

Math 170E – Intro to Probability

University of California, Los Angeles

Duc Vu

Fall 2020

This is math 170E taught by Professor Nguyen. The formal name of the class is **Introduction to Probability and Statistics 1: Probability**. The textbook used for the class is *Probability & Statistical Inference* 10th by *Hogg, Tanis*. We meet weekly on MWF from 10:00 – 10:50 and on Tue at the same time frame for discussion with our TA, Jason Snyder. You can also find other lecture notes at my [github](#). Let me know through my [email](#) if you notice something mathematically wrong/concerning. Thank you!

Contents

1	Lec 1: Oct 2, 2020	4
1.1	Properties of Probability	4
2	Lec 2: Oct 5, 2020	6
2.1	Method of Enumeration	7
3	Lec 3: Oct 7, 2020	9
3.1	Conditional Probability	9
4	Lec 4: Oct 9, 2020	12
4.1	Independent Events	12
5	Lec 5: Oct 12, 2020	14
5.1	Independent Events (cont'd)	14
5.2	Bayes's Theorem	15
6	Lec 6: Oct 14, 2020	17
6.1	Random Variables with Discrete Type	18
7	Lec 7: Oct 16, 2020	21
7.1	Lec 6 (Cont'd)	21
7.2	Expectation & Special Math Expectations	23
8	Lec 8: Oct 19, 2020	26
8.1	Info about 1 st midterm	26
8.2	Lec 7 (Cont'd)	26

9 Lec 9: Oct 21, 2020	30
9.1 Binomial Distribution	30
10 Lec 10: Oct 23, 2020	34
10.1 Practice 6 Problem 3	34
10.2 Hypergeometric Distribution	34
11 Lec 11: Oct 26, 2020	35
11.1 Negative Binomial Distribution	35
11.2 Poisson Distribution	37
12 Lec 12: Oct 28, 2020	38
12.1 Lec 11 (Cont'd)	38
12.2 Binomial Distribution Approximation by Poisson Distribution	40
12.3 Random Variable of Continuous Type	40
13 Lec 13: Oct 30, 2020	42
13.1 Random Variable of Continuous Type (Cont'd)	42
14 Midterm 1: Nov 2, 2020	46
15 Dis 1: Oct 6, 2020	46
15.1 Set Theory	46
16 Dis 2: Oct 13, 2020	48
16.1 Conditional Probabilities	48
16.2 Bayes's Theorem	48
17 Dis 3: Oct 20, 2020	49
17.1 Recap of Terminology/Functions	49
18 Dis 4: Oct 27, 2020	51
18.1 Review of Chapter 2	52
19 Dis 5: Nov 3, 2020	53
19.1 Continuous Random Variables	53
19.2 Types of Distribution for Continuous Variable	55

List of Definitions

1.1 Outcome Space	4
1.4 Probability	5
2.1 "Equally Likely"	7
2.3 Permutation of n objects	7
2.4 Permutation/Combination of size n taken k	8
3.1 Conditional Probability	9
4.2 Independent Events	13
5.1 Mutually Independent Events	15
5.3 Partition of Outcome Space	15

6.2	Random Variable	18
6.5	Probability Mass Function	19
6.7	Cumulative Distribution Function	20
7.2	Mathematical Expectation	24
7.5	Mean, Variance, & Standard Deviation	25
8.1	Moment Generating Function	28
8.5	n^{th} Moment	29
9.1	Bernoulli Trial	30
9.3	Bernoulli Random Variable	31
9.5	Binomial Distribution	32
11.1	Negative Binomial Distribution	35
11.3	Poisson Distribution	38
13.1	Probability Density Function	43
13.3	Percentile	46
15.1	Set	46

§1 | Lec 1: Oct 2, 2020

§1.1 Properties of Probability

Definition 1.1 (Outcome Space) — Consider the outcome of a random experiment, e.g. flipping a coin. The collection of all such outcomes, denoted by $\underbrace{S}_{\omega \text{ in other advanced prob. textbook}}$, is called the outcome space.

- A subset $A \subseteq S$ is called an event.
- If $A_1, A_2, \dots \subseteq S$ satisfy $A_i \cap A_j = \emptyset, i \neq j$ then they are called “disjoint” (mutually exclusive)
- If $A_1, A_2, \dots, A_n \subseteq S$ satisfy $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n = S$. Then $\{A_i\}_{i=1 \dots n}$ are called exhaustive (fully comprehensive).

Example 1.2 1. Flip two coins in order. Denote H = head, T = tail.

$$S = \{HH, HT, TH, TT\}$$

$$A = \{HH\} = \{\text{both coins are head}\}$$

$A \subseteq S$ is an event.

$$B = \{HT, TH\}$$

$B \subseteq S$ is another event.

$A \cap B = \emptyset$, they are disjoint.

2. Flip 2 coins at once.

$$S = \{HH, HT, TT\}$$

$$A = \{\text{one head, one tail}\}$$

$$A = \{HT\}, \text{ is an event.}$$

Probability – A heuristic intro:

Consider an experiment and repeat n times. Let $N(A)$ = number of times A occurs. The ratio $\frac{N(A)}{n}$ is called the relative frequency of A in n repetitions of the experiment.

$$0 \leq \frac{N(A)}{n} \leq 1$$

As $n \rightarrow \infty$,

$$\frac{N(A)}{n} \rightarrow p \in [0, 1]$$

This p is called the prob. that event A occurs.

Example 1.3

(a) Flip a coin

$$S = \{H, T\}$$

$$A = \{H\}$$

What is $P(A)$?

(b) Sometimes, we can also assign prob. based on the nature of the event Pick a random point in the unit circle.

$$A = \{\text{chosen point} \in 1^{\text{st}} \text{quadrant}\}$$

$$P(A) = \frac{\text{Area of first quadrant}}{\text{Area of unit circle}} = \frac{1}{4}$$

(c) Pick a number randomly from $\{0, 1, \dots, 9\}$, $B = \{2 \text{ is picked}\}$

$$P(B) = \frac{1}{10}$$

Table 1: From example 1.3 (a)

n	$N(A)$	$\frac{N(A)}{n}$
50	37	.74
500	333	.66

It is safe to assign $P(A) = 0.66$ **Definition 1.4 (Probability)** — Given an outcome space S , the probability of an event $A \subseteq S$, is a number satisfying:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. $A_1, \dots, A_n \subseteq S$ are disjoint events, i.e. $A_i \cap A_j = \emptyset, i \neq j$, then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) = P(A_1) + \dots + P(A_n)$$

More generally, if $A_1, \dots, A_n, \dots \subseteq S$ are disjoint events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Theorem 1.5 1. Denote A' to be the complement of A in S , i.e.

$$A' \cup A = S$$

$$A' \cap A = \emptyset$$

Then

$$P(A') = 1 - P(A)$$

$$2. P(\emptyset) = 0$$

$$3. \text{ If } A \leq B \text{ then } P(A) \leq P(B)$$

$$4. P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$5. P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

Note: The pattern here is add the prob. of odd event(s) and subtract the prob. of even events. (for prop (4) and (5) of theorem 1.5).

Proof.

$$P(A') = 1 - P(A)$$

Since $A' \cap A = \emptyset$ (by def of A'). By property (c),

$$P(\underbrace{A' \cup A}_S) = P(A') + P(A)$$

$$\underbrace{P(S)}_{1 \text{ (by prop. (b))}} = P(A') + P(A)$$

Thus,

$$P(A') = 1 - P(A)$$

§2 | Lec 2: Oct 5, 2020

Cont'd of Lec 1

(2)

$$\begin{aligned} P(\emptyset) &= 1 - P(S) \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

(3)

$$P(A) \leq P(B)$$

$B \setminus A$ is the set s.t.

$$A \cup (B \setminus A) = B$$

$$A \cap (B \setminus A) = \emptyset$$

something here

implying

$$P(A) \leq P(B)$$

(4)

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

(5) Textbook Section 1.1. □

Definition 2.1 (“Equally Likely”) — Suppose $S = \{e_1, \dots, e_m\}$ where each e_i is a possible outcome. Denote $n(s)$ = number of outcomes = m . If each e_i has the same prob. of occurring, then they are called equally likely. In particular,

$$P(e_i) = \frac{1}{n(s)} = \frac{1}{m}$$

Moreover, if $A \subseteq S$ is an event s.t. $n(A) = k$. Then,

$$P(A) = \frac{n(A)}{n(s)} = \frac{k}{m}$$

Example 2.2

Draw one card from a deck of 52 cards.

$$P(\text{each card is drawn}) = \frac{1}{52}$$

$A = \{\text{a king is drawn}\}$, so $n(A) = 4$. Thus,

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{52}$$

§2.1 Method of Enumeration

Multiplication Principle:

Suppose an experiment E_1 has n_1 outcomes

- For each outcome from E_1 , a 2nd experiment E_2 has n_2 outcomes. Then the composite $E_1 E_2$ has $n_1 \cdot n_2$ outcomes.

Permutation of size n:

Definition 2.3 (Permutation of n objects) — Suppose there are n positions to be filled by n persons. One such arrangement is called a permutation of size n .

FACT: the total number of different such arrangements is given by “ $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$ ”

Proof. • $E_1 =$ fill the 1st position from n persons $\implies n$ outcomes for E_1 .

- E_2 = fill the 2nd pos. from $n - 1$ persons left $\implies n - 1$ outcomes for E_2
- \vdots
- E_n = fill the n^{th} pos. from 1 person left $\implies 1$ outcome for E_n
- One arrangement = $E_1 E_2 \dots E_n$

Thus, total number of arrangements is $n!$. □

Permutation/Combination of n objects taken k :

Definition 2.4 (Permutation/Combination of size n taken k) — Given $k \leq n$ and suppose there are n objects. If k objects are taken from n **with/without** order, then such a selection is called **permutation/combination** of size n taken k .

Note: “Permutation of size n ” = “permutation of size n taken n ”.

Fact 2.1. 1. The total number of permutation n taken k (order is important here) is denoted by ${}^n P_k$ is given by

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

2. The total numbers of combination of n taken k , denoted by ${}^n C_k$ or $\binom{n}{k}$ is given by

$${}^n C_k = \binom{n}{k} = \frac{n!}{(n - k)!k!}$$

Proof. E_1 = fill 1st pos. from $n \implies n$ for E_1

\vdots

E_k = fill k^{th} pos. from $n - k + 1$ persons left. Thus,

$$\text{perm } k = n \cdot \dots \cdot (n - k + 1)$$

(2) Combination of n taken k :

Start with ${}^n P_k$ as follow:

- E_1 = take k from n at once, outcome = ${}^n C_k = \binom{n}{k}$
- E_2 = permute k , outcomes = $k!$. Thus,

$${}^n P_k = \binom{n}{k} \cdot k!$$

implying

$$\binom{n}{k} = \frac{{}^n P_k}{k!} = \frac{n!}{(n - k)!k!}$$

□

Practice 1: https://ccle.ucla.edu/pluginfile.php/3766550/mod_resource/content/1/Practice%201.pdf

1. Consider $S = \{1, \dots, 8\}$

a)

- $E_1 =$ filling 1st pos \implies 8 choices.
- Same for $E_2 \implies$ 8 choices.
- Likewise, E_3 has 8 choices.

Thus, the number of 3 digit numbers can be formed is 8^3

b) “3 distinct digit numbers” = “permutation of size 8 taken 3”

Thus, total such numbers is ${}_8P_3 = \frac{8!}{5!} = 8 \cdot 7 \cdot 6$

c) Considering subset where order is not taken into account

Combination of size 8 taken 3. Thus, the answer is

$$\binom{8}{3} = \frac{8!}{3!5!}$$

d) 3 digit numbers and divisible by 5

- $E_1 =$ choose 5 for the 3rd pos, so 1 choice.
- $E_2 =$ 8 choices
- $E_3 =$ 8 choices

Thus, the total of choices is $8 \cdot 8 = 64$.

e) 4 element subsets of S that has one even digit.

