CSC380: Principles of Data Science

Probability 4

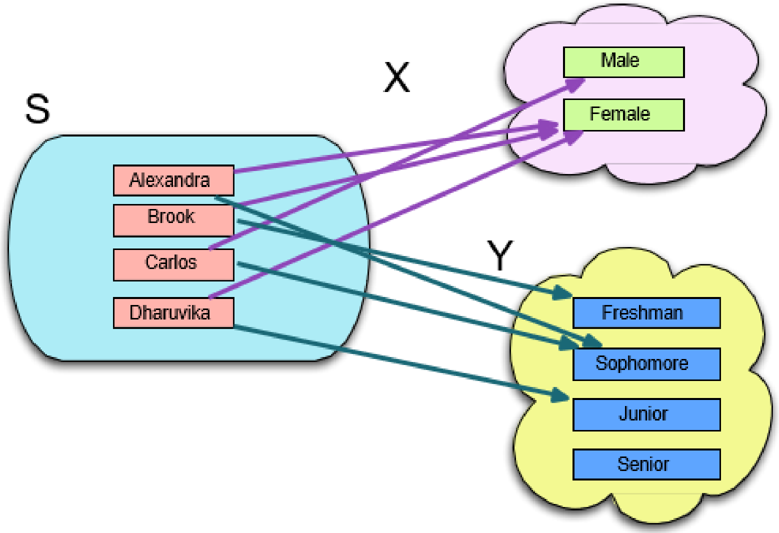
Xinchen Yu

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

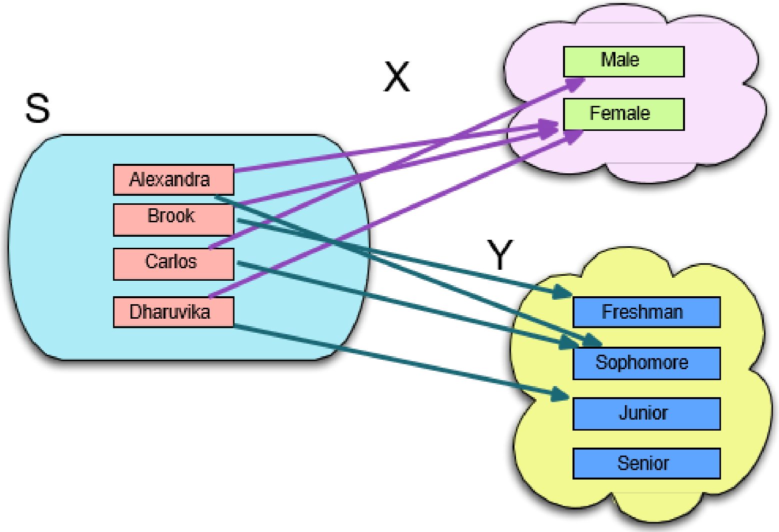
Alexandra

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

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 of them, say X.

- Named the *marginal distribution* of X.

