The Normal Distribution

image: Etsy

Will Monroe
July 19, 2017

with materials by
Mehran Sahami
and Chris Piech

Announcements: Midterm

A week from yesterday:

Tuesday, July 25, 7:00-9:00pm
Building 320-105

One page (both sides) of notes

Material through **today's** lecture

Review session:
Tomorrow, July 20, 2:30-3:20pm
in **Gates B01**

Review: A grid of random variables

	number of **successes**	**time** to get successes	
One trial	$X \sim \mathrm{Ber}(p)$	$X \sim \mathrm{Geo}(p)$	One success
	↑ $n = 1$	↑ $r = 1$	
Several trials	$X \sim \mathrm{Bin}(n, p)$	$X \sim \mathrm{NegBin}(r, p)$	Several successes
Interval of time	$X \sim \mathrm{Poi}(\lambda)$	$X \sim \mathrm{Exp}(\lambda)$ (continuous!)	One success after interval of time

Review: Continuous distributions

A **continuous** random variable has a value that's a **real number** (not necessarily an integer).

Replace sums with **integrals**!

$$P(a < X \le b) = F_X(b) - F_X(a)$$

$$F_X(a) = \int_{x=-\infty}^{a} dx \ f_X(x)$$

Review: Probability density function

The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units of probability *divided by units of X*.
Integrate it to get probabilities!

$$P(a < X \leq b) = \int_{x=a}^{b} dx \, \boxed{f_X(x)}$$

Continuous expectation and variance

Remember: replace sums with **integrals**!

$$E[X] = \sum_{x=-\infty}^{\infty} x \cdot p_X(x) \longrightarrow E[X] = \int_{x=-\infty}^{\infty} dx\, x \cdot f_X(x)$$

$$E[X^2] = \sum_{x=-\infty}^{\infty} x^2 \cdot p_X(x) \longrightarrow E[X^2] = \int_{x=-\infty}^{\infty} dx\, x^2 \cdot f_X(x)$$

$$\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

(still!)

Review: Uniform random variable

A **uniform** random variable is **equally likely** to be any value in a single real number interval.

$$X \sim \text{Uni}(\alpha, \beta)$$

$$f_X(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

Uniform: Fact sheet

$$X \sim \text{Uni}(\alpha, \beta)$$

minimum value: α

maximum value: β

PDF: $$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

CDF: $$F_X(x) = \begin{cases} \frac{x - \alpha}{\beta - \alpha} & \text{if } x \in [\alpha, \beta] \\ 1 & \text{if } x > \beta \\ 0 & \text{otherwise} \end{cases}$$

expectation: $$E[X] = \frac{\alpha + \beta}{2}$$

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

image: Haha169

Review: Exponential random variable

An **exponential** random variable is the **amount of time until the first event** when events occur as in the Poisson distribution.

$$X \sim \text{Exp}(\lambda)$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

image: Adrian Sampson

Exponential: Fact sheet

rate of events per unit time

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

time until first event

PDF: $$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

CDF: $$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

expectation: $$E[X] = \frac{1}{\lambda}$$

variance: $$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

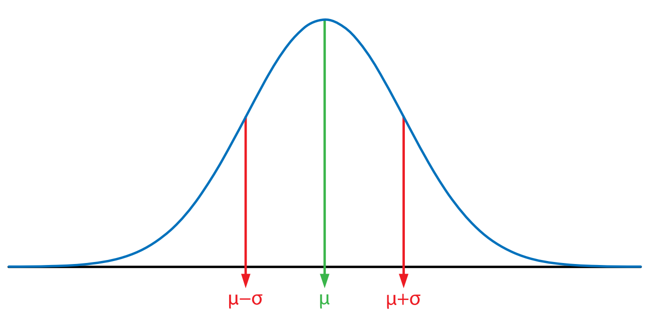
image: Adrian Sampson

Normal random variable

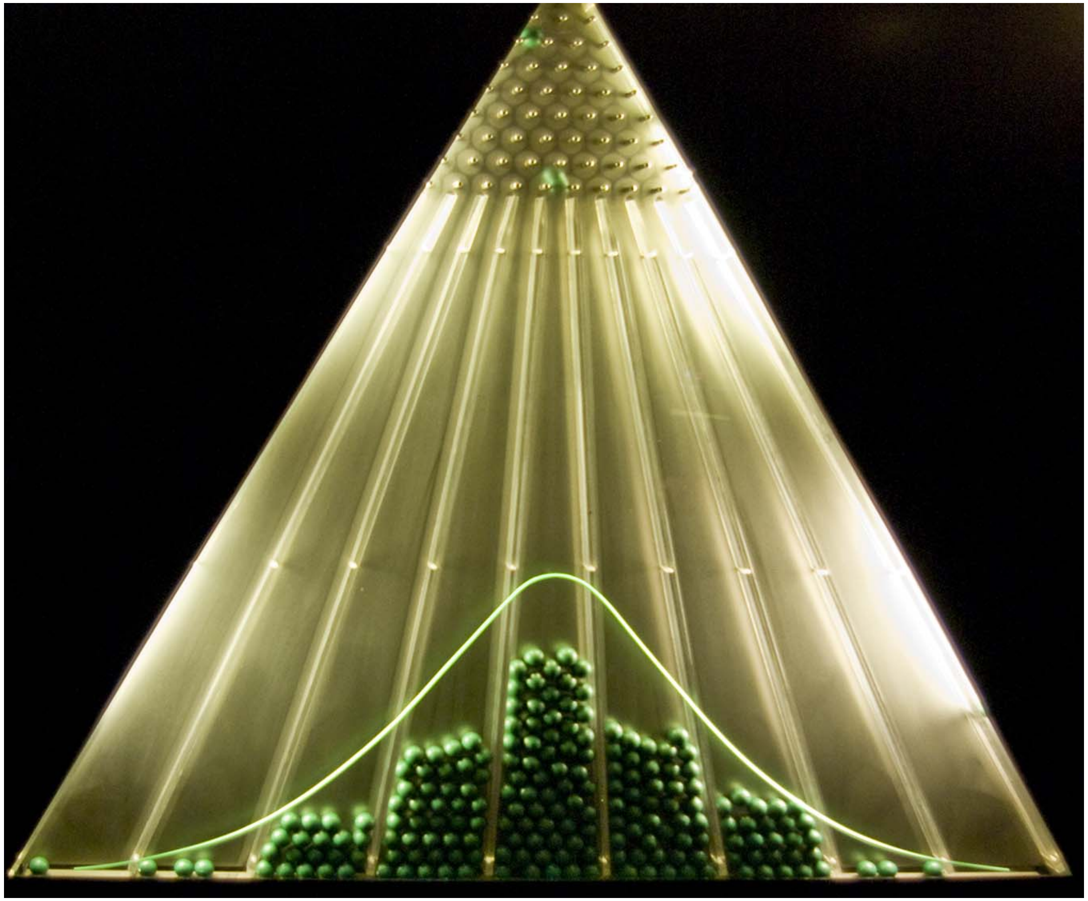
An **normal** (= **Gaussian**) random variable is a good approximation to many other distributions. It often results from **sums or averages** of independent random variables.

