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 Interference* 10^{th} 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

List of Definitions

§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 $A_1, A_2, \ldots \subseteq S$ satisfy $A_i \cap A_j = \emptyset$, $i \neq j$ then they are called "disjoint" (mutually exclusive)
- If $A_1, A_2, \ldots, A_n \subseteq$ satisfy $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \ldots \cup A_n = S$. Then $\{A_i\}_{i=1\ldots n}$ 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 $n \to \infty$,

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

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.

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

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

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

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

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 $A \subseteq S$, is a number satisfying:

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

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

More generally, if $A_1, \ldots, A_n, \ldots \subseteq S$ 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. If $A \leq B$ then $P(A) \leq P(B)$

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

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

<u>*Note*</u>: The pattern here is add the prob. of odd event(s) and substract 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.

$$A \cup (B \setminus A) = B$$
$$A \cap (B \setminus A) = \emptyset$$
$$\text{something here}$$

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 $S = \{e_1, \ldots, e_m\}$ 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 2^{nd} experiment E_2 has n_2 outcomes. Then the composite E_1E_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 "$n! = 1 \cdot 2 \cdot 3 \cdot \ldots n$"

Proof.

- E_1 = fill the 1^{st} position from n persons $\implies$ n outcomes for E_1.
- E_2 = fill the 2^{nd} 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 $= E_1 E_2 \ldots E_n$

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.

<u>*Note*</u>:"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 ${}^nP_k$ is given by
$$ {}^nP_k = \frac{n!}{(n-k)!} $$
2. The total numbers of combination of n taken k, denoted by ${}^nC_k$ or $\binom{n}{k}$ is given by
$$ {}^nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!} $$

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

$\vdots$

E_k = fill k^{th} pos. from $n-k+1$ persons left. Thus,
$$ permk = n \cdot \ldots \cdot (n-k+1) $$
(2) Combination of n taken k :
Start with ${}_nP_k$ as follow:

- E_1 = take k from n at once, outcome $=_n C_k = \binom{n}{k}$
- E_2 = permute k, outcomes $= k!$. Thus,
$$ {}^nP_k = \binom{n}{k} \cdot k! $$
  implying
$$ \binom{n}{k} = \frac{{}^nP_k}{k!} = \frac{n!}{(n-k)!k!} $$
□

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 1^{st} 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 3^{rd} 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"?

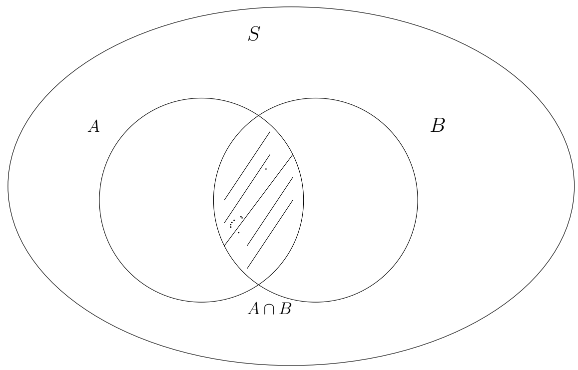
1. Total = exactly "one even" + "two even" + "three even" + "four even"
2. Total = "4-element subset" - "4-element subset with no even digit"

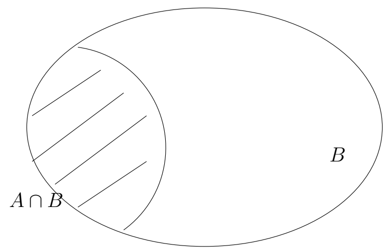
§3 Lec 3: Oct 7, 2020

§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?

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

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

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

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

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

<u>*Note*</u>: 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 $A_1, \ldots, A_n \ldots$ are disjoint events,
  $$P(\bigcup_{k=1}^{\infty} A_k|B) = \sum_{k=1}^{\infty} P(A_k|B)$$

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

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

□

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

□

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 in cluding 13 hearts} $B|A =$ { 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

$$\begin{aligned} S &= \{\text{ HH, HT , TH, TT}\} \\ A &= \left\{1^{\text{st}H}\right\} \\ B &= \left\{2^{\text{nd}}T\right\} \\ C &= \{\text{TT}\} \end{aligned}$$

$C \subseteq B$ "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 independents iff

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

Theorem 4.3

The following are equivalent

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

Proof. Left as an excercise. □

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

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

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

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

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 $A_1, \ldots, A_n, \ldots$ 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 $B_1, \ldots, B_n$ (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: $\bigcup_n^{i=1} B_i = S$

then,

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

Theorem 5.4 (Law of total Probability)
Suppose $B_1, \ldots, B_n$ is a partition of S with $P(B_i) > 0$ for $i = 1, \ldots, n$. 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)

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

□

Practice 5.2. 3 – problem 1:

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

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

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

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

Proof.

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

□

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

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

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

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

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

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

§6.1 Random Variables with Discrete Type

Example 6.1

Flip a coin

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

Define

$$X : S \to \mathbb{R}$$
$$\triangle \mapsto X(s) \in \mathbb{R}$$

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

$$\text{H} \xrightarrow{X} 0$$
$$\text{T} \longrightarrow 1$$

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,

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

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

$$S = \{1, 2, \dots, 6\}$$
$$X : S \to \mathbb{R}$$
$$s \mapsto X(s) = x$$
$$S_x = \{1, 2, \dots, 6\} \, (= S)$$

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:

- $f(x) > 0, \quad x \in S_x$.
- $\sum_{x \in S_x} f(x) = 1$.
- 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 $k = 1, \dots, 6$

$$A = \{1, 2, 3\} = \text{“}X < 4\text{”}$$
$$A \subseteq S_x$$
$$P(X \in A) = \sum_{k \in A} f(k) = \sum_{k=1}^{3} \frac{1}{6} = \frac{1}{2}$$

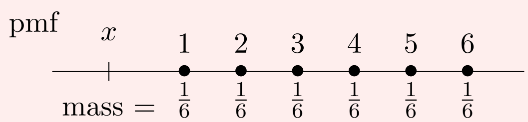
Definition 6.7 (Cumulative Distribution Function) — Cumulative distribution function (cdf) $F(x)$ of a RV x is a function given by

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

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

Example 6.8

Rolling a fair dice



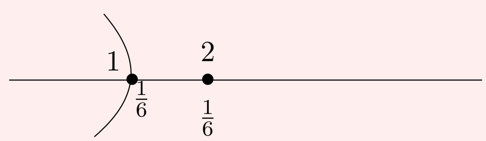
$$\begin{aligned}\text{cdf } F(x) &= P(X \leq x) \\ &= \text{total mass cumulated starting from the left up to x}\end{aligned}$$

$x < 1$,

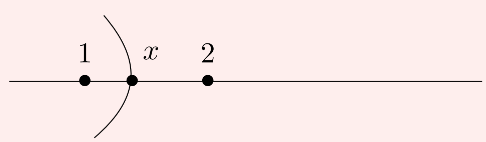
$$\begin{aligned}F(x) &= P(X \leq x) \\ &= 0 \text{ (no mass up to } x < 1)\end{aligned}$$

$x = 1$,

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

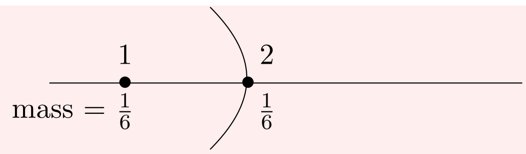


$F(1) = \frac{1}{6}$ (mass up to and including location 1).
$1 < x < 2$



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

$x = 2$



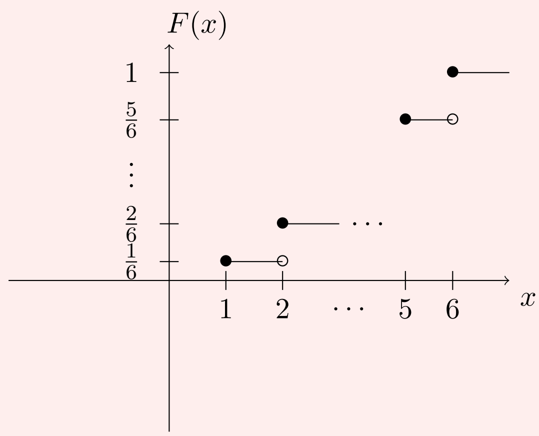
$$\begin{aligned} F(2) &= P(X \leq 2) \\ &= P(X = 1) + P(X = 2) \\ &= \frac{2}{6} \end{aligned}$$

Likewise, $2 < x < 3$

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

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

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

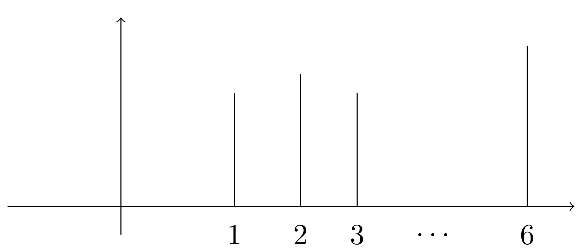


§7 Lec 7: Oct 16, 2020

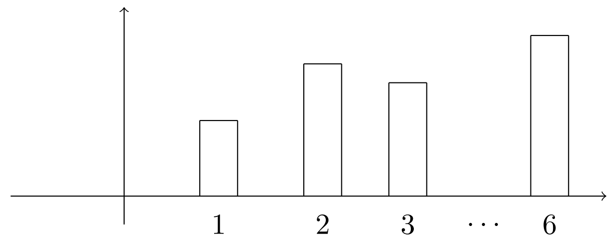
§7.1 Lec 6 (Cont'd)

In order to graph the prob. mass function:

- Line graph



- Histogram



Practice 7.1. 4 – Problem 1:

$$X = \text{ max of two rolls}$$
$$S_X = \{1, 2, \ldots, 6\}$$

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

- 1^{st} approach:

1^{st} roll \ 2^{nd} roll	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	$(6,1)$	$(6,2)$	$\cdots$			

1^{st} roll \ 2^{nd} roll	1	2	3	$\cdots$	6
1	1	2	3	$\cdots$	6
2	2	2	3	$\cdots$	6
3	3	3	3	$\cdots$	6
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
6	6	6	6	$\cdots$	6

$$f(1) = P(X = 1) = \frac{1}{36}$$
$$f(2) = P(X = 2) = \frac{3}{36}$$
$$f(3) = P(X = 3) = \frac{5}{36}$$
$$\vdots$$
$$f(6) = P(X = 6) = \frac{11}{36}$$

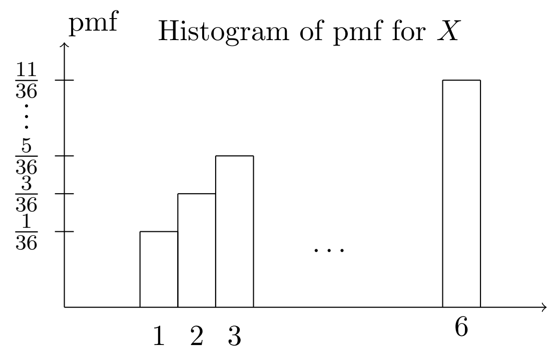
- 2^{nd} approach: for $k = 1, \ldots, 6$ (disjoint sub-events)

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

Thus,

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

<u>*Note*</u>: $\sum_{k=1}^{6} \frac{2k-1}{36} = 1$.



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

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

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.
$$E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)]$$

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

$$E[c_1u_1(X) + c_2u_2(Y)] = c_1E[u_1(X)] + c_2E[u_2(Y)]$$

Proof. Textbook. □

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

- the mean (of X) is denoted by

$$u := E[x]$$

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

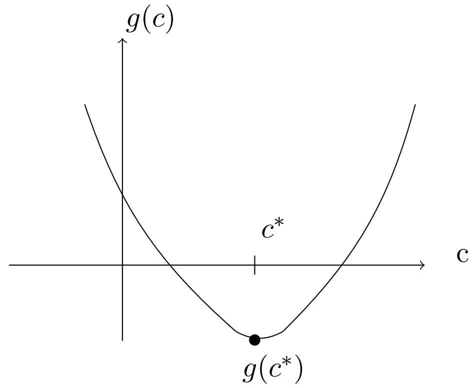
$$\begin{aligned}
\text{mean} = \mu &= E[x] \\
&= \sum_{x \in S_X} x \cdot f(x) \\
&= (-2)\frac{1}{2} + 0\frac{1}{3}1\frac{1}{6} \\
&= -\frac{5}{6} \\
\text{variance} = \sigma^2 &= E[(x-\mu)^2] \\
&= \sum_{x \in S_X} (x-\mu)^2 f(x) \\
&= (-2-(-\frac{5}{6})^2\frac{1}{2} + (0-(-\frac{5}{6}))^2\frac{1}{3} + \dots
\end{aligned}$$

σ^2 intepretation:

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

$$\begin{aligned}
g(c) &= E[(X-c)^2] \\
&= E[X^2 - 2cX + c^2] \\
&= E[X^2] + E[-2cX] + E[c^2] \\
&= E[X^2] - 2cE[X] + c^2 \\
&= c^2 - 2cE[X] \cdot + E[X^2] \\
&= c^2 - 2\mu \cdot c + E[X^2]
\end{aligned}$$

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



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

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

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.,

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

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

§8 Lec 8: Oct 19, 2020

§8.1 Info about 1^{st} midterm

1^{st} Midterm 11/2, Monday, 10am PT. Due: 10am PT – Tuesday 11/3.
2^{nd} Midterm, after Thanksgiving.

§8.2 Lec 7 (Cont'd)

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

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

Differentiating both sides,

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

Practice 8.1. 5 – Problem 2:

$S_X = \{1, 2, \ldots\}$. The pmf $f(f) = P(X = k) = P(1^{\text{st}}$ k-1 shots are missed and k shot succcessful.

a) $E[X] = ?$

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

$$P(X = k) = P\left(\underbrace{A_1' \cap A_2' \cap \ldots \cap A_{k-1}'}_{\text{miss}1^{\text{st}}\ k-1 \text{ shots}} \cap \underbrace{A_k}_{\text{make at} k^{\text{th}} \text{ shots}}\right)$$
$$\overset{\text{independence}}{=} P(A_1')P(A_2')\ldots P(A_{k-1}')P(A_k)$$
$$= q \cdot q \ldots q \cdot p$$
$$= q^{k-1} \cdot p$$

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

$$\sum_{k=1}^{\infty} f(k) = \sum_{k=1}^{\infty} q^{k-1} \cdot p$$
$$= p\left(1 + q + q^2 + \ldots\right)$$
$$= p \cdot \frac{1}{1-q}$$
$$= p \cdot \frac{1}{p}$$
$$= 1$$

Now,

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

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 $k = 1, 2, \ldots, \quad q = 1 - p$. MGF of X is given by

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

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

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

  Thus,

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

- Similarly, 2$^{\text{nd}}$ derivative of $M_X(t)$ given by

$$M_X''(t) = E\left[X^2 e^{tX}\right]$$

$$M_X''(t)\Big|_{t=0} = E[x^2], \text{ second moment of X}$$

- 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

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

which can also be rewritten as:

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

<u>*Note*</u>: A formula of variance

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

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

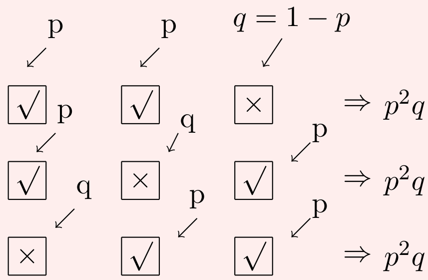
$$\begin{aligned}
\mu = E[X] &= 1 \cdot P(X=1) + 0 \cdot P(X=0) \\
&= p \\
E[X^2] &= 1^2 \cdot P(X=1) + 0^2 \cdot P(X=0) \\
&= p
\end{aligned}$$

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 $= \underbrace{(1-p)\dots(1-p)}_{n-k}$.

