Computer Science

CSC380: Principles of Data Science

Probability 4

Chicheng Zhang

Quiz 5

A company operates a customer support hotline, and the time (in minutes) a customer waits before speaking to an agent X follows an exponential distribution,

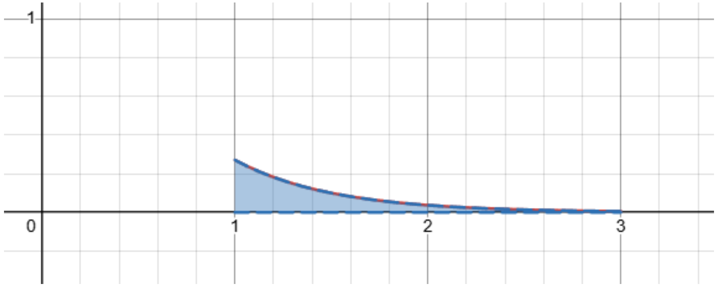
$$f(x) = 2e^{-2x}, x \geq 0$$

What is the probability that a randomly chosen customer waits between 1 and 3 minutes before being connected to an agent? S

(Hint: the antiderivative of $2e^{-2x}$ is $-e^{-2x}$)

Quiz 5

- The question asks for $P(1 \leq X \leq 3)$

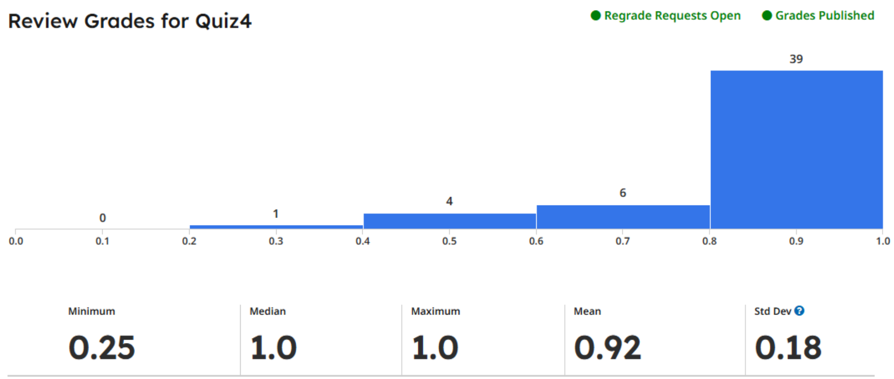


- Integrating PDF, this is

$$\int_1^3 2e^{-2x}dx = -e^{-2x}\Big|_1^3$$
$$= (-e^{-6}) - (-e^{-2})$$
$$= e^{-2} - e^{-6}$$
$$= 0.132$$

Announcements 2/26

- Quiz 4 graded



- Check scores on D2L about Quizzes 1-4 and HW1-3
  - Let us know (by Piazza private post) if your scores are missing / wrong

Announcements 2/26

HWs:

- Merging PDFs
  - We recommend using Adobe Creative Cloud (UA access no cost)

- Sometimes important clarifications on HW questions may be on Piazza, keep a lookout for it..

- We recommend checking out the HW timeline in the syllabus

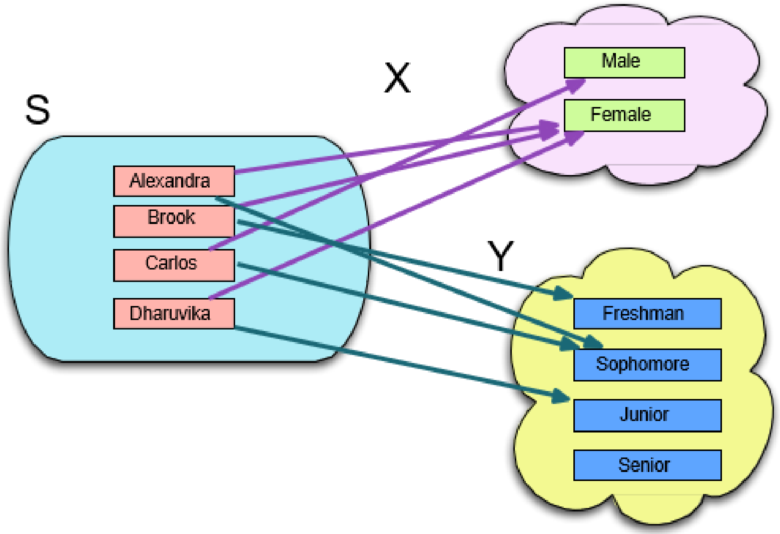
- Generally, writing down more steps help us giving more credits to your solutions more robustly

Outline

- Multivariate Random Variables
  - Joint distribution vs. Marginal distribution
  - Independence of RVs
- Expectation and Variance Revisited
  - Covariance, correlation
- Example multivariate RVs
- Law of Large Numbers
- Central Limit Theorem

Multivariate Random Variables

Multivariate RVs: example



- X: people -> their genders
- Y: people -> their class year
- We'd like to answer questions such as: does X and Y have a correlation?
  - I.e., is a student in higher class year more likely to be male?
- We call (X, Y) a random vector, or a multivariate RV, and will study its *joint* distribution

Joint distribution of discrete RVs

- The joint PMF (probability mass function) of discrete random variables X, Y:

$$f(x,y) = P(X = x, Y = y)$$

Examples

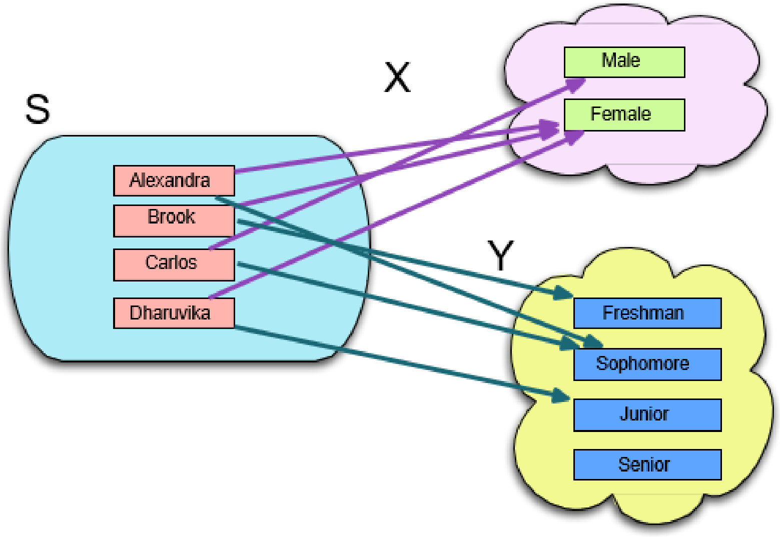
Alexandra

$$P(X = \text{Fem}, Y = \text{Soph}) = \frac{1}{4}$$

Dharuvika

$$P(X = \text{Fem}, Y = \text{Jun}) = \frac{1}{4}$$

…



Joint distribution of discrete RVs

- X: # of cars owned by a randomly selected household
- Y: # of computers owned by the same household
- Joint pmf shown with a table

	y			
x	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

- Probability that a randomly selected household has ≥ 2 cars and ≥ 2 computers?
  - $P(X \geq 2, Y \geq 2) = 0.5$

Marginal distributions

- Given joint distribution of (X, Y), need distribution of one, say X.
- Such a distribution is called the *marginal distribution* of X.

- How to find $P(X = x)$?
- Using law of total probability:

$$f_1(x) = \sum_y f(x, y)$$

	y			
x	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

- This operation is called *marginalization* ('marginalizing out variable Y', or variable elimination)

Marginal distributions

x	y = 1	2	3	4	Total
1	0.1	0	0.1	0	0.2
2	0.3	0	0.1	0.2	0.6
3	0	0.2	0	0	0.2
Total	0.4	0.2	0.2	0.2	1.0

f_1: *marginal distribution* of X

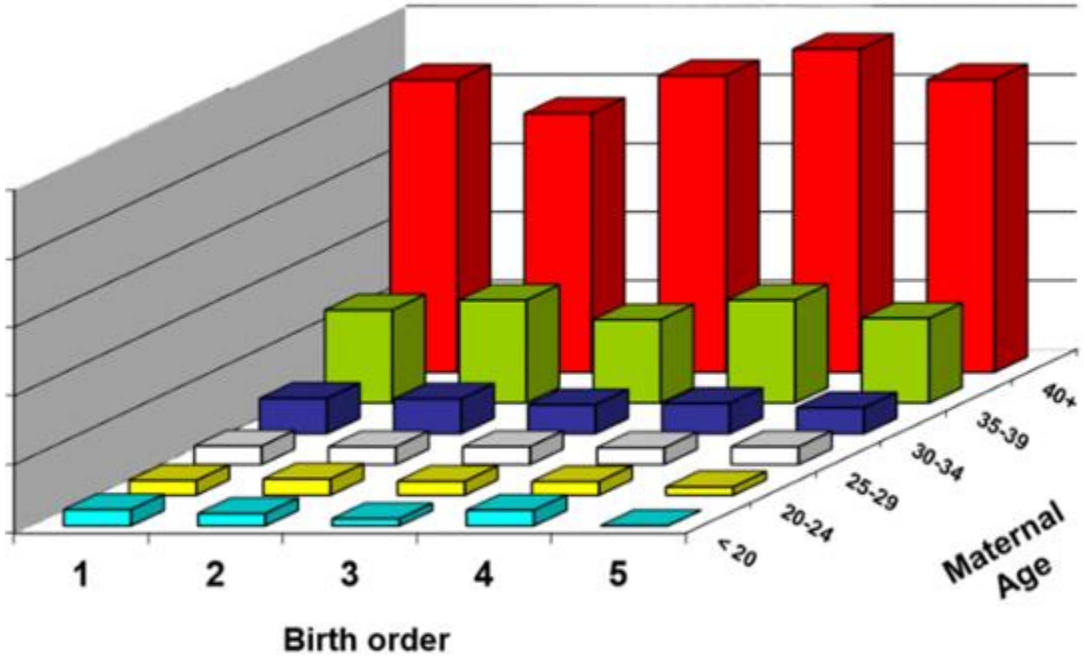
f_2: *marginal distribution* of Y

Marginalization: visualization

Given: joint distribution of (Birth order, Maternal Age) of babies:

- To get marginal probability of ‘Maternal Age’:
  - Stack up all bars of the same color

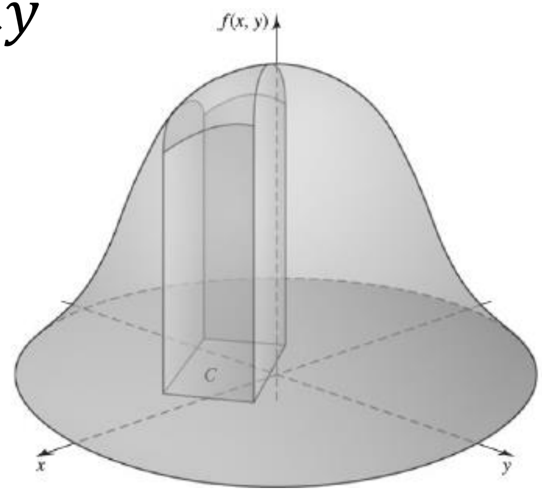
Joint PMF



Joint distribution of continuous RVs

- Any continuous random vector (X,Y) has a *joint probability density function* (PDF) $f(x,y)$, such that for all C,

$$P\big((X,Y) \in C\big) = \iint_C f(x,y)\,dx\,dy$$



$f(x,y)$: represent a 2D surface

This expression (double integral)

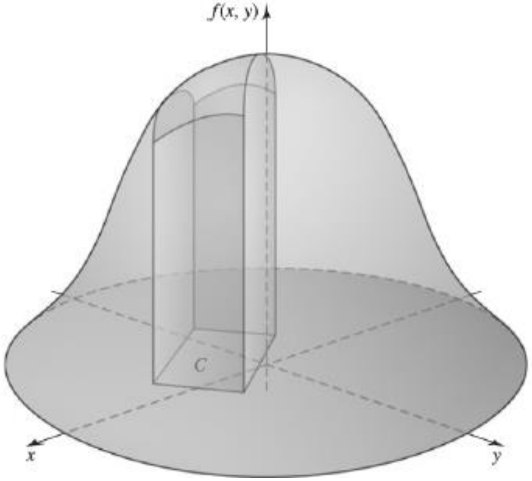
denotes the *volume* under the 2D surface with base C

Joint distribution of continuous RVs

Again:

- the ‘pile of sand’ analogy
- the histogram analogy

are useful to perceive $f(x, y)$

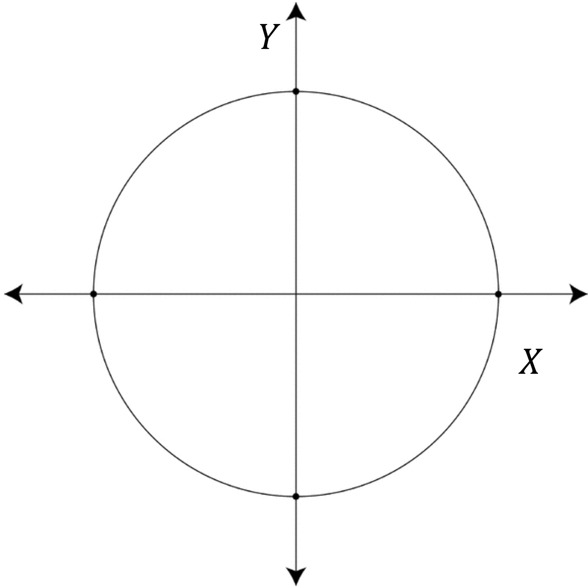


Properties:

- f is nonnegative
- $\iint_{R^2} f(x, y)\, dx\, dy = 1$ (R^2 = the whole x-y plane)
    - $P\big((X, Y) \in R^2\big) = 1$

Example: dartboard

- Dartboard with center (0,0) and radius 1; dart lands uniformly at random on the board



- What is the joint PDF of (X, Y)?

