Introduction to statistics

author: Alejandro Cáceres date:
autosize: true

Barcelona East School of Engineering
Universitat Politècnica de Catalunya (UPC)

Problems in joint probability

Probability summary

Properties:

- $P(S) = 1$
- $P(A) + P(\bar{A}) = 1$
- $P(A \cup B) = P(A) + P(B)$, when $P(A \cap B) = 0$
- $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$
- Sum rule: $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

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

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

- Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Multiplication rule:

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

- A and B are statistically independent if and only if

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

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

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

- Total probability rule

$P(A) = P(A|E1)P(E1) + P(A|E2)P(E2) + \ldots + P(A|En)P(En)$,

When $\{E1, E2, \ldots En\}$ are mutually exclusive

- Baye's theorem $P(Ei|B) = \frac{P(B \cap Ei)}{P(B \cap E1) + \ldots + P(B \cap E1)}$

Problem 1

A brick test is defined by the following events on a brick:

- Defective: D, non-defective: $\bar{D}$
- Pass quality test: E, do no pass quality test: $\bar{E}$

The brick test has:

- sensitivity: $P(E|\bar{D}) = 0.99$
- especificity: $P(\bar{E}|D) = 0.98$
- positive test marginal probability: $P(E) = 0.893$

a. what is the probability that a brick chosen at random is defective, P(D)?

	$\bar{D}$	D
E	P($E|\bar{D}$)=0.99	P($E|D$)=0.02
$\bar{E}$	P($\bar{E}|\bar{D}$)=0.01	P($\bar{E}|D$)=0.98
sum	1	1

	$\bar{D}$	D	**sum**
E	$P(E \cap \bar{D}) = 0.99 \times P(\bar{D})$	$P(E \cap D) = 0.02 \times P(D)$	$P(E) = 0.893$
$\bar{E}$	$P(\bar{E} \cap \bar{D}) = 0.01 \times P(\bar{D})$	$P(\bar{E} \cap D) = 0.98 \times P(D)$	$P(\bar{E})$
sum	$P(\bar{D})$	$P(D)$	1

$$P(E) = P(E \cap \bar{D}) + P(E \cap D)$$

$$= P(E|\bar{D})P(\bar{D}) + P(E|D)P(D)$$

$$= P(E|\bar{D})(1 - P(D)) + P(E|D)P(D)$$

We solve for $P(D)$

$$P(D) = \frac{P(E)-P(E|\bar{D})}{P(E|D)-P(E|\bar{D})} = \frac{0.893-0.99}{0.02-0.99} = 0.1$$

b. The probability that a brick is not defective **and** that it does not pass the test

$$P(\bar{E} \cap \bar{D}) = P(\bar{E}|\bar{D})P(\bar{D}) = 0.01 * (1 - 0.1) = 0.009$$

c. Are D and $\bar{E}$ statistical independent?

- if so then we should have $P(\bar{E}|D) = P(\bar{E})$
- as we have $P(\bar{E}|D) = 0.98$

and $P(\bar{E}) = 1 - P(E) = 0.107$ then D and $\bar{E}$ are not independent.

Problem 2

There are three traffic lights on the road. The probability of finding the first one in red is 0.6. For the other two we have the probabilities $P(R_{j+1}|R_j) = 0.15$ and $P(R_{j+1}|\bar{R}_j) = 0.25$ for $j = 1, 2$, where R_j is the event of finding the j-th traffic light in red. If the probability of the one traffic light depends only on the previous one then what is the probability that when passing by the road you

- find all traffic lights in red
- find at least one traffic light in red
- find only one traffic light in red

The problem gives us the conditional matrices

	R_{j+1}	$\bar{R}_{j+1}$	**sum**
R_j	$P(R_{j+1}|R_j) = 0.15$	$P(\bar{R}_{j+1}|R_j) = 0.85$	1
$\bar{R}_j$	$P(R_{j+1}|\bar{R}_j) = 0.25$	$P(\bar{R}_{j+1}|\bar{R}_j) = 0.75$	1

