MTL 106 (Introduction to Probability Theory and Stochastic Processes)
Assignment 1 Report

Name: Ankish Chandresh Entry Number: 2018EE10447

All questions are made by me from scratch, except few taken inspiration from M ross, papers, articles and modification and intergration of different properties.

1. Basic Probability
Imagine a circle of radius R. What is the probability to Chose three points randomly on the circle such that the center lies inside the triangle. Given that one of the points is chosen at x-axis as reference already and makes no difference in the answer if the if the pints are chosen uniformly or otherwise. Also, upto an angle from 0 to 'a' from the reference the P function in an interval [a,b] is proportional to $1/2^a - 1/2^b$, and uniform from a to 2π.
Let $a = \pi$ for calculations.
Note:
Simple geometry/symmetry techniques may not work since the distributions are not uniform, all assumptions for solving common combinatorial problem thus dont apply, all steps from considering the three axioms and bayes theorem needs to be employed.
Clearly the sample space the set if all two points on the circle with given Probability measure satisfies all three axioms. Check that P function given follows all the axioms, later the proportionality constant can be evaluated for the answer.
Concepts to be considered:
Axioms of probability, Bayes Theorem, Geometry

Solution:
From geometry, we can conlcude that we need to find the probability that the triangle formed is acute, suitable conditinos on the angles created by geometry can be employed for finding the probability.
The first point is already generated and is a point of reference. From this point, we will determine the position of A (the second point, at angle$\alpha$from reference) and then the position of B (the third point, at angle β). Let C be the center of the circle.
Let $A^{'} and B^{'}$ be the two random variables that map the angle and the point on the circle(nonuniform on $[0, a]$ and uniform on $[a, 2\pi]$). For each chosen case There could be "extreme" values of $B^{'}$:
If $0 < \alpha <= \pi$ then the β that allow C to be in the triangle are those in $[\pi, \pi + \alpha]$.
If $\pi < \alpha <= \pi + \alpha$ then the β that allow C to be in the triangle are those in $[\alpha - \pi, \pi]$.
The probability p that C is inside the triangle OAB is thus given by the sum of teh two cases mentioned above which are disjoint, usinf the axioms and also employing bayes thereom:
p = $\sum P(B^{'} in[\pi, \pi + \alpha]|A^{'} = \alpha)P(A^{'} = \alpha) + \sum P(B^{'} in[\alpha - \pi, \pi]|A^{'} = \alpha)P(A^{'} = \alpha)$
Let $a = \pi$
$P(A^{'} = \alpha) = K(1/2^{\alpha} - 1/2^{\alpha+d\alpha}) = Kd\alpha/2^{\alpha}$ (for non-uniform region)
$P(A^{'} = \alpha) = d\alpha/2\pi$ (for uniform region)
Thus sum is converted to integration, with appropriate limits.

p = $\int_0^{\pi} P(B' in[\pi, \pi + \alpha]|A' = \alpha)P(A' = \alpha)d\alpha$ + $\int_{\pi}^{2\pi} P(B' in[\alpha - \pi, \pi]|A' = \alpha)P(A' = \alpha)d\alpha$
Let $a = \pi$ (from question)
p = $\int_0^{\pi}(\alpha/2\pi)Kd\alpha/2^{\alpha}+$ $\int_{\pi}^{2\pi}(\int_{(\alpha-\pi)}^{\pi} K/2^{\theta}d\theta)P(A' = \alpha)d\alpha$
p = $\int_0^{\pi}(\alpha/2\pi)K/2^{\alpha}d\alpha+$ $\int_{\pi}^{2\pi}(K'(1/2^{(\alpha-\pi)} - 1/2^{\pi}))1/2\pi d\alpha$
p = $K0.21 + K'0.14, (where K' is K ln(2))$
Note that the P fuctoin given for 0 to $a(\pi)$ is not uniform but should satisfy the three axioms from which we have K as:
$K(P[0, \pi]) + P[\pi, \pi] = 1$
$K(P[0, \pi]) = K(1/2^0 - 1/2^{\pi}) = 1/2$
$K = 0.56$
$p = 0.17$
one can show, that in a uniform case the a=0 and p is 0.25.

2. Random Variable/Function of a Random Variable

   Imagine four machines generating a sequence of letters A,B,C,D one after the other and independently of the previous letter and independent of each other. The probabilities of generation of next letter is given below for each machine. Decoding a machine means finding out the last letter a machine has generated at that point of time by asking only yes/no type question for one or more letters at a time, (e.g. is the generated letter A? or is the generated letter among A or B?) and the machine returns a yes or no. Asking an exhaustive certain number of questions will eventually decode the machine but the total number of questions for each mahcine may vary based on the probability of generation of letters and strategy of asking the questions. Suppose each question/answer takes 1 minute, find the machines which would take the longest and smallest time on "average" to decode, given you follow the best strategy to decode on "average".

