Chapter 8

Hypothesis Testing

8.1 Introduction

Definition 8.1.1: A ***Hypothesis*** is a statement about a population parameter.

Definition 8.1.2: 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. The general format of the null and alternative hypotheses is

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_0^c$$

where $\Theta_0 \subset \Theta$ and $\Theta_0^c = \Theta \setminus \Theta_0$.

Definition 8.1.3: A *hypothesis testing procedure* or ***Hypothesis Test*** is a rule that specifies:

1. For which sample values the decision is made to accept H_0 as true.
2. For which sample values H_0 is rejected and H_1 is accepted as true.

Remark:

1. The distinction between "rejecting H_0" and "accepting H_1" or between "accepting H_0"and "not rejecting H_0" on a philosophical level will not be concerned.
2. A hypothesis testing problem can be viewed as a problem to make the assertion of H_0 or H_1.

Definition: The subset of the sample space for which H_0 will be rejected is called the ***Rejection Region*** or *critical region*. The complement of the rejection region is called the ***Acceptance Region***.

Definition: A ***Test Statistic*** $W(X_1, \ldots, X_n) = W(\mathbf{X})$ is a function of the sample, in terms of which a hypothesis test is typically specified.

8.2 Methods of Finding Tests

8.2.1 Likelihood Ratio Tests

Definition 8.2.1: The likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{x})}.$$

A ***Likelihood Ratio Test*** **(LRT)** is any test that has a rejection region of the form $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$, where c is any number satisfying $0 \leq c \leq 1$.

LRT and MLE: Let $\hat{\theta}$ be the MLE of θ under the unrestricted parameter space Θ and $\hat{\theta}_0$ be the MLE of θ under the restricted parameter space Θ_0. Then

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

Example 8.2.2: (Normal LRT)
Let $X_1, \dots, X_n$ be iid $\mathrm{n}(\theta, 1)$. We want to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, where θ_0 is a fixed number set by the experimenter. Show that

$$\lambda(\mathbf{x}) = \exp\left[-\frac{n(\bar{x}-\theta_0)^2}{2}\right]$$

so that the LRT rejects H_0 for small values of $\lambda(\mathbf{x})$. Therefore the rejection region is

$$\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$$

which is equivalent to

$$\left\{\mathbf{x} : |\bar{x}-\theta_0| \geq \sqrt{-2(\log c)/n}\right\}.$$

Example 8.2.3: (Exponential LRT)

Let $X_1, \dots, X_n$ be a random sample from an exponential population with pdf

$$f(x|\theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x)$$

where $-\infty < \theta < \infty$.

Considering testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a fixed number set by the experimenter. Show that

$$\lambda(\mathbf{x}) = \begin{cases} 1 & x_{(1)} \leq \theta_0 \\ e^{-n(x_{(1)}-\theta_0)} & x_{(1)} > \theta_0 \end{cases}.$$

Since the LRT rejects H_0 for small values of $\lambda(\mathbf{x})$, the rejection region is

$$\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\} \iff \left\{\mathbf{x} : x_{(1)} \geq \theta_0 - \frac{\log(c)}{n}\right\}$$

Remark: In both of the above examples, the rejection region only depends on the sufficient statistic for θ.

Theorem 8.2.4: If $T(\mathbf{X})$ is a sufficient statistic for θ and $\lambda^*(t)$ and $\lambda(\mathbf{x})$ are the LRT statistics based on T and $\mathbf{X}$, respectively, then $\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x})$ for every $\mathbf{x}$ in the sample space.

Example 8.2.5: (LRT and Sufficiency)

- In Example 8.2.2, we could have used the likelihood associated with the sufficient statistic $\bar{X}$ using the fact that $\bar{X} \sim \mathrm{n}(\theta, 1/n)$, which rejects for large values of $\left|\bar{X} - \theta_0\right|$.
- Similarly, in Example 8.2.3, we can use the likelihood associated with the sufficient statistic $X_{(1)}$, $L(\theta|x_{(1)}) = n \exp\left[-n(x_{(1)} - \theta)\right] I_{[\theta,\infty)}\left(x_{(1)}\right)$, which rejects for large values of $X_{(1)}$.

