Introduction to probability

Suppose an experiment has a finite set $X = \{x_1, x_2, \ldots, x_n\}$ of n possible outcomes. Each time the experiment is performed exactly one on the n outcomes happens. Assign each outcome a real number between 0 and 1, called the *probability* of that outcome. The probability of an outcome is supposed to be proportional to its likelihood of happening.

We want a probability of an outcome being near 1 to mean that that outcome is very likely to be the one that happens, and a probability near 0 to mean that that outcome almost never happens.

Write $p(x_i)$ for the probability of the outcome x_i. The sum of the probabilities of all outcomes in X must be 1 because the outcomes in X are the only possible outcomes, so one of them must happen.

So far we have $0 \leq p(x_i) \leq 1$ for each i and

$$\sum_{i=1}^{n} p(x_i) = 1.$$

Where do the probabilities $p(x_i)$ come from?

Sometimes they come from doing an experiment many times and tabulating the outcomes.

Example. Weather forecasters save enormous tables of weather conditions. The prediction, "There is a 40% chance of rain tomorrow." means that on 40% of the days (listed in the weather records) when the weather conditions were similar to what they are now, it rained the next day.

If outcome x has a 40% chance of happening, then its probability is written $p(x) = 0.4$ so that it will be between 0 and 1.

Where do the probabilities $p(x_i)$ come from?

Frequency probability: Perform an experiment (or look at historical records) and count the good and bad outcomes.

Example: toss a coin 1000 times.

Classical probability: Use a model of how the world works.

Example: a 6-sided die.

Delphi approach: Ask experts, tell them the average of their guesses, let them revise.

Sometimes we expect that all possible outcomes are equally likely because there is no reason to think that some outcomes are more likely than others. In this case, if there are n possible outcomes, then each outcome x has probability $p(x) = 1/n$.

Example. Suppose a deck of 52 cards has been shuffled well and then one card is chosen. The probability that the chosen card is the Six of Hearts is $p(\text{Six of Hearts}) = 1/52$.

This is an example of *equally likely* outcomes. Here is another.

Example. If a coin is properly balanced and tossed well, then the two sides Heads and Tails are equally likely, so each of these outcomes has probability 0.5.

We will often combine some outcomes and ask for the probability that at least one of them happens, but we don't care which one.

A subset E of a set X of all possible outcomes is called an *event*. We say that E "happens" if the outcome of the experiment is one of the outcomes in E.

The probability of an event E is the sum of the probabilities of the outcomes in it. We write $p(E) = \sum_{x \in E} p(x)$.

For any event, $0 \leq p(E) \leq 1$.

The probability that E does not happen is $1 - p(E)$.

Example. In a deck of cards, 13 of the 52 cards are Clubs, so the probability that a Club is drawn from a shuffled deck is

$$13 \times \frac{1}{52} = \frac{1}{4} = 0.25.$$

Four of the cards are Jacks (one from each of the four suits), so the probability that a Jack is drawn is

$$4 \times \frac{1}{52} = \frac{1}{13} \approx 0.076923.$$

Later, we will compute probabilities of events like this one: Suppose two cards are drawn from a deck. What is the probability that they are in the same suit?

Events may be combined using set theory.

The union $E \cup F$ of events E, F, happens if either one of them happens, that is, if the outcome is in either E or F.

The intersection $E \cap F$ of events E, F happens if both happen, that is, if the outcome is in both E and F.

Two events E, F are *mutually exclusive* if they are disjoint sets, that is, $E \cap F$ is empty. In other words, E, F are mutually exclusive if they cannot both happen. When E, F are mutually exclusive,

$$p(E \cup F) = p(E) + p(F).$$

(Recall the definition of $p(E)$.) Ditto for more than two events being mutually exclusive.

Example. The probability that a card drawn from a deck is either a Jack, a Queen or a King is $1/13 + 1/13 + 1/13 = 3/13$ because a card may be at most one of Jack, Queen, King.

