EECS 126 Probability and Random Processes — UC Berkeley, Fall 2023
Jiantao Jiao — December 13, 2023

Final Exam

Last Name	First Name	SID

Left Neighbor First and Last Name	**Right Neighbor** First and Last Name

***Rules*.**

- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- **All work you want to be graded can be on both the front and back of the sheets in the space provided. Both sides will be scanned/graded.**
- You have 10 minutes to read the exam and 160 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $10 \cdot X\%$ time on reading and $160 \cdot X\%$ time on completing the exam).
- This exam is closed-book. You may reference two double-sided handwritten sheets of paper. No calculators or phones are allowed.
- Collaboration with others is strictly prohibited. If you are caught cheating, you may fail the course and face disciplinary consequences.

Problem	out of
Problem 1	25
Problem 2	25
Problem 3	32
Problem 4	20
Problem 5	25
Total	127

1 Two-State Machine [25 points]

Consider a machine that operates for an Exp(μ) amount of time and then fails. Once it fails, it gets repaired. The repair time is an Exp(λ) random variable and is independent of the past. The machine is as good as new after the repair is complete. Let X_t be the state of the machine at time t, 1 if it is up and 0 if it is down. This process is modelled as a continuous-time Markov chain (CTMC).

(a) Write down your SID on the top right corner to get 4 points. (4 points)

(b) Briefly explain why the rate matrix Q of the Markov chain is given by:

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

(3 points)

(c) Let $P(t) = \{p_{ij}(t)\}_{i,j\in[2]}$ denote the transition probability matrix of $X(t)$ ($p_{ij}(t) = \mathbb{P}(X(t) = j|X(0) = i)$). Given

$$P(1) = \begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix} + e^{-(\lambda+\mu)} \cdot \begin{bmatrix} \frac{\lambda}{\lambda+\mu} & -\frac{\lambda}{\lambda+\mu} \\ -\frac{\mu}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \end{bmatrix},$$

compute $P(2)$. (4 points)

(d) Determine the stationary distribution of the CTMC. Is the CTMC reversible? Justify your answer. (6 points)

(e) Suppose the downtime cost of the machine is B per unit time. What is the minimum revenue rate A during the uptime needed to break even in the long run? (4 points)

(f) Suppose the machine is working at time 0. Determine the convergence of the long-run rate of repair completions for this machine (both convergence type and value), i.e.,

$$\frac{\text{number of repairs completed before time } T}{T} \overset{?}{\to} \ ? \ (T \to \infty)$$

(4 points)

(a) SID (4 points).

(b) Omitted. (3 points).

(c) By $P(2) = P(1)P(1)$ (2 points),

$$P(2) = \begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix} + e^{-2(\lambda+\mu)} \cdot \begin{bmatrix} \frac{\lambda}{\lambda+\mu} & -\frac{\lambda}{\lambda+\mu} \\ -\frac{\mu}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \end{bmatrix} \text{ (2 points).}$$

(d) $\pi = (\mu, \lambda)/(\mu + \lambda)$ (3 points). Reversible (3 points).

(e) By big theorem, the average revenue is $A\pi_1 - B\pi_0 = \frac{\lambda A - \mu B}{\mu+\lambda}$ (3 points). To make it positive, $A > \frac{\mu B}{\lambda}$ (1 points).

(f) Let τ_i denote the time between $(i-1)$-th and i-th repair completion, then by memorylessness, τ_i are independent from each other. Notice that τ_i is sum of an $\text{Exp}(\mu)$ random variable and an $\text{Exp}(\lambda)$ random variable, we have $\mathbb{E}[t_i] = \frac{1}{\mu} + \frac{1}{\lambda}$. It follows that

$$\frac{1}{n}\sum_{i=1}^{n} \tau_i \to \frac{1}{\mu} + \frac{1}{\lambda}, \ a.s. \text{ (2 points)}$$

Further notice that if $n =$ number of repairs completed before time T, then

$$\frac{n}{\sum_{i=1}^{n+1} \tau_i} \leq \frac{\text{number of repairs completed before time } T}{T} \leq \frac{n}{\sum_{i=1}^{n} \tau_i}.$$

Therefore the rate converges almost surely (1 points) to $\frac{\lambda\mu}{\mu+\lambda}$ (1 points).

