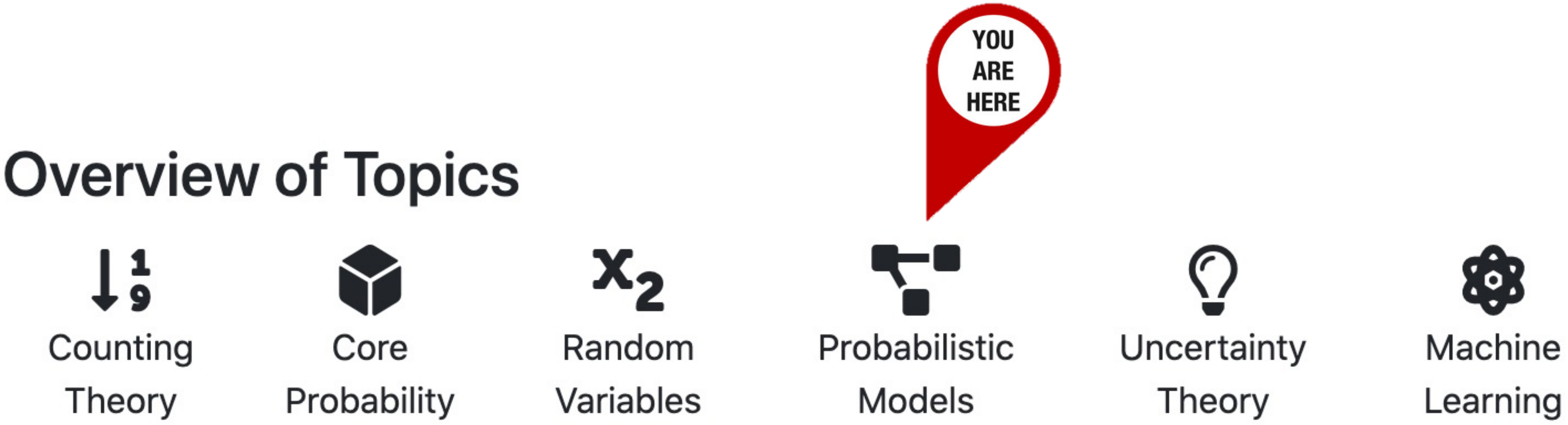
Continuous Joint Distributions

Chris Piech
CS109, Stanford University

Announcements

- Midterm next-next Tuesday.
  - 7-9p in Hewlett 200. Oct 26th
  - Need accommodations? Let us know before Thurs, Oct 21st 9a.
  - Cumulative up until next Friday. Most emphasis will be on the topics in the first 3 psets. Practice starting next Monday.
  - No class the day before the midterm!
- Pset #3 we removed problem 9c.

Where are we in CS109?



CS109 Flow

Today

Discrete Joint Distributions: General Case	Multinomial: A parametric Discrete Joint	Cont. Joint Distributions: General Case

Learning Goals

1. Know how to use a multinomial
2. Be able to calculate large bayes problems using a computer
3. Use a Joint CDF

Products become Sums!

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$\log(\prod_i a_i) = \sum_i \log(a_i)$$

* Spoiler alert: This is important because the product of many small numbers gets hard for computers to represent.

For Multiple RVs, Joint is Complete Information!

$$P(X = a, Y = y)$$

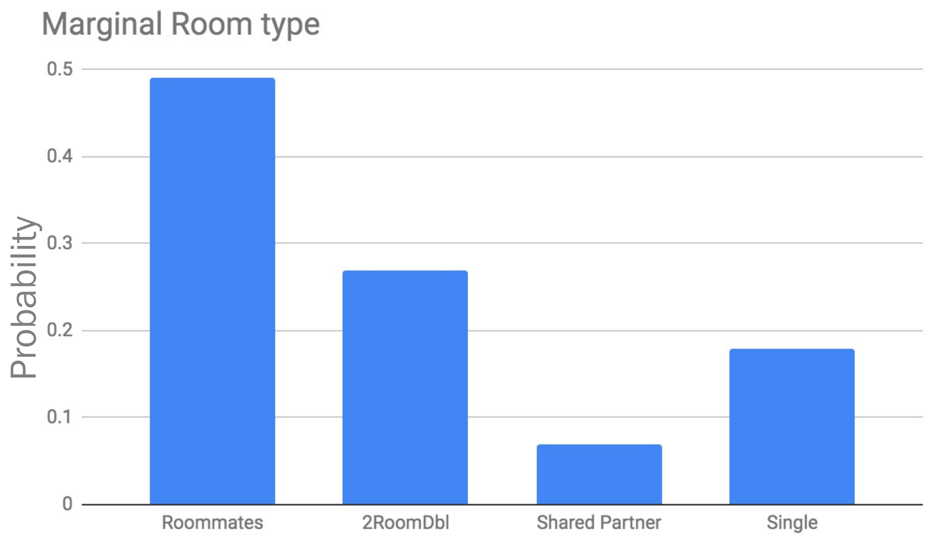
	Single	Relationship	Complicated
Frosh	0.13	0.08	0.02
Soph	0.17	0.11	0.02
Junior	0.09	0.10	0.02
Senior	0.02	0.07	0.01
5+	0.06	0.09	0.04

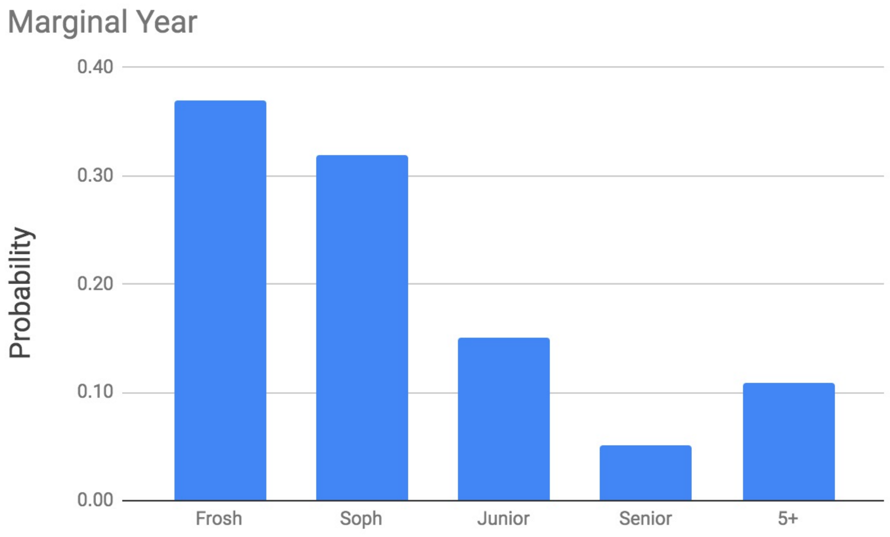
For discrete random variables the joint might look like a table.

A joint distribution is complete information. It can be used to answer any probability question.

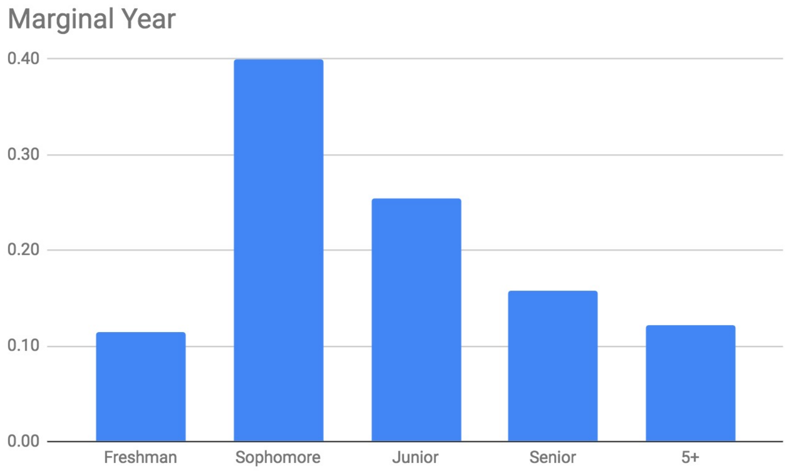
Joint Probability Table

	Roommates	2RoomDbl	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00

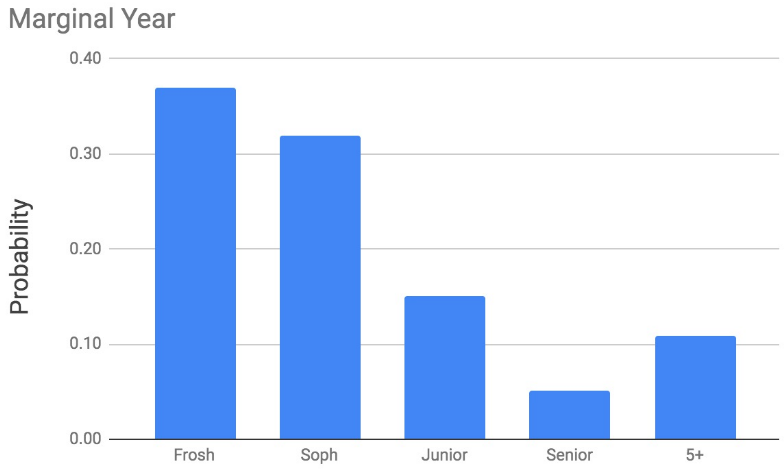




Change in Marginal Year



Fall quarter '18



Spr quarter '19

Can’t always use tables
(size grows exponentially with
number of RVs)

The Multinomial

Multinomial distribution

- *n* independent trials of experiment performed
- Each trial results in one of *m* outcomes, with respective probabilities: $p_1, p_2, \ldots, p_m$ where
- X_i = number of trials with outcome *i*

$$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} p_1^{c_1} p_2^{c_2} \ldots p_m^{c_m}$$

Joint distribution

Multinomial # ways of ordering the successes

Probabilities of each ordering are equal and mutually exclusive

where $\sum_{i=1}^{m} c_i = n$ and $\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1! c_2! \cdots c_m!}$

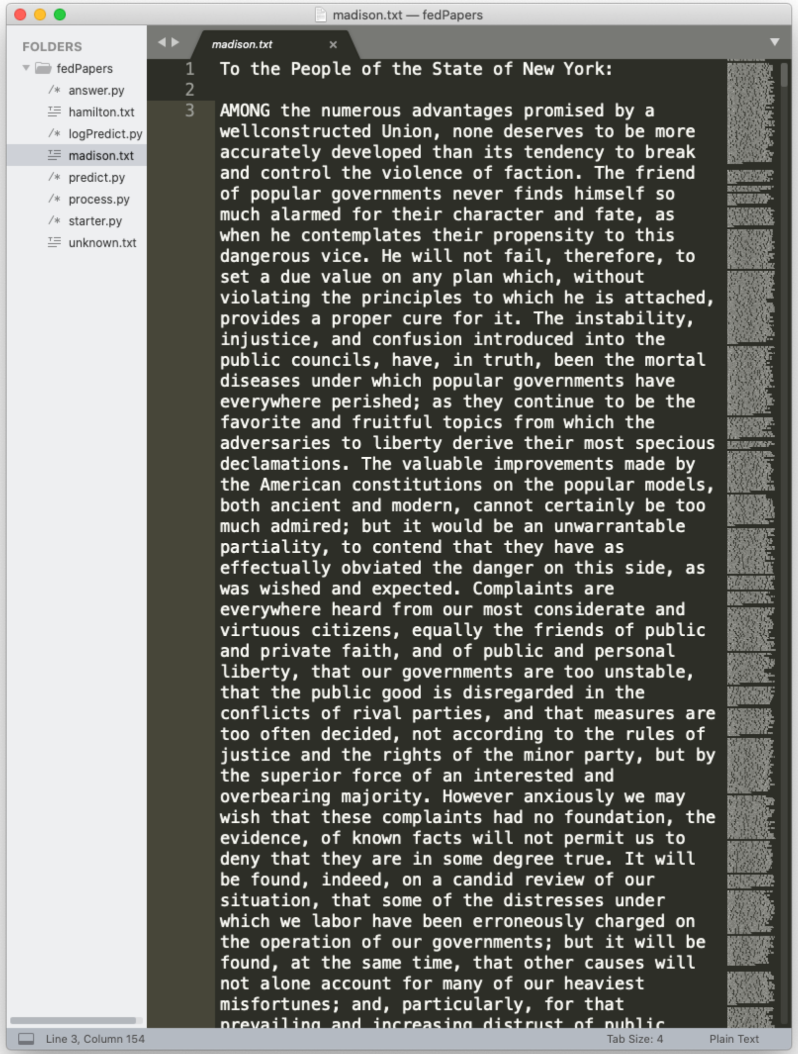
Probabilistic Text Analysis

According to the Global Language Monitor there are 988,968 words in the english language used on the internet.

Who wrote the federalist papers?

Who wrote Federalist Paper 53?

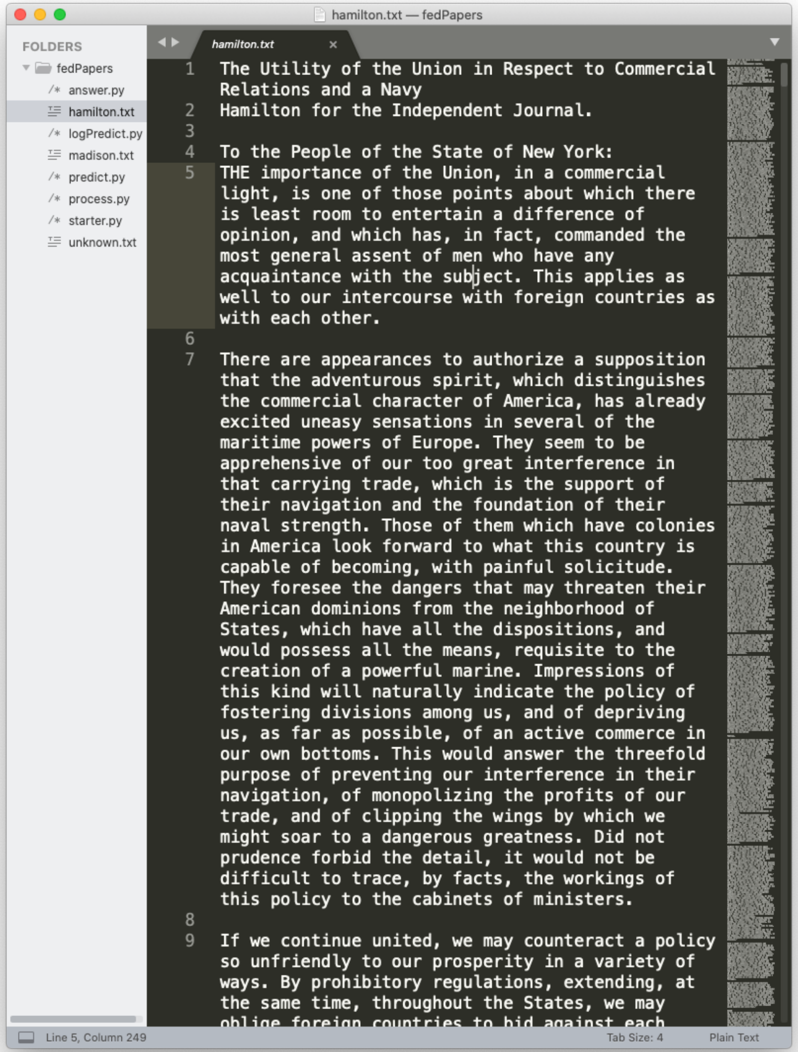
madison.txt



To the People of the State of New York:

AMONG the numerous advantages promised by a wellconstructed Union, none deserves to be more accurately developed than its tendency to break and control the violence of faction. The friend of popular governments never finds himself so much alarmed for their character and fate, as when he contemplates their propensity to this dangerous vice. He will not fail, therefore, to set a due value on any plan which, without violating the principles to which he is attached, provides a proper cure for it. The instability, injustice, and confusion introduced into the public councils, have, in truth, been the mortal diseases under which popular governments have everywhere perished; as they continue to be the favorite and fruitful topics from which the adversaries to liberty derive their most specious declamations. The valuable improvements made by the American constitutions on the popular models, both ancient and modern, cannot certainly be too much admired; but it would be an unwarrantable partiality, to contend that they have as effectually obviated the danger on this side, as was wished and expected. Complaints are everywhere heard from our most considerate and virtuous citizens, equally the friends of public and private faith, and of public and personal liberty, that our governments are too unstable, that the public good is disregarded in the conflicts of rival parties, and that measures are too often decided, not according to the rules of justice and the rights of the minor party, but by the superior force of an interested and overbearing majority. However anxiously we may wish that these complaints had no foundation, the evidence, of known facts will not permit us to deny that they are in some degree true. It will be found, indeed, on a candid review of our situation, that some of the distresses under which we labor have been erroneously charged on the operation of our governments; but it will be found, at the same time, that other causes will not alone account for many of our heaviest misfortunes; and, particularly, for that prevailing and increasing distrust of public

hamilton.txt



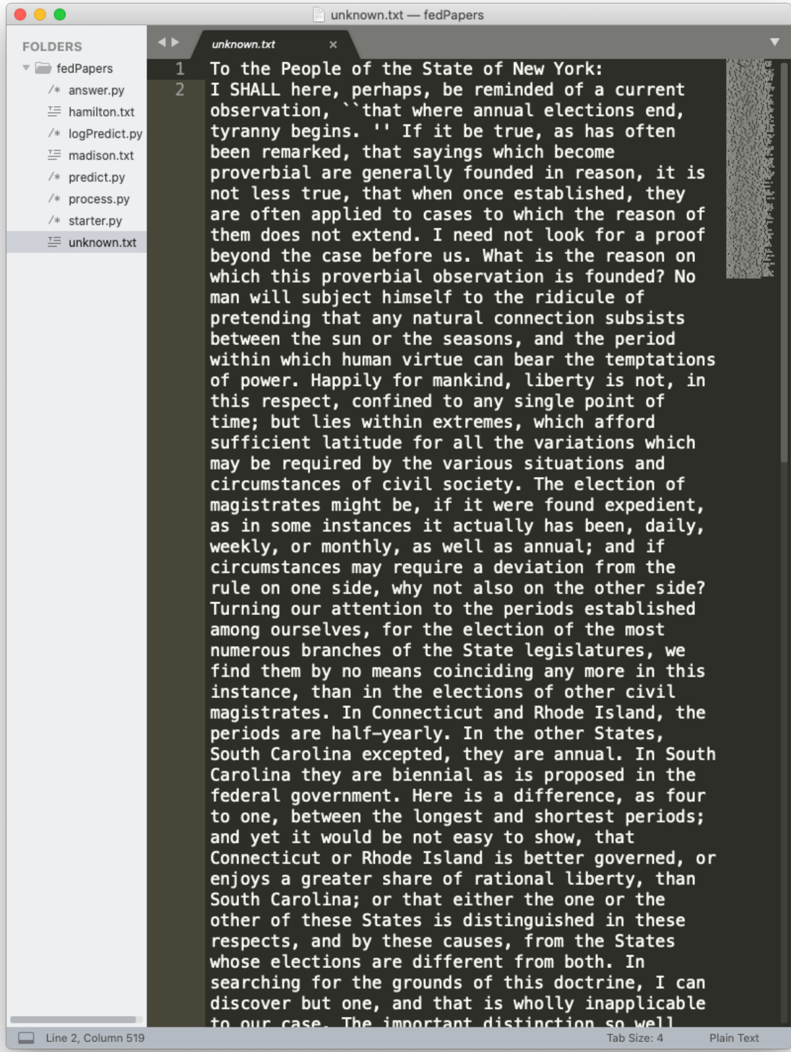
The Utility of the Union in Respect to Commercial Relations and a Navy
Hamilton for the Independent Journal.

