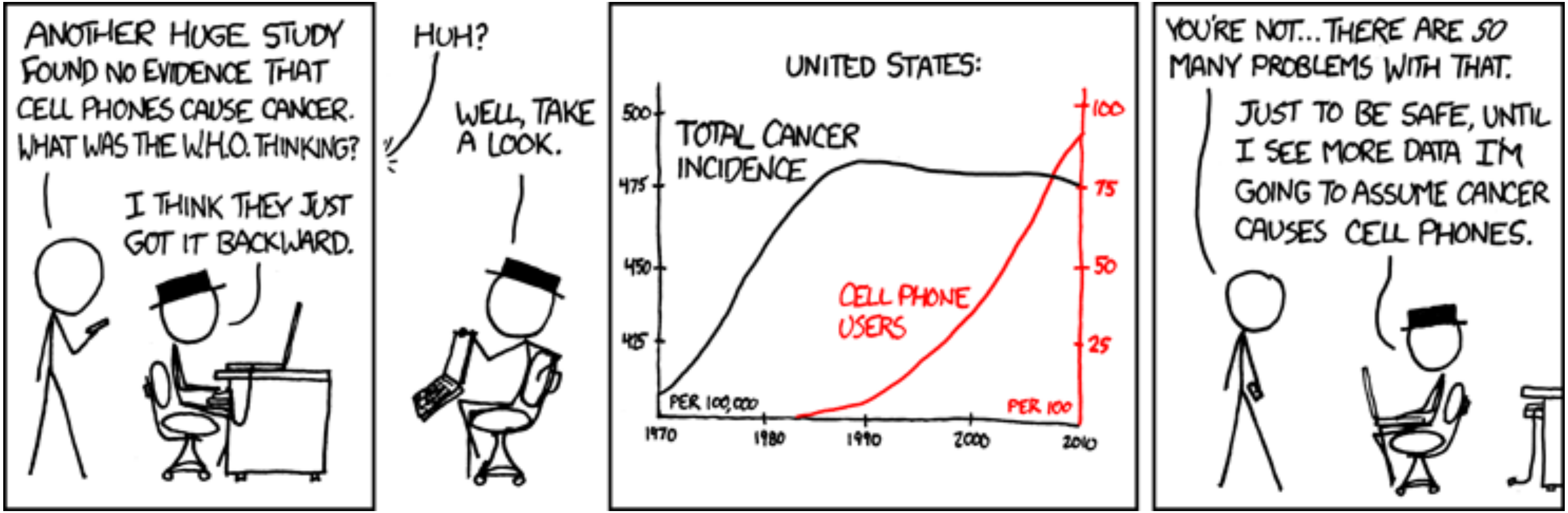


More LTE, Tail Bounds

CSE 312 Su23
Lecture 17

https://xkcd.com/925/

Law of Total Expectation

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

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

Let X, Y be discrete random variables, then

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y = y]\mathbb{P}(Y = y)$$

Similar in form to law of total probability, and the proof goes that way as well.

Elevator Rides

X people enter an elevator on the ground floor, where $X \sim Poi(10)$.

There are N floors above the ground floor, and each person is equally likely to get off at any one of the N floors (independent of where others get off).

What is the expected number of stops the elevator makes before all passengers get off?

Elevator Rides

X people enter an elevator on the ground floor, where $X \sim Poi(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any one of the N floors (independent of where others get off). What is the expected number of stops the elevator makes before all passengers get off?

Elevator Rides

X people enter an elevator on the ground floor, where $X \sim Poi(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any one of the N floors (independent of where others get off). What is the expected number of stops the elevator makes before all passengers get off?

Analogues for continuous

Everything we've seen has a continuous version.

There are "no surprises"– replace PMFs with PDFs and sums with integrals.

	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	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Covariance

Covariance

Measures how one random variable varies with a second.

Can be:

Positive (↑ in one var, ↑ in other)

Zero

Negative (↑ in one var, ↓ in other)

Covariance

We sometimes want to measure how "intertwined" X and Y are – how much knowing about one of them will affect the other.

