Computer
Science

# CSC196: Analyzing Data

## Probability

Jason Pacheco and Cesim Erten

# Outline

- Random Events
- Counting Sample Points
- Random Events and Probability
- Conditional Probability & Independent Events

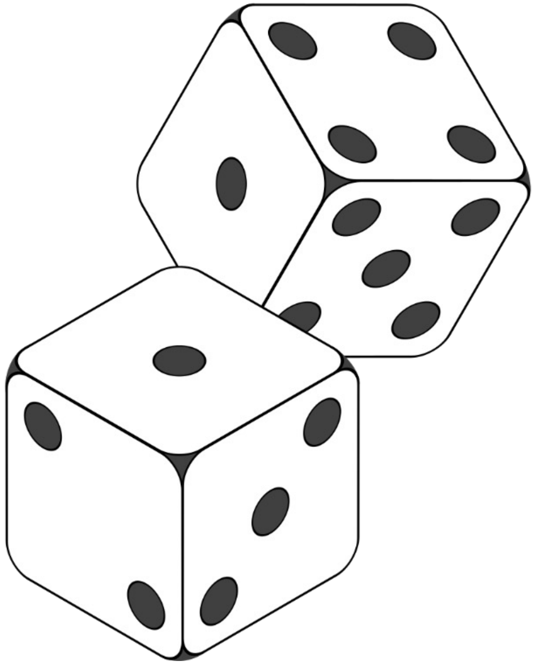
# Outline

- Random Events
- Counting Sample Points
- Random Events and Probability
- Conditional Probability & Independent Events

# Random Events and Probability

***Suppose we roll <u>two fair dice</u>…***

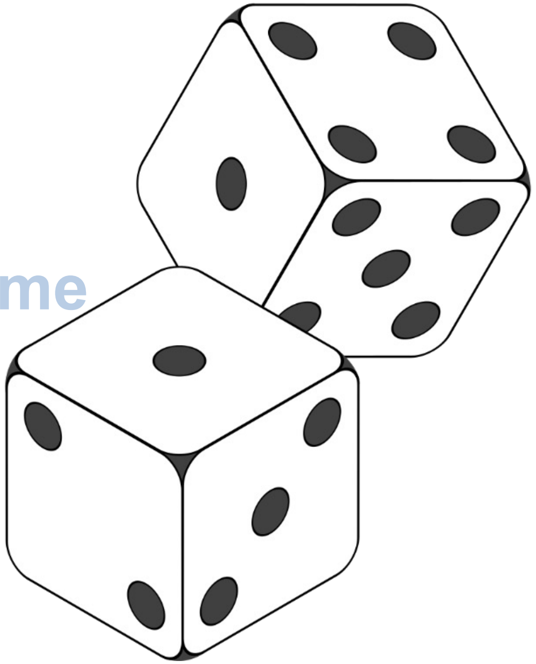
- What are the possible outcomes?
- What is the *probability* of rolling **even** numbers?
- What is the *probability* of rolling **odd** numbers?
- If one die rolls 1, then what is the probability of the second die also rolling 1?
- How to mathematically formulate outcomes and their probabilities?



*…this is an **experiment** or **random process**.*

# Random Events and Probability

**Definition** An **outcome** is a possible result of an experiment or trial, and the collection of all possible outcomes is the **sample space** of the experiment,



Outcome

**Example** $\underbrace{(1,1), (1,2), \ldots, (6,1), (6,2), \ldots, (6,6)}_{\text{Sample Space}}$

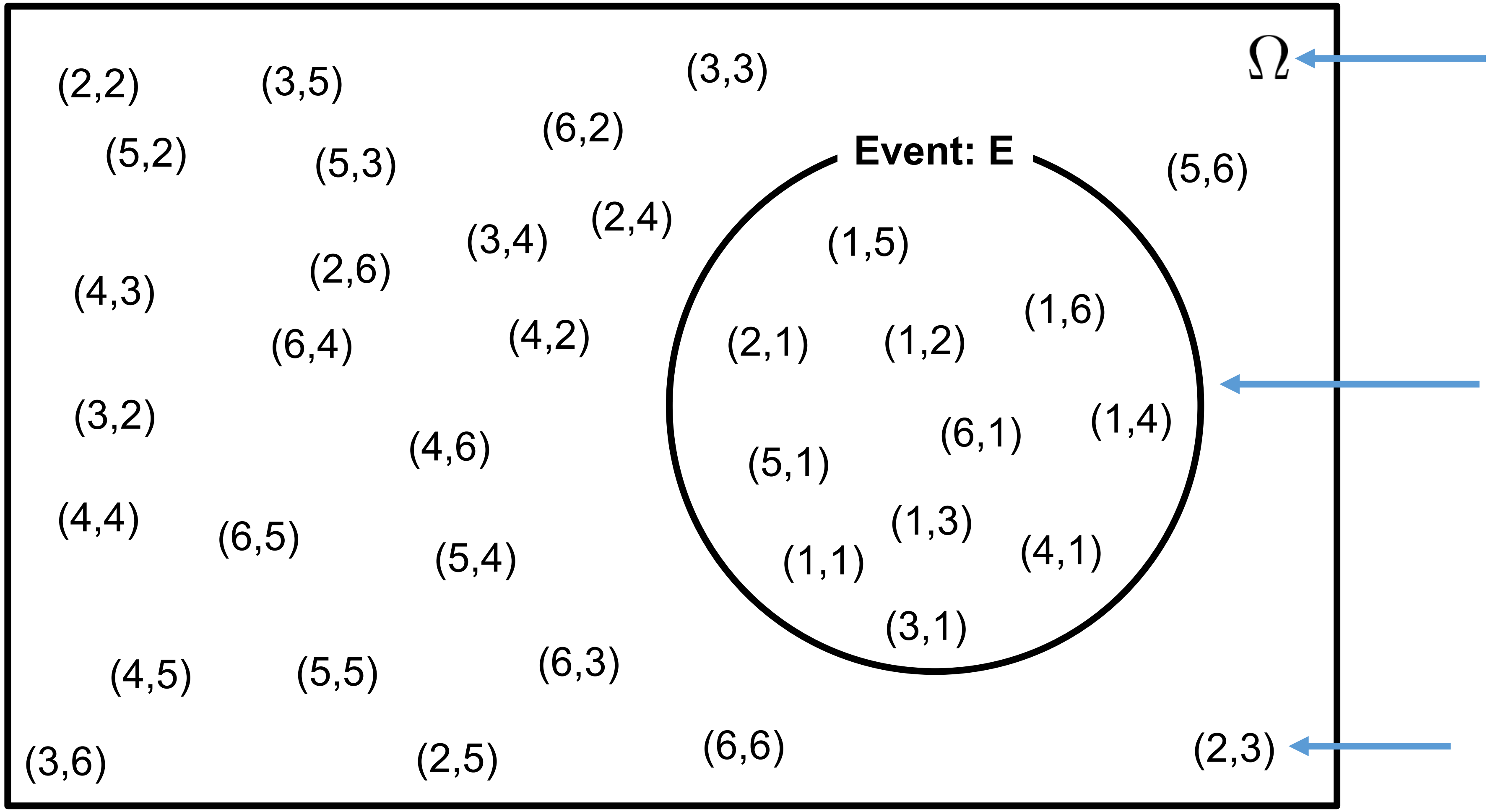
**Definition** An **event** is a *set* of outcomes (a subset of the sample space),

