Mathematical Statistics and Data Analysis

Lecture 2: Review of Probability - Part I

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

September 11, 2019

Outlines

Reading Material

Textbook:

- Rice: Chapter 1;
- Mao: Chapter 1;

Introduction

The mathematical theory of probability has been applied to a wide variety of phenomena:

- Genetics;
- Lengths of various queues;
- Demands on inventories of goods;
- Actuarial science;

Sample Spaces

Probability theory is concerned with situations in which the outcomes occur randomly.

- Such situations are called **experiments**;
- The set of all possible outcomes is the **sample space** corresponding to an experiment.
- The sample space is denoted by Ω, and an element of Ω is denoted by ω.
- The particular subsets of interest of Ω are called **events**.

Sample Spaces

Experiments vs Events

Driving to work, a commuter passes through a sequence of three intersections with traffic lights. At each light, she either stops, s, or continues, c.

- The sample space is the set of all possible outcomes: $\Omega = \{ccc, ccs, css, scs, sss, ssc, scc, scs\}$, where csc denotes the outcome that the commuter continuous through the first light, stops at the second light, and continues through the third light.
- The event that the commuter stops at the first light is the subset of Ω denoted by

$$A = \{sss, ssc, scc, scs\}$$

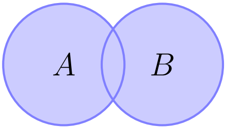
Sample Spaces

Suppose A and B are two events.

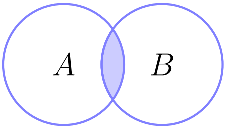
Operations

- The **union** of two events is the event that either A occurs or B occurs or both occur;
- The **intersection** of two events is the event that both A and B occur;

$A \cup B$



$A \cap B$



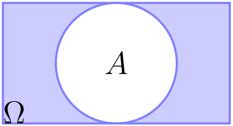
Sample Spaces

Suppose A and B are two events.

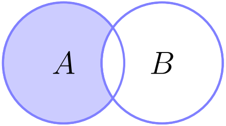
Operations

- The **complement** of an event is the event that A does not occur and thus consists of all those elements in the sample space that are not in A;
- The **difference** of two events is the event that A occurs but B does not occur;

A^c



$A - B = A \cap B^c$



Sample Spaces

Special Set

- Empty set $\emptyset$: the set with no elements;

Example

- A is the event that the commuter stops at the first light, $A = \{scc, scs, ssc, sss\}$;
- C is the event that she continuous through all three lights, $C = \{ccc\}$;
- $A \cap C = \emptyset$;
- A and C are said to be **disjoint**.

Laws of Set Theory

- Commutative Laws:

$$A \cup B = B \cup A;$$
$$A \cap B = B \cap A;$$

- Associative Laws:

$$(A \cup B) \cup C = A \cup (B \cup C);$$
$$(A \cap B) \cap C = A \cap (B \cap C);$$

Laws of Set Theory

- Distributive Laws:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C);$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C);$$

- DeMorgan's Laws:

$$(A \cup B)^c = A^c \cap B^c;$$
$$(A \cap B)^c = A^c \cup B^c;$$

Probability Definition

Sigma Algebra

A collection of subsets of Ω is called a **sigma algebra** (or Borel field), denoted as $\mathcal{F}$. if it satisfies the following three properties :

- $\emptyset \in \mathcal{F}$;
- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;
- If $A_1, A_2, \cdots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

In addition, from DeMorgan's Laws, it follows that $\mathcal{F}$ is closed under countable intersection, i.e.

$$(\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c \in \mathcal{F}$$

Probability Definition

Probability Measure

Given a sample space Ω and an associated sigma algebra $\mathcal{F}$, a **probability measure** is a function P with domain $\mathcal{F}$ that satisfies

- **Non-negativity**: $P(A) \geq 0$ for all $A \in \mathcal{F}$;
- **Normalization**: $P(\Omega) = 1$;
- **Additivity**: If $A_1, A_2, \cdots \in \mathcal{F}$ are pairwise disjoint, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Note that

- **Probability Space**: $(\Omega, \mathcal{F}, P)$.
- The Axioms of probability;
- Proposed by **Kolmogorov**.

