Chapter 1: Probability models and counting

Part 1: Probability model

Probability theory is the mathematical toolbox to describe phenomena or experiments where randomness occur. To have a probability model we need the following ingredients

- A *sample space* S which is the collection of all possible outcomes of the (random) experiment. We shall consider mostly finite sample spaces S.

- A *probability distribution*. To each element $i \in S$ we assign a probability $p(i) \in S$.

  $$p(i) \;=\; \text{Probability that the outcome } i \text{ occurs}$$

  We have

  $$0 \le p(i) \le 1 \qquad \sum_{i \in S} p(i) = 1\,.$$

  By definition probabilities are nonnegative numbers and add up to 1.

- An *event* A is a subset of the sample space S. It describes an experiment or an observation that is compatible with the outcomes $i \in A$. The probability that A occurs, $P(A)$, is given by

  $$P(A) \;=\; \sum_{i \in A} p(i)\,.$$

Example: Thoss three (fair) coins and record if the coin lands on tail (T) or head (H). The sample space is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

and has $8 = 2^3$ elements. For fair coins it is natural assign the probability 1/8 to each outcome.

$$P(HHH) = P(HHT) = \cdots = P(TTT) \;=\; 1/8$$

An example of an event A is that at least two of the coins land on head. Then

$$A = \{HHH, HHT, THH, HHT\} \qquad P(A) = 1/2\,.$$

∎

The *basic operations of set theory* have a direct probabilistic interpretation:

- The event $A \cup B$ is the set of outcomes wich belong either to A or to B. We say that $P(A \cup B)$ is the probability that *either A or B occurs.*
- The event $A \cap B$ is the set of outcomes which belong to A and to B. We say that $P(A \cup B)$ is the probability that A *and* B *occurs.*
- The event $A \setminus B$ is the set of outcomes which belong to A but not to B. We say that $P(A \setminus B)$ is the probability that A *occurs but* B *does not occurs.*
- The event $\overline{A} = S \setminus A$ is the set of outcomes which do not belong to A. We say that $P(\overline{A})$ is the probability that A *does not occur.*

We have the following simple rules to compute probability of events. Check them!

Theorem 1. *Suppose A and B are events. Then we have*

1. $0 \leq P(A) \leq 1$ *for any event* $A \subset S$.
2. $P(A) \leq P(B)$ *if* $A \subset B$.
3. $P(\overline{A}) = 1 - P(A)$
4. $P(A \cup B) = P(A) + P(B)$ *if A and B are disjoint.*
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *for general A and B.*

Proof. We let the reader check 1. to 4. For 5. we can reduce ourselves to 4. by writing $A \cup B$ as the union of two disjoint event, for example

$$A \cup B \,=\, A \cup (B \setminus A)\,.$$

We do have, by 4.

$$P(A \cup B) \,=\, P(A) + P(B \setminus A) \tag{1}$$

On the other hand we can write B has the union of two disjoint sets (the outcomes in B which are also in A or not).

$$B = (B \cap A) \cup (B \setminus A)$$

and so by 4.

$$P(B) = P(B \cap A) + P(B \setminus A) \tag{2}$$

So by combining (1) and (2) we find 5. ∎

Example: Tossing three coins again let A be the event that the first toss is head while B is the event that the second toss is tail. Then $A = \{HHH, HHT, HTH, HTT\}$, and $B = \{HTH, HTT, TTH, TTT\}$, $A \cap B = \{HTH, HTT\}$. We have $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/2 - 1/4 = 3/4$. ∎

There is a generalization of item 5. of the previous theorem to more than 2 events. For example we have for 3 events A_1, A_2 and A_3

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

This can be proved by applying item 5. repeatedly.

The general formula for n events is called the *inclusion-exclusion* theorem and can be proved by induction.

