Conditional Joint Distributions

Noah Arthurs
CS109, Stanford University

Announcements

Midterm:

- Recorded review session on Monday 3:30-4:20 in Skilling (here). Please let me know what you want to see!
- Some practice problems use today's material, but you don't need to know it.
- Keep posting on Piazza!

PS4:

- You will have tools to solve almost all of it on Monday, rest on Wednesday.

Review

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

A Document is a Large *Multinomial*

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

Look Before You Leap

$$\frac{P(H|W)}{P(M|W)} = \frac{P(W|H)P(H)/P(W)}{P(W|M)P(M)/P(W)}$$

$$= \frac{P(W|H)}{P(W|M)}$$

$$= \frac{\binom{n}{c_1,\ldots,c_k}\prod_i h_i^{c_i}}{\binom{n}{c_1,\ldots,c_k}\prod_i m_i^{c_i}}$$

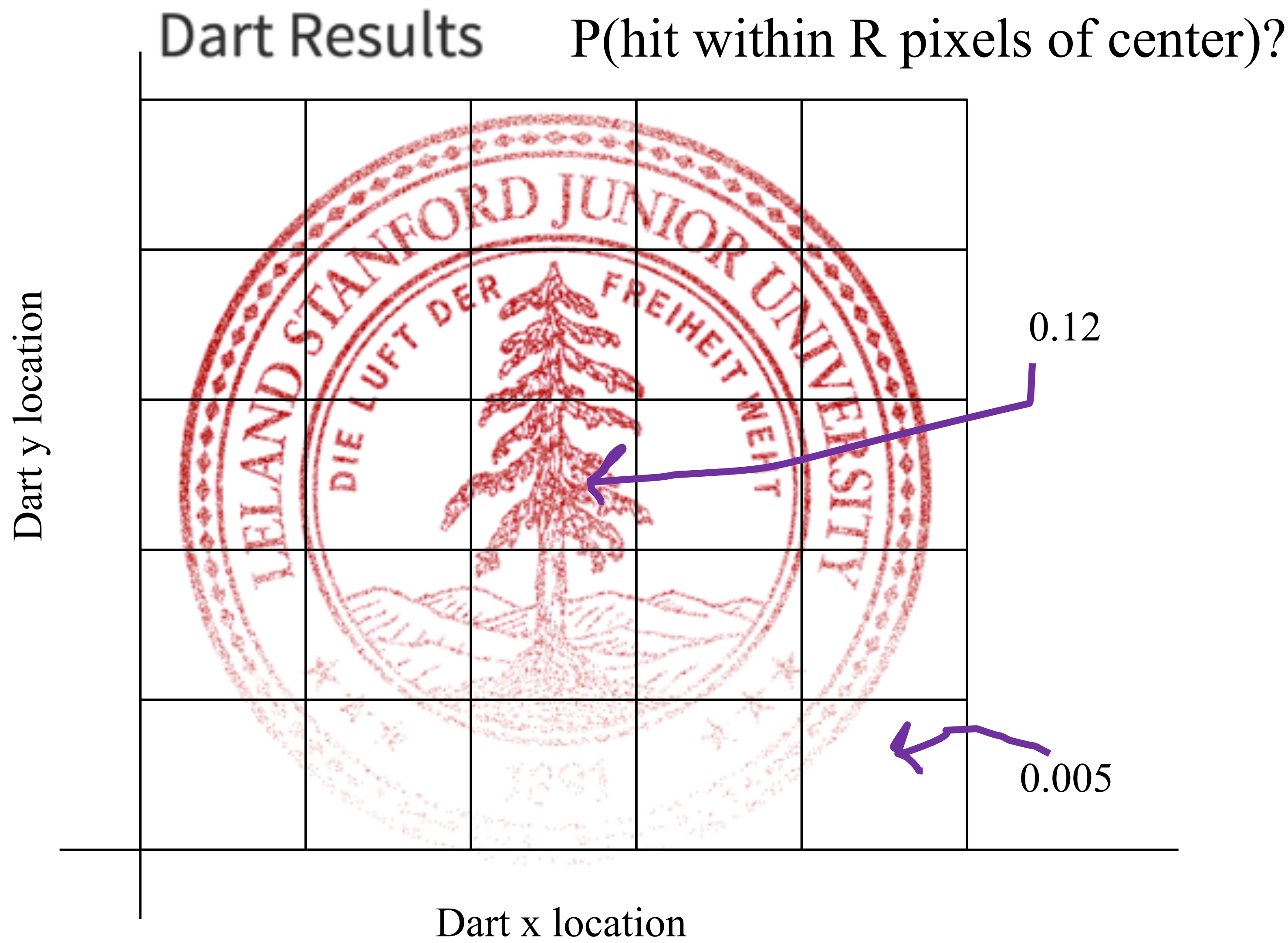
$$= \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}}$$

$$= exp\left(\sum_i c_i \log(h_i) - \sum_i c_i \log(m_i)\right)$$

Continuous Random Variables

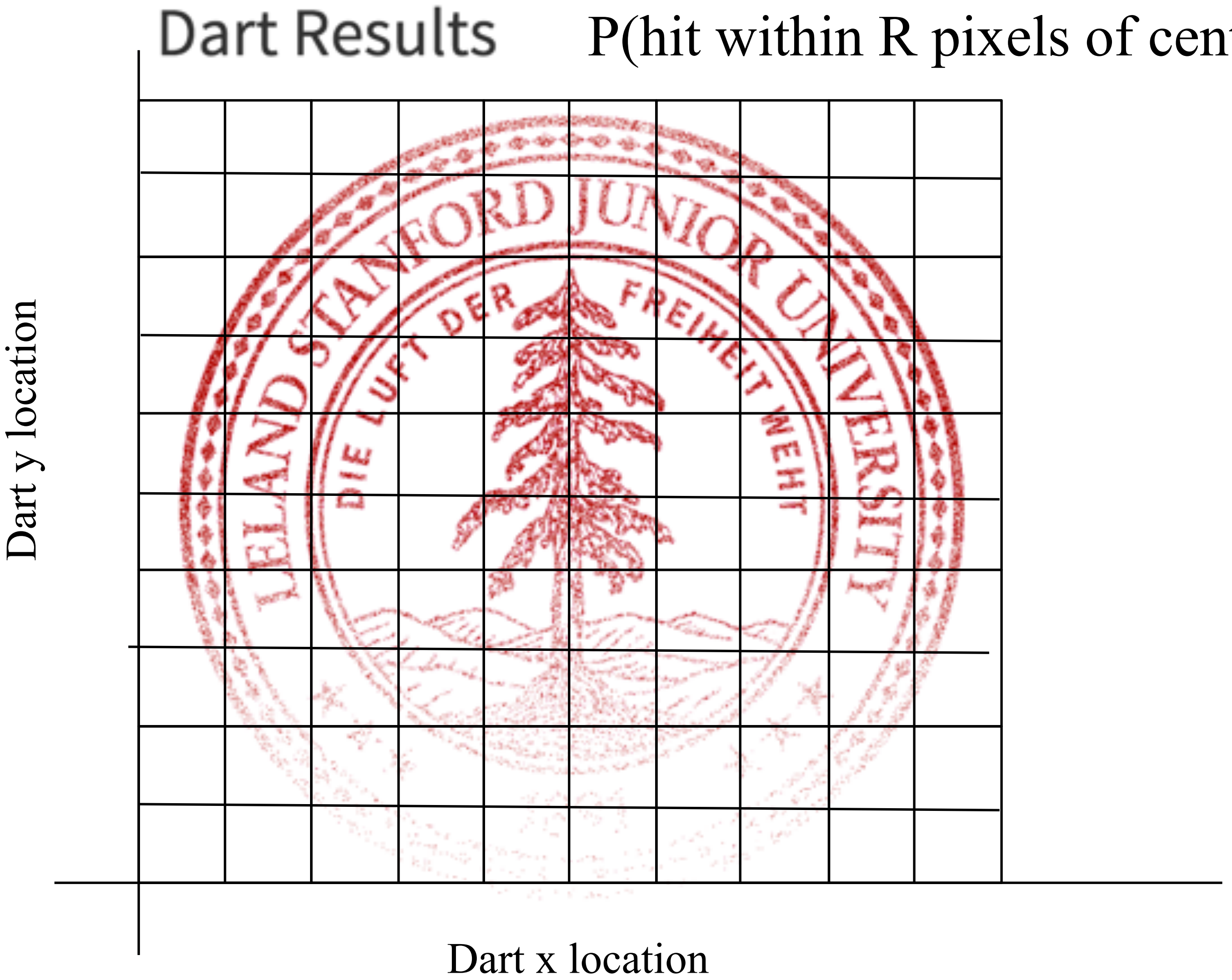
Joint Distributions

Joint Dart Distribution



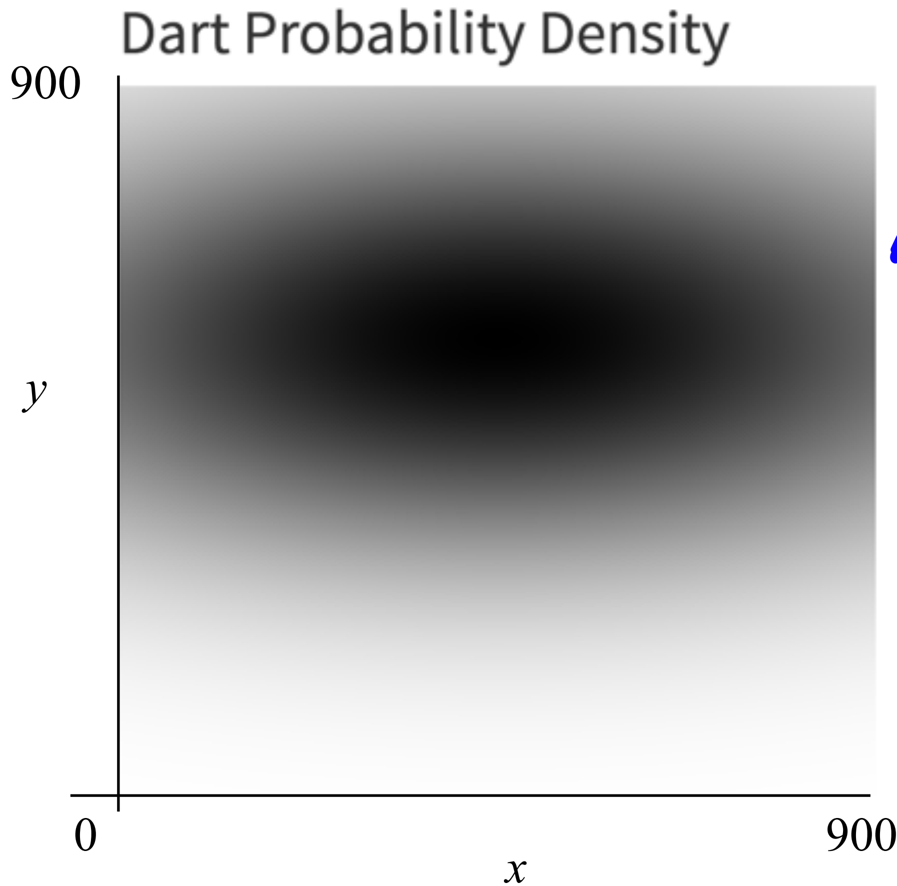
Joint Dart Distribution

P(hit within R pixels of center)?



Joint Dart Distribution

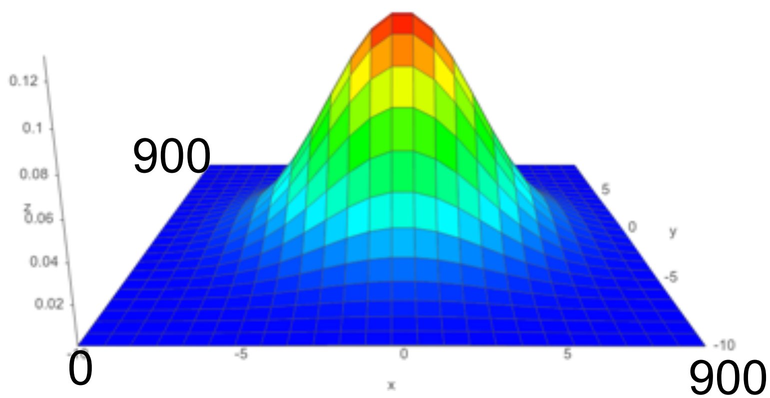
$$f(X = x, Y = y)$$

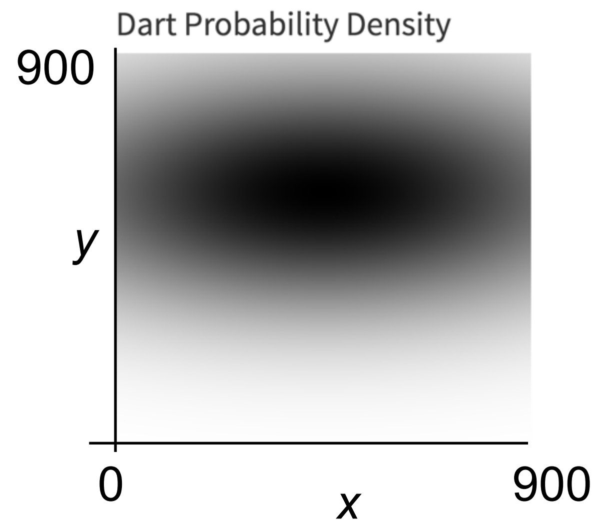


In the limit, as you break down continuous values into intestinally small buckets, you end up with multidimensional probability density

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.

