Stochastic Processes

Midterm 2

Solutions

1. **Bayes Detection** (10+10=20 points)

(a) Suppose the state of nature is $\Omega = \{x_1, \cdots, x_N\}$. The MAP decision rule based on the observation $Y = y$ is given by

$$\begin{aligned}\hat{X}_{\text{MAP}} &= \arg\max_{x_j} P[X = x_j \mid Y = y] \\ &= \arg\max_{x_j} f_Y(y \mid X = x_j) P[X = x_j].\end{aligned}$$

When $c_{i,j} = 1 - \delta_{i,j}$, the average cost is

$$\begin{aligned}C(D) &\triangleq E[C] \\ &= \sum_i \sum_j c_{i,j} P[\text{decide } H_i \text{ and } H_j \text{ is true}] \\ &= \sum_i \sum_j (1 - \delta_{i,j}) P[\text{decide } H_i \text{ and } H_j \text{ is true}] \\ &= \sum_{i \neq j} P[\text{decide } H_i \text{ and } H_j \text{ is true}],\end{aligned}$$

which is exactly the probability of decision error. Bayes detection rule guarantees that this decision error probability is minimized.

(b) The maximum likelihood detection rules states

$$\mathrm{L}(\mathrm{H}_1) \underset{H_2}{\overset{H_1}{\gtrless}} \mathrm{L}(\mathrm{H}_2),$$

where $\mathrm{L}(\mathrm{H}_1)$ and $\mathrm{L}(\mathrm{H}_2)$ are the likelihood of the signal $\mathbf{s}$ associated with hypothesis H_1 and H_2, respectively. More specifically, by taking natural log of both sides, we have

$$-\frac{1}{2}(\mathbf{y} - \mathbf{s})^H \mathbf{K}_n^{-1} (\mathbf{y} - \mathbf{s}) \underset{H_2}{\overset{H_1}{\gtrless}} -\frac{1}{2} \mathbf{y}^H \mathbf{K}_n^{-1} \mathbf{y}.$$

With rearrangement, it follows that

$$\underbrace{\mathbf{s}^H \mathbf{K}_n^{-1}}_{= \mathbf{w}_{\text{ML}}^H} \cdot \mathbf{y} \underset{H_2}{\overset{H_1}{\gtrless}} \frac{\mathbf{s}^H \mathbf{K}_n^{-1} \mathbf{s}}{2}.$$

That is we can write the ML decision rule as

$$\mathbf{w}_{\text{ML}}^H \cdot \mathbf{y} \underset{H_2}{\overset{H_1}{\gtrless}} \eta,$$

where

$$\mathbf{w}_{\text{ML}} = \mathbf{K}_n^{-1} \mathbf{s} \quad \text{and} \quad \eta = \frac{\mathbf{s}^H \mathbf{K}_n^{-1} \mathbf{s}}{2}.$$

2. **Estimation** $(10 + 5 + 5 = 20$ points$)$

Consider the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w},$$

where $\mathbf{H}$ is a known $m \times n$ observation matrix, $\mathbf{x}$ is an $n \times 1$ unknown parameter, and $\mathbf{w}$ is a Gaussian noise vector with $\mathbf{w} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. Assume $\mathbf{H}$ is full rank.

(a) Let's consider a more general case for $\mathbf{w} \sim \mathcal{N}(0, \mathbf{K})$ in the proof, where $\mathbf{K}$ is the covariance matrix of $\mathbf{w}$ that is not necessarily diagonal. The situation in this problem is just a special case that $\mathbf{K} = \sigma^2\mathbf{I}$. So if $\mathbf{w} \sim \mathcal{N}(0, \mathbf{K})$, the ML linear transformation matrix is

$$\mathbf{T}_{ML} = (\mathbf{H}^T\mathbf{K}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{K}^{-1}.$$

Let $\mathbf{K}_1$ and $\mathbf{K}_2$ be the covariance matrices of $\hat{\mathbf{x}}_{ML}$ and any other linear unbiased estimator $\hat{\mathbf{X}} = \mathbf{T} \cdot \mathbf{y}$, respectively. Therefore, this problem requires you to show that $\mathbf{K}_2 - \mathbf{K}_1$ is p.s.d., i.e.

