50.034 - Introduction to Probability and Statistics

Week 6 – Lecture 11

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 11 Assigned Readings: 6.1, 6.2

Outline of Lecture

- Random samples and sample mean
- Markov's inequality
- Chebyshev's inequality
- Convergence in probability
- Law of large numbers
- Almost sure convergence

Sample Mean

Recall: (Lecture 8) If $X_1, \ldots, X_n$ are n **independent** R.V.'s, such that each X_i has the same distribution, then we say that $X_1, \ldots, X_n$ are independent and identically distributed or i.i.d., or iid.

A random sample is a collection $\{X_1, \ldots, X_n\}$ of **iid** R.V.'s, and the number n is called the sample size.

- It is common to use the same variable with positive integer indices to form a random sample. For example, $\{X_1, \ldots, X_n\}$ is a random sample, or $\{Y_1, \ldots, Y_m\}$ is a random sample.

Definition: Let $X_1, \ldots, X_n$ be R.V.'s. The sample mean of $X_1, \ldots, X_n$ is the R.V. $\frac{X_1 + \cdots + X_n}{n}$, i.e. the mean of $\{X_1, \ldots, X_n\}$.

- The sample mean is commonly denoted by $\overline{X}_n$.
- In probability and statistics, the overline (i.e. opposite of underline) is usually reserved to denote "sample mean".
- The subscript n in $\overline{X}_n$ indicates that the size of the given collection of R.V.'s is n.

Mean and Variance of Sample Mean

Note: The sample mean $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$ is a random variable!

- Hence, it makes sense to consider the "mean of sample mean" and the "variance of sample mean".

Let $X_1, \ldots, X_n$ be **iid** R.V.'s, each with mean μ and variance σ^2.

Fact: The mean of $\overline{X}_n$ is $\mathbf{E}[\overline{X}_n] = \mu$.

Proof: Using "mean of sum" = "sum of means", we get that $\mathbf{E}[X_1 + \cdots + X_n] = n\mu$, thus $\mathbf{E}[\overline{X}_n] = \mathbf{E}[\frac{X_1+\cdots+X_n}{n}] = \mu$.

Fact: The variance of $\overline{X}_n$ is $\mathrm{var}(\overline{X}_n) = \frac{\sigma^2}{n}$.

Proof: Since $X_1, \ldots, X_n$ are independent, we can use "variance of sum" = "sum of variances" to get that $\mathrm{var}(X_1 + \cdots + X_n) = n\sigma^2$, hence $\mathrm{var}(\frac{1}{n}(X_1 + \cdots + X_n)) = \frac{1}{n^2}\mathrm{var}(X_1 + \cdots + X_n) = \frac{\sigma^2}{n}$.

Example 1

Coin Toss Experiment: Toss a fair coin 10 times.

- Let $X_1, \ldots, X_{10}$ be R.V.'s, where $X_i = 1$ if the i-th toss yields heads, and $X_i = 0$ if the i-th toss yields tails.
- Every X_i is a Bernoulli R.V. with parameter 0.5. The 10 tosses are 10 independent Bernoulli trials. So $X_1, \ldots, X_{10}$ are iid.
- In other words, $\{X_1, \ldots, X_{10}\}$ is a **random sample** of size 10.

Let $\overline{X}_{10}$ be the sample mean of $X_1, \ldots, X_{10}$. Since the coin is fair, we would expect that $\approx \frac{1}{2}$ of the tosses are heads, i.e. $\overline{X}_{10} \approx 0.5$.

- Indeed, the expected value is $\mathbf{E}[\overline{X}_{10}] = 0.5$.
- However, getting 4 heads ($\overline{X}_{10} = 0.4$) or 6 heads ($\overline{X}_{10} = 0.6$) for the actual experiment is also not too surprising, since the proportion of heads obtained is still quite close to 0.5.

Does sample size matter?

Coin Toss Experiment: Toss a fair coin 100 times.

- Let $Y_1, \dots, Y_{100}$ be R.V.'s, where $Y_i = 1$ if the i-th toss yields heads, and $Y_i = 0$ if the i-th toss yields tails.
- $Y_1, \dots, Y_{100}$ are iid, so $\{Y_1, \dots, Y_{100}\}$ is a **random sample** with sample size 100.

