Probability Theory Notes (2.3 - 2.4)

September 16, 2020

1 Introduction

Old interpretation of probability: "long term frequency of events."

If E is an event, then P(E) = probability that E occurs = $\lim_{n\to\infty} n(E)/n$, where $n(E)$ represents the number of times E occurs in n trials.

Axioms of Probability:

- **Axiom 1**: $0 \leq P(E) \leq 1$, where $P(E)$ = probability of event E.
- **Axiom 2**: $P(S) = 1$, where S represents the sample space
- **Axiom 3**: For any sequence of mutually exclusive events $E_1, E_2, ..., E_\infty$ (i.e. $E_iE_j = \emptyset$ when $i \neq j$), we have that
$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i).$$

Some Simple Propositions:

- **Prop 1**: $P(E^c) = 1 - P(E)$

  Proof: $S = E \cup E^c$, and $E \cap E^c = \emptyset$
  $\implies P(S) = 1 = P(E \cup E^c) = P(E) + P(E^c)$

- **Prop 2**: If $E \subset F$, then $P(E) \leq P(F)$. (Each outcome in E is also an outcome in F).

  Proof: Since $E \subset F$, we can write
  $$F = E \cup (E^cF) \implies P(F) = P(E) + P(E^cF) \implies P(F) \geq P(E) \textit{ since } P(E^cF) \geq 0$$
  (E & E^cF are mutually exclusive since E & E^c are mutually exclusive.)

- **Prop 3**: $P(E \cup F) = P(E) + P(F) - P(EF)$

  (also known as the principle of inclusion-exclusion)

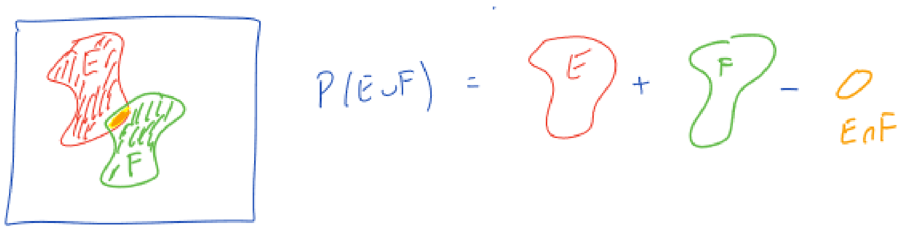

Figure 1: Proposition 3 Venn Diagram.

- **Prop 4**: General inclusion exclusion

  $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i) - \sum_{1\le i_1\le i_2\le n} P(E_{i_1}E_{i_2}) + ... + (-1)^{r+1}\sum_{1\le i_1\le i_2\le ...\le i_r\le n} P(E_{i_1}E_{i_2}...E_{i_r}) + ... + (-1)^{n+1}P(E_1E_2E_3...E_n)$ for all integer $n \ge 2$

  Note that I gave a proof of the general inclusion-exclusion law by induction in the section Problems from Textbook, question 14.

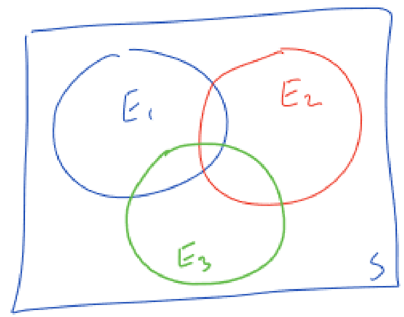

Figure 2: General inclusion-exclusion when n=3.

In Class Examples:

Prop 4.4: (Applying general inclusion-exclusion to Figure 3)

$P(S) = 1$

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

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C) - \cancel{P(A \cap B)} - P(B \cap C) + \cancel{P(A \cap B \cap C)}$. [Note: $P(A \cap B) = 0$ and $P(A \cap B \cap C) = 0$]

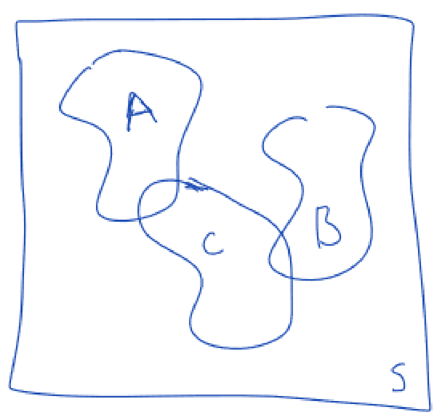

Figure 3: Inclusion-Exclusion for sets A, B, and C

Problem 9 from chapter 2:
24% carry AMEX
61% carry VISA
11% carry both cards

1. What percent carry AMEX or VISA?
   $24\% + 61\% - 11\% = 74\%$.

2. What percent carry AMEX but not VISA?
   $24\% - 11\% = 13\%$

Theoretical Exercise II:
Let $P(E) = 0.9$, $P(F) = 0.8$. Show that $P(E \cap F) = P(EF) \geq 0.7$.

