Spring 2005: EECS126 Final on May 18, 2005

No Collaboration Permitted. You are allowed to use your textbook.

Be clear and precise in your answers if possible

Come to front if you have a question.

Good luck!

Problem 1.1 *(45pts) True or False. Prove or show a counterexample:*

a. 15pts. A Poisson process with arrival rate λ is split into two processes such that odd numbered arrivals are assigned to process A and even arrivals are assigned to process B. **The resulting process A is also a Poisson process.**

False. The inter-arrival times of the process A have Erlang-order(2) distribution, which is a result of a sum of two exponentially distributed inter-arrival times of the original process. Therefore the process A is not a Poisson process.

Note that this model is different than splitting Poisson processes in two processes by tossing a coin. In that case inter-arrival times of the split processes remain exponential. In this model inter-arrivals times become the sum of two exponential, thus Erlang-order(2) distributed.

b. 15pts. If i and j are both recurrent states in a finite state Markov chain, then it is possible to reach state j from state i and vice versa.

False. If i and j are recurrents states that belong to two different recurrent classes then it is not possible to reach state j from state i and vice versa, see example in Figure below.

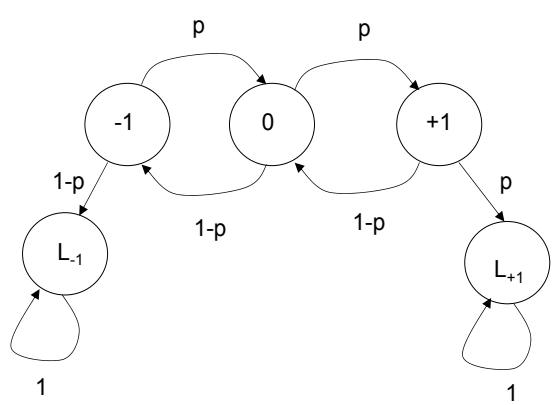

Figure 1: Markov chain with two recurrent classes.

c. 15pts. *X, V, W are i.i.d. standard Normal random variables (0 mean, and variance 1). Let $Y = X + V$ and $Z = X + V + W$. Then the best mean-squared estimate of X given observations of both Y and Z is the same as the best mean-squared estimate of X given an observation only of Y.*

$$E[X|Y = y, Z = z] = E[X|Y = y]$$

True. The easy way to see this is to look at the following Markov Process. Let $X_0 = X$, $X_1 = X_0 + V$, and $X_2 = X_1 + W$. Clearly this is Markov and $X_1 = Y$ and $X_2 = Z$. By Markovianity, the random variables X_0 and X_2 are conditionally independent given X_1. From this, the statement follows immediately.

Worked out in full, we take advantage of everything's obvious joint normality and just notice that $E[X|Y] = \frac{Y}{2}$. Now to see that this is also $\hat{X} = E[X|Y, Z]$, we just look at the residual error $(X - \hat{X}) = (X - \frac{Y}{2}) = (\frac{X}{2} - \frac{V}{2})$ and show that it is independent of Z. To see this:

$$\begin{aligned}
E[Z(X - \frac{Y}{2})] &= E[Z(\frac{X}{2} - \frac{V}{2})] \\
&= E[(X + V + W)(\frac{X}{2} - \frac{V}{2})] \\
&= E[\frac{X^2}{2}] + E[\frac{-V^2}{2}] \\
&= \frac{1}{2} - \frac{1}{2} \\
&= 0
\end{aligned}$$

Because these random variables are both zero-mean and jointly normal, this implies that they are independent. Since the estimation error for $\frac{Y}{2}$ is thus independent of Z, it must be the best estimate given knowledge of both Y and Z.

Problem 1.2 *Choosing investments*

A line of people all want to invest in the technology that is going to be best. There is a choice of technologies, each of which is a-priori *equally likely to turn out to be the best. In this problem we will explore some strategies for choosing the right investment.*

a. *20pts. In the first case, people decide to choose a technology by drawing a common set of boxes of various sizes on the wall and each hiring a monkey to throw a dart uniformly at random towards the wall. They will each invest in whichever technology that the dart hits. It hits the correct technology with probability* $\frac{4}{13}$ *independently of all the throws from other monkeys.*

Suppose that there are 16900 *people investing. A certain online casino offers you a chance to bet on the exact number of people that will end up choosing the correct technology. Your online entry form lets you place upto* 100 *distinct and simultaneous bets on numbers from* 0 *to* 16900 *and will give you a prize if any of your bets hits the true number of people who happen to choose the right technology.*

Where should you place your 100 bets in order to maximize your probability of getting the prize?

