Probabilistic Graphical Models

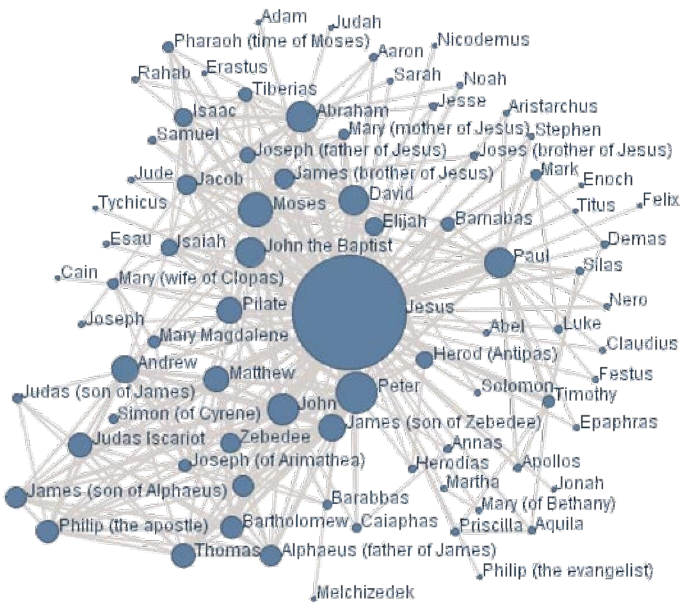
Introduction

Logistics

- Textbooks:
  - Christopher Bishop, Pattern Recognition and Machine Learning
  - Kevin Murphy, Machine Learning: A probabilistic Perspective
  - Daphne Koller and Nir Friedman, Probabilistic Graphical Models
- Class website: https://polyhedron.math.uoc.gr/2223/moodle/course/view.php?id=18
- Grading:
- 2 homework assignments: 20% of grade
  - Theory exercises, Implementation exercises
- Project: 40% of grade
  - Pick a paper (not introductory) from *M.I. Jordan (editor), Learning in Graphical Models*
  - Present it in class.
  - Implement the method
  - Apply it on a real or simulated data set.
  - March 6th :Deadline for project proposal

What Are Graphical Models?

Graph



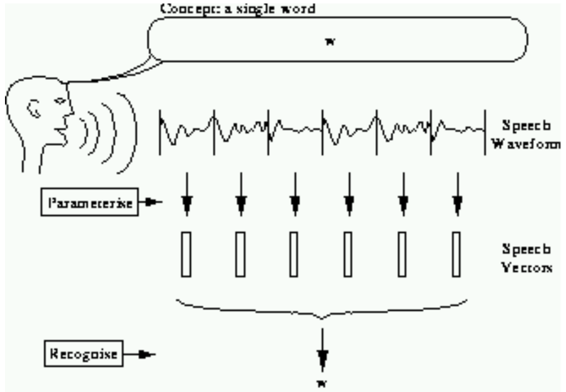
Model

M_G

Data

$$\left\{x_1^{(i)}, \dots, x_n^{i}\right\}_{i=1}^{N}$$

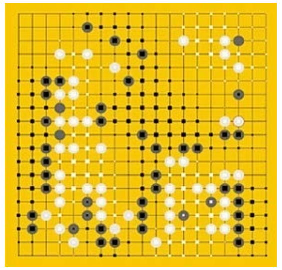
Reasoning under uncertainty!



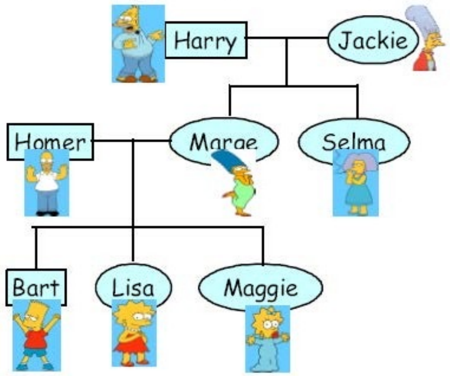
Speech recognition

Information retrieval

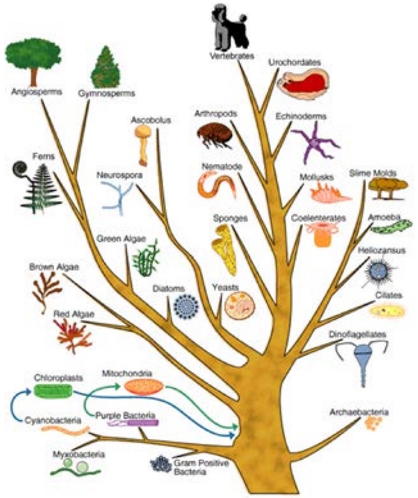
Computer vision



Games



Pedigree



Evolution

Robotic control

Planning

The Fundamental Questions

- Representation
  - How to capture/model uncertainties in possible worlds?
  - How to encode our domain knowledge/assumptions/constraints?
- Inference
  - How do I answer questions/queries according to my model and/or based given data?

    e.g.: $P(X_i \mid \boldsymbol{D})$
- Learning
  - What model is "right" for my data?

