ESE 524 - Detection and Estimation Theory

Midterm, Spring 2019
Total: 100 pts
Problems and Solutions
Duration: 80 mins

Name:________________________________ **Student ID**:___________

(1) (**40 pts**) **MLE and MMSE for Bernoulli distribution**
The outcome of a coin tossing problem is usually modeled as a Bernoulli-distributed random variable X. Let $X = 1$ and $X = 0$ denote "heads" and "tails," respectively. Suppose that for a specific coin, we have $P(X = 1) = p$, i.e.,

$$P(X = x) = p^x(1-p)^{1-x}, \text{ for } x \in \{0, 1\}, \quad 0 < p < 1.$$

Now, we independently toss the coin for N times, and observe the outcomes $\{X_1, X_2, \ldots, X_N\}$.

a) (**10 pts**) From a frequentist's perspective, p is a fixed but unknown parameter. Find the maximum likelihood estimator (MLE) of p.

b) (**10 pts**) Check whether the MLE is unbiased and efficient. Derive the variance of the MLE and the Cramér-Rao bound (CRB) when examining the efficiency of the estimator.

c) (**10 pts**) Is the MLE a sufficient statistic for p? Justify your solution.

d) (**10 pts**) From a Bayesian perspective, we usually have some prior knowledge about p, and the unknown p can be modeled as a random variable with a prior distribution $\pi(p)$. Suppose the prior, $\pi(p)$, is distributed as

$$\pi(p) = 6p(1-p), \text{ for } p \in (0, 1).$$

Find the posterior probability distribution of p based on the N observations, $\{X_1, X_2, \ldots, X_N\}$, i.e., $P(p|X_1, \ldots, X_N)$. Also find the Bayesian minimum mean squared error (MMSE) estimator for p which is given as $\mathbb{E}[p|X_1, \ldots, X_N]$.

Hint: The results can be expressed using the Beta function, which is defined as follows:

$$\text{Beta}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt.$$

Solution:

a) The log-likelihood is given by

$$\begin{aligned}\log P(X_1, X_2, \ldots, X_N; p) &= \log p^{\sum_{n=1}^N X_n}(1-p)^{N-\sum_{n=1}^N X_n} \\ &= \sum_{n=1}^N X_n \log p + \left(N - \sum_{n=1}^N X_n\right)\log(1-p)\end{aligned}$$

Take the derivative,

$$\frac{\partial \log P(X_1, X_2, \ldots, X_N)}{\partial p} = \frac{\sum_{n=1}^N X_n}{p} - \frac{N - \sum_{n=1}^N X_n}{1-p} = 0$$

$$\hat{p}_{\text{MLE}} = \frac{\sum_{n=1}^N X_n}{N}.$$

b) We can compute

$$\mathbb{E}[X_n] = p, \qquad \text{var}[X_n] = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = p(1-p)$$

Then because of the independence, we have

$$\begin{aligned} \mathbb{E}[\hat{p}_{\text{MLE}}] &= \frac{Np}{N} = p, \qquad \rightarrow \text{ unbiased} \\ \text{var}[\hat{p}_{\text{MLE}}] &= \frac{\sum_{n=1}^N \text{var}[X_n]}{N^2} = \frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N}. \end{aligned}$$

The Fisher information is

$$\mathcal{I}(p) = N\mathcal{I}_1(p) = N\mathbb{E}\left[\left(\frac{\partial \log P(X)}{\partial p}\right)^2\right] = N\mathbb{E}\left(\frac{X}{p} - \frac{1-X}{1-p}\right)^2 = \frac{N\mathbb{E}(X-p)^2}{p^2(1-p)^2} = \frac{N}{p(1-p)}.$$

Thus,

$$\text{CRB}(p) = \frac{1}{\mathcal{I}(p)} = \frac{p(1-p)}{N} = \text{var}[\hat{p}_{\text{MLE}}],$$

which implies the efficiency of the MLE.

c)

$$\begin{aligned} P(X_1, X_2, \ldots, X_N) &= p^{\sum_{n=1}^N X_n}(1-p)^{N-\sum_{n=1}^N X_n} \\ &= \underbrace{p^{N\hat{p}_{\text{MLE}}}(1-p)^{N-N\hat{p}_{\text{MLE}}}}_{g(\hat{p}_{\text{MLE}}, p)} \cdot \underbrace{1}_{h(X)}. \end{aligned}$$

Using the factorization theorem, we can easily see $\hat{p}_{\text{MLE}}$ is sufficient for p. Note that here $h(\cdot)$ is a constant function.