	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

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

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

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

Marginal distributions

	y				
x	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

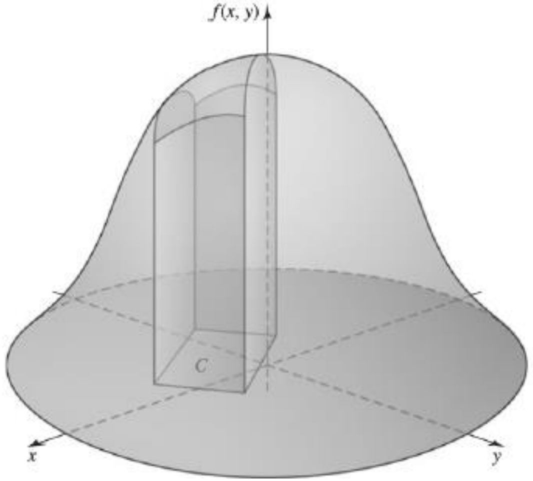
f_2: *marginal distribution* of Y

$$f_1(X=1) = \sum_y f(1,y) = 0.1 + 0 + 0.1 + 0 = 0.2$$

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

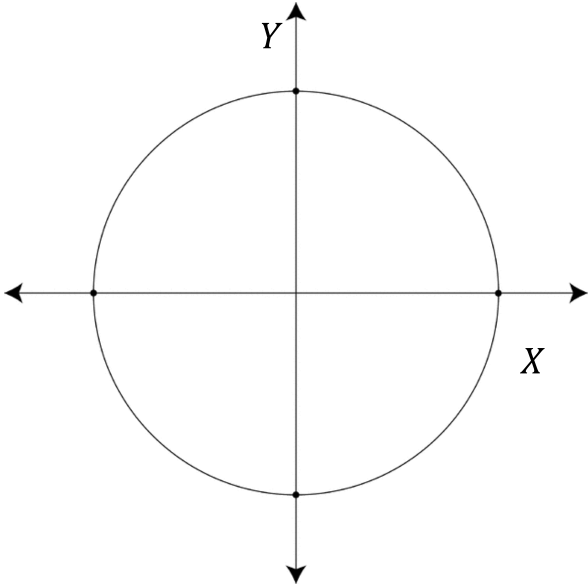
double integral: the *volume* under the surface

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

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

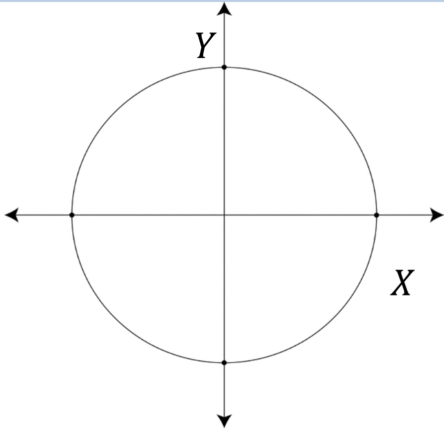
- This is called "the Uniform distribution over the unit disk"

Example: dartboard

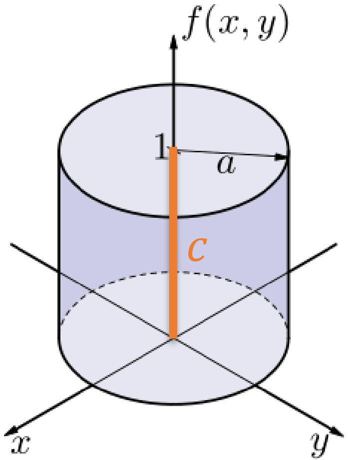
The PDF of X, Y is

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

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), how to find X's PDF f_1?

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

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

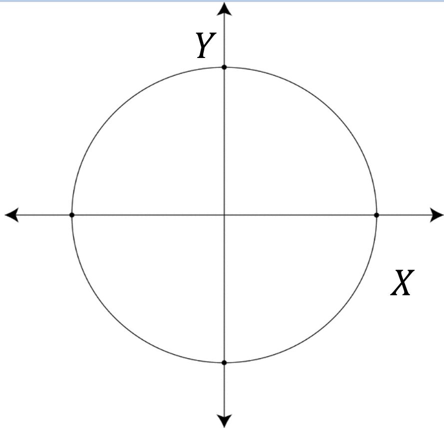
How about Y's PDF f_2?

- Marginalize out X

Example: dartboard

The PDF of X, Y is

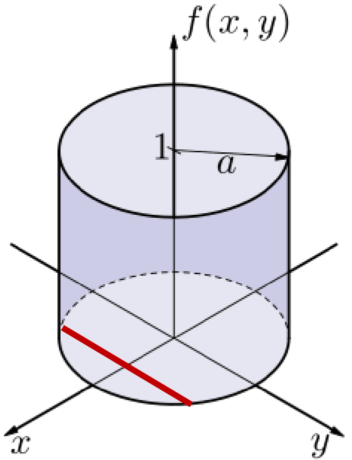
$$f(x,y) = \begin{cases} \frac{1}{\pi}, x^2 + y^2 \leq 1 \\ 0, \text{otherwise} \end{cases}$$



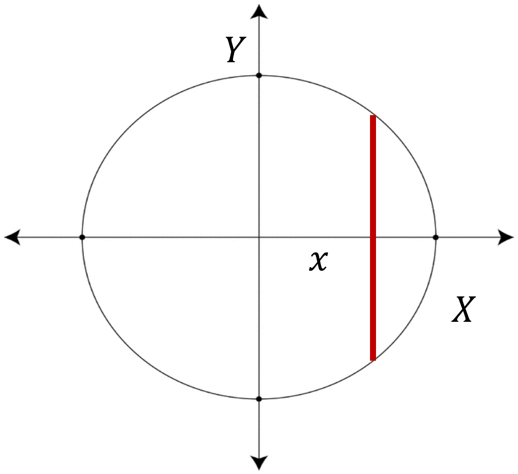
What is the marginal distribution over X?

