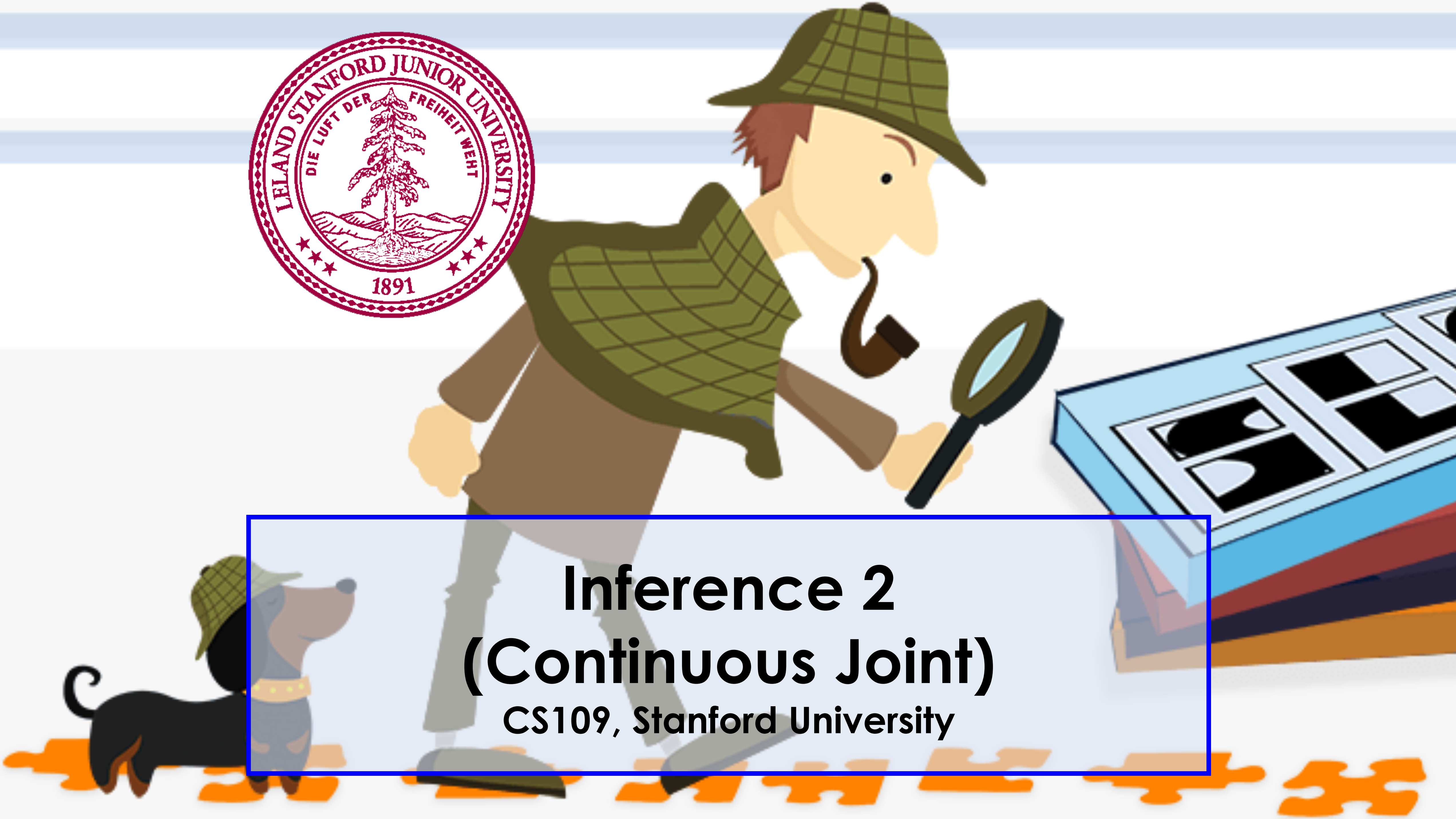


Inference 2 (Continuous Joint)

CS109, Stanford University

Midterm **Tuesday** July 25th, 7~9pm

- Closed computer. 2 sheets of notes, front and back.

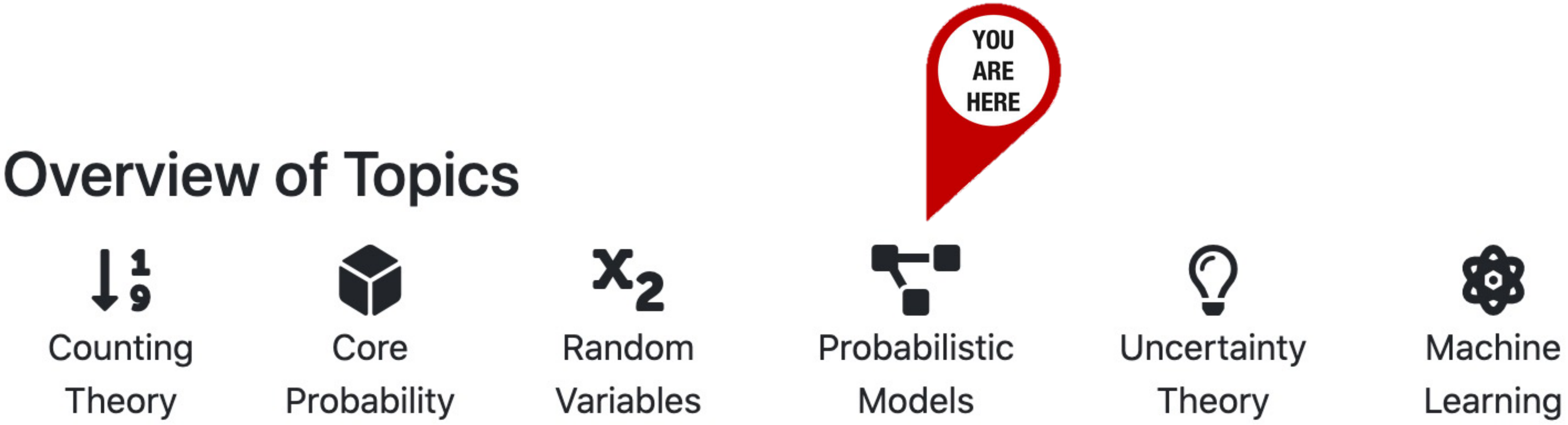
370-370

Learning Goals

1. Understand the Continuous Joint Distribution
2. Combine Bayes Theorem and Continuous Random Variables

Review

Where are we in CS109?



Where are we locally?

Discrete Models: Joints, Multinomial

Inference: Change RV belief from Observations

Modelling: Make your own!

General Inference: Use computers to infer

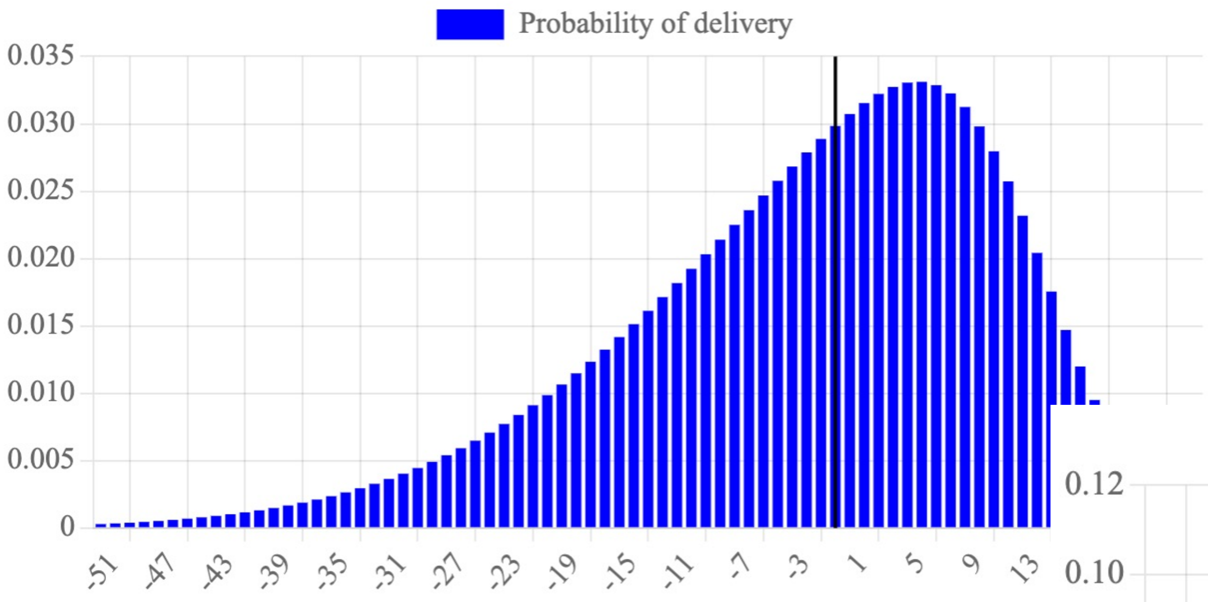
Today: Inference

Inference *noun*

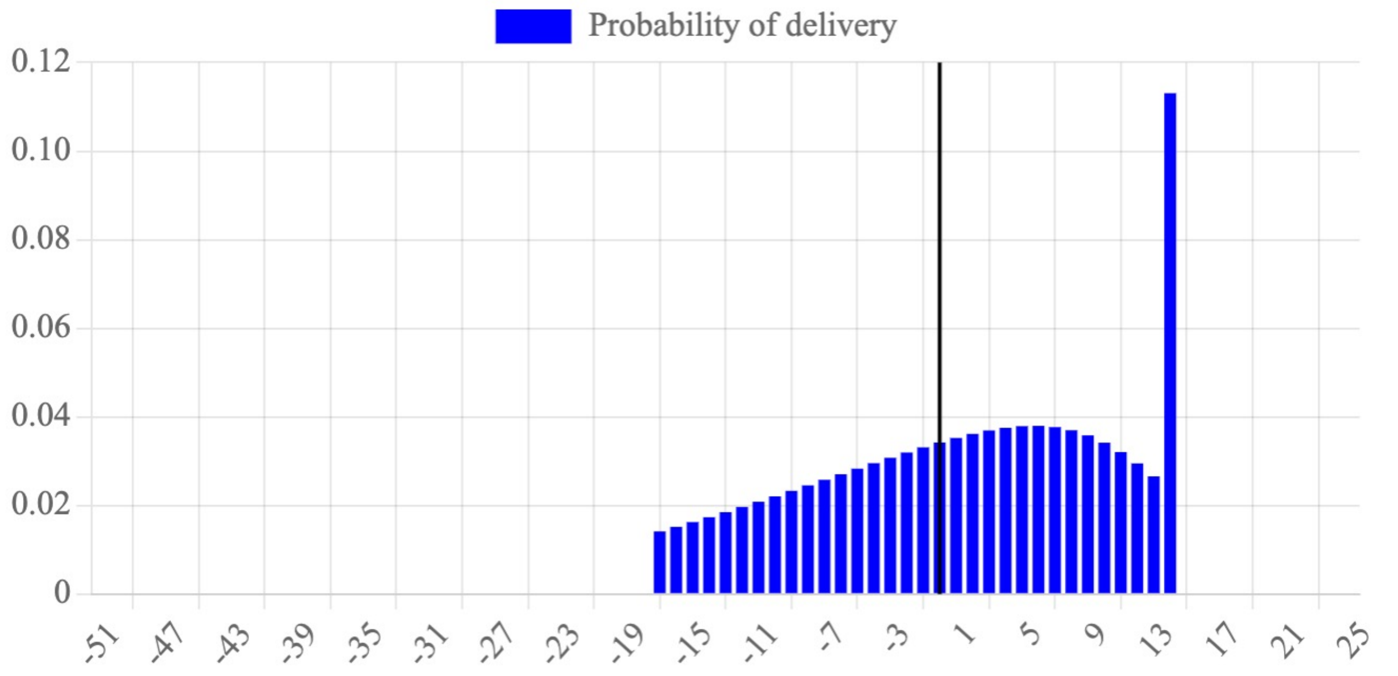
An updated belief about a random variable (or multiple) based on conditional knowledge regarding another random variable (or multiple) in a probabilistic model.

