TEACHING NOTE 99-04:
PROBABILITY AND STATISTICS REVIEW FOR FINANCE: PART I

Version date: October 24, 2000 C:\Classes\Tn99-04.wpd

As scholars of finance, we are frequently required to apply concepts from probability and statistics. Corporate cash flows, exchange rates, stock prices and interest rates are all random variables that must be analyzed by using probability and statistics. The first part of this teaching note provides a brief introductory review of the key concepts while the second part covers a number of important results that arise frequently in financial theories and applications.

Probability is a measure, meaning a numerical description or quantification, of the relative frequency of an event. The probability of getting heads in one toss of a fair coin is obviously 1/2. There are two possibilities and one fits the definition of the desired event. Technically this is called the *marginal probability*, which distinguishes it from the *conditional probability*, which is the probability that something will happen, given that something else has already happened. Finally we have the concept of *joint probability*, which is the probability that two events will happen. We shall cover these concepts in more detail later in this teaching note.

Mathematicians describe probability with great formality and extreme precision. Oftentimes you will see probability described in terms of a space called Ω (upper case Omega), which refers to a set of possible outcomes that might happen. If there are a finite number of possible outcomes, n, then Ω will consist of n elements, each element often described as ω_i (lower case omega). Associated with each ω_i is a probability, p_i. For example, the random event might be throwing a single die. The sample space Ω is the set of numbers $\{1, 2, \ldots, 6\}$, where $\omega_1 = 1, \omega_2 = 2, \ldots, \omega_6 = 6$.

Now suppose there are certain outcomes in which we have a particular interest. For example, we might be interested in the event that the die contains a number greater than 4. Let F be a subset of Ω consisting of the outcomes that $\omega_i > 4$. F is referred to as an *event*. Associated with each event is a complement, denoted as F^c. In our example, F^c is the set of events in which ω_i # 4. The union of sets F and F^c, denoted as F c F^c, is Ω. Consider an event G that 1 # ω_i # 2. The union of F and G, F c G, is a new set H, which would consist of the elements $\omega_i = 1, 2, 5, 6$. The intersection of two sets, such as $F \cap G$, is the set, D, defined as those points in common with both sets. In this case D is the null set $\emptyset$ because no points are in common in F and G. In terms of probabilities, unions of sets represent events in which either F or G occurs. Intersections of sets represent events in which both F and G occur.

More formally, a specific collection of subsets of Ω is denoted as $\mathcal{F}$, which are events that we are interested in and to which we can assign probabilities. In the die example the event might be that we are interested in those outcomes in which the number observed on a single roll of the die is not 3. Thus, the collection of subsets $\mathcal{F}$ is partially represented by the union of sets F and G; however, by admitting only those events contained in F and G, we rule out the possibility of events not contained in F and G. In fact we must include those events not contained in F and G in $\mathcal{F}$ for they do have some probability of occurring. $\mathcal{F}$ can also contain other specifications such as the event of ω_i being between 1 and 3, less than 2, being either 1, 4 or 6, etc. A collection of these subsets $\mathcal{F}$ is called a σ-field, referred to as a *sigma field*.[1] A σ-field can simply be thought of as a collection of events of interest. To such events we must be able to assign probabilities, as represented by a mathematical specification called a measure, often denoted as P. There are, however, other requirements for a σ-field.

Let us specify a collection of sets that will qualify as a σ-field. First we must define the precise requirements of a σ-field. A collection of subsets $\mathcal{F}$ is a σ-field if the following are true:

1. $\Omega \in \mathcal{F}$
2. Let $A_1, A_2, \ldots$, be a sequence of subsets of $\mathcal{F}$. Then the union of these subsets is also contained in $\mathcal{F}$. This means if we combine the subsets of $\mathcal{F}$, the new sets formed by these combinations must also be included in $\mathcal{F}$.
3. For any subset A, its complement, A^c, must also be in $\mathcal{F}$. That is, $A^c \in \mathcal{F}$. In other words, if we are interested in an event, then the result that this event does not occur cannot be omitted from the possible events.

Now let us build a candidate for a σ-field. Consider the following:

$$\mathcal{F} = \{\Omega, \emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}, \{3,4,5,6\}, \{2,4,5,6\},$$

[1]Despite the common use of σ in probability to represent standard deviation, this use of σ has nothing to do with standard deviation.

