11: Joint (Multivariate) Distributions

Jerry Cain

April 20, 2022

Table of Contents

2 Normal Approximation
12 Discrete Joint RVs
25 Multinomial RVs
35 Statistics of Two RVs

Normal Approximation

Normal RVs

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

mean: μ, variance: σ^2

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

- 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 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

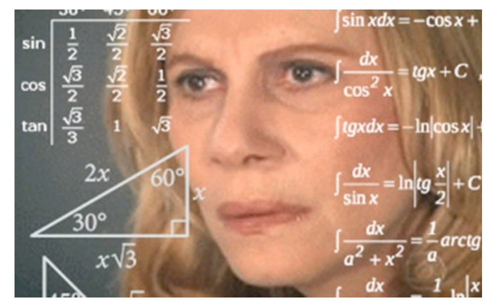
Define

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

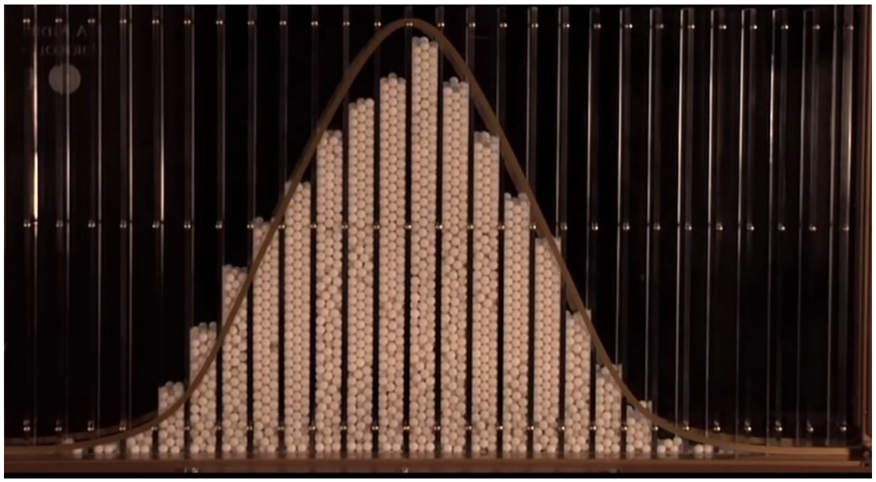
Want: $P(X \geq 65)$

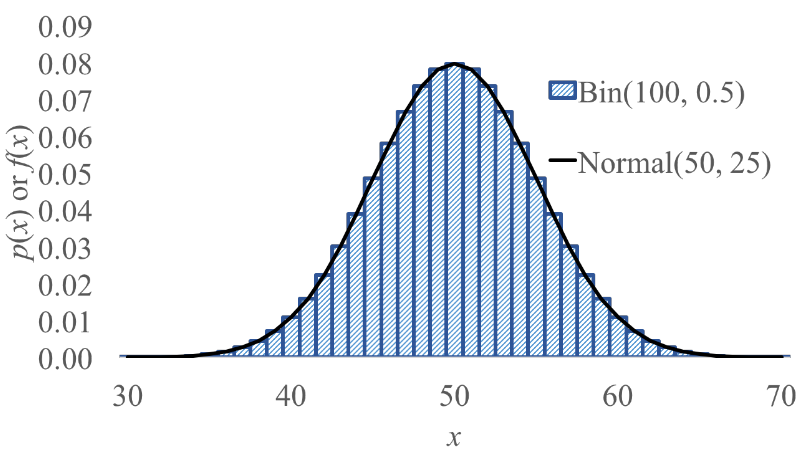
Solve

$$P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} 0.5^i (1 - 0.5)^{100-i} \approx 0.0018$$



Don’t worry, Normal approximates Binomial

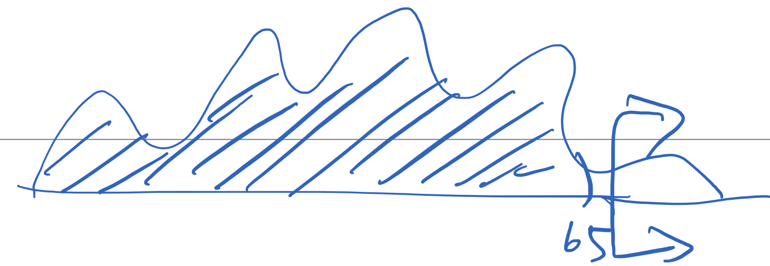




Galton Board

(We’ll explain ***why*** in 2 weeks)

Website testing



- 100 people are given a new website design.
- $X = \#$ people whose time on site increases
- The design actually has no effect, so P(time on site increases) = 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

