EECS 126 Probability and Random Processes
Thomas Courtade

UC Berkeley, Spring 2023
May 11, 2023, 3:10-6:00pm

Final

Last Name	First Name	SID

Left Neighbor Full Name	**Right Neighbor** Full Name	**Room Number**

- **Write your SID on every page to receive 1 point.**
- **All work you want graded should be on the fronts of the sheets in the space provided. Back sides may be used for scratch work, but will not be scanned/graded.**
- **Write all answers clearly. Answers that are not legible will not receive credit.**
- Unless otherwise stated, all your answers need to be justified and your work must be shown.
- You have 170 minutes to complete the exam. (DSP students with $X\,\%$ time accommodation should spend $170 \cdot X\,\%$ time on the exam).
- You are allowed three double-sided sheets of notes. No calculators/phones.
- Remember the Berkeley Honor Code: "*As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.*" Violations may result in sanctions.

Problem	points earned	out of
SID		1
Problem 1		7
Problem 2		8
Problem 3		10
Problem 4		16
Problem 5		19
Problem 6		16
Problem 7		10
Problem 8		10
Problem 9		12
Problem 10		16
Problem 11		15
Total		140

 Student ID: _____

1 In Summer [3+4]

a) What are you looking forward to over the summer?

b) How has your perspective on probability changed after taking this course?

2 Convergence Implies Convergence? [3+3+2]

Let $(X_n)_{n\geq 1}$ be a sequence of random variables defined on a common probability space, satisfying $\lim_{n\to\infty} \mathbb{E}[|X_n|] = 0$. Say whether each of the following statements is true or false. If true, prove it. If false, give a counterexample.

a) $X_n \to 0$ in probability.

b) $X_n \to 0$ in mean-square (i.e., $\lim_{n\to\infty} \mathbb{E}[|X_n|^2] = 0$).

c) $X_n \to 0$ in distribution.

a) True. By Markov's inequality, for any $\epsilon > 0$: $P(|X_n| > \epsilon) \leq \epsilon^{-1}\,\mathbb{E}[|X_n|] \to 0$.

b) False. Let $X_n = 0$ with probability $1 - 1/n^2$ and $X_n = n$ with probability $1/n^2$. Then $\mathbb{E}\,|X_n| = 1/n \to 0$, but $\mathbb{E}\,|X_n|^2 = 1$ for all n.

c) True. Convergence in probability implies convergence in distribution, so "True" follows from part (a).

 Student ID: _____

3 Nuts and Bolts (of Probability) [5+5]

a) Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. zero-mean, unit variance random variables. In this context, precisely state the central limit theorem *and* the definition of its mode of convergence.

b) Suppose a thousand bolts are manufactured in an independent and identical fashion, and you empirically measure the diameters to have mean 1.00cm and standard deviation 0.01cm. If the manufacturing run is extended to 1 million bolts, approximately how many would you expect to find with diameters exceeding 1.05cm? You may leave your answer in terms of the standard normal cdf Φ, but explain your reasoning.

a) Define $S_n = \sum_{i=1}^n X_i$. In the given context, the CLT says $\frac{1}{\sqrt{n}} S_n \to Z$ in distribution, where $Z \sim N(0,1)$. The latter statement about convergence in distribution means

$$\lim_{n\to\infty} P(\tfrac{1}{\sqrt{n}} S_n \leq x) = \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-u^2/2} du, \quad \forall x \in \mathbb{R}.$$

b) We can invoke the CLT to approximate the probability that a bolt has diameter that exceeds the nominal value by 5 times the standard deviation as $\Phi(-5)$. Thus, the number of bolts we expect to find is $10^6 \times \Phi(-5)$. [Remark: numerically, this happens to be approximately 0.29 (i.e., less than one).]

