Joint Distributions

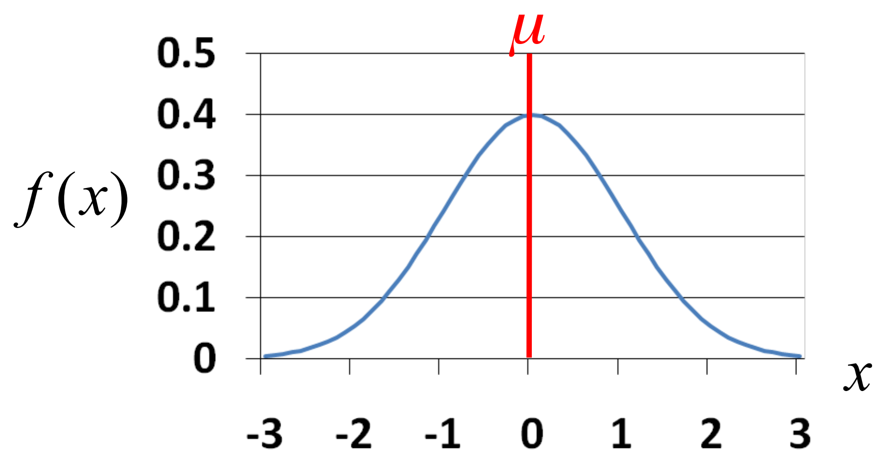
Chris Piech
CS109, Stanford University

Review

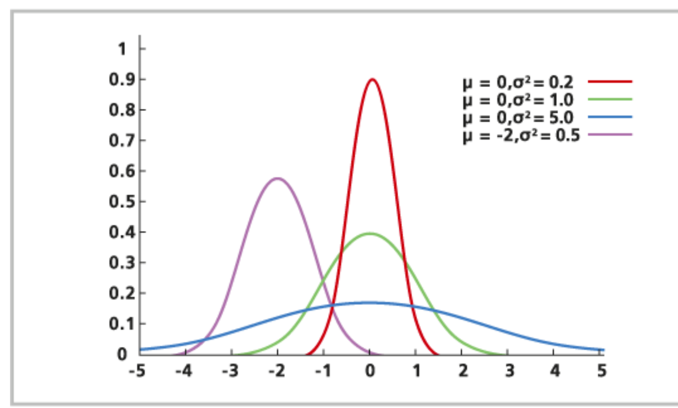
The Normal Distribution

- X is a **Normal Random Variable**: $X \sim N(\mu, \sigma^2)$
  - Probability Density Function (PDF):

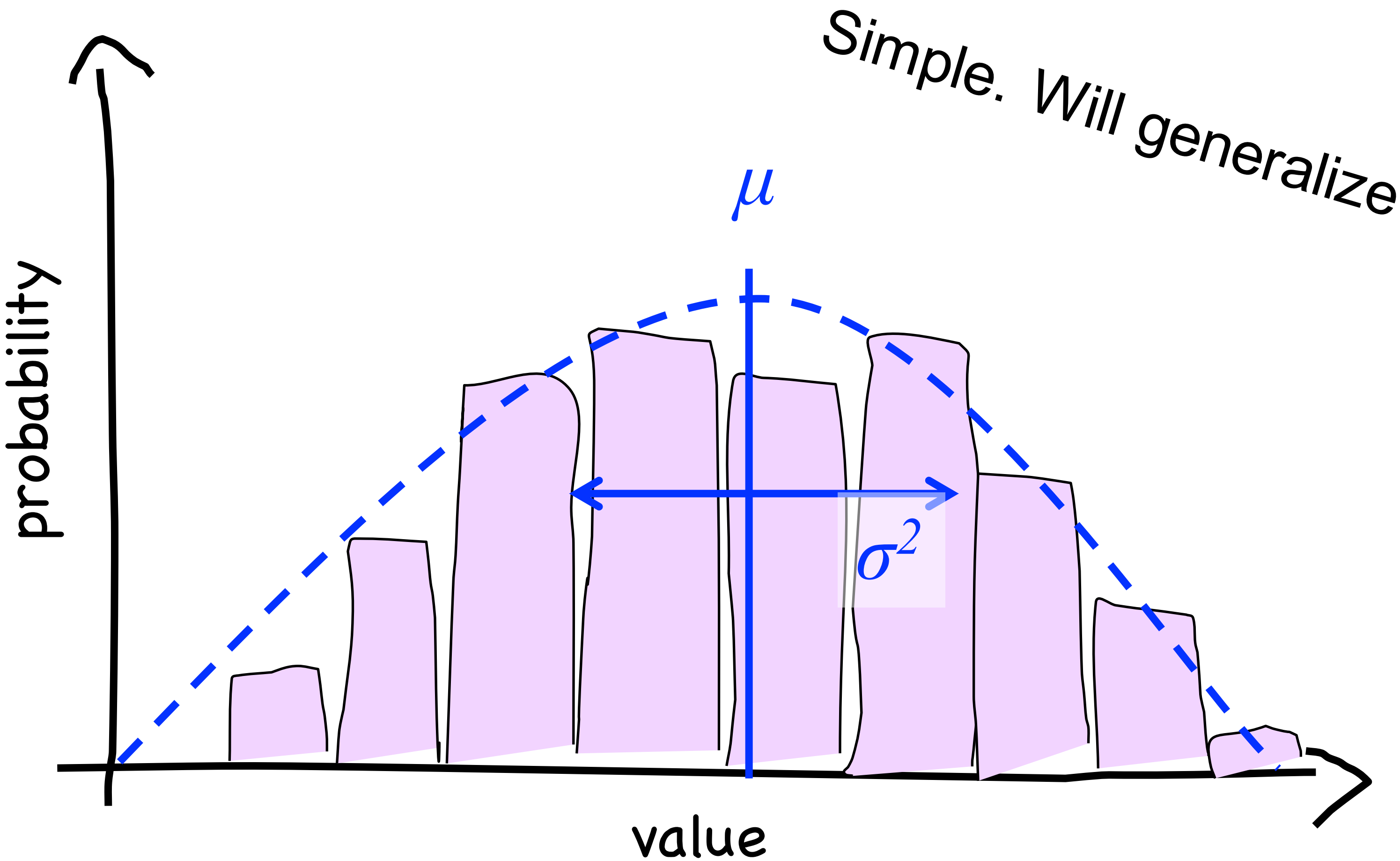
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{where} \quad -\infty < x < \infty$$



  - $E[X] = \mu$
  - $Var(X) = \sigma^2$
  - Also called "Gaussian"
  - Note: $f(x)$ is symmetric about μ



Simplicity is Humble



* A Gaussian maximizes entropy for a given mean and variance

Anatomy of a beautiful equation

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

"exponential"

the distance to the mean

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

probability density at x

a constant

sigma shows up twice

And here we are

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

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

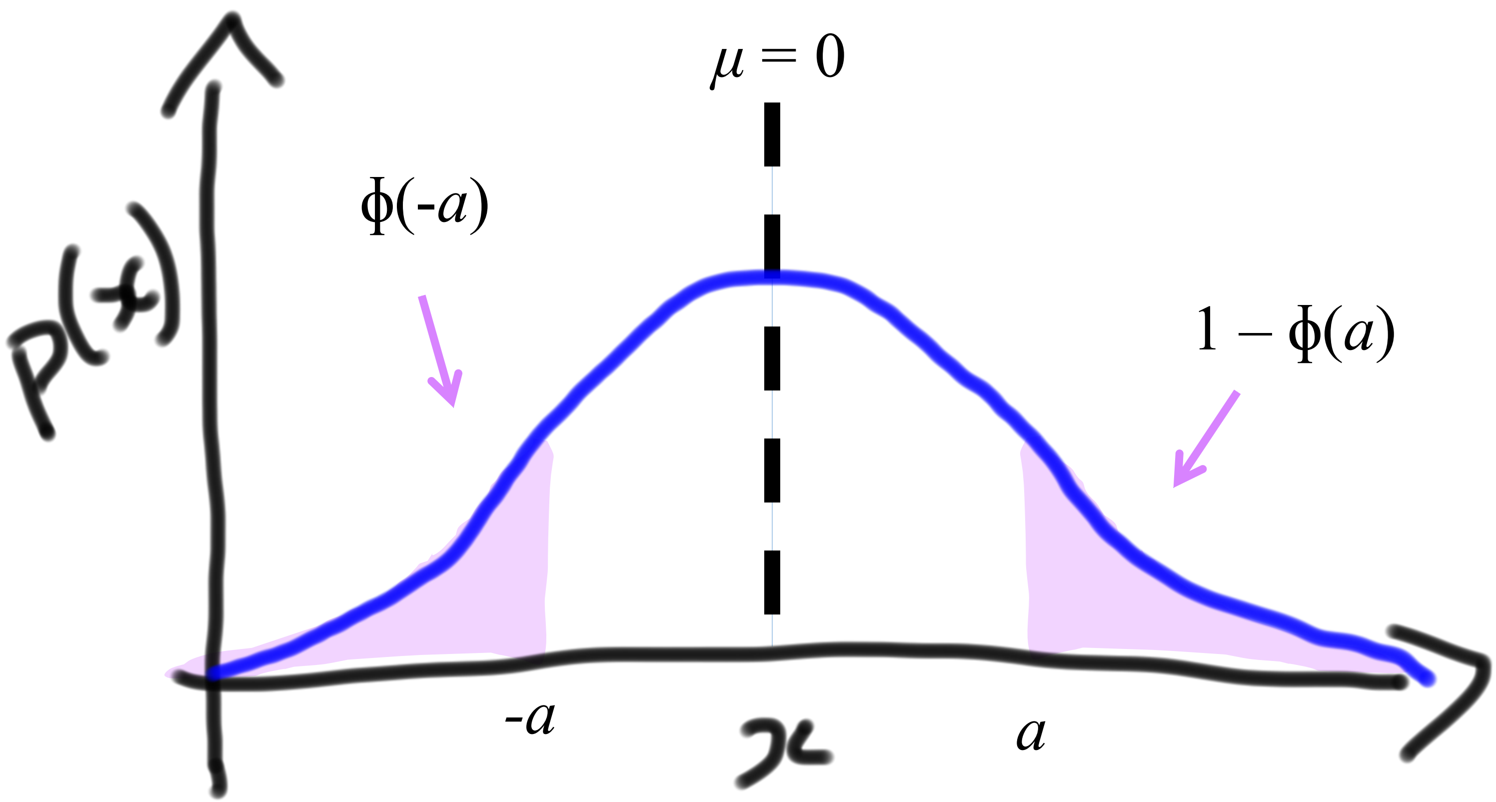
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

