# Topics in Discrete Mathematics 

## George Voutsadakis ${ }^{1}$

${ }^{1}$ Mathematics and Computer Science

Lake Superior State University

## LSSU Math 216

## (1) Probability

- Sample Space
- Events
- Conditional Probability and Independence
- Random Variables
- Expectation


## Introducing Probabilities

- Probability theory provides us with tools for analyzing situations in which events occur at random.
- It is used in a wide range of disciplines (e.g., sociology, physics, genetics, finance).
- It is important to distinguish between mathematical probability theory and its application to problems in the real world.
- In mathematics, a probability is a number associated with some object.
- In applications, the object is some event or uncertain action, and the number is a measure of how frequent or how likely that event is.
- Probabilities are real numbers between 0 and 1 .
- An event with probability 1 is certain to occur, and an event with probability 0 is impossible.
- Probabilities between 0 and 1 reflect the relative likelihood between these two extremes, with unlikely close to 0 , and likely close to 1 .
- Discrete probability problems are often counting problems recast in the language of probability theory.


## Subsection 1

## Sample Space

## Sample Space and Probabilities

- Consider the toss of a die.
- Since, we cannot say in advance which of the six sides of the die will land face up, the outcome of this experiment is unpredictable.
- If the die is fair, we can say that all six outcomes are equally likely.
- Thus, although we cannot predict the outcome, we can describe the likelihood of seeing, for example, a 4 when we roll the die.
- Mathematicians model the roll of a die using the sample space, which consists of two parts:
- It contains a set $S$ of all the outcomes of some experiment.
- It quantifies the likelihood $P(s)$ of each of these outcomes $s \in S$.
- We call the number $P(s)$, with $0 \leq P(s) \leq 1$, the probability of the outcome $s \in S$.
- By convention, we require that the sum of the probabilities of the various possible outcomes be 1 , i.e., $\sum_{s \in S} P(s)=1$.


## Rolling a Fair Die

- Let $S$ be the set of outcomes from the roll of a die.
- We name the outcomes with the integers $1,2,3,4,5$, and 6 , so $S=\{1,2,3,4,5,6\}$.
- We also have a function $P: S \rightarrow \mathbb{R}$, defined by

$$
\begin{array}{lll}
P(1)=\frac{1}{6} & P(2)=\frac{1}{6} & P(3)=\frac{1}{6} \\
P(4)=\frac{1}{6} & P(5)=\frac{1}{6} & P(6)=\frac{1}{6}
\end{array}
$$

- The probabilities are nonnegative real numbers between 0 and 1 .
- Moreover, the sum of the probabilities of all the elements in $S$ is 1 .


## Definition of Sample Space

## Definition (Sample Space)

A sample space is a pair $(S, P)$ where $S$ is a finite, nonempty set and $P$ is a function $P: S \rightarrow \mathbb{R}$, such that $P(s) \geq 0, s \in S$, and $\sum_{s \in S} P(s)=1$.

- Example: Consider the spinner shown: The set of outcomes $S$ is $S=\{1,2,3,4\}$. The probability function $P: S \rightarrow \mathbb{R}$ measures how likely it is for the spinner to land in each of the regions.


Thus we have

$$
P(1)=\frac{1}{2}, \quad P(2)=\frac{1}{4}, \quad P(3)=\frac{1}{8}, \quad P(4)=\frac{1}{8} .
$$

We check that

$$
\sum_{s \in S} P(S)=P(1)+P(2)+P(3)+P(4)=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{8}=1
$$

## A Pair of Dice and a Hand of Poker

- Example: Two dice are tossed. Die 1 can land in any one of 6 equally likely ways, and the same is true for die 2 . We express the outcome of this experiment as an ordered pair $(a, b)$ where $a$ and $b$ are integers between 1 and 6 . There are $6 \times 6=36$ possible outcomes for this experiment. We let $S=\{(1,1),(1,2),(1,3), \ldots,(6,5),(6,6)\}$. Since each of the 36 possible outcomes of this experiment is equally likely, $P(s)=\frac{1}{36}$, for all $s \in S$.
- Example: A hand of poker is a five-element subset of the standard deck of 52 cards. There are $\binom{52}{5}$ different five-element subsets of a 52-element set. The set $S$ consists of all these different five-element subsets. Since they are all equally likely, we have

$$
P(s)=\frac{1}{\binom{52}{5}}, \quad \text { for all } s \in S
$$

## Tossing a Coin

- A fair coin is tossed five times in a row, and the sequence of HEADS and TAILS is recorded. We model this as a sample space. The set $S$ contains all possible outcomes of this experiment. An outcome is denoted by a length-five list of Hs and Ts . There are $2^{5}=32$ such lists, and they are all equally likely. Thus
$S=\{$ TTTTT, TTTTH, TTTHT,.., HHHHT, HHHHH $\}$ and $P(s)=\frac{1}{32}$, for all $s \in S$.
- We can create sample spaces that have no specific physical interpretation:
- Example: Let $S=\{1,2,3,4,5,6\}$ and define $P: S \rightarrow \mathbb{R}$ by

$$
\begin{gathered}
P(1)=0.1, \quad P(2)=0.4, \quad P(3)=0.1 \\
P(4)=0, \quad P(5)=0.2, \quad P(6)=0.2
\end{gathered}
$$

Note that $\sum_{s \in S} P(s)=1$.

## Subsection 2

## Events

## Events and Their Probabilities

- Recall the die-throwing example. In the sample space $(S, P), P$ gives the probability of each of the six possible outcomes of rolling the die.
- Suppose we want the probability that the die will show an even number, i.e., the probability that the die produces a result in the set $\{2,4,6\}$. We call such a set an event.
- The probability of this event is $\frac{1}{2}$. Each of the three outcomes of the die has probability $\frac{1}{6}$, and we add them. We denote the probability of the event $\{2,4,6\}$ as $P(\{2,4,6\})$.
- The function $P$ is a function $P: S \rightarrow \mathbb{R}$. We use the same symbol applied to a subset of $S$. We define this extended use of the symbol $P$ so that $P(\{2,4,6\})=P(2)+P(4)+P(6)$.


## Definition (Event)

Let $(S, P)$ be a sample space. An event $A$ is a subset of $S$ (i.e., $A \subseteq S$ ). The probability of an event $A$, denoted $P(A)$, is $P(A)=\sum_{a \in A} P(a)$.

## Pair of Dice

- Let $(S, P)$ be the sample space representing the toss of a pair of dice. What is the probability that the sum of the numbers on the two dice is 7 ?
Let $A$ denote the event that the numbers on the dice sum to 7 .

$$
\begin{aligned}
A & =\{(a, b) \in S: a+b=7\} \\
& =\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}
\end{aligned}
$$

The probability of this event is

$$
\begin{aligned}
P(A) & =P[(1,6)]+P[(2,5)]+P[(3,4)]+P[(4,3)] \\
& +P[(5,2)]+P[(6,1)] \\
& =\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36} \\
& =\frac{1}{6} .
\end{aligned}
$$

## Tossing a Coin

- Let $(S, P)$ be the sample space that models tossing a coin five times.
- What is the probability that exactly one HEAD shows?

Let $A$ denote the event that exactly one HEAD emerges. We can write this out explicitly as $A=\{$ HTTTT, THTTT, TTHTT, TTTHT, TTTTH $\}$. Note that $A$ contains five outcomes, each of which has probability $\frac{1}{32}$. Therefore $P(A)=\frac{5}{32}$.

- What is the probability that exactly two HEADs show? Let $B$ be the event that exactly two of the coin flips show HEADs. We can write out the elements of $B$ explicitly, but all we really need to know is how many elements are in $B$ (because all elements of $S$ have the same probability). The size of $B$ is $|B|=\binom{5}{2}=10$ (Choose
2 out of the 5 positions for the Hs). Thus, $P(B)=\frac{10}{32}=\frac{5}{16}$.


## Ten Dice

- Ten dice are tossed. What is the probability that none of the dice shows the number 1 ?
We begin by constructing a sample space $(S, P)$. Let $S$ denote the set of all possible outcomes of this experiment. An outcome of this experiment can be expressed as a length-ten list formed from the symbols $1,2,3,4,5$, and 6 . There are $6^{10}$ such lists and they are all equally likely, so $P(s)=6^{-10}$, for all $s \in S$. Let $A$ be the event that none of the dice shows the number 1 . Since all elements of $S$ have the same probability, this problem reduces to finding the number of elements in $A$. The number of outcomes that do not have the number 1 is the number of lists of length ten whose elements are chosen from the symbols $2,3,4,5$, and 6 . The number of such lists is $5^{10}$. Therefore there are $5^{10}$ elements in $A$, all of which have probability $6^{-10}$. Therefore $p(A)=5^{10} \times 6^{-10}=\left(\frac{5}{6}\right)^{10} \approx 0.1615$.


