Computer Science

CSC535: Probabilistic Graphical Models

Probability Primer

Prof. Jason Pacheco

TA: Enfa Rose George

TA: Saiful Islam Salim

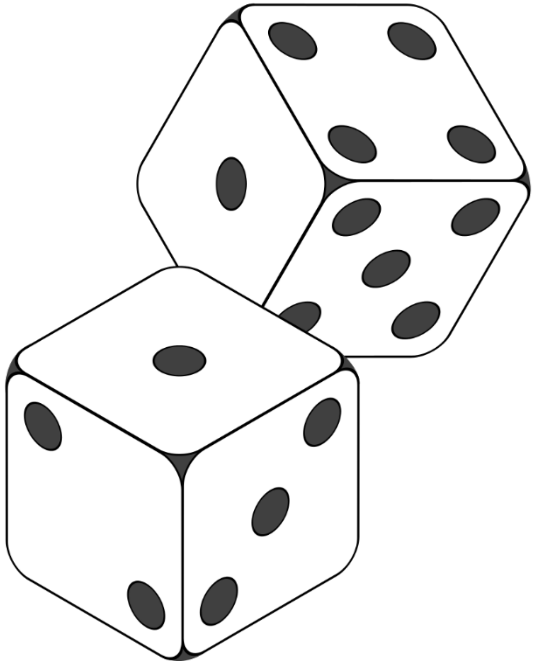
Administrative Items

- Homework 1
  - Will be out next week, Tue 8/31
  - Due Tue 9/7
  - Error in current syllabus
  - We will announce office hours in lecture and on Piazza
- Today's Reading: Wasserman, CH1
- Questions about Book access

Random Events and Probability

Suppose we roll <u>two fair dice</u>…

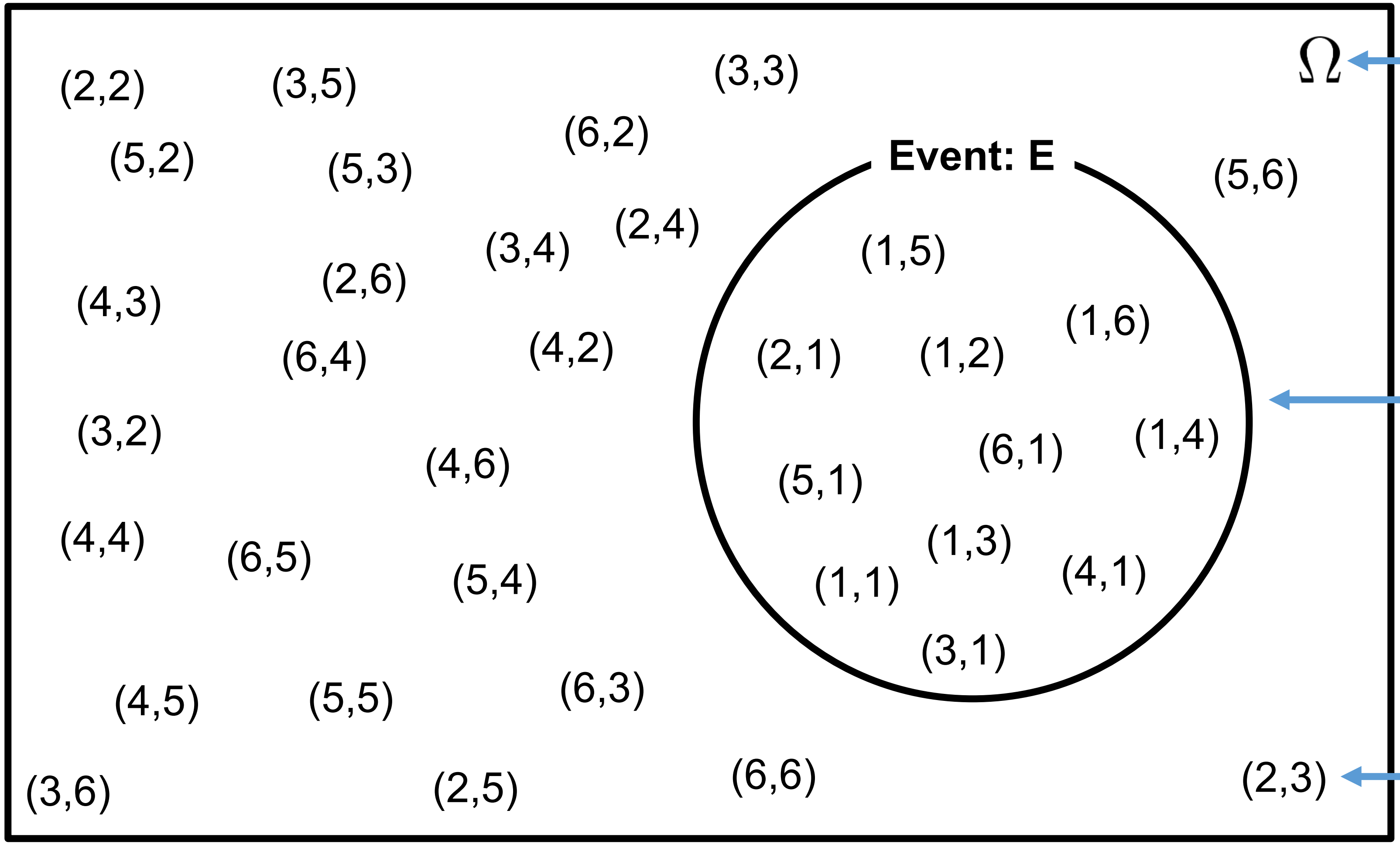
- What are the possible outcomes?
- What is the *probability* of rolling **even** numbers?
- What is the *probability* of rolling **odd** numbers?
- If one die rolls 1, then what is the probability of the second die also rolling 1?
- How to mathematically formulate outcomes and their probabilities?