- $E_1 =$ choose one even digit from S , so 4 choices (2,4,6,8).
- $E_2 =$ choose 3 digits from $\{1, 3, 5, 7\}$ without order, so $\binom{4}{3}$

Thus, total = $E_1 \cdot E_2 = 4 \cdot \binom{4}{3}$.

e') What if “at least one even digit” instead of “exactly one even”?

1. Total = exactly “one even” + “two even” + “three even” + “four even”
2. Total = “4-element subset” - “4-element subset with no even digit”

§3 | Lec 3: Oct 7, 2020

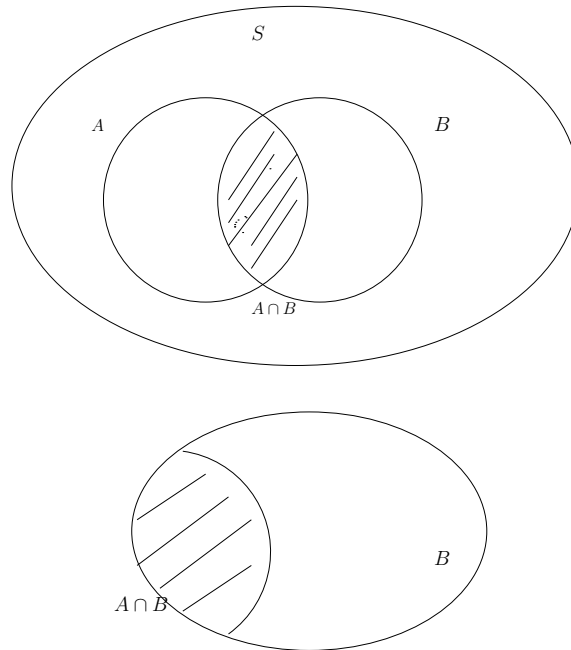
§3.1 Conditional Probability

Definition 3.1 (Conditional Probability) — Let $A, B \subseteq S$ be two events. The conditional prob. of A , given that B has occurred with $P(B) > 0$, is defined as

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

A heuristic explanation: $A \cap B$: “the portion in B that A occurs”

$$P(A|B) = \frac{\text{“area of A in B”}}{\text{“area of B”}}$$



Example 3.2

Suppose my family has two kids. Given that there is at least a boy, what is the prob. my family has two boys?

$$S = \{bb, bg, gb, gg\}$$

Now, let $B = \{\text{at least a boy}\}$. So we only look at the first three outcomes from S (B). Define $A = \{\text{two boys}\}$

$$A \cap B = \{bb\}$$

Note $A = A \cap B$ since $A \subseteq B$. Thus,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Note: We can also consider the alternative outcome space without order as follows

$$S = \left\{ (b, b) - \frac{1}{4}, (b, g) - \frac{1}{2}, (g, g) - \frac{1}{4} \right\}$$

Fact 3.1. $P(A|B)$ satisfies basic properties of probability:

- $P(A|B) \geq 0$
- $P(B|B) = 1$

Moreover, if $B \leq C$ then

$$P(C|B) = 1$$

- If $A_1, \dots, A_n \dots$ are disjoint events,

$$P\left(\bigcup_{k=1}^{\infty} A_k | B\right) = \sum_{k=1}^{\infty} P(A_k | B)$$

Proof. (a) $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$

(b) $P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$
 If $B \subseteq C$ then $B \cap C = B$

$$P(C|B) = \frac{P(B \cap C)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

$B \subseteq C$ means “if B occurs then C must occur”.

(c) $P(\bigcup_{k=1}^{\infty} A_k|B) = \frac{P(\bigcup_{k=1}^{\infty} A_k \cap B)}{P(B)}$. By distributive law,

$$\begin{aligned} &= \frac{P(\bigcup_{k=1}^{\infty} (A_k \cap B))}{P(B)} \\ &= \frac{\sum_{k=1}^{\infty} P(A_k \cap B)}{P(B)} \\ &= \sum_{k=1}^{\infty} P(A_k|B) \end{aligned}$$

□

INSERT: PRACTICE 1 #3 here

Theorem 3.3 1. $P(A \cap B) = P(A|B) \cdot P(B)$ given that $P(B) > 0$
 2. $P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$ given $P(A), P(A \cap B) > 0$.

Proof. 1. By defn of cond. prob.

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

implying

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

2. $P(A \cap B \cap C) = P(C \cap (A \cap B))$. By part 1,

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

□

Practice 3.1. The url: https://ccle.ucla.edu/pluginfile.php/3776692/mod_resource/content/0/Practice%202.pdf

INSERT: Look at the online notes

§4 | Lec 4: Oct 9, 2020

Cont'd (Practice)

3)

$$A = \{\text{spade}\} \quad B = \{\text{heart}\} \quad C = \{\text{diamond}\} \quad D = \{\text{club}\}$$

$P = (A \cap B \cap C \cap D = ?$ So,

$$= P(A)P(B|A)P(C|A \cap B)P(D|A \cap B \cap C)$$

(from problem 2 in practice 2)

- $P(A) = \frac{13}{52}$
- $P(B|A) =$, now restricted to outcome space {51 cards including 13 hearts} $B|A = \{\text{dealing a heart}\}$. Thus,

$$P(B|A) = \frac{13}{51}$$

- Similarly,

$$P(C|A \cap B) = \frac{13}{50}$$

(13 diamond from 50 cards left)

- $P(D|A \cap B \cap C) = \frac{13}{49}$ (13 clubs from 49 cards left).

Hence,

$$P(A \cap B \cap C \cap D) = \frac{13}{52} \frac{13}{51} \frac{13}{50} \frac{13}{49}$$

§4.1 Independent Events

Example 4.1

Flip a fair coin twice

$$S = \{HH, HT, TH, TT\}$$

$$A = \{1^{\text{st}}H\}$$

$$B = \{2^{\text{nd}}T\}$$

$$C = \{TT\}$$

$C \subseteq B$ “2 tails” \implies “2nd is T”. i.e., if C occurs then B must have occurred. Thus,

$$\begin{aligned} P(B|C) &= 1 \\ P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{1}{4}}{\frac{1}{2}} \\ &= \frac{1}{2} \\ P(A) &= \frac{1}{2} \end{aligned}$$

Thus, $P(A|B) = P(A)$, i.e., B occurring does not impact the occurrence of A.

Note also that

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

implying

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

Definition 4.2 (Independent Events) — Given two events A, B which are called independent iff

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

Theorem 4.3

The following are equivalent

- A, B are independent
- $P(A|B) = P(A)$, provided $P(B) > 0$
- $P(B|A) = P(B)$, provided $P(A) > 0$

Proof. Left as an exercise. □

Theorem 4.4 1. If $P(A) = 0$ then A is independent with any event.

2. If A and B are independent then so are the following pairs:

$$A, B' \quad A', B \quad A', B'$$

Proof. 1. Let B an arbitrary event, we need to show $P(A \cap B) = P(A)P(B)$. Since $P(A) = 0$, $P(A)P(B) = 0$.

$$A \cap B \subseteq A$$

imply

$$0 \leq P(A \cap B) \leq P(A) = 0$$

thus $P(A \cap B) = 0$.

2. Textbook(section 1.5)

□

Practice 4.1. Practice 2 – Problem 4:

Let's consider C and D first

$$\begin{aligned} D &= \{ \text{sum of two rolls} = 12 \} \\ &= \{(6, 6)\} \end{aligned}$$

Thus, $D \subseteq C = \{\text{first roll is 6}\}$. Hence, C and D are dependent.

A v.s. B

$$\begin{aligned} P(A) &= \frac{5}{6} \\ B &= \{ \text{sum is even} \} \\ &= \{ \text{first and second roll are even} \} \cup \{ \text{first and second roll are odd} \} \\ P(B) &= P(\text{first even})P(\text{second even}) + P(\text{first odd})P(\text{second odd}) \\ &= \frac{3}{6} \frac{3}{6} + \frac{3}{6} \frac{3}{6} \\ &= \frac{1}{2} \end{aligned}$$

Now, consider $A \cap B = \{1^{\text{st}} \neq 3, \text{sum is even}\}$. So,

$$\begin{aligned} A \cap B &= \{1^{\text{st}} \neq 3, 1^{\text{st}} \text{ odd}, 2^{\text{nd}} \text{ odd}\} \cup \{1^{\text{st}} \neq 3, 1^{\text{st}} \text{ even}, 2^{\text{nd}} \text{ even}\} \\ P(A \cap B) &= P(1^{\text{st}} \neq 3, 1^{\text{st}} \text{ odd})P(2^{\text{nd}} \text{ odd}) + P(1^{\text{st}} \neq 3, 1^{\text{st}} \text{ even})P(2^{\text{nd}} \text{ even}) \\ &= \frac{2}{6} \frac{3}{6} + \frac{3}{6} \frac{3}{6} \\ &= \frac{5}{12} \end{aligned}$$

Since $P(A \cap B) = \frac{5}{12} = \frac{5}{6} \frac{1}{2} = P(A)P(B)$, A and B are independent.

§5 | Lec 5: Oct 12, 2020

§5.1 Independent Events (cont'd)

Definition 5.1 (Mutually Independent Events) — A, B, C are called “mutually independent” if followings hold:

- pairwise independent

$$P(A \cap B) = P(A)P(B) \quad P(B \cap C) = P(B)P(C) \quad P(A \cap C) = P(A)P(C)$$

- “triple” wise independent, i.e.,

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

Note: analogous defn holds for A_1, \dots, A_n, \dots in which any pairs, triple, quadruple and so on must satisfy the similarly multiplication rules. Usually, the term “mutually” is dropped but it is understood that “independence” means “mutually independence”.

Remark 5.2. In general, pairwise independence does not imply triple-wise independence.

Practice 5.1. 2 – Problem 5:

$$A = \{1, 2\}, \quad B = \{1, 3\}, \quad C = \{1, 4\}$$

$$P(A) = \frac{2}{4} = P(B) = P(C)$$

$$A \cap B = \{1\} = B \cap C = A \cap C$$

$$P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{1}{4}$$

Thus,

$$P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

Same for B, C and A, C – so pairwise independent.

Triple:

$$A \cap B \cap C = \{1\}$$

$P(A \cap B \cap C) = \frac{1}{4}$; on the other hand, $P(A)P(B)P(C) = \frac{1}{2} \frac{1}{2} \frac{1}{2} = \frac{1}{8}$. They are not equal! Therefore, A, B, C are not mutually independent.

§5.2 Bayes’s Theorem

Definition 5.3 (Partition of Outcome Space) — The events B_1, \dots, B_n (n may be finite or ∞) are called a partition of the outcome space S if followings hold

- disjoint: $B_i \cap B_k = \emptyset, i \neq k$
- exhausted: $\bigcup_{i=1}^n B_i = S$

then,

$$P(B_1) + \dots + P(B_n) = P(S) = 1$$

Theorem 5.4 (Law of total Probability)

Suppose B_1, \dots, B_n is a partition of S with $P(B_i) > 0$ for $i = 1, \dots, n$. If A is an event in S , then

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

where $P(B_i)$ is called the prior probability.

