More Joint Distributions, Conditional Distributions, Tail Bounds

CSE 312 24Su

Lecture 17

Logistics

> Catch-up/breather lecture on July 31st

> August 2nd and 5th will be MLEs

> After that, we'll be going over some applications which will also allow us to have some more review and practice with the main course content

Today

> One more joint distribution example

> Covariance

> Conditional Distributions

applying things we know about conditioning to random variables (and continuous)
law of total expectation

> Tail Bounds

Markov's Inequality

Chebyshev's Ineqality

Chernoff Bound

(union bound)

We have two **discrete** random variables X and Y
(that may or may not be independent)

Joint **Support/Range** - $\Omega_{X,Y}$

$\Omega_{X,Y} = \{(a, b) : p_{X,Y}(a, b) > 0\} \subseteq \Omega_X \times \Omega_Y$

Joint **PMF** - $p_{X,Y}(a, b)$

$p_{X,Y}(a, b) = \mathbb{P}(X \le a, Y \le b)$
*defined for **all*** $(a, b) \in \mathbb{R} \times \mathbb{R}$

Joint **CDF** - $\mathrm{F}_{X,Y}(a, b)$

$F_{X,Y}(a, b) = \mathbb{P}(X \le a, Y \le b)$
*defined for **all*** $(a, b) \in \mathbb{R} \times \mathbb{R}$

Normalization Property:
$\sum_{(a,b)\in\Omega_{X,Y}} p_{X,Y}(a, b) = 1$

Joint **Independence**

> $p_{X,Y}(a, b) = p_X(a) \cdot p_X(b)$ for all $(a, b) \in \Omega_{X,Y}$
> $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Joint **Expectation**

$\mathbb{E}[g(X, Y)] =$
$\sum_{(a,b)\in\Omega_{X,Y}} g(a, b)\, p_{X,Y}(a, b)$

Marginal **PMF** - $p_X(x)$, $p_Y(y)$

$p_X(x) = \sum_{y\in\Omega_Y} p_{X,Y}(x, y)$
$p_Y(y) = \sum_{x\in\Omega_X} p_{X,Y}(x, y)$

Notice we're summing over what the other RV can be

We have two **continuous** random variables X and Y
(that may or may not be independent)

Joint **Support/Range** - $\Omega_{X,Y}$

$\Omega_{X,Y} = \{(a,b) : f_{X,Y}(a,b) > 0\} \subseteq \Omega_X \times \Omega_Y$

Joint **PDF** - $f_{X,Y}(a,b)$

$f_{X,Y}(a,b)$ *defined for* ***all*** $(a,b) \in \mathbb{R} \times \mathbb{R}$

Normalization Property:

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx\, dy = 1$

Joint **CDF** - $F_{X,Y}(a,b)$

$F_{X,Y}(a,b) = \mathbb{P}(X \leq a, Y \leq b)$
defined for ***all*** $(a,b) \in \mathbb{R} \times \mathbb{R}$

Joint **Independence**

> $f_{X,Y}(a,b) = f_X(a) \cdot f_X(b)$ for all $(a,b) \in \Omega_{X,Y}$
> $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Joint **Expectation**

$\mathbb{E}[g(X,Y)] =$
$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y) dx\, dy$

Marginal **PDF** - $f_X(x)$, $f_Y(y)$

$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$
$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$

Notice we're integrating (summing) over what the other RV can be

Joint Probabilities

To find probability of X and Y being in ranges, we use the joint distribution:

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are discrete....

$$\mathbb{P}(a \leq X \leq b \cap c \leq Y \leq d) = \sum_{x \in a \leq X \leq b} \sum_{y \in c \leq Y \leq d} p_{X,Y}(x, y)$$

sum over the joint PMF for all pairs of x and y that fall in this range

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are continuous...

$$\mathbb{P}(a \leq X \leq b \cap c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dy\, dx$$

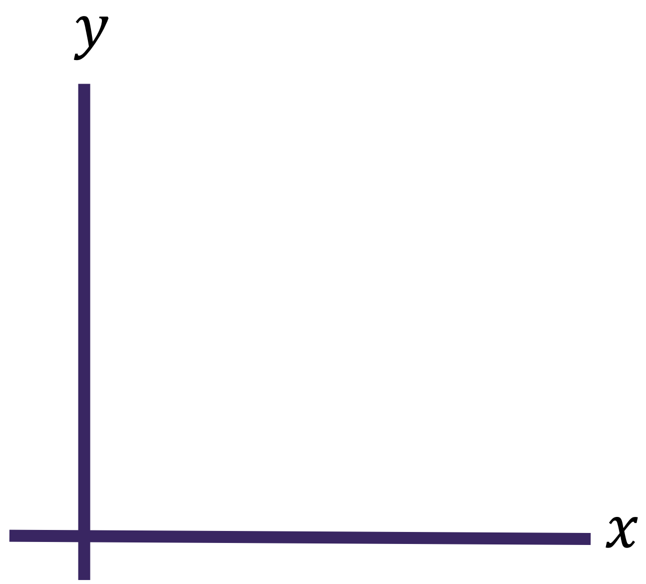
integrate over the joint PDF for all pairs of x and y that fall in this range