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

Random Events and Probability

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

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

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

Random Events and Probability

What does probability of an event P(E) mean?

Two commonly accepted definitions (loosely):

Frequentist The proportion (frequency) of times that E is true in repeated trials.

<u>Example</u> A coin with probability ½ of coming up heads (fair or unbiased coin) means, if we flip an infinite number of times ½ should be heads.

What about events that are not repeatable? Like the probability that someone wins an election?

Random Events and Probability

What does probability of an event P(E) mean?

Two commonly accepted definitions (loosely):

Bayesian A measure of the *degree-of-belief* that an event will be true. Introduces subjectivity, but makes defining probability of non-repeatable events simple.

> Example Coin bias is not a fixed property of the coin. It is random; it has a distribution that encapsulates *all of our uncertainty* about the flip including coin properties, wind effects, etc.

*The difference between **Bayesian** and **frequentist** interpretations won't matter until we address statistical inference*

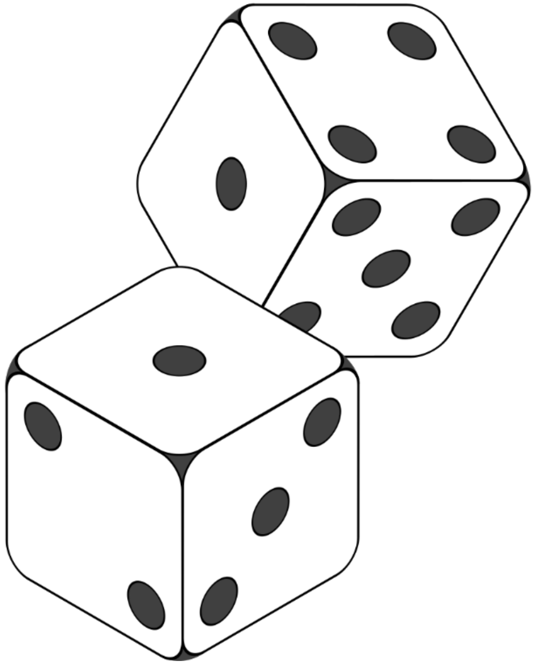
Random Events and Probability

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

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

Number of elements in event set

Number of possible outcomes (36)



This is the uniform probability distribution

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

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

9 Possible outcomes, each with equal probability of occurring

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

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

Number of ways to choose 30 different days from 365 is,

$$\binom{365}{30} \quad \text{where} \quad \binom{N}{k} = \frac{N!}{k! \cdot (N-k)!}$$