TLDR: conditional probability with random variables.

Another example: Baby delivery



Its 19 days until the due date and no baby



For each value *d:*

$$P(D = d | \text{no child so far})$$

$$= \frac{P(\text{no child so far} | D = d) P(D = d)}{P(\text{no child so far})}$$

Bayes Theorm with Discrete

Let M be a **discrete** random variable

Let N be a **discrete** random variable

$$P(M = 2 | N = 3) = \frac{P(N = 3 | M = 2) P(M = 2)}{P(N = 3)}$$

More generally

$$P(M = m | N = n) = \frac{P(N = n | M = m) P(M = m)}{P(N = n)}$$

Shorthand notation

$$P(m | n) = \frac{P(n | m) P(m)}{P(n)}$$

All the Bayes Belong to Us

M,N are discrete. X, Y are continuous

OG Bayes

$$P(m|n) = \frac{P(n|m)P(m)}{P(n)}$$

Mix Bayes #1

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

Mix Bayes #2

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

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

LOTP? Chain Rule? You can play too!

N is discrete. **X is continuous**

$$f(N = n, X = x) = f(X = x | N = n)P(N = n)$$

$$f(X = x) = \sum_n f(X = x | N = n)P(N = n)$$

Inference with Continuous

Q: At birth, girl elephant weights are distributed as a Gaussian with mean = 160kg, std = 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean = 165kg, std = 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl?

Inference with Continuous

Q: What is $P(G = 1 \mid X = 163)$

Let G be an indicator that the elephant is a girl. G is $\text{Bern}(p = 0.5)$

Let X be the distribution of weight of the elephant.

$X \mid G = 1$ is $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$ is $N(\mu = 165, \sigma^2 = 3^2)$

Inference with Continuous

Q: What is P(G = 1 | X = 163)

Let G be an indicator that the elephant is a girl. G is Bern(p = 0.5)

Let X be the distribution of weight of the elephant.

$X \mid G = 1$ is $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$ is $N(\mu = 165, \sigma^2 = 3^2)$

Joint Distribution is Implied:

$$f(G = 1, X = 72.3) = f(X = 72.3|G = 1)P(G = 1)$$

More generally

$$f(G = g, X = x) = f(X = x|G = g)P(G = g)$$

Inference with Continuous

Q: What is $P(G = 1 \mid X = 163)$

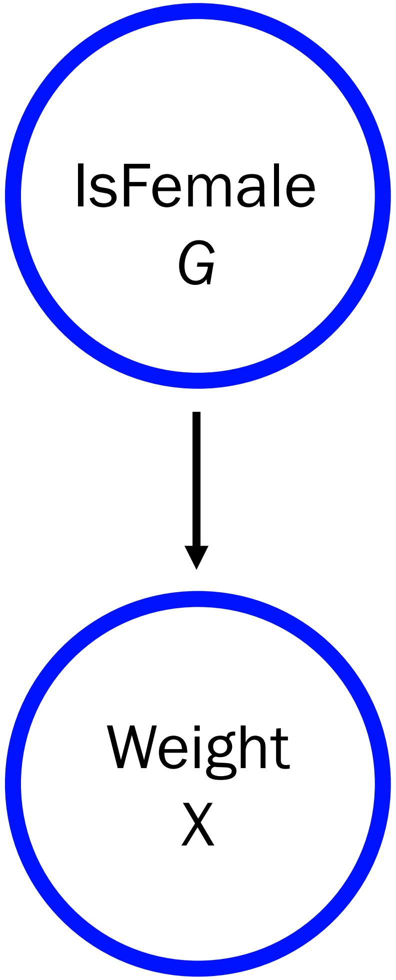
Let G be an indicator that the elephant is a girl. G is Bern($p = 0.5$)

Let X be the distribution of weight of the elephant.

$X \mid G = 1$ is $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$ is $N(\mu = 165, \sigma^2 = 3^2)$

Model Shown Graphically



G = 1 is Bern(p = 0.5)

X | G = 1 is $N(\mu = 160, \sigma^2 = 7^2)$
X | G = 0 is $N(\mu = 165, \sigma^2 = 3^2)$

Does this define the joint?

$$f(G = g, X = x)$$
$$= f(X = x|G = g)P(G = g)$$

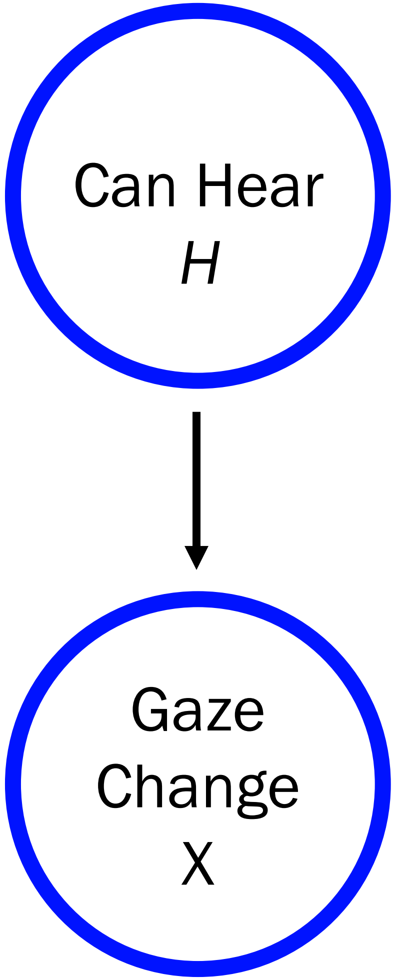
Q: What is P(G = 1 | X = 163)

I Heard That Redux

Normal Assumption: We choose to approximate eye movements with normal distributions. For babies who can hear sounds, we approximate their gaze movement after the sound is played as: $N(\mu = 15, \sigma^2 = 50)$. For babies who can **not** hear sounds, we approximate gaze movement as $N(\mu = 8, \sigma^2 = 50)$.

For a new baby we observe a 14 degree movement after the sound is played. What is your belief that a baby can hear, under The **Normal Assumption**?

I Heard That Redux



$H = 1$ is Bern($p = 0.75$)

$X \mid H = 1$ is $N(\mu = 15, \sigma^2 = 50)$

$X \mid H = 0$ is $N(\mu = 8, \sigma^2 = 50)$

Q: What is $P(H = 1 \mid X = 14)$

Joint is Complete Information!

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

Still true when some variables are continuous

All the Bayes Belong to Us

M,N are discrete. **X, Y are continuous**

OG Bayes

$$P(m|n) = \frac{P(n|m)P(m)}{P(n)}$$

Mix Bayes #1

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

Mix Bayes #2

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

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

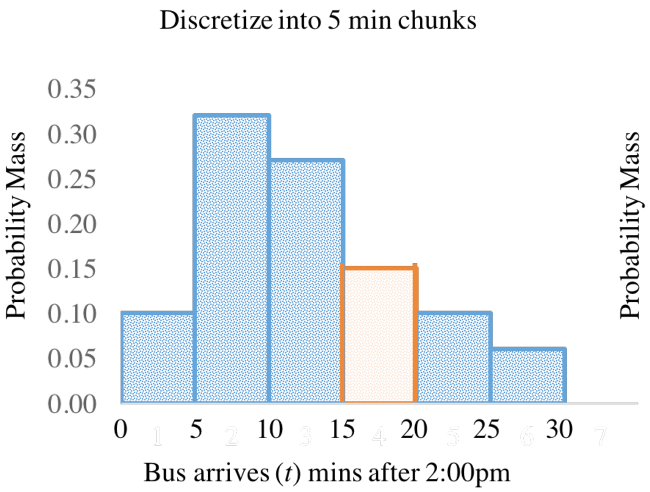
End Review

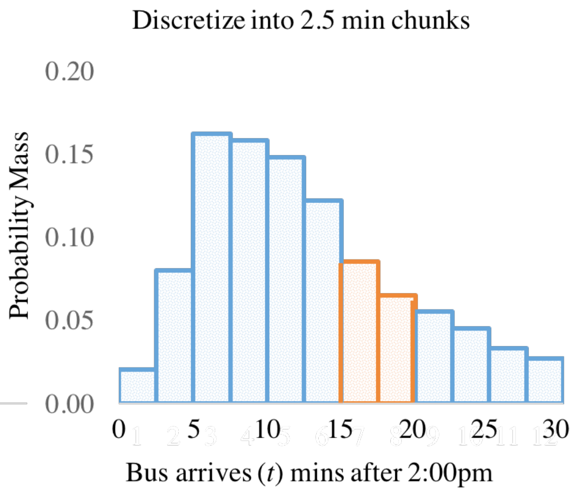
Continuous Joint Distribution

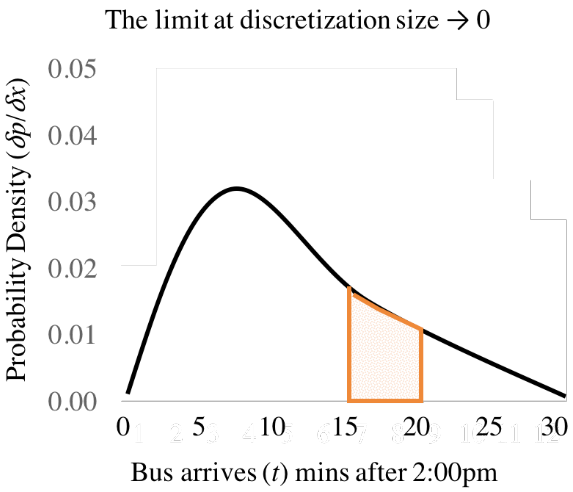
Riding the Marguerite

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)?



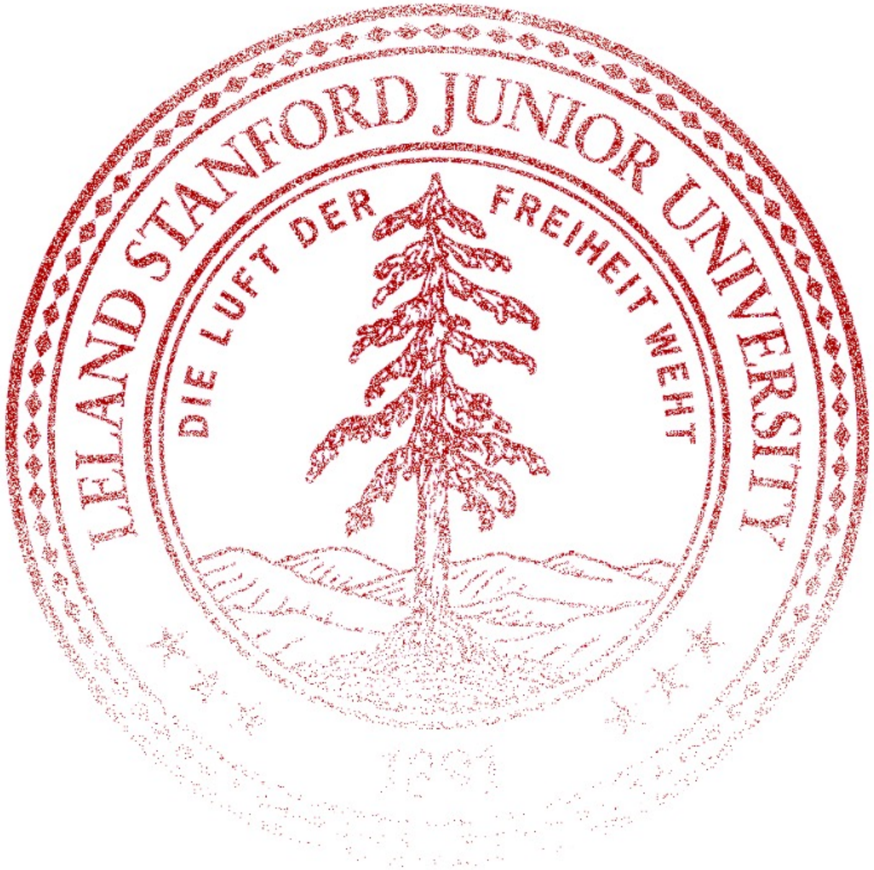




Joint Dart Distribution

Dart Results

P(hit within R pixels of center)?