Theorem 2. *Suppose $A_1, \cdots A_n$ are events . Then we have*

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= \sum_{1 \le i \le n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) \\ &\quad + \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n) \,. \end{aligned}$$

Odds vs probabilities. Often, especially in gambling situations the randomness of the experiment is expressed in terms of *odds* rather than probabilities. For we make a bet at 5 to 1 odds that U of X will beat U of Z in next week basketball game. What is meant is that the probability that X wins is thought to be 5 times greater than the probability that Y wins. That is we have

$$p = P(\text{X wins}) = 5P(\text{Y wins}) = 5P(\text{X looses}) = 5(1 - p)$$

and thus we have $p = 5(p - 1)$ of $p = 5/6$. More generally we have

$$\text{The odds of an event } A \text{ are } r \text{ to } s \iff \frac{P(A)}{1 - P(A)} = \frac{r}{s} \iff P(A) = \frac{r/s}{r/s + 1} = \frac{r}{r + s}$$

Paying odds in gambling. In betting situations, say betting on horse races or betting on basketball or football games, the potential payoff for the gambler is expressed in terms of odds. There are different from the "true odds" so as to permit the betting house (that is the bookmaker to make a profit). The following odds are used

1. **Fractional odds** (common in UK). *Fractional odds* of, for example $\frac{3}{1}$, means that a bet of say \$10, if won, will generate a profit of \$ $\frac{4}{1} \times 10 = 40$ for the gambler. That is the gambler will get \$50 back, his bet of \$10 back plus a profit \$40. If the fraction is less than one, say the odds are $\frac{3}{7}$ then a bet of \$10 will generate a profit $\$\frac{30}{7} = 4.29$.

2. **Decimal odds** (common in Europe). *Decimal odds* are expressed by a number (always greater than 1) which express the amount the betting house will pay for a winning bet of \$1. For example decimal odds of 6 means that for a winning bet of \$10 the player will get \$60 back for a profit of \$50. That is decimal odds of 6 corresponds to a fractional odd of $\frac{5}{1}$. In the same way decimal odds of 1.3 correspond to a fractional odd of $\frac{3}{10}$.

3. **Moneyline odds** (common in the USA). The money line odds are based on a bet of \$ 100 and are expressed either using a positive number greater than 100 (say +250) or as a negative number less than -100 (say -125). Positive and negative number have a slightly different interpretation.

   Odds of +250 means that a bet of \$100 will generate a profit of \$250. So this corresponds to fractional odds of $\frac{250}{100} = \frac{5}{2}$.

   Odds of -125 means that you need to bet \$125 to generate a profit of \$ 100 that is it corresponds to fractional odds of $\frac{100}{125} = \frac{4}{5}$

Computing paying odds. All these odds are of course equivalent but we may wonder how do the betting house fix them? On such procedure is called the *parimutuel betting.* In this model, the betting odds keep chainging as more and more people wage their bets and the odds are computed to reflect the amount of money which is bet on different teams but also taking into account that the betting house takes a cut of any bet. Suppose that

$$T_A = \text{ amount of money bet on Team } A$$

$$T_B = \text{ amount of money bet on Team } B$$

If the betting house would take *no commission* all the money bet $T_A + T_B$ would be redistributed to the winners and so the (fractional) paying odds would be respectively

$$\text{Paying odds if A wins} = \frac{T_B}{T_A} \text{ (no commission)}$$

$$\text{Paying odds if B wins} = \frac{T_A}{T_B} \text{ (no commission)}$$

In practice the betting house takes a fixed percentage of the every bet as commission (let us say %10). Then only %90 of the total amount bet will be returned to the winners that we obtain

$$\text{Paying odds if A wins} = \frac{\frac{9}{10}(T_A + T_B) - T_A}{T_A} = \frac{T_B - \frac{1}{10}(T_A + T_B)}{T_A} \ (\%10 \text{ commission})$$

$$\text{Paying odd if B wins} = \frac{\frac{9}{10}(T_A + T_B) - T_b}{T_B} = \frac{T_A - \frac{1}{10}(T_A + T_B)}{T_B} \ (\%10 \text{ commission})$$

Imagine for example that the bet is on whether team A will win against team B and that \$ 660 has been waged on team A and \$520 on team BB and the commission is %10. then the paying odds will be

$$\text{Paying odds if A wins} = \frac{520 - 118}{660} = 0.609$$

$$\text{Paying odd if B wins} = \frac{660 - 118}{520} = 1.042$$

Expressed in money line odds this would give -164 for A and $+104$ for B.

