

# Week 1

---

## Probability

Given a random experiment (say, rolling a die), a probability measure is a population quantity that summarizes the randomness.

Specifically, probability takes a possible outcome from the experiment and:

- assigns it a number between 0 and 1 ;
- so that the probability that something occurs is 1 and;
- so that the probability of the union of any two sets of outcomes that have nothing in common (mutually exclusive) is the sum of their respective probabilities.

### Rules probability must follow

The probability that nothing occurs is 0 :  $P(\Omega^C) = P(\emptyset) = 0$  .

The probability that something occurs is 1:  $P(\Omega) = 1$  .

The probability of something is 1 minus the probability that the opposite occurs:

$$P(A^C) = 1 - P(A)$$

The probability of at least one of two (or more) things that cannot simultaneously occur (mutually exclusive) is the sum of their respective probabilities:  $A \perp B \Rightarrow P(A \cup B) = P(A) + P(B)$  .

If an event  $A$  implies the occurrence of event  $B$ , then the probability of  $A$  is less than the probability that  $B$  occurs:  $A \in B \Rightarrow P(B) \geq P(A)$  .

For any two events, the probability that at least one occurs is the sum of their probabilities minus their intersection:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ,  $\forall A, B \in \Omega$  .

---

## Random variables

A random variable is a numerical outcome of an event.

- It can be discrete or continuous.

### Probability Mass Function (PMF)

A probability mass function evaluated at a value  $x$  corresponds to the probability that a random variable  $X$  takes that value:  $p(x) = P(X = x)$  .

To be a valid PMF, a function  $p$  must satisfy:

1. It must always be larger or equal to 0:  $p(x) \geq 0$ ,  $\forall x \in \Omega$  .
2. The sum of the possible values that the random variable can take has to add up to one:

$$\sum_{x \in \Omega} p(x) = 1 .$$

### Example:

$X = 0$  represents tails and  $X = 1$  represents heads.

$$p(x) = \theta^x (1 - \theta)^{1-x}, x \in \{0, 1\}$$

### Probability Density Function (PDF)

A probability density function is a function associated with a continuous variable.

To be a valid PDF, a function  $f$  must satisfy:

1. It must be larger than or equal to zero everywhere:  $f(x) \geq 0, \forall x \in \text{Dom}(X)$ .
2. The total area under it must be one:  $\int_{\Omega} f(x)dx = 1$ .

Areas under PDFs correspond to probabilities for that random variable:

$$f(a \leq X \leq b) = \int_a^b f(x)dx.$$

**Example:**

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Special case of the [Beta distribution](#):  $f(x | \alpha, \beta) = c \cdot x^{\alpha-1} (1-x)^{\beta-1}$ , with the constant  $c = 2$ ,  $\alpha = 2$  and  $\beta = 1$ .

### Cumulative Density Function (CDF) and Survival Function

The cumulative distribution function (CDF) of a random variable  $X$  returns the probability that the random variable is less than or equal to the value  $x$ :

$$F(x) = P(X \leq x)$$

It is important to note that  $\frac{d}{dx} F(x) = f(x) \Leftrightarrow F(x) = \int f(x)dx$ .

The survival function of a random variable  $X$  is defined as the probability that the random variable is greater than the value  $x$ :

$$S(x) = P(X > x)$$

Note that  $S(x) = 1 - F(x)$ .

### Quantiles

The  $\alpha$ -th quantile of a distribution with a CDF  $F$  is the point  $x_{\alpha}$  so that:

$$F(x_{\alpha}) = \alpha$$

---

We are referring to population quantities. Therefore, the median being discussed is the population median.

- A probability model connects the data to the population using assumptions.
  - Therefore, the median we are discussing is the estimand, the sample median will be the estimator.
- 

### Conditional probability

Let  $B$  be an event so that  $P(B) > 0$ .

Then, the conditional probability of an event  $A$ , given that  $B$  has occurred is:

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

If  $A$  and  $B$  are independent, that is,  $A \perp B$ :  $P(A \cap B) = P(A)P(B)$ .

- Then,  $P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$ .

### Bayes' rule

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

### Example

Let + and – be the events that the result of a diagnostic test is positive or negative, respectively.

Let  $D$  and  $D^C$  be the events that the subject of the test has or does not have the disease, respectively.

The sensitivity is the probability that the test is positive, given that the subject has the disease, that is,  $\text{Sensitivity} = P(+|D)$ .

The specificity is the probability that the test is negative, given that the subject does not have the disease, *id est*,  $\text{Specificity} = P(-|D^C)$ .

If a positive test is obtained, one may be interested in the positive predictive value:  $P(D|+)$ .

If a negative test is obtained, one may be interested in the negative predictive value:  $P(D^C|-)$ .

In the absence of a test, the probability of presenting the disease  $P(D)$  is called the prevalence of disease.

Suppose that a test for a disease had sensitivity of 99.7% and specificity of 98.5%. Suppose that a subject from a population with a 0.1% prevalence of that disease receives a positive test result. What is the associated positive predictive value?