Probability Properties

- $P(\emptyset) = 0$;
- If $A_1, A_2, \cdots, A_n$ are pairwise disjoint, then

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

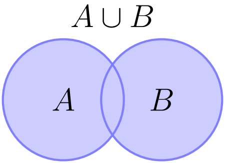
- For each event A, $P(A^c) = 1 - P(A)$;
- If $A \subset B$, then
  - $P(A - B) = P(A) - P(B)$;
  - $P(A) \leq P(B)$;
- Suppose that A and B are two events.

$$P(A - B) = P(A) - P(AB)$$

Probability Properties

For any two events A and B, we have

- **The Addition Formula**:
  $P(A \cup B) = P(A) + P(B) - P(AB)$;
- **Subadditivity**: $P(A \cup B) \leq P(A) + P(B)$;



Counting Method

- The elements of Ω all have equal probability;
- If there are N elements in Ω, each of them has probability $1/N$. If A can occur in any of n mutually exclusive ways, then $P(A) = n/N$, or

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total number of outcomes}}$$

Note that this formula holds only if all the outcomes are equally likely.

Counting Method

Example: Simpson's Paradox

A black urn contains 5 red and 6 green balls, and a white urn contains 3 red and 4 green balls. You are allowed to choose an urn and then choose a ball at random from the urn. If you choose a red ball, you get a prize. Which urn should you choose to draw from?

- If the black urn is drawn, the probability of choosing a red ball is $5/11 = 0.455$;
- If the white urn is drawn, the probability of choosing a red ball is $3/7 = 0.429$;
- So, you should choose to draw from the **black** urn.

Example: Simpson's Paradox (Con'd)

Now consider another game in which a second black urn has 6 red and 3 green balls, and a second white urn has 9 red and 5 green balls. If you choose a red ball, you get a prize. Which urn should you choose to draw from?

- If you draw from the black urn, the probability of a red ball is $6/9 = 0.667$;
- If you choose to draw from the white urn, the probability is $9/14 = 0.634$;
- So, again you should choose to draw from the **black** urn.

Counting Method

Example: Simpson's Paradox (Con'd)

In the final game, the contents of the second black urn are added to the first black urn, and the contents of the second white urn are added to the first white urn. Again, you can choose which urn to draw from. Which should you choose?

- The black urn now contains 11 red and 9 green balls, so the probability of drawing a red ball from it is $11/20 = 0.55$;
- The white urn now contains 12 red and 9 green balls, so the probability of drawing a red ball from it is $12/21 = 0.571$;
- So, again you should choose the **white** urn.

Counting Method

Multiplication Principle

If one experiment has m outcomes and another experiment has n outcomes, then there are mn possible outcomes for the two experiments.

Proof: Denote the outcomes of the first experiment $a_1, \cdots, a_m$ and the outcomes of the second experiment by $b_1, \cdots, b_n$. The outcomes for the two experiments are the ordered pairs (a_i, b_j). These pairs can be exhibited as the entries of an $m \times n$ rectangular array, in which the pair (a_i, b_j) is in the ith row and the jth column. There are mn entries in this array.

Counting Method

Extended Multiplication Principle

If there are p experiments and first has n_1 possible outcomes, the second $n_2, \cdots$, and the pth n_p possible outcomes, then there are a total of $n_1 \times n_2 \times \cdots \times n_p$ possible outcomes for the p experiments.

Proof: We saw that it is true for $p = 2$. Assume that it is true for $p = q$, that is, that there are $n_1 \times n_2 \times \cdots \times n_q$ possible outcomes for the first q experiments. To complete the proof by induction, we must show that it follows that the property holds for $p = q + 1$. We apply the multiplication principle, regarding the first q experiments as a single experiment with $n_1 \times \cdots \times n_q$ outcomes, and conclude that there are $(n_1 \times \cdots \times n_q) \times n_{q+1}$ outcomes for the $q + 1$ experiments.

Counting Method

- A **permutation** is an ordered arrangement of objects.
- Suppose that from the set $C = \{c_1, c_2, \cdots, c_n\}$ we choose r elements and list them in order.
- If no duplication is allowed, we are sampling without replacement.
- If duplication is allowed, we are sample with replacement.

For a set of size n and a sample of size r, there are n^r different ordered samples with replacement and $n(n-1)(n-2)\cdots(n-r+1)$ different ordered samples without replacement.

Counting Method

Question: If r objects are taken from a set of n objects without replacement and disregarding order, how many different samples are possible?
Answer: From the multiplication principle, the number of ordered samples equals the number of unordered samples multiplied by the number of ways to order each sample. Since the number of ordered samples is

$$n(n-1)(n-2)\cdots(n-r+1),$$

and since a sample of size r can be ordered in $r!$ ways, the number of unordered samples is

$$\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!r!}$$

This number is also denoted as $\binom{n}{r}$.

Counting Method

The number of unordered samples of r objects selected from n objects without replacement is $\binom{n}{r}$.