**Example Event** Roll at least a single 1

$\{(1,1), (1,2), (1,3), \ldots, (1,6), \ldots, (6,1)\}$

# Random Events and Probability

*Can formulate / visualize as a space of outcomes and events*



Sample space consists of all possible outcomes

Event is a collection of outcomes and a subset of the sample space
$E \subseteq \Omega$

Each outcome is a sample $\omega \in \Omega$

# Random Events and Probability

**Some examples of events…**

- Roll even numbers,

$$E^{\text{even}} = \{(2,2),(2,4),\ldots,(6,4),(6,6)\}$$

- The sum of both dice is even,

$$E^{\text{sum even}} = \{(1,1),(1,3),(1,5),\ldots,(2,2),(2,4),\ldots\}$$

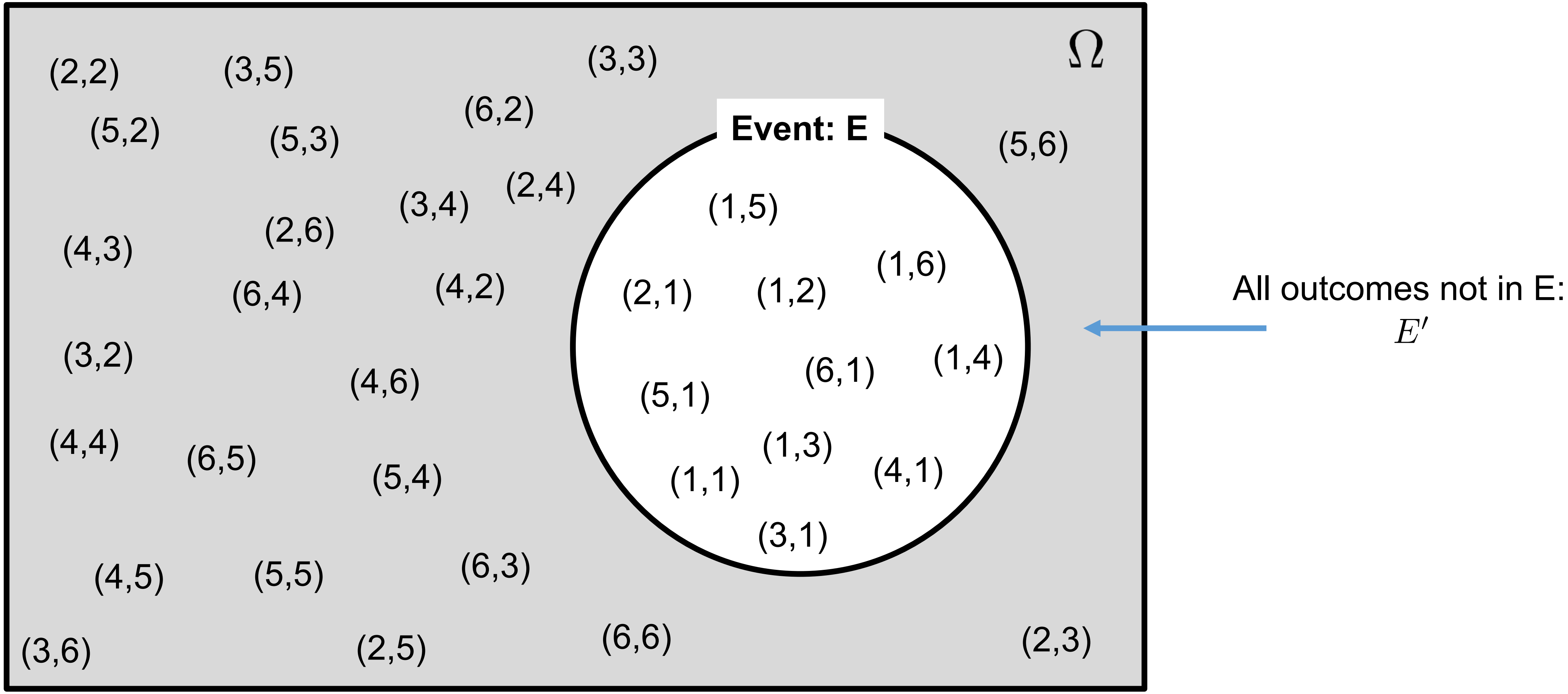
- The sum is greater than 12,

$$E^{\text{sum}>12} = \emptyset$$

