Chapter 9 Hypothesis Testing

9.1 Introduction to Hypothesis Testing

- **Problem:** Given a realization $x_1, \cdots, x_n$ of a random sample $X_1, \cdots, X_n$ from some population distribution $f(x; \theta)$, we want to know whether the true parameter θ belongs to some specific subset Θ_0 of the parameter space Θ.

- **Definition: Hypothesis.** A *hypothesis* is a statement about the population parameter. The two complementary hypotheses in a hypothesis testing problem are called the *null hypothesis* and the *alternative hypothesis*. They are denoted by H_0 and H_1, respectively.

- **Remark:** The general formats of the null and alternative hypotheses are $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^c$, respectively.

9.1 Introduction to Hypothesis Testing

- **Example:** If θ denotes the average change in a patient's blood pressure after taking a drug. We might be interested in testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. The null hypothesis states that, on the average, the drug has no effect on blood pressure. The alternative hypothesis states that there is some effect.

- **Example:** If θ denotes the proportion of defective items. We might be interested in testing $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$, where θ_0 is the maximum acceptable proportion of defective items. The null hypothesis states that the proportion of defective items is unacceptably high.

9.1 Introduction to Hypothesis Testing

- **Definition: Simple Hypothesis Versus Composite Hypothesis.** A hypothesis is simple if and only if it contains exactly one population. If the hypothesis contains more than one population, it is called a composite hypothesis.

- **Definition: Hypothesis Testing.** A hypothesis testing procedure or a hypothesis test is a decide rule that specifies
  - for what sample values $\boldsymbol{x}^n$ the decision is made to accept $\mathbb{H}_0$ as true, and
  - for what sample values $\boldsymbol{x}^n$ $\mathbb{H}_0$ is rejected and $\mathbb{H}_A$ is accepted as true.

9.1 Introduction to Hypothesis Testing

- **Definition: Critical Region or Rejection Region.** The set R of the sample points $\boldsymbol{X}^n$ for which $\mathbb{H}_0$ will be rejected is called the *rejection region* or *critical region*. The complement of the rejection region is called the *acceptance region*.

- **Remarks:**
  - The critical region C does not depend on the parameter θ.
  - Critical region and acceptance region forms a partition of all possible values of $\boldsymbol{x}_n$.
  - A test is determined by its critical region.

9.1 Introduction to Hypothesis Testing

- **Remarks:**

  – If $\theta \in \Theta_0$ (H_0 holds) and the observation $(x_1, \cdots, x_n)$ falls into the critical region C, then a *type I error* is made. The probability of making type I error is

  $$\alpha(\theta) = P_\theta[(X_1, \cdots, X_n) \in C], \quad \theta \in \Theta_0.$$

  – If $\theta \in \Theta_0^c$ (H_1 holds) and the observation $(x_1, \cdots, x_n)$ is in the acceptance region, then a *type II error* is made. The probability of making type II error is

  $$\beta(\theta) = 1 - P_\theta[(X_1, \cdots, X_n) \in C], \quad \theta \in \Theta_0^c.$$

  – If the critical region C decreases, the probability of making type I error decreases, but the probability of making type II error increases. Similarly, if C increases, $\beta(\theta)$ decreases, but $\alpha(\theta)$ increases.

9.1 Introduction to Hypothesis Testing

- **Definition: Power of Tests.** If R is the critical region of a test of the null hypothesis $H_0 : \theta \in \Theta_0$, the function $\pi(\theta) = P_\theta(\boldsymbol{X}^n \in R)$ is called the *power* of test with the rejection region R.

- **Definition: Type I and Type II Errors.** If $H_0 : \theta \in \Theta_0$ holds and the observation $\boldsymbol{x}^n$ falls into the critical region R, then a Type I error is made. The probability of making Type I error is

$$\alpha(\theta) \equiv P_\theta(\boldsymbol{X}^n \in R|H_0).$$

If $H_a : \theta \in \Theta_0^c$ holds and the observation $\boldsymbol{x}^n$ is in the acceptance region, then a Type II error is made. The probability of making Type II error is

$$\beta(\theta) \equiv 1 - P_\theta(\boldsymbol{X}^n \in R|H_A).$$

9.1 Introduction to Hypothesis Testing

- **Definition: Uniformly Most Powerful Tests.** Let $\mathbb{C}$ be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_a : \theta \in \Theta_A$. A test $T(\boldsymbol{X}^n)$ in class $\mathbb{C}$, with power function $\pi(\theta)$, is a uniformly most powerful test over $\mathbb{C}$ if $\pi(\theta) \geq \tilde{\pi}(\theta)$ for all $\theta \in \Theta_A$ and all $\tilde{\pi}(\theta)$ that is a power function of any other test $G(\boldsymbol{X}^n)$ in class $\mathbb{C}$.