Uniform distribution ("Naive probabilities"). In many examples it is natural to assign the same probability to each event in the sample space. If the sample space is S we denote by the cardinality of S by

$$\#S = \text{ number of elements in } S$$

Then for every event $i \in S$ we set

$$p(i) = \frac{1}{\#S},$$

and for any event A we have

$$p(A) = \frac{\#A}{\#S}$$

Example. Throw two fair dice. The sample space is the set of pairs (i, j) with i and j an integer between 1 and 6 and has cardinality 36. We then obtain for example

$$P(\text{Sum of the dice is 2}) = \frac{1}{36}, \quad P(\text{Sum of the dice is 9}) = \frac{4}{36}, \cdots$$

$\blacksquare$

The birthday problem. A classical problem in probability is the following. What is the probability that among N people at least 2 have the same birthday. As it turns out, and at first sight maybe surprisingly, one needs few people to have a high probability of matching birthdays. For example for $N = 23$ there is a probability greater than 1/2 than at least two people have the same birthday. To compute this we will make the simplifying but reasonable assumptions that there is no leap year and that every birthday is equally likely.

If there are N people present, the sample space S is the set of all birthdays of the everyone. Since there is 365 choice for everyone we have

$$\#S \;=\; 365^N$$

We consider the event

$$A = \text{ at least two people have the same birthday}$$

It is easier to consider instead the complementary event

$$B = \overline{A} = \text{ no pair have the same birthday}$$

To compute the cardinality of B we make a list of the N people (the order does not matter). There is 365 choice of birthday for the first one on the list, for the second one on the list, there is only 364 choice of birthday if they do not have the same birthday. Continuing in the same way we find

$$\#B \;=\; 365 \times 364 \times 363 \times \cdots \times (365 - N + 1)$$

and so

$$\begin{aligned} P(B) \;&=\; \frac{365 \times 364 \times 363 \times \cdots \times (365 - (N-1)}{365^N} \\ &=\; 1 \times \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \times \left(1 - \frac{N-1}{365}\right) \end{aligned}$$

To compute this efficiently we recall from calculus (use L'Hospital rule to prove this) that for any number x we have

$$e^x \;=\; \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

We take $n = 365$ which is a reasonably large number and take $x = -1, -2, \cdots$. We have the approximation

$$\left(1 - \frac{j}{365}\right) \approx e^{-j/365}\,.$$

We find then

$$\begin{aligned} P(B) &\approx 1 \times e^{-1/365} \times e^{-2/365} \times e^{-(N-1)/365} \\ &= e^{-(1+2+\cdots N-1)/365} \\ &= e^{-N(N-1)/730} \end{aligned}$$

by using the well-known identity $1 + 2 + \cdots N = \dfrac{N(N+1)}{2}$.

How many people are needed to have a probability of 1/2 of having 2 same birthday. We have

$$P(B) = \frac{1}{2} \approx e^{-N(N-1)/730} \iff N(N-1) = 730 \ln(2)$$

Even for moderately small N, $N(N-1) \approx N^2$ and so we find the approximate answer

$$N \approx \sqrt{730 \ln(2)} \;=\; 22.49$$

That is if there are 23 people in a room, the probability that two have the same birthday is greater than 1/2. Similarly we find that if there are

$$N \approx \sqrt{730 \ln(10)} \;=\; 40.99$$

people in the room there is a probability greater than .9 than two people have the same birthday.

Part 2: Combinatorics

In many problems in probability where one uses uniform distribution (many of them related to gambling's) one needs to count the number of outcomes compatible with a certain event. In order to do this we shall need a few basic facts of combinatorics

Permutations: Suppose you have n objects and you make a list of these objects. There are

$$n! = n(n-1)(n-2)\cdots 1$$

different way to write down this list, since there are n choices for the first on the list, $n-1$ choice for the second, and so on.

The number $n!$ grows very fast with n. Often it is useful to have a good estimate of $n!$ for large n and such an estimate is given by Stirling's formula

$$\text{Stirling's formula} \;\; n! \sim n^n e^{-n}\sqrt{2\pi n}$$

where the symbols $a_n \sim b_n$ means here that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$

Combinations: Suppose you have a group of n objects and you wish to select a j of the n objects. The number of ways you can do this defines the *binomial coefficients*

$$\binom{n}{j} = \#\text{ of ways to pick } j \text{ objects out of } n \text{ objects}$$

and this pronounced "n choose j".

