IES 302: Engineering Statistics **2011/2**

HW Solution 1 — Due: February 1

Lecturer: Prapun Suksompong, Ph.D.

Instructions

(a) ONE part of a question will be graded (5 pt). Of course, you do not know which part will be selected; so you should work on all of them.

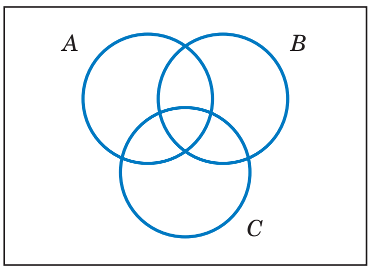
(b) It is important that you try to solve all problems. (5 pt)

(c) Late submission will be heavily penalized.

(d) Write down all the steps that you have done to obtain your answers. You may not get full credit even when your answer is correct without showing how you get your answer.

Problem 1. (**Set Theory**)

(a) Three events are shown on the Venn diagram in the following figure:



Reproduce the figure and shade the region that corresponds to each of the following events.

(i) A^c

(ii) $A \cap B$

(iii) $(A \cap B) \cup C$

(iv) $(B \cup C)^c$

(v) $(A \cap B)^c \cup C$

[Montgomery and Runger, 2010, Q2-19]

(b) Let $\Omega = \{0,1,2,3,4,5,6,7\}$, and put $A = \{1,2,3,4\}$, $B = \{3,4,5,6\}$, and $C = \{5,6\}$. Find $A \cup B$, $A \cap B$, $A \cap C$, A^c, and $B \setminus A$.

For this problem, only answers are needed; you don't have to describe your solution.

Solution:

(a) See Figure 1.1

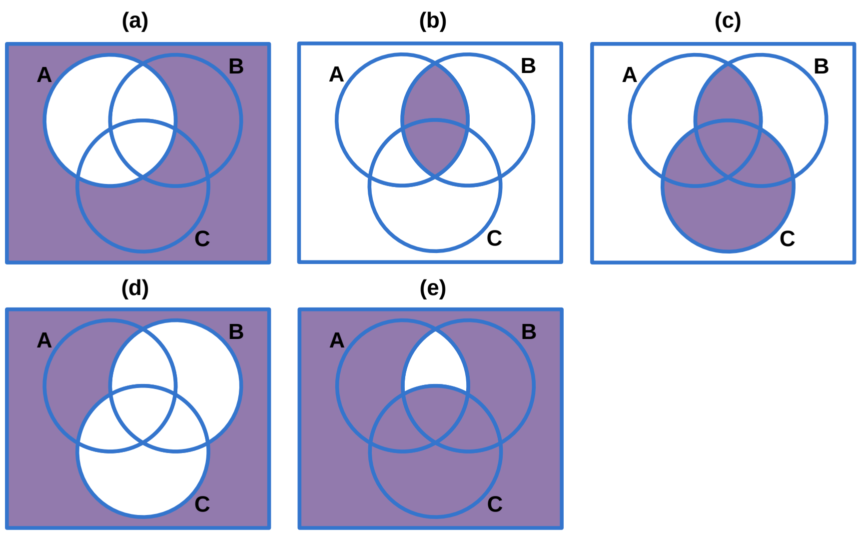


Figure 1.1: Venn diagrams for events in Problem 1

(b) $A \cup B = \{1,2,3,4,5,6\}$, $A \cap B = \{3,4\}$, $A \cap C = \emptyset$, $B \setminus A = \{5,6\} = C$.

Problem 2. (**Classical Probability**) There are three buttons which are painted red on one side and white on the other. If we tosses the buttons into the air, calculate the probability that all three come up the same color.

Remarks: A *wrong* way of thinking about this problem is to say that there are four ways they can fall. All red showing, all white showing, two reds and a white or two whites and a red. Hence, it seems that out of four possibilities, there are two favorable cases and hence the probability is 1/2.

Solution: There are 8 possible outcomes. (The same number of outcomes as tossing three coins.) Among these, only two outcomes will have all three buttons come up the same color. So, the probability is $2/8 = \boxed{1/4}$.

Problem 3. (**Classical Probability**) A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases.

(a) How many different designs are possible? [Montgomery and Runger, 2010, Q2-51]

(b) A specific design is randomly generated by the Web server when you visit the site. If you visit the site five times, what is the probability that you will not see the same design? [Montgomery and Runger, 2010, Q2-71]

Solution:

(a) By the multiplication rule, total number of possible designs

$$= 4 \times 3 \times 5 \times 3 \times 5 = \boxed{900}.$$

(b) From part (a), total number of possible designs is 900. The sample space is now the set of all possible designs that may be seen on five visits. It contains $(900)^5$ outcomes. (This is ordered sampling with replacement.)

The number of outcomes in which all five visits are different can be obtained by realizing that this is ordered sampling without replacement and hence there are $(900)_5$ outcomes. (Alternatively, On the first visit any one of 900 designs may be seen. On the second visit there are 899 remaining designs. On the third visit there are 898 remaining designs. On the fourth and fifth visits there are 897 and 896 remaining designs, respectively. From the multiplication rule, the number of outcomes where all designs are different is $900 \times 899 \times 898 \times 897 \times 896$.)

Therefore, the probability that a design is not seen again is

$$\frac{(900)_5}{900^5} \approx \boxed{0.9889.}$$

Problem 4. (**Combinatorics**) Consider the design of a communication system in the United States.

(a) How many three-digit phone prefixes that are used to represent a particular geographic area (such as an area code) can be created from the digits 0 through 9?

(b) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

(c) As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit?

[Montgomery and Runger, 2010, Q2-45]

Solution:

(a) From the multiplication rule (or by realizing that this is ordered sampling with replacement), $10^3 = \boxed{1,000}$ prefixes are possible

(b) This is ordered sampling without replacement. Therefore $(10)_3 = 10 \times 9 \times 8 = \boxed{720}$ prefixes are possible

(c) From the multiplication rule, $8 \times 2 \times 10 = \boxed{160}$ prefixes are possible.

Problem 5. (**Classical Probability**) A bin of 50 parts contains five that are defective. A sample of two parts is selected at random, without replacement. Determine the probability that both parts in the sample are defective. [Montgomery and Runger, 2010, Q2-49]

Solution: The number of ways to select two parts from 50 is $\binom{50}{2}$ and the number of ways to select two defective parts from the 5 defective ones is $\binom{5}{2}$ Therefore the probability is

$$\frac{\binom{5}{2}}{\binom{50}{2}} = \frac{2}{245} = \boxed{0.0082}.$$

Problem 6. (**Classical Probability**) We all know that the chance of a head (H) or tail (T) coming down after a fair coin is tossed are fifty-fifty. If a fair coin is tossed ten times, then intuition says that five heads are likely to turn up.

Calculate the probability of getting exactly five heads (and hence exactly five tails).

Solution: There are 2^{10} possible outcomes for ten coin tosses. (For each toss, there is two possibilities, H or T). Only $\binom{10}{5}$ among these outcomes have exactly heads and five tails. (Choose 5 positions from 10 position for H. Then, the rest of the positions are automatically T.) The probability of have exactly 5 H and 5 T is

$$\boxed{\frac{\binom{10}{5}}{2^{10}} \approx 0.246.}$$

Note that five heads and five tails will turn up more frequently than any other single combination (one head, nine tails for example) but the sum of all the other possibilities is much greater than the single 5 H, 5 T combination.

Problem 7. (**Classical Probability**) Shuffle a deck of cards and cut it into three piles. What is the probability that (at least) a court card will turn up on top of one of the piles.