   Machine 1: P(A)=0.3,P(B)=0.3,P(C)=0.2,P(D)=0.2
   Machine 2: P(A)=0.25,P(B)=0.25,P(C)=0.25,P(D)=0.25
   Machine 3: P(A)=0.5,P(B)=0.25,P(C)=0.125,P(D)=0.125
   Machine 4: P(A)=0.4,P(B)=0.3,P(C)=0.2,P(D)=0.1
   The question is an explaination/intuitoin for **information theory** and is easy if know beforehand

   Concepts to be considered:

   Kolmogrov axioms of probability, The Problem of Interpretation in probability theory, Informatoin theory(although solution is based on intuitive approach, knnowledge of information theory makes the problem simpler), Expectation value of a random variable, Binary search.

   Note:

   With this problem, in my knowledge,one might get stuck with the problem of interpretability as uncertainity and probability are not the same and may get philosophical at times. It also calls out the need for topics like stochastic processes and expectation values. It also shows the usefulness of shannon entropy and quantifies uncertainity and information.

Sanity checks:

Clearly the given probabilities for all machines follow the three axioms of probabilities.

Best Strategy:

Firstly, the Best strategy for decoding would be to ask such a yes/no question that roughly gets rid of half of the total possible outcomes. And is similar to a binary search approach within a given number of outcomes. e.g. In a simple case Suppose a different strategy For machine 2 :

Ques 1. Is the letter D?(this doesnt eliminate half of the possible outcomes and if it is not D we need ask more ques for determining among A,B and C) Ques 2. Is the letter A?(this too doesnt eliminate half of the possible outcomes and if it is not A we need to ask one more ques for deciding between B and C) Ques 3. Is the letter B?

while for strategy mentioned above asking for a question between A and B and then C and D can take at max 2 questions.

Solution:

We try to model the problem as finding the expected value of the real valued discrete Random Variable X defined on the sample space A,B,C,D (assume a measurable space where X is a random variable) which is equal to the number of yes/no questions needed to determine the letter(A,B,C or D) given we 'Follow the mentioned best strategy'. Since X takes finite values the expectation value exists.

For a given p value no of outcomes would be 1/p and a binary search strategy would take log2(1/p) steps over these possible no of outcomes, or proportional to log2(1/p).

Thus doing a binary search would take log2(1/p) number of questions , where p is probablity of generation of the letter and 1/p is number of outcomes .
Hence, X(A) = log2(1/P(letter=A)) for A in A,B,C,D.
For machine1:
$T1 = E(X) = 0.3 * log(1/0.3) + 0.3 * log(1/0.3) + 0.2 * log(1/0.2) + 0.2 * log(1/0.2) = 1.96$

For machine2:
$T2 = E(X) = 0.25*log(1/0.25)+0.25*log(1/0.25)+0.25*log(1/0.25)+0.25*log(1/0.25) = 2$
For machine3:
$T3 = E(X) = 0.5*log(1/0.5)+0.25*log(1/0.25)+0.125*log(1/0.125)+0.125*log(1/0.125) = 1.75$
For machine4:
$T4 = E(X) = 0.4 * log(1/0.4) + 0.3 * log(1/0.3) + 0.2 * log(1/0.2) + 0.1 * log(1/0.1) = 1.84$

Clearly T2 is maximum and T3 minimum.

Thus machine2 takes the longest time to decode on average and machine3 takes the least time.

3. Two Dimensional Random Variables

Imagine an Octagon placed symmetrically about the origin with two vertices in each quadrant, and four sides prallel to x and y axis respectively with distance between each pair of parallel sided is 2. Now you drop a water drop(a ring can also be considered, but the ring of radius r reduces the distance between two parrallel side by 2r, and rest of the process is same, for calculation we skip this.) from above whose coordiates on the plane are given by (X,Y) where X and Y are real valued independent standard normal distributed random variables. What is the probability that it lands inside the Octagon. Will the answer change if the orienatation of the octagon different.

Major Concepts considered:

Transformation Theorem, Function of rv, Joint distribution of random variables, Conditional Distribution, Symmetry, Kolmogrovs axioms of Probability, Baeys Theorem

Note:

Due to symmetry constraints the question might be easier to solve but for a odd sided polygon(integral is hard to evaluate so i skip to show this) or lack of symmetry its best to convert to polar coordinates using appropriate theorems like Transformation theorem.

