Review

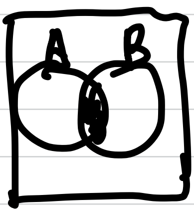
Probability of an event : $p(A)$

1. how likely A happens
2. $p(A) = \dfrac{\text{the number of time A happens}}{\text{the number of trials } (N)}$

$N \to \infty$

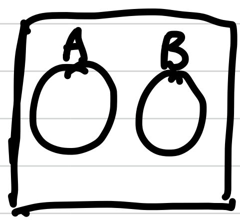
$p(A)$ is often approximated in reality.

Basic Laws



1) If A and B are disjoint,

$$p(A \cup B) = p(A) + p(B)$$



2) If A and B are not disjoint, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$P(A \cap B)$?

1. given in the problems

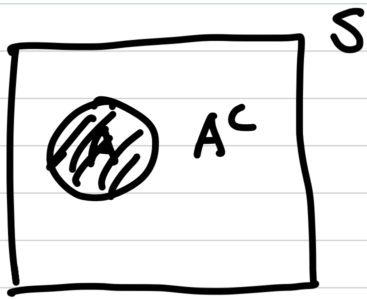
2. $\frac{\text{\# of times } A \cap B \text{ happens}}{N}$

3. If A and B are independent,

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

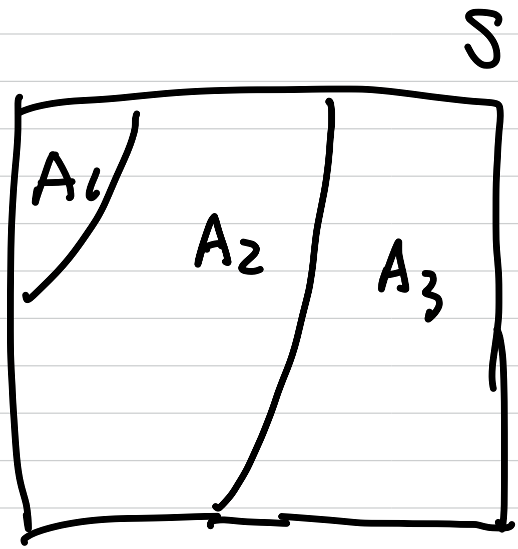
Partition

- A and A^c is a partition of sample space S.

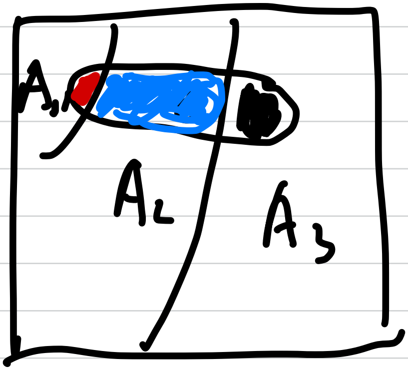


1) $A_1 \cup A_2 \cup A_3 = S$

2) A_1, A_2, A_3 are mutually exclusive/disjoint



$\{A_1, A_2, A_3\}$ is a partition of S.



$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

Total probability rule:

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

$\{A, A^c\}$ is a partition of S.

Conditional probability

$p(A|B)$: the probability of A <u>given</u> that B happens. (the knowledge of whether A happens or not is unkown, but we know B happens)

$p(B|A)$: the probability of B given that A happens.

By assumption,

$$p(A \cap B) = p(A|B)\,p(B)$$

$$= p(B|A)\,p(A)$$

If $P(B) > 0$,

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

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

Bayes' rule: If we know $P(B)$, $P(B^c)$, $P(A|B)$, $P(A|B^c)$, then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

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

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

$$= P(A)$$

$= A \cap B$

$P(A, B \mid C)$: the probability of A and B given that C happens

$P(C \mid A, B)$: the probability of C given that A and B happens

$A \cap B$

$$P(A, B \mid C) = \frac{P(A \cap B \cap C)}{P(C)}$$

$A \cap B$

$$P(C \mid A, B) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

$A \cap B$

Probability, conditional probability, independence, total probability rule, and Bayes' rule