Suppose E and F are two events and F can happen, that is, $p(F) > 0$. Define the *conditional probability of E given F* to be

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}.$$

Example. Find the conditional probability that a card is a Queen given that it is either a Jack, a Queen or a King. Here, E is the event, "the card is a Queen" and F is the event, "the card is either a Jack, a Queen or a King." We have $p(E) = 1/13$, $p(F) = 3/13$ and $p(E \cap F) = 1/13$ because $E \cap F = E$. The answer is

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{1/13}{3/13} = \frac{1}{3}.$$

Bayes' Theorem

Write the definition of conditional probability on the form

$$p(E \cap F) = p(E \mid F)p(F).$$

If we interchange E and F, we get

$$p(F \cap E) = p(F \mid E)p(E).$$

Since $F \cap E = E \cap F$, we have

$$p(E \mid F)p(F) = p(F \mid E)p(E),$$

and we have proved Bayes' Theorem:

Theorem. If both $p(E) > 0$ and $p(F) > 0$, then

$$p(F \mid E) = \frac{p(F)p(E \mid F)}{p(E)}.$$

Example of Bayes' Theorem. Draw one card from a deck. Let F be the event, "the card is a Jack or Queen or King of Spades." Let E be the event, "the card is a Queen." Since the Jack, Queen and King of Spades are 3 of the 52 cards, $p(F) = 3/52$. Since 4 of the 52 cards are Queens, $p(E) = 4/52 = 1/13$.

Let us compute $p(E \mid F)$. If F happens, then the card is one of the three cards: Jack or Queen or King of Spades. One of these is a Queen, so $p(E \mid F) = 1/3$.

By Bayes' Theorem,

$$p(F \mid E) = \frac{p(F)p(E \mid F)}{p(E)} = \frac{(3/52)(1/3)}{1/13} = \frac{1}{4}.$$

This result is easy to verify, because if the card is a Queen, then it has 1 chance in 4 of being the Queen of Spades. Thus, $p(F \mid E) = 1/4$.

Independent Events

Two events E and F are called *independent* if $p(E \mid F) = p(E)$. Intuitively, this says that E and F are *independent* if the probability that E happens does not depend on whether F happens.

When both $p(E) > 0$ and $p(F) > 0$, Bayes' Theorem implies that $p(E \mid F) = p(E)$ if and only if $p(F \mid E) = p(F)$.

The formula $p(E \cap F) = p(E \mid F)p(F)$ and the definition of independent imply that E and F are independent iff $p(E \cap F) = p(E) \cdot p(F)$.

The two events in the Bayes' Theorem example are not independent because $p(E \cap F) =$ 1/52 since the card must be the Queen of Spades, while $p(E) \cdot p(F) =$ (1/13)(3/52) $\neq$ 1/52.

Example. Let E be the event, "the card is a Spade." Let F be the event, "the card is a Queen." Since there are 13 Spades, $p(E) =$ 13/52 = 1/4. Since there are 4 Queens, $p(F) =$ 4/52 = 1/13. The event $E \cap F$ says that the card is the Queen of Spades, which is 1 of 52 cards, so

$$p(E \cap F) = 1/52 = (1/4)(1/13) = p(E)p(F),$$

so the events E and F are independent.

Random Variables

A *sample space* is the set of all possible outcomes x_i, each having a probability $p(x_i)$. (In this class, we assume the number of possible outcomes is finite.)

Example. Draw a card. There are 52 possible outcomes, like $x_i =$ "Queen of Spades". Each has probability $p(x_i) = 1/52$. The sample space is the set of 52 cards.

A *random variable* is a (real-valued) function r defined on a sample space.

Example. Draw a card x_i. Let $r(x_i)$ denote the value of the card, defined as follows: If the card has a number, this number is its value. So $r(\text{Six of Clubs}) = 6$. If the card is an Ace, then its value is 1: $r(\text{Ace of Diamonds}) = 1$. If the card is a Jack, Queen or King, then its value is 10: $r(\text{Queen of Hearts}) = 10$. Then $r(x_i)$ is a random variable define on a deck of cards.

Let r_1, r_2, ..., be all possible values of a random variable r defined on a sample space. (This is a finite number of values.) The *probability distribution of* r is the function f defined by $f(r_j) = p(r(x_i) = r_j)$, that is, $f(r_j)$ is the probability of the event "$r(x_i) = r_j$."

Example. In the example of the random variable on the deck of cards above, $f(i) = 1/13$ for $1 \leq i \leq 9$ because 4 of the 52 cards have value i in this range. However, $f(10) = 4/13$ since 4 cards in each suit have value 10.

Two random variables r, s are *independent* if for any possible values r_1, s_1, they could assume, the probability that "$r(x) = r_1$ and $s(x) = s_1$" equals $p(r(x) = r_1) \cdot p(s(x) = s_1)$.