Proof. (sketch)

$$\begin{aligned} P(A) &= P\left(\bigcup_{i=1}^n (A \cap B_i)\right) \\ &= \sum_{i=1}^n P(A \cap B_i) \\ &= \sum_{i=1}^n P(A|B_i)P(B_i) \end{aligned} \quad \square$$

Practice 5.2. 3 – problem 1:

$$\begin{aligned} P(I) &= .35 \\ P(II) &= .25 \\ P(III) &= .4 \end{aligned}$$

$A = \{ \text{a spring is defective} \}$, $P(A) = ?$ We know

$$\begin{aligned} P(A|I) &= .02 \\ P(A|II) &= .01 \\ P(A|III) &= .03 \end{aligned}$$

By law of total prob:

$$\begin{aligned} P(A) &= P(A|I)P(I) + P(A|II)P(II) + P(A|III)P(III) \\ &= 0.0215 \end{aligned}$$

Theorem 5.5 (Bayes's Theorem)

Suppose $\{B_i\}_{i=1, \dots, n}$ is a partition of S with $P(B_i) > 0$. If A with $P(A) > 0$, then for all $i = 1, \dots, n$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{k=1}^n P(A|B_k)P(B_k)}$$

where $P(B_i|A)$ is called posterior probability.

Proof.

$$\begin{aligned}
 P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} \\
 &= \frac{P(A \cap B_i)}{P(A)} \\
 &= \frac{P(A|B_i)P(B_i)}{P(A)} \\
 &= \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + \dots + P(A|B_n)P(B_n)} \quad \square
 \end{aligned}$$

Practice 5.3. 3 – problem 2: $A = \{ \text{person has disease} \}$, $P(A) = .005$.

$$\begin{aligned}
 + &= \{ \text{test } + \} \\
 - &= \{ \text{test } - \} \\
 P(+|A) &= .99 \\
 P(\underbrace{+|A'}_{\text{false positive}}) &= .03 \\
 P(A|+) &=?
 \end{aligned}$$

By Bayes's Theorem:

$$\begin{aligned}
 P(A|+) &= \frac{P(+|A)P(A)}{P(+|A)P(A) + P(+|A')P(A')} \\
 &= \frac{(.99)(.005)}{(.99)(.005) + (.03)(.995)}
 \end{aligned}$$

$\{A, A'\}$ is a partition of S .

§6 | Lec 6: Oct 14, 2020

Practice 6.1. 3 – Problem 3: Trial: know at least 1 girl

$$P(GG|\text{at least a girl}) = \frac{1}{3}$$

However, the above approach is not correct.

Intuition: The moment the girl opens the door, the first child's gender is determined – which makes the other kid's gender is now independent of the girl. Thus, $P(\text{other kid is girl}) = \frac{1}{2}$.

Correct approach:

$$\begin{aligned}
 A &= \{ \text{a girl opens the door} \} \\
 P(GG|A) &=?
 \end{aligned}$$

- $P(A|GG) = 1$
- $P(A|BB) = 0$
- $P(A|GB) = \frac{1}{2}$

- $P(A|BG) = \frac{1}{2}$

By Bayes' Theorem

$$\begin{aligned} P(GG|A) &= \frac{P(A|GG)P(GG)}{P(A|GG)P(GG) + P(A|BB)P(BB) + P(A|BG)P(BG) + P(A|GB)P(GB)} \\ &= \frac{1}{2} \end{aligned}$$

§6.1 Random Variables with Discrete Type

Example 6.1

Flip a coin

$$S = \{H, T\}$$

Define

$$X : S \rightarrow \mathbb{R}$$

$$\omega \mapsto X(\omega) \in \mathbb{R}$$

$$\text{s.t. } X(H) = 0, \quad X(T) = 1$$

$$H \xrightarrow{X} 0$$

$$T \xrightarrow{\quad} 1$$

The function X is called a random variable (RV). Since S is discrete space, X is called a RV of discrete-type.

Definition 6.2 (Random Variable) — Given an outcome space S , a function X that assigns $X(s) = x \in \mathbb{R}$ for each $s \in S$ is called a random variable. The space(range) of X is the collection of real numbers, denoted by S_x ,

$$S_x = \{x \in \mathbb{R} : \exists s \in S, X(s) = x\}$$

S_x is also called the “support” of X .

When the outcome space S is discrete, then X is called a discrete random variable.

Example above:

$$S_x = \{0, 1\}$$

Note: the space of X is denoted by S in the textbook. Here we will use S_x .

Remark 6.3. Under the above definition, for $x \in S_x$,

$$P(X = x) = P(\{s \in S : X(s) = x\})$$

Example 6.4

Roll a fair dice

$$\begin{aligned} S &= \{1, 2, \dots, 6\} \\ X : S &\rightarrow \mathbb{R} \\ s &\mapsto X(s) = x \\ S_x &= \{1, 2, \dots, 6\} (= S) \end{aligned}$$

For each $k \in S_x$,

$$P(X = k) = P(\{k\}) = \frac{1}{6}$$

Also,

$$\sum_{k \in S_x} P(X = k) = \sum_{k=1}^6 \frac{1}{6} = 1$$

Definition 6.5 (Probability Mass Function) — The probability mass function (pmf) $f(x)$ of a discrete random variable X is a function satisfying the followings:

- $f(x) > 0$, $x \in S_x$.
- $\sum_{x \in S_x} f(x) = 1$.
- If $A \subseteq S_x$,

$$P(X \in A) = \sum_{x \in A} f(x)$$

Note: if $x \notin S_x$, then we assign $f(x) = 0$ ($P(X = x) = 0$).

Example 6.6 (above)

the pmf of X is given by $f(k) = \frac{1}{6}$ for $k = 1, \dots, 6$

$$\begin{aligned} A &= \{1, 2, 3\} = "X < 4" \\ A &\subseteq S_x \end{aligned}$$

$$P(X \in A) = \sum_{k \in A} f(k) = \sum_{k=1}^3 \frac{1}{6} = \frac{1}{2}$$

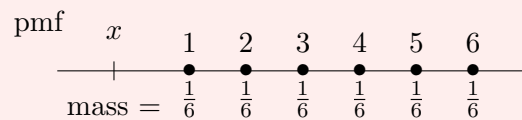
Definition 6.7 (Cumulative Distribution Function) — Cumulative distribution function (cdf) $F(x)$ of a RV x is a function given by

$$F(x) = P(X \leq x), \quad -\infty < x < \infty$$

Note: $F(x)$ is usually called distribution function, “cumulative” is dropped.

Example 6.8

Rolling a fair dice



$$\text{cdf } F(x) = P(X \leq x)$$

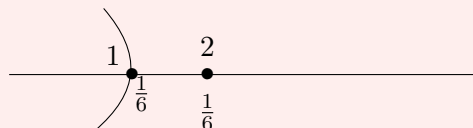
= total mass cumulated starting from the left up to x

$x < 1$,

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= 0 \text{ (no mass up to } x < 1) \end{aligned}$$

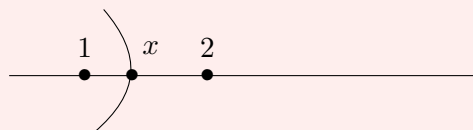
$x = 1$,

$$F(1) = P(X \leq 1)$$



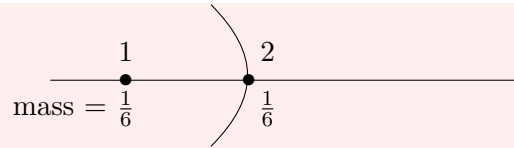
$$F(1) = \frac{1}{6} \text{ (mass up to and including location 1).}$$

$1 < x < 2$



$$\begin{aligned} F(x) &= P(X \leq 1) \\ &= P(X = 1) \\ &= \frac{1}{6} \end{aligned}$$

$x = 2$



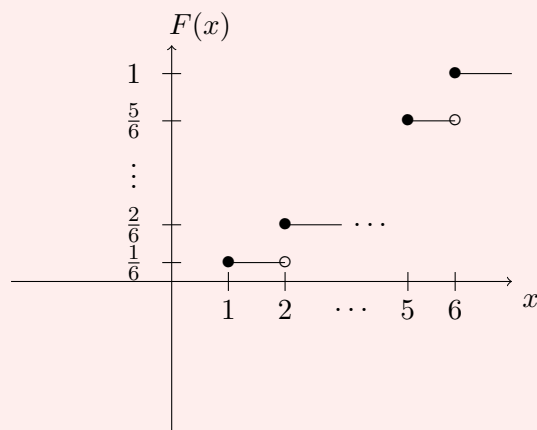
$$\begin{aligned} F(2) &= P(X \leq 2) \\ &= P(X = 1) + P(X = 2) \\ &= \frac{2}{6} \end{aligned}$$

Likewise, $2 < x < 3$

$$F(x) = \frac{2}{6}$$

$$\therefore x = 6, \quad F(X) = P(X \leq 6) = 1$$

$$x > 6, F(x) = 1$$

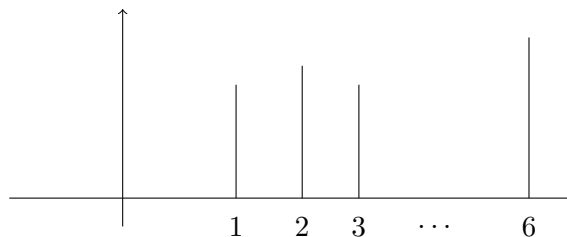


§7 | Lec 7: Oct 16, 2020

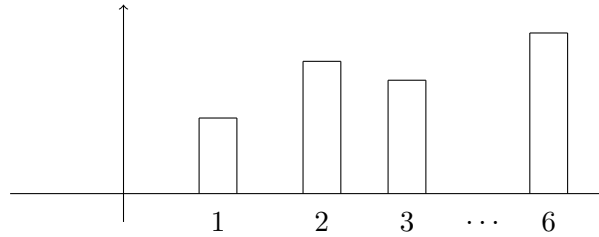
§7.1 Lec 6 (Cont'd)

In order to graph the prob. mass function:

- Line graph



- Histogram



Practice 7.1. 4 – Problem 1:

$$X = \text{max of two rolls}$$

$$S_X = \{1, 2, \dots, 6\}$$

For $k \in S_X$. Determine $f(k) = P(X = k) = ?$

- 1st approach:

$\begin{smallmatrix} 2^{nd} \text{ roll} \\ \diagdown \\ 1^{st} \text{ roll} \end{smallmatrix}$	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
\vdots						
6	(6, 1)	(6, 2)	\dots			

$\begin{smallmatrix} 2^{nd} \text{ roll} \\ \diagdown \\ 1^{st} \text{ roll} \end{smallmatrix}$	1	2	3	\dots	6
1	1	2	3	\dots	6
2	2	2	3	\dots	6
3	3	3	3	\dots	6
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
6	6	6	6	\dots	6

$$f(1) = P(X = 1) = \frac{1}{36}$$

$$f(2) = P(X = 2) = \frac{3}{36}$$

$$f(3) = P(X = 3) = \frac{5}{36}$$

$$\vdots$$

$$f(6) = P(X = 6) = \frac{11}{36}$$

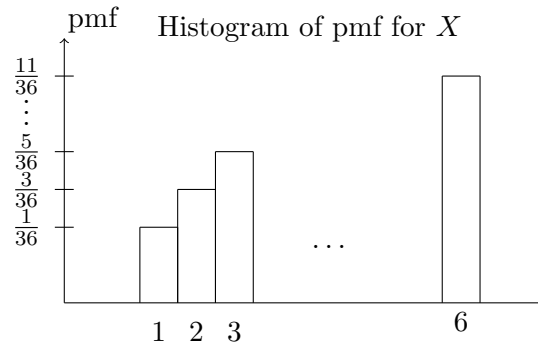
- 2nd approach: for $k = 1, \dots, 6$ (disjoint sub-events)

$$\begin{aligned}\{X = k\} &= \{\max = k\} \\ &= \{1^{\text{st}} \text{roll} = k, 2^{\text{nd}} < k\} \\ &\cup \{1^{\text{st}} \text{roll} < k, 2^{\text{nd}} = k\} \\ &\cup \{1^{\text{st}} \text{roll} = 2^{\text{nd}} = k\}\end{aligned}$$

Thus,

$$\begin{aligned}P(X = k) &= P(1^{\text{st}} \text{roll} = k)P(2^{\text{nd}} < k) + P(1^{\text{st}} < k)P(2^{\text{nd}} = k) + P(1^{\text{st}} = k)P(2^{\text{nd}} = k) \\ &= \frac{1}{6} \frac{k-1}{6} + \frac{k-1}{6} \frac{1}{6} + \frac{1}{6} \frac{1}{6} \\ &= \frac{2k-1}{36}\end{aligned}$$

Note: $\sum_{k=1}^6 \frac{2k-1}{36} = 1$.