Therefore

	R_2	$\bar{R}_2$	sum
R_1	0.15	0.85	1
$\bar{R}_1$	0.25	0.75	1

	R_3	$\bar{R}_3$	sum
R_2	0.15	0.85	1
$\bar{R}_2$	0.25	0.75	1

- $P(R_1) = 0.6$

a. find all traffic lights in red $P(R_1 \cap R_2 \cap R_3)$

- let's write it in terms of conditional probabilities. Let's expand

$P(R_1 \cap R_2 \cap R_3) = P(R_2 \cap R_3|R_1)P(R_1)$ because

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

where $A = R_2 \cap R_3$ and $B = R_1$

- let's expand further

$P(R_1 \cap R_2 \cap R_3) = P(R_2 \cap R_3|R_1)P(R_1)$

$= P(R_3|R_2, R_1)P(R2|R_1)P(R_1)$

because

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

where $A = R_3$, $B = R_2$ and the probability is conditional $P = P(|R_1)$

Now we have it in terms of conditional probabilities

$P(R_1 \cap R_2 \cap R_3) = P(R_3|R_2, R_1)P(R_2|R_1)P(R_1)$

- Since R_3 is independent of R_1 then $P(R_3|R_1) = P(R_3)$, we can then remove the conditional on R_1. Therefore, $P(R_3|R_2, R_1) = P(R_3|R_2)$

Then:

$P(R_1 \cap R_2 \cap R_3) = P(R_3|R_2)P(R_2|R_1)P(R_1)$

Taking the values form the conditional matrices

$P(R_1 \cap R_2 \cap R_3) = P(R_3|R_2)P(R2|R_1)P(R_1)$

$= 0.15 * 0.15 * 0.6 = 0.0135$

b. find at least one traffic light in red $P(R_1 \cup R_2 \cup R_3)$

Let's remember set algebra!

- $\overline{A \cup B} = \bar{A} \cap \bar{B}$

We turn the **unions** into **intersections** into **conditional probabilities**

$P(R_1 \cup R_2 \cup R_3) = 1 - P(\overline{R_1 \cup R_2 \cup R_3})$

$= 1 - P(\bar{R}_1 \cap \bar{R}_2 \cap \bar{R}_3)$

$= 1 - [P(\bar{R}_3|\bar{R}_2)P(\bar{R}_2|\bar{R}_1)P(\bar{R}_1)]$

$= 1 - 0.4 * 0.75 * 0.75 = 0.775$

c. find only one traffic light in red

$P([R_1 \cap \bar{R}_2 \cap \bar{R}_3] \cup [\bar{R}_1 \cap R_2 \cap \bar{R}_3] \cup [\bar{R}_1 \cap \bar{R}_2 \cap R_3])$

$= P(R_1 \cap \bar{R}_2 \cap \bar{R}_3) + P(\bar{R}_1 \cap R_2 \cap \bar{R}_3) + P(\bar{R}_1 \cap \bar{R}_2 \cap R_3)$

Because they are mutually exclusive

$P(R_1 \cap \bar{R}_2 \cap \bar{R}_3) + P(\bar{R}_1 \cap R_2 \cap \bar{R}_3) + (\bar{R}_1 \cap \bar{R}_2 \cap R_3)$

$= P(\bar{R}_3|\bar{R}_2)P(\bar{R}_2|R_1)P(R_1) + P(\bar{R}_3|R_2)P(R_2|\bar{R}_1)P(\bar{R}_1)$

$+P(R_3|\bar{R}_2)P(\bar{R}_2|\bar{R}_1)P(\bar{R}_1)$

remember the general form for conditioning:

$$P(R_1 \cap R_2 \cap R_3) = P(R_3|R_2)P(R_2|R_1)P(R_1)$$

R_1 conditions R_2 that conditions R_3.

We just need to replace all the information that we have so far

$P(\bar{R}_3|\bar{R}_2)P(\bar{R}_2|R_1)P(R_1) + P(\bar{R}_3|R_2)P(R_2|\bar{R}_1)P(\bar{R}_1)$

$+P(R_3|\bar{R}_2)P(\bar{R}_2|\bar{R}_1)P(\bar{R}_1)$

$= 0.75 * 0.85 * 0.6 + 0.85 * 0.25 * 0.4 + 0.25 * 0.75 * 0.4 = 0.5425$

Problem 3

Game of tossing 2 coins and a dice

wining events:

- Two heads and a pair.
- one head one tail and a number greater than 5.

