Confidence Intervals + Joint Distributions I

CSE 312 26Su

Lecture 16

Where Are We?

Last time:

- Central Limit Theorem

Today:

- Polling
- Joint Distributions
  - Joint Support
  - Joint PMF

Central Limit Theorem (Review)

"The sum of **any** independent random variables **approaches** a normal distribution. It becomes closer to normal/the approximation gets better as we sum more RVs together."

Central Limit Theorem

If $X_1, X_2, \dots, X_n$ are i.i.d. random variables, each with mean μ and variance σ^2

Let $Y_n = X_1 + X_2 + \cdots + X_n$

As $n \to \infty$, Y_n approaches a normal distribution $\mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$

(i.e., CDF of Y_n converges to the CDF of $\mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$)

When to use CLT

Use the CLT when:

- The random variable you're interested in is the sum of independent random variables.
- The random variable you're interested in does not have an easily accessible or easy to use pmf/pdf (or the question you're asking doesn't lend it self to easily using the pmf/pdf)
- You only need an approximate answer, and the sum is of at least a moderate number of random variables.

Polling

(Idealized) Polling

Suppose we want to conduct a poll to understand how many people support you in your run for SAC

Poll to determine the fraction p of the population that approve

- Call up a random sample of n people to ask their opinion
- Report the empirical fraction

Is this a good estimate?

How to choose n?

The Reverse Question

Polls are made by sampling n people from a population.

Often reported with "52% of likely voters would vote in favor of proposal if held today (margin of error +/- 3%)"

You are going to run your own poll. And you want a better "margin of error" – you want 2%

How many people do you need to poll?

Let's think about idealized polling – pretend we're really getting a uniformly random person.

Margin of Error

Wait…what’s a “margin of error”

The result of the poll is a random variable – it has a distribution.

You’d like to know something about its variance (Did you poll everyone in the entire country? Just 3 people? How much variance is there in the poll?)

A “margin of error” is an intuitive measurement of the variance of the poll. “If I performed this poll repeatedly, 95% of the time, we’re within true +/- the margin of error.”

Our Goal

Set a target – I want my margin of error to be 2%. That is, at least 95% of the time, your poll's estimate of the fraction of people in favor will be within 2 percentage points of the true value.

So…how many people are you going to need to interview?

Poll Setup

Let p be the true fraction of people who support the proposal.

Let X_i be indicator that the i^{th} person supports the proposal.

Report our estimate of p as $\hat{p} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} X_i$

What is the

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

$\mathrm{Var}(\hat{p}) =$

Poll Setup

Let p be the true fraction of people who support the proposal.

Let $X_i \sim \text{Ber}(p)$ -- indicator that the i^{th} person supports the proposal.

Report our estimate of p as $\hat{p} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} X_i$

What is the

$$\mathbb{E}[\hat{p}] = \frac{1}{n} \cdot \mathbb{E}[\sum X_i] = \frac{pn}{n} = p$$

$$\text{Var}(\hat{p}) = \frac{1}{n^2} \text{Var}(X_i) = \frac{p(1-p)}{n}$$

Poll Setup

Let p be the true fraction of people who support the proposal.

Let $X_i \sim \mathrm{Ber}(p)$ -- indicator that the i^{th} person supports the proposal.

Report our estimate of p as $\hat{p} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} X_i$

What is the

$$\mathbb{E}[\hat{p}] = \frac{1}{n} \cdot \mathbb{E}[\sum X_i] = \frac{pn}{n} = p$$

$$\mathrm{Var}(\hat{p}) = \frac{1}{n^2} \mathrm{Var}(X_i) = \frac{p(1-p)}{n}$$

By CLT, as $n \to \infty$,
$\hat{p} \to N\left(p, \frac{p(1-p)}{n}\right)$

Same question as: for what n is $P(|\hat{p} - p| > 0.02) \leq 0.05$?

