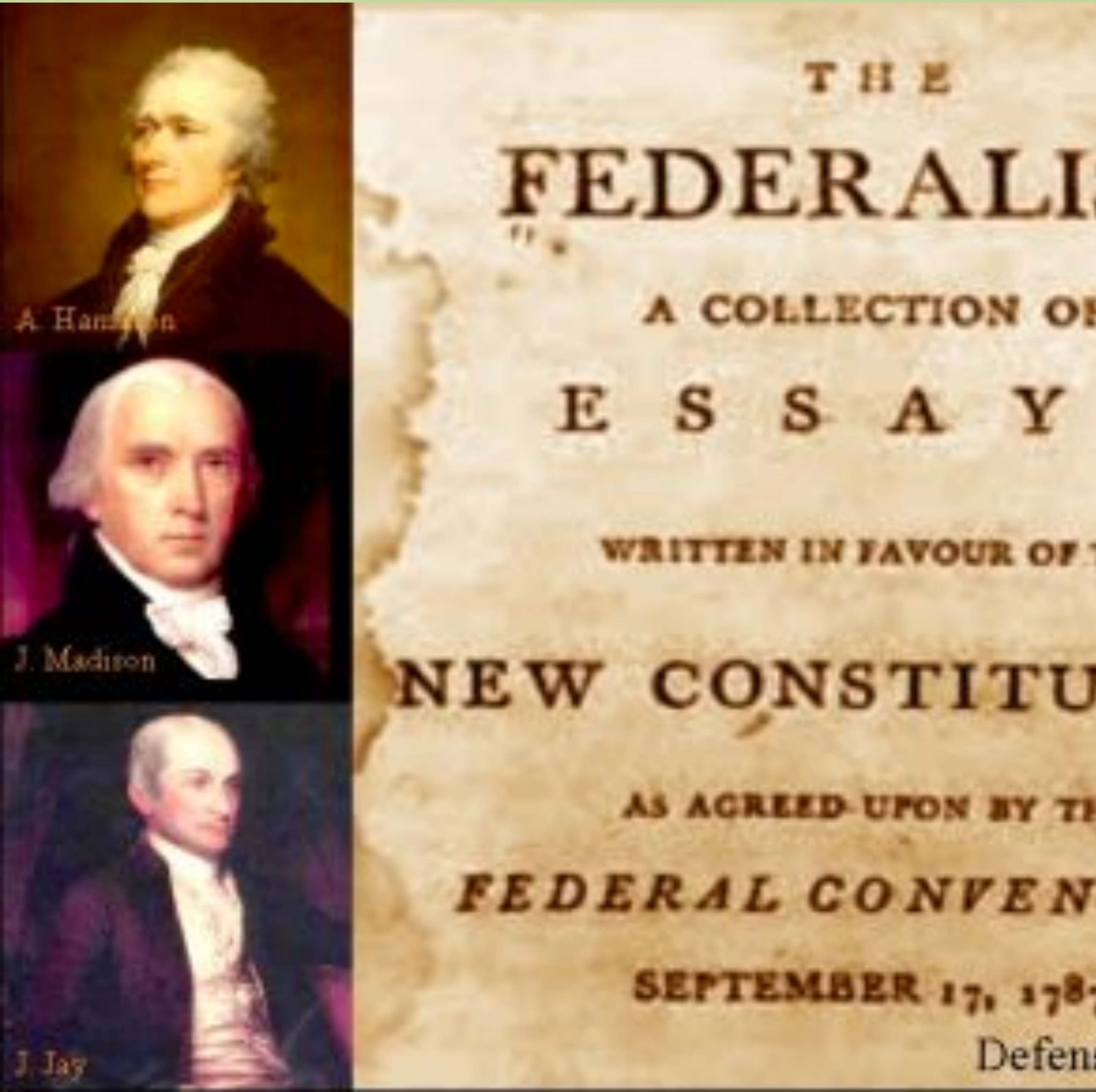
Continuous Joint Distributions

Chris Piech
CS109, Stanford University

Learning Goals

1. Know how to use a multinomial
2. Be able to calculate large bayes problems using a computer
3. Use a Joint CDF

Motivating Examples

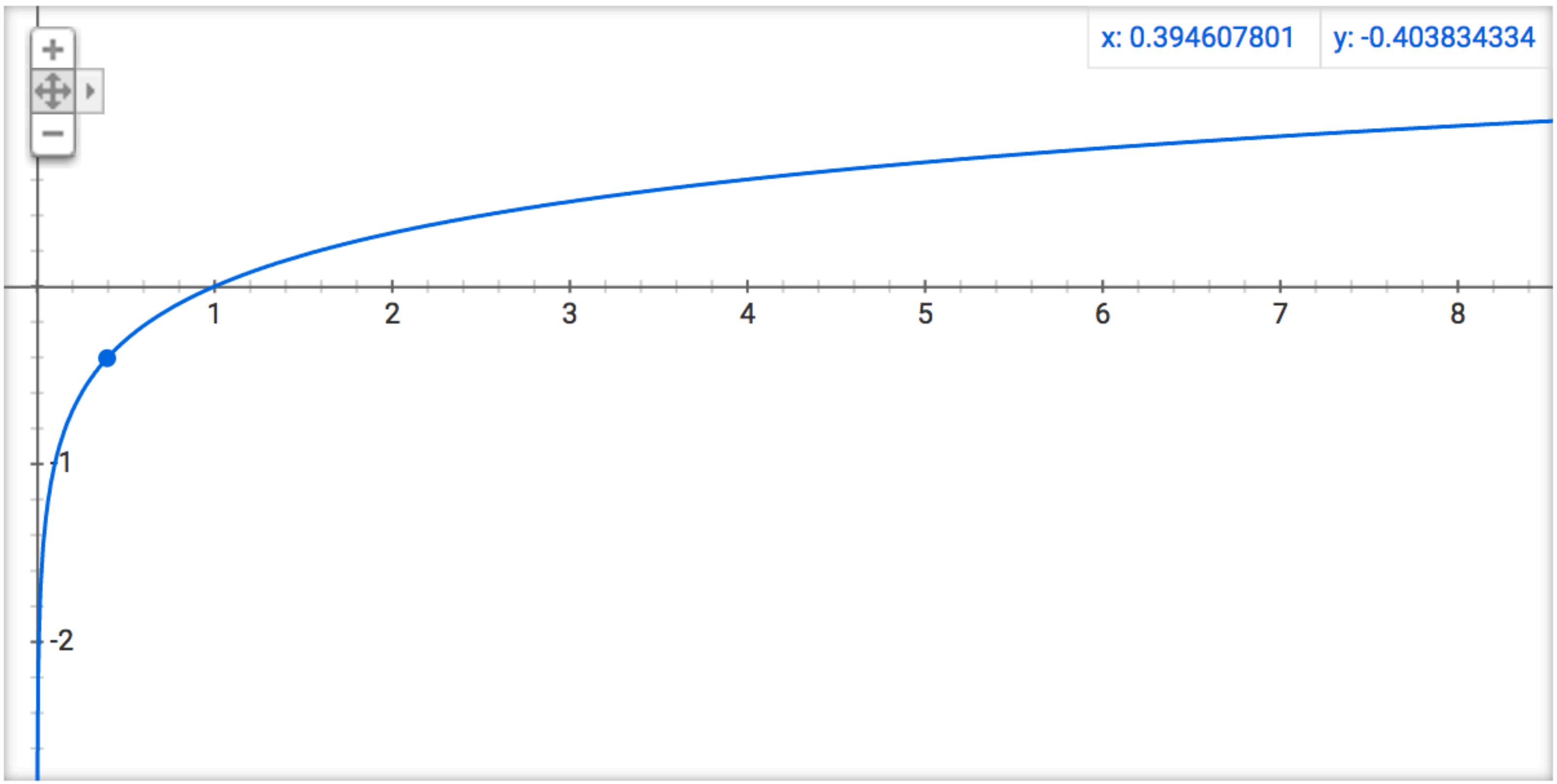


Recall logs

Log Review

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

Graph for log(x)



More info

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!

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

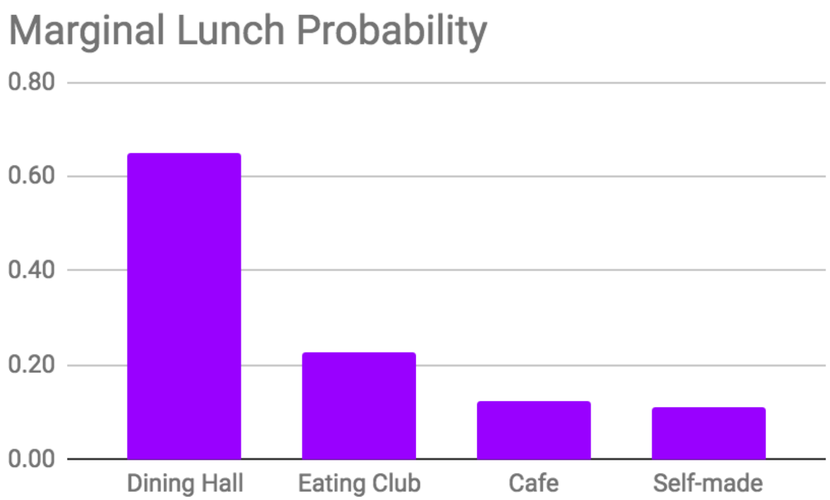
$$\log(\prod_i a_i) = \sum_i \log(a_i)$$

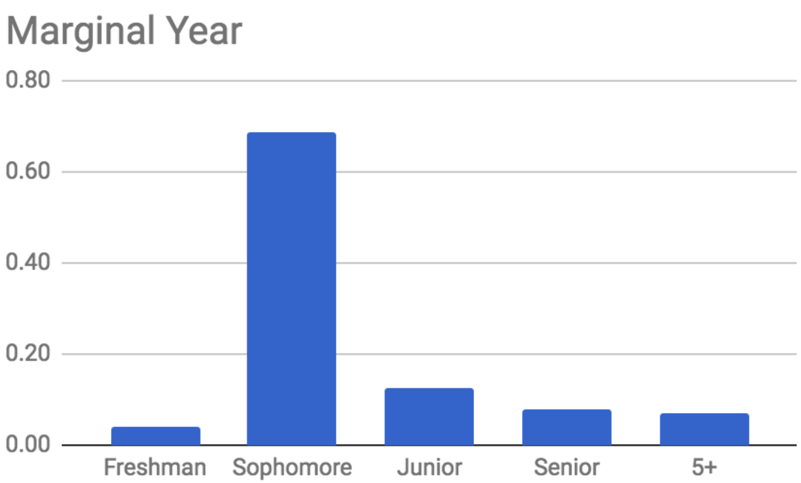
* Spoiler alert: This is important because the product of many small numbers gets hard for computers to represent.

