Probability

Contents

Random process

In a **random process**, we know what outcomes could happen, but we don't know which particular outcome will happen.

Probability

There are several possible interpretations of probability but they almost completely agree on the mathematical rules probability must follow:

$$0 \leq P(A) \leq 1$$

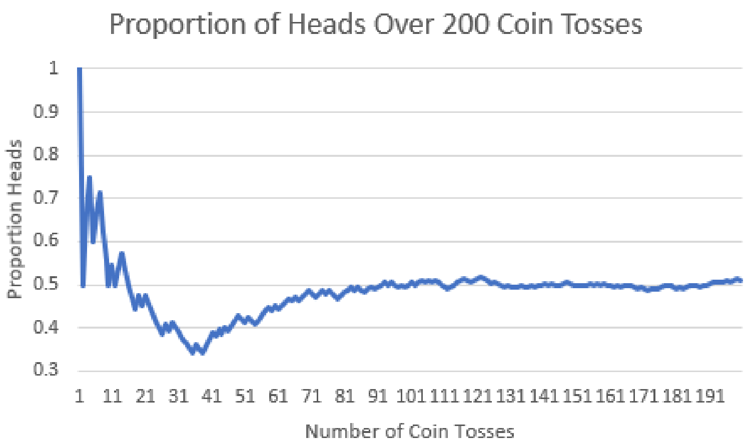
Where $P(A)$ is the probability of event A happening.

- **Frequentist interpretation**: he probability of an outcome is the **proportion** of the times the outcome would occur if we observed the random process an infinite number of times.
- **Bayesian interpretation**: a probability is a subjective degree of belief. For the same event, two separate people could have different viewpoints and so assign different probabilities to it. This interpretation allows for **prior information** to be integrated into the inferential framework.

Defining probability

Law of large numbers

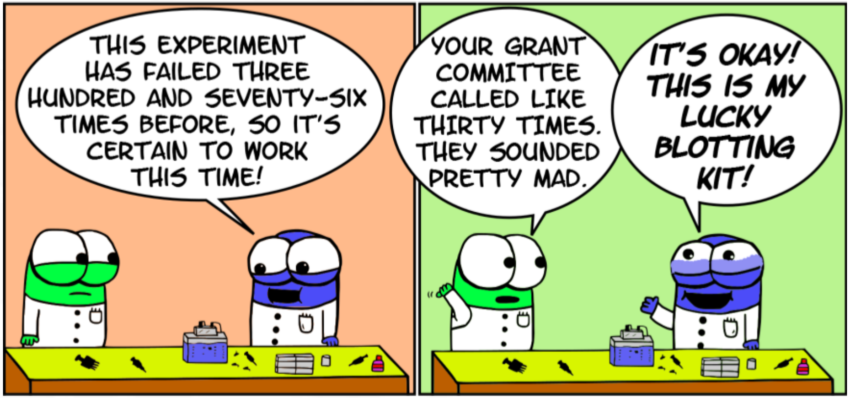
The **law of large numbers** states that as more observations are collected, the **proportion of occurrences** with a particular outcome **converges to the probability** of that outcome. This is why, as we roll a fair die many times, we expect the proportion of say, fives, to settle down to one-sixth.



Defining probability

Gambler’s fallacy

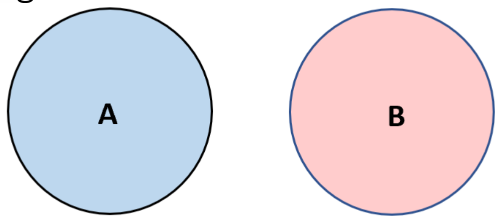
The common misunderstanding of the law of large numbers is that random processes are supposed to compensate for whatever happened in the past. This is called the **gambler's fallacy**, or the law of averages. So, while we know that in a large number of tosses of a coin we would expect about 50% heads and 50% tails, **for any given toss the probability of a head or tail is exactly 0.5**, regardless of what happened in the past.



