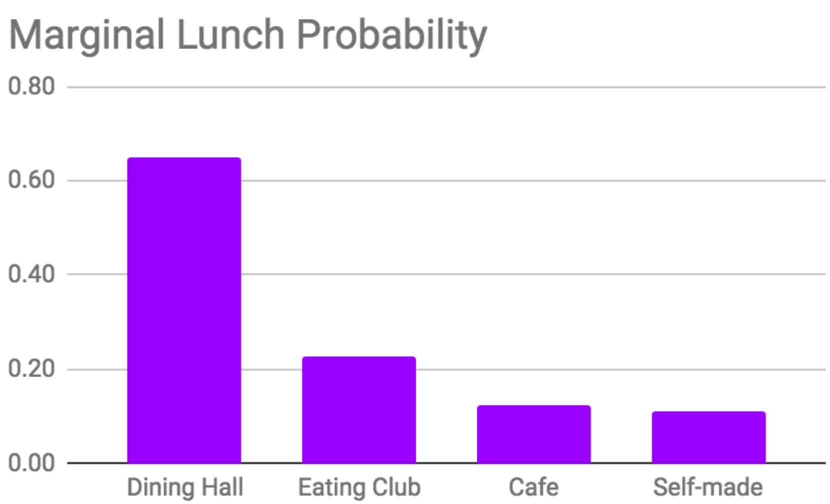
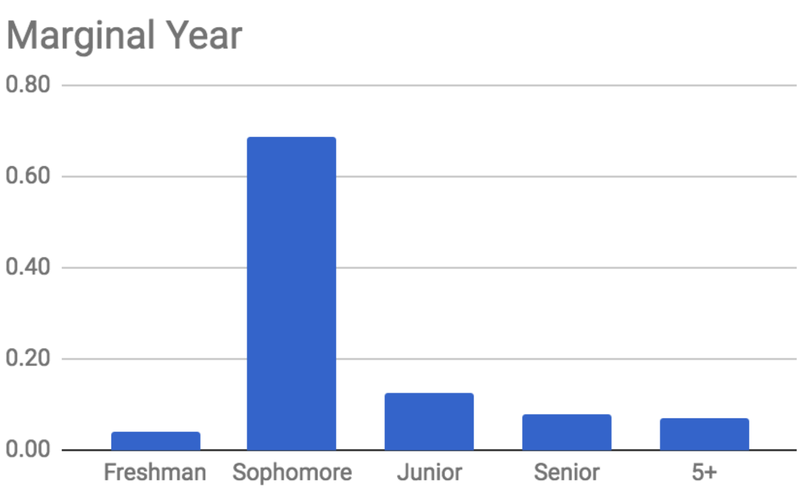
Properties of Joint Distributions

Chris Piech
CS109, Stanford University

Joint Probability Table

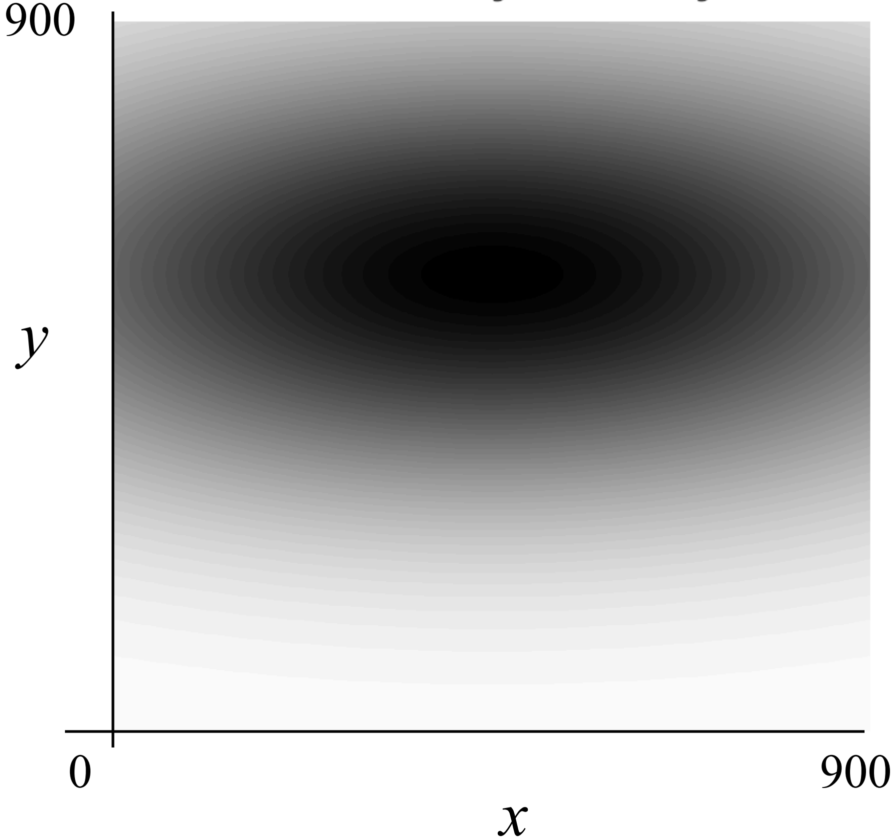
Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.02	0.00	0.02	0.00	0.04
Sophomore	0.51	0.15	0.03	0.03	0.69
Junior	0.08	0.02	0.02	0.02	0.13
Senior	0.02	0.05	0.01	0.01	0.08
5+	0.02	0.01	0.05	0.05	0.07
Marginal Status	0.65	0.23	0.13	0.11	





Continuous Joint Random Variables

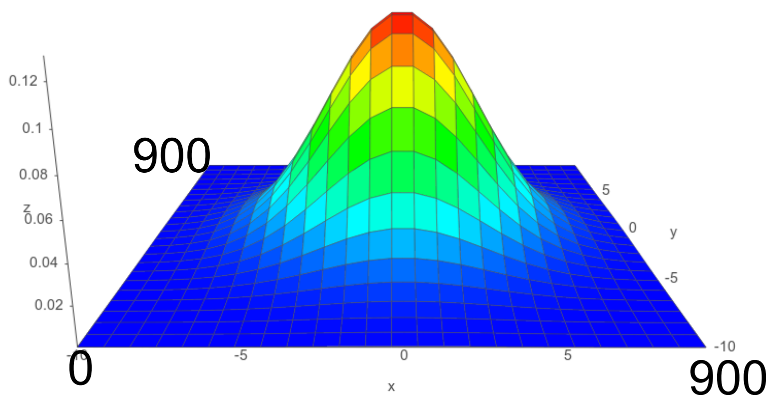
Dart Probability Density

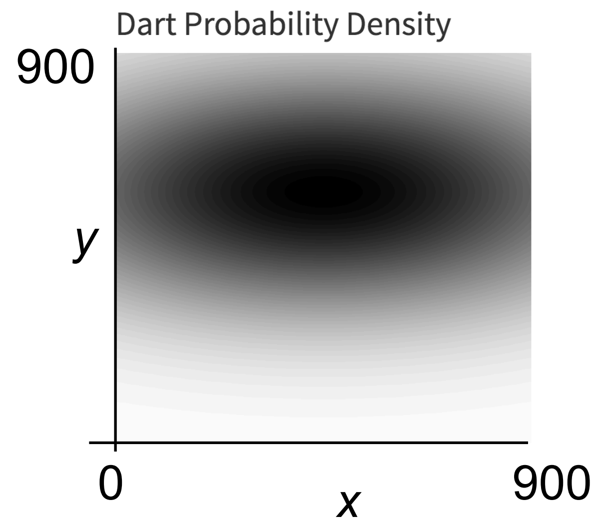


Dart Results

Joint Probability Density Function

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





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

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):

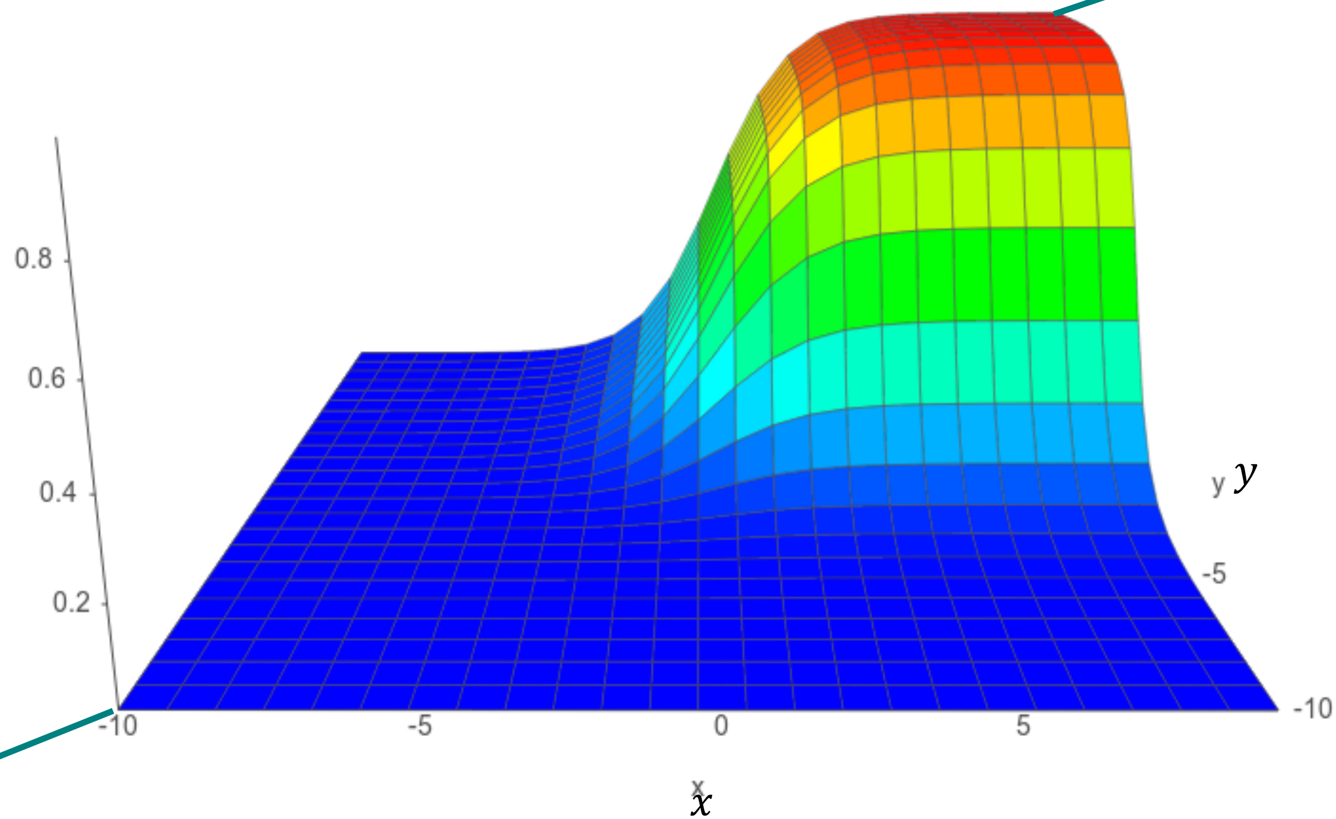
$$F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx$$

