[C245 2014]

BEng and MEng EXAMINATIONS 2014

PART II : STATISTICS (COMPUTING C245)

Date : Friday 16th May 2014 10 - 12

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

Answer ALL questions

Statistical data sheets are provided

[Before starting, please make sure that the paper is complete; there should be a total of FOUR questions. Ask the invigilator for a replacement if your copy is faulty.]

© 2014 Imperial College London

1

[C245 2014]

1. Choose one answer for each part. Partial credit may be awarded for working if an incorrect

answer is selected. There is no negative marking.

(i) For the data below, calculate the sample median, x({n+1}/2); mean, ̄x = Pni=1 xi/n; interquartile range (IQR), x(3{n+1}/4) x({n+1}/4); and standard deviation, sn = pPni=1(xi ̄x)2/n. 0233479.

Which one of the following statements about these sample statistics is true? (a) The median is less than the mean and the IQR is less than the standard deviation; (b) The mean is less than the median and the IQR is less than the standard deviation; (c) The median is less than the mean and the standard deviation is less than the IQR; (d) The mean is less than the median and the standard deviation is less than the IQR; (e) The median is equal to the mean and the IQR is equal to the standard deviation. (ii) For two events A and B, we have P(A)=0.2, P(B)=0.3, and P(A|B)=0.8. What is

P(B|A) equal to?

(a) 0.3; (b) 0.4; (c) 0.5; (d) 0.6; (e) 0.7; (f) 0.8.

(iii) A die is rolled three times. What is the probability that we observe at least one 3?

(a) 0.5; (b) 0.58; (c) 0.42; (d) 0.33; (e) 0.67.

(iv) A bag contains 5 red balls and 1 black ball. Balls are randomly drawn from the bag until the black ball is found, and the number of draws which were required is noted. How much greater is the expected number of draws required if any red ball drawn from the bag is put back in the bag rather than set aside?

(a) 0; (b) 1; (c) 2; (d) 2.5; (e) 3.

(v) Suppose X1,X2,...,Xn are n independent random variables which all follow the same distribution PX which has mean μ and variance σ2. Let Sn = Pni=1 Xi be the sum of the n samples, and define Tn = Sn/pn. By the Central Limit Theorem, which of the following is an approximate distribution for Tn? (a) N(pnμ,σ2); (b) N(μ,nσ2); (c) N( μpn,σ2); (d) N(pnμ, σ2pn); (e) N(μ,σ2n ).

2

[C245 2014]

2. If X is a discrete random variable on the positive integers 11, 2,3,...l, it can be shown that

E(X) =

X1x=1

P(X x).

The Geometric distribution describes such a random variable. Let X ⇠ Geometric(p), so that X has probability mass function

p(x) = p(1 p)x 1, x = 1,2, 3,...

(i) Find P(X x) and hence show that E(X) = 1p. [Hint: Recall that the infinite sum of a

geometric progression with first term a and common ratio r is a/(1 r).] (ii) Find the number of rolls of a die we would expect to require until we finally observe a 6. (iii) Suppose that after five rolls of the die we have not yet observed a 6. If we decide to roll the die another five times, what is the probability that we will have observed a 6? (iv) If Ptwo X 1x=1 summations.]

is P(X a discrete x). [Hint: random Express variable P(X on the x) as positive a summation, integers, and prove then change the formula the order E(X) of the = 3. The number of emails sent to an IT helpdesk in a small department between the hours of 10 a.m. and 11 a.m. was recorded each day for a period of 100 working days. The data are summarised in the following table.

Number of emails 0 1 2 3 4 Frequency 14 39 29 11 7

Before collecting the data, it had been believed that the number of emails in an hour was well described by a Poisson distribution, and that two emails per hour were to be expected on average.

Recall that a Poisson random variable with mean parameter λ has probability mass function

p(x) = λxx! e λ

, x = 0,1, 2,....

(i) For a sample X1,...,Xn from the Poisson distribution, show that the sample mean is

the maximum likelihood estimator (MLE) for λ. (ii) Calculate the value of the MLE for λ for the data above. When calculating the sample

mean of the data, treat the category “ 4” as taking fixed value 4.5. (iii) Using the MLE for λ, perform a goodness-of-fit test at the 95% significance level to assess whether the number of emails per hour does indeed follow a Poisson distribution. (iv) Suppose that instead of estimating λ we assume 2 emails per hour are received on average. Repeat the Poisson goodness-of-fit test under this assumption, and comment on the comparison.

3

[C245 2014]

4. Let X1,...,Xn be a random sample of independent observations from the continuous uniform distribution U(0,b), where the parameter b is unknown. The probability density function for each Xi is, thus,

f(x) =

⇢ 0, 1b, 0  x  b otherwise.

(i) Find the cumulative distribution function, F(x) = P(Xi < x) for x 2 [0,b]. (ii) Let X(n) = maxi Xi be the largest value in the sample. Find the cumulative distribution function P(X(n) < x) for X(n). Hence derive the probability density function of X(n). [Hint: Notice that X(n) < x () 8i, Xi < x]. (iii) Calculate the expected value E(X(n)) as a function of b and n. (iv) Find the bias of the maximum sample value as an estimator for the upper limit of the population b. From this result, suggest a revised estimator of b using X(n) which is unbiased. Remember in constructing your estimator that b is unknown.

END OF PAPER

4

EXAMINATION QUESTIONS/SOLUTIONS 2013-2014 Course

**Comp245**

Question

1.

