

# PLSC30500, Fall 2023

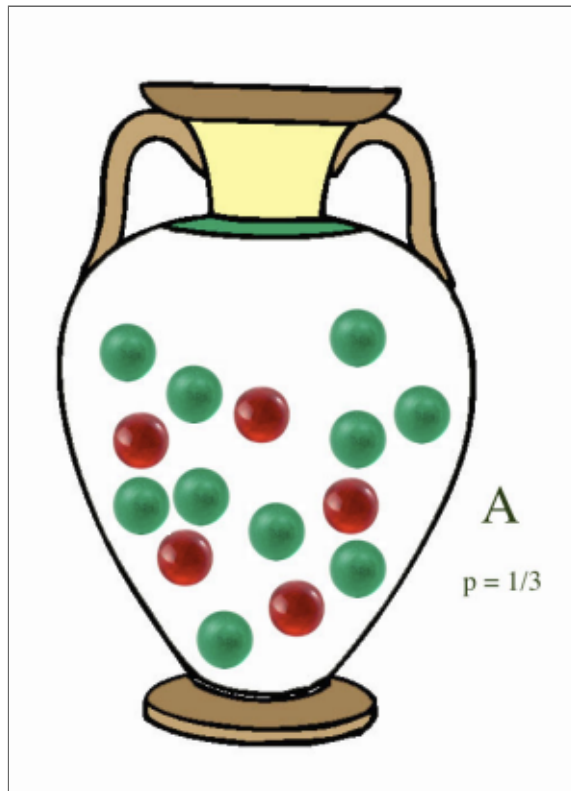
Part 1. Probability Theory (part a)

Andy Eggers

**Motivation & big picture**

## Coin flips and urn problems

You will see a lot of problems about coin flips and selecting balls from urns.  
What does this have to do with social science?



# Sampling

Sometimes a coin flip, an urn, or a similar device actually determines which **units/observations** we see: who gets selected for a survey.

The urn problems help us understand how the **sample** might differ from the **population** (and thus how certain we can be about characteristics of the population using the sample).

## Treatment assignment

Sometimes a coin flip, an urn, or a similar device actually determines which units/observations get a random **treatment**, e.g. in a **randomized experiment** or the Vietnam draft lottery.

The urn problems help us compare differences we see between treatment and control units to differences we might see by chance if the treatment had no effect.

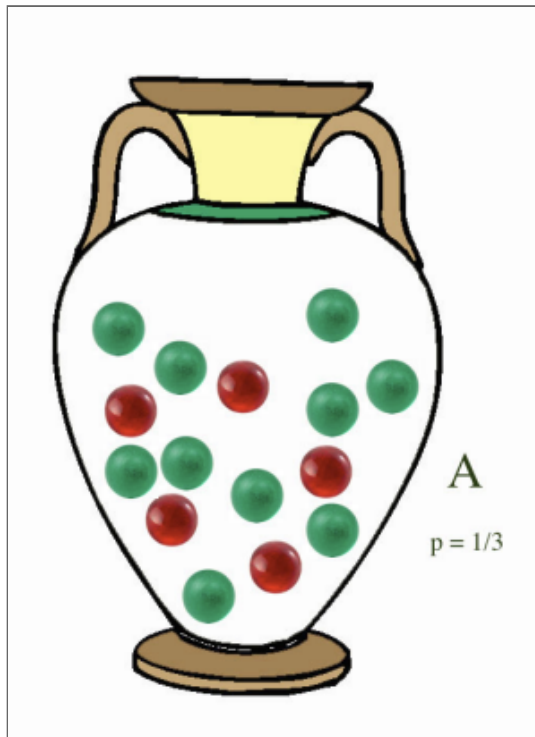
## Urns as metaphors

Even when there was no random selection (e.g. data on all countries) we can *act as if there was*, or act as if the dependent variable (e.g. revolution) has a random component.