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

Jointly CDF

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

to 1 as
x ⟶ +∞,
y ⟶ +∞

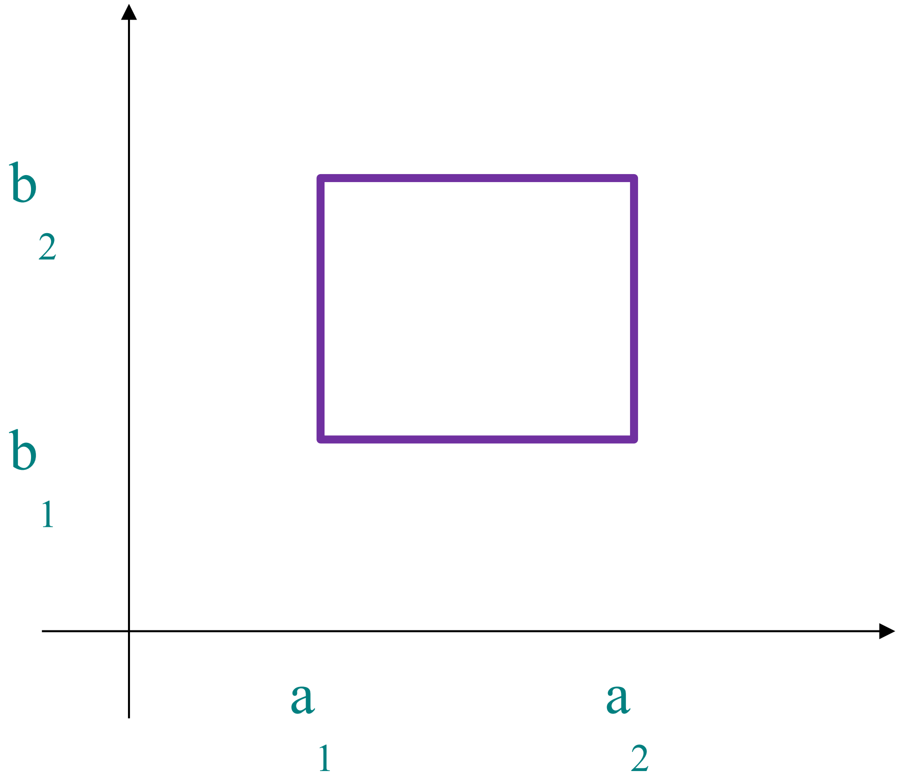


to 0 as
x ⟶ -∞,
y ⟶ -∞

plot by [Academo]

