Computer Science

CSC535: Probabilistic Graphical Models

Probability Primer : Discrete Probability

Prof. Jason Pacheco

Administrative Items

Homework 1

- Out now on D2L
- 4 Problems, worth a total of 5 points on final grade
- Due next Wed, Jan 26 @ 11:59pm (1 week)

Assignment Submissions

- All assignments will be via D2L
- I tried to use Gradescope and it doesn’t easily support the style of assignment in this class (PDF Report + code)

Office Hours

- Fridays, 3-5pm (Zoom)
- Links available in D2L calendar

Outline

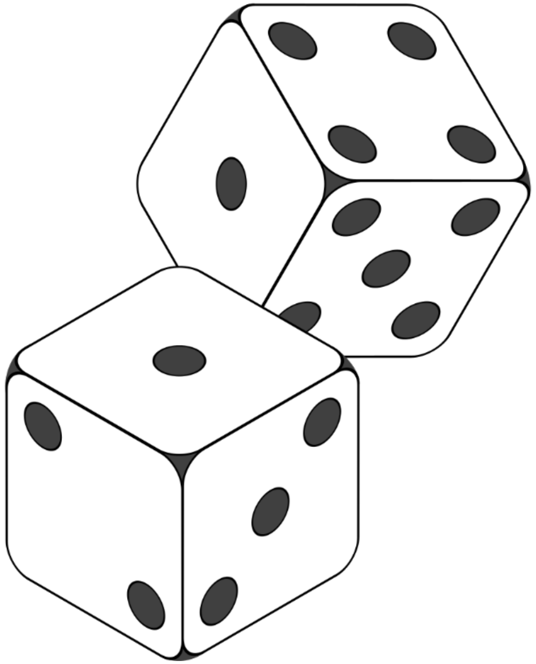
- Random Events and Probability
- Random Variables
- Fundamental Rules of Probability
- Moments and Dependence of Random Variables
- Useful Discrete Distributions

Outline

- Random Events and Probability
- Random Variables
- Fundamental Rules of Probability
- Moments and Dependence of Random Variables
- Useful Discrete Distributions

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?

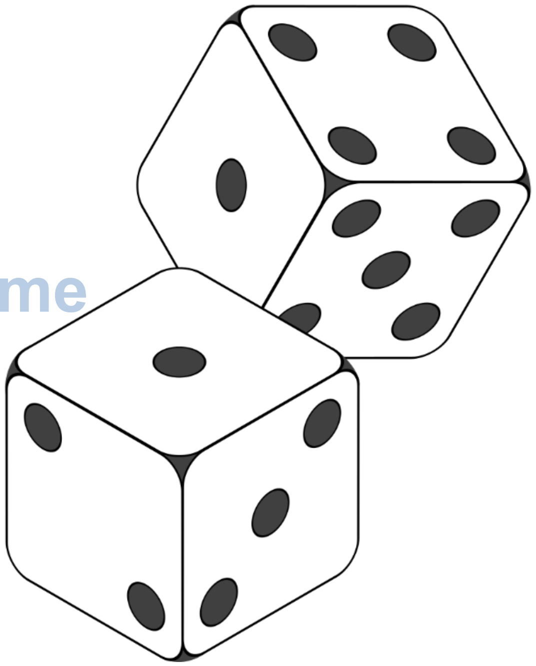
***…probability theory** gives a mathematical formalism to addressing such questions…*

Definition An **experiment** or **trial** is any process that can be repeated with well-defined outcomes. It is *random* if more than one outcome is possible,

Example Roll two fair dice

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 (1,1), (1,2), …, (6,1), (6,2), …, (6,6)

Sample Space

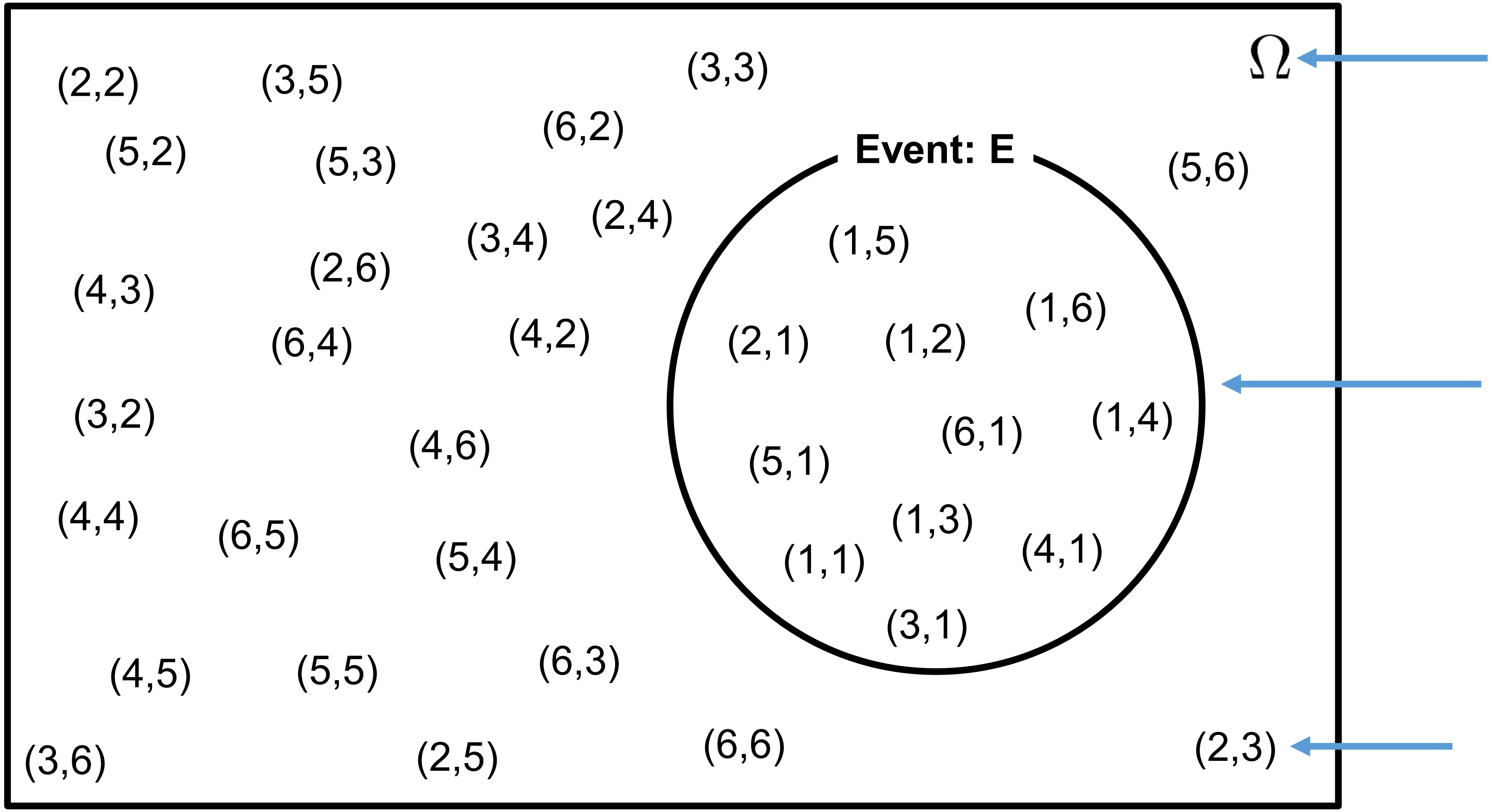
Definition An **event** is a *set* of outcomes (a subset of the sample space),