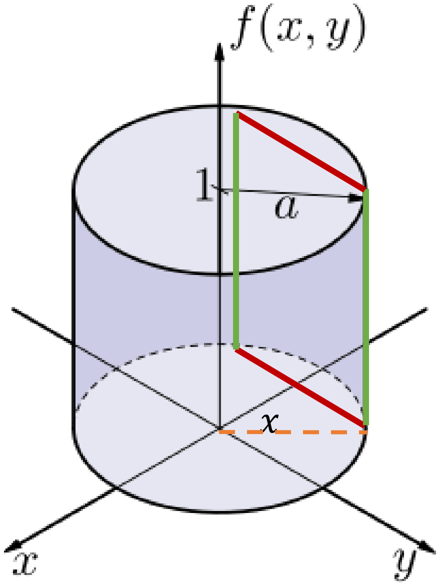
$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$$

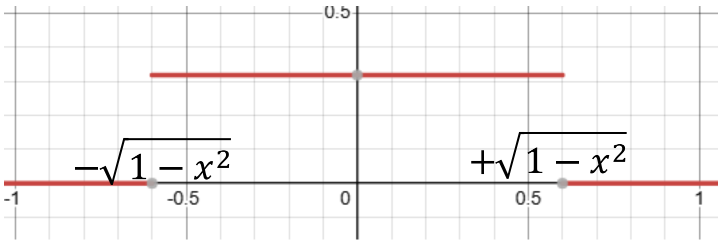
How to find this integral?



Example: dartboard



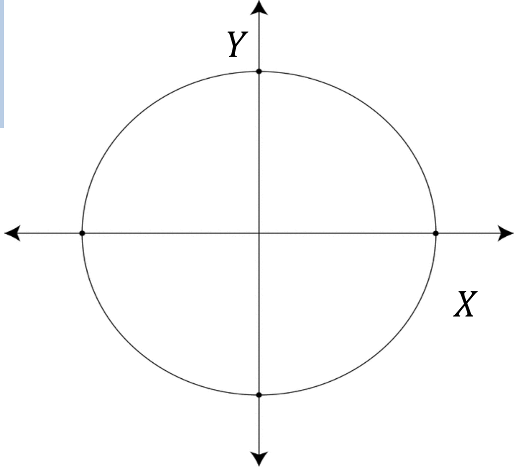




For a fixed $x \in [-1, 1]$, we can think of $f(x)$ is the area of the slice:

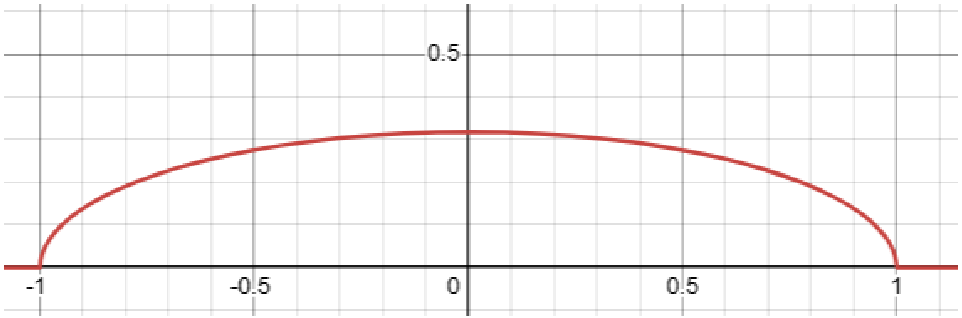
- height: $\frac{1}{\pi}$, width: $2 \cdot \sqrt{1-x^2}$
- $f_1(x) = \frac{2}{\pi} \cdot \sqrt{1-x^2}$

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

Joint distribution of more than 3 RVs

- 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 $P(A=0, B=0), \ldots,\ P(A=1, B=1)$

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

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..

