11: Joint (Multivariate) Distributions

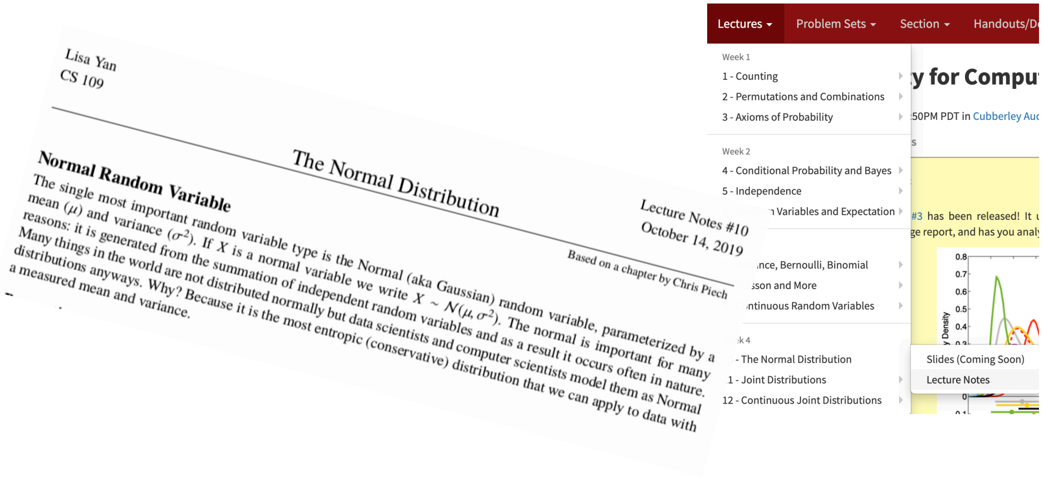
Lisa Yan

October 16, 2019

Concept check feedback

"It is difficult to know which random variable distribution to use when."

"Parts of last lecture were a bit confusing because of typos."



This is a totally understandable and relatable concern!

Problem Set 3 + Section 3 goals:

- Read problems
- Identify random variables.

Thank you for keeping me honest!

- The corrected slides are on website
- Lecture notes have also been updated with explanations for all examples.

Normal RVs

Review

mean variance

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

$$P(X \le x) = F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

⚠ CDF has no closed form

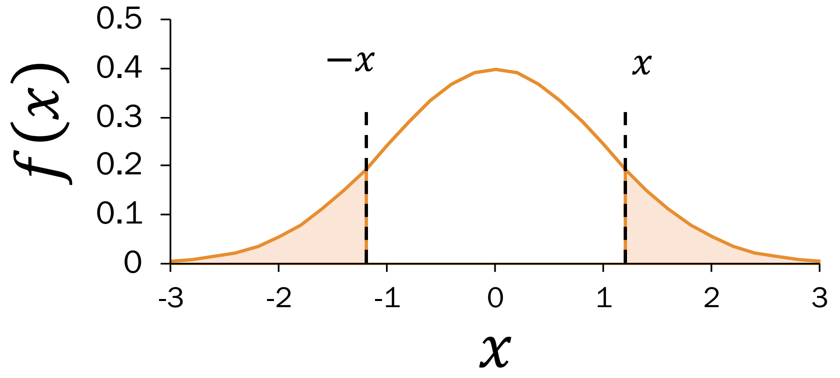
CDF of Standard Normal Z, solved for numerically

CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

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

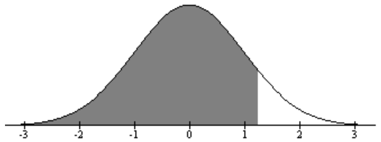
- Symmetry of the PDF of Normal RV implies $\Phi(-x) = 1 - \Phi(x)$



Standard Normal Table

Standard Normal Table

An entry in the table is the area under the curve to the left of z, $P(Z \leq z) = \Phi(z)$.

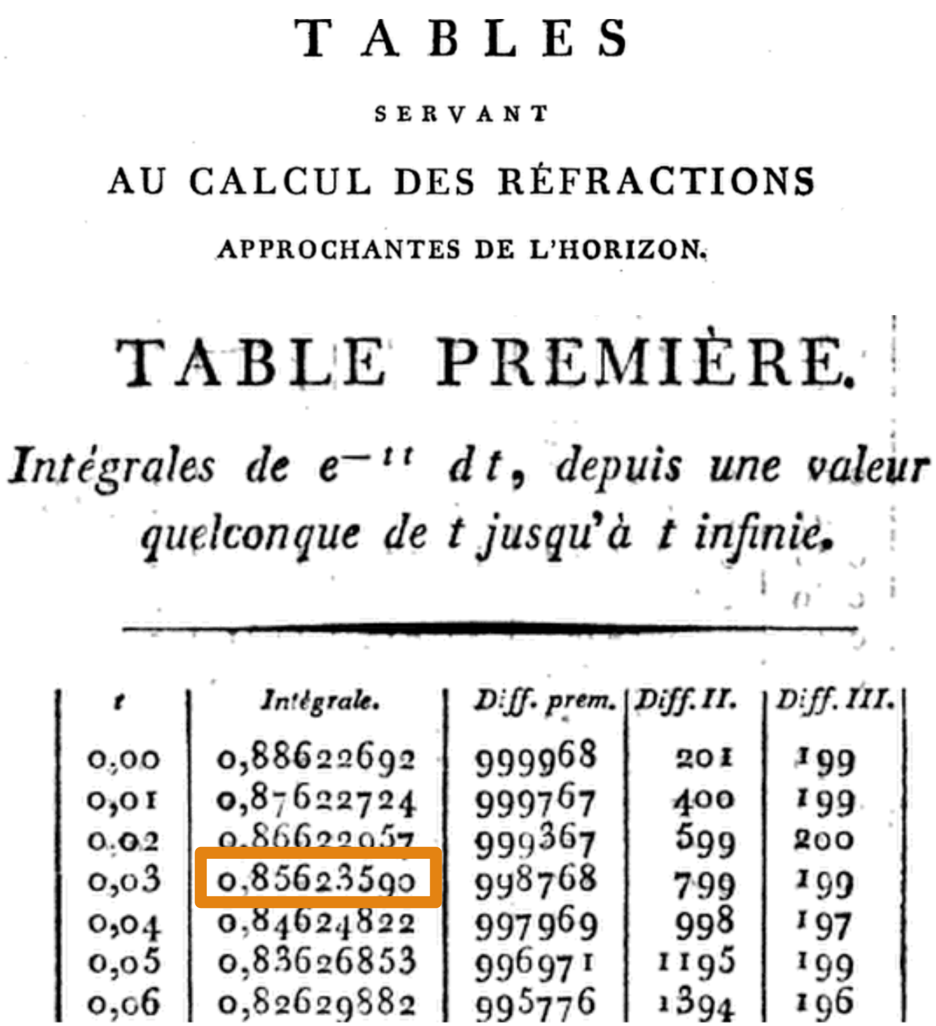


Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	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	0.7703	0.7734	0.7764	0.7793	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	0.8906	0.8925	0.8943	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
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

- $Z \sim \mathcal{N}(0, 1)$ has a numeric lookup table for $\Phi(x)$, where $\boldsymbol{x \geq 0}$.
- Computing implications: saving one lookup table for $\mathcal{N}(0, 1)$ enables you to quickly compute probabilities for general $\mathcal{N}(\mu, \sigma^2)$!