To the People of the State of New York:
THE importance of the Union, in a commercial light, is one of those points about which there is least room to entertain a difference of opinion, and which has, in fact, commanded the most general assent of men who have any acquaintance with the subject. This applies as well to our intercourse with foreign countries as with each other.

There are appearances to authorize a supposition that the adventurous spirit, which distinguishes the commercial character of America, has already excited uneasy sensations in several of the maritime powers of Europe. They seem to be apprehensive of our too great interference in that carrying trade, which is the support of their navigation and the foundation of their naval strength. Those of them which have colonies in America look forward to what this country is capable of becoming, with painful solicitude. They foresee the dangers that may threaten their American dominions from the neighborhood of States, which have all the dispositions, and would possess all the means, requisite to the creation of a powerful marine. Impressions of this kind will naturally indicate the policy of fostering divisions among us, and of depriving us, as far as possible, of an active commerce in our own bottoms. This would answer the threefold purpose of preventing our interference in their navigation, of monopolizing the profits of our trade, and of clipping the wings by which we might soar to a dangerous greatness. Did not prudence forbid the detail, it would not be difficult to trace, by facts, the workings of this policy to the cabinets of ministers.

If we continue united, we may counteract a policy so unfriendly to our prosperity in a variety of ways. By prohibitory regulations, extending, at the same time, throughout the States, we may oblige foreign countries to bid against each

unknown.txt



To the People of the State of New York:
I SHALL here, perhaps, be reminded of a current observation, ``that where annual elections end, tyranny begins. '' If it be true, as has often been remarked, that sayings which become proverbial are generally founded in reason, it is not less true, that when once established, they are often applied to cases to which the reason of them does not extend. I need not look for a proof beyond the case before us. What is the reason on which this proverbial observation is founded? No man will subject himself to the ridicule of pretending that any natural connection subsists between the sun or the seasons, and the period within which human virtue can bear the temptations of power. Happily for mankind, liberty is not, in this respect, confined to any single point of time; but lies within extremes, which afford sufficient latitude for all the variations which may be required by the various situations and circumstances of civil society. The election of magistrates might be, if it were found expedient, as in some instances it actually has been, daily, weekly, or monthly, as well as annual; and if circumstances may require a deviation from the rule on one side, why not also on the other side? Turning our attention to the periods established among ourselves, for the election of the most numerous branches of the State legislatures, we find them by no means coinciding any more in this instance, than in the elections of other civil magistrates. In Connecticut and Rhode Island, the periods are half-yearly. In the other States, South Carolina excepted, they are annual. In South Carolina they are biennial as is proposed in the federal government. Here is a difference, as four to one, between the longest and shortest periods; and yet it would be not easy to show, that Connecticut or Rhode Island is better governed, or enjoys a greater share of rational liberty, than South Carolina; or that either the one or the other of these States is distinguished in these respects, and by these causes, from the States whose elections are different from both. In searching for the grounds of this doctrine, I can discover but one, and that is wholly inapplicable to our case. The important distinction so well

Who wrote Federalist Paper 53?

Prob Document given Hamilton

Prior belief it was Hamilton

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Prob Hamilton given Document

Prob of the document???

Who wrote Federalist Paper 53?

Model document as a multinomial where we care about count of words

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Who wrote Federalist Paper 53?

$$P(H|D) = \frac{\binom{n}{c_1 \dots c_k} \cdot \prod_i h_i^{c_i} \cdot P(H)}{P(D)}$$

Loop over unique words

Prob hamilton would write word i

Prior belief it was Hamilton

Number of times word i is in the doc

Prob Hamilton given Document

Prob of the document???

Who wrote Federalist Paper 53?

Prob that Hamilton wrote it

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

$$= \frac{P(H) \cdot \binom{n}{c_1 \ldots c_m} \cdot \prod_i h_i^{c_i}}{P(D)}$$

Prob that Madison wrote it

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$= \frac{P(M) \cdot \binom{n}{c_1 \ldots c_m} \cdot \prod_i m_i^{c_i}}{P(D)}$$

$$\frac{P(H|D)}{P(M|D)} = \frac{P(M) \cdot \binom{n}{c_1 \ldots c_k} \cdot \prod_i h_i^{c_i}}{P(D)} \Big/ \frac{P(H) \cdot \binom{n}{c_1 \ldots c_k} \cdot \prod_i m_i^{c_i}}{P(D)}$$

$$= \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}}$$

To the code

What happened?

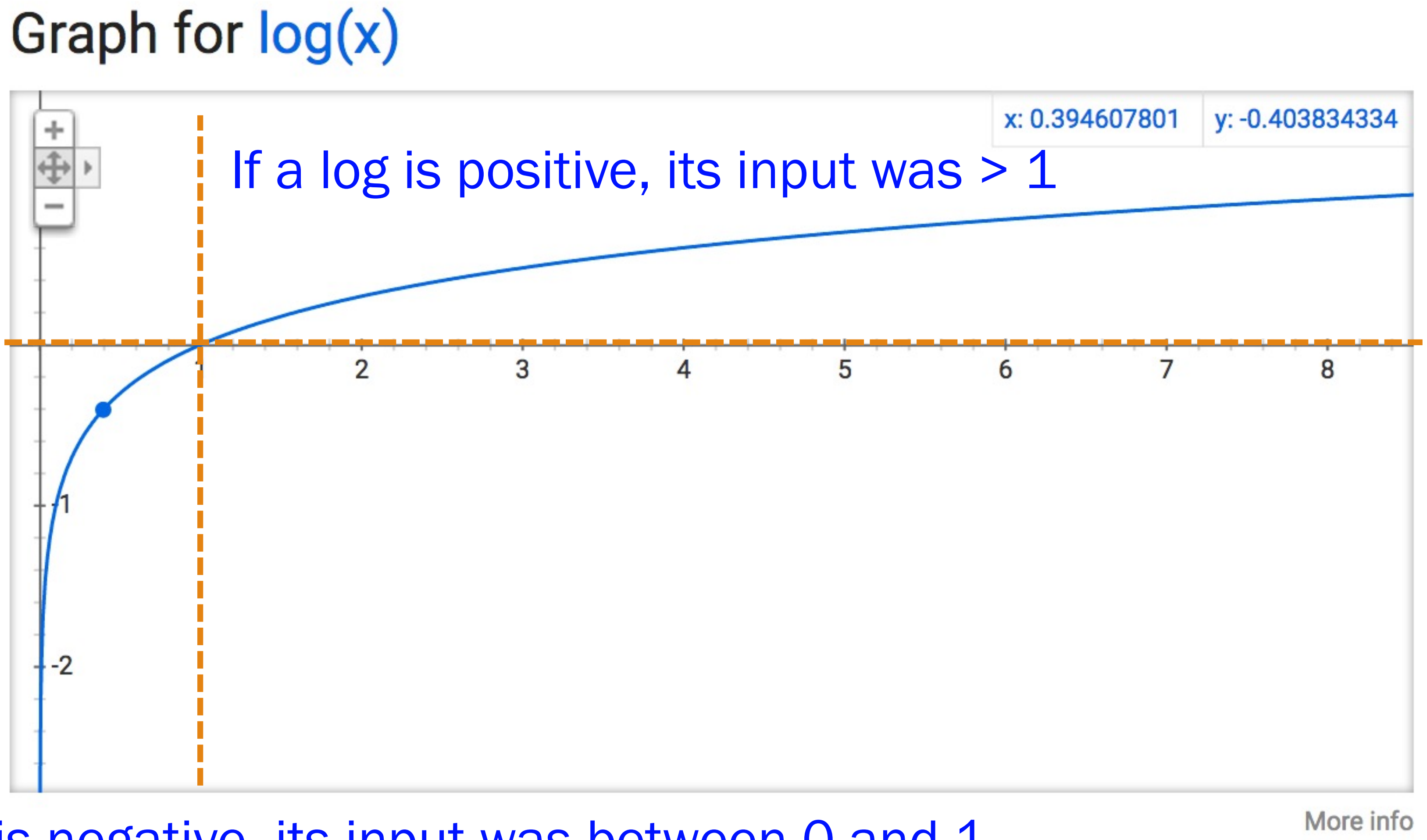
All our probabilities are zero…

Use logs when probabilities become too small!

$$\frac{P(H|D)}{P(M|D)} = \frac{\prod_i m_i^{c_i}}{\prod_i h_i^{c_i}}$$

$$\begin{aligned}\log \frac{P(H|D)}{P(M|D)} &= \log \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}} \\ &= \sum_i \log h_i^{c_i} - \sum_i \log m_i^{c_i} \\ &= \sum_i c_i \cdot \log h_i - \sum_i c_i \log m_i\end{aligned}$$

What does it mean if a log value is positive / negative

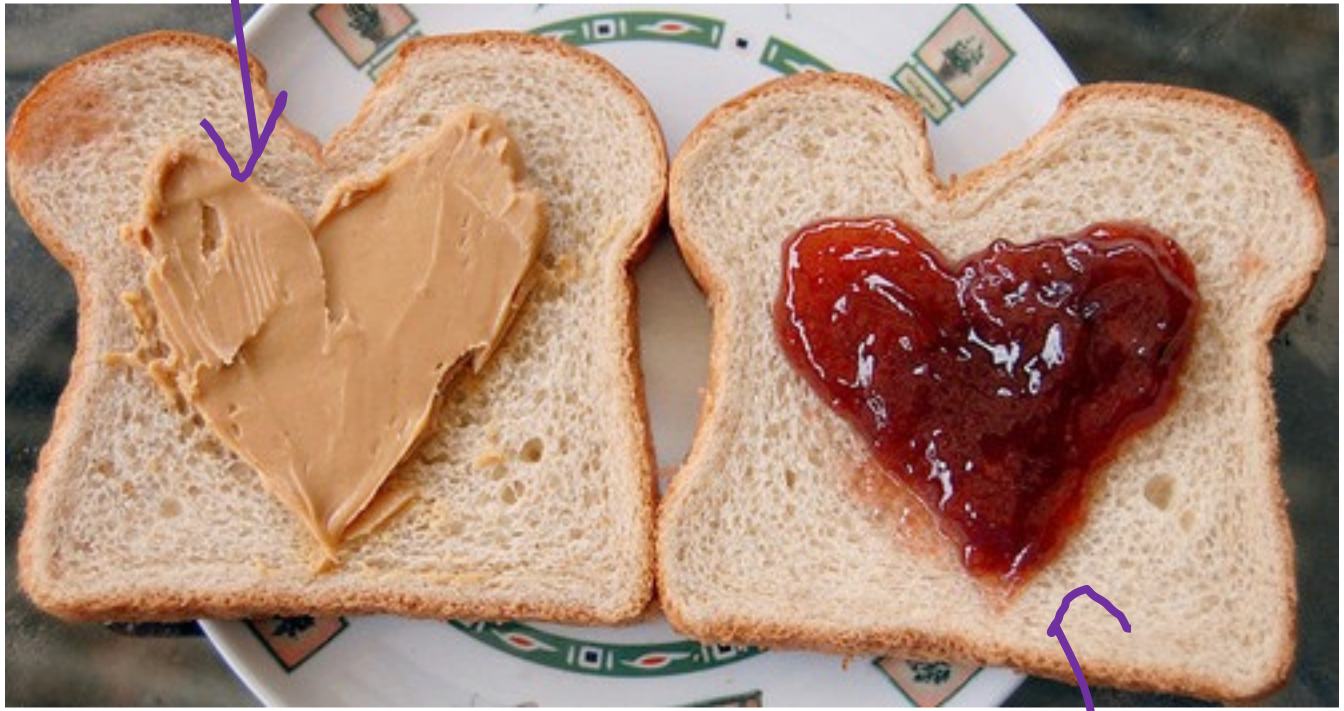


If a log is negative, its input was between 0 and 1 woot

Joint Random Variables

- Use a joint table, density function or CDF to solve probability question
- Think about **conditional** probabilities with joint variables (which might be continuous)
- Use and find **independence** of random variables
- Use and find **expectation** of random variables
- What happens when you **add** random variables?

Continuous Random Variables



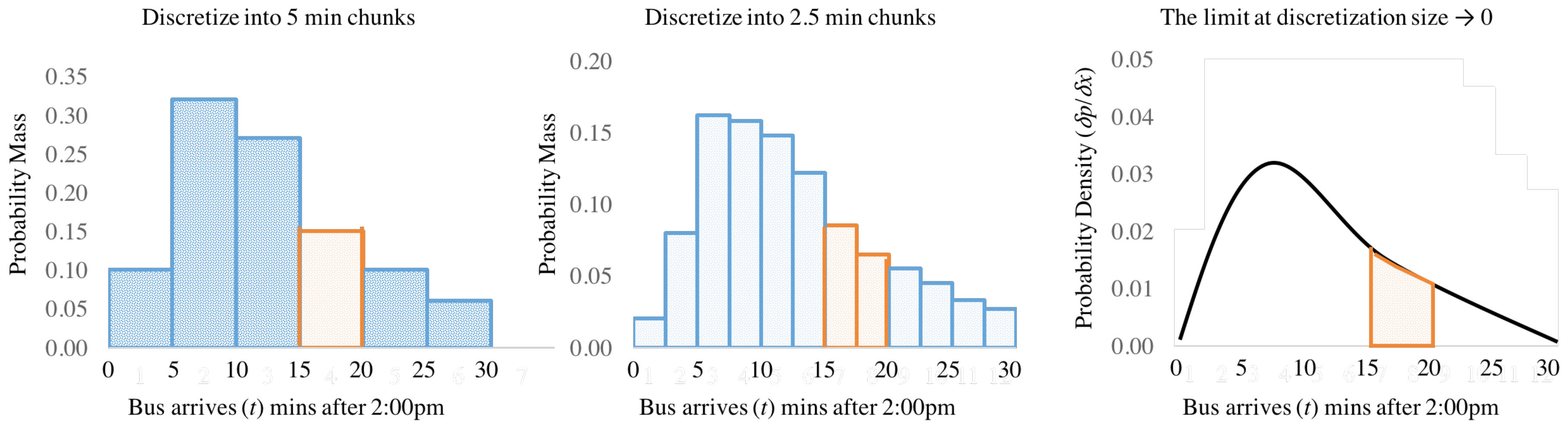
Joint Distributions

Continuous Joint Distribution

Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You arrive at 2:20pm.

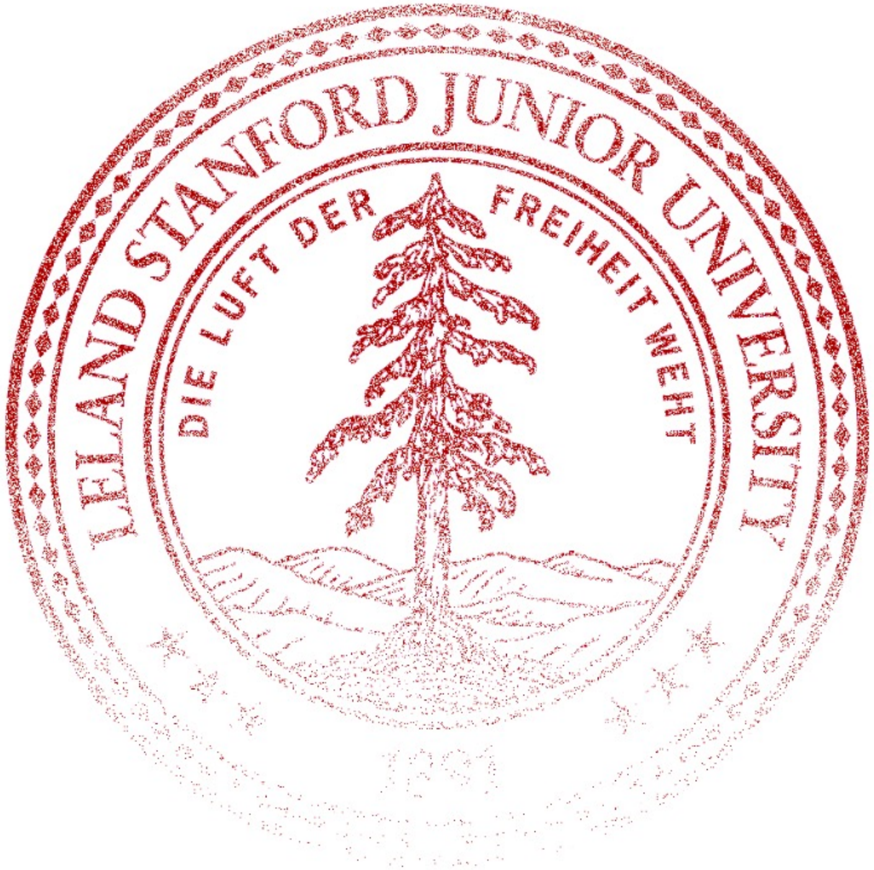
What is P(wait < 5 min)?