Hint: There are 12 court cards (four jacks, four queens and four kings) in the deck.

Solution: In [Lovell, 2006, p. 17–19], this problem is named "Three Lucky Piles". When somebody cuts three piles, they are, in effect, randomly picking three cards from the deck. There are $52 \times 51 \times 50$ possible outcomes. The number of outcomes that do not contain any court card is $40 \times 39 \times 38$. So, the probability of having at least one court card is

$$\boxed{\frac{52 \times 51 \times 50 - 40 \times 39 \times 38}{52 \times 51 \times 50} \approx 0.553.}$$

Problem 8. ***Binomial theorem***: For any positive integer n, we know that

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}. \tag{1.1}$$

(a) What is the coefficient of $x^{12}y^{13}$ in the expansion of $(x+y)^{25}$?

(b) What is the coefficient of $x^{12}y^{13}$ in the expansion of $(2x-3y)^{25}$?

(c) Use the binomial theorem (1.1) to evaluate $\sum_{k=0}^{n} (-1)^k \binom{n}{k}$.

Solution:

(a) $\binom{25}{12} = \boxed{5,200,300}$.

(b) $\binom{25}{12} 2^{12}(-3)^{13} = -\frac{25!}{12!13!} 2^{12} 3^{13} = \boxed{-33959763545702400}$.

(c) From (1.1), set $x = -1$ and $y = 1$, then we have $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = (-1+1)^n = \boxed{0}$.

Problem 9. Each of the possible five outcomes of a random experiment is equally likely. The sample space is $\{a, b, c, d, e\}$. Let A denote the event $\{a, b\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) $P(A)$

(b) $P(B)$

(c) $P(A^c)$

(d) $P(A \cup B)$

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

[Montgomery and Runger, 2010, Q2-54]

Solution: Because the outcomes are equally likely, we can simply use classical probability.

(a) $P(A) = \frac{|A|}{|\Omega|} = \boxed{\frac{2}{5}}$

(b) $P(B) = \frac{|B|}{|\Omega|} = \boxed{\frac{3}{5}}$

(c) $P(A^c) = \frac{|A^c|}{|\Omega|} = \frac{5-2}{5} = \boxed{\frac{3}{5}}$

(d) $P(A \cup B) = \frac{|\{a,b,c,d,e\}|}{|\Omega|} = \frac{5}{5} = \boxed{1}$

(e) $P(A \cap B) = \frac{|\emptyset|}{|\Omega|} = \boxed{0}$

Problem 10. If A, B, and C are disjoint events with $P(A) = 0.2$, $P(B) = 0.3$ and $P(C) = 0.4$, determine the following probabilities:

(a) $P(A \cup B \cup C)$

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

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

(d) $P((A \cup B) \cap C)$

(e) $P(A^c \cap B^c \cap C^c)$

[Montgomery and Runger, 2010, Q2-75]

Solution:

(a) Because A, B, and C are disjoint, $P(A \cup B \cup C) = P(A) + P(B) + P(C) = 0.3 + 0.2 + 0.4 = \boxed{0.9.}$

(b) Because A, B, and C are disjoint, $A \cap B \cap C = \emptyset$ and hence $P(A \cap B \cap C) = P(\emptyset) = \boxed{0}$.

(c) Because A and B are disjoint, $A \cap B = \emptyset$ and hence $P(A \cap B) = P(\emptyset) = \boxed{0}$.

(d) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. By the disjointness among A, B, and C, we have $(A \cap C) \cup (B \cap C) = \emptyset \cup \emptyset = \emptyset$. Therefore, $P((A \cup B) \cap C) = P(\emptyset) = \boxed{0}$.

(e) From $A^c \cap B^c \cap C^c = (A \cup B \cup C)^c$, we have $P(A^c \cap B^c \cap C^c) = 1 - P(A \cup B \cup C) = 1 - 0.9 = \boxed{0.1.}$

IES 302: Engineering Statistics **2011/2**

HW Solution 2 — Due: February 8

Lecturer: Prapun Suksompong, Ph.D.

Instructions

(a) ONE part of a question will be graded (5 pt). Of course, you do not know which part will be selected; so you should work on all of them.

(b) It is important that you try to solve all problems. (5 pt)

(c) Late submission will be heavily penalized.

(d) Write down all the steps that you have done to obtain your answers. You may not get full credit even when your answer is correct without showing how you get your answer.

Problem 1. The sample space of a random experiment is $\{a, b, c, d, e\}$ with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event $\{a, b, c\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) $P(A)$

(b) $P(B)$

(c) $P(A^c)$

(d) $P(A \cup B)$

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

[Montgomery and Runger, 2010, Q2-55]

Solution:

(a) Recall that the probability of a finite or countable event equals the sum of the probabilities of the outcomes in the event. Therefore,

$$\begin{aligned} P(A) &= P(\{a, b, c\}) = P(\{a\}) + P(\{b\}) + P(\{c\}) \\ &= 0.1 + 0.1 + 0.2 = \boxed{0.4.} \end{aligned}$$

(b) Again, the probability of a finite or countable event equals the sum of the probabilities of the outcomes in the event. Thus,

$$P(B) = P(\{c, d, e\}) = P(\{c\}) + P(\{d\}) + P(\{e\}) \\ = 0.2 + 0.4 + 0.2 = \boxed{0.8.}$$

(c) $P(A^c) = 1 - P(A) = 1 - 0.4 = \boxed{0.6.}$

(d) Note that $A \cup B = \Omega$. Hence, $P(A \cup B) = P(\Omega) = \boxed{1.}$

(e) $P(A \cap B) = P(\{c\}) = \boxed{0.2.}$

Problem 2.

(a) Suppose that $P(A) = \frac{1}{2}$ and $P(B) = \frac{2}{3}$. Find the range of the possible value for $P(A \cap B)$. Hint: Smaller than the interval $[0, 1]$. [Capinski and Zastawniak, 2003, Q4.21]

(b) Suppose that $P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{3}$. Find the range of the possible value for $P(A \cup B)$. Hint: Smaller than the interval $[0, 1]$. [Capinski and Zastawniak, 2003, Q4.22]

Solution:

(a) We will first try to bound $P(A \cap B)$. Note that $A \cap B \subset A$ and $A \cap B \subset B$. Hence, we know that $P(A \cap B) \leq P(A)$ and $P(A \cap B) \leq P(B)$. To summarize, we now know that

$$P(A \cap B) \leq \min\{P(A), P(B)\}.$$

On the other hand, we know that

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

Applying the fact that $P(A \cup B) \leq 1$, we then have

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

If the number of the RHS is > 0, then it is a new information. However, if the number on the RHS is negative, it is useless and we will use the fact that $P(A \cap B) \geq 0$. To summarize, we now know that

$$\max\{P(A) + P(B) - 1, 0\} \leq P(A \cap B).$$

In conclusion,

$$\max\{(P(A)+P(B)-1),0\} \leq P(A\cap B) \leq \min\{P(A),P(B)\}.$$

Plugging in the value $P(A)=\frac{1}{2}$ and $P(B)=\frac{2}{3}$ gives the range $\boxed{\left[\frac{1}{6},\frac{1}{2}\right]}$. The upper-bound can be obtained by constructing an example which has $A\subset B$. The lower-bound can be obtained by considering an example where $A\cup B=\Omega$.

