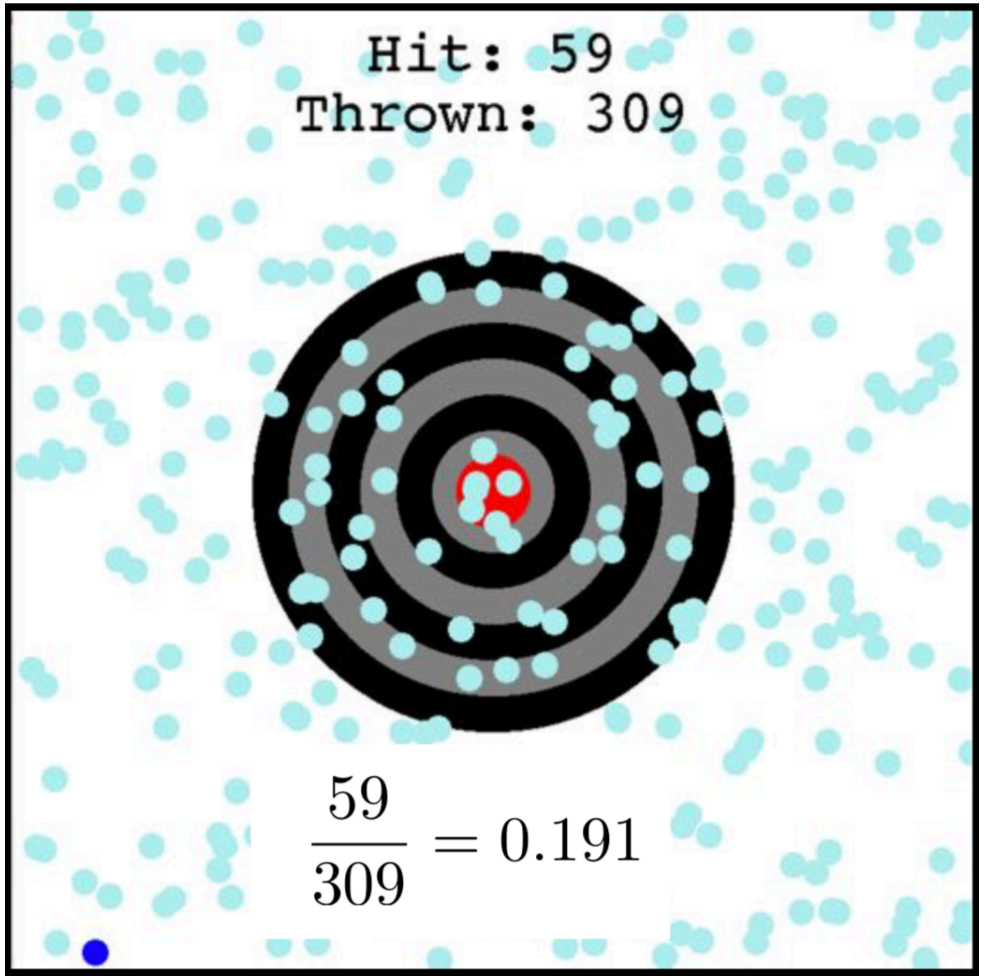
16: Continuous Joint Distributions

Lisa Yan and Jerry Cain
October 19, 2020

Quick slide reference

Continuous joint distributions

Remember target?



Good times…

CS109 logo with darts

The CS109 logo was created by throwing 500,000 darts according to a joint distribution.

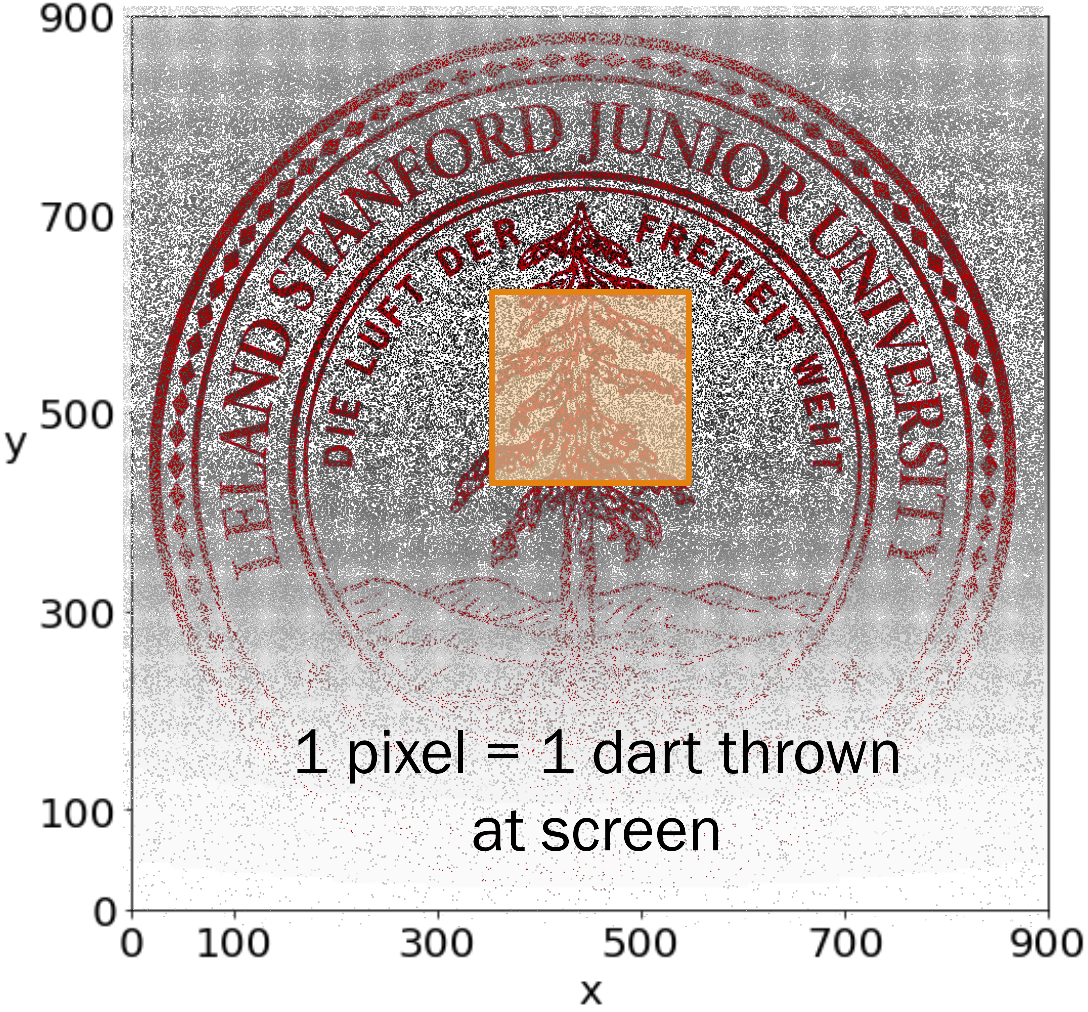
If we throw another dart according to the same distribution, what is

P(dart hits within r pixels of center)?

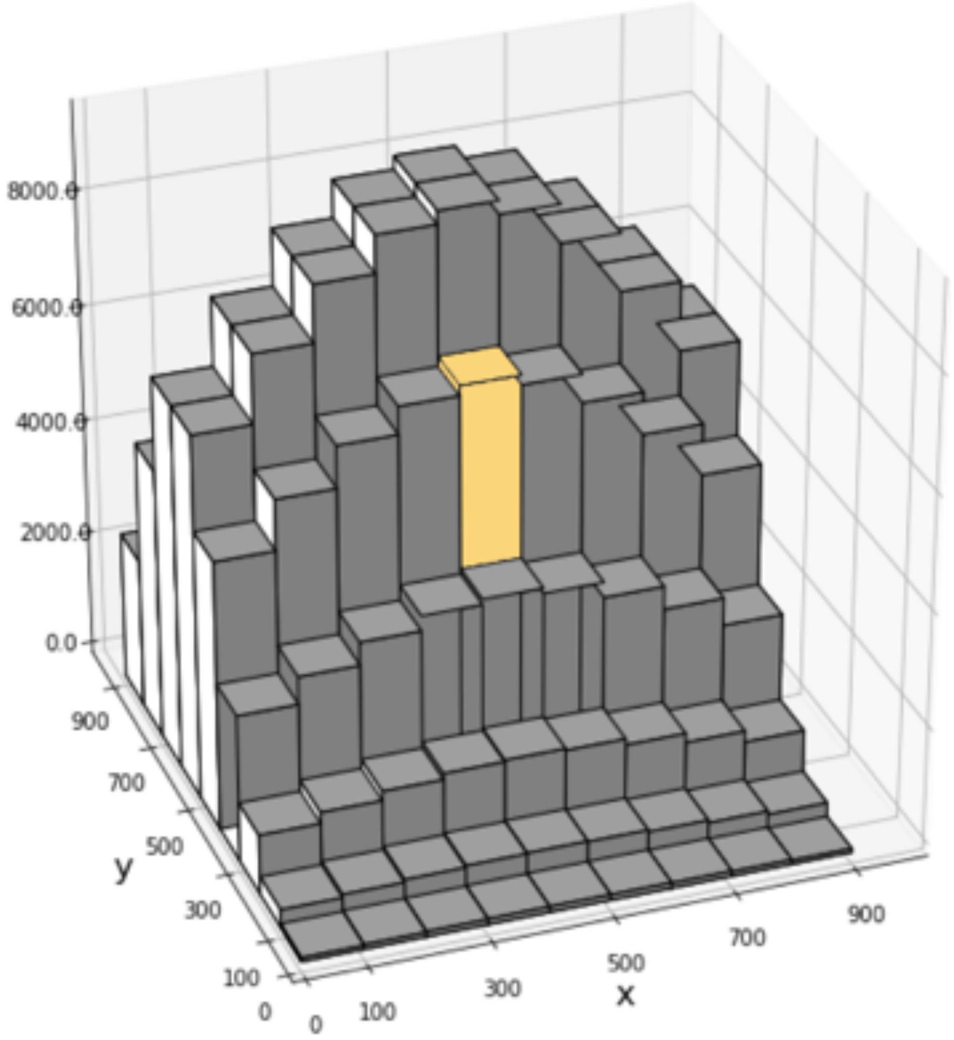
Quick check: What is the probability that a dart hits at (456.2344132343, 532.1865739012)?

CS109 logo with darts

P(dart hits within r pixels of center)?