- Fact: the PDF is

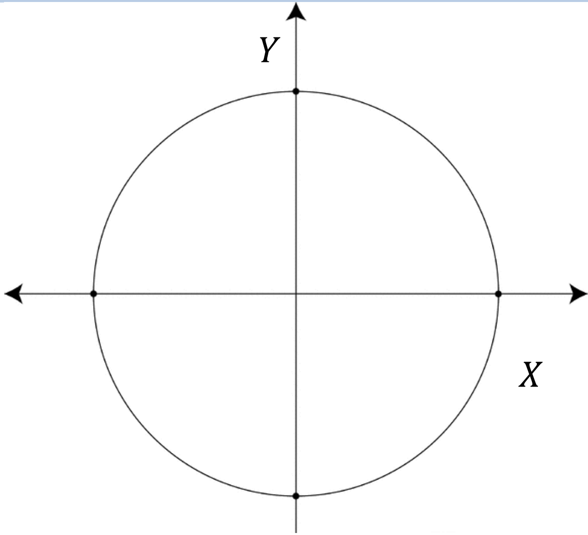
$$f(x, y) = \begin{cases} c, & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

- This is called "the Uniform distribution over the unit disk"

Example: dartboard

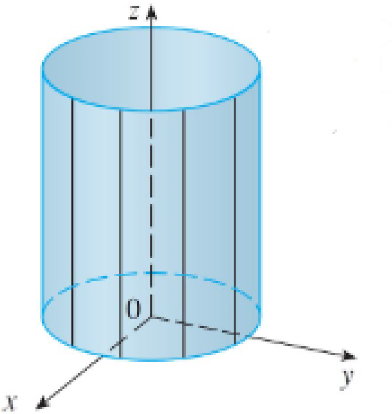
The PDF of X, Y is

$$f(x,y) = \begin{cases} c, x^2 + y^2 \leq 1 \\ 0, \text{otherwise} \end{cases}$$



Can we find c?

Observe: volume under $f(x,y)$ is πc (cylinder) which must also be 1



Therefore, $c = 1/\pi$

Marginal distribution of continuous RV

- Given joint distribution of continuous RV (X, Y), need distribution of one, say X.
- How to find X's PDF f_1?
  - Analogous to discrete case

Fact (marginalization) $f_1(x) = \int_R f(x, y)\, dy$

Replacing summation with integration in the continuous case ('marginalizing / integrating out variable Y')

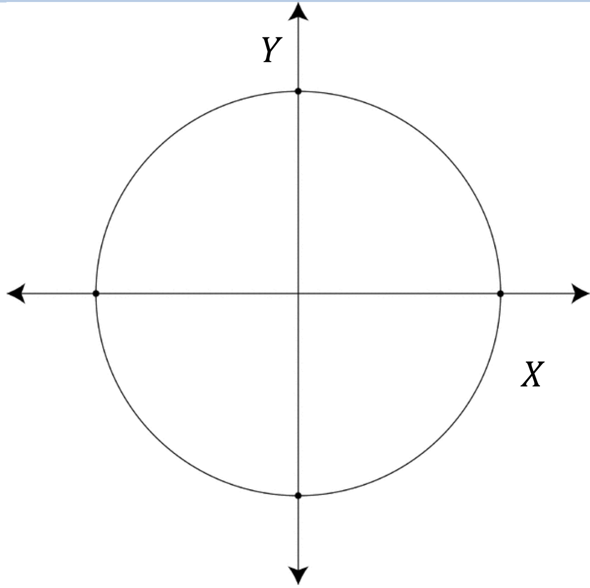
How about Y's PDF f_2?

- Marginalize out X

Example: dartboard

The PDF of X, Y is

$$f(x,y) = \begin{cases} \frac{1}{\pi}, x^2 + y^2 \leq 1 \\ 0, \text{otherwise} \end{cases}$$

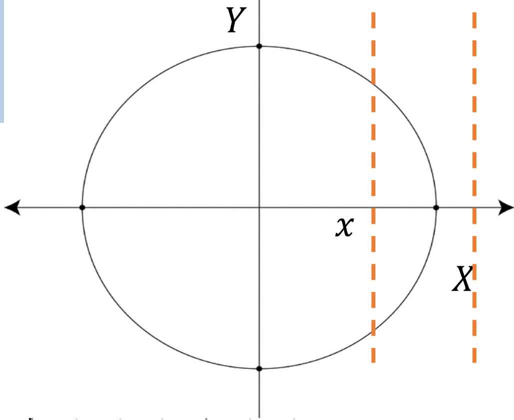


What is the marginal distribution over X?

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$$

How to find this integral?

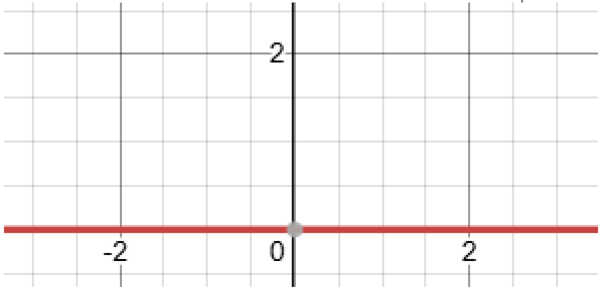
Example: dartboard



$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$$

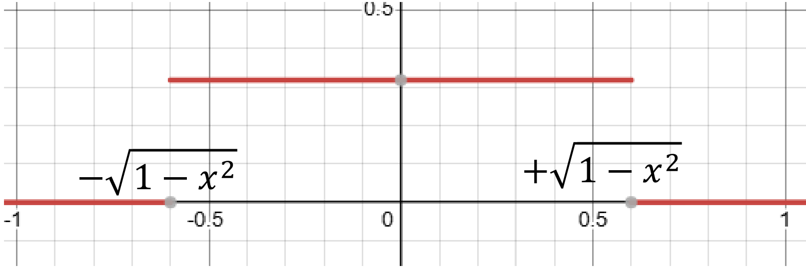
For x outside $[-1,1]$, $f(x,y) = 0$

=> $f_1(x) = 0$

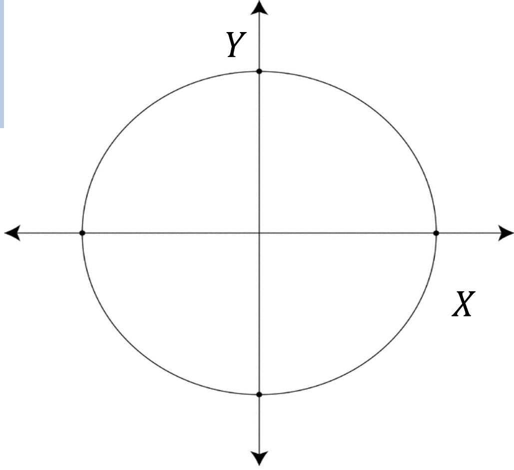


For a fixed $x \in [-1,1]$, $f(x,y)$ looks like

=> $f_1(x) = \frac{2}{\pi} \cdot \sqrt{1 - x^2}$



Example: dartboard



- In summary,

$$f(x) = \begin{cases} \frac{2}{\pi} \cdot \sqrt{1 - x^2}, x \in [-1,1] \\ 0, \qquad \text{otherwise} \end{cases}$$

X's distribution is NOT $\text{Uniform}([-1,1])$!

Actually makes sense: X closer to 1 is harder to be hit

Announcements

- Midterm graded (grade distribution on Piazza)
  - Let me know if you have feedback on curving scheme
- Quiz 5 graded
  - We will have quiz 6 this Wed (3/19)
- Participation: in-class question answering can now serve as equivalent of Piazza participation instance
  - I highly encourage you to get the 5pt bonus!

Announcements

- Lecture plan for the 2nd half of the class
  - Probability (2-3 lectures)
  - Data collection (1 lecture)
  - More on data visualization (1 lecture)
  - Basic statistics (4-5 lectures)
  - Basic data analysis: machine learning (6 lectures)
  - Final review (1 lecture)

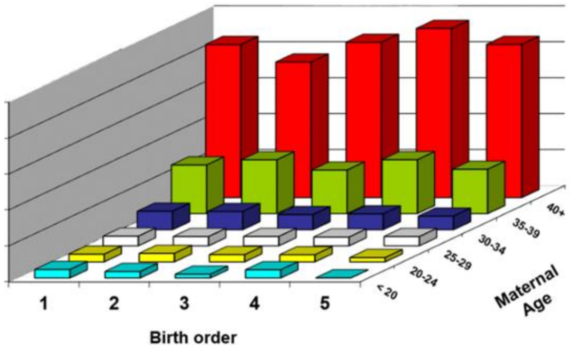
- HW5 may come a bit later (likely next week)
- I will provide project instructions soon (likely today)

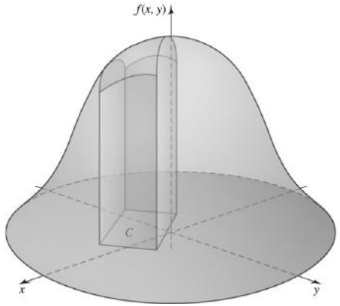
Recap: multivariate RVs

- A pair of RVs: (X, Y)

- If X and Y are both discrete, (X, Y)'s distribution can be characterized by their joint PMFs
  - what values could (X, Y) take
  - For each possible value (x, y), the probability of taking it

- If X and Y continuous, distribution characterized by their joint PDFs





Midterm Q6

Question 6 [17pts]: Suppose we have a discrete random variable X whose cumulative distribution function (CDF) is:

$$F(x) = \begin{cases} 0 & x < -1 \\ 0.2 & -1 \le x < 0 \\ 0.4 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

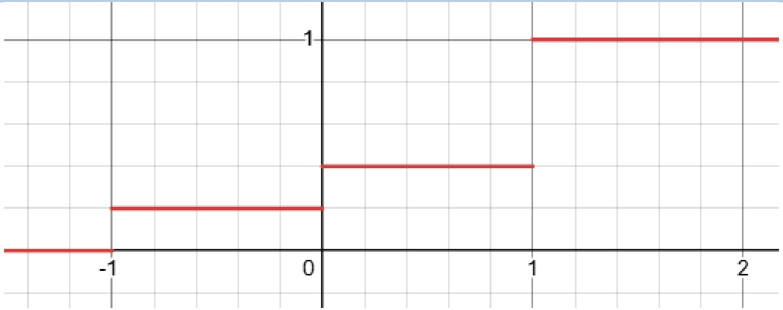
Solve the following problems and justify your answers:

(a) **[7pts]** Find the probability mass function (PMF) of X.

(b) **[6pts]** Let random variable $Y = X^2$. Find the joint PMF of (X, Y).

(c) **[4pts]** Find the marginal PMF of Y.

