CSE 312

Foundations of Computing II

Lecture 19: Joint Distributions

Stream Model – Problem Setup

Input: sequence (aka. "stream") of N elements $x_1, x_2, \ldots, x_N$ from a known universe U (e.g., 8-byte integers).

Goal: perform a computation on the input, in a single left to right pass, where:

- Elements processed in real time
- Can't store the full data ⇒ use minimal amount of storage while maintaining working "summary"

Today: Counting distinct elements

32, 12, 14, 32, 7, 12, 32, 7, 32, 12, 4

Input: sequence (aka. "stream") of N elements $x_1, x_2, \ldots, x_N$ from a known universe U (e.g., 8-byte integers).

Goal: count number of distinct elements

Constraint: Elements processed in real time

– use minimal amount of storage while maintaining working "summary"

Detour – Min of I.I.D. Uniforms

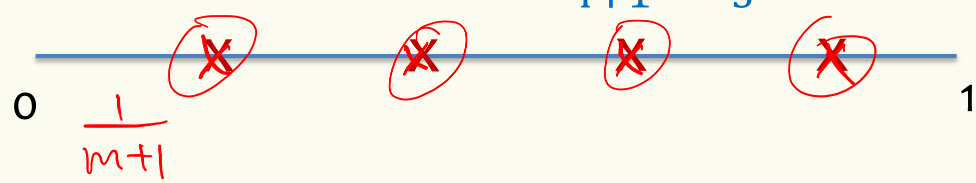
If $Y_1, \cdots, Y_m \sim \text{Unif}(0,1)$ (iid) where do we expect the points to end up?

In general, $\mathbb{E}[\min(Y_1, \cdots, Y_m)] = \frac{1}{m+1}$

$m = 1$ $\qquad \mathbb{E}[\min(Y_1)] = \frac{1}{1+1} = \frac{1}{2}$

$m = 2$ $\qquad \mathbb{E}[\min(Y_1, Y_2)] = \frac{1}{2+1} = \frac{1}{3}$

$m = 4$ $\qquad \mathbb{E}[\min(Y_1, \cdots, Y_4)] = \frac{1}{4+1} = \frac{1}{5}$



Distinct Elements – Hashing into $[0, 1]$

x_1 x_2
$h(x_1)$ $h(x_2)$

Hash function $h: U \to [0,1]$
Assumption: For all $x \in U$, $h(x) \sim \mathrm{Unif}(0,1)$ and mutually independent

$x_1 = 5$	$x_2 = 2$	$x_3 = 27$	$x_4 = 35$	$x_5 = 4$
$h(5)$	$h(2)$	$h(27)$	$h(35)$	$h(4)$

5 distinct elements

$\to$ 5 i.i.d. RVs $h(x_1), \dots, h(x_5) \sim \mathrm{Unif}(0,1)$

$\frac{1}{m+1}$

$$\to \mathbb{E}[\min\{h(x_1), \dots, h(x_5)\}] = \frac{1}{5+1} = \frac{1}{6}$$

Distinct Elements – Hashing into $[0, 1]$

Hash function $h: U \to [0,1]$
Assumption: For all $x \in U$, $h(x) \sim \text{Unif}(0,1)$ and mutually independent

$x_1 = 5$ $x_2 = 2$ $x_3 = 27$ $x_4 = 5$ $x_5 = 4$

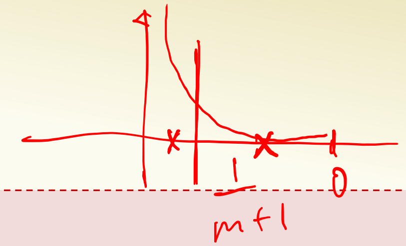
$h(5)$ $h(2)$ $h(27)$ $h(5)$ $h(4)$

4 distinct elements

$\Rightarrow$ 4 i.i.d. RVs $h(x_1), h(x_2), h(x_3), h(x_5) \sim \text{Unif}(0,1)$ and $h(x_1) = h(x_4)$

$\Rightarrow \mathbb{E}[\min\{h(x_1), \dots, h(x_5)\}] = \mathbb{E}[\min\{h(x_1), h(x_2), h(x_3), h(x_5)\}] = \frac{1}{4+1}$

Distinct Elements – Hashing into $[0, 1]$



Hash function $h: U \to [0,1]$

Assumption: For all $x \in U$, $h(x) \sim \mathrm{Unif}(0,1)$ and mutually independent

$x_1, x_2, \ldots, x_N$ contains m distinct elements

$h(x_1), h(x_2), \ldots, h(x_N)$ contains m i.i.d. rvs $\sim \mathrm{Unif}(0,1)$

and $N - m$ repeats

$$\mathbb{E}[\min\{h(x_1), \ldots, h(x_N)\}] = \frac{1}{m+1} \iff m = \frac{1}{\mathbb{E}[\min\{h(x_1), \ldots, h(x_N)\}]} - 1$$

The MinHash Algorithm – Idea

$$m = \frac{1}{\mathbb{E}[\min\{h(x_1), \dots, h(x_N)\}]} - 1$$

pretend.

