
Final

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.
- You have 170 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $170 \cdot X\%$ time on the exam and 10 minutes to submit).
- This exam is not open book. You are permitted to use three double-sided handwritten cheat sheets. No calculator or phones allowed.
- Collaboration is prohibited. If caught cheating, you may fail and face disciplinary actions.
- Write in your SID on every page to receive 1 point.

Problem	points earned	out of
SID		1
Problem 1		22
Problem 2		18
Problem 3		15
Problem 4		9
Problem 5		11
Problem 6		14
Problem 7		11
Problem 8		15
Total		116

1 Another Potpourri of Probability [4 + 7 + 6 + 7 points]**(a) Coin Flips [4 points]**

Let X_1, X_2 be i.i.d. Bernoulli($1/2$) random variables (i.e. fair coin flips).

Show for this choice of X_1 and X_2 that $H(X_1) + H(X_2) \geq H(X_1 + X_2)$.

Since $X_1 + X_2$ has distribution

$$X_1 + X_2 = \begin{cases} 0 & \text{w.p. } 1/4 \\ 1 & \text{w.p. } 1/2 \\ 2 & \text{w.p. } 1/4 \end{cases}$$

the entropy is

$$H(X_1 + X_2) = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2},$$

which is less than $H(X_1) + H(X_2) = 2$.

(b) **Gaussians [2 + 2 + 3 points]**

Let $X = 2Z_1 + 3Z_2$ and $Y = Z_1 + 2Z_2$, where $Z_1, Z_2 \sim_{\text{iid}} N(0, 1)$.

(i) What is the covariance matrix between X and Y , where the entries are

$$\begin{bmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{var}(Y) \end{bmatrix}?$$

(ii) Find $\mathbb{L}[X|Y]$.

(iii) Find $\text{MMSE}[X|Y]$.

(i) As $\begin{pmatrix} X \\ Y \end{pmatrix} = A\vec{Z}$ where $A = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ and \vec{Z} has i.i.d. $N(0, 1)$ entries, we conclude that

$$\text{the covariance matrix is given by } \Sigma = AA^T = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 13 & 8 \\ 8 & 5 \end{bmatrix}.$$

(ii) We compute $\mathbb{L}[X|Y] = \mathbb{E}[X] + \frac{\text{cov}(Y, X)}{\text{var}(Y)}(Y - \mathbb{E}[Y]) = \frac{8}{5}Y$ by observing that X and Y are both zero-mean and reading off the variance and covariance calculated in the previous part.

(iii) Since X and Y are jointly Gaussian, $\text{MMSE}[X|Y] = \mathbb{L}[X|Y] = \frac{8}{5}Y$.

(c) **Poisson MGF [6 points]**

The MGF of $X \sim \text{Poisson}(\lambda)$ is given by $E[e^{tX}] = e^{\lambda(e^t-1)}$. Using this fact, find the distribution of $X + Y$ where $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$, and X is independent of Y .

Note: Finding the distribution without using the MGF will not receive any credit.

$$\begin{aligned} E[e^{t(X+Y)}] &= E[e^{tX} e^{tY}] \\ &= E[e^{tX}] E[e^{tY}] \\ &= e^{\lambda(e^t-1)} e^{\mu(e^t-1)} \\ &= e^{(\lambda+\mu)(e^t-1)} \end{aligned}$$

This is the MGF of a $\text{Poisson}(\lambda + \mu)$ random variable. As MGFs are unique, we conclude $X + Y \sim \text{Poisson}(\lambda + \mu)$.

(d) Tom and Jerry [3 + 4 points]

Tom, Jerry, and 6 of Jerry's other friends are sitting in a room. Outside of the room is a chunk of cheese. At every hour, exactly one of them will get up and exit the room with uniform probability (i.e. if n of them are left, then any one of them will exit with probability $\frac{1}{n}$). Once someone exits the room, they will not return. If Jerry or any one of his friends exits and sees the cheese outside the room, they will eat it completely and leave nothing behind. However, Tom will ignore the cheese.

- (i) What's the probability that Jerry will get to eat the cheese?
- (ii) Now suppose Sohom is also sitting in the room alongside Tom, Jerry, and Jerry's 6 friends. Similar to Tom, Sohom will ignore the cheese once he exits the room. Now what's the probability that Jerry will get to eat the cheese?