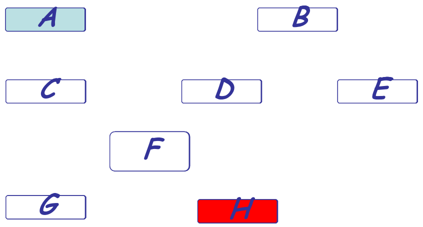
    e.g.: $\mathrm{M} = \arg\max_{\mathrm{M} \in \mathcal{M}} \mathrm{F}(\mathrm{D};\mathrm{M})$

Representing multivariate distributions

- Representation: what is the joint probability dist. on multiple variables?

$$P(A,\ B,\ C,\ D,\ E\ ,F,\ G,\ H)$$

  - How many state configurations in total ? --- 2^8
  - Are they all needed to be represented?
  - Do we get any scientific/medical insight?



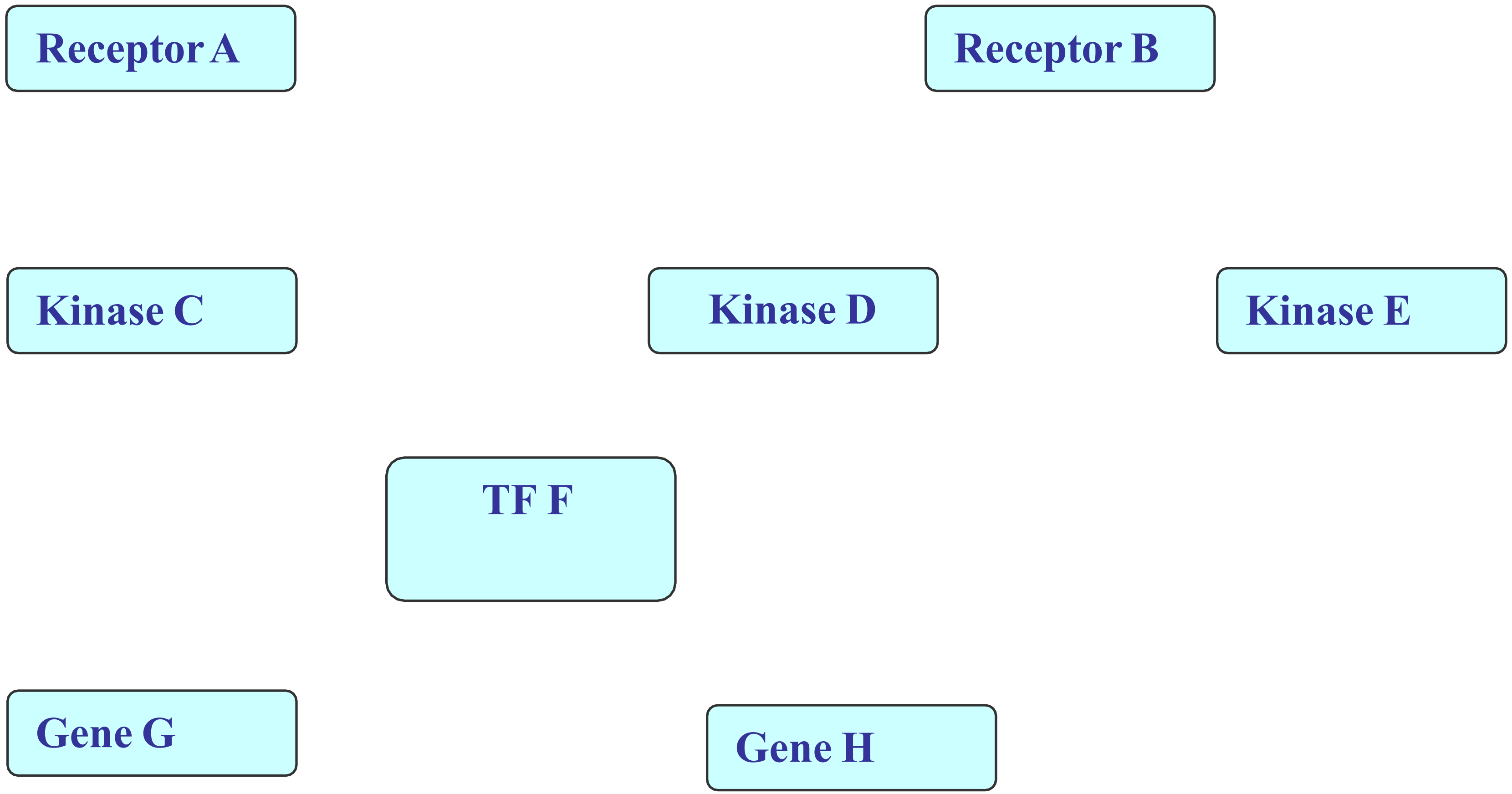
- Learning: where do we get all this probabilities?
  - Maximal-likelihood estimation? but how many data do we need?
  - Are there other est. principles?
  - Where do we put domain knowledge in terms of plausible relationships between variables, and plausible values of the probabilities?

