# Final

| Last Name | First Name | SID |
|---|---|---|
|  |  |  |

- You have 10 minutes to read the exam and 150 minutes to complete this exam.

- The maximum you can score is 130, but 100 points is considered perfect.

- The exam is not open book, but you are allowed to consult the cheat sheet that we provide. No calculators or phones. No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.

- A correct answer without justification will receive little, if any, credit.

- Take into account the points that may be earned for each problem when splitting your time between the problems.

| Problem | points earned | out of |
|---|---|---|
| Problem 1 |  | 25 |
| Problem 2 |  | 20 |
| Problem 3 |  | 20 |
| Problem 4 |  | 15 |
| Problem 5 |  | 25 |
| Problem 6 |  | 10 |
| Problem 7 |  | 15 |
| Total |  | 100 (+30) |

**Problem 1: Answer these questions briefly but clearly. [25]**

**(a) Random Dog Walk [5]**

A dog walks on the integers, possibly reversing direction at each step with probability $p = 0.1$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:

$$(x_0, x_1, x_2, x_3, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, \ldots).$$

**True / False:** The sequence of random variables $(X_n)_{n \geq 0}$ is a Markov chain.

$$\mathbb{P}(X_{n+2} = k + 2 \mid X_{n+1} = k + 1, X_n = k) = 0.9 \neq$$
$$\mathbb{P}(X_{n+2} = k + 2 \mid X_{n+1} = k + 1, X_n = k + 2) = 0.1.$$

**(b) MCMC [5]**

Let $\{0, 1\}$ be the state space of a Markov chain. Suppose that we try to simulate from the distribution $\pi = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ via Metropolis-Hastings using the proposal distribution:

$$p(x, y) = \begin{cases} \dfrac{1}{3}, & y = x \\ \dfrac{2}{3}, & y = 1 - x \end{cases}, \qquad x = 0, 1.$$

In other words, at state $x$ we propose state $y$ with probability $p(x, y)$. What should we set the acceptance probabilities $A(x, y)$ to be in order to make the algorithm correct (i.e., the stationary distribution of the chain is $\pi$)?

In fact we can accept every proposal and the stationary distribution will correctly be $\pi$.

**(c) Book Shop [5]**

A small book shop has room for at most two customers. Potential customers arrive at a Poisson rate of ten customers per hour; they enter if there is room and are turned away, never to return, otherwise. The bookie serves the admitted customers in order, requiring an exponentially distributed time of mean four minutes per customer.

Suppose there are two customers in the store. Find the expected amount of time until the store is empty again.
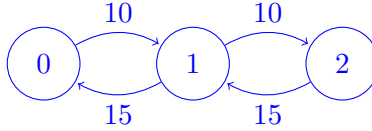
Set up the hitting time equations.

$$\mathbb{E}_2[T_0] = \frac{1}{15} + \mathbb{E}_1[T_0],$$

$$\mathbb{E}_1[T_0] = \frac{1}{25} + \frac{2}{5}\mathbb{E}_2[T_0].$$

Solving these equations yields $\mathbb{E}_2[T_0] = 8/45$.

### (d) Random Graphs [5]

Let $G$ be a random graph generated using the $\mathcal{G}(n,p)$ model. For what value of $p$ is the expected number of cliques of five vertices in $G$ equal to 1? (A *clique* on $n$ vertices is a subset of $n$ vertices such that for every pair of vertices in the clique, the edge between the vertices exists in the graph.)

For a particular subset of five vertices, the probability that every possible edge exists is $p^{\binom{5}{2}} = p^{10}$. There are $\binom{n}{5}$ subsets of five vertices. Thus, the expected number of cliques, by linearity of expectation, is $\binom{n}{5}p^{10}$. Thus, we should take $p = \binom{n}{5}^{-1/10}$.

### (e) Graphical MMSE [5]

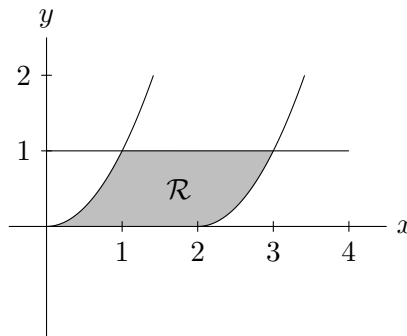Let $(X, Y)$ be a pair of random variables which are uniformly distributed over the region shown in Figure 1.