Joint Dart Distribution

Dart Results

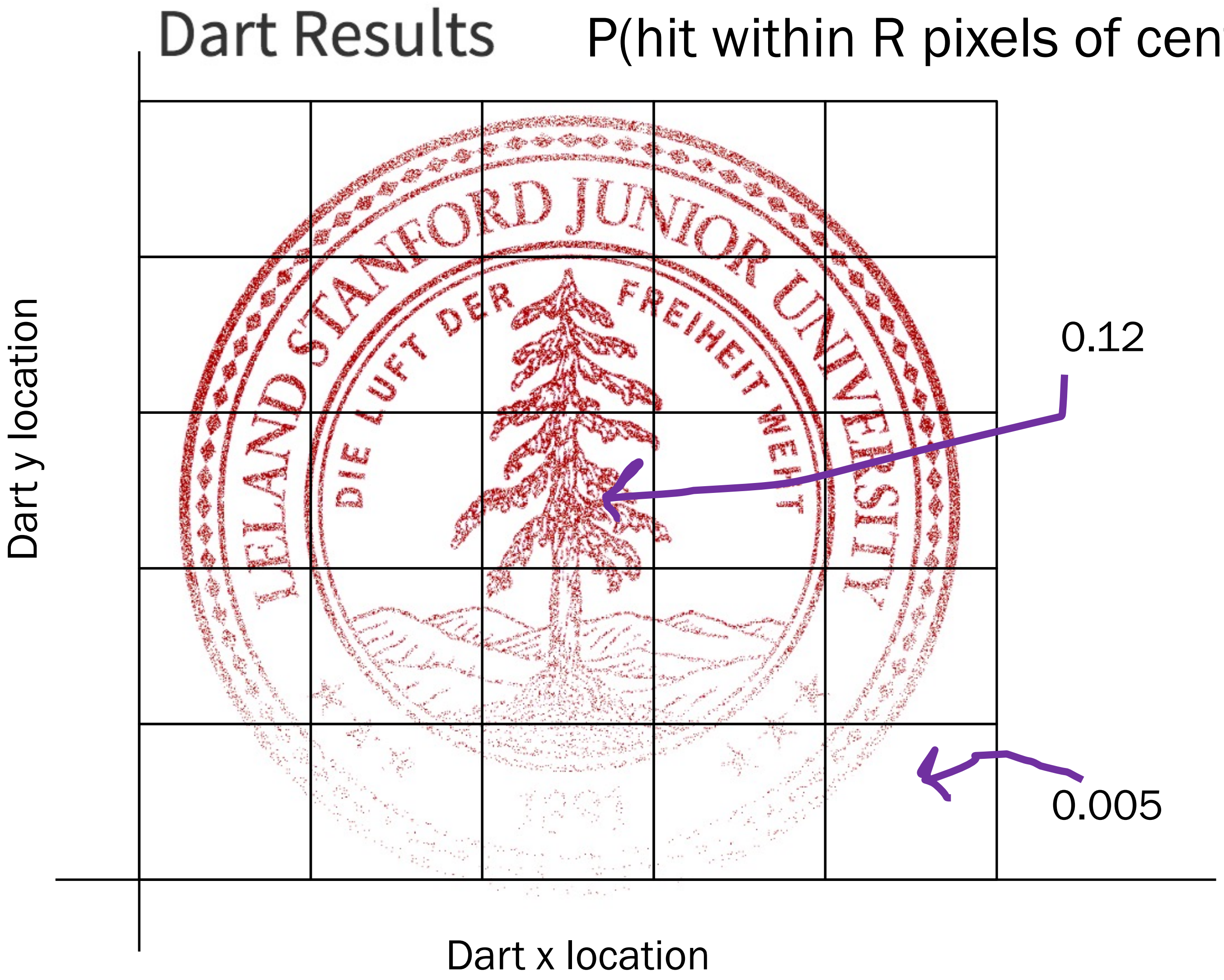
P(hit within R pixels of center)?



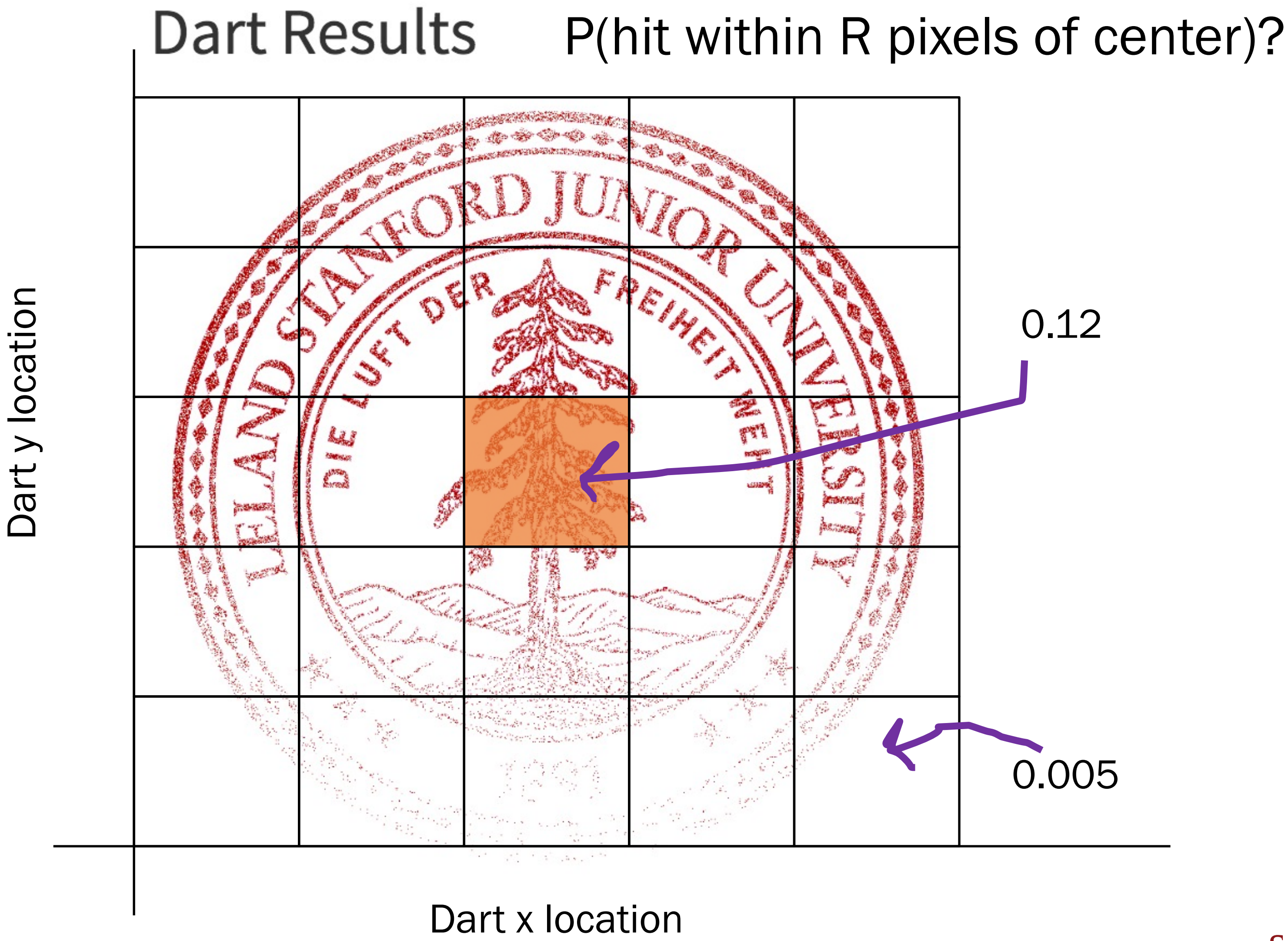
What is the probability that a dart hits at (456.234231234122355, 532.12344123456)?

Joint Dart Distribution

P(hit within R pixels of center)?

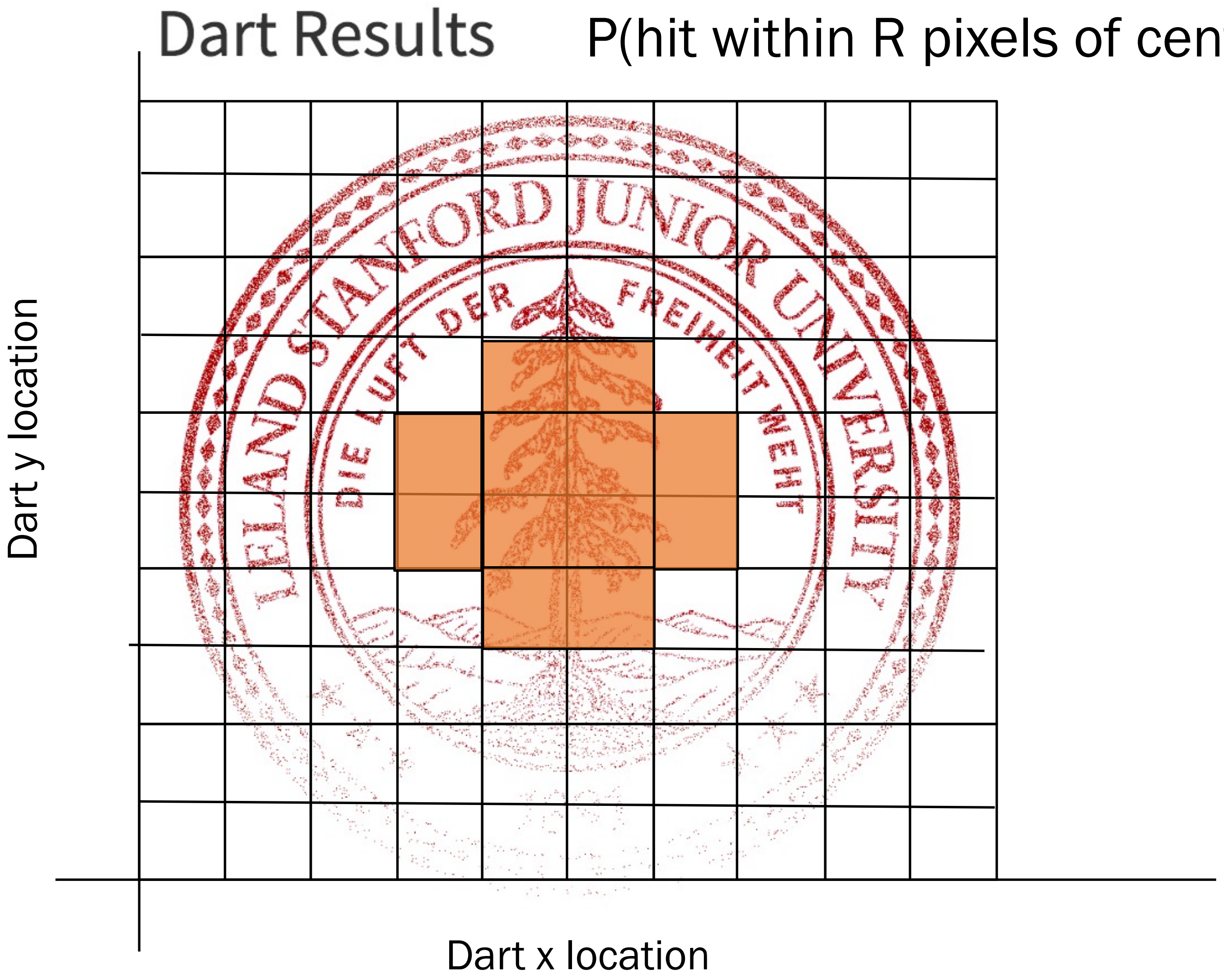


Joint Dart Distribution



Joint Dart Distribution

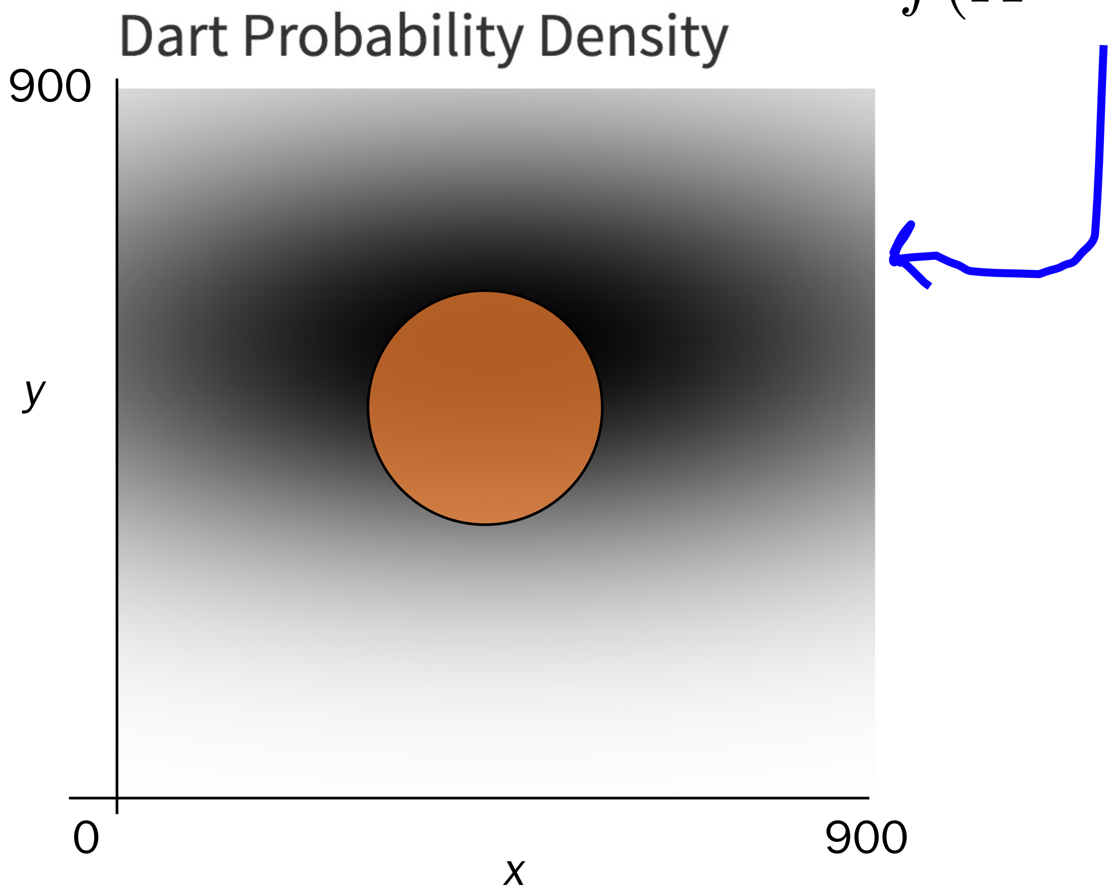
P(hit within R pixels of center)?



Stanford University

Joint Dart Distribution

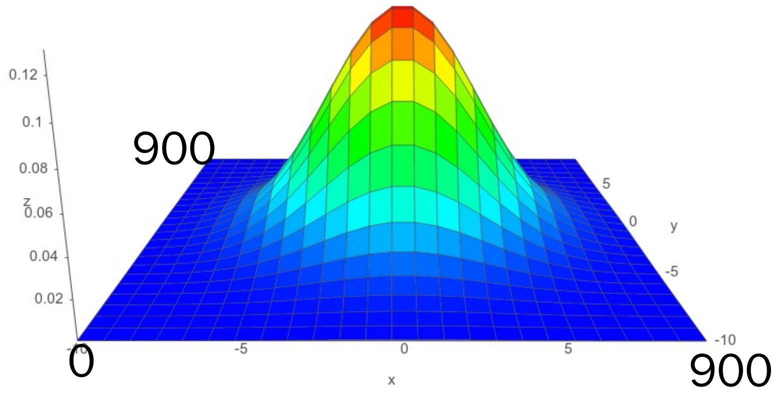
$$f(X = x, Y = y)$$

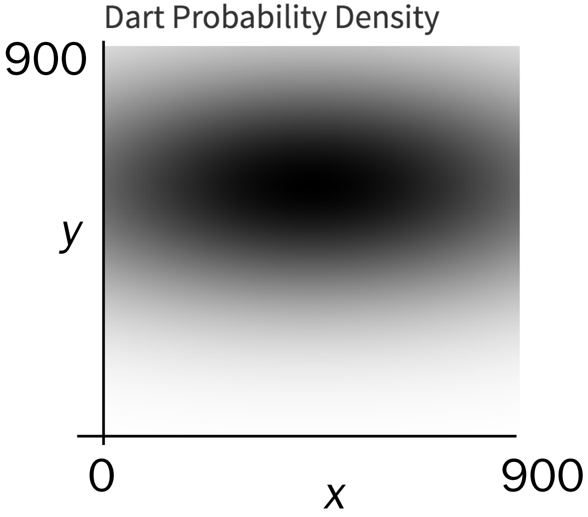


In the limit, as you break down continuous values into intestinally small buckets, you end up with multidimensional probability density

Joint Probability Density Funciton

A **joint probability density function** gives the relative likelihood of **more than one** continuous random variable **each** taking on a specific value.