Example 8.2.6: (Normal LRT with unknown variance)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$ and an experimenter is interested only in inferences about μ, such as testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$. Then the parameter σ is a nuisance parameter. The LRT statistic is

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{\max\limits_{\{\mu,\sigma^2 : \mu \leq \mu_0, \sigma^2 \geq 0\}} L(\mu, \sigma^2 | \mathbf{x})}{\max\limits_{\{\mu,\sigma^2 : -\infty < \mu < \infty, \sigma^2 \geq 0\}} L(\mu, \sigma^2 | \mathbf{x})} \\
&= \frac{\max\limits_{\{\mu,\sigma^2 : \mu \leq \mu_0, \sigma^2 \geq 0\}} L(\mu, \sigma^2 | \mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2 | \mathbf{x})} \\
&= \begin{cases} 1 & \text{if } \hat{\mu} \leq \mu_0 \\ \dfrac{L(\mu_0, \hat{\sigma}_0^2 | \mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2 | \mathbf{x})} & \text{if } \hat{\mu} > \mu_0 \end{cases}
\end{aligned}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the MLEs of μ and σ^2, and $\hat{\sigma}_0^2 = \sum (x_i - \mu_0)^2 / n$.

8.2.2 Bayesian Tests

Bayesian Formulation of Hypothesis Testing

Classical Approach

- The parameter θ is fixed.
- If $\theta \in \Theta_0$ is known, then $P(\theta \in \Theta_0|\mathbf{x}) = 1$ and $P(\theta \in \Theta_0^c|\mathbf{x}) = 0$ for all $\mathbf{x}$.
- If $\theta \in \Theta_0^c$ is known, then $P(\theta \in \Theta_0|\mathbf{x}) = 0$ and $P(\theta \in \Theta_0^c|\mathbf{x}) = 1$.
- In practice, $P(\theta \in \Theta_0|\mathbf{x})$ and $P(\theta \in \Theta_0^c|\mathbf{x})$ are unknown and do not depend on $\mathbf{x}$. Hence these probabilities are not used.

Bayesian Approach

- The parameter θ is random and is assigned a prior distribution.
- The $P(\theta \in \Theta_0|\mathbf{x}) = P(H_0 \text{ is true}|\mathbf{x})$ and $P(\theta \in \Theta_0^c|\mathbf{x}) = P(H_1 \text{ is true}|\mathbf{x})$ can be computed and make sense.
- A way to use posterior distribution to make decisions about H_0 and H_1 is to decide to accept H_0 as true if $P(\theta \in \Theta_0|\mathbf{X}) > P(\theta \in \Theta_0^c|\mathbf{X})$ and to reject H_0 otherwise. The test statistic, a function of the sample, is $P(\theta \in \Theta_0^c|\mathbf{X})$ and the rejection region is $\{\mathbf{x} : P(\theta \in \Theta_0^c|\mathbf{x}) > 1/2\}$.
- One may also define a rejection region as $\{\mathbf{x} : P(\theta \in \Theta_0^c|\mathbf{x}) > c_p\}$, where $0 < c_p < 1$, say $c_p = 0.99$, which is set by the researcher.

Example 8.2.7: (Normal Bayesian Test)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$ and let the prior distribution on θ be $\mathrm{n}(\mu, \tau^2)$, where σ^2, τ^2, μ are known. Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. If we decide to accept H_0 if and only if $P(\theta \in \Theta_0|\mathbf{X}) \geq P(\theta \in \Theta_0^c|\mathbf{X})$, then we will accept H_0 if and only if

$$P(\theta \leq \theta_0|\mathbf{X}) = P(\theta \in \Theta_0|\mathbf{X}) \geq 1/2$$

Recall that the posterior distribution $\pi(\theta|\bar{x})$ is normal with mean $\dfrac{n\tau^2\bar{x} + \sigma^2\mu}{n\tau^2 + \sigma^2}$ and variance $\dfrac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}$. Therefore, H_0 will be accepted as true if

$$\bar{X} \leq \theta_0 + \frac{\sigma^2(\theta_0 - \mu)}{n\tau^2}.$$

8.2.3 Union-Intersection and Intersection-Union Tests

Union-Intersection Method

Let Γ be an arbitrary index set that may be finite or infinite. Define

$$H_0 : \theta \in \bigcap_{\gamma\in\Gamma} \Theta_\gamma.$$

Suppose that for each γ, a test is available for

$$H_{0\gamma} : \theta \in \Theta_\gamma \quad \text{versus} \quad H_{1\gamma} : \theta \in \Theta_\gamma^c,$$

which rejects $H_{0\gamma}$ when $\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma$. Then the rejection region for the union-intersection test is:

$$\bigcup_{\gamma\in\Gamma} \{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}$$