Since X and Y are normally distributed the drop can fall with a non-zero probability at any point on the plane.

Solution:

Clearly, for the given configuration and shape working in cartesian coordinates is difficult first we take steps **to convert to polar coordiantes.**

By **transformation theorem** and the given assumption, we have:

Note that r and $\theta$are not uniquely determined.

Assuming X>0 and Y>0 and since X,Y are symmetric about origin, we can calculate conditional expectation for all four quadrants (X<0 and Y<0..).

$r = g1(x, y) = \sqrt{x^2 + y^2}, \theta = g2(x, y) = tan - 1(y/x)$(using the same notation/convention, here we know $f(z, w) i.e. f(x, y)$ and $f(g1(x, y), g2(x, y)) = f(r, \theta)$ can be calculated going backwards if we know $J(z, w) = J(x, y)$.)

$J(x, y) = 1/\sqrt{(x^2 + y^2)} = 1/r$

$f(x, y/X > 0, Y > 0) = f(x, y)/P(X > 0, Y > 0) = 2e^{-(x^2+y^2)/2}/\pi$(since X and Y are independent $f(x, y) = f(x)f(y)$)

$f(r, \theta/X > 0, Y > 0) = 2re^{-r^2/2}/\pi \ 0 < \theta < \pi/2, 0 < r < inf$

Similarly for all four cases for all four quadrants we get same conditional pdf.(Conditoinal pdf is used for simplification else all four cases can be solved differently and summed upas in case of non-uniqueness.)

In accordance with the axioms of probability, Bayes theorem and symmetry. We can write:

$f(r, \theta) = f(r, \theta/X > 0, Y > 0)/4$

$f(r, \theta) = re^{-r^2/2}/2\pi \ 0 < \theta < 2\pi, 0 < r < inf$

Now, from geometry and for first quadrant we have :

For one of the triangles :

$$\int_0^{\pi/8} \int_0^{d/cos(\theta)} (re^{-r^2/2}/2\pi) dr d\theta$$

Where d=1 (from question).

Taking limits $\pi/8 to 2\pi/8$ and sequentially and accordingly changing the limits for r till we reach 2π, and adding (since these are disjoint sets) we get roughly 0.641 as the required probability.
While seuqentially incrementing limits on θ one might observe probability of all the intervals of size$\pi/8$ are equal, this highlights the fact that θ is independent of r and uniformly distributed, thus one can conclude that the orientation of the shape doesnt matter with the given set up and distribution of X and Y, and answer remains 0.64.

Thus probability of drop landing within the octagon is 0.64 given any orienation if its centerd at the origin.

4. Two Dimensional Random Variables
   You are watching a drunk person walking on a road, taking left and right foot forward covering some distance, succesivley and alternatively. The data points for Left and Right leg are thus recorded, the L and R stepsize and are not necessarily independent.
   Find a rough approximation of the distance covered in m L and n R steps starting from L. (0<=L<=1unit and 0<= R<=1unit)
   For simplicity of calculation we assume m=n=1.
   Note:
   The given question cant be framed and solved here, hence we give an algorithm/way to estimate the distribution of distance covered.
   Pdf of function of two(or more) independent random varibles can simply be given as the convolution of pdf of the two rvs. However, to calculate the pdf of function(X+Y,XY,X/Y) of X and Y we need joint distribution of X and Y, which is trivial for independent case. So the prblem is essentially finding f(x,y) given f(x),f(y) when X and Y are not necessarily indepemnt and have some non-zero correlation coefficient, note that in the question X and Y are absouletly continous.
   A possible way is mentioned in the solution, and is a matter of research.
   Concepts Considered:
   Desnity Estimation methods, Function of random variable, two dimesnional random varible, joint probability distribution, Axioms of probability, Transformation Thoerem.
   Solution:
   Since, the total distance can be given by the distance travelled by the L(X) and then R(Y) step we get:
   Z = h(x,y) = X + Y (where X and Y are not independent and their marginal distrbutions are known and f(x,y) is not know or datapoints hence pdf of Z is not solveable but can be done through density estitimation techniques.(Discretization of continous is also employed in some cases, one reason being implementaion on software))

If the random variables under consideration are discrete, the joint probability distribution is easily obtained by using the relative frequency definition of probability and applying the discrete transformation of variables directly. We are trying to solve for continuous case thus, we follow the following steps:
We employ **multivariate kernel density estimation** method for calculating joint distributoin from the given data points, for the question they are observed and recorded.Which is essentially evaluating a chosen at fixed points and summing up.
The kernel density estimator provides the means by which the joint probability density function may be approximated and it has been demonstrated that even with a relatively small number of observations, good overall approximations of distributions of functions of random variables may be achieved.