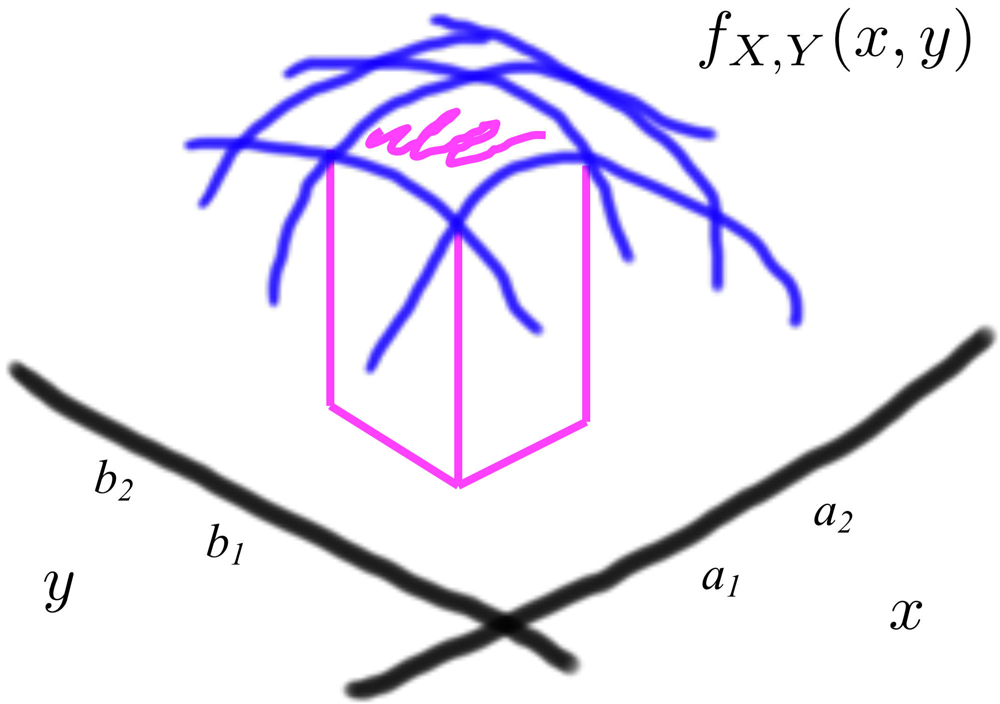




$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$

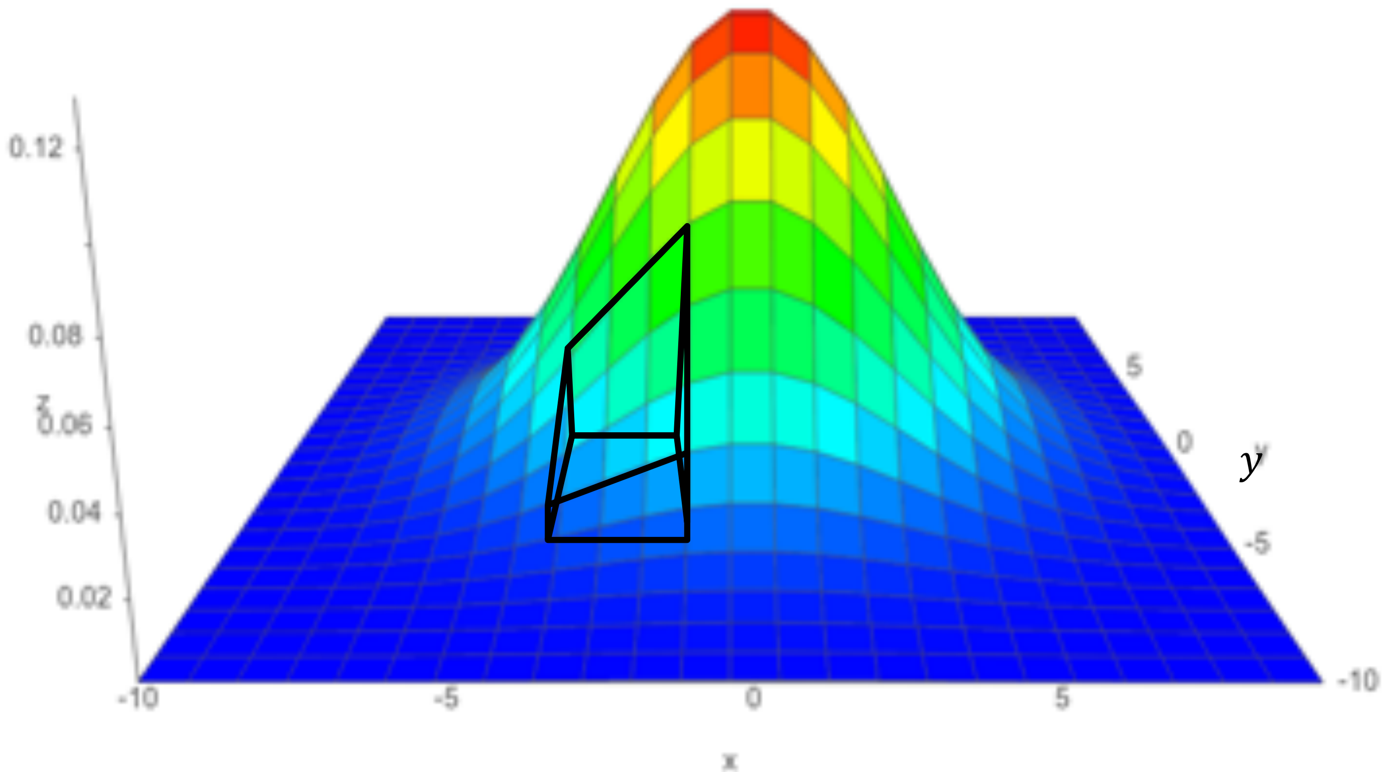
Joint Probability Density Funciton

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



Joint Probability Density Function

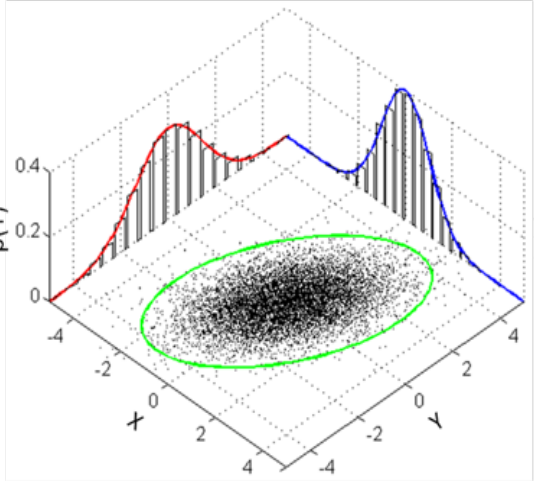
$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$



plot by Academo

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.

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

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

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

Joint Cumulative Density Function

Cumulative Density Function (CDF):