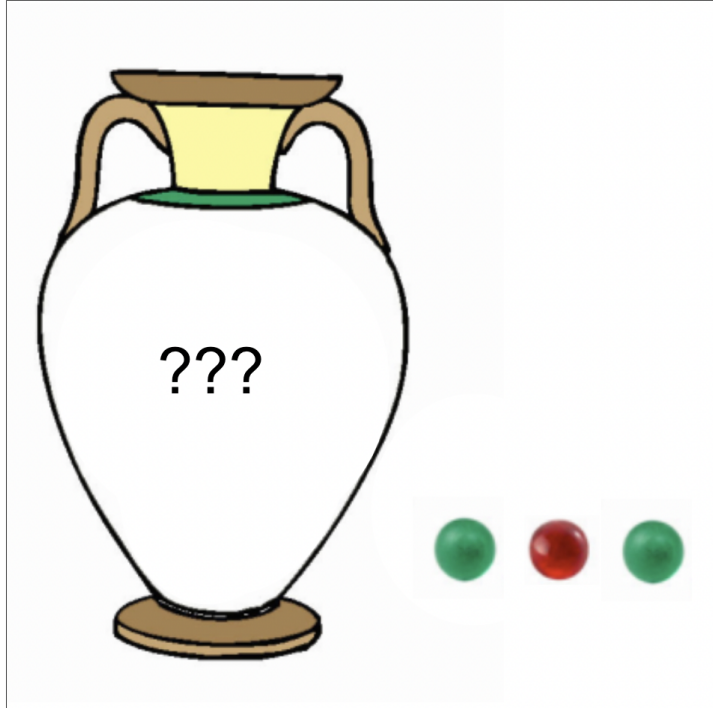
Then the urn problems again help us compare the “sample” to the “population”, or observed reality to what might have happened in an alternate history *if we treat our ignorance as chance*.

## Probability vs statistical inference

In probability problems, we know what's in the urn and we want to describe the possible **draws**.



In many statistics problems, we have one draw and we want to speculate what might be in the urn (i.e. population).





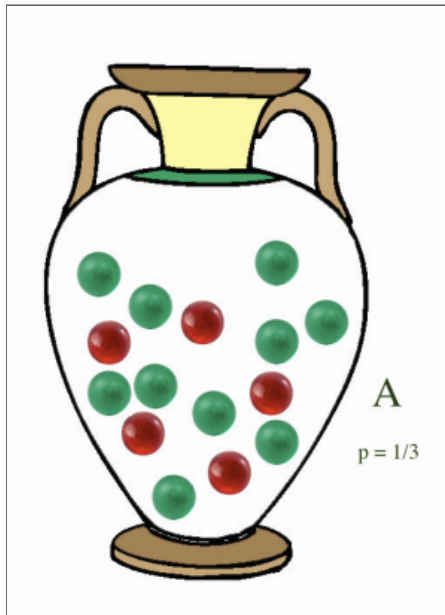
# Probability foundations

# What is probability?

A **random generative process** is a repeatable mechanism that can select an outcome from a set of possible outcomes.

Each **draw** or **realization** of the process may be uncertain (to the typical observer), but the frequency of each **event** can be described.

e.g. flipping a coin, rolling a die, drawing a ball from an urn.



**Frequentist definition of probability:** The **probability** of an event (e.g. “green ball is chosen”) is the proportion of many, many draws producing that event.

**Bayesian definition of probability:** The **probability** of an event is an observer’s degree of belief that the event will happen or has happened. Logical and subjective variants.

## Aside on mathematical notation

# Glossary of Mathematical Notation

In this appendix, we provide a glossary of mathematical notation not defined in the text.

Notation	Definition and Usage
$\in$	<i>Set membership.</i> $s \in A$ (read “ $s$ is an element of $A$ ” or “ $s$ is in $A$ ”) denotes that the object $s$ (which could be a number, a vector, a set, etc.) is an element of the set $A$ . For example, $2 \in \{1, 2, 4\}$ . The negation of $\in$ is denoted by $\notin$ ; $s \notin A$ (read “ $s$ is not an element of $A$ ” or “ $s$ is not in $A$ ”) denotes that the object $s$ is <i>not</i> an element of the set $A$ . For example, $3 \notin \{1, 2, 4\}$ .
$\forall$	<i>For all.</i> Used to state that all elements of a particular set satisfy a given expression. For example, $\forall x \in \{1, 2, 3, 4, 5\}, 2x \leq 10$ . The set may be omitted if it is clear from context. For example, $\forall x, x + 1 > x$ might mean that $x + 1 > x$ is true for any real number $x$ . Similarly, $\forall k > 1, f(k) = 0$ might mean that $f(k) = 0$ is

## Sample space

Sample space  $\Omega$  (“Omega”) is the set of all possible outcomes of the random generative process. Each element  $\omega$  (“omega”) is a unique outcome of the process.

