Joint Distributions

Noah Arthurs

CS109, Stanford University

Announcements

Midterm:

- Tuesday the 23rd 7-9PM
- Covers through today*
- Unlimited notes/textbook, no calculator or computer.
- More review sheets coming today.

PS4:

- Out today! Problems above the line recommended for before midterm.

Review

And here we are

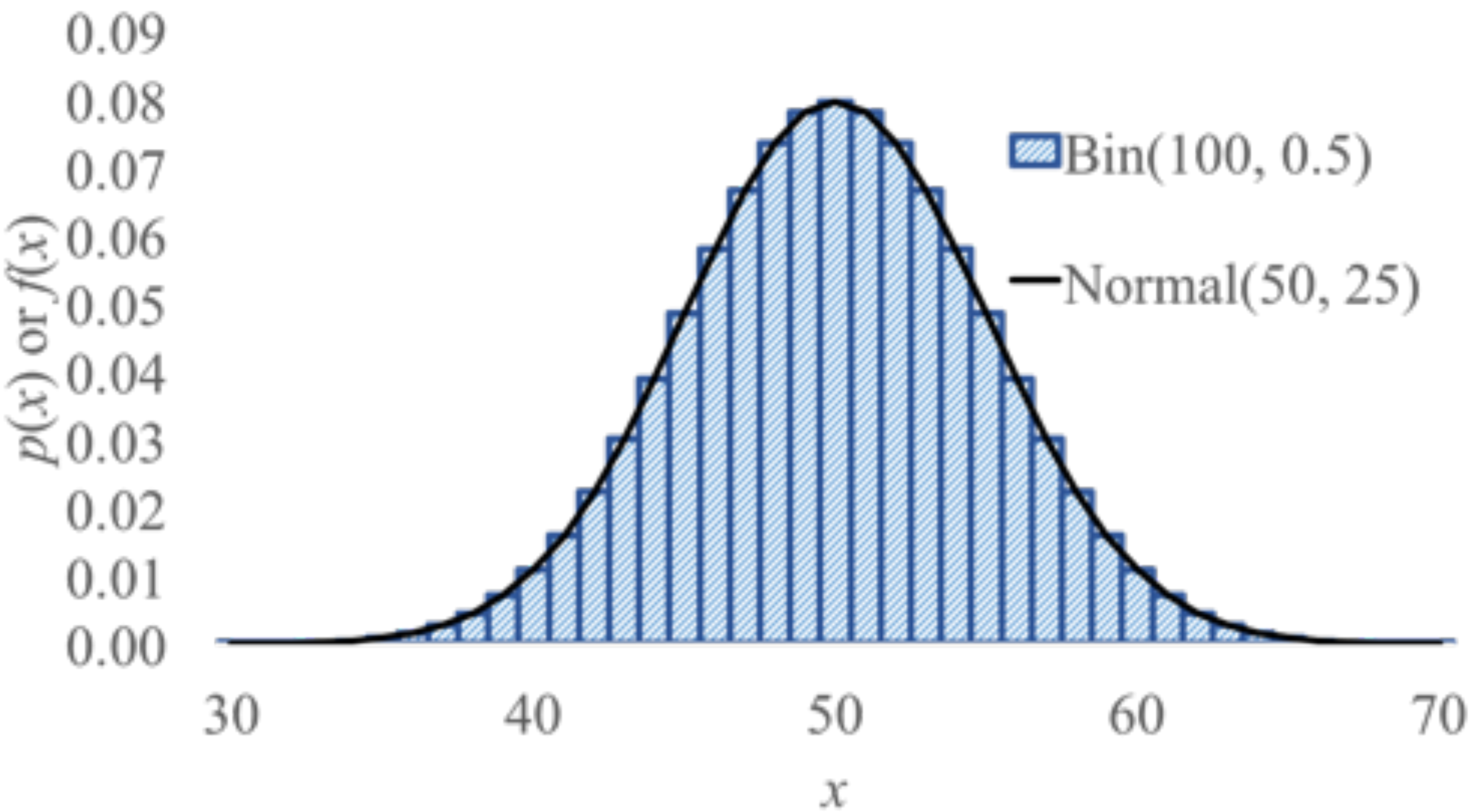
$$\mathcal{N}(\mu, \sigma^2)$$

CDF of Standard Normal: A function that has been solved for numerically

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

The cumulative density function (CDF) of any normal

Normal Approximates Binomial



Continuity Correction

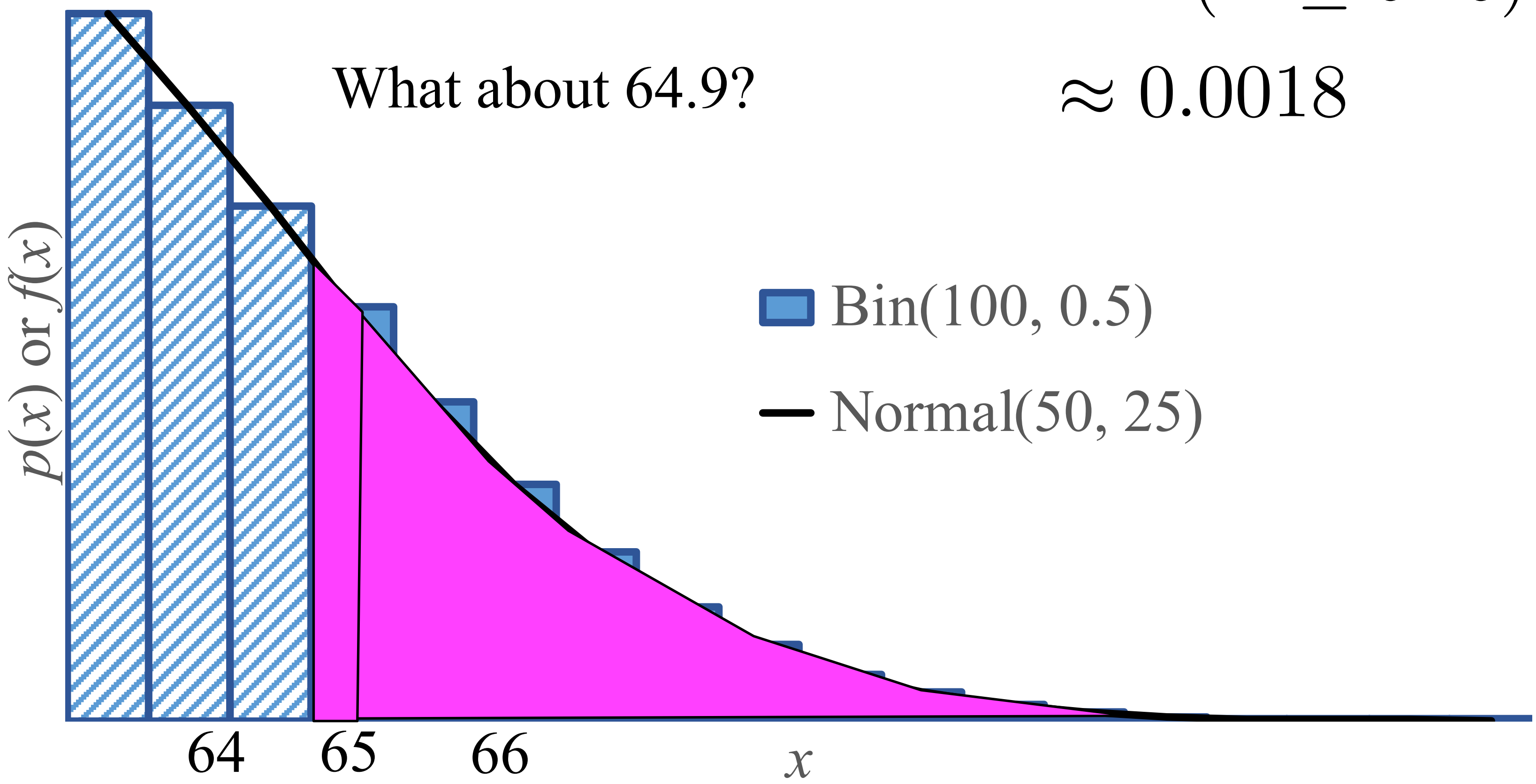
If Y (normal) approximates X (binomial)

$$P(X \geq 65)$$

$$\approx P(Y \geq 64.5)$$

$$\approx 0.0018$$

What about 64.9?

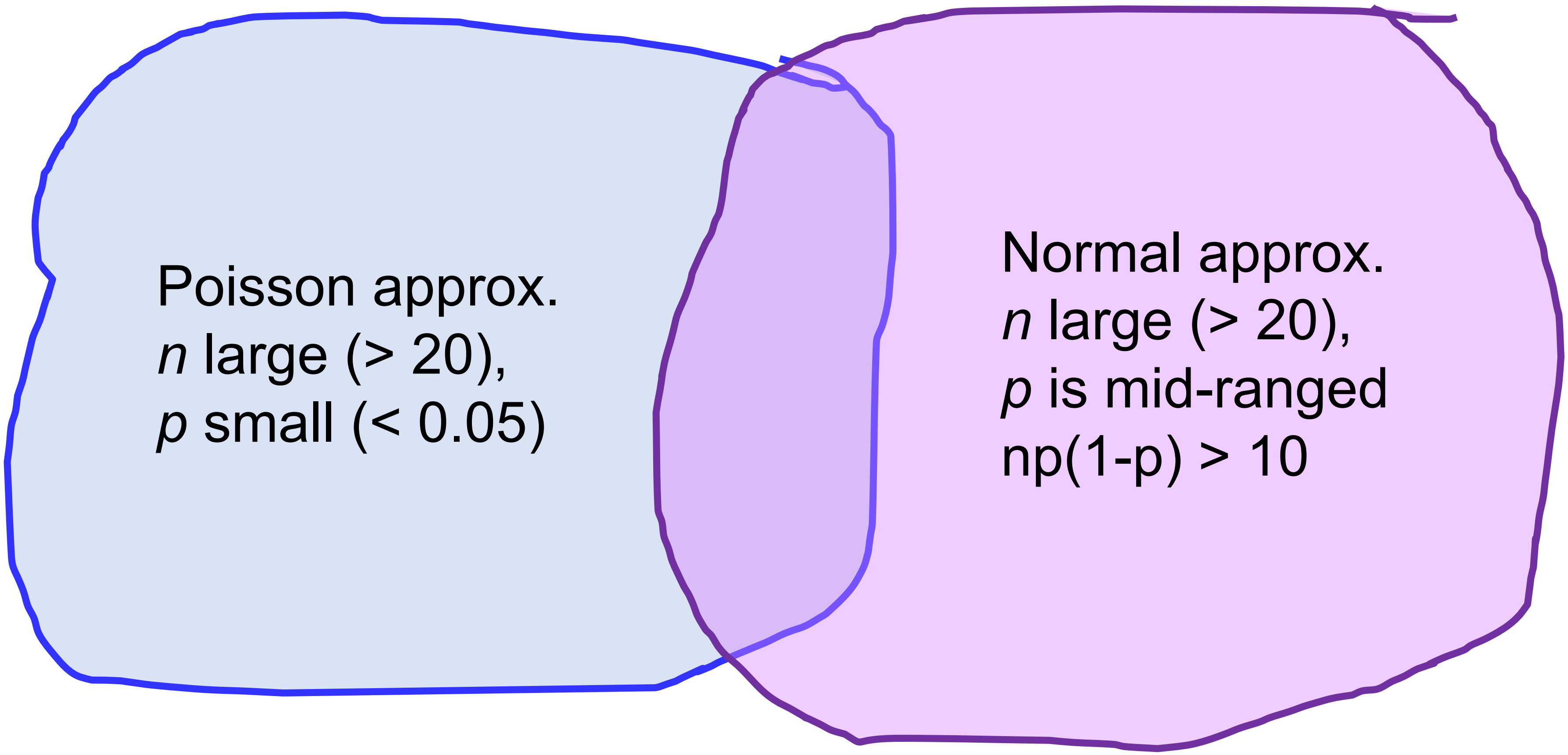


Continuity Correction

Use the continuity correction when approximating a **discrete value** with a **continuous distribution**

Who Gets to Approximate?