Give an exact expression for your probability of winning the prize under your betting strategy.

Give an integral which approximates the probability of winning the prize under your betting strategy.

Let X_i be the random variable representing the outcome of the investment of the ith person based on its monkey dart. Let $X_i = 1$ if the ith person invested in the right technology (success), and $X_i = 0$ if the ith person did not invest in the right technology (failure). According to the model, X_i's are i.i.d $Bernoulli(p = \frac{4}{13})$. The total number of people that invested in the right technology is $N = \sum_{i=1}^{16900} X_i$. N is a Binomial random variable with n=16900, mean $\mu = 16900 * \frac{4}{13} = 5200$ and variance $\sigma^2 = 16900\frac{4}{13}\frac{9}{13} = 3600$. Since the number of people is quite large and N is the sum of the large number of i.i.d. the Central limit theorem tells us that:

$$Z = \frac{X_1 + X_2 + ... + X_{16900} - 16900p}{\sqrt{16900p(1-p)}} = \frac{N - \mu}{\sigma}$$

converges to the standard normal CDF. Since the distribution of N is symmetric and centered around 5200 then the best strategy would be to place your bets uniformly between 5150 and 5250 (for exactly 100 bets you can choose 5149).

For this case the probability of winning the prize would be

$$P(5150 \leq N \leq 5249) = \sum_{k=5150}^{5249} \binom{16900}{k} p^k (1-p)^{(16900-k)}$$

This probability can be approximated by the following:

$$\begin{aligned} P(|N-5200| \leq 50) &= P(\frac{|N-5200|}{60} \leq \frac{50}{60}) \\ &= \Phi(\frac{50}{60}) - \Phi(-\frac{50}{60}) \\ &= 2\Phi(0.0139) - 1 = 2 * 0.798 - 1 = 0.596 \end{aligned}$$

where

$\Phi(z) = \lim_{n\to\infty} P(Z_n \leq z)$ and Z_n is standard normal $N(0,1)$.

b. 20 pts. Case with "Secret Information" and Public Actions

People were dissatisfied with the monkeys' investment advice. With some effort, they are able to collectively reduce the technological possibilities to two equally likely choices, one of which is guaranteed to be correct. Each person decides to hire an independent imperfect analyst that will be able to predict which technology is going to succeed with probability $p > \frac{1}{2}$. *The analysts are therefore each independently wrong with probability* $1 - p$. *The reliability* p *is a constant known by everyone.*

We model the situation as follows: Let Q *be the random variable representing which technology will succeed. It is* $+1$ *with probability* $\frac{1}{2}$ *and* -1 *with probability* $\frac{1}{2}$. *The* i*-th analyst report* A_i *is defined by:*

$$A_i = Q * V_i$$

where the V_i *are i.i.d. and equal to* $+1$ *(representing the event that the analyst gets it right) with probability* p *and* -1 *(representing the event that the analyst gets it wrong) with probability* $1 - p$.

At first, the analysts' reports were covered by non-disclosure agreements and so all that person i *gets to observe are the actual investments of the first* $i - 1$ *people and her own private analyst report. She is perfectly rational and will chose the investment that has the greatest probability of succeeding based on all the information she has. If it is a tie (both investments appear equally likely), then she defers to her private analyst's judgment and will pick whatever he recommends.*

Let X_i *be the investment decision of the* i*-th person. She has access to* A_i *and all the* X_k *for* $k < i$. *She knows that the people before her are following their individually best strategies. It turns out that the best strategy for every* i *is to choose* X_i *to agree with the majority of the combination of the* X_k *that came before us and our private information* A_i. *(i.e. Hold a vote where all the past* X_k *get* 1 *vote along with the current* A_i *which also gets* 1 *vote as well as the ability to break ties.)* (For a substantial bonus, you may argue why this "follow the majority's decision" rule is individually optimal.)