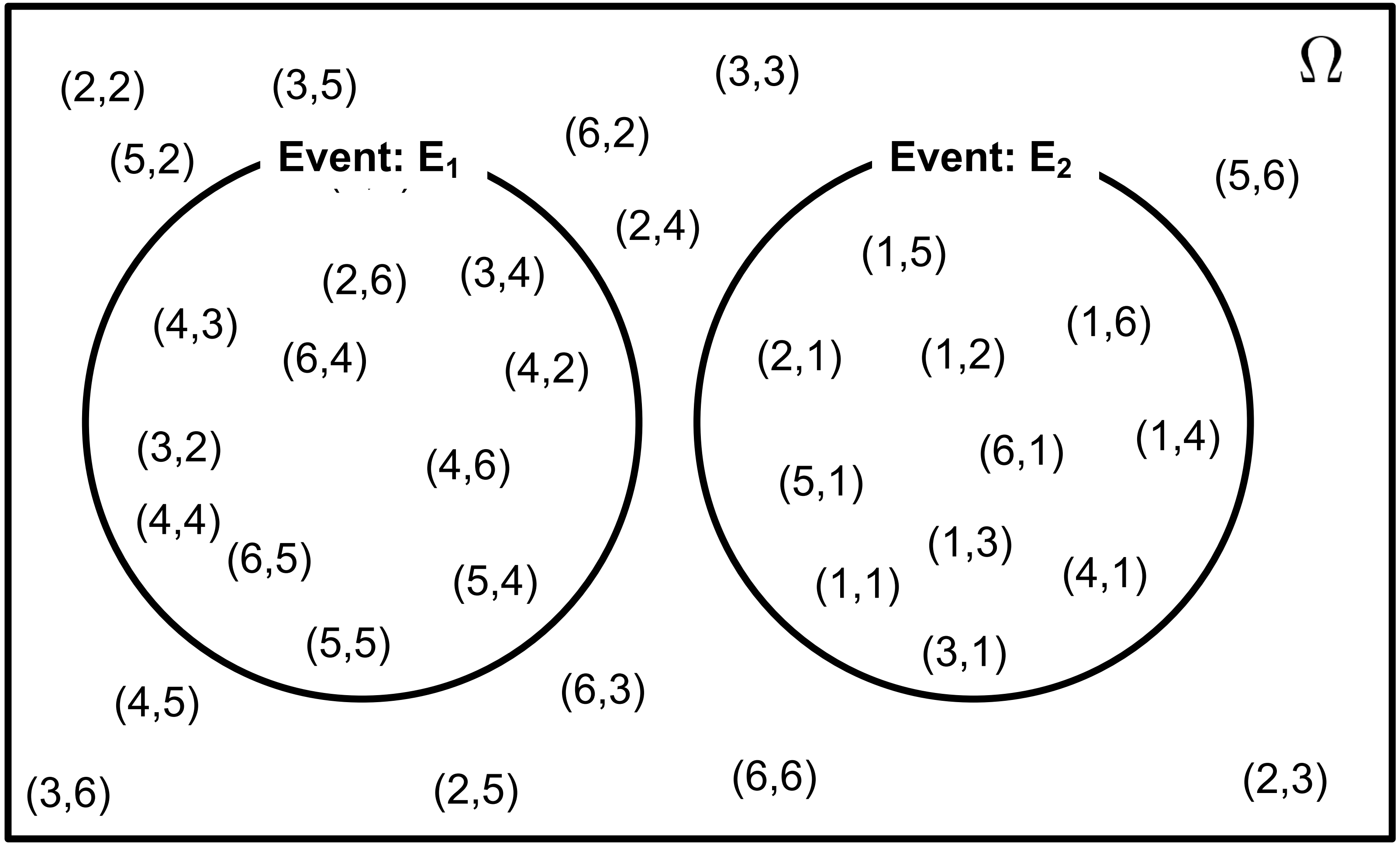
**We can reason about impossible outcomes**

# Random Events and Probability

**Definition** The **complement** of an event E, denoted E’, is the subset of all elements of Ω that are not in E.



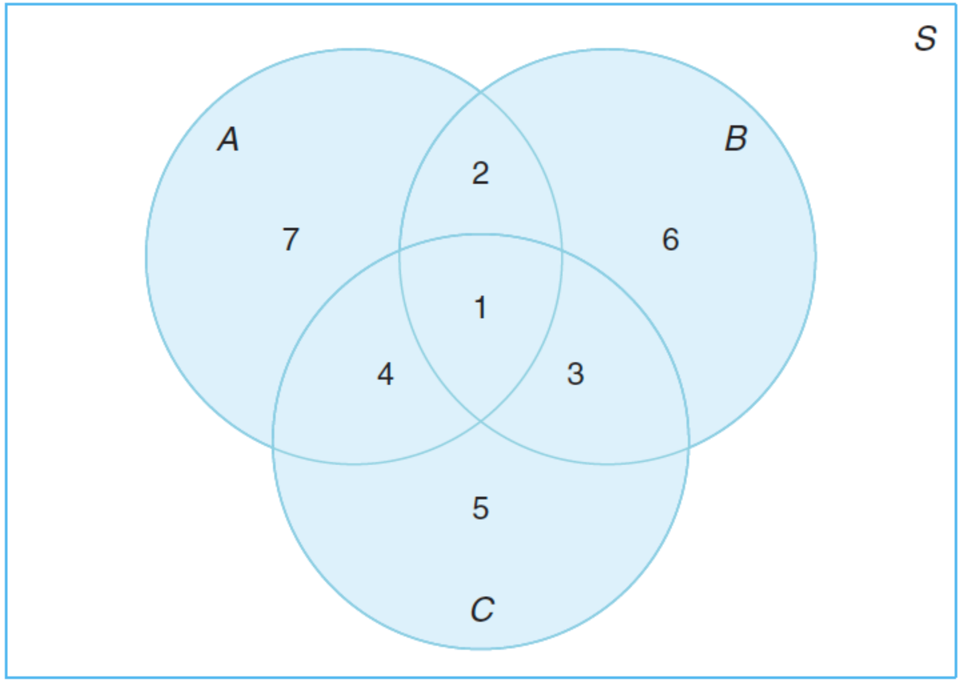
# Random Events and Probability



**Definition** Two events $E_1$ and $E_2$ are called **mutually exclusive** if their intersection is the empty set:

$$E_1 \bigcap E_2 = \emptyset$$

# Random Events and Probability



$$
\begin{aligned}
A \cup C &= \text{regions 1, 2, 3, 4, 5, and 7,}\\
B' \cap A &= \text{regions 4 and 7,}\\
A \cap B \cap C &= \text{region 1,}\\
(A \cup B) \cap C' &= \text{regions 2, 6, and 7,}
\end{aligned}
$$