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

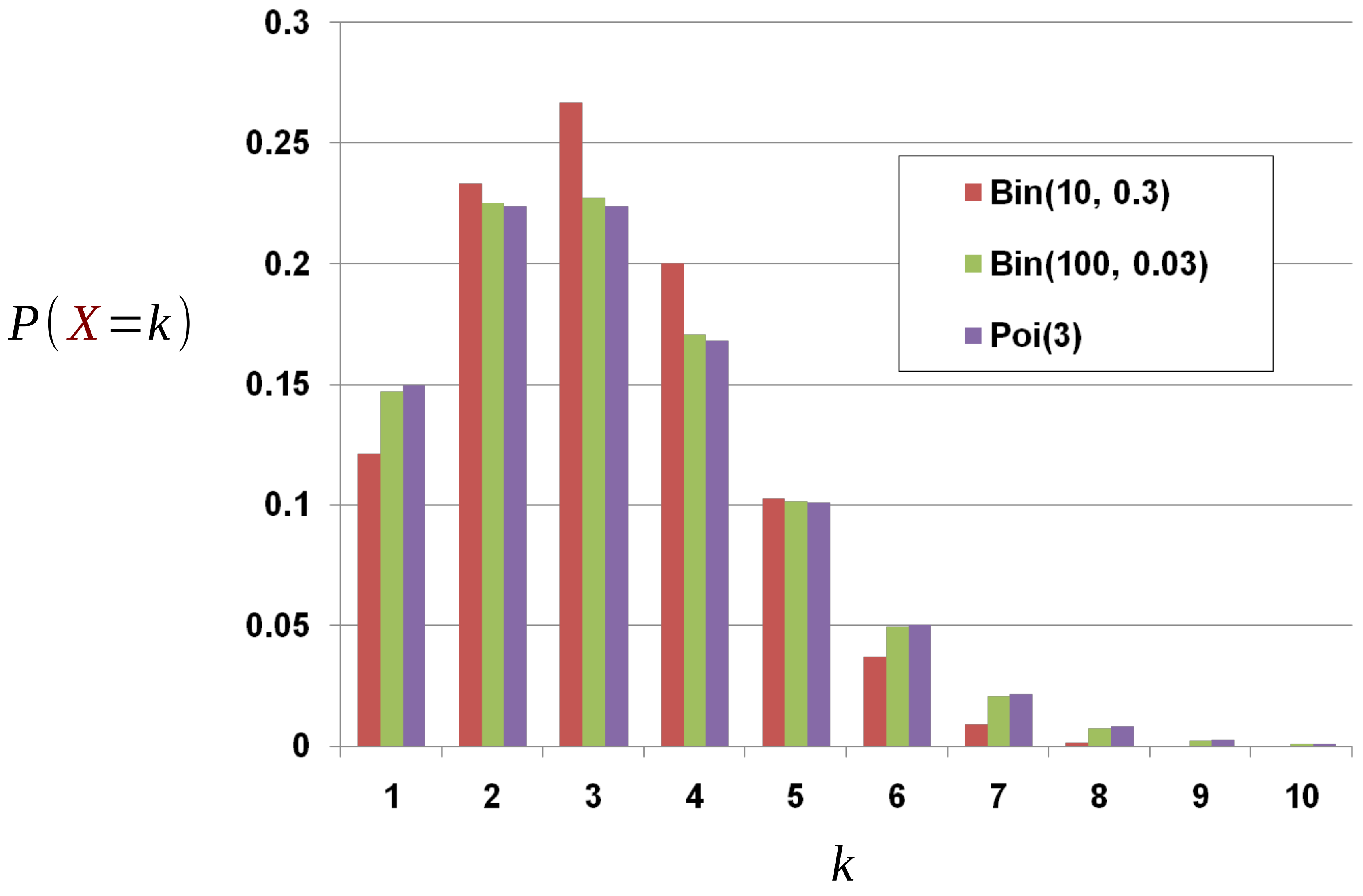
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Déjà vu?

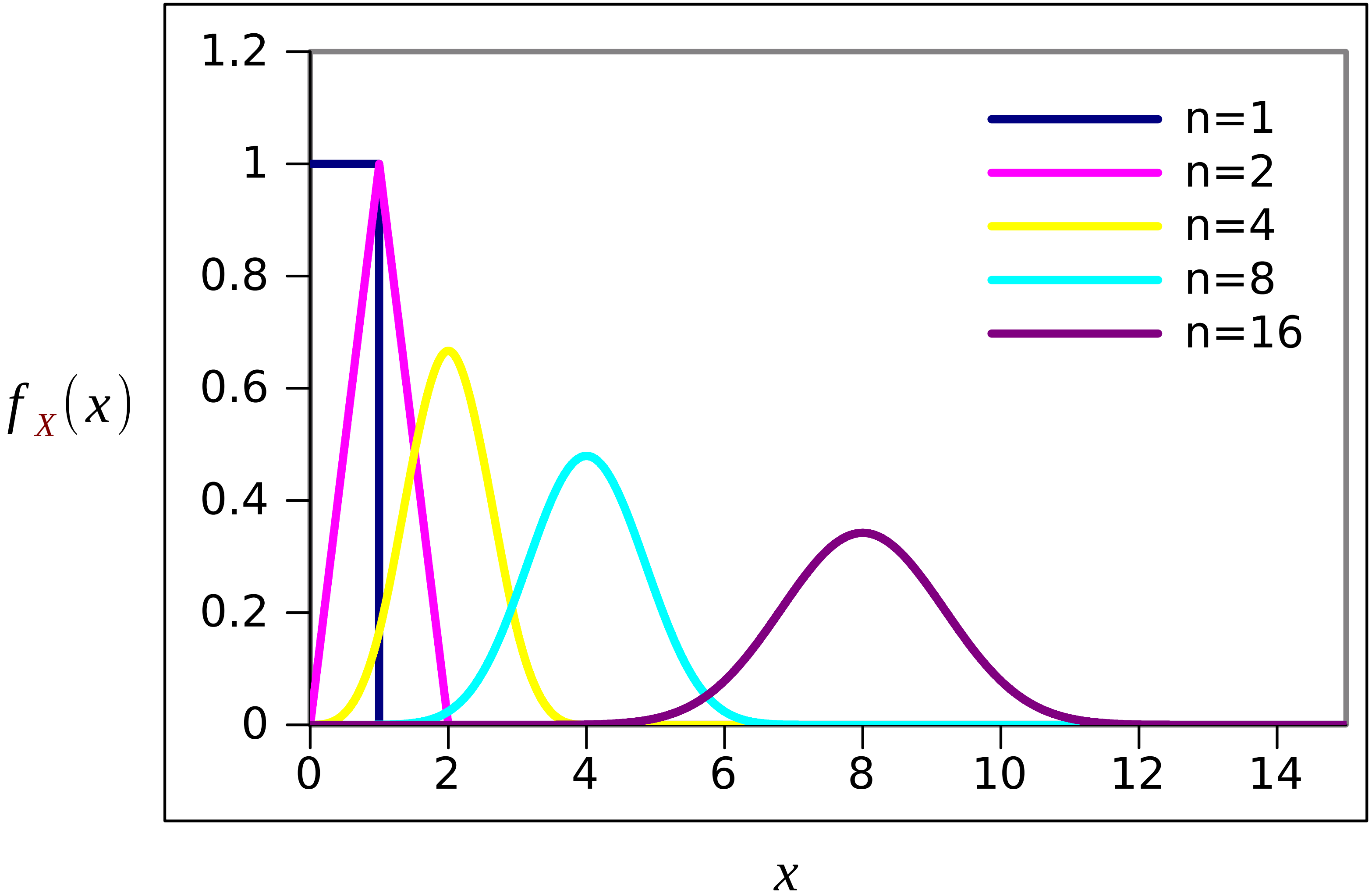


Déjà vu?