If X turns out "big" how likely is it that Y will be "big"? how much do they "vary together"?

Covariance

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

Properties of Covariance

Covariance

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

1. $Cov(X,Y) = Cov(Y,X)$
2. $Var(X) = Cov(X,X)$
3. $Cov(aX+b,Y) = aCov(X,Y)$
4. $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i,Y_j)$

Properties of Covariance

Covariance

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

If $X \perp Y$, what do you expect $Cov(X, Y)$ to be?

Properties of Covariance

Covariance

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

If $X \perp Y$, what do you expect $Cov(X,Y)$ to be? $\mathbf{0}$.

(for X, Y independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).

Properties of Covariance

Does the opposite hold (i.e., does zero covariance imply independence?

$X = \{-1, 0, 1\}$ with equal probability. $Y = 1$ if X=0 and $\mathbf{0}$ otherwise.

$Y \backslash X$	-1	0	1	$p_Y(y)$
0	$1/3$	0	$1/3$	$2/3$
1	0	$1/3$	0	$1/3$
$p_X(x)$	$1/3$	$1/3$	$1/3$	

Properties of Covariance

Does the opposite hold (i.e., does zero covariance imply independence?

$X = \{-1, 0, 1\}$ with equal probability. $Y = 1$ if X=0 and 0 otherwise.

$Y \backslash X$	-1	0	1	$p_Y(y)$
0	$1/3$	0	$1/3$	$2/3$
1	0	$1/3$	0	$1/3$
$p_X(x)$	$1/3$	$1/3$	$1/3$	

$$\mathbb{E}[X] = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$
$$\mathbb{E}[Y] = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$$
$$\mathbb{E}[XY] = (-1 \cdot 0) \cdot \frac{1}{3} + (0 \cdot 1) \cdot \frac{1}{3} + (1 \cdot 0) \cdot \frac{1}{3} = 0$$

Properties of Covariance

Does the opposite hold (i.e., does zero covariance imply independence?

$X = \{-1, 0, 1\}$ with equal probability. $Y = 1$ if X=0 and 0 otherwise.

$Y \backslash X$	-1	0	1	$p_Y(y)$
0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
$p_X(x)$	1/3	1/3	1/3	

$$\mathbb{E}[X] = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$

$$\mathbb{E}[Y] = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$$

$$\mathbb{E}[XY] = (-1 \cdot 0) \cdot \frac{1}{3} + (0 \cdot 1) \cdot \frac{1}{3} + (1 \cdot 0) \cdot \frac{1}{3} = 0$$

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 \cdot \frac{1}{3} = 0$$

$$\mathbb{P}(Y = 1 | X = 0) = 1 \neq \frac{1}{3} = \mathbb{P}(Y = 1)$$

Properties of Covariance

Does the opposite hold (i.e., does zero covariance imply independence?

$X = \{-1, 0, 1\}$ with equal probability. $Y = 1$ if X=0 and 0 otherwise.

$Y \backslash X$	-1	0	1	$p_Y(y)$
0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
$p_X(x)$	1/3	1/3	1/3	

$$\mathbb{E}[X] = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$

$$\mathbb{E}[Y] = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$$

$$\mathbb{E}[XY] = (-1 \cdot 0) \cdot \frac{1}{3} + (0 \cdot 1) \cdot \frac{1}{3} + (1 \cdot 0) \cdot \frac{1}{3} = 0$$

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 \cdot \frac{1}{3} = 0$$

$\mathbb{P}(Y = 1 | X = 0) = 1 \neq \frac{1}{3} = \mathbb{P}(Y = 1)$ (i.e., X and Y are not independent)

So, zero covariance **does not imply** independence.

Variance of sum of RVs (finally!)