# Random Events and Probability

***Two dice example:*** *If* $E_1, E_2 \in \mathcal{F}$ *where,*

$E_1$ : *First* die equals 1 $E_2$ : *Second* die equals 1

$$E_1 = \{(1,1),(1,2),\ldots,(1,6)\} \qquad E_2 = \{(1,1),(2,1),\ldots,(6,1)\}$$

*Then we must include (at least) the following events…*

| Operation | Value | Interpretation |
|---|---|---|
| $E_1 \cup E_2$ | $\{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,1)\}$ | Any die rolls 1 |
| $E_1 \cap E_2$ | $\{(1,1)\}$ | Both dice roll 1 |
| $E_1 - E_2$ | $\{(1,2),(1,3),(1,4),(1,5),(1,6)\}$ | First die rolls 1 only |
| $(E_1 \cup E_2)'$ | $\{(2,2),(2,3),\ldots,(2,6),(3,2),\ldots,(6,6)\}$ | No die rolls 1 |

# Outline

- Random Events
- Counting Sample Points
- Random Events and Probability
- Conditional Probability & Independent Events

# Counting Sample Points

**Multiplication Rule** If an operation can be performed $n_1$ ways, and if for each of these ways a second operation can be performed $n_2$ ways, then the two operations can be performed together in $n_1 n_2$ ways.

**Example** How many sample points are there in the sample space when a pair of dice is thrown once?

**Solution** First die can land face-up in $n_1=6$ ways. For each way, the second die can land face-up $n_2=6$ ways. So, the pair of dice can land in:

$$n_1 n_2 = (6)(6) = 36 \qquad \text{possible ways.}$$

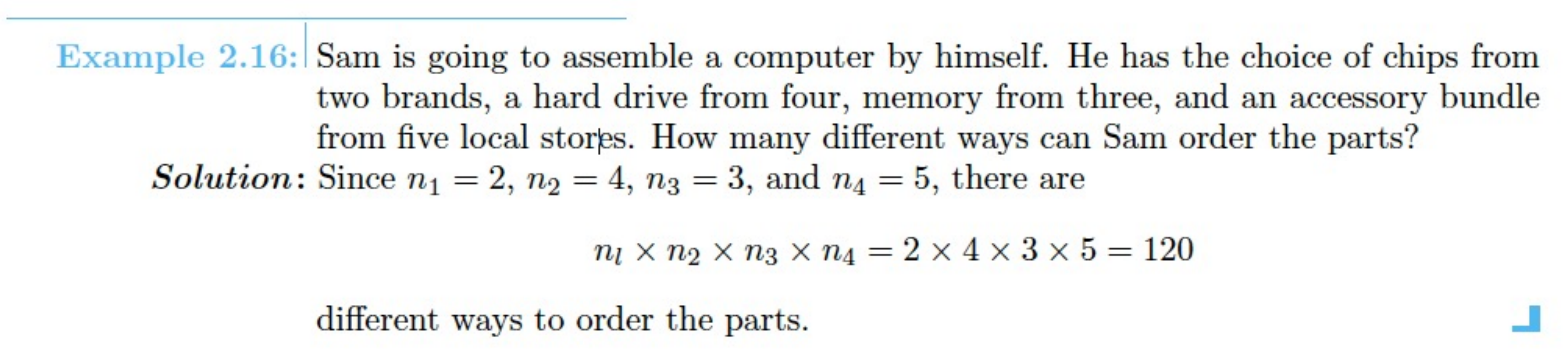
# Counting Sample Points

We can extend this to more operations…

If the first operation can be performed in $n_1$ ways, and the second operation in $n_2$ ways, and the third operation for $n_3$ ways, and so on for k operations, then the sequence can be performed in:

$$n_1 n_2 n_3 \ldots n_k \quad \text{ways}$$

# Counting Sample Points



**Example 2.16:** Sam is going to assemble a computer by himself. He has the choice of chips from two brands, a hard drive from four, memory from three, and an accessory bundle from five local stores. How many different ways can Sam order the parts?

***Solution*:** Since $n_1 = 2$, $n_2 = 4$, $n_3 = 3$, and $n_4 = 5$, there are

$$n_l \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 5 = 120$$

different ways to order the parts.

# Counting Sample Points

**Definition** A **permutation** is an arrangement of all or part of a set of objects.

**Example** Consider letters a, b, and c. How many possible permutations are there for the arrangement of these 3 letters?

$$n_1n_2n_3 = (3)(2)(1) = 6 \text{ ways}$$

# Counting Sample Points

**Example** In general, n distinct objects can be arranged in:

$$n(n-1)(n-2)\ldots(3)(2)(1) \text{ ways.}$$

There is a notation for such a number.

**Definition** For any nonnegative integer n, n! called "n factorial," is defined as:

$$n! = n(n-1)(n-2)\ldots(3)(2)(1)$$

with special case $0!=1$

# Counting Sample Points

**Theorem** The number of permutations of n objects is n!.

**Example** The number of the four letters a, b, c, and d will be:

$$4! = (4)(3)(2)(1) = 24.$$

**Example** Consider the number of permutations by taking only 2 of these 4 letters. These would be:

ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, and dc

There are: $n_1 n_2 = (4)(3) = 12$ possible permutations.

# Counting Sample Points

In general, n distinct objects taken r at a time can be arranged in

n(n-1)(n-2)…(n-r+1) ways.