For a coin flip,  $\Omega = \{H, T\}; \omega \in \{H, T\}$ .

For a single roll of a six-sided die,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

How about for a single roll of two six-sided die?

$$\Omega = \{(x, y) \in \mathbb{Z}^2 : 1 \leq x \leq 6, 1 \leq y \leq 6\}$$

(“Set-builder notation”, used w/o explanation at A&M p. 5)

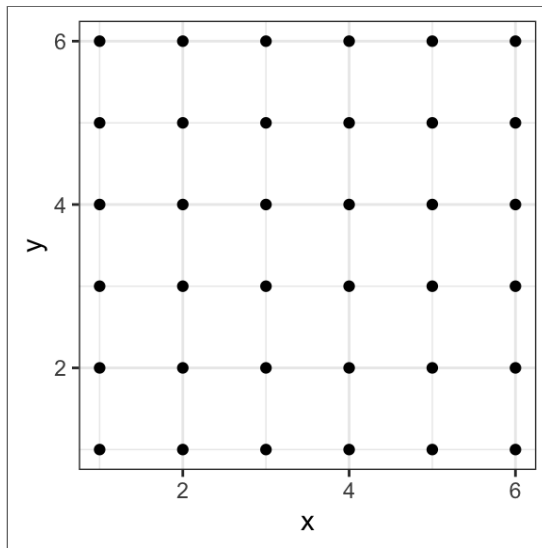
Easier examples:

- $\{x \in \mathbb{R} : x > 0\}$
- $\{x \in \mathbb{Z} : 2 \leq x \leq 5\}$

## Sample space (2)

$$\Omega = \{(x, y) \in \mathbb{Z}^2 : 1 \leq x \leq 6, 1 \leq y \leq 6\}$$

i.e.,  $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (6, 6)\}$



## Events and event spaces

An **event** is a collection of outcomes to which we want to assign a probability. (A subset of the sample space.)

Examples:

- rolling a 3 (Here, an outcome is an event)
- rolling an even number
- in an election, a tie for first between candidates  $a$  and  $b$

An **event space**  $S$  is a set of events composed in a particular way (for technical reasons):

- a set of events of interest (e.g.  $A = a$  wins,  $B = b$  wins,  $T = a$  and  $b$  tie)
- their complements ( $A^C = a$  doesn't win,  $B^C = b$  doesn't win,  $T^C = a$  and  $b$  don't tie)
- the **union** of each subset of events (e.g.  $A \cup B = a$  wins OR  $b$  wins,  $A \cup T = a$  wins or ties, ...)



## Probability measure & Kolmogorov axioms

A **probability measure** is a function  $P : S \rightarrow \mathbb{R}$  that assigns a probability to every event in the event space.

**Kolmogorov axioms:**  $(\Omega, S, P)$  is a **probability space** if it satisfies the following:

- Non-negativity:  $\forall A \in S, P(A) \geq 0$  (probabilities are positive)
- Unitarity:  $P(\Omega) = 1$  (probability that something happens is 1)
- Countable additivity: if  $A_1, A_2, A_3, \dots \in S$  are **pairwise disjoint**, then

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots = \sum_i P(A_i)$$

(for events that cannot co-occur, the probability of one of the event occurring is the sum of the individual probs)

## Basic properties of probability

Let  $(\Omega, S, P)$  be a probability space. Then

- Monotonicity:  $\forall A, B \in S$ , if  $A \subseteq B$ , then  $P(A) \leq P(B)$
- Subtraction rule:  $\forall A, B \in S$ , if  $A \subseteq B$ , then  $P(B \setminus A) = P(B) - P(A)$
- Zero probability of the empty set:  $P(\emptyset) = 0$
- Probability bounds:  $\forall A \in S, 0 \leq P(A) \leq 1$
- Complement rule:  $\forall A \in S, P(A^C) = 1 - P(A)$

Let's prove it! (on board, also see A&M page 8, or slide notes by pressing [s](#))

## An aside on what we're doing

Goal: a strong *system of understanding*; every statement follows from

- definitions (what we mean by  $X$ )
- axioms (assertions taken to be self-evident requirements)
- assumptions (assertions that usefully restrict/simplify)
- logical argument following from above

Every statement supported, no unnecessary assumptions.