Midterm Q6



Answer:

1. Values x that X can take correspond to the "jumps" of the CDF, i.e, -1, 0, 1. The probability that X takes each of these values equals the amount of jump in each of the locations, and therefore, the PMF of X is as follows:

x	-1	0	1
$P(X = x)$	0.2	0.2	0.6

- Given this, how to find the joint PMF of (X, Y), for $Y = X^2$?
- What possible values can (X, Y) take?
  - (-1, 1)
  - (0, 0)
  - (1, 1)

Midterm Q6

- So, a way to write (X,Y)'s PMF is

(x,y)	(-1, 1)	(0, 0)	(1, 1)
P(X=x, Y=y)	0.2	0.2	0.6

- Written in two-way table, it is

	$x=-1$	$x=0$	$x=1$
$y=0$	0	0.2	0
$y=1$	0.2	0	0.6

- How to find the marginal of Y?
  - Take summation over each row

y	0	1
$P(Y=y)$	0.2	0.8

Joint distribution of more than 3 RVs

- Similarly, we can consider the joint distribution of more than 3 random variables,
  - E.g. (A,B,C), A = gender, B = class year, C = blood type
- Discrete RVs: can still define joint PMFs

a	b	c	$P(A=a, B=b, C=c)$
0	0	0	0.06
0	0	1	0.09
0	1	0	0.08
0	1	1	0.12
1	0	0	0.06
1	0	1	0.24
1	1	0	0.10
1	1	1	0.25

Marginalization

a	b	c	$P(A=a,B=b,C=c)$
0	0	0	0.06
0	0	1	0.09
0	1	0	0.08
0	1	1	0.12
1	0	0	0.06
1	0	1	0.24
1	1	0	0.10
1	1	1	0.25

Given the joint distribution of (A, B, C)

- What is the distribution of A?
  - Need to find $P(A = 0)$ and $P(A = 1)$

$$P(A = 0) = \sum_{b,c} P(A = 0, B = b, C = c)$$

Marginalization: summing over irrelevant variables

- What is the joint distribution of (A, B)?
  - Need to find $P(A = 0, B = 0), \ldots,\ P(A = 1, B = 1)$

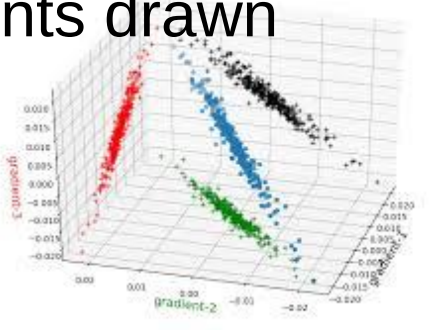
$$P(A = 0, B = 0) = \sum_{c} P(A = 0, B = 0, C = c)$$

Joint distribution of more than 3 RVs

- Continuous RVs: can still define joint PDFs,
- e.g. for (A, B, C), A = blood pressure, B = height, C = weight, they have a joint PDF of

$$f(a, b, c)$$

- Note: 3d PDF Is hard to visualize directly
- Useful to visualize f using a scatterplot of points drawn from joint distribution of (A, B, C)



Marginalization for continuous RVs

Suppose joint PDF of (A, B, C) is $f(a, b, c)$

- What is the PDF of A?

$$f_A(a) = \iint_{R^2} f(a, b, c)\ db\ dc$$

- What is the joint PDF of (A, B)?

$$f_{A,B}(a, b) = \int_R f(a, b, c) dc$$

Marginalization: summing over irrelevant variables

- These operations generalize to joint PDFs of more RVs..

Independence of RVs

Independence of RVs

- RVs X, Y are independent (denoted by $X \perp\!\!\!\perp Y$) if

$$f(x,y) = f_1(x) \cdot f_2(y), \quad \textit{for all } x, y$$

PMF or PDF Marginal of X Marginal of Y

- E.g. for discrete X, Y,

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

Therefore, $\{X = 3\}$ and $\{Y = 4\}$ are independent events

In class activity: checking independence of RVs

- Which of these PMFs correspond to independent $X \perp\!\!\!\perp Y$?

	$Y=0$	$Y=1$	
X=0	1/4	1/4	1/2
X=1	1/4	1/4	1/2
	1/2	1/2	1

X, Y independent

Need to check:
$f_1(0)f_2(0) = f(0,0)$,
..
(4 equalities)

	$Y=0$	$Y=1$	
X=0	1/2	0	1/2
X=1	0	1/2	1/2
	1/2	1/2	1

X, Y not independent

E.g. $f_1(0)f_2(1) = \frac{1}{4}$, whereas $f(0,1) = 0$

only one counterexample suffices to disprove independence!

Independence is invariant under transformations

Fact If X, Y are independent, then $f(X), g(Y)$ are also independent

E.g. X = tomorrow's temperature (in Celsius); Y = tomorrow's NVIDIA stock price (in $)

f(X) = tomorrow's temperature (in Fahrenheit); g(Y) = tomorrow's NVIDIA stock price (in cents)

Independence of more than two RVs

- RVs $X_1, \ldots, X_n$ are independent if their joint PMF or PDF satisfy

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n),$$

PMFs or PDFs　　Marginal for X_1　　Marginal for X_n

for all $x_1, \ldots, x_n$

- This captures many real-world applications:
  - Independent trials: each X_i is $\mathrm{Bernoulli}(p)$
  - Samples: each X_i is an independent sample from a population (foundations of *statistics*)
  - Manufacturing: X_i is the quality of part i

Independence of more than two RVs

Fact If $X_1, \dots, X_n$ are independent, then

- any subset $X_{i_1}, \dots, X_{i_p}$ are independent
  - E.g. X_1, X_3, X_7 are independent

- any disjoint subset $(X_{i_1}, \dots, X_{i_m})$, $(X_{j_1}, \dots, X_{j_l})$ are independent
  - E.g. (X_1, X_2) is independent of X_3
  - (X_1, X_3) is independent of (X_2, X_4)

True or False?

- If I flip 10 coins independently, it is more likely that I see
  HTTTHTHHTHT
  than
  HHHHHHHHHHH

- False

$$f(\text{HTTTHTHHTHT}) = f_1(H) \cdot \ldots \cdot f_{10}(T) = \frac{1}{2^{10}}$$

$$f(\text{HHHHHHHHHHH}) = f_1(H) \cdot \ldots f_{10}(H) = \frac{1}{2^{10}}$$

Conditional distributions of RVs

Conditional distributions (discrete)

- X, Y have joint PMF f. Y has marginal PMF f_2

- Conditional PMF of X given $Y = y$:

$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$

This is actually $\frac{P(X=x,Y=y)}{P(Y=y)} = P(X = x \mid Y = y)$

- Note: $g_1(x|y)$ is best viewed as a function of x; it reads "the conditional distribution of X given $Y = y$"

Conditional distributions (discrete)

Example X=0: car not stolen, X=1: car stolen

Joint PMF of X, Y:

Stolen X	Brand Y 1	2	3	4	5	Total
0	0.129	0.298	0.161	0.280	0.108	0.976
1	0.010	0.010	0.001	0.002	0.001	0.024
Total	0.139	0.308	0.162	0.282	0.109	1.000

Find the table of the conditional PMF of X given Y

Solution

Stolen X	Brand Y 1	2	3	4	5
0	0.928	0.968	0.994	0.993	0.991
1	0.072	0.032	0.006	0.007	0.009

P(X=0 | Y=1)
= 0.129 / 0.139

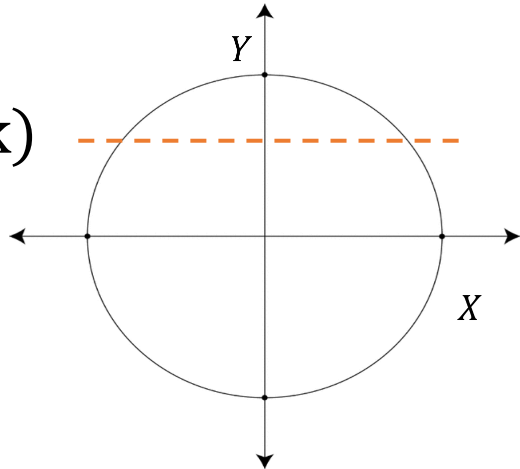
Conditional distributions (continuous)

- X, Y have joint PDF f. Y has marginal PDF f_2

- Conditional PDF of X given Y:

$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$

Example dartboard. $(X, Y) \sim \text{Uniform(unit disk)}$



Conditional distribution of X given $Y = 0.6$:

Uniform($[-0.8, +0.8]$)

Conditional distributions & independence

Fact X, Y are independent
$\Leftrightarrow$ for all y, $g_1(x|y)$ are all equal to $f_1(x)$

Here, g_1, f_1 are PMF or PDF depending on the types of X, Y

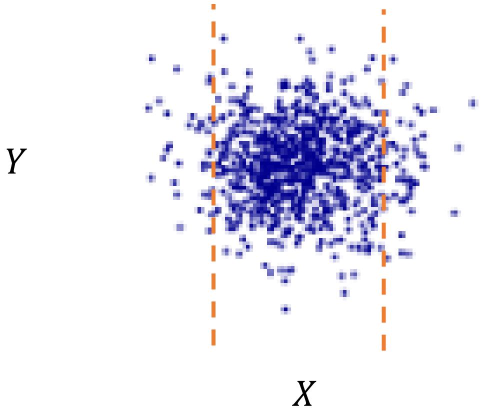
- In other words, knowing Y does not change our belief on X

	Brand Y					
Stolen X	1	2	3	4	5	Total
0	0.928	0.968	0.994	0.993	0.991	0.976
1	0.072	0.032	0.006	0.007	0.009	0.024
	$g_1(x|1)$	$g_1(x|2)$				$f_1(x)$

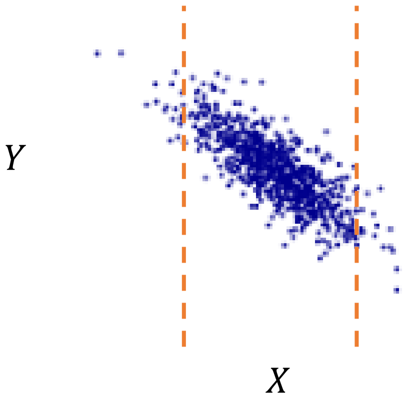
In the car example, X, Y are not independent!

Independence: visualization

- Left: X, Y independent; Right: X, Y not independent



$g_1(y|x=-1)$ $g_1(y|x=+1)$



$g_1(y|x=-1)$ $g_1(y|x=+1)$

True or False?

- If I flip a fair coin repeatedly, and my first 2 trials are both tails. Then my third throw will have a higher chance of showing head.

- This is asking $g_3(\mathrm{H} \mid \mathrm{TT}) = P(X_3 = H | X_1 = T, X_2 = T)$

  Since X_3 is independent of X_1, X_2 $\quad = P(X_3 = H) = 1/2$

  so the claim is false

- This is known as the *gambler's fallacy*
  - Prior losses do not increase the chance of future win

Conditional expectation

Definition The mean of the conditional distribution of X given $Y = y$, is called the *conditional expectation* of X given $Y = y$, denoted as $\mathrm{E}[\,X \mid Y = y\,]$.

$\mathrm{E}[\,X \mid Y = y\,]$ can be found by:

- $\sum_x x\ g_1(x|y)$, if X is discrete

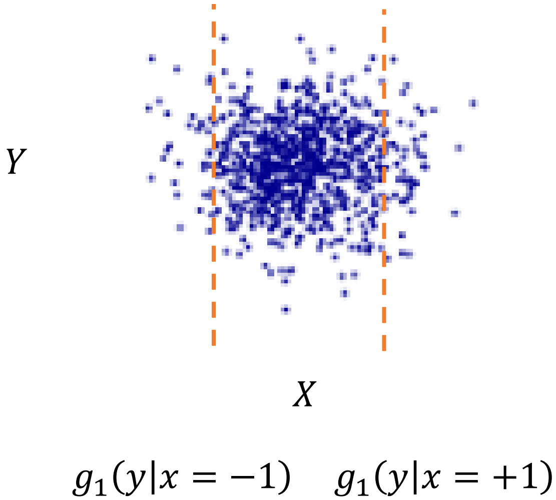
  Conditional PMF

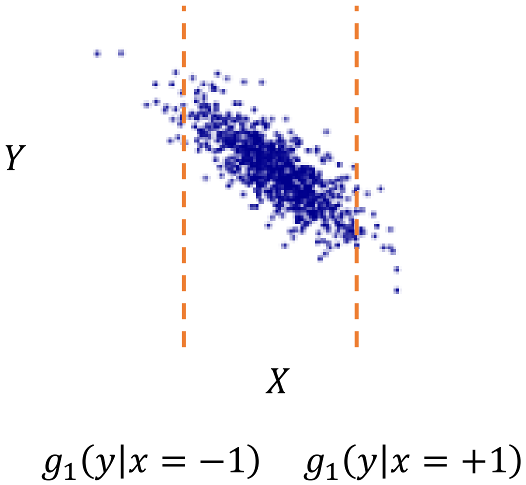
- $\int_{-\infty}^{+\infty} x\ g_1(x|y)\,dx$, if X is continuous

  Conditional PDF

Independence: visualization

- Left: X, Y independent; Right: X, Y not independent





Which one is larger, $\mathrm{E}[Y|X=-1]$ or $\mathrm{E}[Y|X=+1]$?
The former

Conditional expectation

Example Roll 2 fair dice. Expected value of die 1 given that their sum is 5?