$\{1,4,5,6\},\{4,5,6\},\{2,3,4,5,6\},\{1,3,4,5,6\},\{1,2,4,5,6\}\}$

We see that $\mathcal{F}$ has 16 subsets, but $\mathcal{F}$ does not include all possible subsets. For example, the subset {1,5} is not included, as are many others.

Now let us see if $\mathcal{F}$ meets our requirements for a σ-field.

1. $\Omega \in \mathcal{F}$. This is definitely true. It is the first subset.
2. Now we must see if combinations of subsets are elements of $\mathcal{F}$. There are many such combinations. We need only find the union of each pair of subsets. Clearly adding Ω or $\emptyset$ to any other sets produces either Ω or the given subset. So we should just starting adding sequences of subsets and see if what we get also is included in $\mathcal{F}$.

   $\{1\} \cup \{2\} = \{1,2\}$, which is included

   $\{1,2\} \cup \{3\} = \{1,2,3\}$, which is included

   $\{1,2,3\} \cup \{1,3\} = \{1,2,3\}$, which is included

   This continues and soon all combinations equal Ω. In fact any ordered sequence of subsets combines in this manner.
3. We must examine the complements of each subset. We shall do a few here:

   $A_1 = \{1\}$, $A_1^c = \{2,3,4,5,6\} \in \Omega$

   $A_2 = \{2\}$, $A_2^c = \{1,3,4,5,6\} \in \Omega$

   This can be continued for all subsets of Ω and checks out. The complements all belong to $\mathcal{F}$

For $\mathcal{F}$ to not qualify as a σ-field, we could easily just drop one subset. For example, if the subset {1,3} were not included, then $\{1\} \cup \{3\} = \{1,3\} \notin \mathcal{F}$.

Certain rules apply to probabilities associated with σ-fields:

1. $P(\emptyset) = 0$ (something must happen)
2. P(Ω) = 1 (the probability of all events combined is 1)
3. If A is any subset in $\mathcal{F}$, then $P(A) \geq 0$ (A does not have to have any probability of occurrence but probability is not defined as a negative number.)

4. If A and B are disjoint sets, then they have no common elements. For example, let A = {1} and B = {2} be disjoint sets. Then $A \cap B = \{ \emptyset \}$ (the intersection of A and B is the null set), A and B are mutually exclusive, and

$$P(A \cap B) = 0$$

$$P(A \cup B) = P(A) + P(B).$$

The first statement is like saying that the probability of both A and B occurring is zero. This must be true because we have not included a set containing both A and B. The second statement is that the probability of either A and B occurring is the probability of A occurring plus the probability of B occurring.

Mathematical probability theory often describes random events in terms of what is called a *probability space*, written as $(\Omega, \mathcal{F}, P)$, which is a set Ω, a σ-field $\mathcal{F}$, and a probability measure P. In our experiment we take Ω to be the possible outcomes, the numbers 1-6, $\mathcal{F}$ to be a collection of certain subsets in Ω to which we can assign probabilities, and P to be the mathematical specification of the probability associated with the events contained in subset $\mathcal{F}$.

Oftentimes in finance, we work with a *stochastic process*, which is a sequence of probabilistic events, such as stock prices as they evolve on a day to day basis. When modeling such events, we often denote all of the possible sequences of stock prices as Ω. To specify the σ-field, we refer to a construction called a *filtration*, which is a sequence of what we call increasing σ-fields. In other words, we might have $\mathcal{F}_t$, $\mathcal{F}_{t+1}$, $\mathcal{F}_{t+2}$. Each of these sigma fields is said to contain the information in all previous σ-fields, which is what we mean when we say that they are *increasing*. This assumption is analogous in finance to the weak form efficient market hypothesis that the set of current information includes the set of all past information, i.e., no information is forgotten.

Unfortunately, the formality of specifying events and their probabilities in this matter makes the study of probability somewhat harder than it has to be. Mathematicians require a great deal of formality and can often demonstrate that when formality is sacrificed, some imprecision is injected into the problem. In finance we occasionally see problems set up in terms of a probability space such as $(\Omega, \mathcal{F}, P)$. This formality can be confusing, but it should not be. It is merely a mathematician's way of specifying precisely what the uncertainty is. It is quite possible, indeed likely, that a great deal of understanding of probability theory can be obtained without fully grasping this level of mathematical formality.

An excellent article that relates the mathematical formality described here to the commonly used binomial model of option pricing is McIver (1993). The appendix to this teaching note provides a specification of the σ-algebra of the binomial option pricing model.