Disjoint Events

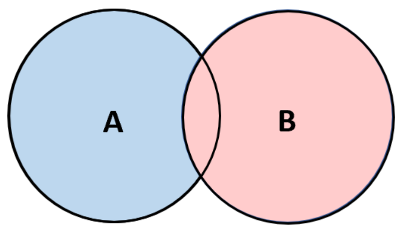
Definition

Disjoint events by definition, cannot happen at the same time. A synonym for this term is **mutually exclusive**. For example, the outcome of a single coin toss cannot be a head and a tail.



$$P(A\ and\ B) = 0$$

Non-disjoint events can happen at the same time.



$$P(A\ and\ B) > 0$$

Disjoint Events

Union of events

For disjoint events, probability of event A or event B happening is given by:

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

For non-disjoint events, probability of event A or event B happening is given by:

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

But... That is the same formula, isn't it?

Second formula is called the **general addition rule**.

Disjoint Events

Union of events

What is the probability that a random sampled student thinks marijuana should be legalized or they agree with their parents' political views?

	Share Parents' Politics		
Legalize MJ	No	Yes	Total
No	11	40	51
Yes	36	78	114
Total	47	118	165

Union of events

What is the probability that a random sampled student thinks marijuana should be legalized or they agree with their parents' political views?

	Share Parents' Politics		
Legalize MJ	No	Yes	Total
No	11	40	51
Yes	36	78	114
Total	47	118	165

$$P(A) = 114/165$$
$$P(B) = 118/165$$
$$P(A \cap B) = 78/165$$
$$P(A \cup B) = \frac{(114 + 118 - 78)}{165} \approx 0.93$$

Disjoint Events

Sample space

A sample space is a collection of **all possible outcomes** of a trial.

The sample space of throwing a dice twice is:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Disjoint Events

Probability distribution

A probability distribution raises from a sample space and **lists all possible outcomes and the probabilities with which they occur**.

Rules:

1. The events must be disjoint.
2. Each probability must be between 0 and 1
3. The sum of the probabilities must total 1.

Independence

Definition

Two processes are independent if knowing the outcome of one provides no useful information about the outcome of the other.

- Outcomes of two tosses of a coin are independent
- Outcomes of two draws from a deck of cards are dependent.

If $P(A \mid B) = P(A)$, then A and B are independent

Product rule for independent events: $P(A \cap B) = P(A) \times P(B)$

Probability example
World Values Surveys

The World Values Survey is an ongoing worldwide survey that polls the world population about perceptions of life, work, family, politics, etc.

The most recent phase of the survey that polled 77,882 people from 57 countries estimates that 36.2% of the world's population agree with the statement "Men should have more right to a job than women."

The survey also estimates that 13.8% of people have a university degree or higher, and that 3.6% of people fit both criteria.

(1) Are agreeing with the statement "Men should have more right to a job than women" and having a university degree or higher disjoint events?

(2) Draw a Venn diagram summarizing the variables and their associated probabilities.

Probability example
World Values Surveys

(3) What is the probability that a randomly drawn person has a university degree or higher or agrees with the statement about men having more right to a job than women?

(4) What percent of the world population do not have a university degree and disagree with the statement about men having more right to a job than women?

(5) Does it appear that the event that someone agrees with the statement is independent of the event that they have a university degree or higher?

(6) What is the probability that at least 1 in 5 randomly selected people agree with the statement about men having more right to a job than women?

Conditional Probabilities
Use case

ADOLESCENTS' UNDERSTANDING OF SOCIAL CLASS

study examining teens' beliefs about social class

sample: 48 working class and 50 upper middle class 16-year-olds

study design:

- "objective" assignment to social class based on self-reported measures of both parents' occupation and education, and household income
- "subjective" association based on survey questions

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Conditional Probabilities
Marginal Probabilities

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Marginal probabilities are probabilities that can be calculated using the margins of the contingency table.

- What is the probability that a student's objective social class position is upper middle class? → $P(obj\ UMC) = \frac{50}{98} \approx 0.51$

Conditional Probabilities

Joint Probabilities

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Joint probabilities are probabilities that can be calculated using data which are at the intersection of the two events at the contingency table.

- What is the probability that a student's objective position and subjective identity are both upper middle class?→ $P(obj\ UMC \cap sub\ UMC) = \frac{37}{98} \approx 0.38$

Conditional Probabilities

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Conditional probabilities are joint probabilities where instead of the total contingency only a subgroup of the population is considered given a known condition.

