21 Q4

(i) The importance of sufficiency and completeness in parameter estimation.

Sufficiency:
A statistic $T(X)$ is said to be sufficient for parameter θ if the conditional distribution of the data X given $T(X)$ does not depend on θ.

Importance: Sufficient statistics extract all the information in the sample about the parameter. Estimation can therefore be based on $T(X)$ without loss of information.

Example: For i.i.d. $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(\theta)$, the sample sum $T = \sum_{i=1}^{n} X_i$ is sufficient for θ.

Completeness:
A sufficient statistic $T(X)$ is complete if any measurable function $g(T)$ with $\mathbb{E}[g(T)] = 0$ for all θ implies $g(T) = 0$ almost surely.

Importance: Completeness ensures uniqueness of unbiased estimators. By the Lehmann–Scheffé theorem, if $T(X)$ is sufficient and complete, then the conditional expectation of any unbiased estimator given $T(X)$ is the unique minimum variance unbiased estimator (MVUE).

Example: In the Bernoulli case above, the sum T is both sufficient and complete for θ
.

(ii) The Conditionality and Likelihood Principles of statistical inference.

Conditionality Principle:
If an experiment is chosen at random from several possible experiments, inference should be based on the experiment actually performed.

Example: If a parameter is estimated from either a Binomial(n, θ) or Negative Binomial(θ) experiment chosen at random, the inference should condition on the experiment actually observed.

Likelihood Principle:
All the evidence about the parameter θ contained in the data is summarized by the likelihood function $L(\theta \mid x)$. Inference should depend only on the likelihood, not on the sampling plan.

Example: If $x = 3$ successes are observed, both Binomial$(n = 10, \theta)$ and Negative Binomial$(r = 3, \theta)$ experiments yield likelihood proportional to $\theta^3(1-\theta)^7$. The Likelihood Principle asserts that inference should be the same, regardless of design.

(iii) Bayesian approaches to hypothesis testing.

In Bayesian hypothesis testing, hypotheses are treated as events with prior probabilities.
For hypotheses H_0 and H_1:

$$P(H_i \mid x) = \frac{P(H_i)L_i(x)}{\sum_j P(H_j)L_j(x)},$$

where $L_i(x)$ is the marginal likelihood under H_i.

Bayes factor:
The ratio of marginal likelihoods

$$B_{01} = \frac{P(x \mid H_0)}{P(x \mid H_1)}$$

quantifies the evidence in favor of H_0 over H_1.

Example: Testing $H_0 : \theta = 0.5$ vs. $H_1 : \theta \neq 0.5$ for Binomial data. Priors are assigned to each hypothesis, and the Bayes factor updates beliefs about which hypothesis is more plausible after observing the data.

(iv) Large sample tests based on maximum likelihood estimators.

Under regularity conditions, the maximum likelihood estimator (MLE) $\hat{\theta}$ has the following properties:

Consistency: $\hat{\theta} \to \theta$.

Asymptotic normality:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(\theta)^{-1}),$$

where $I(\theta)$ is the Fisher information.

These properties allow large-sample hypothesis tests:

Wald test:

$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}} \sim N(0, 1).$$

Likelihood ratio test (LRT):

$$\Lambda = -2 \log \frac{L(\theta_0)}{L(\hat{\theta})} \xrightarrow{d} \chi^2_k,$$

where k is the number of restrictions.

Score test:
Uses the derivative of the log-likelihood evaluated at θ_0.

Example: In logistic regression, large-sample Wald and LRT tests are commonly used to assess significance of predictors.

22 Q4

(i) Admissibility and the James--Stein estimator.

Admissibility:
An estimator $\delta(X)$ of a parameter θ is said to be \emph{admissible} if there does not exist another estimator $\delta'(X)$ such that

$$R(\theta, \delta') \le R(\theta, \delta) \quad \text{for all } \theta,$$

with strict inequality for some θ, where $R(\theta, \delta)$ is the risk function.
If such a δ' exists, then δ is \emph{inadmissible}.

James--Stein Estimator:
Suppose $X \sim N_p(\theta, I_p)$, where $p \ge 3$. The usual estimator of θ is the sample mean vector $\hat{\theta} = X$, which is unbiased and has risk $R(\theta, \hat{\theta}) = p$.
The James--Stein estimator is given by

$$\hat{\theta}_{JS} = \left(1 - \frac{(p-2)}{\|X\|^2}\right) X.$$

It dominates the usual estimator in terms of mean squared error risk for $p \ge 3$, showing that the MLE is inadmissible in this setting.
This is a famous result in decision theory and illustrates the importance of shrinkage estimation.

(ii) Criticisms of the Conditionality Principle.

The Conditionality Principle states that if an experiment is chosen at random from several possible experiments, inference should be based only on the experiment actually performed.
Criticisms include:

- It can conflict with the frequentist emphasis on the sampling distribution, which depends on the design.
- In some complex designs, conditioning may discard relevant information.
- It is not universally accepted when combined with the Sufficiency Principle, since together they imply the Likelihood Principle, which many frequentists reject.