- Inference: If not all variables are observable/observed, how to compute the conditional distribution of latent variables given evidence?
  - Computing $P(H|A)$ would require summing over all 2^6 configurations of the unobserved variables

What is a Graphical Model?

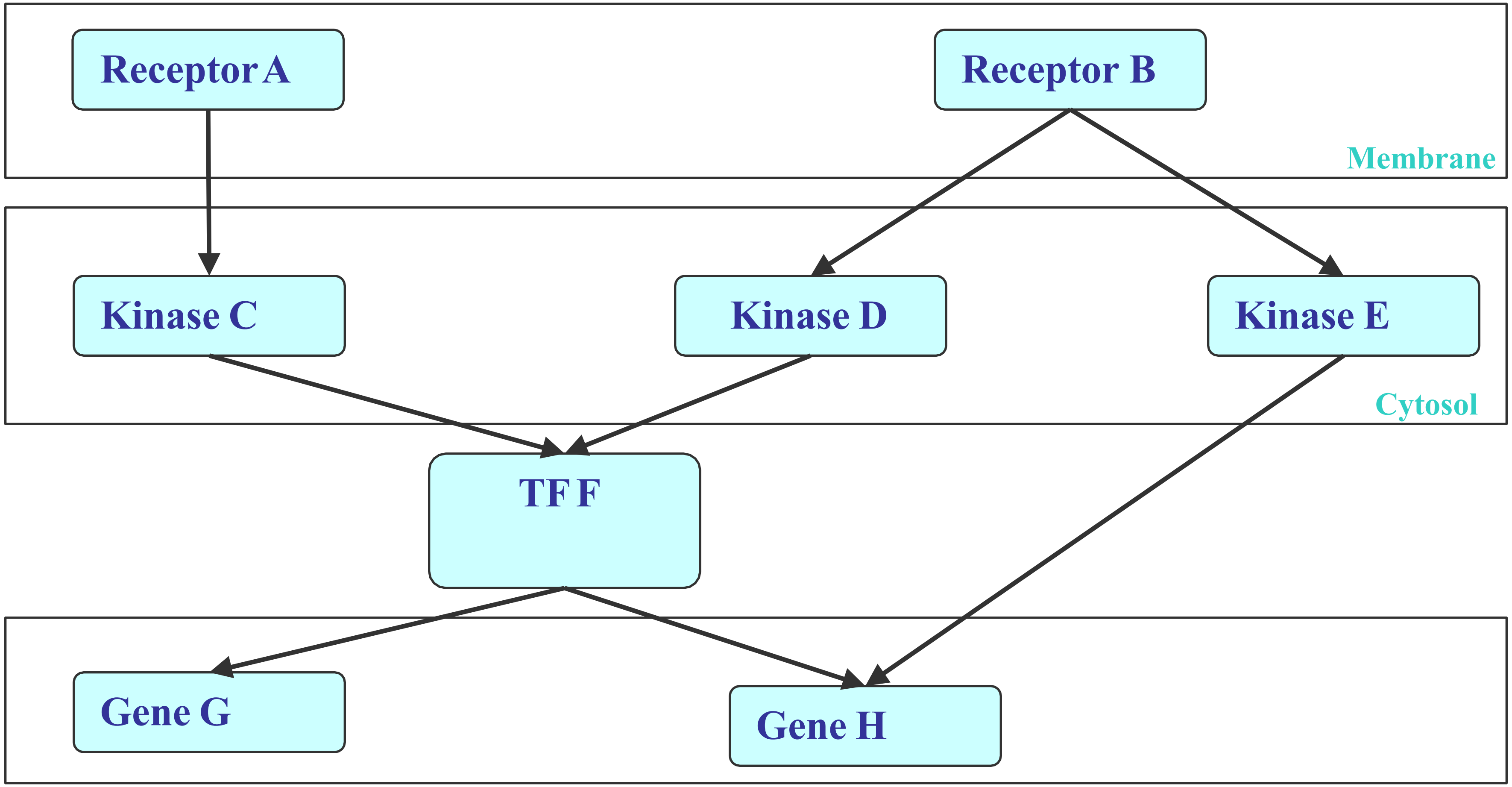
--- Multivariate Distribution in High-D Space