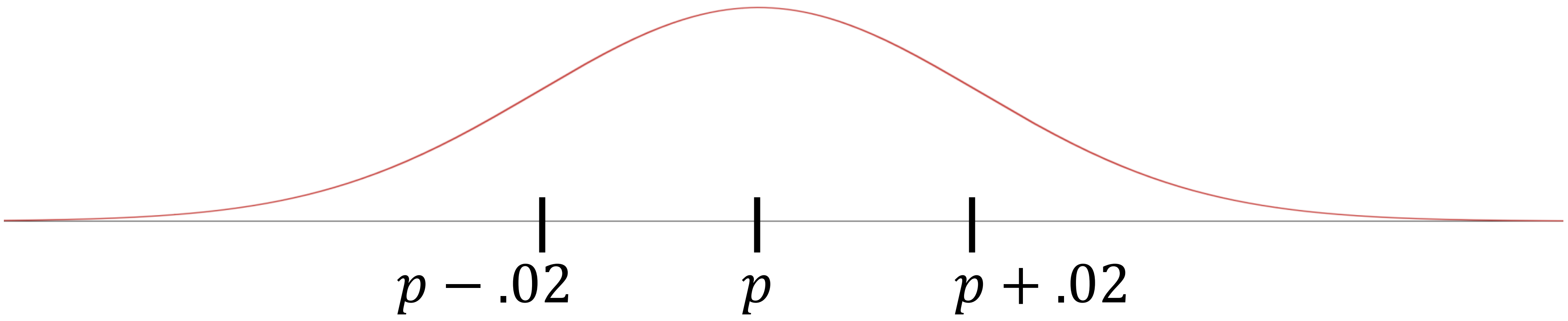
Using the CLT

What are we looking for? Well we have a margin of error:

$\mathbb{P}(p - .02 \leq \hat{p} \leq p + .02) \geq .95$

That says we're within the 2% margin of error at least 95% of the time.

Good estimate if $\hat{p}$ in this region



Same question as: for what n is $P(|\hat{p} - p| > 0.02) \leq 0.05$?

Using the CLT

What are we looking for? Well we have a margin of error:

$\mathbb{P}(p - .02 \leq \hat{p} \leq p + .02) \geq .95$

That says we're within the 2% margin of error at least 95% of the time.

Let's setup to use the CLT.

$$\mathbb{P}\left(\frac{p-.02-p}{\sqrt{p(1-p)/n}} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq \frac{p+.02-p}{\sqrt{p(1-p)/n}}\right) \geq .95$$

Apply the CLT

$$\mathbb{P}\left(\frac{p-.02-p}{\sqrt{p(1-p)/n}} \le \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \le \frac{p+.02-p}{\sqrt{p(1-p)/n}}\right) \ge .95$$

Is well approximated by $\mathbb{P}\left(\frac{-\sqrt{n}\cdot .02}{\sqrt{p(1-p)}} \le Z \le \frac{\sqrt{n}\cdot .02}{\sqrt{p(1-p)}}\right) \ge .95$ for $Z \sim \mathcal{N}(0,1)$

So as n changes, the probability changes. So choose the smallest n for which the probability is at least .95

WAIT, what's $\sqrt{p(1-p)}$? We don't know p. That's *why* we're doing the poll in the first place.

Handling $\sqrt{p(1-p)}$

Justification 1: If we make a mistake, we want it to be making n bigger. (since we're trying to say "take n at least this big, and you'll be safe").

The bigger the standard deviation, the bigger n will need to be to control it. So assume the biggest possible standard deviation.

Justification 2:

As $\sqrt{p(1-p)}$ gets bigger, the interval gets smaller (it's in the denominator), so assuming the biggest value of $\sqrt{p(1-p)}$ gives us the most restricted interval. So no matter what the true interval is we have a subset of it. And if our probability is at least $.95$ then the true probability is at least $.95$.

What's the maximum of $\sqrt{p(1-p)}$?

Worst value of p

Calculus time!