If any one of the $H_{0\gamma}$ is rejected, then H_0 is rejected. Equivalently, H_0 is true only if $H_{0\gamma}$ is true for every γ. In particular, if each test has a rejection region $\{\mathbf{x} : T_\gamma(\mathbf{x}) > c\}$, where c does no depend on γ, then the rejection region for the union-intersection test can be expressed as

$$\bigcup_{\gamma\in\Gamma} \{\mathbf{x} : T_\gamma(\mathbf{x}) > c\} = \left\{\mathbf{x} : \sup_{\gamma\in\Gamma} T_\gamma(\mathbf{x}) > c\right\}$$

Thus the union-intersection test statistic is $T(\mathbf{X}) = \sum_{\gamma\in\Gamma} T_\gamma(\mathbf{X})$.

Example 8.2.8: (Normal Union-Intersection Test)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$. Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. H_0 is equivalent to the intersection of the two sets:

$$H_0 : \{\mu : \mu \leq \mu_0\} \cap \{\mu : \mu \geq \mu_0\}.$$

Test 1: $H_{01} : \mu \leq \mu_0$ versus $H_{11} : \mu > \mu_0$

$$\text{LRT rejects } H_{01} \text{ if } \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq t_{\mathrm{L}}.$$

Test 2: $H_{02} : \mu \geq \mu_0$ versus $H_{12} : \mu < \mu_0$

$$\text{LRT rejects } H_{02} \text{ if } \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq t_{\mathrm{U}}.$$

Thus the union-intersection test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ formed from these two LRTs is

$$\text{reject } H_0 \text{ if } \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq t_{\mathrm{L}} \text{ or } \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq t_{\mathrm{U}}$$

If $t_{\mathrm{L}} = -t_{\mathrm{U}} \geq 0$, the union-intersection test can be simply expressed as

$$\text{reject } H_0 \text{ if } \frac{|\bar{X} - \mu_0|}{S/\sqrt{n}} \geq t_{\mathrm{L}}.$$

<u>**Intersection-Union Method**</u>

Let Γ be an arbitrary index set that may be finite or infinite. Define

$$H_0 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma$$

Suppose that for each γ, a test is available for

$$H_{0\gamma} : \theta \in \Theta_\gamma \quad \text{versus} \quad H_{1\gamma} : \theta \in \Theta_\gamma^c,$$

which rejects $H_{0\gamma}$ when $\{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}$. Then the rejection region for the intersection-union test is

$$\bigcap_{\gamma \in \Gamma} \{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}.$$

i.e., reject H_0 if and only if $H_{0\gamma}$ is rejected for all γ. In particular, if each test has a rejection region $\{\mathbf{x} : T_\gamma(x) \geq c\}$, where c is independent of γ, then the rejection region for H_0

$$\bigcap_{\gamma \in \Gamma} \{\mathbf{x} : T_\gamma(\mathbf{x}) \geq c\} = \left\{\mathbf{x} : \inf_{\gamma \in \Gamma} T_\gamma(\mathbf{x}) \geq c\right\}$$

Thus the intersection-union test statistic is $T(\mathbf{X}) = \inf_{\gamma \in \Gamma} T_\gamma(\mathbf{X})$.

<u>**Example 8.2.9:**</u> Suppose $X_1, \ldots, X_n$ are measurements of breaking strength assumed to be iid $n(\theta_1, \sigma^2)$ and $Y_1, \ldots, Y_m$ are the results of m flammability tests modeled as iid Bernoulli(θ_2), where $Y_i = 1$ if the unit passes the test and $Y_i = 0$ otherwise. Standards to be met: $\theta_1 > 50$ and $\theta_2 > 0.95$, modeled with the hypothesis test

$$H_0 : \{\theta_1 \leq 50 \text{ or } \theta_2 \leq 0.95\} \quad \text{versus} \quad H_1 : \{\theta_1 > 50 \text{ and } \theta_2 > 0.95\}$$

Test 1: $H_{01} : \theta_1 \leq 50$ versus $H_{11} : \theta_1 > 50$

$$\text{LRT rejects } H_{01} \text{ if } \frac{\bar{X} - 50}{S/\sqrt{n}} > t.$$

Test 2: $H_{02} : \theta_2 \leq 0.95$ versus $H_{12} : \theta_2 > 0.95$

$$\text{LRT rejects } H_{02} \text{ if } \sum_{i=1}^{m} Y_i > b.$$

Thus the rejection region for the intersection-union test is given by

$$\left\{(x, y) : \frac{\bar{x} - 50}{s/\sqrt{n}} > t \text{ and } \sum_{i=1}^{m} y_i > b\right\}.$$

8.3 Methods of Evaluating Tests

8.3.1 Error Probabilities and Power Function

Two Types of Error:

- ***Type I Error***: $\theta \in \Theta_0$ but the hypothesis test incorrectly decides to reject H_0.
- ***Type II Error***: $\theta \in \Theta_0^c$ but the test decides to accept H_0.

