Continuous Inference

Chris Piech
CS109, Stanford University

<Review>

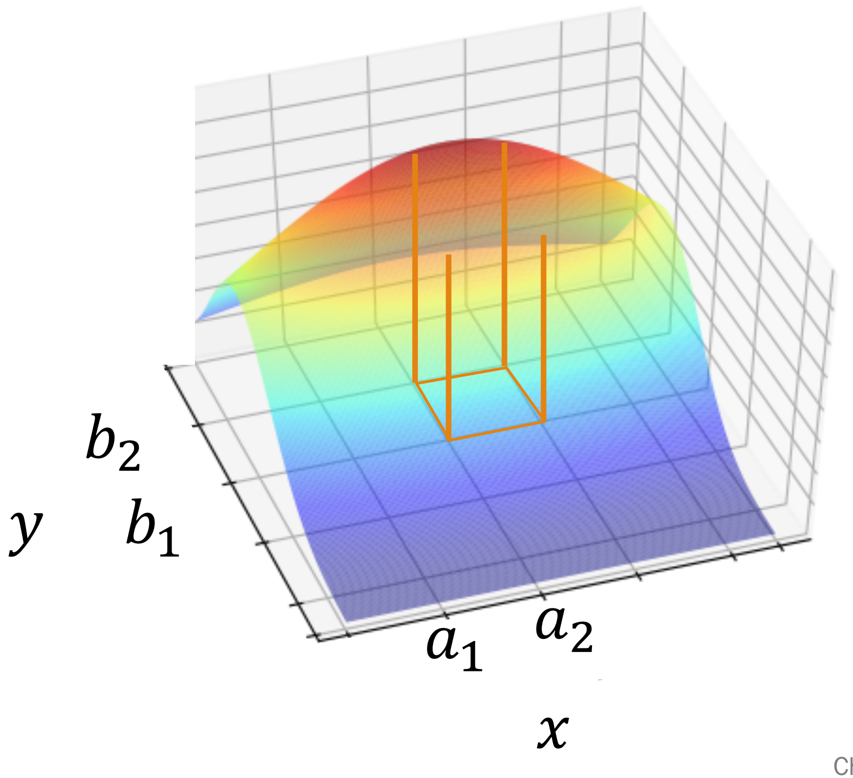
PDFs of Multiple Continuous Variables

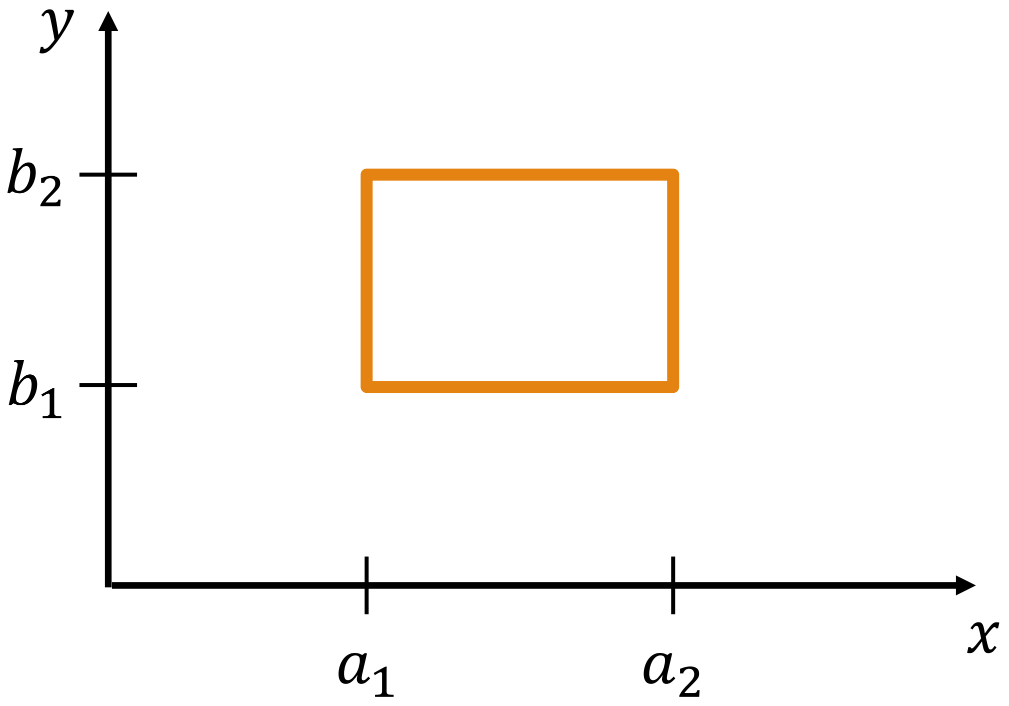
If two random variables X and Y are jointly continuous, then there exists a joint probability density function $f_{X,Y}$ defined over $-\infty < x, y < \infty$ such that:

$$P\left(a_1 \leq X \leq a_{2,} \;\; b_1 \leq Y \leq b_2\right) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y) dy\, dx$$

CDFs of Multiple Continuous Variables

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





Independence of Multiple Continuous Variables

Two continuous random variables X and Y are independent if:

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \qquad \forall x, y$$

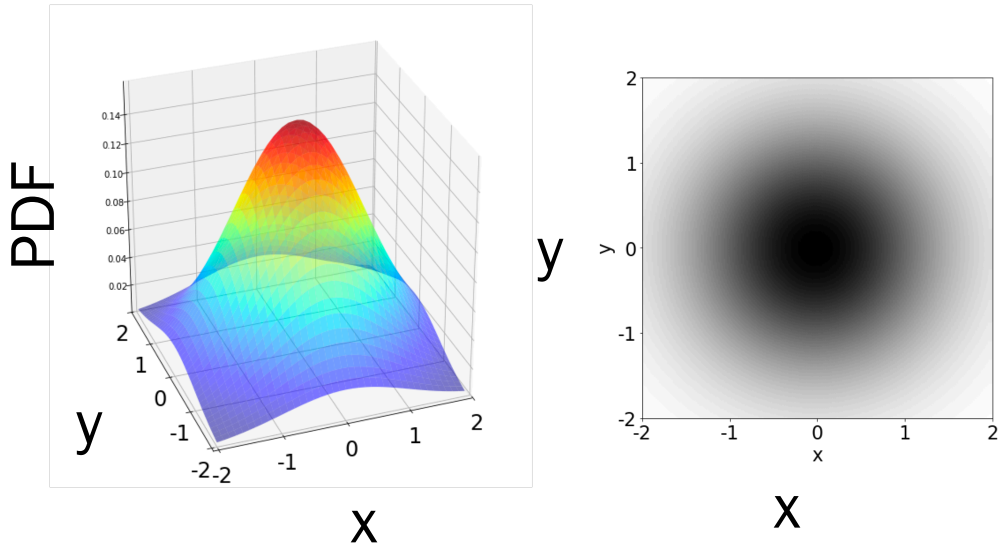
Equivalently:

$$\boxed{\begin{aligned} F_{X,Y}(x, y) &= F_X(x)F_Y(y) \\ f_{X,Y}(x, y) &= f_X(x)f_Y(y) \end{aligned}} \qquad \forall x, y$$

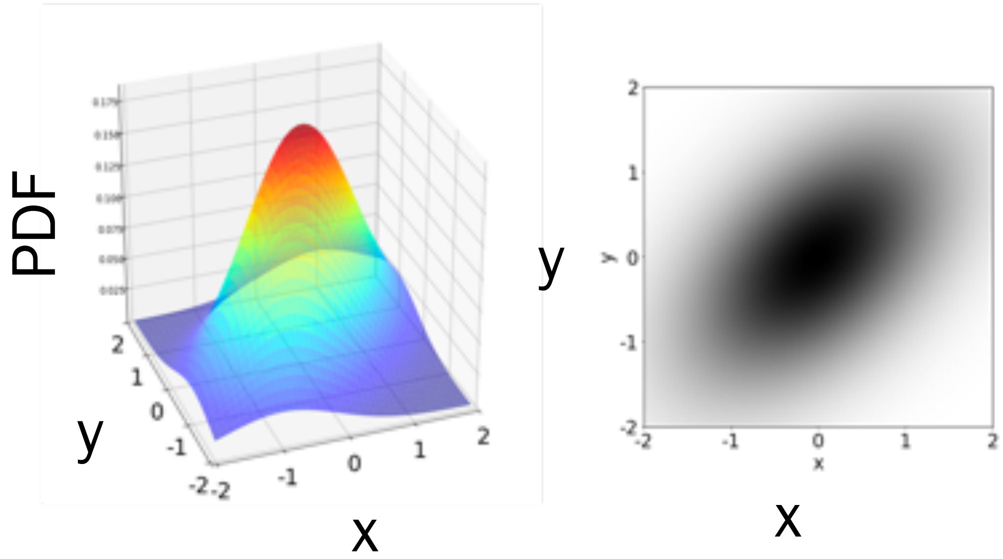
</Review>

Bivariate Gaussians

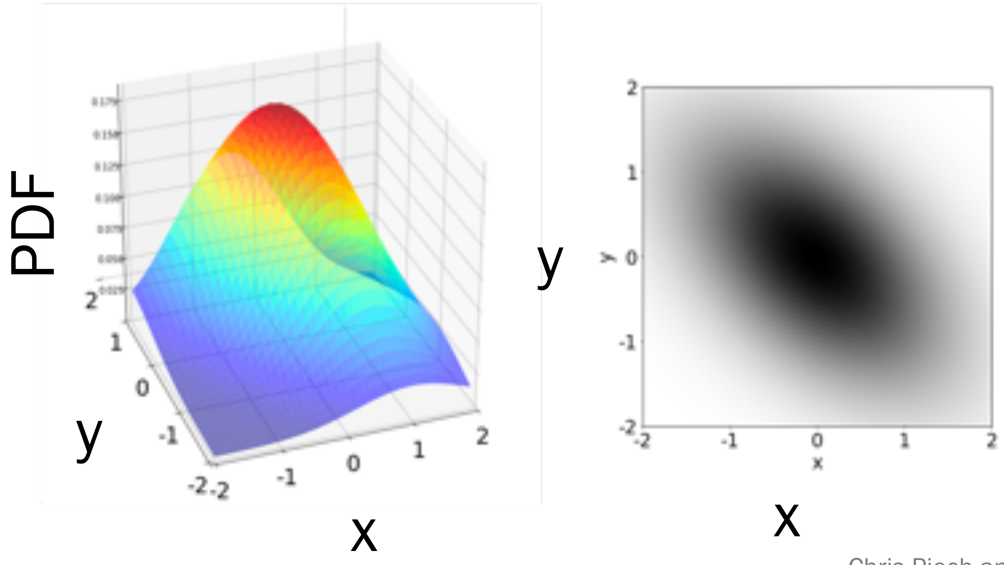
1.