$$\boldsymbol{\alpha}^T(\mathbf{K}_2 - \mathbf{K}_1)\boldsymbol{\alpha} \geq 0, \quad \forall\ \boldsymbol{\alpha}. \tag{1}$$

The key steps have two:

$\rightarrow$ Any covariance matrix $\mathbf{K}$ is p.s.d. (see topic2)

$\rightarrow$ $\mathbf{T}\mathbf{H} = \mathbf{I}$ for any linear unbiased estimator $\hat{\mathbf{x}} = \mathbf{T} \cdot \mathbf{y}$ in the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$.
This can easily be shown that for $E[\hat{\mathbf{x}}] = E[\mathbf{T} \cdot \mathbf{y}] = \mathbf{T}\mathbf{H}\mathbf{x} = \mathbf{x}$, we must have $\mathbf{T}\mathbf{H} = \mathbf{I}$.

We will use these two facts later in the proof.

So, our objective is to show (1) is true. Carrying out the covariance matrices $\mathbf{K}_1$ and $\mathbf{K}_2$, we have

$$\begin{aligned}
\mathbf{K}_1 &= \mathbf{T}_{ML}\mathbf{K}\mathbf{T}_{ML}^T, \\
\mathbf{K}_2 &= \mathbf{T}\mathbf{K}\mathbf{T}^T.
\end{aligned}$$

It follows

$$\mathbf{K}_2 - \mathbf{K}_1 = \mathbf{T}\mathbf{K}\mathbf{T}^T - \mathbf{T}_{ML}\mathbf{K}\mathbf{T}_{ML}^T.$$

Next we can conjecture that

$$\mathbf{K}_2 - \mathbf{K}_1 = (\mathbf{T} - \mathbf{T}_{ML})\mathbf{K}(\mathbf{T} - \mathbf{T}_{ML})^T. \tag{2}$$

If this is the case, then by the fact that $\mathbf{K}$ is p.s.d., we know that

$$\begin{aligned}
\boldsymbol{\alpha}^T(\mathbf{K}_2 - \mathbf{K}_1)\boldsymbol{\alpha} &= \underbrace{\boldsymbol{\alpha}^T(\mathbf{T} - \mathbf{T}_{ML})}_{\boldsymbol{\beta}^T}\mathbf{K}\underbrace{(\mathbf{T} - \mathbf{T}_{ML})^T\boldsymbol{\alpha}}_{\boldsymbol{\beta}} \\
&= \boldsymbol{\beta}^T\mathbf{K}\boldsymbol{\beta} \geq 0,
\end{aligned}$$

which reaches our objective in (1).
So, now the question is whether or not (2) is true.
Expanding (2) gives

$$\mathbf{T}\mathbf{K}\mathbf{T}^T - \mathbf{T}_{ML}\mathbf{K}\mathbf{T}^T - \mathbf{T}\mathbf{K}\mathbf{T}_{ML}^T + \mathbf{T}_{ML}\mathbf{K}\mathbf{T}_{ML}^T, \tag{3}$$

where

$$\begin{aligned}
\mathbf{T}_{ML}\mathbf{K}\mathbf{T}^T &= (\mathbf{H}^T\mathbf{K}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{K}^{-1}\mathbf{K}\mathbf{T}^T \\
&= (\mathbf{H}^T\mathbf{K}^{-1}\mathbf{H})^{-1} \quad \text{(we use } \mathbf{TH=I} \text{ in this equality)} \\
\mathbf{T}_{ML}\mathbf{K}\mathbf{T}_{ML}^T &= (\mathbf{H}^T\mathbf{K}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{K}^{-1}\mathbf{K}\mathbf{K}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{K}^{-1}\mathbf{H})^{-1} \\
&= (\mathbf{H}^T\mathbf{K}^{-1}\mathbf{H})^{-1}.
\end{aligned}$$

Thus, we see the last 3 terms in (3) are identical. It follows that

$$\begin{aligned}
(\mathbf{T}-\mathbf{T}_{ML})\mathbf{K}(\mathbf{T}-\mathbf{T}_{ML})^T &= \mathbf{T}\mathbf{K}\mathbf{T}^T - \mathbf{T}_{ML}\mathbf{K}\mathbf{T}^T - \mathbf{T}\mathbf{K}\mathbf{T}_{ML}^T + \mathbf{T}_{ML}\mathbf{K}\mathbf{T}_{ML}^T \\
&= \mathbf{T}\mathbf{K}\mathbf{T}^T - \mathbf{T}_{ML}\mathbf{K}\mathbf{T}_{ML}^T \\
&= \mathbf{K}_2 - \mathbf{K}_1.
\end{aligned}$$

This completes the proof.