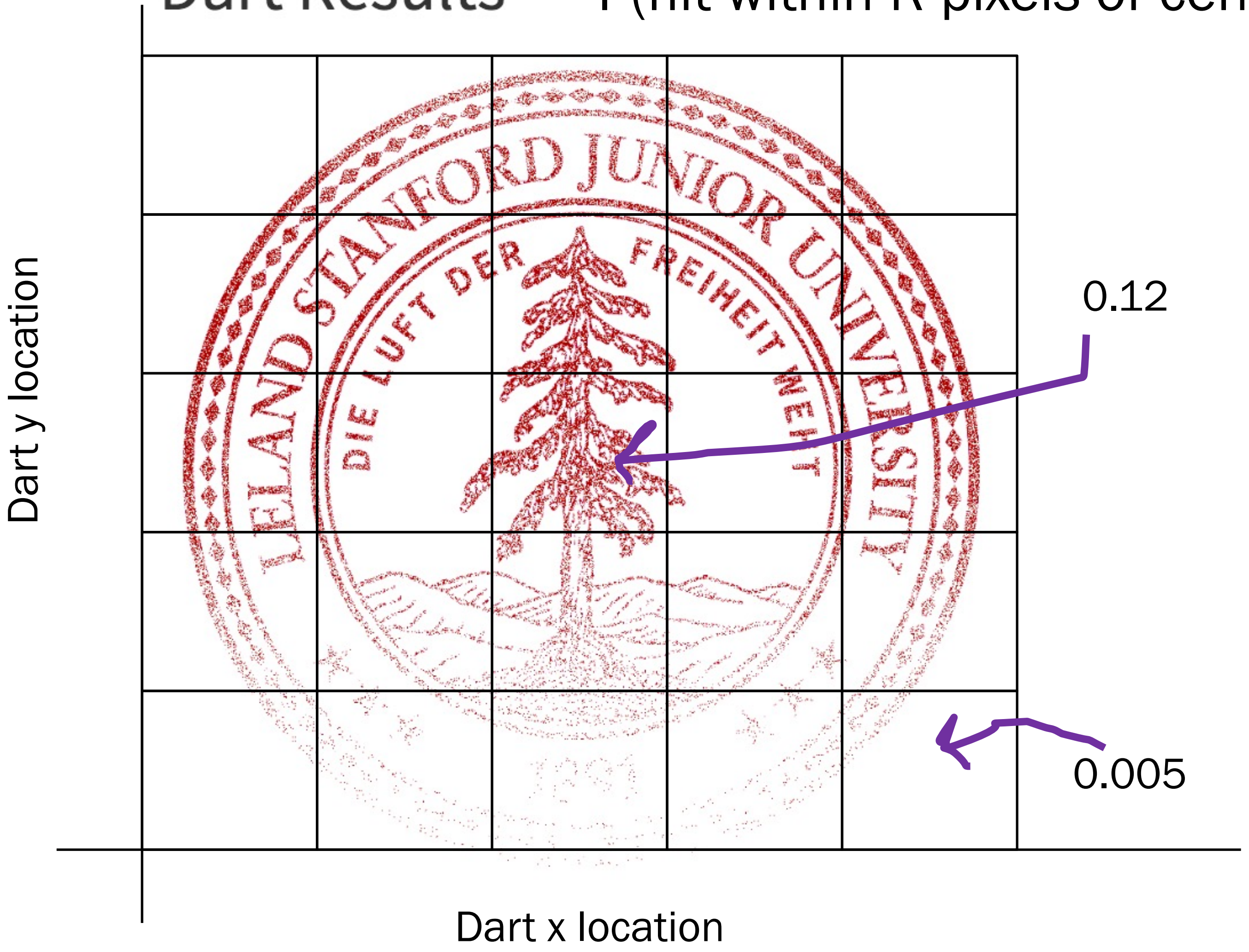


What is the probability that a dart hits at (456.234231234122355, 532.12344123456)?

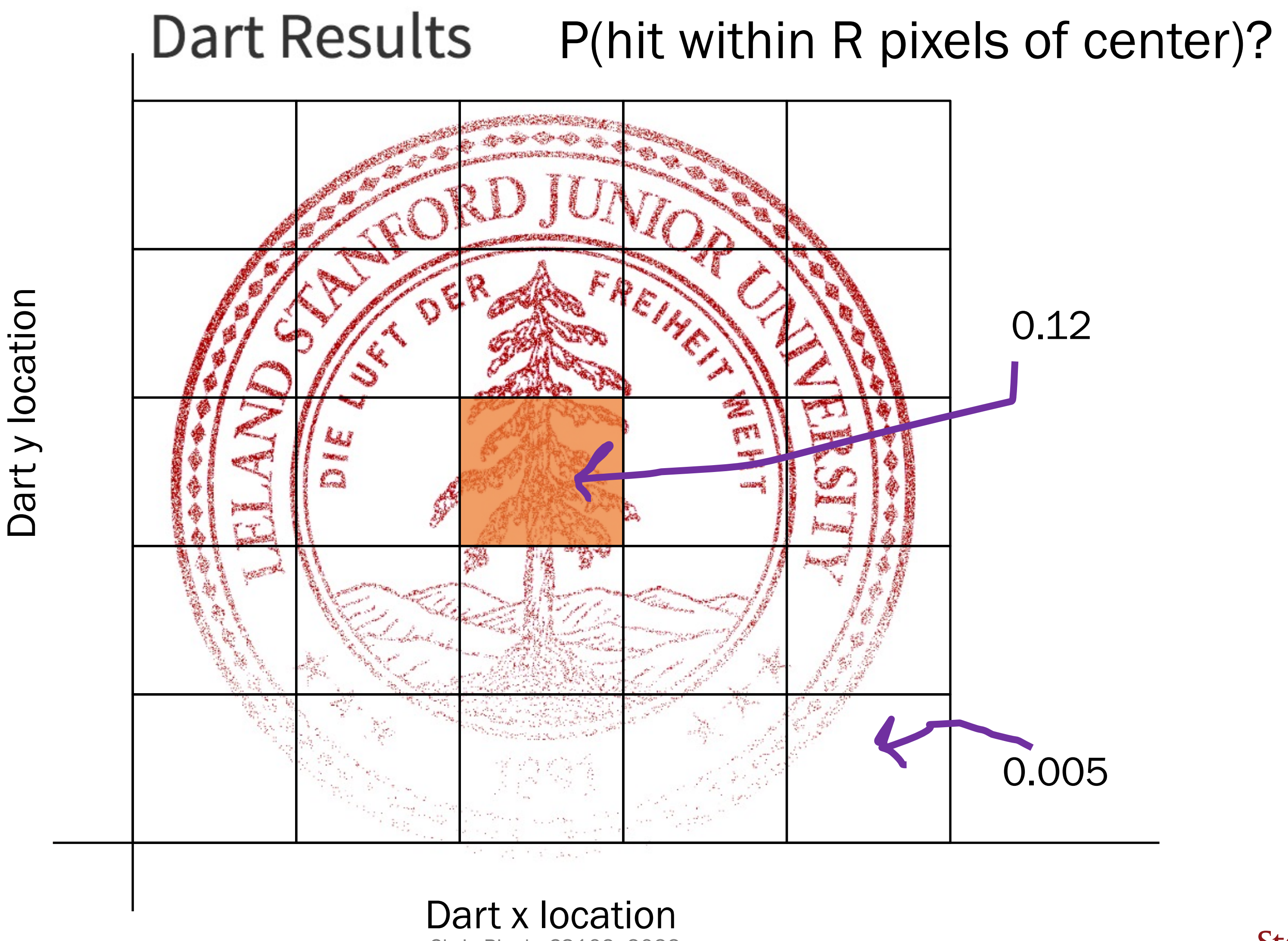
Joint Dart Distribution

Dart Results

P(hit within R pixels of center)?