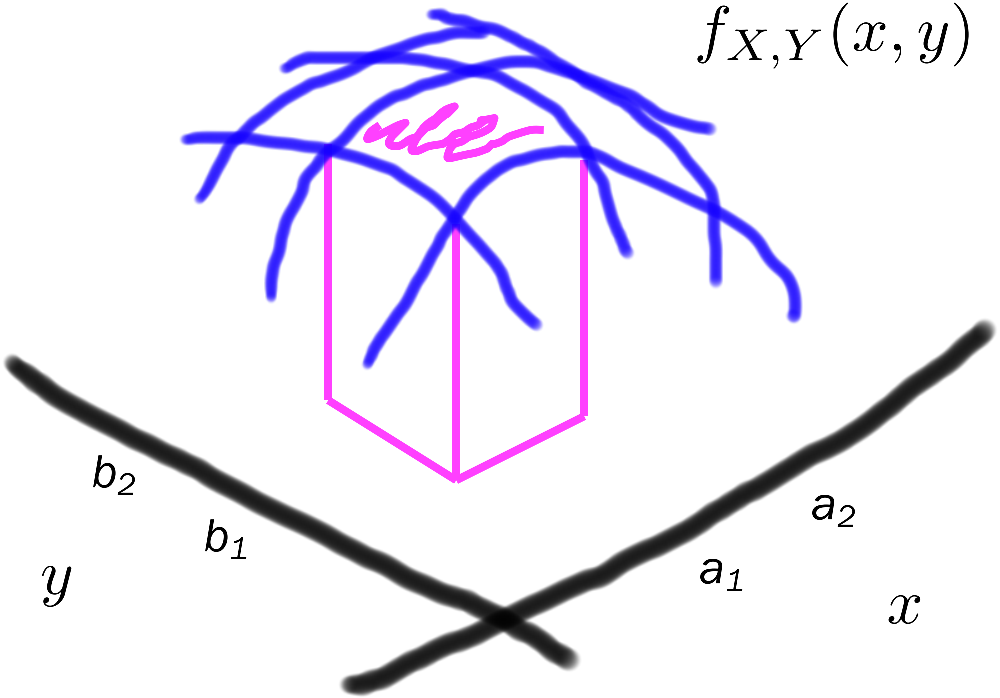




$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$

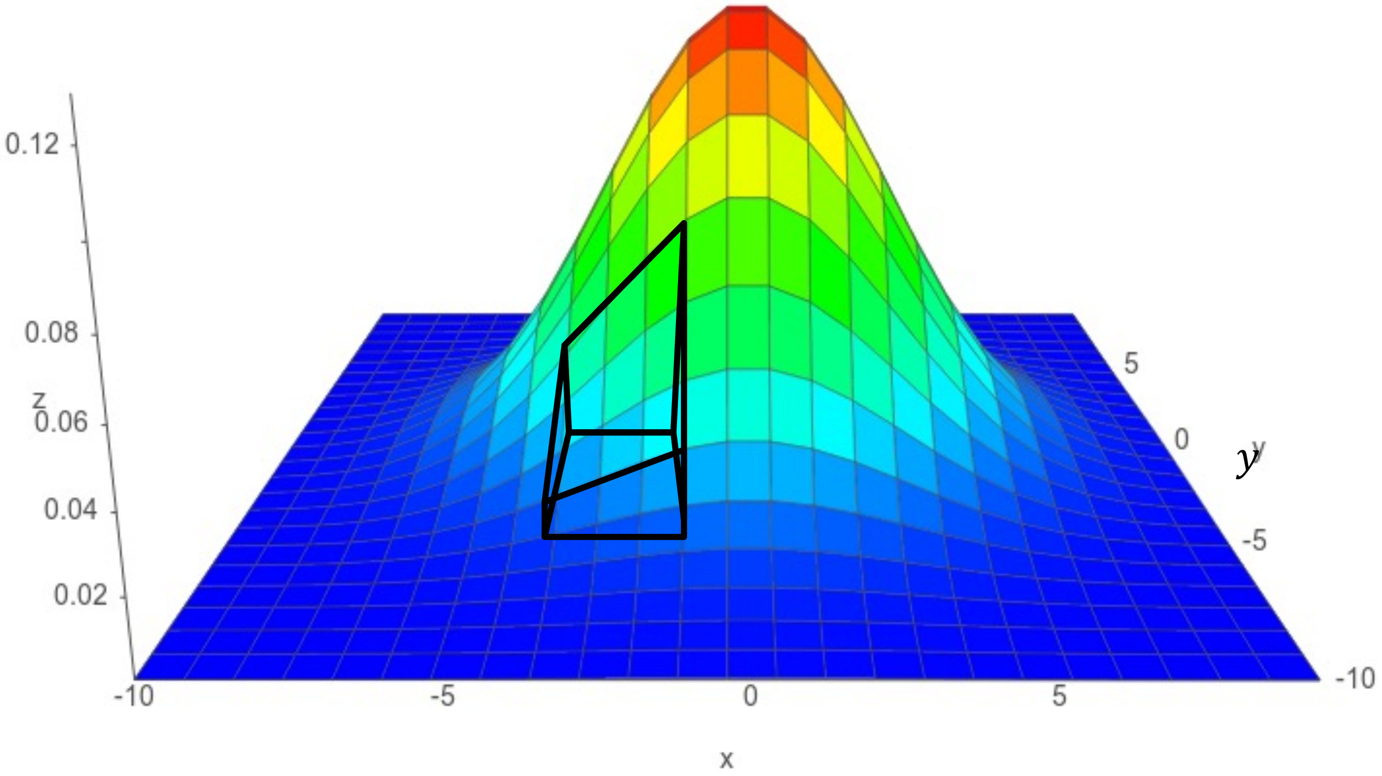
Joint Probability Density Funciton

$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$



Joint Probability Density Function

$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$



plot by Academo

Multiple Integrals Without Tears

Let X and Y be two continuous random variables

- where $0 \le X \le 1$ and $0 \le Y \le 2$

We want to integrate $g(x,y) = xy$ w.r.t. X and Y:

- First, do "innermost" integral (treat y as a constant):

$$\int_{y=0}^{2} \int_{x=0}^{1} xy \, dx \, dy = \int_{y=0}^{2} \left(\int_{x=0}^{1} xy \, dx \right) dy = \int_{y=0}^{2} y \left[\frac{x^2}{2} \right]_0^1 dy = \int_{y=0}^{2} y \frac{1}{2} dy$$

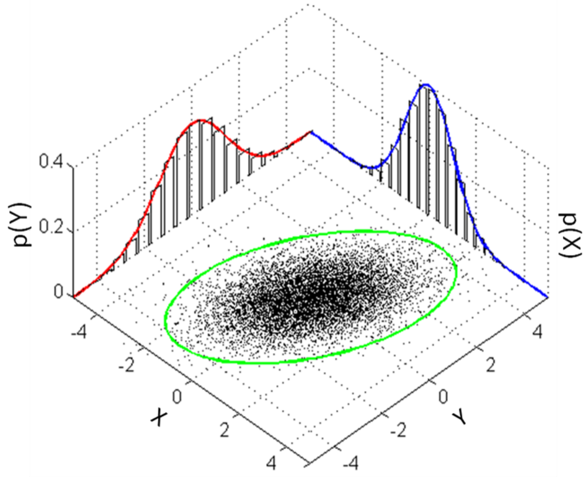
- Then, evaluate remaining (single) integral:

$$\int_{y=0}^{2} y \frac{1}{2} dy = \left[\frac{y^2}{4} \right]_0^2 = 1 - 0 = 1$$

Marginalization

Marginal probabilities give the distribution of **a subset of the variables** (often, just one) of a joint distribution.

Sum/integrate over the variables you don't care about.



$$p_X(a) = \sum_y p_{X,Y}(a, y)$$

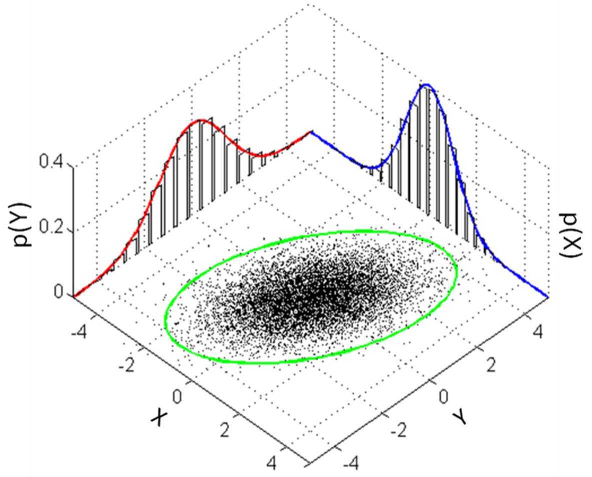
$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y)\ dy$$

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

Marginalization

Marginal probabilities give the distribution of **a subset of the variables** (often, just one) of a joint distribution.

Sum/integrate over the variables you don't care about.



$$p_X(a) = \sum_y P(X = a, Y = y)$$

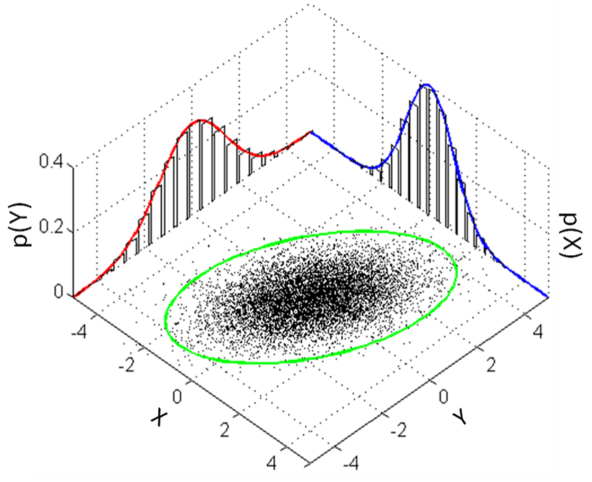
$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y) \, dy$$

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

Marginalization

Marginal probabilities give the distribution of **a subset of the variables** (often, just one) of a joint distribution.

Sum/integrate over the variables you don't care about.

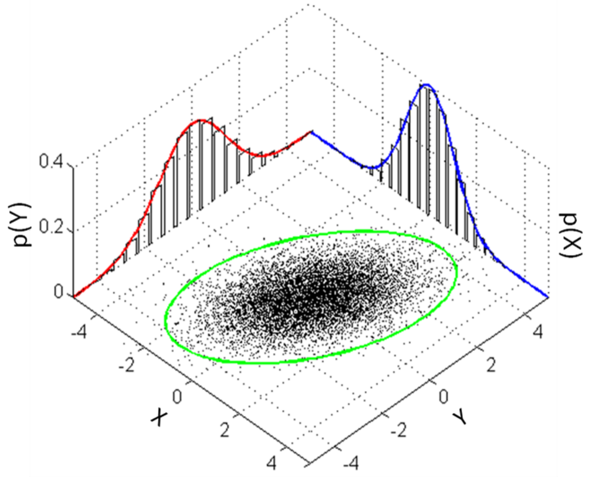


$$p_X(a) = \sum_y P(X = a, Y = y)$$

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

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

Marginalization



Marginal probabilities give the distribution of **a subset of the variables** (often, just one) of a joint distribution.

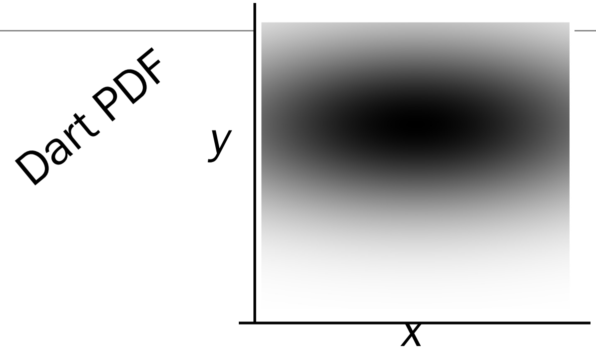
Sum/integrate over the variables you don't care about.

$$P(X = a) = \sum_y P(X = a, Y = y)$$

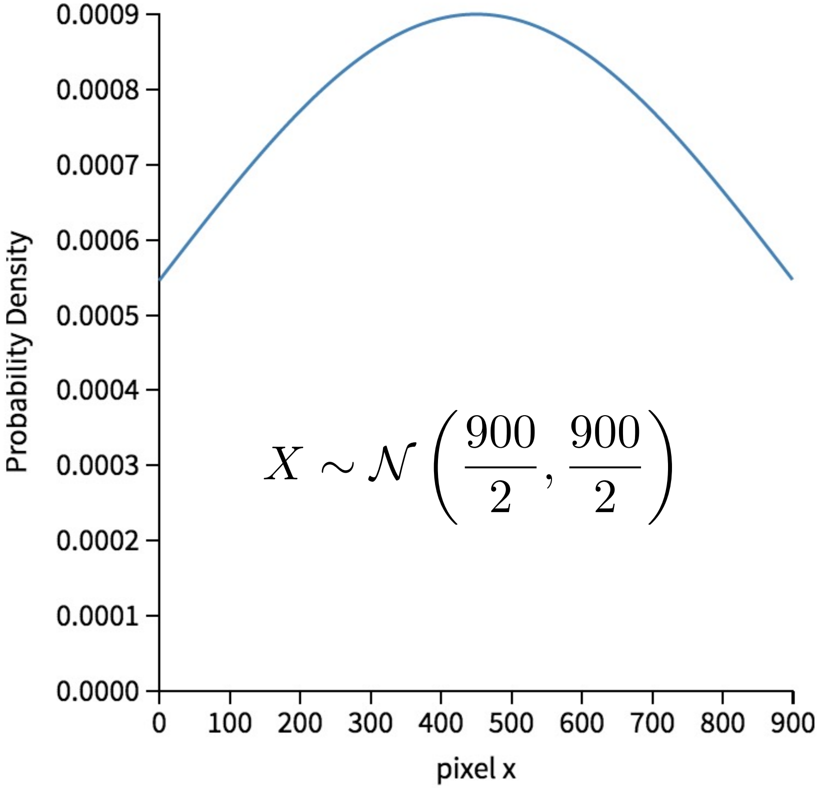
$$f(X = a) = \int_{-\infty}^{\infty} f(X = a, Y = y)$$

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

Darts!

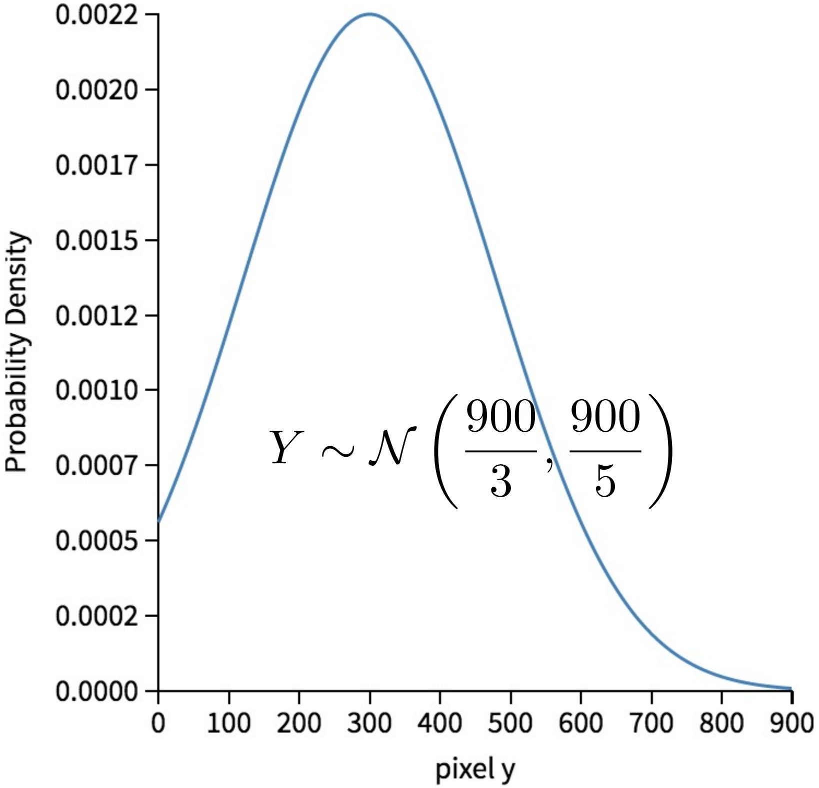


X-Pixel Marginal



$$X \sim \mathcal{N}\left(\frac{900}{2}, \frac{900}{2}\right)$$

Y-Pixel Marginal



$$Y \sim \mathcal{N}\left(\frac{900}{3}, \frac{900}{5}\right)$$

Joint Cumulative Density Function

Cumulative Density Function (CDF):