By Inclusion-Exclusion,

$$P(E \cup F) = P(E) + P(F) - P(EF) = 0.9 + 0.8 - P(EF)$$
$$P(E \cup F) = 1.7 - P(EF)$$

Since $P(E \cup F) \leq 1$, it must be that $P(EF) \geq 0.7$.

More generally, we have Bonferroni's Inequality:

$$P(EF) \geq P(E) + P(F) - 1 \tag{1}$$

Proof:
$P(E \cup F) = P(E) + P(F) - P(EF)$

Since $P(E \cup F) \leq 1$, we have that

$$P(E) + P(F) - P(EF) \leq 1$$
$$\Rightarrow P(EF) \geq P(E) + P(F) - 1$$

Problem 7 from Ch. 2:
There are 15 members of a soccer team: each is either Blue Collar (B) or white collar (W); and either Republican (R), Democrat (D), or Independent (I).

(a) How many outcomes are in the sample space?

To calculate this value, we may consider each person as an independent entity. We know that every individual must fall into one of the six categories: {BR, BD, BI, WR, WD, WI}. Since there are 6 choices for each person and 15 total independent individuals, there is a total of 6^{15} choices.

(b) How many outcomes are in the event that at least one of the team members is a blue collar worker?

Complement: None of the team members are blue collar workers.
We know blue collar worker compose 3 elements of the set: {BR, BD, BI}. Therefore, the answer is

$$6^{15} - \textit{number of outcomes when none are blue collar} = 6^{15} - 3^{15}.$$

(c) How many outcomes are in the event that none of the team members consider themselves independent?

Looking back at part (a), if we restrict the set of possibile identities of the members to {BR, BD, WR, WD}, each individual only has 4 possible choices. Since there are still 15 independent individuals, there are a total of 4^{15} possible choices.

2 Problems from Textbook

Chapter 2 Theoretical Exercises:

12. Show that the probability that exactly one of the events E or F occurs equals $P(E) + P(F) - 2P(EF)$.
proof:
The event that only E occurs is: EF^C. The event that only F occurs is: E^CF.
Therefore, the event that only one of E, F occurs is: $EF^C \cup E^CF$.
The probability that only one occurs is: $P(EF^C \cup E^CF)$.
Now I want to prove that EF^C and E^CF are mutually exclusive:
$\forall x \in EF^C = E \cap F^C$, $x \in E \Rightarrow x \notin E^C \Rightarrow x \notin E^CF$.

Similarly, $\forall y \in E^C F$, $y \notin EF^C$.
Therefore, $EF^C, E^C F$ are mutually exclusive. By Axiom 3:
$P(EF^C \cup E^C F) = P(EF^C) + P(E^C F)$
In problem 13, I have proved: $P(EF^C) = P(E) - P(EF)$
Similarly, we can also prove: $P(E^C F) = P(F) - P(EF)$
Therefore, $P(EF^C \cup E^C F) = P(E) - P(EF) + P(F) - P(EF) = P(E) + P(F) - 2P(EF)$.

13. Prove that $P(EF^C) = P(E) - P(EF)$
proof: $E = EF \cup EF^C \Rightarrow P(E) = P(EF \cup EF^C)$
Since EF and EF^C are mutually exclusive, by axiom 3 of probability, we have:
$P(EF \cup EF^C) = P(EF) + P(EF^C)$
Therefore, $P(E) = P(EF) + P(EF^C)$.
$\Leftrightarrow P(EF^C) = P(E) - P(EF)$.