Where we left off

Joint Probability Table

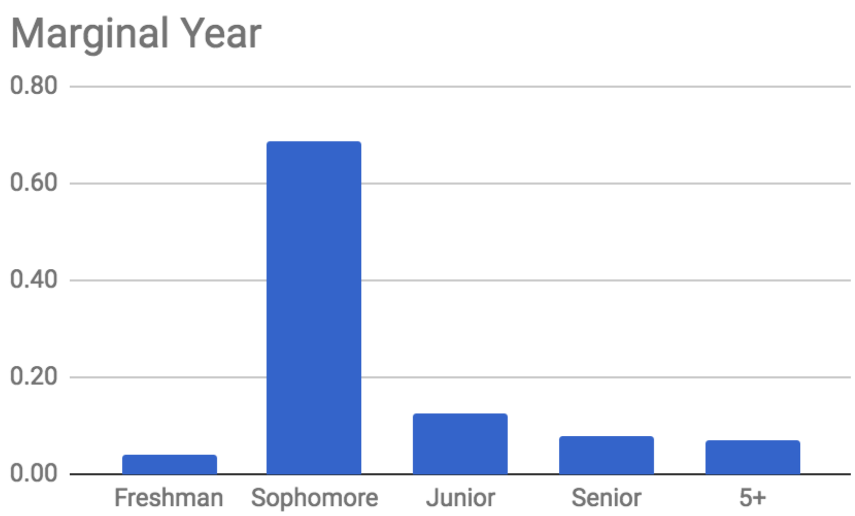
Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.02	0.00	0.02	0.00	0.04
Sophomore	0.51	0.15	0.03	0.03	0.69
Junior	0.08	0.02	0.02	0.02	0.13
Senior	0.02	0.05	0.01	0.01	0.08
5+	0.02	0.01	0.05	0.05	0.07
Marginal Status	0.65	0.23	0.13	0.11	



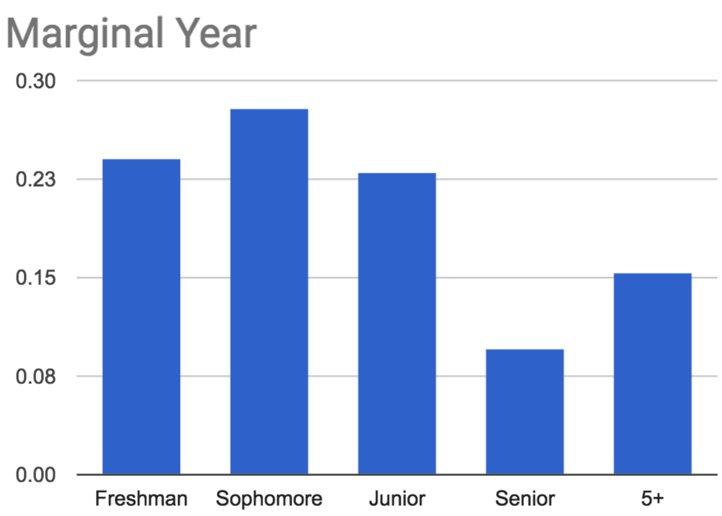


Change in Marginal!

Fall 2017



Spring 2017



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

$$P\left(\begin{array}{l}\text{Viagra} = 2\\ \text{Free} = 2\\ \text{Risk} = 1\\ \text{Credit-card: } 2\\ \ldots\\ \text{For} = 2\end{array} \middle| \text{spam}\right) = \frac{n!}{2!2!\ldots2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \ldots p_{\text{for}}^2$$

It's a Multinomial!

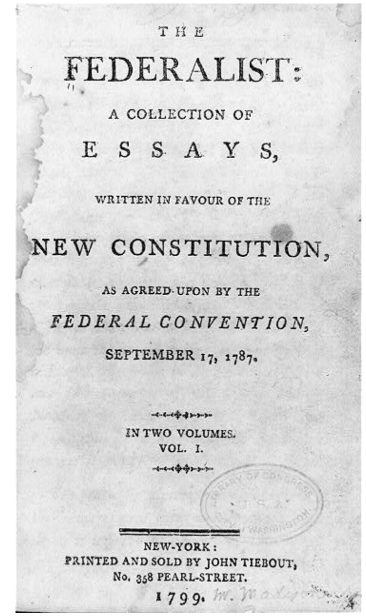
Probability of seeing this document | spam

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!

Joint Expectation

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

- Expectation over a joint isn't nicely defined because it is not clear how to compose the multiple variables:
  - Add them? Multiply them?
- Lemma: For a function g(X,Y) we can calculate the expectation of that function:

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

- Recall, this also holds for single random variables:

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

Expected Values of Sums

Big deal lemma: first stated without proof

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

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

Holds regardless of dependency between X_i's

Skeptical Chris Wants a Proof!

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

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

What a useful lemma

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

By the definition of g(x,y)

Break that sum into parts!

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

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

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

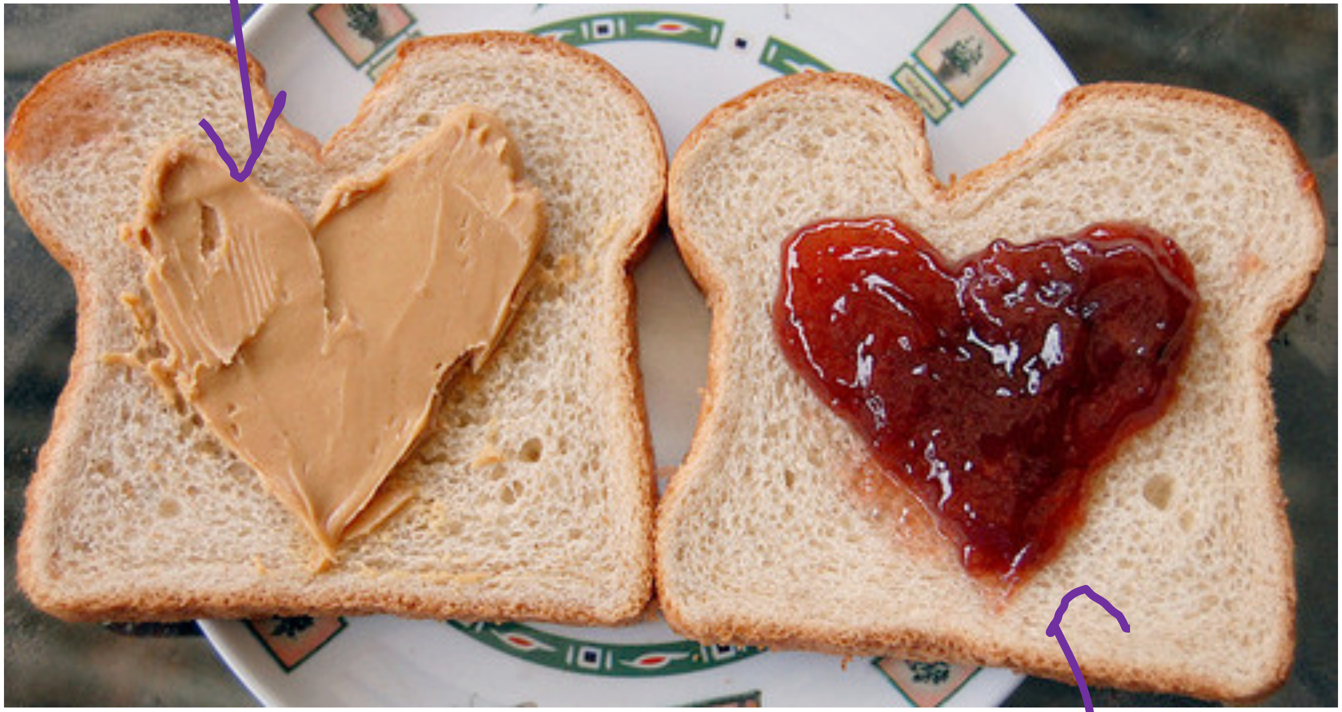
That is the definition of marginal probability

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

That is the definition of expectation

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

Continuous Random Variables



Joint Distributions

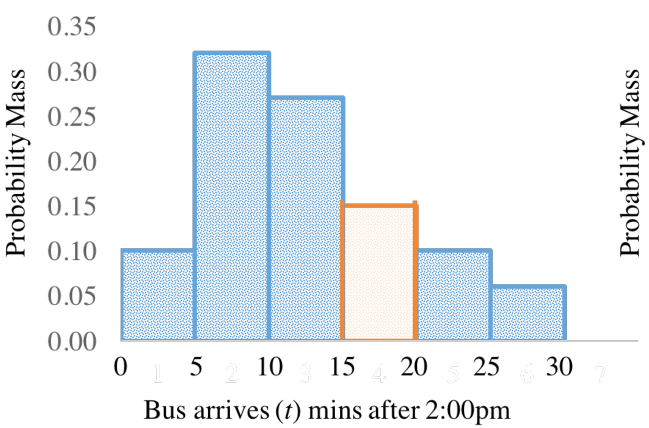
Continuous Joint Distribution

Riding the Marguerite

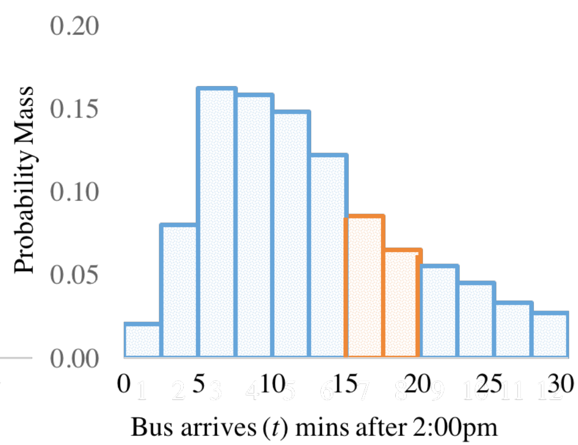
You are running to the bus stop. You don't know exactly when the bus arrives. You arrive at 2:20pm.

What is P(wait < 5 min)?