For some parameters like the kernel, smoothing parameters ,a suitable region of (L,R) datapints,etc can be chaosen and the joint distribution of $(L, R) or (X, Y), f'$ can be calculated by suitable formulae empolying different methods of KDE.
This distribution may further be used for pdf of $Z = h(X, Y)$ by transformation theorem for continuous case, the measure is not accurate and for implementaion in software more discretization of the problem is needed which we skip. Also, for KDE to give good estimates one may need to do data scaling or preprocessing.
Histrogram estimator are also an option but KDE give better results especially in a multivaiate setting.

5. Higher Dimensional Random Variables
Imagine dropping balls in a straight line that dont interact with each other falling over a pyramid of equally sapced wooden pegs(similar to pascals triangle, each smaller triangle replaced by a wooden peg) chopped off from the top with top layer having 3 pegs and last bottom layer has 6 pegs, at constant average rates of $\lambda_1, \lambda_2, \lambda_3$ for 1 unit of time interval,for each peg in top layer respectively and independent of each other, the dropping is however rare to occur. Balls have weight that is sampled from exp(3), but before dropping they pass thorugh a grinder that reudces weight to its sqaure root.
Probability of ball going left is p and going right is q after bouncing of a peg.
(Let p=q=1/2 for simplicity in calculation.) Determine the maximum Expected weight of balls recieved in any of the containers(if $\lambda_1 = \lambda_2 = \lambda_3 = 1$ for easy calculation).
Determine m ,the depth of the pyramid such that Expected weight recieved by third container is maximised.(given $p = q = 1/2$). What happens if the middle peg from the top is removed?
Concepts considered:
Random Sum, Different Probability distributions, Reproductive Property, Conditional expectation, Fucntoin of a rv, Expectation value.
Solution:
With the given description of dropping balls we can assume the dropping of the balls for each peg is poisson distributed for each peg.
Also, for a complete pyramid of pegs, one can show the balls recived at the end of each

layer/container is binomial distributed for each indpendent peg.Thus for the given setup we can assume total balls revieved in the container is the **sum of three such independent binomial distibutions. Note that the three dont simply add up for all pegs, and overlap partially and reprodcutive property for binomial distribution cant be used simply.**
For peg 1 on top layer, distribution of balls recieved from peg1 to peg 4 in bottom layer is as follows:
$p1\ P(3C0\lambda_1/2^3)$
$p2\ P(3C1\lambda_1/2^3)$
$p3\ P(3C2\lambda_1/2^3)$
$p4\ P(3C3\lambda_1/2^3)$
Where P implies poisson distrinution
For peg 2 on top layer, distribution of balls recieved from peg2 to peg 5 in bottom layer is as follows:
$p2\ P(3C0\lambda_2/2^3)$
$p3\ P(3C1\lambda_2/2^3)$
$p4\ P(3C2\lambda_2/2^3)$
$p5\ P(3C3\lambda_2/2^3)$
For peg 3 on top layer, distribution of balls recieved from peg3 to peg 6 in bottom layer is as follows:
$p3\ P(3C0\lambda_3/2^3)$
$p4\ P(3C1\lambda_3/2^3)$
$p5\ P(3C2\lambda_3/2^3)$
$p6\ P(3C3\lambda_3/2^3)$
now adding up respective distributions and using reproductive property of Poisson distribution, we get distributions of no of balls recived for each conatainer:
$p1\ P(3C0\lambda_1/2^3), p4\ P((3C3\lambda_1 + 3C2\lambda_2 + 3C1\lambda_3)/2^3)$
$p2\ P((3C1\lambda_1 + 3C0\lambda_2)/2^3), p5\ P((3C3\lambda_2 + 3C2\lambda_3)/2^3)$
$p3\ P((3C2\lambda_1 + 3C1\lambda_2 + 3C0\lambda_3)/2^3), p6\ P((3C3\lambda_3)/2^3)$
Clearly, the expectation value can now easily evaluted and compared, with $\lambda_1 = \lambda_2 = \lambda_3 = 1$.
Now, if a container receives n balls then total weight, $W = W_1 + W_2 ... W_n$. Note thatW_i are squareroot of actual weights, so its expected value can be calculated by using concpets from funtion of a rv given it exists, and $E(W_i) = 2/3$
Using random sums, we get $E(W) = E(W_i)E(N)(W_i$ can be assumed independent where N is no of balls recived by the container below a peg in the last layer),**for each container, and is maximum for containers below peg3 and peg4, under given conditions.(answer may change if p and q are not equal).**
Also, one may notice, since the Expected weight recived is decided only by the expected no of balls recived from random sums, and for $p = q = 1/2$, expectation of the conatinders below middle peg/s is always maximum. Thus choose $m = 3$(depth the pyramid) , such that the third peg is the middle peg for the second part of the question.
If the middle peg in the top layer is removed then that pegs controbution in the no of balls