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

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 $B(n,p)$:

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

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

$$\left(pe^t + 1 - p\right)^1 = pe^t + 1 - p$$

Now, we can calculate the mean

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

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)

$$\begin{aligned} A &= \{\text{today, no accident}\} \\ B &= \{\text{no accident from day 2 to day 5}\} \\ C &= \{\text{at least one day with accident between day 6 to day 10}\} \\ C' &= \{\text{no accident between day 6 and day 10}\} \end{aligned}$$

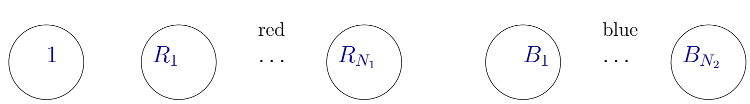
$P(B \cap C|A) =?$ Note that A, B, C are mutually independent. Thus,

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

> **Remark 10.1.** It might be helpful to consider compliment 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.

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

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 $N_2 \implies \binom{N_2}{n-x}$
- $E_1E_2 =$ 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

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

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

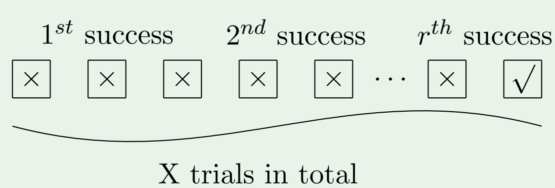
$$\mu = E[X] = n\frac{N_1}{N_1 + N_2}$$

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 $X \sim NB(r, p)$

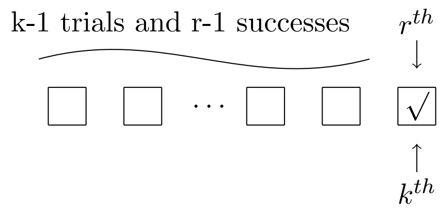
> **Remark 11.2.** When $r = 1$, $X = \#$ needed to observe the first success ($\sim$ Geom (p))

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 $S_X = \{r, r+1, \ldots\}$.

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



$$\begin{aligned}
P(X = k) &= P(\text{in the first k-1 trials, there are exactly r-1 successes}) \\
&\quad \text{and the } k^{\text{th}} \text{ trial is successful} \\
&= P(r-1 \text{ successes from k-1 trials}) \cdot P(k^{\text{th}} \text{ trial is successful}) \\
&= \binom{k-1}{r-1} p^{r-1} (1-p)^{(k-1)-(r-1)} \cdot p \\
&= \binom{k-1}{r-1} p^r (1-p)^{k-r}
\end{aligned}$$

□

<u>*Note*</u>: 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:

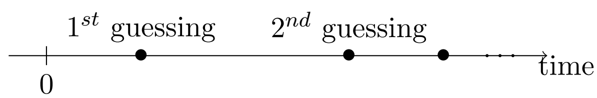
$$\mu = E[X] = \frac{r}{p}$$

Variance:

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

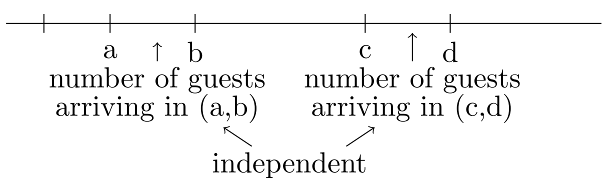
§11.2 Poisson Distribution

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

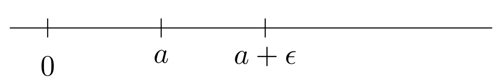


We assume the followings:

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



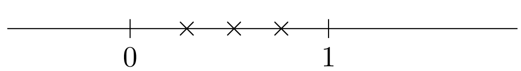
- There exists a fixed $\lambda > 0$ s.t. for all $\epsilon > 0$ efficiently small $P($exactly one arrival in $[a, a+\epsilon]) = \lambda\epsilon$ and $P($at least two arrivals in $[a, a+\epsilon)) = 0$



  Note that we also have

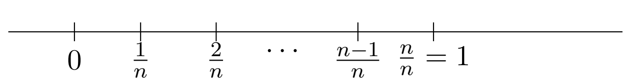
$$P(\text{no arrival in } [a, a+\epsilon)) = 1 - \lambda\epsilon$$

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,

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

" 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 $k \geq 0$

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

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

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

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

Notice that

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

Hence,

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

Definition 11.3 (Poisson Distribution) — Let X be a r.v. taking values in $\{0, 1, 2 \ldots\}$ 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 $X \sim \text{Pois}(A)$

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

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:

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

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

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

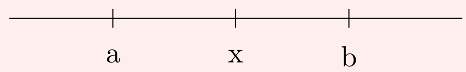
Remark 12.4. The "rule of thump" is that $np \leq 1$. Alternatively, the following is also employed (in other textbooks)

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

§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 $-\infty < a < b < \infty$

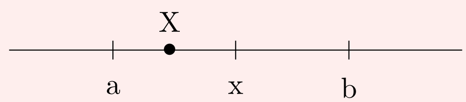


The prob. of X is selected from $[a, x]$ where $a < x < b$ is assigned as

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

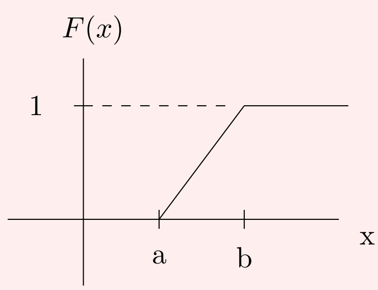
Similarly,

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



The cdf:

$$\begin{aligned} F(x) &= P(X \le x) \\ &= \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases} \\ P(X \le x) &= P(X < a) + P(a \le X \le x) \\ &= 0 + \frac{x-a}{b-a} \end{aligned}$$



Note that the cdf actually satisfies

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

where

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

To see this

- $x < a$

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

- $a \leq x \leq b$

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

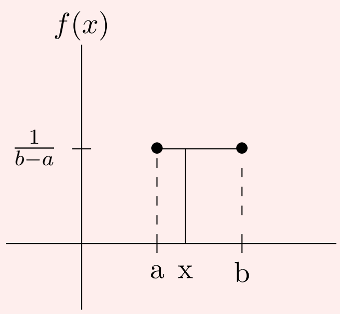
- $x > b$

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

Also, we have

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

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



§13 Lec 13: Oct 30, 2020

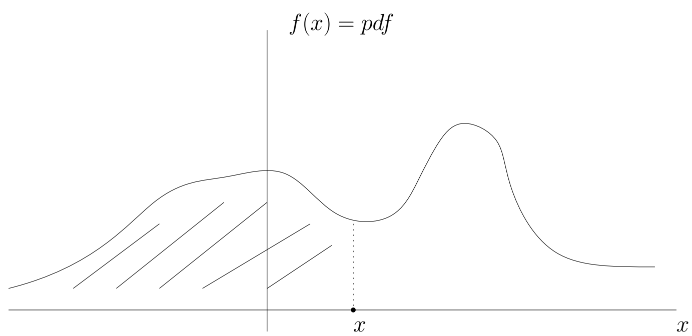
§13.1 Random Variable of Continuous Type (Cont'd)

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)

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



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

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: $E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx$
$\mu = E[x]$ $= \sum_{s \in S_X} xf(x)$ Mgf: $M_X(t) = \sum_{s \in S_X} e^{tx}f(x)$	Mean: $\mu = E[x]$ $= \int_{-\infty}^{\infty} xf(x)dx$ 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:

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

$$\begin{aligned}
\sigma^2 &= E[X^2] - \mu^2 \\
E[X^2] &= \int_a^b x^2 f(x)dx
\end{aligned}$$

$\ldots$ Exercise

Mgf:

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

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,

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

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.

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

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

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

$p = \frac{1}{4}$, $\pi_{0.25} = 25^{\text{th}}$ 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 Lec 14: Nov 4, 2020

§15.1 Exponential Distribution

Definition 15.1 (Exponential Distribution) — A continuous random variable is said to have an exponential distribution if the pdf $f(x)$ is given by for a fixed $\lambda > 0$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0, \\ 0, \text{ otherwise} \end{cases}$$

$S_X = [0, \infty)$. X is denoted as $X \sim \text{Exp}(\lambda)$. Note that in textbook, λ is denoted as $\frac{1}{\theta}$

Remark 15.2. The pdf of $X \sim \text{Exp}(\lambda)$ satisfies

$$\int_0^\infty f(x)dx = \int_0^\infty \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_{x=0}^{x\to\infty} = 1$$

Fact 15.1. If $X \sim \text{Exp}(\lambda)$ then $\mu = \frac{1}{\lambda}$ and $\sigma^2 = \frac{1}{\lambda^2}$.

Indeed,

$$\begin{aligned}\mu = E[x] &= \int_0^\infty xf(x)dx \\ &= \int_0^\infty x\lambda e^{-\lambda x}dx \\ &= x(-e^{-\lambda x})\Big|_{x=0}^{x\to\infty} - \int_0^\infty -e^{-\lambda x}dx \\ &= 0 + -\frac{1}{\lambda}e^{-\lambda x}\Big|_{x=0}^{x\to\infty} \\ &= \frac{1}{\lambda}\end{aligned}$$

Variance:

$$\begin{aligned}\sigma^2 &= E[X^2] - E[X]^2 \\ &= \int_0^\infty x^2\lambda e^{-\lambda x}dx - \frac{1}{\lambda^2} \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2}\end{aligned}$$

Moreover, the mgf of $\text{Exp}(\lambda)$ is given by

$$\begin{aligned}M_X(t) &= \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx \\ &= \lambda\int_0^\infty e^{(t-\lambda)x}dx \\ &= \lambda\int_o^\infty e^{-(\lambda-t)x}dx \\ &= \frac{\lambda}{\lambda - t}\end{aligned}$$

Thus, $M_X(t)$ exists if $t < \lambda$

Practice 15.1. 10 – Problem 1: (Memoryless Property)

$$P(X > t+s|X > t) = P(X > s)$$

- Cdf of $X \sim \text{Exp}(\lambda)$: for $t \geq 0$

$$\begin{aligned}F(t) &= P(X \leq t) \\ &= \int_0^t \lambda e^{-\lambda x}dx \\ &= -e^{-\lambda x}\Big|_{x=0}^{x=t} \\ &= 1 - e^{-\lambda t}\end{aligned}$$

Figure here

$$\begin{aligned}P(X > t) &= 1 - P(X \leq t) \\ &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t}\end{aligned}$$

- $$\begin{aligned} P(X > t+s | X > t) &= \frac{P(\{x > t+s\} \cap \{x > t\})}{P(x > t)} \\ &= \frac{P(X > t+s)}{P(X > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} = P(X > s) \end{aligned}$$

Theorem 15.3

Suppose X is cont r.v. on $[0, \infty)$ s.t. X satisfies the memoryless property above, i.e., for all $t, s > 0$

$$P(X > t+s | X > t) = P(X > s)$$

Then $\exists \lambda$ s.t. $X \sim \text{Exp}(\lambda)$.

§15.2 Poisson Process

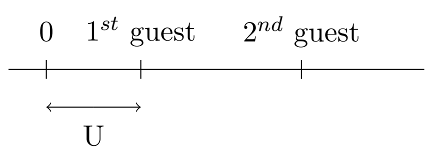
Recall that $X \sim \text{Pois}(\lambda) = \#$ of arrivals in $[0, 1)$ with mean $= \lambda$.

Question 15.1. Denote $N[a, b) = \#$ of guests arrivals in $[a, b)$, $N[a, b) = ?$

Ans: Using a similar approach – $N[a, b) \sim \text{Pois}(\lambda(b - a))$

Definition 15.4 (Poisson Process) — Practice 10.

Practice 15.2. 10 – Problem 3a: $U =$ first arrival time



<u>Goal</u>: Need to find cdf of U.

$$S_U = [0, \infty) \text{ and } U \text{ is a continuous random variable}$$

Given $t \geq 0$

$$\begin{aligned} P(U \leq t) &= 1 - P(U > t) \\ &= 1 - P(\text{"no guest in } [0, t)\text{"}) \\ &= 1 - P(N[0, t) = 0) \\ &= 1 - e^{-\lambda t} \frac{(\lambda t)^0}{0!} \\ &= 1 - e^{-\lambda t}, \text{ cdf of Exp}(\lambda) \end{aligned}$$

Thus, $U \sim \text{Exp}(\lambda)$

§15.3 Gamma Distribution

<u>Notation</u>: Gamma function
For $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

Fact 15.2. If α is positive integer

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

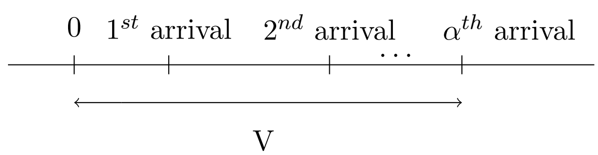
Definition 15.5 (Gamma Distribution) — $X \sim \Gamma(\alpha, \theta)$ if the pdf is given by

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}, x > 0 \\ 0, \text{ otherwise} \end{cases}$$

Remark 15.6. $f(x)$ indeed satisfies

$$\int_0^\infty f(x) dx = 1$$

Practice 15.3. 10 – Problem 3b: $\alpha \in \mathbb{N}$



$$\begin{aligned} P(V \le t) &= 1 - P(V > t) \\ &= 1 - P(\text{“ At most } \alpha - 1 \text{ arrivals before time t”}) \\ &= 1 - P(N[0,t) \le \alpha - 1) \\ &= 1 - \sum_{k=0}^{\alpha-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \\ &= P(V \le t) \end{aligned}$$

Now differentiate with respect to t, we obtain the pdf of V given by

$$f(t) = \frac{t^{\alpha-1} e^{-\frac{t}{\frac{1}{\lambda}}}}{\Gamma(\alpha) \left(\frac{1}{\lambda}\right)^\alpha} \sim \Gamma(\alpha, \theta = \frac{1}{\lambda})$$

$\sim$ Gamma$(\alpha, \theta = \frac{1}{\lambda}$.
Summary:

$$\text{EXP}(\lambda) = \text{arrival time of } 1^{\text{st}} \text{ guest}$$

$$\text{Gamma}(\alpha, \theta = \frac{1}{\lambda}) = \text{arrival time of } \alpha^{\text{th}} \text{ guest}$$

Remark 15.7.
- $\text{Exp}(\lambda\ )$ is a special case of $\text{Gamma}(\alpha, \theta)$ where $\alpha = 1, \theta = \frac{1}{\lambda}$.
- Mean of $\text{Gamma}(\alpha, \theta)$ is $\alpha \cdot \theta$.

§16 Lec 15: Nov 6, 2020

§16.1 Chi – Squared Distribution

Definition 16.1 (Chi – Squared Distribution) — X is called to have a Chi – Squared distribution if $X \sim \text{Gamma}(\alpha = \frac{r}{2}, \theta = 2)$. More specifically, the pdf is given by

$$f(x) = \begin{cases} \frac{x^{\frac{r}{2}-1}e^{-\frac{x}{2}}}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}}, x > 0 \\ 0, \qquad \text{otherwise} \end{cases}$$

X is denoted as

$$X \sim \chi^2(r)$$

and r is called the degree of freedom. (χ^2 dist. with r degree of freedom).