Similarly, we can calculate $Y = \min$ of 2 rolls.

Remark 7.1. Suppose $X = \max\{U, V\}$ where U, V are 2 discrete random variables. Then pmf of X can be calculated as follows:

$$\begin{aligned}f(k) &= P(X = k) \\ &= P(U = k, V < k) + P(U < k, V = k) + P(U = k, V = k)\end{aligned}$$

and we can often use indep. on each of the above events. On the other hand, for $Y = \min\{U, V\}$ then

$$P(Y = k) = P(U = k, V > k) + P(U > k, V = k) + P(U = k, V = k)$$

and use indep. on the above events.

§7.2 Expectation & Special Math Expectations

Definition 7.2 (Mathematical Expectation) — Suppose X is a discrete random variable with S_X , pmf $f(x)$. Let $u(x)$ be a function, then if the sum $\sum_{x \in S_X} u(x)f(x)$ exists (finite) then the sum is mathematical expectation (expected value) of $u(X)$ and is denoted by

$$E[u(X)] := \sum_{x \in S_X} u(x)f(x)$$

Practice 7.2. 5 – Problem 1: $S_X = \{1, \dots, 6\}$. For $x \in S_X$, $u(x) = x - 3.5$

$$\begin{aligned} \text{average income} &= E[u(x)] \\ &= \sum_{x \in S_X} u(x)f(x) \\ &= \sum_{k=1}^6 (k - 3.5) \cdot \frac{1}{6} \\ &= 0 \end{aligned}$$

“After one game, on average, I do not gain/lose any money.”

Theorem 7.3

When it exists, the expectation E satisfies:

- If c is a constant, then

$$E[c] = c$$

- If c is a constant and $u(X)$ is a function, then

$$E[c \cdot u(X)] = cE[u(X)]$$

- If c_1, c_2 are constants and $u_1(X), u_2(X)$ are functions.

$$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$$

Remark 7.4. Part (c) can be generalized for 2 discrete random variables X, Y .

$$E[c_1 u_1(X) + c_2 u_2(Y)] = c_1 E[u_1(X)] + c_2 E[u_2(Y)]$$

Proof. Textbook. □

Definition 7.5 (Mean, Variance, & Standard Deviation) — For a random variable X ,

- the mean (of X) is denoted by

$$\mu := E[x]$$

- the variance (of X) is denoted by

$$\sigma^2 := E[(x - \mu)^2]$$

- the standard deviation

$$\sigma := \sqrt{\sigma^2}$$

Example 7.6

Suppose X has pmf

x	-2	0	1
$f(x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

$$\text{mean} = \mu = E[x]$$

$$= \sum_{x \in S_X} x \cdot f(x)$$

$$= (-2)\frac{1}{2} + 0\frac{1}{3} + 1\frac{1}{6}$$

$$= -\frac{5}{6}$$

$$\text{variance} = \sigma^2 = E[(x - \mu)^2]$$

$$= \sum_{x \in S_X} (x - \mu)^2 f(x)$$

$$= (-2 - (-\frac{5}{6}))^2 \frac{1}{2} + (0 - (-\frac{5}{6}))^2 \frac{1}{3} + \dots$$

σ^2 interpretation:

For a constant $c \in \mathbb{R}$, define $g(c) := E[(x - c)^2]$. Note that

$$g(c) = E[(X - c)^2]$$

$$= E[X^2 - 2cX + c^2]$$

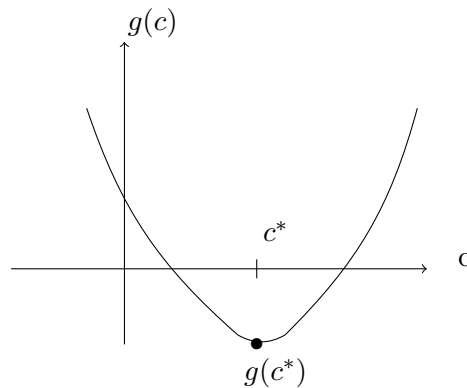
$$= E[X^2] + E[-2cX] + E[c^2]$$

$$= E[X^2] - 2cE[X] + c^2$$

$$= c^2 - 2cE[X] + E[X^2]$$

$$= c^2 - 2\mu \cdot c + E[X^2]$$

“ u and $E[X^2]$ are constant with respect to c ”.



$g(c^*) = \min g(c)$ where c^* satisfies

$$\begin{aligned} g'(c^*) &= 0 \\ g'(x) &= 2c - 2\mu \end{aligned}$$

Thus

$$g'(c^*) = 0 = 2c^* - 2\mu$$

i.e., $c^* = \mu$. Hence,

$$\sigma^2 = E[(x - \mu)^2] = g(\mu)$$

minimizes $g(c) = E[(x - c)^2]$, i.e.,

$$\sigma^2 = \underbrace{\min_{c \in \mathbb{R}} E[(x - c)^2]}_{c \in \mathbb{R}} = E[(x - \mu)^2]$$

“ σ^2 measures fluctuation of X around its mean μ .”

§8 | Lec 8: Oct 19, 2020

§8.1 Info about 1st midterm

1st Midterm 11/2, Monday, 10am PT. Due: 10am PT – Tuesday 11/3.

2nd Midterm, after Thanksgiving.

§8.2 Lec 7 (Cont'd)

Review geometric series: for $|q| < 1$,

$$\sum_{k=0}^{\infty} q^k = 1 + q + q^2 + \dots = \frac{1}{1 - q}$$

Differentiating both sides,

$$\sum_{k=1}^{\infty} kq^{k-1} = 1 + 2q + 3q^2 + \dots = \frac{1}{(1 - q)^2}$$

Practice 8.1. 5 – Problem 2:

$S_X = \{1, 2, \dots\}$. The pmf $f(f) = P(X = k) = P(1^{\text{st}} \text{ k-1 shots are missed and k shot successful.}$
 a) $E[X] = ?$

$$A_k = \left\{ k^{\text{th}} \text{ shot is successful} \right\}$$

$$P(A_k) = p$$

$$P(A'_k) = 1 - p = q = P\left(\left\{ k^{\text{th}} \text{ shot is missed} \right\}\right)$$

$$\begin{aligned} P(X = k) &= P\left(\underbrace{A'_1 \cap A'_2 \cap \dots \cap A'_{k-1}}_{\text{miss } 1^{\text{st}} \text{ } k-1 \text{ shots}} \cap \underbrace{A_k}_{\text{make at } k^{\text{th}} \text{ shots}}\right) \\ &\stackrel{\text{independence}}{=} P(A'_1)P(A'_2) \dots P(A'_{k-1})P(A_k) \\ &= q \cdot q \dots q \cdot p \\ &= q^{k-1} \cdot p \end{aligned}$$

for each $k = 1, 2, 3, \dots$. Note that pmf $f(k) = P(X = k)$ indeed satisfies:

$$\begin{aligned} \sum_{k=1}^{\infty} f(k) &= \sum_{k=1}^{\infty} q^{k-1} \cdot p \\ &= p(1 + q + q^2 + \dots) \\ &= p \cdot \frac{1}{1 - q} \\ &= p \cdot \frac{1}{p} \\ &= 1 \end{aligned}$$

Now,

$$\begin{aligned} \mu = E[x] &= \sum_{x \in S_X} x f(x) \\ &= \sum_{k=1}^{\infty} k \cdot f(k) \\ &= \sum_{k=1}^{\infty} k \cdot q^{k-1} \cdot p \\ &= p \sum_{k=1}^{\infty} k \cdot q^{k-1} \\ &= p \cdot (1 + 2q + 3q^2 + \dots) \\ &= p \cdot \frac{1}{(1 - q)^2} \\ &= p \cdot \frac{1}{p^2} \\ &= \frac{1}{p} \end{aligned}$$

Definition 8.1 (Moment Generating Function) — Given a discrete RV X and δ_X and pmf $f(x)$, if \exists a positive constant h s.t. for all $t \in (-h, h)$, the following expectation function

$$E[e^{tX}] = \sum_{x \in S_X} e^{tx} f(x)$$

exists then $E[e^{tx}]$ is called the mgf of X and is denoted by $M_X(t)$.

Note: $(-h, h)$ needs not be a symmetric interval. But it has to contain the origin 0.

Example 8.2

Suppose X has the following pmf,

x	-2	0	1
$f(x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

$$\begin{aligned} E[e^{tX}] &= M_X(t) = \sum_{x \in S_X} e^{tx} f(x) \\ &= \frac{1}{2}e^{-2t} + \frac{1}{3} + \frac{1}{6}e^t \end{aligned}$$

which is finite for all $t \in \mathbb{R}$.

Theorem 8.3

MGF determines RV X , i.e., if X and Y are 2 RV s.t.

$$M_X(t) = M_Y(t)$$

then

$$S_X = S_Y$$

and

$$\underbrace{f_X(x)}_{\text{pmf of } X} = \underbrace{f_Y(x)}_{\text{pmf of } Y} \quad \text{for } x \in S_X (= S_Y)$$

Example 8.4 (above)

Suppose Y has mgf

$$M_Y(t) = \frac{1}{2}e^{-2t} + \frac{1}{3} + \frac{1}{6}e^t$$

then

$$S_Y = \{-2, 0, 1\}$$

and $f_Y(-2) = \frac{1}{2}$, $f_Y(0) = \frac{1}{3}$, $f_Y(1) = \frac{1}{6}$. So that X and Y have same space and same pmf.

Practice 8.2. 5 – Problem 2b: X has geometric distribution with parameter $p \in [0, 1]$ denoted by $X \sim \text{Geom}(P)$.

with pmf $f(k) = q^{k-1}p$ for $k = 1, 2, \dots$, $q = 1 - p$. MGF of X is given by

$$\begin{aligned} M_X(t) &= \sum_{k=1}^{\infty} e^{tk} f(k) \\ &= \sum_{k=1}^{\infty} e^{tk} q^{k-1} p \\ &= p(e^t + e^{t2}q + e^{t3}q^2 + \dots) \\ &= p \cdot e^t (1 + (e^t q) + (e^t q)^2 + (e^t q)^3 + \dots) \\ &= pe^t \frac{1}{1 - e^t q} \end{aligned}$$

which is finite for t ,

$$\begin{aligned} 0 &< e^t \cdot q < 1 \\ e^t &< \frac{1}{q} \\ t &< \ln\left(\frac{1}{q}\right) \end{aligned}$$

Thus,

$$M_X(t) = \frac{pe^t}{1 - qe^t}, \quad \text{with } t < \ln\left(\frac{1}{q}\right)$$

Definition 8.5 (n^{th} Moment) — For each n positive integer, if $E[X^n] = \sum_{x \in S_X} x^n f(x)$ exists then $E[X^n]$ is called the n^{th} moment of X .

