Victory Lap

CSE 312 23Su
Lecture 24

Announcement

Please fill out the course evaluation!

Show up to the exam on Friday ☺ The staff is available in office hours and over Ed to help you out until then.

My OH today are cancelled.

What’s Next?

CSE 421: Algorithms (design algos – uses some combinatorics)

CSE 422: Toolkit for Modern Algorithms

CSE 426: Cryptography

CSE 427: Computational Biology

CSE 446: Machine Learning (probability + linear algebra)

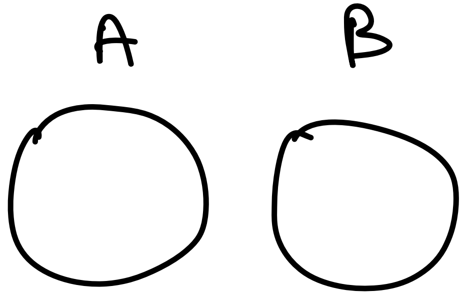
CSE 447: Natural Language Processing

CSE 473: Artificial Intelligence (Bayes nets and such)

CSE 490Q: Quantum Computing

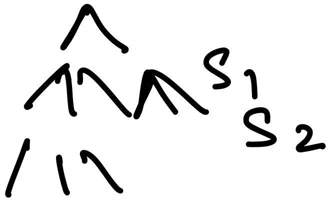
Once upon a time, there was: counting

Sum rule – choose from mutually exclusive (disjoint) sets



$$|A \cup B| = |A| + |B|$$

Product rule – decisions in a sequential process



$$s_1 \cdot s_2 \cdot \ \ldots \ \cdot s_n$$

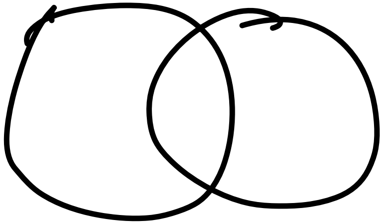
Once upon a time, there was: counting

Complementary counting – sometimes, the complement is easier to count ($|A| = |\mathcal{U}| - |\bar{A}|$)

$\geq x$

not x

Inclusion-exclusion: choose from possibly overlapping sets



$|A \cup B| = |A| + |B| - |A \cap B|$

Once upon a time, there was: counting

$n \cdot n-1 \cdot n-2 \cdot \ldots \cdot 1 = n!$

Permutations – order distinct elements

k-permutation

The number of k-element sequences of distinct symbols from a universe of n symbols is:

$$P(n,k) = \underbrace{n}_{1} \cdot \underbrace{(n-1)}_{2} \cdots \underbrace{(n-k+1)}_{k} = \frac{n!}{(n-k)!}$$

Combinations – number of subsets of a set

k-combination

The number of k-element subsets from a set of n symbols is:

$$C(n,k) = \frac{P(n,k)}{k!} = \frac{n!}{k!\,(n-k)!}$$

Once upon a time, there was: counting

Combinatorial proofs

$X = Y$

$S.$

$$\left.\begin{aligned} X &= |S| \\ Y &= |S| \end{aligned}\right] =$$

Pigeonhole principle

If you have n pigeons and k pigeonholes, then there is **at least one** pigeonhole that has at least $\left\lceil \frac{n}{k} \right\rceil$ pigeons.

x x xx

$\left\lceil \frac{n}{k} \right\rceil$

Stars-and-bars – indistinguishable objects into distinguishable bins

$$\frac{(n+k-1)!}{(k-1)!\,n!} = \binom{n+k-1}{k-1} = \binom{n+k-1}{n}$$

Counting sets leads to… discrete probability

$\omega \in \Omega$

Begin quantifying our uncertainty.

$\mathbb{P}(E) = \sum_{\omega \in \Omega} \mathbb{P}(\omega)$

Probability Space

A (discrete) probability space is a pair $(\Omega, \mathbb{P})$ where:

Ω is the sample space

$\mathbb{P}: \Omega \to [0,1]$ is the probability measure.