Standard Normal Table

optional



TABLES

SERVANT

AU CALCUL DES RÉFRACTIONS

APPROCHANTES DE L'HORIZON.

TABLE PREMIÈRE.

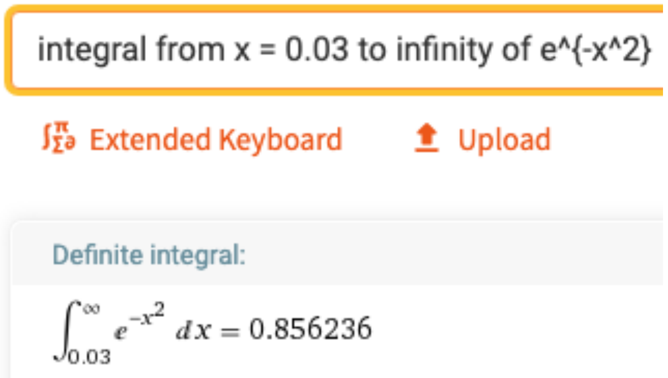
Intégrales de $e^{-tt}\,dt$, depuis une valeur quelconque de t jusqu'à t infinie.

t	Intégrale.	Diff. prem.	Diff. II.	Diff. III.
0,00	0,88622692	999968	201	199
0,01	0,87622724	999767	400	199
0,02	0,86622957	999367	599	200
0,03	$\boxed{0{,}85623590}$	998768	799	199
0,04	0,84624822	997969	998	197
0,05	0,83626853	996971	1195	199
0,06	0,82629882	995776	1394	196

The Standard Normal Table was first computed by Christian Kramp.

French astronomer (1760–1826).
Analyse des Réfractions Astronomiques et Terrestres, 1799

Used a Taylor series expansion to the third power



Today's plan

Normal approximation for Binomial (on pset3)

Joint distributions (discrete)

Multinomial Random Variable

Text analysis

Website testing

(where we left off)

- 100 people are given a new website design.
- X = # people whose time on site increases
- CEO will endorse the new design if $X \geq 65$.
- The design actually has no effect, so P(time on site increases) = 0.5 independently.

What is P(CEO endorses change)? *Give a numerical approximation*.

Strategy:

A. Poisson

B. Bayes' Theorem

C. Binomial

Yes, actually! D. Normal (Gaussian)

E. Uniform

Website testing

- 100 people are given a new website design.
- X = # people whose time on site increases
- CEO will endorse the new design if $X \geq 65$.
- The design actually has no effect, so P(time on site increases) = 0.5 independently.

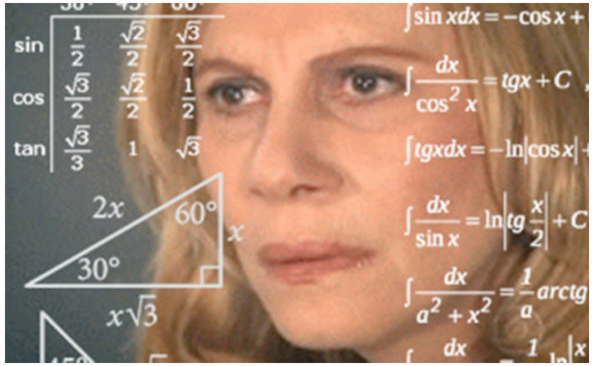
What is $P(\text{CEO endorses change})$? *Give a numerical approximation*.

Approach 1: Binomial

Define

$$X \sim \text{Bin}(n = 100, p = 0.5)$$

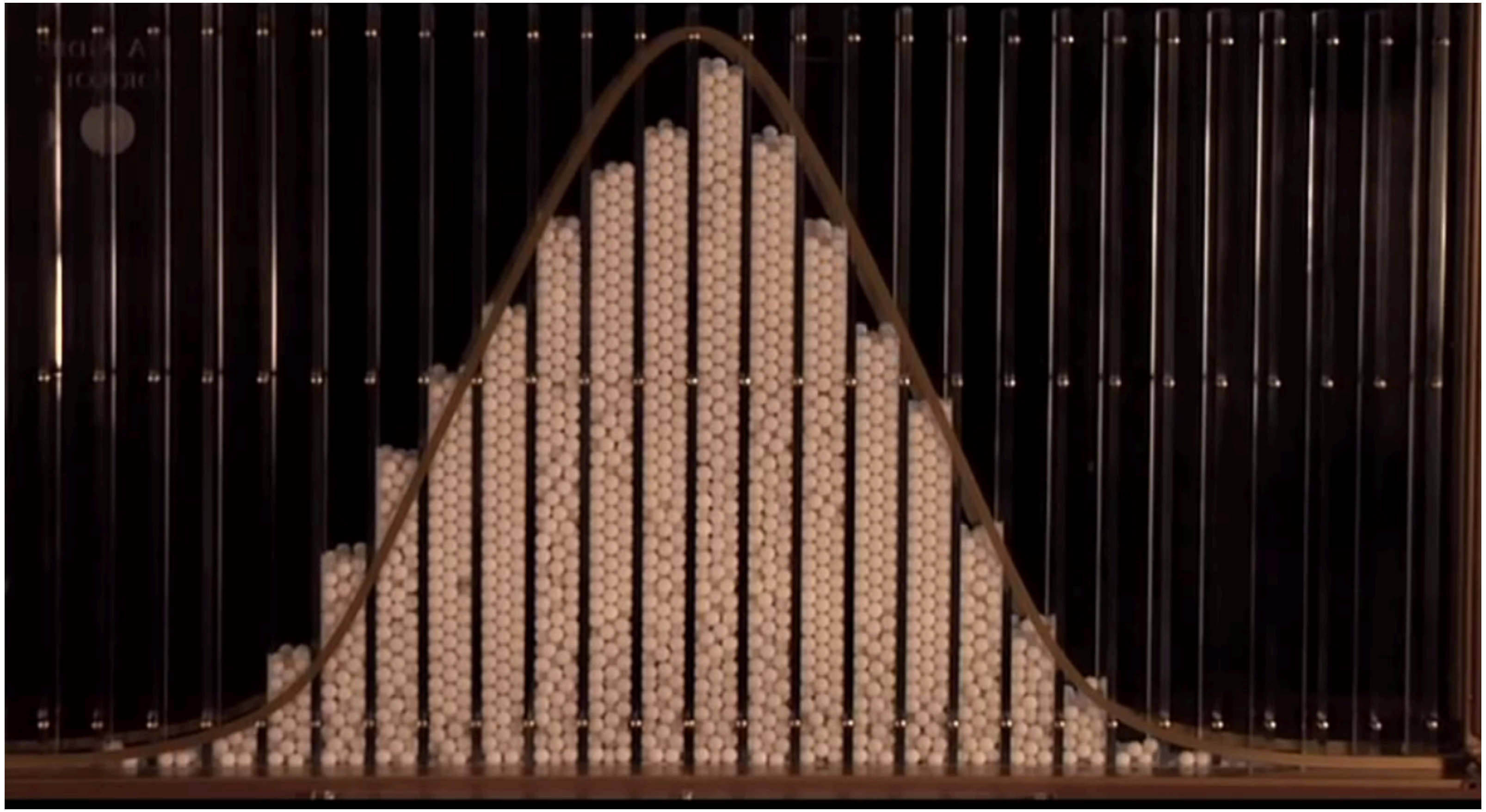
Want: $P(X \geq 65)$



Solve

$$P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} 0.5^i (1 - 0.5)^{100-i}$$

Don’t worry, Normal approximates Binomial



Galton Board

(We’ll explain where this approximation comes from in 2 weeks’ time)

Website testing

- 100 people are given a new website design.
- X = # people whose time on site increases
- CEO will endorse the new design if $X \geq 65$.
- The design actually has no effect, so P(time on site increases) = 0.5 independently.