$$F_{X,Y}(a, b) = P(X < a, Y < b)$$

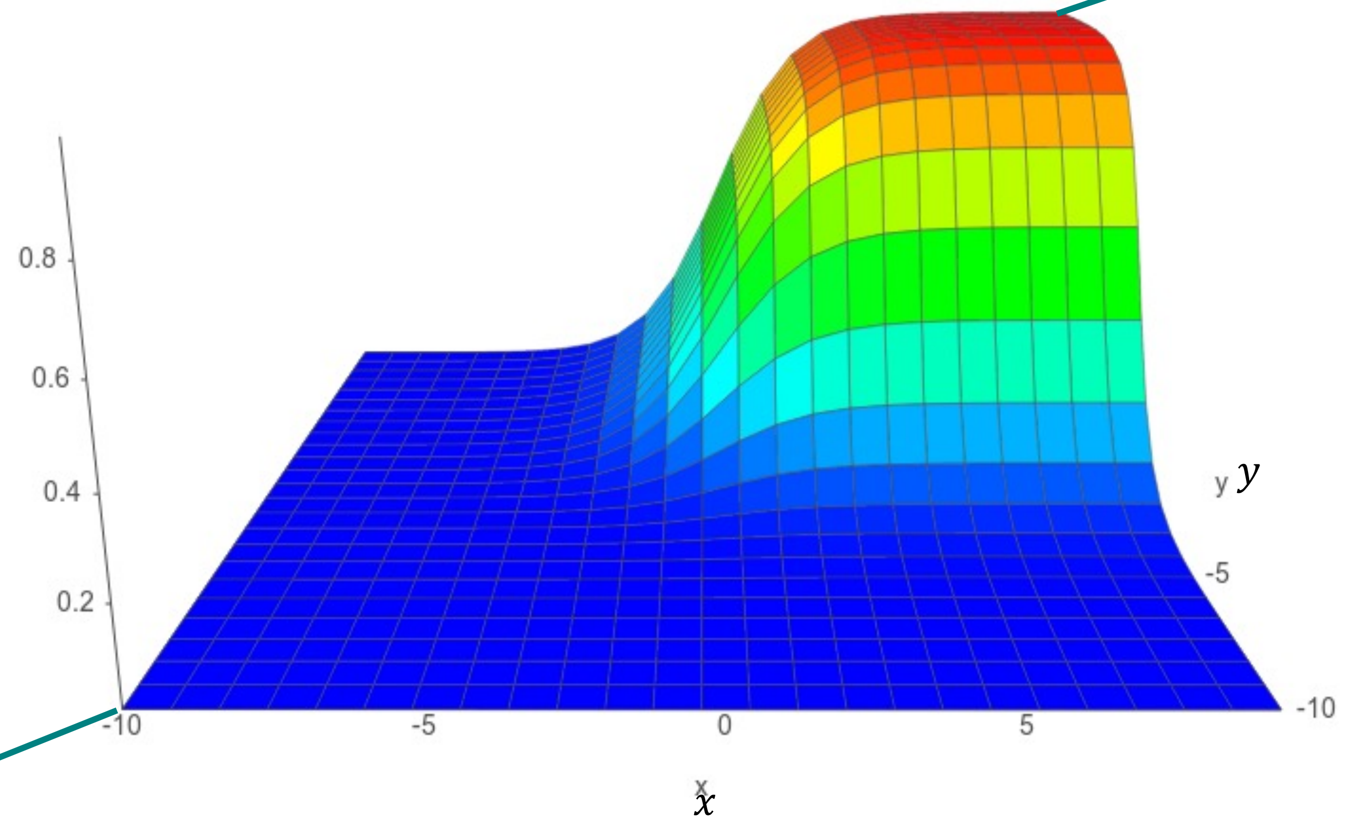
$$F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx$$

