STAT-4300 - Class Notes

June 20, 2024

Chapter 1: Class 1

1.0.1 Set theory review

Definition: A set is an unordered collection of unique items

An element x in set A is denoted

$$x \in A$$

The **empty set** is the set without any elements, denoted

$$\emptyset$$

A is a **subset** of B if every element of A is also an element of B, denoted

$$A \subseteq$$

Note that

$$\emptyset \subseteq A \text{ for any } A$$

Definition: Operation of sets

The **union** of two sets is denoted

$$A \cup B, x \in A \cup B \iff x \in A \text{ or } x \in B$$

The **intersection** of two sets is denoted

$$A \cap B, x \in A \cap B \iff x \in A \text{ and } x \in B$$

The **complement** of a set is denoted

$$A^C, x \in A^C \iff x \notin A$$

Operations of multiple sets
Unions:

$$\begin{aligned}\bigcup_{n=1}^{\infty} A_n &= A_1 \cup A_2 \cup \dots \\ &= \{x : x \in A_n \text{ for some positive integer } n\}\end{aligned}$$

Intersections:

$$\begin{aligned}\bigcap_{n=1}^{\infty} A_n &= A_1 \cap A_2 \cap \dots \\ &= \{x : x \in A_n \text{ for all positive integers } n\}\end{aligned}$$

Disjoint: A and B are disjoint if

$$A \cap B = \emptyset$$

A, B, C are disjoint if

$$A \cap B \cap C = \emptyset$$

1.0.2 De Morgan's Laws

De Morgan's Laws
De Morgan's Laws is the analogue of the distributive property in the contexts of sets.

$$(A \cup B)^C = A^C \cap B^C$$
$$(A \cap B)^C = A^C \cup B^C$$

1.0.3 Sample spaces

Definition:
A sample space S is the set of all possible outcomes

An event A is a subset of S (it can include 1 or more possible outcomes)

1.0.4 Naive definition of probability

Definition:

$$\mathbb{P}_{naive}(A) = \frac{|A|}{|S|} = \frac{\text{number of events in } A}{\text{number of events in } S}$$

Naive because this assumes

- each outcome is equally likely
- finite number of possibilities

Chapter 2: Class 2

2.1 Counting

2.1.1 Multiplication rule

Consider compound experiment, consisting of 2 sub-experiments. Experiment 1 has n outcomes, experiment 2 as n outcomes. Then the compound experiment has mn outcomes

2.1.2 Sampling with / without replacement

Sampling k from n with replacement

$$n^k$$

Sampling k from n without replacement

$$\frac{n!}{(n-k)!}$$

2.1.3 Birthday Paradox

The probability of having a birthday match is

$$\begin{aligned} P(\text{birthday match}) &= \frac{\text{Number of outcomes that match}}{\text{Number of total outcomes}} \\ &= 1 - \frac{\text{Number of outcomes without match}}{\text{Number of total outcomes}} \\ &= 1 - \frac{365!}{(365-k)!365^k} \end{aligned}$$

2.1.4 Binomial coefficients

The number of ways to choose k out of n can be counted by

1. multiplication rule

$$n \times n-1 \times \ldots \times (n-k+1) = \frac{n!}{(n-k)!}$$

2. adjust for overcounting (account for $k!$ reorderings)

$$\frac{1}{k!}\frac{n!}{(n-k)!} = \frac{n!}{k!(n-k)!}$$

More generally,

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

2.1.5 Example with Poker

Find the probability of a royal flush

$$P(\text{ Royal flush }) = \frac{4}{\binom{52}{5}}$$

Find the probability of a flush

$$P(\text{ Flush }) = \frac{4\binom{13}{5}}{\binom{52}{5}} \approx 0.002$$

Find the probability of a full house

$$P(\text{ Full house }) = \binom{13}{2}\binom{2}{1}\binom{4}{3}\binom{4}{2} \div \binom{52}{5} \approx 0.001$$

2.2 A refined definition of probability

Previous definition:

$$\mathbb{P}_{\text{Naive}} \text{ is a function where} \begin{cases} \text{inputs} \\ \text{outputs} \end{cases}$$

Definition

A **probability** space consists of

- Sample space S
- Probability function P
  - Input: any event $A \in S$
  - Output: real number $P(A) \in [0, 1]$

Axiomatic definitions of probability function

- $P(\emptyset) = 0$
- $P(S) = 1$
- For $A_1, A_2, \ldots$ disjoint, then

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

Chapter 3: Class 3

3.1 Properties of probability

Complement rule:
$$P(A^C) = 1 - P(A)$$

Reasoning:

$$\begin{aligned}
&P(A) + P(A^C)\\
=&P(A \cup A^C) \text{ by axiom 2, sum of probability of disjoint events}\\
=&P(S) \text{ by definition of complement}\\
=&1 \text{by axiom 1}
\end{aligned}$$

Subset rule: If $A \in B$, then
$$P(A) \leq P(B)$$

Reasoning:

$$\begin{aligned}
B &= (B \cap A^C) \cup A\\
P(B) &= P((B \cap A^C) \cup A)\\
=&P(B \cap A^C) + P(A) \text{ by axiom 2}\\
\geq&P(A) \text{ since } P(B \cap A^C) \geq 0
\end{aligned}$$

Note that $A \subseteq B$ means that $A \implies B$

Inclusion-exclusion principle

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\begin{aligned}
P(A \cup B \cup C) = &P(A) + P(B) + P(C)\\
&- P(A \cap B) - P(B \cap C) - P(A \cap C)\\
&+ P(A \cap B \cap C)
\end{aligned}$$

3.2 Conditional probability

Definition: Conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- $A|B$ is not an event
- $P(A|B)$ is the probability that A occurs given that B occurs
- $A|B$ and $B|A$ both makes sense, regardless of chronology
- conditional probability concerns the information that one event provides for the other, not about causation

