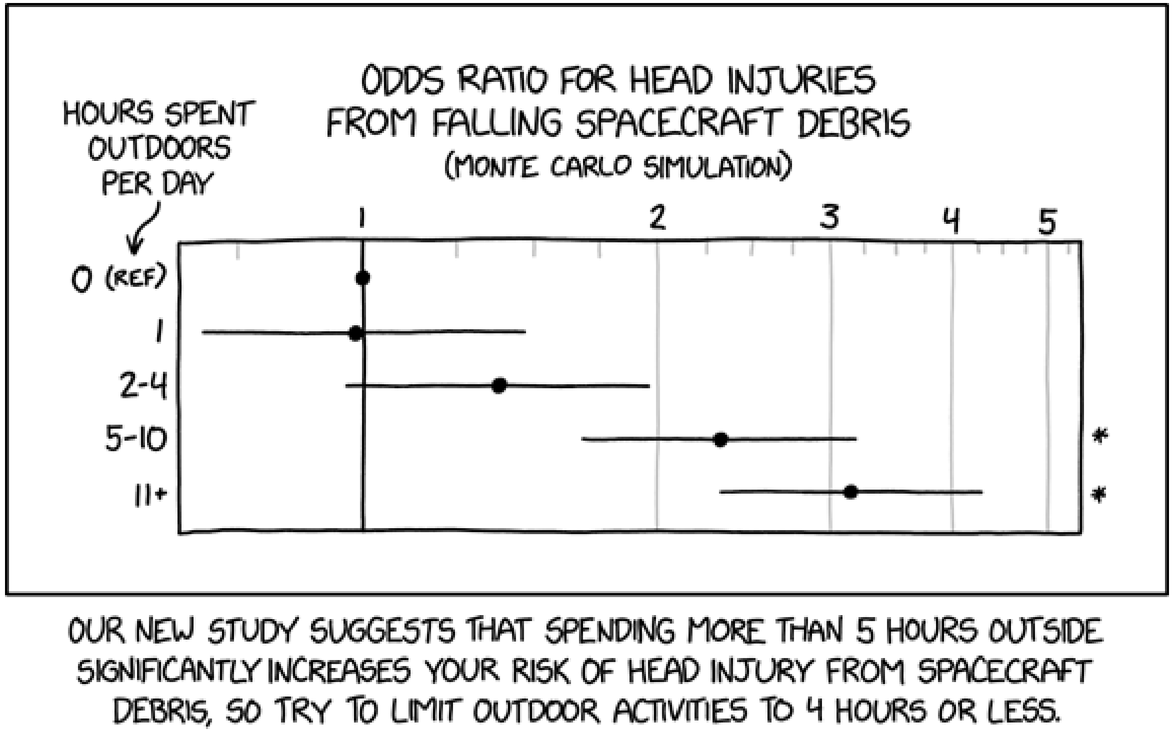



# Final Review + Victory Lap

CSE 312 26Su

Lecture 24

https://xkcd.com/2599/

# Announcements

Please fill out the course evaluation!

Show up to the exam on Thursday & Friday ☺ Staff are available in office hours and over Ed to help you out until then.

My OH today are cancelled.

# Final Logistics

Part 1: Thursday August 20 at 12pm in **HRC 155**

Part 2: Friday August 21 at 12pm in **HRC 155**

Emergency? Email Jolie ASAP + **CC Darin & Krisha**

Notes sheet, reference sheets provided.

60 minutes per part

Full details on website

# What's next?

CSE 446 Machine Learning (probability + linear algebra)

CSE 447 Natural Language Processing

CSE 473 Artificial Intelligence (Bayes nets, MDPs, Markov Models, RL)

CSE 421 Algorithms (design algorithms – uses some combinatorics)

Graduate level course in randomized algorithms with a few more prereqs

CSE 422 Modern Algorithms

CSE 426 Cryptography

CSE 427 Computational Biology

CSE 490Q Quantum Computing

# Victory Lap

# What Have We Done?

Well let’s look back…

# Once upon a time, there was: counting

Sum rule – choose from mutually exclusive (disjoint) sets

Product rule – decisions in a sequential process

# Once upon a time, there was: counting

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

Inclusion-exclusion: choose from possibly overlapping sets

# Once upon a time, there was: counting

Permutations – order distinct elements

## $k$-permutation

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

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

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

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.

Stars-and-bars – indistinguishable objects into distinguishable bins

# Counting sets leads to… discrete probability

Begin quantifying our uncertainty.

