11: Joint (Multivariate) Distributions

Jerry Cain

April 24th, 2024

Lecture Discussion on Ed

Normal Approximation

Normal Random Variables

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

mean: μ variance: σ^2

informally, this means that as a distribution, it commits to as little structure as possible (though just enough so that the mean and the variance are well-defined)

- Used to model many real-life situations because it maximizes entropy (i.e., randomness) for a given mean and variance.

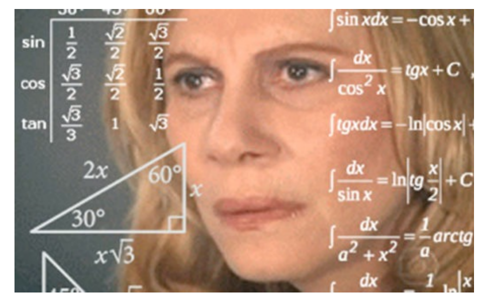
this is about to be discussed right now!

- Also useful for approximating the Binomial random variable!

Website testing

- 100 people are presented with a new website design.
- X = # people whose time on site increases
- PM assumes design has no effect, so assume $\underbrace{\text{P(stickier) = 0.5 independently}}_{\text{this operates as an a priori belief that both designs are equally effective.}}$.
- CEO will endorse the new design if $X \geq 65$.

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



Approach 1: Binomial

Define

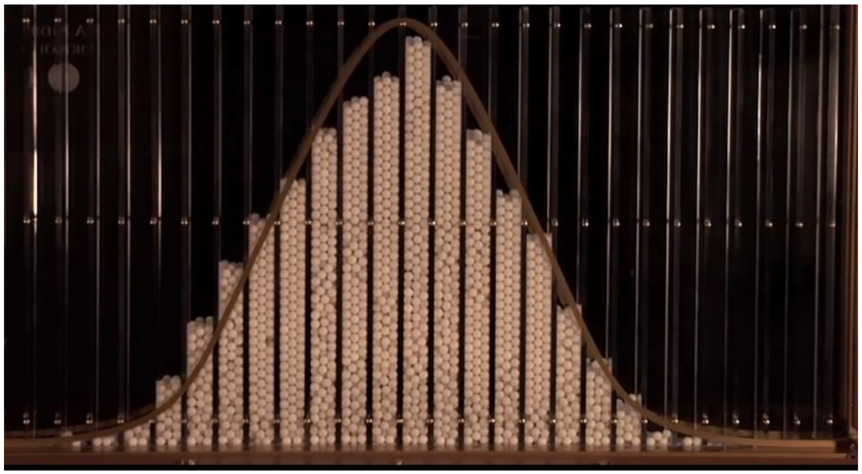
$$X \sim \text{Bin}(n = 100, p = 0.5)$$

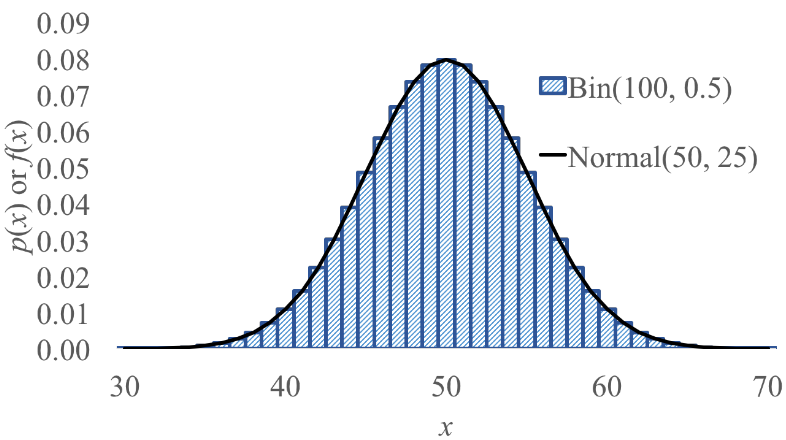
Want: $P(X \geq 65)$

Solve

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

Don't worry, Normal approximates Binomial





Galton Board

(We'll explain **why** in 2 weeks)

Website testing

```
jerry$ python
>>> from scipy.stats import binom, norm
>>> binom.pmf(range(65, 101), n, p).sum()
0.0017588208614850 58
>>> 1 - norm(50, 5).cdf(65)
0.0013498980316301035
```

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

What is P(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

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

computed using python to evaluate that sum.