Set $\frac{d}{dp}\sqrt{p-p^2} = 0$

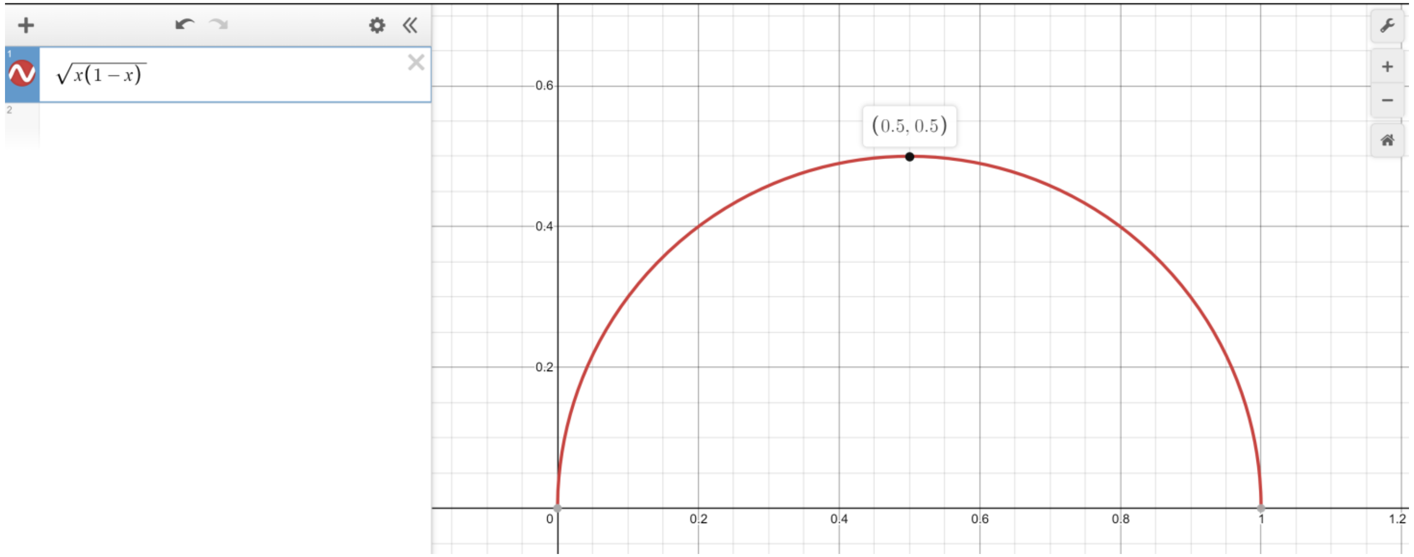
$\frac{1}{\sqrt{p-p^2}}(1-2p) = 0$

$1 - 2p = 0 \to p = 1/2$

Second derivative test will confirm $p = \frac{1}{2}$ is a maximizer

Or just plot it.

$\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)} = \sqrt{1/4}.$



Doing the algebra

$$\mathbb{P}\left(\frac{p-.02-p}{\sqrt{p(1-p)/n}} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq \frac{p+.02-p}{\sqrt{p(1-p)/n}}\right)$$

$$\approx \mathbb{P}\left(\frac{-\sqrt{n}\cdot .02}{\sqrt{p(1-p)}} \leq Z \leq \frac{\sqrt{n}\cdot .02}{\sqrt{p(1-p)}}\right) \text{ by CLT; } Z \sim \mathcal{N}(0,1)$$

$$\geq \mathbb{P}\left(\frac{-\sqrt{n}\cdot .02}{\sqrt{1/4}} \leq Z \leq \frac{\sqrt{n}\cdot .02}{\sqrt{1/4}}\right)$$

$$= \mathbb{P}(-.04\sqrt{n} \leq Z \leq .04\sqrt{n})$$

$$= \Phi(.04\sqrt{n}) - \left(1 - \Phi(.04\sqrt{n})\right) = 2\Phi(.04\sqrt{n}) - 1$$

$$2\Phi(.04\sqrt{n}) - 1 \geq .95 \rightarrow \Phi(.04\sqrt{n}) \geq \frac{1.95}{2}$$

Using the Φ-table

$\Phi(.04\sqrt{n}) \geq .975$

Φ-table says:

$.04\sqrt{n} \geq 1.96$

$\sqrt{n} \geq 49$

$n \geq 2401$. gives 95% confidence interval of +/- 2%.

I.e. 95% of the time, our poll gets a value within 2% of the true value.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

Confidence Intervals

We are trying to estimate some parameter (e.g., p). We output an estimator $\bar{X}$ such that

$$\mathbb{P}(|X - \mu| > \varepsilon) \leq \delta$$

A "confidence interval" tells you the probability (how confident you should be) that your random variable fell in a certain range (interval)

We are $(1 - \delta)$*100% confident that our result is an accurate estimate within ε*100% percent

Idealized Polling

We have just discussed “idealized polling”. Real life is normally not so nice ☹

Assumed we can sample people uniformly at random, not really possible in practice

- Not everyone responds
- Response rates might differ in different groups
- Will people respond truthfully?

Makes polling in real life much more complex than this idealized model!

Multiple Random Variables

Analyzing *Multiple* Random Variables

So far, we've pretty much only analyzed one random variable at a time

We've *worked* with multiple random variables (e.g., the sum of them) but haven't really analyzed their relationships except for independence

Today, we will talk about <u>analyzing the relations between multiple RVs</u>

> **Joint Distributions**
>
> *Joint PMF/PDF, Joint CDF, Joint Expectation*

Analyzing *Multiple* Random Variables

Examples:

- **Economics**: Analyzing the relationship between stock returns and volume helps in understanding market behavior and making investment decisions
- **Healthcare:** Understanding how different factors (like blood pressure and cholesterol levels) affects the probability of medical conditionals
- **Machine learning**: understanding how different features contribute to predicting a label and improving model accuracy (what features are most important)
- **Recommendation systems**: Understanding relationship between user preferences and item features to improve recommendation systems, or other factors like age and choices of products

Joint Distributions

We'll start with the **discrete** case

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

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

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

$\Omega_{X,Y}$ is the set of all possible **pairs** of values X and Y can be together

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

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

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

$\Omega_{X,Y}$ is the set of all possible **pairs** of values X and Y can be together

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

Joint **PMF** *(probability mass function)* - $\mathrm{p}_{X,Y}(a,b)$

$\mathrm{p}_{X,Y}(a,b)$ defines probabilities of X and Y being certain values at once

> $p_{X,Y}(a,b) = \mathbb{P}(X = a \cap Y = b) = \mathbb{P}(X = a, Y = b)$

should be defined for ***all*** *values of a and b*

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

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

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

$\Omega_{X,Y}$ is the set of all possible **pairs** of values X and Y can be together