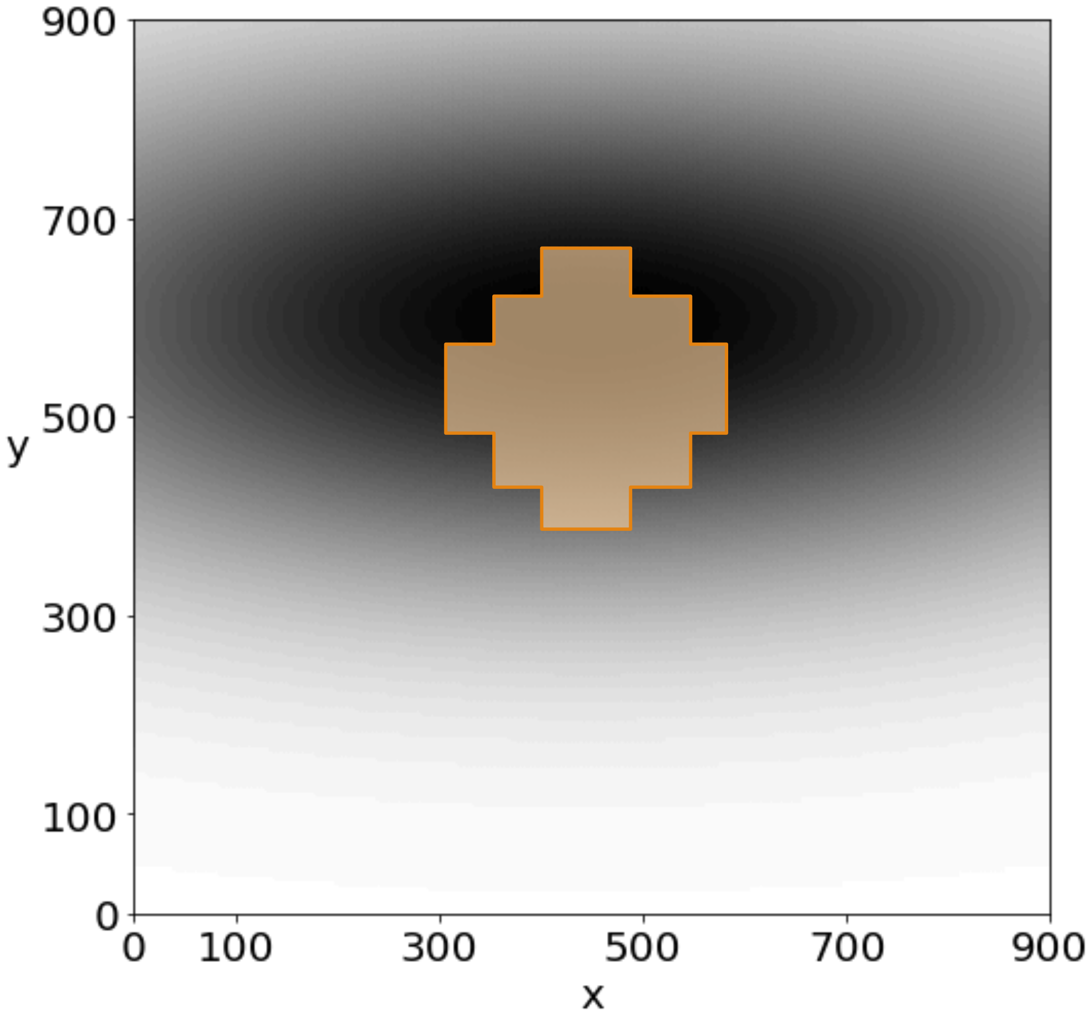
1 pixel = 1 dart thrown at screen

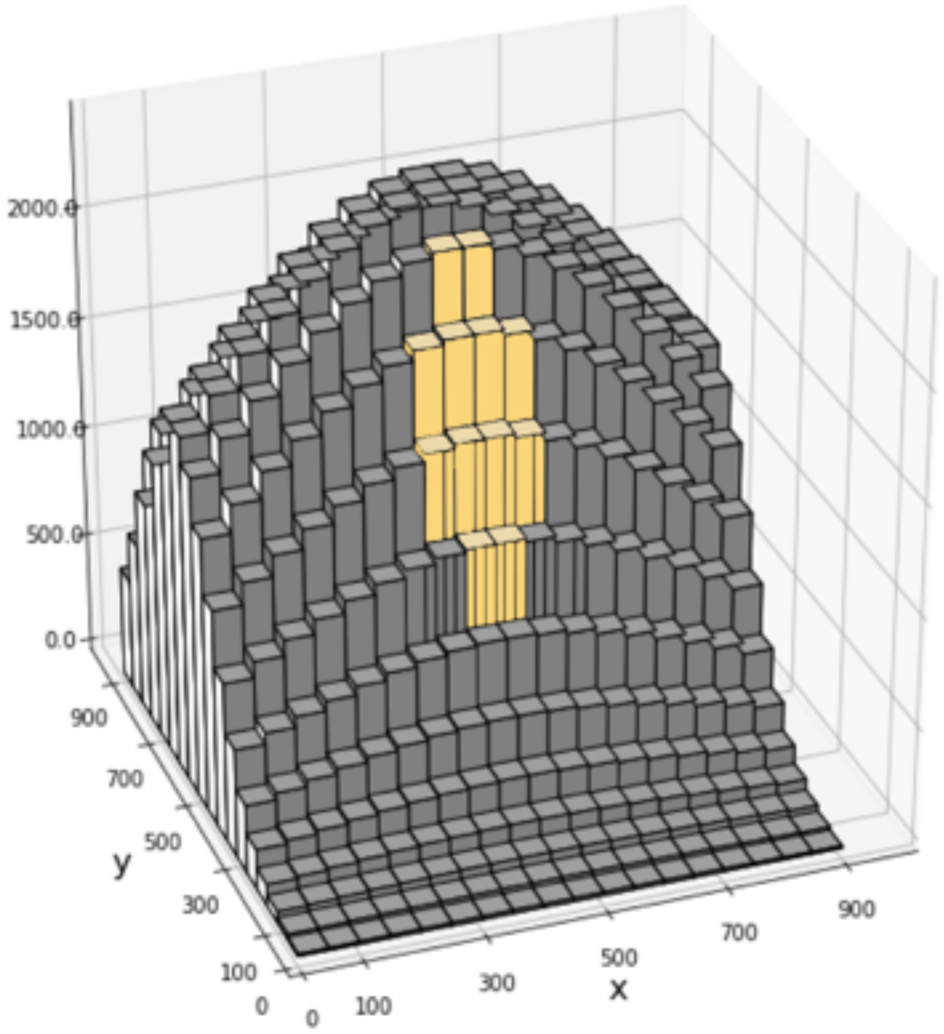


Possible dart counts (in 100x100 boxes)

CS109 logo with darts

P(dart hits within r pixels of center)?

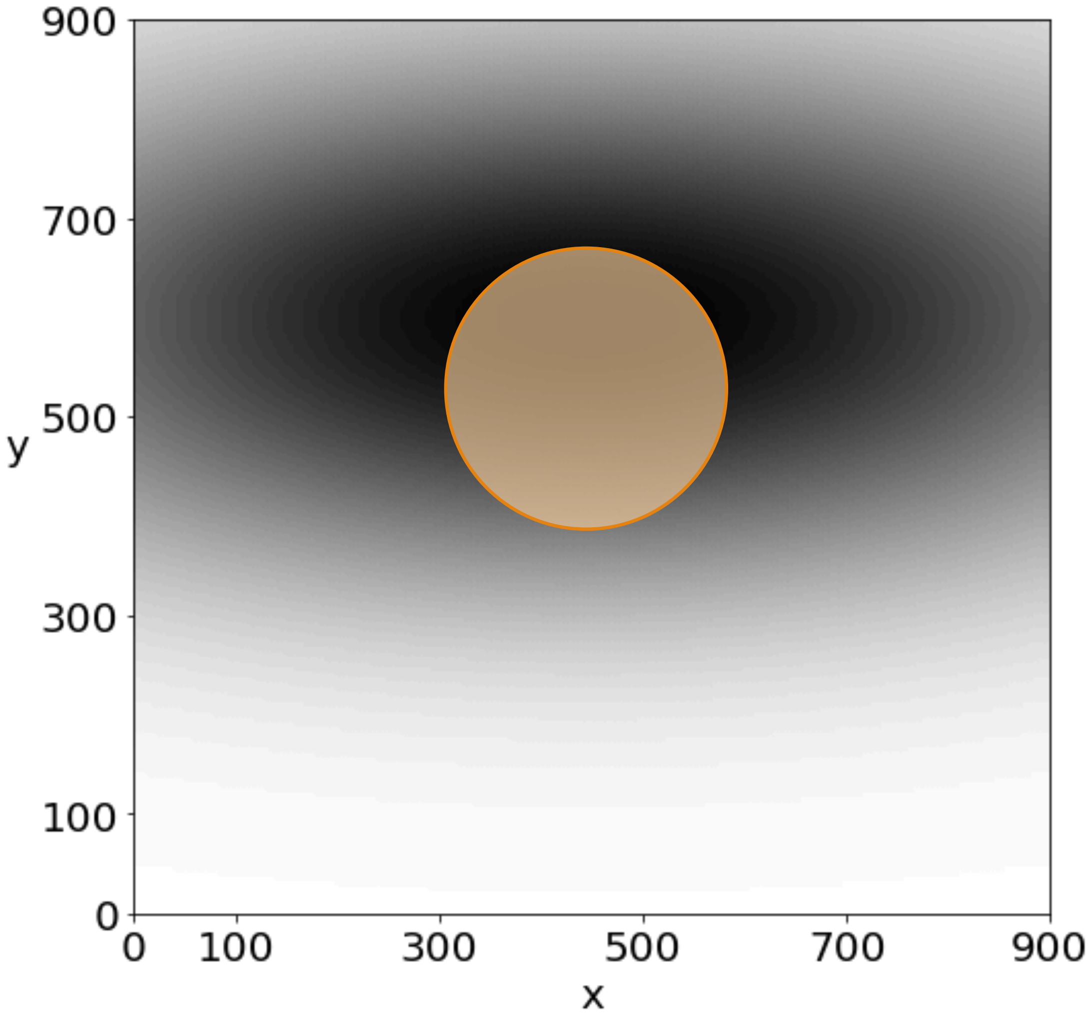


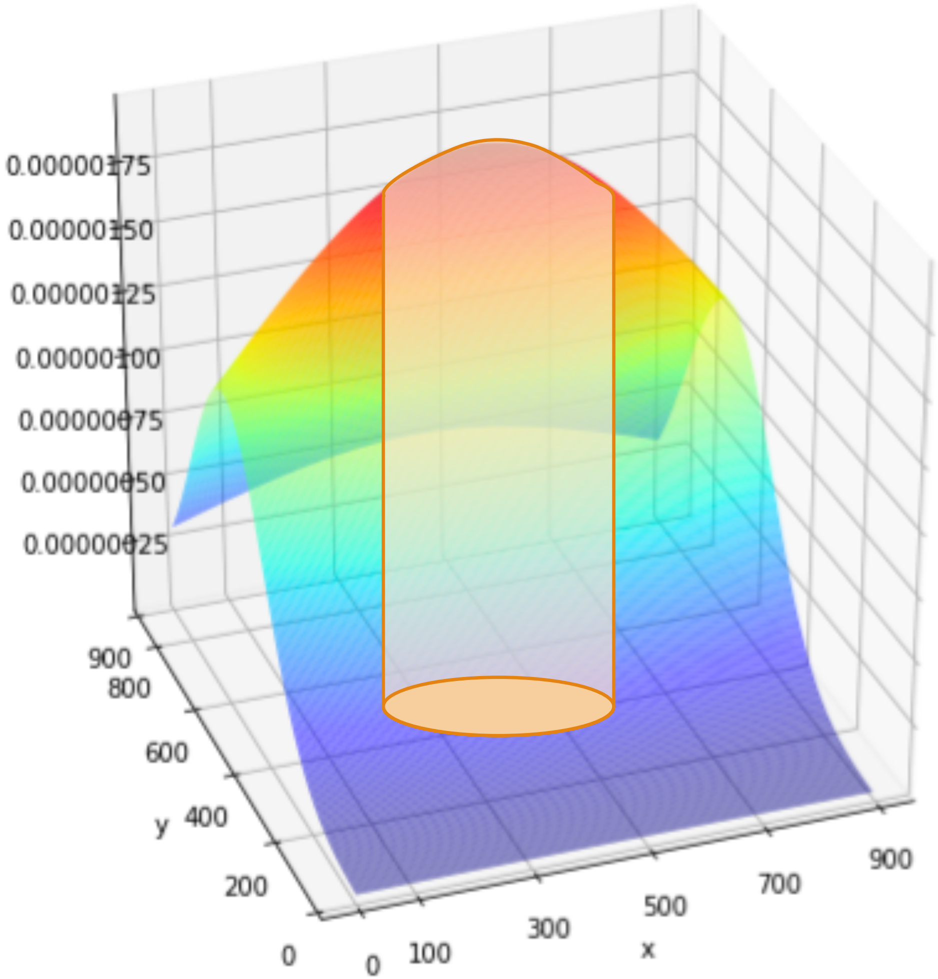


Possible dart counts (in 50x50 boxes)

CS109 logo with darts

P(dart hits within r pixels of center)?





Possible dart counts

(in infinitesimally small boxes)

Continuous joint probability density functions

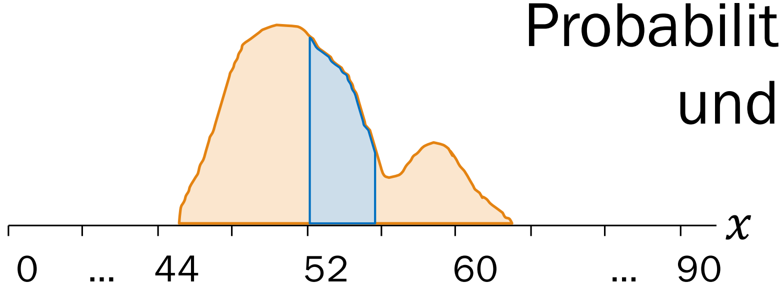
If two random variables X and Y are jointly continuous, then there exists a joint probability density function $f_{X,Y}$ defined over $-\infty < x, y < \infty$ such that:

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

From one continuous RV to jointly continuous RVs

Single continuous RV X

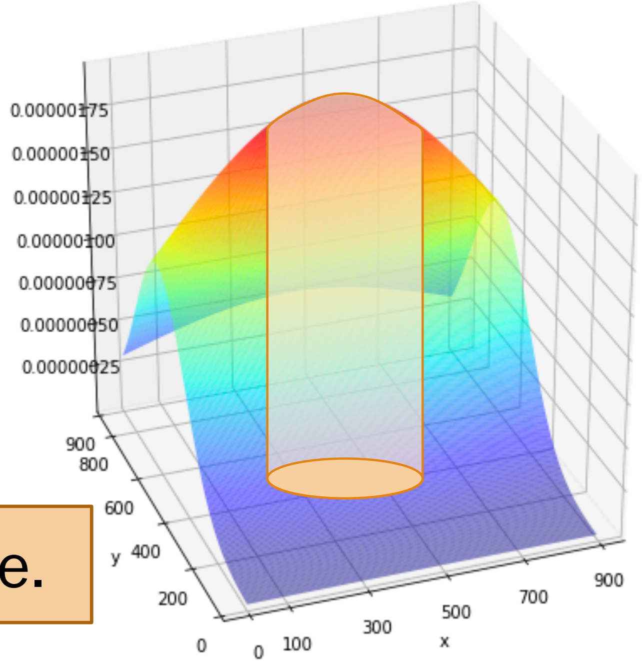
- PDF f_X such that $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- Integrate to get probabilities



Probability = **area** under curve

Jointly continuous RVs X and Y

- PDF $f_{X,Y}$ such that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\,dx = 1$
- Double integrate to get probabilities



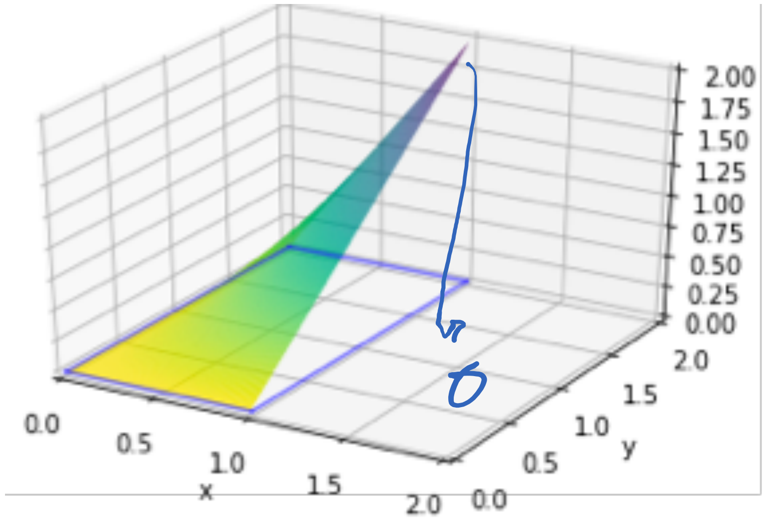
Probability for jointly continuous RVs is **volume** under a surface.