- A possible world for cellular signal transduction:



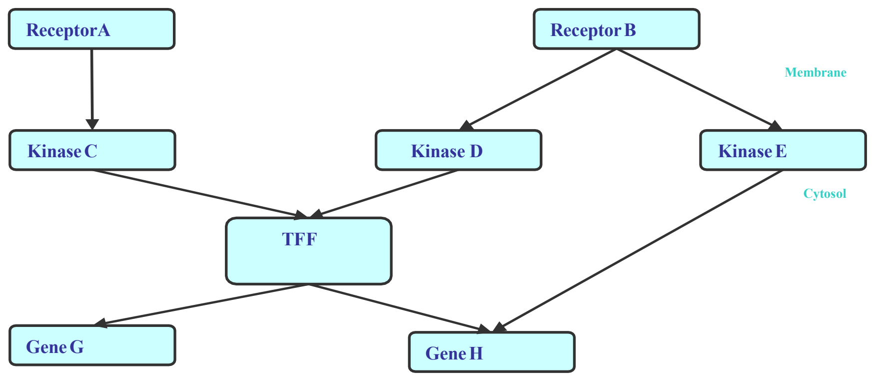
GM: Structure Simplifies Representation

- Dependencies among variables



Probabilistic Graphical Models

- If X_i's are conditionally independent (as described by a PGM), the joint can be factored to a product of simpler terms, e.g.,



$$P(A, B, C, D, E, F, G, H)$$
$$= P(A)\, P(B)\, P(C|\, A)\, P(D|\, B)\, P(E|\, B)\, P(F|C, D)\, P(G|F)\, P(H|\, E, F)$$

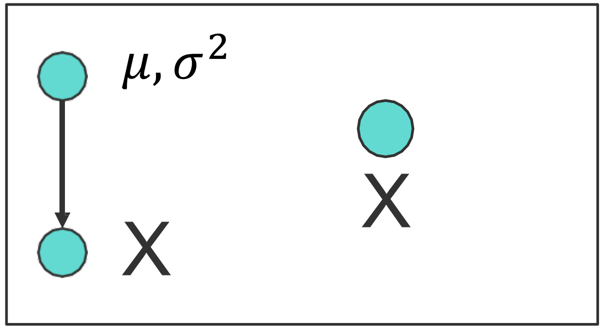
- Why we may favor a PGM?
  - Incorporation of domain knowledge and causal (logical) structures

$1+1+2+2+2+4+2+4=18$, a 16-fold reduction from 2^8 in representation cost !

GMs are your old friends

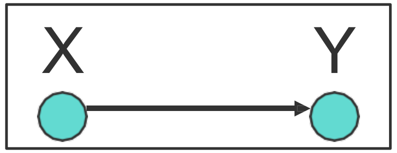
Density estimation

Parametric and nonparametric methods



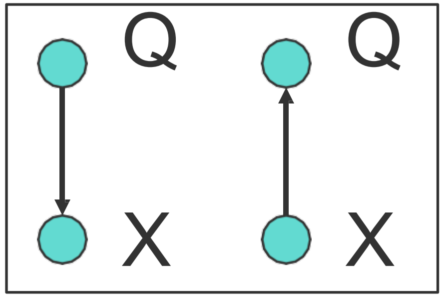
Regression

Linear, conditional mixture, nonparametric



Classification

Generative and discriminative approach



Clustering

Probabilistic Graphical Models

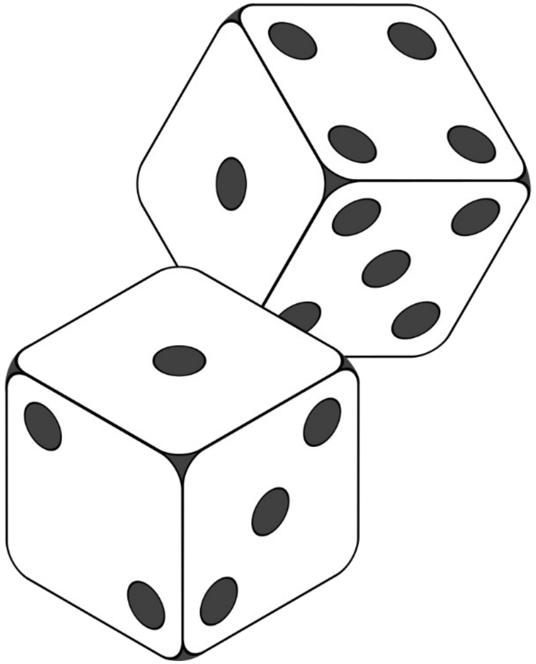
Probability recap

- Reading: Murphy, Secs. 2.1 and 2.2.
- Lots of slides from: Eli Upfal, Jason Panjeho.