Example Event Roll at least a single 1

{(1,1), (1,2), (1,3), …, (1,6), …, (6,1)}

Random Events and Probability

*Can formulate / visualize as **sets** 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 $\omega \in \Omega$

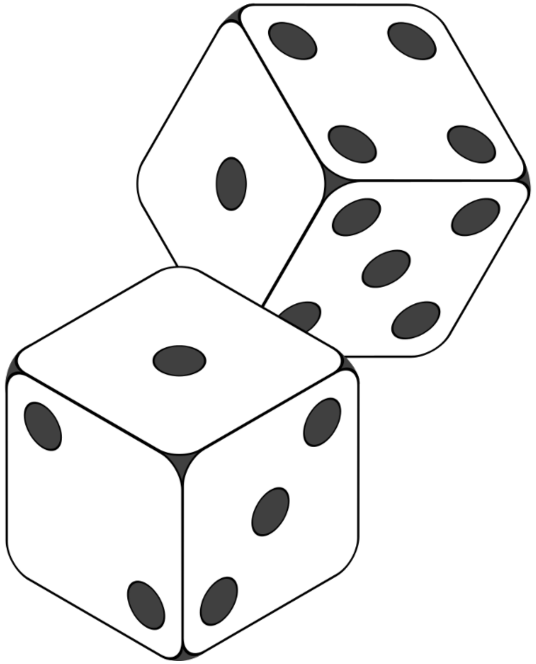
Random Events and Probability

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 greater than 12,

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

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

i.e. we can reason about the probability of impossible outcomes

Random Events and Probability

Events are closed under set operations

E_1 : *First* die equals 1

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

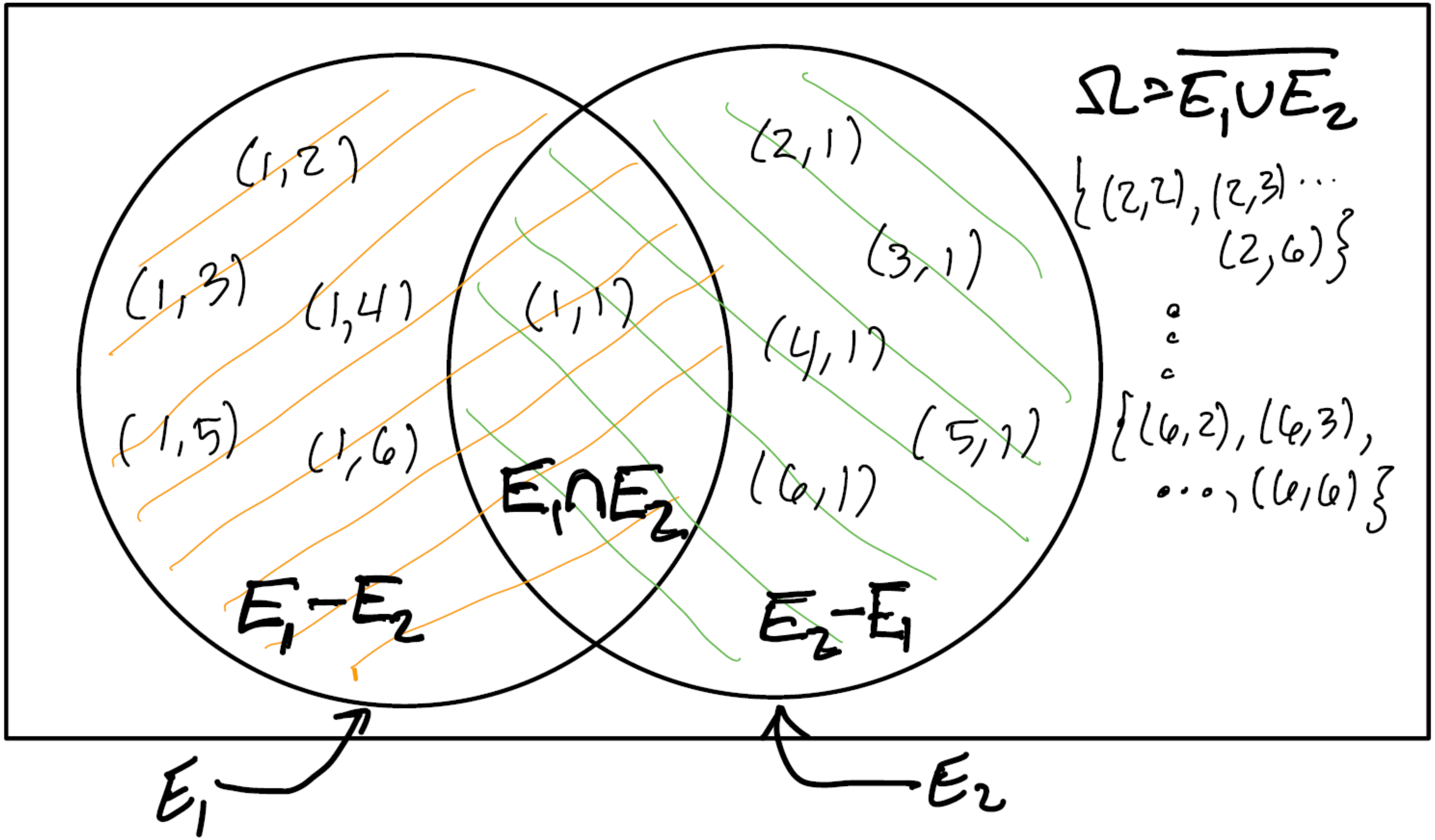
E_2 : *Second* die equals 1

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

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
$\overline{E_1 \cup E_2}$	$\{(2,2),(2,3),\ldots,(2,6),(3,2),\ldots,(6,6)\}$	No die rolls 1