§16.2 Normal Distribution

Definition 16.2 (Normal Distribution) — A continuous random variable is called to have a normal distribution with parameter $\mu \in \mathbb{R}, \sigma^2 > 0$ is the pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}, S_X = \mathbb{R}$$

X is denoted as $X \sim N(\mu, \sigma^2)$.

Remark 16.3. $f(x)$ actually satisfies

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

Fact 16.1.

$$\int_{-\infty}^{\infty} e^{-z^2}dz = \sqrt{\pi}$$

Definition 16.4 —
1. If $Z \sim N(\mu = 0, \sigma^2 = 1)$ then Z is said to have a standard normal distribution.
2. In this case, the cdf of Z is denoted by Φ

$$\Phi(x) = F(z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz$$

Practice 16.1. 11 – Problem 1: Given $x \in \mathbb{R}$, $z = \frac{x-\mu}{\sigma}$

$$\begin{aligned}
P(Z \le x) &= P\left(\frac{x-\mu}{\sigma} \le x\right) \\
&= P(x \le \sigma x + \mu), (\sigma > 0) \\
&= \int_{-\infty}^{\sigma x + \mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt
\end{aligned}$$

Let $z = \frac{t-\mu}{\sigma} \implies dz = \frac{dt}{\sigma}$

$$\begin{aligned}
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2}} \sigma dz \\
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \Phi(x)
\end{aligned}$$

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

Theorem 16.5

If $X \sim N(\mu, \sigma^2)$ then

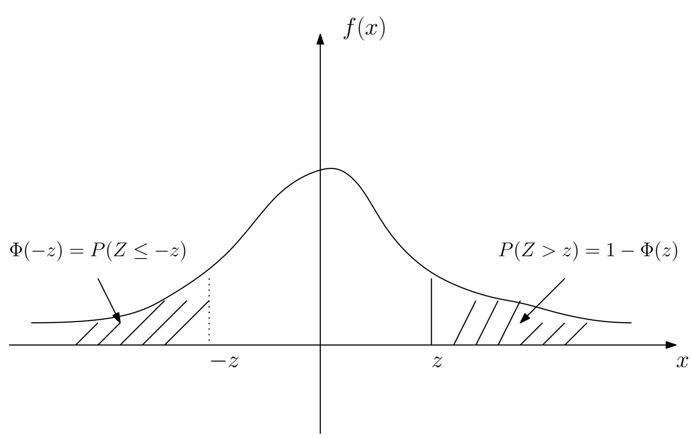
- MGF: $M(t) = \exp(\frac{\mu t + t^2\sigma^2}{2})$.
- $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$

Proof.

$$\begin{aligned}
M(t) &= E[e^{tX}] \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \dots \\
&= e^{\mu t + \frac{\sigma^2 t^2}{2}}
\end{aligned}$$

□

Practice 16.2. 11 – Problem 2: $Z \sim N(0,1)$



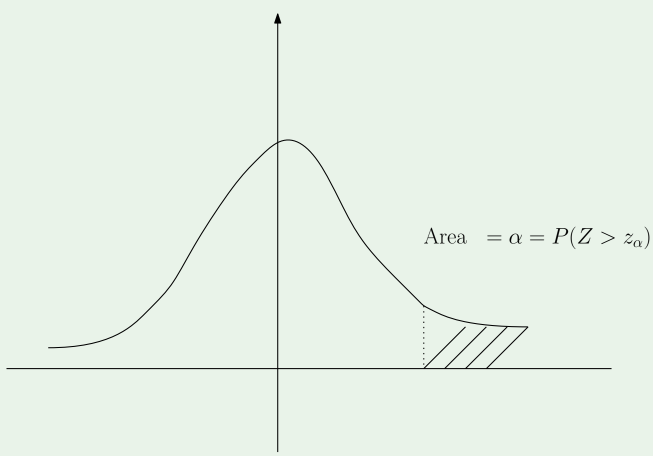
$$\begin{aligned}\Phi(-z) &= P(Z \le -z) \\ &= \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ 1 - \Phi(z) &= 1 - P(Z \le z) \\ &= P(Z > z) \\ &= \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\end{aligned}$$

Now,

$$\begin{aligned}\int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt &= \int_{\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} (-dy) \\ &= \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy\end{aligned}$$

Definition 16.6 — Let $Z \sim N(0,1)$ and $\alpha \in (0,1)$. Then z_α is defined as

$$P(Z > Z_\alpha) = \alpha$$



Practice 16.3. 11 – Problem 3: $Z \sim N(0,1)$

a)

$$\begin{aligned}P(.47 < Z \le 2.13) &= P(Z \le 2.13) - P(Z \le .47) \\ &= \Phi(2.13) - \Phi(.47)\end{aligned}$$

b)

$$\begin{aligned}P(|Z| > 1.5) &= P(\{Z < -1.5\} \cup \{Z > 1.5\}) \\ &= P(Z < -1.5) + P(Z > 1.5) \\ &= 2P(Z > 1.5) = 2 \cdot 0.0668\end{aligned}$$

c) $\alpha = 0.0485 \implies z_\alpha = 1.66$

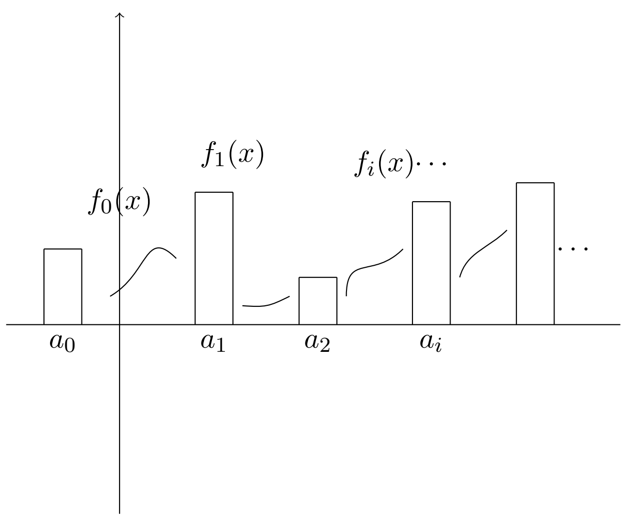
§17 Lec 16: Nov 9, 2020

§17.1 Random Variable of Mixed Type

- Combination of point mass and density

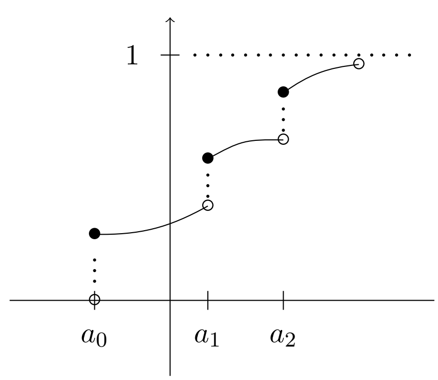
$$a_0 < a_1 < \ldots < a_n$$

- $P(X = a_i) > 0$
- $a_i < a_{i+1}$, density $f_i(x)$



$$\sum_{i=0}^{n} P(X = a_i) + \int_{a_0}^{a_1} f_0(x)dx + \ldots + \int_{a_{n-1}}^{a_n} f_n(x)dx = 1$$

- cdf



- Expectation

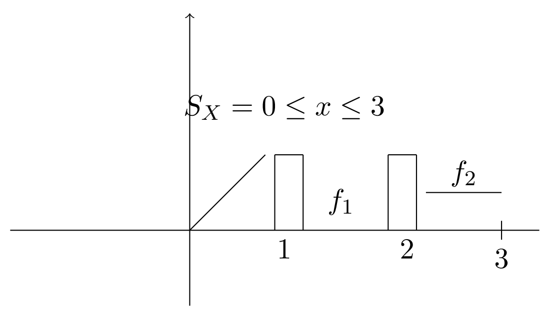
$$E[u(X)] = \sum_{i=0}^{n} u(a_i)P(X = a_i) + \int_{a_0}^{a_1} u(x)f_0(x)dx + \ldots + \int_{a_{n-1}}^{a_n} u(x)f_{n-1}(x)dx$$

Practice 17.1. 12 – Problem 1: find point mass:
$P(X=1)=\frac{1}{2}$

$$\begin{aligned} P(X=2) &= \frac{x}{3}\Big|_{x=2} - \frac{1}{2} \\ &= \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \end{aligned}$$

Find densities (by differentiating cdf)

- $0 \le x < 1$
$$f_0(x) = \left(\frac{x^2}{4}\right)' = \frac{x}{2}$$
- $1 < x < 2$
$$f_1(x) = \left(\frac{1}{2}\right)' = 0$$
- $2 \le x < 3$
$$f_2(x) = \left(\frac{x}{3}\right)' = \frac{1}{3}$$



$$\begin{aligned} E[X] &= 1\cdot P(X=1) + 2\cdot P(X=2) + \int_0^1 xf_0(x)dx + \int_1^2 xf_1(x)dx + \int_2^3 xf_2(x)dx \\ &= 1\cdot\frac{1}{4} + 2\cdot\frac{1}{6} + \int_0^1 x\frac{x}{2}dx + \int_1^2 x\cdot 0dx + \int_2^3 x\cdot\frac{1}{2}dx \\ &= \dots \end{aligned}$$

Practice 17.2. 12 – Problem 2: X = damage (in unit) of car, $S_x = 0 \le x \le 24$,

$$\begin{aligned} P(X=0) &= .95 \\ P(X=24) &= .01 \end{aligned}$$

$0 < x < 24$, $f(x) = \frac{25}{24}\frac{1}{(x+1)^2}$
<u>*Note*</u>:

$$P(X=0) + P(X=24) + \int_0^{24} f(x)dx = 1$$

Define $u(x)$ = insurance payment for damage of x (units).

$$u(x) = \begin{cases} 0, x \le 1 \\ x-1, x > 1 \end{cases}$$

which is due to one-unit deductible policy. Now,

$$\begin{aligned} E\left(u(x)\right) &= u(0)P(X=0) + u(24)P(X=24) + \int_0^{24} u(x)f(x)dx \\ &= 0 \cdot .95 + 23 \cdot .01 + \int_0^1 + \int_1^{24} u(x)f(x)dx \end{aligned}$$

Consider the integral $\int_0^1 = 0$, and

$$\begin{aligned} &= \frac{25}{24}\int_1^{24} \frac{x-1}{(x+1)^2}dx \\ &= \dots \end{aligned}$$

See also Hw 6 #2.

§17.2 Weibull Distribution

Definition 17.1 (Weibull Distribution) — $X \sim \text{Weibull}(\alpha, \beta), \alpha, \beta > 0$ if $S_X = (0, \infty)$ and density is given by

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}, x > 0$$

Remark 17.2. Let $G(x) = \left(\frac{x}{\beta}\right)^\alpha$, then

$$f(x) = G'(x)e^{-G(x)}$$

In contrast, for $Y \sim \text{Exp}(\lambda)$

$$f_Y(x) = \lambda e^{-\lambda x}$$

with $G_2 = \lambda x$, then

$$f_Y(x) = G_2'(x)e^{-G_2(x)}$$

Practice 17.3. 12 – Problem 12: $X \sim \text{Weibull}(\alpha, \beta)$, $E[X] =?$

The MGF approach is not really helpful – See also HW 6 # 5.

$$\begin{aligned} E[X] &= \int_0^\infty xf(x)dx \\ &= \int_0^\infty x\frac{\alpha}{\beta^\alpha}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^\alpha}dx \\ &= \frac{\alpha}{\beta^\alpha}\int_0^\infty x^\alpha e^{-\left(\frac{x}{\beta}\right)^\alpha}dx \\ &= \alpha\beta\int_0^\infty u^\alpha e^{-u^\alpha}dx \end{aligned}$$

Let $z = u^\alpha$

$$\begin{aligned}
&= \alpha\beta \int_0^\infty z e^{-z} \frac{dz}{\alpha z^{1-\frac{1}{\alpha}}} dz \\
&= \beta \int_0^\infty z^{\frac{1}{\alpha}} e^{-z} dz \\
&= \beta \int_0^\infty \frac{z^{(\frac{1}{\alpha}+1)-1} e^{-\frac{z}{1}}}{\Gamma(\frac{1}{\alpha}+1) 1^{\frac{1}{\alpha}+1}} dz \Gamma(\frac{1}{\alpha}+1) \\
&= \beta \Gamma\left(\frac{1}{\alpha}+1\right)
\end{aligned}$$

§18 Veterans Day: Nov 11, 2020

No class :D

§19 Lec 17: Nov 13, 2020

§19.1 Bivariate Distribution of Discrete Type

Definition 19.1 (Joint pmf) — Let X, Y be discrete random variables

1. $S_{X \times Y}$: the two – dimensional space of $X \times Y$.
2. The joint $PMF, f(x,y)$ for each $x \times y \in S_{X \times Y}$ is given by

$$f(x,y) = P(X = x, Y = y)$$

   satisfying the followings:

   - $f(x,y) \geq 0$
   - $\sum_{(x,y) \in S_{X \times Y}} f(x,y) = 1$
   - $P((X,Y) \in A) = \sum_{(x,y) \in A} f(x,y)$ where $A \subseteq S_{X \times Y}$

Example 19.2

Roll a dice twice. Denote X = min of 2 rolls, Y = max of 2 rolls.
e.g., roll (1,3) then $X = 1, Y = 3$.

Table of outcomes of rolls with equal probability $\frac{1}{36}$ each. TBA

$$
\begin{aligned}
f(x,y) &= P(X=x, Y=y) \\
&= \begin{cases} \frac{1}{36}, x=y \\ \frac{2}{36}, x<y \\ 0, x>y \end{cases} \\
&= \sum_{(x,y)\in S_{X,Y}} f(x,y) = 1
\end{aligned}
$$

Definition 19.3 (Marginal Pmf) — Given a joint pmf of X, Y on $S_{X\times Y}$, the pmf of X itself is called the marginal pmf of X and given by

$$f_X(x) = P(X=x) = \sum_y f(x,y)$$

where $x \in S_X$. Similarly for the marginal pmf of Y.

Remark 19.4. We have

$$
\begin{aligned}
P(X=x) &= \sum_{y\in S_Y} P(X=x, Y=y) \\
&= \sum_{y\in S_y} f(x,y)
\end{aligned}
$$

Definition 19.5 (Independent for Multivariable) — X, Y are independent if

$$P(X=x, Y=y) = P(X=x)P(Y=y)$$

i.e., $f(x,y) = f_X(x)f_Y(y)$

Example 19.6 (above)

Marginal of Y

$$
\begin{aligned}
f_Y(1) &= \frac{1}{36} \\
f_Y(2) &= \frac{2}{36} + \frac{1}{36} = \frac{3}{36}
\end{aligned}
$$

Marginal of X

$$f_X(1) = \sum_{y \in S_Y} f(1, y) = \frac{11}{36}$$

$$f_X(2) = \sum_{2^{\text{nd}} \text{ column}} f(2, y) = \frac{9}{36}$$

Question 19.1. X, Y independent?

$$f(1, 1) = \frac{1}{36} \neq \frac{1}{36} \cdot \frac{11}{36} = f_X(1) f_Y(1)$$

Thus, not independent.

Or, an alternative way:

$$f(2, 1) = 0 \neq \frac{9}{36} \cdot \frac{1}{36} = f_X(2) f_Y(1)$$

Remark 19.7.
1. If the joint pmf table is not "full" then X, Y are dependent.
2. If the table is "full", i.e., all entries are non-zero, it does NOT imply independence.