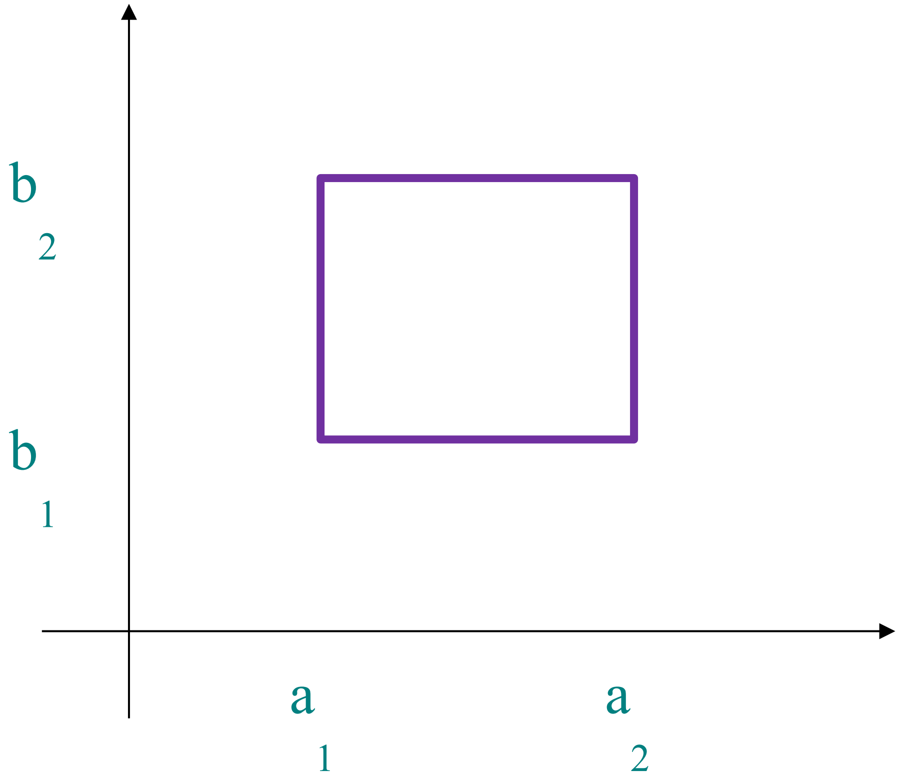
Probabilities from Joint CDF

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



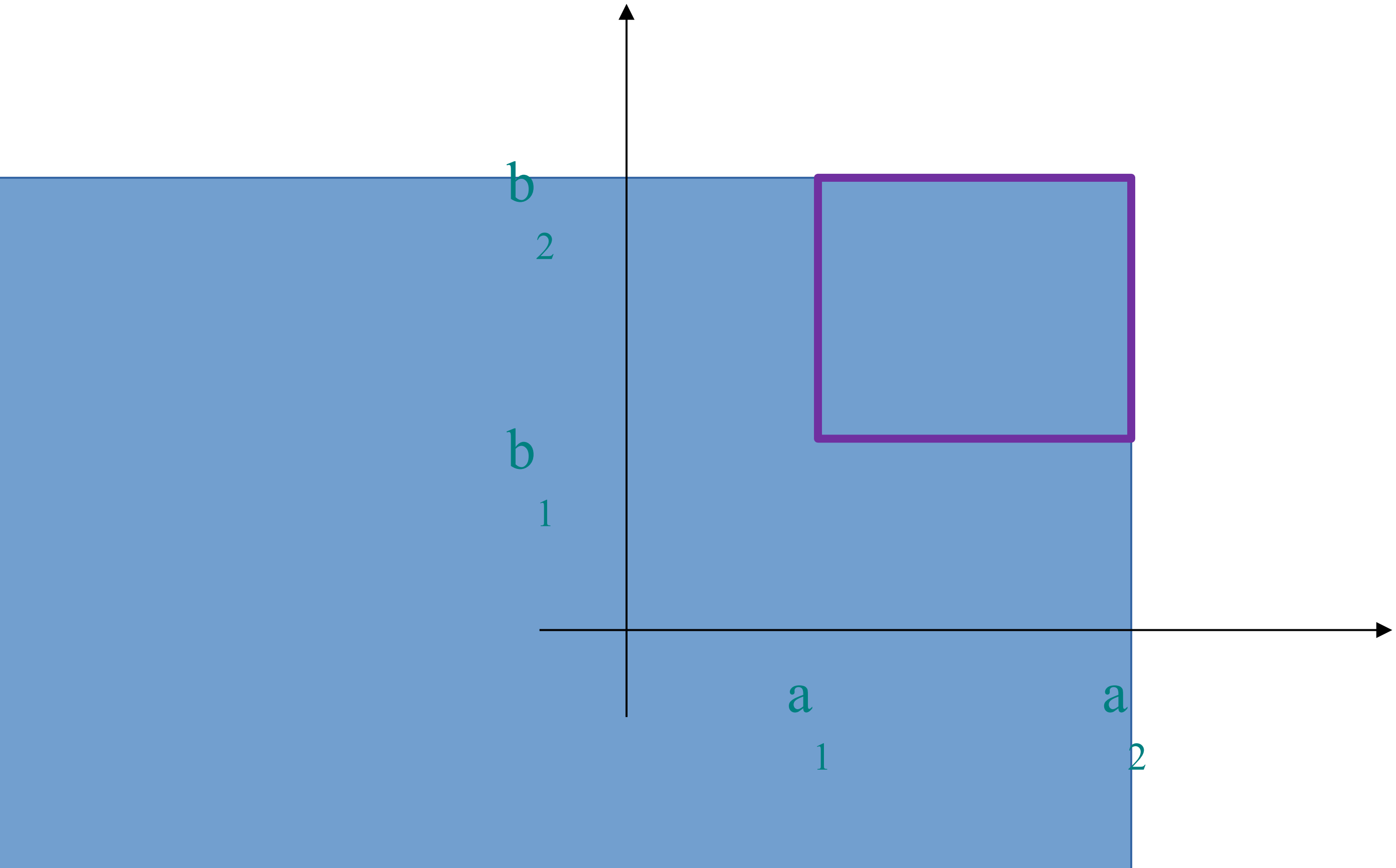
Probabilities from Joint CDF

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



Probabilities from Joint CDF

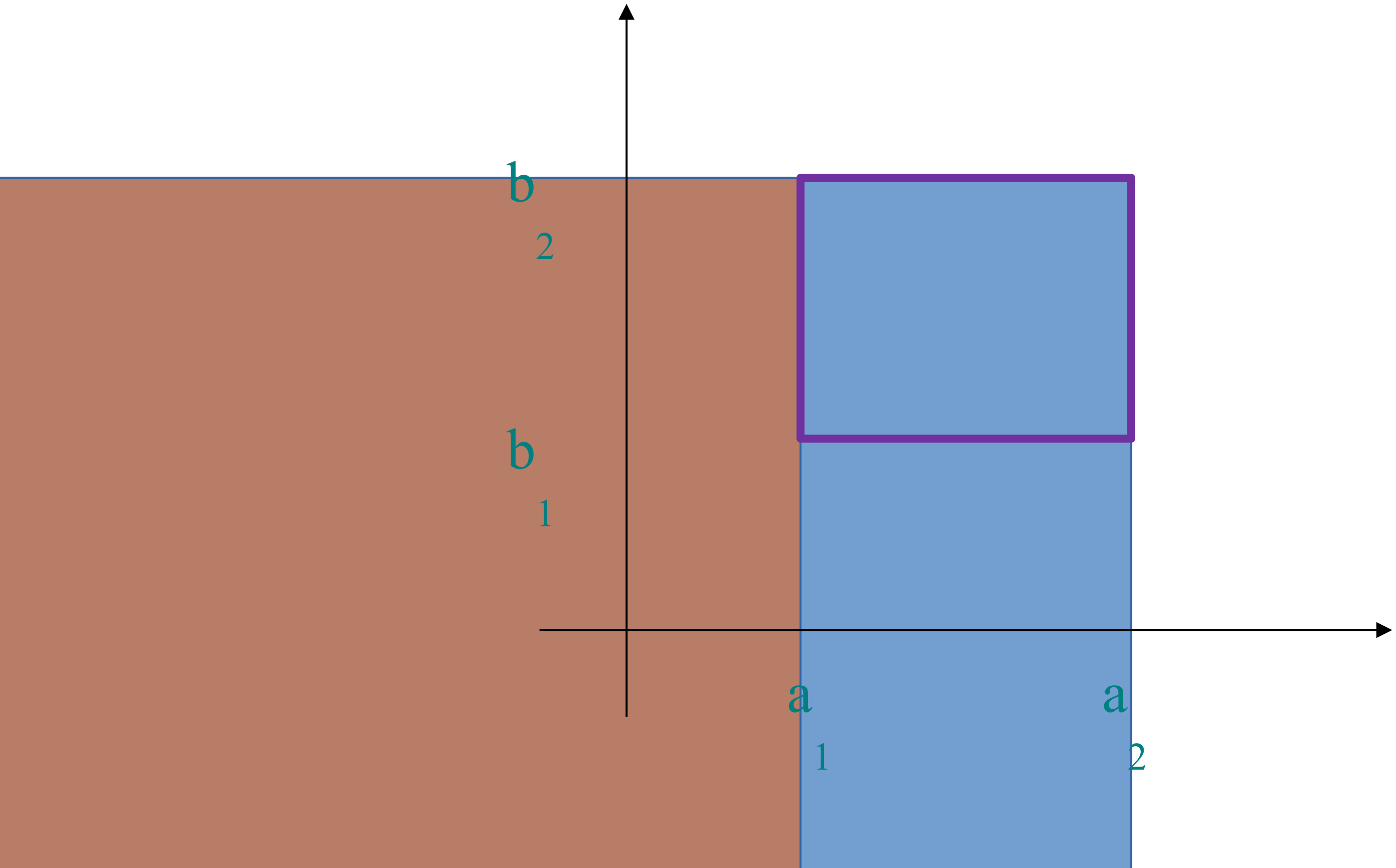
$$P\left(a_1 < X \leq a_2, b_1 < Y \leq b_2\right) = F_{X,Y}\left(a_{2,} b_2\right)$$



