Economics 520, Fall 2006

Lecture Note 1: Elementary Probability Theory and Combinatorics[1]

You should already be familiar with basic set-theoretic notation and operations, such as unions, intersections, complementation, empty set.

Definition 1 *The sample space (denoted by Ω) is the set of all possible outcomes of an experiment.*

Example 1: Tossing a die. There are six outcomes in the sample space, corresponding to the number on top of the die, so we can take $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Definition 2 *An event (denoted by E) is a collection of possible outcomes of an experiment, that is, a subset of the sample space.*

Example 1 continued: Possible events include "an odd number", $E_1 = \{1, 3, 5\}$, "an even number", $E_2 = \{2, 4, 6\}$, or "a number less than 3", $E_3 = \{1, 2\}$.

Definition 3 *Two events E_1 and E_2 are disjoint if their intersection $E_1 \cap E_2$ is equal to the empty set $\emptyset$.*

Example 1 continued: E_1 and E_2 are disjoint because their intersection is empty, but E_1 and E_3 are not disjoint because their intersection is $\{1\}$.

Definition 4 *If the sets $E_1, E_2, \ldots$ are pairwise disjoint and their union $\cup_i E_i$ is equal to the sample space, the collection $E_1, E_2, \ldots$ forms a partition of the sample space.*

Example 1 continued: E_1 and E_2 form a partition: they are disjoint and their union is the entire sample space Ω.

Example 2: Relative humidity on a randomly selected day. In this case we might take $\Omega = [0, 1]$, the unit interval. (In practice, the relevant quantity might only be measured with finite precision, but it is often a convenient fiction to suppose that the quantity can take on a continuum of values.) The events

$$E_1 = [0, .1), E_2 = [.1, .7), E_3 = [.7, 1]$$

form a partition of the sample space. Of course there are many other possible ways to partition the unit interval.

Loosely speaking, a probability distribution assigns probabilities between 0 and 1 to different possible events. In Example 1, a fair die would assign probability 1/6 to the event $\{1\}$ which corresponds to rolling a "1." So we want to define a set of events over which probabilities will be assigned:

[1] This note and many of the later lecture notes are based on notes written by Guido Imbens. I thank him for permission to use his material in this course.

Definition 5 *A collection $\mathcal{B}$ of subsets of Ω is a sigma-algebra*[2] *if it satisfies the following three conditions:*

1. *The empty set $\emptyset$ is contained in $\mathcal{B}$.*

2. *If $E \in \mathcal{B}$ then its complement $E^c = \{\omega \in \Omega | \omega \notin E\}$ is also in $\mathcal{B}$.*

3. *$\mathcal{B}$ is closed under countable unions, that is, if $E_1, E_2, \ldots$ are all in $\mathcal{B}$, then so is $\cup_i E_i$.*

One possible sigma-algebra (in fact the smallest possible one) is $\mathcal{B}_1 = \{\emptyset, \Omega\}$. This works for any Ω, but this is not a very interesting sigma-algebra.

For Example 1, a possible sigma-algebra is $\mathcal{B}_2 = \{\emptyset, \Omega, \{1\}, \{2, 3, 4, 5, 6\}\}$. You should verify that this satisfies the three conditions of the definition.

Another possible sigma-algebra for Example 1 is the power set, the set of all subsets of Ω.

For a given sample space, we'd like the probability distribution to assign probabilities to as many different events as possible. In Example 1, it is possible to assign probabilities consistently to every member of the power set. For example, with a fair die the probability of $E = \{1, 2\}$ is equal to the probability of $\{1\}$ plus the probability of $\{2\}$, i.e. $1/6 + 1/6 = 1/3$.

However, in Example 2 when $\Omega = [0, 1]$, it turns out that there is no way to do this for *every* possible subset of Ω. In this case, the power set is so large that we have to limit our attention to a large, but technically more manageable, collection of sets. (For more on this, see Billingsley, P., *Probability and Measure*, p.45).

Definition 6 *(Kolmogorov Axioms) Given a sample space Ω and an associated sigma-algebra $\mathcal{B}$, a probability function is a function P from $\mathcal{B}$ to the real line satisfying:*

1. *(Nonnegativity) For all $E \in \mathcal{B}$, $P(E) \geq 0$.*

2. *(Unit probability for the sample space) $P(\Omega) = 1$.*

3. *(Additivity of probability of disjoint sets) If $E_1, E_2, \ldots$ are pairwise disjoint, then $P(\cup_i E_i) = \sum_i P(E_i)$.*

Remarks:

1. The empty set $\emptyset$ is contained in $\mathcal{B}$ and therefore its complement $\Omega = \emptyset^c$ is also contained in $\mathcal{B}$. Hence the second condition is well defined.