$\mathbb{P}$ satisfies:

- $\mathbb{P}(x) \geq 0$ for all x (*non-negativity*)
- $\sum_{x \in \Omega} \mathbb{P}(x) = 1$ (*normalization)*
- If $E, F \subseteq \Omega$ and $E \cap F = \emptyset$ then $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$ (*countable additivity*)

Counting sets leads to… discrete probability

Equally likely outcomes $\omega \in \Omega$

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

Corollaries: complementation, monotonicity, inclusion-exclusion

- $\mathbb{P}(\bar{E}) = 1 - \mathbb{P}(E)$ (complementation)
- If $E \subseteq F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$ (monotonicity)
- $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$ (inclusion-exclusion)

Sometimes we have extra information…

Restrict the sample space

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

Bayes uses a prior to calculate a posterior

prior

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$

posterior

Sometimes we have extra information…

Law of Total Probability

partition Ω: mutually exclusive
exhaustive

$$\mathbb{P}(B) = \sum_{E_i} \mathbb{P}(B|E_i)\,\mathbb{P}(E_i) = \sum_{E_i} \mathbb{P}(B \cap E_i)$$

Chain rule

Chain Rule

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}) \cdot \mathbb{P}(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}) \cdots \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_1)$$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B|A)\,\mathbb{P}(C|A, B)$$

Sometimes that extra is irrelevant…

Independence

Pairwise Independence

Events $A_1, A_2, \dots, A_n$ are pairwise independent if
$\mathbb{P}(\boldsymbol{A_i} \cap \boldsymbol{A_j}) = \mathbb{P}(\boldsymbol{A_i}) \cdot \mathbb{P}(\boldsymbol{A_j})$ for all $\boldsymbol{i}, \boldsymbol{j}$ where $\boldsymbol{i} \neq \boldsymbol{j}$

$\binom{n}{2}$

Mutual Independence

Events $A_1, A_2, \dots, A_n$ are mutually independent if
$\mathbb{P}(\boldsymbol{A_{i_1}} \cap \boldsymbol{A_{i_2}} \cap \cdots \cap \boldsymbol{A_{i_k}}) = \mathbb{P}(\boldsymbol{A_{i_1}}) \cdot \mathbb{P}(\boldsymbol{A_{i_2}}) \cdots \mathbb{P}(\boldsymbol{A_{i_k}})$
for every subset $\{\boldsymbol{i_1}, \boldsymbol{i_2}, \dots, \boldsymbol{i_k}\}$ of $\{\boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{n}\}$.

$\binom{n}{1} \dots \binom{n}{n}$

Sometimes that extra is irrelevant…

Conditional independence

$\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C)$

We say A and B are conditionally independent on C if

$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$

Summarize Numerical Info from Ω

(Discrete) Random variables

> Range Ω_X $\quad p_X(x) > 0$

> PMF $\quad p_X(x) = \mathbb{P}(X = x) \qquad \mathbb{R}$

> CDF $\quad F_X(x) = \mathbb{P}(X \leq x)$

> Expectation

$$\mathbb{E}[X] = \sum_{x \in \Omega} x \cdot \mathbb{P}(X = x)$$

Summarize Numerical Info from Ω

(Discrete) Random variables

> Linearity of Expectation

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$X_i = \begin{cases} 1 & \text{event } A \\ 0 & \text{o.w} \end{cases} \qquad X = \sum_{i=1}^{n} X_i$$

Summarize Numerical Info from Ω

(Discrete) Random variables

> Variance

Variance

The variance of a random variable X is

$$\mathrm{Var}(X) = \sum_{\omega} \mathbb{P}(\omega) \cdot (X(\omega) - \mathbb{E}[X])^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Summarize Numerical Info from Ω

$X \sim \mathrm{Unif}(a,b)$	$X \sim \mathrm{Ber}(p)$	$X \sim \mathrm{Bin}(n,p)$	$X \sim \mathrm{Geo}(p)$
$p_X(k) = \frac{1}{b-a+1}$ $\mathbb{E}[X] = \frac{a+b}{2}$ $\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$	$p_X(0) = 1-p;$ $p_X(1) = p$ $\mathbb{E}[X] = p$ $\mathrm{Var}(X) = p(1-p)$	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $\mathbb{E}[X] = np$ $\mathrm{Var}(X) = np(1-p)$	$p_X(k) = (1-p)^{k-1}p$ $\mathbb{E}[X] = \frac{1}{p}$ $\mathrm{Var}(X) = \frac{1-p}{p^2}$

$X \sim \mathrm{NegBin}(r,p)$	$X \sim \mathrm{HypGeo}(N,K,n)$	$X \sim \mathrm{Poi}(\lambda)$
$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ $\mathbb{E}[X] = \frac{r}{p}$ $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$	$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$ $\mathbb{E}[X] = n\frac{K}{N}$ $\mathrm{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$	$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ $\mathbb{E}[X] = \lambda$ $\mathrm{Var}(X) = \lambda$

Our sample spaces may be uncountable!

