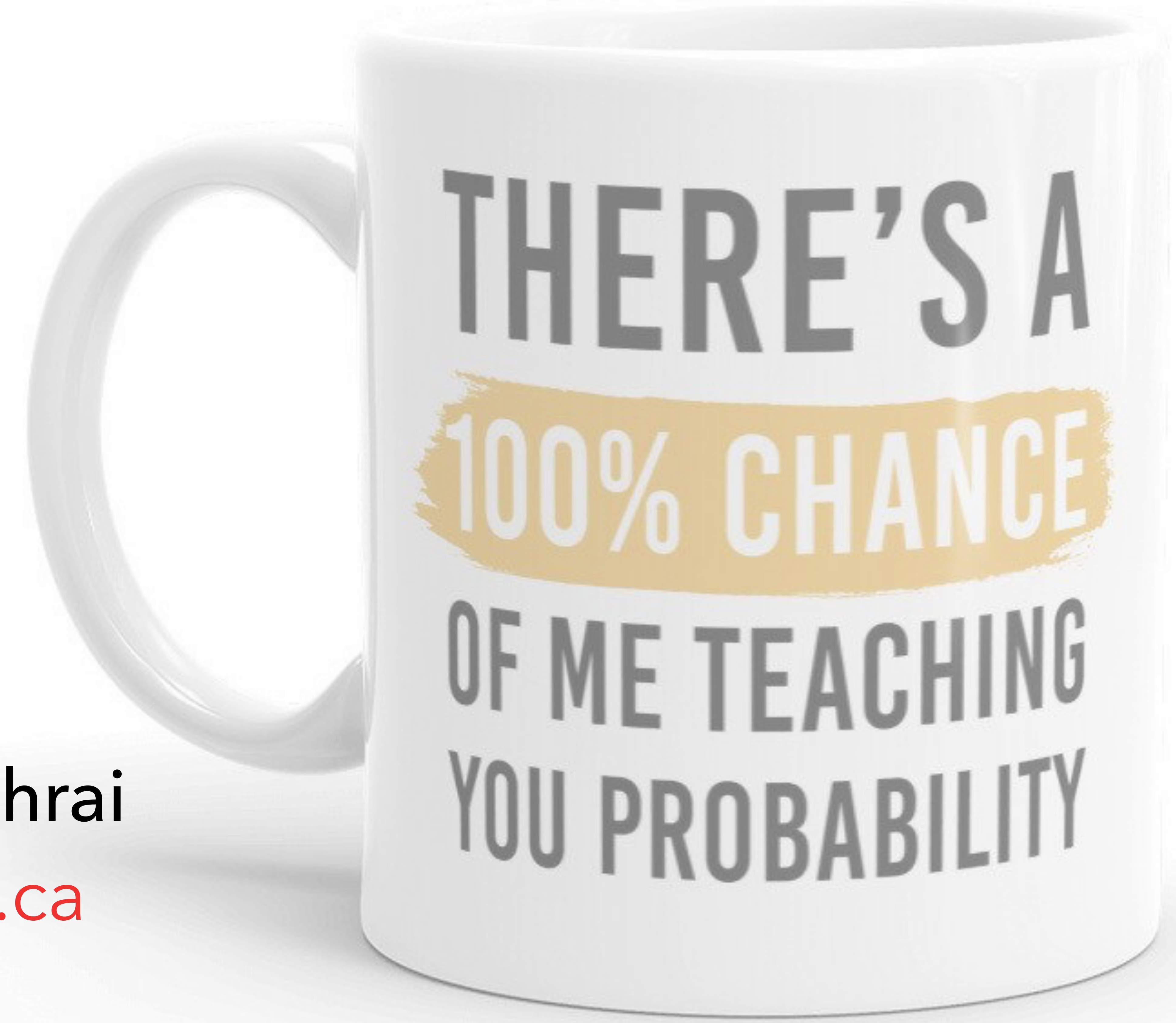
3A – PROBABILITY: REVIEW



Derek Nowrouzezahrai
derek@cim.mcgill.ca

Probability Basics

Probability Basics

Fundamental notion in probability is that of a **random experiment:**

- an experiment whose outcome cannot be determined in advance, but is nevertheless still subject to analysis

Probability Basics

Fundamental notion in probability is that of a **random experiment**:

- an experiment whose outcome cannot be determined in advance, but is nevertheless still subject to analysis

1) tossing a die

2) flipping a coin

3) counting the number of left-handers in this class

4) measuring the amount of snow in Montreal in February

Probability Basics

The **sample space** Ω of an experiment is the **set** of all possible **outcomes**

1) tossing a die: $\Omega = \{1,2,3,4,5,6\}$

2) tossing two dice: $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots(6,6)\}$

3) flipping a coin: $\Omega = \{H,T\}$

4) counting the number of left-handers here: $\Omega = \{0,1,\ldots\} = \mathbb{Z}_+$

5) measuring the amount of snow in Montreal: $\Omega = \mathbb{R}_+$

Probability Basics

An **event** A is a **subset** of the sample space of a random experiment

1) when the sum of two dice is 10 or more
   $\Omega = \{(1,1),\ldots,(1,6),(2,1),\ldots(6,6)\}$ & $A = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}$

2) when flipping a coin 3 times yields exactly one head
   $\Omega = \{(H,H,H), (H,H,T), \ldots, (T,T,T)\}$ & $A = \{(H,T,T), (T,H,T), (T,T,H)\}$

3) the number of left-handers in this class is 3
   $\Omega = \{0,1,2,\ldots\} = \mathbb{Z}_+$ & $A = \{3\}$

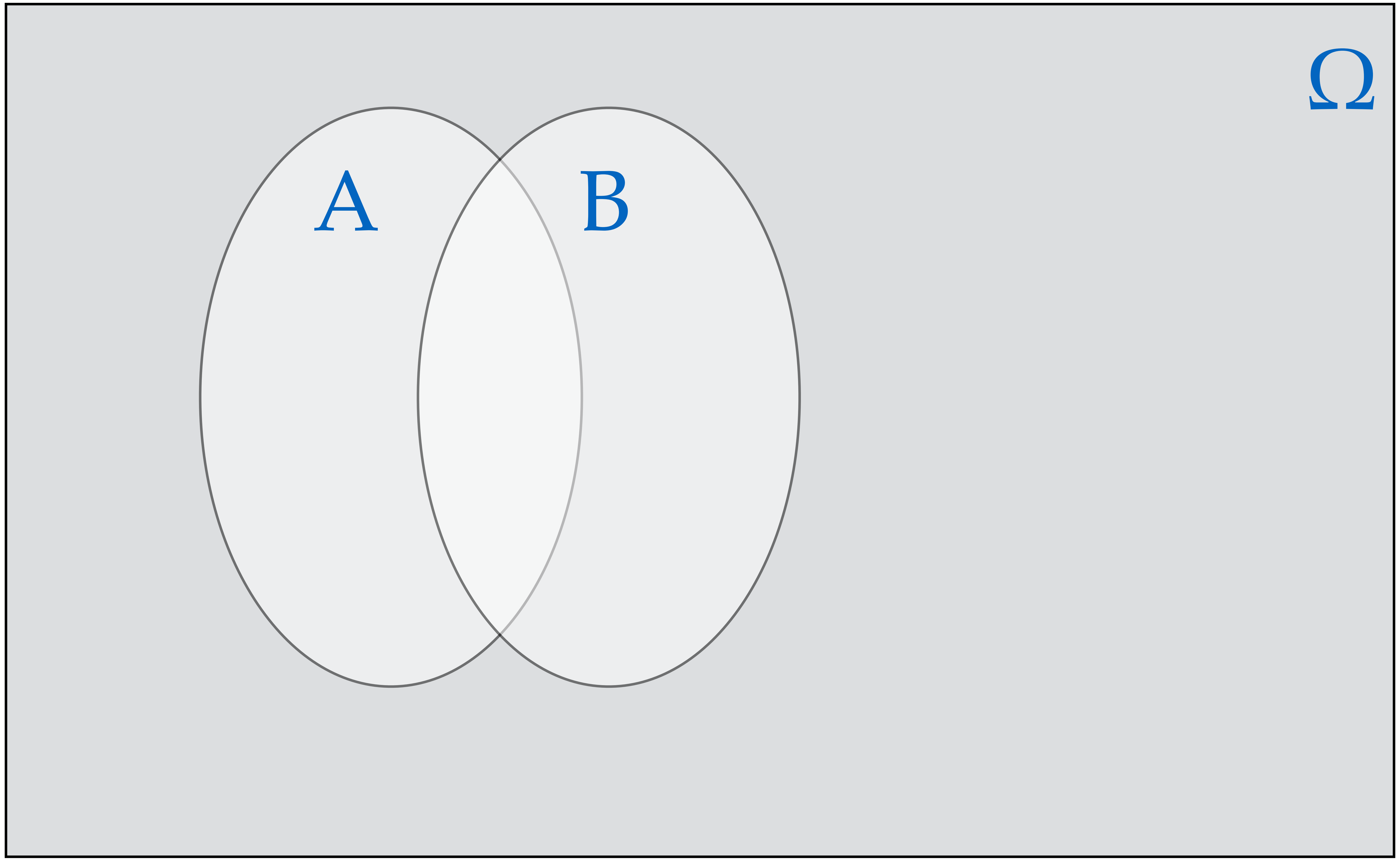
Probability Basics

An **event** A is a **subset** of the sample space of a random experiment

- the set $\mathrm{A} \cup \mathrm{B}$ (**union**) is the event that *either* A *or* B (or both) occur
- the set $\mathrm{A} \cap \mathrm{B}$ (**intersection**) is the event that A *and* B both occur

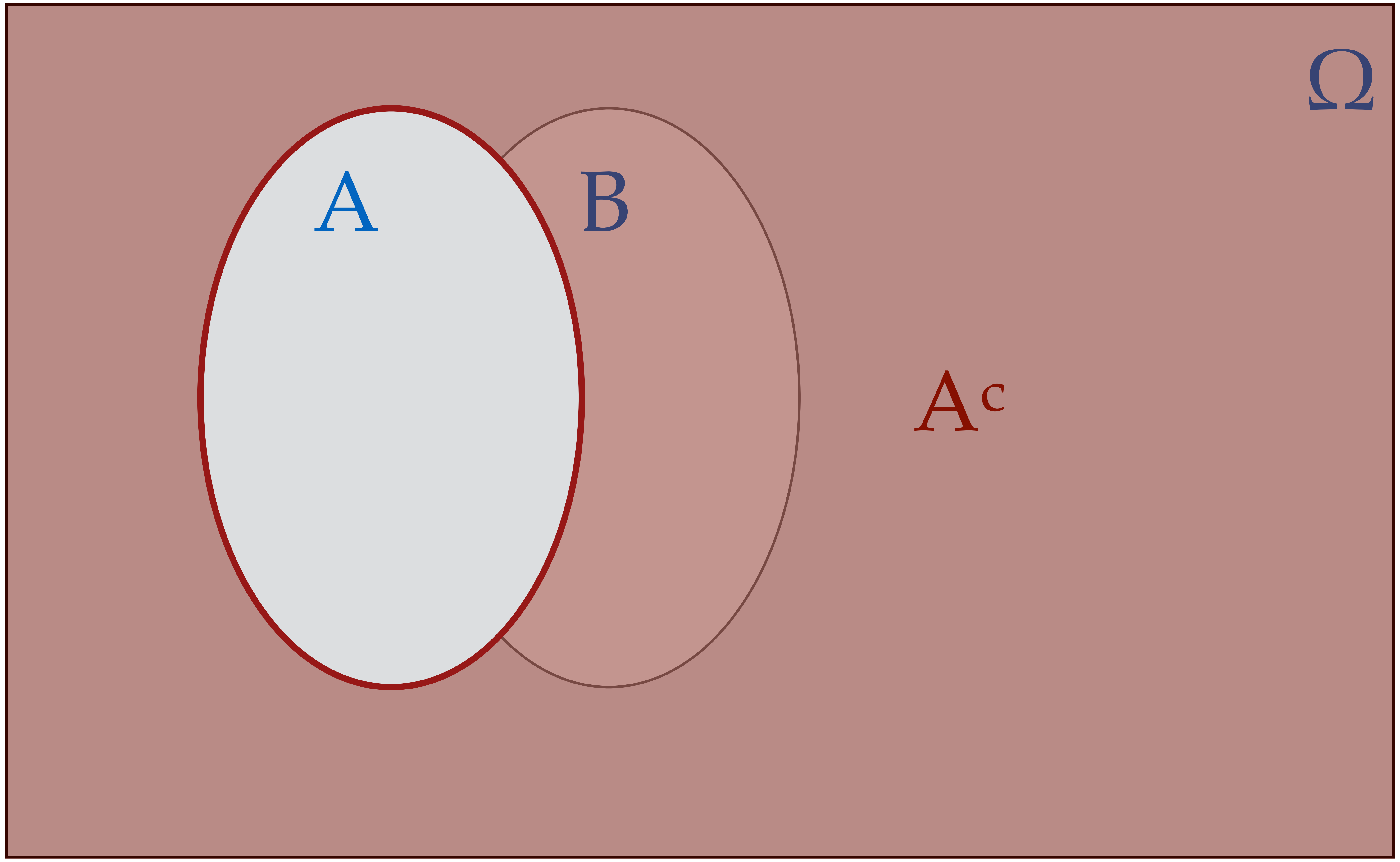
Probability Basics

An **event** A is a **subset** of the sample space of a random experiment



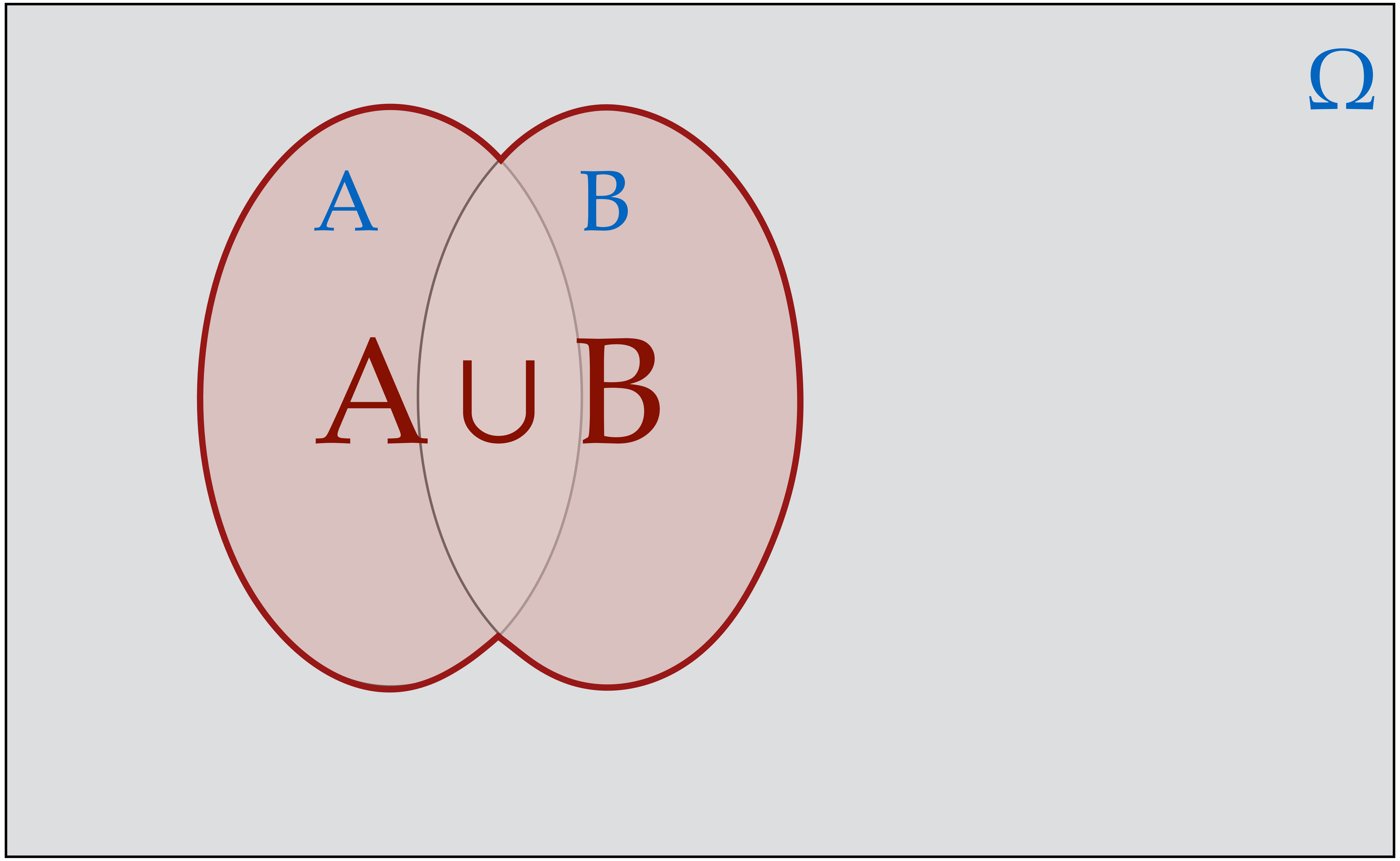
Probability Basics

An **event** A is a **subset** of the sample space of a random experiment



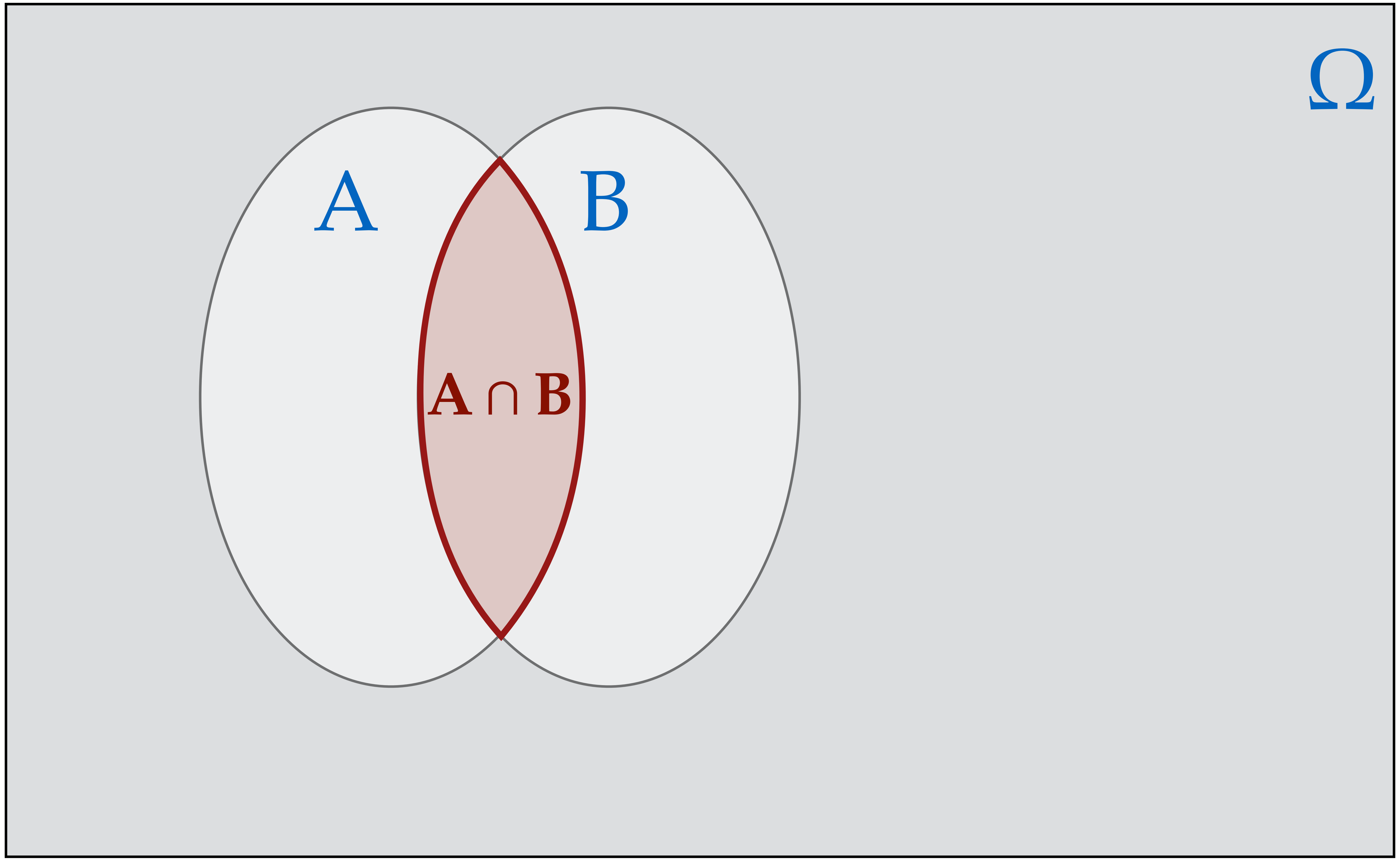
Probability Basics

An **event** A is a **subset** of the sample space of a random experiment



Probability Basics

An **event** A is a **subset** of the sample space of a random experiment



Probability Basics

Cast two dice consecutively. The sample space is

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$

Let the event A be when the first die is 6

- $A = \{(6,1),\ldots,(6,6)\}$

Let the event B be when the second die is 6

- $B = \{(1,6),\ldots,(6,6)\}$

The intersection is the event that both die are 6

- $A \cap B = \{(6,1),\ldots,(6,6)\} \cap \{(1,6),\ldots,(6,6)\} = \{(6,6)\}$

Probability Basics

An **event** A is a **subset** of the sample space of a random experiment

- the set $A \cup B$ (**union**) is the event that either A or B (or both) occur
- the set $A \cap B$ (**intersection**) is the event that A and B both occur
- the event A^c (**complement**) is the event that A does ***not*** occur