(b) By monotonicity we must have

$$P(A\cup B) \geq \max\{P(A),P(B)\}.$$

On the other hand, we know that

$$P(A\cup B) \leq P(A)+P(B).$$

If the RHS is >1, then the inequality is useless and we simply use the fact that it must be ≤ 1. To summarize, we have

$$P(A\cup B) \leq \min\{(P(A)+P(B)),1\}.$$

In conclusion,

$$\max\{P(A),P(B)\} \leq P(A\cup B) \leq \min\{(P(A)+P(B)),1\}.$$

Plugging in the value $P(A)=\frac{1}{2}$ and $P(B)=\frac{1}{3}$, we have

$$P(A\cup B) \in \boxed{\left[\frac{1}{2},\frac{5}{6}\right]}.$$

The upper-bound can be obtained by making $A\perp B$. The lower-bound is achieved when $B\subset A$.

Problem 3. Let A and B be events for which $P(A)$, $P(B)$, and $P(A\cup B)$ are known. Express the following probabilities in terms of the three known probabilities above.

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

(b) $P(A\cap B^c)$

(c) $P\left(B\cup(A\cap B^c)\right)$

(d) $P(A^c \cap B^c)$

Solution:

(a) $P(A \cap B) = \boxed{P(A) + P(B) - P(A \cup B)}$. This property is shown in class.

(b) We have seen in class that $P(A \cap B^c) = P(A) - P(A \cap B)$. Plugging in the expression for $P(A \cap B)$ from the previous part, we have

$$P(A \cap B^c) = P(A) - (P(A) + P(B) - P(A \cup B)) = \boxed{P(A \cup B) - P(B)}.$$

Alternatively, we can start from scratch with the set identity $A \cup B = B \cup (A \cap B^c)$ whose union is a disjoint union. Hence,

$$P(A \cup B) = P(B) + P(A \cap B^c).$$

Moving $P(B)$ to the LHS finishes the proof.

(c) $P(B \cup (A \cap B^c)) = \boxed{P(A \cup B)}$ because $A \cup B = B \cup (A \cap B^c)$.

(d) $P(A^c \cap B^c) = \boxed{1 - P(A \cup B)}$ because $A^c \cap B^c = (A \cup B)^c$.

Problem 4.

(a) Suppose that $P(A|B) = 0.4$ and $P(B) = 0.5$ Determine the following:

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

(ii) $P(A^c \cap B)$

[Montgomery and Runger, 2010, Q2-105]

(b) Suppose that $P(A|B) = 0.2$, $P(A|B^c) = 0.3$ and $P(B) = 0.8$ What is $P(A)$? [Montgomery and Runger, 2010, Q2-106]

Solution:

(a) Recall that $P(A \cap B) = P(A|B)P(B)$. Therefore,

(i) $P(A \cap B) = 0.4 \times 0.5 = \boxed{0.2.}$

(ii) $P(A^c \cap B) = P(B \setminus A) = P(B) - P(A \cap B) = 0.5 - 0.2 = \boxed{0.3.}$
Alternatively, $P(A^c \cap B) = P(A^c|B)P(B) = (1 - P(A|B))P(B) = (1 - 0.4) \times 0.5 = 0.3$.

(b) By the total probability formula, $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = 0.2 \times 0.8 + 0.3 \times (1 - 0.8) = \boxed{0.22.}$

Problem 5. [Gubner, 2006, Q2.60] You have five computer chips, two of which are known to be defective.

(a) You test one of the chips; what is the probability that it is defective?

(b) Your friend tests two chips at random and reports that one is defective and one is not. Given this information, you test one of the three remaining chips at random; what is the conditional probability that the chip you test is defective?

Solution:

(a) $\boxed{\frac{2}{5}}$ (two of five chips are defective.)

(b) Among the three remaining chips, only one is defective. So, the conditional probability that the chosen chip is defective is $\boxed{\frac{1}{3}.}$

Problem 6. Due to an Internet configuration error, packets sent from New York to Los Angeles are routed through El Paso, Texas with probability 3/4. Given that a packet is routed through El Paso, suppose it has conditional probability 1/3 of being dropped. Given that a packet is not routed through El Paso, suppose it has conditional probability 1/4 of being dropped.

(a) Find the probability that a packet is dropped.

(b) Find the conditional probability that a packet is routed through El Paso given that it is not dropped.

[Gubner, 2006, Ex.1.20]

Solution: To solve this problem, we use the notation $E = \{\text{routed through El Paso}\}$ and $D = \{\text{packet is dropped}\}$. With this notation, it is easy to interpret the problem as telling us that

$$P(D|E) = 1/3, \quad P(D|E^c) = 1/4, \quad \text{and } P(E) = 3/4.$$

(a) By the law of total probability,

$$\begin{aligned} P(D) &= P(D|E)P(E) + P(D|E^c)P(E^c) = (1/3)(3/4) + (1/4)(1 - 3/4) \\ &= 1/4 + 1/16 = \boxed{5/16}. \end{aligned}$$

(b) $P(E|D^c) = \frac{P(E\cap D^c)}{P(D^c)} = \frac{P(D^c|E)P(E)}{P(D^c)} = \frac{(1-1/3)(3/4)}{1-5/16} = \boxed{\frac{8}{11}}$.

Problem 7. You have two coins, a fair one with probability of heads $\frac{1}{2}$ and an unfair one with probability of heads $\frac{1}{3}$, but otherwise identical. A coin is selected at random and tossed, falling heads up. How likely is it that it is the fair one? [Capinski and Zastawniak, 2003, Q7.28]

Solution: Let F, U, and H be the events that "the selected coin is fair", "the selected coin is unfair", and "the coin lands heads up", respectively.

Because the coin is selected at random, the probability $P(F)$ of selecting the fair coin is $P(F) = \frac{1}{2}$. For fair coin, the conditional probability $P(H|F)$ of heads is $\frac{1}{2}$ For the unfair coin, $P(U) = 1 - P(F) = \frac{1}{2}$ and $P(H|U) = \frac{1}{3}$.

By the Bayes' formula, the probability that the fair coin has been selected given that it lands heads up is

$$P(F|H) = \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|U)P(U)} = \frac{\frac{1}{2}\times\frac{1}{2}}{\frac{1}{2}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{2}} = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}} = \frac{1}{1+\frac{2}{3}} = \boxed{\frac{3}{5}}.$$

Problem 8. You have three coins in your pocket, two fair ones but the third biased with probability of heads p and tails $1-p$. One coin selected at random drops to the floor, landing heads up. How likely is it that it is one of the fair coins? [Capinski and Zastawniak, 2003, Q7.29]

Solution: Let F, U, and H be the events that "the selected coin is fair", "the selected coin is unfair", and "the coin lands heads up", respectively. We are given that

$$P(F) = \frac{2}{3}, \quad P(U) = \frac{1}{3}, \quad P(H|F) = \frac{1}{2}, P(H|U) = p.$$

By the Bayes' formula, the probability that the fair coin has been selected given that it lands heads up is

$$P(F|H) = \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|U)P(U)} = \frac{\frac{1}{2}\times\frac{2}{3}}{\frac{1}{2}\times\frac{2}{3}+p\times\frac{1}{3}} = \boxed{\frac{1}{1+p}}.$$

Problem 9. Someone has rolled a fair die twice. You know that one of the rolls turned up a face value of six. What is the probability that the other roll turned up a six as well?

Hint: Not $\frac{1}{6}$.

