ECO 317 – Economics of Uncertainty – Fall Term 2009
Notes for lectures
1. Reminder and Review of Probability Concepts

1 States and Events

In an uncertain situation, any one of several possible outcomes may be realized. Treatment of this in probability theory begins by listing all the logical possibilities. These are called the *elementary events* in probability and statistical theories, and in our economic applications we will often call them *elementary states of the world or states of nature*, or sometimes in financial economics, all possible *scenarios*. Each elementary event or state is intended to be an exhaustive description of exactly one possible outcome; an elementary state of nature is intended to be an exhaustive description of exactly one set of circumstances. How fine or coarse the distinction is made, or what is exogenous and what is not, depends on the context of each specific application. As far as the mathematics of probability is concerned, the elementary events or states are just general abstract entities – they are among the basic concepts or "primitives" of the theory, and are assumed to follow certain specified rules and axioms. The logical deductions from this structure of rules constitutes the theory.

The set of all elementary events is called the *sample space* or *probability space*; in the economic context we will simply call it the *set of all states of nature*. Each subset of this space is called an *event*; singleton subsets have already been termed elementary events. In economic applications, the sample space (the full list of elementary events) should be exogenous, but individuals may control or affect the probabilities that attach to these events. The primary instances of this are mixed strategies and moral hazard.

Examples: [1] When a die is rolled, there are six elementary events corresponding to the number of dots on the top face. An event such as "the number is even" or "the number exceeds 4" are composite events. [2] When two coins are tossed, each can land either heads up or tails up. If the two coins are distinguishable, for example one is a quarter and the other a dime, or a coin is tossed twice and we keep track of the order in which the heads or tails appeared, then there are four elementary events: HH, TT, HT, TH. But if the coins are not distinguishable, then there are only three elementary events: two heads, two tails, and one head one tail. Thus the concept of elementary event can be specific to the context. [3] The upcoming interest rate choice of the Fed can be random from the perspective of an individual investor. (From a larger perspective that includes the Fed as a participant the interest rate may not be random; thus the very idea of randomness and probabilities may be specific to a context of application.) Suppose interest rates must be positive, they are denominated in basis points (one hundredth of a percentage point) and are absolutely sure to be less than or equal to 12 percent. Then there are 1200 elementary events. We could say that the Fed only ever uses 25-basis-point increments, so only 60 of the events are relevant and we should discard the rest from the sample space. But some day technology may allow finer adjustments of monetary policy so we may want to retain all 1200. Such decisions about practical modeling are matters for context-specific judgments.

2 Probability

The idea is to formalize the intuitive concept of how likely or unlikely is one event among the possible outcomes of an uncertain situation. Again the mathematics stands in the abstract, governed by the rules or axioms we impose. But these assumptions are usually made with some application in mind, and probability theory has behind it one of three motivations or interpretations:

Classical: Probability describes the physical chance that a controlled experiment produces some outcome. Example: radioactive decay or various phenomena in quantum physics. Sometimes events that are in principle deterministic but in reality too complex to calculate may be better modeled as probabilistic, e.g. some classical statistical mechanics.

Frequentist: Probability corresponds to the long-run frequency of an event in an experiment that can be repeated independently under identical conditions. Example: coin tosses.

Subjectivist: Probability is a numerical representation of an individual's subjective belief about the likelihood of an event. Example: who will win the Super Bowl.

We will generally adopt either a frequentist (objective) or subjectivist interpretation as suits our applications in a pragmatic way.

Let S denote the sample space, 2^S the set of its subsets (the set of all logically conceivable events), and R_+ the set of non-negative real numbers. For the moment, suppose S is finite.

We define *probability* as a function $Pr: \; 2^S \mapsto R_+$ with the following properties:

1. $Pr(\emptyset) = 0$, $Pr(S) = 1$.
2. If $A, B \subset S$ and $A \cap B = \emptyset$, then $Pr(A \cup B) = Pr(A) + Pr(B)$.

From this, one can immediately prove that if $A_i \subset S$ for $i = 1, 2, \ldots n$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} Pr(A_i)\,. \tag{1}$$

Just to give you an idea of how such proofs proceed, I will sketch this one.

When $n = 2$, this is just property 2 in the definition. Next we show that if the result is true for any n, the it is also true for $(n+1)$.

