Probability and Statistics

Fall 2000

Lect. 1: Elementary Probability Theory and Combinatorics (CB 1.1–1.2)

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

As an example consider the experiment of tossing a die. There are six outcomes in the sample space, corresponding to the number on top of the die (ruling out the possibility that it lands on a point), or $\Omega = \{1, 2, 3, 4, 5, 6\}$.

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

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 $\emptyset$.*

In this example 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 $\cup_i$ is equal to the sample space, the collection $E_1, E_2, \ldots$ forms a partition of the sample space.*

Definition 5 *A collection $\mathcal{B}$ of subsets of Ω is a Borel field if it satisfies the following three conditions:*

1. *The empty set 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 Borel field (in fact the smallest possible Borel field) is $\mathcal{B}_1 = \{\emptyset, \Omega\}$. This always works, that is in any experiment, but this is not a very interesting Borel field. Another Borel field in the die example is $\mathcal{B}_2 = \{\emptyset, \Omega, \{1\}, \{2,3,4,5,6\}\}$. The most natural Borel field in this example is the set of all subsets of the sample space. Although the set of all subsets is always a Borel field, it is not always convenient for defining probability measures. For an example of this see Billingsley, *Probability and Measure*, page 41.

Definition 6 *(Kolmogorov Axioms)*
Given a sample space Ω and an associated Borel field $\mathcal{B}$, 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. *(Addivity of probability of disjoint sets) If $E_1, E_2, \ldots$ are pairwise disjoint, then $P(\cup_i E_i) = \sum_i P(E_i)$.*

Two remarks:
(i) 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.
(ii) If the Borel field is finite, Condition 3 need only hold for finite unions.

A natural choice for the probability function in the die example is

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

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:

$$P(\emptyset) = 0.$$

More indirect is the following result:

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

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:

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

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.

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.

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.

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.

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

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}.$$

Result 4 *(unordered, with replacement) The total number of ways k objects can be drawn out of a set of n without replacement is $(n+k-1)!/(k!(n-1)!)$.*

This one is messier than the others. Reformulate the problem (and this is a common solution technique: reformulate the problem into one that is solvable) 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.

Now let us consider a more interesting, historical, example, one that was the subject of a debate between Newton and Pepys in the 1800's. 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.