Probability & Statistics

Dongjiao Ge
djge@cityu.edu.mo

Session 01

Assessment

- **Attendance: 30% share**
- **Mid-term group assignment: 30% share**
- **Exam (open-book exam): 40% share** --- You can bring the printed lecture notes and the notes (hard copies only!) you prepared for the exam. Any electronic devices are strictly banned during the exam.

Part 1
Probability

Reading list:
https://webspace.maths.qmul.ac.uk/p.j.cameron/notes/prob.pdf
https://www.math.ucdavis.edu/~gravner/MAT135A/resources/lecturenotes.pdf
Lecture notes for Probability & Statistics II (QMUL).

Note that this lecture note is made based on the above materials.

This module is called “probability and statistics”, so you may ask what is probability/statistics. While, may be firstly to ask what is “probability”.

------ We won’t answer this question at the moment

At the very beginning of this module, we will look at

- some basic **axioms** which probability must satisfy, and making deductions from these.
- different kinds of **sampling**, and examine what it means for events to be independent

Chapter 1 Preparation

1.1 Sample space, events

*We perform an **experiment** which can have many different outcomes.*
*The **sample space** is **the set of all possible outcomes of the experiment**.*

We usually S or Ω to present the sample space.

Example: I plant ten bean seeds and count the number that germinate, the sample space is

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}.$$

I toss a coin three times and record the result, the sample space is

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

Remark: H -- head, T -- tail

Remarks:

- Sometimes we can assume that all the outcomes are equally likely.

  (Don't assume this unless either you are told to, or there is some physical reason for assuming it. In the beans example, it is most unlikely. In the example of tossing the coin, the assumption will hold if the coin is 'fair'.)

  <u>Anyway, Don't just assume that all outcomes are equally likely, especially when you are given enough information to calculate their probabilities!</u>

- If all outcomes are equally likely, then each has probability $1/|S|$.
  ($|S|$ is the number of elements in the set S).

An **event** is a subset of S.

We can specify an event by **listing all the outcomes that make it up**.

In the coin tossing example, let A be the event '*more heads than tails*' and B the event '*heads on last throw*'. Then,

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

$$\begin{aligned} A &= \{HHH, HHT, HTH, THH\}, \\ B &= \{HHH, HTH, THH, TTH\}. \end{aligned}$$

The **probability of an event** is calculated by *adding up the probabilities of all the outcomes comprising that event*. So, if all outcomes are equally likely, we have

$$P(A) = \frac{|A|}{|S|}.$$

An **event** is *simple* if it consists of **just a single outcome**, and is *compound* otherwise.

In the coin-tossing example, A and B are *compound* events.
The event ‘heads on every throw’ is *simple* (as a set, it is $\{HHH\}$).

If $A = \{a\}$ is a **simple event**, then the **probability of A** is just the probability of the outcome a, and we usually write $P(a)$, which is simpler to write than $P(\{a\})$.

We can build new events from old ones:

- $A \cup B$ (read ‘A union B’) consists of all the outcomes in A or in B (or both!)
- $A \cap B$ (read ‘A intersection B’) consists of all the outcomes in both A and B;
- $A \setminus B$ (read ‘A minus B’) consists of all the outcomes in A but not in B;
- A' (read ‘A complement’) consists of all outcomes not in A (that is, $S \setminus A$);
- $\emptyset$ (read ‘empty set’) for the event which doesn’t contain any outcomes.

Remember an event is a set!

Q1 (15 minutes) Roll two dice. What is the most likely sum?

sum 2----12

点数和最大可能为7
（1，6）（2，5）（3，4）（6，1）（5，2）（4，3）【the number of outcomes】

1.2 What is probability?

Many different definitions have tried to describe the probability

- **Some people think of it as '*limiting frequency*'.**
  That is, to say that the probability of getting heads when a coin is tossed means that, if the coin is tossed many times, it is likely to come down heads about half the time.
  Tossing a coin 1000 times, you are not likely to get exactly 500 heads.

- **Some people would say that you can work out probability by physical arguments, like the one we used for a fair coin**.
  But this argument doesn't work in all cases, and it doesn't explain what probability means.

