2.3 INDEPENDENCE

Conditional Probability: We tried to understand how information AFFECTS probabilities.

What if information didn't matter?

use definition

If A indep of B then

$$= \mathbb{P}(A \mid B) = P(A)$$

Ex:

Urn: 3R + 2G. We sample 2 balls with replacement. Let

B_1 = 1st ball red.

B_2 = 2nd ball red.

Is B_1 independent of B_2?

What if we sampled without replacement?

POLL (Do it first with, then without)

Is B_1 independent of B_2

Yes

No

The only way to know for sure is by computing ① $P(B_2)$ and $P(B_2 | B_1)$

Lets start with

$$P(B_2 | B_1) = \frac{P(B_2 \cap B_1)^{②}}{P(B_1)③}$$

With replacement:

$$P(B_2) = \frac{5 \cdot 3}{5 \cdot 5} = \frac{3}{5} \qquad P(B_1) = \frac{3}{5}$$

$$P(B_1 \cap B_2) = \frac{3 \cdot 3}{5 \cdot 5}$$

$$P(B_2 | B_1) = \frac{\frac{3 \cdot 3}{5 \cdot 5}}{3/5} = \frac{3}{5}.$$ ← What does this tell you?

Without replacement:

$$P(B_1) = \frac{3}{5}$$

$$P(B_1 \cap B_2) = \frac{3 \cdot 2}{5 \cdot 4}$$ ← without.

$$\Rightarrow \quad P(B_2 | B_1) =$$

Last thing to calculate is $P(B_2)$.

POLL

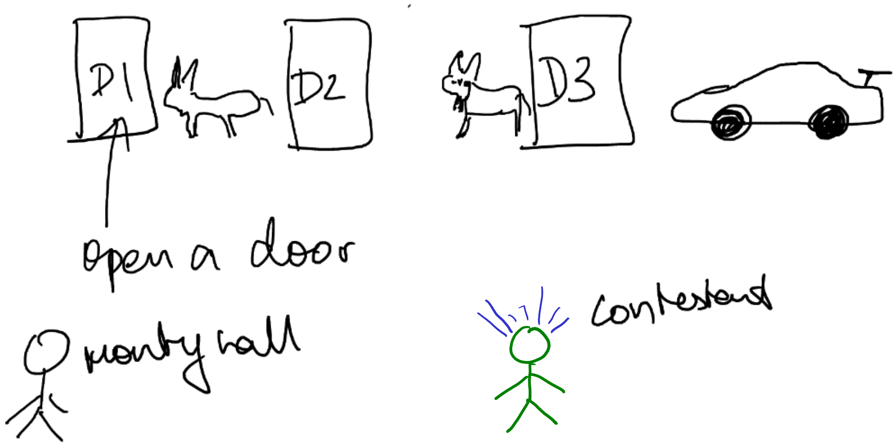
$P(B_1) \neq P(B_2)$	$P(B_2) = P(B_1)$

$$P(B_2) = P(B_2, B_1) + P(B_2, B_1^c)$$

$$= P(B_2 | B_1) P(B_1) + P(B_2 | B_1^c) P(B_1^c)$$

Remove one ball, DO NOT LOOK AT IT. THEN Remove your second ball. Now you could exchange the 1st ball and the 2nd ball. A little thinking shows that probabilities for the 2nd ball are the same as the first.

Famous example: Monty Hall Problem.



1) Contestant picks a door
2) Monty Hall shows them a door that has a goat behind it.
3) Now contestant has to decide whether or not to change their choice.

Argument 1 : There are two doors both equally likely to contain a car. There is no point in changing

Argument 2 : Might as well change, got nothing better to do.

Argument 3 : Label the doors $\{1,2,3\}$

$\Omega = \{(\text{car}, \text{choice}) : \text{car} \in \{1,2,3\}, \text{choice} \in \{1,2,3\}\}$

$|\Omega| =$

$C = \{\omega \in \Omega : \text{win by changing your choice}$

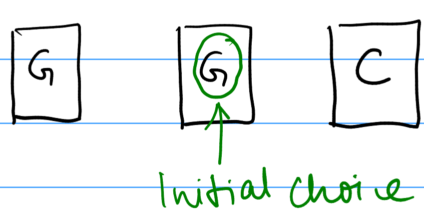
$NC = \{\omega \in \Omega : \text{win without changing your choice}\}$

$P(NC) = \mathbb{P}(\{(\text{car}, \text{choice}) : \text{car} = \text{choice}\})$

$=$

$P(C) = P((\text{car}, \text{choice}) : \text{car} \neq \text{choice})$

?