3.2.1 Probability of intersections

Theorem: If A, B have positive probability, then

$$P(A \cap B) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A)$$

Note the following shorthand in notation

- $P(A|B, C) = P(A|B \cap C)$
- $P(A, B) = P(A \cap B)$
- $P(A, B, C) = P(A \cap B \cap C)$

For 3 events, if $P(A, B) > 0$

$$P(A, B, C) = P(A) \cdot P(B|A) \cdot P(C|A, B)$$

In general,

$$P(A_1, \dots A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1, A_2) \dots P(A_n|A_1, \dots A_{n-1})$$

3.2.2 Bayes' Law and the law of total probability

Theorem: Bayes Theorem states that if $P(A), P(B) > 0$, then

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Theorem: Law of total probability states that for $A_1, \dots A_n$ that is a partition of sample space S

Where partition implies

- A_i are disjoint, and their union is S
- A_i are mutually exclusive and collectively exhaustive

Then, if $P(A_i) > 0$ for all i,

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Chapter 4: Class 4

4.1 Example 1 - Flipping Conditioning

1 fair and 1 biased coin (heads with probability 0.75). Pick one coin, flip 3 times, and observe HHH. What is the probability that the coin is fair?

Solution:
let F be the event that the fair coin was picked
let A be the event that the result is HHH

$$P(F|A) = \frac{P(A|F)P(F)}{P(A)}$$

4.2 Example 2 - Base Rate fallacy

Whartonitis problem.

1% of Wharton has Whartonitis. Test has 5% false positive rate, 5% false negative rate. If a test returns positive, what is the probability that the student has Whartonitis?

Let W be the event of Whartonitis
Let T be the event of testing positive.

$$\begin{aligned} P(W) &= 0.01 \\ P(T|W^C) &= 0.05 \\ P(T^C|W) &= 0.05 \end{aligned}$$

Then

$$\begin{aligned} P(W|T) &= \frac{P(T|W)P(W)}{P(T)} \\ &= \frac{P(T|W)P(W)}{P(T|W)P(W) + P(T|W^C)P(W^C)} \\ &= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} && = 0.16 \end{aligned}$$

4.3 Conditional probabilities are probabilities

Note that

- $A|B$ is not an event
- $P(\cdot)$ is not the same probability function as $P(\cdot|B)$
- In conditional probability, when $P(B) > 0$, we define a function $P(\cdot|B)$ that takes A as an input, and outputs $P(A|B)$

- Conditional probabilities functions are valid probability functions because they satisfies the axioms

Theorem: If $P(\cdot)$ satisfies the rules and $P(B) > 0$, then $P(\cdot|B)$ is a valid probability function, and all the rules of probability applies.

Complement Rule:

$$P(A|B) = 1 - P(A^C|B)$$

Inclusion-exclusion:

$$P\left((A_1 \cup A_2)\,|B\right) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$$

Bayes' Rule

$$P(A|B, C) = \frac{P(C|A, B)P(A|B)}{P(C|B)}$$

Law of total probability: If $A_1, \dots A_n$ partition S, then

$$P(C|B) = \sum_{i=1}^{n} P(C|A_i, B) \cdot P(A_i|B)$$

Chapter 5: Class 5

5.1 Independence

Definition: Two events are independent if

$$P(A \cap B) = P(A) \cdot P(B)$$

Equivalent definitions: if $P(A) > 0$ and $P(B) > 0$, then the following are equivalent

- A, B independent
- $P(A|B) = P(A)$
- $P(B|A) = P(B)$

Note: In the degenerate case, $P(A) = 0$, A is independent to any other event.
Note: Independence is completely different from disjointedness.

- if A, B disjoint, knowing A occurs gives **a lot** of information about B
- the only exception is in the edge case, where $P(A) = 0$ or $P(B) = 0$

Theorem: The following are equivalent

- A and B are independent
- A^C and B are independent
- A and B^C are independent
- A^C and B^C are independent

5.2 Independence of multiple events

Events A, B, C independent if

- $P(A \cap B) = P(A) \cdot P(B)$
- $P(B \cap C) = P(B) \cdot P(C)$
- $P(A \cap C) = P(A) \cdot P(C)$
- $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

Example: Recall the fair-unfair dice example.

let A be the event that first flip is heads
let B be the event that second flip is heads
let C be the event that the first and second flips are the same

Note that A, B, C are **pairwise independent**, but knowing 2 gives complete information about the second.

5.3 Independence of multiple events

Definition A, B, C are independent if

- $P(A \cap B) = P(A) \cdot P(B)$
- $P(B \cap C) = P(B) \cdot P(C)$
- $P(C \cap A) = P(C) \cdot P(A)$
- $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

A collection of sets $A_1, A_2, \ldots A_n$ independent if

- $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$
- $P(A_i \cap A_j \cap A_k) = P(A_i) \cdot P(A_j) \cap P(A_k)$

Note: additional reading, look for "information theory"

5.4 Conditional independence

Definition: Events A, B are conditionally independent given E if

$$P(A \cap B|E) = P(A|E) \cdot P(B|E)$$

Note

- Two events can be conditionally independent given E but not when given E^C
  - Recall good Wharton class example
- Conditional independence does not imply independence.
  - Recall the "FAIR / UNFAIR COIN" example, coin tosses only independent given fair or unfair coin
- Independence does not imply conditional independence.
  - Recall "FAIR / UNFAIR COIN" example, (A, B independent, but A, B are not conditionally independent when given C

Chapter 6: Class 6

6.1 Problem solving strategies using conditioning

6.1.1 Monty Hall

Main takeaway - Remaining door has gone through some selection process, hence it has a higher probability of having the car.