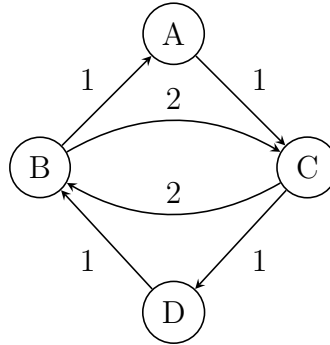
(i) $p = \frac{6! \times 8}{8!} = \frac{1}{7}$. There are $8!$ total orders in which everybody can leave. If we only consider Jerry and his friends, to make sure Jerry gets the cheese, Jerry needs to always be the first among them to go out. In this case, there are $6!$ ways to order the rest of Jerry's friends. Finally, Tom can leave at any time. Once the others' times are fixed, there are 8 positions where we can place Tom into the sequence. Therefore there are $6! \times 8$ sequences in which Jerry can get the cheese.

(ii) $p = \frac{1}{7}$. We actually don't care about when Tom and Sohom leave, since they won't eat the cheese. So we only require Jerry to leave before his 6 friends. Each of the 7 people has the same probability to leave at each hour, so the probability that Jerry leaves first among the 7 people is $\frac{1}{7}$ by symmetry.

2 Markov Chain(s) [8 + 10 points]

(a) CTMC [4 + 4 points]

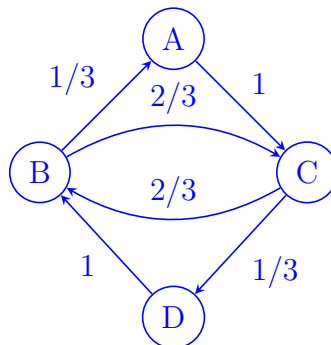
Consider the following CTMC:



- (i) Compute its stationary distribution π_{CTMC} using the associated jump chain.
- (ii) Compute its stationary distribution π_{CTMC} using uniformization.

Note: Finding π_{CTMC} without using the specified method will not receive any credit.

(i) The jump chain looks like:



We can solve to get $\pi_{\text{Jump}} = [1/8, 3/8, 3/8, 1/8]$ (corresponding to states A, B, C, D, respectively). Then the stationary distribution is given by

$$\pi_{\text{CTMC}}(x) = \frac{\frac{1}{Q(x)}\pi_{\text{Jump}}(x)}{\sum_y \frac{1}{Q(y)}\pi_{\text{Jump}}(y)}$$

In particular, we can calculate this to be

$$\begin{aligned} \pi_{\text{CTMC}} &= \frac{1}{1/8 + 1/8 + 1/8 + 1/8} [1/8, 1/8, 1/8, 1/8] \\ &= [1/4, 1/4, 1/4, 1/4] \end{aligned}$$

(ii) Alternatively, using uniformization, we can pick a $q \geq \max_x Q(x)$, i.e. pick $q = 3$, and

compute the stationary distribution of the matrix

$$P = I + \frac{1}{q}Q = \begin{bmatrix} 2/3 & 0 & 1/3 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 2/3 \end{bmatrix}$$

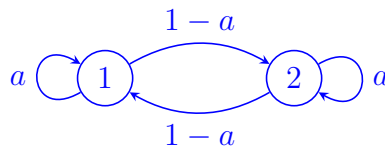
Solving, we get $\pi_{\text{CTMC}} = \pi_{\text{Uniformized}} = [1/4, 1/4, 1/4, 1/4]$.

(b) **DTMC [4 + 6 points]**

- (i) Construct an irreducible discrete time Markov chain with a stationary distribution $(\frac{1}{2}, \frac{1}{2})$. Verify your solution by showing $\pi P = \pi$ where $\pi = (\frac{1}{2}, \frac{1}{2})$.
- (ii) Construct an irreducible discrete time Markov chain such that the stationary distribution is $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. Verify your solution by showing $\pi P = \pi$ where $\pi = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

Note: No credit will be given for this question if you do not explicitly show that $\pi P = \pi$.

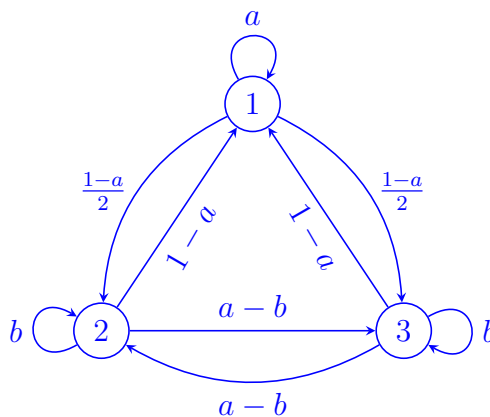
(i) By symmetry, any DTMC with the transition matrix of the form $\begin{bmatrix} a & 1-a \\ 1-a & a \end{bmatrix}$ works.