Plan

- Multivariate RVs
  - $f_1(x) = \sum_y f(x, y)$ for discrete X, Y
  - $f_1(x) = \int_R f(x, y)\, dy$ for continuous X, Y
- Independence of RVs
- Conditional distribution of RVs
- Mean of conditional distribution
- Finding distribution of X + Y when they are independent

Independence of RVs

Independence of two RVs

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

$$\underbrace{f(x,y)}_{\text{PMF or PDF}} = \underbrace{f_1(x)}_{\text{Marginal of X}} \cdot \underbrace{f_2(y)}_{\text{Marginal of Y}}, \quad \textit{for all } x, y$$

- E.g. for discrete X, Y,

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

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	$\boxed{0}$	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 $X_1, \dots, X_n$ are independent if their joint PMF or PDF satisfy

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

PMFs or PDFs   Marginal for X_1 Marginal for X_n

for all $x_1, \dots, x_n$

This captures many real-world applications:

- Independent trials: each X_i is $\mathrm{Bernoulli}(p)$
  - Flip 10 coins: $x_1, x_2, \dots, x_{10}$

True or False?

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

- False

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

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

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)

Conditional distributions of RVs

Conditional distributions (discrete)

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

- Conditional PMF of X given $Y = y$:
$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$
Same as $\frac{P(X=x, Y=y)}{P(Y=y)} = P(X = x \mid Y = y)$

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

In-class activity (discrete case)

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

Joint PMF of X, Y, find $\mathrm{P}(\mathrm{X}=0|\mathrm{Y}=1)$

	Brand Y					
Stolen X	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

Solution

$$\mathrm{P}(\mathrm{X}=0|\mathrm{Y}=1)=\frac{\mathrm{P}(\mathrm{X}=0,\mathrm{Y}=1)}{\mathrm{P}(\mathrm{Y}=1)}=\frac{0.129}{0.139}=0.928$$

In-class activity (discrete case)

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

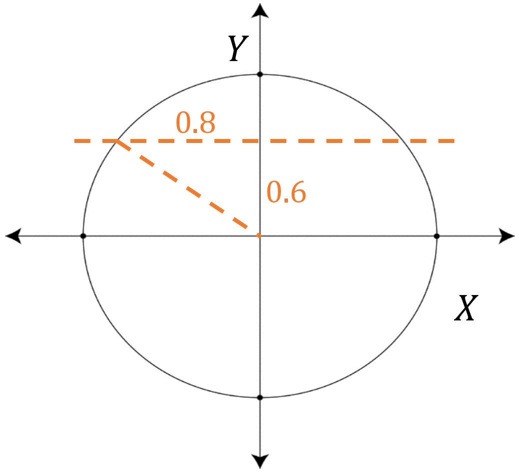
Stolen X	Brand Y 1	2	3	4	5
0	0.928	0.968	0.994	0.993	0.991
1	0.072	0.032	0.006	0.007	0.009

Conditional distributions (continuous)

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

- Conditional PDF of X given Y:

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



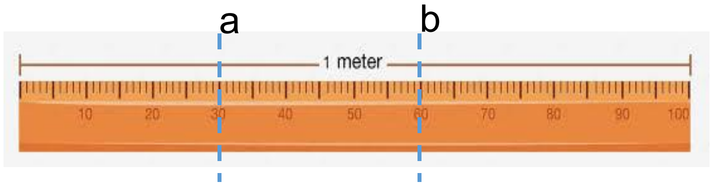
Example Conditional distribution of X given $Y = 0.6$:

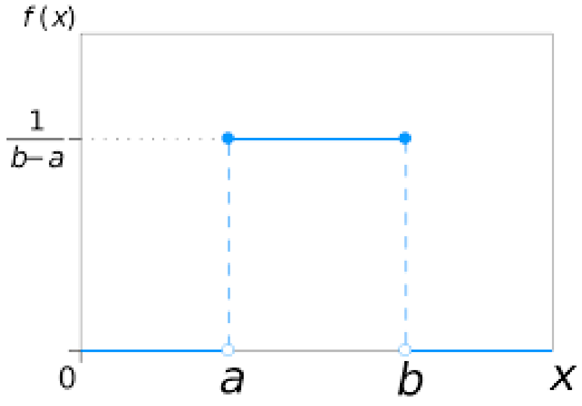
Answer: Uniform($[-0.8, +0.8]$), $\quad f(x) = \frac{1}{0.8+0.8} = \frac{1}{1.6}$