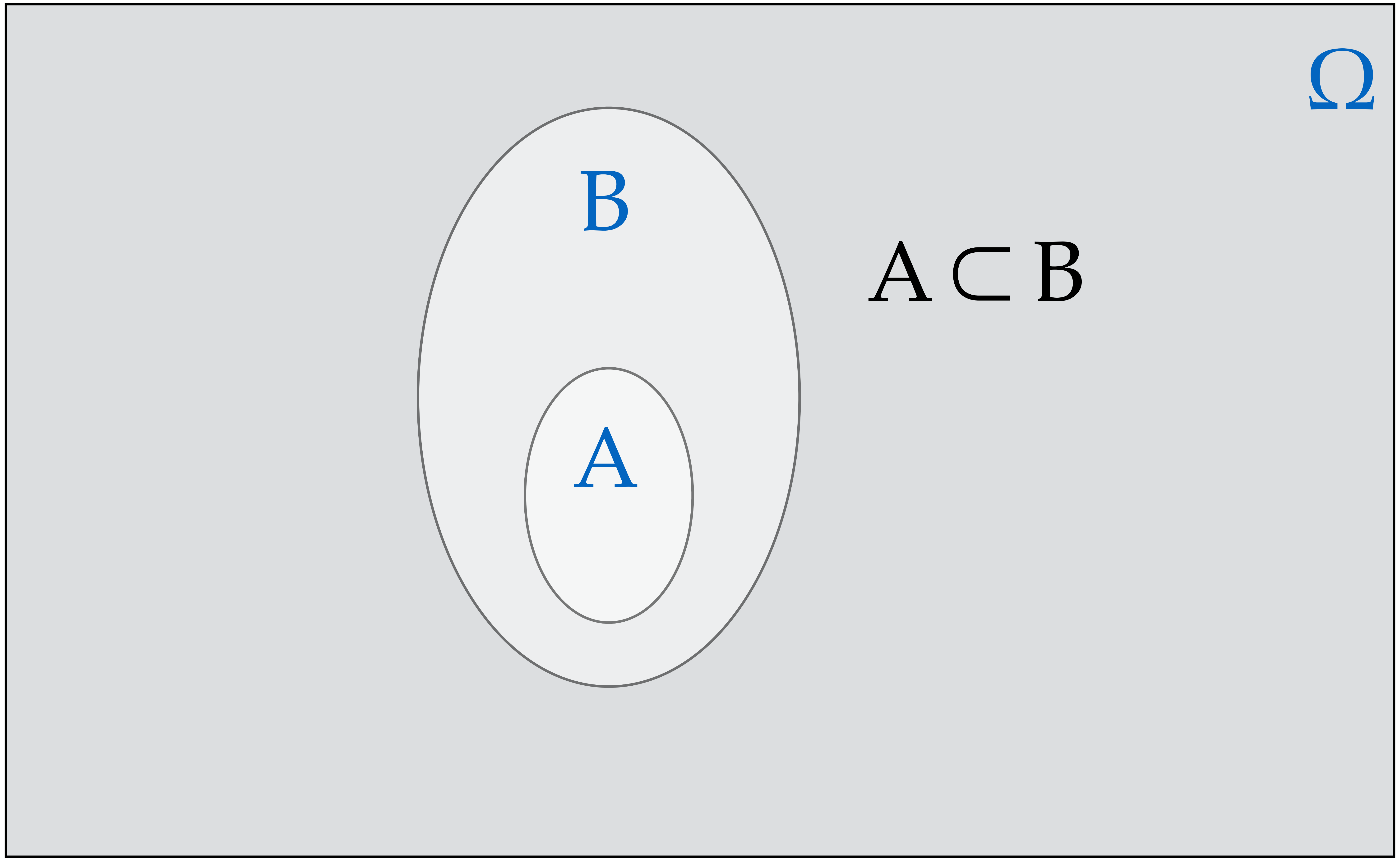
Probability Basics

An **event** A is a **subset** of the sample space of a random experiment

- the set $A \cup B$ (**union**) is the event that either A or B (or both) occur
- the set $A \cap B$ (**intersection**) is the event that A and B both occur
- the event A^c (**complement**) is the event that A does ***not*** occur
- if $A \subset B$ (**subset**) then the events in A are contained in B

Probability Basics

An **event** A is a **subset** of the sample space of a random experiment

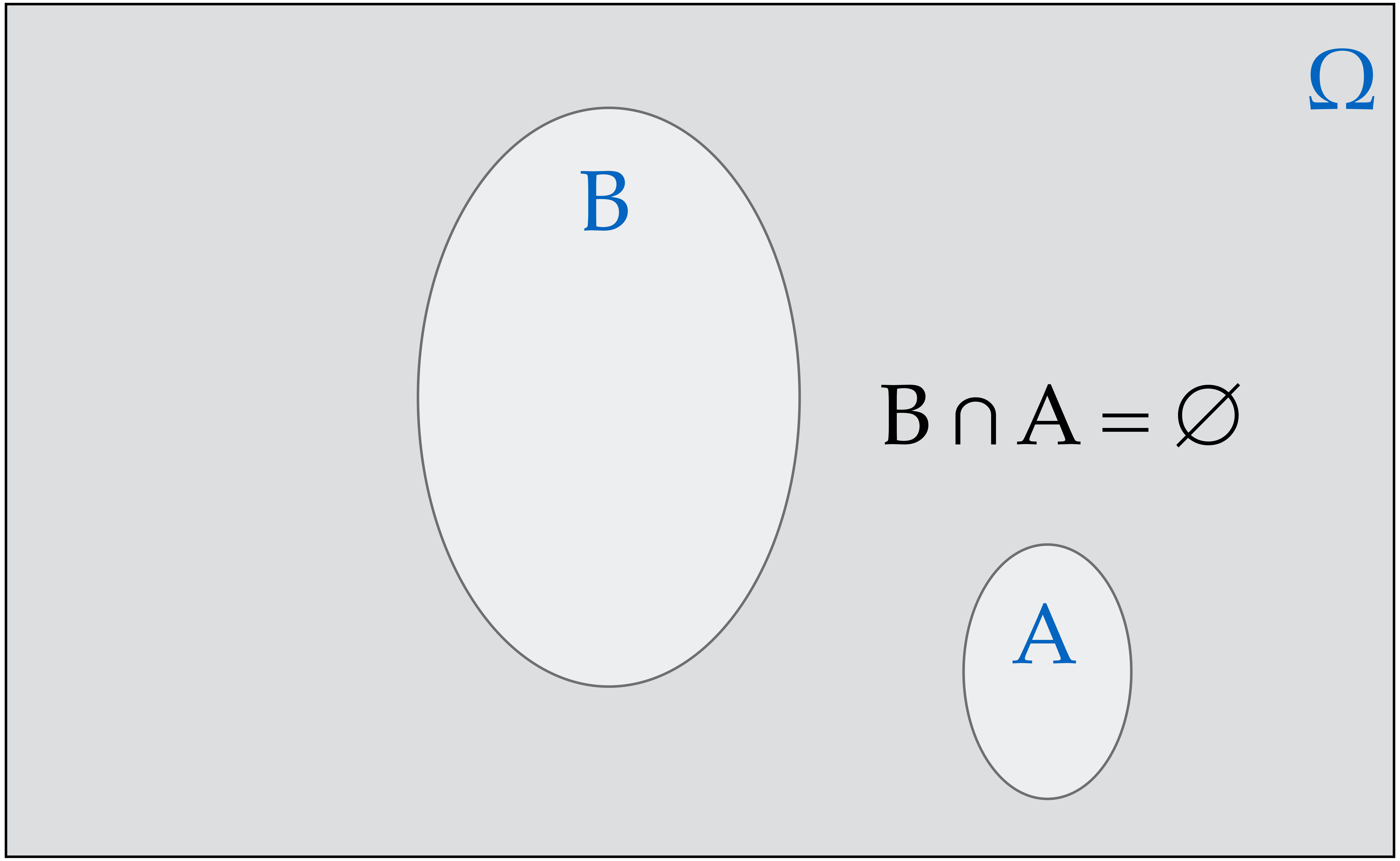


Probability Basics

An **event** A is a **subset** of the sample space of a random experiment

- the set $A \cup B$ (**union**) is the event that either A or B (or both) occur
- the set $A \cap B$ (**intersection**) is the event that A and B both occur
- the event A^c (**complement**) is the event that A does ***not*** occur
- if $A \subset B$ (**subset**) then the events in A are contained in B
- A and B are **disjoint** if $A \cap B = \varnothing$

