b) of values for the random variable is then

$$P[X \in (a,b)] = \int_a^b f(x)\,dx. \tag{2}$$

Random variables that take on no single numerical value with positive probability, but have a pdf over the real line are called **continuously distributed**, while those that take on a list of possible values, each with positive probability, are called **discretely distributed**. There can also be random variables that mix these two categories.

A set of random variables $\{X_1, \ldots, X_n\}$ may have a **joint distribution**. The simplest sort of example would be the joint distribution of the values of two dice rolled together. Each can take on the values 1 to 6. The joint probability space for them is the set of pairs of numbers (n_1, n_2), with each of n_1 and n_2 taking on values 1 to 6. We can display this probability space graphically as:

		die 2					
		1	2	3	4	5	6
	1	•	•	•	•	•	•
	2	•	•	•	•	•	•
die 1	3	•	•	•	•	•	•
	4	•	•	•	•	•	•
	5	•	•	•	•	•	•
	6	•	•	•	•	•	•

If each side of a single die has probability $\frac{1}{6}$ and they are thrown fairly, we usually assume that each of the 36 dots in this diagram has the same probability. Since there are 36 dots and the probabilities add to one, each has probability $\frac{1}{36}$. Note that now we can see why it does not make sense to give equal probability to all possible sums of values on the dice. The sum of the two dice is the same along diagonal rows of the diagram, running from upper right to lower left. The sum is two just at the diagram's upper left corner, 12 just at the lower right corner, and 7 along the longest diagonal, running from lower left to upper right. So a sum of 7 occurs at 6 points and has a total probability of $\frac{1}{6}$, while 2 has probability $\frac{1}{36}$.

For a pair of continuously distributed random variables $\{X_1, X_2\}$ the joint distribution is described by a **joint pdf** $f(\cdot,\cdot)$. The probability of a region of values in which

X_1 lies in (a,b) and X_2 lies in (c,d) is the integral of $f(x,y)$ over the rectangle defined by $a < x < b$ and $c < y < d$, i.e. it is

$$P[X_1 \in (a,b) \,\text{and}\, X_2 \in (c,d)] = \int_{\langle x,y\rangle \in (a,b)\times(b,c)} f(x,y)\,dx\,dy$$
$$= \int_a^b \int_c^d f(x,y)\,dx\,dy. \quad (3)$$

In (3) we introduce the notation $A \times B$ for the collection of all pairs of points $\langle x,y\rangle$ such that x is in A and y is in B. When A and B are intervals on the real line, $A \times B$ is just a rectangle.

If you have not had multivariable calculus, the double integral above may be a new concept. For our purposes, it is only very slightly new. You can think of $\int f(x,y)\,dx\,dy$ as just an ordinary one-variable integration with respect to x, with y held constant, followed by an ordinary one-dimensional integration with respect to y.

2. Expectations

For any function g of a random variable X with pdf f we can form the **expectation** of $g(X)$ as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx\,. \quad (4)$$

Expectation is actually a more general idea than probability. The probability of an interval (a,b) is just the expectation of the **indicator function** for the interval, defined as the function

$$\mathcal{I}_{(a,b)}(x) = \begin{cases} 1 & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases}\,.$$

With this definition,

$$P[X \in (a,b)] = E[\mathcal{I}_{(a,b)}(X)]\,. \quad (5)$$

These definitions all generalize in the obvious way to the case where X is a vector of random variables. That is, we can replace intervals with rectangles and one-dimensional integration with multivariate integration.

