ECE 6504: Advanced Topics in Machine Learning

Probabilistic Graphical Models and Large-Scale Learning

Topics:

- Bayes Nets: Representation/Semantics

Readings: Barber 3.3.1; KF 3.1-3.2.2

Dhruv Batra

Virginia Tech

Administrativia

- Scholar
  - Anybody not have access?
  - Still have problems reading/submitting? Resolve ASAP.
  - Please post questions on Scholar Forum.

- Reading/Material/Pointers
  - Slides PPT on Scholar; PDF on Public Website
  - Scanned handwritten notes on Scholar
  - Readings/Video pointers on Public Website

Plan for today

- Review
  - Graph Concepts
  - Probability Concepts

- Bayesian Networks
  - Directed Acyclic Graphs (DAGs)
  - Conditional Probability Tables (CPTs)
  - Conditional Independence

Plan for the semester

- Start with Bayes Nets
  - Finish
    - Representation
    - Learning
    - Inference
  - Simple setting
    - Discrete Variables
    - Fully Observed
    - Exact Inference & Learning
- Then generalize
  - Undirected Models (MRFs / CRFs)
  - Approximate Inference & Learning
  - Hidden Variables (missing from training data) – EM
  - Continuous Variables (Gaussians, Gaussians everywhere…)

A.M. TURING CENTENARY CELEBRATION WEBCAST

Search TYPE HERE

MORE ACM AWARDS

A.M. TURING AWARD

A.M. TURING AWARD WINNERS BY...

ALPHABETICAL LISTING | YEAR OF THE AWARD | RESEARCH SUBJECT

Photo-Essay

BIRTH:

September 4, 1936, Tel Aviv.

EDUCATION:

B.S., Electrical Engineering (Technion, 1960); M.S., Electronics (Newark College of Engineering, 1961); M.S., Physics (Rutgers University, 1965); Ph.D., Electrical Engineering (Polytechnic Institute of Brooklyn, 1965).

EXPERIENCE:

Research Engineer, New York University Medical School (1960–1961); Instructor,

JUDEA PEARL

United States – 2011

CITATION

For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

SHORT ANNOTATED BIBLIOGRAPHY | ACM DL AUTHOR PROFILE | ACM TURING AWARD LECTURE VIDEO | RESEARCH SUBJECTS | ADDITIONAL MATERIALS

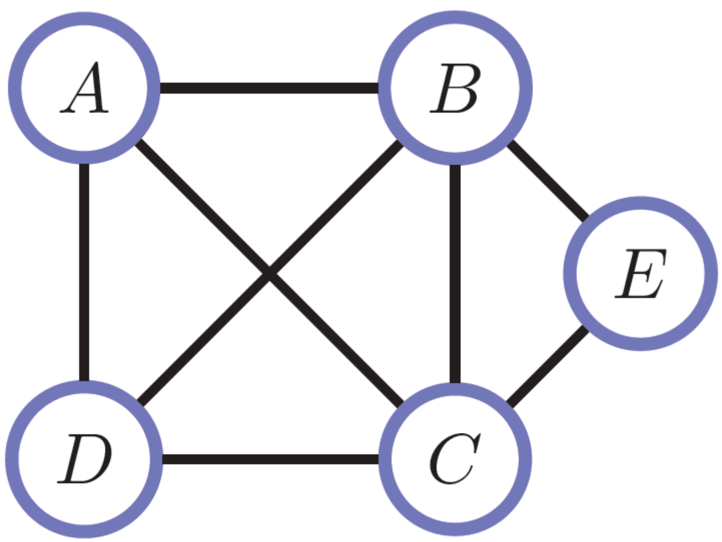
Judea Pearl created the representational and computational foundation for the processing of information under uncertainty.

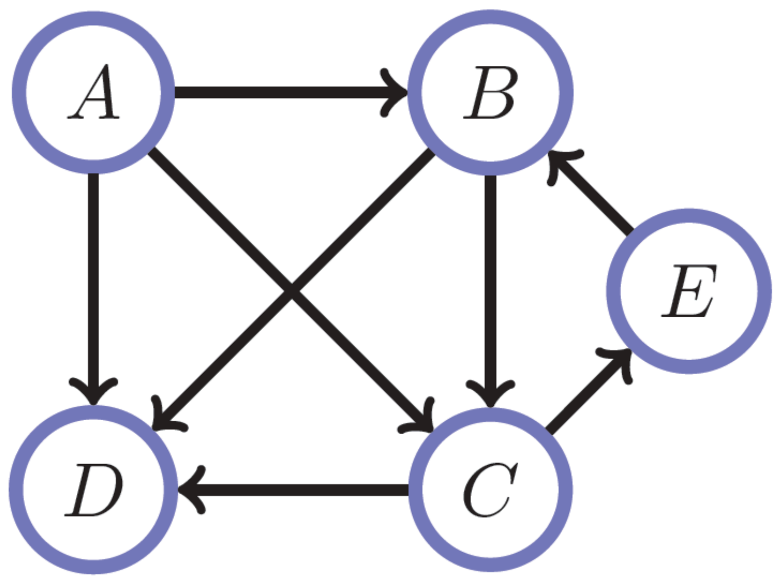
He is credited with the invention of *Bayesian networks*, a mathematical formalism for defining complex probability models, as well as the principal algorithms used for inference in these models. This work not only revolutionized the field of artificial intelligence but also became an important tool for many other branches of engineering and the natural sciences. He later created a mathematical framework for causal inference that has had significant impact in the social sciences.