## Poker: Four of a Kind

- Recall the poker hand sample space. A poker hand is called a four of a kind if four of the five cards show the same value (e.g., all 7 s or all kings). What is the probability that a poker hand is a four of a kind?

Let $A$ be the event that the poker hand is a four of a kind. Since every poker hand has probability $\frac{1}{\binom{52}{5}}$, we simply need to calculate $|A|$.
There are 13 choices for which value is repeated four times. Given that value, there are 48 choices for the fifth card. Thus,
$P(A)=\frac{13 \times 48}{\binom{52}{5}}=\frac{1}{4165} \approx 0.00024$.

## Four Children

- A couple has four children. Which is more likely: They have two boys and two girls, or they have three of one gender and one of the other?

Let $S$ be the set of all possible lists of genders the couple might have. We can represent the genders of the children as a list of length four drawn from the symbols $b$ and $g$. There are $2^{4}=16$ such lists, and they are all equally likely.
Let $A$ be the event that the couple has two boys and two girls. Then $A=\{g g b b, g b g b, g b b g, b b g g, b g b g, b g g b\}$. So
$P(A)=\frac{6}{16}=\frac{3}{8}=0.375$.
Let $B$ be the event that the couple has three of one gender and one of the other. Thus, $B=\{$ gggb,ggbg,gbgg,bggg,bbbg,bbgb,bgbb,gbbb $\}$. So $P(B)=\frac{8}{16}=\frac{1}{2}=0.5$.
Since $P(B)>P(A)$, it is more likely for the couple to have three of one gender and one of the other than to have two boys and two girls.

## Combining Events: Union and Intersection

- Since events are subsets of a probability space, we can use the operations of set theory (union, intersection, etc.) to combine events.
- Let $(S, P)$ be a sample space.
- If $A$ and $B$ are events, so is $A \cup B$. We can think of $A \cup B$ as the event that " $A$ or $B$ occurs".
- Example: If $A$ is the event "a die shows an even number" and $B$ is the event that "the die shows a prime number", then $A \cup B$ is the event that "the die shows a number that is even or prime", so $A \cup B=\{2,4,6\} \cup\{2,3,5\}=\{2,3,4,5,6\}$. The probability of the event $A \cup B$ is $\frac{5}{6}$.
- Likewise, $A \cap B$ is the event that "both $A$ and $B$ occur".
- If $A$ is the event that "a die shows an even number" and $B$ is the event that "it shows a prime number", then
$A \cap B=\{2,4,6\} \cap\{2,3,5\}=\{2\}$ and $P(A \cap B)=\frac{1}{6}$.


## Combining Events: Set Difference

- The set $A-B$ is the event that " $A$ occurs but $B$ does not".
- For the die rolling example, $A-B=\{2,4,6\}-\{2,3,5\}=\{4,6\}$. The probability of "rolling a number that is even but not prime" is $P(A-B)=\frac{1}{3}$.
- Since the set $S$ of a sample space is the "universe" of all outcomes, it is sensible to write $\bar{A}$ to stand for the set $S-A$. The set $\bar{A}$ represents the event " $A$ does not occur".
- For the die rolling example, $\bar{A}$ is the event that "we do not roll an even number", so $P(\bar{A})=P(\{1,3,5\})=\frac{1}{2}$.
- Can we find $P(A \cup B)$ if we know only $P(A)$ and $P(B)$ ? No!
- Consider these two examples (from rolling a die):
- Let $A=\{2,4,6\}$ and $B=\{2,3,5\}$. We have $P(A)=P(B)=\frac{1}{2}$ and $P(A \cup B)=P(\{2,3,4,5,6\})=\frac{5}{6}$.
- Let $A=\{2,4,6\}$ and let $B=\{1,3,5\}$. Then $P(A)=P(B)=\frac{1}{2}$ and $P(A \cup B)=P(S)=1$.
- So $P(A)=P(B)=\frac{1}{2}$ is not enough to determine $P(A \cup B)$.


## An Important Proposition

## Proposition

Let $A$ and $B$ be events in a sample space $(S, P)$. Then $P(A)+P(B)=P(A \cup B)+P(A \cap B)$.

- Consider the sums $P(A)+P(B)$ and $P(A \cup B)+P(A \cap B)$.
- The first is $P(A)+P(B)=\sum_{s \in A} P(s)+\sum_{s \in B} P(s)$.
- The second is $P(A \cup B)+P(A \cap B)=\sum_{s \in A \cup B} P(s)+\sum_{s \in A \cap B} P(s)$.

Consider an arbitrary element $s \in S$. There are four possibilities:

- $s$ is in neither $A$ nor $B$. In this case, the term $P(s)$ does not enter either side of the equation.
- $s$ is in $A$ but not in $B$. In this case, $P(s)$ enters exactly once into both sides of the equation (once in $P(A)$ and once in $P(A \cup B)$ ).
- $s$ is in B but not in A . As before, $P(s)$ enters exactly once into both sides of the equation.
- $s$ is in both $A$ and $B$. In this case, $P(s)$ appears twice on each side of the equation (once each in $P(A), P(B), P(A \cup B)$ and $P(A \cap B)$ ).
- Therefore, $P(A)+P(B)=P(A \cup B)+P(A \cap B)$ are equal.


## A Corollary and Mutually Exclusive Events

## Proposition

Let $(S, P)$ be a sample space and let $A$ and $B$ be events. We have the following:

- If $A \cap B=\emptyset$, then $P(A \cup B)=P(A)+P(B)$.
- $P(A \cup B) \leq P(A)+P(B)$.
- $P(S)=1$.
- $P(\emptyset)=0$.
- $P(\bar{A})=1-P(A)$.
- Two events whose intersection is the empty set are called mutually exclusive.


## The Birthday Problem for Four People

- Four people are chosen at random. What is the probability that two (or more) of them have the same birthday?
We make two simplifying assumptions:
- We ignore the possibility of a February 29 birthday.
- We assume that all birthdays are equally likely, i.e., the probability a random person is born on a given day of the year is $\frac{1}{365}$.
The sample space $(S, P)$ consists of all length 4 lists of days of the year, represented as $\left(d_{1}, d_{2}, d_{3}, d_{4}\right)$, where the $d_{i}$ are integers from 1 to 365 . All such lists are equally likely with probability $365^{-4}$. Let $A$ be the event that two (or more) of the people have the same birthday. It is easier to calculate $P(\bar{A})$, the probability they all have different birthdays. We can choose the first date in 365, the second in 364, the third in 363, and the last in 362 ways. Therefore, $P(\bar{A})=\frac{365 \cdot 364 \cdot 363 \cdot 362}{365^{4}}$.
Thus, $P(A)=1-P(\bar{A})=1-\frac{365 \cdot 364 \cdot 363 \cdot 362}{365^{4}} \approx 0.0164$. It is rather unlikely that two of them have the same birthday.


## The Birthday Problem for Twenty-Three People

- Suppose that 23 people are chosen at random. What is the probability that some of them have the same birthday?
It would seem, since 23 is much smaller than 365 , that this is also an unlikely event.
Consider the sample space $(S, P)$ where $S$ contains all length 23 lists $\left(d_{1}, d_{2}, \ldots, d_{23}\right)$ where each of the $d_{i}$ is an integer from 1 to 365 . We assign probability $365^{-23}$ to each of these lists.
Let $A$ be the event that two (or more) of the $d_{i} \mathrm{~s}$ are equal. It is again easier to calculate the probability of $\bar{A}$. The number of length 23 repetition-free lists we can form from 365 different symbols is $(365)_{23}$. Therefore, $P(\bar{A})=\frac{(365)_{23}}{365^{23}}$ and so $P(A)=1-\frac{(365)_{23}}{365^{23}}$.
Calculating, $P(A)=0.5073$, so it is more likely that two (or more) people have the same birthday than it is that no two of them do!


## Subsection 3

## Conditional Probability and Independence

## Introduction

- Let $A$ represent the event "a student misses the school bus".
- Let $B$ represent the event "the student's alarm clock malfunctions".
- They have low probability, i.e., $P(A)$ and $P(B)$ are small numbers.
- Consider the probability that the "student misses the school bus given that the alarm clock malfunctioned".
- It is now very likely the student will miss the bus!
- We denote this probability as $P(A \mid B)$ : the probability that event $A$ occurs given that event $B$ occurs.
- We illustrate events as regions in a diagram:
- The box $S$ represents the entire sample space.
- Regions $A$ and $B$ represent the two events.

