DS203: Programming for Data Sciences

Solution to Assignment 1

Q1

Let us denote quality of sample by the r.v $S = 0$(defective) and 1(good product). Also the event of a product being manufactured by a factory i is denoted by the r.v F and $F = 0$(for A) and 1(for B).

Part a We have

$$\begin{aligned} P(S=0) &= \sum_{i\in\{0,1\}} P(S=0|F=i) * P(F=i) \\ &= 0.8 * 0.3 + 0.2 * 0.1 \\ &= 0.26 \end{aligned}$$

Part b We have to find out

$$P(F=0|S=0) = \frac{P(S=0|F=0) * P(F=0)}{P(S=0)} = \frac{0.3 * 0.8}{0.26} = \frac{12}{13} \approx 0.92$$

Q2

Let A denote the r.v that access was successful(1) or not(0). Let S denote the r.v that server is working(1) or not(0). Given $P(S = 1) = 0.8$ and $P(A = 1|S = 1) = 0.9$

Part a

$$\begin{aligned} P(A=0) &= \sum_{i\in\{0,1\}} P(A=0|S=i) * P(S=i) \\ &= 0.1 * 0.8 + 1 * 0.2 = 0.28 \end{aligned}$$

Part b

$$P(S=1|A=0) = \frac{P(A=0|S=1) * P(S=1)}{P(A=0)} = \frac{0.1 * 0.8}{0.28} = \frac{2}{7} \approx 0.29$$

Part c

$$\begin{aligned} P(A_2=0|A_1=0) &= \frac{P(A_2=0, A_1=0)}{P(A_1=0)} \\ &= \frac{\sum_{i\in\{0,1\}} P(A_2=0, A_1=0|S=i)}{0.28} \\ &= \frac{0.2 + 0.8 * 0.1 * 0.1}{0.28} \approx 0.78 \end{aligned}$$

Part d

$$\begin{aligned}
P(S=1|A_2=0,A_1=0) =& \frac{P(S=1,A_2=0,A_1=0)}{P(A_2=0,A_1=0)} \\
=& \frac{P(S=1,A_2=0,A_1=0)}{P(A_2=0|A_1=0)*P(A_1=0)} \\
\approx& \frac{0.8*0.1*0.1}{0.78*0.28} \approx 0.038
\end{aligned}$$

Q3

Let the r.v associated with dice 1 be X_1 and for dice 2 be X_2.

Part a To find

$$\begin{aligned}
P(X_1=6 \text{ or } X_2=6) =& P(X_1=6)+P(X_2=6)-P(X_1=6,X_2=6) \\
=& \frac{1}{6}+\frac{1}{6}-\frac{1}{6}*\frac{1}{6}=\frac{11}{36}\approx 0.31
\end{aligned}$$

Here $X_1=6, X_2=6$ is 1 of the 36 total possibilities and therefore P$(X_1=6,X_2=6)=\frac{1}{36}$

Part b

$$\begin{aligned}
P(X_1=6 \text{ or } X_2=6|X_1\neq X_2) =& \frac{P(X_1=6 \text{ or } X_2=6, X_1\neq X_2)}{P(X_1\neq X_2)} \\
=& \frac{P(X_1=6,X_2\neq 6)+P(X_1\neq 6,X_2=6)}{\sum_{i=1}^{6} P(X_1=i)*P(X_2\neq i)} \\
=& \frac{\frac{1}{6}*\frac{5}{6}+\frac{5}{6}*\frac{1}{6}}{6*\frac{1}{6}*\frac{5}{6}}=\frac{1}{3}\approx 0.33
\end{aligned}$$

Q4

Let M, F and C denote that events that the randomly chosen person is a male, female and colorblind respectively. We want to find $P(M|C)$.
Applying Bayes Rule, we have

$$P(M|C)=\frac{P(C|M)\cdot P(M)}{P(C)}$$

Applying Total Probability Rule, we have

$$P(C)=(P(C|M)\cdot P(M))+(P(C|F)\cdot P(F))$$

Note that, here we are assuming M and F are mutually exclusive and exhaustive.
We know, $P(C|M)=0.05$, $P(C|F)=0.01$, and $P(M)=P(F)=0.5$. Plugging in the values, we have

$$P(C)=0.03$$

Hence,

$$P(M|C)=\frac{0.05*0.5}{0.03}=0.83333$$

Q5

Part a We know, for independent events A and B

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

Substituting A and B by E, we have

$$P(E \cap E) = P(E) \cdot P(E)$$

Notice that, $E \cap E = E$. Hence,

$$P(E) = P(E) \cdot P(E)$$
$$P(E)(P(E) - 1) = 0$$

Hence, $P(E) = 0$ or $P(E) = 1$.