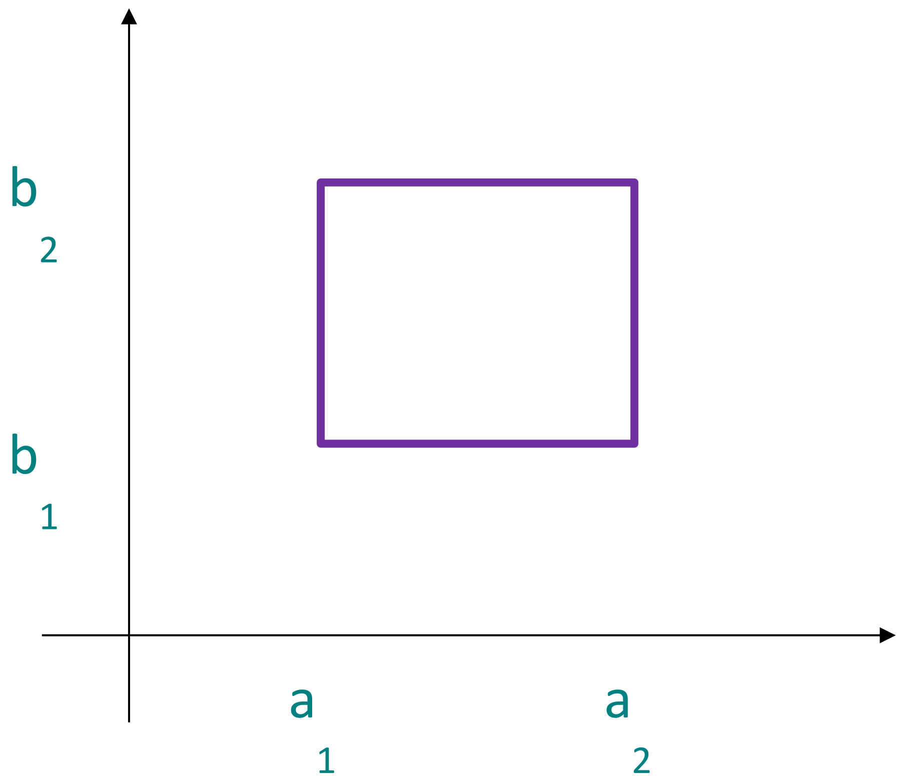
$$F_{X,Y}(a,b) = P(X < a, Y < b)$$

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

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

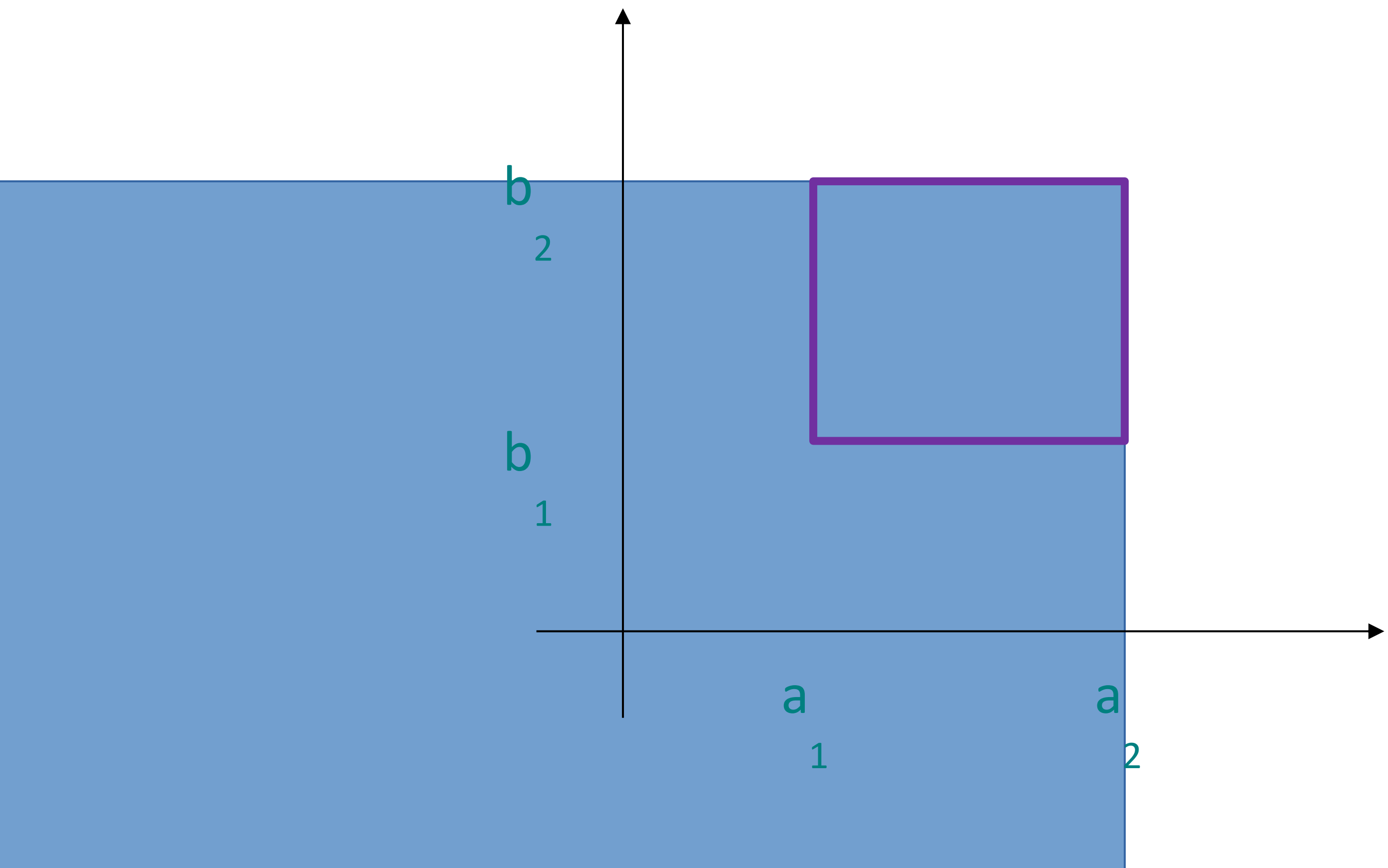
Probabilities from Joint CDF

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



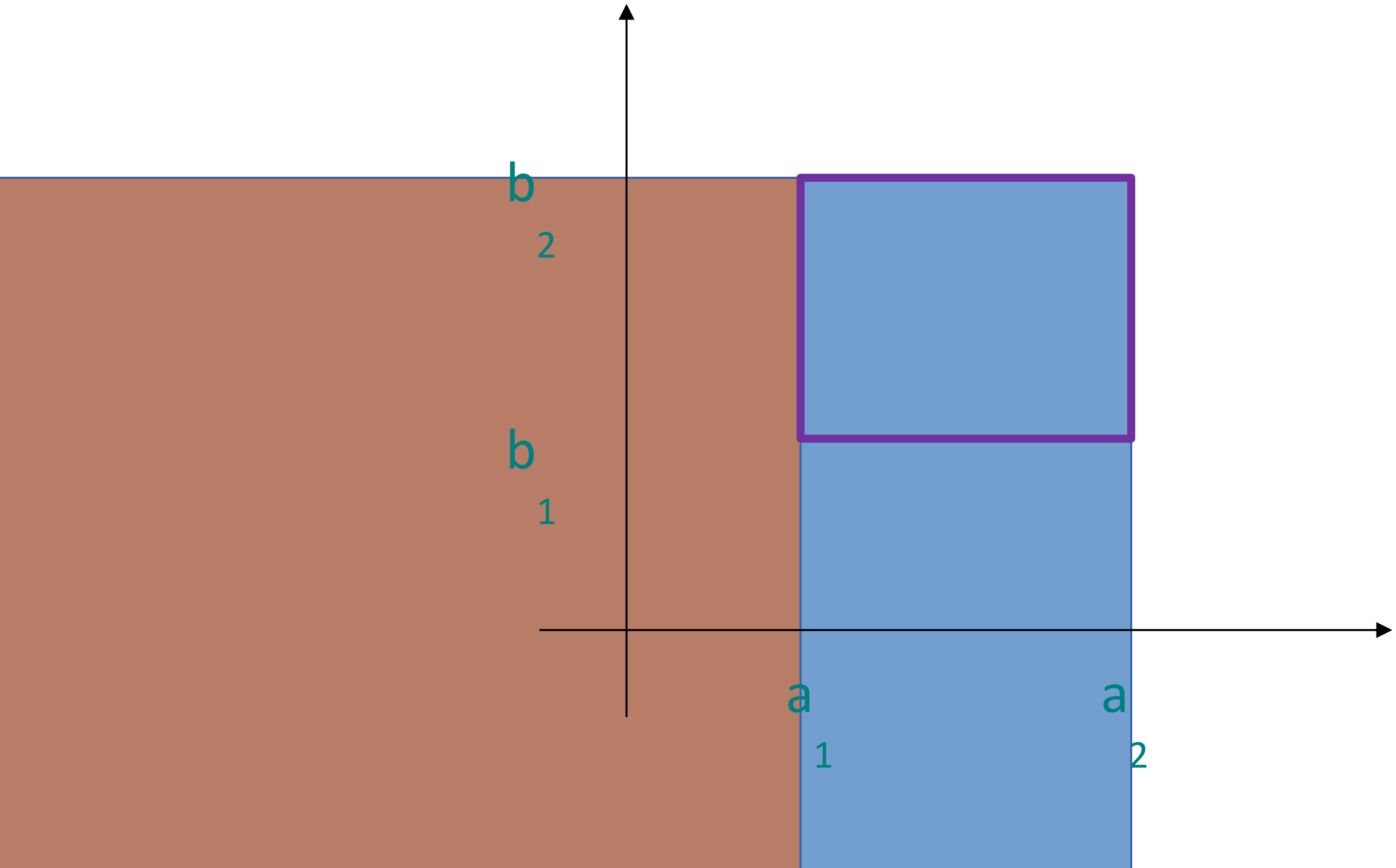
Probabilities from Joint CDF

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



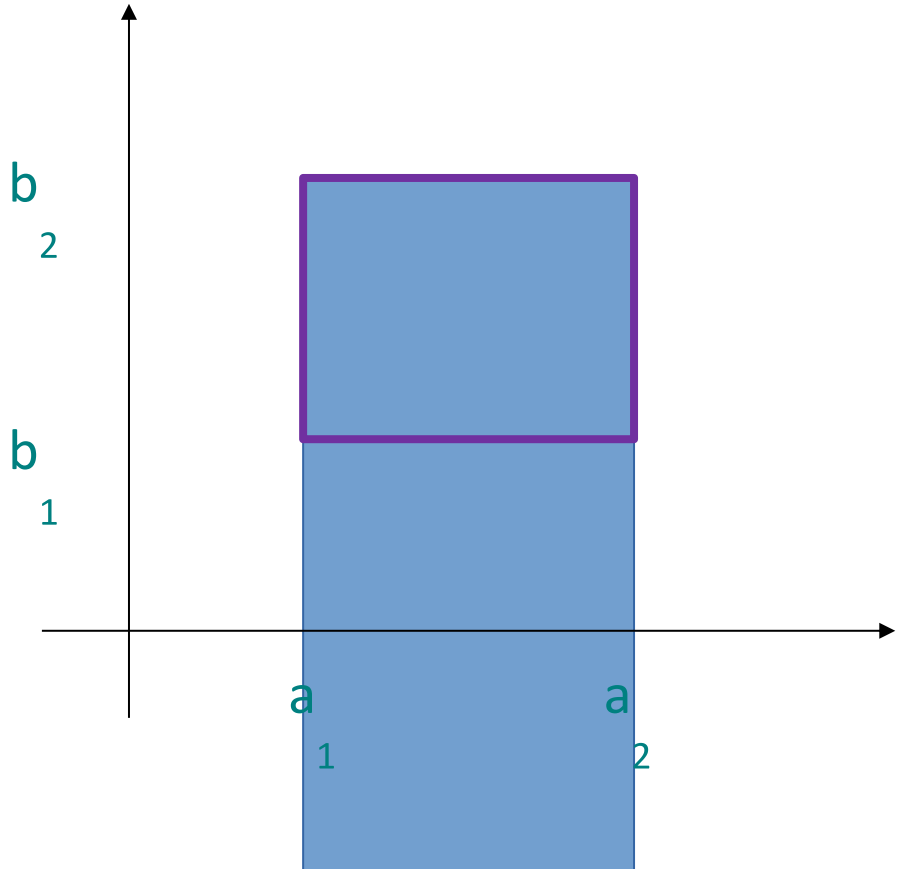
Probabilities from Joint CDF

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



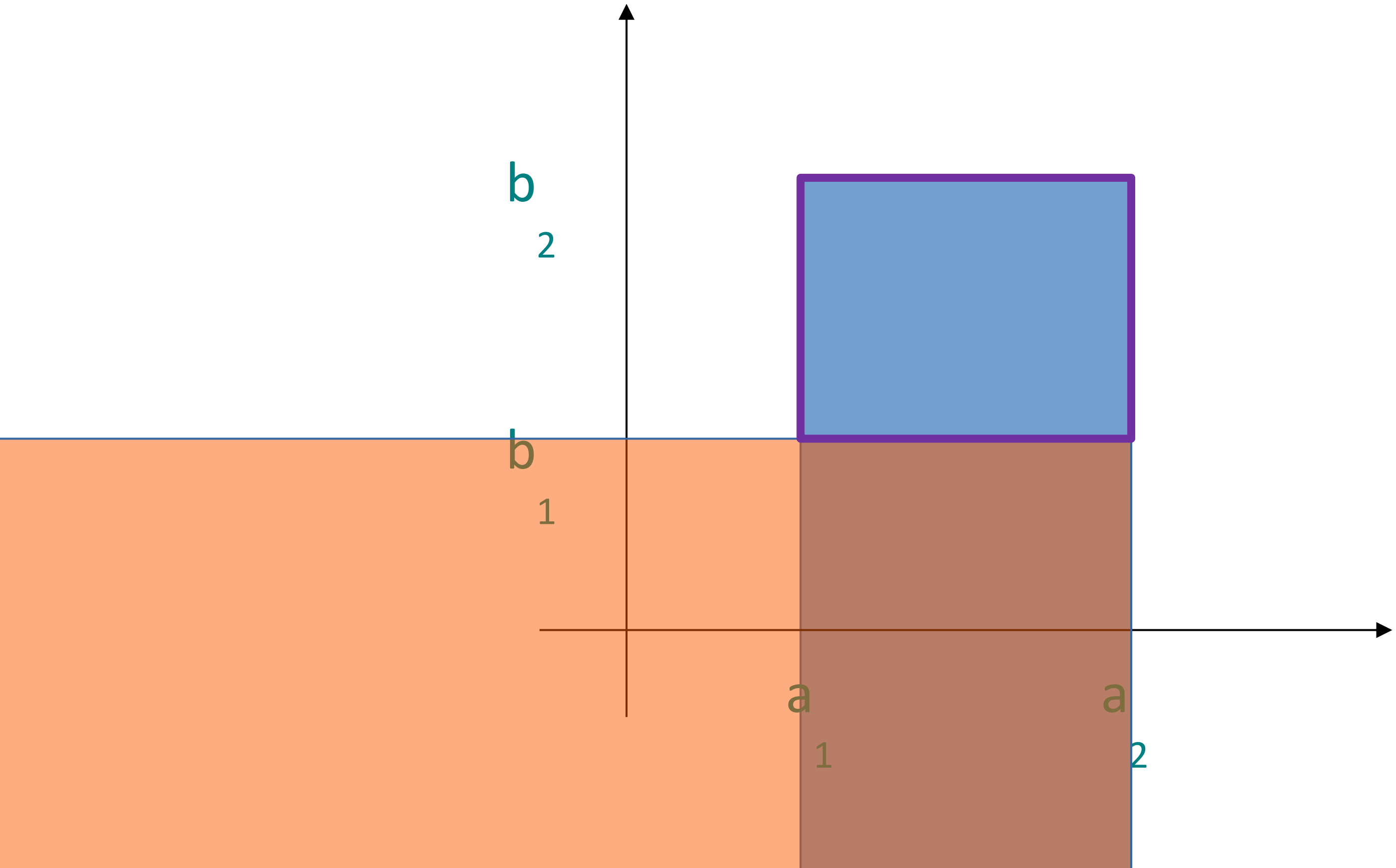
Probabilities from Joint CDF

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



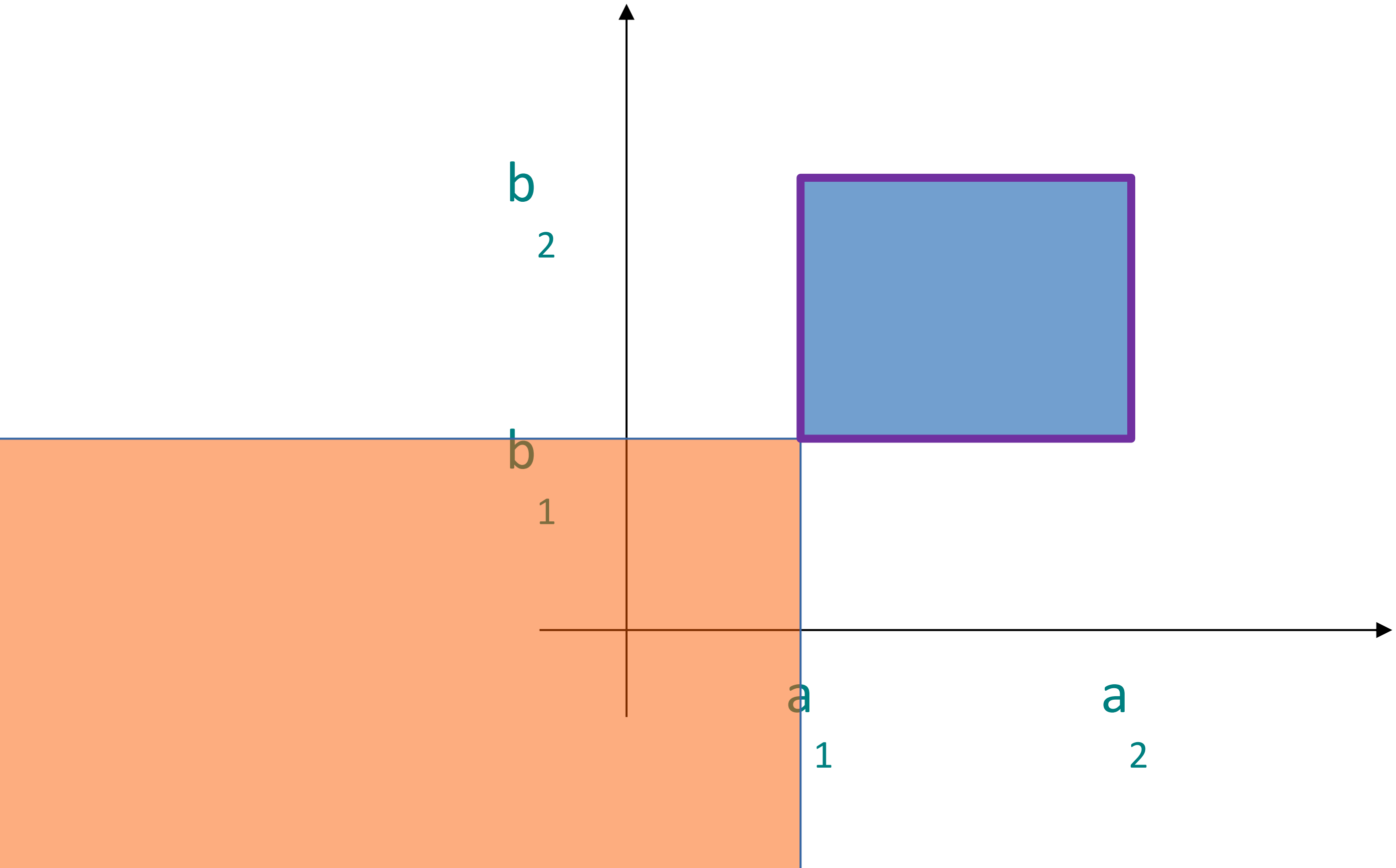
Probabilities from Joint CDF

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



Probabilities from Joint CDF

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



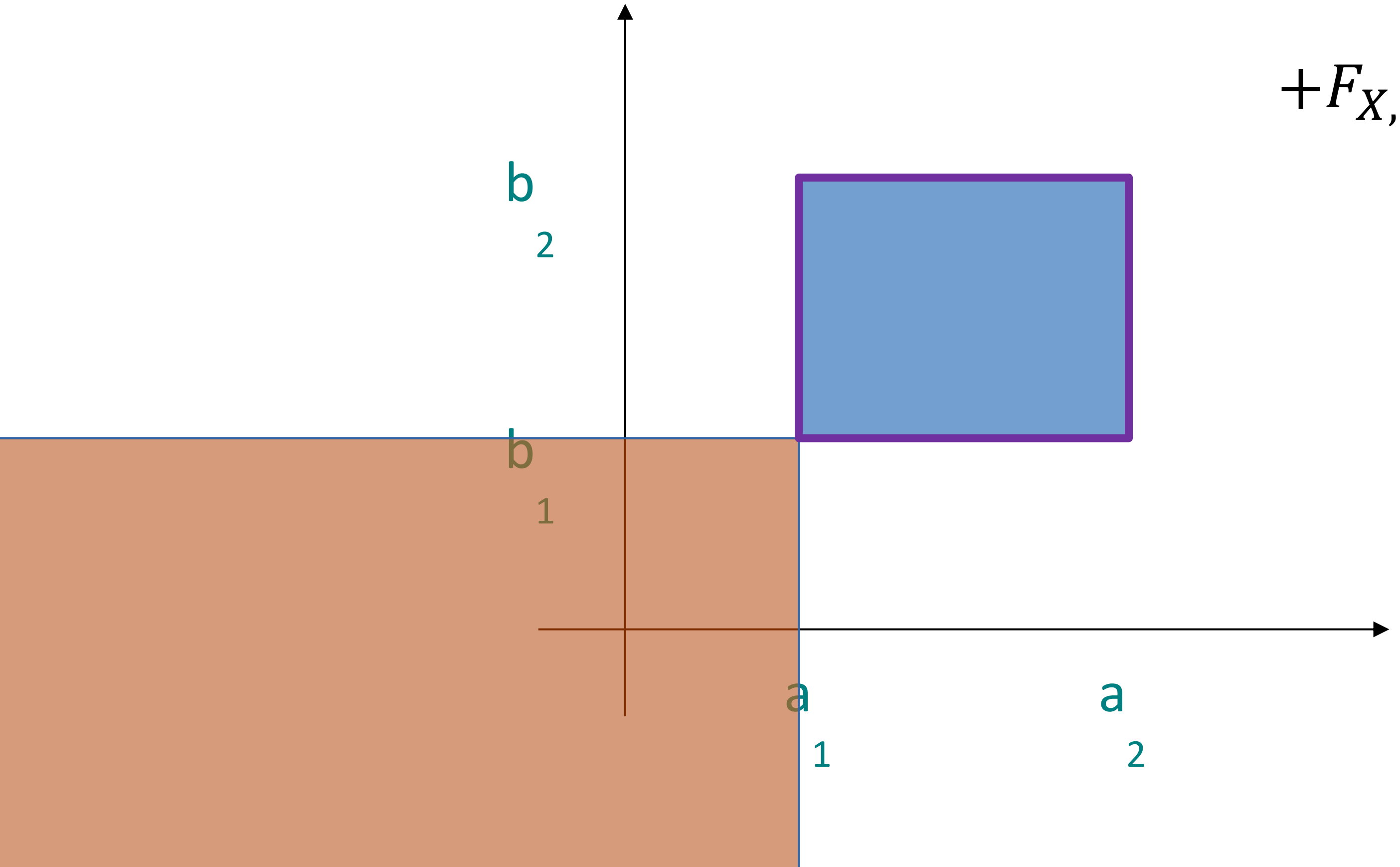
Probabilities from Joint CDF

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

$$-F_{X,Y}(a_1, b_2)$$

$$-F_{X,Y}(a_2, b_1)$$