For these rules, what is $P(X_n \neq Q)$ **in the limit of large** n**?**

HINT: You might want to model the net-state of information as a Markov chain. Where do you start? Are there any recurrent states? Any transient ones?

The state of the information before we make a new observation A_i is the currently running tally of how many votes for the first investment minus how many votes for the second investment. Let us call this $Y_{i-1} = \sum_{k=1}^{i-1} X_k$.

When person i gets the a new observation A_i, she will combine it with the previous Y_{i-1} to generate her X_i as follows:

$$X_i = \begin{cases} +1 & \text{if } Y_{i-1} + A_i > 0 \\ A_i & \text{if } Y_{i-1} + A_i = 0 \\ -1 & \text{if } Y_{i-1} + A_i < 0 \end{cases}$$

Then we can generate the next Y_i by the rule

$$Y_i = Y_{i-1} + X_i$$

We also know that we start out with $Y_0 = 0$ and so this looks like a birth-death Markov Process. But we can actually simplify this somewhat by taking a closer look at the state dependence of X_i.

We notice that if $Y_{i-1} \geq 2$, then no matter what value of A_i that comes up, we are always going to choose $X_i = +1$. Similarly, if $Y_{i-1} \leq -2$, then no matter what value of A_i that comes up, we are always going to choose $X_i = -1$. As a result, for all practical purposes, the states $Y_i \geq 2$ are equivalent and so we can group them all together into a single state that we call L_{+1}. Similarly, we can group all the states $Y_i \leq -2$ together into a single state that we call L_{-1}. We call these states L to remind us of "Lemmings" because once in this state, the people will just follow the person ahead of them in a completely mindless fashion.

On the other hand, if Y_{i-1} is either $+1, 0, -1$, then $X_i = A_i$. So these states are distinguishable. Since Q can be either $+1$ or -1 in an equally likely fashion and the rules above are symmetric to these possibilities, we can assume for the purpose of analysis that $Q = +1$. So we now we have a 5 state birth death Markov chain set up as follows:

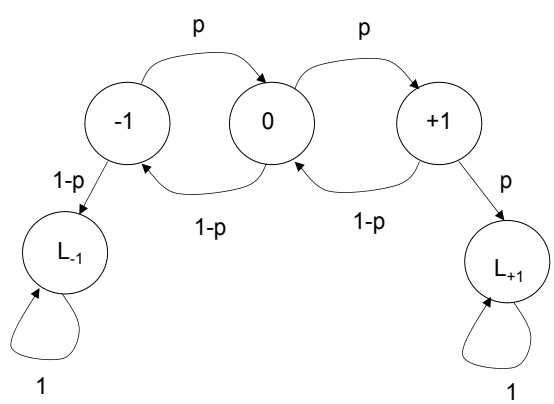

Figure 2: 5 state Markov chain

L_{-1} No births or deaths ever occur in this state. Once here, we stay forever.

-1 A birth occurs whenever we move towards state 0. This happens when $A_i = Q$ and so has probability p. A death occurs whenever we fall into L_{-1} which happens when $A_i = -Q$ and so has probability $1 - p$.

0 A birth occurs whenever we move towards state $+1$. This happens when $A_i = Q$ and so has probability p. A death occurs whenever go to -1 which happens when $A_i = -Q$ and so has probability $1-p$.

$+1$ A birth occurs whenever we fall into state L_{+1}. This happens when $A_i = Q$ and so has probability p. A death occurs whenever go to 0 which happens when $A_i = -Q$ and so has probability $1-p$.

L_{+1} No births or deaths ever occur in this state. Once here, we stay forever.

It is clear that this Markov chain has two recurrent classes. One consisting of L_{-1} by itself and the other consisting of L_{+1} by itself. Eventually, we are going to fall into one of the two Lemming modes. If we fall into L_{-1}, we will make erroneous investments forever. If we fall into L_{+1}, we will make correct investments forever. As long as we are in the transient states $-1, 0, +1$, we make correct investments with probability p.

