2. A Brief IntRoduction to Probability and Probability Distributions

Massimiliano Pastore (& Gianmarco Altoè)
University of Padova

Contents

Introduction

- In very general terms, probability could be viewed as a mesaure of uncertainty related to the occurrence of an event
- The theory of probability is the basis for inferential statistics (both frequentist and bayesian inferential statistics)
- So before moving to bayesian inference, let's recall the basic properties of probability

Note. For a more in depth introduction to the theory of probability see: Etz & Vandekerckhove (2017). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, 1-30
https://link.springer.com/article/10.3758/s13423-017-1262-3

Basics: Sample Space, Outcomes and Events

- Given a random phenomenon (e.g., a roll of a die), the set of all possible *outcomes* (results) is called **sample space** (S)
- The elements of S must be *mutually exclusive* and *collectively exhaustive*
- An *event* (E) is a subset of S
- Example:
  Random phenomenon: A roll of a six-sided dice
  Sample space: $S = \{1, 2, 3, 4, 5, 6\}$
  Event: "An even number will come up", $E = \{2, 4, 6\}$

Discrete and continuous sample spaces

- A sample space is called **discrete** if its elements are countable:

  EXAMPLE
  Phenomenon: Number of errors in a task
  Sample space: $S = \{0, 1, 2, 3, \ldots\}$

- A sample space is called **continuous** if its elements are not countable, but rather they represent a continuum of values

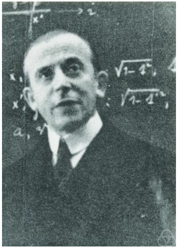
  EXAMPLE
  Phenomenon: Proportion of fixation time to a stimulus (θ)
  Sample space: $S = \{0 \leq \theta \leq 1\}$

Probability definitions

- The first formal approaches to probability theory were made in the second half of the 17th century (Pascal, Fermat, Bernoulli).
- The first areas of application for probability theory were in the fields of games of chance and insurance problems.
- One first attempt at rigorous formalization was made by Laplace (1812); other important developments were made by De Moivre (1667–1754), Gauss (1777–1855) and Poisson (1781–1840).
- From the second half of the 19th century until the 1920s, important contributions were made by Chebyshev (1821–1894), Markov (1856–1922), and Lyapunov (1857–1918).

Probability definitions

1. Classic (Laplace)
2. Frequentist (**Von Mises**, Reichebach, Castelnuovo)
3. Subjective (**De Finetti**, Ramsey, Savage)
4. Axiomatic (**Kolmogorov**)



Classical definition

The probability of an event E is the ratio of the number of cases favorable to its occurrence (s) to the total number of possible cases ($n = |S|$), assuming that all outcomes are equally probable.

Formally:

$$p(E) = \frac{s}{|S|}$$

where $|S|$ is the cardinality of the sample space S.

Limitations:

- It is not always possible to determine $|S|$, the total number of possible outcomes in the sample space, especially for infinite or very large sample spaces.
- The assumption that all possible outcomes are equally probable does not hold in many real-world scenarios.

Frequentist definition

The probability of an event E is defined as the value to which the relative frequency of occurrence of the event tends to converge as the number of trials approaches infinity

Formally:

$$p(E) = \lim_{n \to \infty} \frac{f(E)}{n}$$

where $f(E)$ is the number of times event E occurs in n trials.

Limitations:

- It is not possible to determine the exact probability, only an estimate based on observed frequencies.
- The reliability of the estimate depends on tests being performed under consistent conditions; if conditions vary, the estimate may not be reliable.

Subjective definition

The subjective definition of probability was developed to address the limitations of classical and frequentist approaches. According to this definition:

The probability of an event is the price that an individual considers fair to pay for a contract that pays:

- 1 unit of currency if the event occurs
- 0 if the event does not occur

This definition reflects the degree of belief an individual has in the occurrence of the event, based on their personal knowledge and judgment.

Assiomatic definition

The axiomatic definition of probability is not operational in nature. Unlike classical or frequentist definitions, it does not provide specific guidance on how to calculate probabilities. Instead, it establishes a formal mathematical framework that can accommodate both objectivist and subjectivist interpretations of probability.

The term 'axiomatic' refers to the process of axiomatization used in its development. This process involves:

- Identifying primitive (undefined) concepts in probability theory.
- Establishing a set of axioms or postulates based on these concepts. Deriving theorems and properties of probability from these axioms.

