
Final Exam

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

Rules.

- 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 170 minutes to complete it.
- The exam is not open book; we are giving you a cheat sheet. No calculators or phones allowed.
- Unless otherwise stated, all your answers need to be justified. Show all your work to get partial credit.
- Maximum you can score is 132 but 100 points is considered perfect.

Problem	points earned	out of
Problem 1		50
Problem 2		18
Problem 3		12
Problem 4		12
Problem 5		20
Problem 6		20
Total		132

Problem 1: Answer these questions briefly but clearly.

(a)[6] **CLT True / False:** No justification is required.

- $\lim_{n \rightarrow \infty} P(\text{Binomial}(n, p) > np) = p$ ☐ True ☐ False
- $\lim_{n \rightarrow \infty} P(\text{Poisson}(n) > n) = \frac{1}{2}$ ☐ True ☐ False
- $\lim_{n \rightarrow \infty} P(\text{Exponential}(n) > \frac{1}{n}) = \frac{1}{2}$ ☐ True ☐ False

- **False** because it should be $\frac{1}{2}$.
- **True** because a Poisson RV with rate n is really the sum of n independent Poisson RVs with rate 1. So if we let $X_i \stackrel{iid}{\sim} \text{Poisson}(1)$ then we can rewrite limit as

$$\lim_{n \rightarrow \infty} P(n \cdot \frac{1}{n} \sum_{i=1}^n X_i > n \cdot 1) = \lim_{n \rightarrow \infty} P(\frac{1}{n} \sum_{i=1}^n X_i > 1)$$

In the limit, this is equal to the probability that a normally distributed RV is larger than its mean, which is $\frac{1}{2}$.

We can also write it differently to make CLT explicit: (since $\frac{1}{n} \sum_{i=1}^n X_i$ converges to a constant a.s.; not to normal)

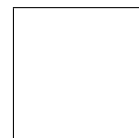
$$\lim_{n \rightarrow \infty} P(\sum_{i=1}^n (X_i - 1) > 0) = \lim_{n \rightarrow \infty} P(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - 1) > 0)$$

- **False** because by the CDF of the exponential distribution this is equal to

$$\lim_{n \rightarrow \infty} 1 - e^{-n \cdot \frac{1}{n}} = 1 - e^{-1}$$

Unlike the previous part, here the exponential distribution can't be decomposed into independent sums which means we can't apply the CLT.

(b)[3] **Order Statistic:** Given that the 5th arrival time of a $\text{Poisson}(\lambda)$ process with $\lambda = 10$ occurs at time $t = 1$ second, what is the expected arrival time of the 2nd arrival?



The expected arrival time is 0.4 seconds, as the four arrivals in between are arranged according to the order-statistic of 4 uniform random variables between 0 and 1.

(c)[3] **MMSE Sanity Check:** Assume that X and Y are two random variables such that $\mathbb{E}[X|Y] = L[X|Y]$. Then it must be that (choose the correct answers, if any):

- ☐ X and Y are jointly Gaussian.
- ☐ X can be written as $X = aY + Z$, where Z is a random variable independent of Y .
- ☐ $\mathbb{E}((X - L[X|Y])Y^k) = 0 \quad \forall k \geq 0$
- ☒ **False.** If $X = Y$ then $E[X | Y] = L[X | Y]$ but X, Y can be anything.
- ☒ **False,** $X = YZ$ where Z is independent of Y .
- ☒ **True.** Since $L[X | Y] = E[X | Y]$, we know that $E[(X - L[X | Y])f(Y)] = 0$ for any function f of Y . Y^k is a function of Y .

(d)[4] **MMSE:** Let X and Y be independent Gaussian random variables each with mean zero, and $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$. Find $\mathbb{E}[X|e^{X+Y}]$.

Since X and Y are jointly Gaussian, we have

$$\mathbb{E}[X|X+Y] = L[X|X+Y] = \mathbb{E}[X] + \frac{\mathbb{E}[(X+Y)X]}{Var(X+Y)}(X+Y) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}(X+Y).$$

Also, since $X+Y$ and e^{X+Y} are one-to-one mappings, the MMSE estimate of $\mathbb{E}[X|e^{X+Y}] = \mathbb{E}[X|X+Y]$. Let $Z = e^{X+Y}$ and, hence, $X+Y = \log Z$. Thus,

$$\mathbb{E}[X|e^{X+Y} = Z] = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}(\log Z).$$

(e)[5] **Random Graph on a Random Graph:** Suppose we generate a random graph by starting with an Erdos-Renyi graph $G(n, p)$. Then, we generate a random graph using the Erdos-Renyi model again on the subgraph of singletons (that is, each edge between two singletons is added with probability p). Calculate the expected number of edges in total.

Every edge has probability $p + (1 - p)^{2n-3}p$ of being in the graph. The first term p comes from the initial random graph. For the subgraph, both of the two vertices need to be singletons, which mean that none of the $(n - 1) + (n - 1) - 1 = 2n - 3$ edges connected to the the two vertices can be in the initial graph. And finally we multiply by p for being connected in the subgraph. By linearity of expectation, this leads to an answer of $\binom{n}{2}(p + (1 - p)^{2n-3}p)$.

(f)[3] **MMSE:** Given three i.i.d. random variables X, Y, Z , what is $\mathbb{E}[X|X + Y + Z]$?



By symmetry, since $\mathbb{E}[X|X + Y + Z] = \mathbb{E}[Y|X + Y + Z] = \mathbb{E}[Z|X + Y + Z]$ and $\mathbb{E}[X + Y + Z|X + Y + Z] = X + Y + Z$, then $\mathbb{E}[X|X + Y + Z] = \frac{1}{3}(X + Y + Z)$.

(g)[3+3] **Jointly Gaussian:** Let X, Y and Z be jointly Gaussian random variables having covariance matrix

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

and mean vector $[0 \ 10 \ 0]^T$.