Random Events and Probability

Can interpret these operations as a Venn diagram…



Random Events and Probability

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

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

Proof:

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

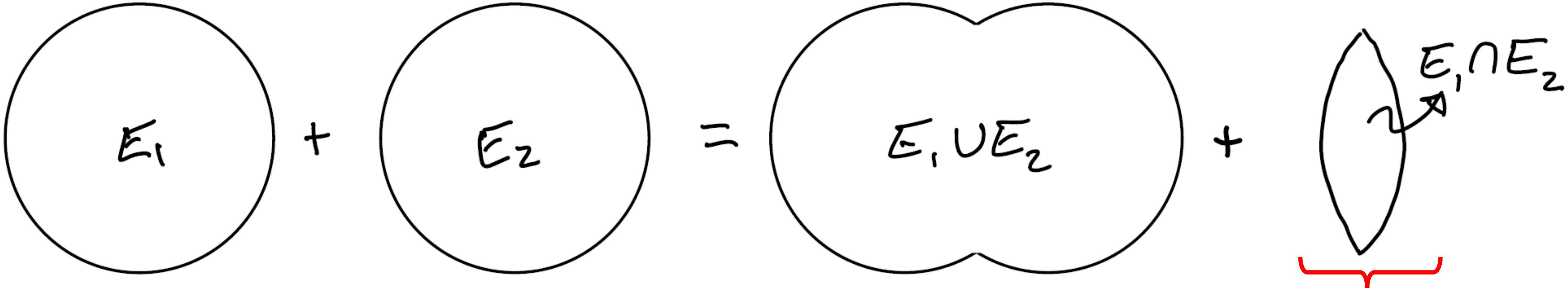
$$P(E_2) = P(E_2 - (E_1 \cap E_2)) + P(E_1 \cap E_2)$$

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

Lemma: *For any two events* E_1 *and* E_2 ,

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

Graphical Proof:



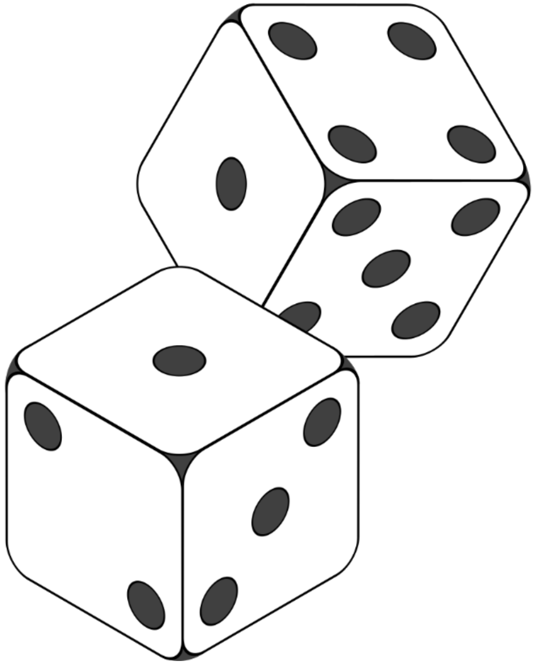
Subtract from both sides to avoid double counting

Outline

- Random Events and Probability
- Random Variables
- Fundamental Rules of Probability
- Moments and Dependence of Random Variables
- Useful Discrete Distributions

Random Variables

Suppose we are interested in a distribution over the sum of dice…



Option 1 Let E_i be event that the sum equals *i*

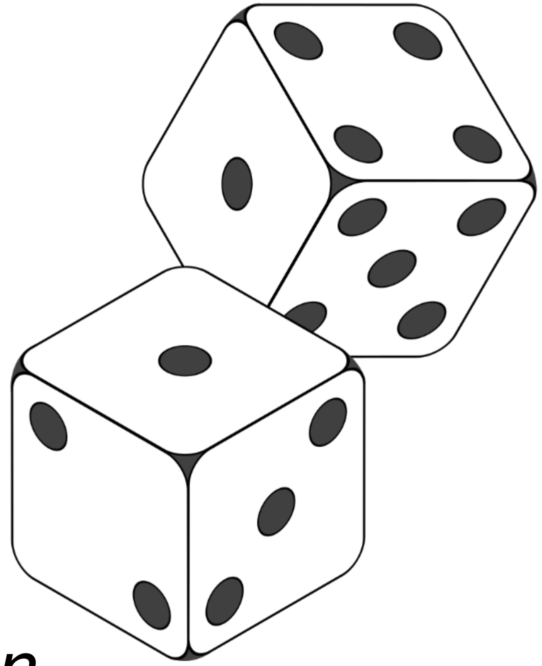
Two dice example:

$$E_2 = \{(1,1)\} \quad E_3 = \{(1,2),(2,1)\} \quad E_4 = \{(1,3),(2,2),(3,1)\}$$

$$E_5 = \{(1,4),(2,3),(3,2),(4,1)\} \quad E_6 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

Enumerate all possible means of obtaining desired sum. Gets cumbersome for N>2 dice…

Random Variables



Suppose we are interested in a distribution over the sum of dice…

Option 2 Use a function of sample space…

Definition *A* **random variable** X *is a real-valued function* $X : \Omega \rightarrow \mathbb{R}$. *We say X is a* ***discrete random variable*** *if it takes on only a finite or countably infinite number of values.*

Example X is the *sum of two dice* with values,

$$X \in \{2, 3, 4, \ldots, 12\}$$

Discrete vs. Continuous Probability

- Discrete RVs take on a finite or countably infinite set of values; continuous RVs take an uncountably infinite set of values

- Representing / interpreting / computing probabilities becomes more complicated in the continuous setting

- Thus, we will focus on discrete RVs for now… it will simplify presentation of the fundamental rules of probability and probability measures

