etherpad.wikimedia.org/p/312 for (anonymous) questions/comments!

Catch-up/Review :D

CSE 312 24Su

Lecture 19

Announcements

> Fill out **final conflict form** if you have an unavoidable conflict during the current final slot

> Details and resources for the final will be posted later this week/early next week

> Typo on current HW in 3b (net return Is -15 if the design fails)

Where can I find more practice?

> Alex Tsun's textbook (linked on website) has more practice

> Practice exams - practice midterms, practice finals (coming soon)

> Section handouts

Outline for today

> Review discrete vs. continuous random variables

> Normal distributions

> Practice with CLT

> Practice with discrete and continuous joint distributions

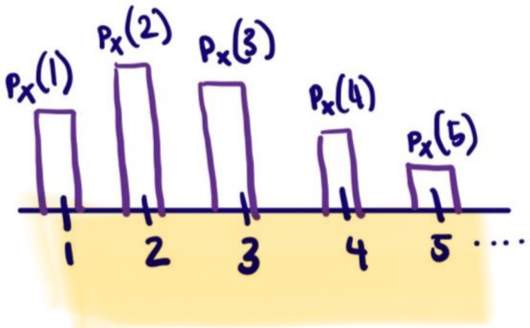
> Practice with Law of total Expectation

> *we won't review tail bounds today, you'll get lots of practice in section tomorrow, and we'll do another problem at the start on Friday!*

Discrete RVs

Support is finite/countably infinite (e.g. *integers*)

Probability mass function $p_X(k)$ gives probability of each value in support



$$\sum_{k \in \Omega_X} p_X(k) = 1$$

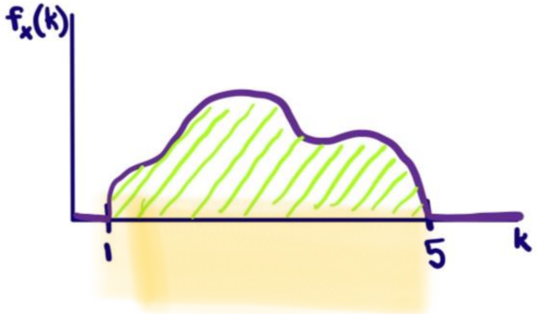
$$0 \leq p_X(k) \leq 1$$

Continuous RVs

Support is uncountably infinite (e.g., *real numbers*)

$\mathbb{P}(\boldsymbol{X} = \boldsymbol{k}) = \frac{1}{\infty} = \mathbf{0}$ so we don't use PMF. instead...

Probability density function $\boldsymbol{f_X(k)}$ describes relative chances of taking values around k



$$\int_{-\infty}^{\infty} f_X(k)\ \mathrm{d}k = 1$$

$$f_X(k) \geq 0$$

Cumulative Distribution Function (CDF) is the function $F_X(k) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{k})$

Sum up the probabilities of values $\leq k$

Integrate over values $\leq k$: $F_X(k) = \int_{-\infty}^{k} f_X(z) dz$

Expectation: $\mathbb{E}[X] = \sum_{k \in \Omega_X} (k \cdot p_X(k))$

$\mathbb{E}[g(X)] = \sum_{k \in \Omega_X} (g(k) \cdot p_X(k))$

Expectation: $\mathbb{E}[X] = \int_{-\infty}^{\infty} z \cdot f_X(z)\ \mathrm{d}z$

$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(z) \cdot f_X(z)\ \mathrm{d}z$

Variance is $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Linearity of expectation and properties of expectation and variance applies in both!

We saw three continuous RVs in the zoo…

> Uniform distribution: $\text{Unif}(a, b)$ is a random variables that takes on a real number uniformly at random between a and b

> Exponential distribution: $\text{Exp}(\lambda)$ is a random variables that tells us the time till the first success

> Normal distribution: $N(\mu, \sigma^2)$ is…

Normal Distributions

Normal Random Variable *(AKA Gaussian)*

There's not a single scenario that follows a normal distribution…
But we're going to see that it shows up in a lot of real world situations!

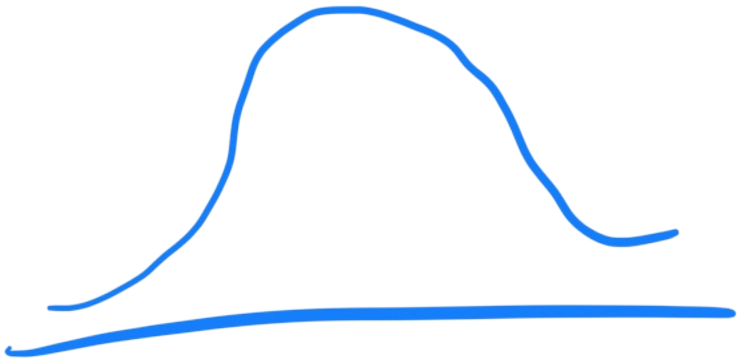
A **normal random variable** $X \sim \mathcal{N}(\mu, \sigma^2)$:

- $\mu = \mathbb{E}[X]$ is the mean
- $\sigma^2 = \mathrm{Var}(X)$ is the variance

  $\sigma = \sqrt{Var(X)}$ is *standard deviation*

and follows this *probability density function*:

$$f_X(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}}$$



Normal Distributions

The *CDF* has no closed form, so instead, we have a table containing values of the CDF for a standard normal random variable $\mathcal{N}(0,1)$.