The numbers $\binom{n}{k}$, called the binomial coefficients, occur in the expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

In particular,

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

Counting Method

Example: Birthday Problem

Problem: Suppose that a room contains n people. What is the probability that at least two of them have a common birthday?

Solution: This is a famous problem with a counterintuitive answer. Assume that every day of the year is equally likely to be a birthday, disregard leap years, and denote by A the event that at least two people have a common birthday. As is sometimes the case, finding $P(A^c)$ is easier than finding $P(A)$. This is because A can happen in many ways, whereas A^c is much simpler. There are 365^n possible outcomes, and A^c can happen in $365 \times 364 \times \cdots \times (365 - n + 1)$ ways. Thus,

$$P(A^c) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$

Example: Birthday Problem

Problem: Suppose that a room contains n people. What is the probability that at least two of them have a common birthday?

Solution: This is a famous problem with a counterintuitive answer. Assume that every day of the year is equally likely to be a birthday, disregard leap years, and denote by A the event that at least two people have a common birthday. As is sometimes the case, finding $P(A^c)$ is easier than finding $P(A)$. This is because A can happen in many ways, whereas A^c is much simpler. There are 365^n possible outcomes, and A^c can happen in $365 \times 364 \times \cdots \times (365 - n + 1)$ ways. Thus,

$$P(A^c) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$

Example: Birthday Problem (Con'd)

The desired probability is

$$\begin{aligned} P(A^c) &= \frac{365!}{365^n(365-n)!} \\ &= \left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{n-1}{365}\right) \end{aligned}$$

Another Problem: How to obtain an approximate value of this probability?

Example: Birthday Problem (Con'd)

The desired probability is

$$\begin{aligned} P(A^c) &= \frac{365!}{365^n(365-n)!} \\ &= \left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{n-1}{365}\right) \end{aligned}$$

Another Problem: How to obtain an approximate value of this probability?

Example: Birthday Problem (Con'd)

Solution:

- When n is not large, the terms $\frac{i}{365} \times \frac{j}{365}$ can be omitted. It means

$$P(A^c) \approx 1 - \frac{1 + 2 + \cdots + (n-1)}{365} = 1 - \frac{n(n-1)}{730}$$

- When n is large, we have $\ln(1-x) \approx -x$ for a small positive number x. It means

$$\ln P(A) \approx -\frac{1 + 2 + \cdots + (n-1)}{365} = -\frac{n(n-1)}{730}$$

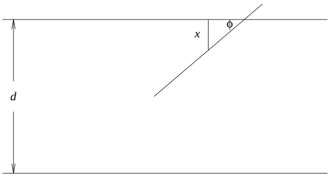
Geometric method

Buffon needle problem

A needle of length l is dropped randomly on a plane ruled with parallel lines that are a distance d apart, where $d > l$. What is the probability that the needle comes to rest crossing a line?

Solution:

Let x be the distance from the middle point of the needle to the nearest parallel line and ϕ be the angle between the needle and the line.



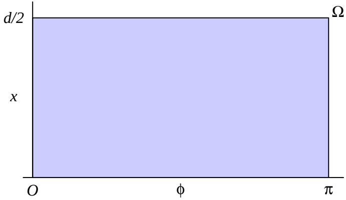
Geometric method

Buffon needle problem (Con'd)

It is obvious that

$$0 \leq x \leq d/2 \text{ and } 0 \leq \phi \leq \pi$$

The sample space is a blue rectangle, shown as follows:



The area of sample space is

$$S_\Omega = d\pi/2$$

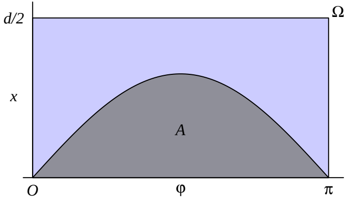
Geometric method

Buffon needle problem (Con'd)

The event of interest is that the needle comes to rest crossing a line, if and only if

$$x \leq \frac{l}{2}\sin\phi.$$

The area of interest is shaded in the following figure.



This area is

$$S_A = \int_0^{\pi} \frac{l}{2}\sin\varphi \mathrm{d}\varphi = l$$

Buffon needle problem (Con'd)

Then the probability of interest is

$$P(A) = \frac{S_A}{S_\Omega} = \frac{l}{d\pi/2} = \frac{2l}{d\pi}.$$

Suppose that l and d are known.

- If π is known, it is easy to calculate the probability $P(A)$;
- Given $P(A)$, we can obtain the approximation of π: The needles is dropped N times and it crosses a line n times. The probability $P(A)$ is estimated by n/N. Thus,

$$\frac{n}{N} \approx P(A) = \frac{2l}{d\pi} \Rightarrow \pi \approx \frac{2lN}{dn}$$

Conditional Probability

Definition

Let Ω be the sample space, and A and B be two events with $P(B) \neq 0$.