2 Graphs and Testing [25 points]

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) An Erdos-Renyi random graph $G = (V, E)$ is sampled from $\mathcal{G}(n, p)$. We observe the nodes V and the edges E. We also have a prior belief that $p \sim \mathsf{Beta}(\alpha, \beta)$. For this question, leave your answers in terms of α, β, n, and $e = |E|$ (the number of edges observed).

(i) The posterior distribution of p given the observation G is $\mathsf{Beta}(a, b)$, find a and b. (5 points)

(ii) Find the MAP estimator for p given G in terms of a and b. (4 points)

(iii) Find the MMSE estimator for p given G in terms of a and b. (3 points)

Hint: $\mathsf{Beta}(\alpha, \beta)$ *is a distribution over the interval* $[0, 1]$ *with the pdf at point* x *being* $c(\alpha, \beta)x^{\alpha-1}(1-x)^{\beta-1}$ *for some normalizing constant* $c(\alpha, \beta)$ *that depends on* α *and* β*. The mean of* $\mathsf{Beta}(\alpha, \beta)$ *is* $\frac{\alpha}{\alpha+\beta}$.

(c) Let V be a set of n nodes. An Erdos-Renyi random graph $G = (V, E)$ is sampled from $\mathcal{G}(n, p)$ with the nodes labeled as V. We observe only the edges E and not the set of nodes V. For example, suppose $n = 6$ and $V = (a, b, f, k, x, 2)$. For a random graph $G = (V, E)$, we only observe the edge set like $E = \{(a, 2), (f, a), (k, 2)\}$. We do not know V or n; by looking at E, we can conclude that there V has at least the four elements $a, 2, f, k$.

Given the set E, find the joint MLE estimate for n and p. That is, find $(\hat{n}, \hat{p}) = \operatorname{argmax}_{n,p} P(E|n, p)$. (5 points)

Hint: leave your answers in terms of $e = |E|$ *(the number of edges observed) and* m*, where* m *is the number of distinct nodes seen in the edge set* E*.*

(d) In a bin, there are four balls of color red, blue, yellow, and green. According to the null hypothesis, the probability of picking a ball Y is given as

$$H_0\text{: Y is } \begin{cases} \text{Red} & \text{w.p. } 0.1 \\ \text{Blue} & \text{w.p. } 0.2 \\ \text{Yellow} & \text{w.p. } 0.3 \\ \text{Green} & \text{w.p. } 0.4 \end{cases} \tag{1}$$

while according to the alternate hypothesis, the probability of picking a ball Y is given as

$$H_1\text{: Y is } \begin{cases} \text{Red} & \text{w.p. } 0.15 \\ \text{Blue} & \text{w.p. } 0.3 \\ \text{Yellow} & \text{w.p. } 0.5 \\ \text{Green} & \text{w.p. } 0.05. \end{cases} \tag{2}$$

Find an optimal test (and write it clearly in terms of the observation Y) that maximizes Probability of Correct Detection (PCD) subject to Probability of False Alarm (PFA) ≤ 0.5 (5 points).

(a) SID (3 points).