4 Counting Triangles [3+5+8]

A *triangle* in a graph $G = (V, E)$ is a collection of three vertices $u, v, w \in V$ which are joined by three edges. Let N_Δ denote the number of triangles in $G \sim \mathcal{G}(n, p)$.

a) Compute $\mathbb{E}[N_\Delta]$.

b) Explain where each of the four terms in the following expression comes from:

$$\mathbb{E}[N_\Delta^2] = \binom{n}{3}\binom{n-3}{3}p^6 + 3\binom{n}{3}\binom{n-3}{2}p^6 + 3(n-3)\binom{n}{3}p^5 + \binom{n}{3}p^3.$$

c) Find the sharp threshold $t(n)$ (of the form $t(n) = n^{-\alpha}$ for some $\alpha > 0$) such that $p(n) \gg t(n)$ ensures that G contains a triangle with high probability, and $p(n) \ll t(n)$ ensures that G is triangle-free with high probability. Show all steps in your derivation.

(The inequality $P(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}[X])^2}$ might be helpful for part of this.)

a) Let $p = p(n)$. Using indicators, we have $\mathbb{E}[N_\Delta] = \binom{n}{3}p^3$.

b) Let $\binom{V}{3}$ denote the collection of triples of vertices, and let X_S be the indicator that $S \in \binom{V}{3}$ forms a triangle. By opening the square, we have

$$N_\Delta^2 = \sum_{S,S' \in \binom{V}{3}} X_S X_{S'}.$$

It is possible for S, S' to share all three vertices (in which case $X_S = X_S'$); or for S, S' to share one or zero vertices (in which case X_S, X_S' are independent, since they share no edges); or for S, S' to share two vertices, in which case they share an edge, and $\mathbb{E}[X_S X_{S'}] = p^5$. Taking expectations and enumerating these cases gives

$$\mathbb{E}[N_\Delta^2] = \binom{n}{3}p^3 + 3\binom{n}{3}\binom{n-3}{2}p^6 + \binom{n}{3}\binom{n-3}{3}p^6 + 3(n-3)\binom{n}{3}p^5.$$

c) First, Markov's inequality implies $P(N_\Delta \geq 1) \leq \mathbb{E}[N_\Delta] \leq n^3 p^3$. The RHS vanishes if $p = p(n) \ll n^{-1}$, so that G will be triangle-free in this case with high probability.

Using the hint, the probability the graph is triangle free is at most

$$P(N_\Delta = 0) \leq \frac{1}{\binom{n}{3}p^3} + 3\frac{\binom{n-3}{2}}{\binom{n}{3}} + \frac{\binom{n-3}{3}}{\binom{n}{3}} + 3\frac{n-3}{\binom{n}{3}p} - 1 \lesssim \frac{1}{n^3p^3} + \frac{1}{n^2p}.$$

This tends to zero if $p = p(n) \gg n^{-1}$. Hence, G will contain at least one triangle with high probability. So, $t(n) = n^{-1}$.

5 Sisyphean Chain [3+8+8]

Throughout this problem, $p_0, p_1, \ldots$ is a given sequence of numbers in the interval $[0, 1]$.

a) Let $(F_n)_{n\geq 0}$ be independent Bernoulli trials, with $F_n \sim \text{Bernoulli}(1-p_n)$ for each $n \geq 0$. Let $N = \min\{n \geq 0 : F_n = 1\}$ be the time of the first success. Compute $\mathbb{E}[N]$ in terms of the given sequence $p_0, p_1, \ldots$.

b) Let $(X_n)_{n\geq 0}$ be a Markov chain on state space $\{0, 1, 2, \ldots\}$ with nonzero transition probabilities given by

$$P_{n,0} = 1 - p_n, \quad P_{n,n+1} = p_n, \quad n \geq 0.$$

Find necessary and sufficient conditions on $p_0, p_1, \ldots$ for this chain to be irreducible and positive recurrent.

c) Assuming the conditions of part (b) hold, what is the stationary distribution for the chain $(X_n)_{n\geq 0}$ in terms of $p_0, p_1, \ldots$?

a) We can use indicators or the tail sum formula. The latter gives

$$\mathbb{E}[N] = \sum_{n\geq 0} P(N > n) = \sum_{n\geq 0} P(F_0 = 0, \ldots, F_n = 0) = \sum_{n\geq 0} \prod_{k=0}^{n} p_k.$$

b) For irreducibility, we need $p_n > 0$ for all $n \geq 0$. For positive recurrence, starting in state 0, we need expected return time to be finite. The expected return time is one plus what we computed in part (a), so necessary conditions are:

$$p_n > 0 \ \forall n \geq 0, \quad \text{and} \quad \sum_{n\geq 0} \prod_{k=0}^{n} p_k < \infty.$$