- What is the probability that a student who is objectively in the working class associates with upper middle class?→ $P(sub\ UMC \mid obj\ WC) = \frac{8}{48} \approx 0.17$

Conditional Probabilities

Bayes Theorem

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Formally, conditional probabilities are calculated using the Bayes theorem:

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

- What is the probability that a student who is objectively in the working class associates with upper middle class?→

$$P(sub\ UMC \mid obj\ WC) = \frac{8/98}{48/98} \approx 0.17$$

Conditional Probabilities

Bayes Theorem

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Formally, conditional probabilities are calculated using the Bayes theorem:

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

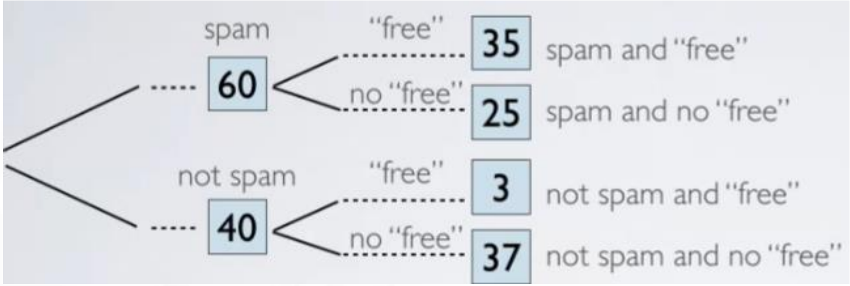
From this expression, general product rule is obtained without assuming independence of events:

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

Conditional Probabilities

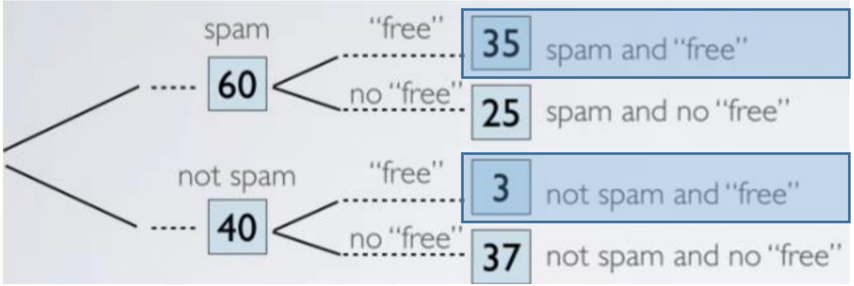
Probabilities tree

You have 100 emails in your inbox. 60 of them are spam, and 40 are not.Of the 60 spam emails, 35 contain the word free. Of the rest, only three contain the word free. If an email contains the word free, what is the probability that it is spam?



Probabilities tree

You have 100 emails in your inbox. 60 of them are spam, and 40 are not.Of the 60 spam emails, 35 contain the word free. Of the rest, only three contain the word free. If an email contains the word free, what is the probability that it is spam?

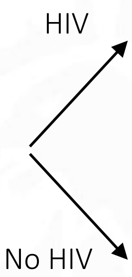


$$P(spam \mid "free") = \frac{35}{35+3} = 0.92$$

Conditional Probabilities

Probabilities tree

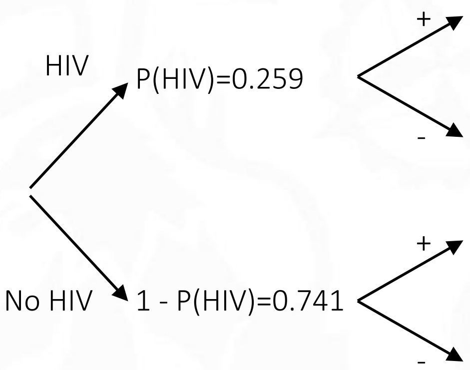
Swaziland, has the highest HIV problems in the world. 25.9% of this country's population is infected with HIV. The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. Note, by the way, that these probabilities are estimates. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?



