50.034 - Introduction to Probability and Statistics

Week 3 – Cohort Class

January–May Term, 2019

CC Week 3 Assigned Readings: (Revise: 3.3, 3.8, 4.1, 4.2, 4.3)

Outline of Cohort Class

- Recap: Cumulative distribution function (cdf)
- Recap: Expectation and variance
- **Mini-quiz 1**

Exercises on the following topics:

- Geometric distribution
- Poisson distribution
- Exponential distribution

Cumulative distribution functions

Recall: The cumulative distribution function (cdf) of *any* R.V. X is the function $F(x) = \Pr(X \leq x)$, for $-\infty < x < \infty$.

- $F(x)$ is the probability that the observed X-value is at most x.
- This "cdf" makes sense for any R.V.!

Special Cases:

- If X is a **discrete** R.V. with **pmf** $p(x)$, then its cdf is

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

- If X is a **continuous** R.V. with **pdf** $f(x)$, then its cdf is

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

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 $\pm\infty$:

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

Expectation of discrete R.V.

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

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

The expectation or expected value or mean of X is

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

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.

Question: What happens if exactly one of the sums is infinite?

Expectation of discrete R.V.

Case 1: $\sum_{x \in D_{\geq 0}} x \cdot p(x) = \infty$ and $\sum_{x \in D_{<0}} (-x) \cdot p(x) < \infty$.

By the definition of expectation,

$$\begin{aligned}
\mathbf{E}[X] = \sum_{x \in D} x \cdot p(x) &= \sum_{x \in D_{\geq 0}} x \cdot p(x) + \sum_{x \in D_{<0}} x \cdot p(x) \\
&= \infty + (\text{some finite negative number}) \\
&= \infty.
\end{aligned}$$

Case 2: $\sum_{x \in D_{\geq 0}} x \cdot p(x) < \infty$ and $\sum_{x \in D_{<0}} (-x) \cdot p(x) = \infty$.

By the definition of expectation,

$$\begin{aligned}
\mathbf{E}[X] = \sum_{x \in D} x \cdot p(x) &= \sum_{x \in D_{\geq 0}} x \cdot p(x) + \sum_{x \in D_{<0}} x \cdot p(x) \\
&= (\text{some finite positive number}) + (-\infty) \\
&= -\infty.
\end{aligned}$$

In other words, the definition allows for $\mathbf{E}[X]$ to be $\pm\infty$.

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 is

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

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.
- The definition allows for $\mathbf{E}[X]$ to be $\pm\infty$.

Expectation of a function of a R.V.

Let $h : \mathbb{R} \to \mathbb{R}$ be any function.

Theorem: Let X be a **discrete** R.V. with possible values in D and **pmf** $p(x)$. If $\mathbf{E}[h(X)]$ exists, then we can calculate it as follows:

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

Theorem: Let X be a **continuous** R.V. with **pdf** $f(x)$. If $\mathbf{E}[h(X)]$ exists, then we can calculate it as follows:

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

Variance of a R.V.

Definition: 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.

- 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.

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

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

Exercise 1 (20 mins)

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

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

where c is some unspecified constant.

1. Find the value of c.
2. Find the mean $\mathbf{E}[X]$.
3. Find the variance $\mathrm{var}(X)$.

Exercise 1 - Solution

1. Since $f(x)$ is a pdf,

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{1}^{\infty} \tfrac{c}{x^3}\,dx = \Big[-\tfrac{c}{2x^2} \Big]_{x=1}^{x=\infty} = \tfrac{c}{2},$$

which implies that $c = 2$.

2. $$\begin{aligned} \mathbf{E}[X] &= \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{1}^{\infty} \tfrac{2}{x^2}\,dx \\ &= \Big[-\tfrac{2}{x} \Big]_{x=1}^{x=\infty} \\ &= 2. \end{aligned}$$

3. First, we calculate $\mathbf{E}[X^2]$:

$$\begin{aligned} \mathbf{E}[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_{1}^{\infty} \tfrac{2}{x}\,dx = \Big[2\ln x \Big]_{x=1}^{x=\infty} \\ &= \infty. \end{aligned}$$

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

Mini-quiz 1 (15 mins)

Only writing materials are allowed. No calculators, notes, books, or cheat sheets are allowed. Don't worry, you won't need calculators.