1. Compute $\text{val} = \min\{h(x_1), \dots, h(x_N)\}$
2. Assume that $\text{val} \approx \mathbb{E}[\min\{h(x_1), \dots, h(x_N)\}]$
3. Output $\text{round}\left(\frac{1}{\text{val}} - 1\right)$

The MinHash Algorithm – Implementation

Algorithm **`MinHash`**$(x_1, x_2, \ldots, x_N)$

$\text{val} \leftarrow \infty$

for $i = 1$ **to** N **do**

  $\text{val} \leftarrow \min\{\text{val}, h(x_i)\}$

return $\text{round}\left(\frac{1}{\text{val}} - 1\right)$

Memory cost = just remember val (with sufficient precision)

MinHash Example

Stream: 13, 25, 19, 25, 19, 19

Hashes: 0.51, 0.26, 0.79, 0.26, 0.79, 0.79

What does MinHash return?

Poll: pollev.com/rachel312

a. 1
b. 3
c. 5
d. No idea

$\text{Round}\left(\frac{1}{0.26} - 1\right)$ < 3

MinHash Example II

Stream: 11, 34, 89, 11, 89, 23

Hashes: 0.5, 0.21, 0.94, 0.5, 0.94, 0.1

Output is $\frac{1}{0.1} - 1 = 9$

Clearly, not a very good answer!

Not unlikely: $P(h(x) < 0.1) = 0.1$

The MinHash Algorithm – Problem

Algorithm **MinHash**$(x_1, x_2, \dots, x_N)$

$\text{val} \leftarrow \infty$

for $i = 1$ **to** N **do**

  $\text{val} \leftarrow \min\{\text{val}, h(x_i)\}$

return $\text{round}\left(\frac{1}{\text{val}} - 1\right)$

But, val is not $\mathbb{E}[\text{val}]$!

How far is val from $\mathbb{E}[\text{val}]$?

$$\text{Var}(\text{val}) \approx \frac{1}{(m+1)^2}$$

$\text{val} = \min\{h(x_1), \dots, h(x_N)\}$ $\mathbb{E}[\text{val}] = \frac{1}{m+1}$

How can we reduce the variance?

Idea: Repetition to reduce variance!

Use k **independent** hash functions $h^1, h^2, \cdots h^k$

Algorithm **MinHash**$(x_1, x_2, \ldots, x_N)$

$\mathrm{val}_1, \ldots, \mathrm{val}_k \leftarrow \infty$

for $i = 1$ **to** N **do**

$\quad \mathrm{val}_1 \leftarrow \min\{\mathrm{val}_1, h^1(x_i)\}, \ldots, \mathrm{val}_k \leftarrow \min\{\mathrm{val}_k, h^k(x_i)\}$

$\mathrm{val} \leftarrow \frac{1}{k}\sum_{i=1}^{k} \mathrm{val}_i$

return $\mathrm{round}\left(\frac{1}{\mathrm{val}} - 1\right)$

$$\mathrm{Var}(\mathrm{val}) = \frac{1}{k}\frac{1}{(m+1)^2}$$

MinHash and Estimating # of Distinct Elements in Practice

- MinHash in practice:
  - One also stores the element that has the minimum hash value for each of the k hash functions
    - Then, just given separate MinHashes for sets A and B, can also estimate
      - what fraction of $A \cup B$ is in $A \cap B$; i.e., how similar A and B are

- Another randomized data structure for distinct elements in practice:
  - HyperLoglog - even more space efficient but doesn't have the set combination properties of MinHash

Agenda

- Joint Distributions
  - Cartesian Products
  - Joint PMFs and Joint Range
  - Marginal Distribution
- Conditional Expectation and Law of Total Expectation
- Conditional expectation and LTE for continuous RVs
- Covariance

Why joint distributions?