Conditional Probabilities

Probabilities tree

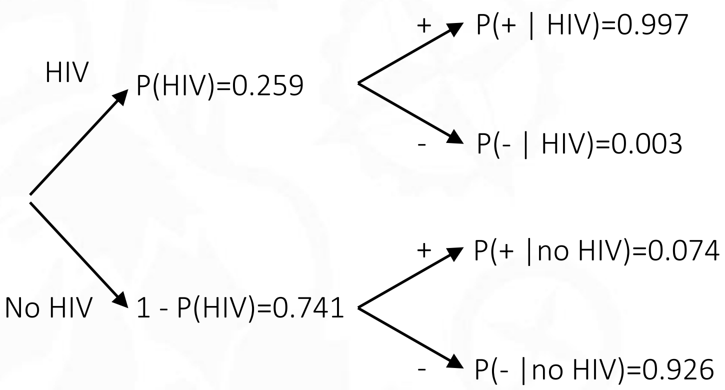
Swaziland, has the highest HIV problems in the world. 25.9% of this country's population is infected with HIV. The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. Note, by the way, that these probabilities are estimates. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?



Conditional Probabilities

Probabilities tree

Swaziland, has the highest HIV problems in the world. 25.9% of this country's population is infected with HIV. The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. Note, by the way, that these probabilities are estimates. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?

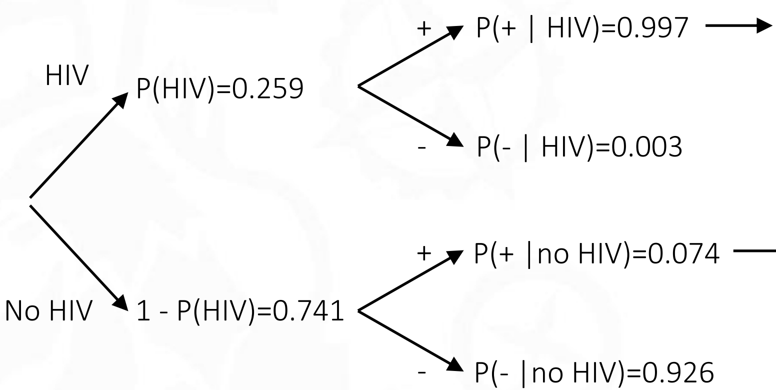


$$P(HIV|+) = \frac{P(HIV \cap +)}{P(+)}$$

Conditional Probabilities

Probabilities tree

Swaziland, has the highest HIV problems in the world. 25.9% of this country's population is infected with HIV. The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. Note, by the way, that these probabilities are estimates. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?



$$P(HIV \cap +) = 0.997 \cdot 0.259 = 0.2582$$

$$P(no\ HIV \cap +) = 0.741 \cdot 0.074 = 0.0548$$

$$P(HIV|+) = \frac{P(HIV \cap +)}{P(+)} = \frac{0.2582}{0.2582 + 0.0548}$$

Set-up

We have two dices, one in each hand:

6-sided dice

12-sided dice

The ultimate goal of the game is to guess which hand is holding which die, but this is more than just a guessing game. Before you make a final decision, you will be able to collect data by asking me to roll the die in one hand, and I'll tell you whether the outcome of the roll is greater than or equal to 4.

What it means to roll a number greater than or equal to 4?

Set-up

6-sided dice

$$S = \{1, 2, 3, 4, 5, 6\}$$
$$P(x \geq 4 \mid s = 6) =$$

12-sided dice

$$S = \{1, 2, 3, 4, 5, 6, 7, 8.9, 10, 11, 12\}$$
$$P(x \geq 4 \mid s = 12) =$$

Set-up

6-sided dice

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$P(\geq 4 \mid s = 6) = \frac{3}{6} = \frac{1}{2} = 0.5$$

12-sided dice

$$S = \{1, 2, 3, 4, 5, 6, 7, 8. 9, 10, 11, 12\}$$

$$P(x \geq 4 \mid s = 12) = \frac{8}{12} = \frac{3}{4} = 0.75$$

Rules

The ultimate goal of the game is to guess which hand is holding which die.

1. I keep one die in the left hand and the other in the right.
2. You pick a hand, left or right. I roll it; and I tell you if the outcome is greater or equal to 4 or not. **I do not tell you the actual number.**
3. Based on that piece of information, you make a decision as to which hand holds which dice.
4. You could ask me to roll again, but each round cost you money.

Bayesian Inference

Rules

The ultimate goal of the game is to guess which hand is holding which die.

