IE 7374 ST: Machine Learning in Engineering
HW-1

Rohit Bokade
NUID: 001280767

February 7, 2021

1 Problem 1

$$p(x) = \begin{cases} \frac{x}{10}, & \text{if } x = 1, 2, 3 \\ 0, & \text{otherwise} \end{cases}$$

$$p(y|x) = \left(\frac{x+1}{2x}\right)^{y} \left(1 - \frac{x+1}{2x}\right)^{1-y}$$

where $y \in \{0, 1\}$

1.

$$\mathbb{E}(X) = \sum_x xp(x) = \sum_{x=1}^{4} x\left(\frac{x}{10}\right) = \boxed{3}$$

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{x=1}^{4} x^2\left(\frac{x}{10}\right) - 3^2 = 10 - 9 = \boxed{1}$$

2. By using the conditional equality $p(x, y) = p(y|x)p(x)$, we can find

$$p(x,y) = \left(\frac{x+1}{2x}\right)^{y} \left(1 - \frac{x+1}{2x}\right)^{1-y} \frac{x}{10}$$

and the marginal p(y) can be found by summing the joint distribution over the domain of x

$$p(y) = \sum_x p(x,y) = \sum_{x=1}^{4} \left[\left(\frac{x+1}{2x}\right)^{y} \left(1 - \frac{x+1}{2x}\right)^{1-y} \frac{x}{10}\right]$$

Next, exploiting the conditional equality again, we can find $p(x|y)$

$$\boxed{p(x|y) = \frac{p(x,y)}{p(y)} = \frac{\left(\frac{x+1}{2x}\right)^{y} \left(1 - \frac{x+1}{2x}\right)^{1-y} \frac{x}{10}}{\sum_{x=1}^{4} \left[\left(\frac{x+1}{2x}\right)^{y} \left(1 - \frac{x+1}{2x}\right)^{1-y} \frac{x}{10}\right]}}$$

3.

$$\mathbb{E}[X|Y=1] = \sum_{x=1}^{4} xp(x|y=1)$$
$$= \sum_{x=1}^{4} x \left(\frac{\left(\frac{x+1}{2x}\right)\frac{x}{10}}{\sum_{x=1}^{4}\left[\left(\frac{x+1}{2x}\right)\frac{x}{10}\right]} \right) = \frac{20}{7}$$

2 Problem 2: Legal reasoning

1. Let `crime_blood_type` represent that the blood type is the one that was found at the scene of crime and `guilty` represent that the defendant is guilty. Probability that anyone has the blood type is $p(\texttt{crime\_blood\_type}) = 0.1$ and therefore probability that the defendant has the blood type `crime_blood_type` is 0.1.

   However, there is no evidence provided to assess the dependence between `blood_type` and `guilty`. Therefore, it would be safe and unbiased to assume that $p(\texttt{guilty}|\texttt{crime\_blood\_type}) = p(\texttt{innocent}|\texttt{crime\_blood\_type}) = 0.5$ and $p(\texttt{guilty}) = p(\texttt{innocent}) = 0.5$.

   Thus,
   $p(\texttt{crime\_blood\_type}|\texttt{innocent}) = \frac{p(\texttt{innocent}|\texttt{crime\_blood\_type})p(\texttt{crime\_blood\_type})}{p(\texttt{innocent})} = \frac{0.5 \cdot 0.1}{0.5} = 0.1$

   - The claim made by the prosecutor is $p(\texttt{crime\_blood\_type}|\texttt{innocent}) = 0.1$ and $p(\texttt{guilty}) = 0.99$. Thus, $p(\texttt{crime\_blood\_type}|\texttt{innocent}) = 0.1$ might hold true if we make unbiased assumptions, which includes $p(\texttt{guilty}) = 0.5$ and inferring $p(\texttt{guilty})$ would be incorrect.

2. The only evidence available is that 1 in 8000 people would have the crime blood type. This would **only imply that the defendant has 1 in 8000 chance of having the crime blood type and not 1 in 8000 chance that the defendant is guilty**. The fault of the defender's argument is making an assumption between the dependence between these two random variables.