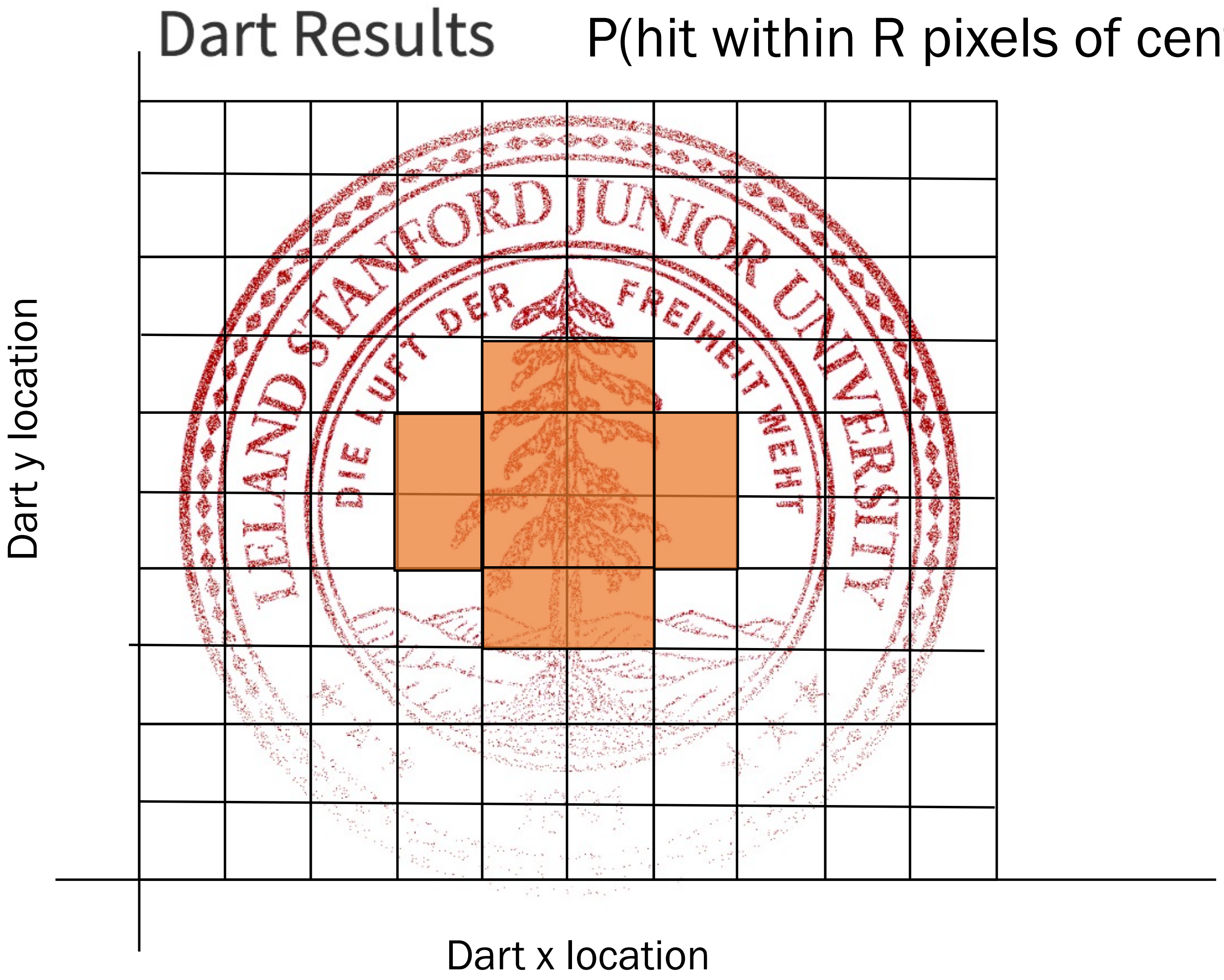


Joint Dart Distribution



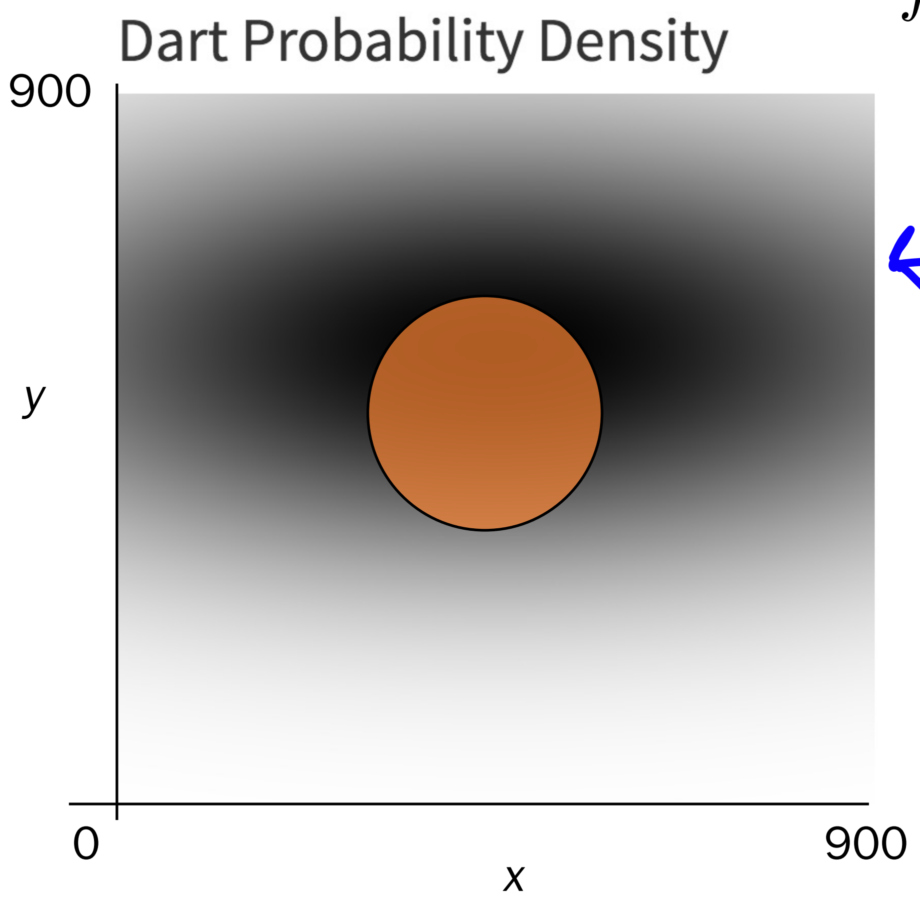
Joint Dart Distribution

P(hit within R pixels of center)?



Joint Dart Distribution

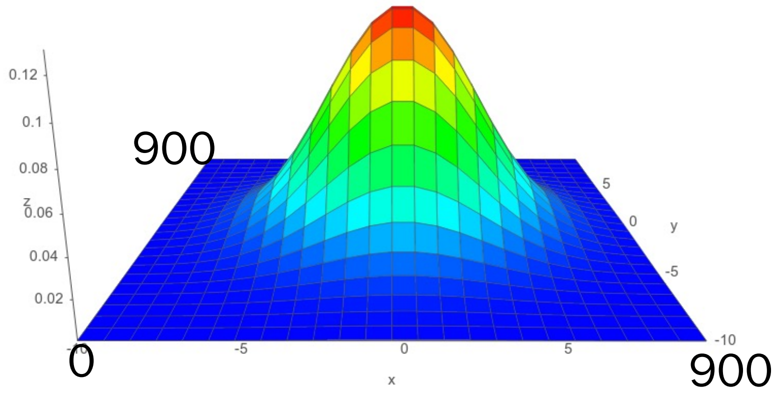
$$f(X = x, Y = y)$$

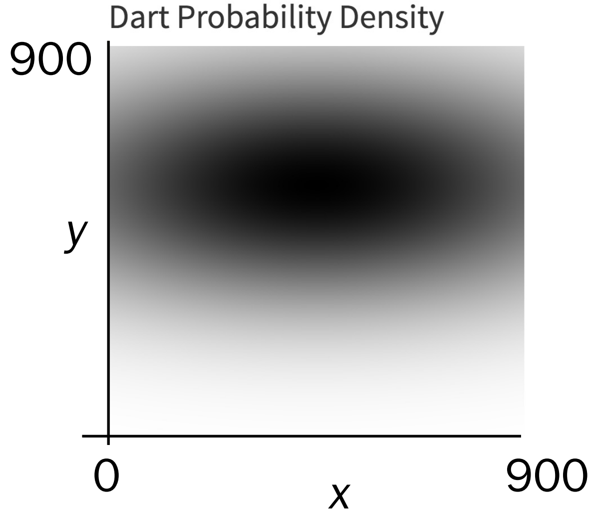


In the limit, as you break down continuous values into intestinally small buckets, you end up with multidimensional probability density

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.



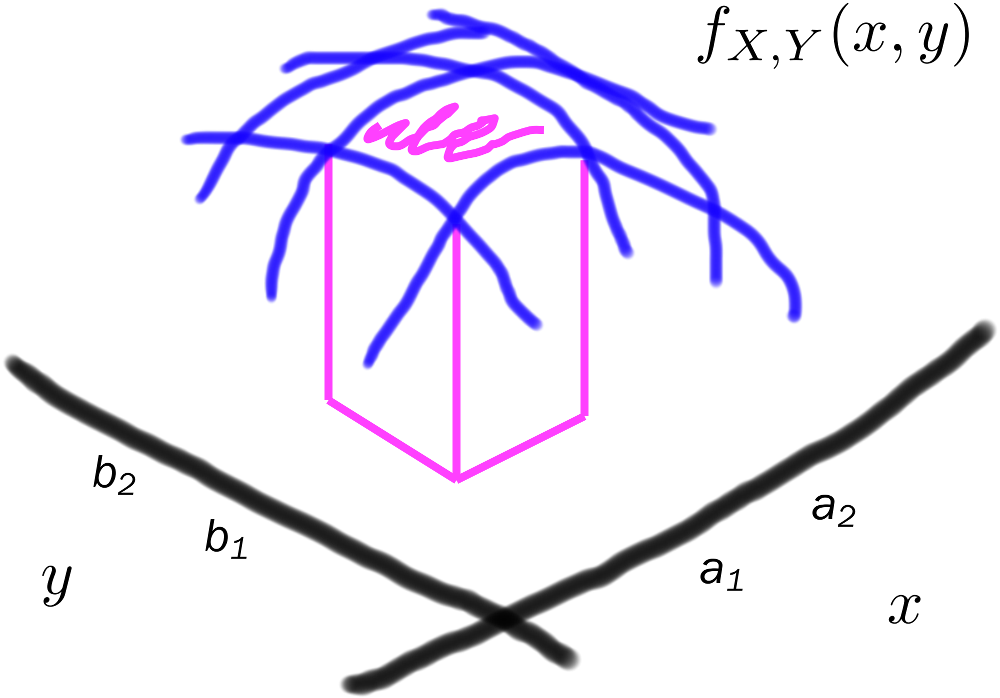


$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$

Joint Probability Density Funciton

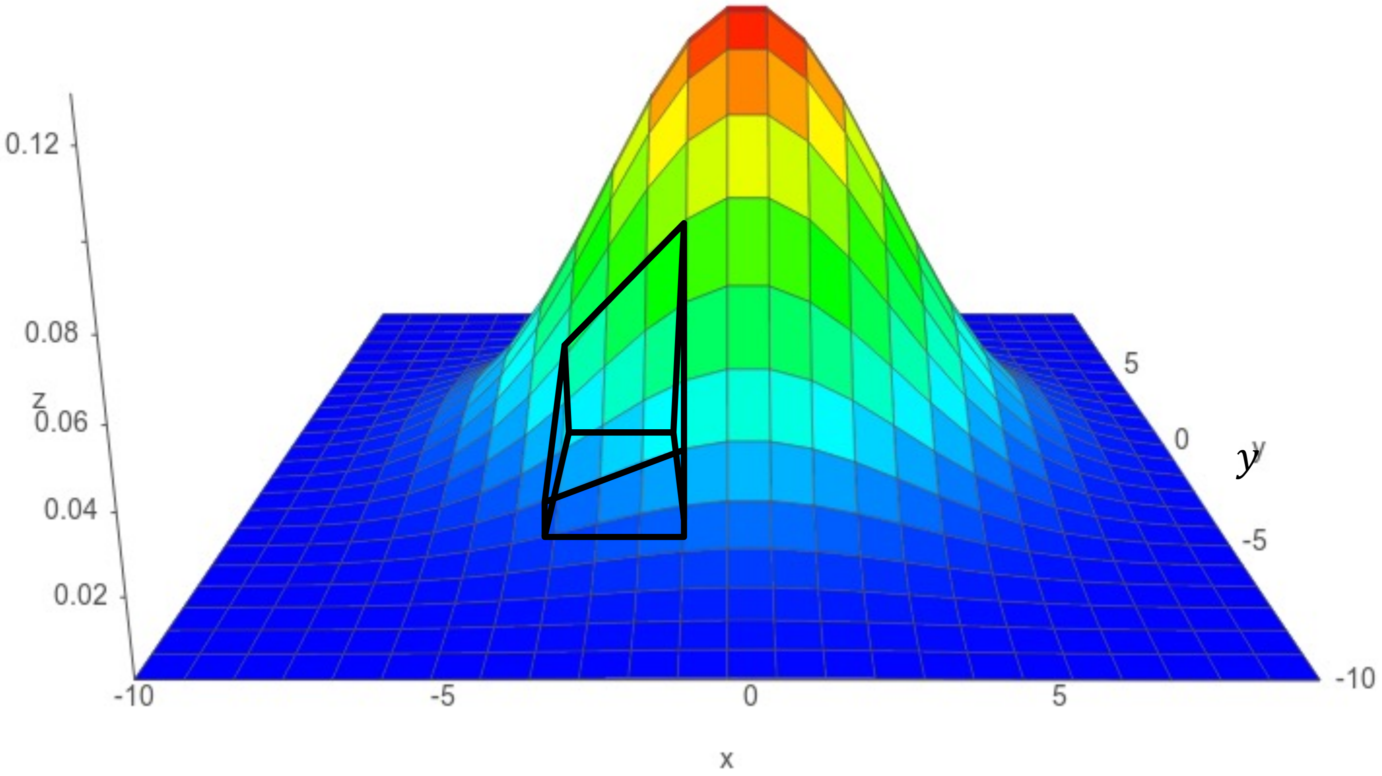
$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$

$f_{X,Y}(x, y)$



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):

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

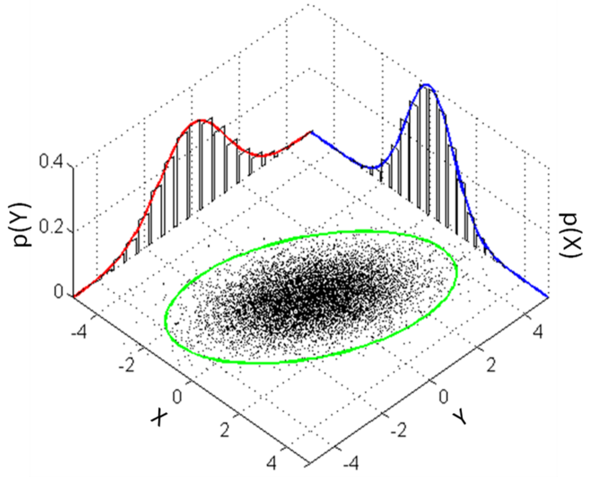
- Then, evaluate remaining (single) integral:

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

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.