Random Events and Probability

Suppose we roll two fair dice…

- 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?

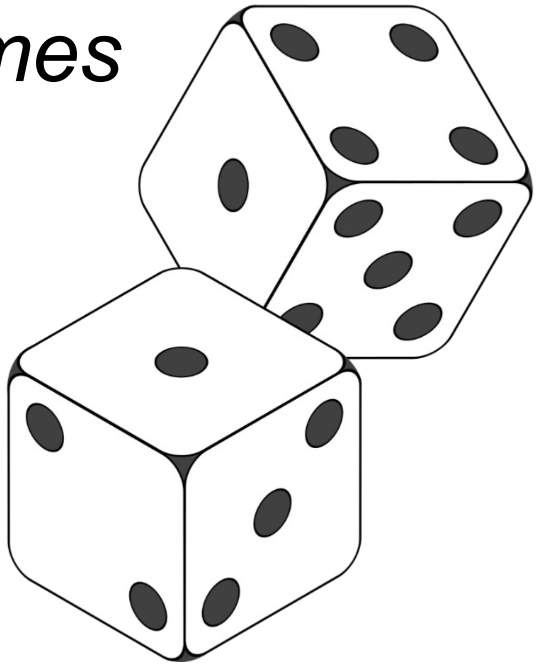


*…this is an **experiment** or **random process**.*

Formulate as probability space having 3 components

Random Events and Probability

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



Dice Example: All combinations of dice rolls,

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

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,

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

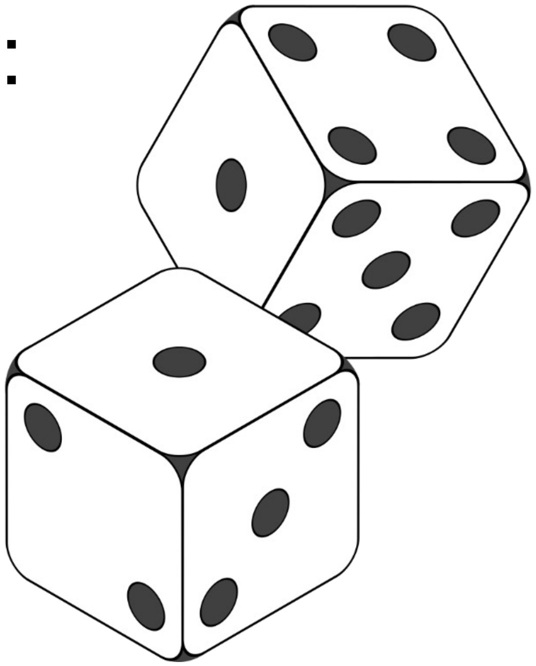
Random Events and Probability

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

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

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

Axioms of Probability



(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 countable unions, countable intersections and complement: if $E_1, E_2 \in \mathcal{F}$ then:
  - $E_1 \cup E_2 \in \mathcal{F}$
  - $E_1 \cap E_2 \in \mathcal{F}$
  - $\overline{E_1} = \Omega - E_1 \in \mathcal{F}$