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

Joint **PMF** *(probability mass function)* - $\mathrm{p}_{X,Y}(a, b)$

$\mathrm{p}_{X,Y}(a, b)$ defines probabilities of X and Y being certain values at once

> $p_{X,Y}(a, b) = \mathbb{P}(X = a \cap Y = b) = \mathbb{P}(X = a, Y = b)$

should be defined for ***all*** *values of a and b*

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

> $F_{X,Y}(a, b) = \mathbb{P}(X \le a \cap Y \le b) = \mathbb{P}(X \le a, Y \le b)$

should be defined for ***all*** *values of a and b*

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

(Joint) **Independence**

X and Y are independent if:

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

> $\Omega_{X,} = \Omega_X \times \Omega_Y$
if you find that $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$, that is enough to show that they aren't independent

This is the same definition we've seen before for independence of random variables
$\mathbb{P}(X = a, Y = b) = \mathbb{P}(X = a) \cdot \mathbb{P}(X = b)$

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let X be the value of the 1st dice
Let Y be the value of the 2nd dice

$\Omega_X = \Omega_Y = \{1,2,3,4\}$

The **joint support** is $\Omega_{X,Y} = \Omega_X \times \Omega_Y$
because all combinations are possible

What is the **joint PMF?**

$p_{X,Y}(a, b) =$

$p_{X,Y}$	Y=1	Y=2	Y=3	Y=4
X=1				
X=2				
X=3				
X=4				

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let X be the value of the 1st dice
Let Y be the value of the 2nd dice

$\Omega_X = \Omega_Y = \{1,2,3,4\}$

The **joint support** is $\Omega_{X,Y} = \Omega_X \times \Omega_Y$
because all combinations are possible

What is the **joint PMF**?

$$p_{X,Y}(a,b) = \begin{cases} 1/16 & \text{if } a,b \in \Omega_{X,Y} \\ 0 & \text{otherwise} \end{cases}$$

$p_{X,Y}$	Y=1	Y=2	Y=3	Y=4
X=1	1/16	1/16	1/16	1/16
X=2	1/16	1/16	1/16	1/16
X=3	1/16	1/16	1/16	1/16
X=4	1/16	1/16	1/16	1/16

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$
Let W be the max value $W = \max(X, Y)$

$\Omega_U = \Omega_W = \{1,2,3,4\}$

The **joint support** is:

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1				
U=2				
U=3				
U=4				

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$
Let W be the max value $W = \max(X, Y)$

$\Omega_U = \Omega_W = \{1,2,3,4\}$

The **joint support** is:
$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W; u \leq w\}$

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1				
U=2				
U=3				
U=4				

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$
Let W be the max value $W = \max(X, Y)$

$\Omega_U = \Omega_W = \{1,2,3,4\}$

The **joint support** is:
$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W; u \le w\}$

What is the **joint PMF?**

$p_{U,W}(u, w) =$

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1				
U=2				
U=3				
U=4				

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$
Let W be the max value $W = \max(X, Y)$

$\Omega_U = \Omega_W = \{1,2,3,4\}$

The **joint support** is:
$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W ; u \leq w\}$

What is the **joint PMF?**

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

$$p_{U,W}(u, w) = \begin{cases} 1/16 & \text{if } (u, w) \in \Omega_{U,W} \text{ and } u = w \\ 2/16 & \text{if } (u, w) \in \Omega_{U,W} \text{ and } u < w \\ 0 & \text{otherwise} \end{cases}$$

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$
Let W be the max value $W = \max(X, Y)$

$\Omega_U = \Omega_W = \{1,2,3,4\}$

The **joint support** is:
$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W; u \leq w\}$

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

What is the **joint PMF?**

$$p_{U,W}(u, w) = \begin{cases} 1/16 & \text{if } (u, w) \in \Omega_{U,W} \text{ and } u = w \\ 2/16 & \text{if } (u, w) \in \Omega_{U,W} \text{ and } u < w \\ 0 & \text{otherwise} \end{cases}$$

U and W are **NOT independent** because $\Omega_{U,W} \neq \Omega_U \times \Omega_W$

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$. Let W be the max value $W = \max(X, Y)$

Suppose I didn't know how to compute $p_U(u) = \mathbb{P}(U = u)$ *directly.*
Can we find it from the joint PMF $p_{U,V}(u, v)$*?*

We know $\mathbb{P}(U = 1 \cap V = v)$ for all v...