- The proportion of $B$ 's area that is overlapped by $A$ represents those days on which the student misses the bus and the alarm clock fails.


## An Example on Conditional Probability

- The proportion of box $B$ covered by the overlap region is $\frac{P(A \cap B)}{P(B)}$.
- This represents the frequency with which the student misses the bus on days the alarm clock fails.
- The conditional probability of event $A$ given $B$ is $P(A \mid B)=\frac{P(A \cap B)}{P(B)}$.
- Example: Let $(S, P)$ be the pair-of-die sample space. Consider the events $A$ and $B$ defined by:
- Event $A$ : the numbers on the dice sum to 8 .
- Event B: the numbers on the dice are both even.

$$
\begin{aligned}
& A=\{(2,6),(3,5),(4,4),(5,3),(6,2)\} \\
& B=\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}
\end{aligned}
$$

$$
\text { So } P(A)=\frac{5}{36} \text { and } P(B)=\frac{9}{36}=\frac{1}{4} \text {. }
$$

What is the probability the dice sum to 8 given that both dice show even numbers? Of the nine, equally likely dice rolls in set $B$, three of them sum to 8 . Therefore $P(A \mid B)=\frac{3}{9}=\frac{1}{3}$. Notice

$$
P(A \cap B)=\frac{3}{36}=\frac{1}{12} \text { and } P(B)=\frac{9}{36}=\frac{1}{4}, \text { so } \frac{P(A \cap B)}{P(B)}=\frac{1 / 12}{1 / 4}=\frac{1}{3} .
$$

## Conditional Probability

## Definition (Conditional Probability)

Let $A$ and $B$ be events in a sample space $(S, P)$ and suppose $P(B) \neq 0$. The conditional probability $P(A \mid B)$, the probability of $A$ given $B$, is

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

- Consider the spinner: Let $A$ be the event that we spin to a 1 and let $B$ be the event that the pointer ends in a shaded region.


What is the probability that we spin to a 1 given that the pointer ends in a shaded region?
Notice that region 1 consumes $\frac{4}{5}$ of the colored portion of the diagram. Also: $P(A \mid B)=\frac{P(A \cap B)}{P(B)}=\frac{P(\{1\})}{P(\{1,3\})}=\frac{1 / 2}{5 / 8}=\frac{4}{5}$.

## Flipping a Coin Five Times

- A coin is flipped five times. What is the probability that the first flip is a TAIL given that exactly three HEADS are flipped?
Let
- $A$ be the event that the first flip is TAILS;
- $B$ be the event that we flip exactly three HEADS.

We calculate $P(A)=\frac{2^{4}}{2^{5}}=\frac{1}{2}$ and $P(B)=\frac{\binom{5}{3}}{2^{5}}=\frac{10}{32}=\frac{5}{16}$. To calculate $P(A \mid B)$, we also need to know $P(A \cap B)$. The set $A \cap B$ contains exactly $\binom{4}{3}=4$ sequences. So $P(A \cap B)=\frac{4}{32}=\frac{1}{8}$. Therefore, $P(A \mid B)=\frac{P(A \cap B)}{P(B)}=\frac{1 / 8}{5 / 16}=\frac{2}{5}$.

## Example on Independence

- A coin is flipped five times. What is the probability that the first flip comes up HEADS given that the last flip comes up HEADS?
Let
- $A$ be the event that the first flip comes up HEADS;
- $B$ be the event that the last flip comes up HEADS.

We have $P(A)=\frac{2^{4}}{2^{5}}=\frac{1}{2}, P(B)=\frac{2^{4}}{2^{5}}=\frac{1}{2}$ and $P(A \cap B)=\frac{2^{3}}{2^{5}}=\frac{1}{4}$.
Therefore, $P(A \mid B)=\frac{P(A \cap B)}{P(B)}=\frac{1 / 4}{1 / 2}=\frac{1}{2}$.
Notice that $P(A \mid B)=P(A)$. The probability the first flip comes up HEADS has nothing to do with the last flip. We call such events independent.

## Independent Events

- We work out some consequences of the equation $P(A \mid B)=P(A)$.
- We have $P(A \mid B)=\frac{P(A \cap B)}{P(B)}=P(A)$ and, if we multiply by $P(B)$, we get $P(A \cap B)=P(A) P(B)$.
- If $P(A) \neq 0$, we can divide by $P(A): P(B \mid A)=\frac{P(A \cap B)}{P(A)}=P(B)$.


## Proposition

Let $A, B$ be events in a sample space $(S, P)$ and suppose $P(A), P(B) \neq 0$. Then the following statements are equivalent:
(3) $P(A \mid B)=P(A)$.
(2) $P(B \mid A)=P(B)$.

- $P(A \cap B)=P(A) P(B)$.


## Definition (Independent events)

Let $A$ and $B$ be events in a sample space. We say that these events are independent provided $P(A \cap B)=P(A) P(B)$. Events that are not independent are called dependent.

## Drawing Balls

- A bag contains ten red and ten blue balls. Two balls are drawn. Let
- $A$ be the event that the first ball drawn is red;
- $B$ be the event that the second ball drawn is red.

Are these events independent?
The answer depends on whether or not we replace the first ball before drawing the second.

- Suppose we replace the first ball before drawing the second. Then there are $20 \times 20$ ways to pick the two balls, of which $10 \times 20$ have the property that the first ball is red. Thus $P(A)=\frac{200}{400}=\frac{1}{2}$. Likewise, $P(B)=\frac{1}{2}$. Finally, there are $10 \times 10$ ways to draw the balls such that both the first and second balls are red. Therefore $P(A \cap B)=\frac{100}{400}=\frac{1}{4}$. Since $P(A \cap B)=\frac{1}{4}=\frac{1}{2} \cdot \frac{1}{2}=P(A) P(B), A$ and $B$ are independent.
- Suppose we do not replace the first ball once it is drawn. There are $20 \times 19=380$ different ways to draw two balls. There are $10 \times 19$ ways in which the first ball is red; hence $P(A)=\frac{190}{380}=\frac{1}{2}$. Similarly, $P(B)=\frac{1}{2}$. However, there are only $10 \times 9$ ways to pick the balls such that both are red. Therefore $P(A \cap B)=\frac{90}{380}=\frac{9}{38} \neq \frac{1}{4}=P(A) P(B)$ and so the events are dependent.


## Example on Repeated Independent Trials

- Consider again the spinner: Suppose we spin the needle twice. Instead of 4 possible outcomes, there are 16: $(1,1), \ldots,(4,4)$. What is the probability that we spin a 3 and then we spin a 2 ?


The first spin of the spinner and the second spin are independent of one another: the number that comes up on the second spin is not in any way dependent on the first number that appears. We think of "first spin a 3 " and "next spin a 2" as independent events with probabilities $\frac{1}{8}$ and $\frac{1}{4}$, respectively. Then the probability that we spin a 3 and then a 2 ought to be $\frac{1}{8} \cdot \frac{1}{4}=\frac{1}{32}$.

- This is an example of repeated independent trials. We have a sample space $(S, P)$ and, instead of drawing a single $s \in S$, we consider a sequence of events $s_{1}, s_{2}, \ldots, s_{n}$ each drawn at random from $S$.


## Repeated Trials

## Definition (Repeated Trials)

Let $(S, P)$ be a sample space and let $n$ be a positive integer. Let $S^{n}$ denote the set of all length $n$ lists of elements in $S$. Then $\left(S^{n}, P\right)$ is the $n$-fold repeated trial sample space in which

$$
P\left[\left(s_{1}, \ldots, s_{n}\right)\right]=P\left(s_{1}\right) P\left(s_{2}\right) \cdots P\left(s_{n}\right)
$$

- Example: The pair-of-dice sample space can be considered a repeated trial on a single die. Let $(S, P)$ be the sample space with $S=\{1,2,3,4,5,6\}$ and $P(s)=\frac{1}{6}$, for all $s \in S$. Then $\left(S^{2}, P\right)$ represents the "roll of two dice" sample space. The elements of $S^{2}$ are all possible results for rolling a pair of dice, from $(1,1)$ through $(6,6)$, all with probability $\frac{1}{6} \cdot \frac{1}{6}=\frac{1}{36}$.