Solution X: outcome of die 1; Y: sum of 2 dice, $E[X \mid Y = 5]$

Let's find the conditional distribution of X given $Y = 5$ first..

$$g_1(\, x \mid 5 \,) = P(\, X = x \mid Y = 5 \,)$$
$$= \frac{P(X=x, Y=5)}{P(Y=5)}$$

When is this nonzero?

Conditional expectation

$$g_1(x \mid 5) = P(X = x \mid Y = 5)$$
$$= \frac{P(X=x, Y=5)}{P(Y=5)}$$

When is this nonzero?

$x = 1,2,3,4$

$$\frac{4}{36} = \frac{1}{9}$$

Thus, the conditional distribution of X given $Y = 5$ is

x	1	2	3	4
P(X=x|Y=5) = $g_1(x \mid 5)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Therefore, $E[X \mid Y = 5]$ is

$$\frac{1}{4}(1 + 2 + 3 + 4) = 2.5$$

Quiz 6

Roll 2 fair dice independently. What is the expected value of die 1 given that their product is 4?

Quiz 6

Example Roll 2 fair dice. Expected value of die 1 given that their product is 4?

Solution X: outcome of die 1; Y: product of 2 dice, $E[X \mid Y = 4]$

Let's find the conditional distribution of X given $Y = 4$ first..

$$g_1(\, x \mid 5 \,) = P(\, X = x \mid Y = 4 \,)$$
$$= \frac{P(X=x, Y=4)}{P(Y=4)}$$

When is this nonzero?

Conditional expectation

$$g_1(\, x \mid 4\,) = P(\, X = x \mid Y = 4\,)$$
$$= \frac{P(X=x, Y=5)}{P(Y=4)}$$

$$\frac{3}{36} = \frac{1}{12}$$

When is this nonzero?

$x = 1,2,4$

Thus, the conditional distribution of X given $Y = 4$ is

x	1	2	4
P(X=x|Y=4) = $g_1(\, x \mid 4\,)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Therefore, $E[X \mid Y = 4]$ is

$$\frac{1}{3}(1 + 2 + 4) = \frac{7}{3}$$

Announcements

- Project information is out (on Piazza)
  - Let me know if you need help finding teammates
- We will release HW5 next week
  - My goal: teach basic machine learning next week
- Midterm question review..

Midterm Q7

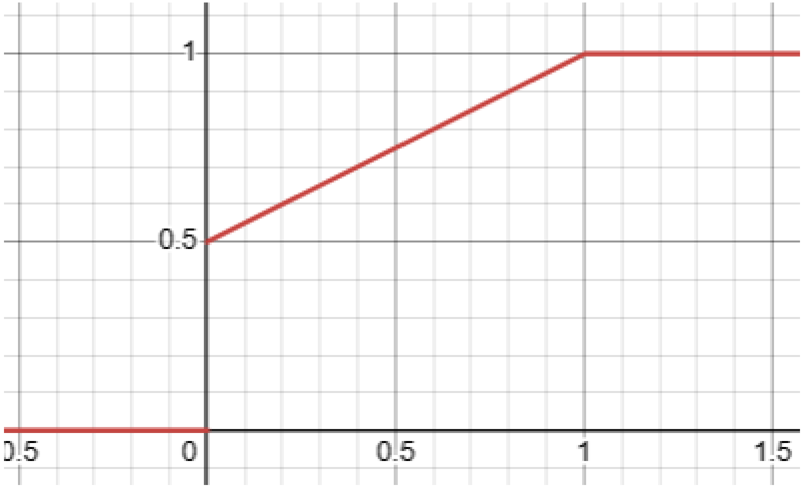
- Tools to characterize RVs

	Discrete RVs	Continuous RVs
CDF	√ (staircase)	√ (continuous)
PMF	√	×
PDF	×	√

- There are RVs that have neither PMF or PDFs
  - So they are neither discrete nor continuous
  - Example: mixture of discrete and continuous distributions (next slide)

Midterm Q7

- Consider the following example:
  - Flip a fair coin
  - If head, return X = 0
  - If tail, return X $\sim \text{Uniform}([0,1])$



- Then X's CDF is:
  - Neither a staircase
  - Nor continuous

Finding distributions of RVs

Finding distributions of random variables

- Oftentimes we are tasked with finding distributions of some complex random variable, e.g.
  - Total cost Z = X + Y, where X = food expenses, Y = transportation cost
  - Energy bill Z = X Y, where X = #hours at home, Y = power of all electrical devices

- How to find distributions of such Z?
  - We will learn how to do this when Z is discrete

Finding distributions of random variables

- How to find the distribution of a discrete RV Z:
  - Step 1: find what values Z can take
  - Step 2: find the probability that Z takes each possible value
- For continuous Z:
  - We can simulate drawing samples from Z and draw histogram!
  - For exact calculation, we will only state important facts

Finding distributions of random variables

Example Suppose $X \sim \text{Uniform}(\{1,2\}), Y \sim \text{Uniform}(\{1,2,3\})$, and $X \perp\!\!\!\perp Y$. Find the distribution of $Z = X + Y$.

Solution

Step 1: what values can Z take?

2, 3, 4, 5

Step 2: for each possible value, what is the probability that Z takes it?

Finding distributions of random variables

Example Suppose $X \sim \mathrm{Uniform}(\{1,2\}), Y \sim \mathrm{Uniform}(\{1,2,3\})$, and $X \perp\!\!\!\perp Y$. Find the distribution of $Z = X + Y$.

Solution

Step 2: what is the probability that Z takes 2? 3? 4? 5?

$$P(Z = 2) = P(X = 1, Y = 1) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$P(Z = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = \frac{1}{3}$$

…

z	2	3	4	5
P(Z=z)	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

Rule of Lazy Statistician

- If we are only interested in finding $\mathrm{E}[r(X,Y)]$, we can bypass finding $r(X,Y)$'s distribution using *the rule of lazy statistician*

- E.g. when X, Y are discrete:

$$\mathrm{E}[r(X,Y)] = \sum_{x,y} r(x,y) \cdot P(X = x, Y = y)$$

- Similar formulae hold for more than 3 RVs / continuous RVs
- We will see examples soon

Expectation and Variance revisited

Linearity of expectation

Fact Expectation of sum is sum of expectations

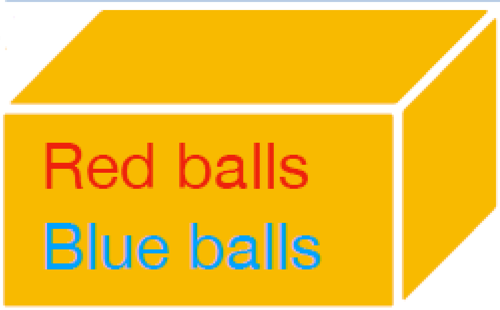
$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

Example: betting on two games

Note: generalizes to n variables

This property, together with the previously known $E[aX + b] = aE[X] + b$, are called the *linearity of expectation*

Linearity of expectation



Example Proportion of R balls is p

- Randomly sample n balls with replacement
- X: # R balls in the sample. $\mathrm{E}[X] =?$
- (We already knew the answer from binomial distribution..)

Solution Let $X_i = 1$ if i-th ball is R, and 0 otherwise

$\Rightarrow X = X_1 + \cdots + X_n$

Each X_i has expectation p

$\Rightarrow \mathrm{E}[X] = E[X_1] + \cdots + E[X_n] = np$

Linearity of Variance?

Is $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$?

- It depends..
  - Case 1: when $Y = -X$,

$$\mathrm{Var}[X+Y] = 0$$

$$\mathrm{Var}[Y] = \mathrm{Var}[X]$$

  => LHS < RHS

  - Case 2: when $Y = X$,

$$\mathrm{Var}[X+Y] = \mathrm{Var}[2X] = 4\,\mathrm{Var}[X]$$

$$\mathrm{Var}[Y] = \mathrm{Var}[X]$$

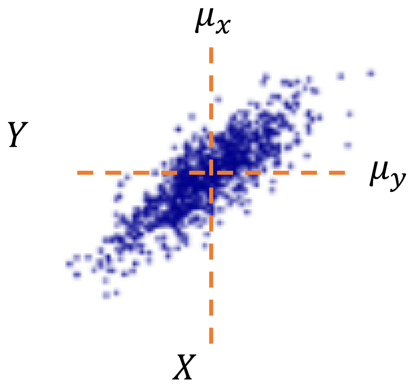
  => LHS > RHS

- Observation: extra correction is needed to balance the equation

Covariance

- Covariance of X, Y: numerical measure of the degree to which X, Y vary together. Let $\mathrm{E}[X] = \mu_x, \mathrm{E}[Y] = \mu_y$,

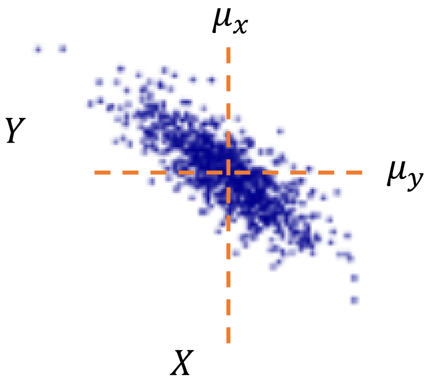
$$\mathrm{Cov}(X, Y) = \mathrm{E}\big[(X - \mu_x)(Y - \mu_y)\big]$$



$\mathrm{Cov}(X, Y) > 0$

Positive correlation:

X, Y simultaneously large or small

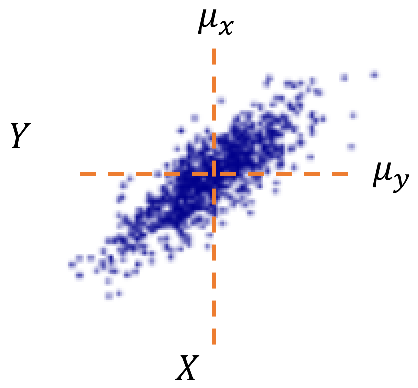


$\mathrm{Cov}(X, Y) < 0$

Covariance

Let $\mathrm{E}[X] = \mu_x, \mathrm{E}[Y] = \mu_y$,

$$\mathrm{Cov}(X, Y) = \mathrm{E}\big[(X - \mu_x)(Y - \mu_y)\big]$$



Properties

- $\mathrm{Cov}(X, X) = \mathrm{Var}[X]$
- $\mathrm{Cov}(X + a, Y + b) = \mathrm{Cov}(X, Y)$ — Covariance is invariant to shifting
- $\mathrm{Cov}(cX, dY) = cd\,\mathrm{Cov}(X, Y)$ — Covariance is sensitive to scaling

Covariance

Fact (alternative formula) $\text{Cov}(X, Y) = E[XY] - \mu_x\mu_y$

Example Find $\text{Cov}(X, Y)$ given PMF

	$Y=0$	$Y=1$	
X=0	1/2	0	1/2
X=1	0	1/2	1/2
	1/2	1/2	1

$E[XY] = \sum_{x,y} xy \; P(X = x, Y = y) = 0 \cdot 0 \cdot \frac{1}{2} + 1 \cdot 1 \cdot \frac{1}{2} = \frac{1}{2}$

$\mu_x = \frac{1}{2}, \mu_y = \frac{1}{2}$

$\text{Cov}(X, Y) = \frac{1}{2} - \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Correlation coefficient

- Covariance is sensitive to scaling, e.g.

$$\mathrm{Cov}(100X, Y) = 100\,\mathrm{Cov}(X, Y)$$

- Better measure, independent of changes in scales

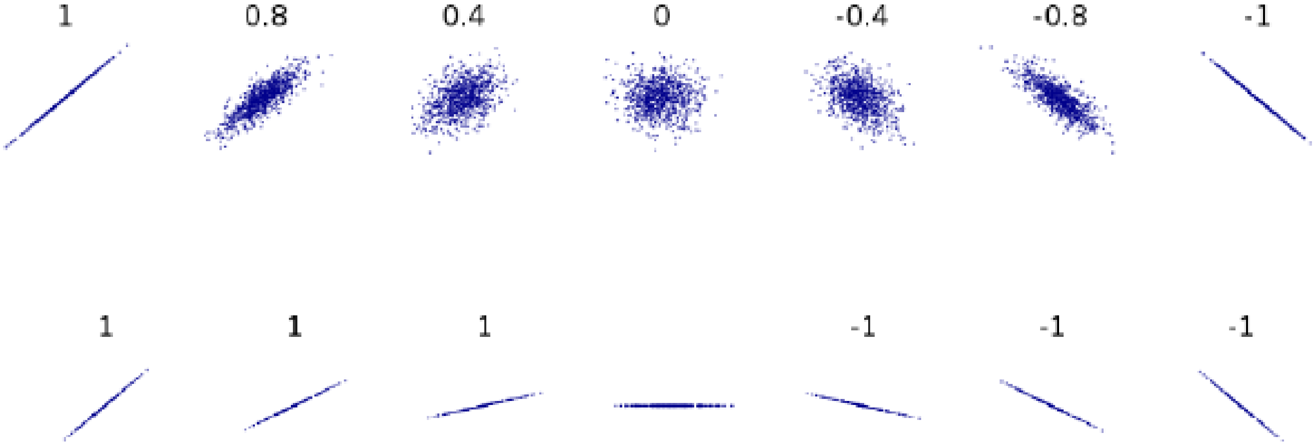
$$\text{Correlation of } X, Y = \rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Standard deviation (i.e. square root variance) of X and Y

- Measures linear association of X, Y. Always in [-1,1].

