50.034 - Introduction to Probability and Statistics

Week 3 – Lecture 5

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 5 Assigned Readings: 3.3, 3.8, 4.1, 4.2, 4.3

Outline of Lecture

- Cumulative distribution function (cdf)
- Functions of a random variable
- Expectation of a random variable
- Properties of expectation
- Variance of a random variable

Cumulative distribution function (cdf)

Recall: The probability distribution of *any* random variable X is the collection of all probabilities of the form $\Pr(X \in C)$.

- Each $\{X \in C\}$ is an event, where C is a set of real numbers.
- Let x be any real number. The probability of $\{X \in C\}$ when C is the interval $(-\infty, x]$ is usually written as $\Pr(X \leq x)$.
- Define $\Pr(X < x)$, $\Pr(X \geq x)$ and $\Pr(X > x)$ analogously.

The cumulative distribution function (cdf) of a random variable X is the function

$$F(x) = \Pr(X \leq x), \qquad \text{for } -\infty < x < \infty.$$

(This definition is for *any* R.V., not just discrete or continuous R.V.'s.)

Interpretation:
$F(x)$ is the probability that the observed value of X is at most x.

Properties of the cdf of every R.V.

Let X be *any* (discrete, continuous, or mixed) R.V. with cdf $F(x)$.

- For any real number a,

$$\Pr(X > a) = 1 - F(a).$$

- The function $F(x)$ is **non-decreasing**, i.e.

$$\text{If } x_1 < x_2, \text{ then } F(x_1) \leq F(x_2).$$

- For any two real numbers a and b satisfying $a < b$,

$$\Pr(a < X \leq b) = F(b) - F(a).$$

- The limits of $F(x)$ at $\pm\infty$:

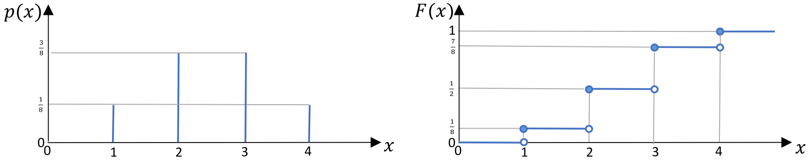
$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$

The cdf of a discrete R.V.

Let X be a discrete R.V. with pmf $p(x)$.
The cdf $F(x)$ of X is

$$F(x) = \Pr(X \le x) = \sum_{y : y \le x} p(y).$$

$F(x)$ is the probability that the observed value of X is at most x.
The graph of $F(x)$ is a **step function**:



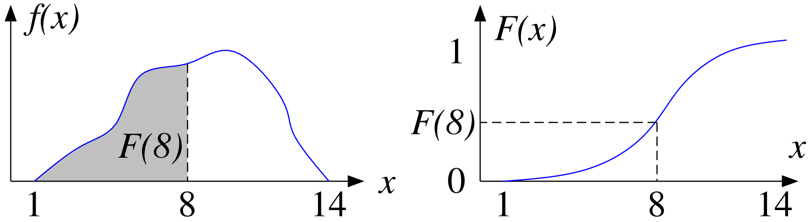
Fact: The cdf $F(x)$ of a **discrete** R.V. has "discrete jumps", so it is never a continuous function.

The cdf of a continuous R.V.

Let X be a continuous R.V. with pdf $f(x)$.
The cdf $F(x)$ of X is defined for every real number x:

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)\, du.$$

For each x, the value of $F(x)$ is the area under the density curve to the left of x.

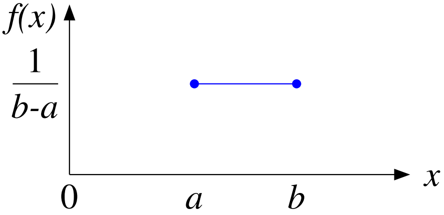


Fact: The cdf $F(x)$ of a **continuous** R.V. is always a **continuous** function, even if the pdf $f(x)$ is not continuous.

Example 1

Let X be a continuous R.V. that has a uniform distribution, such that its pdf is

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq x \leq b; \\ 0, & \text{otherwise.} \end{cases}$$

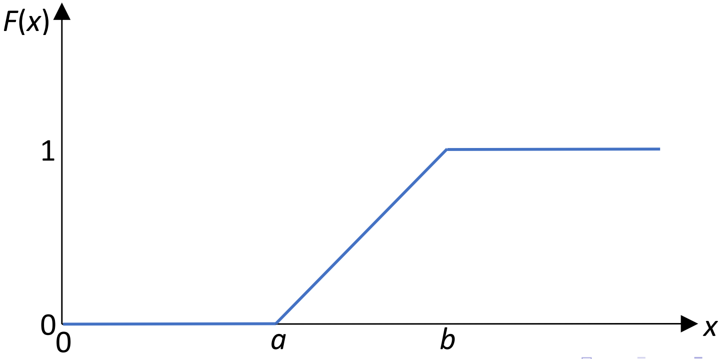


Find and plot the cdf of X.