$$+F_{X,Y}(a_1, b_1)$$



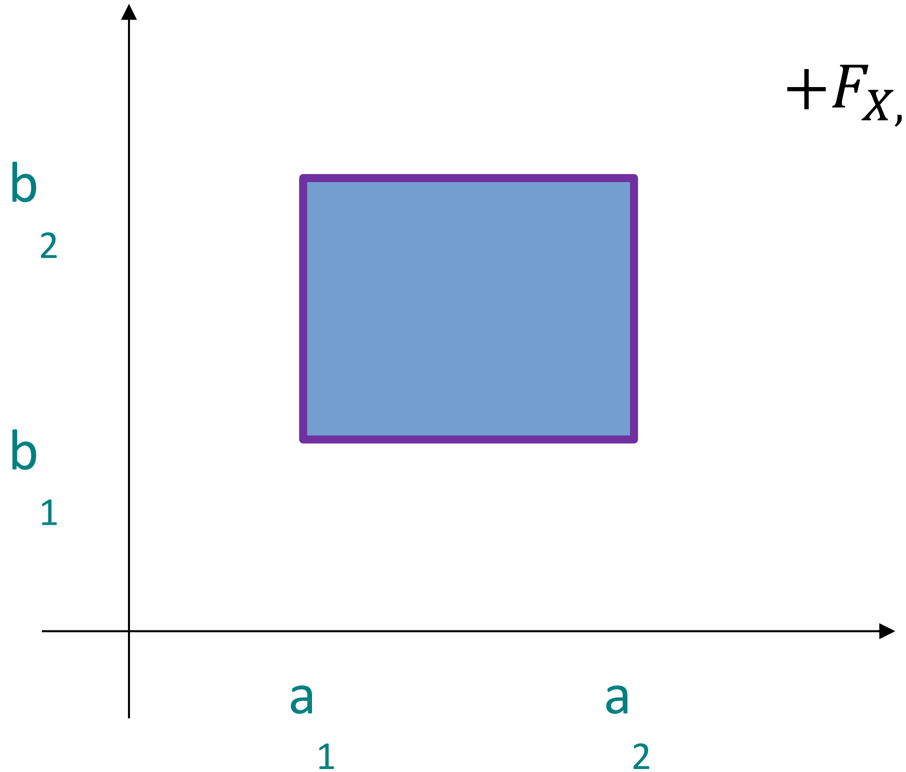
Probabilities from Joint CDF

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

$$-F_{X,Y}(a_1, b_2)$$

$$-F_{X,Y}(a_2, b_1)$$

$$+F_{X,Y}(a_1, b_1)$$



End Review

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?

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 random variables
- [ ] What happens when you **add** random variables?

Boolean Operation on Variable = Event

Recall: any boolean question about a random variable makes for an event. For example:

$$P(X \leq 5)$$

$$P(Y = 6)$$

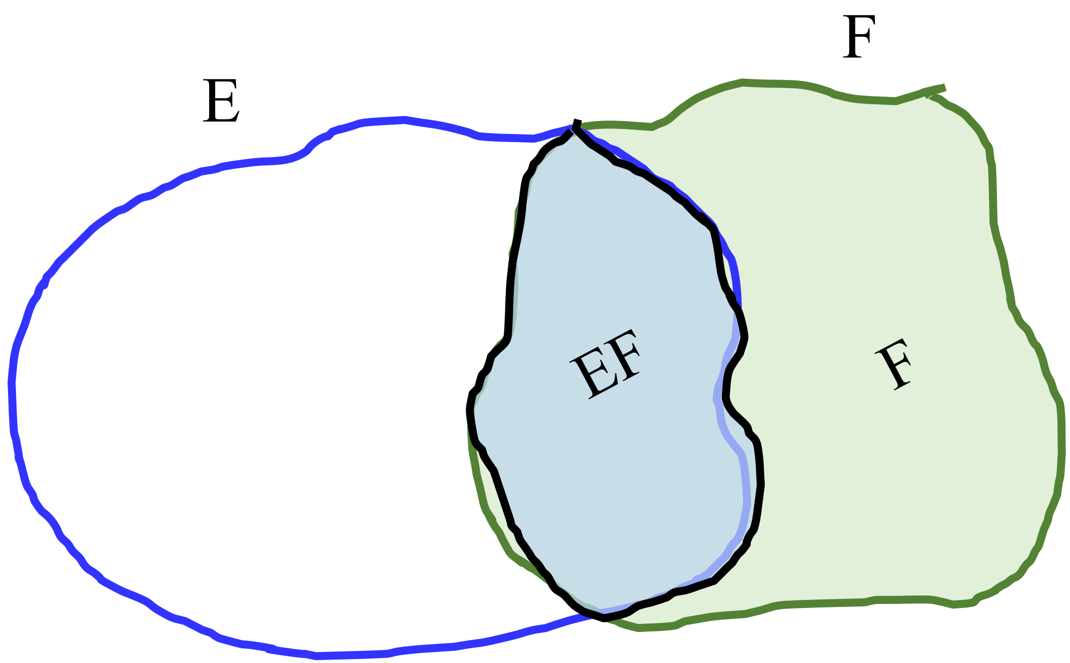
$$P(5 \leq Z \leq 10)$$

Conditionals with multiple variables

Discrete Conditional Distribution

- 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:

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

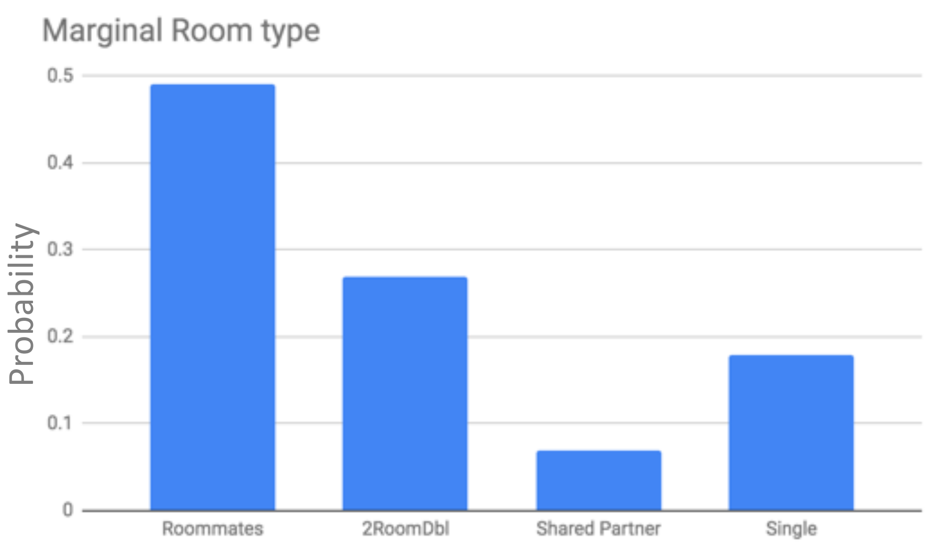
- Now, have X and Y as **discrete** random variables
  - **Conditional PMF** of X given Y:

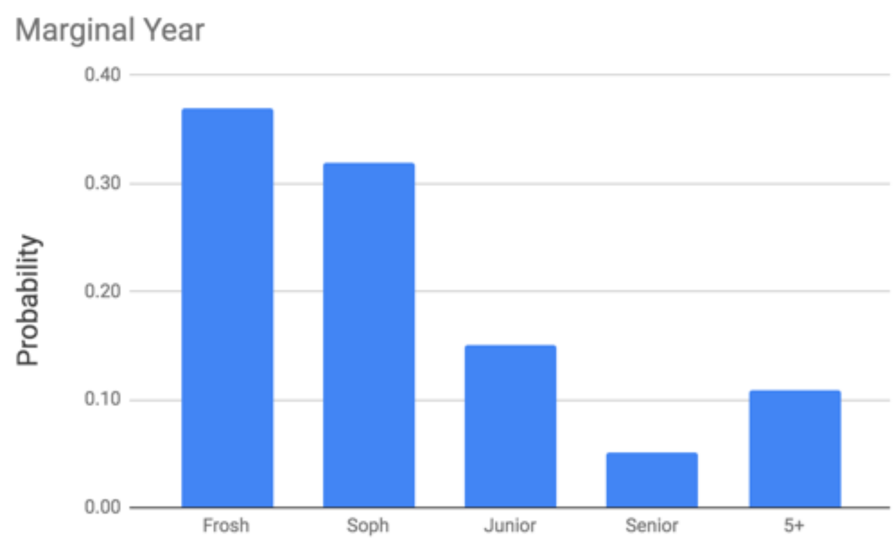
$$P_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Different notations,
same idea.