Example. Draw a card. Record its value as r. Replace the card in the deck. Shuffle the deck again and draw a second card. Record its value as s. Then r and s are independent random variables with the same probability distribution.

There are several concise ways to describe the probability distribution of a random variable by giving a "typical" value of it.

The *median* of the probability distribution f of a random variable r is a value r_m so that the probability of $r(x) > r_m$ is as close to 0.5 as possible. (f is used to compute this probability.) The median is the "middle value" of $r(x)$.

Example. Suppose r has this probability distribution:

r	3	6	8	9
$f(r)$	0.2	0.2	0.4	0.2

The median of r is 6 because $p(r > 6)$ is 0.6 while $p(r > 8)$ is 0.2 and 0.6 is closer to 0.5.

Another (more useful) typical value is the mean or average or expected value.

The *mean* or *expected value* of a random variable r with values r_1, r_2, ... and probability distribution f is

$$\mu = \mathbf{E}(r) = \sum_i r_i f(r_i).$$

Example. The mean of the value of cards in the example above is

$$\begin{aligned} 1 \cdot \frac{1}{13} + 2 \cdot \frac{1}{13} + \cdots 9 \cdot \frac{1}{13} + 10 \cdot \frac{4}{13} &= \\ \left(\frac{9 \cdot 10}{2}\right)\left(\frac{1}{13}\right) + 10 \cdot \frac{4}{13} &= \frac{85}{13}. \end{aligned}$$

This number, $85/13 \approx 6.5$, is the average value of a card.

If F is a real function of a real variable, and r is a random variable, then $F(r)$ is another random variable. It has value $F(r(x))$ on outcome x. Its expected value is

$$\mu = \mathbf{E}(F(r)) = \sum_i F(r_i) f(r_i).$$

The k-th moment of a random variable r is the expected value of $F(r) = r^k$.

The *variance* of a random variable r with expected value μ is

$$\mathbf{Var}(r) = \mathbf{E}((r-\mu)^2) = \mathbf{E}(r^2) - \mu^2.$$

The last equation is a simple theorem.

The square root of the variance of r is the *standard deviation* of r. It measures how much $r(x)$ varies from the mean μ.

Example. Suppose r has this probability distribution:

r	3	6	8	9
$f(r)$	0.2	0.2	0.4	0.2

The mean of $F(r) = r$ is

$$\mu = 3 \cdot 0.2 + 6 \cdot 0.2 + 8 \cdot 0.4 + 9 \cdot 0.2 = 6.8.$$

The second moment of r is $\mathbf{E}(r^2) =$

$$3^2 \cdot 0.2 + 6^2 \cdot 0.2 + 8^2 \cdot 0.4 + 9^2 \cdot 0.2 = 50.8.$$

The variance of r is

$$\mathbf{Var}(r) = \mathbf{E}(r^2) - \mu^2 = 50.8 - (6.8)^2 = 4.56$$

and the standard deviation is $\sigma = \sqrt{4.56} = 2.14$.

An important and common probability distribution is the *uniform distribution* in which each possible value for the random variable has the same probability. If there are n possible values, then each has probability $1/n$ of occurring. We say the values are *equally likely*.

Recall the question we asked earlier: What is the probability that two cards drawn at random from a deck are in the same suit?

First suppose that the first card drawn is not replaced. Then there are 51 cards remaining and 12 of them are in the same suit as the first card. The probability is 12/51. (This is called sampling without replacement.)

Now suppose that the suit of the first card is noted and then it is replaced in the deck and the deck reshuffled before the second card is drawn. Then the second card is one of 52 cards of which 13 are in the same suit as the first card. The probability is $13/52 = 1/4 = 0.25$. (This is called sampling with replacement.)

The Birthday Paradox

What is the smallest positive integer k so that the probability is > 0.5 that at least two people in a group of k people have the same birthday?

The surprising answer is $k = 23$.

The explanation is complicated and will come later.

If there were n birthdays, rather than 365 or 366, the answer would be that we need $k \approx 1.18\sqrt{n}$ people to get a 50% chance that two have the same birthday.

This fact is needed for the study of hash functions.

The overlap between two sets

A related fact we need for hash functions is this:

What is the smallest positive integer k so that the probability is > 0.5 that in two groups of k people, at least one person in the first group has the same birthday as at least one person in the second group?