$$f_{X,Y}(a,b) = \frac{\partial^2}{\partial a\, \partial b} F_{X,Y}(a,b)$$

Joint CDF

$$F_{X,Y}(a,b) = P(X < a, Y < b)$$

to 1 as
x → +∞,
y → +∞

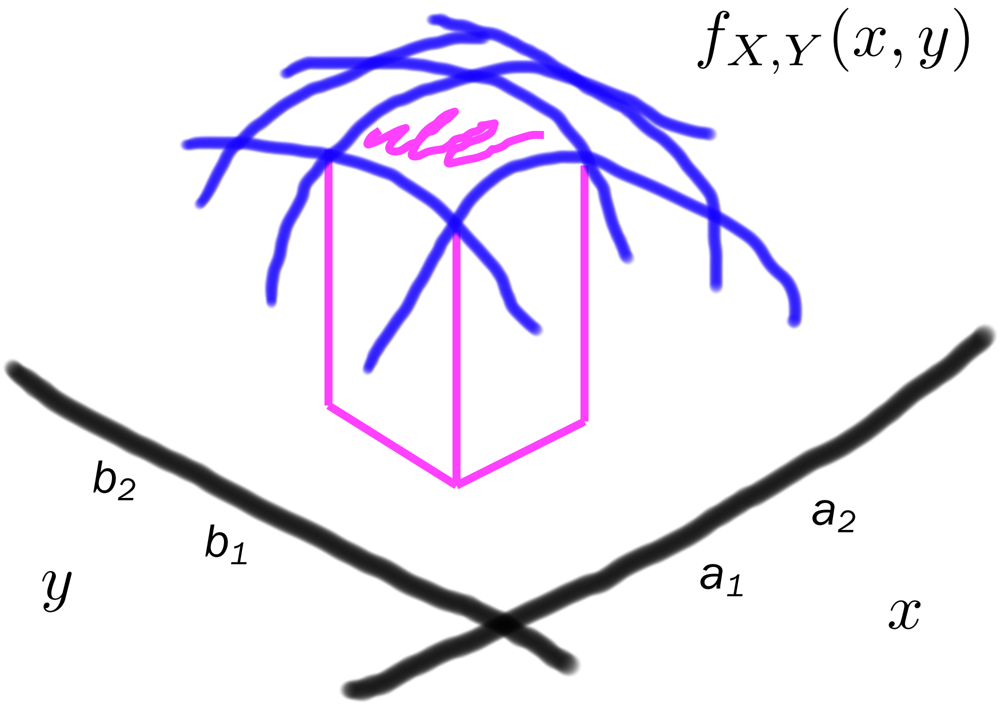


to 0 as
x → -∞,
y → -∞

plot by Academo

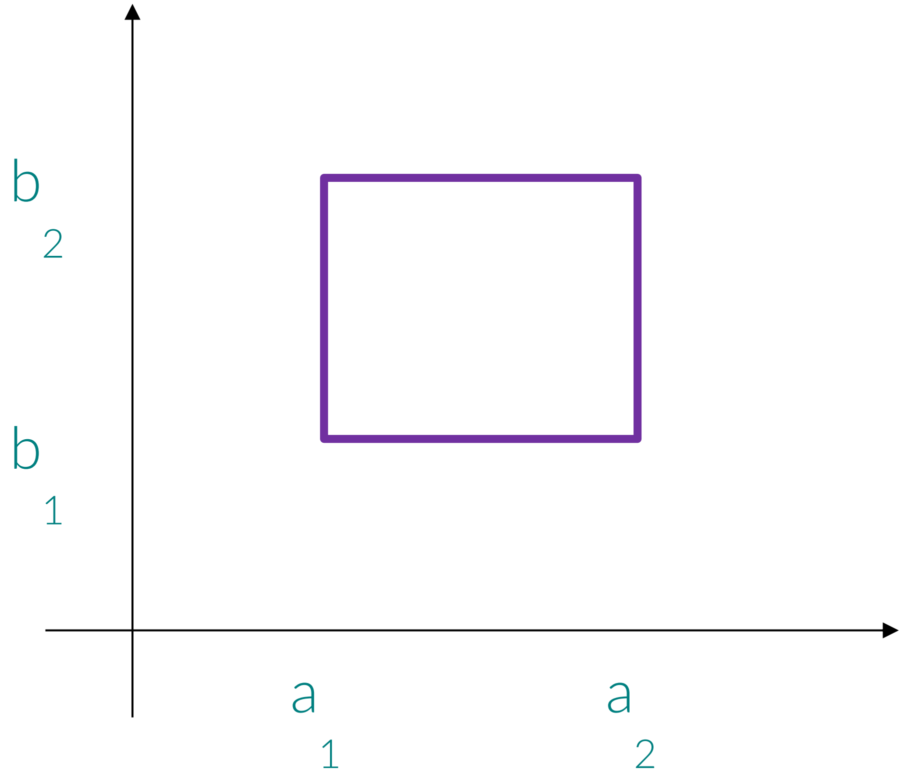
Jointly Continuous

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x,y)\, dy\, dx$$



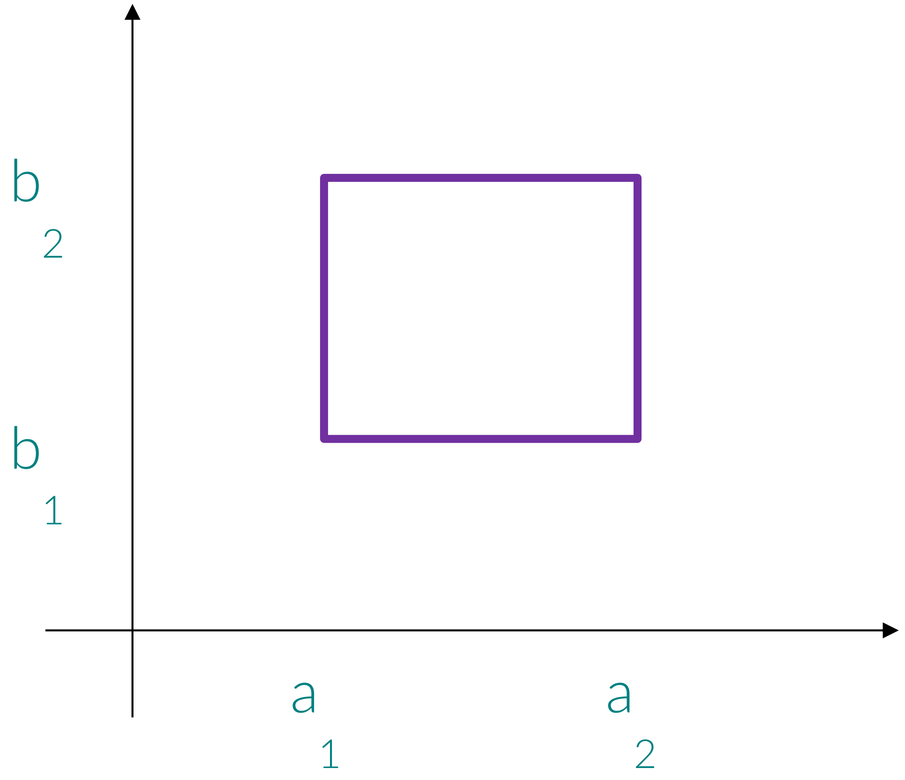
Probabilities from Joint CDF

$$P(a_1 < X \leq a_{2,} b_1 < Y \leq b_2)$$



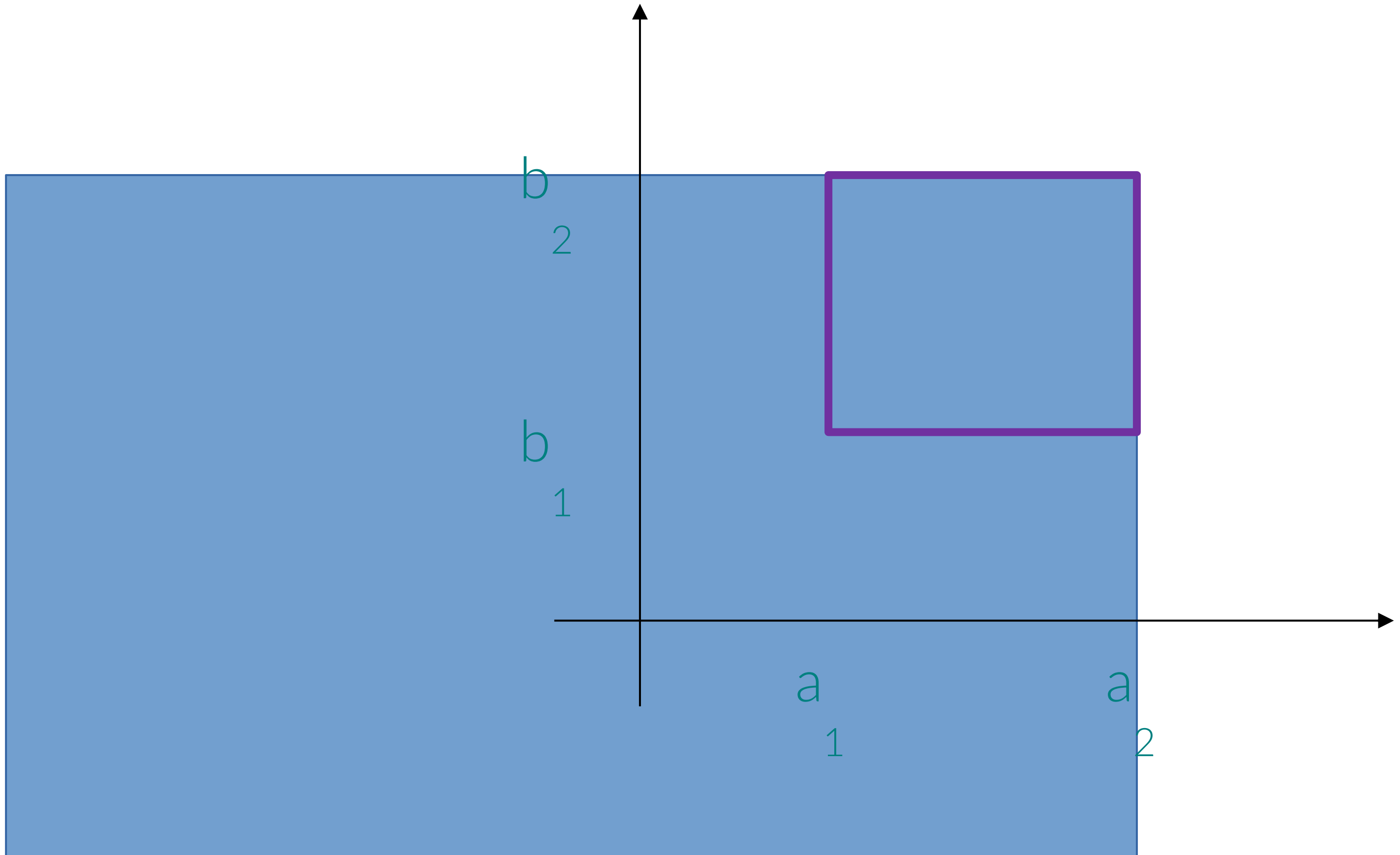
Probabilities from Joint CDF

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$



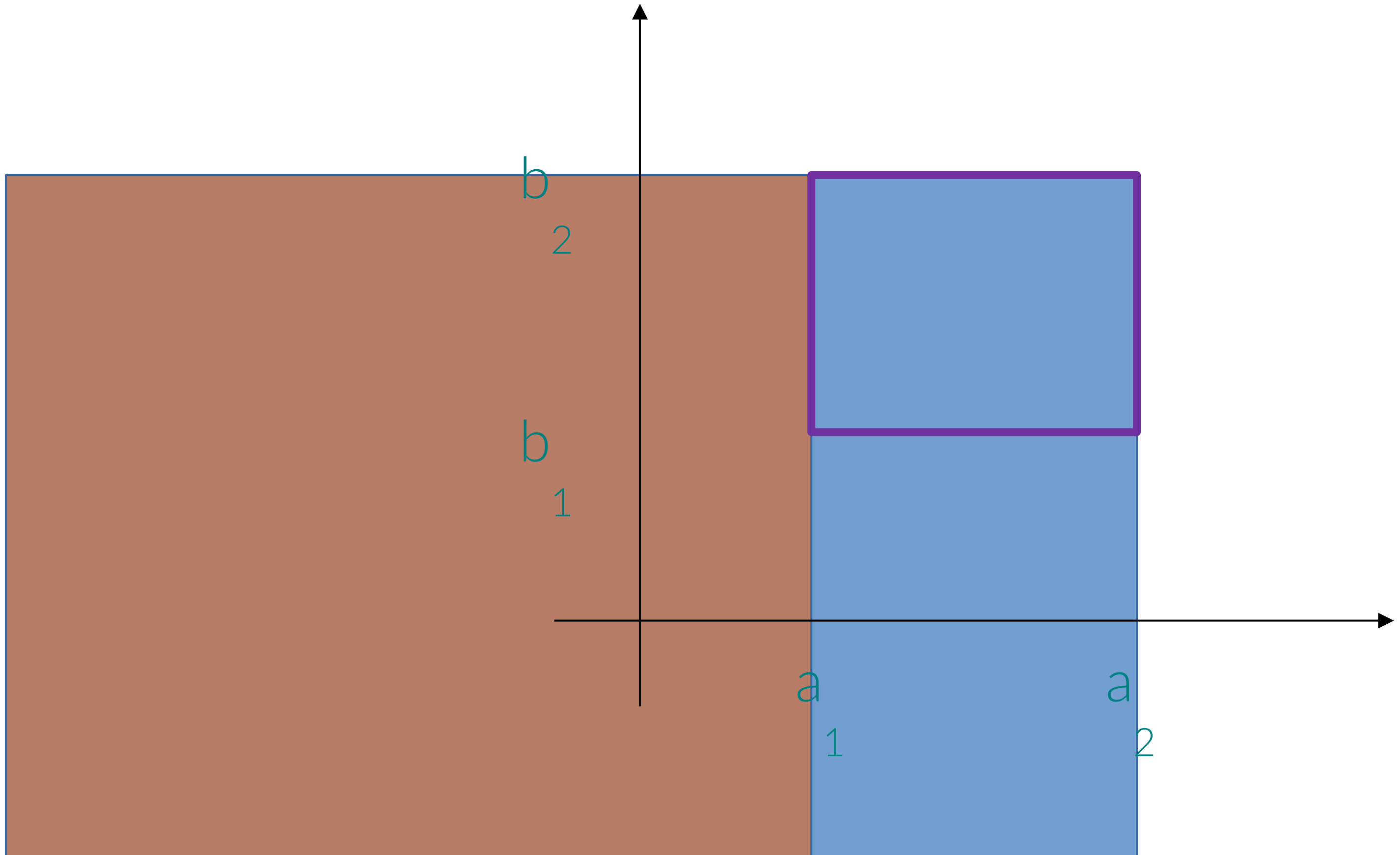
Probabilities from Joint CDF

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$



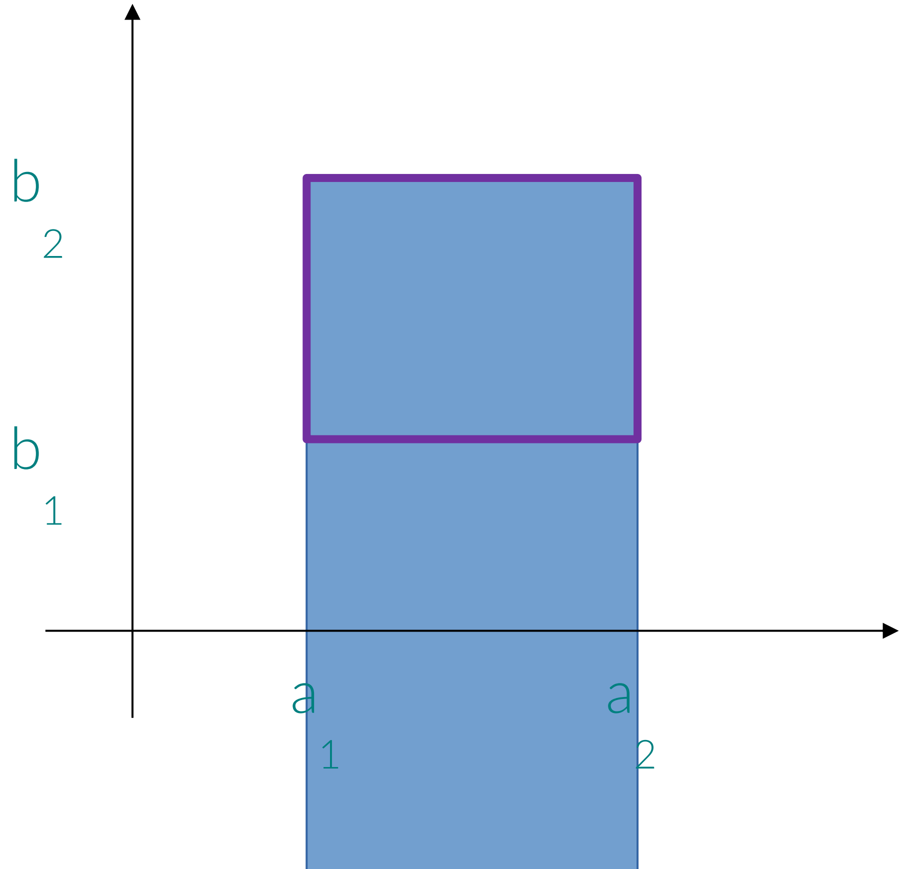
Probabilities from Joint CDF

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$



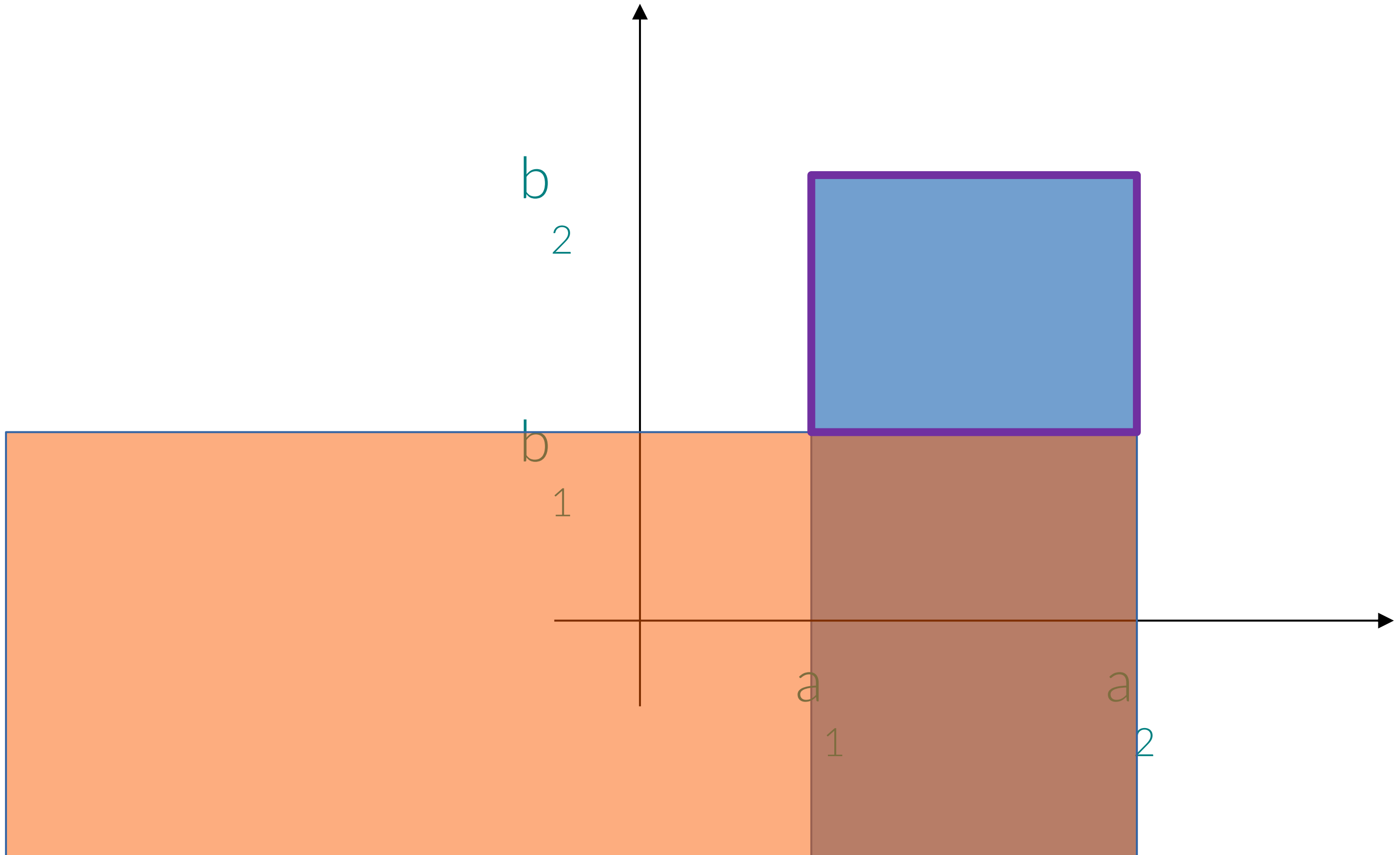
Probabilities from Joint CDF

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$
$$-F_{X,Y}(a_1, b_2)$$



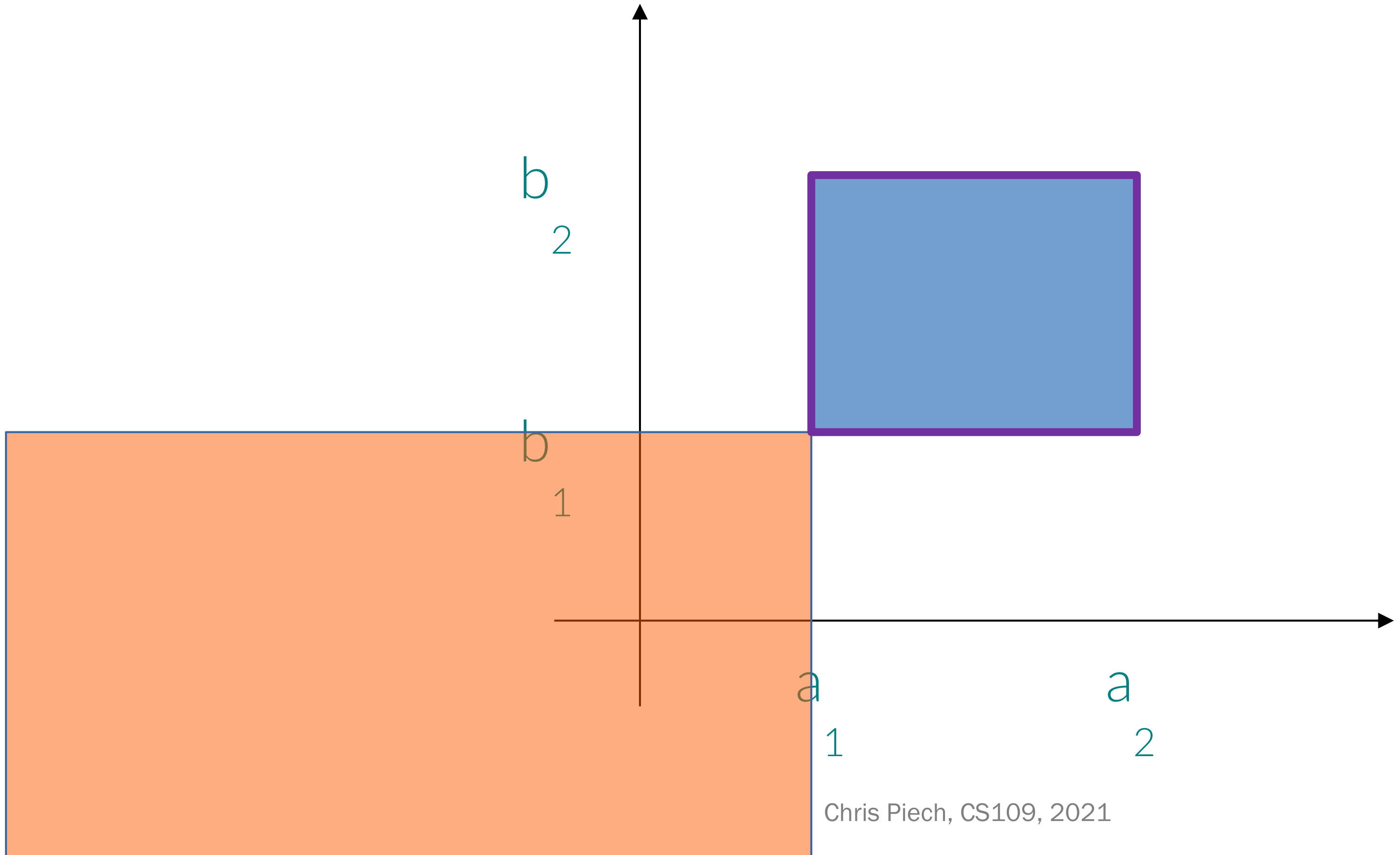
Probabilities from Joint CDF

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$
$$-F_{X,Y}(a_1, b_2)$$
$$-F_{X,Y}(a_2, b_1)$$



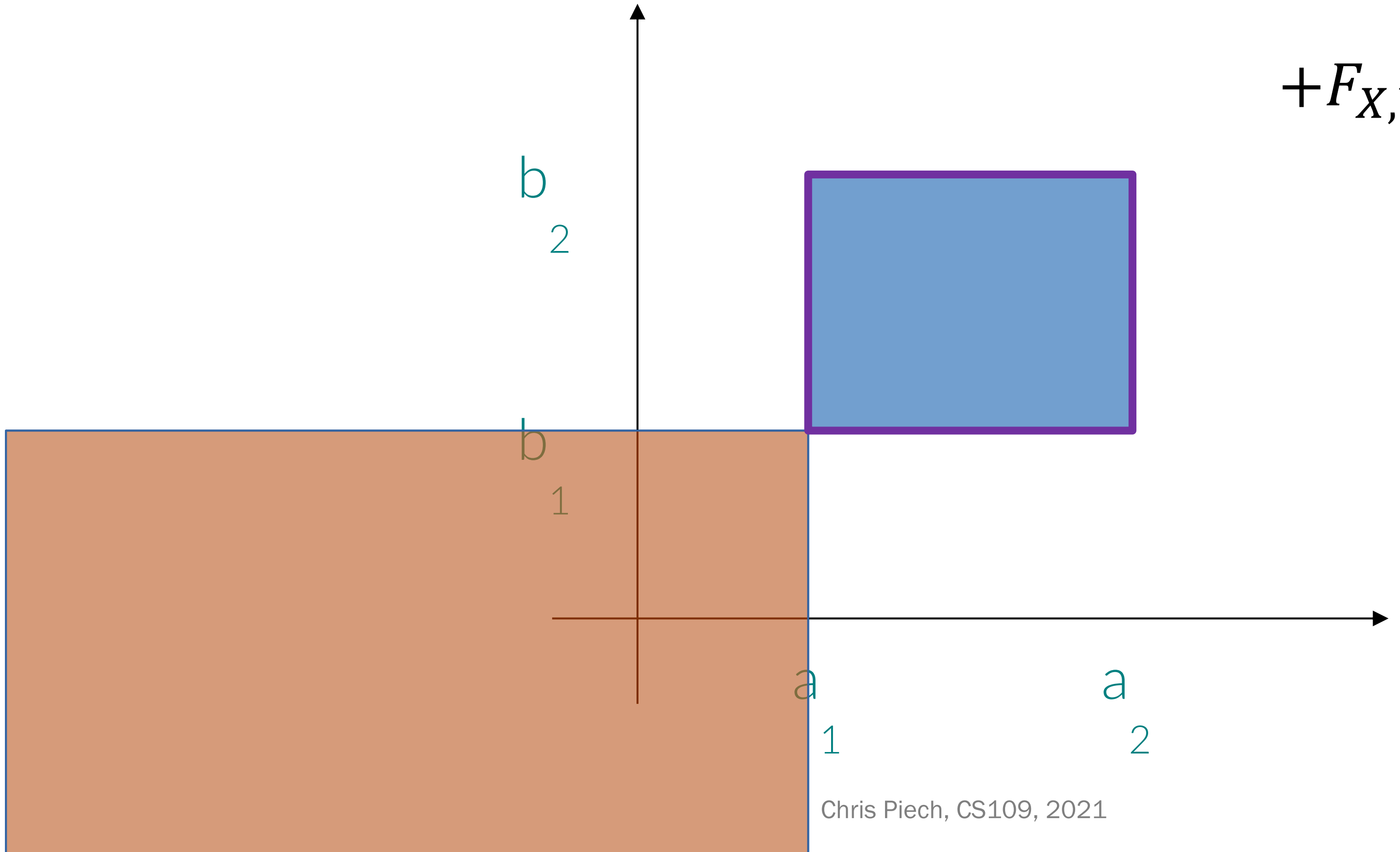
Probabilities from Joint CDF

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$
$$-F_{X,Y}(a_1, b_2)$$
$$-F_{X,Y}(a_2, b_1)$$



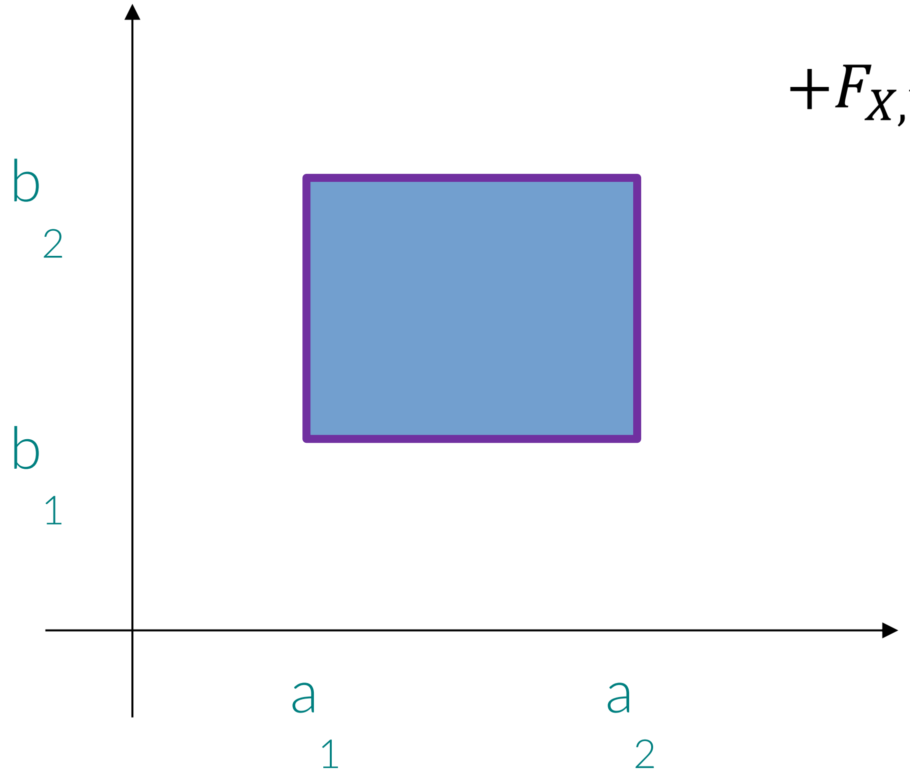
Probabilities from Joint CDF

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$
$$-F_{X,Y}(a_1, b_2)$$
$$-F_{X,Y}(a_2, b_1)$$
$$+F_{X,Y}(a_1, b_1)$$



Probabilities from Joint CDF

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$
$$-F_{X,Y}(a_1, b_2)$$
$$-F_{X,Y}(a_2, b_1)$$
$$+F_{X,Y}(a_1, b_1)$$



Probability for Instagram!

Gaussian Blur

In image processing, a Gaussian blur is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise.