Definition 19.8 (Expectation for Multivarible) —
1. The expectation $E\left[u(X, Y)\right]$ is given by

$$E\left[u(X, Y)\right] = \sum_{(x,y) \in S_{X \times Y}} u(x, y) f(x, y)$$

2. Marginal mean

$$\mu_X = E[X], \quad u(X, Y) = X$$

Marginal variance

$$\sigma_X^2 = E\left[(x - \mu_X)^2\right], \quad u(X, Y) = (X - \mu_X)^2$$

and similar notions for Y.

Practice 19.1. 13 – Problem 1: Left as exercise.
13 – Problem 2: $X = \#A$ students, $Y = \#B$ students
a)

$$S_{X \times Y} = \begin{cases} (x, y) : x \geq 0, \\ \quad y \geq 0, \\ \quad x \leq 30, \\ \quad y \leq 60, \\ \quad x + y \leq 40 \end{cases}$$

b) The total number of ways to choose 40 from 200 is $\binom{200}{40}$. Given $(x, y) \in S_{X\times Y}$

- Choose x students from 30 students with A which is $\binom{30}{x}$.
- Choose y students from 60 B which is $\binom{60}{y}$.
- Choose $40-x-y$ students from 110 students with C, D, F, which is $\binom{110}{40-x-y}$.

Thus,

$$P(X = x, Y = y) = \frac{\binom{30}{x}\binom{60}{y}\binom{110}{40-x-y}}{\binom{200}{40}}$$

c) $X = \#A$ students from a random of 40, $n = 40$.

$$\begin{aligned} N_1 &= \#\text{A students } = 30 \\ N_2 &= \#\text{ non A students } = 170 \end{aligned}$$

$X \sim \text{Hypergeom}(N_1 = 30, N_2 = 170, n = 40)$

$$P(X = x) = \frac{\binom{30}{x}\binom{170}{40-x}}{\binom{200}{40}}$$

and

$$S_X = \begin{cases} x \geq 0, x \leq 30 \\ \quad 40 - x \leq 170 \end{cases}$$

Practice 19.2. 13 – Problem 3: $X = \#$ sweet cups, $Y = \#$ bland cups. Each trial (cup) has 3 outcomes

1. sweet with prob $p_1 = .26$
2. bland with prob $p_2 = .04$
3. perfect with prob $p_3 = .7$

- Choose x cups from 25 to assign sweet which is $\binom{25}{x}P_1^x$
- Choose y cups from $25-x$ to assign "bland" which is $\binom{25-x}{y}P_2^y$
- Choose $25-x-y$ cups from $25-x-y$ to assign "perfect" which is $\binom{25-x-y}{25-x-y}P_3^{25-x-y}$. Thus, $P(X = x, Y = y)$

$$\begin{aligned} &= \binom{25}{x}\binom{25-x}{y}\binom{25-x-y}{25-x-y}P_1^xP_2^y\left(1-P_1-P_2\right)^{25-x-y} \\ &= \frac{25!}{x!(25-x)!}\cdot\frac{(25-x)!}{y!(25-x-y)!}\cdot 1\cdot\ldots \\ &= \frac{25!}{x!y!(25-x-y)!}P_1^xP_2^y(1-P_1-P_2)^{25-x-y} \\ &= \binom{25}{x, y, 25-x-y}P_1^xP_2^y\left(1-P_1-P_2\right)^{25-x-y} \end{aligned}$$

$$S_{X\times Y} = \begin{cases} (x, y) : x + y \leq 25 \\ \quad x \geq 0, y \geq 0 \end{cases}$$

Note: Marginal of $X \sim \text{Binom}(n = 25, P_1 = .26)$,
Marginal of $Y \sim \text{Binom}(n = 25, P_2 = 0.04)$.
b) $P(X \geq 2 \text{ or } Y \geq 1)$ which is equal to $1 - P(X \leq 1, Y = 0) = 1 - f(0,0) - f(1,0)$.

§20 Lec 18: Nov 16, 2020

§20.1 Correlation Coefficient

Recall that if (X, Y) has a joint pmf $f(x, y)$ then $\mu_X = E[X], \mu_y = E[Y]$ and the variance $\sigma_X^2 = E(X - \mu_X)^2, \sigma_Y^2 = E(Y - \mu_Y)^2$.

Definition 20.1 (Covariance – Correlation Coefficient) — 1. The covariance, denoted by $\text{cov}(X, Y) := \sigma_{XY}$ is given by

$$\text{cov}(X, Y) = \sigma_{XY} = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

2. The correlation coefficient, denoted P, is given by

$$P = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Theorem 20.2 1. The covariance σ_{XY} is given by

$$\sigma_{XY} = \text{cov}(X, Y) = E[XY] - \mu_X \mu_Y$$

2. If X, Y are independent, then
   - $E\left[u(X)v(Y)\right] = E\left[u(X)\right] E\left[v(Y)\right]$ for any $u(x)$ and $v(y)$.
   - $\sigma_{XY} = 0$.

   In general, $\sigma_{XY} = 0$, then X, Y are called uncorrelated.

Proof. 1.

$$\begin{aligned}
\sigma_{XY} &= E\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= E\left[XY - \mu_X \cdot Y - X \cdot \mu_Y + \mu_X \mu_Y\right] \\
&= E\left[XY\right] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y \\
&= E[XY] - \mu_X \mu_Y
\end{aligned}$$

2. Recall X, Y independent means $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all

$(x, y) \in S_{X\times Y}$. We have

$$\begin{aligned} E\left[u(X)v(Y)\right] &= \sum_{(x,y)\in S_{X\times Y}} u(x)v(y)P(X=x, Y=y) \\ &= \sum u(x)v(y)P(X=x)P(Y=y) \\ &= \sum_{x\in S_X}\sum_{y\in S_Y} u(x)P(X=x)v(y)P(Y=y) \\ &= \sum_{x\in S_X} u(x)P(X=x) \sum_{y\in S_Y} v(y)P(Y=y) \\ &= E[u(X)]E[v(Y)] \end{aligned}$$

Also,

$$\begin{aligned} \operatorname{cov}(X,Y) &= \sigma_{XY} \\ &= E[XY] - \mu_X\mu_Y \\ &= E[X]E[Y] - \mu_X\mu_Y \\ &= 0 \end{aligned}$$

□

Remark 20.3.
1. Note that in general, $\operatorname{cov}(X,Y) = 0$ does not imply independence. Example: figure here $f(1,1) = 0$ but $f_X(1) = f_Y(1) = \frac{1}{3}$ and thus $\frac{1}{3^2} \neq 0$. So, X, Y are dependent. However, notice that $\operatorname{cov}(X,Y) = 0$.
2. The correlation coefficient $p = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$ satisfies $-1 \leq p \leq 1$ i.e., $|p| \leq 1$. (σ_{XY} maybe negative in general)

Practice 20.1. 14 – Problem 1: a) $(X,Y) \sim$ Trinom(n, p_1, p_2)
Each trial:

- X occurs with prob p_1
- X does not occur with prob $1 - p_1$

$X \sim$ Binom(n, p_1)

$$\begin{aligned} \mu_X &= np_1 \\ \sigma_X^2 &= np_1(1-p_1) \end{aligned}$$

Likewise, $Y \sim$ Binom(n, p_2).
b) Left as exercise.

Note: For a derivation of ρ the correlation coefficient, see textbook section 4.2.

§20.2 Conditional Distribution

Consider (X, Y) with joint $f(x, y)$ and marginal f_X, f_Y. Define $A = \{X = x\}, B = \{Y = y\}$. Then

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{f(x, y)}{f_Y(y)}, \end{aligned}$$

provided $f_Y(y) > 0$.

Definition 20.4 (Conditional pmf) — 1. The conditional pmf of X given $Y = y$, is defined as

$$g(x|y) := \frac{f(x, y)}{f_Y(y)}, \text{ provided } f_Y(y) > 0$$

2. Likewise, the conditional pmf of Y, given $X = x$, is given by

$$h(y|x) := \frac{f(x, y)}{f_X(x)}, \text{ provided } f_X(x) > 0$$

Example 20.5

Flip a coin with faces $\{0, 1\}$ twice. Define $X =$ smaller value, $Y =$ larger value. figure here $y = 0, X|Y = 0$ is a RV with pmf

$$g(x|0) = \frac{f(x, 0)}{f_Y(0)} = \begin{cases} \frac{\frac{1}{4}}{\frac{1}{4}} = 1, & \text{if } x = 0 \\ \frac{0}{\frac{1}{4}} = 0, & \text{if } x = 1 \end{cases}$$

Given max $= 0$, the min must be 0 with prob 1.
$y = 1, X|Y = 1$ is a RV with pmf

$$g(x|1) = \frac{f(x, 1)}{f_Y(1)} = \begin{cases} \frac{\frac{2}{4}}{\frac{3}{4}} = \frac{2}{3}, & \text{if } x = 0 \\ \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}, & \text{if } x = 1 \end{cases}$$

Note that in both cases,

$$\sum_{x \in S_X} g(x|0) = 1 = \sum_{x \in S_X} g(x|1)$$

Similarly, when either $x = 0$ or 1

$$\sum_{y \in S_Y} h(y|x = 0) = \sum_{y \in S_Y} h(y|x = 1) = 1$$

Proposition 20.6

The conditional pmf $g(x|y)$ and $h(y|x)$ satisfy

$$\sum_{x \in S_X} g(x|y) = 1$$

and

$$\sum_{y \in S_Y} h(y|x) = 1$$

Proof. Given $X = x$,

$$\begin{aligned}\sum_{y \in S_Y} h(y|x) &= \sum_{y \in S_Y} \frac{f(x,y)}{f_X(x)} \\ &= \frac{\sum_{y \in S_Y} f(x,y)}{f_X(x)} \\ &= \frac{f_X(x)}{f_X(x)} \\ &= 1\end{aligned}$$

Similarly for $\sum_{x \in S_X} g(x|y) = 1$. □

§21 Lec 19: Nov 18, 2020

§21.1 Lec 18 (Cont'd)

Recall that $(X|Y = y)$ is discrete RV with the pmf

$$g(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Definition 21.1 (Conditional Expectation) — The conditional expectation of X, given $\{Y = y\}$, is defined as

$$\begin{aligned} E[X|Y = y] &:= E[X|y] \\ &:= \sum_{x \in S_X} x g(x|y) \end{aligned}$$

More generally, given $Y = y$,

$$\begin{aligned} E[u(X)|Y = y] &:= E[u(X)|y] \\ &:= \sum_{x \in S_X} u(x) g(x|y) \end{aligned}$$

We denote

$$\begin{aligned} \mu_{X|y} &= E[X|y] \\ \sigma^2_{X|y} &= E\left[(X - \mu_{X|y})^2 | y\right] \end{aligned}$$

Proposition 21.2

$$\sigma^2_{X|y} = E[X^2|y] - \left(\mu_{X|y}\right)^2$$

Proof. Left as exercise. □

Example 21.3 (Previous Lecture)

$X = \min$, $Y = \max$

- $y = 0$,

$$\begin{aligned} g(x|0) &= 1 \text{ when } x = 0 \\ \mu_{X|0} &= 0 \\ \sigma^2_{X|0} &= E[X^2|0] - \left(\mu_{X|0}\right)^2 \\ &= 0 - 0^2 = 0 \end{aligned}$$

- $y = 1$,

$$g(x|1) = \begin{cases} \frac{2}{3}, & \text{when } x = 0 \\ \frac{1}{3}, & \text{when } x = 1 \end{cases}$$

$$\mu_{X|1} = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$$

$$E[X^2|1] = 0^2 \cdot \frac{2}{3} + 1^2 \cdot \frac{1}{3} = \frac{1}{3}$$

$$\begin{aligned} \sigma^2_{X|1} &= E[X^2|1] - \left(\mu_{X|1}\right)^2 \\ &= \frac{1}{3} - \left(\frac{1}{3}\right)^2 \\ &= \frac{2}{9} \end{aligned}$$

- In summary,

$$\mu_{X|Y} = \begin{cases} 0, & y = 0 \\ \frac{1}{3}, & y = 1 \end{cases}$$

i.e.,

$$\mu_{X|Y} = E[X|y] \text{ is a function of } y$$

Practice 21.1. 14 – Problem 2: a) Find $h(y|x)$. 1 trial:

- Success
- Normal
- Failure

$X = \#$ successes, $Y = \#$ failures. $(X, Y) \sim$ trinom(n, p_1, p_2). Given $\{X = x\} = \{\text{there are} x \text{ successes among n trials}\}$. A heuristics argument: there are x successes among n trials – there are n-x non-success trials left, each happens with prob. $1 - p_1$.

$$\{Y = y|X = x\} = \{y \text{ failures among n-x non-success trials}\}$$

$$Y|X = x \sim \text{ Binom}(n - x, \frac{p_2}{1 - p_1})$$

Rigorous calculation:

$$\begin{aligned} h(y|x) &= P\left(Y = y|X = x\right) \\ &= \frac{f(x, y)}{f_X(x)}, \quad X \sim \text{ Binom}(n, p_1) \\ &= \frac{\frac{n!}{x!y!(n-x-y)!} \cdot p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}}{\frac{n!}{x!(n-x)!} \cdot p_1^x (1 - p_1)^{n-x}} \\ &= \frac{(n - x)!}{y!(n - x - y)!} \cdot \left(\frac{p_2}{1 - p_1}\right)^y \cdot \left(1 - \frac{p_2}{1 - p_1}\right)^{n-x-y} \end{aligned}$$

$(Y|X = x) \sim \text{binom}(n - x, \frac{p_2}{1-p_1})$. Thus,

$$\mu_{Y|x} = (n - x) \cdot \frac{p_2}{1 - p_1}$$

$B(n,p)$ then $\mu = np$.

Notice that

$$\frac{p_2}{1-p_1} \leq 1 \text{ since } p_2 + p_1 \leq 1$$

§21.2 Conditional Expectation as a Random Variable

Example 21.4 1. X is a RV then $u(X)$ is too.

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

x	-1	0	2
$u(x)$	-1	-2	2

$u(x) = x^2 - 2$

$u(x)$	-1	-2	2
$f_u(x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

pmf of $u(x)$

i.e., $u(X) = X^2 - 2$ is a discrete random variable with the abode pmf.

$$P(u(X) = u(x)) = P(X = x)$$

2. Trinomial distribution: Define

$$u(x) \coloneqq E[Y|x] = (n-x)\frac{p_2}{1-p_1}$$

which is a function of x. Thus, $u(X) \coloneqq E[Y|X]$ is a random variable with pmf

$$P(u|X) = E[Y|x] = (n-x)\frac{p_2}{1-p_1} = P(X=x) = \frac{n!}{x!(n-x)!}p_1^x(1-p_1)^{n-x}$$

Definition 21.5 (Conditional Expectation as a RV) — Given (X,Y) jointly distributed, define

$$u(x) \coloneqq E[Y|x] = E[Y|X=x]$$

Then $u(X)$, denoted by $E[Y|X]$, is a RV with the space of values $S = \{E[Y|x] : x \in S_X\}$, with pmf

$$P\left(u(X) = E[Y|x]\right) = P(X=x)$$

Example 21.6 (Trinomial Distribution)
$E[Y|X]$ is a discrete RV with pmf

$$P\left(E[Y|X] = E[Y|x]\right) = P(X=x)$$

Now,

$$\begin{aligned}
E\left[E[Y|X]\right] &= \sum E[Y|x] \cdot P(E[Y|X] = E[Y|x]) \\
&= \sum (n-x) \frac{p_2}{1-p_1} \frac{n!}{x!(n-x)!} p_1^x (1-p_1)^{n-x} \\
&= n \cdot p_2 \\
&= E[Y], \quad (Y \sim \text{ Binom}(n, p_2))
\end{aligned}$$

Theorem 21.7

$E\left[E[Y|X]\right] = E[Y]$ (Practice 14 – Problem 3).