$\mathbb{P}(U = 1)$
$=$

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$. Let W be the max value $W = \max(X, Y)$

Suppose I didn't know how to compute $p_U(u) = \mathbb{P}(U = u)$ *directly.*
Can we find it from the joint PMF $p_{U,V}(u, v)$*?*

We know $\mathbb{P}(U = 1 \cap V = v)$ for all v...

$$\mathbb{P}(U = 1)$$
$$= \mathbb{P}(U = 1 \cap V = 1) + \mathbb{P}(U = 1 \cap V = 2) + \mathbb{P}(U = 1 \cap V = 3) + \mathbb{P}(U = 1 \cap V = 4)$$
$$= \frac{1}{16} + \frac{2}{16} + \frac{2}{16} + \frac{2}{16} = \frac{7}{16}$$

Use LoTP because the events $V = 1, V = 2, V = 3, V = 4$ ***partition the sample space***

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$. Let W be the max value $W = \max(X, Y)$

Suppose I didn't know how to compute $p_U(u) = \mathbb{P}(U = u)$ *directly.*
Can we find it from the joint PMF $p_{U,V}(u, v)$*?*

In general…

$$\mathbb{P}(U = u)$$
$$= \mathbb{P}(U = u \cap V = 1) + \mathbb{P}(U = u \cap V = 2)$$
$$+ \mathbb{P}(U = u \cap V = 3) + \mathbb{P}(U = u \cap V = 4)$$
$$= \sum_{v \in \Omega_V} p_{U,V}(u, v)$$

Use LoTP because the events $V = 1$, $V = 2$, $V = 3$, $V = 4$ partition the sample space

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$. Let W be the max value $W = \max(X, Y)$

Suppose I didn't know how to compute $p_U(u) = \mathbb{P}(U = u)$ directly.
Can we find it from the joint PMF $p_{U,V}(u, v)$?

In general...

$\mathbb{P}(U = u) = \sum_{v \in \Omega_V} p_{U,V}(u, v)$

Use LoTP because the events $V = 1$, $V = 2$, $V = 3$, $V = 4$ partition the sample space

$p_U(u)$ is the "marginal" PMF for U
(because we "marginalized" V)

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

Marginal Distribution

If you have the joint PMF for two random variables, you can find the PMF of one of them by **using the law of total probability**, **partitioning on the values of the other random variable:**

$$\boxed{p_X(a) = \sum_{b \in \Omega_Y} p_{X,Y}(a,b)}$$

^ it's the same PMF we've talked about before, the "marginal" is just there to indicate is was derived from a joint PMF

Example: Weird Dice

Roll 2 fair 4-sided dice independently

Let U be the min value $U = \min(X, Y)$. Let W be the max value $W = \max(X, Y)$

Suppose I didn't know how to compute $p_U(u) = \mathbb{P}(U = u)$ *directly.*
Can we find it from the joint PMF $p_{U,V}(u, v)$*?*

$$p_U(u) = \begin{cases} \frac{7}{16} & \text{if } u = 1 \\ \frac{5}{16} & \text{if } u = 2 \\ \frac{3}{16} & \text{if } u = 3 \\ \frac{1}{16} & \text{if } u = 4 \\ 0 & \text{otherwise} \end{cases}$$

$p_{U,V}$	V=1	V=2	V=3	V=4
U=1	1/16	2/16	2/16	2/16
U=2	0	1/16	2/16	2/16
U=3	0	0	1/16	2/16
U=4	0	0	0	1/16

Joint **Expectation**

Joint Expectations

For a function $\boldsymbol{g(X,Y)}$, the expectation can be written in terms of the joint PMF.

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

Same ideas as before!

Examples of joint functions:

$g(X,Y) = X + Y$, $g(X,Y) = XY$, $g(X,Y) = X^Y$, $g(X,Y) = \binom{X}{Y}$, *etc.*

Joint Distributions

For **continuous** joint distributions, everything will look very similar but we *replace summations with integrals*, and use a *density function instead of the PMF*

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 $(a, b) \in \Omega_{X,Y}$
> $\Omega_{X,} = \Omega_X \times \Omega_Y$

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

Example: Darts

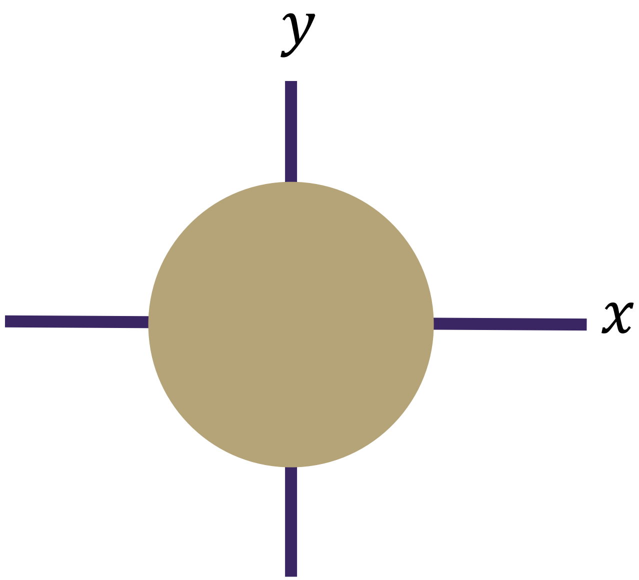
We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