**Theorem** The number of permutations of n distinct objects taken r at a time is,

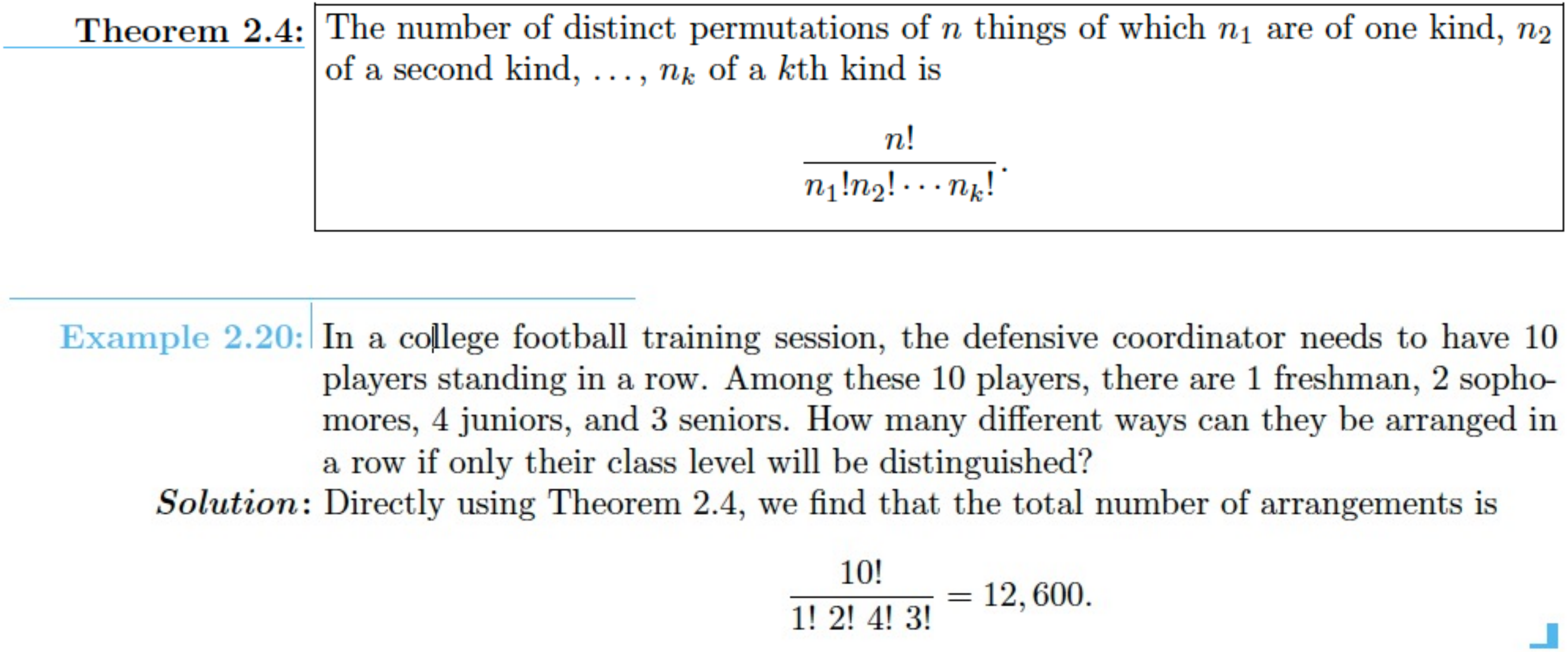
$${}_{n}P_{r} = \frac{n!}{(n-r)!}$$

# Counting Sample Points

**Example** How many different ways can three awards (research, teaching, service) be given to a class of 25 graduate students where each student can receive at most one award?

$${}_{25}P_3 = \frac{25!}{22!} = (25)(24)(23) = 13,800$$

# Counting Sample Points



**Theorem 2.4:** The number of distinct permutations of $n$ things of which $n_1$ are of one kind, $n_2$ of a second kind, ..., $n_k$ of a $k$th kind is

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

**Example 2.20:** In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?

***Solution*:** Directly using Theorem 2.4, we find that the total number of arrangements is

$$\frac{10!}{1!\ 2!\ 4!\ 3!} = 12,600.$$

# Counting Sample Points

We are often interested in selecting r objects from n *without regard to order.* We call such selections **combinations.**

**Theorem** The number of combinations of n distinct objects taken r at a time is,

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

# Counting Sample Points

**Example** How many pairs of letters can we make from the letters a, b, c, and d?

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$$

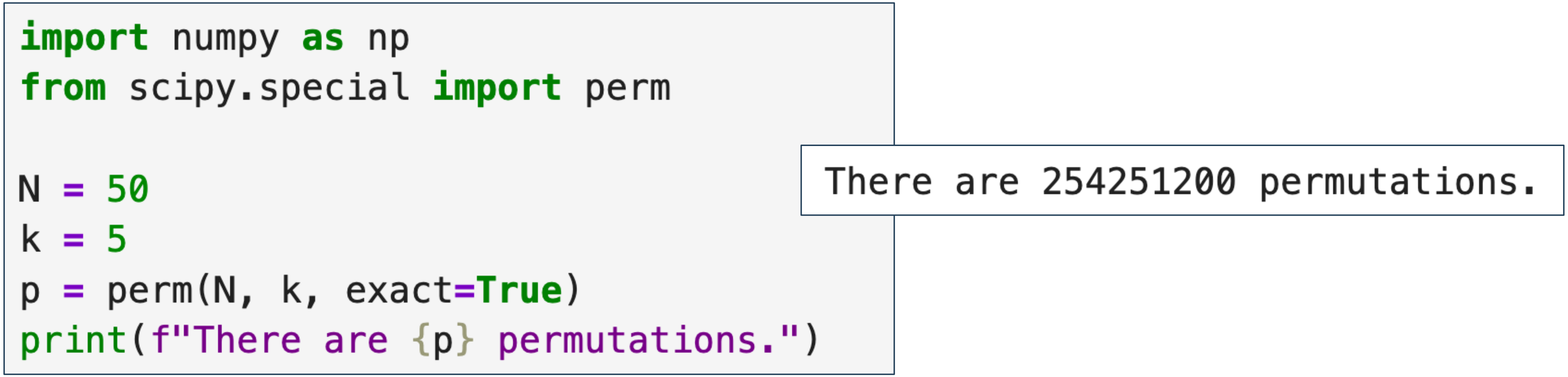
**ab, ac, ad, bc, bd, cd**

Note that we do not consider order so “ab” and “ba” are the same element.