Random Variables and Probability

Capitol letters represent random variables

Lowercase letters are realized *values*

$$X = x$$

$X = x$ is the **event** that X takes the value x

Example Let X be the random variable (RV) representing the sum of two dice with values,

$$X \in \{2, 3, 4, \ldots, 12\}$$

X=5 is the *event* that the dice sum to 5,

$$X = 5 \quad \Leftrightarrow \quad E^5 = \{(1,4), (2,3), (3,2), (4,1)\}$$

Random Variables and Probability

For *discrete* RVs $X = x$ is an **event** with **probability mass function**:

$$p(X = x) = \sum_{\omega \in \Omega \,:\, X(\omega) = x} P(\omega)$$

All outcomes that cause the event X=x

Example Let X be the sum of two fair dice. The probability of rolling a sum of 6 is:

“fair” is code for “uniform distribution”

$$p(X = 6) = P((1,5)) + P((2,4)) + P((3,3)) + P((4,2)) + P((5,1))$$

$$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{5}{36}$$

Probability Mass Function

A function $p(X)$ is a **probability mass function (PMF)** of a discrete random variable if the following conditions hold:

(a) It is nonnegative for all values in the support,

$$p(X = x) \geq 0$$

(b) The sum over all values in the support is 1,

$$\sum_x p(X = x) = 1$$

Intuition Probability mass is conserved, just as in physical mass. Reducing probability mass of one event must increase probability mass of other events so that the definition holds...

Probability Mass Function

Example Let X be the outcome of a single fair die. It has the PMF,

$$p(X = x) = \frac{1}{36} \qquad \text{for } x = 1, \ldots, 6$$

Uniform Distribution

Example We can often represent the PMF as a vector. Let S be an RV that is the *sum of two fair dice*. The PMF is then,

$$p(S) = \begin{pmatrix} p(S=2) \\ p(S=3) \\ p(S=4) \\ \vdots \\ p(S=12) \end{pmatrix} = \begin{pmatrix} 1/36 \\ 1/18 \\ 1/2 \\ \vdots \\ 1/36 \end{pmatrix}$$

Observe that S does <u>not</u> follow a uniform distribution

Functions of Random Variables

Any function $f(X)$ of a random variable X is also a random variable and it has a probability distribution

Example Let X_1 be an RV that represents the result of a fair die, and let X_2 be the result of another fair die. Then,

$$S = X_1 + X_2$$

Is an RV that is the *sum of two fair dice* with PMF $p(S)$.

NOTE Even if we know the PMF $p(X)$ and we know that the PMF $p(f(X))$ exists, it is not always easy to calculate!

PMF Notation

- We use $P(E)$ for the probability distribution of events, but lowercase $p(X)$ for PMF for reasons that will be clear later
- We use $p(X)$ to refer to the probability mass *function* (i.e. a function of the RV X)
- We use $p(X=x)$ to refer to the probability of the *event* $X=x$
- We will often use $p(x)$ as shorthand for $p(X=x)$

Outline

- Random Events and Probability
- Random Variables
- Fundamental Rules of Probability
- Moments and Dependence of Random Variables
- Useful Discrete Distributions

Joint Probability

Definition Two (discrete) RVs X and Y have a *joint PMF* denoted by $p(X, Y)$ and the probability of the event X=x and Y=y denoted by $p(X = x, Y = y)$ where,

(a) It is nonnegative for all values in the support,

$$p(X = x, Y = y) \geq 0$$

(b) The sum over all values in the support is 1,

$$\sum_x \sum_y p(X = x, Y = y) = 1$$

Joint Probability

Let X and Y be *binary RVs.* We can represent the joint PMF p(X,Y) as a 2x2 array (table):

		Y = 0	Y = 1
X	0	0.04	0.36
	1	0.30	0.30

All values are nonnegative

Joint Probability

Let X and Y be *binary RVs.* We can represent the joint PMF p(X,Y) as a 2x2 array (table):

		Y = 0	Y = 1
X	0	0.04	0.36
	1	0.30	0.30

The sum over all values is 1:
0.04 + 0.36 + 0.30 + 0.30 = 1

Joint Probability

Let X and Y be *binary RVs.* We can represent the joint PMF p(X,Y) as a 2x2 array (table):

		Y = 0	Y = 1
X	0	0.04	0.36
	1	0.30	0.30

P(X=1, Y=0) = 0.30

Fundamental Rules of Probability

Given two RVs X and Y the **conditional distribution** is:

$$p(X \mid Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(X,Y)}{\sum_x p(X=x,Y)}$$

Multiply both sides by $p(Y)$ to obtain the **probability chain rule**:

$$p(X,Y) = p(Y)p(X \mid Y)$$

For N RVs $X_1, X_2, \ldots, X_N$:

$$p(X_1, X_2, \ldots, X_N) = p(X_1)p(X_2 \mid X_1) \ldots p(X_N \mid X_{N-1}, \ldots, X_1)$$

$$= p(X_1) \prod_{i=2}^{N} p(X_i \mid X_{i-1}, \ldots, X_1)$$

Chain rule valid for any ordering

Fundamental Rules of Probability

Law of total probability

$$p(Y) = \sum_x p(Y, X = x)$$

Proof $$\sum_x p(Y, X = x) = \sum_x p(Y) p(X = x \mid Y) \quad \text{(chain rule)}$$

$$= p(Y) \sum_x p(X = x \mid Y) \quad \text{(distributive property)}$$

$$= p(Y) \quad \text{(axiom of probability)}$$

Generalization for conditionals:

$$p(Y \mid Z) = \textstyle\sum_x p(Y, X = x \mid Z)$$

Tabular Method

Let X, Y be binary RVs with the joint probability table

For Binomial use K-by-K probability table.

X \ Y	y_1	y_2	$P(x)$
x_1	0.04	0.36	0.4 ($P(x_1)$)
x_2	0.30	0.30	0.6 ($P(x_2)$)
$P(y)$	0.34 ($P(y_1)$)	0.66 ($P(y_2)$)	

$P(y_1)=P(x_1,y_1)+P(x_2,y_1)$
$P(y_2)=P(x_1,y_2)+P(x_2,y_2)$
[i.e., sum down columns]

$P(x_1)=P(x_1,y_1)+P(x_1,y_2)$
$P(x_2)=P(x_2,y_1)+P(x_2,y_2)$
[i.e., sum across rows]

Tabular Method

We don't care about event $Y=y_2$

		Y	
		y_1	y_2
X	x_1	0.04	
	x_2	0.30	
		0.34	

Censored!