(b) (i) Let $f(x|\alpha,\beta) = c(\alpha,\beta)x^{\alpha-1}(1-x)^{\beta-1}$ be the pdf of $\mathsf{Beta}(\alpha,\beta)$.

$$\begin{aligned}
P(p|G) &= \frac{P(G|p)P(p)}{P(G)} \text{ (1 points)} \\
&= \frac{p^e(1-p)^{\binom{n}{2}-e} \times c(\alpha,\beta)p^{\alpha-1}(1-p)^{\beta-1}}{P(G)} \text{ (1 point)} \\
&= \frac{c(\alpha,\beta)p^{e+\alpha-1}(1-p)^{\binom{n}{2}-e+\beta-1}}{P(G)} \\
&= f\left(p|\alpha+e, \beta+\binom{n}{2}-e\right) \text{ (2 points)},
\end{aligned}$$

where the last equality follows since we know $P(p|G)$ is a valid probability distribution and is proportional to $p^{e+\alpha-1}(1-p)^{\binom{n}{2}-e+\beta-1}$ so it must be a $\mathsf{Beta}(\alpha+e, \beta+\binom{n}{2}-e)$ distribution (1 point).

(ii) $\hat{p}_{MAP}$ maximizes $P(p|G) \propto p^{e+\alpha-1}(1-p)^{\binom{n}{2}-e+\beta-1}$ (1 point). On taking the *log* and optimizing, we get $\hat{p}_{MAP} = \frac{e+\alpha-1}{\binom{n}{2}+\alpha+\beta-2}$ (3 point).

(iii) $\hat{p}_{MMSE} = \mathbb{E}[p|G]$ (1 point).
Since the posterior of p given G is $\mathsf{Beta}(\alpha+e, \beta+\binom{n}{2}-e)$, we get $E[p|G] = \frac{\alpha+e}{\alpha+\beta+\binom{n}{2}}$.
(2 point)

(c)

$$P(E|n,p) = p^e(1-p)^{\binom{n}{2}-e} I\{n \geq m\} \text{ (3 points)}.$$

For the joint optimization of n, p, we can first optimize over n to get $\hat{n} = m$. (1 point)

Next, optimizing over p leads to $\hat{p} = \frac{e}{\binom{m}{2}}$. (1 point)

(d) We see that (2 points)

$$L(Y) = \begin{cases} 5/3 & \text{Y = Yellow} \\ 3/2 & \text{Y = Red or Blue} \\ 1/8 & \text{Y = Green} \end{cases} \tag{3}$$

If we set $L(Y) \geq 5/3$ as the rule for choosing H_1, then PFA=P_0(Yellow) $= 0.3 < 0.5$. If we set $L(Y) \geq 3/2$ as the rule for choosing H_1, then PFA=P_0(Yellow or Red or Blue) $= 0.6 > 0.5$.

Observe that the rule that chooses H_1 when $Y =$ Yellow or Blue, and H_0 otherwise satisfies the constraint exactly and is optimal (3 points).

Alternate: It can be noted that the above is not a Neyman-Pearson type test. In order to get the optimal Neyman-Pearson test, consider the following rule

$$\hat{X}(y) = \begin{cases} 1 & \text{y = Yellow} \\ 1\text{w.p. } \gamma & \text{y = Red or Blue} \\ 0 & \text{y = Green} \end{cases} \tag{4}$$

Then, PFA $= P_0(\text{yellow}) + \gamma P_0(\text{red or blue}) = 0.3 + \gamma 0.3 = 0.5$. Therefore, set $\gamma = 2/3$.

In fact, it is possible to create infinite number of tests by exploiting the 'functional' equivalence of red and blue outcome.

3 Gaussian Estimation from Two Perspectives [32 points]

Let $X_1, \dots, X_n$ ($n \geq 2$) be i.i.d. samples from $\mathcal{N}(\mu, \sigma^2)$. In this problem, we consider Frequentist and Bayesian approaches to estimate μ and σ^2.