To find what is $P(X_n \neq Q)$ in the limit of large N, we just need to find out what is the probability that we will end up in L_{-1} given that we start in 0. This is a question of absorption probabilities and can be answered by letting a_s reflect the probability of eventually being absorbed by L_{-1} given that we are in state s. Then we know that the following system of equations is satisfied:

$$\begin{aligned}
a_{L_{-1}} &= 1 \\
a_{-1} &= (1-p)a_{L_{-1}} + pa_0 \\
a_0 &= (1-p)a_{-1} + pa_{+1} \\
a_{-1} &= (1-p)a_0 + pa_{L_{+1}} \\
a_{L_{+1}} &= 0
\end{aligned}$$

which simplifies immediately to

$$\begin{aligned}
a_{-1} &= (1-p) + pa_0 \\
a_0 &= (1-p)a_{-1} + pa_{+1} \\
a_{-1} &= (1-p)a_0
\end{aligned}$$

by applying the substitutions in terms of a_0 given in the last equations we then get:

$$a_0 = (1-p)((1-p) + pa_0) + p(1-p)a_0$$

which solves as

$$a_0 = \frac{(1-p)^2}{1-2p(1-p)}$$

This is the long term probability of making a mistake. Notice that it does not go to 0 no matter how long we wait.

c. 20pts. Case with "Freedom of Information"

After their experience in part (b), the society had the good sense to ban non-disclosure agreements. In this new case, all information is public and every person gets to observe both the actions and the underlying information available to everyone ahead of them. But since the actions are strictly a function of the underlying information, it is only the information that matters.

Let X_i be the investment decision of the i-th person. It is allowed to be a function of the $A_1, A_2, \ldots, A_i$.

It turns out in this case that the best thing to do is to choose X_i to agree with the majority of the A_k observed up to this time. Ties can be broken arbitrarily — you can pick whatever rule you like for tiebreaking.

If people follow this rule, give an exact expression for and then the best upper bound you can for $P(X_i \neq Q)$?

Hint: distinguish between the case of even and odd i in writing the answer

It is clear in this case that the best thing to do is to choose X_i to agree with the majority of the A_k observed up to this time. Ties can be broken arbitrarily — you can pick whatever rule you like for tiebreaking.

If people follow this rule, give an exact expression for and then the best upper bound you can for $P(X_i \neq Q)$?

Let us break ties by just picking $+1$ each time. It is clear that as a result, we will be wrong with probability $\frac{1}{2}$ since half the time the correct investment is -1.

It is now clear that the event

$$\{X_i \neq Q\} = \{\sum_{k=1}^{i} V_i < 0\} \cup (\{\sum_{k=1}^{i} V_i = 0\} \cap \{Q = -1\})$$

This is because for us to make a mistake either more than exactly half of the analysts have to give the wrong answer, or exactly half give the wrong answer and the right answer happens to be -1. The nice thing about this decomposition into two sets is that they are disjoint and hence the additive rule of probability applies:

$$\begin{aligned}
P(X_i \neq Q) &= P(\sum_{k=1}^{i} V_i < 0) + P(\{\sum_{k=1}^{i} V_i = 0\} \cap \{Q = -1\}) \\
&= P(\sum_{k=1}^{i} V_i < 0) + P(\sum_{k=1}^{i} V_i = 0)P(Q = -1) \\
&= P(\sum_{k=1}^{i} V_i < 0) + \frac{P(\sum_{k=1}^{i} V_i = 0)}{2} \\
&= P(\sum_{k=1}^{i} \frac{V_i + 1}{2} < \frac{i}{2}) + \frac{P(\sum_{k=1}^{i} \frac{V_i+1}{2} = \frac{i}{2})}{2}
\end{aligned}$$

Notice that the $\frac{V_i+1}{2}$ are i.i.d. Bernoulli(p) random variables and hence we have binomial probabilities above. As such, it divides into even and odd cases for i since if i is odd, then $\frac{i}{2}$ is a non integer and hence the Bernoulli can not take that value. So we have: (Letting S_i be a Binomial random variable with probability p and i terms)

$$\begin{aligned}
P(X_i \neq Q) &= P(S_i < \frac{i}{2}) + \frac{P(S_i = \frac{i}{2})}{2} \\
&= \begin{cases} P(S_i \leq \frac{i}{2} - 1) + \frac{P(S_i = \frac{i}{2})}{2} & \text{if } i \text{ is even.} \\ P(S_i \leq \lfloor \frac{i}{2} \rfloor) & \text{if } i \text{ is odd.} \end{cases} \\
&= \begin{cases} (\sum_{k=0}^{\frac{i}{2}-1} \binom{k}{i} p^k (1-p)^{i-k}) + \frac{\binom{\frac{i}{2}}{i} p^{\frac{i}{2}} (1-p)^{\frac{i}{2}}}{2} & \text{if } i \text{ is even.} \\ (\sum_{k=0}^{\lfloor \frac{i}{2} \rfloor} \binom{k}{i} p^k (1-p)^{i-k}) & \text{if } i \text{ is odd.} \end{cases}
\end{aligned}$$