2. If the Borel field is finite, Condition 3 need only hold for finite unions.

[2] Note: CB uses the term Borel field interchangeably with sigma-algebra. However, some mathematics texts use Borel field more narrowly, to refer to a certain type of sigma-algebra that is generated by the Borel topology.

Example 1 continued: For a fair die,

$$P(E) = \frac{\text{number of outcomes in } E}{\text{total number of outcomes in } \Omega}.$$

This assigns probability 1/6 to each of the six outcomes. Alternatively we can assign any other nonnegative number to each of the six outcomes provided they add up to one.

An immediate implication of the Kolmogorov axioms is that

$$P(E^c) = 1 - P(E),$$

because

$$1 = P(\Omega) = P(E) + P(E^c).$$

Therefore:

$$P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 0.$$

Another useful result: for any events E_1 and E_2,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

The proof, typical for this type of result, relies on creating pairwise disjoint sets for which one can add up the probabities by the third axiom:

$$\begin{aligned} P(E_1 \cup E_2) &= P\Big((E_1 \cap E_2^c) \cup (E_1^c \cap E_2) \cup (E_1 \cap E_2)\Big) \\ &= P(E_1 \cap E_2^c) + P(E_1^c \cap E_2) + P(E_1 \cap E_2). \qquad (1) \end{aligned}$$

Also:

$$P(E_1) = P(E_1 \cap E_2) + P(E_1 \cap E_2^c),$$

which, after rearranging, gives

$$P(E_1 \cap E_2^c) = P(E_1) - P(E_1 \cap E_2),$$

which after substituting in (1) gives the desired result.

You should read CB 1.2.2 for further results of this type, which help in calculating probabilities for complicated events.

Counting and Probability

Early problems in the history of probability often involved games of chance where the probabilities for basic outcomes were clear but the probabilities of interesting events were difficult to calculate because of the large number of basic outcomes for events of interest. A number of these problems can be formulated as problems of drawing k objects with and without replacement out of a set of n while being or not being concerned with the ordering. Solving them requires counting the ways in which you can do this. As an example we consider the case where we have $n = 4$ objects, labelled A, B, C, and D, and wish to draw $k = 2$. Recall that $n! = n \times (n-1) \times (n-2) \times \cdots \times 1$ is called <u>n factorial</u>.

Result 1 *(ordered, with replacement) The total number of ways k objects can be drawn out of a set of n with replacement is n^k.*

For the first draw there are n choices, for the second one there are again n choices and so on. In the example, the set of outcomes is $\{AA, AB, AC, AD, BA, BB, BC, BD, CA, CB, CC, CD, DA, DB, DC, DD\}$, with sixteen elements.

Result 2 *(ordered, without replacement) The total number of ways k objects can be drawn out of a set of n without replacement is $n \times (n-1) \times (n-2) \times \ldots \times (n-k+1) = n!/(n-k)!$.*

For the first draw there are n choices, for the second one there are $n-1$ choices and so on. In the example the set of outcomes is $\{AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC\}$, with twelve elements.

Result 3 *(unordered, without replacement) The total number of ways k objects can be drawn out of a set of n without replacement is $n!/(k!(n-k)!)$.*

The total number of ways k objects can be drawn out of a set of n *with* replacement is $n \times (n-1) \times (n-2) \times \ldots \times (n-k+1) = n!/k!$. If we do not care about the ordering, we have to take account of the number of different ways we can order k objects. This is $k!$, by Result 2, so we have to divide this into the $n!/(n-k)!$ to get

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}.$$

The left hand side notation is read as "n choose k." In our example, the set of outcomes is $\{AB, AC, AD, BC, BD, CD\}$, with six elements.