- Given all of its user's ratings for different movies, and any preferences you have expressed, Netflix wants to recommend a new movie for you.
- Given a large amount of medical data correlating symptoms and personal history with diseases, predict what is ailing a person with a particular medical history and set of symptoms.
- Given current traffic, pedestrian locations, weather, lights, etc. decide whether a self-driving car should slow down or come to a stop

Review Cartesian Product

Definition. Let A and B be sets. The **Cartesian product** of A and B is denoted

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

Example.

$$\{1,2,3\} \times \{4, 5\} = \{(1, 4), (1,5), (2,4), (2,5), (3, 4), (3,5)\}$$

If A and B are finite sets, then $|A \times B| = |A| \cdot |B|$.

The sets don't need to be finite! You can have $\mathbb{R} \times \mathbb{R}$ (often denoted $\mathbb{R}^2$)

Joint PMFs and Joint Range

Definition. Let X and Y be discrete random variables. The **Joint PMF** of X and Y is

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

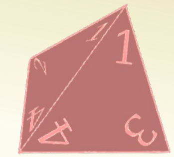
Definition. The **joint range** of $p_{X,Y}$ is

$$\Omega_{X,Y} = \{(c,d) : p_{X,Y}(c,d) > 0\} \subseteq \Omega_X \times \Omega_Y$$

Note that

$$\sum_{(s,t)\in\Omega_{X,Y}} p_{X,Y}(s,t) = 1$$

Example – Weird Dice




Suppose I roll two fair 4-sided die independently. Let X be the value of the first die, and Y be the value of the second die.

$\Omega_X = \{1,2,3,4\}$ and $\Omega_Y = \{1,2,3,4\}$

X\Y	1	2	3	4
1	1/16	1/16	1/16	1/16
2	1/16	1/16	1/16	1/16
3	1/16	1/16	1/16	1/16
4	1/16	1/16	1/16	1/16

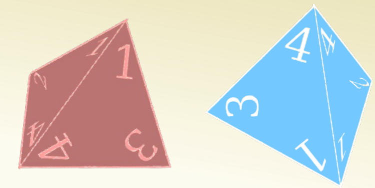
In this problem, the joint PMF is if

$$p_{X,Y}(x,y) = \begin{cases} 1/16 & \text{if } x,y \in \Omega_{X,Y} \\ 0 & \text{otherwise} \end{cases}$$

and the joint range is (since all combinations have non-zero probability)

$\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Example – Weirder Dice



Suppose I roll two fair 4-sided die independently. Let X be the value of the first die, and Y be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

$\Omega_U = \{1,2,3,4\}$ and $\Omega_W = \{1,2,3,4\}$

$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W : u \leq w\} \neq \Omega_U \times \Omega_W$

Poll: pollev.com/rachel312

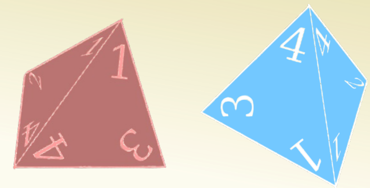
What is $p_{U,W}(1, 3) = P(U = 1, W = 3)$?

a. 1/16
b. 2/16
c. 1/2
d. Not sure

(1, 3) (3, 1) 2/16

U\W	1	2	3	4
1				
2				
3				
4				

Example – Weirder Dice



Suppose I roll two fair 4-sided die independently. Let X be the value of the first die, and Y be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

$\Omega_U = \{1,2,3,4\}$ and $\Omega_W = \{1,2,3,4\}$

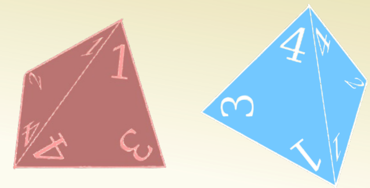
$$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W : u \leq w\} \neq \Omega_U \times \Omega_W$$

The joint PMF $p_{U,W}(u, w) = P(U = u, W = w)$ is

$$p_{U,W}(u, w) = \begin{cases} 2/16 & \text{if } (u, w) \in \Omega_U \times \Omega_W \text{ where } w > u \\ 1/16 & \text{if } (u, w) \in \Omega_U \times \Omega_W \text{ where } w = u \\ 0 & \text{otherwise} \end{cases}$$

U\W	1	2	3	4
1	1/16	2/16	2/16	2/16
2	0	1/16	2/16	2/16
3	0	0	1/16	2/16
4	0	0	0	1/16

Example – Weirder Dice



Suppose I roll two fair 4-sided die independently. Let X be the value of the first die, and Y be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

Suppose we didn't know how to compute $P(U = u)$ directly. Can we figure it out if we know $p_{U,W}(u, w)$?

