CS5841/EE5841 Machine Learning

Lecture 2: Probability Review

Evan Lucas

Overview

- Basic probability
- Baye's Rule
- Random variables and distributions

Presentation contains material from Dr. Tim Havens, used with permission

Class updates

- Friday (January 12) class is cancelled in person
- We will have a video tutorial and extra credit assignment to introduce Julia
- Discord server link is on Canvas front page
- Two surveys due Friday!

Related reading

- Strongly suggested
  - Section 1.2 in Bishop
- Additional
  - Chapters 2-4 in Murphy
  - Chapter 3 in Goodfellow

Probability taxonomy

- **Experiment**
  - How is data being collected?
- **Sample Space**
  - What are the possible outcomes of the experiment?
- **Event**
  - A subset of the possible outcomes
- **Probability of an event**
  - The likelihood of an event occurring

Probability example

- **<u>Experiment</u>**
  - Flip 2 coins
- **<u>Sample Space</u>**
  - S = {HH, TT, HT, TH}
- **<u>Event</u>**
  - A = {HH}; B = {HT, TH}
- **<u>Probability of an event</u>**
  - Axiom 1: Pr(A) ⩾ 0
  - Axiom 2: Pr(S) = 1
  - Axiom 3: For every sequence of disjoint events

$$Pr\left(\bigcup_i A_i\right) = \sum_i Pr(A_i)$$

Frequentist statistics

- Count things!
- Example: Pr(A) = n(A)/N (frequentist statistics)

Joint Probability

- For events A and B, joint probability Pr(AB) [also shown as Pr(A,B)] denotes the probability that both events happen
- Example: A = {HH}, B = {HT, TH}, what is the joint probability Pr(AB)?

Independence

- Two events A and B are independent iff
  - $Pr(AB) = Pr(A)Pr(B)$
- A set of events $\{A_i\}$ is independent iff

$$Pr\left(\bigcap_i A_i\right) = \prod_i Pr(A_i)$$

Independence, continued

Consider the experiment of tossing a coin twice

- Example 1:
  - A = {HT, HH}, B = {HT}
  - Is A independent from event B?
- Example 2:
  - A = {HT}, B = {TH}
  - Is A independent from event B?
- Disjoint ≠ independence
- If A is independent from B, B is independent from C, is A independent from C?

Conditioning

- If A and B are events with Pr(A) > 0, the **conditional probability of B given A** is

$$Pr(B|A) = \frac{Pr(AB)}{Pr(A)}$$

- If A is independent from B, what is the relationship between Pr(A|B) and Pr(A)?

Conditional Probability Example

$$Pr(B|A) = \frac{Pr(AB)}{Pr(A)}$$

	Female	Male
Success	200	1800
Failure	1800	200

Pr(A) = n(A)/N

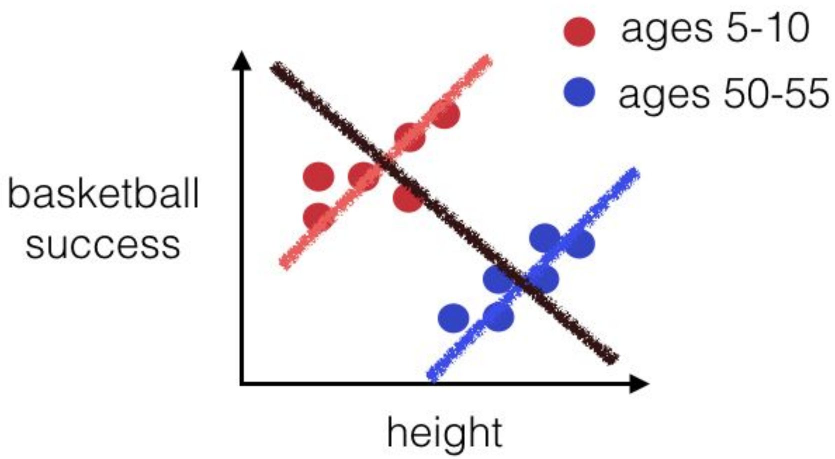
A = {patient is female}
B = {drug fails}
Pr(B|A) = ?
Pr(A|B) = ?

Conditional Probability Example, cont.