# Python Example: Permutations

**Example** Suppose a bin contains 50 balls, each numbered 1, 2, 3, …, 50. Suppose we draw 5 balls at random in sequence. How many unique sequences of 5 balls can occur?

$$_{50}P_5 = \frac{50!}{45!} = (50)(49)(48)(47)(46) = 254,251,200$$



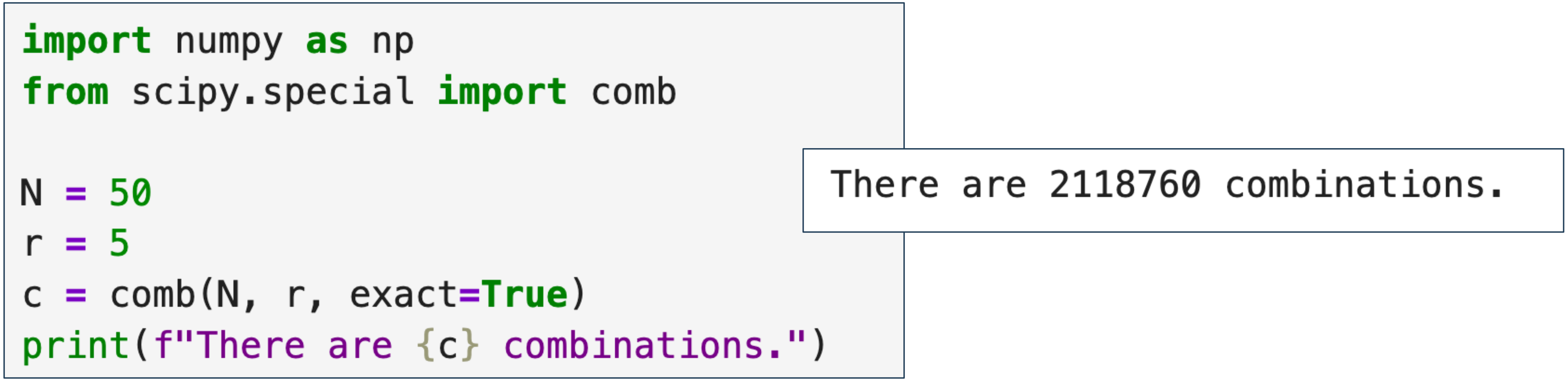
```python
import numpy as np
from scipy.special import perm

N = 50
k = 5
p = perm(N, k, exact=True)
print(f"There are {p} permutations.")
```

```
There are 254251200 permutations.
```

# Python Example: Combinations

**Example** Suppose a bin contains 50 balls, each numbered 1, 2, 3, …, 50. Suppose we draw 5 balls at random. Ignoring ordering, how many unique combinations of 5 balls can occur?

$$\binom{50}{5} = \frac{50!}{5!(50-5)!} = 2,118,760$$



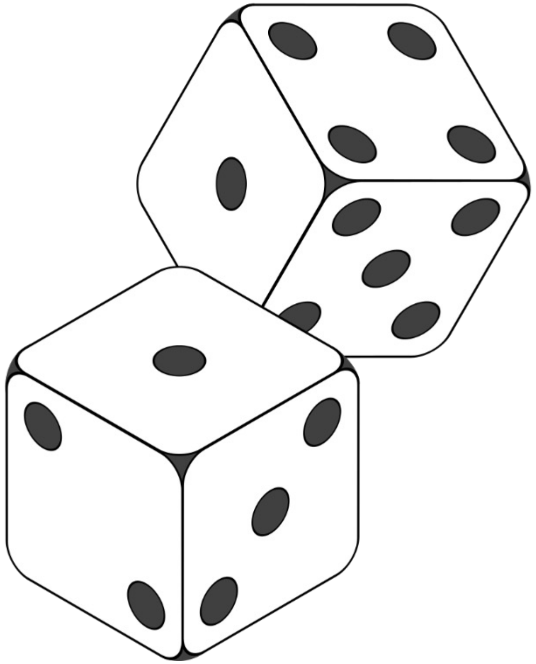
```python
import numpy as np
from scipy.special import comb

N = 50
r = 5
c = comb(N, r, exact=True)
print(f"There are {c} combinations.")
```

```
There are 2118760 combinations.
```

# Outline

- Random Events
- Counting Sample Points
- **Random Events and Probability**
- Conditional Probability & Independent Events

# Random Events and Probability



Assume each outcome is equally likely, and sample space is finite, then the probability of event is:

$$P(E) = \frac{|E|}{|\Omega|}$$

Number of outcomes in event set

Number of possible outcomes in sample space

This is the **uniform probability distribution**

**Example** Probability that we roll *only* even numbers,

$$E^{\text{even}} = \{(2,2),(2,4),\ldots,(6,4),(6,6)\}$$

$$P(E^{\text{even}}) = \frac{|E^{\text{even}}|}{|\Omega|} = \frac{9}{36}$$

# Random Events and Probability

**Example** Probability that the *sum of both dice* is even,

$$E^{\text{sum even}} = \{(1,1),(1,3),(1,5),\ldots,(2,2),(2,4),\ldots\}$$

$$P(E^{\text{sum even}}) = \frac{|E^{\text{sum even}}|}{|\Omega|} = \frac{18}{36} = \frac{1}{2}$$

**Example** Probability that the *sum of both dice* is <u>greater than 12,</u>

$$E^{>12} = \emptyset$$

$$P(E^{>12}) = \frac{|E^{>12}|}{|\Omega|} = 0$$

*i.e. we can reason about the probability of <u>impossible</u> outcomes*

# Random Events and Probability

To measure the *probability* of an event…

- Function P(E) maps events to probabilities in interval [0,1]
- P(E) known as a **probability distribution**
- Follows the **axioms of probability**,

1. For any event $E, 0 \leq P(E) \leq 1$
2. $P(\Omega) = 1 \text{ and } P(\emptyset) = 0$
3. For any *finite* or *countably infinite* sequence of <u>pairwise mutually disjoint events</u> $E_1, E_2, E_3, \ldots$

$$P\Big(\bigcup_{i \geq 1} E_i\Big) = \sum_{i \geq 1} P(E_i)$$

*How likely is it that two people share the same birthday here?*

*The probability of random events is not always intuitive.*

# Birthday Paradox

**Assumptions**

- 30 people in the room (there are more)
- Birthday uniformly distributed over 365 days (this is a simplification but easy)
- Ignore leap year effects

