**MTL 106 (Introduction to Probability Theory and Stochastic Processes)**
**Assignment 2 Report**

Name: Ankish Chandresh Entry Number: 2018EE10447

1. Basic Probability
Following is a problem on Bayesian Inference.
Nature chooses some parameter $\theta$ for $B(1,\theta)$ and generates n indpendent samples $X_1, ..., X_n$. Based on your intuitions or belief prior distribution is believed to be uniform $\pi(\theta) = 1$. You also chose the model $p(X/\theta)$that reflects your beliefs by the likelihood function $Ln(\theta) = \prod_i p(X_i/\theta)$(since samples are independent). You then observe the samples and make updates in your belief by computing the**posterior** distribution of the parameter and thus using Bayes theorem you find bayes posterior estimator for some confidence interval(dont evalute the interval). Find the estimator such that your "risk function" $\theta\ E\left[(\widehat{\theta}(x) - \theta)^2\right]$ is minimised.
**Solution:**
Let$(X_1, X_2...X_n) = D_n$
$p(\theta/D_n) = p(D_n/\theta) * \pi(\theta)/p(D_n) = Ln(\theta) * \pi(\theta)/p(D_n)$
$p(\theta/D_n) \propto \pi(\theta)Ln(\theta) = \theta^{Sn}(1-\theta)^{n-Sn} = \theta^{Sn+1-1}(1-\theta)^{n-Sn+1-1}$
Where $S_n = \sum_{i=1}^{n} X_i$
Note that the posterior distribution for $\theta$ is a Beta distribution with parameters$Sn+1andn-Sn+1$.
Thus, $p(\theta/D_n) = \Gamma(n+2)/\Gamma(S_n+1)\Gamma(n-S_n+1)\theta^{(Sn+1)-1}(1-\theta)^{(n-Sn+1)-1}$
$\theta/D_n \quad Beta(Sn+1, n-Sn+1)$
The estimator(mean) minimises the given risk measure and is a possible estimator others being mode etc.
Now, The mean of a Beta(a,b) distribution is given by a/(a + b), thus the Bayes posterior estimator(mean) is $\widehat{\theta} = S_n + 1/(n+2)$. Ans thus is the required answer.
Clearly, the change in prior beilefs e.g. $\pi(\theta)$ can be change the estimator value.

2. Random Variable/Function of a Random Variable
Consider choosing two random points on ac circle of radius r, uniformly. Find the distribution of the length of the chord so formed. Also, find the disytrbution. of area of triangle formed by the origin and the two chosen points. Note that the chosien points have angular separation less than $\pi/2$.
**Note:**
The problem can be modeled as finctions of random varibales $\alpha$ and $\beta$ , the angular positions of the randomly chosen points(assumed independently).
**Solution:**
Let two points on the circle we have:
$r^2 + r^2 - 2r^2cos(\theta) <= l^2$ where $\theta$ is the angualar separation and l is the length of the chord.
$cos(\theta) >= 1 - l^2/r^2$
Thus $P(L <= l/L <= \sqrt{2}) = P(2cos(\theta) >= 1 - l^2/r^2) = P(|\theta| <= cos^{-1}((1 - l^2/r^2)/2)$
Notice that inverse can be taken since the angular separation is less than $\pi/2$.
Now,

$P(|\theta| <= cos^{-1}((1-l^2/r^2)/2) =$

$$\int_0^{2\pi}\int_{\alpha-\theta}^{\alpha+\theta} f(\alpha,\beta)d\beta d\alpha$$

where $f(\alpha,\beta) = 1/2\pi 2\pi$
Thus, $P(|\theta| <= cos^{-1}((1-l^2/r^2)/2) = \theta/\pi = (cos^{-1}((1-l^2/r^2)/2)/\pi$
$P(L <= l/L <= \sqrt{2}) = (cos^{-1}((1-l^2/r^2)/2))/\pi$
Now, we notice the area $A = r^2 sin(|\theta|)/2$ where $\theta$ is the angular separation.
Similarly, $P(A <= a/A <= r^2/2) = P(|\theta| <= sin^{-1}(2a/r^2))$
$P(|\theta| <= sin^{-1}(2a/r^2)) = \theta/\pi = sin^{-1}(2a/r^2)/\pi$
Thus, $P(A <= a/A <= r^2/2) = sin^{-1}(2a/r^2)/\pi$
Note that to avoid mutiple cases and condtions angular separatoin is taken to be less than $\pi/2$.
An another interesting version could be choose three points and find the ditribution for the area of the triangle thus formed, given we found distributino of area of one of the three smaller triangles formed, the process is similar as mentioned with some conditional probabilities.