Let R denote the rejection region for a test. Then

$$P_\theta(\mathbf{X} \in R) = \begin{cases} P(\text{Type I Error}) & \text{if } \theta \in \Theta_0 \\ 1 - P(\text{Type II Error}) & \text{if } \theta \in \Theta_0^c \end{cases}.$$

Definition 8.3.1: The ***Power Function*** of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\mathbf{X} \in R)$.

Example 8.3.2: (Binomial Power Function)
Let $X \sim \text{binomial}(5, \theta)$. Consider:

$$H_0 : \theta \leq 1/2 \text{ versus } H_1 : \theta > 1/2.$$

Test 1: $R = \{\text{All "successes" are observed}\}$.
Test 2: $R = \{X = 3, 4, \text{ or } 5\}$.

Example 8.3.3: (Normal Power Function)

Let $X_1, \dots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$ where σ^2 is known. An LRT of $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ is a test that rejects H_0 if $(\bar{X} - \theta_0)/(\sigma/\sqrt{n}) > c$. The power function of this test is

$$\begin{aligned}\beta(\theta) &= P_\theta\left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c\right) \\ &= P_\theta\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right),\end{aligned}$$

where Z is a standard normal random variable.

Example 8.3.4: (Continuation of Example 8.3.3)

Suppose the experimenter wishes to have a maximum Type I Error probability of 0.1 and a maximum Type II Error probability of 0.2 if $\theta \ge \theta_0 + \sigma$. How do we choose c and n?

Definition 8.3.5: For $0 \le \alpha \le 1$, a test with power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$.

Definition 8.3.6: For $0 \le \alpha \le 1$, a test with power function $\beta(\theta)$ is a level α test if $\sup_{\theta \in \Theta_0} \beta(\theta) \le \alpha$.

Remark:

1. Some authors use the terms level and size α interchangeably.
2. The set of level α tests contains the set of size α tests.
3. The distinction becomes important in complicated testing situations (e.g., intersection-union and union-intersection tests), where it is often computationally impossible to construct a size α test and an experimenter have to settle for a level α test.
4. The commonly used α values in practice are 0.1, 0.05 and 0.01.
5. Fixing the level of a test is controlling the Type I error but not Type II error.
6. H_0 and H_1 should be set up properly so that the more important error to control is the Type I error.
7. H_1 is typically the hypothesis that we expect the data to support, and hope to prove. (The alternative hypothesis is hence sometimes called the ***Research Hypothesis*** in this context.)

Example 8.3.7: (Size of LRT)
A size α LRT is constructed by choosing the appropriate c such that

$$\sup_{\theta\in\Theta_0} P_\theta\big(\lambda(X) \le c\big) = \alpha.$$

In Example 8.2.2, the test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ with following rejection region , is the size α LRT.

$$R = \left\{\mathbf{x} : |\bar{x} - \theta_0| \ge \frac{z_{\alpha/2}}{\sqrt{n}}\right\}$$

In Example 8.2.3, finding a size α is more complicated because $H_0 : \theta \le \theta_0$ consists of more than one point. Since θ is a location parameter for $X_{(1)}$,

$$P_\theta(X_{(1)} \ge c) \le P_{\theta_0}(X_{(1)} \ge c), \qquad \text{for any } \theta \le \theta_0.$$

Thus

$$\sup_{\theta\in\Theta_0} \beta(\theta) = \sup_{\theta\le\theta_0} P_\theta(X_{(1)} \ge c) = P_{\theta_0}(X_{(1)} \ge c) = e^{-n(c-\theta_0)} = \alpha$$

which implies that $c = -\dfrac{\log(\alpha)}{n} + \theta_0$ yields the size α LRT.