Getting back to the less formal set up, we noted above the marginal probability of an event, such as the probability of a head on one toss of a coin being 1/2. Now suppose we toss two coins. What is the probability of a head? Now we have to define more precisely what we mean. Do we mean one head, two heads, or at least one head? We also have to be assured that the coins are independent, for if not, it does affect our answer. Assuming the two coins are independent, the probability of one head is the probability that we toss a head on coin one and a tail on coin two, which is (1/2)(1/2) = 1/4. This is what we mean by the joint probability. There are two ways to toss one head, however: a head on the first and a tail on the second or a tail on the first and a head on the second. So that probability is 1/2. If we want the probability of two heads, then we must toss a head on both coins. Since the events are independent, the joint probability is the product of the marginal probabilities. Thus, the probability of two heads is (1/2)(1/2) = 1/4. If we want the probability of at least one head, we require the probability of exactly one head plus the probability of two heads, which is 1/2 + 1/4 = 3/4. We can easily verify these results by summarizing the possible outcomes:

1. Head on first coin, head on second coin
2. Head on first coin, tail on second coin
3. Tail on first coin, head on second coin
4. Tail on first coin, tail on second coin

These are equally likely events so each has a probability of 1/4. The outcome of one head is represented by events 2 and 3 so the probability is 1/2. For two heads, there is just one way to do it, event 1. So that probability is 1/4. If we want the probability of at least one head, we include events 1, 2 and 3 so that probability is 3/4. Alternatively, the probability of at least one head can be restated as 1 minus the probability of no heads. There is only one event here with no heads, # 4, so the answer is 1 - 1/4 = 3/4 or alternatively, 1 - the probability of a tail on both tosses, which is 1 - 1/4 = 3/4.

The probability of two events occurring, as noted above, is called the joint probability. Defining the events as A and B, we state the joint probability for independent events in the following manner:

$$\text{Prob}(A \ \& \ B) = \text{Prob}(A)\text{Prob}(B), \text{ given independence of A and B,}$$

or several similar variations.[2]

For events that are dependent, we must make some modifications. First what do we mean by independence and dependence? The coin tosses are clearly independent. The outcome of one coin is unrelated to the outcome of the other. Consider the following experiment. Toss a coin one time. If heads occurs, record the value +1; if tails occurs, record the value - 1. Then toss the coin again. Once again record +1 or -1 and add it to the amount recorded after the first coin toss. We are interested in the probabilities of the numerical outcomes. Here are the equally likely possibilities:

1. Head on first coin, head on second coin; sum = +2
2. Head on first coin, tail on second coin: sum = 0
3. Tail on first coin, head on second coin: sum = 0
4. Tail on first coin, tail on second coin: sum = -2

Thus, the probability distribution is

Outcome	Probability
+2	¼
0	½
-2	¼

These are the marginal probabilities. We could also speak in terms of conditional probabilities. For example, we might wish to know the probabilities of +2, 0 and -2 given that a head occurred on the first toss. These probabilities are:

[2]For example, we sometimes use *Pr* or *P* for Prob.

Outcome	Probability[3]
+2	½
0	½
-2	0

If a tail occurs on the first toss, the probabilities are[4]

Outcome	Probability
+2	0
0	½
-2	½

The probabilities of +2 and -2 outcomes are quite different if a head or tail has occurred on the first toss. This is what we mean by the *conditional probability*. The events, defined specifically as the sum of the +1 or -1 assigned to the head or tail tossed, are not independent. The occurrence of a head or a tail, however, is independent from one toss to the other. It is interesting to note how the same sample experiment can give rise to a quite different set of probabilities for different events.

In general, we have the following result,

$$\text{Prob}(A \ \& \ B) = \text{Prob}(B{*}A)\text{Prob}(A) = \text{Prob}(A{*}B)\text{Prob}(B).$$

To cast our event in this context, think of the two tosses as two coins, coin # 1 and coin # 2. The outcome of a +2 can occur only in the joint condition that both coins come up heads. The probability of this occurring is obtained from the conditional probabilities as follows: the probability of a head on the second coin conditional on a head on the first coin (1/2) times the probability of a head on the first coin (1/2). Alternatively, it is the probability of a head on the first coin conditional on a head on the second coin (1/2) times the probability of a head on the second coin (1/2).[5]

---

[3]If we got a head on the first toss, we currently have a total of +1. Thus, there is a probability of ½ that we get another head, giving us a total of +2, and a probability of ½ that we get a tail, giving us a total of 0. There is zero probability that we end up with a total of 0.

