Discrete Mathematics and Its Applications 2 (CS147)

Lecture 9: Conditional probability, independence

Fanghui Liu

Department of Computer Science, University of Warwick, UK

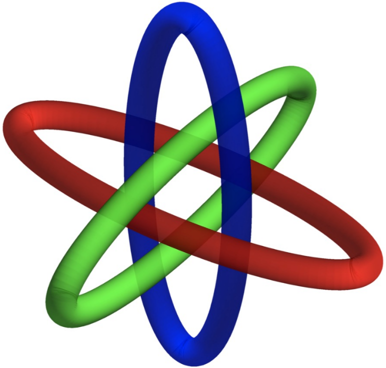


Figure: Borromean rings.

Recall Probability...

- a metric/measure/function f of "event A occurs"

Definition (Probability)

Probability $\Pr : \mathcal{F} \to [0, 1]$ is a function that assigns a value to events

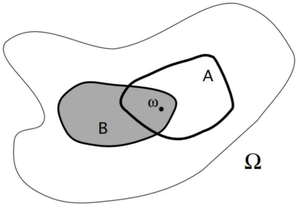
- nonnegativity: $\Pr(A) \geq 0$
- normalization: $\Pr(\Omega) = 1$
- countable additivity: if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$

- $(\Omega, \mathcal{F})$ is a measurable space
- $(\Omega, \mathcal{F}, \Pr)$ is a probability space

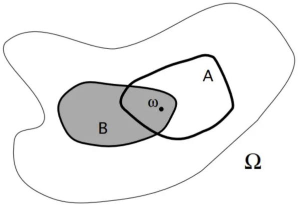
Conditional probability

Problem

If the event B occurs, then what is the probability of event A?



Conditional probability



Problem

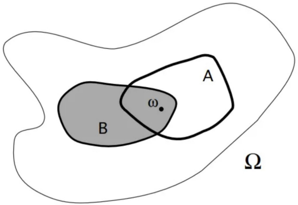
If the event B occurs, then what is the probability of event A?

Remark: Given additional **information**, we infer the outcome of a random trial.

Conditional probability

Problem

If the event B occurs, then what is the probability of event A?



Remark: Given additional **information**, we infer the outcome of a random trial.

Definition (Conditional probability)

Consider any two events $A, B \subseteq \Omega$, if $\Pr(B) > 0$, the conditional probability is

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

Remark: the probability of event A occurs given that event B occurs.

Partition theorem

Definition (Partition)

$\{B_1, \ldots, B_n\} \subseteq \Omega$ be a partition of the sample space Ω if

- $\Omega = \cup_{i=1}^{n} B_i$.
- $\Pr(B_i) > 0, \ \forall i \in [n]$.
- $B_i \cap B_j = \emptyset \ \forall i \neq j$.

Partition theorem

Definition (Partition)

$\{B_1, \ldots, B_n\} \subseteq \Omega$ be a partition of the sample space Ω if

- $\Omega = \cup_{i=1}^n B_i$.
- $\Pr(B_i) > 0, \ \forall i \in [n]$.
- $B_i \cap B_j = \emptyset \ \forall i \neq j$.

Definition (Law of total probability)

Let $\{B_1, \ldots, B_n\} \subseteq \Omega$ be a partition of the sample space Ω. Consider any event $A \subseteq \Omega$, we have

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A|B_i)\Pr(B_i) \, .$$

Remark: a special case: $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c)$.

From reason to result, from result to reason...

- law of total probability: from reason to result
  - A: result/phenomenon

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

  - $\{B_i\}_{i=1}^{n}$: reason

From reason to result, from result to reason...

◦ law of total probability: from reason to result

- A: result/phenomenon

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

- $\{B_i\}_{i=1}^{n}$: reason

In practice, we observe some phenomenon, and then infer which reason(s) cause this.

From reason to result, from result to reason...

- law of total probability: from reason to result
  - A: result/phenomenon

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

  - $\{B_i\}_{i=1}^{n}$: reason

In practice, we observe some phenomenon, and then infer which reason(s) cause this.

- Bayes's theorem: from result to reason
  - $\Pr(B_i|A)$: event A occurs, infer the probability that the event is caused by B_i
  - $\Pr(B_i)$: prior probability

Bayes's theorem

Theorem

Let $\{B_1, \ldots, B_n\} \subseteq \Omega$ *be a partition of the sample space* Ω *such that* $\Pr(B_i) > 0, \ \forall i \in [n]$. *Consider any event* $A \subseteq \Omega$, *we have*

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

Remark: special case with $n = 2$: $\Omega = B \cup B^c$.