Let C_i be the event that the car is in door i, $i = 1, 2, 3$

$$= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$$

$= 2_{\overline{3}}$

$P(\text{Switching strategy gets car}) = \sum_{i=1}^{3} P(\text{switching strategy gets car — C}_i\text{switching strategy gets car — C}_i\text{switching strategy}$

6.1.2 Amoeba Problem

There is a single amoeba which in 1 minute, either stays the same, splits in 2, or dies with equal probability. Each amoeba is independent of all other amoebas. What is the probability it eventually dies out?

Let P be the event that the amoeba population eventually dies out.
Let A_0 be the event that a single amoeba dies in the next minute.
Let A_1 be the event that a single amoeba stays the same.
Let A_2 be the event that it splits in 2.

$$\begin{aligned} P(B) &= P(B|A_0)P(A_0) + P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \\ &= 1 \cdot \frac{1}{3} + P(B)\frac{1}{3} + (P(B))^2\frac{1}{3} \\ x &= \frac{1}{3} + \frac{x}{3} + \frac{x^2}{3} \\ x &= 1 \end{aligned}$$

Chapter 7: Class 7

7.1 Random Variables

Definition: Given an experiment with sample space S, a random variable is a function that maps each possible outcome $s \in S$ to a real number

7.2 Distribution

Definition: A r.v. X is discrete if there is a finite list of values $a_1, \dots a_n$ or an infinite list of values such that $a_1, a_2 \dots$ such that $P(X = a_j \text{ for some } j) = 1$

Definition: Support of X is the set of all values x such that $P(X = x) > 0$

Definition: The probability mass function of a discrete r.v. X specifies the probability of any particular value of X

$$p_X(x) = P(X = x)$$

For p_X to be a valid pmf,

- sum to 1 : $\sum_x p_X(x) = 1$
- non-negative : $p_X(x) \geq 0$ for any x

7.3 Bernoulli & Binomial Distributions

Definition A random variable X has the bernoulli distribution with parameter p if

- $P(X = 1) = p$
- $P(X = 0) = 1 - p$

X is distributed as Bernoulli with parameter p

$$X \sim Ber(p)$$

For any event A, we can define an indicator random variable I_A that equals 1 if A occurs, 0 otherwise

$$I_A \sim Ber(p), p = P(A)$$

Expectation

$$E[I_A] = p$$

Variance

$$Var(I_A) = p(1 - p)$$

Definition Suppose n independent Bernoulli trials are performed, each with the same success probability p, let X be the number of successes. The distribution of X is called the Binomial distribution with parameters n and p

$$X \sim Bin(n, p)$$

Theorem: If $X \sim Bin(n, p)$, then the PMF of X is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 0, 1, \dots n$$

Expectation

$$E[X] = np$$

Variance

$$Var(X) = np(1-p)$$

This is a valid PMF because

- non-negative

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \geq 0$$

- sums to 1

$$\begin{aligned}
\sum_{k=0}^{n} P(X = k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\
&= [p + (1-p)]^n \textbf{ By Binomial Theorem} \\
&= 1^n \\
&= 1
\end{aligned}$$

7.4 Cumulative Distribution Functions

Definition: The CDF of a r.v. X is

$$F_X(u) = P(X \leq u)$$

Any CDF satisfies the following properties

- increasing

$$u_1 \leq u_2 \implies F(u_1) \leq F(u_2)$$

- Right continuous
- Converges to 0 and 1 in the limits

$$\lim_{u \to -\infty} F(u) = 0$$

$$\lim_{u \to \infty} F(u) = 1$$

CDF can be calculated from the PMF

$$P(X \leq 2.5) = P(X = 0) + P(X = 1) + P(X = 2)$$

PMF can be calculated by taking the difference of the CDF¿

Chapter 8: Lesson 8

Placeholder

Chapter 9: Chapter 9

9.1 Hypergeometric Distribution

Drawing from a bag of w white balls and b black balls yields a Hypergeometric distribution with parameters w, b, n for the total number of white balls

$$HGeom(w, b, n)$$

Theorem: The support of HGeom distribution is all integer k such that

$$\begin{aligned} 0 \leq k \leq w \\ 0 \leq n - k \leq b \end{aligned}$$

Theorem: The PMF of $X \sim HGeom(w, b, n)$ is

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

The expected value is

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=i}^{n} \frac{w}{w+b} = \frac{nw}{w+b}$$

Note: let the support be K,

$$\begin{aligned} \sum_{k \in K} P(X = k) = 1 \\ \sum_{k \in K} \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = 1 \\ \sum_{k} \binom{w}{k}\binom{b}{n-k} = \binom{w+b}{n} \end{aligned}$$

Chapter 10: Lesson 10

10.1 Functions of random variables

A function of a r.v. is a r.v. Given sample space S, random variable X and function $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ is a r.v. that takes the value $g(X(s))$ for $s \in S$

10.2 Independence of r.v.s

Definition: X, Y independent if

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

$X_1, \dots X_n$ independent if

$$P(X_1 \leq x_1 \dots X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots P(X_n \leq x_n)$$

Theorem: If X, Y independent then any function of X is independent of any function of Y

$$X \perp Y \implies f(X) \perp g(Y) \text{ for any } f, g$$

Definition: $X_1, \dots X_n$ are i.i.d. if

- $X_1 \dots X_n$ independent
- $X_1 \dots X_n$ have the same distribution

Chapter 11: Class 12

11.1 Expectation

Definition: the expected value of a discrete r.v. X whose possible values are $x_1, x_2, \dots$ is

$$E[X] = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

11.1.1 Linearity of expectation

Theorem: For any X, Y and constant c,

$$E\left[X + Y\right] = E\left[X\right] + E\left[Y\right]$$

$$E\left[cX\right] = cE\left[X\right]$$

11.2 Geometric distribution

Definition: For a sequence of independent Bernoulli trials, each with the same success probability, let X denote the number of trials before the first success, X has the geometric distribution with parameter p

$$X \sim Geom(p)$$

$$X + 1 \sim FS(p)$$

The PMF is

$$P(X = k) = (1 - p)^k \cdot p$$

The expectation is

$$E[X] = \frac{1 - p}{p}$$

Chapter 12: Class 12

12.1 Indicator R.V.s

Definition: The indicator r.v. for an event A, I_A takes value 1 if A occurs and 0 if A does not occur.

