DSC 40A

Theoretical Foundations of Data Science I

Today

- More examples of using combinatorics to solve probability questions.

HW 6 released today, due next Mon

Question

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!

If the direct link doesn't work, click the "Lecture Questions" link in the top right corner of dsc40a.com.

Counting as a Tool for Probability

Example 8. What is the probability that a randomly generated bitstring of length 10 is the string 0011001101?

0 0 1 1 _ _ _ _ _ _

bit: '0' or '1'

2 options

each bit has prob $\frac{1}{2}$

possible bit strings of length $10 = 2^{10}$

uniform probability = all equally likely

$$P = \frac{1}{2^{10}} = \left(\frac{1}{2}\right)^{10}$$

$$P(n,k) = (n)(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

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

Counting as a Tool for Probability

Example 7. What is the probability that a randomly generated bitstring of length 10 contains an equal number of zeros and ones?

0 0 0 0 0 1 1 1 1 1

0 1 0 1 0 1 0 1 0 1

_ _ _ _ _ _ _ _ _ _

5 0's 5 1's in 10 positions

5 positions for 0's = {1, 3, 5, 7, 10} ⟺ 0 1 0 1 0 1 0 1 1 0

{1, 2, 3, 4, 5}

$$\frac{\#\text{ sets of 5 numbers out of 10}}{\#\text{ bitstrings of length 10}} = \frac{C(10,5)}{2^{10}}$$

$n = 10$, $k = 5$

$$P(n,k) = (n)(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

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

Why is the positions of 5 0's a set, not a sequence?

The following sets are all the same:

$$\{1, 3, 5, 7, 10\} = \{5, 3, 1, 7, 10\} = \{10, 7, 5, 3, 1\} = \dots$$

The order in which the positions are specified doesn't matter, therefore it's a set and it's equivalent to:

bitstring:	0	1	0	1	0	1	0	1	1	0
	1		3		5		7			10

How to solve using permutations?

(Recall:

$\#\text{sequences} = \#\text{sets} \cdot \#\text{orderings}$)

10 options, 9 options, 8 options, . . . 1 option

we have to place: 5 0's 5 1's in 10 places _ _ _ _ _ _ _ _ _ _

so we could calculate all possible permutations $= 10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$

However since the objects we are permuting (bits) aren't unique note the following:

Let's start out with 0 0 0 0 0 1 1 1 1 1

if we swap first and last bits we get

1 0 0 0 0 1 1 1 1 0

which is a different string from the previous one we had

if we shuffle the first two bits we get

0 0 0 0 0 1 1 1 1 1 which is the same

So all permutations of the 0's amongst themselves are equivalent $= 5!$

and all permutations of the 1's amongst themselves are equivalent $= 5!$

Therefore # of unique bitstrings $= \dfrac{10!}{5!\,5!} = C(10,5)$

which is the same answer as using sets

Different version of this question

1) We have a string of length 10 composed of characters from a, b, c, d, e, f, g, h, i, j

What is the number of strings of length 10 such that all characters in the string are unique.

Ex: abcdefghij , jihgfedcba but not bbccadefji

Answer _ _ _ _ _ _ _ _ _ _ $\Rightarrow 10 \cdot 9 \cdot 8 \cdot \_ \ldots = 10!$

2) We have a string of length 10 composed of characters from a, b, c, d, e, f

What is the number of strings that have 5 a's and the rest of the characters are all different?

ex: aaaaabcdef abacadaeaf

<u>Solution 1:</u> How many permutations? 10! (In first slot we select from 5 a's, 1 b, 1 c, 1 d, 1 e 1 f)

How many of these are unique strings?

There are 5! orderings of the 5 a's

$$\left.\begin{array}{c} a\,a\,a\,a\,a \\ a\,a\,a\,a\,a \\ a\,a\,a\,a\,a \\ a\,a\,a\,a\,a \\ \vdots \end{array}\right\} = a\ a\ a\ a\ a$$

Therefore: $\frac{10!}{5!} = P(10, 5)$

<u>Solution 2:</u> If we select the positions for b, c, d, e, f then the remaining are all a's

To assign positions we need to select at random a <u>sequence</u> of 5 numbers from 1-10 without replacement

$\Rightarrow P(10, 5)$

$(1, 2, 3, 4, 5) \Rightarrow$ b c d e f a a a a a

$(5, 4, 3, 2, 1) \Rightarrow$ f d e c b a a a a a

$(2, 4, 6, 8, 10) \Rightarrow$ a b a c a d a e a f

Counting as a Tool for Probability

Example 10. What is the probability that a fair coin flipped 10 times turns up HHTTHHTTHT?

_ _ _ _ _ _ _ _ _ _

H or T

$\frac{1}{2}$

$$P = \left(\frac{1}{2}\right)^{10} = \frac{1}{2^{10}}$$

uniform distribution

$$P(n,k) = (n)(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

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

Counting as a Tool for Probability

Example 9. What is the probability that a fair coin flipped 10 times turns up an equal number of heads and tails?