X ~ Bin(*n*, *p*)



If there is a choice, go with the normal approximation

Probability Table for Discrete

- States all possible outcomes with several discrete variables
- A probability table is not "parametric"
- If #variables is > 2, you can have a probability table, but you can't draw it on a slide

All values of A

All values of B

	0	1	2
0			
1		$P(A = 1, B = 1)$	
2			

Every outcome falls into a bucket

Here "," means "and"

Discrete Joint Mass Function

- For two discrete random variables *X* and *Y*, the **Joint Probability Mass Function** is:

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

- Marginal distributions:

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

$$p_Y(b) = P(Y = b) = \sum_x p_{X,Y}(x, b)$$

- Example: X = value of die D_1, Y = value of die D_2

$$P(X = 1) = \sum_{y=1}^{6} p_{X,Y}(1, y) = \sum_{y=1}^{6} \frac{1}{36} = \frac{1}{6}$$

A Computer (or Three) In Every House

- Consider households in Silicon Valley
  - A household has X Macs and Y PCs
  - Can't have more than 3 Macs or 3 PCs

Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	?	0.04	
1	0.12	0.14	0.12	0	
2	0.07	0.12	0	0	
3	0.04	0	0	0	
$p_X(x)$					

A Computer (or Three) In Every House

- Consider households in Silicon Valley
  - A household has X Macs and Y PCs
  - Can't have more than 3 Macs or 3 PCs

Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	0.07	0.04	
1	0.12	0.14	0.12	0	
2	0.07	0.12	0	0	
3	0.04	0	0	0	
$p_X(x)$					

A Computer (or Three) In Every House

- Consider households in Silicon Valley
  - A household has X Macs and Y PCs
  - Can't have more than 3 Macs or 3 PCs

Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	0.07	0.04	0.39
1	0.12	0.14	0.12	0	0.38
2	0.07	0.12	0	0	0.19
3	0.04	0	0	0	0.04
$p_X(x)$	0.39	0.38	0.19	0.04	1.00

Marginal distributions

End Review

Permutations

How many ways are there to order n distinct objects?

$$n!$$

Binomial

How many ways are there to make an unordered selection of r objects from n objects?

How many ways are there to order n objects such that:
r are the same (indistinguishable)
$(n-r)$ are the same (indistinguishable)?

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Called the "binomial" because of something from Algebra

Binomial Distribution

- Consider *n* independent trials of Ber(p) rand. var.
  - X is number of successes in *n* trials
  - X is a **Binomial** Random Variable: X ~ Bin(n, p)

Binomial # ways of ordering the successes

$$P(X=i) = p(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0,1,\ldots,n$$

Probability of exactly *i* successes

Probability of each ordering of *i* successes is equal + mutually exclusive

Multinomial

How many ways are there to order n objects such that:

n_1 are the same (indistinguishable)

n_2 are the same (indistinguishable)

$\dots$

n_r are the same (indistinguishable)?

$$\frac{n!}{n_1! n_2! \dots n_r!} = \binom{n}{n_1, n_2, \dots, n_r}$$

Note: Multinomial > Binomial

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 $\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1! c_2! \cdots c_m!}$

Hello Die Rolls, My Old Friends

- 6-sided die is rolled 7 times
  - Roll results: 1 one, 1 two, 0 three, 2 four, 0 five, 3 six

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3)$$

$$= \frac{7!}{1!1!0!2!0!3!}\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^3 = 420\left(\frac{1}{6}\right)^7$$

- This is generalization of Binomial distribution
  - Binomial: each trial had 2 possible outcomes
  - Multinomial: each trial has *m* possible outcomes

Probabilistic Text Analysis

- Ignoring order of words, what is probability of any given word you write in English?
  - P(word = “the”) > P(word = “transatlantic”)
  - P(word = “Stanford”) > P(word = “Cal”)
  - Probability of each word is just multinomial distribution
- What about probability of those same words in someone else’s writing?
  - P(word = “probability” | writer = you) >

    P(word = “probability” | writer = non-CS109 student)
  - After estimating P(word | writer) from known writings, use Bayes’ Theorem to determine P(writer | word) for new writings!

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.

Text is a Multinomial

Example document:
"Pay for Viagra with a credit-card. Viagra is great. So are credit-cards. Risk free Viagra. Click for free."
n = 18

It's a Multinomial!

$$P\left(\begin{array}{l}\text{Viagra} = 2\\ \text{Free} = 2\\ \text{Risk} = 1\\ \text{Credit-card: } 2\\ \dots\\ \text{For} = 2\end{array} \middle| \text{spam}\right) = \frac{n!}{2!2!\dots2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \dots p_{\text{for}}^2$$

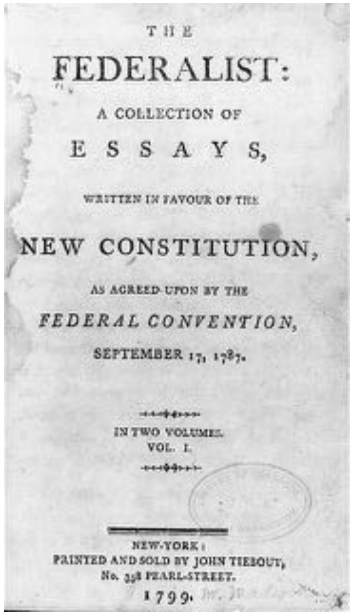
Probability of seeing this document | spam

The probability of a word in spam email being viagra

Who wrote the federalist papers?

Old and New Analysis

- Authorship of "Federalist Papers"
  - 85 essays advocating ratification of US constitution
  - Written under pseudonym "Publius"
    - Really, Alexander Hamilton, James Madison and John Jay
  - Who wrote which essays?
    - Analyzed probability of words in each essay versus word distributions from known writings of three authors

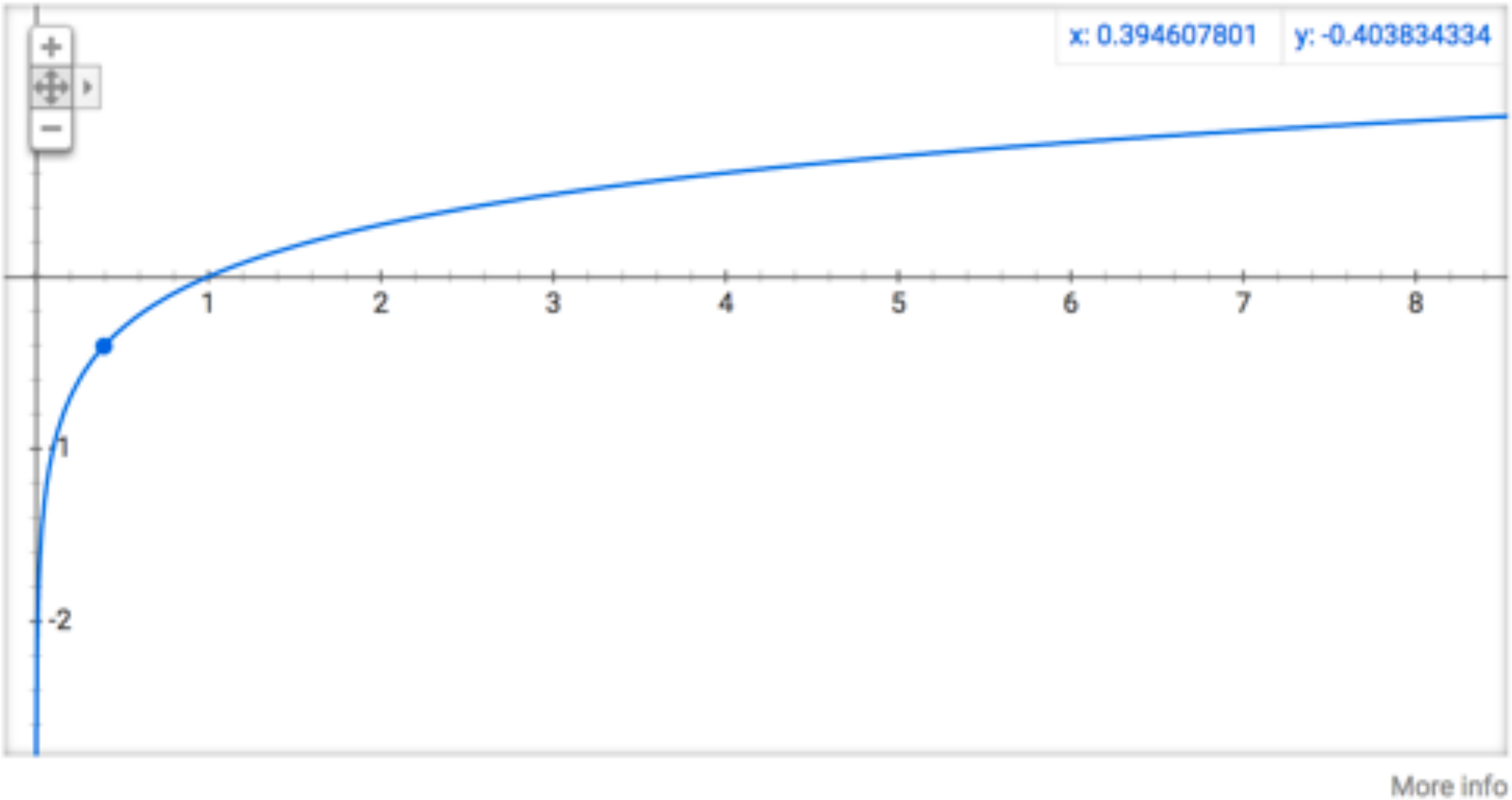

THE
FEDERALIST:
A COLLECTION OF
ESSAYS,
WRITTEN IN FAVOUR OF THE
NEW CONSTITUTION,
AS AGREED UPON BY THE
FEDERAL CONVENTION,
SEPTEMBER 17, 1787.
IN TWO VOLUMES.
VOL. I.
NEW-YORK:
PRINTED AND SOLD BY JOHN TIEBOUT,
No. 358 PEARL-STREET.
1799.

Let’s write a program!

Log Review

$$e^y = x \qquad \log(x) = y$$

Graph for log(x)



Log Identities

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

$$\log(a/b) = \log(a) - \log(b)$$

$$\log(a^n) = n \cdot \log(a)$$

Products become Sums!

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

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

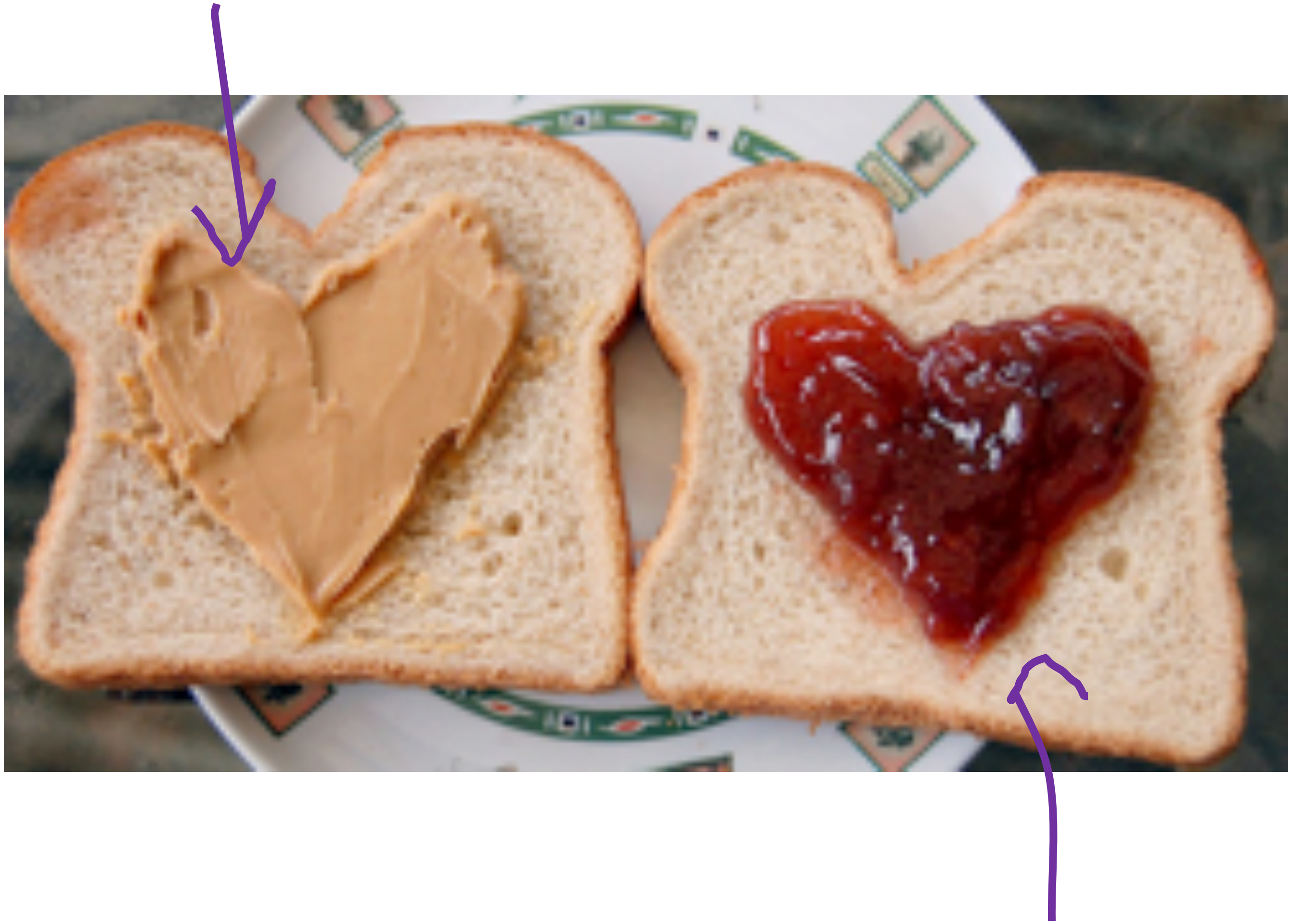
This is important because the product of many small numbers gets hard for computers to represent.