$$I_A \sim Ber(p), p = P(A)$$

Fundamental bridge between probability and expectation

$$E[I_A] = p = P(A)$$

Properties

- $I_{A^C} = 1 - I_A$
- $I_{A \cap B} = I_A \cdot I_B$
- $I_{A \cup B} = I_A + I_B - I_A I_B$

12.2 Expectation via Survival Function / Darth Vader rule

Theorem: Let X be a r.v. with support $0, 1, 2 \dots$

$$E[X] = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n)$$

Special case when X takes as its support $[0, 1, \dots N]$

$$E[X] = \sum_{N=0}^{N-1} P(X > n) = \sum_{n=1}^{N} P(X \geq n)$$

12.3 Law of the Unconscious Statistician (LOTUS)

Theorem:

$$E[g(X)] = \sum_{x} g(x) P(X = x)$$

12.4 Variance

Definition: The Variance of a r.v. X is

$$Var(X) = E\left[(X - E[X])^2\right] = E\left[X^{2\text{¿}}\right] - E[X]^2$$

The standard deviation of X is

$$sd(X) = \sqrt{Var(X)}$$

Properties

- $Var(X) \geq 0$
- $Var(X + c) = Var(X)$
- $Var(cX) = c^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y)$ if $X \perp Y$

Chapter 13: Poisson Distribution

13.1 Poisson Distribution

Poisson Distribution is a common model for count data with support over all the non-negative integers. It is useful for modelling the number of successes in a sequence of Bernoulli trials where

- number of trials is large
- success probability is low

The Poisson distribution provides a simple and good enough approximation to a Binomial distribution when p is small and n is large.

Definition: A r.v. X has the Poisson distribution with parameter $\lambda > 0$ if its PMF is

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \implies X \sim Pois(X)$$

Expectation

$$E[X] = \lambda$$

Approach 1: the mean of $Y \sim Bin(n, p)$ is np, which stays constant for a Poisson distribution.

Approach 2:

$$\begin{aligned}
E[X] &= \sum_{k=1}^{\infty} kP(X = k) \\
&= \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \cdot \lambda \cdot e^{\lambda}
\end{aligned}$$

This is a valid PMF because
1) PMF is non-negative for all values in the support 2) PMF sums to 1 over the entire support

$$\begin{aligned}
\sum_{k=0}^{\infty} e^{\lambda}\frac{\lambda^k}{k!} &= e^{\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}e^{\lambda} \\
&= 1
\end{aligned}$$

13.2 Continuous RVs

Definition: X is a continuous r.v. if it has a continuous distribution. The distribution is ontinuous if its CDF $F(u) = P(X \leq u)$ is

- differentiable everywhere
- continuous everywhere, and differentiable at all but a finite number of points

Definition: The probability density function of X with CDF F is

$$f(u) = F'(u)$$

The support is the set of all u such that $f(u) > 0$
The CDF can be found via

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Chapter 14: Class 14

14.1 Continuous Distributions

From CDF to PDF:

$$f'(x) = F'(x)$$

From PDF to CDF

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

14.2 Rayleigh Distribution

Definition: The Rayleigh distribution has CDF

$$F(x) = \begin{cases} 1 - e^{-\frac{x^2}{2}} \text{ for } x \geq 0 \\ 0 \text{ for } x \leq 0 \end{cases}$$

The PDF is

$$f(x) = \begin{cases} xe^{-\frac{x^2}{2}} \text{ for } x \geq 0 \\ 0 \text{ for } x \leq 0 \end{cases}$$

14.3 Uniform distribution

Definition U has the uniform distribution on the interval (a, b)

$$U \sim Unif(a, b)$$

If its PDF is

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ for } a < x < b \\ 0 \text{ otherwise} \end{cases}$$

and CDF is

$$F(x) = \begin{cases} 0 \text{ if } x \leq a \\ \frac{x-a}{b-a} \text{ if } a < x < b \\ 1 \text{ if } x \geq b \end{cases}$$

Expectation

$$E[U] = \int_{\infty}^{\infty} xf(x)dx = \frac{1}{b-a}\int_{a}^{b} xdx = \frac{b+a}{2}$$

Variance

$$E\left[U^2\right] = \int_{\infty}^{\infty} x^2 f(x)dx = \frac{b^2+ab+a^2}{3}$$

Key property: subintervals within (a, b) have probability proportional to their length.

For $U \sim Unif(0, 1)$

$$P(u \in [0, \frac{1}{4}]) = \frac{1}{4}$$

14.3.1 Location-scale transform of uniform r.v.s

Given uniform *r.v.* X with interval (a, b)

$$X \sim Unif(a, b)$$

The location transform of X has distribution

$$cX + d \sim Unif(ca + d, cb + d)$$

14.3.2 Universality of uniform distribution

Theorem: let F be a continuous CDF taht is strictly increasing over the support of the distribution

- $X = F^{-1}(U)$ is a r.v. with CDF F
- $F(X)$ has the $Unif(0, 1)$ distribution

Chapter 15: Class 15

15.1 Gaussian distribution

15.1.1 Standard Gaussian distribution

Definition: The Gaussian distribution with mean 0 and variance 1, denoted

$$Z \sim N(0,1)$$

has PDF

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{ for } \infty < z < \infty$$

and CDF:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Expectation

$$E\left[Z\right] = 0$$

Variance

$$Var(Z) = 1$$

Properties

- Symmetrical PDF

$$\varphi(z) = \varphi(-z)$$

- Symmetric tail probabilities

$$\Phi(z) = 1 - Phi(-z)$$

- Symmetric r.v.