0.3
0.25
0.2
0.15
0.1
0.05
0
$P(X=k)$
Bin(10, 0.3)
Bin(100, 0.03)
Poi(3)
1
2
3
4
5
6
7
8
9
10
k

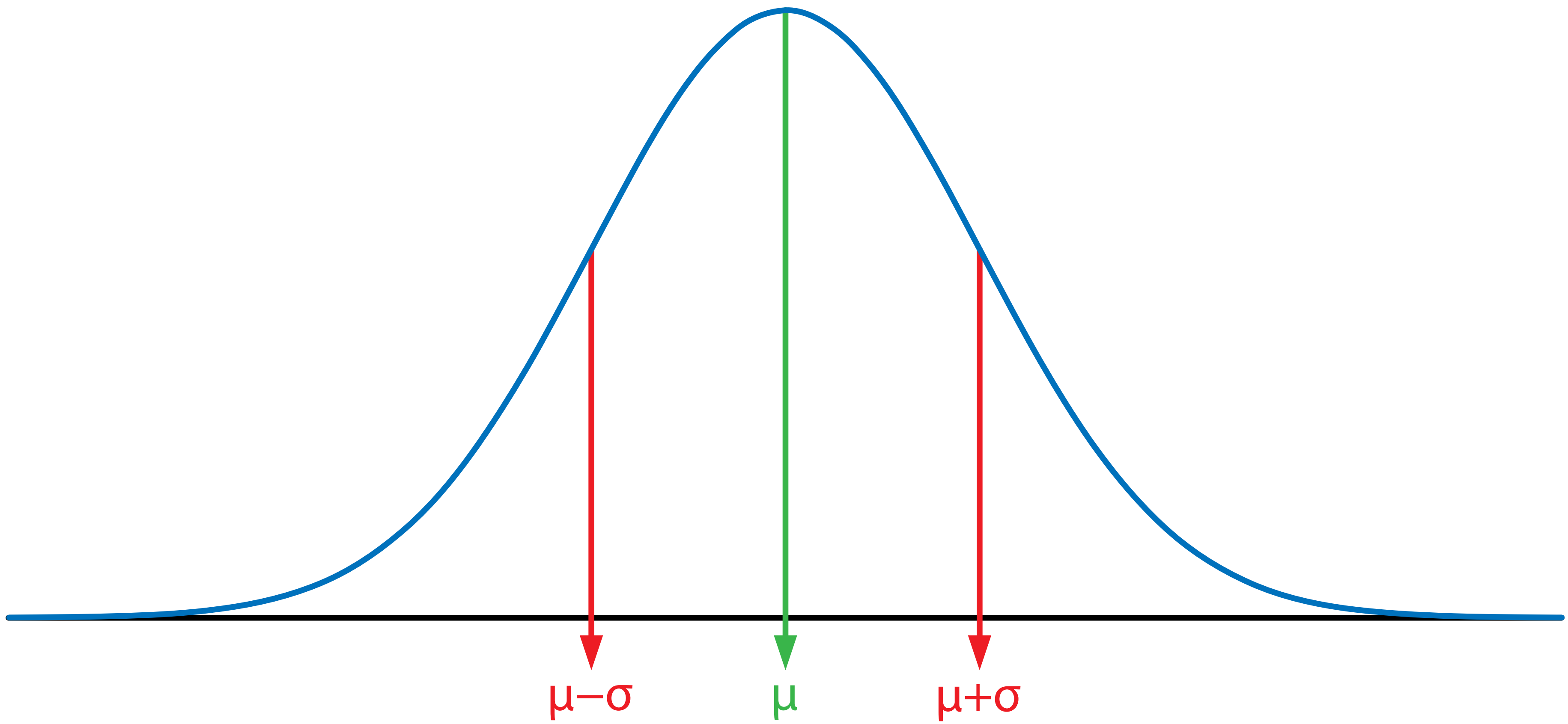
Déjà vu?



X = sum of *n* independent Uni(0, 1) variables

image: Thomasda

“The normal distribution”



Also known as: Gaussian distribution

Shape: bell curve

Personality: easygoing

What is normally distributed?

Natural phenomena: heights, weights…

Noise in measurements

(approximately)

Sums/averages of many random variables

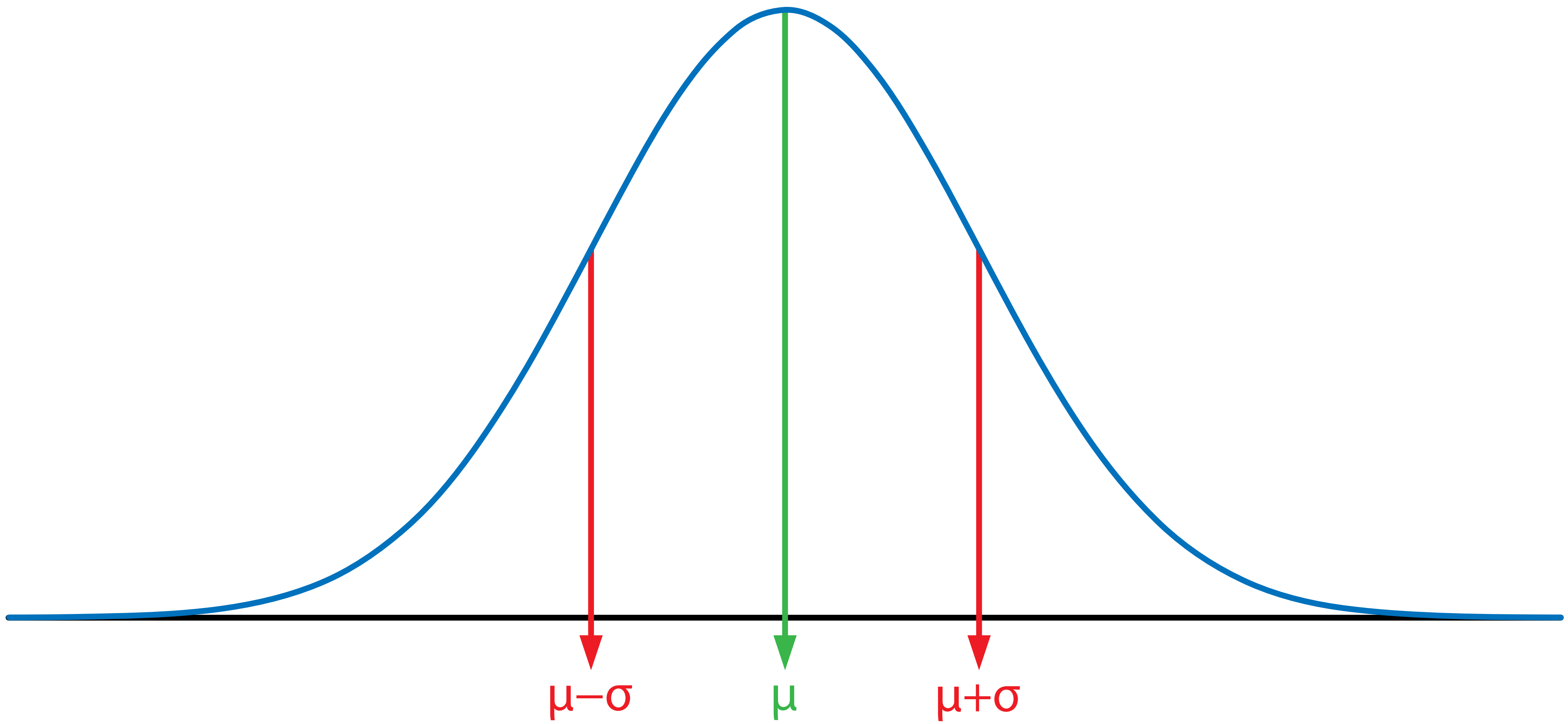
(caveats: independence, equal weighting, continuity…)

