Solutions

Fall 2012: EECS 126 Final

Dec 11, 2012

FINAL EXAM

No collaboration permitted.

Three sheets of notes, both sides, permitted. Turn them in with your exam.

No access to the internet, textbook, notes beyond the three sheets above is permitted.

You have **till 6PM** to complete the exam.

Be clear and precise in your answers. Illegible exams cannot be graded.

Write your name and student ID on EVERY sheet.

There are a total of 179 points on this exam. 120 is a good score.

Goodluck!

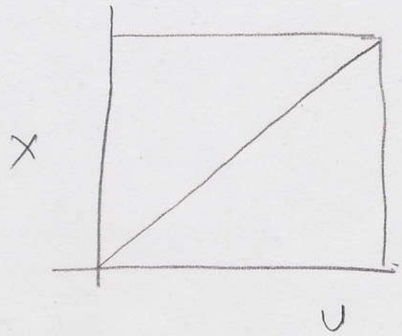
Note: Feel free to leave answers to any questions in terms of $\Phi(x)$ or $\Phi^{-1}(x)$.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

1. (24 points) True or False? Prove or give a counterexample.

   a. (12 points) If U and V are iid Uniform $[0, 1]$ random variables, and $X = \max(U, V)$, then (U, X) are jointly continuous random variables.

False.



If X and U were jointly continuous, then the tot. probability along the line $X = U$ would have to be zero. But clearly $P(X=U) = 1/2$, and hence there is a delta along that line.

Alternatively, you can also calculate the density as in discussion to show this.

b. (12 points) If X_i are iid Exponential random variables with parameter λ and N is an independent Geometric random variable with parameter $0 < p < 1$, then $Y = \sum_{i=1}^{N} X_i$ is an Exponential random variable.

True

Consider a poisson process of rate λ. Now, split it using a coin which is heads with probability p. Then Y is the first arrival of the "heads" split of the process, and hence is Exponential.

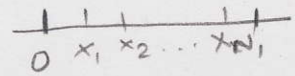
2. (20 points) Consider the interval $[0, 1]$. An arrival process is constructed for this interval as follows:

- First, a random variable N is drawn with PMF $Pr(N = i) = e^{-1}\frac{1}{i!}$ (for integer $i \geq 0$, the PMF is zero for all other values.)
- Then, N arrival times X_i are chosen iid Uniform$[0, 1]$

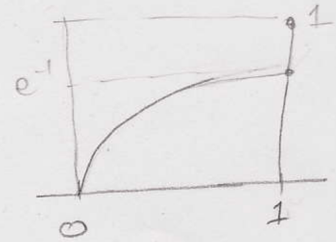
Let T be the time till the first arrival. (If $N = 0$, then $T = 1$.)

What is the distribution for T?

(HINT: Think about Poisson Processes to get an intuition. Then try and calculate $P(T > t)$ using conditioning as appropriate.)



$$f_T(t) = \sum_{n=0}^{\infty} f_{T|N}(t|N=n) \cdot P(N=n)$$



$f_{T|N}(t|N=n) \Rightarrow$ first calculate the CDF

$$P(T \leq t | N=n) = 1 - P(T \geq t | N=n)$$

$$= \begin{cases} 1 - (1-t)^n & 0 \leq t \leq 1 \\ 1 & t \geq 1 \end{cases}$$

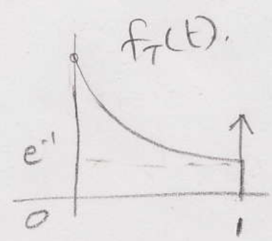
$$f_{T|N}(t|N=n) = \begin{cases} n(1-t)^{n-1} & 0 \leq t < 1. \\ (1 - n(1-t)^{n-1}) \cdot \delta(t-1) & t = 1 \\ 0 & t > 1. \end{cases}$$

$$f_T(t) = \sum_{n=0}^{\infty} n(1-t)^{n-1} \cdot \frac{1}{e} \cdot \frac{1}{n!} \qquad 0 \leq t < 1$$

$$= \sum_{n=0}^{\infty} e^{-1(1-t)} e^{-t} \frac{(1-t)^{n-1}}{(n-1)!}$$

$$= e^{-t} \cdot \sum_{n=0}^{\infty} e^{-(1-t)} \frac{(1-t)^{n-1}}{n-1!}$$

$$= e^{-t} \cdot 1 \qquad 0 \leq t = 1.$$



$$f_T(1) = 1 - \int_0^1 e^{-t} dt = 1 + e^{-t}\Big|_0^1 = 1 + e^{-1} - 1 = e^{-1}$$