Example: Continuous Servers

The time until server 1 crashes is $X \sim \text{Exp}(u)$, and the time until server 2 crashes is $Y \sim \text{Exp}(v)$. Both servers are independent of each other.

What is the probability server 1 crashes before server 2?

$\mathbb{P}(X < Y) =$



Fill out the poll everywhere:
pollev.com/cse312

Example: Continuous Servers

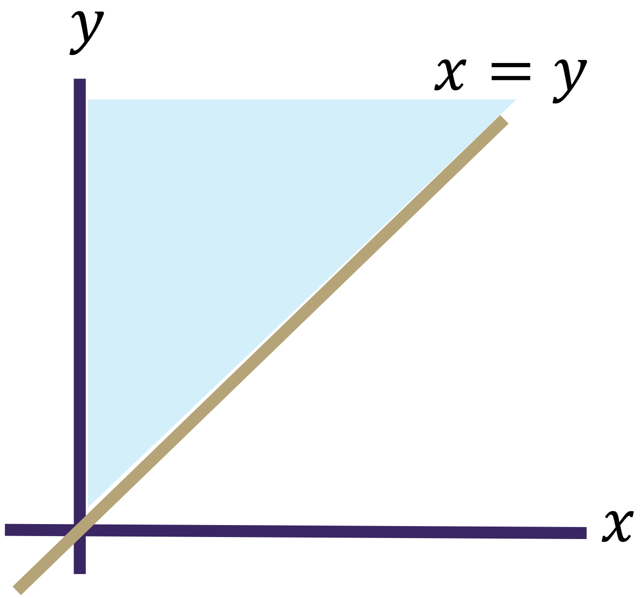
The time until server 1 crashes is $X \sim \text{Exp}(u)$, and the time until server 2 crashes is $Y \sim \text{Exp}(v)$. Both servers are independent of each other.

What is the probability server 1 crashes before server 2?

$\mathbb{P}(X < Y)$

$= \int_0^\infty \int_x^\infty f_{X,Y}(x, y)\, dy\, dx$

$= \int_0^\infty \int_x^\infty f_X(x) f_Y(y)\, dy\, dx$ *by independence*



Discrete vs. Continuous Joint Distributions

	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	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
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

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

$\text{Cov}(X, Y)$ measure the dependence between X and Y

> Covariance is **positive** -> they are *positively correlated*

If X increases, Y tends to also increase

> Covariance is **negative** -> they are *negatively correlated*

If X increases, Y tends to decrease

Covariance

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

Properties of Covariance

Covariance

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

If X and Y are independent, what is $\mathrm{Cov}(X, Y)$? $\mathbf{0}$.
because $\mathbb{E}[\boldsymbol{XY}] = \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{Y}]$ *if independent*

⚠️ This isn't always true the other way! The are some dependent random variables X **and** $\boldsymbol{Y}$ **where** $\mathrm{Cov}(X, Y) = 0$

Properties of Covariance

Covariance

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

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

> $\mathrm{Cov(X, X)} = \mathrm{Var}(X)$
because when you plug in X, X above, we get $\mathbb{E}[(X - \mathbb{E}[X])^2]$ which is the variance

> $\mathrm{Cov(aX + b, Y)} = \mathrm{a} \cdot \mathrm{Cov(X, Y)}$
linearity of expectation

> $\mathrm{Cov}(\sum_{i=1}^{n} X_i , \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$

Covariance *used for variance of a sum*

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

^ *we do* ***not*** *need* X and Y to be independent to use this formula!

Proof:

$$\begin{aligned}
&\mathrm{Var}(X+Y)\\
&=\mathrm{Cov}(X+Y,X+Y)\\
&=\mathrm{Cov}(X,X)+\mathrm{Cov}(X,Y)+\mathrm{Cov}(Y,X)+\mathrm{Cov}(Y,Y)\\
&=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)
\end{aligned}$$

Covariance *used for variance of a sum*

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

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

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

Fill out the poll everywhere:
pollev.com/cse312

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

Covariance *used for variance of a sum*

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

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

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

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

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

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

Covariance

The **magnitude** of covariance is affected by the units of the random variables involved because $\mathrm{Cov}(2X, Y) = 2\mathrm{Cov}(X, Y)$, so we can't really compare and it's not very helpful

is covariance big because of the units or because of a very strong relationship?