Probability Basics

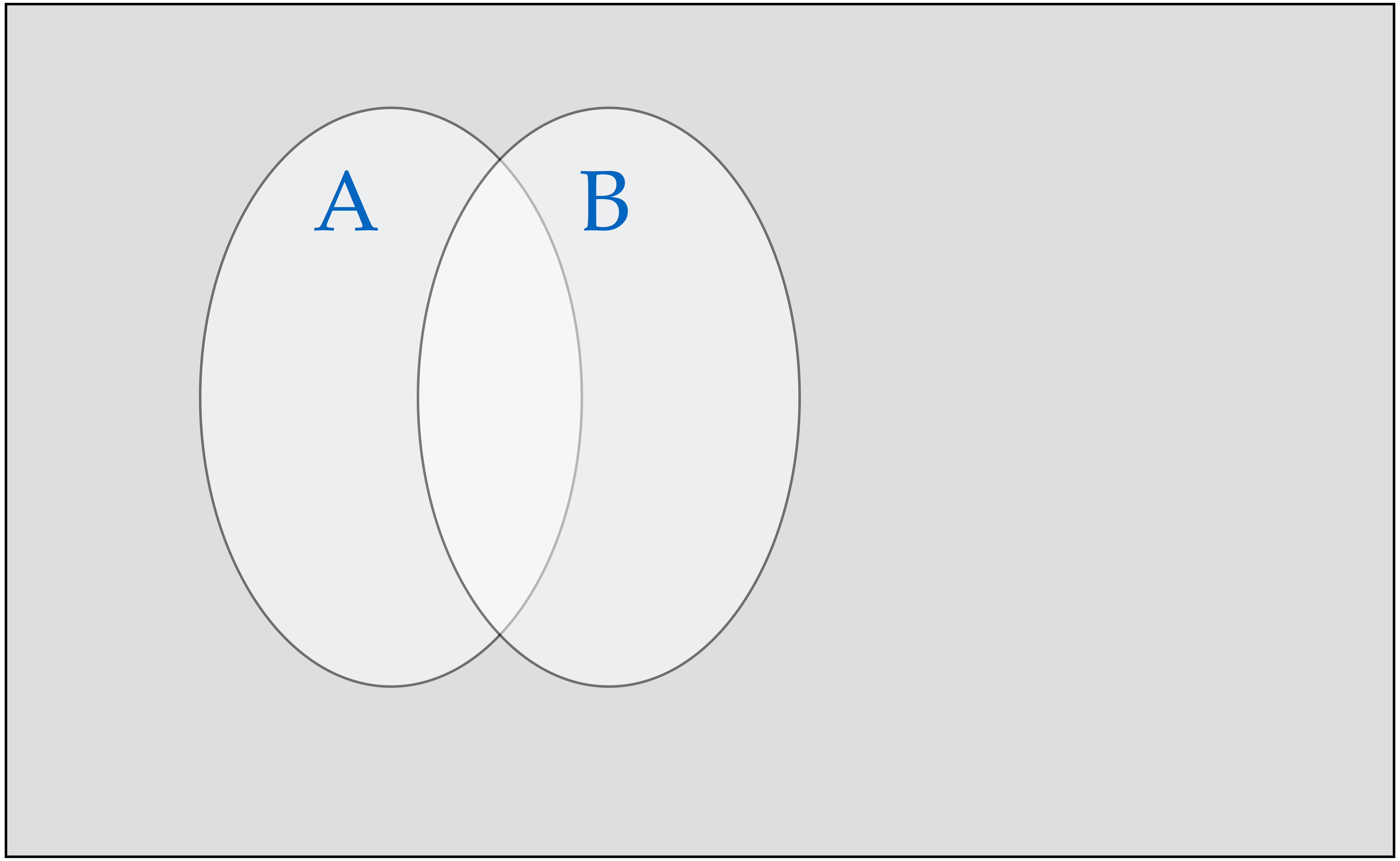


De Morgan's Laws

$$(A \cap B)^c = A^c \cup B^c$$

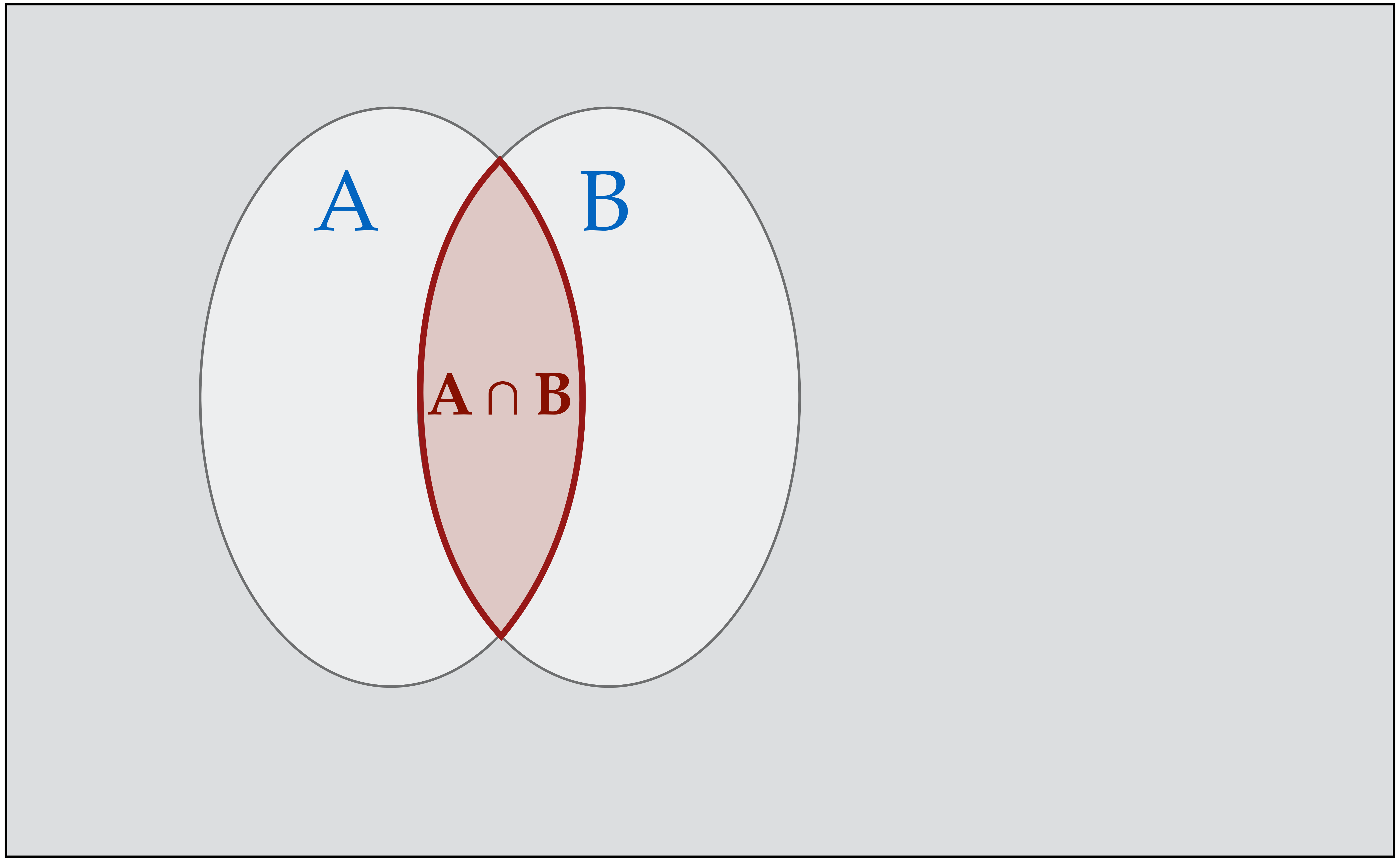
De Morgan’s Laws

$$(A \cap B)^c = A^c \cup B^c$$



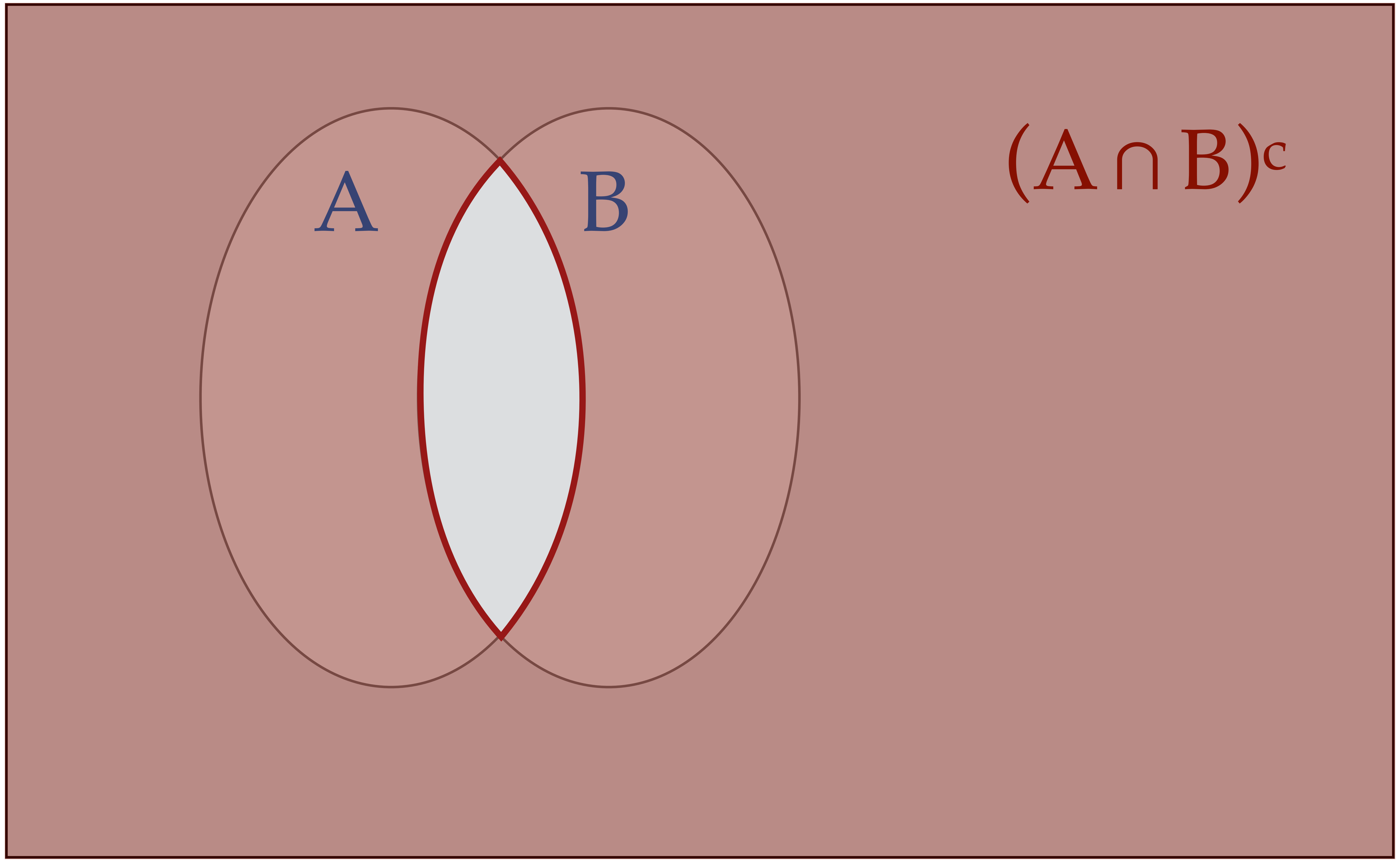
De Morgan’s Laws

$$(A \cap B)^c = A^c \cup B^c$$



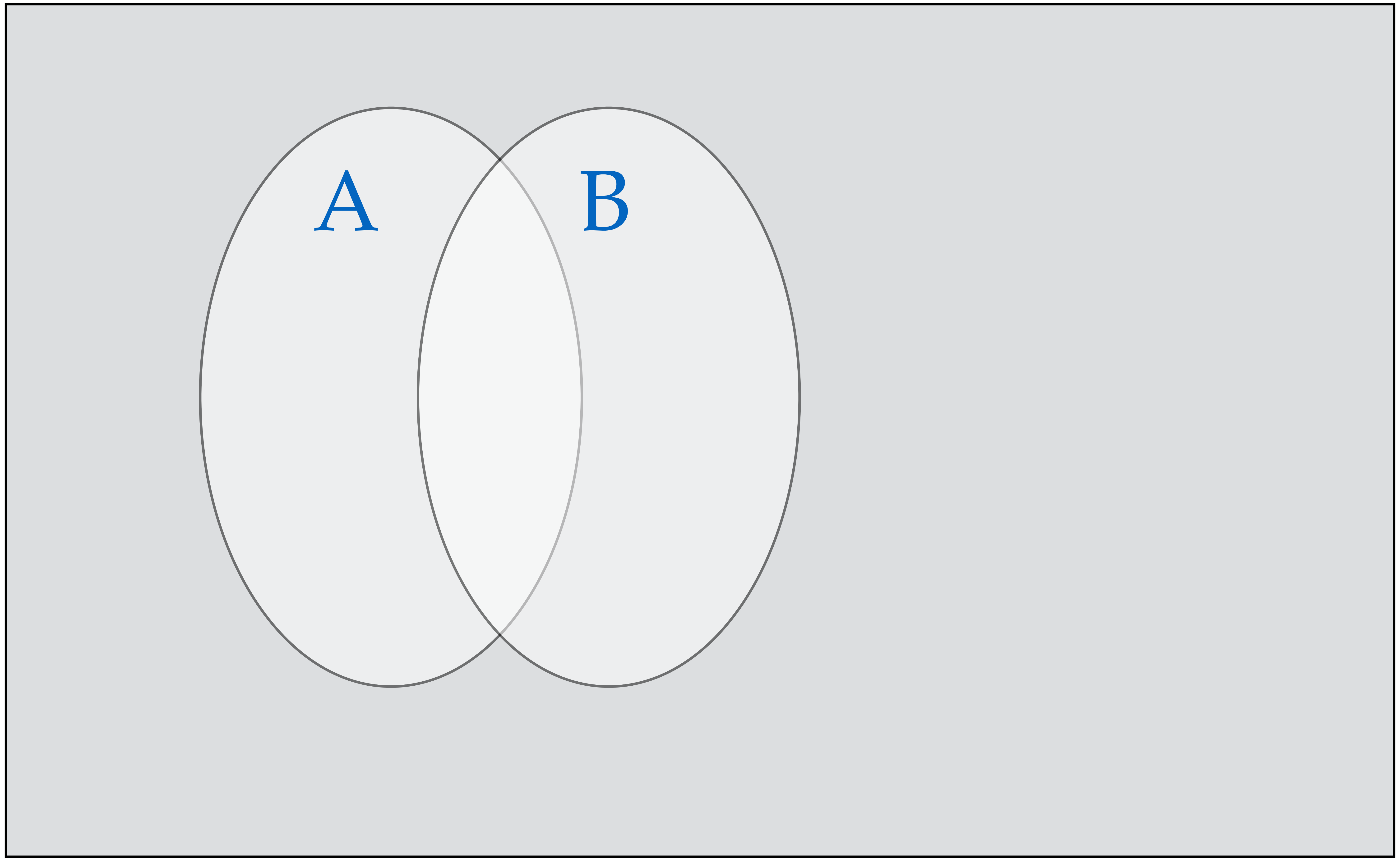
De Morgan’s Laws

$$(A \cap B)^c = A^c \cup B^c$$



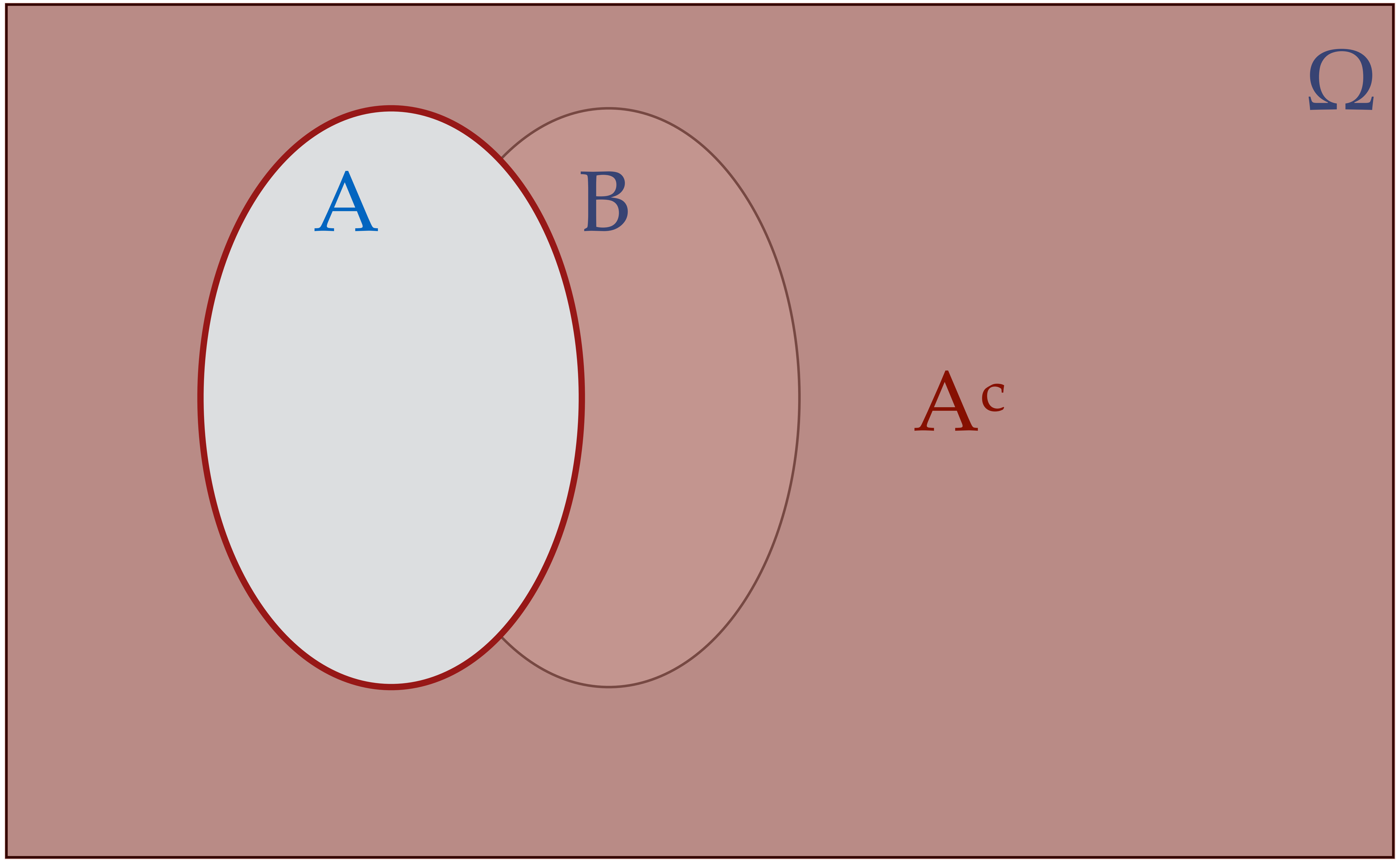
De Morgan's Laws

$$(A \cap B)^c = A^c \cup B^c$$



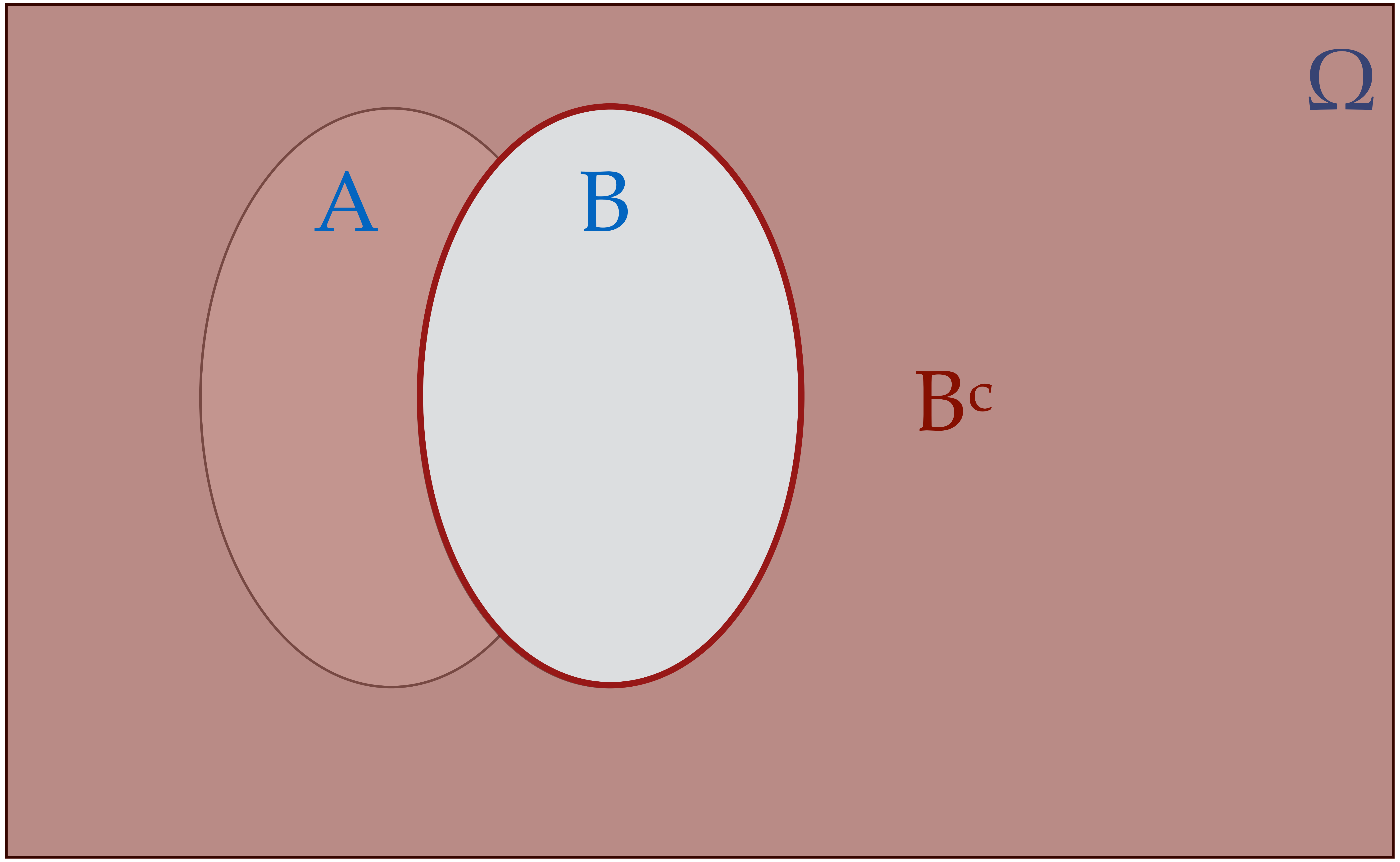
De Morgan’s Laws

$$(A \cap B)^c = A^c \cup B^c$$



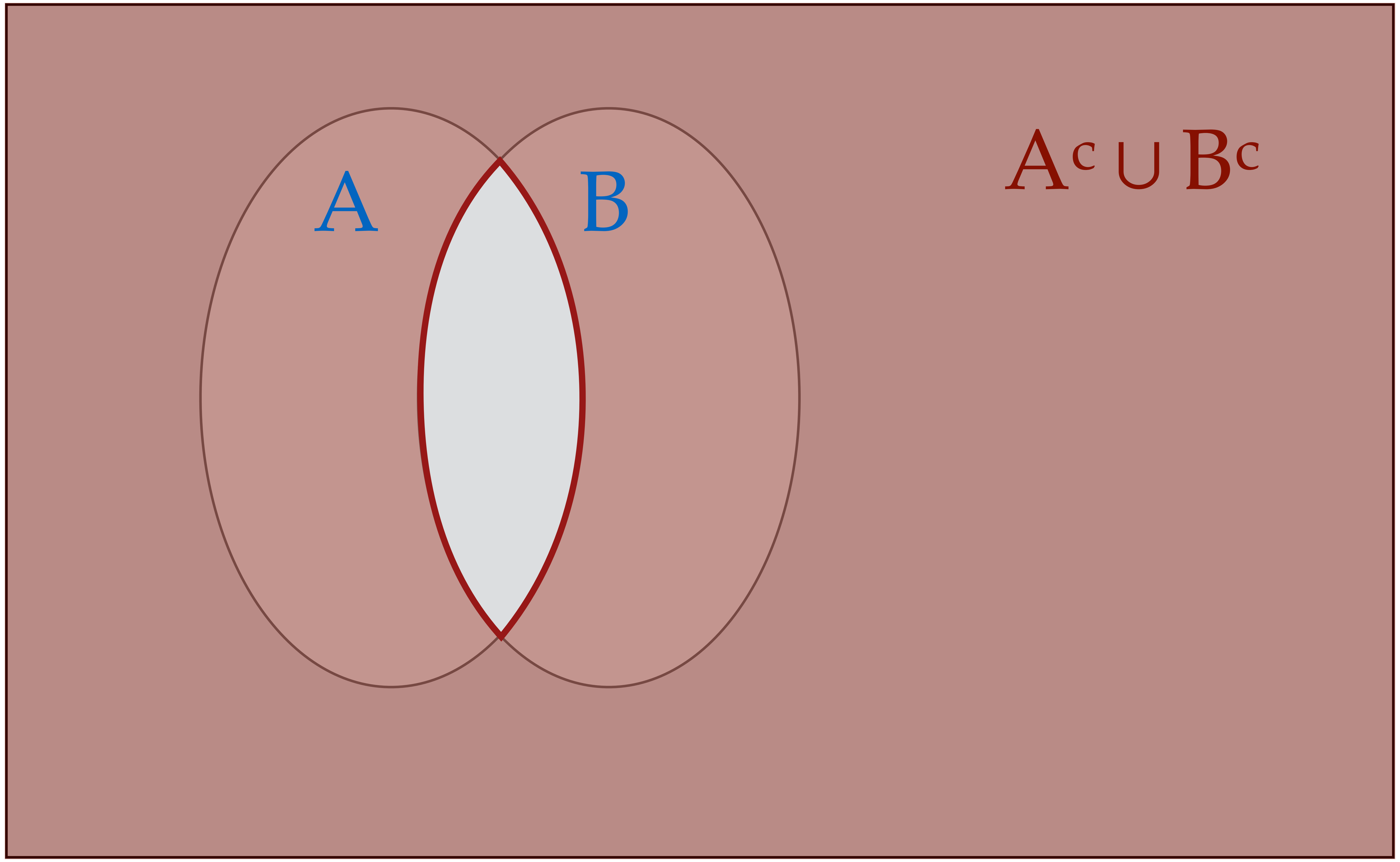
De Morgan's Laws

$$(A \cap B)^c = A^c \cup B^c$$



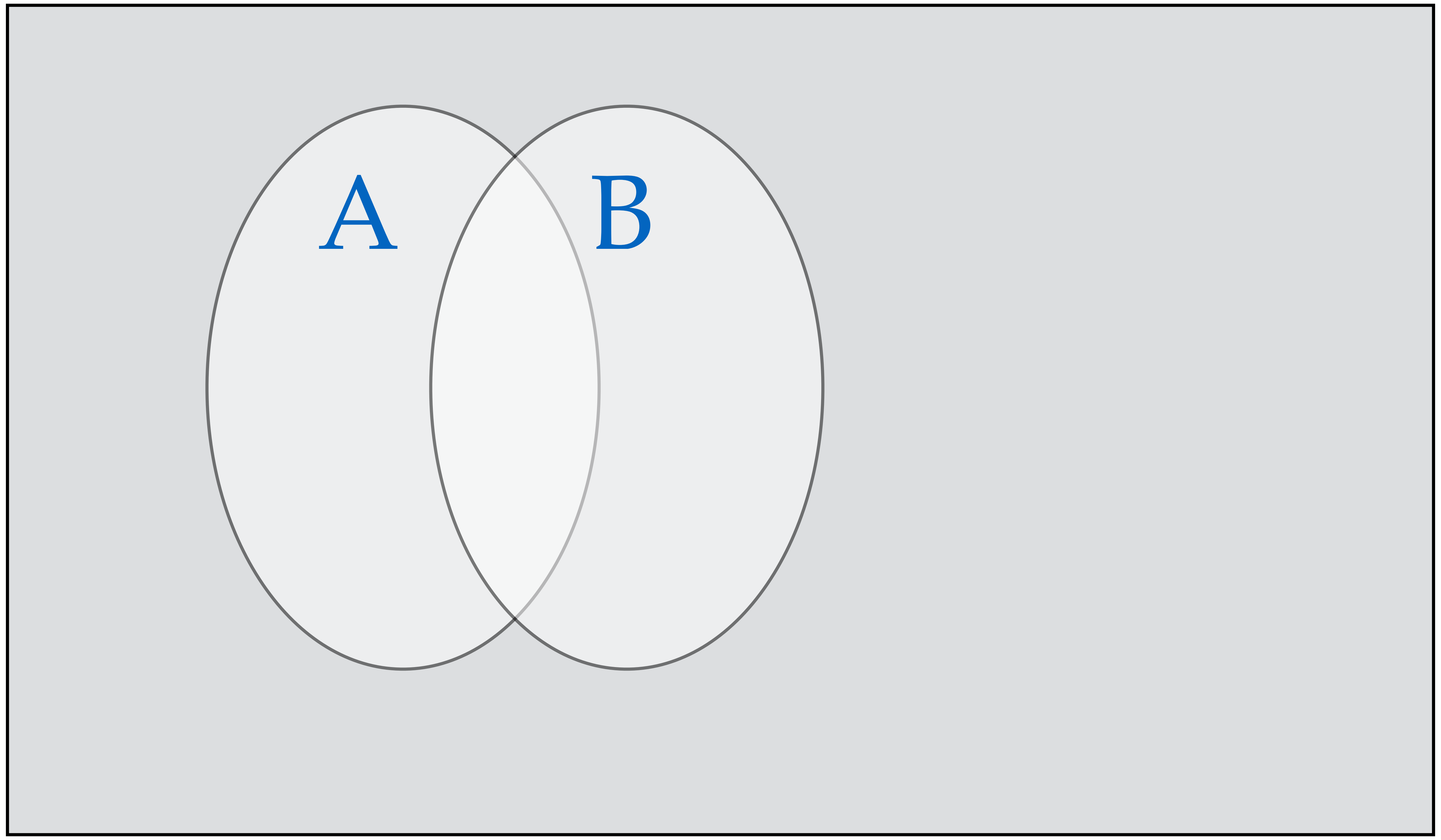
De Morgan's Laws

$$(A \cap B)^c = A^c \cup B^c$$



De Morgan's Laws

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$



Discrete Random Variables

Discrete Random Variables

Discrete Random Variable

- "discrete": countable/listable set of outcomes
- "random" because value depends on a random outcome
- "variable" because we will treat it as a variable; can add/subtract, etc.
- e.g., X: the value on one die; Y: the value on a second die; $Z = X + Y$
- For any value a we write $Z = a$ to mean the *event* consisting of all ways where $Z = a$, e.g., $Z = 5$ (all ways two dice add up to 5)
- $P(Z = a)$: probability that sum of two dice is a

Discrete Random Variables

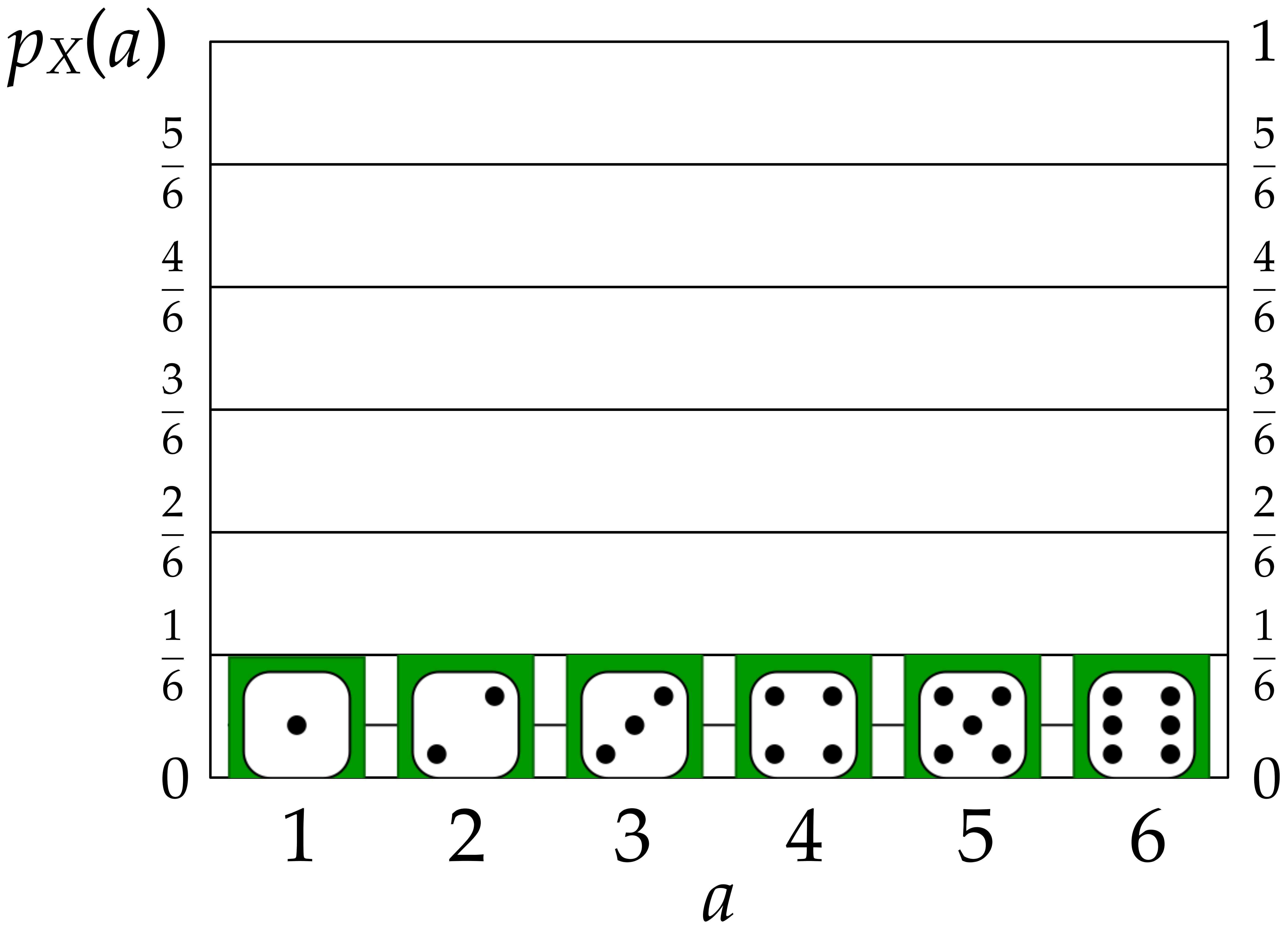
Probability mass function (*pmf*) of a discrete random variable X:

- $p_X(a) = P(X = a)$, or simply $p(a) = P(X = a)$
- sums to one (i.e., is normalized):

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

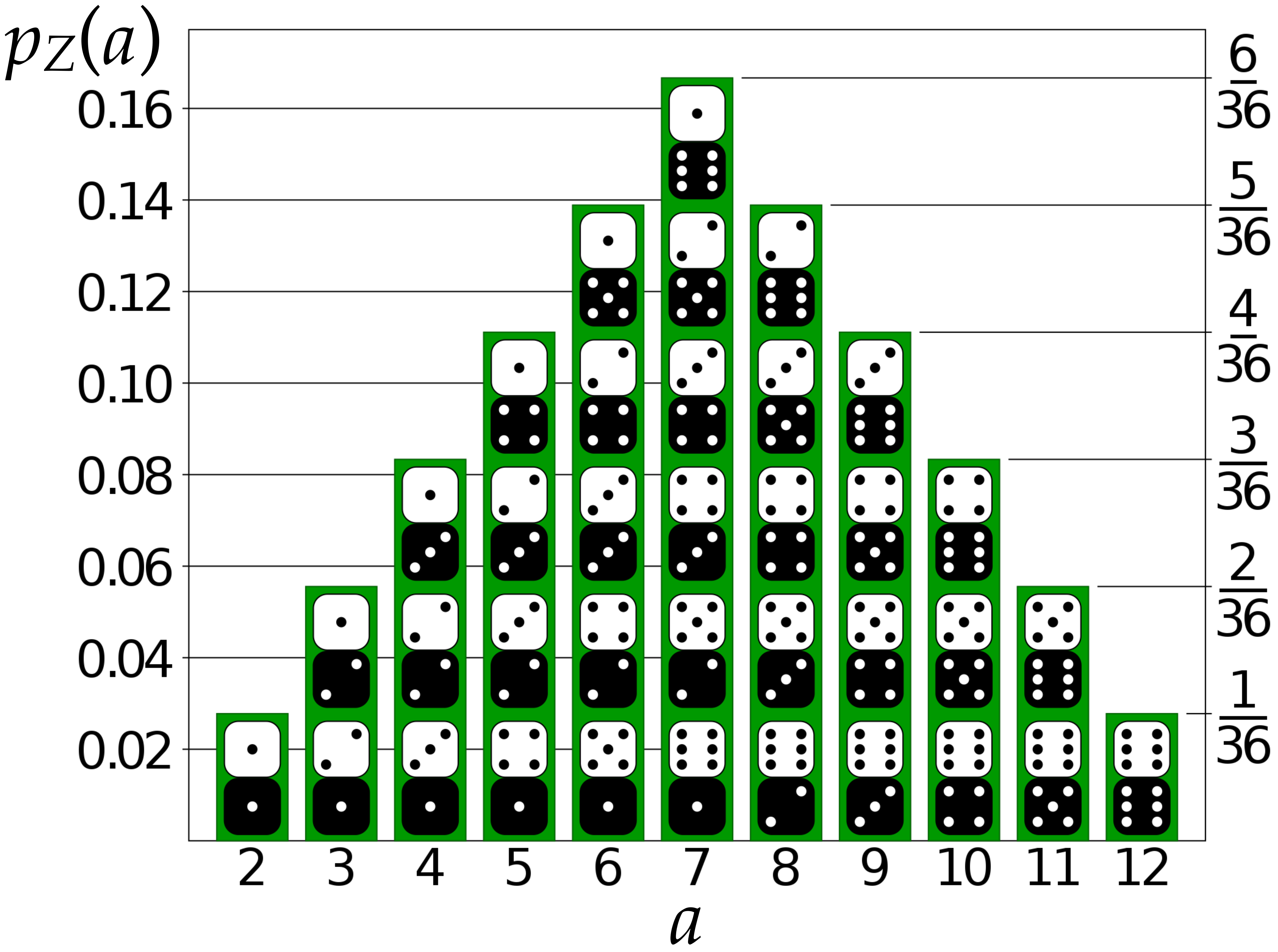
Probability mass function — example

X: the value of one fair die



Probability mass function – example

Z: the sum of two fair dice



Probability mass function

A probability mass function p is a function which assigns a positive number to each discrete event, and satisfies:

1. $p(a) \geq 0$, for all events a
2. $p(\Omega) = 1$, where Ω is sample space
3. for any sequence of **<u>disjoint</u>** events $a_1, a_2, \ldots, a_n$

$$p\left(\bigcup_{i=1}^{n} a_i\right) = \sum_{i=1}^{n} p(a_i)$$

Probability

A probability mass function p is a **function** which assigns a **positive number** to each discrete **event**:

- throw a single die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- let a be any event, in this case: $p(a) = |a| / 6$

Probability

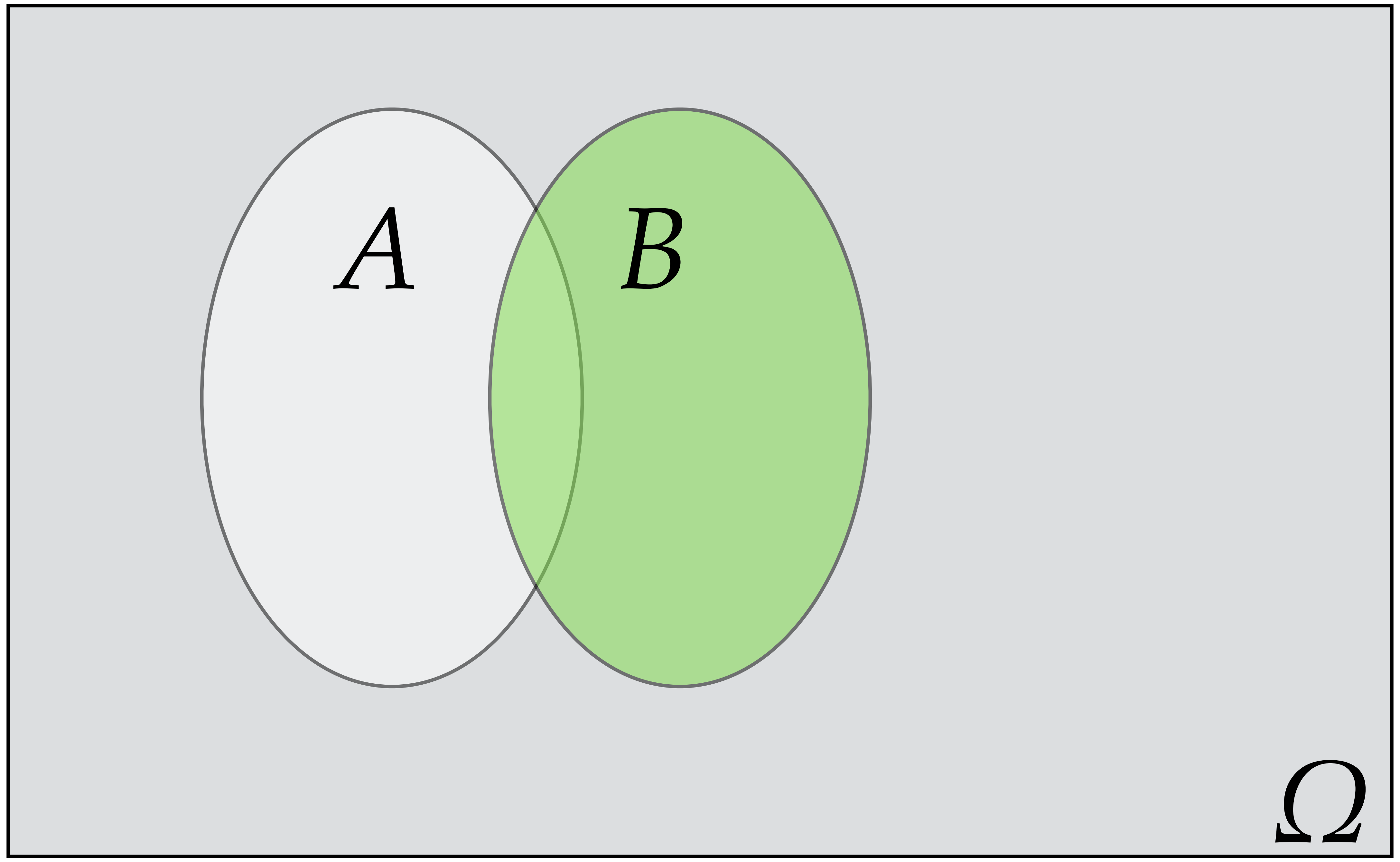
A probability mass function p is a **function** which assigns a **positive number** to each discrete **event**:

- throw a single die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- let a be the event where the die's value is even: $p(a) = 1/2$

Conditional Probability

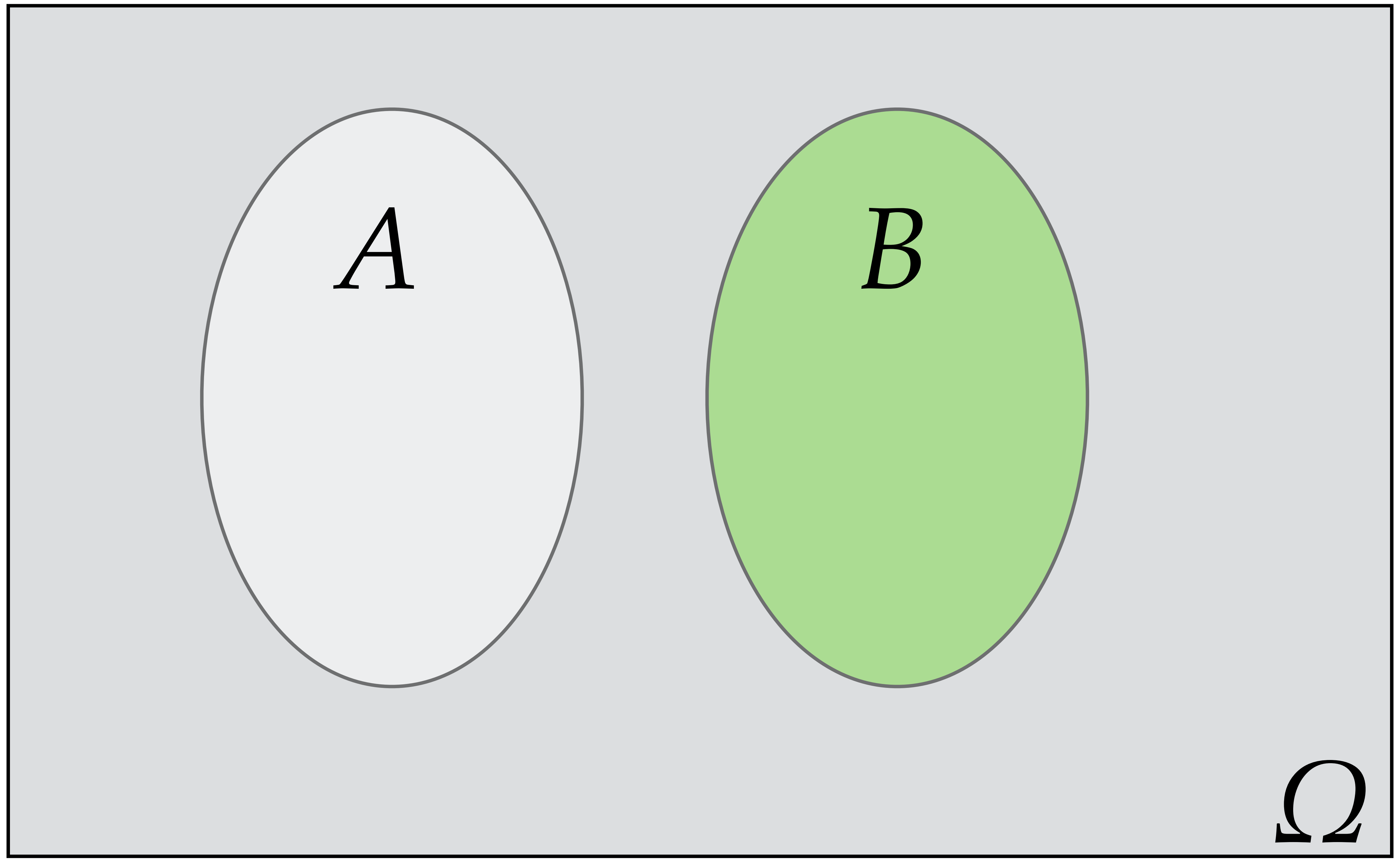
Conditional Probability

How does the probability of event $A \subset \Omega$ change when we know something about another event $B \subset \Omega$?



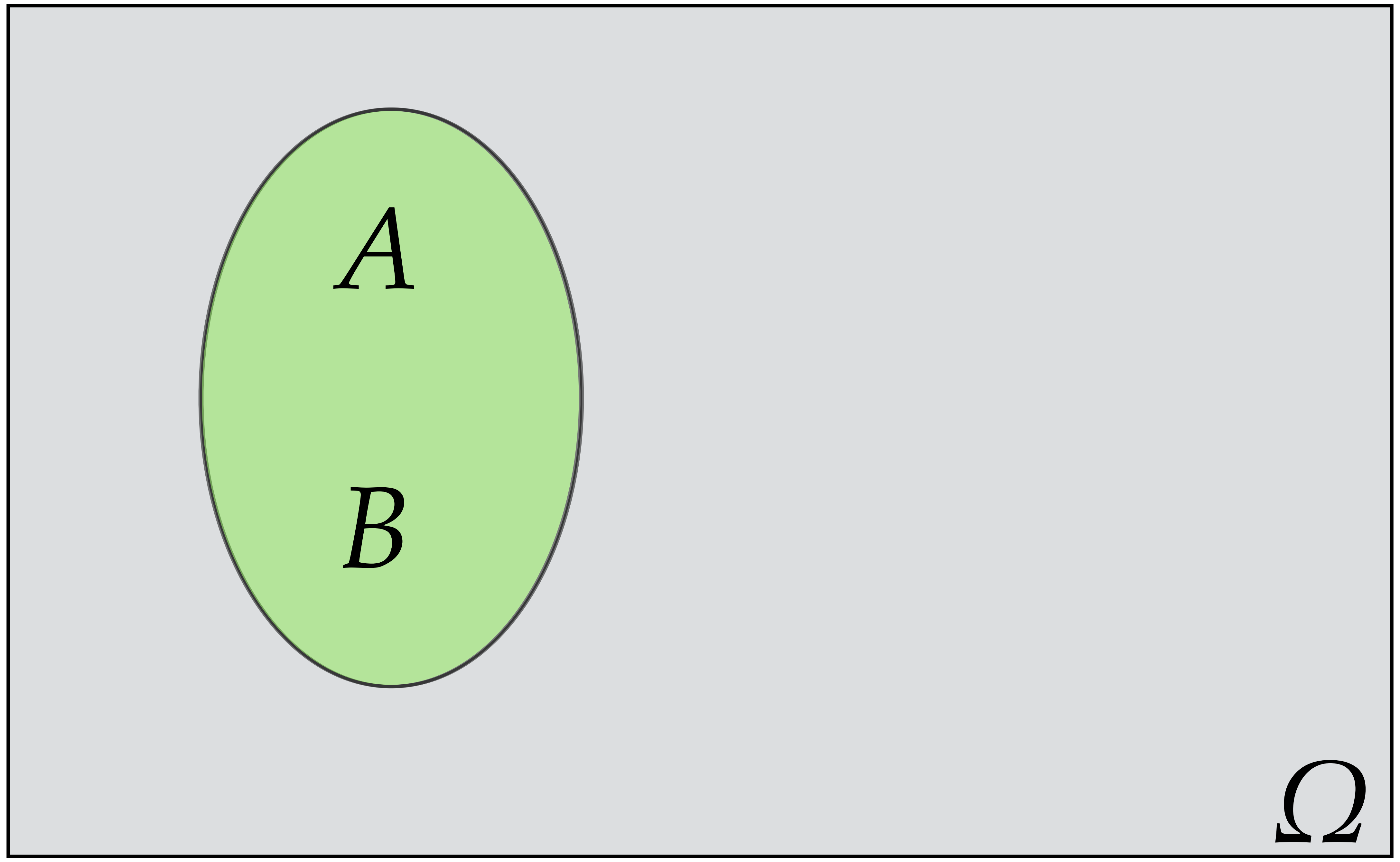
Conditional Probability

How does the probability of event $A \subset \Omega$ change when we know something about another event $B \subset \Omega$?



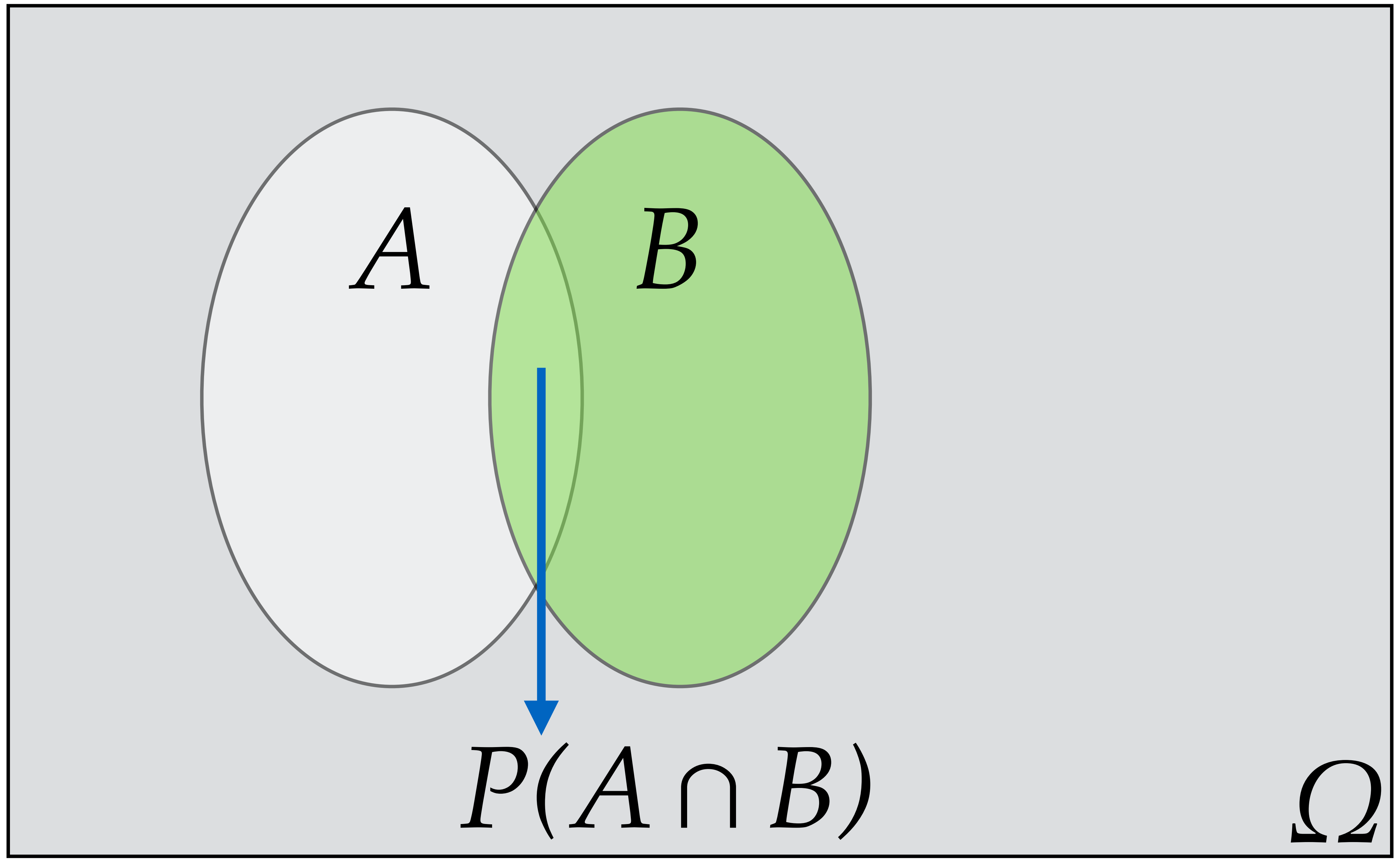
Conditional Probability

How does the probability of event $A \subset \Omega$ change when we know something about another event $B \subset \Omega$?



Conditional Probability

How does the probability of event $A \subset \Omega$ change when we know something about another event $B \subset \Omega$?



Given that B occurs, then A will occur only if $A \cap B$ occurs

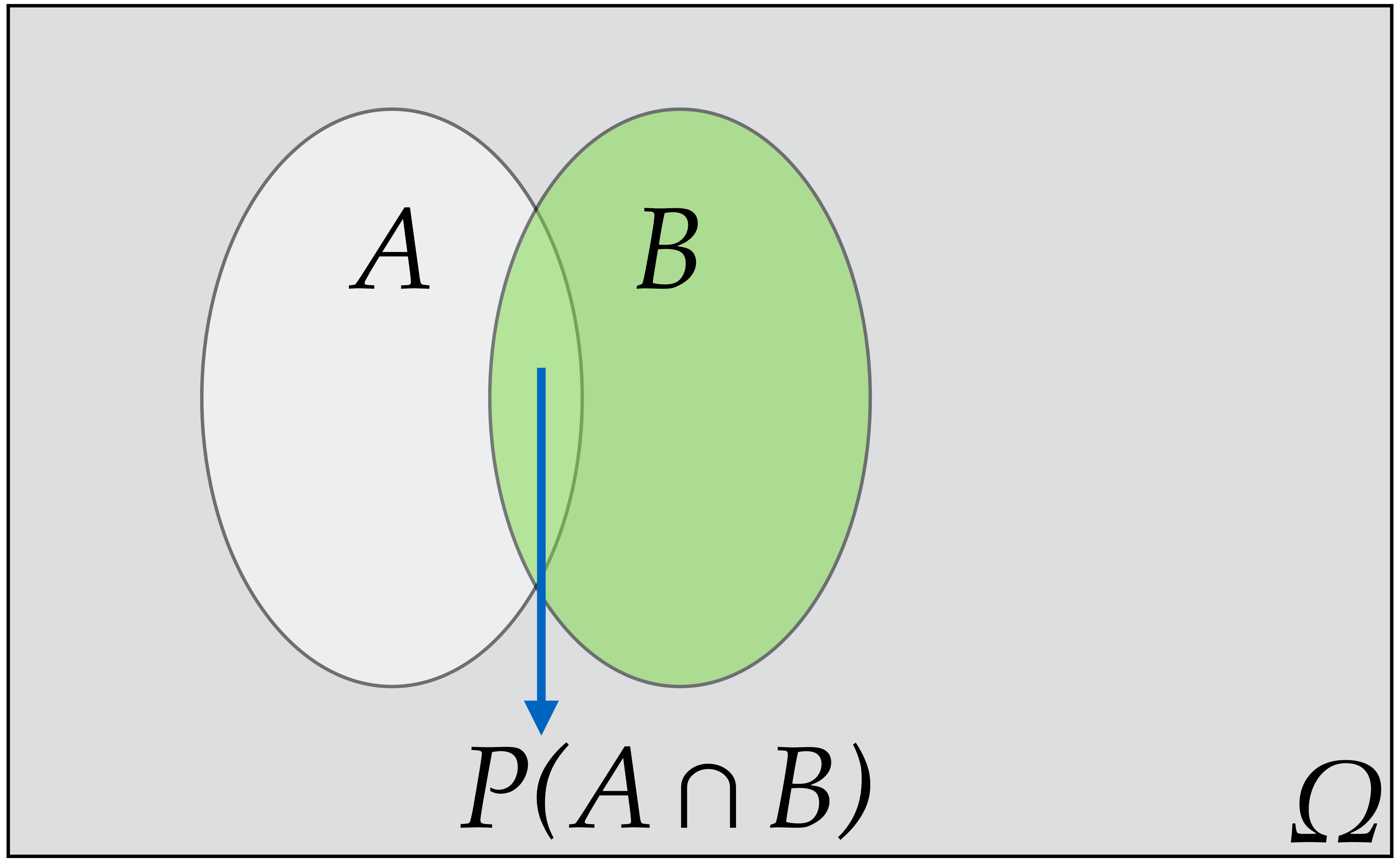
The probability of A occurring is:

$$P(A \cap B)/P(B)$$

(i.e., the sample space is now just B)

Conditional Probability

How does the probability of event $A \subset \Omega$ change when we know something about another event $B \subset \Omega$?



Conditional probability:

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

"the probability of A given B"

Conditional Probability – example

Two dice sum to 10. What's the probability that one 6 is cast?

- the size of the sample space is $|\Omega| = 36$
- let B be the event that the sum is 10: $B = \{(4,6), (5,5), (6,4)\}$
- let A be the event that one 6 is cast, $A = \{(1,6), (2,6), \ldots, (5,6)\ldots\}$
- the intersection of A and B is: $A \cap B = \{(4,6), (6,4)\}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{3/36} = \frac{2}{3}$$

Conditional Probability – example

Two dice sum to 10. What's the probability that one 1 is cast?

- the size of the sample space is $|\Omega| = 36$
- let B be the event that the sum is 10: $B = \{(4,6), (5,5), (6,4)\}$
- let A be the event that one 1 is cast, $A = \{(1,2), (1,3), \ldots, (1,6)\ldots\}$
- the intersection of A and B is: $A \cap B = \{\}$ (A and B are disjoint)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{3/36} = 0$$

Conditional Probability – example

Conditional Probability — example

A

B

C

Conditional Probability — example

A B C

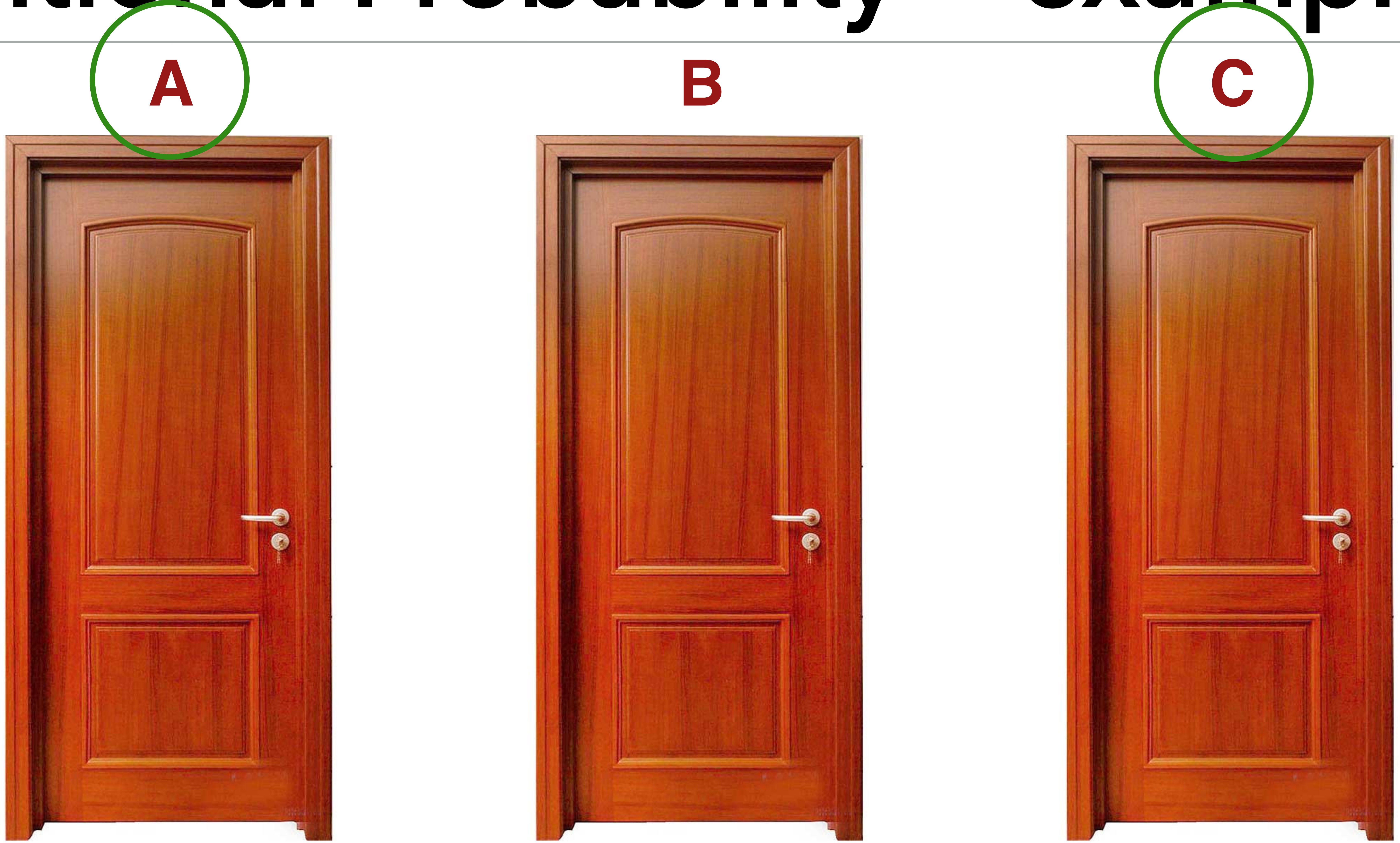
you choose door A

Conditional Probability – example

you choose door A

the host opens, say, door C

Conditional Probability – example



you choose door A

the host opens, say, door C

do you stay with A or switch to B?

