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.**
- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- You have 10 minutes to read the exam and 160 minutes to complete it.
- The exam is not open book. No calculators or phones allowed.

Problem	points earned	out of
Problem 1		62
Problem 2		16
Problem 3		16
Problem 4		16
Problem 5		16
Total		126

1 Assorted Problems [62]

(a) **Tossing Coins [5]**

A fair coin is tossed eleven times. What is the probability that the sequence of outcomes is a palindrome (i.e. a sequence that is the same when reversed)?

Since the coin is fair, each sequence of 11 flips is equally likely to happen (with probability $\frac{1}{2^{11}}$). We now count the number of palindromes. Note that if we know the first 5 outcomes, we also know the outcomes of tosses 7 through 11. Note that the sixth coin toss isn't restricted. So the number of permutations is $2^5 \cdot 2 = 2^6$. Therefore the probability is $\frac{1}{2^5}$.

(b) **Sketchy Bounds [5]**

As we derived in the homework, the element-wise expectation and variance of $\hat{I} = S^T S$ where S is a $d \times n$ Gaussian sketch matrix (i.e. $S_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \frac{1}{d})$) are

$$\mathrm{E}[\hat{I}_{ij}] = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}, \mathrm{Var}[\hat{I}_{ij}] = \begin{cases} 2/d, & \text{if } i = j \\ 1/d, & \text{otherwise} \end{cases}$$

We want to reduce noise by increasing the dimension d of the sketching matrix. Using Chebyshev's inequality, find a lower bound on d in terms of ϵ such that for the diagonal entry $\hat{I}_{ii}$, we have the bound $P(\hat{I}_{ii} \in [1-\epsilon, 1+\epsilon]) \geq \frac{3}{4}$.

$$P(|\hat{I}_{ii} - 1| > \epsilon) \leq \frac{\mathrm{Var}(\hat{I}_{ii})}{\epsilon^2} \leq \frac{1}{4}$$
$$\frac{1}{4} \geq \frac{2}{d\epsilon^2}$$
$$d \geq \frac{8}{\epsilon^2}$$

(c) **Exponential Shifting [5]**

Let X, Y be distributed independently as Exponential(λ), Exponential(μ). What is the probability that X is less than Y by at least some fixed amount $c \geq 0$?

If $Y - c \leq 0$, then $P(X < Y - c \mid Y - c \leq 0) = 0$. This means we only need to consider the other case where $Y - c > 0$. We can use the memorylessness of exponentials and observe that $P(X < Y - c \mid Y - c > 0) = \frac{\lambda}{\lambda+\mu}$. Combining the above gives

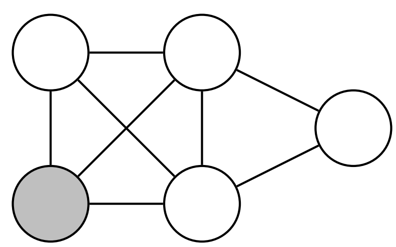
$$\begin{aligned} P(X < Y - c) &= P(X < Y - c \mid Y - c > 0) \cdot P(Y - c > 0) \\ &= \frac{\lambda}{\lambda + \mu} e^{-\mu c} \end{aligned}$$

Alternatively, we use the same setup from homework.

$$\begin{aligned} P(X < Y - c) &= \int_c^\infty P(X < y - c \mid Y = y) \cdot f_Y(y)\, dy \\ &= \int_c^\infty (1 - e^{-\lambda(y-c)}) \mu e^{-\mu y}\, dy \\ &= e^{-\mu c} - \frac{\mu e^{-\mu c}}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu} e^{-\mu c} \end{aligned}$$

(d) **Undirected Markov Chain [5]**

Consider a Markov chain defined on the following undirected graph. At each time step, you pick one of your neighbors (you cannot pick yourself) uniformly at random to move to. What is the stationary distribution probability of the shaded state?



For undirected graphs,

$$\pi(i) = \frac{d(i)}{\sum_j d(j)} = \frac{d(i)}{2E}$$

where E is the number of edges in the graph. Since the shaded state has 3 neighbors and there are 8 edges, the answer is $\frac{3}{16}$.

(e) **Leaving [6]**

Suppose people enter a waiting room according to a Poisson Process with rate λ. Upon a new arrival, each person in the waiting room before the arrival leaves independently with probability $p < 1$. At time 0, the room is empty. At time T, what is the expected number of people in the waiting room?