To find the probability of a normal RV $X \sim \mathcal{N}(\mu, \sigma^2)$ being in some range…

1. **Standardize** the normal random variable: $Z = \frac{X-\mu}{\sigma}$

   note: when we standardize, the numbers left are called z-scores (the number of standard deviations away from the mean (e.g., $\mathbb{P}(Z \geq 2)$ *means we're finding probability of being more than 2 standard deviations away from the mean)*

2. **Write probability expression in terms of** $\boldsymbol{\Phi}(\mathbf{z}) = \mathbb{P}(Z \leq z)$

3. **Look up the value(s)** in the table

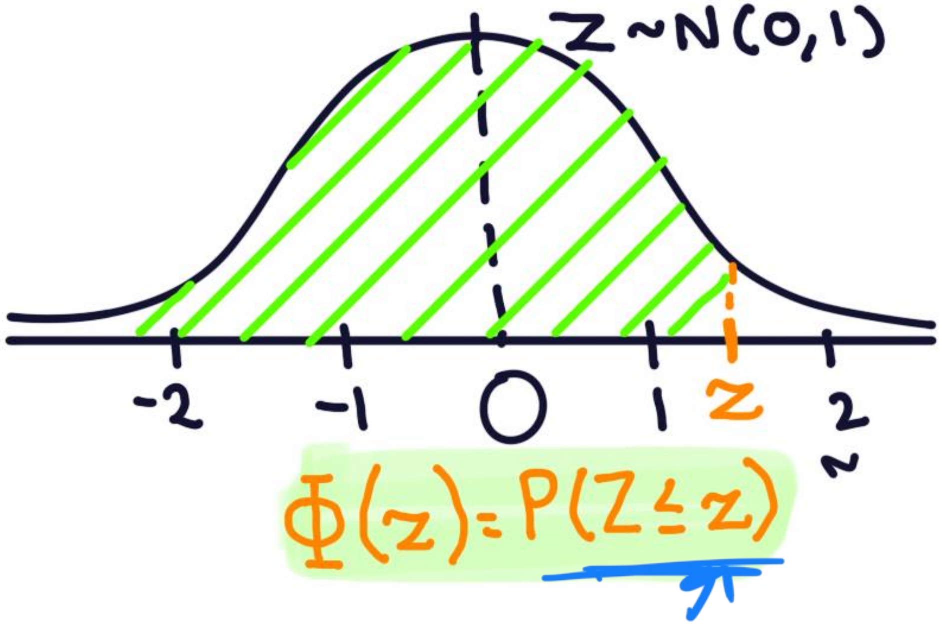
We have a table with precomputed values!

AKA the "z-table", "phi-table"

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

We have a table containing values for the CDF of the **standard normal random variable $\boldsymbol{Z} \sim \mathcal{N}(0,1)$**

> Φ is a function for CDF of $\mathcal{N}(0,1)$
> $\Phi(z) = F_Z(z) = \mathbb{P}(Z \leq z)$



Examples

Let $Y \sim N(10, 5)$

What is $\mathbb{P}(8 \leq Y \leq 12)$?

$12 \quad 14$

1. Standardize. $P\left(\frac{12-10}{\sqrt{5}} < Z < \frac{14-10}{\sqrt{5}}\right) = P(\ldots \leq$

$P(1 \leq Z \leq 2)$

2. $P(Z \leq 2) - P(Z \leq 1)$

$\Phi(2) - \Phi(1)$

What values of c will give $\mathbb{P}(Y \geq c) \leq 0.2$?

Examples

Let $Y \sim N(10, 4)$

What is $\mathbb{P}(12 \leq Y \leq 14)$?

1. Standardize: *subtract the mean, divide by the standard deviation to get* $Z \sim \mathcal{N}(0,1)$

$$\mathbb{P}(8 \leq Y \leq 10) = \mathbb{P}\left(\frac{12-10}{\sqrt{4}} \leq \frac{Y-10}{\sqrt{4}} \leq \frac{14-10}{\sqrt{4}}\right) = \mathbb{P}\left(\frac{12-10}{\sqrt{4}} \leq Z \leq \frac{14-10}{\sqrt{4}}\right) = \mathbb{P}(1 \leq Z \leq 2)$$

2. Write in terms of $\boldsymbol{\Phi}$: $\mathbb{P}(1 \leq Z \leq 2) = \mathbb{P}(Z \leq 2) - \mathbb{P}(Z \leq 1) = \Phi(2) - \Phi(1)$

3. Plug into the z-table: $\mathbb{P}(12 \leq Y \leq 14) \approx 0.97725 - 0.84134 = 0.13591$

What value of $\boldsymbol{c}$ gives $\mathbb{P}(Y \geq c) \geq 0.7$

Examples

Let $Y \sim N(10, 4)$

What is $\mathbb{P}(12 \leq Y \leq 14)$?