## Probability Space

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

$\Omega$ 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

Corollaries: complementation, monotonicity, inclusion-exclusion

- $\mathbb{P}(E) = 1 - \mathbb{P}(E)$ (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

Bayes uses a prior to calculate a posterior

# Sometimes we have extra information…

Law of Total Probability

Chain rule

**Chain Rule**

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

# Sometimes that extra is irrelevant…

Independence

## Pairwise Independence

Events $A_1, A_2, \ldots, A_n$ 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 $\boldsymbol{i, j}$ where $\boldsymbol{i \neq j}$

## Mutual Independence

Events $A_1, A_2, \ldots, A_n$ 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 $\{\boldsymbol{i_1}, \boldsymbol{i_2}, \ldots, \boldsymbol{i_k}\}$ of $\{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{n}\}$.

# Sometimes that extra is irrelevant…

Conditional independence

We say $A$ and $B$ are conditionally independent on $C$ if

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

# Summarize Numerical Info from $\Omega$

(Discrete) Random variables

> Range

> PMF

> CDF

> Expectation

# Summarize Numerical Info from $\Omega$

(Discrete) Random variables

> Linearity of Expectation

# Summarize Numerical Info from $\Omega$

(Discrete) Random variables

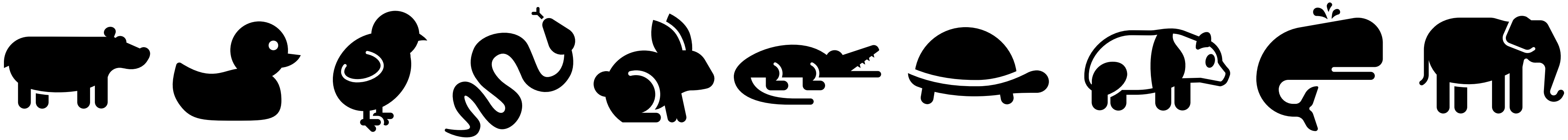
> Variance

## Variance

**The variance of a random variable $X$ is**

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

# Zoo!



## $X \sim \mathrm{Unif}(a, b)$

$$p_X(k) = \frac{1}{b-a+1}$$

$$\mathbb{E}[X] = \frac{a+b}{2}$$

$$\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

## $X \sim \mathrm{Ber}(p)$

$$p_X(0) = 1-p;$$
$$p_X(1) = p$$

$$\mathbb{E}[X] = p$$

$$\mathrm{Var}(X) = p(1-p)$$

## $X \sim \mathrm{Bin}(n, p)$

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbb{E}[X] = np$$

$$\mathrm{Var}(X) = np(1-p)$$

## $X \sim \mathrm{Geo}(p)$

$$p_X(k) = (1-p)^{k-1} p$$

$$\mathbb{E}[X] = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

## $X \sim \mathrm{NegBin}(r, p)$

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

$$\mathbb{E}[X] = \frac{r}{p}$$

$$\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

## $X \sim \mathrm{HypGeo}(N, K, n)$

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}[X] = n\frac{K}{N}$$

$$\mathrm{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$$

## $X \sim \mathrm{Poi}(\lambda)$

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\mathbb{E}[X] = \lambda$$

$$\mathrm{Var}(X) = \lambda$$

# Our sample spaces may be uncountable!

Continuous RVs

PDFs

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

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

# Our sample spaces may be uncountable!

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

# An Approximation…

## Central Limit Theorem

Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables, with mean $\mu$ and variance $\sigma^2$. 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 $\mathcal{N}(0, 1)$

# CLT E.g.

# 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

Law of Total Expectation

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

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

# 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 $t > 0$

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

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

Tail bounds / concentration inequalities

## (Multiplicative) Chernoff Bound

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

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

*LEFT TAIL* *RIGHT TAIL*

# 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 function $\mathcal{L}(E;\theta)$

$\prod_{i=1}^{n} \mathbb{P}\,(x_i;\theta)$ for discrete and $\prod_{i=1}^{n} f(x_i;\theta)$ for continuous

# 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

Distinct Elements

Bloom Filters

Polling

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!

# How (not) to lie with statistics

You now know a lot of the tools that people use to lie with statistics. (See also: INFO 270)

Some patterns to watch out for:

My smoke alarm is going off, please pay for my new house! (analogy from Matt Parker)

Make a model, find that an event that occurred had small probability/fails some statistical test, claim that the **only** explanation is something nefarious occurred.

Better response: could the model be wrong? Is this statistical test appropriate? Once in 100 year events do happen…about once in every hundred years, is this just the one?

# How (not) to lie with statistics

See a story about testing?

Remember from Bayes' Rule that you need three numbers to understand a test. (3 of prior, posterior, false positive rate, false negative rate).

Headlines usually give you one number, that often isn't even one of the ones you need for Bayes ("this test is less accurate than a coin flip!").

The article itself, if you're lucky, might give you one or two of the numbers for Bayes – don't forget the prior!

# 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!