(i) Find $\mathbb{E}[Y|X, Z]$.

$$\mathbb{E}[Y|X, Z] = 10 + aX + bZ.$$

$$\begin{aligned} \mathbb{E}[(Y - 10 - aX - bZ)X] &= \mathbb{E}[(Y - 10)X] - a\mathbb{E}[X^2] - b\mathbb{E}[ZX] = 0 \\ &\Rightarrow 1 - 2a = 0. \end{aligned}$$

Also,

$$\begin{aligned} \mathbb{E}[(Y - 10 - aX - bZ)Z] &= \mathbb{E}[(Y - 10)Z] - a\mathbb{E}[XZ] - b\mathbb{E}[Z^2] = 0 \\ &\Rightarrow 1 - 2b = 0. \end{aligned}$$

Thus, $\mathbb{E}[Y|X, Z] = 10 + (X + Z)/2$.

(ii) Find $\mathbb{E}[(e^X - e^{-X})Y(\sin Z)]$. (*Hint:* Condition on (X, Z) .)

Write

$$\mathbb{E}[(e^X - e^{-X})Y(\sin Z)] = \mathbb{E}[\mathbb{E}[(e^X - e^{-X})Y(\sin Z)|X, Z]] = \mathbb{E}[(e^X - e^{-X})(\sin Z) \mathbb{E}[Y|X, Z]].$$

Further, we have $\mathbb{E}[Y|X, Z] = 10 + aX + bZ$ for some constants a and b . We thus get

$$\mathbb{E}[(e^X - e^{-X})Y(\sin Z)] = 10 \mathbb{E}[(e^X - e^{-X})(\sin Z)] + a \mathbb{E}[X(e^X - e^{-X})(\sin Z)] + b \mathbb{E}[(e^X - e^{-X})Z(\sin Z)].$$

From the structure of the covariance matrix, we see that X and Z are independent, so this can be written as

$$\mathbb{E}[(e^X - e^{-X})Y(\sin Z)] = 2 \mathbb{E}[(e^X - e^{-X})] \mathbb{E}[(\sin Z)] + a \mathbb{E}[X(e^X - e^{-X})] \mathbb{E}[(\sin Z)] + b \mathbb{E}[(e^X - e^{-X})] \mathbb{E}[Z(\sin Z)].$$

Also, note that $\mathbb{E}[(e^X - e^{-X})]$ and $\mathbb{E}[(\sin Z)]$ are both zero because of both X and Z are zero-mean and symmetric around zero as well and the functions $\sin(Z)$ and $e^X - e^{-X}$ are odd, that is, symmetric around origin. Thus all the three summation terms in the expression of $\mathbb{E}[(e^X - e^{-X})Y(\sin Z)]$ are zero.

(h)[5] Neyman Pearson testing: Ray's posts on Piazza can be modeled as a Poisson process. Let its rate be λ_0 according to the null hypothesis H_0 and λ_1 according to the alternate hypothesis H_1 , where $\lambda_1 > \lambda_0$. Say you observe the first post at time y_1 . Describe the optimal Neyman Pearson (NP) hypothesis test for this problem. Assume the maximum probability of false alarm is ϵ , where $0 < \epsilon < 1$.

The likelihood function

$$\ell(y_1) = \frac{f_{H_1}(y_1)}{f_{H_0}(y_1)} = \frac{\lambda_1 e^{-\lambda_1 y_1}}{\lambda_0 e^{-\lambda_0 y_1}} = \left(\frac{\lambda_1}{\lambda_0} \right) e^{-(\lambda_1 - \lambda_0) y_1}.$$

Hence, the likelihood function is a decreasing function of y_1 . Let the optimal threshold be \hat{y}_1 , where we say H_1 is true if $y_1 < \hat{y}_1$ and say H_0 is true if $y_1 > \hat{y}_1$ according to the NP test. Note that the probability that $y_1 = \hat{y}_1$ is zero since we are dealing with continuous random variables, and hence randomization in the NP test is not required.

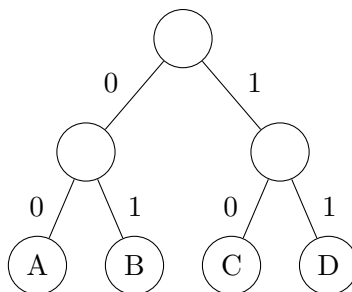
Now, to find y_1 , we use the fact that the probability of false alarm is ϵ , that is

$$\begin{aligned} \mathbb{P}_{H_0}(y_1 < \hat{y}_1) &= \epsilon \\ \Rightarrow \int_0^{\hat{y}_1} f_{H_0}(y_1) dy_1 &= \epsilon \\ \Rightarrow \int_0^{\hat{y}_1} \lambda_0 e^{-\lambda_0 y_1} dy_1 &= \epsilon \\ \Rightarrow -e^{-\lambda_0 y_1} \Big|_0^{\hat{y}_1} &= \epsilon \\ \Rightarrow 1 - e^{-\lambda_0 \hat{y}_1} &= \epsilon \\ \Rightarrow \hat{y}_1 &= \frac{-1}{\lambda_0} \log(1 - \epsilon). \end{aligned}$$

Alternate solution: Can also solve for \hat{y}_1 using the fact that false alarm occurs when the number of arrivals (or Piazza posts) in time $[0, \hat{y}_1]$ is ≥ 1 when H_0 is true, that is, $\mathbb{P}(\text{Poisson}(\hat{y}_1 \lambda_0) \geq 1) = \epsilon$.

(i)[3] Huffman Tree:

Let X be a discrete random variable taking on 4 values A , B , C , and D . If we were to encode X with Huffman encoding, the resulting Huffman tree would look like this.



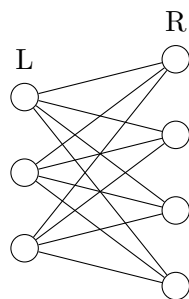
True or False: $H(X) \geq 1$? Justify.

The length of the codewords for A , B , C , and D are all 2, and therefore the average length of the code is 2. Since we know that Huffman coding will produce a average codeword length that is between $H(X)$ and $H(X) + 1$, we know that $H(X) \geq 2 - 1 = 1$.

Alternate answer without invoking optimality: From the shape of the tree, we know that $P(X = x) \leq \frac{1}{2}$ for all x (otherwise, if there was an x that had such a high probability, it would be a leaf attached to the root.) This tells us that $-\log_2 P(X = x) \geq 1$ for each x , and multiplying by $P(X = x)$ and summing up over all x gives us the answer.

(j)[3] Bipartite Markov Chain: Suppose in a bipartite graph you have two sets of nodes, L and R , of sizes m and n , such that for each $u \in L$ and $v \in R$, the transition probability from u to v is $1/n$, and the transition probability from v to u is $1/m$. Calculate the stationary distribution.