## More Examples on Repeated Trials

- Example: Consider the sample space representing five flips of a fair coin.
We can reformulate this situation as follows: Let $(S, P)$ be the sample space in which $S=\{$ HEADS, TAILS $\}$ and $P(s)=\frac{1}{2}$, for both $s \in S$. The "toss-five-times" sample space is simply $\left(S^{5}, P\right)$. The set $S^{5}$ contains all length-five lists of the symbols HEADS and TAILS. All such lists are equally likely with probability $\frac{1}{32}$.
- Example: (Tossing an unfair coin) Imagine a coin that is not fairly balanced; that is, it does not turn up HEADS and TAILS with the same frequencies. We model this with a sample space $(S, P)$ where $S=\{$ HEADS, TAILS $\}$, but $P($ HEADS $)=p$ and $P($ TAILS $)=1-p$, where $p$ is a number with $0 \leq p \leq 1$. If we toss this coin five times, what is the probability that we see (in this order): HEADS, HEADS, TAILS, TAILS, HEADS? The answer is $P(\mathrm{HHTTH})=P(\mathrm{H}) P(\mathrm{H}) P(\mathrm{~T}) P(\mathrm{~T}) P(\mathrm{H})=p \cdot p \cdot(1-p) \cdot(1-p) \cdot p$.


## The TV Show "Let's Make a Deal"

- A contestant is presented with a choice of three doors. Behind exactly one of these doors is a terrific prize and the other doors conceal items of considerably less value. The contestant is asked to choose a door. At this point, the host Monty Hall shows the contestant one of the worthless prizes behind one of the other doors and offers the contestant the opportunity to switch to the other closed door. Is it helpful to switch to the other door, or does it actually matter?
An incorrect analysis: The probability that the prize is behind the door originally picked by the contestant is $\frac{1}{3}$. But now that one door has been revealed, the probability that the valuable prize is behind either of the two remaining doors is $\frac{1}{2}$, so it does not matter whether the contestant switches to the other door.
The error in this argument is that the contestant knows more than the fact that the prize is not behind a certain door. The door the host opens depends on which door the contestant originally chose!


## A Correct Mathematical Model for "Let's Make a Deal"

- Let us model this situation with a sample space.
- Suppose the contestant chooses door 1 .
- The prize might be behind door 1 , in which case the host will show door 2 or 3 . Let us suppose the host is equally likely to pick either.
- If the prize is behind door 2 , then the host will certainly show door 3.
- If the prize is behind door 3, then the host will certainly show door 2.
- Let us write "P1:S2" to stand for "the prize is behind door 1 and the host shows door 2." With this notation, the four possible occurrences are P1:S2, P1:S3, P2:S3, P3:S2. We model this as a sample space by assigning the following probabilities:

$$
P(\mathrm{P} 1: \mathrm{S} 2)=\frac{1}{6}, P(\mathrm{P} 1: \mathrm{S} 3)=\frac{1}{6}, P(\mathrm{P} 2: \mathrm{S} 3)=\frac{1}{3}, P(\mathrm{P} 3: \mathrm{S} 2)=\frac{1}{3}
$$

## A Correct Decision for "Let's Make a Deal"

- Suppose after the contestant picks door 1 , the host reveals the worthless item behind door 2 . Should the contestant switch to door 3 ?
Consider the following three events:
- $A$ : "the prize is behind door 1", i.e., $A=\{\mathrm{P} 1: \mathrm{S} 2, \mathrm{P} 1: \mathrm{S} 3\}$.
- $B$ : "the prize is behind door 3 ", i.e., $B=\{\mathrm{P} 3: \mathrm{S} 2\}$.
- $C$ : "the host reveals door 2", i.e., $C=\{\mathrm{P} 1: \mathrm{S} 2, \mathrm{P} 3: \mathrm{S} 2\}$.

Note that $P(A)=P(B)=\frac{1}{3}$. If the host did not reveal a door, there is no reason to switch. On the other hand,
$P(A \mid C)=\frac{P(A \cap C)}{P(C)}=\frac{P(\{\mathrm{P} 1: \mathrm{S} 2\})}{P(\{\mathrm{P}: \mathrm{S} 2 \mathrm{P} 3: \mathrm{S} 2\})}=\frac{1 / 6}{1 / 6+1 / 3}=\frac{1}{3}$ and
$P(B \mid C)=\frac{P(B \cap C)}{P(C)}=\frac{P(\{P 3: \mathrm{S} 2\})}{P(\{P 1: S 2, \mathrm{P} 3: \mathrm{S} 2\})}=\frac{1 / 3}{1 / 6+1 / 3}=\frac{2}{3}$.
Therefore it is twice as likely that the contestant will win the big prize by switching doors than by staying with the original choice.

## Subsection 4

## Random Variables

## Random Variables

- Let $(S, P)$ be a sample space.
- A random variable $X$ associates a number with each outcome in a sample space $(S, P)$, i.e., $X(s)$ is a number that depends on $s \in S$.
- For example, $X$ might represent the number of HEADS observed in ten flips of a coin. So, if $s=$ HHTHTTTTHT, then $X(s)=4$.
- The proper way to express this idea is to say that $X$ is a function.
- The domain of $X$ is the set $S$ of a sample space $(S, P)$.
- Each $s \in S$ has a value $X(s)$ that is usually a real number.
- In this case, we have $X: S \rightarrow \mathbb{R}$.


## Definition (Random Variable)

A random variable is a function defined on a probability space; that is, if $(S, P)$ is a sample space, then a random variable is a function $X: S \rightarrow V$ (for some set $V$ ).

## Examples

- Let $(S, P)$ be the pair-of-dice sample space. Let $X: S \rightarrow \mathbb{N}$ be the random variable that gives the sum of the numbers on the two dice. For example, $X[(1,2)]=3, X[(5,5)]=10$, and $X[(6,2)]=8$.
- Let $(S, P)$ be the sample space representing ten tosses of a fair coin. Let $X: S \rightarrow \mathbb{Z}$ be the random variable that gives the number of HEADS minus the number of TAILS. For example, $X($ HHTHTTTTHT $)=4-6=-2$.
We can also define random variables $X_{H}$ and $X_{T}$ as the number of HEADS and the number of TAILS in an outcome. For example, $X_{\mathrm{H}}($ HHTHTTTTHT $)=4$ and $X_{\mathrm{T}}($ HHTHTTTTHT $)=6$. Notice that $X=X_{\mathrm{H}}-X_{\mathrm{T}}$, i.e., for any $s \in S, X(s)=X_{\mathrm{H}}(s)-X_{\mathrm{T}}(s)$.


## A Non-Numerical Random Variable

- An example of a random variable whose values are not numbers:
- Let $(S, P)$ be the sample space representing ten tosses of a fair coin. For $s \in S$, let $Z(s)$ denote the set of positions where HEADS is observed. For example, $Z($ HHTHTTTTHT $)=\{1,2,4,9\}$ because the HEADS are in positions $1,2,4$, and 9 . We call $Z$ a set-valued random variable because $Z(s)$ is a set.
- The random variable $X_{\mathrm{H}}$ from the previous example is closely related to $Z$. We have $X_{\mathrm{H}}=|Z|$. This means that $X_{\mathrm{H}}(s)=|Z(s)|$, for all $s \in S$.


## Events Through Random Variables

- Let $X$ be a random variable defined on a sample space $(S, P)$. We might like to know the probability that $X$ assumes a value $v$.
- Example: If we roll a pair of dice, what is the probability that the sum of the numbers is 8 ?
We can express this question in two ways:
- Let $A$ be the event that the two dice sum to 8 ; that is, $A=\{(2,6)$, $(3,5),(4,4),(5,3),(6,2)\}$. We then ask: What is $P(A)$ ?
- Define a random variable $X$ to be the sum of the numbers on the dice. We can then ask: What is the probability that $X=8$ ? We write this as $P(X=8)$.
- In writing $P(X=8)$ we extend the $P(\bullet)$ notation beyond its previous scope. So far, we allowed two sorts of objects to follow the $P$.
- We may write $P(s)$ where $s \in S$;
- We may write $P(A)$ where $A$ is an event.
- The way to read the expression $P(X=8)$ is to interpret the " $X=8$ " as an event. So $X=8$ is shorthand for $\{s \in S: X(s)=8\}$.


