Chapter 2. Probability

Probability: a number between 0 and 1 to quantifying uncertainty in a mathematical manner.

2.1 Sample space and events

Sample Space Ω: a set of possible outcomes of an experiment.

Sample outcome, realization or **element**: a point in a sample space, denoted by $\omega \in \Omega$.

Event or **random event**: a subset of Ω, i.e. an assemble of some sample outcomes

Example 1. Experiment – Toss a coin two times.

Sample space = $\{HH, HT, TH, TT\}$.

$A \equiv \{HH, HT\} = \{\text{1st toss is head}\}$ is an event.

What is the sample space if we toss a coin for ever? — *the Bernoulli trial*.

Background of tossing a coin: success or failure, up or down, better or worse, boy or girl, 1 or 0 and etc.

Example 2. Find the sample space in each of the following cases

- number of insect damaged leaves on a plant
- lifetime (in hours) of a light bulb
- weight of a 10-hour old infant
- exchange rate of pounds sterling to US-dollars today next year
- directional movement S&P500 index price tomorrow

Complement of event A: $A^c = \{\omega \in \Omega : \omega \notin A\}$. Obviously $\Omega^c = \varnothing$ (the empty set).

Union of events A and B: $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$. Then $A \cup B = B \cup A$, $A \cup A^c = \Omega$.

Intersection of events A and B: $A \cap B \equiv AB = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$. Then $A \cap B = B \cap A$, $A \cap A^c = \varnothing$.

If $A_1, A_2, \cdots$ is a sequence of events,

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for at least one } i\}, \quad \bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for all } i\}.$$

Difference of events A and B: $A - B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$. Obviously $A - B \neq B - A$.

Inclusion: occurrence of event A implies that of B, we say $A \subset B$.

Summary of Terminology

Ω	Sample space, true event (always true)
$\varnothing$	null event (always false)
ω	outcome, realization or element
A^c	complement of A (not A)
$A \cup B$	union (A or B)
$A \cap B$ or AB	intersection (A and B)
$A - B$ or $A \backslash B$	set difference
$A \subset B$	set inclusion

Mutually exclusive or **disjoint**: A and B are mutually exclusive if $A \cap B = \varnothing$. Obviously A and A^c are mutually exclusive.

Partition of Ω: a sequence $\underline{\text{disjoint}}$ events $A_1, A_2, \cdots$ such that

$$\bigcup_{i=1}^{\infty} A_i = \Omega.$$

Indicator of A: $I_A \equiv I_A(\omega)$ — a function defined on $\omega \in \Omega$:

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise,} \end{cases} \quad \text{or equivalently} \quad I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Limits of a sequence of monotonic events:
(i) A sequence $A_1, A_2, \cdots$ is monotone increasing if $A_1 \subset A_2 \subset \cdots$. We define $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i$.
(ii) A sequence $A_1, A_2, \cdots$ is monotone decreasing if $A_1 \supset A_2 \supset \cdots$. We define $\lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i$.
In both cases, we may write $A_n \to A$, where A denotes its limit.

Example 3. Let $\Omega = (-\infty, \infty)$, $A_i = [0, 1/i)$. Then

$$\cup_{i=1}^{\infty} A_i = [0, 1), \qquad \cap_{i=1}^{\infty} A_i = \{0\}.$$

If we change to $A_i = (0, 1/i)$, then $\cup_{i=1}^{\infty} A_i = (0, 1)$ and $\cap_{i=1}^{\infty} A_i = \emptyset$.

For $A_i = (-i, i)$, $\cup_{i=1}^{\infty} A_i = \Omega$.