If we know that someone has won, what is the probability that he did it with the first event?

Probability for the following events

- getting a pair number on the dice: $P(A) = 1/2$
- getting a number greater than 5 in the dice: $P(B) = 2/6$
- getting two heads from the coins: $P(C) = 1/2 * 1/2 = 1/4$
- getting one head and one tail: $P(D) = P(H,T) + P(T,H) = 1/4 + 1/4 = 1/2$

The probability of winning is:

$P(G) = P(C \cap A \cup D \cap B)$

$= P(C \cap A) + P(D \cap B) = P(C)P(A) + P(D)P(B) = 7/24$

and the probability of having two heads when wining (or wining by two heads)

$P(C|G) = \frac{P(C \cap A)}{P(G)} = \frac{1/8}{7/24} = 3/7$

Problem 4

Capacitors are stored as following

μ F	box 1	box 2	box 3	Total
0.01	20	95	25	140
0.1	55	35	75	165
1.0	70	80	145	295
Total	145	210	245	600

Conditional matrix from the experiment

μ F	box 1	box 2	box 3
0.01	20/145	95/210	25/245
0.1	55/145	35/210	75/245
1.0	70/145	80/210	145/245
sum	1	1	1

$P(c1) = P(c2) = P(c3) = 1/3$

if we choose a box and a capacitor at random

a. what is the probability that we select box 2 (c2) and a capacitor of $0.1\mu F$?

$P(0.1\mu F \cap c2) = P(0.1\mu F|c2)P(c2) = 35/210 * 1/3 = 210/600$

b. what is the probability to select a capacitor of $0.1\mu F$?

$P(0.1\mu F) = P(0.1\mu F|c1)P(c1) + P(0.1\mu F|c2)P(c2) + P(0.1\mu F|c3)P(c3)$

$= \frac{20}{145} * 1/3 + \frac{95}{210} * 1/3 + \frac{25}{245} * 1/3 = 0.23078$

Problem 5

A pack of 50 washer rings, 30 of which exceed the required thickness.

a. if three rings are picked randomly, what is the probability that the three rings exceed the required thickness?

if A_i is the event that the i-th ring exceeds the required thickness then

$P(A_1 \cap A_2 \cap A_3) = P(A_3 \cap A_2|A_1)P(A_1) = P(A_3|A_2, A_1)P(A_2|A_1)P(A_1) = 28/48 * 29/49 * 30/50 = 0.20714$ is the probability that the three rings exceed the required thickness.

Problem 5

b. what is the probability that the third ring exceeds the required thickness?

$P(A_3) = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap \bar{A}_2 \cap A_3) + P(\bar{A}_1 \cap A_2 \cap A_3) + P(\bar{A}_1 \cap \bar{A}_2 \cap A_3)$

$P(A_3 \cap A_2 \cap A_1) = P(A_3|A_2, A_1)P(A_2|A_1)P(A_1) = 28/48 * 29/49 * 30/50$

$P(A_3 \cap \bar{A}_2 \cap A_1) = P(A_3|\bar{A}_2, A_1)P(\bar{A}_2|A_1)P(A_1) = 29/48 * 20/49 * 30/50$

$P(A_3 \cap A_2 \cap \bar{A}_1) = P(A_3|A_2, \bar{A}_1)P(A_2|\bar{A}_1)P(\bar{A}_1) = 29/48 * 30/49 * 20/50$

$P(A_3 \cap \bar{A}_2 \cap \bar{A}_1) = P(A_3|\bar{A}_2, \bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(\bar{A}_1) = 30/48 * 19/49 * 20/50$

Summing up $P(A3) = 0.6$

Problem 6

What is the probability of a successful satellite launch from Florida P(A)?

- The probability of a successful satellite launch from Florida P(A) is greater than a launch from Japan P(B)
- The probability that one satellite from Florida **OR** one from Japan is successful is $P(A \cup B) = 0.626$
- The probability that one satellite from Florida **AND** one from Japan are successful is $P(A \cap B) = 0.144$

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.626$
2. $P(A \cap B) = P(A)P(B) = 0.144$

1 and 2. $-P(A)^2 + (0.626 + 0.144)P(A) - 0.144 = 0$ whose solutions are {0.45, 0.32}, since $P(A) < P(B)$ then $P(A) = 0.45$ and $P(B) = 0.32$.