Hint: Condition on the number of arrivals, N. The MGF of a Poisson random variable Z with rate λ is $M_Z(s) = e^{\lambda(e^s-1)}$

Given that there were N arrivals, the first person who arrived is still in the room with probability $(1-p)^{N-1}$, and the second person is in the room with probability $(1-p)^{N-2}$, and so on with the N'th person being in the room with probability 1. Then by linearity of expectation on indicators, the expected number of people in the room is $\sum_{i=0}^{N-1}(1-p)^i = \frac{1-(1-p)^N}{p}$.

At time T, the number of arrivals N is distributed as Poisson(λT). Thus, we are interested in calculating $\mathbb{E}[\frac{1-(1-p)^N}{p}]$.

$$\begin{aligned}\frac{1}{p}\mathbb{E}[1-(1-p)^N] &= \frac{1}{p}(1-\mathbb{E}[e^{\log(1-p)N}]) \\ &= \frac{1}{p}(1-M_N(\log(1-p))) \\ &= \frac{1-\exp(\lambda T(e^{\log(1-p)}-1))}{p} \\ &= \frac{1-e^{-\lambda pT}}{p}\end{aligned}$$

(f) **Messages [5]**

Justin and Hong are continuously sending messages to you. Each of their messages arrive according to a Poisson Process, and their rates are λ_1 and λ_2, respectively. What is the expected amount of time, T, until you see a message from Justin directly followed by a message from Hong?

Example: For example you will record HJH if you get an arrival from Hong at time T_a, then from Justin at time T_b, then from Hong at time T_c, at which point you will have seen the pattern. Here, $T = T_c$.

Starting from time 0, we know that it will take time $\frac{1}{\lambda_1}$ until we see a message from Justin. Afterwards, additional arrivals from Justin before seeing an arrival from Hong doesn't matter. It will take $\frac{1}{\lambda_2}$ time to see the arrival from Hong. Thus, the answer is $\frac{1}{\lambda_1} + \frac{1}{\lambda_2}$.

(g) **Revisiting ER** [5]

Consider a set of N vertices of a graph. Each vertex i is associated with a $X_i \sim \mathcal{N}(0,1)$ RV, all i.i.d. Suppose we draw an edge between vertex i and vertex j if $X_i + X_j > c$ for some constant c.

(i) What is the probability that a particular edge exists? Your answers to the following questions may be expressed in terms of Φ, the standard Gaussian CDF.

$$P(X_i + X_j > c) = P(\mathcal{N}(0,2) > c) = P(\mathcal{N}(0,1) < -\frac{c}{\sqrt{2}}) = \Phi(-c/\sqrt{2})$$

(ii) Is this an ER random graph? Justify your answer.

◯ True ◯ False

No. Because the distribution of X_i changes if we know that $X_i + X_j > c$, which means that the edges do not appear independently of each other.

(h) **Really Random Binomial MAP** [5]

Suppose you have a binomial variable $X \sim \text{Binomial}(n, U)$. If you know U is distributed as follows

$$U = \begin{cases} 0 \text{ w.p } 0.2 \\ 0.5 \text{ w.p } 0.6 \\ 1 \text{ w.p } 0.2 \end{cases}$$

what is the MAP estimate of U given X.

We want to choose maximize $P(X = x \mid U = u) \cdot P(U = u) = \binom{n}{x} u^x (1-u)^{n-x} \cdot P(U = u)$. Let us look at different choices of u and x. We have to handle the edge cases where $x = 0, n$ separately. The table below computes the posterior $P(X = x \mid U = u) \cdot P(U = u)$.

	$x = 0$	$0 < x < n$	$x = n$
u=0	0.2	0	0
u=0.5	$0.5^n \cdot 0.6$	$0.5^n \cdot 0.6$	$0.5^n \cdot 0.6$
u=1	0	0	0.2

Based on this, we should choose $U = 0$ when $x = 0$ and $n > 1$, $U = 0.5$ when $0 < x < n$ or $n = 1$, and $U = n$ when $x = n$ and $n > 1$. For $n = 1$, we still want to guess $U = 0.5$ since $0.2 < 0.5 \cdot 0.6 = 0.3$.

(i) **Two Sided Hypothesis Test [5]**

Suppose we observe n samples $X_1, X_2, \ldots, X_n$ where $X_i \sim \mathcal{N}(0, \sigma^2)$. We know

$$X = \begin{cases} 0 & \text{if } \sigma = \sigma_0 \\ 1 & \text{if } \sigma = \sigma_1 \end{cases}$$

where $\sigma_0 < \sigma_1$. Show that the test $\hat{X}$ that maximizes $P(\hat{X} = 1 \mid X = 1)$ while ensuring $P(\hat{X} = 1 \mid X = 0)$ is at most β is of the form

$$\hat{X} = \begin{cases} 1 & \text{if } \sum_{i=1}^n X_i^2 > c \\ 0 & \text{otherwise} \end{cases}$$

You do not need to find the specific value of c.