3. Stochastic Processes
Consider a baby learning from the surroundings and becoming adult eventually. Is the learning a stochastic process, if yes how? The process is indeed a stochastic process, consider the amount of learning to be given by the number of connections formed in the brain. Since the connections formed can be very large assume the process of learning (number of connections formed) to be a poisson process for each half portion of the brain with rate 2 and 3 repectively(left and right portion of the brain are independent). Find the probability that up to age 3 the person forms more than 10 total connections given that upto age 1 he has already formed 5 connections.
What is the probability that the total learning by the person from 10 to 20 years is non-zero. What is the probability that the total learning in the first 5 years and last 1 year are same(give the expression).
**Solution:**
Since, the input(sensed through eyes, touch etc..) distribution also changes with time, the process of learning described is a stochastic process.(or the input distribution is stochastic)
Since the total learning is given by the sum of left and right portions of the brain which are themselves poisson processes with rate 2 and 3 and are independent the total learning is also a Poisson process with rate 2+3 or 5.
For first part:
$P(X(3) > 10/X(1) = 5) = P(X(3-1) > 5)$(by Poisson process property)
Thus we get$P(X(2) > 5) = 1 - P(X(2) <= 5)$
Where$P(X(2) <= 5) = \sum_{k=0}^{5} P(X(2) = k)$whith $rate = 5$.
So for second question, $t = 30 - 20 = 10$ (Poisson process property) and $\lambda = 5$
Non-zero total learning in t time interval $= P(X(t) > 0) = 1 - P(X(t) = 0) = 1 - \exp -5t = 1 - \exp -5 * 10$

Now, for first 5 years, $t1 = 5$, and last 1 year, $t2 = 1$(Poisson process property.)
For total learning to be equal we have$\sum_{k=0}^{inf} P(X(t1) = k, X(t2) = k) = \sum_{k=0}^{inf} P(X(t1) = k)P(X(t2) = k)$ since the total learning is independent if the intervals don't overlap(assuming intervals don't overlap.) Also, Note that the answers are different when intervals overlap(the person lives for 5 years) and is equivalent to no learning for first 4years.

4. Stochastic Processes
Suppose you walk on X-axis with the positon after n steps given by $S_n = \sum_{i=1}^{n} X_i$ which is a non-symmetric random walk, with$P(X_i = 1) = p$ and $P(X_i = -1) = q$. At each position $S_n$ there are $|S_n|$ people, if position is positive you sample a price from a uniform distribution such that you give maximum p/q money for each person at that position, if the position is negative the maximum changes to q/p. Find the step at which the expected probability of the given description to happen is maximum i.e You evaluate expectation $M_{n+1} = a^{S_n+1}$ after covering n steps where $a = p/q$ or $q/p$. You are given $X = \{X_1, X_2..X_{10}\} = \{1, 1, -1, 1, -1, -1, 1, -1, 1, 1\}$ for 10 steps.
**Solution:**
For the Given description $M_n$ is simply $(q/p)^{S_n}$.
Since the probaility for sampling from the unform ditribution is 1/(a) and ($a = p/q$ or $a = q/p$) and for negative positions the inverse is inversed again.
Note that $M_n$ follows martinagle proeprty as shown fro a random walk($X_i$ are independent):
Also for the description we get the following conditional expectation:
$E[M_{n+1}/X_1 = x_1, X_2 = x_2...X_n = x_n] = E((q/p)^{S_{n+1}}/X_1 = x_1, X_2 = x_2...X_n = x_n) = ((q/p)^{S_n} E[(q/p)^{X_{n+1}}/X_1 = x_1, X_2 = x_2...X_n = x_n] = (q/p)^{S_n} E[(q/p)^{X_{n+1}}]$
See that $E[(q/p)^{X_{n+1}}] = q/p * p + p/q * q = 1$, thus:
$E[M_{n+1}/X_1 = x_1, X_2 = x_2...X_n = x_n] = (q/p)^{s_n}$
or equivalently $E(M_{n+1}/S_n = s_n, S_{n-1} = s_{n-1}...S_1 = s_1) = (q/p)^{s_n} = M_n$
Thus, for maximum expectation value, we simply find maximum and minimum(depending on wether $p > q$ or $q > p$) prefix sum for the given X vector. Which is 2 and -1 respectively and occurs for more than one psoitnos.
Thus anwer would be either $(q/p)^2$ or p/q(p not equal to q).