If you are not present in class at the start of the quiz, you will not be given additional time to finish the quiz.

Remarks:

- There are no make-up mini-quizzes! If you arrive in class after the mini-quiz ends, or do not attend that cohort class, you will not have a chance to take the mini-quiz.
- To take into account unforeseen circumstances (e.g. mini-quiz missed due to illness), only the **best 3 of 4** mini-quiz scores will be counted towards your final grade.

Exercise 2 (15 mins)

Suppose NBA superstar Steph Curry is throwing basketballs towards the hoop from the three-pointer line, and he keeps on throwing until his first miss (i.e. until the first throw such that the basketball does not go into the hoop). His previous record at a contest was 13 consecutive successful throws before his first miss.

Every throw he makes is a Bernoulli process with success rate 0.438.

What is the probability that he repeats his record, i.e. his first miss is on his 14th throw?

Exercise 2 - Solution

Let X be the the number of successful throws, immediately before the first unsuccessful throw.

Then X can be modeled as a **geometric** R.V. with parameter $p = 1 - 0.438 = 0.562$.

In other words, X has the following pmf:

$$p(x) = 0.562 \cdot (0.438)^x.$$

Hence,

$$\begin{aligned}
&\Pr(\text{first unsuccessful throw on 14th try}) \\
&= \Pr(X = 13) = p(13) = 0.562 \cdot (0.438)^{13} \\
&\approx 0.00001227.
\end{aligned}$$

Exercise 3 (25 mins)

Suppose that internet users visit the SUTD main website following a Poisson distribution with a rate of 30 visitors per hour.

1. What is the probability that there will be exactly 15 visitors in a particular half-an-hour period?
2. One day, the SUTD website maintainer decides to see how fast it takes for her to wait before the first visitor arrives at the SUTD main website. What is the probability that she has to wait no more than 1 minute?
3. Given that the SUTD website maintainer did not observe any visitors in the first 10 minutes of her tracking, what is the probability that there will be no visitors in the next one minute?

Exercise 3 - Solution

1. Let X be the number of visitors per 30-minute period. Since the number of visitors per hour follows a Poisson distribution with parameter 30, we infer that X is a Poisson R.V. with parameter 15. The pmf of X is:

$$p(x) = \frac{15^x e^{-15}}{x!}.$$

Thus $\Pr(X = 15) = \frac{15^{15} e^{-15}}{15!} \approx 0.1024$.

2. Let Y be the waiting time in minutes until the first visitor arrives. Y can be modeled as an exponential R.V. with parameter $\lambda = \frac{30}{60} = \frac{1}{2}$ (i.e. the occurrence rate is $\frac{1}{2}$). The pdf of Y is

$$f(y) = \begin{cases} \frac{1}{2} e^{-y/2}, & \text{if } y \geq 0; \\ 0, & \text{if } y < 0; \end{cases}$$

therefore

$$\Pr(Y \leq 1) = \int_0^1 \frac{1}{2} e^{-y/2}\, dy = \left[-e^{-0.5y} \right]_{y=0}^{y=1} = 1 - e^{-0.5} \approx 0.3935.$$

Exercise 3 - Solution

3. As before, let Y be the waiting time in minutes until the first visitor arrives. From the previous part, Y is an exponential R.V. with parameter $\lambda = \frac{1}{2}$.

Also, we have computed that $\Pr(Y \leq 1) = 1 - e^{-0.5} \approx 0.3935$.

Hence, $\Pr(Y \geq 1) = 1 - \Pr(Y < 1) = 1 - \Pr(Y \leq 1) = e^{-0.5}$.

Since every exponential distribution has the **memoryless property**, it follows that

$$\Pr(Y \geq 31 | Y \geq 30) = \Pr(Y \geq 1) = e^{-0.5} \approx 0.6065,$$

i.e. the conditional probability there will be no visitors in the next one minute (after 30 minutes of no visitors) is ≈ 0.6065.

Summary

- Recap: Cumulative distribution function (cdf)
- Recap: Expectation and variance
- **Mini-quiz 1**

Exercises on the following topics:

- Geometric distribution
- Poisson distribution
- Exponential distribution

Reminder: Homework Set 3 is due next Cohort Class.