Probability and Statistics
MTH202 End-Sem Exam 27 April 2016

Write your name and/or registration number in the box provided.
Read *all* questions carefully before answering *any* question.
Calculators are *not* allowed. Make paper and pencil estimates.
Write the formulas *first* and then your calculations.
Justify your answers.
Write your answer in the space provided.
You have 3 hours to complete this exam.

Name:________________________ *Reg. No:*________________________

Question:	1	2	3	4	5	6	7	8	Total
Points:	6	8	6	6	6	6	8	4	50
Score:									

1. Suppose that out of 180 students who join an Institute in a particular batch, the following are the results after core years:

   - The number of students with poor ('D' or less) grades in Biology is 20. Similarly, it is 35 and 40 in Chemistry and Physics, respectively.
   - The number of students with poor grades in both Biology and Physics is 5. Similarly, it is 15 for Biology and Chemistry and 15 for Physics and Chemistry.
   - Finally, there are *no* students with poor grades in all three types of courses.

(2 marks) (a) Suppose that a student is chosen at random from this batch. What is the probability that, *given* that this student has a poor grade in Physics, this student *also* has a poor grade in Biology.

> **Solution:** We have $P(P = 0) = 40/180$ and $P(B = 0 = P) = 5/180$ so
>
> $$P(B = 0|P = 0) = P(B = 0 = P)/P(P = 0) = 5/40 = 1/8$$

(2 marks) (b) For a randomly chosen student, let B be the random variable that takes the value 0 if he/she has a poor grade in Biology and 1 otherwise. Similarly define C for Chemistry and P for Physics. Are there two of these random variables B, C and P which are independent?

Solution: We have

$$\begin{aligned} P(B=0=P) = 5/180 &\neq P(B=0)\cdot P(P=0) = (20\cdot 40)/180^2 \\ P(B=0=C) = 15/180 &\neq P(B=0)\cdot P(C=0) = (20\cdot 35)/180^2 \\ P(P=0=C) = 15/180 &\neq P(P=0)\cdot P(C=0) = (40\cdot 35)/180^2 \end{aligned}$$

Hence these are not pairwise independent.

(2 marks) (c) What is the probability that a randomly chosen student has no poor grade in any of these three types of courses?

Solution: We have $P(B=P=C=1) = 1 - P(B=0 \cup C=0 \cup P=0)$

$$\begin{aligned} P(B=0\cup C=0\cup P=0) &= P(B=0)+P(C=0)+P(P=0) \\ &- P(B=0=C) - P(B=0=P) - P(C=0=P) + P(B=C=P=0) \end{aligned}$$

So we have

$$P(B=C=P=1) = \frac{180-20-35-40+5+15+15}{180} = \frac{2}{3}$$

2. A donkey climbs a hill by jumping. The probability is 1/5, that the jump takes it 40 centimetres up and the probability is 4/5, that the jump takes it 10 centimeters down. Assume that all the jumps are independent of each other and of the position of the donkey. The donkey starts at the mid-way point (5000 centimetres) of a 10000 centimetre high hill,

(2 marks) (a) How many jumps must there be so that there is a non-zero probability that the donkey reaches the top of the hill.

Solution: As soon as $N \geq (10000-5000)/40$ there is a non-zero probability that the donkey has reached the top of the hill.

(2 marks) (b) Let X_N be the random variable giving the position of the donkey on the hill (in centimetres) after N jumps. What is $E(X_N)$?

Solution: Let J_i be the random variable indicating the distance travelled in the i-th jump. Then

$$E(J_i) = 40(1/5) - 10(4/5) = 0$$

Since $X_N = 5000 + \sum_i J_i$, we have $E(X_N) = 5000$.

(2 marks) (c) What is $\sigma^2(X_N)$?

Solution: Since J_i are independent, we have $\sigma^2(X_N) = \sum_i \sigma^2(J_i)$. We calculate

$$\sigma^2(J_i) = 40^2(1/5) + 10^2(4/5) = 400$$

Hence, $\sigma^2(X_N) = 400 \cdot N$.