[4]If we got a tail on the first toss, we currently have a total of -1. Thus, there is a probability of ½ that we get a head, giving us a total of 0, and a probability of ½ that we get another tail, giving us a total of -2. There is zero probability that we end up with a total of +2.

[5]In finance sometimes conditional expectations are written in the form, $E[X_{t+j}{*}I_t]$, meaning that the expectation of X at time t+j is based on the information set, I, available at time t. Mathematicians will often write this as $E[X_{t+j}{*}\,ö_t]$ where the information set is represented by the σ-field $ö_t$.

A probability distribution is a mathematical specification of the probabilities associated with events. For the example shown above, we easily laid out the probability distribution. Many types of random outcomes can be characterized with an exact mathematical formulation that gives either the probability of an event occurring or the probability of a range of events occurring. Coin tosses are from what is called a *binomial probability distribution*. This is one of a family of distributions that are called *discrete*, which means that only a finite number of outcomes can occur. Another family of distributions is described as *continuous*, meaning that the number of possible outcomes is infinite. Anything that can be measured with fractional precision is continuous. For example, the return on a stock is continuous. A stock bought at 99 7/8 and sold at 102 ½ has a return of 2.628285 .... %. There are an infinite number of possible returns, provided the return is measured with decimal precision and not rounded off to a certain number of decimal places. One major example of a continuous random variable is the familiar normal or bell-shaped distribution.

Consider a random variable, X. An expression such as Prob(X = a) is one example of information revealed by a discrete probability distribution. In some cases we wish to know the cumulative probability. For example consider the above problem in which we wish to know the probability of achieving a specific total after two coin tosses where a head counts as +1 and a tail counts as -1. Suppose we wish to know the probability that the total is non-negative. Then we have Prob(X $ 0), which equals Prob(X = 0) + Prob(X = 2) = ½ + ¼ = ¾.

If a random variable is continuous, we cannot specify a probability in terms of a specific value. For example, in the normal distribution we cannot specify Prob(X = 0). Because there are an infinite number of outcomes, the probability of any one outcome occurring is zero. We can, however, specify the probability of a range of outcomes occurring. Statements like Prob(X > 0) are quite acceptable.[6] The answer is 0.5, owing to the symmetry of the distribution and the fact that its expected value is zero. Likewise, we can use statements like Prob(b < X < a), which is easily found as Prob(X < a) - Prob(X < b).

Since the probability of a specific value of the random variable occurring is zero, there is no mathematical specification that gives such results as Prob(X = a). There is, however, a mathematical

[6]Note that since Prob(X = 0) = 0, then Prob(X $ 0) = Prob(X > 0).

specification that can lead to such statements as Prob(X < a). We start with a mathematical function referred to as the *probability density function*. If we plot such a function, specified as f(x), we observe a graph of values of a variable f(x) in terms of the random variable X.[7] The variable f(x) has no particular interpretation, but the area under the curve generated by f(x) is the probability we seek. Thus, the area under the curve and to the left of a particular value f(a) is Prob(X < a). Integrating the function, meaning to accumulate its values over a range of values of x, gives us the desired probability. Such a function is called the *probability distribution function,* or just the *distribution function*, and sometimes the *cumulative density function*.[8]

Probability distributions are often characterized not only with a density or distribution function, but also with a *moment-generating function*. This is a mathematical specification that incorporates the density or distribution function and yields what are called the moments of the distribution. The k^{th} raw moment of a distribution is defined as $E(X^k)$. As we shall see in the next section, the first moment (k=1) is the expected value and the second moment (k=2) is closely related to the variance. The third moment, $E(X^3)$, is closely related to the concept of *skewness*, which measures the symmetry of a probability distribution. The fourth moment, $E(X^4)$, is closely related to the concept of *kurtosis*, which measures the extent to which a distribution is peaked or flat.

Not all probability distributions have a distribution function, but all have a *characteristic function*, which although requiring use of complex numbers, can be used to yield many useful results. Characteristic functions and to some extent moment generating functions are occasionally but not often used in finance.