Averages of samples from a population

(with sufficient sample sizes)

The Know-Nothing Distribution

"maximum entropy"



The normal is the most spread-out distribution
with a fixed expectation and variance.

If you know E[X] and Var(X) but *nothing else*,
a normal is probably a good starting point!

Normal: Fact sheet

mean

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

variance (σ = standard deviation)

PDF: $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The Standard Normal

$$Z \sim N(0, 1)$$

μ $\quad$ σ^2

$$X \sim N(\mu, \sigma^2) \rightarrow X = \sigma Z + \mu$$

$$Z = \frac{X - \mu}{\sigma}$$

De-scarifying the normal PDF

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

De-scarifying the normal PDF

$$f_Z(z) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2}$$

De-scarifying the normal PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

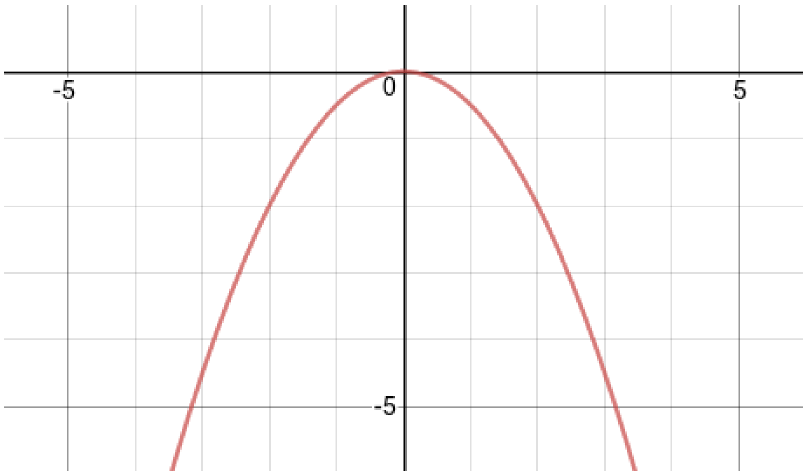
De-scarifying the normal PDF

$$f_Z(z) = C e^{-\frac{1}{2} z^2}$$

De-scarifying the normal PDF

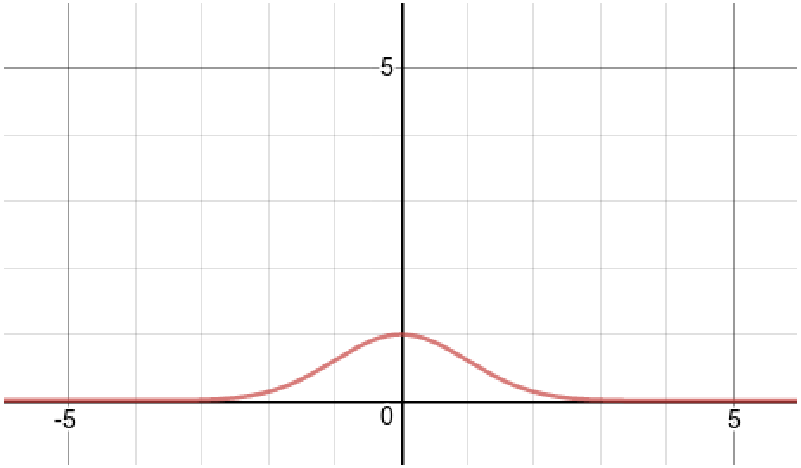
$$f_Z(z) = C e^{-\frac{1}{2} z^2}$$

$$-\frac{1}{2} z^2$$

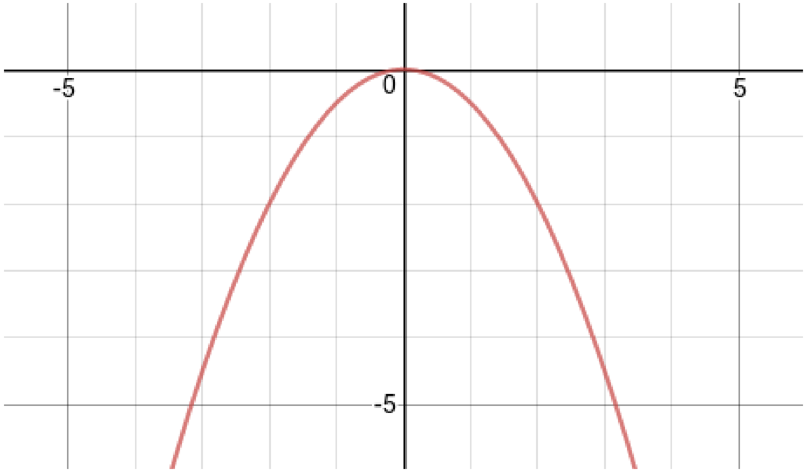


De-scarifying the normal PDF

$$f_Z(z) = C e^{-\frac{1}{2} z^2}$$



$$-\frac{1}{2} z^2$$



De-scarifying the normal PDF

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

normalizing
constant

$$Z = \frac{X-\mu}{\sigma}$$

Normal: Fact sheet

mean

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

variance (σ = standard deviation)

PDF: $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

CDF: $$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{x} dx\, f_X(x)$$

(no closed form)

