

Seminar 2

```
In [1]: import numpy as np
import pandas as pd

import IPython.display as dp
import matplotlib.pyplot as plt
import seaborn as sns

dp.set_matplotlib_formats("retina")
sns.set(style="whitegrid", font_scale=1.5)
sns.despine()

%matplotlib inline
```

```
/var/folders/33/j0cl7y453td68qb96j7bqcj4cf41kc/T/ipykernel_40555/321256244
3.py:8: DeprecationWarning: `set_matplotlib_formats` is deprecated since I
Python 7.23, directly use `matplotlib_inline.backend_inline.set_matplotlib
_formats()`
  dp.set_matplotlib_formats("retina")
```

Last time

- Product rule for sampling with replacement when order matters
- Factorial rule for sampling without replacement when order matters
- Naive definition of probability
- Finding the probability of complement

This time

- Binomial coefficient for sampling without replacement when order doesn't matter
- "Stars and bars" for sampling with replacement where the order doesn't matter
- Axiomatic definition of probability

Binomial coefficient

A binomial coefficient counts the number of subsets of a certain size for a set, such as the number of ways to choose a committee of size k from a set of n people. Sets and subsets are by definition unordered, e.g., $\{3, 1, 4\} = \{4, 1, 3\}$, so we are counting the number of ways to choose k objects out of n , without replacement and without distinguishing between the different orders in which they could be chosen.

For any nonnegative integers k and n , the binomial coefficient $\binom{n}{k}$, read as " n choose k ", is the number of subsets of size k for a set of size n . For $k \leq n$,

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

Problem 1

How many paths are there from the point (0,0) to the point (110,111) in the plane such that each step either consists of going one unit up or one unit to the right?

Solution 1

Reminder: We will encode a path as a sequence of letters U (for up step) and R (for right step), like $URURURU \dots UURUR$. The sequence must consist of 110 R s and 111 U s.

Note that to fully describe the sequence we actually only need to specify where the R s are located. This falls under binomial coefficient definition. So there are

$\binom{110+111}{110}$ possible paths.

Problem 2

How many ways are there to split a dozen people into 3 teams, where each team has 4 people?

Solution 2

Let's randomly pick the first team, then randomly pick the second and claim the remaining people the third team.

This gives us $\binom{12}{4} \cdot \binom{8}{4}$ possibilities. Is it correct?

It is not correct, because we overcounted due the fact that we do not actually care which team is the first, second or third. So we need to divide the expression by $3!$.

The final answer is:

$$\frac{1}{3!} \cdot \binom{12}{4} \cdot \binom{8}{4} = \frac{1}{3!} \cdot \frac{12!}{4!8!} \cdot \frac{8!}{4!4!} = \frac{12!}{4!4!4!3!}$$

If we cared which team is which, we would obtain $\frac{12!}{4!4!4!}$, which is called a **multinomial coefficient**. The only difference is that we choose more than one subset from one total.

Problem 3

How many ways are there to put 10 balls into 3 boxes?

Solution 3

Let's visualize the balls as o and borders of boxes as |:

|oooo|oooo|oo|

How many balls do we have? How many borders do we have? How many total objects do we have?

We have:

- 10 balls
- 3 + 1 borders. Note however, that 2 borders are fixed to be leftmost and rightmost. So the free borders are actually 3 - 1.
- 10 + 3 - 1 objects total

Then, we need to choose the positions of the borders:

$$\binom{10 + 3 - 1}{3 - 1} = \binom{n + k - 1}{k - 1}$$

Recap of counting

Sampling k objects from n choices:

With replacement	Order matters	Formula	
Yes	Yes	n^k	(1)
Yes	No	$\binom{n + k - 1}{k - 1}$	(2)
No	Yes	$n!$	(3)

No

No

$$\binom{n}{k}$$

(4)

Problem 4

There are 100 passengers lined up to board an airplane with 100 seats (with each seat assigned to one of the passengers). The first passenger in line crazily decides to sit in a randomly chosen seat (with all seats equally likely). Each subsequent passenger takes their assigned seat if available, and otherwise sits in a random available seat. What is the probability that the last passenger in line gets to sit in their assigned seat?

Solution 4

Denote i -th passenger true seat as i , regardless of its position in the plane.

Notice that there is always only one seat occupied incorrectly. Then consider a passenger j . He could sit in any other available seat, but he decided to sit into seat 1. If he sits into seat 1, it means that his place j is taken.

pax\seat	1	2	3	4	5
			1		
	2	1			
	2	1	3		
	2	1	3	4	
	5	2	1	3	4

pax\seat	1	2	3	4	5
			1		
	2	1			
	2	1	3		
	4	2	1	3	
	4	2	1	3	5

pax\seat	1	2	3	4	5
					1

	2	1			
	2	1		3	
	2	1	4	3	
	5	2	1	4	3
pax\seat					
	1	2	3	4	5
			1		
		2	1		
	3	2	1		
	3	2	1	4	
	3	2	1	4	5

By sitting into seat 1, he removes the source of permutation, because now all seats are occupied correctly. After that, all the passengers that enter the plane will be able to sit in their true seats. It is important that it always happens and can happen with any passenger j .

Generally, the last 100-th passenger may observe two cases:

- The permutation was removed, then he has the option to sit into his true 100-th seat
- The permutation was not removed, then he is the one to remove the permutation and take seat 1.

We can now reduce the problem to just two seats: 1-st and 100-th. One of the passengers seating on these seats is the last 100-th passenger, the other is any other passenger j .

Since $j < 100$, it means that both 1-st and 100-th seats were empty when he or she boarded the plane! And the probabilities to sit in any of them is equal.

If j sat in 1 then the last passenger ended up sitting in 100 and the resulting configuration of passengers sitting in the 100 seats is the same as if j had sat in 100 except for the fact that the passengers in 1 and 100 are swapped. Therefore these two configurations occur with the same probability and exactly one of them has the last passenger in his or her seat 100.

This implies that all the final configurations of passengers can be paired such that the two configurations in any pair occur with the same probability and exactly one has the last passenger in his or her seat.

This implies that the probability that the last passenger is in her seat is 0.5.

Non-naive definition

Definition

A probability space consists of a sample space S and a probability function P which takes an event $A \subseteq S$ as input and returns $P(A)$, a real number between 0 and 1, as output. The function P must satisfy the following axioms:

- $P(\emptyset) = 0, P(S) = 1$
- If A_1, A_2, \dots are disjoint ($A_i \cap A_j = \emptyset, i \neq j$) events, then

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

Properties

1. $P(A^c) = 1 - P(A)$
2. If $A \subseteq B$, then $P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Problem 5

Consider set S of all (how many?) subsets of set $M = \{1, \dots, N\}$. We take two sets randomly and independently two sets $A, B \in S$. Find the probability that $A \cap B = \emptyset$.

Solution 5

Take any element $e \in M$ from the original set.

- $\mathbb{P}(e \in A) = p_1 = \frac{1}{2}$, by construction of A
- $\mathbb{P}(e \in B) = p_2 = \frac{1}{2}$, by construction of B
- $\mathbb{P}(e \in A \cap B) = p_{12} = p_1 \cdot p_2 = \frac{1}{4}$
- $\mathbb{P}(e \notin A \cap B) = 1 - \mathbb{P}(e \in A \cap B) = 1 - p_{12} = \frac{3}{4}$

Repeat for every $e \in M$ to obtain:

$$\mathbb{P}(A \cap B = \emptyset) = \left(\frac{3}{4}\right)^N$$

Inclusion-exclusion formula

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

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_i P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^r$$