Quiz 1: Gradescope due Friday (posted ~T-W)

COMPSCI 688: Probabilistic Graphical Models

Lecture 2: More Probability and Directed Graphical Models

Dan Sheldon

Manning College of Information and Computer Sciences
University of Massachusetts Amherst

Partially based on materials by Benjamin M. Marlin (marlin@cs.umass.edu) and Justin Domke (domke@cs.umass.edu)

Review

Discrete Distributions

- Sample space Ω
- Atomic probability $p(\omega)$ for all $\omega \in \Omega$

$$p(\omega) \geq 0, \quad \sum_{\omega \in \Omega} p(\omega) = 1$$

- Events $A \subseteq \Omega$ (only things that have probabilities!)

$$P(A) = \sum_{\omega \in A} p(\omega)$$

- Random variable $X : \Omega \to \mathrm{Val}(X)$ has probability mass function (PMF)

$p_Y(1)$ $p_Y(2)$
$p_X(1)$ $p_X(2)$
$p(1)$ $p(2)$

$$p_{\cancel{X}}(x) = P(X(\omega) = x) = P(X = x)$$

$p(x)$

Events vs Random Variables

- A random variable X is a a mapping from Ω to $\mathrm{Val}(X)$
- **But**: for any random variable X, we can also define the probability distribution with sample space $\Omega = \mathrm{Val}(X)$ and atomic probabilities $p_X(x)$. This is the **distribution** of X.
- If we only care about events involving X, it's easier to just define the distribution of X without using a different underlying probability space
- If we care about multiple random variables, we can similarly define their **joint distribution**

Joint Distributions

Random Variables and Data Sets

In ML and stats, probability distributions are defined over records described by multiple attributes modeled as random variables. This leads to joint distributions.

X_1	X_2	attribute X_3	X_4
Gender	Blood Pressure	Cholesterol	Heart Disease
Male	Med	Low	No
record → Male	Hi	Hi	Yes
Male	Med	Med	Yes
Male	Med	Hi	No
Female	Med	Low	No
Male	Low	Med	No

Joint Probability Distributions

- The *joint distribution* of random variables $X_1, \ldots, X_N$ is a probability distribution over their *canonical sample space*
- The *canonical sample space* Ω of $X_1, \ldots, X_N$ is the Cartesian product of their domains $\Omega = \mathrm{Val}(X_1) \times \ldots \times \mathrm{Val}(X_N)$.
- An element of Ω is a joint assignment $(x_1, \ldots, x_N)$
- The joint probability mass function of $X_1, \ldots, X_N$ is

$$p(x_1, \ldots, x_N) = P(X_1 = x_1, \ldots, X_N = x_N)$$

"and"

$$P(X=a, Y=b) = P([X=a] \cap [Y=b])$$

Joint Distributions: Heart Disease Example

Example: The joint distribution over random variables *Gender*, *BloodPressure*, *Cholesterol* and *HeartDisease* is given by a table like this:

all possible records

Ω

Gender	BloodPressure	Cholesterol	HeartDisease	P
F	L	L	N	0.0127
F	L	L	Y	0.0007
F	L	M	N	0.0098
F	L	M	Y	0.0009
F	L	H	N	0.0087
F	L	H	Y	0.0010
...	...	...	...	...

exponential in # RVs

Random Vectors

- It's convenient to use vector-valued random variables $\mathbf{X} = (X_1, ..., X_N)$ (or "random vectors") and assignments $\mathbf{x} = (x_1, ..., x_N)$:

$$P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, ..., X_N = x_N)$$

- The PMF is $p_{\mathbf{X}}(\mathbf{x})$ or just $p(\mathbf{x})$
- This is just notation: it means the same thing as a joint distribution over $(X_1, \ldots, X_N)$
- **Notation**: use $\mathbf{X}_{-i}$ and $\mathbf{x}_{-i}$ for vectors excluding X_i or x_i

$(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N)$

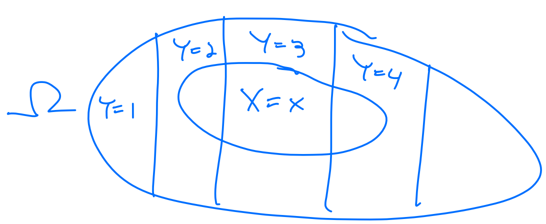
Rules of Probability

Marginal Distributions

$X_1, \ldots, X_N \quad Y_1, \ldots, Y_m$

- Suppose we have a joint distribution $P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})$.
- $P(\mathbf{X} = \mathbf{x})$ is called a *marginal distribution*. How can we find $P(\mathbf{X} = \mathbf{x})$?

$$P(X=x) = \sum_{y \in Val(Y)} P(X=x, Y=y)$$



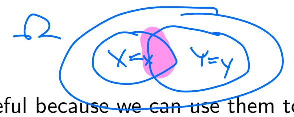
Marginal Distributions: Heart Disease Example

Given a joint distribution on G, BP, C, HD, we obtain the marginal probability $P(G = M, BP = H, C = H)$ as follows:

Gender	BloodPressure	Cholesterol	HeartDisease	P
M	H	H	Y	0.050
M	H	H	N	0.005
M	H	M	Y	0.045
M	H	M	N	0.008
...	...	...	...	...