The Standard Normal

$$Z \sim N(0, 1)$$

μ σ^2

$$X \sim N(\mu, \sigma^2) \rightarrow X = \sigma Z + \mu$$

$$Z = \frac{X - \mu}{\sigma}$$

$$\Phi(z) = F_Z(z) = P(Z \leq z)$$

Symmetry of the normal

$$P(X \le \mu - x) = P(X \ge \mu + x)$$

and don't forget:

$$P(X > x) = 1 - P(X \le x)$$

Symmetry of the normal

$$P(Z \leq -z) = P(Z \geq z)$$

and don't forget:

$$P(Z > z) = 1 - P(Z \leq z)$$

Symmetry of the normal

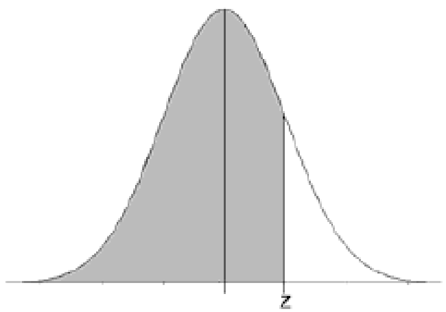
$$\Phi(-z) = P(Z \geq z)$$

and don't forget:

$$P(Z > z) = 1 - \Phi(z)$$

The standard normal table

Standard Normal Cumulative Probability Table



Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

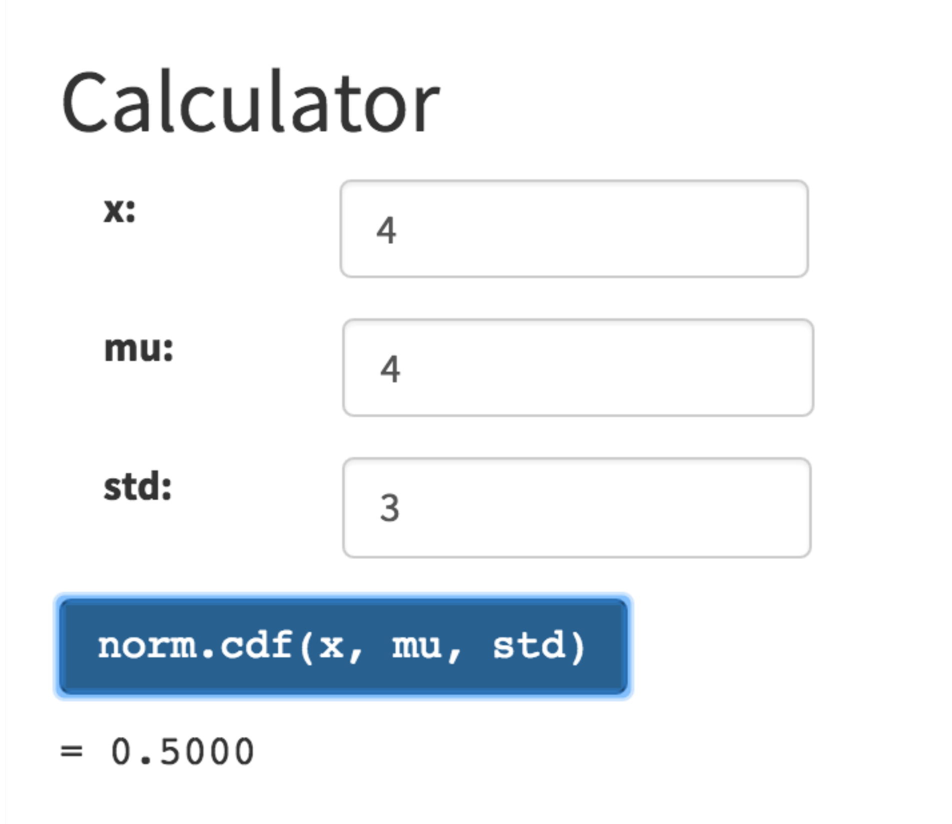
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

$$\Phi(0.54) = P(Z \leq 0.54) = 0.7054$$

With today's technology

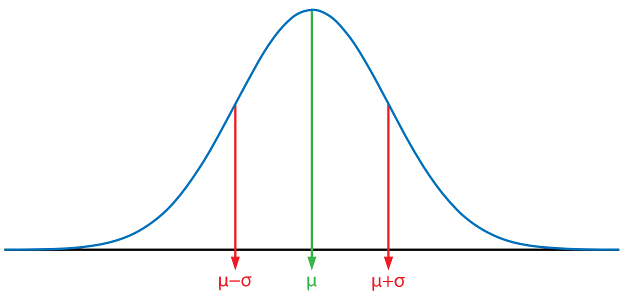
```
scipy.stats.norm(mean, std).cdf(x)
```

standard deviation! not variance.
you might need `math.sqrt` here.



Break time!

Practice with the Gaussian



$X \sim N(3, 16)$

$\mu = 3$

$\sigma^2 = 16$

$\sigma = 4$

$$P(X>0) = P\left(\frac{X-3}{4} > \frac{0-3}{4}\right)$$

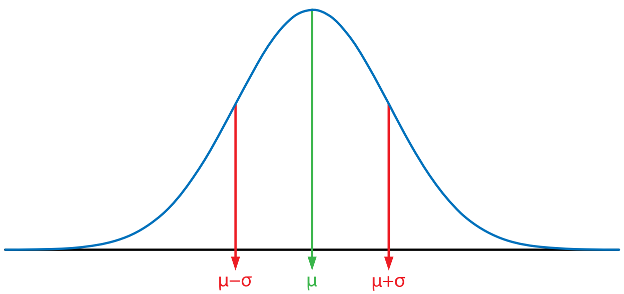
$$= P\left(Z > -\frac{3}{4}\right)$$

$$= 1 - P\left(Z \leq -\frac{3}{4}\right) = 1 - \Phi\left(-\frac{3}{4}\right)$$

$$= 1 - \left(1 - \Phi\left(\frac{3}{4}\right)\right)$$

$$= \Phi\left(\frac{3}{4}\right) \approx 0.7734$$

Practice with the Gaussian



$X \sim N(3, 16)$

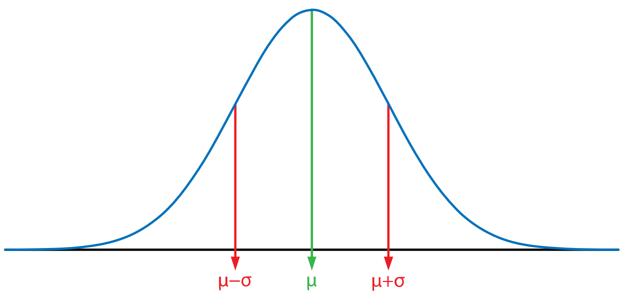
$\mu = 3$

$\sigma^2 = 16$

$\sigma = 4$

$$
\begin{aligned}
P(|X-3|>4) &= P(X<-1)+P(X>7) \\
&= P\left(\frac{X-3}{4}<\frac{-1-3}{4}\right)+P\left(\frac{X-3}{4}>\frac{7-3}{4}\right) \\
&= P(Z<-1)+P(Z>1) \\
&= \Phi(-1)+(1-\Phi(1)) \\
&= (1-\Phi(1))+(1-\Phi(1)) \\
&\approx 2\cdot(1-0.8413) \\
&= 0.3173
\end{aligned}
$$