Random Events and Probability

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

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

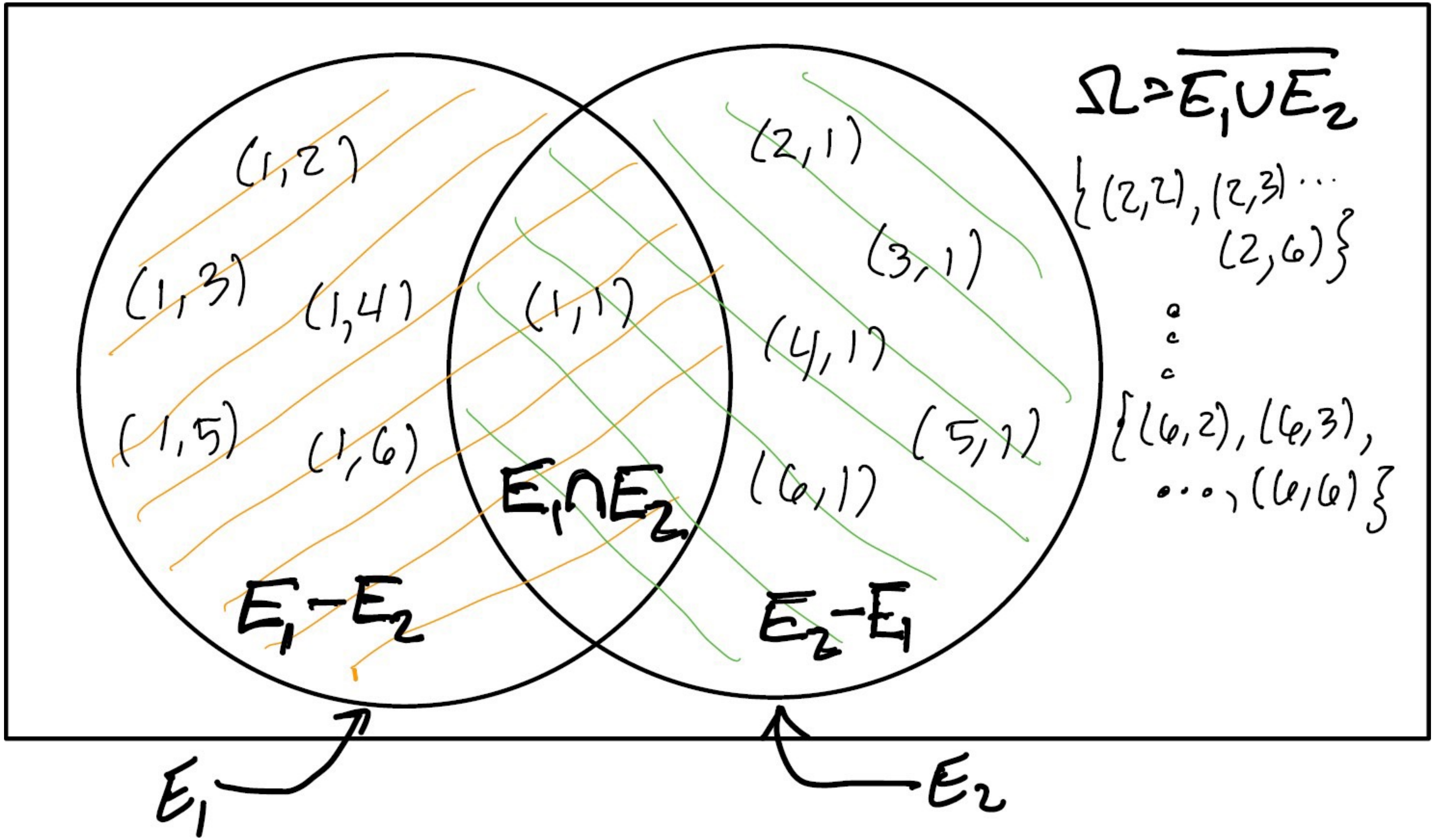
$$E_1 = \{(1,1),(1,2),\ldots,(1,6)\} \qquad E_2 = \{(1,1),(2,1),\ldots,(6,1)\}$$

Then we must include 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

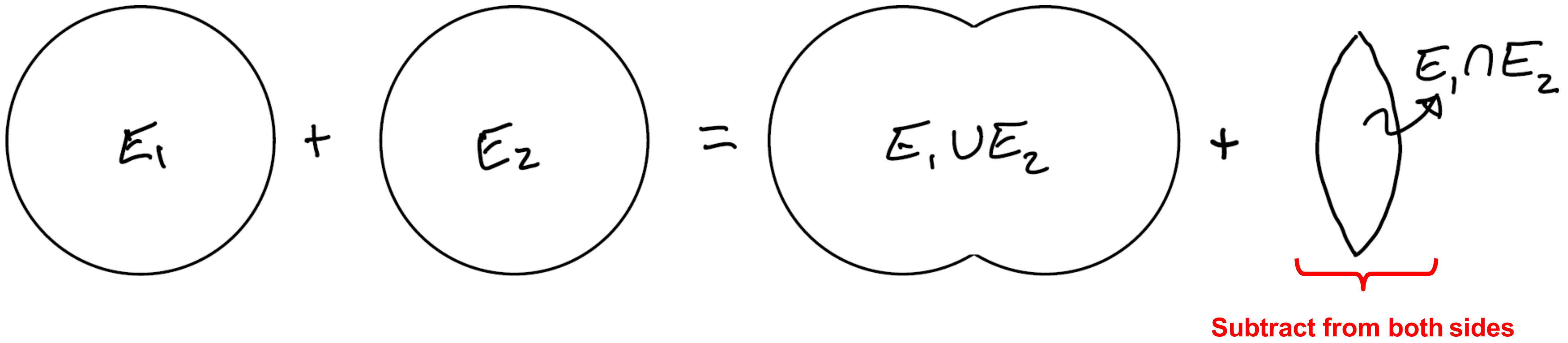
Can interpret these operations as a Venn diagram…