recieved is a binomial distiribution with n parameter reduced by 2, for the given case the parameter reduces from 3 to 1, similar calculations can then be done, interesting patterns
Higher Dimensional Random Variables
Consider a random vector X = (X1,X2,X3) sampled from a distributoin with mgf = $(1+e^t)/2$. A weighted dimensionality reduction matrix $W(2*3)$ then exp matrix that is exponentiation of the input(W) matrix converts X to Y of dimension 2. Weights of row 1 and 2 of W are independently sampled from exponential distribution wirh parameter a and b respectibely and then transformations are applied. Find the range of values of a if b=2 such that the elements in the column of expectation vector of Y are decreasing.(For ease of calculation demensions are kept low). Find the joint pdf f(a,b) if a and b are chosen independently from U[2,5], satisfying the given condition.
Concepts Considered:
Order statistics, Linear Algebra, Random Vector, Conditional expectation for random vectors, MGF if it exists, Law of unconcious statistics.
Solution:
Check that the given mgf is a valid mgf by checking its positive and log-convex,and M(0)=1.
All results on condditional expectation can also be applied for matrices/vectors. Thus we get:
$E(\mathbf{Y}) = E(E(\mathbf{Y}/\mathbf{W}))$ (Y and W could be vectors or matrices,its interesting to note that the dimensions of the Y and W may be different)
After the first linear transformation and expoentiaition we get:
$W'^T = (e^{w1X1+w2X2+w3X3}, e^{w4X1+w4X2+w5X3})$
$E(W'^T) = (E(e^{w1X1+w2X2+w3X3}), E(e^{w4X1+w4X2+w5X3}))$
$E(W'^T) = (E(e^{w1X1})E(e^{w2X2})E(e^{w3X3}), E(e^{w4X1})E(e^{w4X2})E(e^{w5X3})))$
This means evaluation of mgf of X at w1,w2...
$E(W'^T) = ((1+e^{w1})(1+e^{w_2})(1+e^{w_3})/8, (1+e^{w_4})(1+e^{w_5})(1+e^{w_6})/8)$
$E(W') = E(\mathbf{Y}/\mathbf{W})$
$E(E(W')) = E(Y^T)$
$E(Y^T) = (E((1+e^{w1})(1+e^{w_2})(1+e^{w_3})/8), E((1+e^{w_4})(1+e^{w_5})(1+e^{w_6})/8))$
Expectation(if it exists) of F(w1,w2,w3) can then be know if f(w1,w2,w3) is known, here they are sampled indpendelty from exp(a) and exp(b) so we get, the following result.
$E(Y^T) = (E(1+e^{w1})E(1+e^{w_2})E(1+e^{w_3})/8, E(1+e^{w_4})E(1+e^{w_5})E(1+e^{w_6})/8))$
$E(Y^T) = (((2a-1)/2(a-1))^3, ((2b-1)/2(b-1))^3)$
Clearly the elements in the vector are an decreasing function of a or b.
Thus we chose a<b such that the requirements are met, thus a in (1,b) is the required range.
Note that in order for the expectation to exist a>1 and b>1.
Thus, by order statistics f(a,b) =2!f(a)f(b).Where f(a)=1/4=f(b) and a<b.

6. Cross Moments
You are instructed to buy two strings from the two different shops that sell string of length greater than 1, considering many factors like location , size of the shop etc, length bought

given by continous rv X and Y(from each shop). Clearly X and Y may not be independent. After observing the shops you see the lengths sold are close to the expected length of string sold by each shop. The cost of buying is 2ln(X) and 4ln(Y) and are uncoorelated.
The continuous Random Variables X and Y taking values greater than 1 have expectations μ_1, μ_2 and variance σ_1^2, σ_2^2 respectively, given all moments exists. Note that X an d Y may not be iid.
You then divide each string in half and stretch each string segment raised to some positive interger power a >=1 and b >=1 respectively to form a square (ie square of dimensions $(X/2)^a * (Y/2)^b$) , find the expected area of the square and the cost, take asssumption for suitable approximations.