This approach provides a rigorous mathematical foundation for probability theory, allowing for consistent application across various fields and interpretations.

Summarizing

- **Probability theory** serves as a general system for representing plausibility. It applies both to enumerable events in the world (by counting the ways in which they can occur) and to theoretical constructs defined by **parameters** (McElreath, 2016).
- **Probability is a measure of the degree of uncertainty about the occurrence of an event.**

The probability function

The probability function, $Pr()$, is a mathematichal function that assigns numbers to each outcome of a sample space.
The numbers, called probabilities, need to satisfy three properties (Kolmogorov, 1956):

1. A probability value must be non-negative (i.e., zero or positive)
2. The sum of the probabilities across all outcomes in the entire sample space must be 1
3. For any two mutually exclusive events, the probability that one or the other occurs is the sum of their individual probabilities

Probability distributions

Probability distribution

A probability distribution is a list of all outcomes of the sample space and their corresponding probabilities

Example

- **Phenomenon**: X = a single toss of a fair coin; where X is a *random variable* that can take the following values: $x = 0$ (Head) e $x = 1$ (Tail)
- **Sample space**: $S = \{0, 1\}$
- **Probability distribution**: $Pr(X = x) = \theta^x(1 - \theta)^{1-x}$ where θ (i.e., the parameter that defines the probability distribution) is equal to .5
- **Properties of the probability distribution**: $Pr(X = 0) = .5$; $Pr(X = 1) = .5 \geq 0$ $Pr(X = 0) + Pr(X = 1) = 1$

Discrete and continuous probability distributions

- A probability distribution associated with a *discrete sample space* is called **discrete probability distribution** or **mass probability function**
- A probability distribution associated with a *continuous sample space* is called **continuous probability distribution** or **density probability function**

Properties of discrete probability distributions

Let $Pr(X = x)$ be a discrete probability distribution with sample space $S = \{X\}$, then:

$$0 \leq Pr(X = x) \leq 1 \quad \forall x \tag{1}$$

$$\sum_{X} Pr(X = x) = 1 \tag{2}$$

Some discrete probability distributions

Some discrete probability distributions

Name	S	Param.	Mass	Mean	Variance
Bernoulli	$x = 0, 1$	θ	$\theta^x (1-\theta)^{1-x}$	θ	$\theta(1-\theta)$
Binomial	$x = 0, 1, \dots, n$	θ	$\left(\begin{array}{c} n \\ x \end{array} \right) \theta^x (1-\theta)^{n-x}$	$n\theta$	$n\theta(1-\theta)$
Poisson	$x = 0, 1, 2, \dots$	λ	$e^{-\lambda} \frac{\lambda^n}{n!}$	λ	λ

Properties of continuous probability distributions

Let $Pr(X = x)$ be a continuous probability distribution with sample space $S = \{X\}$, then:

$$p(X = x) \geq 0 \quad \forall x \tag{3}$$

$$\int_X p(X = x)dx = 1 \tag{4}$$

Properties of continuous probability distributions

Let $Pr(X = x)$ be a continuous probability distribution with sample space $S = \{X\}$, then:

$$p(X = x) \geq 0 \quad \forall x \tag{3}$$

$$\int_X p(X = x)dx = 1 \tag{4}$$

Note. Probability densities can be greater than 1, whereas probability masses cannot be greater than 1. In any case, the overall area under the probability function (i.e. the overall probability associated with the sample space) is always 1.
. . . try

```
> curve( dnorm(x,mean=0,sd=.1), from=-.5, to=.5 )
```

Some continuous probability distributions

Some continuous probability distributions

Name	S	Param.	Density	Mean	Variance
Uniform	$a \leq x \leq b$	a, b	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$-\infty \leq x \leq +\infty$	μ, σ	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$	μ	σ^2
Beta	$0 \leq x \leq 1$	a, b	$\frac{x^{(a-1)}(1-x)^{(b-1)}}{B(a,b)}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$

Probability distributions and R

- R makes it easy to work with probability distributions
- There are four functions available for probability distributions:
    - `d*`: probability mass for discrete distribution and probability density for continuous distributions
    - `p*`: the cumulative probability for a given quantile
    - `q*`: the quantile for a given cumulative probability
    - `r*`: draw a random sample for a given distribution

Note. * distribution names in R (e.g., Normal: `norm`)