Double integrals without tears

Let X and Y be two continuous random variables.

- Support: $0 \leq X \leq 1$, $0 \leq Y \leq 2$.

Is $g(x, y) = xy$ a valid joint PDF over X and Y?



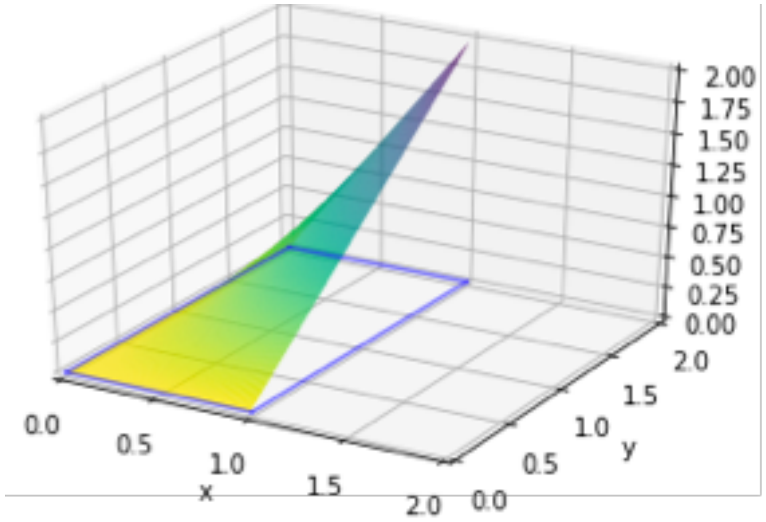
Write down the definite double integral that must integrate to 1:

Double integrals without tears

Let X and Y be two continuous random variables.

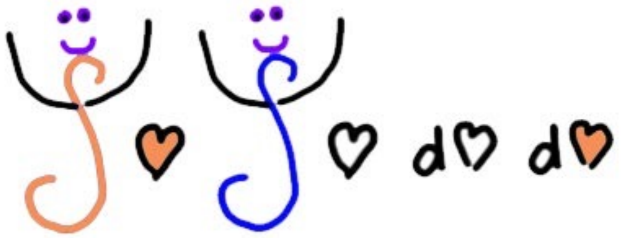
- Support: $0 \leq X \leq 1, 0 \leq Y \leq 2.$

Is $g(x, y) = xy$ a valid joint PDF over X and Y?

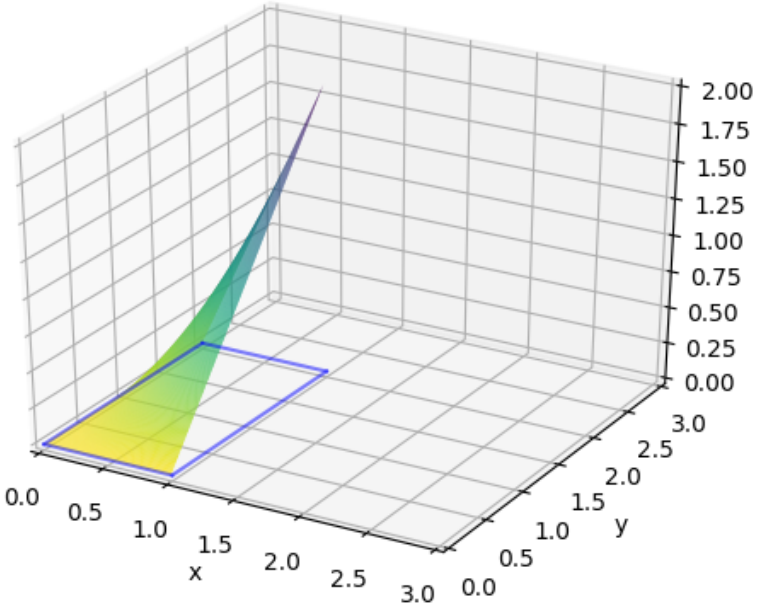


Write down the definite double integral that must integrate to 1:

$$\underbrace{\int_{y=0}^{2}\int_{x=0}^{1} xy\,dx\,dy = 1}_{\text{(used in next slide)}} \qquad \text{or} \qquad \int_{x=0}^{1}\int_{y=0}^{2} xy\,dy\,dx = 1$$



Double integrals without tears



Let X and Y be two continuous random variables.

- Support: $0 \le X \le 1$, $0 \le Y \le 2$.

Is $g(x, y) = xy$ a valid joint PDF over X and Y?

0. Set up integral:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dx\,dy = \int_{y=0}^{2}\int_{x=0}^{1} xy\,dx\,dy$$

1. Evaluate inside integral by treating y as a constant:

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

2. Evaluate remaining (single) integral:

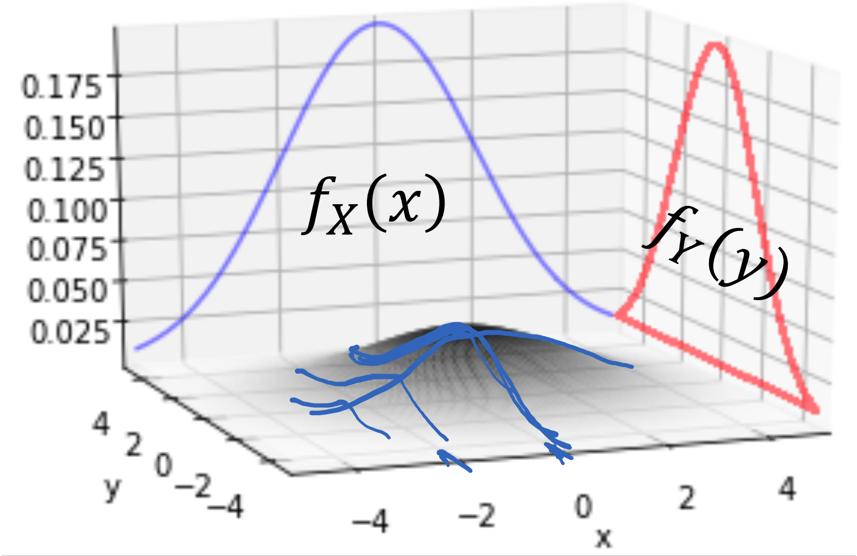
$$\int_{y=0}^{2} y\frac{1}{2}dy = \left[\frac{y^2}{4}\right]_{y=0}^{2} = 1 - 0 = 1$$

Yes, $g(x,y)$ is a valid joint PDF because it integrates to 1.

Marginal distributions

Suppose X and Y are continuous random variables with joint PDF:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\,dx = 1$$



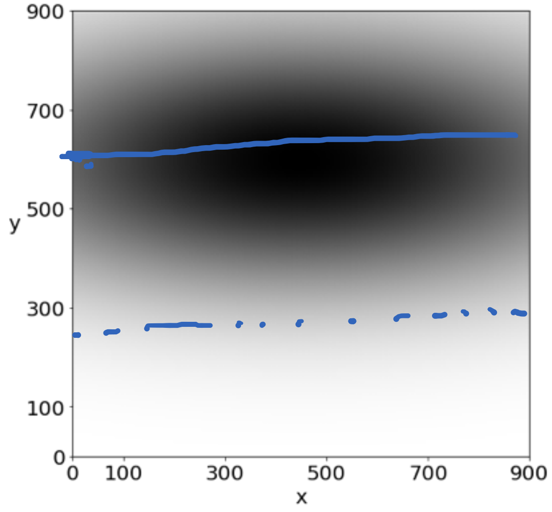
The marginal density functions (**marginal PDFs**) are therefore:

$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(\underset{\uparrow}{a},y)dy \qquad f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x,b)dx$$

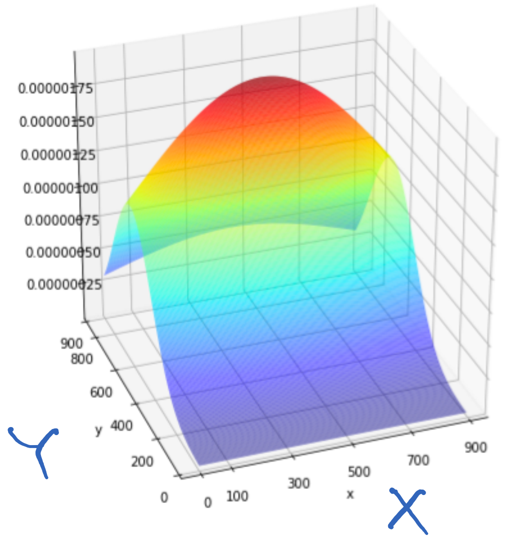
$$P_X(a) = \sum_{y} P_{X,Y}(a,y)$$

Back to darts!

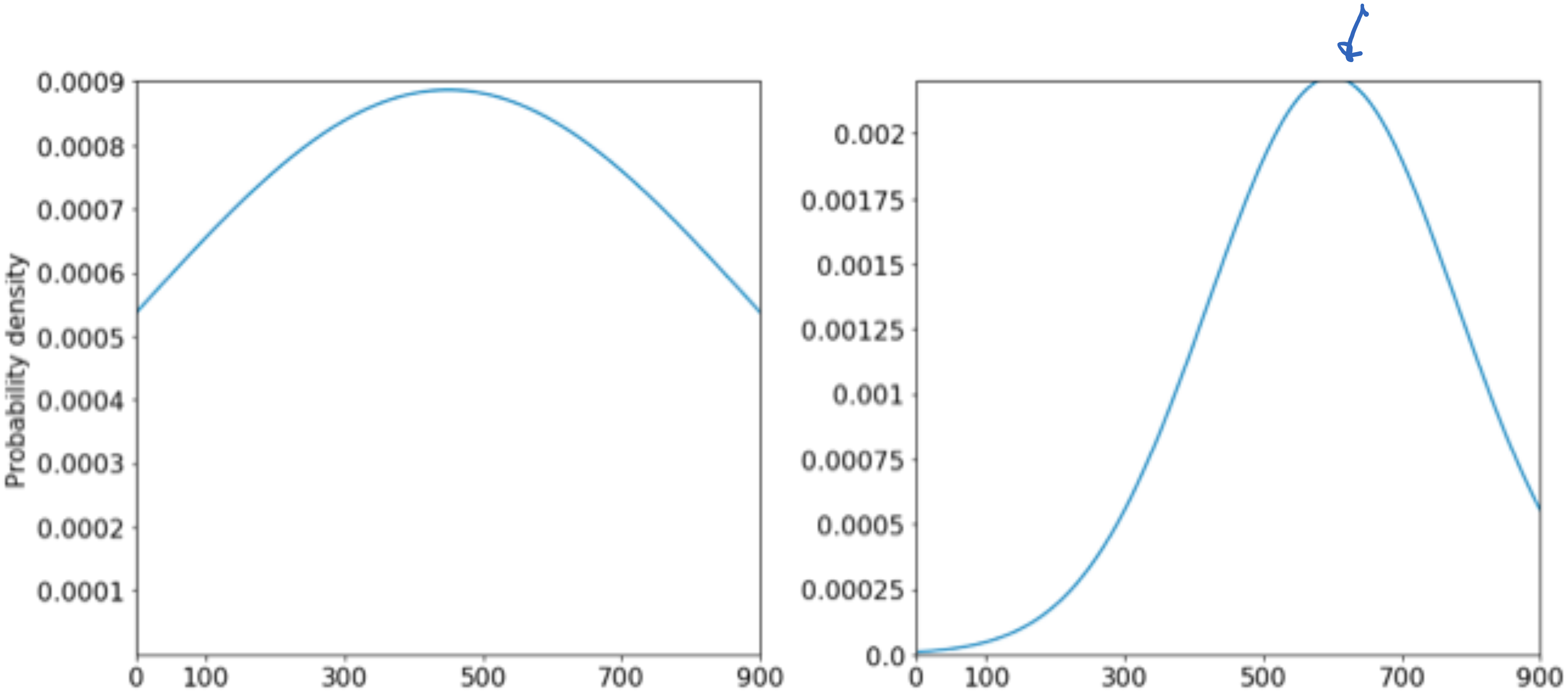
(top-down)



(side view)

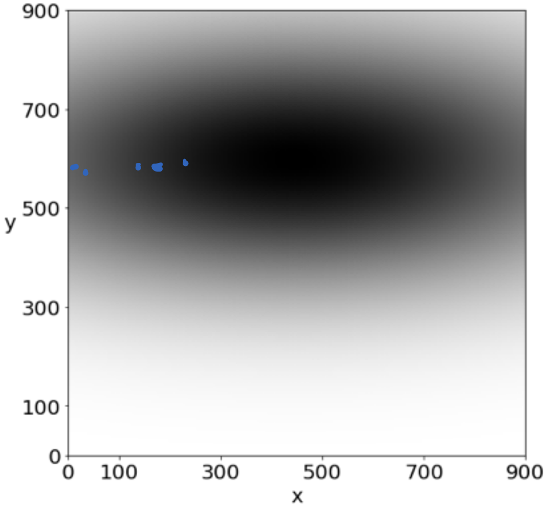


Match X and Y to their respective marginal PDFs:

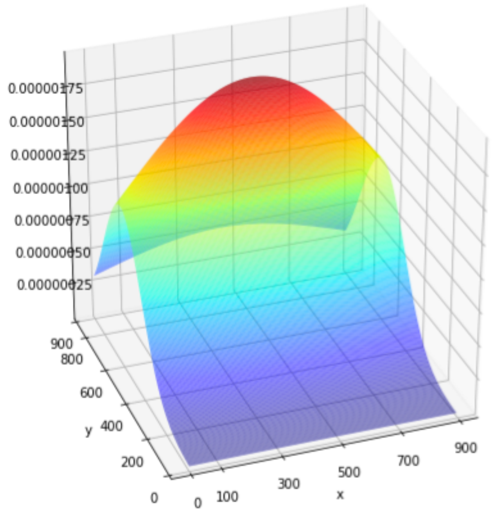


Back to darts!