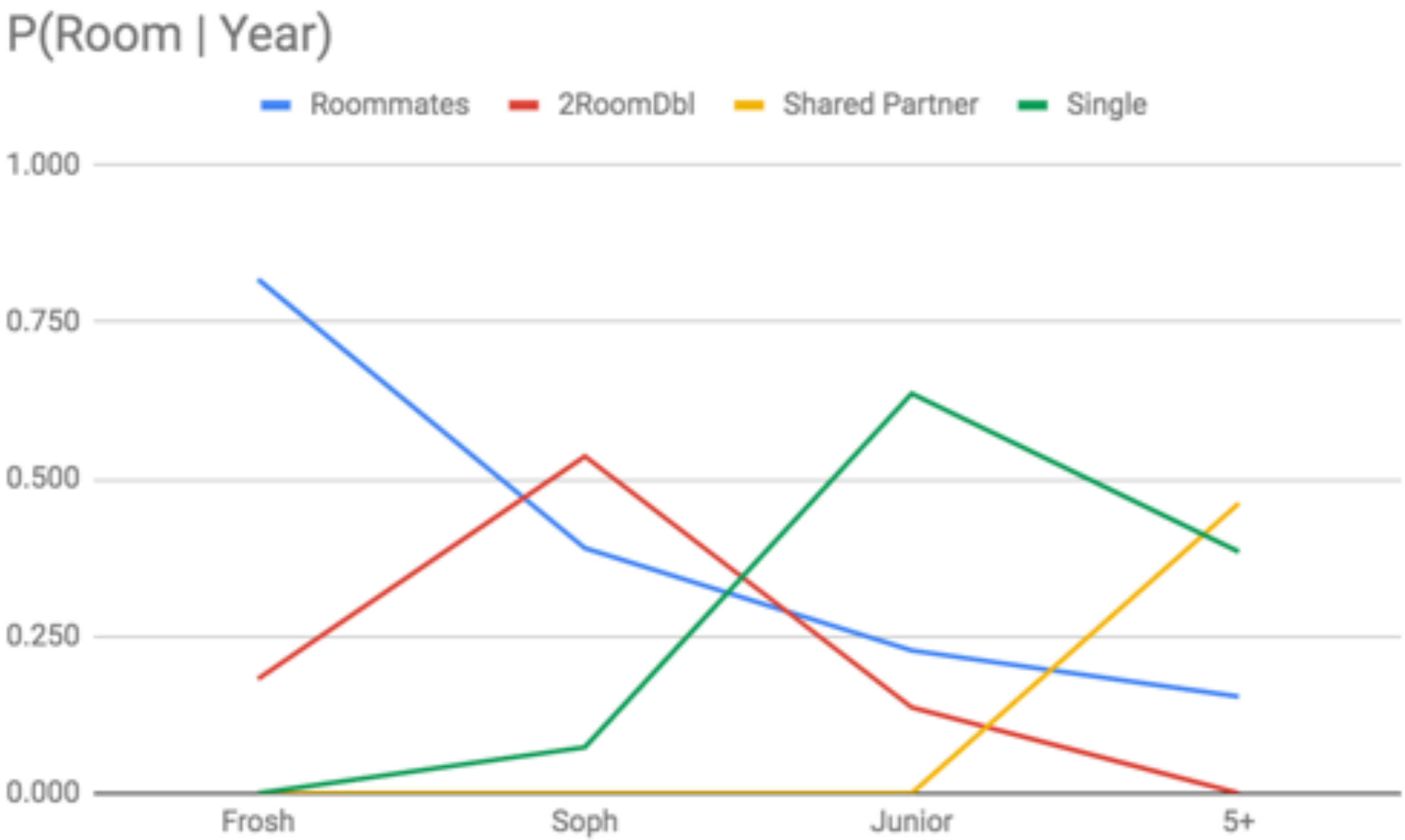
Joint Probability Table

	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



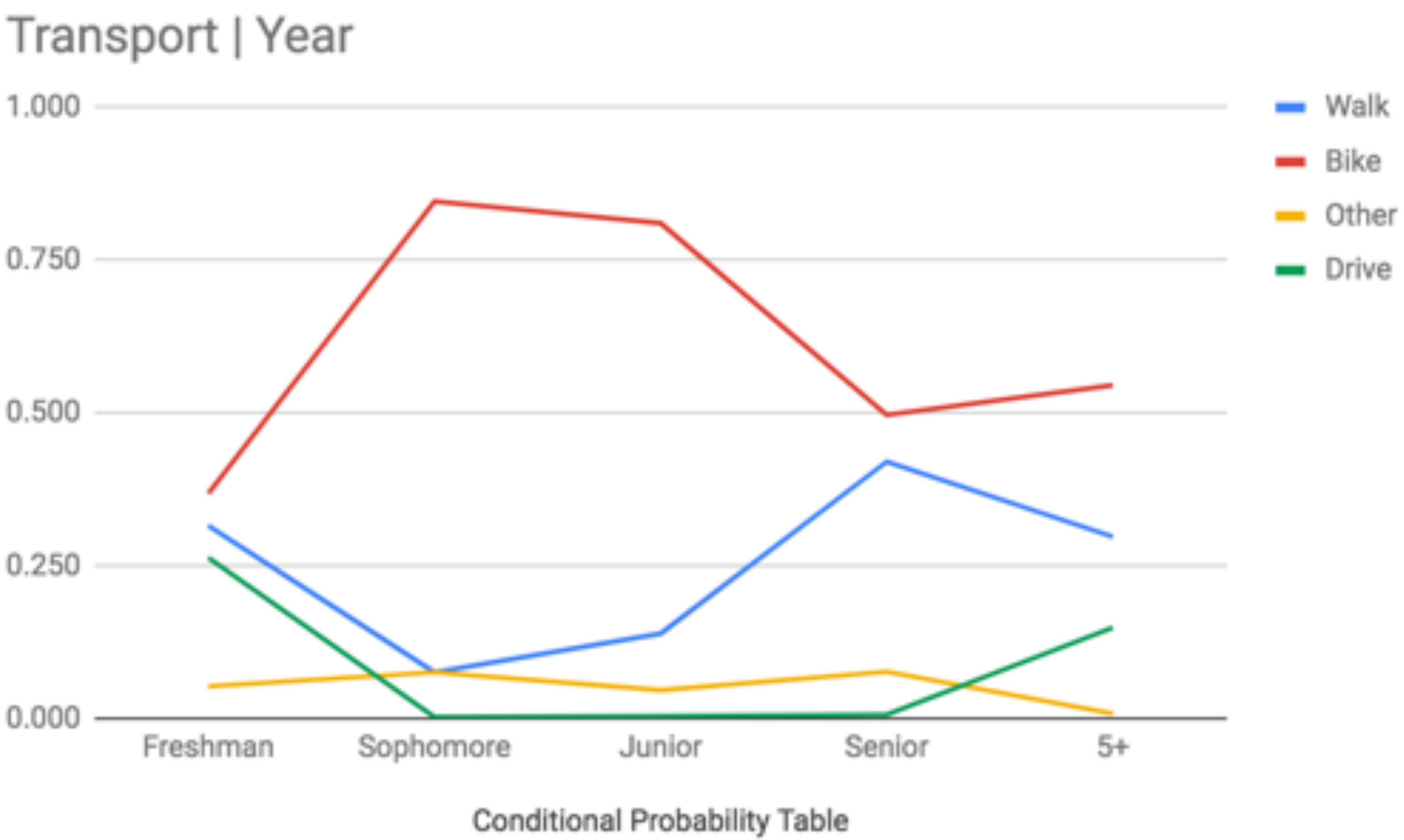


Room | Year

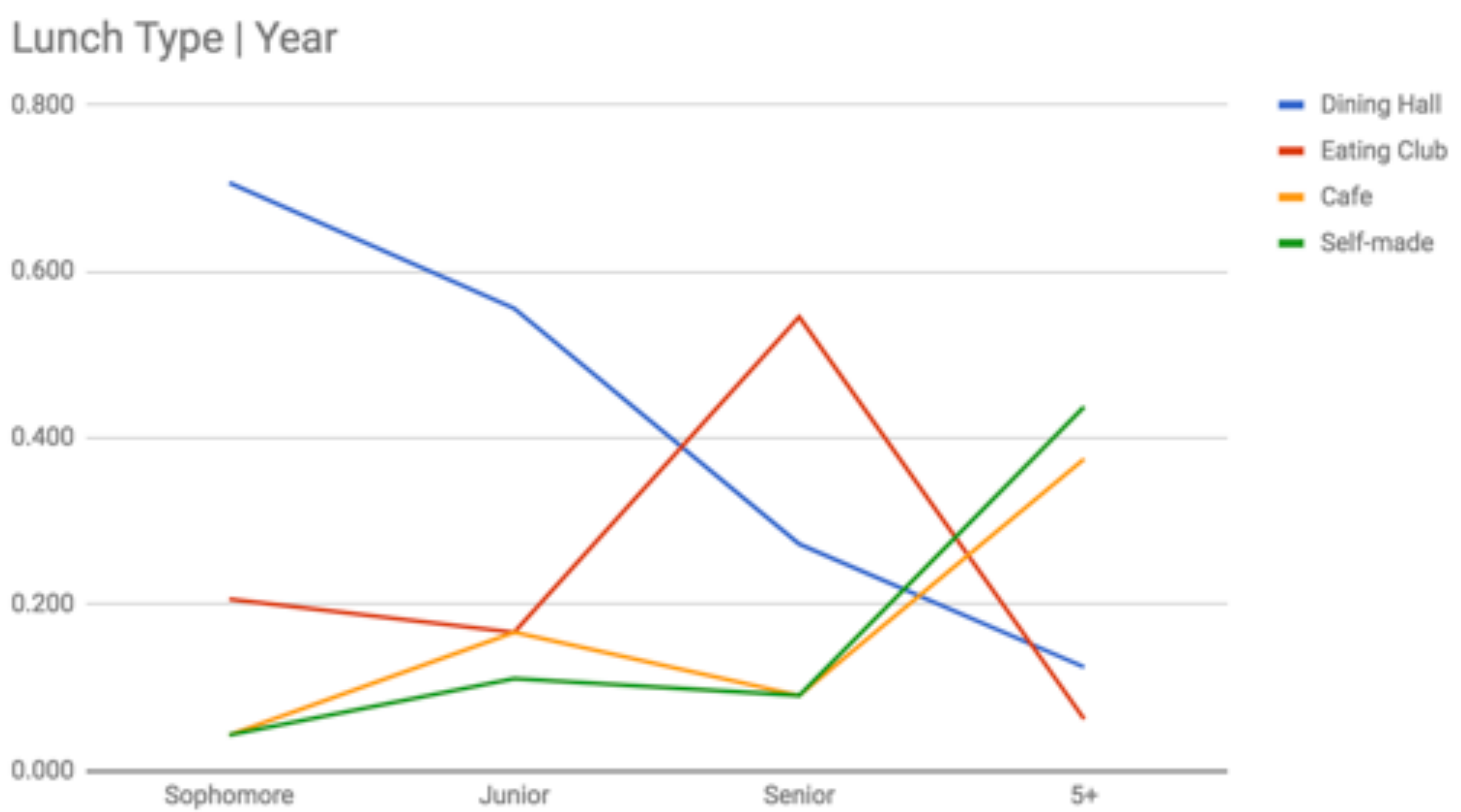


P(Room | Year)
Roommates
2RoomDbl
Shared Partner
Single
1.000
0.750
0.500
0.250
0.000
Frosh
Soph
Junior
5+

Transport | Year

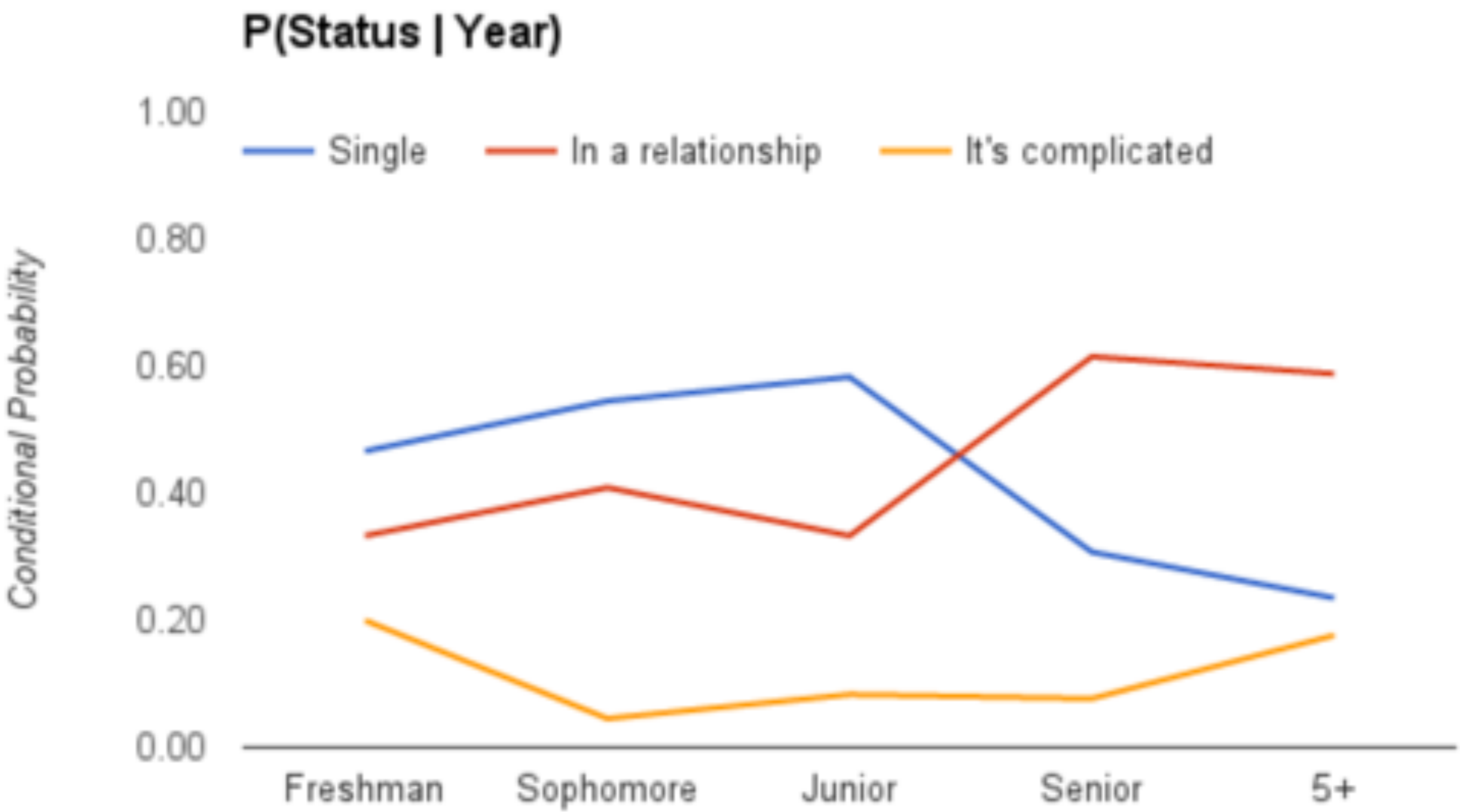


Lunch | Year



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)

Continuous Conditional Distributions

Let X and Y be continuous random variables

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Epsilons are just for derivation

$$\boxed{f_{X|Y}(x|y) \cdot \epsilon_x = \frac{f_{X,Y}(x,y) \cdot \epsilon_x \cdot \epsilon_y}{f_Y(y) \cdot \epsilon_y}}$$

You can skip to this version

$$\boxed{f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}}$$

Warmup: Bayes Revisited

Posterior belief

Likelihood of evidence

Prior belief

$$P(B|E) = \frac{P(E|B) \quad P(B)}{P(E)}$$

Normalization constant

Mixing Discrete and Continuous

Let N be a discrete random variable

Discrete X

$$P(X = x | N = n) = \frac{P(N = n | X = x) P(X = x)}{P(N = n)}$$

$$P_{X|N}(x|n) = \frac{P_{N|X}(n|x) P_X(x)}{P_N(n)}$$

Continuous X

$$f_{X|N}(x|n) \cdot \epsilon_x = \frac{P_{N|X}(n|x) f_X(x) \cdot \epsilon_x}{P_N(n)}$$

$$\boxed{f_{X|N}(x|n) = \frac{P_{N|X}(n|x) f_X(x)}{P_N(n)}}$$

All the Bayes Belong to Us

M,N are discrete. X, Y are continuous

OG Bayes

$$p_{M|N}(m|n) = \frac{P_{N|M}(n|m)p_M(m)}{p_N(n)}$$

Mix Bayes #1

$$f_{X|N}(x|n) = \frac{P_{N|X}(n|x)f_X(x)}{P_N(n)}$$

Mix Bayes #2

$$p_{N|X}(n|x) = \frac{f_{X|N}(x|n)p_N(n)}{f_X(x)}$$

All Continuous

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Warmup: Bayes Revisited

Likelihood of evidence

Prior belief

Posterior belief

$$P(B|E) = \frac{P(E|B) \quad P(B)}{P(E)}$$

Normalization constant

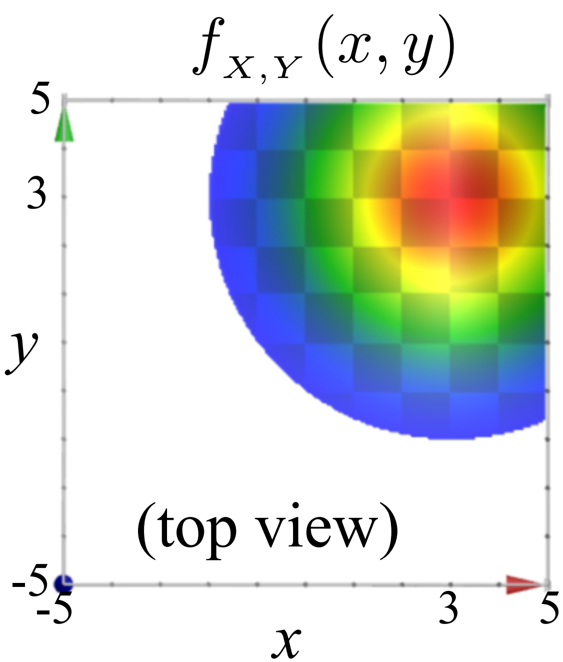
Warmup: Bivariate Normal

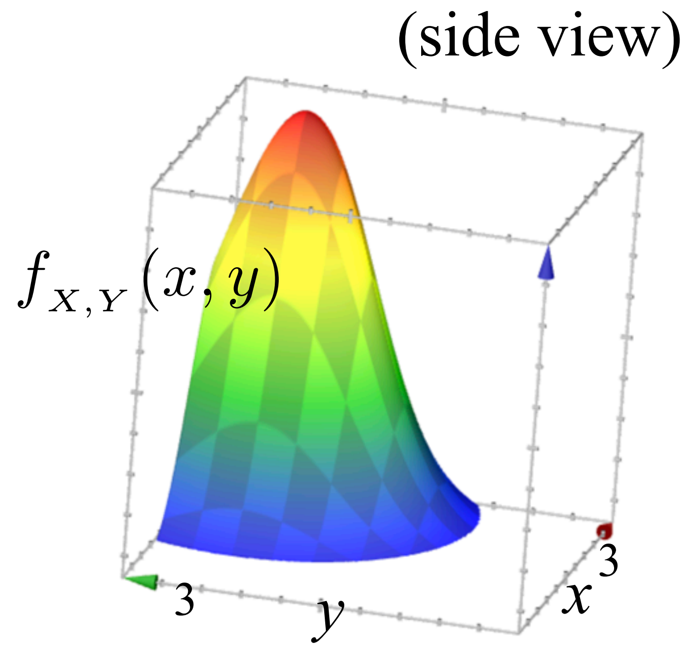
- X, Y follow a symmetric bivariate normal distribution if they have joint PDF:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{[(x-\mu_x)^2+(y-\mu_y)^2]}{2\cdot\sigma^2}}$$