Let's find and sketch the **joint range** $\Omega_{X,Y}$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

Let's find and sketch the **joint range** $\Omega_{X,Y}$

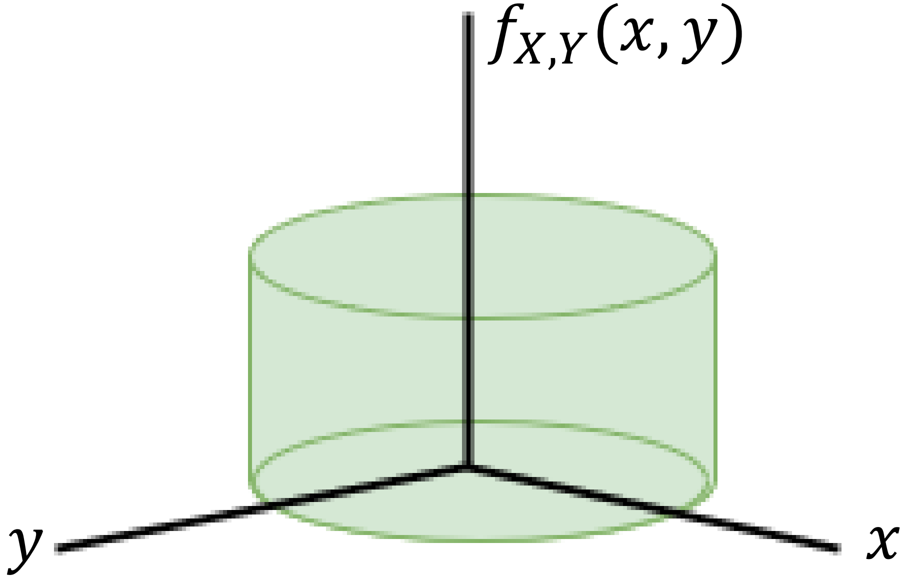


$$\Omega_{X,Y} = \{(x, y) \in \mathbb{R} \times \mathbb{R}; \mathrm{x}^2 + \mathrm{y}^2 \leq r\}$$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

Let's find and sketch the **joint PDF** $f_{X,Y}(x,y)$

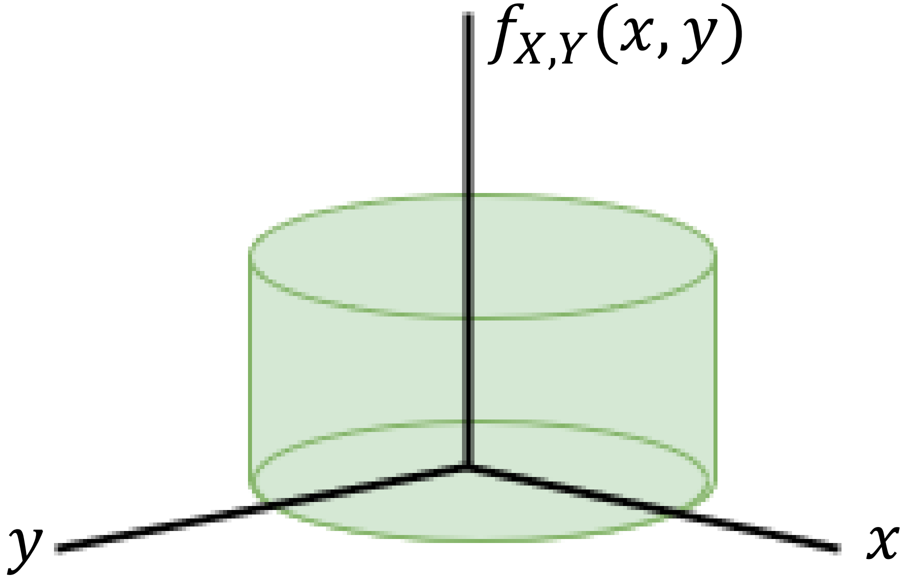


$$f_{X,Y}(x,y) = \begin{cases} \\ \\ \end{cases}$$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