Li Li

Department of Mathematics and Statistics

Probability

Probability of an event

- Classical probability is a specific value (one value).
  - Event A, N identical trials
  - Probability of Event: $P(A)$
    - $P(A) = \lim_{N\to\infty}$ (number of time A occurs / N trials)
    - To know $P(A)$ precisely requires infinite number of identical trials
- With known $P(A)$, there is still uncertainty in the outcome
- Classical probability is an objective concept: probability is a frequency ratio, not a physical rate but a dimensionless ratio

The Kolmogorov axioms

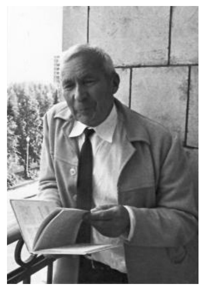


Figure: Andrey Nikolaevich Kolmogorov (25 April 1903 - 20 October 1987): Soviet mathematician who made significant contributions to the mathematics of probability theory.

The Kolmogorov axioms for probabilities:

- $0 \leq P(A) \leq 1$ for any event A
- $P(S) = 1$
- If $A_1, A_2, \ldots, A_n$ are pairwise disjoint events, then the probability of the union (logical OR) of all the events is the sum of the probabilities of each event:
  $P(A_1 \text{ or } A_2 \text{ or } A_3 ... \text{ or } A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$

Some Consequences of Probability Measure

- $P(\text{null set}) = 0$
- If event A is a subset of event B, then $P(A) \leq P(B)$

Notation

- Logical OR denoted with $\cup$ (union)
- Logical AND denoted with $\cap$ (intersection)
- Will NOT use $+$ and $*$ for OR and AND; we only use + and * only for addition and multiplication
- Why? Avoid confusion–as discussed later
  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Calculus of Probabilities

- In general, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - If A and B are mutually exclusive $A \cap B = \emptyset$ and $P(A \cap B) = 0$

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

  - If A and B are independent, $P(A \cap B) = P(A) * P(B)$

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

- Mutually exclusive is NOT Independent; in fact, Mutually exclusive is “Totally Dependent” in general.

Example

Show the following:

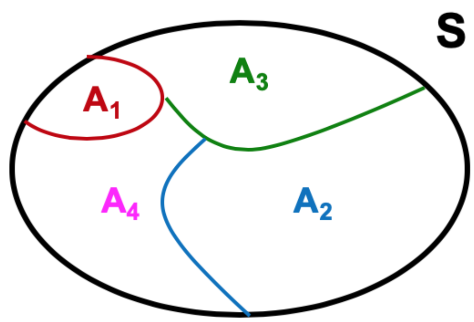
- $P(A) + P(A^c) = 1$.
- $P(B \cap A) + P(B \cap A^c) = P(B)$

Example

Partition

Partition is a set of mutually exclusive events that covers the sample space

- $A_1, A_2, \ldots, A_n$ form a partition over S if $\cup_{i=1}^{n} A_i = S$ and $A_j \cap A_i = \emptyset$ for any A_j and A_i in the collection of $\{A_i\}$.



- The set of all outcomes is a partition of the sample space.

Calculus of Probabilities

If $A_1, A_2, \ldots, A_n$ form a partition of the sample space S,

- $P(A_1) + P(A_2) + \cdots + P(A_n) = 1$.
- $P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n) = P(B)$

Probability of an event

- **Probability of an event from a discrete sample space**: Denote the outcomes in the sample space as $s_1, \ldots, s_n$. Let $p_1, \ldots, p_n$ be the corresponding probabilities that sum to 1. For any event A,

$$P(A) = \sum_{i:s_i \in A} p_i$$

  For example, if $S = \{2, 3, 5, 6\}$ and the probabilities for these outcomes are $\{0.1, 0.3, 0.4, 0.2\}$ respectively. Then if $A = \{3, 5\}$, then $P(A) = 0.3 + 0.4 = 0.7$

- If all outcomes are equal likely, then $P(A) = \frac{\text{The number of outcomes in A}}{\text{The number of outcomes in S}}$