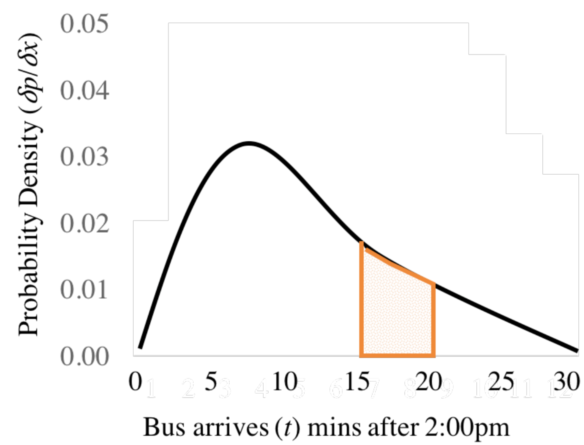
Discretize into 5 min chunks



Discretize into 2.5 min chunks



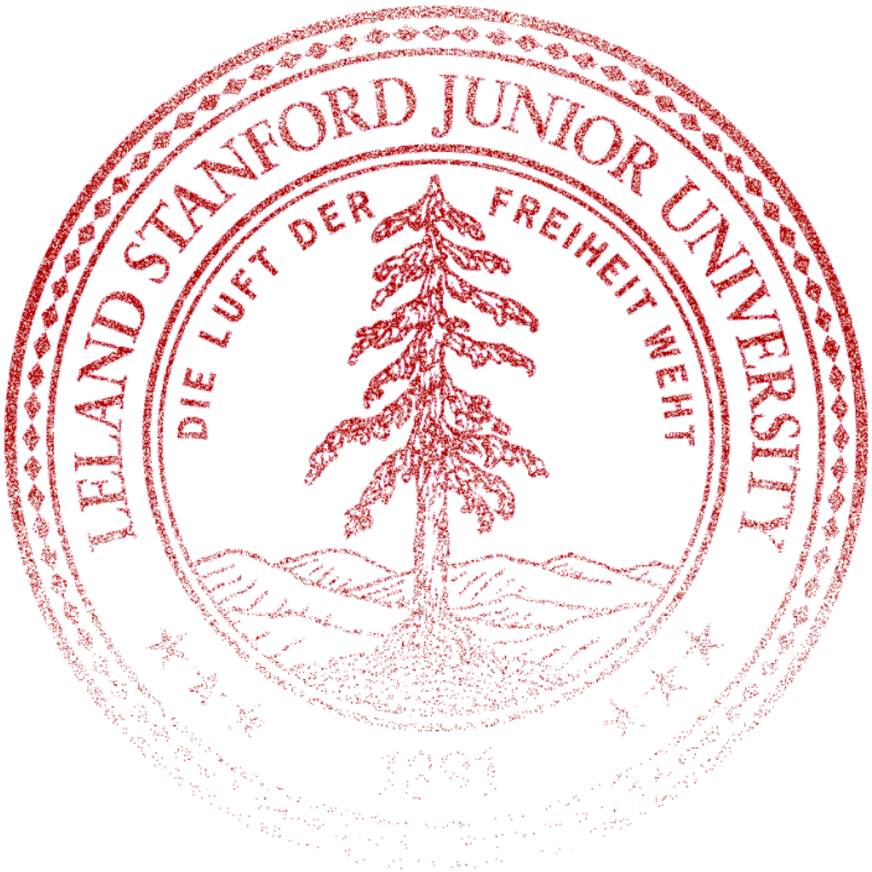
The limit at discretization size → 0



Joint Dart Distribution

Dart Results

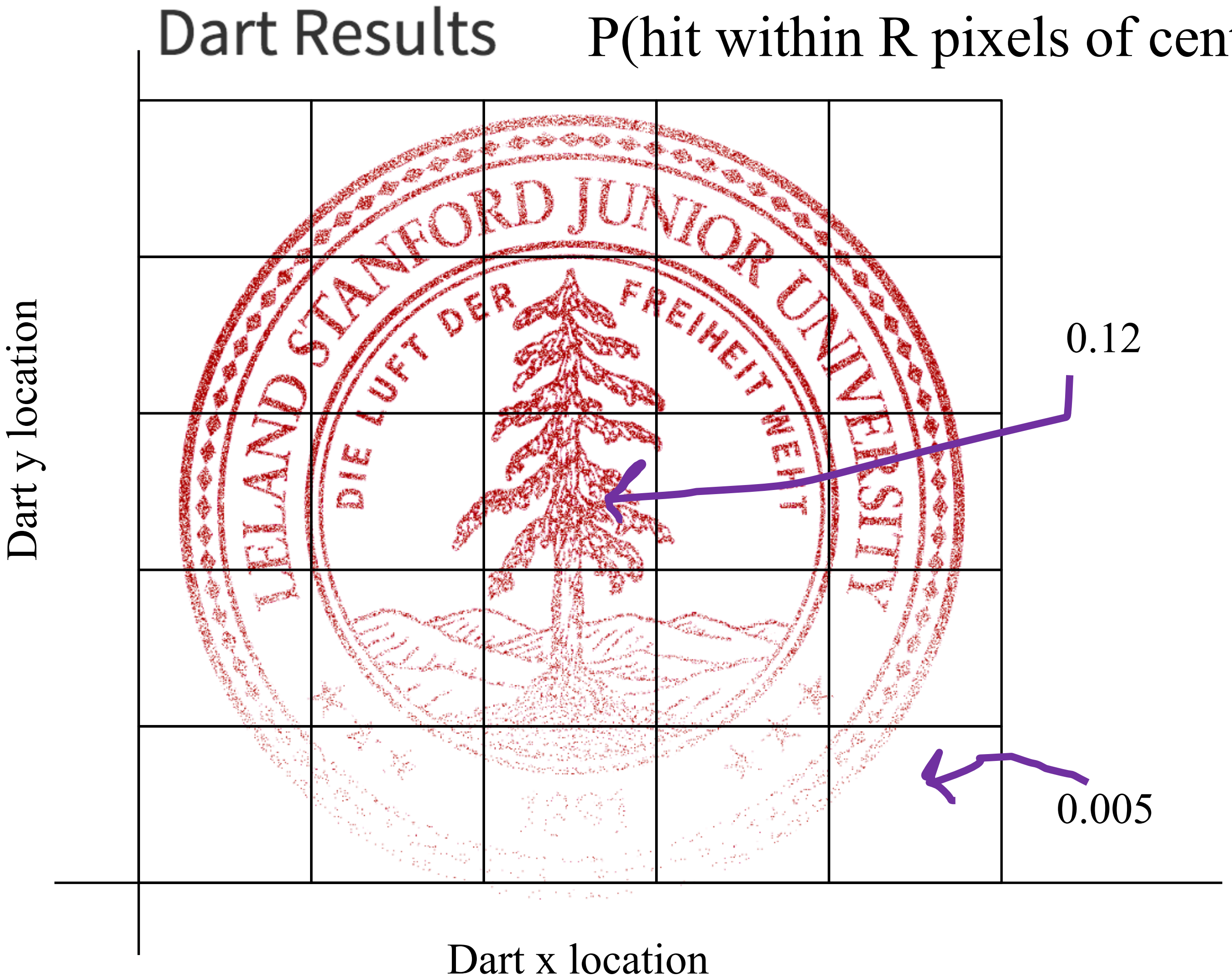
P(hit within R pixels of center)?



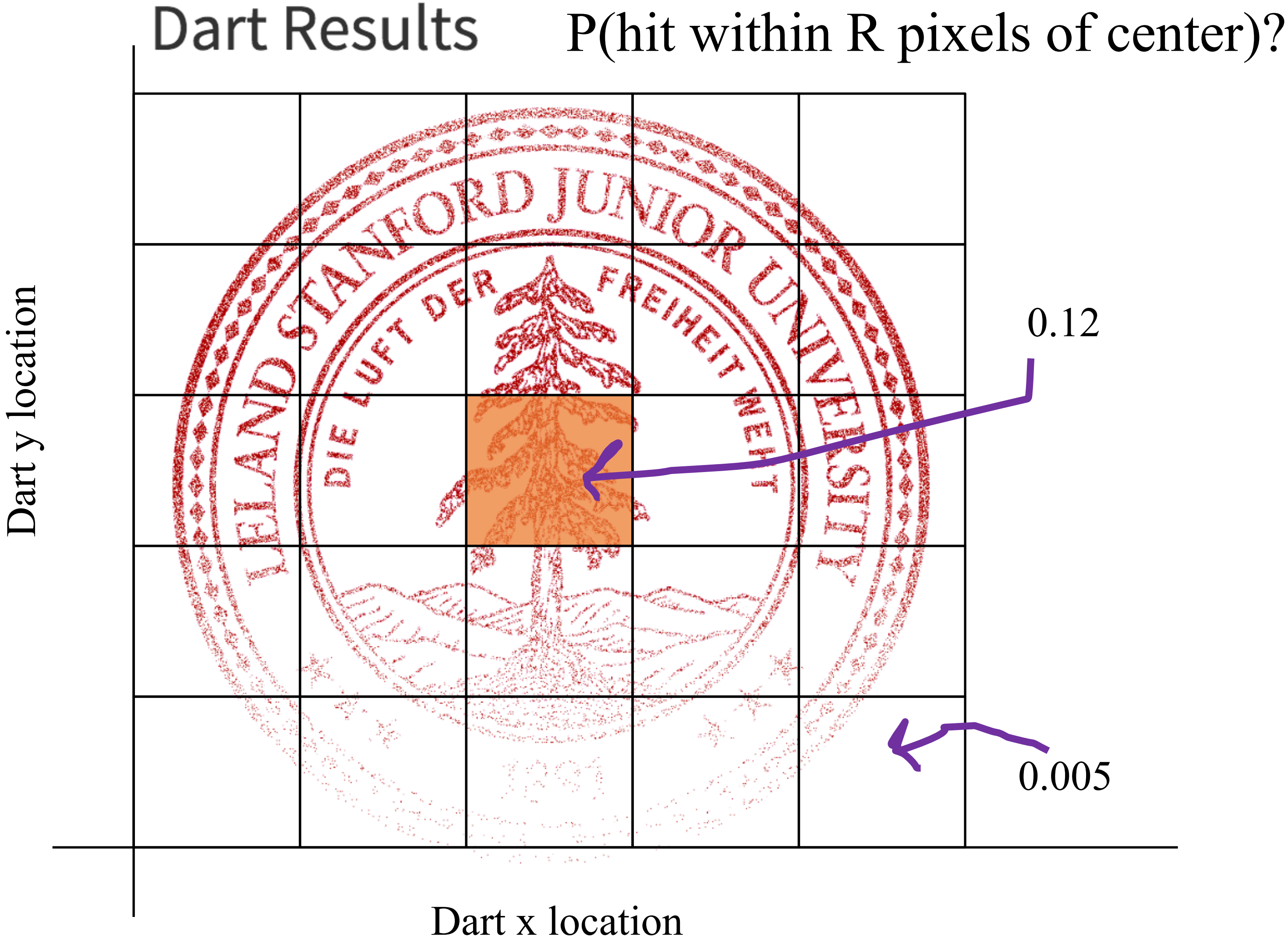
What is the probability that a dart hits at (456.234231234122355, 532.12344123456)?

Joint Dart Distribution

P(hit within R pixels of center)?