Part b We know, for events A and B,

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

If A and B are independent, $P(A \cap B) = P(A) \cdot P(B)$. Hence,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - (P(A) \cdot P(B)) \\ &= 0.3 + 0.4 - (0.3 \cdot 0.4) \\ &= 0.58 \end{aligned}$$

If A and B are mutually exclusive, $P(A \cap B) = 0$.

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \\ &= 0.3 + 0.4 \\ &= 0.7 \end{aligned}$$

Part c If A and B were to be independent,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - (P(A) \cdot P(B)) \\ &= 0.6 + 0.8 - (0.6 \cdot 0.8) \\ &= 0.92 \\ &< 1 \end{aligned}$$

Since, there isn't any direct contradiction, we could say A and B can be independent.
If A and B were to be mutually exclusive,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \\ &= 0.6 + 0.8 \\ &= 1.4 \\ &> 1 \end{aligned}$$

Since, Probability of an event cannot be greater than 1, we arrive at a contradiction. Hence, A and B cannot be mutually exclusive.

Q6

We know, a valid CDF F must be non-decreasing, and right-continuous, and must satisfy

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1$$

Part 1 The given F satisfies all the properties of a valid CDF.

$$\begin{aligned} P(X^2 > 5) &= P(X > \sqrt{5} \cup X < -\sqrt{5}) \\ &= P(X > \sqrt{5}) + P(X < -\sqrt{5}) \end{aligned}$$

Notice that, $P(X < -\sqrt{5}) = F(-\sqrt{5}) = e^{-5}/4$.

$$\begin{aligned} P(X > \sqrt{5}) &= 1 - P(X \leq \sqrt{5}) \\ &= 1 - F(\sqrt{5}) & = 1 - (1 - e^{-5}/4) = e^{-5}/4 \end{aligned}$$

Hence,

$$P(X^2 > 5) = e^{-5}/4 + e^{-5}/4 = e^{-5}/2$$

Part 2 Notice that

$$\begin{aligned} F(0) &= 1.5 \\ &> 1 \end{aligned}$$

Hence, F is not a valid CDF.

Part 3 The given F satisfies all the properties of a valid CDF.

$$P(X < -\sqrt{5}) = F(-\sqrt{5}) = 0$$

$$\begin{aligned} P(X > \sqrt{5}) &= 1 - P(X \leq \sqrt{5}) \\ &= 1 - F(\sqrt{5}) & = 1 - (0.5 + \sqrt{5}/20) = 0.5 - \sqrt{5}/20 \end{aligned}$$

Hence,

$$P(X^2 > 5) = 0 + 0.5 - \sqrt{5}/20 = 0.5 - \sqrt{5}/20$$

Q7

Given the CDF ($F_X(x)$) plot (see that it is right continuous) ; we'll first convert to equation form:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 0.5, & \text{if } 0 \leq x \leq 1 \\ 0.5x, & \text{if } 1 \leq x \leq 2 \\ 1, & \text{otherwise} \end{cases}$$

Part 1 To find $P(X \leq 0.8)$:
From definition of cumulative density function (CDF):

$$\begin{aligned} Pr\{X \leq 0.8\} =& F_X(0.8) \\ =& 0.5 \times 0.8 \\ =& 0.4 \end{aligned}$$

Part 2 To find $E(X)$:
Since CDF is non-continuous (at points x=0); the X is a mixture of discrete and continuous random variable:
To calculate the discrete value probabilities; we look for points with jumps i.e. x=0 in the graph :

$$\begin{aligned} P(X = 0) &= lim_{a \to 0}(F_X(0) - F_X(a)) \\ &= 0.5 \end{aligned}$$

With this the probability density function can be defined as:

$$f_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 0, & \text{if } 0 < x < 1 \\ 0.5, & \text{if } 1 \leq x \leq 2 \\ 0, & \text{if } 2 < x \end{cases}$$