(a) Write down your SID on the top right corner to get 4 points. (4 points)

(b) **Frequentist:**

  (a) Find the maximum likelihood estimation (MLE) of μ and σ^2. (8 points)

  (b) Are the above MLE estimators unbiased? Justify your claim. (6 points)

(c) **Bayesian:**

  (a) Suppose σ^2 is known and we have prior $\mu \sim \mathcal{N}(\theta, \tau^2)$. Find the maximum a posteriori (MAP) estimator of μ. (4 points)

  (b) We introduce the following inverse-χ^2 distribution:

  The density of inverse-χ^2 distribution $\text{Inv}-\chi^2(\nu, \sigma^2)$ is given by

$$p(x|\nu, \sigma^2) = \begin{cases} \frac{(\nu\sigma^2/2)^{\nu/2}}{\Gamma(\nu/2)} x^{-\left(1+\frac{\nu}{2}\right)} e^{-\frac{\nu\sigma^2}{2x}}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

  Here $\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt$ is the gamma function. The mode and mean of $\text{Inv}-\chi^2(\nu, \sigma^2)$ are $\frac{\nu\sigma^2}{\nu+2}$ and $\frac{\nu\sigma^2}{\nu-2}$ ($\nu > 2$) respectively.

  Suppose μ is known and we have prior $\sigma^2 \sim \text{Inv}-\chi^2(\theta, \tau^2)$. Find the MAP estimator of σ^2. (4 points) *Hint: Does the posterior of σ^2 also follow inverse-χ^2 distribution?*

  (c) Are the above MAP estimators minimum mean square error (MMSE) estimators? Justify your claim. (6 points)

(a) SID (4 points).

(b) (a) The likelihood is given by

$$f(X_1, \dots, X_n | \mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right). \text{ (2 points)}$$

  Maximizing the logarithm of the above (2 points) yields $\hat{\mu}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^n X_i$ (2 points), $\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})^2$ (2 points).

  (b) By simple arithmetic, $\hat{\mu}_{\text{MLE}}$ is (3 points), $\hat{\sigma}^2_{\text{MLE}}$ is not (3 points).

(c) (a) The posterior is

$$f(\mu|X_1,\ldots,X_n,\sigma^2) = C\cdot\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{(\mu-\theta)^2}{2\tau^2}\right)$$
$$= C_1\cdot\exp\left(-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\mu^2 + \left(\frac{\sum_{i=1}^{n}X_i}{\sigma^2}+\frac{\theta}{\tau^2}\right)\mu + C_2\right)$$

where C, C_1, C_2 are independent of μ. It follows that $\hat{\mu}_{\text{MAP}} = \frac{\frac{\sum_{i=1}^{n}X_i}{\sigma^2}+\frac{\theta}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$ (4 points).

(b) By simple arithmetic

$$\sigma^2|X_1,\ldots,X_n \sim \text{Inv}-\chi^2\left(\theta+n, \frac{\theta\tau^2+\sum_{i=1}^{n}(X_i-\mu)^2}{n+\theta}\right) \text{ (2 points)}$$

Thus $\hat{\sigma}^2_{\text{MAP}} = \frac{\theta\tau^2+\sum_{i=1}^{n}(X_i-\mu)^2}{\theta+n+2}$ (2 points).

(c) Since mean and mode are identical in Gaussian distribution, $\hat{\mu}_{\text{MAP}}$ is (3 points). Since mean and mode are not identical in inverse-χ^2 distribution, $\hat{\sigma}^2_{\text{MAP}}$ is not (3 points).

4 Two Hypotheses [20 points]

Under H_0, a random variable has the cumulative distribution function $F_0(x) = x^2, 0 \le x \le 1$; and under H_1, it has the cumulative distribution function $F_1(x) = x, 0 \le x \le 1$.

(a) Write down your SID on the top right corner to get 4 points. (4 points)

(b) Let X be sampled from H_0 and Y be sampled from H_1, independently from each other. Find the linear least squares estimator (LLSE) $\mathbb{L}(X + Y|X - Y)$. (6 points)

(c) What is the Neyman-Pearson test of H_0 vs. H_1, such that the probability of false alarm (PFA) is α? (7 points) What is the probability of correct detection (PCD) of the above test? (3 points)

The density functions are $f_0(x) = 2x$ vs. $f_1(x) = 1$.