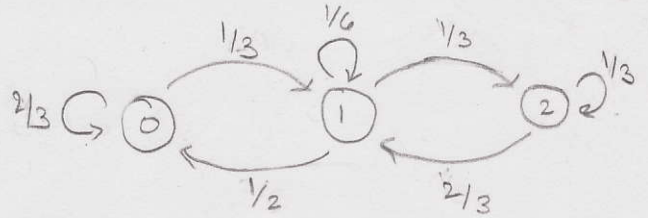
f(t) = e^(-t) 0 <= t < 1

= e^(-1) delta(t-1) t=1

EXTRA SPARE PAGE

3. (20 points) This problem asks you to construct Markov Chains having certain properties. No justification is required.

   a. (10 points) A discrete-time Markov Chain with a unique stationary distribution of $(\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$. *(HINT: you can solve both a and b with a single solution if you want.)*



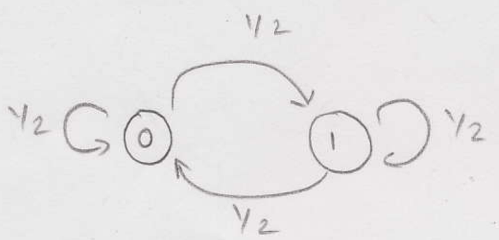
$\pi_0 \cdot \frac{1}{3} = \pi_1 \cdot \frac{1}{2} \Rightarrow 2\pi_0 = 3\pi_1$

$\pi_1 \cdot \frac{1}{3} = \pi_2 \cdot \frac{2}{3} \Rightarrow \pi_1 = 2\pi_2$

$\pi_0 + \frac{2}{3}\pi_0 + \frac{1}{3}\pi_0 = 1.$

$\therefore \pi_0 = \frac{1}{2}, \pi_1 = \frac{1}{3}, \pi_2 = \frac{1}{6}$

   b. (10 points) A discrete-time Markov Chain with an eigenvalue of 0 associated with its probability transition matrix.

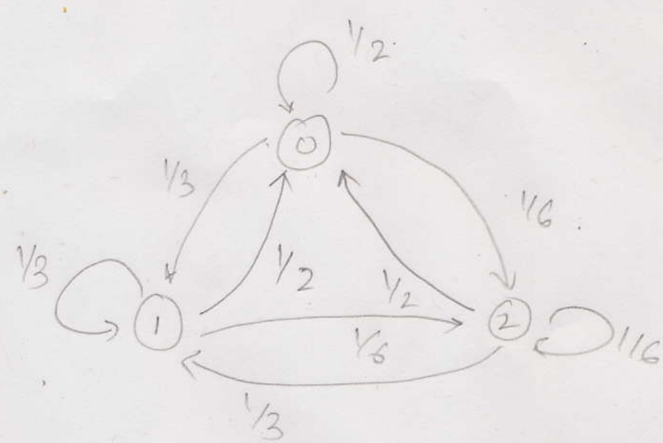


$P = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$

$\because$ det(P) = 0, 0 is an eigenvalue.

The combined solution: iid.

All states are independant



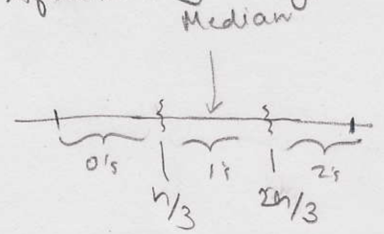
4. (20 points) Consider X_i iid discrete random variables with a uniform probability over $\{0, 1, 2\}$. Prove

$$\lim_{n\to\infty} Pr(\text{EmpiricalMedian}(X_1, \ldots, X_n) = 1) = 1.$$

Recall that the Empirical Median of a sequence is the number in the middle of the sequence after you sort the sequence. Feel free to assume that n is odd.

(HINT: Use the AEP form of the WLLN to see why this must be true. Then prove it however you want.)

After sorting a typical



Let $Y_i = 1$ if $X_i = 0$, $Y_i = 0$ o.w.

$Z_i = 1$ if $X_i = 1$, $Z_i = 0$ o.w.

$W_i = 1$ if $X_i = 2$, $W_i = 0$ o.w.

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Z_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{3}$$ a.s. by [redacted] SLLN

Let the Empirical Median $= M$.