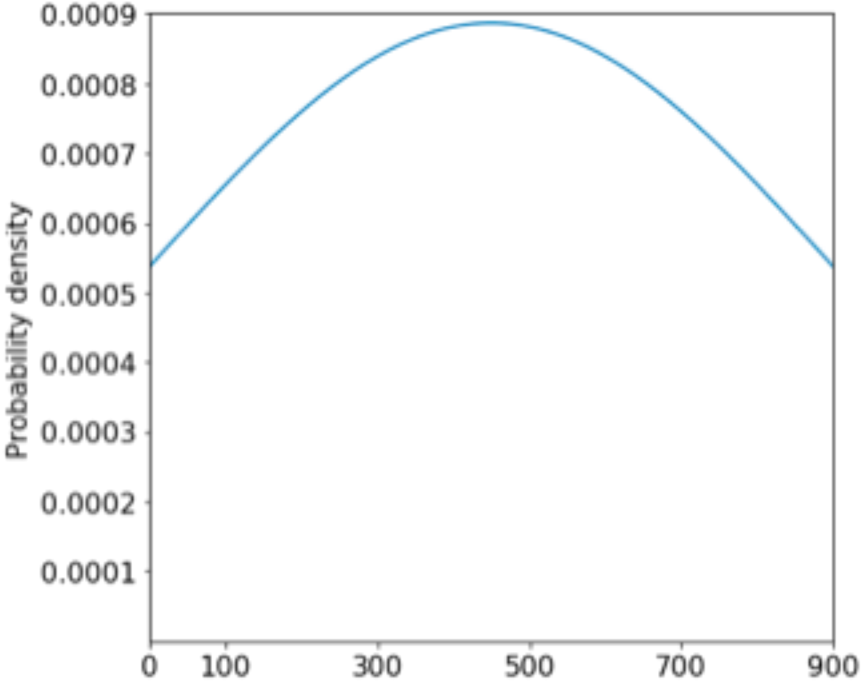
(top-down)



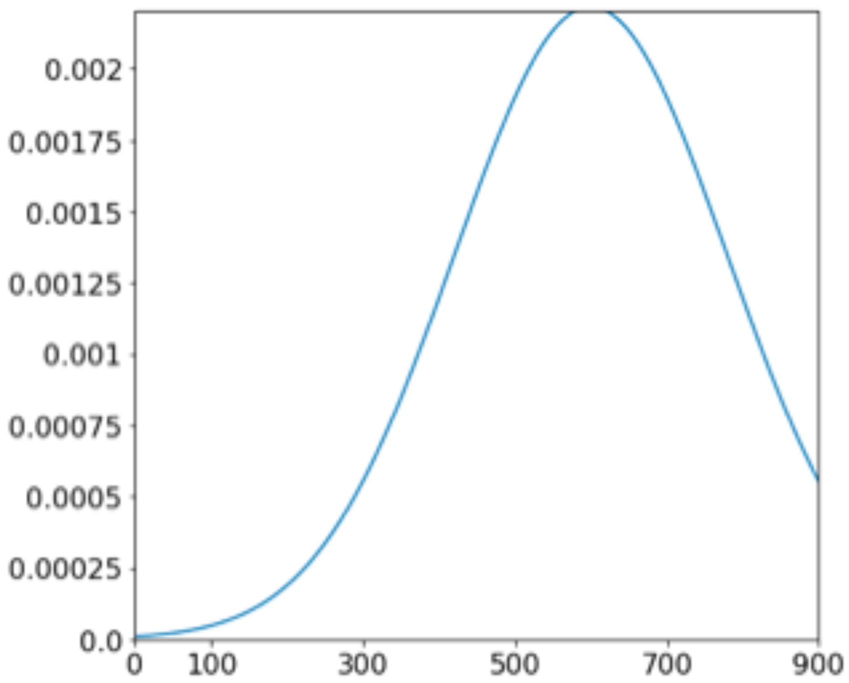
(side view)



Match X and Y to their respective marginal PDFs:



pixel x



pixel y

Extra slides

If you want more practice with double integrals,
I’ve included two exercises at the end of this lecture.

16b_joint_cdfs

Joint CDFs

An observation: Connecting CDF to PDF

For a continuous random variable X with PDF f, the CDF (cumulative distribution function) is

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

The density f is therefore the derivative of the CDF, F:

$$\boxed{f(a) = \frac{d}{da}F(a)}$$

(Fundamental Theorem of Calculus)

Joint cumulative distribution function

For two random variables X and Y, there can be a **joint cumulative distribution function** $F_{X,Y}$:

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

For discrete X and Y:

$$F_{X,Y}(a,b) = \sum_{x \le a} \sum_{y \le b} p_{X,Y}(x,y)$$

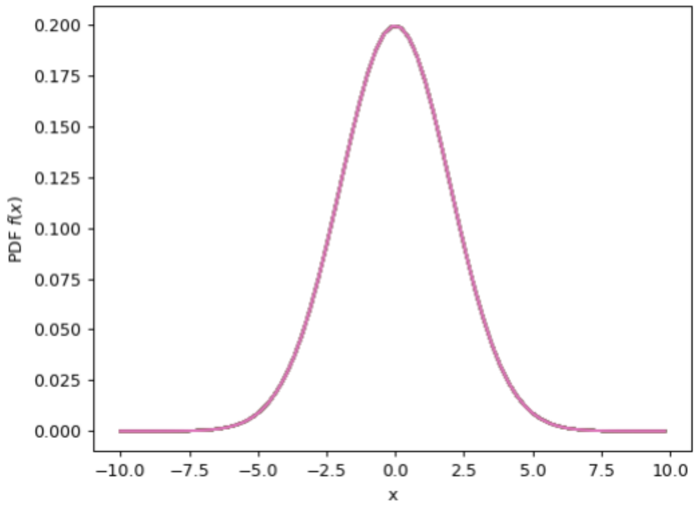
For continuous X and Y:

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

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

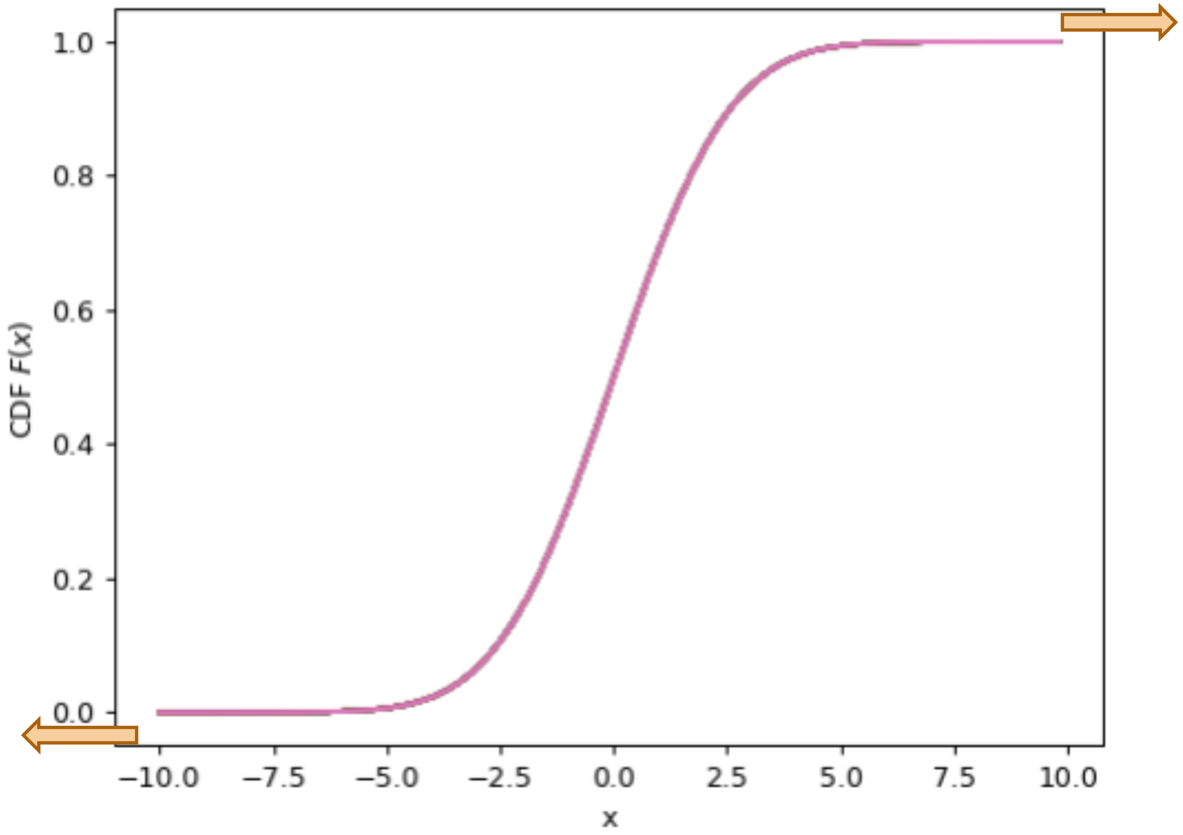
Single variable CDF, graphically

Review



$$f_X(x)$$

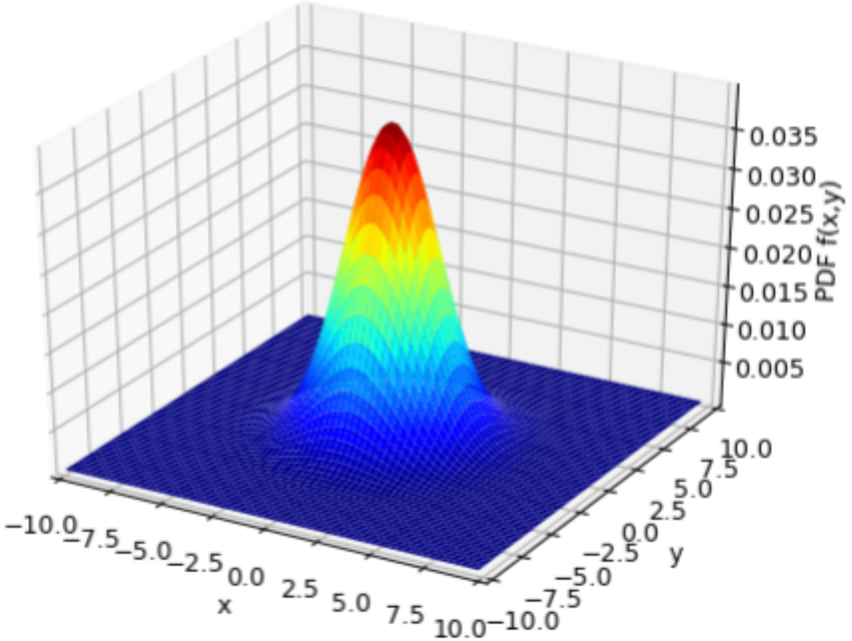
$$\lim_{x \to +\infty} F_X(x) = 1$$



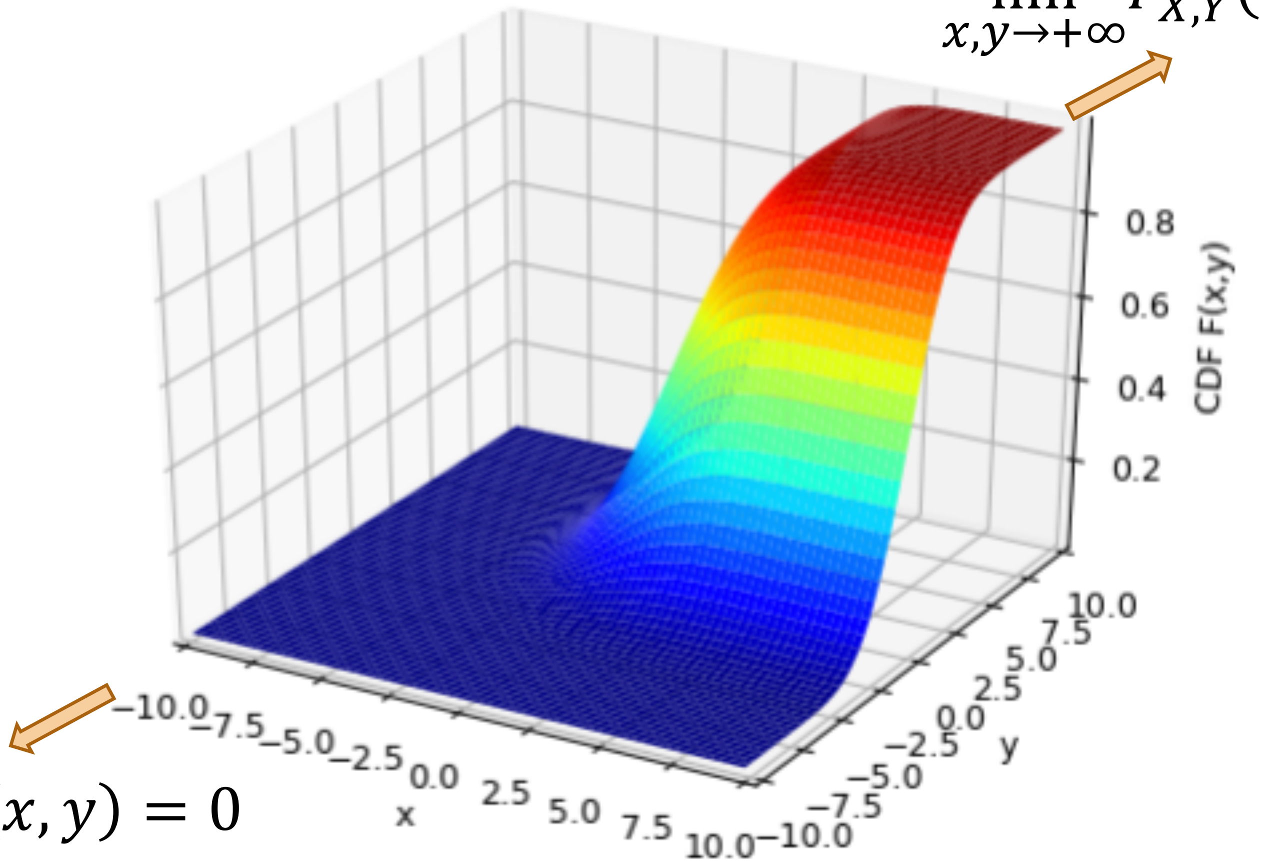
$$\lim_{x \to -\infty} F_X(x) = 0$$

$$F_X(x) = P(X \leq x)$$

Joint CDF, graphically



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



$$\lim_{x,y\to+\infty} F_{X,Y}(x,y) = 1$$

$$\lim_{x,y\to-\infty} F_{X,Y}(x,y) = 0$$

$$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

16c_indep_cont_rvs

Independent Continuous RVs

Independent continuous RVs

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:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$
$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Proof of PDF:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_X(x)F_Y(y) \qquad \forall x, y$$

$$= \frac{\partial}{\partial x}\underline{\frac{\partial}{\partial y} F_X(x)F_Y(y)} \qquad = \frac{\partial}{\partial x}F_X(x)\frac{\partial}{\partial y}F_Y(y)$$

$$= f_X(x)f_Y(y)$$

Independent continuous RVs

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

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

Equivalently:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$
$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

More generally, X and Y are independent if joint density factors separately:

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

Pop quiz! (just kidding)

$f_{X,Y}(x,y) = g(x)h(y)$,
where $-\infty < x, y < \infty$ ⇒ independent X and Y

Are X and Y independent in the following cases?

1. $f_{X,Y}(x,y) = 6e^{-3x}e^{-2y}$
   where $0 < x, y < \infty$
2. $f_{X,Y}(x,y) = 4xy$
   where $0 < x, y < 1$
3. $f_{X,Y}(x,y) = 24xy$
   where $0 < x + y < 1$

Pop quiz! (just kidding)

$f_{X,Y}(x,y) = g(x)h(y)$, where $-\infty < x, y < \infty$ → independent X and Y

Are X and Y independent in the following cases?

✓ 1. $f_{X,Y}(x,y) = 6e^{-3x}e^{-2y}$ where $0 < x, y < \infty$

Separable functions: $g(x) = 3e^{-3x}$

$h(y) = 2e^{-2y}$

✓ 2. $f_{X,Y}(x,y) = 4xy$ where $0 < x, y < 1$

Separable functions: $g(x) = 2x$

$h(y) = 2y$

$g(x) = 4x$

$h(y) = y$

✗ 3. $f_{X,Y}(x,y) = 24xy$ where $0 < x + y < 1$

Cannot capture constraint on $x + y$ into factorization!

$$\int_0^1\int_0^1 4xy = 1 = \underbrace{\int_0^1 c\cdot 4x\,dx}_{1} \cdot \underbrace{\int_0^1 \frac{1}{c}\cdot y\,dy}_{1}$$

If you can factor densities over all of the support, you have independence.

Bivariate Normal Distribution

Bivariate Normal Distribution

$X \sim N(\mu, \sigma^2)$ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

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

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

Can show that $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ (Ross chapter 6, example 5d)

Often written as:

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

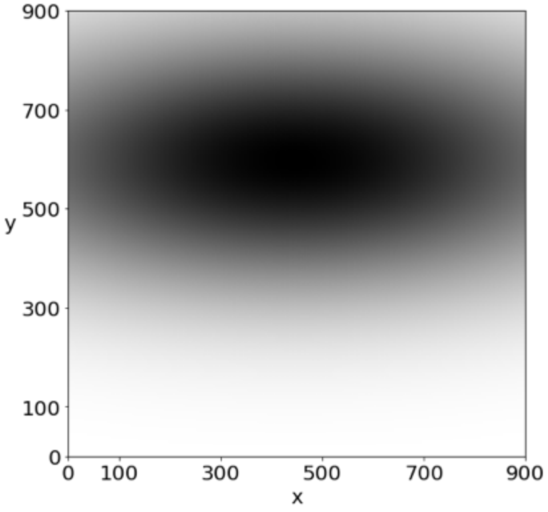
- Vector $\boldsymbol{X} = (X_1, X_2)$
- Mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$, Covariance matrix: $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) \end{bmatrix}$

Recall correlation: $\rho = \frac{\text{Cov}(X_1,X_2)}{\sigma_1\sigma_2}$

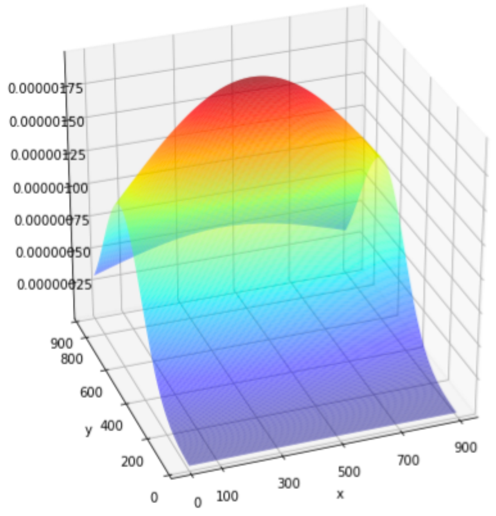
We will focus on understanding the **shape** of a bivariate Normal RV.

Back to darts

(top-down)



(side view)



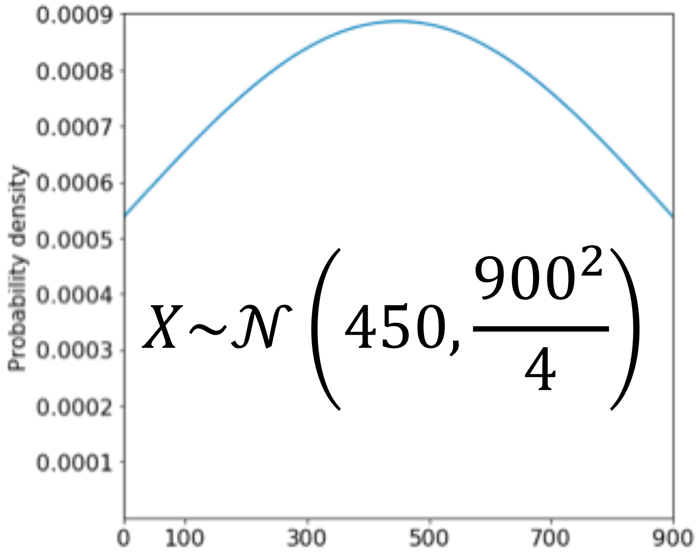
These darts were actually thrown according to a bivariate normal distribution:

$$(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} = (450, 600)$$

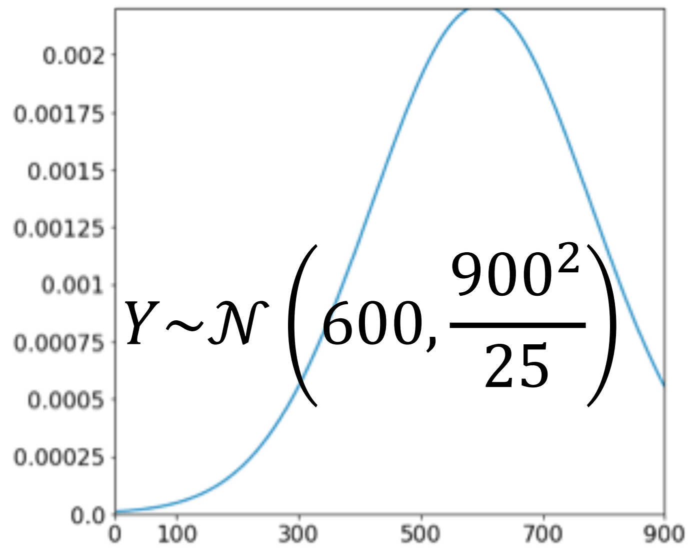
$$\boldsymbol{\Sigma} = \begin{bmatrix} 900^2/4 & 0 \\ 0 & 900^2/25 \end{bmatrix}$$

Marginal PDFs:



$$X \sim \mathcal{N}\left(450, \frac{900^2}{4}\right)$$

pixel x



$$Y \sim \mathcal{N}\left(600, \frac{900^2}{25}\right)$$

pixel y

A diagonal covariance matrix

Let $\boldsymbol{X} = (X_1, X_2)$ follow a bivariate normal distribution $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = (\mu_1, \mu_2), \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$\mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$
$= 0$

Are X_1 and X_2 independent?

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

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

(Note covariance: $\rho\sigma_1\sigma_2 = 0$)

$\rho = 0$

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

X_1 and X_2 are **independent** with marginal distributions $X_1 \sim \mathcal{N}(\mu_1\ \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2\ \sigma_2^2)$

(live)
16: Continuous Joint Distributions (I)

Lisa Yan and Jerry Cain

October 19, 2020

Jointly continuous RVs

Review

X and Y are jointly continuous if they have a joint PDF:

$$f_{X,Y}(x, y) \text{ such that } \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)dy\,dx = 1$$

Most things we've learned about discrete joint distributions translate:

Marginal distributions	$p_X(a) = \sum_y p_{X,Y}(a, y)$	$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y)dy$
Independent RVs	$p_{X,Y}(x, y) = p_X(x)p_Y(y)$	$f_{X,Y}(x, y) = f_X(x)f_Y(y)$
Expectation (e.g., LOTUS)	$E[g(X, Y)] = \sum_x \sum_y g(x, y)p_{X,Y}(x, y)$	$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dy\,dx$

…etc.

Big ideas today

Basics of jointly continuous RVs

- Independence, marginal PDFs
- Compute probability (i.e., definite double integrals)

Jointly distributed normal RVs

- Bivariate Normal ←
- Sum of independent Normals (part of next class's pre-lecture 17b)

Think

Slide 36 has a question to go over by yourself.