(b) Since $\mathbf{K}_2 - \mathbf{K}_1$ is p.s.d., it's diagonal term is non-negative. This is because that, if we choose the vector

$$\boldsymbol{\alpha} = [0, \cdots, 0, \underbrace{1}_{i\text{th position}}, 0, \cdots, 0]^T,$$

the vector with a '1' in the ith position and zero in others, then

$$\boldsymbol{\alpha}^T(\mathbf{K}_2 - \mathbf{K}_1)\boldsymbol{\alpha} = (\mathbf{K}_2 - \mathbf{K}_1)_{ii} \geq 0, \tag{4}$$

where $(\mathbf{K}_2 - \mathbf{K}_1)_{ii}$ is the ith diagonal term in $\mathbf{K}_2 - \mathbf{K}_1$.
Since $(\mathbf{K}_2 - \mathbf{K}_1)_{ii} = (\mathbf{K}_2)_{ii} - (\mathbf{K}_1)_{ii}$, and the ith diagonal term of $\mathbf{K}_2$ and $\mathbf{K}_1$ is the variance of the ith component of $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}_{ML}$, respectively. It follows from (4) that

$$(\mathbf{K}_2 - \mathbf{K}_1)_{ii} = (\mathbf{K}_2)_{ii} - (\mathbf{K}_1)_{ii} \geq 0.$$

So the variance of the ith component of $\hat{\mathbf{x}}$ is greater than or equal to that of $\hat{\mathbf{x}}_{ML}$.

(c) The least squares estimator is defined as

$$\hat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2.$$

This is equivalent to finding a vector $\mathbf{H}\hat{\mathbf{x}}_{LS}$ in the column space of $\mathbf{H}$ that is closest to $\mathbf{y}$. Thus, $\mathbf{H}\hat{\mathbf{x}}_{LS}$ is the orthogonal projection of $\mathbf{y}$ onto the space spanned by the columns of $\mathbf{H}$.
Geometrically, we see $\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_{LS}$ will be orthogonal to $\mathbf{H}\mathbf{x}$ for all $\mathbf{x}$. Mathematically, this is

$$(\mathbf{H}\mathbf{x})^T(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_{LS}) = 0 \quad \text{for all } \mathbf{x}.$$

This leads to the normal equation

$$\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}_{LS}) = 0$$

3. **Conditional Expectation of Jointly Gaussian** (10+10=20 points)

(a) The cross-covariance matrix $\mathbf{K}_{y\hat{z}} = E\Big[(\mathbf{y} - \mathbf{m}_y)(\hat{\mathbf{z}} - E[\hat{\mathbf{z}}])^T\Big]$ can be obtained as

$$\begin{aligned}
\mathbf{K}_{y\hat{z}} &= E\Big[(\mathbf{y} - \mathbf{m}_y)(\hat{\mathbf{z}} - E[\hat{\mathbf{z}}])^T\Big] \\
&= E[(\mathbf{y} - \mathbf{m}_y)\hat{\mathbf{z}}^T] \\
&= E[\mathbf{y}\hat{\mathbf{z}}^T] \qquad \text{(since } E[\hat{\mathbf{z}}] = 0) \\
&= E[\mathbf{y}\mathbf{z}^T] - E\Big[\mathbf{y}E[\mathbf{z}^T|\mathbf{y}]\Big] \\
&= E[\mathbf{y}\mathbf{z}^T] - E\Big[E[\mathbf{y}\mathbf{z}^T|\mathbf{y}]\Big] \\
&= E[\mathbf{y}\mathbf{z}^T] - E[\mathbf{y}\mathbf{z}^T] \\
&= \mathbf{0}.
\end{aligned}$$

(b) (Not graded) See hw2 solutions.

4. **Gaussian Sample** ($10 \times 5 = 50$ points)

(a) First, it is clear to see $E[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$.