- Pr(A) = n(A)/N = 2000/4000 = 0.5
- Pr(B) = n(B)/N = 2000/4000 = 0.5
- Pr(AB) = n(AB)/N = 1800/4000 = 0.45
- Pr(B|A) = Pr(AB)/Pr(A) = 0.45/0.5 = 0.9
- Pr(A|B) = Pr(AB)/Pr(B) = 0.45/0.5 = 0.9

Simpson's Paradox

A trend in data is reversed when data is grouped differently



Simpson's Paradox example

	Female		Male	
	Drug 1	Drug 2	Drug 1	Drug 2
Success	200	10	19	1000
Failure	1800	190	1	1000

Case 1: Grouped by drugs

Case 2: Grouped by gender

Simpson's Paradox example, Case 1

	Female		Male	
	Drug 1	Drug 2	Drug 1	Drug 2
Success	200	10	19	1000
Failure	1800	190	1	1000

Drug 2 is better than Drug 1

	Drug 1	Drug 2
Success	219	1010
Failure	1801	1190

A = {Using Drug 1}
B = {Using Drug 2}
C = {Drug succeeds}
Pr(C|A) is about 10%
Pr(C|B) is about 50%

Simpson's Paradox example, Case 2

	Female	
	Drug 1	Drug 2
Success	200	10
Failure	1800	190

Male	
Drug 1	Drug 2
19	1000
1	1000

A = {Using Drug 1}
B = {Using Drug 2}
C = {Drug succeeds}
Pr(C|A) is about 20%
Pr(C|B) is about 5%

A = {Using Drug 1}
B = {Using Drug 2}
C = {Drug succeeds}
Pr(C|A) is about 100%
Pr(C|B) is about 50%

Drug 1 is better than Drug 2

Conditional Independence

- Event A and B are conditionally independent given C iff
  - Pr(AB|C) = Pr(A|C)Pr(B|C)
- A set of events {Ai} is conditionally independent given C iff

$$Pr\left(\bigcup_i A_i | c\right) = \prod_i Pr(A_i|C)$$

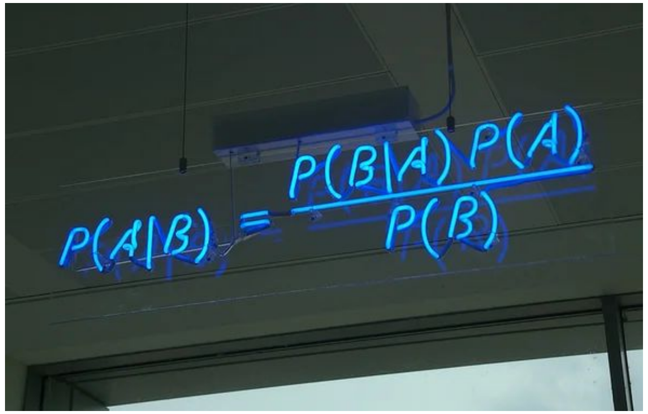
Conditional Independence, continued

- Example: Consider three events: A, B, C
  - $\Pr(A) = \Pr(B) = \Pr(C) = 1/5$
  - $\Pr(A,C) = \Pr(B,C) = 1/25$, $\Pr(A,B) = 1/10$
  - $\Pr(A,B,C) = 1/125$
  - Are A and B independent?
  - Are A and B conditionally independent given C?
- A and B are independent ≠ A and B are conditionally independent

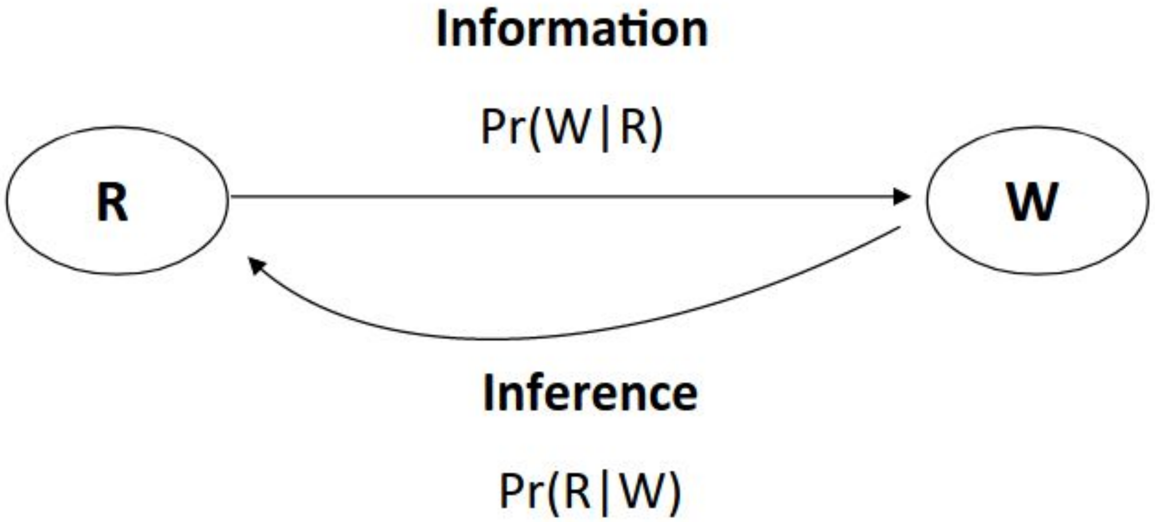
Baye's Rule

- Given two events A and B, where Pr(A)>0, then:

$$Pr(B|A) = \frac{Pr(AB)}{Pr(A)} = \frac{Pr(A|B)Pr(B)}{Pr(A)}$$



Baye’s Rule, continued



$$\underbrace{Pr(B|A)}_{\text{Posterior}} = \frac{Pr(AB)}{Pr(A)} = \frac{\overbrace{Pr(A|B)}^{\text{Likelihood}}\overbrace{Pr(B)}^{\text{Prior}}}{Pr(A)}$$

Baye's Rule example

$$Pr(B|A) = \frac{Pr(AB)}{Pr(A)} = \frac{Pr(A|B)Pr(B)}{Pr(A)}$$

Example:

- R: It's rainy
- W: Grass is wet
- Pr(R) = 0.8
- Pr(R|W) = ?

Pr(W|R)	R	~R
W	0.7	0.4
~W	0.3	0.6

Interlude: How do we get Pr(A)

- Law of Total Probability

$$P(B) = P(B \mid A_1)\,P(A_1) + \ldots + P(B \mid A_n)\,P(A_n).$$

Baye's Rule example worked

$$Pr(B|A) = \frac{Pr(AB)}{Pr(A)} = \frac{Pr(A|B)Pr(B)}{Pr(A)}$$

Prior: Pr(R) = 0.8

Likelihood: Pr(W|R) = 0.7

Normalization: Pr(W) = Pr(W|R)*P(R) + Pr(W|~R)*P(~R) =

0.7*0.8 + 0.4*0.2 = 0.64

All together: Pr(R|W) = (0.7*0.8)/0.64 =0.875

Random variables and distributions

- A random variable X is a numerical outcome of a random experiment
- The distribution of a random variable is the collection of possible outcomes along with their probabilities
  - Discrete case: $Pr(X = x) = p_\theta(x)$
  - Continuous case: $Pr(a \leq X \leq b) = \int_a^b p_\theta(x)dx$

PDFs, CDFs, etc.

Probability density function (PDF): a continuous function that describes the possible values and likelihoods of a random variable, represented as $f_x(x)$

Probability mass function (PMF): PDF specifically for discrete random variables, often represented as $p_x(x)$

Cumulative distribution function (CDF): a continuous function that describes the probability that a random variable is less than or equal to a given value, represented as $F_x(x)$

Random variable example

- Let S be the set of all sequences of two rolls of a fair 6-sided die. Let X be the sum of both rolls
  - What are all possible values for X?
  - Pr(X=2) = ?
    - Frequentist statistics!
    - 1 possible sequence sums to 2, 36 different sequences possible
    - 1/36
  - Pr(X=7) = ?

Expected value (mean)

- Given a random variable X~Pr(X=x)
- Expectation is: $E[X] = \sum_x xPr(X = x)$
    - For an empirical sample: $x_1, x_2, \ldots, x_n$: $E[X] = \frac{1}{N}\sum_{i=1}^{N} x_i$
- Continuous case:

$$E[X] = \int_{-\infty}^{\infty} xp_\theta(x)dx$$

- Expectation of a sum of random variables:

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

Expectation examples

$$E[X] = \sum_{x} xPr(X = x)$$

- Let X be the sum of two rolls of a fair 6-sided die
  - What is E[x]?
  - E[x] = 2*1/36 + 3* 2/36 + …
-