Remark 8.6. Properties of MGF $M_X(t)$

- $t = 0$, $M_X(0) = E[e^{0 \cdot X}] = E[1] = 1$.
- Derivatives of $M_X(t)$ is given by

$$\begin{aligned}\frac{d}{dt}[M_X(t)] &= \frac{d}{dt} [E[e^{tX}]] \\ &= E \left[\frac{d}{dt} e^{tX} \right] \quad \text{assume } \frac{d}{dt} \text{ and } E \text{ are interchangeable} \\ M'_X(t) &= E[Xe^{tX}]\end{aligned}$$

Thus,

$$M'_X(t) \Big|_{t=0} = E[Xe^{0 \cdot X}] = E[X], \text{ first moment of } X$$

- Similarly, 2nd derivative of $M_X(t)$ given by

$$\begin{aligned}M''_X(t) &= E[X^2 e^{tX}] \\ M''_X(t) \Big|_{t=0} &= E[x^2], \text{ second moment of } X\end{aligned}$$

- More generally, the n^{th} - derivative of M_X satisfies

$$M_X^{(n)}(t) \Big|_{t=0} = E[x^n]$$

hence the name “mgf”.

Example 8.7

$X \sim \text{Geom}(p)$.

$$\begin{aligned}M_X(t) &= \frac{pe^t}{1 - qe^t}, \quad q = 1 - p \\ M'_X(t) &= \frac{pe^t}{(1 - qe^t)^2} \\ M'_X(0) &= \frac{p}{(1 - q)^2} = \frac{p}{p^2} = \frac{1}{p} = E[x]\end{aligned}$$

§9 | Lec 9: Oct 21, 2020

§9.1 Binomial Distribution

Definition 9.1 (Bernoulli Trial) — Bernoulli trial is a random experiment such that the outcomes can be classified in one of two mutually exclusive and exhaustive ways.

Example 9.2 1. Flipping a coin $S = \{H, T\}$.

2. A sequence of Bernoulli trials occurs when the experiment is performed several times and the prob. of success is the same in every trial and the trials are independent.
3. A player shooting the throws in basket ball
 - Making the shots has prob. $p \in (0, 1)$.
 - Missing.

Each throw is a Bernoulli trial. A sequence of throw is a sequence of Bernoulli trial.

Definition 9.3 (Bernoulli Random Variable) — Let X be the random variable associated with a Bernoulli trial. Then X is called a Bernoulli R.V with the pmf

$$\begin{aligned} P(X = 1(\text{success})) &= p \\ P(X = 0(\text{failure})) &= 1 - p \end{aligned}$$

which can also be rewritten as:

$$f(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}$$

Note: A formula of variance

$$\begin{aligned} \sigma^2 &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - (E[X])^2 \\ &= M_X''(0) - (M_X'(0))^2 \end{aligned}$$

Practice 9.1. 6 – Problem 1: Let $X \sim \text{Bernoulli R.V}$ with p

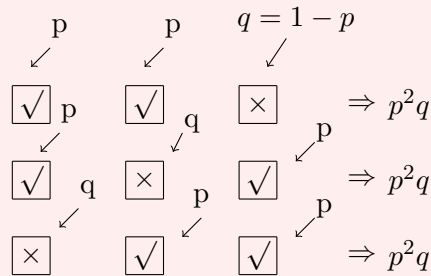
$$\begin{aligned} \mu &= E[X] = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) \\ &= p \\ E[X^2] &= 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) \\ &= p \end{aligned}$$

Thus,

$$\begin{aligned} \sigma^2 &= E[X^2] - (E[X])^2 \\ &= p - p^2 \\ &= p(1 - p) \\ &= pq \end{aligned}$$

Example 9.4

Suppose the player shoots three times. Let X be the number of times of making the shot. $P(X = 2) = ?$



In total

$$P(X = 2) = 3p^2q = \binom{3}{2}p^2q$$

Definition 9.5 (Binomial Distribution) — Given a Bernoulli trial, let X be the number of successes in n Bernoulli trials. Then X is called the binomial distribution and is denoted by

$$X \sim B(n, p) \quad \text{or} \quad X \sim \text{Binom}(n, p)$$

The pmf of X is given by

$$\begin{aligned} f(k) &= P(X = k), \quad k \in S_X = \{0, \dots, n\} \\ &= \binom{n}{k} p^k (1 - p)^{n-k} \end{aligned}$$

Explanation:

- choose k trials for success:

$$\# \text{ ways} = \binom{n}{k}$$

- for each choice, prob of success = $\underbrace{p \cdot p \dots p}_{k \text{ times}}$ and failures = $\underbrace{(1 - p) \dots (1 - p)}_{n-k}$.

$$\implies \binom{n}{k} p^k (1 - p)^{n-k}$$

Note: the pmf of $B(n, p)$ satisfies

$$\begin{aligned} \sum_{k=0}^n f(k) &= \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \\ &= (p + 1 - p)^n \quad \text{by Binomial Expansion Formula} \\ &= 1 \end{aligned}$$

Practice 9.2. 6 – Problem 2: mgf of $B(n, p)$:

$$\begin{aligned}
 E[e^{tX}] &= \sum_{k=0}^n e^{tk} P(X = k) \\
 &= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=0}^n \binom{n}{k} e^{tk} p^k (1-p)^{n-k} \\
 &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\
 &= (pe^t + 1 - p)^n \quad \text{by Binomial Expansion}
 \end{aligned}$$

Note that $n = 1$, $B(1, p)$ is simply a Bernoulli trial mgf if Bernoulli trial is given by

$$(pe^t + 1 - p)^1 = pe^t + 1 - p$$

Now, we can calculate the mean

$$\begin{aligned}
 \mu = E[X] &= \sum_{x \in S_X} x f(x) \\
 &= \underbrace{\sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}}_{\text{time consuming but doable}}
 \end{aligned}$$

MGF approach:

$$\begin{aligned}
 \mu = E[X] &= M'_X(t) \Big|_{t=0} \\
 M_X(t) &= (pe^t + 1 - p)^n \\
 M'_X(0) &= np
 \end{aligned}$$

Variance:

$$\begin{aligned}
 \sigma^2 &= E[X^2] - (E[X])^2 \\
 E[X^2] &= M''_X(0) \\
 M''_X(0) &= n(n-1)p^2 + np
 \end{aligned}$$

Thus,

$$\begin{aligned}
 \sigma^2 &= E[X^2] - (E[X])^2 \\
 &= n(n-1)p^2 + np - (np)^2 \\
 &= np(1-p)
 \end{aligned}$$

“Recalling variance of Bernoulli trial is $p(1-p)$.”

§10 | Lec 10: Oct 23, 2020

§10.1 Practice 6 Problem 3

Practice 10.1. 6 – Problem 3: $p = 0.95$

a) Let X be the number of days without an accident in next 7 days. Then $X \sim B(n = 7, p = 0.95)$.

$$\begin{aligned} P(X = 7) &= \binom{7}{7} .95^7 (1 - .95)^{7-7} \\ &= .95^7 \end{aligned}$$

b) Y = number of days in October without accident. $Y \sim B(n = 31, p = .95)$.

$$P(Y = 29) = \binom{31}{29} .95^{29} (.05)^2$$

c)

$A = \{\text{today, no accident}\}$

$B = \{\text{no accident from day 2 to day 5}\}$

$C = \{\text{at least one day with accident between day 6 to day 10}\}$

$C' = \{\text{no accident between day 6 and day 10}\}$

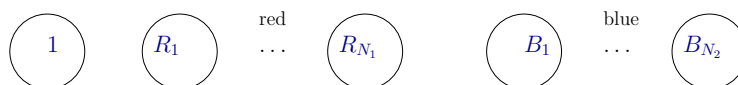
$P(B \cap C|A) = ?$ Note that A, B, C are mutually independent. Thus,

$$\begin{aligned} P(B \cap C|A) &= P(B \cap C) \\ &= \underbrace{P(B)}_{(n=4, p=0.95)} \underbrace{P(C)}_{(n=5, p=.95)} \\ &= \binom{4}{4} (.95)^4 (.05)^0 [1 - P(C')] \left[1 - \binom{5}{5} (.95)^5 (.05)^0 \right] \\ &= (.95)^4 [1 - (.95)^5] \end{aligned}$$

Remark 10.1. It might be helpful to consider complement when dealing with “at least” event.

§10.2 Hypergeometric Distribution

Practice 10.2. 7 – Problem 1: draw $n = x$ reds + $(n - x)$ blues



Denote $X = \#$ red balls from n drawn.

$$S_X = \begin{cases} x \in \mathbb{N} : 0 \leq x \leq n, \\ 0 \leq x \leq N_1, \\ 0 \leq n - x \leq N_2 \end{cases}$$

For $x \in S_X$, $P(X = x) = ?$

Ways to drawn n balls from $N_1 + N_2$: $\binom{N_1+N_2}{n}$

- E_1 = pick x reds from N_1 which is $\binom{N_1}{x}$
- E_2 = pick $n - x$ blues from $N_2 \implies \binom{N_2}{n-x}$
- $E_1 E_2$ = number of ways to pick n balls from $N_1 + N_2$ and pick exactly x red balls.
 $\implies \binom{N_1}{x} \binom{N_2}{n-x}$. Thus,

$$P(X = x) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N_1+N_2}{n}}$$

Note that X is denoted as $X \sim HG(N_1, N_2, n)$. The pmf indeed satisfies

$$\sum_{x \in S_X} f(x) = \sum_{x \in S_X} \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N_1+N_2}{n}} = 1$$

Fact 10.1. Let $X \sim HG(N_1, N_2, n)$ then

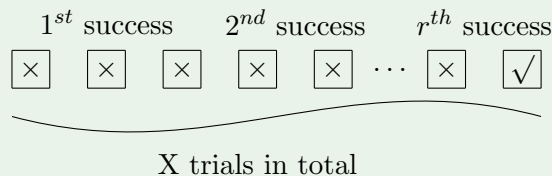
$$\mu = E[X] = n \frac{N_1}{N_1 + N_2}$$

Proof. See textbook 2.5. □

§11 | Lec 11: Oct 26, 2020

§11.1 Negative Binomial Distribution

Definition 11.1 (Negative Binomial Distribution) — Considering the experiment of performing Bernoulli trials until r successes occur (r is a fixed pos. integer). X = number needed to observe the r^{th} success. Then X is called a negative binomial distribution.



X is denoted as $X \sim NB(r, p)$

Remark 11.2. When $r = 1$, $X = \#$ needed to observe the first success ($\sim \text{Geom}(p)$)

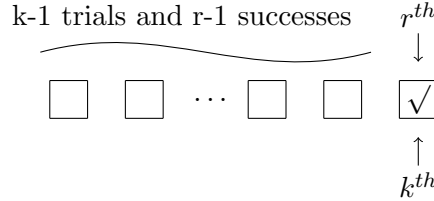
Fact 11.1. The pmf of $X \sim NB(r, p)$ is given by

for $k \geq r$,

$$f(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

where p is the probability of success (from Bernoulli trial). The space $S_X = \{r, r+1, \dots\}$.