- **Definition: Size of Tests.** The *size* of test C is defined as

$$\overline{\alpha}_C = \sup_{\theta \in \Theta_0} \pi_C(\theta).$$

For $0 < \alpha \leq 1$, a test C satisfying $\overline{\alpha}_C \leq \alpha$ will be called an *level* α test. α is called the *significance level*.

9.2 Neyman-Pearson Lemma

- **Theorem: Neyman-Pearson Lemma.** Let $\boldsymbol{x}_n = (x_1, \cdots, x_n)$ be an observation from distribution with joint PMF/PDF $f(\boldsymbol{x}_n; \theta)$. Consider testing a **simple** null hypothesis $H_0 : \theta = \theta_0$ versus a **simple** alternative $H_1 : \theta = \theta_1$. Suppose a test with rejection and acceptance regions $\mathbb{R}_n(c)$ and $\mathbb{A}_n(c)$ defined as follows

$$\mathbb{R}_n(c) = \Big\{\boldsymbol{x}^n : \frac{f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta_1)}{f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta_0)} > c\Big\}$$

and

$$\mathbb{A}_n(c) = \Big\{\boldsymbol{x}^n : \frac{f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta_1)}{f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta_0)} \leq c\Big\}$$

for some $c \geq 0$, and

$$P\Big[\boldsymbol{X}^n \in \mathbb{R}_n(c) | H_0\Big] = \alpha.$$

9.2 Neyman-Pearson Lemma

- Then

  (1) [**Sufficiency**] Any test that satisfies the above conditions is a uniformly most powerful level α test.

  (2) [**Necessity**] If there exists a test satisfying the above conditions with $c > 0$, then every uniformly most powerful level α test is a size α, and every uniformly most powerful level α test satisfies the first two conditions except perhaps on a set A satisfying $P(\boldsymbol{X}^n \in A|H_0) = P(\boldsymbol{X}^n \in A|H_A) = 0$.

9.2 Neyman-Pearson Lemma

- **Remarks:**
  - $f(\boldsymbol{X}_n;\theta_1)/f(\boldsymbol{X}_n;\theta_0)$ is called the *likelihood ratio statistic*.
  - The test $C = \{\boldsymbol{x}_n : f(\boldsymbol{x}_n;\theta_1)/f(\boldsymbol{x}_n;\theta_0) \geq k\}$ is called the *likelihood ratio test*.
  - By Neyman-Pearson lemma, we want the likelihood ratio to be as large as possible in the critical region.

9.2 Neyman-Pearson Lemma

- **Corollary.** Suppose $T(\mathbf{X}^n)$ is a sufficient statistic for θ and $g(t, \theta_i)$ is the PMF/PDF of $T(\boldsymbol{X}^n)$ corresponding to $\theta_i, i = 0, 1$. Then any test based on $T(\boldsymbol{X}^n)$ with rejection region $\mathbb{R}_n$ is a uniformly most powerful level α test for $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_A : \theta = \theta_1$ if the test has the rejection and acceptance regions
$$\mathbb{R}_n(c) = \left\{t : \frac{g(t, \theta_1)}{g(t, \theta_0)} > c\right\}$$
and
$$\mathbb{A}_n(c) = \left\{t : \frac{g(t, \theta_1)}{g(t, \theta_0)} \leq c\right\}$$
for some $c \geq 0$, where $P[T(\boldsymbol{X}^n) \in \mathbb{R}_n(c) | \mathbb{H}_0] = \alpha$.