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

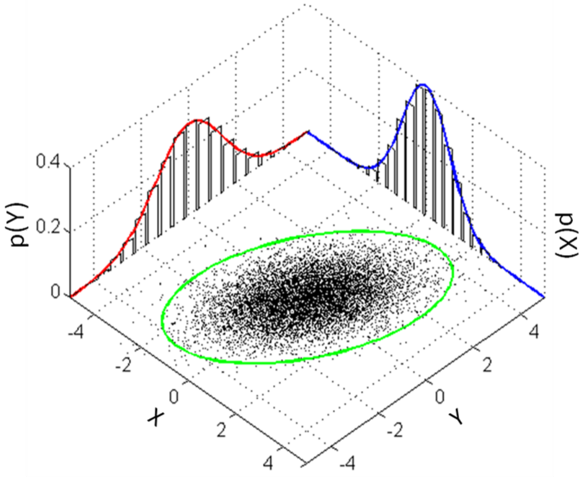
$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y)\ dy$$

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

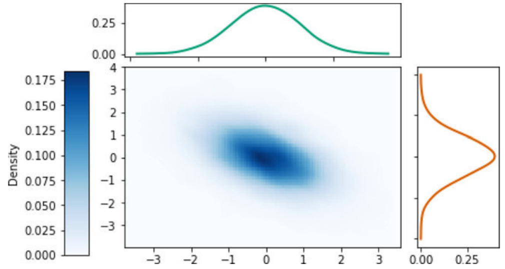
Marginalization

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.



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



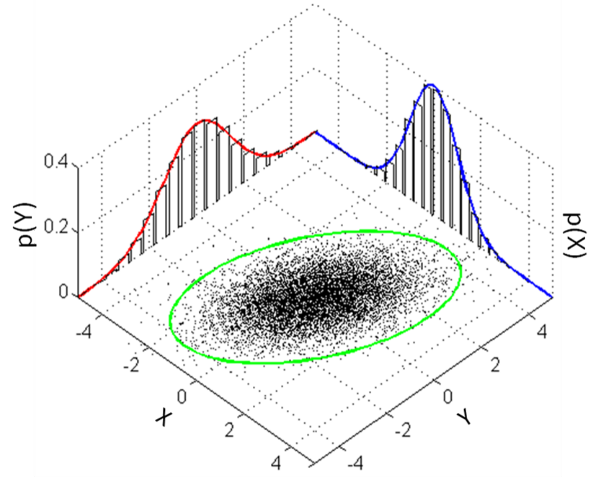
$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y)\ dy$$

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

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.



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

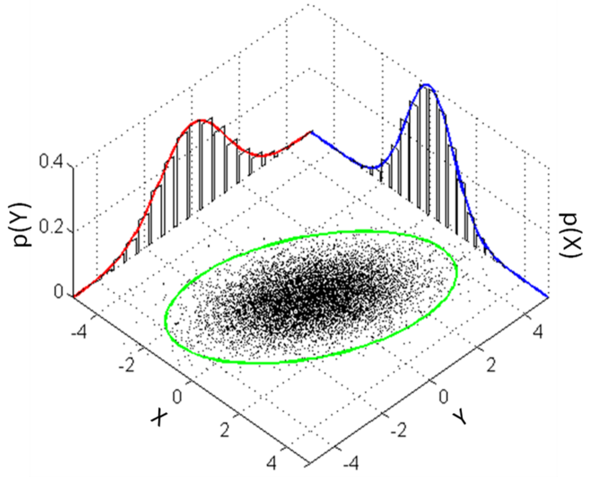
$$f_X(a) = \int_{-\infty}^{\infty} f(X = a, Y = y)$$

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

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.

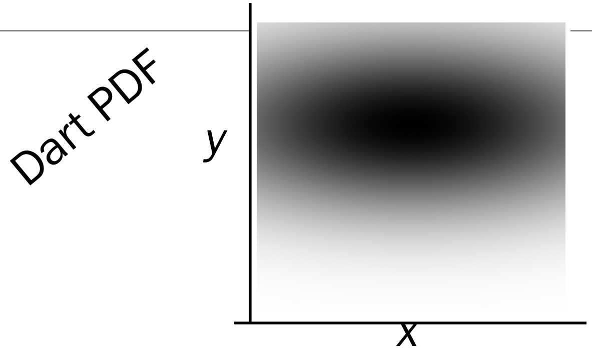


$$P(X = a) = \sum_y P(X = a, Y = y)$$

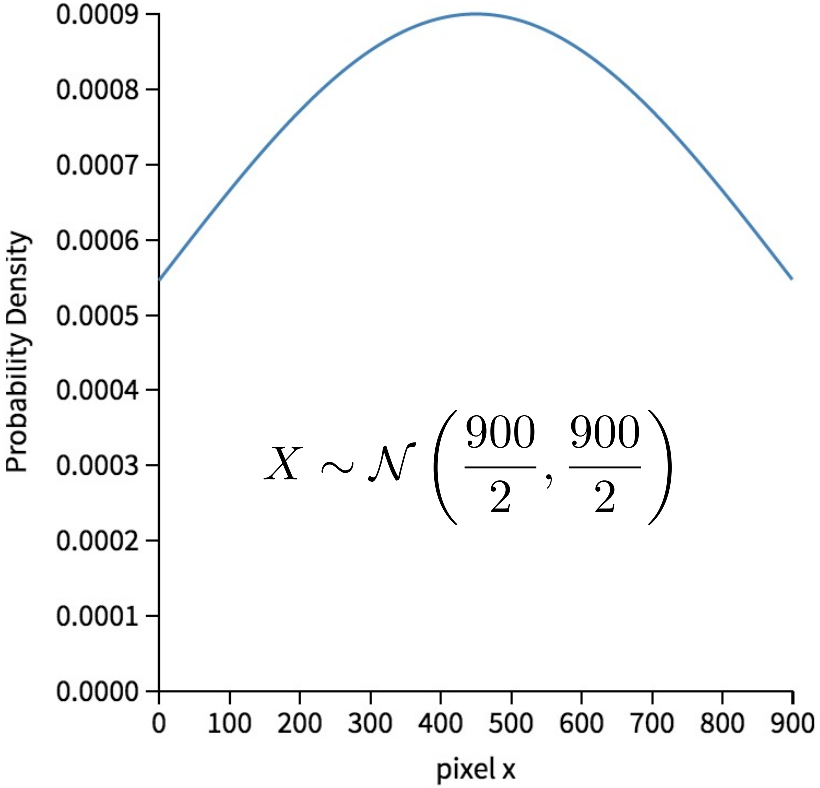
$$f(X = a) = \int_{-\infty}^{\infty} f(X = a, Y = y)$$

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

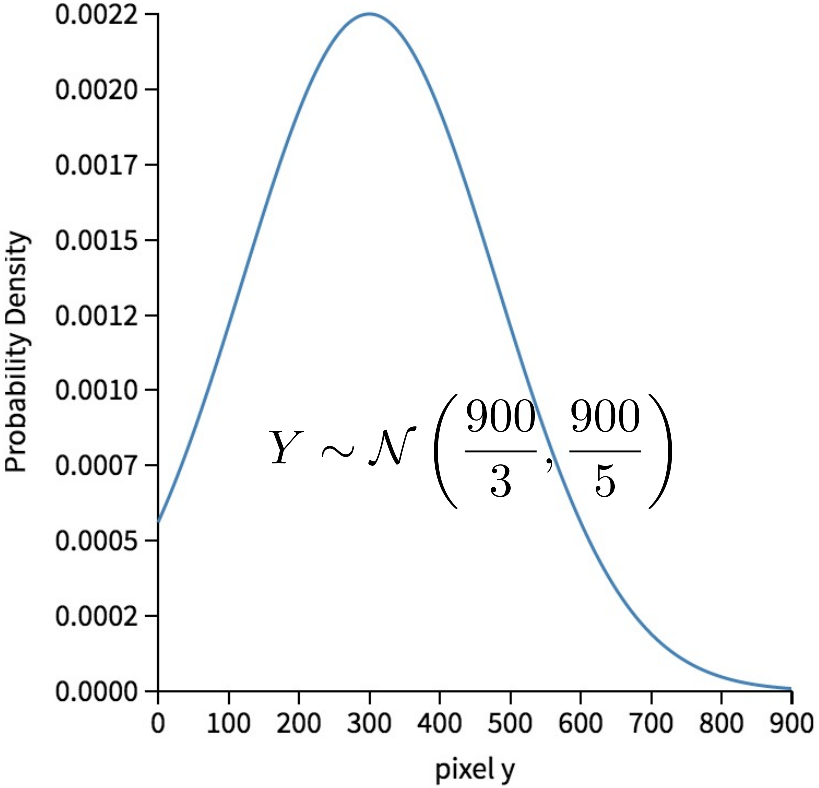
Darts!



X-Pixel Marginal



Y-Pixel Marginal



Inference in the Wild!