0.055

Conditional Distributions



- Joint distributions are useful because we can use them to answer queries like "What is the probability that $\mathbf{Y} = \mathbf{y}$ given that I observed $\mathbf{X} = \mathbf{x}$?":

$\forall\, x, y$

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{P(\mathbf{X} = \mathbf{x})}$$

$$= \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{\sum_{\mathbf{y} \in \mathrm{Val}(\mathbf{Y})} P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}$$

free fix

- Write $p(\mathbf{y}|\mathbf{x})$ to denote the PMF of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$

$p_{Y|X}(y|x)$

Conditional Distributions: Heart Disease Example

$$P(HD = Y | G = M, BP = H, C = H) = \frac{P(G = M, BP = H, C = H, HD = Y)}{P(G = M, BP = H, C = H)}$$

$$= \frac{0.050}{0.050 + 0.005} = 0.91$$

Gender	BloodPressure	Cholesterol	HeartDisease	P
M	H	H	$\boxed{Y}$	$\boxed{0.050}$
M	H	H	N	0.005
M	H	M	Y	0.045
M	H	M	N	0.008
...	...	...	...	...

Chain Rule

- By rearranging the definition of conditional probability, we get the chain rule:

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

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

$p_{X,Y}(x,y) = p_{Y|X}(y|x)\, p_X(x)$

- Applying the chain rule repeatedly to a random vector $\mathbf{X}$ gives:

$x_1, x_2, \ldots, x_N$

$$p(\mathbf{x}) = p(x_N | x_1, ..., x_{N-1}) p(x_1, ..., x_{N-1})$$
$$\vdots$$
$$= p(x_N | x_1, ..., x_{N-1}) p(x_{N-1} | x_1, ..., x_{N-2}) \cdots p(x_3 | x_2, x_1) p(x_2 | x_1) p(x_1)$$

Chain Rule: Heart Disease Example

$p(\underbrace{g}_{X}, \underbrace{bp, c, hd}_{Y}) = p(g)\, p(bp, c, hd | g)$

$= p(g)\, p(bp|g)\, p(c, hd | g, bp)$

$= p(g)\, p(bp|g)\, p(c|g, bp)\, p(hd|g, c, bp)$

We can apply the chain rule using any ordering of the variables:

$$p(g, bp, c, hd) = p(hd|c, bp, g)p(c|bp, g)p(bp|g)p(g)$$
$$p(g, bp, c, hd) = p(g|bp, c, hd)p(bp|c, hd)p(c|hd)p(hd)$$
$$p(g, bp, c, hd) = p(c|hd, g, bp)p(hd|g, bp)p(g|bp)p(bp)$$

Bayes' Rule

hidden $\quad$ observed

$X \longrightarrow Y$

$p(x) \quad p(y|x) \longrightarrow p(x,y)$

- By using the definition of conditional probability twice, we obtain one of the most important equations in probability theory, Bayes' Rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(y|x)p(x)}{\sum_x p(x)p(y|x)}$$

- Bayes' rule lets us compute $p(\mathbf{x}|\mathbf{y})$ from a joint distribution specified by $p(\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x})$.

Conditional Independence

Probabilistic Models

The solution to the problem of exponential-sized joint distributions is the use of **compact** probabilistic models.

- Bayesian networks achieve compactness by exploiting the chain rule and asserting (conditional) independence relations
- As a result, Bayesian networks can express high-dimensional distributions as products of simpler factors.