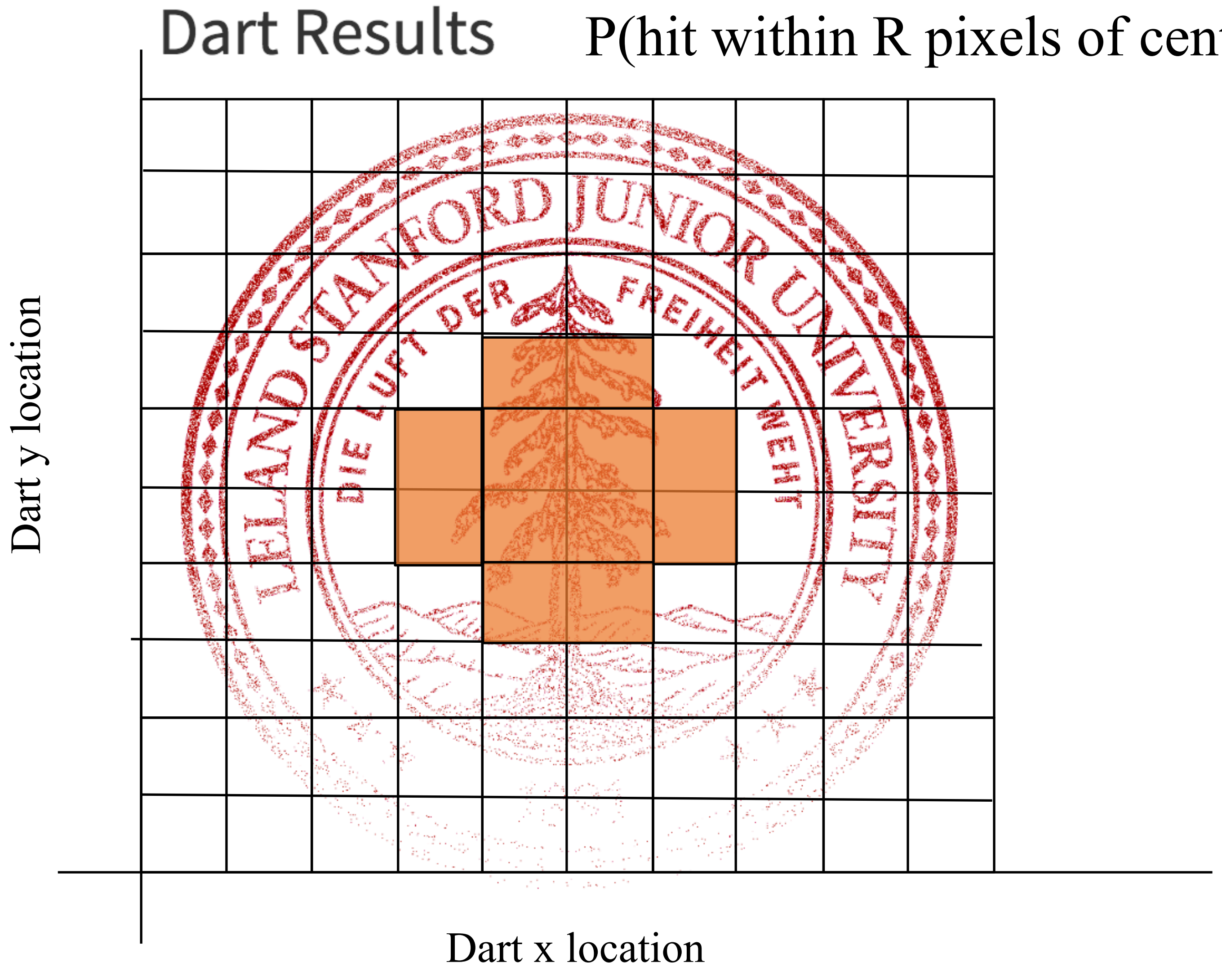


Joint Dart Distribution

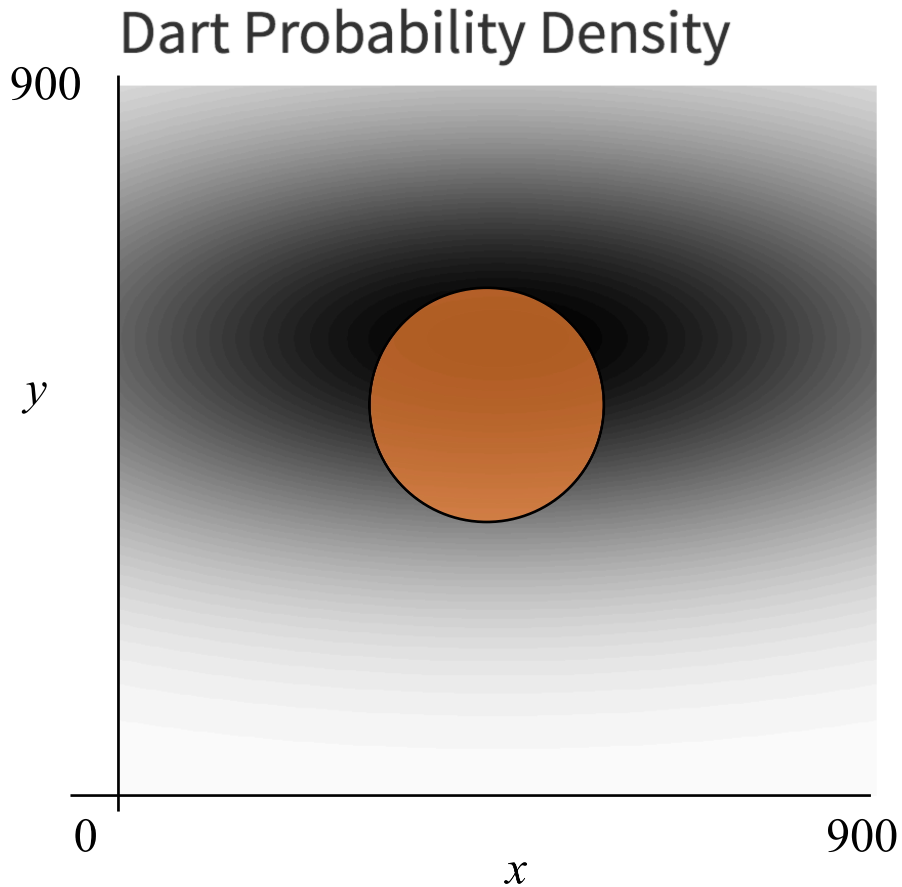


Joint Dart Distribution

P(hit within R pixels of center)?



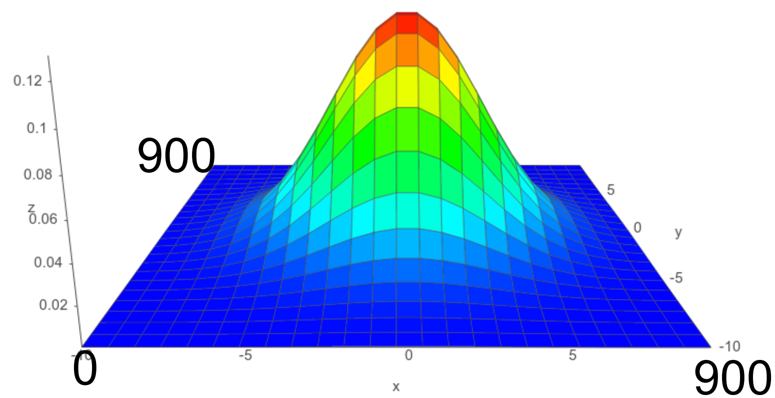
Joint Dart Distribution

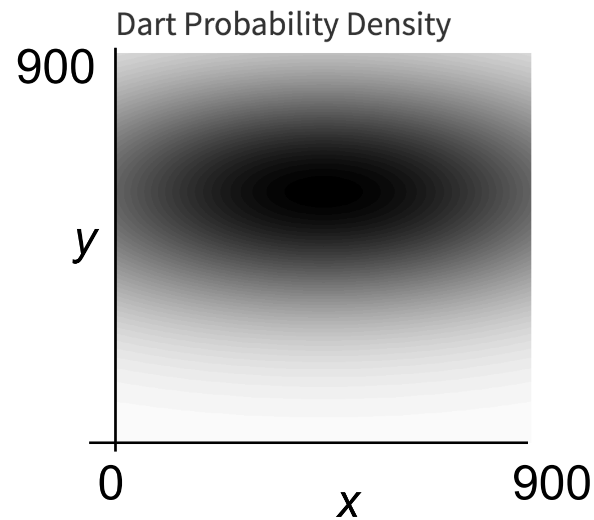


In the limit, as you break down continuous values into intestinally small buckets, you end up with multidimensional probability density

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.

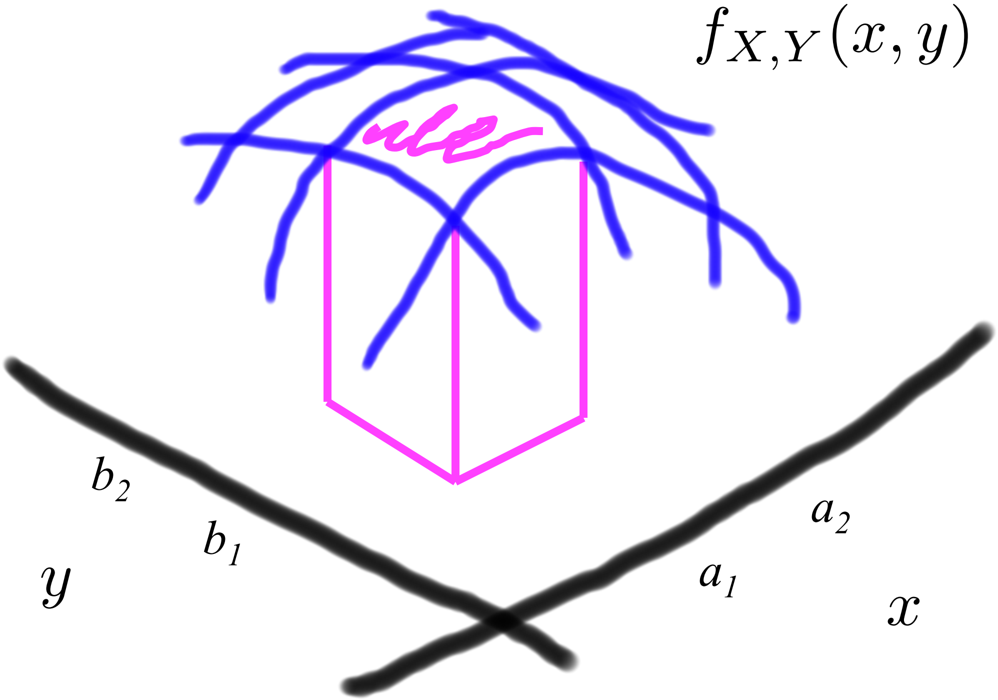




$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y)\, dy\, dx$$

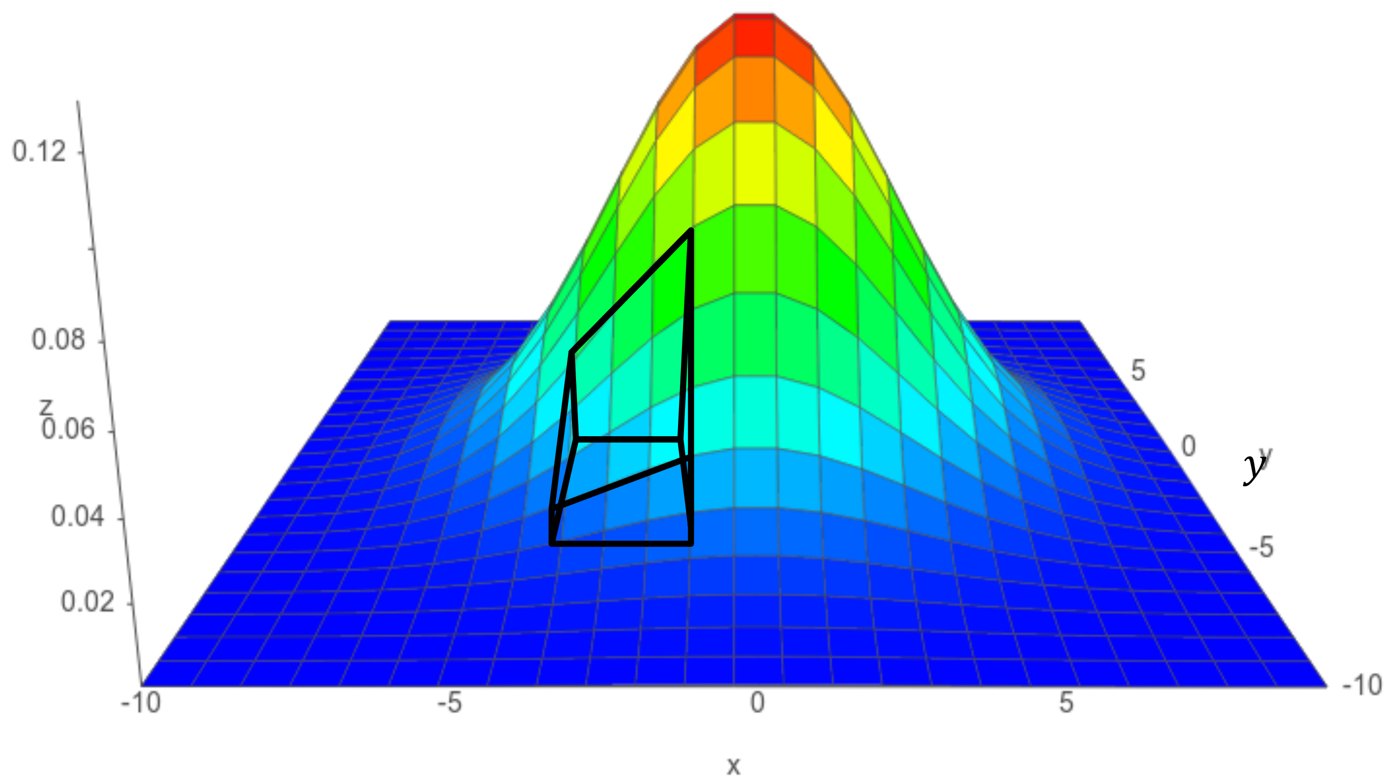
Joint Probability Density Funciton

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



Joint Probability Density Funciton

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



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):