There are “rules” (e.g. addition rule) and “laws” (e.g. law of total probability) but no one made them (directly).

## Joint probability and the addition rule

**Def 1.1.5:** For  $A, B \in S$ , the **joint probability** of  $A$  and  $B$  is the probability that both  $A$  and  $B$  happen, i.e.  $P(A \cap B)$

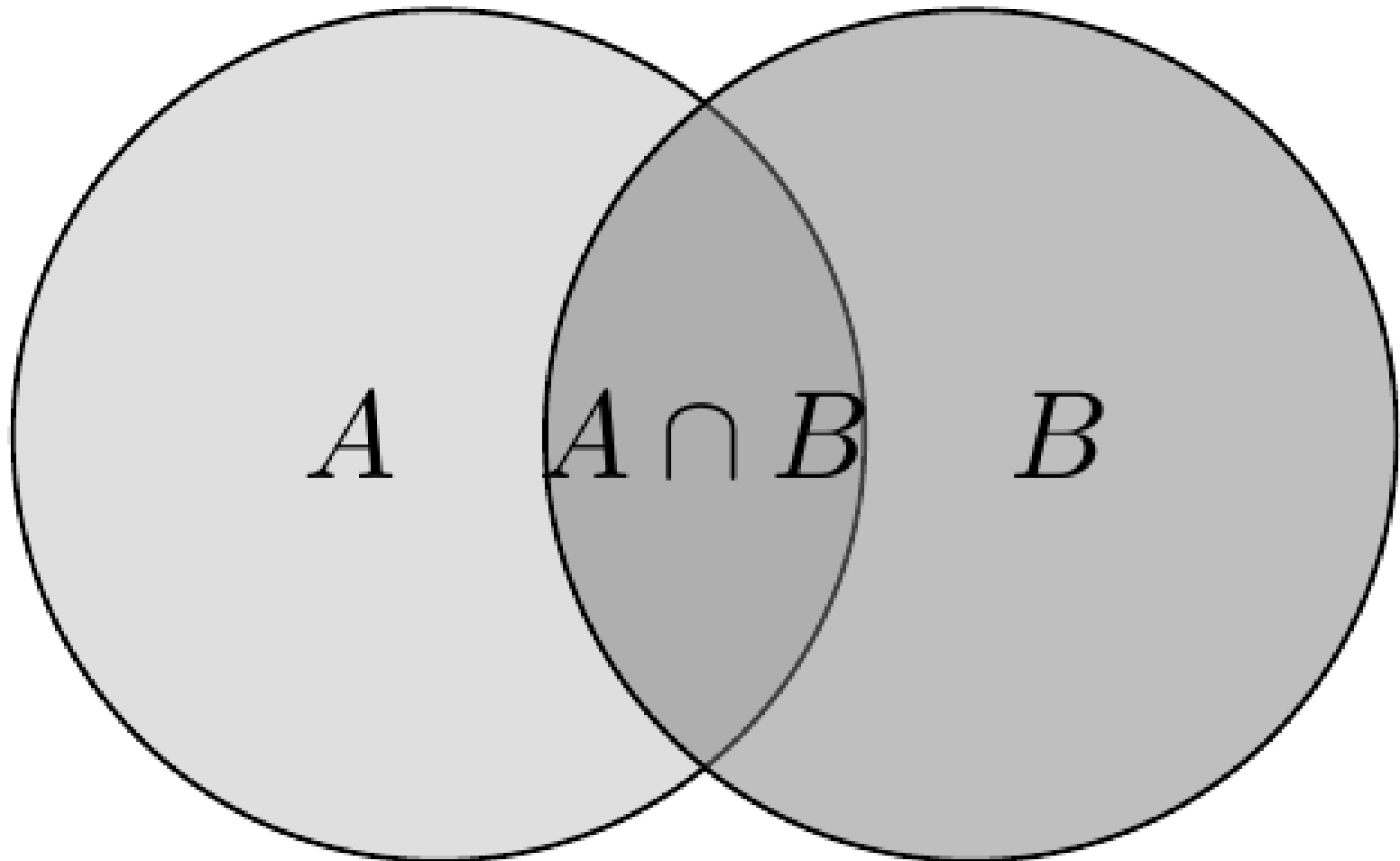
**Addition rule:** For  $A, B \in S$ ,

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

Let's prove it! (Board, or book or **S**)