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

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

Graphical Proof:



Random Events and Probability

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

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

Proof:

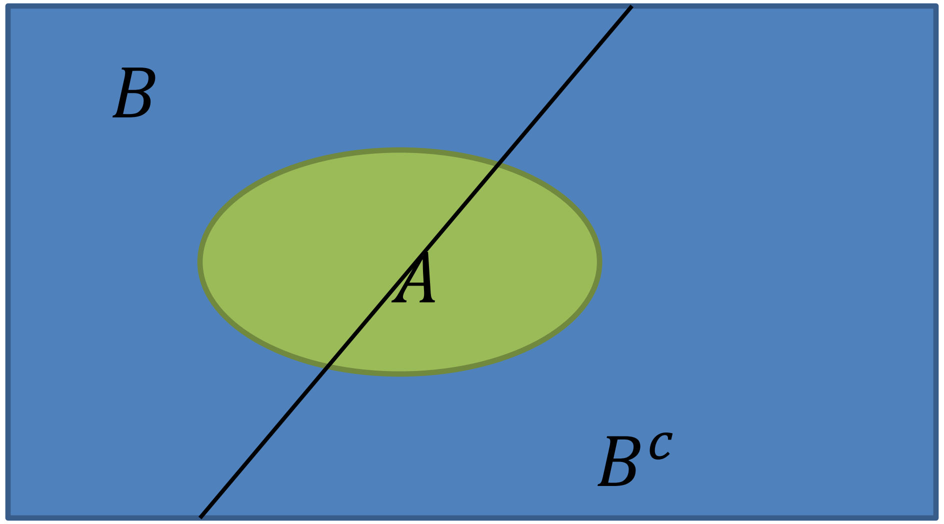
$$P(E_1) = P(E_1 \setminus (E_1 \cap E_2)) + P(E_1 \cap E_2)$$

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

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

Law of Total Probability

$$P(A) = P(A \cap B) + P(A \cap B^c)$$



Independence

Definition

Two events A and B are independent if

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

We denote this as

$$A \perp B$$

Example

You roll a die. Consider the following two events

- A : "The outcome is an even number"
- B : "The outcome is one of the numbers {1, 2, 3, 4}"
- C : "The outcome is one of the numbers {1, 2, 3}"

Find $P(A), P(B), and\ P(A \cap B)$

Are A and B independent?

Find $P(A), P(C), and\ P(A \cap C)$

Are A and C independent?

Conditional Probability

One way to interpret the independence of events is as follows:

- Consider again the following two events:
  - A : "The outcome is an even number”
  - B : "The outcome is one of the numbers {1, 2, 3, 4}"

- You want to bet on event A. How much are you willing to bet?
- I roll the die and tell you that event B has happened (hence, the outcome is one of {1, 2, 3, 4}.
- How much are you willing to bet now?
- We just described the conditional probability

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

Conditional Probability

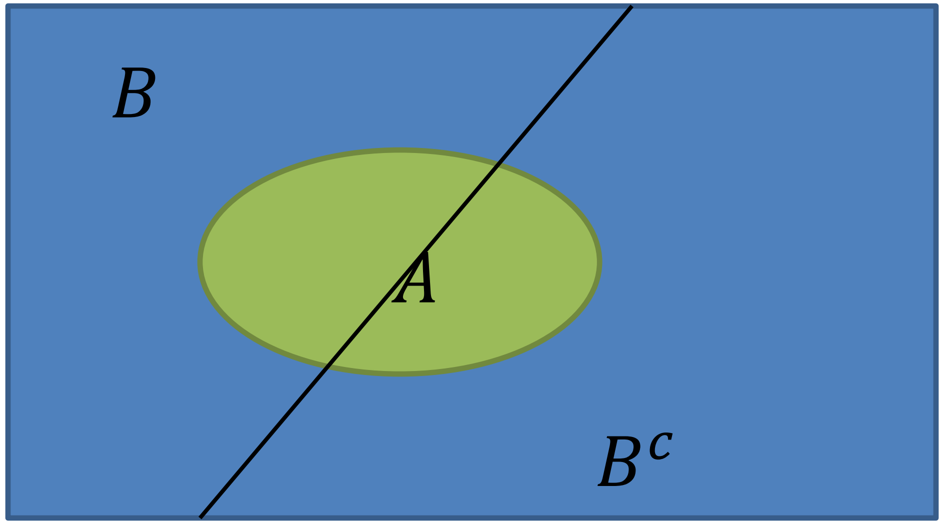
Conditional Probability:

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

Law of Total Probability

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$



Bayes Rule

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

Very often, people confuse P(A|B) and P(B|A). These can be VERY different.