Approach 2: approximate with Normal

Define

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

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

derived from Binomial, used as params for Gaussian

Solve

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

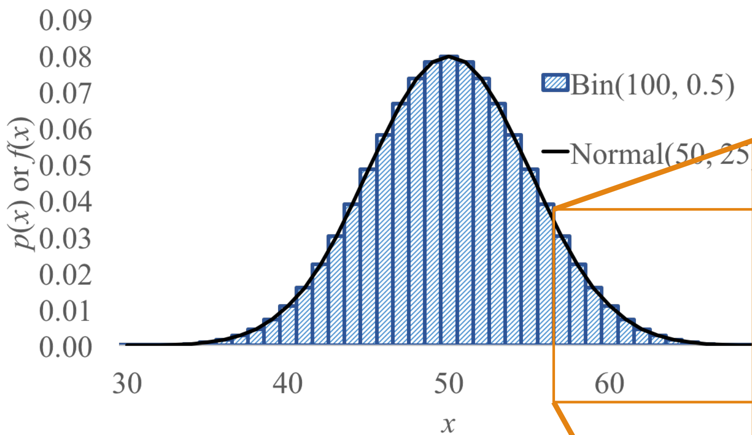
$$= 1 - \Phi\left(\frac{65-50}{5}\right) = 1 - \Phi(3) \approx 0.0013 \text{ ?}$$

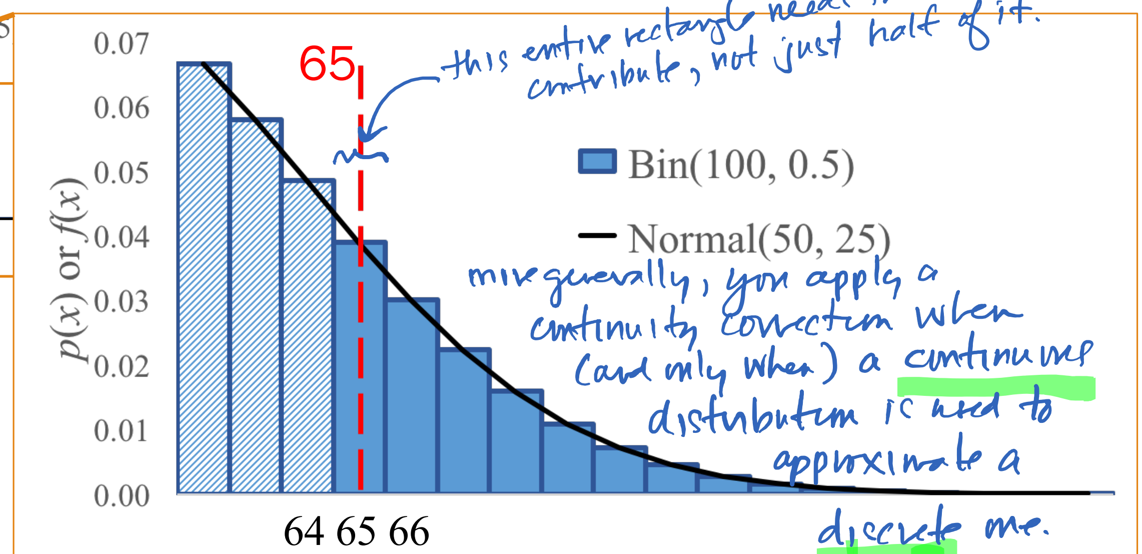
not all that great! what's up?