Judea Pearl was born on September 4, 1936, in Tel Aviv, which was at that time administered under the British Mandate for Palestine. He grew up in Bnei Brak, a Biblical town his grandfather went to reestablish in 1924. In 1956, after serving in the Israeli army and joining a Kibbutz, Judea decided to study engineering. He attended the Technion, where he met his wife, Ruth, and received a B.S. degree in Electrical Engineering in 1960. Recalling the Technion faculty members in a 2012 interview in the *Technion Magazine*, he emphasized the thrill of discovery:

Graphs

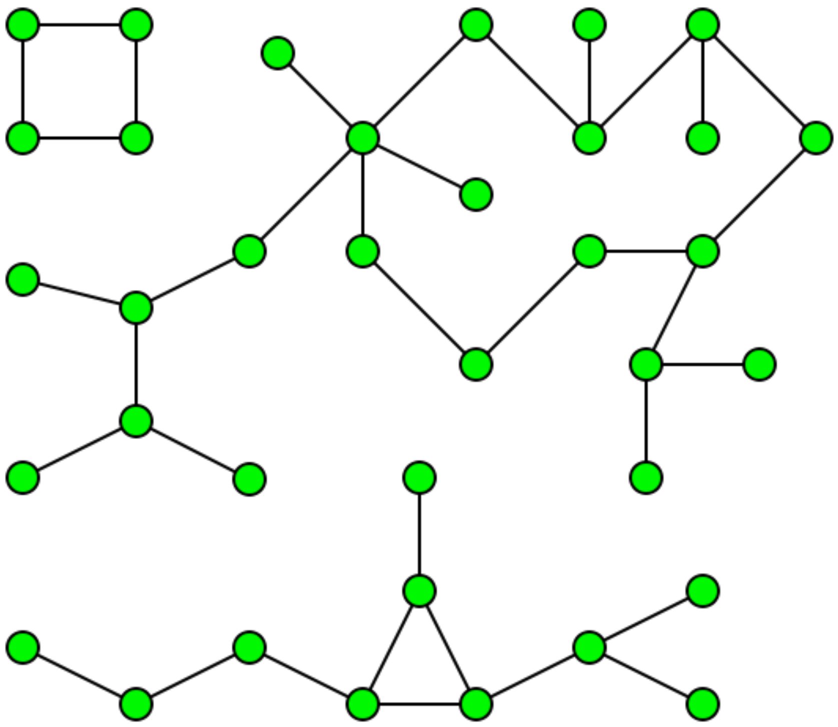
- Concepts:
  - Definition of G
  - Vertices/Nodes
  - Edges
  - Directed vs Undirected graphs
  - Neighbours vs Parent/Child
  - Degree vs In/Out-degree
  - Walk vs Path vs Cycle





Concepts

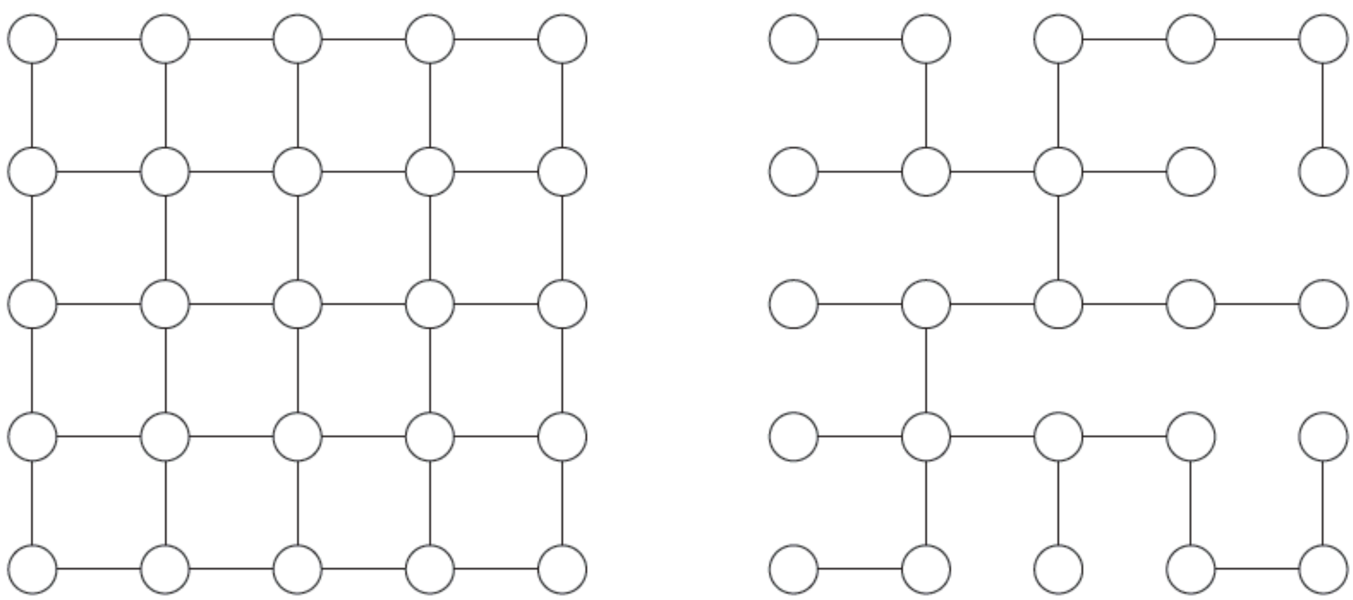
- Connected Components
  - Largest subset of nodes with paths to each other
- Connected vs Disconnected Graphs
  - Single components vs multiple components



