ON HYPOTHESIS TESTING

FADOUA BALABDAOUI

Spring 2018

Notes: David Markwalder

CONTENTS

Part 1. Introduction and some fundamentals
1. Posing the problem 1
2. The fundamental lemma on hypothesis testing 2
3. Composite hypothese for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ 6
3.1. Karlin-Rubin Theorem 6
4. P-Values 9
5. Brief look at multiple testing 11
5.1. Bonferroni’s correction 12

Part 2. Further methods for constructing tests
1. Likelihood Ratio Tests 12
2. Gaussian vectors and related distributions 13
2.1. Multivariate Gaussian distribution 13
2.2. Gamma-function 13
2.3. $\chi^2_k$: Chi-square distribution with $k$ degrees of freedom 13
2.4. Distribution of Student(t-) of $k$ degrees of freedom 13
2.5. F-distribution 13
3. Example for LRT 13
3.1. Example a 13
3.2. Cochran’s Theorem 15
3.3. Example b 16
3.4. Example c 19
3.5. Example d 20
4. F-tests and application in linear regression 20
4.1. Regression model 20
4.2. Linear Regression 21
4.3. Least Squares Estimator 21
4.4. Properties of the LSE 22
4.5. $\chi^2$- and F-tests for variable selection 24

Part 1. Introduction and some fundamentals

1. Posing the problem

Let $X : (\Omega, \mathcal{A}, P) \rightarrow (\chi, \mathcal{B})$ be a random variable, $(\Omega, \mathcal{A}, P)$ a probability space, $(\chi, \mathcal{B})$ a measurable space.

Result: $X$ induces the probability measure $P_X$ on $(\chi, \mathcal{B})$ given by $P_X(B) = P(X \in B)$ for all $B \in \mathcal{B}$.
Example: Suppose $X \sim N(\theta, 1)$ with $\theta \in \mathbb{R}$. Then
\[ P_X(B) = \int_B \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \theta)^2\right) dx, \quad \forall B \in \mathcal{B}. \]

We are going to assume that $P_X$ belongs to some parametric family, that is, that there exists some parameter space $\Theta$ such that $P_X \in \{P_\theta : \theta \in \Theta\}$. Here, for all $\theta \in \Theta$, $P_\theta$ is a probability measure on $(\chi, \mathcal{B})$. In the previous example, $\Theta = \mathbb{R}$.

**Example:** $X \sim \text{Pois}(\theta)$, $\theta \in (0, +\infty)$. Then
\[ P_X(B) = \sum_{x \in B} \exp(\theta x) x^\theta \bigg/ \theta^\theta \theta! \quad \forall \theta \in 2^\mathbb{N} \]
the ensemble of all subsets of $\mathbb{N}$.

**Problem:** Let $\Theta_0$ and $\Theta_1$ be two subsets of $\Theta$ such that $\Theta_0 \cap \Theta_1 = \emptyset$.

**Goal:** We want, based on observed realisation of $X_1$, be able to decide between $\Theta_0$ and $\Theta_1$. This is a testing problem which can be formalized as follows:
\[ H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1, \]
where $H_0$ denotes the null- and $H_1$ denotes the alternative hypothesis.

**Definition 1.1. **critical function** We call a critical function any function $\Phi$ such that $\Phi(x) \in [0, 1]$ for all $x \in \chi$.

**Definition 1.2. **test function** A test function is a critical function $\Phi$ such that for all $x \in \chi$ we either accept $H_0$ with probability $1 - \Phi(x)$ or we reject $H_0$ with probability $\Phi(x)$.

**Definition 1.3. **type-I error, power, type-II error**

(i) for $\theta \in \Theta_0$, the function $\theta \mapsto \mathbb{E}_\theta[\Phi(X)]$ is called Type-I error.

(ii) for $\theta \in \Theta_1$, the same function is called power (usually denoted by $\beta(\theta)$).

(iii) $1 - \beta(\theta)$ is called type-II error.

<table>
<thead>
<tr>
<th>Truth \ Decision</th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta_0$</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>$\Theta_1$</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

The goal is to find a test function $\Phi$ such that $\sup_{\theta \in \Theta_0} E_\theta(\Phi(X)) \leq \alpha$ for some given $\alpha \in (0, 1)$ $\beta(\theta)$ is maximal $\forall \theta \in \Theta_1$.

**Goal:** Find a function $\Phi$ such that Type-I error is controlled if and only if $\sup_{\theta \in \Theta_0} E_\theta[\Phi(x)] \leq \alpha$ (for some given $\alpha \in (0, 1)$).

The power of $\Phi$ is the largest among all other testing functions $\Phi^*(x)$ satisfying $\sup_{\theta \in \Theta_0} E_\theta[\Phi(x)] \leq \alpha$ if and only if for all $\theta \in \Theta_1, \beta(\theta) = E_\theta(\Phi(x)) \geq E_\theta(\Phi^*(x)) = \beta^*(\theta)$.

**Definition 1.4. **We say that $H_0$ or $H_1$ is

(i) simple if $\Theta_0 = \{\theta_0\}$ or $\Theta_1 = \{\theta_1\}$.