- **Example:** Suppose random variables $X_1, \cdots, X_n$ are i.i.d. from the EXP(λ) distribution with pdf $f(x; \lambda) = \lambda e^{-\lambda x}$ ($E(X_i) = 1/\lambda$). Test hypothesis $H_0 : \lambda = 5$ against $H_1 : \lambda = 8$ at the significance level α.
  - The likelihood ratio is
  $$\frac{f(\boldsymbol{x}_n; \lambda = 8)}{f(\boldsymbol{x}_n; \lambda = 5)} = \frac{\prod_{i=1}^n 8e^{-8x_i}}{\prod_{i=1}^n 5e^{-5x_i}} = \left(\frac{8}{5}\right)^n e^{-3\sum_{i=1}^n x_i}.$$
  - The most powerful test has the form of $\{\frac{f(\boldsymbol{x}_n;\lambda=8)}{f(\boldsymbol{x}_n;\lambda=5)} > k\}$, which is equivalent to $\{\sum_{i=1}^n x_i < k^*\}$.
  - Find k^* such that $P_{\lambda=5}(\sum_{i=1}^n X_i < k^*) = \alpha$, where $2\lambda \sum_{i=1}^n X_i$ follows a χ^2_{2n} distribution. If $n = 10$, $\alpha = 0.1$, then we can obtain $k^* = 1.244$ (0.1-quantile of χ^2_{20} is 12.44).
  - $\{\sum_{i=1}^n x_i < k^*\}$ is the most powerful test at significant level α.
  - The probability of making type II error is 0.4637.

9.2 Neyman-Pearson Lemma

- **Remarks:**

  - Here $\sum_{i=1}^n X_i$ is called a *test statistic*, and k^* is called the *critical value*.
  - In this example, if we observed a realization $(x_1, \cdots, x_n)$, $P_{\lambda=5}(\sum_{i=1}^n X_i < \sum_{i=1}^n x_i)$ is called the *p-value* of the observation.
  - The p value is the probability of obtaining a test statistic at least as extreme as the one that was observed. H_0 will be accepted if p value of the observation is larger than α, H_0 will be rejected if p value of the observation is less than α.
  - In this example, consider testing $H_0 : \lambda = 8$ against $H_1 : \lambda = 5$. The most powerful critical region at significance level $\alpha = 0.1$ is $\{\sum_{i=1}^n x_i > 1.7757\}$. It is possible to find some observations that are accepted by both $H_0 : \lambda = 5$ and $H_0 : \lambda = 8$ at significance level $\alpha = 0.1$.

- **Example:** Suppose $X_1, \cdots, X_n$ are i.i.d. from normal distribution $N(\mu, \sigma^2)$ with σ^2 known. Consider testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$ at significance level α.
  - For any fixed $\mu' > 0$, consider the likelihood ratio test for $H_0 : \mu = 0$ against $H_1 : \mu = \mu'$.
  - We have
$$\frac{f(\boldsymbol{x}_n; \mu = \mu')}{f(\boldsymbol{x}_n; \mu = 0)} \propto exp\left\{(\mu'/\sigma^2)\sum_{i=1}^{n} x_i\right\}.$$
  - Because $\mu' > 0$, the most powerful test has the form of $\{\sum_{i=1}^n x_i > k\}$.
  - Find k^* such that $P_{\mu=0}(\sum_{i=1}^n X_i > k^*) = \alpha$, where $\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n X_i$ follows a standard normal distribution.
  - Because the critical regions (or k^*) are the same for all $\mu' > 0$, $C = \{\sum_{i=1}^n X_i > k^*\}$ is the UMP test at significant level α.

- **Example:** Suppose $X_1, \cdots, X_n$ are i.i.d. from normal distribution $N(\mu, \sigma^2)$ with σ^2 known. Consider testing $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ at significance level α.
  - Fixed $\mu' \neq 0$, consider the likelihood ratio test for $H_0 : \mu = 0$ against $H_1 : \mu = \mu'$.
  - The critical regions obtained by using the likelihood ratio test are $\{\sum_{i=1}^n X_i > k^*\}$ and $\{\sum_{i=1}^n X_i < k^*\}$ for cases $\mu' > 0$ and $\mu' < 0$, respectively. The critical regions are different for different μ'. The UMP test does not exist in this case.
  - We can use $\{|\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n X_i| > k^*\}$, where $k^* > 0$ is the value that makes $P_{\mu=0}(|\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n X_i| > k^*) = \alpha$.
  - However, $\{|\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n X_i| > k^*\}$ is not a UMP test.

9.3 Wald Test and Lagrangian Multiplier Test

Wald Test: Consider testing

$$H_0 : g(\theta) = 0 \quad \text{against} \quad H_1 : g(\theta) \neq 0.$$

$$H_0 : g(\theta_0) = 0 \quad \text{against} \quad H_1 : g(\theta_0) \neq 0.$$

- **Assumptions:**
  - **B.1:** $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V)$, where V is a $p \times p$ symmetric bounded and non-singular covariance matrix, θ_0 is the true parameter value which is an interior point in Θ, and Θ is a compact parameter space.
  - **B.2:** $\widehat{V}_n \xrightarrow{p} V$, where $\widehat{V}_n$ is a function of $X_1, \cdots, X_n$.
  - **B.3:** $g : \mathbb{R}^p \to \mathbb{R}^q$ $(q \leq p)$ is a continuously differentiable vector-valued function of θ. The first derivative $G(\theta_0) \triangleq \frac{\partial}{\partial \theta} g(\theta_0)$ is a $q \times p$ matrix and has rank q.
- **Remark:** $\widehat{\theta}_n$ can be the MLE or MME.