The **sign** of the covariance (positive or negative) is helpful but it only tells us the direction.

We want to understand the ***strength of the relationship***!

Pearson Correlation *(normalized covariance!)*

To understand the strength, we normalize the covariance!

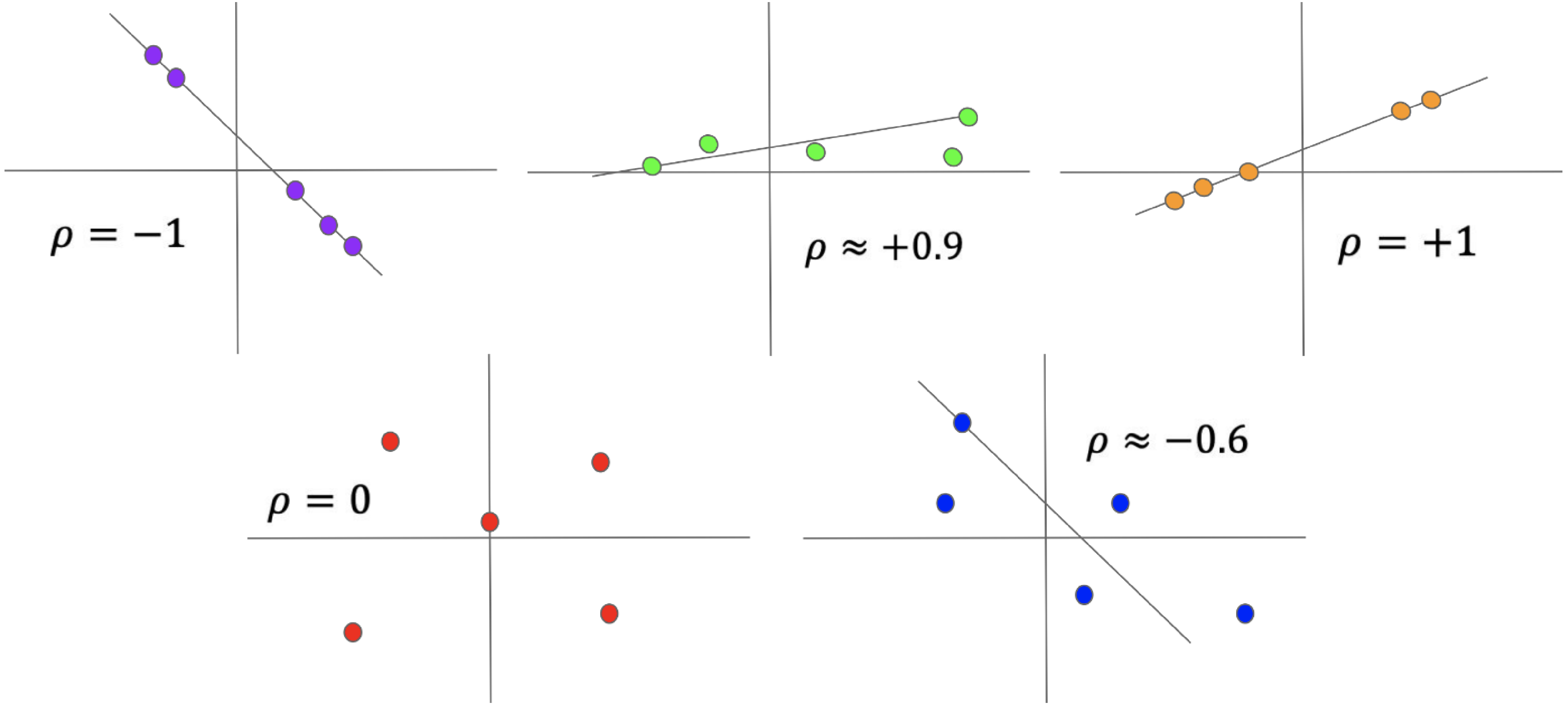
Pearson correlation: $\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$