d) We have

$$\begin{aligned} P(p|X_1, X_2, \ldots, X_N) &= \frac{P(X_1, X_2, \ldots, X_N|p)\pi(p)}{\int_0^1 P(X_1, X_2, \ldots, X_N|p)\pi(p)dp} \\ &= \frac{p^{1+\sum_{n=1}^N X_n}(1-p)^{N+1-\sum_{n=1}^N X_n}}{\int_0^1 p^{1+\sum_{n=1}^N X_n}(1-p)^{N+1-\sum_{n=1}^N X_n}dp} \\ &= \frac{p^{1+\sum_{n=1}^N X_n}(1-p)^{N+1-\sum_{n=1}^N X_n}}{\text{Beta}(2+\sum_{n=1}^N X_n, N+2-\sum_{n=1}^N X_n)} \end{aligned}$$

where $p \in (0,1)$. The Bayesian MMSE estimator is the posterior mean,

$$\begin{aligned} \mathbb{E}(p|X_1, X_2, \ldots, X_N) &= \int_0^1 pP(p|X_1, \ldots, X_N)dp \\ &= \frac{\int_0^1 p^{2+\sum_{n=1}^N X_n}(1-p)^{N+1-\sum_{n=1}^N X_n}dp}{\text{Beta}(2+\sum_{n=1}^N X_n, N+2-\sum_{n=1}^N X_n)} \\ &= \frac{\text{Beta}(3+\sum_{n=1}^N X_n, N+2-\sum_{n=1}^N X_n)}{\text{Beta}(2+\sum_{n=1}^N X_n, N+2-\sum_{n=1}^N X_n)} \end{aligned}$$

(2) **(20 pts) Sufficient Statistics**

Let $\boldsymbol{x} = [x[0], \ldots, x[N-1]]$ be a set of i.i.d. Gaussian random variables with mean μ and variance μ^2, i.e., $x[n] \sim \mathcal{N}(\mu, \mu^2), \forall n \in \{0, \ldots, N-1\}$.

a) (**10 pts**) Show that $\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ is not a sufficient statistic for μ by demonstrating that the conditional joint probability density function of $\boldsymbol{x}$ given $\bar{x}$ is a function of μ.

Hint 1: The conditional density is:

$$p(\boldsymbol{x}|\bar{x}=\bar{x}_0;\mu)=\frac{p(\boldsymbol{x};\mu)\delta(\sum_{n=0}^{N-1}x[n]-\bar{x}_0)}{p(\bar{x}=\bar{x}_0;\mu)}$$

where $\delta(\cdot)$ is the Dirac delta function.

Hint 2: The distribution of $p(\bar{x}=\bar{x}_0;\mu)$ is $\mathcal{N}\left(\mu,\mu^2/N\right)$

b) (**10 pts**) Find a two-dimensional sufficient statistic for μ.

Solution:

a) From the hint, the conditional density is

$$\frac{\frac{1}{(2\pi\mu^2)^{N/2}}\exp(-\frac{1}{2\mu^2}\sum_{n=0}^{N-1}(x[n]-\mu)^2)\delta(\sum_{n=0}^{N-1}x[n]-\bar{x}_0)}{\frac{1}{(2\pi\mu^2/N)^{1/2}}\exp(-\frac{N}{2\mu^2}(\bar{\mathbf{x}}_0-\mu)^2)}=$$

$$\frac{\sqrt{N}}{(2\pi)^{(N-1)/2}\mu^{N-1}}\exp(-\frac{1}{2\mu^2}\sum_{n=0}^{N-1}(x[n]-\mu)^2+\frac{N}{2\mu^2}(\bar{\mathbf{x}}_0-\mu)^2))\delta(\sum_{n=0}^{N-1}x[n]-\bar{x}_0)$$

In the case when the dirac delta function is non-zero, there is no way to cancel out the μ from the denominator, so by the definition of sufficient statistics, $\bar{x}$ can't be a sufficient statistic.

b) By the factorization theorem, $T(x)=\begin{bmatrix}\sum_{n=0}^{N-1}x[n]\\ \sum_{n=0}^{N-1}x[n]^2\end{bmatrix}$ is jointly sufficient for μ

(3) (**20 pts**) **Statistical Theory**

a) (**10 pts**) Let $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ and $\boldsymbol{\epsilon}$ be a random vector drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0},\boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is a $p\times p$ Identity matrix. Show that

$$\mathbb{E}\left[\|\boldsymbol{X}\boldsymbol{\epsilon}\|_2^2\right]=\|\boldsymbol{X}\|_{\mathrm{F}}^2\,.$$

Hint 1: If $\boldsymbol{x}\in\mathbb{R}^{n\times 1}$, then $\|\boldsymbol{x}\|_2^2=\mathrm{Tr}(\boldsymbol{x}\boldsymbol{x}^T)$. If $\boldsymbol{X}\in\mathbb{R}^{n\times p}$, then $\|\boldsymbol{X}\|_{\mathrm{F}}^2=\mathrm{Tr}(\boldsymbol{X}\boldsymbol{X}^T)$.