- **Theorem: Asymptotic Distribution of Wald Test Statistic.** Suppose Assumptions **B.1-B.3** hold. Under H_0,

$$\widehat{W}_n \triangleq n g(\widehat{\theta}_n)' \left[G(\widehat{\theta}_n) \widehat{V}_n G(\widehat{\theta}_n)' \right]^{-1} g(\widehat{\theta}_n) \xrightarrow{d} \chi_q^2.$$

  – **Remarks:**

  – In Wald test, we only need to consider parameter estimation in the whole parameter space Θ.

  – Test $C = \{\boldsymbol{x}_n : \widehat{W}_n > \chi^2_{\alpha,q}\}$ is an asymptotically level α test, where $\chi^2_{\alpha,q}$ is the upper α-quantile of χ^2_q distribution..

9.3 Wald Test and Lagrangian Multiplier Test

Lagrangian Multiplier Test:

- Consider maximizing the normalized log-likelihood $l(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i;\theta)$ under $H_0 : g(\theta) = 0$.

- The Lagrangian function is

$$L(\theta;\lambda) = l(\theta) + \lambda' g(\theta),$$

  where λ is a $q \times 1$ vector.

- $\widetilde{\theta}_n$ and $\widetilde{\lambda}_n$ are the solutions of the first order conditions

$$\begin{aligned} \frac{\partial L(\widetilde{\theta}_n, \widetilde{\lambda}_n)}{\partial \theta} &= \frac{\partial l(\widetilde{\theta}_n)}{\partial \theta} + \widetilde{\lambda}_n' G(\widetilde{\theta}_n) = 0, \\ \frac{\partial L(\widetilde{\theta}_n, \widetilde{\lambda}_n)}{\partial \lambda} &= g(\widetilde{\theta}_n) = 0. \end{aligned}$$

9.3 Wald Test and Lagrangian Multiplier Test

- **Theorem: Asymptotic Distribution of Lagrangian Multiplier Test Statistic.** Suppose Assumptions **A.1**-**A.6** hold, and Assumption B.3 holds. Under H_0,

$$LM \triangleq -n\widetilde{\lambda}_n' G(\widetilde{\theta}_n) \left[\widehat{H}(\widetilde{\theta}_n)\right]^{-1} G(\widetilde{\theta}_n)'\widetilde{\lambda}_n \xrightarrow{d} \chi_q^2,$$

- **Remarks:**
  - An asymptotically size α LM test will reject the null hypothesis $\mathbb{H}_0$: $g(\theta_0) = 0$ when LM exceeds the $(1-\alpha)$th quantile of the χ^2 distribution.
  - In Lagrangian multiplier test, we only need to maximize the likelihood function in the null parameter space $\Theta_0 = \{\theta : g(\theta) = 0\}$.
  - Test $C = \{\boldsymbol{x}_n : LM_n > \chi^2_{\alpha,q}\}$ is an asymptotically level α test.
  - In many cases, Wald test and Lagrangian multiplier test are not as powerful as the generalized likelihood ratio test.

9.4 The Likelihood Ratio Test

- **Likelihood Ratio(LR) Test.** Consider a LR test for $\mathbb{H}_0 : g(\theta) = 0$. Suppose $\boldsymbol{X}^n$ is an IID random sample from the population $f(x, \theta_0)$, where θ_0 is an unknown parameter value in Θ. Then the likelihood function of $\boldsymbol{X}^n$ is

$$f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta) = \Pi_{i=1}^n f(X_i, \theta).$$

The likelihood ratio statistic is defined as

$$\hat{\Lambda} = \frac{\max_{\theta \in \Theta} f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta)}{\max_{\theta \in \Theta_0} f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta)} = \frac{\prod_{i=1}^n f(X_i, \hat{\theta}_n)}{\prod_{i=1}^n f(X_i, \tilde{\theta}_n)},$$

where $\hat{\theta}_n$ and $\tilde{\theta}_n$ are the unconstrained and constrained MLE estimators respectively, that is $\hat{\theta}_n = \arg\max_{\theta \in \Theta} \hat{l}(\theta)$, and $\tilde{\theta}_n = \arg\max_{\theta \in \Theta_0} \hat{l}(\theta)$, with $\hat{l}(\theta) = \frac{1}{n} \sum_{i=1}^n \ln f(X_i, \theta)$. And Θ_0 is the parameter space Θ subject to the constraint $g(\theta) = 0$.

