DGB Sample Question paper

Disclaimer: The following model question paper is provided solely for illustrative purposes to demonstrate the structure and format of the actual examination. Please note that the difficulty level and topic-wise distribution of questions in the final exam may vary.

The question paper is divided into two groups: Biology Group and Non-Biology Group. Please choose ANY ONE group and answer the questions from that group only, according to the instructions.

Biology group

Answer Question 1 (15 marks) and any 5 from the other questions (5×15 = 75 marks).

1. a) Suppose a variable assumes the values *0, 1, 2, …, n* with equal frequencies 1 for all the values. Find the variance of the variable.

   b) The harmonic mean and geometric mean of two numbers are *a* and *b*, respectively. Prove that the arithmetic of the numbers is $\frac{b^2}{a}$.

   c) If P(*A*) = 0.6 and P(*B*) = 0.5, find the range of possible values for P($A \cap B$).

   d) A website receives an average of 20 hits per minute. What is the probability that exactly 18 hits will occur in a particular minute?

   e) What is the probability of a Type-I error at a 10% significance level for any statistical hypothesis test?

   [3+3+3+3+3]

2. Consider a recessive disorder controlled by the minor allele of an autosomal biallelic locus that conforms to Hardy-Weinberg Equilibrium. If 42% of the population are carriers, what is the percentage of unaffected individuals in the population? What is the chance that the offspring of a pair of carriers is unaffected?

   [10+5]

3. a) A linear DNA molecule is subjected to single and double digestions with restriction endonucleases, and the following results are obtained:

Enzymes	**Fragment Sizes (in kb)**
EcoRI	2.9, 4.5, 7.4, 8.0
HindIII	3.9, 6.0, 12.9
EcoRI and HindIII	1.0, 2.0, 2.9, 3.5, 6.0, 7.4

   Draw the restriction map defined by these data.

   b) Two couples want to know their risk of conceiving a child with cystic fibrosis. In one couple, neither partner has a family history of the disease; in the other, one partner knows he is a carrier. How do their risks differ?

   c) Inherited cancers like retinoblastoma show a dominant pattern of inheritance. However, the underlying genetic defect is a recessive loss-of-function mutation—often the result of a deletion. How can the dominant pattern of inheritance be reconciled with the recessive nature of the mutation?

   [6+5+4]

4. a) Describe one method used to identify DNA-binding proteins and the location of DNA control elements in regulatory regions of genes.

b) Compare the events involved in transcription initiation for prokaryotic and eukaryotic mRNA synthesis.

c) Explain the mechanism of drug resistance in bacteria.

[5+5+5]

5. a) Explain the following differences between prokaryotes and eukaryotes:
   i. Poly-cistronic mRNAs are observed in prokaryotes but not in eukaryotes
   ii. Transcription and translation are a simultaneous process in prokaryotes, but not in eukaryotes.

   b) Name three histone modifications that are believed to affect gene expression. Briefly describe their roles in altering the chromatin structure and gene expression.

   c) State a few reasons why DNA is a better material than RNA for storage of genetic information.

[3+3+2+4+3]

6. a) During each cell division, telomere is shortened. Suppose this process limits cell division by 50 generations. Explain how can tumor be formed in human in spite of the above phenomenon?

   b) Some miRNAs function as oncomir in one type of cancer, while the same miRNA can act as a tumor suppressor miRNA in another type of cancer. Explain how is this possible?

[7+8]

7. a) Japanese living in Japan has high incidence of stomach cancer compared to Japanese living in Hawaii, USA. Explain this observation with biological mechanisms.

   b) State a few reasons why DNA is a better material than RNA for storage of genetic information.

[8+7]

8. a) What is the difference between reverse transcriptase PCR (RT-PCR) and a standard PCR?

   b) You are cloning the genome of a new DNA virus into pUC18. You plate out your transformants on ampicillin plates containing X-gal and pick one blue colony and one white colony. When you check the size of the inserts in each plasmid (blue and white), you are surprised to find that the plasmid from the blue colony contains a very small insert of approximately 60 bp, while the plasmid from the white colony does not appear to contain any insert at all. Explain these results.

   c) Briefly explain Northern blot. Explain the purpose of performing a northern blot.

[4+6+5]

Non-Biology group

Answer Question 1 (15 marks) and any 5 from the other questions (5×15 = 75 marks).

1. a) Suppose a variable assumes the values 0, 1, 2, …, n with equal frequencies 1 for all the values. Find the variance of the variable.
   b) The harmonic mean and geometric mean of two numbers are *a* and *b*, respectively. Prove that the arithmetic of the numbers is $\frac{b^2}{a}$.
   c) If P(*A*) = 0.6 and P(*B*) = 0.5, find the range of possible values for P($A \cap B$).
   d) A website receives an average of 20 hits per minute. What is the probability that exactly 18 hits will occur in a particular minute?
   e) What is the probability of a Type-I error at a 10% significance level for any statistical hypothesis test?