We regard probability as a mathematical construction satisfying some axioms (devised by the Russian mathematician A. N. Kolmogorov). --- a mathematical definition

What are those axioms that define the probability?

1.3 Kolmogorov's Axioms

Kolmogorov's axioms tell us each event A has a probability $P(A)$, which is a real number. These numbers satisfy three axioms:

Axiom 1: For any event A, we have $P(A) \geq 0$.

Axiom 2: $P(S) = 1$.

Axiom 3: If the events $A_1, A_2, \ldots$ are pairwise disjoint, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

$\forall i\, j \in N,\ A_i \cap A_j = \phi$

Remark 1: Axiom 3a if there are only finitely many events $A_1, A_2, \ldots, A_n$, so that we have

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

Remark 2: A number of events say $A_1, A_2, \ldots,$ are called **mutually disjoint** or **pairwise disjoint** if $A_i \cap A_j = \phi$ for any two of the events A_i and A_j; that is, no two of the events overlap.

1.4 Proving things from the axioms

Here are some examples of things that can be proved from the axioms.

Proposition 1.1 *If the event A contains only a finite number of outcomes, say* $A = \{a_1, a_2, \ldots, a_n\}$, *then*

$$P(A) = P(a_1) + P(a_2) + \cdots + P(a_n).$$

Proof

To prove the proposition, we define a new event A_i containing only the outcome a_i, that is, $A_i = \{a_i\}$, for $i = 1, \ldots, n$. Then $A_1, \ldots, A_n$ are mutually disjoint (each contains only one element which is in none of the others), and $A_1 \cup A_2 \cup \cdots \cup A_n = A$; so by Axiom 3a, we have

$$P(A) = P(a_1) + P(a_2) + \cdots + P(a_n).$$

Corollary 1.2 *If the sample space $\mathcal{S}$ is finite, say $\mathcal{S} = \{a_1, \ldots, a_n\}$, then*

$$P(a_1) + P(a_2) + \cdots + P(a_n) = 1.$$

Proof

Proposition 1.1 $\Longrightarrow$ $P(a_1) + P(a_2) + \cdots + P(a_n) = P(\mathcal{S})$

Axiom 2 $\Longrightarrow$ $P(\mathcal{S}) = 1$

If all the n outcomes are equally likely, and their probabilities sum to 1, then each has probability $1/n$, that is, $1/|S|$.

Going back to Proposition 1.1, we see that, if *all outcomes are equally likely*, then

$$P(A) = \frac{|A|}{|\mathcal{S}|}$$

for any event A, justifying the principle we used earlier.

Proposition 1.3 $P(A') = 1 - P(A)$ *for any event* A.

Proof

Let $A_1 = A$ and $A_2 = A'$ (the complement of A). Then $A_1 \cap A_2 = \emptyset$ (that is, the events A_1 and A_2 are disjoint), and $A_1 \cup A_2 = \mathcal{S}$. So

$$\begin{aligned} P(A_1) + P(A_2) &= P(A_1 \cup A_2) \quad \text{(Axiom 3)} \\ &= P(\mathcal{S}) \\ &= 1 \quad \text{(Axiom 2).} \end{aligned}$$

So $P(A) = P(A_1) = 1 - P(A_2)$.

Corollary 1.4 $P(A) \leq 1$ *for any event A.*

Proof

For $1 - P(A) = P(A')$ by Proposition 1.3, and $P(A') \geq 0$ by Axiom 1; so $1 - P(A) \geq 0$, from which we get $P(A) \leq 1$.

Corollary 1.5 $P(\emptyset) = 0$.

Proof

For $\emptyset = \mathcal{S}'$, so $P(\emptyset) = 1 - P(\mathcal{S})$ by Proposition 1.3; and $P(\mathcal{S}) = 1$ by Axiom 2, so $P(\emptyset) = 0$.