They are also sufficient. The first gives $0 \to n$, and with the second (that return time to zero is finite) implies $n \to 0$, so $0 \leftrightarrow n$ for all n; irreducibility follows. Positive recurrence (a class property) follows since the average round-trip time from state 0 is finite.

c) We know a SD exists and is unique, so we should solve $\pi P = \pi$, which reads as:

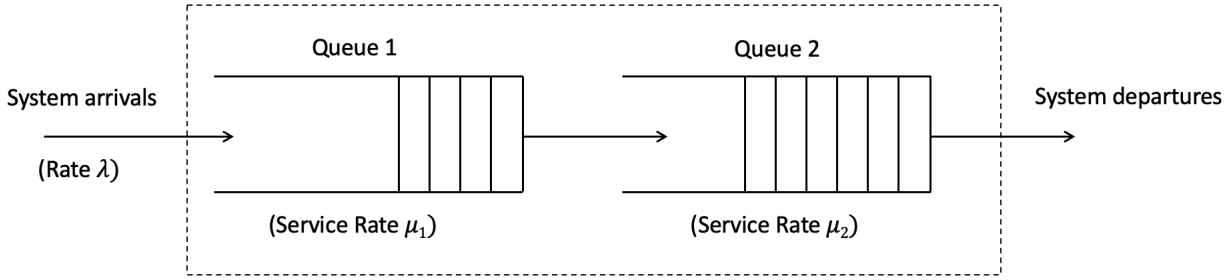
$$\sum_{n\geq 0} \pi_n(1 - p_n) = \pi_0, \quad \text{and} \quad \pi_n p_n = \pi_{n+1} \ \ \forall n \geq 0.$$

Note that iterating the second implies

$$\pi_n p_n = \pi_{n-1} p_{n-1} p_n = \cdots = \pi_0 \prod_{k=0}^{n} p_k \quad \Rightarrow \quad \pi_n = \pi_0 \prod_{k=0}^{n-1} p_k.$$

Now, π_0 can be computed as the inverse of the expected return time: $\pi_0 = \frac{1}{1+\sum_{n\geq 0} \prod_{k=0}^{n} p_k}$.

6 Waiting in Line, Again [6+4+6]



Consider a system of two M/M/1 queues connected as illustrated above. The arrival rate to the combined system is assumed to be a Poisson process of rate λ, and the service times in the first and second queues are i.i.d. Exp(μ_1) and Exp(μ_2), respectively. Assume that $0 < \lambda < \min\{\mu_1, \mu_2\}$.

Let $N_{i,t}$ be the number of customers in queue $i \in \{1, 2\}$ at time $t \geq 0$. The state of the system at time $t \geq 0$ is represented by the pair $X_t := (N_{1,t}, N_{2,t})$.

a) Model the process $(X_t)_{t\geq 0}$ as a CTMC. In particular, draw a state transition diagram with clearly labeled transition rates between states.

b) Let $q_{(m,n),(m',n')}$ denote the transition rate from state (m, n) to (m', n'). Argue that if a distribution π satisfies

$$\pi_{(m,n)} q_{(m,n),(m+1,n)} = \pi_{(m+1,n)} q_{(m+1,n),(m,n+1)} = \pi_{(m,n+1)} q_{(m,n+1),(m,n)} \quad \forall m, n \geq 0,$$

then it is a stationary distribution.

(Hint: With the help of your diagram from part (a), you can answer this question without doing any math.)

c) Compute the stationary distribution π, and formulate a simple probabilistic model for the system in steady state.