$=$

You can think about this another way.

$$P(C) = P(C \mid \text{choice} = \text{car})\, P(\text{choice} = \text{car})$$
$$+ P(C \mid \text{choice} \neq \text{car})\, P(\text{choice} \neq \text{car})$$

I read an article in the American mathematical monthly by someone who worked in "decision trees". He claims that he explained this problem to Paul Erdős and Erdős got the answer wrong.

In fact Erdős didn't believe the decision engineer's explanation, and finally accepted the result when the engineer showed him a simulation.

Later on, Erdős' friend Graham gave him a more mathematical explanation, and Erdős wrote back to the engineer saying that he understood it now.

Lesson: The problem is very nonintuitive, and professional mathematicians frequently get it wrong.

More on independence

1) We saw

$$P(A|B) = P(A)$$

means A is independent of B.

It also implies

$$P(A \cap B) = P(A)P(B)$$

Is

$$P(B|A) = P(B) \quad \text{as well?}$$

POLL

A	B
True	False

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B)\cancel{P(A)}}{\cancel{P(A)}} \checkmark$$

2) What about B^c and A?

In terms of information,

A does not affect the probability of B happening. Knowing whether B happened or not is the same as knowing whether B^c happened or not.

So we must have

$$P(B^c \cap A) = P(B^c)\,P(A)$$

Lets check: (to be certain)

$$P(B^c \cap A) + P(B \cap A) = P(A)$$
$$P(B^c \cap A) + P(B)P(A) = P(A)$$

$$P(B^c \cap A)$$

Mutually exclusive events and Independence

If A and B are independent, are they mutually exclusive?

How many of you think there are distinct concepts?

POLL

YES	NO.

Suppose A and B are independent, then

$$P(A \cap B) = P(A)P(B).$$

If both $P(A) > 0$, $P(B) > 0$, then

$$P(A \cap B) > 0.$$

$\Rightarrow$ $A \cap B \neq \phi$ (null set).

So A and B cannot be mutually exclusive.

Independence of More than two events

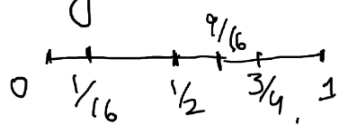
Suppose we are given 3 events A, B, C. How do we check if they are independent? Natural guess:

$$P(A \cap B \cap C) \stackrel{?}{=}$$

<u>Ex 2.24</u> $A = [1/2, 1]$ $B = [1/2, 3/4]$

$C = \left[\frac{1}{16}, \frac{9}{16}\right]$. $\Omega = [0, 1]$ and

$\mathbb{P}$ is the uniform measure.



$$P(A \cap B \cap C) = P([1/2, 9/16]) = \frac{1}{16}$$

$$P(A) = \frac{1}{2} \quad P(B) = \frac{1}{4} \quad P(C) = \frac{1}{2}$$

$$\Rightarrow P(A \cap B \cap C) = P(A)\, P(B)\, P(C).$$

However

$$P(A \cap B) = P(B) = \frac{1}{4} \neq P(A)\, P(B) = \frac{1}{8}$$

Lesson: For 3 events to be independent, we need to have

$$P(A \cap B) = P(A)\, P(B)$$
$$P(B \cap C) = P(B)\, P(C)$$

$$P(C \cap A) = P(C)\ P(A)$$
$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C).$$

In general, n events

$A_1 \ldots\ldots A_n$

Let $A_{i_1}, A_2 \ldots A_{i_k}$ be a subcollection.