<u>Example 8.3.8</u>: (Size of Union-Intersection Test)

The problem of finding size α union-intersection test in Example 8.2.8 involves finding t_{L} and t_{U} such that

$$\sup_{\theta\in\Theta_0} P_\theta\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \geq t_{\mathrm{L}} \text{ or } \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq t_{\mathrm{U}}\right) = \alpha.$$

For any $(\mu, \sigma^2) \in \Theta$, $\mu = \mu_0$ and thus $\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$ has a Student's t distribution with $n-1$ degrees of freedom. So any choice of $t_{\mathrm{U}} = t_{n-1,1-\alpha_1}$ and $t_{\mathrm{L}} = t_{n-1,\alpha_2}$, with $\alpha_1 + \alpha_2 = \alpha$, will yield a test with Type I Error probability of exactly α for all $\theta \in \Theta_0$. The usual choice is $t_{\mathrm{L}} = -t_{\mathrm{U}} = t_{n-1,\alpha/2}$.

<u>Definition 8.3.9</u>: A test with power function $\beta(\theta)$ is unbiased if $\beta(\theta') \geq \beta(\theta'')$ for every $\theta' \in \Theta_0^c$ and $\theta'' \in \Theta_0$.

<u>Example 8.3.10</u>: (Conclusion of Example 8.3.3)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$ where σ^2 is known. An LRT of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ has power function

$$\beta(\theta) = P\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)$$

where $Z \sim \mathrm{n}(0,1)$. Since $\beta(\theta)$ is an increasing function of θ, it follows that

$$\beta(\theta) > \beta(\theta_0) = \max_{t\leq\theta_0} \beta(t), \text{ for all } \theta > \theta_0.$$

Thus the test is unbiased.

8.3.2 Most Powerful Tests

Definition 8.3.11: Let $\mathcal{C}$ be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a ***Uniformly Most Powerful*** (**UMP**) *class $\mathcal{C}$ test* if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$.

Theorem 8.3.12: (Neyman-Pearson Lemma)
Considering testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf or pmf corresponding to θ_i if $f(\mathbf{x}|\theta_i)$, $i = 0, 1$, using a test with rejection R that satisfies for some $k \geq 0$,

$$\begin{aligned} \mathbf{x} \in R \ \text{ if } \ f(\mathbf{x}|\theta_1) > kf(\mathbf{x}|\theta_0) \\ \mathbf{x} \in R^c \ \text{ if } \ f(\mathbf{x}|\theta_1) < kf(\mathbf{x}|\theta_0) \end{aligned} \tag{8.3.1}$$

and

$$\alpha = P_{\theta_0}(\mathbf{X} \in R). \tag{8.3.2}$$

Then

a. (Sufficiency) Any test that satisfies (8.3.1) and (8.3.2) is a UMP level α test.

b. (Necessity) If there exists a test satisfying (8.3.1) and (8.3.2) with $k > 0$, then every UMP level α test is a size α test (satisfies (8.3.2)) and every UMP level α test satisfies (8.3.1) except perhaps on a set A satisfying $P_{\theta_0}(\mathbf{X} \in A) = P_{\theta_1}(\mathbf{X} \in A) = 0$.

Corollary 8.3.13: Consider the hypothesis problem posed in Theorem 8.3.12. Suppose $T(\mathbf{X})$ is a sufficient statistic for θ and $g(t|\theta_i)$ is the pdf or pmf of T corresponding to θ_i, $i = 0, 1$. Then any test based on T with rejection region S (a subset of the sample space of T) is a UMP level α test if it satisfies for some $k \geq 0$,

$$\begin{aligned} t \in S & \text{ if } g(t|\theta_1) > kg(t|\theta_0) \\ t \in S^c & \text{ if } g(t|\theta_1) < kg(t|\theta_0) \end{aligned}$$

and

$$\alpha = P_{\theta_0}(T \in S).$$

Example 8.3.14: (UMP Binomial Test)

Let $X \sim \text{binomial}(2, \theta)$. We want to test $H_0 : \theta = 1/2$ versus $H_1 : \theta = 3/4$.

$$\frac{f(0|\theta = 3/4)}{f(0|\theta = 1/2)} = \frac{1}{4}, \quad \frac{f(1|\theta = 3/4)}{f(1|\theta = 1/2)} = \frac{3}{4}, \quad \frac{f(2|\theta = 3/4)}{f(2|\theta = 1/2)} = \frac{9}{4}.$$