Just apply LTP over the possible values of W:

$p_U(1) = 7/16$

$p_U(2) = 5/16$

$p_U(3) = 3/16$

$p_U(4) = 1/16$

$P(U=2) = \sum_{i=1}^{4} P(U=2, W=i)$

U\W	1	2	3	4
1	1/16	2/16	2/16	2/16
2	0	1/16	2/16	2/16
3	0	0	1/16	2/16
4	0	0	0	1/16

Marginal PMF

Definition. Let X and Y be discrete random variables and $p_{X,Y}(a,b)$ their joint PMF. The **marginal PMF** of X

$$p_X(a) = \sum_{b \in \Omega_Y} p_{X,Y}(a,b)$$

Similarly, $p_Y(b) = \sum_{a \in \Omega_X} p_{X,Y}(a,b)$

Continuous distributions on $\mathbb{R}\times\mathbb{R}$

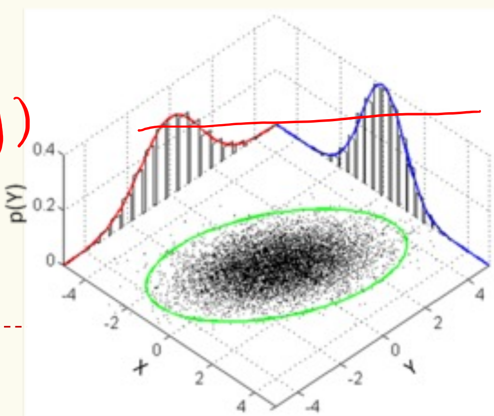
Definition. The **joint probability density function (PDF)** of continuous random variables X and Y is a function $f_{X,Y}$ defined on $\mathbb{R}\times\mathbb{R}$ such that

- $f_{X,Y}(x,y) \geq 0$ for all $x, y \in \mathbb{R}$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x\,\mathrm{d}y = 1$

for $A \subseteq \mathbb{R}\times\mathbb{R}$ the probability that $(X,Y) \in A$ is $\iint_A f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y$

The **(marginal) PDFs** f_X and f_Y are given by

- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x$



Independence and joint distributions

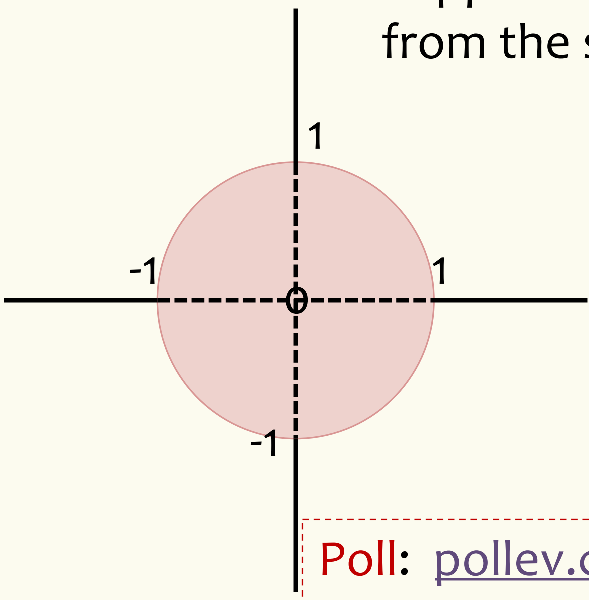
Definition. Discrete random variables X and Y are **independent** iff

- $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all $x \in \Omega_X, y \in \Omega_Y$

Definition. Continuous random variables X and Y are **independent** iff

- $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ for all $x, y \in \mathbb{R}$

Example – Uniform distribution on a unit disk



Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Poll: pollev.com/rachel312

Are X and Y independent?

a. Yes

b. No

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} \mathrm{d}y$$

$$= 2\sqrt{1 - x^2}/\pi$$

Joint Expectation

Definition. Let X and Y be discrete random variables and $p_{X,Y}(a,b)$ their joint PMF. The **expectation** of some function $g(x,y)$ with inputs X and Y

$$\mathbb{E}[g(X,Y)] = \sum_{a \in \Omega_X} \sum_{b \in \Omega_Y} g(a,b) \cdot p_{X,Y}(a,b)$$