Solution: Take as sample space the set $\{(i,j)|i,j = 1,\ldots,6\}$, where i and j denote the outcomes of the first and second rolls. A probability of 1/36 is assigned to each element of the sample space. The event of two sixes is given by $A = \{(6,6)\}$ and the event of at least one six is given by $B = (1,6),\ldots,(5,6),(6,6),(6,5),\ldots,(6,1)$. Applying the definition of conditional probability gives

$$P(A|B) = P(A \cap B)/P(B) = \frac{1/36}{11/36}.$$

Hence the desired probability is $\boxed{1/11}$. [Tijms, 2007, Example 8.1, p. 244]

Problem 10. An article in the British Medical Journal ["Comparison of Treatment of Renal Calculi by Operative Surgery, Percutaneous Nephrolithotomy, and Extracorporeal Shock Wave Lithotripsy" (1986, Vol. 82, pp. 879892)] provided the following discussion of success rates in kidney stone removals. Open surgery (OS) had a success rate of 78% (273/350) while a newer method, percutaneous nephrolithotomy (PN), had a success rate of 83% (289/350). This newer method looked better, but the results changed when stone diameter was considered. For stones with diameters less than two centimeters, 93% (81/87) of cases of open surgery were successful compared with only 87% (234/270) of cases of PN. For stones greater than or equal to two centimeters, the success rates were 73% (192/263) and 69% (55/80) for open surgery and PN, respectively. Open surgery is better for both stone sizes, but less successful in total. In 1951, E. H. Simpson pointed out this apparent contradiction (known as Simpsons Paradox) but the hazard still persists today. Explain how open surgery can be better for both stone sizes but worse in total. [Montgomery and Runger, 2010, Q2-115]

Solution: First, let's recall the total probability theorem:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)\,P(B) + P(A|B^c)\,P(B^c). \end{aligned}$$

We can see that $P(A)$ does not depend only on $P(A \cap B)$ and $P(A|B^c)$. It also depends on $P(B)$ and $P(B^c)$. In the extreme case, we may imagine the case with $P(B) = 1$ in which $P(A) = P(A|B)$. At another extreme, we may imagine the case with $P(B) = 0$ in which $P(A) = P(A|B^c)$. Therefore, depending on the value of $P(B)$, the value of $P(A)$ can be anywhere between $P(A|B)$ and $P(A|B^c)$.

Now, let's consider events A_1, B_1, A_2, and B_2. Let $P(A_1|B_1) = 0.93$ and $P(A_1|B_1^c) = 0.73$. Therefore, $P(A_1) \in [0.73, 0.93]$. On the other hand, let $P(A_2|B_2) = 0.87$ and $P(A_2|B_2^c) = 0.69$. Therefore, $P(A_2) \in [0.69, 0.87]$. With small value of $P(B_1)$, the value of $P(A_1)$ can be 0.78 which is closer to its lower end of the bound. With large value of $P(B_2)$, the value of $P(A_2)$ can be 0.83 which is closer to its upper end of the bound. Therefore, even though $P(A_1|B_1) > P(A_2|B_2) = 0.87$ and $P(A_1|B_1^c) > P(A_2|B_2^c)$, it is possible that $P(A_1) < P(A_2)$.

In the context of the paradox under consideration, note that the success rate of PN with small stones (87%) is higher than the success rate of OS with large stones (73%). Therefore, by having a lot of large stone cases to be tested under OS and also have a lot of small stone cases to be tested under PN, we can create a situation where the overall success rate of PN comes out to be better then the success rate of OS. This is exactly what happened in the study as shown in Table 2.1.

Open surgery	success	failure	sample size	sample percentage	conditional success rate
large stone	192	71	263	75%	73%
small stone	81	6	87	25%	93%
overall summary	273	77	350	100%	78%

PN	success	failure	sample size	sample percentage	conditional success rate
large stone	55	25	80	23%	69%
small stone	234	36	270	77%	87%
overall summary	289	61	350	100%	83%

Table 2.1: Success rates in kidney stone removals.

IES 302: Engineering Statistics **2011/2**

HW Solution 3 — Due: February 15

Lecturer: Prapun Suksompong, Ph.D.

Instructions

(a) ONE part of a question will be graded (5 pt). Of course, you do not know which part will be selected; so you should work on all of them.

(b) It is important that you try to solve all problems. (5 pt)

(c) Late submission will be heavily penalized.

(d) Write down all the steps that you have done to obtain your answers. You may not get full credit even when your answer is correct without showing how you get your answer.

Problem 1. Suppose that for the general population, 1 in 5000 people carries the human immunodeficiency virus (HIV). A test for the presence of HIV yields either a positive (+) or negative (-) response. Suppose the test gives the correct answer 99% of the time.

(a) What is $P(-|H)$, the conditional probability that a person tests negative given that the person does have the HIV virus?

(b) What is $P(H|+)$, the conditional probability that a randomly chosen person has the HIV virus given that the person tests positive?

Solution:

(a) Because the test is correct 99% of the time,

$$P(-|H) = P(+|H^c) = \boxed{0.01}.$$

(b) Using Bayes' formula, $P(H|+) = \frac{P(+|H)P(H)}{P(+)}$, where $P(+)$ can be evaluated by the total probability formula:

$$P(+) = P(+|H)P(H) + P(+|H^c)P(H^c) = 0.99 \times 0.0002 + 0.01 \times 0.9998.$$

Plugging this back into the Bayes' formula gives

$$P(H|+) = \frac{0.99 \times 0.0002}{0.99 \times 0.0002 + 0.01 \times 0.9998} \approx \boxed{0.0194}.$$

Thus, even though the test is correct 99% of the time, the probability that a random person who tests positive actually has HIV is less than 2%. The reason this probability is so low is that the a priori probability that a person has HIV is very small.

Problem 2.

(a) Suppose that $P(A|B) = 1/3$ and $P(A|B^c) = 1/4$. Find the range of the possible values for $P(A)$.

(b) Suppose that C_1, C_2, and C_3 partition Ω. Furthermore, suppose we know that $P(A|C_1) = 1/3$, $P(A|C_2) = 1/4$ and $P(A|C_3) = 1/5$. Find the range of the possible values for $P(A)$.

Solution: First recall the total probability theorem: Suppose we have a collection of events $B_1, B_2, \ldots, B_n$ which partitions Ω. Then,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots P(A \cap B_n) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots P(A|B_n)P(B_n) \end{aligned}$$

(a) Note that B and B^c partition Ω. So, we can apply the total probability theorem:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = \frac{1}{3}P(B) + \frac{1}{4}(1 - P(B)).$$

You may check that, by varying the value of $P(B)$ from 0 to 1, we can get the value of $P(A)$ to be any number in the range $\left[\frac{1}{4}, \frac{1}{3}\right]$. Technically, we can not use $P(B) = 0$ because that would make $P(A|B)$ not well-defined. Similarly, we can not use $P(B) = 1$ because that would mean $P(B^c) = 0$ and hence make $P(A|B^c)$ not well-defined. Therfore, the range of $P(A)$ is $\boxed{\left(\frac{1}{4}, \frac{1}{3}\right).}$

Note that larger value of $P(A)$ is not possible because

$$P(A) = \frac{1}{3}P(B) + \frac{1}{4}(1 - P(B)) < \frac{1}{3}P(B) + \frac{1}{3}(1 - P(B)) = \frac{1}{3}.$$

Similarly, smaller value of $P(A)$ is not possible because

$$P(A) = \frac{1}{3}P(B) + \frac{1}{4}(1 - P(B)) > \frac{1}{4}P(B) + \frac{1}{3}(1 - P(B)) = \frac{1}{4}.$$

(b) Again, we apply the total probability theorem:

$$\begin{aligned} P(A) &= P(A|C_1)P(C_1) + P(A|C_2)P(C_2) + P(A|C_3)P(C_3) \\ &= \frac{1}{3}P(C_1) + \frac{1}{4}P(C_2) + \frac{1}{5}P(C_3). \end{aligned}$$

Because C_1, C_2, and C_3 partition Ω, we know that $P(C_1) + P(C_2) + P(C_3) = 1$. Now,

$$P(A) = \frac{1}{3}P(C_1) + \frac{1}{4}P(C_2) + \frac{1}{5}P(C_3) < \frac{1}{3}P(C_1) + \frac{1}{3}P(C_2) + \frac{1}{3}P(C_3) = \frac{1}{3}.$$

Similarly,

$$P(A) = \frac{1}{3}P(C_1) + \frac{1}{4}P(C_2) + \frac{1}{5}P(C_3) > \frac{1}{5}P(C_1) + \frac{1}{5}P(C_2) + \frac{1}{5}P(C_3) = \frac{1}{5}.$$

Therefore, $P(A)$ must be inside $\left(\frac{1}{5}, \frac{1}{3}\right)$.