$P(y_1)$

$P(x|y_1)=?$

Tabular Method

		$Y=y_1$
X	x_1	0.04
	x_2	0.30
		0.34

$P(y_1)$

$\Rightarrow$

$P(x \mid y_1)$
0.04 / 0.34
0.30 / 0.34

These sum to one:
A conditional probability distribution is still a probability distribution

Administrative Items

- HW 1 Due Wednesday
- HW 2 Out Wednesday
- Special Office Hours Today
  - 4-5pm
  - Use Zoom link in D2L for normal (Friday) office hours

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,*

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

Which of the following are true?

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

b) $p(X_1 = 1 | X_2 = 1) = p(X_1 = 1)$ Outcome of die 2 doesn't *affect* die 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) $p(X_1 = 1 | X_3 = 3) < p(X_1 = 1)$

Only 2 ways to get $X_3 = 3$ *, each with equal probability:*

$(X_1 = 1, X_2 = 2)$ *or* $(X_1 = 2, X_2 = 1)$

so

$$p(X_1 = 1 \mid X_3 = 3) = \frac{1}{2} > \frac{1}{6} = p(X_1 = 1)$$

Outline

- Random Events and Probability
- Random Variables
- Fundamental Rules of Probability
- Moments and Dependence of Random Variables
- Useful Discrete Distributions

Independence of RVs

Definition *Two random variables X and Y are independent if and only if,*

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

for all values x and y, and we say $X \perp Y$.

Definition RVs $X_1, X_2, \ldots, X_N$ are mutually independent if and only if,

$$p(X_1 = x_1, \ldots, X_N = x_N) = \prod_{i=1}^{N} p(X_i = x_i)$$

- Independence is *symmetric:* $X \perp Y \Leftrightarrow Y \perp X$
- Equivalent definition of independence: $p(X \mid Y) = p(X)$

Independence of RVs

Intuition…

Consider $P(B|A)$ 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,

$$P(B|A) \neq P(B)$$

Independence of RVs

Definition *Two random variables X and Y are <u>conditionally independent</u> given Z if and only if,*

$$p(X = x, Y = y \mid Z = z) = p(X = x \mid Z = z) p(Y = y \mid Z = z)$$

for all values x, y, and z, and we say that $X \perp Y \mid Z$.

- N RVs conditionally independent, given Z, if and only if:

$$p(X_1, \ldots, X_N \mid Z) = \prod_{i=1}^{N} p(X_i \mid Z)$$

Shorthand notation
Implies for all *x, y, z*

- Equivalent def'n of conditional independence: $p(X \mid Y, Z) = p(X \mid Z)$
- Symmetric: $X \perp Y \mid Z \Leftrightarrow Y \perp X \mid Z$

Moments of RVs

Definition *The <u>expectation</u> of a discrete RV* X*, denoted by* $\mathbf{E}[X]$*, is:*

$$\mathbf{E}[X] = \sum_x x\, p(X = x)$$

Summation over all values in domain of X

Example Let X be the sum of two fair dice, then:

$$\mathbf{E}[X] = \frac{1}{36} \cdot 2 + \frac{1}{18} \cdot 3 + \ldots + \frac{1}{36} \cdot 12 = 7$$

Theorem (Linearity of Expectations) *For any finite collection of discrete RVs* $X_1, X_2, \ldots, X_N$ *with finite expectations,*

Corollary *For any constant c*
$\mathbf{E}[cX] = c\mathbf{E}[X]$

$$\mathbf{E}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbf{E}[X_i]$$

E.g. for two RVs X and Y
$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$

Moments of RVs

Law of Total Expectation *Let X and Y be discrete RVs with finite expectations, then:*

$$\mathbf{E}[X] = \mathbf{E}_Y[\mathbf{E}_X[X \mid Y]]$$

Proof

$$\mathbf{E}_Y[\mathbf{E}_X[X \mid Y]] = \mathbf{E}_Y\left[\sum_x x \cdot p(x \mid Y)\right]$$

$$= \sum_y \left[\sum_x x \cdot p(x \mid y)\right] \cdot p(y)$$ **(Definition of expectation)**

$$= \sum_y \sum_x x \cdot p(x, y)$$ **(Probability chain rule)**

$$= \sum_x x \sum_y \cdot p(x, y)$$ **(Linearity of expectations)**

$$= \sum_x x \cdot p(x) = \mathbf{E}[X]$$ **(Law of total probability)**

Moments of RVs

Theorem: *If* $X \perp Y$ *then* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.

Proof:

$$\mathbf{E}[XY] = \sum_x \sum_y (x \cdot y) p(X = x, Y = y)$$

$$= \sum_x \sum_y (x \cdot y) p(X = x) p(Y = y)$$ **(Independence)**

$$= \left(\sum_x x \cdot p(X = x) \right) \left(\sum_y y \cdot p(Y = y) \right) = \mathbf{E}[X]\mathbf{E}[Y]$$ **(Linearity of Expectation)**

Example *Let* $X_1, X_2 \in \{1, \ldots, 6\}$ *be RVs representing the result of rolling two fair standard die.* *What is the mean of their product?*

$$\mathbf{E}[X_1 X_2] = \mathbf{E}[X_1]\mathbf{E}[X_2] = 3.5^2 = 12 \cdot \frac{1}{4}$$

Moments of RVs

Definition *The conditional expectation of a discrete RV X, given Y is:*

$$\mathbf{E}[X \mid Y = y] = \sum_x x\, p(X = x \mid Y = y)$$

Example Roll two standard six-sided dice and let X be the result of the first die and let Y be the sum of both dice, then:

$$\mathbf{E}[X_1 \mid Y = 5] = \sum_{x=1}^{4} x\, p(X_1 = x \mid Y = 5)$$

$$= \sum_{x=1}^{4} x \frac{p(X_1 = x, Y = 5)}{p(Y = 5)} = \sum_{x=1}^{4} x \frac{1/36}{4/36} = \frac{5}{2}$$

Conditional expectation follows properties of expectation (linearity, etc.)

Moments of RVs

Definition *The* *variance* *of a RV* X *is defined as,*

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$ (X-units)^2

The *standard deviation* *is* $\sigma[X] = \sqrt{\mathbf{Var}[X]}$. (X-units)

Lemma An equivalent form of variance is:

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

Proof Keep in mind that $E[X]$ is a constant,

$$\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2]$$ **(Distributive property)**

$$= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2$$ **(Linearity of expectations)**

$$= \mathbf{E}[X^2] - \mathbf{E}[X]^2$$ **(Algebra)**

Moments of RVs

Definition *The covariance of two RVs* X *and* Y *is defined as,*

$$\mathbf{Cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

Lemma *For any two RVs* X *and* Y*,*

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

e.g. variance is not a linear operator.