(2 marks) (d) Write down an approximate formula for the probability that the donkey has moved at least a and at most b centimeters, after a large number N of jumps.

Solution: We use the Central Limit Theorem (or de Moivre Formula) to approximate by the normal distribution

$$P(A < \frac{X_N}{20\sqrt{N}} \leq B) = \int_A^B \frac{\exp(-s^2/2)}{\sqrt{2\pi}} ds$$

where we take $A = a/20\sqrt{N}$ and $B = b/20\sqrt{N}$.

3. Consider the functions $\phi_{a,b}(t)$ given by the formula:

$$\phi_{a,b}(t) = a^2 \cos(t) + b^2 \sin(t)/t$$

where a and b are some real numbers.

(2 marks) (a) What are the conditions on a and b so that $\phi_{a,b}$ is the characteristic function of a random variable?

Solution: Since $\cos(t)$ and $\sin(t)/t$ are both characteristic functions, the combination is a characteristic function if and only if $a^2 + b^2 = 1$. (We need to have a convex combination of characteeristic functions.)

(2 marks) (b) Write a formula (in terms of a and b) for the mean of the random variable for which $\phi_{a,b}$ is a characteristic function.

Solution: The mean is

$$(1/i)\frac{d\phi}{dt}|_{t=0} = 0$$

(2 marks) (c) Write a formula (in terms of a and b) for the variance of the random variable for which $\phi_{a,b}$ is a characteristic function.

Solution: The variance is

$$-\frac{d^2\phi}{dt^2}|_{t=0} = a^2 + b^2/3$$

4. A coin is flipped 10000 times and a Head shows up 5100 times.

(2 marks) (a) What is the maximum likelihood estimate for the probability of getting a Head on this coin.

Solution: The maximum likelihood estimate is given by the ratio of the number of heads to the total number of coin flips, i. e. 51/100.

(2 marks) (b) What is the log-likelihood ratio of this estimate against an unbiased coin?

Solution: The log likelihood ratio is

$$5100 \log(51/50) + (4900) \log(49/50) \simeq 100(51 \cdot (1/50) - 49 \cdot (1/50)) = 4$$

(2 marks) (c) Justify why you will (or will not) consider this to be a biased coin.

Solution: We can consider that this likelihood ratio is like that of $4/\log(2) \simeq 7$ successive heads on a unbiased coin. That is why it can be considered as an unbiased coin.

5. Based on theoretical calculations, a certain radioactive source should be emitting an alpha particle 0.20 times per second. To examine this, we bring a Geiger counter near this source (which clicks once for each emission).

(2 marks) (a) What is the probability distribution for the waiting time for hearing 8 clicks?

Solution: Let W be the random variable whose value is the amount of time before hearing 8 clicks. Then we have $P(W < 0) = 0$ and

$$P(W \leq t) = \int_0^t \frac{c^8 e^{-cs}}{8!} ds$$

where $c = 0.2$.

(2 marks) (b) The gaps between the click of the counter are (in seconds): 5.41, 1.21, 10.42, 4.31, 3.70, 0.57, 5.93, 0.45.

What is the maximum likelihood estimate for the frequency of emission of an alpha particle by this radiation source.

Solution: The maximum likelihood estimate for the frequency is

$$c' = \frac{\text{number of clicks heard}}{\text{total time}} = 8/32 = 0.25$$

since

$$5.41 + 1.21 + 10.42 + 4.31 + 3.70 + 0.57 + 5.93 + 0.45 = 32$$

(2 marks) (c) Justify why you will (or will not) accept the theoretical calculations.

Solution: The log likelihood ratio is

$$\begin{aligned} 8\log(c'/c) - 32(c' - c) &= 8\log(5/4) - 1.6 \\ &\simeq 8(1/4 - 1/32 + 1/192 - 1/5) \\ &= 8 \cdot \frac{240 - 30 + 5}{960} = \frac{43}{240} \simeq 0.18 \end{aligned}$$

This is very low and so the theoretical calculations seem to be quite good based on *this* experiment. (Actually, the sample size in the experiment is too small to really disprove the theory.)