(ii) Note that if we group nodes 2 and 3 together, then it is a DTMC with a stationary distribution of $(\frac{1}{2}, \frac{1}{2})$. This motivates us to try to decompose the problem using the construction from the previous part. Thus we consider a DTMC where node 1 has a self-loop of probability a and where nodes 2 and 3 have probability of $1-a$ of transitioning back to 1. Then, we add self-loops and connections between 2 and 3 so that they are symmetric. Thus, we find that any DTMC with the transition matrix of the form

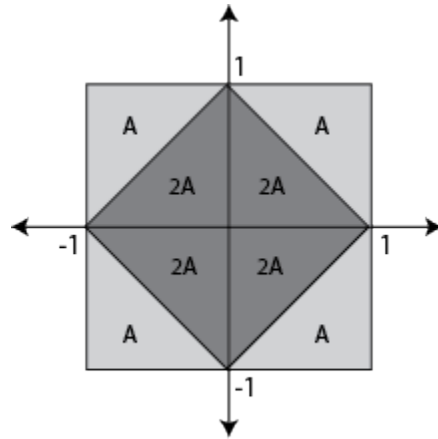
$$\begin{bmatrix} a & \frac{(1-a)}{2} & \frac{(1-a)}{2} \\ 1-a & b & a-b \\ 1-a & a-b & b \end{bmatrix}$$

works, provided that all entries in the matrix are nonnegative. However, there are many other constructions that you can consider.



3 Graphical Density [6 + 4 + 6 points]

Consider the joint density $f_{X,Y}$ shown below:



- Find the value of A and determine closed-form expressions for f_X and f_Y .
- Compute $E[X|Y = y]$ for $-1 \leq y \leq 1$.
- Are the random variables X and Y independent? Are they uncorrelated?

- (a) Integrating over the total area shown, we get $\frac{1}{2}(A + A + A + A) + 2(2A) = 1$ so $A = \frac{1}{6}$. To get $f_X(x)$, we note that the density increases linearly from -1 to 0 and decreases linearly at the same rate from 0 to 1 . The values of f_X evaluated at $x = -1, 0$, and 1 should be $2 * A : 2 * 2A : 2 * A = \frac{1}{3} : \frac{2}{3} : \frac{1}{3}$. Hence the (normalized) density would be

$$f_X(x) = \mathbf{1}_{-1 \leq x \leq 0} \left(\frac{1}{3}x + \frac{2}{3} \right) + \mathbf{1}_{0 < x \leq 1} \left(-\frac{1}{3}x + \frac{2}{3} \right) = \mathbf{1}_{-1 \leq x \leq 1} \left(-\frac{1}{3}|x| + \frac{2}{3} \right)$$

Since the joint density is symmetric, we know that f_Y takes the same form, but as a function of y instead.

- For any value of y , it's clear that the conditional density of X is symmetric about the origin, so $E[X|Y = y] = 0$ for any $-1 \leq y \leq 1$.
- No, they are not independent. If $y = 0$ then $X \sim \text{Uniform}[-1, 1]$, but if $y = 1/2$, X is twice as likely to fall in the range $[-1/2, 1/2]$ than outside of it. They are, however, uncorrelated. Since $E[X|Y = y] = 0 = E[X]$, we can write

$$E[XY] = \int_{-1}^1 E[XY|Y = y]f_Y(y)dy = \int_{-1}^1 yf_Y(y)E[X|Y = y]dy = 0 = E[X]E[Y].$$

4 Fishin' Processes [4 + 5 points]

Akshit is fishing and observes that salmon arrive to his fishing hook according to a Poisson process with rate λ_s per minute, and tuna arrive according to a Poisson process with rate λ_t per minute. These two Poisson processes are independent.

- (a) What is the probability that at least 2 salmon will arrive in one hour?

The arrival process of salmon is distributed according to $PP(\lambda_s)$, so the number of arrivals in a time period is given by $\text{Poisson}(\lambda_s t) = \text{Poisson}(60\lambda_s)$.

$$\Pr(\text{Poisson}(60\lambda_s) \geq 2) = 1 - \Pr(\text{Poisson}(60\lambda_s) \in \{0, 1\}) = 1 - e^{-60\lambda_s} - 60\lambda_s e^{-60\lambda_s}$$

- (b) Akshit decides to start selling salmon and tuna! He learns that some of the salmon and tuna are poor quality, so with probability 0.1 he will discard a fish that he catches, independent of each other. Suppose Albert is waiting to buy fish from Akshit. Assuming there are no previously caught fish available, how long can Albert expect to wait for the next non-discarded fish?