What is $P(\text{CEO endorses change})$? *Give a numerical approximation.*

Approach 1: Binomial

Define

$X \sim \text{Bin}(n = 100, p = 0.5)$

Want: $P(X \geq 65)$

Solve

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

Approach 2: approximate with Normal

Define

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

$\mu = np = 50$
$\sigma^2 = np(1-p) = 25$
$\sigma = \sqrt{25} = 5$

Solve

$$P(X \geq 65) \approx P(Y \geq 65) = 1 - F_Y(65)$$ 🤨

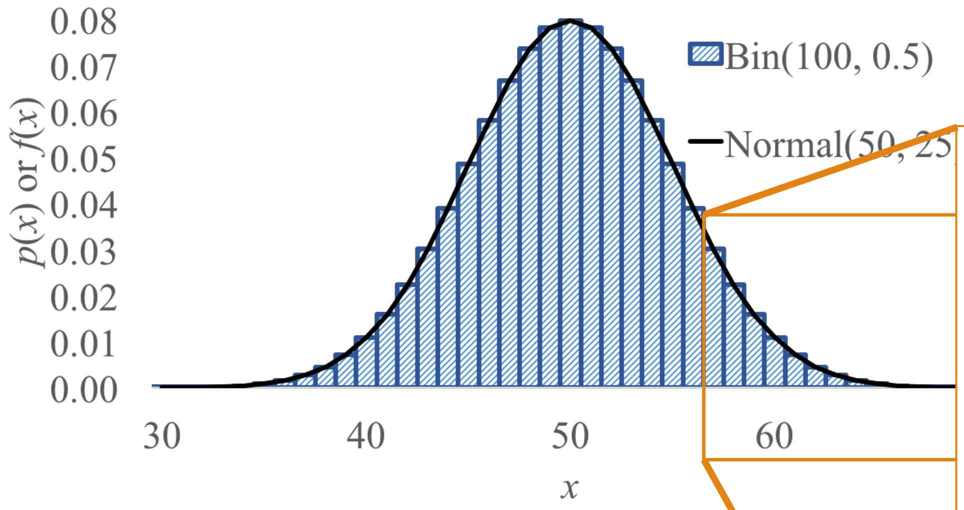
$$= 1 - \Phi\left(\frac{65-50}{5}\right) = 1 - \Phi(3) \approx 0.0013 \text{ ?}$$

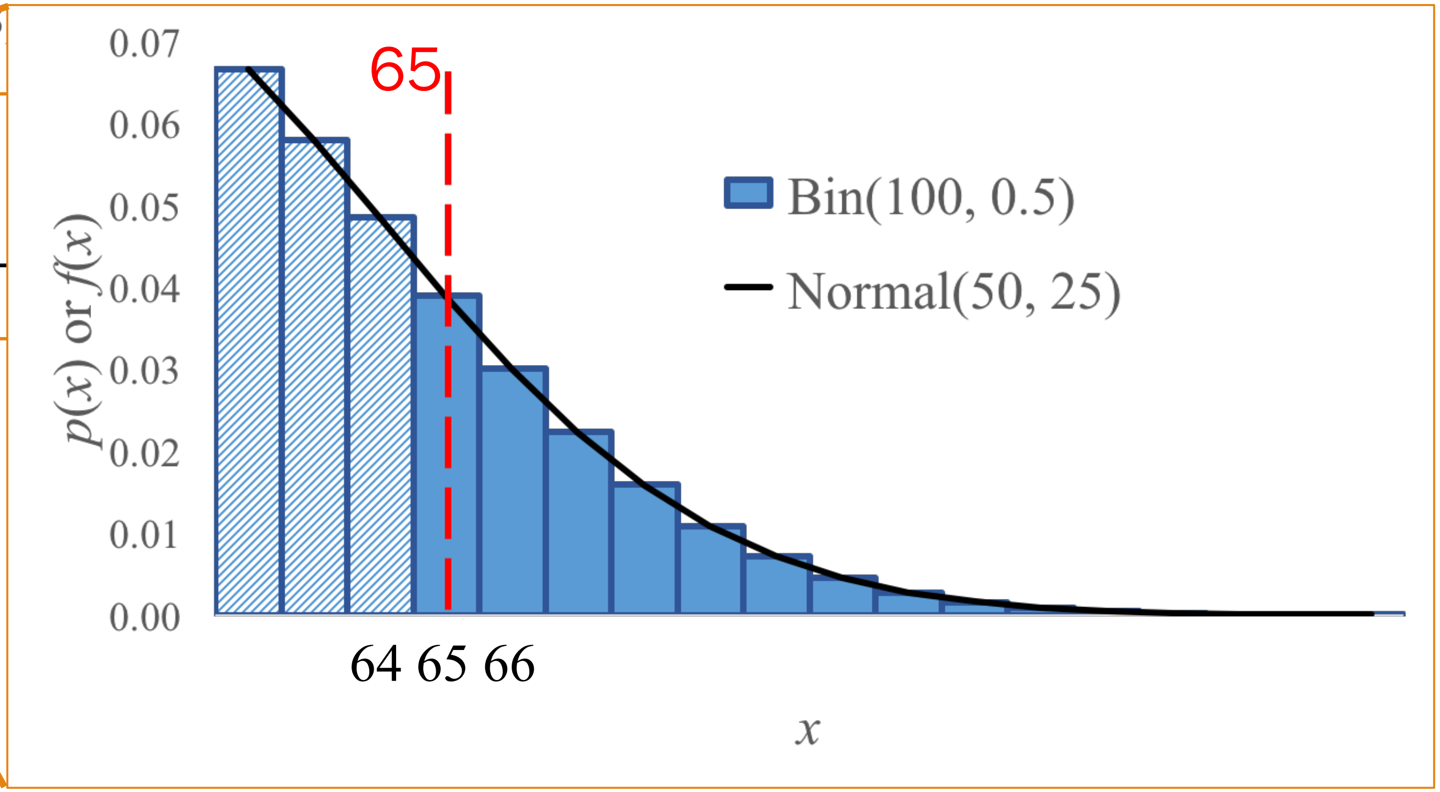
⚠️ ⚠️ (this approach is actually missing something)

Website testing with continuity correction

You must perform a continuity correction when approximating a discrete RV with a continuous RV.

$Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$





$P(X \geq 65)$ Binomial

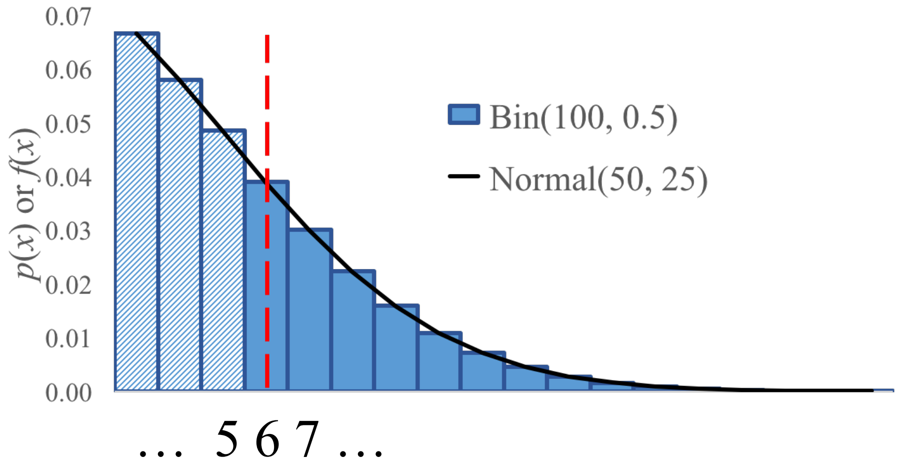
$\approx P(Y \geq 64.5)$ Normal

≈ 0.0018 ✅

Continuity correction

If $Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

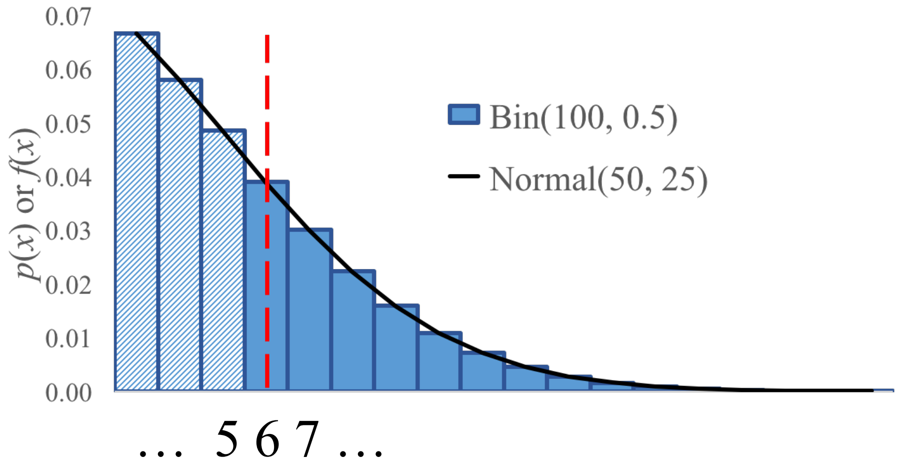
Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	
$P(X \geq 6)$	
$P(X > 6)$	
$P(X < 6)$	
$P(X \leq 6)$	



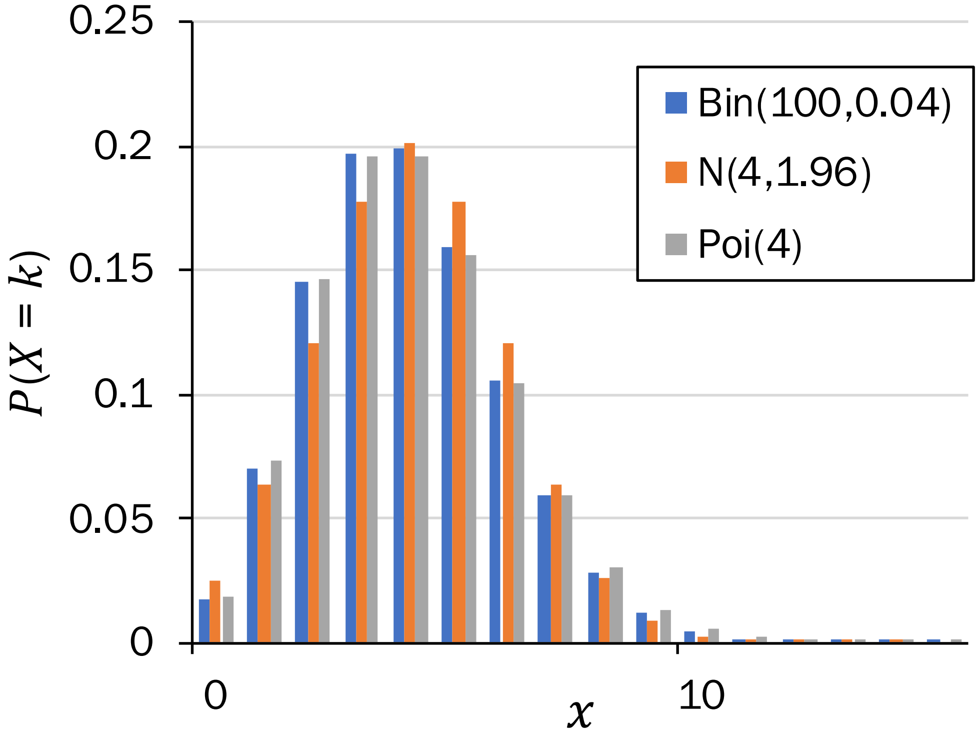
Continuity correction

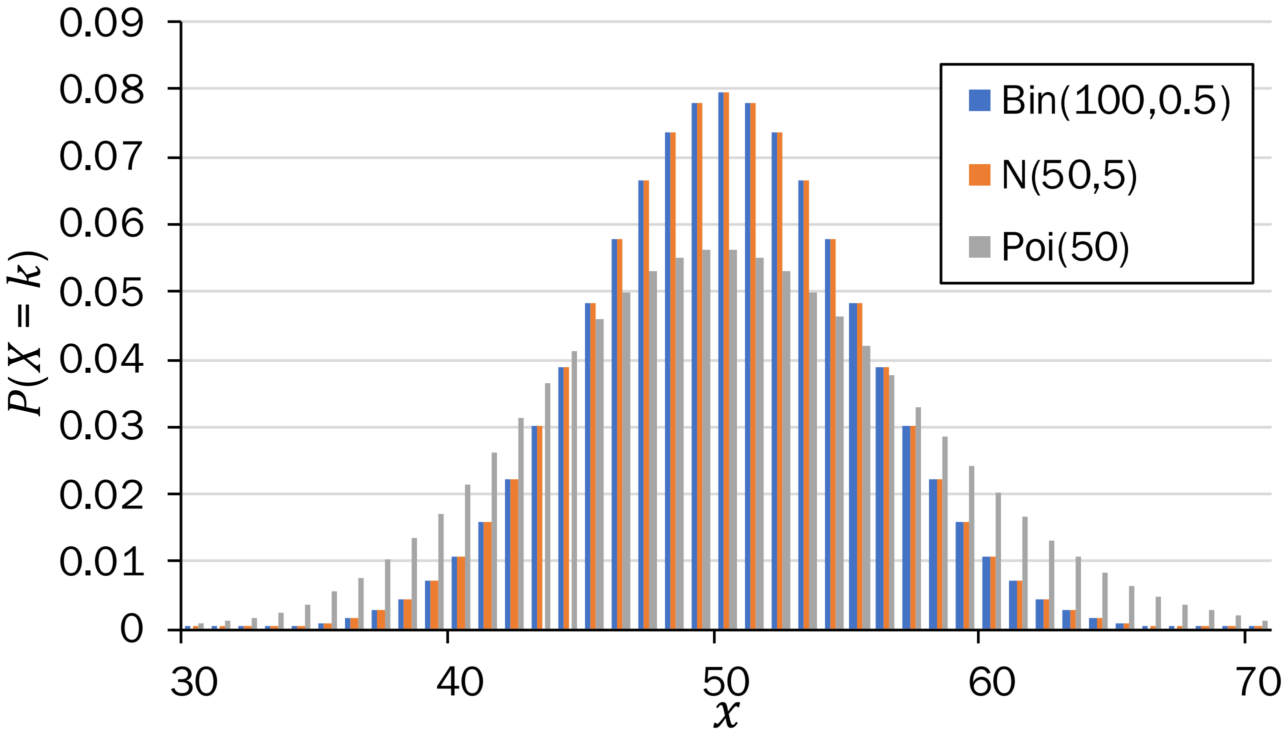
If $Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	$P(5.5 \le Y \le 6.5)$
$P(X \ge 6)$	$P(Y \ge 5.5)$
$P(X > 6)$	$P(Y \ge 6.5)$
$P(X < 6)$	$P(Y \le 5.5)$
$P(X \le 6)$	$P(Y \le 6.5)$



Who gets to approximate?