If we are given a joint distribution for two vectors of random variables X and Y, we may want to know what is the implied distribution for one of them by itself. The distribution for X alone, extracted from a joint distribution, is know as the **marginal distribution** of X. If X and Y have a discrete joint distribution (like that of the two dice), and if we use the notation

$$p(\langle x,y\rangle) = P[X = x \,\text{and}\, Y = y]\,, \quad (6)$$

then the marginal distribution of X can be calculated from

$$P[X = x] = \sum_{\text{all possible values } y \text{ of } Y} p(\langle x, y \rangle) \,. \tag{7}$$

When instead X and Y are jointly continuously distributed, so they have a pdf $f(x, y)$, the marginal pdf $g(x)$ of X is found from

$$g(x) = \int f(x, y)\, dy \,. \tag{8}$$

3. Conditional Expectation and Conditional Probability

If we have two vectors (i.e. lists) of random variables, X and Y, and a function $g(X, Y)$ of them both, the **conditional expectation of** $g(X, Y)$ **given** X is a function of X written as $E[g(X, Y) \,|X\,]$. It is defined by the property that for any function $h(X)$ of X alone such that $E[g(X, Y)h(X)]$ exists, we have

$$E[g(X, Y)h(X)] = E\big[E[g(X, Y) \,|X\,]h(X)\big] \,. \tag{9}$$

This definition is often not very helpful if we are trying to calculate conditional expectations, but it makes it easy to understand the main applications of conditional expectations in finance.

Just as we can take the expectation of an indicator function to get the probability of a set of numbers (as in as in (5)), we can take the conditional expectation of an indicator function to get what is called a **conditional probability**. We write

$$P[Y \in (a, b) \,|X\,] = E[\mathcal{I}_{(a,b)}(Y) \,|X\,]. \tag{10}$$

Except in special circumstances that we will probably not encounter, conditional probabilities defined this way behave like ordinary probability distributions when we hold the conditioning variable (X in (10)) fixed and consider how they behave as we vary the set ((a, b) in (10)). In particular, if X and Y are jointly continuously distributed, so they have a pdf $f(x, y)$, then usually the conditional distribution of Y given X also has a pdf, which we might write as $f(y \,|x\,)$. We can nearly always find the conditional probability from the formula

$$f(y \,|x\,) = \frac{f(x, y)}{\int f(x, y)\, dy} \,. \tag{11}$$

Notice that the denominator of the expression in (11) is just the marginal pdf of X. So in words, the conditional pdf of Y given X is the ratio of the joint pdf of Y and X to the marginal pdf of X.[1]

[1] This formula only works "nearly" always because it is possible, even though f integrates to one in its two arguments jointly, that for a few isolated values of x the integral in the denominator of (11) might not exist. We'll try to avoid encountering such cases in this course.

For discretely distributed variables, conditional probabilities can always be formed by the discrete analogue of (11):

$$P[Y = y|\, X = x] = \frac{P[X = x \,\text{and}\, Y = y]}{P[X = x]} \,. \tag{12}$$

At least in classroom problem sets, the simplest way to calculate conditional expectations for continuous distributions is usually to use (11) to form a conditional pdf, then to integrate the product of the conditional pdf with the thing whose expectation is being taken. That is, use

$$E[g(X,Y)\,|x] = \int g(x,y) f(y\,|x) dy \,. \tag{13}$$

For discrete distributions the analogous formula is

$$E[g(X,Y)\,|x] = \sum_y g(x,y) P(Y = y\,|X = x) \,. \tag{14}$$

Often instead of starting from a joint pdf, we start with a model that describes a conditional pdf. Then we might want to construct a joint pdf from a given marginal pdf for X and conditional pdf for $Y|X$. It is easy to see that (12) can be turned around to read

$$f(x,y) = f(y|x) f(x) \,. \tag{15}$$

Note that here we are relying on the different argument lists for f to make it clear that there are in fact three different functions that are all denoted f. The left-hand side of (15) is the joint pdf of Y and X, the first f on the right is the conditional pdf of $Y|X$, and the second f on the right is the marginal pdf of X. This is sometimes handy notation, but it can lead to confusion, especially if, say, we are dealing with both a $Y|X$ pdf and a $X|Y$ pdf in the same discussion.