Continuous RVs

PDFs

$\mathbb{P}(X = x) \quad \frac{1}{\infty} \to 0$

$f_X(x)$

$F_X(k) = \mathbb{P}(X \le k) = \int_{-\infty}^{k} f_X(z)\,\mathrm{d}z$

$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(z) \cdot f_X(z)\,\mathrm{d}z$

$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,dx$

$= F_X(b) - F_X(a)$

Our sample spaces may be uncountable!

$X \sim \text{Unif}(a, b)$	$X \sim \text{Exp}(\lambda)$	$X \sim \mathcal{N}(\mu, \sigma^2)$
$f_X(k) = \frac{1}{b-a}$ $\mathbb{E}[X] = \frac{a+b}{2}$ $\text{Var}(X) = \frac{(b-a)^2}{12}$	$f_X(k) = \lambda e^{-\lambda k}$ for $k \geq 0$ $\mathbb{E}[X] = \frac{1}{\lambda}$ $\text{Var}(X) = \frac{1}{\lambda^2}$	$f_X(k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $\mathbb{E}[X] = \mu$ $\text{Var}(X) = \sigma^2$

$\mathbb{P}(X \leq 5)$

$\mathbb{P}\left(\frac{X - \mu}{\sigma} \leq \frac{5-\mu}{\sigma}\right)$

An Approximation…

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, with mean μ and variance σ^2. Let $Y_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$

As $n \to \infty$, the CDF of Y_n converges to the CDF of $\mathcal{N}(0, 1)$

i.i.d

$x = x_1 + x_2 + \ldots + x_n$

$x \sim \mathcal{N}(n\mu, n\sigma^2)$

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$\bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

CLT E.g.

$$X = \sum_{i=1}^{n} X_i$$

$$\sim N(n\mu, n\sigma^2)$$

$$\mathbb{P}(a \le X \le b)$$

$$= \mathbb{P}\left(\underbrace{\frac{a - n\mu}{\sqrt{n\sigma^2}}}_{x_1} \le \frac{X - n\mu}{\sqrt{n\sigma^2}} \le \underbrace{\frac{b - n\mu}{\sqrt{n\sigma^2}}}_{x_2} \right)$$

$$\approx P(x_1 \le Z \le x_2) = \Phi(x_2) - \Phi(x_1)$$

More than 1 RV? You got it!

	Discrete	**Continuous**
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Conditional PMF/PDF	$p_{X \mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_{x} x p_{X \mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y) dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

More than 1 RV? You got it!

Calculating probabilities

$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d)$ $\int_a^b \int_c^d$

Law of Total Expectation

Let $A_1, A_2, \ldots, A_k$ be a partition of the sample space, then

$$\mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{E}[X|A_i]\mathbb{P}(A_i)$$

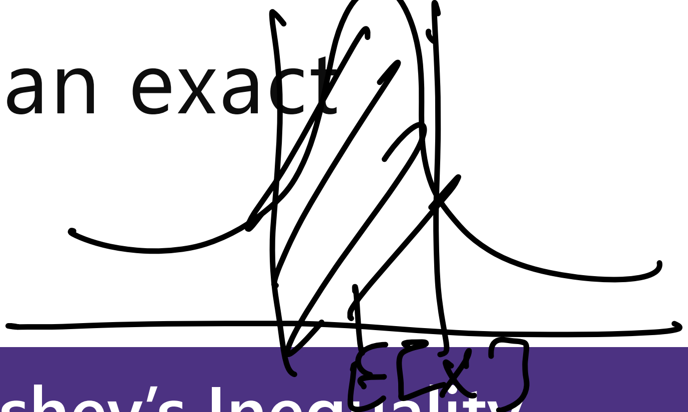
Variance of the sum of RVs?

Covariance

$$\mathbf{Cov}(\mathbf{X},\mathbf{Y}) = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])] = \mathbb{E}[\boldsymbol{XY}] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{Y}]$$

$$Var(X+Y) = Var(X) + 2 \cdot Cov(X,Y) + Var(Y)$$

What if we can't/don't need an exact probability calculation?