Stretch!

Continuous Joint

Noah Arthurs
CS109, Stanford University

Continuous Random Variables



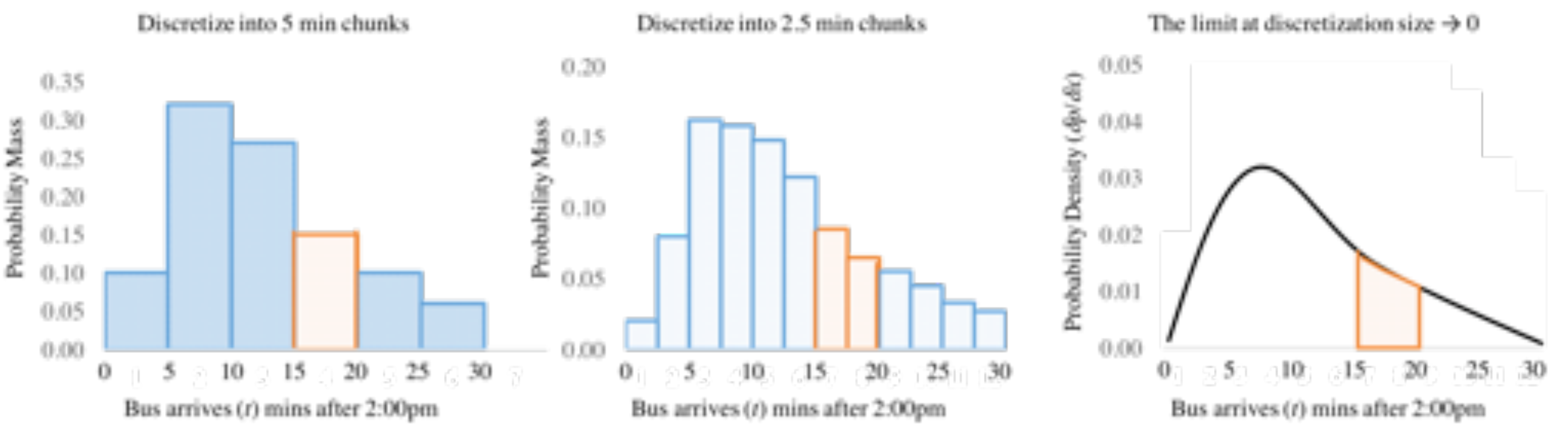
Joint Distributions

Continuous Joint Distribution

Riding the Marguerite

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

What is P(wait < 5 min)?



Joint Dart Distribution

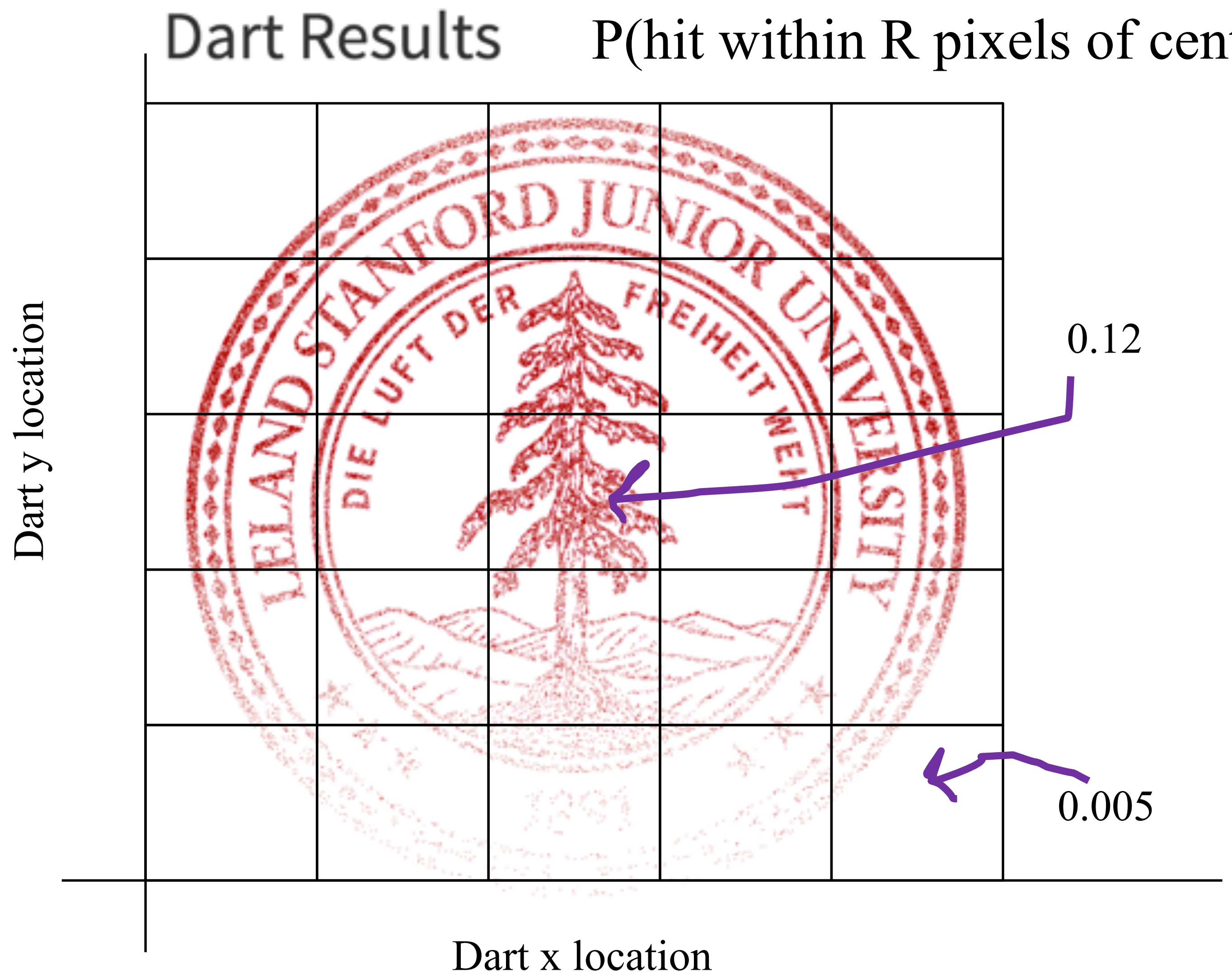
Dart Results

P(hit within R pixels of center)?

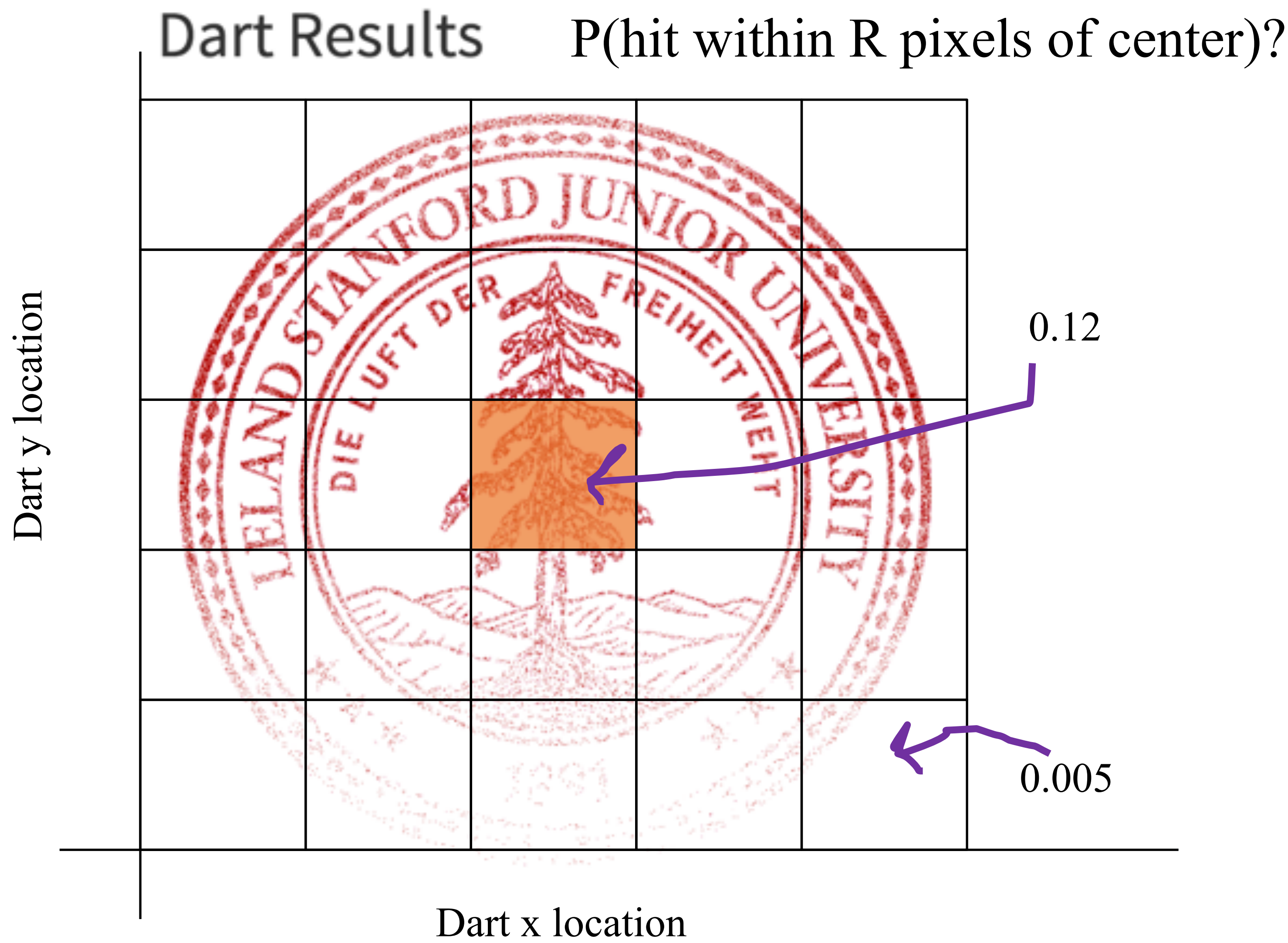
What is the probability that a dart hits at (456.234231234122355, 532.12344123456)?

Joint Dart Distribution

P(hit within R pixels of center)?