Special Graphs

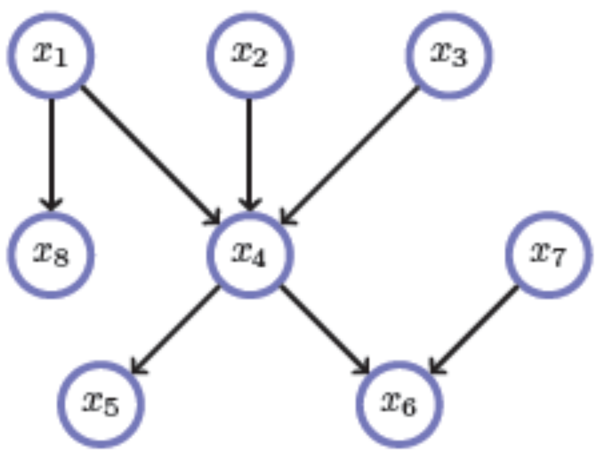
- Tree
  – Undirected graph; no cycles

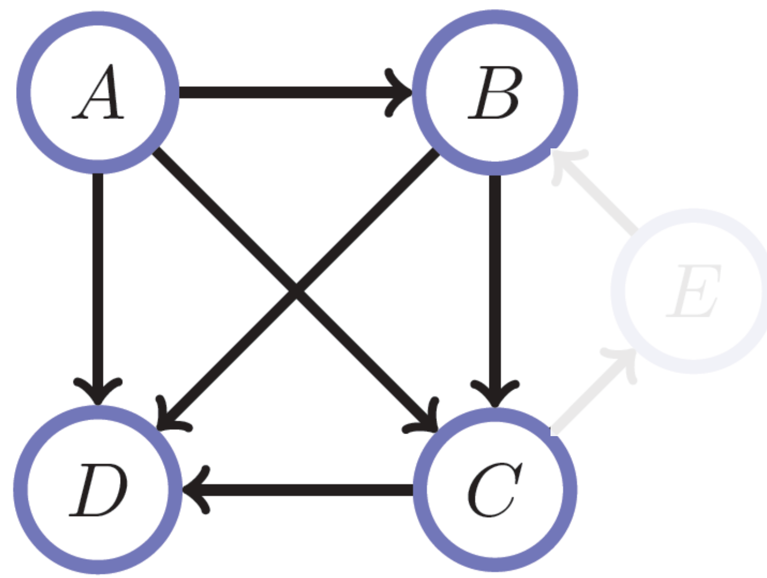
- Spanning Tree
  – Same vertices as G; subset of edges; no cycles; connected



Directed Acyclic Graphs (DAGs)

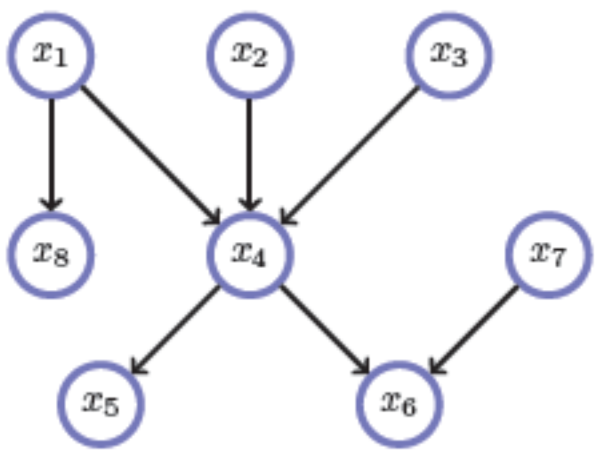
- Exactly what the name suggests
  - Directed edges
  - No (directed) cycles
  - Underlying undirected cycles okay

