ECE531 Spring 2009 Final Examination

Instructions: This exam is worth a total of 500 points. You may consult two double-sided letter-sized sheets of notes (in your own handwriting) and you may use a calculator during the exam. Please show your work on each problem and box/circle your final answers. Points may be deducted for a disorderly presentation of your solution. The exam is closed-book.

1. 200 points. Suppose that an unknown scalar parameter Θ is known to have a prior distribution of

$$p_\Theta(\theta) = \pi(\theta) = \begin{cases} e^{-\theta} & \theta \geq 0 \\ 0 & \theta < 0. \end{cases}$$

You receive a scalar observation

$$Y = \Theta + U$$

with U independent of Θ and possessing the distribution

$$p_U(u) = \begin{cases} e^{-u} & u \geq 0 \\ 0 & u < 0. \end{cases}$$

(a) (50 points) Find the MMSE estimator of the unknown scalar parameter.

(b) (50 points) Find the LMMSE (linear MMSE) estimator of the unknown scalar parameter. Compare your answer to Part (a).

(c) (50 points) Find the MAP estimator of the unknown scalar parameter. Hint: Is the MAP estimator unique?

(d) (50 points) Find the ML estimator of the unknown scalar parameter and compare your answer to Part (c). Hint: The ML estimator is appropriate for non-random parameter estimation, hence your analysis here should not use the prior on Θ.

Solution: See next two pages...

Problem 1

a) To find the MMSE estimator, we need to compute the conditional mean which requires us to know the conditional density

$$P_{\Theta}(\theta|y) = \frac{P_Y(y|\theta)\, P_{\Theta}(\theta)}{P_Y(y)}$$

We are given $P_{\Theta}(\theta) = \begin{cases} e^{-\theta} & \theta \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Based on the observation model $Y = \Theta + U$, we can also write

$$P_Y(y|\theta) = \begin{cases} \exp(\theta - y) & y \geq \theta \\ 0 & \text{otherwise.} \end{cases}$$

Finally $P_Y(y) = \int_{-\infty}^{\infty} P_Y(y|\theta)\, P_{\Theta}(\theta)\, d\theta$

$$= \int_0^y e^{-y} e^{\theta} e^{-\theta}\, d\theta$$

$$= \begin{cases} y e^{-y} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$ (you can check that this is a valid density).

Hence $P_{\Theta}(\theta|y) = \dfrac{e^{\theta} e^{-y} e^{-\theta}}{y e^{-y}} = \begin{cases} \frac{1}{y} & 0 < \theta \leq y \\ 0 & \text{otherwise} \end{cases}$

(note that this conditional distribution is uniform.)

The conditional mean is now easy to compute:

$$\boxed{\hat{\theta}_{MMSE}(y) = E[\Theta | Y = y] = \frac{y}{2}}$$

b) The LMMSE estimator can be found without any calculation since the MMSE estimator is linear in the observations. In general,

$$\hat{\theta}_{LMMSE}(y) = hy + c$$

We know that $h = \frac{1}{2}$ and $c = 0$ forms the MMSE estimator, hence

$$\boxed{\hat{\theta}_{LMMSE}(y) = \frac{y}{2}} \quad \leftarrow \text{same as } \hat{\theta}_{MMSE}(y).$$

This must be true because LMMSE is a subclass of MMSE.

c) The MAP estimator selects the value of θ that maximizes $p_{\Theta}(\theta|y)$. Since the conditional distribution is uniform, then

$$\boxed{\hat{\theta}_{MAP}(y) = z \quad \text{for any } z \in (0, y]}$$

is a valid MAP estimator.

d) The ML estimator does not use the prior (recall that ML estimation is a non-random parameter estimation technique). Instead, the ML estimator picks the value of θ that makes the observation most likely.

We know that $p_Y(y|\theta) = p_Y(y;\theta) = \begin{cases} \exp\{\theta - y\} & y \geq \theta \\ 0 & \text{otherwise.} \end{cases}$

Hence $\boxed{\hat{\theta}_{ML}(y) = y}$.