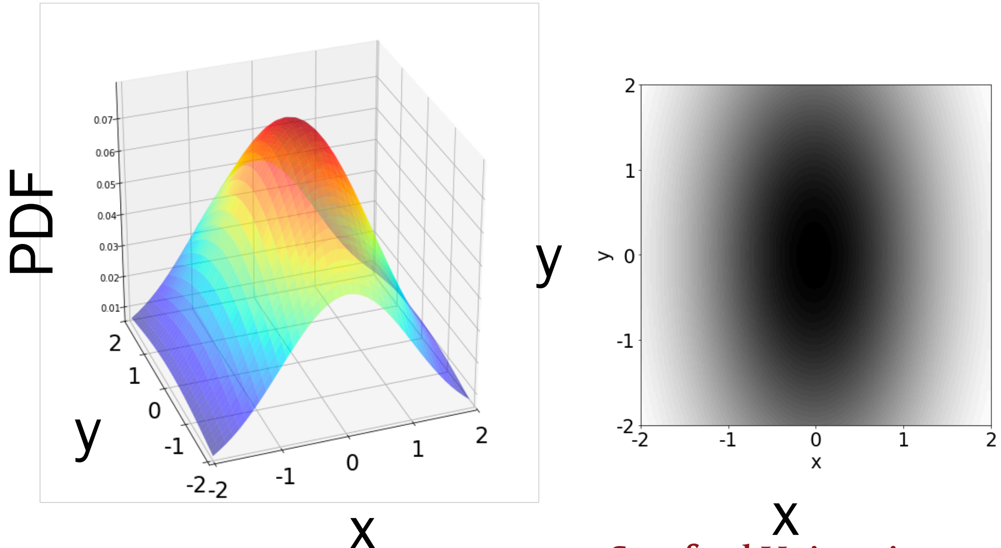
2.



3.



4.



Bivariate Gaussians

Two continuous variables

Bivariate Gaussian

Variance of each RV

$$X_1, X_2 \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

Means of each RV

Correlation of the two RVs

Sometimes written equivalently as:

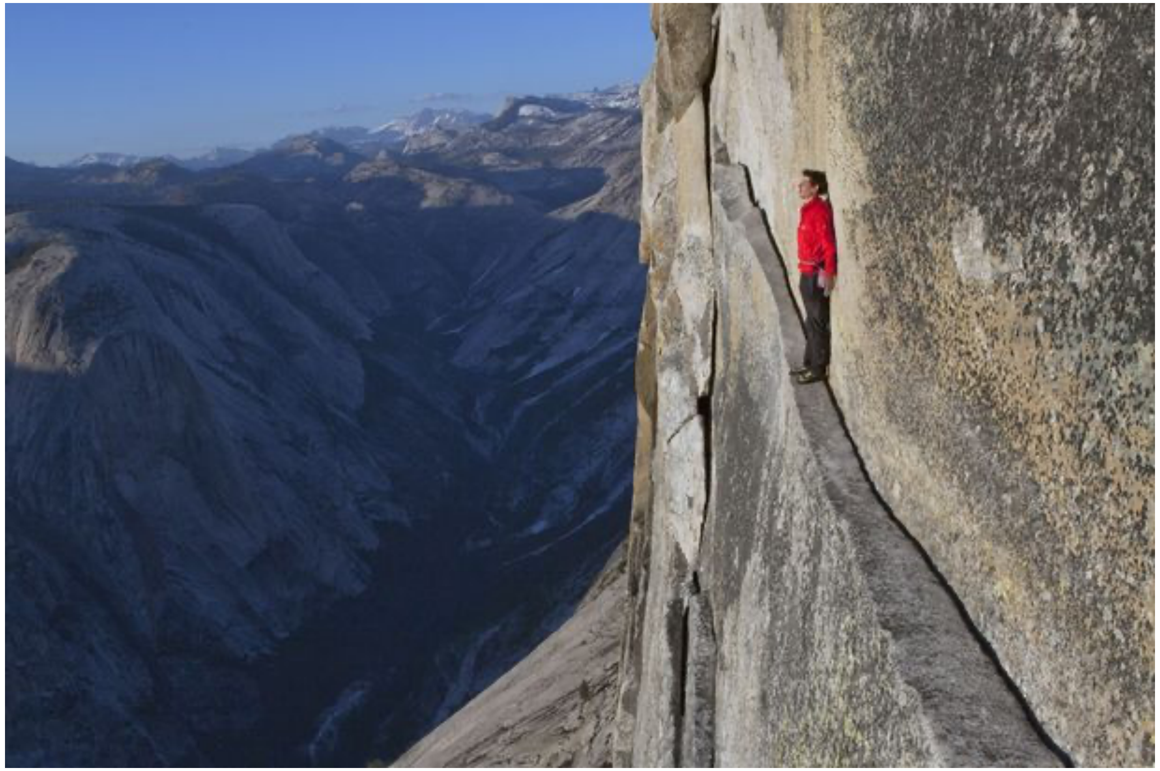
$$\vec{X} \sim N_k(\vec{\mu}, \vec{\Sigma})$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Bivariate Gaussians PDF

X_1 and X_2 follow a bivariate normal distribution if their joint PDF f is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$



But what if $\rho = 0$?

X_1 and X_2 follow a bivariate normal distribution if their joint PDF f is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$f(x_1, x_2) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x_1-\mu_1)^2/2\sigma_1^2} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x_2-\mu_2)^2/2\sigma_2^2}$$

Independent Multivariate Gaussians

X_1 and X_2 are **independent** with marginal distributions. Or equivalently if ρ = 0, $X_1 \sim \mathcal{N}(\mu_1\ \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2\ \sigma_2^2)$

Joint PDF

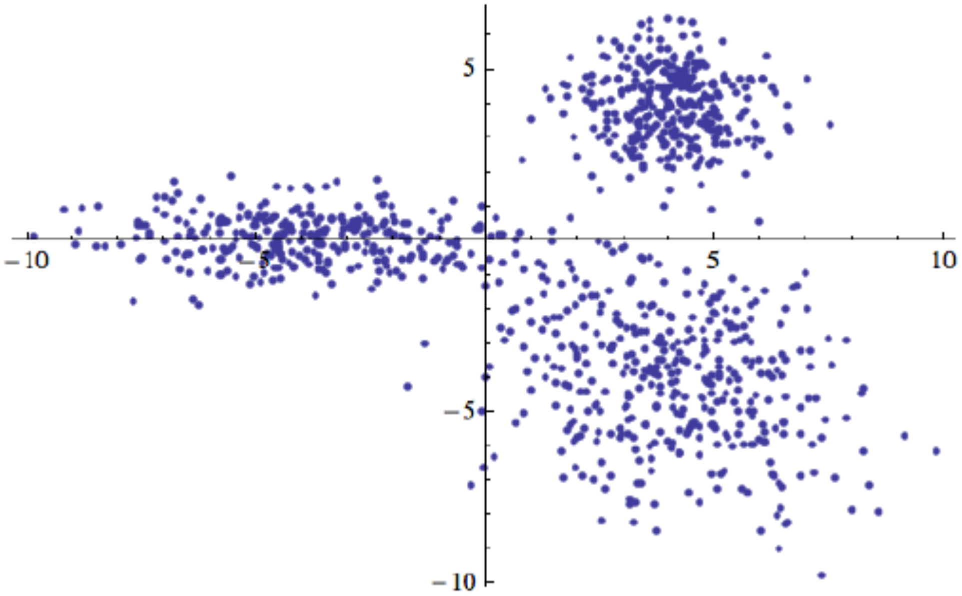
$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

Joint CDF

$$F(x_1, x_2) = \Phi\left(\frac{x_1 - \mu_1}{\sigma_1}\right) \cdot \Phi\left(\frac{x_2 - \mu_2}{\sigma_2}\right)$$

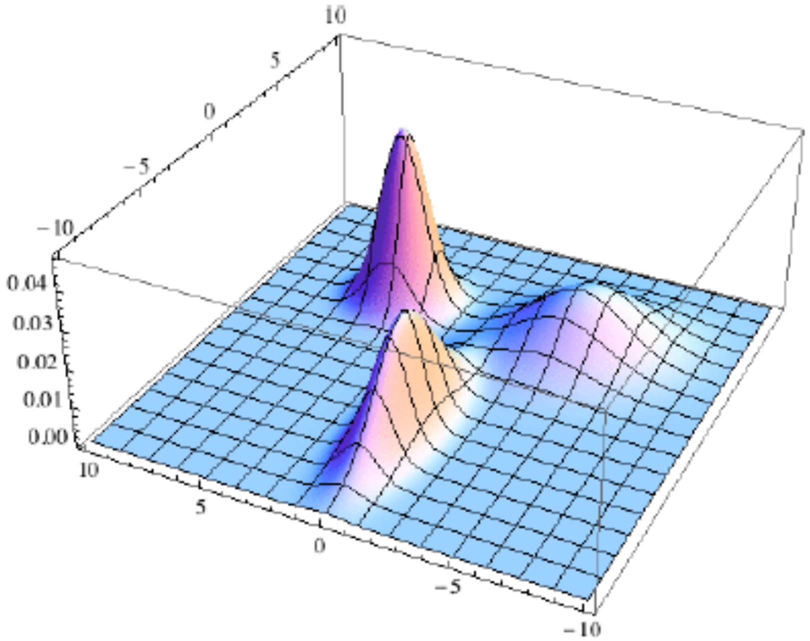
Multivariate Normal Distributions are Rad ☺

The Data



A 2D scatter plot with three "clusters"

The Joint Model



A mixture of 3 bi-variate Normal distributions

See you in CS221, Gaussian Mixture Models

Probability with Instagram!