We know that $\hat{X} = 1$ if

$$\ell(x) = \frac{f(x|\sigma_1)}{f(x|\sigma_0)} > \lambda$$

The likelihood ratio is

$$\begin{aligned} \ell(x) &= \frac{\prod_{i=1}^n P(\mathcal{N}(0, \sigma_1^2) = x_i)}{\prod_{i=1}^n P(\mathcal{N}(0, \sigma_0^2) = x_i)} \\ &= \prod_{i=1}^n \frac{\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{x_i^2}{2\sigma_1^2}\right]}{\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{x_i^2}{2\sigma_0^2}\right]} \\ &= \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\frac{1}{2} \sum_i x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right] \end{aligned}$$

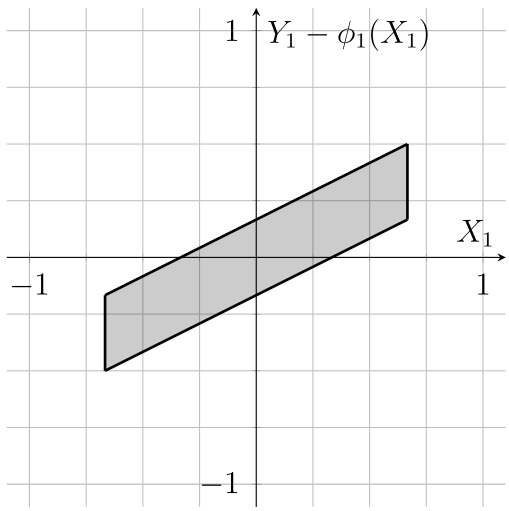
Manipulating the inequality yields:

$$\begin{aligned} \exp\left[\frac{1}{2} \sum_i x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right] &> \lambda\left(\frac{\sigma_1}{\sigma_0}\right)^n \\ \frac{1}{2} \sum_i x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) &> \log(\lambda(\sigma_1/\sigma_0)^n) \\ \sum_i x_i^2 &> \frac{2\log(\lambda(\sigma_1/\sigma_0)^n)}{\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)} = c \end{aligned}$$

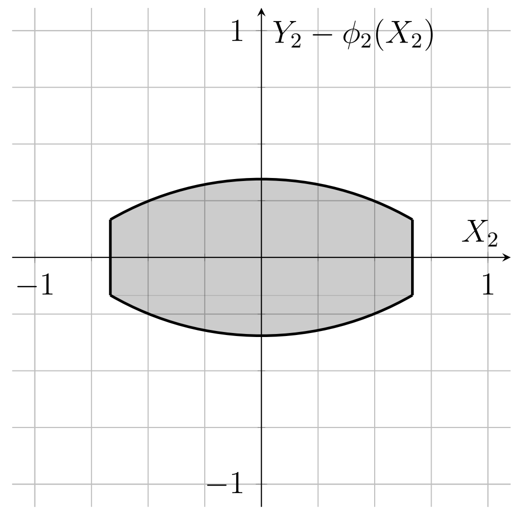
for some c. Note we used the fact that $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} > 0$.

(j) **Graphical Estimators [6]**

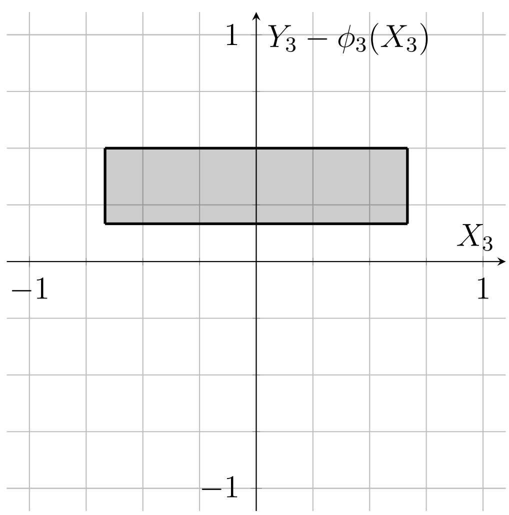
Each of the following 4 plots correspond to an estimator $\phi_i(X_i)$ of Y_i given X_i. In each plot, the joint density of $(X_i, Y_i - \phi_i(X_i))$ is shown. Assuming the density is the uniform distribution on the shaded area, could $\phi_i(X_i)$ be the LLSE and/or the MMSE? Use properties you know about the LLSE and MMSE. **No justification is necessary.**



