# CSE 312 Section 7

CLT, Joint Distributions
Law of Total Expectation

# Administrivia

# Announcements & Reminders

- Quiz 6
    - Tomorrow, 12pm
    - CLT & Joint Distributions

- HW6
    - Released last night
    - Due Wednesday 8/13 @ 11:59pm

# Review & Questions

# Any lingering questions from this last week?

Each week in section, we’ll be reviewing the main concepts from this week and putting them into action by  going through some practice problems together. But before we get into that review, we’ll try to start off each section with some time for you to ask questions. Was anything particularly confusing this week? Is there anything we can clarify before we dive into the review? This is your chance to clear things up!

# Problem 1 – Content Review Questions

# 1 – Content Review

a) $X \sim N(n\mu, n\sigma^2)$

# 1 – Content Review

a) $X \sim N(n\mu, n\sigma^2)$

b) $\sum_c c \cdot p_{C|D}(c|d)$

# 1 – Content Review

a) $X \sim N(n\mu, n\sigma^2)$

b) $\sum_c c \cdot p_{C|D}(c|d)$

c) $13$

# Problem 4 – Tweets

# 4 – Tweets

A prolific Twitter user tweets approximately 350 tweets per week. Let’s assume for simplicity that the tweets are independent and each consists of a uniformly random number of characters between 10 and 140. Thus, the CLT implies that the number of characters tweeted by this user is approximately normal. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).

Work on this problem with the people around you, and then we’ll go over it together!

# 4 – Tweets

- Let $X$ be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the $i$th tweet (since the start of the week).

# 4 – Tweets

- Let $X$ be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the $i$th tweet (since the start of the week).
  - $X$ is the sum of 350 IID RVs with $\mu = 75$ and $\sigma^2 = 1430$ (recall that $E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a+1)^2-1}{12}$ for the discrete uniform RV), so $X \approx N \sim Normal(350 \cdot 75, 350 \cdot 1430)$.

# 4 – Tweets

- Let $X$ be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the $i$th tweet (since the start of the week).
  - $X$ is the sum of 350 IID RVs with $\mu = 75$ and $\sigma^2 = 1430$ (recall that $E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a+1)^2-1}{12}$ for the discrete uniform RV), so $X \approx N \sim Normal(350 \cdot 75, 350 \cdot 1430)$.
- We find: $P(26{,}000 \leq X \leq 27{,}000) = P(25{,}999.5 \leq X \leq 27{,}000.5)$
  - Notice the continuity correction is an equality, *not* an approximation

# 4 – Tweets

- Let $X$ be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the $i$th tweet (since the start of the week).
  - $X$ is the sum of 350 IID RVs with $\mu = 75$ and $\sigma^2 = 1430$ (recall that $E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a+1)^2-1}{12}$ for the discrete uniform RV), so $X \approx N \sim Normal(350 \cdot 75, 350 \cdot 1430)$.
- We find: $P(26{,}000 \leq X \leq 27{,}000) = P(25{,}999.5 \leq X \leq 27{,}000.5)$
  - Apply CLT: $P(25{,}999.5 \leq X \leq 27{,}000.5) \approx P(25{,}999.5 \leq N \leq 27{,}000.5)$
  - Standardize: $P(25{,}999.5 \leq N \leq 27{,}000.5) = P(\frac{25999.5-350\cdot 75}{\sqrt{350\cdot 1430}} \leq \frac{N-350\cdot 75}{\sqrt{350\cdot 1430}} \leq \frac{27000.5-350\cdot 75}{\sqrt{350\cdot 1430}})$
    - Note this is an <u>equality</u>, *not* an approximation