Marks & seen/unseen Parts

(i) x({n+1}/2)=3, ̄x=4, IQR=x(6) x(2)=5, sn =2p2. seen sim.

Answer is (c).

(ii) Using Bayes Theorem,

P(B|A) = P(A P(A) \ B)

= P(A|B)P(B)

P(A) = (1 P(A|B))(1 P(A)

P(B))

= (1 0.8)(1 0.2 0.3)

= 0.7.

Answer is (e). seen sim.

(iii) The probability of not observing a 3 is (5/6)3 ⇡ 0.58, so the probability of at least one

3 is 1 0.58 = 0.42. seen sim. Answer is (c).

(iv) Let X be the number of draws until the black ball is found. Without replacement, X is equally likely to be any number between 1 and 6, so E(X)=3.5. With replacement, X ⇠ Geometric(1/6), so E(X)=6. unseen Answer is (d).

(v) By distribution the Central of TLimit n = STheorem, n/ pn is then, an approximate approximately, distribution N(pnμ, for 2). Sn is N(nμ,n 2). The

unseen Answer is (a).

4 marks each

Setter’s initials Checker’s initials Page number

MM 1 of 4

EXAMINATION QUESTIONS/SOLUTIONS 2013-2014 Course

**Comp245**

Question

2.

Marks & seen/unseen Parts

X1X1(i) P(X x) =

p(k) = p (1 k=x k=xSetter’s initials Checker’s initials Page number p)k 1 = p ⇥ 1 (1 p)x 1

(1 p) = (1 p)x 1. seen

4 marks

X1We then have E(X) = x=1

P(X X1x) =

(1 x=1p)x 1 = 1 1

(1 p) = 1p. unseen

4 marks

(ii) If X is the number of rolls required, then X ⇠ Geom⇣16⌘, so E(X) = 116

= 6. seen sim. 3 marks

(iii) This is the same as the probability of failing to roll a 6 in five attempts; the first five unsuccessful rolls have no bearing on subsequent attempts (the Geometric is memoryless). The probability is P(X  10|X > 5) = 1 P(X ✓11|X > 5) = 1

1 16◆5

= 0.598.

seen sim.

4 marks X1(iv)

x=1

X1x=1

X1k=x

X1k=1

Xkx=1 X1P(X x) = p(k) =

p(k) =

kp(k) = E(X). unseen

k=1

5 marks

MM 2 of 4

EXAMINATION QUESTIONS/SOLUTIONS 2013-2014 Course

**Comp245**

Question

3.

Marks & seen/unseen Parts

(i)

L( ) =

Setter’s initials Checker’s initials Page number

YnPi=1 xie xi! =

Qni=1 ni=1 xie xi! n Xn=) `( ) = (

xi) log( ) n i=1

Ynlog(

xi!) i=1 =) d d`( ) =

Pni=1 xi n =) ˆ =

Pni=1 n xi

= ̄x.

This is a maximum since =) d d2

2`( ) =

Pni=1 xi

2

seen 7 marks is negative for all > 0.

(ii) ˆ = ̄x =1.615. 3 marks

(iii) Assuming a Poisson(1.61) distribution, we would have the following expected counts.

Number of emails (i) 0 1 2 3 4 Observed Frequency (Oi) 14 39 29 11 7 Expected Frequency (Ei) 19.89 32.12 25.94 13.96 8.09

These give a chi-square statistic of

X2 =

X4i=0

(Oi Ei)2

Ei = 4.35

which is less than 23,0.95=7.81. So insufficient evidence to reject the null hypothesis seen sim. that the email counts follow a Poisson distribution. 5 marks (iv) Now assuming a Poisson(2) distribution, we would have the following expected

counts.

Number of emails (i) 0 1 2 3 4 Expected Frequency (Ei) 13.53 27.07 27.07 18.04 14.29

Here,

X2 =

X4i=0

(Oi Ei)2

Ei = 11.88

which is greater than 24,0.95=9.49, so we reject the null hypothesis that the email seen sim. counts follow a Poisson(2) distribution. This suggests the mean number of emails per 5 marks hour is not close to 2.

MM 3 of 4

EXAMINATION QUESTIONS/SOLUTIONS 2013-2014 Course

**Comp245**

Question

4.

Marks & seen/unseen Parts

(i) Xi ⇠ U(0,b) has cumulative distribution function seen

FXi(x) =

8>>><>>>:

0, x  0

2 marks

1, xb, 0 < x < b x b.

(ii) By independence of the {Xi}, unseen

3 marks FX(n)(x) = P(X(n) < x) =

Setter’s initials Checker’s initials Page number

Yni=1

✓xb◆n P(Xi < x) =

, 0  x  b.

The density is the first derivative of this distribution function, giving seen

fX(n)(x) = dxdFX(n)(x) = dx

d✓xb4 marks

◆n = nb

✓xb◆n 1 , 0  x  b.

(iii)

seen 5 marks E(X(n)|b) =

Z 1 1 x fX(n)(x)dx = nZ b0

✓xb◆n dx = n + n

1

✓ n

n + 1◆b.

(iv) As an estimator of b, unseen

Bias(X(n)) = E(X(n)|b) b = b

n + 1. 3 marks Since E(X(n)|b) =

xn+1 bn

b= 0 ✓ n n + 1◆b, consider a revised estimator T =

✓n + 1 n

◆X(n). unseen

Then clearly E(T|b) = b, so T is now an unbiased estimator for b.

3 marks MM 4 of 4