1. I keep one die in the left hand and the other in the right.
2. You pick a hand, left or right. I roll it; and I tell you if the outcome is greater or equal to 4 or not. **I do not tell you the actual number.**
3. Based on that piece of information, you make a decision as to which hand holds which dice.
4. You could ask me to roll again, but each round cost you money.

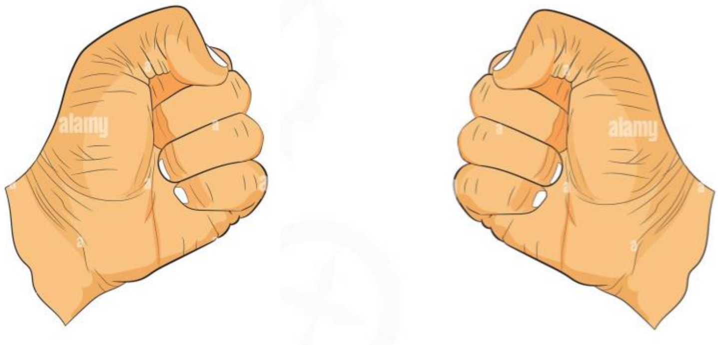
		Truth	
		Right = 12	Left = 12
Decision	Right = 12	You win	You lose
	Left = 12	You lose	You win

Prior probabilities

Two hypothesis:

- H1 = 12 sided dice on the Right
- H2 = 12 sided dice on the Left

$$P(H_1) = P(H_2) = 0.5$$

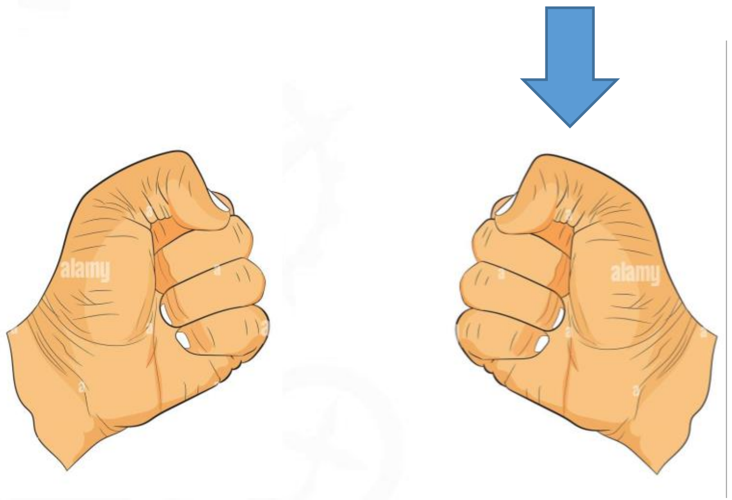


Take a sample

Two hypothesis:

- H1 = 12 sided dice on the Right
- H2 = 12 sided dice on the Left

$$P(H_1) = P(H_2) = 0.5$$

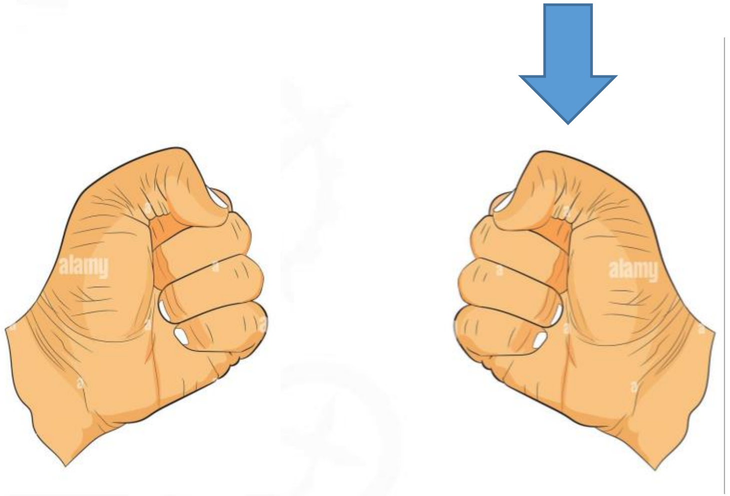


Bayesian Inference
Take a sample

Two hypothesis:

- H1 = 12 sided dice on the Right
- H2 = 12 sided dice on the Left

$$P(H_1) = P(H_2) = 0.5$$

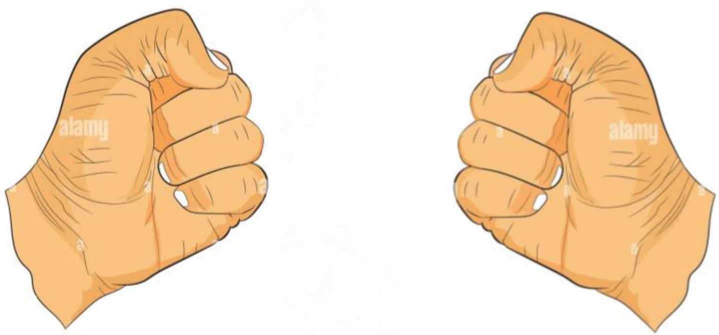


$\geq \mathbf{4}$

Posterior probabilities

Having observed the data point, how, if at all, do the probabilities of the hypothesis change?

- $P(H_1)$ is more than 0.5, $P(H_2)$ is less than 0.5
- $P(H_1)$ is less than 0.5, $P(H_2)$ is more than 0.5
- Probabilities remain the same



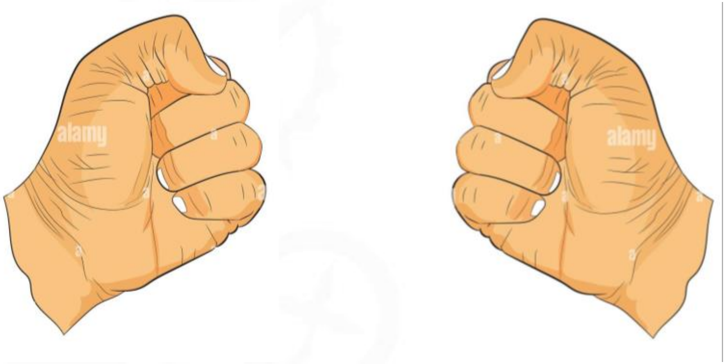
Bayesian Inference

Posterior probabilities

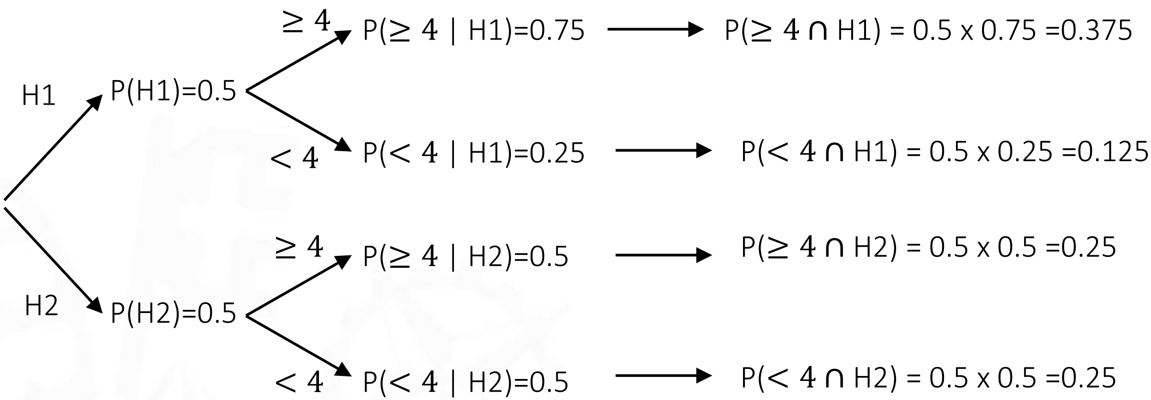
Having observed the data point, how, if at all, do the probabilities of the hypothesis change?

- $P(H_1)$ is more than 0.5, $P(H_2)$ is less than 0.5
- $P(H_1)$ is less than 0.5, $P(H_2)$ is more than 0.5
- Probabilities remain the same

But… how much have they changed?