Addition rule, visually

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



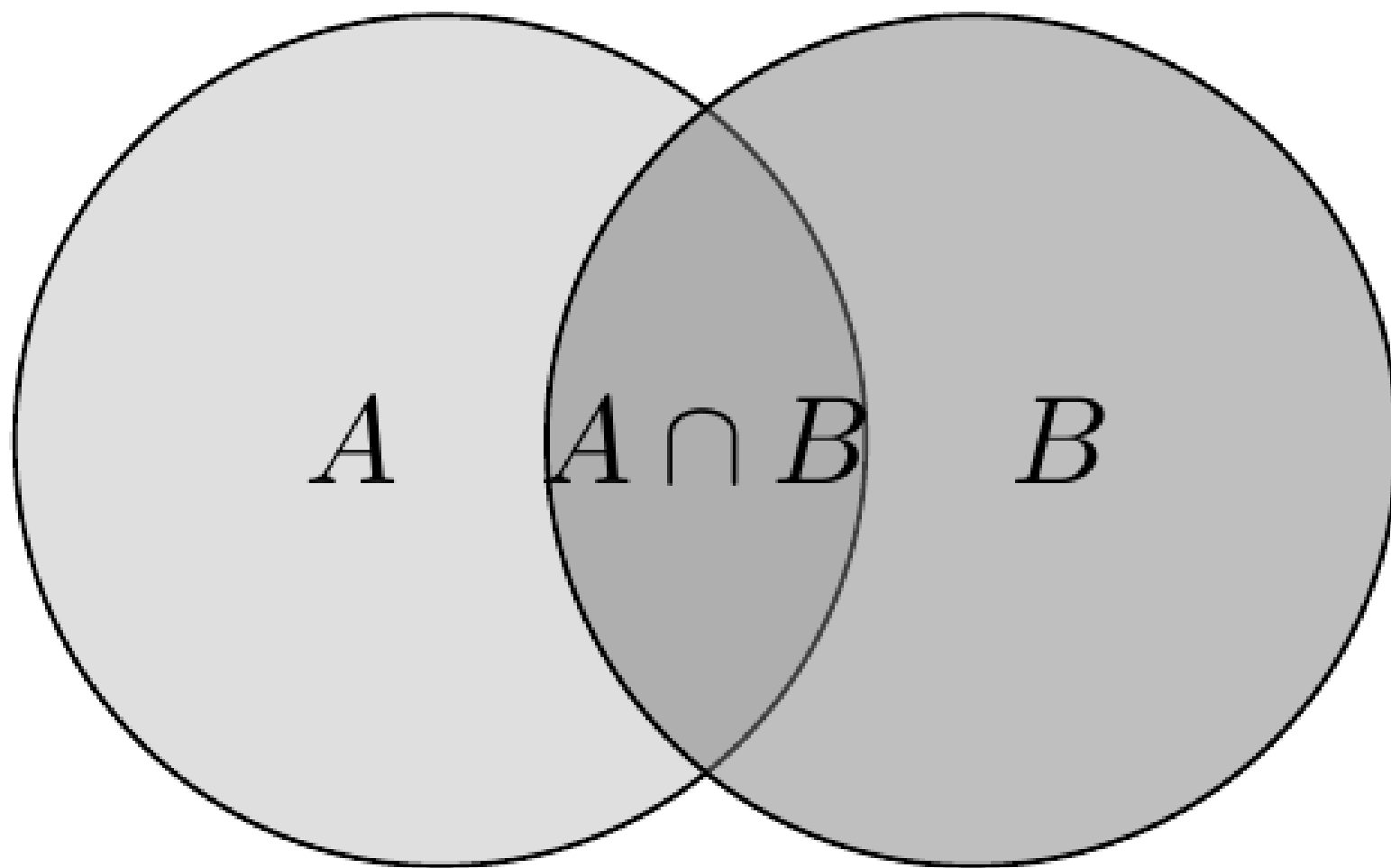
Assumptions behind this use of Venn diagram:

- think of each pixel in the rectangle as an outcome  $\omega$  in the sample space  $\Omega$
- each outcome has equal probability
- $A$  and  $B$  are events (sets of outcomes)

## Conditional probability

**Def 1.1.8:** For  $A, B$  with  $P(B) > 0$ , the **conditional probability** of  $A$  given  $B$  is

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

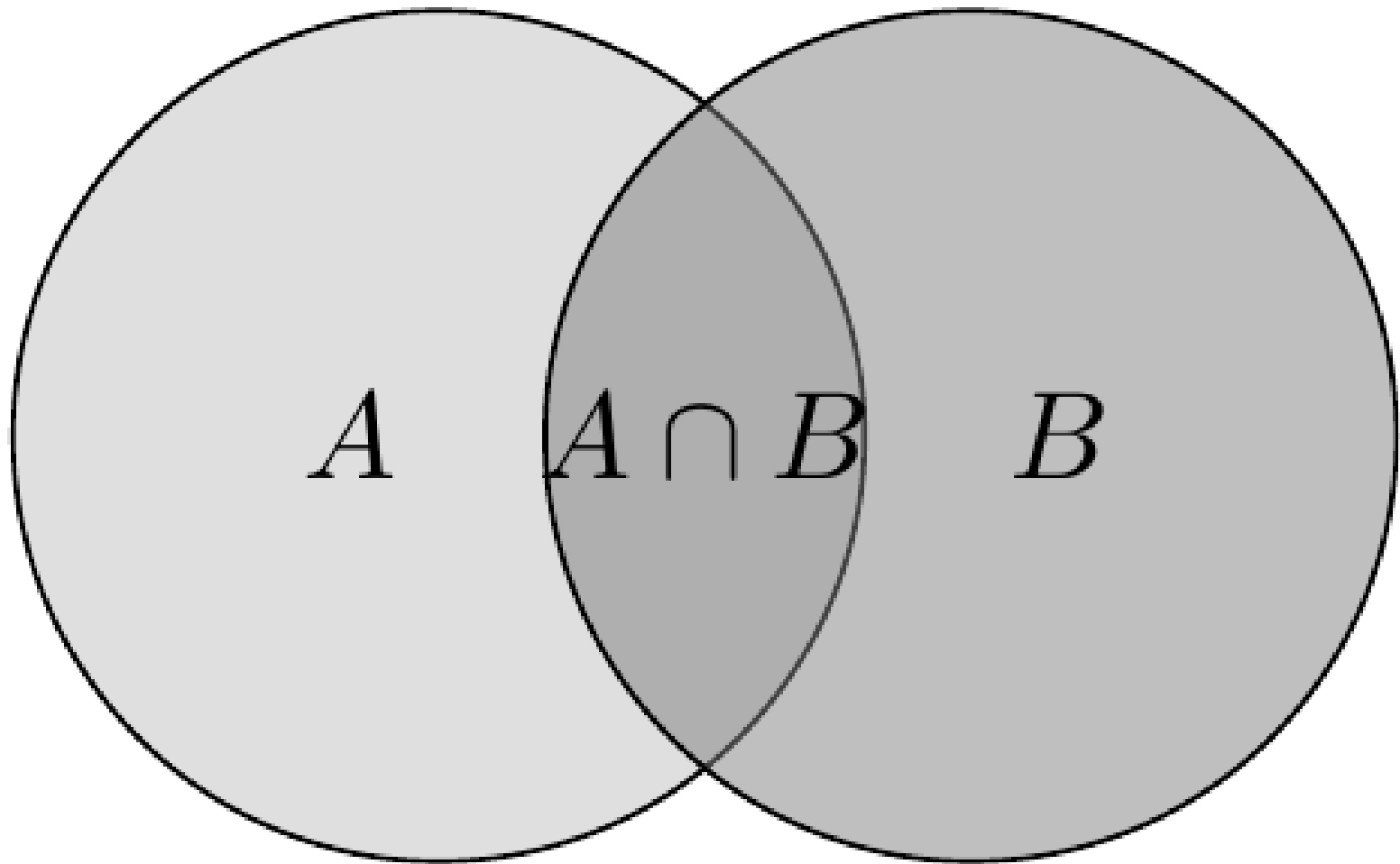


Read  $P(A \mid B)$  as “probability of  $A$  given  $B$ ”.