Probability of an event

- **Probability of an event from a continuous sample space**: Suppose the sample space is an interval $[a, b]$ where a and b are real numbers. Let $f(x)dx$ be the probability of a small interval dx in the sample space. For any event A,
$$P(A) = \int_A f(x)dx$$
- For example, suppose the probability for a light bulb to fail in a small time interval dx (in years) is $e^{-x}dx$, then the probability of the light bulb to fail in $[1, 2]$ years is $\int_1^2 e^{-x}dx = e^{-1} - e^{-2} = 0.233$

In practice, the probabilities of events are typically unknown. Statistical methods are often used to estimate the probabilities.

Conditional probability

Where is my bag?

- Suppose you've just landed in Albuquerque after making a tight connection in Denver, and you're waiting for your suitcase to appear on the carousel.
- Suppose other passengers have started to get their bags.
- If the bags have just started to show up on the carousel, perhaps you should be patient and wait a little longer.
- If the you've been waiting a long time, then things are perhaps looking bad.

Use the notations,

$$p(BP \mid T_x, C^c)$$

Where is my bag?

If x minutes have passed and I still haven't gotten my bag, what is the probability that it was on the plane?

The question is :

P(bag on the plane = T | time elapsed = x, Carousel = F)

Denote BP as the event the bag on the plane

T_x as the event that time elapsed is x minutes

C as the event that the bag does showing up Carousel

$$P(B_p \mid T_x, C^c)$$

$$= \frac{P(B_p \cap T_x \cap C^c)}{P(T_x \cap C^c)}$$

Information given:
- $P(C^c \mid B_p, T_x)$
- $P(C \mid B_p, T_x)$
- $P(C \mid B_p^c, T_x)$
- $P(C^c \mid B_p^c, T_x)$
- $P(B_p) = 0.5$

Final answer

$$= \frac{(1-0.1x)\cdot 0.5}{(1-0.1x)\cdot 0.5 + 0.5}$$

$x=0$	0.5
$x=1$	0.4737
$x=2$	0.444
$x=3$	0.4118
$x=4$	0.375

$$P(B_p \cap T_x \cap C^c)$$

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

$$= P(C^c \mid B_p, T_x)\, P(B_p, T_x)$$

$$= (1-0.1x)\cdot P(B_p)\cdot P(T_x)$$

Independence assumption

$$= (1-0.1x)\cdot 0.5\cdot P(T_x)$$

$P(T_x \cap C^c)$: T_x and C^c are not independent

$P(A) = P(A \mid B)\,P(B) + P(A \mid B^c)\cdot P(B^c)$

$$= P(C^c \mid T_x)\, P(T_x)$$

$$= \left[P(C^c \mid T_x, B_p)\, P(B_p \mid T_x) + P(C^c \mid T_x, B_p^c)\, P(B_p^c \mid T_x)\right] P(T_x)$$

$$= \left[(1-0.1x)\cdot 0.5 + 1\cdot 0.5\right]\cdot P(T_x)$$

Where is my bag?

$p(C^c | BP^c, T_0) = 1$

$p(C | BP^c, T_0) = 0$

TABLE 3.3. A more complicated conditional probability table.

Probability of →, Given ↓			
bag on plane	*time elapsed*	*carousel* = false	*carousel* = true
False	0	100	0
False	1	100	0
False	2	100	0
False	3	100	0
False	4	100	0
False	5	100	0
False	6	100	0
False	7	100	0
False	8	100	0
False	9	100	0
False	10	100	0
True	0	100	0
True	1	90	10
True	2	80	20
True	3	70	30
True	4	60	40
True	5	50	50
True	6	40	60
True	7	30	70
True	8	20	80
True	9	10	90
True	10	0	100

$p(C^c | BP^c, T_1) = 1$

$p(C^c | BP, T_0) = 1$

$p(C^c | BP, T_5)$

$1 - 0.1 \cdot x$

Where is my bag?

The key to answer the question is the Bayes' rule.

Where is my bag?