Conditional Probability – example

A B C

you choose door A

the host opens either door B or C
and reveals junk

Conditional Probability — example

A

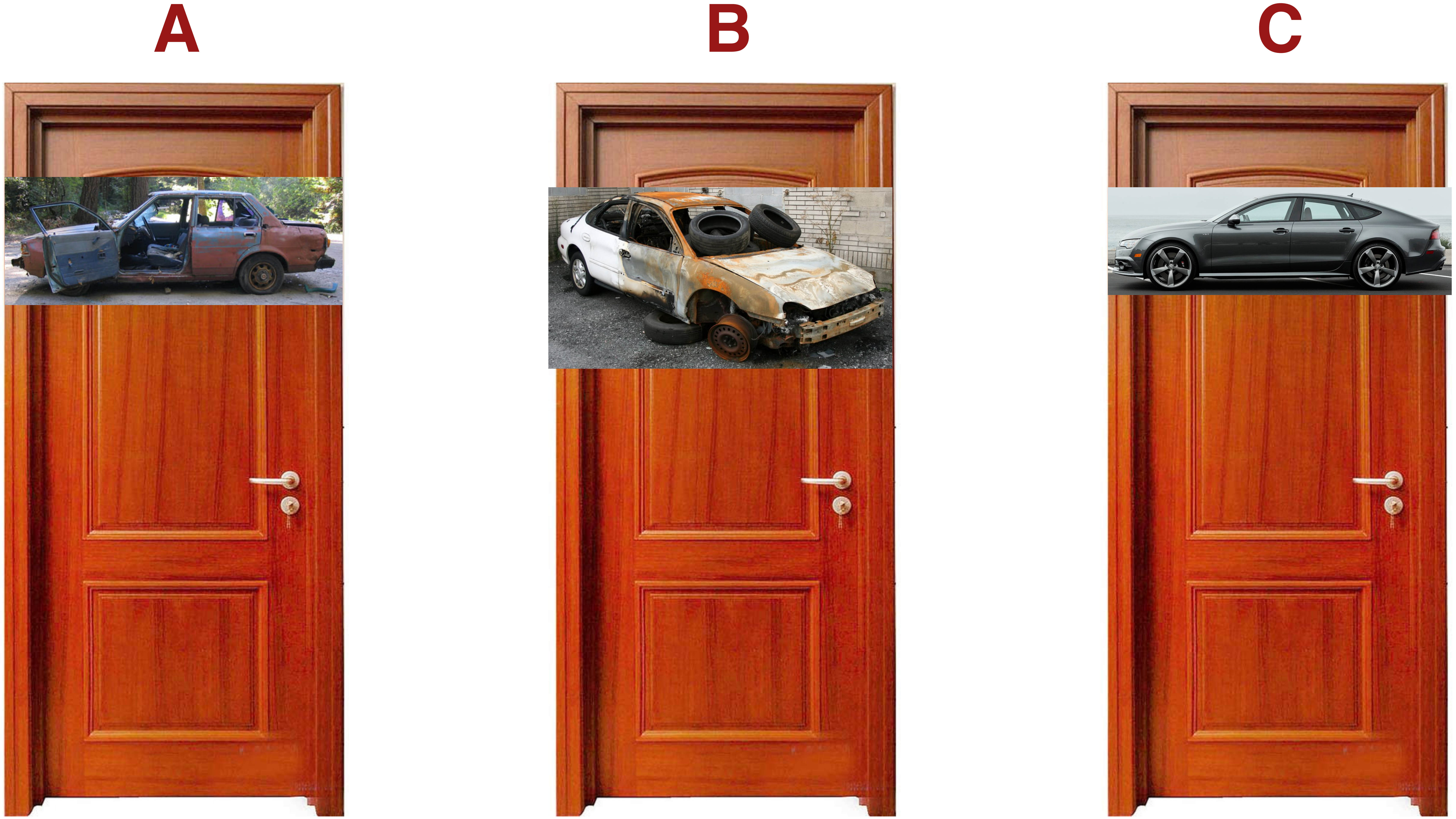
B

C

you choose door A

the host opens door C
and reveals junk

Conditional Probability – example



you choose door A

the host opens
door B and reveals junk

Conditional Probability — example

A B C

you choose door A

the host opens door B

$P(A \mid B \text{ is opened}) = ?$

$P(C \mid B \text{ is opened}) = ?$

Conditional Probability — example

Without loss of generality, you pick door A

A_c: prize is behind A, C is opened | $P(A_{prize}) = 1/3$
A_b: prize is behind A, B is opened |

B_c: prize is behind B, C is opened | $P(B_{prize}) = 1/3$

C_b: prize is behind C, B is opened | $P(C_{prize}) = 1/3$

Conditional Probability — example

Without loss of generality, you pick door A

A_c: prize is behind A, C is opened
A_b: prize is behind A, B is opened | $P(A_{prize}) = 1/3$

B_c: prize is behind B, C is opened | $P(B_{prize}) = 1/3$

C_b: prize is behind C, B is opened | $P(C_{prize}) = 1/3$

$P(A_c) = 1/6$
$P(A_b) = 1/6$

Conditional Probability — example

Without loss of generality, you pick door A

A_c: prize is behind A, C is opened | $P(A_{prize}) = 1/3$
A_b: prize is behind A, B is opened |
B_c: prize is behind B, C is opened | $P(B_{prize}) = 1/3$
C_b: prize is behind C, B is opened | $P(C_{prize}) = 1/3$

$P(A_c) = 1/6$
$P(A_b) = 1/6$

Don't switch

$$
\begin{aligned}
& P(A_{prize} \mid B_{open}) \\
&= P(A_{prize} \cap B_{open}) / P(B_{open}) \\
&= P(\{A_c, A_b\} \cap \{A_b, C_b\}) / P(\{A_b, C_b\}) \\
&= P(A_b) / P(\{A_b, C_b\}) \\
&= (1/6) / (1/6 + 1/3) \\
&= (1/6) / (1/2) \\
&= 1/3
\end{aligned}
$$

Conditional Probability — example

Without loss of generality, you pick door A

A_c: prize is behind A, C is opened | $P(A_{prize}) = 1/3$
A_b: prize is behind A, B is opened |
B_c: prize is behind B, C is opened | $P(B_{prize}) = 1/3$
C_b: prize is behind C, B is opened | $P(C_{prize}) = 1/3$

$P(A_c) = 1/6$
$P(A_b) = 1/6$

Don't switch

$P(A_{prize} | C_{open})$

Conditional Probability — example

Without loss of generality, you pick door A

A_c: prize is behind A, C is opened
A_b: prize is behind A, B is opened | $P(A_{prize}) = 1/3$

B_c: prize is behind B, C is opened | $P(B_{prize}) = 1/3$

C_b: prize is behind C, B is opened | $P(C_{prize}) = 1/3$

$P(A_c) = 1/6$
$P(A_b) = 1/6$

Don't switch

$$
\begin{aligned}
&P(A_{prize} \mid C_{open}) \\
&\quad = P(A_{prize} \cap C_{open}) \,/\, P(C_{open}) \\
&\quad = P(\{A_c, A_b\} \cap \{A_c, B_c\}) \,/\, P(\{A_c, B_c\}) \\
&\quad = P(A_c) \,/\, P(\{A_b, B_c\}) \\
&\quad = (1/6) \,/\, (1/6 + 1/3) \\
&\quad = (1/6) \,/\, (1/2) \\
&\quad = 1/3
\end{aligned}
$$

Conditional Probability — example

Without loss of generality, you pick door A

A_c: prize is behind A, C is opened | $P(A_{prize}) = 1/3$

A_b: prize is behind A, B is opened |

B_c: prize is behind B, C is opened | $P(B_{prize}) = 1/3$

C_b: prize is behind C, B is opened | $P(C_{prize}) = 1/3$

$P(A_c) = 1/6$

$P(A_b) = 1/6$

Switch

$P(C_{prize} \mid B_{open})$

$= P(C_{prize} \cap B_{open}) / P(B_{open})$

$= P(\{C_b\} \cap \{A_b, C_b\}) / P(\{A_b, C_b\})$

$= P(C_b) / P(\{A_b, C_b\})$

$= (1/3) / (1/6 + 1/3)$

$= (1/3) / (1/2)$

$= 2/3$

Conditional Probability — Recap

Conditional Probability

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

The events A and B are independent if knowledge of B does not have an effect on A (or vice versa)

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

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

Conditional Probability — Recap

Conditional Probability

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

The events A and B are independent if knowledge of B does not have an effect on A (or vice versa)

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

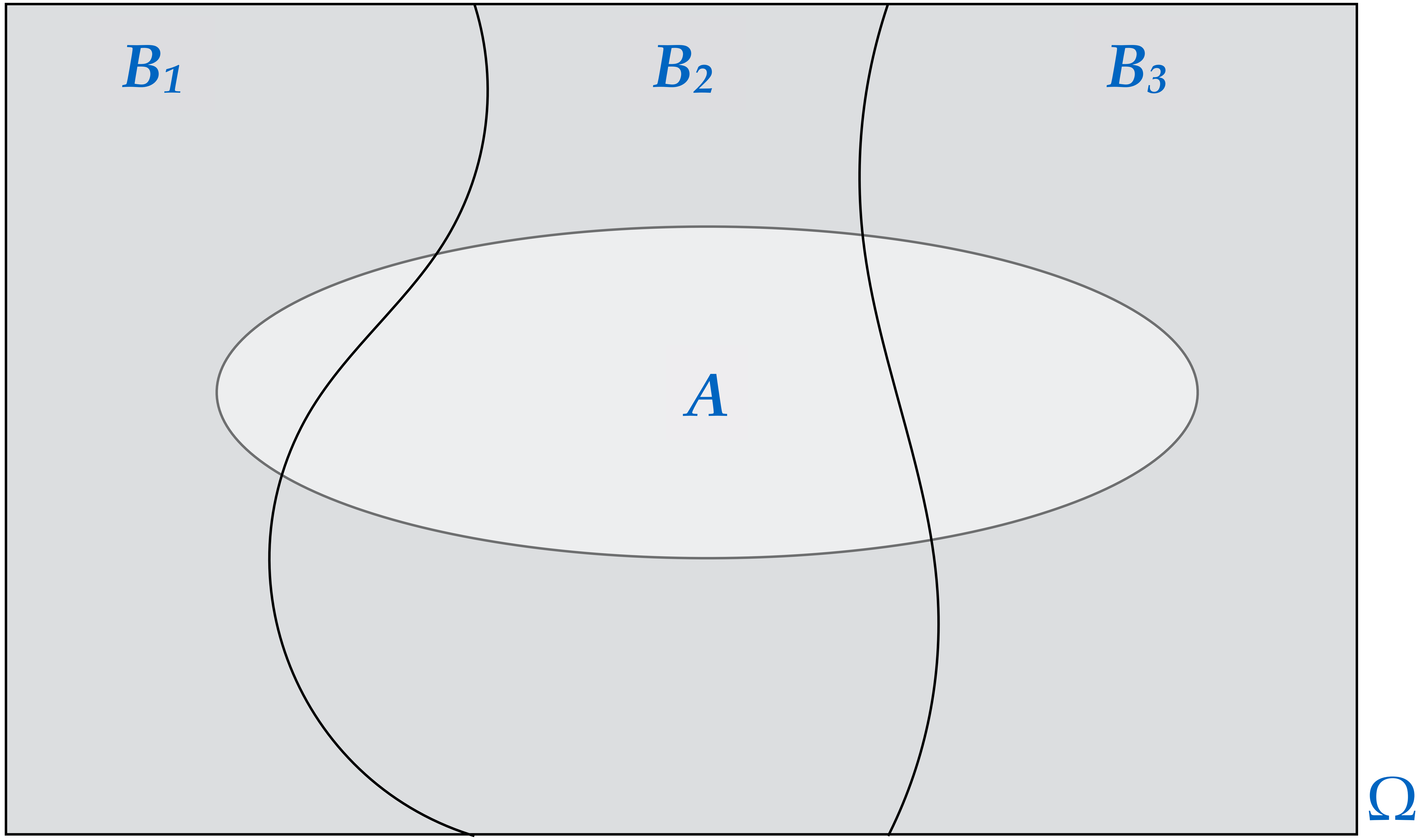
$$P(A|B) = \frac{P(A)P(B)}{P(B)} \longrightarrow P(A \cap B) = P(A)P(B)$$

Total Probability

Total Probability

B_1, B_2, and B_3 are disjoint and their union is Ω

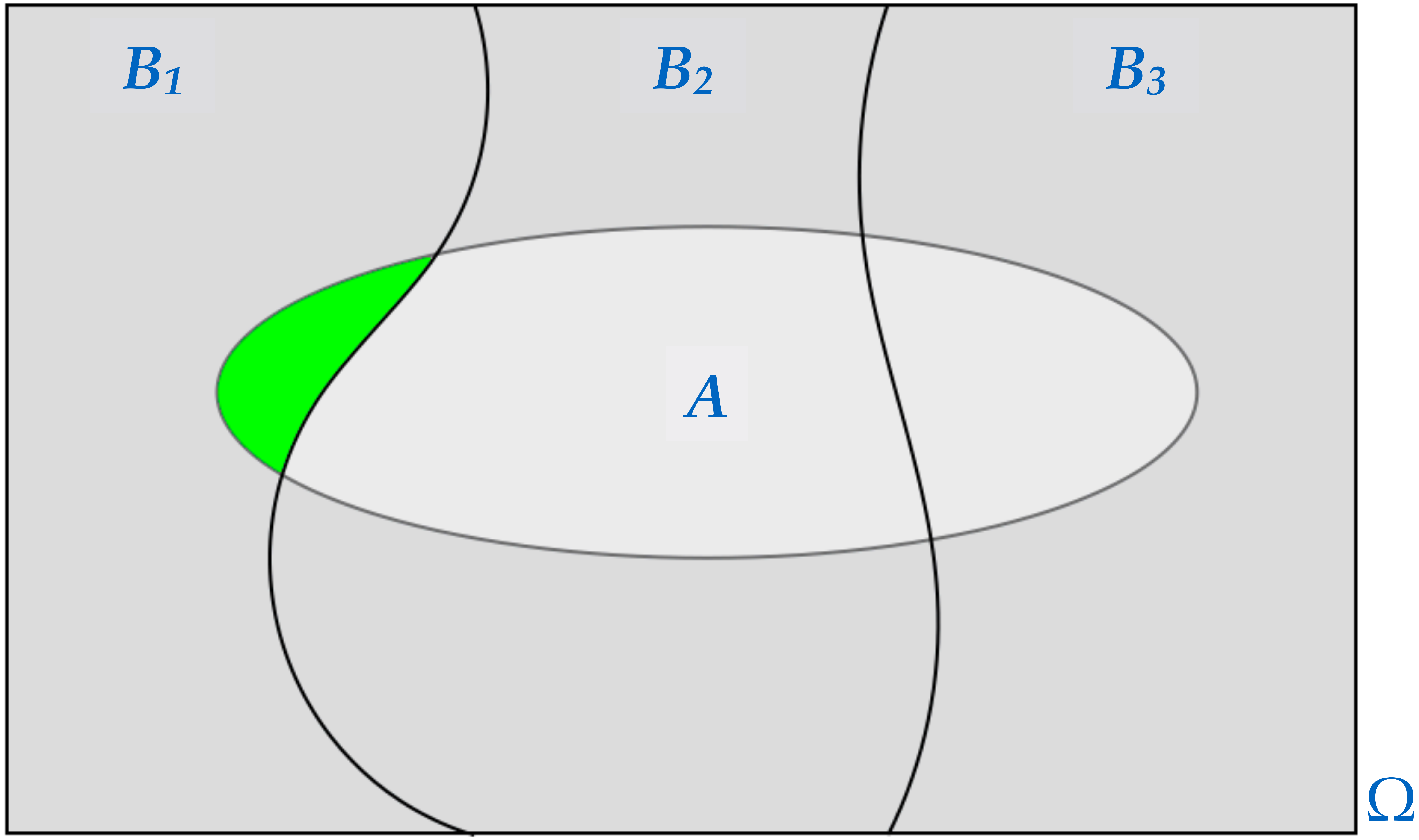
$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$



Total Probability

B_1, B_2, and B_3 are disjoint and their union is Ω

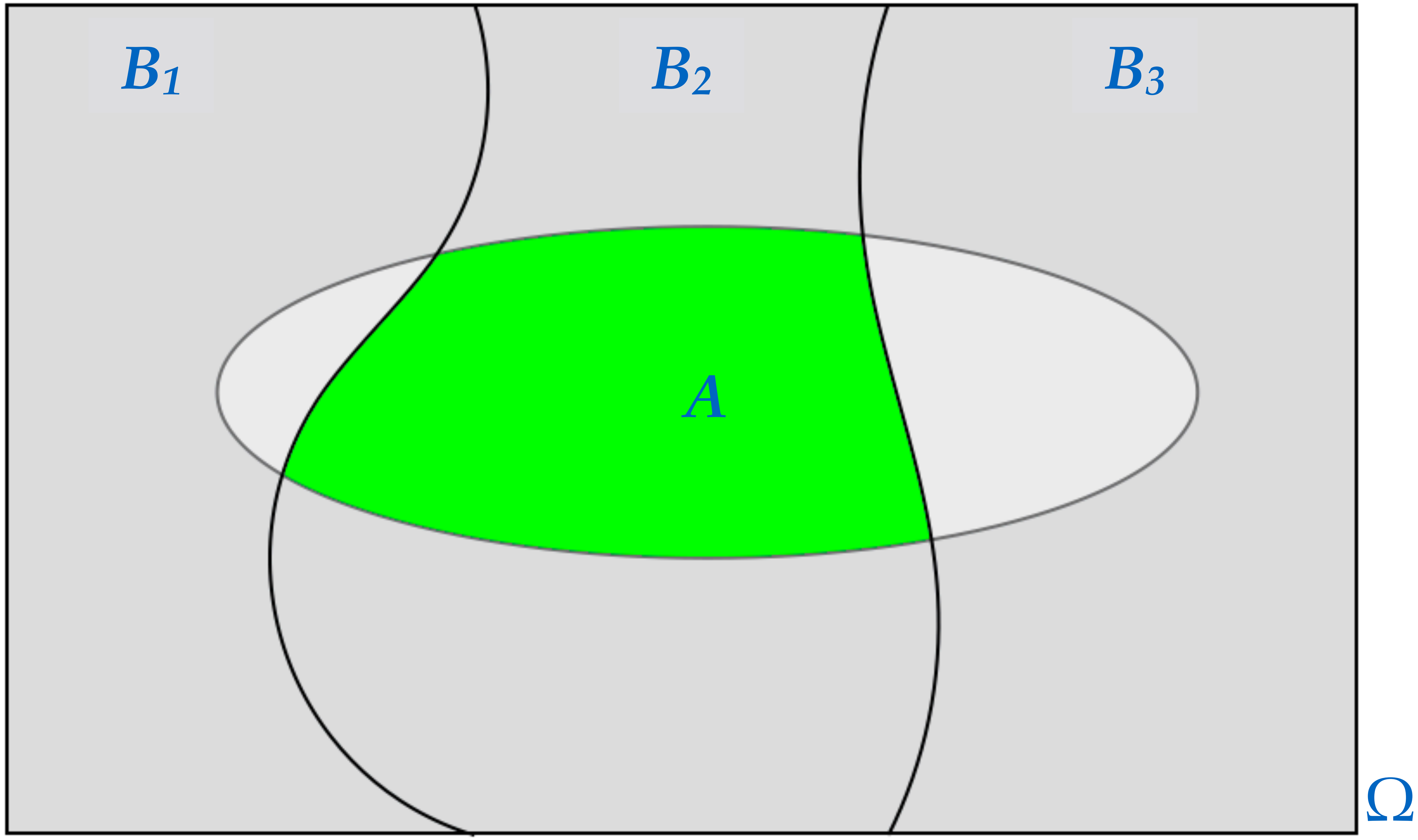
$$P(A) = \sum_{i=1}^{n} \boxed{P(A \cap B_i)}$$