The cumulative density function (CDF) of any normal

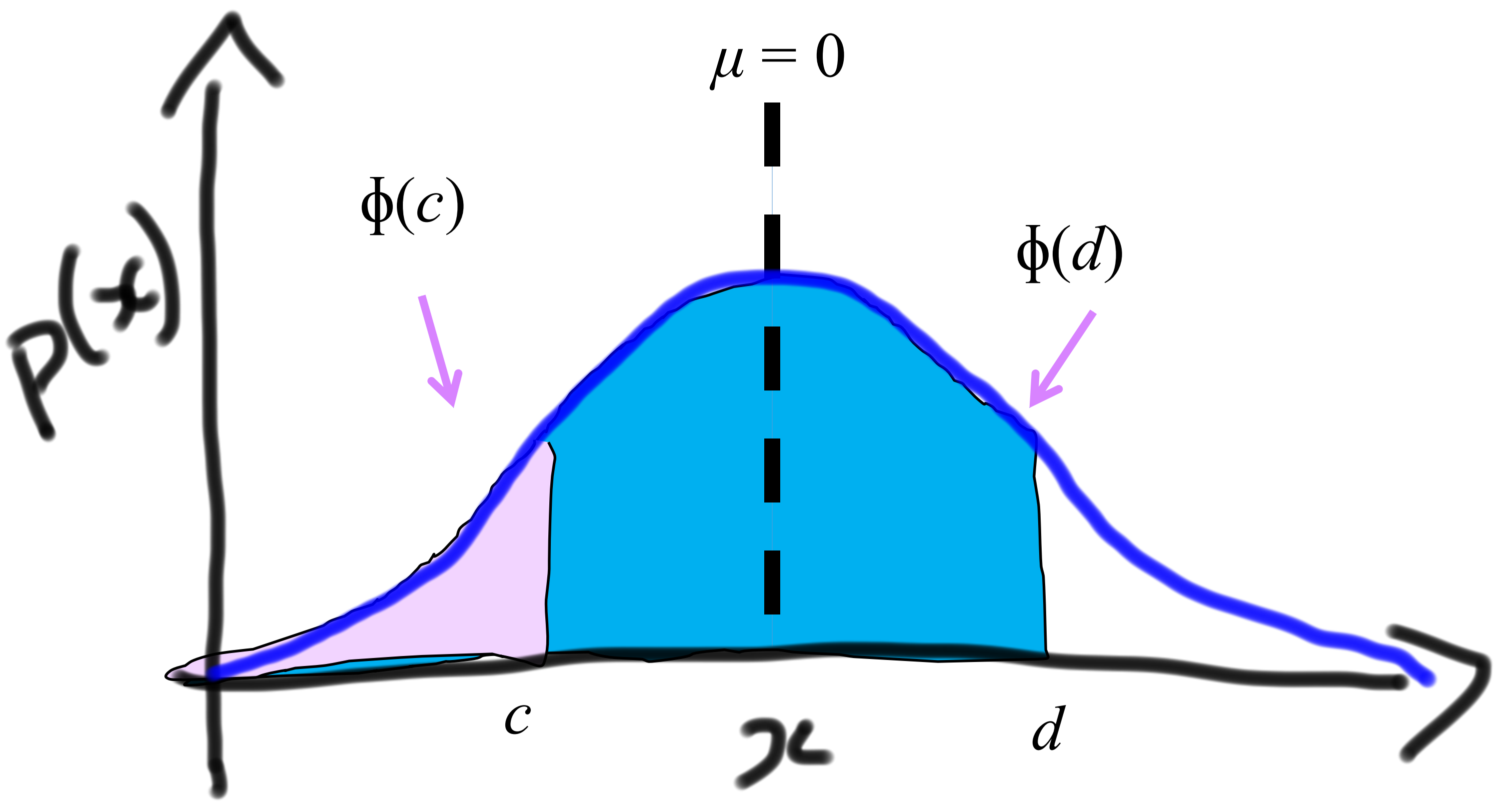
Table of $\Phi(z)$ values in textbook, p. 201 and handout

Symmetry of Phi

$$\phi(-a) = 1 - \phi(a)$$

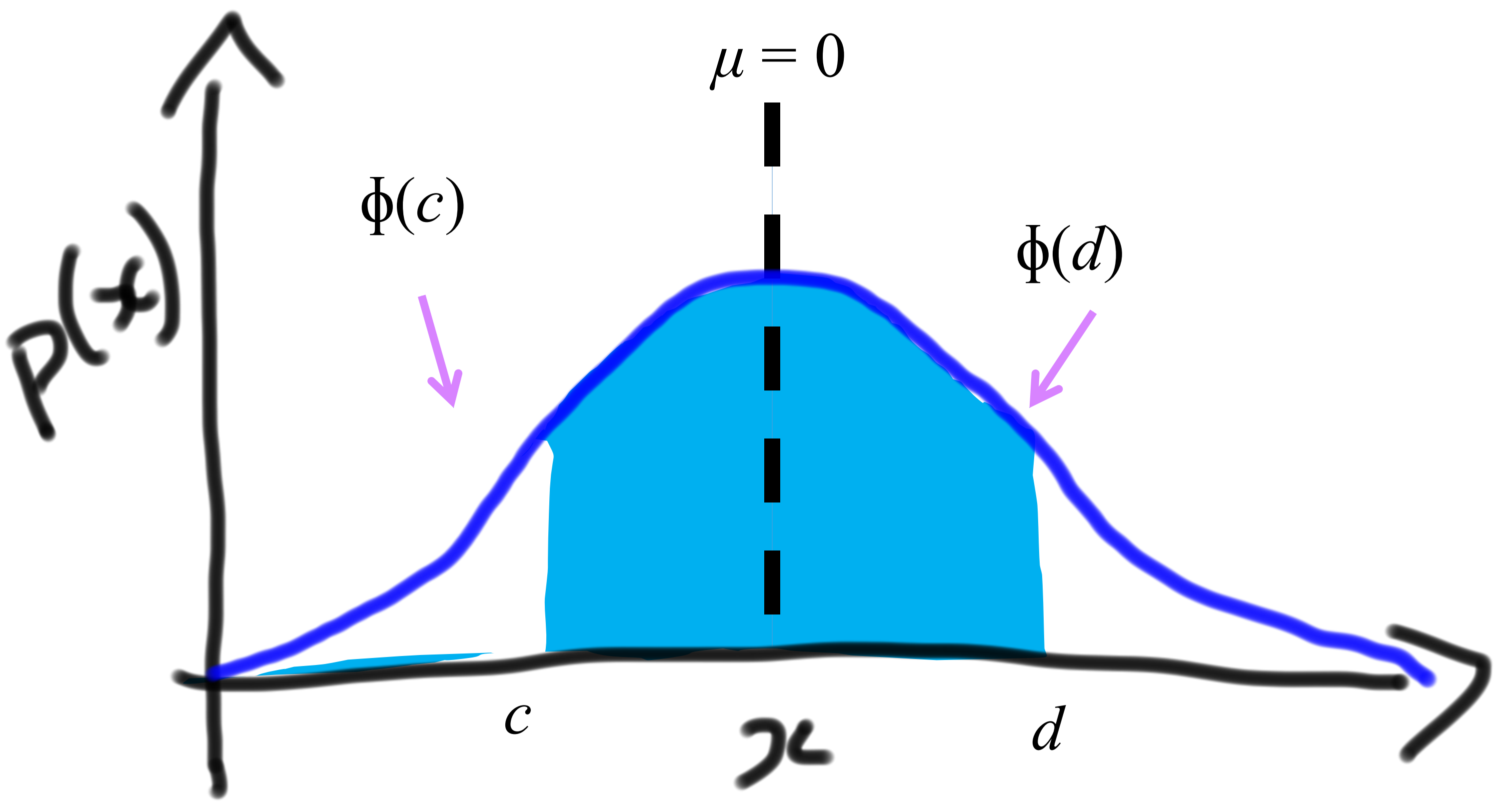


Interval of Phi



Interval of Phi

$$\phi(d) - \phi(c)$$



Great questions!