Note that the ML estimator is also a MAP estimator, but the MAP estimator is not unique. Also note that $\hat{\theta}_{ML}(y) = 2\,\hat{\theta}_{MMSE}(y)$.

2. 100 points. Suppose that θ is a scalar parameter that you wish to estimate from N i.i.d. observations received according to the marginal pdf

$$p(y_k\,;\,\theta) \quad = \quad \begin{cases} \exp(\theta - y_k) & y_k \geq \theta \\ 0 & y_k < \theta. \end{cases}$$

Given $N = 2$ i.i.d. observations, i.e. $y = \{y_0, y_1\}$, find

(a) (50 points) a *scalar* sufficient statistic $T(y) \in \mathbb{R}$ for the scalar parameter θ and

(b) (50 points) an MVU estimator for the scalar parameter θ. Given the time limits of the exam, you can assume that your scalar sufficient statistic is complete here.

Solution:

(a) Since the observations are i.i.d. the joint density of Y parameterized by the non-random parameter θ can be written as

$$\begin{aligned} p(y\,;\,\theta) &= \left\{ \prod_{k=0}^{1} \exp(\theta - y_k) \right\} \mathbb{I}_{\{y_0 \geq \theta\}} \mathbb{I}_{\{y_1 \geq \theta\}} \\ &= \exp\left(-(y_0 + y_1)\right) \exp\left(2\theta\right) \mathbb{I}_{\{\min\{y_0, y_1\} \geq \theta\}} \end{aligned}$$

where $\mathbb{I}_{\{x \geq a\}}$ is the indicator function that is equal to one when $x \geq a$ and equal to zero otherwise. The Neyman-Fisher factorization theorem tells us that $T(y) = \min\{y_0, y_1\}$ is a sufficient statistic.

(b) To find the MVU estimator, we can use the RBLS theorem since we are allowed to assume the sufficient statistic is complete. Let $Z = \min\{Y_0, Y_1\}$. It can be shown (see, e.g. the handouts from the Papoulis textboook) that the density of Z can be written as

$$p_Z(z) = \begin{cases} 2\exp\{-2(z - \theta)\} & z \geq \theta \\ 0 & \text{otherwise.} \end{cases}$$

Hence, since $\mathrm{E}_\theta\{Z\} = \theta + \frac{1}{2}$, an unbiased estimator of the parameter θ given the observation $y = \{y_0, y_1\}$ is

$$\hat{\theta}(y) \quad = \quad \min\{y_0, y_1\} - \frac{1}{2}.$$

All that remains is to take the conditional expectation to find the MVU estimator:

$$\begin{aligned} \hat{\theta}_{\mathsf{MVU}}(y) &= \mathrm{E}\{\hat{\theta}(Y)\,|\,T(Y) = t\} \\ &= \mathrm{E}\{\hat{\theta}(Y)\,|\,\min\{Y_0, Y_1\} = \min\{y_0, y_1\}\} \\ &= \mathrm{E}\{\min\{Y_0, Y_1\} - \frac{1}{2}\,|\,\min\{Y_0, Y_1\} = \min\{y_0, y_1\}\} \\ &= \min\{y_k\} - \frac{1}{2} \end{aligned}$$

This result should make intuitive sense since the support of the i.i.d. observations is lower bounded by θ. The offset of $-1/2$ is needed to make the estimator unbiased.

3. 100 points. Consider the *scalar* dynamical system

$$X[\ell+1] \quad = \quad fX[\ell] + U[\ell] \text{ for } \ell = 0, 1, \dots$$

where f is a known scalar constant, and the observation model $Y[\ell] = X[\ell] + V[\ell]$ for $\ell = 0, 1, \dots$. Assume that $\{U[0], U[1], \dots\}$ and $\{V[0], V[1], \dots\}$ are independent sequences of i.i.d. $\mathcal{N}(0, 1)$ random variables. Also assume that the initial state $X[0] \sim \mathcal{N}(0, \sigma^2)$ and is independent of all $U[\ell]$ and $V[\ell]$.