Agenda

- Joint Distributions
  - Cartesian Products
  - Joint PMFs and Joint Range
  - Marginal Distribution
- Conditional Expectation and Law of Total Expectation
- Conditional expectation and LTE for continuous RVs
- Covariance

Conditional Expectation

Definition. Let X be a discrete random variable then the **conditional expectation** of X given event A is

$$\mathbb{E}[X \mid A] = \sum_{x \in \Omega_X} x \cdot P(X = x \mid A)$$

Notes:

- Can be phrased as a "random variable version"

$$\mathbb{E}[X | Y = y]$$

- Linearity of expectation still applies here

$$\mathbb{E}[aX + bY + c \mid A] = a\,\mathbb{E}[X \mid A] + b\,\mathbb{E}[Y \mid A] + c$$

Law of Total Expectation

Law of Total Expectation (event version). Let X be a random variable and let events $A_1, \ldots, A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot P(A_i)$$

Law of Total Expectation (random variable version). Let X be a random variable and Y be a discrete random variable. Then,

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X \mid Y = y] \cdot P(Y = y)$$

Proof of Law of Total Expectation

Follows from Law of Total Probability and manipulating sums

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$= \sum_{x \in \Omega_X} x \cdot \sum_{i=1}^{n} P(X = x \mid A_i) \cdot P(A_i)$$ (by LTP)

$$= \sum_{i=1}^{n} P(A_i) \sum_{x \in \Omega_X} x \cdot P(X = x \mid A_i)$$ (change order of sums)

$$= \sum_{i=1}^{n} P(A_i) \cdot \mathbb{E}[X \mid A_i]$$ (def of cond. expect.)

Example – Flipping a Random Number of Coins

Suppose someone gave us $Y \sim \mathrm{Poi}(5)$ fair coins and we wanted to compute the expected number of heads X from flipping those coins.

By the Law of Total Expectation

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{E}[X \mid Y = i] \cdot P(Y = i) = \sum_{i=0}^{\infty} \frac{i}{2} \cdot P(Y = i)$$

$$= \frac{1}{2} \cdot \sum_{i=0}^{\infty} i \cdot P(Y = i)$$

$$= \frac{1}{2} \cdot \mathbb{E}[Y] = \frac{1}{2} \cdot 5 = 2.5$$

Example – Computer Failures

Suppose your computer operates in a sequence of steps, and that at each step i your computer will fail with probability p (independently of other steps).

Let X be the number of steps it takes your computer to fail.

What is $\mathbb{E}[X]$?

Let Y be the indicator random variable for the event of failure in step 1

Then by LTE, $\mathbb{E}[X] = \mathbb{E}[X \mid Y = 1] \cdot P(Y = 1) + \mathbb{E}[X \mid Y = 0] \cdot P(Y = 0)$

$$= 1 \cdot p + \mathbb{E}[X \mid Y = 0] \cdot (1 - p)$$

$$= p + (1 + \mathbb{E}[X]) \cdot (1 - p)$$

since if $Y = 0$ experiment starting at step 2 looks like original experiment

Solving we get $\mathbb{E}[X] = 1/p$

Agenda

- Joint Distributions
  - Cartesian Products
  - Joint PMFs and Joint Range
  - Marginal Distribution
- Conditional Expectation and Law of Total Expectation
- Conditional expectation and LTE for continuous RVs
- Covariance

Conditional Expectation again…

Definition. Let X be a discrete random variable; then the **conditional expectation** of X given event A is

$$\mathbb{E}[X \mid A] = \sum_{x \in \Omega_X} x \cdot P(X = x \mid A)$$

Therefore for X and Y discrete random variables, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \sum_{x \in \Omega_X} x \cdot P(X = x \mid Y = y) = \sum_{x \in \Omega_X} x \cdot p_{X|Y}(x|y)$$

where we **define** $p_{X|Y}(x|y) = P(X = x \mid Y = y) = \dfrac{p_{X,Y}(x, y)}{p_Y(y)}$

Conditional Expectation – Discrete & Continuous

Discrete: Conditional PMF: $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$

Conditional Expectation: $\mathbb{E}[X \mid Y = y] = \sum_{x \in \Omega_X} x \cdot p_{X|Y}(x|y)$

Continuous: Conditional PDF: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Conditional Expectation: $\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx$

Law of Total Expectation - continuous

Law of Total Expectation (event version). Let X be a random variable and let events $A_1, \ldots, A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot P(A_i)$$