Some probability distributions in R

Distribution	Type	Name
Binomial	discrete	`@binom`
Poisson	discrete	`@pois`
Uniform	continuous	`@unif`
Normal	continuous	`@norm`
Student's t	continuous	`@t`
Beta	continuous	`@beta`

Note: `@` = prefix: `d` or `p` or `q` or `r`
For example `rnorm()` draw random samples from a Normal distribution

. . . from theory to practice

Let's practice with some probability distributions:

- Poisson
- Normal

Poisson distribution

Suppose that the number of errors committed in a cognitive task has a Poisson distribution with parameter $\lambda = 3$:

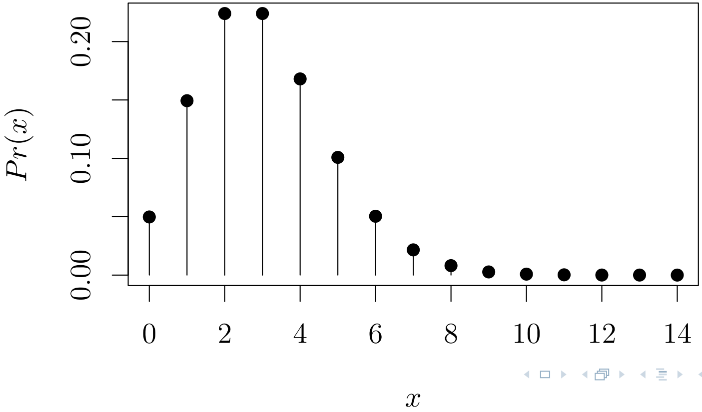
$$Pr(X = x) \sim P(x; \lambda), \quad x = 0, 1, 2, 3, 4, 5, \ldots$$

Graphical representation

```
> x <- 0:14
> plot( x, dpois( x, lambda = 3), type = "h", ylab = "Pr(x)" )
> points( x, dpois( x, lambda = 3 ), pch = 19 )
```

Graphical representation

```
> x <- 0:14
> plot( x, dpois( x, lambda = 3), type = "h", ylab = "Pr(x)" )
> points( x, dpois( x, lambda = 3 ), pch = 19 )
```



Exercises

Exercises

- What is the probability that a subject makes 4 errors?

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

- What is the probability that a subject makes less than 2 errors?

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

- What is the probability that a subject makes less than 2 errors?

```
> ppois( 1, 3 )
[1] 0.1991483
```

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

- What is the probability that a subject makes less than 2 errors?

```
> ppois( 1, 3 )
[1] 0.1991483
```

- What is the probability that a subject makes more than 2 errors?

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

- What is the probability that a subject makes less than 2 errors?

```
> ppois( 1, 3 )
[1] 0.1991483
```

- What is the probability that a subject makes more than 2 errors?

```
> 1 - ppois( 2, 3 )
[1] 0.5768099
```

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

- What is the probability that a subject makes less than 2 errors?

```
> ppois( 1, 3 )
[1] 0.1991483
```

- What is the probability that a subject makes more than 2 errors?

```
> 1 - ppois( 2, 3 )
[1] 0.5768099
```

- What is the probability that a subject makes at least 1 error?

Exercises

- What is the probability that a subject makes 4 errors?

```
> dpois( 4, 3 )
[1] 0.1680314
```

- What is the probability that a subject makes 3 or 4 errors?

```
> dpois( 3, 3 ) + dpois( 4, 3 )
[1] 0.3920732
```

- What is the probability that a subject makes less than 2 errors?

```
> ppois( 1, 3 )
[1] 0.1991483
```

- What is the probability that a subject makes more than 2 errors?

```
> 1 - ppois( 2, 3 )
[1] 0.5768099
```

- What is the probability that a subject makes at least 1 error?

```
> 1 - ppois( 0, 3 )
[1] 0.9502129
```

Drawing random samples

A random samples of 10 performance

Drawing random samples

A random samples of 10 performance

```
> rpois( 10, 3 )

 [1] 3 2 5 4 2 6 5 4 5 3
```

Drawing random samples

A random samples of 10 performance

```
> rpois( 10, 3 )

 [1] 3 2 5 4 2 6 5 4 5 3
```

A random samples of 1000000 performance

Drawing random samples

A random samples of 10 performance

```
> rpois( 10, 3 )

 [1] 3 2 5 4 2 6 5 4 5 3
```

A random samples of 1000000 performance

```
> large_sample <- rpois( 1e6, 3 )
```