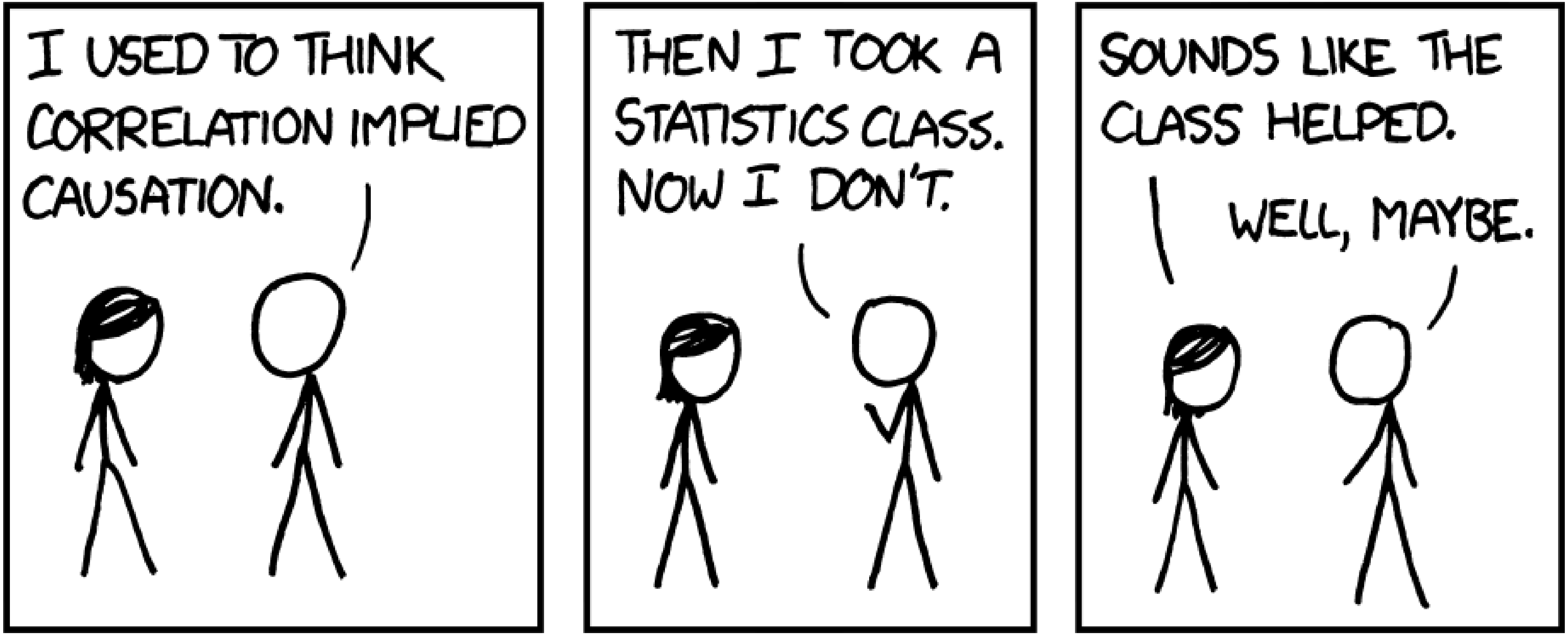
> divide the covariance by the product of the standard deviation of X and the standard deviation of Y

> a value in the range $-1 \le \rho(X,Y) \le 1$

-1 means STRONG **negative** correlation, $+1$ means STRONG **positive** correlation

Pearson Correlation *(normalized covariance!)*





https://xkcd.com/552/

Some Miscellaneous Topics…

Extending things we've learned about before to random variables (and the continuous case)

Conditional Distributions

More formulae…but really, explicitly shifting our
knowledge of conditional probability to random variables

Conditional PMFs/PDFs

Waaaaaay back, we said conditioning on an event creates a new probability space, with all the laws holding.

When we look at $\boldsymbol{X|A}$ where $\boldsymbol{A}$ is some event, we're redefining a random variable $\boldsymbol{X}$ inside that restricted probability space conditioning on $\boldsymbol{A}$

Conditional PMF: $p_{X|Y}(a|b) = \mathbb{P}(X = a|Y = b) = \frac{p_{X,Y}(a,b)}{p_Y(b)} = \frac{p_{Y|X}(b,a)\, p_X(a)}{p_Y(b)}$

Conditional PDF: $f_{X|Y}(a|b) = \frac{f_{X,Y}(a,b)}{f_Y(b)} = \frac{f_{Y|X}(b,a)\, f_X(a)}{f_Y(b)}$

Conditional Expectation

Waaaaaay back when, we said conditioning on an event creates a new probability space, with all the laws holding.

So, we can define things like "**conditional expectations**" which is the expectation of a random variable in that new probability space.

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

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

or $\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} k \cdot f_{X|Y}(k, y)\, dk$ **if continuous**

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

or if continuous, $\mathbb{E}[X] = \int_{-\infty}^{\infty} k \cdot f_X(k)\, dx$

Conditional Expectation

All your favorite theorems are still true.