- **Test 1**: If we choose $\frac{3}{4} < k < \frac{9}{4}$, then by Neyman-Pearson Lemma, the UMP test rejects H_0 when $X = 2$. The corresponding size of this test is
$$\alpha = P(X = 2|\theta = \frac{1}{2}) = \frac{1}{4}.$$
- **Test 2**: If we choose $\frac{1}{4} < k < \frac{3}{4}$, then by Neyman-Pearson Lemma, the UMP test rejects H_0 when $X = 1$ or 2. The corresponding size of this test is
$$\alpha = P(X = 1 \text{ or } 2|\theta = \frac{1}{2}) = \frac{3}{4}.$$
- **Test 3**: If we choose $k < \frac{1}{4}$ or $k > \frac{9}{4}$, the corresponding UMP tests have size $\alpha = 1$ and $\alpha = 0$, respectively.
- **Test 4**: If we choose $k = \frac{3}{4}$, then the UMP test must reject H_0 for $x = 2$ and accept H_0 for $x = 0$, but leaves the actions for $x = 1$ undetermined. We can put $X = 1$ in either the rejection or acceptance region. The resulting size of the test will depend on this choice because this is a discrete setting.

Example 8.3.15: (UMP Normal Test)
Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$, where σ^2 is known. We want to obtain the UMP test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where $\theta_0 > \theta_1$. Recall that $\bar{X}$ is a sufficient statistic for θ and $\bar{X} \sim \mathrm{n}(\theta, \sigma^2/n)$, which implies that

$$\frac{g(\bar{x}|\theta_1)}{g(\bar{x}|\theta_0)} = \exp\left[\frac{n}{2\sigma^2}\left(2\bar{x}(\theta_1 - \theta_0) - (\theta_1^2 - \theta_0^2)\right)\right].$$

Thus, by Corollary 8.3.13, the UMP test rejects H_0 when $g(\bar{x}|\theta_1) > kg(\bar{x}|\theta_0)$, which is equivalent to

$$\bar{x} < \frac{(2\sigma^2 \log k)/n - \theta_0^2 + \theta_1^2}{2(\theta_1 - \theta_0)}.$$

The test with rejection region $\bar{x} < c$ is the UMP level α test, where

$$\alpha = P_{\theta_0}(\bar{X} < c) = P\left(Z < \frac{c - \theta_0}{\sigma/\sqrt{n}}\right) \Longrightarrow c = \frac{-z_\alpha \sigma}{\sqrt{n}} + \theta_0.$$

Types of Hypotheses:

1. ***Simple Hypothesis***: that specify only one possible distribution for the sample, $H : \theta = \theta_0$.
2. ***Composite Hypothesis***: that specify more than one possible distribution for the sample
   a. One-sided Hypothesis, $H : \theta \leq \theta_0$
   b. Two-sided Hypothesis, $H : \theta \neq \theta_0$

Question: How to find the UMP level α test for composite hypotheses?

Definition 8.3.16: A family of pdfs or pmfs $\{g(t|\theta) : \theta \in \Theta\}$ for a univariate random variable T with real-valued parameter θ has a ***Monotone Likelihood Ratio*** (**MLR**) if, for every $\theta_2 > \theta_1$, $g(t|\theta_2)/g(t|\theta_1)$ is a monotone (nonincreasing or nondecreasing) function of t on $\{t : g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$. Note that $c/0$ is defined as ∞ if $c > 0$.

Note: Any regular exponential family with $g(t|\theta) = h(t)c(\theta)e^{w(\theta)t}$ has an MLR if $w(\theta)$ is a nondecreasing or nonincreasing function.

<u>Theorem 8.3.17</u>: (Karlin-Rubin)
Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Suppose that T is a sufficient statistic for θ and the family of pdfs or pmfs $\{g(t|\theta) : \theta \in \Theta\}$ of T has an MLR, i.e., $g(t|\theta_2)/g(t|\theta_1)$ is a monotone nondecreasing function of t for every $\theta_2 > \theta_1$. Then for any t_0, the test that rejects H_0 if and only if $T > t_0$ is a UMP level α test, where $\alpha = P_{\theta_0}(T > t_0)$.