# 4 – Tweets

- Let $X$ be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the $i$th tweet (since the start of the week).
  - $X$ is the sum of 350 IID RVs with $\mu = 75$ and $\sigma^2 = 1430$ (recall that $E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a+1)^2-1}{12}$ for the discrete uniform RV), so $X \approx N \sim Normal(350 \cdot 75, 350 \cdot 1430)$.
- We find: $P(26{,}000 \leq X \leq 27{,}000) = P(25{,}999.5 \leq X \leq 27{,}000.5)$
  - Apply CLT: $P(25{,}999.5 \leq X \leq 27{,}000.5) \approx P(25{,}999.5 \leq N \leq 27{,}000.5)$
  - Standardize: $P(25{,}999.5 \leq N \leq 27{,}000.5) = P(\frac{25999.5-350\cdot75}{\sqrt{350\cdot1430}} \leq \frac{N-350\cdot75}{\sqrt{350\cdot1430}} \leq \frac{27000.5-350\cdot75}{\sqrt{350\cdot1430}})$
  - $\approx P\left(-0.3541 \leq \frac{N-350\cdot75}{\sqrt{350\cdot1430}} \leq 1.0608\right)$ (approximation due to <u>rounding</u>)
  - This is $= \Phi(1.0608) - \Phi(-0.3541) \approx 0.4923$ (equals, then approximation due to rounding in $\Phi$ table)

# Problem 9 – Trapped Miner

# Problem 9 – Trapped Miner

A miner is trapped in a mine containing 3 doors.

- $D_1$: The 1st door leads to a tunnel that will take him to safety after 3 hours.
- $D_2$: The 2nd door leads to a tunnel that returns him to the mine after 5 hours.
- $D_3$: The 3rd door leads to a tunnel that returns him to the mine after a number of hours that is Binomial with parameters $(12, \frac{1}{3})$.

At all times, he is equally likely to choose any one of the doors. What is the expected number of hours for this miner to reach safety?

# Problem 9 – Trapped Miner

Let T = number of hours for the miner to reach safety. (T is a random variable)
Let $D_i$ be the event the $i^{th}$ door is chosen. $i \in \{1, 2, 3\}$. Finally, let $T_3$ be the time it takes to return to the mine in the third case only (a random variable). Note that the expectation of $T_3$ is $12 * \frac{1}{3}$ because it is binomially distributed with parameters $n = 12, p = \frac{1}{3}$. By Law of Total Expectation, linearity of expectation, and by applying the conditional expectations given by the problem statement:

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[T|D_1]\,\mathbb{P}(D_1) + \mathbb{E}[T|D_2]\,\mathbb{P}(D_2) + \mathbb{E}[T|D_3]\,\mathbb{P}(D_3) \\
&= 3 \cdot \frac{1}{3} + (5 + \mathbb{E}[T]) \cdot \frac{1}{3} + (\mathbb{E}[T_3 + T]) \cdot \frac{1}{3} \\
&= 3 \cdot \frac{1}{3} + (5 + \mathbb{E}[T]) \cdot \frac{1}{3} + (\mathbb{E}[T_3] + \mathbb{E}[T]) \cdot \frac{1}{3} \\
&= 3 \cdot \frac{1}{3} + (5 + \mathbb{E}[T]) \cdot \frac{1}{3} + (4 + \mathbb{E}[T]) \cdot \frac{1}{3}
\end{aligned}$$

Solving this equation for $\mathbb{E}[T]$, we get

$$\mathbb{E}[T] = 12$$

Therefore, the expected number of hours for this miner to reach safety is 12.

# Problem 10 – Lemonade Stand

# Problem 10 – Lemonade Stand

- Suppose I run a lemonade stand, which costs me $100 a day to operate. I sell a drink of lemonade for $20. Every person who walks by my stand either buys a drink or doesn't (no one buys more than one). If it is raining, $n_1$ people walk by my stand, and each buys a drink independently with probability $p_1$. If it isn't raining, $n_2$ people walk by my stand, and each buys a drink independently with probability $p_2$. It rains each day with probability $p_3$, independently of every other day. Let $X$ be my profit over the next week. In terms of $n_1, n_2, p_1, p_2,$ and $p_3$, what is $E[X]$?

Work on this with the people around you and then we'll go over it together!

