Midterm Exam #1

Ruixin Guo

October 8, 2021

1. Given that $p_j = P(W_j|C_1), r_j = P(W_j|C_2), P(C_1) = 0.7, P(C_2) = 0.3$. Let W_j^c be the event that an email does not contain the jth word.

$$\begin{aligned}
P(W_j^c|C_1) &= 1 - P(W_j|C_1) = 1 - p_j \\
P(W_j^c|C_2) &= 1 - P(W_j|C_2) = 1 - r_j
\end{aligned}$$

The probability that a new email that contains the first 3 words is spam can be expressed as the conditional probability $P(C_1|W_1, W_2, W_3, W_4^c, ..., W_{50}^c)$. Knowing that $W_1, W_2, W_3, ..., W_{50}$ are conditionally independent. By the Bayes' rule,

$$\begin{aligned}
P(C_1|W_1, W_2, W_3, W_4^c, ..., W_{50}^c) &= \frac{P(W_1, W_2, W_3, W_4^c, ..., W_{50}^c|C_1)P(C_1)}{P(W_1, W_2, W_3, W_4^c, ..., W_{50}^c)} \\
&= \frac{P(W_1|C_1)P(W_2|C_1)P(W_3|C_1)P(W_4^c|C_1)...P(W_{50}^c|C_1)P(C_1)}{P(W_1, W_2, W_3, W_4^c, ..., W_{50}^c|C_1)P(C_1) + P(W_1, W_2, W_3, W_4^c, ..., W_{50}^c|C_2)P(C_2)} \\
&= \frac{0.7p_1p_2p_3\prod_{i=4}^{50}(1-p_i)}{0.7p_1p_2p_3\prod_{i=4}^{50}(1-p_i) + 0.3r_1r_2r_3\prod_{i=4}^{50}(1-r_i)}
\end{aligned}$$

2.

a. Since $E(X_i) = \mu$,

$$\begin{aligned}
E(\bar{X}_n) &= E(\frac{1}{n}(X_1 + X_2 + ... + X_n)) \\
&= \frac{1}{n}(E(X_1) + E(X_2) + ... + E(X_n)) \\
&= \frac{1}{n}n\mu \\
&= \mu
\end{aligned}$$

b. Since $X_1, X_2, ..., X_n$ are independent and identically distributed, $Cov(X_i, X_j) = 0$ for any $\{i \neq j | 1 \leq$

$i, j \le n\}$. Given that $Var(X_i) = \sigma^2$,

$$\begin{aligned}
Var(\bar{X}_n) &= Var(\frac{1}{n}(X_1 + X_2 + ... + X_n)) \\
&= \frac{1}{n^2}(Var(X_1) + Var(X_2) + ... + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)) \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

3.

a. Since both X and Y are exponentially distributed with mean $1/5$, the PDF of X and Y is $f(x) = 5e^{-5x}$ where $x > 0$.

$$\begin{aligned}
P(X \ge 5Y) &= \int_0^\infty P(Y \le \frac{X}{5} | X = x) P(X = x) dx \\
&= \int_0^\infty (\int_0^{x/5} 5e^{-5y} dy) 5e^{-5x} dx \\
&= \int_0^\infty (1 - e^{-x}) 5e^{-5x} dx \\
&= [-e^{-5x} + \frac{5}{6} e^{-6x}]_0^\infty \\
&= \frac{1}{6}
\end{aligned}$$

b. Using the survival function. Let $Z = \max\{X, Y\}$, the CDF of Z is

$$\begin{aligned}
F_Z(z) &= P(Z \le z) \\
&= P(\max\{X, Y\} \le z) \\
&= P(X \le z, Y \le z) \\
&= P(X \le z) P(Y \le z) \\
&= (1 - e^{-5z})(1 - e^{-5z})
\end{aligned}$$

Thus the PDF of Z is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = 10e^{-5z} - 10e^{-10z}$$

c. The CDF of Q is

$$F_Q(q) = P(Q < q) = P(\sqrt{X} < q) = P(X < q^2) = \int_0^{q^2} 5e^{-5x} dx = 1 - e^{-5q^2}$$

Thus the PDF of Q is

$$f_Q(q) = \frac{d}{dq} F_Q(q) = 10qe^{-5q^2}$$

4.

a. Since $X_i|P_i \sim \text{Bern}(P_i)$, the PMF of $X_i|P_i$ is

$$P(X_i = x|P_i) = P_i^x(1-P_i)^{1-x}$$

where $x = 0, 1$.

Since $P_i \sim \text{Beta}(a, b)$, by the law of total probability,

$$\begin{aligned} P(X_i = x) &= \int_0^1 P(X_i = x|P_i)P(P_i)dP_i \\ &= \int_0^1 P_i^x(1-P_i)^{1-x}\frac{1}{\beta(a,b)}P_i^{a-1}(1-P_i)^{b-1}dP_i \\ &= \frac{\beta(a+x, b-x+1)}{\beta(a,b)} \end{aligned}$$

When $x = 0$, $P(X_i) = \frac{b}{a+b}$; when $x = 1$, $P(X_i) = \frac{a}{a+b}$. Thus $X_i \sim \text{Bern}(\frac{a}{a+b})$. $E(X_i) = \frac{a}{a+b}$, $Var(X_i) = \frac{ab}{(a+b)^2}$.