Set of positions that H's are in:

$n = 10$ possible positions

$k = 5$ positions for H's $\to$ this determines T's

$$\text{prob} = \frac{C(10,5)}{2^{10}} = C(10,5) \cdot \left(\frac{1}{2}\right)^{10} = C(10,5) \cdot \left(\frac{1}{2}\right)^{5} \left(\frac{1}{2}\right)^{5}$$

Example: 6H, 4T: $\frac{C(10,6)}{2^{10}} = \frac{C(10,4)}{2^{10}}$

General principle

$$C(n,k) = C(n, n-k)$$

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

$$P(n,k) = (n)(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

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

Counting as a Tool for Probability

Example 12. What is the probability that a biased coin with $Prob(H) = \frac{1}{3}$ flipped 10 times turns up HHTTHHTTHT?

not uniform probability

$\cancel{\frac{1}{2^{10}}}$?

$$prob(HH \ldots H) = \left(\frac{1}{3}\right)^{10} \qquad prob(TT \ldots T) = \left(\frac{2}{3}\right)^{10}$$

H	H	T	T	H	H	T	T	H	T
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$

$$prob = \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5$$

$$P(n,k) = (n)(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!} \qquad C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Counting as a Tool for Probability

Example 11. What is the probability that a biased coin with $Prob(H) = \frac{1}{3}$ flipped 10 times turns up an equal number of heads and tails?

$S =$ coin toss sequences of length 10

$E =$ coin toss sequences of equal num of H's, T's

$prob(s \in E) = \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5$ as in previous slide

$$p(E) = \sum_{s \in E} p(s) = \underbrace{C(10,5)}_{\substack{\#\text{ outcomes in } E,\\ n=10,\ k=5}} \underbrace{\left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5}_{p(s)} \neq \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S} = \frac{|E|}{|S|}$$

not uniform prob

because $p(s) \neq \frac{1}{|S|}$

$$P(n,k) = (n)(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

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

Example 2: $\underbrace{6H, 4T}_{E}$ biased coin $p(H) = \frac{1}{3}$

$$p(s \in E) = \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^4$$

$$p(E) = C(10,6) \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^4$$

these are not equal

Example 3: 6T, 4H

$$p(s \in E) = \left(\frac{2}{3}\right)^6 \left(\frac{1}{3}\right)^4$$

$$p(E) = C(10,6) \left(\frac{2}{3}\right)^6 \left(\frac{1}{3}\right)^4$$

The Easy Way

Example 6. There are 20 students in a class. A computer program selects a random sample of students by drawing 5 students at random **without replacement**. What is the chance that a particular student is among the 5 selected students?

We solved twice

- sequences → complement
- sets

The Easy Way

Example 6. There are 20 students in a class. A computer program selects a random sample of students by drawing 5 students at random **without replacement**. What is the chance that a particular student is among the 5 selected students?

Another way to think of sampling without replacement:

1. randomly shuffle all 20 students
2. take the first 5

S = possible positions for student #17

E = student #17 is in the first 5

$|S| = 20 \quad |E| = 5 \quad Pr = \frac{5}{20} = \frac{1}{4}$

$$\boxed{\underset{1}{\_\_}\ \underset{2}{\_\_}\ \underset{3}{\_\_}\ \underset{4}{\_\_}\ \underset{5}{\_\_}}\ \_\_\ \_\_\ \cdots\ \_\_\ \_\_\ \underset{19}{\_\_}\ \underset{20}{\_\_}$$

from [Theory Meets Data](#) by Ani Adhikari, Chapter 4

Practice Problems

Example 13. You were one of 238 individuals who reported for jury duty. If 54 of these people will be assigned to a courtroom, what is the probability that you get assigned to a courtroom?

Sampling without replacement

sample space: 238 positions

shuffle and select first 54 $\Rightarrow$ prob $\frac{54}{238}$

Alternative: S is set of 54 chosen from 238

$C(238, 54)$ $\leftarrow$ denominator

Practice Problems

Example 13. You were one of 238 individuals who reported for jury duty. If 54 of these people will be assigned to a courtroom, what is the probability that you get assigned to a courtroom?

set of 54 including you $\Rightarrow$ need to select 53 out of 237

How many sets of 54 individuals include you?

A. C(238, 54)

B. C(237, 54)

C. C(238, 53)

D. C(237, 53)

$238-1$ $\quad$ $54-1$

$$\frac{C(237,53)}{C(238,54)} = \frac{54}{238}$$

Practice Problems

Example 14. You were one of 238 individuals who reported for jury duty. Suppose 28 of these individuals are doctors. If 54 of these people will be assigned to a courtroom, what is the probability that exactly 5 doctors get assigned to a courtroom?

S = sets of 54 chosen from 238 $\qquad |S| = C(238, 54)$

238 − 28 $\qquad$ 54 − 5

$$\text{Prob}\left(\begin{array}{c}\text{5 doctors selected} \\ \text{in 54 out of} \\ 238\end{array}\right) = \frac{\#\text{ sets in } S \text{ with 5 doctors}}{\#\text{ sets in } S} = \frac{\underbrace{C(28,5)}_{\text{choose 5 doctors}}\,\underbrace{C(210,49)}_{\text{choose remaining 49 non-doctors}}}{C(238,54)}$$

why product
$C(28,5)\cdot C(210,49)$? $\Longrightarrow$

d1, d2, d3, n1, n2, n3
d1, d2, d3, n4, n5, n7
d4, d5, d6, n4, n5, n7

Practice Problems

Example 15. What is the probability that your five-card poker hand is a straight?

Practice Problems

Example 16. Suppose you look at your first card as it is dealt, and you see that it is a Queen. What is the probability that your five-card hand is a straight?

Summary

- Counting is a useful tool for probability.
- Sometimes there's an easy way!
- **Next time:** Bayes Theorem