Proposition 1.6 *If $A \subseteq B$, then $P(A) \leq P(B)$.*

Proof

This time, take $A_1 = A$, $A_2 = B \setminus A$. Again we have $A_1 \cap A_2 = \emptyset$ (since the elements of $B \setminus A$ are, by definition, not in A), and $A_1 \cup A_2 = B$. So by Axiom 3,

$$P(A_1) + P(A_2) = P(A_1 \cup A_2) = P(B).$$

In other words, $P(A) + P(B \setminus A) = P(B)$. Now $P(B \setminus A) \geq 0$ by Axiom 1; so

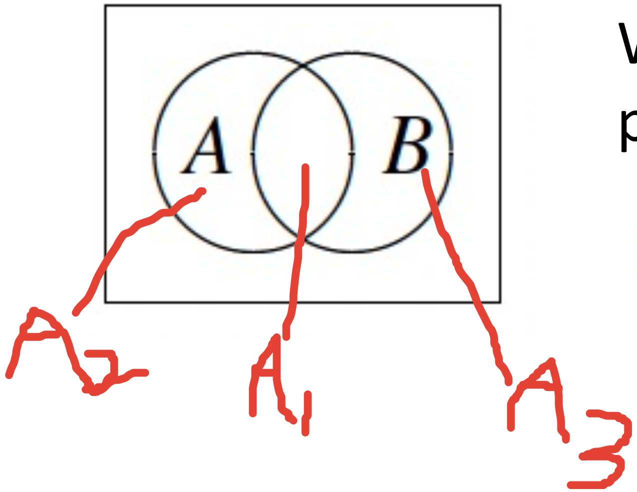
$$P(A) \leq P(B),$$

as we had to show.

1.5 Inclusion-Exclusion Principle

Proposition 1.7

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



We can use Venn diagram like this to help us to prove the above proposition

$A \cup B$ is made up of three parts, namely

$$A_1 = A \cap B, \qquad A_2 = A \setminus B, \qquad A_3 = B \setminus A.$$

We have $A \cup B = A_1 \cup A_2 \cup A_3$.

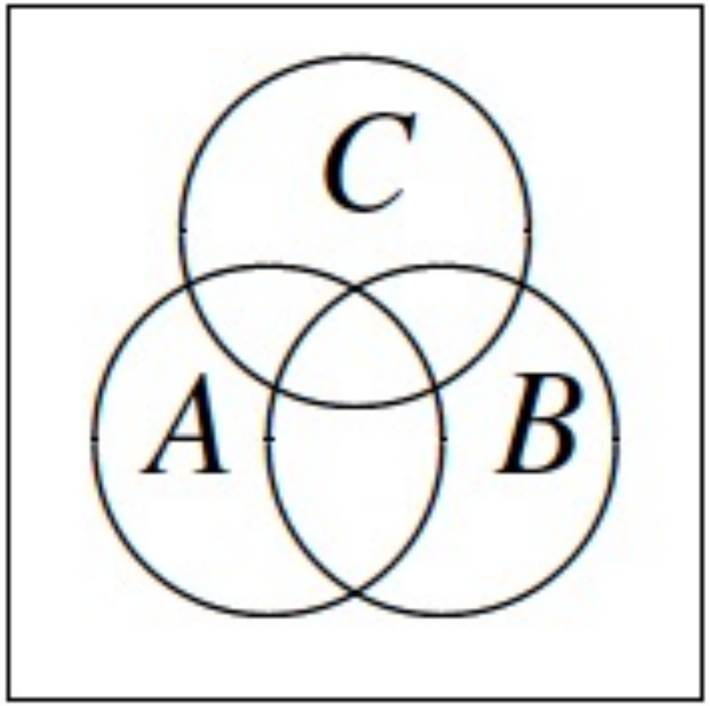
The sets A_1, A_2, A_3 are mutually disjoint. 互不相交

So, by Axiom 3, we have

$$\begin{aligned} P(A) &= P(A_1)+P(A_2), \\ P(B) &= P(A_1)+P(A_3), \\ P(A\cup B) &= P(A_1)+P(A_2)+P(A_3). \end{aligned}$$

From this we obtain

$$\begin{aligned} P(A)+P(B)-P(A\cap B) &= (P(A_1)+P(A_2))+(P(A_1)+P(A_3))-P(A_1) \\ &= P(A_1)+P(A_2)+P(A_3) \\ &= P(A\cup B) \end{aligned}$$



The Inclusion-Exclusion Principle extends to more than two events, but more complicated.