(a) SID (4 points).

(b) Given the probability density functions $f_X(x) = 2x$ and $f_Y(y) = 1$ for X and Y respectively, the expected values and variances are calculated as follows: (2 points)

$$\begin{aligned}
E[X] &= \int_0^1 x f_X(x)\,dx = \frac{2}{3},\\
E[Y] &= \int_0^1 y f_Y(y)\,dy = \frac{1}{2},\\
E[X^2] &= \int_0^1 x^2 f_X(x)\,dx = \frac{1}{2},\\
E[Y^2] &= \int_0^1 y^2 f_Y(y)\,dy = \frac{1}{3},\\
\text{var}(X) &= E[X^2] - (E[X])^2 = \frac{1}{18},\\
\text{var}(Y) &= E[Y^2] - (E[Y])^2 = \frac{1}{12}.
\end{aligned}$$

For $U = X + Y$ and $V = X - Y$, we have: (2 points)

$$\begin{aligned}
E[U] &= E[X] + E[Y] = \frac{7}{6},\\
E[V] &= E[X] - E[Y] = \frac{1}{6},\\
\text{var}(V) &= \text{var}(X) + \text{var}(Y) = \frac{5}{36},\\
\text{cov}(U, V) &= \text{var}(X) - \text{var}(Y) = -\frac{1}{36}.
\end{aligned}$$

Therefore, the coefficients for the linear least squares estimator $\mathbb{L}(U|V)$ are (2 points):

$$a = \frac{\text{cov}(U,V)}{\text{var}(V)} = -\frac{1}{5},$$
$$b = E[U] - a \cdot E[V] = \frac{6}{5}.$$

The linear least squares estimator $\mathbb{L}(U|V)$ is given by:

$$\mathbb{L}(U|V) = \frac{6}{5} - \frac{1}{5}V = \frac{6}{5} - \frac{1}{5}(X - Y).$$

(c) To determine the Neyman-Pearson (NP) test of level α, we set up the likelihood ratio:

$$\lambda(x) = \frac{f_1(x)}{f_0(x)} = \frac{1}{2x} \text{ (2 points)}$$

We reject H_0 in favor of H_1 when $\lambda(x)$ is large, or equivalently, when x is small. The rejection region is therefore $x \le c$, where c is determined by the test size α:

$$\alpha = P(X \le c | H_0) = \int_0^c 2x\,dx = c^2 \text{ (3 points)}$$

Solving for c, we find:

$$c = \sqrt{\alpha}.$$

It follows that the reject region is $\{x \le \sqrt{\alpha}\}$ (2 points).

The probability of correct detection (PCD) is the power of the test, which is the probability that the test correctly rejects H_0 when H_1 is true. The PCD is given by:

$$PCD = P(X \le c | H_1) = \int_0^c 1\,dx = c = \sqrt{\alpha} \text{ (3 points)}$$

5 Kalman Filters [25 points]

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) What color is the Pink Panther? (1 point)

(c) Consider the standard Kalman Filter state updates but with a slight change. The observations have a constant and unknown bias Ω, and no other noise. Concretely, $\forall\ i \geq 0$,

$$X_{i+1} = aX_i \tag{5}$$
$$Y_i = X_i + \Omega. \tag{6}$$

It is given that X_0 and Ω are zero-mean random variable and independent, with variances σ_X^2 and σ_Ω^2 respectively. We want to do Kalman Prediction, i.e., obtain $\hat{x}_{i|i-1} = \mathbb{L}[X_i|Y_0, \ldots, Y_{i-1}]$ and the corresponding errors $\sigma_{i|i-1}^2 = \mathbb{E}[(X_i - \hat{x}_{i|i-1})^2]$. In order to do so for this problem, we would need to keep track of three additional quantities which we define below:

$$\hat{\Omega}_i = \mathbb{L}[\Omega|Y_0, \ldots, Y_i]$$
$$\lambda_i^2 = \mathbb{E}[(\Omega - \hat{\Omega}_i)^2]$$
$$\rho_i = \mathbb{E}[(X_{i+1} - \hat{x}_{i+1|i})(\Omega - \hat{\Omega}_i)]$$