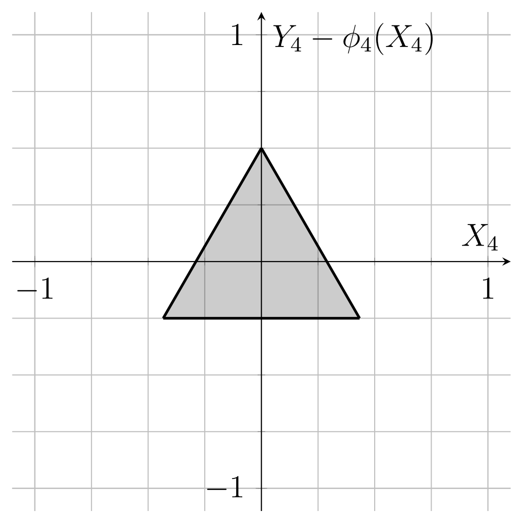
○ Both ○ MMSE ○ LLSE ○ None



○ Both ○ MMSE ○ LLSE ○ None



○ Both ○ MMSE ○ LLSE ○ None



○ Both ○ MMSE ○ LLSE ○ None

$\phi_i(X_i)$ is the LLSE if and only if $\mathbb{E}[Y_i - \phi_i(X_i)] = 0$ and $\mathbb{E}[(Y_i - \phi_i(X_i))X_i] = 0$.
$\phi(X_i)$ is the MMSE if and only if $\mathbb{E}[Y_i - \phi_i(X_i)|X_i] = 0 \quad \forall x$.

Thus, ϕ_1 is neither (correlated), ϕ_2 could be the MMSE/LLSE, ϕ_3 is (not 0 mean), and ϕ_4 could be the LLSE.

(k) **Jointly Gaussian Probability [5]**

Let X be distributed as $\mathcal{N}(0,1)$ and Y be distributed as $\mathcal{N}(1,1)$ with covariance 0.5. Define $W = X - Y$. Find $\Pr(W > Y)$ in terms of the standard Gaussian CDF, Φ.

> W and Y are jointly gaussian, so $W - Y = X - 2Y$ is Gaussian.
>
> $$\begin{aligned} \mathrm{E}[X - 2Y] &= \mathrm{E}[X] - 2\,\mathrm{E}[Y] = 0 - 2 \cdot 1 = -2 \\ \mathrm{var}(X - 2Y) &= \mathrm{cov}(X - 2Y, X - 2Y) \\ &= \mathrm{var}(X) - 4\,\mathrm{cov}(X, Y) + 4\,\mathrm{var}(Y) \\ &= 1 - 4 \cdot 0.5 + 4 \cdot 1 = 3 \end{aligned}$$
>
> Hence, $P(W - Y > 0) = P(\mathcal{N}(-2, 3) > 0) = P(\mathcal{N}(0,1) > \frac{2}{\sqrt{3}}) = \Phi(-\frac{2}{\sqrt{3}})$.

(l) **Jointly Gaussian True or False [5]**

For the following two questions, justify your answer or describe a counterexample.

(i) If two random variables X and Y are marginally Gaussian, then they are jointly Gaussian.

◯ True ◯ False

> **False**. Consider $U, V \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. We can define X and Y by zeroing out the density in the second and fourth quadrants and doubling the density in the first and third. The marginal densities stay the same but the RVs are no longer jointly Gaussian. Check out https://stats.stackexchange.com/questions/30159/is-it-possible-to-have-a-pair-of-gaussian-random-variables-for-which-the-joint-d

(ii) If two independent random variables X and Y are marginally Gaussian, then they are jointly Gaussian.

