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 `ductuanvu.wordpress.com/notes/`. Let me know through my email if you notice something mathematically wrong/concerning. Thank you!

Contents

§1 Lec 1: Oct 2, 2020

§1.1 Properties of Probability

Definition 1.1 — 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 — 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. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

Note: 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)

$$P(\emptyset) = 1 - P(S)$$
$$= 1 - 1$$
$$= 0$$

(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

$$P(A) \le P(B)$$

(4)

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

(5) Textbook Section 1.1. □

Definition 2.1 — 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 — 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 $= E_1E_2\dots E_n$

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

Permutation/Combination of n objects taken k:

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

Note:"Permutation of size n" = "permutation of size n taken n".

Fact 2.1.

1. The total number of permutation n taken k (order is important here) is denoted by ${}^nP_k$ 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 $S = \{1, \ldots, 8\}$

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

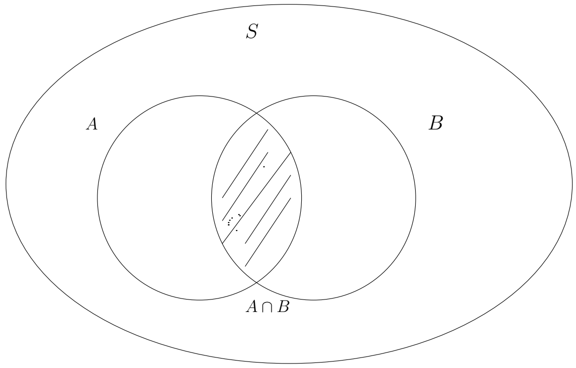
§3.1 Conditional Probability

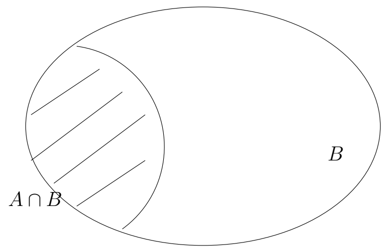
Definition 3.1 — Let $A, B \subseteq S$ be two events. The conditional prob. of A, given that B has occurred with $P(B) > 0$, is defined as

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

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,

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

□

INSERT: PRACTICE 1 #3 here

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

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

Proof. 1. By defn of cond. prob.

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

implying

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

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

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

□

Practice 3.1. The url: `https://ccle.ucla.edu/pluginfile.php/3776692/mod_resource/content/0/Practice%202.pdf`
INSERT: Look at the online notes

§4 Lec 4: Oct 9, 2020

Cont'd (Practice)
3)

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

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

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

(from problem 2 in practice 2)

- $P(A) = \frac{13}{52}$
- $P(B|A) =$, now restricted to outcome space {51 cards 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 — 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)

$\square$

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, $D \subsetneq C = \{\text{first roll is 6}\}$. 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 — 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 — 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$\}$, $\quad P(A) =$? 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 — 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

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

For each $k \in S_x$,

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

Also,

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

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

- $f(x) > 0, \quad x \in S_x$.
- $\sum_{x \in S_x} f(x) = 1$.
- If $A \subseteq S_x$,
$$P(X \in A) = \sum_{x \in A} f(x)$$
<u>*Note*</u>: 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 $k = 1, \ldots, 6$

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

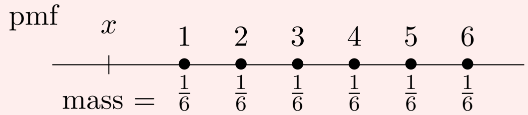
Definition 6.7 — Cumulative distribution function (cdf) $F(x)$ of a RV x is a function given by

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

<u>*Note*</u>: $F(x)$ is usually called distribution function, “cumulative” is dropped.