Proposition 1.8 *For any three events A, B, C, we have*

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

1.6 Other results about sets

Let us have a look at some examples that other standard results about sets which are useful in probability theory.

Proposition 1.9 *Let A, B, C be subsets of $\mathcal{S}$.*

Distributive laws: $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ *and* $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

De Morgan's Laws: $(A \cup B)' = A' \cap B'$ *and* $(A \cap B)' = A' \cup B'$.

1.7 Sampling

We have four pens in the desk drawer; they are **R**ed, **G**reen, **B**lue, and **P**urple. We draw a pen; each pen has the same chance of being selected. So, we have the sample space $S = \{R, G, B, P\}$

A is the event 'red or green pen chosen', then

$$P(A) = \frac{|A|}{|S|} = \frac{2}{4} = \frac{1}{2}.$$

More generally, if I have a set of n objects and choose one, with each one equally likely to be chosen, then each of the n outcomes has probability $1/n$, and an event consisting of m of the outcomes has probability m/n.

What if we choose more than one pen/object?

We need to be more careful to specify the sample space.
Firstly, we need to be clear about whether we are

- *sampling with replacement*
- *sampling without replacement*

Sampling with replacement means that we choose a pen, note its colour, put it back and shake the drawer, then choose a pen again (which may be the same pen as before or a different one), and so on until the required number of pens have been chosen. If we choose **two pens** with replacements, the sample space is

$$\begin{array}{llll} \{RR, & RG, & RB, & RP, \\ GR, & GG, & GB, & GP, \\ BR, & BG, & BB, & BP, \\ PR, & PG, & PB, & PP\} \end{array}$$

*The first draw taking one from **4**.*
e.g., we take the purple one
*(**Then we put it back!**)*

*The second draw we still taking one from **4**.*

The event ‘at least one red pen’ is $\{RR, RG, RB, RP, GR, BR, PR\}$, and has probability $7/16$.

Sampling without replacement means that we choose a pen but do not put it back so that our final selection cannot include two pens of the same colour. In this case, the sample space for choosing two pens is

$$\begin{array}{cccc} \{ & RG, & RB, & RP, \\ GR, & & GB, & GP, \\ BR, & BG, & & BP, \\ PR, & PG, & PB & \} \end{array}$$

The first draw taking one from 4. e.g., we take the purple one

The second draw taking one from thee remaining 3.

and the event ‘at least one red pen’ is $\{RG, RB, RP, GR, BR, PR\}$, with probability $6/12 = 1/2$.

Another issue is that whether we care about the order in which the pens are chosen. We will only consider this in the case of sampling without replacement.
So, if we think don't care about the order, i.e., e.g., {R,G}={G,R} , in this case the sample space is

$$\{\{R,G\},\{R,B\},\{R,P\},\{G,B\},\{G,P\},\{B,P\}\}$$

Only **six** elements then!

The event 'at least one red pen' is $\{\{R,G\},\{R,B\},\{R,P\}\}$, with probability 3/6 = 1/2.

We can use the following formulas to get the size of the sample space

$$\begin{aligned} n! &= n(n-1)(n-2)\cdots 1 \\ {}^{n}P_k &= n(n-1)(n-2)\cdots(n-k+1) \\ {}^{n}C_k &= {}^{n}P_k/k! \end{aligned}$$

Permutation

Combination

排列数与组合数

Note that $n!$ is the product of all the whole numbers from 1 to n; and

$${}^{n}P_k = \frac{n!}{(n-k)!},$$

so that

$${}^{n}C_k = \frac{n!}{k!(n-k)!}.$$

Theorem 1.10 *The number of selections of k objects from a set of n objects is given in the following table.*

	with replacement	*without replacement*
ordered sample	n^k	${}^nP_k$
unordered sample		${}^nC_k$

First, for sampling with replacement and ordered samples, there are n choices for the first object and, n choices for the second, and so on; we multiply the choices for different objects. The product of k factors each equal to n is n^k.

