Maxwell Hanks – n12405710

MXB101 Problem Solving Task 2 (10%)

Question 1a.

How many unique ways can this webpage be produced in the system?

Solution:

There are 12 shirts and 8 trousers, hence, there are 12×8 total combinations of shirts and trousers, therefore there is 96 possible distinct outfits.

The store developed a tool that randomly produces a webpage with 5 complete outfits, which is an **ordered** sequence with distinct outfits (order matters with no repeats). Hence, a permutation will be used:

$$
\begin{aligned}
{}_nP_r &= \frac{n!}{(n-r)!} \\
{}_{96}P_5 &= \frac{96!}{91!} \\
&= 7{,}334{,}887{,}680
\end{aligned}
$$

Therefore, there are 7,334,887,680 unique ways the webpage can be displayed.

Question 1b.

What is the probability that the two outfits that the system marks as suggested are entirely brown?

Solution:

There are 8 brown shirts, and 3 brown pants, inferring there 8×3; 24 total combinations of brown outfits. There are two outfits that can be marked as suggested, **without replacement.** Hence,

$$
\begin{aligned}
\Pr(\mathit{Both\ brown}) &= \frac{24}{96} \times \frac{24-1}{96-1} \\
&= \frac{552}{9120} \\
&= 6.0526\%
\end{aligned}
$$

Therefore, the probability that the two suggested outfits are entirely brown is 6%.

Question 2a.

The total number of such errors present in a message is given by the discrete random variable X with probability mass function $\Pr(X = x) = p(x), x \in \{0,1, \dots ,8\}$ where $p(x)$ is given by

$$p(x) = \begin{cases} \frac{1}{3}, \text{if } x = 0 \\ \frac{k}{x(x+1)}, \text{if } x \in \{1,2, \dots ,8\} \end{cases}$$

Find k such that $p(x)$ is a valid probability mass function.

Solution:

For $p(x)$ to be a valid probability mass function (pmf), the sum of the probabilities must be equal to 1:

$$\sum_{x=0}^{8} p(x) = 1$$

The probability of $p(0)$ is $\frac{1}{3}$, and the probability of $p(x) = \frac{k}{x(x+1)}, x \in \{1,2, \dots ,8\}, k \in \mathbb{R}$:

$$\begin{aligned} \sum_{x=0}^{8} p(x) = \frac{1}{3} + \sum_{x=1}^{8} p(x) &= 1 \\ \frac{1}{3} + \sum_{x=1}^{8} \frac{k}{x(x+1)} &= 1 \\ \frac{1}{3} + k\sum_{x=1}^{8} \frac{1}{x(x+1)} &= 1 \end{aligned}$$

By partial fractions decomposition:

$$\begin{aligned} \frac{1}{3} + k\sum_{x=1}^{8} \left(\frac{1}{x} - \frac{1}{x+1}\right) &= 1 \\ \frac{1}{3} + k\left[\left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \dots + \left(\frac{1}{8} - \frac{1}{9}\right)\right] &= 1 \end{aligned}$$

The series cancels out the middle terms:

$$\begin{aligned} \frac{1}{3} + k\left[1 - \frac{1}{9}\right] &= 1 \\ \frac{1}{3} + \frac{8}{9}k &= 1 \\ \frac{8}{9}k &= \frac{2}{3} \\ \therefore k &= \frac{3}{4} \end{aligned}$$

Therefore, $k = \frac{3}{4}$ such that $p(x)$ is a valid probability mass function.

Question 2b.

Calculate the standard deviation of X.

Solution:

Let $p(x)$ be defined with the following probability mass function such that $k = \frac{3}{4}$

$$p(x) = \begin{cases} \frac{1}{3}, \text{if } x = 0 \\ \frac{3}{4} \cdot \frac{1}{x(x+1)}, \text{if } x \in \{1,2, \dots ,8\} \end{cases}$$

To calculate standard deviation, firstly, the expected value must first be established:

$$\mathbb{E}[X] = \sum_{x} x \cdot \Pr(X = x)$$

Whereby:

$$\mathbb{E}[X] = \sum_{x=0}^{8} x \cdot p(x)$$

For $x = 0, p(x) = \frac{1}{3}, x \cdot p(x) = 0$

$$\begin{aligned} \therefore \mathbb{E}[X] &= \sum_{x=1}^{8} x \cdot p(x) \\ &= \sum_{x=1}^{8} x \cdot \frac{3}{4} \cdot \frac{1}{x(x+1)} \\ &= \frac{3}{4} \sum_{x=1}^{8} \frac{1}{x+1} \\ &= \frac{3}{4} \left(\frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{9}\right) \\ &\approx 1.3717 \end{aligned}$$

Therefore, the mean is approximately 1.3717. Variance of a discrete probability function is denoted:

$$\begin{aligned} Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \sum_{x=0}^{8} x^2 \cdot p(x) - (1.3717)^2 \\ &= \frac{3}{4} \sum_{x=1}^{8} \frac{x}{x+1} - (1.3717)^2 \\ &= 4.6283 - (1.3717)^2 \\ \therefore Var[X] &\approx 2.7467 \end{aligned}$$

Standard deviation is denoted:

$$\begin{aligned} \sigma = \sqrt{Var[X]} &= \sqrt{2.7467} \\ &\approx 1.6573 \end{aligned}$$

Therefore, the standard deviation of X is approximately 1.6573.

Question 2c.