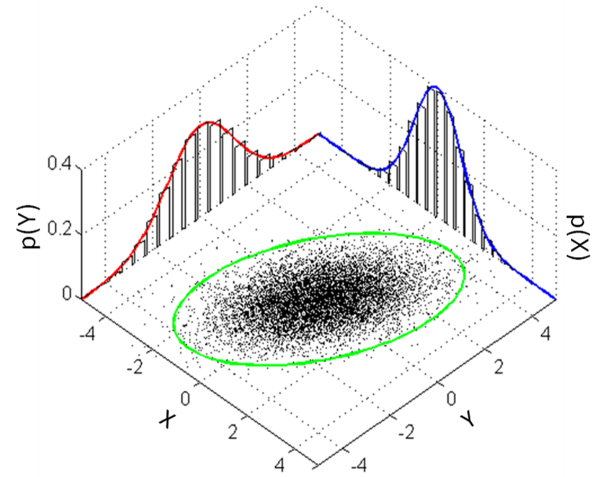
$$\int_{y=0}^{2}\int_{x=0}^{1} xy\,dx\,dy = \int_{y=0}^{2}\left(\int_{x=0}^{1} xy\,dx\right)dy = \int_{y=0}^{2} y\left[\frac{x^2}{2}\right]_0^1 dy = \int_{y=0}^{2} y\,\frac{1}{2}\,dy$$

  - 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.



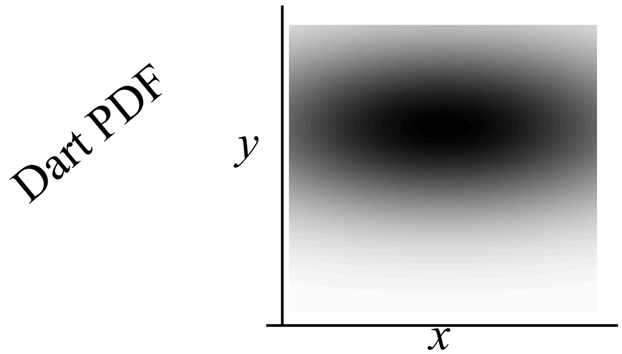
Sum/integrate over the variables you don't care about.

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

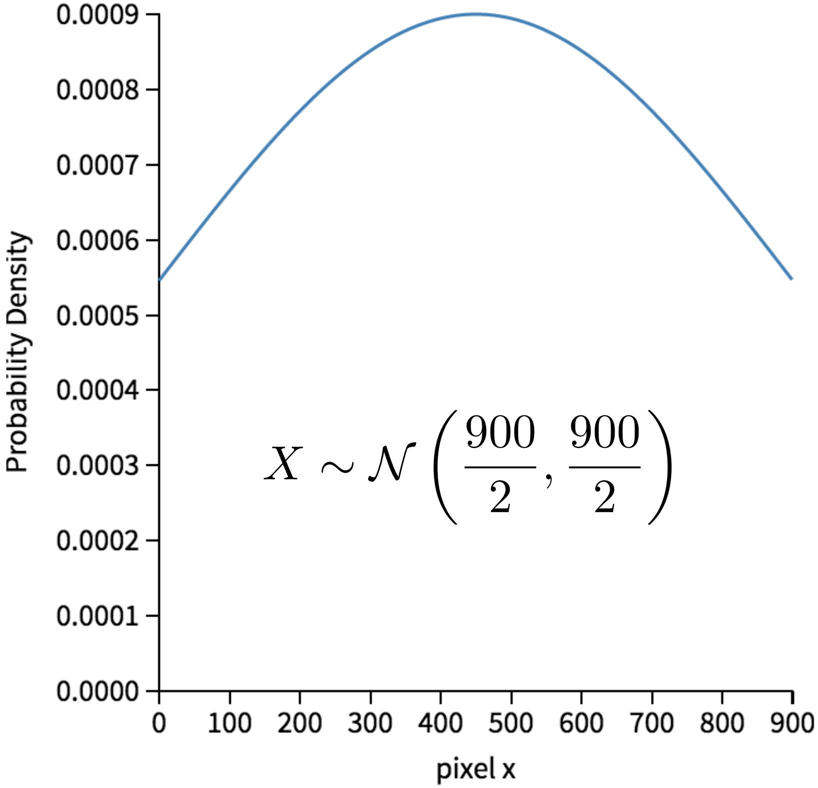
$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y)\ dy$$

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

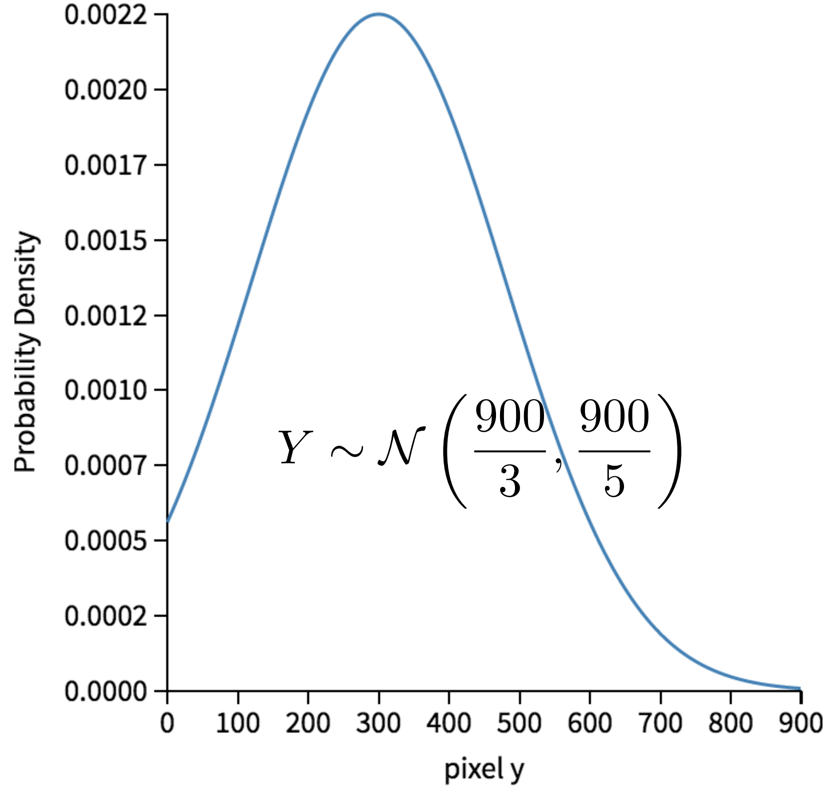
Darts!



X-Pixel Marginal



Y-Pixel Marginal



Jointly Continuous

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

- Cumulative Density Function (CDF):

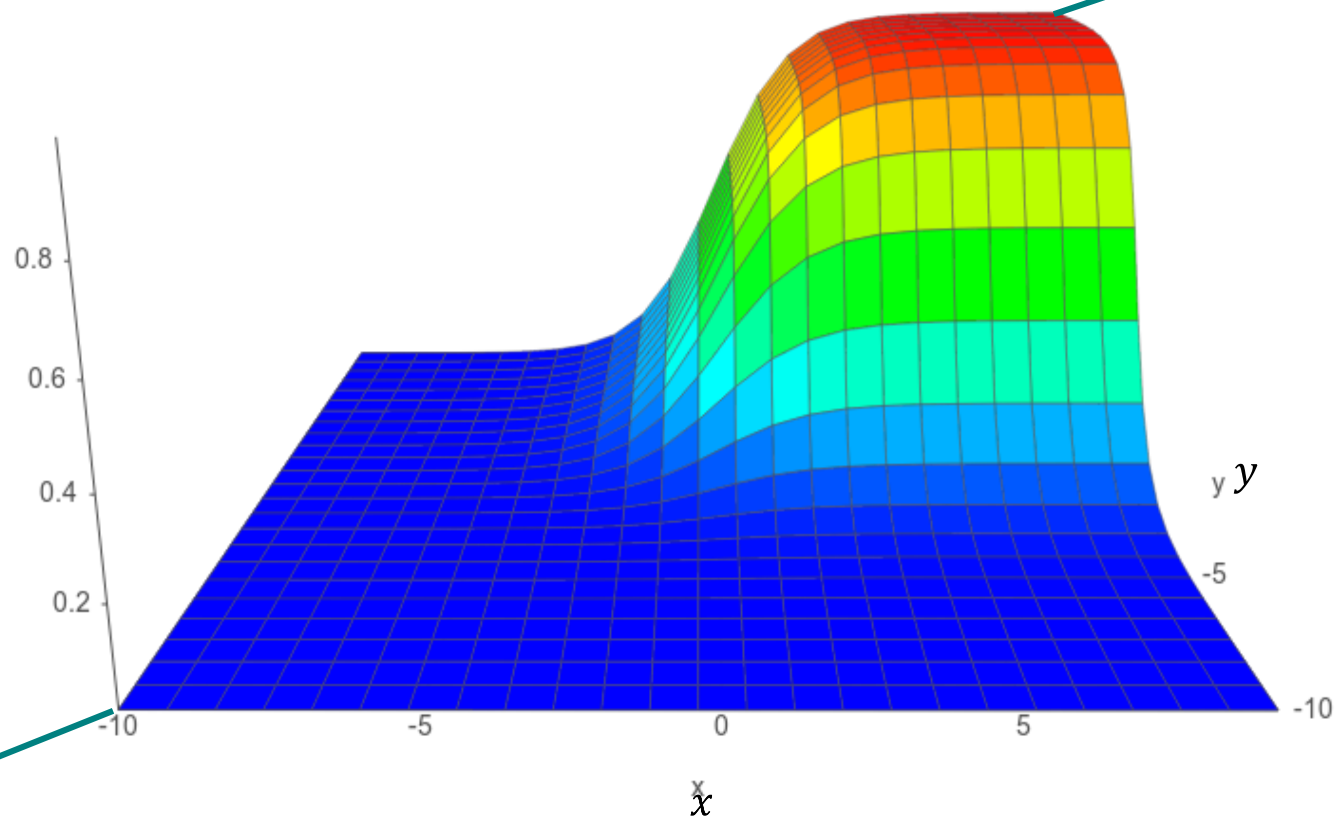
$$F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx$$