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

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 blur

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

In a Gaussian blur, for every pixel:

- Weight each pixel by the probability that X and Y are both within the pixel bounds
- The weighting function is a Bivariate Gaussian (Normal) standard deviation parameter σ

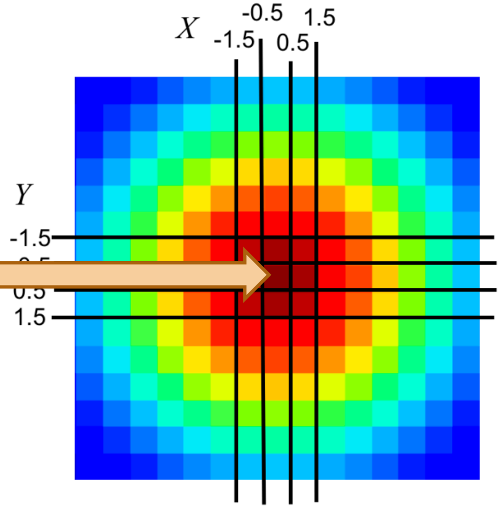
Weight matrix:

Center pixel: (0, 0)

Pixel bounds:

$-0.5 < x \leq 0.5$

$-0.5 < y \leq 0.5$



→ Independently: $X \sim \mathcal{N}(0, 3^2), Y \sim \mathcal{N}(0, 3^2)$

→ Jointly: $X, Y \sim N_2(\mu_x = 0, \mu_y = 0, \sigma_x^2 = 3^2, \sigma_y^2 = 3^2, \rho = 0)$

What is the weight of the center pixel?

Gaussian blur

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

In a Gaussian blur, for every pixel:

- Weight each pixel by the probability that X and Y are both within the pixel bounds
- The weighting function is a Bivariate Gaussian (Normal) standard deviation parameter σ

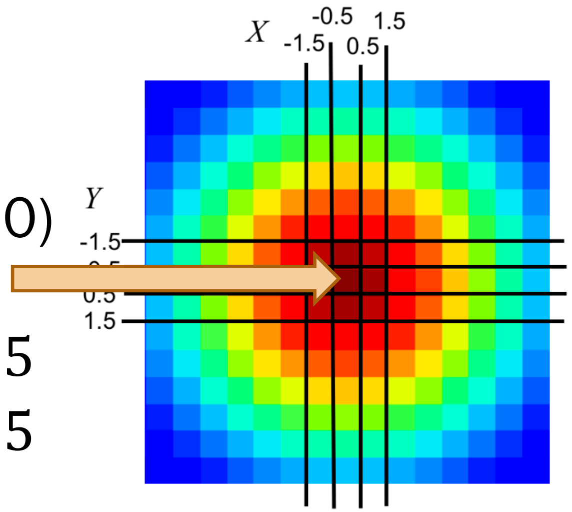
Weight matrix:

Center pixel: (0, 0)

Pixel bounds:

$-0.5 < x \leq 0.5$

$-0.5 < y \leq 0.5$



→ Independently: $X \sim \mathcal{N}(0, 3^2), Y \sim \mathcal{N}(0, 3^2)$

→ Jointly: $X, Y \sim N_2(\mu_x = 0, \mu_y = 0, \sigma_x^2 = 3^2, \sigma_y^2 = 3^2, \rho = 0)$

What is the weight of the center pixel?

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

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

Gaussian blur

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

In a Gaussian blur, for every pixel:

- Weight each pixel by the probability that X and Y are both within the pixel bounds
- The weighting function is a Bivariate Gaussian (Normal) standard deviation parameter σ

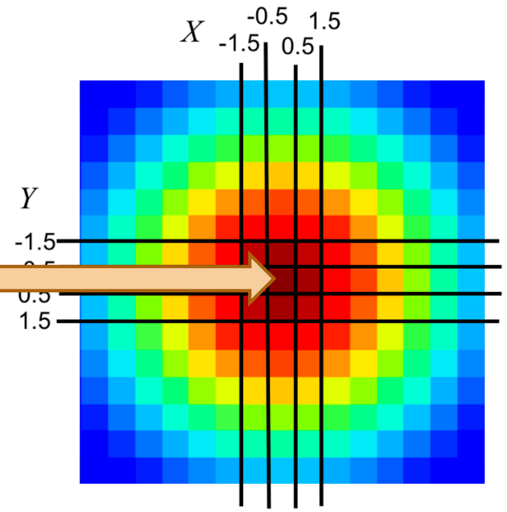
Weight matrix:

Center pixel: (0, 0)

Pixel bounds:

$-0.5 < x \le 0.5$

$-0.5 < y \le 0.5$



→ Independently: $X \sim \mathcal{N}(0, 3^2), Y \sim \mathcal{N}(0, 3^2)$

→ Jointly: $X, Y \sim N_2(\mu_x = 0, \mu_y = 0, \sigma_x^2 = 3^2, \sigma_y^2 = 3^2, \rho = 0)$

What is the weight of the center pixel?

$$P(-0.5 < X \le 0.5, -0.5 < Y \le 0.5) =$$

$= 0.206$

Is there a version of Bayes' Theorem for continuous Random Variables?

Conditional probability and Bayes' Theorem

Definition

$$P(F|E) = \frac{P(E \cap F)}{\underbrace{P(E)}_{\text{Scaling to the correct sample space}}}$$

Scaling to the correct sample space

Independence

E, F independent

$$P(F|E) = \underbrace{P(F)}_{\text{Sample space doesn't need to be scaled}}$$

Sample space doesn't need to be scaled

Bayes' Theorem

Prior: some prob. of event F

$$P(F|E) = \frac{P(F)P(E|F)}{\underbrace{P(E)}_{\text{Scaling to the correct sample space}}}$$

Likelihood

Posterior: prob. of F knowing that E happened

Scaling to the correct sample space

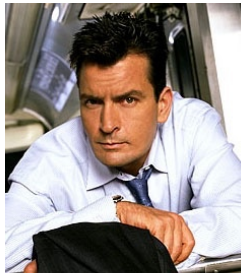
Multiple Bayes' Theorems

with
events

$$P(F|E) = \frac{P(F)P(E|F)}{P(E)}$$

with
discrete RVs

$$p_{Y|X}(y|x) = \frac{p_Y(y)p_{X|Y}(x|y)}{p_X(x)}$$

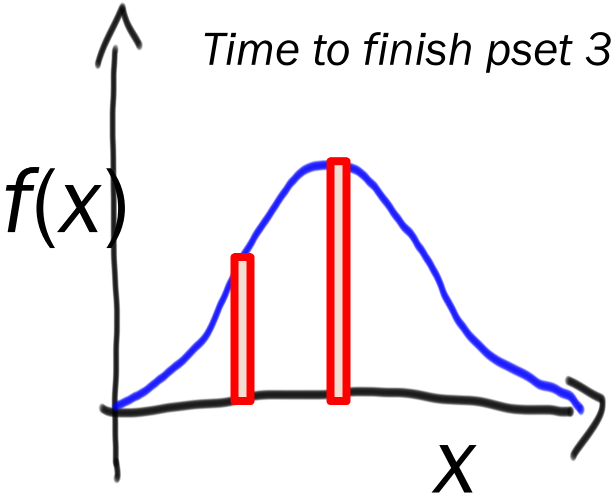


with
continuous RVs

Relative Probability of Continuous Variables

X = time to finish pset 3

X ~ N(10, 2)



How much more likely are you to complete in 10 hours than in 5?

$$\frac{P(X=10)}{P(X=5)} = \frac{\varepsilon f(X=10)}{\varepsilon f(X=5)}$$

$$= \frac{f(X=10)}{f(X=5)}$$

$$= \frac{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}}$$

$$= \frac{\frac{1}{\sqrt{4\pi}} e^{-\frac{(10-10)^2}{4}}}{\frac{1}{\sqrt{4\pi}} e^{-\frac{(5-10)^2}{4}}}$$

$$= \frac{e^0}{e^{-\frac{25}{4}}} = 518$$

Bayes with Continuous Random Variables

Let X and Y be continuous random variables

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

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

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

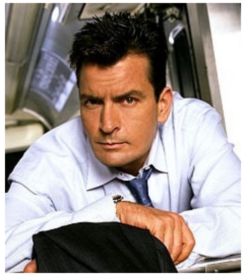
Multiple Bayes' Theorems

with events

$$P(F|E) = \frac{P(F)P(E|F)}{P(E)}$$

with discrete RVs

$$p_{Y|X}(y|x) = \frac{p_Y(y)p_{X|Y}(x|y)}{p_X(x)}$$



with continuous RVs

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

Really all the same idea!