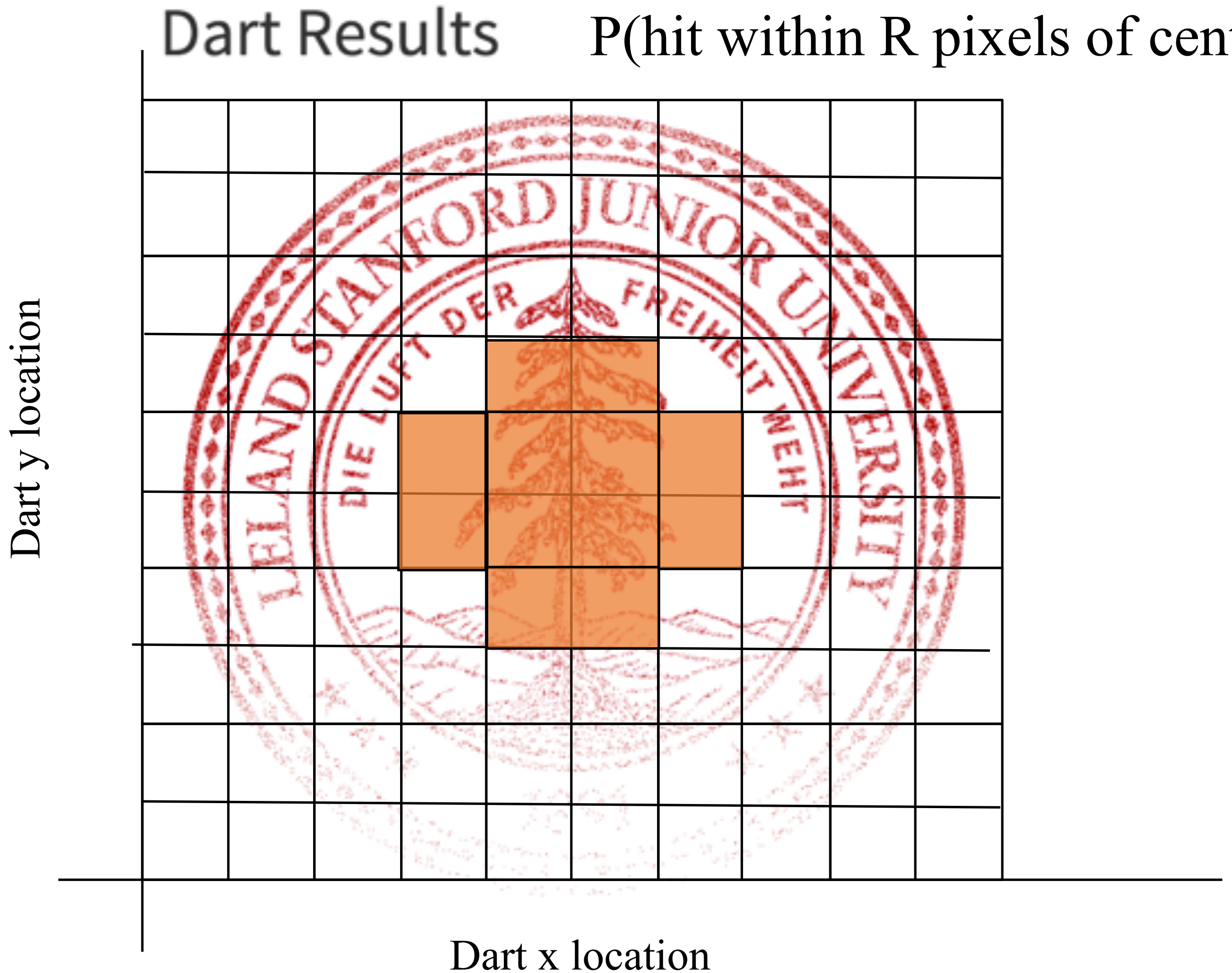


Joint Dart Distribution

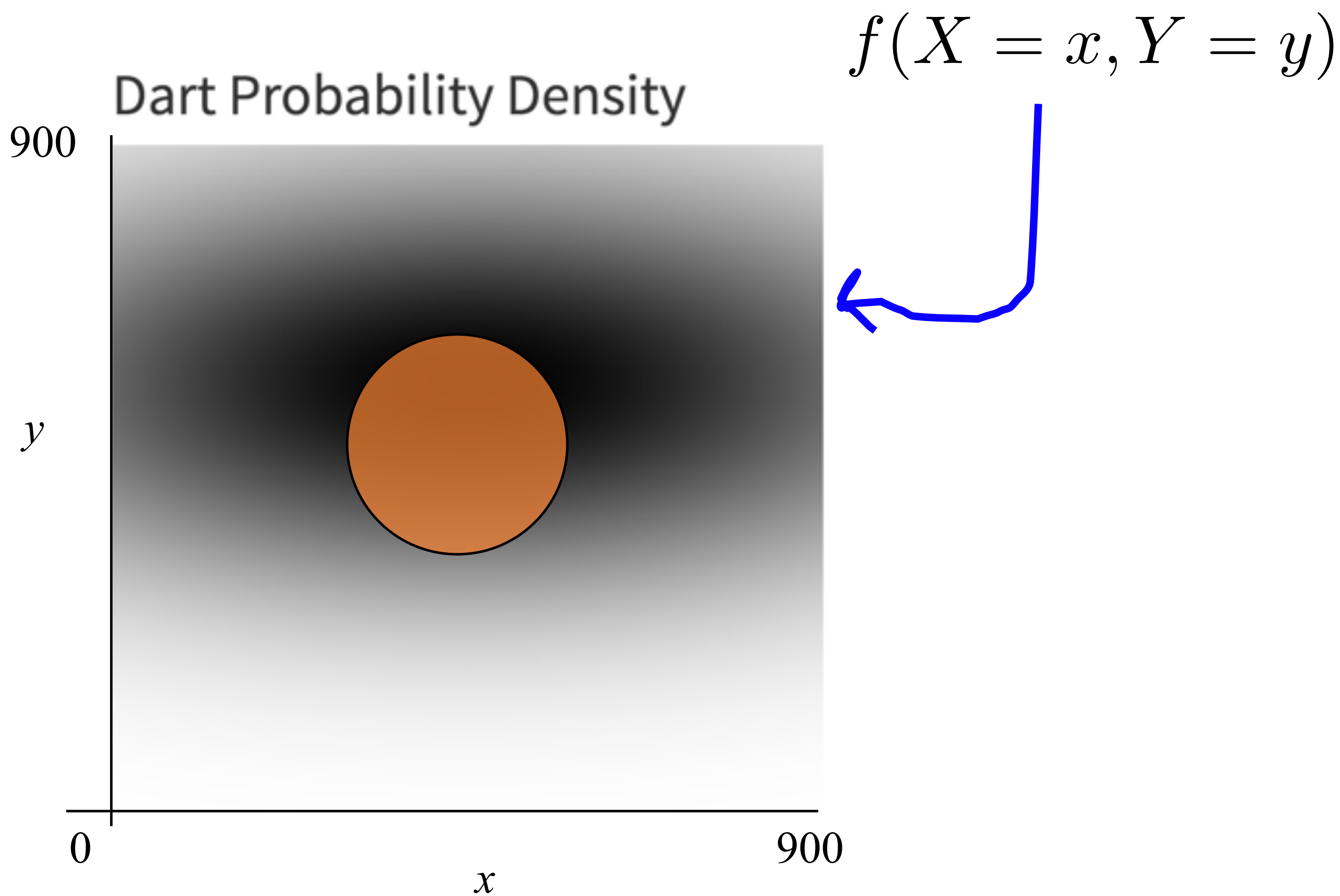


Joint Dart Distribution

P(hit within R pixels of center)?



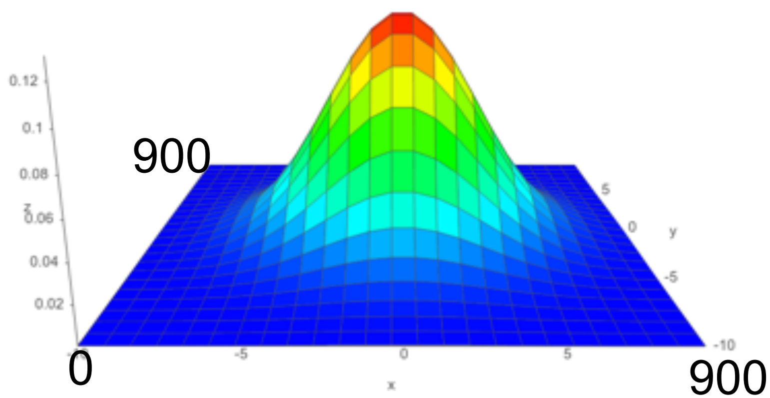
Joint Dart Distribution

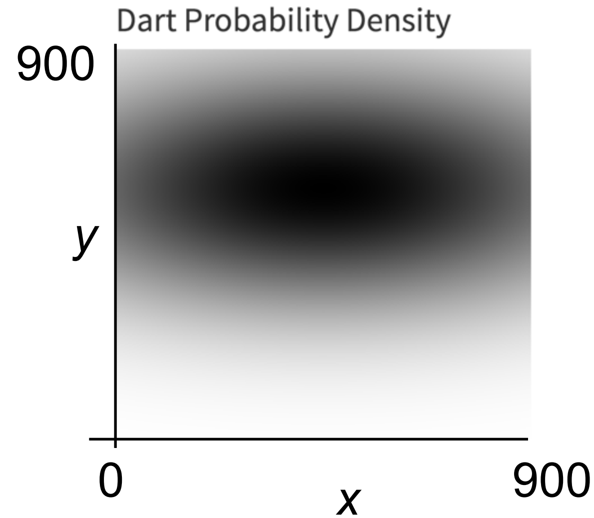


In the limit, as you break down continuous values into intestinally small buckets, you end up with multidimensional probability density