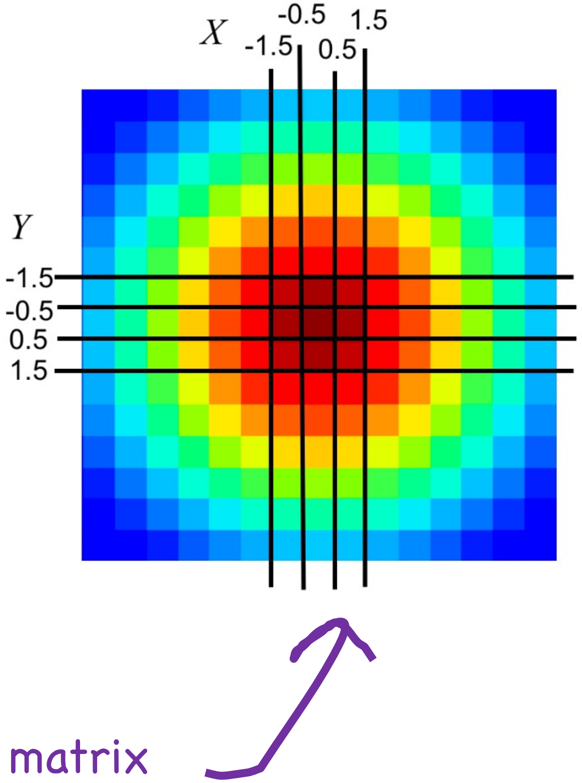
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

Joint PDF

$$f_{X,Y}(x,y) = \frac{1}{2\pi \cdot 3^2} e^{-\frac{x^2+y^2}{2\cdot 3^2}}$$

Joint CDF

$$F_{X,Y}(x,y) = \Phi\left(\frac{x}{3}\right) \cdot \Phi\left(\frac{y}{3}\right)$$



Used to generate this weight matrix

Gaussian Blur

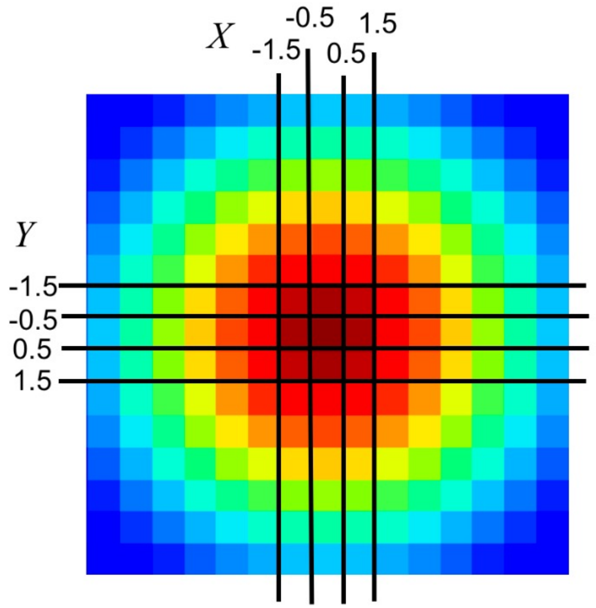
Joint PDF

$$f_{X,Y}(x,y) = \frac{1}{2\pi \cdot 3^2} e^{-\frac{x^2+y^2}{2\cdot 3^2}}$$

Joint CDF

$$F_{X,Y}(x,y) = \Phi\left(\frac{x}{3}\right) \cdot \Phi\left(\frac{y}{3}\right)$$

Weight Matrix



Each pixel is given a weight equal to the probability that X and Y are both within the pixel bounds. The center pixel covers the area where

$-0.5 \leq x \leq 0.5$ and $-0.5 \leq y \leq 0.5$

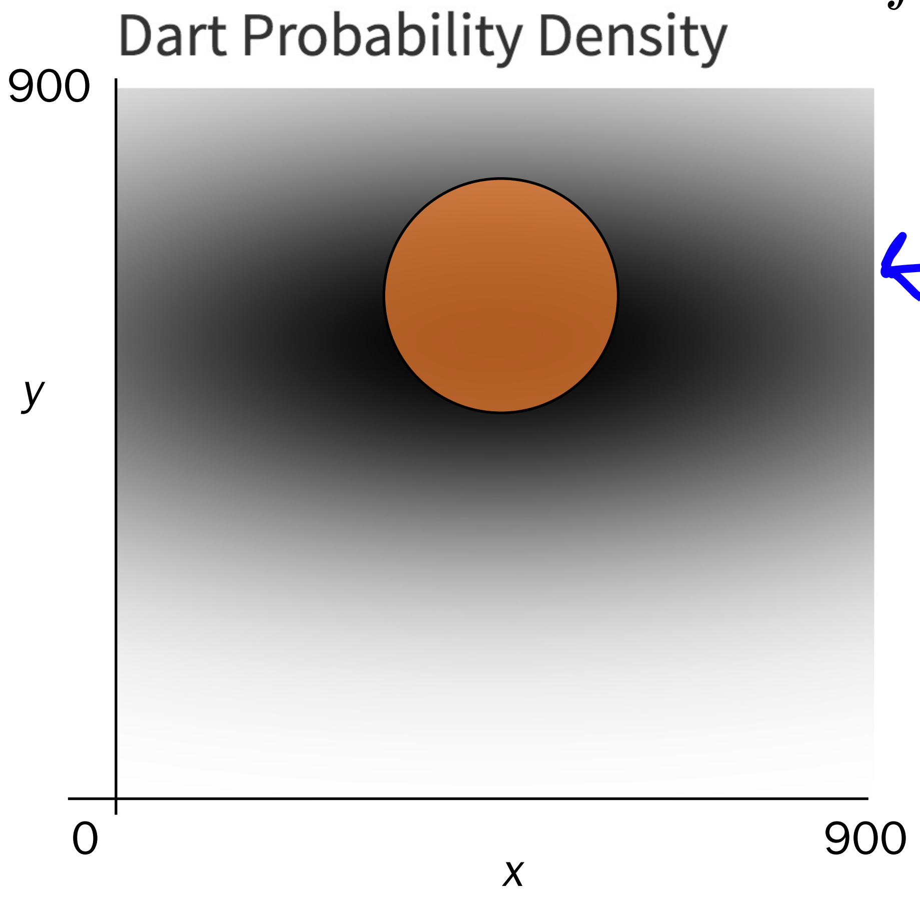
What is the weight of the center pixel?

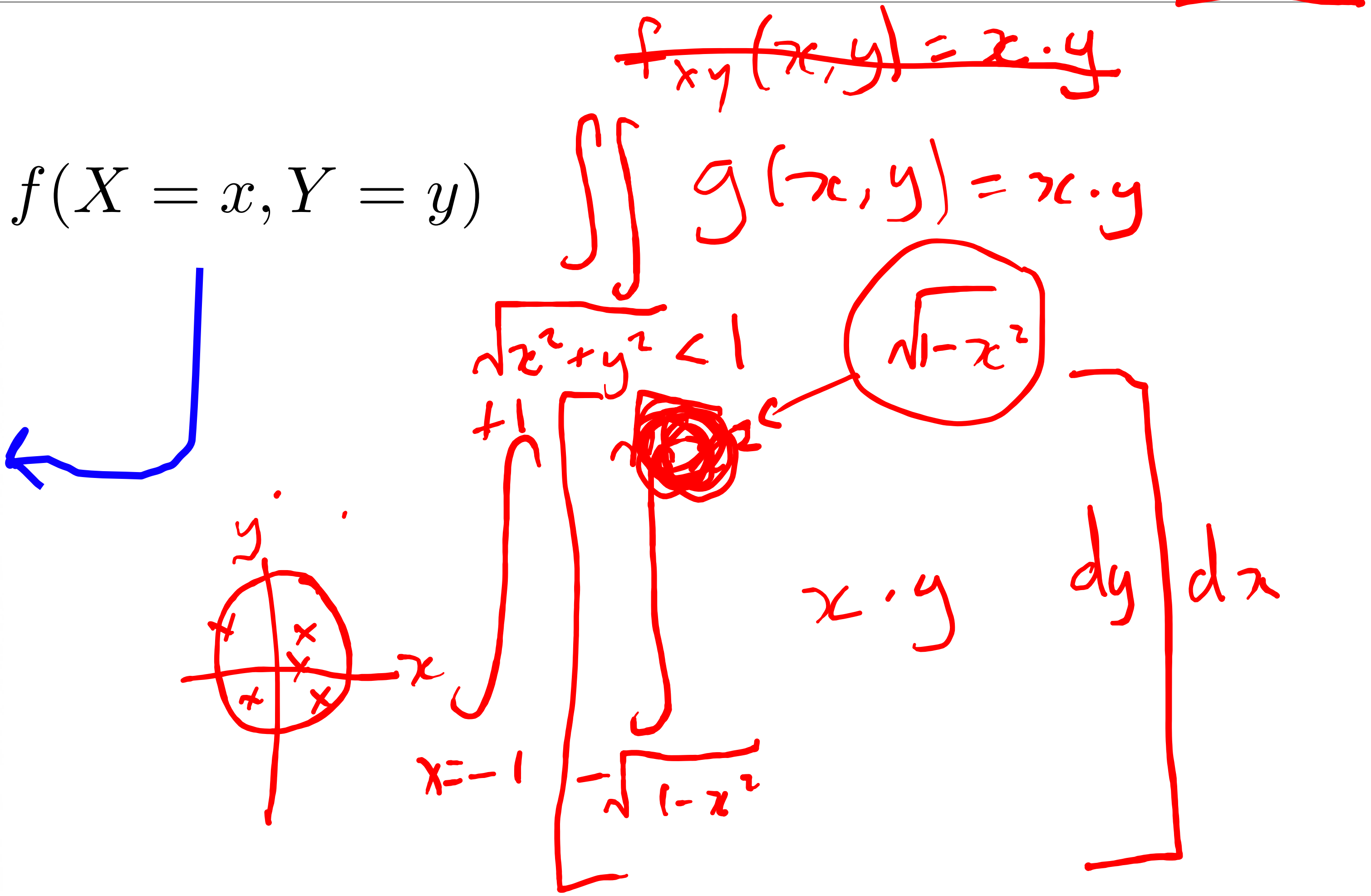
$$P(-0.5 < X < 0.5, -0.5 < Y < 0.5)$$
$$= P(X < 0.5, Y < 0.5) - P(X < 0.5, Y < -0.5)$$
$$- P(X < -0.5, Y < 0.5) + P(X < -0.5, Y < -0.5)$$
$$= \phi\left(\frac{0.5}{3}\right) \cdot \phi\left(\frac{0.5}{3}\right) - 2\phi\left(\frac{0.5}{3}\right) \cdot \phi\left(\frac{-0.5}{3}\right)$$
$$+ \phi\left(\frac{-0.5}{3}\right) \cdot \phi\left(\frac{-0.5}{3}\right)$$
$$= 0.5662^2 - 2 \cdot 0.5662 \cdot 0.4338 + 0.4338^2 = 0.206$$

Special case of a Multivariate Gaussian. Each gaussian is independent!

How do you integrate under a circle?

$$f(X = x, Y = y)$$





Pedagogic Pause

Joint Random Variables

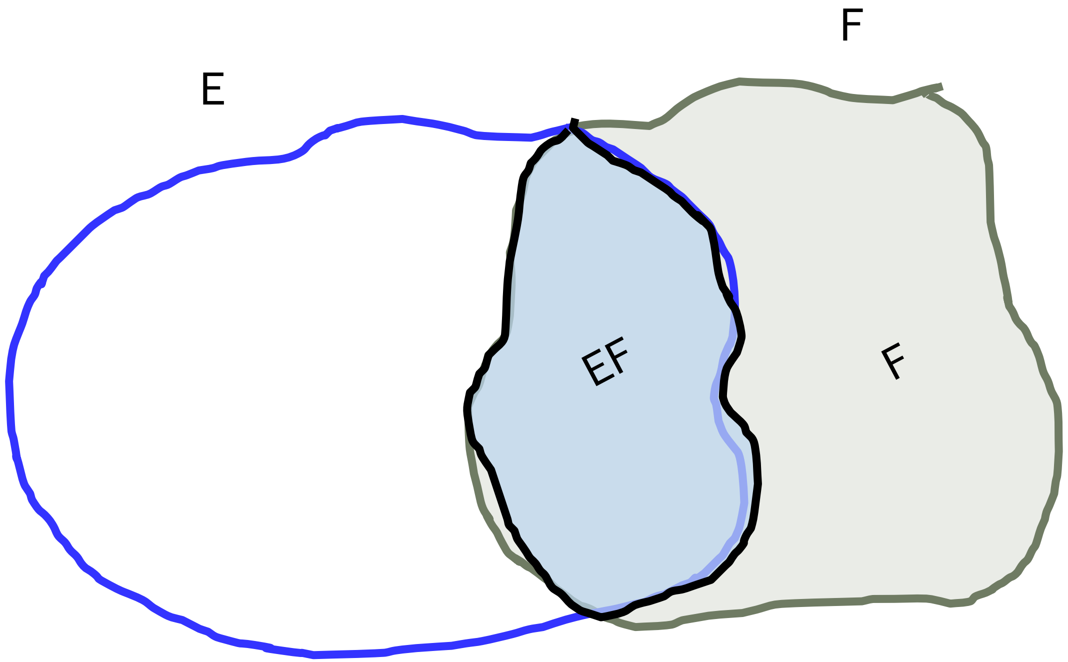
- [x] Use a joint table, density function or CDF to solve probability question
- [ ] Think about **conditional** probabilities with joint variables (which might be continuous)
- [ ] Use and find **independence** of random variables
- [ ] Use and find **expectation** of multiple random variables
- [ ] What happens when you **add** random variables?

Conditionals with multiple variables

Conditionals with Events

Recall that for *events* E and F:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where} \quad P(F) > 0$$



Discrete Conditional Distributions

Recall that for *events* E and F:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where} \quad P(F) > 0$$

Now, have X and Y as **discrete** random variables

- **Conditional PMF** of X given Y:

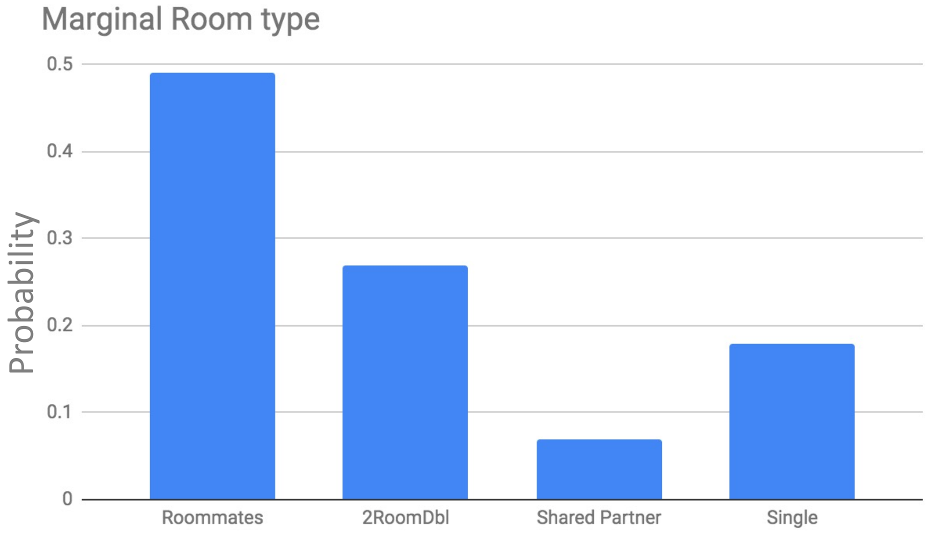
$$P_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

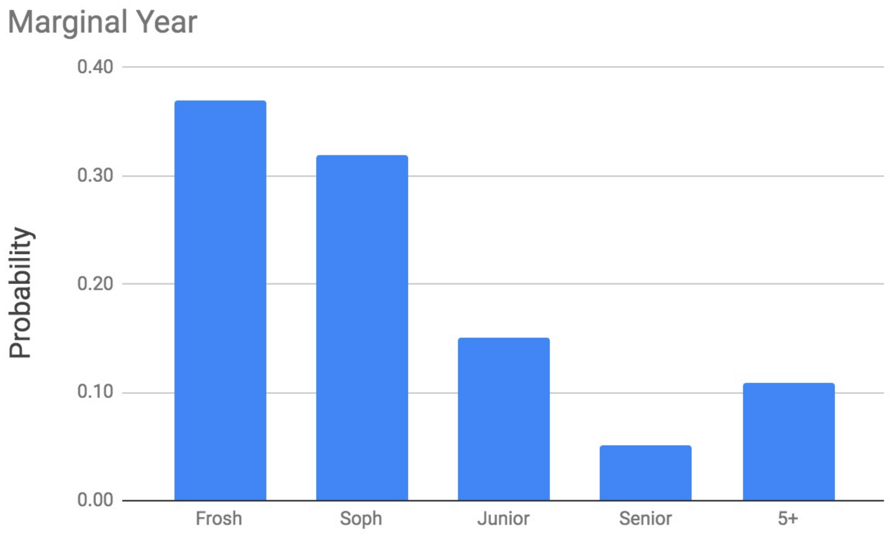
Different notations, same idea.