Poisson approximation
n large (> 20), p small (< 0.05)
slight dependence okay

Normal approximation
n large (> 20), p mid-ranged ($np(1-p) > 10$)
independence

1. If there is a choice, use Normal to approx.
2. When using Normal to approximate a discrete RV, use a continuity correction.

Stanford Admissions (a while back)

Stanford accepts 2480 students.

- Each accepted student has 68% chance of attending (independent trials)
- Let X = # of students who will attend

What is $P(X > 1745)$? *Give a numerical approximation*.

Strategy:

A. Just Binomial
B. Poisson
C. Normal
D. None/other

Stanford Admissions (a while back)

Stanford accepts 2480 students.

- Each accepted student has 68% chance of attending (independent trials)
- Let X = # of students who will attend

What is $P(X > 1745)$? *Give a numerical approximation*.

Strategy:

A. Just Binomial — $n = 2480$, computationally expensive
B. Poisson — $p = 0.68$, not small enough
C. Normal — ✅ Variance $np(1-p) = 540 > 10$
D. None/other

Define an approximation

Let $Y \sim \mathcal{N}\left(E[X], \text{Var}(X)\right)$

$E[X] = np = 1686$
$\text{Var}(X) = np(1-p) \approx 540 \rightarrow \sigma = 23.3$

$P(X > 1745) \approx P(Y \geq 1745.5)$ ⚠️ Continuity correction

Solve

$$P(Y \geq 1745.5) = 1 - F(1745.5)$$
$$= 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right)$$
$$= 1 - \Phi(2.54) \approx 0.0055$$

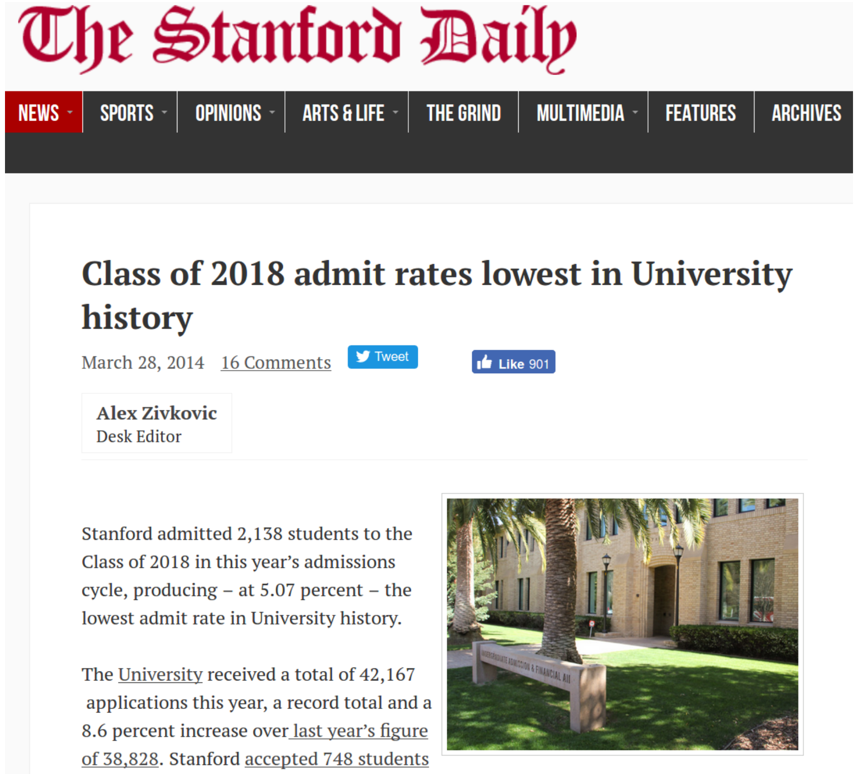
Changes in Stanford Admissions

Stanford accepts 2480 students.

- Each accepted student has 68% chance of attending (independent trials)
- Let X = # of students who will attend

Yield rate 20 years ago

What is $P(X > 1745)$? *Give a numerical approximation*.



The Stanford Daily

NEWS SPORTS OPINIONS ARTS & LIFE THE GRIND MULTIMEDIA FEATURES ARCHIVES

Class of 2018 admit rates lowest in University history

March 28, 2014 16 Comments Tweet Like 901

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

Overview for the Class of 2022

- Total Applicants: 47,451
- Total Admits: 2,071
- Total Enrolled: 1,706

Admit rate: 4.3%

Yield rate: 81.9%

People love coming to Stanford!

Today's plan

Normal approximation for Binomial

Cool normal facts

Joint distributions (discrete)

Multinomial Random Variable

Text analysis

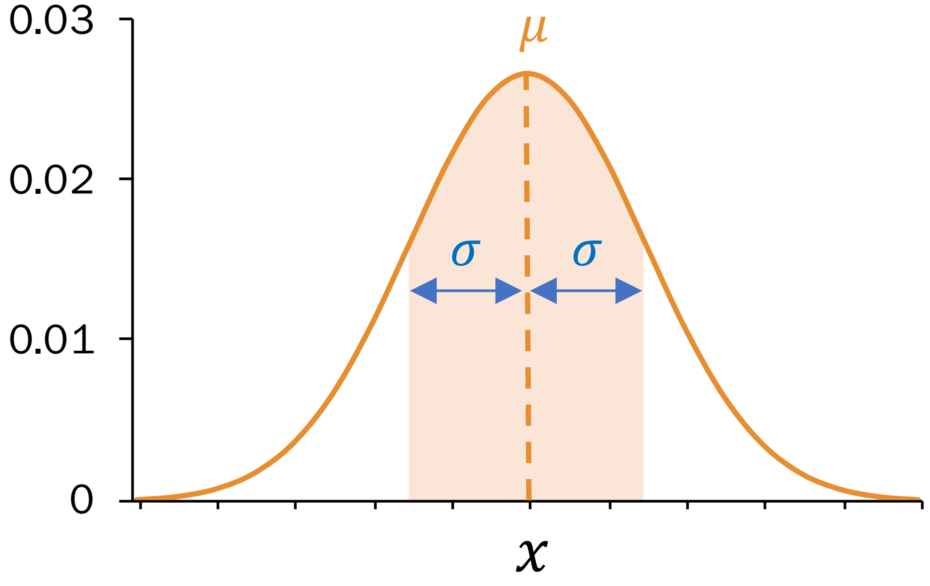
68% rule

You may have heard the statement:

"68% of the class will fall within 1 standard deviation of the exam average."