Let $\bigcup_{i=1}^{n} A_i = B$. We claim that $B \cap A_{n+1} = \emptyset$. For suppose not. Then there exists $x \in B \cap A_{n+1}$. Therefore $x \in B$, and therefore $x \in A_i$ for at least one $i = 1, 2, \ldots n$. But from the supposition we are making temporarily, we also have $x \in A_{n+1}$. Therefore $x \in A_i \cap A_{n+1}$ for this i, which contradicts $A_i \cap A_{n+1} = \emptyset$. Therefore our supposition must be false; this proves the claim.

Now we can use property 2 in the definition to write

$$Pr(B \cup A_{n+1}) = Pr(B) + Pr(A_{n+1})\,.$$

But

$$B \cup A_{n+1} = \bigcup_{i=1}^{n} A_i \cup A_{n+1} = \bigcup_{i=1}^{n+1} A_i\,,$$

and the assumption that the result is true for n gives us

$$Pr(B) = \sum_{i=1}^{n} Pr(A_i)\,.$$

Substituting into (1), we see that the result is true for $n+1$. QED

This is proof by mathematical induction, with an inner step (lemma, if you like) proved by contradiction. Such techniques will be useful from time to time.

This proof is so trivial that we could have directly made the n case the definition without loss of generality, but doing it this way served as a simple introduction to "proof math," which will appear from time to time in this course.

Since Pr is defined over subsets of S, we should write $Pr(\{s\})$ for the probability of an elementary event $s \in S$, but we will more simply write $Pr(s)$ without much risk of confusion.

What if S is a countably or uncountably infinite set? We might want to require the probability function to be countably or uncountably additive over pairwise disjoint events, that is, extend (1) to the case of countably or uncountably infinite n. But this is problematic. Uncountable additivity is clearly too much to ask in the standard system of analysis: if a real number is being picked randomly from the interval (0,1), then we will want $Pr(s) = 0$ for any elementary event or number $s \in (0,1)$, but $Pr(\,(0,1)\,) = 1$. Even countable additivity can be problematic. In general it is not possible to define probability over all subsets of an infinite S and get countable additivity. The definition must be restricted to subclasses of 2^S that are called σ-fields or σ-algebras. Such a structure must have the following properties: it should contain the empty set and the whole space, and unions and intersections of countable families of sets already in it. If our event space is the real line or an interval, the usual σ-field is constructed by taking countable unions, intersections, and complements of all intervals; this is called the Borel σ-field. We will denote the σ-field or class of subsets of our sample space S over which probabilities are defined by $\mathcal{A}$. Thus $\mathcal{A} \subset 2^S$, and if S is finite, we will usually take $\mathcal{A} = 2^S$.

To sum up, the foundations of our probability theory are the triple $(S, \mathcal{A}, Pr)$, where S is the sample space, $\mathcal{A}$ a σ-field over S, and a function $Pr : \mathcal{A} \mapsto R_+$ satisfying

1. $Pr(\emptyset) = 0$, $Pr(S) = 1$.
2. If $A_i \subset S$ for $i = 1, 2, \ldots$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr(A_i)\,. \tag{2}$$

Luckily for our purpose in this course, the details of such measure-theoretic foundations of probability are largely irrelevant. But some related ideas will occasionally crop up, and you will need them in more detail if you go on to do advanced courses such as rocket-science finance where the sample space of the stochastic processes under study consists of time-paths (functions) and the σ-field over which probabilities are defined must evolve in time as the process itself unfolds.

The mathematical structure as set out above is independent of any interpretation. But in each application one must assign probability numbers to events. How is this done? Here are some examples. [1] In some situations, the physics or other science of a situation gives us probabilities; radioactive decay is an example. Some would argue that in fact everything is deterministic, and probabilistic modeling is only an imperfect way of dealing with our inability to calculate a highly complex reality. We leave such debates about the true meaning of determinism or randomness to philosophers, who have a long tradition of arguing questions and never finding any answers. [2] In independently repeatable experiments we can assign probabilities by performing the experiment numerous times and observing frequencies. A similar method underlies most forecasting based on estimation of models. [3] Sometimes a "principle of insufficient reason" is invoked: if we have no good reason to believe that one event is more likely than another, then we regard them as equally likely. But this is fallible. For example, the three elementary outcomes of tossing two indistinguishable coins are not equally likely; it would be a mistake to forget the underlying process that makes "one head, one tail" twice as likely as either two heads or two tails. [4] How can we quantify an individual's purely subjective assessment that "one event is more likely than another"? If the individual has a complete ordering of the class of events, and if the class includes (or can be augmented to include by adding sequences of coin tosses) events that have objective probabilities $k/2^n$, then the probability $Pr(E)$ of any event E can be found to any desired level of accurancy by repeatedly asking the individual to make comparisons to establish bounds

$$(k-1)/2^n < Pr(E) \leq k/2^n \, .$$

This is basically Savage's theory of subjective probabilities. Kreps (1988, chapter 8) gives a good account.