<u>Note</u>: There are four possible cases:

1. $H_0 : \theta \leq \theta_0$ v.s. $H_1 : \theta > \theta_0$ and nondecreasing $\Rightarrow$ UMP: $T > t_0$
2. $H_0 : \theta \leq \theta_0$ v.s. $H_1 : \theta > \theta_0$ and nonincreasing $\Rightarrow$ UMP: $T < t_0$
3. $H_0 : \theta \geq \theta_0$ v.s. $H_1 : \theta < \theta_0$ and nondecreasing $\Rightarrow$ UMP: $T < t_0$
4. $H_0 : \theta \geq \theta_0$ v.s. $H_1 : \theta < \theta_0$ and nonincreasing $\Rightarrow$ UMP: $T > t_0$

<u>Example 8.3.18</u>: (Continuation of Example 8.3.15)
Consider testing $H_0' : \theta \geq \theta_0$ versus $H_1' : \theta < \theta_0$ using the test that rejects H_0 if

$$\bar{X} < \frac{-z_\alpha \sigma}{\sqrt{n}} + \theta_0.$$

Note that $T = \bar{X} \sim \mathrm{n}(\theta, \sigma^2/n)$ has an MLR. To show this, assume $\theta_2 > \theta_1$, we have

$$\begin{aligned}\frac{g(t|\theta_2)}{g(t|\theta_1)} &= \exp\left\{-\frac{n}{2\sigma^2}\left[(t-\theta_2)^2 - (t-\theta_1)^2\right]\right\} \\ &= \exp\left\{\frac{n(\theta_1^2 - \theta_2^2)}{2\sigma^2}\right\}\exp\left\{\frac{nt(\theta_2 - \theta_1)}{\sigma^2}\right\}.\end{aligned}$$

Since $\theta_2 - \theta_1 > 0$, this ratio is an increasing function of t. By Karlin-Rubin Theorem, the test above is a UMP level α test.

As the power function of this test,

$$\beta(\theta) = P_\theta\left(\bar{X} < -\frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\right) = P\left(Z < -z_\alpha + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right),$$

is a decreasing function of θ, the value of α is give by

$$\sup_{\theta \geq \theta_0} \beta(\theta) = \beta(\theta_0) = \alpha.$$

Example 8.3.19: (Nonexistence of UMP Test)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$ where σ^2 is known. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. A level α test for this problem is any test that satisfies

$$P_{\theta_0}(\text{reject } H_0) \leq \alpha$$

Test 1: Consider $\theta_1 < \theta_0$. Using the same argument as in Example 8.3.18, the test that rejects H_0 if $\bar{X} < -\sigma z_\alpha/\sqrt{n} + \theta_0$ has the highest power at θ_1. Furthermore, by part (b) of the Neyman-Pearson Lemma, any other level α test that has as high a power as Test 1 at θ_1 must have the same rejection region as Test 1 except for a set A satisfying $\int_A f(\mathbf{x}|\theta_i)d\mathbf{x} = 0$.

Test 2: Consider the test that rejects H_0 if $\bar{X} > \sigma z_\alpha/\sqrt{n} + \theta_0$. Let $\beta_1(\theta)$ and $\beta_2(\theta)$ be the power function Test 1 and 2, respectively. Then for any $\theta_2 > \theta_0$, we have

$$\begin{aligned}
\beta_2(\theta_2) &= P_{\theta_2}\left(\bar{X} > \frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\right) \\
&= P_{\theta_2}\left(\frac{\bar{X} - \theta_2}{\sigma/\sqrt{n}} > z_\alpha + \frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}\right) \\
&> P(Z > z_\alpha) \\
&= P(Z < -z_\alpha) \\
&> P_{\theta_2}\left(\frac{\bar{X} - \theta_2}{\sigma/\sqrt{n}} < -z_\alpha + \frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}\right) \\
&= P_{\theta_2}\left(\bar{X} < -\frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\right) \\
&= \beta_1(\theta_2).
\end{aligned}$$

From Neyman-Person Lemma, the UMP level α test would have to be Test 1, but Test 2 has a higher power than Test 1 at θ_2, which implies that there exists no UMP level α test in this problem.

Example 8.3.20: (Unbiased Test)

When no UMP level α test exists within the class of all tests, the next best thing is to find a UMP level α test the class of unbiased tests. Test 1 and Test 2 are not unbiased tests.

Test 3: Rejects $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta \neq \theta_0$ if and only if

$$\bar{X} > \sigma z_{\alpha/2}/\sqrt{n} + \theta_0 \quad \text{or} \quad \bar{X} < -\sigma z_{\alpha/2}/\sqrt{n} + \theta_0.$$

Although Test 1 and Test 2 have slightly higher power than Test 3 for some values of θ, it turns out that Test 3 is a UMP unbiased level α test, that is, it is the UMP in the class of unbiased tests.