Thus:

$$\begin{aligned} E(X) =& Pr(X = 0) \times 0 + \int_{-\infty}^{\infty} f_X(t) t dt \\ =& 0 + \int_1^2 0.5 t dt \\ =& 0.25[t^2]_1^2 \\ =& 0.75 \end{aligned}$$

Part 2 To find $Var(X)$:

Since $Var(X) = E(X^2) - (E(X))^2$

$$\begin{aligned} E(X^2) =& Pr(X = 0) \times 0^2 + \int_{-\infty}^{\infty} f_X(t) t^2 dt \\ =& 0 + \int_1^2 0.5 t^2 dt \\ =& \frac{[t^3]_1^2}{6} \\ =& \frac{7}{6} \end{aligned}$$

Thus:

$$
\begin{aligned}
Var(X) =& E(X^2) - (E(X))^2 \\
=& \frac{7}{6} - (0.75)^2 \\
=& 1.1667 - 0.5625 \\
=& 0.6042
\end{aligned}
$$

Q8

Given probability density function:

$$
f_X(x) = \begin{cases} ce^{-2x}, & \text{if } 0 \leq x \leq \infty \\ 0, & \text{if } x < 0 \end{cases}
$$

Let us first calculate the CDF $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$,
Since c is unknown we'll break the CDF into 2 parts at x=0;
Thus if x > 0 (for x < 0 ; $F_X(x) = 0$):

$$
\begin{aligned}
F_X(x) =& \int_{-\infty}^{x} f_X(t)dt \\
=& \int_{-\infty}^{0} f_X(t)dt + \int_{0}^{x} f_X(t)dt \\
=& 0 + \int_{0}^{x} ce^{-2t}dt \\
=& -0.5c[e^{-2t}]_0^x \\
=& 0.5c(1 - e^{-2x})
\end{aligned}
$$

To find c we apply the property:

$$
\begin{aligned}
lim_{(a \to \infty)} F_X(a) = 1 \\
lim_{(a \to \infty)} 0.5c(1 - e^{-2a}) = 1 \\
0.5c \times (1 - 0) = 1 \\
c = 2
\end{aligned}
$$

Also we need to find $P(X > 2)$:

$$
\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) \\
&= 1 - F_X(2) \\
&= 1 - (1 - e^{-2x})_{x=2} \\
&= 1 - (1 - e^{-4}) \\
&= e^{-4}
\end{aligned}
$$

Q9

A coin with $p_{heads} = 0.7$ is tossed 3 times. To find pmf of the number of heads denoted by H_c:
Thus possible number of heads $= 0, 1, 2, 3$
Considering case wise; where (X_1, X_2, X_3) denote the tosses:

Case- 0 Heads:
Only 1 event in which we can have 0 heads

$$\begin{aligned} P(H_c = 0) =& P(T, T, T) \\ =& (1 - p_{heads})^3 \\ =& (0.3)^3 \\ =& 0.027 \end{aligned}$$

Case- 1 Heads:
There are $\binom{3}{1}$ events in which we can have 1 heads

$$\begin{aligned} P(H_c = 2) =& \binom{3}{1} \times P(H, T, T) \\ =& 3 \times p_{heads} \times (1 - p_{heads})^2 \\ =& 3 \times (0.7) \times (0.3)^2 \\ =& 0.189 \end{aligned}$$

Case- 2 Heads:
There are $\binom{3}{2}$ events in which we can have 2 heads

$$\begin{aligned} P(H_c = 2) =& \binom{3}{2} \times P(H, H, T) \\ =& 3 \times p_{heads}^2 \times (1 - p_{heads}) \\ =& 3 \times (0.7)^2 \times (0.3) \\ =& 0.441 \end{aligned}$$

Case- 3 Heads:
Only 1 event in which we can have 3 heads

$$\begin{aligned} P(H_c = 3) =& P(H, H, H) \\ =& p_{heads}^3 \\ =& (0.7)^3 \\ =& 0.343 \end{aligned}$$

Thus the pmf of H_c is

$$p_{H_c}(x) = \begin{cases} 0.027, & \text{if } x = 0 \\ 0.189, & \text{if } x = 1 \\ 0.441, & \text{if } x = 2 \\ 0.343, & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Q10