## An Additional Example

- In the rolling of a pair of dice $P(X=8)=P(\{s \in S: X(s)=8\})=$ $P(\{(2,6),(3,5),(4,4),(5,3),(6,2)\})=\frac{5}{36}$.
- What does $P(X \geq 8)$ mean?
$P(X \geq 8)=P(\{s \in S: X(s) \geq 8\})=\frac{5+4+3+2+1}{36}=\frac{15}{36}=\frac{5}{12}$.
- Recall the random variables $X_{\mathrm{H}}$ and $X_{\mathrm{T}}$ that count the number of HEADS and of TAILS, respectively, in ten flips of a fair coin. What is the probability that there are at least four HEADS and at least four TAILS in ten flips of the coin?
This question can be expressed in these various ways:
- $P\left(X_{\mathrm{H}} \geq 4\right.$ and $\left.X_{\mathrm{T}} \geq 4\right)$
- $P\left(X_{\mathrm{H}} \geq 4 \wedge X_{\mathrm{T}} \geq 4\right)$
- $P\left(X_{\mathrm{H}} \geq 4 \cap X_{\mathrm{T}} \geq 4\right)$
- $P\left(4 \leq X_{\mathrm{H}} \leq 6\right)$.

We seek the probability of $\left\{s \in S: X_{\mathrm{H}}(s) \geq 4\right.$ and $\left.X_{\mathrm{T}}(s) \geq 4\right\}$. We have $P\left(X_{\mathrm{H}} \geq 4 \wedge X_{\mathrm{T}} \geq 4\right)=\frac{\binom{10}{4}+\binom{10}{5}+\binom{10}{6}}{2^{10}}=\frac{672}{1024}=\frac{21}{32}$.

## Binomial Random Variables

- Consider the unfair coin which produces HEADS with probability $p$ and TAILS with probability $1-p$.
- The coin is flipped $n$ times.
- Let $X$ denote the number of times that we see HEADS and $h$ be an integer. What is $P(X=h)$ ?
- If $h<0$ or $h>n$, it is impossible for $X(s)=h$, so $P(X=h)=0$.
- Suppose $0 \leq h \leq n$. There are exactly $\binom{n}{h}$ sequences of $n$ flips with exactly $h$ HEADS. All of these sequences have the same probability: $p^{h}(1-p)^{n-h}$. Therefore

$$
P(X=h)=\binom{n}{h} p^{h}(1-p)^{n-h} .
$$

- We call $X$ a binomial random variable for the following reason:
- If we expand the expression $(p+q)^{n}$ using the binomial theorem, one of the terms in the expansion is $\binom{n}{h} p^{h} q^{n-h}$.
- If we set $q=1-p$, this is exactly $P(X=h)$.


## Conditional Probabilities Using Random Variables

- Recall the pair-of-dice sample space.
- For this sample space, we define two random variables, $X_{1}$ and $X_{2}$
- $X_{1}(s)$ is the number on the first die, e.g., $X_{1}[(5,3)]=5$.
- $X_{2}(s)$ is the number on the second die, e.g., $X_{2}[(5,3)]=3$.
- Let $X=X_{1}+X_{2}$, i.e., $X$ is the sum of the numbers on the dice, e.g., $X[(5,3)]=8$.
- Knowledge of $X_{2}$ tells us some information about $X$.
- For example, if we know that $X_{2}(s)=4$, then $X(s)=4$ is impossible.
- If we know that $X_{2}(s)=4$, then the probability that $X(s)=5$ is $\frac{1}{6}$. We can express this as $P\left(X=5 \mid X_{2}=4\right)=\frac{1}{6}$. The meaning of $P\left(X=5 \mid X_{2}=4\right)$ is the usual meaning of conditional probability:

$$
P\left(X=5 \mid X_{2}=4\right)=\frac{P\left(X=5 \text { and } X_{2}=4\right)}{P\left(X_{2}=4\right)}=\frac{1 / 36}{1 / 6}=\frac{1}{6} .
$$

## Independent Random Variables

- We continue with the pair-of-dice and $X_{1}, X_{2}$ and $X=X_{1}+X_{2}$.
- Knowledge of $X_{2}$ tells us nothing about $X_{1}$. If $a$ and $b$ are integers from 1 to 6 , we have $P\left(X_{1}=a \mid X_{2}=b\right)=\frac{P\left(X_{1}=a \text { and } X_{2}=b\right)}{P\left(X_{2}=b\right)}=\frac{1}{6}$. Since $P\left(X_{1}=a\right.$ and $\left.X_{2}=b\right)=\frac{1}{36}=\frac{1}{6} \cdot \frac{1}{6}=P\left(X_{1}=a\right) P\left(X_{2}=b\right)$ the events " $X_{1}=a$ " and " $X_{2}=b$ " are independent.


## Definition (Independent Random Variables)

Let $(S, P)$ be a sample space and let $X$ and $Y$ be random variables defined on $(S, P)$. We say that $X$ and $Y$ are independent if, for all $a, b$,

$$
P(X=a \text { and } Y=b)=P(X=a) P(Y=b)
$$

- The random variables $X$ and $Y$ are functions defined on $(S, P)$. Therefore we may write $X: S \rightarrow A$ and $Y: S \rightarrow B$ for some sets $A$ and $B$. Since $X$ cannot take on values outside of $A$ nor can $Y$ outside of $B$, the condition can be rewritten

$$
\forall a \in A, \forall b \in B, P(X=a \text { and } Y=b)=P(X=a) P(Y=b)
$$

## Subsection 5

## Expectation

## Average Value

- When a random variable yields numerical results, we can ask questions such as:
- What is the average value this random variable might take?
- How widely spread are its values?
- The expected value of real-valued random variable can be interpreted as the average value of the random variable.
- Example: Recall the spinner:

Define $X$ to be the number of the region in which the pointer lands. Thus $P(X=1)=\frac{1}{2}$, $P(X=2)=\frac{1}{4}, P(X=3)=P(X=4)=\frac{1}{8}$.


- What is the average value of $X$ ? An incorrect reply would take the average $\frac{1+2+3+4}{4}=\frac{5}{2}$.


## Average Value: The Spinner

- The needle lands in region 1 far more often than in region 4.
- In spinning the pointer many times, many more 1's and 2's than 3's and 4's would show.
- So we would get an average value less than 2.5.
- If we were to spin the pointer a huge number $N$ times, we would expect to see (roughly) $\frac{N}{2} 1$ 's, $\frac{N}{4}$ 2's. $\frac{N}{8} 3$ 's and $\frac{N}{8} 4$ 's.
- So we would expect an average value of

$$
\frac{\frac{N}{2} \cdot 1+\frac{N}{4} \cdot 2+\frac{N}{8} \cdot 3+\frac{N}{8} \cdot 4}{N}=\frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{1}{2}=\frac{15}{8}<\frac{5}{2} .
$$

- What we have calculated is a weighted average of the values of $X$. The value $a$ is counted a number of times that is proportional to how often a appears.
- We call this weighted average of the values of $X$ the expected value or expectation of $X$.


## Expectation

## Definition (Expectation)

Let $X$ be a real-valued random variable defined on a sample space $(S, P)$. The expectation (or the expected value) of $X$ is

$$
E(X)=\sum_{s \in S} X(s) P(s)
$$

- The expected value of $X$ is also called the mean value of $X$.
- The letter $\mu$ is often used for the expectation of a random variable.
- Example: Let $X$ be the number that appears on the spinner. Its expected value is

$$
\begin{aligned}
E(X) & =\sum_{a=1}^{4} X(a) P(a) \\
& =X(1) P(1)+X(2) P(2)+X(3) P(3)+X(4) P(4) \\
& =1 \cdot \frac{1}{2}+2 \cdot \frac{1}{4}+3 \cdot \frac{1}{8}+4 \cdot \frac{1}{8}=\frac{15}{8} .
\end{aligned}
$$

## Examples

- A die is tossed. Let $X$ denote the number that shows. What is the expected value of $X$ ?
The expected value is $E(X)=\sum_{a=1}^{6} X(a) P(a)=X(1) P(1)+$ $X(2) P(2)+X(3) P(3)+X(4) P(4)+X(5) P(5)+X(6) P(6)=$ $1 \cdot \frac{1}{6}+2 \cdot \frac{1}{6}+3 \cdot \frac{1}{6}+4 \cdot \frac{1}{6}+5 \cdot \frac{1}{6}+6 \cdot \frac{1}{6}=\frac{1+2+3+4+5+6}{6}=3.5$.
- Suppose we roll a pair of dice. Let $X$ be the sum of the numbers on the two dice. What is the expected value of $X$ ?
There are 36 different outcomes in the set $S$. Instead of writing out all 36 terms in the sum $\sum_{s \in S} X(s) P(s)$, we collect like terms: $E(X)=2 P(X=2)+3 P(X=3)+\cdots+11 P(X=11)+12 P(X=12)$.
Therefore, $E(X)=$
$2 P(X=2)+3 P(X=3)+\cdots+11 P(X=11)+12 P(X=12)=$ $2 \frac{1}{36}+3 \frac{2}{36}+4 \frac{3}{36}+5 \frac{4}{36}+6 \frac{5}{36}+7 \frac{6}{36}+8 \frac{5}{36}+9 \frac{4}{36}+10 \frac{3}{36}+11 \frac{2}{36}+12 \frac{1}{36}=$ $\frac{2+6+12+20+30+42+40+36+30+22+12}{36}=\frac{252}{36}=7$.