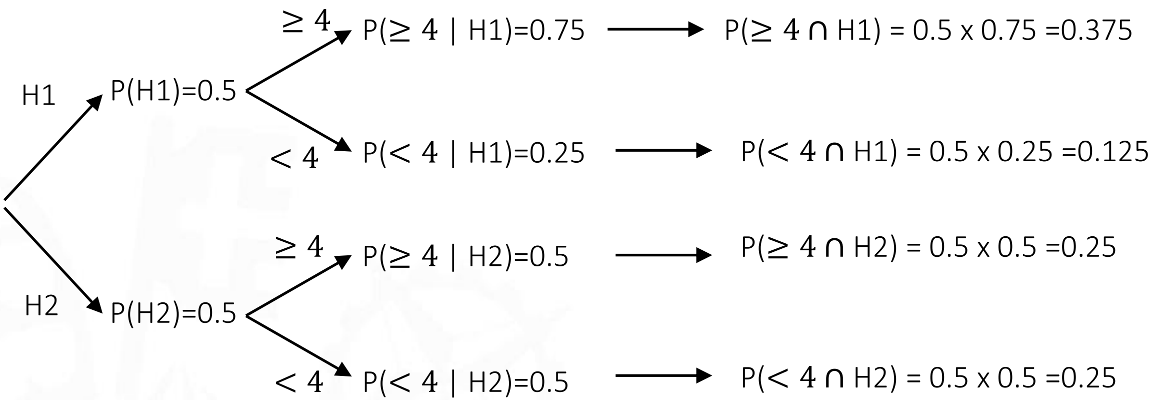


Posterior probabilities



$$P(H_1 \mid \geq 4) = \frac{P(\geq 4 \cap H_1)}{P(\geq 4)} = \frac{P(\geq 4 \cap H_1)}{P(\geq 4 \cap H_1) + P(\geq 4 \cap H_2)} = \frac{0.375}{0.375 + 0.25} = 0.6$$

Posterior probabilities



$$P(H_1 \mid \geq 4) = \frac{P(\geq 4 \cap H_1)}{P(\geq 4)} = \frac{P(\geq 4 \cap H_1)}{P(\geq 4 \cap H_1) + P(\geq 4 \cap H_2)} = \frac{0.375}{0.375 + 0.25} = 0.6$$

Posterior probability is generally defined as P(Hypothesis | data). This is different than the p-value, which is P(data | hypothesis).

Bayesian Inference

Posterior probabilities

H1 → P(H1)=0.6

- ≥ 4 → P(≥ 4 | H1)=0.75 ⟶ P($\geq 4 \cap$ H1) = 0.6 x 0.75 =0.45
- < 4 → P(< 4 | H1)=0.25 ⟶ P($< 4 \cap$ H1) = 0.6 x 0.25 =0.15

H2 → P(H2)=0.4

- ≥ 4 → P(≥ 4 | H2)=0.5 ⟶ P($\geq 4 \cap$ H2) = 0.4 x 0.5 =0.2
- < 4 → P(< 4 | H2)=0.5 ⟶ P($< 4 \cap$ H2) = 0.4 x 0.5 =0.2

$$P(H_1 \mid \geq 4) = \frac{P(\geq 4 \cap H_1)}{P(\geq 4)} = \frac{P(\geq 4 \cap H_1)}{P(\geq 4 \cap H_1) + P(\geq 4 \cap H_2)} = \frac{0.45}{0.45 + 0.2} \approx 0.7$$

Definition

A probability distribution is a mathematical function that describes the **probability of different possible values of a random variable**. Probability distributions are often depicted using graphs or probability tables.

A probability distribution is an **idealized frequency distribution**. A frequency distribution describes a specific sample or dataset. It is the number of times each possible value of a variable occurs in the dataset. More specifically, **the probability of a value is its relative frequency in an infinitely large sample**.

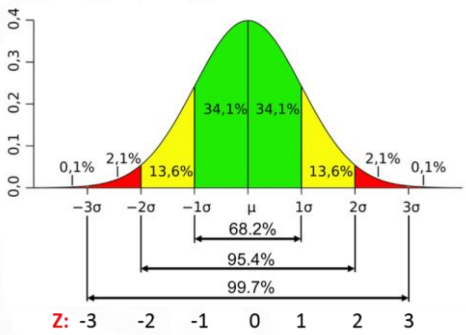
Normal distribution

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

$$P(x) = \frac{1}{\sigma\sqrt{2\cdot\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$$