4. The Law of Iterated Expectations and a Model for Information Flow

A property of conditional expectations that plays a central role in finance models is the Law of Iterated Expectations, which we have actually presented as the definition of conditional expectation in (9). It is often presented in a form that follows from (9):

$$E[Y|X] = E[E[Y|X,Z]|X]. \tag{16}$$

In economic and financial applications, we often think of information emerging at a time t as a vector X of random variables that can be seen at t, but not before t. The information available at t is then the set of all X_s vectors with dates $s \le t$, denoted $\{X_s|s \le t\}$. (Note that this implicitly assumes that information once seen is never forgotten.) If P_{t+s} is the price of an asset at time $t+s$, we then often want to consider its expectation conditional on all the information available at t. This situation occurs so often that we introduce special notation for it:

$$E_t[P_{t+s}] = E[P_{t+s}|X_s,\ \text{all } s \leq t]. \tag{17}$$

It is then a consequence of the law of iterated expectations that for $s \leq t$,

$$E_s[P_v] = E_s[E_t[P_v]]\,. \tag{18}$$

Note that for market participants, it must be that P_t is in the information set available at t—they couldn't trade at the price P_t if they didn't know what it was. This means that we always include P_t in the X_t vector, and that therefore $E_t[P_v] = P_v$ for any $v \leq t$.[2]

Note also that in introducing the "E_t" notation, we avoid the tedious writing out of the information set, which can be dangerous. The same equation, written in E_t notation, can have different implications according to what is being implicitly assumed about the information set.

5. The Normal, or Gaussian, pdf

A constantly recurring pdf in financial models is the Gaussian pdf. For a one-dimensional random variable X the Gaussian pdf is

$$\phi(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,. \tag{19}$$

The parameter μ turns out to be the expected value, or **mean**, of X, and the parameter σ is the **standard deviation** of X, which is defined as $\sqrt{E[(X-E[X])^2]}$. The **variance** of a random variable is its squared standard deviation, so for a Gaussian random variable it is σ^2. If X has the pdf given in (19) we say it has a normal, or gaussian, distribution, or that it has $N(\mu,\sigma^2)$ distribution, or more concisely just that $X \sim N(\mu,\sigma^2)$.

If X is an n-dimensional jointly normal vector of random variables, its joint pdf is, in matrix notation,

$$\phi(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}} \left|\Sigma\right|^{-1/2} \exp\left(-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)\,. \tag{20}$$

In this formula $|\Sigma|$ is the determinant of Σ, a notion you should have encountered before and that we will not define here. The matrix notation in the exponent can be written out as

$$x'\Sigma^{-1}x = \sum_{i,j=1}^{n} x_i x_j \sigma^{ij}\,, \tag{21}$$

[2]This may seem intuitively obvious, but you might see if you can prove it from our definition of conditional expectation in (9).

where σ^{ij} is the element in the i'th row, j'th column of Σ^{-1}. [3] In (20) the vector μ is the mean of the X vector and Σ is the **covariance matrix** of X, defined as

$$\Sigma = \mathrm{Cov}(X) = E[(X-\mu)'(X-\mu)] \,. \tag{22}$$

Avoiding matrix product notation, we can write this out by stating that σ_{ij}, the i'th row, j'th column element of Σ, is

$$\sigma_{ij} = E[(X_i-\mu_i)\cdot(X_j-\mu_j)] \,. \tag{23}$$

One reason the normal distribution occurs so often is that it allows calculation of marginal and conditional distributions directly, without explicitly evaluating integrals of pdf's. So in practice it is less important to know (20) than to know that if X_1 and X_2 are two jointly normal random variables with 2×1 mean vector μ and 2×2 covariance matrix Σ,

$$E[X_1|X_2] = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(X_2-\mu_2) \,, \tag{24}$$

and that, furthermore, the conditional distribution of $X_1|X_2$ is itself normal. The marginal distribution of X_1 in this case is just $N(\mu_1,\sigma_{11})$.