Second, we can use Chebyshev inequality to examine consistency of any unbiased estimators. Here, we have

$$P\left[\,\left|\bar{X}_n - \mu\right| > k\right] \leq \frac{\mathrm{Var}(\bar{X}_n)}{k^2}, \quad \text{for any} \quad k > 0. \tag{5}$$

The variance of the sample mean is

$$\begin{aligned}
\mathrm{Var}(\bar{X}_n) &= \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2.
\end{aligned}$$

Therefore, by plugging the above result into (5), we can easily show the consistency of sample mean, i.e. for any $k > 0$,

$$\lim_{n\to\infty} P\left[\,\left|\bar{X}_n - \mu\right| > k\right] = 0.$$

(b) The mean and variance are both unknown. We need simultaneously find their ML estimator jointly. That is the vector of parameter to be estimated is $\boldsymbol{\theta} = [\mu, \sigma^2]^T$.

The likelihood function of $\boldsymbol{\theta}$ given $\mathsf{x} = [X_1 = \mathsf{x}_1, X_2 = \mathsf{x}_2, \cdots, X_n = \mathsf{x}_n]$ is

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\theta}|\mathsf{x}) &= f_{\mathbf{x}}(\mathsf{x};\theta) = \prod_{i=1}^{n} f_{X_i}(\mathsf{x}_i;\theta) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathsf{x}_i - \mu)^2\right).
\end{aligned}$$

And, the log likelihood function is

$$\log\mathrm{L}(\theta|\mathsf{x}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathsf{x}_i - \theta)^2.$$

Taking partial derivative and solving the following equations simultaneously gives us possible candidates of the ML estimators. It follows that

$$\frac{\partial \log \mathrm{L}(\theta|\mathsf{x})}{\partial \mu} = 0 \longrightarrow \hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n,$$
$$\frac{\partial \log \mathrm{L}(\theta|\mathsf{x})}{\partial \sigma^2} = 0 \longrightarrow \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2,$$

which are indeed the ML estimators after checking with the boundaries.

So, the answer to this problem is

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

(c) We see the relation between $\hat{\sigma}^2_{ML}$ and S_n^2 is given by

$$\hat{\sigma}^2_{ML} = \frac{n-1}{n} S_n^2.$$

And, since S_n^2 is unbiased, we thus conclude $\hat{\sigma}^2_{ML}$ is not unbiased.

(d) The MSE between $\hat{\sigma}^2_{ML}$ and σ^2 is

$$\begin{aligned}
E\Big[(\hat{\sigma}^2_{ML} - \sigma^2)^2\Big] &= E\left[\left(\frac{n-1}{n} S_n^2 - \sigma^2\right)^2\right] \\
&= E\left[\left(\left(\frac{n-1}{n} S_n^2 - \frac{n-1}{n}\sigma^2\right) - \frac{1}{n}\sigma^2\right)^2\right] \\
&= E\left[\left(\frac{n-1}{n} S_n^2 - \frac{n-1}{n}\sigma^2\right)^2\right] + \left(\frac{1}{n}\sigma^2\right)^2 \\
&= \left(\frac{n-1}{n}\right)^2 \mathrm{Var}\left(S_n^2\right) + \left(\frac{1}{n}\sigma^2\right)^2 \\
&= \frac{2n-1}{n^2}\sigma^4.
\end{aligned}$$

With a little more algebraic efforts, it is not difficult to see with positive integer n

$$\frac{2n-1}{n^2}\sigma^4 < \mathrm{Var}(S_n^2) = \frac{2\sigma^4}{n-1}.$$

This means the ML estimator of σ^2 is more accurate in terms of MSE than the sample variance estimator, although MLE is not unbiased.

(e) We know

$$\begin{aligned}
1 - 2F_{T_{n-1}}(-z) &= P\left[-z \le \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \le z\right] \\
&= P\left[\bar{X}_n - \frac{S_n z}{\sqrt{n}} \le \mu \le \bar{X}_n + \frac{S_n z}{\sqrt{n}}\right] \\
&= 1 - \alpha.
\end{aligned}$$

For $1 - \alpha = 0.95$, we need $F_{T_{n-1}}(-z) = 0.025$, where $F_{T_{n-1}}(-z) = P\left[T_{n-1} \le -z\right]$ is the cumulative distribution function of T_{n-1}. This yields $z = -\beta$. It follows the interval with confidence level 0.95 is

$$\left[\bar{X}_n + \frac{S_n\beta}{\sqrt{n}}, \bar{X}_n - \frac{S_n\beta}{\sqrt{n}}\right].$$