Using Bayes' formula:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^C)P(D^C)}$$

Knowing that  $P(+|D^C) = 1 - P(-|D^C)$  and  $P(D^C) = 1 - P(D)$ :

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + [1 - P(-|D^C)][1 - P(D)]} \\ &= \frac{.997 \times .001}{.997 \times .001 + .015 \times .999} = .062 \end{aligned}$$

### Likelihood ratios

$$\frac{P(D|+)}{P(D^C|+)} = \frac{P(+|D)}{P(+|D^C)} \times \frac{P(D)}{P(D^C)}$$

- $\frac{P(D|+)}{P(D^C|+)}$ : odds of disease given a positive test result.
- $\frac{P(D)}{P(D^C)}$ : odds of disease in the absence of a test result.
- $\frac{P(+|D)}{P(+|D^C)}$ : diagnostic likelihood ratio for a positive test result  $\text{DLR}_+$ .

post-test odds of  $D = \text{DLR}_+ \times \text{pre-test odds of } D$

The diagnostic likelihood ratio of a positive test result  $\text{DLR}_+$  is the factor by which you multiply your odds, in the presence of a positive test, to obtain your post-test odds.

## Example

Returning to the previous example:

$$\text{DLR}_+ = \frac{.997}{1 - .985} \approx 66$$

The hypothesis of disease is 66 times more supported by the data than the hypothesis of no disease, given a positive test result.

$$\text{DLR}_- = \frac{1 - .997}{.985} \approx .003$$

The hypothesis of disease is supported 0.003 times that of the hypothesis of the absence of disease, given the negative test result.

## Independence

Event  $A$  is independent of event  $B$  if  $P(A \cap B) = P(A)P(B) \Leftrightarrow P(A|B) = P(A)$

### Independent and identically distributed (IID) variables

Independent: statistically unrelated from one another.

Identically distributed: all having been drawn from the same population distribution.

---

## Expected values

Our sample expected values will estimate the population versions.

### The population mean

The expected value  $E[X]$  or mean of a random variable  $X$  is the center of its distribution.

$E[X]$  represents the center of mass of a collection of locations and weights,  $\{x, p(x)\}$ .

For a discrete random variable  $X$ :  $E[X] = \sum_x xp(x)$ .

For a continuous random variable  $X$ :  $E[X] = \int_{-\infty}^{\infty} xf(x)dx$ .

### The sample mean

The center of mass of the data is the empirical mean.

For a random variable  $X$ :  $\bar{X} = \sum_{i=1}^n x_i p(x_i)$ .

### Examples:

Coin flip: suppose that random variable  $X$  is so that  $P(X = 1) = p$  and  $P(X = 0) = 1 - p$ .

- Then,  $E[X] = 0 \times (1 - p) + 1 \times p = p$ .

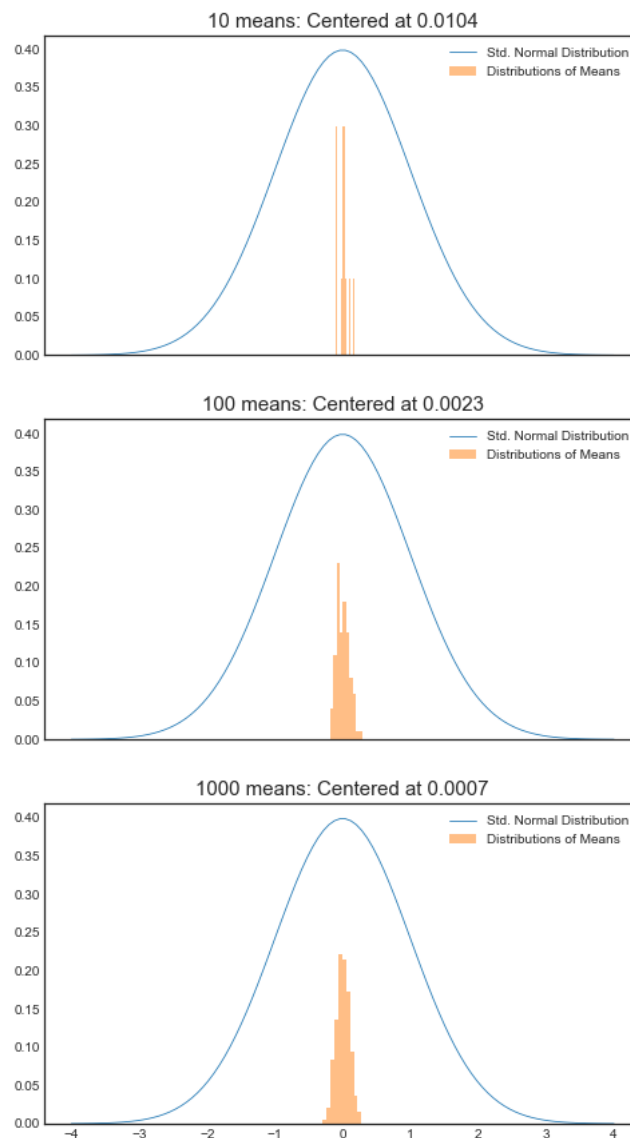
Die roll: suppose that a die is rolled and  $X$  is the number face up.

- Then,  $E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$

### Facts about expected values

- Expected values are properties of distributions.

- The average of random variables is itself a random variable and its associated distribution has an expected value.
  - The center of this distribution is the same as that of the original distribution.
  - The expected value of the sample mean is exactly the population mean that it is trying to estimate:  $E[\bar{X}] = E[X]$ .
  - The more data goes into the sample mean, the more concentrated its density/mass function is around the population mean.



When the distribution of an estimator  $\hat{\theta}$  is centered at the same point as the estimand  $\theta$ , that is, when  $E[\hat{\theta}] = \theta$ , the estimator is said to be unbiased. More [here](#).

- The expected value is a linear function, so:
  - $E[c \cdot X] = c \cdot E[X]$ , where  $c$  is a constant;
  - $E[X + Y] = E[X] + E[Y]$ ;
  - More generally,  $E[a \cdot X + b \cdot Y] = a \cdot E[X] + b \cdot E[Y]$