Problems in conditional probability

Problem 1

a. What is the probability that a brick that has passed the test is really defective?

$P(D|E) = \frac{P(E|D)P(D)}{P(E)} = \frac{P(E|D)P(D)}{P(E)}$

$P(D|E) = \frac{0.02*0.01}{0.893} = 0.0022$

Problem 2

Comunication chanel with

- input {a,b} and output {1,0}
- probabilities: $Pr(a) = 0.6$, $Pr(b) = 0.4$, $Pr(0|a) = 0.2$ and $Pr(0|b) = 0.6$
- rule: $P(input = x|output = z) \geq P(input = y|output = z)$ then input x is assigned to output z (i.e. $P(b|0) \geq P(a|0)$ then b is assiged to 0).

	a	**b**
0	P(0|a)=0.2	P(0|b)=0.6
1	P(1|a)=0.8	P(1|b)=0.4
sum	1	1

a. If the output was 0 what was the input?

	a	**b**	**sum**
0	P(a|0)	P(b|0)	1
1	P(a|1)	P(b|1)	1

- $P(a|0) = \frac{P(0|a)P(a)}{P(0)} = \frac{0.2*0.6}{P(0)}$
- $P(b|0) = \frac{P(0|b)P(b)}{P(0)} = \frac{0.6*0.4}{P(0)}$

then $P(b|0) > P(a|0)$ and we asign b to 0

b. If the output was 1 what was the input?

- $P(a|1) = \frac{P(1|a)P(a)}{P(1)} = \frac{[1-P(0|a)]P(a)}{P(1)} = \frac{0.8*0.6}{P(1)}$
- $P(b|1) = \frac{P(1|b)P(b)}{P(1)} = \frac{[1-P(0|b)]P(a)}{P(1)} = \frac{0.4*0.4}{P(1)}$ then $P(a|1) > P(b|1)$ and we asign a to 1

c. What is the error rate?

	a	**b**	**sum**
0	P(0,a)	**P(0,b)**	P(0)
1	**P(1,a)**	P(1,b)	P(1)

	a	b	sum
sum	P(a)	P(b)	1

probabiity of error: $P(0, a) + P(1, b)$

$P(0, a) + P(1, b) = P(0 \cap a) + P(1 \cap b)$

$= P(0|a)P(a) + P(1|b)P(b) = 0.2 * 0.6 + 0.4 * 0.4 = 0.28$

Problem 3

20 item-pack of electronic devices is distributed to costumers with the following features

if D_i is the event that a pack contains i defective items

- $D_i \in 0, 1, 2$
- $P(D_0) = 0.6$
- $P(D_2) = 0.1$

a. What is the probability of any number of items being defective?

$P(\bar{D_0}) = 1 - P(D_0) = 1 - 0.6 = 0 - 4$

b. if we choose two items randomly in a pack, what is the probability that both items are defective?

d_i is the event of selecting a defective item in a pack.

Probability of first selecting a pack with two defective items **AND** then selecting the two defective items in two random draws: $P(D_2 \cap d_2)$

$P(D_2 \cap d_2) = P(d2|D2)P(D2)$

$= (\frac{2}{20} * \frac{1}{19}) * 0.1 = 0.0005263$

c. After a pack is randomly chosen and then two items are also randomly chosen, no defective item is observed. What is the proability that the pack does not have any defective item?

$P(D_0|d_0) = \frac{P(d_0|D_0)P(D_0)}{P(d_0)}$

What is $P(d_0)$?

$P(d_0) = P(d_0|D_0)P(D_0) + P(d_0|D_1)P(D_1) + P(d_0|D_2)P(D_2)$

$= 1 * 0.6 + (\frac{19}{20}\frac{18}{19}) * 0.3 + (\frac{18}{20}\frac{17}{19}) * 0.1$

then $P(D_0|d_0) = \frac{0.6}{1*0.6+(\frac{19}{20}\frac{18}{19})*0.3+(\frac{18}{20}\frac{17}{19})*0.1} = 0.6312$

Problem 4

We have a test for detecting bacteria in a water sample. P is the event of a positive test and C is the event that the water is contaminated. The diagnosis has the following features

- $P(P|C) = 0.7$ then $P(\bar{P}|C) = 0.3$
- $P(\bar{P}|\bar{C}) = 0.6$ then $P(P|\bar{C}) = 0.4$
- $P(C) = 0.2$ then $P(\bar{C}) = 0.8$

a. What is the probability that if the test is positive the sample has a bacteria, $P(C|P)$?