Correlation coefficient

- Example instances of $\rho(X, Y)$:



What happens to this distribution?
$\sigma_Y = 0$, making $\rho(X, Y)$ undefined

Property of Variance – Corrected formula

Fact

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y)$$

Sanity check:

- When $Y = -X$: $2\mathrm{Cov}(X,Y) = -2\mathrm{Cov}(X,X) = -2\,\mathrm{Var}[X]$
  - LHS = RHS = 0
- When $Y = X$: $2\mathrm{Cov}(X,Y) = 2\mathrm{Var}[X]$
  - LHS = RHS = $4\,\mathrm{Var}[X]$
- What happens when X, Y are independent?

Independent RVs: important properties

Fact When $X \perp\!\!\!\perp Y$, $E[XY] = E[X]E[Y]$. As a result,

$$\mathrm{Cov}(X, Y) = 0 \text{ and } \mathrm{Var}(X + Y) = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Justification

independence

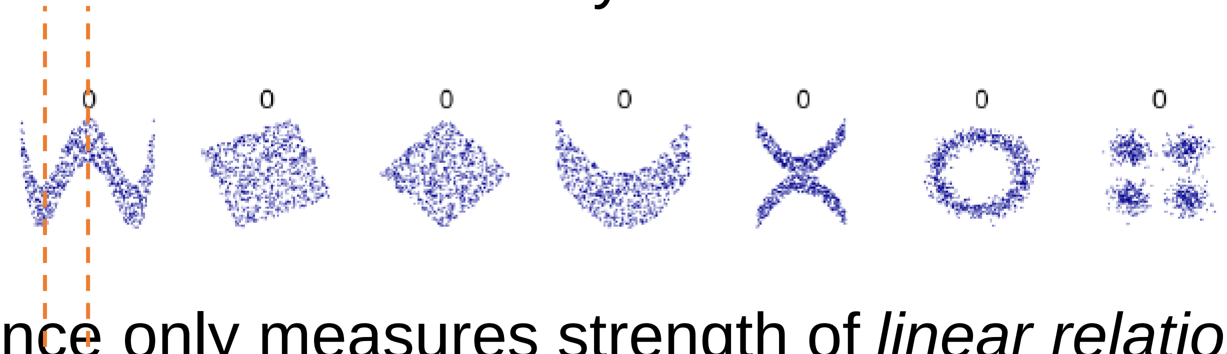
$$E[XY] = \sum_x \sum_y x\, y\, f(x, y) = \sum_x \sum_y x\, y\, f_1(x) f_2(y)$$

Rule of Lazy Statistician

$$= \sum_x x f_1(x)\, \mu_y = \mu_x \mu_y$$

Independence vs. Zero Covariance

- Independence implies zero covariance.
- Does zero covariance imply independence?
  - No!
- When $\text{Cov}(X, Y) = 0$, i.e., $\rho(X, Y) = 0$, X, Y can still be dependent on all kinds of ways:



- Covariance only measures strength of *linear relationship* between X, Y

In class exercise: a concrete counterexample

Example $X \sim \mathrm{Uniform}(\{-1,0,1\})$. $Y = X^2$.

Are X, Y independent?

Is $\mathrm{Cov}(X, Y) = 0$?

Announcements 3/24

- Quiz 6 graded – let us know if you have questions

- We will have quiz 7 this Wednesday

Recap 3/24

- Expectation of RVs' sum

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

holds *in general* – does not require independence!

- Variance of RVs' sum $\mathrm{Var}[X+Y]$

$$\mathrm{Var}[X+Y]=\mathrm{Var}[X]+\mathrm{Var}[Y] \quad +2\mathrm{Cov}[X,Y]$$

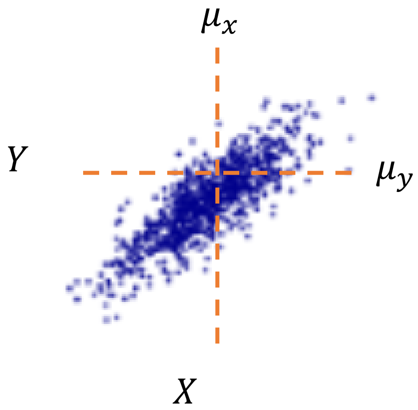
Covariance: measure the correlation of X, Y

Recap 3/24

$$\text{Cov}(X,Y) = E\left[(X-\mu_x)\left(Y-\mu_y\right)\right]$$

$\mu_x = E[X]$
$\mu_y = E[Y]$

$\text{Cov}(X,Y) > 0$ if more X, Y deviates from μ_x, μ_y in the same direction (simultaneously large or simultaneously small)



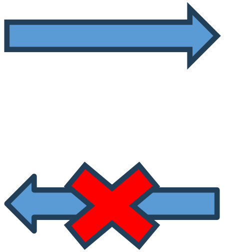
$\text{Cov}(X,Y) = \text{Cov}(Y,X)$

Fact (alternative formula) $\text{Cov}(X,Y) = E[XY] - \mu_x\mu_y$

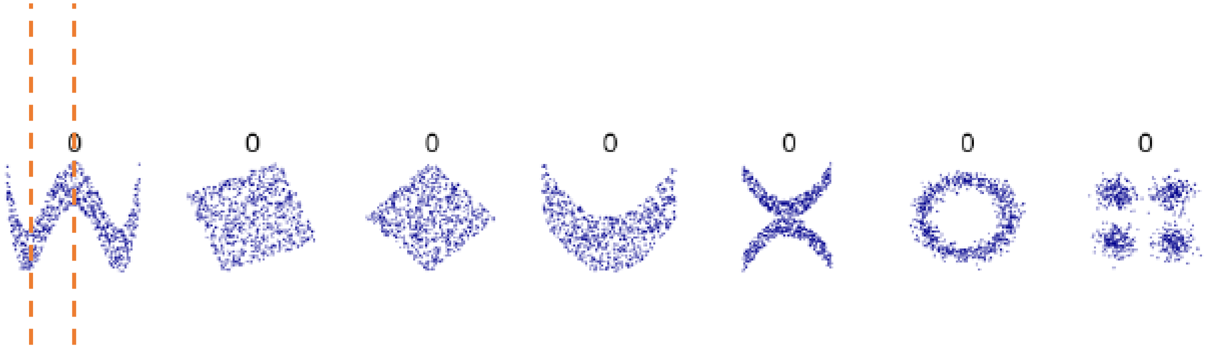
Recap 3/24

- $\mathrm{Cov}(X, Y) = E\left[(X - \mu_x)\left(Y - \mu_y\right)\right]$

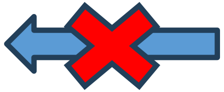
X, Y are independent



$\mathrm{Cov}(X, Y) = 0$



In class exercise: a concrete counterexample

X, Y are not independent  $\mathrm{Cov}(X, Y) = 0$

Example $X \sim \mathrm{Uniform}(\{-1,0,1\})$. $Y = X^2$.

Show: X, Y are not independent, but $\mathrm{Cov}(X, Y) = 0$

In class exercise: a concrete counterexample

Example $X \sim \text{Uniform}(\{-1,0,1\})$. $Y = X^2$.

Why are X, Y not independent?

- $Y \mid X = 0$ and $Y \mid X = 1$ have different distributions

	x=-1	x=0	x=1
y=0	0	1/3	0
y=1	1/3	0	1/3

Why is $\text{Cov}(X, Y) = 0$?

- $\mu_x = 0, \mu_y = \frac{2}{3}$
- $\text{E}[XY] = \text{E}[X^3] = 0$
- $\text{Cov}(X, Y) = \text{E}[XY] - \mu_x \mu_y = 0$

The covariance matrix

The *covariance matrix* of RVs A, B is a 2x2 array, with its entries being

$$\left[\begin{array}{c|c} \begin{matrix} Cov(A,A) \\ Var(A) \end{matrix} & Cov(A,B) \\ \hline Cov(B,A) & \begin{matrix} Cov(B,B) \\ Var(B) \end{matrix} \end{array}\right]$$

Matrix: 2d array of elements

The covariance matrix of RVs $(X_1, \ldots, X_n)$ is a nxn array, with its entries being

$$\begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \cdots & \mathrm{Cov}(X_n, X_n) \end{bmatrix}$$

(we will see examples soon..)

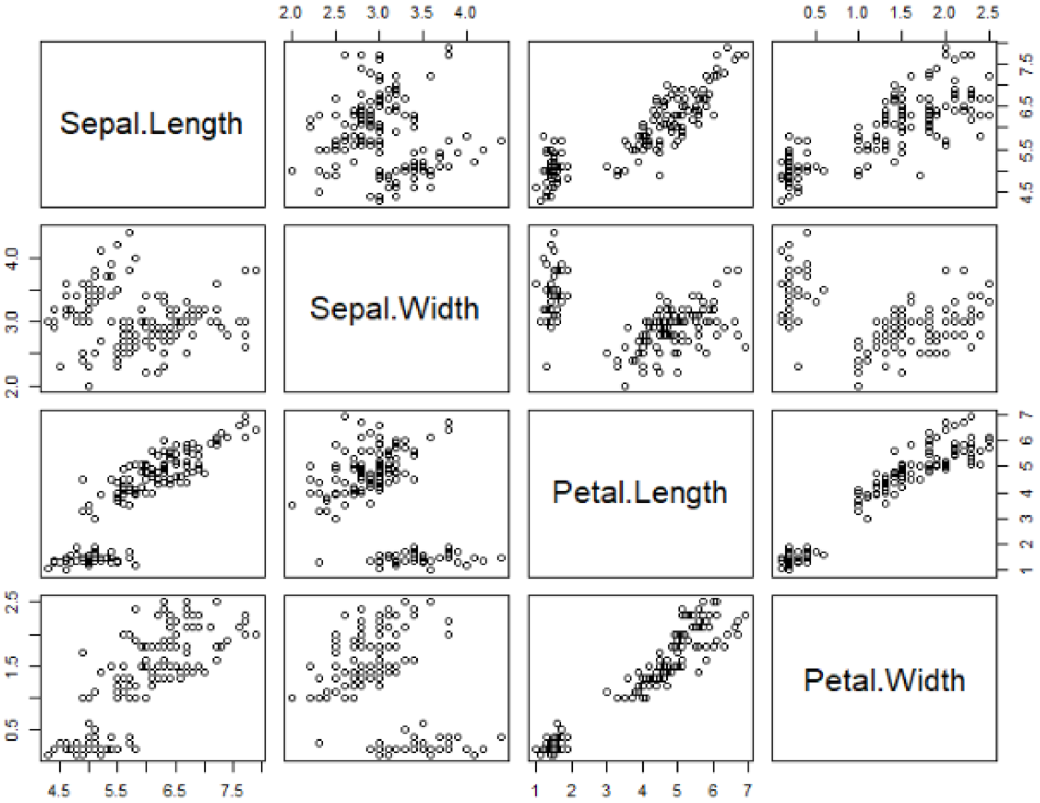
Aside: visualizing correlations between variables