Mixing Discrete and Continuous

Let X be a continuous random variable

Let N be a discrete random variable

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

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

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

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

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

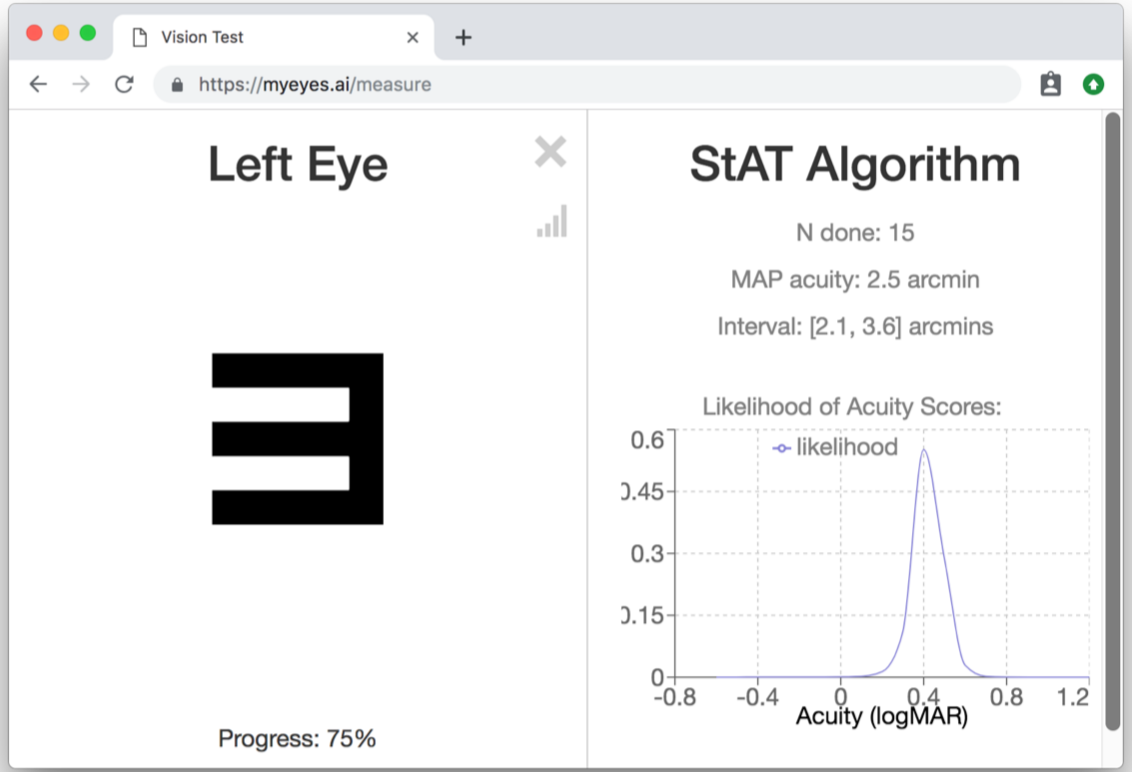
$$p_{N|X}(n|x) = \frac{f_{X|N}(x|n)p_N(n)}{f_X(x)}$$

All Continuous

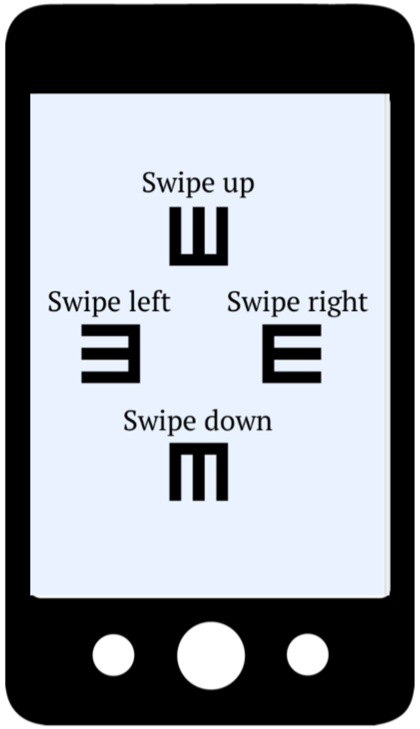
$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Stanford Acuity Test

1 Take an eye exam on this website



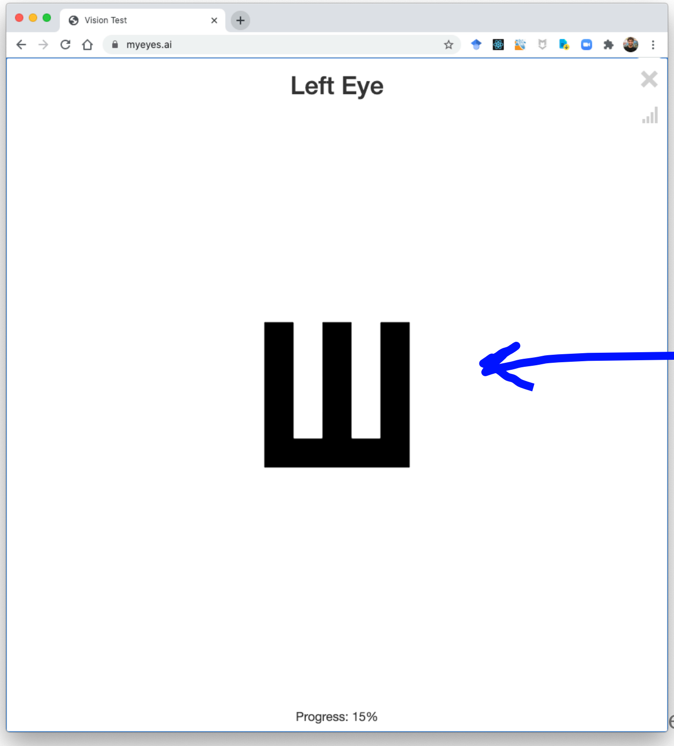
2 Connect your phone



3 Visualize the math

Stanford Acuity Test

A user is shown a letter at font size 3 and gets it wrong. What is your new belief that their visual ability is 3?



Aside: font size = 3, means it is 3x what someone with healthy vision can see clearly. Visual ability = 3 means the person can see font size 3 with 80% accuracy

First lets Define a Few Variables

Random Variables

A Ability to see

Y Indicates that they answered a question correct

Numbers

x Font size shown

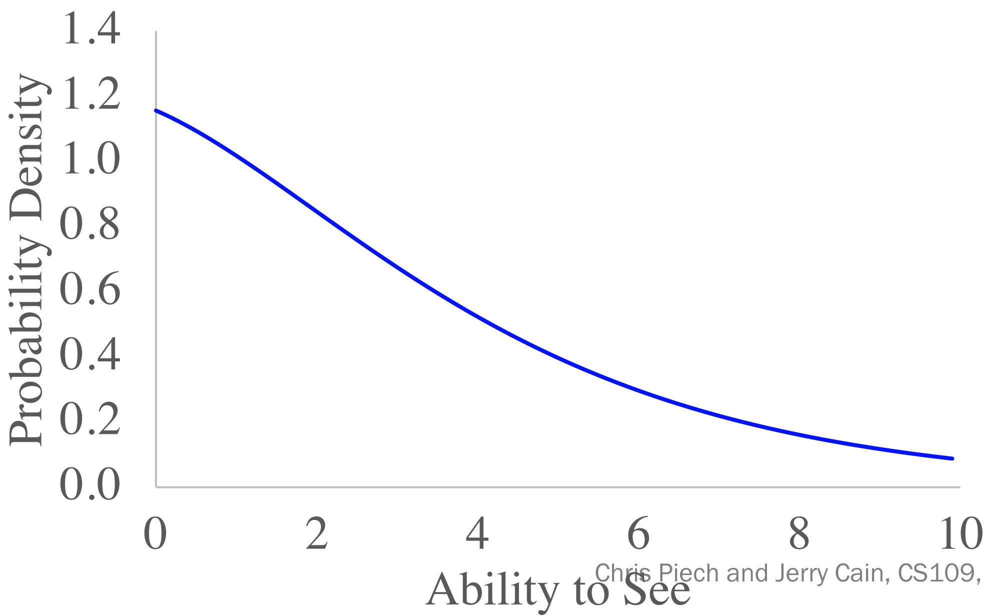
Probability Question

$$f(A = a | Y = 0)$$

Stanford Acuity Test

Your prior belief in ability to see is distributed as a gumbel

$$f(A = a) = 3.3 \cdot e^{-0.3a - 0.33 - e^{(-0.3a - 0.33)}}$$

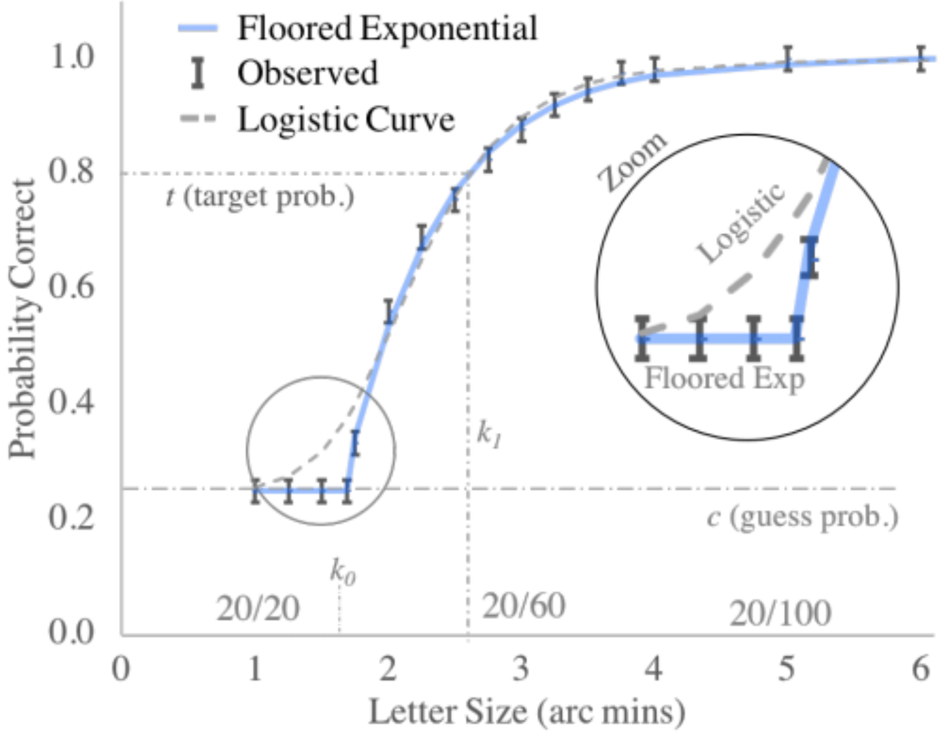


Curious?

https://en.wikipedia.org/wiki/Gumbel_distribution

What is the probability you are correct, given ability

$$P(Y = 1 | A = a) = \max(0.25, 1 - 0.75 \cdot 0.27^{\frac{x-a}{0.2a}})$$



Bring it all together.