◯ True ◯ False

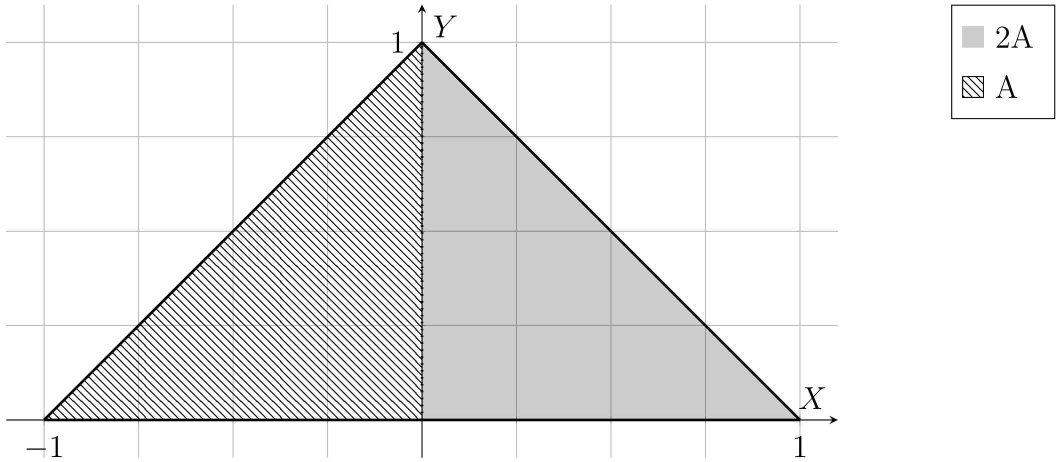
> **True**. We can express $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ as
>
> $$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \sigma_X & 0 \\ 0 & \sigma_Y \end{bmatrix} \begin{bmatrix} \frac{X - \mu_X}{\sigma_X} \\ \frac{Y - \mu_Y}{\sigma_Y} \end{bmatrix} + \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$
>
> Here, $\frac{X - \mu_X}{\sigma_X}$ and $\frac{Y - \mu_Y}{\sigma_Y}$ are independent standard normal RVs.

(iii) If $L[X \mid Y] = E[X \mid Y]$, then X and Y are jointly Gaussian.

◯ True ◯ False

> **False**. If X and Y are independent, then $E[X \mid Y] = E[X]$, which is linear. However, X and Y could have arbitrary distributions.

2 Graphical Density [16]



(a) **MMSE [4]**

Find the MMSE of Y given X.

$$E[Y|X = x] = \begin{cases} \frac{1}{2} + \frac{X}{2} & \text{if } x \le 0 \\ \frac{1}{2} - \frac{X}{2} & \text{if } x > 0 \end{cases}$$

(b) **Covariance [6]**

Show that $\text{cov}(X, Y) = -\frac{\text{var}(Y)}{6}$.

To find the covariance we need to find $E[XY] - E[X]\,E[Y]$. To find both $E[XY]$ and $E[X]$, we will use the tower property. Note that

$$E[X] = E[E[X \mid Y = y]]$$
$$E[XY] = E[E[XY|Y = y]] = E[Y\,E[X|Y = y]]$$

To compute $E[X|Y = y]$, we can look at the left and right halves separately. The average X for a given $Y = y$ is $\frac{1-y}{2}$ in the right half and $-\frac{1-y}{2}$ in the left half. Weighting these by $\frac{2}{3}$ and $\frac{1}{3}$, we get

$$E[X \mid Y = y] = \frac{2}{3} \cdot \frac{1-y}{2} - \frac{1}{3} \cdot \frac{1-y}{2} = \frac{1-y}{6}$$
$$E[X] = E[\frac{1-Y}{6}] = \frac{1-E[Y]}{6}$$
$$E[XY] = E[\frac{Y(1-Y)}{6}] = \frac{E[Y] - E[Y^2]}{6}$$

Combining everything together, we get

$$\begin{aligned}\operatorname{cov}(X,Y) &= \frac{\mathrm{E}[Y]-\mathrm{E}[Y^2]}{6} - \frac{1-\mathrm{E}[Y]}{6}\mathrm{E}[Y] \\ &= \frac{-\mathrm{E}[Y^2]+\mathrm{E}[Y]^2}{6} = -\frac{\operatorname{var}(Y)}{6}\end{aligned}$$

(c) **LLSE [6]**

Find the LLSE of Y given X, $L[Y \mid X]$. Hint: To find var(X), think about the random variable —X—.

Note: There was a mistake in the solutions originally.

The formula of LLSE is

$$\begin{aligned}L[Y|X] &= \mathrm{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(X-\mathrm{E}[X]) \\ &= \mathrm{E}[Y] + \frac{-\operatorname{var}(Y)}{6\operatorname{var}(X)}(X-\mathrm{E}[X])\end{aligned}$$

To find $\mathrm{E}[Y]$ we should find the marginal PDF for Y. A look at the the graph tells us its linearly decreasing from 0 to 1 and therefore should be of the form $f_Y(y) = \alpha(1-y)$. The normalizing constant ends up being 2, and therefore,

$$\mathrm{E}[Y] = \int_0^1 2(1-y)\cdot y\,dy = \int_0^1 2y - 2y^2\,dy = \frac{2}{2} - \frac{2}{3} = \frac{1}{3}$$

