Throughout this chapter we consider a sample $X$ taken from a population indexed by $\theta \in \Theta \subset \mathbb{R}^k$.

Instead of estimating the unknown parameter, we want to make a decision on whether the unknown $\theta$ is in $\Theta_0$, a subset of $\Theta$.

**Definition 8.1.1.**

A hypothesis is a statement about the population parameter $\theta$. It is specified as $H : \theta \in \Theta_0$ for a $\Theta_0 \subset \Theta$, where $H$ stands for a hypothesis.

**Definition 8.1.2.**

For any hypothesis $H_0 : \theta \in \Theta_0$, its complementary hypothesis is $H_1 : \theta \in \Theta_1 = \Theta_0^c$. $H_0$ is called the null hypothesis and $H_1$ is called the alternative hypothesis.

Based on a sample from the population, we want to decide which of the two complementary hypotheses is true, i.e., to test

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1 = \Theta_0^c$$
Definition 8.1.3.

A hypothesis test (or simply a test) is a rule that specifies for which sample values $H_0$ is accepted or rejected ($H_1$ is accepted). The subset of the sample space for which $H_0$ is rejected is called the rejection region or critical region. Its complement is called the acceptance region.

- In this course, rejecting $H_0$ results in accepting $H_1$ and accepting $H_0$ results in rejecting $H_1$. Also, "not rejecting" is the same as accepting.
- Typically, a test is specified in terms of a test statistic $T(X) = T(X_1, \ldots, X_n)$, a function of the sample $X$. For example, a test might specify that $H_0$ is to be rejected if the sample mean $\bar{X}$ is greater than 3.

We introduce a method of using the likelihood function to construct tests, which is applicable as long as a likelihood is available. Properties and optimality of these tests will be studied later.
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

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0} L(\theta | X)}{\sup_{\theta \in \Theta} L(\theta | X)},$$

where $L(\theta | x)$ is the likelihood function based on $X = x$.

A likelihood ratio test (LRT) is any test that has a rejection region of the form

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

where $c$ is a constant satisfying $0 \leq c \leq 1$.

- The rationale behind LRTs is that $\lambda(x)$ is likely to be small if there are parameter points in $\Theta_0^c$ for which $x$ is much more likely than for any parameter in $\Theta_0$.
- Note that in the denominator of the likelihood ratio, the sup is taken over $\Theta$, not $\Theta_0^c$.
- The method of determining $c$ is discussed later.
The likelihood ratio method is related to the MLE discussed in Section 7.2.2.
Suppose that \( \hat{\theta} \) is the MLE of \( \theta \) and \( \hat{\theta}_0 \) is the MLE of \( \theta \) when \( \Theta_0 \) is treated as the parameter space (the so-called restricted MLE). Then

\[
\lambda(X) = \frac{L(\hat{\theta}_0|X)}{L(\hat{\theta}|X)}
\]

**Example 8.2.2.**

Let \( X_1, \ldots, X_n \) be iid from \( N(\mu, \sigma^2) \) with unknown \( \mu \in \mathbb{R} \).
Let \( \mu_0 \) be a constant and consider

\[
H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0
\]

First, assume \( \sigma^2 = 1 \).
\( \Theta_0 \) has only one element, \( \mu_0 \), and the numerator of \( \lambda(x) \) is \( L(\mu_0|x) \).
Since the MLE of \( \mu \) is \( \bar{X} \),

\[
\lambda(x) = \frac{(2\pi)^{-n/2} \exp(-\sum_{i=1}^n (x_i - \mu_0)^2/2)}{(2\pi)^{-n/2} \exp(-\sum_{i=1}^n (x_i - \bar{x})^2/2)}
\]
\[= \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_0)^2 + \frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) = \exp \left( -\frac{n}{2} (\bar{x} - \mu_0)^2 \right)\]

Hence,
\[
\{ x : \lambda(x) \leq c \} = \left\{ x : |\bar{x} - \mu_0| \geq \sqrt{-2(\log c)/n} \right\}
\]

Next, we consider the situation where \( \sigma^2 \) is unknown, i.e., \( \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty) \) and \( \Theta_0 = \{ \mu_0 \} \times (0, \infty) \).