Proof.

$$\begin{aligned}
E\left[E[Y|X]\right] &= \sum E[Y|x] \cdot P\left(E[Y|X] = E[Y|x]\right) \\
&= \sum_{x \in S_X} E[Y|x] \cdot f_X(x) \\
&= \sum_{x \in S_X} \left[\sum_y y h(y|x)\right] f_X(x) \\
&= \sum_{x \in S_X} \left[\sum_{y \in S_Y} y \frac{f(x,y)}{f_X(x)}\right] f_X(x) \\
&= \sum_x \left[\sum_y y \cdot f(x,y)\right] \\
&= \sum_y \sum_x y f(x,y) \\
&= \sum_y y \left[\sum_x f(x,y)\right] \\
&= \sum y f_Y(y) \\
&= E[Y]
\end{aligned}$$

□

§22 Lec 20: Nov 20, 2020

§22.1 Continuous Bivariate Random Variable

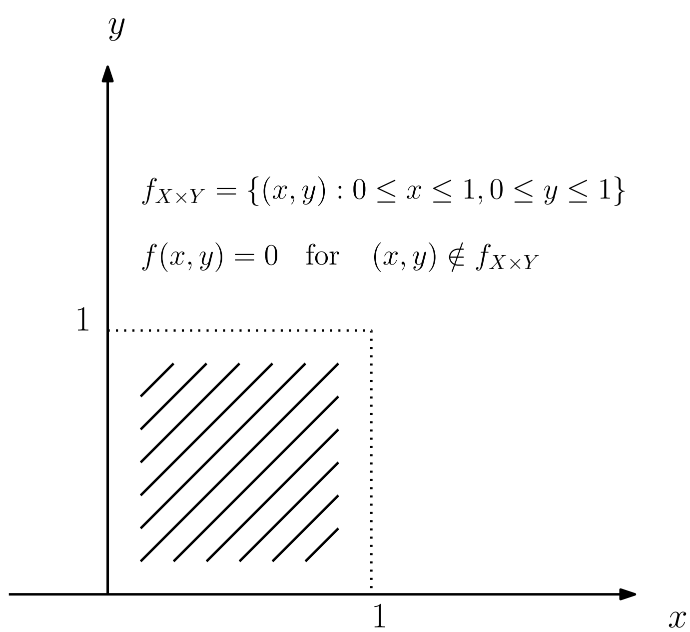
Definition 22.1 — 1. The joint pdf of a continuous bivariate RV (X, Y) is an integrable function $f(x, y)$ s.t.

- $f(x, y) \geq 0, (x, y) \in S_{X\times Y}$ and $f(x, y) = 0$ if $(x, y) \notin S_{X\times Y}$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$
- For $A \subseteq S_{X\times Y}$, $\iint_A f(x, y) dx dy = P((X, Y) \in A)$

2. The marginal pdf's of X, Y are given

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, x \in S_X$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx, y \in S_Y$$

Problem 22.1. 15 – Problem 1a): $f(x, y) = \frac{4}{3}(1 - xy)$



- Check $f(x, y) \geq 0$ for $(x, y) \in S_{X\times Y}$ since $0 \leq x, y \leq 1, xy \leq 1$ thus $\frac{4}{3}(1 - xy) \geq 0$

- Check

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1 \\ &= \int_0^1\int_0^1 \frac{4}{3}(1-xy)\,dx\,dy \\ &= \int_0^1 \left[\frac{4}{3}x - \frac{4}{3}y\cdot\frac{x^2}{2}\right]_{x=0}^{x=1} dy \\ &= \int_0^1 \frac{4}{3} - \frac{2}{3}y dy \\ &= \frac{4}{3}y - \frac{1}{3}y^2\Big|_{y=0}^{y=1} \\ &= 1\end{aligned}$$

Remark 22.2. For double integral, the order of integration does not matter, i.e.,

$$\iint f(x,y)\,dx\,dy = \iint f(x,y)\,dy\,dx$$

under "advance" condition. However, one direction might be easier than the other.

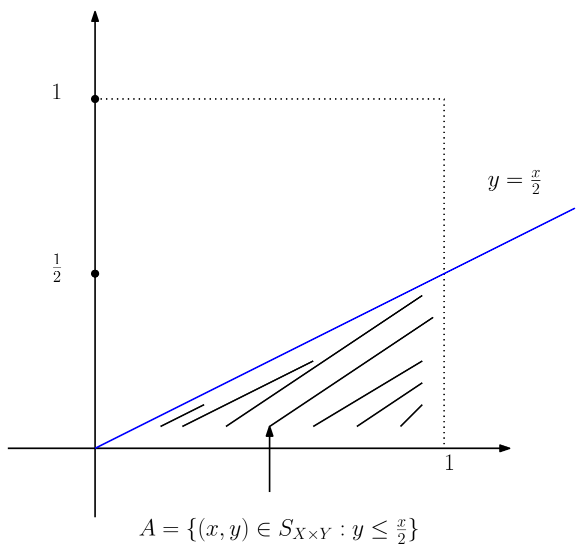
Problem 22.2. 15 – 1a) (cont'd) for each $x \in [0,1] = S_X$

$$\begin{aligned}f_X(x) &= \int_{\mathbb{R}} f(x,y)dy \\ &= \int_0^1 f(x,y)dy \\ &= \int_0^1 \frac{4}{3}(1-xy)dy \\ &= \frac{4}{3}y - \frac{4}{3}x\cdot\frac{y^2}{2}\Big|_{y=0}^{y=1} \\ &= \frac{4}{3} - \frac{2}{3}x\end{aligned}$$

Likewise,

$$f_Y(y) = \int_0^1 f(x,y)dx = \frac{4}{3} - \frac{2}{3}y$$

b) $P(Y \le \frac{X}{2})$



$$P(Y \leq \frac{X}{2}) = \iint_A f(x,y)\,dx\,dy$$
$$= \int_0^{\frac{1}{2}} \int_{2y}^1 f(x,y)\,dx\,dy$$

Note that we also have

$$P(Y \leq \frac{X}{2}) = \int_0^1 \int_0^{\frac{x}{2}} f(x,y)\,dy\,dx$$
$$= \int_0^1 \frac{2}{3}x - \frac{1}{6}x^3 dx$$
$$= \frac{7}{24}$$

c)

$$E\left[\underbrace{X^2 - Y}_{u(X,Y)}\right] = \int_0^1 \int_0^1 (x^2 - y) f(x,y) dx dy$$
$$= \int_0^1 \int_0^1 (x^2 - y) \frac{4}{3}(1 - xy) dx dy$$
$$= \dots$$
$$= \frac{1}{6}$$

§23 Lec 21: Nov 23, 2020

§23.1 Lec 20 (Cont'd)

Midterm 2 covers chapters 3 & 4.

Practice 23.1. 15 – Problem 2a): Note that $f(x,y) = 4 > 0$ in $S_{X\times Y}$ and

$$1 = \int_0^{\frac{1}{2}} \int_{2y}^{1} f(x,y)\, dx\, dy = \int_0^1 \int_0^{\frac{x}{2}} f(x,y)\, dy\, dx$$

b) Marginal:

$$\begin{aligned} f_X(x) &= \int_0^{\frac{x}{2}} f(x,y) dy \\ &= \int_0^{\frac{x}{2}} 4 dy \\ &= 2x \end{aligned}$$

And

$$\begin{aligned} f_Y(y) &= \int_{2y}^{1} f(x,y) dx \\ &= \int_{2y}^{1} 4 dx \\ &= 4 - 8y \end{aligned}$$

c) $S_{X\times Y} = \left\{0 \le X, Y \le \frac{1}{2}\right\}$

$$\begin{aligned} P(0 \le X, Y \le \frac{1}{2}) &= \iint f(x,y)\, dx\, dy \\ &= \int_0^{\frac{1}{4}} \int_{2y}^{\frac{1}{2}} 4\, dx\, dy \\ &= \dots \text{ (algebra)} \\ &= \frac{1}{4} \end{aligned}$$

§23.2 Independence

Definition 23.1 (Independent Continuous Bivariate RV) — Let (X,Y) be a continuous random bivariate random variables. Then X, Y are said to be independent if for any $A \subseteq S_X, B \subseteq S_Y$,

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

Theorem 23.2

Let (X,Y) be a continuous bivariate RV. Then X, Y independent iff $f(x,y) = f_X(x) f_Y(y)$.

Note:

Discrete case:

- Independence $\iff \underbrace{f(x,y)}_{\text{joint pmf}} = \underbrace{f_X(x)f_Y(y)}_{\text{marginal pmf}}$.
- pmf table is not "full" $\implies$ dependence.
- However, "full" does not imply independence

Continuous case:

- Independence $\iff f(x,y) = f_X(x)f_Y(y)$
- Domain $S_{X\times Y}$ is not a "rectangle" $\implies$ dependence.
- However, "rectangle" $S_{X\times Y}$ does not imply independence.

Example 23.3 (Practice 15)
2) $S_{X\times Y}$ is a triangle, X, Y are dependent.

$$\begin{aligned} f(x,y) &= 4 \\ f_X(x) &= 2x \\ f_Y(y) &= 4(1-2y) \\ 4 &\neq 2x(4-8y) \end{aligned}$$

for $(X,Y) \in S_{X\times Y}$. Thus, X, Y are dependent.

Theorem 23.4 1. X, Y are independent iff for any $u(X), v(Y)$

$$E\left[u(X)v(Y)\right] = E[u(X)]E[v(Y)]$$

2. If X, Y are independent then

$$\text{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E[XY] - \mu_X\mu_Y = 0$$

Practice 23.2. 15 – Problem 3: Recall in 1D, $X \sim \text{Unif}(a,b)$ then $f(x) = \frac{1}{b-a}, x \in (a,b)$. So, in 2D, (X,Y) is said to have a uniform dist with a joint pdf

$$f(x,y) = \frac{1}{\text{area of the domain}}$$

In this problem,

$$f(x,y) = \frac{1}{2}, \quad (x,y) \in \{0 \le x \le 1, 0 \le y \le 2\}$$

Note that

$$\iint_{S_{X\times Y}} f(x,y)dxdy = \int_0^2 \int_0^1 \frac{1}{2} dxdy = 1$$

Now, to verify independence,

$$f_X(x) = \int_0^2 f(x,y)dy = \int_0^2 \frac{1}{2}dy = 1$$

for $X \in S_X = (0,1)$. Thus, $X \sim \text{Unif}(0,1)$.

$$f_Y(y) = \int_0^1 f(x,y)\,dx = \int_0^1 \frac{1}{2}\,dx = \frac{1}{2}$$

for $Y \in S_Y = (0,2)$. Thus, $Y \sim \text{Unif}(0,2)$. Now,

$$f(x,y) = \frac{1}{2} = 1 \cdot \frac{1}{2} = f_X(x) f_Y(y)$$

Thus, X, Y are independent.

§23.3 Conditional Expectation

Discrete RV:

- Conditional pmf: $X|Y = y$ is a RV

$$\begin{aligned} g(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= P(X = x | Y = y) \end{aligned}$$

- $\sum_{x \in S_X} g(x|y) = 1$.
- Expectation:

$$E\left[u(X)|y\right] = \sum_{x \in S_X} u(x) g(x|y)$$

  In particular,
$$E\left[X|y\right] = \sum x g(x|y)$$

Continuous RV:

- $X|Y = y$ is a continuous RV with conditional pdf

$$g(x|y) = \frac{f(x,y)}{f_Y(y)}$$

-

$$\begin{aligned} 1 &= \int_{\mathbb{R}} g(x|y)\,dx \\ &= \int_{\mathbb{R}} \frac{f(x,y)}{f_Y(y)} \\ &= \frac{1}{f_Y(y)} f_Y(y) \\ &= 1 \end{aligned}$$

- Expectation:

$$E\left[u(X)|Y\right] = \int_{\mathbb{R}} u(x)g(x|y)dx$$

  In particular,

$$E[X|y] = \int_{\mathbb{R}} xg(x|y)dx$$

Theorem 23.5

If (X, Y) are conditional bivariate random variable, then

$$E[X] = \int E[X|y]f_Y(y)\,dy$$
$$E[Y] = \int E[Y|x]f_X(x)\,dx$$

§24 Lec 22: Nov 25, 2020

§24.1 Lec 21 (Cont'd)

Recalling $Y|X = x$ is a continuous RV with the pdf

$$h(y|x) = \frac{f(x,y)}{f_X(x)}$$

and

$$E\left[u(Y)|x\right] = \int_{\mathbb{R}} u(y)h(y|x)\,dy$$

Practice 24.1. 15 – Problem 2: $Y|X = x$

- $$h(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2}{x}$$
- $S_{Y|x} = \left\{0 \le y \le \frac{x}{2}\right\}$. Thus, $Y|X = x \sim$ Unif$\left(0, \frac{x}{2}\right)$.
- $E\left[Y|X = x\right] = \frac{0+\frac{x}{2}}{2} = \frac{x}{4}$.
- Likewise, $X|Y = y$ is a RV with pdf

$$g(x|y) = \frac{1}{1-2y}$$

  and $S_{X|y} = \{2y \le x \le 1\}$. Thus, $X|Y = y \sim$ Unif$(2y, 1)$ and $E[X|y] = \frac{2y+1}{2}$.

§24.2 Bivariate Normal Distribution

Definition 24.1 — (X, Y) is called to have a bivariate normal distribution if any linear combination of X, Y has a normal distribution, i.e., for all constants a and b in $\mathbb{R}, a, b$ both not zero.

$$a \cdot X + b \cdot Y \sim N\left(\mu_{ab}, \sigma_{ab}^2\right)$$

where $\mu_{ab} \in \mathbb{R}, \sigma_{ab}^2 > 0$ depending on a, b.

Remark 24.2.

1. It follows from the definition that

$$X = 1 \cdot X + 0 \cdot Y \sim N(\mu_X, \sigma_X^2)$$
$$Y = 0 \cdot X + 1 \cdot Y \sim N(\mu_Y, \sigma_Y^2)$$

2. In general, X, Y are normal does NOT imply (X, Y) bivariate normal, i.e., $aX + bY$ is normal (per defn).

Example 24.3

$X \sim N(0, 1), Y := -X$. Then $Y \sim N(0, 1)$. However, $1 \cdot X + 1 \cdot Y = X + (-X) = 0$ which is not normal.

Theorem 24.4

Refer to Theorem 1, Practice 16.

Practice 24.2. 16 – Problem 1:

- Recall that if X, Y are independent then $\text{Cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right] = 0$. However, in general, $\text{Cov} = 0$ does not imply independence.
- Now, consider (X, Y) bivariate normal, Independence $= f(x, y) = f_X(x) f_Y(y)$. Since $\text{Cov}(X, Y) = 0$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = 0$$

Now,

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x - \mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X\sigma_Y} + \frac{(y - \mu_Y)^2}{\sigma_Y^2}\right]\right\} \\ &= \dots \\ &= f_X(x) f_Y(y) \end{aligned}$$

Practice 24.3. 16 – Problem 2: See §4.5, textbook.