Warmup: Bayes Revisited

Posterior belief

Likelihood of evidence

Prior belief

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

Normalization constant

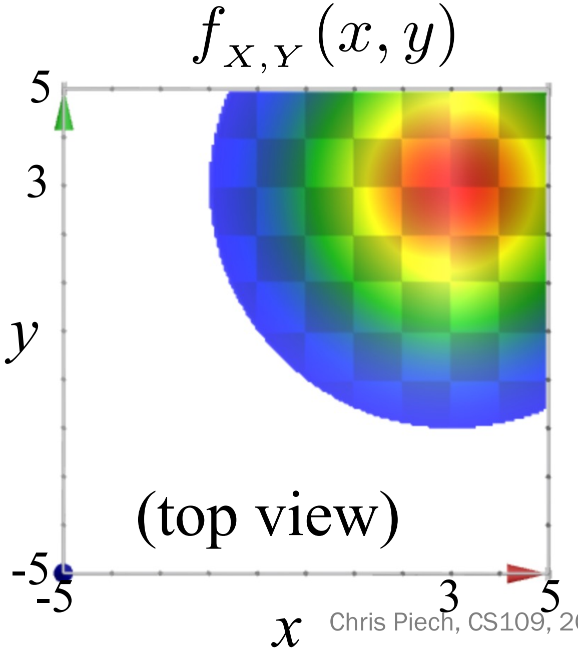
Warmup: Bivariate Normal

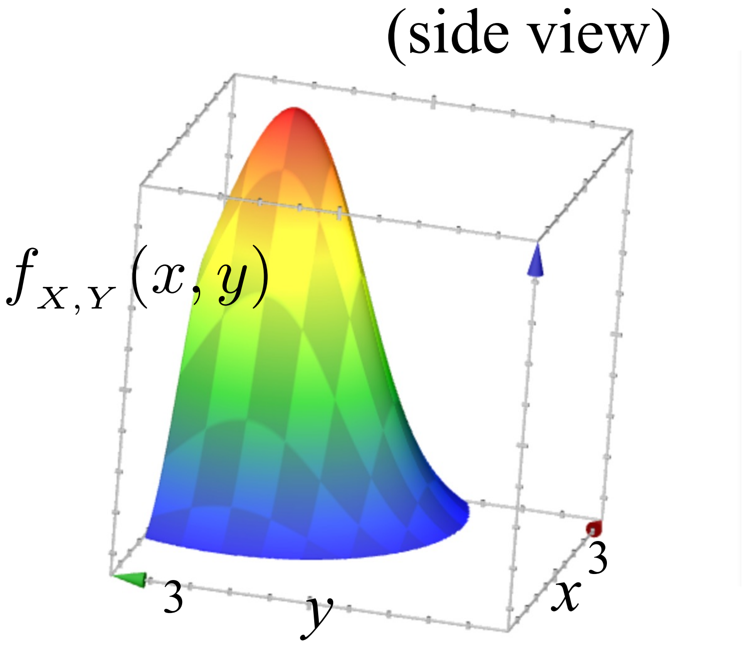
X, Y follow a symmetric bivariate normal distribution if they have joint PDF:

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

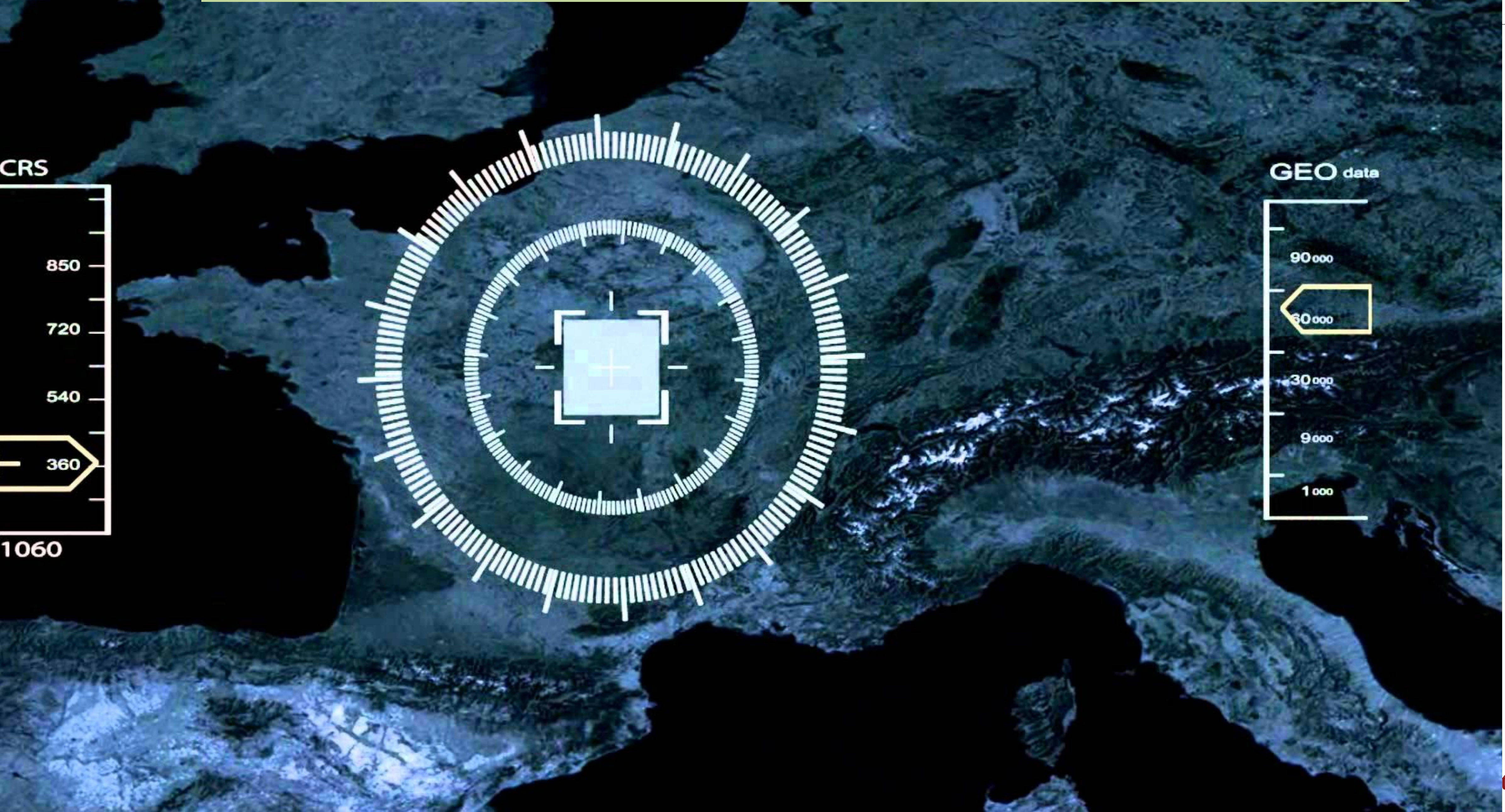
Here is an example where:

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





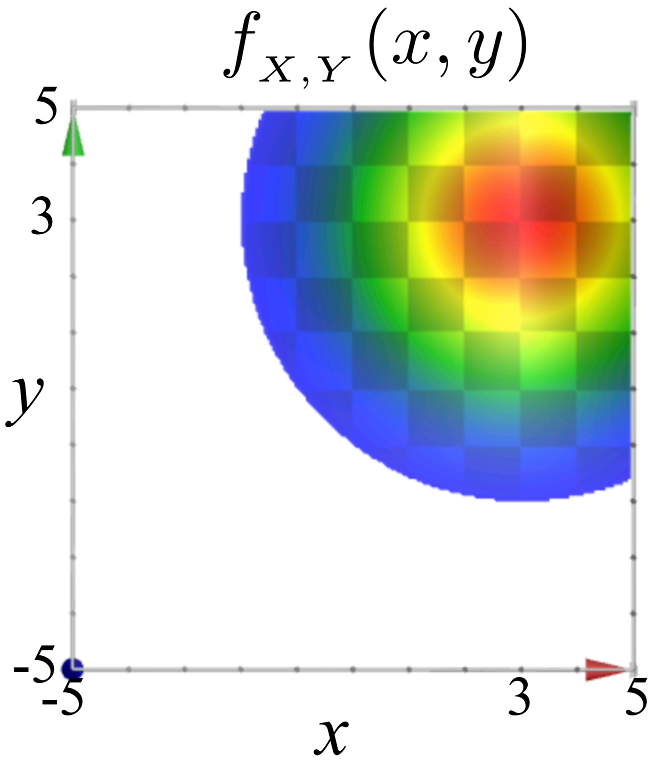
Tracking in 2D Space?



ty

Tracking in 2D Space: Prior

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



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

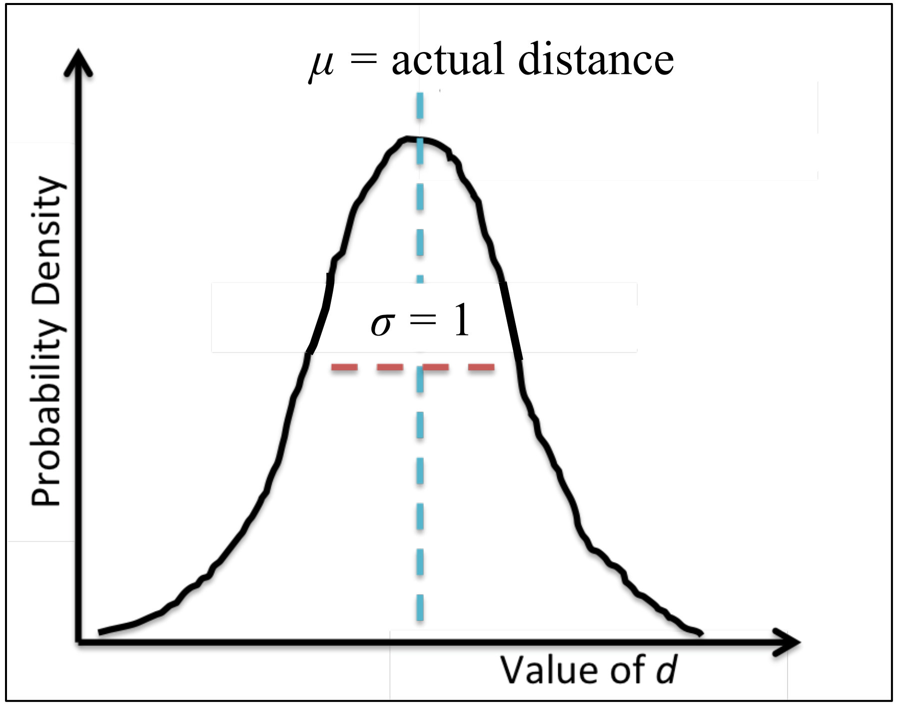
Prior belief with K: $$f_{X,Y}(x,y) = K \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

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…

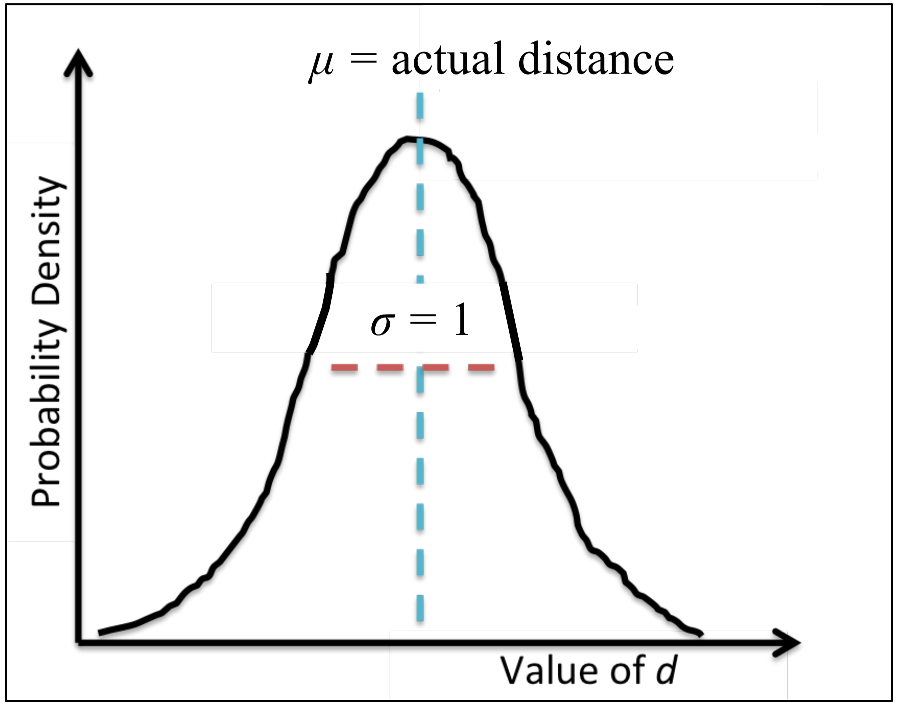
f(D | X, Y) is knowable!