Binomial coefficient

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

$$|E^c| = \binom{365}{30} 30!$$

30! Orderings of each set of birthdays

Birthday Paradox

Total number of possible combinations of birthdays among 30 people,

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

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

$$P(E^c) = \frac{|E^c|}{|\Omega|} = \frac{\binom{365}{30}}{365^{30}} = 0.294$$

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

$$P(E) = 1 - P(E^c) = 0.706$$

With the 58 people registered
P(E) = 0.992

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

Random Events and Probability

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

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

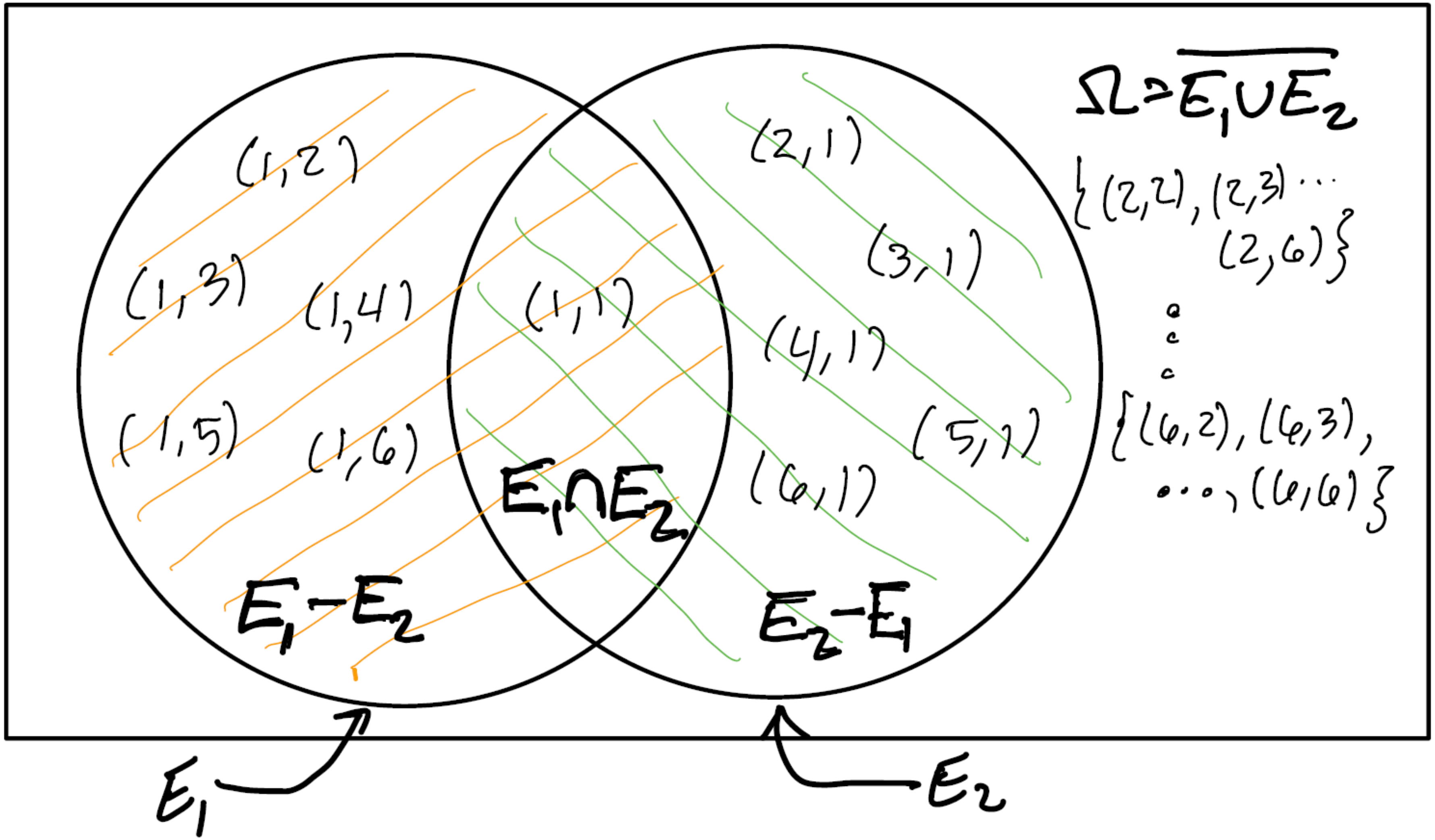
$$E_1 = \{(1,1),(1,2),\ldots,(1,6)\} \qquad E_2 = \{(1,1),(2,1),\ldots,(6,1)\}$$