68% rule only for Gaussians?

68% Rule?

What is the probability that a normal variable $X \sim N(\mu, \sigma^2)$ has a value within one standard deviation of its mean?

$$P(\mu - \sigma < X < \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + \sigma - \mu}{\sigma}\right)$$

$$= P(-1 < Z < 1)$$

$$= \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - [1 - \Phi(1)]$$

$$= 2\Phi(1) - 1$$

$$= 2[0.8413] - 1 = 0.683$$

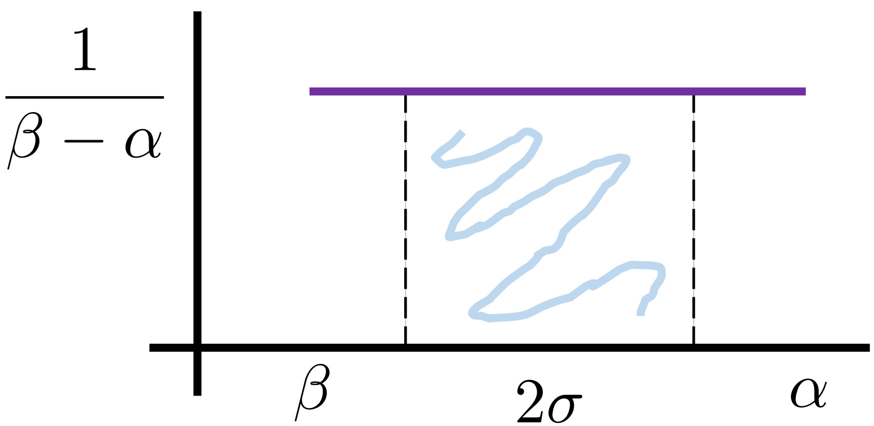
Only applies to normal

68% Rule?

Counter example: Uniform $X \sim Uni(\alpha, \beta)$

$$Var(X) = \frac{(\beta - \alpha)^2}{12}$$

$$\sigma = \sqrt{Var(X)} = \frac{\beta - \alpha}{\sqrt{12}}$$



$$P(\mu - \sigma < X < \mu + \sigma) = \frac{1}{\beta - \alpha}\left[\frac{2(\beta - \alpha)}{\sqrt{12}}\right] = \frac{2}{\sqrt{12}} = 0.58$$

How does python sample from a Gaussian?

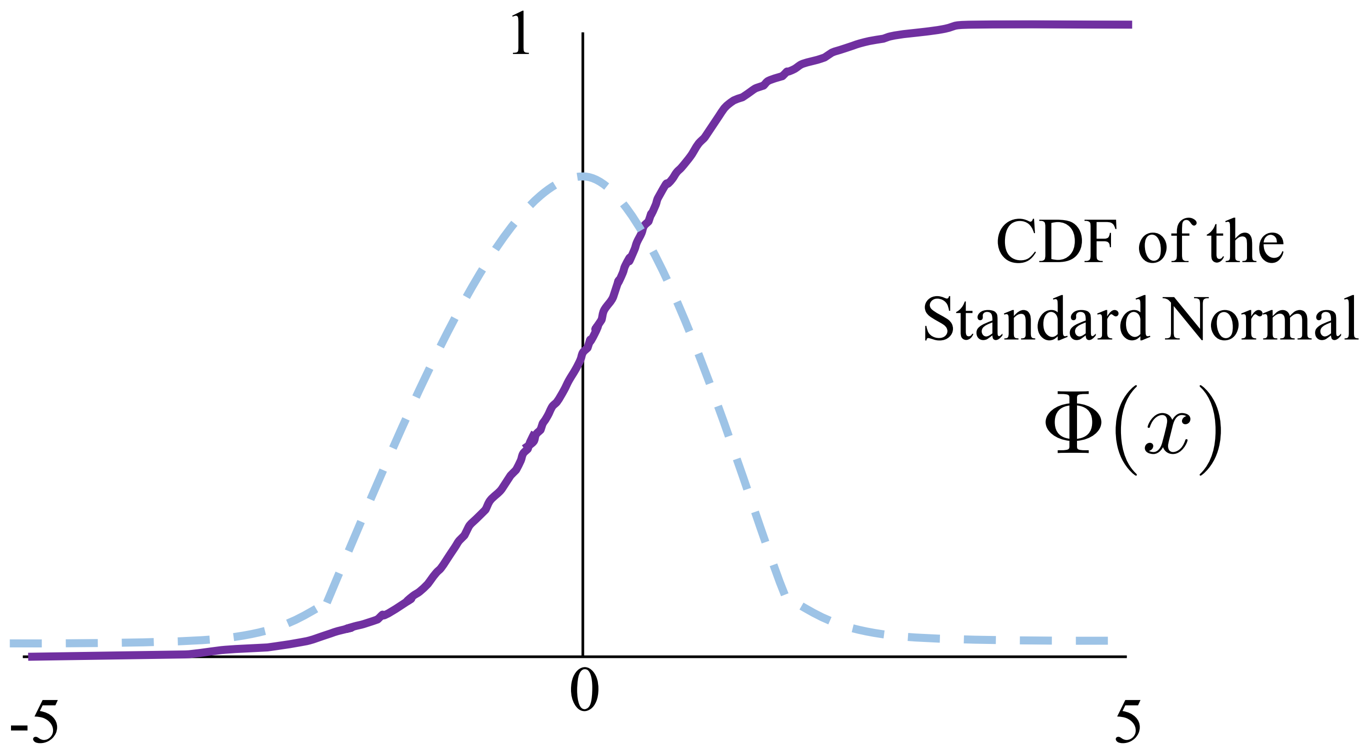
```python
from random import *

for i in range(10):
    mean = 5
    std = 1
    sample = gauss(mean, std)
    print sample
```

How does this work?

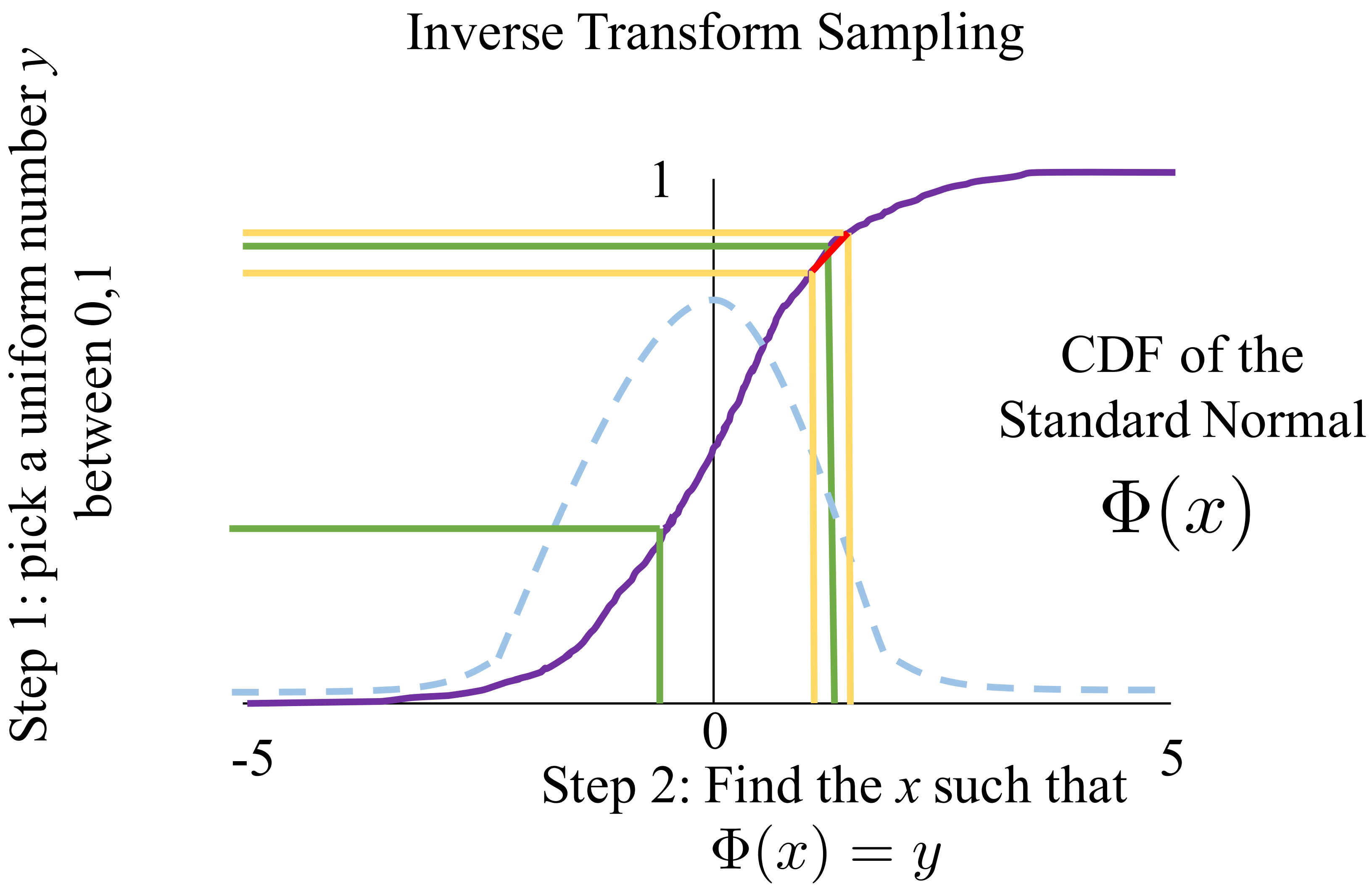
```
3.79317794179
5.19104589315
4.209360629
5.39633891584
7.10044176511
6.72655475942
5.51485158841
4.94570606131
6.14724644482
4.73774184354
```