$$Z \sim N(0,1) \implies -Z \sim N(0,1)$$

15.1.2 General Gaussian distribution

Definition: if $Z \sim N(0,1)$, then $X = \mu + \sigma Z$ has the Gaussian distribution with mean μ and variance σ^2,

$$\mu + \sigma Z \sim N(\mu, \sigma^2)$$

Standardization: the standardized version, or Z score, of $X = \mu + \sigma Z$ is

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

The PDF is

$$f(x) = \varphi(\frac{x-\mu}{\sigma})\frac{1}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The CDF is

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

15.1.3 Empirical rule

For $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

- $P(-1 < Z < 1) \approx 0.68$
- $P(-2 < Z < 2) \approx 0.95$
- $P(-3 < Z < 3) \approx 0.997$

15.2 Exponential Distribution

Definition: Exponential distribution with parameter λ

$$X \sim Expo(\lambda)$$

has PDF

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0$$

and CDF

$$F(x) = \int_0^x f(t)dt$$

Scaling transform: if $X \sim Expo(1)$ and $Y = \frac{X}{\lambda}$ then

$$Y \sim Expo(\lambda)$$

Expectation and variance of standard exponential

$$E[X] = 1$$

$$Var(X) = 1$$

Expectation and variance of general exponential

$$E[Y] = E\left[\frac{X}{\lambda}\right] = \frac{1}{\lambda}$$

$$Var(Y) = Var\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda^2}Var(X) = \frac{1}{\lambda^2}$$

Memorylessness

$$P(X \geq s+t | X \geq s) = P(X \geq t) \text{ for all } s, t \geq 0$$

Chapter 16: Class 16

Midterm 2 review

Chapter 17: Class 17

17.1 Moments

Definition: For any r.v. X and $n = 1, 2, 3 \ldots$

n-th moment:

$$E[X^n]$$

n-th central moment:

$$E[(X-\mu)^n]$$

n-th standardized moment

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^n\right]$$

17.2 Moment Generating Functions

Definition: The MGF of X is

$$M_X(t) = E\left[e^{tX}\right]$$

17.2.1 MGF of Bernoulli

The MGF of $X \sim Ber(p)$, for all $t \in \mathbb{R}$

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] \\ &= \sum_x e^{tX} P(X = x) \\ &= e^{t\cdot 1} p + e^{t\cdot 0}(1-p) \\ &= pe^t + (1-p) \end{aligned}$$

17.2.2 MGF of Uniform

The MGF of $U \sim Unif(a, b)$, for all $t \in \mathbb{R}$

$$
\begin{aligned}
M_U(t) &= E\left[^{tU}\right] \\
&= \int_{\infty}^{\infty} e^{tU} f(u)du \\
&= \frac{1}{b-a} \int_a^b e^{tu} du \\
&= \frac{1}{b-a} \left(\frac{e^{tu}}{t}\right)\Bigg|_a^b \\
&= \frac{e^{tb} - e^{ta}}{t(b-a)}
\end{aligned}
$$

$$
M_U(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} \text{ if } t \neq 0 \\ 1 \text{ if } t = 0 \end{cases}
$$

17.2.3 MGF of Exponential

For $X \sim Expo(1)$ for $t < 1$

$$
\begin{aligned}
M_X(t) &= E\left[e^{tX}\right] \\
&= \int_{-\infty}^{\infty} e^{tx} f(x)dx \\
&= \int_0^{\infty} e^{tx} e^{-x} dx \\
&= \int_0^{\infty} e^{(t-1)x} dx \\
&= \frac{e^{(t-1)x}}{t-1}\Bigg|_0^{\infty} \\
&= \int_{e^{(t-1)\infty}-1}^{t-1} \\
&= \begin{cases} \infty \text{ if } t \geq 1 \\ \frac{1}{1-t} \text{ if } t < 1 \end{cases}
\end{aligned}
$$

17.2.4 MGF of standard gaussian

For $Z \sim N(0,1)$ for all $t \in \mathbb{R}$

$$
M_Z(t) = E\left[e^{tZ}\right] = e^{\frac{t^2}{2}}
$$

For general Gaussian, $X \sim N(\mu, \sigma^2), X = \mu + \sigma Z$

$$
\begin{aligned}
M_X(t) &= M_{\mu+\sigma Z}(t) \\
&= e^{\mu t} M_Z(\sigma t) \\
&= e^{\mu t} e^{\frac{(\sigma t^2)}{2}} \\
&= e^{\mu t + \frac{\sigma^2 t^2}{2}}
\end{aligned}
$$

For X_1, X_2 independent where

- $X_1 \sim N(\mu_1, \sigma_1^2$
- $X_2 \sim N(\mu_2, \sigma_2^2$

Their sum has MGF

$$\begin{aligned} M_{X_1+X_2}(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \\ &= e^{(\mu_1+\mu_2)t + \frac{\left(\sigma_1^2+\sigma_2^2\right)t^2}{2}} \end{aligned}$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

17.2.5 Properties of MGF

Derivatives of MGF gives the moments

Theorem: IF the MGF exists

$$E\left[X^n\right] = M^{(n)}(0)$$

MGF determines the distribution.

Theorem: If X, Y have the same MGF then they have the same distribution.