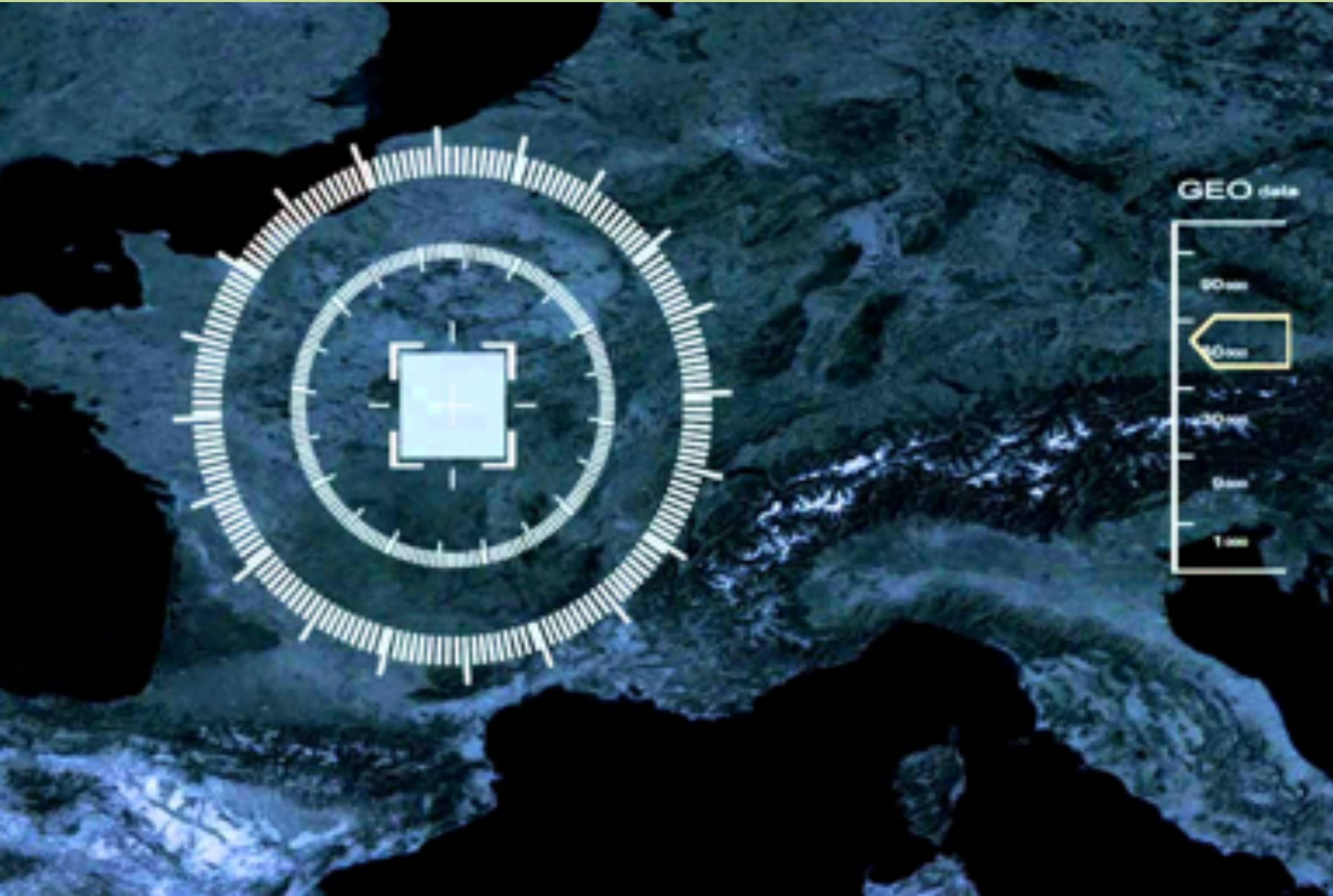
Here is an example where:

$$\mu_x = 3$$
$$\mu_y = 3$$
$$\sigma = 2$$



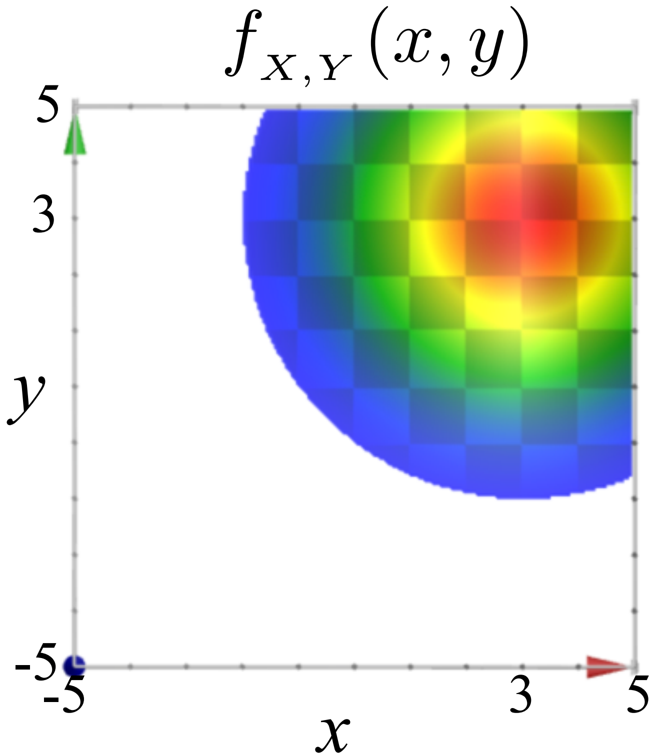


Tracking in 2D Space?



Tracking in 2D Space: Prior

Prior belief: $$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{[(x-\mu_x)^2+(y-\mu_y)^2]}{2\cdot\sigma^2}}$$



Satellite is at (0, 0)

$$\mu_x = 3$$
$$\mu_y = 3$$
$$\sigma = 2$$

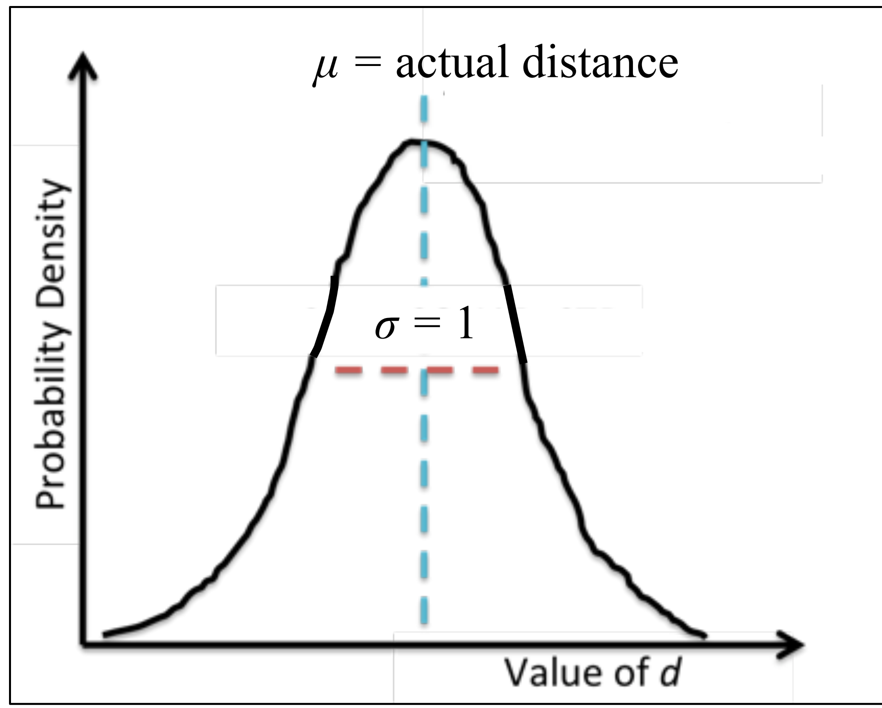
Prior belief with K: $$f_{X,Y}(x,y) = K \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

Tracking in 2D Space: Observation!

You now observe a noisy distance reading.
It says that your object is distance D away

We can say how likely that reading is if we know the actual location of the object…

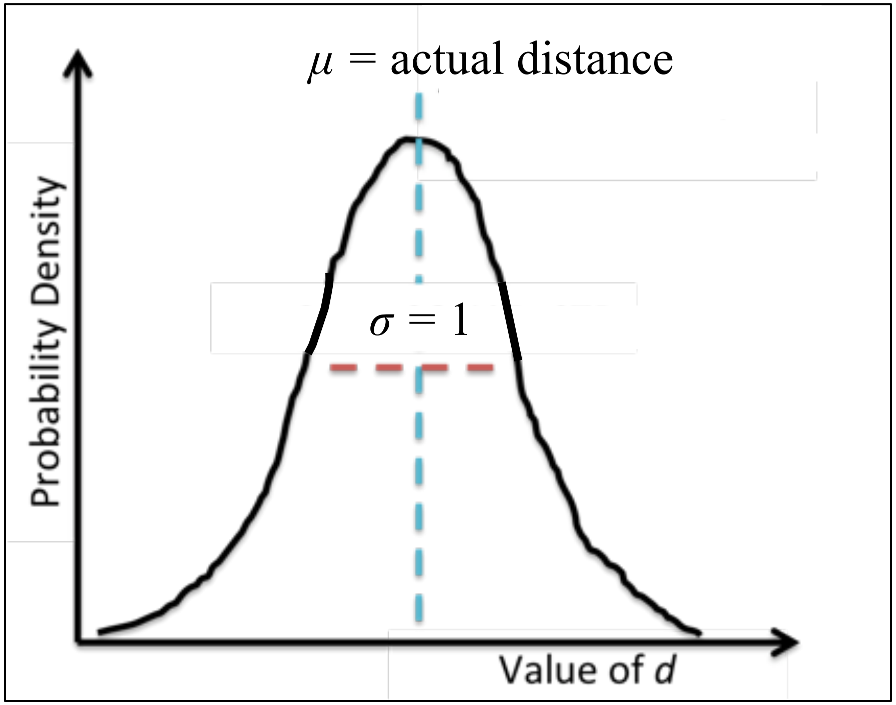
f(D | X, Y) is knowable!



Tracking in 2D Space: Observation!

Observe a ping of the object that is distance D away from satellite!

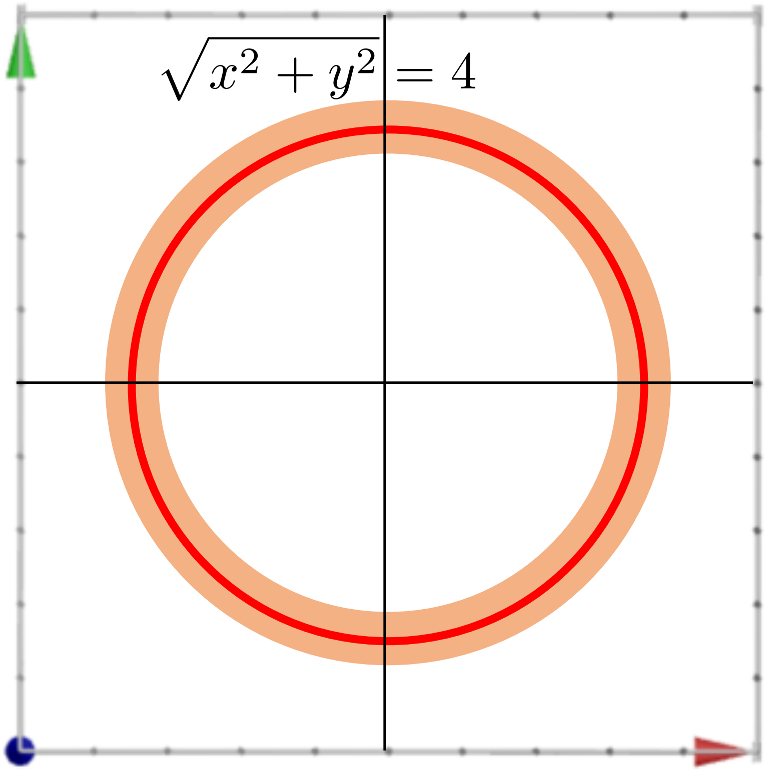
$$D|X,Y \sim N(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)$$

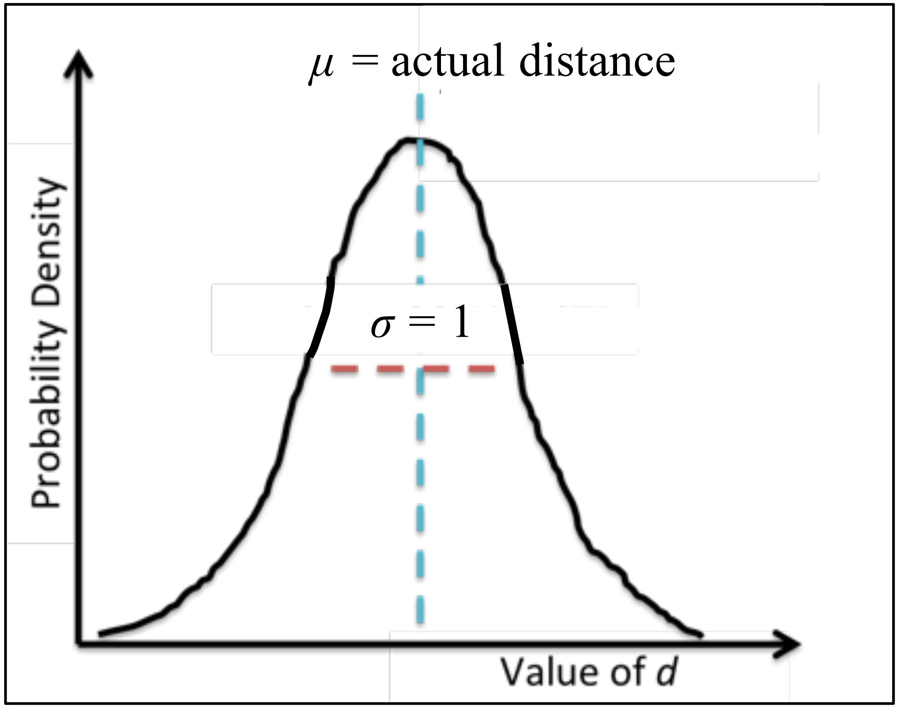


Know that the distance of a ping is normal with respect to the true distance.

