Independence

Lemma. If A and B are independent, then:

(a) A and B^c are independent,

(b) A^c and B are independent,

(c) A^c and B^c are independent

Proof. We prove (a), the rest can be proved in a similar way.

Independent vs. disjoint

One common mistake is to confuse independence and being disjoint. These are **different** concepts.

When two events *A* and *B* are **disjoint** it means that if one of them occurs, the other one cannot occur, that is,

$$A \cap B = \emptyset$$

There is already some sort of dependence, because $A \subseteq B^c$.

Lemma. Consider two events A and B, with $P(A) \neq 0$ and $P(B) \neq 0$. If *A* and *B* are disjoint, then they are not independent.

Proof. Since and are disjoint, we have

$$P(A \cap B) = 0 \neq P(A)P(B)$$

Thus, *A* and *B* are not independent. ☐

Law of **Total Probability**

For any two events A and B,

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

and using the definition of conditional probability,

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

we can write

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

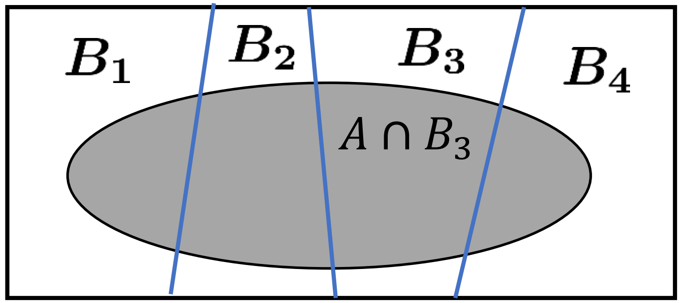
Law of **Total Probability**

If

$$\{B_i \mid i \in I\}$$

is a **partition** of the sample space S, then for any event , we have

$$P(A) = \sum_{i \in I} P(A \cap B_i) = \sum_{i \in I} P(A|B_i)P(B_i)$$



Example

I have three bags that each contain 100 marbles:

- Bag 1 has 90 red and 10 blue marbles
- Bag 2 has 58 red and 42 blue marbles
- Bag 3 has 25 red and 75 blue marbles

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

Bayes' rule

We are ready to state one of the most useful results in conditional probability: **Bayes' rule.**

Suppose that we know $P(A|B)$, but we want to know the probability $P(B|A)$?

Using the definition of conditional probability, we have

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

We can solve

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

Example

Let us continue the marbles example. Suppose we observe that the chosen marble is red. What is the probability that Bag 1 was chosen?

Solution: We find $\mathrm{P}(B_1|R)$ using Bayes' rule.

Bayes' rule (general version)

Often, in order to find $P(A)$ in Bayes' formula we will use the law of total probability, so sometimes Bayes' rule is stated as

$$P\left(B_j \middle| A\right) = \frac{P\left(A \middle| B_j\right) P(B_j)}{\sum_{i=1}^{n} P(A|B_i) P(B_i)}$$

where $B_1, B_2, B_3, \dots, B_n$ form a partition of the sample space

Spam detection

Assume we have learned that the word ”free” appears in 36% of the emails marked as spam.

On the other hand, 5% of non-spam mail includes the same word ”free”.

Assume that 30% of all emails received by a certain user is **spam mail**. Find the probability that a mail is spam when the word ”free” appears in it.

- $P(\text{free} \mid \text{spam}) = 0.36$ (spam with word ”free”)
- $P(\text{free} \mid \text{nonspam}) = 0.05$ (non-spam with ”free”)
- $P(\text{spam}) = 0.30 \implies P(\text{nonspam}) = 0.70$

By Bayes formula:

$$P(\text{spam} \mid \text{free}) = \frac{P(\text{spam}) \times P(\text{free} \mid \text{spam})}{P(\text{free})}$$

Spam detection

P(free) is the total probability:

$$P(\text{free}) = P(\text{spam}) \times P(\text{free} \mid \text{spam}) + P(\text{nonspam}) \times P(\text{free} \mid \text{nonspam})$$
$$= 0.30 \times 0.36 + 0.70 \times 0.05 = 0.143$$

The probability that a mail containing ”free” is spam:

$$P(\text{spam} \mid \text{free}) = (0.30 \times 0.36) \,/\, 0.143 = 0.755$$

The probability that a mail containing ”free” is not spam:

$$P(\text{nonspam} \mid \text{free}) = (0.70 \times 0.05) \,/\, 0.143 = 0.245$$

Supervised classification means identifying to which category a new observation belongs, on the basis of a training set of data containing observations whose category membership is known (human-guided)

3.1 Random variables

A ***random variable*** *X* gets its value from a random experiment.

Example. Toss a coin four times:

$$S = \{TTTT, TTTH, \cdots, HHHH\}.$$

The random variable *X* = “the number of heads” can have the five values: $0, 1, 2, 3, 4$

Range

Random variables are denoted by capital letters X, Y, Z

Range of X is the set of possible values for X.
In the previous example, range of *X* is $\{0,1,2,3,4\}$.

A random variable *X* is a ***discrete random variable***, if its range is *countable*. This means that range of *X* is *finite* or the range is *countably infinite*:

$$\{x_1, x_2, x_3, \dots\}$$

Range

Let X be a random variable.

We denote the range of X simply by R_X.

This means that R_X are the possible values of X.