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

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

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

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

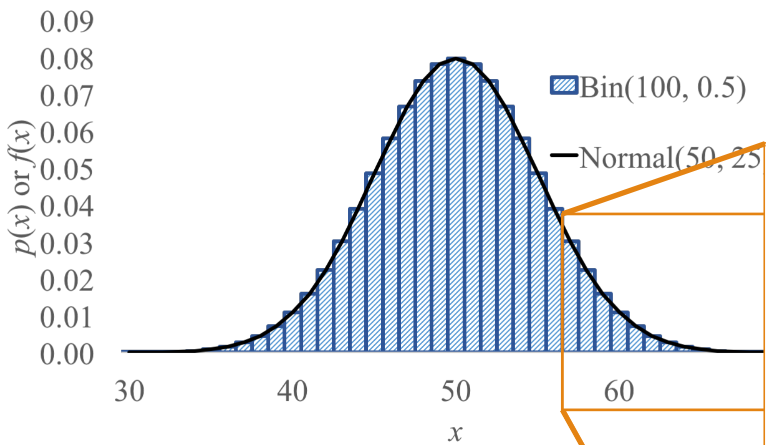
(this approach is missing something important)

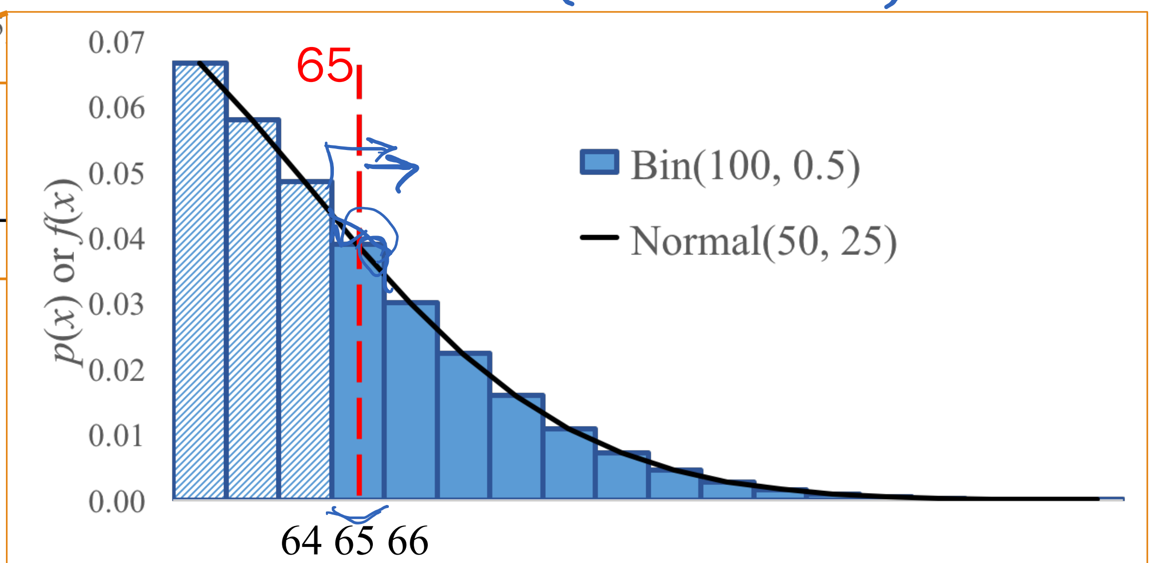
Website testing (with continuity correction)

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

$P(X = 65) \overset{??}{=} P(Y = 65)$

$\Rightarrow P(64.5 \le Y \le 65.5)$





$P(X \geq 65)$ Binomial

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

≈ 0.0018

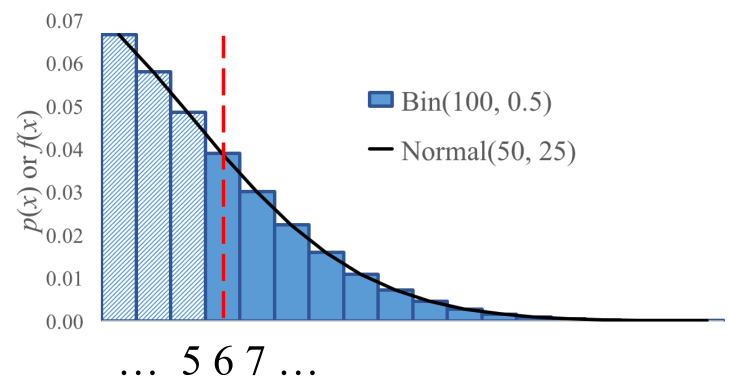
the better Approach 2

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?