For example, **linearity of expectation** still holds

$$\mathbb{E}[(aX + bY + c) \mid A] = a\mathbb{E}[X|A] + b\mathbb{E}[Y|A] + c$$

Law of Total Expectation (LTE)

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

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

Let $\boldsymbol{X}, \boldsymbol{Y}$ be discrete RVs, then,

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

$\boldsymbol{X}, \boldsymbol{Y}$ are continuous RVs, then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y)$$

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

LTE Example: Exponential Coins

You flip 2 (independent, fair coins). X is the number of heads. Then, the random variable Y follows the distribution $\mathrm{Exp}(X+1)$.

What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

LTE Example: Exponential Coins

You flip 2 (independent, fair coins). X is the number of heads. Then, the random variable Y follows the distribution $\mathrm{Exp}(X+1)$.

What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

$$\mathbb{E}[Y]$$

$$= \mathbb{E}[Y|X=0]\mathbb{P}(X=0) + \mathbb{E}[Y|X=1]\mathbb{P}(X=1) + \mathbb{E}[Y|X=2]\mathbb{P}(X=2)$$

LTE Example: Exponential Coins

You flip 2 (independent, fair coins). X is the number of heads. Then, the random variable Y follows the distribution $\mathrm{Exp}(X+1)$.

What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

$$\mathbb{E}[Y]$$
$$= \mathbb{E}[Y|X=0]\mathbb{P}(X=0) + \mathbb{E}[Y|X=1]\mathbb{P}(X=1) + \mathbb{E}[Y|X=2]\mathbb{P}(X=2)$$
$$= \mathbb{E}[Y|X=0]\cdot\frac{1}{4} + \mathbb{E}[Y|X=1]\cdot\frac{1}{2} + \mathbb{E}[Y|X=2]\cdot\frac{1}{4}$$
$$= \frac{1}{0+1}\cdot\frac{1}{4} + \frac{1}{1+1}\cdot\frac{1}{2} + \frac{1}{2+1}\cdot\frac{1}{4} = \frac{7}{12}.$$

LTE Example: Elevator Rides

The number of people who enter an elevator on the ground floor is $X \sim \text{Poi}(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any of the N floors, independently of others. What is the **expected number of stops the elevator will make before discharging all the passengers**?

Y is the number of stops the elevator makes. What is $\mathbb{E}[Y]$?

Again, Y *depends on what the value of* X *is. So, use LTE, partitioning on* X.

LTE Example: Elevator Rides

The number of people who enter an elevator on the ground floor is $X \sim \mathrm{Poi}(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any of the N floors, independently of others. What is the **expected number of stops the elevator will make before discharging all the passengers**?

Y is the number of stops the elevator makes. What is $\mathbb{E}[Y]$?

Again, *Y depends on what the value of X is. So, use LTE, partitioning on X.*

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\,e^{-10}\frac{10^i}{i!}$$

Law of Total Probability

We've seen law of total probability before. We can use the notation we've learned to write LoTP for random variables:

Let $\boldsymbol{X}, \boldsymbol{Y}$ be discrete RVs, then,

$$p_X(x) = \sum_{y \in \Omega_Y} p_{X|Y}(x|y)\mathbb{P}(Y = y)$$

$\boldsymbol{X}, \boldsymbol{Y}$ are continuous RVs, then,

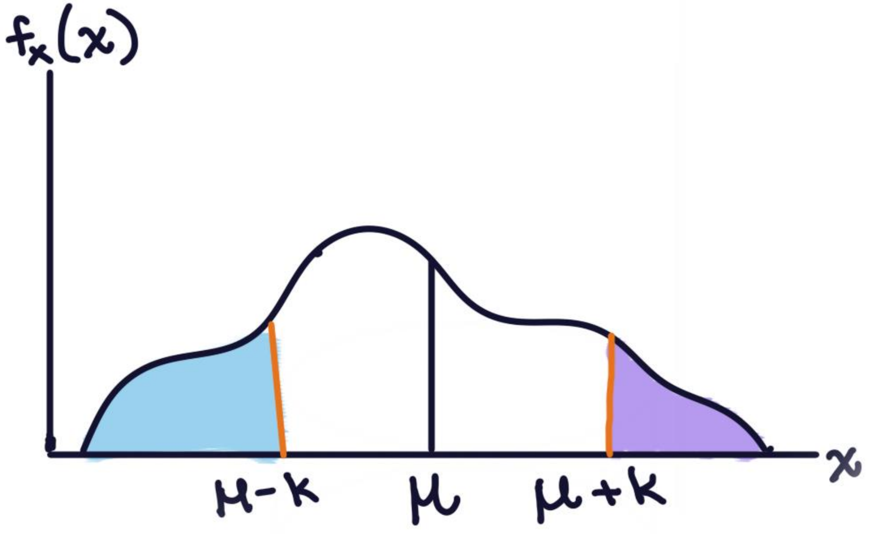
$$f_{X(x)} = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)$$

Tail Bounds

What's a Tail Bound?