Under \( H_0 \), the MLE of \( \theta \) is \( (\mu_0, \hat{\sigma}^2_0) \), and the MLE without restriction is \( \hat{\theta} = (\bar{X}, \hat{\sigma}^2) \), where
\[
\hat{\sigma}^2_0 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2
\]

Then
\[
\lambda(x) = \frac{(2\pi \hat{\sigma}^2_0)^{-n/2} \exp(-\sum_{i=1}^{n} (x_i - \mu_0)^2 / 2\hat{\sigma}^2_0)}{(2\pi \hat{\sigma}^2)^{-n/2} \exp(-\sum_{i=1}^{n} (x_i - \bar{x})^2 / 2\hat{\sigma}^2)} = \frac{\hat{\sigma}^n \exp(-n/2)}{\hat{\sigma}^n_0 \exp(-n/2)}
\]
\[
= \left[ \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2} \right]^n = \left[ 1 + \frac{n(\bar{x} - \mu_0)^2}{(n-1)s^2} \right]^{-n}
\]
where \( s^2 \) is the observed value of the sample variance \( S^2 \). Hence,

\[
\{ x : \lambda(x) \leq c \} = \left\{ x : \frac{\sqrt{n}|\bar{x} - \mu_0|}{\sqrt{n-1}s} \geq \sqrt{c^{-1/n} - 1} \right\}
\]

**Example 8.2.3.**

Let \( X_1, \ldots, X_n \) be iid from \( \text{exponential}(\theta, 1) \) with unknown \( \theta \in \mathbb{R} \) and pdf \( f_\theta(x) = e^{-(x-\theta)}, \ x > \theta \).

Let \( \theta_0 \) be a fixed constant and the hypotheses be

\[
H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0
\]

The likelihood function with \( X = x = (x_1, \ldots, x_n) \) is

\[
L(\theta|x) = \begin{cases} 
    e^{-(x_1-\theta)} \cdots e^{-(x_n-\theta)} & \theta \leq x_i, \ i = 1, \ldots, n, \\
    0 & \text{otherwise}
\end{cases}
\]

\[
= \begin{cases} 
    \exp(- \sum_{i=1}^n x_i + n\theta) & \theta \leq x_{(1)} \\
    0 & \theta > x_{(1)}
\end{cases}
\]

Clearly, \( L(\theta|x) \) is an increasing function of \( \theta \) on \((-\infty, x_{(1)}]\).
Thus, the unrestricted MLE of $\theta$ is $X_{(1)}$ and

$$
\sup_{\theta \in \mathcal{R}} L(\theta|x) = \exp \left( - \sum_{i=1}^{n} x_i + nx_{(1)} \right)
$$

The restricted MLE of $\theta$ depends on whether $x_{(1)} \leq \theta_0$; if yes, the restricted MLE is still $x_{(1)}$; if $x_{(1)} > \theta_0$, the restricted MLE is $\theta_0$ since $L(\theta|x)$ increases for $\theta \leq x_{(1)}$.

Hence,

$$
\sup_{\theta \leq \theta_0} L(\theta|x) = \begin{cases} 
\exp \left( - \sum_{i=1}^{n} x_i + nx_{(1)} \right) & x_{(1)} \leq \theta_0 \\
\exp \left( - \sum_{i=1}^{n} x_i + n\theta_0 \right) & x_{(1)} > \theta_0
\end{cases}
$$

$$
\lambda(x) = \begin{cases} 
1 & x_{(1)} \leq \theta_0 \\
\exp \left( -n(x_{(1)} - \theta_0) \right) & x_{(1)} > \theta_0
\end{cases}
$$

$$
\{ x : \lambda(x) \leq c \} = \{ x : \exp \left( -n(x_{(1)} - \theta_0) \right) \leq c \} = \{ x : x_{(1)} \geq \theta_0 - n^{-1} \log c \}
$$

In both examples, $\lambda(x)$ is a function of sufficient statistics, which is true in general.
Theorem 8.2.4.