Given probability density function:

$$f_X(x) = \begin{cases} 2x, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Cumulative probability function for $x \in [0, 1]$:

$$\begin{aligned} F_X(x) =& Pr\{X \leq x\} \\ =& \int_0^x f_X(t)dt \\ =& \int_0^x 2tdt \\ =& x^2 \end{aligned}$$

$$\begin{aligned} P(X \geq 0.4 | X \leq 0.8) =& \frac{P(X \geq 0.4 \text{ and } X \leq 0.8)}{P(X \leq 0.8)} \\ =& \frac{F_X(0.8) - F_X(0.4)}{F_X(0.8)} \\ =& \frac{3}{4} \end{aligned}$$

Q11

A continuous random variable X is said to have an exponential distribution with parameter $\lambda > 0$, shown as $X \sim Exponential(\lambda)$, if its PDF is given by

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

Lets compute the below probability for $a \geq 0$:

$$\begin{aligned} P(X > a) =& \int_a^\infty f_X(x)dx \\ =& \int_a^\infty \lambda e^{-\lambda x}dx \\ =& e^{-\lambda a} \end{aligned}$$

Now:

$$\begin{aligned} P(X > a+b | X > a) =& \frac{P(X > a+b \text{ and } X > a)}{P(X > a)} \\ =& \frac{P(X > a+b)}{P(X > a)} \\ =& \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\ =& e^{-\lambda b} \end{aligned}$$

The exponential distribution is **memoryless** because the past has no bearing on its future behavior.

Q12

I_E equals 1, if event E occurs. Event E is the event in which all the five fair coins land heads. The Probability of that happening is:

$$\begin{aligned} P\{I_E = 1\} &= P(\text{ head in a fair coin toss })^5 \\ &= \frac{1}{2^5} \end{aligned}$$

Q13

We know that CDF $F_X(x)$ is defined as $P(X \leq x)$. Hence $P(X = x)$ is consequently $F_X(x) - F_X(x-1)$.
$F_X(x) - F_X(x-1)$ is 0 everywhere except at $0, 1$ where it takes the values $0.5, 0.5$ respectively. Hence

$$P(X = x) = \begin{cases} 0.5 & \text{if } x \in \{0, 1\} \\ 0 & \text{otherwise} \end{cases}$$

Q14

Let the color of the ball drawn at time-step t be given by random variable (RV) X_t. The probability distribution of all X_t will be identical to (say) X. The RV X is 0 if the ball drawn at time t is black and is 1 if the ball is white.

$$P(X = 0) = P(X = 1) = 3/6 = 0.5$$

$$P(X_1 = i_1, X_2 = i_2, X_3 = i_3, X_4 = i_4) = 0.5^4 = 0.0625$$

This shows that any sequence of drawing 4 balls is equally likely to 0.0625. There are $\binom{4}{2}$ required sequences of interest hence,

$$\begin{aligned} \sum_{\sum_{j=1}^{4} i_j = 2} P(X_1 = i_1, X_2 = i_2, X_3 = i_3, X_4 = i_4) =& \binom{4}{2} * 0.625 \\ =& 0.375 \end{aligned}$$

Q15

Note that the last toss of the n flips has to be a head. Let X_i denote that i flip. The probability of the flips X_i will be identical to (say) X. Let 0 denote tails and 1 heads. Then the probability of the sequence $(i_1, ..., i_n)$ where $i_k \in \{0, 1\}$

under the constraint $\sum_{j=1}^{n} i_j = r$ is

$$P(X_1 = i_1, ..., X_n = i_n) = p^r(1-p)^{n-r}$$

All sequence of desired characteristics i.e. $\sum_{j=1}^{n} i_j = r$ and $i_n = 1$ are equally likely given by equation above. The number of such sequences by combinatorics is $\binom{n-1}{r-1}$ (i_n is fixed and i_1,...,i_{n-1} need to choose r-1). Hence the required probability is

$$\sum_{\sum_{j=1}^{n} i_j = r, i_n = 1} P(X_1 = i_1, ..., X_n = i_n) = \binom{n-1}{r-1} * p^r(1-p)^{n-r}$$

Q16

Let X denote the number of errors in the page. Then $P(X > 0) = 1 - P(X = 0)$.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Substituting λ as 1 and k as 0 we get $P(X = 0) = e^{-1}$. Hence the answer is $1 - e^{-1}$