When the X_i's are organized in increasing order

$\lim_{n\to\infty} P(\text{first } 1/3\text{rd} = 0\text{'s}) = 1$ +/- epsilon

$\lim_{n\to\infty} P(\text{second } 1/3\text{rd} = 1\text{'s}) = 1$ +/- epsilon

$\lim_{n\to\infty} P(\text{third } 1/3\text{rd} = 2\text{'s}) = 1$ +/- epsilon

Hence, the empirical median must be in middle $1/3$rd.

$$\therefore \lim_{n\to\infty} P_r(M = 1) = 1.$$

When we count from the smallest element, as n tends to infinity the probability that the middle element will be 1 tends to 1, since the first third will be zeros and the last third will be 2s (with some epsilon slop)

5. (40 points) X is a standard Gaussian $N(0,1)$ random variable. U_i are iid standard Gaussian $N(0,1)$ random variables that are also independent of X.

a. (5 points) Let $Y_1 = X + U_1$. **What is the MMSE estimate for X given Y_1? And calculate the variance of the estimation error.**

By symmetry $E[X|Y_1] = Y_1/2$.

$E[(X - \frac{Y_1}{2})^2] = E[(\frac{X}{2} - \frac{U_1}{2})^2] = \frac{1}{4} E[X^2] + \frac{1}{4} E[U^2] = \frac{1}{2}$.

b. (15 points) Suppose that you also have access to $Y_2 = U_1 + U_2$. **What is the MMSE estimate for X given both Y_1, Y_2? And calculate the variance of the estimation error.**

$$E[X|Y_1, Y_2] = K_{XY} K_Y^{-1} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$= [\mathrm{Cov}(X,Y_1) \quad \mathrm{Cov}(X,Y_2)] \begin{bmatrix} \mathrm{var}(Y_1) & \mathrm{Cov}(Y_1,Y_2) \\ \mathrm{Cov}(Y_1,Y_2) & \mathrm{var}(Y_2) \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$= [1 \quad 0] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = [1 \quad 0] \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$= \frac{2}{3} Y_1 - \frac{1}{3} Y_2.$$

$$E[(X - \tfrac{2}{3}Y_1 + \tfrac{1}{3}Y_2)^2] = E[(X - \tfrac{2}{3}X - \tfrac{2}{3}U_1 + \tfrac{1}{3}U_1 + \tfrac{1}{3}U_2)^2]$$

$$= E[(\tfrac{1}{3}X - \tfrac{1}{3}U_1 + \tfrac{1}{3}U_2)^2]$$

$$= \tfrac{1}{9}(E[X^2] + E[U_1^2] + E[U_2^2]) = \tfrac{1}{3}$$

c. (20 points) Now suppose that $Y_i = U_{i-1} + U_i$ for $i \geq 2$. **What is the MMSE estimate for X given Y_1, Y_2, Y_3? And calculate the variance of the estimation error. What happens to the optimal estimation error as the number of observations $n \to \infty$? Why?**

(HINT: There are many ways to proceed. Remember the key properties of jointly Gaussian estimation: the MMSE estimate and the estimation error are independent of each other. And if you get a new observation, you don't have to recalculate everything from scratch...)

(Second HINT: Problems are easier in the right coordinate system. So as a trick, think about first causally transforming the observations Y_i into a form Y_i' in which the "noise" is independent from observation to observation and the "signal" X is present in each observation...)

$Y_1 = X + U_1$ Let $Z_1 = Y_1 = X + U_1$

$Y_2 = U_1 + U_2$ $Z_2 = Y_1 - Y_2 = X - U_2$

$Y_3 = U_2 + U_3$ $Z_3 = Y_1 - Y_2 + Y_3 = X + U_3$

So now $E[X | Y_1, Y_2, Y_3] = E[X | Z_1, Z_2, Z_3]$ clearly. All information is the same.