5. DTMC
Consider an Artificial Neural Network(ANN) with an input layer and three hidden layers and the output is not in consideration for this problem. Clearly you feed the input carrying some information $I_0$ and each layer transforms the input data and changing the output information content (say $I'$). Firstly is ANNs a markov chain mention reasons and the type of the markov chain specify the states. Now you are a researcher and allow the layers to communicate or form recurrent connections(i.e. layer L-1 can have input from layer L or any other layer). You know the goal of ANNs is to minimise the information content for the last layer and is later used by output layer(not in consideration) for classification, suppose after training the first version(without recuurent connections) you obtain weights such that first hidden layer gains information content 1/2 of its input(not necessarily the intial input data), second layer gains 1/3 information of its input and third layer gains 1/4, i.e. if input has $I_0$

information, first,second and third hidden layer has $I_0/2$, $I_0/2*3$, $I_0/2*3*4$ information respectively. Now if recurrent connections are allowed from the given one step-transition matrix(given below) find the state for initial input($I_0$) and the final state($\alpha I_0, \alpha < 1$) to be used for classifier layer after three steps(since first version also uses three steps so comparison is possible) of transition and modifying information content $I_0$ for maximum performance with the given resources and no more training/updates, compare the two versions.

$$P = [0.25, 0.5, 0.25]$$

$$[0.2, 0.2, 0.6]$$

$$[0.3, 0.4, 0.3]$$

**Solution:**
Yes, a typical ANN is indeed a dicrete time markov chain,with fixed transition one-step probailities, and the states are actiavtions(n-dimensional vector representaion) of the hidden layers carying some information, clearly the states (a vector) are discrete and the time( ith layer) is also discrete.
Now, since we are required to find the expected information conetent after three transitions we didvide the elements by the factor of reduction in the one-step transiotion probabilit matrix.(We assume for a self-loop the information remains unaltered, thus diagonal elements remain unchanged).

$$P = [0.25, 0.5/3, 0.25/4]$$

$$[0.2/2, 0.2, 0.6/4]$$

$$[0.3/2, 0.4/3, 0.3]$$

Now,

$$P^3 = [0.0394, 0.0393, 0.0358]$$

$$[0.038, 0.037, 0.04]$$

$$[0.051, 0.050, 0.0555]$$

Clearly the the expected value after three step transition from 1 to 3 is the least(could be different if P is taken differetly) and thus has expected informatio as $0.0358 * I_0$ compared to first version which is roughly $0.04 * I_0$.
Thus we see small improvement in the performance by allowing recurrent connections(terminology not the same as used in DTMC).

6. DTMC
Consider a $8 * 8$ chess board. Place two knights one at one of corners of the board and the other at the usual starting position. The pieces are then allowed to move, find which piece takes lesser average number of steps required to get back at their respective starting points.