Example 1

The cdf of X is

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \begin{cases} 0, & \text{if } x < a; \\ \dfrac{x-a}{b-a}, & \text{if } a \le x \le b; \\ 1, & \text{if } x > b. \end{cases}$$



Obtaining $f(x)$ from $F(x)$

Theorem: If X is a continuous R.V. with pdf $f(x)$ and cdf $F(x)$, and if the derivative $F'(x)$ exists at $x = x_0$, then $f(x_0) = F'(x_0)$.

- $F'(x)$ exists at $x = x_0$ whenever $f(x)$ is continuous at $x = x_0$; this is a consequence of the fundamental theorem of calculus.
- If $f(x)$ is a continuous function, then $F(x)$ is differentiable, and $f(x) = F'(x)$.

Example: Given the cdf $F(x)$ of X, find the pdf $f(x)$ of X.

$$F(x) = \begin{cases} 1 - e^{-x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

$$f(x) = F'(x) = \begin{cases} e^{-x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

Intuition for Expected Value

Experiment: Toss a fair coin 3 times.
Let X be the number of heads in the experiment. (discrete R.V.)
We have computed the pmf $p(x)$ of X in Example 1 of Lecture 4:

$$p(0) = \tfrac{1}{8},\ p(1) = \tfrac{3}{8},\ p(2) = \tfrac{3}{8},\ p(3) = \tfrac{1}{8}.$$

Question: If you repeat the experiment multiple times, then on average, what is the number of heads you should expect?

Answer: The expected number of heads is

$$\sum_{x=0}^{3} x \cdot p(x) = 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = 1.5$$

- More generally, for a discrete R.V. with finitely many possible values, we can always calculuate its expected value.
- What happens when there are infinitely many possible values?

Expected Value/Expectation of discrete R.V.

Let X be a **discrete** R.V. with possible values in D and pmf $p(x)$. Suppose we partition D into two subsets $D_{\geq 0}$ and $D_{<0}$:

- $D_{\geq 0}$ contains all the non-negative values in D;
- $D_{<0}$ contains all the strictly negative values in D.

The expectation or expected value or mean of X, denoted by $\mathbf{E}[X]$, is defined to be

$$\mathbf{E}[X] = \sum_{x \in D} x \cdot p(x),$$

provided that

$$\sum_{x \in D_{\geq 0}} x \cdot p(x) < \infty \quad \text{or} \quad \sum_{x \in D_{<0}} (-x) \cdot p(x) < \infty \quad \text{(or both)}.$$

(We use "expectation", "expected value" and "mean" interchangeably.)

- If both sums are infinite, then the expectation is undefined.
- **Why this technical condition?** Read Chap. 4.1 of textbook.

Other notation: $\mathbf{E}[X]$ is sometimes also denoted by μ_X or μ.

Example 2

Experiment: Count how many cars enter SUTD in the morning. Let X be the number of cars in the experiment. (discrete R.V.)

Given some parameter $\lambda > 0$, we could model X with the pmf:

$$p(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & \text{if } x = 0, 1, 2, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

- The resulting probability distribution is commonly called the Poisson distribution with parameter λ. (More in Lecture 6..)
- We say that X is a Poisson R.V. with parameter λ.

Question: What is the expected value of X?
(Hint: $\displaystyle\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = 1$.)

Example 2

We are given that

$$p(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & \text{if } x = 0, 1, 2, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

Solution:

$$\mathbf{E}[X] = \sum_{x=0}^{\infty} x \cdot p(x) = \sum_{x=0}^{\infty} \frac{x\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}$$

Using the hint $\displaystyle\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = 1$, we get

$$\mathbf{E}[X] = \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} = \lambda \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = \lambda.$$

More generally, the same calculation tells us that:

$$\mathbf{E}[\text{Poisson R.V. with parameter } \lambda] = \lambda.$$

Expectation of continuous R.V.