For *sampling without replacement and ordered samples*, there are still n choices for the first object, but now only $n-1$ choices for the second (since we do not replace the first), and $n-2$ for the third, and so on; there are $n-k+1$ choices for the k-th object, since $k-1$ have previously been removed and $n-(k-1)$ remain. We multiply them and get ${}^nP_k$.

For *sampling without replacement and unordered sample*, think first of choosing an ordered sample, which we can do in ${}^nP_k$ ways. But each unordered sample could be obtained by drawing it in $k!$ different orders. So we divide by $k!$ to remove the order, obtaining ${}^nP_k/k! = {}^nC_k$ choices.

*Note that, if we use the phrase 'sampling without replacement, ordered sample', or any other combination, **we are assuming that all outcomes are equally likely**.*

Example A six-sided die is rolled twice. What is the probability that the sum of the numbers is at least 10?

We are **sampling with replacement**, since the two numbers may be the same or different. So the number of elements in the sample space is $6^2 = 36$.

To obtain a sum of 10 or more, the possibilities for the two numbers are (4,6), (5,5), (6,4), (5,6), (6,5) or (6,6). So the probability of the event is $6/36 = 1/6$.

Example A box contains 20 balls, of which 10 are red and 10 are blue. We draw ten balls from the box, and we are interested in the event that exactly 5 of the balls are red and 5 are blue. Do you think that this is more likely to occur if the draws are made with or without replacement?

Compare the probabilities

Let S be the sample space, and A the event that five balls are red and five are blue.

Sampling with replacement: $|S| = 20^{10}$.

$|A|$? The number of ways in which we can choose first five red balls and then five blue ones (that is, RRRRRBBBBB), is $10^5 \cdot 10^5 = 10^{10}$. But there are many other ways to get five red and five blue balls. In fact, the five red balls could appear in any five of the ten draws. This means that there are ${}^{10}C_5 = 252$ different patterns of five Rs and Bs.

$$|A| = 252 \cdot 10^{10}, \qquad P(A) = \frac{252 \cdot 10^{10}}{20^{10}} = 0.246.$$

Sampling without replacement

If we regard **the sample as being ordered**, then $|S| = {}^{20}P_{10}$

${}^{10}P_5$ --- The number of ways to choose five of the ten red balls, and the same for the ten blue balls.

${}^{10}C_5$ --- The number of different patterns of five red balls and blue balls.

$$|A| = ({}^{10}P_5)^2 \cdot {}^{10}C_5, \quad P(A) = \frac{({}^{10}P_5)^2 \cdot {}^{10}C_5}{{}^{20}P_{10}} = 0.343\,.$$

If we regard **the sample as being ~~ordered~~**, then $|S| = {}^{20}C_{10}$

${}^{10}C_5$ --- The number of choices for the five red/ blue balls.

We no longer have to count the patterns of the balls since their locations no longer matter.

$$|A| = ({}^{10}C_5)^2, \qquad P(A) = \frac{({}^{10}C_5)^2}{{}^{20}C_{10}} = 0.343\,.$$

<u>Our final answer should be that the event is more likely if we sample without replacement.</u>

In a sampling problem, you should ……

- decide whether the sampling is **with or without replacement**
- If it is **without replacement**, decide whether the sample is ordered (e.g. does the question say anything about the first object drawn?).
    - If so, then use the **formula for ordered samples**.
    - If not, then use **either ordered or unordered samples**, whichever is convenient; they should give the same answer.
- If it is **with replacement**, or if it involves throwing a die or coin several times, then use the **formula for sampling with replacement**.

Q2 A bag has 6 pieces of paper, each with one of the letters, E, E, P, P, P, and R, on it. Pull 6 pieces at random out of the bag (1) without, and (2) with replacement. What is the probability that these pieces, in order, spell PEPPER? (You should consider both with and without replacement cases.)

Exercises

EX1. A full deck of 52 cards contains 13 hearts. Pick 7 cards from the deck at random (a) without replacement and (b) with replacement. In each case, compute the probability that you get no hearts.

EX2 Compute the probability of a 2:2 boy-girl split in a four-children family. Please give the sample space too.

$P(2:2)=6/16=3/8$