Let's find and sketch the **joint PDF** $f_{X,Y}(x,y)$



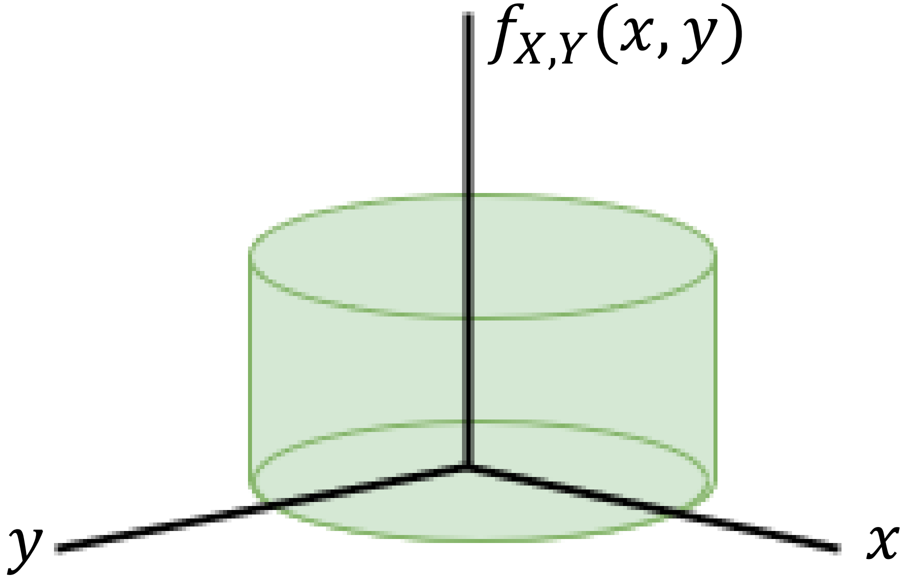
$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi r^2} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi r^2} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the marginal PDF for X and Y?



$f_X(a) =$

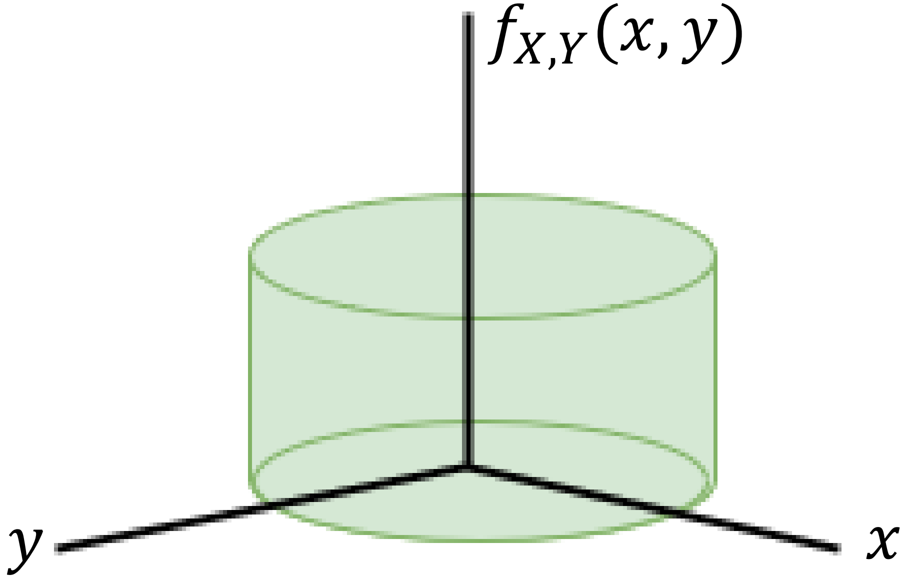
$f_Y(b) =$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi r^2} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the marginal PDF for X and Y?



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

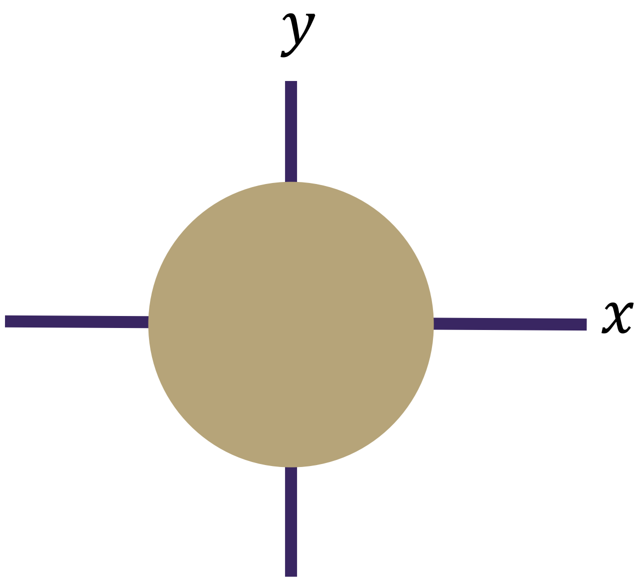
$$f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x,b)dx$$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi r^2} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the marginal PDF for X and Y?