Tail bounds / concentration inequalities

Markov's Inequality

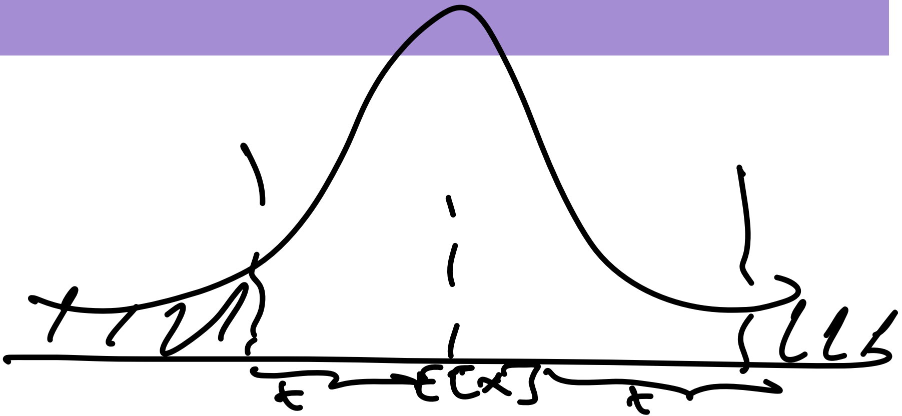
Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

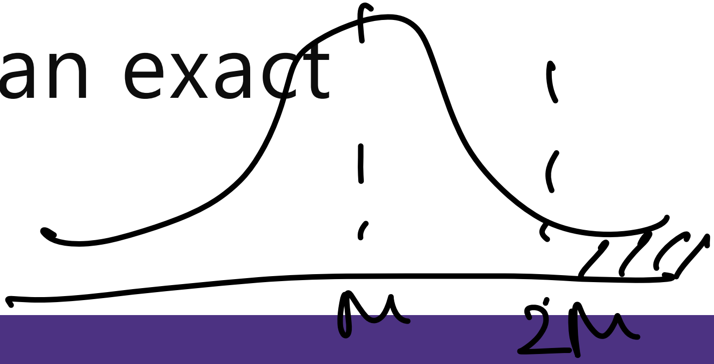
Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$



What if we can't/don't need an exact probability calculation?



Tail bounds / concentration inequalities

(Multiplicative) Chernoff Bound

Let $X_1, X_2, \ldots, X_n$ be *independent* Bernoulli random variables.
Let $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \le \delta \le 1$

$$\mathbb{P}(X \ge (1+\delta)\mu) \le \exp\left(-\frac{\delta^2 \mu}{3}\right) \text{ and } \mathbb{P}(X \le (1-\delta)\mu) \le \exp\left(-\frac{\delta^2 \mu}{2}\right)$$

Right Left

What if we can't/don't need an exact probability calculation?

Union Bound

For any events E, F

$$\mathbb{P}(\boldsymbol{E} \cup \boldsymbol{F}) \leq \mathbb{P}(\boldsymbol{E}) + \mathbb{P}(\boldsymbol{F})$$

What if we have data but not the model parameters?

Maximum Likelihood Estimation

Likelihood

The likelihood of i.i.d observations $x_1, \dots, x_n$ from a discrete distribution is

$$\mathcal{L}(x_1, \dots, x_n\,; \theta) = \prod_{i=1}^{n} \mathbb{P}(x_i; \theta)$$

$\mathbb{P}(\text{see data}\,; \theta)$

ln

$\frac{\partial}{\partial \theta}$ $\xrightarrow{\text{argmax}}$ $= 0$

solve for $\hat{\theta}$

What if we have data but not the model parameters?

Maximum Likelihood Estimation

An estimator $\hat{\theta}$ is "unbiased" if

$$\mathbb{E}\left[\hat{\theta}\right] = \theta$$

An estimator $\hat{\theta}$ is "consistent" if

$$\lim_{n \to \infty} \mathbb{E}\left[\hat{\theta}\right] = \theta$$

Some Extras! ☺

Naïve Bayes Spam Filtering

Bloom Filters

Polling

Algorithmic Privacy

Markov Chains (and Markov Chain Monte Carlo algorithms)

Multi-armed Bandits

Python + Latex!

Your Takeaways ☺

Better precise mathematical communication.

Strong problem-solving skills.

A solid foundation to learn more about algorithms, machine learning, data structures, networks, and so on!

Your Takeaways ☺

Better intuition (but more often, knowledge that intuition often fails in the probabilistic world)

Know that people often use statistics as a tool to support their (dubious) claims. There are a lot of different tricks, but here's my one-stop method: if you don't completely understand what a number is and where it came from, then you can't be sure of what it means.

Just because something **can** be quantified doesn't mean we **should**.

Thank you for a wonderful quarter!