To get an upper bound for these probabilities, we first notice that

$$\begin{aligned}
P(X_i \neq Q) &= P(\sum_{k=1}^{i} V_i < 0) + \frac{P(\sum_{k=1}^{i} V_i = 0)}{2} \\
&\leq P(\sum_{k=1}^{i} V_i \leq 0)
\end{aligned}$$

and then apply Chernoff style bounding. This is done by applying Markov's inequality using $Z = e^{s \sum_{k=1}^{i} V_i}$ for a free parameter $s > 0$. Then we have:

$$\begin{aligned}
P(X_i \neq Q) &\leq P(\sum_{k=1}^{i} V_i \leq 0) \\
&= P(e^{s \sum_{k=1}^{i} V_i} \leq e^0) \\
&\leq \frac{E[e^{s \sum_{k=1}^{i} V_i}]}{e^0} \\
&= E[\prod_{k=1}^{i} e^{sV_i}] \\
&= \prod_{k=1}^{i} E[e^{sV_i}] \\
&= (pe^s + (1-p)e^{-s})^i
\end{aligned}$$

Since this holds for every $s > 0$ we are free to minimize it over s. We do this by taking the derivative and setting it to zero. This tells us that the unique critical point is the solution to

$$pe^s = (1-p)e^{-s}$$

and so

$$e^s = \sqrt{\frac{1-p}{p}}$$

Plugging that in we get:

$$P(X_i \neq Q) \leq (2\sqrt{p(1-p)})^i$$

for our bound.

This goes to zero exponentially in i.

Problem 1.3 *(55pts) Quantization Error and Dithering*

You are designing a system for doing signal processing, but you need to encode your sampled real-valued signal X and send some representative $\hat{X}$ over your limited capacity communication channel. Doing this will entail some error and so you decide to think of $\hat{X}$ as $\hat{X} = X - \tilde{X}$ where $\tilde{X}$ represents the random error that your encoding introduces.

To do this, you have decided to employ a module $Q(a)$ that quantizes a to the nearest integer q — in other words it chooses the integer q that minimizes $|q - a|$.(breaking ties however you would like)

The difficulty is that you do not know the distribution for X. It could be anything.

a. 10pts. Your first attempt is to just use the module directly and let $\hat{X} = Q(X)$. Show that for this choice of encoding, the error $\tilde{X}$ need not even be zero-mean if X has an unfortunate distribution.

Use the constant random variable $X = 1.2$. For this $\hat{X} = Q(1.2) = 1$ and hence the error $\tilde{X} = 1.2 - 1 = 0.2$ is also a nonzero constant. A nonzero constant certainly does not have a zero mean.

b. 20pts. You realize that you have access to a uniform random variable V on $[-0.5, 0.5]$ at the encoder that is independent of X. You decide to use this to dither and then quantize $\hat{X} = Q(X+V)$. Show that for this choice of encoding, the error $\tilde{X}$ is always zero-mean, but it need not be independent of X.

We first show that it need not to be independent of $\tilde{X}$ by the following example. Let X be a random variable that is 1.2 with probability $\frac{1}{2}$ and -1.5 with probability $\frac{1}{2}$.