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

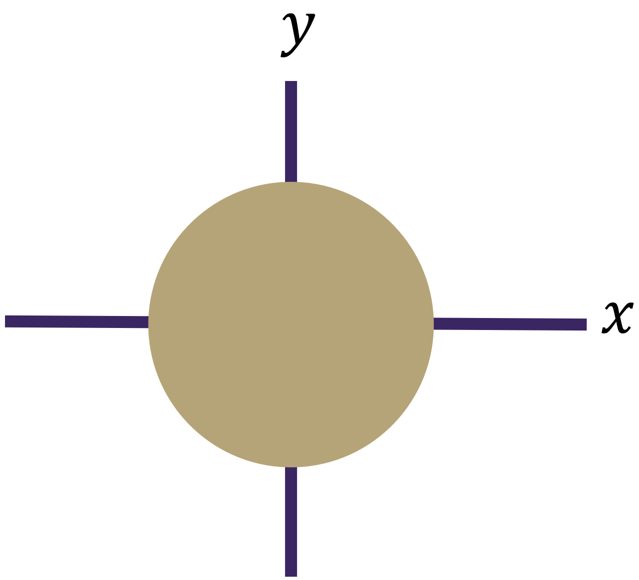
$$f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x,b)dx$$

Example: Darts

We throw a dart uniformly at random onto a circle of radius r centered around the version. X and Y are the x and y coordinates of the point the dart lands at.

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi r^2} & x^2 + y^2 \leq r \\ 0 & \text{otherwise} \end{cases}$$

What is the marginal PDF for X and Y?



$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a,y)dy = \int_{-\sqrt{r-a^2}}^{\sqrt{r-a^2}} \frac{1}{\pi r^2} dy = \frac{2\sqrt{r-a^2}}{\pi r^2}$$

$$f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x,b)dx = \int_{-\sqrt{r-b^2}}^{\sqrt{r-b^2}} \frac{1}{\pi r^2} dx = \frac{2\sqrt{r-b^2}}{\pi r^2}$$

Joint Continuous Probabilities

Just like we've done with PDFs for single random variables…

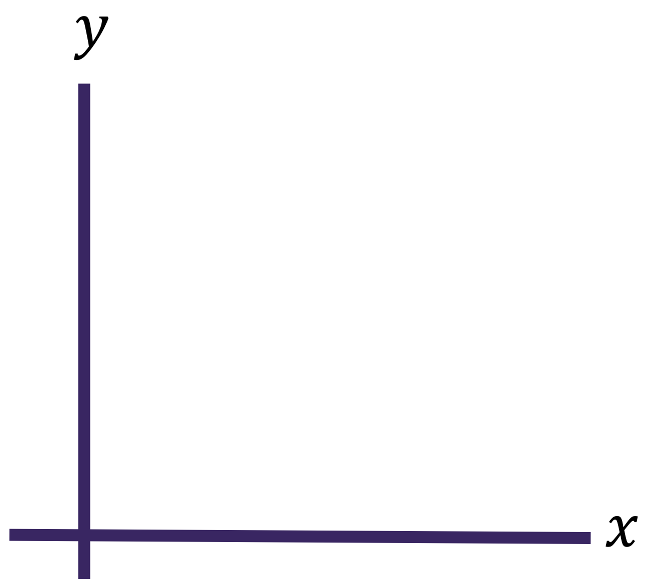
$$\mathbb{P}(a \leq X \leq b \cap c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dy\, dx$$

Example: Continuous Servers

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

What is the probability server 1 crashes before server 2?

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

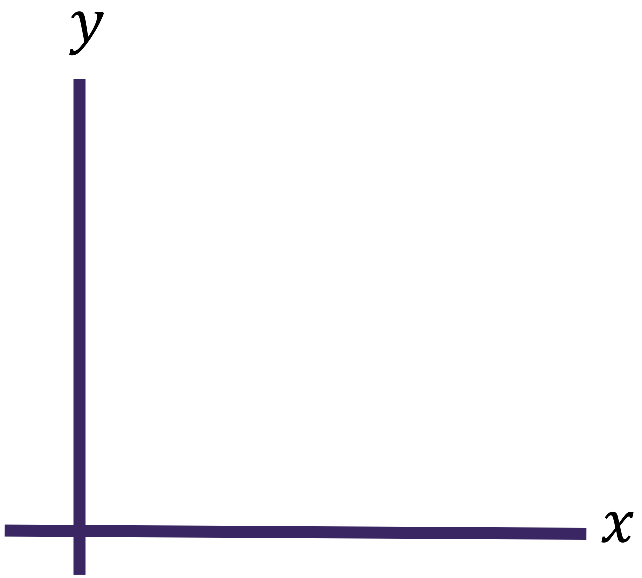


Example: Continuous Servers

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

What is the probability server 1 crashes before server 2?

$$\mathbb{P}(X < Y)$$
$$= \int_0^\infty \int_x^\infty f_{X,Y}(x, y)\, dy\, dx$$
$$= \int_0^\infty \int_x^\infty f_X(x) f_Y(y)\, dy\, dx \text{ by independence}$$



Analogues for continuous

Everything we saw today has a continuous version.

There are "no surprises"– replace pmf with pdf 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	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)dsdt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$
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	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y)p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy$
Independence	$\forall x,y, p_{X,Y}(x,y) = p_X(x)p_Y(y)$	$\forall x,y, f_{X,Y}(x,y) = f_X(x)f_Y(y)$