Recap: Uniform Distribution

- $X \sim \text{Uniform}([a, b])$

$$f(x) = \begin{cases} 0, & y < a \\ \frac{1}{b-a}, & y \in [a, b] \\ 0, & y > b \end{cases}$$





Conditional distributions & independence

Fact X,Y are independent

$\Leftrightarrow$ for all y, $g(x|y)$ are all equal to $f(x)$

Here, g, f are PMF or PDF depending on the types of X,Y

Assume Y can only take the value 1, 2, and 3. We say X,Y are independent when

- $f(X = x) = g(X = x|Y = 1)$, and
- $f(X = x) = g(X = x|Y = 2)$, and
- $f(X = x) = g(X = x|Y = 3)$

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

In-class activity

Joint PMF

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

$f(x)$

conditional PMF of X, Y

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

$g(x|1)$ $g(x|2)$

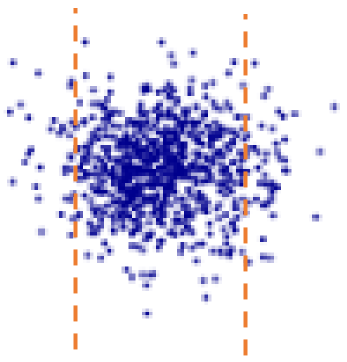
Question: are X,Y independent?

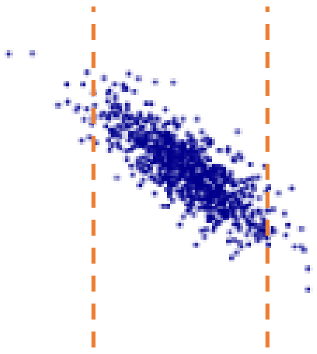
$$g(x = 0|1) = 0.928$$
$$f(x = 0) = 0.976$$

Not equal, so not independent

Independence: visualization

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





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:

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

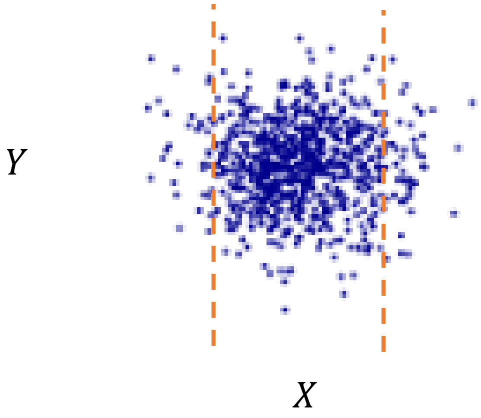
  Conditional PMF

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

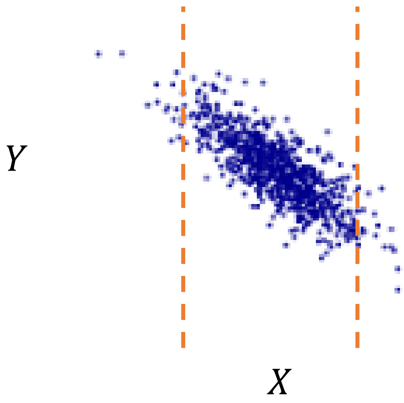
  Conditional PDF

Independence: visualization

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



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



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

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

Recap

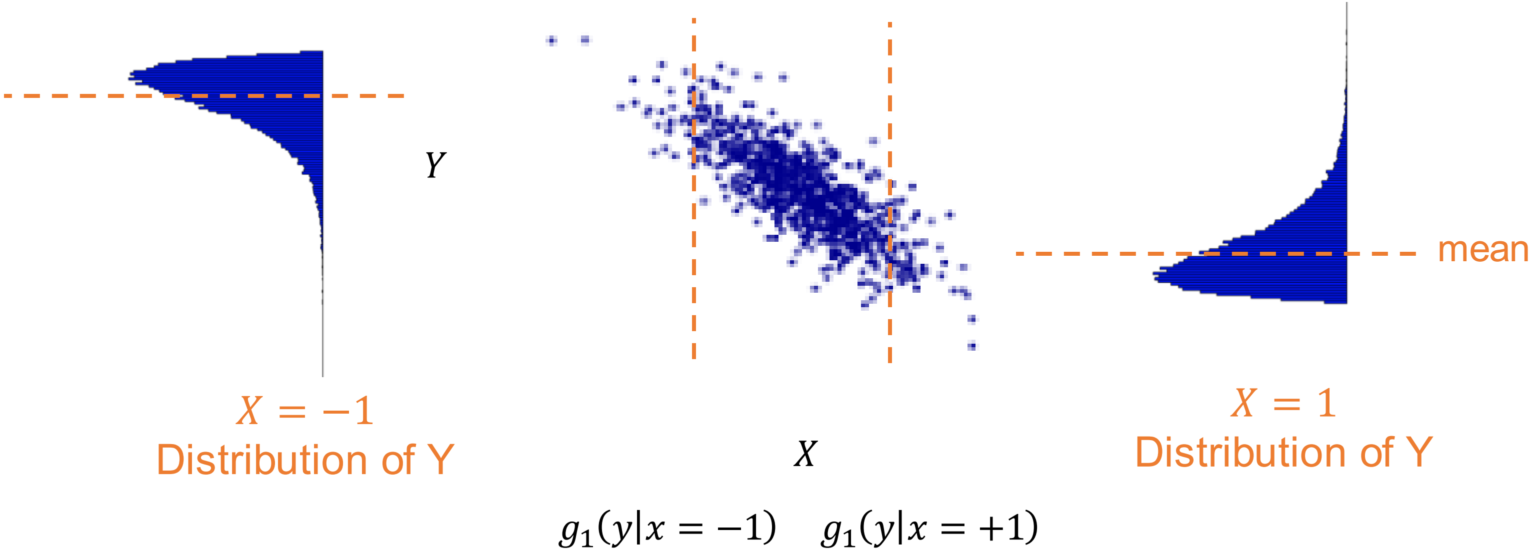
- RVs $X_1, \dots, X_n$ are independent if their joint PMF or PDF satisfy
$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n),$$

- Conditional PDF of X given Y:
$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$

- X,Y are independent ⇔ for all y, $g(x \mid y) = f(x)$

Independence: visualization

X, Y not independent



Which one is larger, $\mathrm{E}[Y|X = -1]$ or $\mathrm{E}[Y|X = +1]$?
Answer: compare the mean of the conditional distribution, so the former has higher mean

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?

$x = 1,2,3,4$

$$\frac{P(X = 1, Y = 5)}{P(Y = 5)} = \frac{1}{4}$$

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

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

Finding distributions of RVs

Finding distributions of random variables

Assume $Z = r(X, Y) = X + Y$, how to find distribution of Z ?

- Example: Total cost $Z = X + Y$, where X = food expenses, Y = transportation cost

- Step 1: find potential values of Z

- Step 2: find the probability that Z takes each possible value

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 1: what values can $X + Y$ take?

2, 3, 4, 5

Step 2: for each possible value, what is the probability?

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 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)	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

Finding distributions of random variables

- 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

Finding distributions of random variables

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

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

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

$$= (1+1) \cdot P(X = 1, Y = 1) + \cdots + (2+3) \cdot P(X = 2, Y = 3)$$

$$= 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} + 5 \cdot \frac{1}{6} = \frac{7}{2}$$

Expectation and Variance revisited

Recap: expectation and variance

- Mean
  - $E[a \cdot X] = a \cdot E[X]$
  - $E[a \cdot X + b] = a \cdot E[X] + b$
  - $E[X \cdot Y] = E[X] \cdot E[Y]$ when X, Y are independent
- Variance
  - $Var(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2$
  - $Var(a \cdot X) = a^2 \cdot Var(X)$

- Plan
  - $E[X + Y]$?
  - $Var[X + Y]$?

Linearity of expectation

Fact Expectation of sum is sum of expectations

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

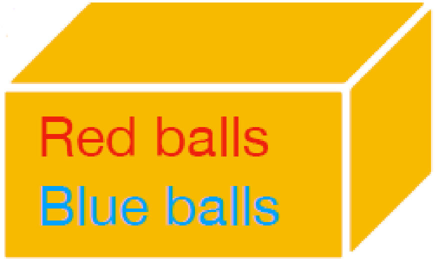
Example: betting on two games

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 $p = 20\%$

- Randomly sample $n = 100$ balls with replacement
- X: number of R balls in the sample.
- Let $X_i = 1$ if i-th ball is R, and 0 otherwise
- $\mathrm{E}[X] =?$

Solution

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

Each X_i has expectation p

$\Rightarrow \mathrm{E}[X] = E[X_1] + \cdots + E[X_n] = np = 20$

Linearity of Variance?