Example 6.8

Rolling a fair dice



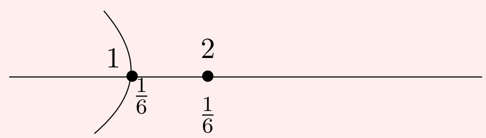
$$\begin{aligned} \text{cdf } F(x) &= P(X \le x) \\ &= \text{total mass cumulated starting from the left up to x} \end{aligned}$$

$x < 1$,

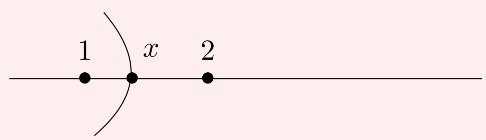
$$\begin{aligned} F(x) &= P(X \le x) \\ &= 0 \text{ (no mass up to } x < 1) \end{aligned}$$

$x = 1$,

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

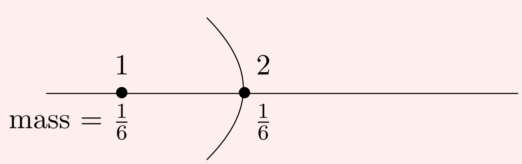


$F(1) = \frac{1}{6}$ (mass up to and including location 1).
$1 < x < 2$



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

$x = 2$



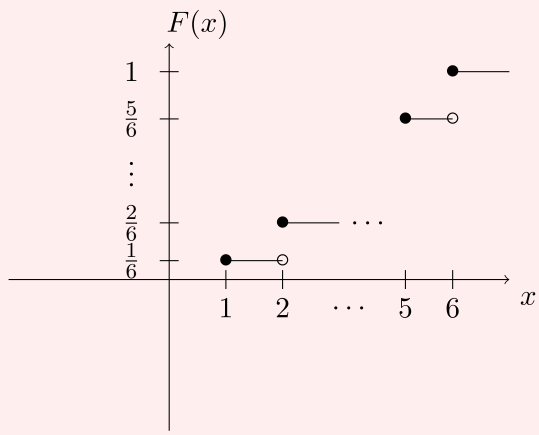
$$\begin{aligned} F(2) &= P(X \le 2) \\ &= P(X = 1) + P(X = 2) \\ &= \frac{2}{6} \end{aligned}$$

Likewise, $2 < x < 3$

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

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

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

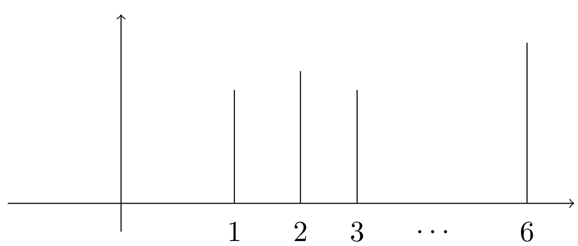


§7 Lec 7: Oct 16, 2020

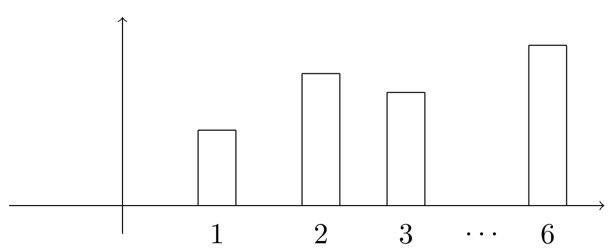
§7.1 Lec 6 (Cont'd)

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:

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

2^{nd} roll / 1^{st} 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

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

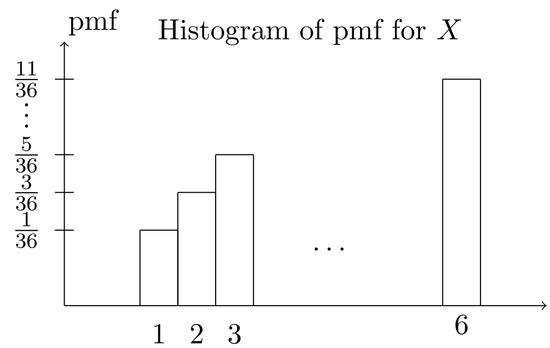
- 2^{nd} approach: for $k = 1, \dots, 6$ (disjoint sub-events)