(a) (50 points). Determine a value of the initial state variance $\sigma^2 > 0$ such that the Kalman gain is a constant for all $\ell = 0, 1, \dots$, i.e.,

$$K[\ell] = \Sigma[\ell \,|\, \ell-1] H^\top[\ell] \left(H[\ell] \Sigma[\ell \,|\, \ell-1] H^\top[\ell] + R[\ell] \right)^{-1} \equiv K.$$

(b) (50 points). Find expressions for the mean-squared prediction error and the mean-squared filtering error for the case derived in Part (a). Comment on the behavior of these errors as $|f| \to 0$ and $|f| \to \infty$.

Solution:

(a) We need $K[\ell+1] = K[\ell] \equiv K$ for all ℓ. It is clear that

$$K[0] \quad = \quad \frac{\sigma^2}{\sigma^2 + 1}$$

and some straightforward calculations lead to an expression for $K[1]$ as

$$K[1] \quad = \quad \frac{(f^2+1)\sigma^2 + 1}{(f^2+2)\sigma^2 + 2}$$

Equating $K[0]$ and $K[1]$ and simplifying, we can say that $K[0] = K[1]$ if

$$\sigma^4 - f^2\sigma^2 - 1 = 0.$$

There are two roots but only one positive root for σ^2. The positive root is given as

$$\sigma^2 \quad = \quad \frac{f^2 + \sqrt{f^4+4}}{2}$$

Using this value for σ^2 causes $K[\ell] = K = \frac{\sigma^2}{\sigma^2+1}$ for all ℓ.

(b) Since K is a constant, the mean squared prediction error in this case is

$$\Sigma[\ell+1 \,|\, \ell] = \sigma^2 = \frac{f^2 + \sqrt{f^4+4}}{2}.$$

The mean squared filtering error in this case is

$$\Sigma[\ell \,|\, \ell] = K = \frac{\sigma^2}{\sigma^2+1} = \frac{f^2 + \sqrt{f^4+4}}{f^2 + \sqrt{f^4+4} + 2}.$$

In the limit as $|f| \to 0$, the prediction error goes to one and the filtering error goes to one half. In this case, we have a state that is updated only by the Gaussian input;

the state essentially has no memory, hence predictions tend to be much worse than the estimates. In the limit as $|f| \to \infty$, the prediction error goes to infinity and the filtering error goes to one. In this extreme, the prediction errors tend to be very large because the state updates are large and any error in the previous estimate is magnified by the large magnitude of f. The filtering error is much better that the prediction error here because the observation $Y[\ell]$ removes the uncertainty caused by the large state update from $X[\ell - 1]$ to $X[\ell]$, and only leaves the uncertainty caused by the observation noise.

4. 100 points. Suppose you observe a scalar random variable Y given by

$$Y = N + \theta\lambda$$

where θ is either 0 or 1, λ is a fixed (known) number between 0 and 2, and where N is a random variable that has a uniform density on the interval $(-1, 1)$. We wish to decide between the hypotheses

$$H_0 : \theta = 0$$

versus

$$H_1 : \theta = 1.$$

(a) (50 points). Find the Neyman-Pearson decision rule for false-alarm probability $\alpha \in [0, 1]$.

(b) (50 points). Find the power of the Neyman-Pearson decision rule as a function of the false-positive probability and the known parameter λ. Sketch the receiver operating characteristic for the cases $\lambda = 0$, $\lambda = 1$, and $\lambda = 2$.

Solution: See next two pages...

Problem 4

a) $Y = N + \theta\lambda$ $\qquad 0 \le \lambda \le 2$ known

$$p_\theta(y) = \begin{cases} \frac{1}{2} & -1 < y < 1 \quad \text{when } \theta = 0 \\ \frac{1}{2} & -1+\lambda < y < 1+\lambda \quad \text{when } \theta = 1 \end{cases}$$

The likelihood ratio here is

$$L(y) = \frac{p_1(y)}{p_0(y)} = \begin{cases} 0 & -1 < y \le -1+\lambda \\ 1 & -1+\lambda < y \le 1 \\ \infty & 1 < y \le 1+\lambda \end{cases}$$

monotone increasing, hence the N-P decision rule will be of the form

$$\rho(y) = \begin{cases} 1 & y > v \\ \gamma & y = v \\ 0 & y < v \end{cases}$$

We just need to find the threshold v and the randomization γ such that $P_{fp} = \alpha$.