Z-score is the number of standard deviation of which a samples is away from the mean:

$$z = \frac{x-\mu}{\sigma}$$

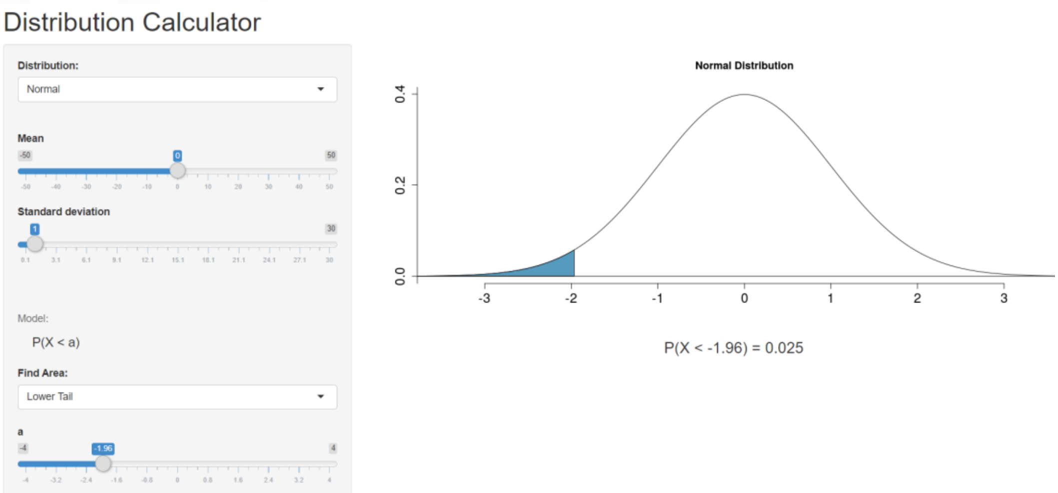


Probability distribution

Percentile

A percentile is the value in a distribution that has a specified percentage of observations below it.

Check them here: https://gallery.shinyapps.io/dist_calc/



Binomial distribution

A Bernoulli trial (or binomial trial) is a random experiment with exactly two possible outcomes, “success” and “failure”, in which the probability of success is the same every time the experiment is conducted.

The binomial distribution describes the probability of having exactly **k** successes in **n** independent Bernoulli trials with probability of success **p**.

$$P(k\ successes\ in\ n\ trials) = \binom{n}{k} p^k \cdot (1-p)^{(n-k)}$$

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Trials must follow the binomial conditions:

1. Trials must be independent
2. The number of trials must be fixed
3. Each trial outcome must be classified as a success or a failure
4. The probability of success must be the same for every trial.

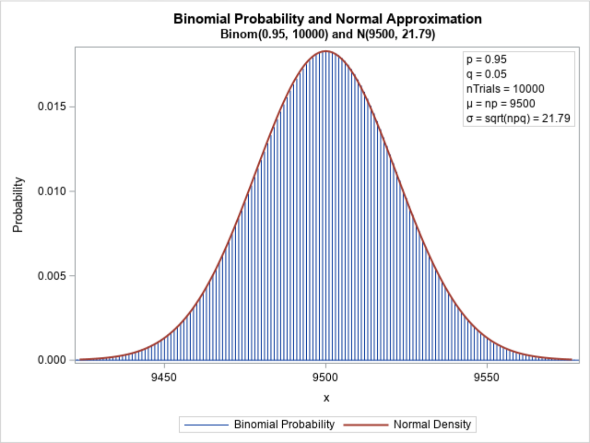
Some statistics: $\mu = n \cdot p$ $\quad \sigma = \sqrt{n \cdot p \cdot (1-p)}$

Normal approximation to binomial

As number of trials starts to grow, binomial tends to follow a normal distribution. This could be useful when computing binomial probabilities needs a lot of combinatorial processes. As a rule of thumb, a binomial can be approximated by a normal distribution if:

$$n \cdot p \geq 10$$
$$n \cdot (1-p) \geq 10$$

If those conditions holds, a binomial distribution can be approximated by a Normal distribution with $\mu = n \cdot p$ and $\sigma = \sqrt{n \cdot p \cdot (1-p)}$



Alberto Aguilera 23, E-28015 Madrid - Tel: +34 91 542 2800 - http://www.iit.comillas.edu