Post any clarifications here or in Zoom chat!

https://us.edstem.org/courses/2678/discussion/153770

Think by yourself: 2 min

🤔 (by yourself)

Warmup exercise

X and Y have the following joint PDF:

$$f_{X,Y}(x, y) = 3e^{-3x}$$
where $0 < x < \infty, 1 < y < 2$

1. Are X and Y independent?
2. What is the marginal PDF of X? Of Y?
3. What is $E[X + Y]$?

🤔 (by yourself)

Warmup exercise

X and Y have the following joint PDF:

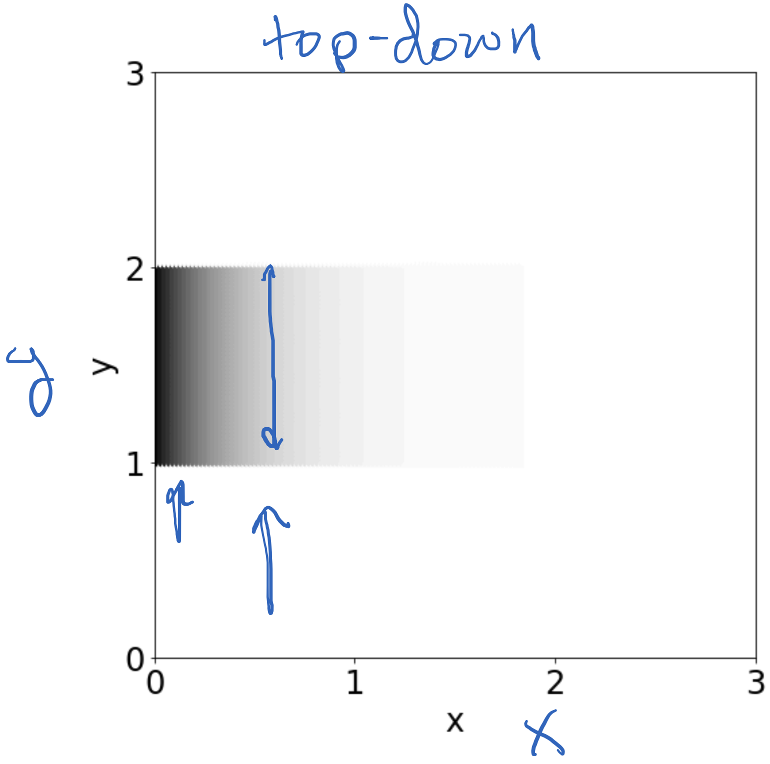
$$f_{X,Y}(x,y) = 3e^{-3x}$$
where $0 < x < \infty, 1 < y < 2$

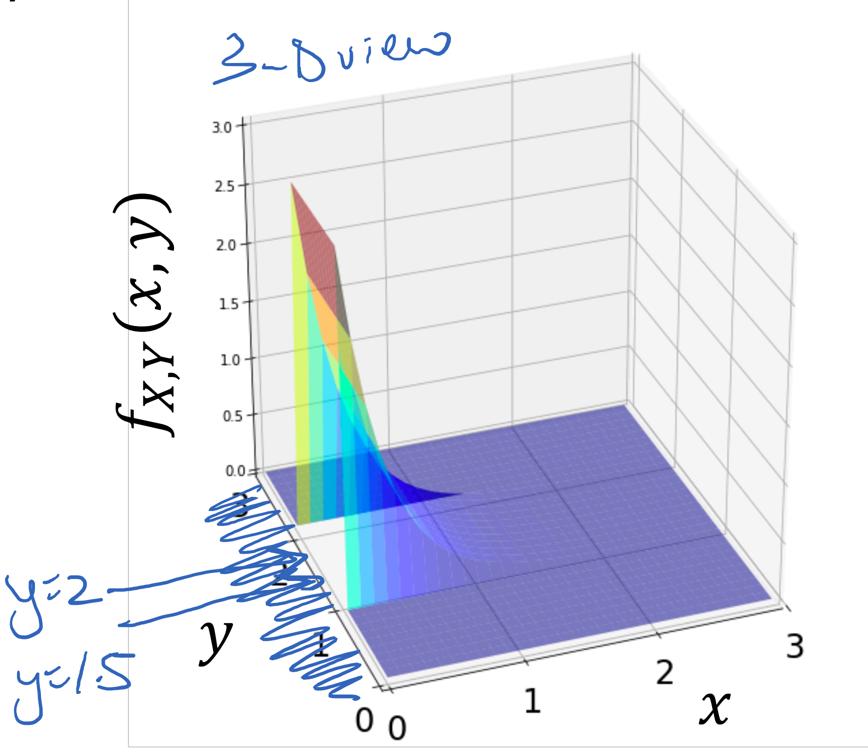
1. Are X and Y independent? ✅

$g(x) = 3Ce^{-3x}, 0 < x < \infty$
$h(y) = 1/C, \quad 1 < y < 2$

C is a constant

2. What is the marginal PDF of X? Of Y?
3. What is $E[X + Y]$?





Warmup exercise

X and Y have the following joint PDF:

$$f_{X,Y}(x, y) = 3e^{-3x}$$
where $0 < x < \infty, 1 < y < 2$

1. Are X and Y independent? ✅

$g(x) = 3Ce^{-3x}, 0 < x < \infty$
$h(y) = 1/C, \quad 1 < y < 2$

C is a constant

↑ $\int_1^2 1/c \, dy = 1 \Rightarrow C = 1$

2. What is the marginal PDF of X? Of Y?

$\underline{1 < y < 2}$ $\quad f_Y(y) = h(y) = \underline{1} \iff Y \sim Uni(a=1, b=2)$

$0 < x < \infty$ $\quad f_X(x) = g(x) = 3e^{-3x} \Rightarrow X \sim Exp(\lambda = 3)$

3. What is $E[X + Y]$?

<u>Strat 1</u> $E[g(X,Y)] = \iint g(x,y) f_{X,Y}(x,y)\, dx\, dy$

$= \iint (x+y) f_{X,Y}(x,y)\, dx\, dy$ LOTUS

<u>Strat 2</u> $E[X+Y] = E[X] + E[Y]$ Linearity

$= 1/3 + 3/2$

Breakout Rooms

Check out the question on the next slide (Slide 40). Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/153770

Breakout rooms: 4 min. Introduce yourself!

The joy of meetings

Two people set up a meeting time. Each arrives independently at a time uniformly distributed between 12pm and 12:30pm.

Define X = # minutes past 12pm that person 1 arrives. $X \sim \text{Unif}(0, 30)$

Y = # minutes past 12pm that person 2 arrives. $Y \sim \text{Unif}(0, 30)$

What is the probability that the first to arrive waits >10 mins for the other?

Compute: $P(X + 10 < Y) + P(Y + 10 < X) = 2P(X + 10 < Y)$ (by symmetry)

1. What is "symmetry" here?
2. How do we integrate to compute this probability?

$f_{X,Y}(x,y) = \left(\frac{1}{30}\right)^2$

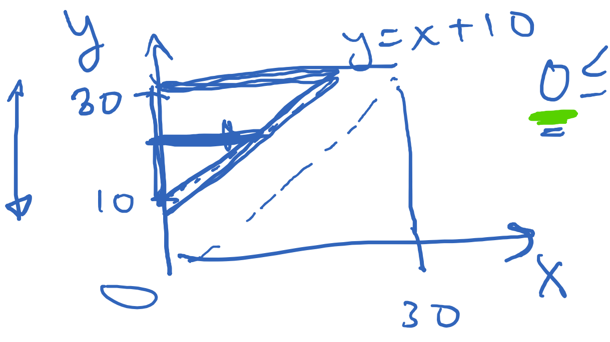
$$\underbrace{\iint}_{\substack{0 \le x, y \le 30 \\ x+10<y}} \left(\frac{1}{30}\right)^2 dx\, dy + \iint_{\substack{0 \le x, y \le 30 \\ y+10<x}} \left(\frac{1}{30}\right)^2 dy\, dx$$

Double integrals: A guide

From last slide: $$2P(X+10<Y) = 2\cdot \iint\limits_{\substack{x+10<y,\\ 0\le x,y,\le 30}} (1/30)^2 dxdy$$ (by symmetry, independence)

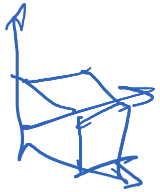
<u>Steps</u>:

1. Draw a picture.
2. Set bounds “from outside in.”
   - Outer integral bounds should be full range possible
   - Inner integral can depend on integration variable of outer integral



$0 \le x \le 30$

$x < y-10$



$$\boxed{= \frac{2}{30^2}\int_{10}^{30}\boxed{\int_0^{y-10} 1\, dx}dy}$$

$$= \frac{2}{30^2}\int_{10}^{30}\boxed{(y-10)}dy \qquad = \cdots = \frac{4}{9}$$

Su	M	Tu	W	Th	F	Sa

strong days??

Interlude for jokes/announcements

Announcements

Mid-quarter feedback form

link

Open until: this Friday 10/23

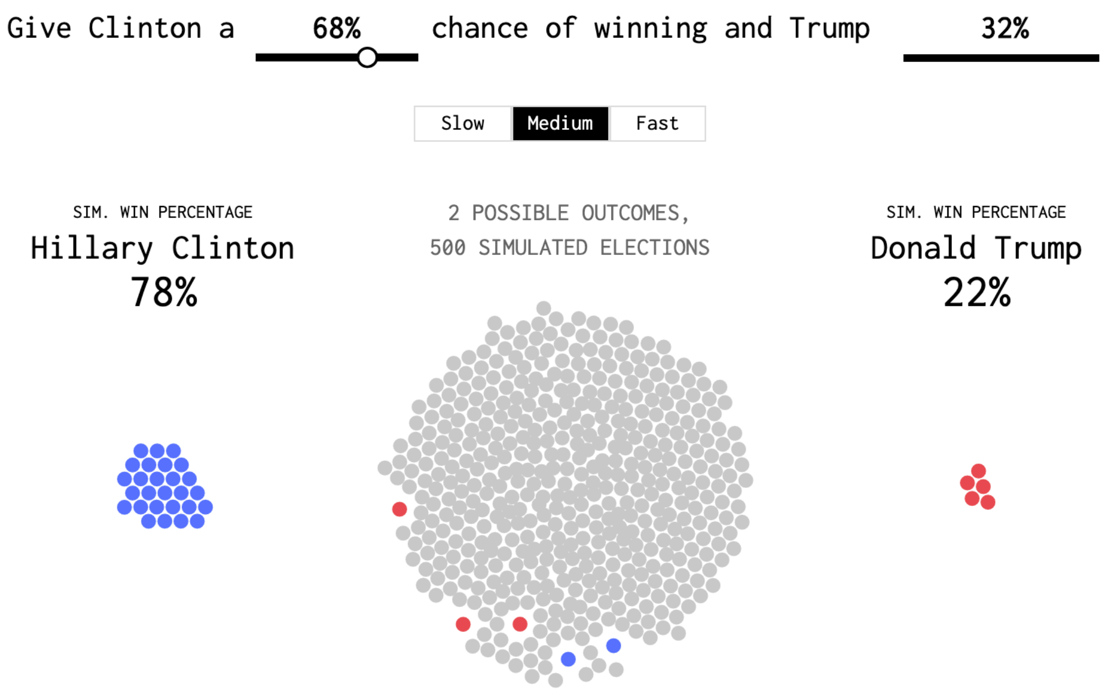
Python tutorial #3

When: Today (Mon) 6-7**pm** PT
Recorded? Yes
Covers: PS4-PS6 content
Notes: to be posted online
Zoom link: link

Interesting probability news

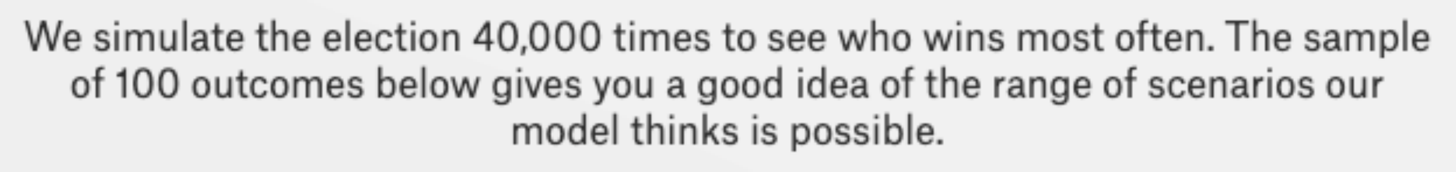
What That Election Probability Means

Even when you shift the probability far left or far right, the opposing candidate still gets some wins. That doesn't mean a forecast was wrong. That's just randomness and uncertainty at play.



Give Clinton a 68% chance of winning and Trump 32%

FiveThirtyEight 2020



We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

https://flowingdata.com/2016/07/28/what-that-election-probability-means/

Frequentist definition of probability!

Big ideas today

Basics of jointly continuous RVs

- Independence, marginal PDFs
- Compute probability (i.e., definite double integrals)

Jointly distributed normal RVs

- Bivariate Normal
- Sum of independent Normals (part of next class's pre-lecture 17b)

Bivariate normal distribution

Review

The bivariate normal distribution of $\boldsymbol{X} = (X_1, X_2)$:

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$

Cov(X,Y)

- Covariance matrix: $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ $\quad \text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1) = \rho\sigma_1\sigma_2$

Cov(Y,X) Cov(Y,Y)=Var(Y)

- Marginal distributions: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
- For bivariate normals in particular, $\text{Cov}(X_1, X_2) = 0$ implies X_1, X_2 **independent.**

We will focus on understanding the **shape** of a bivariate Normal RV.

Think

Check out the question on the next slide (Slide 47). Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/153770

Think by yourself: 1 min

🤔 (by yourself)

don't press Enter, Chat   1 D, 2 B, ...