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

Proof:

$$Var(X+Y) = \mathrm{Cov}(\mathrm{X}+\mathrm{Y}, \mathrm{X}+\mathrm{Y})$$

$$= \mathrm{Cov}(\mathrm{X}, \mathrm{X}) + \mathrm{Cov}(\mathrm{X}, \mathrm{Y}) + \mathrm{Cov}(\mathrm{Y}, \mathrm{X}) + \mathrm{Cov}(\mathrm{Y}, \mathrm{Y})$$

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

This aligns with what we know, in the case that X and Y are independent.

Covariance

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\text{Var}(X + Y)$?

Before you calculate, make a prediction. What should it be?

Covariance

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

$$\mathrm{Var}(X+Y) = Var(X) + 2Cov(X,Y) + Var(Y)$$

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

Covariance

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\mathrm{Var}(X+Y)$?

$\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1 - 0^2 = 1$

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

$\mathbb{E}[XY] = \frac{1}{2} \cdot (-1 \cdot 1) + \frac{1}{2}(1 \cdot -1) = -1$

$\mathrm{Cov}(X,Y) = -1 - 0 \cdot 0 = -1.$

$\mathrm{Var}(X+Y) = 1 + 1 + 2 \cdot -1 = 0$

Covariance

The **magnitude** of covariance is affected by the units of the random variables involved (because $Cov(2X, Y) = 2Cov(X, Y)$).

The **sign** is more meaningful, but only tells us about the direction of the relationship.

We would like to understand the strength of the relationship!

Correlation

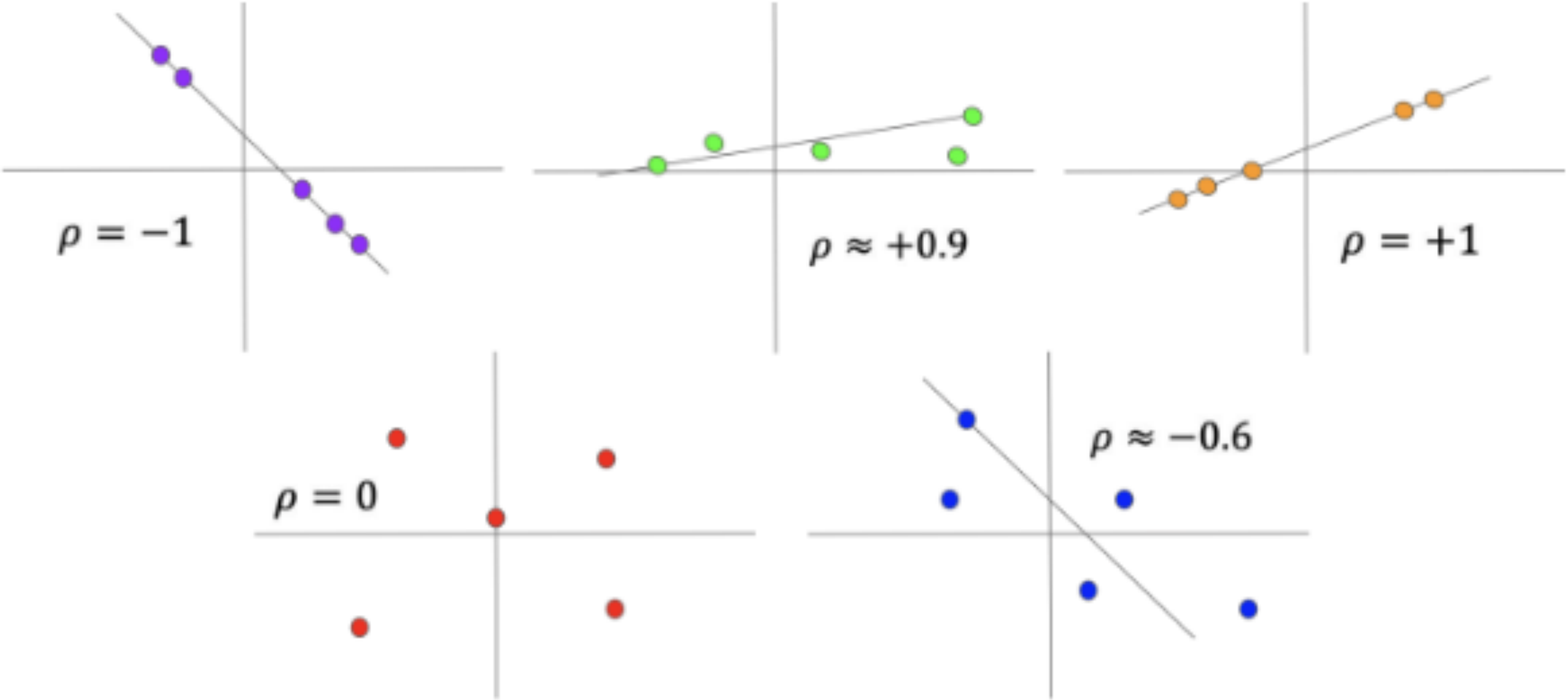
The (Pearson) correlation of two random variables X and Y is:

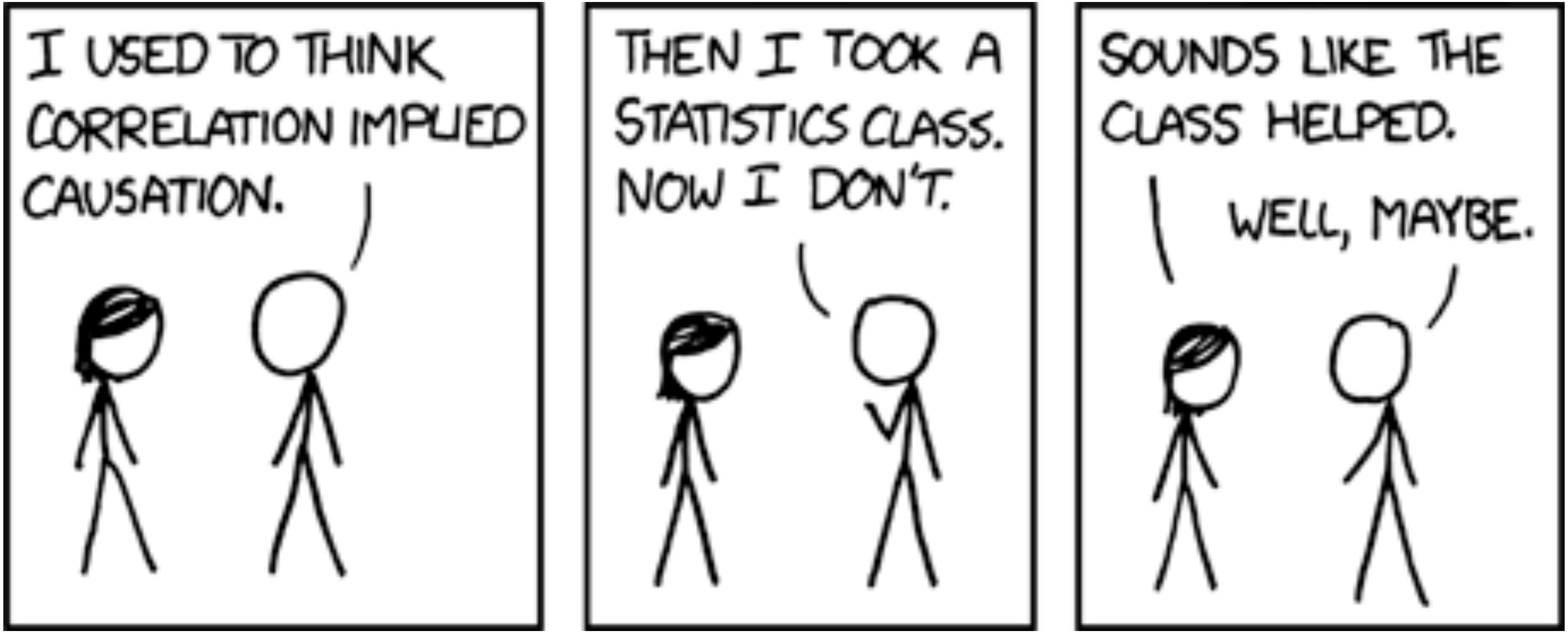
$$\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Where $-1 \leq \rho(X, Y) \leq 1$.