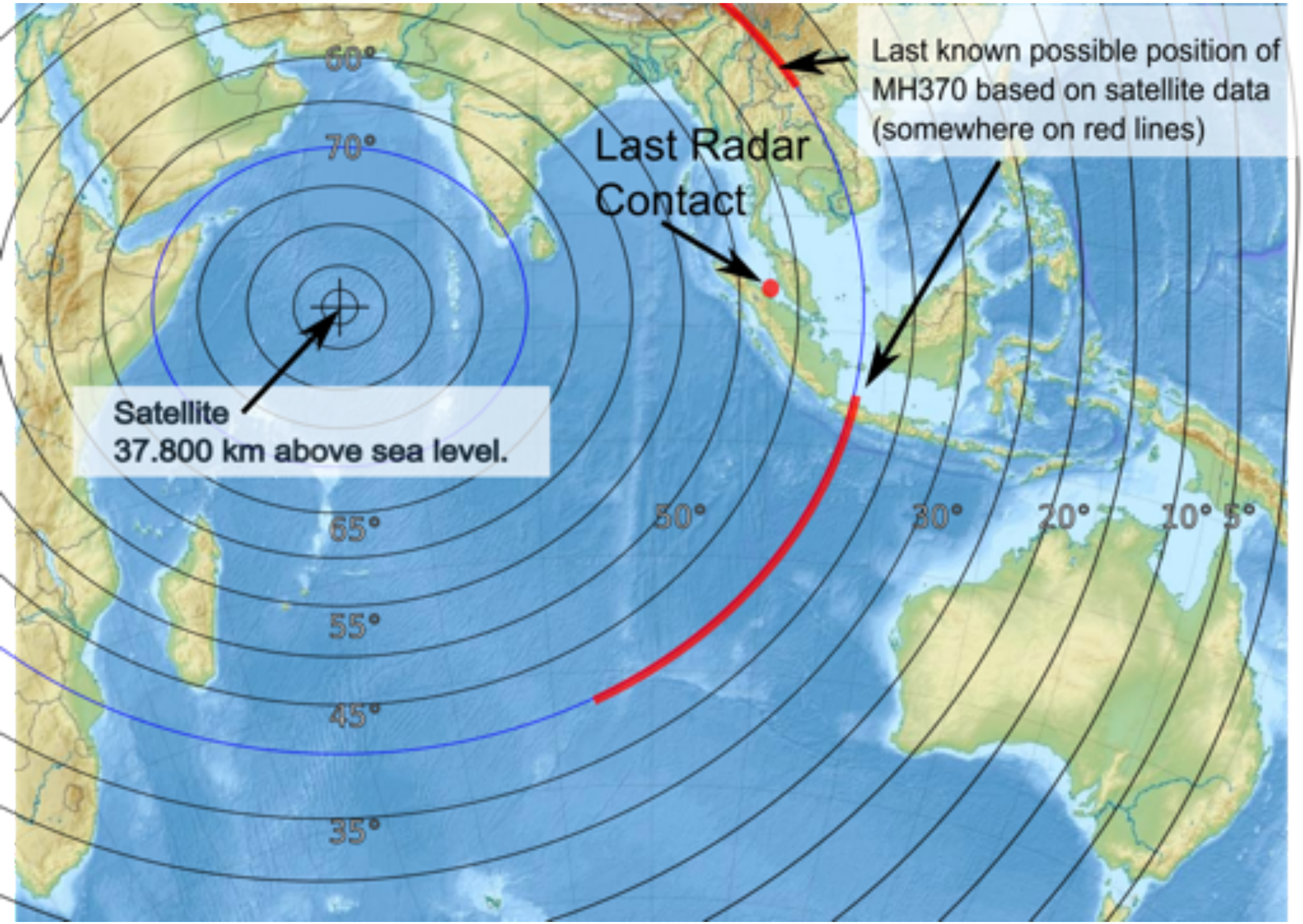
Tracking in 2D Space: Observation!

Observe a ping of the object that is distance $D = 4$ away!



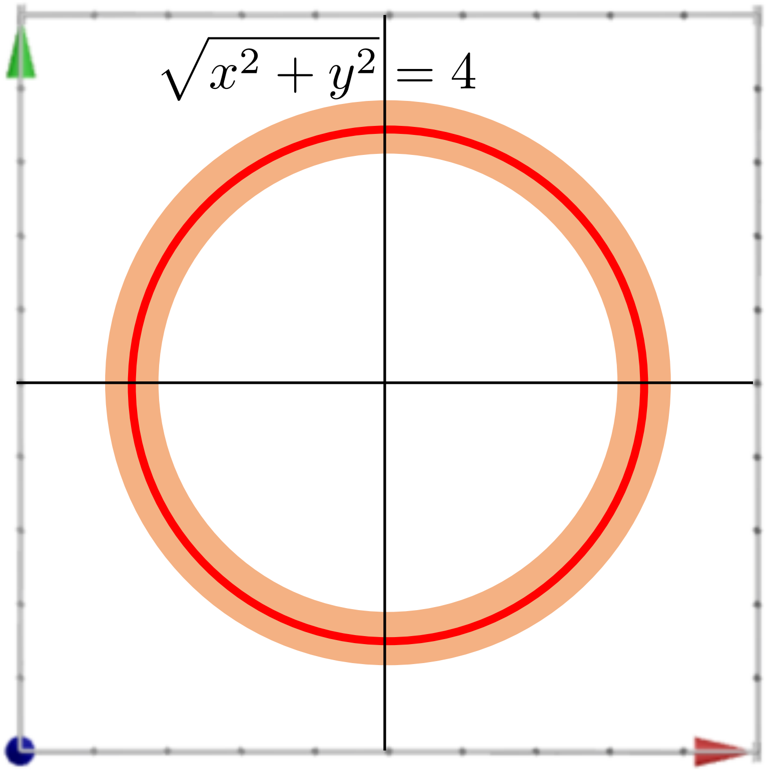


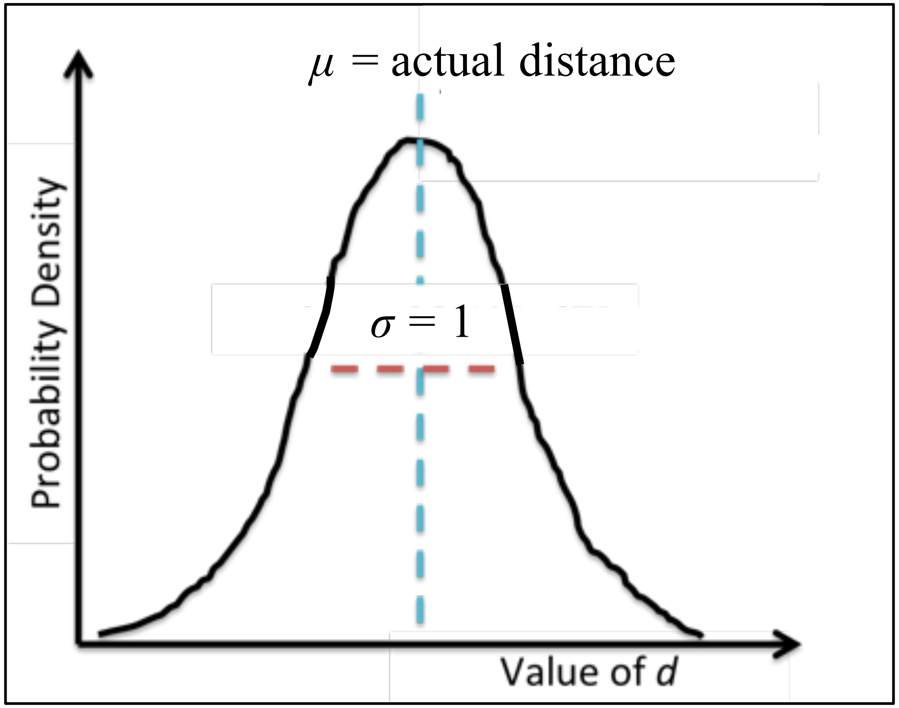
Know that the distance of a ping is normal with respect to the true distance


Last known possible position of
MH370 based on satellite data
(somewhere on red lines)
Last Radar
Contact
Satellite
37.800 km above sea level.
60°
70°
65°
55°
45°
35°
50°
30°
20°
10°
5°

Tracking in 2D Space: Observation!

Observe a ping of the object that is distance $D = 4$ away!





Know that the distance of a ping is normal with respect to the true distance

Tracking in 2D Space: Observation!

Observe a ping of the object that is distance *D = 4* away from satellite!

$$D|X,Y \sim N(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)$$

$$f(D = d|X = x, Y = y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(d-\mu)^2}{2\sigma^2}}$$

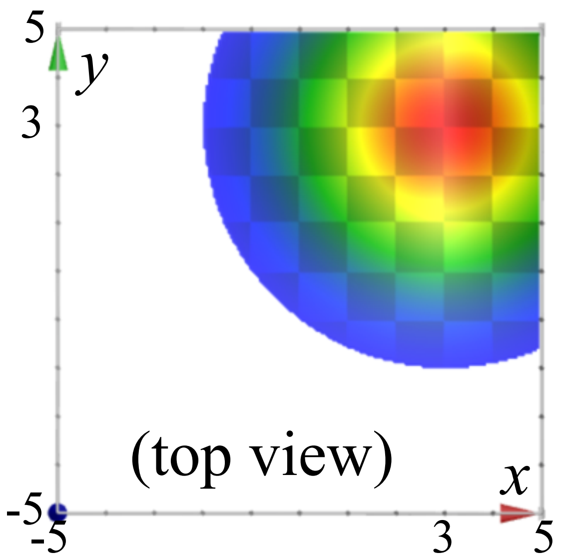
$$= \frac{1}{\sqrt{2\pi}} e^{\frac{-(d-\mu)^2}{2}}$$

$$= K_2 \cdot e^{\frac{-(d-\mu)^2}{2}}$$

$$= K_2 \cdot e^{\frac{-(d-\sqrt{x^2+y^2})^2}{2}}$$

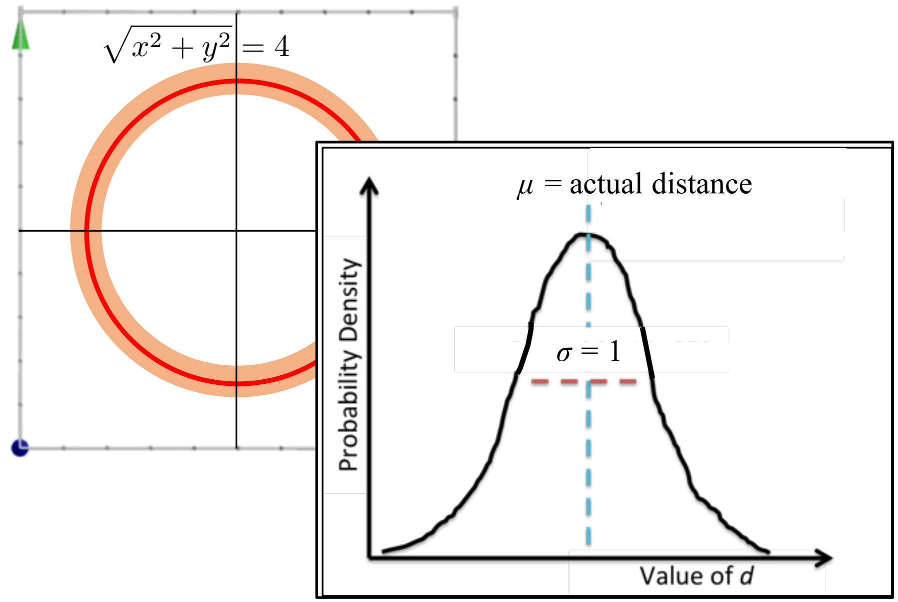
Tracking in 2D Space: New Belief

$$f(X=x, Y=y) = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$



Prior

Observation



$$f(D=d \mid X=x, Y=y) = K_2 \cdot e^{-\frac{[d-\sqrt{x^2+y^2}]^2}{2}}$$

What is your ***new*** belief for the location of the object being tracked? Your joint probability density function can be expressed with a constant