Then when $X = 1.2$, we have $\hat{X} = 1, \tilde{X} = -0.2$ whenever $Q(1.2 + V) = 1$ which occurs whenever $0.5 < 1.2 + V \leq 1.5$ which occurs whenever $-0.7 < V \leq 0.3$. This has a probability $0.5 + 0.3 = 0.8$. The other 0.2 of the time we have $\hat{X} = 2, \tilde{X} = 0.8$. Meanwhile, when $X = -1.5$, we have $\hat{X} = -1, \tilde{X} = 0.5$ whenever $Q(-1.5 + V) = 1$ which occurs whenever $-1.5 < -1.5 + V \leq -0.5$ which occurs whenever $0 < V \leq 0.5$. This has probability $0.5+0 = 0.5$. The other 0.5 of the time we have $\hat{X} = -2, \tilde{X} = -0.5$

In this example, knowing $\hat{X} > 0$ tells us that $P(\tilde{X} = 0.5) = 0$ while knowing $\hat{X} < 0$ tells us that $P(\tilde{X} = 0.5) = \frac{1}{2}$. So they are not independent in this example.

To see that the error $\tilde{X}$ is zero mean, we first condition on X and show that the conditional expectation $E[\tilde{X}|X = x] = 0$ no matter what x is. Then $E[\tilde{X}] = E_x[E[\tilde{X}|X = x]] = E_x[0] = 0$. To see that $E[\tilde{X}|X = x] = 0$:

$$\begin{aligned}
E[\tilde{X}|X = x] &= E[X - \hat{X}|X = x] \\
&= x - E[Q(x + V)|X = x] \\
&= x - \int_{-0.5}^{0.5} Q(x + v)dv
\end{aligned}$$

Hence $Q(x + v) = \lfloor x \rfloor$ whenever $x + v \leq \lfloor x \rfloor + 0.5$ and $Q(x + v) = \lfloor x \rfloor + 1$ whenever $x + v > \lfloor x \rfloor + 0.5$. We can pull the common $\lfloor x \rfloor$ out to get:

$$\begin{aligned}
E[\tilde{X}|X = x] &= x - \int_{-0.5}^{0.5} Q(x + v)dv \\
&= x - \lfloor x \rfloor - \int_{x+v>\lfloor x \rfloor+0.5} 1dv \\
&= x - \lfloor x \rfloor - \int_{\lfloor x \rfloor-x+0.5}^{0.5} 1dv \\
&= (x - \lfloor x \rfloor) + \int_{0}^{\lfloor x \rfloor-x} 1dv \\
&= (x - \lfloor x \rfloor) + (\lfloor x \rfloor - x) = 0
\end{aligned}$$

c. 25pts. *A flash of insight hits you as you realize that the decoder can also be made to have access to V. So you decide to subtract the dither at the end and in effect encode $\hat{X} = Q(X+V) - V$. Show that for this choice of encoding, the error $\tilde{X}$ is zero mean and independent of X regardless of what distribution X has.*

The zero mean is easy to see since:

$$E[\tilde{X}] = E[(X - Q(X+V)) + V] = E[X - Q(X+V)] + E[V]$$

and from part (b) we know that $E[X - Q(X+V)] = 0$ and $E[V] = 0$, since it is uniform and symmetric around 0.

To see that $\tilde{X}$ is independent of X, we will show that no matter what the value for $X = x$ is, $\tilde{X}$ conditioned on it is a uniform random variable on [-0.5, 0.5]. To see this we are going to use convex combinations in an interesting way. But first, we will introduce a little notation to make it easier to see what is going on. For any real x, let $\tilde{X}_x$ be the random variable defined as follows:

$$\tilde{X}_x = x - Q(x+V) + V$$

It should be clear that $\tilde{X}_x$ has the same density as the density for $\tilde{X}$ conditioned on $X = x$. From the reasoning in part (b), we know that $Q(x+V)$ is a discrete random variable that can only take values $\lfloor x \rfloor$ and $(\lfloor x \rfloor + 1)$. Furthermore, it takes them on the events $\{-0.5 \geq v \leq \lfloor x \rfloor + 0.5 - x\}$ and $\{\lfloor x \rfloor + 0.5 - x] \geq v \leq 0.5\}$ respectively. As such, we can define a new indicator random variable I_x that is 0 whenever $\{-0.5 \geq v \leq \lfloor x \rfloor + 0.5 - x\}$ and 1 otherwise. Then, we can use this to rewrite X_x as:

$$\tilde{X}_x = x - \lfloor x \rfloor - I_x + V$$

and use conditioning to get:

$$\begin{aligned} f_{\tilde{X}|X}(\tilde{x}|x) = &= P(I_x = 1) f_{\tilde{X}|X,I_x=1}(\tilde{x}|x) + P(I_x = 0) f_{\tilde{X}|X,I_x=0}(\tilde{x}|x) \\ &= (\lfloor x \rfloor + 1 - x) f_{\tilde{X}|X,I_x=1}(\tilde{x}|x) + (x - \lfloor x \rfloor) f_{\tilde{X}|X,I_x=0}(\tilde{x}|x) \end{aligned}$$

