GROUP A

1. Consider a sequence $\{r_k\}_{k\geq 1}$ of rational numbers lying in the interval $(0, 1)$. Further, assume that r_k converges to an irrational number as $k \to \infty$. Suppose $r_k = \frac{m_k}{n_k}$, for all $k \geq 1$, where m_k and n_k are positive integers with no common divisors. Show that the set of integers $\{n_k : k \geq 1\}$ is not bounded.

2. Consider a 4×4 real matrix A which has positive trace and negative determinant.

   (a) Show that A must have at least two real eigenvalues.

   (b) Show that A can have non-real eigenvalues.

3. Let P be a regular polygon with 24 sides. Consider all the triangles whose vertices are also vertices of P. Find the number of such triangles that are neither isosceles nor equilateral.

GROUP B

4. Suppose U and V are independent and identically distributed random variables following a binomial distribution with parameters n and $\frac{1}{2}$. Let the random variable T denote the number of distinct real roots of the quadratic equation

$$x^2 + 2Ux + V^2 = 0.$$

   Find $E(T)$ and $Var(T)$.

5. Suppose $r \geq 1$ distinct books are distributed at random among $n \geq 3$ children.

(a) For each $j \in \{0, 1, 2, \ldots, r\}$, compute the probability that the first child gets exactly j books.

(b) Let X be the number of children who do not get any book, and Y be the number of children who get exactly one book. Show that

$$\mathrm{Cov}(X, Y) = \frac{r(n-1)(n-2)^{r-1}}{n^{r-1}} - \frac{r(n-1)^{2r-1}}{n^{2r-2}}.$$

6. Consider two random variables (X, Y) distributed as bivariate normal with parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Based on a random sample from this bivariate distribution, the fitted least squares regression line of Y on X and that of X on Y were as follows:

$$Y = 22 - 3X$$

$$X = 5.84 - 0.12Y$$

(a) Compute the maximum likelihood estimates of the following parametric functions:

(i) $\min\{\mu_1, \mu_2\}$ (ii) $\frac{\sigma_1}{\sigma_2}$ (iii) ρ

You are not required to derive the expressions for the maximum likelihood estimators of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ.

(b) If a new observation (5, 5) is included in the above set of observations, determine whether each of the maximum likelihood estimates obtained in (a) will increase, decrease or remain unchanged.

7. Suppose N students arriving at a college are all equally likely to have a particular disease with an unknown probability p. The disease status (affected / not affected) of all students are independent. Blood samples are collected from all N students. In order to estimate p, two strategies are proposed.

Strategy 1 Test all samples separately to obtain the status for all N students.

Strategy 2 Randomly partition the students into m disjoint groups, each comprising $K = N/m$ students (with $K \geq 2$ being an integer). For each group, pool (mix) the blood samples from all K students within the group and test the pooled sample. If the pooled sample tests positive, then at least one student within that group is affected. If the pooled sample tests negative, all students within that group are unaffected.

(a) Based on the data obtained from Strategy 2, find a real-valued sufficient statistic for p and the maximum likelihood estimator of p.

(b) If a group tests positive, then all students within that group are further tested individually. Suppose that each test (for individual sample or pooled sample) has equal cost. Then, which of the two strategies would you prefer to identify all students affected with the disease when the underlying $p = 0.5$, $N = 200$, and $m = 20$?

8. Based on historical data, a positive random variable X arising from an unknown distribution is believed to have the chi-square distribution with 1 degree of freedom. A new theory suggests that it may be better to model $\sqrt{X}$ as exponentially distributed with mean λ, where λ is such that $\mathrm{E}(X)$ for this model is the same as that for the earlier model.

(a) Compute λ.

(b) Suppose $X_1, X_2, \ldots, X_n$ are independent observations from this unknown distribution. For $\alpha \in (0,1)$, consider the most powerful level α test for the null hypothesis that the earlier model is correct against the alternative that the new model is correct. Show that the rejection region of this test is

$$\left\{(x_1, x_2, \ldots, x_n) : \sum_i x_i - 2\sqrt{2}\sum_i \sqrt{x_i} > c\right\}$$

where c is a constant that depends on α.

9. The following ANOVA table for three factors A, B, and C was obtained (under a suitable model) from some data, but several values were illegible and marked '∗'. It is known that A had two levels and there were 24 observations in all.

Source	df	Sum of squares	Mean Square	F ratio
A	∗	∗	∗	∗
B	∗	∗	∗	0.62
C	∗	∗	25.4	∗
AB	2	∗	5.4	0.54
BC	2	∗	4.2	∗
Error	∗	∗	∗	
Total	∗	257.4		

(a) Fill in the missing values, providing suitable justification.

(b) Suppose that a student has access only to tables for t distributions. Can she test the hypothesis of equality of effects of levels of A and that of equality of effects of levels of B, both at 5% level of significance? Justify your answer.