This is only true of normal distributions:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF F.



$$P(|X-\mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$$
$$= F(\mu+\sigma) - F(\mu-\sigma)$$
$$= \Phi\left(\frac{(\mu+\sigma)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-\sigma)-\mu}{\sigma}\right)$$
$$= \Phi(1) - \Phi(-1) = \Phi(1) - \left(1 - \Phi(1)\right)$$
$$= 2\Phi(1) - 1 \quad \approx 2(0.8413) - 1 = 0.6826$$

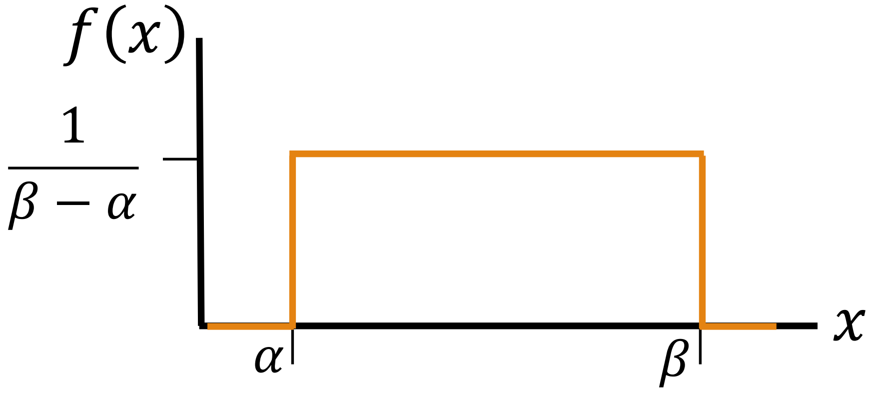
68% rule

You may have heard the statement:

"68% of the class will fall within 1 standard deviation of the exam average."

This is only true of normal distributions:

Counterexample: Let $X \sim \text{Unif}(\alpha, \beta)$.



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

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12} \qquad \sigma = \text{SD}(X) = \frac{\beta - \alpha}{\sqrt{12}}$$

$$\begin{aligned} P(|X - \mu| < \sigma) &= P(\mu - \sigma < X < \mu + \sigma) \\ &= \frac{1}{\beta - \alpha} \cdot [(\mu + \sigma) - (\mu - \sigma)] \\ &= \frac{1}{\beta - \alpha}[2\sigma] = \frac{1}{\beta - \alpha} \cdot \left[2 \cdot \frac{\beta - \alpha}{\sqrt{12}}\right] \\ &= 2/\sqrt{12} \approx 0.58 \end{aligned}$$

How does a computer sample the Normal?

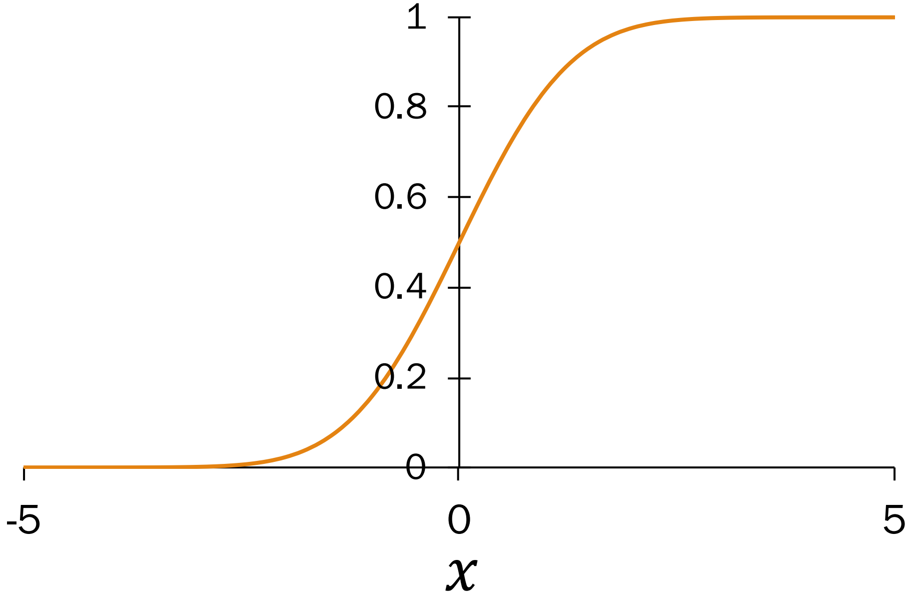
optional

How does Python generate random values according to a Normal distribution?

```
from scipy import stats
mean = 0
std = 1
for i in range(10):
  sample = stats.norm.rvs(mean, std)
  print(sample)
```

```
-1.5213511002970745
 1.3986457271717916
 2.1661966495582745
-0.09612045842653026
-0.6504681012424954
-0.6614649985106745
-1.1273650614139048
-1.8898482565694437
-2.4804202575017054
 0.8141949960752278
```

CDF of Standard Normal, $\Phi(x)$



Inverse transform sampling

1. Generate a random probability u from $U \sim \text{Unif}(0,1)$.
2. Find x such that $\Phi(x) = u$. In other words, compute $x = \Phi^{-1}(u)$.

(Since Φ^{-1} has no analytical solution, look up Box-Muller transform for further reading)

Today's plan

Normal approximation for Binomial

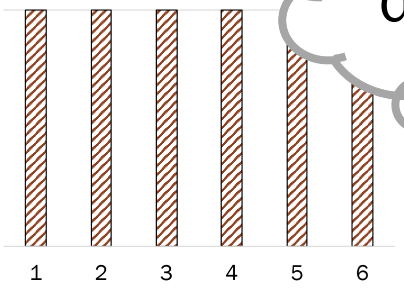
Joint distributions (discrete)

Multinomial Random Variable

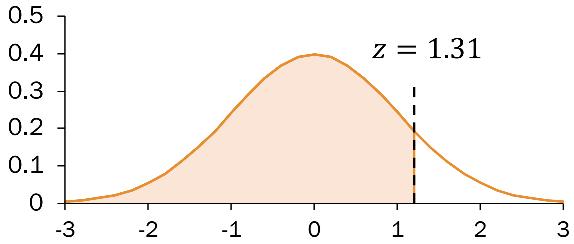
Text analysis

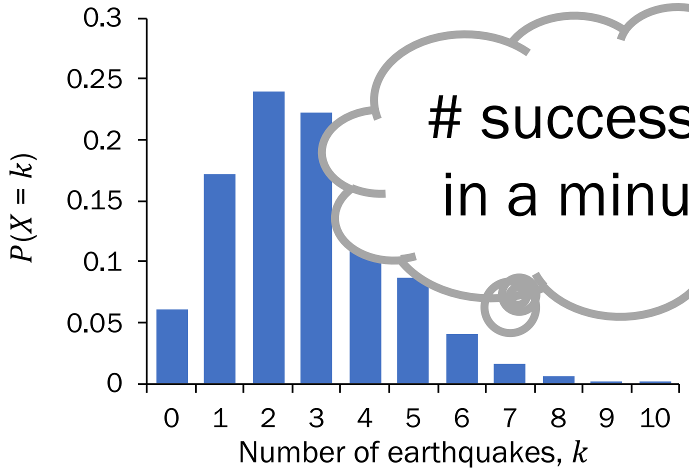
Joint distributions

So far, we have only worked with 1-dimensional random variables:



Single dice roll





successes in a minute

However, in the real world, events often occur with other events.

Outcomes on two dice rolls

2 successes in minute 1,
none in minutes 2-4,
3 successes in minute 5

ELO ratings

What is the probability that the Warriors win?
How do you model zero-sum games?