(X, Y) Matching (all have $\boldsymbol{\mu} = (0, 0)$)

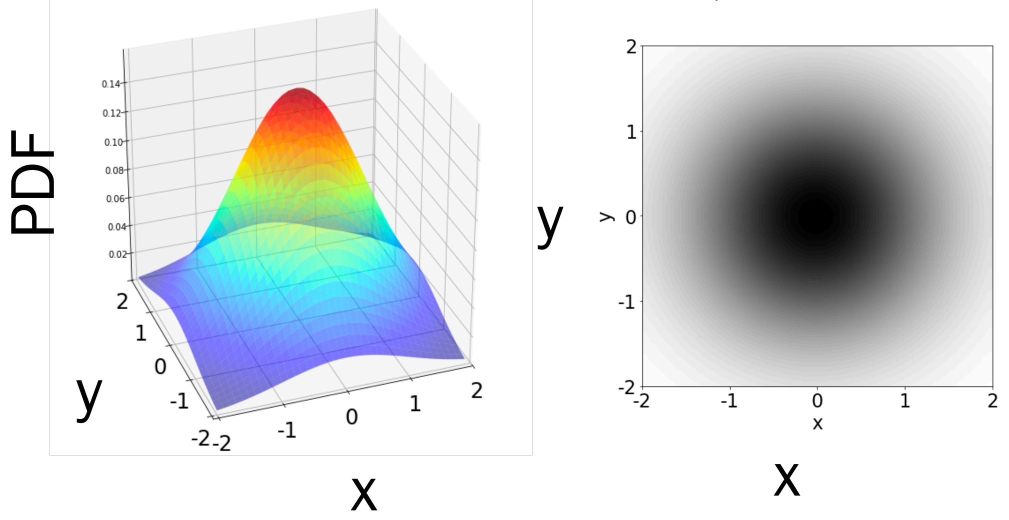
(by yourself)

A. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ B. $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

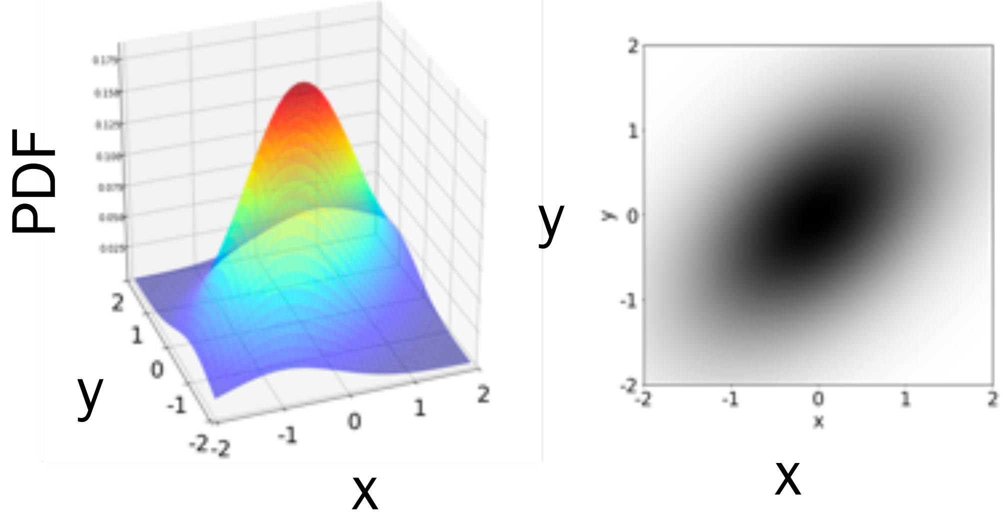
C. $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ D. $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$

3-D view   top-down view

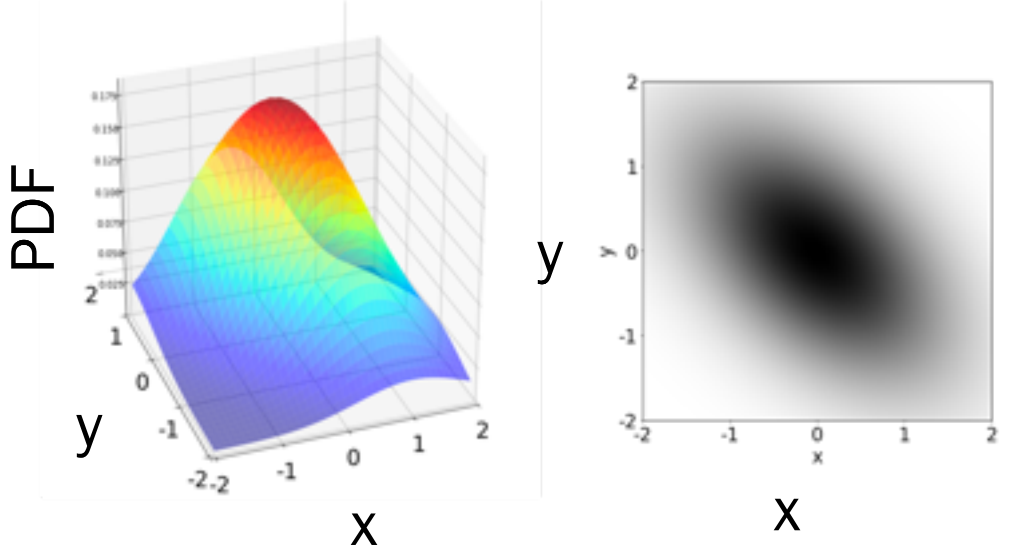
1.



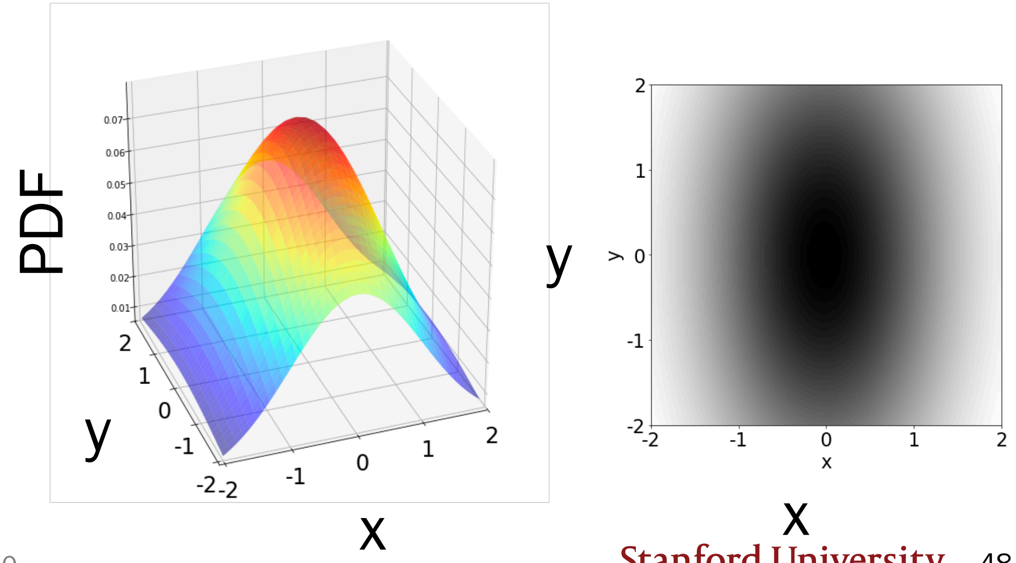
2.



3.



4.

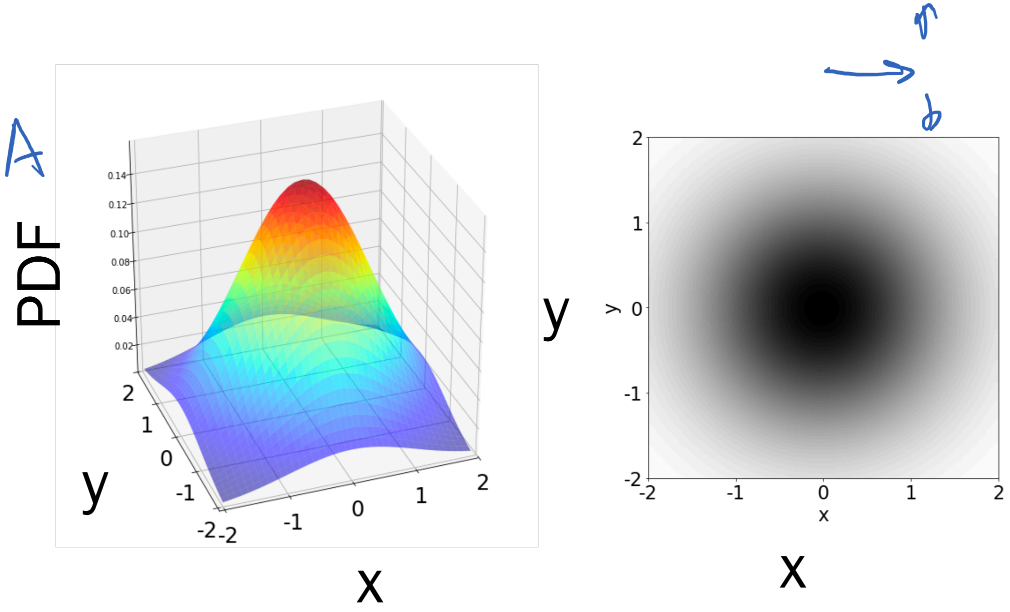


(X, Y) Matching (all have $\boldsymbol{\mu} = (0,0)$)

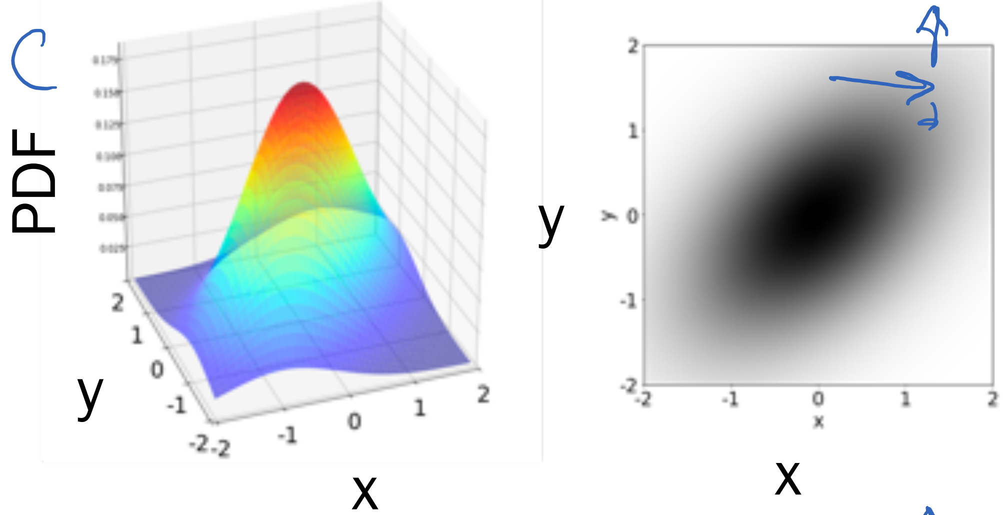
A. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ B. $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