Let $\overline{Y}_{100}$ be the sample mean of $Y_1, \dots, Y_{100}$.
Again, since the coin is fair, we would expect that $\overline{Y}_{100} \approx 0.5$.

- Indeed, the expected value is $\mathbf{E}[\overline{Y}_{100}] = 0.5$.
- Getting 49 heads ($\overline{Y}_{100} = 0.49$) or 51 heads ($\overline{Y}_{100} = 0.51$) is not too surprising, since the proportion of heads obtained is still quite close to 0.5.
- **Question:** Would you be surprised if you got 40 heads out of 100 tosses? Which do you think is more surprising: getting 4 heads out of 10 tosses, or getting 40 heads out of 100 tosses?

A closer look at the coin toss experiment

$X = X_1 + \cdots + X_{10}$ is a binomial R.V. (parameters 10 and 0.5).
$Y = Y_1 + \cdots + Y_{100}$ is a binomial R.V. (parameters 100 and 0.5).

- $\Pr(0.4 \leq \overline{X}_{10} \leq 0.6) = \sum_{k=4}^{6} \Pr(X = k) \approx 0.6563$.
- $\Pr(0.4 \leq \overline{Y}_{100} \leq 0.6) = \sum_{k=40}^{60} \Pr(Y = k) \approx 0.9648$.
- The greater the number of tosses, the higher the probability that the proportion of heads obtained is between 0.4 and 0.6.

What about getting heads for **exactly** half of the tosses?

- $\Pr(\overline{X}_{10} = 0.5) = \binom{10}{5}(0.5)^5(1-0.5)^5 \approx 0.2461$.
- $\Pr(\overline{Y}_{100} = 0.5) = \binom{100}{50}(0.5)^{50}(1-0.5)^{50} \approx 0.0796$.
- The greater the number of tosses, the lower the probability that the proportion of heads obtained is **exactly** 0.5.

Intuition: As the sample size n grows, the probability of getting heads for **exactly** half of n tosses gets smaller, but the probability of getting heads for **roughly** half of n tosses gets larger.

Markov's inequality

Theorem: (**Markov's inequality**) Let X be any R.V. satisfying $\Pr(X \geq 0) = 1$. If the expectation $\mathbf{E}[X]$ exists, then for every real number $t > 0$,

$$\Pr(X \geq t) \leq \frac{\mathbf{E}[X]}{t}.$$

Intuition: For a non-negative R.V. X, we would expect the event $\{X \geq t\}$ to have a small probability when t is very large. Markov's inequality gives a bound for how small this probability could be.

Interesting consequences of Markov's inequality:

- No more than $\frac{1}{5}$ of Singapore's population can have more than 5 times the average income in Singapore.
  - We assume that incomes are non-negative.
- In the canteen, no more than half of the people would spend more than twice the average queuing time to queue for food.

Example 2

Seismic data indicate that Indonesia suffers a major earthquake of magnitude ≥ 8.0 on average once every 10 years. What can we say about the probability that there will be another major earthquake in Indonesia of magnitude ≥ 8.0 in the next 30 years?

Solution: Let X be the waiting time (in years) until the next major earthquake in Indonesia of magnitude ≥ 8.0 occurs.

By Markov's inequality,

$$\Pr(X \geq 30) \leq \frac{\mathbf{E}[X]}{30} = \frac{1}{3}.$$

Thus, the probability that there **will be** another major earthquake in Indonesia of magnitude ≥ 8.0 in the next 30 years is

$$\Pr(X < 30) = 1 - \Pr(X \geq 30) \geq \frac{2}{3}.$$

Markov's inequality as a crude upper bound

Note: Markov's inequality holds for **any** R.V.!

- This inequality is very general and holds for all possible R.V.'s, so we should not expect it to give a good bound all the time.

Example: Let X be the number of heads obtained in 100 tosses of a fair coin. Then Markov's inequality yields:

$$\Pr(X \geq 99) \leq \tfrac{50}{99} \approx 0.5051.$$

We know that X is binomial with parameters 100 and 0.5, hence

$$\Pr(X \geq 99) = \binom{100}{99}(0.5)^{99}(0.5)^{1} + \binom{100}{100}(0.5)^{100} \approx 7.967 \times 10^{-27}.$$

- $\Pr(X \geq 99) \leq 0.5$ is quite different from $\Pr(X \geq 99) \approx 0$.
- **Note:** $\Pr(X \geq 99) \leq 0.5$ is still accurate! It's just that Markov's inequality is not too useful in this example.