1. Standardize: *subtract the mean, divide by the standard deviation to get* $Z \sim \mathcal{N}(0,1)$
$\mathbb{P}(8 \leq Y \leq 10) = \mathbb{P}\left(\frac{12-10}{\sqrt{4}} \leq \frac{Y-10}{\sqrt{4}} \leq \frac{14-10}{\sqrt{4}}\right) = \mathbb{P}\left(\frac{12-10}{\sqrt{4}} \leq Z \leq \frac{14-10}{\sqrt{4}}\right) = \mathbb{P}(1 \leq Z \leq 2)$

2. Write in terms of $\mathbf{\Phi}$: $\mathbb{P}(1 \leq Z \leq 2) = \mathbb{P}(Z \leq 2) - \mathbb{P}(Z \leq 1) = \Phi(2) - \Phi(1)$

3. Plug into the z-table: $\mathbb{P}(12 \leq Y \leq 14) \approx 0.97725 - 0.84134 = 0.13591$

What value of $\boldsymbol{c}$ gives $\mathbb{P}(Y \geq c) \geq 0.7$

1. Standardize: $\mathbb{P}(Y \geq c) = \mathbb{P}(Z \geq \frac{c-10}{\sqrt{4}})$

2. Write in terms of $\mathbf{\Phi}$: $\mathbb{P}\left(Z \geq \frac{c-10}{\sqrt{4}}\right) = \mathbb{P}\left(Z \leq -\frac{c-10}{\sqrt{4}}\right) = \Phi(\frac{c-10}{\sqrt{4}}) \geq 0.7$

3. Reverse z-table lookup: $\Phi(-\frac{c-10}{\sqrt{4}}) \geq 0.7$ --> $-\frac{c-10}{\sqrt{4}} \geq 0.53$ -> $\boldsymbol{c \geq 8.94}$

Examples

Let $Y \sim N(10, 4)$

What values of c will give $\mathbb{P}(|Y - 10| \geq c) \leq 0.3$?

Examples

Let $Y \sim N(10, 4)$

What values of c will give $\mathbb{P}(|Y - 10| \geq c) \leq 0.3$?

get rid of that absolute value… $\mathbb{P}(|Y - 10| \geq c) = \mathbb{P}(Y - 10 \geq c) + \mathbb{P}(Y - 10 \leq -c)$

isolate the Y… $\mathbb{P}(Y \geq c + 10) + \mathbb{P}(Y \leq -c + 10)$

1. Standardize

$$\mathbb{P}(Y \geq c + 10) + \mathbb{P}(Y \leq -c + 10) = \mathbb{P}\left(\frac{Y-10}{\sqrt{4}} \geq \frac{c+10-10}{\sqrt{4}}\right) + \mathbb{P}\left(\frac{Y-10}{\sqrt{4}} \leq \frac{-c+10-10}{\sqrt{4}}\right) =$$

$$= \mathbb{P}\left(Z \geq \frac{c+10-10}{\sqrt{4}}\right) + \mathbb{P}\left(Z \leq \frac{-c+10-10}{\sqrt{4}}\right) = \mathbb{P}\left(Z \geq \frac{c}{\sqrt{4}}\right) + \mathbb{P}\left(Z \leq \frac{-c}{\sqrt{4}}\right)$$

2 Write in terms of Φ: $\mathbb{P}\left(Z \geq \frac{c}{\sqrt{4}}\right) + \mathbb{P}\left(Z \leq \frac{-c}{\sqrt{4}}\right) = 2 \cdot \mathbb{P}\left(Z \geq \frac{c}{\sqrt{4}}\right) = 2 \cdot \left(1 - \Phi\left(\frac{c}{\sqrt{4}}\right)\right) \leq 0.3$

3. Reverse z-table lookup to solve for c:

$$2 \cdot \left(1 - \Phi\left(\frac{c}{\sqrt{4}}\right)\right) \leq 0.3 \text{ --> } \Phi\left(\frac{c}{\sqrt{4}}\right) \geq 0.85 \text{ --> } \frac{c}{\sqrt{4}} \geq 1.4 \text{ --> } \boldsymbol{c \geq 2.8}$$

Central Limit Theorem

~~Proof by double counting~~

What is the Central Limit Theorem?