3 Problem 3: Maximum Likelihood Estimation (MLE)

$$p(x|\theta_0) = \begin{cases} e^{-\sum_{i=1}^{n}(x_i-\theta_0)}, & x \geq \theta_0 \\ 0, & \text{otherwise} \end{cases}$$

$$L(\theta) = \prod_{i=1}^{n} p(x|\theta)\ \forall x \geq \theta = \prod_{i=1}^{n} e^{-\sum_{i=1}^{n}(x_i-\theta)}$$
$$\log L(\theta) = \sum_{i=1}^{n} \log e^{-\sum_{i=1}^{n}(x_i-\theta)} = -\sum_{i=1}^{n} x_i + n\theta$$

Hence,

$$\arg\max_{\theta} = \arg\max_{\theta}\{N\theta - \sum^{N} X\}\ s.t.\ \theta \leq X$$
$$= \min\{X\} = \boxed{\min\{x_i\}\ \forall i \in \{1 \cdots n\}}$$

4 Problem 4: Maximum Likelihood Estimation (MLE)

X is a Bernoulli random variable with probability distribution

1.

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x = \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

And the likelihood function can be written as

$$L(p) = \prod_{i=1}^{n} f(x;p)$$

$$\log L(p) = \log \sum_{i=1}^{n} f(x;p) = \log \sum_{i=1}^{n} p^x(1-p)^{1-x} = \log p \sum_{i=1}^{n} x \log(1-p) \sum_{i=1}^{n} (1-x)$$

$$\frac{\partial \ell(p)}{\partial p} = \frac{\sum_{i=1}^{n} x}{p} - \frac{\sum_{i=1}(1-x)}{\log(1-p)} \overset{\text{set}}{=} 0$$

$$\sum_{i=1}^{n} x_i - p \sum_{i=1}^{n} x_i = p \sum_{i=1}^{n} (1-x)$$

$$\boxed{p = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

2.

$$\mathbb{E}[\hat{p}] = \mathbb{E}[\frac{1}{n} \sum_{i=1}^{n} x_i] = \frac{1}{n} (\mathbb{E}[x_i]) = \frac{1}{n} np = p$$

Therefore, $\hat{p}$ is an unbiased estimator of p

3. Expected square error of $\hat{p}$ in terms of p can be written as

$$\begin{aligned} \mathbb{E}[(p - L(p))^2] = \mathbb{E}[(p-\hat{p})^2] = \mathbb{E}[p^2 - 2p\hat{p} + \hat{p}^2] &= \mathbb{E}[p^2] - \mathbb{E}[2p\hat{p}] + \mathbb{E}[\hat{p}^2] \\ &= p^2 - \frac{2p}{n} \sum_{i=1}^{n} \mathbb{E}[x_{i=1}^{n}] + \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[x_{i=1}^{n}] \\ &= p^2 - \frac{2np^2}{n} + \frac{np}{n^2}((n-1)p+1) \\ &= p^2 - 2p^2 - \frac{p^2}{n} + \frac{p}{n} + p^2 \\ &\boxed{\text{MSE}(\hat{p}) = \frac{p(1-p)}{n}} \end{aligned}$$

4. We know that $p \in \{\frac{1}{4}, \frac{3}{4}\}$ and $n = 3$. $R(\theta, \delta) = \frac{p(1-p)}{n}$. For $p \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, we have the following values for $R(\theta, \delta) = \{0.0625, 0.0833, 0.0625\}$. $\hat{p}$ would be an inadmissible estimator of p if there exists an $R(\theta, \delta)$ that gives a better estimate of MSE. Intuitively, the best estimator of p would have $MSE = 0$ (lowest uncertainty). Hence, in this case, $\hat{p}$ would be an inadmissible estimator for p.

5 Problem 5

$$\mu_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\sigma^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_{\text{MLE}})^2$$

-

$$\begin{aligned}
\mathbb{E}[\sigma^2_{\text{MLE}}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_{\text{MLE}})^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(x_i^2) - 2n\bar{x}^2 + n\bar{x}^2\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(x_i^2) - n\bar{x}^2\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}x_i^2\right] - \mathbb{E}\left[\bar{x}^2\right] = \mathbb{E}\left[x_i^2\right] - \mathbb{E}\left[\bar{x}^2\right]
\end{aligned}$$

We know that $Var[X] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$ from the definition of variance.