C. $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ D. $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$

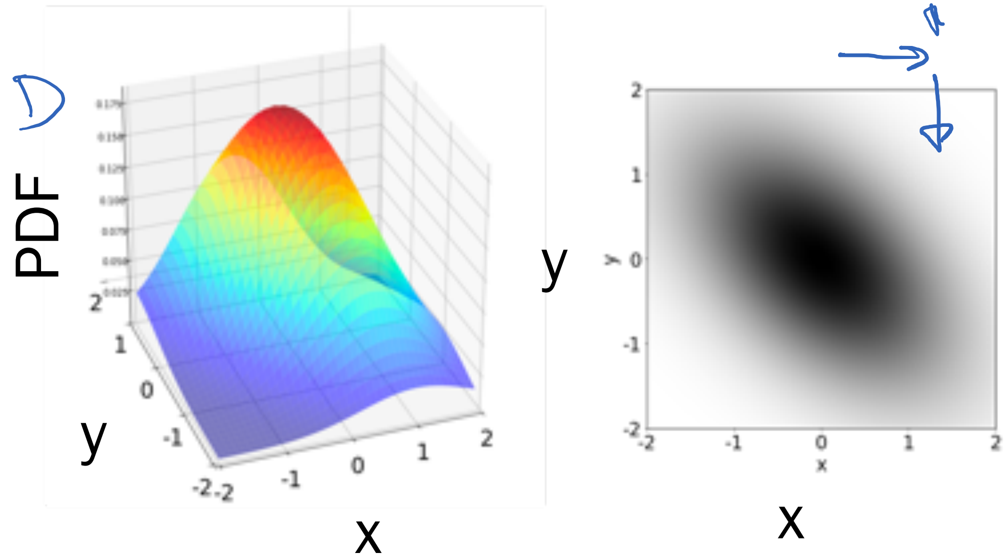
1. A



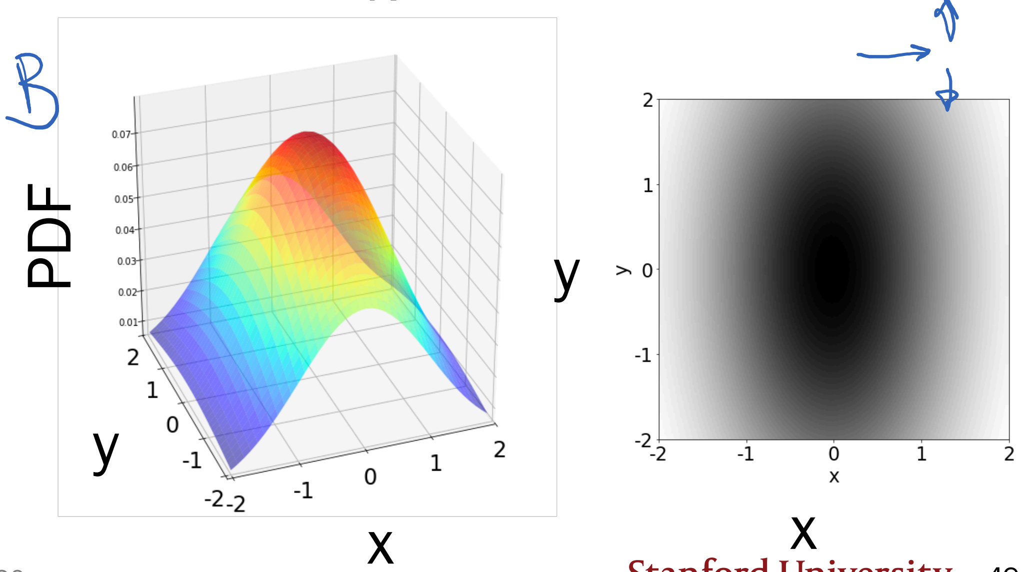
2. C



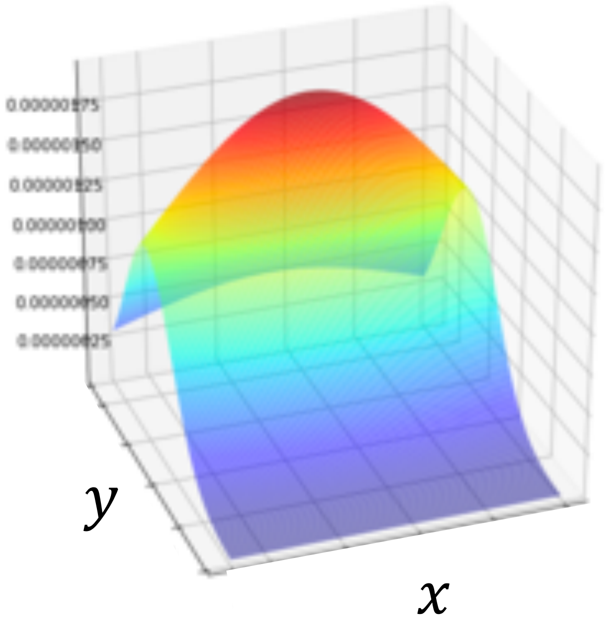
3. D



4. B



Why are joint PDFs useful?



- How 2 continuous RVs vary with each other

$P(X < Y)$,
$\text{Cov}(X, Y), \rho(X, Y)$

- How continuous RV is distributed given evidence (next time)

Given $Y = y$, the distribution of X

- How a continuous RV can be decomposed into 2 RVs (or vice versa)

Distribution of $Z = X + Y$
(which is a 1-D RV!)

Independence
2-D support
Joint PDF
Joint CDF
Marginal PDF
(next time) Conditional PDF

Sum of independent Normals

(covered more in
pre-lecture video 17b)

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$,
$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
X, X_2 independent

$$X_1 + X_2 \sim \underbrace{\mathcal{N}}_{}\left(\overbrace{\mu_1 + \mu_2}^{E[X_1+X_2] = E[X_1] + E[X_2]}, \overbrace{\sigma_1^2 + \sigma_2^2}^{\text{indep } X_1, X_2 \ Var(X_1+X_2)}\right)$$

(proof left to Wikipedia)

Wait, how is this related to linear transformations of Normals?

Recall: $X \sim N(\mu, \sigma^2)$

$$\text{If } Y = aX + b, \text{ then } Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Linear transforms vs. independence

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = X + X$. What is the distribution of Y?

- Are both approaches valid?

Independent RVs approach

Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
be independent.
Then $Y = X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Linear transform approach

Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
If $Y = aX + b$,
then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Linear transforms vs. independence ⚠

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = X + X$. What is the distribution of Y?

- Are both approaches valid?

Independent RVs approach ✗

Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
be independent.
Then $Y = X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$$Y = X + X$$

$$X + X \sim \mathcal{N}(\mu + \mu, \sigma^2 + \sigma^2)?$$

$$Y \sim \mathcal{N}(2\mu, \underline{2\sigma^2})?$$

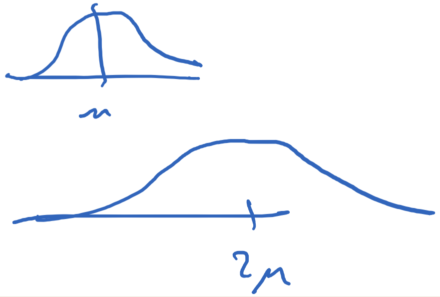
X is NOT independent of X!

Linear transform approach ✅

Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
If $Y = aX + b$,
then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

$$Y = \overset{a}{\overset{\uparrow}{2}}X$$

$$Y \sim \mathcal{N}(2\mu, 4\sigma^2)$$



For <u>independent</u> $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$,
$aX_1 + bX_2 + c \sim \mathcal{N}(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$

Breakout Rooms

(If time, otherwise we'll get to it next time)

Check out the question on the next slide (Slide 55). Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/153770

Breakout rooms: 4 min. Introduce yourself!

Virus infections

Suppose you are working with the WHO to plan a response to the initial conditions of a virus. There are two exposed groups:

- G1: 200 people, each independently infected with $p_1 = 0.1$
- G2: 100 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 55)$? An approximation is okay.

1. Define RVs & state goal

Let $A =$ # infected in G1.

$A \sim \text{Bin}(200, 0.1)$

$B =$ # infected in G2.

$B \sim \text{Bin}(100, 0.4)$

Want: $P(A + B \geq 55)$

Strategy:

A. Sum of indep. Binomials
B. (approximate) Sum of indep. Poissons
C. (approximate) Sum of indep. Normals
D. None/other

Virus infections

Suppose you are working with the WHO to plan a response to the initial conditions of a virus. There are two exposed groups:

- G1: 200 people, each independently infected with $p_1 = 0.1$
- G2: 100 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 55)$? An approximation is okay.

1. Define RVs & state goal

Let A = # infected in G1.

$A \sim \text{Bin}(200, 0.1)$

B = # infected in G2.

$B \sim \text{Bin}(100, 0.4)$

Want: $P(A + B \geq 55)$

Strategy:

A. Sum of indep. Binomials

B. (approximate) Sum of indep. Poissons

C. (approximate) Sum of indep. Normals

D. None/other

Virus infections

Suppose you are working with the WHO to plan a response to the initial conditions of a virus. There are two exposed groups:

- G1: 200 people, each independently infected with $p_1 = 0.1$
- G2: 100 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 55)$?

1. Define RVs & state goal

Let A = # infected in G1.

$A \sim \text{Bin}(200, 0.1)$

B = # infected in G2.

$B \sim \text{Bin}(100, 0.4)$

Want: $P(A + B \geq 55)$

2. Approximate as sum of Normals

$A \approx X \sim \mathcal{N}(20, 18) \qquad B \approx Y \sim \mathcal{N}(40, 24)$

$P(A + B \geq 55) \approx P(X + Y \geq 54.5)$ continuity correction

3. Solve

Virus infections

Suppose you are working with the WHO to plan a response to the initial conditions of a virus. There are two exposed groups:

- G1: 200 people, each independently infected with $p_1 = 0.1$
- G2: 100 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 55)$?

1. Define RVs & state goal

Let $A = \#$ infected in G1.

$A \sim \text{Bin}(200, 0.1)$

$B = \#$ infected in G2.

$B \sim \text{Bin}(100, 0.4)$

Want: $P(A + B \geq 55)$

2. Approximate as sum of Normals

$A \approx X \sim \mathcal{N}(20, 18) \qquad B \approx Y \sim \mathcal{N}(40, 24)$

$P(A + B \geq 55) \approx P(X + Y \geq 54.5)$ continuity correction

3. Solve

Let $W = X + Y \sim \mathcal{N}(20 + 40 = 60, 18 + 24 = 42)$

$$P(W \geq 54.5) = 1 - \Phi\left(\frac{54.5 - 60}{\sqrt{42}}\right) \approx 1 - \Phi(-0.85)$$

≈ 0.8023

Extra

1. Integral practice

Let X and Y be two continuous random variables with joint PDF:

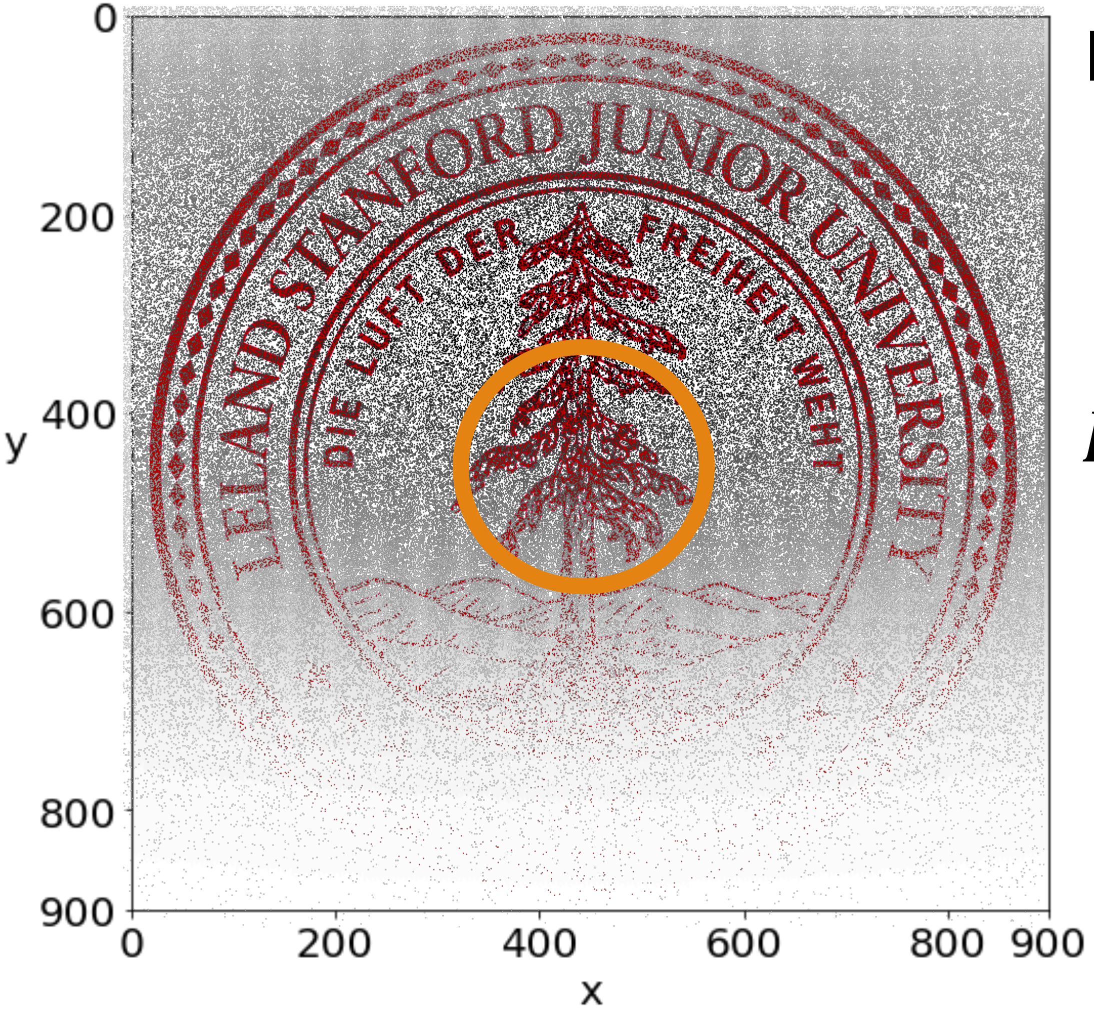
$$f(x,y) = \begin{cases} 4xy & 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X \le Y)$?

$$P(X \le Y) = \iint\limits_{\substack{x \le y, \\ 0 \le x, y \le 1}} 4xy\, dx\, dy = \int_{y=0}^{1} \int_{x \le y} 4xy\, dx\, dy \quad = \int_{y=0}^{1} \int_{x=0}^{y} 4xy\, dx\, dy$$

$$= \int_{y=0}^{1} 4y \left[\frac{x^2}{2}\right]_0^y dy \quad = \int_{y=0}^{1} 2y^3 dy \quad = \left[\frac{2}{4} y^4\right]_0^1 \quad = \frac{1}{2}$$

2. How do you integrate over a circle?



P(dart hits within $r = 10$ pixels of center)?

$$P(x^2 + y^2 \leq 10^2) = \iint_{x^2+y^2 \leq 10^2} f_{X,Y}(x, y) dy\, dx$$

Let’s try an example that doesn’t involve integrating a Normal RV ☺

2. Imperfection on Disk

You have a disk surface, a circle of radius R. Suppose you have a single point imperfection uniformly distributed on the disk.

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

What are the marginal distributions of X and Y? Are X and Y independent?

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{1}{\pi R^2} \int_{x^2+y^2 \leq R^2} dy$$ where $-R \leq x \leq R$

$$= \frac{1}{\pi R^2} \int_{y=-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \qquad = \frac{2\sqrt{R^2 - x^2}}{\pi R^2}$$

$$f_Y(y) = \frac{2\sqrt{R^2 - y}}{\pi R^2}$$ where $-R \leq y \leq R$, by symmetry

No, X and Y are **dependent.**

$$f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$$