The **central limit theorem** tells us that a **sum of i.i.d** (independent and identically distributed) **random variables** can be **approximated** as a **normal distribution.** *This approximation gets more accurate as we sum more and more random variables togethers.*

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables, each with mean μ and variance σ^2

Let $Y_n = X_1 + X_2 + \cdots + X_n$

As $n \to \infty$, Y_n approaches a normal distribution $\mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$

(i.e., CDF of Y_n converges to the CDF of $\mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$)

Outline of CLT steps

1. Setup the problem (e.g., $X = \sum_{i=1}^{n} X_i$, X_i are i.i.d., and we want $\mathbb{P}(X \leq k)$)
   Write event you are interested in, in terms of sum of random variables.

   ⭐ Apply *continuity correction here* if RVs are discrete.

2. Apply CLT (e.g., approx X as $Y \sim N(n\mu, n\sigma^2)$ -> $\mathbb{P}(X \leq k) \approx \mathbb{P}(Y \leq k)$)
   Approximate sum of RVs as normal with appropriate mean and variance

from here, we're working with a normal distribution, which we've worked with before!

3. Compute probability approximation using Phi table

   > *Standardize* ($Z = \frac{N-\mu}{\sigma}$) -> $\mathbb{P}(Y \leq k) = \mathbb{P}\left(\frac{Y-\mu}{\sigma} \leq \frac{k-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{k-\mu}{\sigma}\right)$

   > *Write in terms of* $\Phi(z) = \mathbb{P}(Z \leq z)$

   > *Look up in table*

You and Bo are at an art exhibition with 100 people. Each person's will buy one of yours and one of Bo's. People buy Bo's art with an average of $50 and a variance of $10. People buy your art with an average of $50 and variance of $a (depending on will you are to bargain). What does the your variance per art piece need to be for probability of the difference in total earnings between you both being *more* than $50 to be less than 0.10? All art purchases are independent.

1. Setup the problem (*make sure to clearly define random variables, and write as a sum*)

$A_i \sim$ how i^{th} spend your, $B_i \sim i^{th}$ spend on Bo's art

$$A = \sum_{i=1}^{100} A_i, \quad B = \sum_{i=1}^{100} B_i$$

difference

$$D = A - B = \sum_{i=1}^{100} A_i - \sum_{i=1}^{100} B_i = \sum_{i=1}^{100} (A_i - B_i)$$

2. Apply CLT.

$$P(|D| > 50) < 0.10$$

3. Solve.

You and Bo are at an art exhibition with 100 people. Each person's will buy one of yours and one of Bo's. People buy Bo's art with an average of $\$50$ and a variance of $\$10$. People buy your art with an average of $\$50$ and variance of $\$a$ (depending on will you are to bargain). What does the your variance per art piece need to be for probability of the difference in total earnings between you both being *more* than \$50 to be less than 0.10? All art purchases are independent.

1. Setup the problem (*make sure to clearly define random variables, and write as a sum)*

A_i is how much the i'th person spends on your art, and B_i be how much they spend on Bo's art.
Your total earnings is $A = \sum_{i=1}^{100} A_i$ and Bo's total earnings is $B = \sum_{i=1}^{100} B_i$
The difference in earnings is $D = A - B = \sum_{i=1}^{100} A_i - \sum_{i=1}^{100} B_i = \sum_{i=1}^{100} (A_i - B_i) = \sum_{i=1}^{100} D_i$

Our goal is: $\mathbb{P}(|D| \geq 50) \leq 0.10$

2. Apply CLT.

$P(|D| > 50) \approx P(|Y| > 50)$

$Y \sim N(100\mu, 100 \cdot \sigma^2) \to D \approx Y$

$\mu = E[A_i - B_i] = E[A_i] - E[B_i] = 50 - 50 = 0$

3. Solve.

$\sigma^2 = Var(A_i - B_i) = Var(A_i) + Var(-B_i) = Var(A_i) + Var$...
$= a + 10$

You and Bo are at an art exhibition with 100 people. Each person's will buy one of yours and one of Bo's. People buy Bo's art with an average of $\$50$ and a variance of $\$10$. People buy your art with an average of $\$50$ and variance of $\$a$ (depending on will you are to bargain). What does the your variance per art piece need to be for probability of the difference in total earnings between you both being *more* than \$50 to be less than 0.10? All art purchases are independent.

1. Setup the problem (*make sure to clearly define random variables, and write as a sum)*

A_i is how much the i'th person spends on your art, and B_i be how much they spend on Bo's art.
Your total earnings is $A = \sum_{i=1}^{100} A_i$ and Bo's total earnings is $B = \sum_{i=1}^{100} B_i$
The difference in earnings is $D = A - B = \sum_{i=1}^{100} A_i - \sum_{i=1}^{100} B_i = \sum_{i=1}^{100} (A_i - B_i) = \sum_{i=1}^{100} D_i$