How Does a Computer Sample Normal?

Inverse Transform Sampling



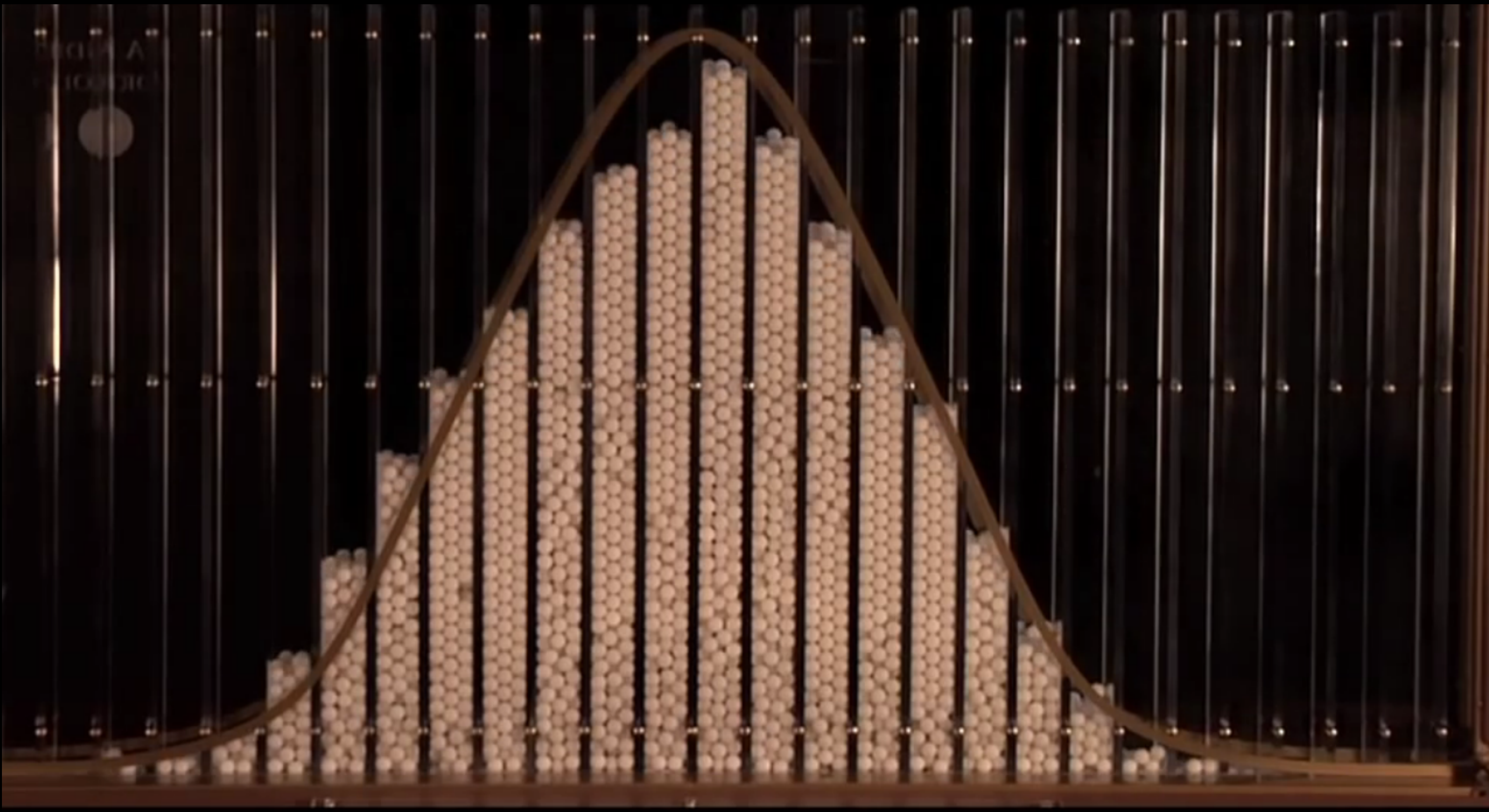
How Does a Computer Sample Normal?



Further reading: Box–Muller transform

End Review

Normal Approximates Binomial

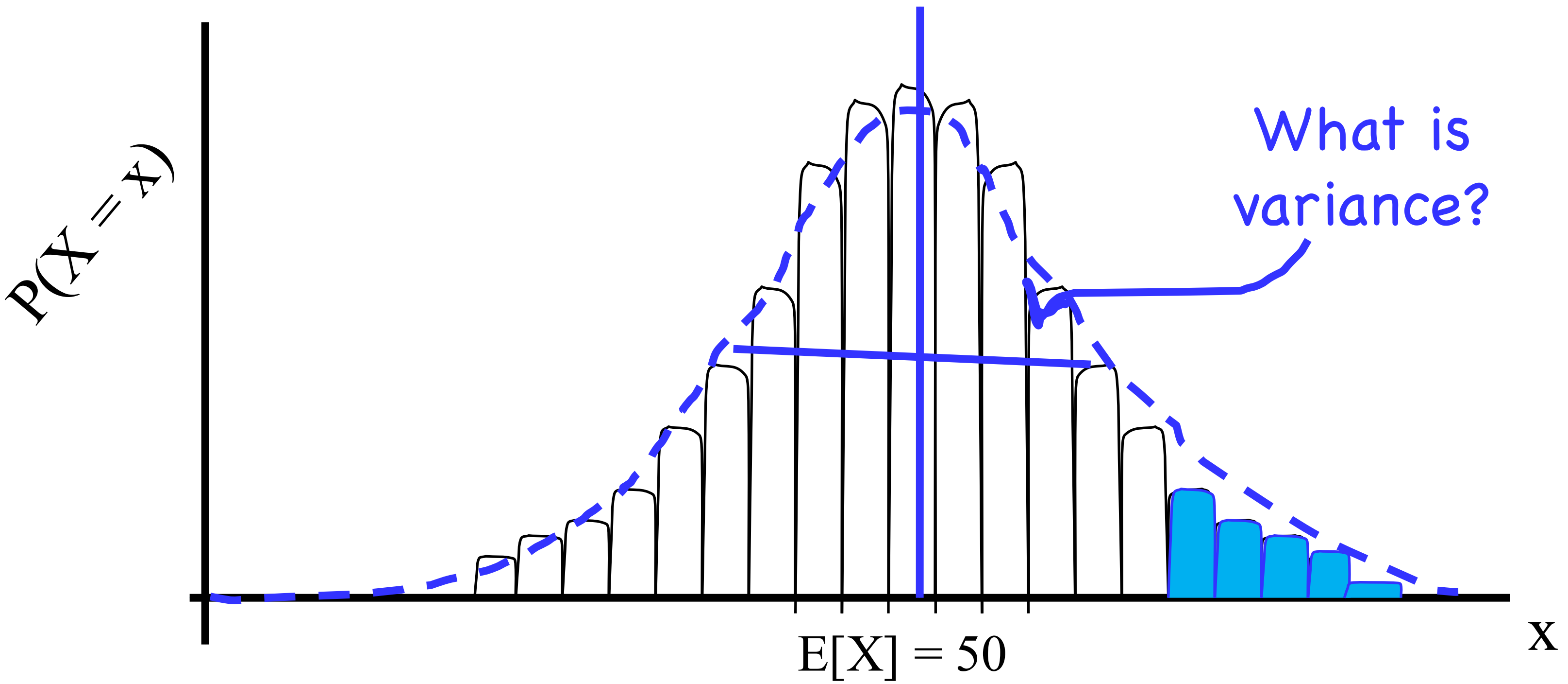


There is a deep reason for the Binomial/Normal similarity…

Let’s invent it

Website Testing

- 100 people are given a new website design
  - X = # people whose time on site increases
  - CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
  - X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)