If $T(X)$ is a sufficient statistic for $\theta$ and $\lambda_*(T)$ and $\lambda(X)$ are the likelihood ratios based on $T$ and $X$, respectively, then $\lambda_*(T(x)) = \lambda(x)$ for every $x \in \mathcal{X}$ (the sample space).

Proof.

From the factorization theorem, $L(\theta|x) = g_\theta(T(x))h(x)$ for some functions $g_\theta$ and $h$.

Then

$$
\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta|x)}{\sup_{\theta \in \Theta} L(\theta|x)} = \frac{\sup_{\theta \in \Theta_0} g_\theta(T(x))h(x)}{\sup_{\theta \in \Theta} g_\theta(T(x))h(x)} = \frac{\sup_{\theta \in \Theta_0} g_\theta(T(x))}{\sup_{\theta \in \Theta} g_\theta(T(x))}
$$

$$
= \frac{\sup_{\theta \in \Theta_0} L_*(\theta|T(x))}{\sup_{\theta \in \Theta} L_*(\theta|T(x))} = \lambda_*(T(x))
$$

where the 3rd equality holds because $h(x)$ does not depend on $\theta$ and the 4th equality holds because it can be shown that $g_\theta(x)$ is proportional to the pdf or pmf of $T(X)$. 
One way analysis of variance (ANOVA)

Consider independent data

\[ Y_{i1}, \ldots, Y_{in_i} \sim N(\mu_i, \sigma^2), \quad i = 1, \ldots, k \]

We want to test whether the populations have a common mean,

\[ H_0 : \mu_i = \mu \text{ for all } i \quad \text{vs} \quad H_1 : \mu_i \text{'s are not all equal} \]

We now derive the LRT for testing \( H_0 \).

Under \( H_0 \), \( Y_{ij} \)'s are iid \( N(\mu, \sigma^2) \), and the result about normal population tells us that the MLE of \( (\mu, \sigma^2) \) under \( H_0 \) is \( \hat{\theta}_0 = (\bar{Y}, \text{SST}/N) \), where

\[ \bar{Y} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} \quad \text{SST} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 \]

are respectively the sample mean of the total sample, the sample variance of the total sample multiplied by \( n - 1 \), and \( n = \sum_{i=1}^{k} n_i \) is the total sample size.

In general, the MLE of \( (\mu_1, \ldots, \mu_k, \sigma^2) \) is \( (\bar{Y}_1, \ldots, \bar{Y}_k, \text{SSW}/n) \), where

\[ \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{SSW} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \]
are respectively the sample mean of the \( i \)th sample (from population \( i \)) and the so-called within population (group) sum of squares. Then, the likelihood ratio based on \( X = (Y_{ij}, j = 1, ..., n_i, i = 1, ..., k) \) is

\[
\lambda(X) = \left( \frac{2\pi \frac{SST}{n}}{2\pi \frac{SSW}{n}} \right)^{-n/2} \frac{e^{-n/2}}{e^{-n/2}} = \left( \frac{SST}{SSW} \right)^{-n/2}
\]

To derive the cut-off value \( c \) for the LRT that rejects \( H_0 \) iff \( \lambda(X) < c \), we need to find the distribution of a monotone function of \( \lambda(X) \).

Note that

\[
SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})^2 = SSW + SSB
\]

where \( SSB \) is the sum of squares between population (groups). This identity is a special case of the so-called ANVOA decomposition. Then

\[
\lambda(X) = \left( 1 + \frac{(k-1)F}{n-k} \right)^{-n/2}, \quad F = \frac{SSB/(k-1)}{SSW/(n-k)}
\]

i.e., \( \lambda(X) \) is a strictly decreasing function of \( F \).
We may show that $F$ has an $F$-distribution with degrees of freedom $k - 1$ and $n - k$, and this $F$-distribution is central under $H_0$ so that we reject $H_0$ if $F >$ the upper $\alpha$ quantile of the central $F$-distribution.