Proof $\mathbf{Var}[X+Y] = \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2]$

(Linearity of expectation) $= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2]$

(Distributive property) $= \mathbf{E}[(X-\mathbf{E}[X])^2 + (Y-\mathbf{E}[Y])^2 + 2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$

(Linearity of expectation) $= \mathbf{E}[(X-\mathbf{E}[X])^2] + \mathbf{E}[(Y-\mathbf{E}[Y])^2] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$

(Definition of Var / Cov) $= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$

Moments of RVs

Question: *What is the variance of the sum of <u>independent RVs</u>*

$$
\begin{aligned}
\mathbf{Var}[X_1 + X_2] &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{Cov}(X_1, X_2) \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])(X_2 - \mathbf{E}[X_2])] \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])]\mathbf{E}[(X_2 - \mathbf{E}[X_2])] \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\left(\mathbf{E}[X_1] - \mathbf{E}[X_1]\right)\left(\mathbf{E}[X_2] - \mathbf{E}[X_2]\right) \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2]
\end{aligned}
$$

E.g. variance is a *linear operator* for <u>independent RVs</u>

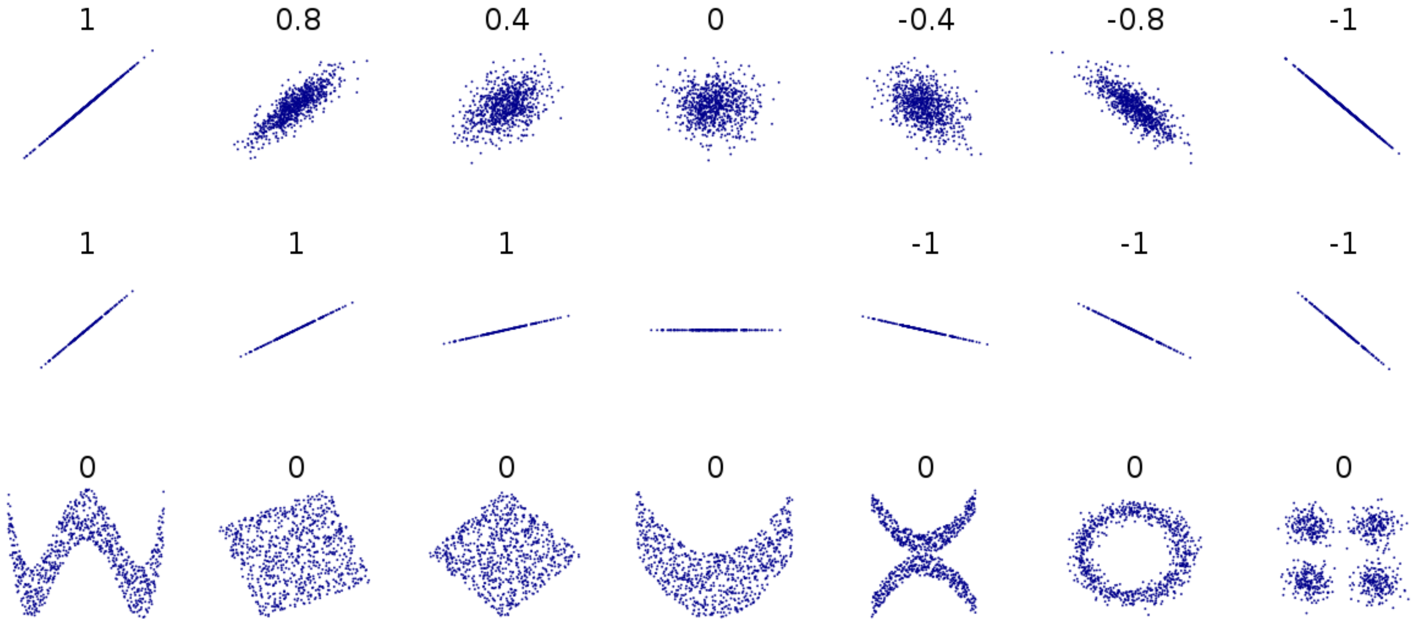
Theorem: *If* $X \perp Y$ *then* $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$

Corollary: *If* $X \perp Y$ *then* $\mathbf{Cov}(X, Y) = 0$

Correlation

Definition *The correlation of two RVs X and Y is given by,*

$$\mathbf{Corr}(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \textit{where} \quad \sigma_X = \sqrt{\mathbf{Var}(X)}$$



Like covariance, only expresses linear relationships!

Outline

- Random Events and Probability
- Random Variables
- Fundamental Rules of Probability
- Moments and Dependence of Random Variables
- Useful Discrete Distributions

Useful Discrete Distributions

Bernoulli *A.k.a. the* ***coinflip*** *distribution on* *binary* *RVs* $X \in \{0,1\}$

$$p(X) = \pi^X (1-\pi)^{(1-X)}$$

Where π *is the probability of* ***success*** *(e.g. heads), and also the mean*

$$\mathbf{E}[X] = \pi \cdot 1 + (1-\pi) \cdot 0 = \pi$$

Suppose we flip N independent coins $X_1, X_2, \ldots, X_N$, what is the distribution over their sum $Y = \sum_{i=1}^N X_i$

Num. "successes" out of N trials

Num. ways to obtain k successes out of N

Binomial Dist. $$p(Y=k) = \binom{N}{k} \pi^k (1-\pi)^{N-k}$$

Binomial Mean: $\mathbf{E}[Y] = N \cdot \pi$ Sum of means for N indep. Bernoulli RVs

Useful Discrete Distributions

Question: How many flips until we observe a success?

Geometric *Distribution on number of independent draws of* $X \sim \text{Bernoulli}(\pi)$ *until success:*

$$p(Y = n) = (1 - \pi)^{n-1}\pi \qquad \mathbf{E}[Y] = \frac{1}{\pi}$$

E.g. for fair coin $\pi = 1/2$ takes two flips on avg.

e.g. there must be n-1 failures (tails) before a success (heads).

Question: How many more flips of we have already seen k failures?