Website Testing

- 100 people are given a new website design
  - X = # people whose time on site increases
  - CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
  - X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)

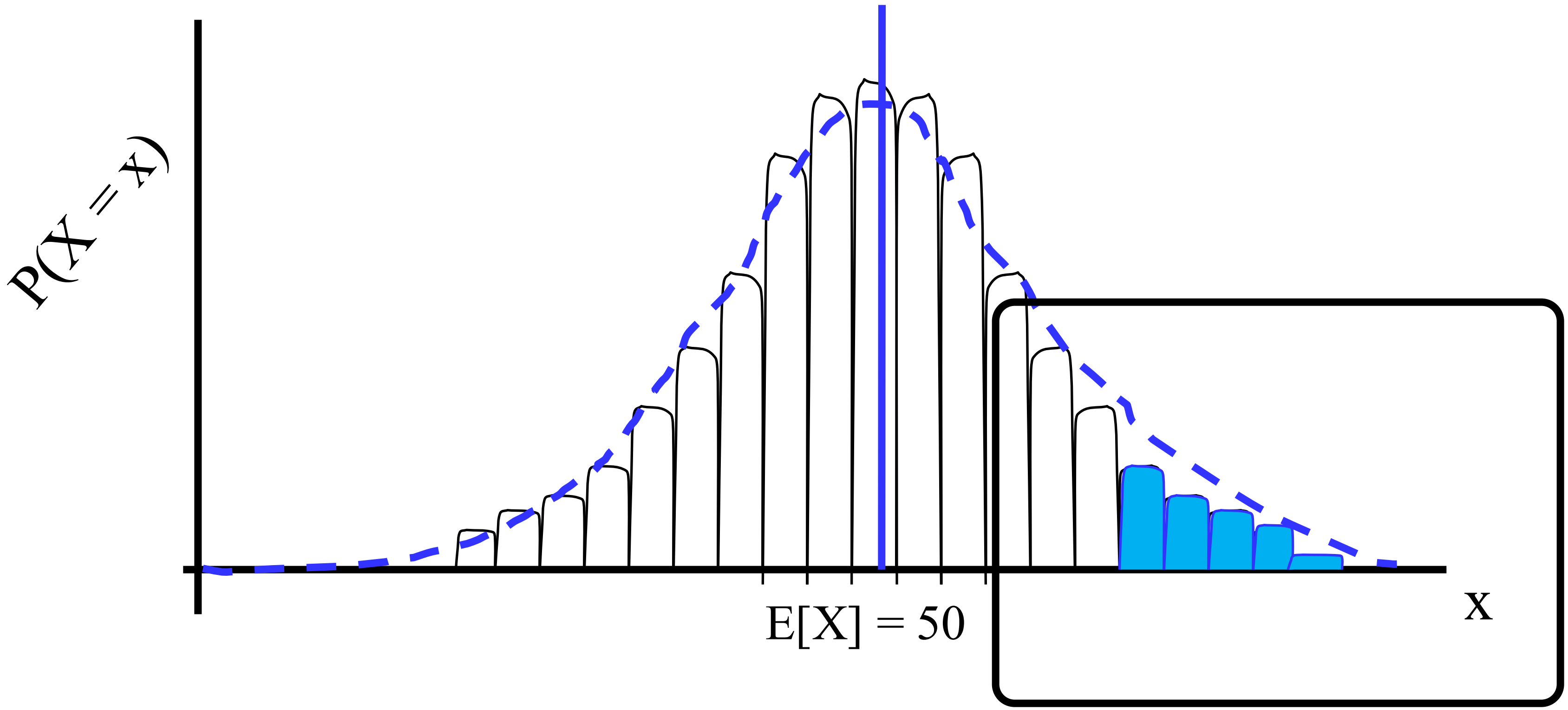
$$np = 50 \quad np(1-p) = 25 \quad \sqrt{np(1-p)} = 5$$

  - Use Normal approximation: Y ~ N(50, 25)

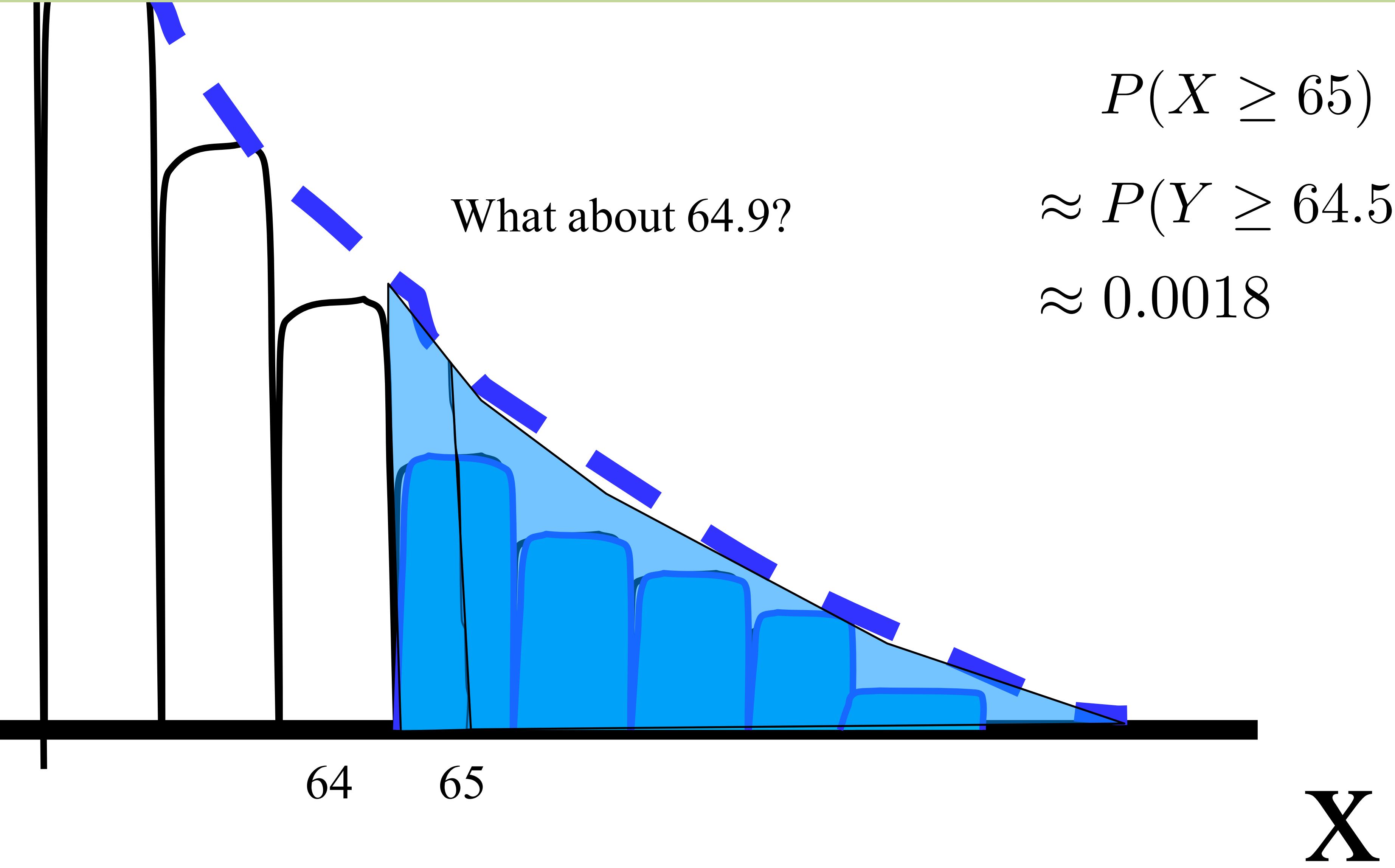
$$P(Y \geq 65) = P\left(\frac{Y-50}{5} > \frac{65-50}{5}\right) = P(Z > 3) = 1 - \phi(3) \approx 0.002$$