Suppose that the bag is on the plane with 50% of chance from prior belief.

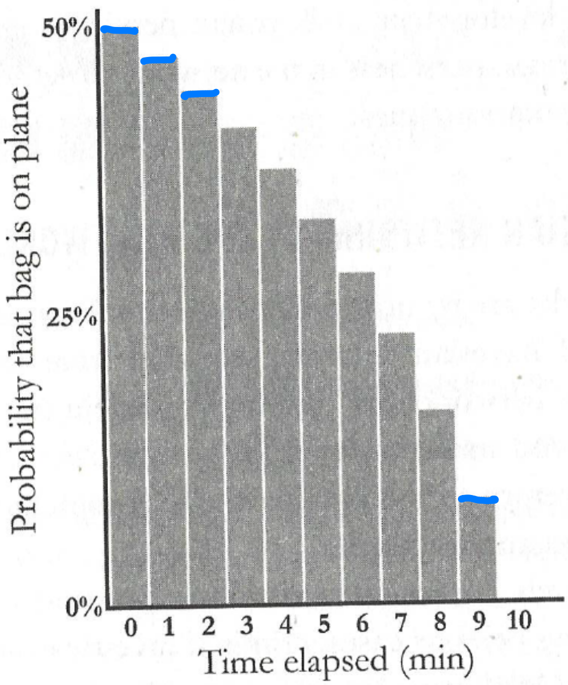


FIGURE 3.6. The probability of seeing your bag on the carousel decreases slowly at first, then more rapidly. (*Source:* Graph by Maayan Harel, data from Stefan Conrady and Lionel Jouffe.)

Definition of conditional probability

- Conditional probability is a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred.
- **Conditional probability of** A **given** B: is written $P(A|B)$.
- **Conditional probability of** B **given** A: is written $P(B|A)$.

Example

Two cards are dealt from the top of a well-shuffled deck that has 52 cards and are play in order. Denote A as the event for the first card being a diamond and B as the event for the second card being a spade.

- What is the probability of B given A?
- What is the probability that the first card is a diamond and the second card is a spade?

Conditional probability rules

In general,

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

Conditional probability rules

- If $P(A) > 0$, then

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

- If $P(B) > 0$, then

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

- If $P(B) > 0$ and $P(A) > 0$, then

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

Conditional probability rules

- **Total probability rule:** For any events A and B,

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

- **Bayes' rule:** For any events A and B,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Proofs for Total probability rule and Bayes' rule

Exercises

$P(A^c) = 1 - P(A) = 0.6$

If $P(A) = 0.4$, $P(B|A) = 0.1$, and $P(B|A^c) = 0.2$, what are $P(B)$ and $P(A|B)$?

$P(B) = P(B \cap A) + P(B \cap A^c) = 0.16$

$P(B \cap A) = P(B|A)\,P(A) = 0.1 \cdot 0.4 = 0.04$

$P(B \cap A^c) = P(B|A^c)\,P(A^c) = 0.2 \cdot 0.6 = 0.12$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.04}{0.16} = 0.25$$

Total probability rule and Bayes' rule

Assume $A_1, A_2, \cdots, A_m$ be a partition of the sample space and let B be any set.

- **Total probability rule for multiple events:**

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_m)P(A_m)$$

- **Bayes' rule for multiple events:**

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_m)P(A_m)}$$

Multiple events and multiple conditions

- Multiple events:
  - **Conditional probability of** A **and** B **given** C: is written $P(A, B|C)$.
  - **Conditional probability of** B **and** C **given** A: is written $P(B, C|A)$.
- Multiple conditions:
  - **Conditional probability of** A **given** B **and** C: is written $P(A|B, C)$.
  - **Conditional probability of** B **given** A **and** C: is written $P(B|A, C)$.