**Note:**
The problem is a standard one, many variants can be developed by changing the chess piece or the starting position which doesn't affect the general process.
**Solution:**
Picture the problem as a markov chain. Where a step(discrete) is a transition from one square/state to the other. Each position on the board being a state(discrete) and having some transition probability for moving to the next position. This is thus a graph, some states or squares are reachable in a single move consider an edge in such case with one step transition probability 1/(degree of the state or vertex) else no edge.
Every pair of possible states/vertices is connected by a path and thus the graph is connected. Since the graph is connected this simply means the markov chain is irreducible, thus we have the following theorem:
The ith element of the stationary distribution is give by $\pi_i = d_i / \sum d_j$ and is unique. where $d_i$ is the degree or poissble number of moves from the ith square or state and $\sum d_j$ is the sum of all degree of all vertices/states (assume each state is represented by i from 1 to 64 ).
Also since the DTMC is irreducible and positive reccurent(by observation for the chess board) we have the expected return time or the average number of steps needed to return , for the ith state/square as $E_i = 1/\pi_i$ (by result/theorem).
Clearly, for the given problem we need to evaluate $d_1$ and $d_2$ that is the corner square is state 1 and the usual posotion is state 2.
Thus $d_1 = 2$ and $d_2 = 3$. Now, there are 8 possible directions a knight can move, and each direction can start from a 6x7 square(note that the graph is bidirectional that is the states communnicate), so we can know the number of edges/degrees buy only counting for the bigger sqauere which allows all 8 moves thus there will be $8 * 6 * 7 = 336$ possible moves.
Now,$E_1 = \sum d_j/d_1 = 336/2 = 168$ also ,$E_2 = \sum d_j/d_1 = 336/3 = 112$

Thus clearly knight at usual position takes lesser number of averge steps.
One can notice the states with higher degree or possible state transitions take lesser average return time since there more likely to return intutively.

7. CTMC
Consider two molecular reactions going on between A,B and C. One A generates two C molecules and one B molecule. A to B has forward rate $\mu_1$, B to A has backward rate $\lambda_1$ and A to C has forward rate $\mu_2$, C to A has backward rate $\lambda_2$. You start with 10 A molecules and zero B and C molecules. Find the distribution of number of A molecules after a long time.
Now, you restrict all backward reactions(generation of A doesnt happen), thus show that the probability of number of A molecules is a binomial distribution with parameters $\mu_1 + \mu_2 and 10$, considering forward rate is linearly proportional to the state (number of A molecules

at that time).

**Solution:**

Clearly, the problem is a modification of birth-death process with $q_{i,i+1} = \lambda_1 + 2\lambda_2$ and $q_{i,i-1} = \mu_1 + \mu_2$.(Note that $2\lambda_2$ is used since the 2 C molecules do backward reaction to genertate A each with rate $\lambda_2$).

Thus, Q = $\begin{pmatrix} \text{-}(\lambda_1 + 2\lambda_2) & \lambda_1 + 2\lambda_2 & 0 & .. & \\ \mu_1 + \mu_2 & -(\lambda_1 + 2\lambda_2 + \mu_1 + \mu_2) & \lambda_1 + 2\lambda_2 \ ... & .. & .. \end{pmatrix}$

Thus , soving the equation $\pi Q = 0(\pi$ is unique for a birth-death process)
we get the result as : $\pi_i = (\lambda_1 + 2\lambda_2)^i \pi_0/(\mu_1 + \mu_2)^i$

where $\pi_0 = 1/(1+\sum_{i=1}^{n}(\lambda_1+2\lambda_2)^i\pi_0/(\mu_1+\mu_2)^i)$. Note that n = 10 , since the state space(Number of A molecules is 10) is [0,10]

For second part,$\mu_i = i(\mu_1 + \mu_2)$, this is similar to a linear pure death process which is binomially distriubuted with parameters $\mu andN$(terminology as used in the course book ) thus, we replace $\mu$ by $(\mu_1 + \mu_2)$ and get the desire answer.

8. CTMC

   Consider a CTMC with given P(t) and Q matrices.
   Given the equation $2P(t)^k + 3P(t)^{k-1} = R(t)Q^{k-1}$.Solve for R

   **Solution:**

   since we know $P(t)^1 = P(t)Q$, we have $P^2 = P(t)^1Q = P(t)Q^2$ , following this recurvise relation.
   $P(t)^{k-1} = P(t)^1Q^{k-1}$.
   $P(t)^k = P(t)Q^k$.
   Also, 2P(t)$^k = 2P(t)Q^k$
   and 3P(t)$^{k-1} = 3P(t)^1Q^{k-1}$
   adding the equations we get $2P(t)^k + 3P(t)^{k-1} = 2P(t)Q^k + 3P(t)^1Q^{k-1}$
   $2P(t)^k + 3P(t)^{k-1} = 2P(t)Q^k + 3P(t)^1Q^{k-1} = (2P(t)Q + 3P(t)^1)Q^{k-1} = 5P(t)^1Q^{k-1}$
   Which is the same as given in the question, this implies $R(t) = 5P(t)^1$ is a solution of the given higher order differential equation by comparison.
   Thus, $R(t) = 5P(t)^1$
   where $P(t)^1$ is given simply by diagonalisation as $ADA^{-1}$, where D is a diagonal matrix with entire $\beta_i exp \beta_i t$