(Note: A bipartite graph has two sets of nodes where nodes in each set are only connected to the nodes in the other set.)



Complete Bipartite Graph for $m = 3, n = 4$

By symmetry, it is $\frac{1}{2m}$ on L and $\frac{1}{2n}$ on R , since the chain spends half the time on L and the other half on R , and is uniform within each set.

(k)[4] MLE with Numbered Balls: A box is filled with N balls numbered 1 through N . I randomly select K balls from the box. I order the balls in ascending order of their numbers and find them to be x_1, x_2, \dots, x_K . What is the maximum likelihood estimate of N given my K observations? Justify your answer to get credit.

We want to maximize $\mathbb{P}(x_1, \dots, x_K \mid N)$. N must be at least the highest number x_K , but beyond that, every sequence x_1, \dots, x_K is equally likely given N . That is, $\mathbb{P}(x_1, \dots, x_K \mid N) = \frac{1}{\binom{N}{K}}$ (because for each choice of K , there is exactly one ascending order and each choice is equally likely) This is a decreasing function in N so we want N to be as small as possible, so $MLE[N \mid x_1, \dots, x_K] = x_K$.

(1)[2+3] Fun with Gaussians:

(i) Show that the sum of two independent Gaussian random variables is Gaussian.

Consider two Gaussians $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

We know that $M_{X+Y}(s) = M_X(s)M_Y(s)$ as X, Y are independent. So we have

$$M_{X+Y}(s) = \exp\{\mu_1 s + \sigma_1^2 s^2/2\} \exp\{\mu_2 s + \sigma_2^2 s^2/2\} = \exp\{(\mu_1 + \mu_2)s + (\sigma_1^2 + \sigma_2^2)s^2/2\}$$

which we identify as the MGF of $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

(ii) Show that the sum of two jointly Gaussian random variables is Gaussian, starting from the definition that two jointly Gaussian random variables X, Y can be written as linear combinations of underlying independent standard Gaussians Z_1, Z_2 , i.e,

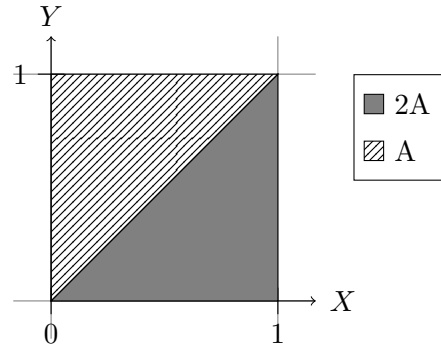
$$X = a_X Z_1 + b_X Z_2 + \mu_X$$

$$Y = a_Y Z_1 + b_Y Z_2 + \mu_Y$$

$X + Y = (a_X + a_Y)Z_1 + (b_X + b_Y)Z_2 + (\mu_X + \mu_Y)$ Since scaling Z_1 and Z_2 still yields independent Gaussians, the sum of the first two terms is Gaussian. The last term simply shifts the mean so $X + Y$ is Gaussian.

Problem 2 [2+5+4+7]: Graphical Density

Let X and Y have joint PDF as depicted below.



- (a) Determine the value of A .

For this to be a valid density, we need $A \left(\frac{1}{2}\right) + 2A \left(\frac{1}{2}\right) = 1$, so $A = \frac{2}{3}$.

- (b) Compute $E[X|Y]$.

For $0 \leq y \leq 1$,

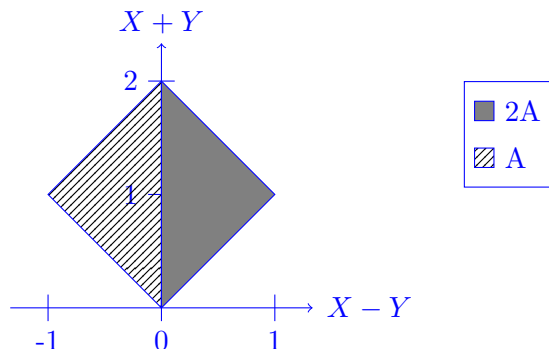
$$\mathbb{E}[X|Y = y] = \int_0^1 x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx = \int_0^y \frac{2/3}{4/3 - 2y/3} x dx + \int_y^1 \frac{4/3}{4/3 - 2y/3} x dx = \frac{2 - y^2}{4 - 2y}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y) dx = \int_0^y \frac{2}{3} dx + \int_y^1 \frac{4}{3} dx = \frac{4}{3} - \frac{2}{3}y$$

Therefore, $\mathbb{E}[X|Y] = \frac{2-Y^2}{4-2Y}$.

- (c) Compute $L[X + Y|X - Y]$.

Solution 1: The joint density of $X - Y$ and $X + Y$ is shown below:



We see from the diagram that $\mathbb{E}[X+Y|X-Y] = 1$, which is linear, so $L[X+Y|X-Y] = 1$.

Solution 2:

$$L[X+Y|X-Y] = \mathbb{E}[X+Y] + \frac{\text{cov}(X+Y, X-Y)}{\text{Var}(Y)}(X-Y - \mathbb{E}[X-Y])$$

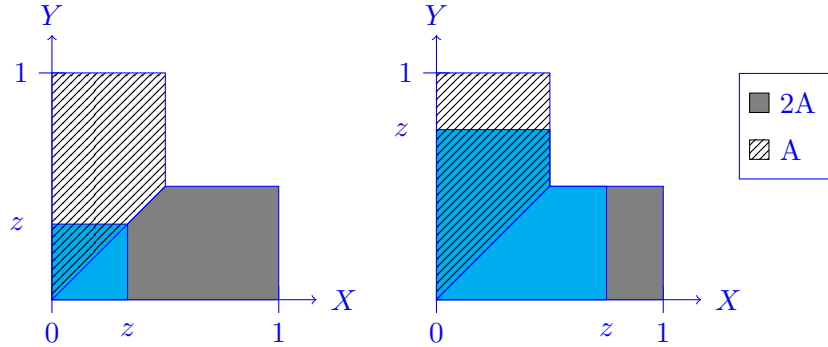
We see from the original diagram that Y and $1-X$ are identically distributed, so $\mathbb{E}[Y] = 1 - \mathbb{E}[X]$, which means $\mathbb{E}[X+Y] = 1$. Furthermore, we must also have $\text{Var}(Y) = \text{Var}(1-X)$, so $\text{cov}(X+Y, X-Y) = \text{Var}(X) - \text{Var}(Y) = \text{Var}(X) - \text{Var}(1-X) = \text{Var}(X) - \text{Var}(X) = 0$.