You may check that any value of $P(A)$ in the range $\boxed{\left(\frac{1}{5}, \frac{1}{3}\right)}$ can be obtained by first setting the value of $P(C_2)$ to be close to 0 and varying the value of $P(C_1)$ from 0 to 1.

Problem 3. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let A denote the event that the design color is red and let B denote the event that the font size is not the smallest one. Calculate the following probabilities.

(a) $P(A \cup B)$

(b) $P(A \cup B^c)$

(c) $P(A^c \cup B^c)$

[Montgomery and Runger, 2010, Q2-84]

Solution:

(a) First recall that $P(A \cup B) = P(A) + P(B) - P(A \cap C)$. For this problem, $P(A) = 1/4$ (red is one of the four colors) and $P(B) = 4/5$ (four of the five fonts can be used). Because the design is randomly generated, events A and B are independent. Hence, $P(A \cap B) = \frac{1}{4}\frac{4}{5} = \frac{1}{5} = 0.2$. Therefore, $P(A \cup B) = \frac{1}{4} + \frac{4}{5} - \frac{1}{5} = \boxed{\frac{17}{20} = 0.85.}$

(b) $P(A^c \cup B^c) = 1 - P(A \cap B) = 1 - 0.2 = \boxed{0.8.}$

(c) $P(A \cup B^c) = 1 - P(A^c \cap B)$. Because $A \perp\!\!\!\perp B$, we also have $A^c \perp\!\!\!\perp B$. Hence, $P(A^c \cup B^c) = 1 - P(A^c)P(B) = 1 - \frac{3}{4}\frac{4}{5} = \frac{2}{5} = \boxed{0.4.}$

Problem 4. Anne and Betty go fishing. Find the conditional probability that Anne catches no fish given that at least one of them catches no fish. Assume they catch fish independently and that each has probability $0 < p < 1$ of catching no fish. [Gubner, 2006, Q2.62]

Solution: Let A be the event that Anne catches no fish and B be the event that Betty catches no fish. From the question, we know that A and B are independent. The event "at least one of the two women catches nothing" can be represented by $A \cup B$. So we have

$$P(A|A \cup B) = \frac{P(A \cap (A \cup B))}{P(A \cup B)} = \frac{P(A)}{P(A) + P(B) - P(A)P(B)} = \frac{p}{2p - p^2} = \boxed{\frac{1}{2-p}}.$$

Problem 5. In this question, each experiment has equiprobable outcomes.

(a) Let $\Omega = \{1, 2, 3, 4\}$, $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{2, 3\}$.

- (i) Determine whether $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$.
- (ii) Check whether $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$.
- (iii) Are A_1, A_2, and A_3 independent?

(b) Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A_1 = \{1, 2, 3, 4\}$, $A_2 = A_3 = \{4, 5, 6\}$.

- (i) Check whether $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$.
- (ii) Check whether $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$.
- (iii) Are A_1, A_2, and A_3 independent?

Solution:

(a) We have $P(A_i) = \frac{1}{2}$ and $P(A_i \cap A_j) = \frac{1}{4}$.

- (i) $P(A_i \cap A_j) = P(A_i)P(A_j)$ for any $i \neq j$.
- (ii) $A_1 \cap A_2 \cap A_3 = \emptyset$. Hence, $P(A_1 \cap A_2 \cap A_3) = 0$, which is *not* the same as $P(A_1)P(A_2)P(A_3)$.
- (iii) No.

(b) We have $P(A_1) = \frac{4}{6} = \frac{2}{3}$ and $P(A_2) = P(A_3) = \frac{3}{6} = \frac{1}{2}$.

- (i) $A_1 \cap A_2 \cap A_3 = \{4\}$. Hence, $P(A_1 \cap A_2 \cap A_3) = \frac{1}{6}$.
  $P(A_1)P(A_2)P(A_3) = \frac{2}{3}\frac{1}{2}\frac{1}{2} = \frac{1}{6}$.
  Hence, $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$.

(ii) $P(A_2 \cap A_3) = P(A_2) = \frac{1}{2} \neq P(A_2)P(A_3)$
$P(A_1 \cap A_2) = p(4) = \frac{1}{6} \neq P(A_1)P(A_2)$
$P(A_1 \cap A_3) = p(4) = \frac{1}{6} \neq P(A_1)P(A_3)$
Hence, $P(A_i \cap A_j) \neq P(A_i) P(A_j)$ for all $i \neq j$.

(iii) No.

Problem 6. A certain binary communication system has a bit-error rate of 0.1; i.e., in transmitting a single bit, the probability of receiving the bit in error is 0.1. To transmit messages, a three-bit repetition code is used. In other words, to send the message 1, 111 is transmitted, and to send the message 0, 000 is transmitted. At the receiver, if two or more 1s are received, the decoder decides that message 1 was sent; otherwise, i.e., if two or more zeros are received, it decides that message 0 was sent.

Assuming bit errors occur independently, find the probability that the decoder puts out the wrong message.

[Gubner, 2006, Q2.62]

Solution: Let $p = 0.1$ be the bit error rate. Error event $\mathcal{E}$ occurs if there are at least two bit errors. Therefore

$$P(\mathcal{E}) = \binom{3}{2} p^2(1-p) + \binom{3}{3} p^3 = p^2(3-2p).$$

When $p = 0.1$, $P(\mathcal{E}) \approx \boxed{0.028}$.

Problem 7. In an experiment, A, B, C, and D are events with probabilities $P(A \cup B) = \frac{5}{8}$, $P(A) = \frac{3}{8}$, $P(C \cap D) = \frac{1}{3}$, and $P(C) = \frac{1}{2}$. Furthermore, A and B are disjoint, while C and D are independent.

(a) Find

  (i) $P(A \cap B)$
  (ii) $P(B)$
  (iii) $P(A \cap B^c)$
  (iv) $P(A \cup B^c)$

(b) Are A and B independent?

(c) Find

  (i) $P(D)$
  (ii) $P(C \cap D^c)$

(iii) $P(C^c \cap D^c)$

(iv) $P(C|D)$

(v) $P(C \cup D)$

(vi) $P(C \cup D^c)$

(d) Are C and D^c independent?

Solution:

(a)

(i) Because $A \perp B$, we have $A \cap B = \emptyset$ and hence $P(A \cap B) = \boxed{0}$.

(ii) Recall that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Hence, $P(B) = P(A \cup B) - P(A) + P(A \cap B) = 5/8 - 3/8 + 0 = 2/8 = boxed 1/4$.

(iii) $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) = \boxed{3/8}$.

(iv) Start with $P(A \cup B^c) = 1 - P(A^c \cap B)$. Now, $P(A^c \cap B) = P(B) - P(A \cap B) = P(B) = 1/4$. Hence, $P(A \cup B^c) = 1 - 1/4 = \boxed{3/4}$.