Proof. Given $k \geq r$, $P(X = k) = ?$



$P(X = k) = P(\text{in the first } k-1 \text{ trials, there are exactly } r-1 \text{ successes})$

and the k^{th} trial is successful

$= P(r-1 \text{ successes from } k-1 \text{ trials}) \cdot P(k^{\text{th}} \text{ trial is successful})$

$$= \binom{k-1}{r-1} p^{r-1} (1-p)^{(k-1)-(r-1)} \cdot p$$

$$= \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

□

Note: The pmf of $NB(r, p)$ satisfies

$$\sum_{k=r}^{\infty} f(k) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = 1$$

We need Taylor expansion for the above formula, for $|w| < 1$,

$$\frac{1}{(1-w)^r} = \sum_{k=1}^{\infty} \binom{k-1}{r-1} w^{k-r}$$

So,

$$\begin{aligned} \sum_{k=r}^{\infty} f(k) &= p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (1-p)^{k-r} \\ &= p^r \frac{1}{(1-(1-p))^r} \\ &= 1 \end{aligned}$$

Fact 11.2. $X \sim NB(r, p)$ then

$$M_X(t) = \left[\frac{pe^t}{1 - (1-p)e^t} \right]^r$$

Mean:

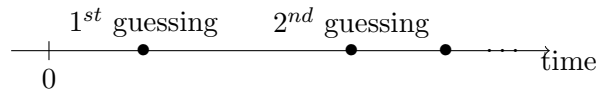
$$\mu = E[X] = \frac{r}{p}$$

Variance:

$$\sigma^2 = \text{Var}(X) = \frac{r(1-p)}{p^2}$$

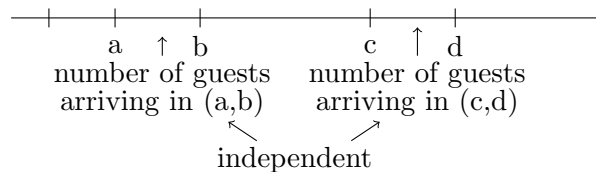
§11.2 Poisson Distribution

Motivation: Considering the arrivals (of guests at a bank or a restaurant, etc) in a continuous time interval

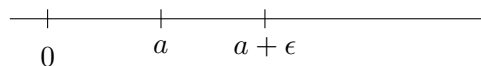


We assume the followings:

- The number of arrivals in non-overlapping intervals are mutually independent.



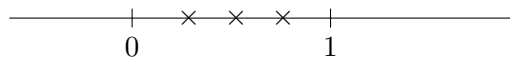
- There exists a fixed $\lambda > 0$ s.t. for all $\epsilon > 0$ efficiently small $P(\text{exactly one arrival in } [a, a + \epsilon]) = \lambda\epsilon$ and $P(\text{at least two arrivals in } [a, a + \epsilon]) = 0$



Note that we also have

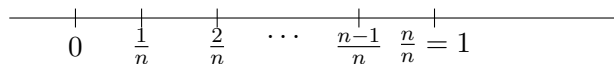
$$P(\text{no arrival in } [a, a + \epsilon]) = 1 - \lambda\epsilon$$

Question 11.1. $X = \#$ arrivals in one hour



$$P(X = k) = ?$$

Approach: for n large



By the second assumption,

$$P(\text{one arrival in one subinterval}) = \lambda \cdot \frac{1}{n} = \frac{\lambda}{n}$$

By the first assumption, subintervals arrivals are independent. Thus,

$$P(X = k) \cong P(k \text{ subintervals have one arrival each, among } n \text{ subintervals})$$

“ a subinterval having one arrival is a success with prob. $\frac{\lambda}{n}$ ” where

$$P(X = k) \cong \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Practice 11.1. 8 – Problem 1: For $k \geq 0$

$$S_n \sim B(n, \frac{\lambda}{n})$$

$$\lim_{n \rightarrow \infty} P(S_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Everything converges to 1 except $\frac{\alpha^k}{\lambda^k}$ and

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\alpha}{n}\right)^n = \lim_{y \rightarrow \infty} \left[\left(1 - \frac{1}{y}\right)^y \right]^\lambda$$

Notice that

$$\lim_{y \rightarrow \infty} \left[\left(1 - \frac{1}{y}\right)^y \right]^\lambda = (e^{-1})^\lambda = e^{-\lambda}$$

Hence,

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Definition 11.3 (Poisson Distribution) — Let X be a r.v. taking values in $\{0, 1, 2, \dots\}$ with pmf $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for a fixed $\lambda > 0$. Thus X is called a Poisson distribution, $X \sim \text{Pois}(\lambda)$

§12 | Lec 12: Oct 28, 2020

§12.1 Lec 11 (Cont'd)

Remark 12.1. The pmf of Poisson distribution satisfies

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

Practice 12.1. 8 – Problem 2: Calculate the MGF of $X \sim \text{Pois}(\lambda)$

$$\begin{aligned}
 M_X(t) &= E[e^{tX}] \\
 &= \sum_{k \geq 0} e^{tk} f(k) \\
 &= \sum_{k \geq 0} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} \frac{e^{t\lambda}}{k!} \\
 &= e^{-\lambda} \sum \frac{(e^t \lambda)^k}{k!} \\
 &= e^{\lambda(e^t - 1)}
 \end{aligned}$$

Note: $M_X(t)$ exists for all $t \in \mathbb{R}$.

Now,

$$\begin{aligned}
 \mu &= E[x] = M'_X(t) \Big|_{t=0} \\
 M'_X(t) &= \lambda e^t e^{\lambda(e^t - 1)} \\
 M'_X(0) &= \lambda
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 \sigma^2 &= E[x - \mu]^2 \\
 &= E[X^2] - \mu^2 \\
 &= M''_X(t) \Big|_{t=0} - \mu^2 \\
 &= \lambda
 \end{aligned}$$

Another approach:

$$\begin{aligned}
 M_X(t) &:= E[e^{tX}] \\
 &= E \left[1 + tX + \frac{t^2 X^2}{2!} + \dots \right] \\
 &= 1 + tM'_X(0) + \frac{t^2}{2} M''_X(0) + \dots
 \end{aligned}$$

Remark 12.2. $X \sim \text{Pois}(\lambda)$ “represents” the number of arrivals in one hour and $\mu = E[X] = \lambda$. Thus, on average, we expect to have λ arrivals in one hour.

Practice 12.2. 8 – Problem 3:

$$\begin{aligned}
 X &= \# \text{ goals scored in one game} \\
 S_X &= \{0, 1, 2, 3, \dots\}
 \end{aligned}$$

$X \sim \text{Pois}(\lambda)$ where α is TBD. Know: $P(X \geq 1) = \frac{1}{2}$, so what's $P(X = 3)$?

Find λ

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - e^{-\lambda} \frac{\lambda^0}{0!} \\ \frac{1}{2} &= 1 - e^{-\lambda} \\ \lambda &= \ln 2 \end{aligned}$$

$$\begin{aligned} P(X = 3) &= e^{-\lambda} \frac{\lambda^3}{3!} \\ &= \frac{1}{2} \frac{(\ln 2)^3}{3!} \end{aligned}$$

§12.2 Binomial Distribution Approximation by Poisson Distribution

Suppose $Y \sim B(n, p)$ where $p \ll n$. Then we can approximate Y by $X \sim \text{Pois}(\alpha = np)$, i.e.,

$$\begin{aligned} P(Y = k) &\cong e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-np} \frac{(np)^k}{k!} \end{aligned}$$

Example 12.3

Suppose $Y \sim \text{Binom}(n = 1000, p = .001)$, so $np = 1$.

$$P(Y \leq 2) \cong P(X \leq 2)$$

where $X \sim \text{Pois}(\lambda = np = 1)$

$$\begin{aligned} P(Y \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-1} \frac{1^0}{0!} + e^{-1} \frac{1}{1!} + e^{-1} \frac{1^2}{2!} \\ &= \frac{5}{2} e^{-1} \end{aligned}$$

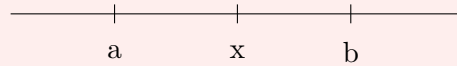
Remark 12.4. The “rule of thumb” is that $np \leq 1$. Alternatively, the following is also employed (in other textbooks)

$$np(1 - p) \leq 1$$

§12.3 Random Variable of Continuous Type

Example 12.5 (Motivation)

Let X denote the outcome of selecting a point randomly from the interval $[a, b]$ where $-\infty < a < b < \infty$

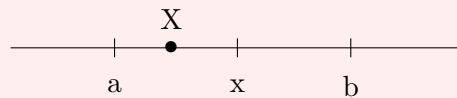


The prob. of X is selected from $[a, x]$ where $a < x < b$ is assigned as

$$P(a \leq X \leq x) = \frac{x - a}{b - a}$$

Similarly,

$$P(a \leq X \leq b) = \frac{b - a}{b - a} = 1$$

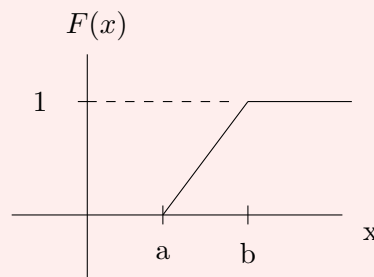


The cdf:

$$F(x) = P(X \leq x)$$

$$= \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

$$\begin{aligned} P(X \leq x) &= P(X < a) + P(a \leq X \leq x) \\ &= 0 + \frac{x - a}{b - a} \end{aligned}$$



Note that the cdf actually satisfies

$$F(x) = \int_{-\infty}^x f(y) dy$$

where

$$f(y) = \begin{cases} \frac{1}{b-a}, & a \leq y \leq b \\ 0, & \text{otherwise} \end{cases}$$

To see this

- $x < a$

$$\int_{-\infty}^x f(y)dy = \int_{-\infty}^x 0dy = 0 = F(x)$$

- $a \leq x \leq b$

$$\begin{aligned}\int_{-\infty}^x f(y)dy &= \int_{-\infty}^a f(y) + \int_a^x f(y)dy \\ &= 0 + \int_a^x \frac{1}{b-a} \\ &= \frac{x-a}{b-a} \\ &= F(x)\end{aligned}$$

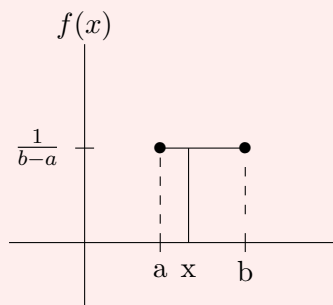
- $x > b$

$$\begin{aligned}\int_{-\infty}^x &= \int_{-\infty}^a + \int_a^b + \int_b^x f(y)dy \\ &= \int_a^b f(y)dy \\ &= \int_a^b \frac{1}{b-a} \\ &= 1\end{aligned}$$

Also, we have

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

$f(x)$ is called the “probability density function”.