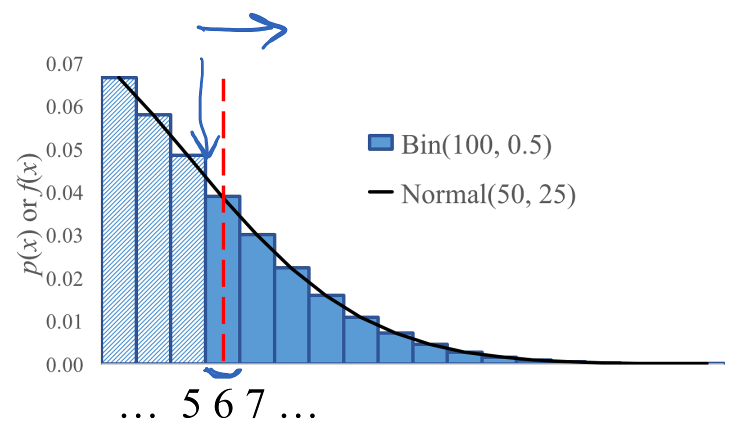
Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	
$P(X \geq 6)$	
$P(X > 6)$	
$P(X < 6)$	
$P(X \leq 6)$	



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?

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



Who gets to approximate?

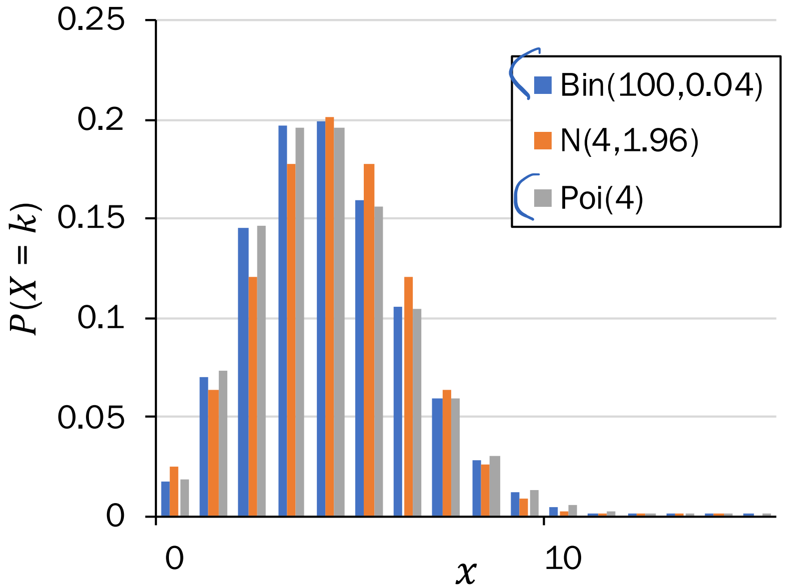
$$X \sim \text{Bin}(n, p)$$
$$E[X] = np$$
$$\text{Var}(X) = np(1-p)$$

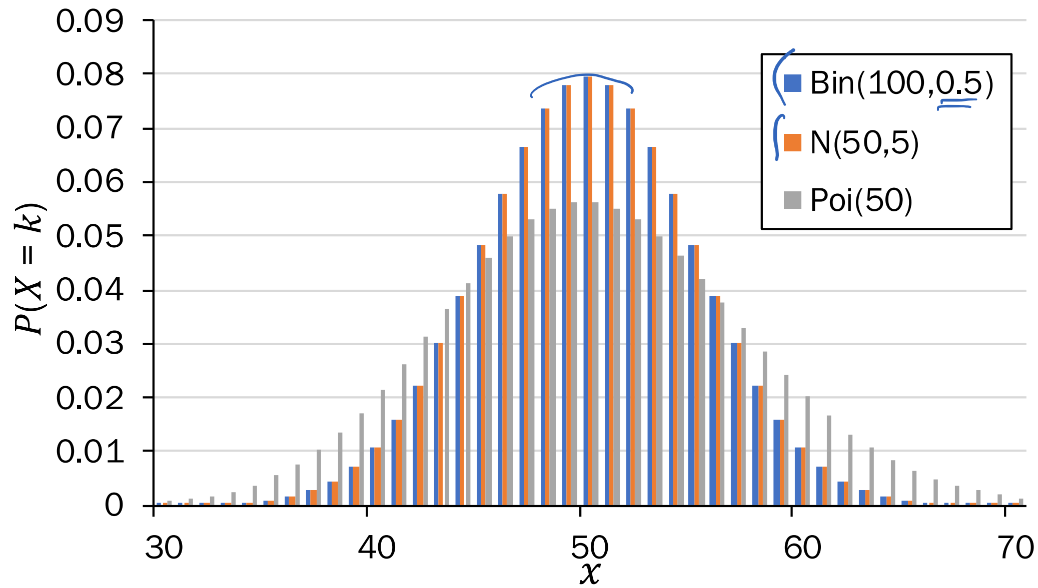
?