Example: The set $U = \{a, b, c\}$ has 3 elements. The subsets of U are

$$\emptyset\,, \quad \{a\}\,, \quad \{b\}\,, \quad \{c\}\,, \quad \{a,b\}\,, \quad \{a,c\}\,, \quad \{b,c\}\,, \quad \{a,b,c\}$$

and there are $\binom{3}{0} = 1$ subset with 0 elements, $\binom{3}{1} = 3$ subset with 1 elements, $\binom{3}{2} = 3$ subset with 2 elements, and $\binom{3}{3} = 1$ subset with 3 elements. ■

Formulas involving binomial coefficients and "Storyproofs": There are many relations between binomial coefficients. One can prove these relation using the formula for the coefficients derived a bit later. A very elegant alternative is often to use the meaning of the coefficients and to make up a "story". We give a few example

1. We have the equality
$$\binom{n}{k} = \binom{n}{n-k} \tag{3}$$
For example we have $\binom{10}{6} = \binom{10}{4}$.

   To see why this is true think of forming a group of k people out of n people. You can do by selecting k people with $\binom{n}{k}$ choices. Alternatively you can for the group by selecting all the people who are **not** among the group, that is you select $n-k$ people not in the group and there $\binom{n}{n-k}$ ways of doing this.

2. **Recursion relation for the binomial coefficients:and Pascal triangle:** There is a simple recursion relation for the binomial coefficients $\binom{n}{j}$ in terms of the binomial coefficients $\binom{n-1}{k}$:
$$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1} \tag{4}$$
for $0 < j < n$. To use this recursion one needs to know that $\binom{n}{0} = \binom{n}{n} = 1$.

   To see why the formula (4) holds think of a group of n people. The left hand side of (4) is the number of ways to form groups of j people out of those n people. Now let us pick one distinguished individual among the n, let us say we pick Bob. Then $\binom{n-1}{j}$ is the number of way to choose a group of j people which do not include Bob (pick j out of the remaining $n-1$) while $\binom{n-1}{j-1}$ is the number of ways to pick a group of j people which does include Bob (pick Bob and then pick $j-1$ out of the remaining $n-1$. Adding these two we obtain the right hand side of (4) .

3. We have the relation
$$k\binom{n}{k} = n\binom{n-1}{k-1} \tag{5}$$
To with this holds imagine selecting a team of k (out of n) and selecting also a captain for the team. Then you can either pick first the team ($\binom{n}{k}$ ways) and then selecting the captain (k choices). This gives the left hand side of (5). Alternatively you can pick first the captain (n choices) and then select the rest of the team ($\binom{n-1}{k-1}$ ways).This gives the right hand side of (5).

Formula for the binomial coefficients To find an explicit formula for $\binom{n}{j}$ we note first that
$$n(n-1)\cdots n-(j-1)$$
is the number of ways to write an *ordered* list of j objects out of n objects since there are n choices for the first one on the list, $n-1$ choices for the second one and so on. Many

of these lists contain the same objects but arranged in a different order and there are $j!$ ways to write a list of the same j objects in different orders. So we have

$$\binom{n}{j} = \frac{n(n-1)\cdots n-(j-1)}{j!}$$

which we can rewrite as

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}$$

Poker hands. We will compute the probability of certain poker hands. A poker hands consists of a 5 randomly chosen cards out of a deck of 52. So we have

$$\text{Total number of poker hands} \;=\; \binom{52}{5} \;=\; 2598960$$

Four of a kind: This hands consists of 4 cards of the same values (say 4 seven). To compute the probability of a four of a kinf note that there are 13 choices for the choice of values of the four of a kind. Then there are 48 cards left and so 48 choice for the remaining cards. So

$$\text{Probability of a four of a kind} \;=\; \frac{13 \times 48}{\binom{52}{5}} \;=\; \frac{624}{2598960} \;=\; 0.00024$$

Full house: This hands consists three cards of the same value and two cards of an another value (e.g. 3 kings and 2 eights). There are 13 ways to choose the value of three of a kind and once this value is chose there is $\binom{4}{3}$ to select the three cards out of the four of same value. There are then 12 values left to choose from for the pair and there $\binom{4}{2}$ to select the the pair. So we have

$$\text{Probability of a four of a full house} \;=\; \frac{13 \times \binom{4}{3} \times 12 \times \binom{4}{2}}{\binom{52}{5}} \;=\; \frac{3744}{2598960} \;=\; 0.0014$$

So the full house is 6 times as likely as the four of a kind.