## Collecting Like Terms

## Proposition

Let $(S, P)$ be a sample space and let $X$ be a real-valued random variable defined on $S$. Then $E(X)=\sum_{a \in \mathbb{R}} a P(X=a)$.

- Even though it seems we are computing an infinite sum, since $S$ is finite, there are only finitely many different values that $X(s)$ can actually take.
- Let $X$ be a real-valued random variable defined on a sample space $(S, P)$. The expected value of $X$ is $E(X)=\sum_{s \in S} X(s) P(s)$. We can rearrange the order of the terms by collecting those terms with a common value for $X(s): E(X)=\sum_{a \in \mathbb{R}}\left[\sum_{s \in S: X(s)=a} X(s) P(s)\right]$. Because $X(s)=a$ for all $s$ in the inner sum, $E(X)=$ $\sum_{a \in \mathbb{R}}\left[\sum_{s \in S: X(s)=a} a P(s)\right]=\sum_{a \in \mathbb{R}}\left[a \sum_{s \in S: X(s)=a} P(s)\right]=$
$\sum_{a \in \mathbb{R}}[a P(X=a)]$.


## A Spinning Game

- Consider a game in which we spin the spinner receiving $\$ 10$ for spinning an odd number and $\$ 20$ for spinning an even number. Let $X$ be the payout from this game. What is the expected value of $X$ ?


We calculate the answer in two ways.

- By Definition: $E(X)=\sum_{s \in S} X(s) P(s)=X(1) P(1)+X(2) P(2)+$ $X(3) P(3)+X(4) P(4)=10 \frac{1}{2}+20 \frac{1}{4}+10 \frac{1}{8}+20 \frac{1}{8}=\frac{110}{8}=13.75$.
- Alternatively, by the Proposition $E(X)=\sum_{a \in \mathbb{R}} a P(X=a)=$

$$
10 \cdot P(X=10)+20 \cdot P(X=20)=10 \frac{5}{8}+20 \frac{3}{8}=\frac{110}{8}=13.75
$$

- If we play this game repeatedly, we expect to receive an average of $\$ 13.75$ per spin.


## Rolling a Pair of Dice

- Consider again the pair-of-dice sample space. The random variable $X$ is the absolute value of the difference of the numbers on the two dice. What is the expected value of $X$ ?

$$
\begin{aligned}
E(X)= & \sum_{a \in \mathbb{R}} a P(X=a) \\
= & 0 \cdot P(X=0)+1 \cdot P(X=1)+2 \cdot P(X=2) \\
& +3 \cdot P(X=3)+4 \cdot P(X=4)+5 \cdot P(X=5) \\
= & 0 \frac{6}{36}+1 \frac{10}{36}+2 \frac{8}{36}+3 \frac{3}{36}+4 \frac{4}{36}+5 \frac{2}{36} \\
= & \frac{10+16+18+16+10}{36}=\frac{70}{36} \approx 1.944 .
\end{aligned}
$$

## Algebraic Combinations of Random Variables

- Suppose $X$ and $Y$ are real-valued random variables defined on a sample space $(S, P)$.
- We can form a new random variable $Z=X+Y$, i.e., the value of $Z$ evaluated at $s$ is simply the sum of the values $X(s)$ and $Y(s)$.
- Similarly, if $X$ and $Y$ are real-valued random variables on a sample space $(S, P)$, then $X Y$ is the random variable whose value at $s$ is $X(s) Y(s)$.
- Likewise we can define $X-Y$ and so on.
- If $c$ is a number and $X$ is a real-valued random variable, then $c X$ is the random variable whose value at $s$ is $c X(s)$.
- If we know the expected value of $X$ and $Y$, can we determine the expected value of $X+Y, X Y$, or some other algebraic combination of $X$ and $Y$ ?


## Expectation of a Sum of Random Variables

## Theorem (Expectation of a Sum)

Suppose $X$ and $Y$ are real-valued random variables defined on a sample space $(S, P)$. Then $E(X+Y)=E(X)+E(Y)$.

- Let $Z=X+Y$. We have $E(Z)=\sum_{s \in S} Z(s) P(s)=$ $\sum_{s \in S}[X(s)+Y(s)] P(s)=\sum_{s \in S}[X(s) P(s)+Y(s) P(s)]=$ $\sum_{s \in S} X(s) P(s)+\sum_{s \in S} Y(s) P(s)=E(X)+E(Y)$.
- Example: Let $(S, P)$ be the pair-of-dice sample space and let $Z$ be the random variable giving the sum of the values on the two dice.
What is $E(X)$ ?
Let $X_{1}$ be the value on the first die and $X_{2}$ the value on the second. Note that $Z=X_{1}+X_{2}$. We have computed $E\left(X_{1}\right)=E\left(X_{2}\right)=\frac{7}{2}$, so $E(Z)=E\left(X_{1}\right)+E\left(X_{2}\right)=7$.


## Another Example

- A basket holds 100 chips that are labeled 1 through 100 . Two chips are drawn at random from the basket (without replacement). What is the expected value of their sum, $X$ ?
There are three ways we can approach this problem:
- First, we can apply the definition of expectation to find $E(X)=\sum_{s \in S} X(s) P(s)$. This summation involves 9900 terms (there are 100 choices for the first chip times 99 choices for the second chip).
- Second, we can compute $E(X)=\sum_{a \in \mathbb{R}} a P(X=a)$. The possible sums range from 3 to 199 , so this sum has nearly 200 terms.
- Third, let $X_{1}$ be the number on the first chip and $X_{2}$ the number on the second chip. Both $X_{1}$ and $X_{2}$ can be any value from 1 to 100 and they are all equally likely. So $E\left(X_{1}\right)=E\left(X_{2}\right)=\frac{1+2+\cdots+100}{100}=\frac{5050}{100}=$ 50.5. Since $X=X_{1}+X_{2}$, we have

$$
E(X)=E\left(X_{1}+X_{2}\right)=E\left(X_{1}\right)+E\left(X_{2}\right)=50.5+50.5=101 .
$$

- It is important to note that $X_{1}$ and $X_{2}$ are dependent random variables but the last proposition is still applicable.


## Expectation of a Real Multiple of a Random Variable

- What happens in the case of multiplication of random variables?
- Consider, first, a real-valued random variable $X$ on a sample space $(S, P)$, and a real number $c$. The expected value of $c X$ is $E(c X)=\sum_{s \in S}(c X)(s) P(s)=\sum_{s \in S}[c X(s)] P(s)=$ $c \sum_{s \in S} X(s) P(s)=c E(X)$.


## Proposition

Let $X$ be a real-valued random variable on a sample space $(S, P)$ and let $c$ be a real number. Then $E(c X)=c E(X)$.

- Restated the Proposition says that, if the average value of $X$ is some number $a$, then the average value of $c X$ is $c a$.


## Linearity of Expectation

## Theorem (Linearity of Expectation)

Suppose $X$ and $Y$ are real-valued random variables on a sample space $(S, P)$ and suppose $a$ and $b$ are real numbers. Then

$$
E(a X+b Y)=a E(X)+b E(Y)
$$

- $E(a X+b Y)=E(a X)+E(b Y)=a E(X)+b E(Y)$.
- In fact, if $X_{1}, X_{2}, \ldots, X_{n}$ are random variables defined on $(S, P)$, and $c_{1}, c_{2}, \ldots, c_{n}$ are real numbers, then

$$
E\left[c_{1} X_{1}+\cdots+c_{n} X_{n}\right]=c_{1} E\left[X_{1}\right]+\cdots+c_{n} E\left[X_{n}\right]
$$

## Example: Tossing a Coin 10 Times

- A coin is tossed 10 times. Let $X$ be the number of times we observe TAILS immediately after seeing HEADS. What is the expected value of $X$ ?
To compute $E(X)$, we express $X$ as the sum of other random variables whose expectations are easier to calculate.
- Let $X_{1}$ be the random variable whose value is one if the first two tosses are HEADS-TAILS and is zero otherwise.
- Let $X_{2}$ be the random variable that is one if the second and third tosses come up HEADS-TAILS and is zero otherwise.
- More generally, let $X_{k}$ be the random variable defined as follows: $X_{k}= \begin{cases}1, & \text { if toss } k \text { is HEADS and toss } k+1 \text { is TAILS } \\ 0, & \text { otherwise }\end{cases}$
The random variable $X_{k}$ can take on only two values, one and zero, so $E\left(X_{k}\right)=0 P(X=0)+1 P(X=1)=P(X=1)$ and the probability we see HEADS-TAILS in positions $k, k+1$ is exactly $\frac{1}{4}$. Therefore $E\left(X_{k}\right)=\frac{1}{4}$ for each $k$ with $1 \leq k \leq 9$.
Now we get $E(X)=E\left(X_{1}\right)+\cdots+E\left(X_{9}\right)=\frac{9}{4}$.