Example 4. Let Ω be the salaries earned by the graduates from a Business School. We may choose $\Omega = [0, \infty)$. Based on the dataset "Jobs.txt", we extract some *interesting* events/subsets. Recall the info of the dataset:

- C1: ID number
- C2: Job type, 1 - accounting, 2 - finance, 3 - management, 4 - marketing and sales, 5 -others
- C3: Sex, 1 - male, 2 - female
- C4: Job satisfaction, 1 - very satisfied, 2 - satisfied, 3 - not satisfied
- C5: Salary (in thousand pounds)

C6: No. of jobs after graduation

We have defined salaries of male and female graduates respectively as follows:

```
> jobs <- read.table("Jobs.txt", header=T, row.names=1)
> mSalary <- jobs[,4][jobs[,2]==1]
> fSalary <- jobs[,4][jobs[,2]==2]
```

Similarly we may extract the salaries from finance sector or accounting:

```
> finSalary <- jobs[,4][jobs[,1]==2]; summary(finSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  37.00   46.00   53.00   52.08   58.00   65.00 
> accSalary <- jobs[,4][jobs[,1]==1]; summary(accSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  40.00   47.00   51.00   50.45   54.00   62.00 
```

According to this dataset, finance pays slightly higher than accounting. We may also extract the salaries for males (females) in accounting:

```
> maccSalary <- jobs[,4][(jobs[,1]==1) & (jobs[,2]==1)]
       # &' stands for logic operation and'
> summary(mfinSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  44.00   48.00   51.00   51.31   55.00   62.00
> faccSalary<- jobs[,4][(jobs[,1]==1) & (jobs[,2]==2)]
> summary(ffinSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  40.00   45.25   49.50   49.66   53.00   61.00
```

To extract the salaries for males in both finance and accounting:

```
>  mfinaccSalary <- jobs[,4][ (jobs[,2]==1) & ( (jobs[,1]==1) |
       (jobs[,1]==2) ) ]     # |' stands for logic operation or'
```

To remove (unwanted) objects:

```
> rm(mSalary, fSalary, accSalary, finSalary, maccSalary,
          faccSalary, mfinaccSalary)
```

2.2 Probability

Definition. Probability a function P that assigns a real number P(A) to each event in a sample space, which satisfies the three conditions:

(i) $P(A) \geq 0$ for any event A,
(ii) $P(\Omega) = 1$, and
(iii) For disjoint events $A_1, A_2, \cdots,\ P\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i)$.

Let $A_1 = \Omega,\ A_2 = A_3 = \cdots = \varnothing$. By (iii) and (ii), $P(\varnothing) = 0$.
Hence for any disjoint A and B, $P(A \cup B) = P(A) + P(B)$.

More properties of probability:

1. $P(A^c) = 1 - P(A)$.

2. If $A \subset B$, $P(B) = P(A) + P(B - A) \geq P(A)$.

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$.
4. Boole inequality: $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$.
5. If $A_n \to A$, $P(A_n) \to P(A)$.

Proof. 1. $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$.

3. $A \cup B = (AB) \cup (AB^c) \cup (A^c B)$, and the 3 events on the RHS are disjoint. Hence

$$P(A \cup B) = P(AB) + P(AB^c) + P(A^c B).$$

Since $A = (AB) \cup (AB^c)$, $P(A) = P(AB) + P(AB^c)$. Similarly $P(B) = P(AB) + P(A^c B)$. Therefore
$P(A \cup B) = P(AB) + \{P(A) - P(AB)\} + \{P(B) - P(AB)\} = P(A) + P(B) - P(AB)$.

4. is obtained by applying 3. repeatedly.

The proof of 5. is a bit more involved, we refer to p.7 of Wasserman (2004).

Example 5. Toss a fair 6-sided die, there are 6 possible outcomes each with probability 1/6. If we toss it two times, the sample space is $\Omega = \{(i,j) : i,j = 1,\cdots,6\}$. Since each outcome is equally likely,

$$P(A) = \frac{\text{No. of elements in } A}{36}, \qquad A \in \Omega.$$

For example, $P(A) = 2/36 = 1/18$ for $A = \{\text{the sum is 3}\}$, and $P(A) = 3/36 = 1/12$ for $A = \{\text{the sum is 4}\}$.

2.3 Independence

Definition. k events $A_1, \cdots, A_k$ are independent if

$$P(A_{i_1} A_{i_2} \cdots A_{i_j}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_j})$$

for any $1 \leq i_1 < i_2 < \cdots < i_j \leq k$ and $2 \leq j \leq k$.