Marginal Independence

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

"X is indept. of Y"

$$\mathbf{X} \perp \mathbf{Y} \iff p(\mathbf{x},\mathbf{y}) = p(\mathbf{x})p(\mathbf{y}) \quad \forall x, y$$

$= p(y)p(x|y)$

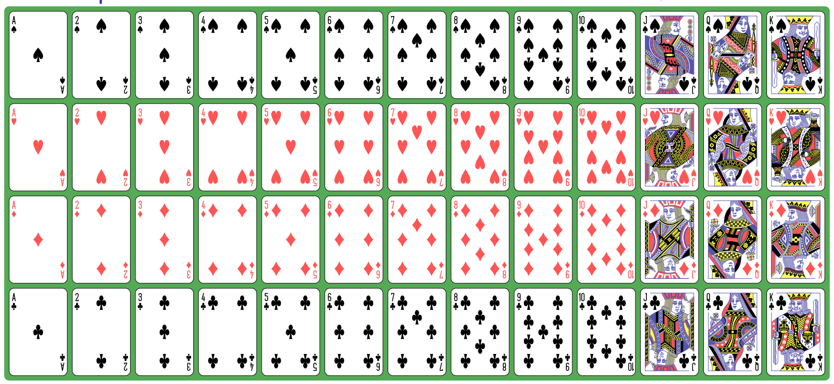
X, Y, Z

$$\mathbf{X} \perp \mathbf{Y} \iff p(x|y) = p(x)$$

$$\mathbf{X} \perp \mathbf{Y} \iff p(y|x) = p(y)$$

Card Example I

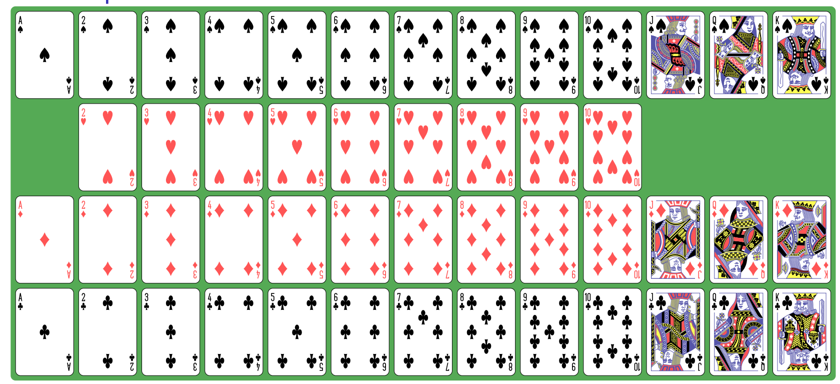
$p(\text{value}=2 \mid \text{col}=r) = \frac{1}{13}$ $\quad p(\text{value}=2) = \frac{1}{13}$



Draw a random card: is value $\perp$ color?

Card Example II

$p(\text{val}=J \mid \text{color}=\text{red}) \neq p(\text{val}=J)$



What about with this deck? Is value $\perp$ color? No

Conditional Independence

$p_{Y,X|Z}(y,x|z) = p_{X|Z}(x|z)\, p_{Y|Z}(y|z)$

"X is conditionally indept of Y given Z"

$$\mathbf{X} \perp \mathbf{Y} | \mathbf{Z} \iff p(\mathbf{y},\mathbf{x}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z}) \quad \forall x,y,z$$

$$\mathbf{X} \perp \mathbf{Y} | \mathbf{Z} \iff p(x|y,z) = p(x|z)$$

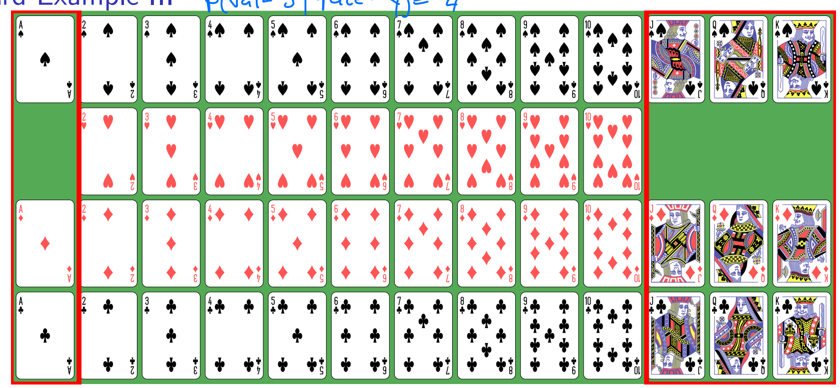
$$\mathbf{X} \perp \mathbf{Y} | \mathbf{Z} \iff p(y|x,z) = p(y|z)$$

"Given z, knowing y tells you nothing (else) about dist of x"

Card Example III

$p(\text{val}=J \mid \text{col}=\text{bl}, \text{face}=y) = \frac{1}{4}$

$p(\text{val}=J \mid \text{face}=y) = \frac{1}{4}$



Is value $\perp$ color | facecard? $\boxed{\text{Yes}}$

$p(\text{val} \mid \text{col}, \text{face})$

$p(\text{val} \mid \text{face})$