Joint Probability Density Funciton

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

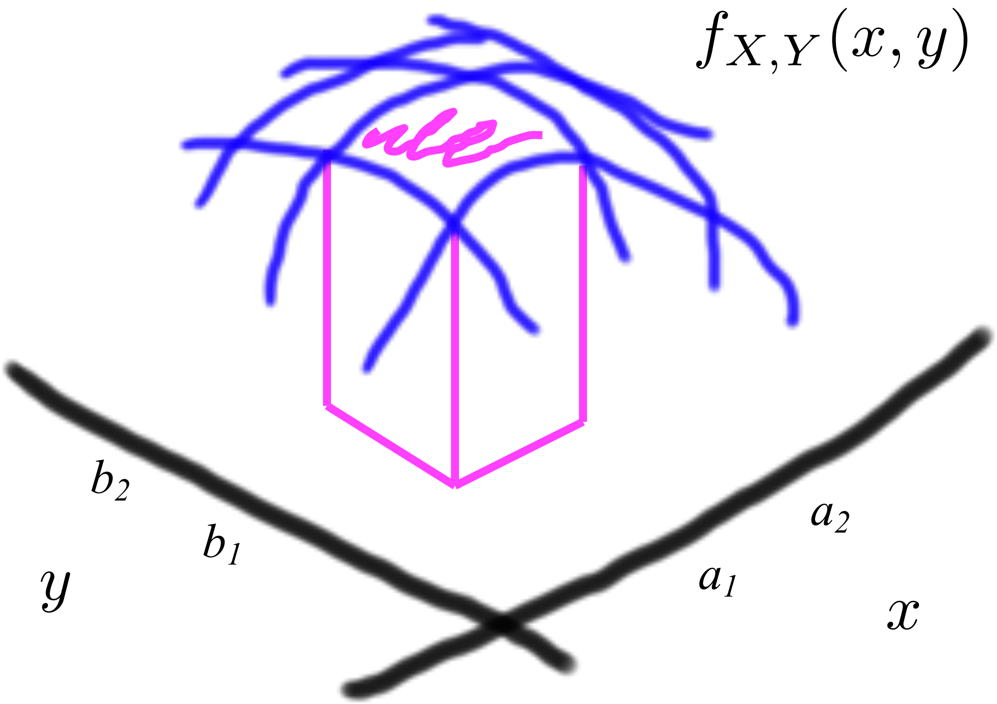




$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$

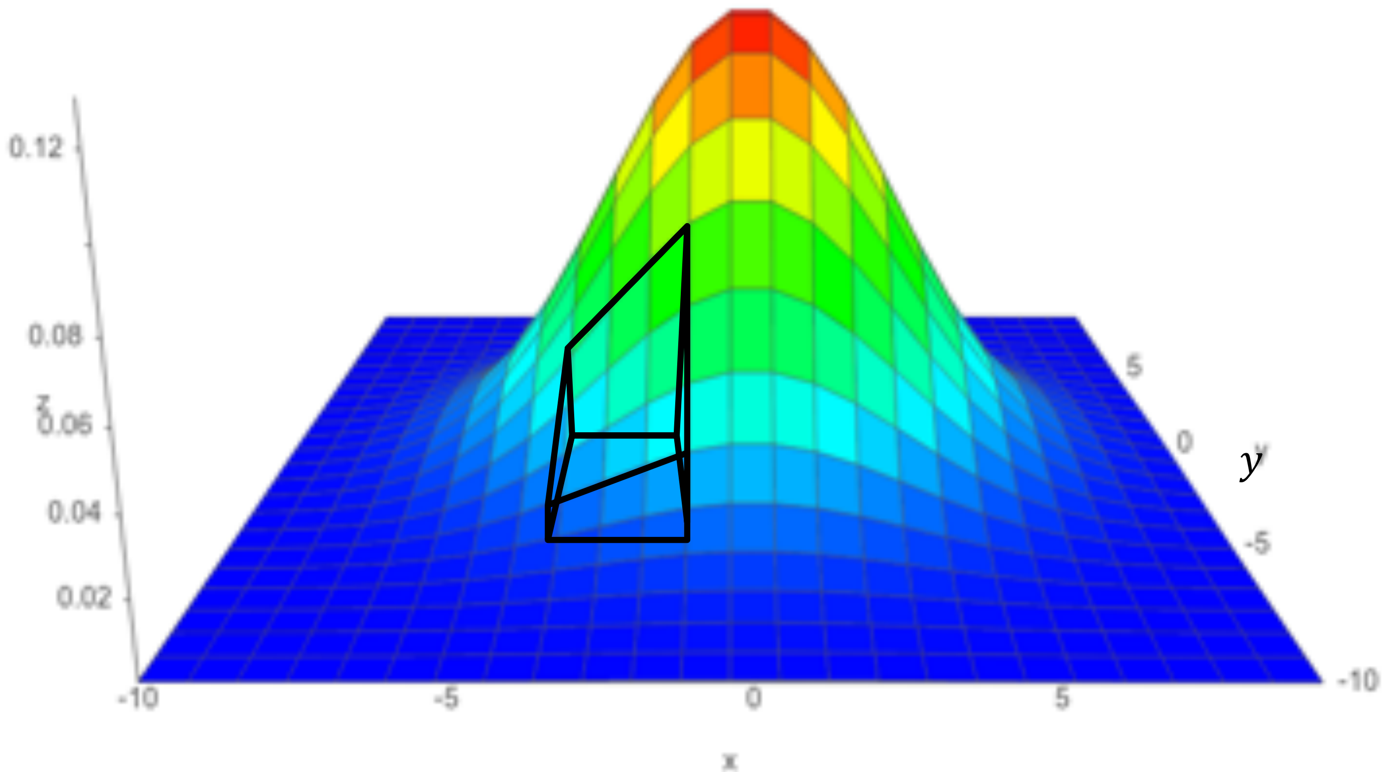
Joint Probability Density Funciton

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



Joint Probability Density Function

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



plot by Academo

Multiple Integrals Without Tears

- Let *X* and *Y* be two continuous random variables
  - where $0 \le X \le 1$ and $0 \le Y \le 2$
- We want to integrate $g(x,y) = xy$ w.r.t. *X* and *Y*:
  - First, do "innermost" integral (treat *y* as a constant):

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

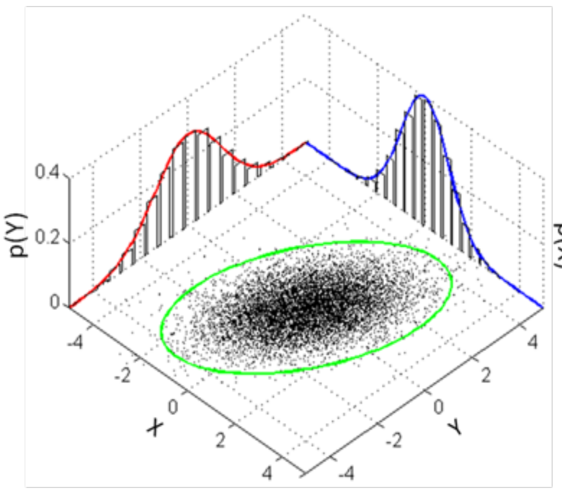
  - Then, evaluate remaining (single) integral:

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

Marginalization

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

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

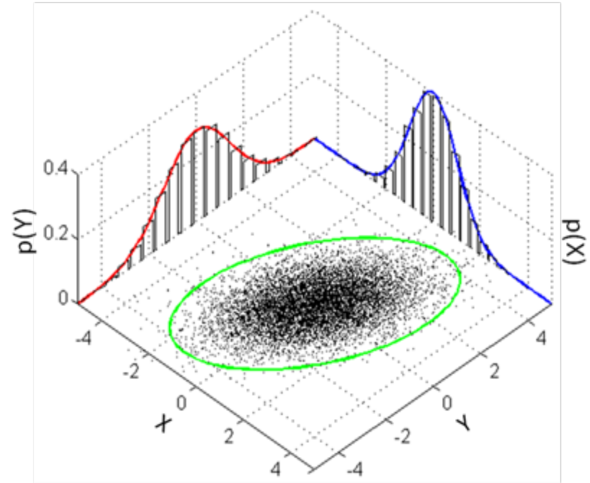


$$p_X(a) = \sum_y p_{X,Y}(a, y)$$

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

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

Marginalization



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

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

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

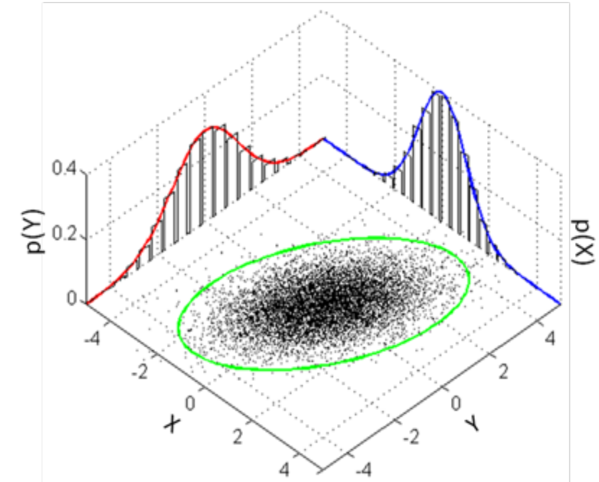
$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y)\ dy$$

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

Marginalization

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

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



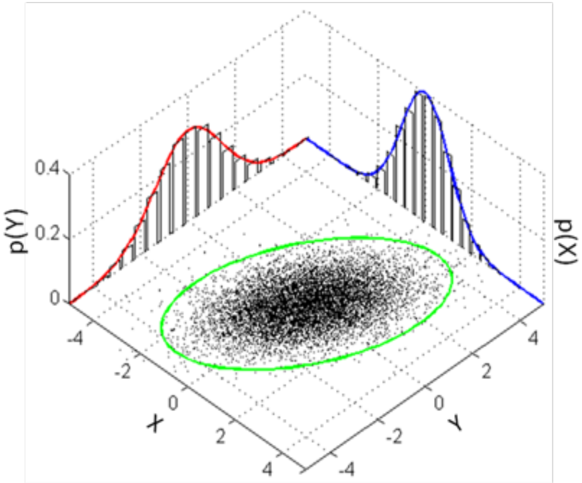
$$p_X(a) = \sum_y P(X = a, Y = y)$$

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

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

Marginalization

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



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

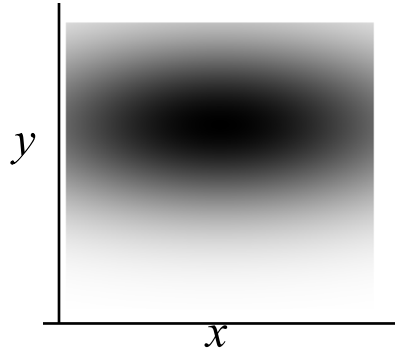
$$P(X = a) = \sum_{y} P(X = a, Y = y)$$

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

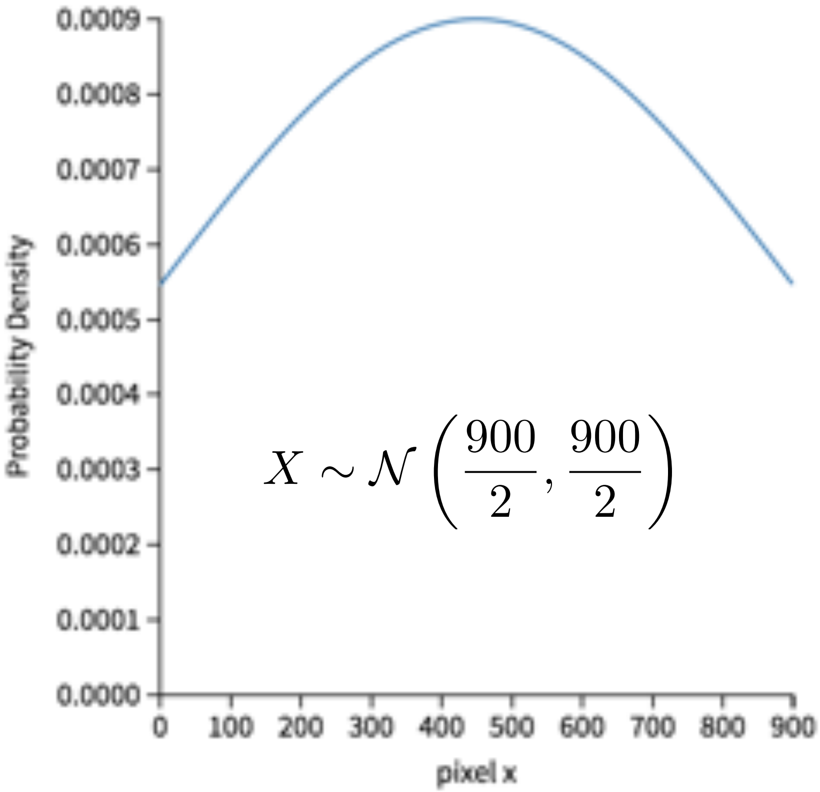
$$f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x, b)\ dx$$

Darts!

Dart PDF

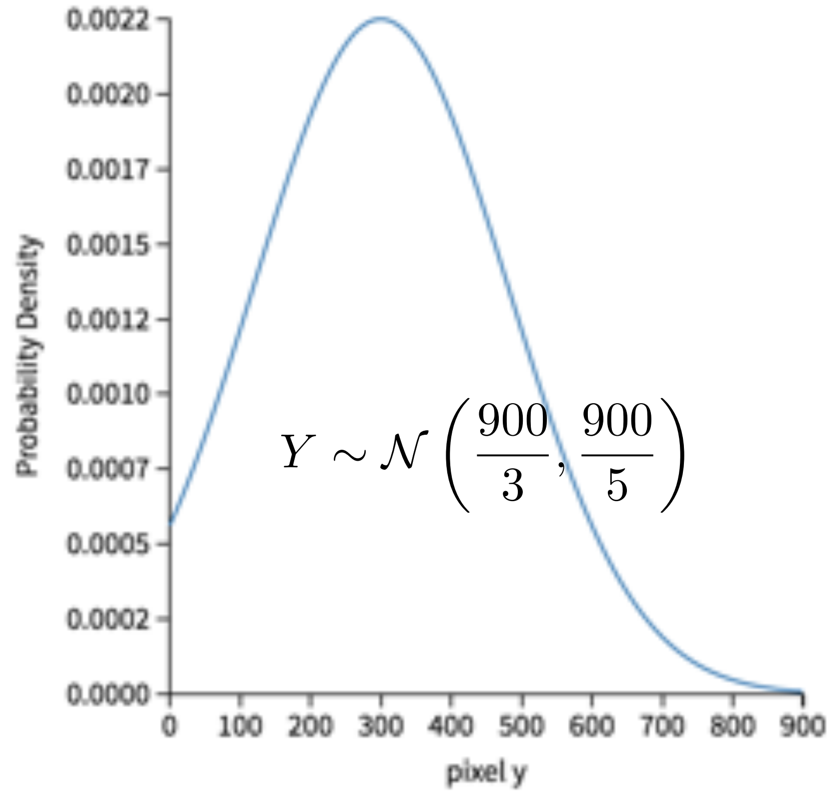


X-Pixel Marginal



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

Y-Pixel Marginal



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

Joint Cumulative Density Function

Cumulative Density Function (CDF):