For $k \in X$, we are interested in knowing the probability of $X = k$

We denote this probability by $P(X = k)$ ("the probability that X gets the value k")

Probability mass function

The function

$$P: R_X \to [0,1]$$

is called the ***probability mass function*** (PMF) of X

$P(X = x_k)$ is denoted also by $P_X(k)$

Properties of the mass function

1. For all $k \in R_X,\ 0 \leq P(X = k) \leq 1$

2. $\sum_{k \in R_X} P(X = k) = 1$

3. For any set $A \subseteq R_X, \sum_{k \in A} P(X = k) \leq 1$

Various distributions

- There are some specific distributions that are given special names.

- There is a certain kind of random experiment behind each of these distributions.

- We will provide PMFs for all of these special random variables. You should try to understand the random experiment behind each of them. If you understand the random experiments, you can simply derive the PMFs when you need them.

- Even it seems that there are a lot of formulas in this section, there are in fact very few new concepts.

Bernoulli distribution

A ***Bernoulli random variable*** is a random variable that can only take two possible values, usually 0 and 1, sometimes referred to as "success" and "failure"

Examples

- You take an exam. You either pass (resulting in $X = 1$) or fail (resulting in $X = 0$).
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender is either male or female.

Bernoulli distribution

Denote:

$$\boldsymbol{X \sim Bernoulli(p)}$$

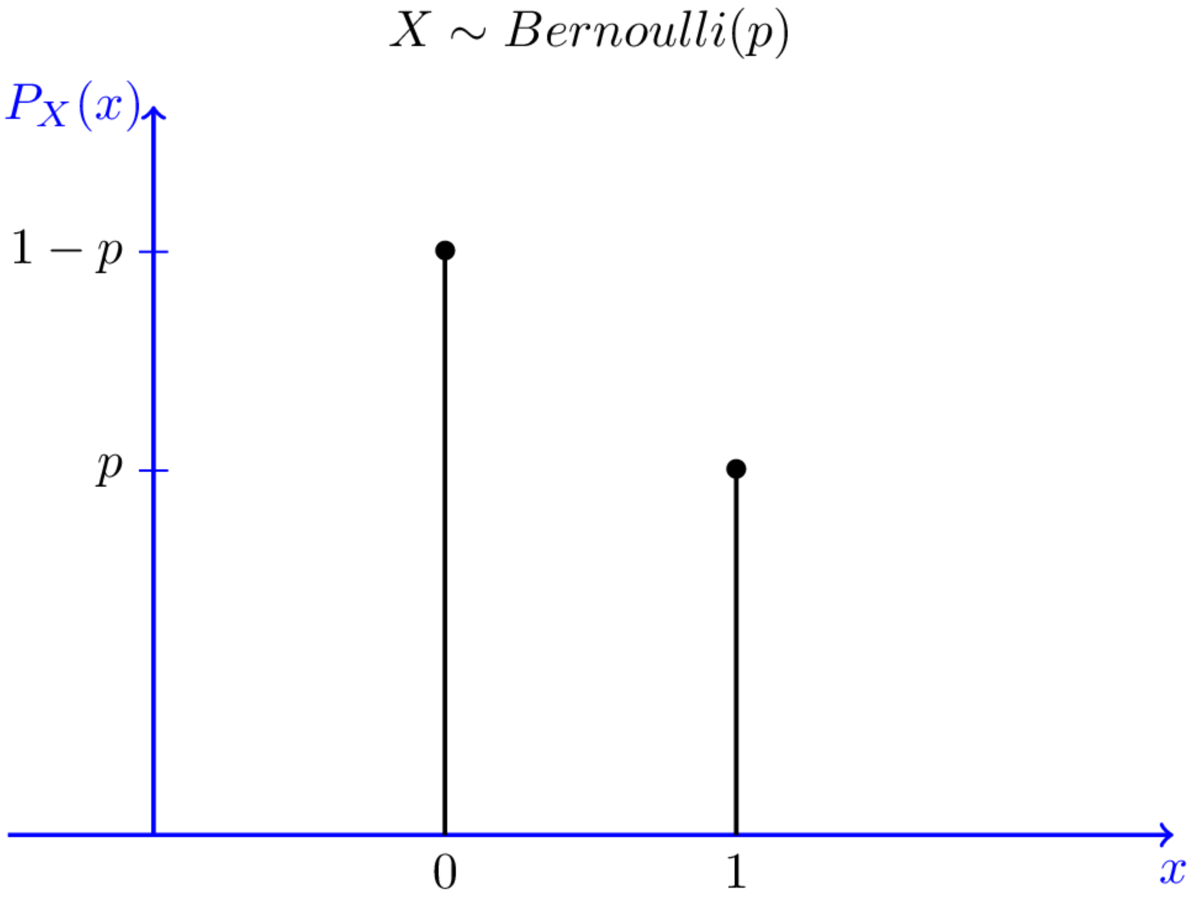
Probability mass function:

$$P_X(x) = \begin{cases} p & \text{for } x = 1 \\ 1-p & \text{for } x = 0 \end{cases} \quad \text{Range}(X) = \{0, 1\}.$$

that is,

$$P_X(0) = 1 - p, \quad P_X(1) = p.$$

Bernoulli distribution



Geometric distribution

Suppose that I have a coin (not necessarily fair) with $P(H) = p$. I toss the coin **until** I observe the first heads. Define *X* as the total number of coin tosses in this experiment.

$$\boldsymbol{X \sim Geometric(p)}$$

The range of *X* is $R_X = \{1,2,3, \ldots\}$.

Probability mass function:

$$\boldsymbol{P_X(k) = p(1-p)^{k-1}, \qquad k = 1, 2, 3, \ldots}$$

Geometric distribution