The "tails" of a probability distribution are the extreme regions to the left or right of the expectation

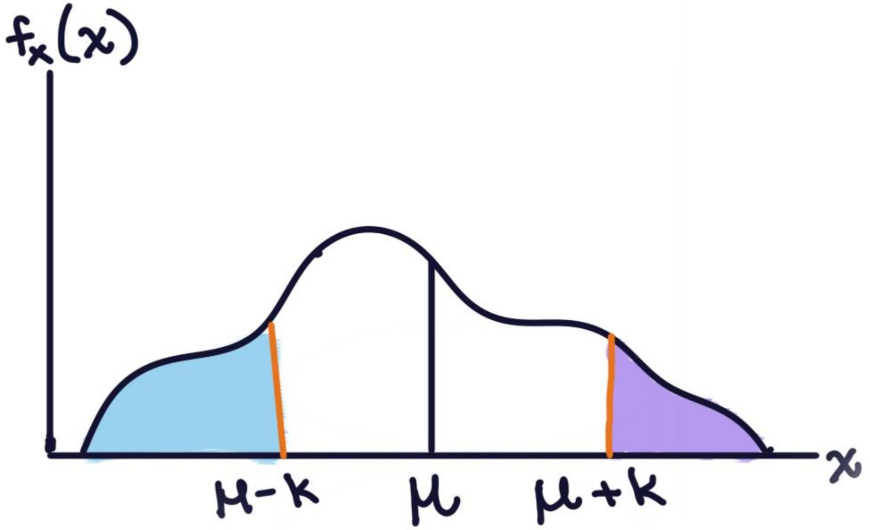
e.g., the shaded regions $X \leq \mu - k$ *and* $X \geq \mu + k$ *are "tails" of the distribution*



What's a Tail Bound?

The "tails" of a probability distribution are the extreme regions to the left or right of the expectation

e.g., the shaded regions $X \leq \mu + k$ *and* $X \geq \mu - k$ *are "tails" of the distribution*



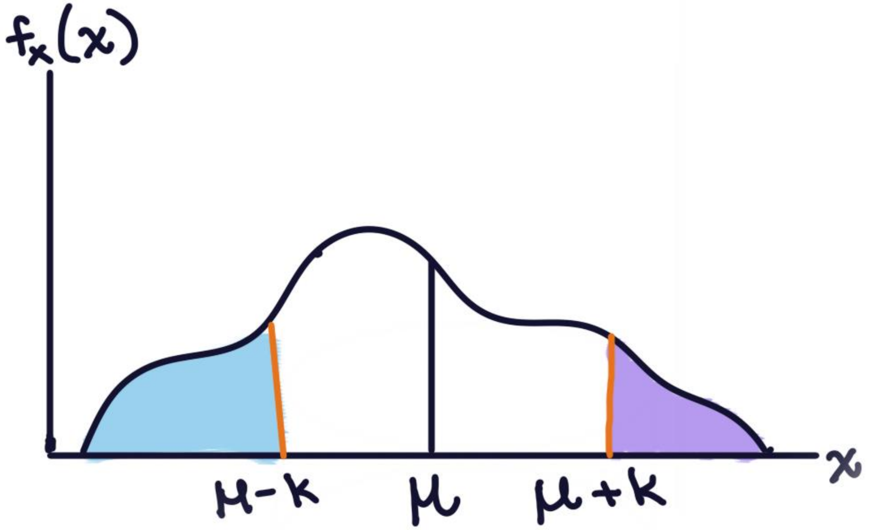
Often, we want to make some guarantees about the probability of being in a tail is (e.g., $\mathbb{P}(X \geq k) \leq ??$)

guarantees about the running time (the chance of being > 5sec is no more than __)

What's a Tail Bound?

The "tails" of a probability distribution are the extreme regions to the left or right of the expectation

e.g., the shaded regions $X \leq \mu + k$ *and* $X \geq \mu - k$ *are "tails" of the distribution*



Often, we want to make some guarantees about the probability of being in a tail is (e.g., $\mathbb{P}(X \geq k) \leq ??$)

guarantees about the running time (the chance of being > 5sec is no more than __)

A **tail bound** (or concentration inequality) is a statement that bounds the probability in the "tails" of the distribution (e.g., there's little probability far from the center) or (equivalently) the probability is concentrated near the expectation.