Probabilities from Joint CDF

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

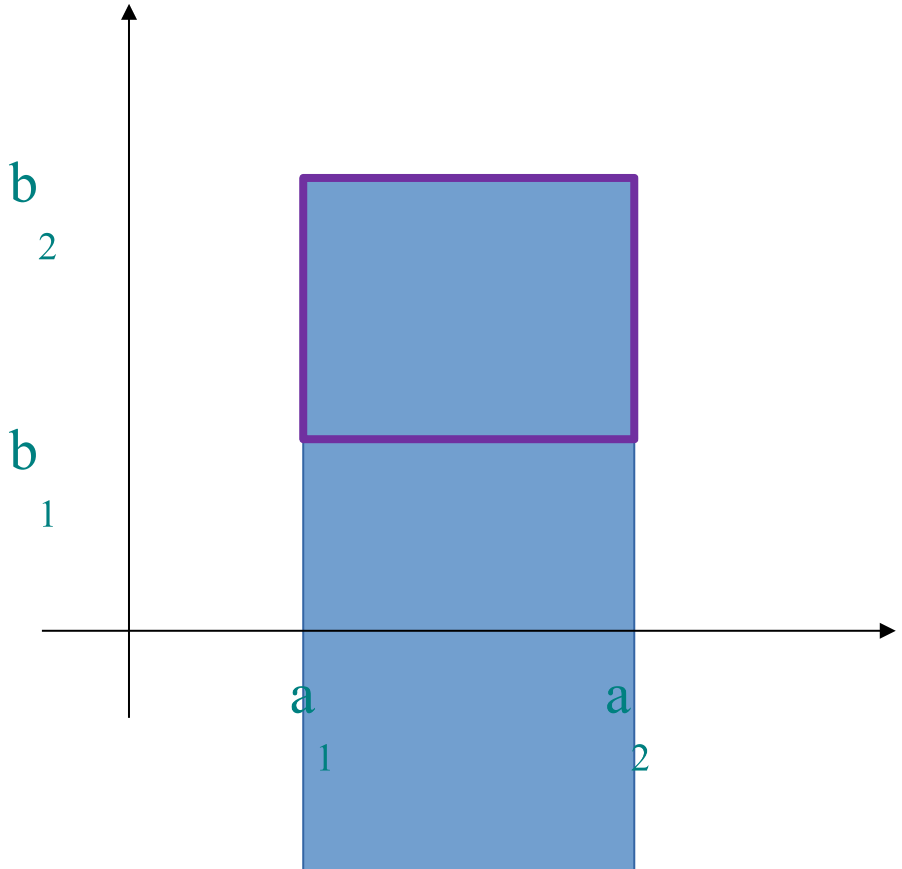
$$-F_{X,Y}\left(a_1, b_2\right)$$



Probabilities from Joint CDF

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

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

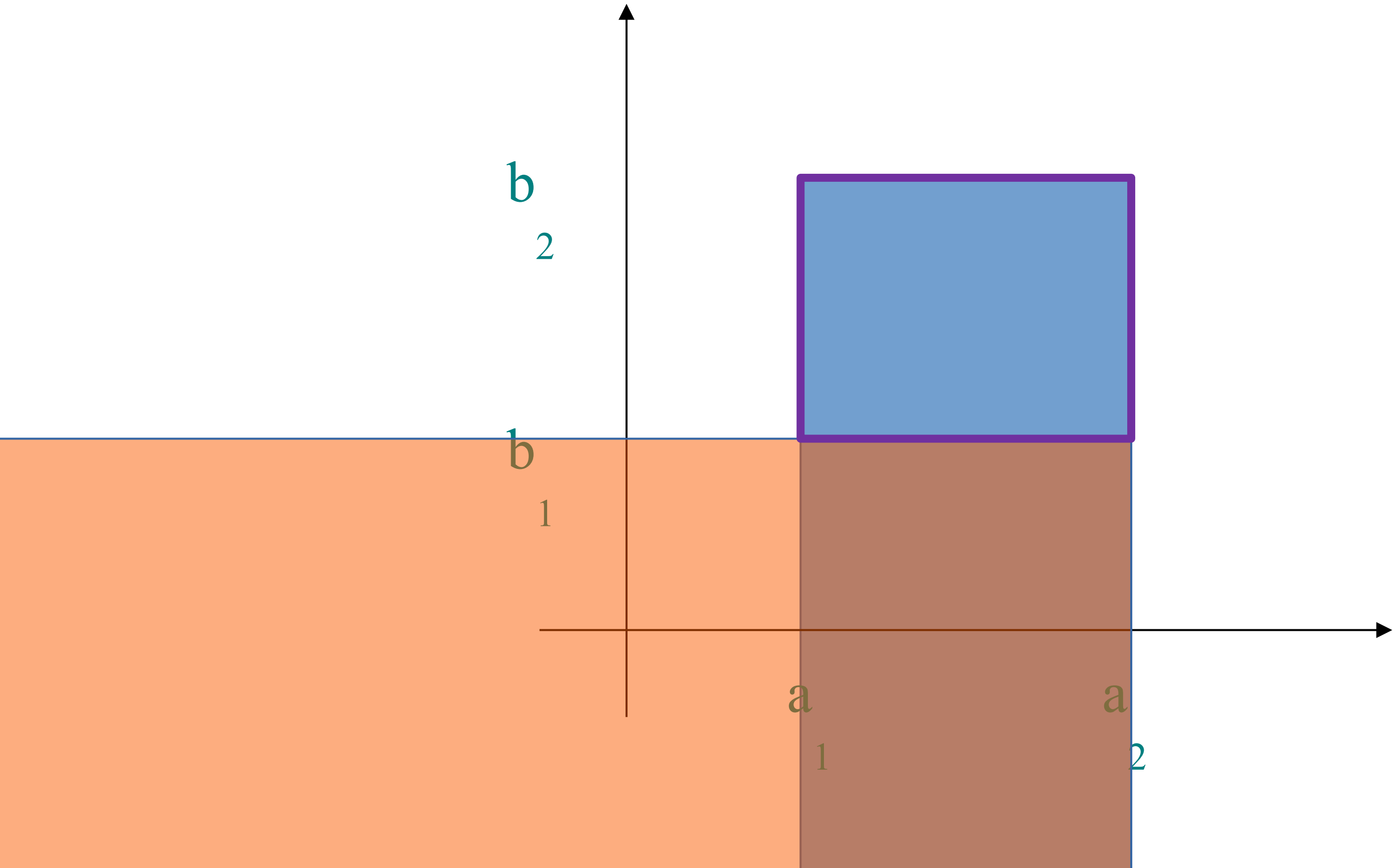


Probabilities from Joint CDF

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

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

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

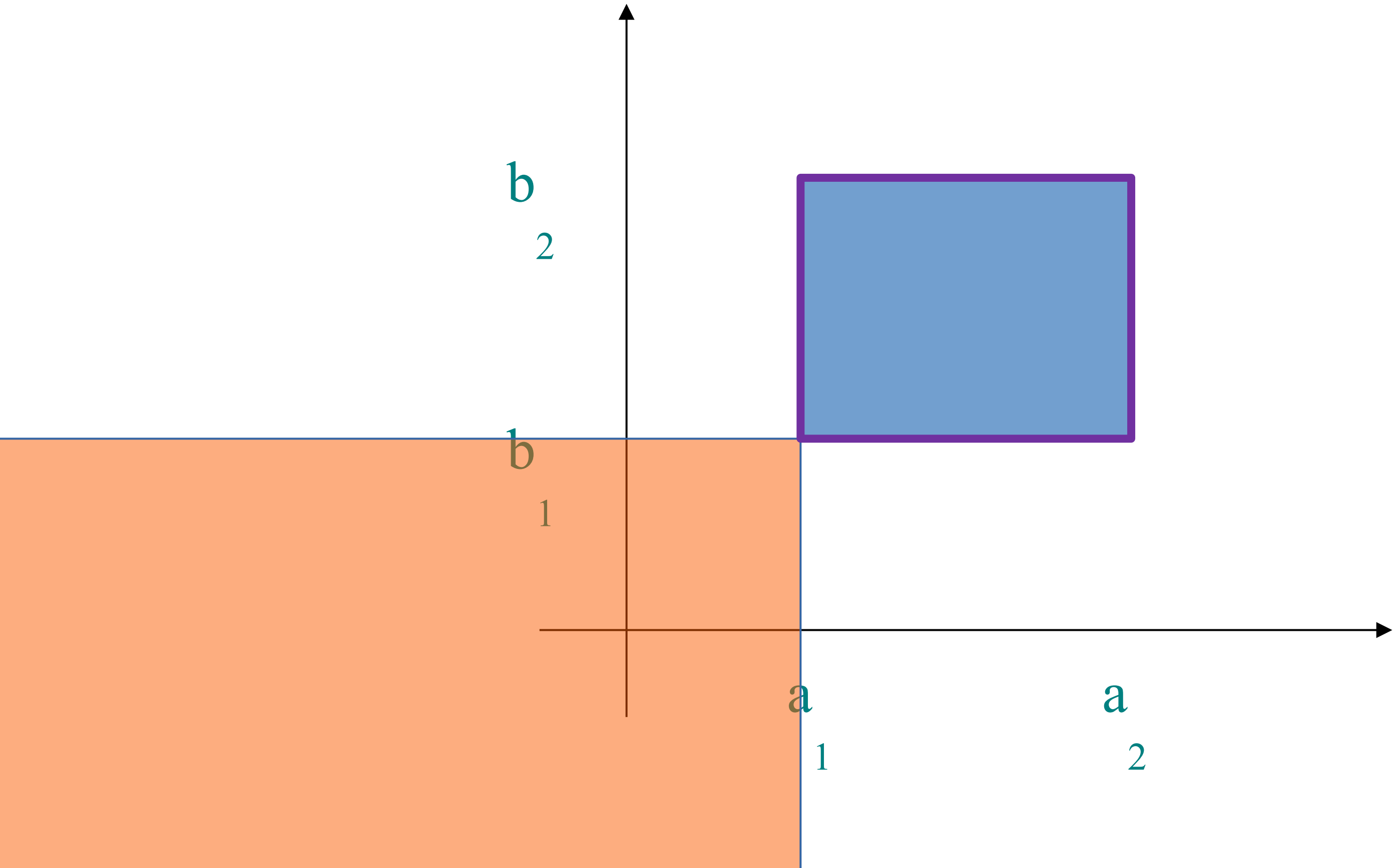


Probabilities from Joint CDF

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

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

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



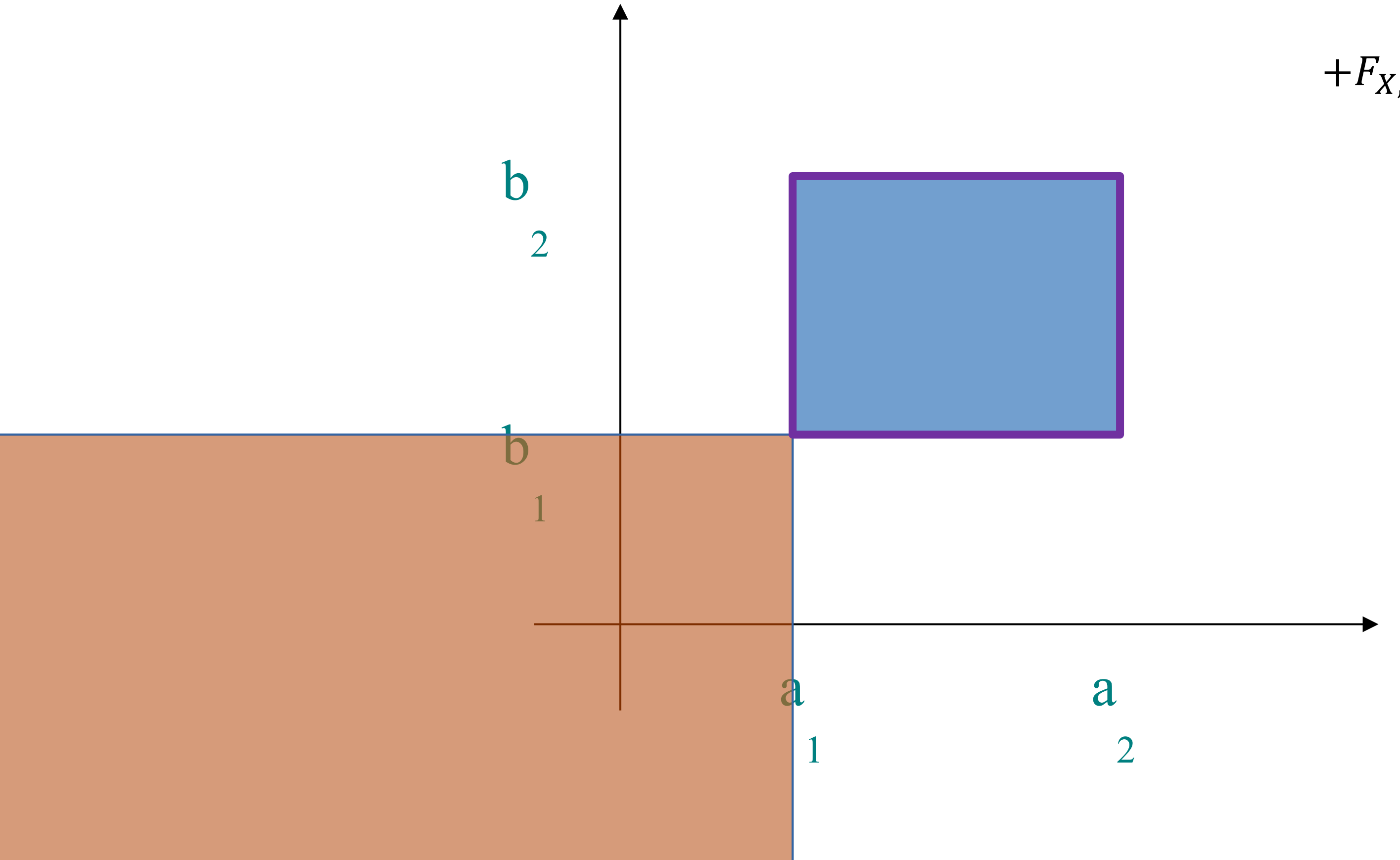
Probabilities from Joint CDF

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

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

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

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



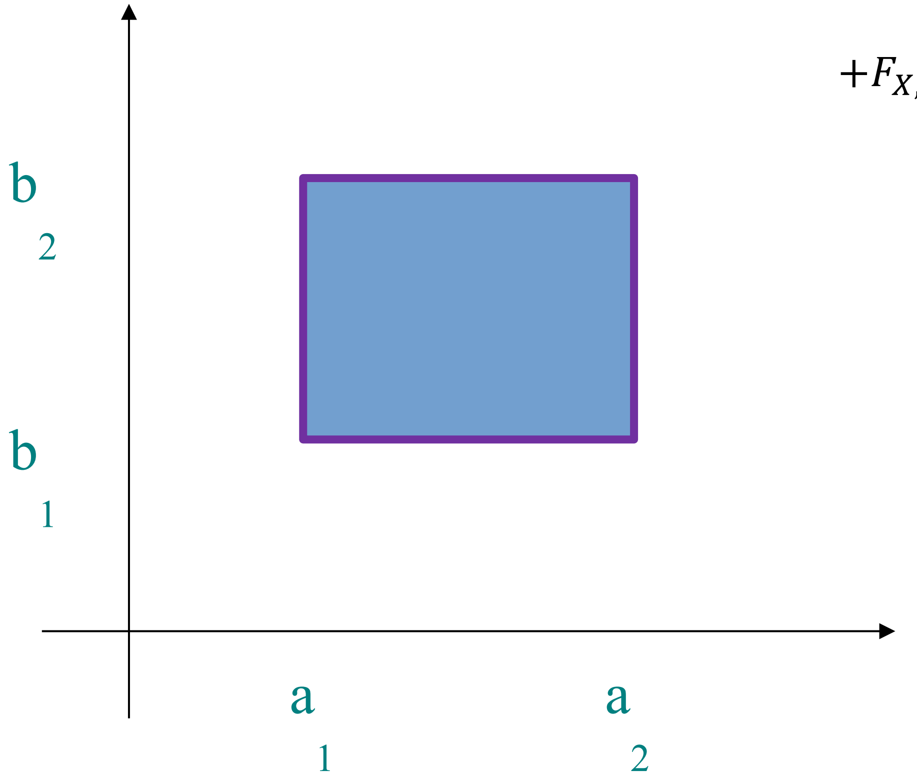
Probabilities from Joint CDF

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

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

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

$$+F_{X,Y}(a_{1,} b_1)$$

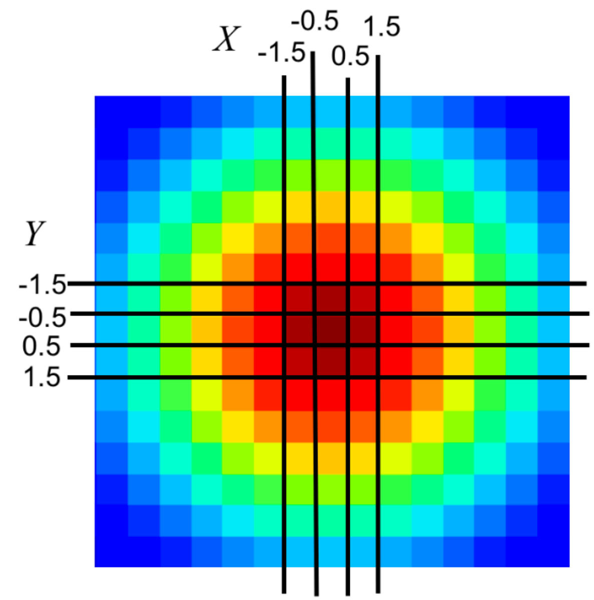


Probability for Instagram!

