MATH 3070 Introduction to Probability and Statistics
Lecture notes
Probability

Objectives:

1. Learn the basic concepts of probability
2. Learn the basic vocabulary for probability
3. Identify the sample space and outcomes for experiments
4. Identity simple and compound events
5. Differentiate between independent and mutually exclusive events
6. Identify random variables and derive probability distributions
7. Learn the Law of Large Numbers

Basic Concepts of Probability

Things happen. Can we make a reasonable estimation as to how and when?

Probability is **empirical**, that is, it is based on observation. The probability that an event will occur is the relative frequency with which that event can be expected to occur. **Relative frequency** is the observed occurrences of an event. If we assume that the event is random then individual outcomes are uncertain, but there is nonetheless a regular distribution of outcomes in a large number of repetitions.

But what gives this any validity? The fact is, chance behavior is unpredictable in the short run, but has a regular and predictable pattern in the long run. So we can call a phenomenon random if the individual outcomes are uncertain, but a regular distribution appears with a large number of repetitions.

The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.

Vocabulary of Probability

Experiment Any process that yields a result or an observation. A process that leads to well-defined results called outcomes.

Outcome A particular result of an experiment. The result of a single trial of an experiment.

Sample Space The set of all possible outcomes of an experiment $\{S\}$. Each element in the set is a sample point.

Event Any subset of the sample space. Can be one or more outcomes of the same experiment.

Probability Model A mathematical description of a random phenomenon consisting of a sample space (S) and a way of assigning probabilities to events.

Complement The opposite of an event, or, the set of all sample points in the sample space that do not belong to event A. We denote this as "A-bar" $\bar{A}$.

Note: Outcomes must not overlap (mutually exclusive) and there cannot be any exceptions (inclusive).

Examples of experiments and sample spaces (from Bluman, p. 142):

Experiment	Sample Space
Toss one coin	Head, Tail
Roll a die	1,2,3,4,5,6
Answer a true-false question	True, False
Toss two coins	Head-head, Tail-tail, Head-Tail, Tail-Head

We can represent the sample space in several ways: tree diagram, set of ordered tuples, a list, a grid, etc.

The outcomes can be affected by whether the experiment is conducted **with replacement** or **without replacement**. If you select an item and return it so that it may be selected again that is **with replacement**. If you do not return the item such that it can only be selected once, that is **without replacement**.

In a sample space containing sample points that are equally likely to occur, the probability $P(A)$ of an event A is the ratio of the number $n(A)$ of points that satisfy the definition of even A to the number $n(S)$ of sample points in the entire sample space:

$$P(A) = \frac{\text{(number of times } A \text{ occurs in sample space)}}{\text{(number of elements in } S)}$$

or said another way

$$P(A) = \frac{\text{(number of ways } A \text{ can occur)}}{\text{(total number of outcomes in the sample space)}}$$

Rules of Probability

There are some rules to Probability:

1. All probabilities must be between 0 and 1, inclusive.

$$0 \leq P(A) \leq 1 \text{ for any event } A$$

2. All probabilities must add up to 1.

3. The probability that an event does not occur is 1 minus the probability that the event does occur $(1 - P(A))$.

4. If events are mutually exclusive, the probability that one of them happens is the sum of their individual probabilities. (Also known as the **addition rule for disjoint events**.

So let's look at an example of an experiment and the assignment of probabilities to each of the outcomes. This example is from *A First Course in Statistics*, by McClave and Dietrich (p. 105).

A fair die is tossed, and the up face is observed. If the face is even, you win $1. Otherwise, you lose $1. What is the probability you win?

Recall that the sample space for this experiment contains six outcomes:

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

Since the die is balanced (fair), we assign a probability of $\frac{1}{6}$ to each of the outcomes in this sample space. An even number will occur if one of the outcomes 2, 4, or 6, occurs. A collection of outcomes such as this is called an event and we denote this event by the letter A. Since the event A (an even number is observed) contains three outcomes, all with the probability of $\frac{1}{6}$ and since no outcomes can occur simultaneously, we reason that the probability of A is the sum of the probabilities of the outcomes in A. Thus, the probability of A is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$. This implies that in the long run you will win \$1 half the time and lose \$ half the time.

Random Variables and Probability Distributions

The outcome of any trial (or experiment) can take on any of the possible values in the sample space. In an experiment where we flip two coins and record the results, we could have any of the following: HH, HT, TH, or TT. When we roll a single fair die the results could be any number between 1 and 6, inclusive. Since the values can change from any given trial to any other this makes the results a variable. And since we don't know what the value might be that makes it random. The "box" that holds this random value is called a **random variable**. The value in this box depends on chance and can be any of the values possible. These values could be either integer (whole) values or values such as "heads" or "tails". In these cases we call these values **discrete** and further classify the variables as discrete random variables. If the values could be more precisely defined to a more granular degree (written in real, or floating point, notation, with a decimal point) we call them continuous random variables. For our purposes we will concentrate on the discrete random variables (**d.r.v.**) for now. Some examples of d.r.v. are

- The number of cars through the GA 400 toll booth in one day.
- Passengers through Hartsfield-Jackson Airport during Thanksgiving.
- The score on the math portion of the SAT or ACT.