Useful tool: Pair plot

Example iris data

each data point has 4 features

$$X_1, X_2, X_3, X_4$$



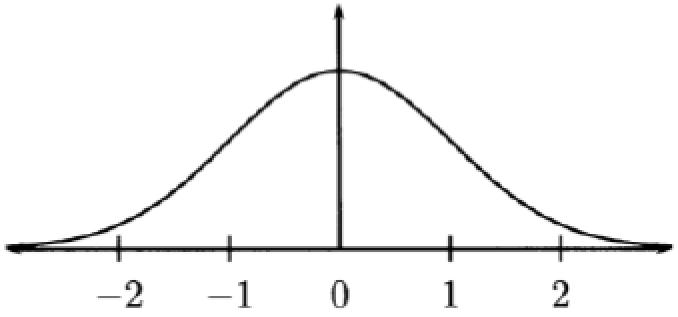
Terminology: moments

$E[X^k]$: k-th order *raw* moments

- Notable example: $k = 1$ => mean

$E[(X - \mu)^k]$: k-th order *central* moments

- Notable example: $k = 2$ => variance
- $k = 3$:
  - Skewness – degree of asymmetry
- $k = 4$:
  - Kurtosis – frequency of outliers



$E[(X - \mu_x)^m (Y - \mu_y)^n]$: cross moments

Moments are useful summaries of distributions of RVs

Example multivariate random variables

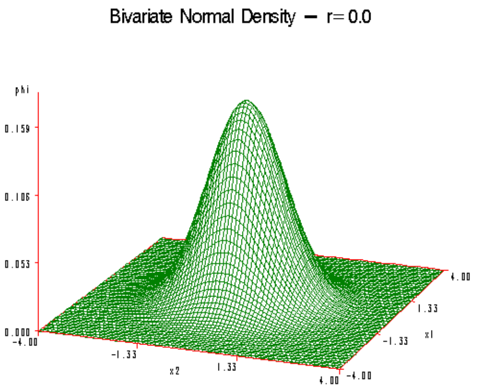
The 2d standard Gaussian distribution

Suppose $X \sim N(0,1)$, $Y \sim N(0,1)$, and $X \perp\!\!\!\perp Y$, (X,Y) is said to be drawn from the *two-dimensional standard Gaussian distribution*

What is its PDF?

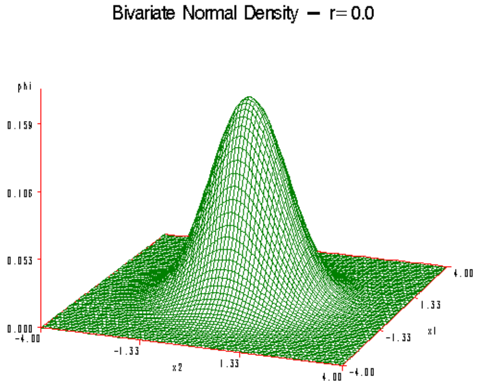
$$f(x,y) = f_1(x)\, f_2(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$$

$$= \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)$$



It is a *bell-shaped surface*

The 2d standard Gaussian distribution



Standard Gaussian PDF

$$f(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)$$

What is its mean vector

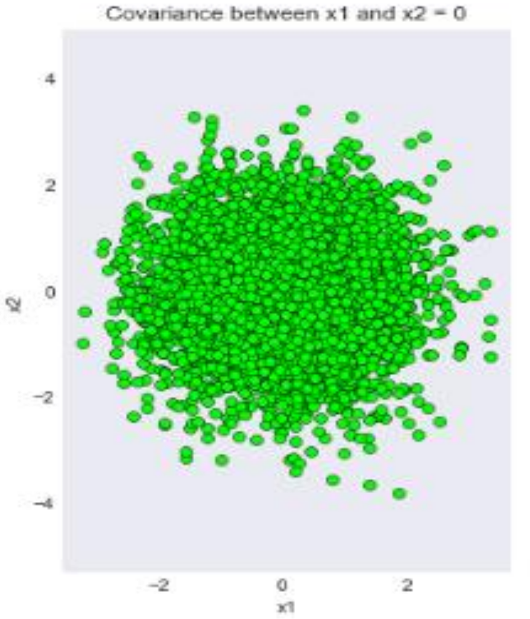
$$\begin{bmatrix} \mathrm{E}[X] \\ \mathrm{E}[Y] \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

What is its covariance matrix?

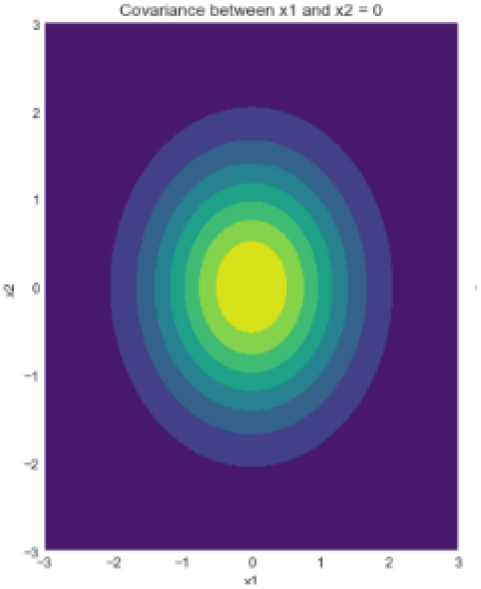
$$\begin{bmatrix} \mathrm{Cov}(X,X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Cov}(Y,Y) \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$: *identity matrix*

The 2d standard Gaussian distribution

Scatter plot of the samples



Contour of the PDF

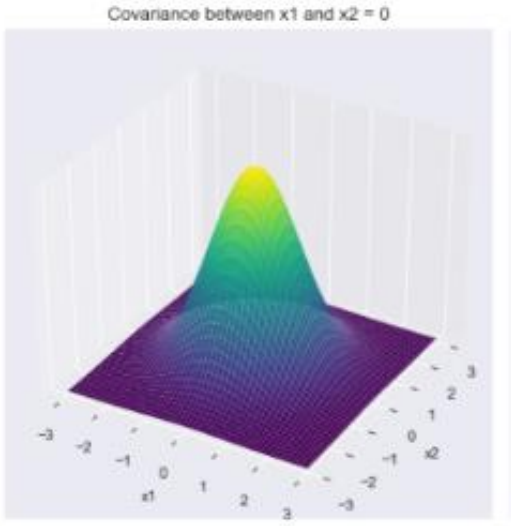


- isotropic: identical variations across different directions
- samples look like a “spherical” point cloud

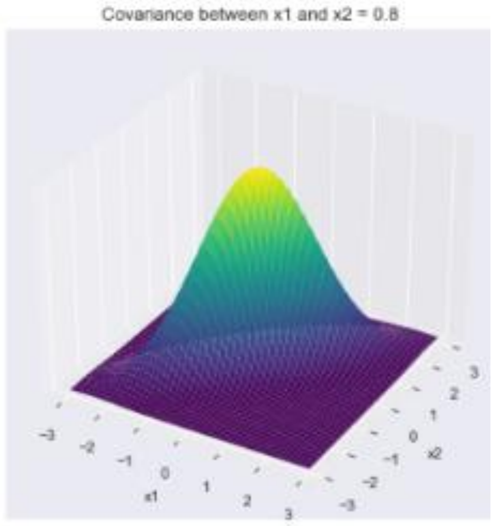
2d general Gaussian distributions

Fact For any 2x1 mean vector μ and a 2x2 covariance matrix Σ, there is a two-dimensional Gaussian distribution associated with it, denoted as $N(\mu, \Sigma)$.

Standard Gaussian PDF

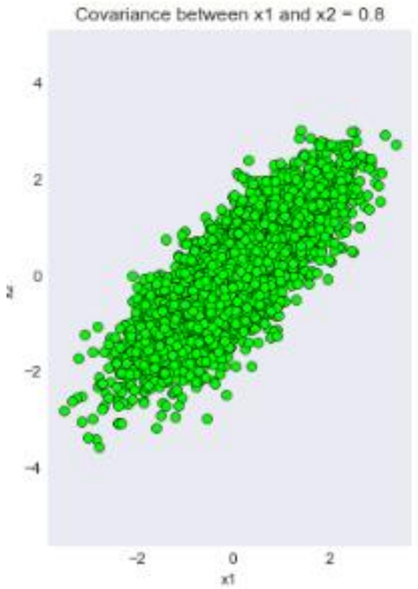


general Gaussian PDF

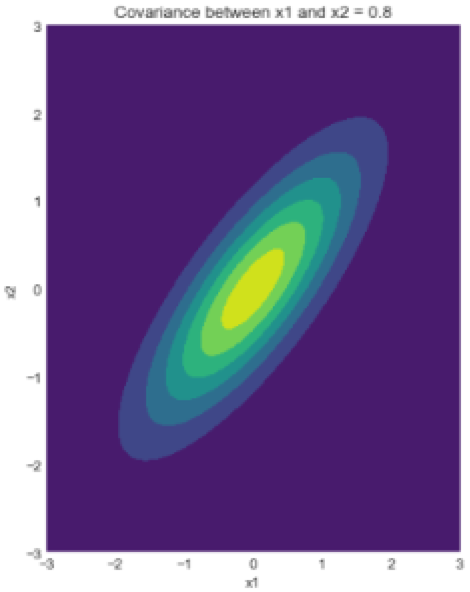


2d general Gaussian distributions

Scatter plot of the samples



Contour of the PDF

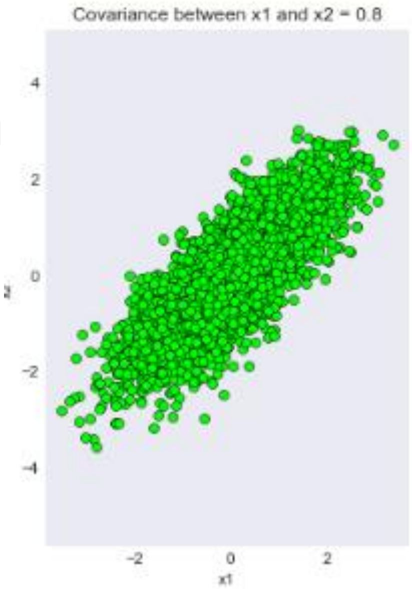


Elevation contours of general Gaussian PDFs are *ellipses*

2d Gaussian distribution

Real-world examples:

- Temperature and Pressure at random location
- Height and Weight of Individuals
- Stock Market Returns of Two Companies

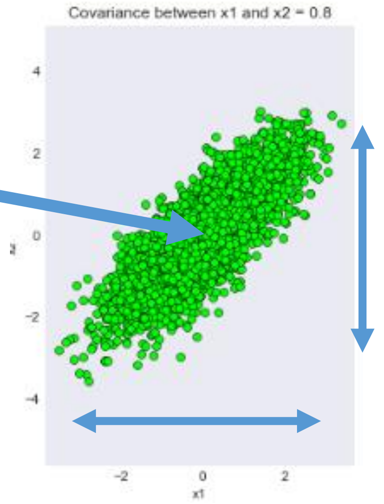


2d general Gaussian distributions

2d Gaussian distribution $N(\mu, \Sigma)$, How many parameters?

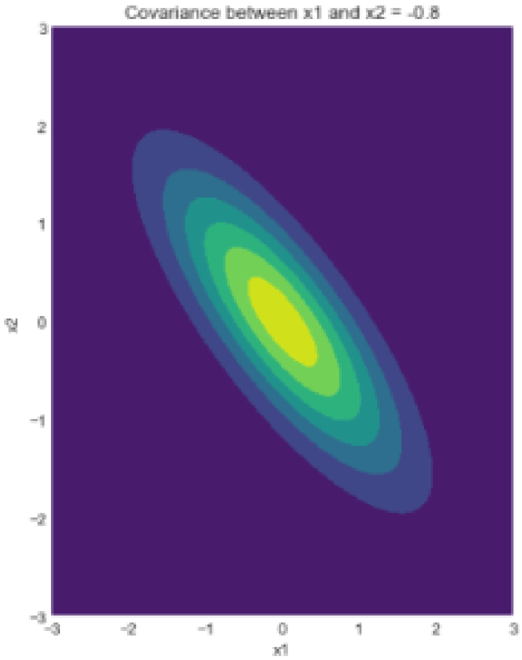
- $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, 6 parameters
- What are the meanings of the parameters?