The arrival process of salmon and tuna can be modeled by a merged Poisson process with rate $\lambda_s + \lambda_t$. As fish are independently discarded, we can split this process (with probability 0.9) to get that undiscarded fish appear according to $PP(0.9(\lambda_s + \lambda_t))$. Thus the arrival time of the first undiscarded fish T_1 is distributed according to $\text{Exp}(0.9(\lambda_s + \lambda_t))$, so Albert can therefore expect to wait $E[T_1] = \frac{1}{0.9(\lambda_s + \lambda_t)}$ minutes.

5 Tennis Distribution [5 + 6 points]

Clark is playing tennis! The rate at which he hits depends on the quality of his tennis balls. Suppose that the tennis ball quality is distributed as $\Lambda \sim \text{Geometric}(p)$, for some fixed $p \in (0, 1)$ and given that $\Lambda = \lambda$, the number of balls he hits is distributed as $X \sim \text{Poisson}(\lambda)$.

Hint: If $f(x) > 0$ for all x , then $\arg \max_x f(x) = \arg \max_x \ln(f(x))$.

(a) What is $\text{MLE}[\Lambda|X]$?

(b) What is $\text{MAP}[\Lambda|X]$?

(a) Supposing that we observe $X = x$, we can compute the log-likelihood as

$$\begin{aligned}\ell(\lambda) &= \ln P(X = x | \Lambda = \lambda) \\ &= \ln \left(\frac{\lambda^x e^{-\lambda}}{x!} \right) \\ &= -\ln(x!) + x \ln \lambda - \lambda \\ \ell'(\lambda) &= \frac{x}{\lambda} - 1\end{aligned}$$

Setting the derivative of the log-likelihood to 0 gives $\text{MLE}[\Lambda | X] = X$.

(b) As in the previous part, we compute the log-likelihood:

$$\begin{aligned}\ell(\lambda) &= \ln P(X = x | \Lambda = \lambda) P(\Lambda = \lambda) \\ &= \ln \left(\frac{\lambda^x e^{-\lambda}}{x!} p(1-p)^{\lambda-1} \right) \\ &= -\ln(x!) + x \ln \lambda + \ln \left(\frac{p}{1-p} \right) - \lambda \ln \left(\frac{e}{1-p} \right) \\ \ell'(\lambda) &= \frac{x}{\lambda} - \ln \left(\frac{e}{1-p} \right)\end{aligned}$$

Setting the derivative of the log-likelihood to 0 gives $\text{MAP}[\Lambda | X] = X \left(1 + \ln \left(\frac{1}{1-p} \right) \right)^{-1}$. As $\ell(\lambda)$ is a concave function and Λ is integer valued, we conclude that the MAP estimate of Λ is $\max(\lfloor \lambda^* \rfloor, \lceil \lambda^* \rceil)$ where $\lambda^* = X(1 - \ln(1 - p))^{-1}$.

6 Go Bears! [4 + 5 + 6 points]

Suppose that Cal wins the Big Game with probability p and Stanford with probability $1 - p$, independent of any previous year's result. Your friend at Stanford suggests that $p = \frac{1}{3}$, but you think that $p = \frac{2}{3}$. To decide who's right, you plan to observe the result of three games. Let Y be the number of games Cal wins, and let X be a binary random variable indicating whether you are correct. That is, $X = 0 \iff p = \frac{1}{3}$ and $X = 1 \iff p = \frac{2}{3}$.

Follow the steps to construct a Neyman-Pearson decision rule to maximize $\Pr\{\hat{X} = 1|X = 1\}$ under the constraint that $\Pr\{\hat{X} = 1|X = 0\} \leq \frac{1}{3}$, where \hat{X} is the output of the decision rule.

