

M.E. INFORMATION TECHNOLOGY FIRST YEAR FIRST SEMESTER – 2025
PROBABILITY, STATISTICS AND STOCHASTIC PROCESS

Times: 3hr

Marks: 100

CO1: (I) 30 Marks

- A. Suppose that X takes values between 0 and 1 and has probability density function $2x$. Compute $\text{Var}(X)$ and $\text{Var}(X^2)$. (5)
- B. Suppose 100 people all toss a hat into a box and then proceed to randomly pick out of a hat. What is the expected number of people to get their own hat back? (8)
- C. Suppose you are taking a multiple-choice test with c choices for each question. In answering a question on this test, the probability that you know the answer is p . If you don't know the answer, you choose one at random. What is the probability that you knew the answer to a question, given that you answered it correctly? (8)
- D. The Randomizer holds the 6-sided die in one fist and the 8-sided die in the other. The Roller selects one of the Randomizer's fists and covertly takes the die. The Roller rolls the die in secret and reports the result to the table. Given the reported number, what is the probability that the 6-sided die was chosen? (Find the probability for each possible reported number.) (9)

CO2: (II) 20 Marks

- A. Suppose that the cdf of X is given by:

$$F(a) = \begin{cases} 0 & \text{for } a < 0 \\ \frac{1}{8} & \text{for } 0 \leq a < 2 \\ \frac{3}{8} & \text{for } 2 \leq a < 4 \\ 1 & \text{for } a \geq 4. \end{cases}$$

Determine the pmf of X . (3)

- B. If birth weights in a population are normally distributed with a mean of 109 oz and a standard deviation of 13 oz, (6)
 - a. What is the chance of obtaining a birth weight of 141 oz or heavier when sampling birth records at random?

b. What is the chance of obtaining a birth weight of 120 or lighter?

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9959	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

C. If a disease is fairly rare and the antibody test is fairly expensive, in a resource-poor region, one strategy is to take half of the serum from each sample and pool it with n other halved samples, and test the pooled lot. If the pooled lot is negative, this saves $n-1$ tests. If it's positive, then you go back and test each sample individually, requiring $n+1$ tests total. (6)

- Suppose a particular disease has a prevalence of 10% in a third-world population and you have 500 blood samples to screen. If you pool 20 samples at a time (25 lots), how many tests do you expect to have to run (assuming the test is perfect!)?
- What if you pool only 10 samples at a time?
- 5 samples at a time?

D. Find the probability of getting at least 5 times head-on tossing an unbiased coin for 6 times by using the binomial distribution. (5)

CO3+ CO4 : (III) 20 Marks

- A. Explain the concept of random sampling and measure central tendency, correlation, and regression. (5)
- B. A teacher recorded the following test scores for her class: [85, 78, 92, 88, 84, 90, 79, 81, 87, 93, 78, 84]. (6)
 - a. Calculate the mean score.
 - b. Determine the median score.
 - c. Identify the mode(s) if any.
- C. Find covariance for following data $x = \{6, 5, 3, 4, 2\}$, $y = \{12, 10, 8, 6, 4\}$ (3)
- D. Which of the following is true for the coefficient of correlation? (Explain) (2)
 - a. The coefficient of correlation is not dependent on the change of scale
 - b. The coefficient of correlation is not dependent on the change of origin
 - c. The coefficient of correlation is not dependent on both the change of scale and change of origin
 - d. None of the above
- E. Which of the following techniques is an analysis of the relationship between two variables to help provide the prediction mechanism? (Discuss) (2)
 - a. Standard error
 - b. Correlation
 - c. Regression
 - d. None of the above
- F. If change in one variable results in a corresponding change in the other variable, then the variables are..... (Explain) (2)
 - a. Correlated
 - b. Not correlated
 - c. Any of the above
 - d. None of the above

CO5: (IV) 30 Marks

- A. I have 4 umbrellas, some at home, some in the office. I keep moving between home and the office. I take an umbrella with me only if it rains. If it does not rain I leave the umbrella behind (at home or in the office). It may happen that all umbrellas are in one place, I am at the other, it starts raining and must leave, so I get wet.
 - a. If the probability of rain is p , what is the probability that I get wet?
 - b. Current estimates show that $p = 0.6$ in Edinburgh. How many umbrellas should I have so that, if I follow the strategy above, the probability I get wet is less than 0.1? (12)

OR

Assume that a man's profession can be classified as professional, skilled labourer, or unskilled labourer. Assume that, of the sons of professional men, 80 percent are professional, 10 percent are skilled labourers, and 10 percent are unskilled labourers. In the case of sons of skilled labourers, 60 percent are skilled labourers, 20 percent are professional, and 20 percent are unskilled. Finally, in the case of unskilled labourers, 50 percent of the sons are unskilled labourers, and 25 percent each are in the other two categories. Assume that every man has at least one son, and form a Markov chain by following the profession of a randomly chosen son of a given family through several generations. Set up the matrix of transition probabilities. Find the probability that a randomly chosen grandson of an unskilled labourer is a professional man. (12)

B. Discuss the nature of the following Process:

A bus taking students back and forth between the dormitory complex and the student union arrives at the student union several times during the day. The bus has a capacity to seat K persons. If there are K or fewer waiting persons when the bus arrives, it departs with the available number. If there are more than K waiting, the driver allows the first K to board, and the remaining persons must wait for the next bus. The university administrators would like to know at each time period during the day how many students are left behind when a bus departs. (9)

C. Illustrate an application of the Hidden Markov Model in Natural Language Processing with an example. (9)