Figure 1: The region $\mathcal{R}$ is enclosed by the curves: $y = 0$, $y = 1$, $y = x^2$, $y = (x-2)^2$.

Find $\mathbb{E}[X \mid Y]$.

The MMSE is the curve $y = (x-1)^2$, which connects the horizontal midpoints of the shaded region, hence

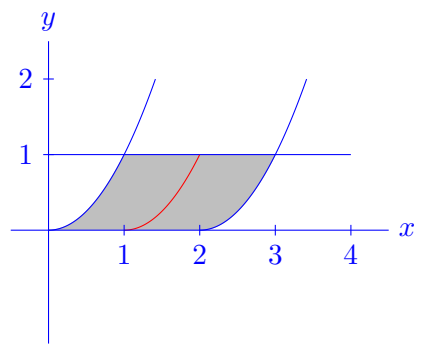$$\mathbb{E}[X \mid Y] = \sqrt{Y} + 1.$$

3

Figure 2: The MMSE is shown in red.

**Problem 2: Waiting in Line for Infinity War (no spoilers) [20]**

At 6:30 PM, there are 9 people waiting in line outside of the theater to watch *Avengers: Infinity War*. People arrive and wait in line according to a Poisson process with rate 2 per minute.

**(a) [5]** Starting from 6:30 PM, the theater admits exactly one person every minute (i.e., at 6:31 PM, 6:32 PM, 6:33 PM, etc.). What is the expected number of people remaining in line the instant after 6:40 PM?

Let $N_{10}$ represent the number of people who have arrived between 6:30 PM and 6:40 PM. Then, we are looking at $\mathbb{E}[\max\{N_{10} - 1, 0\}]$. With probability $\exp(-20)$, no one arrives between 6:30 PM and 6:40 PM. So, since $N_{10} \sim \mathrm{Poisson}(20)$,

$$\mathbb{E}[\max\{N_{10} - 1, 0\}] = \sum_{k=1}^{\infty} \frac{\exp(-20)20^k}{k!}(k-1) = \sum_{k=1}^{\infty} \frac{\exp(-20)20^k}{k!}k - \sum_{k=1}^{\infty} \frac{\exp(-20)20^k}{k!}$$
$$= 20 - [1 - \exp(-20)].$$

**(b) [5]** Suppose we observe the time $Y$ at which the third new person arrives in line. What is the MLE estimate for the time $X$ at which the fourth new person arrives in line? (Assume that $X$ and $Y$ are measured in minutes since 6:30 PM.)

Given $X = x$, we know that $Y$ is the maximum of three i.i.d. Uniform$[0, x]$ random variables. Thus, for $y \in [0, x]$, $\mathbb{P}(Y \le y \mid X = x) = (y/x)^3$, and so $f_{Y|X}(y \mid x) = 3y^2/x^3$. Thus, the MLE estimate $\hat{X}$ for $X$ given $Y$ is $\arg\max_{x>0}(3y^2/x^3)\mathbb{1}\{x \ge Y\} = Y$.

**(c) [5]** Starting from 6:40 PM, the theater stops letting in more people. The people in line become fed up: each person in line, independently, waits an Exponential(1) amount of time and then leaves the line. (People are still arriving in line according to the Poisson process of rate 2 per minute. If a new person arrives, he or she also waits an Exponential(1) amount of time before leaving the line.) After a long period of time, what is the average number of people in line?

The number of people in line is an $M/M/\infty$ queue with forward rate 2 and backwards rate 1; the stationary distribution is Poisson with mean 2.

**(d) [5]** Meanwhile, inside the theater, the movie has already begun. Thanos appears on the big screen according to a Poisson process of rate 1 per minute. On each of his appearances, independently with probability 1/3 he makes the audience scream. An observer outside the theater observes $S$, the number of times that the audience screams in the first 30 minutes. What is the LLSE estimate of $T$, the number of times that Thanos appeared on the big screen in the first 30 minutes, given $S$?

