Chapter 4: Probability

4.1: Randomness

If you toss a coin (or even if you draw an SRS), it is impossible to predict the outcome in advance because the results will vary each time you run a trial.

Probabilities only describe what happens in the long term. While it is tempting to predict probabilities based on short-term outcomes, these short-term outcomes are often inaccurate in describing the actual long-term probability. Back to our coin-tossing example: In the long run, we know that half of the trials will produce heads and half of the trials will produce tails. However, if you only toss the coin 10 times, the results may not be five heads and five tails.

Important Terms

- A phenomenon is *random* if you cannot predict individual outcomes, but the individual outcomes still make up a regular distribution over many trials.
- The probability of an outcome of a random phenomenon is the proportion of times the outcome will occur, long term.
- Going back to the coin example: We know that the probability of the coin landing on heads or tails is 50%, or 0.5, however in the real world, coins have minor imperfections that make the probabilities of it landing on each side when flipped, a little bit more or a little bit less than 0.5. A coin is *fair* if the probability of it landing on either side is exactly 0.5.

Randomness

There are several important things to remember about randomness:

- For something to be random, it must be observed over many *independent* trials. This means that the results of one trial do not affect results from another trial.
- Probabilities are empirical. Simulations start with a given probability and imitate random trials, but we can only estimate a probability of everyday life by observing many repetitions of the same act.
- Having said that, simulations are still very useful because running many trials manually is not an effective use of resources.

4.2 Probability Models

A probability model is the description of a random phenomenon in mathematical terms. A model always includes:

- A list of all possible outcomes (for a coin: heads, tails.)
- The probability of each outcome (heads and tails both have a chance of 0.5.)

A sample space (S) of a random phenomenon is the set of all possible outcomes. For a coin, this would be: S = (head, tail).

An event is an individual outcome or set of outcomes for a random phenomenon. An event is a part of a sample space. The probability of a coin landing on heads exactly two out of four times is an example of an event. Therefore, the representation of this event, A, is:

A = (HHTT, HTHT, HTTH, THHT, THTH, TTHH)

Facts About Probability

- Probabilities are always between 0 and 1. If the probability of an event is 0, it means that this event never occurs. If the probability of an event is 1, the event always occurs. An event with probability 0.5 will occur half the time.
- The sum of the probabilities of all possible outcomes is 1.
- If two events do not have a common outcome, the probability that one of the two events will occur is the sum of their individual probabilities.
- The probability that an event does not occur is 1 minus the probability that the event does occur.

Rules About Probability

There are several rules to follow when dealing with probabilities:

- $0 \leq P(A) \leq 1$, where A is the event and P(A) is the probability that that event will happen. This means that the probability is between 0 and 1.
- If S is the sample space, then: P (S) = 1.
- Two events A and B are *disjoint* if they have no common outcomes. Events that are disjoint cannot coexist. In these cases, P (A or B) = P (A) + P (B). This is sometimes called the *addition rule for disjoint events*.
- The *complement* of any event, A, is the probability that the event will not occur. This is written as A^c. $P(A^c) = 1 - P(A)$

Probabilities in Finite Space

A *finite space* is where there are only a limited number of possible outcomes, and all possible events are disjoint. In finite spaces, the previously discussed rules apply:

- Each individual outcome has a probability between 0 and 1.
- Since the events in a finite space are disjoint, the sum of the probabilities of each individual event is 1.

Equally Likely Outcomes

Sometimes, two events have an equal chance of occurring. While we must always observe many trials in order to determine the probabilities of certain events, some phenomena have an element of balance that allows us to assume equal chances of each outcome happening. For example, in a coin toss, we assume that there is a probability of 0.5 that the coin will land on heads or tails because of the physical properties of a coin - a coin has two sides and it is equally likely for the coin to land on either side.

In assigning probabilities for events that are equally likely to occur, the following applies:

$$P(A) = \frac{\text{number of outcomes in A}}{k}$$

where k is the total number of outcomes in the space.

Independence and the Multiplication Rule

Two events are considered *independent* if the occurrence of one does not change the probability that the other will occur. If event A and event B are independent,

P(A and B) = P (A) P (B).

This is known as the *multiplication rule for independent events.*

4.3 Random Variables

Sample Spaces do not necessarily have to consist of numbers. If you toss a coin 4 times, for example, you could write the outcome as a string of four letters (eg: HTTH). In statistics, it is more interesting and applicable to look at numerical outcomes, for example, how many times the coin landed on 'heads" out of 4 tosses.

- A *random variable* has a numerical value associated with a random phenomenon. It is random because its values vary with repeated trials.
- Random variables are often written in *capital letters*, such as X or Y.

Discrete Random Variables

A variable is *discrete* if it has numerical outcomes. The *probability distribution* of discrete variable X lists the values and probabilities of each possible value. The values of X can be written as $x_1, x_2, x_3 \ldots x_{k'}$ and their corresponding probabilities as $p_1, p_2, p_3 \ldots p_k$.

It is important to remember that the rules of probability still apply in this case: p_k must be between 0 and 1, and $p_1 + p_2 + ... + p_k = 1$.