But this result follows from the general result we will establish next. In applications, it is common to use the following ANOVA table.

### ANOVA table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>between</td>
<td>$k - 1$</td>
<td>SSB</td>
<td>MSB=SSB/($k - 1$)</td>
<td>MSB/MSW</td>
</tr>
<tr>
<td>within</td>
<td>$n - k$</td>
<td>SSW</td>
<td>MSW=SSW/($n - k$)</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>$n - 1$</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

df = degrees of freedom

---

### Example 11.2.12

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>between</td>
<td>3</td>
<td>995.90</td>
<td>331.97</td>
<td>26.09</td>
</tr>
<tr>
<td>within</td>
<td>15</td>
<td>190.83</td>
<td>12.72</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>18</td>
<td>1186.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypotheses concerning linear functions of $\beta$ in a linear model

Under the linear model with $Y \sim N(X\beta, \sigma^2 I_n)$, we consider

$$H_0 : L\beta = 0 \quad \text{versus} \quad H_1 : L\beta \neq 0,$$

where $L$ is an $s \times p$ matrix of rank $s \leq p$.

For the special case of one-way ANOVA with $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, $p = k$, $s = k - 1$ and

$$L = \begin{pmatrix}
1 & 0 & \cdots & 0 & -1 \\
0 & 1 & \cdots & 0 & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & -1
\end{pmatrix}$$

We consider the likelihood ratio test.

The likelihood function is

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{||Y - X\beta||^2}{2\sigma^2} \right\}$$

$$\leq \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{||Y - X\hat{\beta}||^2}{2\sigma^2} \right\}$$
since \( \| Y - X\beta \|^2 \geq \| Y - X\hat{\beta} \|^2 \) for any \( \beta \) and the LSE \( \hat{\beta} \).

Treating the right-hand side of this expression as a function of \( \sigma^2 \), it is easy to show that it has a maximum at \( \sigma^2 = \hat{\sigma}^2 = \| Y - X\hat{\beta} \|^2 / n \) and

\[
\sup_{\beta, \sigma^2} L(\beta, \sigma^2) = (2\pi \hat{\sigma}^2)^{-n/2} e^{-n/2}.
\]

Similarly, let \( \hat{\beta}_{H_0} \) be the LSE under \( H_0 \) and \( \hat{\sigma}^2_{H_0} = \| Y - X\hat{\beta}_{H_0} \|^2 / n \).

Then

\[
\sup_{L\beta = 0, \sigma^2} L(\beta, \sigma^2) = (2\pi \hat{\sigma}^2_{H_0})^{-n/2} e^{-n/2}.
\]

Thus, the LR is

\[
\lambda = \left( \frac{\hat{\sigma}^2}{\hat{\sigma}^2_{H_0}} \right)^{n/2} = \frac{\| Y - X\hat{\beta} \|^n}{\| Y - X\hat{\beta}_{H_0} \|^n}
\]

Next, we derive the distribution of a monotone function of \( \lambda \) under \( L\beta = 0 \) so that we can obtain a size \( \alpha \) test.

For the given \( L \), there exists a \((p - s) \times p\) matrix \( M \) such that \( G' = (M' \quad L') \) is nonsingular.

Let \( \gamma = G\beta \), whose last \( s \) components are 0’s when \( L\beta = 0 \).
Note that

\[ X\beta = XG^{-1}G\beta = (XG^{-1})\gamma = W\gamma, \quad W = XG^{-1} \]

Then the problem reduces to testing whether the last \( s \) components of \( \gamma \) are 0's with \( Y \sim N(W\gamma, \sigma^2 I_n) \), \( \hat{\gamma} = G\hat{\beta} \), and \( X\hat{\beta} = W\hat{\gamma} \).