Total Probability

B_1, B_2, and B_3 are disjoint and their union is Ω

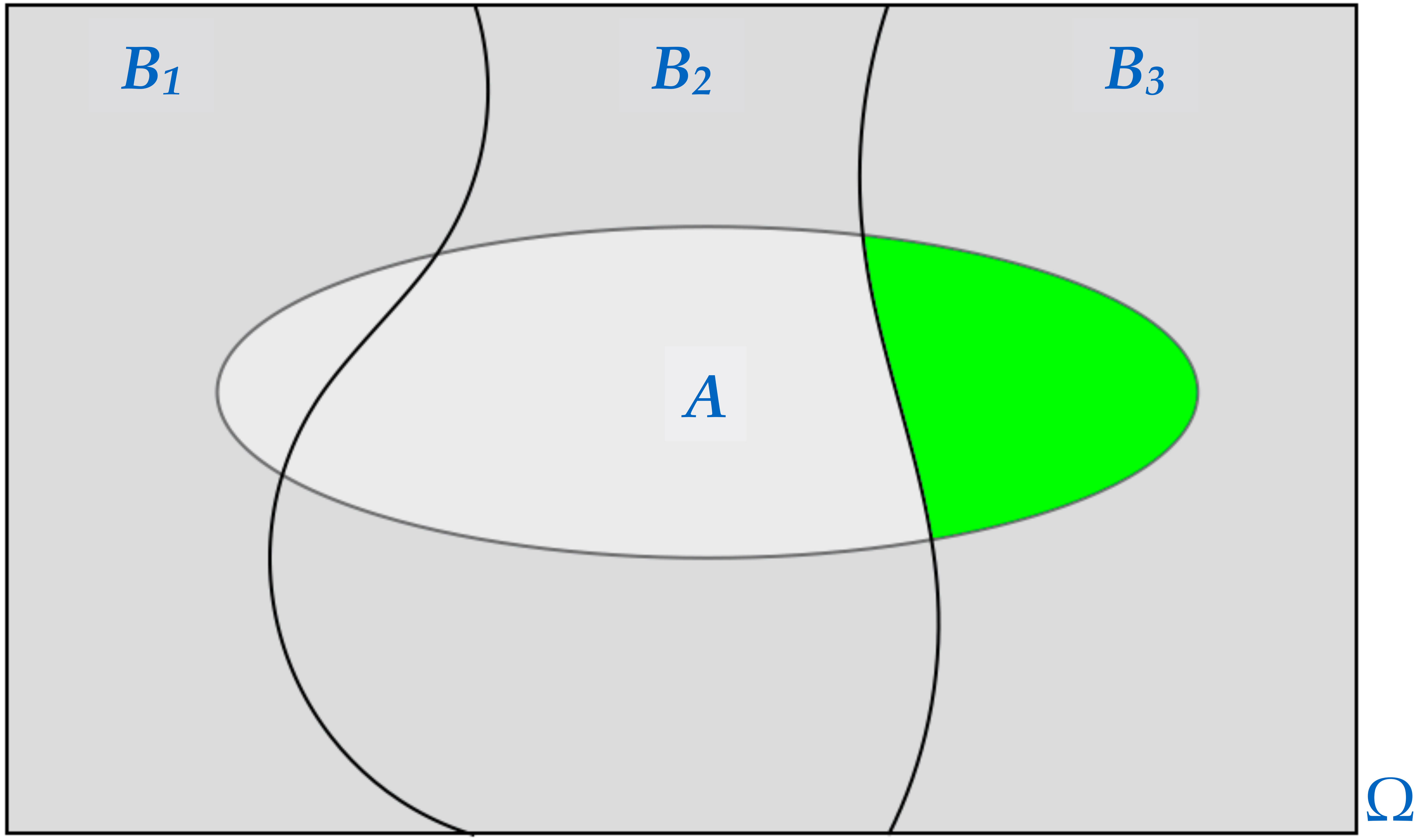
$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$



Total Probability

B_1, B_2, and B_3 are disjoint and their union is Ω

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$



Total Probability

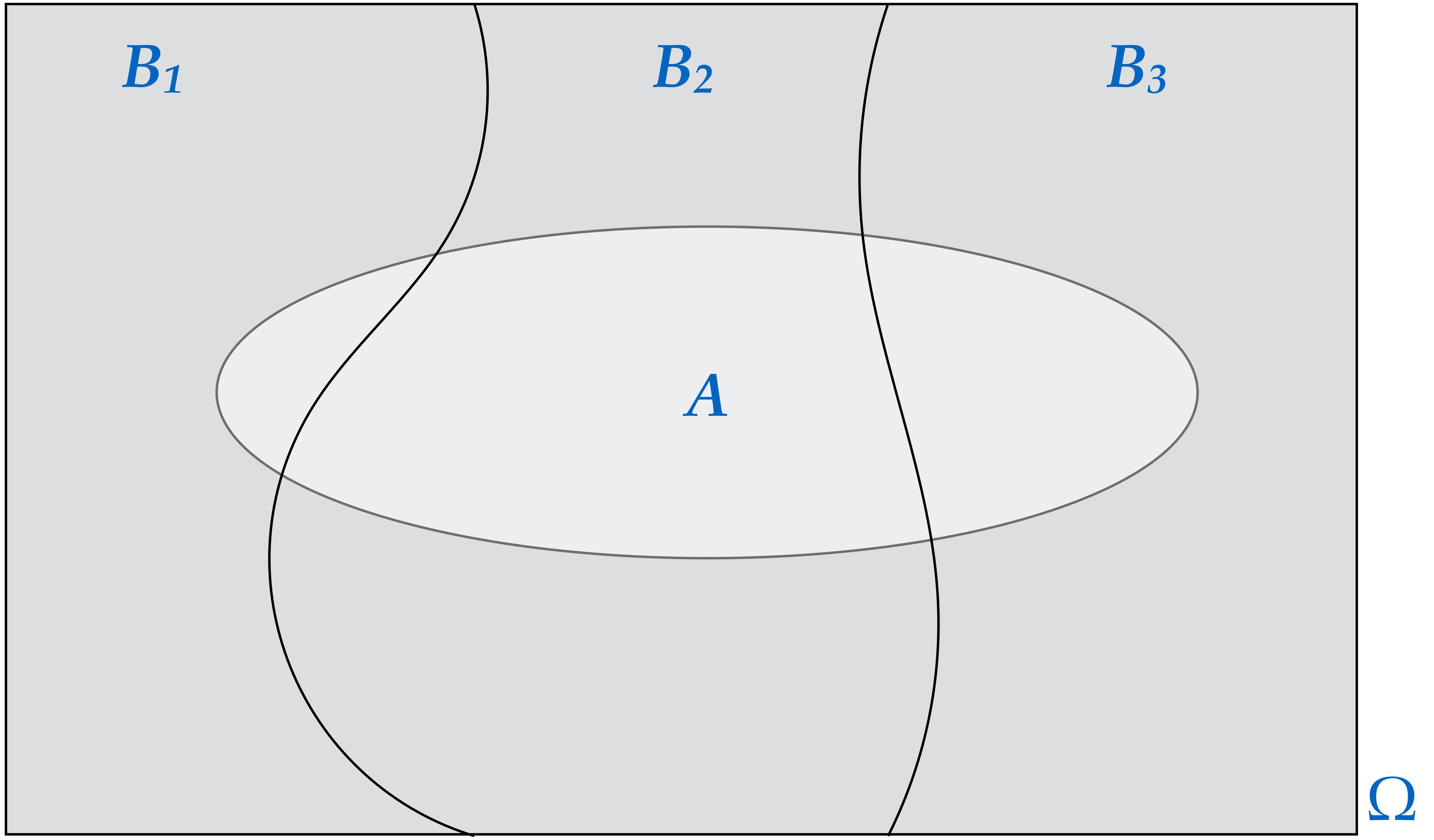
B_1, B_2, and B_3 are disjoint and their union is Ω

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

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

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Law of total probability

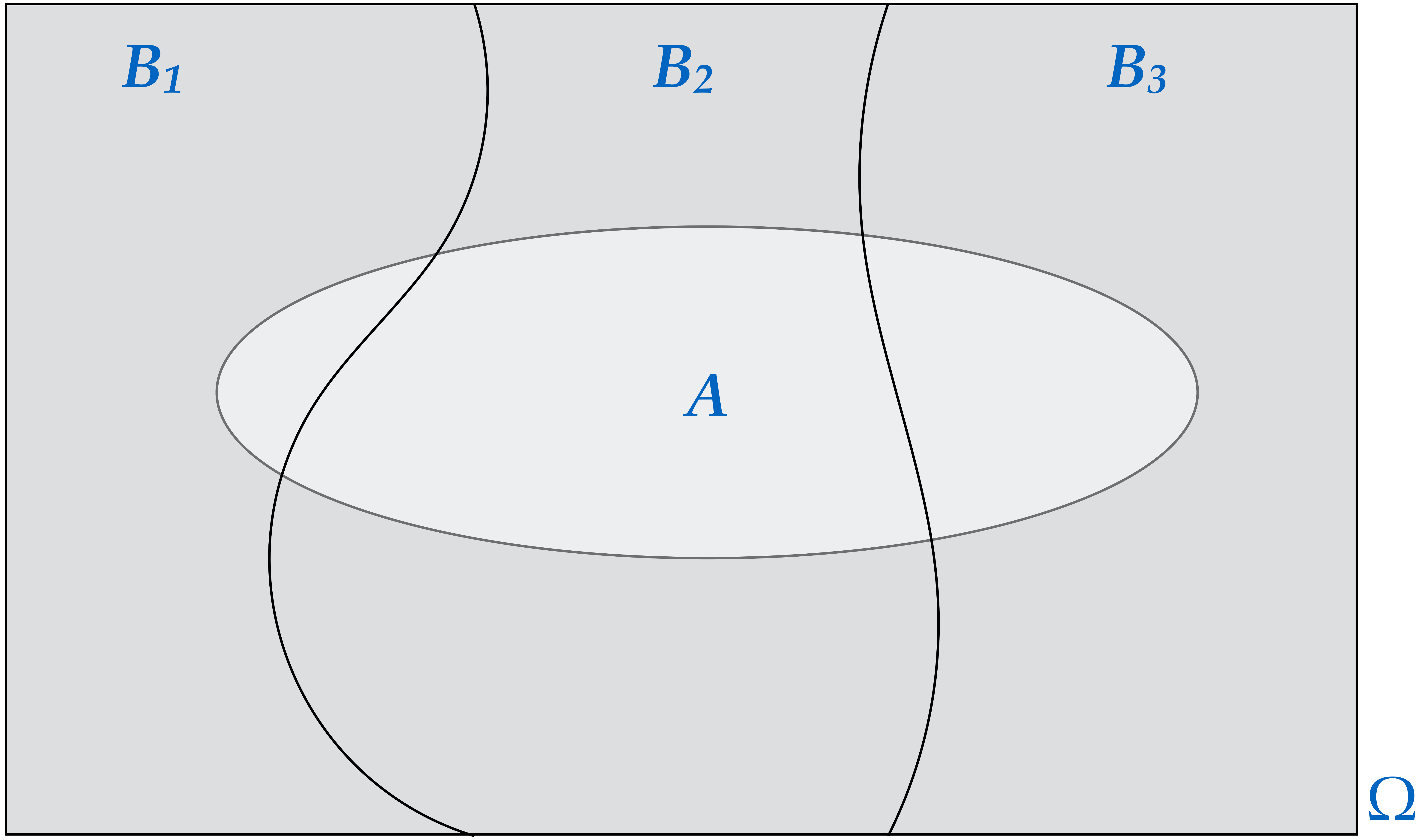


Bayes' Theorem

Bayes' Theorem

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

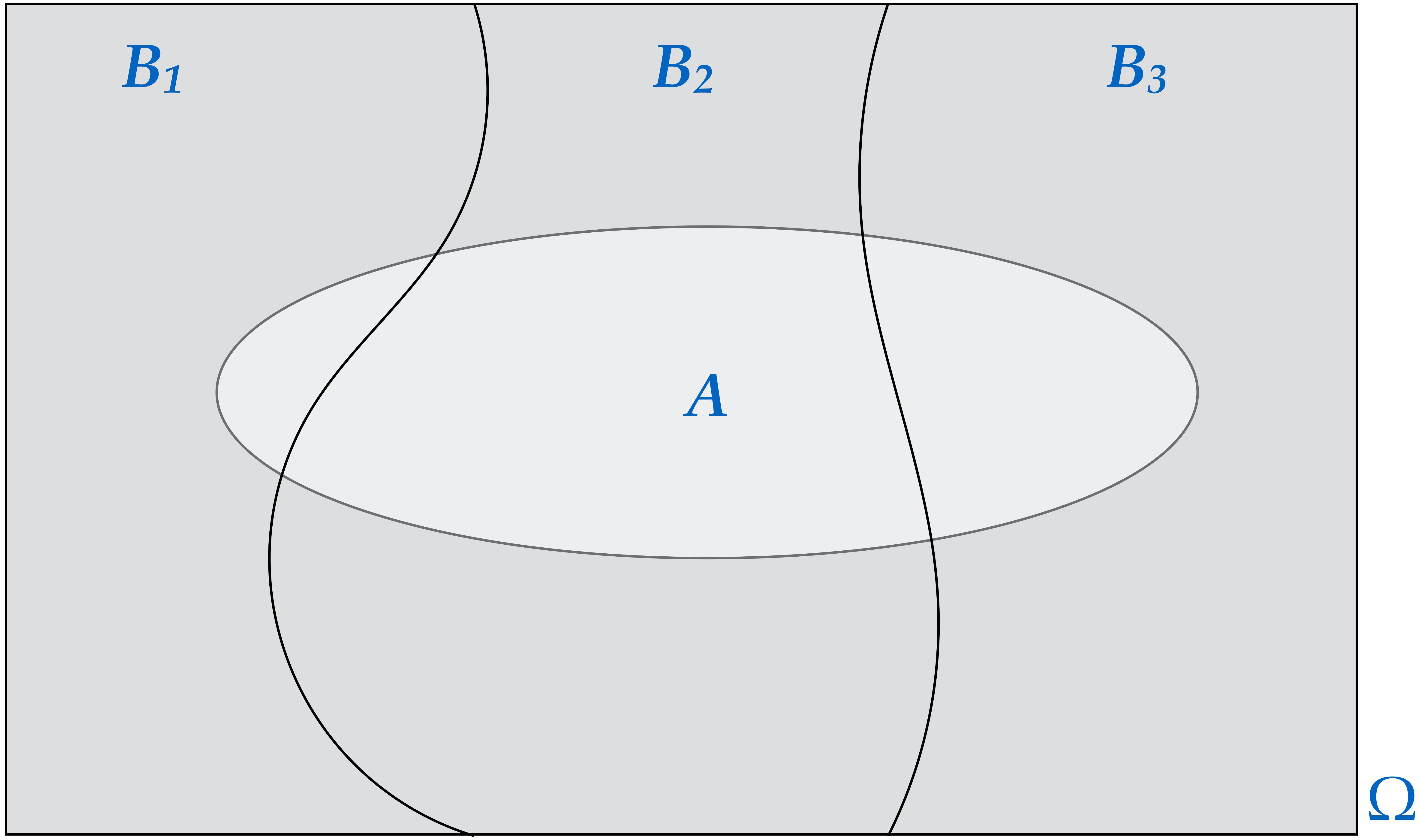
$$P(A|B_j) = \frac{P(A \cap B_j)}{P(B_j)}$$



Bayes' Theorem

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

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

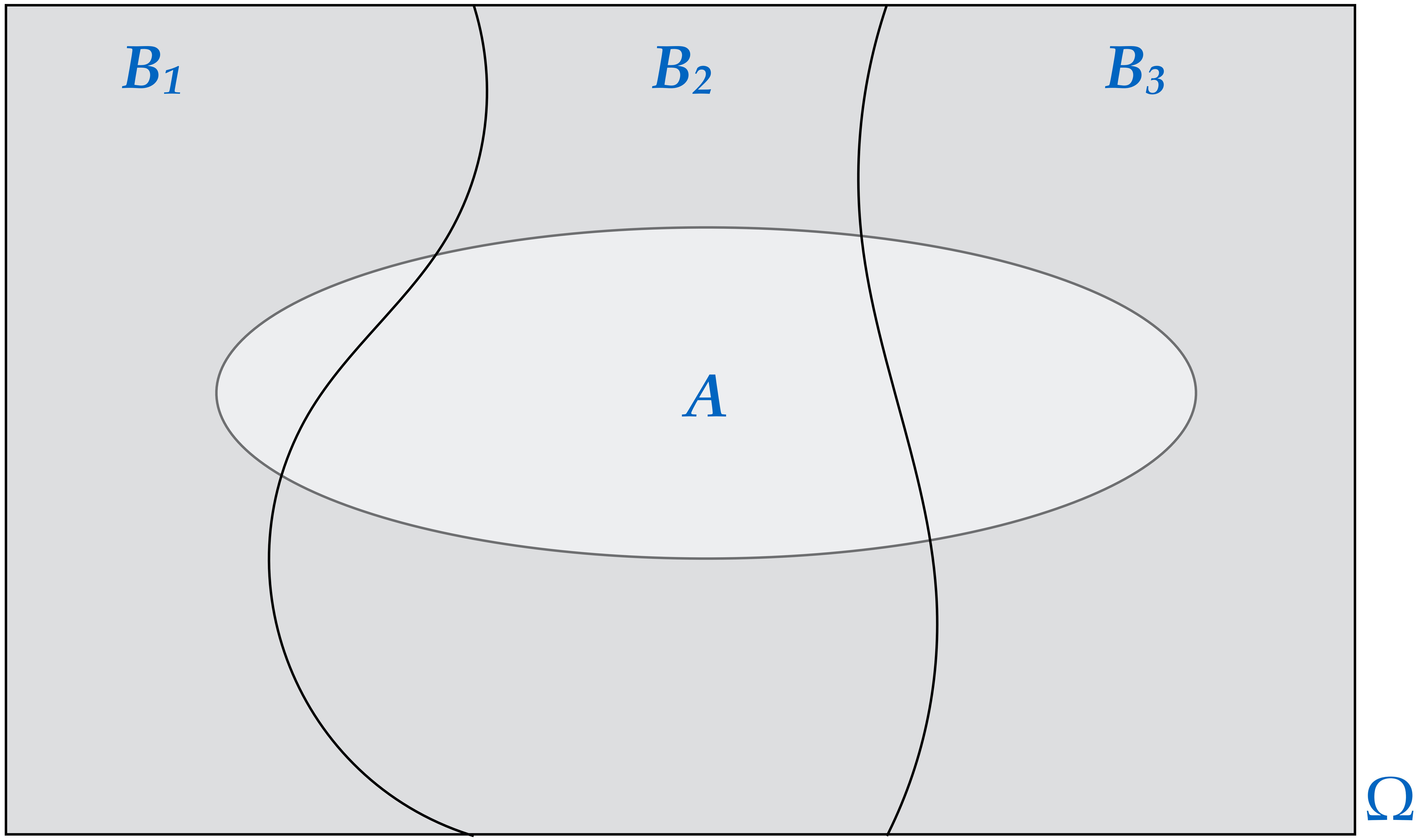


Bayes' Theorem

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

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

$$P(B_j|A)P(A) = P(A|B_j)P(B_j)$$

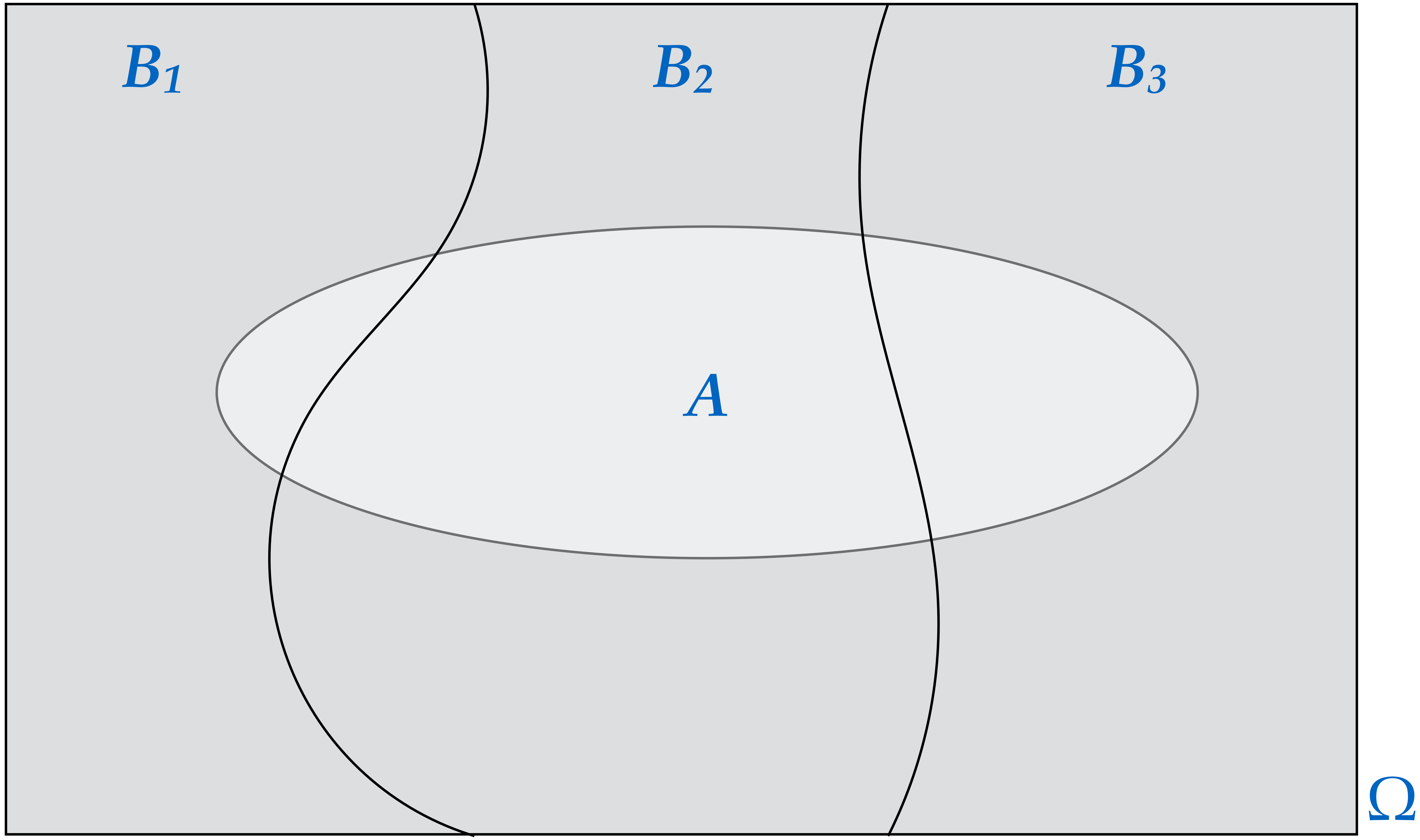


Bayes' Theorem

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

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

$$P(B_j|A)P(A) = P(A|B_j)P(B_j)$$



Bayes' Theorem

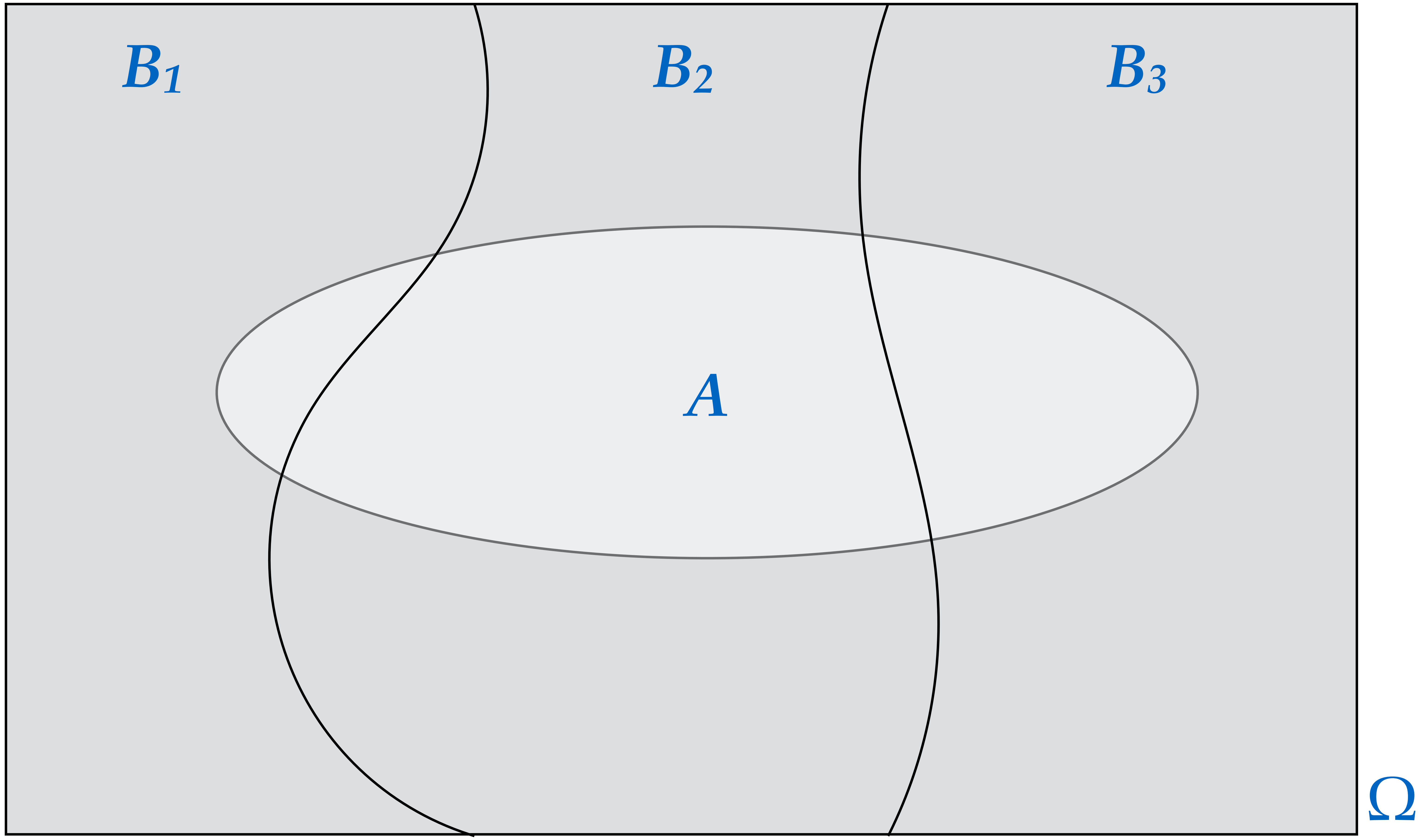
$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}$$