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

- It depends..
  - when $Y = -X$,

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

$$\mathrm{Var}[\mathrm{Y}] = \mathrm{Var}[-1 \cdot \mathrm{X}] = 1^2 \cdot \mathrm{Var}[\mathrm{X}] = \mathrm{Var}[\mathrm{X}]$$

  => Left-hand side < Right-hand side
  - when $Y = X$,

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

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

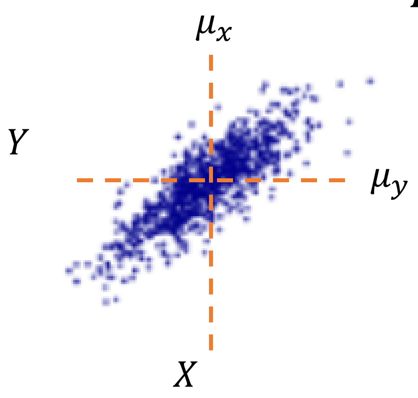
  => Left-hand side > Right-hand side
- Extra correction is needed to balance the equation: covariance!

Covariance

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

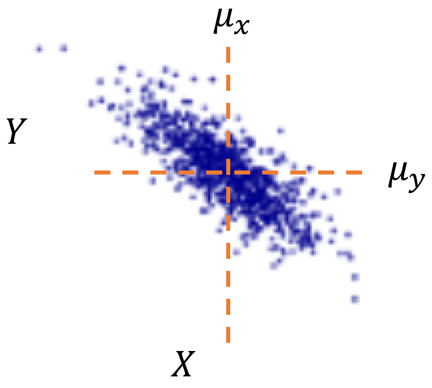
$$\mathrm{Cov}(X,Y) = \mathrm{E}\big[(X-\mu_x)(Y-\mu_y)\big]$$

$$= \ E[XY] - \mu_x \mu_y$$



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

Positive correlation: X, Y simultaneously large or small



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

Calculating covariance

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

Example Find $\mathrm{Cov}(X, Y)$ 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

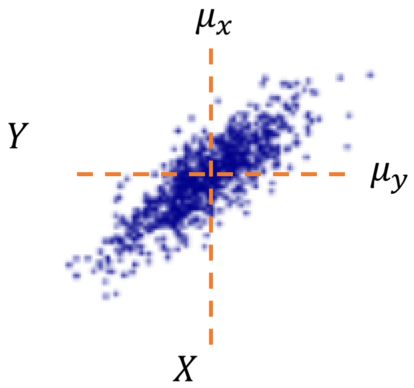
$$E[XY] = \sum_{x,y} xy \; P(X = x, Y = y) = 0 \cdot 0 \cdot \frac{1}{2} + 1 \cdot 1 \cdot \frac{1}{2} = \frac{1}{2}$$

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

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

Properties of Covariance

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



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

Properties

- $\mathrm{Cov}(X,X) = \mathrm{E}[(X-\mu_x)^2] = \mathrm{Var}[X]$
- $\mathrm{Cov}(X+a, Y+b) = \mathrm{Cov}(X,Y)$ Covariance is invariant to shifting
- $\mathrm{Cov}(cX, dY) = cd\,\mathrm{Cov}(X,Y)$ Covariance is sensitive to scaling

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

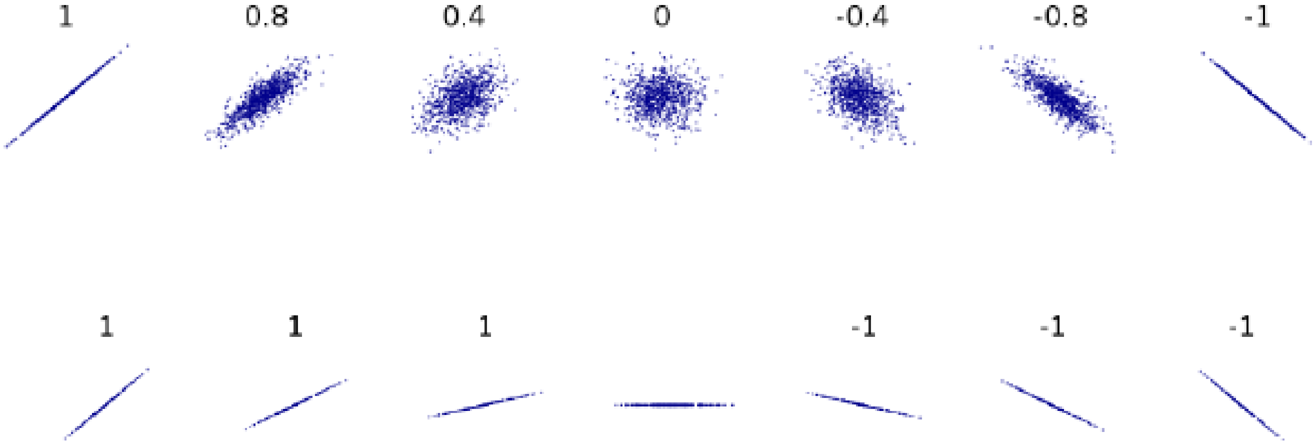
$$\text{Correlation of } X, Y = \rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