§24.3 Functions of One Dimension RV

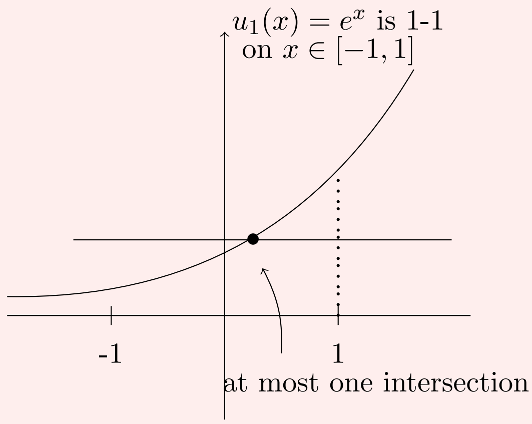
Question 24.1. Given a continuous RV X with pdf $f(x)$ and S_X, define

$$Y = u(X)$$

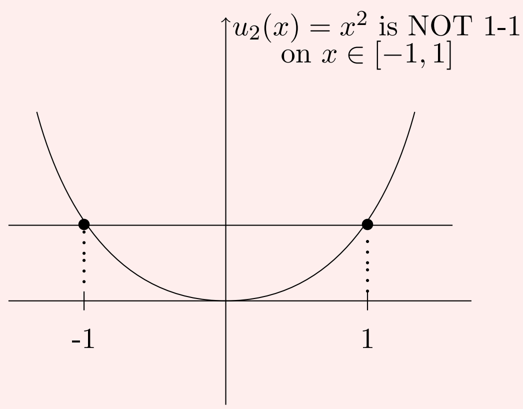
where $u(x)$ is a one-to-one and increasing function on S_X. Find the distribution of Y?

Example 24.5 (1-1 Function)

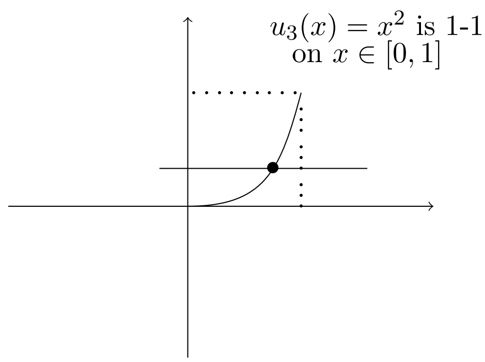
$u_1(x) = e^x$ is 1-1 on $x \in [-1, 1]$.



$u_2(x) = x^2$ is NOT 1-1 on $x \in [-1, 1]$.



Defn of 1-1: The eqn $u(x) = a$ for a constant has at most one root. Note that



Remark 24.6. If $y = u(x)$ is 1-1 on S_X, then it admits an inverse function $x = v(y)$.

Example 24.7

$y = u_3(x) = x^2$ on $[0, 1]$. Then $x = \sqrt{y}$ on $y \in [0, 1]$.

Distribution Technique:

- Find the cdf of $Y = u(X)$

$$\begin{aligned} P(Y \le y) &= P\left(u(X) \le y\right) \\ &= P(X \le v(y)) \\ &= \int_{-\infty}^{v(y)} f(x)\,dx \end{aligned}$$

- Find the density of Y

$$g(y) = \frac{d}{dy}\left[\int_{-\infty}^{v(y)} f(x)\,dx\right]$$

  Define

$$F(y) = \int_{-\infty}^{y} f(x)\,dx \implies F'(y) = f(y)$$

  Then

$$\begin{aligned} \int_{-\infty}^{v(y)} f(x)dx &= F\left(v(y)\right) \\ g(y) = \frac{d}{dy}F(v(y)) &= F'(v(y))v'(y) \\ g(y) &= f(v(y))v'(y) \end{aligned}$$

Change-of-variable Technique:

Given $u(x)$ 1-1 function on S_X then $Y = u(X)$ has a density $g(y) = f\left(v(y)|v'(y)\right)$ where $v(y) = x$ is the inverse of $y = u(x)$.

§25 Lec 23: Nov 30, 2020

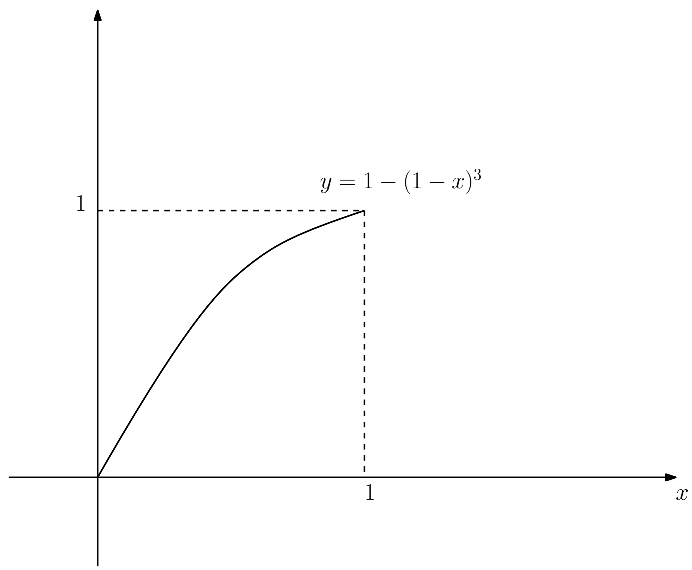
§25.1 Lec 22 (Cont'd)

Recall: Given $u(x)$ 1-1 function of f_X then $Y = u(X)$ has a density

$$g(y) = f\left(v(y)\right) \cdot \left|v'(y)\right|$$

where $v(y) = x$ is the inverse of $y = u(x)$.

Practice 25.1. 17 – Problem 2: $f(x) = 3(1-x)^2, 0 < x < 1$. So, $y = 1-(1-x)^3$



- $S_Y = 0 < y < 1$
- Solve for x in terms of y.

$$\begin{aligned} y &= 1-(1-x)^3 \\ 1-x &= (1-y)^{\frac{1}{3}} \\ v(y) = x &= 1-(1-y)^{\frac{1}{3}} \end{aligned}$$

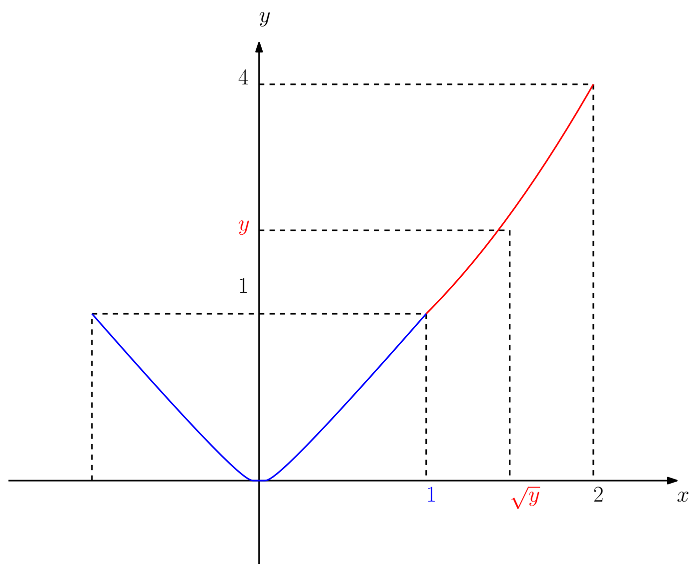
- $$\begin{aligned} g(y) &= f(v(y)) \cdot \left|v'(y)\right| \\ &= 3\left(1-(1-(1-y)^{\frac{1}{3}})\right)^2 + \frac{1}{3}(1-y)^{-\frac{2}{3}} \\ &= 1 \end{aligned}$$

Thus, $Y \sim$ Unif$(0,1)$.

Question 25.1. What if u is not 1-1?

Ans: No universal approach – Use case-by-case basis which is using the cdf approach.

Practice 25.2. 17 – Problem 3:



- $S_Y = \{0 < y < 4\}$
- 1-1 part: $1 \leq y \leq 4$

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(X \leq \sqrt{y}) \end{aligned}$$

Thus,

$$\begin{aligned} g(y) &= f\left(v(y)\right) \cdot \left|v'(y)\right| \\ &= \frac{2}{9}(\sqrt{y}+1) \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{1}{9}\frac{\sqrt{y}+1}{\sqrt{y}} \end{aligned}$$

- Non 1-1 part: $0 \leq y \leq 1$. Cdf technique:

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(|X| \leq \sqrt{y}) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) \\ &= F(\sqrt{y}) - F(-\sqrt{y}) \end{aligned}$$

Thus,

$$\begin{aligned} g(y) &= \frac{d}{dy}\left[F(\sqrt{y}) - F(-\sqrt{y})\right] \\ &= \frac{2}{9}(\sqrt{y}+1) \cdot \frac{1}{2\sqrt{y}} - \frac{2}{9}(-\sqrt{y}+1) - \frac{1}{2\sqrt{y}} \\ &= \frac{2}{9\sqrt{y}} \end{aligned}$$

In summary,

$$g(y) = \begin{cases} \frac{1}{9}\frac{\sqrt{y}+1}{\sqrt{y}}, 1 \le y \le 4 \\ \frac{2}{9\sqrt{y}}, 0 \le y \le 1 \end{cases}$$

§25.2 Transformations of 2 Random Variables

Partial Derivatives & Jacobian matrix: Let $f(x,y), g(x,y)$ be 2 functions.

1. Partial derivative :

   $$\frac{\partial}{\partial x} f(x,y) := \frac{d}{dx} f(x,y)$$

   Similarly,

   $$\frac{\partial}{\partial y} f(x,y) := \frac{d}{dy} f(x,y)$$

> **Example 25.1**
>
> $f(x,y) = \sin(x \cdot y^2)$
>
> $$\frac{\partial f}{\partial x} = \cos(x \cdot y^2) y^2$$
> $$\frac{\partial f}{\partial y} = \cos(xy^2) \cdot 2xy$$

2. Jacobian Matrix of f, g

   $$J(x,y) := \begin{pmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \\ \frac{\partial g}{\partial x}(x,y) & \frac{\partial g}{\partial y}(x,y) \end{pmatrix}$$

   determinant of J is defined as

   $$\det J = \frac{\partial f}{\partial x} \cdot \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \cdot \frac{\partial g}{\partial x}$$

   is a function of x and y.

Problem: Suppose X, Y is a bivariate random variables with joint pdf $f(x,y)$. Suppose

$$U = u(X,Y)$$
$$V = v(X,Y)$$

Under what condition of $u(x,y)$ and $v(x,y)$, can we determine the joint pdf of (U,V)?
Change-of-variable technique: Suppose $u(x,y), v(x,y)$ have inverse functions, i.e., x, y can be solved in terms of u and v

$$x = x(u,v)$$
$$y = y(u,v)$$

Then, $g(u,v)$ the joint pdf of (U,V) is given by the formula

$$g(u,v) = f\left(x(u,v), y(u,v)\right) \cdot |\det J|$$

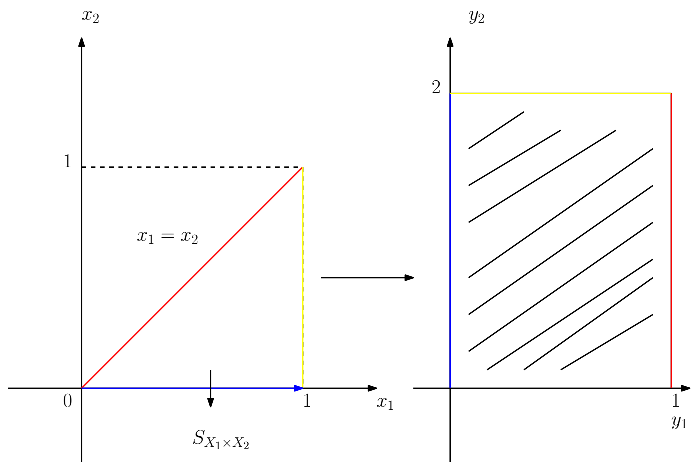
where $\det J = |\frac{\partial x}{\partial u} \cdot \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \cdot \frac{\partial y}{\partial u}|$. For 1-1 function $Y = u(X)$:

$$g(y) = f(v(y)) \cdot |v'(y)|$$

where $v(y) = x$ is the inverse of $x = u(y)$.

Practice 25.3. 18 – Problem 1:

1. Step 1: Find $S_{Y_1 \times Y_2}$



   - $x_1 \in (0, 1)$ and $x_2 = 0$, then

$$y_1 = \frac{x_2}{x_1} = 0$$
$$y_2 = 2x_1 \in (0, 2)$$

   - $x_1 = 1, 0 < x_2 < 1$ then

$$y_1 = \frac{x_2}{x_1} = x_2 \in (0, 1)$$
$$y_2 = 2x_1 = 2$$

   - $x_1 = x_2 \in (0, 1)$ then

$$y_1 = \frac{x_2}{x_1} = 1$$
$$y_2 = 2x_1 \in (0, 2)$$

   Thus,

$$S_{Y_1 \times Y_2} = \begin{cases} 0 < y_1 < 1, \\ 0 < y_2 < 2 \end{cases}$$

2. Step 2: Find the inverse function

$$x_1 = x_1(y_1, y_2)$$
$$x_2 = x_2(y_1, y_2)$$

$$\begin{cases} y_1 &= \frac{x_2}{x_1} \\ y_2 &= 2 \cdot 1 \end{cases} \implies \begin{cases} x_2 &= x_1 \cdot y_1 \\ x_1 &= \frac{y_2}{2} \end{cases}$$
$$\implies \begin{cases} x_2 &= \frac{y_2}{2} \cdot y_1 \\ x_1 &= \frac{y_2}{2} \end{cases}$$

3. Step 3: Find $|\det J|$

$$\det J = \frac{y_2}{4}$$

4. Step 4: joint pdf of (Y_1, Y_2)

$$\begin{aligned} g(y_1, y_2) &= f\left(x_1(y_1, y_2), x_2(y_1, y_2)\right) |\det J| \\ &= \frac{y_2}{2} \end{aligned}$$

Note: g is indeed a joint pdf

$$\int_0^2 \int_0^1 g dy_1 dy_2 = \int_0^2 \int_0^1 \frac{y_2}{2} dy_1 dy_2 = 1$$

§26 Lec 24: Dec 2, 2020

§26.1 Change of Variable Techniques to find pdf

1–Dimension: Given X, $Y = u(X)$

1. Find S_Y
2. If $u(x($ is 1-1, find the inverse x in terms of y, i.e.,

$$x = v(y)$$

3. Pdf of Y

$$g(y) = f\left(v(y)\right) |v'(y)|$$

where f is pdf of X.

2–Dimension: given (X, Y)

$$\begin{aligned} U &= u(x, y) \\ V &= v(x, y) \end{aligned}$$

1. Find $S_{U \times V}$
2. Find the inverse (x, y) in terms of u, v i.e.,

$$\begin{aligned} x &= x(u, v) \\ y &= y(u, v) \end{aligned}$$

3. Pdf of (u, v) is

$$g(u, v) = f\left(x(u, v), y(u, v)\right) \cdot |\det J|$$

where $J = \frac{\partial(x,y)}{\partial(u,v)}$ and

$$\det J = \frac{\partial x}{\partial u} \cdot \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \cdot \frac{\partial y}{\partial u}$$

§26.2 Several Random Variables

Definition 26.1 (Independent and Identically Distributed Sequence) — Given a sequence of RVs, $X_1, X_2, \ldots$ that are (mutually) independent and have same distribution

1. $\{X_k\}_{k \geq 1}$ is called an "i.i.d" (independent and identically distributed) sequence of RVs.