9.4 Likelihood Ratio Test

- **Example:** Suppose $X_1, \cdots, X_n$ are i.i.d. from normal distribution $N(\mu, \sigma^2)$. Consider testing $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. Find the likelihood ratio test (1) when σ^2 is known; (2) when σ^2 is unknown.

- (1) When σ^2 is known,

$$
\begin{aligned}
\lambda(\boldsymbol{x}_n) &= exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\} \\
&= exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\} \\
&= exp\left\{\frac{n}{2\sigma^2}\overline{x}_n^2\right\},
\end{aligned}
$$

the corresponding GLR test is $\{\boldsymbol{x}_n : \lambda(\boldsymbol{x}_n) > k\}$, which is equivalent to $\{\boldsymbol{x}_n : |\frac{\sqrt{n}}{\sigma}\overline{x}_n| > k^*\}$. Under H_0, the test statistic $\frac{\sqrt{n}}{\sigma}\overline{X}_n$ follows a $N(0,1)$ distribution.

- (2) When σ^2 is unknown,

$$\sup_{\mu,\sigma^2} f(\boldsymbol{x}_n;\mu,\sigma^2) = \left[\frac{n}{2\pi\sum_{i=1}^n (x_i-\overline{x}_n)^2}\right]^{n/2} e^{-n/2},$$

$$\sup_{\mu=0,\sigma^2} f(\boldsymbol{x}_n;\mu,\sigma^2) = \left[\frac{n}{2\pi\sum_{i=1}^n x_i^2}\right]^{n/2} e^{-n/2},$$

we have

$$\begin{aligned}\lambda(\boldsymbol{x}_n) &= \left[\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i-\overline{x}_n)^2}\right]^{n/2} \\ &= \left[1+\frac{n\overline{x}_n^2}{\sum_{i=1}^n (x_i-\overline{x}_n)^2}\right]^{n/2}.\end{aligned}$$

The corresponding GLR test is $\left\{\boldsymbol{x}_n : \frac{\sqrt{n(n-1)}|\overline{x}_n|}{\sqrt{\sum_{i=1}^n (x_i-\overline{x}_n)^2}} > k\right\}$. Under H_0, the test statistic $\frac{\sqrt{n(n-1)}\,\overline{X}_n}{\sqrt{\sum_{i=1}^n (X_i-\overline{X}_n)^2}}$ follows a t_{n-1} distribution.

9.4 Likelihood Ratio Test

- **Theorem: LR Test.** Suppose Assumptions **A.1**-**A.6**, and **B.3** hold. Consider testing

$$H_0 : g(\theta_0) = 0 \quad \text{against} \quad H_1 : g(\theta_0) \neq 0.$$

Under H_0,

$$LR \stackrel{\triangle}{=} 2 \ln \hat{\Lambda} = 2n \left[\widehat{l}_n(\widehat{\theta}_n) - \widehat{l}_n(\widetilde{\theta}_n) \right] \xrightarrow{d} \chi^2_q.$$

- **Remarks:**
  - GLR test statistic is a function of any sufficient statistic T.
  - To construct an asymptotically level α test, we can let critical region $C = \{\boldsymbol{x}_n : 2 \log \lambda(\boldsymbol{x}_n) > \chi^2_{\alpha,q}\}$, where $\chi^2_{\alpha,q}$ is the upper α-quantile of χ^2_q distribution.
  - We can show that if $g(\theta_0) \neq 0$ (H_1 holds), the power function of $C = \{\boldsymbol{x}_n : 2 \log \lambda(\boldsymbol{x}_n) > \chi^2_{\alpha,q}\}$ at θ_0 tends to 1 as $n \to \infty$.

9.4 Likelihood Ratio Test

- **Theorem.** If $T(\boldsymbol{X}^n)$ is a sufficient statistic for θ, and $LR(\boldsymbol{X}^n)$ and $LR[T(\boldsymbol{X}^n)]$ are the likelihood ratio based on $\boldsymbol{X}^n$ and $T(\boldsymbol{X}^n)$ respectively, then

$$LR(\boldsymbol{X}^n) = LR[T(\boldsymbol{X}^n)].$$

- **Examples.**

- **Conclusion.**