But conditioned on $I_x = 1$, we know that V is a uniform random variable on $[\lfloor x \rfloor - x + 0.5, 0.5]$ and so $\tilde{X}_x = (x - \lfloor x \rfloor) - 1 + V$ is a uniform random variable on $[-0.5, x - \lfloor x \rfloor - 0.5]$. Meanwhile, conditioned on $I_x = 0$, we know that V is a uniform random variable on $[0.5, \lfloor x \rfloor - x + 0.5)$ and so $\tilde{X}_x = (x - \lfloor x \rfloor) - 1 + V$ is a uniform random variable on $[x - \lfloor x \rfloor - 0.5, 0.5)$. So taking the convex combination of them we have:

$$
\begin{aligned}
f_{\tilde{X}|X}(\tilde{x}|x) = &= (\lfloor x \rfloor + 1 - x) f_{\tilde{X}|X,I_x=1}(\tilde{x}|x) + (x - \lfloor x \rfloor) f_{\tilde{X}|X,I_x=0}(\tilde{x}|x) \\
&= \begin{cases} 0 & \textit{if } \tilde{x} < -0.5 \\ (\lfloor x \rfloor + 1 - x)(\frac{1}{\lfloor x \rfloor + 0.5 - x - (-0.5)}) & \textit{if } -0.5 \leq \tilde{x} \leq \lfloor x \rfloor + 0.5 - x \\ (x - \lfloor x \rfloor)(\frac{1}{0.5 - (\lfloor x \rfloor + 0.5 - x)}) & \textit{if } \lfloor x \rfloor + 0.5 - x < \tilde{x} \leq 0.5 \\ 0 & \textit{if } \tilde{x} > 0.5 \end{cases} \\
&= \begin{cases} 0 & \textit{if } \tilde{x} < -0.5 \\ 1 & \textit{if } -0.5 \leq \tilde{x} \leq 0.5 \\ 0 & \textit{if } \tilde{x} > 0.5 \end{cases}
\end{aligned}
$$

Thus $\tilde{X}_x$ is uniform between $[-0.5, 0.5]$ no matter what the value for $X = x$ is. Since the density does not change with x, the two random variables are independent.

Problem 1.4 *(45 pts) Least Squares Estimation*

The continuous random variables X and Y have joint pdf given by:

$$f_{X,Y}(x,y) = \begin{cases} c & \textit{if (x,y) belongs to the shaded area} \\ 0 & \textit{otherwise} \end{cases}$$

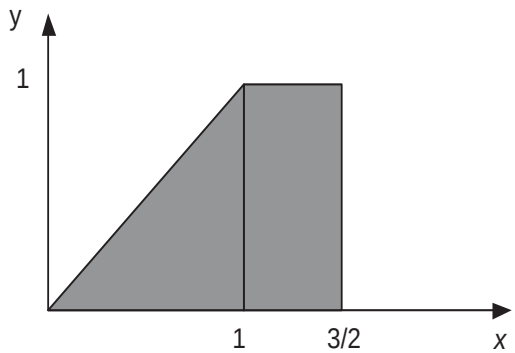

Figure 3: Joint PDF.

a. 15pts. *Find the best minimum mean squares estimate (MMSE) of Y given that $X = x$, for all possible values of x.*

X, Y are jointly uniform over the shaded area S, so the constant c can be easily found from the normalization constraint:

$$\iint_{(x,y)\in S} f(x,y)dxdy = 1 \Rightarrow c = 1$$

The best minimum mean squares estimate of Y given $X = x$ is $E[Y|X = x]$. The conditional distribution of Y give $X = x$ is uniform over the range of possible values of Y, and the conditional expectation is the midpoint. In particular:

$$E[Y|X = x] = \begin{cases} \frac{1}{2}x & \textit{if } 0 \le x \le 1 \\ \frac{1}{2} & \textit{if } 1 \le x \le \frac{3}{2} \end{cases}$$

b. 15pts. Find the best linear least squares estimate (LLSE) of Y given that $X = x$, for all possible values of x.