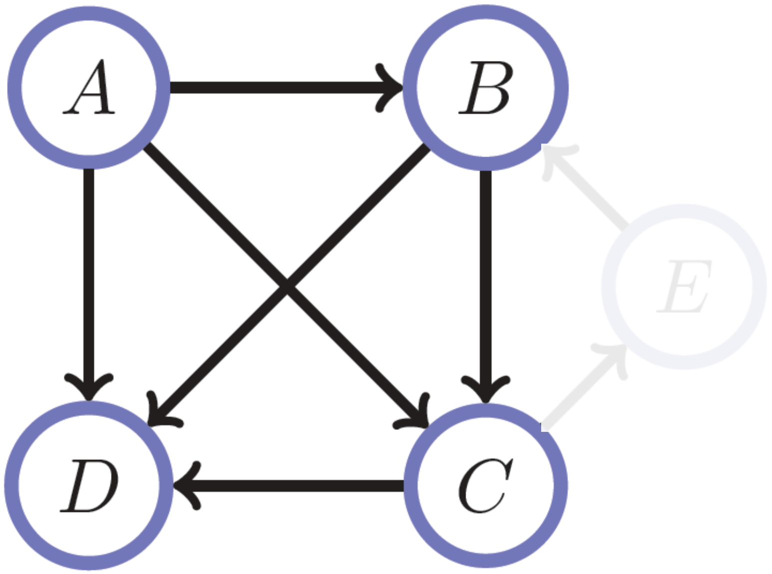




Directed Acyclic Graphs (DAGs)

- Concept
  - Ancestors vs Descendants





Probability

- A is non-deterministic event
  - Can think of A as a boolean-valued variable

- Examples
  - A = your next patient has cancer
  - A = Rafael Nada wins Australian Open 2014

Interpreting Probabilities

- What does P(A) mean?

- Frequentist View
  - limit $N \rightarrow \infty$ #(A is true)/N
  - limiting frequency of a repeating non-deterministic event

- Bayesian View
  - P(A) is your "belief" about A

- Market Design View
  - P(A) tells you how much you would bet

Discrete Random Variables

X → discrete random variable

$\mathcal{X}$ or Val(X) → sample space of possible outcomes, which may be finite or countably infinite

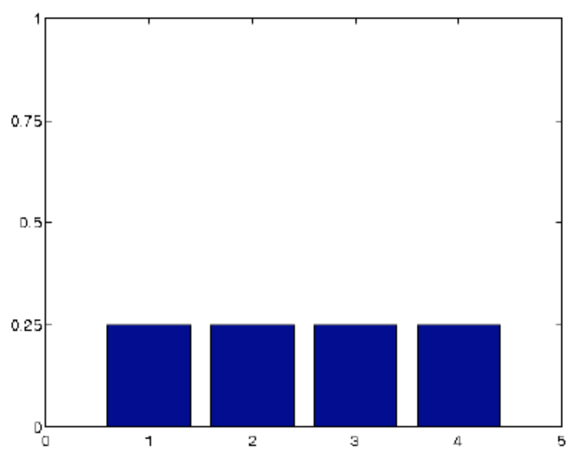
$x \in \mathcal{X}$ → outcome of sample of discrete random variable

$p(X = x)$ → probability distribution (probability mass function)

$p(x)$ → shorthand used when no ambiguity

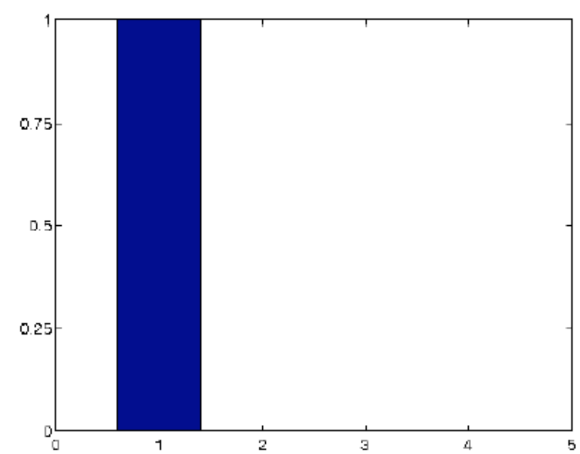
$$0 \leq p(x) \leq 1 \text{ for all } x \in \mathcal{X}$$

