Problem 1

(a) First, we compute $P(A)$, $P(B)$, $P(A \cap B)$

$$\begin{aligned} P(A) &= 1-p \\ P(B) &= p\varepsilon_0 + (1-p)\varepsilon_1 \\ P(A \cap B) &= (1-p)\varepsilon_1 \end{aligned}$$

Next, we check when $P(A \cap B) = P(A) \cdot P(B)$

$$\begin{aligned} &(1-p)\varepsilon_1 = (1-p)p\varepsilon_0 + (1-p)^2\varepsilon_1 \\ \Rightarrow &\, p(1-p)(\varepsilon_0 - \varepsilon_1) = 0 \end{aligned}$$

Therefore, A and B are independent if $p = 0$, $p = 1$, or $\varepsilon_0 = \varepsilon_1$.

(b) First, we compute $P(A|C)$, $P(B|C)$, $P(A \cap B|C)$

$$\begin{aligned} P(A|C) &= \frac{P(A \cap C)}{P(C)} = \frac{(1-p)(\alpha_1 + \varepsilon_1)}{p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)} \\ P(B|C) &= \frac{P(B \cap C)}{P(C)} = \frac{p\varepsilon_0 + (1-p)\varepsilon_1}{p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)} \\ P(A \cap B|C) &= \frac{P((A \cap B) \cap C)}{P(C)} = \frac{(1-p)\varepsilon_1}{p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)} \end{aligned}$$

Next, we check if $P(A \cap B|C) = P(A|C) \cdot P(B|C)$

$$\begin{aligned} P(A \cap B|C) &= \frac{(1-p)\varepsilon_1}{p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)} \\ P(A|C) \cdot P(B|C) &= \frac{(1-p)(\alpha_1 + \varepsilon_1)}{p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)} \cdot \frac{p\varepsilon_0 + (1-p)\varepsilon_1}{p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)} \end{aligned}$$

We can simplify the check to

$$\begin{aligned} &P((A \cap B) \cap C) \cdot P(C) \\ =&(1-p)\varepsilon_1 \cdot [p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)] \\ =&(1-p)[p\varepsilon_1\alpha_0 + p\varepsilon_0\varepsilon_1 + (1-p)\varepsilon_1\alpha_1 + (1-p)\varepsilon_1^2] \\ &P(A \cap C) \cdot P(B \cap C) \\ =&(1-p)[p\varepsilon_0\alpha_1 + p\varepsilon_0\varepsilon_1 + (1-p)\varepsilon_1\alpha_1 + (1-p)\varepsilon_1^2] \\ =&(1-p)(\alpha_1 + \varepsilon_1)[p\varepsilon_0 + (1-p)\varepsilon_1] \end{aligned}$$

Therefore, events A and B are conditionally independent given the event C if $p = 0$, $p = 1$, or $\varepsilon_0\alpha_1 = \varepsilon_1\alpha_0$.

(c) By total probability theorem, we have

$$P(\text{Bit delivered correctly}) = p(a - \varepsilon_0 - \alpha_0) + (1-p)(1 - \varepsilon_1 - \alpha_1)$$

Problem 2

(a) Let X be the number of `"1"`s transmitted in the time interval. Given $Z = n$, X follows a binomial distribution

$$X|(Z = n) \sim \text{Binomial}(n, p)$$

Then, the probability that $X = k$ is

$$\begin{aligned}
P(X = k) &= \sum_{n=k}^{\infty} P(X = k|Z = n)P(Z = n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \frac{e^{-\lambda}p^k\lambda^k}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}\lambda^{n-k}}{(n-k)!} \\
&= \frac{e^{-\lambda}p^k\lambda^k}{k!} \sum_{n=0}^{\infty} \frac{(1-p)^{n}\lambda^{n}}{n!} \\
&= \frac{e^{-\lambda}p^k\lambda^k}{k!} e^{\lambda(1-p)} \\
&= \frac{e^{-\lambda p}(\lambda p)^k}{k!}
\end{aligned}$$

Therefore, $X \sim$ Poisson(λp).