Under \( H_0 \), \( X\hat{\beta}_{H_0} = W_1(W_1'W_1)^{-1}W_1'Y \), where \( W_1 \) is the first \( p - s \) columns of \( W \), i.e., \( W = (W_1 \quad W_2) \).

Then

\[ \| Y - X\hat{\beta} \|^2 = \| Y - W(W'W)^{-1}W'Y \|^2 = Y'(I_n - H_w)Y, \]

where \( H_w = W(W'W)^{-1}W' \), a projection matrix of rank \( p \), and

\[ \| Y - X\hat{\beta}_{H_0} \|^2 = \| Y - W_1(W_1'W_1)^{-1}W_1'Y \|^2 = Y'(I_n - H_{w1})Y, \]

where \( H_{w1} = W_1(W_1'W_1)^{-1}W_1' \), a projection matrix of rank \( p - s \).

Since

\[ Y'Y = Y'H_{w1}Y + Y'(H_w - H_{w1})Y + Y'(I_n - H_w)Y \]

and the sum of the ranks of \( H_{w1}, H_w - H_{w1}, \) and \( I_n - H_w \) is \( p - s + s + n - p = n \), by Cochran's theorem, \( Y'(H_w - H_{w1})Y \) and \( Y'(I_n - H_w)Y \) are independently chi-square distributed.
Hence

\[
F = \frac{Y'(H_w - H_{w1}) Y / s}{Y'(I_n - H_w) Y / (n - p)} = \frac{[\| Y - X\hat{\beta}_{H_0} \|^2 - \| Y - X\hat{\beta} \|^2] / s}{\| Y - X\hat{\beta} \|^2 / (n - p)}
\]

which is a decreasing function of the LR \( \lambda \), has the noncentral F-distribution with degrees of freedom \( s \) and \( n - p \) and the noncentrality parameter \( \delta = \gamma' W' (H_w - H_{w1}) W \gamma / \sigma^2 \).

Under \( H_0 \), \( \delta = 0 \) (exercise) and, hence, the LR test of size \( \alpha \) rejects \( H_0 \) when \( F > F_{s,n-p,\alpha} \), and the power of this test is related to the noncentral F-distribution.

Examples

For the one-way ANOVA with \( H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \), the result holds with \( s = k - 1 \) and \( p = k \).

Consider the two-way balanced ANOVA with \( H_0 : \alpha_1 = \cdots = \alpha_{a-1} = 0 \).

The statistic \( F \) can be calculated using the SSR under the two-way balanced ANOVA model and the SSR under the reduced model with \( \alpha_i \)'s equal to 0.
Alternatively, we can use the following decomposition:

\[
\sum_{i,j,k} (Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i,j,k} \left( Y_{ijk} - \bar{Y}_{ij.} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij} \right)^2
\]

\[
= bc \sum_i \hat{\alpha}_i^2 + ac \sum_j \hat{\beta}_j^2 + c \sum_{i,j} \hat{\gamma}_{ij}^2 + \sum_{i,j,k} \left( Y_{ijk} - \bar{Y}_{ij.} \right)^2
\]

\[
= SSA + SSB + SSAB + SSR
\]

The LR tests of size \( \alpha \) can be obtained using the following table.

**ANOVA table**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>( a - 1 )</td>
<td>SSA</td>
<td>MSA = ( \frac{SSA}{a-1} )</td>
<td>MSA/MSR</td>
</tr>
<tr>
<td>B</td>
<td>( b - 1 )</td>
<td>SSB</td>
<td>MSB = ( \frac{SSB}{b-1} )</td>
<td>MSB/MSR</td>
</tr>
<tr>
<td>AB</td>
<td>( (a - 1)(b - 1) )</td>
<td>SSAB</td>
<td>MSAB = ( \frac{SSAB}{(a-1)(b-1)} )</td>
<td>MSAB/MSR</td>
</tr>
<tr>
<td>Residual</td>
<td>( ab(c - 1) )</td>
<td>SSR</td>
<td>MSR = ( \frac{SSR}{ab(c-1)} )</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>( abc - 1 )</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( df = \) degrees of freedom