6. Let X_i be a sequence of independent random variables with $P(X_i = 1) = 1/2$ and $P(X_i = -1) = 1/2$. Let $S_N = \sum_{i=1}^{N} X_i$.

(2 marks) (a) Calculate $E(S_N)$.

Solution: We have $E(X_i) = 0$ and hence $E(S_N) = 0$ as well.

(2 marks) (b) Calculate $\sigma^2(S_N)$.

Solution: We have $\sigma^2(X_i) = 1$ and since the variables are independent $\sigma^2(S_N) = N$ as well.

(2 marks) (c) Suppose that $k(p)$ is a function of p so that

$$\int_{-k(p)}^{k(p)} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} = p$$

What is the 80% confidence interval for S_N? (In terms of k.)

Solution: We need to take $p = 0.8$, hence the confidence interval is

$$\left[-k(0.8)\sqrt{N}, k(0.8)\sqrt{N}\right]$$

7. In a certain class of 200 students, 10 names are called at random (all names are equally likely) during each class. Assume that there are 20 classes.

(2 marks) (a) What is the probability that a certain student's name is never called?

Solution: The probability that the student's name is not called in a particular class is

$$p = \frac{\text{\# ways of picking 10 students out of 199}}{\text{\# ways of picking 10 students out of 200}}$$

$$= \frac{\binom{199}{10}}{\binom{200}{10}} = \frac{199!190!}{189!200!} = 19/20$$

The probability that this is repeated (independently) 20 times is p^{20}.

$$p^{20} = (1 - 1/20)^{20} \simeq e^{-1} \simeq 0.37$$

Hence, there is a 37% chance that a certain name will never be called.

Alternative Solution: We can take the names in each class "with replacement". (Since the numbers are large, this will not make much of a difference to the answer.) In this case, we get $p = (1 - 1/200)^{10}$. It follows that the answer we want is

$$p^{20} = \left(1 - \frac{1}{200}\right)^{200} \simeq e^{-1} \simeq 0.37$$

Alternative Solution: Let X be the random variable counting the number of classes when the student's name is called. We want $P(X = 0)$. Now X is a Binomial random variable with probability $10/200 = 1/20$ of being called in a given class. Thus

$$P(X = 0) = \binom{20}{0}(1/20)^0(19/20)^{20} = (19/20)^{20} \simeq e^{-1}$$

Alternative Solution: Use the Poisson distribution with $c = 20 \times 10/200 = 1$ as the expected number of times a certain student's name is called. The probability that the student's name is never called is $(c^0 e^{-c})/0! = e^{-1}$.

Alternative Solution: Consider the time to wait before a student's name is called. The name is called with frequency $c = 10/200 = 1/20$ per class. The probablity that the waiting time is more than 20 classes is

$$\int_{20}^{\infty} dt(ce^{-ct}) = e^{-(1/20)20} = e^{-1}$$

(2 marks) (b) What is the probability that there is at least one student whose name is never called?

Solution: In order that *all* names are called we *must* call different names each time. The probability of this (with replacement) is

$$\frac{\text{\# number of permutations of 200 students}}{\text{\# 200 pickings with replacement from 200}} = \frac{200!}{200^{200}}$$

If we take 10 names each time without replacement, the probability is

$$\frac{\text{\# number of permutations of 200 students}}{\text{\# 10 pickings without replacement from 200 done 10 times}} = \frac{200!}{(200 \cdot 199 \cdots 191)^{20}} \frac{200!}{(10!\binom{200}{10})^{20}}$$

The second fraction is bigger than the first one. So we estimate it. Using the Stirling approximation we see that this is approximately

$$\sqrt{2\pi \cdot 200}(200/e)^{200} \cdot \left(\frac{\sqrt{2\pi \cdot 190}(190/e)^{190}}{\sqrt{2\pi \cdot 200}(200/e)^{200}} \right)^{20} = \sqrt{2\pi \cdot 200} \cdot \left(\frac{19}{20} \right)^{3810} \simeq \sqrt{2\pi \cdot 200}e^{-195}$$

In the first approach we will get a similar answer with 195 replaced by 200.