Let X be a **continuous** R.V. with pdf $f(x)$.
The expectation or expected value or mean of X, denoted by $\mathbf{E}[X]$, is defined to be

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx,$$

provided that

$$\int_{0}^{\infty} x \cdot f(x)\, dx < \infty \quad \text{or} \quad \int_{-\infty}^{0} (-x) \cdot f(x)\, dx < \infty \quad \text{(or both)}.$$

(We use "expectation", "expected value" and "mean" interchangeably.)

Remarks similar to the discrete R.V. case:

- If both integrals are infinite, then the expectation is undefined.
- Chap. 4.1 of textbook explains this technical condition.
- $\mathbf{E}[X]$ is sometimes also denoted by μ_X or μ.

Example 3

A laptop has a warranty of 1 year. Let X be a continuous R.V. that represents the time (in years) after which the laptop fails, and let $f(x)$ be the pdf of X, given by:

$$f(x) = \begin{cases} \frac{3}{2x^2\sqrt{x}}, & \text{if } x \geq 1; \\ 0, & \text{otherwise.} \end{cases}$$

What is the expectation of X?

Solution:

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{1}^{\infty} \frac{3}{2x\sqrt{x}}\,dx = \left[-\frac{3}{\sqrt{x}} \right]_{x=1}^{x=\infty} = 0 - (-3)$$
$$= 3.$$

Expectation of arbitrary R.V.

The expectation or expected value or mean of any R.V. X.

- Notation: $\mathbf{E}[X]$ or μ_X (or μ when context is clear).
- So far, we have seen the definitions of $\mathbf{E}[X]$ when X is either a **discrete** R.V. or **continuous** R.V.
- There is a more general definition of $\mathbf{E}[X]$ when X is *any* R.V.
  - This general definition is in terms of the "density" of cdf of X.
  - Recall: The cdf $F(x)$ of X is defined for *any* R.V.
  - There is also a technical condition for when $\mathbf{E}[X]$ exists.
  - Precise definition of $\mathbf{E}[X]$ (general case) is out of syllabus.

Functions of random variables

Given a R.V. X, one may be interested in another R.V. $Y = h(X)$, which is a function of X. Here are some examples:

- X = wind speed (in ms^{-1});

  Y = sensor output, e.g. $Y(\text{speed}) = \begin{cases} 0, & \text{if speed} \le 5; \\ 1, & \text{if } 5 < \text{speed} \le 20; \\ 2, & \text{if speed} > 20. \end{cases}$

  (Sensor output values $0, 1, 2$ represent "low", "medium", "high".)

  $$h(t) = \begin{cases} 0, & \text{if } t \le 5; \\ 1, & \text{if } 5 < t \le 20; \\ 2, & \text{if } t > 20. \end{cases}.$$

- X = game outcome ($X(\text{win}) = 1$, $X(\text{lose}) = -1$, $X(\text{draw}) = 0$);

  Y = payoff (in dollars) from betting on the game.

  e.g. $h(t)$ could be a function on domain $\{-1, 0, 1\}$, given by $h(1) = 95$, $h(-1) = -5$, $h(0) = -3$.

  (Betting ticket costs \$5; \$100 prize for win; \$2 prize for draw.)

Note: Any real-valued function of a R.V. is always a R.V.

Expectation of a function of a discrete R.V.

Let X be a **discrete** R.V. with possible values in D and pmf $p(x)$. Let $h : \mathbb{R} \to \mathbb{R}$ be any function.

- From the previous slide, we know that $h(X)$ is another R.V.
- We can compute $\mathbf{E}[X]$ (if it exists). What about $\mathbf{E}[h(X)]$?
- $\mathbf{E}[h(X)]$ (or $\mu_{h(X)}$) denotes the expectation of the R.V. $h(X)$.
- Recall: $\mathbf{E}[X] = \sum_{x \in D} x \cdot p(x)$.

Theorem: If $\mathbf{E}[h(X)]$ exists, then we can calculate it:

$$\mathbf{E}[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Note: For the special case when $h(t) = t$ is the identity function, we get the usual mean of X.

Expectation of a function of a continuous R.V.

Let X be a **continuous** R.V. with pdf $f(x)$.
Let $h : \mathbb{R} \to \mathbb{R}$ be any function.

- X is a R.V., so we can compute $\mathbf{E}[X]$ (if it exists).
- $h(X)$ is a R.V., so we can also compute $\mathbf{E}[h(X)]$ (if it exists).
- $\mathbf{E}[h(X)]$ (or $\mu_{h(X)}$) denotes the expectation of the R.V. $h(X)$.
- Recall: $\mathbf{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$.

Theorem: If $\mathbf{E}[h(X)]$ exists, then we can calculate it:

$$\mathbf{E}[h(X)] = \mathbf{E}[X] = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

Note: For the special case when $h(t) = t$ is the identity function, we get the usual mean of X.