MGF of sum of independent R.Vs

Theorem: If X, Y are independent and their MGF exist, then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Location-scale transform of MGFs

If X has MGF $M_X(t)$, then a + bX has MGF

$$\begin{aligned} M_{a+bX}(t) &= E\left[e^{t(a+bX)}\right] \\ &= E\left[e^{ta}e^{tbX}\right] \\ &= e^{ta}E\left[e^{tbX}\right] \\ &= e^{ta}M_X(tb) \end{aligned}$$

Chapter 18: Class 18

18.1 Discrete Joint Distributions

Definition: The joint PMF of X, Y is the function $P_{X,Y}$ given by

$$P_{X,Y}(x, y) = P(X = x, Y = y)$$

Definition: The marginal PMF of X is the function p_X given by

$$P_X(x) = P(X = x) = \sum_{y \in Val(Y)} P(X = x, Y = y)$$

Where $val(Y)$ denotes the support of Y

Definition: The conditional PMF of Y given $X = x$ is the function $P_{Y|X}(y|x)$

$$P_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

18.2 Continuous Joint Distributions

Definition: The joint CDF of X, Y is the function $F_{X,Y}$ given by

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

Definition: The joint PDF of X, Y is the function $f_{X,Y}$ given by

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

Definition: The marginal PDF of X is the function f_X given by

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

Definition: The conditional PDF of Y given $X = x$ is the function $f_{Y|X=x}$ given by

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Chapter 19: Class 19

19.1 Independence of continuous R.Vs

Definition: X, Y independent if

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \text{ for all } x, y \in \mathbb{R}$$

OR

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$$

Theorem: Let X, Y be r.v.s with joint PDf $f_{X,Y}$, then the following are equivalent

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

Theorem: If X, Y independent then

$$E[XY] = E[X] E[Y]$$

Proof: using LOTUS in 2D

$$\begin{aligned}
E[XY] &= \iint_{\mathbb{R}^2} xyf_{X,Y}(x, y)dxdy \\
&= \iint_{\mathbb{R}^2} xf_X(x) \cdot yf_Y(y)dxdy \text{ by independence} \\
&= \int_{\mathbb{R}} xf_X(x)dx \cdot \int_{\mathbb{R}} yf_Y(y)dy \text{ by factoring} \\
&= E[X] \cdot E[Y]
\end{aligned}$$

Chapter 20: Class 20

20.1 LOTUS in multivariate distributions

20.1.1 Discrete case

$$E\left[g(X,Y)\right] = \sum_x \sum_y g(x,y) P_{X,Y}(x,y)$$

20.1.2 Continuous case

$$E\left[g(X,Y_0\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$$

Chapter 21: Class 21

21.1 Covariance

Definition: For r.v. X, Y, the **covariance** is defined as

$$Cov(X,Y) = E\left[(X - E[X]) \cdot (Y - E[Y])\right]$$

Equivalently,

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

Note that

$$Var(X) = Cov(X,X)$$

Properties of covariance

- Independent r.v.s have zero covariance

$$X \perp Y \implies Cov(X,Y)$$

- Symmetry

$$Cov(X,Y) = Cov(Y,X)$$

- Covariance between r.v. and constant is zero

$$Cov(X,c) = 0 \text{ for any constant } c$$

- Scaling by a constant

$$Cov(aX,Y) = aCov(X,Y)$$

- Sum of two random variables

$$Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$$

- Covariance of sum of r.v.s

$$Cov(X_1 + X_2, Y_1 + Y_2) = Cov(X_1, Y_1) + Cov(X_1, Y_2) + Cov(X_2, Y_1) + Cov(X_2, Y_2)$$

- More generally,

$$Cov\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} Cov(X_i, Y_j)$$

Proofs of properties

Proof of (1):

$$\begin{aligned} X \perp Y &\implies E[XY] = E[X]E[Y] \\ &\implies Cov(X,Y) = E[XY] - E[X]E[Y] = E[XY] - E[XY] = 0 \end{aligned}$$

Proof of (2)

$$\begin{aligned}
Cov(X, Y) &= E[XY] - E[X]E[Y] \\
&= E[YX] - E[Y]E[X] \\
&= Cov(Y, X)
\end{aligned}$$

Proof of (3)

$$\begin{aligned}
Cov(X, c) &= E[cX] - E[c]E[X] \\
&= cE[X] - cE[X] \\
&= 0
\end{aligned}$$

Proof of (4)

$$\begin{aligned}
Cov(aX, Y) &= E[aXY] - E[aX]E[Y] \\
&= aE[XY] - aE[X]E[Y] \\
&= a\left(E[XY] - E[X]E[Y]\right) \\
&= aCov(X, Y)
\end{aligned}$$

Proof of (5)

$$\begin{aligned}
Cov(X_1 + X_2, Y) &= E[(X_1 + X_2)Y] - E[X_1 + X_2]E[Y] \\
&= E[X_1Y + X_2Y] - E[X_1 + X_2]E[Y] \\
&= E[X_1Y] + E[X_2Y] - E[X_1]E[Y] - E[X_2]E[Y] \\
&= E[X_1Y] - E[X_1]E[Y] + E[X_2Y] - E[X_2]E[Y] \\
&= Cov(X_1, Y) + Cov(X_2, Y)
\end{aligned}$$

Proof of (6)

$$\begin{aligned}
&Cov(X_1 + X_2, Y_1 + Y_2) \\
=&Cov(X_1, Y_1 + Y_2) + Cov(X_2, Y_1 + Y_2) \quad \text{by Property 5} \\
=&Cov(X_1, Y_1) + Cov(X_1, Y_2) + Cov(X_2, Y_1) + Cov(X_2, Y_2)
\end{aligned}$$

21.2 General case for variance of a sum of r.v.s

Theorem: For any X, Y

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

Proof

$$\begin{aligned}
Var(X + Y) &= Cov(X + Y, X + Y) \\
&= E\left[(X + Y)^2\right] - E[X + Y]^2 \\
&= E\left[X^2 + 2XY + Y^2\right] - (E[X + Y])^2 \\
&= E[X^2] + E[2XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2 \\
&= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2\left(E[XY] - E[X]E[Y]\right) \\
&= Var(X) + Var(Y) + 2Cov(X, Y)
\end{aligned}$$

OR

$$\begin{aligned}
&Cov(X + Y, X + Y) \\
=&Cov(X, Y) + Cov(X, Y) + Cov(Y, X) + Cov(Y, Y) \\
=&Var(X) + 2Cov(X, Y) + Var(Y)
\end{aligned}$$