Tracking in 2D Space: New Belief

$$f(X=x, Y=y|D=4) = \frac{f(D=4|X=x,Y=y)\cdot f(X=x,Y=y)}{f(D=4)}$$

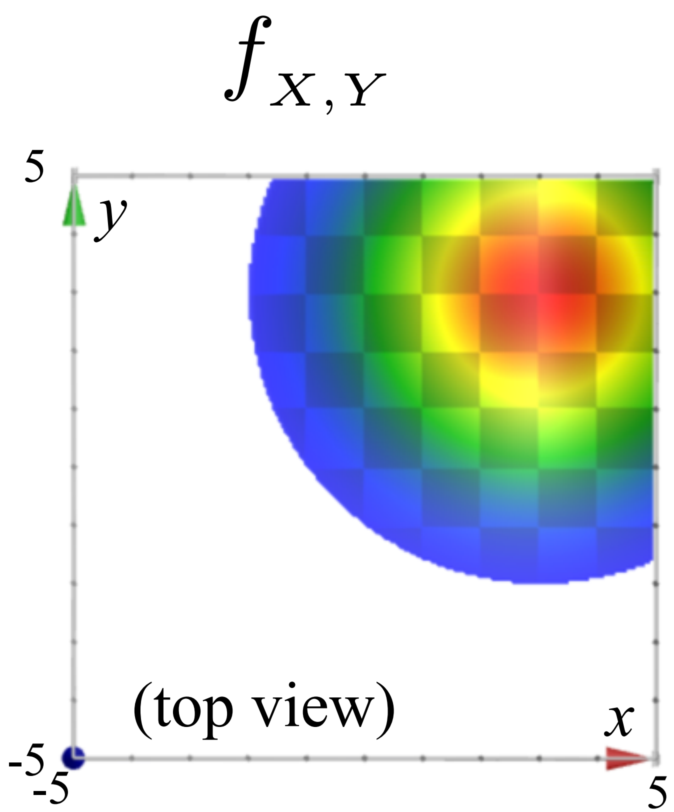
$$= \frac{K_1 \cdot e^{-\frac{[4-\sqrt{x^2+y^2})^2]}{2}} \cdot K_2 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}}{f(D=4)}$$

$$= \frac{K_3 \cdot e^{-\left[\frac{[4-\sqrt{x^2+y^2})^2]}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]}}{f(D=4)}$$

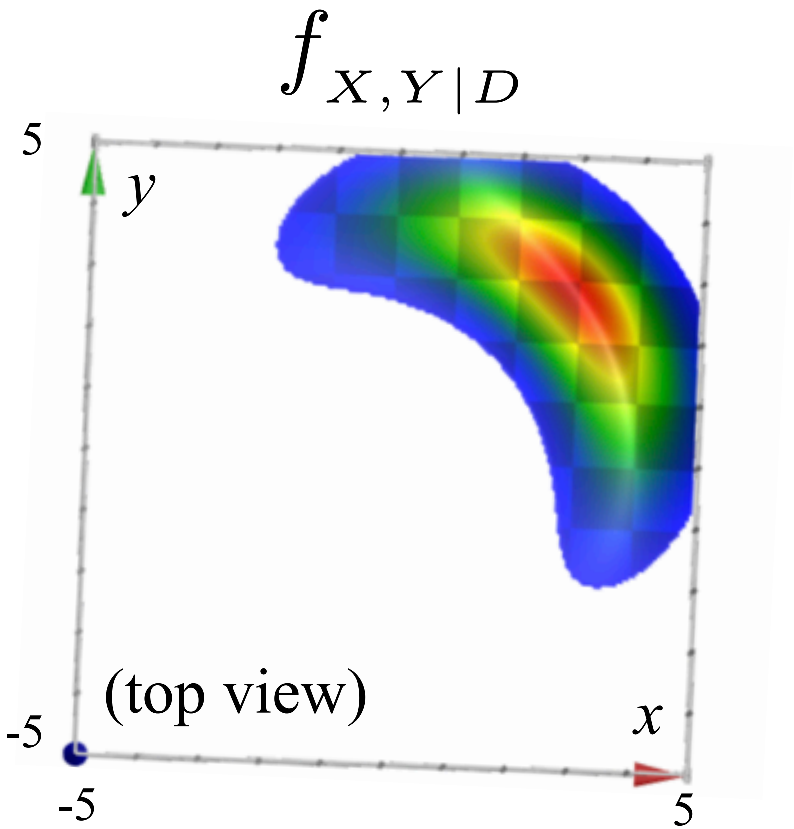
$$= K_4 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]}$$

For your notes…

Tracking in 2D Space: Posterior

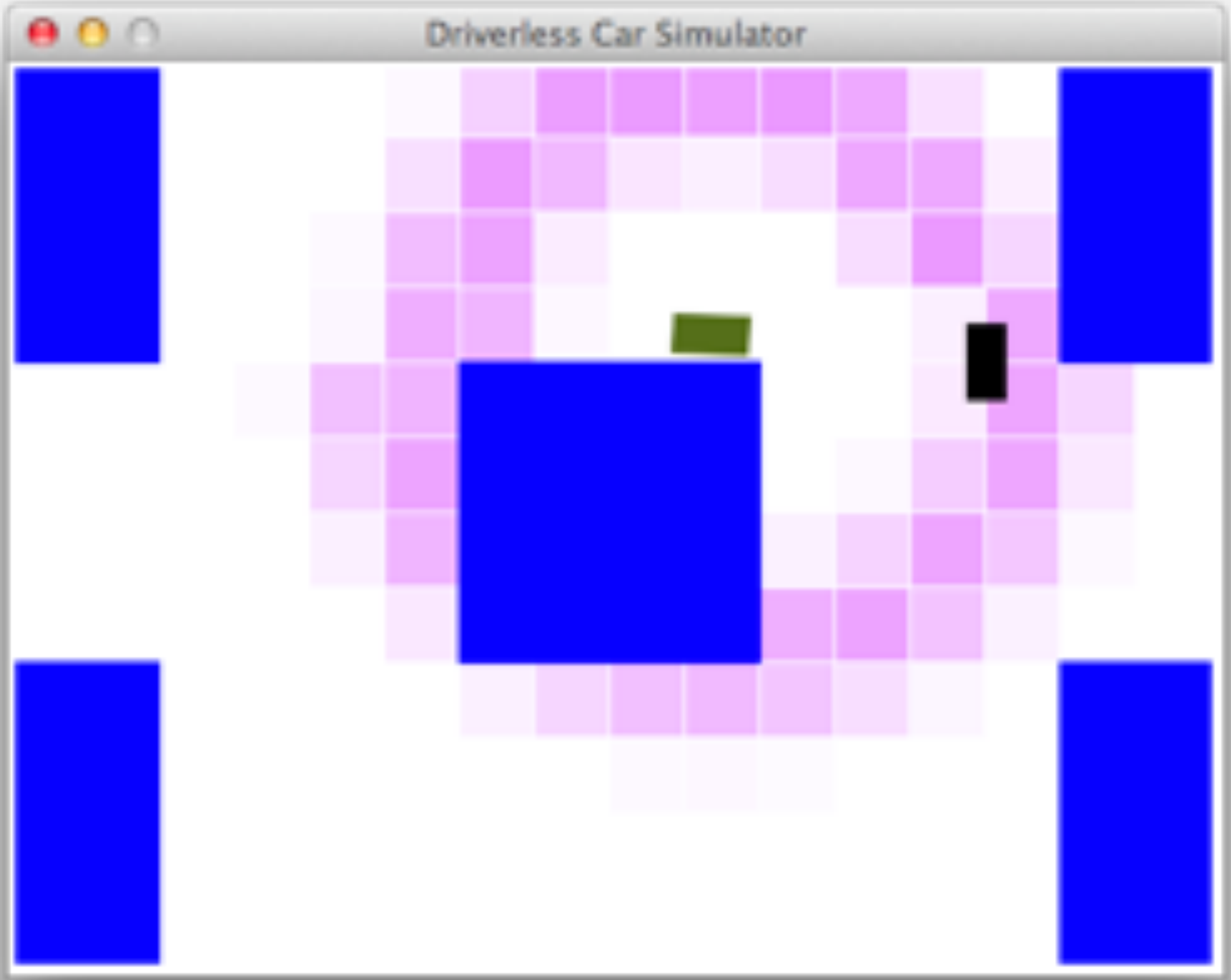


Prior



Posterior

Tracking in 2D Space: CS221



Stretch!

Properties of Joints:

Expectation, Independence, Convolution

Noah Arthurs
CS109, Stanford University

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 **expectation** of multiple RVS
- Use and find **independence** of multiple RVS
- What happens when you **add** random variables?

Expectation of Multiple RVs

Joint Expectation

$$E[X] = \sum_x xp(x)$$

- Expectation over a joint isn't nicely defined because it is not clear how to compose the multiple variables:
  - Add them? Multiply them?

- Lemma: For a function g(X,Y) we can calculate the expectation of that function:

$$E[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

- Recall, this also holds for single random variables:

$$E[g(X)] = \sum_x g(x)p(x)$$

Expected Values of Sums

Big deal lemma: first
stated without proof

E[X + Y] = E[X] + E[Y]

Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Holds regardless of dependency between X_i's

We Want a Proof!

Let g(X,Y) = [X + Y]

$$E[X+Y] = E[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

What a useful lemma

By the definition of g(x,y)

$$= \sum_{x,y} [x+y]p(x,y)$$

Break that sum into parts!

$$= \sum_{x,y} xp(x,y) + \sum_{x,y} yp(x,y)$$

Change the sum of (x,y) into separate sums

$$= \sum_{x} x \sum_{y} p(x,y) + \sum_{y} y \sum_{x} p(x,y)$$

That is the definition of marginal probability

$$= \sum_{x} xp(x) + \sum_{y} yp(y)$$

That is the definition of expectation

$$= E[X] + E[Y]$$

Independence and Random Variables

Independent Discrete Variables

- Two discrete random variables X and Y are called **<u>independent</u>** if:

$$p(x, y) = p_X(x) p_Y(y) \quad \text{for all } x, y$$

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

- Intuitively: knowing the value of X tells us nothing about the distribution of Y (and vice versa)
  - If two variables are **<u>not</u>** independent, they are called ***<u>dependent</u>***

- Similar conceptually to independent *events*, but we are dealing with multiple ***<u>variables</u>***
  - Keep your events and variables distinct (and clear)!

Is Year Independent of Lunch?

Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.03	0.00	0.02	0.00	0.05
Sophomore	0.50	0.15	0.03	0.03	0.68
Junior	0.08	0.02	0.02	0.02	0.12
Senior	0.02	0.05	0.01	0.01	0.08
5+	0.02	0.01	0.05	0.05	0.07
Marginal Status	0.65	0.22	0.12	0.11	

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Lunch} = s)}_{0.50} = \underbrace{P(\text{Year} = y)}_{0.68}\underbrace{P(\text{Lunch} = s)}_{0.65}$$

Yes!

Is Year Independent of Lunch?

Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.03	0.00	0.02	0.00	0.05
Sophomore	0.50	0.15	$\boxed{0.03}$	0.03	$\boxed{0.68}$
Junior	0.08	0.02	0.02	0.02	0.12
Senior	0.02	0.05	0.01	0.01	0.08
5+	0.02	0.01	0.05	0.05	0.07
Marginal Status	0.65	0.22	$\boxed{0.12}$	0.11	

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Lunch} = s)}_{0.03} = \underbrace{\underbrace{P(\text{Year} = y)}_{0.68}\underbrace{P(\text{Lunch} = s)}_{0.12}}_{0.08}$$

No ☹

Coin Flips

- Flip coin with probability *p* of "heads"
  - Flip coin a total of *n* + *m* times
  - Let X = number of heads in first *n* flips
  - Let Y = number of heads in next *m* flips

$$P(X = x, Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}$$

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

  - X and Y are independent
  - Let Z = number of total heads in *n* + *m* flips
  - Are X and Z independent?
    - What if you are told Z = 0?

Independent Continuous Variables

- Two continuous random variables X and Y are called **independent** if:

  P(X ≤ *a*, Y ≤ *b*) = P(X ≤ *a*) P(Y ≤ *b*) for any *a*, *b*

- Equivalently:

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \quad \text{for all } a,b$$

$$f_{X,Y}(a,b) = f_X(a)f_Y(b) \quad \text{for all } a,b$$

- More generally, joint density factors separately:

$$f_{X,Y}(x,y) = h(x)g(y) \quad \text{where } -\infty < x, y < \infty$$

Is the Blur Distribution Independent?

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.

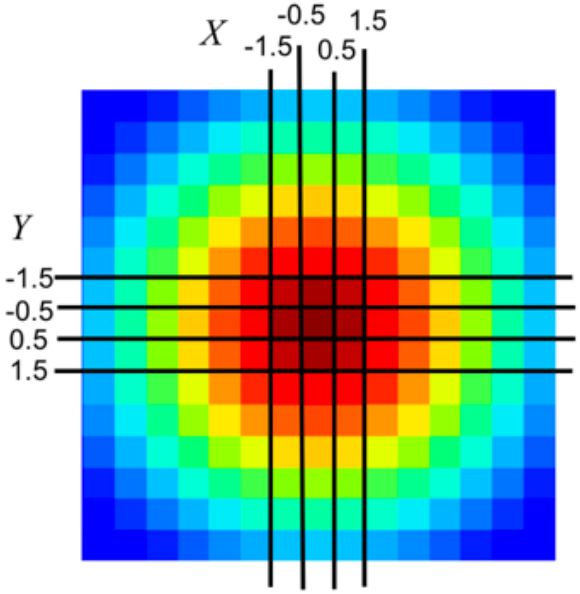
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

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

Pop Quiz (just kidding)

- Consider joint density function of X and Y:

$$f_{X,Y}(x,y) = 6e^{-3x}e^{-2y} \quad \text{for } 0 < x, y < \infty$$

  - Are X and Y independent? Yes!

Let $h(x) = 3e^{-3x}$ and $g(y) = 2e^{-2y}$, so $f_{X,Y}(x,y) = h(x)g(y)$

- Consider joint density function of X and Y:

$$f_{X,Y}(x,y) = 4xy \quad \text{for } 0 < x, y < 1$$

  - Are X and Y independent? Yes!

Let $h(x) = 2x$ and $g(y) = 2y$, so $f_{X,Y}(x,y) = h(x)g(y)$

  - Now add constraint that: $0 < (x + y) < 1$
  - Are X and Y independent? No!
    - Cannot capture constraint on x + y in factorization!