Bayes theorem $P(C|P) = \frac{P(P|C)*P(C)}{P(P|C)*P(C)+P(P|\bar{C})*P(\bar{C})}$

$P(C|P) = \frac{0.7*0.2}{0.7*0.2+0.4*0.8} = 0.3043$

Problem 5

Epidemiological facts on people with hypertension (H) and knowing they are hypertense (S):

- $P(H) = 0.2$
- $P(S|H) = 0.7$
- $P(\bar{S}|\bar{H}) = 0.6$

If a person does not know she is hypertense what is the probability that she is hypertense?

$$P(H|\bar{S}) = \frac{P(\bar{S}|H)P(H)}{P(\bar{S})}$$

$$= \frac{P(\bar{S}|H)P(H)}{P(\bar{S}|H)P(H)+P(\bar{S}|\bar{H})P(\bar{H})} = \frac{[1-P(S|H)]P(H)}{[1-P(S|H)]P(H)+P(\bar{S}|\bar{H})[1-P(H)]} = 0.111$$

Problem 6

See solution of problem 4 from joint probability exercises.

a. if av$1.0\mu F$ has been selected, what is the probability that it belonged to box 1?

$$P(c1|1.0\mu F) = \frac{P(1.0\mu F|c1)P(c1)}{P(1.0\mu F)} = \frac{70/145*1/3}{P(1.0\mu F)}$$

We then compute

$$P(1.0\mu F) = \frac{70}{145} * 1/3 + \frac{80}{210} * 1/3 + \frac{145}{245} * 1/3 = 0.4851$$

therefore

$$P(c1|1.0\mu F) = 0.33167$$

Problem 7

See solution to problem 6 of joint probability exercises

a. what is the probability that the system was not tested if it was found to fail, $P(O|F)$?

- not tested: O, tested $\bar{O}$
- failed: F, not failed $\bar{F}$

Probabilities:

$P(O) = 2/3$, probability of not testing.

$P(F|O) = 1/4$, probability of failing if not tested.

$P(F|\bar{O}) = P(\bar{F}|\bar{O})$ probabilities of failure and not failure are equall when tested. Then $P(F|\bar{O}) = 1/2$ because $P(F|\bar{O}) + P(\bar{F}|\bar{O}) = 1$.

From Baye's theorem we can then compute

$$P(O|\bar{F}) = \frac{P(\bar{F}|O)P(O)}{P(\bar{F})} = \frac{[1-P(F|O)]P(O)}{1-P(F)}$$

We miss $P(F)$

$$P(F) = P(F|O)P(O) + P(F|\bar{O})P(\bar{O})$$

$$= P(F|O)P(O) + P(F|\bar{O})[1 - P(O)] = 1/3,$$

then $P(O|\bar{F}) = 3/4$

Problems discrete variables

Problem 1

- $x \in \{0, 1, 2, 3\}$
- $P(X = 0) = 2P(X = 1)$
- $E(X) = 2$
- $E(X^2) = 5$

a. what is $f(x) = P(X = x)$?

We solve a system of 4 equations
i) nromalization: $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$
ii) expectation: $P(X = 0) * 0 + P(X = 1) * 1 + P(X = 2) * 2 + P(X = 3) * 3 = 2$
iii) variance: $P(X = 0) * 0^2 + P(X = 1) * 1^2 + P(X = 2) * 2^2 + P(X = 3)3^2 = 5$
iv) $P(X = 0) - 2P(X = 1) = 0$
whose solution is
$f(0) = P(X = 0) = \frac{1}{7}$
$f(1) = P(X = 1) = \frac{1}{14}$
$f(2) = P(X = 2) = \frac{3}{7}$
$f(3) = P(X = 3) = \frac{5}{14}$

b. The variance can ebe obtained from $V(X) = E(X^2) - E(X)^2 = 5 - 2^2 = 1$

c. What is the range of the variable $Y = 2X - 1$?
$Y \in -1, 1, 3, 5$

- $P(3 < Y \leq 5) = P(2 < X \leq 3) = P(X = 3) = \frac{5}{14}$
- $E(Y) = 2E(X) - 1 = 3$
- $V(Y) = 2^2 V(X) = 4$