Practice with the Gaussian



$X \sim N(3, 16)$

$\mu = 3$

$\sigma^2 = 16$

$\sigma = 4$

$$
\begin{aligned}
P(|X-\mu|>\sigma) &= P(X<\mu-\sigma)+P(X>\mu+\sigma) \\
&= P\left(\frac{X-\mu}{\sigma}<\frac{\mu-\sigma-\mu}{\sigma}\right)+P\left(\frac{X-\mu}{\sigma}>\frac{\mu+\sigma-\mu}{\sigma}\right) \\
&= P(Z<-1)+P(Z>1) \\
&= \Phi(-1)+(1-\Phi(1)) \\
&= (1-\Phi(1))+(1-\Phi(1)) \\
&\approx 2\cdot(1-0.8413) \\
&= 0.3173
\end{aligned}
$$

Normal: Fact sheet

mean

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

variance (σ = standard deviation)

PDF: $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

CDF: $$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{x} dx\, f_X(x)$$

(no closed form)

expectation: $$E[X] = \mu$$

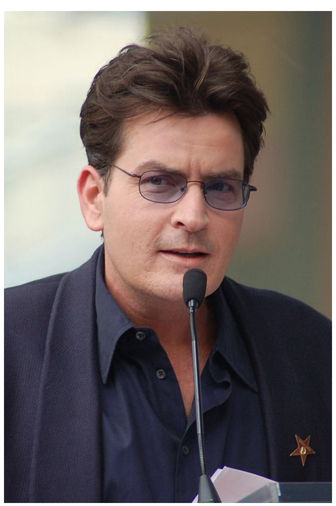
variance: $$\mathrm{Var}(X) = \sigma^2$$

Carl Friedrich Gauss

(1775-1855)—remarkably influential German mathematician

Started doing groundbreaking math as a teenager

Didn't invent the normal distribution (but popularized it)



Noisy wires

Send a voltage of X = 2 or -2 on a wire.
+2 represents 1, -2 represents 0.

Receive voltage of X + Y on other end,
where Y ~ *N*(0, 1).

If X + Y ≥ 0.5, then output 1, else 0.

P(incorrect output | original bit = 1) =

$$
\begin{aligned}
P(2+Y<0.5) &= P(Y<-1.5) \\
&= \Phi(-1.5) \\
&= 1-\Phi(1.5) \approx 0.0668
\end{aligned}
$$

Noisy wires

Send a voltage of $X = 2$ or -2 on a wire.
$+2$ represents 1, -2 represents 0.

Receive voltage of $X + Y$ on other end,
where $Y \sim \mathcal{N}(0, 1)$.

If $X + Y \geq 0.5$, then output 1, else 0.

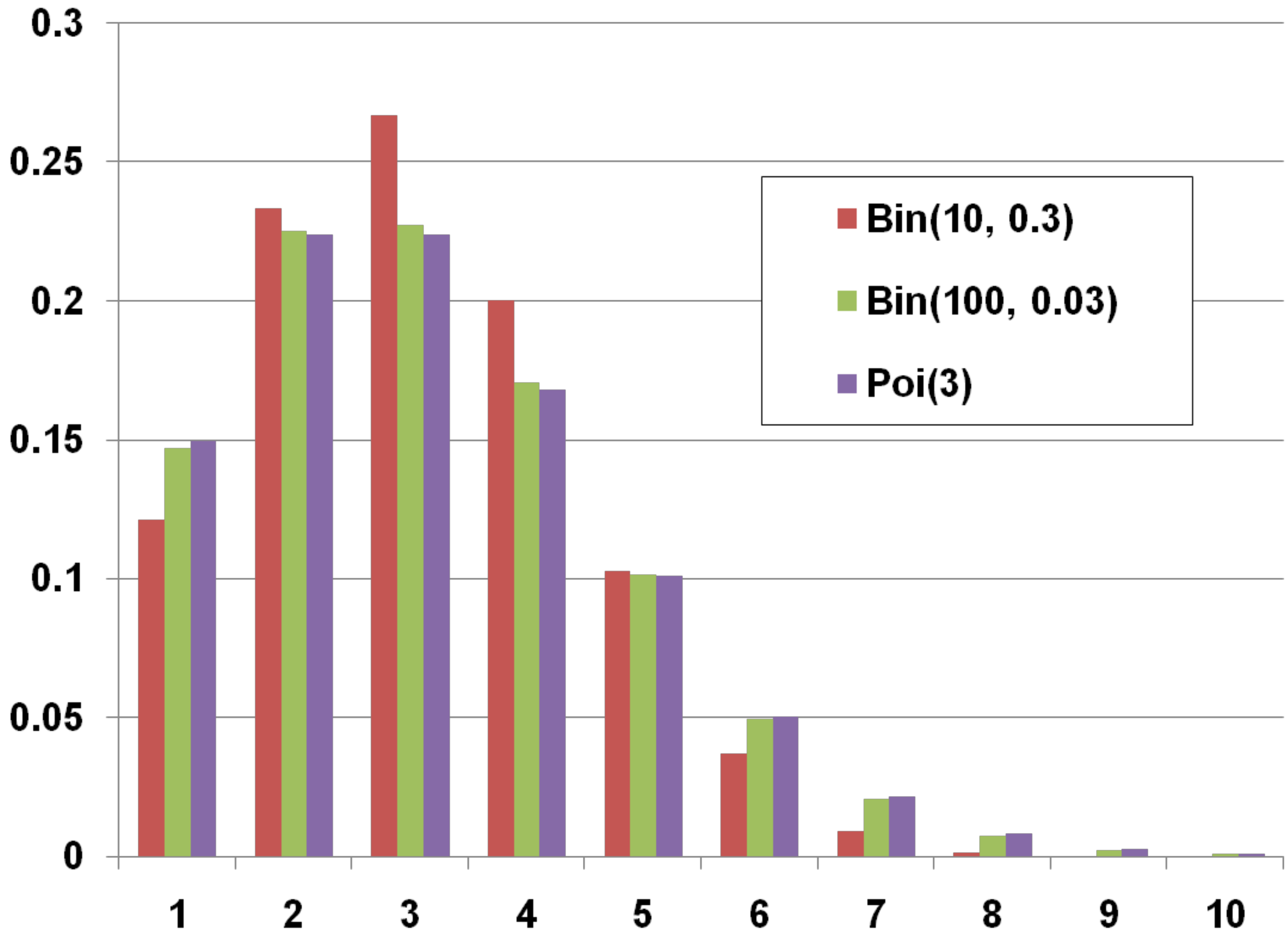
P(incorrect output | original bit = 0) =

$$P\left(-2+Y \geq 0.5\right) = P\left(Y \geq 2.5\right)$$
$$= 1 - P\left(Y < 2.5\right)$$
$$= 1 - \Phi\left(2.5\right) \approx 0.0062$$