ELO ratings

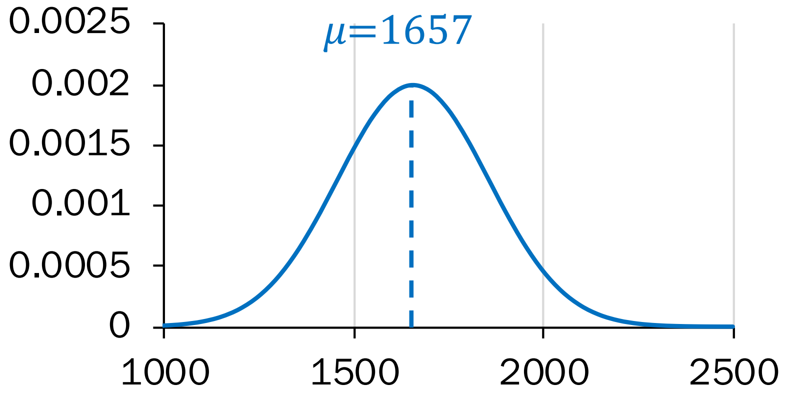
Review

Want: $P(\text{Warriors win}) = P(A_W > A_B)$

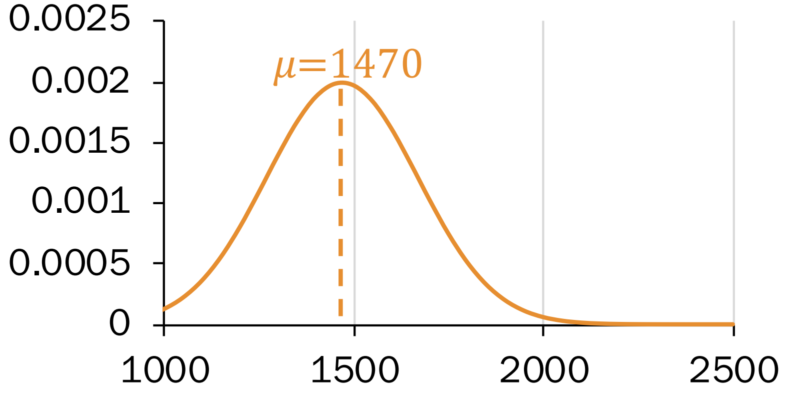
```python
from scipy import stats
WARRIORS_ELO = 1657
OPPONENT_ELO = 1470
STDEV = 200
NTRIALS = 10000

nSuccess = 0
for i in range(NTRIALS):
  w = stats.norm.rvs(WARRIORS_ELO, STDEV)
  b = stats.norm.rvs(OPPONENT_ELO, STDEV)
  if w > b:
    nSuccess += 1
print("Warriors sampled win fraction",
      float(nSuccess) / NTRIALS)
```

≈ 0.7488, calculated by sampling

Warriors $A_W \sim \mathcal{N}(S = 1657, 200^2)$



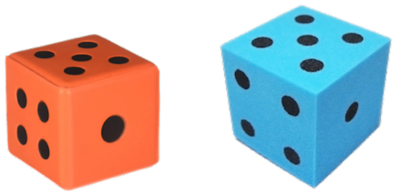
Opponents $A_B \sim \mathcal{N}(S = 1470, 200^2)$



CS109 Goal: Reason about probabilities involving multiple random variables.

Joint probability mass functions

Roll two 6-sided dice, yielding values X and Y.



X

random variable

$P(X = 1)$

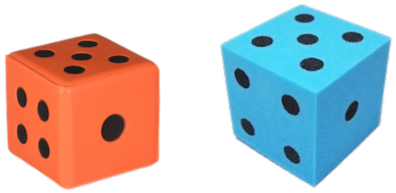
probability of an event

$P(X = k)$

probability mass function

Joint probability mass functions

Roll two 6-sided dice, yielding values X and Y.



X random variable	$P(X = 1)$ probability of an event	$P(X = k)$ probability mass function
X, Y random variables	$P(X = 1 \cap Y = 6)$ $P(X = 1, Y = 6)$ new notation: the comma probability of the intersection of two events	$P(X = a, Y = b)$ joint probability mass function

Discrete joint distributions

For two discrete joint random variables X and Y,
the joint probability mass function is defined as:

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

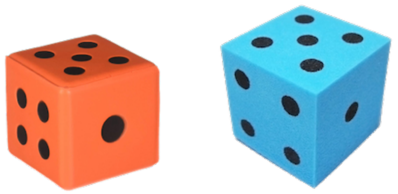
The marginal distributions of the joint PMF are defined as:

$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a, y) \qquad p_Y(b) = P(Y = b) = \sum_x p_{X,Y}(x, b)$$

👉 Use marginal distributions to get a 1-D RV from a joint PMF.

Two dice

Roll two 6-sided dice, yielding values X and Y.



1. What is the joint PMF of X and Y?

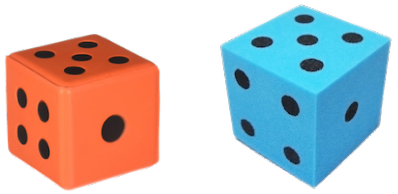
$$p_{X,Y}(a,b) = 1/36 \qquad (a,b) \in \{(1,1),\ldots,(6,6)\}$$

2. What is the marginal PMF of X?

$$p_X(a) = P(X=a) = \sum_y p_{X,Y}(a,y) = \sum_{y=1}^{6} \frac{1}{36} = \frac{1}{6} \qquad a \in \{1,\ldots,6\}$$

Two dice

Roll two 6-sided dice, yielding values X and Y.



1. What is the joint PMF of X and Y? $p_{X,Y}(a,b) = 1/36$

		X					
		1	2	3	4	5	6
	1	1/36	…	…	…	…	1/36
	2	…	…	…	…	…	…
	3	…	…	…	…	…	…
Y	4	…	…	…	…	…	…
	5	…	…	…	…	…	…
	6	1/36	…	…	…	…	1/36

$P(X = 4, Y = 2)$

Probability table

- All possible outcomes for several discrete RVs
- Not parametric (e.g., parameter p in Ber(p))

2. What is the marginal PMF of X?

Break for jokes/ announcements

Announcements

Concept checks

Due date: Tuesdays 1:00pm

Selected anonymous answers (with consent)

Late days

Free: 2 free class days

No late days after last day of quarter (Fri 12/7)

(note PS#6 due Wed 12/5)

Problem Set 1

Problem 16 solutions posted

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

What is $P(X = 1, Y = 0)$, the missing entry in the probability table?

A. $1 - (.16 + .12 + .07 + \cdots + .04) = 0.12$

B. $.24 - (.12) = 0.12$

C. $0.5(.24) = 0.12$

D. All of the above

E. None/other

X (# Macs)

Y (# PCs)	0	1	2	3
0	.16	?	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

What is $P(X = 1, Y = 0)$, the missing entry in the probability table?

A. $1 - (.16 + .12 + .07 + \cdots + .04) = 0.12$

B. $.24 - (.12) = 0.12$

C. $0.5(.24) = 0.12$

D. All of the above

E. None/other

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	?	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

A joint probability table must sum to 1.

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

Which entries in the probability table correspond to $P(C = 3)$?

A.

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

B.

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

C.

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

Which entries in the probability table correspond to $P(C = 3)$?

A.

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

B.

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

C.

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

How do you compute $P(X = 0, Y = 3)$?

$P(X = 0, Y = 3)$

Law of Total Probability $= P(X = 0, Y = 3 | C = 3)P(C = 3) + P(X = 0, Y = 3 | C \neq 3)P(C \neq 3)$

Bin(n=3,p=0.5) $= \binom{3}{0} 0.5^0 0.5^3 \cdot (0.32) + 0$

$= 0.04$

Y (# PCs) \ X (# Macs)	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

Which entries in the probability table correspond to the marginal PMF of X?