Our goal is: $\mathbb{P}(|D| \geq 50) \leq 0.10$

2. Apply CLT.

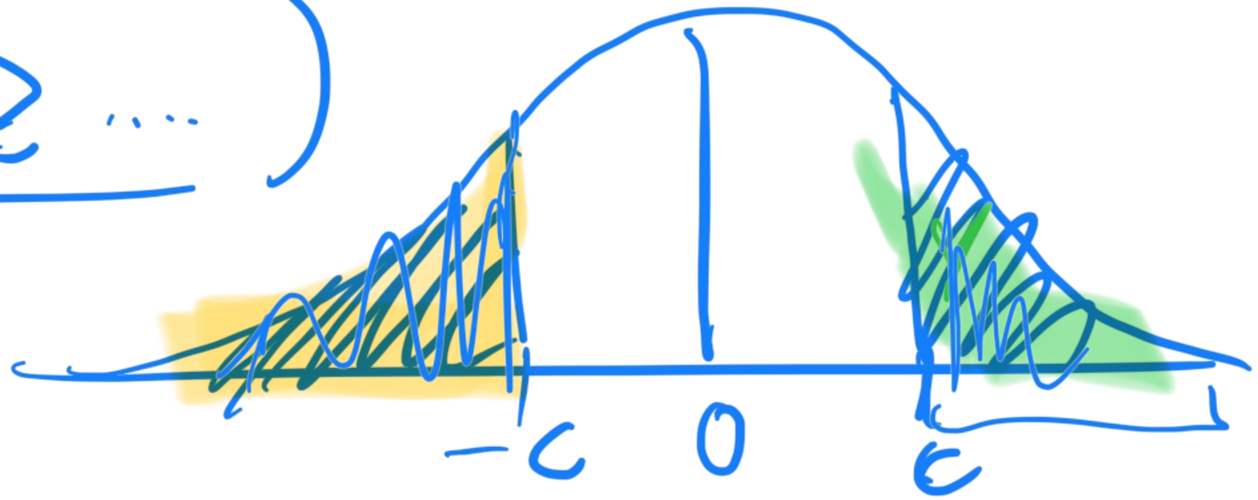
Summing together 100 of $(A_i - B_i)$ each with mean $\mu = \mathbb{E}[A_i - B_i] = \mathbb{E}[A_i] - \mathbb{E}[B_i] = 50 - 50$
and variance $\sigma^2 = \text{Var}(A_i - B_i) = \text{Var}(A_i) - \text{Var}(B_i) = a + 10$

So, we can approximate $D = \sum_{i=1}^{100} D_i$ as $Y \sim N(100 \cdot \mu, 100 \cdot \sigma^2)$, and $\mathbb{P}(|D| \geq 50) \approx \mathbb{P}(|Y| \geq 50)$

3. Solve.

$\mathbb{P}(|Y| \geq 50) = P\left(\left|\frac{Y - 100 \cdot \mu}{\sqrt{100 \cdot \sigma^2}}\right| \geq \frac{50 - 100 \cdot \mu}{\sqrt{100 \cdot \sigma^2}}\right) = P\left(|Z| \geq \frac{50 - 100 \cdot \mu}{\sqrt{100 \cdot \sigma^2}}\right)$

$$= P\left(Z \le -\left(\frac{50-100\cdot M}{\sqrt{\ldots}}\right)\right) + P(Z \ge \ldots)$$



$$= 2 \cdot P\left(Z \ge \frac{50-100M}{\sqrt{\ldots}}\right)$$

$$= 2 \cdot \left(1 - P(Z \le \ldots)\right)$$

$$= 2\left(1 - \Phi\left(\frac{500-100M}{\sqrt{\ldots}}\right)\right)$$

You and Bo are at an art exhibition with 100 people. Each person's will buy one of yours and one of Bo's. People buy Bo's art with an average of $\$50$ and a variance of $\$10$. People buy your art with an average of $\$50$ and variance of $\$a$ (depending on will you are to bargain). What does the your variance per art piece need to be for probability of the difference in total earnings between you both being *more* than \$50 to be less than 0.10? All art purchases are independent.

1. Setup the problem (*make sure to clearly define random variables, and write as a sum)*

A_i is how much the i'th person spends on your art, and B_i be how much they spend on Bo's art.
Your total earnings is $A = \sum_{i=1}^{100} A_i$ and Bo's total earnings is $B = \sum_{i=1}^{100} B_i$
The difference in earnings is $D = A - B = \sum_{i=1}^{100} A_i - \sum_{i=1}^{100} B_i = \sum_{i=1}^{100} (A_i - B_i) = \sum_{i=1}^{100} D_i$

Our goal is: $\mathbb{P}(|D| \geq 50) \leq 0.10$

2. Apply CLT.
Summing together 100 of $(A_i - B_i)$ each with mean $\mu = \mathbb{E}[A_i - B_i] = \mathbb{E}[A_i] - \mathbb{E}[B_i] = 50 - 50$ and variance $\sigma^2 = \mathrm{Var}(A_i - B_i) = \mathrm{Var}(A_i) - \mathrm{Var}(B_i) = a - 10$

So, we can approximate $D = \sum_{i=1}^{100} D_i$ as $Y \sim N(100 \cdot \mu, 100 \cdot \sigma^2)$, and $\mathbb{P}(|D| \geq 50) \approx \mathbb{P}(|Y| \geq 50)$