Now, U_1, $-U_2$ and U_3 are all $N(0,1)$.

$$E[X | Z_1, Z_2, Z_3] = K_{XZ} K_Z^{-1} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}$$

$$= [1 \quad 1 \quad 1] \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}$$

$$= [1 \quad 1 \quad 1] \begin{bmatrix} 3/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & 3/4 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}$$

$$= \frac{1}{4} Z_1 + \frac{1}{4} Z_2 + \frac{1}{4} Z_3 = \frac{1}{4} Y_1 + \frac{1}{4}(Y_1 - Y_2) + \frac{1}{4}(Y_1 - Y_2 + Y_3)$$

$$= \frac{3}{4} Y_1 - \frac{1}{2} Y_2 + \frac{1}{4} Y_3 .$$

$$E[(X - \tfrac{3}{4}X - \tfrac{3}{4}U_1 + \tfrac{1}{2}U_1 + \tfrac{1}{2}U_2 - \tfrac{1}{4}U_2 - \tfrac{1}{4}U_3)^2] = E[(\tfrac{1}{4}X - \tfrac{1}{4}U_1 - \tfrac{1}{4}U_2 - \tfrac{1}{4}U_3)^2] = \frac{1}{4}.$$

Similarly, as $n \to \infty$, we will get more and more Z_i, i.e. noisy observations of X, and the error variance will go to zero as $\frac{1}{n}$

Blank Page for Work

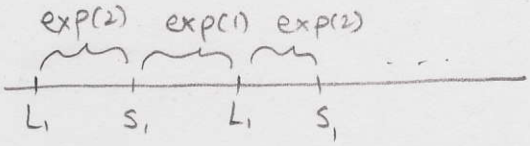
6. (30 points) You are attending a party and are immediately introduced to the host and begin a conversation about the election. After some time, they introduce you to someone new (that you haven't met before) and you begin a new conversation. Suppose that there is a social convention that a long-haired person will only introduce you to a random short-haired person and a short-haired person will only introduce you to a random long-haired person. (The host is long-haired.)

   Further suppose that a Long-haired person introduces you to someone new in a random amount of time that is distributed like an Exponential random variable with parameter $\lambda = 2$. For Short-haired people, the parameter is $\lambda = 1$.

   a. (10 points) **Argue that the proportion of long-haired people that you've met will be very close to 0.5 of the total people that you've met.**
   (HINT: This is easy.)

Since you necessarily alternate between meeting long and short haired people, as the number of people grows large, you will have met an equal number of long and short haired people.

b. (20 points) **About how long a time should you budget for yourself at the party to have met at least 101 people total with a probability of at least 80%?**
(HINT: First calculate the expected amount of time till you meet the 101th person. Then, what do you think the pdf of the time to meet the 101st person looks like?)
(Second HINT: This is part (b) for a reason.)



Let X_i be the Long → Short-haired interarrival times and
Y_i be the Short → long haired "

Then T, time to meet 101st person $\qquad$ X_i's and Y_i's are independent.

$$T = X_1 + Y_1 + X_2 + Y_2 + \cdots + X_{50} + Y_{50}.$$
$$= \sum_{i=1}^{50} X_i + \sum_{i=1}^{50} Y_i$$

$\therefore E[T] = 50 \cdot 1 + 50 \cdot \frac{1}{2} = 75$. The pdf of T is the sum of two Erlang random variables. $\quad 50 \cdot \frac{1}{1^2} \downarrow$

But we can also approximate $\sum_{i=1}^{50} X_i$ by a Gaussian $N(50, 50)$
and $\sum_{i=1}^{50} Y_i$ by $N(25, 25/2)$. Hence, T is approx $N(75, \frac{125}{2})$.
$\uparrow 50 \cdot \frac{1}{4}$

So I want to find t, s.t.

$P(T \le t) = 80\%$. ie. I will have met 101 people in time t with prob. 80%.

$$P(N(75, \tfrac{125}{2}) \le t) = P(N(0, \tfrac{125}{2}) \le t - 75)$$
$$= P(N(0,1) \cdot \sqrt{\tfrac{125}{2}} \le t - 75)$$
$$= P(N(0,1) \le \frac{t-75}{5} \cdot \sqrt{\tfrac{2}{5}}).$$
$$= \Phi\left(\frac{t-75}{5} \sqrt{\tfrac{2}{5}}\right) = 80\%$$

$$\therefore \frac{t-75}{5} \sqrt{\frac{2}{5}} = \Phi^{-1}\left(\frac{4}{5}\right)$$

$$\boxed{\therefore t = \Phi^{-1}\left(\frac{4}{5}\right) \sqrt{\frac{5}{2}} \cdot 5 + 75}$$

Blank page for work.

7. (25 points) We want to estimate what fraction of people like Psy's Gangnam Style video. The population consists of 50% Long-haired people and 50% Short-haired people.