⚠️ ⚠️ (this approach is missing something important)

Website testing (with continuity correction)

In our website testing, $Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$.





this entire rectangle needs to contribute, not just half of it.

more generally, you apply a continuity correction when (and only when) a continuous distribution is used to approximate a discrete one.

$P(X \geq 65)$ Binomial

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

≈ 0.0018 ✅ the better

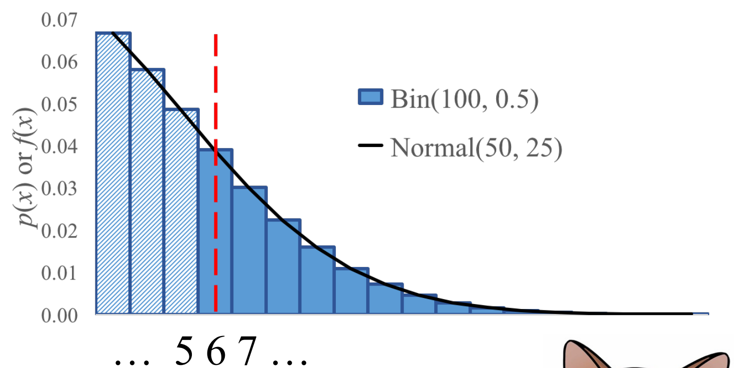
Approach 2

You must perform a **continuity correction** when approximating a Binomial RV with a Normal RV.

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?

those underlined in green should include X=6 when approximating.

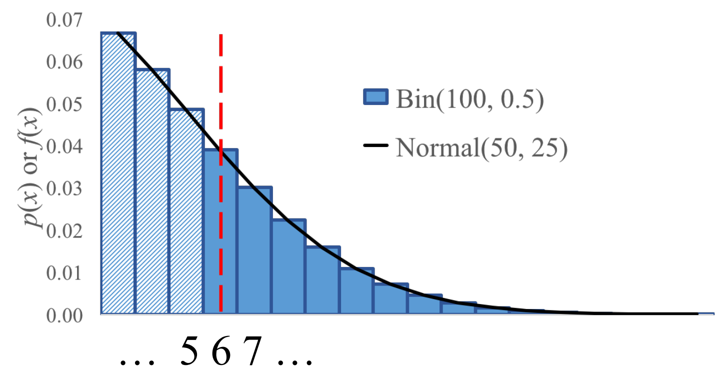
those underlined in red want nothing to do with X=6, so be sure to exclude its contribution when approximating.

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

means ≥ 7

means ≤ 5

helpful to frame inequalities in terms of $\le$ and $\ge$



Who gets to approximate?

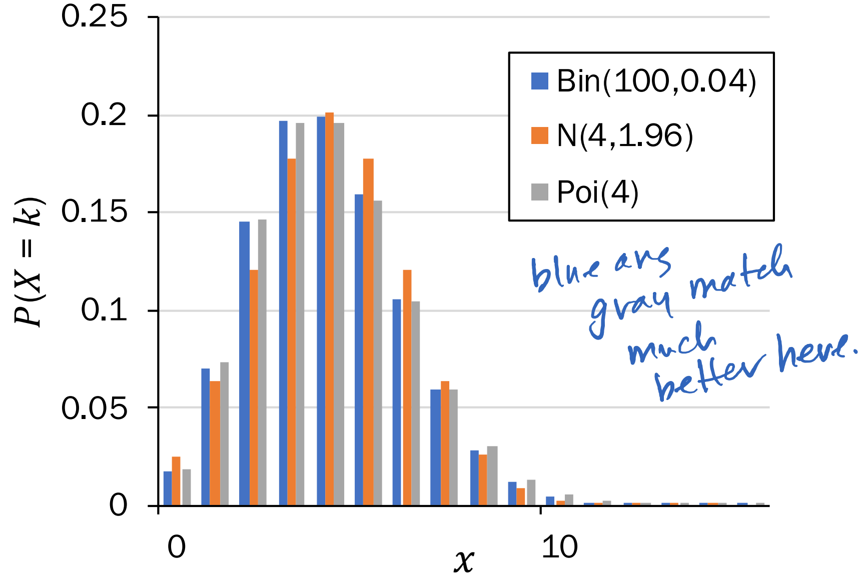
$$X \sim \text{Bin}(n, p)$$
$$E[X] = np$$
$$\text{Var}(X) = np(1-p)$$