Drawing random samples

A random samples of 10 performance

```
> rpois( 10, 3 )

 [1] 3 2 5 4 2 6 5 4 5 3
```

A random samples of 1000000 performance

```
> large_sample <- rpois( 1e6, 3 )
```

Sample mean and sample variance

Drawing random samples

A random samples of 10 performance

```
> rpois( 10, 3 )

 [1] 3 2 5 4 2 6 5 4 5 3
```

A random samples of 1000000 performance

```
> large_sample <- rpois( 1e6, 3 )
```

Sample mean and sample variance

```
> mean( large_sample )

[1] 3.000185

> var( large_sample )

[1] 3.002194
```

Normal distribution

Suppose that the time (in seconds) spent on by a guinea pig to complete a maze is normally distributed with mean $\mu = 180$ e standard deviation $\sigma = 40$:

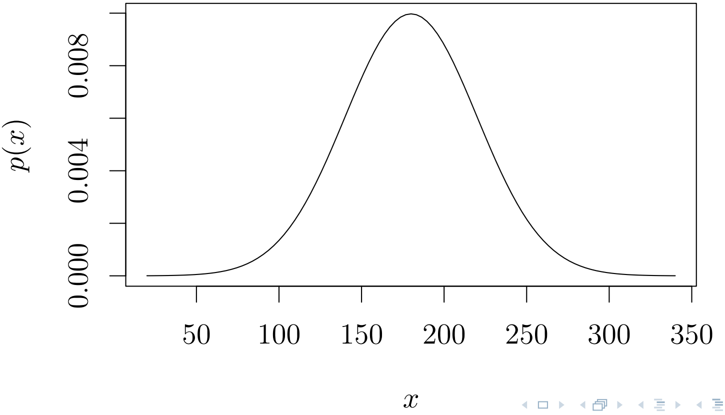
$$p(X = x) \sim \mathcal{N}(x; 180, 40)$$

Graphical representation

```
> curve( dnorm( x, 180, 40 ), from = 180-40*4, to = 180+40*4,
+        ylab = "p(x)" )
```

Graphical representation

```
> curve( dnorm( x, 180, 40 ), from = 180-40*4, to = 180+40*4,
+         ylab = "p(x)" )
```



Exercises

1. What is the probability that a randomly selected guinea pig spends less than 200 seconds to complete the maze?
2. What is the probability that a randomly selected guinea pig spends between 140 and 220 seconds to complete the maze?
3. Graphically represent 3 normal distributions with a fixed mean of 180 and a standard deviation of 20, 40, 80 respectively.
   Hint: To add different graphs on the same plot use the option `add = TRUE` (see, `?curve`)

Bivariate probability distributions

Bivariate probability distributions

- There are many situations in which we are interested in the conjunction of two (or more) outcomes regarding two(or more) random variables
- For example:
  - What is the probability of meeting a person with both red hair and green eyes?
  - What is the probability that "Linda is a bank teller and is active in the feminist movement?" (Kahneman e Tversky, 1974)
- **Bivariate** (or **multivariate**) probability distributions are used to answer these kinds of questions

Tossing a fair coin

As an example for developing our new ideas, imagine tossing a fair coin three times in a row:

Sample_space	Probability
HHH	1/8
HHT	1/8
HTH	1/8
HTT	1/8
THH	1/8
THT	1/8
TTH	1/8
TTT	1/8

Note. T = Tail, H = Head

Let's now consider two random variables

- X = number of heads, where:
  $S_X = \{0, 1, 2, 3\}$
- Y = Number of switches between heads and tails, where:
  $S_X = \{0, 1, 2\}$

Table of the bivariate probability distribution

The bivariate (or conjoint) probability distribution of X and Y:

$$Pr(X = x, Y = y) = Pr(x, y)$$

	X (nr. of heads)			
Y (nr. of switches)	0	1	2	3
0				
1				
2				

The probability of two events happening together is called their conjoint probability. Example: $Pr(X = 1, Y = 2) = 1/8 = .125$

Table of the bivariate probability distribution

The bivariate (or conjoint) probability distribution of X and Y:

$$Pr(X = x, Y = y) = Pr(x, y)$$

	X (nr. of heads)			
Y (nr. of switches)	0	1	2	3
0	1/8	0	0	1/8
1	0	2/8	2/8	0
2	0	1/8	1/8	0

The probability of two events happening together is called their conjoint probability. Example: $Pr(X = 1, Y = 2) = 1/8 = .125$

Marginal probability distribution

Marginal probability distributions give us the probability of obtaining one variable outcome regardless of the value of the other variable(s).