3 Conditional Probabilities

Suppose we start with a sample space S, and then are told that a particular event $E \in \mathcal{A}$ with $\Pr(E) > 0$ has happened. So uncertainty is partially resolved: we know that the actual outcome is some elementary event in E, but do not know precisely which. We should now redefine probabilities taking into account of the fact that E has happened. These are called *conditional probabilities given* E, and are defined as

$$Pr(A|E) = \frac{Pr(A \cap E)}{Pr(E)} \tag{3}$$

(The definition is meaningless if $Pr(E) = 0$: then $Pr(A \cap E) = 0$ also (Why?), so the ratio is $0/0$. The definition remains valid but trivial if E is an elementary event so that its occurance resolves the uncertainty completely. We can restrict the definition of these probabilities just to sub-events of E, with the σ-field appropriately restricted. Or we could define these probabilities for all A in the original σ-field, and say that $Pr(A|E) = 0$ if $A \cap E = \emptyset$. For our applications it makes little difference which usage we adopt.)

Example: Consider the roll of a die, and suppose all outcomes are equally likely with probability $\frac{1}{6}$ each. Define E to be the event that the outcome is odd (1 or 3 or 5), so

$Pr(E) = \frac{1}{2}$. Let A be the event that the outcome is less than or equal to 3, so $Pr(A) = \frac{1}{2}$. Then $A \cap E = \{1, 3\}$, and $Pr(A \cap E) = \frac{1}{3}$. Therefore

$$Pr(A|E) = \frac{1/3}{1/2} = \frac{2}{3}\,.$$

Observe that $Pr(A|E) \neq Pr(A)$ in this example. That is because the knowledge that E has occurred conveys some extra information about A (here the occurrance of E tells us that the outcome cannot be 2. If such is not the case, so that the conditional probability of the second event equals its unconditional probability, we call the events independent. Formally, we define events A_1, A_2, $\ldots$ A_n to be *independent* if, for any selection of any k distinct events, say A_{i_1}, A_{i_2}, $\ldots$ A_{i_k} (including the case where $k = n$ so all are selected),

$$Pr\left(A_{i_1} \cap A_{i_2} \cap A_{i_3} \ldots \cap A_{i_k}\right) = Pr(A_{i_1})\,Pr(A_{i_2})\,Pr(A_{i_3})\,\ldots\,Pr(A_{i_k})\,. \tag{4}$$

Then (3) immediately shows that if A and B are independent events,

$$Pr(A|B) = \frac{Pr(A)\ Pr(B)}{Pr(B)} = Pr(A)\,.$$

More generally, for any $k+1$ independent events A_1, A_2, $\ldots$ A_k, A_{k+1},

$$Pr(A_{k+1}|A_1 \cap A_2 \cap \ldots A_k) = Pr(A_{k+1})$$

Observe that in making all these statements I omitted to say that all these events have to be in the σ-algebra $\mathcal{A}$. Since the σ-algebra structure is not of much importance in our applications, I will continue to avoid such pedantry from now on.

For applications in economics, perhaps the most important result about conditional probabilities is Bayes' theorem or formula. It enables us to extract information about the probabilities of some underlying events (possible causes?) by observing some other events (effects?). The situation is usually as follows. Suppose E_1, E_2, $\ldots$ E_m is a *partition* of S, that is, collection of mutually exclusive and exhaustive events:

$$E_i \cap E_j = \emptyset \ \text{ for } i \neq j; \qquad \bigcup_{i=1}^{n} E_i = S\,.$$