$$Y \sim \text{Poi}(\lambda)$$
$$\lambda = np$$

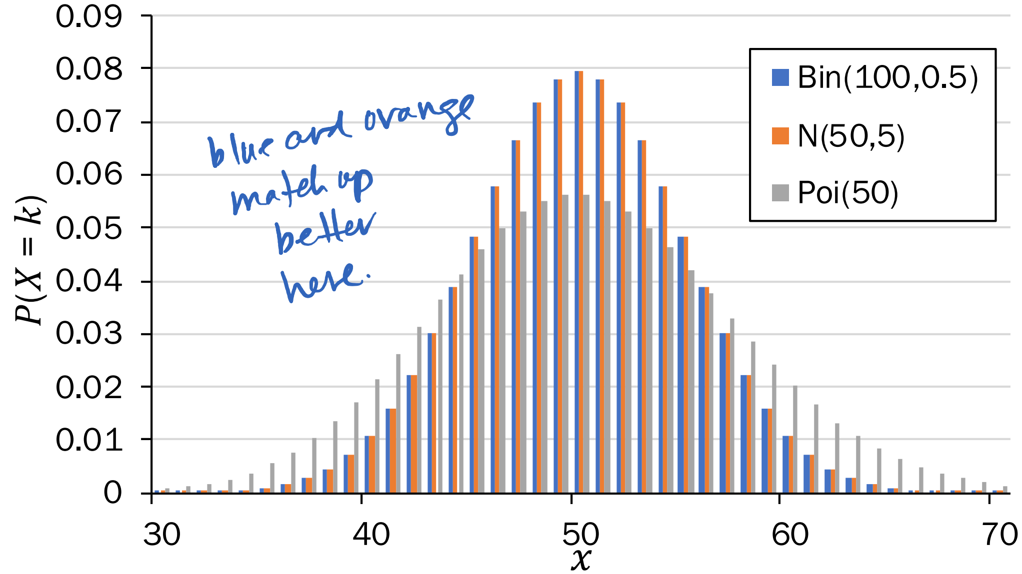
?

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$
$$\mu = np$$
$$\sigma^2 = np(1-p)$$

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 Gaussian to approximate.
2. When using Normal to approximate a discrete RV, use a continuity correction.

Stanford Admissions (a while back)

Stanford accepts 2480 students.

- Each admitted student matriculates with p = 0.68 (independently)
- 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 admitted student matriculates with p = 0.68 (independently)
- Let X = # of students who will attend

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

Strategy:

A. Just Binomial — computationally expensive (also not an approximation)

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

$$P(Y \geq 1745.5) = 1 - F(1745.5)$$

$$= 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right)$$

$$= 1 - \Phi(2.54) \approx 0.0055$$

Discrete Joint RVs

From last slide deck

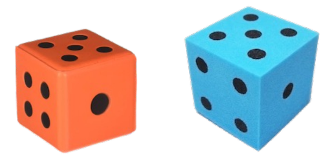
Review

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

$$P(A_W > A_B)$$

This is a probability of an event involving **two** random variables!

Joint probability mass functions



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

X

random variable

$P(X = 1)$

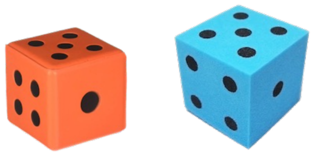
probability of an event

$P(X = k)$

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:

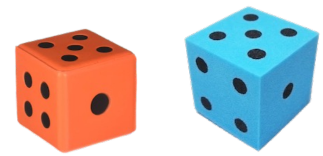
$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a,y)$$

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

Use marginal distributions to extract a 1D 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?