8.3.4 p-Values

Definition 8.3.26: A *p-value* $p(\mathbf{X})$ is a est statistic satisfying $0 \le p(\mathbf{x}) \le 1$ for every sample point $\mathbf{x}$. Small values of $p(\mathbf{X})$ give evidence that H_1 is true. A p-value is *valid* if, for every $\theta \in \Theta_0$ and every $0 \le \alpha \le 1$,

$$P_\theta(p(\mathbf{X}) \le \alpha) \le \alpha.$$

Remark:

1. Given a valid p-value $p(\mathbf{X})$, a level α test rejects H_0 if and only if $p(\mathbf{X}) \le \alpha$.
2. A p-value reports the results of a test in a more continuous scale, rather than the dichotomous decision "accept H_0" or "Reject H_0", through which each read can choose the α he or she considers appropriate.
3. The smaller p-value is, the stronger the evidence for rejecting H_0.

Theorem 8.3.27: Let $W(\mathbf{X})$ be a test statistic such that large values of W give evidence that H_1 is true. For each sample point $\mathbf{x}$, define

$$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P_\theta\big(W(\mathbf{X}) \ge W(\mathbf{x})\big).$$

Then, $p(\mathbf{X})$ is a valid p-value.

Example 8.3.28: (Two-Sided Normal p-Value)
Let $X_1, \dots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$. The LRT for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$

$$\text{rejects } H_0 \text{ if } W(\mathbf{X}) = \frac{|\bar{X} - \mu_0|}{S/\sqrt{n}} \text{ is large.}$$

Under $H_0 : \mu = \mu_0$, regardless of the value of σ^2, $\dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$. Thus, the p-value for this two-sided t test is

$$p(\mathbf{x}) = 2P\left(T_{n-1} > \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}\right)$$

where T_{n-1} has Student's t distribution with $n-1$ degrees of freedom.

Example 8.3.29: (One-Sided Normal p-Value)
Let $X_1, \dots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$. The LRT for testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$

$$\text{rejects } H_0 \text{ if } \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \text{ is large.}$$

Note that for all $\mu \leq \mu_0$,

$$\begin{aligned}
P_{\mu,\sigma^2}\big(W(\mathbf{X}) \geq W(\mathbf{x})\big) &= P_{\mu,\sigma^2}\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq W(\mathbf{x})\right) \\
&= P_{\mu,\sigma^2}\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq W(\mathbf{x}) + \frac{\mu_0 - \mu}{S/\sqrt{n}}\right) \\
&= P_{\mu,\sigma^2}\left(T_{n-1} \geq W(\mathbf{x}) + \frac{\mu_0 - \mu}{S/\sqrt{n}}\right) \\
&\leq P\big(T_{n-1} \geq W(\mathbf{x})\big),
\end{aligned}$$

which achieves the supremum at (μ_0, σ^2). Thus, the p-value for this one-sided t test is

$$p(\mathbf{x}) = P\big(T_{n-1} \geq W(\mathbf{x})\big) = P\left(T_{n-1} \geq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right).$$

Alternative Method for Finding p-Values

Let $X(\mathbf{X})$ be a sufficient statistic only for the model $\{f(\mathbf{x}|\theta) : \theta \in \Theta_0\}$. Then for each sample point $\mathbf{x}$ define

$$p(\mathbf{x}) = P(W(\mathbf{X}) \geq W(\mathbf{x}) | S = S(\mathbf{x})).$$

Note that this is a valid p-value because

$$P_\theta(p(\mathbf{x}) \leq \alpha) = \sum_s P(p(\mathbf{x}) \leq \alpha | S = s) P_\theta(S = s) \leq \sum_s \alpha P_\theta(S = s) \leq \alpha.$$

Sums can be replaced by integrals for continuous S, but this alternative method is usually for discrete S.

Example 8.3.30: (Fisher's Exact Test)

Let S_1 and S_2 be independent observations with $S_1 \sim \text{binomial}(n_1, p_1)$ and $S_2 \sim \text{binomial}(n_2, p_2)$. Consider testing $H_0 : p_1 = p_2$ versus $H_1 : p_1 > p_2$. How to construct a valid p-value using the alternative method above? Hint: $S = S_1 + S_2$ is a sufficient statistic under H_0.