What we **know**

$$f(A=a) = 3.3 \cdot e^{-0.3a-0.33-e^{(-0.3a-0.33)}}$$

$$P(Y=1|A=a) = \max(0.25, 1-0.75 \cdot 0.27^{\frac{x-a}{0.2a}})$$

Font-size was 3

What we **want**

$$f(A=3|Y=0) = \frac{[1-P(Y=1|A=3)] \cdot f(A=3)}{P(Y=0)}$$

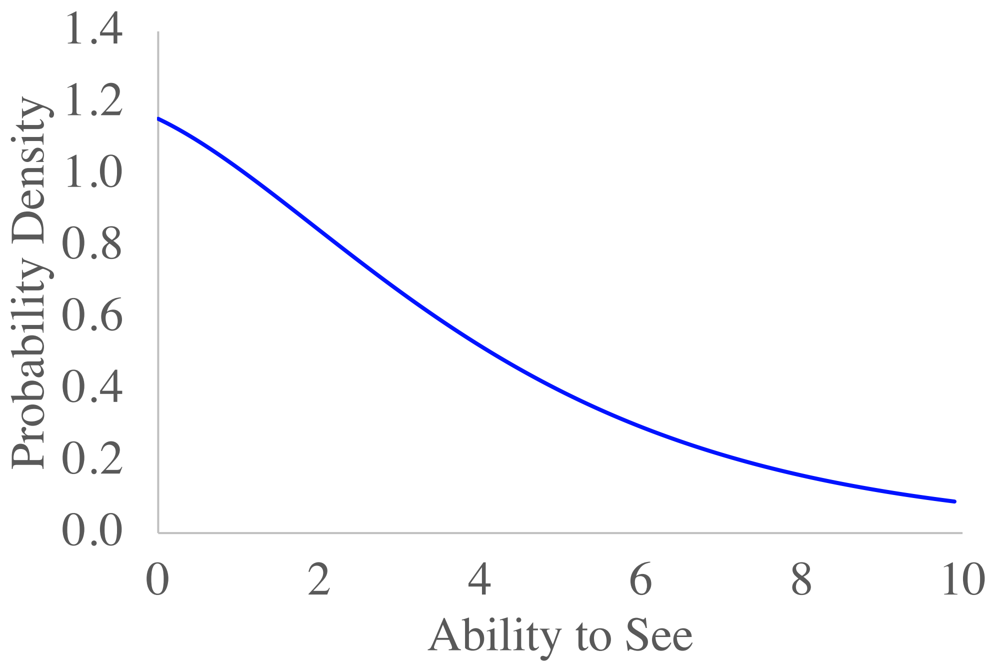
$$= \frac{[1-P(Y=1|A=3)] \cdot f(A=3)}{\int_a P(Y=0|A=a)\ da}$$

$$= K \cdot [1-P(Y=1|A=3)] \cdot f(A=3)$$

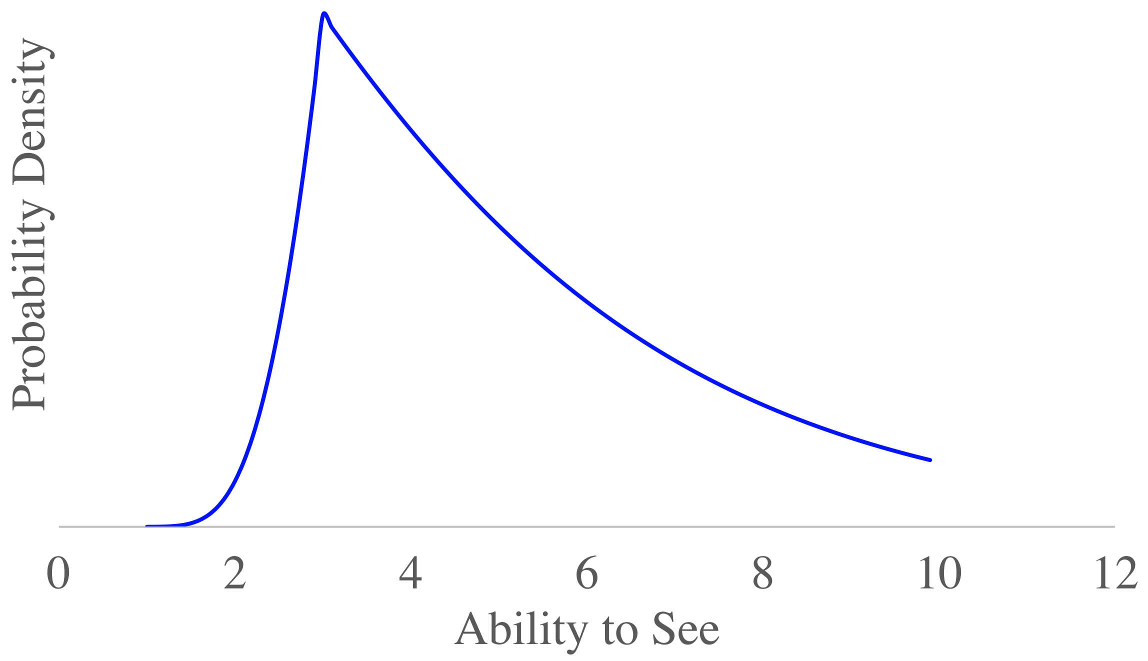
An Updated Belief

A user is shown a letter at **font size 3** and gets it **wrong**.
What is your new belief that their **visual ability is 3**?