2. A finite sub collection $\{X_1, \ldots, X_n\}$ from the above i.i.d sequence is called " a random sample of size n".

Theorem 26.2
$X_1, \ldots, X_n$ are independent if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n)$$

where $f, f_i, i = 1, \ldots, n$ are either joint pdf and marginal pmf (for discrete) or joint pdf and marginal pdf (for continuous).

Theorem 26.3
If $X_1, X_2, \ldots, X_n$ are independent then for $u_1(x_1), \ldots, u_n(x_n)$ are functions,

$$E\left[\prod_{i=1}^{n} u_i(X_i)\right] = \prod_{i=1}^{n} E\left[u_i(X_i)\right]$$

Proof. A consequence of the above theorem (see textbook §5.3). □

Theorem 26.4
If $X_1, \ldots, X_n$ are independent with means $\mu_1, \ldots, \mu_n$, and variances $\sigma_1^2, \ldots, \sigma_n^2$. Let $Y = \sum_{i=1}^{n} a_i X_i$ where $a_i \in \mathbb{R}$ if constant. Then

1. $E[Y] = \sum_{i=1}^{n} a_i \mu_i$

2. $\text{Var}[Y] = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

Proof. 1.

$$\begin{aligned} E[Y] &= E\left[\sum a_i X_i\right] \\ &= \sum a_i E[X_i] \\ &= \sum a_i \mu_i \end{aligned}$$

2. $\text{Var}[Y] = E\left[Y - \mu_Y\right]^2$. We have

$$\begin{aligned}(Y-\mu_Y)^2 &= \left(\sum a_i X_i - \sum a_i \mu_i\right)^2 \\ &= \left(\sum_{i=1}^n a_i (X_i - \mu_i)\right)^2 \\ &= \sum_{i=1}^n a_i^2 \left(X_i - \mu_i\right)^2 + 2 \sum_{1 \le j, i \le n, i \ne j} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\end{aligned}$$

Thus,

$$E\left(Y - \mu_Y\right)^2 = \sum_{i=1}^n a_i^2 \sigma_i^2 + 2 \sum_{i \ne j} a_i a_j \text{Cov}(X_i, X_j)$$

So, $\text{var}(Y) = \sum a_i^2 \sigma_i^2$.

□

Remark 26.5. Independence was only employed in $\text{Var}(Y)$. $E[Y] = \sum a_i \mu_i$ always holds regardless of independence.

Example 26.6

Suppose $\{x_k\}$ is iid with mean μ and variance σ^2. Consider

$$\overline{X_n} := \frac{X_1 + \ldots + X_n}{n} = \sum_{i=1}^n \frac{1}{n} X_i$$

Then

$$\begin{aligned}E\left[\overline{X_n}\right] &= \sum_{i=1}^n \left[\frac{1}{n}\mu\right] = \mu \\ \text{Var}\left[\overline{X_n}\right] &= \sum_{i=1}^n \frac{1}{n^2}\sigma^2 = \frac{sima^2}{n}\end{aligned}$$

Practice 26.1. 19 – Problem 1:

$$\begin{gathered}X_1 \sim \text{ Exp}(\lambda_1), f_1(x) = \lambda_1 e^{-\lambda_1 x}, \\ X_2 \sim \text{ Exp}(\lambda_2), f_2(x) = \lambda_2 e^{-\lambda_2 x}, x > 0\end{gathered}$$

a) $W = \min(X_1, X_2)$.
<u>Goal</u>: Find pdf of W from the cdf of w. Note that $S_W = \{w > 0\}$ since $w = \min\{x_1, x_2\}$

and $x_1, x_2 > 0$.

$$\begin{aligned} P(W > t) &= P(\min(X_1, X_2) > t) \\ &= P(\text{both greater than t}) \\ &= P(\{X_1 > t\} \cap \{X_2 > t\}) \\ &= P(X_1 > t)P(X_2 > t) \\ &= e^{-\lambda_1 t} e^{-\lambda_2 t} \\ &= e^{-(\lambda_1 + \lambda_2)t} \\ &= P(w > t) \end{aligned}$$

Thus, $W \sim \text{ Exp}(\lambda_1 + \lambda_2)$.

> **Remark 26.7.**
> 1. If $X_i \sim \text{ Exp}(\lambda_i), i = 1, \ldots, n$ are independent. Then $W_n = \min\{X_1, \ldots, X_n\} \sim \text{Exp}(\lambda_1 + \ldots + \lambda_n)$.
> 2. $$\begin{aligned} \{\min(X_1, \ldots, X_n) > t\} &= \{X_1 > t\} \cap \ldots \cap \{X_n > t\} \\ \{\max(X_1, \ldots, X_n) < t\} &= \{X_1 < t\} \cap \ldots \cap \{X_n < t\} \end{aligned}$$

I is a Bernoulli dist. $I \in \{1, 2\}$.

$$\begin{aligned} P(I = 1) &= P\left(\min\{X_1, X_2\} = X_1\right) \\ &= P(X_1 < X_2) \end{aligned}$$

By independence, joint pdf

$$\begin{aligned} f(x_1, x_2) &= f_1(x_1) f_2(x_2) \\ &= \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 x_2} \end{aligned}$$

So,

$$\begin{aligned} \iint_{x_1 < x_2} f(x_1, x_2)\, dx_1\, dx_2 &= P(X_1 < X_2) \\ \int_0^\infty \int_0^{x_2} \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 x_2}\, dx_1\, dx_2 &= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} (1 - e^{-\lambda_1 x_2})\, dx_2 \\ &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \\ &= P(I = 1) \end{aligned}$$

> **Remark 26.8.** If $X_i \sim \text{ Exp}(\lambda_1$, then
> $$P(\min\{X_1, \ldots, X_n\} = X_j) = \frac{\lambda_j}{\lambda_1 + \ldots + \lambda_n}$$

§27 Lec 25: Dec 4, 2020

§27.1 MGF Technique

Recall mgf $M(t)$ determines the distribution, e.g.,

$$M(t) = e^{\mu t + \sigma^2 \frac{t^2}{2}}$$

$\implies X \sim N(\mu, \sigma^2)$.

Theorem 27.1

Given $X_1, \ldots, X_n$ are independent RVs with MGF $M_1(t), \ldots, M_n(t)$, respectively, define $Y = a_1X_1 + \ldots + a_nX_n$ where $a_i's$ are real constants. Then

$$M_Y(t) = \prod_{i=1}^{n} M_i(a_it)$$

Proof.

$$\begin{aligned} M_Y(t) &= E\left[e^{tY}\right] \\ &= E\left[e^{t\sum a_iX_i}\right] \\ &= E\left[e^{\sum a_itX_i}\right] \\ &= E\left[e^{a_1tX_1}\right] \ldots E\left[e^{a_ntX_n}\right] \\ &= M_1(a_1t)M_2(a_2t)\ldots M_n(a_nt) \end{aligned}$$

□

Corollary 27.2

If $\{X_k\}_{k\geq 1}$ iid with the same mgf $M_X(t)$

1. IF $Y = \sum_{i=1}^n X_i, (a_i = 1)$ then $M_Y(t) = [M_X(t)]^n$
2. If $\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n} = \sum_{i=1}^n \frac{1}{n} X_i$. Then

$$M_{\overline{X_n}}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$$

Example 27.3 1. $\{X_i\}$ iid Bernoulli RVs

$$X_i = \begin{cases} 1 \text{ wp } p \in (0,1) \\ 0 \text{ wp } 1-p \end{cases}$$

$Y = \sum_{i=1}^n X_i = \#$ of successes in n trials ($\sim$ Binom$(0,1)$). We have the mgf of

X_i given by

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

By the above Corollary, $Y = \sum_{i=1}^n X_i$

$$\begin{aligned} M_Y(t) &= [M_X(t)]^n \\ &= \left[pe^t + 1 - p\right]^n \end{aligned}$$

which is the MGF of Binom(n,p). Thus,

$$Y \sim \text{ Binom}(n, p).$$

Example 27.4 2. X_i iid $\sim$ Exp(λ). Recall from Poisson process with rate λ

- Arrival of first guest

$$T_1 \sim \text{ Exp}(\lambda) = \text{ Gamma}(1, \lambda)$$

- arrivals of nth guest

$$T_n \sim \text{ Gamma}(n, \lambda)$$

-

$$\begin{aligned} x_2 &= T_2 - T_1 \sim \text{ Exp}(\lambda) \\ x_n &= T_n - T_{n-1} \sim \text{ Exp}(\lambda) \end{aligned}$$

$x_1, \ldots, x_n$ are indep. i.e., $\{x_k\}$ iid $\sim$ Exp(λ).

Now, $T_n = \sum_{k=1}^n X_k \sim$ Gamma(n, λ). Mgf approach:

$$\begin{aligned} M_{X_k}(t) &= E\left[e^{tX}\right] \\ &= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx \\ &= \frac{\lambda}{\lambda - t} \\ &= \frac{1}{1 - \frac{t}{\lambda}} \\ M_{T_n}(t) &= [M_{X_k}(t)]^n \\ &= \left[\frac{1}{1 - \frac{t}{\lambda}}\right]^n \end{aligned}$$

Thus, $T_n \sim$ Gamma(n, λ).

Practice 27.1. 19 – Problem 2: Similar to 5.2-12 if using cdf approach.

$$X \sim N(\mu_1, \sigma_1^2), M_X(t) = e^{\mu_1 t + \sigma_1^2 \frac{t^2}{2}}$$
$$Y \sim N(\mu_2, \sigma_2^2), M_Y(t) = e^{\mu_2 t + \sigma_2^2 \frac{t^2}{2}}$$

Mgf: By independence,

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \\ &= e^{\mu_1 t + \sigma^2 \frac{t^2}{2}} e^{\mu_2 t + \sigma_2^2 \frac{t^2}{2}} \\ &= e^{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)\frac{t^2}{2}} \end{aligned}$$

Thus,

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

> **Remark 27.5.** Since X, Y are independent, (X, Y) bivariate normal.
>
> $$\begin{aligned} f(x,y) &= f_X(x) f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\left[\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(y-\mu_2)^2}{2\sigma_2^2}\right]} \end{aligned}$$
>
> Recall that biv. normal means any $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_x^2 + b^2\sigma_y^2$. Thus, $X + Y = 1 \cdot X + 1 \cdot Y \sim N(\mu_X = \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Practice 27.2. 20 – Problem 1: Recall Gamma(α, β) then $\chi^2(r) = \text{Gamma}(\frac{r}{2}, \frac{1}{2})$.
a) $X \sim N(0,1), X^2 \sim ?$. Mgf:

$$\begin{aligned} M_{X^2}(t) &= E\left[e^{tX^2}\right] \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tx^2} e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int e^{-(\frac{1}{2} - t)x^2}\, dx \\ &= \dots \\ &= \left(\frac{1}{1-2t}\right)^{\frac{1}{2}} \sim \text{MGF of } \chi^2(r=1) \end{aligned}$$

b) $\{X_k\}$ iid $\sim N(0,1)$ then X_k^2 are independent and $\sim \chi^2(1)$.

$$\begin{aligned} Y &= \sum_{k=1}^n X_k^2 \\ M_Y(t) &= \left[M_{X_k^2}(t)\right]^n \\ &= \left[\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}\right]^n \\ &= \left(\frac{1}{1-2t}\right)^{\frac{n}{2}} \end{aligned}$$

Thus, $Y \sim \chi^2(n)$.

Note: There is a similar but more advance method called "characteristics function" using complex analysis.

§28 Lec 26: Dec 7, 2020

§28.1 Random Functions Associated with Normal Distributions

Recall that

1. $X \sim N(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma} \sim N(0,1)$.
2. If $\{X_k\}$ iid $\sim N(0,1)$ then

$$Y = \sum_{k=1}^{n} X_k^2 \sim \chi^2(n)$$

Definition 28.1 (Sample Mean & Sample Variance) — Given $\{X_k\}$ iid with mean μ& variance σ^2.

- Sample mean

$$\overline{X_n} := \frac{X_1 + \ldots + X_n}{n} = \frac{1}{n} \sum_{k=1}^{n} X_k$$

- Sample variance:

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left(X_k - \overline{X_n}\right)^2$$

Remark 28.2. 1.

$$\begin{aligned} E[\overline{X_n}] &= \frac{1}{n} \sum_{k=1}^{n} E[X_k] \\ &= \mu \end{aligned}$$

2. $E[S^2] = \sigma^2$.

Theorem 28.3
If $\{X_k\}$ independent, $X_k \sim N(\mu_k, \sigma_k^2)$, then

$$Y = \sum_{k=1}^{n} a_k X_k \sim N\left(\sum_{k=1}^{n} a_k \mu_k, \sum_{k=1}^{n} a_k \sigma_k^2\right)$$

Proof. MGF technique: Recall $X_K \sim N(\mu_k, \sigma_k^2)$. MGF:

$$M_k(t) = e^{\mu_k t + \frac{\sigma_k^2}{2} t^2}$$

Now,

$$\begin{aligned}M_Y(t) &= E\left[e^{t\sum a_k X_k}\right] \\ &= E\left[e^{ta_1X_1}\right]\dots E\left[e^{ta_nX_n}\right] \\ &= e^{\mu_1 a_1 t + \frac{\sigma_1^2}{2}a_1^2t^2}\dots e^{\mu_n a_n t + \frac{\sigma_n^2}{2}a_n^2t^2} \\ &= e^{(\sum \mu_k a_k)t + \left(\sum \sigma_k^2 a_k^2\right)\frac{t^2}{2}}\end{aligned}$$

Thus,

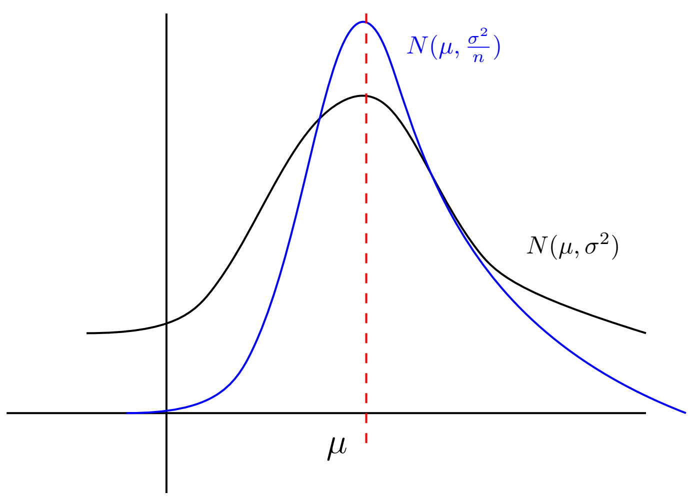
$$Y \sim N\left(\sum a_k\mu_k, \sum a_k^2\sigma_k^2\right)$$

□

Corollary 28.4

If X_k iid $\sim N(\mu, \sigma^2)$, then

$$\overline{X_n} = \sum_{k=1}^{n} \frac{1}{n} X_k \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



"$\overline{X_n}$ is used to estimate μ in practice"

Fact 28.1.
1. If $\{X_k\}$ iid $\sim N(\mu, \sigma^2)$ then for all $k = 1, \dots, n$, $X_k - \overline{X_n}$ independent of $\overline{X_n}$.