21.2.1 Example

Let $X, Y \sim Unif(0,1)$, find $Cov(X+Y, X-Y)$

$$\begin{aligned} Cov(X+Y, X-Y) &= Var(X) - Var(Y) \\ &= 0 \end{aligned}$$

$X+Y, X-Y$ not independent

$$X+Y=2 \iff X=Y=1 \iff X-Y=0$$

21.3 Correlation

*not on Finals

Definition: Correlation between X, Y is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y}} = \frac{Cov(X,Y)}{sd(X)\cdot sd(Y)}$$

*Correlation is the rescaled version of $Cov(X,Y)$

Properties

- Location-scale transforms have no effect on correlations

$$Corr(aX+b, cY+d) = Corr(X,Y) \text{ for any } a>0, c>0$$

- Correlation doesn't depend on units of measurement
- Correlation takes values from -1 to 1

$$-1 \leq Corr(X,Y) \leq 1$$

- Perfect correlation implies perfect linear relationship

$$|Corr(X,Y)| = 1 \implies Y = aX+b$$

21.4 Multivariate Gaussian Distribution

*not on Finals

Definition: a **random vector** is an ordered list of random variables

Definition: a random vector $(X_1 \dots X_k)$ has the Multivariate Gaussian distribution if every linear combination of the X_j s has a Gaussian distribution.

$$(X_1 \dots X_k) \sim MVG \text{ if } \forall \mathbf{t} \in \mathbb{R}^k$$
$$\left(\sum_{i=1}^{k} t_i X_i\right) \sim N(\mu, \sigma^2) \text{ for some } \mu, \sigma$$

Chapter 22: Class 22 - Transformations

22.1 Discrete random variables

Discrete: Given the distribution of X, to compute the distribution of $Y = g(X)$

$$P(Y = y) = P(g(X) = y) = \sum_{x s.t. g(x)=y} P(X = x)$$

If g invertible,

$$P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y))$$

22.2 Continuous random variables

22.2.1 Via CDF

Continuous:

Method 1

Step 1: Find the CDF of Y

For order-preserving, invertible g,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y) = F_X(g^{-1}(y))$$

Step 2: differentiate to get the PDF of Y

22.2.2 Via change in variable

Theorem: Suppose X is a continuous r.v. with PDF f_X, let $Y = g(X)$ where g is differentiable, and strictly increasing or decreasing on the support of X, then the PDF of y is

$$f_Y(y) = f_X(x) \cdot \left| \frac{dx}{dy} \right| \text{ where } x = g^{-1}(y)$$

Note

- $$\frac{dx}{dy} = \frac{1}{dy/dx}$$
- Support of Y is the range of g whose input is restricted to the support of X

$$support(Y) = \{y = g(x) \text{ for some } x \in support(X)\}$$

In multivariable setting

Given $\vec{Y} = (Y_1, Y_2 \dots Y_n)$, $\vec{y} = (y_1, y_2 \dots y_n)$,

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x}) \cdot \left|\frac{\partial \vec{x}}{\partial \vec{y}}\right|$$

Chapter 23: Convolutions and Conditional expectation

23.1 Convolutions

Theorem [Discrete] Let X, Y be independent r.v.s, the PMF of $T = X + Y$ is

$$P(T = t) = \sum_x P(Y = t - x) \cdot P(X = x)$$

Justification: Law of Total Probability

$$\text{LOTP} : P(B) = \sum_i P(B|A_i)P(A_i)$$

In this case

$$\begin{aligned}
P(T = t) &= P(X + Y = t) \text{ by definition} \\
&= \sum_x P(X + Y = t|X = x) \cdot P(X = x) \text{ LOTP} \\
&= \sum_x P(Y = t - x|X = x) \cdot P(X = x) \text{ by rearranging} \\
&= \sum_x P(Y = t - x) \cdot P(X = x) \text{ by independence of } X, Y
\end{aligned}$$

Theorem [Continuous] Let X, Y be independent r.v.s, the PDF of $T = X + Y$ is

$$f_T(t) = \int_{\infty}^{\infty} f_X(x) \cdot f_Y(t - x)dx$$

Chapter 24: Conditional Expectation

24.1 Definition of $E[Y|X=x]$

Definition Conditional given event

- discrete Y

$$E[Y|A] = \sum_y yP(Y = y|A)$$

- continuous Y

$$E[Y|A] = \int_{\infty}^{\infty} yf_Y(y|A)dy$$

Definition Conditional given the value of a random variable X

- If X is discrete

$$E[Y|X = x] = E[Y|A] \text{ where } A \text{ is the event that } X = x$$

- if X, Y continuous

$$E[Y|X = x] = \int_{\infty}^{\infty} yf_{Y|X}(y|x)dy$$

- if X continuous, Y discrete

$$E[Y|X = x] = \sum_y yP(Y = y|X = x)$$

24.2 Definition of $E[Y|X]$

The **conditional expectation** of Y, given X is the r.v. $g(X)$

$$E[Y|X] = g(X)$$

such that for each x in the support of X, $g(X)$ takes the value of

$$E[Y|X = x]$$

24.3 Properties of conditional expectation

Ignoring what's independent

$$X \perp Y \to E[Y|X] = E[Y]$$

Taking out what's known:

$$E[h(X)Y|X] = h(X) \cdot E[Y|X]$$

Linearity

$$E[Y_1 + Y_2|X] = E[Y_1|X] + E[Y_2|X]$$

Law of Iterated Expectation / Law of Total Expectations / Adam's Law / Tower Law

$$E[Y] = E\left[[Y|X]\right]$$

Chapter 25: Lesson 25

25.1 Markov and Chebyshev Inequalities

25.1.1 Markov's inequality

Theorem For any r.v. X and any constant $a > 0$

$$P\left(|X| \geq a\right) \leq \frac{E\left[|X|\right]}{a}$$

Proof: Let $Y = |X|$, show that

$$aP(Y \geq a) \leq E[Y]$$

Using Tower's Law

$$\begin{aligned} E[Y] &= E\left[Y|Y \geq a\right] P(Y \geq a) + E\left[Y|Y < a\right] P(Y < a) \\ E\left[Y|Y \geq a\right] &\geq a \\ E\left[Y|Y < a\right] &\geq 0 \\ P(Y \geq a) &\geq 0 \\ P(Y < a) &\geq 0 E[Y] \geq aP(Y \geq a) \end{aligned}$$