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

Example

Example

Consider a disease with an incidence rate of 1 in 10^5 among the population. There is a diagnostic test the disease. For one person:

- If (s)he has this disease, this test is positive with probability at 9/10
- If (s)he doesn't have this disease, the test is positive with probability at 1/20

Question: tested with positive now, what is the probability that he/she has this disease?

Example

Example

Consider a disease with an incidence rate of 1 in 10^5 among the population. There is a diagnostic test the disease. For one person:

- If (s)he has this disease, this test is positive with probability at 9/10
- If (s)he doesn't have this disease, the test is positive with probability at 1/20

Question: tested with positive now, what is the probability that he/she has this disease?

- result/phenomenon: tested with positive (event A)
- reasons:
  - 1) has this disease (event B)
  - 2) mistakes by the test (false alarm)

Example

Example

Consider a disease with an incidence rate of 1 in 10^5 among the population. There is a diagnostic test the disease. For one person:

- If (s)he has this disease, this test is positive with probability at 9/10
- If (s)he doesn't have this disease, the test is positive with probability at 1/20

Question: tested with positive now, what is the probability that he/she has this disease?

- result/phenomenon: tested with positive (event A)
- reasons:
  - 1) has this disease (event B)
  - 2) mistakes by the test (false alarm)

Target: estimate $\Pr(B|A)$

- prior: $\Pr(B) = 10^{-5}$
- $\Pr(A|B) = 0.9$
- $\Pr(A|B^c) = 0.05$

Solutions

- prior: $\Pr(B) = 10^{-5}$
- $\Pr(A|B) = 0.9$
- $\Pr(A|B^c) = 0.05$

Solution

Denote $A =$ *event that he/she is tested with positive;* $B =$ *event that he/she has this disease.*

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

$$= \frac{0.9 \times 10^{-5}}{0.9 \times 10^{-5} + 0.05 \times (1 - 10^{-5})} \approx 0.00018$$

Independence

$\Pr(A|B)$ changes when B changes

Definition (Independence)

Event A and B are **independent** if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

Independence

$\Pr(A|B)$ changes when B changes

Definition (Independence)

Event A and B are **independent** if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

Definition (mutually independent)

A collection of events $A_1, A_2, \ldots, A_k \subseteq \Omega$ are independent if and only if

$$\forall I \subseteq [1,k], \quad \Pr(\cap_{j \in I} A_j) = \prod_{j \in I} \Pr(A_j)\,.$$

Independence

$\Pr(A|B)$ changes when B changes

Definition (Independence)

Event A and B are **independent** if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

Definition (mutually independent)

A collection of events $A_1, A_2, \ldots, A_k \subseteq \Omega$ are independent if and only if

$$\forall I \subseteq [1, k], \quad \Pr(\cap_{j \in I} A_j) = \prod_{j \in I} \Pr(A_j)\,.$$

Definition (pairwise independent)

A collection of events $A_1, A_2, \ldots, A_k \subseteq \Omega$ are pairwise independent if and only if

$$\forall i, j \subseteq [1, k], i \neq j, \quad \Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)\,.$$

Relationship between mutually independent and pairwise independent

- mutually independent $\Rightarrow$ pairwise independent
- pairwise independent $\not\Rightarrow$ mutually independent

Statement

Intuitive idea: two events A, B occur, leading to the case that C occurs

Relationship between mutually independent and pairwise independent

- ∘ mutually independent $\Rightarrow$ pairwise independent
- ∘ pairwise independent $\not\Rightarrow$ mutually independent

Statement

Intuitive idea: *two events A, B occur, leading to the case that C occurs*
mutually independent: $\Pr(ABC) = \Pr(A)\Pr(B)\Pr(C)$
pairwise independent: $\Pr(ABC) = \Pr(A|BC)\Pr(BC) = \Pr(A|BC)\Pr(B)\Pr(C)$

Relationship between mutually independent and pairwise independent

- ∘ mutually independent $\Rightarrow$ pairwise independent
- ∘ pairwise independent $\nRightarrow$ mutually independent

Statement

Intuitive idea: *two events A, B occur, leading to the case that C occurs*
mutually independent: $\Pr(ABC) = \Pr(A)\Pr(B)\Pr(C)$
pairwise independent: $\Pr(ABC) = \Pr(A|BC)\Pr(BC) = \Pr(A|BC)\Pr(B)\Pr(C)$