(Hint: The stationary distribution for this chain takes the form $\pi_{(m,n)} = f(m)g(n)$ for some functions f, g.)

a) The diagram should look like a grid on state space $\{0, 1, 2, \dots\}^2$ with one diagonal arrow in each square. Assuming queue 1's state is on the x-axis, the vertical, horizontal and diagonal edges are oriented as $\downarrow$, $\rightarrow$, and $\nwarrow$, respectively. The transition rates are

$$q_{(m,n),(m+1,n)} = \lambda; \qquad q_{(m+1,n),(m,n+1)} = \mu_1; \qquad q_{(m,n+1),(m,n)} = \mu_2, \quad m, n \geq 0.$$

b) The given equations say that, under π, the "flow" is conserved within the triangles in the state transition diagram shaped like $\downarrow\!\nwarrow\!\rightarrow$.

This local conservation of flow implies the global balance equation $\pi Q = 0$, which expresses that total flow into each state is equal to total flow out.

c) The product structure means the number of people in each queue will be independent

under π, with f (resp. g) the stationary distribution of queue 1 (resp. queue 2). The first queue is oblivious to the second queue, so we know its stationary distribution from MT2 is Geom$(1-\lambda/\mu_1)$, supported on $\{0,1,\dots\}$. This tells us $f(m) = (1-\lambda/\mu_1)(\lambda/\mu_1)^m$. Plugging into the balance condition given in part b) says we should solve for g satisfying

$$\begin{aligned}(1-\lambda/\mu_1)(\lambda/\mu_1)^m g(n)\lambda &= (1-\lambda/\mu_1)(\lambda/\mu_1)^{m+1} g(n)\mu_1 \\ &= (1-\lambda/\mu_1)(\lambda/\mu_1)^m g(n+1)\mu_2 \quad \forall m,n \geq 0.\end{aligned}$$

Dividing through by common factors reveals

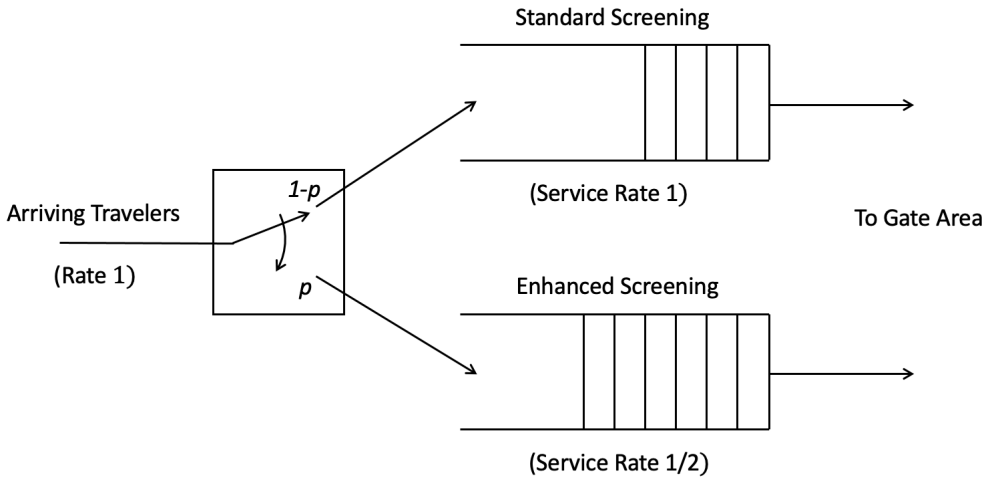
$$g(n+1) = \frac{\mu_2}{\lambda} g(n) = \cdots = \left(\frac{\mu_2}{\lambda}\right)^{n+1} g(0).$$

Thus, $g(n) = \left(1-\frac{\mu_2}{\lambda}\right)\left(\frac{\mu_2}{\lambda}\right)^n$. In particular, we conclude:

$$\pi_{(m,n)} = \left(1-\frac{\mu_1}{\lambda}\right)\left(\frac{\mu_1}{\lambda}\right)^m \times \left(1-\frac{\mu_2}{\lambda}\right)\left(\frac{\mu_2}{\lambda}\right)^n.$$

Therefore, in steady state, the number of individuals in queue $i \in \{1,2\}$ can be modeled as independent Geom$(1-\lambda/\mu_i)$ random variables!