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

Jointly CDF

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

to 1 as
x ⟶ +∞,
y ⟶ +∞

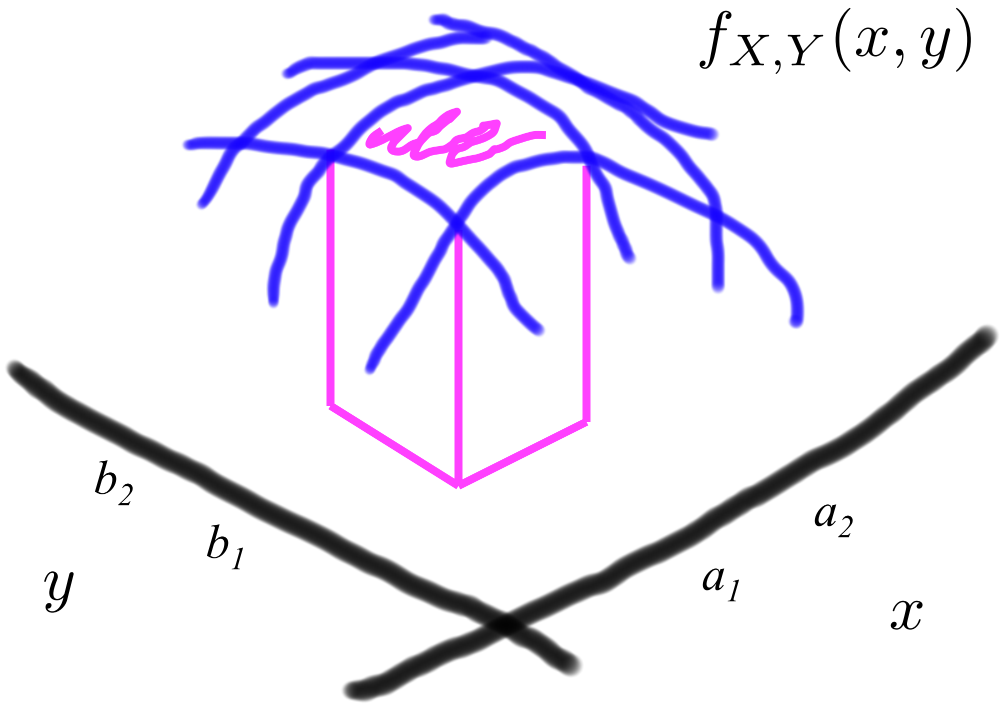


to 0 as
x ⟶ -∞,
y ⟶ -∞

plot by Academo

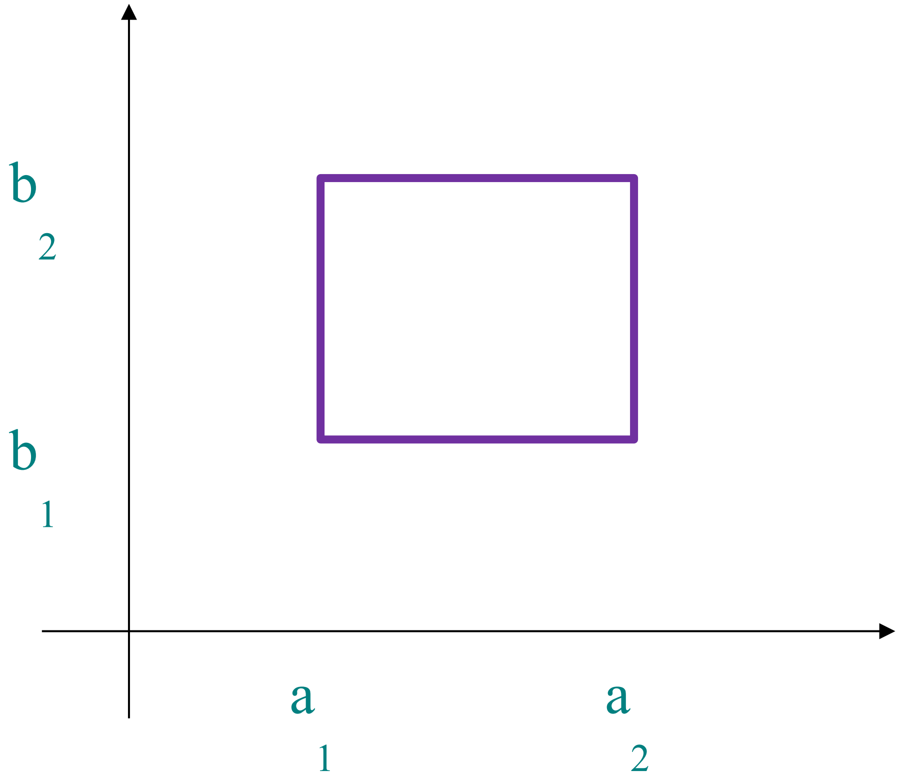
Jointly Continuous

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



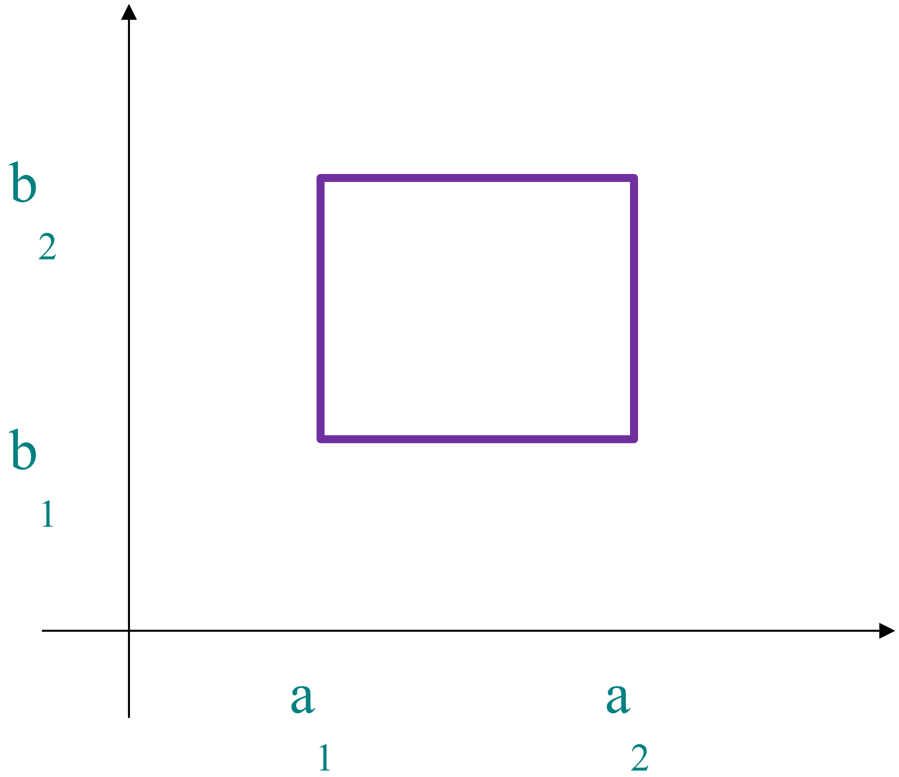
Probabilities from Joint CDF

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



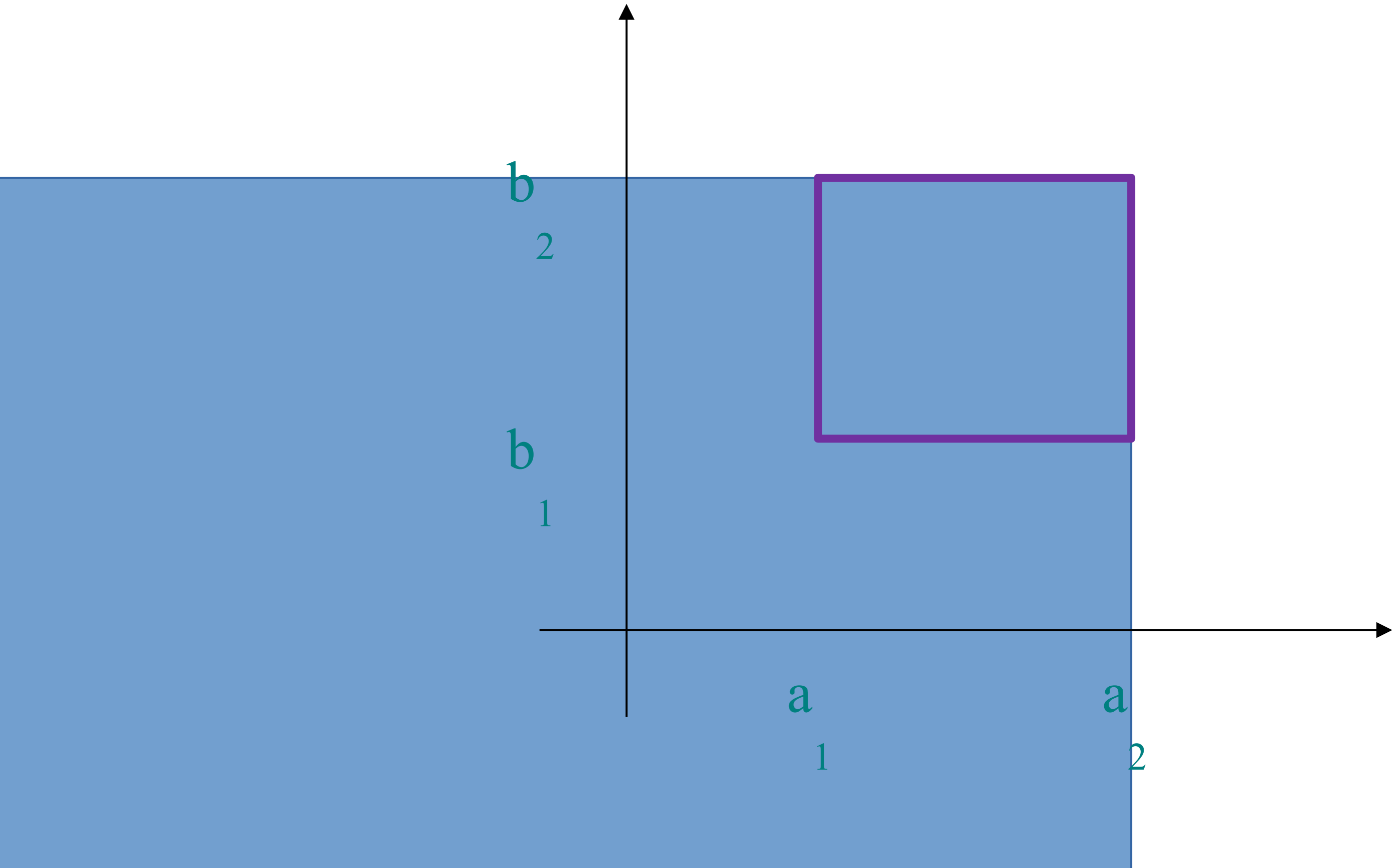
Probabilities from Joint CDF

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



Probabilities from Joint CDF

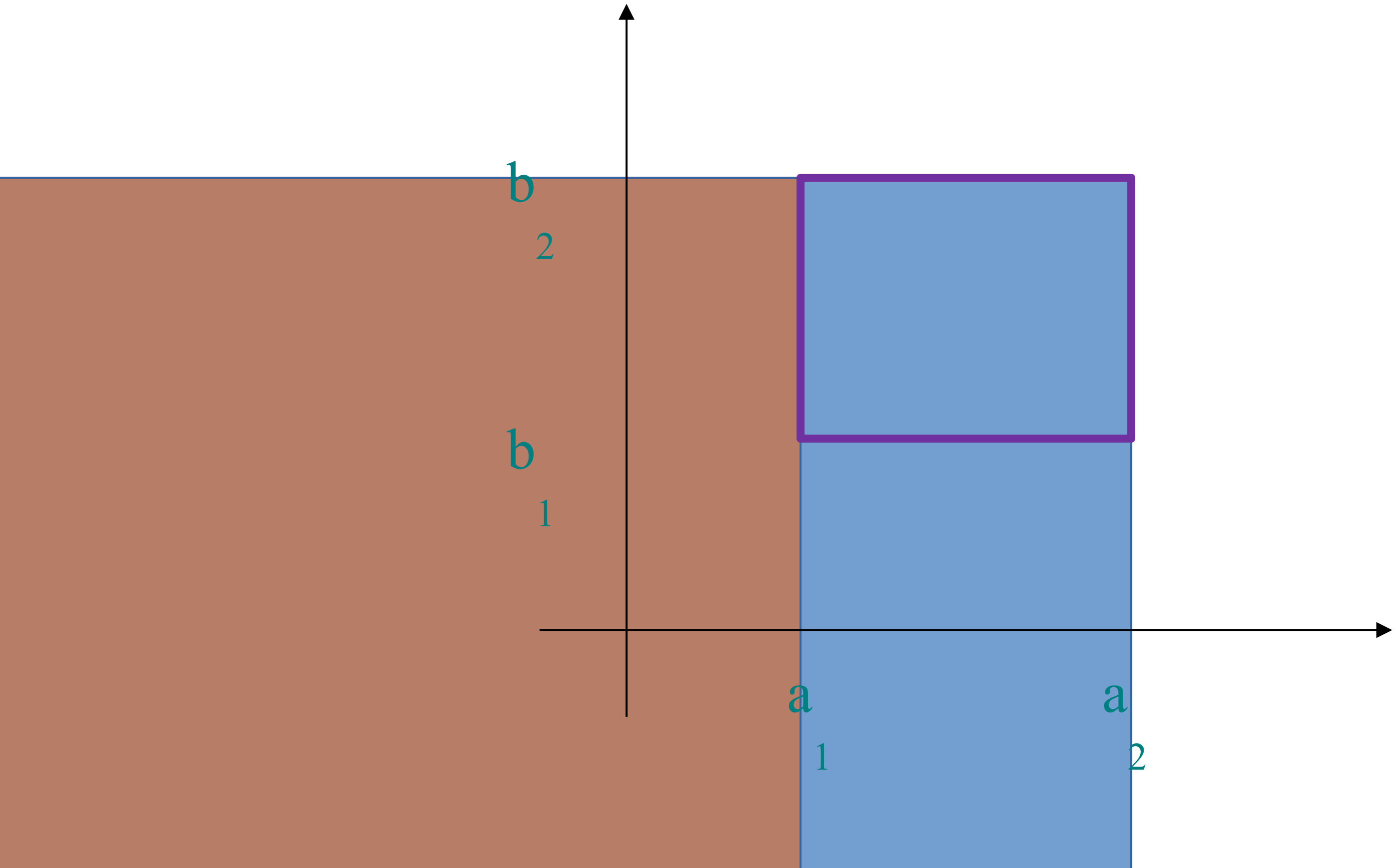
$$P\left(a_1 < X \le a_2, b_1 < Y \le b_2\right) = F_{X,Y}\left(a_{2,} b_2\right)$$