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.

0.0000	0.0000	0.0000	0.0001	0.0001	0
0.0000	0.0001	0.0005	0.0020	0.0032	0
0.0000	0.0005	0.0052	0.0206	0.0326	0
0.0001	0.0020	0.0206	0.0821	0.1300	0
0.0001	0.0032	0.0326	0.1300	**0.2060**	0
0.0001	0.0020	0.0206	0.0821	0.1300	0
0.0000	0.0005	0.0052	0.0206	0.0326	0
0.0000	0.0001	0.0005	0.0020	0.0032	0
0.0000	0.0000	0.0000	0.0001	0.0001	0



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.

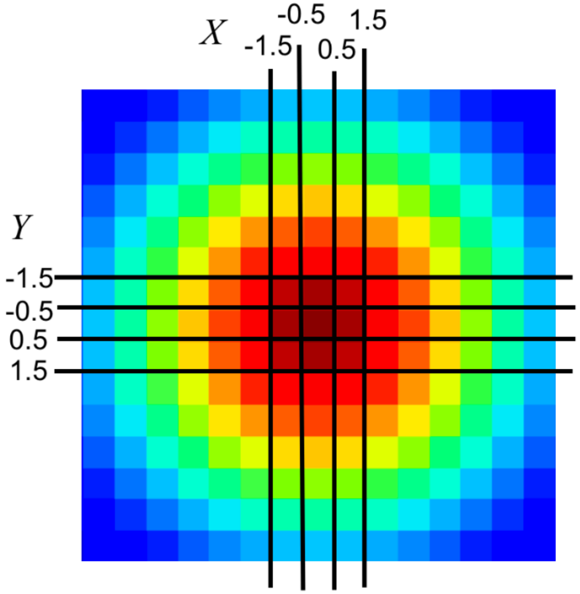
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

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

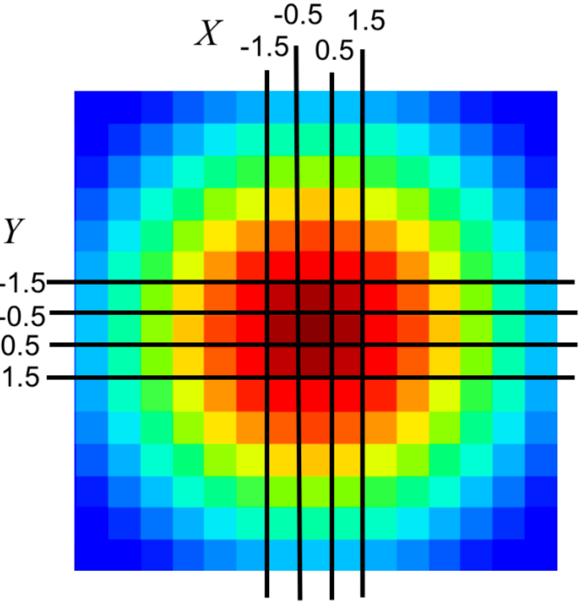
Joint PDF

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

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?

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

Properties of Joint Distributions

Boolean Operation on Variable = Event

Recall: any boolean question about a random variable makes for an event. For example:

$$P(X \leq 5)$$

$$P(Y = 6)$$

$$P(5 \leq Z \leq 10)$$

Independence and Random Variables

Independent Discrete Variables

- Two discrete random variables X and Y are called **independent** if:

$$p(x, y) = p_X(x)p_Y(y) \quad \text{for all } x, y$$

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

- Intuitively: knowing the value of X tells us nothing about the distribution of Y (and vice versa)
  - If two variables are **not** independent, they are called ***dependent***
- Similar conceptually to independent *events*, but we are dealing with multiple ***variables***
  - Keep your events and variables distinct (and clear)!

Is Year Independent of Lunch?

Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.03	0.00	0.02	0.00	0.05
Sophomore	0.50	0.15	0.03	0.03	0.68
Junior	0.08	0.02	0.02	0.02	0.12
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	0.22	0.12	0.11	

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Lunch} = s)}_{0.50} = \underbrace{P(\text{Year} = y)}_{0.68}\underbrace{P(\text{Lunch} = s)}_{0.65}$$

Yes!

Is Year Independent of Lunch?

Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.03	0.00	0.02	0.00	0.05
Sophomore	0.50	0.15	0.03	0.03	0.68
Junior	0.08	0.02	0.02	0.02	0.12
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	0.22	0.12	0.11	

For all values of Year, Status:

$$P(\text{Year} = y, \text{Lunch}= s) = P(\text{Year} = y)P(\text{Lunch} = s)$$

$$\underset{0.03}{} \qquad \underbrace{0.68 \quad 0.12}_{0.08}$$

No ☹

Aside: Butterfly Effect

Coin Flips

- Flip coin with probability *p* of "heads"
  - Flip coin a total of *n* + *m* times
  - Let X = number of heads in first *n* flips
  - Let Y = number of heads in next *m* flips

$$P(X = x, Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}$$

$$= P(X = x)P(Y = y)$$

  - X and Y are independent
  - Let Z = number of total heads in *n* + *m* flips
  - Are X and Z independent?
    - What if you are told Z = 0?

Recall: Poisson Random Variable

- X is a **Poisson** Random Variable: the number of occurrences in a fixed interval of time.

$$X \sim \mathrm{Poi}(\lambda)$$

  - λ is the "rate"
  - X takes on values 0, 1, 2…
  - has distribution (PMF):

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = p) or from a "bot" (probability = $(1 – p)$), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, p)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - p)

$$
\begin{aligned}
P(X=i,Y=j) = {} & P(X=i,Y=j \mid X+Y=i+j)P(X+Y=i+j) \\
& + P(X=i,Y=j \mid X+Y \neq i+j)P(X+Y \neq i+j)
\end{aligned}
$$

Probability of *i* human requests and *j* bot requests

Probability of *i* human requests and *j* bot requests | we got *i + j* requests

Probability of number of requests in a day was i + j

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$
\begin{aligned}
P(X = i, Y = j) = {} & P(X = i, Y = j \mid X + Y = i + j) P(X + Y = i + j) \\
& + P(X = i, Y = j \mid X + Y \neq i + j) P(X + Y \neq i + j)
\end{aligned}
$$

  - Note: $P(X = i, Y = j \mid X + Y \neq i + j) = 0$

You got *i* human requests and *j* bot requests

You did not get *i* + *j* requests

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = p) or from a "bot" (probability = (1 – p)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, p)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - p)

$$P(X=i, Y=j) = P(X=i, Y=j \mid X+Y=i+j)P(X+Y=i+j)$$

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$P(X=i,Y=j)=P(X=i,Y=j\mid X+Y=i+j)P(X+Y=i+j)$$

$$P(X=i,Y=j\mid X+Y=i+j)=\binom{i+j}{i}p^i(1-p)^j$$

Binomial

$$P(X+Y=i+j)=e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}$$

Poisson

$$P(X=i,Y=j)=\binom{i+j}{i}p^i(1-p)^j e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}$$

Joint

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$P(X=i,Y=j)=\frac{(i+j)!}{i!\,j!}p^i(1-p)^j e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}=e^{-\lambda}\frac{(\lambda p)^i}{i!}\cdot\frac{(\lambda(1-p))^j}{j!}$$

Reorder terms

$$=e^{-\lambda p}\frac{(\lambda p)^i}{i!}\cdot e^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}=P(X=i)P(Y=j)$$

  - Where X ~ Poi(λp) and Y ~ Poi(λ(1 – *p*))
  - X and Y are independent!

Independent Continuous Variables

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

  P(X ≤ *a*, Y ≤ *b*) = P(X ≤ *a*) P(Y ≤ *b*) for any *a*, *b*

- Equivalently:

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \quad \text{for all } a,b$$
$$f_{X,Y}(a,b) = f_X(a)f_Y(b) \quad \text{for all } a,b$$

- More generally, joint density factors separately:

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

Is the Blur Distribution Independent?

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.

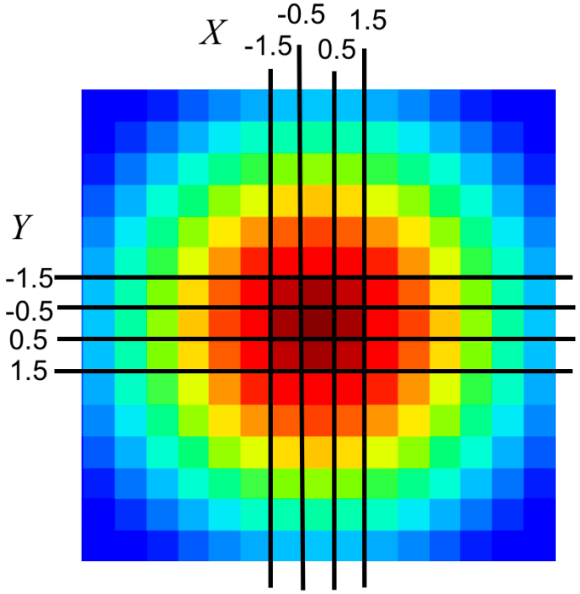
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

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

Pop Quiz (just kidding)

- Consider joint density function of X and Y:

$$f_{X,Y}(x, y) = 6e^{-3x}e^{-2y} \quad \text{for } \ 0 < x, y < \infty$$

  - Are X and Y independent? Yes!