Normalized version of covariance.

Correlation





https://xkcd.com/552/

Bayesian Networks

So far…

We've looked closely at how to represent probabilistic models with multiple random variables.

Generally, we want to perform **inference**: i.e., we want to update our belief about one random variable, conditioned on new information on others.

(This was our motivation for this topic!)

Medical modelling

For e.g., WebMD's symptom checker is a probabilistic model with symptoms, risk factors and diseases.

Ideally, how would we specify the relationship between all these random variables?

A full joint distribution table will let us answer any question about their relation.

Medical modelling

How feasible is a full joint distribution table?

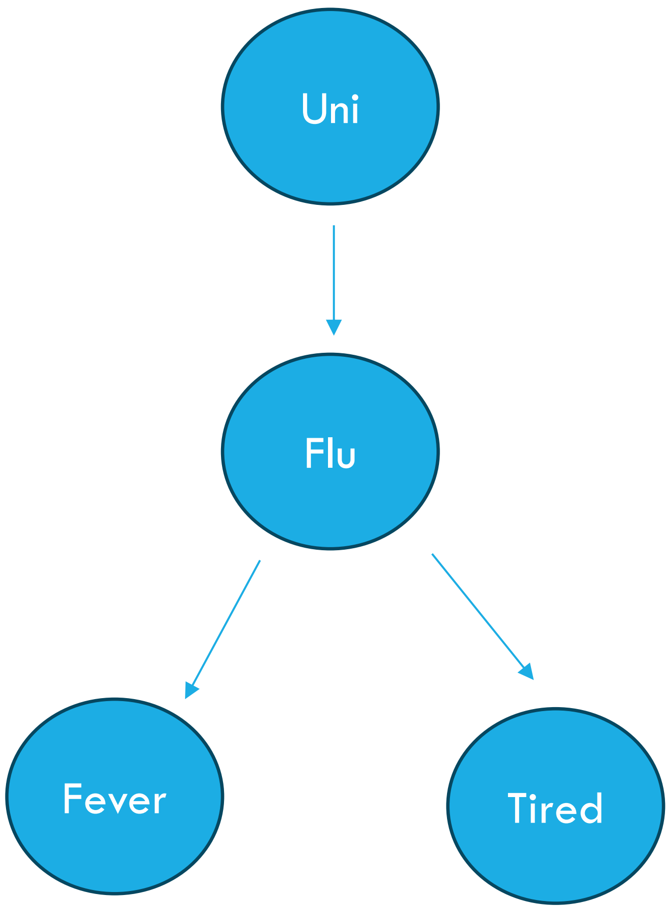
How many probabilities would we have to assign for $N = 3$ binary random variables? 2^3

What if $N = 100$? $2^{100} > 10^{30}$ (there are $\sim 10^{70}$ atoms in the universe)