Problem 2

x	$P(X = x)$
10	0.1
12	0.3
14	0.25
15	0.15
17	
20	0.15

a. what is $E(X)$ and $\sigma = \sqrt{V(X)}$?
fist we compute $P(X = 17) = 1 - P(X \neq 17) = 0.05$

Then we explicitly calculate $E(X) = \sum_i^m x_i * P(X = x_i)$

$$E(X) = 0.1 * 10 + 0.3 * 12 + 0.25 * 14 + 0.15 * 15 + 0.05 * 17 + 0.15 * 20 = 14.2$$

and $V(X)$ from the expression: $V(X) = E(X^2) - E(X)^2$

$E(X^2) = 0.1 * 10^2 + 0.3 * 12^2 + 0.25 * 14^2 + 0.15 * 15^2 + 0.05 * 17^2 + 0.15 * 20^2 = 22.96$

then $V(X) = E(X^2) - E(X)^2 = 22.96 - 14.2 = 8.76$ and $\sigma = \sqrt{8.76} = 2.96$

Problem 3

Given the cumulative distribution for a discrete variable X

$$F(x) = \begin{cases} 0, & x \leq -1 \\ 0.2, & x \in [-1, 0) \\ 0.35, & x \in [0, 1) \\ 0.45, & x \in [1, 2) \\ 1, & x \geq 2 \end{cases}$$

a. find $E(X)$ and $V(X)$

$E(X) = \sum_i^4 x_i * f(x_i)$ therefore we need to find the probability mass distribution $f(x_i)$

$$f(x) = \begin{cases} 0, & x < -1 \\ F(0) - F(-1) = 0.2, & x = -1 \\ F(1) - F(0) = 0.15, & x = 0 \\ F(2) - F(1) = 0.10, & x = 1 \\ 1 - F(2) = 0.55, & x = 2 \end{cases}$$

therefore:

- $E(X) = -0.2 * 1 + 0.5 * 0 + 0.1 * 1 + 0.55 * 2 = 1$
- $V(X) = E(X^2) - E(X)^2 = 0.2 * 1^2 + 0.5 * 0^2 + 0.1 * 1^2 + 0.55 * 2^2 - 1 = 1.5$ then $\sigma = \sqrt{1.5}$

If $Y = 2X - 3$ then

- $E(Y) = E(2X - 3) = 2E(X) - 3 = -1$
- $V(Y) = V(2X - 3) = 2^2 V(X) = 4 * 1.5 = 6$

Problems continuous variables

Problem 1

- $$f(x) = \begin{cases} k(a + bx), & \text{if } x \in (0, 1) \\ 0, & x \notin (0, 1) \end{cases}$$
- $E[X] = 4/7$

a. Find a, b, k and $V(X)$.

Problem 1

We solve for a and b using two equations:

i. $E(X) = \int_0^1 k(a + bx)x dx = \frac{1}{2}kax^2 + \frac{1}{3}kbx^3\Big|_0^1$

$\frac{1}{2}ka + \frac{1}{3}kb = \frac{4}{7}$

ii. $\int_0^1 k(a+bx)dx = kax + \frac{1}{2}kbx^2\Big|_0^1$

$ka + \frac{1}{2}kb = 1$

Whose solution of b and k as a function of a is

1. $b = \frac{3}{2}a$
2. $k = \frac{4}{7a}$

Replacing

$$f(x) = \begin{cases} \frac{4}{7} + \frac{12}{14}x, & \text{if } x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

$a \neq 0$

The variance can be computed from the relation

$V(X) = E(X^2) - E(X)^2 = \int_0^1 (\frac{4}{7} + \frac{12}{14}x)x^2 dx - (\frac{4}{7})^2$

$= \frac{1}{3}x^3\frac{4}{7}\big|_0^1 + \frac{1}{4}x^4\frac{12}{14}\big|_0^1 - \frac{16}{49}$