(b) Events A and B are not independent because $P(A \cap B) \neq P(A)P(B)$.

(c)

(i) Because $C \perp\!\!\!\perp D$, we have $P(C \cap D) = P(C)P(D)$. Hence, $P(D) = \frac{P(C \cap D)}{P(C)} = \frac{1/3}{1/2} = \boxed{2/3}$.

(ii) $P(C \cap D^c) = P(C) - P(C \cap D) = 1/2 - 1/3 = \boxed{1/6}$.
Alternatively, because $C \perp\!\!\!\perp D$, we know that $C \perp\!\!\!\perp D^c$. Hence, $P(C \cap D^c) = P(C)P(D^c) = \frac{1}{2}\left(1 - \frac{2}{3}\right) = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$.

(iii) First, we find $P(C \cup D) = P(C) + P(D) - P(C \cap D) = 1/2 + 2/3 - 1/3 = 5/6$. Hence, $P(C^c \cap D^c) = 1 - P(C \cup D) = 1 - 5/6 = \boxed{1/6}$.
Alternatively, because $C \perp\!\!\!\perp D$, we know that $C^c \perp\!\!\!\perp D^c$. Hence, $P(C^c \cap D^c) = P(C^c)P(D^c) = \left(1 - \frac{1}{2}\right)\left(1 - \frac{2}{3}\right) = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$.

(iv) Because $C \perp\!\!\!\perp D$, we have $P(C|D) = P(C) = \boxed{1/2}$.

(v) In part (iii), we already found $P(C \cup D) = P(C) + P(D) - P(C \cap D) = 1/2 + 2/3 - 1/3 = \boxed{5/6}$.

(vi) $P(C \cup D^c) = 1 - P(C^c \cap D) = 1 - P(C^c)P(D) = 1 - \frac{1}{2}\frac{2}{3} = \boxed{2/3}$. Note that we use the fact that $C^c \perp\!\!\!\perp D$ to get the second equality.

Alternatively, $P(C \cup D^c) = P(C) + P(D^c) - P(C \cap D^C)$. From (i), we have $P(D) = 2/3$. Hence, $P(D^c) = 1-2/3 = 1/3$. From (ii), we have $P(C \cap D^C) = 1/6$. Therefore, $P(C \cup D^c) = 1/2 + 1/3 - 1/6 = 2/3$.

(d) Yes. We know that if $C \perp\!\!\!\perp D$, then $C \perp\!\!\!\perp D^c$.

Problem 8. Consider the sample space $\Omega = \{-2, -1, 0, 1, 2, 3, 4\}$. For an event $A \subset \Omega$, suppose that $P(A) = |A|/|\Omega|$. Define the random variable $X(\omega) = \omega^2$. Find the probability mass function of X.

Solution: Because $|\Omega| = 7$, we have $p(\omega) = 1/7$. The random variable maps the outcomes $-2, -1, 0, 1, 2, 3, 4$ to numbers $4, 1, 0, 1, 4, 9, 16$, respectively. Therefore,

$$
\begin{aligned}
p_X(0) &= P(\{0\}) = \frac{1}{7}, \\
p_X(1) &= P(\{-1, 1\}) = \frac{2}{7}, \\
p_X(4) &= P(\{-2, 2\}) = \frac{2}{7}, \\
p_X(9) &= P(\{3\}) = \frac{1}{7}, \text{ and} \\
p_X(16) &= P(\{4\}) = \frac{1}{7}.
\end{aligned}
$$

The pmf can then be expressed as

$$
p_X(x) = \boxed{\begin{cases} \frac{1}{7}, & x = 0, 9, 16 \\ \frac{2}{7}, & x = 1, 4 \\ 0, & \text{otherwise.} \end{cases}}
$$

Problem 9. Suppose X is a random variable whose pmf at $x = 0, 1, 2, 3, 4$ is given by $p_X(x) = \frac{2x+1}{25}$.

Remark: Note that the statement above does not specify the value of the $p_X(x)$ at the value of x that is not 0,1,2,3, or 4.

(a) What is $p_X(5)$?

(b) Determine the following probabilities:

(i) $P[X = 4]$

(ii) $P[X \le 1]$

(iii) $P[2 \leq X < 4]$

(iv) $P[X > -10]$

Solution:

(a) First, we calculate

$$\sum_{x=0}^{4} p_X(x) = \sum_{x=0}^{4} \frac{2x+1}{25} = \frac{25}{25} = 1.$$

Therefore, there can't be any other x with $p_X(x) > 0$. At $x = 5$, we then conclude that $p_X(5) = \boxed{0.}$ The sam reasoning also implies that $p_X(x) = 0$ at any x that is not 0,1,2,3, or 4.

(b) Recall that, for discrete random variable X, the probability

$$P[\text{some condition(s) on } X]$$

can be calculated by adding $p_X(x)$ for all x in the support of X that satisfies the given condition(s).

(i) $P[X = 4] = p_X(4) = \frac{2\times4+1}{25} = \boxed{\frac{9}{25}.}$

(ii) $P[X \leq 1] = p_X(0) + p_X(1) = \frac{2\times0+1}{25} + \frac{2\times1+1}{25} = \frac{1}{25} + \frac{3}{25} = \boxed{\frac{4}{25}.}$

(iii) $P[2 \leq X < 4] = p_X(2) + p_X(3) = \frac{2\times2+1}{25} + \frac{2\times3+1}{25} = \frac{5}{25} + \frac{7}{25} = \boxed{\frac{12}{25}.}$

(iv) $P[X > -10] = \boxed{1}$ because all the x in the support of X satisfies $x > -10$.

IES 302: Engineering Statistics	2011/2

HW Solution 4 — Not Due

Lecturer: Prapun Suksompong, Ph.D.

Problem 1. The random variable V has pmf

$$p_V(v) = \begin{cases} cv^2, & v = 1, 2, 3, 4, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of the constant c.

(b) Find $P[V \in \{u^2 : u = 1, 2, 3, \dots\}]$.

(c) Find the probability that V is an even number.

(d) Find $P[V > 2]$.

(e) Sketch $p_V(v)$.

(f) Sketch $F_V(v)$.

Solution: [Y&G, Q2.2.3]

(a) We choose c so that the pmf sums to one:

$$\sum_v p_V(v) = c(1^2 + 2^2 + 3^2 + 4^2) = 30c = 1.$$

Hence, $c = \boxed{1/30}$.

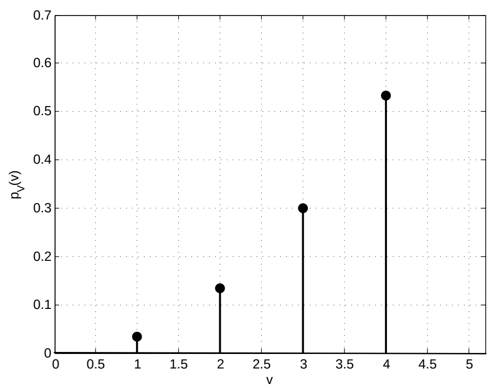
(b) $P[V \in \{u^2 : u = 1, 2, 3, \dots\}] = p_V(1) + p_V(4) = c(1^2 + 4^2) = \boxed{17/30}$.