(a) Find the likelihood ratio $L(y)$ for $y \in \{0, 1, 2, 3\}$

Using the binomial distribution for $p = \frac{1}{3}$ and $p = \frac{2}{3}$, we compute

$$L(y) = \frac{\binom{3}{y} \left(\frac{2}{3}\right)^y \left(\frac{1}{3}\right)^{3-y}}{\binom{3}{y} \left(\frac{1}{3}\right)^y \left(\frac{2}{3}\right)^{3-y}} = \begin{cases} \frac{1}{8} & y = 0 \\ \frac{1}{2} & y = 1 \\ 2 & y = 2 \\ 8 & y = 3 \end{cases}$$

(b) Given $X = 0$, find the values that $L(Y)$ takes on and the associated probabilities.

We first find the distribution of Y , then combine it with the result from the previous part. With $p = \frac{1}{3}$, Y follows a binomial distribution with parameters $n = 3$ and $p = \frac{1}{3}$.

$$\Pr\{Y = y\} = \begin{cases} \frac{8}{27} & y = 0 \\ \frac{12}{27} & y = 1 \\ \frac{6}{27} & y = 2 \\ \frac{1}{27} & y = 3 \end{cases} \implies L(Y) = \begin{cases} \frac{1}{8} & \text{w.p. } \frac{8}{27} \\ \frac{1}{2} & \text{w.p. } \frac{12}{27} \\ 2 & \text{w.p. } \frac{6}{27} \\ 8 & \text{w.p. } \frac{1}{27} \end{cases}$$

(c) Construct the Neyman-Pearson decision rule.

We seek to find c and γ such that under H_0 ,

$$\Pr\{L(Y) > c\} + \gamma \Pr\{L(Y) = c\} = \frac{1}{3}$$

From the distribution of $L(Y)$ in the previous part, we find that $c = \frac{1}{2}$ and $\gamma = \frac{1}{6}$. The optimal decision rule is then

1. if $Y \in \{2, 3\}$, choose $\hat{X} = 1$
2. if $Y = 1$, $\hat{X} = 1$ with probability $\frac{1}{6}$ and $\hat{X} = 0$ with probability $\frac{5}{6}$

3. if $Y = 0$, $\hat{X} = 0$

7 Bacteria LLSE [11 points]

We have a colony of bacteria with initial population $X \sim \text{Poisson}(\lambda)$. Overnight, each bacterium produces a $\text{Poisson}(\lambda)$ number of offspring independently of the others before it passes away. Let Z be the number of bacteria at beginning of the next day.

Mathematically, we can represent this process by letting $X \sim \text{Poisson}(\lambda)$, $Y_1, Y_2, \dots \sim_{\text{iid}} \text{Poisson}(\lambda)$ independent of X , and $Z = \sum_{i=1}^X Y_i$.

Compute $L[X|Z]$.

Hint: The law of total variance, $\text{var}(Y) = E[\text{var}(Y|X)] + \text{var}(E[Y|X])$, may be helpful.

Recall that

$$L[X|Z] = E[X] + \frac{\text{cov}(X, Z)}{\text{var}(Z)}(Z - E[Z])$$

We compute each of these quantities in turn. Note that $E[X] = \lambda$ as the mean of a Poisson distribution and $E[Z] = E\left[\sum_{i=1}^X Y_i\right] = E[X] E[Y_1] = \lambda^2$ by Wald's identity. For the covariance, observe

$$\text{cov}(X, Z) = E[ZX] - E[Z] E[X] = E[X E[Z|X]] - \lambda^3 = E[\lambda X^2] - \lambda^3 = \lambda^2$$

Finally, using the law of total variance, we can compute

$$\begin{aligned} \text{var}(E[Z|X]) &= \text{var}(\lambda X) = \lambda^3 \\ E[\text{var}(Z|X)] &= E\left[\text{var}\left(\sum_{i=1}^X Y_i|X\right)\right] = E[X \text{var}(Y_1)] = \lambda^2 \\ \text{var}(Z) &= \text{var}(E[Z|X]) + E[\text{var}(Z|X)] = \lambda^2 + \lambda^3 \end{aligned}$$

where in the calculation of $E[\text{var}(Z|X)]$ we exploit the independence of the different Y_i . Putting it all together, we get

$$\begin{aligned} L[X|Z] &= \lambda + \frac{1}{1 + \lambda}(Z - \lambda^2) \\ &= \frac{Z + \lambda}{1 + \lambda} \end{aligned}$$