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

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

Sanity check:

- When $Y = -X$: $2\mathrm{Cov}(X, Y) = -2\,\mathrm{Var}[X]$
  - LHS = RHS = 0

$\mathrm{Cov}(X, Y) = E[X \cdot Y] - \mu_x \mu_y$
$= E[X \cdot -X] - E[X] \cdot E[-X]$
$= -E[X^2] + (E[X])^2 = -Var[X]$

- 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,

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0$$
$$\mathrm{Var}(X+Y) = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Justification

$$E[XY] = \sum_x \sum_y x\, y\, f(x,y) \overset{\text{independence}}{=} \sum_x \sum_y x\, y\, f_1(x) f_2(y)$$
$$= \sum_x x f_1(x) \sum_y y\, f_2(y) = \sum_x x f_1(x)\, \mu_y = \mu_x \mu_y$$

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.

Find the parameters of Z's distribution: $Z \sim N(?\,,?\,)$

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

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

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

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

$$E[a \cdot X] = a \cdot E[X]$$
$$Var[a \cdot X] = a^2 \cdot Var[X]$$

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

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

$$\frac{1}{3}(X_1 + X_2 + X_3) \sim N\left(\frac{120}{3}, \frac{3 \times 0.1^2}{3^2}\right) = N\left(40, \frac{0.1^2}{3}\right)$$

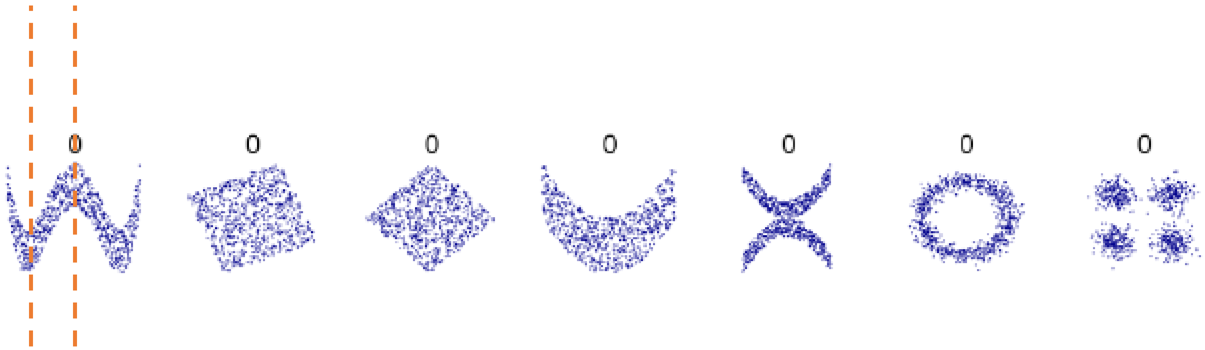
Averaging over multiple measurements reduces measurement error!

In class exercise: a concrete counterexample

- Does zero covariance imply independence?
  - No: covariance only measures strength of *linear relationship* between X, Y

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

$\not\Longleftarrow$



In class exercise: a concrete counterexample

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

Counterexample $X \sim \mathrm{Uniform}(\{-1,0,1\})$. $Y = X^2$. Check independence and find covariance.

Step 1: Fill out the PMF table for X and Y

	x=-1	x=0	x=1
y=0			
y=1			

In class exercise: a concrete counterexample

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

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

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

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

$$X_1, X_2, X_3, X_4$$