Poisson approximation to binomial

large *n*, small *p*

$P(X=k)$

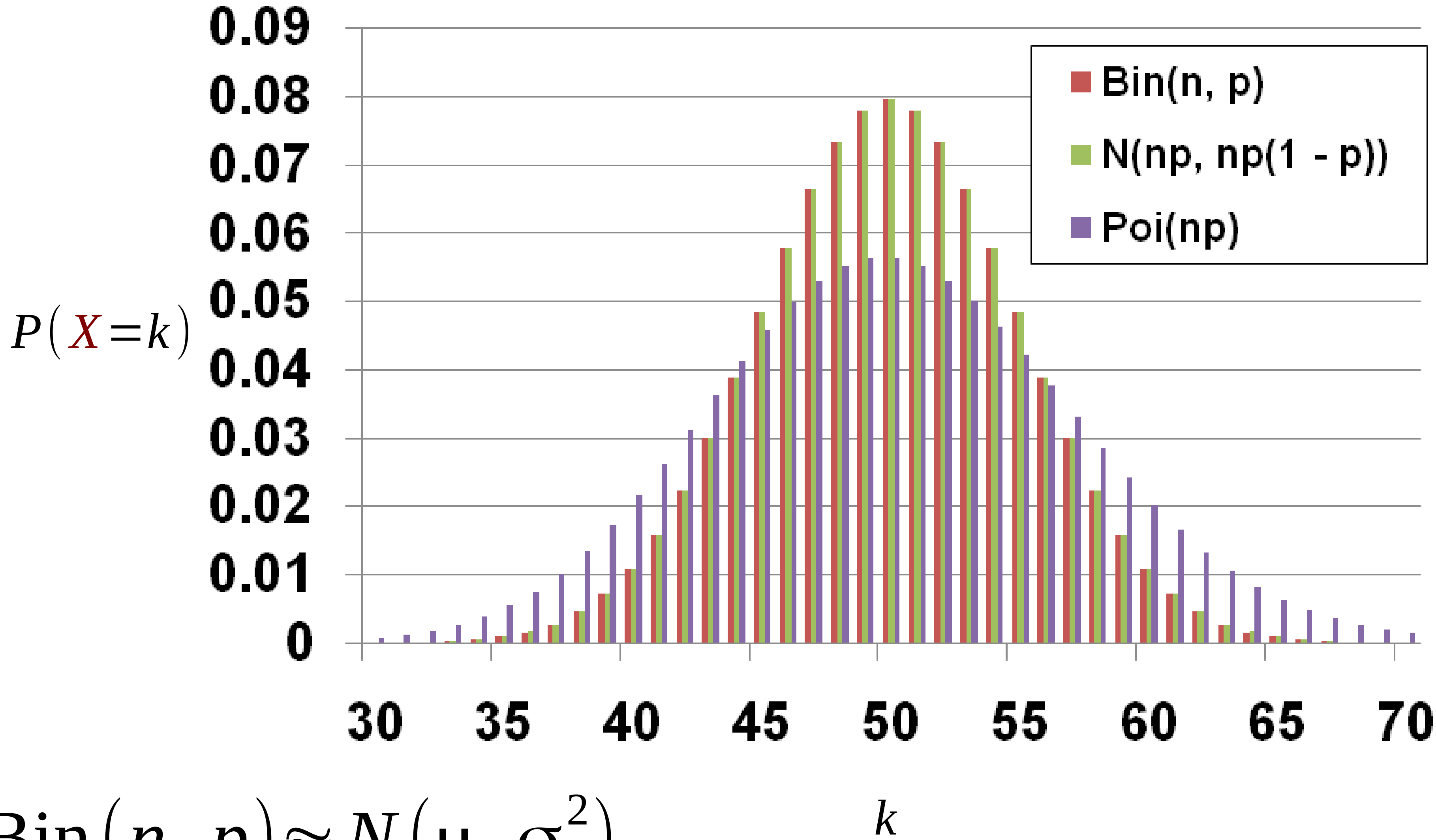


k

$$\text{Bin}(n,p) \approx \text{Poi}(\lambda)$$

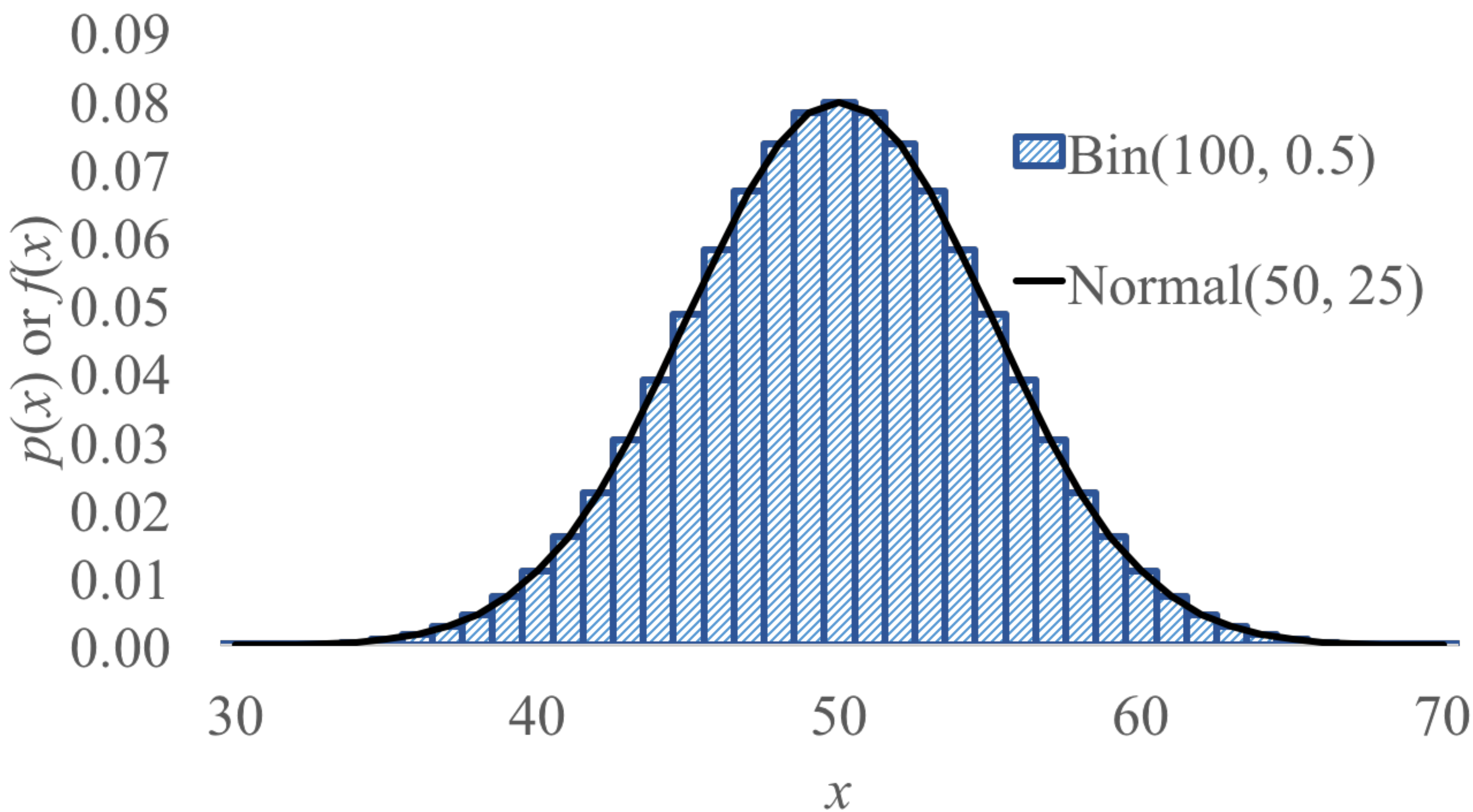
Normal approximation to binomial

large *n*, **medium** *p*



$$\mathrm{Bin}(n,p) \approx N(\mu, \sigma^2)$$

Something is strange...