What's a Tail Bound?

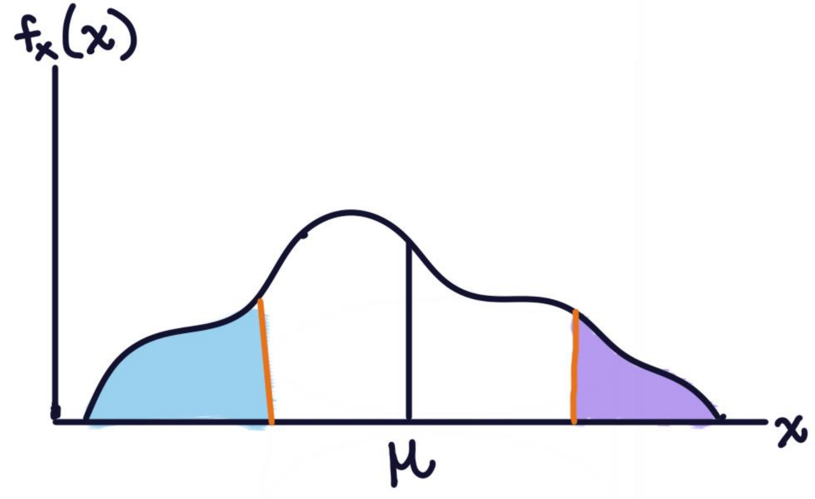
A **tail bound** (or concentration inequality) bounds the probability in the "tails" of the distribution. e.g., statements like $\mathbb{P}(X \geq 4) \leq 0.8$, $\mathbb{P}(X \geq 4) \leq 0.8$

What's a Tail Bound?

A **tail bound** (or concentration inequality) bounds the probability in the "tails" of the distribution. e.g., statements like $\mathbb{P}(X \geq 4) \leq 0.8$, $\mathbb{P}(X \geq 4) \leq 0.8$

We've seen this before! We can:

- Compute these probabilities exactly in some cases
- Approximate X as normal using CLT if X is the sum of a bunch of i.i.d random variables

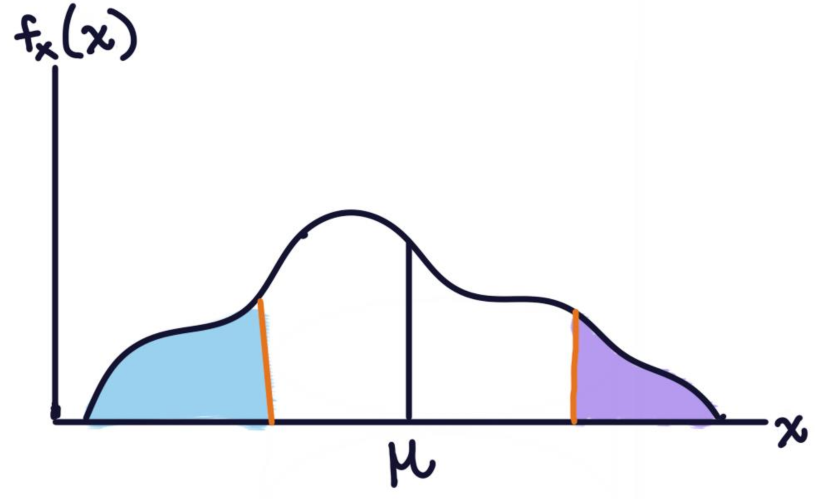


What’s a Tail Bound?

A **tail bound** (or concentration inequality) bounds the probability in the “tails” of the distribution. e.g., statements like $\mathbb{P}(X \geq 4) \leq 0.8$, $\mathbb{P}(X \geq 4) \leq 0.8$

We’ve seen this before! We can:

- Compute these probabilities exactly in some cases
- Approximate X as normal using CLT if X is the sum of a bunch of i.i.d random variables



But what if we barely know anything about X and it doesn't fit into the frameworks we've learned about? Can we still make some tail bound guarantees?