Continuous Random Variables

A random variable is continuous if it can take an infinite amount of values. The *probability distribution* of continuous random variables is described by a *density curve*. The probability of the occurrence of an event is the *area under the curve* and above the values of X associated with the event. Using this method, probabilities are assigned to intervals of outcomes, rather than individual outcomes themselves. Individual outcomes are automatically assigned a probability of 0. The density curve that is most frequently used for continuing random variables is the *Normal Distribution*. Normal distributions are probability distributions. If X has a distribution of N (μ, σ), the standardized variable is then: $z = (x-\mu) / \sigma$. This standardized variable has a mean of 0 and a standard deviation of 1: N (0,1). *See Ch. 1.4 for more information on Normal Distributions.*

4.4 Means and Variances of Random Variables

Mean

The mean of a probability distribution is denoted by μ. To remind ourselves that we are talking about the average of X (and not the average of the population), we use the notation μ_x. Sometimes the mean in this context is also referred to as the *expected value* of X. This term can be misleading, as a value of X does not necessarily have to be close to the mean of X.

To find the mean of a discrete random variable, multiply each possible value by its probability and sum the products:

$\mu_x = x_1p_1 + x_2p_2 + ... + x_kp_k$

$\mu_x = \sum x_ip_i$

The Law of Large Numbers

The law of large numbers states that if the number of observations increases, the mean of the observed values, x-bar, will approach the mean of the population μ.

How large a sample should be is not an easy question to answer. The number of observations that need to be made will depend on the distribution of the random outcomes. The more spread observed in the results, the more observations are necessary in order to guarantee that the mean of the observed values will be close to μ.

The Law of Small Numbers

The Law of Large Numbers describes what happens in the long term. If you throw a coin four times, then you many very well get four tails, while we know that there is a probability of 0.5 of the coin landing on tails. The law of Small Numbers states that the short term results of a probability experiment are not reliable in predicting long term results.

Rules for Means

There are three rules for the means of random variables:

- If X is a random variable and a and b are fixed numbers, $\mu_{a+bX} = a + b\mu_X$.
- If X and Y are random variables, $\mu_{x+y} = \mu_x + \mu_y$.
- Similarly, if X and Y are random variables, $\mu_{x-y} = \mu_x - \mu_y$.

Rules for Variance and Standard Deviation

The variance of a data set is denoted by s^2, while the variance of a random variable is written as σ_x^2. For a discrete variable, X, the variance can be calculated using the following formula:

$\sigma_x^2 = (x_1-\mu_x)^2 p_1 + (x_2-\mu_x)^2 p_2 + (x_3-\mu_x)^2 p_3 + \ldots + (x_k-\mu_x)^2 p_k$

The standard deviation, σ_x can be found by finding the square root of the variance.

The main rules for variances and standard deviations are as follows:

- If X is a random variable and a and b are fixed numbers, then: $\sigma^2_{a+bX} = b^2\sigma^2_X$
- If X and Y are independent random variables $\sigma^2_{X+Y} = \sigma_x^2 + \sigma_y^2$ and $\sigma^2_{X-Y} = \sigma_x^2 - \sigma_y^2$. This is called the *addition rule for variances of independent random variables.*
- If X and Y have a correlation ρ, $\sigma^2_{X+Y} = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_X\sigma_y$ and $\sigma^2_{X-Y} = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_y$
- This is the *general addition rule for variances of random variables.*

4.5 General Probability Rules

In addition to the rules that were discussed in the previous sections, there are several other rules to keep in mind when dealing with probability:

- The *union* of two events is the probability that at least of the two events will occur.
- We already know that if two events are disjoint, the probability of one or the other happening, P (A or B), is the sum of the separate probabilities: P (A) + P (B).

In cases where there are more than two events or if the events are not disjoint, the following rules apply:

- If events A, B and C are disjoint and therefore have no outcomes in common with each other, P (one or more of A, B, C) = P (A) + P (B) + P (C). This rule is also applicable in situations where there are many more events.
- The probability of event A or B occurring can also be found: P (A or B) = P (A) + P (B) - P (A and B). If A and B are disjoint, the probability P (A and B) is zero. This event, P(A or B) is called an *empty event* and can be left out of the equation.

Conditional Probability

A *conditional probability* is when we look at the probability of a particular event, given that another event has occurred. A conditional probability is written as P (A | B). This translates to the likelihood of event A occurring, given the event B has already occurred.

The probability that events A and B occur together is calculated by the formula:

- P (A and B) = P (A) P (B | A)

When the probability of event A is greater than 0, the *conditional probability* of B given A, is found using the formula:

- P (B | A) = P (A and B) / P (A)
- If events A and B are independent, P (B | A) = P (B).

The *intersection* of a set of events is the event that all events in the set occur. The intersection for event A, B and C is therefore:

- P (A and B and C) = P (A) P (B | A) P (C | A and B

Bayes's Rule

Suppose that A_1, A_2,.., A_k are disjoint events, all of which are have a probability greater than 0 and that add up to 1. Suppose that C is a different event, whose probability is not 0 or 1. In that case, the Bayes rule can be applied:

$$P(A_i|C) = \frac{P(C \mid A_i)\, P(A_i)}{P(C \mid A_1)\, P(A_1) + P(C \mid A_2)\, P(A_2) + ... + P(C \mid A_k)\, P(A_k)}$$

Independent events

Two events A and B with positive chances are dependent when:

$P(B \mid A) = P(B)$