   [3 + 3 + 3 + 3 + 3]

2. a) If X and Y are two random variables such that 2X-3Y and 2X+3Y are independent, find the ratio Var (X)/Var(Y)
   b) If X follows a Poisson distribution such that P(X=4) = P(X=5), find the variance of the passion distribution.
   c) If X_1 and X_2 are iid N(0,1) random variables, find the probability P ($X_1+X_2 \geq 2$)

   [5+5+5]

3. a) Suppose that there are five pairs of shoes in a closet and four shoes are taken out at a random. What is the probability that, among the four which are taken out, there is at least one complete pair?
   b) In a tennis match between two players A and B, A wins a set with probability 0.6. A best of five match stops as soon as a player wins 3 sets. Assume that the outcomes of the sets are independent.
      I. If the match lasts for 5 sets, what is the probability that A will win the match?
      II. If B wins the match, what is the probability that the match lasts for 5 sets?

   c)Let f: [0, ∞] → [0, ∞] be an increasing function. For all t>0, define

   $g(t) = \left(\int_0^t f(x)\,dx\right) / t$. Show that g is also an increasing function.

   [5+5+5]

4. a) The sun is about 25,000 light years away from the center of our galaxy milky way and is rotating around the center in an approximately circular path with a period of 170,000,000 years. The earth is 8 light minutes away from the sun and is rotating around the sun in an approximately circular path in one year. From this data alone, find out the approximate mass of our galaxy in units of sun's mass.
   b) Two particles of mass m and M respectively are at a distance d apart.

i. Calculate the moment of inertia of the system about an axis passing through the centre of mass and perpendicular to the line joining the two masses.

ii. If γ is the frequency of revolution, show that the rotational kinetic energy of the system is

$$2\pi^2\gamma^2 d^2 \frac{2mM}{m+M}$$

c) One kilogram of water at 273K is brought into contact with a heat reservoir at 373K. When the water has reached 373K, what is entropy change of

i. water,

ii. heat reservoir and

iii. the universe?

d) A Carnot cycle operates as a heat engine between two bodies of same mass and same specific heat C, until their temperatures are equal. If the initial temperatures of the bodies were T_1 and T_2 ($T_1 > T_2$) respectively,

i. show that the final temperature is $(T_1T_2)^{1/2}$

ii. work done by the engine is $W = C[T_1+T_2-2(T_1T_2)^{1/2}]$.

[4+(2+2) +3+(2+2)]

5. A gene is considered ON based on the following conditions: (i) Transcription factor A must be present, and (ii) At least one of transcription factors B or C must be present. Each transcription factor can be either present (1) or absent (0).

a) Write a Boolean expression, F (A, B, C) that represents the gene being ON.

b) Construct the complete truth table for the function F.

c) Express F in canonical Sum of Products (SOP) form by listing the corresponding minterms.

d) Use Karnaugh Map Minimization to obtain the minimized Boolean expression.

[3+4+3+5]

6. a) A sequence is a list of characters. For example: S = [A, B, C, A, B, A, C]. Write a pseudocode to count how many times the characters A or C appear in the sequence. [do not assume any build-in function to perform the said counting]

b) An object (for instance, a biological organism) starts with 1 unit. Every step, the number of units' doubles.

For example:

Step 0 → 1 unit

Step 1 → 2 units

Step 2 → 4 units

Write a recursive pseudocode to compute the number of units after n steps.

[7+8]

7. In a simple biological experiment, several samples are collected from different conditions. Each sample is assigned a numeric score representing its activity level (higher score means higher activity)

   Example data:

Sample IDs:	S1	S2	S3	S4	S5	…	S50
Scores:	23	5	17	9	30	…	12

   a) Write a high-level pseudocode to obtain the top 10 samples. State the average time complexity of your algorithm.

   b) Now suppose the experiment generates one million samples. Would your approach in parts till work efficiently? If not, write pseudocode for an efficient algorithm suitable in this case and state its average time complexity. If yes, justify.

   [8+7]

8. A long sequence of symbols is used to represent biological information. Each symbol is a single character, and the entire sequence can be treated as a simple array of characters.

   Example sequence: S = [A, B, A, C, B, A, D, C, A, B]

   Suppose we want to check whether a specific pattern appears in the sequence. A pattern is a short sequence of characters. Example pattern: P = [A, C, B]

   a) Write a pseudocode to check whether the pattern P occurs in S.

   b) If the length of S is n and the length of P is m, what is the worst-case time complexity of this method? Justify your answer.

   [8+7]

9. a) You are given a list of numbers representing scores: scores = [10, 20, 30, 40, 50]. Write a pseudocode that counts how many scores are greater than the average.

   b) Consider the following pseudocode:

```
function f(n):
    if n == 0:
        return 1
    else:
        return 2 * f(n - 1)
```

   i. Briefly explain what the function f(n) computes.

   ii. What is the value returned by f(4)? Show the intermediate steps.

   iii. How many times is the function f called (including the initial call) when f(4) is evaluated?

   [5+(3+3+4)]

10. a) If $2x+3y+4z=1$, find the maximum possible value of xyz (here x, y and z are three positive real numbers)

    b) If the arithmetic mean of some observation is 8, and all these observations lie on the interval [0,10], show that the standard deviation of the observation cannot exceed 4.

    c) If $u_{\sim}$ is a Di dimensional unit vector (i.e. $||u||=1$) and I is the dxd identity matrix, show that I-uú is a positive semi-definite matrix. [5+5+5]