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

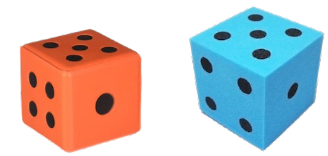
		X					
		1	2	3	4	5	6
Y	1	1/36	…	…	…	…	1/36
	2	…	…	…	…	…	…
	3	…	…	…	…	…	…
	4	…	…	…	…	…	…
	5	…	…	…	…	…	…
	6	1/36	…	…	…	…	1/36

$P(X = 4, Y = 3)$

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

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?

note that x is constrained to be a everywhere, whereas y varies over its full support.

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

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has X Macs and Y PCs.
- Each house has a maximum of 3 computers total (Macs + PCs).

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

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

symmetry above the diagonal is incidental, not a requirement.

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has X Macs and Y PCs.
- Each house has a maximum of 3 computers total (Macs + PCs).

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

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 joint PMF must sum to 1:

$$\sum_x \sum_y p_{X,Y}(x, y) = 1$$

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has X Macs and Y PCs.
- Each house has a maximum of 3 computers total (Macs + PCs).

2. How do you compute the marginal PMF of X?

Y (# PCs) \ X (# Macs)	0	1	2	3	sum rows here
0 A	.16	.12	.07	.04	.39
1	.12	.14	.12	0	38
2	.07	.12	0	0	.19
3	.04	0	0	0	.04
B (sum cols here)	.39	.38	.19	.04	C

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has X Macs and Y PCs.
- Each house has a maximum of 3 computers total (Macs + PCs).

2. How do you compute the marginal PMF of X?

Y (# PCs) \ X (# Macs)	0	1	2	3	sum rows here
0 A	.16	.12	.07	.04	.39
1	.12	.14	.12	0	.38
2	.07	.12	0	0	.19
3	.04	0	0	0	.04
B sum cols here	.39	.38	.19	.04	

C

A. $p_{X,Y}(x,0) = P(X = x, Y = 0)$

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

C. Marginal PMF of Y $\quad p_Y(y) = \sum_x p_{X,Y}(x,y)$

To find a marginal distribution over one variable, sum over all other variables in the joint PMF.

A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has X Macs and Y PCs.
- Each house has a maximum of 3 computers total (Macs + PCs).

3. Let $C = X + Y$. What is $P(C = 3)$?

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 X Macs and Y PCs.
- Each house has a maximum of 3 computers total (Macs + PCs).

3. Let $C = X + Y$. What is $P(C = 3)$?

these conditional probabilities are all either 0 (when $x+y \neq 3$) or 1 (when $x+y=3$)

X (# Macs) across the columns, Y (# PCs) down the rows:

Y \ X	0	1	2	3
0	.16	.12	.07	.04
1	.12	.14	.12	0
2	.07	.12	0	0
3	.04	0	0	0

$$P(C = 3) = P(X + Y = 3)$$

Law of Total Probability

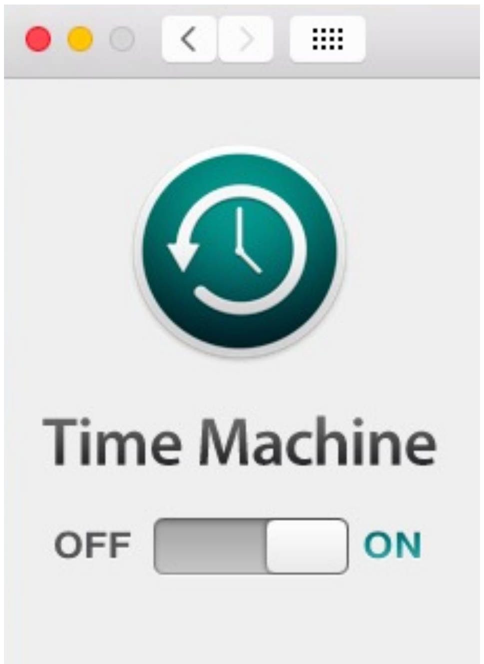
$$= \sum_x \sum_y P(X + Y = 3 | X = x, Y = y) P(X = x, Y = y)$$

$$= P(X = 0, Y = 3) + P(X = 1, Y = 2) + P(X = 2, Y = 1) + P(X = 3, Y = 0) = 0.32$$

We'll come back to sums of RVs next lecture!