Probabilities from Joint CDF

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

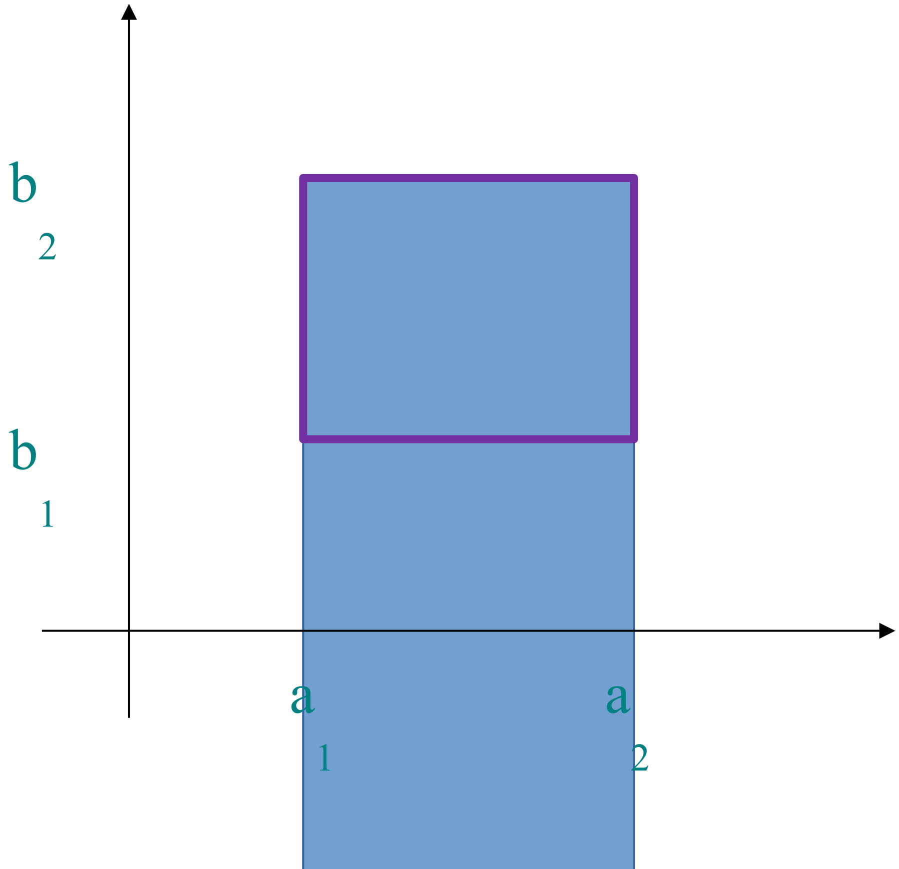
$$-F_{X,Y}\left(a_1, b_2\right)$$



Probabilities from Joint CDF

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

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

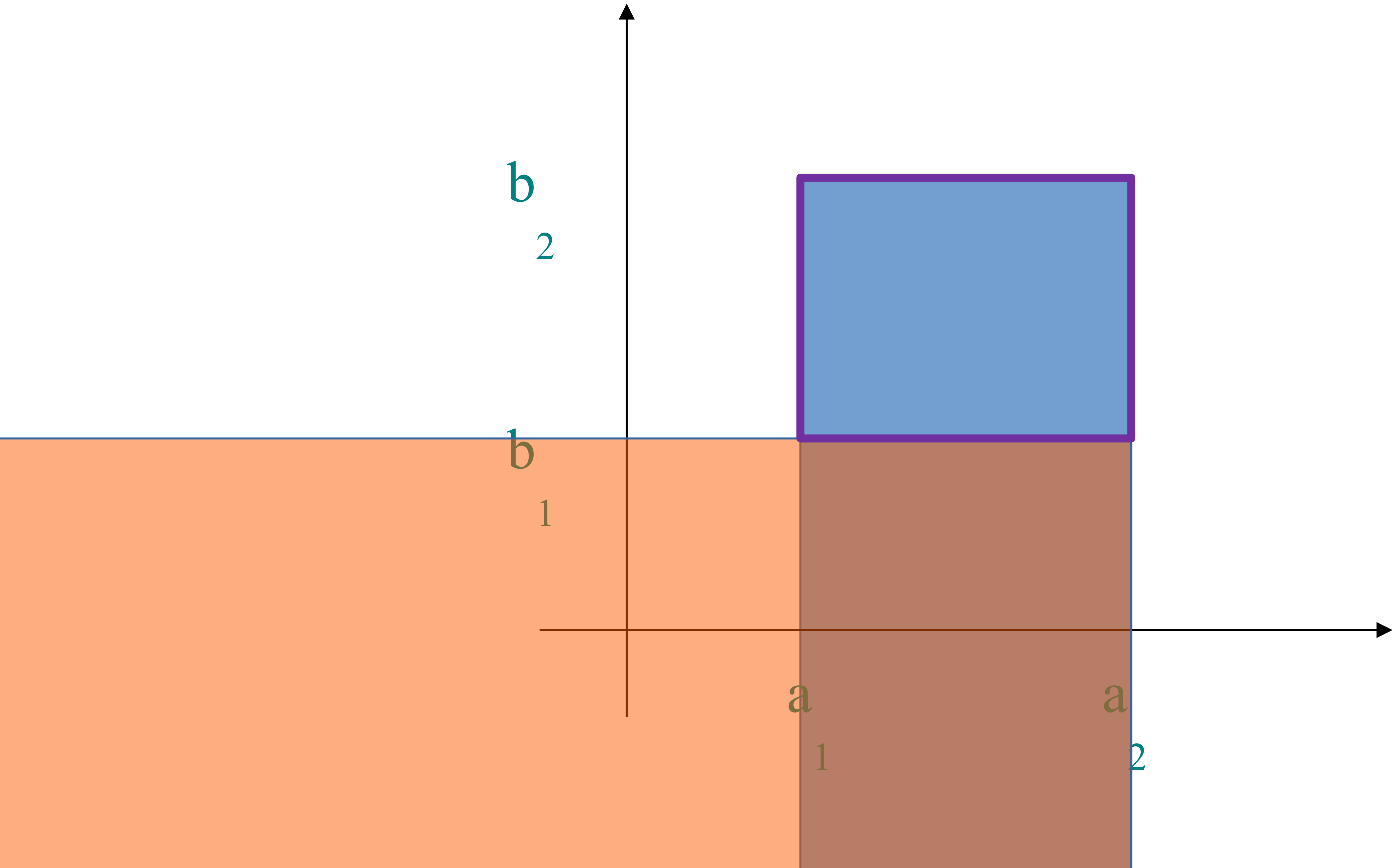


Probabilities from Joint CDF

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

$$-F_{X,Y}(a_{1,} b_2)$$

$$-F_{X,Y}(a_{2,} b_1)$$

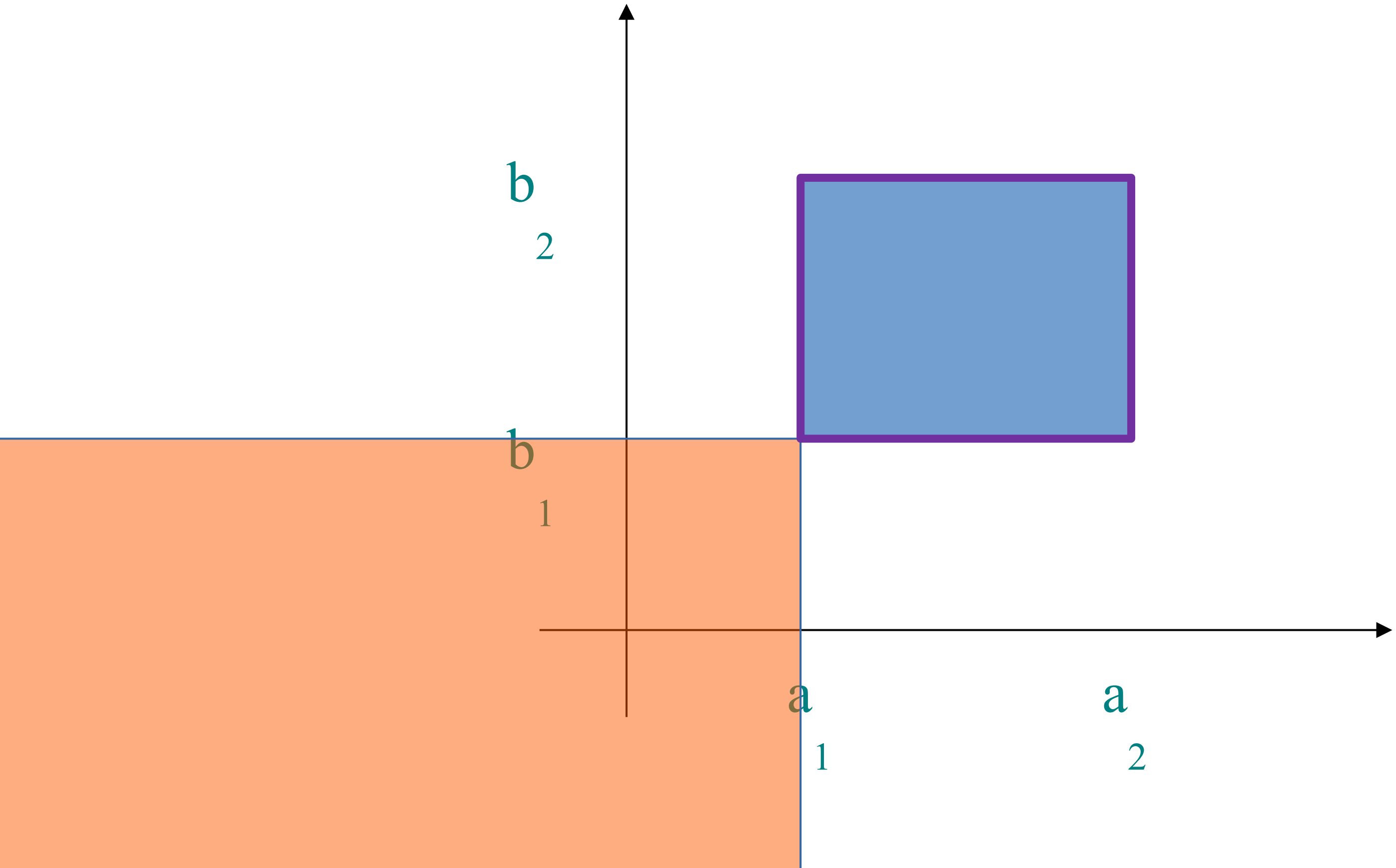


Probabilities from Joint CDF

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

$$-F_{X,Y}(a_{1,} b_2)$$

$$-F_{X,Y}(a_{2,} b_1)$$



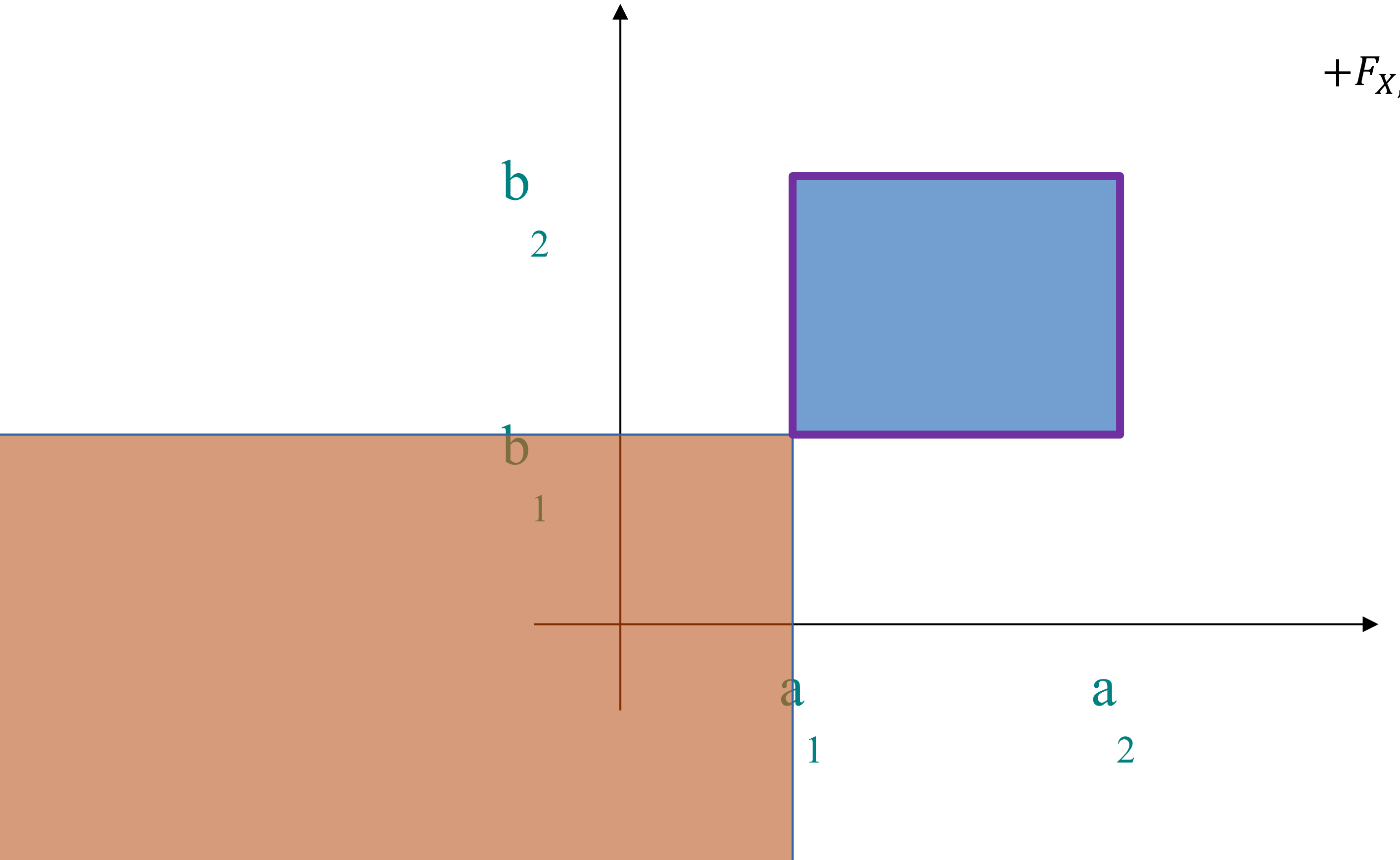
Probabilities from Joint CDF

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

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

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