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

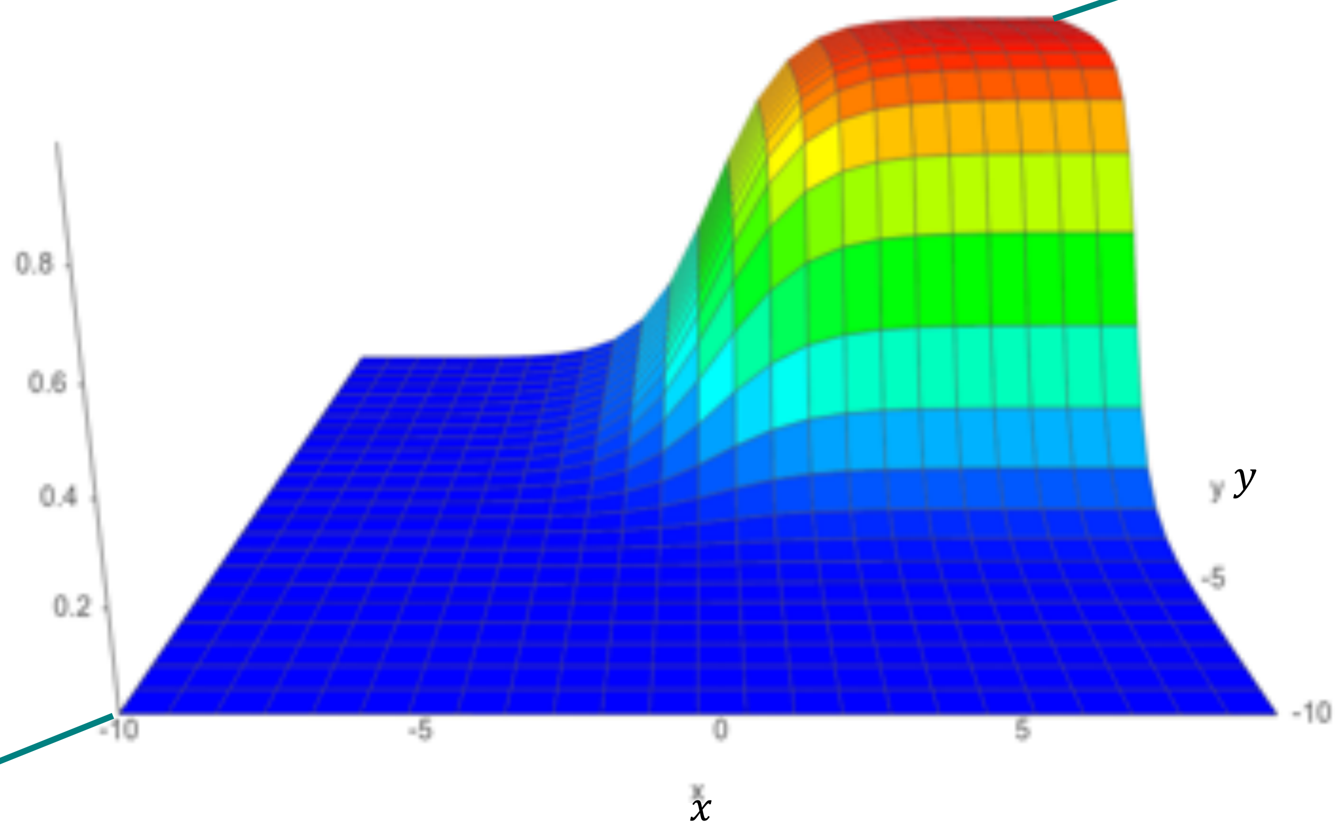
$$F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx$$