When X_1 and X_2 are themselves vectors, it is again true that the conditional distributions of one given the other are normal. We split the mean vector of the joint distribution into two pieces, and the covariance matrix into four pieces, corresponding to the division of the full vector made up of X_1 and X_2 into its two components. Then in matrix notation we can write

$$E[X_1|X_2] = \mu_1 + (X_2-\mu_2)'\Sigma_{22}^{-1}\Sigma_{21} \,. \tag{25}$$

In this expression we are using the notation that

$$\Sigma_{ij} = E[(X_i-\mu_i)\cdot(X_j-\mu_j)'] \,,\ i,j=1,2 \,. \tag{26}$$

6. The Simple Version of the Efficient Markets Hypothesis: The Martingale Model

A crude **arbitrage** argument supports the claim that, to a good approximation, asset prices should satisfy

$$E_t[P_{t+s}] = P_t \,,\ \text{ all } s\geq 0 \,. \tag{27}$$

The argument is that if this condition failed, the asset price P would be expected to be higher in the future than it is now, and that this would lead market participants to want to buy it now and sell it later. The incentive to buy now would push the current price up until (27) was satisfied. An arbitrage argument is any argument that, like this one, claims that something is impossible because, if it were true, there would

[3] We're not assuming a previous course in linear algebra for this course, but here I am assuming that people know what a matrix, its inverse, and its determinant are. My impression is that if you have taken econometrics or you should have encountered these notions already. If this is not a good assumption in your particular case, let me know promptly.

be a profit opportunity in the market that market participants would exploit to make the profit opportunity disappear. We will see later that there are many reasons why (27) doesn't hold exactly for asset prices, but it does usually hold to a pretty good approximation, and we will be working out the implications as if it did hold exactly.

The condition in (27) is often labeled the (simple) **efficient markets hypothesis**, though there are actually several related definitions of this term. In mathematics, (27) is the definition of a **martingale**. So saying P satisfies (27) is the same thing as saying P is a martingale.

7. Exercises

(1) Suppose that time is discrete, i.e. that t takes on only integer values, and that we replace (27) by the apparently weaker assertion that for all t, $E_t[P_{t+1}] = P_t$. Use the law of iterated expectations to show that this implies (27).

(2) There are only four time periods, $t = 1, 2, 3, 4$. We know that there are only four possible time paths for the asset price P_t, and we can list them as

P_t paths:

	t			
path no.	1	2	3	4
1	2	3	3	3
2	2	1	2	3
3	2	1	2	1
4	2	1	0	0

Here each row represents a possible time path for P_t. All paths have $P_1 = 2$, reflecting the fact that there is no uncertainty about P_1, but after that the course of prices is uncertain. The probabilities of the four paths are π_1, π_2, π_3, π_4. The information known at t consists only of values of P_s for $s \leq t$. If the four probabilities are $\pi_1 = .5$, $\pi_2 = .125$, $\pi_3 = .125$, and $\pi_4 = .25$, show that P is a martingale. Is P also a martingale if we change π_3 to .25 and π_4 to .125? If not, what is the profit opportunity implied by this change in the probabilities, and at what time would it appear? Note that what you have been given here is the joint distribution of $P_1, \ldots, P_4$, so that you will have to do some summing to get marginal distributions, form conditional distributions, and form expectations.

(3) Suppose an asset price P_t at three dates, $t = 1, 2, 3$, has a joint normal distribution with mean 2 at all three dates (i.e. a mean vector consisting of three 2's stacked up) and one of the following three covariance matrices:

(a) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 2 \\ 0 & 2 & 5 \end{bmatrix}$

(b) $\begin{bmatrix} 3 & 3 & 3 \\ 3 & 5 & 5 \\ 3 & 5 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} 6 & 5 & 3 \\ 5 & 5 & 3 \\ 3 & 3 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 3 & 3 \\ 3 & 4 & 3 \\ 3 & 3 & 5 \end{bmatrix}$

We again assume that information available at t is only current and past prices. For each of these covariance matrices, determine whether P_t is a martingale. Here you will want to use (25). (The distributions we assume here imply nonzero probabilities of negative P, which is unrealistic. Ignore this lack of realism in doing the exercise.)