## Equivalently: product rule

For  $A, B$  with  $P(B) > 0$ ,  
 $P(A \cap B) = P(B)P(A|B)$



Here, product rule follows from definition of conditional probability; in logical approach (Cox's Theorem) it follows from consistency axioms.

## Partition

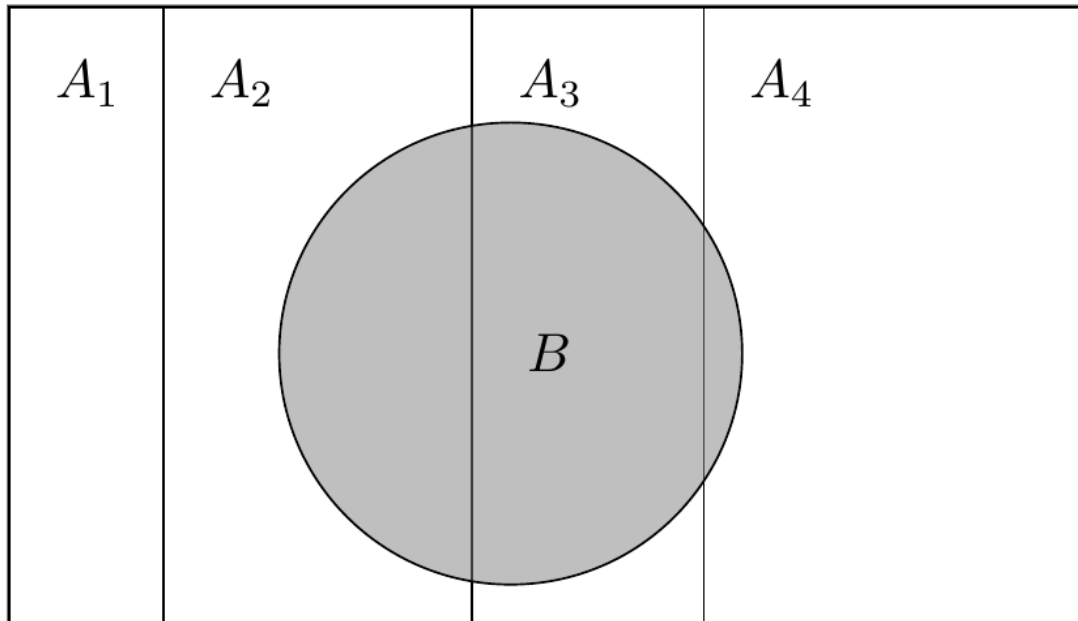
If  $A_1, A_2, \dots \in \mathcal{S}$  is nonempty and pairwise disjoint, and  $\Omega = A_1 \cup A_2 \cup \dots$ , then  $A_1, A_2, \dots$  is a **partition** of  $\Omega$ .