Let $h(x) = 3e^{-3x}$ and $g(y) = 2e^{-2y}$, so $f_{X,Y}(x, y) = h(x)g(y)$

- Consider joint density function of X and Y:

$$f_{X,Y}(x, y) = 4xy \quad \text{for } \ 0 < x, y < 1$$

  - Are X and Y independent? Yes!

Let $h(x) = 2x$ and $g(y) = 2y$, so $f_{X,Y}(x, y) = h(x)g(y)$

  - Now add constraint that: $0 < (x + y) < 1$
  - Are X and Y independent? No!
    - Cannot capture constraint on x + y in factorization!

Independence of Multiple Variables

- *n* random variables X_1, X_2, …, X_n are called **independent** if:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \quad \text{for all subsets of } x_1, x_2, \ldots, x_n$$

- Analogously, for continuous random variables:

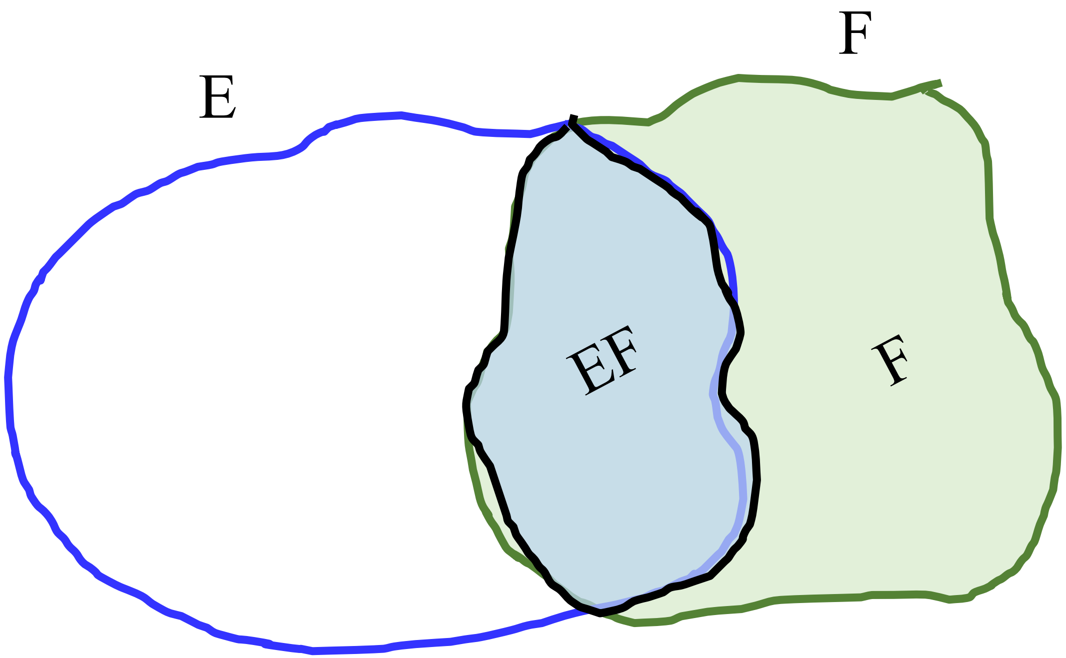
$$P(X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n) = \prod_{i=1}^{n} P(X_i \le a_i) \quad \text{for all subsets of } a_1, a_2, \ldots, a_n$$

Conditionals with multiple variables

Discrete Conditional Distribution

- Recall that for *events* E and F:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where } P(F) > 0$$



Discrete Conditional Distributions

- Recall that for *events* E and F:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where } P(F) > 0$$

- Now, have X and Y as discrete random variables
  - Conditional PMF of X given Y (where $p_Y(y) > 0$):

$$P_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

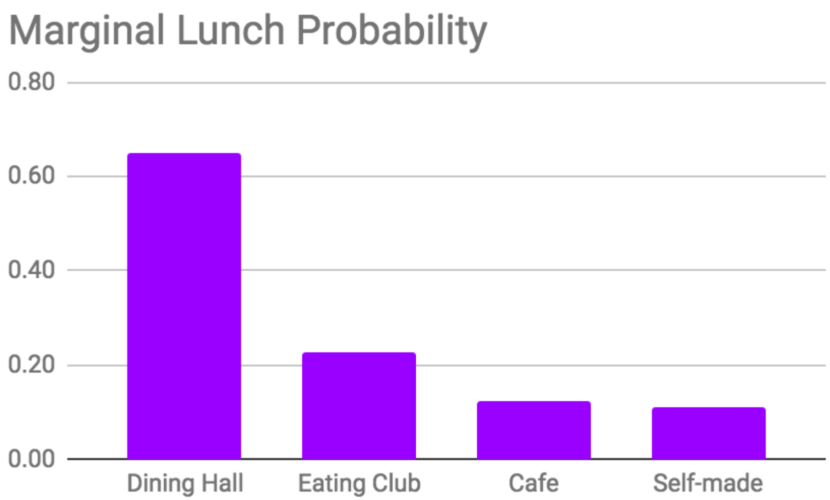
  - Conditional CDF of X given Y (where $p_Y(y) > 0$):

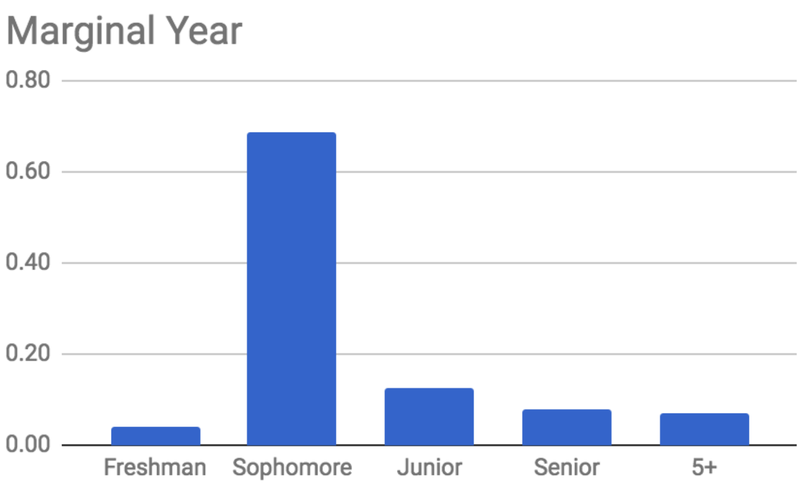
$$F_{X|Y}(a \mid y) = P(X \le a \mid Y = y) = \frac{P(X \le a, Y = y)}{P(Y = y)}$$

$$= \frac{\sum_{x \le a} p_{X,Y}(x, y)}{p_Y(y)} = \sum_{x \le a} p_{X|Y}(x \mid y)$$

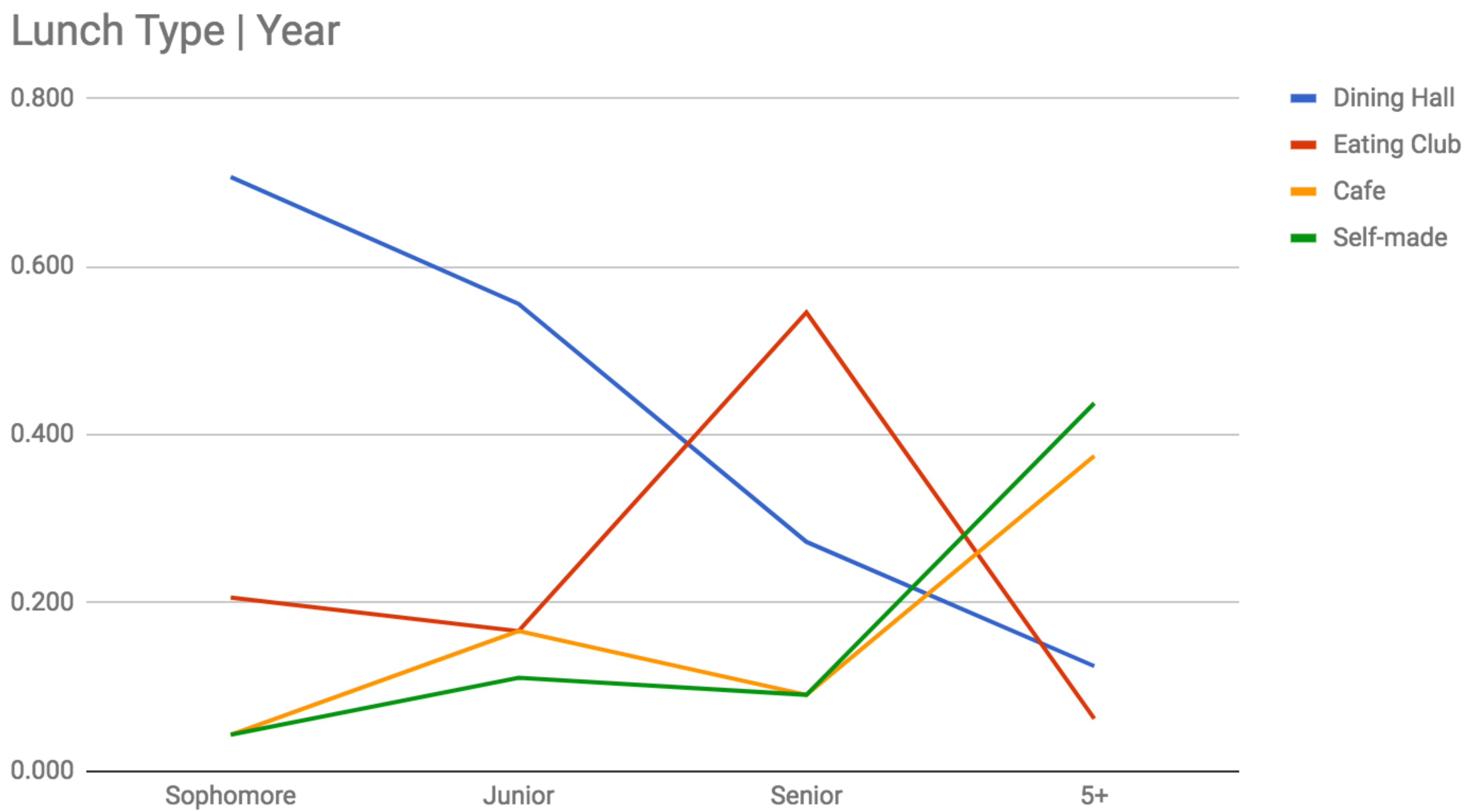
Joint Probability Table

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	





Lunch | Year



Lunch Type | Year
0.800
0.600
0.400
0.200
0.000
Sophomore
Junior
Senior
5+
Dining Hall
Eating Club
Cafe
Self-made