The *entropy* of a random variable represents the average level of uncertainty in the variable. For a discrete random variable, entropy is given by:

$$\mathbb{E}[-\ln(p(X))] = -\sum_{x\in\Omega} p(x)\ln p(x).$$

Calculate the entropy of X.

Solution:

Given the definition of entropy of a discrete random variable,

$$\mathbb{E}[-\ln(p(X))] = -\sum_{x\in\Omega} p(x)\ln p(x)$$

The entropy of X can be calculated, let us denote $\mathbb{E}[-\ln(p(X))]$ as $\mathbb{H}[X]$.

$$\begin{aligned}
\mathbb{H}[X] &= -\sum_{x\in\Omega} p(x)\ln p(x) \\
&= -\left[p(0)\ln p(0) + \sum_{x=1}^{8} p(x)\ln p(x)\right] \\
&= -\left[\frac{1}{3}\ln\left(\frac{1}{3}\right) + \sum_{x=1}^{8}\frac{3}{4x(x+1)}\ln\left(\frac{3}{4x(x+1)}\right)\right]
\end{aligned}$$

Using Desmos, the solution can be found as $\mathbb{H}[X] \approx 1.55977$.

$$-\left[\frac{1}{3}\ln\left(\frac{1}{3}\right) + \sum_{x=1}^{8}\left(\frac{3}{4x(x+1)}\right)\ln\left(\frac{3}{4x(x+1)}\right)\right]$$

$$= 1.55977249718$$

Question 3a.

Consider the continuous random variable, X, defined by the quantile function:

$$Q(p) = \sqrt{-\ln(1-p)}$$

Find the cumulative distribution function (CDF) of $X, F(x) = \Pr(X \leq x)$, and the range of possible values X can take.

Solution:

Given the quantile function $Q(p) = \sqrt{-\ln(1-p)}$, where $p \in [0{,}1]$ yields to a value x such that $\Pr(X \leq x) = p$. To find the CDF of X, $F(x)$, the function needs to be inverted where $x = F^{-1}(p) = Q(p)$.

$$\begin{aligned} x &= \sqrt{-\ln(1-p)} \\ x^2 &= -\ln(1-p) \\ -x^2 &= \ln(1-p) \\ e^{-x^2} &= 1-p \\ p &= 1-e^{-x^2} \end{aligned}$$

Therefore the CDF is:

$$F(x) = \Pr(X \leq x) = \begin{cases} 0, & \text{if } x < 0 \\ 1-e^{-x^2}, & \text{if } x \geq 0 \end{cases}$$

To find the range of X, the limits of the function $Q(p)$ can be evaluated at the bounds where $p \in [0{,}1]$. For lower bound, $\lim_{p \to 0} Q(p) = \lim_{p \to 0} \sqrt{-\ln(1-p)} = 0$. For the upper bound, $\lim_{p \to 1} Q(p) = \lim_{p \to 1} \sqrt{-\ln(1-p)} = \infty$.

Hence, the range of possible values X can take is $[0, \infty)$ or $X \geq 0$.

Question 3b.

Find $\mathbb{E}[X]$. You must validate the assumptions behind any rules you apply. *Note that you may find it useful to know that*

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

Solution:

Since X is a continuous function, the expected value can be calculated as

$$\mathbb{E}[X] = \int_x x f(x) dx$$

By applying integration by parts, another result can be obtained for a continuous function X:

$$\mathbb{E}[X] = -\int_{x<0} F(x)\,dx + \int_{x>0} (1 - F(x))\,dx$$

Since X is a non-negative function, the $-\int_{x<0} F(x)\,dx$ integral cancels out, hence:

$$\mathbb{E}[X] = \int_{x>0} \big(1 - F(x)\big)\,dx$$

Substituting the limits and CDF:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{\infty} \big(1 - F(x)\big) dx \\ &= \int_0^{\infty} \left(1 - \left(1 - e^{-x^2}\right)\right) dx \\ &= \int_0^{\infty} e^{-x^2}\,dx\end{aligned}$$

By use of the Gaussian Integral:

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

The Gaussian Integral is symmetric about $x = 0$, and integrating on a real domain from $-\infty$ to ∞ is $\sqrt{\pi}$, therefore, since X is non-negative hence excluding $x > 0$, integrating from 0 to ∞ is half of the Gaussian Integral, hence:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{\infty} e^{-x^2}\,dx \\ &= \frac{1}{2}\int_{-\infty}^{\infty} e^{-x^2}\,dx \\ &= \frac{\sqrt{\pi}}{2}\end{aligned}$$

Therefore the expected value $\mathbb{E}[X] = \frac{\sqrt{\pi}}{2}$.

Question 3c.

Calculate the interquartile range of X, that is, the difference between the upper and lower quartiles.

Solution:

Given interquartile range (IQR) is given by $Q_3 - Q_1$, since Q is denoted the quantile function, therefore IQR is $Q(0.75) - Q(0.25)$, recall $Q(p) = \sqrt{-\ln(1-p)}$:

- $Q(0.75) = \sqrt{-\ln(1-0.75)} = \sqrt{-\ln(0.25)} = \sqrt{\ln 4}$
- $Q(0.25) = \sqrt{-\ln(1-0.25)} = \sqrt{-\ln(0.75)}$

Hence, IQR is denoted:

$$\begin{aligned} \text{IQR} &= Q(0.75) - Q(0.25) \\ &= \sqrt{\ln 4} - \sqrt{-\ln(0.75)} \\ &\approx 0.64015 \end{aligned}$$

Therefore, the IQR of X is approximately 0.64.