In general, Markov's inequality is useful as a crude upper bound for $\Pr(X \geq t)$, especially if we only know the mean of X, or if t is very large compared to $\mathbf{E}[X]$.

Chebyshev's Inequality

Theorem: (**Chebyshev's inequality**) Let X be any R.V. with a finite mean. If its variance $\mathrm{var}(X)$ exists, then for every real $t > 0$,

$$\Pr(|X - \mathbf{E}[X]| \geq t) \leq \frac{\mathrm{var}(X)}{t^2}.$$

Proof: Consider the R.V. $Y = (X - \mathbf{E}[X])^2$.

Note that $\Pr(Y \geq 0) = 1$, and $\mathbf{E}[Y] = \mathrm{var}(X)$.

Thus, by Markov's inequality applied to Y,

$$\Pr(|X - \mathbf{E}[X]| \geq t) = \Pr(Y \geq t^2) \leq \frac{\mathrm{var}(X)}{t^2}.$$

- In other words, Chebyshev's inequality is a special case of Markov's inequality!

Similarly, Chebyshev's inequality is useful as a crude upper bound for $\Pr(|X - \mathbf{E}[X]| \geq t)$, especially if we only know the variance of X, or if t is very large compared to $\mathrm{var}(X)$.

Interpretation of Chebyshev's inequality

If X is a R.V. with finite mean and variance, then Chebyshev's inequality tells us that we should expect the deviation from the mean, i.e. $|X - \mathbf{E}[X]|$, to be not too large.

- An actual observed X-value that deviates a lot from the expected value of X has a small probability of happening.
- In fact, for fixed $\mathrm{var}(X)$, the probability $\Pr(|X - \mathbf{E}[X]| \geq t)$ approaches 0 as $t \to \infty$.

Very Important Consequence:
Suppose $\overline{X}_n$ is the sample mean of n iid R.V.'s $X_1, \ldots, X_n$, each with mean μ and variance σ^2.

- **Fact:** $\mathbf{E}[\overline{X}_n] = \mu$ and $\mathrm{var}(\overline{X}_n) = \frac{\sigma^2}{n}$.
- Chebyshev's inequality says that $\Pr(|\overline{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}$.
- In particular, for fixed $\varepsilon > 0$, as the sample size $n \to \infty$, the probability $\Pr(|\overline{X}_n - \mu| \geq \varepsilon)$ approaches 0, or equivalently, the probability $\Pr(|\overline{X}_n - \mu| < \varepsilon)$ approaches 1.

Convergence in Probability

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s.

Definition: Suppose r is a real number. We say that the sequence $X_1, X_2, X_3, \ldots$ converges in probability to r if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr(|X_n - r| < \varepsilon) = 1.$$

- More simply, we say that X_n converges in probability to r, and we write $X_n \xrightarrow{\text{p}} r$.
  - By writing X_n, we are treating the subscript n as a variable index, so that we can take the limit $n \to \infty$. This means we are implicitly assuming that X_n is the n-th R.V. of some infinite sequence of R.V.'s.
- "converges to r in probability" = "converges in probability to r".

Technicality: When we are given a finite sequence $X_1, \ldots, X_n$ of **iid** R.V.'s, it still makes sense to take the limit $n \to \infty$. Since every X_i has the exact same distribution, we could extend our sequence to an infinite sequence $X_1, X_2, \ldots$ of independent R.V.'s, with each R.V. having this same distribution.

Law of Large Numbers

Theorem: Let $\overline{X}_n$ be the sample mean of n iid R.V.'s $X_1, \ldots, X_n$, each with finite variance and with mean μ. Then $\overline{X}_n \xrightarrow{p} \mu$.

- This theorem is famously called the law of large numbers.
- As shown earlier, this theorem can be proven by applying **Chebyshev's inequality** to $\overline{X}_n$, then taking the limit $n \to \infty$.

Interpretation:

- Given any threshold $\varepsilon > 0$, no matter how small ε is, the probability that $|\overline{X}_n - \mu|$ does not exceed the given threshold ε would be high, as long as the sample size is sufficiently large.
- Hence, by taking a larger and larger random sample, the sample mean becomes closer and closer to the expected value.