Multinomial RV

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?

assume, of course, that $n_1 + n_2 + \cdots + n_r = n$

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

assumes the groups are labeled.

Multinomials generalize Binomials 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 of n independent trials:

- Each trial results in one of m outcomes. $P(\text{outcome } i) = p_i$, $\sum_{i=1}^{m} p_i = 1$
- Let X_i= # trials with outcome i

Joint PMF

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

where $\sum_{i=1}^{m} c_i = n$ and $\sum_{i=1}^{m} p_i = 1$

generalization of Binomial

Multinomial # of ways of ordering the outcomes

Probability of each ordering is equal + mutually exclusive

Hello dice rolls, my old friends

A fair, six-sided die is rolled 7 times.

What is the probability of getting:

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

Hello dice rolls, my old friends

A fair, six-sided die is rolled 7 times.

What is the probability of getting:

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

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

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

Hello dice rolls, my old friends

A fair, six-sided die is rolled 7 times.

What is the probability of getting:

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

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, \overbrace{X_6 = 3}^{\text{\# of times a six appears}})$$

$$= \binom{7}{1,1,0,2,0,\underbrace{3}_{\text{choose where the sixes appear}}} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^0 \left(\underbrace{\frac{1}{6}}_{\text{probability of rolling a six}}\right)^{\overset{\text{this many times}}{3}} = 420 \left(\frac{1}{6}\right)^7$$

Probabilistic text analysis

Ignoring the order of words…

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

- $P(\text{word} = \text{"the"}) > P(\text{word} = \text{"susurration"})$
- $P(\text{word} = \text{"Stanford"}) > P(\text{word} = \text{"Cal"})$

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 overseas 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 Heavens work as my wish."

$$P\left(\begin{array}{l} \text{bank} = 1 \\ \text{fund} = 1 \\ \text{money} = 1 \\ \text{wish} = 1 \\ \dots \\ \boxed{\text{to} = 3} \end{array} \middle| \text{spam} \right) = \frac{48!}{1!\,1!\,1!\,1!\cdots \boxed{3!}} p_{\text{bank}}^{1} p_{\text{fund}}^{1} \cdots p_{\text{to}}^{\boxed{3}}$$

not concerned with order or placement within document! just frequencies.

Note: $P(\text{bank}|\text{spam}) \gg P(\text{bank}|\text{legit})$

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
- Curious what the analysis is? Read this!

Statistics of Two RVs

Expectation and Covariance

In real life, we often have many RVs interacting at once.

- We've seen some simpler cases (e.g., sum of independent Bernoullis).
- Come Friday, we'll discuss sums of Binomials, Poissons, etc.
- In general, manipulating joint PMFs is difficult.
- Fortunately, **you don't need to model** joint RVs completely all the time.

Instead, we'll focus next on reporting **statistics** of multiple RVs:

- **Expectation of sums** (you've seen some of this, more on Friday)
- **Covariance**: measure of how two random variable vary with each other (more next Monday and Wednesday)

Properties of Expectation, extended to two RVs

1. Linearity:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

2. Expectation of a sum = sum of expectation:

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

we've seen this!
we'll prove momentarily.

3. Unconscious statistician:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

True for both independent and dependent random variables!

Proof of expectation of a sum of RVs

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

$$E[X+Y] = \sum_x \sum_y (x+y) p_{X,Y}(x,y)$$

LOTUS,
$g(X,Y) = X + Y$

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

switch ☺

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

Linearity of summations
(cont. case: linearity of integrals)

$$= \sum_x x p_X(x) + \sum_y y p_Y(y)$$

Marginal PMFs for X and Y

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