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



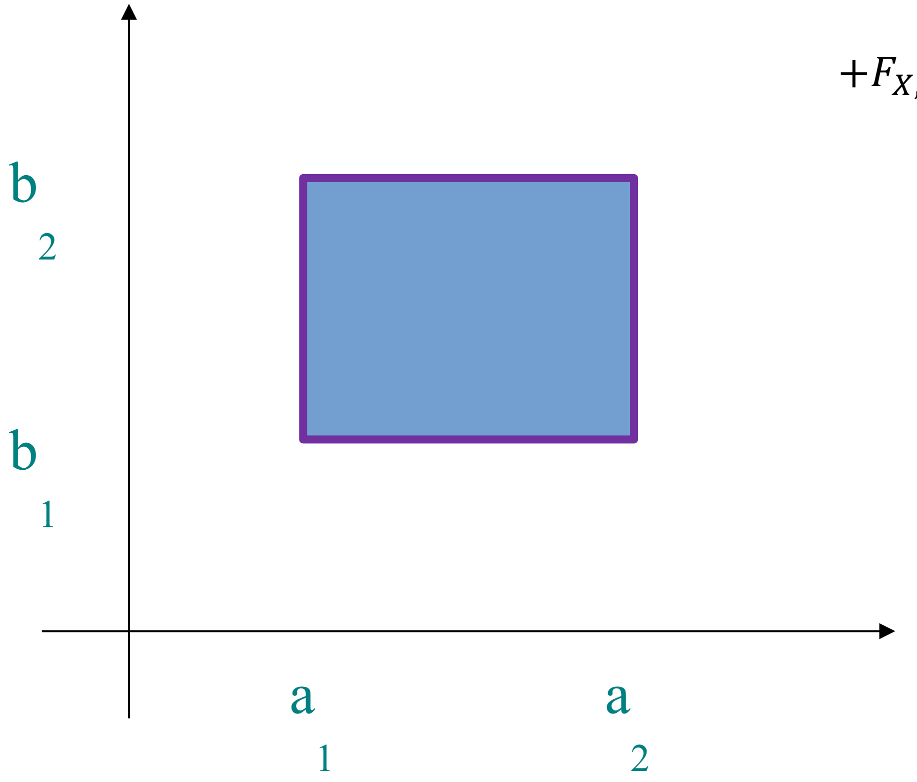
Probabilities from Joint CDF

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

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

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

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



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.

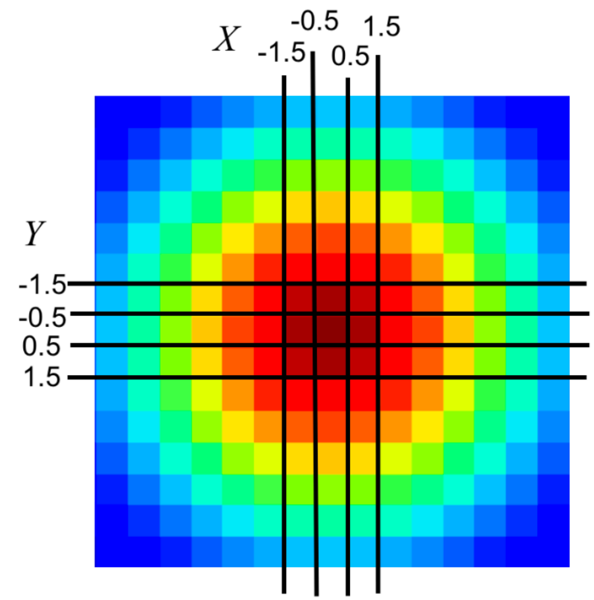
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

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

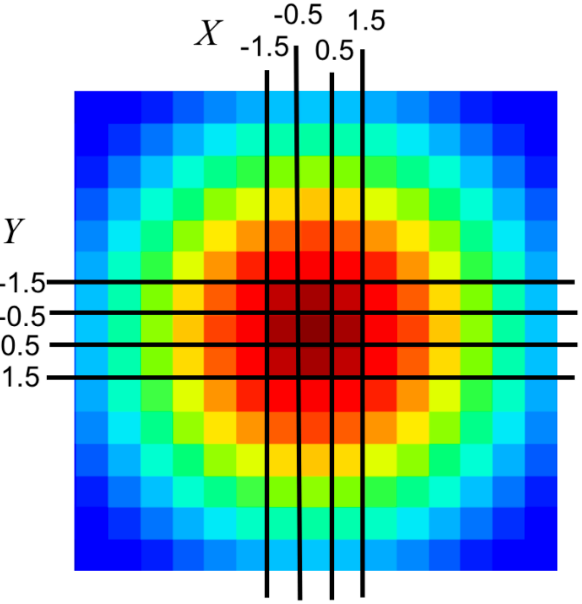
Joint PDF

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

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

$-0.5 \leq x \leq 0.5$ and $-0.5 \leq y \leq 0.5$

What is the weight of the center pixel?

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