Law of Total Expectation (random variable version). Let X and Y be continuous random variables. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y) \; dy$$

Using LTE for Continuous RVs

PDF for $Exp(\lambda)$ is $\begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{o.w.} \end{cases}$

Expectation is $1/\lambda$

Suppose that we first choose $Y \sim Exp(1/2)$ and then choose $X \sim Exp(Y)$.   What is $\mathbb{E}[X]$?

$$f_{X|Y}(x|y) = y\, e^{-x/y}$$

y is fixed here

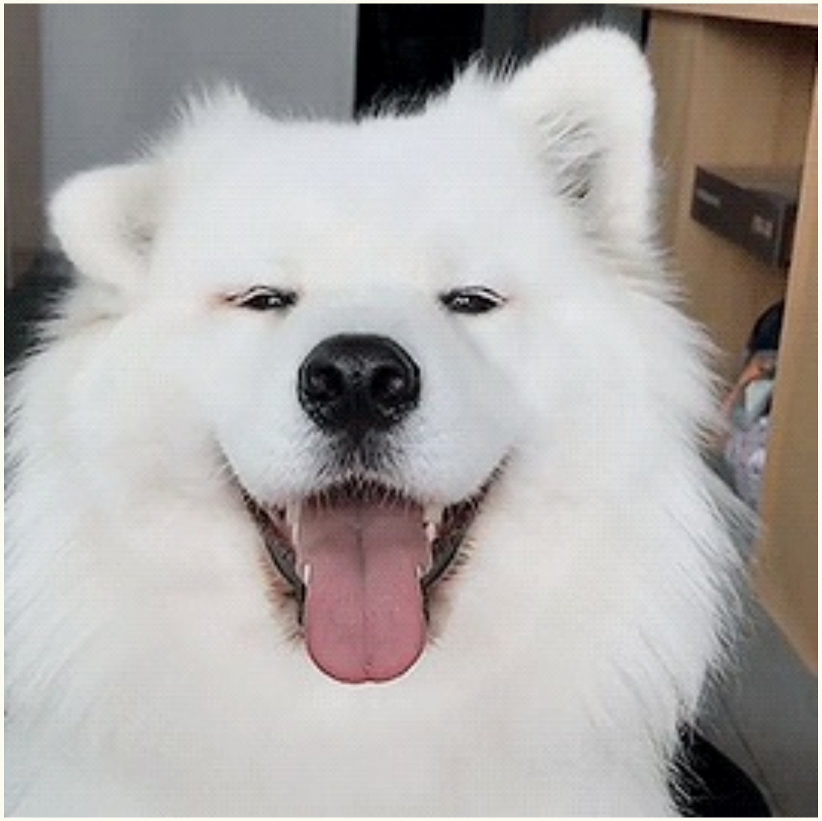
$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx = \int_{-\infty}^{\infty} x \cdot y\, e^{-x/y} dx = y$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y]\, f_Y(y)\, dy = \int_{-\infty}^{\infty} y \cdot 2\, e^{-y/2} dx = \ 2$$

Reference Sheet (with continuous RVs)

	Discrete	**Continuous**
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X=x, Y=y)$	$f_{X,Y}(x,y) \neq P(X=x, Y=y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
Conditional PMF/PDF	$p_{X \mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_{x} x p_{X \mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y)\,dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Brain Break



Agenda

- Joint Distributions
  - Cartesian Products
  - Joint PMFs and Joint Range
  - Marginal Distribution
- Conditional Expectation and Law of Total Expectation
- Conditional expectation and LTE for continuous RVs
- Covariance

Covariance: How correlated are X and Y?

Recall that if X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Definition: The **covariance** of random variables X and Y,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Unlike variance, covariance can be positive or negative. It has has value 0 if the random variables are independent.

Two Covariance examples:

$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Suppose $X \sim \mathrm{Bernoulli}(p)$

If random variable $Y = X$ then

$$\mathrm{Cov}(X, Y) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Var}(X) = p(1 - p)$$

If random variable $Z = -X$ then

$$\begin{aligned} \mathrm{Cov}(X, Z) &= \mathbb{E}[XZ] - \mathbb{E}[X] \cdot \mathbb{E}[Z] \\ &= \mathbb{E}[-X^2] - \mathbb{E}[X] \cdot \mathbb{E}[-X] \\ &= -\mathbb{E}[X^2] + \mathbb{E}[X]^2 = -\mathrm{Var}(X) = -p(1 - p) \end{aligned}$$