## Indicator Random Variables

- Indicator random variables take on only two values: zero and one. Such random variables are also called zero-one random variables.


## Proposition

Let $X$ be a zero-one random variable. Then $E(X)=P(X=1)$.

- Example: Let $\pi$ be a random permutation of the numbers $\{1,2, \ldots, n\}$. In other words, the sample space is $\left(S_{n}, P\right)$ where all permutations $\pi \in S_{n}$ have probability $P(\pi)=\frac{1}{n!}$. Let $X(\pi)$ be the number of values $k$ such that $\pi(k)=k$. Such a value $k$ is called a fixed point of the permutation. What is the expected value of $X$ ? For $k$ with $1 \leq k \leq n$, let $X_{k}(\pi)=\left\{\begin{array}{ll}1, & \text { if } \pi(k)=k \\ 0, & \text { otherwise }\end{array}\right.$ Note that $X=X_{1}+X_{2}+\cdots+X_{n}$. Since $X_{k}$ is a zero-one random variable, $E\left(X_{k}\right)=P\left(X_{k}=1\right)=\frac{1}{n}$. Therefore $E(X)=E\left(X_{1}\right)+\cdots+E\left(X_{n}\right)=$ $\frac{1}{n}+\cdots+\frac{1}{n}=1$.
On average, a random permutation has exactly one fixed point.


## Product of Independent Random Variables: Example

- A pair of dice are tossed. Let $X$ be the product of the numbers on the two dice. What is the expected value of $X$ ?
We can express $X$ as the product of $X_{1}$ (the number on the first die) and $X_{2}$ (the number on the second die), with $E\left(X_{1}\right)=E\left(X_{2}\right)=\frac{7}{2}$. We evaluate $E(X)$ by computing $\sum_{\in \mathbb{R}} a P(X=a)$ :

| $a$ | $P(X=a)$ | $a P(X=a)$ | $a$ | $P(X=a)$ | $a P(X=a)$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | $1 / 36$ | $1 / 36$ | 12 | $4 / 36$ | $48 / 36$ |
| 2 | $2 / 36$ | $4 / 36$ | 15 | $2 / 36$ | $30 / 36$ |
| 3 | $2 / 36$ | $6 / 36$ | 16 | $1 / 36$ | $16 / 36$ |
| 4 | $3 / 36$ | $12 / 36$ | 18 | $2 / 36$ | $36 / 36$ |
| 5 | $2 / 36$ | $10 / 36$ | 20 | $2 / 36$ | $40 / 36$ |
| 6 | $4 / 36$ | $24 / 36$ | 24 | $2 / 36$ | $48 / 36$ |
| 8 | $2 / 36$ | $16 / 36$ | 25 | $1 / 36$ | $25 / 36$ |
| 9 | $1 / 36$ | $9 / 36$ | 30 | $2 / 36$ | $60 / 36$ |
| 10 | $2 / 36$ | $20 / 36$ | 36 | $1 / 36$ | $36 / 36$ |
|  |  | $102 / 36$ |  |  | $339 / 36$ |

Therefore $E(X)=\frac{441}{36}=\left(\frac{7}{2}\right)^{2}$. So in this case $E(X)=E\left(X_{1}\right) E\left(X_{2}\right)$.

## Product of Dependent Random Variables: Example

- A fair coin is tossed twice. Let $X_{H}$ be the number of HEADS and let $X_{\mathrm{T}}$ be the number of TAILS observed. Let $Z=X_{\mathrm{H}} X_{\mathrm{T}}$. What is $E(Z)$ ?
Note that $E\left(X_{\mathrm{H}}\right)=E\left(X_{\mathrm{T}}\right)=2 \frac{1}{4}+1 \frac{1}{2}+0 \frac{1}{4}=1$.
However, we get

$$
\begin{aligned}
E(Z) & =\sum_{a \in \mathbb{R}} a P(Z=a) \\
& =0 \cdot P(Z=0)+1 \cdot P(Z=1) \\
& =P(Z=1) \\
& =P(\{\mathrm{HT}, \mathrm{TH}\})=\frac{1}{2} .
\end{aligned}
$$

This example shows that $E(X Y)=E(X) E(Y)$ is incorrect in general.

## Expectation: Product of Independent Random Variables

## Theorem (Product of Independent Random Variables)

Let $X$ and $Y$ be independent, real-valued random variables defined on a sample space $(S, P)$. Then $E(X Y)=E(X) E(Y)$.

- Let $Z=X Y$. Then

$$
\begin{aligned}
E(Z) & =\sum_{a \in \mathbb{R}} a P(Z=a) \\
& =\sum_{a \in \mathbb{R}} a\left[\sum_{b, c \in \mathbb{R}: b c=a} P(X=b \wedge Y=c)\right] \quad(Z=X Y) \\
& =\sum_{a \in \mathbb{R}} a\left[\sum_{b, c \in \mathbb{R}: b c=a} P(X=b) P(Y=c)\right] \quad \text { (Indep.) } \\
& =\sum_{a \in \mathbb{R}}\left[\sum_{b, c \in \mathbb{R}: b c=a} b c P(X=b) P(Y=c)\right] \quad(a=b c) \\
& =\sum_{b \in \mathbb{R}}\left[\sum_{c \in \mathbb{R}} b P(X=b) c P(Y=c)\right] \\
& =\sum_{b \in \mathbb{R}} b P(X=b)\left[\sum_{c \in \mathbb{R}} c P(Y=c)\right] \\
& =\left[\sum_{b \in \mathbb{R}} b P(X=b)\right]\left[\sum_{c \in \mathbb{R}} c P(Y=c)\right] \\
& =E(X) E(Y) .
\end{aligned}
$$

## Counterexample for the Converse

- If $X$ and $Y$ satisfy $E(X Y)=E(X) E(Y)$, we may not conclude that $X$ and $Y$ are independent! The following example showcases this:
- Example: Let $(S, P)$ be the sample space with $S=\{a, b, c\}$ in which all three elements have probability $\frac{1}{3}$. Define random variables $X$ and $Y$ according to:

| $s$ | $X(s)$ | $Y(s)$ |
| :---: | :---: | :---: |
| $a$ | 1 | 0 |
| $b$ | 0 | 1 |
| $c$ | -1 | 0 |

Note that $X$ and $Y$ are not independent because $P(X=0)=\frac{1}{3}$, $P(Y=0)=\frac{2}{3}$ and $P(X=0 \wedge Y=0)=0 \neq P(X=0) P(Y=0)$. On the other hand, for all $s \in S$, we have $X(s) Y(s)=0$. Therefore $E(X)=0, E(Y)=\frac{1}{3}$ and $E(X Y)=0=E(X) E(Y)$.

## Expected Value as a Measure of Centrality

- The expected value of a real-valued random variable is in the "middle" of all the values $X(s)$.
- Example: Consider the sample space $(S, P)$ where $S=\{1,2, \ldots, 10\}$ and $P(s)=\frac{1}{10}$ for all $s \in S$. Define $X$ by:

| $s$ | $X(s)$ | $s$ | $X(s)$ |
| :---: | :---: | :---: | :---: |
| 1 | 1 | 6 | 2 |
| 2 | 1 | 7 | 8 |
| 3 | 1 | 8 | 8 |
| 4 | 1 | 9 | 8 |
| 5 | 2 | 10 | 8 |

Then

$$
E(X)=\sum_{a \in \mathbb{R}} a P(X=a)=1 \cdot 0.4+2 \cdot 0.2+8 \cdot 0.4=4
$$

## Measuring of Centrality: Another Example

- Imagine a long horizontal plank along which we place weights. We place a weight at position a provided $P(X=a)>0$. The weight we place at a is $P(X=a)$ kilograms. For the random variable $X$ of the table in the previous slide we get


At what point does this device balance?
If it balances at a point $\ell$, the amounts of twist-torque applied in either side should balance. Therefore:

$$
\begin{aligned}
& \sum_{a \in \mathbb{R}} P(X=a)(a-\ell)=0 \\
& \Rightarrow \quad \sum_{a \in \mathbb{R}} a P(X=a)=\ell \sum_{a \in \mathbb{R}} P(X=a)=\ell \\
& \Rightarrow \quad \ell=E(X) .
\end{aligned}
$$

In the figure, the balancing point is at $\ell=4$.