- (μ_1, μ_2): center of the distribution
- Σ_{11}: variance of X_1
- Σ_{22}: variance of X_2
- Σ_{12}: covariance of X_1, X_2



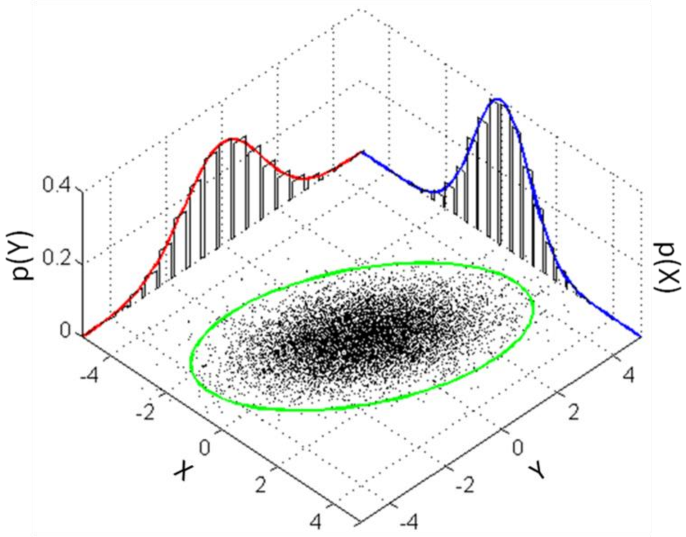
In-class activity: 2d general Gaussian distributions

- Can you draw the contour of some 2d Gaussian distribution with $\mathrm{cov}(X_1, X_2) < 0$?



2d general Gaussian distributions

- **Fact** Suppose (X, Y) follows the 2d Gaussian distribution $N(\mu, \Sigma)$. Then both X and Y's marginal distributions are Gaussian.



- What are X's mean & variance?
  - $\mu_1, \ \Sigma_{11}$
  - X's marginal distribution is $N(\mu_1, \Sigma_{11})$

- What about Y?
  - Y's marginal distribution is $N(\mu_2, \Sigma_{22})$

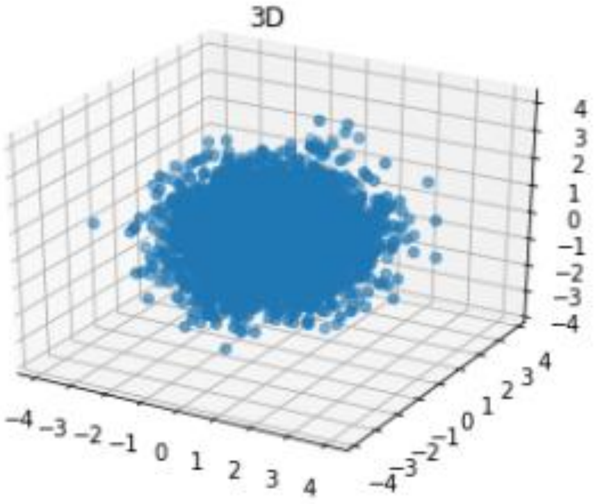
The general n-dimensional Gaussian distribution

Fact For any nx1 mean vector μ and a nxn covariance matrix Σ, there is a n-dimensional Gaussian distribution associated with it, denoted as $N(\mu, \Sigma)$**.**

Meaning of parameters:

$$\mu = \begin{bmatrix} \mathrm{E}[X_1] \\ \ldots \\ \mathrm{E}[X_n] \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \cdots & Cov(X_n, X_n) \end{bmatrix}$$



Gaussian is closed under addition

Fact If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, and $X \perp\!\!\!\perp Y$**,** then $Z = X + Y$ is also Gaussian.

Can you find the parameters of Z's Gaussian distribution?

$$\mathrm{E}[Z] = \mathrm{E}[X] + \mathrm{E}[Y] = \mu_X + \mu_Y$$

$$\mathrm{Var}[Z] = \mathrm{Var}[X] + \mathrm{Var}[Y] = \sigma_X^2 + \sigma_Y^2$$

Thus, $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Gaussian is closed under addition

Example Suppose X_1, X_2, X_3 are 3 independent measurements of the length of a table (in cm), which follow distribution $N(40, 0.1^2)$. Find the distribution of sample mean

true length of table

$$\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$$

Solution

$$X_1 + X_2 \sim N(80, 2 \times 0.1^2)$$

Since $X_2 \perp\!\!\!\perp X_1$

$$X_1 + X_2 + X_3 \sim N(120, 3 \times 0.1^2)$$

Since $X_3 \perp\!\!\!\perp (X_1, X_2)$ (and thus $X_3 \perp\!\!\!\perp X_1 + X_2$)

Gaussian is closed under addition

Example Suppose X_1, X_2, X_3 are 3 independent measurements of the length of a table (in cm), which follow distribution $N(40, 0.1^2)$. Find the distribution of sample mean

$$\bar{X} = \frac{1}{3}\,(X_1 + X_2 + X_3)$$

Solution

$$X_1 + X_2 + X_3 \sim N(120, 3 \times 0.1^2)$$

$$\bar{X} \sim N\left(\frac{120}{3}, 3 \times \frac{0.1^2}{3^2}\right) = N\left(40, \frac{0.1^2}{3}\right)$$

Conclusion: averaging over multiple measurements reduces measurement error

Law of Large Numbers

Motivation: measurement

- Suppose we use a ruler to measure the width of a tumor and collect readings such as (in cm):

1.132, 1.136, 1.127, 1.119

X_1 X_2 X_3 X_4

- These readings can be viewed as the random draws of RV X with mean μ (μ is the true width of the tumor)

- The sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ should approach μ?

Motivation: insurance payments

- Suppose we are a health insurance company that serves 1 million policyholders
- Each holder will file insurance claims in year 2025
- We'd like to estimate the total payments we make this year

- Luckily, we know that all holders' claim amount X are independent, and follow the $\mathrm{Uniform}([0, 1000])$ distribution

Terminology: IID sample

Definition $X_1, \dots, X_n$ is an *independent & identically distributed (IID, iid)* sample of X if:

- each X_i has the same distribution as X
- $X_1, \dots, X_n$ are independent

Note: a sample is a collection of many data points!

More examples:

- Randomly draw 10 students from UA student database *with replacement*
- Make 3 independent measurements

Law of Large Numbers (LLN)

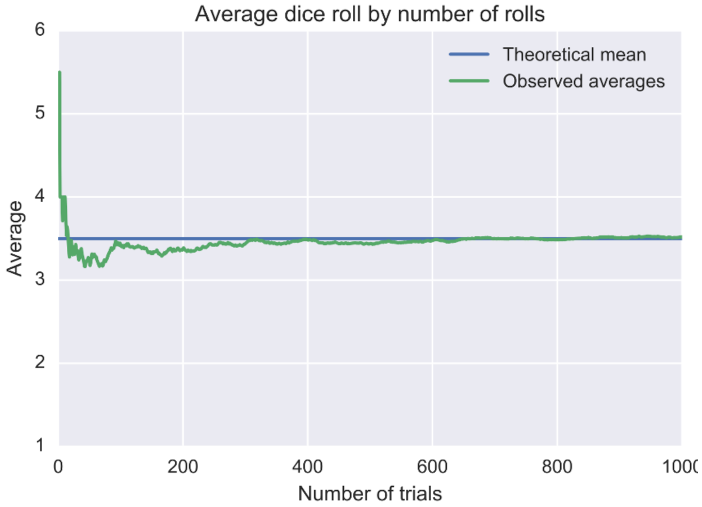
Law of Large Numbers Let $X_1, \ldots, X_n$ be an iid sample of random variable X. Let $\bar{X}_n$ be sample mean, and $\mu = \mathrm{E}[X]$. Then

$$\bar{X}_n \to \mu \text{ as } n \to \infty$$

Example: dice roll

$$X \sim \text{Uniform}(\{1, .., 6\})$$

$$\mu = \mathrm{E}[X] = 3.5$$



Example: insurance payments

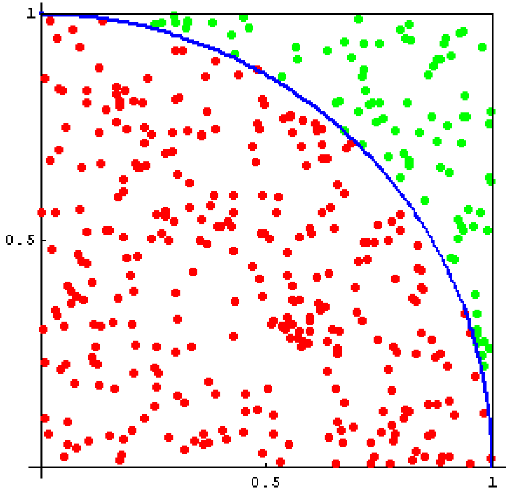
- Suppose we are a health insurance company that serves 1 million policyholders
- We know that all holders' claim amount X follows a distribution, $\mathrm{Uniform}([0,1000])$

- Suppose $X_1, \ldots, X_n$ are the payments we make to each holder, $n = 1M$.
- LLN => $\frac{1}{n}(X_1 + \cdots + X_n) \approx \mathrm{E}[X] = 500$
- We should prepare $X_1 + \cdots + X_n \approx 500M$ for payments

Application: Monte Carlo methods

- LLN has many other cool applications!
- Monte Carlo methods: use randomization to compute probabilities or expectations of interest

Example estimate π by sampling



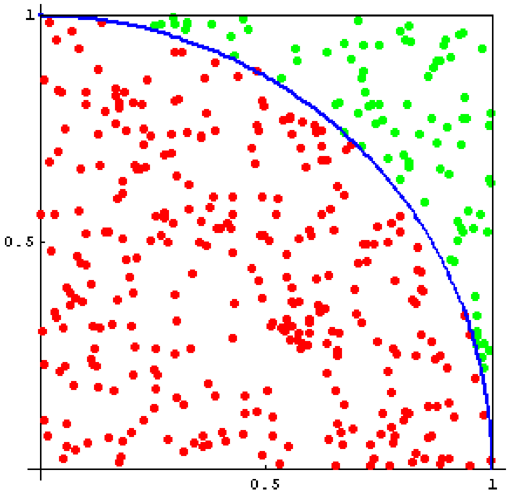
Application: Monte Carlo methods

```python
from random import random
from math import sqrt
# Number of random points:
N = 10000
# Counter of points inside:
I = 0
for i in range(N):
    # Generate random point
    # in the 1 x 1 square:
    x = random()
    y = random()
    # Is it inside the circle?
    r = sqrt(x**2 + y**2)
    if r < 1: I += 1
# Calculate Pi:
print(4 * I / N)
```

L_i

LLN => $\frac{1}{10000}(L_1 + \cdots + L_{10000}) \approx \mathrm{E}[L]$

$$\mathrm{E}[L] = \pi/4$$



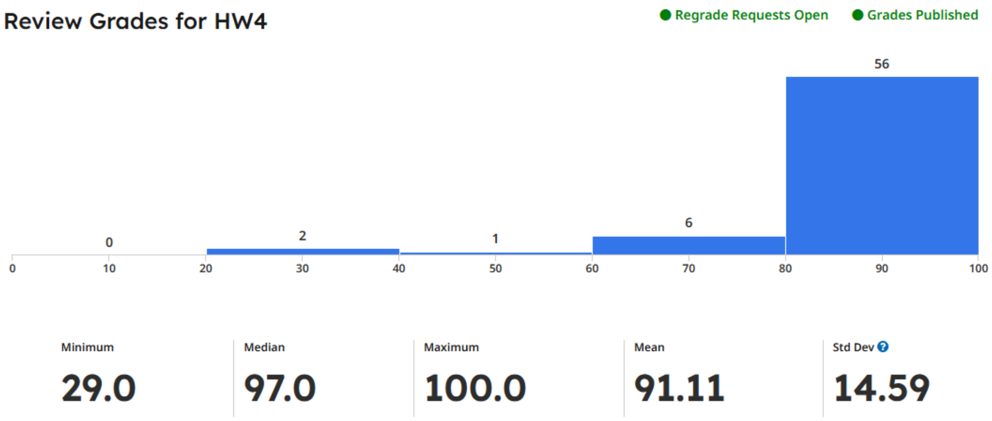
Results of 5 runs:

3.1768
3.1496
3.1644
3.1504
3.1384

Central Limit Theorem

Announcements 3/26

- HW4 graded



- We are working on uploading the (curved) midterm scores on D2L this week
- Participation award mechanism change (to fractional)
- HW5 will be up soon..