Then, using the fact that $\mathrm{E}[X] = \frac{1-\mathrm{E}[Y]}{6}$ from the previous part, we get

$$\mathrm{E}[X] = \frac{1-\frac{1}{3}}{6} = \frac{1}{9}$$

Now that we have $\mathrm{E}[X]$ and $\mathrm{E}[Y]$, to calculate $\operatorname{var}(X)$ and $\operatorname{var}(Y)$, we need to find $\mathrm{E}[X^2]$ and $\mathrm{E}[Y^2]$. To find $\mathrm{E}[Y^2]$, we use the definition.

$$\mathrm{E}[Y^2] = \int_0^1 2(1-y)\cdot y^2\,dy = 2(\frac{1}{3} - \frac{1}{4}) = \frac{1}{6}$$

To find $\mathrm{E}[X^2]$, we use the fact that $\mathrm{E}[X^2] = \mathrm{E}[|X|^2]$. However $|X|$ is just X "folded" on itself, and it looks exactly like Y afterwards. So $\mathrm{E}[X^2] = \mathrm{E}[Y^2] = \frac{1}{6}$. To put everything together,

$$\begin{aligned}L[Y \mid X] &= \frac{1}{3} + -\frac{1}{6}\Big(\frac{\frac{1}{6} - (\frac{1}{3})^2}{\frac{1}{6} - (\frac{1}{9})^2}\Big)(X - \frac{1}{9}) \\ &= \frac{1}{3} - \frac{3}{50}(X - \frac{1}{9})\end{aligned}$$

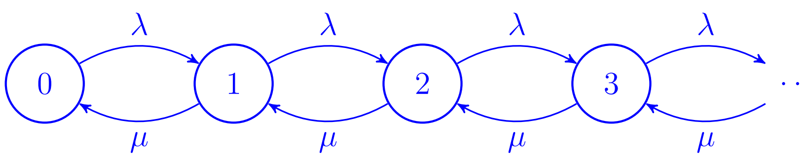
3 Operating Systems [16]

In an operating system, new tasks arrive to a queue according to a PP with rate λ. Furthermore, suppose tasks are processed one by one in the order they arrived, and that the processing time for each task independently has an exponential distribution with rate μ. For all of the parts, suppose $\lambda < \mu$ and that a long time has passed.

(a) **Expected Number of Tasks [6]**

What is the expected number of tasks $\mathrm{E}[X]$ in the queue?

We can model the number of tasks in the queue with the following CTMC.



To calculate the expected number of tasks after a long time, we can calculate the stationary distribution and use the definition. The stationary distribution is

$$\pi(k) = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k$$

The expectation is

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=1}^{\infty} k \cdot \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) \\
&= \frac{\lambda}{\mu} \cdot \left[\sum_{k=1}^{\infty} k \cdot \left(\frac{\lambda}{\mu}\right)^{k-1} \left(1 - \frac{\lambda}{\mu}\right)\right] \\
&= \frac{\lambda}{\mu} \cdot \frac{1}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda}
\end{aligned}$$

(b) **Expected Delay [5]**

A new task arrives on the queue. What is the expected delay $\mathrm{E}[T]$ before it is done processing? You may write your answer in terms of your answer to the previous question, $\mathrm{E}[X]$.

Once a task arrives in the queue, it just needs to wait for each task ahead of it to be processed, then itself. Each task takes time Exponential(μ), so

$$\mathrm{E}[T] = \mathrm{E}[\mathrm{E}[T \mid X]] = \mathrm{E}\left[\frac{X+1}{\mu}\right]$$

$$= \frac{1}{\mu} \cdot \left(\frac{\lambda}{\mu - \lambda} + \frac{\mu - \lambda}{\mu - \lambda}\right)$$

$$= \frac{1}{\mu} \cdot \frac{\mu}{\mu - \lambda} = \frac{1}{\mu - \lambda}$$

(c) **LLSE [5]**

Suppose the level of the noise your computer fan makes depends on the number of tasks in the queue. In particular, the noise level $Z \sim \text{Poisson}(X)$. What is $L[X \mid Z]$? You may write your answer in terms of $\mathrm{E}[X]$ and $\mathrm{var}(X)$.

We need to calculate $\mathrm{cov}(X, Z)$, $\mathrm{var}(Z)$, and $\mathrm{E}[Z]$.