## Variance: The Main Idea

- Consider three random variables $X, Y$, and $Z$. They take on real values as follows:

| $X$ | $P(X)$ |
| :---: | :---: |
| -2 | $\frac{1}{2}$ |
| 2 | $\frac{1}{2}$ |


| $Y$ | $P(Y)$ |
| :---: | :---: |
| -10 | 0.001 |
| 0 | 0.998 |
| 10 | 0.001 |


| $Z$ | $P(Z)$ |
| :---: | :---: |
| -5 | $\frac{1}{3}$ |
| 0 | $\frac{1}{3}$ |
| 5 | $\frac{1}{3}$ |

All three random variables have an expected value equal to zero.
Which of these is more "spread out"?
We calculate how far away each value of $X$ is from $\mu=E(X)$, but count it only proportional to its probability. That is, we add up $[X(s)-\mu] P(s)$. Unfortunately,
$\sum_{s \in S}[X(s)-\mu] P(s)=\sum_{s \in S} X(s) P(s)-\sum_{s \in S} \mu P(s)=\mu-\mu \cdot 1=0$.
The problem is that values to the right of $\mu$ are exactly canceled by values to the left. To prevent this cancelation, we square the distances between $X$ and $\mu$, counting them proportional to their probability.

## Variance: Finishing the Example

## Definition (Variance)

Let $X$ be a real-valued random variable on a sample space $(S, P)$ and $\mu=E(X)$. The variance of $X$ is $\operatorname{Var}(X)=E\left[(X-\mu)^{2}\right]$.

- Consider again $X, Y$, and $Z$ :

| $X$ | $P(X)$ |
| :---: | :---: |
| -2 | $\frac{1}{2}$ |
| 2 | $\frac{1}{2}$ |


| $Y$ | $P(Y)$ |
| :---: | :---: |
| -10 | 0.001 |
| 0 | 0.998 |
| 10 | 0.001 |


| $Z$ | $P(Z)$ |
| :---: | :---: |
| -5 | $\frac{1}{3}$ |
| 0 | $\frac{1}{3}$ |
| 5 | $\frac{1}{3}$ |

We calculate their variances as follows:

$$
\begin{aligned}
& \operatorname{Var}(X)=E\left[(X-\mu)^{2}\right]=E\left(X^{2}\right)=(-2)^{2} \cdot 0.5+2^{2} \cdot 0.5=4 \\
& \operatorname{Var}(Y)=E\left[(Y-\mu)^{2}\right]=E\left(Y^{2}\right)=0.2 \\
& \operatorname{Var}(Z)=E\left[(Z-\mu)^{2}\right]=E\left(Z^{2}\right)=\frac{50}{3} \approx 16.67 .
\end{aligned}
$$

By this measure, $Z$ is the most spread out and $Y$ is the most concentrated.

## Another Example

- A die is tossed. Let $X$ denote the number that appears on the die. What is the variance of $X$ ?
Let $\mu=E(X)=\frac{7}{2}$. Then

$$
\begin{aligned}
\operatorname{Var}(X)= & E\left[(X-\mu)^{2}\right] \\
= & E\left[\left(X-\frac{7}{2}\right)^{2}\right] \\
= & \left(1-\frac{7}{2}\right)^{2} \cdot \frac{1}{6}+\left(2-\frac{7}{2}\right)^{2} \cdot \frac{1}{6}+\left(3-\frac{7}{2}\right)^{2} \cdot \frac{1}{6} \\
& +\left(4-\frac{7}{2}\right)^{2} \cdot \frac{1}{6}+\left(5-\frac{7}{2}\right)^{2} \cdot \frac{1}{6}+\left(1-\frac{7}{2}\right)^{2} \cdot \frac{1}{6} \\
= & \frac{25}{24}+\frac{3}{8}+\frac{1}{24}+\frac{1}{24}+\frac{3}{8}+\frac{25}{24} \\
= & \frac{35}{12} \approx 2.9167 .
\end{aligned}
$$

## A Proposition

## Proposition

Let $X$ be a real-valued random variable. Then

$$
\operatorname{Var}(X)=E\left[X^{2}\right]-E[X]^{2}
$$

- Let $\mu=E(X)$. By definition, $\operatorname{Var}(X)=E\left[(X-\mu)^{2}\right]$. We can write $(X-\mu)^{2}=X^{2}-2 \mu X+\mu^{2}$. We can think of this as the sum of three random variables: $X^{2},-2 \mu X$ and $\mu^{2}$. If we evaluate these at an element $s$ of the sample space, we get $[X(s)]^{2},-2 \mu X(s)$, and $\mu^{2}$, respectively. Since, as a random variable, the value of $\mu^{2}$ at every $s$ is the constant $\mu^{2}, E\left(\mu^{2}\right)=\mu^{2}$. Now, we calculate

$$
\begin{aligned}
\operatorname{Var}(X) & =E\left[(X-\mu)^{2}\right]=E\left[X^{2}-2 \mu X+\mu^{2}\right] \\
& =E\left[X^{2}\right]-2 \mu E[X]+E\left[\mu^{2}\right]=E\left[X^{2}\right]-2 \mu^{2}+\mu^{2} \\
& =E\left[X^{2}\right]-\mu^{2}=E\left[X^{2}\right]-E[X]^{2} .
\end{aligned}
$$

## Another Example

- Let $X$ be the number showing on a random toss of a die. What is $\operatorname{Var}(X)$ ?
We get $\operatorname{Var}(X)=E\left[X^{2}\right]-E[X]^{2}$. Note that

$$
E[X]^{2}=\left(\frac{7}{2}\right)^{2}=\frac{49}{4}
$$

Also,

$$
\begin{aligned}
E\left[X^{2}\right] & =1^{2} \cdot \frac{1}{6}+2^{2} \cdot \frac{1}{6}+3^{2} \cdot \frac{1}{6}+4^{2} \cdot \frac{1}{6}+5^{2} \cdot \frac{1}{6}+6^{2} \cdot \frac{1}{6} \\
& =\frac{1^{2}+2^{2}+3^{2}+4^{2}+5^{2}+6^{2}}{6} \\
& =\frac{91}{6} .
\end{aligned}
$$

Therefore $\operatorname{Var}(X)=E\left[X^{2}\right]-E[X]^{2}=\frac{91}{6}-\frac{49}{4}=\frac{35}{12}$.

## Binomial Random Variables

- Consider an unfair coin that produces HEADS with probability $p$ and TAILS with probability $1-p$ flipped $n$ times. Let $X$ denote the number of times we see HEADS. We have $E(X)=n p$. What is the variance of $X$ ?
We can express $X$ as the sum of the zero-one indicator random variables $X_{j}=\left\{\begin{array}{ll}1, & \text { if the } j \text { th flip comes up HEADS } \\ 0, & \text { if the } j \text { th flip comes up TAILS }\end{array}\right.$ Then $X=X_{1}+X_{2}+\cdots+X_{n}$. Since $E[X]=n p$, we have $E[X]^{2}=n^{2} p^{2}$. To calculate $E\left[X^{2}\right]$ notice that $X^{2}=\left[X_{1}+X_{2}+\cdots+X_{n}\right]^{2}=$ $X_{1} X_{1}+X_{1} X_{2}+\cdots+X_{1} X_{n}+X_{2} X_{1}+\cdots+X_{n} X_{n}=\sum_{i=1}^{n} X_{i}^{2}+\sum_{i \neq j} X_{i} X_{j}$. By linearity of expectation

$$
\begin{aligned}
E\left[X^{2}\right] & =E\left[\sum_{i=1}^{n} X_{i}^{2}+\sum_{i \neq j} X_{i} X_{j}\right] \\
& =\sum_{i=1}^{n} E\left[X_{i}^{2}\right]+\sum_{i \neq j} E\left[X_{i} X_{j}\right]=n p+n(n-1) p^{2}
\end{aligned}
$$

Therefore, $\operatorname{Var}[X]=E\left[X^{2}\right]-E[X]^{2}=n p+n(n-1) p^{2}-n^{2} p^{2}=$ $n p+n^{2} p^{2}-n p^{2}-n^{2} p^{2}=n p-n p^{2}=n p(1-p)$.