By the properties of the Poisson process, $T \sim \mathrm{Poisson}(30)$ and $S \sim \mathrm{Poisson}(10)$. Calculate $\mathbb{E}[ST] = \mathbb{E}[\mathbb{E}(ST \mid T)] = \mathbb{E}[T\,\mathbb{E}(S \mid T)] = \mathbb{E}[T^2]/3 = 930/3 = 310$. Also, $\mathbb{E}[T] = 30$ and

$\mathbb{E}[S] = 10$, so $\text{cov}(S, T) = 310 - 300 = 10$. Also, $\text{var } S = 10$, so

$$L[T \mid S] = 30 + (S - 10) = 20 + S.$$

**Problem 3: Hypothesis Testing [20]**

**(a) [10]** We want to test two hypotheses; we have prior knowledge that $X \sim \text{Bernoulli}(2/3)$. In both cases, our observation $Y$ has the Laplace distribution, but with different mean and shape. In particular

$$f_{Y|X}(y \mid 0) = \frac{1}{2}e^{-|y|}, \ y \in \mathbb{R}$$

$$f_{Y|X}(y \mid 1) = \frac{1}{4}e^{-\frac{|y-2|}{2}}, \ y \in \mathbb{R}.$$

Construct a decision rule $r : \mathbb{R} \to \{0, 1\}$ that minimizes $\mathbb{P}(r(Y) \neq X) = \mathbb{E}[I\{r(Y) \neq X\}]$ (use the MAP rule).

$$\frac{\mathbb{P}(X = 0 \mid Y)}{\mathbb{P}(X = 1 \mid Y)} = \frac{f_{Y|X}(Y \mid 0)\mathbb{P}(X = 0)}{f_{Y|X}(Y \mid 1)\mathbb{P}(X = 1)} = \frac{\frac{1}{3} \cdot \frac{1}{2}e^{-|Y|}}{\frac{2}{3} \cdot \frac{1}{4}e^{-|Y-2|/2}} = e^{|Y-2|/2-|Y|}$$

Therefore,

$$\mathbb{P}(X = 0 \mid Y) \leq \mathbb{P}(X = 1 \mid Y) \iff e^{|Y-2|/2-|Y|} \leq 1 \iff \frac{|Y - 2|}{2} \leq |Y|$$

$$\iff Y \leq -2 \text{ or } Y \geq \frac{2}{3}.$$

Thus the optimal decision rule is:

$$r(Y) = \begin{cases} 1 & \text{if } Y \leq -2 \\ 0 & \text{if } -2 < Y < \frac{2}{3} \\ 1 & \text{if } Y \geq \frac{2}{3} \end{cases}.$$

Full points were only given for simplified decision rules.

**(b) [10]** We consider a standard hypothesis testing problem in which we have

$$Y \mid X = 0 = \begin{cases} a, & \text{w.p. } \frac{1}{4} \\ b, & \text{w.p. } \frac{1}{4} \\ c, & \text{w.p. } \frac{1}{2}, \end{cases} \quad \text{and} \quad Y \mid X = 1 = \begin{cases} a, & \text{w.p. } \frac{1}{2} \\ b, & \text{w.p. } \frac{1}{4} \\ c, & \text{w.p. } \frac{1}{4}. \end{cases}$$

Design a randomized decision rule $r : \{a, b, c\} \to \{0, 1\}$ in order to minimize $\mathbb{P}(r(Y) = 0 \mid X = 1)$ subject to $\mathbb{P}(r(Y) = 1 \mid X = 0) \leq 0.3$.

The likelihood ratio is

$$L(a) = 2,$$
$$L(b) = 1,$$

7

$$L(c) = \frac{1}{2}.$$

Looking at the likelihood ratio, we can see that $L(a) > L(b) > L(c)$. We solve for the optimal decision rule using neyman pearson, finding that $\lambda = 1$. Thus, if $Y = a$ we should always raise the alarm. If we always raise the alarm when $Y = b$, then we would exceed the constraint on the PFA, so we set the threshold at $L(Y) = 1$. The optimal randomized decision rule, by the Neyman-Pearson Lemma, is

$$r(Y) = \begin{cases} 1 & \text{if } Y = a \\ \gamma & \text{if } Y = b \\ 0 & \text{if } Y = c \end{cases}$$

where the randomization $\gamma$ is found by $\mathbb{P}(Y = a \mid X = 0) + \gamma \mathbb{P}(Y = b \mid X = 0) = 0.3$, or $\gamma = 0.2$. Full points were only given for properly justified answers.