(i) With the understanding that $\hat{x}_{0|-1} = \mathbb{E}[X_0]$ and $\hat{\Omega}_{-1} = \mathbb{E}[\Omega]$, write down the expression for $\sigma_{0|-1}^2$, λ_{-1}^2, and ρ_{-1}. (3 points)

(ii) Find the Kalman update for $\hat{x}_{i+1|i}$. In other words, find the term K_i and $\hat{Y}_i$ below in terms of $a, \hat{\sigma}_{i|i-1}^2, \rho_i, \lambda_{i-1}^2, \hat{x}_{i|i-1}, \hat{\Omega}_{i-1}$ and Y_i. (6 points)

$$\hat{x}_{i+1|i} = a\hat{x}_{i|i-1} + K_i\hat{Y}_i$$

(iii) Find the Kalman update for $\hat{\Omega}_i$. In other words, find the term L_i and $\hat{Y}_i$ below in terms of $a, \hat{\sigma}_{i|i-1}^2, \rho_i, \lambda_{i-1}^2, \hat{x}_{i|i-1}, \hat{\Omega}_{i-1}$ and Y_i. (2 points)

$$\hat{\Omega}_i = \hat{\Omega}_{i-1} + L_i\hat{Y}_i$$

(iv) Find the Kalman update for $\sigma_{i+1|i}^2$. In other words, find the term α_i below in terms of $a, \sigma_{i|i-1}^2, \rho_{i-1}, \lambda_{i-1}^2$. (3 points)

$$\sigma_{i+1|i}^2 = a^2\sigma_{i|i-1}^2 - K_i\alpha_i$$

(v) Find the Kalman update for λ_i^2. In other words, find the term β_i below in terms of $a, \sigma_{i|i-1}^2, \rho_{i-1}, \lambda_{i-1}^2$. (3 points)

$$\lambda_i^2 = \lambda_{i-1}^2 - L_i\beta_i$$

(vi) Find the Kalman update for ρ_i. In other words, find the term γ_i below in terms of $a, \sigma_{i|i-1}^2, \rho_{i-1}, \lambda_{i-1}^2$. (4 points)

$$\rho_i = a\rho_{i-1} - L_i\alpha_i - K_i\beta_i + K_iL_i\gamma_i$$

(a) SID (3 points).

(b) Pink (1 point)

(c) (i) Since $\hat{x}_{0|-1} = \mathbb{E}[X_0]$, $\sigma^2_{0|-1} = \sigma^2_X$. (1 point)
Similarly, $\lambda^2_{-1} = \sigma^2_\Omega$ (1 point) and
$\rho_{-1} = \text{Cov}(X_0, \Omega) = 0$ (1 point).

(ii) We note that $\hat{x}_{i+1|i} = a\hat{x}_{i|i-1} + \frac{\text{Cov}(X_{i+1}, \hat{Y}_i)}{\text{Var}(\hat{Y}_i)}\hat{Y}_i$. (1 point).
We note that that $\hat{Y}_i = Y_i - \mathbb{L}[X_i + \Omega | Y_0, \dots, Y_{i-1}] = Y_i - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}$. (1 point)
Further,

$$\begin{aligned}
\text{Cov}(X_{i+1}, \hat{Y}_i) &= \text{Cov}(aX_i, Y_i - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}) \\
&= \text{Cov}(aX_i, Y_i - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}) \\
&= \text{Cov}(aX_i, X_i + \Omega - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}) \\
&= a\sigma^2_{i|i-1} + a\rho_{i-1}. \text{ (1 point)}
\end{aligned}$$