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

$$P(B_j|A)P(A) = P(A|B_j)P(B_j)$$

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

Bayes' theorem



Bayes' Theorem

Conditional probability, Likelihood

Posterior probability

Prior

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

Observations

Evidence (often fixed)

Probability Basics Recap

Experiment: a repeatable procedure with well-defined possible **outcomes** (ω)

Sample space: the set of all possible outcomes, e.g. $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$

Event: a subset of the sample space, e.g. $A=\{\omega_1, \omega_3\}$

Probability Basics Recap

Union ($A \cup B$): either A or B (or both) occur

Intersection ($A \cap B$): A and B both occur

Complement (A^c): A does *not* occur

Subset ($A \subset B$): events in A are contained in B

Disjoint ($A \cap B = \varnothing$): no overlap

Probability Basics Recap

Probability function: a function giving the probability for each outcome/event

- The probability of an event E is $P(E) = \sum_{\omega \in E} P(\omega)$
- $0 \leq P(A) \leq 1; \quad P(\Omega) = 1$

Conditional Probability: $P(A \mid B) = P(A \cap B) / P(B)$

- "probability of A given B" has occurred

Total Probability: if B_1, B_2, and B_3 are disjoint and their union is Ω:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

Independent events: $P(A \mid B) = P(A)P(B)$

Probability Basics Recap

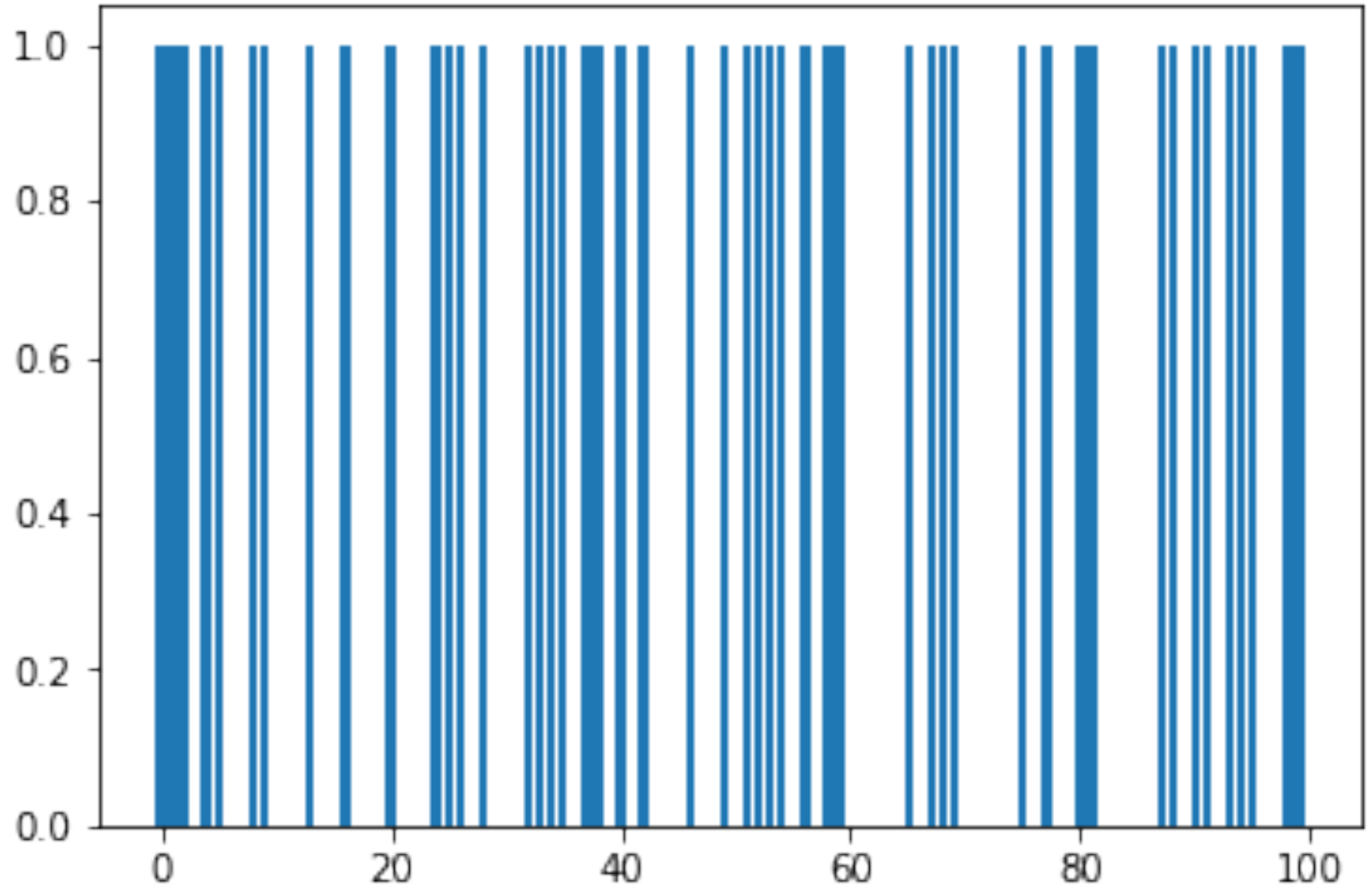
Bayes Theorem/Rule: $P(A \mid B) = P(B \mid A) \cdot P(A) \,/\, P(B)$

- allows you to find $P(A \mid B)$ from $P(B \mid A)$, i.e. to "invert" conditional probabilities
- often compute the denominator $P(B)$ using the law of total probability

Probability – Empirical Validation

Empirical Validation – Code ⇔ Math

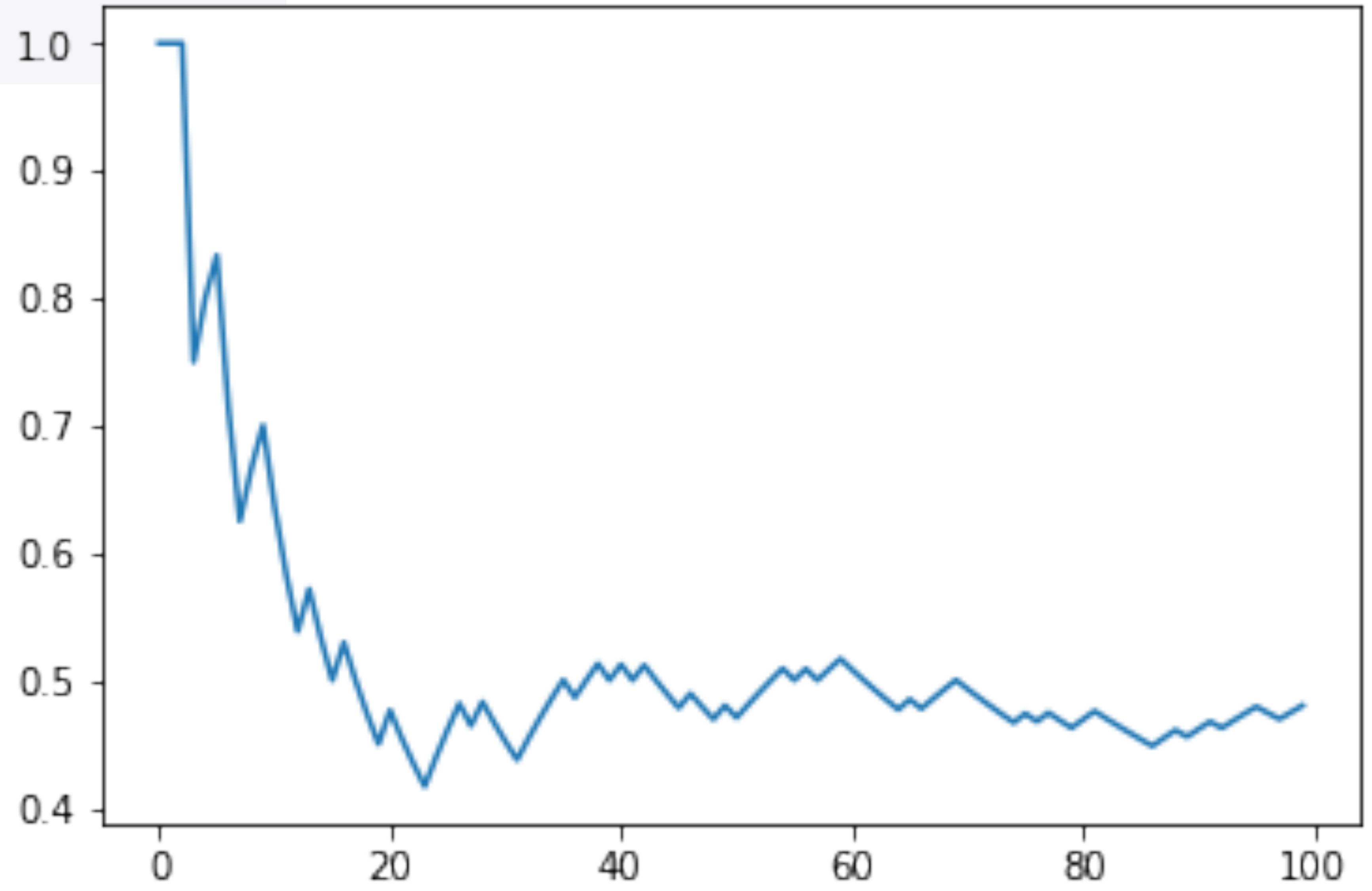
```
# simulate a bunch of coin tosses
import numpy as np
import matplotlib.pyplot as plt
n = 100 # number of coin tosses
x = (np.random.random(n) < 1/2)
plt.bar(np.arange(n),x)
```



Empirical Validation – Code ⇔ Math

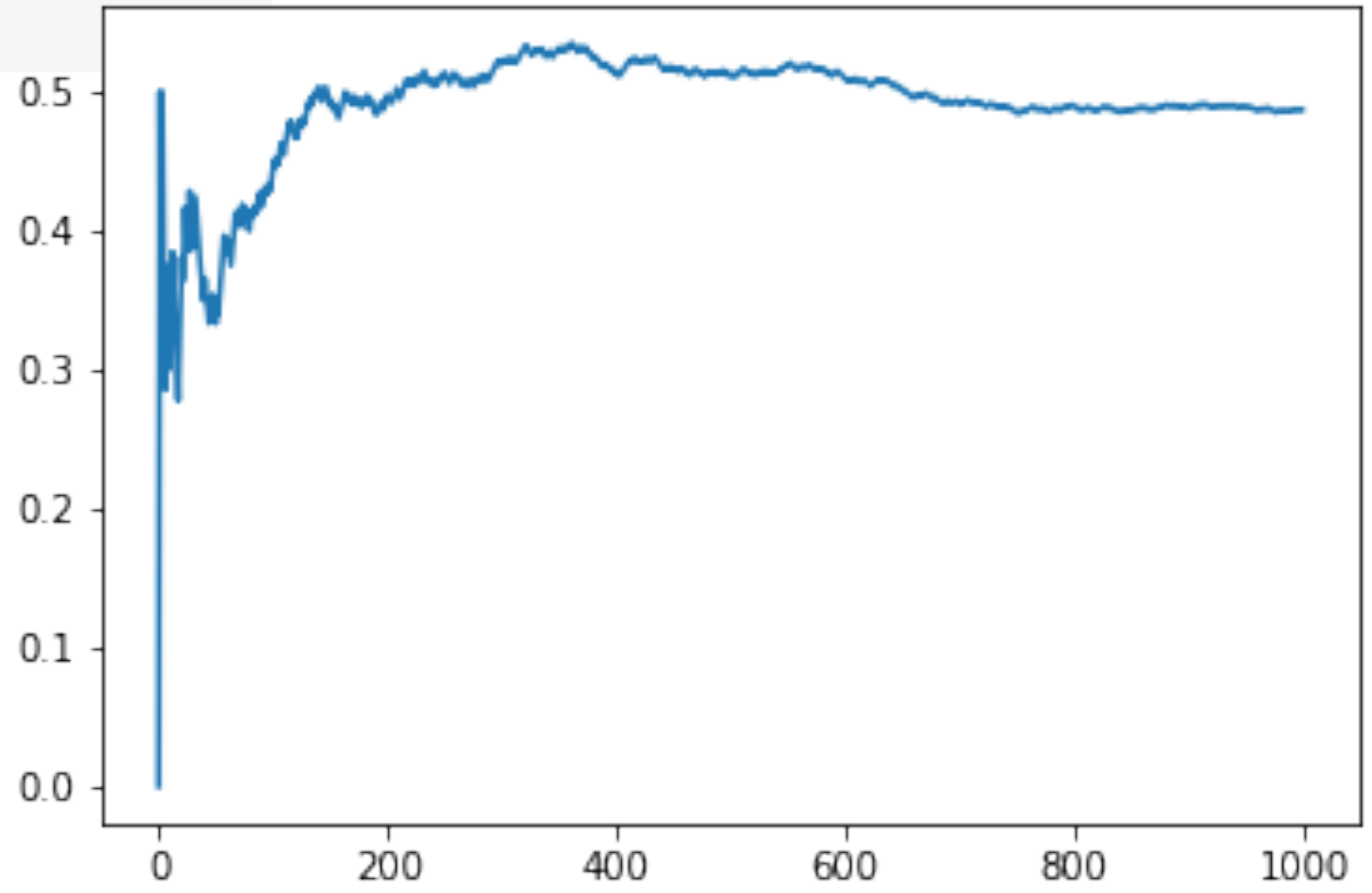
```
# simulate a bunch of coin tosses
import numpy as np
import matplotlib.pyplot as plt
n = 100 # number of coin tosses
x = (np.random.random(n) < 1/2)
cs = np.cumsum(x) * (1.0/np.arange(1,n+1))
plt.plot(cs)
```

$$C_x(k) = \frac{1}{k}\sum_{i=1}^{k} x_k$$

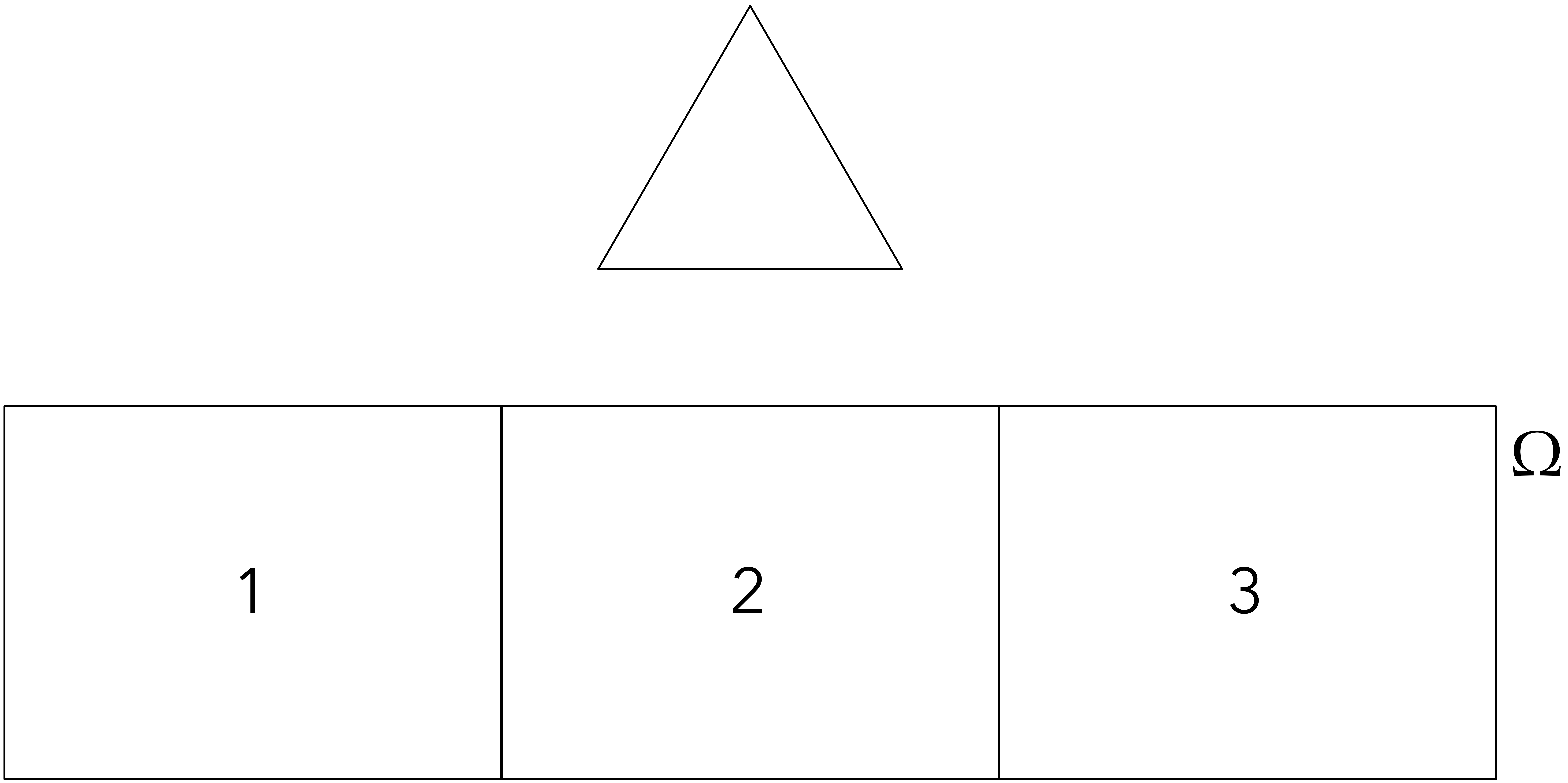


Empirical Validation – Code ⇔ Math

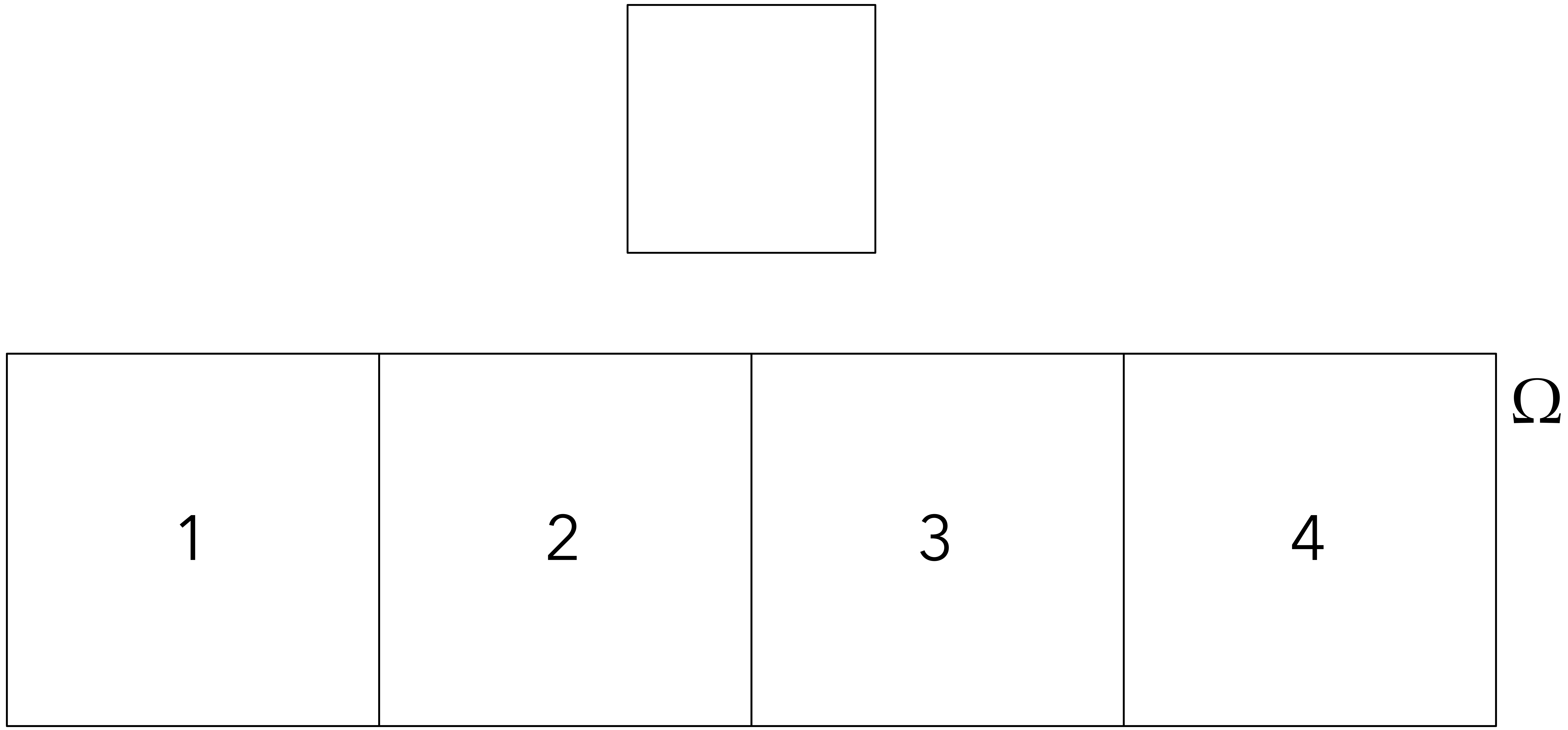
```python
# simulate a bunch of coin tosses
import numpy as np
import matplotlib.pyplot as plt
n = 1000 # number of coin tosses
x = (np.random.random(n) < 1/2)
cs = np.cumsum(x) * (1.0/np.arange(1,n+1))
plt.plot(cs)
```