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

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

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

Stanford Admissions (a while back)

based on info late 90's

Stanford accepts 2480 students.

- Each admitted student matriculates w.p. 0.68 (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)

$\binom{2480}{1746}(.68)^{1746}(.32)^{2480-1746}$

Stanford accepts 2480 students.

- Each admitted student matriculates w.p. 0.68 (independent trials)
- 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 $Y \sim \mathcal{N}(E[X], \text{Var}(X))$

$E[X] = np = 1686$

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

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

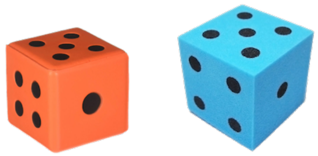
What is the probability that the Warriors win?
How do you model zero-sum games?

$$P(A_W > A_B)$$

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

Joint probability mass functions

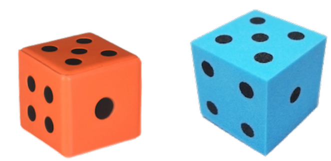
Roll two 6-sided dice, yielding values X and Y.



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:

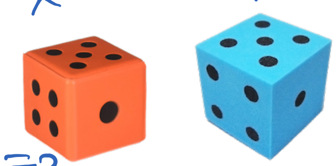
$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a,y)$$

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

Use marginal distributions to 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?

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

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

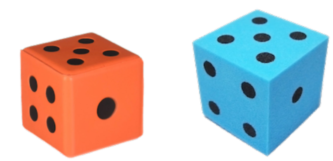
$P(X = 4, Y = 2)$

Probability table

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

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?

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

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 (Macs + PCs) in the house.

.12

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

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 (Macs + PCs) in the house.

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 (Macs + PCs) in the house.

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

Y (# PCs) \ X (# Macs)	0	1	2	3	sum cols here (C)
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 rows here)	.39	.38	.19	.04	

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 (Macs + PCs) in the house.

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

Y (# PCs) \ X (# Macs)	0	1	2	3	sum cols 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 rows 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 (Macs + PCs) in the house.

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 (Macs + PCs) in the house.

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

	X (# Macs) 0	1	2	3
Y (# PCs) 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

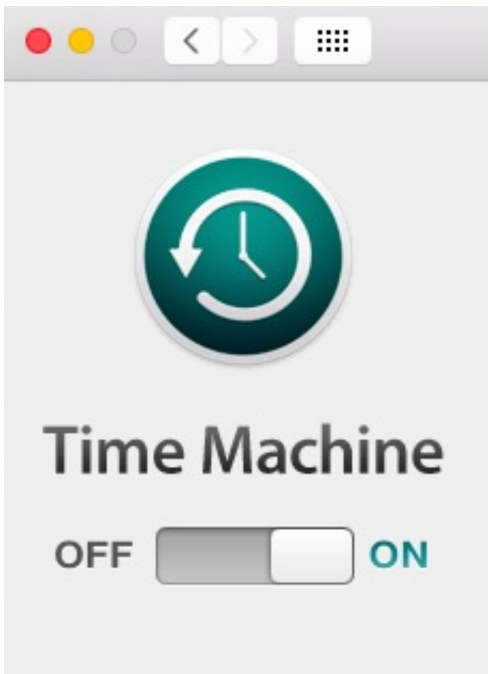
$$= \sum_x \sum_y P(X + Y = 3 | X = x, Y = y) P(X = x, Y = y)$$

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

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?

$\binom{n}{k, n-k}$

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

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 $n_1, n_2, \ldots, n_r$ respectively?

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

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 $\sum_{i=1}^{m} p_i = 1$

Multinomial # of ways of ordering the outcomes

Probability of each ordering is equal + mutually exclusive

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

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

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

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

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

# of times a six appears

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

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

choose where the sixes appear

probability of rolling a six

this many times

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, sums of Poissons, etc.
- Computing joint PMFs in general is hard!
- 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 RVs vary with each other (more on this next week)

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 today!)

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

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

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

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

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

Expectations of common RVs: Binomial

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

# of successes in n independent trials with probability of success p

Recall: $\text{Bin}(1, p) = \text{Ber}(p)$

$$X = \sum_{i=1}^{n} X_i$$

Let $X_i = i$th trial is heads
$X_i \sim \text{Ber}(p), E[X_i] = p$

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$