Concepts Considered:
Cross Moments, Meyers Approximation, Function of a random variable, Expectation and Variance of sum of two random variables if it exists, convergence of moments

Sanity Checks:
Question is find approximate measure of the cross momet $E\left(X^aY^b\right)/(2^a2^b)$(constant can be taken out by property) centered at zero, where a and b are positive integers.(ignore the constant $1/(2^a2^b)$) for now)
Clearly since all moments exists this implies $E(X^a), E(Y^b)$ also exists.
Also note the random variables take values greater than 1, which means expectation and variance of a Random Variable $X^{'} = ln(X)$(also for Y) exists, since it is absolutely summable/intergrable since |ln(x)|<x for x>1 and thus intergral/summation converges.

Also note X and Y may not be iid, zero covariance don't imply independence.
$Cov\left(ln(X), ln(Y)\right) = 0$(the constants 2,4 can be neglected using properties of Covariance.)

Solution:
Clearly its hard to get an accurate measure of $E\left(X^aY^b\right)$ but we can use Meyer Approximations to get a rough estimate of this value. Any bounding trick like Cauchy-swarchtz inequality wont work too since the moments of higher order.

Let
$$X^{'} = ln(X),\ Y^{'} = ln(Y)$$
and
$$A = ln(X^aY^b) = aln(X) + bln(Y)$$

Since expectation of ln(X) and ln(Y) exists, the expectation of their linear combination also exists and can be proved, also can be evaluted separately and comibned linearly. $i.e. E(A) = aE(ln(X)) + bE(ln(Y))$

For getting $Eln(X)$ roughly, we use Meyer Approximation,since $ln(X)$ is doubly differentiable for all values under given assumptions and the $X - E(X)$ is small(also for Y) , as follows

$$E(X^{'}) = ln(\mu_1) - \sigma_1^2/2\mu_1^2$$

$$E(Y^{'}) = ln(\mu_2) - \sigma_2^2/2\mu_2^2$$

Similarly for variance,

$$Var(X^{'}) = \sigma_1^2/\mu_1^2$$

$$Var(Y^{'}) = \sigma_2^2/\mu_2^2$$

Thus,
$E(A) = a(ln(\mu_1) - \sigma_1^2/2\mu_1^2) + b(ln(\mu_2) - \sigma_2^2/2\mu_2^2)$
$Var(A) = a^2\sigma_1^2/\mu_1^2 + b^2\sigma_2^2/\mu_2^2$ ($ln(x)$ and $ln(Y)$ are uncorrelated)
Now the required random variable is clearly $A^{'} = e^A$
Again using **Meyer Approximation**,(since assumptions still hold) as exponential is doubly differential for all values and $A - E(A)$(from the assumption that X-E(X) and Y-E(Y) are small)is small, as follows:
$E\left(X^aY^b\right) = E(A^{'}) = e^{E(A)} + Var(A)e^{E(A)}/2$
$E\left(X^aY^b\right) = e^{a(ln(\mu_1)-\sigma_1^2/2\mu_1^2)+b(ln(\mu_2)-\sigma_2^2/2\mu_2^2)}(1 + (a^2\sigma_1^2/\mu_1^2 + b^2\sigma_2^2/\mu_2^2)/2)$
$E\left(X^aY^b\right) = (\mu_1^a\mu_2^b)(e^{-\sigma_1^2/2\mu_1^2-\sigma_2^2/2\mu_2^2})(1 + (a^2\sigma_1^2/\mu_1^2 + b^2\sigma_2^2/\mu_2^2)/2))/2^a2^b$
Hence, expected area of square formed by the strechted strings is
$(\mu_1^a\mu_2^b)(e^{-\sigma_1^2/2\mu_1^2-\sigma_2^2/2\mu_2^2})(1 + (a^2\sigma_1^2/\mu_1^2 + b^2\sigma_2^2/\mu_2^2)/2))$
Hence, expected cost of square formed by the strechted strings is $E(ln(XY))$ ie E(A) when $a = b = 1$

7. Cross Moments
You have N signals with M samples. You linearly Combine the given N signals to get N new signals.You want that the new signals are uncorrelated and trace of coovariance matrix of the signals is N. Comment if such process can be done, if yes how, if no when?
Suppose N»M(very large) when signal only has 1/0 as elements, what can you conclude?Can we chose signals such that the process is always followed?
Assume E{M} means averaging over each column of matrix M.
Concepts Considered:
Linear Algebra, Eigenvalue decomposition, Covariance matrix, Signal Processing, Random Vector, whitening matrix and techniques, pegeonhole principle, diagoanlisability of coovariance matrix
Note:
Suppose X is a random vector of dimension n and zero mean with non-singular covariance matrix V.
The problem is essentially finding L(whitening matrix),some linear combination, such that Y = LX and $L^TL = V^{-1}$. Covriance matrix is I and so Y follows all the given conditions in the question, and is thus one possbile answer.
There are different methods employed for finding the whitening matrix L.
One of them is $L = V^{-1/2}$. Others being the Cholesky decomposition of V^{-1}.