Three of a kind: There are $13\binom{4}{3}$ ways to pick a three of kind. There are then 48 cards left from which to choose the remaining last 2 cards and there are $\binom{48}{2}$ ways to do this. But we are then also allowing to pick a pair for the remaining two cards which would give a full house. Therefore we have

$$\text{Probability of a three of a kind} \;=\; \frac{13 \times \binom{4}{3} \times \binom{48}{2} - 13 \times \binom{4}{3} \times 12 \times \binom{4}{2}}{\binom{52}{5}}$$

Another way to compute this probability is to note that among the 48 remaining cards we should choose two different values (so as not to have a pair) and then pick a card of that value. This gives

$$\text{Probability of a three of a kind} = \frac{13 \times \binom{4}{3} \times \binom{12}{2} \times \binom{4}{1} \times \binom{4}{1}}{\binom{52}{5}}$$

Either way this gives a probability $\frac{54912}{2598960} = 0.0211$

Keno. This is a popular form of lottery played in casinos as well as often in bars and restaurants. For example in Massachusetts, the numbers are drawn every four minutes and appear on screens.

The game is played with the numbers $\{1, 2, 3, \cdots, 80\}$ and the casino randomly selects 20 numbers out of those. Clearly there are $\binom{80}{20}$ choices. The player plays by selecting m numbers out of 80. The number m varies and typically the game allows for the player to choose m (for example in Massachusetts any m with $1 \le m \le 12$ is allowed). One say that a player gets a *catch of* k if k of his m numbers matches some of the 20 numbers selected by the casino.

Let us take $m = 8$ and compute the probability of a catch of k in this case. Think now of the 80 numbers divided into two groups the 8 "good numbers" selected by the player and the 72 bad numbers which are not. For the player to get a catch of k, k of his numbers must be selected by the casino from his 8 "good numbers" and the remaining $20 - k$ numbers are selected from the "bad numbers". So we find

$$P(\text{ catch of } k \text{ if playing 8 numbers}) = \frac{\binom{8}{k}\binom{72}{20-k}}{\binom{80}{20}}, \quad k = 0, 1, \cdots 8$$

In general we find

$$P(\text{ catch of } k \text{ if playing } m \text{ numbers}) = \frac{\binom{m}{k}\binom{80-m}{20-k}}{\binom{80}{20}} \quad k = 0, 1, \cdots m$$

We will revisit the game of Keno later: the list of all odds and payouts as well as the the detailed rules for Keno as played Massachusetts can be found at `http://www.masslottery.com/games/keno.html`.

A lottery game. Consider the following lottery game. There are N tickets which are numbered 1 to N. To each ticket one assigns a randomly chosen between 1 and N (so that all N numbers are chosen) and the numbers are hidden behind a scratch surface. The winner are the one with both numbers matching and the question we ask is to find the probability that we have at least one winner at this game. The problem occurs in various context and is often called the"matching hat" problem since you can imagine N gentlemen dropping all their hats in a big bucket and then picking them at random. The question is what is the probability to pick your own hat.

To answer this question we use inclusion-exclusion. We write A_i for the event that the ticket with number i is a winner and so we need to compute

$$P(A_1 \cup A_2 \cup \cdots A_N)$$

We have $P(A_1) = \frac{1}{N}$ since there are N choice for the number and by symmetry we have $P(A_i) = \frac{1}{N}$ for all i. For two events $A_1 \cap A_2$ we find $P(A_1 \cap A_2) = \frac{1}{N(N-1)}$ since there are 1 out of N choice for the player 1 and then 1 out of $N-1$ choices for the second player. Again by symmetry we have $P(A_i \cap A_j) = \frac{1}{N(N-1)}$. Similarly we find for any collection

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) \;=\; \frac{1}{N(N-1)\cdots(N-k+1)}\,.$$

if all the i_l are distinct. Getting back to the inclusion-exclusion we find that

$$\begin{aligned}
& P(A_1 \cup A_2 \cup \cdots A_N) \\
= \;& N\frac{1}{N} - \binom{N}{2}\frac{1}{N(N-1)} + \binom{N}{3}\frac{1}{N(N-1)(N-2)} \cdots (-1)^{N-1}\binom{N}{N}\frac{1}{N!} \\
= \;& 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots (-1)^{N-1}\frac{1}{N!} \qquad (6)
\end{aligned}$$

To get an idea recall that the series for e^{-1} is

$$e^{-1} \;=\; \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \;=\; 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots$$

so that we have, approximately,

$$P(A_1 \cup A_2 \cup \cdots A_N) \approx 1 - \frac{1}{e} = 0.6321$$

This approximation is however extremely good since we have the true values

N	$P(A_1 \cup A_2 \cup \cdots A_N)$
1	1
2	0.5
3	0. 6666
4	0.6250
5	0.6333
6	0.6319
7	0.6321

so that the approximation is essentially exact for very small number!