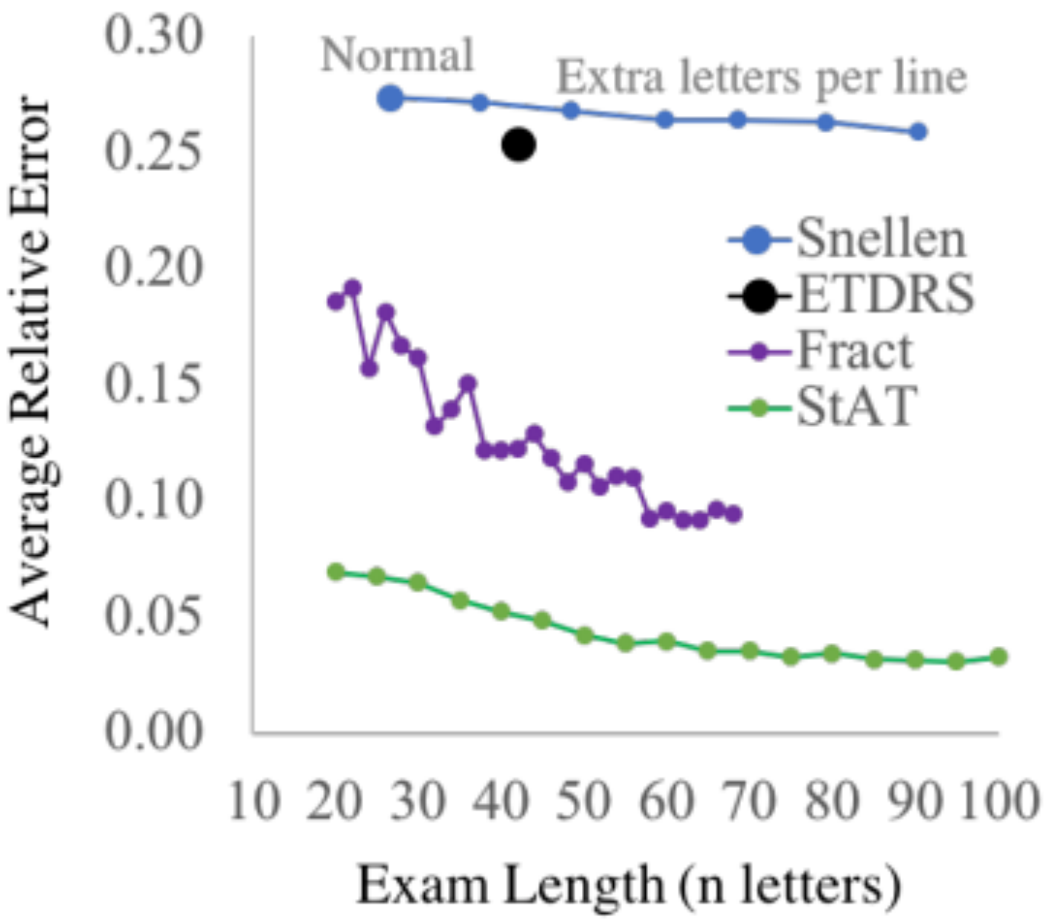
Prior



Posterior

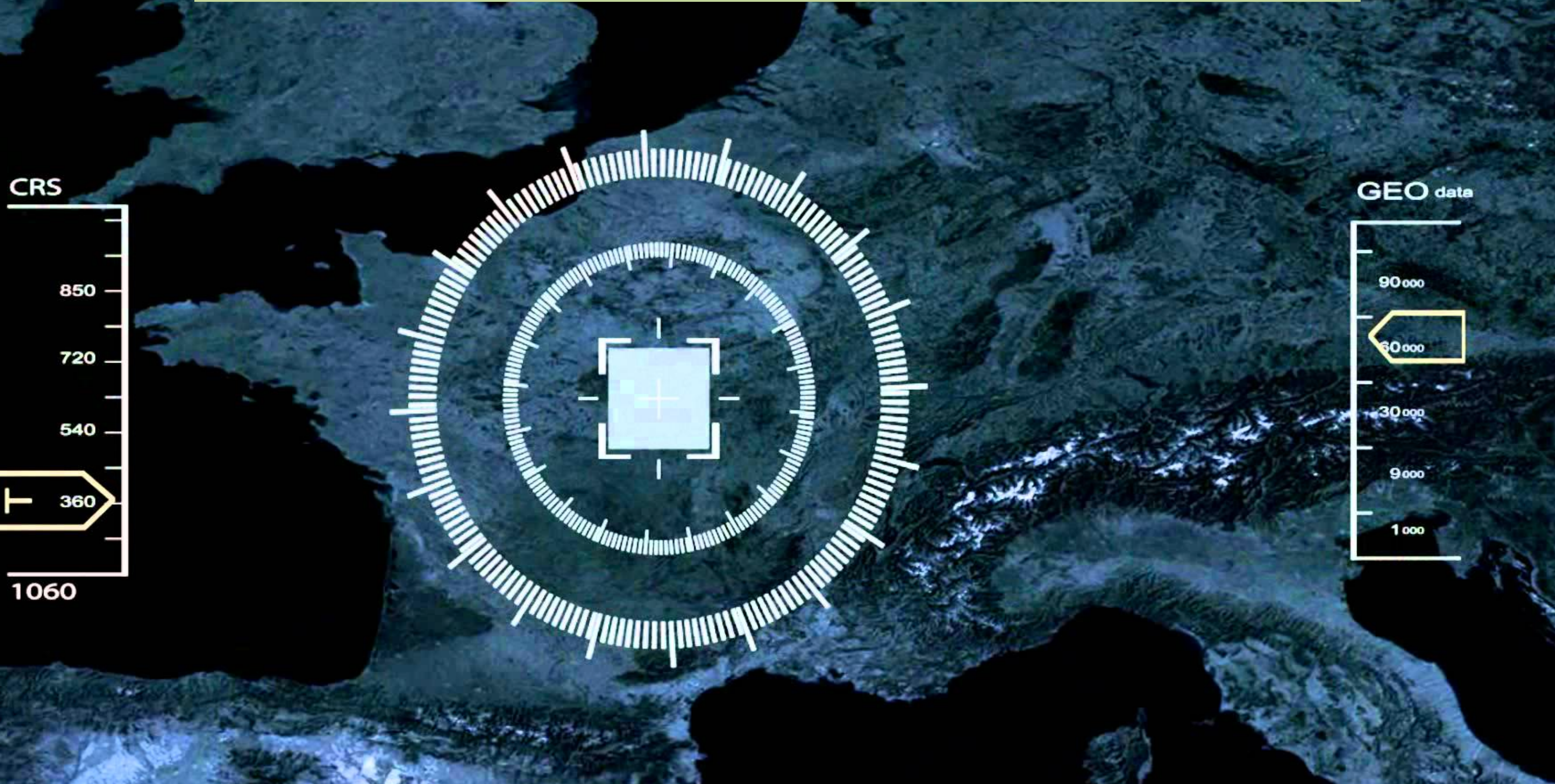


So what?



Actual model also included
+ a probability of "slip"
+ an intelligent algorithm for
choosing the next letter size

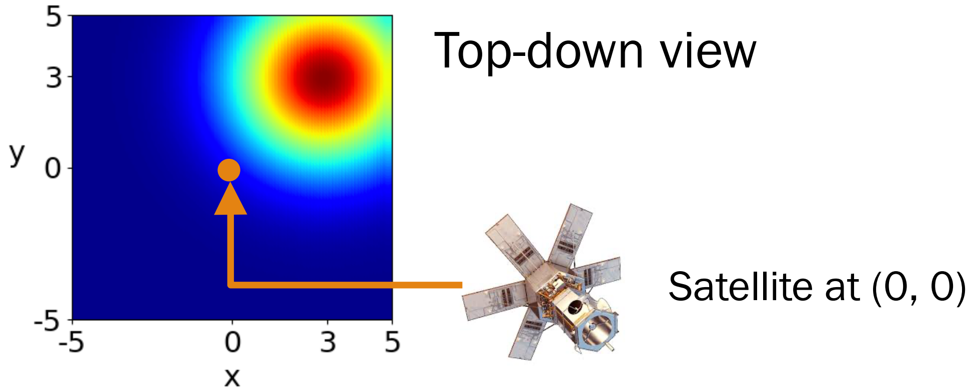
Tracking in 2D Space?



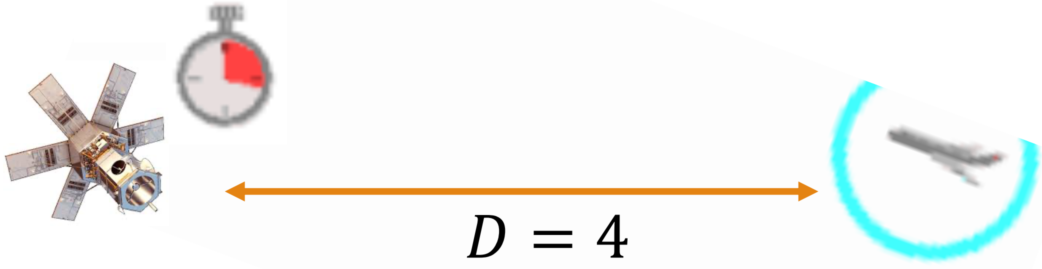

Give this a try!

Tracking in 2-D space

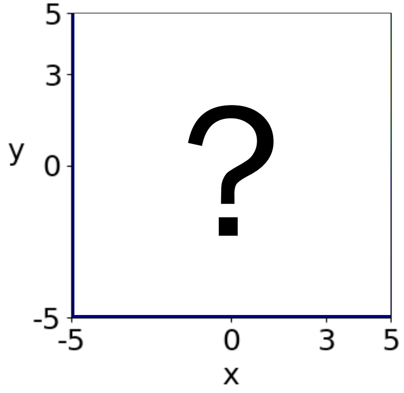
- Before measuring, we have some **prior belief** about the 2-D location of an object, (X, Y).



- We observe some noisy **measurement** $D = 4$, the Euclidean distance of the object to a satellite.



- After the measurement, what is our **updated (posterior) belief** of the 2-D location of the object?

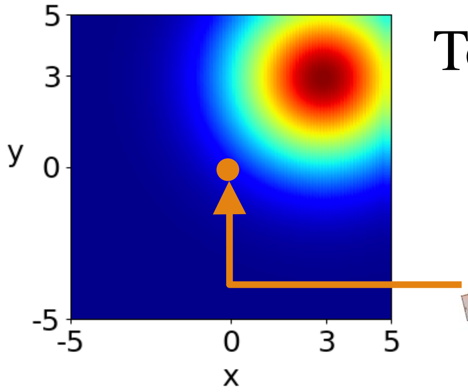


Tracking in 2-D space

- Before measuring, we have some **prior belief** about the 2-D location of an object, (X, Y).

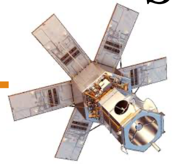
X,Y ~ Independent Bivariate normal

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



Top-down view

Satellite at $(0, 0)$

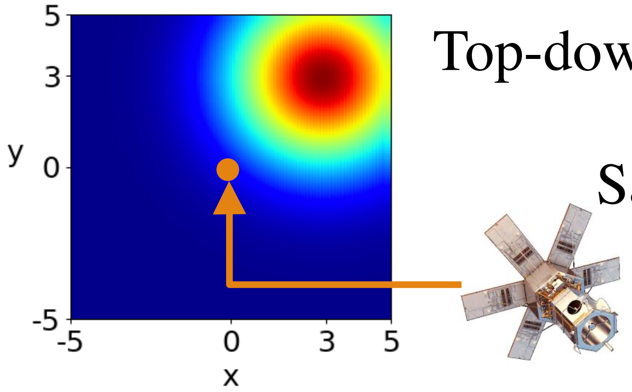


Tracking in 2-D space

- Before measuring, we have some **prior belief** about the 2-D location of an object, (X, Y).

X,Y ~ Independent Bivariate normal

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

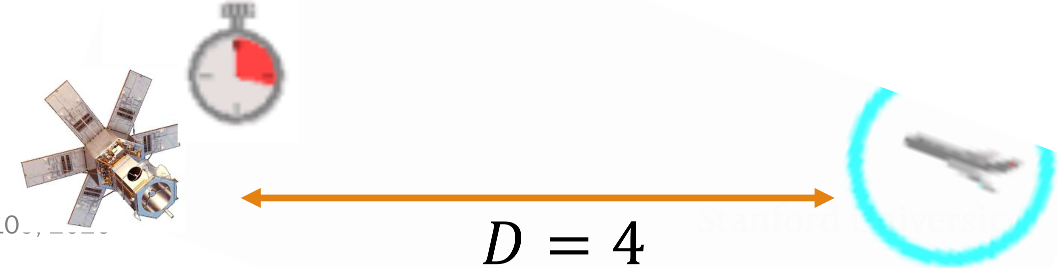


Top-down view

Satellite at $(0, 0)$

- We observe some noisy **measurement** $D = 4$, the Euclidean distance of the object to a satellite.

Let D = observed distance from the satellite.
Observed distance is true distance plus noise.
Noise is a **standard normal**.



Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Recall Bayes terminology:

posterior belief

likelihood (of evidence)

prior belief

$$f_{X,Y|D}(x, y|d) = \frac{f_{D|X,Y}(d|x, y) f_{X,Y}(x, y)}{f_D(d)}$$

normalization constant

1. Define prior

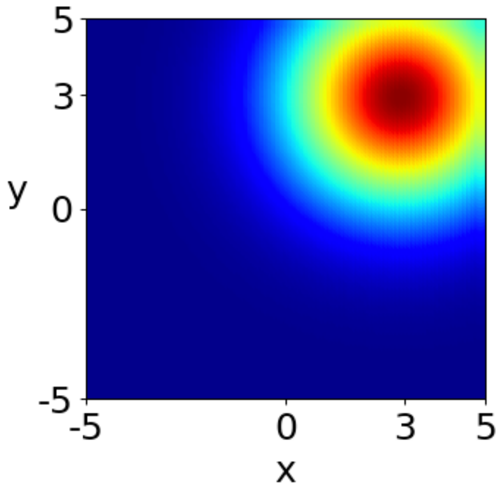
$$f_{X,Y|D}(x,y|d) = \frac{f_{D|X,Y}(d|x,y)\,\boxed{f_{X,Y}(x,y)}}{f_D(d)}$$

You have a prior belief about the 2-D location of an object, (X, Y).

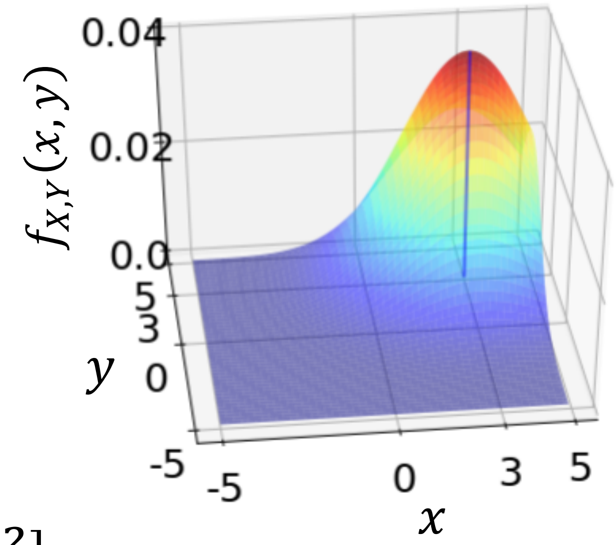
Let (X, Y) = object's 2-D location.
(your satellite is at (0,0))

Suppose the prior distribution is a symmetric bivariate normal distribution:

Top-down view



3-D view



$$f_{X,Y}(x,y) = \frac{1}{2\pi 2^2} e^{-\frac{[(x-3)^2+(y-3)^2]}{2(2^2)}} = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

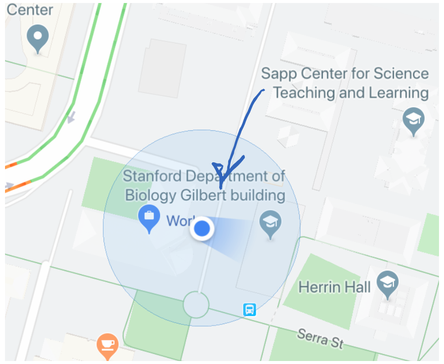
normalizing constant

2. Define likelihood

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

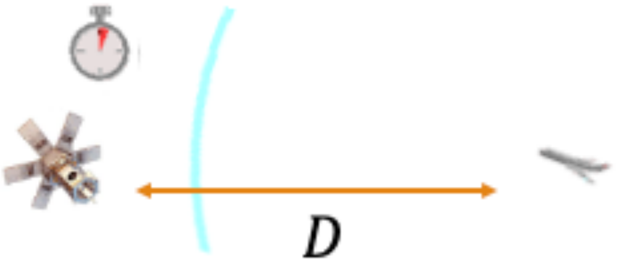
You observe a noisy distance measurement, $D = 4$.

If you knew your actual location (x, y), you could say how likely a measurement $D = 4$ is:



Let D = distance from the satellite (radially).
Suppose you knew your actual position: (x, y).

- D is still noisy! Suppose noise is **standard normal.**
- On average, D is your true Euclidean distance: $\sqrt{x^2 + y^2}$



2. Define likelihood

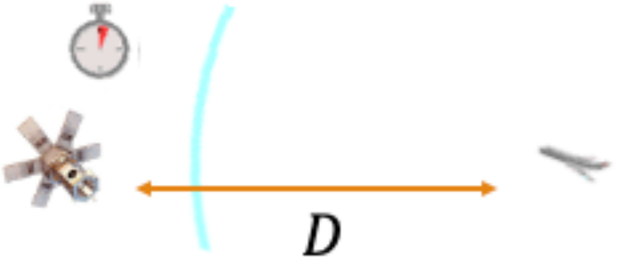
$$f_{X,Y|D}(x,y|d) = \frac{\boxed{f_{D|X,Y}(d|x,y)}\, f_{X,Y}(x,y)}{f_D(d)}$$

You observe a noisy distance measurement, $D = 4$.

If you knew your actual location (x, y), you could say how likely a measurement $D = 4$ is:

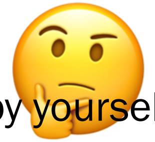
Let D = distance from the satellite (radially).
Suppose you knew your actual position: (x, y).

- D is still noisy! Suppose noise is **standard normal.**
- On average, D is your true Euclidean distance: $\sqrt{x^2 + y^2}$



$$D|X,Y \sim N(\mu = \ (A)\ , \sigma^2 = \ (B)\ )$$

$$f_{D|X,Y}(D = d|X = x, Y = y) = \frac{1}{(C)\ \sqrt{2\pi}} e^{\{\ \ (D)\ \ \}}$$



(by yourself)

2. Define likelihood

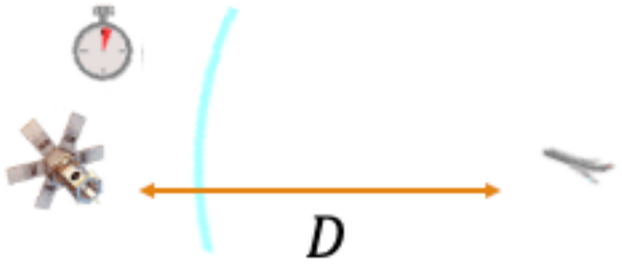
$$f_{X,Y|D}(x,y|d) = \frac{\boxed{f_{D|X,Y}(d|x,y)}\, f_{X,Y}(x,y)}{f_D(d)}$$

You observe a noisy distance measurement, $D = 4$.

If you knew your actual location (x, y), you could say how likely a measurement $D = 4$ is:

Let D = distance from the satellite (radially).
Suppose you knew your actual position: (x, y).

- D is still noisy! Suppose noise is **standard normal.**
- On average, D is your true Euclidean distance: $\sqrt{x^2 + y^2}$



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

$$f_{D|X,Y}(D = d|X = x, Y = y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\left(d-\sqrt{x^2+y^2}\right)^2}{2}} = K_2 \cdot e^{\frac{-\left(d-\sqrt{x^2+y^2}\right)^2}{2}}$$

normalizing constant

3. Compute posterior

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

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

<u>Compute:</u>

Posterior belief

$$f_{X,Y|D}(x,y|4) = f_{X,Y|D}(X=x, Y=y|D=4)$$

3. Compute posterior

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

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Compute:

Posterior belief

$$f_{X,Y|D}(x,y|4) = f_{X,Y|D}(X = x, Y = y|D = 4)$$

Know:

Prior belief

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

Observation likelihood

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

Tips
- Use Bayes' Theorem!
- $f_D(4)$ is just a scaling constant. Why?
- How can we approximate the final scaling constant with a computer?

Deep breath

Tracking in 2-D space

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

likelihood of $D = 4$ prior belief

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

Bayes' Theorem

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

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

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

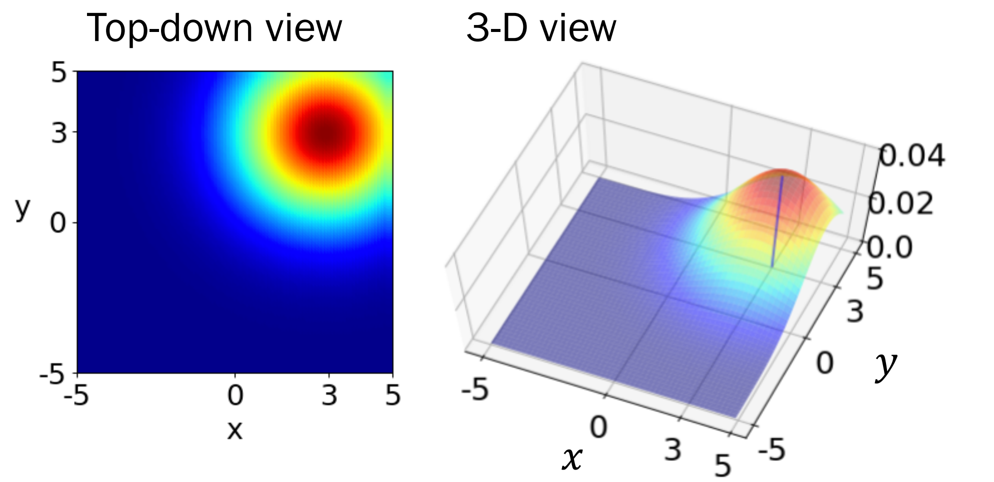
For your notes...