Expectation of a linear function

Theorem: Let X be *any* R.V., and let a and b be finite constants. Then,

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$$

Proof: (Discrete R.V. case)
Let X be a discrete R.V. with possible values in D and pmf $p(x)$.

$$\begin{aligned}\mathbf{E}[aX + b] &= \sum_{x \in D} (ax + b)p(x) \\ &= a \sum_{x \in D} x \cdot p(x) + b \sum_{x \in D} p(x) \\ &= a\mathbf{E}[X] + b.\end{aligned}$$

The last equality holds because $\sum_{x \in D} p(x) = 1$ for any pmf $p(x)$.

Expectation of a linear function

Proof: (Continuous R.V. case)
Let X be a continuous R.V. with pdf $f(x)$.

$$\begin{aligned}
\mathbf{E}[aX+b] &= \int_{-\infty}^{\infty}(ax+b)f(x)\,dx \\
&= a\int_{-\infty}^{\infty} x\cdot f(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx \\
&= a\mathbf{E}[X]+b.
\end{aligned}$$

The last equality holds because $\int_{-\infty}^{\infty} f(x)\,dx = 1$ for any pdf $f(x)$.

- A similar argument holds for the general case, so the theorem is true for **any** R.V., not just discrete or continuous R.V.'s.
- Equivalent terminology: $\mu_{(aX+b)} = a\mu_X + b$.

Example 4

A 1-meter stick is randomly cut into two pieces. The length of the first part X is uniformly distributed on the interval $[0, 1]$, i.e.

$$f(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

Let $Y = \min\{X, 1 - X\}$ be the length of the shorter part, i.e.

$$Y = h(X) = \begin{cases} x, & \text{if } 0 \leq x \leq 0.5; \\ 1 - x, & \text{if } 0.5 < x \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

What is the mean of Y?

Example 4

Solution: The mean of Y is

$$\begin{aligned}
\mathbf{E}[Y] &= \mathbf{E}[h(X)] \\
&= \int_{-\infty}^{\infty} h(x)f(x)dx \\
&= \int_{0}^{0.5} x \cdot f(x)dx + \int_{0.5}^{1} (1-x)f(x)dx \\
&= \left[\frac{x^2}{2}\right]_{x=0}^{x=0.5} + 0.5 - \left[\frac{x^2}{2}\right]_{x=0.5}^{x=1} \\
&= 0.25
\end{aligned}$$

Expectation of sum or linear combination of R.V.'s

Theorem: Let $X_1, \ldots, X_n$ be n arbitrary R.V.'s, not necessarily independent, such that the means $\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n]$ are all finite. Then $X_1 + \ldots X_n$ is a R.V. with mean

$$\mathbf{E}[X_1 + \cdots + X_n] = \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n].$$

- Remember: "mean of sum" = "sum of means".

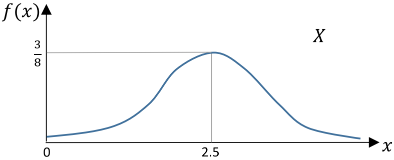
Corollary: Let $X_1, \ldots, X_n$ be n arbitrary R.V.'s, not necessarily independent, such that the means $\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n]$ are all finite. Let $a_1, \ldots, a_n, b$ be finite constants. Then $a_1 X_1 + \ldots a_n X_n + b$ is a R.V. with mean

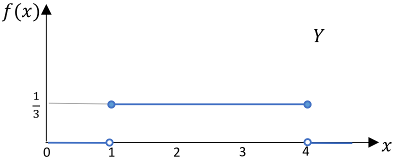
$$\mathbf{E}[a_1 X_1 + \cdots + a_n X_n + b] = a_1 \mathbf{E}[X_1] + \cdots + a_n \mathbf{E}[X_n] + b.$$

["Corollary" means a result that is easily proven using previous results.]

Intuition of Variance

Let X and Y be continuous R.V.'s with pdf's given by:





These two R.V.'s have very different probability distributions, yet both have the same mean $\mathbf{E}[X] = \mathbf{E}[Y] = 2.5$.

The **variance** of *any* R.V. (continuous or not) is a measure of how **spread out** its probability distribution is.

- The graph of the pdf of X is more spread out than the graph of the pdf of Y, so we should expect the variance of X to be larger than the variance of Y.

Variance of discrete R.V.