Example 3: How many bridge hands of 13 cards can be formed from a 52-card deck? Since the order does not matter, Result 3 is relevant. The number is

$$\binom{52}{13} = \frac{52!}{13!(52-13)!} = 635,013,559,600.$$

Result 4 *(unordered, with replacement) The total number of ways k objects can be drawn out of a set of n with replacement is $(n+k-1)!/(k!(n-1)!)$.*

This one is messier than the others. Reformulate the problem as follows: put k objects in n bins, allowing for more than one object per bin. We can describe the result by the k numbers or by a sequence of $n-1$ zeros and k ones, where a zero indicates that we have finished with one of the n bins, and one indicates one of the objects in that particular bin. For example, with $n = 3$ bins and $k = 2$ objects we could have the following outcome: (2,3) which would be coded as 0101: 0 (because the first bin is empty), 1 (because there is an object in the second bin), 0 (because there is no additional object in the second bin), 1 because there is an object in the third bin. We do not record the last 0 because it would always end in a zero. In the same example (2,2) would be coded as 0110: 0 because the first bin is empty, 1 because there is an object in the second bin, 1 because there is a second object in this bin, 0 because there is no additional object in this bin.

Now the problem is one of choosing a set of k ones out of a set of $n+k-1$ which can be done in $(n+k-1)!/(k!(n-1)!)$ different ways.

In the example the set is $\{AA, AB, AC, AD, BB, BC, BD, CC, CD, DD\}$, with ten elements.

Example 4: Isaac Newton and Samuel Pepys debated the following question: is the probability of tossing at least one six in six tosses with a fair die smaller than, equal to, or larger than the probability of tossing at least two sixes in twelve tosses?

Consider the probability of the first event. There are 6^6 different outcomes for the six tosses. Each has probability $1/6^6$. The question is how many of the 6^6 outcomes are favorable, that is, how many have at least one six. It is easier to answer the opposite: how many outcomes have no six at all. There are five possibilities for each toss in that case, so 5^6 have no six. Hence the probability of at least one six is

$$Pr(\text{at least one six}) = 1 - Pr(\text{no six}) = 1 - 5^6/6^6 \approx 0.665.$$

Consider the probability of the second event. There are 6^{12} different outcomes, again each with the same probability $1/6^{12}$. How many have at least two sixes is 6^{12} minus the number that have at most one six. At most one six is either no six or exactly one six. The number of outcomes with no six is 5^{12}. The event of exactly one six can be partitioned into twelve events depending on the location of the six. The number of outcomes with a six in the first toss is 5^{11}. Hence the number of outcomes with a single six is $12 \cdot 5^{11}$. Hence the probability of at least two sixes is

$$Pr(\text{at least two sixes}) = 1 - Pr(\text{no six}) - Pr(\text{one six})$$

$$= 1 - 5^{12}/6^{12} - 12 \cdot 5^{11}/6^{12} \approx 0.619.$$

Hence, at least two sixes in twelve tosses is less likely than at least one six in six tosses.

Example 5: Suppose there are 25 people with dogs at a park, and suppose they all choose randomly from a set of n breeds of dogs. How large should n be to ensure the chances of everybody having a different breed of dog is at least 0.5? (As n increases this probability clearly goes up.)

The number of ways 25 people can choose from n types of dogs is n^{25}. All these outcomes are equally likely, with probability $1/n^{25}$. The number of ways they can pick with a different dog breed for everybody is $n!/(n-25)!$. Now we want to know n such that

$$\frac{n!/(n-25)!}{n^{25}} \geq 0.5, \qquad \text{and} \quad \frac{(n-1)!/(n-1-25)!}{(n-1)^{25}} < 0.5.$$

At $n = 442$:

$$\frac{n!/(n-25)!}{n^{25}} = 0.5008,$$

and

$$\frac{(n-1)!/(n-1-25)!}{(n-1)^{25}} = 0.49996.$$

Hence $n = 442$ is the solution.

Further Reading

The axiomatic development of probability theory initiated by Kolmogorov is built upon measure theory. We will not go into measure theory much further in this class, but if you are planning to do research in econometric theory or some aspects of micro theory, some exposure to measure-theoretic probability may be helpful. Some good texts include:

- Billingsley, P., *Probability and Measure*, 3rd ed., Wiley.
- Dudley, R., *Real Analysis and Probability*, Cambridge University Press.
- Pollard, D., *A User's Guide to Measure-Theoretic Probability*, Cambridge University Press.

The Billingsley book is a classic reference. The book by Dudley is excellent and very modern in its treatment, but very advanced.

If you are interested in applications of probability theory to games of chance and other recreations, you might look at:

- Epstein, R. A., *The Theory of Gambling and Statistical Logic*, Academic Press.
- Mosteller, H., *Fifty Challenging Problems in Probability with Solutions*, Dover.