Solution:
Solution of the above problem is prewhitening of signals and is very commmon in signal processing.
Let X is the N*M matrix for the original signals, M=1 for explanation above.
Finding if such process can possibly be done is equivalent to finding a linear transformation L on centered X such that $E\{L(X_c)L^T(X_c)\} = I_n$. We employ a common approach among many of eignevalue decomposition on the covariance matrix of the centered data(Center each row).
$E\{X_cX_c^T\} = EDE^T$ where X_c means centered X.
where E is the matrix of eigenvectors and D is the diagonal matrix of eigenvalues. Mutliply X by E^Tgives a uncoorelated covariance matrix, for the trace condition in qquestion one solution can be that the coovariance matrix is identity, thus trace is N.
The next step is whitening step.
The whitened data matrix is defined thus by $Y = D^{-1/2}E^TX_c$(can be derived from D-1D = I)
$ThusL = D^{-1/2}E^T$.
Also, The given process can be achieved thorugh eigenvalue decomposition only if the matrix $E\{X_cX_c^T\}$ is diagonalizable and can be verified for any data given.
Now, for second part, even if we establish by piegonhole principle that atleast one row is repreated (one row in linear combination of other rows(i.e singular)), interestingly enough nothing can be concluded for exitence of required L, since $E\{X_cX_c^T\}$ matrix is also singular but a singular matirx can still be diagonalizebale.
For third part, one can prove if the matrix is symmetric,real or positive semdefinite (by definition and from the question), it is always diagonalisable, and so is a coovarinace matrix, hence the above process is always followed since the coovariane matrix is always diagonalisable.

8. Limiting Distributions
You measure two types of error in a component everyday given by X_iandY_i, and then fix it and error nextday changes such that X_i and Y_i based on some sequnce, after a long time(about a year) almost surely converges to X and Y repectively and both are normally distributed $N(1, 2), N(2, 1)$ repectively. However evaluation of a compomnent is differently done, On a day you take the X_iandY_i and multiply to get a term with three errors in every possible way such that order also matters and add the them ie $(XXX + XXY + YXX..)$, the total error evaluation value at the end of the year are again added for new components for 50 years and average is calculated, find the probability taht the average is greater than 0.
Major Concepts Considered:
CLT, modes of convergence, ditribution of function of a rv, moment genenrating fucntion, Chernoff bound, result on almost surely convergenve of f(Xn) and $X_n + Y_n$.
Solution:
Since X_iandY_i converge as to X and Y, thus they also converge in distribution.

Also, the evaluation value is simply $(X_i + Y_i)^3$ for a day, we have the following colclusoins:

$X_i + Y_i$ as congeres to $X + Y$
$F(X_i + Y_i)$ as converges to $F(X + Y)$(by result and cane be proved)
Thus, $(X_i + Y_i)^3$ as converges to $(X + Y)^3$ at the end of the year, based on the description.
Thus, we know, the distribution and all moments(if it exists) and relevant information of $Z = (X + Y)^3$, since we know surely about X and Y, and also that X+Y are normally distirbuted with μ =3 and$variance = 3$
Now, since Z_i at the end of each year are independent and 50 is large enough using CLT we get,
$Er = (Z1 + Z2..Z50)/50$ (Note that the distribution of Z_i are iid and distribution could be very different from a normal distribtuon but CLT approximation gives a good estimate)
Er converges in distribution to normal distributoin (Expetaion and Var are to calcualted)
$P(Er > 0) = P((Er - E(Er))/\sqrt{Var(Er)})) > -E(Er)/\sqrt{Var(Er)}) = 1 - \phi(-E(Er)/\sqrt{Var(Er)})$
$P(Er > 0) = 1 - \phi(-E(Er)/\sqrt{Var(Er)})$
For E(Er) and Var(Er) we employ different methods, **mgf/fourier transform if it exists, Law of unnconsicoius statsictics, binomial relation between nth and 1st order moments**. Since the X an Y are nomally distributed, we get.
We get $E(Er) = \mu^3 + 3\mu\sigma^2$ and $50Var(Er) = \mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6 - (\mu^3 + 3\mu\sigma^2)^2$
Where $\mu = 3, \sigma^2 = 3$
(from reproductive property)