Important Consequence: If the exact value of the mean μ is unknown, then we can take a large random sample, and use the sample mean as an estimate for μ.

Tossing many many coins

Coin Toss Experiment: Toss a coin 10000 times.

- Let $X_1, \ldots, X_{10000}$ be R.V.'s, where $X_i = 1$ if the i-th toss yields heads, and $X_i = 0$ if the i-th toss yields tails.
- Let $\overline{X}_{10000}$ be the sample mean of $X_1, \ldots, X_{10000}$.

Suppose among the 10000 tosses, you obtained 5501 heads.

- This means that $\overline{X}_{10000} = 0.5501$.
- Hence, 0.5501 is a good estimate for the mean $\mathbf{E}[\overline{X}_{10000}]$.

Question: Do you think the coin is fair?

- Notice that this question is not clear at all!
- Is it a "yes-or-no" question?
- If "yes", what exactly does it mean by being "fair" enough?
- How do you determine what is considered "yes", and what is considered "no"?

Making the coin toss question more precise

To determine whether the coin is fair, we need to state **thresholds** that we are willing to accept.

- We know that the sample mean should be close to the mean.
- If the coin is indeed fair, then the deviation of the sample mean $\overline{X}_{10000}$ from 0.5 should be small.
- What does it mean for the deviation to be small?
  - We need to decide on some threshold $\varepsilon > 0$, where any non-negative real number $< \varepsilon$ is considered "small", and any non-negative real number $\geq \varepsilon$ is considered "not small".
  - We want the probability of having a small deviation (i.e. $|\overline{X}_{10000} - 0.5| < \varepsilon$) to be large.
- What does it mean for this probability to be large?
  - Again, we need to decide on some threshold $0 \leq p_0 \leq 1$, where any probability $p > p_0$ is considered "large", and any probability $p \leq p_0$ is considered "not large".
- In other words, we need to decide on the thresholds ε and p_0, so as to make the question (fair coin or not) more precise.

Tossing many many coins (revisited)

Let's agree that any probability $p > 0.99$ is considered "large".

New Question:
What should ε be, so that $\Pr(|\overline{X}_{10000} - 0.5| < \varepsilon) \geq 0.99$?

- What should the threshold ε be, so that the probability of having a small deviation is large?
- What should ε be, so that $\Pr(|\overline{X}_{10000} - 0.5| \geq \varepsilon) \leq 0.01$?

Solution: If the coin is indeed fair (i.e. each X_i is Bernoulli with parameter $p = 0.5$), then by Chebyshev's inequality,

$$\Pr(|\overline{X}_{10000} - 0.5| \geq \varepsilon) \leq \frac{(0.5)(1-0.5)}{(10000)\varepsilon^2}.$$

(Recall: Chebyshev's inequality says that $\Pr(|\overline{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}$.)
Thus, solving for $\frac{(0.5)(1-0.5)}{(10000)\varepsilon^2} = 0.01$, we get $\varepsilon = 0.05$.

Tossing many many coins (revisited)

Therefore, Chebyshev's inequality gives:

$$\Pr(|\overline{X}_{10000} - 0.5| \geq 0.05) \leq 0.01.$$

- For the probability of having a small deviation to be large (i.e. ≥ 0.99), the threshold for the deviation should be $\varepsilon = 0.05$.

If the coin is fair, then based on our threshold that any $p \geq 0.99$ is a large probability, and based on our criterion that the probability of having a small deviation must be large, we deduced that the deviation $|\overline{X}_{10000} - 0.5|$ should not exceed $\varepsilon = 0.05$.

We are given that $|\overline{X}_{10000} - 0.5| = 0.0501 > 0.05$.

Therefore, we conclude that the coin is NOT fair (based on the threshold 0.99 that we chose).

Example 3

Let μ be the average age (in years) of all 5.6 million people in Singapore. We do not know what the exact value of μ is, and suppose we want to get an estimate of the value of μ.

We decide to conduct a poll on n randomly selected people to record their ages. Suppose it is known that the age of any randomly selected person in Singapore, treated as a R.V., has a standard deviation of 20 years.

Find a value of n for which the average age of these n polled people is strictly within 2 years of the actual average age (μ years), with probability at least 0.99.