Tracking in 2D Space: Observation!

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

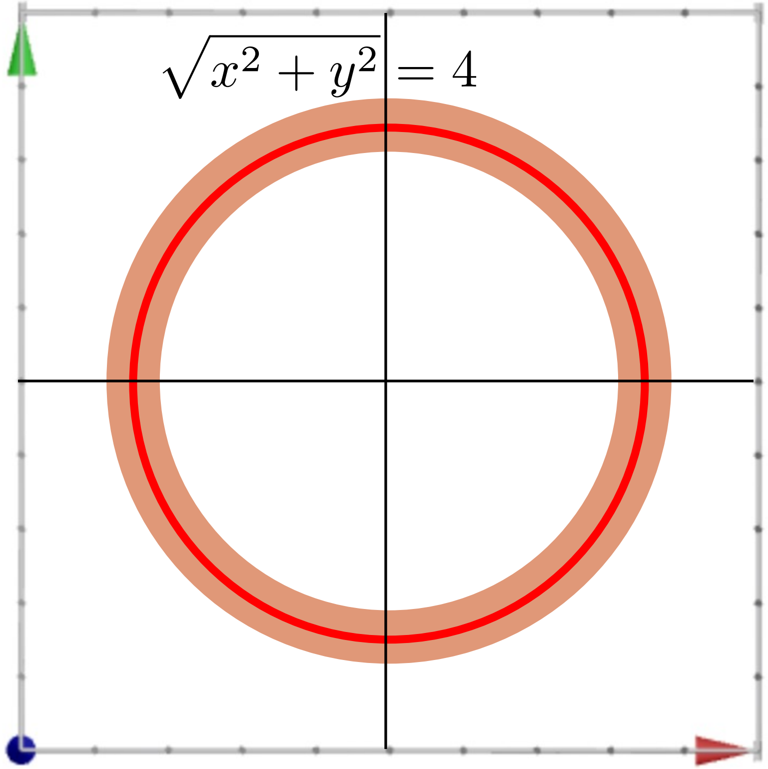
$$D|X,Y \sim N(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)$$

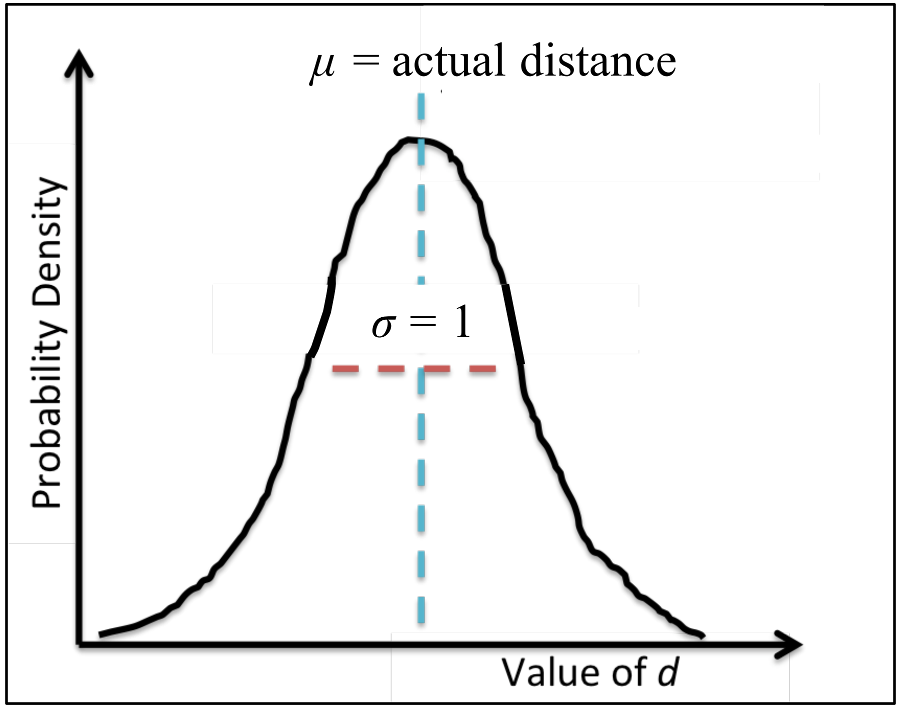


Know that the distance of a ping is normal with respect to the true distance.

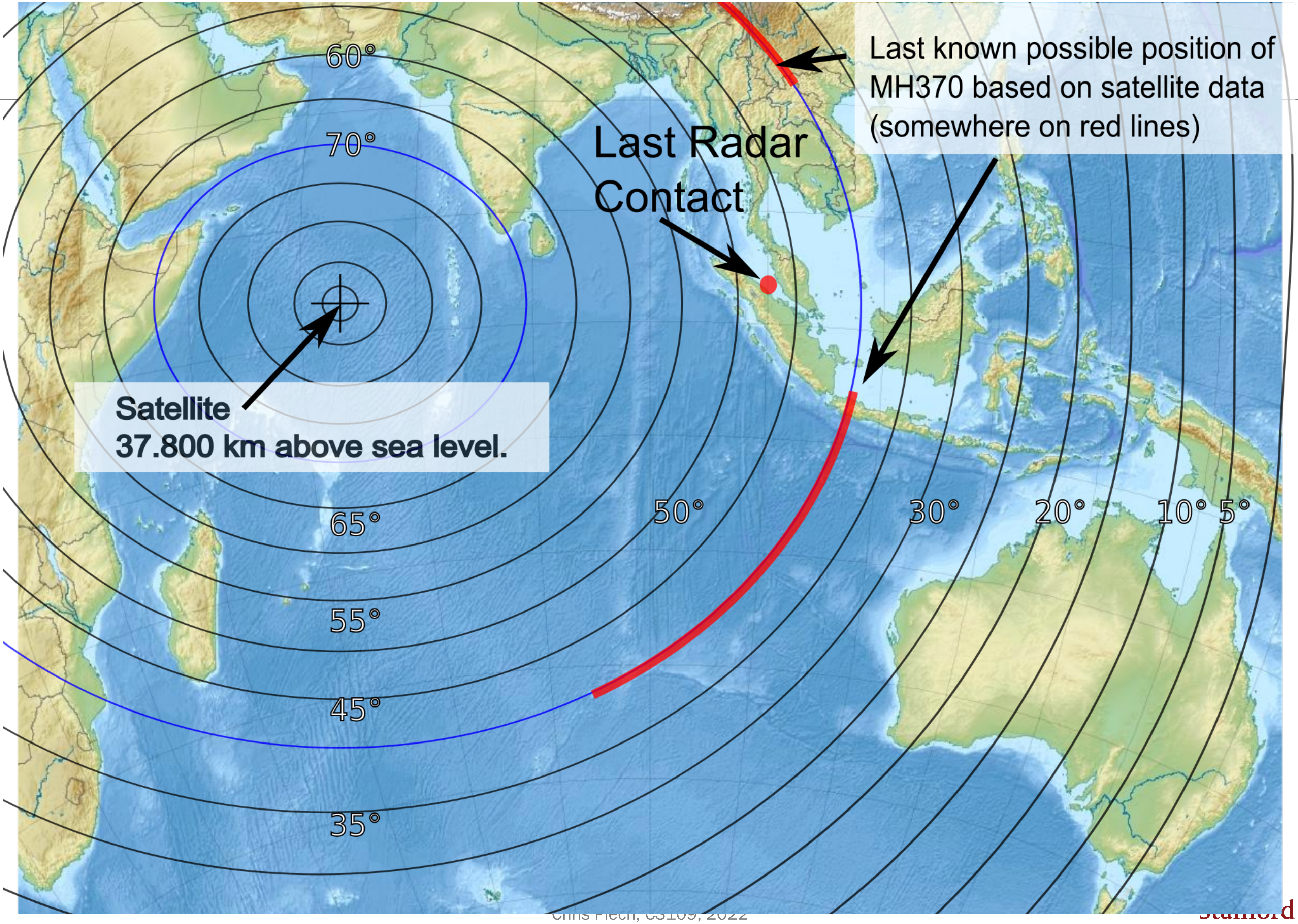
Tracking in 2D Space: Observation!

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





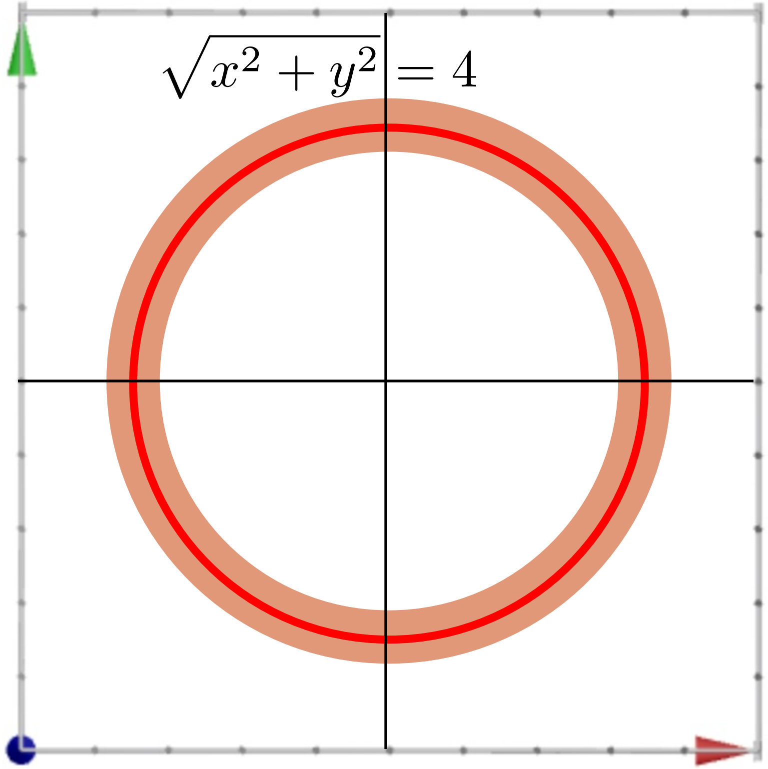
Know that the distance of a ping is normal with respect to the true distance

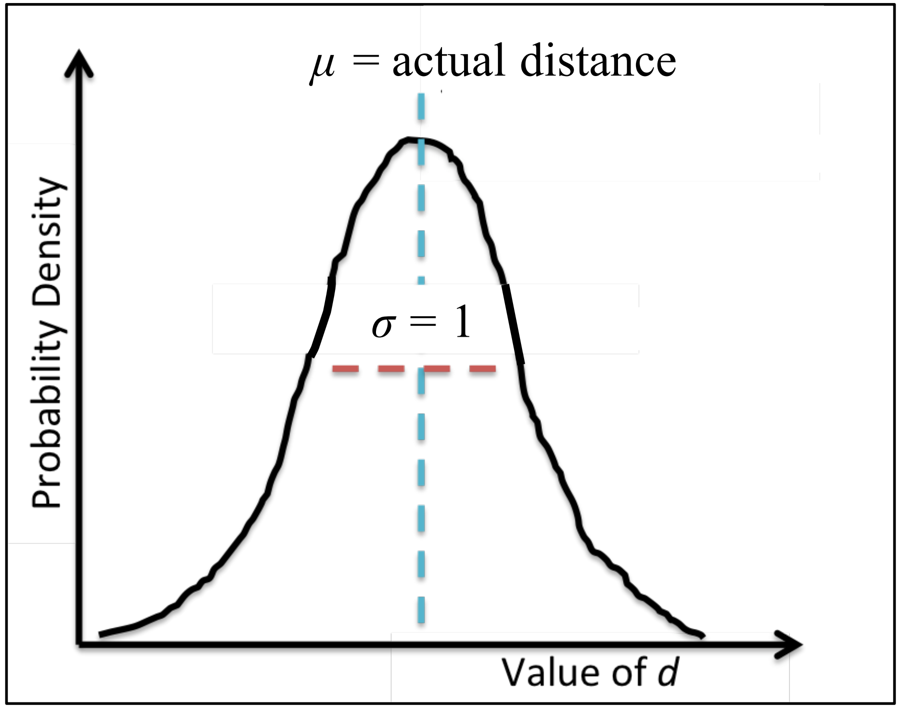


Last known possible position of MH370 based on satellite data (somewhere on red lines)
Last Radar Contact
Satellite
37.800 km above sea level.
60°
70°
65°
55°
45°
35°
50°
30°
20°
10°
5°

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!