A better way

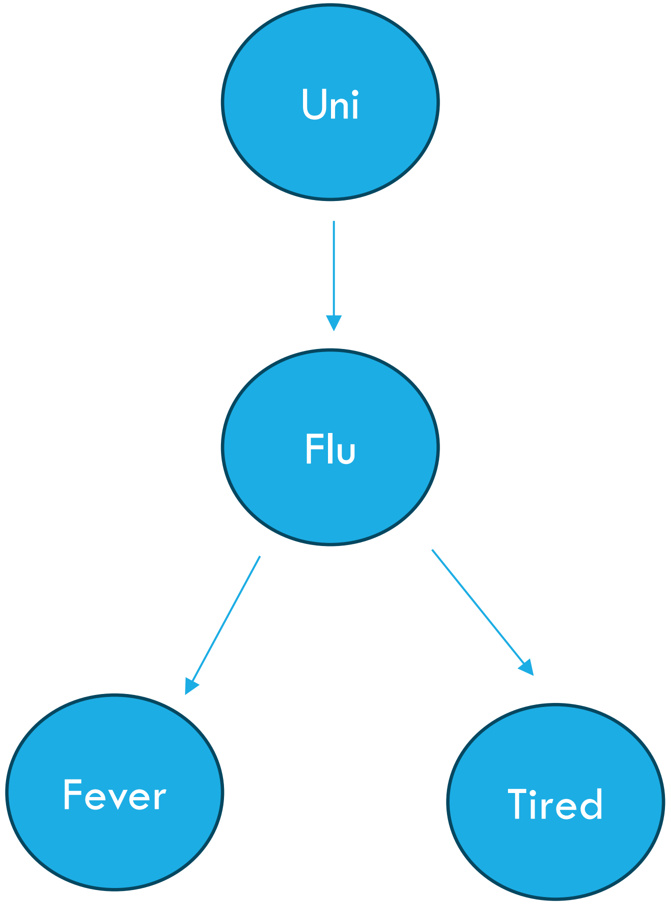
Our task is simpler if we know the "generative" process that creates a joint assignment.

For diseases, there is a directed flow of influence from demographics -> conditions -> symptoms.



Bayesian Networks

In a Bayes Net, an arrow from one random variable X to another Y represents our assumption that X directly influences the likelihood of Y – X is a **parent** of Y.



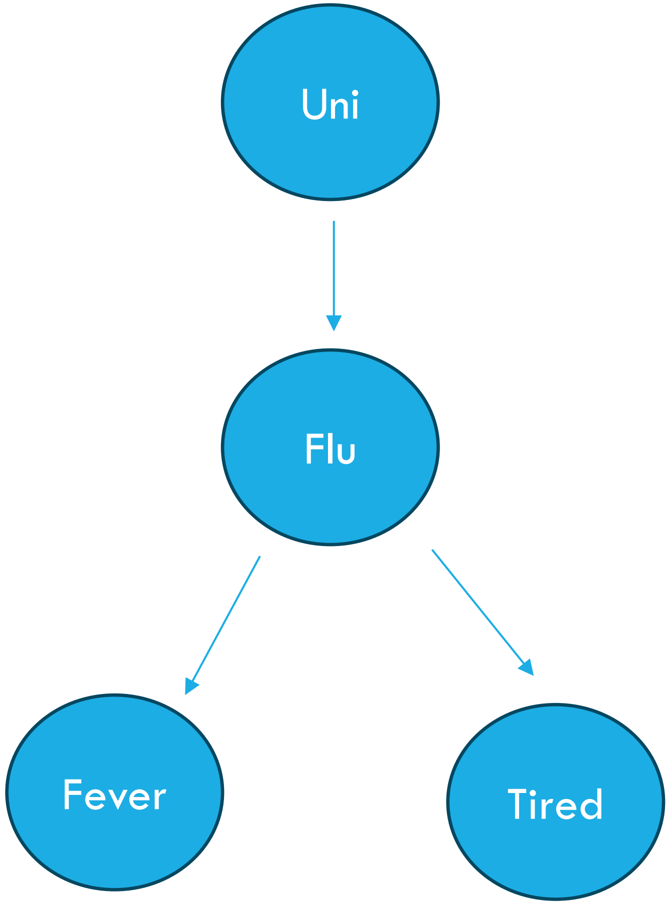
Bayesian Networks

A Bayes Net provides the probability of each random variable conditioned on knowing the value of all their parents.

