1. (a) State the weak law of large number (WLLN) and the ordinary central limit theorem (CLT). [4 marks]

   (b) State the Chebychev inequality and use it to prove a WLLN. [4 marks]

   (c) Let $X_1, \ldots, X_n$ be a random sample of size n from the uniform distribution on $(0, 1)$ with probability density function

$$f(x) = I(0 < x < 1).$$

      i. Compute the mean and variance of this distribution. [2 marks]

      ii. Compute $P[X_1 \leq 0.25]$. [2 marks]

      iii. Let $\bar{X} = (\sum_{i=1}^{n} X_i)/n$. For $n = 100$. [3 marks]

      iv. Let $X_{(1)}$ denote the minimum of $\{X_1, X_2, \ldots, X_n\}$. For $n = 100$, compute $P[X_{(1)} \leq 0.25]$. [5 marks]

2. Let X and Y be two random variables.

   (a) Under appropriate conditions (to be stated by you), prove that

$$\text{Variance}\{X\} = \text{Variance}\{E[X|Y]\} + E\{\text{Variance}[X|Y]\}.$$

   You may use without proof the result that $E\{E[X|Y]\} = E(X)$. [6 marks]

   (b) State and prove the Rao-Blackwell theorem. Explain how can it be used to improve upon (reduce) the variance of an unbiased estimator. [8 marks]

   (c) The total number of misprints per page in a book can often be modelled as a Poisson random variable X with parameter λ and the probability mass function (pmf) given by

$$f(x, \lambda) = e^{-\lambda}\lambda^x/x!, \quad x = 0, 1, 2, \ldots.$$

   Suppose that the publisher is interested in estimating the probability of only one misprint per page.

      i. Obtain an unbiased estimator of $P[X = 1]$ based on X_1 only. [3 marks]
      [Hint: Use indicator function.]

      ii. Use the Rao-Blackwell theorem with sufficient statistic $\sum_{i=1}^{n} X_i$ to obtain an improved estimator of $P[X = 1]$. You don't need to show the explicit expression of the estimator in terms of $\sum_{i=1}^{n} X_i$. [3 marks]

3. Suppose $X_1, \ldots, X_n$ is a random sample of size n from the Poisson distribution with parameter $\lambda > 0$ and the probability mass function (pmf) given by

$$f(x, \lambda) = e^{-\lambda}\lambda^x/x!, \quad x = 0, 1, 2, \ldots.$$

   (a) Prove directly using the definition of sufficiency that $\sum_{i=1}^n X_i$ is a sufficient statistic for λ. Here you are not allowed to use the factorisation theorem. However, you can use the fact that the sum of independent Poisson random variables is Poisson with parameters added. [10 marks]

   (b) Write down the expressions for the mean and variance of the sample mean $\bar{X}$ in terms of λ. [2 marks]

   (c) Prove directly using the definition of completeness that $\sum_{i=1}^n X_i$ is a complete sufficient statistic for λ. Hence obtain the uniformly minimum variance unbiased estimator (UMVUE) of λ. [8 marks]

4. (a) State the Cramer-Rao lower bound (CRLB) for the variance of an unbiased estimator based on a random sample of size n. You must explain your notations clearly. [5 marks]

   (b) Let θ be the proportion of the population who support a certain political issue, $0 < \theta < 1$. Let $X_1, \ldots, X_n$ denote a random sample of size n from the population where $X_i = 1$ if the i-th person supports the issue and $X_i = 0$ if the i-th person does not support the issue. Note that the probability mass function (pmf) of each X_i is given by

$$f(x, \theta) = \theta^x (1 - \theta)^{1-x} I(x = 0, 1).$$

   Compute the mean and variance of each X_i. [3 marks]

   (c) Suppose we are interested in estimating the proportion θ based on $X_1, \ldots, X_n$.

      i. Derive the CRLB (Cramer-Rao lower bound) for estimating $\tau(\theta) = \theta$. [6 marks]

      ii. Show that $\bar{X}$ is the uniformly minimum variance unbiased estimator (UMVUE) of θ. [2 marks]

      iii. What is the UMVUE of $\tau(\theta) = \theta/2$? [2 marks]

      iv. Suppose that $n = 5$ and the first and the fifth person support the issue while others do not. Compute the UMVU estimate of θ and $\tau(\theta) = \theta/2$. [2 marks]

5. Let $X_1, \ldots, X_n$ denote a random sample of size n from the uniform distribution on $(0, \theta)$ with the following probability density function (pdf)

$$f(x, \theta) = \theta^{-1} I(0 < x \le \theta),$$

where $\theta > 0$ is the unknown parameter.