Therefore, $L[X+Y|X-Y] = 1$.

- (d) Compute $\mathbb{E}[\max(X, Y) | \min(X, Y) \leq 0.5]$.

Solution 1:

Let $Z = \max(X, Y)$. We will first graphically compute the conditional CDF of Z , i.e. $\mathbb{P}(Z \leq z | \min(X, Y) \leq 0.5)$. As shown below, there are two cases to consider: $0 \leq z < 0.5$ and $0.5 \leq z \leq 1$.



The blue region occupies the same fraction in both the regions with pdfs A and $2A$, so it suffices to take ratios of areas in one of the regions.

Therefore,

$$\begin{aligned} \mathbb{P}(Z \leq z | \min(X, Y) \leq 0.5) &= \begin{cases} \frac{z^2/2}{3/8} & 0 \leq z < \frac{1}{2} \\ \frac{1/8 + (z-1/2)/2}{3/8} & \frac{1}{2} \leq z \leq 1 \end{cases} \\ &= \begin{cases} \frac{4}{3}z^2 & 0 \leq z < \frac{1}{2} \\ \frac{4}{3}z - \frac{1}{3} & \frac{1}{2} \leq z \leq 1 \end{cases} \end{aligned}$$

$$f_{Z | \min(X, Y) \leq 0.5}(z) = \begin{cases} \frac{8}{3}z & 0 \leq z < \frac{1}{2} \\ \frac{4}{3} & \frac{1}{2} \leq z \leq 1 \end{cases}$$

and the expectation is

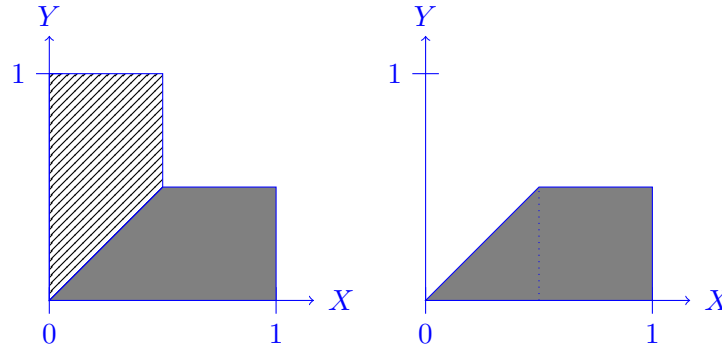
$$\int_0^{1/2} z \left(\frac{8}{3}z \right) dz + \int_{1/2}^1 \frac{4}{3}z dz = \frac{11}{18}$$

Solution 2:

By the law of total expectation,

$$\begin{aligned}
\mathbb{E}[Z | \min(X, Y) \leq 0.5] &= \mathbb{E}[Z | \min(X, Y) \leq 0.5, X < Y] \mathbb{P}(X < Y | \min(X, Y) \leq 0.5) + \\
&\quad \mathbb{E}[Z | \min(X, Y) \leq 0.5, X \geq Y] \mathbb{P}(X \geq Y | \min(X, Y) \leq 0.5) \\
&= \mathbb{E}[Y | X \leq 0.5, X < Y] \mathbb{P}(X < Y | \min(X, Y) \leq 0.5) + \\
&\quad \mathbb{E}[X | Y \leq 0.5, X \geq Y] \mathbb{P}(X \geq Y | \min(X, Y) \leq 0.5)
\end{aligned}$$

By symmetry across the line $Y = X$ (the conditional pdf of Y given $X \leq 0.5, X < Y$ is the same as the same as the conditional pdf of X given $Y \leq 0.5, Y \leq X$), the two conditional expectations above are equal.



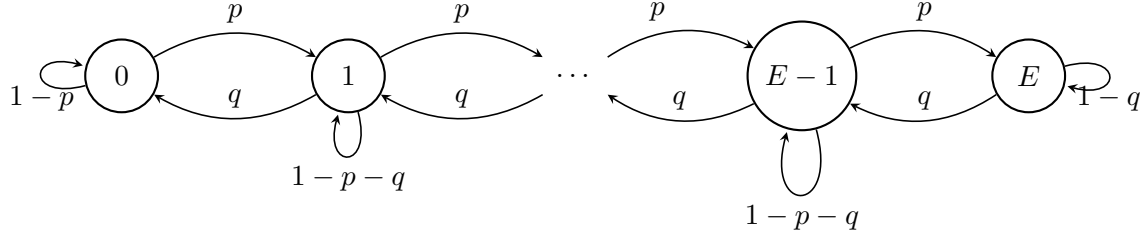
Therefore, the desired expectation is $\mathbb{E}[X | Y \leq 0.5, X \geq Y]$. Applying the law of total expectation again, $\mathbb{E}[X | Y \leq 0.5, X \geq Y] = \mathbb{E}[X | Y \leq 0.5, X \geq Y, X < 0.5] \mathbb{P}(X < 0.5 | Y \leq 0.5, X \geq Y) + \mathbb{E}[X | Y \leq 0.5, X \geq Y, X \geq 0.5] \mathbb{P}(X \geq 0.5 | Y \leq 0.5, X \geq Y)$ (splitting into the triangle and square as in the right diagram).

We graphically see $\mathbb{P}(X < 0.5 | Y \leq 0.5, X \geq Y) = \frac{1}{3}$, $\mathbb{P}(X > 0.5 | Y \leq 0.5, X \geq Y) = \frac{2}{3}$, and $\mathbb{E}[X | Y \leq 0.5, X \geq Y, X \geq 0.5] = \frac{3}{4}$. Furthermore, we can compute $\mathbb{E}[X | Y \leq 0.5, X \geq Y, X < 0.5] = \int_0^{\frac{1}{2}} (8x)x \, dx = \frac{1}{3}$.

So the desired expectation is $\frac{1}{3} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{2}{3} = \frac{11}{18}$.