- The **conditional probability** of A given B is defined to be
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Problem:

Is $P(\cdot|B)$ a probability?

Conditional Probability

Definition

Let Ω be the sample space, and A and B be two events with $P(B) \neq 0$.

- The **conditional probability** of A given B is defined to be

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

Problem:

Is $P(\cdot|B)$ a probability?

Conditional Probability

Properties

A conditional probability is a probability. Equivalently, if $P(B) > 0$, then

- $P(A|B) \geq 0$, for any $A \in \mathcal{F}$;
- $P(\Omega|B) = 1$;
- If $A_1, A_2, \cdots, A_n \in \mathcal{F}$ are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{n} A_i | B\right) = \sum_{i=1}^{n} P(A_i|B)$$

Multiplication Law

- If $P(B) \neq 0$, then

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

- If $P(A_1A_2 \cdots A_{n-1}) > 0$, then

$$P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1A_2 \cdots A_{n-1})$$

Multiplication Law

The Urn Model

- In a urn, there are b **blue** balls and r **red** balls.
- A ball is drawn randomly, and then return it to the urn.
- Meanwhile, c balls with the same color and d with the opposite color are added to the urn.

At ith withdrawal, the black ball is denoted by B_i and the red ball is denoted by R_i. Consider three sequential balls are withdrawn from a urn. Two balls are red and one is black.

The Urn Model (Con'd)

- **Blue** - **Red** - **Red**

$$\begin{aligned} P(B_1R_2R_3) &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2) \\ &= \frac{b}{b+r} \cdot \frac{r+d}{b+r+c+d} \cdot \frac{r+d+c}{b+r+2c+2d} \end{aligned}$$

- **Red** - **Blue** - **Red**

$$\begin{aligned} P(R_1B_2R_3) &= P(R_1)P(B_2|R_1)P(R_3|R_1R_2) \\ &= \frac{r}{b+r} \cdot \frac{b+d}{b+r+c+d} \cdot \frac{r+d+c}{b+r+2c+2d} \end{aligned}$$

- **Red** - **Red** - **Blue**

$$\begin{aligned} P(R_1R_2B_3) &= P(R_1)P(R_2|R_1)P(B_3|R_1R_2) \\ &= \frac{r}{b+r} \cdot \frac{r+c}{b+r+c+d} \cdot \frac{b+2d}{b+r+2c+2d} \end{aligned}$$

Multiplication Law

The Urn Model (Con'd)

- The urn modl is also called **Pólya model**.
- There are some special cases as follows:
  - $c = -1, d = 0 \Leftrightarrow$ Sampling without replacement.
  - $c = 0, d = 0 \Leftrightarrow$ Sampling with replacement.
  - $c > 0, d = 0 \Leftrightarrow$ Epidemic model;

  Note that $P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3)$ if $d = 0$.
  - $c = 0, d > 0 \Leftrightarrow$ Security model:
    - $P(B_1R_2R_3) = \frac{b}{b+r} \cdot \frac{r+d}{b+r+d} \cdot \frac{r+d}{b+r+2d}$;
    - $P(R_1B_2R_3) = \frac{r}{b+r} \cdot \frac{b+d}{b+r+d} \cdot \frac{r+d}{b+r+2d}$;
    - $P(R_1R_2B_3) = \frac{r}{b+r} \cdot \frac{r}{b+r+d} \cdot \frac{b+2d}{b+r+2d}$;

Law of Total Probability

Definition

$\mathcal{B} = \{B_1, B_2, \cdots, B_n\}$ is a **partition** of the sample space Ω, which satisfies

- For any two events B_i and B_j, B_i and B_j are disjoint;
- $\cup_{i=1}^{n} B_i = \Omega$;

Property

Suppose that $\mathcal{B} = \{B_1, B_2, \cdots, B_n\}$ is a partition of the sample space Ω. If $P(B_i) > 0, i = 1, 2, \cdots, n$, then, for each event A, we have

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A|B_i)$$

Law of Total Probability

The conditions in the law of total probability are follows:

- $B_1, B_2, \cdots, B_n$ are pairwise disjoint;
- $\bigcup_{i=1}^n B_i = \Omega$;
- $P(B_i) > 0$.