3. Solve.
$\mathbb{P}(|Y| \geq 50) = \mathbb{P}\left(|Z| \geq \frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right)$ *(standardize)* $= \mathbb{P}\left(Z \leq -\frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right) + \mathbb{P}\left(Z \geq \frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right)$
$= 2 \cdot \mathbb{P}\left(Z \geq \frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right) = 2 \cdot \left(1 - \Phi\left(\frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right)\right)$ *(write in terms of* $\boldsymbol{\Phi}$*)*

You and Bo are at an art exhibition with 100 people. Each person's will buy one of yours and one of Bo's. People buy Bo's art with an average of $\$50$ and a variance of $\$10$. People buy your art with an average of $\$50$ and variance of $\$a$ (depending on will you are to bargain). What does the your variance per art piece need to be for probability of the difference in total earnings between you both being *more* than \$50 to be less than 0.10? All art purchases are independent.

1. Setup the problem (*make sure to clearly define random variables, and write as a sum)*

....

Our goal is: $\mathbb{P}(|D| \geq 50) \leq 0.10$

2. Apply CLT.

$\mu = \mathbb{E}[A_i - B_i] = a - 50$ and variance $\sigma^2 = \mathrm{Var}(A_i - B_i) = 10 - 5 = 5.$

So, we can approximate $D = \sum_{i=1}^{100} D_i$ as $Y \sim N(100 \cdot \mu, 100 \cdot \sigma^2)$, and $\mathbb{P}(|D| \geq 50) \approx \mathbb{P}(|Y| \geq 50)$

3. Solve.

$$\mathbb{P}(|Y| \geq 50) = \mathbb{P}\left(|Z| \geq \frac{500 - 100 \cdot \mu}{\sqrt{100 \cdot \sigma^2}}\right) \text{ (standardize)}$$

$$= 2 \cdot \mathbb{P}\left(Z \geq \frac{50 - 100 \cdot \mu}{\sqrt{100 \cdot \sigma^2}}\right) = 2 \cdot \left(1 - \Phi\left(\frac{50 - 100 \cdot \mu}{\sqrt{100 \cdot \sigma^2}}\right)\right) \text{ (write in terms of } \boldsymbol{\Phi}\text{)}$$

You and Bo are at an art exhibition with 100 people. Each person's will buy one of yours and one of Bo's. People buy Bo's art with an average of $\$50$ and a variance of $\$10$. People buy your art with an average of $\$50$ and variance of $\$a$ (depending on will you are to bargain). What does the your variance per art piece need to be for probability of the difference in total earnings between you both being *more* than $50 to be less than 0.10? All art purchases are independent.

1. Setup the problem (*make sure to clearly define random variables, and write as a sum)*

Our goal is: $\mathbb{P}(|D| \geq 50) \leq 0.10$

2. Apply CLT.

$\mu = \mathbb{E}[A_i - B_i] = a - 50$ and variance $\sigma^2 = \text{Var}(A_i - B_i) = 10 - 5 = 5.$

So, we can approximate $D = \sum_{i=1}^{100} D_i$ as $Y \sim N(100 \cdot \mu, 100 \cdot \sigma^2)$, and $\mathbb{P}(|D| \geq 50) \approx \mathbb{P}(|Y| \geq 50)$

3. Solve.

$$\mathbb{P}(|Y| \geq 50) = \mathbb{P}\left(|Z| \geq \frac{500-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right) \text{ (standardize)}$$

$$= 2 \cdot \mathbb{P}\left(Z \geq \frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right) = 2 \cdot \left(1 - \Phi\left(\frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right)\right) \text{ (write in terms of } \boldsymbol{\Phi})$$

$$\leq \mathbf{0.1}$$

Solve for $\boldsymbol{\Phi}$: $2 \cdot \left(1 - \Phi\left(\frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right)\right) \leq 0.1$ -> $\Phi\left(\frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}}\right) \geq 0.95$

Reverse z-table lookup: $\frac{50-100\cdot\mu}{\sqrt{100\cdot\sigma^2}} = \frac{50-100\cdot(0)}{\sqrt{100\cdot(a+10)}} \geq 1.65$ *Algebra solving for* $\boldsymbol{a}$: $\boldsymbol{a \geq 19.2}$

Joint Distributions

We have two **discrete** random variables X and Y
(that may or may not be independent)

Joint **Support/Range** - $\Omega_{X,Y}$

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

Joint **PMF** - $p_{X,Y}(a,b)$

$p_{X,Y}(a,b) = \mathbb{P}(X \le a, Y \le b)$
*defined for **all*** $(a,b) \in \mathbb{R} \times \mathbb{R}$

Joint **CDF** - $F_{X,Y}(a,b)$

$F_{X,Y}(a,b) = \mathbb{P}(X \le a, Y \le b)$
*defined for **all*** $(a,b) \in \mathbb{R} \times \mathbb{R}$