Variance

- The variance of a random variable X is the expectation of (X-E[X])2

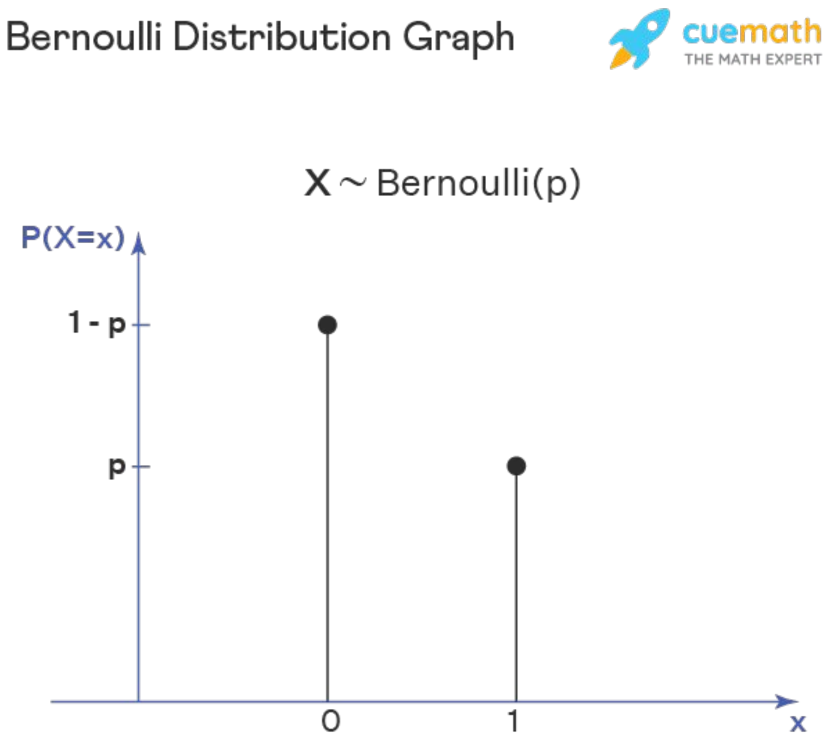
$$
\begin{aligned}
Var(X) &= E\left[(X - E[X])^2\right] \\
&= E\left[X^2 + E[X]^2 - 2XE[X]\right] \\
&= E\left[X^2 - E[X]^2\right] \\
&= E\left[X^2\right] - E[X]^2
\end{aligned}
$$

Some common distributions

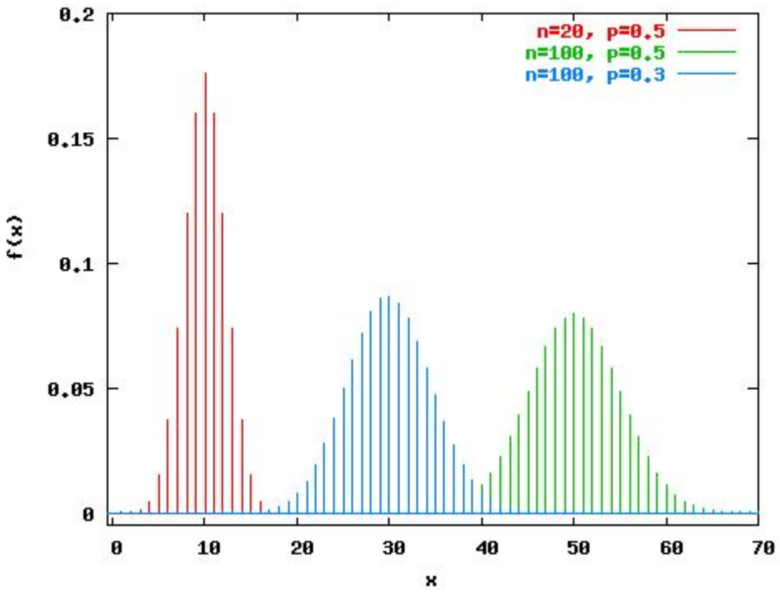
- Bernoulli
- Binomial
- Poisson
- Gaussian (aka. Normal)

Bernoulli Distribution

- The outcome of an experiment is a success or a failure
  - Pr(X=1) = p, Pr(X=0) = 1-p



Binomial Distribution



- n draws of a Bernoulli distribution
  - X_i~Bernoulli(p), X~Bin(p,n)
- Random experiment X stands for number of successes

$$Pr(X=x) = p_\theta(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 1, 2, \ldots, n \\ 0 & \text{else} \end{cases}$$

- E[X] = np, Var(X) = np(1-p)

Poisson Distribution

- Comes from Binomial Distribution
  - Fix expectation E[x] = λ
  - Let number of trials n >> λ
  - Binomial distribution becomes a Poisson distribution

$$Pr(X = x) = p_\theta(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

- E[x] = Var[x] = λ

Gaussian (Normal) Distribution

- $X \sim N(\mu, \sigma^2)$

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$$

$$Pr(a \leq X \leq b) = \int_a^b p_\theta(x)$$

- E[x] = ?, Var(X) = ?

Questions + Comments?