(a) Sketch the likelihood function and find the maximum likelihood estimator (MLE) of θ. [5 marks]

(b) i. Find the cumulative distribution function (cdf) of X_1. [2 marks]

ii. Derive an expression for the cdf of the MLE. [3 marks]

iii. Hence find the probability density function (pdf) of the MLE. [2 marks]

iv. Find the mean of the MLE. Is the MLE unbiased for θ? If not, construct an unbiased estimator of θ based on the MLE. [3 marks]

(c) Show that the MLE converges in probability to θ. [5 marks]

6. The lifetime, X, of a certain electronic component is modelled by the exponential distribution with the probability density function (pdf)

$$f(x) = \frac{1}{\theta} e^{-x/\theta} I(x > 0),$$

where $\theta > 0$ is the unknown parameter. Let $X_1, \cdots, X_n$ be independent and identically distributed with the above pdf. Supose we want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, where θ_0 and θ_1 are known and $\theta_0 < \theta_1$.

(a) Use the Neyman-Pearson Lemma to show that the critical region of the most powerful test is of the form $\{\sum_{i=1}^{n} x_i \geq k\}$, where k is a constant. [4 marks]

(b) It can be shown that $Y = 2X/\theta$ has a chi-squared distribution with 2 degrees of freedom (d.f.) where X has the above exponential distribution. Using 'sum of independent chi-squares is a chi-squared random variable with d.f.'s added', show that the most powerful size α test has critical region

$$\{\sum_{i=1}^{n} x_i \geq k\}, \text{ where } P[\chi^2_{2n} \geq 2k/\theta_0] = \alpha.$$

[3 marks]

(c) In a particular study, suppose that $n = 20$ and we are interested in testing $H_0 : \theta_0 = 3$ against $H_1 : \theta_1 = 5$ at 5% level. Find the value of k. Here and elsewhere you may use the fact that for large d.f. d, $\chi^2_d \approx N[d, 2d]$. [4 marks]

(d) If the observed value is $\sum_{i=1}^{n} x_i = 75$, what is your conclusion? [1 mark]

(e) Compute the approximate power of your test at $\theta_1 = 5$. [4 marks]

(f) Give an expression for a two-sided $100(1 - \alpha)\%$ confidence interval for θ in terms of the percentiles of the χ^2_{40} distribution. [3 marks]

(g) Compute a two-sided 95% confidence interval for θ based on the observed value of $\sum_{i=1}^{n} x_i = 75$. Here you may use $\chi^2_{40,0.025} = 59.34$ and $\chi^2_{40,0.975} = 24.43$, where $\chi^2_{d,p}$ is defined as $P[\chi^2_d > \chi^2_{d,p}] = p$. [1 mark]

7. Suppose that the service-time, X, for a customer in a bank (in minutes) is an exponential random variable with parameter $\theta > 0$ and with the following probability density function (pdf)

$$f(x|\theta) = \theta e^{-\theta x} I(x > 0).$$

Suppose $x_1, \ldots, x_n$ are the observed service times of randomly selected n customers from the above bank. Consider the gamma prior distribution (with parameters $\alpha > 0$ and $\beta > 0$) for θ with the following pdf:

$$f(\theta) = \frac{\beta^{\alpha} \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)} I(\theta > 0).$$

Recall that the mean and variance of the gamma distribution are given by $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$.

(a) Obtain the posterior distribution of θ. [9 marks]

(b) Find the Bayes estimator of θ as the posterior mean. [3 marks]

(c) Suppose that the prior mean and the prior variance of θ are 0.2 and 1 respectively. A sample of $n = 20$ customers has mean 3.8 minutes. What will be the posterior distribution and the Bayes estimators of θ based on these data? [8 marks]