# Problem 10 – Lemonade Stand

Let $R$ be the event it rains. Let $X_i$ be how many drinks I sell on day $i$ for $i = 1, ..., 7$. We are interested in $X = \sum_{i=1}^{7} (20X_i - 100)$. We have $X_i|R \sim$ Binomial$(n_1, p_1)$, so $\mathbb{E}[X_i|R] = n_1 p_1$. Similarly, $X_i|R^C \sim$ Binomial$(n_2, p_2)$, so $\mathbb{E}\left[X_i|R^C\right] = n_2 p_2$. By the law of total expectation,

$$\mu = \mathbb{E}[X_i] = \mathbb{E}[X_i|R]\,\mathbb{P}(R) + \mathbb{E}\left[X_i|R^C\right]\mathbb{P}(R^C) = n_1 p_1 p_3 + n_2 p_2 (1 - p_3)$$

# Problem 10 – Lemonade Stand

Let $R$ be the event it rains. Let $X_i$ be how many drinks I sell on day $i$ for $i = 1, ..., 7$. We are interested in $X = \sum_{i=1}^{7} (20X_i - 100)$. We have $X_i|R \sim$ Binomial$(n_1, p_1)$, so $\mathbb{E}[X_i|R] = n_1 p_1$. Similarly, $X_i|R^C \sim$ Binomial$(n_2, p_2)$, so $\mathbb{E}\left[X_i|R^C\right] = n_2 p_2$. By the law of total expectation,

$$\mu = \mathbb{E}[X_i] = \mathbb{E}[X_i|R]\,\mathbb{P}(R) + \mathbb{E}\left[X_i|R^C\right]\mathbb{P}(R^C) = n_1 p_1 p_3 + n_2 p_2 (1 - p_3)$$

Hence, by linearity of expectation,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{7} (20X_i - 100)\right] = 20 \sum_{i=1}^{7} \mathbb{E}[X_i] - 700 = 140\mu - 700$$

$$= 140 \cdot (n_1 p_1 p_3 + n_2 p_2 (1 - p_3)) - 700.$$

# Problem 2 – Round off error

## 2 – Round off error

Let $X$ be the sum of 100 real numbers, and let $Y$ be the same sum, but with each number rounded to the nearest integer before summing. If the roundoff errors are independent and uniformly distributed between -0.5 and 0.5, what is the approximate probability that $|X - Y| > 3$?

Work on this problem with the people around you, and then we'll go over it together!

## 2 – Round off error

- Notation: $X = \sum_{i=1}^{100} X_i$ and $Y = \sum_{i=1}^{100} r(X_i)$, where $r(\cdot)$ rounds to the nearest integer. Then

$$X - Y = \sum_{i=1}^{100} X_i - r(X_i)$$

## 2 – Round off error

- Notation: $X = \sum_{i=1}^{100} X_i$ and $Y = \sum_{i=1}^{100} r(X_i)$, where $r(\cdot)$ rounds to the nearest integer. Then

$$X - Y = \sum_{i=1}^{100} X_i - r(X_i)$$

- Each $X_i - r(X_i)$ is round off error which is distributed as $Uni(-0.5, 0.5)$. We know the expectation and variance of a continuous uniform distribution: $\mu = 0$ and $\sigma^2 = \frac{1}{12}$

## 2 – Round off error

- Notation: $X = \sum_{i=1}^{100} X_i$ and $Y = \sum_{i=1}^{100} r(X_i)$, where $r(\cdot)$ rounds to the nearest integer. Then
$$X - Y = \sum_{i=1}^{100} X_i - r(X_i)$$
- Each $X_i - r(X_i)$ is round off error which is distributed as $Uni(-0.5, 0.5)$. We know the expectation and variance of a continuous uniform distribution: $\mu = 0$ and $\sigma^2 = \frac{1}{12}$
- We then use the CLT: $X - Y \approx W \sim N\left(0, \frac{100}{12}\right)$
- Standardize:
$$P(|X - Y| > 3) \approx P(|W| > 3) = 2P(W > 3) \approx 2P(Z > 1.039)$$
$$\approx 0.29834$$

# That's All, Folks!

**Thanks for coming to section this week!**
**Any questions?**