§13 | Lec 13: Oct 30, 2020

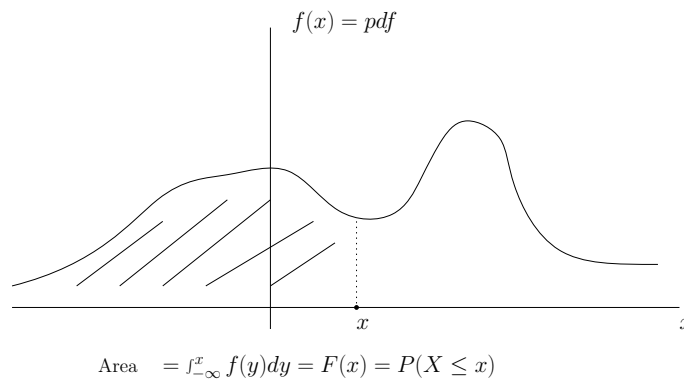
§13.1 Random Variable of Continuous Type (Cont'd)

Definition 13.1 (Probability Density Function) — The probability density function (pdf) of a continuous random variable X on a space S_X is an integrable function s.t. the followings hold:

- $f(x) \geq 0, x \in S_X$
- $\int_{-\infty}^{\infty} f(x) = 1$
- If $(a, b) \in S_X$, then $P(a < X < b) = \int_a^b f(x)dx$

The cumulative distribution function (cdf)

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^x f(y)dy \end{aligned}$$



Remark 13.2. 1. If X is a continuous RV with a pdf, $f(x)$, then

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X < b) \\ &= \int_a^b f(x)dx \end{aligned}$$

i.e., a continuous RV does NOT have point mass, which can be seen

$$P(X = a) = \int_a^a f(x)dx = 0$$

2. By calculus, the cdf $F(x)$ is a continuous function

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(y)dy \\ F'(x) &= f(x) \end{aligned}$$

Discrete RV	Continuous RV
pmf (mass func) $f(x) = P(X = x)$ $f(x) \geq 0, x \in S_X$ $\sum_{s \in S_X} f(x) = 1$ $P(X \in A) = \sum_{x \in A} f(x)$	pdf (density function): $f(x) \geq 0, x \in S_X$ $\int_{-\infty}^{\infty} f(x) dx = 1$ $P(a \leq X \leq b) = \int_a^b f(x) dx$
Cdf $F(x) = P(X \leq x)$ cumulative mass from the left up to and including x.	Cdf $F(x) = P(X \leq x)$ $= \int_{-\infty}^x f(x) dy$
Expectation: $E[u(X)] = \sum_{x \in S_X} u(x)f(x)$	Expectation: $E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x) dx$
$\mu = E[x]$ Mgf: $M_X(t) = \sum_{s \in S_X} e^{tx} f(x)$ $= \sum_{s \in S_X} x f(x)$	Mean: $\mu = E[x]$ $= \int_{-\infty}^{\infty} x f(x) dx$ Mgf: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$

Practice 13.1. 9 – Problem 1: $X \sim \text{Unif}(a, b)$ if X has the pdf

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

Mean:

$$\begin{aligned}
 \mu &= E[X] \\
 &= \int_{-\infty}^{\infty} x f(x) dx \\
 &= \int_{-\infty}^a + \int_a^b + \int_b^{\infty} x f(x) dx \\
 &= \int_a^b x f(x) dx \\
 &= \int_a^b x \frac{1}{b-a} dx \\
 \mu &= \frac{a+b}{2}
 \end{aligned}$$

$$\sigma^2 = E[X^2] - \mu^2$$

$$E[X^2] = \int_a^b x^2 f(x) dx$$

... Exercise

Mgf:

$$\begin{aligned}
 M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\
 &= \int_a^b e^{tx} \frac{1}{b-a} dx \\
 &= \frac{1}{b-a} \left. \frac{e^{tx}}{t} \right|_{x=a}^{x=b} \\
 &= \frac{1}{b-a} \frac{e^{tb} - e^{ta}}{t}
 \end{aligned}$$

Note that $M_X(t)$ is well-defined for all $t \in \mathbb{R}$

$$M_X(t) = \begin{cases} \frac{1}{b-a} \frac{e^{tb} - e^{ta}}{t}, & t \neq 0 \\ \int_{-\infty}^{\infty} e^{0 \cdot x} f(x) dx = 1, & t = 0 \end{cases}$$

Also,

$$\lim_{t \rightarrow 0} \frac{1}{b-a} \frac{e^{tb} - e^{ta}}{t} = 1$$

Practice 13.2. 9 – Problem 2: Need to verify 2 condition:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
 - $f_3(x)$ is not a pdf because $\sin x$ changes sign.
 - $f_1(x) \geq 0$, note that $S_X = [1, \infty)$

$$\begin{aligned}
 \int_{-\infty}^{\infty} f(x) dx &= \int_1^{\infty} \frac{1}{x^2} dx \\
 &= 1
 \end{aligned}$$

$f_1(x)$ is a pdf.

- $f_2(x)$: If $b \leq 0$ then f_2 is NOT a pdf. If $b > 0$, then we have to find a, b s.t. $\int_{-a}^a f(x) dx = 1$.

$$\int_{-a}^a b \sqrt{a^2 - x^2} dx = b \int_{-a}^a \sqrt{a^2 - x^2}$$

Thus,

$$\int_{-a}^a b \sqrt{a^2 - x^2} dx = 1 = b \cdot \frac{\pi a^2}{2}$$

implying

$$a^2 b = \frac{2}{\pi}$$

Definition 13.3 (Percentile) — Given $p \in [0, 1]$, the $100.p^{\text{th}}$ percentile is a number π_p s.t.

$$F(\pi_p) = \int_{-\infty}^{\pi_p} f(x)dx = p$$

$p = \frac{1}{2}$, = 50th percentile, $\pi_{0.5}$ is called the median

$$F(\pi_{0.5}) = P(X \leq \pi_{0.5}) = \frac{1}{2}$$

$p = \frac{1}{4}$, $\pi_{0.25}$ = 25th percentile is called the first quartile

$$F(\pi_{0.25}) = P(X \leq \pi_{0.25}) = \frac{1}{4}$$

§14 | Midterm 1: Nov 2, 2020

NO CLASS :D

§15 | Dis 1: Oct 6, 2020

§15.1 Set Theory

Definition 15.1 (Set) — A set is a collection of items.

Example 15.2

$$T = \{1, 2, 3, \text{red}, \text{blue}\}$$

$$S = \{1, 3, \text{red}\}$$

$$R = \{1, 2, 4\}$$

$$S \subseteq T$$

$$S' = S^c = \{2, \text{blue}\}$$

$$R \not\subseteq T$$

$$3 \underbrace{\in}_{\text{is an element of}} T$$

$$\{3\} \subseteq T$$

Example 15.3

$$\begin{aligned}
 A &= \{1, 3, 7\} & A \cup B &= \{1, 2, 3, 4, 7\} \\
 B &= \{2, 3, 4\} & A \cap B &= \{3\} \\
 A \setminus B &= \{1, 7\} & B \setminus A &= \{2, 4\}
 \end{aligned}$$

De Morgan Laws:

$$\begin{aligned}
 (A \cup B)' &= A' \cap B' \\
 (A_1 \cup A_2 \cup \dots \cup A_n)' &= A_1' \cap A_2' \cap \dots \cap A_n' \\
 (A \cap B)' &= A' \cup B'
 \end{aligned}$$

If have a sample space S , and subset of S are called events. A probability function is a function \mathbb{P} that assigns a real number each event with three rules:

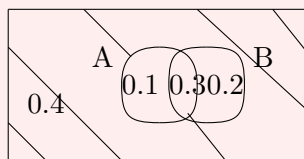
1. $P(A) \geq 0$
2. $P(S) = 1$
3. A_1, A_2, \dots, A_n with $A_i \cap A_j = \emptyset = \{\}$, then $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$

Example 15.4

1.1-6 (from the book): $P(A) = 0.4$, $P(B) = 0.5$, $P(A \cap B) = 0.3$

Find

- $P(A \cup B) = .1 + .3 + .2 = .6$
- $P(A \cap B)' = .1$
- $P(A' \cap B) = .2$



Note: (P, S) : probability space on all subsets of S

Example 15.5

1.2-5: How many four letter codes can be made from the letters in IOWA if

- Letters may not be repeat: $4! = 24$ ways.
- Letters may repeat: $4^4 = 256$ ways.

§16 | Dis 2: Oct 13, 2020

1.4.16: An urn has 5 balls. One is marked “win” and the other are marked “lose”. You and another player each take balls out one at a time until somebody picks win. You pick first.

W/o replacement: $P(\text{winning}) = \frac{1}{5} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{3}{5}$

With replacement:

$$\begin{aligned} P(\text{winning}) &= \frac{1}{5} + \frac{4}{5} \cdot \frac{4}{5} \cdot \frac{1}{5} + \frac{4}{5} \cdot \frac{4}{5} \cdot \frac{4}{5} \cdot \frac{1}{5} + \dots \\ &= \frac{\frac{1}{5}}{1 - \frac{16}{25}} = \frac{5}{9} \end{aligned}$$

§16.1 Conditional Probabilities

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

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

Example 16.1

1.3.7: An urn has 4 balls. 2 are red and 2 are blue. We pull out 2 balls. We are told that at least one is red. What’s the probability that they’re both red?

$$P(\text{both red} | \text{at least one red}) = \frac{P(\text{both red and at least one red})}{P(\text{at least one red})} = \frac{P(\text{both red})}{P(\text{at least red})} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}.$$

§16.2 Bayes’s Theorem

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

Example 16.2

1.5-8: Four types of tablets: B_1, B_2, B_3, B_4 with % of sales 0.4, 0.3, 0.2, 0.1 and % tablet needs repair 0.1, 0.05, 0.03, 0.02, respectively. What is the probability that a

tablet needing repair is B_i ?

$$\begin{aligned}
 P(B_1|\text{need repair}) &= \frac{P(\text{need repair}|B_1) \cdot P(B_1)}{P(\text{need repair})} \\
 &= \frac{(0.1)(0.4)}{(0.40)(0.10) + (0.30)(0.05) + (0.20)(0.03) + (0.10)(0.02)} \\
 &\approx 63.5\% \\
 P(B_2|\text{need repair}) &= \frac{(0.30)(0.05)}{0.063} \approx 23.8\% \\
 P(B_3|\text{need repair}) &\approx 9.5\% \\
 P(B_4|\text{need repair}) &\approx 3.2\%
 \end{aligned}$$

§17 | Dis 3: Oct 20,2020

§17.1 Recap of Terminology/Functions

We have a situation with a set of possible outcomes

- This set is called the **sample space** denoted S or Ω .
- Elements of S are called **outcomes**.
- Subsets of S are called **events**.
- A **probability function** is a function where

$$P : \{ \text{subset of } S \} \rightarrow [0, 1]$$

Example 17.1

$$\begin{aligned}
 S &= \{HH, HT, TH, TT\} \\
 A &= \{HH, HT\} \\
 B &= \{HH\} \\
 P(A) &= 0.5 \\
 P(B) &= P(\{HH\}) = 0.25
 \end{aligned}$$

A **random variable**, denoted X , is a function

$$X : \underbrace{S}_{\text{sample space}} \rightarrow \underbrace{S_X}_{\text{the space support}} \subseteq \mathbb{R}$$

$$"X = a" \leftrightarrow \{w \in S \text{ s.t. } X(w) = a\}.$$

Example 17.2

Define $X(w)$ to be the number of tails in the outcome w .

$$X(HH) = 0$$

$$X(HT) = 1$$

$$X(TH) = 1$$

$$X(TT) = 2$$

$$(X = 1) = \{HT, TH\}$$

$$(X = 2) = \{TT\}$$

$$(X = 0) = \{HH\}$$

$$(X = 3) = \emptyset$$

The **probability mass function** or pmf of a r.v. X is a function $f_x : S_X \rightarrow [0, 1]$ defined by

$$f(x) = P(X = x)$$

$$f(a) = P(X = a)$$

Example 17.3

$$f_x(a) = \begin{cases} 0.25 & a = 0 \\ 0.5 & a = 1 \\ 0.25 & a = 2 \end{cases}$$

Also,

$$P(X = 1) = P(\{HT, TH\}) = 0.5$$

$$P(X < 2) = P(\{HH, HT, TH\}) = 0.75$$

The **cumulative distribution function** or cdf of a r.v. X is a function $F_x : S_x \rightarrow [0, 1]$ defined by

$$F(a) = P(X \leq a)$$

Example 17.4

$$F_x(a) = \begin{cases} 0.25 & a = 0 \\ 0.75 & a = 1 \\ 1 & a = 2 \end{cases}$$

The **expectation** or **mean** of X is

$$E[x] = \sum_{a \in S_x} a f(a)$$

$$E[g(x)] = \sum_{a \in S_x} g(a) f(a)$$

Example 17.5 (above)

$$E[x] = (0)0.25 + (1)0.5 + (2)0.25$$

$$= 1$$

$$E[x^2] = (0^2)0.25 + (1^2)0.5 + (2^2)0.25$$

$$= 2.5 \neq E[x]^2$$

The **moment of generating function** or mgf of X is

$$M_x(t) = E[e^{tX}] = \sum_{a \in S_x} e^{ta} f(a)$$

Example 17.6

$M(t) = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t} = \sum_{a \in \{1,2,3\}} e^{at} f(a)$. Find mean, variance, pmf.
 $S_x = \{1, 2, 3\}$. The pmf is

$$f_x(a) = \begin{cases} \frac{2}{5} & a = 1 \\ \frac{1}{5} & a = 2 \\ \frac{2}{5} & a = 3 \end{cases}$$

The mean is

$$E[x] = (1)\frac{2}{5} + (2)\frac{1}{5} + (3)\frac{2}{5} = 2$$

Variance is

$$\sigma^2 = \text{Var}(X) = E[x^2] - E[x]^2$$

$$= \left((1^2)\frac{2}{5} + (2^2)\frac{1}{5} + (3^2)\frac{2}{5} \right) - 2^2$$

$$= \frac{4}{5}$$

§18 | Dis 4: Oct 27, 2020

First half of chapter 2: Concepts relating discrete RVs X

- $E[X]$
- pmf, cdf, mgf
- moments
- plots

§18.1 Review of Chapter 2

Example 18.1 (Binomial)

Test w/ 100 multiple choice questions (A,B,C,D) and you guess on every question. $X = \#$ correct answers. $X \sim b(100, 0.25)$. What is the prob. of :

1. Getting exactly 25 right? ($f(25) = P(X = 25)$)

$$f(25) = P(X = 25) = \binom{100}{25} \left(\frac{1}{4}\right)^{25} \left(\frac{3}{4}\right)^{75}$$

2. Getting at least 25 right?

$$\begin{aligned} P(X \geq 25) &= \sum_{k=25}^{100} \binom{100}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{100-k} \\ &= 1 - \sum_{k=0}^{24} \binom{100}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{100-k} \end{aligned}$$

Example 18.2 (above – Negative Binomial)

What's the probability it takes us 50 questions to get 10 right?