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

Random Events and Probability

Can interpret these operations as a Venn diagram…

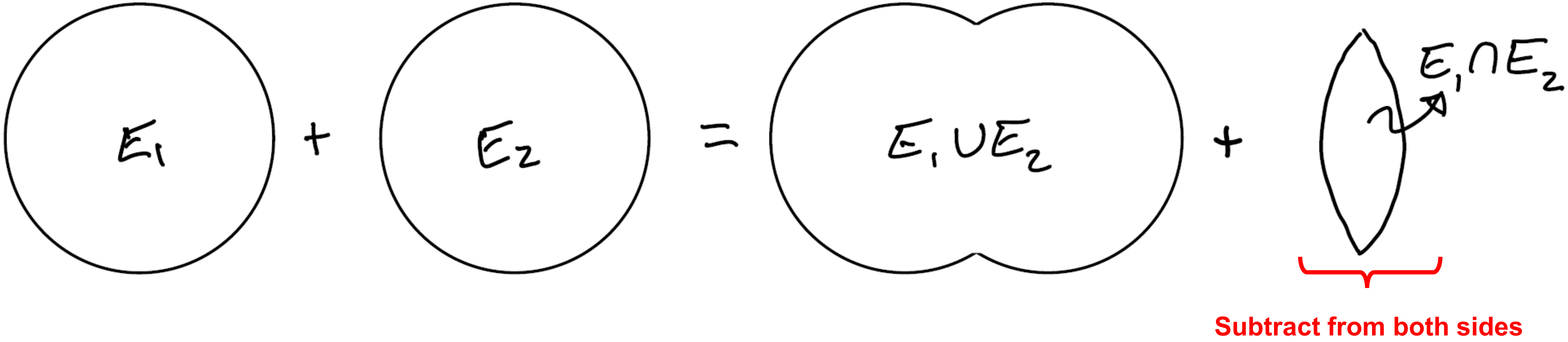


Random Events and Probability

Lemma: *For 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:



Random Events and Probability

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

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

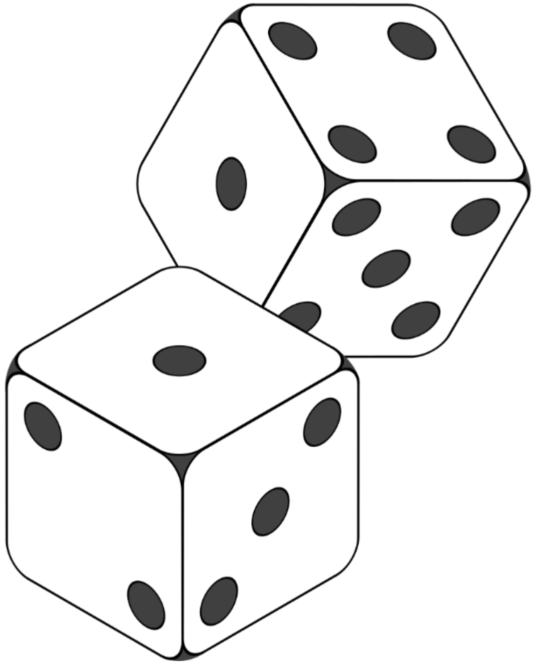
Proof:

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

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

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

Random Variables and Probability



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

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

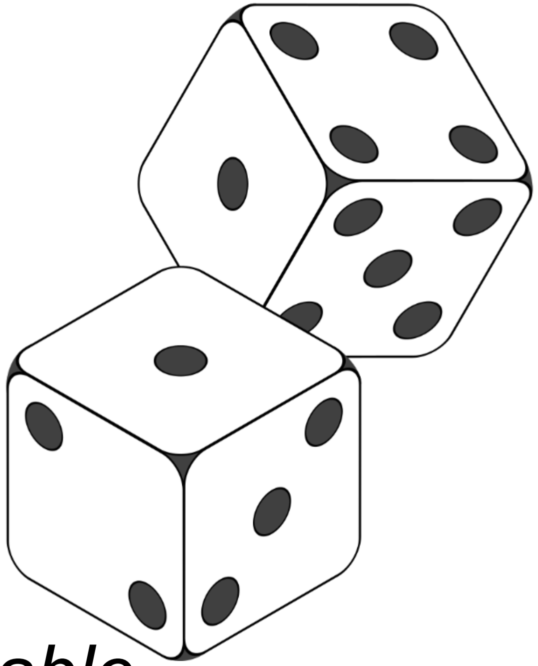
Two dice example:

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

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

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

Random Variables and Probability



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

Option 2 Use a function of sample space…

Definition *A* random variable $X(\omega)$ *for* $\omega \in \Omega$ *is a* *real-valued function* $X : \Omega \rightarrow \mathbb{R}$. *A discrete random variable takes on only a finite or countably infinite number of values.*

For *discrete* $X = x$ is an **event** with **probability mass function (PMF)**:

We will use the shorthand X instead of $X(\omega)$

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

Random Variables and Probability

Given random variables X, Y the *joint probability distribution,*

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

is the probability that **both** events $X = x$ and $Y = y$ occur

Example Suppose we flip 2 fair coins represented by random variables X and Y. What is the probability that they are both heads?

There are 4 possible outcomes, each with equal prob.

$$(X = T, Y = T), (X = H, Y = T), (X = T, Y = H), (X = H, Y = H)$$

Joint probability is,

$$P(X = H, Y = H) = \frac{|X=H, Y=H|}{4} = \frac{1}{4}$$

Random Variables and Probability

Some notes on notation for random variables (RVs)…

- We denote the RV by capital X and its realization by lowercase x
- Generally use shorthand X instead of $X(\omega)$
- Other common shorthand: $p(x) = p(X = x)$
- Any function $f(X)$ of an RV is also an RV, e.g. $Y = f(X)$

Fundamental Rules of Probability

The ***law of total probability*** is,

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

p(y) is known as the ***marginal probability.***

Example Roll two fair dice and let X be the outcome of the first die. Let Y be the sum of both dice. What is the probability that both dice sum to Y=6?

$$p(Y = 6) = \sum_{x=1}^{6} p(Y = 6, X = x)$$

$$= p(Y = 6, X = 1) + p(Y = 6, X = 2) + \ldots + p(Y = 6, X = 6)$$

$$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + 0 = \frac{5}{36}$$

Fundamental Rules of Probability

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

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

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]

Fundamental Rules of Probability

The **probability chain rule** is,

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

Marginal → $p(Y)$; Conditional → $p(X \mid Y)$

Proof By definition of the conditional distribution,

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

Multiply both sides by p(Y),

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

Fundamental Rules of Probability

Suppose we have a collection of N random variables,

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

The probability chain rule for these random variables is,

$$p(X_1, X_2, \ldots, X_N) = p(X_1)p(X_2 \mid X_1) \ldots p(X_N \mid X_{N-1}, \ldots, X_1)$$
$$= p(X_1) \prod_{i=2}^{N} p(X_i \mid X_{i-1}, \ldots, X_1)$$

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

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

Fundamental Rules of Probability

We have enough tools to prove the law of total probability

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

Proof

$$\sum_x p(Y, X = x) = \sum_x p(Y) p(X = x \mid Y)$$ (chain rule)

$$= p(Y) \sum_x p(X = x \mid Y)$$ (distributive property)

$$= p(Y)$$ (axiom of probability)

Also works for conditional probabilities,

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

Tabular Calculations

Tabular representation of two binary RVs

Use K-by-K probability table for K-valued discrete RVs

Y

X		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 Calculations

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

		Y	
		y_1	y_2
X	x_1	0.04	Censored!
	x_2	0.30	
		0.34	

0.34 ← $P(y_1)$

$P(x|y_1)=?$

Tabular Calculations

		$Y=y_1$
X	x_1	0.04
	x_2	0.30

0.34

$P(y_1)$

$\Rightarrow$

0.04 / 0.34

0.30 / 0.34

$P(x|y_1)$

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