(b) (i) This event occurs when all X_i are less than or equal to k.
CDF of X is $F_X(k) = \prod_{i=1}^{n} P(X_i \le k) = \prod_{i=1}^{n} 1 - P(X_i > k) = [1 - (1-p)^k]^n$
PMF of X is $P(X = k) = F_X(k) - F_X(k-1) = [1 - (1-p)^k]^n - [1 - (1-p)^{k-1}]^n$

(ii) This event occurs when at least one X_i is less than or equal to k.
CDF of Y is $F_Y(k) = 1 - P(\text{min}(X_1, \ldots, X_n) > k) = 1 - [(1-p)^k]^n = 1 - (1-p)^{nk}$
PMF of Y is $P(Y = k) = F_Y(k) - F_Y(k-1) = (1-p)^{n(k-1)} - (1-p)^{nk}$

(iii) $P(Y = k) = (1-p)^{n(k-1)} - (1-p)^{nk} = (1-p)^{n(k-1)}[1 - (1-p)^n]$
Let $p' = 1 - (1-p)^n$, then $P(Y = k) = (1-p')^{(k-1)}p'$, this is the PMF of a Geometric random variable with success probability p'.

(c) According to the Binomial distribution, we have

$$P(X_{S_2} = k) = \binom{123}{k} p_T^k (1 - p_T)^{123-k} \text{ for } k = 0, 1, \cdots, 123$$

Problem 3

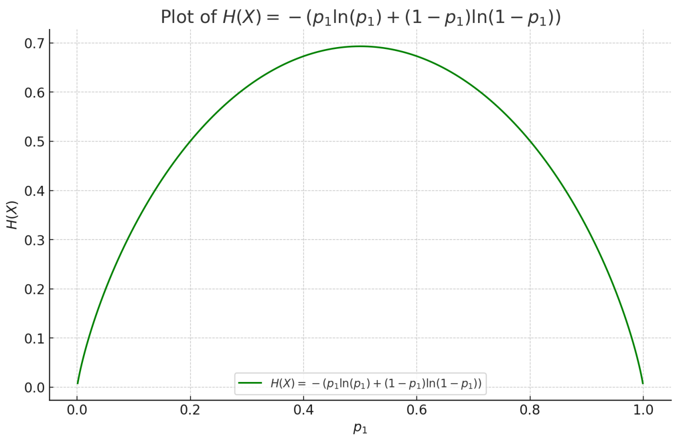
(a)

$$\begin{aligned}
H(\frac{1}{2},\frac{1}{3},\frac{1}{6}) &= -(\frac{1}{2}\ln\frac{1}{2}+\frac{1}{3}\ln\frac{1}{3}+\frac{1}{6}\ln\frac{1}{6}) \\
&= (\frac{1}{2}+\frac{1}{6})\ln 2 + (\frac{1}{3}+\frac{1}{6}\ln 3) \\
&= \frac{2}{3}\ln 2 + \frac{1}{2}\ln 3 \\
H(\frac{1}{2},\frac{1}{2}) + \frac{1}{2}\cdot H(\frac{2}{3},\frac{1}{3}) &= -(\frac{1}{2}\ln\frac{1}{2}+\frac{1}{2}\ln\frac{1}{2}) - \frac{1}{2}\cdot(\frac{2}{3}\ln\frac{2}{3}+\frac{1}{3}\ln\frac{1}{3}) \\
&= \ln 2 + \frac{1}{2}(\frac{2}{3}(\ln 3 - \ln 2) + \frac{1}{3}\ln 3) \\
&= \frac{2}{3}\ln 2 + \frac{1}{2}\ln 3
\end{aligned}$$

Therefore, $H(\frac{1}{2},\frac{1}{3},\frac{1}{6}) = H(\frac{1}{2},\frac{1}{2}) + \frac{1}{2}\cdot H(\frac{2}{3},\frac{1}{3})$.

(b) Maximum entropy: $H_{max} = \ln 2$ at $p_1 = 0.5$

Minimum entropy: $H_{min} = 0$ at $p_1 = 1$ or $p_1 = 0$



(c) Let $w_i = p_i$ and $x_i = \frac{1}{p_i}$.

$$\frac{p_1\frac{1}{p_1}+\cdots+p_n\frac{1}{p_n}}{1} = n \geq \prod_{i=1}^{n}(\frac{1}{p_i})^{p_i}$$

$$\Rightarrow \ln n \geq \ln(\prod_{i=1}^{n}(\frac{1}{p_i})^{p_i}) = -\sum_{i=1}^{n} p_i \ln p_i = H(X)$$

The equality holds when $p_1 = \cdots = p_n = \frac{1}{n}$. Therefore, $H(X)_{max} = \ln n$ when $\{p_i = \frac{1}{n}\}_{i=1}^{n}$.