$$\sum_{x \in \mathcal{X}} p(x) = 1$$



uniform distribution

$\mathcal{X} = \{1, 2, 3, 4\}$



degenerate distribution

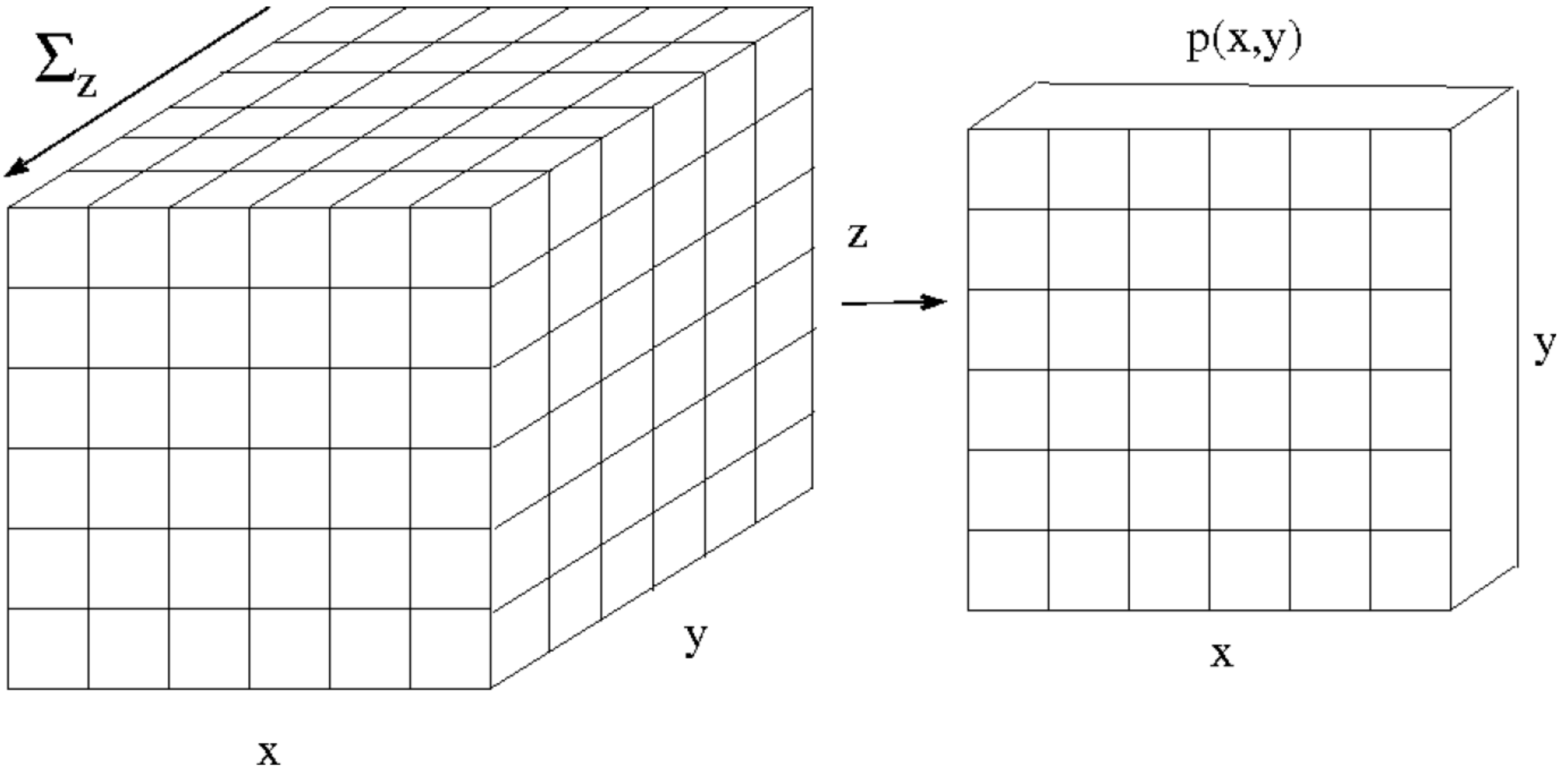
Most Important Concepts

- Marginal distributions / Marginalization
- Conditional distribution / Chain Rule
- Bayes Rule

Marginalization

- Marginalization
  - Events: P(A) = P(A and B) + P(A and not B)
  - Random variables $P(X=x)=\sum_y P(X=x,Y=y)$

Marginal Distributions



$$p(x, y) = \sum_{z \in \mathcal{Z}} p(x, y, z)$$

$$p(x) = \sum_{y \in \mathcal{Y}} p(x, y)$$

Conditional Probabilities

- $P(Y=y \mid X=x)$

- What do you believe about Y=y, if I tell you X=x?

- P(Rafael Nadal wins Australian Open)?