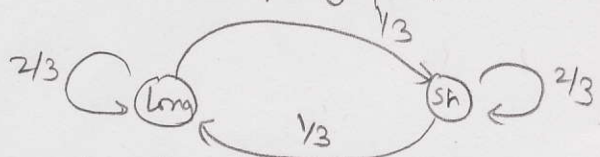
We are not allowed to poll people by drawing them randomly from the population. Instead, we start with one randomly selected person, and from then on, we randomly pick one of that person's friends. Next we sample that person's friends to pick a person. And so on....

The challenge is that long-haired people tend to have long-haired friends and short-haired people tend to have short-haired friends. Suppose that a long-haired person has $\frac{2}{3}$ long-haired friends and $\frac{1}{3}$ short-haired friends. Similarly, a short-haired person has $\frac{2}{3}$ short-haired friends and $\frac{1}{3}$ long-haired friends.

How many people should we sample using this process to get at least 100 long-haired people with confidence at least 80%?

(HINT: Think about using a Markov chain in your model. What effect do you think this has? Think about the HW where you had to prove a WLLN for Markov chains...)

We'll model the seq. of people by a Markov Chain.



$$P = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Let $X_i = 1$ if state at time i is Long haired, 0 otherwise

We know MC have a WLLN, but they converge more slowly due to memory. So we'll have to wait a bit longer than we would have had to for iid X_i

We want $P(\sum_{i=1}^{n} X_i \le 100) \le 0.05$.

To calc. $\text{var}(\sum X_i)$, first, we calc. $E[X_iX_j] = E[X_j|X_i=1]\cdot P(X_i=1)$ $j>i$

$= P(X_j=1|X_i=1)\cdot 1/2$

To calculate the probability we take P to the $(j-i)$ th power, and get $P(X_j=1|X_i=1) = \frac{1}{2}(1+(\frac{1}{3})^{j-i})$.

$\therefore E[X_iX_j] = \frac{1}{4} + \frac{1}{4}(\frac{1}{3})^{j-i}$

$E[X_i] = \frac{1}{2}$, $E[\sum_{i=1}^{n} X_i]^2 = n^2 \cdot \frac{1}{4}$.

$$E[(\sum_{i=1}^{n} X_i)^2] = \sum_{i=1}^{n} E[X_i^2] + 2\cdot\sum_{i\ne j} E[X_iX_j] = n\cdot\frac{1}{2} + n(n-1)\cdot\frac{1}{4} + \frac{1}{4}\sum_{j\ne i}(\tfrac{1}{3})^{|j-i|}$$

$$\le \frac{n}{2} + \frac{n(n-1)}{4} + \frac{n}{2}\cdot\sum_{k=1}^{\infty}(\tfrac{1}{3})^k$$ (n geometric series, we add some extra terms)

$$= \frac{n}{2} + \frac{n(n-1)}{4} + \frac{n}{2}\left(\frac{1/3}{1-1/3}\right)$$

$$= \frac{n}{2} + \frac{(n)(n-1)}{4} + \frac{n}{4} = \frac{n}{2} + \frac{n^2}{4}$$

Hence, $\text{var}(\sum_{i=1}^{n} X_i) \le \frac{n}{2} + \frac{n^2}{4} - \frac{n^2}{4}$

$= \frac{n}{2}$

Blank page for work So, now we can apply the CLT to find n.

$\to$ Note that the variance is converging and so we can apply CLT as though these are iid.

[redacted]

The additional cross terms in the higher moments will also just sum-up to converge rather than blowing up to dominate.

$$P\left(\sum_{i=1}^{n} X_i \le 100\right)$$

$$= P\left(\frac{\sum_{i=1}^{n} X_i - \frac{n}{2}}{\sqrt{n/2}} \le \frac{100 - n/2}{\sqrt{\frac{n}{2}}}\right).$$

$$= P\left(N(0,1) \le \frac{100 - n/2}{\sqrt{n/2}}\right)$$

$$= \phi\left(\frac{\left(100 - \frac{n}{2}\right)\cdot\sqrt{2}}{\sqrt{n}}\right) = 0.8$$

$$\therefore \left(100 - \frac{n}{2}\right)\sqrt{2} = \phi^{-1}(0.8)\sqrt{n}.$$

$$\frac{n}{\sqrt{2}} + \phi^{-1}(0.8)\cdot\sqrt{n} - 100\sqrt{2} = 0.$$

$$\sqrt{n} = \frac{-\phi^{-1}(0.8) \pm \sqrt{\phi^{-1}(0.8)^2 + 400}}{\sqrt{2}}$$

Clearly, we take the positive root for n, so choose the + sign above.