Problem 3 [4+4+4]: Markov Gainz

Ray has an energy level $X_n \in \{0, 1, \dots, E\}$ units on the n -th day. Every day, with probability p , he takes a good rest which increases his energy level by 1 unit, and with probability q , he parties which decreases his energy level by 1 unit. Otherwise, the energy level remains the same. The energy level X_n can be described by the following birth-death chain:



Ray goes to the gym every day, and does $Y_n = \begin{cases} \text{Poisson}(X_n) & \text{if } X_n > 0 \\ 0 & \text{if } X_n = 0 \end{cases}$ bench presses.

- (a) Ray has an energy level of 0 units today. How many days on average will it take for him to have energy level of 2 units (assume that $E > 2$)?

We can set the hitting time equations and solve for them to get that the hitting time is $h(0) = p * h(1) + (1-p)h(0) + 1$, $h(1) = q * h(0) + (1-p-q) * h(1) + 1$, and we get that $h(0) = h(1) + \frac{1}{p}$, $h(1) = q(h(1) + \frac{1}{p}) + (1-p-q)h(1) + 1$, and then we get $h(0) = \frac{1}{p} + \frac{1}{p}(1 + \frac{q}{p})$

- (b) **For this part only**, assume that $E = \infty$, and $p/q \leq 1$. Suppose Ray has been going to the gym for a very long time. How many bench presses will Ray do today in expectation?

We first note that $\mathbb{E}[Y_n] = \mathbb{E}[\mathbb{E}[Y_n | X_n]] = \mathbb{E}[X_n]$ by law of iterated expectation. So, all we need to do is to find the expectation of a birth-death chain. We know that the stationary distribution is given by a geometric distribution with parameter $1 - \frac{p}{q}$ (offset by 1, since we are starting from 0) which you can deduce by using detailed balance equations (i.e. reversibility). See discussion 7 solutions for this derivation.

So, in particular, the expectation is the reciprocal of that minus 1.

$$\mathbb{E}[X_n] - 1 = \frac{1}{1 - \frac{p}{q}} - 1 = \frac{q}{q-p} - 1 = \frac{p}{q-p}$$

.

Alternatively, you can compute the expectation from the distribution itself:

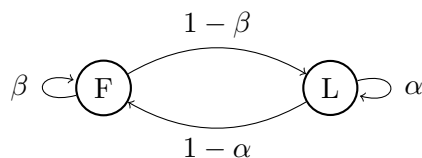
$$\pi(j) = (1 - \frac{p}{q})(\frac{p}{q})^j, j \in \mathbb{Z}$$

- (c) Ray did 0 bench presses on the n th day (that is, $Y_n = 0$). Find the ratio p/q such that the posterior of X_n (that is, $P(X_n|Y_n = 0)$) is uniform over all energy levels $\{0, \dots, E\}$. Assume that his prior on the energy levels (that is, $P(X_n)$) is the stationary distribution.

Much like before, π_i is proportional to $(\frac{p}{q})^i$. Also, probability of a Poisson being 0 is $e^{-\lambda}$, so it is e^{-i} in this example. As such, we need that $\frac{p}{q} = e$.

Problem 4 [3+5+4]: Fair and Loaded Coins

We have two indistinguishable coins, one fair and one loaded. The fair coin (F) has probability of heads 0.5 and the loaded coin (L) has probability of heads 1. We do n coin flips. For the first coin flip, we choose one of the F or L coins with equal probability. For every subsequent coin flip, the coin is chosen according to the following Markov chain:



e.g., if on coin flip j , you flipped the F coin, on coin flip $j + 1$, you have a β chance of flipping of the F coin and a $(1 - \beta)$ chance of flipping the the L coin. We want to find the MLSE of the label of the coins given an observed Heads/Tails sequence.

- (a) Suppose you observed T in the current state and H in the next state. Populate the one-stage trellis diagram shown below with appropriate costs (negative log-likelihoods as seen in lecture).

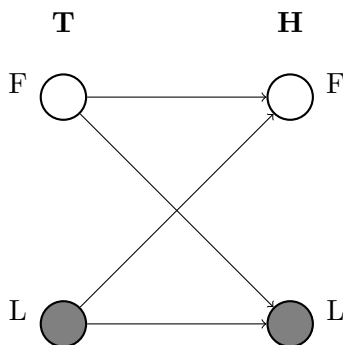


Figure 1: One “Stage” of the Trellis Diagram

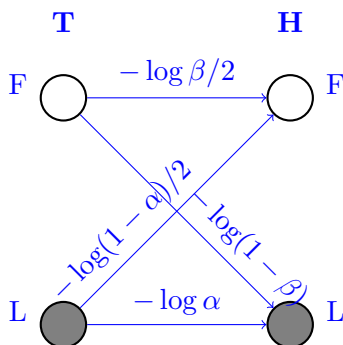


Figure 2: One “Stage” of the Trellis Diagram

- (b) Given that $n = 3$, $\beta = 3/4$, $\alpha = 1/2$ and the sequence $\{\mathbf{T}, \mathbf{H}, \mathbf{H}\}$ is observed, draw the corresponding trellis diagram for estimating the sequence of which coin was used for each flip and write down the MLSE (maximum likelihood sequence estimate) of the label of coins. (Take $\log_2 3 = 1.6$ for easier calculations)