Each value in a random variable has a portion of the total probability assigned to it. We know the total has to be 1 when we sum all the probabilities and each share of the total probability is between zero and one. The assignment of probability to each value is referred to as the **probability distribution**. This can be represented as either a function with output values or as a table.

> it Probability distribution: A listing of the possible values and corresponding probabilities of a discrete random variable; or a formula for the probabilities. (Weiss, p 291)

To construct a probability distribution first identify the values for the discrete random variable. Then compute the probabilities for each of those values. Then construct a table with the values for the d.r.v. and the corresponding probability.

For example, if X is the d.r.v. in an experiment involving rolling one fair die, the possible values for the d.r.v. are 1,2,3,4,5, and 6. Since the die is fair we know each value is equiprobable (has equal probability) so we can simply divide the total probability by six to compute the amount of probability to assign to each value ($\frac{1}{6}$). Constructing our table we get something like this:

roll of die X	1	2	3	4	5	6
probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Of course, the probability does not have to be evenly distributed. If we look at students in high school, grades 9 through 12, and treat the grade level as a random variable (X) then we could construct a probability distribution that looks like this

grade level X	9	10	11	12
probability	0.300	0.262	0.230	0.208

Each of these probabilities is between zero and one and the sum is $.300 + .262 + .230 + .208 = 1.0$. This is a valid probability distribution.

Law of Large Numbers

Empirical, or experimental, probability is the observed relative frequency with which an event occurs. The value assigned to the probability of event A as a result of experimentation is:

$P\prime(A)$ = (number of times A occurred) / (number of trials)

We use $P\prime$ (P prime) to denote empirical probability.

The more experiments we run the larger the set of results we collect. The larger the set of results the closer to the "true" probability of occurrence. This leads us to the law of large numbers which states:

> As the number of times an experiment is repeated increases, the ratio of the number of successful occurrences to the number of trials will tend to approach the theoretical probability of the outcome for an individual trial.

That is, the larger the number of experimental trials n, the closer the empirical (observed) probability $P\prime(A)$ is expected to be to the true or theoretical probability $P(A)$.

Compound events and probability

We have talked about instances where an item must belong to one, and only one, group. These are called **mutually exclusive**. Events, outcomes of an experiment, can also be mutually exclusive, that is, an event can be defined in such a way that the occurrence of one event precludes the occurrence of any of the other events. (JK) These can also be called **disjoint events**. (DM)

There are three possible compound event outcomes:

1. Either event A or event B will occur, $P(A \text{ or } B)$.
2. Both event A and event B will occur, $P(A \text{ and } B)$.
3. Event A will occur given that event B has occurred, $P(A \mid B)$.

Each has a different rule applied to it for various reasons.

An example: Roll two dice and look at the sum of the numbers. Let there be three outcomes of interest:

A The sum is seven.
B The sum is ten.
C The two numbers are the same.

Clearly events **B** and **C** can happen at the same time since ten can be the sum of five and five. But **A** and **C** cannot happen together since seven is not evenly divisible by two. (JK)

We have special rules for computing the probability of mutually exclusive events. These are the **General Addition Rule** and the **Special Addition Rule**.

The **General Addition Rule** states:

Let A and B be two events defined in a sample space S. Then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

The **Special Addition Rule** states:

Let A and B be two events defined in a sample space S. If A and B are mutually exclusive events, then

$$P(A \text{ or } B) = P(A) + P(B)$$

Its simple, really. If two (or more) events are mutually exclusive then there cannot be overlap. Double counting cannot occur. We can then add the probabilities together to determine the probability.

Addition is not always the answer, though.

Consider the experiment of flipping a coin. Its been shown that the probability of either side of a coin resulting from a flip is one-half (0.5). If we toss a coin twice, the probability does not change because we have flipped the coin previously. The probability of the second toss is still 0.5. The probability of the same face twice in a row would then be 0.25, or 0.5×0.5. This is because the two events are independent.

Two events A and B are independent events is and only if the occurrence (or nonoccurrence) of one does not affect the probability assigned to the occurrence of the other. If this can be proven or reasonably assumed, then we can use the Special Multiplication Rule to calculate probability, which is

$$P(A \text{ and } B) = P(A)P(B)$$

If one event can affect the outcome of another, the events are dependent. And example would be event A being "sum of 10" on a roll of two dice and event B being "doubles." These two events are dependent in nature since there is only one outcome that satisfies both, (5,5). (JK)

Another example of dependence is the color of a card dealt from a standard deck. There are 52 cards, 26 each of black and red. If the first card dealt is red, then the probability of the next card being red has been changed from 26/52 (0.5) to 25/51 (0.49).

If we want to know the probability of an event occurring given that a previous event occurred, then we have conditional probability which we write as $P(A \mid B)$, the probability that A will occur given B.

Conditional probability is computed by taking the number of elements in the intersection of the two events (A and B) and dividing by the number of elements in the contingency (B). That is,

$$P(A \mid B) = P(A \text{ and } B)/P(B)$$

If there is dependence between the events, the computation of the events in sequence is determined by the General Multiplicative Rule, which states

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$