We know the probabilities $Pr(E_i)$, but don't get any direct information about the occurance or otherwise of these events. However, there is another family of mutually exclusive and exhaustive events A_1, A_2, $\ldots$ A_n. We know the conditional probabilities $Pr(A_j|E_i)$, and we do get to observe which of the A_j actually occurs. Then Bayes' formula gives us the "reverse probabilities"

$$Pr(E_i|A_j) = \frac{Pr(A_j|E_i)\ Pr(E_i)}{\sum_{k=1}^{m}\ Pr(A_j|E_k)\ Pr(E_k)}\,. \tag{5}$$

In many applications, the probabilities $Pr(E_i)$ are the ones initially held, which are then revised in the light of the information as to which of the A_j occurs. Therefore the $Pr(E_i)$ are called the *prior* probabilities and the relevant $Pr(E_i|A_j)$ the *posterior* probabilities.

To prove Bayes' formula, begin with the definition (3) to write

$$Pr(E_i|A_j) = \frac{Pr(E_i \cap A_j)}{Pr(A_j)} .$$

In the numerator, use the definition of the conditional probability in the other direction to write

$$Pr(E_i \cap A_j) = Pr(A_j \cap E_i) = Pr(A_j|E_i)\ Pr(E_i) .$$

Turning to the denominator, observe that the event A_j can be partitioned into mutually exclusive and exhaustive subevents:

$$A_j = \bigcup_{k=1}^{m} (A_j \cap E_k) ,$$

and therefore

$$Pr(A_j) = \sum_{k=1}^{n} Pr(A_j \cap E_k) = \sum_{k=1}^{n} Pr(A_j|E_k)\ Pr(E_k)$$

This completes the proof. Observe that I have used the symbol k for the index of summation to distinguish the E_k here from the E_i for a specific i in the numerator of Bayes' formula.

Example: Suppose the world consists of good guys and bad guys, and your prior is that the probability of a random person being good is 70%. The two types have different temptations to cheat you; a bad guy will cheat you 80% of the time and a good guy will cheat you only 10% of the time. You interact with a person who cheats you. What is your posterior probability that this person is bad?

Let E_1 = the person is bad, E_2 = the person is good. So your prior probabilities are $Pr(E_1) = 0.3$, $Pr(E_2) = 0.7$. Let A_1 = you get cheated, A_2 = you are not cheated. The conditional probabilities are

$$Pr(A_1|E_1) = 0.8, \quad Pr(A_2|E_1) = 0.2, \qquad Pr(A_1|E_2) = 0.1, \quad Pr(A_2|E_2) = 0.9 .$$

Then by Bayes' formula the required posterior probability is

$$\begin{aligned} Pr(E_1|A_1) &= \frac{Pr(A_1|E_1)\ Pr(E_1)}{Pr(A_1|E_1)\ Pr(E_1) + Pr(A_1|E_2)\ Pr(E_2)} \\ &= \frac{0.8\ *\ 0.3}{0.8\ *\ 0.3 + 0.1\ * 0.7} = \frac{0.24}{0.24 + 0.07} = \frac{0.24}{0.31} = 0.774 \end{aligned}$$

I usually find it convenient to display this calculation in a matrix where the rows are the unobservable E_i, the columns are the observable A_j, the cells are the intersections $E_i \cap A_j$, and the cell entries are the probabilities. Then, once an A_j is observed, we are restricted to that column. The conditional probabilities of the various E_i are the ratios of the probabilities in each row divided by the sum of the rows in that column:

		Observables (effects)	
		A_1 (cheated)	A_2 (not cheated)
Unobserved	E_1 (bad, prior prob 0.3)	$0.3 * 0.8 = 0.24$	$0.3 * 0.2 = 0.06$
Causes	E_2 (good, prior prob. 0.7)	$0.7 * 0.1 = 0.07$	$0.7 * 0.9 = 0.63$
	Sum over rows	0.31	0.69

In this example, if you were not cheated, you would revise the probability of the person being bad down to $0.06/0.69 = 0.087$.

Such calculations will appear often when solving games with asymmetric information. There the conditional probabilities $Pr(A_j|E_i)$ will be the (mixed) strategies of the types of players, and their equilibrium values will of course have to be found as a part of the solution.

4 Random Variables

Sample spaces in general can be quite abstract or can consist of objects like cards in a deck. But often, and especially in economic applications, numerical magnitudes are associated with events. To study these mathematically, we define the concept of a *random variable* as a real valued function on a sample space, $X : S \mapsto R$. Thus a random variable is actually neither random nor variable; it is just a function.