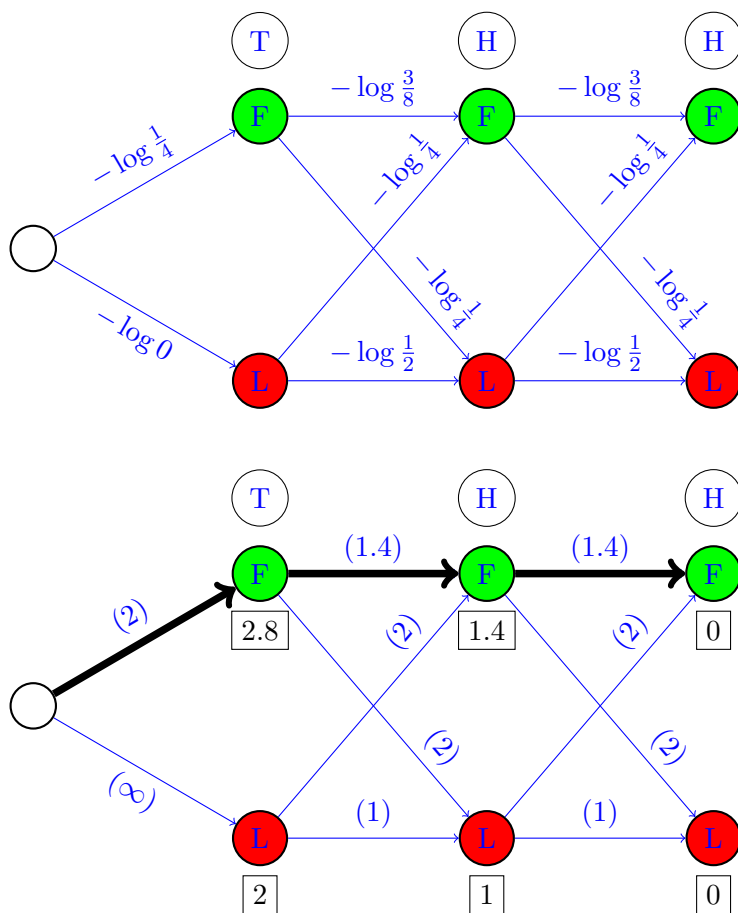


Figure 3: Obtaining the MLSE Estimate

Hence the MLSE estimate is FFF.

- (c) For part (b), what is the MLE of the coin label for the second flip?

We want to choose the max of $P(\text{Observing THH} \mid \text{2nd coin is F})$ and $P(\text{Observing THH} \mid \text{2nd coin is L})$. One thing to note is that the 1st coin must be F. $P(\text{THH} \mid \text{F})$ (abbreviated notation) is $\frac{1}{2} \cdot \frac{1}{2}$ for the first two flips then $\frac{3}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 1$ for the 3rd coin flip. So we get $\frac{1}{2} \cdot \frac{1}{2} \cdot (\frac{3}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 1) = \frac{5}{32}$. On the other hand $P(\text{THH} \mid \text{L}) = \frac{1}{2} \cdot 1 \cdot (\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2}) = \frac{3}{8} = \frac{12}{32}$. So the MLE for the 2nd coin is L, even though in the MLSE it is F.

Problem 5 [3+3+3+6+5]: Kalman Filters and LLSE

Suppose we have the following dynamical system of equations:

$$X_n = \rho X_{n-1} + V_n, \quad n = 2, 3, \dots \text{ (with } X_1 = V_1)$$

$$Y_n = X_n + W_n, \quad n = 1, 2, \dots$$

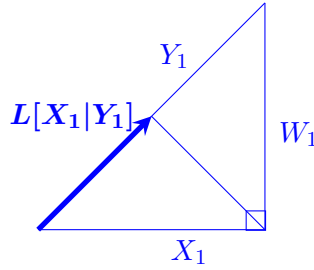
where V_n , and W_n for $n = 1, 2, \dots$ are i.i.d $\mathcal{N}(0, 1)$ noise random variables and $|\rho| < 1$.

- (a) What is the variance of X_n as $n \rightarrow \infty$?

$\text{var}(X_n) = \rho^2 \text{var}(X_{n-1}) + 1$. If we expand this recurrence relation, we get a geometric series with common ratio ρ^2 . So $\lim_{n \rightarrow \infty} \text{var}(X_n) = \frac{1}{1-\rho^2}$

- (b) (i) Find $L[X_1|Y_1]$ geometrically through a vector space representation of the random variables X_1, Y_1 , and W_1 . Mark your plot clearly.

$$L[X_1|Y_1] = \frac{Y_1}{2}$$



- (ii) Find the expected mean-squared estimation error in estimating X_1 given Y_1 .

$$\sigma_{1|1}^2 = \text{var}(X_1 - \frac{Y_1}{2}) = \text{var}(\frac{X_1}{2} - \frac{W_1}{2}) = 1/4 + 1/4 = 1/2$$

- (c) Find the prediction estimate of X_2 given Y_1 , i.e. $L[X_2|Y_1]$, as well as the expected mean-squared estimation error in estimating X_2 given Y_1 .

$$L[X_2|Y_1] = L[\rho X_1 + V_2|Y_1] = \frac{\rho}{2} Y_1$$

$$\sigma_{2|1}^2 = \text{var}(X_2 - \frac{\rho Y_1}{2}) = \text{var}(\rho X_1 + V_2 - \frac{\rho Y_1}{2}) = \text{var}(\rho(X_1 - \frac{Y_1}{2}) + V_2) = \rho^2 \sigma_{1|1}^2 + 1 = \rho^2/2 + 1.$$

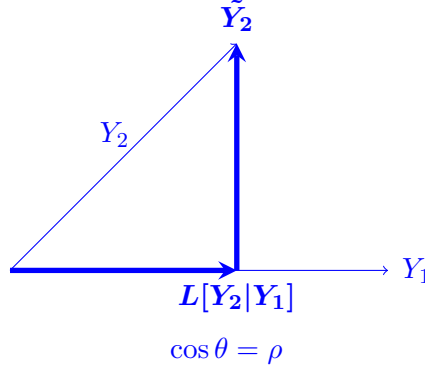
This can also be seen by directly plugging into the appropriate Kalman update equation or drawing the geometric picture as drawn in the note.

- (d) Now you want to update your estimate of X_2 given Y_1 and Y_2 by forming:

$$L[X_2|Y_1, Y_2] = L[X_2|Y_1] + L[X_2|\tilde{Y}_2]$$

- (i) What is \tilde{Y}_2 in the above equation? Express it geometrically in terms of Y_1 and Y_2 .

$$\text{The innovation } \tilde{Y}_2 = Y_2 - L[Y_2|Y_1] = Y_2 - \frac{\text{cov}(\rho X_1 + V_2 + W_2, X_1 + W_1)}{\text{var}(X_1 + W_1)} Y_1 = Y_2 - \frac{\rho}{2} Y_1$$



(ii) Find $L[X_2|\tilde{Y}_2]$ and the MMSE estimate of X_2 given Y_1 and Y_2 .

From the Kalman update equations, we know $L[X_2|\tilde{Y}_2] = k_2\tilde{Y}_2$, where the Kalman gain $k_2 = \frac{\sigma_{2|1}^2}{\sigma_{2|1}^2 + 1} = \frac{\rho^2/2+1}{\rho^2/2+2}$. This can also be seen by reconstructing the geometric picture from the notes.