Total number of possible combinations of birthdays among 30 people,

$$|\Omega| = 365^{30}$$

Let E' be event **no two people** share a birthday—number of ways is,

$$|E'| =_{365} P_{30}$$

# Birthday Paradox

Probability of having **no two matching birthdays**,

$$P(E') = \frac{|E'|}{|\Omega|} = \frac{_{365}P_{30}}{365^{30}} = 0.294$$

Let E be the event that at least two people share a birthday,

$$P(E) = 1 - P(E') = 0.706$$

With the 75 people registered
P(E) = 0.9997

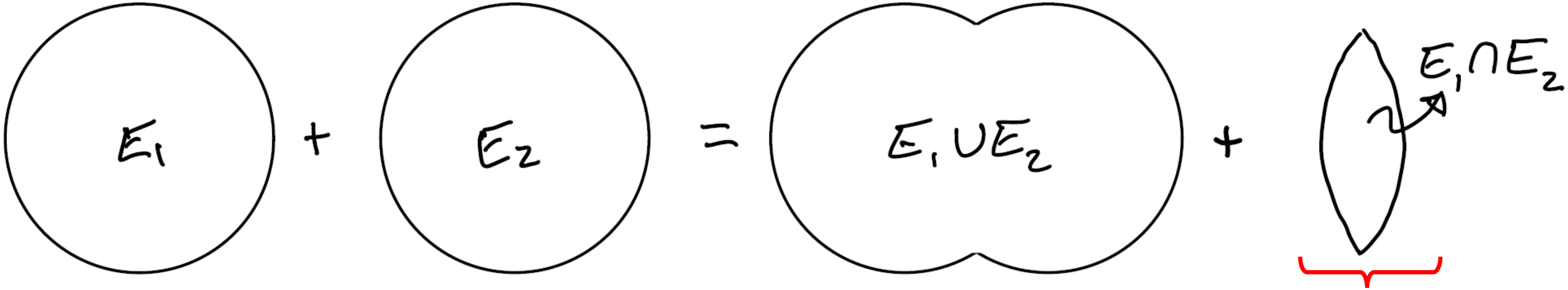
*With only 30 people there is over 70% chance of shared birthdays*

# Random Events and Probability

**Lemma:** *For <u>any</u> two events* $E_1$ *and* $E_2$ *,*

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

**Graphical Proof:**



Subtract from both sides

# Random Events and Probability

**Lemma:** *For <u>any</u> two events* $E_1$ *and* $E_2$,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

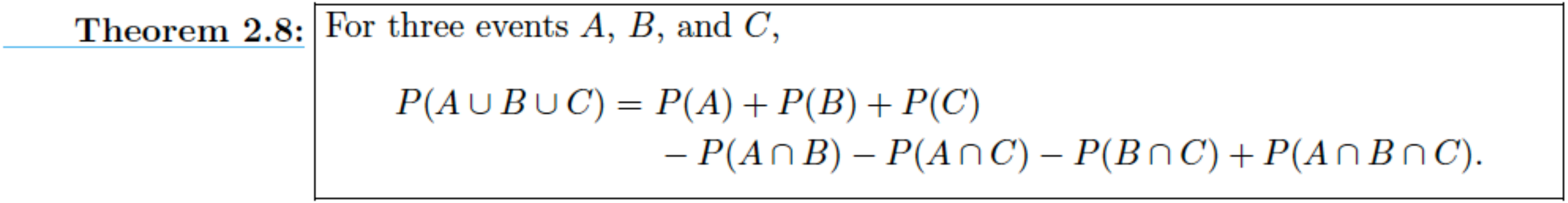
**Proof:**

$$P(E_1) = P(E_1 - (E_1 \cap E_2)) + P(E_1 \cap E_2)$$

$$P(E_2) = P(E_2 - (E_1 \cap E_2)) + P(E_1 \cap E_2)$$

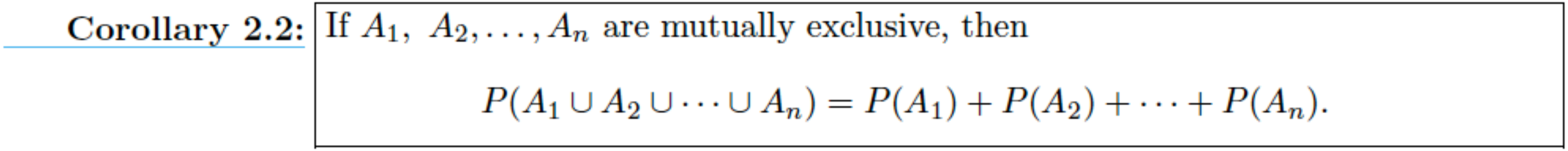
$$P(E_1 \cup E_2) = P(E_1 - (E_1 \cap E_2)) + P(E_2 - (E_1 \cap E_2)) + P(E_1 \cap E_2)$$

# Random Events and Probability



**Theorem 2.8:** For three events $A$, $B$, and $C$,

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



**Corollary 2.2:** If $A_1,\ A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

# Outline

- Random Events
- Counting Sample Points
- Random Events and Probability
- Conditional Probability & Independent Events

# Conditional Probability

The probability of an event B occurring when it is known that some event A has occurred is called the **conditional probability** and is denoted,

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

Usually read as “the probability of B, given A”.