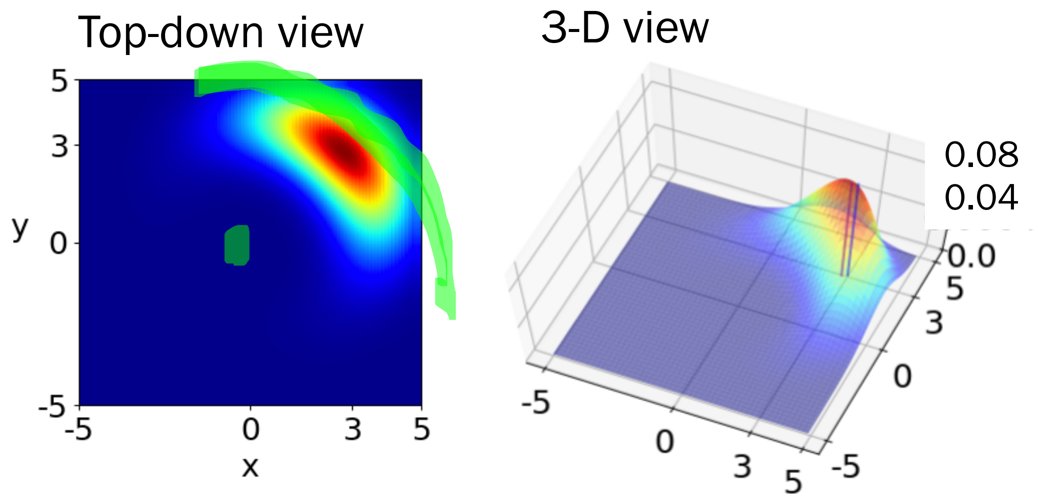
Key: Once we know the part dependent on x, y, we can computationally approximate K_4 such that $f_{X,Y|D}$ is a valid PDF.

Tracking in 2-D space

With this continuous version of Bayes' theorem, we can explore new domains.

- Before measuring, we have some **prior belief** about the 2-D location of an object, (X, Y).
- We observe some noisy **measurement** of the distance of the object to a satellite.
- After the measurement, what is our **updated (posterior) belief** of the 2-D location of the object?

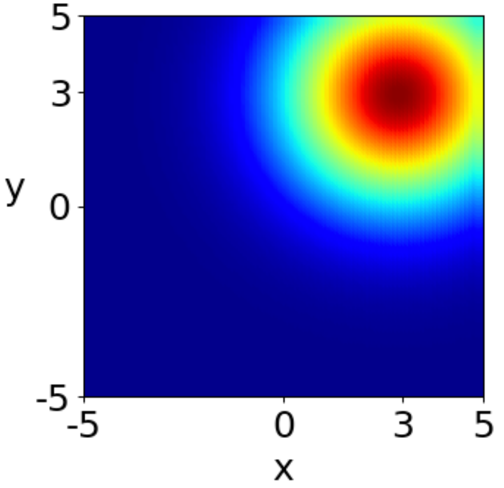




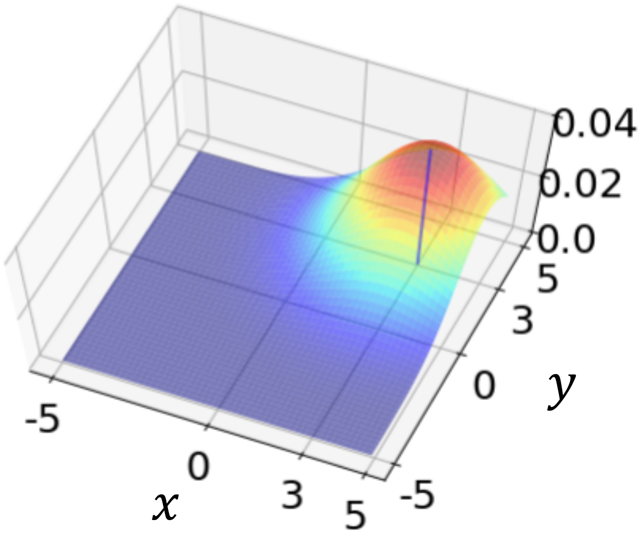
Tracking in 2-D space: Posterior belief

Prior belief

Top-down view



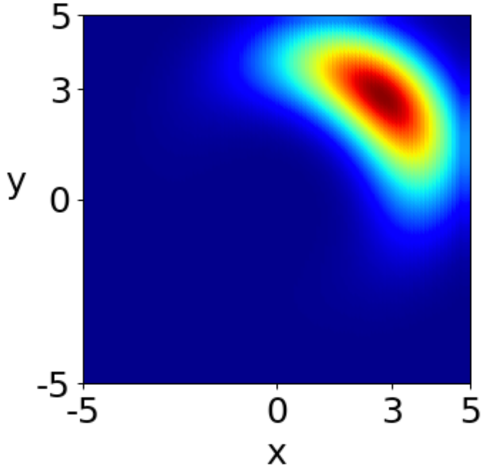
3-D view



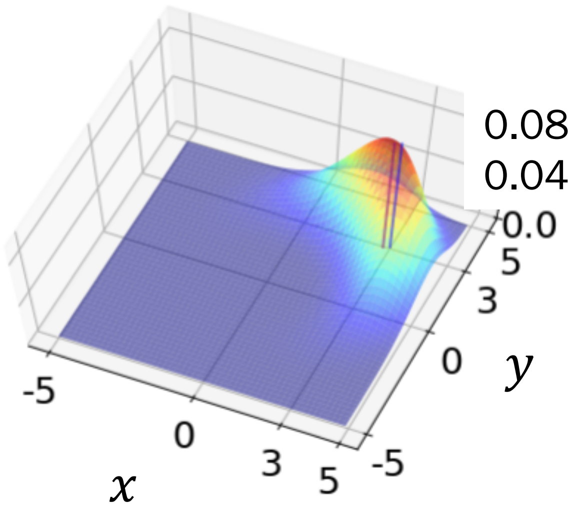
$$f_{X,Y}(x,y) = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

Posterior belief

Top-down view



3-D view



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

How'd you compute that K_4?

To be a valid conditional PDF, $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y|D}(x,y|4)\,dx\,dy = 1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_4 \cdot e^{-\left[\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]} dx\,dy = 1$$

$$\frac{1}{K_4} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\left[\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]} dx\,dy$$

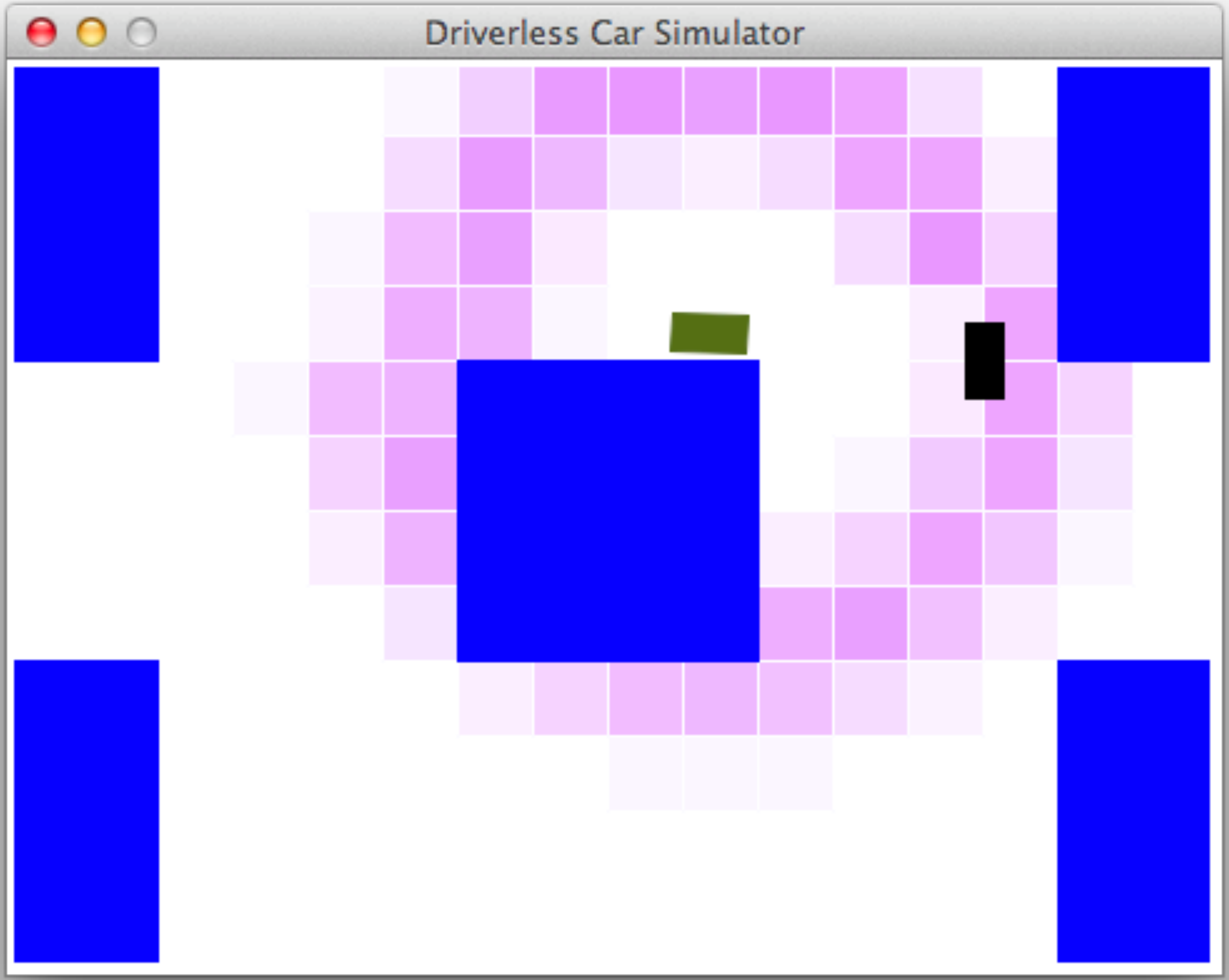
(pull out K_4, divide)

Approximate:

$$\frac{1}{K_4} \approx \sum_x \sum_y e^{-\left[\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]} \Delta x \Delta y$$

Use a computer!

Tracking in 2D Space




Give this a try!