$$\begin{aligned}
\mathbb{E}\left[x_i^2\right] - \mathbb{E}\left[\bar{x}^2\right] &= \sigma_x^2 + \mathbb{E}[x_i]^2 - \sigma_x^2 - \mathbb{E}[x_i]^2 = \sigma_x^2 - \sigma_{\bar{x}}^2 \\
&= \sigma_x^2 - Var[\bar{x}] = \sigma_x^2 - Var\left[\frac{1}{n}\sum_{i=1}^{n}x_i\right] \\
&= \sigma_x^2 - \left(\frac{1}{n}\right)^2 Var\left[\sum_{i=1}^{n}x_i\right]
\end{aligned}$$

Now,

$$\sigma_x^2 - \left(\frac{1}{n}\right)^2 Var\left[\sum_{i=1}^{n}x_i\right] = \sigma_x^2 - \left(\frac{1}{n}\right)^2 n\sigma_x^2 - \frac{1}{n}\sigma_x^2 = \boxed{\frac{n-1}{n}\sigma_x^2 < \sigma_x^2}$$

Thus, σ^2_{MLE} is an unbiased estimator of σ^2.

-

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(x_i - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i^2 - 2x_i\mu + \mu^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[x_i^2] - 2\sum_{i=1}^{n}\mathbb{E}[x_i\mu] + \sum_{i=1}^{n}\mathbb{E}[\mu^2]\right] = \frac{1}{n}[n\sigma^2 + n^2\mu^2 + n^2\mu 2 - 2n^2\mu^2] = \frac{1}{n}[n\sigma^2] = \sigma^2
\end{aligned}$$

Hence, $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

6 Problem 6

Consider a covariance matrix Σ for any random variable X. A covariance matrix is symmetric by definition. Next. consider a vector $Z \to (z_1, \cdots, z_n)$ in n space.

$$\begin{aligned} Z^T &= Z^T \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T] \\ &= \mathbb{E}[((x - \mathbb{E}[x])^T \cdot Z)^T((x - \mathbb{E}[x])^T \cdot Z)] \\ &= \mathbb{E}[\underbrace{((x - \mathbb{E}[x])^T \cdot Z)^T}_{\text{Non-negative}}] \geq 0 \end{aligned}$$

Since, the value of the term within the expectation is non-negative, the expectation must be non-negative. Therefore, $Z^T \Sigma Z \geq 0$. Now, if X were to be a Gaussian random variable, we know that the covariance matrix Σ must be invertivle and full rank. Also, $Z^T \Sigma Z > 0$ and is positive definite.

7 Problem 7

$$||X^T w - y||^2 = \underbrace{(w - \hat{w})^T X^T X(w - \hat{w})}_{(x - M^{-1}b)^T x^T x(x - M^{-1}b)} + ||X^T \hat{w} - y||^2$$

We can show that $||X^T w - y||^2 - ||X^T \hat{w} - y||^2 = (w - \hat{w})^T X^T X(w - \hat{w})$

Let $x = w, M = X^T X, b = X^T y$. Then $(w - \hat{w})^T X^T X(w - \hat{w})$ can be written as $(x - M^{-1}b)^T x^T x(x - M^{-1}b)$. Also, we know that

$$(x - M^{-1}b)^T x^T x(x - M^{-1}b) = x^T M x - 2b^T x + b^T M^{-1} b$$

Therefore

$$\begin{aligned} (w - \hat{w})^T X^T X(w - \hat{w}) &= w^T X^T X w - 2(X^T y)^T w + (X^T y)^T (X^T X)^{-1} (X^T y) \\ &= w^T X^T X w - 2(X^T y)^T w + (X^T y)^T \hat{w} (X^T y)^{-1} (X^T y) \end{aligned}$$

Next,

$$\begin{aligned} ||X^T w - y||^2 - ||X^T \hat{w} - y||^2 &= (X^T w)^T (X^T w) - 2y^T (X^T w) + y^T y - (X^T \hat{w})^T (X^T \hat{w}) + 2y^T (X^T \hat{w}) - y^T y \\ &= (X^T w)^T (X^T w) - (X^T \hat{w})^T (X^T \hat{w}) - 2y^T (X^T w) + 2y^T (X^T \hat{w}) \end{aligned}$$