Example 3 - Solution

Let X_i be the age (in years) of the i-th person polled.
We are given that $\{X_1, \ldots, X_n\}$ forms a random sample.
We are also given that each X_i has standard deviation $\sigma = 20$.

We want to determine the minimum possible value for n, so that the sample mean $\overline{X}_n$ of $X_1, \ldots, X_n$ satisfies

$$\Pr(|\overline{X}_n - \mu| < 2) \geq 0.99,$$

or equivalently, $\Pr(|\overline{X}_n - \mu| \geq 2) \leq 0.01$.

By Chebyshev's inequality, $\Pr(|\overline{X}_n - \mu| \geq 2) \leq \frac{\sigma^2}{n(2^2)} = \frac{100}{n}$.
Equating $\frac{100}{n} = 0.01$, we get $n = 10000$.

Therefore, using only the given information that X_i has a known standard deviation $\sigma = 20$, we conclude that polling $n = 10000$ people would **guarantee** that $\Pr(|\overline{X}_n - \mu| < 2) \geq 0.99$.

Sampling from the same distribution is important!

Let $\overline{X}_n$ be the sample mean of R.V.'s $X_1, \ldots, X_n$.

For us to be able to apply **Chebyshev's inequality** on $\overline{X}_n$, we must ensure that $X_1, \ldots, X_n$ are iid R.V.'s.

For real-world sampling (e.g. polls and surveys), we usually analyze a random sample, so as to deduce properties about the population.

- e.g. sample the ages of 10000 randomly selected people to get an estimate of the true average age of people in Singapore.

For us to accurately deduce something about the entire population, the distribution of each R.V. in the random sample must be the same as the distribution of the entire population.

- For example, if most of the people polled are from SUTD, then we are actually sampling from a different population, which has a different distribution.

Real-world Challenge: How to ensure that $X_1, \ldots, X_n$ are sampled from the same distribution as the population's?

- How can we eliminate sampling bias?

Convergence in Probability to a R.V.

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s.
Earlier today, we saw "convergence in probability" of $X_1, X_2, X_3, \ldots$ to a constant real number. Now, we shall extend this definition.

Definition: Let X be a random variable. We say that the sequence $X_1, X_2, X_3, \ldots$ converges in probability to X if for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr(|X_n - X| < \varepsilon) = 1.$$

- More simply, we say X_n converges in probability to X, and write $X_n \xrightarrow{p} X$.

Note: This definition extends our previous definition!

- A real number r can be treated as a discrete R.V. with exactly one possible value r.
  - In particular, $\mathbf{E}[r] = r$, and $\operatorname{var}(r) = 0$.

Note: In addition to "convergence in probability", there are also several other kinds of convergence.

- e.g. convergence in distribution, almost sure convergence, etc.

Convergence in probability under continuous functions

Theorem: Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s, let X be a R.V., and suppose that $X_n \xrightarrow{p} X$. If $h(t)$ is a real-valued **continuous** function on the reals, then $h(X_n) \xrightarrow{p} h(X)$.

- In other words, for every $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n - X| < \varepsilon) = 1$.

Corollary: If $X_n \xrightarrow{p} r$ for some $r \in \mathbb{R}$, then $h(X_n) \xrightarrow{p} h(r)$.

Corollary: Let $\overline{X}_n$ be the sample mean of n iid R.V.'s $X_1, \ldots, X_n$, each with finite variance and with mean μ. If $h(t)$ is a real-valued **continuous** function on the reals, then $h(\overline{X}_n) \xrightarrow{p} h(\mu)$.

Almost sure convergence

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s.

Definition: Suppose X is a random variable. We say that the sequence $X_1, X_2, X_3, \ldots$ converges almost surely to X if

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

- More simply, we say X_n converges almost surely to X.
- We usually write X_n converges a.s. to X, or write $X_n \xrightarrow{\text{a.s.}} X$.

More precisely, $\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1$ means that

$$\Pr(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

- $\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ is the event containing all outcomes ω such that $X_n(\omega)$ approaches $X(\omega)$ as $n \to \infty$.
- We are assuming that all R.V.'s here X and $X_1, X_2, X_3, \ldots$ have the same sample space Ω.

Remarks on almost sure convergence

"almost surely" here has a very precise mathematical meaning!

- "almost surely" = "with probability 1".
- **Example:** If you are asked to choose a random number in the range $[1, \infty)$, you would **almost surely** not choose 500.