Here the answer is $k \approx 16$.

If there were n birthdays, rather than 365 or 366, the answer would be that we need $k \approx 0.83\sqrt{n}$ people to get a 50% chance that one person from the first group has the same birthday as one person from the second group.

Now we derive the results just mentioned, beginning with the Birthday Paradox.

Ignore February 29.

Assume each birthday is equally likely.

The probability that k people all have different birthdays is

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365-k+1}{365}$$

which is

$$\frac{365!}{(365-k)! \times (365)^k}.$$

Thus the probability that at least two of k people have the same birthday is

$$P(k) = 1 - \frac{365!}{(365-k)! \times (365)^k}.$$

More generally, suppose we are given an integer-valued random variable with uniform distribution between 1 and n. Choose k instances of this random variable. What is the probability $P(n,k)$ that at least two of the k instances are the same value?

As for birthdays, we find

$$P(n,k) = 1 - \frac{n!}{(n-k)!n^k}.$$

Write this as

$$P(n,k) = 1 - (1 - \frac{1}{n})(1 - \frac{2}{n}) \times \cdots \times (1 - \frac{k-1}{n}).$$

To estimate this, note that $1 - x \approx e^{-x}$ when x is small.

This gives

$$P(n,k) \approx 1 - e^{-1/n}e^{-2/n}e^{-3/n} \times \cdots \times e^{-(k-1)/n}$$

$$P(n,k) \approx 1 - e^{-(1/n+2/n+3/n+\cdots+(k-1)/n}$$

$$P(n,k) \approx 1 - e^{-k(k-1)/(2n)}.$$

We will have $P(n,k) \approx 0.5$ when

$$0.5 \approx 1 - e^{-k(k-1)/(2n)}$$

or $2 \approx e^{k(k-1)/(2n)}$, that is, when

$$\ln 2 \approx k(k-1)/(2n).$$

When k is large, the percentage difference between k and $k-1$ is small, and we may approximate $k-1 \approx k$. This gives $k^2 \approx 2n \ln 2$ or

$$k \approx \sqrt{2(\ln 2)n} \approx 1.18\sqrt{n}.$$

For $n = 365$, we find

$$k \approx 1.18\sqrt{365} \approx 22.54,$$

or $k \approx 23$.

Suppose $H(M)$ is a hash function with m-bit output. There are $n = 2^m$ possible hash values.

If H is applied to k random inputs, the probability of finding a duplicate ($H(M) = H(M')$) is $P(2^m, k)$. The minimum number of k needed for a duplicate to occur with probability > 0.5 is about

$$k = 1.18\sqrt{2^m} = 1.18 \times 2^{m/2}.$$

The overlap between two sets

Given an integer random variable with uniform distribution between 1 and n, and two sets of k ($k \leq n$) instances of the random variable, what is the probability $R(n,k)$ that the two sets overlap, that is, at least one of the n values appears in both sets?

We assume k is small enough ($k < \sqrt{n}$) so that the k instances of the random variable in each set are all different. (A few duplicates won't hurt this analysis.)

The probability that one given element of the first set does not match any element of the second set is $(1 - \frac{1}{n})^k$.

The probability that the two sets are disjoint is

$$((1 - \frac{1}{n})^k)^k = (1 - \frac{1}{n})^{k^2}$$

so $R(n,k) = 1 - (1 - \frac{1}{n})^{k^2}$.

Using $1 - x \approx e^{-x}$, we get

$$R(n,k) \approx 1 - (e^{-1/n})^{k^2} = 1 - e^{-k^2/n}.$$

We will have $R(n,k) \approx 0.5$ when $\frac{1}{2} \approx 1 - e^{-k^2/n}$ or $2 \approx e^{k^2/n}$ or $\ln 2 \approx k^2/n$ or

$$k \approx \sqrt{(\ln 2)n} \approx 0.83\sqrt{n}$$

Suppose a hash function H with $n = 2^m$ possible values is applied to k random inputs to produce a set X of hash values and again to k additional random inputs to produce another set Y of hash values. What is the minimum value of k so that the probability is at least 0.5 of finding at least one match between the two sets, that is, $H(x) = H(y)$, where $x \in X$ and $y \in Y$? Using the approximation above, the minimum k is about

$$k \approx 0.83\sqrt{2^m} = 0.83 \times 2^{m/2}.$$