$A_1$	$A_2$	$A_3$	$A_4$
-------	-------	-------	-------

## Law of total probability

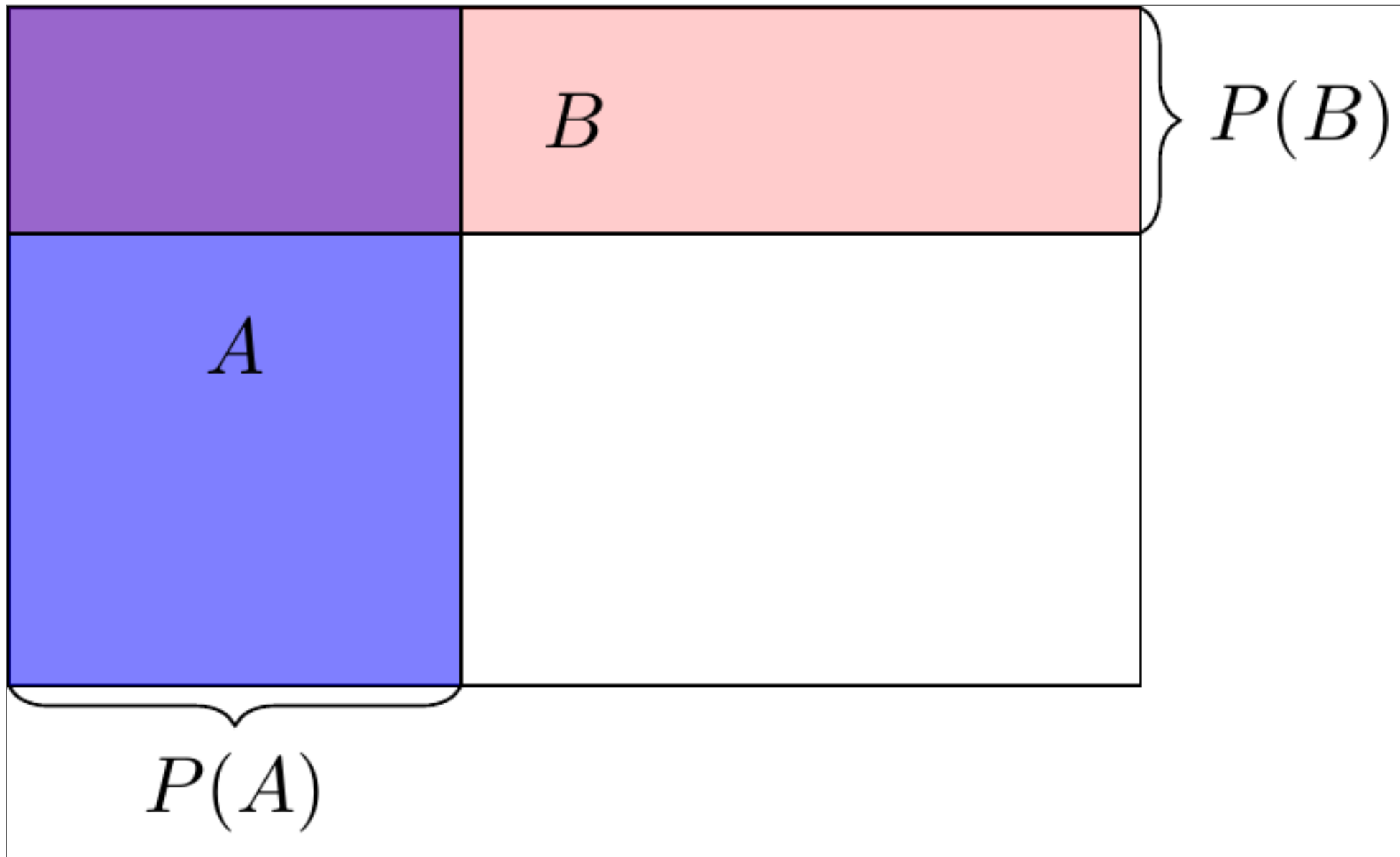
If  $\{A_1, A_2, \dots\}$  is a partition of  $\Omega$  and  $P(A_i) > 0 \forall i$ , then

$$P(B) = \sum_i P(B \cap A_i) = \sum_i P(B | A_i)P(A_i)$$



## Independence of events

**Definition:** Events  $A, B \in \mathcal{S}$  are independent if  $P(A \cap B) = P(A)P(B)$



Informally, knowing  $A$  occurs does not tell you anything about whether  $B$  occurs.

## Independence and conditional probability

**Theorem 1.1.16** For  $A, B \in \mathcal{S}$  with  $P(B) > 0$ ,  $A$  and  $B$  are independent (i.e.  $A \perp B$ ) if and only if  $P(A \mid B) = P(A)$ .

**Proof:**

$$A \perp B \iff P(A \cap B) = P(A)P(B) \text{ (definition)}$$

$$A \perp B \iff P(A \mid B)P(B) = P(A)P(B) \text{ (product rule)}$$

$$A \perp B \iff P(A \mid B) = P(A). \blacksquare \text{ (divide by } P(B))$$

In subjective terms, knowing  $B$  occurred does not affect our assessment of probability that  $A$  occurred.

## Independent?

(Recall: events  $A$  and  $B$  independent means  $P(A \cap B) = P(A)P(B)$  and  $P(A|B) = P(A)$ )

Pick a student at random from the university. Are  $A$  and  $B$  independent in the following examples?

- $A$  = student is undergraduate,  $B$  = student is under 22 years old
- $A$  = student is studying biology,  $B$  = student is breathing
- $A$  = student is woman,  $B$  = student has brown hair