Multiple events

- If $P(C) \neq 0$,

$$P(A, B|C) = \frac{P(A \cap B \cap C)}{P(C)}$$

Multiple conditions

- If $P(B \cap C) \neq 0$,

$$P(A|B,C) = \frac{P(A \cap B \cap C)}{P(B \cap C)}$$

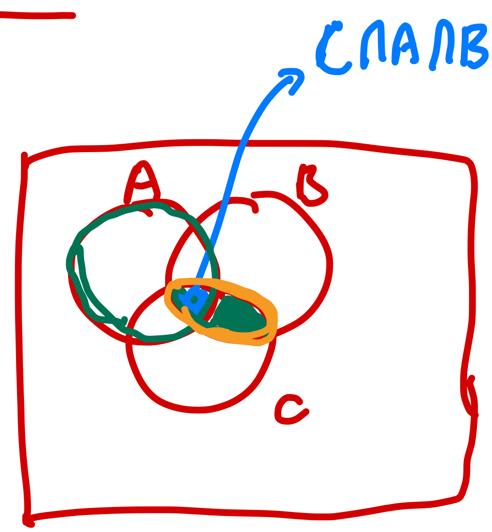
- If $P(C) \neq 0$, we can further have

$$P(A|B,C) = \frac{P(A \cap B|C)P(C)}{P(B|C)P(C)} = \frac{P(A,B|C)}{P(B|C)}$$

Example

Show that

$$P(A|B,C) = \frac{P(C|A,B)P(A|B)P(B)}{P(C|A,B)P(A|B)P(B) + P(C|A^c,B)P(A^c|B)P(B)}$$



Where is my bag re-visit

Where is my bag re-visit

Bayes' rule used in real world

Enzyme immunoassay tests are used to screen blood specimens for the presence of antibodies to HIV, the virus that causes AIDS. Antibodies indicate the presence of the virus. The test is quite accurate but is not always correct. Here are approximate probabilities of positive and negative test results when the blood test does and does not contain antibodies to HIV:

Ap: Antibodies present

+ : Positive

− : Negative

Aa : Antibodies absent

Probability of →, Given ↓	Test = Positive	Test = Negative
Antibodies present	0.9985	0.0015
Antibodies absent	0.0060	0.9940

According to estimated HIV incidence among persons aged between 13 to 24 years from CDC (2010-2016 data), the prevalence is 18.9 persons per 100, 000 population.

$P(+ \mid A_p) = 0.9985$; $P(- \mid A_p) = 0.0015$

$P(+ \mid A_a) = 0.0060$; $P(- \mid A_a) = 0.9940$

$$P(A_p) = \frac{18.9}{100{,}000} = 0.000189$$

$$P(A_a) = 1 - P(A_p) = 0.999811$$

Example

$P(+)$ vs $P(+ \mid A_p)$

(a) What is the probability that the test is positive for a randomly chosen person from this population?

(b) If a person is tested positive, what is the probability that a person truly has antibodies?

(a) $P(+) = P(+ \mid A_p) \cdot P(A_p) + P(+ \mid A_a) P(A_a)$ ②

$= P(\text{'+' and } A_p) + P(\text{'+' and } A_a)$ ①

$= 0.9985 \cdot 0.000189 + 0.006 \cdot 0.999811$

$= 0.00619$

(b) $P(A_p \mid +) = \dfrac{P(A_p \text{ and '+'})}{P(+)} = \dfrac{0.0001887}{0.00619}$

$= 0.03049$

Diagnostic test

Diagnostic test

Independence

Definition

suppose the actions are independent

- Suppose in a random experiment, a coin is flipped 3 times. Suppose the probability of showing up "Head" is 0.5 for each flip and the result of flips do not interfere each other. What is the probability of seeing three heads?
- Since the results of flips do not interfere each other, the probability of seeing three heads is $0.5 * 0.5 * 0.5 = 0.125$.

A_1: the first coin showing up head

A_2: the second coin × × ×

A_3: the third coin × × ×

$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$

$= 0.5 \cdot 0.5 \cdot 0.5 = 0.125$

Independence

- Independence of two events means that the occurrence of one event does not affect whether another event occurs or vice versa.
- Using conditional probability, if A and B are independent, then

We know B happens.

$$P(A|B) = P(A) \quad (1)$$
$$P(B|A) = P(B) \quad (2)$$

- Both equations (1) and (2) imply

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

Independence

In general,

- Independence of two events does not imply mutual exclusion (disjoint)
- Mutual exclusion (disjoint) of two events does not imply independence

Failure to understand the difference is a fundamental mistake!