Thus, answer is $1 - \phi(-(2\sqrt{(}10/3)))$

9. Limiting Distributions
Your are throwing a ball from a fixed point (origin) n(>1) times on a straight line, such that postion of the ball is normally dustributed(zero-mean) and $variane = 1/ln(n)$, independence should not be assumed. Find n such that the Probability that maximum distance from the origin of the ball is gretaer than 3 is upper bounded by 0.5.
Concepts considered:
Jensens inequality(generalisation of the result of E(X2)>=(E(X))2), Moment Generating functions, Random vector, Expectation Value, Union Bound,Markov inequality.
Solution:

Suppose a real valued random variable $Z = max\{|X1|, |X2|, ...|Xn|\}$, which is the required rv for the problem.
First we prove $E(Z) <= 2\sqrt{\sigma^2 ln(n)}$
Following a bounding trick, particularly using Union Bound(which comes as a result of the three axioms of probability) and jensens inequality seems a viable option, here.
$e^{\lambda E\{max\{|X1|,|X2|,...\}\}} <= E\{e^{\lambda max\{|X1|,|X2|..\}}\}$(exp is. convex ,$\lambda > 0$)
or
$e^{\lambda E\{max\{|X1|,|X2|,...\}\}} <= E\{max\{e^{\lambda|X1|}, e^{\lambda|X2|}..\}\}$
Then By **union bound**(all conditions are satisfied) we have
$e^{\lambda E\{max\{|X1|,|X2|,|X3|...\}\}} <= \sum_{i=1} E\{e^{\lambda|X_i|}\}$

$e^{\lambda E\{max\{|X1|,|X2|,...\}\}} <= ne^{\lambda^2\sigma^2}$
Taking ln and differentiating wrt λ for tighter bound we get:
$E\{max\{|X1|,|X2|,|X3|...\}\} <= ln(n)/\lambda + \lambda\sigma^2 <= 2\sqrt{\sigma^2 ln(n)}$
$E\{Z\} <= 2\sqrt{\sigma^2 ln(n)}$
where $\lambda = \sqrt{ln(n)}/\sigma$
now , we prove $P(Z >= 2\sqrt{\sigma^2 ln(n)} + k) <= e^{-k^2/2\sigma^2}$
Using **markov inequlity**, since Z is non-negative rv.
$P[Z >= 2\sqrt{\sigma^2 logn} + k] = P[e^{\lambda Z} >= e^{\lambda(2\sqrt{\sigma^2 logn}+k)}]$
$P[e^{\lambda Z} >= e^{\lambda(2\sqrt{\sigma^2 logn}+k)}] <= E[e^{\lambda Z}](e^{-}\lambda 2\sqrt{\sigma^2 ln(n)})(e^{-}k\lambda)$
Using **jenese inequality** for e^x and above derived result for E(Z) :
$P[e^{\lambda Z} >= e^{\lambda(2\sqrt{\sigma^2 logn}+k)}] <= 2nexp((\lambda^2\sigma^2)(\lambda(k + 2\sqrt{\sigma^2 logn})))$
Differentiating wrt to λ we get a tigher bound
$P[e^{\lambda Z} >= e^{\lambda(2\sqrt{\sigma^2 logn}+k)}] <= 2nexp(-((k + 2\sqrt{\sigma^2 logn})^2)/2\sigma^2) = 2exp(-1/2\sigma^2[2\sigma^2 logn + k^2 + 4k\sqrt{\sigma^2 logn} + 4\sigma^2 logn])$
With a loose upperbound
$P[Z >= 2\sqrt{\sigma^2 logn} + k] = P[e^{\lambda Z} >= e^{\lambda(2\sqrt{\sigma^2 logn}+k)}] <= 2exp(-k^2/2\sigma^2)$
According to the question:
$\sigma^2 ln(n) = 1$
Let k =1
$P[Z >= 2 + k] = P[Z >= 3] <= 2exp(-/2\sigma^2) = 2exp(-ln(n)/2) = 2/\sqrt{(n)}$ Thus n can. be chosen as 16 for the upper bound to be 0.5. Note that for the inequality above the loose upperbound a much lower n can be chosen for the tighter bound if calculations are easier, however the proved inequlity with loose bound gives a possible answer.