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

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

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

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

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

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

Theorem 7.3

When it exists, the expectation E satisfies:

- If c is a constant, then
$$E[c] = c$$
- If c is a constant and $u(X)$ is a function, then
$$E[c \cdot u(X)] = cE[u(X)]$$
- If c_1, c_2 are constants and $u_1(X), u_2(X)$ are functions.
$$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$$

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

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

Proof. Textbook. □

Definition 7.5 — 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
$f(x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

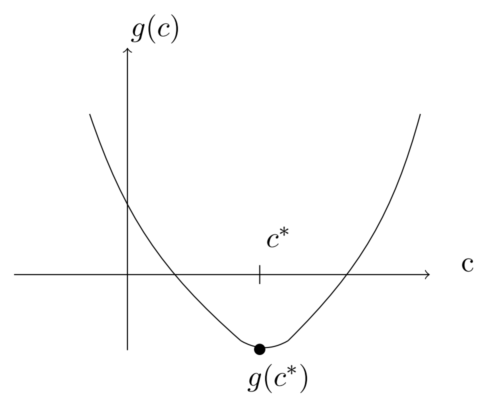
$$\begin{aligned}
\text{mean} = \mu &= E[x] \\
&= \sum_{x \in S_X} x \cdot f(x) \\
&= (-2)\frac{1}{2} + 0\frac{1}{3}1\frac{1}{6} \\
&= -\frac{5}{6} \\
\text{variance} = \sigma^2 &= E[(x-\mu)^2] \\
&= \sum_{x \in S_X} (x-\mu)^2 f(x) \\
&= (-2-(-\frac{5}{6})^2\frac{1}{2} + (0-(-\frac{5}{6}))^2\frac{1}{3}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

σ^2 intepretation:

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

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

"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 Dis 1: Oct 6, 2020

§8.1 Set Theory

Definition 8.1 — A set is a collection of items.

Example 8.2

$$T = \{1, 2, 3, \text{red}, \text{blue}\}$$
$$S = \{1, 3, \text{red}\}$$
$$R = \{1, 2, 4\}$$
$$S \subseteq T$$
$$S' = S^c = \{2, \text{blue}\}$$
$$R \nsubseteq T$$
$$3 \quad \underbrace{\in}_{\text{is an element of}} \quad T$$
$$\{3\} \subseteq T$$

Example 8.3

$$A = \{1, 3, 7\} \qquad A \cup B = \{1, 2, 3, 4, 7\}$$
$$B = \{2, 3, 4\} \qquad A \cap B = \{3\}$$
$$A \setminus B = \{1, 7\} \qquad B \setminus A = \{2, 4\}$$

De Morgan Laws:

$$(A \cup B)' = A' \cap B'$$
$$(A_1 \cup A_2 \cup \ldots \cup A_n) = A_1' \cap A_2' \cap \ldots \cap A_n'$$
$$(A \cap B)' = A' \cup B'$$

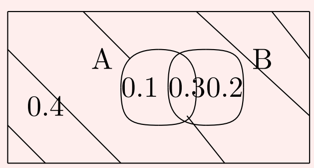
If have a sample space S, and subset of S are called events. A probability function is a function $\mathbb{P}$ that assigns a real number each event with three rules:

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

Example 8.4

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

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



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

Example 8.5

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

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

§9 Dis 2: Oct 13, 2020

1.4.16: An urn has 5 balls. One is marked "win" and the other are marked "lose". You and another player each take balls out one at a time until somebody picks win. You pick first.
W/o replacement: $P(\text{winning}) = \frac{1}{5} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{3}{5}$

With replacement:

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

§9.1 Conditional Probabilities

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

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

Example 9.1

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

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

§9.2 Bayes's Theorem

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

Example 9.2

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

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