- $\mathrm{E}[Z] = \mathrm{E}[\mathrm{E}[Z \mid X]] = \mathrm{E}[X]$.
- $\mathrm{cov}(X, Z) = \mathrm{E}[XZ] - \mathrm{E}[X]\,\mathrm{E}[Z] = \mathrm{E}[\mathrm{E}[XZ \mid X]] - \mathrm{E}[X]^2 = \mathrm{E}[X\,\mathrm{E}[Z \mid X]] - \mathrm{E}[X]^2 = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \mathrm{var}(X)$.
- $\mathrm{var}(Z) = \mathrm{E}[\mathrm{var}(Z \mid X)] + \mathrm{var}(\mathrm{E}[Z \mid X]) = \mathrm{E}[X] + \mathrm{var}(X)$.

The answer is

$$L[X \mid Z] = \mathrm{E}[X] + \frac{\mathrm{var}(X)}{\mathrm{var}(X) + \mathrm{E}[X]}(Z - \mathrm{E}[X])$$

4 Gaussian Poker [16]

Justin and Will are playing poker. Justin wants to estimate the true value of Will's hand from his bets. The game can be modeled as follows.

- Initially his hand has value $X_0 \sim \mathcal{N}(0, 3)$. It can actually be negative.
- Every round, a new card is drawn, and the value of his hand changes to $X_n = X_{n-1} + V_n$, where $V_n \sim \mathcal{N}(0, 1)$ $(n = 1, 2, \ldots)$.
- After each round, he bets $Y_n = X_n + W_n$, where $W_n \sim \mathcal{N}(0, \sigma_w^2)$ $(n = 1, 2, \ldots)$. These can also be negative.
- X_0, $V_1, V_2, \ldots$, and $W_1, W_2, \ldots$ are all independent.

(a) **First Round** [5]

Suppose $\sigma_w^2 = 2$. What is $E[X_1 \mid Y_1]$?

We can initialize the Kalman Filter equations with $\hat{x}_{0|0} = E[X_0] = 0$ and $\sigma_{0|0}^2 = \text{var}(X_0) = 3$. Then,

$$\hat{x}_{1|0} = \hat{x}_{0|0} = 0$$
$$\sigma_{1|0}^2 = \sigma_{0|0}^2 + \sigma_v^2 = 3 + 1 = 4$$
$$k_1 = \frac{\sigma_{1|0}^2}{\sigma_{1|0}^2 + \sigma_w^2} = \frac{4}{4+2} = \frac{2}{3}$$
$$\hat{x}_{1|1} = \hat{x}_{1|0} + k_1(Y_1 - \hat{x}_{1|0}) = \frac{2}{3}Y_1$$

(b) **Second Round** [6]

Again suppose $\sigma_w^2 = 2$. What is $E[X_2 \mid Y_1, Y_2]$?

We can continue the Kalman Filter equations as follows.

$$\sigma_{1|1}^2 = \sigma_{1|0}^2(1 - k_1) = 4\left(1 - \frac{2}{3}\right) = \frac{4}{3}$$
$$\hat{x}_{2|1} = \hat{x}_{1|1} = \frac{2}{3}Y_1$$
$$\sigma_{2|1}^2 = \sigma_{1|1}^2 + \sigma_v^2 = \frac{4}{3} + 1 = \frac{7}{3}$$
$$k_2 = \frac{\sigma_{2|1}^2}{\sigma_{2|1}^2 + \sigma_w^2} = \frac{\frac{7}{3}}{\frac{7}{3} + 2} = \frac{7}{13}$$
$$\hat{x}_{2|2} = \hat{x}_{2|1} + k_2(Y_2 - \hat{x}_{2|1})$$
$$= \frac{2}{3}Y_1 + \frac{7}{13}\left(Y_2 - \frac{2}{3}Y_1\right)$$

$$= \frac{12}{39}Y_1 + \frac{7}{13}Y_2$$

(c) **He's Bluffing [5]**

Suppose Justin didn't actually know Will's betting habits, but wanted to estimate his spread. Given a record of Will's bets $Y_1, \ldots, Y_n$ and the value of his hand on the corresponding rounds $X_1, \ldots, X_n$ find the MLE of the standard deviation σ_w.