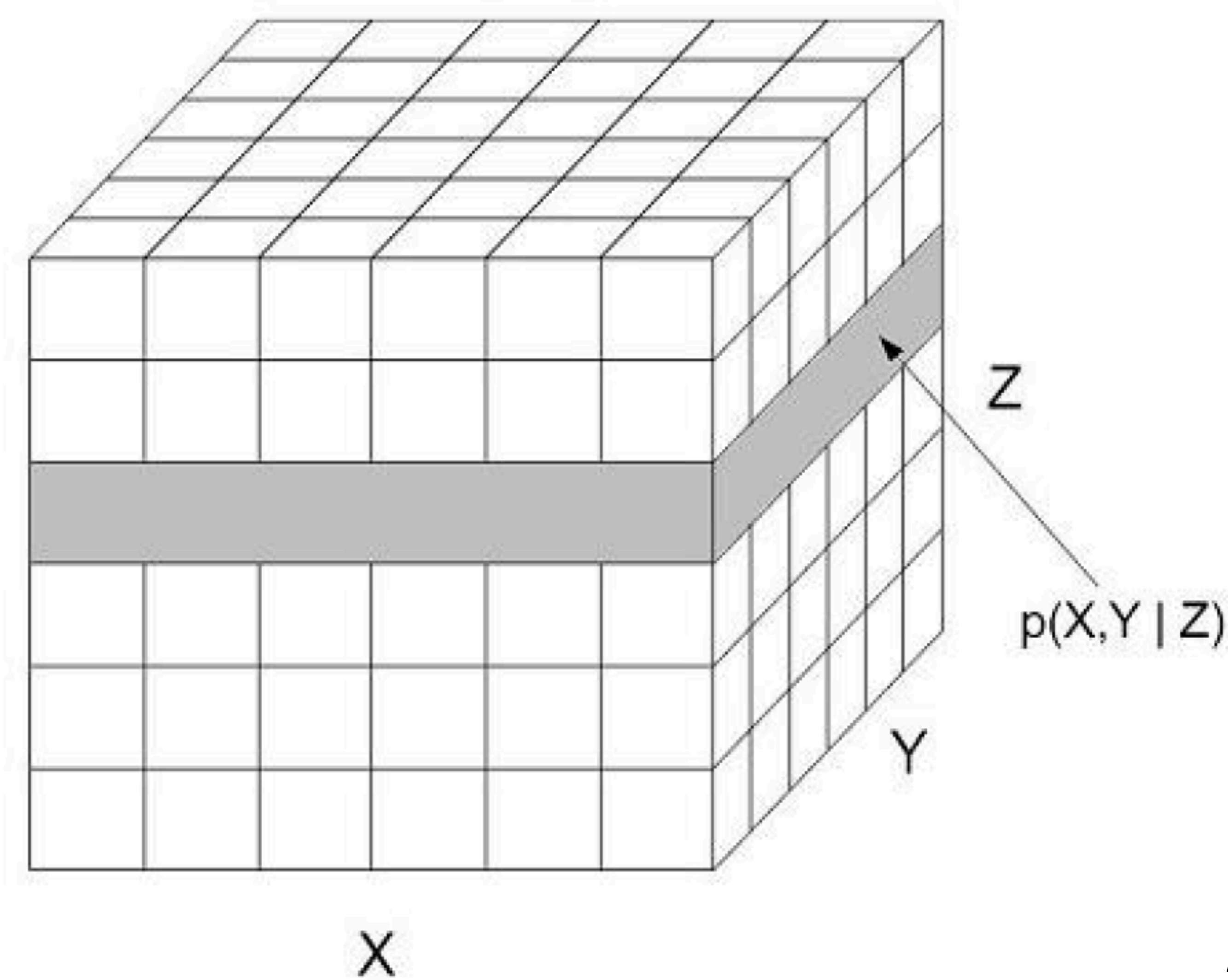
- What if I tell you:
  - He won US Open 2013 and is currently rank 1
  - I offered a similar analysis last fall and he won