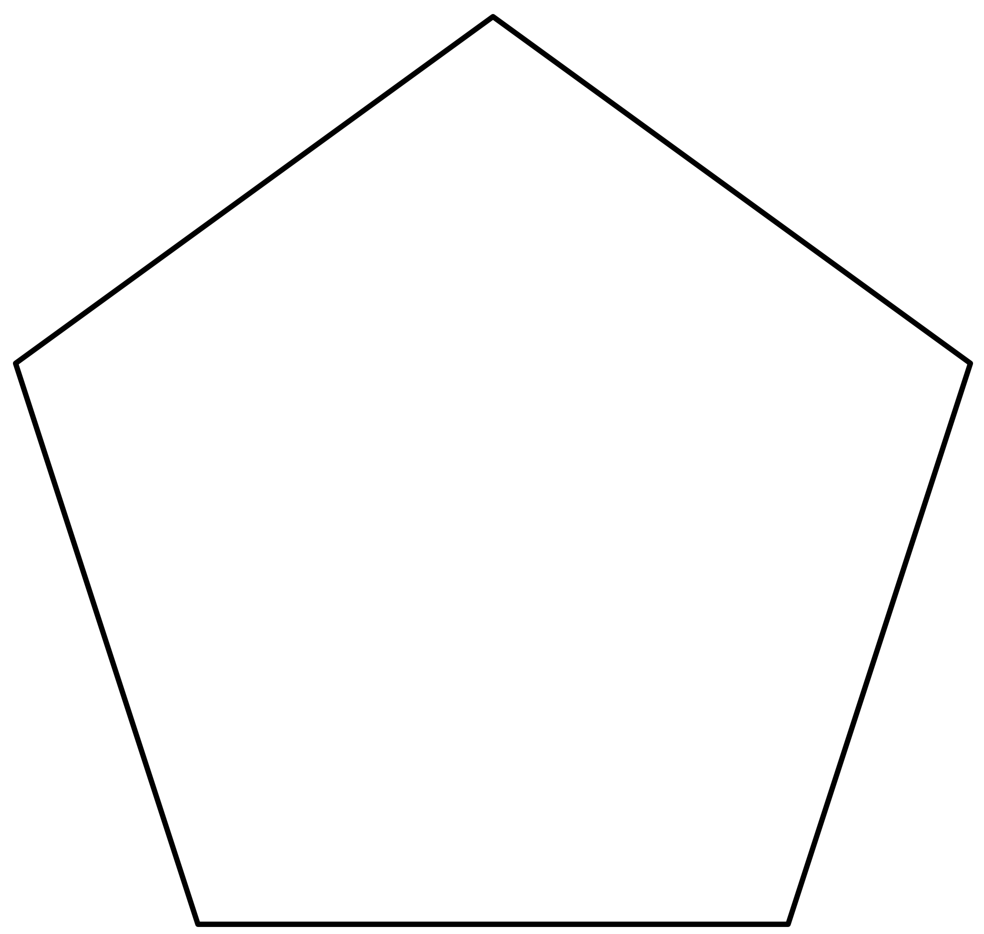
Discrete Probability Spaces

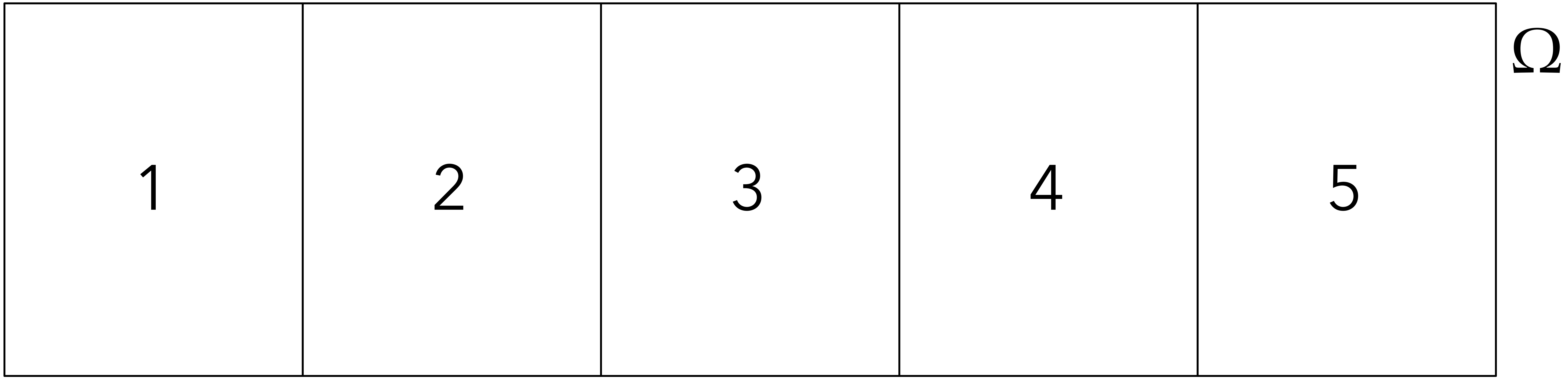


Discrete Probability Spaces

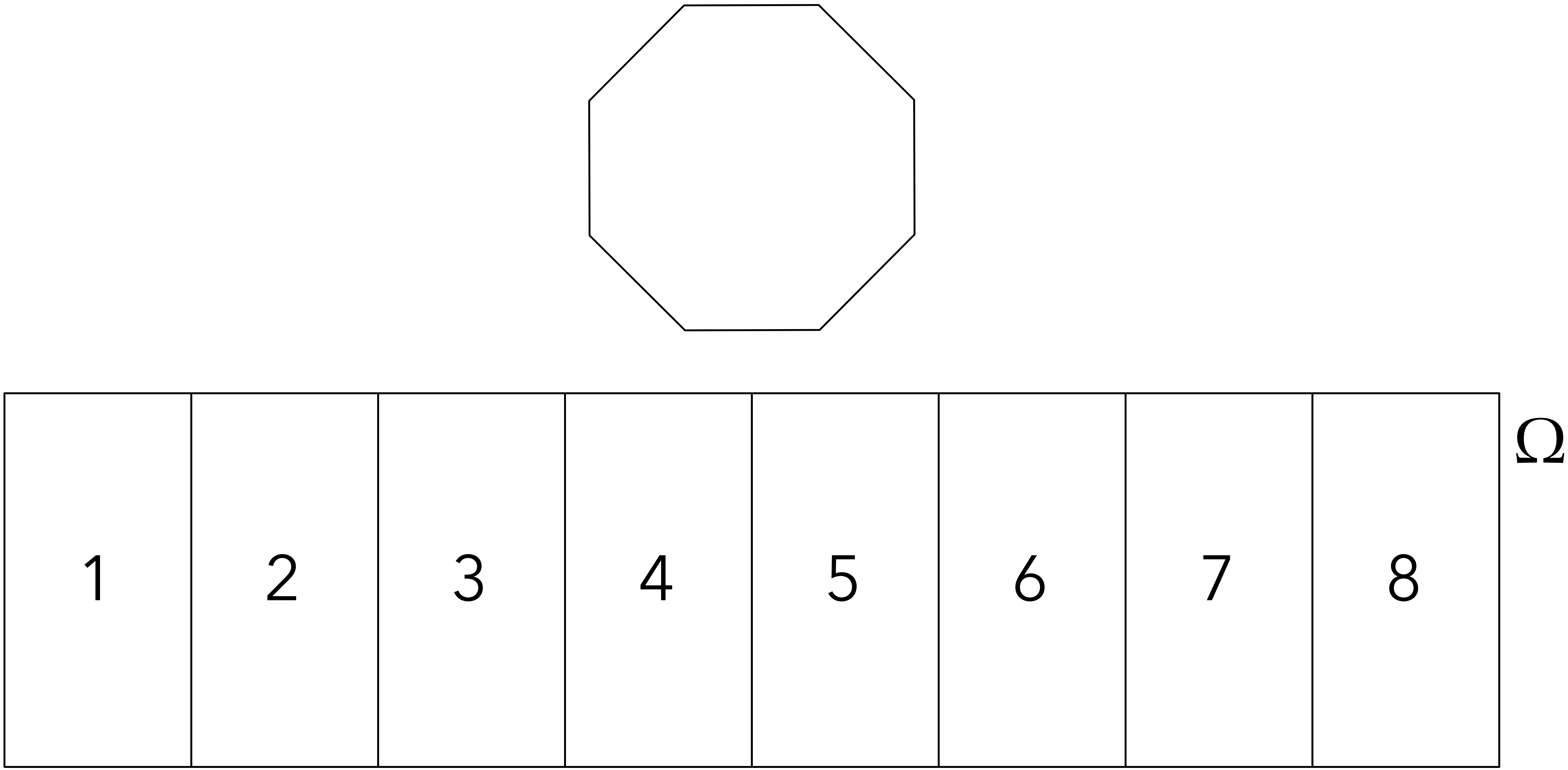


Discrete Probability Spaces

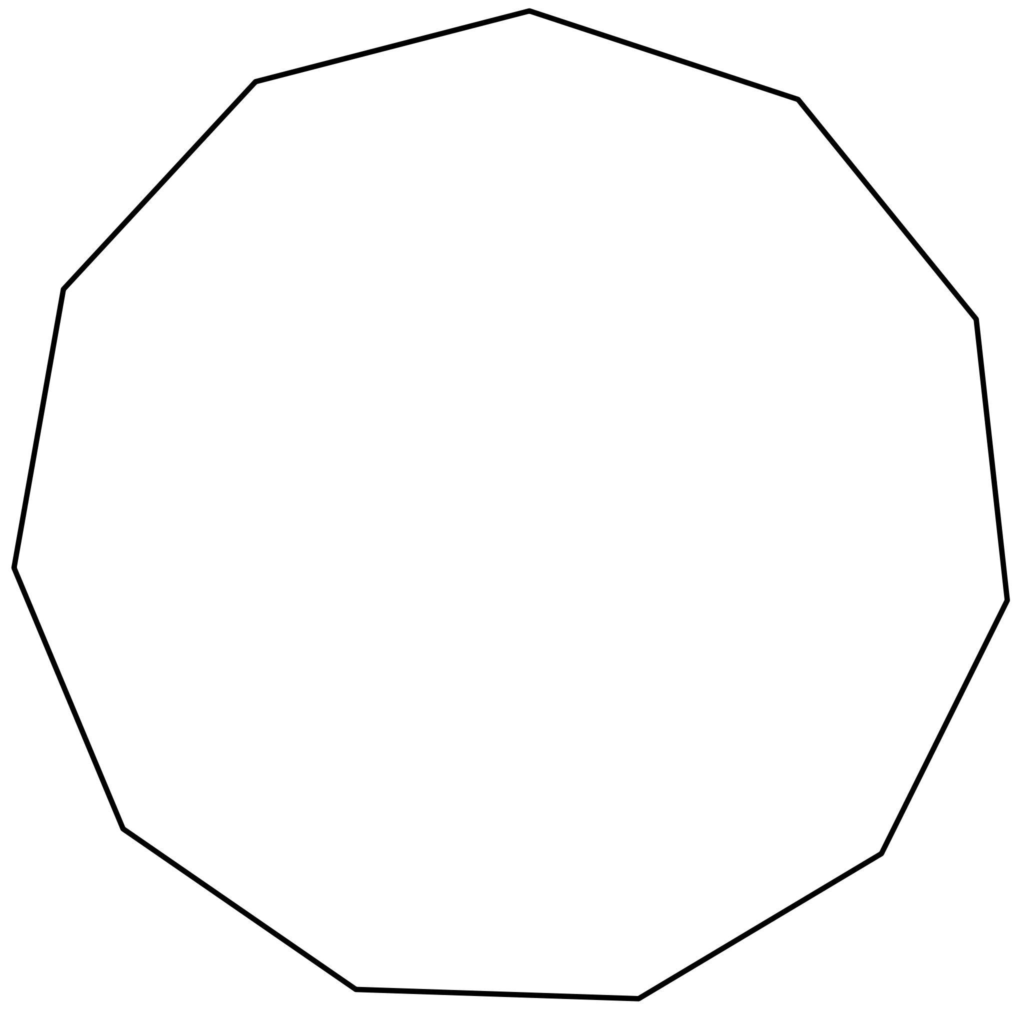




Discrete Probability Spaces



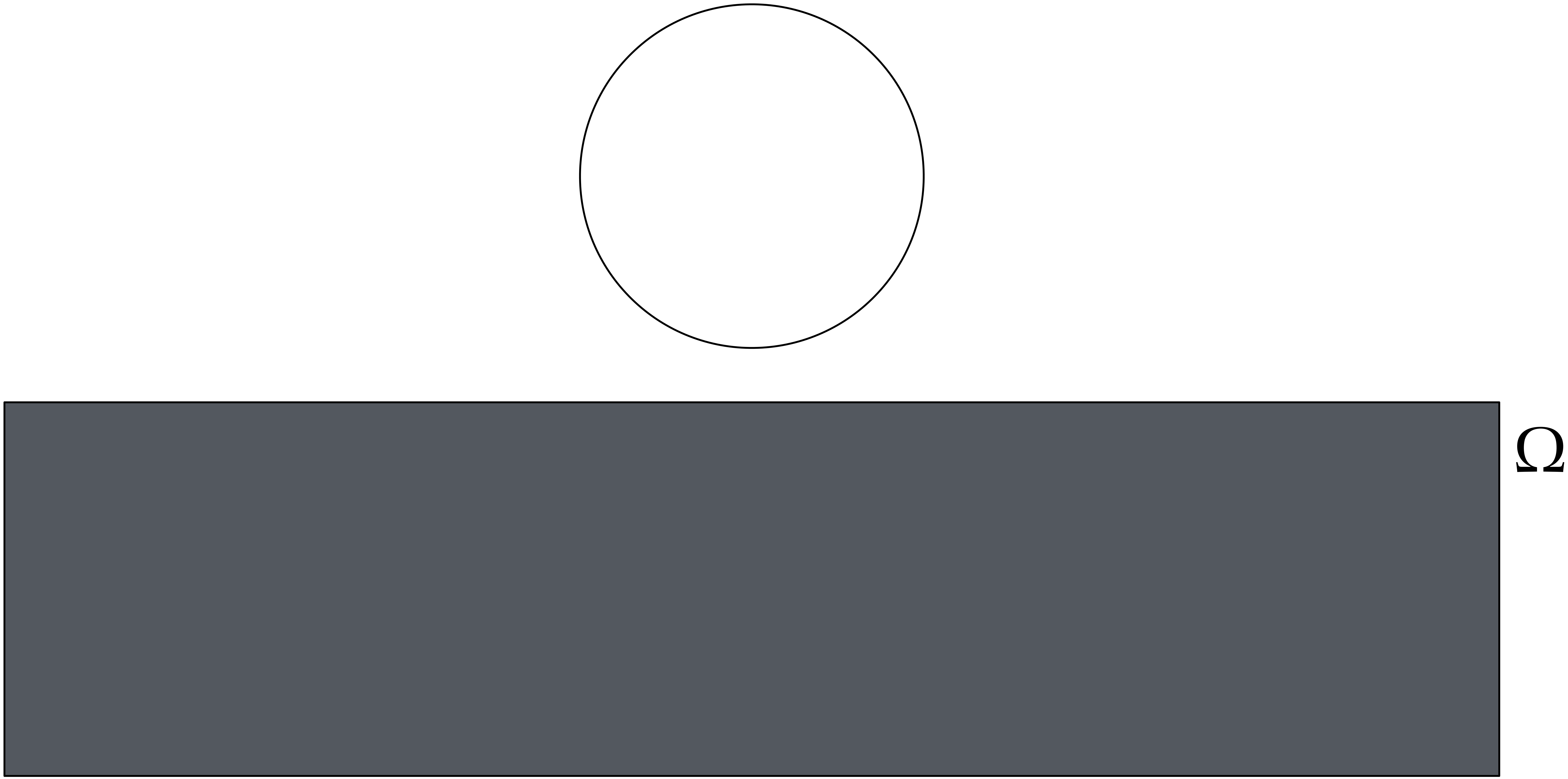
Discrete Probability Spaces



1	2	3	4	5	6	7	8	9	10	11

Ω

What happens in the limit?



Continuous
Random Variables

Random Variables – discrete vs. continuous

Random Variable: assigns a number to each outcome

- “discrete”: **countable/listable** set of outcomes, e.g.,
    - value on a die, number of left-handers
- “continuous”: **measurable**, infinite possible outcomes, e.g.,
    - height of a person, amount of rainfall

Continuous Random Variables

Probability density function (*pdf*) of X:

- **not** a probability; have to integrate to obtain probability:

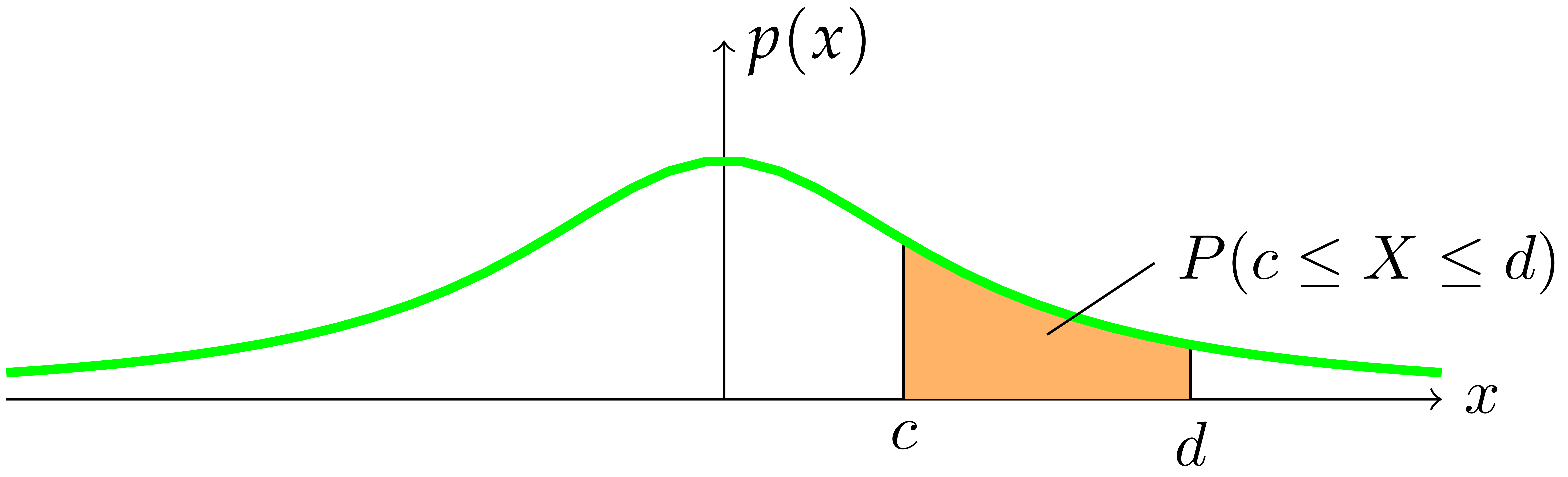
$$P(a \leq X \leq b) = \int_a^b p(x)\mathrm{d}x$$

  - the probability that X exactly equals some value is zero!

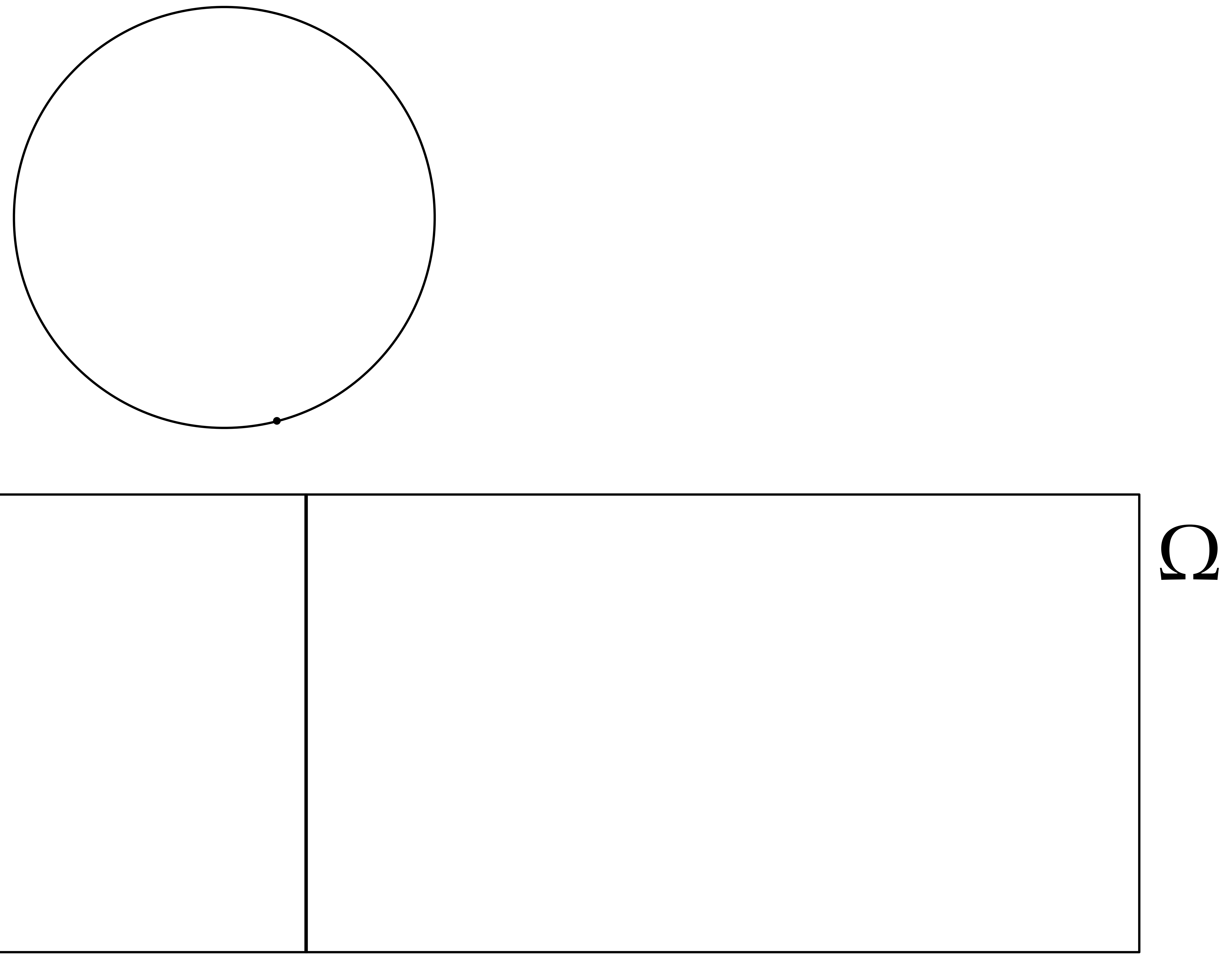
- integrates to one: $\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$
- no restriction that $p(x) < 1$

Probability density function

$$\boxed{P(c \leq X \leq d) = \text{area under the graph between } c \text{ and } d.}$$



Probability of landing at a point is 0



Probability of landing within a range