(d) The entropy is minimized when the random variable X is deterministic. For a specific $i \in \{1, 2, \ldots, n\}$

$$H(X) = -(1\cdot\ln 1 + \sum_{j\neq i}^{n} 0\cdot\ln 0) = 0$$

Therefore, $H(X)_{\text{min}} = 0$ is achieved by any PMF where exactly one $i \in \{1, 2, \ldots, n\}$, $p_i = 1$ and $p_j = 0$ for all $j \neq i$.

Problem 4

(a) (i)

$$E[X] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p = p \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

(ii)

$$\begin{aligned} E[e^{tX}] &= \sum_{k=1}^{\infty} e^{tk} \cdot (1-p)^{k-1} p \\ &= p \sum_{k=1}^{\infty} e^{tk} \cdot (1-p)^{k-1} \\ &= pe^t \sum_{k=1}^{\infty} [e^t(1-p)]^{k-1} \\ &= pe^t \cdot \frac{1}{1-(1-p)e^t} = \frac{pe^t}{1-(1-p)e^t} \end{aligned}$$

Additionally, we should verify the convergence condition

$$|e^t(1-p)| < 1 \Rightarrow \; e^t < \frac{1}{1-p} \; \Rightarrow \; t < \ln \frac{1}{1-p} = -\ln(1-p)$$

(b) (i) To find $Var[Z] = E[Z^2] - (E[Z])^2$, we first compute $E[Z]$ and $E[Z^2]$

$$E[Z^2] = \sum_{n=1}^{\infty} z_n^2 \cdot p_Z(z_n) = \sum_{n=1}^{\infty} n \cdot \frac{6}{(\pi n)^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Therefore, since $E[Z^2] = \infty$, the variance $Var[Z]$ does not exist.

(ii)

$$\sum_{n=1}^{\infty} z_n^3 \cdot p_Z(z_n) = \sum_{n=1}^{\infty} (-1)^n (\sqrt{n})^3 \cdot \frac{6}{(\pi n)^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} (-1)^n \frac{1}{\sqrt{n}} = -\frac{6}{\pi^2} \eta(\frac{1}{2}) \approx -\frac{6}{\pi^2} \cdot 0.6049$$

(iii)

$$E[|Z^3|] = \sum_{n=1}^{\infty} |z_n^3| \cdot p_Z(z_n) = \sum_{n=1}^{\infty} n^{\frac{3}{2}} \cdot \frac{6}{(\pi n)^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty$$

Since $E[|Z^3|]$ is infinite, $E[Z^3]$ does not exist.

Problem 5

(a) Partition: `[0.7, 0.3]` / Prior: `Uniform`
  - Test accuracy: 98.21%
  - Observations:
    - The classifier performs exceptionally well for the ham class, with a precision and recall close to 100%.
    - The spam detection is slightly weaker, with a recall of 89%, indicating that some spam messages were missed.

(b)
  - Partition: `[0.7, 0.3]` / Prior `[0.5, 0.5]`
    - Test accuracy: 97.67%
    - Observations:
      * This prior slightly lowered the overall accuracy compared to the uniform prior.
      * The spam detection was more balanced, achieving equal precision and recall, which implies the model's confidence in detecting both classes became more stable.
  - Partition: `[0.7, 0.3]` / Prior `[0.9, 0.1]`
    - Test accuracy: 98.27%
    - Observations:
      * Setting a heavier prior towards ham improved the model's confidence in predicting spam but didn't significantly change the recall.
      * The accuracy remains similar to the uniform prior but with slightly better precision for spam.
  - Partition: `[0.8, 0.2]` / Prior: `Uniform`
    - Test accuracy: 98.39%
    - Observations:
      * Increasing the training data led to a slight improvement in accuracy.
      * Spam detection precision increased to 99%, showing that more training data helps the model learn the nuances better.
  - Partition: `[0.6, 0.4]` / Prior: `Uniform`
    - Test accuracy: 98.34%
    - Observations:
      * A larger test set still resulted in high accuracy.
      * Spam detection improved slightly, demonstrating the model's robustness even with less training data.