Conditional Probabilities

- Definition

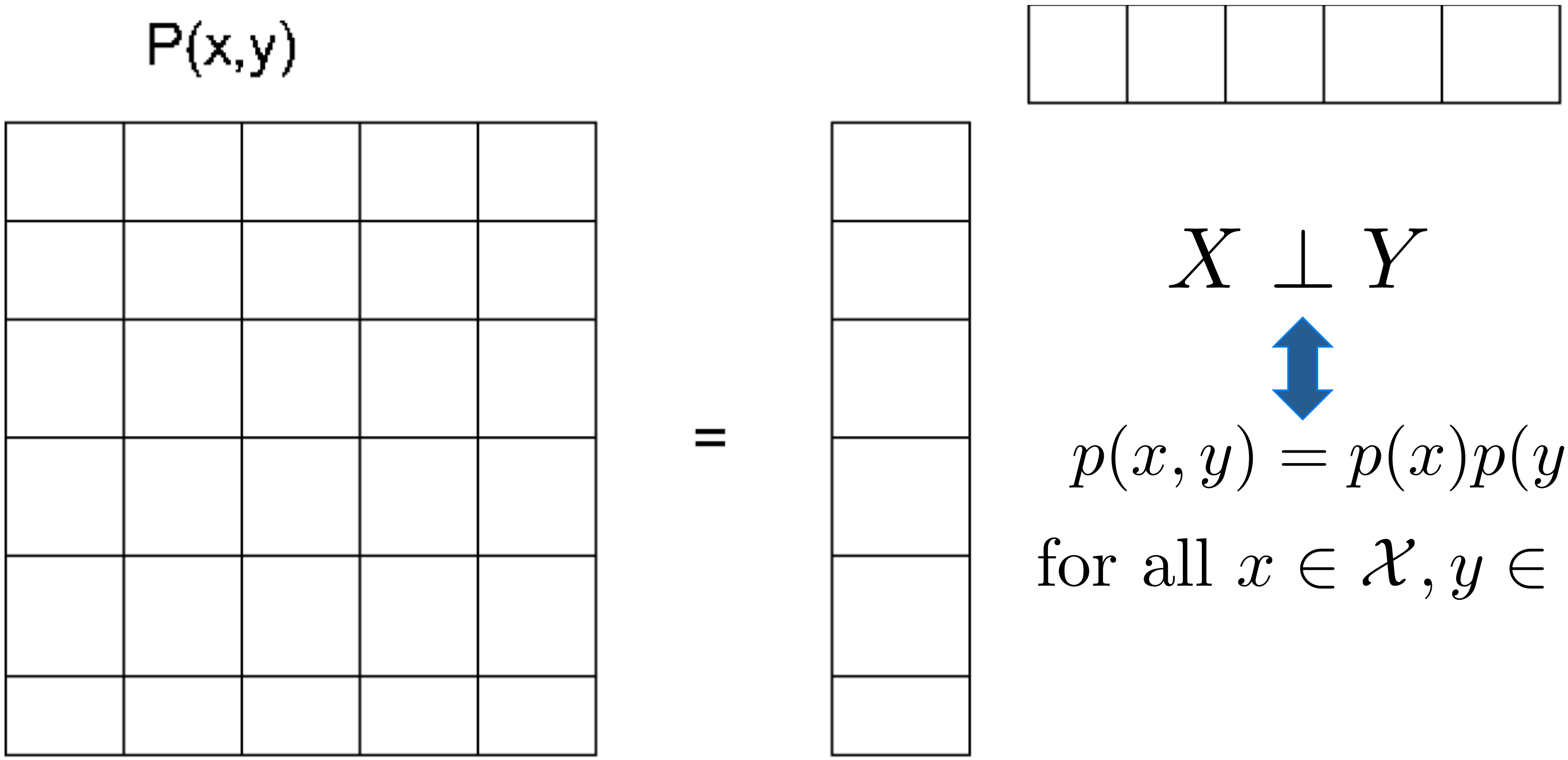
- Corollary: Chain Rule

Conditional Distributions



$$p(x, y \mid Z = z) = \frac{p(x, y, z)}{p(z)}$$

Independent Random Variables



$$X \perp Y$$

$\Updownarrow$

$$p(x,y) = p(x)p(y)$$
$$\text{for all } x \in \mathcal{X}, y \in \mathcal{Y}$$

Marginal Independence

- **Sets** of variables **X**, **Y**

- **X** is independent of **Y**
  - Shorthand: $P \vdash (\mathbf{X} \perp \mathbf{Y})$

- **Proposition:** P satisfies $(\mathbf{X} \perp \mathbf{Y})$ if and only if
  - $P(\mathbf{X=x},\mathbf{Y=y}) = P(\mathbf{X=x})\, P(\mathbf{Y=y})$, $\quad \forall \mathbf{x} \in Val(\mathbf{X}),\ \mathbf{y} \in Val(\mathbf{Y})$

Conditional independence

- **Sets** of variables **X**, **Y**, **Z**

- **X** is independent of **Y** given **Z** if
  - Shorthand: $P \vdash (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$
  - For $P \vdash (\mathbf{X} \perp \mathbf{Y} \mid \emptyset)$, write $P \vdash (\mathbf{X} \perp \mathbf{Y})$

- **Proposition:** *P* satisfies $(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$ if and only if
  - $P(\mathbf{X},\mathbf{Y}|\mathbf{Z}) = P(\mathbf{X}|\mathbf{Z})\, P(\mathbf{Y}|\mathbf{Z})$, $\forall \mathbf{x} \in Val(\mathbf{X}), \mathbf{y} \in Val(\mathbf{Y}), \mathbf{z} \in Val(\mathbf{Z})$

Bayes Rule

- Simple yet profound
  - Using Bayes Rules doesn't make your analysis Bayesian!

- Concepts:
  - Likelihood
    - How much does a certain hypothesis explain the data?
  - Prior
    - What do you believe before seeing any data?
  - Posterior
    - What do we believe after seeing the data?