9. Queueing Models

   In a supply-chain, products follow inter arrival time as exponential distribution with parameter $\lambda$, you have a processing station with a capacity of N products for processing of the product already arrived at a time, other products simply don't enter and are routed to some other processing station, unless there is some a vacancy in the current station, the processing time is also exponential distribution with parameter $\mu$ and follows FIFO discipline. Find the average time spent by a product for processing at the given station.

   **Solution:**

   Clearly, based on the description, problem is a queueing model or specifically a M/M/1/N queing model.

We will use littles theorem for avergae time spent, note , however that arrival rate is $\lambda$ but due to derouting and products getting blocked from entering the effective arrival rate is different and is given by $\lambda_{eff} = \lambda(1 - P_N)$, where $P_N$ is the blocking probability(for this state ,N, prodcts dont enter the station).
Thus, we find the $P_N$ and $L_s$(average number of products in the station) as follows:
$P_n = (1 - \rho)\rho^n/(1 - \rho^{N+1})$ if $\rho$ not equal to 1.
else $P_n = 1/(N + 1)$ and is derived as the usual procedure for finding equilibrium probbailites.
Thus, $L_s = \sum_{n=0}^{N} nP_n$
$L_s = \rho/(\rho - 1) - (N + 1)\rho^{N+1}/(1 - \rho^{N+1})$ if $\rho$ not equal to 1 else $L_s = N/2$
Also, $\lambda_{eff} = \lambda(1 - P_N) = \lambda(1 - (1 - \rho)\rho^N/(1 - \rho^{N+1}))$

Thus, required answer is $T_s = L_s/\lambda_{eff}$

10. Queueing Models
Tandem Queue:
Consider a queueing model with two servers in tandem:
A customer with Poisson arrival rate $\lambda$, waits in line for the service from the first single-server facility and after completion, swithces to wait in line for the second single-server facility. After second completion too, the customer departs the system/facility. Assume that the first facility is a FIFO M/M/1, and the second server has exponential service times and is also FIFO. The service rates are $\mu_1$ and $\mu_2$ repectively . Assume independence of service times over all. Let X(t) = (X1(t),X2(t)), show that as $t- > inf$ $X_i(t)$ become independet rvs with goemetric ditributoin and that the second facility also behaves like M/M/1 queue.
**Solution:**
Writing the steady state Balance equations to solve for stationary probabilities:
Let $P_{n,m}$ denote steady state joint probability when there are n customers at the first facility and m at the second.
$\lambda P_{0,0} = \mu_2 P_{0,1}$(only possible transitoin)
also for $n >= 1, m >= 1$, we have
$\lambda + \mu_1 + \mu_2)P_{n,m} = \lambda P_{n-1,m} + \mu_1 P_{n+1,m-1} + \mu_2 P_{n,m+1}$(similar to M/M/1).
Letting $\rho_i = \lambda/\mu_i$, $i = 1, 2$ , and is leass than 1, we have:
$P_{n,m} = (1\rho_1)\rho_1^n * (1\rho_2)\rho_2^m$ ( $n >= 0, m >= 0$)
Now, we interpret the result as:

As $t- > inf$ , X1(t) and X2(t) become independent rvs each with a geometric distribution. Notice that they actually arent indpendent at any time t. Also notice second facility behaves like a M/M/1 queue for $t- > inf$. The result is interesting to see.
Also, Note that the first facilty is a M/M/1 queue so X1(t) is a CTMC with stationary distribution $Pn = (1\rho_1)\rho_1^n$ , $n >= 0$. If we start off X1(0) with this stationary distribution ($P(X1(0) = n) = Pn, n >= 0$), then we know that X1(t) will have this same distribution for all $t >= 0$, that is, X1(t) is stationary.Also note that when stationary, the departure process is itself a Poisson process at rate $\lambda$, and so the second facility (in isolation) can be treated itself as an M/M/1 queue when X1(t) is stationary. This sheds some light on the result.