$= \frac{25}{294}$

b. Find F(X)

$F(x) = \int_0^x (\frac{4}{7} + \frac{12}{14}t)dt$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \int_0^x (\frac{4}{7} + \frac{12}{14}t)dt = \frac{4}{7}t\big|_0^x + \frac{1}{2}\frac{12}{14}t^2\big|_0^x = \frac{4}{7}x + \frac{3}{7}x^2, & x \in (0,1) \\ 0, & x \geq 1 \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{4}{7}x + \frac{3}{7}x^2, & x \in (0,1) \\ 0, & x \geq 1 \end{cases}$$

c. median of X

$F(q_{0.5}) = \frac{1}{2} = \frac{4}{7}q_{0.5} + \frac{3}{7}q_{0.5}^2$

solving the quadratic $-7 + 8q_{0.5} + 6q_{0.5}^2 = 0$

gives solutions $q_{0.5} = -1.9, 0.6$

0.6 is in the correct range.

Problem 2

Given

$$F(x) = \begin{cases} 0, & x \leq 2 \\ \frac{x^2-a}{b}, & x \in (2,4) \\ 0, & x > 4 \end{cases}$$

a. find a and b

1. from normalization, that is, the sum of probabilities across the range of x add to 1
   $F(4) - F(2) = 1$
   $\frac{4^2-a}{b} - \frac{2^2-a}{b} = 1$ solving for b
   $b = 12$

2. from continuity at x=2

$F(2) = 0$ then $\frac{2^2-a}{b} = 0$ and $a = 4$.

b. what is $\sigma = \sqrt{V(X)}$?

We first find the density function by differentiation.

$f(x) = \frac{dF(x)}{dx} = \frac{d}{dx}\frac{x^2-4}{12} = \frac{x}{6}$

then we compute the variance
$V(X) = E(X^2) - E(X)^2 = \int_2^4 x^2 \frac{x}{6} dx - [\int_2^4 x \frac{x}{6} dx]^2$
$= \frac{1}{4}\frac{x^4}{6}\Big|_2^4 - [\frac{1}{3}\frac{x^3}{6}]^2$
$= 0.32$

then $\sigma = \sqrt{0.32} = 0.56$

c. $P(X < 3.5|X > 3) = \frac{P(X<3.5 \cap X>3)}{P(X>3)} = \frac{P(3<X<3.5)}{1-P(X\leq 3)} = \frac{F(3.5)-F(3)}{1-F(3)} = 0.46$

Problem 3

Given

$$f(x) = \begin{cases} \frac{a}{\pi(1+x^2)}, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

a. compute a.

By normalization we have $\int_0^1 f(x)dx = \frac{a}{\pi}\tan^{-1}(x)\Big|_0^1 = \frac{a}{\pi}\frac{\pi}{4} = 1$ then $a = 4$

b. compute $E(X) = \int_0^1 \frac{ax}{\pi(1+x^2)} dx$

$= \frac{4}{\pi}\tan^{-1}(x)x\Big|_0^1 - \frac{4}{\pi}\int_0^1 \tan^{-1}(x)dx$

$= \frac{4}{\pi}\tan^{-1}(x)x\Big|_0^1 - \frac{4}{\pi}\tan^{-1}(x)x\Big|_0^1 + \frac{4}{\pi}\frac{1}{2}\log(x^2+1)\Big|_0^1 = 0.441$

c. compute $F(X)$

We had already computed the indefinite integral: $F(x) = \int f(t)dt$, we now set the limits of intergration from 0 to x
$F(x) = \int_0^x f(t)dt$ within (0,1).

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{4}{\pi}\tan^{-1}(x), & x \in (0,1) \\ 0, & x \geq 1 \end{cases}$$

Problem 4

Given

$$f(x) = \begin{cases} \frac{x}{\theta^2}e^{-x^2/(2\theta^2)}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

a. prove that it is a probability density function.