Marginal probability distribution

Marginal probability distributions give us the probability of obtaining one variable outcome regardless of the value of the other variable(s).

	X (nr. of heads)				
Y (nr. of switches)	0	1	2	3	$X_.$
0	1/8	0	0	1/8	2/8
1	0	2/8	2/8	0	4/8
2	0	1/8	1/8	0	2/8
$Y_.$	1/8	3/8	3/8	1/8	

Discrete and continuous marginal distributions

- To compute the mariginal probability distribution of X, we sum $Pr(x,y)$ across all values of y
- When the X and Y variables are discrete, we will have:

$$Pr(x) = \sum_y Pr(x,y) \qquad \text{(probability mass)}$$

- When the X and Y variables are continuous, we will have:

$$p(x) = \int_y p(x,y)dy, \qquad \text{(probability density)}$$

Note. The process described above is called marginalizing over Y or integrating out the variable Y.

Conditional probability

- We often want to know the probability of one event, given that we know another event is true
  - What is the probability that you will pass the statistics exam given that you scored 20/30 on the first assignment?
  - What is the probability of the observed data given that the *Null Hypothesis* is true? What is the probability of the *Null Hypothesis* being true given the observed data?!
- Conditional probabilities are used to answer these kinds of questions

Conditional probabilities: The idea

- Let's go back to our example of three coin flips ...
- What is the probability that a sequence of three coin flips has 1 switch ($Y = 1$) given that it has 1 head ($X = 1$)?

Conditional probabilities: The idea

- Let's go back to our example of three coin flips ...
- What is the probability that a sequence of three coin flips has 1 switch ($Y = 1$) given that it has 1 head ($X = 1$)?

$$Pr(Y = 1|X = 1) =?$$

(where "|" is read as: "given that")

- Intuitively, we will must consider only the cases in which a single Head occurs. And then we will have to compare the probability of obtaining a Switch to the overall probability of having a single Head

Conditional probabilities: The calculation

$$Pr(Y = 1|X = 1) = \frac{Pr(Y{=}1,X{=}1)}{Pr(Y{=}0,X{=}1)+Pr(Y{=}1,X{=}1)+Pr(Y{=}2,X{=}1)}$$

Conditional probabilities: The calculation

$$
\begin{aligned}
Pr(Y = 1|X = 1) &= \frac{Pr(Y{=}1,X{=}1)}{Pr(Y{=}0,X{=}1)+Pr(Y{=}1,X{=}1)+Pr(Y{=}2,X{=}1)} \\
&= \frac{2/8}{0+2/8+1/8} \\
&= \frac{2/8}{3/8} \\
&= \frac{2}{3} = 0.667
\end{aligned}
$$

Conditional probabilities: The calculation

$$\begin{aligned} Pr(Y=1|X=1) &= \frac{Pr(Y{=}1,X{=}1)}{Pr(Y{=}0,X{=}1)+Pr(Y{=}1,X{=}1)+Pr(Y{=}2,X{=}1)} \\ &= \frac{2/8}{0+2/8+1/8} \\ &= \frac{2/8}{3/8} \\ &= \frac{2}{3} = 0.667 \end{aligned}$$

Question: $Pr(X=1|Y=1) = Pr(Y=1|X=1)$?

Conditional probabilities: The Formalization

- When the X and Y variables are discrete, we will have:

$$Pr(y|x) = \frac{Pr(y,x)}{\sum_y Pr(y,x)} = \frac{Pr(y,x)}{Pr(x)}$$

- When the X and Y variables are continuous, we will have:

$$p(y|x) = \frac{p(y,x)}{\int_y p(y,x)dy} = \frac{p(y,x)}{p(x)}$$

Independent probability distributions

Two probability distribution, $p(x)$ and $p(y)$, are said to be independent if and only if:

$$p(y|x) = p(y), \text{ for each } y \text{ and } x$$

and equivalently

$$p(x|y) = p(x), \text{ for each } x \text{ and } y$$

Fundamental property of independent distributions

If and only if, the probability distribution of X and Y are indipendent, we will have:

$$p(x, y) = p(x)p(y), \text{ for each } y \text{ e } x$$

Fundamental property of independent distributions

If and only if, the probability distribution of X and Y are indipendent, we will have:

$$p(x, y) = p(x)p(y), \text{ for each } y \text{ e } x$$

Exercises:

1. Demonstrate this simple property
2. In our example of the three coin flips, are X e Y indipendent? Prove it.

Before bayesian inference . . . remember that

- **In general:** $Pr(x|y) \neq Pr(y|x)$
  If someone smiles at you, what is the probability that they love you?
  If someone loves you, what is the probability that they will smile at you?
- **There is no temporal order in conditional probabilities**
  When we say "the probability of x given y" we do not mean that y has already happened and x has yet to happen. All we mean is that we are restricting our calculations of probability to a particular subset of possible events.

Bonus: Going Bayes

- We know that:

$$Pr(y|x) = \frac{Pr(y,x)}{Pr(x)}, \quad Pr(x|y) = \frac{Pr(y,x)}{Pr(y)}$$

Bonus: Going Bayes

- We know that:

$$Pr(y|x) = \frac{Pr(y,x)}{Pr(x)}, \quad Pr(x|y) = \frac{Pr(y,x)}{Pr(y)}$$

- Obtain $Pr(y,x)$ from both equations and equalize the results:

$$Pr(y|x)Pr(x) = Pr(x|y)Pr(y)$$

Bonus: Going Bayes

- We know that:

$$Pr(y|x) = \frac{Pr(y,x)}{Pr(x)}, \quad Pr(x|y) = \frac{Pr(y,x)}{Pr(y)}$$

- Obtain $Pr(y, x)$ from both equations and equalize the results:

$$Pr(y|x)Pr(x) = Pr(x|y)Pr(y)$$

- Divide both sides of the equation by $Pr(x)$:

$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{Pr(x)}$$

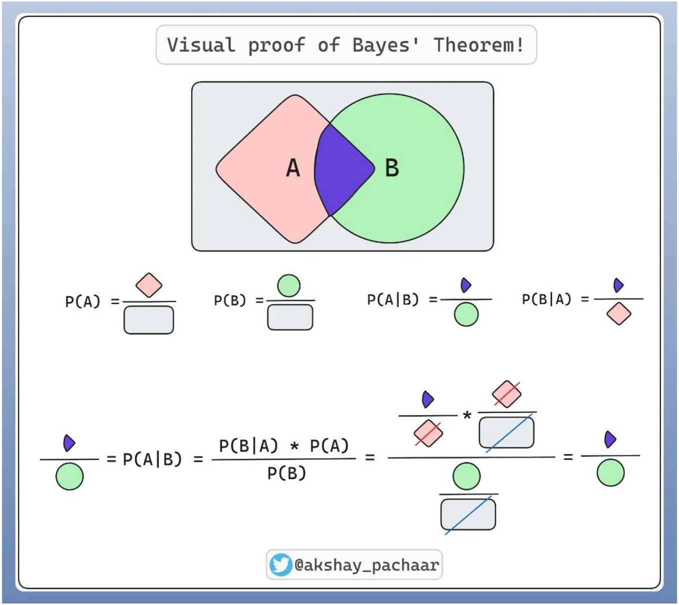
Bonus: Going Bayes

- Now, remembering that $Pr(x) = \sum_y Pr(x, y)$, and knowing that $Pr(x, y) = Pr(x|y)Pr(y)$

Bonus: Going Bayes

- Now, remembering that $Pr(x) = \sum_y Pr(x, y)$, and knowing that $Pr(x, y) = Pr(x|y)Pr(y)$ … here we have the **Bayes Theorem**:

$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{\sum_y Pr(x, y)}$$


Visual proof of Bayes' Theorem!
A
B
P(A) =
P(B) =
P(A|B) =
P(B|A) =
= P(A|B) =
P(B|A) * P(A)
P(B)
=
*
=
=
@akshay_pachaar

Used R packages

- `gtools`. Warnes G, Bolker B, Lumley T, Magnusson A, Venables B, Ryodan G, Moeller S (2023).
- `knitr`. Xie Y (2025).
- `R`. R Core Team (2025).
- `report`. Makowski D, Lüdecke D, Patil I, Thériault R, Ben-Shachar M, Wiernik B (2023).
- `xtable`. Dahl D, Scott D, Roosen C, Magnusson A, Swinton J (2019).

massimiliano.pastore@unipd.it
https://psicostat.dpss.psy.unipd.it/