7 TSA Checkpoint [10]



Airport security can be modeled as two parallel M/M/1 queues, with the first queue corresponding to standard screening, and the second queue corresponding to enhanced screening. The arrival rate of travelers to the checkpoint is 1/min, the average service time for standard screening is 1 min, and the average service time for enhanced screening is 2 min. (As always, *service time* only counts the time "in service", and not time spent waiting in the screening queue.)

Security officers direct travelers to enhanced screening with probability p, independent of all other travelers. What value of $p \in [0, 1]$ minimizes the expected time a customer spends (waiting and in service) in airport security? Assume the system is in steady state for your computation.

Recall that the expected time spent in a stationary M/M/1 queue with arrival rate λ and service rate μ is equal to $1/(\mu - \lambda)$ (you showed this on MT2, and in the subsequent HW where you reworked the problems).

The switch thins the arrival process, so that the travelers arriving to standard screening (resp. enhanced screening) form a Poisson process with rate $1 - p$ (resp. p). Thus, the average time a traveler spends in the standard screening queue will be $\frac{1}{1-(1-p)} = \frac{1}{p}$, and the average time a traveler spends in the enhanced screening queue will be $\frac{1}{1/2-p} = \frac{2}{1-2p}$ if $p < 1/2$ and $+\infty$ if $p \geq 1/2$. However, a fraction $1 - p$ (resp. p) customers experiences standard (resp. enhanced) screening, so the total average time spent in security is

$$\frac{1-p}{p} + \frac{2p}{1-2p} = \frac{1}{p} + \frac{1}{1-2p} - 2, \quad 0 \leq p < 1/2.$$

If $p \geq 1/2$, then the average time is infinite, so we can restrict attention to $p < 1/2$. Differentiating with respect to p and setting the derivative equal to zero implies the optimal p should satisfy

$$2\frac{1}{(1-2p)^2} = \frac{1}{p^2} \quad \Rightarrow \quad 1 - 2p = \sqrt{2}p \quad \Rightarrow p^* = \frac{1}{2+\sqrt{2}}.$$

8 MAP Estimation [10]

Let X have density

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Conditioned on $\{X = x\}$, you observe $Y \sim \text{Geom}(x)$ (supported on $\{1, 2, \dots\}$). What is the MAP estimate of X given observation $Y = k$?

We should compute

$$\hat{X}_{MAP} = \arg\max_x p_{Y|X=x}(Y) f_X(x) = \arg\max_x x(1-x)^{Y-1} 2x = \arg\max_x x^2(1-x)^{Y-1}.$$

Differentiating the objective with respect to x and setting equal to zero gives

$$2x(1-x)^{Y-1} = (Y-1)x^2(1-x)^{Y-2} \quad \Rightarrow \quad 2(1-x) = (Y-1)x \Rightarrow x = \frac{2}{Y+1}.$$

Hence, the MAP estimate given observation $Y = k$ is equal to $2/(k+1)$.

9 Rectangle or Triangle? [2+8+2]

Consider two densities f_0, f_1 as defined below:

$$f_0(y) = \begin{cases} \frac{1}{2} & -1 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}, \quad \text{and} \quad f_1(y) = \begin{cases} 1-|y| & -1 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

You observe Y, and would like to discriminate between the hypotheses:

$$H_0 : Y \sim f_0, \quad \text{and} \quad H_1 : Y \sim f_1.$$

a) What is likelihood ratio $L(y)$ for this hypothesis testing problem?

b) What is the Type II error rate of the optimal test, subject to Type I error rate at most α?

c) Sketch the error curve, and place an 'X' corresponding to the performance of the MLE test. Clearly label the x-y coordinates of your 'X'.

a) By definition,

$$L(y) = \frac{f_1(y)}{f_0(y)} = 2(1-|y|), \quad |y| \le 1.$$

b) A threshold test with threshold η achieves Type I error rate:

$$\alpha(\eta) = P_{H_0}\{L(Y) \ge \eta\} = P_{H_0}\{|Y| \le 1-\eta/2\} = 2\int_0^{1-\eta/2} \frac{1}{2} dy = 1-\eta/2.$$

The Type II error rate for the same test is:

$$\beta(\eta) = P_{H_1}(L(Y) < \eta) = P_{H_1}\{|Y| > 1-\eta/2\} = 2\int_{1-\eta/2}^{1} (1-y)dy = 2\int_0^{\eta/2} u du = \eta^2/4.$$

Since threshold tests are optimal, we can set η to achieve our desired Type I error rate, and the resulting β will be the Type II error rate of the most powerful test. I.e., this gives

$$\beta(\alpha) = 1-\alpha^2.$$

c) The error curve is just the plot of $\beta(\alpha)$ above as a function of $\alpha \in [0,1]$. The ML test is the threshold test with $\eta = 1$, so our 'X' will be at $(\alpha, \beta) = (1/2, 1/4)$.