- End of Reviews

- Start of Bayes Nets

Bayes Nets: Example 1

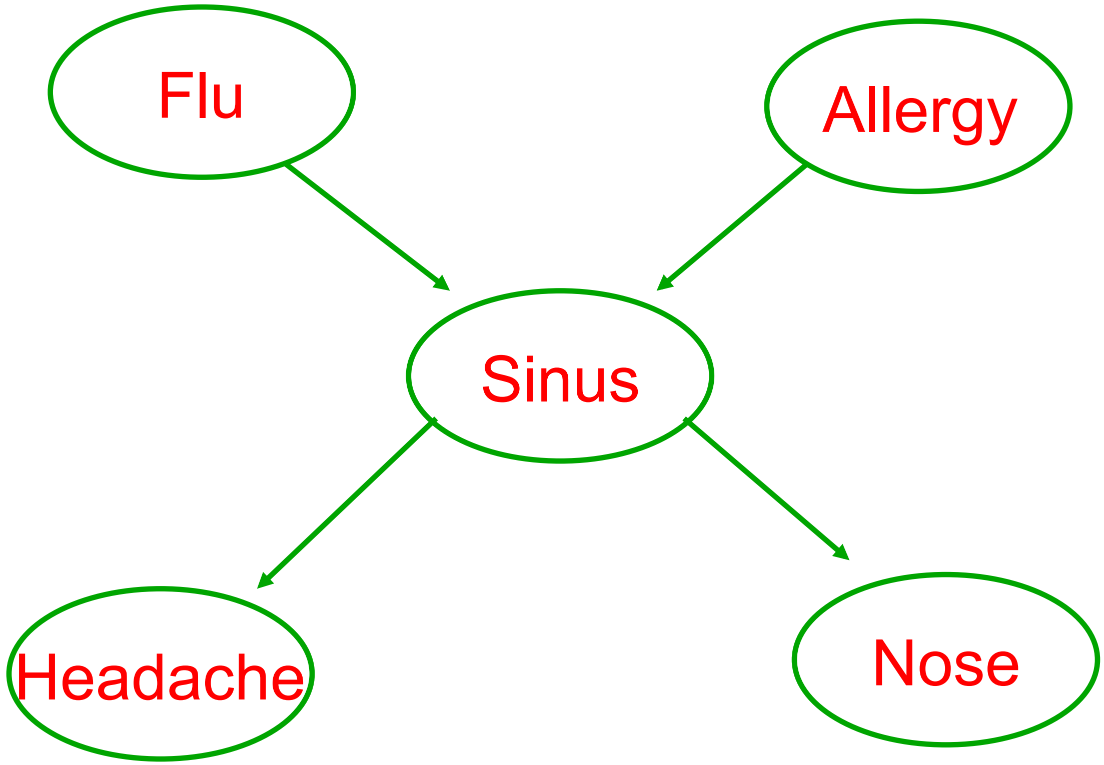
- Compare chain rule to Markov chain
  - On board

Bayes Nets: Example 2

- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches

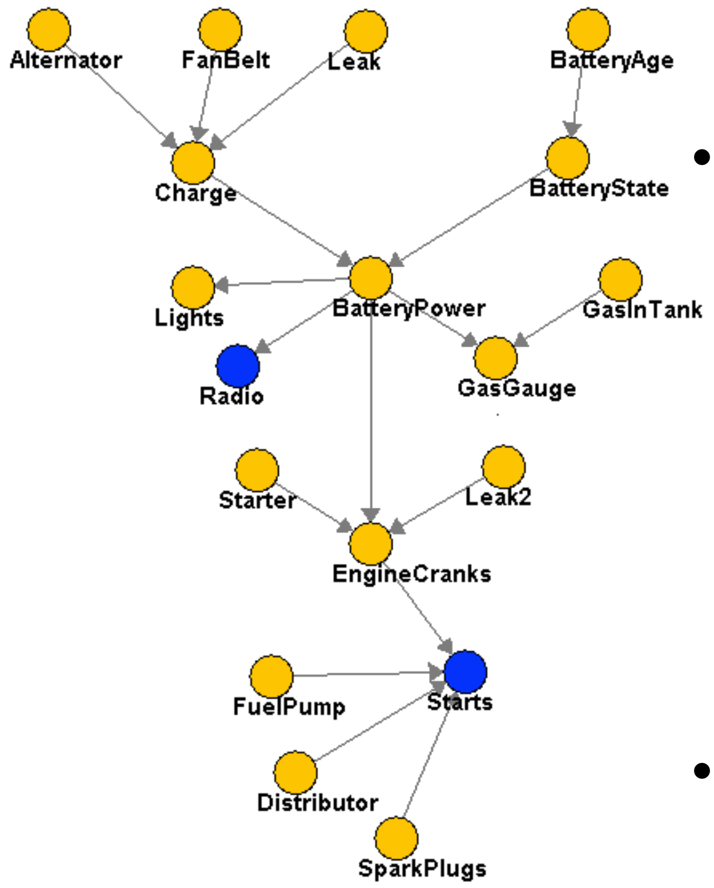
- How are these connected?

Bayes Nets: Example 2



- On board
  - Factored distribution; Count parameters

Car starts BN



- 18 binary attributes
- Inference
  - P(BatteryAge|Starts=f)
- 2^{18} terms, why so fast?

A general Bayes net

- Set of random variables

- Directed acyclic graph
  - Encodes independence assumptions

- CPTs
  - Conditional Probability Tables

- Joint distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

How many parameters in a BN?

- Discrete variables $X_1, \ldots, X_n$

- Graph
  - Defines parents of X_i, $\mathbf{Pa}_{X_i}$

- CPTs – $P(X_i | \mathbf{Pa}_{X_i})$

Bayes Nets

- BN encode (conditional) independence assumptions.

- Which ones?
- And how can we easily read them?

Local Structures

- Causal Trail
  - X → Y → Z
- Evidential Trail
  - X ← Y ← Z
- Common Cause
  - X ← Y → Z
- Common Effect (v-structure)
  - X → Y ← Z