Tail Bounds

We're going to learn about 3 tail bounds that we can use when all we know about X is it's expected value and/or variance:

- Markov's Inequality
- Chebysev's Inequality
- Chernoff Bound

Markov's Inequality

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 $t > 0$

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

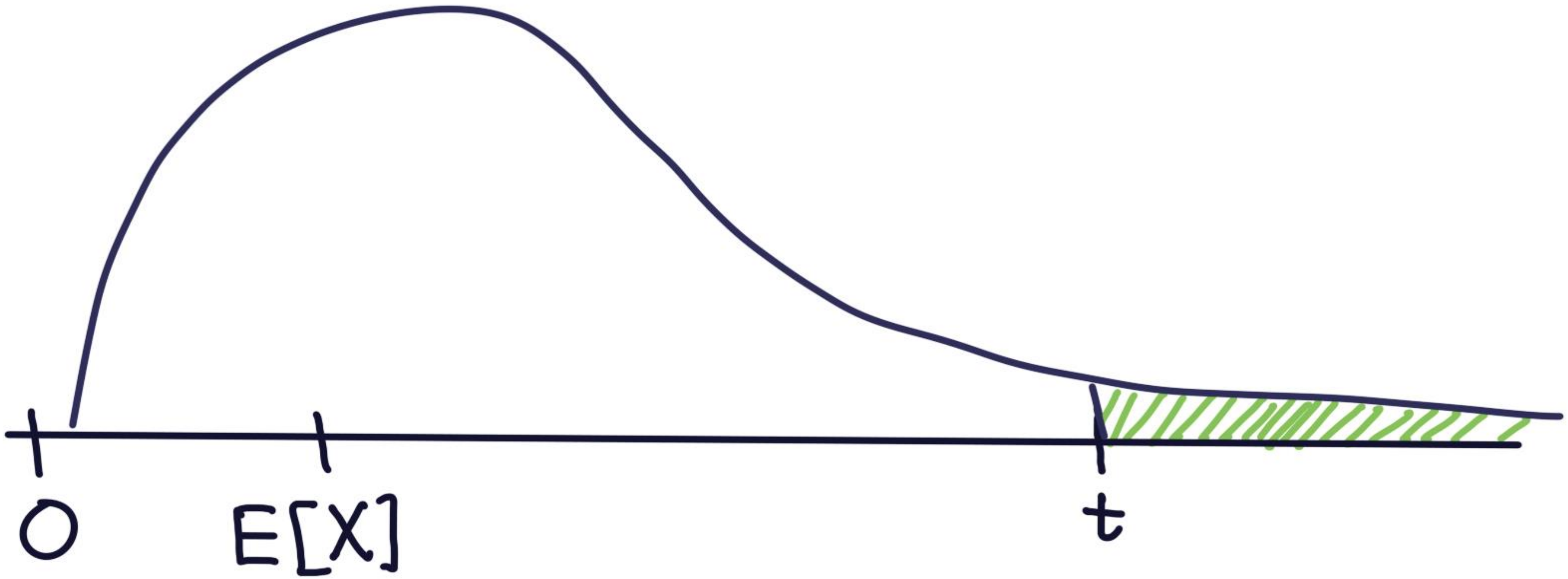
Markov's Inequality

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

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

Requirements:

1. X must be non-negative
2. We know the expectation of X


0
E[X]
t

Proof

$$\mathbb{E}[X] = \mathbb{E}[X|X < t]\mathbb{P}(X < t) + \mathbb{E}[X|X \geq t]\mathbb{P}(X \geq t)$$

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

$\mathbb{E}[X|X \geq t]\mathbb{P}(X \geq t) \geq 0$ if X is non-negative

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

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

Doing some algebra…we get exactly what's in Markov's inequality! →

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: Let's see how good this bound is…

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls. Bound the probability that $X \geq 12$.

$X \sim \mathrm{Geo}\left(\frac{1}{6}\right)$, so $\mathbb{E}[X] = 1/(\frac{1}{6}) = 6$

Applying Markov's Inequality…

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

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: Let's see how good this bound is…

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls. Bound the probability that $X \geq 12$.

$X \sim \text{Geo}\left(\frac{1}{6}\right)$, so $\mathbb{E}[X] = 1/(\frac{1}{6}) = 6$

Applying Markov's Inequality…

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

Exact probability?

$$1 - \mathbb{P}(X < 12) \approx 1 - 0.865 = 0.135$$

Markov's Inequality

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

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

Example: Ads

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 $t > 0$

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

Example: Ads

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 $t > 0$

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

Example: More Ads

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.

Fill out the poll everywhere:
pollev.com/cse312

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: More Ads

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!

Upper vs. Lower Bound

If we find something like $\mathbb{P}(A) \leq b$, we found an upper bound

This *highest/"uppermost"* value the probability of A could be is b

If we find something like $\mathbb{P}(A) \geq b$, we found a lower bound

This *lowest/smallest* value the probability of A could be is b