Note that the distributions are continuous here, so γ is arbitrary. To find v, we can write

$$\int_v^1 p_0(y)\,dy = \alpha$$

$$(1-v)\frac{1}{2} = \alpha \quad \Rightarrow \quad \boxed{v = 1 - 2\alpha}$$

check: $\alpha = 0 \Rightarrow v = 1 \Rightarrow$ decide $\mathcal{H}_0$ unless we get an observation larger than 1. ✓

$\alpha = 1 \Rightarrow v = -1 \Rightarrow$ always decide $\mathcal{H}_1$ ✓

hence, the N-P decision rule is

$$\rho(y) = \begin{cases} 1 & y \geq 1-2\alpha \\ 0 & y < 1-2\alpha \end{cases}$$

(not a function of λ).

b) The power of the N-P decision rule is

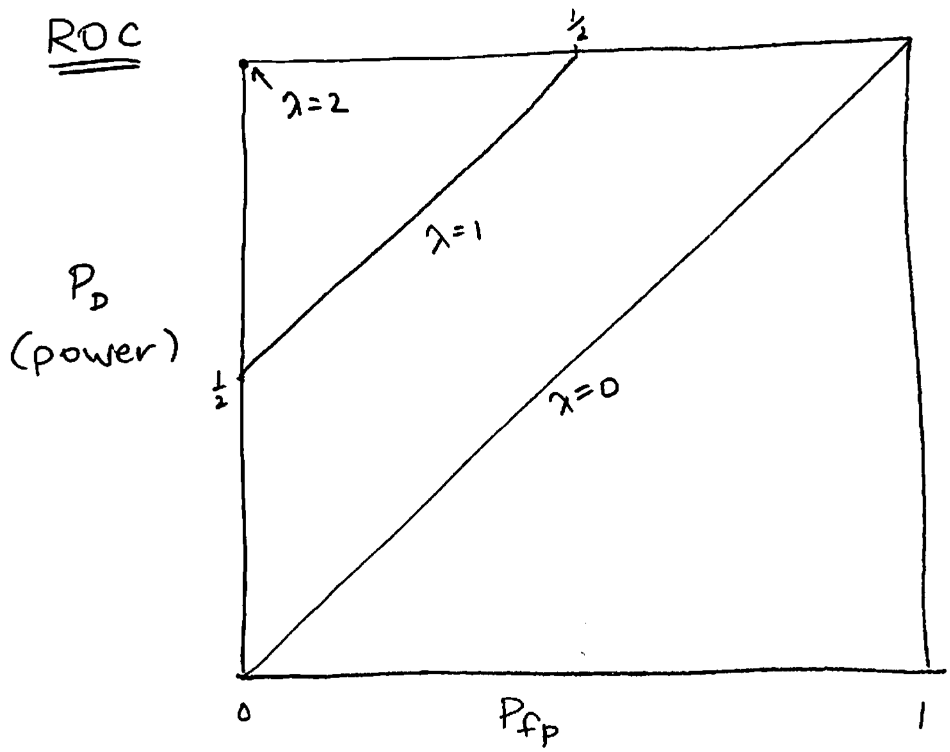
Prob(decide $\mathcal{H}_1$ | true state is x_1)

$= \text{Prob}(y \geq 1-2\alpha |$ true state is x_1)

$$= \int_{1-2\alpha}^{1+\lambda} \frac{1}{2}\, dy = \frac{1}{2}\left[(1+\lambda) - (1-2\alpha)\right] = \boxed{\frac{\lambda}{2} + \alpha}$$

Here is where λ has an effect on the N-P detector.

ROC



ROC curve.

When $\lambda = 0$, it is impossible to distinguish between hypotheses, hence $P_{fp} = P_D$. As λ increases to 2, the conditional distributions become more distinct. When $\lambda = 2$, they are so distinct that you can always make the correct decision. (end of solution)