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

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

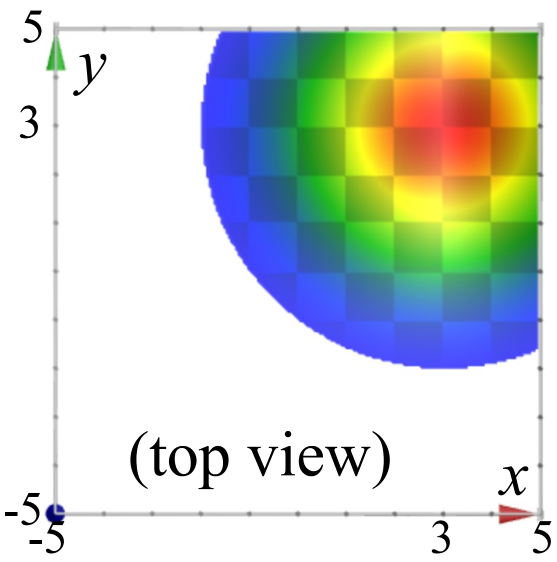
$$= \frac{1}{\sqrt{2\pi}} e^{\frac{-(d-\mu)^2}{2}}$$

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

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

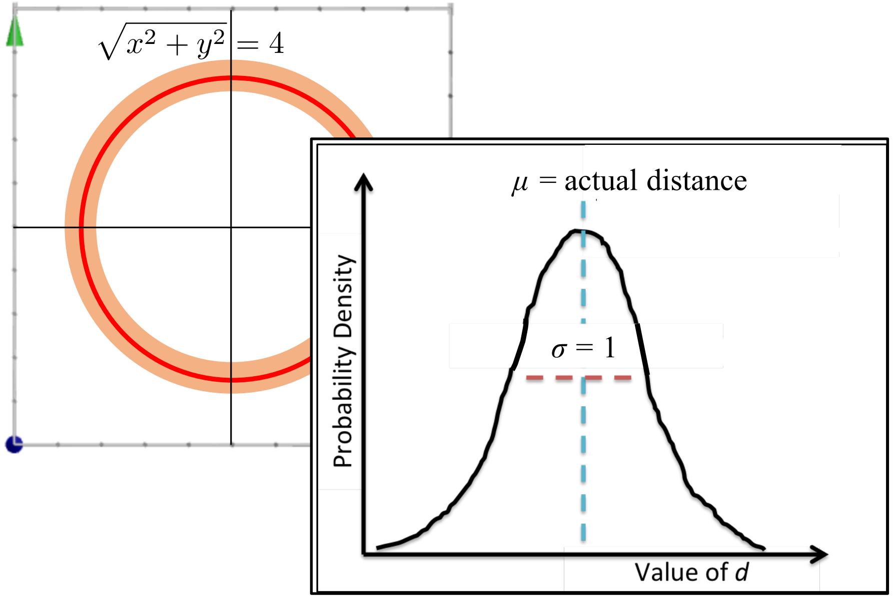
Tracking in 2D Space: New Belief

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



Prior

Observation



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

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

Prior

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

Observation

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

Tracking in 2D Space: New Belief

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

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

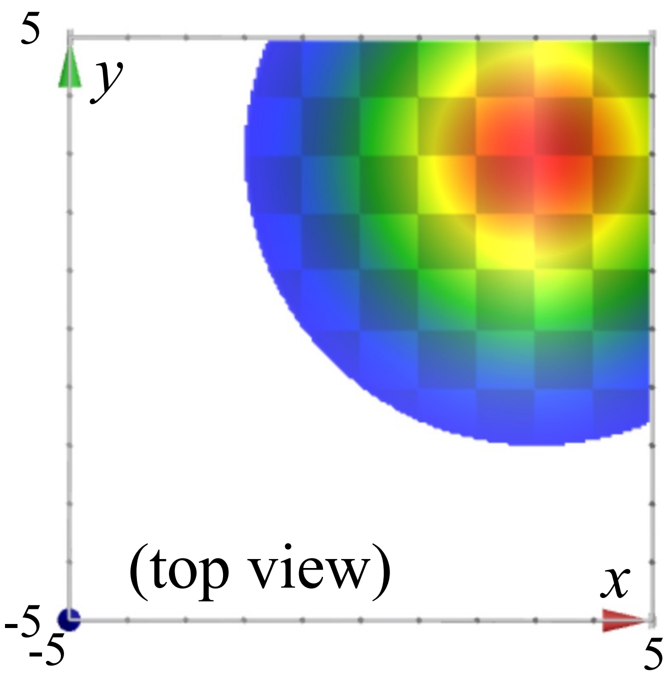
$$=\frac{K_3\cdot e^{-\left[\frac{[4-\sqrt{x^2+y^2})^2]}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]}}{f(D=4)}$$

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

For your notes…

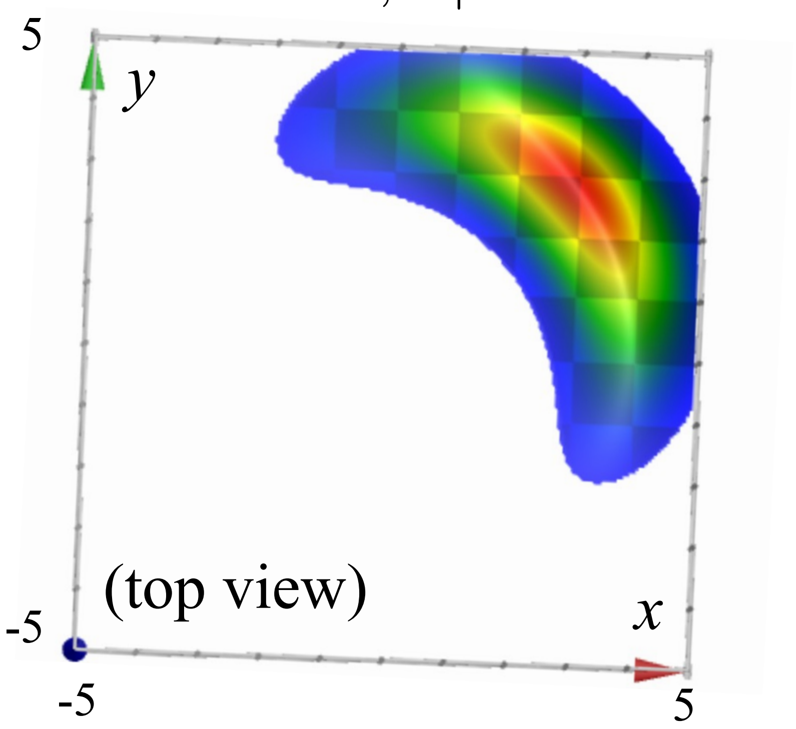
Tracking in 2D Space: Posterior

$f_{X,Y}$



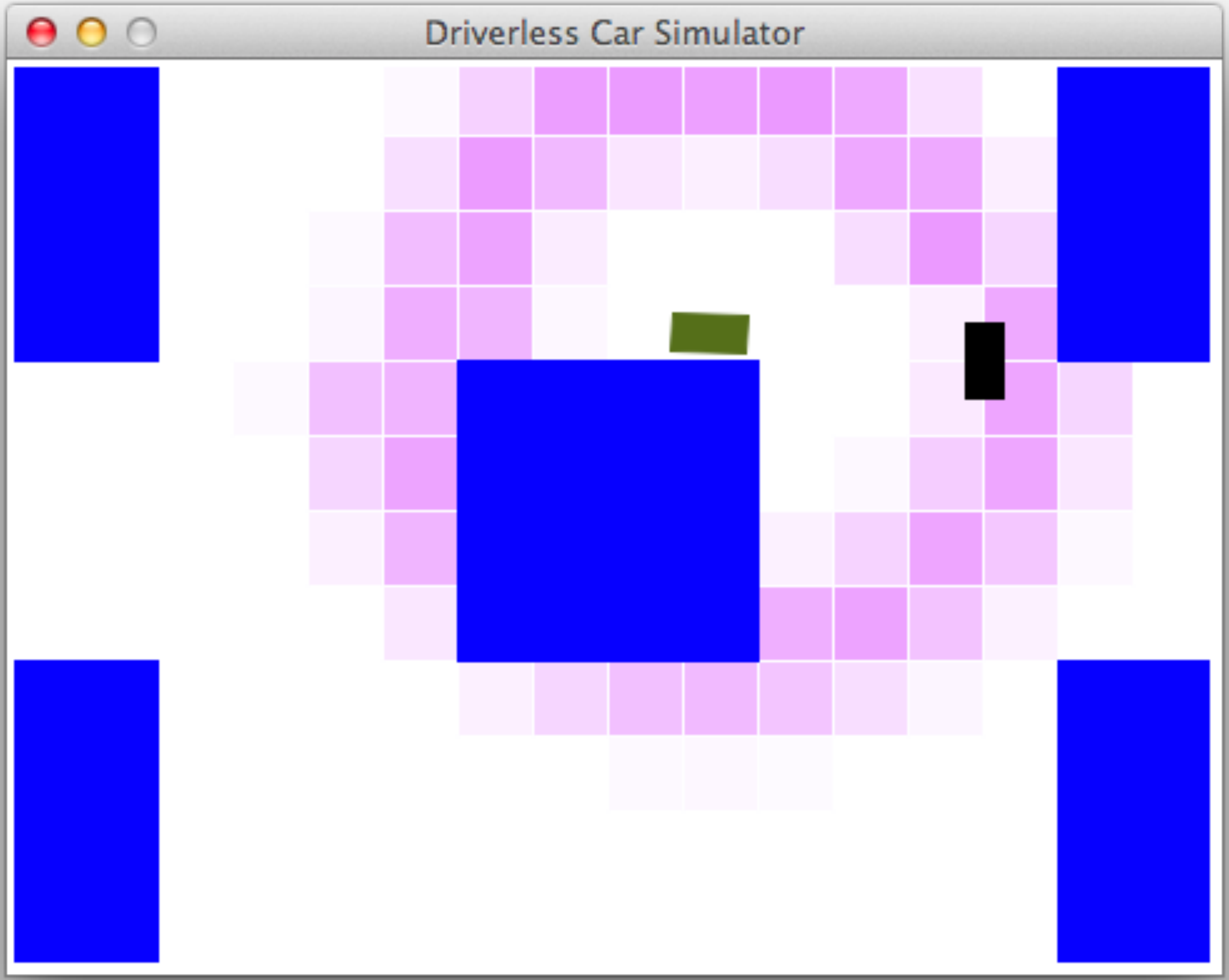
Prior

$f_{X,Y|D}$



Posterior

Tracking in 2D Space: CS221



Joint Random Variables

- [x] Use a joint table, or joint function to solve probability question
- [x] 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