10 Gaussian Parameters and Estimation [6+10]

The following subparts are independent of one another.

a) Let (X, Y) be jointly Gaussian vectors, with respective marginal means μ_X, μ_Y, marginal covariance matrices Σ_X, Σ_Y, and covariance $\Sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)^T]$. For (constant) matrices A, B of compatible dimensions and a vector ζ, determine the distribution of

$$Z = AX + BY + \zeta.$$

b) Suppose you observe i.i.d. $X_1, X_2, \ldots X_n$, which are assumed to be $N(\mu, \Sigma)$ with (μ, Σ) unknown. What is the maximum likelihood estimate of the pair (μ, Σ) given the observations $X_1, \ldots, X_n$?

(You may may assume μ, Σ are scalars for full credit. However, the multidimensional case is no more difficult if you are familiar with basic multivariable calculus. In particular, for a $d \times d$ positive definite matrix A and a vector $x \in \mathbb{R}^d$, you may freely use the following gradient expressions: $\nabla_A \log\det(A) = A^{-1}$. For $f(A, x) := x^T A x$, we have the partial derivatives $\nabla_A f(A, x) = xx^T$ and $\nabla_x f(A, x) = 2Ax$.)

a) Since Z is an affine transformation of the Gaussian vector (X, Y), it is itself Gaussian. Thus, we only need to determine its mean and covariance, since these parametrize the distribution. By linearity of expectation,

$$\mu_Z = A\mu_X + B\mu_Y + \zeta.$$

Likewise,

$$\begin{aligned}\Sigma_Z &= \mathbb{E}[(A(X - \mu_X) + B(Y - \mu_Y))(A(X - \mu_X) + B(Y - \mu_Y))^T] \\ &= A\Sigma_X A^T + B\Sigma_Y B^T + A\Sigma_{XY} B^T + B\Sigma_{XY}^T A^T.\end{aligned}$$

b) Let d denote the dimension of Σ. The likelihood of the observations is given by

$$\begin{aligned}&\prod_{i=1}^n \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(X_i - \mu)^T\Sigma^{-1}(X_i - \mu)\right) \\ &= \frac{1}{(2\pi)^{nd/2}\det(\Sigma)^{n/2}} \exp\left(-\sum_{i=1}^n \frac{1}{2}(X_i - \mu)^T\Sigma^{-1}(X_i - \mu)\right).\end{aligned}$$

Maximizing the likelihood is equivalent to minimizing the negative log-likelihood times $1/n$, meaning we should solve

$$\arg\min_{\mu,\Sigma} \frac{1}{n}\sum_{i=1}^n \frac{1}{2}(X_i - \mu)^T\Sigma^{-1}(X_i - \mu) - \frac{1}{2}\log\det\left(\Sigma^{-1}\right).$$

Taking the gradient of the objective with respect to Σ^{-1} and setting it equal to zero gives

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^T$$

Likewise, taking the gradient of the objective with respect to μ and setting it equal to zero gives

$$\Sigma^{-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) = 0.$$

This has solution $\mu = \sum_{i=1}^{n} X_i$. Note: you might recognize these quantities as what appears in PCA.

11 Hacking the Kalman Filter [5+10]

Consider the scalar state-space and observation model

$$X_n = aX_{n-1} + V_n$$
$$Y_n = X_n + W_n, \quad n \geq 1,$$

with $X_0 = 0, \text{Var}(V_n) = \sigma_V^2, \text{Var}(W_n) = \sigma_W^2$, and the usual assumption of uncorrelated, zero-mean noise processes. Assume $a \neq 0$.