**Problem 4: Estimating the Slope of a Line [15]**

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d., where $X_1 \sim \mathcal{N}(0, 1)$ and $Y_1 \mid X_1 = x \sim \mathcal{N}(x\theta, 1)$ where $\theta \in \mathbb{R}$ is unknown.

**(a) [8]** Compute the MLE of $\theta$ given the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Given $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}$, the joint density of the observations is

$$p\big((x_1, y_1), \ldots, (x_n, y_n); \theta\big) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x_i^2}{2}\Big) \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{(y_i - x_i\theta)^2}{2}\Big)$$

so the log-likelihood is

$$\ell\big(\theta; (X_1, Y_1), \ldots, (X_n, Y_n)\big) = \text{constant} - \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\theta)^2.$$

Differentiating w.r.t. $\theta$, we have $\sum_{i=1}^{n} X_i(Y_i - X_i\theta) = 0$ which gives

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{n^{-1}\sum_{i=1}^{n} X_i Y_i}{n^{-1}\sum_{i=1}^{n} X_i^2}.$$

Check that this is indeed a global maximum. Does this look familiar? The MLE is exactly the empirical covariance divided by the empirical variance of $X$, which should remind you of the formula for the slope of the LLSE.

**(b) [7]** Compute $\mathbb{E}[X_1 \mid Y_1]$.

First, observe that $(X_1, Y_1)$ is joint Gaussian so $\mathbb{E}[X_1 \mid Y_1] = L[X_1 \mid Y_1]$. Then, $X_1$ is zero mean and $\mathbb{E}[Y_1] = \mathbb{E}[\mathbb{E}[Y_1 \mid X_1]] = \mathbb{E}[\theta X_1] = 0$ as well. So, calculating

$$\text{cov}(X_1, Y_1) = \mathbb{E}[X_1 Y_1] = \mathbb{E}\big[\mathbb{E}[X_1 Y_1 \mid X_1]\big] = \mathbb{E}\big[X_1 \mathbb{E}[Y_1 \mid X_1]\big] = \mathbb{E}[\theta X_1^2] = \theta$$

and $\text{var } Y_1 = \mathbb{E}[Y_1^2] = \mathbb{E}[\mathbb{E}[Y_1^2 \mid X_1]] = \mathbb{E}[\theta^2 X_1^2 + 1] = \theta^2 + 1$, we then have
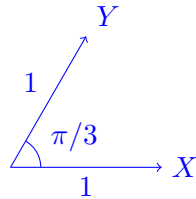
$$\mathbb{E}[X_1 \mid Y_1] = \frac{\theta}{\theta^2 + 1} Y_1.$$

**Problem 5: Estimating Jointly Gaussian Random Variables [25]**

Consider the Gaussian random vector

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right).$$

**(a) [4]** Draw on the plane the random variables as vectors in the Hilbert space of random variables. Make sure to note the length of the vectors, as well as the angle that they form.



**(b) [4]** Find $\mathbb{E}[X \mid Y]$.

$$\mathbb{E}[X \mid Y] = L[X \mid Y] = \frac{\text{cov}(X, Y)}{\text{var}\, Y} Y = \frac{Y}{2}.$$

**(c) [5]** Create a random variable $Z$ which is a function of $X$ and $Y$ such that

- $Z$ and $Y$ are independent.

- The vector $\begin{bmatrix} Z \\ Y \end{bmatrix}$ has exactly the same information content as the vector $\begin{bmatrix} X \\ Y \end{bmatrix}$, in the sense that you can recover the latter vector from the former.

Show that your $Z$ satisfies these properties.

$$Z = X - \mathbb{E}[X \mid Y] = X - L[X \mid Y] = X - \frac{Y}{2}.$$

By construction $\mathbb{E}[ZY] = 0$, and because $Z, Y$ are jointly Gaussian this implies that $Z$ and $Y$ are independent.

In addition from the vector $\begin{bmatrix} Z \\ Y \end{bmatrix}$ we can restore $\begin{bmatrix} X \\ Y \end{bmatrix}$ as follows

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Z \\ Y \end{bmatrix}.$$

**(d) [6]** Let $U = Y^3$. Find $\mathbb{E}[X \mid U]$.

*Hint:* This requires almost no computation, but make sure you justify your answer.