(iii) Difficulties with Bayes factors.

Bayes factors are widely used in Bayesian hypothesis testing, but they suffer from several difficulties:

1. Dependence on priors: The marginal likelihood $P(x \mid H)$ requires a prior under each hypothesis. The Bayes factor can be highly sensitive to the choice of priors, especially for parameters under H_1.
2. Improper priors: If noninformative or improper priors are used, the marginal likelihood may not be well-defined, making the Bayes factor unusable.
3. Interpretation: While Bayes factors provide evidence in favor of one hypothesis over another, their interpretation in practice can be subjective, relying on arbitrary thresholds (e.g., Jeffreys' scale).
4. Computational complexity: For high-dimensional models, evaluating marginal likelihoods can be very difficult.

(iv) Inference based on large-sample properties of maximum likelihood estimators.

Under regularity conditions, the maximum likelihood estimator (MLE) $\hat{\theta}$ has the following properties:

Consistency: $\hat{\theta} \to \theta$.

Asymptotic normality:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(\theta)^{-1}),$$

where $I(\theta)$ is the Fisher information.

These properties allow construction of large-sample confidence intervals and hypothesis tests such as the Wald test, likelihood ratio test, and score test.

23 Q4

(ii) Optimal point estimation in the Inverse Gaussian distribution.

The Inverse Gaussian distribution with parameters $\mu > 0$ and $\lambda > 0$ has density

$$f(x \mid \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$$

For a random sample $X_1, \ldots, X_n \sim IG(\mu, \lambda)$, the log-likelihood function is

$$\ell(\mu, \lambda) = \frac{n}{2}\log\lambda - \frac{3n}{2}\log(2\pi) - \frac{3}{2}\sum_{i=1}^{n}\log X_i - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{X_i}.$$

Maximizing with respect to μ and λ, the maximum likelihood estimators are

$$\hat{\mu} = \bar{X}, \qquad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\bar{X}^2 X_i}}.$$

Thus, the optimal point estimator (in the sense of maximum likelihood) for μ is the sample mean, and for λ is the reciprocal of a weighted sample variance term.

(iii) Principles of decision theory.

Decision theory provides a framework for statistical inference by formalizing the process of making choices under uncertainty. Its main components are:

1. Action space: The set of all possible decisions (e.g., choosing an estimator).
2. Parameter space: The set of possible values of the unknown parameter θ.
3. Loss function: $L(\theta, a)$ measures the cost of taking action a when the true parameter is θ.
4. Risk function: The expected loss,

$$R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))],$$

   where $\delta(X)$ is a decision rule (estimator).

Principles include:

- Minimax principle: Choose the decision rule that minimizes the maximum risk.
- Admissibility: An estimator is admissible if no other estimator dominates it uniformly in risk.
- Bayes principle: Choose the rule that minimizes the posterior expected loss with respect to a prior distribution on θ.

These principles guide the evaluation and selection of statistical procedures.

24 Q4

(ii) Optimal point estimation of the parameters of an exponential family distribution.

A one-parameter exponential family has density

$$f(x \mid \theta) = h(x) \exp\{\eta(\theta) T(x) - A(\theta)\},$$

where $T(x)$ is the sufficient statistic.

For a random sample $X_1, \ldots, X_n$, the joint density is

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^n h(x_i) \exp\left\{\eta(\theta) \sum_{i=1}^n T(x_i) - nA(\theta)\right\}.$$

The log-likelihood is

$$\ell(\theta) = \eta(\theta) \sum_{i=1}^n T(x_i) - nA(\theta) + \sum_{i=1}^n \log h(x_i).$$

Maximizing gives the likelihood equation

$$\frac{\partial \ell}{\partial \theta} = \left(\frac{d\eta}{d\theta}\right) \sum_{i=1}^{n} T(x_i) - nA'(\theta) = 0.$$

Since $A'(\theta) = \mathbb{E}_\theta[T(X)]$, the MLE $\hat{\theta}$ satisfies

$$\frac{1}{n} \sum_{i=1}^{n} T(x_i) = \mathbb{E}_{\hat{\theta}}[T(X)].$$

Thus the optimal point estimator equates the observed sufficient statistic with its expectation under the model.

Example: For the exponential distribution with mean $1/\lambda$, the sufficient statistic is $\sum X_i$, and the MLE is $\hat{\lambda} = n / \sum X_i$.

(iv) Asymptotic properties of maximum likelihood estimators.

Under regularity conditions, the MLE $\hat{\theta}$ has the following asymptotic properties:

- Consistency: $\hat{\theta} \to \theta$ as $n \to \infty$.
- Asymptotic normality:

  $$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(\theta)^{-1}),$$

  where $I(\theta)$ is the Fisher information.
- Efficiency: The MLE achieves the Cramér–Rao lower bound asymptotically.

These properties make the MLE a central tool in large-sample inference.

```
In [ ]:
```

21 Q4