The MLE estimator maximizes the log likelihood $\ell(\sigma_w | X_n, Y_n)$:

$$\begin{aligned}
\ell(\sigma_w | X_n, Y_n) &= \sum_{k=1}^{n} \log f(Y_k - X_k | \sigma_w) \\
&= \sum_{k=1}^{n} \log\Big[\frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{(Y_k - X_k)^2}{2\sigma_w^2}}\Big] \\
&\propto -\sum_{k=1}^{n}\Big[\log \sigma_w + \frac{1}{2}\Big(\frac{Y_k - X_k}{\sigma_w}\Big)^2\Big]
\end{aligned}$$

Taking the partial derivative with respect to σ_w:

$$\begin{aligned}
-\Big[\frac{n}{\sigma_w} - \sum_{k=1}^{n} \frac{(Y_n - X_n)^2}{\sigma_w^3}\Big] &= 0 \\
n\sigma_w^2 - \sum_{k=1}^{n} (Y_n - X_n)^2 &= 0 \\
\hat{\sigma}_w &= \sqrt{\frac{1}{n}\sum_{k=1}^{n} (Y_k - X_k)^2}
\end{aligned}$$

5 Gaussian Process [16]

We define a **Gaussian Process** to be a sequence of random variables $X_1, X_2, \dots$ such that any finite subset $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ is jointly Gaussian. (Thus, this further implies that each of the random variables X_i are Gaussian.) Suppose that the mean of each variable is $\mathbb{E}[X_i] = 0$, and the covariance between two of the Gaussians X_i, X_j is

$$\operatorname{cov}(X_i, X_j) = \frac{1}{ij}\min(i, j).$$

(a) **Covariance Matrix [5]**

Compute the covariance matrix of the random vector $\begin{bmatrix} X_i \\ X_j \end{bmatrix}$ where $i < j$.

$$\begin{bmatrix} \frac{1}{i} & \frac{1}{j} \\ \frac{1}{j} & \frac{1}{j} \end{bmatrix}$$

(b) **Orthogonalization [6]**

For $i < j$, find the random vector $\begin{bmatrix} Y_i \\ Y_j \end{bmatrix}$ such that Y_i, Y_j are linear combinations $a_i X_i + b_i X_j, a_j X_i + b_j X_j$ of the random variables X_i, X_j and the covariance matrix of $\begin{bmatrix} Y_i \\ Y_j \end{bmatrix}$ is the identity matrix I.

We first normalize X_i by letting $Y_i = \frac{X_i}{\sqrt{\operatorname{var}(X_i)}} = \sqrt{i}X_i$. We then want to get Y_j by finding the part of X_j that is orthogonal to X_i, then normalizing. The orthogonal component is

$$X_j - \frac{\langle X_j, X_i\rangle}{\langle X_i, X_i\rangle}X_i = X_j - \frac{\operatorname{cov}(X_j, X_i)}{\operatorname{var}(X_i)}X_i = X_j - \frac{i}{j}X_i$$

To normalize this, we find the variance

$$\begin{aligned}\operatorname{var}(X_j - \frac{i}{j}X_i) &= \operatorname{var}(X_j) - 2\frac{i}{j}\operatorname{cov}(X_i, X_j) + \frac{i^2}{j^2}\operatorname{var}(X_i) \\ &= \frac{1}{j} - \frac{2i}{j^2} + \frac{i}{j^2} = \frac{j-i}{j^2} \\ Y_j &= \frac{j}{\sqrt{j-i}}\Big(X_j - \frac{i}{j}X_i\Big)\end{aligned}$$

This leads to

$$Y_i = \sqrt{i}X_i, Y_j = \frac{jX_j - iX_i}{\sqrt{j-i}}$$

(c) **Process Convergence [5]**

Prove that the sequence of random variables $\{X_i\}_{i=1}^{\infty}$ converges to 0 almost surely. Hint: Choose $i = j - 1$ in part (b) and see if you can rewrite the Gaussian Process $\{X_i\}_{i=1}^{\infty}$ so that we can apply SLLN.

Part (b) tells us that $Z_j = jX_j - (j-1)X_{j-1}$ is a $\mathcal{N}(0,1)$ random variable. Now, comparing two different Z_i and Z_j (WLOG assuming $i < j$), note that

$$\begin{aligned}\text{cov}(Z_i, Z_j) &= \text{cov}(iX_i - (i-1)X_{i-1}, jX_j - (j-1)X_{j-1}) \\ &= \frac{ij}{j} - \frac{i(j-1)}{j-1} - \frac{(i-1)j}{j} + \frac{(i-1)(j-1)}{j-1} \\ &= i - i - (i-1) + (i-1) = 0\end{aligned}$$

Since all Z_i are uncorrelated and jointly Gaussian, they're independent. And since it turns out that

$$\begin{aligned}X_n &= \frac{1}{n}(nX_n) \\ &= \frac{1}{n}\sum_{i=1}^{n} iX_i - (i-1)X_{i-1} \\ &= \frac{1}{n}\sum_{i=1}^{n} Z_i\end{aligned}$$

by the strong law of large numbers this converges almost surely to 0.