Examples of independent events in real applications

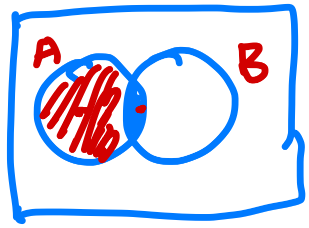
- In a clinical trial, a treatment or a placebo is randomly assigned to each patient. Whether a patient receives treatment is independent of whether the patient is young or old.
- In a lab, specimen tubs are broken due to random reasons and result in missing values for the measurement. Whether the measurement is missing is independent of the actual level had the tub being measured .

How to evaluate independence?

Two events, A and B are statistically independent if you can verify one of the following:

- $P(A \cap B) = P(A)P(B)$, then A and B are independent
- $P(A|B) = P(A)$, then A and B are independent
- $P(B|A) = P(B)$

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



$$P(A) - P(B^c \cap A) = P(A)\cdot P(B)$$

$$P(A)(1-P(B)) = P(B^c\cap A) \Rightarrow A \text{ and } B^c \text{ are independent}$$

$1-P(B) = P(B^c)$

Example

If A and B are independent, then A and B^c are independent.

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

Since A and B are independent, $P(A\cap B) = P(A)\cdot P(B)$

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

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

$$P(B)(1-P(A)) = P(A\cap B^c)$$

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

$\Rightarrow$ B and A^c are independent

Independence of
- two events A and B: the occurrence of A does not affect the occurrence of B
- two random variables
- two distributions

Independence does not imply $A \cap B = \emptyset$

It implies
$$\begin{cases} P(A|B) = P(A) \\ P(B|A) = P(B) \end{cases} \Longrightarrow P(A \cap B) = P(A)P(B)$$

Independence of multiple events

A collection of events $A_1, A_2, \ldots, A_n$ are mutually independent if for any sub collection $A_{i1}, \ldots A_{ik}$, we have

$$P(\cap_{j=1}^{k} A_{ij}) = \prod_{j=1}^{k} P(A_{ij}).$$

Suppose n=3, then A_1, A_2, A_3 are mutually independent if all of following statements are true

$$\begin{cases} P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) \quad ① \\ P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) \\ P(A_1 \cap A_3) = P(A_1) \cdot P(A_3) \\ P(A_2 \cap A_3) = P(A_2) \cdot P(A_3) \end{cases}$$

If $A_1, A_2, A_3, \ldots A_k$ are mutually independent,

$$P(A_1 \cap A_2 \cap A_3 \ldots \cap A_k) = P(A_1) \cdot P(A_2) \ldots P(A_k)$$

- The probability of getting exactly 2 heads is: $0.6 \cdot 0.6 \cdot 0.4^8 \cdot \binom{10}{2}$
- The probability of getting less than 2 heads?

Example

$0.4^{10} + 10 \cdot 0.6 \cdot 0.4^9$

Consider flipping an unfair coin 10 times independently. The probability of "Head" is 0.6.

- What is the probability of getting 10 heads? 0.6^{10}
- What is the probability of getting exactly 1 head?

$0.6 \cdot 0.4^9 \cdot \binom{10}{1}$

$A = \{\text{getting exactly 1 head}\}$

$A = A_1 \cup A_2 \cup A_3 \cdots A_{10}$

A_i: the ith flip showing up head, and the rest showing up tail.

$p(A_1) = 0.6 \cdot 0.4^9$

$p(A_2) = 0.4 \cdot 0.6 \cdot 0.4 \cdots 0.4 = 0.6 \cdot 0.4^9$

$p(A_i) = 0.6 \cdot 0.4^9$

$p(A) = p(A_1) + p(A_2) + \cdots p(A_{10}) = 10 \cdot 0.6 \cdot 0.4^9$

Example