Normalization Property:

$\sum_{(a,b)\in\Omega_{X,Y}} p_{X,Y}(a,b) = 1$

Joint **Independence**

> $p_{X,Y}(a,b) = p_X(a) \cdot p_X(b)$ for all $(a,b) \in \Omega_{X,Y}$
> $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Joint **Expectation**

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

Marginal **PMF** - $p_X(x)$, $p_Y(y)$

$p_X(x) = \sum_{y\in\Omega_Y} p_{X,Y}(x,y)$
$p_Y(y) = \sum_{x\in\Omega_X} p_{X,Y}(x,y)$

Notice we're summing over what the other RV can be

We have two **continuous** random variables X and Y
(that may or may not be independent)

Joint **Support/Range** - $\Omega_{X,Y}$

$\Omega_{X,Y} = \{(a,b) : f_{X,Y}(a,b) > 0\} \subseteq \Omega_X \times \Omega_Y$

Joint **PDF** - $f_{X,Y}(a,b)$

$f_{X,Y}(a,b)$ *defined for **all*** $(a,b) \in \mathbb{R} \times \mathbb{R}$

Normalization Property:

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx\, dy = 1$

Joint **CDF** - $F_{X,Y}(a,b)$

$F_{X,Y}(a,b) = \mathbb{P}(X \leq a, Y \leq b)$
*defined for **all*** $(a,b) \in \mathbb{R} \times \mathbb{R}$

Joint **Independence**

> $f_{X,Y}(a,b) = f_X(a) \cdot f_X(b)$ for all $(a,b) \in \Omega_{X,Y}$
> $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Joint **Expectation**

$\mathbb{E}[g(X,Y)] =$
$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y) dx\, dy$

Marginal **PDF** - $f_X(x)$, $f_Y(y)$

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

Notice we're integrating (summing) over what the other RV can be

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x, y) = \begin{cases} cy & x, y \in \mathbb{N}, 0 \leq x \leq y \leq 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion of the container's volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x, y) = \begin{cases} cy & \text{if } 0 \leq x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$

Normalization Property

probabilities must sum to 1 / the density function must integrate to 1

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \leq x \leq y \leq 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion of the container's volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} cy & \text{if } 0 \leq x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$

Normalization Property

probabilities must sum to 1 / the density function must integrate to 1

$\sum_{y=0}^{100} \sum_{x=0}^{y} cy = 1$

doing a bunch of algebra… $c = \frac{1}{343400}$

$\int_0^1 \int_0^y cy \, dx \, dy = 1$ or

$\int_0^1 \int_x^1 cy \, dy \, dx = 1$

evaluating the integral… $c = 3$

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion* of the container's *volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Finding probabilities (e.g., CDF)

sum/integrate over all the pairs of x and y in the desired region

What is $F_X(60, 50) =$ $P(X \le 60, Y \le 50)$

What is $F_X(0.5, 0.6) =$?

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion of the container's volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Finding probabilities (e.g., CDF)

sum/integrate over all the pairs of x and y in the desired region

What is $F_{X,Y}(60,50) = \mathbb{P}(X \le 60 \cap Y \le 50)$?

$\sum_{y=0}^{50} \sum_{x=0}^{60} cy$

What is $F_{X,Y}(0.5,0.6) = \mathbb{P}(X \le 0.5 \cap Y \le 0.6)$

$\int_0^{0.6} \int_0^{0.5} 3y \, dx \, dy = 0.27$

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion* of the container's *volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Finding the *marginal* PMF's (PDF) of X and Y

use law of total probability, partitioning on the values of the other random variable

What is $p_X(x)$ and $p_Y(y)$?

What is $f_X(x)$ and $f_Y(y)$?

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion of the container's volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Finding the *marginal* PMF's of X and Y

use law of total probability, partitioning on the values of the other random variable

What is $p_X(x)$ and $p_Y(y)$?

$p_X(x) = \sum_{y=x}^{100} cy$

$p_Y(y) = \sum_{x=0}^{y} cy$

What is $f_X(x)$ and $f_Y(y)$?

$f_X(x) = \int_x^1 3y \, dy = \frac{3^2}{2} - \frac{3x^2}{2}$

$f_Y(y) = \int_0^y 3y \, dx = 3y^2$

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion* of the container's *volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Finding the *joint expectation* $\mathbb{E}[XY^2]$

*go through ALL pairs (x,y) in the joint support, and sum/integrate over the joint pmf * the function*

$\mathbb{E}[XY^2] =$

$\mathbb{E}[XY^2] =$

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \leq x \leq y \leq 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion* of the container's *volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \leq x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$

Finding the *joint expectation* $\mathbb{E}[XY^2]$

*go through ALL pairs (x,y) in the joint support, and sum/integrate over the joint pmf * the function*

$\mathbb{E}[XY^2] = \sum_{y=0}^{100} \sum_{x=0}^{y} (xy^2 \cdot cy) = 1$

$$\mathbb{E}[XY^2] = \int_0^1 \int_0^y xy^2 \cdot cy \, dx \, dy$$
$$= \int_0^1 \int_x^1 xy^2 \cdot cy \, dy \, dx$$

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion* of the container's *volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Are X and Y independent?