Interpretation of almost sure convergence:

"X_n converges almost surely to X" means that if we consider all outcomes $\omega \in \Omega$ such that $\lim_{n\to\infty} X_n(\omega) \neq X(\omega)$, then the set of all these "exceptional" outcomes would form an event with probability zero, even if this set is non-empty. (Recall: An event is a subset of outcomes.)

Fact: $X_n \xrightarrow{\text{a.s.}} X$ implies $X_n \xrightarrow{\text{p}} X$, but not conversely.

Equivalent terminology: The following mean exactly the same.

- X_n converges almost surely to X.
- X_n converges almost everywhere to X.
- X_n converges with probability 1 to X.
- X_n converges strongly to X.

Law of Large Numbers: Different Versions

Recall: The **law of large numbers** (on slide 13) states that if $\overline{X}_n$ is the sample mean of n iid R.V.'s $X_1, \ldots, X_n$, each with finite variance and with mean μ, then $\overline{X}_n \xrightarrow{p} \mu$.

- This is sometimes called the **weak law of large numbers**.

There is a "stronger" version of this "law of large numbers", called the **strong law of large numbers**:

Theorem: (**Strong law of large numbers**) If $\overline{X}_n$ is the sample mean of n iid R.V.'s $X_1, \ldots, X_n$, each with finite variance and with mean μ, then $\overline{X}_n \xrightarrow{\text{a.s.}} \mu$.

Note: Almost sure convergence implies convergence in probability ($X_n \xrightarrow{\text{a.s.}} X$ implies $X_n \xrightarrow{p} X$), hence

$$\boxed{\begin{array}{c}\text{Strong law of}\\ \text{large numbers}\end{array}} \Longrightarrow \boxed{\begin{array}{c}\text{Weak law of}\\ \text{large numbers}\end{array}}$$

Announcement: Class Participation Assignment

As part of your final grade, 5% will go towards **class participation**.

How to get this 5% class participation?

- You would have to do the **class participation assignment**, which was posted this morning (4th Mar 2019).
- You will get a chance to do something fun!
- **Due Date:** Week 8 Cohort Class (21 March or 22 March).

The assignment is based on something you learned last week.

Let X, Y, Z be R.V.'s with finite means and finite variances.
Fact: Even if both (X, Y) and (Y, Z) are positively correlated, it is NOT NECESSARILY true that (X, Z) are positively correlated.

- We saw a "real-world" basketball example of this fact.

Announcement: Class Participation Assignment

Key Details of Class Participation Assignment

1. Form a **group** of 2–5 people.
2. As a group, write a hard-copy **report** describing <u>one</u> real-world example of a triple (X, Y, Z) of R.V.'s such that $\rho(X, Y) > 0$ and $\rho(Y, Z) > 0$, but $\rho(X, Z) < 0$.
   - You would have to precisely describe what each of X, Y, Z is.
   - Based on your selected X, Y, Z, you would have to justify why $\rho(X, Y)$ and $\rho(Y, Z)$ are both positive, but $\rho(X, Z)$ is negative.
3. As a group, give a **presentation** of your real-world example.
   - Your group presentation can be in **any** format.
   - e.g. powerpoint presentation, performance skit, etc.
   - All group members must participate!
   - Your audience: the rest of your cohort classmates.
   - Maximum time: **4 minutes** for your presentation.
4. Impress your classmates with your creativity!

See the class participation assignment posted on eDimension for more details.

Announcement: Class Participation Assignment

Grading for Class Participation Assignment

- In your group report, the names and student IDs of **all** group members must be clearly written.
- Every group member will get **the same score**.
  - Based on names of group members written in report.
  - If you do not submit a hard-copy report, your score is 0.
- If you submit your group report, then your score (max: 5%) depends **only** on class participation:

$$\text{Score } = \frac{k}{n} \times 5\%,$$

where

$$k = \text{number of group members participating in group presentation, or are absent with valid MC,}$$

$$n = \text{total number of group members.}$$

Example: If your group has 4 people, and on your presentation day, one of your group members is absent without a valid MC, then all 4 of you will each get 3.75% (out of the max 5%).

Summary

- Random samples and sample mean
- Markov's inequality
- Chebyshev's inequality
- Convergence in probability
- Law of large numbers
- Almost sure convergence