The above paragraphs provide only a brief review, but should be sufficient to refresh our memory of previous encounters with this material. In some cases certain concepts are being encountered for the first time. The reader will in all likelihood be required to refer to more specific material to fill in gaps and extend knowledge. We now turn to the primary operations used with

---

[7]At this point, you may have noticed that in some cases we use a capital X and in others, we use a lower case x. It is somewhat common to let the capital X represent the random variable, while the lower case x represents a specific outcome of the random variable. The function f(x) maps a specific outcome of X to a real number that, when integrated over, gives a probability.

[8]While f(x) is used for the density function, F(x) is often used for the distribution function.

random variables in finance courses, which are the taking of expectations, variances and covariances. This material is found in TN00-06.

Appendix: The σ-algebra of the Binomial Model

In the world of finance, binomial models play an important role. This is particularly true for pricing options. The binomial model is a special structure consistent with a σ-algebra. The article by McIver (1993) is, to my knowledge, the only published item where the formal mathematical theory of σ-fields is tied into the option pricing model. The McIver model is helpful but not sufficiently detailed to provide an understanding of this important linkage.

A stochastic process is a σ-algebra represented by an increasing sequence of σ-fields, which we shall designate as $\mathcal{F}_0$, $\mathcal{F}_1$, $\mathcal{F}_2$, ...etc. Each set $\mathcal{F}_t$, $t \geq 0$, is a σ-field representing the collective state of information as of time t. Let us use a two-period binomial model as our stochastic process and then proceed to define the elements of $\mathcal{F}_0$, $\mathcal{F}_1$, and $\mathcal{F}_2$.

First we define Ω as {uu,ud,du,dd}. This is the set of all possible events. Each σ-field must contain Ω as well as the null set $\emptyset$. At time 0, the σ-field consists of

$$\mathcal{F}_0 = \{\Omega, \emptyset\}.$$

Given the current state of information, all we know is that each path is possible.

Now we move to time 1, at which time we will have either gone up or down. Consequently, it will be meaningless to talk about events such as {du,ud} and {ud,du}; therefore, our σ-field is

$$\mathcal{F}_1 = \{\Omega, \emptyset, \{uu,ud\}, \{du,dd\}\}.$$

Let us make sure this qualifies as σ-field. Naturally it contains Ω and $\emptyset$. If we consider increasing sequences of its subsets, we would obtain $\{uu,ud\} \cup \{du,dd\} = \Omega$. Lastly, the subsets {uu,ud} and {du,dd} are mutual complements. Thus, the above specification is indeed a legitimate σ-field.

Now we move to time 2. We would now know which path was taken. Thus, we can start by saying that

$$\mathcal{F}_2 = \{\Omega, \emptyset, \{uu\}, \{ud\}, \{du\}, \{dd\}, ...\},$$

but we must leave open the possibility that there may be elements of $\mathcal{F}_2$ that we have not yet included. Remember that we must include combinations of sequences of subsets. Thus, we must also have

$$\{uu\} \cup \{ud\} = \{uu,ud\}$$
$$\{uu\} \cup \{du\} = \{uu,du\}$$
$$\{uu\} \cup \{dd\} = \{uu,dd\}$$
$$\{ud\} \cup \{du\} = \{ud,du\}$$
$$\{ud\} \cup \{dd\} = \{ud,dd\}$$
$$\{du\} \cup \{dd\} = \{du,dd\}.$$

We also know that the complements must be included. Therefore, we must have

$$\{uu\}^c = \{ud,du,dd\}$$
$$\{ud\}^c = \{uu,du,dd\}$$
$$\{du\}^c = \{uu,ud,dd\}$$
$$\{dd\}^c = \{uu,ud,du\}.$$

Consequently, the correct and final specification of our σ-field at time 2 is

$$\ddot{o}_2 = \{\Omega, i, \{uu\},\{ud\},\{du\},\{dd\},\{uu,ud\},\{uu,du\},\{uu,dd\},\{ud,du\},\{ud,dd\},\{du,dd\},$$
$$\{ud,du,dd\},\{uu,du,dd\},\{uu,ud,dd\},\{uu,ud,du\}\}.$$

This, as it turns out, is the set of all possible subsets.

References

Of course for a rigorous treatment, one should consult books on probability theory. For items specifically oriented toward finance, see

McIver, J. "Tree Power." *Risk* 6 (December, 1993), 58-64.

Martin, J. D., S. H. Cox, and R. D. MacMinn. *The Theory of Finance: Evidence and Applications*. Chicago: Dryden Press (1988), Appendix.