*go through ALL pairs (x,y) in the joint support, and sum/integrate over the joint pmf * the function*

Two requirements:

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

> $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

Two requirements:

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

> $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

At a coffee shop, beans are stocked in a storage container each week.

Discrete

X is the *number of coffee beans* that is stocked at the beginning of the week.

Y is the *number of coffee beans* used to make coffee in the same week.

Joint PMF: $p_{X,Y}(x,y) = \begin{cases} cy & x,y \in \mathbb{N}, 0 \le x \le y \le 100 \\ 0 & \text{otherwise} \end{cases}$

Continuous

X is the *proportion* of the container's *volume* filled with coffee at the week's start

Y is the *proportion of the container's volume* used to make coffee in the week.

Joint PDF: $f_{X,Y}(x,y) = \begin{cases} 3y & \text{if } 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Are X and Y independent?

*go through ALL pairs (x,y) in the joint support, and sum/integrate over the joint pmf * the function*

Discrete

Two requirements:

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

> $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

Continuous

Two requirements:

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

> $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

Law of Total Expectation

Law of Total Expectation (LTE)

Let $\boldsymbol{A_1}, \boldsymbol{A_2}, \dots, \boldsymbol{A_k}$ be a partition of the sample space, then

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

Let $\boldsymbol{X}, \boldsymbol{Y}$ be discrete RVs, then,

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

$\boldsymbol{X}, \boldsymbol{Y}$ are continuous RVs, then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y)$$

Similar in form/idea to *law of total probability*, and the proof goes that way as well.

Reminder: conditional expectation

Everything looks the same, we're just adding on that event we're conditioning on:

$$\mathbb{E}[X|A] = \sum_{k \in \Omega} k \cdot \mathbb{P}(X = k|A)$$

$\mathbb{E}[X|Y = y] = \sum_{k \in \Omega_X} k \cdot \mathbb{P}(X = k|Y = y)$
or $\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} k \cdot f_{X|Y}(k, y)\, dk$ **if continuous**

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

or if continuous, $\mathbb{E}[X] = \int_{-\infty}^{\infty} k \cdot f_X(k)\, dx$

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

Example: Elevator Rides 🛗

The number of people who enter an elevator on the ground floor is $X \sim \mathrm{Poi}(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any of the N floors, independently of others. What is the **expected number of stops the elevator will make before discharging all the passengers**?

Y is the number of stops the elevator makes. What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

Example: Elevator Rides

The number of people who enter an elevator on the ground floor is $X \sim \mathrm{Poi}(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any of the N floors, independently of others. What is the **expected number of stops the elevator will make before discharging all the passengers**?

Y is the number of stops the elevator makes. What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\, e^{-10}\frac{10^i}{i!}$$

Fill out the poll everywhere:
pollev.com/cse312

Example: Elevator Rides 🛗

The number of people who enter an elevator on the ground floor is $X \sim \mathrm{Poi}(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any of the N floors, independently of others. What is the **expected number of stops the elevator will make before discharging all the passengers**?

Y is the number of stops the elevator makes. What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\, \mathbb{P}(X=k) = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\, e^{-10} \frac{10^i}{i!}$$

Example: Elevator Rides

The number of people who enter an elevator on the ground floor is $X \sim \text{Poi}(10)$. There are N floors above the ground floor, and each person is equally likely to get off at any of the N floors, independently of others. What is the **expected number of stops the elevator will make before discharging all the passengers**?

Y is the number of stops the elevator makes. What is $\mathbb{E}[Y]$?

Y depends on what the value of X is. So, use LTE, partitioning on X.

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty} \mathbb{E}[Y|X=k]\, e^{-10}\frac{10^i}{i!}$$

To find $\mathbb{E}[Y|X=k]$, we will use ***linearity of expectation***

Decompose: Let $Y_i = \begin{cases} 1 & \text{if stops on i'th floor} \\ 0 & \text{otherwise} \end{cases}$ --> $Y = \sum_i^N Y_i$

Apply LoE: $\mathbb{E}[Y|X=k] = \mathbb{E}\left[\sum_i^N Y_i \,\middle|X=k\right] = \sum_i^N \mathbb{E}[Y_i|X=k] = \sum_i^N \mathbb{P}(Y_i=1|X=k)$

Conquer: $\mathbb{P}(Y_i=1|X=k) = 1 - \mathbb{P}(Y_i=0|X=k) = 1 - \left(\frac{N-1}{N}\right)^{\mathrm{k}}$ --> $\mathbb{E}[Y|X=k] = N(1 - \left(\frac{N-1}{N}\right)^{\mathrm{k}})$