$U$ and $Y$ contain exactly the same information, since from $U$ we can restore $Y$ as $U^{1/3}$, because the function $y \mapsto y^3$ is invertible. Hence

$$\mathbb{E}[X \mid U] = \mathbb{E}\big[\mathbb{E}[X \mid Y, U] \mid U\big]$$
$$= \mathbb{E}\big[\mathbb{E}[X \mid Y] \mid U\big]$$
$$= \mathbb{E}\Big[\frac{Y}{2} \,\Big|\, U\Big]$$
$$= \frac{U^{1/3}}{2}.$$

**(e) [6]** Let $V = Y^2$. Find $\mathbb{E}[X \mid V]$.

*Hint:* This requires almost no computation as well, but first understand why it is different from (c).

The difference from (a) is that now $Y$ contains more information than $V$, since the function $y \mapsto y^2$ is not invertible, so now we need to consider two cases.

$$\mathbb{E}[X \mid V] = \mathbb{E}\big[\mathbb{E}[X \mid Y, V] \mid V\big]$$
$$= \mathbb{E}\big[\mathbb{E}[X \mid Y] \mid V\big]$$
$$= \mathbb{E}\Big[\frac{Y}{2} \,\Big|\, V\Big]$$
$$= \mathbb{E}\Big[\frac{Y}{2} \,\Big|\, V, Y \geq 0\Big]\mathbb{P}(Y \geq 0) + \mathbb{E}\Big[\frac{Y}{2} \,\Big|\, V, Y < 0\Big]\mathbb{P}(Y < 0)$$
$$= \frac{1}{4}\sqrt{V} - \frac{1}{4}\sqrt{V}$$
$$= 0.$$

(Note that the event $\{Y \geq 0\}$ is independent of $V$.)

## Problem 6: Recursive Parameter Estimation [10]

Consider the stochastic dynamical system

$$X_{n+1} = X_n,$$
$$Y_n = X_n + W_n,$$

for each $n \in \mathbb{N}$, where $X_0, W_0, W_1, W_2, \ldots$ are i.i.d. $\mathcal{N}(0,1)$. At time $n - 1$, suppose that we have computed the state estimate $\hat{X}_{n-1|n-1}$. Then, at time $n$, we see the observation $Y_n$.

**(a) [3]** How do we compute the innovation at time $n$?

The innovation is

$$Y_n - \mathbb{E}[Y_n \mid Y_0, Y_1, \ldots, Y_{n-1}] = Y_n - \mathbb{E}[X_n + W_n \mid Y_0, Y_1, \ldots, Y_{n-1}]$$
$$= Y_n - \mathbb{E}[X_{n-1} \mid Y_0, Y_1, \ldots, Y_{n-1}] = Y_n - \hat{X}_{n-1|n-1}.$$

**(b) [7]** We know that the Kalman filter equations recursively calculate $\hat{X}_{n|n}$ in "real time". Suppose we are willing to tolerate a delay of one sample in our estimate, i.e. we can wait until *after* observing $Y_{n+1}$ to estimate $X_n$ for $n = 1, 2, 3, \ldots$. Let $\hat{X}_{n|n}$ be the MMSE of $X_n$ at time $n$ (i.e. based on $Y^{(n)} = Y_0, Y_1, \ldots, Y_n$), and let $\sigma_{n|n}^2$ be our estimation error at time $n$ (i.e. $\sigma_{n|n}^2 = \mathbb{E}[(X_n - \hat{X}_{n|n})^2]$). Consider the update equation

$$\hat{X}_{n|n+1} = \hat{X}_{n|n} + a(Y_{n+1} - \hat{X}_{n|n}).$$

What is the optimal value of $a$ to minimize $\sigma_{n|n+1}^2 = \mathbb{E}[(X_n - \hat{X}_{n|n+1})^2]$?