Quiz 7

Suppose we measure the ping time to a server 100 times under ideal network conditions. The results (in ms) are normally distributed with mean $\mu = 15$, standard deviation $\sigma = 5$.

```
(base) chichengz@DESKTOP-TDF1D2U:/mnt/c/Users/zcc13$ ping arizona.edu
PING arizona.edu (151.101.2.133) 56(84) bytes of data.
64 bytes from 151.101.2.133 (151.101.2.133): icmp_seq=1 ttl=51 time=14.3 ms
64 bytes from 151.101.2.133 (151.101.2.133): icmp_seq=2 ttl=51 time=16.9 ms
64 bytes from 151.101.2.133 (151.101.2.133): icmp_seq=3 ttl=51 time=17.1 ms
```

What is the distribution of the average ping time?

Quiz 7

What is the distribution of the average ping time?

We have a sample $X_1, \dots, X_{100}$, each is $N(15, 5^2)$, problem asking about distribution of $\bar{X} = \frac{1}{100}(X_1 + \cdots + X_{100})$

$$X_1 + \cdots + X_{100} \sim N(1500, 5^2 \times 100)$$

So

$$\bar{X} \sim N\left(15, \frac{5^2 \times 100}{100^2}\right) = N\left(15, \frac{5^2}{100}\right) = N(15, 0.5^2)$$

Central limit theorem (CLT)

- Informally: given an iid sample of X, sample mean $\hat{\mu}_n$ has approximately *Gaussian distribution* (with appropriate scaling)

- Note: this happens for *any* distribution of X!
  - X can be discrete, continuous (e.g. Bernoulli, exponential, …)

- This highlights the *central role* of Gaussian distribution in probability and statistics
  - One distribution to rule them all

Central limit theorem

Formal statement Let $X_1, \ldots, X_n$ be an iid sample with mean μ and variance σ^2. Then for (moderately large) n:

$$\bar{X}_n \text{ approximately} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Equivalently,

$$\bar{X}_n - \mu \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$\sqrt{n}(\bar{X}_n - \mu) \sim N(0, \sigma^2)$$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$$

hold approximately

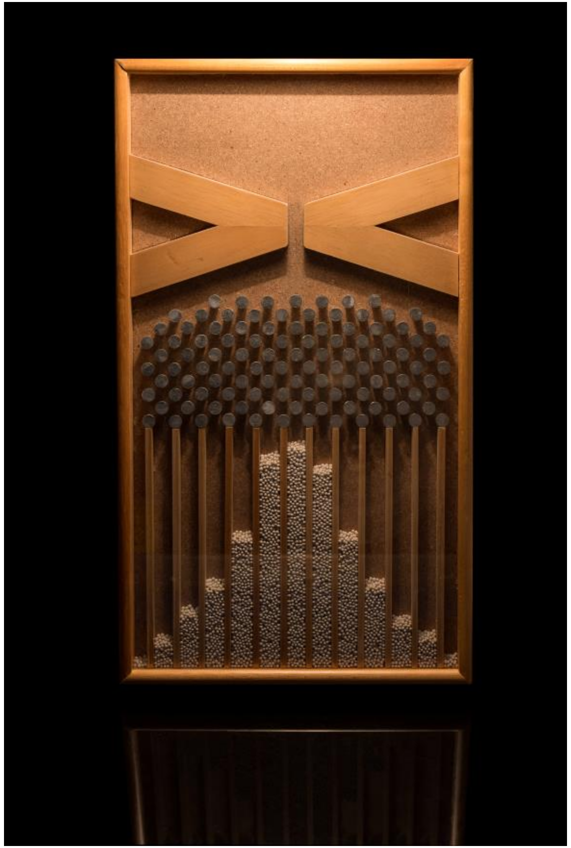
Experimental validation 1: Galton Boards

- Bead has 10 chances hitting a peg
- each time a peg is hit, bead randomly bounces to the left or the right with equal probabilities

- We can represent the final location of the bead as

$$X_1 + \cdots + X_{10}$$

where X_i's is an IID sample of $\mathrm{Uniform}(\{-1, +1\})$

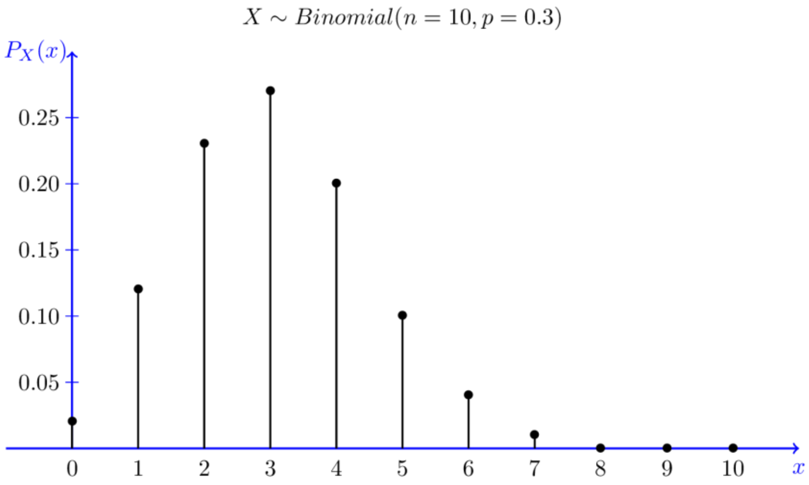


Galton boards

Sir Francis Galton demonstrates his "Galton board" or "quincunx" at the Royal Institution. He saw this pinball-like apparatus as an analogy for the inheritance of genetic traits like stature. The pinballs accumulate in a bell-shaped curve that is similar to the distribution of human heights. The puzzle of why human heights

Binomial distribution ≈ Normal distribution?

- Binomial distribution looking similar to normal distribution is not a coincidence



Example

- $X \sim \text{Bin}(10, 0.3)$

- Equivalent to $X = X_1 + \cdots + X_{10}$, each $X_i \sim \text{Bernoulli}(0.3)$

which is close to a normal distribution by CLT

Experimental validation 2: python simulations

```
# Parameters
sample_size = 30  # Number of observations per sample (n)
num_samples = 1000  # Number of samples to take
a, b = 0, 10  # Uniform distribution parameters (U[a, b])

# Draw multiple samples from a uniform distribution and compute their means
sample_means = [np.mean(np.random.uniform(a, b, sample_size)) for _ in range(num_samples)]
```

$X \sim \text{Uniform}([0,10])$

Each element is a separate $\bar{X}_{30}$ induced by a different sample

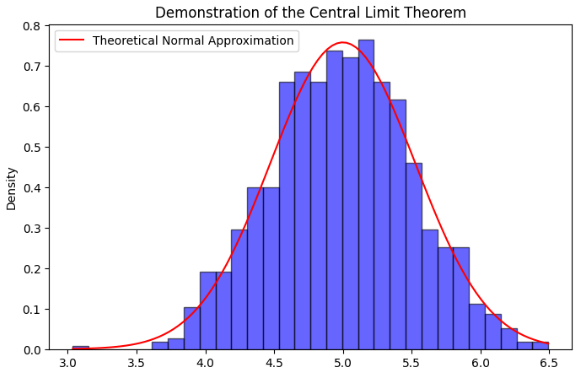
We have 1000 samples => 1000 $\bar{X}_{30}$'s

- What does CLT predict about the distribution of $\bar{X}_{30}$?
  - Approximately $N\left(\mu, \frac{\sigma^2}{30}\right) = N(5, 0.52^2)$
- Let's see if this prediction is accurate..

Experimental validation 2: python simulations

```
# Plot the histogram of sample means
plt.figure(figsize=(8, 5))
plt.hist(sample_means, bins=30, density=True, alpha=0.6, color='b', edgecolor='black')
plt.title("Demonstration of the Central Limit Theorem")
plt.xlabel("Sample Mean")
plt.ylabel("Density")
plt.legend()
plt.show()
```

- CLT predicted that elements from sample_means are roughly $N(5, 0.52^2)$
- Let's see..



- Experiments agree pretty well with theory

Central limit theorem: application

Example X_i: customer spending with $\mu = 80$, $\sigma = 40$.
Approximate the probability that the average spending of 100 customers is 10% below expected value

$$P(\bar{X}_n \leq 72)$$

Solution by CLT, $\bar{X}_n \sim N\left(80, \frac{40^2}{100}\right) = N(80, 4^2)$ approximately

Therefore, $Z = \frac{\bar{X}_n - 80}{4} \sim N(0,1)$ approximately

$$P(\bar{X}_n \leq 72) = P(Z \leq -2) \approx \Phi(-2) = 0.023$$

Review

We have covered a lot of ground on probability…

Discrete Random Variables

- Definition of sample space / random events
- Axioms of probability
- Uniform probability of random event
- Fundamental rules of probability (chain rule, conditional, law of total probability)

Probability Distributions

- Random Variables
- Useful discrete probability mass functions
- Introduction to continuous probability
- Useful probability density functions

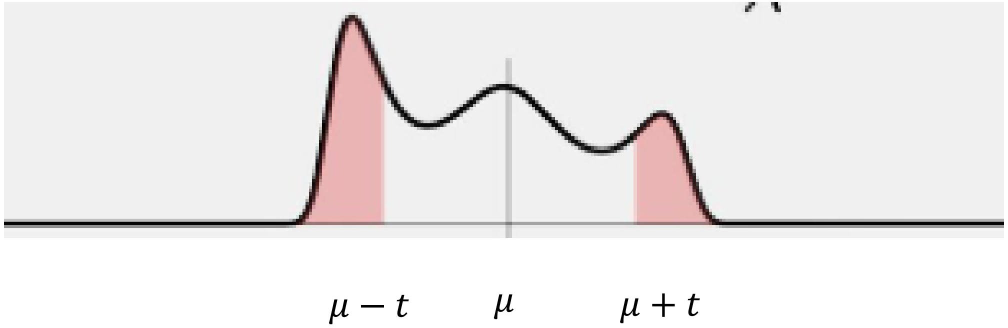
Moments / Independence

- Expected Value
- Linearity
- Variance, Covariance, Corr.
- Dependent / Independent RVs

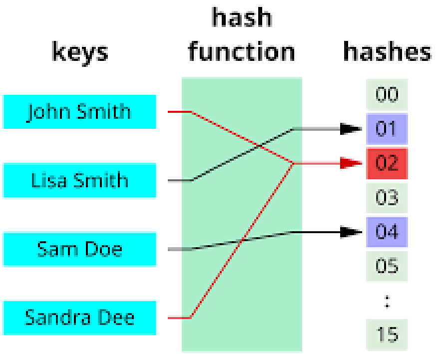
Probability: closing thoughts

Markov and Chebyshev's inequalities: how to make inferences on where X lies when we do not know its distribution exactly?

$$P(|X-\mu| \geq t) \leq \frac{\text{Var}[X]}{t^2}$$



Probability: closing thoughts

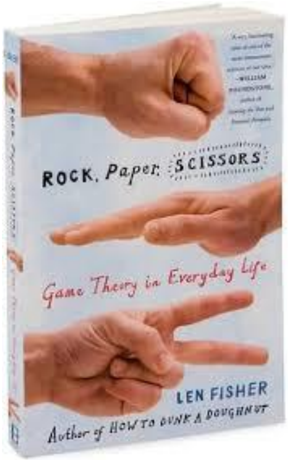


Randomization is also a useful tool for algorithm design

- Example: Hashing
  - Using randomization to mitigate collision

Randomization is fundamental in playing games

- Examples: rock paper scissors, penalty kick



Probability: philosophical remarks

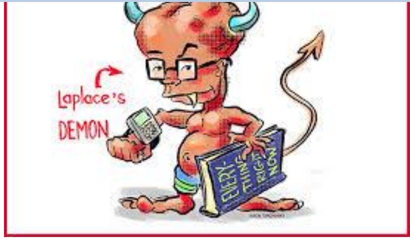
- Pierre-Simon Laplace (1812): thought experiment

- I toss the coin, you guess how it will land
- Probability of predicting correctly: 1

- Laplace’s view: probability does not actually exist; it is a useful way to quantify human ignorance though

Avi Wigderson: Randomness

Probability: philosophical remarks



- Laplace’s demon
  - a **hypothetical intelligence** that knows the exact position & momentum of every particle in the universe.
  - Using the laws of classical mechanics, it could predict the entire future with absolute certainty.
  - Suggests a fully deterministic universe, where free will is an illusion.

- Under debate: Heisenberg’s Uncertainty Principle (1927) precludes such perfect knowledge of particles
- “God does not play dice with the universe” – Einstein, 1926