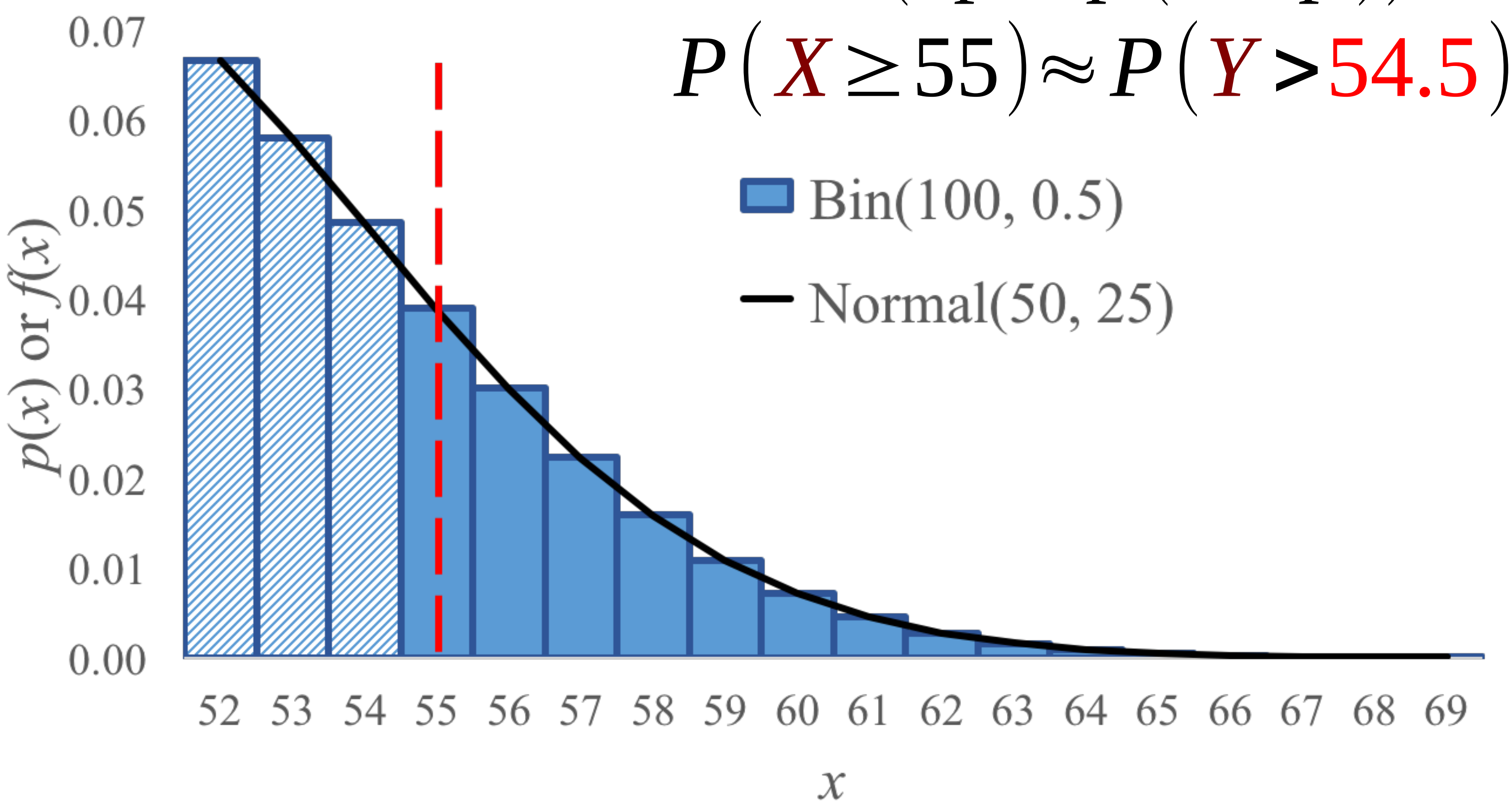


Continuity correction

$$X \sim \text{Bin}(n, p)$$

$$Y \sim N(np, np(1-p))$$

$$P(X \geq 55) \approx P(Y > 54.5)$$



When approximating a **discrete** distribution with a **continuous** distribution, adjust the bounds by 0.5 to account for the missing half-bar.

Miracle diets

100 people placed on a special diet.

Doctor will endorse diet if ≥ 65 people have cholesterol levels decrease.

What is P(doctor endorses | diet has no effect)?

X: # people whose cholesterol decreases

X ~ Bin(100, 0.5)
$np = 50$
$np(1 - p) = 50(1 - 0.5) = 25$

$\approx Y \sim N(50, 25)$

$$P(Y > 64.5) = P\left(\frac{Y - 50}{5} > \frac{64.5 - 50}{5}\right)$$

$$= P(Z > 2.9) = 1 - \Phi(2.9) \approx 0.00187$$

Stanford admissions

Stanford accepts 2480 students.
Each student independently
decides to attend with $p = 0.68$.

What is
P(at least 1750 students attend)?

X: # of students who will attend.
$X \sim \text{Bin}(2480, 0.68)$
$np = 1686.4$
$\sigma^2 = np(1-p) \approx 539.65$

$\approx Y \sim N(1686.4, 539.65)$

$$P(Y > 1749.5) = P\left(\frac{Y - 1686.4}{\sqrt{539.65}} > \frac{1749.5 - 1686.4}{\sqrt{539.65}}\right)$$

$$\approx P(Z > 2.54) = 1 - \Phi(2.54) \approx 0.0053$$

image: Victor Gane

Stanford admissions changes

The Stanford Daily

Class of 2018 admit rates lowest in University history

March 28, 2014 16 Comments Tweet

Alex Zivkovic
Desk Editor

Stanford admitted 2,138 students to the Class of 2018 in this year's admissions cycle, producing – at 5.07 percent – the lowest admit rate in University history.

The University received a total of 42,167 applications this year, a record total and a 8.6 percent increase over last year's figure of 38,828. Stanford accepted 748 students

Record 81.1 percent yield rate reported for Class of 2019

June 9, 2015 24 Comments Tweet

Victor Xu

The Office of Undergraduate Admission reports that 81.1 percent of admitted undergraduates enrolled as students in the Class of 2019, up from 78.2 percent last year. The yield rate is the highest in Stanford history.