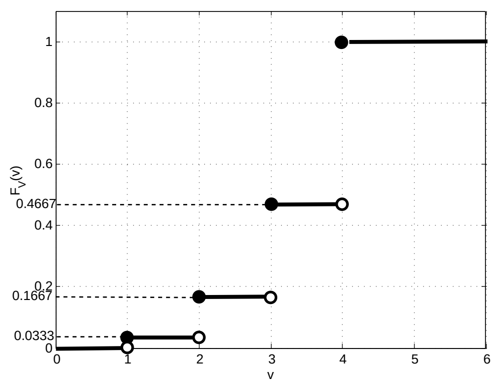
(c) $P[V \text{ even}] = p_V(2) + p_V(4) = c(2^2 + 4^2) = 20/30 = \boxed{2/3}$.

(d) $P[V > 2] = p_V(3) + p_V(4) = c(3^2 + 4^2) = 25/30 = \boxed{5/6}$.

(e) Sketch of $p_V(v)$:

(f) Sketch of $F_V(v)$:





Problem 2. An optical inspection system is to distinguish among different part types. The probability of a correct classification of any part is 0.98. Suppose that three parts are inspected and that the classifications are independent.

(a) Let the random variable X denote the number of parts that are correctly classified. Determine the probability mass function of X. [Montgomery and Runger, 2010, Q3-20]

(b) Let the random variable Y denote the number of parts that are incorrectly classified. Determine the probability mass function of Y.

Solution:

(a) X is a binomial random variable with $n = 3$ and $p = 0.98$. Hence,

$$p_X(x) = \begin{cases} \binom{3}{x} 0.98^x (0.02)^{3-x}, & x \in \{0,1,2,3\}, \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

In particular, $p_X(0) = 8 \times 10^{-6}$, $p_X(1) = 0.001176$, $p_X(2) = 0.057624$, and $p_X(3) = 0.941192$. Note that in `MATLAB`, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.98)`.

(b) Y is a binomial random variable with $n = 3$ and $p = 0.02$. Hence,

$$p_Y(y) = \begin{cases} \binom{3}{y} 0.02^y (0.98)^{3-y}, & y \in \{0,1,2,3\}, \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

In particular, $p_Y(0) = 0.941192$, $p_Y(1) = 0.057624$, $p_Y(2) = 0.001176$, and $p_Y(3) = 8 \times 10^{-6}$. Note that in `MATLAB`, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.02)`.

Alternatively, note that there are three parts. If X of them are classified correctly, then the number of incorrectly classified parts is $n - X$, which is what we defined as Y. Therefore, $Y = 3 - X$. Hence, $p_Y(y) = P[Y = y] = P[3 - X = y] = P[X = 3 - y] = p_X(3 - y)$.

Problem 3. The thickness of the wood paneling (in inches) that a customer orders is a random variable with the following cdf:

$$F_X(x) = \begin{cases} 0, & x < \frac{1}{8} \\ 0.2, & \frac{1}{8} \le x < \frac{1}{4} \\ 0.9, & \frac{1}{4} \le x < \frac{3}{8} \\ 1 & x \ge \frac{3}{8} \end{cases}$$

Determine the following probabilities:

(a) $P[X \le 1/18]$

(b) $P[X \le 1/4]$

(c) $P[X \le 5/16]$

(d) $P[X > 1/4]$

(e) $P[X \le 1/2]$

[Montgomery and Runger, 2010, Q3-42]

Solution:

(a) $P[X \le 1/18] = F_X(1/18) = 0$ because $\frac{1}{18} < \frac{1}{8}$.

(b) $P[X \le 1/4] = F_X(1/4) = 0.9$

(c) $P[X \le 5/16] = F_X(5/16) = 0.9$ because $\frac{1}{4} < \frac{5}{16} < \frac{1}{8}$.

(d) $P[X > 1/4] = 1 - P[X \le 1/4] = 1 - F_X(1/4) = 1 - 0.9 = 0.1$.

(e) $P[X \leq 1/2] = F_X(1/2) = 1$ because $\frac{1}{2} > \frac{3}{8}$.

Alternatively, we can also derive the pmf first and then calculate the probabilities.

Problem 4. Plot the Poisson pmf for $\alpha = 10$, 30, and 50.

Solution: See Figure 4.1.

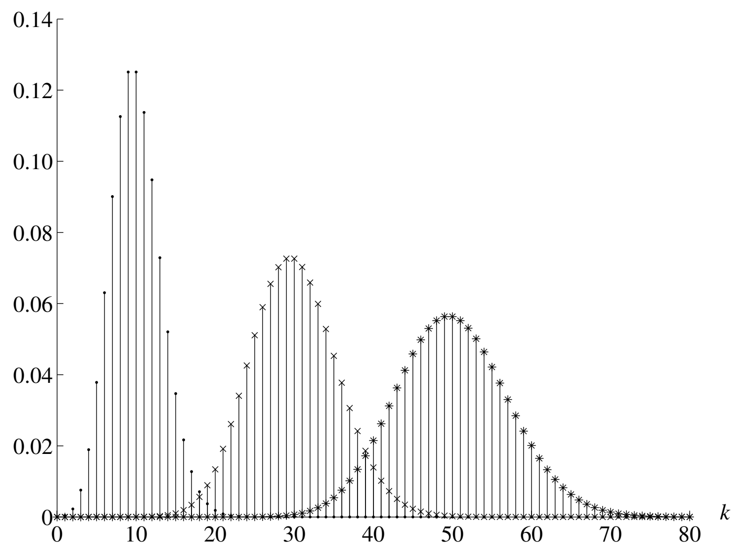


Figure 4.1: The Poisson pmf for $\alpha = 10$, 30, and 50 from left to right, respectively. [Gubner, 2006, Figure 2.5]

Problem 5. Let $X \sim \mathcal{P}(\alpha)$.

(a) Evaluate $P[X > 1]$. Your answer should be in terms of α.

(b) Compute the numerical value of $P[X > 1]$ when $\alpha = 1$.

Solution:

(a) $P[X > 1] = 1 - P[X \leq 1] = 1 - (P[X = 0] + P[X = 1]) = \boxed{1 - e^{-\alpha}(1 + \alpha)}$.

(b) $\boxed{0.264}$.

Problem 6. When n is large, binomial distribution Binomial(n, p) becomes difficult to compute directly because of the need to calculate factorial terms. In this question, we will consider an approximation when p is close to 0. In such case, the binomial can be approximated by the Poisson distribution with parameter $\alpha = np$.

More specifically, suppose X_n has a binomial distribution with parameters n and p_n. If $p_n \to 0$ and $np_n \to \alpha$ as $n \to \infty$, then

$$P[X_n = k] \to e^{-\alpha}\frac{\alpha^k}{k!}.$$

(a) Let $X \sim$ Binomial$(12, 1/36)$. (For example, roll two dice 12 times and let X be the number of times a double 6 appears.) Evaluate $p_X(x)$ for $x = 0, 1, 2$.

(b) Compare your answers in the previous part with the Poisson approximation.

(c) Compare the plot of $p_X(x)$ and $\mathcal{P}(np)$.

Solution:

(a) 0.7132, 0.2445, 0.0384.

(b) 0.7165, 0.2388, 0.0398.

(c) See Figure 4.2.

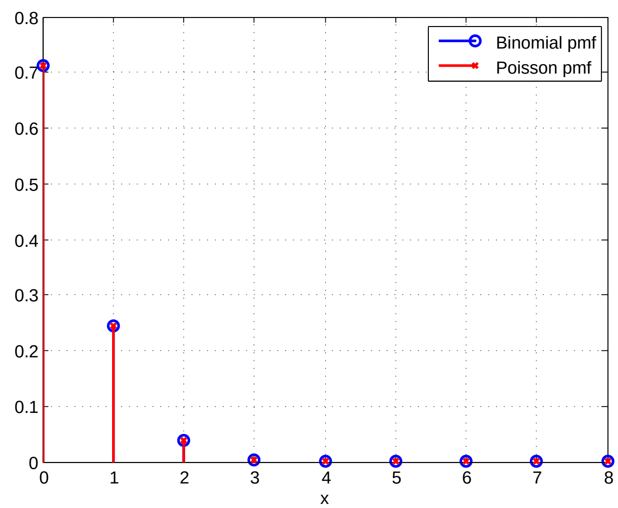


Figure 4.2: Poisson Approximation

Problem 7. In one of the New York state lottery games, a number is chosen at random between 0 and 999. Suppose you play this game 250 times. Use the Poisson approximation to estimate the probability that you will never win and compare this with the exact answer.