   Proof. (sketch) prove for $n = 2$

   $$\overline{X_n} = \overline{X_2} = \frac{X_1 + X_2}{2}$$

   and $k = 1, 2$

   $$\begin{aligned}X_1 - \overline{X_2} &= \frac{X_1 - X_2}{2} \\ X_2 - \overline{X_2} &= \frac{X_2 - X_1}{2}\end{aligned}$$

Goal: Show $X_1 - X_2$ is independent of $X_1 + X_2$. Using change-of-variable technique. ☐

2. As a consequence,

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2 \text{ is independent of } \overline{X_n}$$

Theorem 28.5 (Practice 20, # 2)

$$\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$$

Proof. We have

$$\begin{aligned}
\frac{n-1}{\sigma^2}\delta^2 &= \sum_{k=1}^{n}\frac{\left(X_k - \overline{X_n}\right)^2}{\sigma^2} \\
&= \sum \frac{(x_k - \mu + \mu - \overline{x_n})^2}{\sigma^2} \\
&= \sum \frac{(x_k-\mu)^2}{\sigma^2} + \frac{2(x_k-\mu)(\mu-\overline{x_n})}{\sigma^2} + \frac{(\mu-\overline{x_n})^2}{\sigma^2} \\
&= \sum_{k=1}^{n}\left(\frac{x_k-\mu}{\sigma}\right)^2 + \frac{2}{\sigma^2}\sum_{k=1}^{n}(x_k-\mu)(\mu-\overline{x_n}) + n\frac{(\mu-x_n)^2}{\sigma^2} \\
&= \dots \\
&= \sum\left(\frac{x_k-\mu}{\sigma}\right)^2 - n\left(\frac{x_n-\mu}{\sigma}\right)^2
\end{aligned}$$

Thus, $\frac{n-1}{\sigma}\delta^2 = \sum\left(\frac{x_k-\mu}{\sigma}\right)^2 - n\left(\frac{\overline{x_n}-\mu}{\sigma}\right)^2$. It follows that

$$\frac{n-1}{\sigma^2}\delta^2 + n\left(\frac{\overline{x_n}-\mu}{\sigma}\right)^2 = \sum\left(\frac{x_k-\mu}{\sigma}\right)^2$$

i.e.,

$$\frac{n-1}{\sigma^2}\delta^2 + \left(\frac{\overline{x_n}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 = \sum\left(\frac{x_k-\mu}{\sigma}\right)^2$$

Recall $\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n}) \implies \frac{\overline{X_n}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ and

$$\left(\frac{\overline{X_n}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi^2(1)$$

$X_k \sim N(\mu, \sigma^2) \implies \frac{X_k - \mu}{\sigma} \sim N(0,1)$ and $\sum_{k=1}^{n} \left(\frac{x_k-\mu}{\sigma}\right)^2 \sim \chi^2(n)$. Thus, since A independent of B, by MGF

$$\begin{aligned} E\left[e^{t(A+B)}\right] &= E\left[e^{tA}\right] E\left[e^{tB}\right] \\ E\left[e^{tC}\right] &= E\left[e^{tA}\right] E\left[e^{tB}\right] \\ \frac{1}{(1-2t)^n} &= E\left[e^{tA}\right]\left(\frac{1}{1-2t}\right)^1 \end{aligned}$$

Hence,

$$E\left[e^{tA}\right] = \left(\frac{1}{1-2t}\right)^{n-1}$$

and thus

$$A = \frac{n-1}{\sigma^2}\delta^2 \sim \chi^2(n-1)$$

□

Practice 28.1. 20 – Problem 3: See Thm 5.5-3, §5.5 Textbook. $P(T < t)$ since Z independent of U

$$S_{Z\times U} = \{-\infty < z < \infty, 0 < u\}$$

and

$$f(z,u) = f_Z(z) f_U(u)$$

And

$$\{T < t\} = \left\{\frac{z}{\sqrt{\frac{u}{r}}} < t\right\} = \left\{z < t\sqrt{\frac{u}{r}}\right\}$$

So,

$$\begin{aligned} P(T<t) &= P(Z < \frac{t}{\sqrt{r}}\sqrt{U}) \\ &= \int_0^\infty \int_{-\infty}^{\frac{t}{\sqrt{r}}\sqrt{u}} f(z,u)\,dz\,du \end{aligned}$$

Thus, the pdf of T is given by

$$\frac{d}{dt}\left[\int_0^\infty \int_{-\infty}^{\frac{t}{\sqrt{r}}\sqrt{u}} f(z,u)\,dz\,du\right]$$

§29 Lec 27: Dec 9, 2020

§29.1 Central Limit Theorem

Recall

1. If (X,Y) biv. normal then $\text{Cov}(X,Y) = 0 \implies X, Y$ independent.

2. If $\{X_1\}$ iid $\sim N(\mu, \sigma^2)$ then for all $i = 1, \dots, n$, $X_i - \overline{X}_n$ is independent of $\overline{X}_n$

$$\overline{X}_n = \frac{\sum_{i=1}^n x_i}{n}$$

and sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \text{ independent of } \overline{X}_n$$

Question 29.1. Why "normal"?

Recall: If $\{X_k\}$ iid with mean μ and variance σ^2 then

$$\begin{aligned} E[\overline{X}_n] &= E\left[\frac{1}{n}\sum_{k=1}^n X_k\right] = \mu \\ \text{Var}[\overline{X}_n] &= \text{Var}\left[\frac{1}{n}\sum X_u\right] \\ &= \sum_{k=1}^n \text{Var}[X_k] \\ &= n \cdot \frac{1}{n^2}\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Definition 29.1 (Converge in Distribution) — A sequence of RV $\{X_n\}_{n\geq 1}$ is said to "converge in distribution" to a RV X if for all fixed $t \in \mathbb{R}$,

$$P(X_n \leq t) \overset{n\to\infty}{\longrightarrow} P(X \leq t)$$

i.e., $F_n(t) \overset{n\to\infty}{\longrightarrow} F(t)$ where F_n : cdf of X_n, F : cdf of X.

Lemma 29.2

If $X_n \overset{\text{in distribution}}{\longrightarrow} X$ then for all fixed $s < t$

$$P(s < X_n \leq t) \overset{n\to\infty}{\longrightarrow} P(s < X \leq t)$$

Proof.

$$\begin{aligned} P(s < X_n \leq t) &= P(X_n \leq t) - P(X_n \leq s) \\ &= P(X \leq t) - P(X \leq s) \text{ as } n \to \infty \\ &= P(s < X \leq t) \end{aligned}$$

□

Remark 29.3. "Convergence in Dist" does not indicate anything about independence, space of values, etc between $\{X_n\}$ and X. It only says some limiting behavior of the functions cdf.

Theorem 29.4 (Central Limit)

Suppose $\{X_k\}_{k\geq 1}$ iid with mean μ. Then $\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}$, where $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$, converges in distribution to $Z \sim N(0,1)$, i.e., for all $t \in \mathbb{R}$,

$$P\left(\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} \leq t\right) \overset{n\to\infty}{\longrightarrow} P(Z \leq t) = \Phi(t)$$

where $\Phi(t)$ is the cdf of $N(0,1)$.

Proof. (Skipped) □

Remark 29.5.

1. All sample means $\overline{X}_n$ with the right scaling must "normalize" to $N(0,1)$. Hence, the name "normal".
2. We note that

$$\begin{aligned}\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} &= \frac{\frac{1}{n}\sum_{k=1}^n X_u - \mu}{\frac{\sigma}{\sqrt{n}}}\cdot\frac{n}{n}\\ &= \frac{\sum_{k=1}^n X_u - n\mu}{\sigma\sqrt{n}}\end{aligned}$$

Define $S_n := \sum_{k=1}^n X_k$, then CLT can be stated as follows:

$$P\left(\frac{S_n-n\mu}{\sigma\sqrt{n}} \leq t\right) \overset{n\to\infty}{\longrightarrow} P(Z \leq t) = \Phi(t)$$

Practice 29.1. 21 – Problem 1: $n = 25, \mu = 15, \sigma^2 = 4$, $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k, \{X_k\}$ iid

$$\begin{aligned}P(14.4 < \overline{X}_n < 15.6) &= P\left(\frac{14.4-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{15.6-\mu}{\frac{\sigma}{\sqrt{n}}}\right)\\ &\approx P(\frac{14.4-15}{\frac{2}{5}} < Z < \frac{15.6-15}{\frac{2}{5}}\\ &= P(-1.5 < Z < 1.5)\\ &= P(Z < 1.5) - P(Z < -1.5)\\ &= P(Z < 1.5) - [1 - P(Z < 1.5)]\\ &= 2\Phi(1.5) - 1\end{aligned}$$

Practice 29.2. 21 – Problem 2: $X_k = \#$ sick days of a worker, $\{X_k\}$ iid, $\mu = 10, \sigma = 2, n = 20, S_n = \sum_{k=1}^n X_k$. Denote $A = \#$ sick days budgeted by the firm.

$$P(S_n > A) < 0.2$$

Question 29.2. Find A

We have

$$\begin{aligned} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{A - n\mu}{\sigma\sqrt{n}}\right) &< 0.2 \\ \approx \left(Z > \frac{A - n\mu}{\sigma\sqrt{n}}\right) &< 0.2 \\ P(Z > .85) = .1977 &< .2 \\ P(Z > .84) = .2005 &> .2 \end{aligned}$$

Thus,

$$\begin{aligned} \frac{A - n\mu}{\sigma\sqrt{n}} &\approx .85 \\ \frac{A - 20 \cdot 10}{\sigma\sqrt{20}} &\approx .85 \end{aligned}$$

Thus,

$$\begin{aligned} A &\approx .85 \cdot 2\sqrt{20} + 200 \\ &= 233.65 \\ &\approx 234 \text{ sick days the firm should budget} \end{aligned}$$

Practice 29.3. 21 – Problem 3:

$$\begin{aligned} E[X_k] = 30, &\quad \text{Var}(X_k) = 52 \\ E[Y_k] = 50, &\quad \text{Var}(Y_k) = 64 \\ \text{Cov}(X_k, Y_k) &= 14 \end{aligned}$$

$$\begin{aligned} Z_k &= \#\text{ hours a kid watching movies cartoon} \\ &= X_k + Y_k \\ Z &= \sum_{k=1}^{n} Z_k = \text{total } \#\text{ hours} \\ P(197 < Z < 2090) &=? \end{aligned}$$

We have $\{Z_k\}$ iid

$$\begin{aligned} \mu_{Z_k} = E[Z_k] &= E[X_k + Y_k] \\ &= 30 + 50 = 80 \\ \text{Var}(Z_u) &= \text{Var}(X_u + Y_u) \\ &= \text{var } X_u + \text{var } Y_u + 2\text{Cov}(X_k, Y_k) \\ &= 52 + 64 + 2 \cdot 14 \\ &= 144 \end{aligned}$$

So, $\text{Var}(Z_u) = 12^2 = \sigma_{Z_u}^2$. Now, CLT

$$P\left(\dots < \frac{Z - n\mu_{Z_k}}{\sigma_{Z_u}\sqrt{n}} < \dots\right)$$

§30 Lec 28: Dec 11, 2020

§30.1 Chebyshev's Inequality & Convergence in Probability

Theorem 30.1 (Markov's Inequality)

(Assignment 22, # 1) Given a non-negative RV $X(S_X = [0, \infty))$ and a finite n^{th}-moment for some $n, E[X^n] < \infty$, then for all $t > 0$

$$P(X \geq t) \leq \frac{EX^n}{t^n}$$

Remark 30.2. Markov's Inequality is only useful when t is large (hence, "total estimate"). If t is very close to zero,

$$P(X \geq t) \leq \frac{E[X^n]}{t^n} \leftarrow \text{ a large number}$$

Proof. (Of theorem) Cont. RV

$$\begin{aligned}
E[X^n] &= \int_0^\infty x^n f(x)\,dx \\
&= \int_0^t + \int_t^\infty x^n f(x)\,dx \\
&\geq \int_t^\infty x^n f(x)\,dx \\
x \geq t \text{ for } x \in [t, \infty) \\
&\geq \int_t^\infty t^n f(x)\,dx \\
&= t^n \int_t^\infty f(x)\,dx \\
&= t^n P(X \geq t)
\end{aligned}$$

Thus,

$$\frac{E[X^n]}{t^n} \geq P(X \geq t)$$

□

Discrete RV:

$$\begin{aligned} E[X^n] &= \sum_{x>0} x^n p(x), \quad p(x) \text{ is the pmf} \\ &= \sum_{0<x<t} + \sum_{x\geq t} x^n p(x) \\ &\geq \sum_{x\geq t} x^n p(x) \\ &\geq \sum_{x\geq t} t^n p(x) \\ &= t^n \sum_{x\geq t} p(x) \\ &= t^n P(X \geq t) \end{aligned}$$

Thus,...

Corollary 30.3 (Chebyshev's Inequality)
If X is a RV with mean μ and finite variance σ^2, then for all $k \geq 1$,

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

Note: If $k < 1$, then the estimate is trivial.

Proof. The RV $\frac{|X-\mu|}{\sigma}$ is non-negative and the second momemnt

$$E\left[\frac{|X-\mu|}{\sigma}\right]^2 = \frac{E|X-\mu|^2}{\sigma^2} = 1$$

By Markov's Inequality, $(n = 2)$

$$P\left(\frac{|X-\mu|}{\sigma} \geq k\right) \leq \frac{E\left(\frac{|X-\mu|}{\sigma}\right)^2}{k^2} = \frac{1}{k^2}$$

□

Remark 30.4. Chebyshev's Inequality can also be stated as follows, for $k \geq 1$,

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

Replacing k by $\frac{k}{\sigma}$ in the original Chebyshev:

$$P(|x-\mu| \geq \frac{k}{\sigma} \cdot \sigma) \leq \frac{1}{\left(\frac{k}{\sigma}\right)^2}$$

or

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

Practice 30.1. 22 – Problem 2: $\mu = 33, \sigma^2 = 16, \sigma = 4$
b)

$$\begin{aligned} P(|x-\mu| \geq k) &\leq \frac{\sigma^2}{k^2} \\ P(|x-33| \geq 14) &\leq \frac{16}{14^2} = \frac{4}{49} \end{aligned}$$

"In the final, the final answer must a simplified fraction".
a)

$$\begin{aligned} P(23 < X < 43) &= P(23-33 < X-33 < 43-33) \\ &= P(|X-33| < 10) \\ &= 1 - P(|X-33| \geq 10) \\ &\geq 1 - \frac{4}{25} = \frac{21}{25} \end{aligned}$$

Definition 30.5 (Convergence in Probability) — $X_n \to X$ (in probability) if for all $\epsilon > 0, P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

Theorem 30.6
Convergence in Probability implies convergence in distribution.

Remark 30.7. The converse is not true in general, i.e., convergence in distribution does not imply convergence in probability.

Theorem 30.8 (The weak law of Large number)
(Practice 22, # 3) Suppose $\{X_n\}$ iid with mean μ and variance σ^2. Then $\overline{X}_n = \frac{\sum_{k=1}^n X_k}{n} \overset{\text{in prob}}{\longrightarrow} \mu$,i.e., for all $\epsilon > 0$,

$$P\left(|\overline{x}_n - \mu| > \epsilon\right) \overset{n\to\infty}{\longrightarrow} 0$$

Proof. Recall $\{X_n\}$ iid then $E[\overline{X}_n] = \mu$, Var $\overline{X}_n = \frac{\sigma^2}{n}$. By Chebyshev, for all $\epsilon > 0$,

$$P\left(|\overline{x}_n - \mu| > \epsilon\right) \leq \frac{\frac{\sigma^2}{n}}{\epsilon^2} = \frac{1}{n}\frac{\sigma^2}{\epsilon^2} = 0$$

as $n \to \infty$. □