A.

Y (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	
1	.12	.14	.12	0	
2	.07	.12	0	0	
3	.04	0	0	0	

B.

Y (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	
1	.12	.14	.12	0	
2	.07	.12	0	0	
3	.04	0	0	0	
	Sum rows here				

C.

Y (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	sum cols here
1	.12	.14	.12	0	
2	.07	.12	0	0	
3	.04	0	0	0	

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a Mac or PC.

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

Which entries in the probability table correspond to the marginal PMF of X?

A.

Y (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	
1	.12	.14	.12	0	
2	.07	.12	0	0	
3	.04	0	0	0	

B.

Y (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	
1	.12	.14	.12	0	
2	.07	.12	0	0	
3	.04	0	0	0	
	Sum rows here				

C.

Y (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	sum cols here
1	.12	.14	.12	0	
2	.07	.12	0	0	
3	.04	0	0	0	

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has C computers, where $C = X$ Macs + Y PCs.
- Each computer in a household is equally likely to be a 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 (# PCs) \ X (# Macs)	0	1	2	3	
0	.16	.12	.07	.04	.39
1	.12	.14	.12	0	.38
2	.07	.12	0	0	.19
3	.04	0	0	0	.04
	.39	.38	.19	.04	

Marginal PMF of Y,

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

Marginal PMF of X, $p_X(x) = \sum_y p_{X,Y}(x, y)$

To find a marginal distribution over one variable, sum over all other variables.

Today's plan

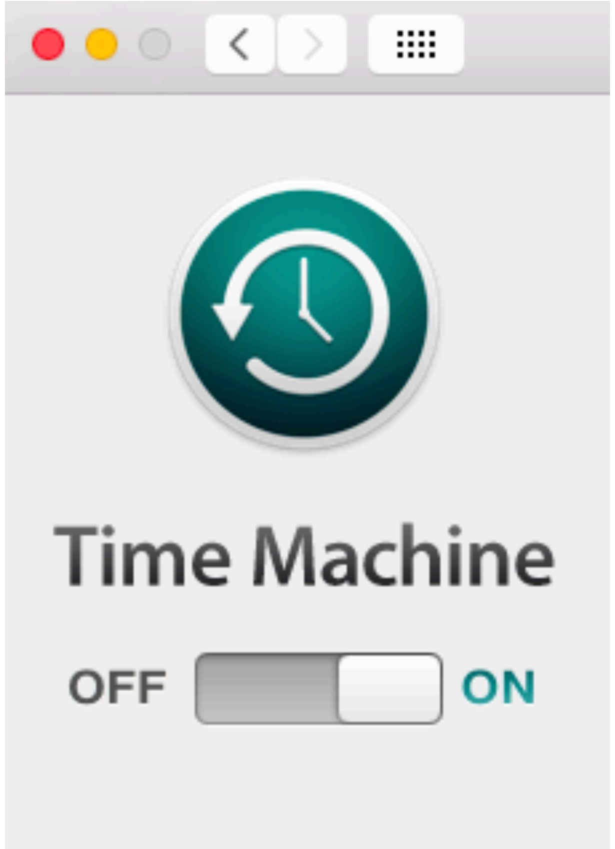
Normal approximation for Binomial

Joint distributions (discrete)

Multinomial Random Variable

Text analysis

Recall the good times



Permutations

$n!$

How many ways are there to order n objects?

Counting unordered objects

Binomial coefficient

How many ways are there to group n objects into two groups of size k and $n-k$, respectively?

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

Called the binomial coefficient because of something from Algebra

Multinomial coefficient

How many ways are there to group n objects into r groups of sizes n_1, n_2, …, n_r respectively?

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

👉 Multinomial generalizes Binomial for counting.

Probability

Binomial RV

What is the probability of getting k successes and $n - k$ failures in n trials?

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Binomial # of ways of ordering the successes

Probability of each ordering of k successes is equal + mutually exclusive

Multinomial RV

What is the probability of getting c_1 of outcome 1, c_2 of outcome 2, …, and c_m of outcome m in n trials?

Multinomial RVs also generalize Binomial RVs for probability!

Multinomial Random Variable

Consider an experiment:

- n independent trials
- Each trial results in one of m outcomes with respective probabilities $p_1, p_2, \ldots, p_m$ where $\sum_{i=1}^{m} p_i = 1$
- Let $X_i =$# of trials with outcome i.

def A **Multinomial** random variable X is defined as follows:

Joint PMF

$$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} \boxed{p_1^{c_1} p_2^{c_2} \cdots p_m^{c_m}}$$

Multinomial # of ways of ordering the outcomes

Probability of each ordering is equal + mutually exclusive

where $\sum_{i=1}^{m} c_i = n$ and $\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1! c_2! \cdots c_m!}$ is a multinomial coefficient

Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes

Strategy (choose all that apply):

A. Law of total probability
B. Counting
C. Multinomial RV
D. Binomial RV
E. None/other

🤔

Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes

Strategy (choose all that apply):

A. Law of total probability
B. Counting
C. Multinomial RV
D. Binomial RV
E. None/other

Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes

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

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

Today's plan

Normal approximation for Binomial

Joint distributions (discrete)

Multinomial Random Variable

Text analysis

Probabilistic text analysis

Ignoring the order of words…

What is the probability of any given word that you write in English?

- P(word = "the") $>$ P(word = "pokemon")
- P(word = "Stanford") $>$ P(word = "Cal")

Probabilities of *counts* of words = Multinomial distribution 👉

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.)

Probabilistic text analysis

Probabilities of *counts* of words = Multinomial distribution

Example document:

#words: $n = 48$

"When my late husband was alive he deposited some amount of Money with china Bank in which the amount will be declared to you once you respond to this message indicating your interest in helping to receive the fund and use it for Gods work as my wish."

$$P\left(\begin{array}{l}\text{bank} = 1\\ \text{fund} = 1\\ \text{money} = 1\\ \text{wish} = 1\\ \dots\\ \text{to} = 3\end{array}\middle|\,\text{spam}\right) = \frac{n!}{1!\,1!\,1!\,1!\cdots 3!}\, p_{\text{bank}}^{1}\, p_{\text{fund}}^{1} \cdots p_{\text{to}}^{3}$$

Note: $P\left(\text{bank}\,\middle|\,\begin{array}{c}\text{spam}\\ \text{writer}\end{array}\right) \gg P\left(\text{bank}\,\middle|\,\begin{array}{c}\text{writer=}\\ \text{you}\end{array}\right)$

Probabilistic text analysis

Probabilities of *counts* of words = Multinomial distribution

What about probability of those same words in someone else's writing?

- $P\left(\text{word = “probability”} \,\middle|\, \begin{matrix}\text{writer =} \\ \text{you}\end{matrix}\right) > P\left(\text{word = “probability”} \,\middle|\, \begin{matrix}\text{writer =} \\ \text{non−CS109 student}\end{matrix}\right)$

To determine authorship:

1. Estimate $P(\text{word}|\text{writer})$ from known writings
2. Use Bayes' Theorem to determine $P(\text{writer}|\text{document})$ for a new writing!

Who wrote The Federalist Papers?

Old and New Analysis

Authorship of The Federalist Papers

- 85 essays advocating ratification of the US constitution
- Written under the pseudonym “Publius” (really, Alexander Hamilton, James Madison, John Jay)

Who wrote which essays?

- Analyze probability of words in each essay and compare against word distributions from known writings of three authors

Let’s write a program!