CS215 **Data Analytics and Interpretation** **Endsem**

19/11/23 **Time- 05.30 pm to 7.30 pm (2 hours' duration)** **Marks: 40**

(There are 10 questions. First 8 questions are MCQs carrying 3 marks each. You get full marks for ticking ALL correct options, else you get -1. The last two questions are subjective; you have to write your answers on two sides of only ONE A4 sheet which is provided. Write your name and roll no on **both** sides of the A4 size paper.)

Q1. Let $E(X)$ denote the expectation of a random variable and be equal to 12. The variance $V(X)$ is 9. Then $E[(X+3)^2]$ is:

(a) *215*
(b) *234*
(c) *81*
(d) *96*

Ans: (b)

Justification: $E[X^2+6X+9]=E(X^2)+6E(X)+9= E(X^2)-(E(X))^2+(E(X))^2+6E(X)+9=V(X)+[E(X)+3]^2=9+(12+3)^2=234$

Q2. Suppose Z_1 and Z_2 are two independent random variables with distribution Uniform(0,1). What is $E(\min(Z_1, Z_2))$?

(a) ½
(b) 1/3
(c) ¼
(d) None of the above

Ans: (**b**)

Q3. An exponential random variable X has a pdf which is given as $f_X(x) = \lambda e^{-\lambda x}$ where $x \in [0, \infty)$ and $\lambda > 0$. What is the pmf of floor(X)? Recall that floor(X) is the largest integer less than or equal to X.

(a) $e^{-\lambda(n+1)}$
(b) $e^{-\lambda n}$
(c) $e^{-\lambda n}(1-e^{-\lambda})$
(d) *None of the above*

Ans: (**c**)

Justification: $P(\text{floor}(X) = n) = P(n \leq X < n + 1) = F_X(n + 1) - F_X(n) = (1-e^{-\lambda(n+1)}) - (1-e^{-\lambda n}) = e^{-\lambda n}(1-e^{-\lambda})$.

Q4. Let X and Y be random variables with MGFs $\phi_X(t)$ and $\phi_Y(t)$ respectively. Let Z be a third r.v. which is equal to X with probability p, and equal to Y with probability $1-p$. Then the MGF of Z in terms of ϕ_X, ϕ_Y, t and p is given by:

(a) $\phi_Z(t) = \phi_X(t) + (1-p)\phi_Y(t)$
(b) $\phi_Z(t) = p\phi_X(t) + \phi_Y(t)$
(c) $\phi_Z(t) = 0.5p\phi_X(t) + (1-2p)\phi_Y(t)$
(d) *None of the above*

Ans: (**d**)

Justification: $\phi_Z(t) = p\phi_X(t) + (1-p)\phi_Y(t)$, because

$P_Z(r)=p.P_X(r)+(1-p).P_Y(r)$

$\varphi_X(t)=\Sigma e^{tr}.P_X(r)$

$\varphi_Y(t)=\Sigma e^{tr}.P_Y(r)$

$\varphi_Z(t)=\Sigma e^{tr}.P_Z(r)= \Sigma e^{tr}.[p.P_X(r)+ (1-p).P_Y(r)]$

$= p.\Sigma e^{tr}.P_X(r)+ (1-p). \Sigma e^{tr}.P_Y(r)$

$=p.\varphi_X(t)+(1-p).\varphi_Y(t)$

Q5. Let X be a random variable in such a way that $\log X \sim N(3, 4)$, i.e., *logX* has a Gaussian distribution with mean 3 and variance 4. Then the median of X is:

(a) e^3
(b) $0.5e^3$
(c) e^4
(d) *None of the above*

Ans: (a)

Justification: Let $Y = \log X$. The median of Y must equal 3 as it is Gaussian distributed with mean 3. This implies $P(Y \geq 3) = 0.5$, that is $P(\log X \geq 3) = 0.5$ and hence $P(X \geq e^3) = 0.5$. Thus the median of X is e^3.

Q6: There are two brands of light bulbs: A and B. The mean lifetime of A light bulbs is 1400 hours and that of B bulbs is 1200 hours. The corresponding standard deviations are 120 hours and 80 hours respectively. 50 samples of each brand have been considered. In making the statement that "the difference of mean life time between A bulbs and B bulbs lies in the range 165 to 235 hours", my confidence level will be closest to:

(a) 90%
(b) 95%
(c) 99%
(d) None of the above

Ans: **(a)**

Justification:

$X_A-X_B=1400-1200=200$, where X_A and X_B are sample means.

sqrt(120^2/50+80^2/50)=20.4

90% confidence limits are 200+1.33*20.4= 227.13 and 200-1.33*20.4=172.87, i.e. **(172,228)**

95% confidence limits are 200+1.96*20.4= 239.98 and 200-1.96*20.4=160.01, i.e. **(160,240)**

99% confidence limits are 200+2.58*20.4= 252.63 and 200-2.58*20.4=147.37, i.e. **(147,253)**

The observed values fall only within the 90% C.I.