We probe that $f(x)$ satisfies the the normalization condition $\int_0^\infty f(x)dx = 1$

Change variable $z = \frac{x^2}{2\theta^2}$ then the differential changes to $dz = \frac{x}{\theta^2}dx$

$\int_0^\infty f(x)dx = \int_0^\infty e^{-z}dz = -e^{-z}\Big|_0^\infty = 1$

b. compute $F(x)$

- for $z > 0$

$F(z) = \int_0^z e^{-t}dt = -e^{-t}\Big|_0^z = -e^{-z} + 1$

changing variables back

$F(x) = 1 - e^{-x^2/(2\theta^2)}$

- for $x \leq 0$, $F(x) = 0$

Problem 5

Given

$$f(x) = \begin{cases} 0, & x < 0 \\ ax, & x \in [0,3] \\ b, & x \in (3,5) \\ \frac{b}{3}(8-x), & x \in [5,8] \\ 0, & x > 8 \end{cases}$$

a. compute a and b such that $f(x)$ is a probability density function.

b. by continuity $\lim_{x\to 3} f(x) = f(3)$ from both sides then:
$3 * a = b$

ii. by normalization $\int_{-\infty}^\infty f(x)dx = 1$ and i)

$\int_0^3 axdx + \int_3^5 3adx + \int_5^8 a(8-x)dx$

$\frac{a}{2}x^2\Big|_0^3 + 3ax\Big|_3^5 + a8x\Big|_5^8 - \frac{a}{2}x^2|_5^8 = 1$

replacing and solving for a then $a = \frac{1}{15}$ and by i) $b = \frac{1}{5}$

b. compute $P(3 < X < 5|3 < X)$

$P(3 < X < 5|3 < X) = \frac{P(3<X<5 \cap 3<X)}{P(3<X)}$

$= \frac{P(3<X<5)}{P(3<X)} = \frac{P(3<X<5)}{1-P(X\leq 3)}$

on one hand: $P(3 < X < 5) = \int_3^5 \frac{1}{5}dx = \frac{1}{5}x\Big|_3^5 = \frac{2}{5}$

on the other hand: $P(X \leq 3) = \int_0^3 \frac{1}{15}xdx = \frac{1}{30}x^2\Big|_0^3 = \frac{3}{10}$

then $P(3 < X < 5|3 < X) = \frac{2/5}{1-3/10} = 4/7$

Problem 6

Given the cumulative distribution for a random variable X

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{80}(17 + 16x - x^2), & x \in [-1,7) \\ 1, & x \geq 7 \end{cases}$$

a. compute $P(X > 0)$

$P(X > 0) = 1 - P(X \leq 0) = 1 - \frac{17}{80} = \frac{63}{80}$

b. compute $E(X) = \int_{-\infty}^{\infty} x f(x) dx$

We need to find $f(x)$ by differentiation

$$f(x) = \begin{cases} 0, & x \notin (1,7) \\ \frac{16}{80} - \frac{2x}{80}, & x \in (-1,7) \end{cases}$$

Then: $E(X) = \int_{-1}^{7} \frac{16}{80}x - \frac{2}{80}x^2 dx$

$= \frac{16}{80}\frac{x^2}{2} - \frac{2}{80}\frac{x^3}{3}\Big|_1^7$

$= \frac{48}{10} - \frac{344}{120} = 1.933$

c. compute $P(X > 0|X < 2)$

$P(X > 0|X < 2) = \frac{P(X>0 \cap X<2)}{P(X<2)}$

$= \frac{P(0<X<2)}{P(X<2)}$

$= \frac{F(2)-F(1)}{F(2)} = \frac{45-17}{17} = \frac{28}{45}$

Problem 7

Given

$$f(x) = \begin{cases} \frac{c}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

a. compute c so that the function is a probability density function and then $E(X)$.
$\int_1^{\infty} \frac{c}{x^3} = -\frac{1}{2}cx^{-2}\Big|_1^{\infty} = \frac{1}{2}c = 1$ then $c = 2$

then
$E(X) = \int_1^{\infty} \frac{2}{x^2} dx = -2x^{-1}\Big|_1^{\infty} = 2$

b. compute $F(X)$

$F(X) = \int_1^{x} \frac{2}{t^3} dt = \frac{1}{2}2t^{-2}\Big|_1^{x} = -\frac{1}{x^2} + 1$

b. compute $\frac{P(X \leq 2.5)}{P(X \leq 10)}$

$\frac{P(x \leq 2.5)}{P(x \leq 10)} = \frac{-1/(2.5^2)+1}{-1/(10^2)+1} = 0.84$