And It Applies to Books Too

P(Buy Book Y | Bought Book X)

Continuous Conditional Distributions

Let X and Y be continuous random variables

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

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

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

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

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

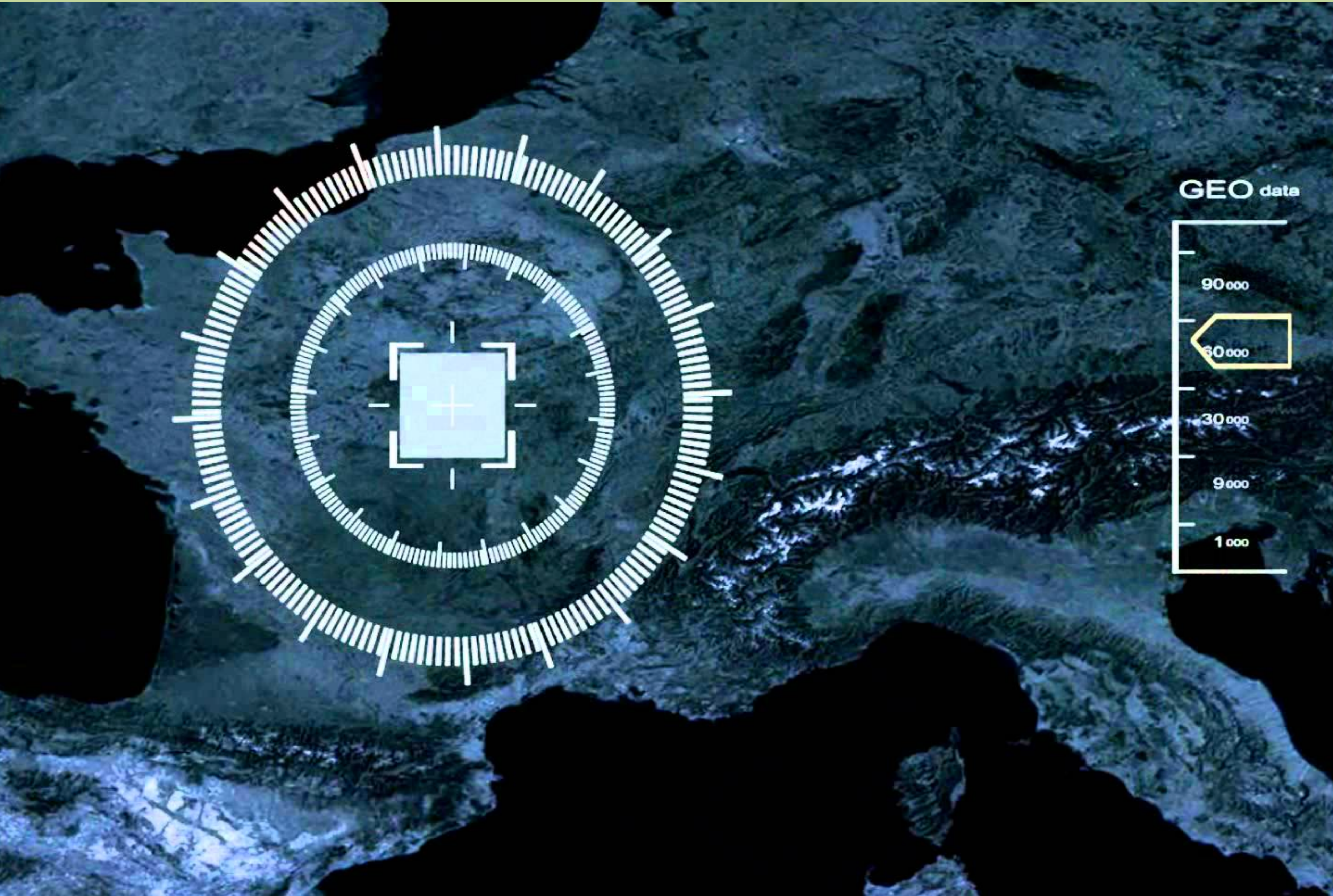
Mix Bayes #2

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

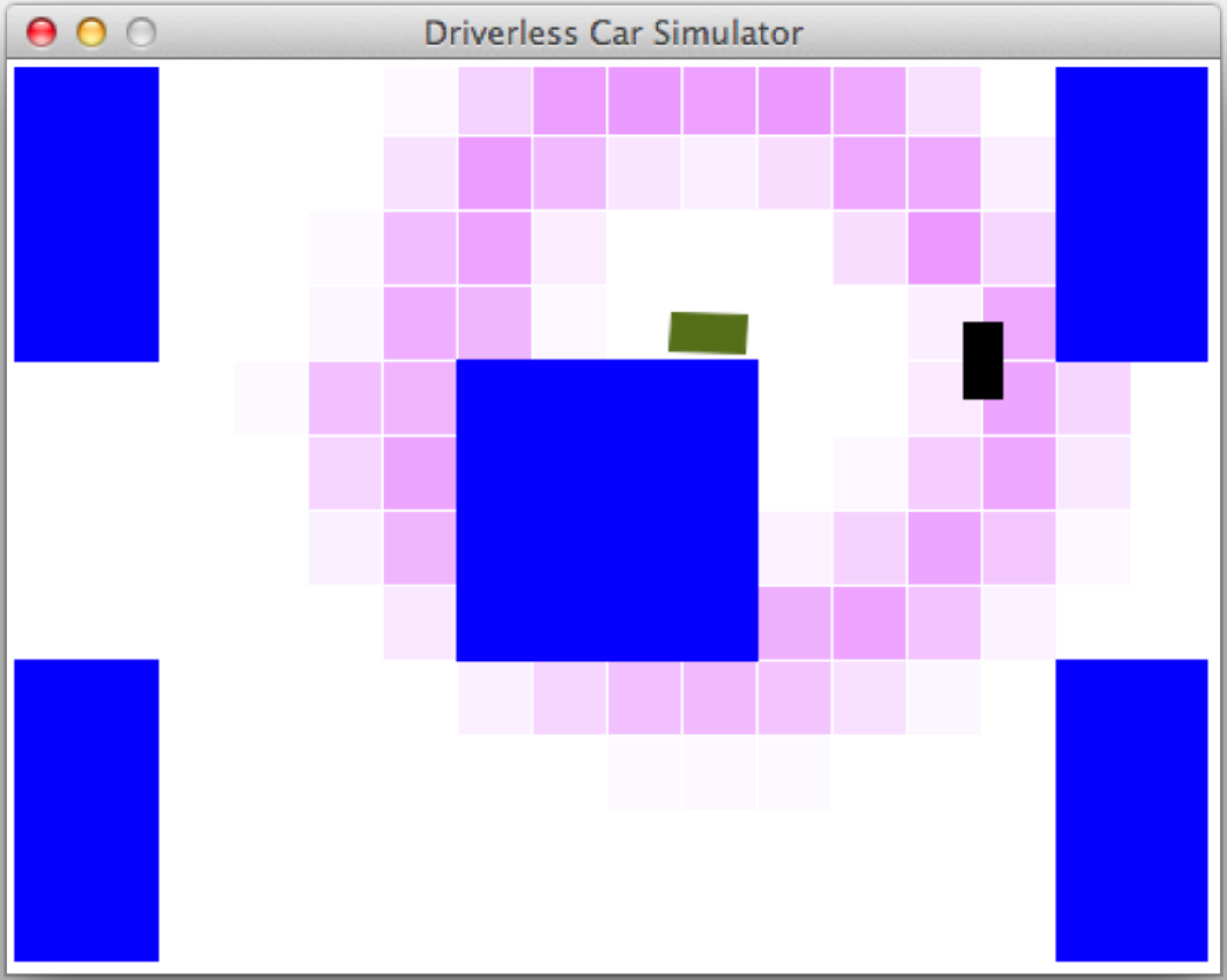
All Continuous

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

Tracking in 2D Space?