Let X be a **discrete** R.V. with possible values in D, pmf $p(x)$. Then the variance of X, denoted by $\mathrm{var}(X)$ or σ_X^2 (or simply σ^2), is

$$\mathrm{var}(X) = \sum_{x \in D} (x - \mu_X)^2 p(x) = E[(x - \mu_X)^2],$$

provided that $\mathbf{E}[(x - \mu_X)^2]$ exists.

- μ_X must exist and be finite, for $\mathrm{var}(X)$ to make sense.
- If $\mu_X = \pm\infty$ or does not exist, then $\mathrm{var}(X)$ does not exist.

The standard deviation of X, denoted by σ_X, is

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathrm{var}(X)}$$

- The standard deviation of X is the non-negative square root of the variance of X (if it exists).

Example 5

Let X be a Bernoulli R.V. with parameter p.

- Recall (from Lecture 4): This means that X takes on values 0 and 1 with probabilities $1 - p$ and p respectively.

Question: What are the mean $\mathbf{E}[X]$ and the variance $\mathrm{var}(X)$?

Solution: The mean of X is

$$\mathbf{E}[X] = 1 \times p + 0 \times (1 - p) = p$$

Let $p(x)$ be the pmf of X. Note: $p(0) = 1 - p$, $p(1) = p$. Thus,

$$\begin{aligned}\mathrm{var}(X) &= \sum_{x=0}^{1} (x - p)^2 p(x) \\ &= p^2(1 - p) + (1 - p)^2 p \\ &= p(1 - p)\end{aligned}$$

- If $p = 0$ or $p = 1$, then $\mathrm{var}(X)$ is 0.
- This makes sense because in either case, the pmf $p(x)$ is concentrated at a single point with zero variance.

Variance of continuous R.V.

Let X be a **continuous** R.V. with pdf $f(x)$.
Then the variance of X, denoted by $\mathrm{var}(X)$ or σ_X^2 (or simply σ^2), is

$$\mathrm{var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx = \mathbf{E}[(X - \mu_X)^2],$$

provided that $\mathbf{E}[(X - \mu_X)^2]$ exists.

- μ_X must exist and be finite, for $\mathrm{var}(X)$ to make sense.
- If $\mu_X = \pm\infty$ or does not exist, then $\mathrm{var}(X)$ does not exist.

The standard deviation of X, denoted by σ_X, is

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathrm{var}(X)}$$

- The standard deviation of X is the non-negative square root of the variance of X (if it exists).

Formula for variance of arbitrary R.V.

Let X be an **arbitrary** R.V. with finite mean μ_X.
Then the variance of X, denoted by $\mathrm{var}(X)$ or σ_X^2 (or simply σ^2), is

$$\mathrm{var}(X) = \mathbf{E}[(X - \mu_X)^2],$$

provided that $\mathbf{E}[(X - \mu_X)^2]$ exists.

Very useful formula: $\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$.

Proof: By definition,

$$\begin{aligned}\mathrm{var}(X) &= \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}\left[X^2 - 2\mathbf{E}[X]X + (\mathbf{E}[X])^2\right] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + (\mathbf{E}[X])^2 \\ &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2.\end{aligned}$$

Note: The formula holds for all (discrete, continuous, or mixed) R.V.'s X.

Example 6

Let X be a continuous R.V. with pdf $f(x)$ given by:

$$f(x) = \begin{cases} e^{-x}, & \text{if } x \geq 0; \\ 0, & \textit{otherwise}. \end{cases}$$

Find the mean $\mathbf{E}[X]$, and use the formula

$$\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

to calculate $\mathrm{var}(X)$.

Hint: Use integration by parts.

Example 6

It is time to review and practice integration.

$$\begin{aligned} \mathbf{E}[X] &= \int_0^\infty x e^{-x} dx \\ &= \Big[-xe^{-x} - e^{-x} \Big]_{x=0}^{x=\infty} = 1. \end{aligned}$$

$$\begin{aligned} \mathbf{E}[X^2] &= \int_0^\infty x^2 e^{-x} dx \\ &= \Big[-x^2 e^{-x} - 2xe^{-x} - 2e^{-x} \Big]_{x=0}^{x=\infty} = 2. \end{aligned}$$

Therefore, by the formula,

$$\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = 1.$$

Summary

- Cumulative distribution function (cdf)
- Functions of a random variable
- Expectation of a random variable
- Properties of expectation
- Variance of a random variable

Reminder:
There is **mini-quiz 1** (15mins) this week during Cohort Class.

- Tested on materials from Lecture 1 up to and including Slide 7 ("Mean and variance of binomial R.V.") of next lecture.