where we used $\text{Cov}(X_i, X_i - \hat{x}_{i|i-1} + \Omega - \hat{\Omega}_{i-1}) = \text{Cov}(X_i, X_i - \hat{x}_{i|i-1}) + \text{Cov}(X_i, \Omega - \hat{\Omega}_{i-1}) = \text{Cov}(X_i - \hat{x}_{i|i-1}, X_i - \hat{x}_{i|i-1}) + \text{Cov}(X_i - \hat{x}_{i|i-1}, \Omega - \hat{\Omega}_{i-1})$ by orthogonality (2 points).
Similarly,

$$\begin{aligned}
\text{Var}(\hat{Y}_i) &= \text{Var}(X_i + \Omega - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}) \\
&= \sigma^2_{i|i-1} + \lambda^2_{i-1} + 2\rho_{i-1}. \text{ (1 point)}
\end{aligned}$$

Therefore, $K_i = \frac{a\sigma^2_{i|i-1} + a\rho_{i-1}}{\sigma^2_{i|i-1} + \lambda^2_{i-1} + 2\rho_{i-1}}$.

(iii) We have $\hat{\Omega}_i = \hat{\Omega}_{i-1} + \frac{\text{Cov}(\Omega, \hat{Y}_i)}{\text{Var}(\hat{Y}_i)}\hat{Y}_i$. (1 point).
Simplifying,

$$\begin{aligned}
\text{Cov}(\Omega, \hat{Y}_i) &= \text{Cov}(\Omega, Y_i - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}) \\
&= \text{Cov}(\Omega, X_i + \Omega - \hat{x}_{i|i-1} - \hat{\Omega}_{i-1}) \\
&= \lambda^2_{i-1} + \rho_{i-1}. \text{ (1 point)}
\end{aligned}$$

Therefore, $L_i = \frac{\lambda^2_{i-1} + \rho_{i-1}}{\sigma^2_{i|i-1} + \lambda^2_{i-1} + 2\rho_{i-1}}$.

(iv) We have $\sigma^2_{i+1|i} = \sigma^2_{i+1|i-1} - K_i \text{Cov}(X_{i+1}, \hat{Y}_i)$. (2 point)
Thus, $\sigma^2_{i+1|i} = \sigma^2_{i+1|i-1} - aK_i(\sigma^2_{i|i-1} + \rho_{i-1})$. (1 point)
Further, $\sigma^2_{i+1|i-1} = \text{Var}(X_{i+1} - \hat{x}_{i+1|i-1}) = \text{Var}(aX_i - a\hat{x}_{i|i-1}) = a^2\sigma^2_{i|i-1}$.
$\alpha_i = a(\sigma^2_{i|i-1} + \rho_{i-1})$.

(v) We have $\lambda_i^2 = \lambda_{i-1}^2 - L_i \mathrm{Cov}(\Omega, \hat{Y}_i)$. (2 point)
Thus, $\lambda_i^2 = \lambda_{i-1}^2 - L_i(\lambda_{i-1}^2 + \rho_{i-1})$. (1 point)
$\beta_i = (\lambda_{i-1}^2 + \rho_{i-1})$.

(vi)

$$\begin{aligned}\rho_i &= \mathbb{E}[(X_{i+1} - \hat{x}_{i+1|i})(\Omega - \hat{\Omega}_i)] \\ &= \mathbb{E}[(aX_i - a\hat{x}_{i|i-1} - K_i\hat{Y}_i)(\Omega - \hat{\Omega}_{i-1} - L_i\hat{Y}_i)] \text{ (2 points)} \\ &= a\rho_{i-1} - aL_i(\sigma_{i|i-1}^2 + \rho_{i-1}) - K_i(\lambda_{i-1}^2 + \rho_{i-1}) + K_iL_i(\lambda_{i-1}^2 + \sigma_{i|i-1}^2 + 2\rho_{i-1}) \text{ (2 points).}\end{aligned}$$