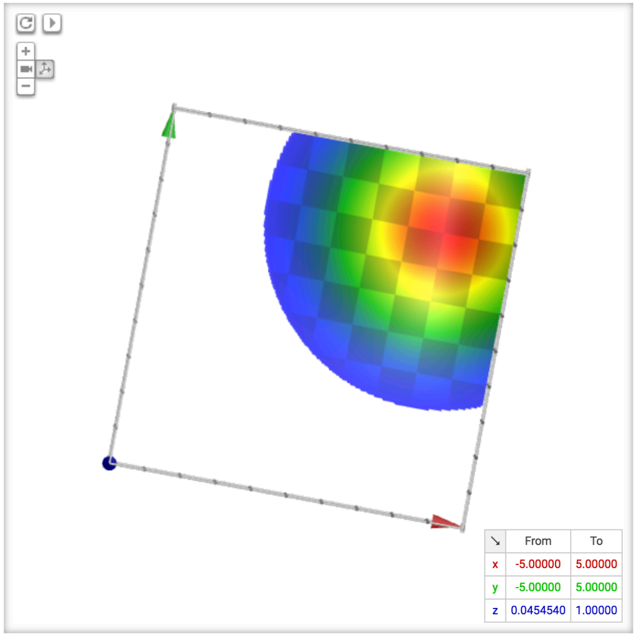
Tracking in 2D Space: CS221



Bivariate Normal

- X, Y follow a symmetric bivariate normal distribution if it has PDF:

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

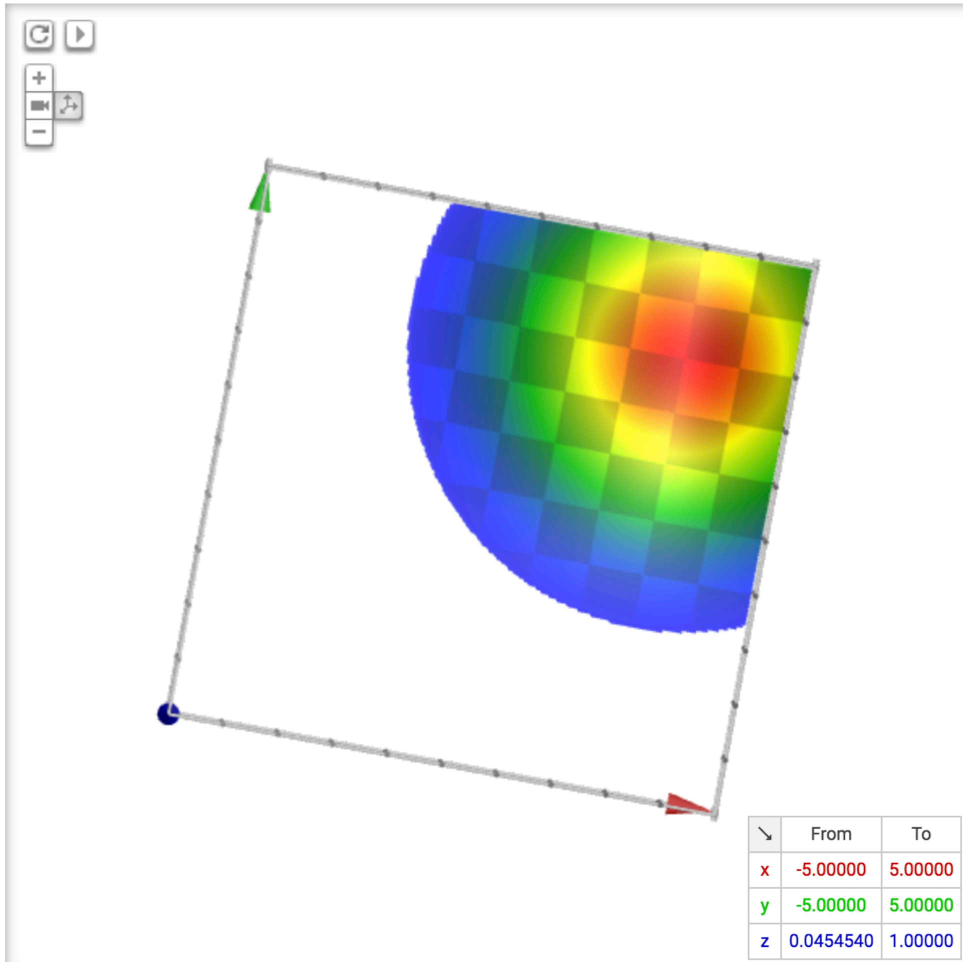


Here is an example where

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

Tracking in 2D Space: Prior

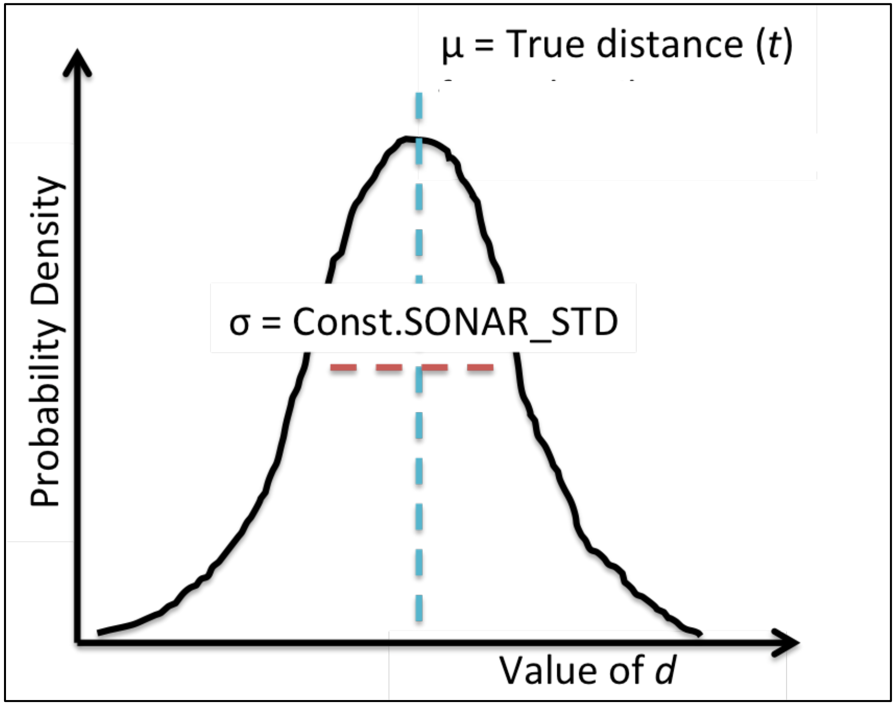
$$f_{X,Y}(x,y) = K \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$



Tracking in 2D Space: Observation!

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

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

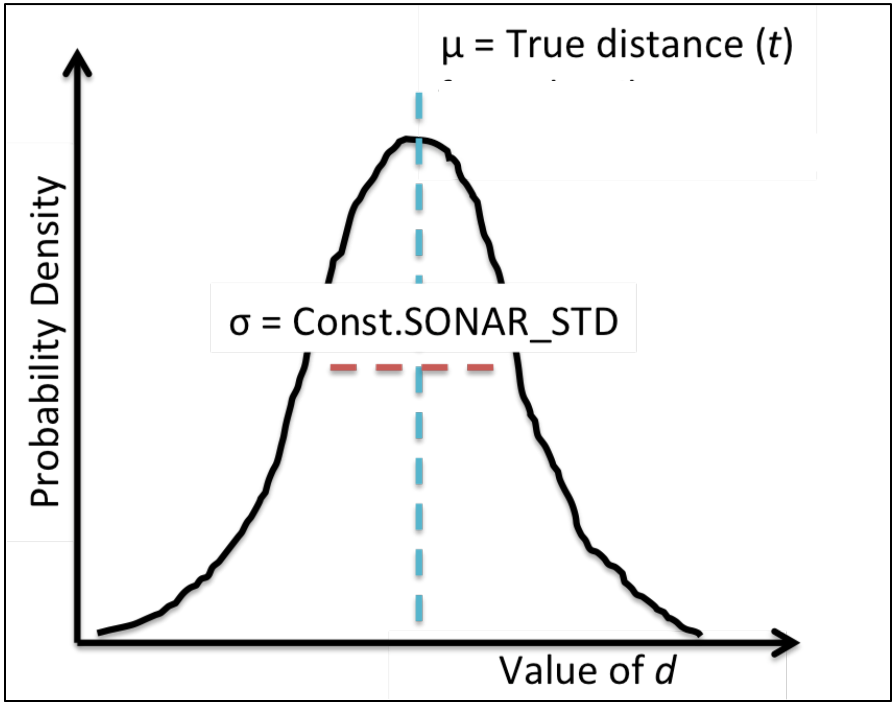


What is your new belief for the location of the object being tracked? Your probability density function can be expressed with a constant

Tracking in 2D Space: Observation!

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

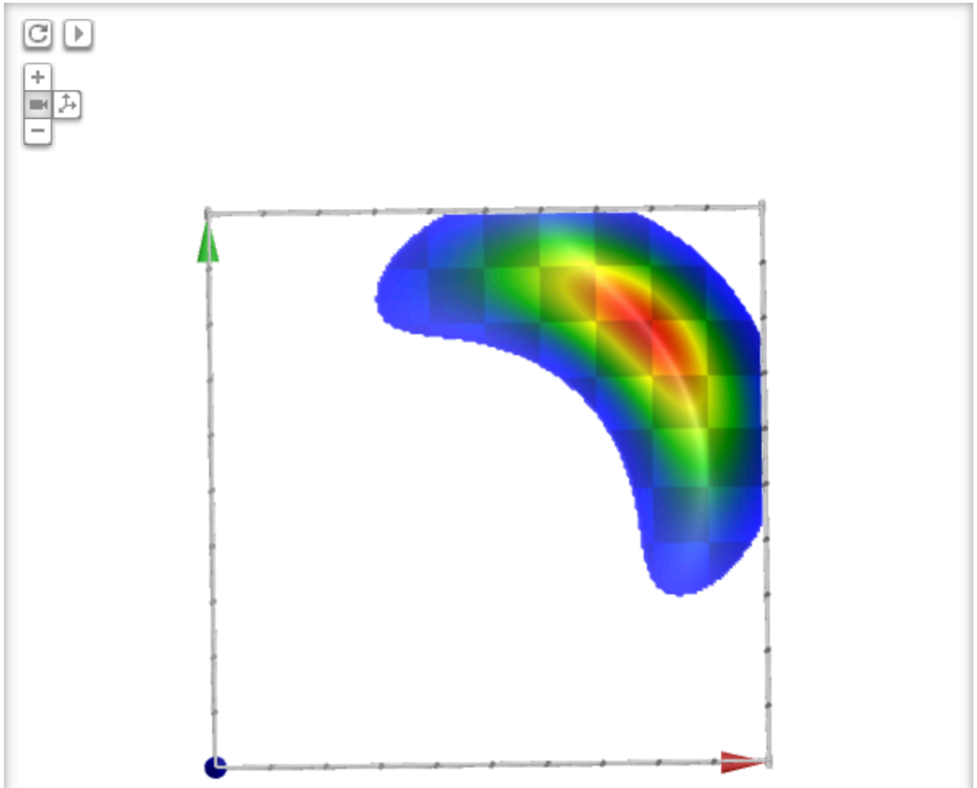
$$f_{D|X,Y}(d|x,y) = K \cdot e^{-[d-\sqrt{x^2+y^2}]^2}$$



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

Tracking in 2D Space: Posterior

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



	From	To
x	-5.00000	5.00000
y	-5.00000	5.00000
z	0.0454545	1.00000