Before we compute the best linear estimator lets compute the densities for X and Y.

First conditional pdf of X given Y is uniform and given by:

$f_{X|Y}(x|y) = \frac{1}{1.5-y}$ for $x \in [y, 1.5]$

Similarly conditional pdf of Y given X is :

$f_{Y|X}(y|x) = \frac{1}{x}$ for $y \in [0, x]$ and $f_{Y|X}(y|x) = 1$ for $x \in [1, 1.5]$

Then pdf of X and pdf of Y can be found as:

$$f_X(x) = \frac{f_{X,Y}(x,y)}{f_{Y|X}(y|x)} = \begin{cases} x & \textit{if } 0 \le x \le 1 \\ 1 & \textit{if } 1 \le x \le \frac{3}{2} \\ 0 & \textit{otherwise} \end{cases}$$

$$f_Y(y) = \frac{f_{X,Y}(x,y)}{f_{X|Y}(x|y)} = \begin{cases} 1.5 - y & \textit{if } 0 \le y \le 1 \\ 0 & \textit{otherwise} \end{cases}$$

The best linear least squares estimator has a form of $\hat{Y} = aX + b$. Optimal coefficients a, b can be found by minimizing the mean square error:

$$E[(Y - aX - b)^2]$$

i.e. by setting its partial derivatives with respect to a and b to zero, which results into $E[(Y - aX - b)X] = 0$ and $E[Y - aX - b] = 0$ (i.e. orthogonality principle).

$$aE[X^2] + bE[X] = E[XY]$$
$$aE[X] + b = E[Y]$$

$$E[X] = \int_0^1 x^2 dx + \int_1^{1.5} x dx = 0.96$$
$$E[X^2] = \int_0^1 x^3 dx + \int_1^{1.5} x^2 dx = 1.04$$
$$E[Y] = \int_0^1 (1.5 - y) y dy = 0.42$$
$$E[XY] = \iint_{(x,y)\in S} xy dx dy = \int_0^1 \int_0^x xy dx dy + \int_1^{1.5} \int_0^1 xy dx dy = 0.44$$

By solving this system of linear equations we find that optimal $a = 0.31$ and $b = 0.12$, and the best linear estimator of Y given $X = x$ is $\hat{Y} = 0.31x + 0.12$

c. 15pts. *Find the best quadratic least squares estimate (QLSE) of Y given that $X = x$, for all possible values of x. The quadratic least squares estimator is allowed to have the form $ax^2 + bx + c$ where a, b, c are real numbers.*

Similarly to part b. the best quadratic least squares estimate $\hat{Y} = aX^2 + bX + c$ and its optimal coefficients a, b, c can be found by minimizing the mean square error:

$$E[(Y - aX^2 - bX - c)^2]$$

i.e. by setting its partial derivatives with respect to a, b and c to zero, which results into $E[(Y - aX^2 - bX - c)X^2] = 0$, $E[(Y - aX^2 - bX - c)X] = 0$ and $E[Y - aX^2 - bX - c] = 0$ (i.e. orthogonality principle).

$$aE[X^4] + bE[X^3] + cE[X^2] = E[X^2Y]$$
$$aE[X^3] + bE[X^2] + cE[X] = E[XY]$$
$$aE[X^2] + bE[X] + c = E[Y]$$

In order to solve this system we need to compute additional moments of X and Y

$$E[X^3] = \int_0^1 x^4 dx + \int_1^{1.5} x^3 dx = 1.22$$
$$E[X^4] = \int_0^1 x^5 dx + \int_1^{1.5} x^4 dx = 1.48$$
$$E[X^2Y] = \iint_{(x,y)\in S} x^2 y dx dy = \int_0^1 \int_0^x x^2 y dx dy + \int_1^{1.5} \int_0^1 x^2 y dx dy = 0.5$$

By solving this system of linear equations we find that optimal $a = 0.35$, $b = -0.34$, and $c = 0.38$ and the best linear estimator of Y given $X = x$ is $\hat{Y} = 0.35x^2 - 0.34x + 0.38$