Think about it:

You read in the paper: "Half of the people hospitalized with covid-19 are fully vaccinated". Do you think that getting the vaccine lowers your chances of getting hospitalized?

Example

A box contains two coins: a regular coin and one fake two-headed coin ($P(H) = 1$). I choose a coin at random and toss it twice. Define the following events.

- A = First coin toss results in an H.
- B = Second coin toss results in an H.
- C= Coin 1 (regular) has been selected.

Find $P(A|C), P(B|C), P(A \cap B|C), P(A), P(B), and\ P(A \cap B)$

Are A and B independent?

Are A and B independent given C?

Bayes Rule

- *Vacc: True if vaccinated, false otherwise*
- *Hosp: True if hospitalized, false otherwise.*
- $P(Hosp|Vacc) = 0.01$
- $P(Hosp|Vacc^c) = 0.2$
- *Consider these different possibilities:* $P(Vacc) =$ *0.5, 0.99*

Let' s use Bayes rule to compute $P(Vacc|Hosp)$ *for both cases.*

Conditional Independence

Definition

Two events A and B are independent given an event C if

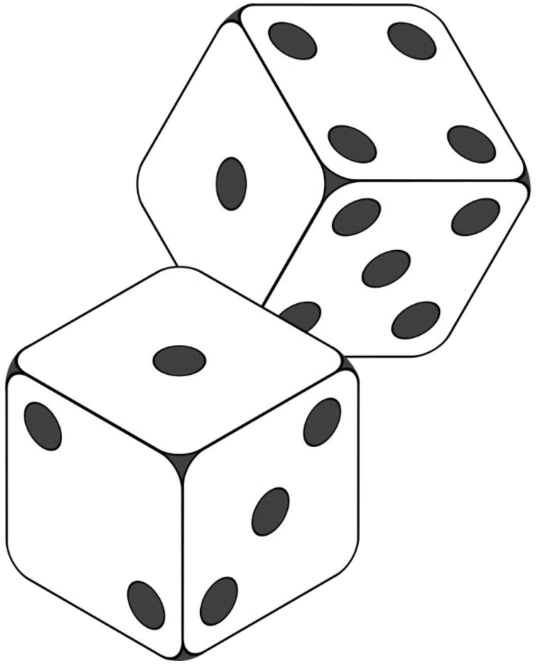
$$P(A \cap B|C) = P(A|C)P(B|C).$$

We denote this as

$$A \perp B|\mathrm{C}$$

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*

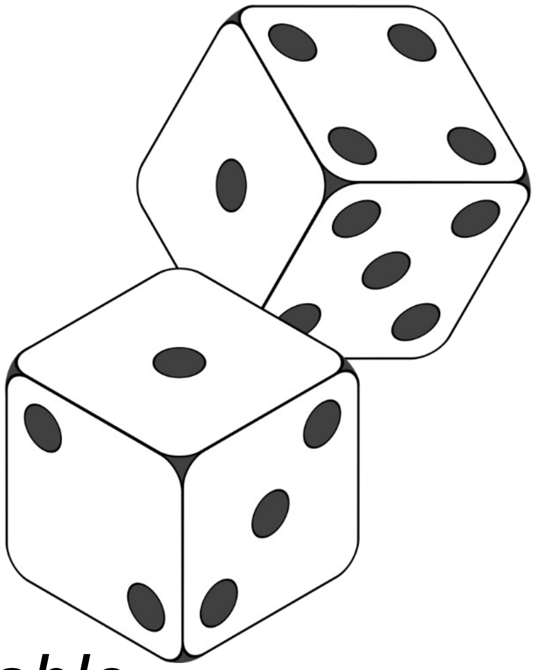
Two dice example:

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

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

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 <u>sum of dice</u>…

<u>Option 2</u> Use a function of sample space…

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

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

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

Random Variables

Some notes on 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. $Y(\omega) = f(X(\omega))$
- More shorthand: the joint distribution of RVs $p(X, Y) = p(X \cap Y)$
- We will use *"distribution"* loosely to refer to distributions, PMFs, probability density and cumulative distribution functions (defined later)