Observe that $L[Y_{n+1} \mid Y^{(n)}] = L[X_n + W_{n+1} \mid Y^{(n)}] = L[X_n \mid Y^{(n)}] = \hat{X}_{n|n}$ so $Y_{n+1} - \hat{X}_{n|n}$ is the innovation. Write $Y_{n+1} - \hat{X}_{n|n} = X_{n+1} + W_{n+1} - \hat{X}_{n|n} = X_n - \hat{X}_{n|n} + W_{n+1}$ where $X_n - \hat{X}_{n|n}$ is orthogonal and independent of $W_{n+1}$ since $X_n - \hat{X}_{n|n}$ only depends on $(X_0, W_0, W_1, \ldots, W_n)$. From the orthogonal update formula,

$$\hat{X}_{n|n+1} = L[X_n \mid Y^{(n)}] + L[X_n \mid Y_{n+1} - \hat{X}_{n|n}] = \hat{X}_{n|n} + L[X_n \mid X_n - \hat{X}_{n|n} + W_{n+1}]$$

where $\mathbb{E}[X_n(X_n - \hat{X}_{n|n} + W_{n+1})] = \mathbb{E}[X_n(X_n - \hat{X}_{n|n})] = \mathbb{E}[(X_n - \hat{X}_{n|n})^2] = \sigma_{n|n}^2$ (recall that $X_{n|n}$ is orthogonal to $X_n - \hat{X}_{n|n}$ by the property of the MMSE). Also,

$$\mathrm{var}(X_n - \hat{X}_{n|n} + W_{n+1}) = \mathrm{var}(X_n - \hat{X}_{n|n}) + \mathrm{var}\, W_{n+1} = \sigma_{n|n}^2 + 1$$

so using the formula for the LLSE,

$$a = \frac{\sigma_{n|n}^2}{\sigma_{n|n}^2 + 1}.$$

12

**Problem 7: HMM Estimation [15]**

Consider a HMM $(X_n)_{n\in\mathbb{N}}$ with state space $\{0,1\}$ and transitions $P(0,1) = P(1,0) = p \in [0,1]$. The hidden state is observed through a Binary Symmetric Channel (BSC) with error probability $1/3$. Assume that the initial state is equally likely to be 0 or 1. The sequence of observations is $(1,1,1)$.

**(a) [5]** Suppose that it is known that $p = 3/4$. What is the most likely sequence of hidden states?

It is $(1,0,1)$. Let $p_k(x)$ be the largest probability of any sequence of states starting from state $x$ at time $k$. We solve for these quantities backwards as in the Viterbi algorithm.

$$p_2(0) = \frac{1}{3},$$
$$p_2(1) = \frac{2}{3}.$$

These probabilities are just the probabilities of the observations.

$$p_1(0) = \max\left\{\frac{1}{4}p_2(0), \frac{3}{4}p_2(1)\right\} \cdot \frac{1}{3} = \frac{1}{6} \text{ achieved by } 0 \to 1,$$
$$p_1(1) = \max\left\{\frac{3}{4}p_2(0), \frac{1}{4}p_2(1)\right\} \cdot \frac{2}{3} = \frac{1}{6} \text{ achieved by } 1 \to 0,$$
$$p_0(0) = \max\left\{\frac{1}{4}p_1(0), \frac{3}{4}p_1(1)\right\} \cdot \frac{1}{3} = \frac{1}{24} \text{ achieved by } 0 \to 1 \to 0,$$
$$p_0(1) = \max\left\{\frac{3}{4}p_1(0), \frac{1}{4}p_1(1)\right\} \cdot \frac{2}{3} = \frac{1}{12} \text{ achieved by } 1 \to 0 \to 1.$$

The most likely sequence of states is $(1,0,1)$ with probability $1/12$.

**(b) [5]** Now, suppose that $p$ is unknown. We will try applying the hard EM algorithm to this problem. We start by initializing $\hat{p}^{(0)} = 3/4$. In the E step, we fill in the most likely values of the hidden variables $(X_0, X_1, X_2)$ given the observations, assuming that the parameter is $\hat{p}^{(0)}$. Carry out the E step.

Using our answer from the previous part, we fill in $(X_0, X_1, X_2) = (1,0,1)$.

**(c) [5]** In the M step, we compute the MLE of $p$ assuming the filled-in values for the hidden variables computed in the E step. Carry out the M step.

For the hidden states $(X_0, X_1, X_2) = (1,0,1)$, there are 2 switches of the hidden state, so the likelihood of $p$ is $p^2$. Thus the MLE is $\hat{p}^{(1)} = 1$.