Need $\mathbb{P}(A_{i_1} A_{i_2} \ldots A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$

For example, if we set $k=3$

$i_1 = 1\ i_2 = 3\ i_3 = 5$

$$\mathbb{P}(A_1 A_3 A_5) = \mathbb{P}(A_1)\,\mathbb{P}(A_3)\,\mathbb{P}(A_5)$$

$2^n - 1$ such equations, one for each subset of events.

The textbook gives a weirder equation

$$\mathbb{P}(A_1^* \cdots A_n^*) = \prod_{i=1}^{n} \mathbb{P}(A_i^*)$$

where $A_1^* = A_1$ OR A_1^c.

<u>Open question</u>: Can you find an example for each n such that $A_1 \cdots A_n$ satisfy some specified set of independence equations but not all.

In other words,

Are the equations themselves independent?

Independent Random Variables

Take an urn with 10 balls and sample 4 with replacement. Label the balls $\{1, 2, \dots, 10\}$

Let $X_i = $ # of i^{th} ball.

$$\mathbb{P}(X_1 = 3, X_2 = 3, \dots, X_4 = 3)$$

$$= \frac{1 \cdot 1 \cdot 1 \cdots \cdots 1}{n \cdot n \cdot \dots \cdot n}$$

$$= \mathbb{P}(X_1 = 3)\, \mathbb{P}(X_2 = 3) \cdots \mathbb{P}(X_4 = 3).$$

Then for ANY sequence of numbers $a_1 \dots a_4$

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \dots X_4 = a)$$

$$= \prod_{i=1}^{4} \mathbb{P}(X_i = a_i) \qquad (\star)$$

This is a general principle:

Discrete Random variables are indep. iff $\star$ holds for all $(a_1, \dots a_4)$.

INDEPENDENT RANDOM VARIABLES

Suppose we had 10 balls in an urn, and we draw 4 with replacement. Let X_i = ball drawn on i^{th} draw.

We have X_1, X_2, X_3, X_4. Are the draws independent?

Are the rvs independent?

How do we check?

$$P(X_1 = 1, \quad X_2 = 2, X_3 = 2, X_4 = 1)$$

$$= \quad \underline{1 \cdot 1 \cdot 1 \cdot 1}$$

$$P(X_1 = 1) = P(X_2 = 2) = P(X_3 = 3) = P(X_4 = 4) =$$

In general, have to check

$$P(X_1 = a_1, \ldots, X_4 = a_4)$$

$$= P(X_1 = a_1) \ldots P(X_4 = a_4)$$

for <u>all</u> $(a_1, a_2, a_3, a_4) \in [10]^4$

Let us do the same problem, but
WITHOUT REPLACEMENT

Originally we used Ω_1

$\Omega_1 = \{$ draw k balls from $\{1, \ldots, n\}\}$ with replacement

$k = 4$, n

WHAT IF WE USE THIS SAMPLE SPACE.

$\Omega_2 = \{$ draw k balls without replacement$\}$

Define X_i as before

POLL

Are the X_i independent?

A YES

B NO

Let us do the same problem, but WITHOUT REPLACEMENT

$\Omega_1 = \{$ draw k balls from $\{1, \dots, n\}\}$ with replacement

WHAT IF WE USE THIS SAMPLE SPACE.

$\Omega_2 = \{$ draw k balls <u>without</u> replacement $\}$

Then in particular

$$P(X_1 = a_1, X_2 = a_1, \dots X_n = a_1)$$

$$= 0 \; !$$

(*1)

(Since we cannot draw balls we have already drawn)

However

$$P(X_1 = a_1) = \frac{1}{n} \qquad (\star 2)$$

Recall our argument that says

"draw one ball. Dont look at it, draw the second. Now throw the 1st ball back in and ask, what is the probability that the 1st ball is a_1?" It follows that.

$$P(X_2 = a_1) = \frac{1}{n}$$

One can also PROVE this by looking at

$$P(X_2 = a_1) = \sum_{i=2}^{n} P(X_2 = a_1, X_1 = a_i)$$

(LAW OF TOTAL PROBABILITY)

Therefore, from (★1) and (★2) we have

$$0 \neq \frac{1}{n} \cdot \frac{1}{n} \cdot \cdots \frac{1}{n}$$

Thus, we have also mathematically proved it.