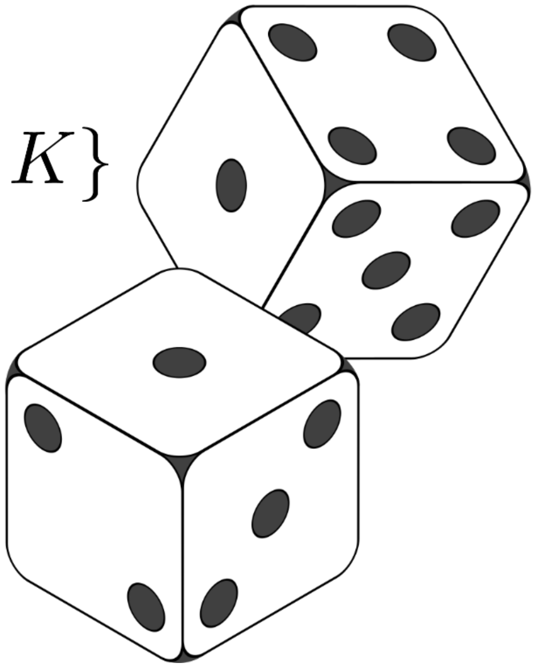
$$p(Y = n + k \mid Y > k) = \frac{p(Y{=}n{+}k, Y{>}k)}{p(Y{>}k)} = \frac{p(Y{=}n{+}k)}{p(Y{>}k)}$$

$$= \frac{(1{-}\pi)^{n+k-1}\pi}{\sum_{i=k}^{\infty}(1{-}\pi)^i\pi} = \frac{(1{-}\pi)^{n+k-1}\pi}{(1{-}\pi)^k} = (1 - \pi)^{n-1}\pi = p(Y = n)$$

For $0 < x < 1, \sum_{i=k}^{\infty} x^i = x^k/(1 - x)$

Corollary: $p(Y > k) = (1 - \pi)^{k-1}$

Useful Discrete Distributions



Categorical *Distribution on integer-valued RV* $X \in \{1, \ldots, K\}$

$$p(X) = \prod_{k=1}^{K} \pi_k^{\mathbf{I}(X=k)} \quad \text{or} \quad p(X) = \sum_{k=1}^{K} \mathbf{I}(X=k) \cdot \pi_k$$

with parameter $p(X=k) = \pi_k$ *and Kronecker delta:*

$$\mathbf{I}(X=k) = \begin{cases} 1, & \text{If } X = k \\ 0, & \text{Otherwise} \end{cases}$$

Can also represent X as *one-hot* binary vector,

$$X \in \{0,1\}^K \quad \text{where} \quad \sum_{k=1}^{K} X_k = 1 \quad \text{then} \quad p(X) = \prod_{k=1}^{K} \pi_k^{X_k}$$

This representation is special case of the **multinomial distribution**

Useful Discrete Distributions

What if we count outcomes of *N* independent categorical RVs?

Multinomial *Distribution on K-vector* $X \in \{0, N\}^K$ *of counts of N repeated trials* $\sum_{k=1}^K X_k = N$ *with PMF:*

$$p(x_1, \ldots, x_K) = \binom{n}{x_1 x_2 \ldots x_K} \prod_{k=1}^{K} \pi_k^{x_k}$$

Number of ways to partition N objects into K groups:

$$\binom{n}{x_1 x_2 \ldots x_K} = \frac{n!}{x_1! x_2! \ldots x_K!}$$

Leading term ensures PMF is properly normalized:

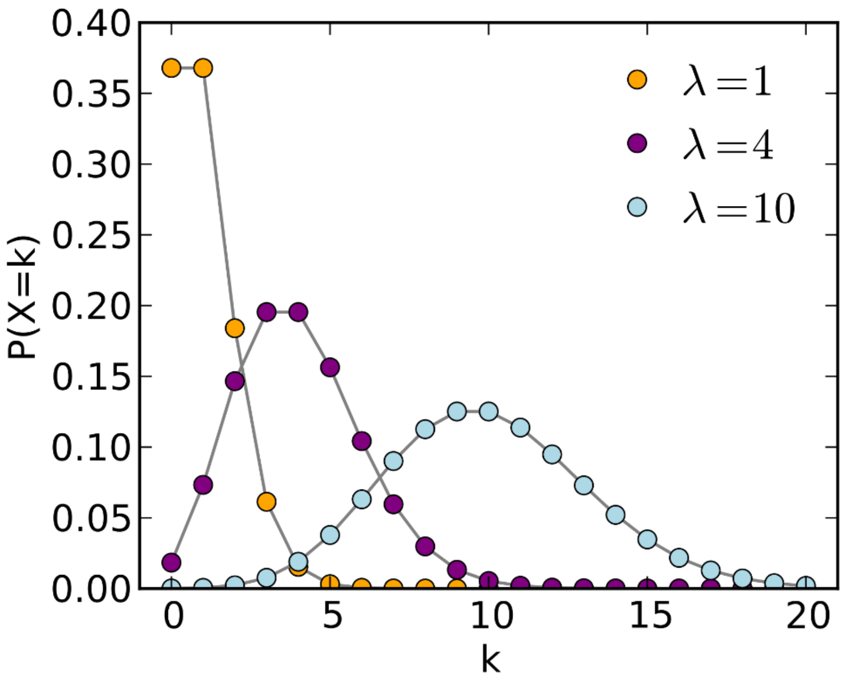
$$\sum_{x_1} \sum_{x_2} \ldots \sum_{x_K} p(x_1, x_2, \ldots, x_K) = 1$$

Useful Discrete Distributions

*A **Poisson** RV X with <u>rate</u> parameter λ has the following distribution:*

Mean and variance both scale with parameter

$$p(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad \mathbf{E}[X] = \mathbf{Var}[X] = \lambda$$



Represents number of times an *event* occurs in an interval of time or space.

Ex. *Probability of overflow floods in 100 years,*

$$p(k \text{overflow floods in 100 yrs}) = \frac{e^{-1}1^k}{k!}$$

Avg. 1 overflow flood every 100 years, makes setting rate parameter easy.