  - Using Binomial: $P(X \geq 65) \approx 0.0018$

Website Testing

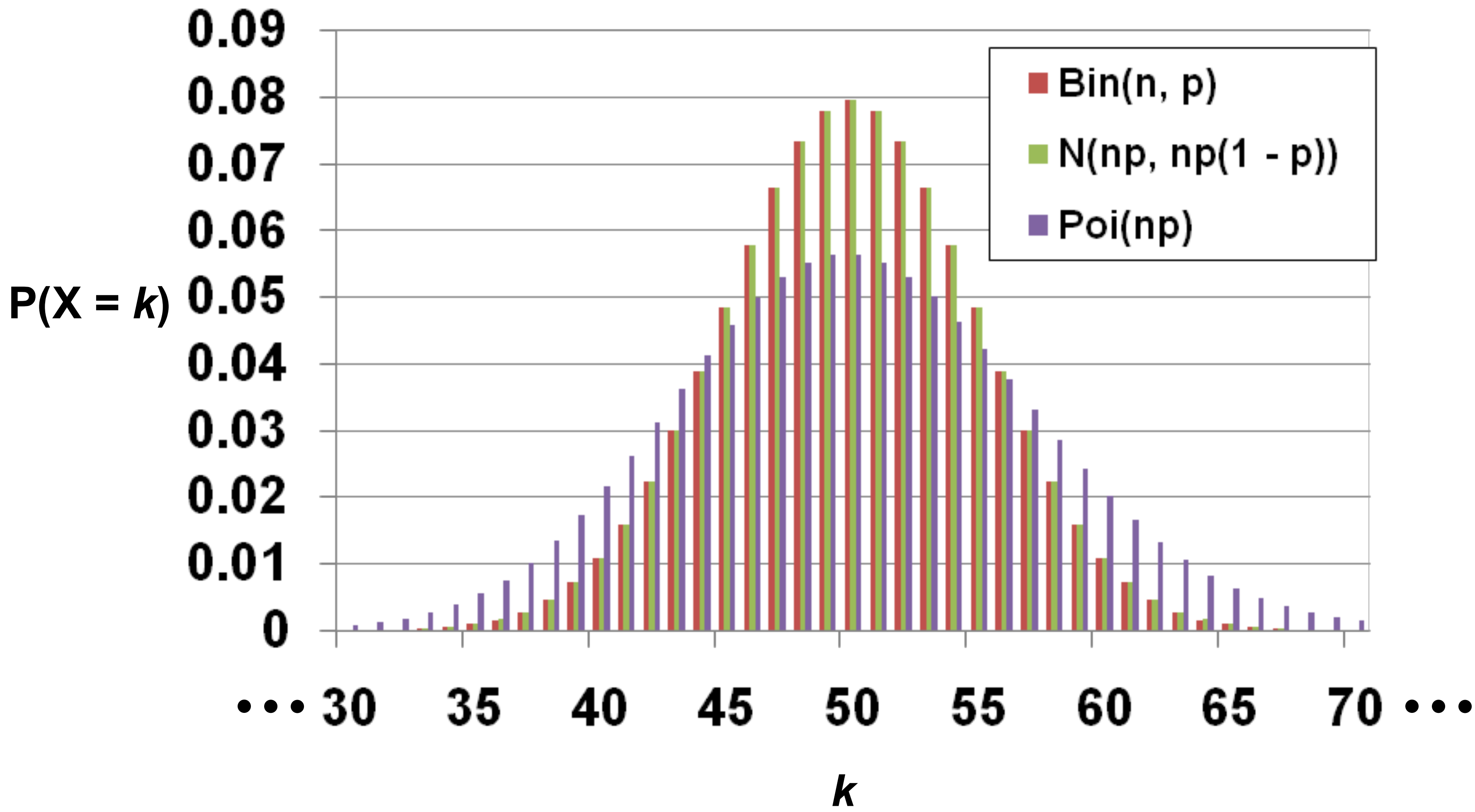


Continuity Correction



$$P(X \geq 65)$$
$$\approx P(Y \geq 64.5)$$
$$\approx 0.0018$$

Comparison when *n* = 100, *p* = 0.5



Normal Approximation of Binomial

- X ~ Bin(*n*, *p*)
  - E[X] = *np*   Var(X) = *np*(1 – *p*)
  - Poisson approx. good: *n* large (> 20), *p* small (< 0.05)
  - For large *n:* X ≈ Y ~ N(E[X], Var(X)) = N(*np*, *np*(1 – *p*))
  - Normal approx. good : Var(X) = *np*(1 – *p*) ≥ 10

$$P(X = k) \approx P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

"Continuity correction"

Continuity Correction

Discrete (eg Binomial) probability question	Continuous (Normal) probability question
x = 6	5.5 < x < 6.5
x >= 6	x > 5.5
x > 6	x > 6.5
x < 6	x < 5.5
x <= 6	x < 6.5

Stanford Admissions

- Stanford accepts 2050 students this year
  - Each accepted student has 84% chance of attending
  - X = # students who will attend.   X ~ Bin(2050, 0.84)
  - What is P(X > 1745)?

$$np = 1722 \qquad np(1-p) = 276 \qquad \sqrt{np(1-p)} = 16.6$$

  - Use Normal approximation: Y ~ N(1722, 276)

$$P(X > 1745) \approx P(Y > 1745.5)$$

$$P(Y \geq 1745.5) = P\left(\frac{Y - 1722}{16.6} > \frac{1745.5 - 1722}{16.6}\right) = P(Z > 1.4)$$

$$\approx 0.08$$

Changes in Stanford Admissions

Class of 2020 Admit Rates Lowest in University History

"Fewer students were admitted to the Class of 2020 than the Class of 2019, due to the increase in Stanford's yield rate which has increased over 5 percent in the past four years, according to Colleen Lim M.A. '80, Director of Undergraduate Admission."

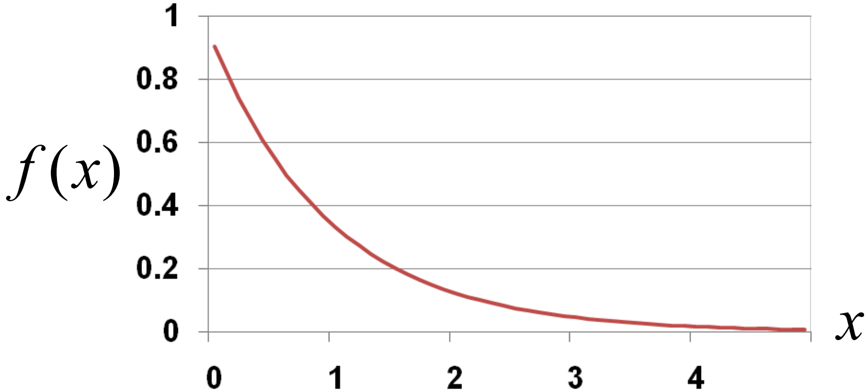
68% 10 years ago 84% last year

More practice with continuous random variables: Exponential

Exponential Random Variable

- X is an **<u>Exponential RV</u>**: $X \sim \text{Exp}(\lambda)$ Rate: $\lambda > 0$
  - Probability Density Function (PDF):

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad \text{where} \quad -\infty < x < \infty$$



  - $E[X] = \frac{1}{\lambda}$
  - $Var(X) = \frac{1}{\lambda^2}$
  - Cumulative distribution function (CDF), F(X) = P($X \leq x$):

$$F(x) = 1 - e^{-\lambda x} \quad \text{where } x \geq 0$$

  - Represents time until some event
    - Earthquake, request to web server, end cell phone contract, etc.

Visits to a Website

- Say visitor to your web site leaves after X minutes
  - On average, visitors leave site after 5 minutes
  - Assume length of stay is Exponentially distributed
  - X ~ Exp(λ = 1/5), since E[X] = 1/λ = 5
  - What is P(X > 10)?

$$P(X > 10) = 1 - F(10) = 1 - (1 - e^{-\lambda 10}) = e^{-2} \approx 0.1353$$

  - What is P(10 < X < 20)?

$$P(10 < X < 20) = F(20) - F(10) = (1 - e^{-4}) - (1 - e^{-2}) \approx 0.1170$$

Replacing Your Laptop

- X = # hours of use until your laptop dies
  - On average, laptops die after 5000 hours of use
  - X ~ Exp(λ = 1/5000), since E[X] = 1/λ = 5000
  - You use your laptop 5 hours/day.
  - What is P(your laptop lasts 4 years)?
  - That is: P(X > (5)(365)(4) = 7300)

$$P(X > 7300) = 1 - F(7300) = 1 - (1 - e^{-7300/5000}) = e^{-1.46} \approx 0.2322$$

  - Better plan ahead... especially if you are coterming:

$$P(X > 9125) = 1 - F(9125) = e^{-1.825} \approx 0.1612 \quad \text{(5 year plan)}$$

$$P(X > 10950) = 1 - F(10950) = e^{-2.19} \approx 0.1119 \quad \text{(6 year plan)}$$

Exponential is Memoryless

- X = time until some event occurs
  - X ~ Exp(λ)
  - What is P(X > s + t | X > s)?

$$P(X > s + t \mid X > s) = \frac{P(X > s + t \text{ and } X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}$$

$$\frac{P(X > s + t)}{P(X > s)} = \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P(X > t)$$

$$\text{So, } P(X > s + t \mid X > s) = P(X > t)$$

  - After initial period of time *s*, P(X > t | •) for waiting another *t* units of time until event is same as at start
  - "Memoryless" = no impact from preceding period *s*

Continuous Random Variables

Uniform Random Variable $X \sim Uni(\alpha, \beta)$

All values of *x* between alpha and beta are equally likely.