With much less space, we can calculate:

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_i \mathbb{P}(X_i = x_i | values\ of\ parents\ of\ X_i).$$

(Formally, we assume each node is conditionally independent of its non-descendents, given its parents)



Tail Bounds

Using the CLT

$$\mathbb{P}(p - 0.05 \le \bar{X} \le p + 0.05)$$

$$= \mathbb{P}\left(\frac{p-0.05-p}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{(p+0.05-p)}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$= \mathbb{P}\left(\frac{-0.05}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{0.05}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\approx \mathbb{P}\left(\frac{-0.05}{\sqrt{\frac{p(1-p)}{n}}} \le Z \le \frac{0.05}{\sqrt{\frac{p(1-p)}{n}}}\right) \text{ (CLT)}$$

$$= \mathbb{P}\left(\frac{-0.05\sqrt{n}}{\sqrt{p(1-p)}} \le Z \le \frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

What’s a Tail Bound?

When we were finding our margin of error, we didn’t need an exact calculation of the probability.

We needed an inequality: the probability of being outside the margin of error was **at most** 2%.

A tail bound (or concentration inequality) is a statement that bounds the probability in the “tails” of the distribution (says there’s very little probability far from the center) or (equivalently) says that the probability is concentrated near the expectation.

Our First bound

Two statements are equivalent.
Left form is often easier to use.
Right form is more intuitive.

Markov's Inequality

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

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

Markov's Inequality

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

$$\mathbb{P}(X \geq k\mathbb{E}[X]) \leq \frac{1}{k}$$

To apply this bound you only need to know:

1. it's non-negative
2. Its expectation.

Proof

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

Markov's Inequality

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

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

Proof

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

$$= \sum_{x:x\geq t} x \cdot \mathbb{P}(X = x) + \sum_{x:x<t} x \cdot \mathbb{P}(X = x)$$

Markov's Inequality

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

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

Proof

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

$$= \sum_{x:x \geq t} x \cdot \mathbb{P}(X = x) + \sum_{x:x<t} x \cdot \mathbb{P}(X = x)$$

$$\geq \sum_{x:x \geq t} x \cdot \mathbb{P}(X = x) + 0$$

$\boxed{x \geq 0 \text{ whenever } \mathbb{P}(X = x) > 0}$

Markov's Inequality

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

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

Proof

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

$$= \sum_{x:x\geq t} x \cdot \mathbb{P}(X = x) + \sum_{x:x<t} x \cdot \mathbb{P}(X = x)$$

$$\geq \sum_{x:x\geq t} x \cdot \mathbb{P}(X = x) + 0$$

$$\geq \sum_{x:x\geq t} t \cdot \mathbb{P}(X = x)$$

$x \geq 0$ whenever $\mathbb{P}(X = x) > 0$

Markov's Inequality

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

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

Proof

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

$$= \sum_{x:x\geq t} x \cdot \mathbb{P}(X = x) + \sum_{x:x<t} x \cdot \mathbb{P}(X = x)$$

$$\geq \sum_{x:x\geq t} x \cdot \mathbb{P}(X = x) + 0$$

$x \geq 0$ whenever $\mathbb{P}(X = x) > 0$

$$\geq \sum_{x:x\geq t} t \cdot \mathbb{P}(X = x)$$

$$= t \cdot \sum_{x:x\geq t} \mathbb{P}(X = x)$$

$$= t \cdot \mathbb{P}(X \geq t)$$

$$\boxed{\mathbb{E}[X] \geq t \cdot \mathbb{P}(X \geq t)}$$

Markov's Inequality

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

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

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

Markov's Inequality

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

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

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

$$\mathbb{P}(X \geq 12) \leq \frac{\mathbb{E}[X]}{12} = \frac{6}{12} = \frac{1}{2}.$$

Exact probability?

$$1 - \mathbb{P}(X < 12) \approx 1 - 0.865 = .135$$

Markov's Inequality

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

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

A Second Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

Markov's Inequality

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

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

A Second Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}[X]}{75} = \frac{25}{75} = \frac{1}{3}$$

Markov's Inequality

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

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

Useless Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

Markov's Inequality

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

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

Useless Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

$$\mathbb{P}(X \geq 20) \leq \frac{\mathbb{E}[X]}{20} = \frac{25}{20} = 1.25$$

Well, that's…true. Technically.