Lemma (additive closure) *The sum of a finite number of Poisson RVs is a Poisson RV.*

$$X \sim \text{Poisson}(\lambda_1), \quad Y \sim \text{Poisson}(\lambda_2), \quad X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

Recap

- A **random variable** is a function of samples to real values: $X : \Omega \to \mathbb{R}$
- $X = x$ Is an event with probability: $p(X = x) = \sum_{\omega \in \Omega : X(\omega) = x} P(\omega)$
- p(X) is a **probability mass function** (PMF) satisfying

$$p(X = x) \geq 0 \qquad \sum_x p(X = x) = 1$$

- Some fundamental rules of probability:
  - Conditional: $p(X \mid Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(X,Y)}{\sum_x p(X=x,Y)}$
  - Law of total probability: $p(Y) = \sum_x p(Y, X = x)$
  - Probability chain rule: $p(X, Y) = p(Y)p(X \mid Y)$

Recap

- Independence of RVs:
  - Two RVs X & Y are independent iff: $p(X \mid Y) = p(X)$
  - Equivalently: $p(X, Y) = p(X)p(Y)$
  - X & Y are conditionally independent given Z iff: $p(X \mid Y, Z) = p(X \mid Z)$
  - Equivalently: $p(X, Y \mid Z) = p(X \mid Z)p(Y \mid Z)$

- Moments and Expected Value
  - Expected value of a discrete RV: $\mathbf{E}[X] = \sum_x x\, p(X = x)$
  - Expectation is a linear operator $\mathbf{E}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbf{E}[X_i]$
  - Variance of a RV: $\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$
  - Variance is **not** a linear operator (unless RVs are independent)

Useful Discrete Distributions

Theorem *Let* $X \sim \text{Binomial}(n, \pi(n))$ *where* $\pi(n)$ *is a function of n and* $\lim_{n\to\infty} n \cdot \pi(n) = \lambda$ *for some constant* λ. *Then for any fixed k:*

$$\lim_{n\to\infty} \text{Binomial}(X \mid n, \pi(n)) = \text{Poisson}(X \mid \lambda)$$

Proof Sketch Use Taylor expansion of e^x and $(1-\pi)^k \geq (1-\pi k)$ to upper and lower bound Binomial probability as a function of n:

$$\underbrace{\frac{e^{\pi n}((n-k+1)\pi)^k}{k!}(1-\pi^2 n)}_{\textit{LB(n)}} \leq \text{Binomial}(X=k \mid n, \pi) \leq \underbrace{\frac{e^{-\pi n}(n\pi)^k}{k!}\frac{1}{1-\pi k}}_{\textit{UB(n)}}$$

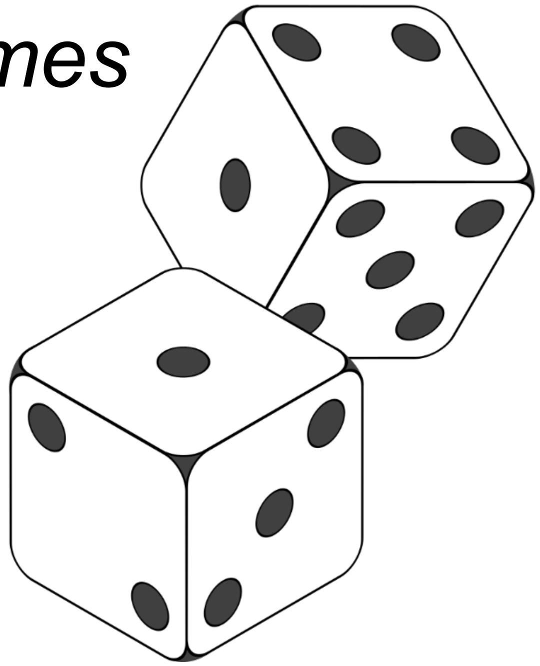
As $n \to \infty$ it must be that $\pi(n) \to 0$ so that $\lim_{n\to\infty} n \cdot \pi(n) = \lambda$ is constant. Then $1/(1-\pi k) \to 1$ and $1 - \pi^2 n \to 1$. The difference $[(n-k+1)\pi] - n\pi$ approaches 0. Therefore:

$$\lim_{n\to\infty} \text{LB}(n) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{and} \quad \lim_{n\to\infty} \text{UB}(n) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Bounds converge so result holds.

Random Events and Probability

1. A **sample space** Ω: *set of all possible outcomes* of the experiment.



Dice Example: All combinations of dice rolls,

$$\Omega = \{(1,1),(1,2),\dots,(6,5),(6,6)\}$$

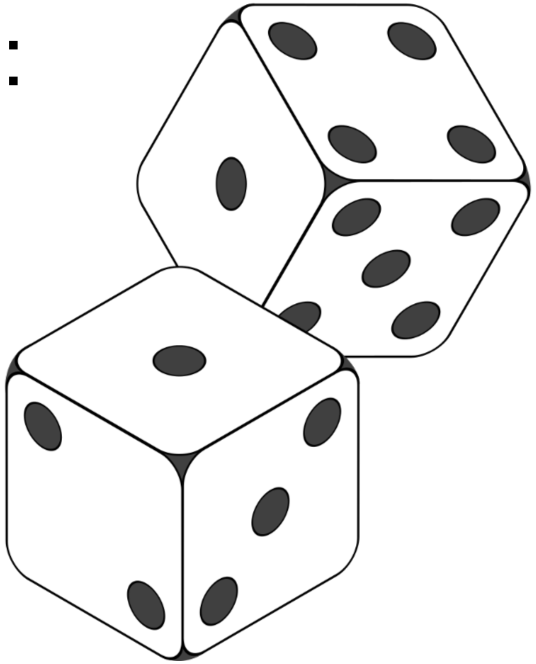
2. An **event space** $\mathcal{F}$: Family of sets representing allowable events, where each set in $\mathcal{F}$ is a subset of the sample space Ω.

Dice Example: Event that we roll even numbers,

$$E = \{(2,2),(2,4),\dots,(6,4),(6,6)\} \in \mathcal{F}$$

Random Events and Probability

3 A **probability function** $P : \mathcal{F} \rightarrow \mathbf{R}$ satisfying:



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 pairwise mutually disjoint events $E_1, E_2, E_3, \ldots$

Axioms of Probability

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

(Fair) Dice Example: Probability that we roll even numbers,

$$P((2,2) \cup (2,4) \cup \ldots \cup (6,6)) = P((2,2)) + P((2,4)) + \ldots + P((6,6))$$

$$= \frac{1}{36} + \frac{1}{36} + \ldots + \frac{1}{36} = \frac{9}{36}$$

9 Possible outcomes, each with equal probability of occurring

Random Events and Probability

Some rules regarding set of event space $\mathcal{F}$…

- $\mathcal{F}$ must include $\emptyset$ and Ω
- $\mathcal{F}$ is **closed** under set operations, if $E_1, E_2 \in \mathcal{F}$ then:
  - $E_1 \cup E_2 \in \mathcal{F}$
  - $E_1 \cap E_2 \in \mathcal{F}$
  - $\overline{E_1} = \Omega - E_1 \in \mathcal{F}$