Note that

- **The simplest formula**: If $0 < P(B) < 1$, then

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

- $\bigcup_{i=1}^n B_i = \Omega$ can be replaced with

$$A \subset \bigcup_{i=1}^n B_i$$

Bayes' Rule

Definition

Suppose that A is an event from a certain sample space Ω and $\mathcal{B} = \{B_1, B_2, \cdots, B_n\}$ is a partition of the sample space. Let $P(A) \neq 0$ and $P(B_i) \neq 0$, $i = 1, 2, \cdots, n$. Then

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

The simplest formula:

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

Bayes' Rule

Definition

Suppose that A is an event from a certain sample space Ω and $\mathcal{B} = \{B_1, B_2, \cdots, B_n\}$ is a partition of the sample space. Let $P(A) \neq 0$ and $P(B_i) \neq 0$, $i = 1, 2, \cdots, n$. Then

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

The simplest formula:

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

Bayes' Rule

The Boy Who Cried Wolf

A shepherd boy, who tended his flock not far from a village, used to amuse himself at times in crying out "Wolf! Wolf!" Twice or thrice his trick succeeded; the whole village came running out to his assistance, when all the return they got was to be laughed at for their pains.

At last one day the wolf came indeed. The boy cried out in earnest. His neighbors, supposing him to be at his old sport, paid no heed to his cries, and the wolf devoured the sheep. So the boy learned, when it was too late, that liars are not believed even when they tell the truth.

The Boy Who Cried Wolf (Con'd)

- A: The boy told a lie;
- B: The boy was trustworthy.

Suppose that

$$P(B) = 0.8 \text{ and } P(B^c) = 0.2$$

and

$$P(A|B) = 0.1 \text{ and } P(A|B^c) = 0.5.$$

The Boy Who Cried Wolf (Con'd)

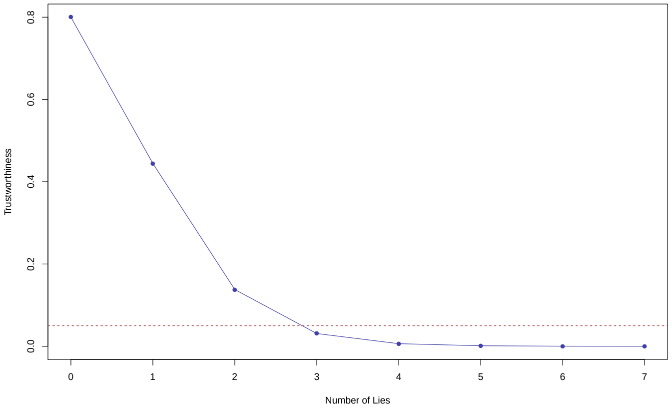
- After the first trick succeeded, the probability that he was trustworthy is

$$
\begin{aligned}
P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(B^c)P(A|B^c)} \\
&= \frac{0.8\times 0.1}{0.8\times 0.1+0.2\times 0.5} = 0.444
\end{aligned}
$$

- After the second trick succeeded, the probability that he was trustworthy is

$$
\begin{aligned}
P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(B^c)P(A|B^c)} \\
&= \frac{0.444\times 0.1}{0.444\times 0.1+0.556\times 0.5} = 0.138
\end{aligned}
$$

The Boy Who Cried Wolf (Con'd)



Independence

Definition

A and B are said to be **independent events** if

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

Property

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

Proof: According to the property of probability, we know

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

Since A and B are independent, $P(A \cap B) = P(A)P(B)$. Thus, $P(AB^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$. This indicates that A and B^c are independent.

Independence

Definition

A and B are said to be **independent events** if

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

Property

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

Proof: According to the property of probability, we know

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

Since A and B are independent, $P(A \cap B) = P(A)P(B)$. Thus, $P(AB^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$. This indicates that A and B^c are independent.

Independence

Definition

A and B are said to be **independent events** if

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

Property

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

Proof: According to the property of probability, we know

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

Since A and B are independent, $P(A \cap B) = P(A)P(B)$. Thus, $P(AB^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$. This indicates that A and B^c are independent.

Independence

Definition

Suppose that A, B and C are three events.

- If

$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{cases}$$

  then A, B and C are said to be **pairwise independent**.

- If

$$P(ABC) = P(A)P(B)P(C)$$

  then A, B and C are said to be **mutually independent**.

Independence

Definition

Suppose that $A_1, A_2, \cdots, A_n$ are n events.

- For every $1 \leq i < j < k < \cdots \leq n$, if the following equality

$$\begin{cases} P(A_iA_j) = P(A_i)P(A_j), \\ P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k), ; \\ \vdots \\ P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n) \end{cases}$$

hold, then A, B and C are said to be **mutually independent**.