We then have $L[X_2|Y_1, Y_2] = \hat{X}_{2|2} = L[X_2|Y_1] + L[X_2|\tilde{Y}_2] = \frac{\rho}{2}Y_1 + \frac{\rho^2+2}{\rho^2+4}(Y_2 - \frac{\rho}{2}Y_1)$

(iii) What is the expected mean-squared-error in estimating X_2 given Y_1 and Y_2 ? How does it compare to the estimation error in part (c)?

From the Kalman update equations, we have $\sigma_{2|2}^2 = (1 - k_2)\sigma_{2|1}^2 = \frac{2}{\rho^2+4}(\frac{\rho^2}{2} + 1) = \frac{\rho^2+2}{\rho^2+4}$. This can also be seen by reconstructing the geometric picture from the notes or by writing out the variance by hand (this is somewhat tedious).

The estimation error is less as expected because we have more observations.

(e) Now you want to further update your estimate of X_2 given Y_1 , Y_2 and Y_3 . Find $L[X_2|Y_1, Y_2, Y_3]$ and the expected mean-squared estimation error in estimating X_2 given Y_1, Y_2 , and Y_3 . How does this compare to the estimation error in parts 3 and 4?

$$\begin{aligned}
L[X_2|Y_1, Y_2, Y_3] &= L[X_2|Y_1, Y_2] + L[X_2|Y_3 - L[Y_3|Y_1, Y_2]] \\
&= \hat{X}_{2|2} + L[X_2|Y_3 - \rho\hat{X}_{2|2}] \text{ (since } L[Y_3|Y_1, Y_2] = L[\rho X_2 + V_3 + W_3|Y_1, Y_2] = \rho\hat{X}_{2|2}) \\
&= \hat{X}_{2|2} + \frac{\text{cov}(X_2, Y_3 - \rho\hat{X}_{2|2})}{\text{var}(Y_3 - \rho\hat{X}_{2|2})}(Y_3 - \rho\hat{X}_{2|2})
\end{aligned}$$

$$\begin{aligned}
\text{cov}(X_2, Y_3 - \rho\hat{X}_{2|2}) &= \text{cov}(X_2 - \hat{X}_{2|2}, Y_3 - \rho\hat{X}_{2|2}) \text{ (since } \tilde{Y}_3 \text{ is orthogonal to functions of } Y_1, Y_2 \text{ like } \hat{X}_{2|2}) \\
&= \text{cov}(X_2 - \hat{X}_{2|2}, \rho X_2 + V_3 + W_3 - \rho\hat{X}_{2|2}) \\
&= \text{cov}(X_2 - \hat{X}_{2|2}, \rho(X_2 - \hat{X}_{2|2})) \\
&= \rho\sigma_{2|2}^2
\end{aligned}$$

$$\text{var}(Y_3 - \rho\hat{X}_{2|2}) = \text{var}(\rho X_2 + V_3 + W_3 - \rho\hat{X}_{2|2}) = \text{var}(\rho(X_2 - \hat{X}_{2|2}) + V_3 + W_3) = \rho^2\sigma_{2|2}^2 + 2$$

So we have

$$L[X_2|Y_1, Y_2, Y_3] = \hat{X}_{2|2} + \frac{\rho\sigma_{2|2}^2}{\rho^2\sigma_{2|2}^2 + 2}(Y_3 - \rho\hat{X}_{2|2})$$

For the mean squared error,

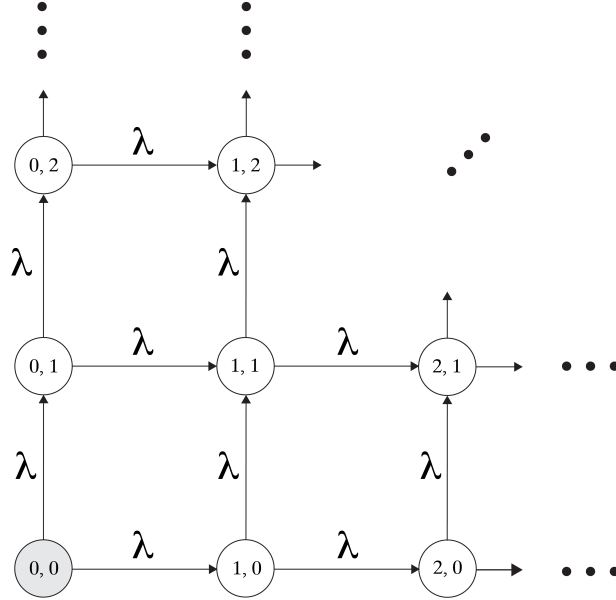
$$\begin{aligned} \text{var}(X_2 - L[X_2|Y_1, Y_2, Y_3]) &= \text{var}(X_2 - \hat{X}_{2|2} - k_{23}(Y_3 - \rho\hat{X}_{2|2})) \text{ (where } k_{23} = \frac{\rho\sigma_{2|2}^2}{\rho^2\sigma_{2|2}^2 + 2}\text{)} \\ &= \text{var}((X_2 - \hat{X}_{2|2}) - k_{23}(\rho X_2 + V_3 + W_3 - \rho\hat{X}_{2|2})) \\ &= \text{var}((X_2 - \hat{X}_{2|2})(1 - k_{23}) - k_{23}(V_1 + W_1)) \\ &= (1 - k_{23})^2\sigma_{2|2}^2 + 2k_{23}^2 \\ &= \frac{4\sigma_{2|2}^2}{(\rho^2\sigma_{2|2}^2 + 2)^2} + \frac{2\rho^2\sigma_{2|2}^2}{(\rho^2\sigma_{2|2}^2 + 2)^2} \\ &= \frac{2\sigma_{2|2}^2}{\rho^2\sigma_{2|2}^2 + 2} \end{aligned}$$

Again, as expected, we see this is less than $\sigma_{2|2}^2$, the mean squared error from 5d(iii).

*Note: The calculations for the MSE were very tedious and not in the spirit of what we meant to test. Full credit was given if the estimator and the relative size of the error compared to previous parts were correct.