Intuition. If A and B are independent, the occurrence of A has nothing to do with the occurrence of B. For example, two persons toss two coins: two outcomes are independent with each other.

Example 6. Toss a fair coin 10 times. Let A = "at least one head". Define T_j be the event that tail occurs on the j-th toss. Then

$$\begin{aligned} P(A) &= 1 - P(A^c) = 1 - P(T_1 \cdots T_{10}) = 1 - P(T_1)P(T_2)\cdots P(T_{10}) \\ &= 1 - (0.5)^{10} \approx 0.9999. \end{aligned}$$

Example 7. John and Peter play each other in the final of a tennis tournament. Whoever wins 2 out of 3 games will win the tournament. Suppose that John is higher ranked player who beats Peter in a single game with probability 0.6, and each game will be played independently. Find the probability that John will win the tournament.

Let A_i ="John wins the i-th game", and A ="John wins the tournament". Then

$$A = (A_1A_2) \cup (A_1A_2^cA_3) \cup (A_1^cA_2A_3),$$

and the 3 events on the RHS are disjoint. Hence

$$\begin{aligned} P(A) &= P(A_1 A_2) + P(A_1 A_2^c A_3) + P(A_1^c A_2 A_3) \\ &= (0.6)^2 + 2 \times (0.6)^2 \times 0.4 = 0.648, \end{aligned}$$

which is greater than the probability for John to win a single game.

Question. Would John prefer to play the maximum 5 (instead of 3) games in the final?

2.4 Conditional Probability

Example 8. Five people take one ball each out of a bag containing 4 white balls and one red ball.

Obviously the Probability for the 1st person to take the red ball is 1/5. What is the probability for the 2nd, 3rd, 4th or the last person to take the red ball?

Definition. If $P(B) > 0$, the conditional probability of A given B is

$$P(A|B) = P(AB)/P(B),$$

which is the probability of event A given the condition that event B occurs already.

Remark. (i) If A and B are independent, $P(A|B) = P(A)$.

(ii) In general $P(AB) = P(A|B)P(B)$.

Example 8. (Continue)

$$\begin{aligned}
&P(\text{2nd person takes R}) = P(\text{1st person takes W, 2nd Person take R})\\
&= P(\text{1st person takes W}) \times P(\text{2nd person takes R}|\text{1st person takes W})\\
&= \frac{4}{5} \times \frac{1}{4} = 1/5,
\end{aligned}$$

which is the same as the probability for the 1st person to take the red.

Let $A_1, \cdots, A_k$ be a partition of Ω.

Law of Total Probability. For any event B,

$$P(B) = P(BA_1) + \cdots + P(BA_k).$$

Proof. $B = B\Omega = B(\cup_i A_i) = \cup_i (BA_i)$. Since $BA_1, \cdots, BA_k$ are disjoint, the law holds.

Bayes' Formula. Let $P(B) > 0$ and $P(A_i) > 0$ for $i = 1, \cdots, k$. Then

$$P(A_j|B) = P(B|A_j)P(A_j) \Big/ \sum_{i=1}^{k} P(B|A_i)P(A_i).$$

Proof. $P(A_j|B) = P(A_jB)\big/P(B) = P(B|A_j)P(A_j)\big/P(B)$. Replacing $P(B)$ using the law of total probability, we obtain Bayes' Formula.

Example 9. Larry divides his emails into 3 categories: A_1 ="spam", A_2 ="low priority" and A_3 ="high priority". From previous experience he concludes

$$P(A_1) = 0.7, \quad P(A_2) = 0.2, \quad P(A_3) = 0.1.$$

Let B be the event that an email contains the word 'free'. Again based on previous experience,

$$P(B|A_1) = 0.9, \quad P(B|A_2) = 0.1, \quad P(B|A_3) = 0.1.$$

He receives a new email with word 'free'. What is the probability that it is spam?

By Bayes' theorem,

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} = 0.955.$$