# Conditional Probability

**Example** Two dice are rolled, and the sum is equal to 6. What is the probability that the first die is 3?

Let S be the event that the sum is 6. We have,

|S| = |{5,1},{1,5},{4,2},{2,4},{3,3}| = 5

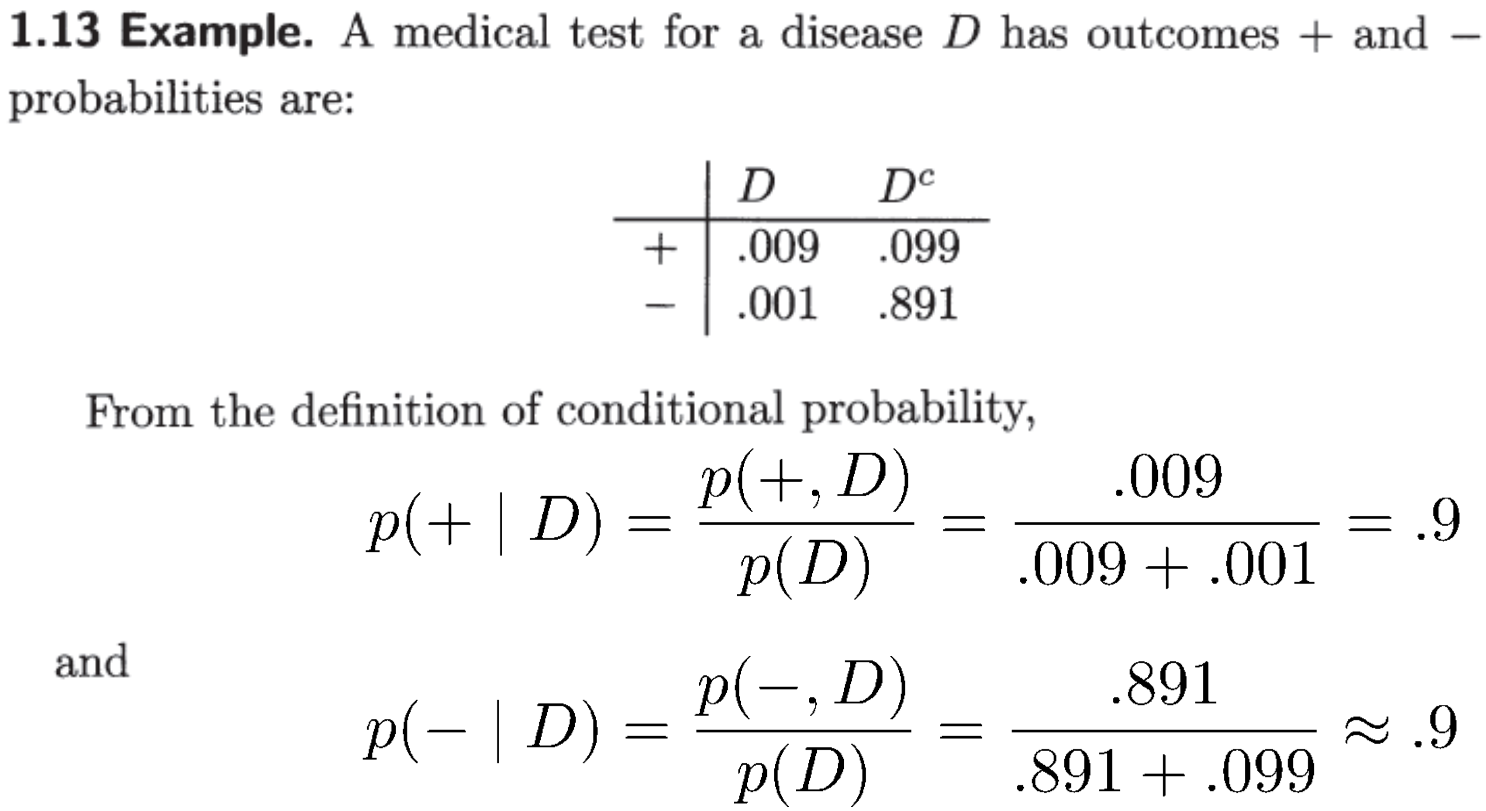
Let E be the event that the first die is a 3, then:

$$P(E \mid S) = \frac{P(E \cap S)}{P(S)} = \frac{1}{5}$$

# Fundamental Rules of Probability

Given two random events A and B the **conditional distribution** is:

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



**1.13 Example.** A medical test for a disease $D$ has outcomes $+$ and $-$. The probabilities are:

|  | $D$ | $D^c$ |
|---|---|---|
| $+$ | .009 | .099 |
| $-$ | .001 | .891 |

From the definition of conditional probability,

$$p(+ \mid D) = \frac{p(+, D)}{p(D)} = \frac{.009}{.009 + .001} = .9$$

and

$$p(- \mid D) = \frac{p(-, D)}{p(D)} = \frac{.891}{.891 + .099} \approx .9$$

[ Source: Wasserman, L. 2004 ]

# Probability Chain Rule

The **probability chain rule** is,

$$P(X \cap Y) = \underbrace{P(Y)}_{\text{Marginal}}\underbrace{P(X \mid Y)}_{\text{Conditional}}$$

**Proof** By definition of the conditional distribution,

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

Multiply both sides by P(Y),

$$P(X \cap Y) = P(Y)P(X \mid Y)$$

# Probability Chain Rule

Suppose we have a collection of N random events,

$$X_1, X_2, \ldots, X_N$$

The probability chain rule for these random variables is,

$$\begin{aligned} P(X_1 \cap X_2 \cap \ldots \cap X_N) &= P(X_1)P(X_2 \mid X_1) \ldots P(X_N \mid X_{N-1}, \ldots, X_1) \\ &= P(X_1) \prod_{i=2}^{N} P(X_i \mid X_{i-1}, \ldots, X_1) \end{aligned}$$

The chain rule is valid for any ordering of RVs, for example:

$$P(X_1 \cap \ldots \cap X_N) = P(X_2)P(X_3 \mid X_2)P(X_1 \mid X_2, X_3) \ldots P(X_7 \mid X_1, \ldots, X_6, X_8, \ldots, X_N)$$

# Intuition Check

*Question: Roll two dice and let their outcomes be* $X_1, X_2 \in \{1, \ldots, 6\}$ *for die 1 and die 2, respectively. Recall the definition of conditional probability,*

$$p(X_1 \mid X_2) = \frac{p(X_1 \cap X_2)}{p(X_2)}$$

*Which of the following are true?*

a) $p(X_1 = 1 | X_2 = 1) > p(X_1 = 1)$

b) $p(X_1 = 1 | X_2 = 1) = p(X_1 = 1)$

c) $p(X_1 = 1 | X_2 = 1) < p(X_1 = 1)$

# Intuition Check

*__Question:__ Let* $X_1 \in \{1, \ldots, 6\}$ *be outcome of die 1, as before. Now let* $X_3 \in \{2, 3, \ldots, 12\}$ *be the* sum of both dice. Which of the following are true?

- *a)* $p(X_1 = 1 | X_3 = 3) > p(X_1 = 1)$
- *b)* $p(X_1 = 1 | X_3 = 3) = p(X_1 = 1)$
- *c)* $p(X_1 = 1 | X_3 = 3) < p(X_1 = 1)$

# Dependence of RVs

Intuition…

Consider $P\left(B \middle| A\right)$ where you want to bet on $B$

Should you pay to know A?

In general you would pay something for A if it changed your belief about B. In other words if,

$$P\left(B \middle| A\right) \neq P\left(B\right)$$

# Independent Events

There is a special case when,

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

In this case we refer to A and B as **independent events.**

**Example** Roll two fair dice. Let $E_1$ be the event that the first die is a 1. What is the probability that the second die is a 1 (denoted $E_2$)?

$$P(E_2 \mid E_1) = P(E_2) = \frac{6}{36} = 1/6$$