Hint 2: Use the commutative property of expectation and trace of a matrix.

b) (**10 pts**) Let $\boldsymbol{\epsilon}$ be a random vector drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0},\boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is a $p\times p$ Identity matrix. If $\boldsymbol{X}$ is a projection matrix, i.e., $\boldsymbol{X}=\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ where $\boldsymbol{A}\in\mathbb{R}^{n\times p}$ and $\mathrm{rank}(\boldsymbol{A})=p$, then show that

$$\mathbb{E}\left[\|\boldsymbol{X}\boldsymbol{\epsilon}\|_2^2\right]=p.$$

Solution:

a) Expanding the $\|\cdot\|_2^2$ as follows:

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{X}\boldsymbol{\epsilon}\|_2^2\right]&=\mathbb{E}\left[\boldsymbol{\epsilon}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\epsilon}\right]\\&=\mathbb{E}\left[\mathrm{Tr}(\boldsymbol{X}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\boldsymbol{X}^T)\right]\\&=\mathrm{Tr}\left\{\mathbb{E}[\boldsymbol{X}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\boldsymbol{X}^T]\right\}\\&=\mathrm{Tr}\left\{\boldsymbol{X}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T]\boldsymbol{X}^T\right\}\\&=\mathrm{Tr}\left\{\boldsymbol{X}\boldsymbol{I}_p\boldsymbol{X}^T\right\}\\&=\|\boldsymbol{X}\|_{\mathrm{F}}^2\end{aligned}$$

Here, we used we used $\mathrm{Tr}(\boldsymbol{A}\boldsymbol{B})=\mathrm{Tr}(\boldsymbol{B}\boldsymbol{A})$, and the commutative property of trace and expectation.

b) Using the result in (b), we get

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{X}\boldsymbol{\epsilon}\|_2^2\right] &= \|\boldsymbol{X}\|_{\mathrm{F}}^2 \\ &= \mathrm{Tr}(\boldsymbol{X}\boldsymbol{X}^T) \\ &= \mathrm{Tr}\left[\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\right] \\ &= \mathrm{Tr}[\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T] \\ &= p\end{aligned}$$

(4) **(20 pts) Best Linear Unbiased Estimator (BLUE)**
Given a linear model, $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{w}$, where $\mathbb{E}(\boldsymbol{w}) = \boldsymbol{0}$ and $\mathbb{E}(\boldsymbol{w}\boldsymbol{w}^T) = \boldsymbol{C}$, we wish to estimate

$$\boldsymbol{\alpha} = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{b}$$

where $\boldsymbol{B}$ is a known $p \times p$ invertible matrix and $\boldsymbol{b}$ is a known $p \times 1$ vector. Prove that the BLUE of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{B}\hat{\boldsymbol{\theta}} + \boldsymbol{b}$$

where $\hat{\boldsymbol{\theta}}$ is the BLUE for $\boldsymbol{\theta}$.

Hint 1: Replace $\boldsymbol{\theta}$ in the linear model with $\boldsymbol{B}^{-1}(\boldsymbol{\alpha} - \boldsymbol{b})$ and rewrite as a new linear model.

Hint 2: The BLUE of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x}$.

Solution: As matrix $\boldsymbol{B}$ is invertible, $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{\theta} = \boldsymbol{B}^{-1}(\boldsymbol{\alpha} - \boldsymbol{b})$$

Thus, $\boldsymbol{x}$ can be written as

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{H}\boldsymbol{B}^{-1}(\boldsymbol{\alpha} - \boldsymbol{b}) + \boldsymbol{w} \\ &= \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{\alpha} - \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b} + \boldsymbol{w}\end{aligned}$$

Implies

$$\boldsymbol{x} + \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b} = \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{\alpha} + \boldsymbol{w}$$

Let $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b}$ and $\boldsymbol{H}\boldsymbol{B}^{-1} = \boldsymbol{H}'$. Then,

$$\boldsymbol{x}' = \boldsymbol{H}'\boldsymbol{\alpha} + \boldsymbol{w}$$

The BLUE is given as

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{x}'$$

Substituting $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b}$, we get

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= (\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}'^T\boldsymbol{C}^{-1}(\boldsymbol{x} + \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b}) \\ &= (\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{x} + (\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}'^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b} \\ &= (\boldsymbol{B}^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{B}^{-1})^{-1}\boldsymbol{B}^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} + (\boldsymbol{B}^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{B}^{-1})^{-1}\boldsymbol{B}^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b} \\ &= \boldsymbol{B}(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H})^{-1}\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H})^{-1}\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{b} \\ &= \boldsymbol{B}\hat{\boldsymbol{\theta}} + \boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{b} \\ &= \boldsymbol{B}\hat{\boldsymbol{\theta}} + \boldsymbol{b}\end{aligned}$$