8 Filter Finale [4 + 5 + 6 points]

Consider the system

$$\begin{aligned} X_n &= aX_{n-1} + V_n \\ Y_n &= X_n + W_n \\ n &\geq 1 \end{aligned}$$

where X_0 is zero mean, $(V_n)_{n \geq 1} \sim_{\text{iid}} N(0, \sigma_v^2)$, and $(W_n)_{n \geq 1} \sim_{\text{iid}} N(0, \sigma_w^2)$, all independent of each other. In class, we have seen the Kalman Filter, which estimates X_n given (Y_1, \dots, Y_n) , or $\hat{X}_{n|n}$, as observations stream in. We now wish to work with the Kalman Predictor, which estimates X_{n+1} given (Y_1, \dots, Y_n) , or $\hat{X}_{n+1|n}$. The update equations are shown below (with one missing part). Note that k_n represents the usual Kalman Filter gain.

$$\begin{aligned} \hat{X}_{n+1|n} &\leftarrow a\hat{X}_{n|n-1} + k_n \cdot \textcircled{1} \\ \sigma_{n+1|n}^2 &\leftarrow a^2\sigma_{n|n}^2 + \sigma_v^2 \\ k_n &\leftarrow \sigma_{n|n-1}^2 \cdot (\sigma_{n|n-1}^2 + \sigma_w^2)^{-1} \\ \sigma_{n|n}^2 &\leftarrow (1 - k_n)\sigma_{n|n-1}^2 \end{aligned}$$

- (a) Find the linear innovation $\tilde{Y}_n = Y_n - L[Y_n|Y_1, \dots, Y_{n-1}]$. Express your answer in terms of Y_n and $\hat{X}_{n|n-1}$.

Answer: $(Y_n - \hat{X}_{n|n-1})$

$$\begin{aligned} \tilde{Y}_n &= Y_n - L[Y_n|Y_1, \dots, Y_{n-1}] \\ &= Y_n - L[X_n + W_n|Y_1, \dots, Y_{n-1}] \\ &= Y_n - L[X_n|Y_1, \dots, Y_{n-1}] \\ &= Y_n - \hat{X}_{n|n-1} \end{aligned}$$

- (b) Show that $a\tilde{Y}_n$ fills blank $\textcircled{1}$.

Note that

$$\begin{aligned} \hat{X}_{n+1|n} &= L[X_{n+1}|Y_1 \dots Y_n] \\ &= L[aX_n + W_n|Y_1 \dots Y_n] \\ &= aL[X_n|Y_1 \dots Y_n] \\ &= a(L[X_n|Y_1 \dots Y_{n-1}] + L[X_n|Y_n - L[Y_n|Y_1 \dots Y_{n-1}]]) \\ &= a(\hat{X}_{n|n-1} + k_n\tilde{Y}_n) \end{aligned}$$

Hence, $a\tilde{Y}_n$ fills blank $\textcircled{1}$.

- (c) Suppose $X_0 = 0$, $a = 1$, and $\sigma_v^2 = \sigma_w^2 = 2$. Initialize $\hat{X}_{1|0} = 0$ and $\sigma_{0|0}^2 = 0$. Use the update equations above to express $\hat{X}_{3|2}$ as a linear function of Y_1 and Y_2 .

Answer: $\hat{X}_{3|2} = \frac{1}{5}Y_1 + \frac{3}{5}Y_2$.

First, let's find $\hat{X}_{2|1}$.

$$\sigma_{1|0}^2 = a^2\sigma_{0|0}^2 + \sigma_v^2 = 2$$

$$k_1 = \sigma_{1|0}^2(\sigma_{1|0}^2 + \sigma_w^2)^{-1} = \frac{1}{2}$$

$$\begin{aligned}\hat{X}_{2|1} &= a\hat{X}_{1|0} + k_1a\tilde{Y}_1 \\ &= a\hat{X}_{1|0} + k_1aY_1 - k_1a\hat{X}_{1|0} \\ &= \frac{1}{2}Y_1\end{aligned}$$

Now computing $\hat{X}_{3|2}$.

$$\sigma_{1|1}^2 = (1 - k_1)\sigma_{1|0}^2 = 1$$

$$\sigma_{2|1}^2 = a^2\sigma_{1|1}^2 + \sigma_v^2 = 3$$

$$k_2 = \sigma_{2|1}^2(\sigma_{2|1}^2 + \sigma_w^2)^{-1} = \frac{3}{5}$$

$$\begin{aligned}\hat{X}_{3|2} &= a\hat{X}_{2|1} + k_2a\tilde{Y}_2 \\ &= \frac{1}{2}Y_1 + \frac{3}{5}(Y_2 - \hat{X}_{2|1}) \\ &= \frac{1}{5}Y_1 + \frac{3}{5}Y_2\end{aligned}$$