This is quite small. Thus, the probability that there is at least one student whose name is never called is almost 1!

(2 marks) (c) What is the probability that a certain student's name is called in two successive classes?

Solution: The probability that a certain student's name is called in a given class is $q = 1 - p = 1/20$.

Now, if the student's name is not called in successive classes, then whenever the student's name is called in a class, it is not called in the preceding class or the following class.

Suppose the student's name is called in r classes. Then each such class *must* be followed by a class where the name is not called except when there is no following class. There are $\binom{20-r}{r}$ was of choosing r positions for such pairs. In addition, there are $\binom{19-r+1}{r-1}$ ways of choosing $r - 1$ pairs followed by one class where the name is called.

Thus, the probability of the students name *not* being called in successive classes is

$$Q = \sum_{r=0}^{10} \left(\binom{20-r}{r} + \binom{19-r+1}{r-1} \right) \frac{19^{20-r}}{20^{20}}$$

The probability of being called in successive classes is $1 - Q$.

Alternative Solution: The probability of being called in two chosen classes is $q^2 = 1/400$.

There are 19 pairs of successive classes. *Assume* that the calling of the student in each pair is independent of any other pair. This is not strictly true but the probability of a triple of successive classes is 1/8000 which is small so we ignore it! Under this assumption, the probability of being called in a pair of successive classes is $19q^2 = 0.0475$. (This is not bad since the above exact answer turns out to be 0.044!)

(2 marks) (d) Assuming that a student randomly chooses 10 classes to be absent. What is the probability that the student's name is called when he/she is absent?

Solution: First we calculate the probability of a student's name *not* being called when the student is absent.

The probability of a student being absent in a specific set of 10 classes is $(1/2)^{20}$. The probability of this student's name not being called in any particular class is p. Assuming that the choice of classes is independent of the choice of names for the roll call, the probability of the student's name not being called during a specific set of 10 classes and the student being abset in that set is $p^{10} \cdot (1/2)^{20}$. The number of choices of this set os $\binom{20}{10}$. Thus, the probability of a student's name not being called when she/he is absent for *some* set of 10 classes is

$$\binom{20}{10} p^{10} (1/2)^{20}$$

Now, as calculated above p^{10} is essentially $(1 - 1/20)^{10} \simeq e^{-1/2}$. The other term can be calculated using the Stirling approximation

$$(1/2)^{20} \binom{20}{10} \simeq \frac{20^{20+1/2}}{2^{20} 19^{2(10+1/2)} \sqrt{2\pi}} = \frac{1}{\sqrt{10\pi}}$$

We thus get the estimate $1/\sqrt{10e\pi}$ which is about 1/9. The probability that the students name *is* called during one of the classes she/he is absent is $1 - 1/9 = 8/9$.

8. Before a game of (singles) badminton between A and B, we have no information about the quality of each player, so we give each of them an equal chance of winning any particular point. At the end of the first game, the score is 21 to 15 in favour of A. (The

rules are simplified and the first person to reach 21 wins.)

(2 marks) (a) What is the probability that we assign to the chance of A winning a point over B *after* the first game?

> **Solution:** The maximum likelihood estimate for the probability of A winning a point over B after the first game is $21/(21+15) = 7/12$.

(2 marks) (b) What is the most probable score in the second game?

> **Solution:** The probability of B scoring r points before A scores 21 points is given by the negative binomial distribution
>
> $$\binom{21+r-1}{r}(7/12)^{21}(5/12)^r$$
>
> We can guess that this should be around $r = 15$. Since the shape of this has a single "bump", we check for $r = 14$ and $r = 13$ by Stirling approximation and find that the maximum is attained for $r = 14$.