Suppose someone has already implemented a Kalman filter for you. That is, at iteration n, you have the quantities Y_n, $(\hat{X}_{n|n}, \sigma^2_{n|n})$, and $(\hat{X}_{n-1|n-1}, \sigma^2_{n-1|n-1})$ available to you.

a) What extra updates should you add to do one-step prediction? I.e., what equations should you add to the Kalman filter to compute $\hat{X}_{n+1|n}$ on iteration n?

b) What extra updates should you add to do one-step smoothing? I.e., what equations should you add to the Kalman filter to compute $\hat{X}_{n-1|n}$ on iteration n?

Answers should be in terms of Y_n, $(\hat{X}_{n|n}, \sigma^2_{n|n})$, and $(\hat{X}_{n-1|n-1}, \sigma^2_{n-1|n-1})$, and any parameters of the state-space model. You may use $\text{Var}(\tilde{Y}_n) = a^2\sigma^2_{n-1|n-1} + \sigma_V^2$, from lecture.

a) Observe that

$$\hat{X}_{n+1|n} = \mathbb{L}[aX_n + V_{n+1}|Y^n] = a\mathbb{L}[X_n|Y^n] = a\hat{X}_{n|n}.$$

So, we just need to add the equation $\hat{X}_{n+1|n} = a\hat{X}_{n|n}$ inside the Kalman loop to do one-step prediction.

b) For one-step smoothing, observe that

$$\hat{X}_{n|n} = \mathbb{L}[aX_{n-1} + V_n|Y^n] = a\hat{X}_{n-1|n} + \mathbb{L}[V_n|Y^n] = a\hat{X}_{n-1|n} + \mathbb{L}[V_n|\tilde{Y}^n].$$

Now, we need to do some simplification. In particular, note that V_n is uncorrelated with each $Y_1, \ldots, Y_{n-1}$, and is therefore uncorrelated with each $\tilde{Y}_1, \ldots, \tilde{Y}_{n-1}$. Thus, only $\tilde{Y}_n$ is useful for estimating V_n; i.e.,

$$\mathbb{L}[V_n|\tilde{Y}^n] = \mathbb{L}[V_n|\tilde{Y}^{n-1}] + \mathbb{L}[V_n|\tilde{Y}_n] = 0 + \frac{\text{Cov}(V_n, \tilde{Y}_n)}{\text{Var}(\tilde{Y}_n)}\tilde{Y}_n = \frac{\text{Cov}(V_n, \tilde{Y}_n)}{a^2\sigma^2_{n-1|n-1} + \sigma_V^2}\tilde{Y}_n$$

As we did in the derivation of the KF, we have

$$\tilde{Y}_n = Y_n - \mathbb{L}[Y_n|Y^{n-1}] = Y_n - a\hat{X}_{n-1|n-1}.$$

So,

$$\begin{aligned}\text{Cov}(V_n, \tilde{Y}_n) &= \text{Cov}(V_n, Y_n) - a\,\text{Cov}(V_n, \hat{X}_{n-1|n-1}) \\ &= \text{Cov}(V_n, X_n + V_n) - a\,\text{Cov}(V_n, \hat{X}_{n-1|n-1}) = \sigma_V^2,\end{aligned}$$

where we used in the last step that V_n is uncorrelated with X_n and $\hat{X}_{n-1|n-1}$ (the latter since V_n is uncorrelated with all $Y_1, \ldots, Y_{n-1}$). Thus, we conclude

$$\hat{X}_{n|n} = a\hat{X}_{n-1|n} + \frac{\sigma_V^2}{a^2\sigma_{n-1|n-1}^2 + \sigma_V^2}(Y_n - a\hat{X}_{n-1|n-1}).$$

Rearrange to obtain $\hat{X}_{n-1|n}$ in terms of the desired quantities:

$$\hat{X}_{n-1|n} = a^{-1}\hat{X}_{n|n} - \frac{\sigma_V^2}{a^2\sigma_{n-1|n-1}^2 + \sigma_V^2}(a^{-1}Y_n - \hat{X}_{n-1|n-1}).$$