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

Joint CDF

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

to 1 as
x → +∞,
y → +∞

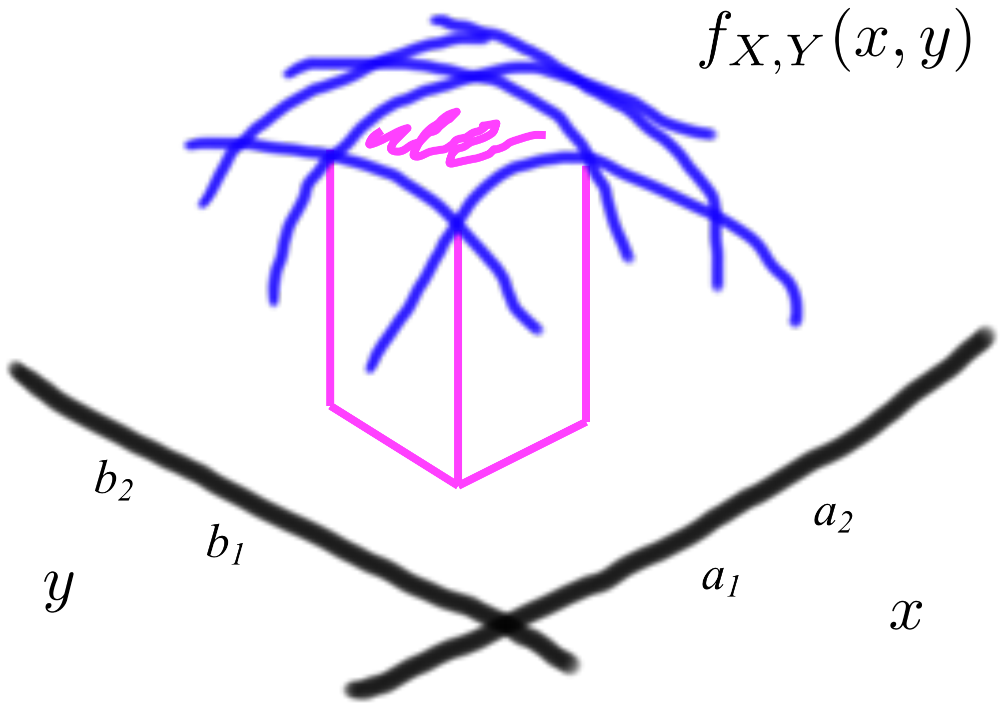


to 0 as
x → -∞,
y → -∞

plot by Academo

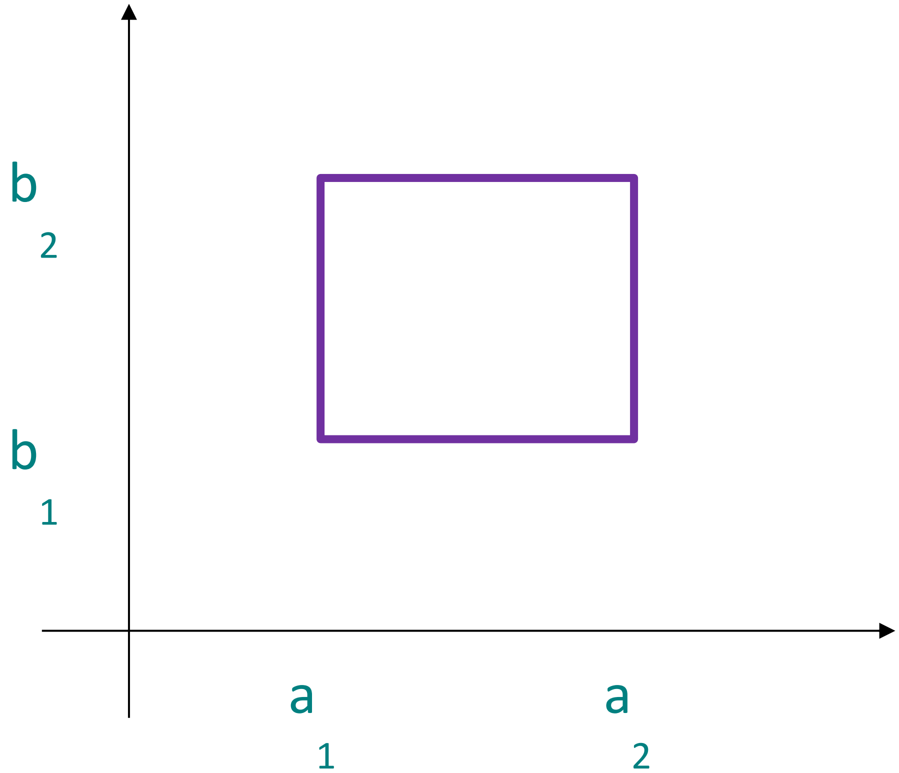
Jointly Continuous

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



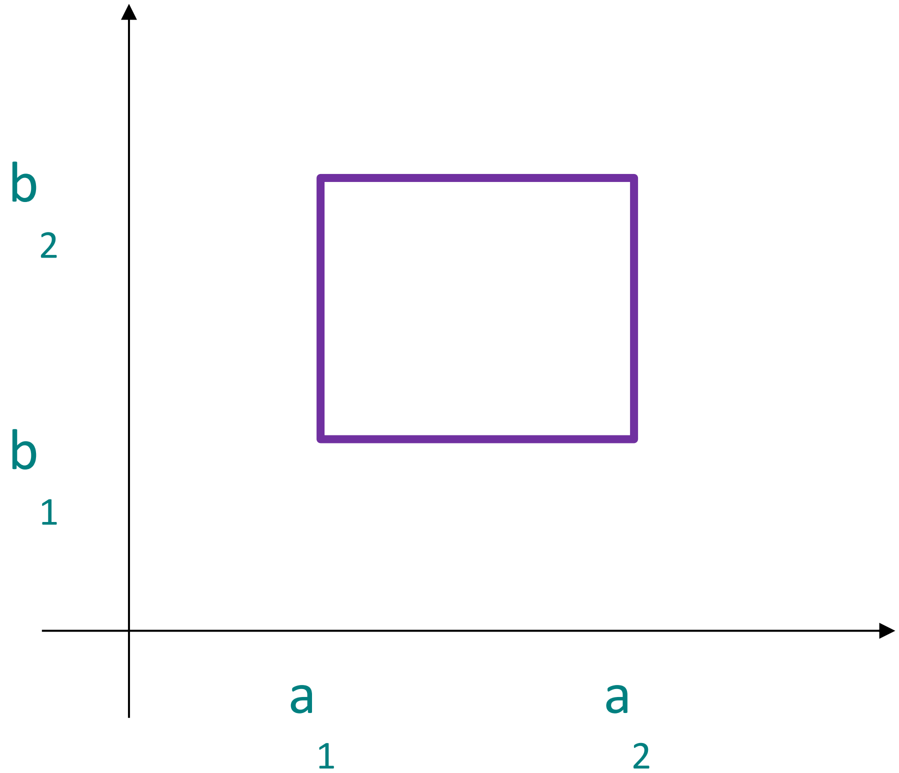
Probabilities from Joint CDF

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



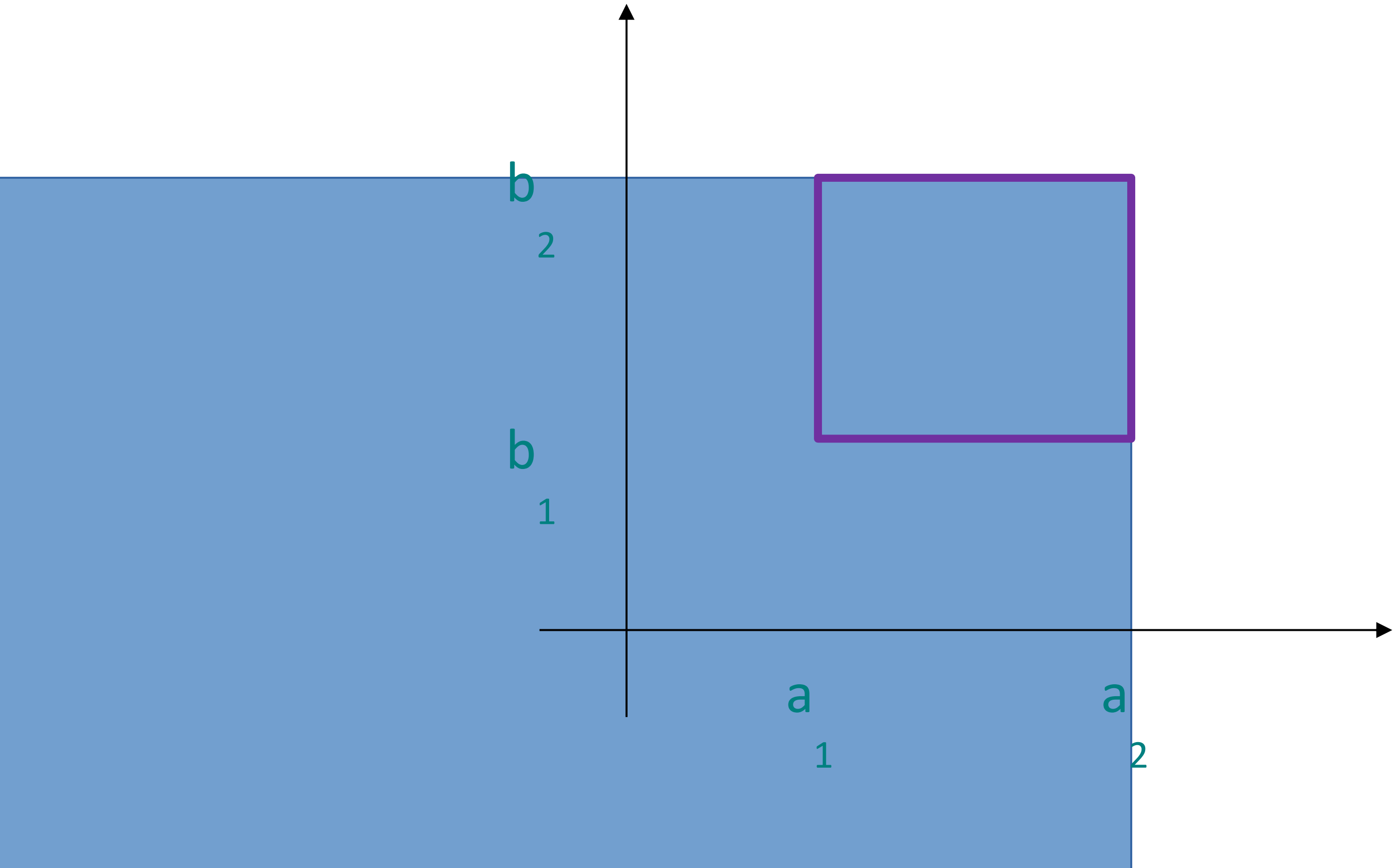
Probabilities from Joint CDF

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



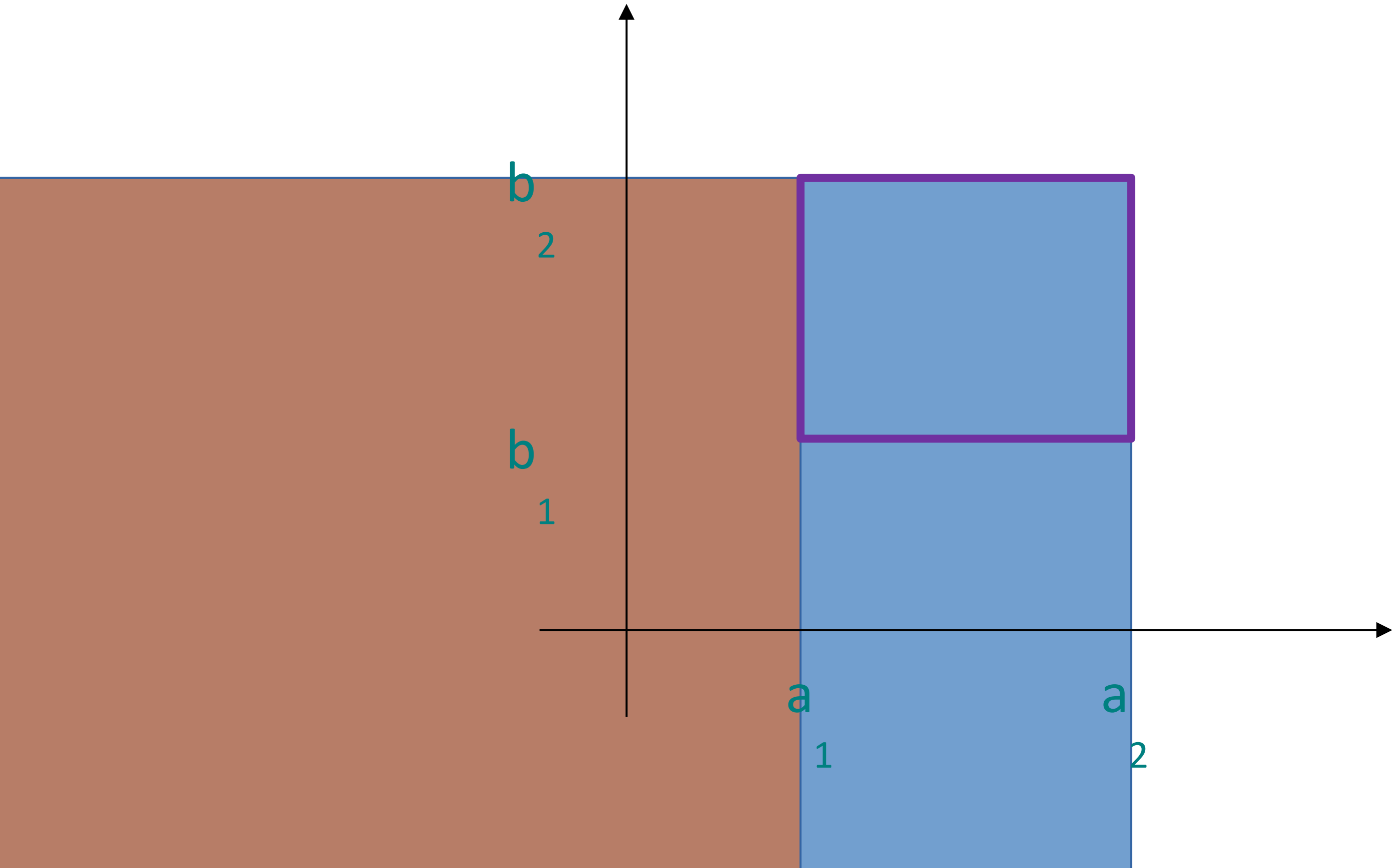
Probabilities from Joint CDF

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



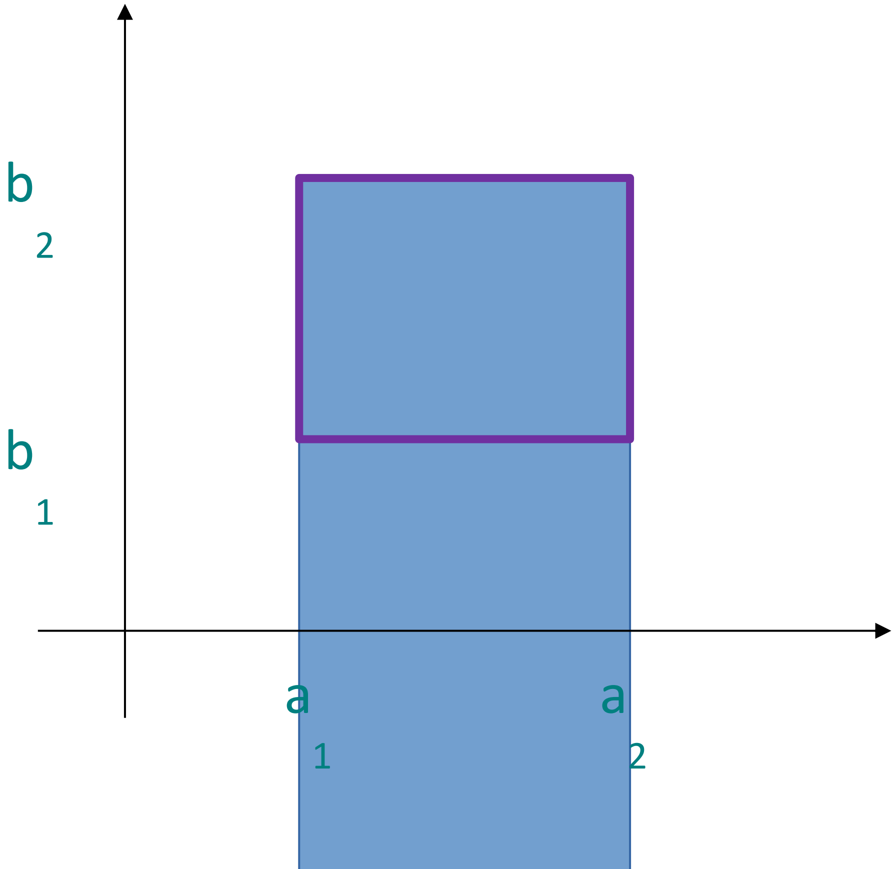
Probabilities from Joint CDF

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



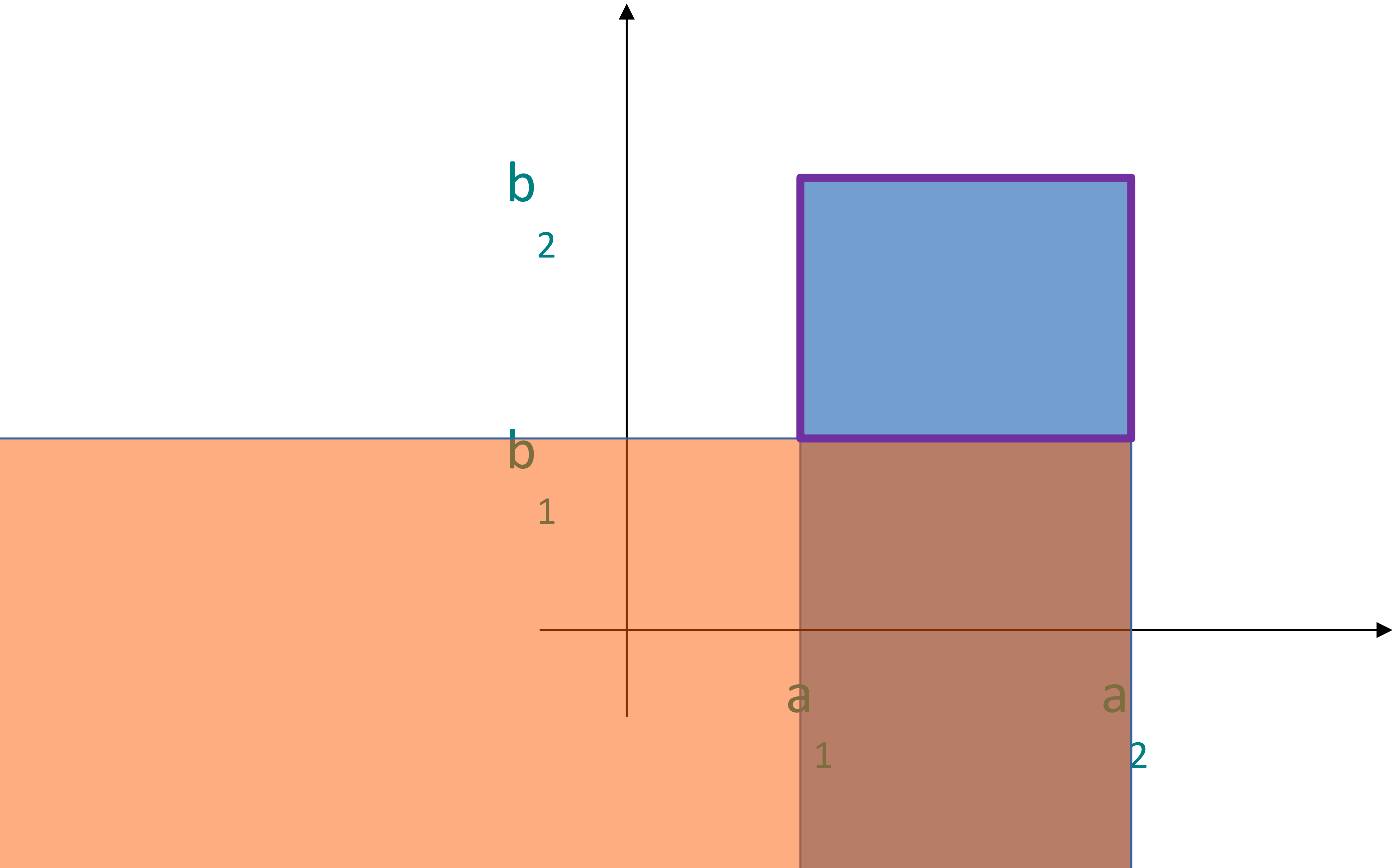
Probabilities from Joint CDF

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



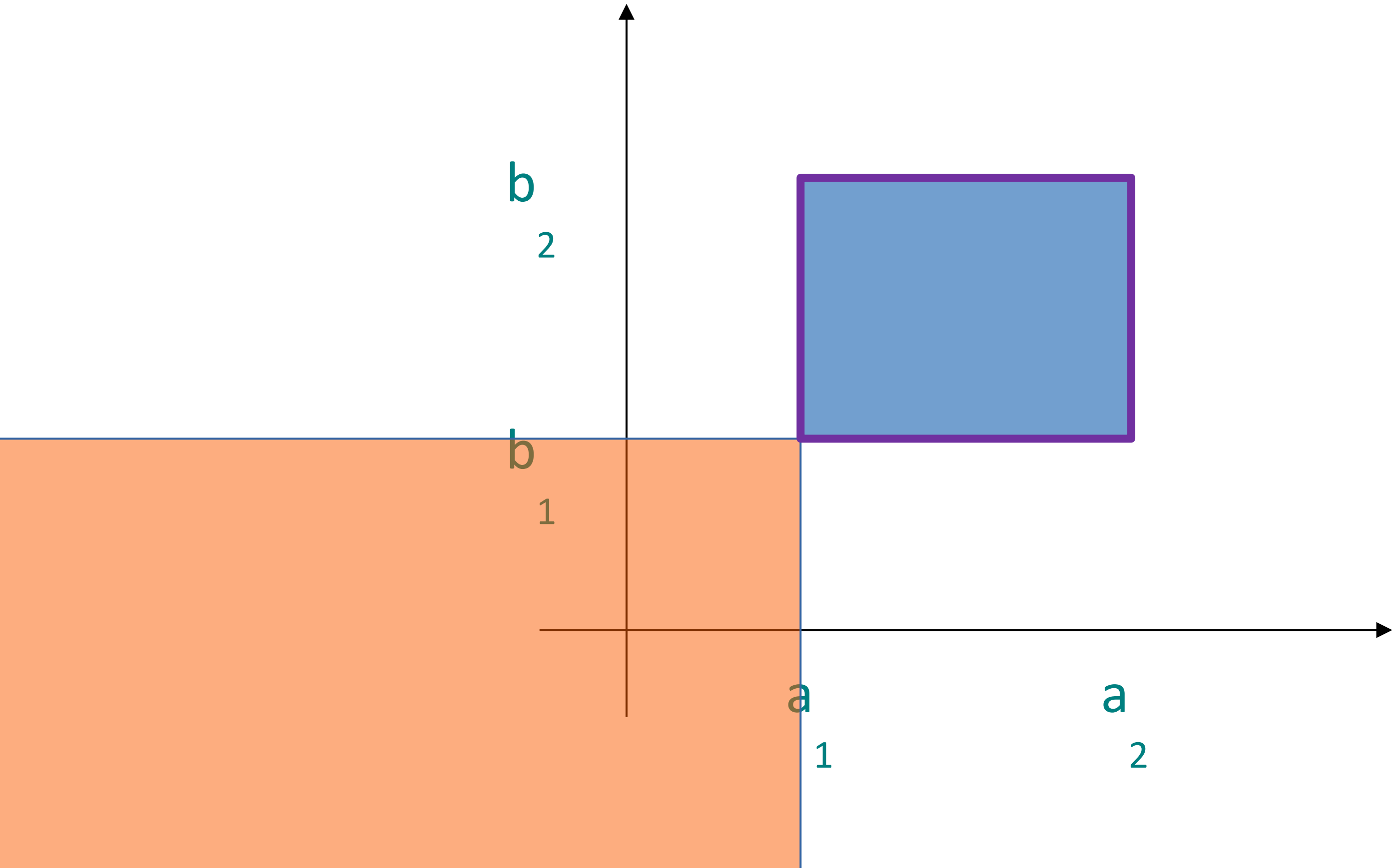
Probabilities from Joint CDF

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



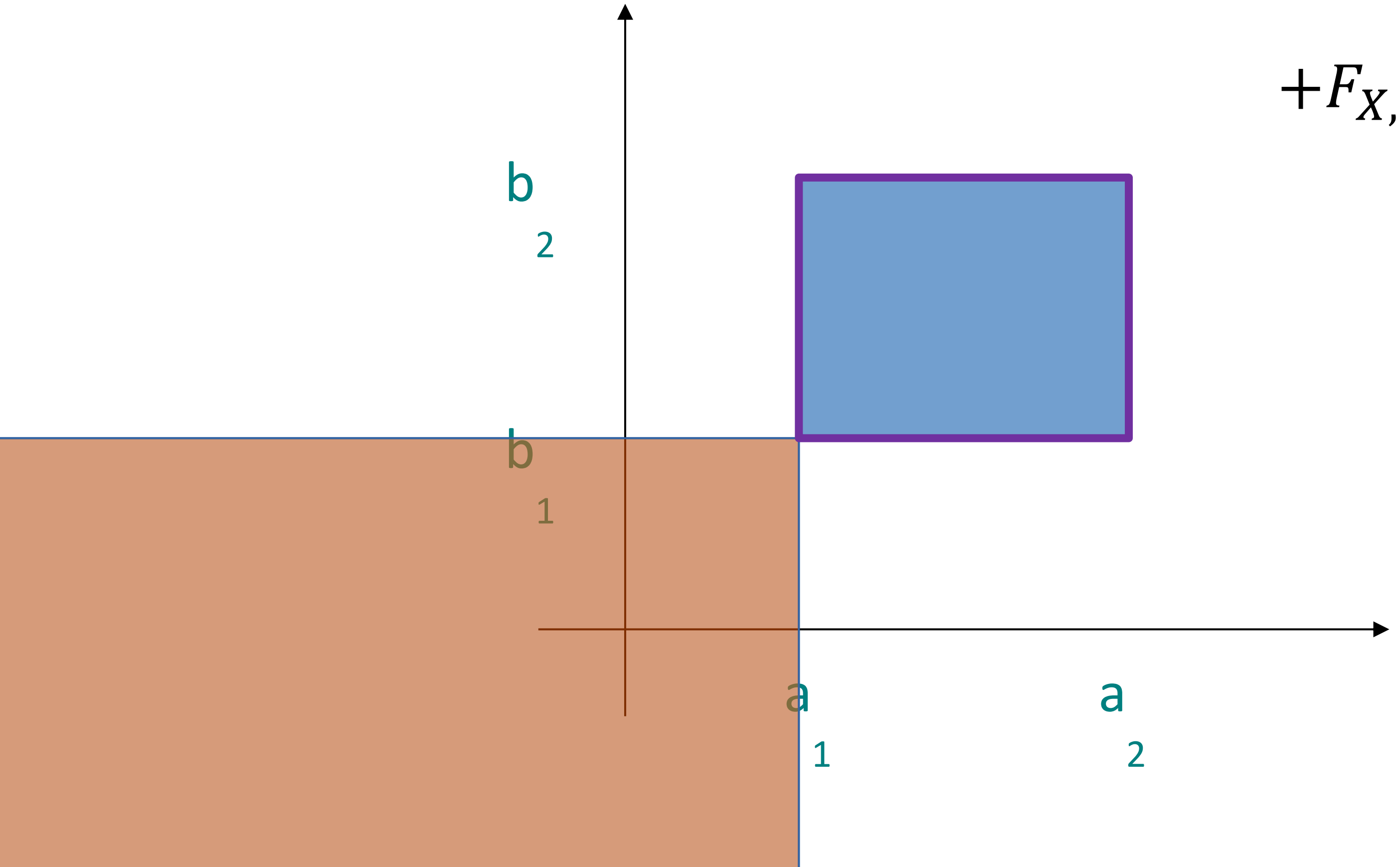