$Y = \#$ of questions until we get 10 right.

$$P(Y = 50) = \binom{49}{9} \left(\frac{1}{4}\right)^{10} \left(\frac{3}{4}\right)^{40}$$

Example 18.3 (Hypergeometric)

50 objects, 2 of which are special. If we pick 5 of random, what's the probability:

- none are special: $\frac{\binom{48}{5}}{\binom{50}{5}} = P(X = 0)$
- one is special: $\frac{\binom{2}{1} \binom{48}{4}}{\binom{50}{5}} = P(X = 1)$
- two are special: $\frac{\binom{2}{2} \binom{48}{3}}{\binom{50}{5}} = P(X = 2)$

Poisson:

sort of a “continuous version” of Bernoulli trials.

Table X Discrete Distributions					
Probability Distribution and Parameter Values	Probability Mass Function	Moment-Generating Function	Mean $E(X)$	Variance $\text{Var}(X)$	Examples
Bernoulli $0 < p < 1$ $q = 1 - p$	$p^x q^{1-x}, x = 0, 1$	$q + pe^t, -\infty < t < \infty$	p	pq	Experiment with two possible outcomes, say success and failure, $p = P(\text{success})$
Binomial $n = 1, 2, 3, \dots$ $0 < p < 1$	$\binom{n}{x} p^x q^{n-x}, x = 0, 1, \dots, n$	$(q + pe^t)^n, -\infty < t < \infty$	np	npq	Number of successes in a sequence of n Bernoulli trials, $p = P(\text{success})$
Geometric $0 < p < 1$ $q = 1 - p$	$q^{x-1} p, x = 1, 2, \dots$	$\frac{pe^t}{1 - qe^t}, t < -\ln(1 - p)$	$\frac{1}{p}$	$\frac{q}{p^2}$	The number of trials to obtain the first success in a sequence of Bernoulli trials
Hypergeometric $x \leq n, x \leq N_1$ $n - x \leq N_2$ $N = N_1 + N_2$ $N_1 > 0, N_2 > 0$	$\frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}}$		$n \frac{N_1}{N}$	$n \frac{N_1}{N} \frac{N_2}{N} \frac{N-n}{N-1}$	Selecting n objects at random without replacement from a set composed of two types of objects
Negative Binomial $r = 1, 2, 3, \dots$ $0 < p < 1$	$\binom{x-1}{r-1} p^r q^{x-r}, x = r, r+1, \dots$	$\frac{(pe^t)^r}{(1 - qe^t)^r}, t < -\ln(1 - p)$	$\frac{r}{p}$	$\frac{rq}{p^2}$	The number of trials to obtain the r th success in a sequence of Bernoulli trials
Poisson $\lambda > 0$	$\frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, \dots$	$e^{\lambda(e^t - 1)}, -\infty < t < \infty$	λ	λ	Number of events occurring in a unit interval, events are occurring randomly at a mean rate of λ per unit interval
Uniform $m > 0$	$\frac{1}{m}, x = 1, 2, \dots, m$		$\frac{m+1}{2}$	$\frac{m^2 - 1}{12}$	Select an integer randomly from $1, 2, \dots, m$

Figure 1: A Summary of Chapter 2

Example 18.4 (Poisson)

People entering a store. We expect to see one person per 10 minutes. One hour passes, What's the prob. of:

Let $X = \#$ of people we see in the hour – $X \sim \text{Poisson}(6)$

- Seeing exactly 5 people: $P(X = 5) = \frac{6^5 e^{-6}}{5!}$
- At most two people: $P(X \leq 2) = \frac{6^0 e^{-6}}{0!} + \frac{6^1 e^{-6}}{1!} + \frac{6^2 e^{-6}}{2!}$

Example 18.5

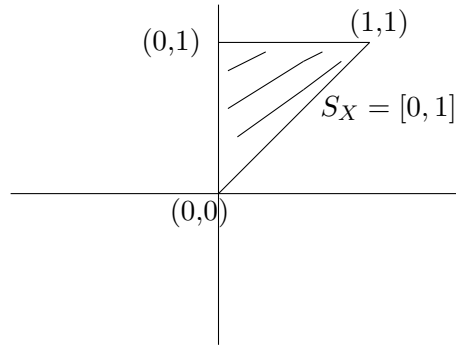
You have 0.001 chance of winning lottery. If you play 2000 times, what's the prob. you win at least once?

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - \frac{2^0 e^{-2}}{0!} = 1 - \frac{1}{e^2}$$

§19 | Dis 5: Nov 3, 2020

§19.1 Continuous Random Variables

$X : S \rightarrow S_X \subseteq \mathbb{R}$. Example X is the x-coordinate of a point randomly chosen from



To a continuous random variable, we can associate:

- Cumulative Distribution Function (cdf)

$$F_X(a) = P(X \leq a) = \int_{-\infty}^a f_X(t) dt$$

- Probability density function (pdf)

$$f_X(a) = F'(a)$$

when F is differentiable. Note that

$$f_X(a) \neq P(X = a)$$

$$\int_{-\infty}^{\infty} f_X(t) dt = 1 \iff \sum_{a \in S_X} f_X(a) = 1$$

- Moment generating function (mgf)

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

- Expectation:

$$E[X] = \int_{-\infty}^{\infty} t f(t) dt$$

- Percentiles: p^{th} percentile = π_p

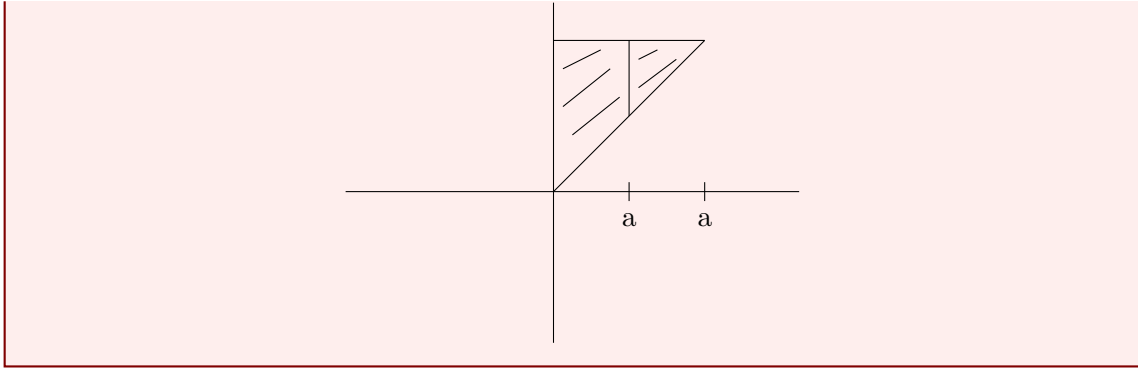
$$\frac{p}{100} = \int_{-\infty}^{\pi_p} f(t) dt$$

Example 19.1 • Cdf

$$F(a) = \frac{\int_0^a 1 - x dx}{\frac{1}{2}} = 2a - a^2$$

- pdf

$$f(a) = 2 - 2a$$



Problem 19.1. 3.1 – 10: cdf of X is $f(x) = \frac{c}{x^2}, 1 < x < \infty$

a. Find c

$1 = \int_{-\infty}^{\infty} f(x)dx$, so

$$\begin{aligned} 1 &= \int_1^{\infty} \frac{c}{x^2} dx \\ &= \lim_{b \rightarrow \infty} \int_1^b \frac{c}{x^2} dx \\ &= \lim_{b \rightarrow \infty} \left(\frac{-c}{b} + \frac{c}{1} \right) \\ &= c \end{aligned}$$

b. $E[x] = \infty$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} t f(t) dt \\ &= \int_1^{\infty} t \frac{1}{t^2} dt \\ &= \int_1^{\infty} \frac{1}{t} dt \\ &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{t} dt \\ &= \lim_{b \rightarrow \infty} \log(b) \\ &= \infty \end{aligned}$$

§19.2 Types of Distribution for Continuous Variable

Uniform: $X \sim \text{Uniform}(0,17)$. Find the probability $P(X^2 - 4 \leq 5)$

$$\begin{aligned} x^2 - 4 &\leq 5 \\ x &\leq 3 \end{aligned}$$

$$\Rightarrow P = \frac{3}{17}$$

Exponential:

Example 19.2

A store opens at 8am. On average, it gets one customer every 10 minutes.

$X = \#$ people that arrive in first hour – $X \sim \text{Poisson}(6)$

$Y =$ time (in minutes) of the first arrival – $Y \sim \text{Exp}(10)$ or $Y \sim \text{Exp}(\frac{1}{10})$ (just notation convention)

$$f_Y(t) = \frac{1}{10}e^{-\frac{t}{10}}, \quad t \geq 0$$

What's the probability no one comes in during the first 30 minutes?

$$\begin{aligned} P(Y > 30) &= 1 - P(Y \leq 30) \\ &= 1 - \int_0^{30} f_Y(t) \\ &= 1 - \int_0^{30} \frac{1}{10}e^{-\frac{t}{10}} dt \\ &= 1 - (-e^{-3} + 1) \\ &= e^{-3} \approx 5\% \end{aligned}$$

Gamma Distribution:

$Z =$ time (in minutes) until the fifth person arrives – $Z \sim \text{Gamma}(10,5)$ or $Z \sim \text{Gamma}(\frac{1}{10}, 5)$. Note that, in this case,

$$\Gamma(a) = (a-1)!$$

$$f_Z(a) = \frac{a^4 e^{-\frac{a}{10}}}{\Gamma(5)10^5}$$

What is the probability that there are at least five people in the first hour?

$$\begin{aligned} P(Z \leq 60) &= \int_0^{60} \frac{x^4 e^{-\frac{x}{10}}}{\Gamma(5)10^5} dx \\ &= \frac{1}{2400000} \int_0^{60} x^4 e^{-\frac{x}{10}} \\ &= \frac{1}{2400000} \left(240000 \left(1 - \frac{115}{e^6} \right) \right) \\ &= 1 - \frac{115}{e^6} \approx 71\% \end{aligned}$$