Example: The sample space has two elementary events, "earthquake in LA" and "no earthquake in LA," and the random variable maps each event to the aggregate of property values in LA in that event.

Given a random variable X, we define its *(cumulative) distribution function* (CDF): for any real number t, this takes the value equal to the probability that the value of the random variable is less than or equal to t. Symbolically, the CDF $F : R \mapsto [0, 1]$ is defined by the rule

$$F(t) = Pr(X^{-1}(-\infty, t]) \tag{6}$$

(For this to be meaningful, the set in the sample space that is mapped to $(-\infty, t]$ by the random variable X, namely the set of preimages $X^{-1}(-\infty, t]$), has to be in the σ-field over which probabilities are defined; for this, X has to be what is called a measurable function. But in our applications in this course we can disregard this mathematical complexity.)

The CDF of any random variable must be non-decreasing (prove this). But it may be flat over some ranges of R. If S is finite, then X can take on only a finite number of values, and its CDF will be flat between the values with jumps at these values. Even if S is an uncountably infinite continuum, X may have gaps in the values it takes and may take some real values with positive probability, so the CDF may have flats and/or jumps.

If the CDF is differentiable, its derivative function $f(t) = F'(t)$ is called the *probability density function* (PDF) of the random variable X. This definition can be generalized by allowing suitably infinite derivatives for step functions (Dirac Delta functions); we may occasionally do this rather heuristically.

The *support* of the distribution of a random variable is the subset of the real line corresponding to the values the variable can take with positive probability or density. In most of our applications, we will find that the value $F(t)$ of the CDF is zero over an interval $(-\infty, t_L)$, then it increases either continuously or in jumps, finally reaching 1 at t_H and then staying there over (t_H, ∞). Then $[t_L, t_H]$ is the support. In special cases we may get $t_L = -\infty$ and/or $t_H = \infty$. More generally, mathematical complications arise in rigorous definition and treatment of the concept of the support. We do not need these in our economic applications, therefore I will omit this.

The *(mathematical) expectation* or *expected value* of a random variable X over a finite sample space S is defined as

$$\mathrm{E}[X] = \sum_{s \in S} X(s)\ Pr(s)\,.$$

Important – "expected value" has no connotation of anticipation or entitlement. It is just a mathematical term. Note that the expected value is a single number.

For an infinite sample space, we can define a corresponding integral

$$\mathrm{E}[X] = \int_{s \in S} X(s)\ Pr(ds)\,,$$

but a more convenient representation is by means of the CDF or the PDF F of X,

$$\mathrm{E}[X] = \int_{t_L}^{t_H} t\ dF(t) = \int_{t_L}^{t_H} t\ f(t)\ dt\,.$$

If the CDF has jumps, the first form of the integral has to be defined and calculated appropriately; we won't go into the general theory of this but will explain it in specific examples when (if) the issue arises.

Write $\mathrm{E}[X] = \mu$ for short; then the variance of X is defined as

$$\mathrm{V}[X] = \mathrm{E}[\ (X-\mu)^2\ ]\,.$$

Prove that

$$\mathrm{V}[X] = \mathrm{E}[X^2] - \mu^2\,.$$

Other moments are similarly defined. Some further formulas for expectations of other functions of a random variable (e.g. exponential) and for CDFs, PDFs, means, variances etc. of particular random variables (negative exponential, normal, etc.) will be needed from time to time. You should know most of these from your probability and statistics courses, and we will develop others as needed.

More reminders or wake-up calls for matters from probability theory and statistics will appear in the first problem set.

5 Further Reading

The only required reading for this background is your textbook in your prerequisite statistics course, ECO202 (old ECO200) or ORF245. For those interested, here is some more

Feller, William. 1968. *An Introduction to Probability Theory and Its Applications: Volume I.* Third Edition. New York: Wiley.

The Introduction chapter of this gives an outstanding discussion of the conceptual foundations. Chapters I, V, and IX cover the above material for finite sample spaces.

Billingsley, Patrick. 1986. *Probability and Measure.* Second Edition. New York: Wiley.

This gives the rigorous general theory for arbitrary sample spaces.

If you want to find out more about the Savage approach to subjective probability, read

Kreps, David. 1988. *Notes on the Theory of Choice.* Boulder, CO: Westview Press. Chapter 8.