14. Prove the general Inclusion-Exclusion Law by mathematical induction.
Claim: $P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) - \sum_{1 \le i_1 \le i_2 \le n} P(E_{i_1} E_{i_2}) + ... + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le n} P(E_{i_1} E_{i_2} ... E_{i_r}) + ... + (-1)^{n+1} P(E_1 E_2 E_3 ... E_n)$ for all integer $n \ge 2$
proof:
1. $n = 2$: By Prop 3, we have $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 E_2)$.
2. Suppose for $n = k, k \ge 2$, we have:
$P(\bigcup_{i=1}^k E_i) = \sum_{i=1}^k P(E_i) - \sum_{1 \le i_1 \le i_2 \le k} P(E_{i_1} E_{i_2}) + ... + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k} P(E_{i_1} E_{i_2} ... E_{i_r}) + ... + (-1)^{k+1} P(E_1 E_2 E_3 ... E_k)$
3. For $n = k + 1$:
$P(\bigcup_{i=1}^{k+1} E_i) = P(\bigcup_{i=1}^k E_i \cup E_{k+1}) = P(\bigcup_{i=1}^k E_i) + P(E_{k+1}) - P(\bigcup_{i=1}^k E_i \cap E_{k+1})$ by Prop 3.
$P(\bigcup_{i=1}^k E_i \cap E_{k+1}) = P((E_1 \cup E_2 \cup ... E_{k-1} \cup E_k) \cap E_{k+1})$
By distribution law over $\cap$ and $\cup$:
$P(\bigcup_{i=1}^k E_i \cap E_{k+1}) = P(E_1 E_{k+1} \cup E_2 E_{k+1} \cup ... \cup E_k E_{k+1}) = P(\bigcup_{i=1}^k E_i E_{k+1})$
Using the assumption in 2.
$= \sum_{i=1}^k P(E_i E_{k+1}) - \sum_{1 \le i_1 \le i_2 \le k} P(E_{i_1} E_{i_2} E_{k+1}) + (-1)^r) \sum_{1 \le i_1 \le i_2 \le ... \le i_{r-1} \le k} P(E_{i_1} E_{i_2} ... E_{i_{r-1}} E_{k+1}) + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k} P(E_{i_1} E_{i_2} ... E_{i_r} E_{k+1}) + ... + (-1)^k \sum_{1 \le i_1 \le i_2 \le ... \le i_{k-1} \le k} P(E_{i_1} E_{i_2} ... E_{i_{k-1}} E_{k+1}) + (-1)^{k+1} P(E_1 E_2 E_3 ... E_k E_{k+1})$ †
Therefore, since $P(\bigcup_{i=1}^{k+1} E_i) = P(\bigcup_{i=1}^k E_i \cup E_{k+1}) = P(\bigcup_{i=1}^k E_i) + P(E_{k+1}) - P(\bigcup_{i=1}^k E_i \cap E_{k+1})$,
$P(\bigcup_{i=1}^{k+1} E_i) = P(\bigcup_{i=1}^k E_i \cup E_{k+1}) = P(\bigcup_{i=1}^k E_i) + P(E_{k+1}) - \dagger$
Expanding $P(\bigcup_{i=1}^k E_i)$:
$= \{\sum_{i=1}^k P(E_i) - \sum_{1 \le i_1 \le i_2 \le k} P(E_{i_1} E_{i_2}) + ... + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k} P(E_{i_1} E_{i_2} ... E_{i_r}) + ... + (-1)^{k+1} P(E_1 E_2 E_3 ... E_k)\} + P(E_{k+1}) - \dagger$
$= \sum_{i=1}^{k+1} P(E_i) + \{-\sum_{1 \le i_1 \le i_2 \le k} P(E_{i_1} E_{i_2}) + ... + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k} P(E_{i_1} E_{i_2} ... E_{i_r}) + ... + (-1)^{k+1} P(E_1 E_2 E_3 ... E_k)\} - \dagger$
Pair the first term with first term, second term with second term,.., $r - 1^{th}$ term with $r - 1^{th}$ term in † and in what is in {}, we get:
$P(\bigcup_{i=1}^{k+1} E_i) = \sum_{i=1}^{k+1} -(\sum_{1 \le i_1 \le i_2 \le k} P(E_{i_1} E_{i_2}) + \sum_{i=1}^k P(E_i E_{k+1}) + (\sum_{1 \le i_1 \le i_1 \le i_3 \le k} P(E_{i_1} E_{i_2} E_{i_3}) + \sum_{1 \le i_1 \le i_2 \le k} P(E_{i_1} E_{i_2} E_{k+1}))$
$+ ... + \{(-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k} P(E_{i_1} E_{i_2} ... E_{i_r}) + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_{r-1} \le k} P(E_{i_1} E_{i_2} ... E_{r-1} E_{k+1})\}$
$+ ... + \{(-1)^{k+1} P(E_1 E_2 ... E_k) + (-1)^{k+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_{k-1} \le k} P(E_{i_1} E_{i_2} ... E_{i_{k-1}} E_{k+1})\} + (-1)^{k+2} P(E_1 E_2 E_3 ... E_k E_{k+1})$
Note that the general term:

$$(-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k} P(E_{i_1} E_{i_2} ... E_{i_r}) + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_{r-1} \le k} P(E_{i_1} E_{i_2} ... E_{r-1} E_{k+1})$$

is equal to:

$$(-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k+1} P(E_{i_1} E_{i_2} E_{i_3} ... Ei_r)$$

Therefore, $P(\bigcup_{i=1}^{k+1} E_i) = \sum_{i=1}^{k+1} P(E_i) - \sum_{1 \le i_1 \le i_2 \le k+1} P(E_{i_1} E_{i_2}) + ... + (-1)^{r+1} \sum_{1 \le i_1 \le i_2 \le ... \le i_r \le k+1} P(E_{i_1} E_{i_2} ... E_{i_r}) + ... + (-1)^{k+2} P(E_1 E_2 E_3 ... E_k E_{k+1})$
The claim is proved.