Probabilities from Joint CDF

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



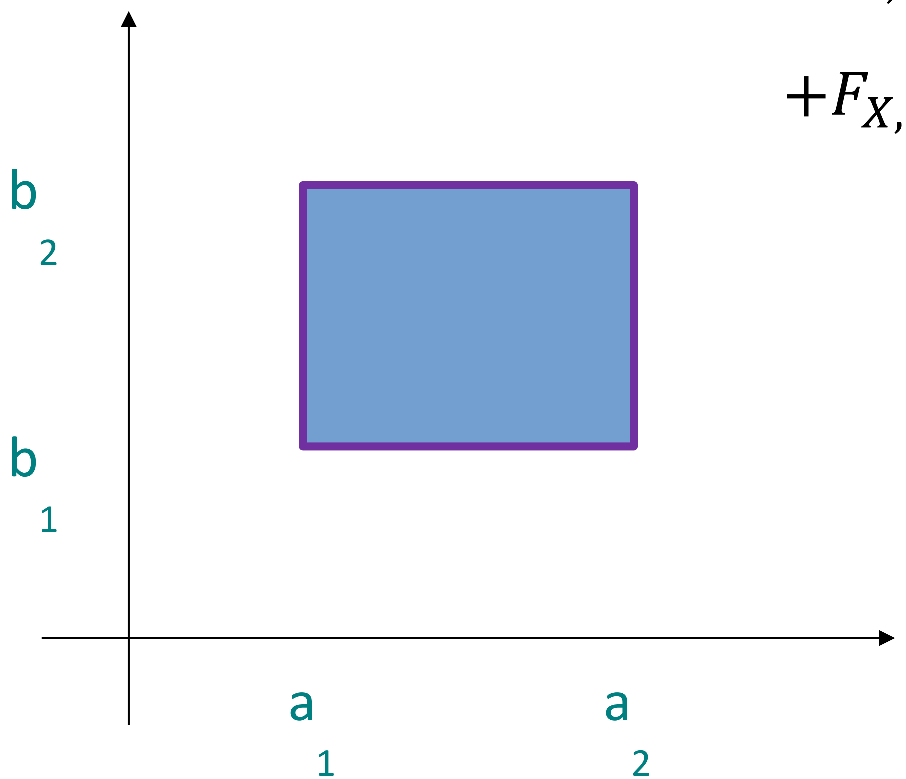
Probabilities from Joint CDF

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



Probabilities from Joint CDF

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



Probability for Instagram!

Gaussian Blur

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

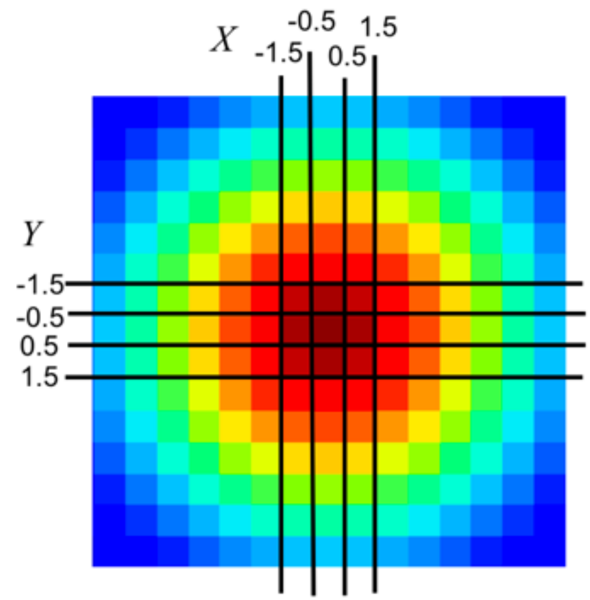
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

Joint PDF

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

Joint CDF

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



Used to generate this weight matrix

Gaussian Blur

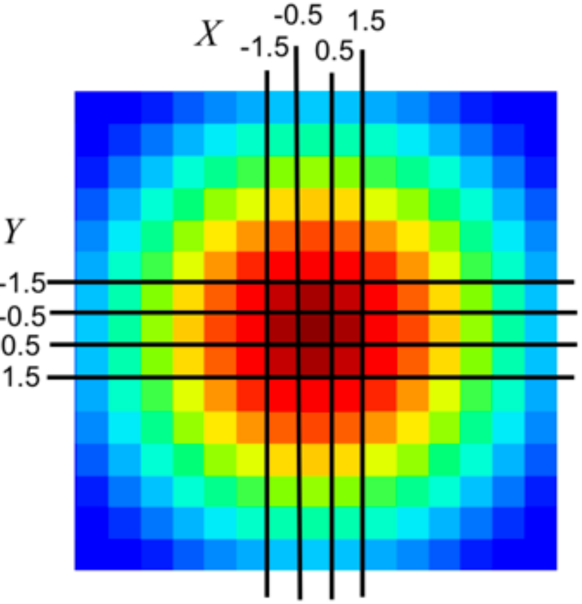
Joint PDF

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

Joint CDF

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

Weight Matrix



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

$-0.5 \le x \le 0.5$ and $-0.5 \le y \le 0.5$

What is the weight of the center pixel?

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

How do you integrate under a circle?

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