Example

- Two independent fair coin tosses
  - ∘ A: First toss is H
  - ∘ B: Second toss is H
- C: the two tosses had the same result

Relationship between mutually independent and pairwise independent

∘ mutually independent $\Rightarrow$ pairwise independent
∘ pairwise independent $\nRightarrow$ mutually independent

Statement

Intuitive idea: *two events A, B occur, leading to the case that C occurs*
mutually independent: $\Pr(ABC) = \Pr(A)\Pr(B)\Pr(C)$
pairwise independent: $\Pr(ABC) = \Pr(A|BC)\Pr(BC) = \Pr(A|BC)\Pr(B)\Pr(C)$

Example

- Two independent fair coin tosses
  - ∘ A: First toss is H
  - ∘ B: Second toss is H
- C: the two tosses had the same result

$\Pr(A \cap B) = \frac{1}{4} = \Pr(A)\Pr(B)$
$\Pr(A \cap C) = \frac{1}{4} = \Pr(A)\Pr(C)$ (similar to B)
$\Pr(A \cap B \cap C) = \frac{1}{4} \neq \Pr(A)\Pr(B)\Pr(C) = \frac{1}{8}$

Union bound

Statement

We have

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)\,.$$

Union bound

Statement

We have

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)\,.$$

Definition

Consider any events $A_1, A_2, \ldots, A_k \subseteq \Omega$, then

$$\Pr(A_1 \cup A_2 \cup \ldots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \ldots + \Pr(A_k)\,.$$

One example in Seminar

Problem

Suppose that in your inbox, 70% of all email is spam, 90% of spam emails contain the word "lottery", and 5% of non-spam emails contain the word "lottery". What is the probability that an email selected uniformly at random is actually spam given that it contains the word "lottery"?

*Naive Bayes classifier - Illustration

◦ train a binary classifier h on training data

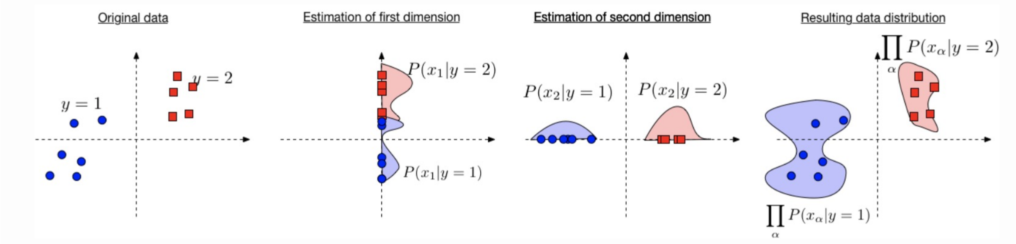


Figure: Illustration behind the Naive Bayes algorithm. source from link.

*Naive Bayes classifier - Illustration

∘ train a binary classifier h on training data

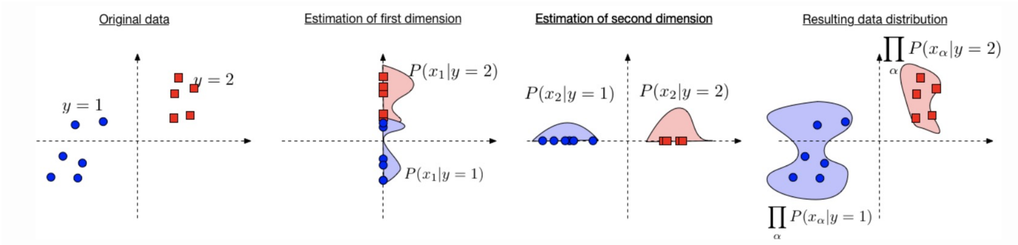


Figure: Illustration behind the Naive Bayes algorithm. source from link.

$$h(\boldsymbol{x}) = \arg\max_y \Pr(y|\boldsymbol{x}) = \arg\max_y \frac{\Pr(\boldsymbol{x}|y)\Pr(y)}{\Pr(\boldsymbol{x})} = \arg\max_y \prod_{\alpha=1}^{d} \Pr(x_\alpha|y)\Pr(y)\,.$$

- density estimation for $\Pr(\boldsymbol{x}|y) \rightarrow$ curse of dimensionality
- Assumption: features are conditionally independent given the label