Joint Probability Table

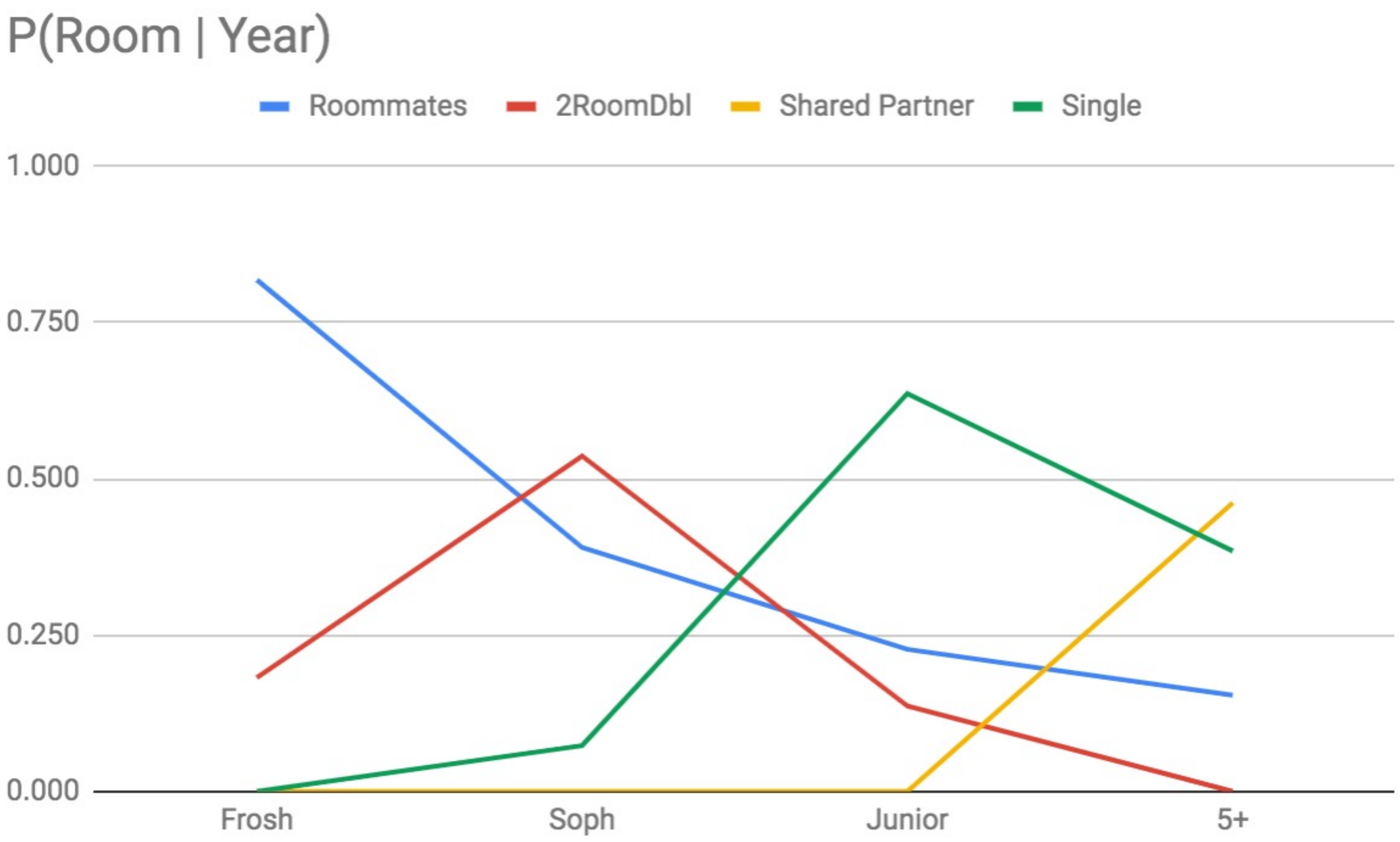
	Roommates	2RoomDbl	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00

$$\frac{P(2D, F)}{P(F)}$$



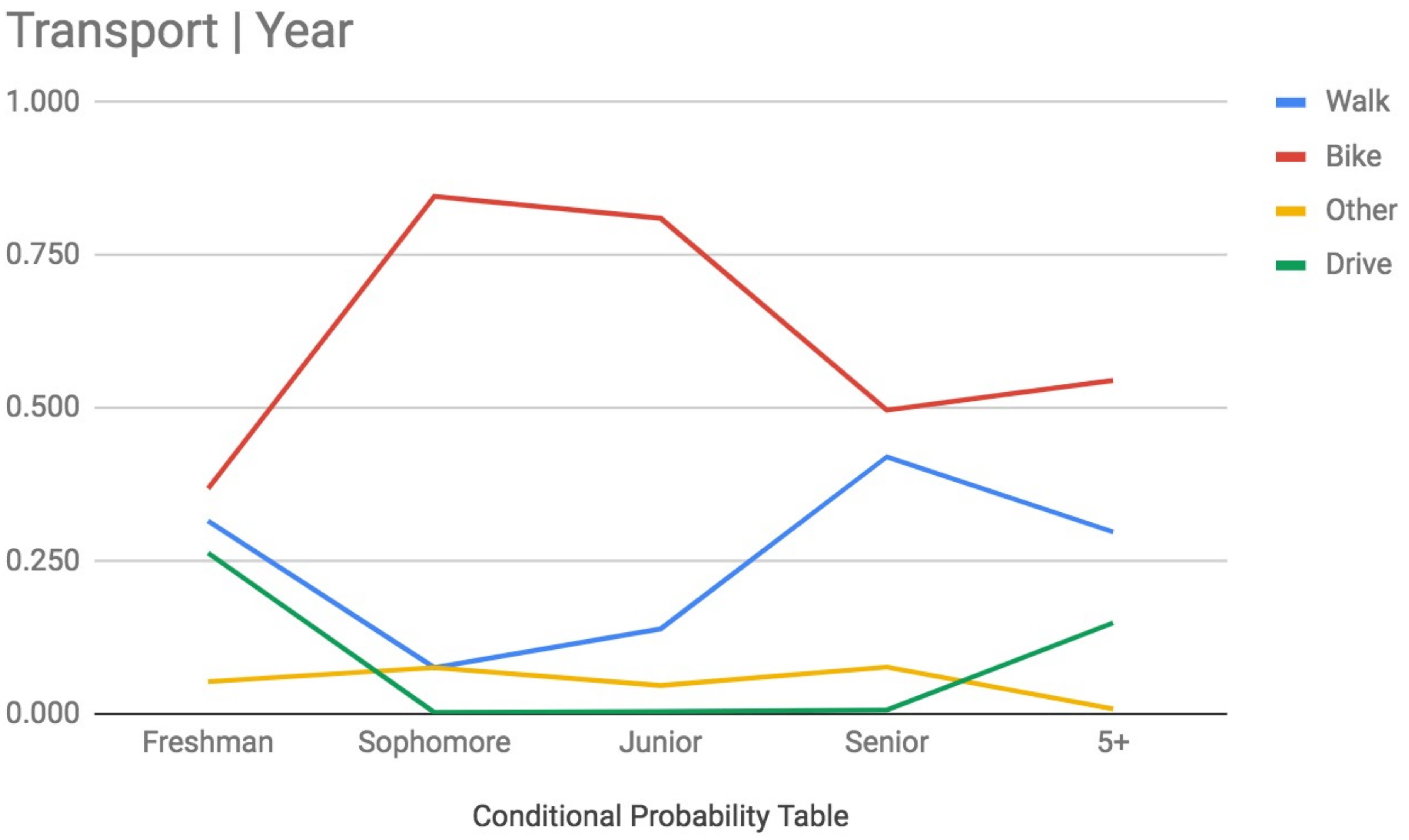


Room | Year

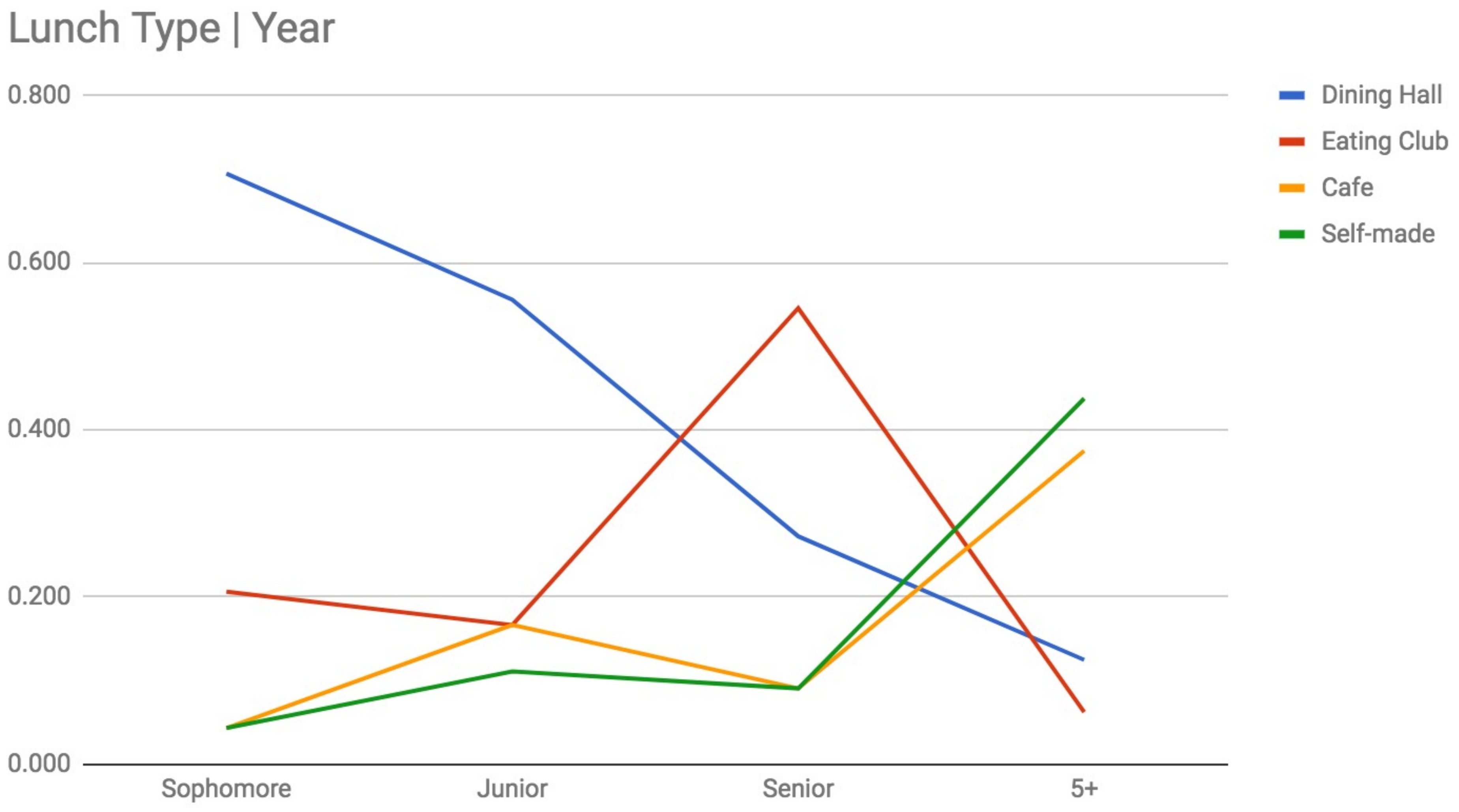


P(Room | Year)
Roommates
2RoomDbl
Shared Partner
Single
1.000
0.750
0.500
0.250
0.000
Frosh
Soph
Junior
5+

Transport | Year

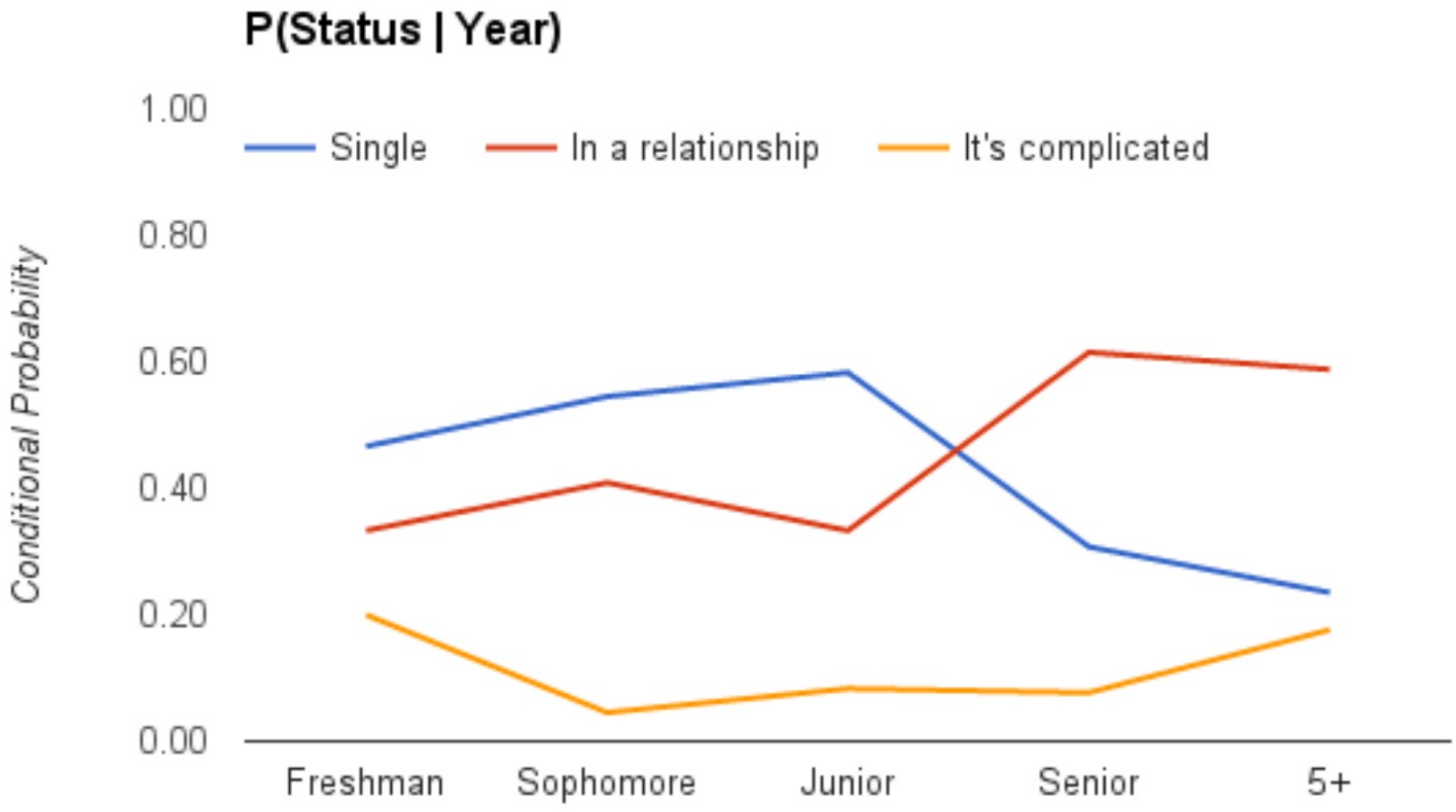


Lunch | Year



Lunch Type | Year
0.800
0.600
0.400
0.200
0.000
Sophomore
Junior
Senior
5+
Dining Hall
Eating Club
Cafe
Self-made

Relationship Status | Year



Number or Function?

$$P(X = 2 | Y = 5)$$

Number

Number or Function?

$$P(X = 2 | Y = y)$$

Function

(or 1D table)

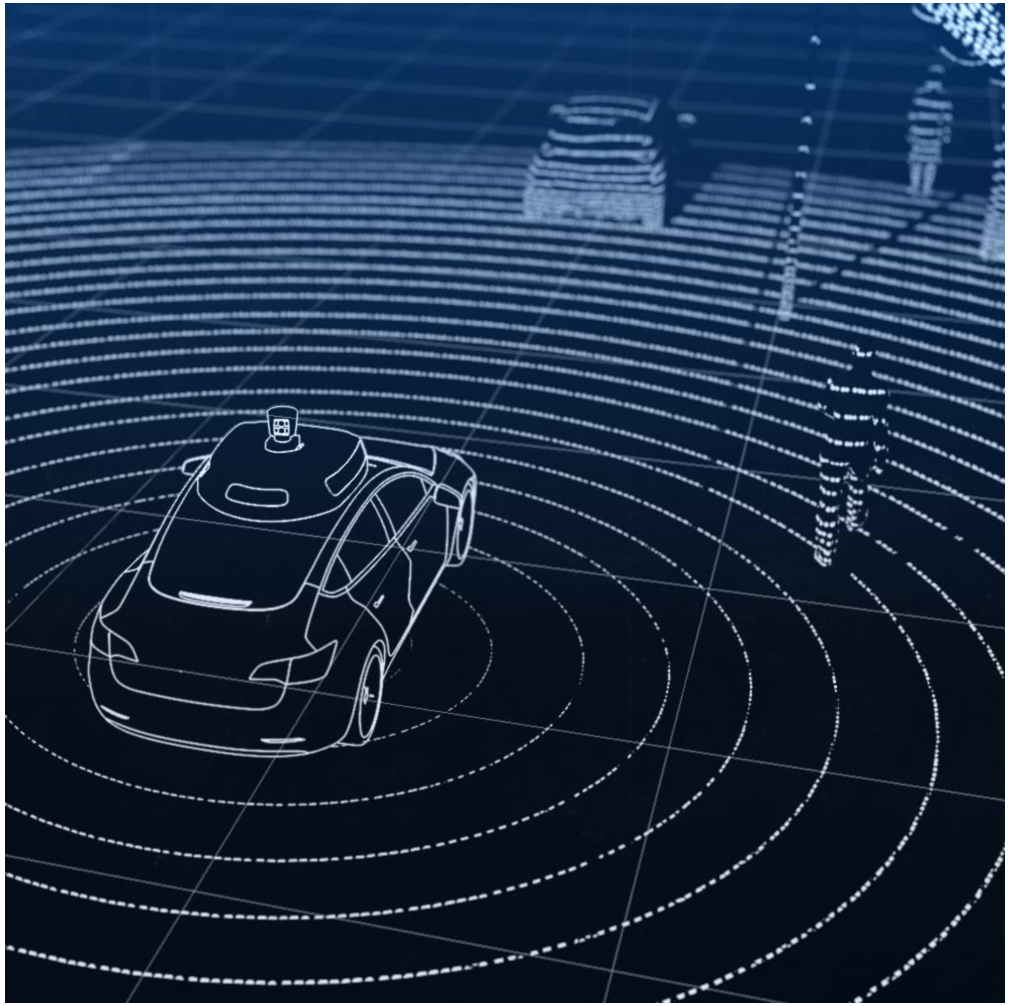
Number or Function?

$$P(X = x | Y = y)$$

2D Function

(or 2D table)

Conditional Probability with Random Variables



Have a Great Weekend