Q7. The breaking strength of cables produced by a manufacturer is 800 kg with standard deviation of 35 kg. The manufacturer claims that by a change of technology, he has been able to increase this breaking strength. To test this claim, a sample of 49 cables is tested. Based on this test, I support the claim of the manufacturer, and I am confident that I will be wrong in only 1 out of 100 cases. What mean breaking strength of cables have I observed in the sample, which gives me this confidence? Choose the closest option from below.

(a) At least 810
(b) At least 812
(c) Between 705 and 805
(d) None of the above

Ans: (d)

Justification: The question implies 99% confidence interval. Upper 1-side test neds to be done, with Z_c=2.33

H_0: μ>=800; H_A: μ<800

Under H0, Z_o=(X-800)/(35/sqrt(49))=(X-800)/5, where X is the observed sample mean.

Since I support the claim of the manufacturer with 99% confidence, Z_o falls with the confidence interval, i.e., Z_o is not greater than 2.33.

⇨ Z0<=2.33
⇨ X<=800+5*2.33=811.65

i.e., X is at most 811.65.

Q8. Suppose a fuzzy rule is given as *R:"those who are tall are fat"*. Let the profiles of '*tall*' and '*fat*' be given respectively by the curves $Y_1=min(1,X_1/6)$ and $Y_2=min(1,X_2/60)$, where X_1 is height in ft and X_2 is weight in kg. A person Ram has a height of 5.4 ft. If the fuzzy truth value of *R* is 0.8, then the weight of Ram is:

(a) between 40 and 50 kg
(b) between 50 and 70 kg
(c) between 70 and 90 kg
(d) None of the above

Ans: (a)

Justification: from Y_1, μ_{tall}(5.4ft)=0.9, use Lukasiewitz rule: t(P→Q)=min(1,1-t(P)+t(Q))

0.8=min(1, 1-0.9+t(Q))=min(1,0.1+t(Q))

➔ 0.8=0.1+t(Q)

➔ t(Q)=0.7

⇨ Y_2=0.7=X2/60

⇨ X_2=42kg

Q9. The Modus Tolens rule for logic is "*given P →Q and ~Q, infer ~P*". Give the fuzzy Modus Tolens Rule. Post the answer in the A4 sheet provided. -8 marks

Ans:

Start with the goal that if t(Q)<=b, what is t(P) -2 marks

Use Lukasiewitz rule, t(P→Q)=min(1,1-t(P)+t(Q)) - 1 mark

Suppose t(P→Q)=c

c=min(1,1-t(P)+t(Q))

<u>Case-1:</u>

c=1

➔1-t(P)+t(Q)<=1

- ⇨ t(P)<=t(Q)
- ⇨ t(P)<=b -2 marks

Case-2:

c<1

➔1-t(P)+t(Q)=c

- ⇨ t(P)=1+t(Q)-c
- ⇨ t(P)<=1+b-c -2 marks

Combining case-1 and 2,

t(P)<=min(1,1+b-c) - 1 mark

***<u>Alternative Method</u>*:**

One can start from t(~Q→~P) and go up to t(~P)=max(0,c-b). But proof that t(~Q→~P) can be used will have to be given. Else 2+1+2=5.

Q10. Derive the mean and variance of chisquare distribution. Give a rigorous proof. Post the answer in the A4 sheet provided. -3+5=8 marks

Ans:

An r.v. X follows chisquare distribution with d.f. n, if

$X=Z^2_1+ Z^2_2+ Z^2_3+\ldots Z^2_n$

$E(X)=E(Z^2_1+ Z^2_2+ Z^2_3+\ldots Z^2_n)$

$=E(Z^2{}_1)+ E(Z^2{}_2)+ E(Z^2{}_3)+\ldots E(Z^2{}_n)$

Now, wlg, $E(Z^2{}_1)$=second moment of MGF of Z_1=2nd derivative of the MGF at (t=0)=(1+t)e^t^2=1

Hence **E(X)=n** **-** 3 marks

$V(X)= V(Z^2{}_1)+ V(Z^2{}_2)+ V(Z^2{}_3)+\ldots V(Z^2{}_n)$

Wlg, $V(Z^2{}_1)= E(Z^4{}_1)-[E(Z^2{}_1)]^2$

$E(Z^4{}_1)$=4th moment of MGF of Z1=4th derivative of the MGF at (t=0)=3

$V(Z^2{}_1)= E(Z^4{}_1)-[E(Z^2{}_1)]^2=3-1=2$

So, **V(X)=2n** **-**5 marks

If somebody has proved the result using MGF of chisquare, give half the marks, i.e., 4.

==================end=================