25.1.2 Chebyshev's Inequality

Theorem: Let X have mean μ and variance σ^2, then for any $a > 0$

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Usage:

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

Proof

$$\begin{aligned} P(|X - \mu| \geq a) &= P((X - \mu)^2 \geq a^2) \\ &\leq \frac{E\left[(X - \mu)^2\right]}{a^2} \\ &= \frac{Var(X)}{a^2} \end{aligned}$$

25.2 Limit Theorem

25.2.1 Weak law of large numbers

Theorem: for $X_1, X_2 \ldots$ i.i.d with finite expectation μ and variance σ^2, for each $n = 1, 2, \ldots$, let $\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$

$$\forall \epsilon > 0, \lim_{n\to\infty} P(|\overline{X_n} - \mu| \leq 1) = 1]$$

Proof:

VERY IMPORTANT

Step 1: $E[\overline{X_n}] = \mu$

$$E\left[\overline{X_n}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

Step 2: $Var(\overline{X_n}) = \frac{\sigma^2}{n}$

$$Var\left(\overline{X_n}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

Step 3: Use the Chebyshev's inequality on $\overline{X_n}$

$$P\left(\left|\overline{X_n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Hence

$$\lim_{n\to\infty} P\left(\left|\overline{X_n} - \mu\right| \leq \epsilon\right) = 1$$

Theorem Central Limit Theorem states that for X_1, X_2 i.i.d from some distribution with finite mean μ and variance σ, for any $n = 1, 2, 3 \ldots$,

$$\overline{X}_n = \frac{X_1 \ldots X_n}{n}$$

and

$$Z_n = \frac{\overline{X}_n - \mu}{\frac{sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right)$$

Chapter 26: Class 26 - Review

26.1 Q1

Problem: Suppose $Y_1 \dots Y_n$ are iid $Ber(p)$, what is the approximate distribution of the sample mean $\overline{Y}_n = \frac{1}{n}\sum_{n=1}^{100} Y_i$ for n large.

Solution: By CLT, approximately

$$Y_i \approx N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(p, \frac{p(1-p)}{n}\right)$$

This is useful because

$$\begin{aligned}
&P\text{(more than 60 heads in 100)} \\
=&P\left(N\left(p, \frac{p(1-p)}{n}\right) \geq 0.6\right) \\
=&P\left(N\left(0.5, \frac{1}{400}\right) \geq 0.6\right) \\
=&P\left(N\left(0, \frac{1}{400}\right) \geq 0.1\right) \\
=&P\left(N(0,1) \geq 2\right) \\
=&2.5\%
\end{aligned}$$

26.2 Q2

Problem : Same as above, but now $Y_i = aX_i + b$

Solution: By CLT, approximately

$$Y_i \approx N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(ap+b, \frac{a^2p(1-p)}{n}\right)$$

26.3 Q3

Problem:Same as above, but now $Y_i = e^{-X_i}$ where $X_i \sim Expo(1)$
Solution:

$$\begin{aligned}
E\left[Y_i\right] &= E\left[e^{-X_i}\right] \\
&= \int_0^\infty e^{-x}e^{-x}dx \\
&= \frac{1}{2}E\left[Y_i^2\right] = \dots = \frac{1}{3}
\end{aligned}$$

$$Y_i \approx N\left(\frac{1}{2}\frac{1}{12n}\right)$$

26.4 Q4

Flip a coin until first heads. Let X denote number of tails seen.

$$
\begin{aligned}
E\left[X|X \leq 2\right] &= \sum_{k=0}^{\infty} kP(X = k|X \leq 2) \\
E\left[X|X \leq 2\right] &= \sum_{k=0}^{2} k\frac{P(X = k)}{7/8} \\
&= \frac{8}{7}\sum_{k=0}^{2} k\frac{1}{2^{k+1}} \\
&= \frac{8}{7}\frac{1}{2} \\
&= \frac{4}{7}
\end{aligned}
$$

26.5 Q5

Problem: Given 2 coins, one with probability a of H, the other with b. Choosing one uniformly at random and flipping until first heads, what is the expected number of flips?
Solution:

$$E\left[X\right] = E\left[X|A\right] P(A) + E\left[X|A^C\right] P(A^C) = \frac{1}{2}\left(\frac{1}{a} + \frac{1}{b}\right)$$

26.6 Q6

Given a coin with probability of head p, flip it once, and let X be the number of additional flip until it flips the original flip again. Compute $E\left[X\right]$

$$
\begin{aligned}
E\left[X\right] &= E\left[X|A\right] P(A) + E\left[X|A^C\right] P(A^C) \\
&= \frac{1}{p}(p) + \frac{1}{1-p}(1-p) \\
&= 2
\end{aligned}
$$

26.7 Q7

Let $A, B \sim Unif(0, 1)$, what is the probability that $Ax^2 + Bx + 1$ has at least 1 real root?

$$
\begin{aligned}
&P(B^2 - 4A \geq 0) \\
=&P\left(A \leq \frac{B^2}{4}\right) \\
=&\int_0^1 \int_0^{\frac{b^2}{4}} 1 da db \\
=&\int_0^1 \frac{b^2}{4} db \\
=&\frac{1}{12}
\end{aligned}
$$