Problem 6 [4+5+5+6]: Continuous Random Walk on a Grid

An ant performs a continuous time random walk on the non-negative integer lattice. At any time $t \geq 0$, the position of the ant $Z(t)$ is a tuple $(X(t), Y(t))$. The ant starts in state $(0, 0)$. At any time, the ant moves to the right with rate λ and up also with rate λ , so that the position of the ant is described by an infinite CTMC on the state space $\mathbb{N} \times \mathbb{N}$, as pictured below.



- (a) Argue that $X(t)$ and $Y(t)$ are independent Poisson Processes and write down their rates.

Every move that the ant makes is either an increment to $X(t)$ or to $Y(t)$. Changes in the ant's position arrive according to a Poisson process of rate 2λ . Each move is independently a move up with probability $1/2$ and right with probability $1/2$. By Poisson splitting, the process that counts upward moves and the process that counts right moves are independent Poisson processes with rate λ each. These are exactly the processes $X(t)$ and $Y(t)$.

- (b) At time $t = 1$, the ant is at position $(3, 1)$. What is the probability that at time $t = 0.75$ the ant was at position $(3, 0)$?

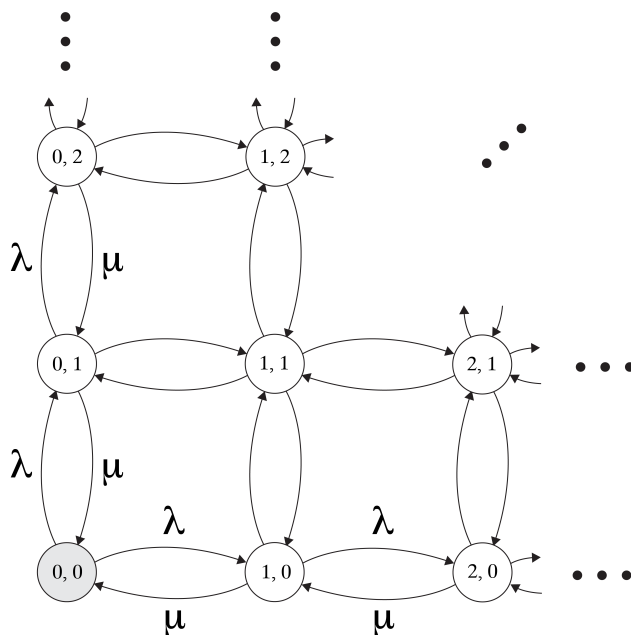
We are given that at time 1, there have been 4 arrivals to the process that counts changes to the ant's position. 3 of these are right moves and 1 is an up move. Conditioned on there being 4 arrivals, these arrivals are uniformly distributed on the interval $(0, 1)$. We want the probability that the three right moves are in $(0, 0.75)$ and the one up move is in $(0.75, 1)$. This probability is $(3/4)^3(1/4)$.

- (c) Denote by V_n the ant's average speed at time $t = n$, that is, $V_n = (X(n) + Y(n))/n$. Does the sequence $(V_n)_{i=1}^{\infty}$ converge a.s.? If not, prove it. If yes, specify what it converges to and justify (assume n to be an integer).

Proof: Let $N(s, t)$ denote the number of changes in position that the ant undergoes in the time interval $[s, t)$. Then

$$V_n = \frac{1}{n}N(0, n) = \frac{1}{n} \sum_{i=1}^n N(i-1, i)$$

(d) Now, modify the walk by allowing the ant to also move left (if possible) at rate μ and down (if possible) also with rate μ . The new CTMC is pictured below (all down and left arrows have rate μ and all up and right arrows have rate λ). For $\lambda < \mu$, find the stationary distribution of the corresponding CTMC (Hint: Use a symmetry argument that parallels your argument in part (a)).


$$\pi_X(x) = \left(\frac{\lambda}{\mu}\right)^x \left(1 - \frac{\lambda}{\mu}\right)$$

Which is the same for Y . Then finally we can find the stationary distribution of the original chain using independence

$$\pi(x, y) = \pi_X(x)\pi_Y(y) = \left(\frac{\lambda}{\mu}\right)^{x+y} \left(1 - \frac{\lambda}{\mu}\right)^2$$

It is easily checked that this solves detailed balance for the original chain.

Solution 2: By symmetry, the value of the stationary distribution must agree on all states x, y that are the same manhattan distance away from the origin. Let $\alpha(d)$ be the value of the stationary distribution at any state (x, y) for which $x + y = d$ so that $\pi(x, y) = \alpha(x + y)$. It suffices to compute α .

Consider the cut $S_d = \{(x, y) : x + y \leq d\}$. The stationary distribution must satisfy the relation that flow leaving S_d equals flow entering S_d .

For the flow leaving S_d , there are $d + 1$ states for which $x + y = d$. Each of these states move out of S_d at rate 2λ so flow out of S_d is $2\lambda(d + 1)\alpha(d)$.

For the flow entering S_d , there are d states that flow into S_d at rate 2μ and an additional 2 states that flow into S_d at rate μ each (these are the states $(0, d + 1)$ and $(d + 1, 0)$). Total flow into S_d is then $2\mu(d + 1)\alpha(d + 1)$.

Equating the two, we get $\alpha(d + 1) = (\lambda/\mu)\alpha(d)$ so that

$$\alpha(d) = \left(\frac{\lambda}{\mu}\right)^d \alpha(0)$$

Finally, enforce that $\sum \pi(x, y) = 1$:

$$\sum_{x,y} \pi(x, y) = \sum_{x,y} \alpha(x + y) = \sum_{d=0}^{\infty} (d + 1)\alpha(d) = \alpha(0) \sum_{d=0}^{\infty} (d + 1) \left(\frac{\lambda}{\mu}\right)^d = \frac{\alpha(0)}{(1 - (\lambda/\mu))^2}$$

The value of the final summation of the form $\sum (d+1)r^d$ can be computed by differentiating a geometric series $\sum r^d = 1/(1 - r)$ with respect to r within the radius of convergence.

Equating this to 1 readily gives $\alpha(d) = (\lambda/\mu)^d (1 - \lambda/\mu)^2$. Now we can write $\pi(x, y)$

$$\pi(x, y) = \alpha(x + y) = \left(\frac{\lambda}{\mu}\right)^{x+y} \left(1 - \frac{\lambda}{\mu}\right)^2$$