But without more information we couldn't hope to do much better. What if every page gives exactly 25 ads? Then the probability really is 1.

So…what do we do?

A better inequality!

We’re trying to bound the tails of the distribution.

What parameter of a random variable describes the tails?

The variance!

Chebyshev's Inequality

Two statements are equivalent.
Left form is often easier to use.
Right form is more intuitive.

Chebyshev's Inequality

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

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

Chebyshev's Inequality

Let X be a random variable. For any $\mathrm{k} > 0$

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq k\sqrt{\mathrm{Var}(X)}\right) \leq \frac{1}{k^2}$$

Proof of Chebyshev

Chebyshev's Inequality

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

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

Let $Z = X - \mathbb{E}[X]$

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

Proof of Chebyshev

Chebyshev's Inequality

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

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

Let $Z = X - \mathbb{E}[X]$

Markov's Inequality

$\mathbb{E}[Z] = 0$

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

Inequalities are equivalent (square each side).

Z is just X shifted. Variance is unchanged.

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

Chebyshev's Inequality

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

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

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

$$\mathbb{P}(X \geq 12) \leq \mathbb{P}(|X-6| \geq 6) \leq \frac{\frac{5/6}{1/36}}{6^2} = \frac{5}{6}$$

Not any better than Markov ☹

Chebyshev's Inequality

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

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

Example with geometric RV

Let X be a geometric rv with parameter p

Bound the probability that $X \geq \frac{2}{p}$

Chebyshev's Inequality

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

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

Example with geometric RV

Let X be a geometric rv with parameter p

Bound the probability that $X \geq \frac{2}{p}$

$$\mathbb{P}(X \geq 2/p) \leq \mathbb{P}(|X - 1/p| \geq 1/p) \leq \frac{\frac{1-p}{p^2}}{1/p^2} = 1 - p$$

Markov gives:

$$\mathbb{P}\left(X \geq \frac{2}{p}\right) = \frac{\mathbb{E}[X]}{2/p} = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2}.$$

For large p, Chebyshev is better.

Chebyshev's Inequality

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

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

Better Example

Suppose the average number of ads you see on a website is 25. And the variance of the number of ads is 16. Give an upper bound on the probability of seeing a website with 30 or more ads.

Better Example

Suppose the average number of ads you see on a website is 25. And the variance of the number of ads is 16. Give an upper bound on the probability of seeing a website with 30 or more ads.

$$\mathbb{P}(X \geq 30) \leq \mathbb{P}(|X - 25| \geq 5) \leq \frac{16}{25}$$

Chebyshev's – Repeated Experiments

How many coin flips (each head with probability p) are needed until you get n heads.

Let X be the number necessary. What is probability $X \geq 2n/p$?

Markov

Chebyshev

Chebyshev's – Repeated Experiments

How many coin flips (each head with probability p) are needed until you get n heads.

Let X be the number necessary. What is probability $X \geq 2n/p$?

Markov $\quad \mathbb{P}\left(X \geq \frac{2n}{p}\right) \leq \frac{n/p}{2n/p} = \frac{1}{2}$

Chebyshev $\quad \mathbb{P}\left(X \geq \frac{2n}{p}\right) \leq \mathbb{P}\left(\left|X - \frac{n}{p}\right| \geq \frac{n}{p}\right) \leq \frac{\mathrm{Var}(X)}{n^2/p^2} = \frac{n(1-p)/p^2}{n^2/p^2} = \frac{1-p}{n}$

Tail Bounds – Takeaways

Useful when an experiment is complicated and you just need the probability to be small (you don't need the exact value).

Choosing a minimum n for a poll – don't need exact probability of failure, just to make sure it's small.

Designing probabilistic algorithms – just need a guarantee that they'll be extremely accurate

Learning more about the situation (e.g. learning variance instead of just mean) usually lets you get more accurate bounds.

Next time: more assumptions to get better bounds.