NATIONAL UNIVERSITY OF SINGAPORE

ST2131/MA2216 PROBABILITY

(Semester 2 : AY 2016/2017)

Time Allowed : 2 Hours

INSTRUCTIONS TO STUDENTS

1. This assessment paper contains **FOUR (4)** questions and comprises **FOURTEEN (14)** printed pages.
2. Students are required to answer **ALL** questions. The total mark for this paper is 50.
3. You are allowed one double-sided, A4-sized formula sheet.
4. Non-programmable calculators may be used.
5. If you need extra space, you may use the blank pages on pages 13 and 14, but please label your answers clearly.
6. Please write your student number only in the space below. **Do not write your name.**

Student No: ______________________________

Total Marks (50)	

Question/ sub-question	Max. marks	Marks scored
1(a)	6	
1(b)	4	
2(a)	5	
2(b)	5	
3(a)	3	
3(b)	3	
3(c)	3	
3(d)	6	
4(a)	6	
4(b)	6	
4(c)	3	

1. (*10 points*) Let X have the pdf

$$f_X(x) = \frac{1}{2}(1+x), \quad -1 < x < 1$$

a. (*6 pts*) Find the pdf of $Y = X^2$.

b. (*4 pts*) Find $E(Y)$ and $\mathrm{var}(Y)$.

2. (*10 points*) Stirling's formula is a way to approximate factorials, for large n.

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

In this question, we shall prove Stirling's formula using probabilistic notions. Let Y be a Poisson random variable with parameter n.

a. (*5 pts*) Use the central limit theorem to show that, when n is large,

$$P(Y = n) \approx \Phi\left(\frac{1}{2\sqrt{n}}\right) - \Phi\left(\frac{-1}{2\sqrt{n}}\right) \tag{1}$$

b. (*5 pts*) Approximate the area represented by the right-hand-side in equation (1) on the previous page to derive Stirling's formula (*Remember that n is large!*). Feel free to use a sketch of the $N(0,1)$ pdf to illustrate your point.

3. (*15 points*) Let X_1 and X_2 have a bivariate normal distribution with pdf

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x_1^2 + x_2^2 - 2x_1x_2\rho)\right]$$

for $-\infty < x_1,\ x_2 < \infty$. Let $Y = \min(X_1, X_2)$, and $Z = \max(X_1, X_2)$.

a. (*3 pts*) Write down the following quantities: $E(X_1)$, $\mathrm{var}(X_1)$ and $\mathrm{cov}(X_1,\ X_2)$.

b. (*3 pts*) Let $Z = \max(X_1, X_2)$. Show that

$$E(Z) = E(X_1) + E(X_2) - E(Y)$$

c. (*3 pts*) Use Chebyshev's inequality to find an upper bound for $P(X_1 \geq 3)$.

d. (*6 pts*) By partitioning on whether $Y = X_1$ or $Y = X_2$, show that the pdf of Y is

$$f_Y(y) = 2\phi(y)\Phi\left(\frac{-y + \rho y}{\sqrt{1-\rho^2}}\right), \quad -\infty < y < \infty$$

where ϕ and Φ are the pdf and cdf of the $N(0, 1)$ distribution.

4. (*15 points*) The police from the district of Farmington have seized 496 packets of an unknown white powder. The police believe that all 496 packets contain cocaine. Due to resource contraints, however, they are only able to test 4 of these packets in a laboratory – all 4 tested positive for cocaine. In case you are wondering, the test is a conclusive one; it is never wrong.

If instead, contrary to what the police believe, there is a mixture of cocaine and non-cocaine packets within the 496, then we can presume a Hypergeometric distribution for the number of cocaine packets picked:

- $N = 496$ packets
- $m =$ the number of packets that are actually cocaine. This quantity is not random, but it is unknown.
- $n =$ the number of packets selected. In the paragraph above, $n = 4$.
- X is the random variable representing the number of cocaine packets selected.

The pmf of X is

$$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \quad x = 0, 1, \ldots, N$$

a. (*6 pts*) Here's yet another way to prove that a $H(n, N, m)$ distribution can be approximated by a $Bin(n, m/N)$ distribution. Work out

1. $\frac{P(X=x)}{P(X=x-1)}$ for $X \sim H(n, N, m)$.

2. $\frac{P(Y=x)}{P(Y=x-1)}$ for $Y \sim Bin(n, p)$.

Show that, if $m, N \to \infty$ such that $m/N = p$, then

$$\lim_{m,N\to\infty} \frac{P(X = x)}{P(X = x - 1)} = \frac{P(Y = x)}{P(Y = x - 1)}$$

Of course, we still have to show that $P(X = 0) \to P(Y = 0)$, but we'll skip that part for this exam. For the remaining parts of this question, approximate $H(n, N, m)$ with $Bin(n, m/N)$ whenever you need to.

b. (*6 pts*) The police decide to use the remaining packets to entrap drug offenders. They select 2 of the remaining untested packets at random, and sell them to a willing buyer. They then arrest the buyer at his home. Unfortunately, he had been tipped off and had disposed of the two packets!!

The lawyer for the buyer argues that it is possible that the 2 packets that he was sold did not contain cocaine, and that hence he was innocent. He argues that the probability that this was indeed the case is as high as 0.0223.

How did he arrive at that number? Maximise the probability that we first obtained 4 cocaine, and then 2 non-cocaine packets in the two consecutive experiments.

c. (*3 pts*) The police argue back! They challenge the lawyer to provide a probability p. They volunteer to test as many packets as necessary such that it reduces the maximal probability that the batch of packets is mixed (i.e. the probability in the previous question) *to below* p.

The lawyer picks $p = 0.001$. Use the table below to decide how many more packets the police need to test (with the belief that they will all test positive for cocaine).

Packets tested	**Maximal probability**
17	0.0012
18	0.0011
19	0.0010
20	0.0009
21	0.0009

– END OF QUESTIONS –

Extra space:

Extra space:

– END OF PAPER –