Normal Random Variable $X \sim \mathcal{N}(\mu, \sigma^2)$

Aka Gaussian. Defined by mean and variance. Goldilocks distribution.

Exponential Random Variable $X \sim Exp(\lambda)$

Time until an event happens. Parameterized by lambda (same as Poisson).

Alpha Beta Random Variable

How mysterious and curious. You must wait a few classes ☺.

Joint Distributions

Events occur with other events

Probability Table

- States all possible outcomes with several discrete variables
- Often 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

Remember "," means "and"

It's Complicated Demo

Go to this URL: **https://goo.gl/jCMY18**

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 C computers: C = X Macs + Y PCs
  - Assume each computer equally likely to be Mac or PC

$$P(C=c)=\begin{cases}0.16 & c=0\\ 0.24 & c=1\\ 0.28 & c=2\\ 0.32 & c=3\end{cases}$$

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 C computers: C = X Macs + Y PCs
  - Assume each computer equally likely to be Mac or PC

$$P(C=c)=\begin{cases}0.16 & c=0\\ 0.24 & c=1\\ 0.28 & c=2\\ 0.32 & c=3\end{cases}$$

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 C computers: C = X Macs + Y PCs
  - Assume each computer equally likely to be Mac or PC

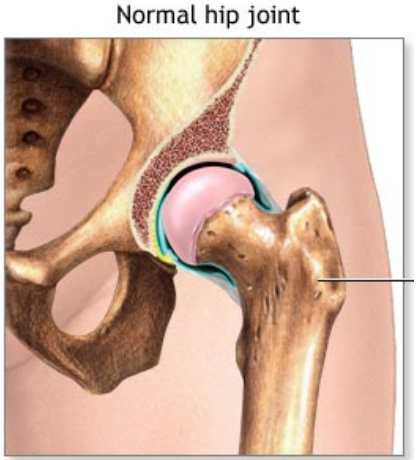
$$P(C=c)=\begin{cases}0.16 & c=0\\ 0.24 & c=1\\ 0.28 & c=2\\ 0.32 & c=3\end{cases}$$

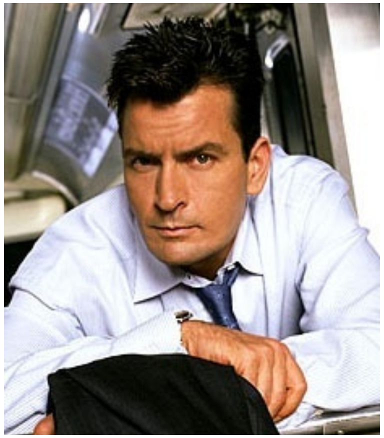
Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	0.07	0.04	0.39
1	0.12	0.14	0.12	0	0.38
2	0.07	0.12	0	0	0.19
3	0.04	0	0	0	0.04
$p_X(x)$	0.39	0.38	0.19	0.04	1.00

Marginal distributions

Joint

- This is a joint





- A joint is not a mathematician
  - It did not start doing mathematics at an early age
  - It is not the reason we have “joint distributions”
  - And, no, Charlie Sheen does not look like a joint
    - But he does have them…
    - He also has **joint** custody of his children with Denise Richards

What about the continuous world?

Jointly Continuous

- Random variables *X* and *Y*, are **Jointly Continuous** if there exists PDF $f_{X,Y}(x, y)$ defined over $-\infty < x, y < \infty$ such that:

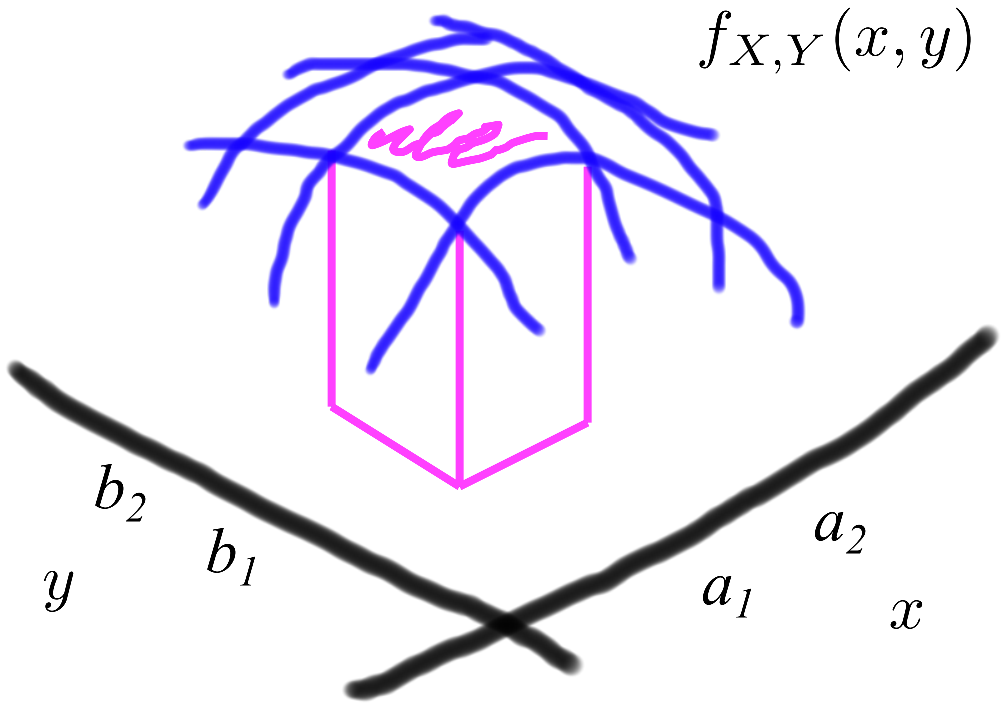
$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y)\, dy\, dx$$

Let's look at one:

Demo

Jointly Continuous

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



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):

$$F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx \qquad f_{X,Y}(a,b) = \frac{\partial^2}{\partial a\, \partial b} F_{X,Y}(a,b)$$

- Marginal density functions:

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

Continuous Joint Distribution Functions

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

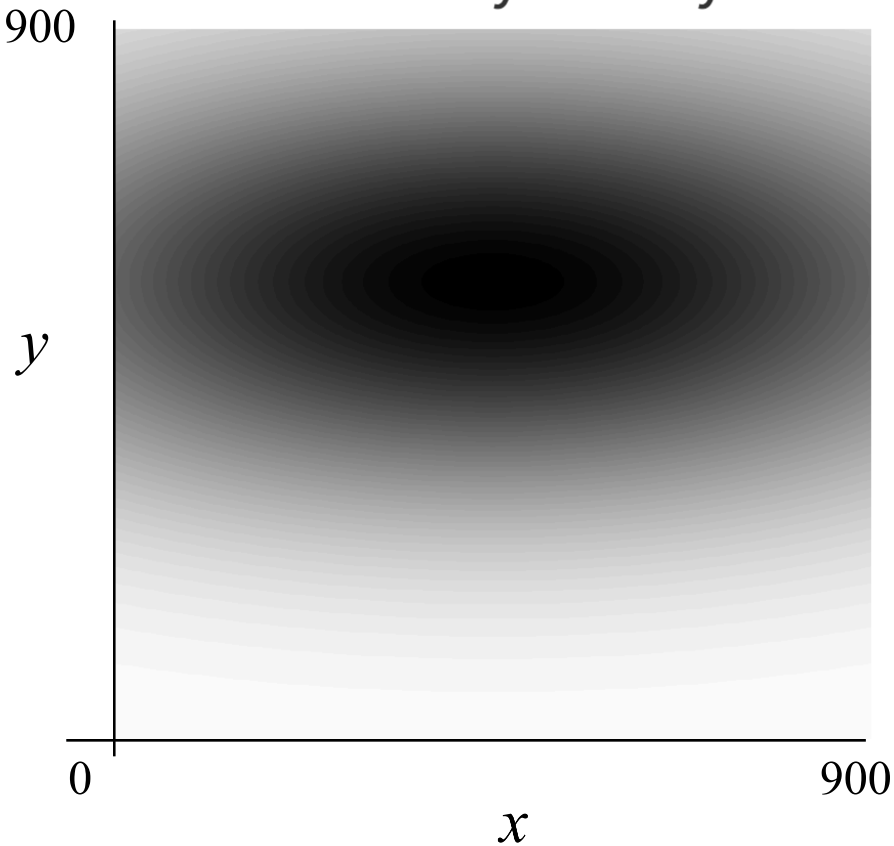
- Marginal distributions:

$$F_X(a) = P(X \le a) = \mathrm{P}(X \le a, Y < \infty) = F_{X,Y}(a,\infty)$$

$$F_Y(b) = P(Y \le b) = \mathrm{P}(X < \infty, Y \le b) = F_{X,Y}(\infty,b)$$

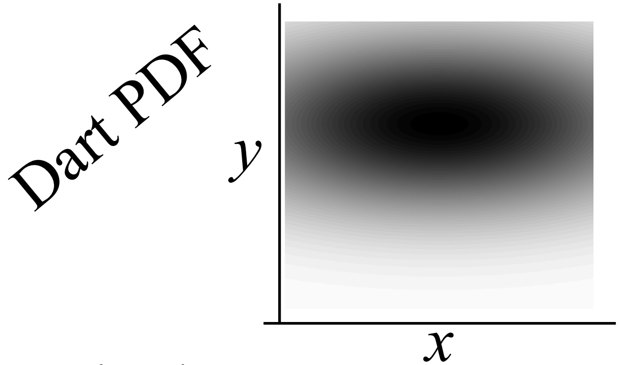
Joint Dart Distribution

Dart Probability Density

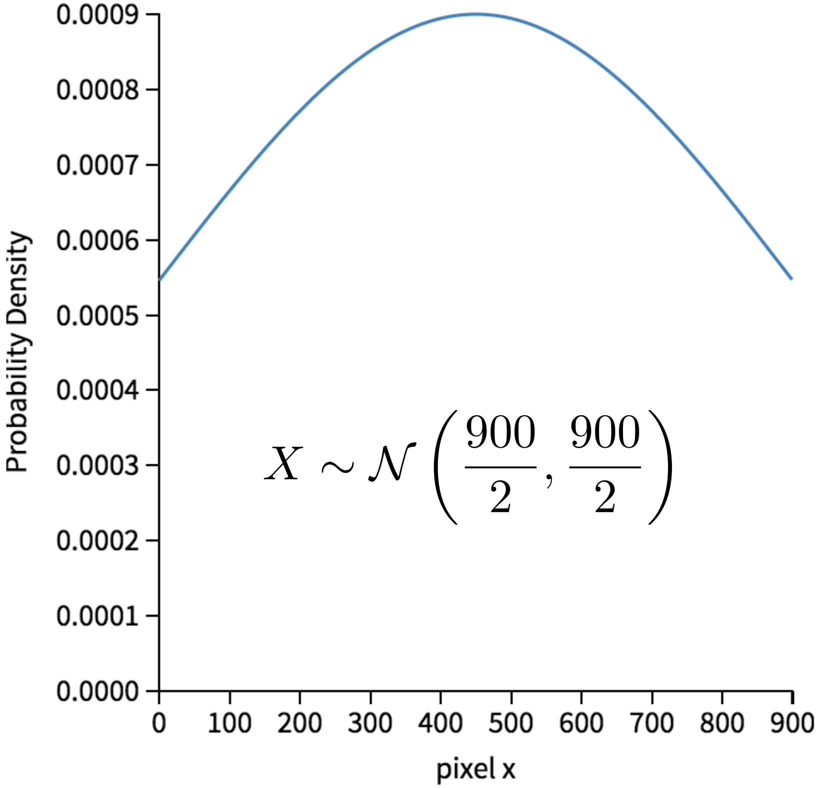


Dart Results

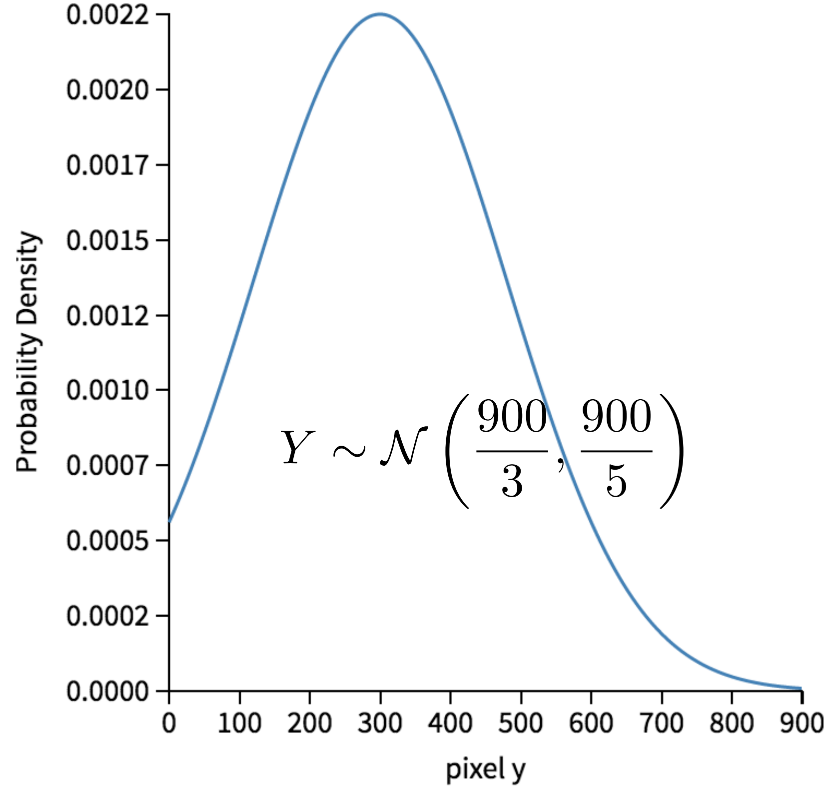
Darts!



X-Pixel Marginal



Y-Pixel Marginal



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:

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

Computing Joint Probabilities

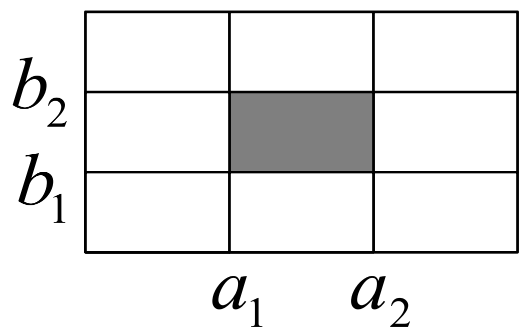
Let $F_{X,Y}(x, y)$ be joint CDF for X and Y

$$
\begin{aligned}
\mathrm{P}(X > a, Y > b) &= 1 - P((X > a, Y > b)^c) \\
&= 1 - P((X > a)^c \cup (Y > b)^c) \\
&= 1 - P((X \le a) \cup (Y \le b)) \\
&= 1 - (P(X \le a) + P(Y \le b) - P(X \le a, Y \le b)) \\
&= 1 - F_X(a) - F_Y(b) + F_{X,Y}(a,b)
\end{aligned}
$$

The General Rule Given Joint CDF

Let $F_{X,Y}(x,y)$ be joint CDF for X and Y

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



Lovely Lemma

- Y is a <u>non-negative</u> continuous random variable
  - Probability Density Function: $f_Y(y)$
  - Already knew that:

$$E[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$$

  - But, did you know that:

$$E[Y] = \int_{0}^{\infty} P(Y > y)\, dy \text{ ?!?}$$

  - Analogously, in the discrete case, where $X = 1, 2, \ldots, n$

$$E[X] = \sum_{i=1}^{n} P(X \geq i)$$

How this lemma was made

In the discrete case, where $X = 1, 2, \ldots, n$

$$E[X] = \sum_{i=1}^{n} P(X \geq i)$$

$$\sum_{i=1}^{n} P(X \geq i) =$$

Each row is an expansion for one value of i

$$\begin{aligned}
P(X = 1) + P(X = 2) + P(X = 3) + \cdots + P(X = n)& \\
+ P(X = 2) + P(X = 3) + \cdots + P(X = n)& \\
+ P(X = 3) + \cdots + P(X = n)& \\
\vdots\quad& \\
+ P(X = n)&
\end{aligned}$$

$$\begin{aligned}
&= 1P(X = 1) + 2P(X = 2) + \cdots + n(PX = n) \\
&= E[X]
\end{aligned}$$

Life gives you lemmas,
make lemmanade!

Imperfections on a Disk

- Disk surface is a circle of radius R
  - A single point imperfection uniformly distributed on disk

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi R^2} & \text{if } x^2 + y^2 \le R^2 \\ 0 & \text{if } x^2 + y^2 > R^2 \end{cases} \quad \text{where} \quad -\infty < x, y < \infty$$

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

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

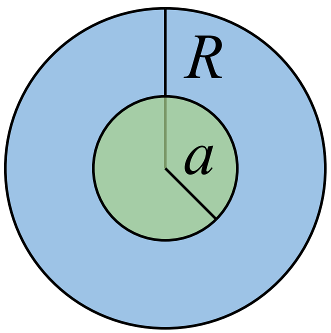
$$= \frac{2\sqrt{R^2 - x^2}}{\pi R^2}$$

Only integrate over the support range

Marginal of Y is the same by symmetry

Imperfections on a Disk

- Disk surface is a circle of radius R
  - A single point imperfection uniformly distributed on disk
  - Distance to origin: $D = \sqrt{X^2 + Y^2}$
  - What is *E[D]*?



$$P(D \le a) = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}$$

Because of equally likely outcomes

$$E[D] = \int_0^R P(D > a)da = \int_0^R 1 - P(D \le a)da$$

$$= \int_0^R 1 - \frac{a^2}{R^2}da$$

$$= \left[a - \frac{a^3}{3R^2}\right]_0^R = \frac{2R}{3}$$