Solution: [Durrett, 2009, Q2.41] Let W be the number of wins. Then, $W \sim \text{Binomial}(250, p)$ where $p = 1/1000$. Hence,

$$P[W = 0] = \binom{250}{0} p^0 (1-p)^{250} \approx 0.7787.$$

If we approximate W by $\Lambda \sim \mathcal{P}(\alpha)$. Then we need to set

$$\alpha = np = \frac{250}{1000} = \frac{1}{4}.$$

In which case,

$$P[\Lambda = 0] = e^{-\alpha} \frac{\alpha^0}{0!} = e^{-\alpha} \approx 0.7788$$

which is very close to the answer from direct calculation.

Problem 8. Suppose X is a random variable whose pmf at $x = 0, 1, 2, 3, 4$ is given by $p_X(x) = \frac{2x+1}{25}$. Determine its expected value and variance. [Montgomery and Runger, 2010, Q3-51]

Solution:

$$\begin{aligned}\mathbb{E}X &= \sum_{x=0}^{4} x p_X(x) = \sum_{x=0}^{4} x \frac{2x+1}{25} = 0 + 1\left(\frac{3}{25}\right) + 2\left(\frac{5}{25}\right) + 3\left(\frac{7}{25}\right) + 4\left(\frac{9}{25}\right) \\ &= \frac{70}{25} = \frac{14}{5} = \boxed{2.8.}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= \sum_{x=0}^{4} x^2 p_X(x) = \sum_{x=0}^{4} x^2 \frac{2x+1}{25} = 0 + 1^2\left(\frac{3}{25}\right) + 2^2\left(\frac{5}{25}\right) + 3^2\left(\frac{7}{25}\right) + 4^2\left(\frac{9}{25}\right) \\ &= \frac{230}{25} = \frac{46}{5} = 9.2\end{aligned}$$

$$\operatorname{Var} X = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 = 9.2 - 2.8^2 = \boxed{1.36.}$$

Problem 9. An article in Information Security Technical Report ["Malicious Software—Past, Present and Future" (2004, Vol. 9, pp. 618)] provided the data (shown in Figure 4.3) on the top ten malicious software instances for 2002. The clear leader in the number of registered incidences for the year 2002 was the Internet worm "Klez". This virus was first

Place	Name	% Instances
1	I-Worm.Klez	61.22%
2	I-Worm.Lentin	20.52%
3	I-Worm.Tanatos	2.09%
4	I-Worm.BadtransII	1.31%
5	Macro.Word97.Thus	1.19%
6	I-Worm.Hybris	0.60%
7	I-Worm.Bridex	0.32%
8	I-Worm.Magistr	0.30%
9	Win95.CIH	0.27%
10	I-Worm.Sircam	0.24%

Figure 4.3: The 10 most widespread malicious programs for 2002 (Source—Kaspersky Labs).

detected on 26 October 2001, and it has held the top spot among malicious software for the longest period in the history of virology.

Suppose that 20 malicious software instances are reported. Assume that the malicious sources can be assumed to be inde- pendent.

(a) What is the probability that at least one instance is "Klez"?

(b) What is the probability that three or more instances are "Klez"?

(c) What are the expected value and standard deviation of the number of "Klez" instances among the 20 reported?

Solution: Let N be the number of instances (among the 20) that are "Klez". Then, $N \sim$binomial(n,p) where $n = 20$ and $p = 0.6122$.

(a) $P[N \geq 1] = 1 - P[N < 1] = 1 - P[N = 0] = 1 - p_N(0) = 1 - \binom{20}{0} \times 0.6122^0 \times 0.3878^{20} \approx 0.9999999941 \approx 1$.

(b)

$$P[N \geq 3] = 1 - P[N < 3] = 1 - (P[N = 0] + P[N = 1] + P[N = 2])$$
$$= 1 - \sum_{k=0}^{2} \binom{20}{k} (0.6122)^k (0.3878)^{20-k} \approx 0.999997$$

(c) $\mathbb{E}N = np = 20 \times 0.6122 = 12.244$.
$\sigma_N = \sqrt{\operatorname{Var} N} = \sqrt{np(1-p)} = \sqrt{20 \times 0.6122 \times 0.3878} \approx 2.179$.

Problem 10. The random variable V has pmf

$$p_V(v) = \begin{cases} \frac{1}{v^2} + c, & v \in \{-2, 2, 3\} \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of the constant c.

(b) Find $P[V > 3]$.

(c) Find $P[V < 3]$.

(d) Find $P[V^2 > 1]$.

(e) Let $W = V^2 - V + 1$. Find the pmf of W.

(f) Find $\mathbb{E}V$

(g) Find $\mathbb{E}[V^2]$

(h) Find $\operatorname{Var} V$

(i) Find σ_V

(j) Find $\mathbb{E}W$

Solution:

(a) The pmf must sum to 1. Hence,

$$\frac{1}{(-2)^2} + c + \frac{1}{(2)^2} + c + \frac{1}{(3)^2} + c = 1.$$

The value of c must be

$$c = \frac{1}{3}\left(1 - \frac{1}{4} - \frac{1}{4} - \frac{1}{9}\right) = \boxed{\frac{7}{54}} \approx 0.1296$$

Note that this gives

$$p_V(-2) = p_V(2) = \frac{41}{108} \approx 0.38 \quad \text{and} \quad p_V(3) = \frac{13}{54} \approx 0.241.$$

(b) $P[V > 3] = \boxed{0}$ because all elements in the support of V are ≤ 3.

(c) $P[V < 3] = 1 - p_V(3) = \frac{41}{54} \approx 0.759$.

(d) $P[V^2 > 1] = \boxed{1}$ because the square of any element in the support of V is > 1.

(e) $W = V^2 - V + 1$. So, when $V = -2, 2, 3$, we have $W = 7, 3, 7$, respectively. Hence, W takes only two values, 7 and 3. the corresponding probabilities are

$$P[W = 7] = p_V(-2) + p_V(3) = \frac{67}{108} \approx 0.62.$$

and

$$P[W = 3] = p_V(2) = \frac{41}{108} \approx 0.38.$$

Hence, the pmf of W is given by

$$p_W(w) = \begin{cases} \frac{41}{108}, & w = 3, \\ \frac{67}{108}, & w = 7, \\ 0, & \text{otherwise.} \end{cases} \approx \boxed{\begin{cases} 0.38, & w = 3, \\ 0.62, & w = 7, \\ 0, & \text{otherwise.} \end{cases}}$$

(f) $\mathbb{E}V = \frac{13}{18} \approx 0.7222$

(g) $\mathbb{E}V^2 = \frac{281}{54} \approx 5.2037$

(h) $\operatorname{Var} V = \mathbb{E}V^2 - (\mathbb{E}V)^2 = \frac{1517}{324} \approx 4.682$.

(i) $\sigma_V = \sqrt{\operatorname{Var} V} \approx 2.1638$

(j) $\mathbb{E}W = 5.4815$