Since $Y = \sum_{i=1}^n X_i$, we have

$$E(Y) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \frac{na}{a+b}$$

$$Var(Y) = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j) = \sum_{i=1}^n Var(X_i) = \frac{nab}{(a+b)^2}$$

The third equality in inducing $Var(Y)$ is because $X_1, X_2, ..., X_n$ are independent. Thus for any $1 \le i, j \le n$, $Cov(X_i, X_j) = 0$.

b. Since $X_1, X_2, ..., X_n$ are i.i.d $\text{Bern}(\frac{a}{a+b})$. By the definition of binomial distribution, $Y \sim \text{Bin}(n, \frac{a}{a+b})$.

c. By the law of total expectation,

$$\begin{aligned} E(Y) &= E(\sum_{i=1}^k X_i) \\ &= E(\sum_{i=1}^k E(X_i|P_i)) \\ &= E(\sum_{i=1}^k n_iP_i) \quad [X_i|P_i \sim \text{Bin}(n_i, P_i), E(X_i|P_i) = n_iP_i] \\ &= \sum_{i=1}^k n_iE(P_i) \\ &= \frac{a}{a+b}\sum_{i=1}^k n_i \end{aligned}$$

By the law of total variance,

$$\begin{aligned}
Var(Y) &= Var(\sum_{i=1}^{k} X_i) \\
&= E(Var((\sum_{i=1}^{k} X_i)|P_i)) + Var(E((\sum_{i=1}^{k} X_i)|P_i)) \\
&= E(\sum_{i=1}^{k} Var(X_i|P_i)) + Var(\sum_{i=1}^{k} E(X_i|P_i)) \\
&= E(\sum_{i=1}^{k} n_i P_i(1-P_i)) + Var(\sum_{i=1}^{k} n_i P_i) \\
&= \sum_{i=1}^{k} n_i E(P_i(1-P_i)) + \sum_{i=1}^{k} n_i^2 Var(P_i) \\
&= \sum_{i=1}^{k} n_i \frac{ab}{(a+b)(a+b+1)} + \sum_{i=1}^{k} n_i^2 \frac{ab}{(a+b)^2(a+b+1)} \\
&= \sum_{i=1}^{k} n_i \frac{ab(a+b+n_i)}{(a+b)^2(a+b+1)}
\end{aligned}$$

The 6th equality in inducing $Var(Y)$ is because $E(P_i(1-P_i)) = \int_0^1 P_i(1-P_i)\frac{1}{\beta(a,b)}P_i^{a-1}(1-P_i)^{b-1} = \frac{\beta(a+1,b+1)}{\beta(a,b)} = \frac{ab}{(a+b)(a+b+1)}$ and $Var(P_i) = E(P_i^2) - (E(P_i))^2 = \frac{\beta(a+2,b)}{\beta(a,b)} - (\frac{\beta(a+1,b)}{\beta(a,b)})^2 = \frac{ab}{(a+b)^2(a+b+1)}$.

5.

a.

$$\begin{aligned}
E(\hat{\theta}-\theta)^2 &= E(\hat{\theta}^2) - 2E(\hat{\theta}\theta) + E(\theta^2) \\
&= E(\hat{\theta}^2) - 2E(E(\hat{\theta}\theta|\theta)) + E(\theta^2) \quad \text{[Law of total expectation]} \\
&= E(\hat{\theta}^2) - 2E(\theta E(\hat{\theta}|\theta)) + E(\theta^2) \quad [E(h(X)Y|X) = h(X)E(Y|X)] \\
&= E(\hat{\theta}^2) - 2E(\theta^2) + E(\theta^2) \quad [\text{Given that } E(\hat{\theta}|\theta) = \theta] \\
&= E(\hat{\theta}^2) - E(\theta^2) \qquad (1)
\end{aligned}$$

b.

$$\begin{aligned}
E(\hat{\theta}-\theta)^2 &= E(\hat{\theta}^2) - 2E(\hat{\theta}\theta) + E(\theta^2) \\
&= E(\hat{\theta}^2) - 2E(E(\hat{\theta}\theta|X)) + E(\theta^2) \quad \text{[Law of total expectation]} \\
&= E(\hat{\theta}^2) - 2E(\hat{\theta}E(\theta|X)) + E(\theta^2) \quad [E(h(X)Y|X) = h(X)E(Y|X)] \\
&= E(\hat{\theta}^2) - 2E(\hat{\theta}^2) + E(\theta^2) \quad [\text{Given that } E(\theta|X) = \hat{\theta}] \\
&= E(\theta^2) - E(\hat{\theta}^2) \qquad (2)
\end{aligned}$$

c. By combining Equation (1) and Equation (2) we have $E(\hat{\theta}^2) - E(\theta^2) = -(E(\theta^2) - E(\hat{\theta}^2))$, which shows that Equation (1) and Equation (2) will not be equal unless $E(\hat{\theta}^2) = E(\theta^2)$. Thus it is impossible for $\hat{\theta}$ to be both the Bayes Procedure and unbiased unless $\hat{\theta}$ perfectly matches θ, which can only be achieved by perfectly observing X.