(ii) composite if $\text{card}(\Theta_0) > 1$ or $\text{card}(\Theta_1) > 1$.

**Example:** $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$

then we are testing a simple hypothesis against a simple hypothesis.

$H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta \geq \theta_1$

2. The fundamental lemma on hypothesis testing

**Definition 2.1. **UMP A test $\Phi$ is is said to be uniformly most powerful of level $\alpha$ (UMP of level $\alpha$) if $\sup_{\theta \in \Theta_0} E_\theta[\Phi(X)] \leq \alpha$ and for any other test $\Phi^*$ such that $\sup_{\theta \in \Theta_0} E_\theta[\Phi^*(X)] \leq \alpha$ we have
\[ E_\theta[\Phi^*(X)] \leq E_\theta[\Phi(X)] \]
for all $\theta \in \Theta_1$. 

Theorem 2.2. Neyman-Pearson-Lemma Let \( P_0 \) and \( P_1 \) be two probability measures on \((\chi, \mathcal{B})\) such that \( P_0 \) and \( P_1 \) admit densities \( p_0 \) and \( p_1 \) with respect to some \( \sigma \)-finite measure \( \mu \). Let \( \alpha \in (0, 1) \) and consider the problem \( H_0 : p = p_0 \) vs. \( H_1 : p = p_1 \).

(i) There exists \( k_\alpha \in (0, \infty) \) such that the test
\[
\Phi(x) := \begin{cases} 
1 & \text{if } p_1(x) > k_\alpha p_0(x) \\
0 & \text{if } p_1(x) < k_\alpha p_0(x) 
\end{cases}
\]
satisfies \( E_{p_0} [\Phi(x)] = \alpha \) and \( \Phi \) is UMP of level \( \alpha \) (existence).

(ii) If \( \Phi \) is a UMP test of level \( \alpha \) (for the same problem), then it must be given by \( \Phi \)-m.a.e. (uniqueness).

Lemma 2.3. Let \( f \) be some measurable function on \((\chi, \mathcal{B})\) such that \( f(x) > 0 \) for all \( x \in S \) (\( s \) is a set in \( \mathcal{B} \)). Also let \( \mu \) be some \( \sigma \)-finite measure on \((\chi, \mathcal{B})\). Then \( \int_S f d\mu = 0 \Rightarrow \mu(S) = 0 \).

Proof. Define \( S_n := \{ x \in S : f(x) \geq 1/n \} \), \( n > 0 \). By definition of \( S \) \((f(x) > 0 \) for all \( x \in S \)), we have \( S \subset \bigcup_{n>0} S_n \). But, using the properties of measures we see that \( \mu(S) = \sum_{n>0} \mu(S_n) \). But \( \mu(S_n) \leq n \int_{S_n} f d\mu \) because \( f \geq \frac{1}{n} \mu(S) \) which implies \( \int_{S_n} f d\mu \geq \frac{1}{n} \mu(S) \). So
\[
S_n \subset S \Rightarrow \int_{S_n} f d\mu \leq \frac{1}{n} \mu(S) = 0
\]
by assumption. We conclude that \( \mu(S) \leq 0 \) if and only if \( \mu(S) = 0 \). \(\square\)

Proof. We first show i) (existence) Consider the random variable \( Y = \frac{p_1(x)}{p_0(x)} \) which, under \( H_0 \) is almost surely defined and we have \( P_0(p_0(x) = 0) = \int_{\{p_0(x) = 0\}} \chi_x \). Let \( F_0 \) be the cdf of \( Y \) under \( H_0 : p = p_0 \) and let \( k_\alpha = \inf\{y : F_0(y) \geq 1 - \alpha\} \) be the \((1 - \alpha)\) quantile of \( F_0 \). Let us consider the following test function
\[
\Phi(x) := \begin{cases} 
1 & \text{if } \frac{p_1(x)}{p_0(x)} > k_\alpha \\
\gamma_\alpha & \text{if } \frac{p_1(x)}{p_0(x)} = k_\alpha \\
0 & \text{if } \frac{p_1(x)}{p_0(x)} < k_\alpha 
\end{cases}
\]
such that \( \gamma_\alpha \) satisfies \( E_{p_0} [\Phi(x)] = \alpha \). This means that
\[
1 \cdot P_{p_0} \left( \frac{p_1(x)}{p_0(x)} > k_\alpha \right) + \gamma_\alpha \cdot P_{p_0} \left( \frac{p_1(x)}{p_0(x)} = k_\alpha \right) + 0 \cdot P_{p_0} \left( \frac{p_1(x)}{p_0(x)} < k_\alpha \right) = \alpha
\]
or equivalently
\[
1 - F_0(k_\alpha) + \gamma_\alpha \left( F_0(k_\alpha) - F_0(k_\alpha^-) \right) = \alpha.
\]
Now define
\[
\gamma_\alpha := \begin{cases} 
\frac{\alpha - (1 - F_0(k_\alpha))}{F_0(k_\alpha^-) - F_0(k_\alpha^-)} & \text{if } F_0(k_\alpha) > F_0(k_\alpha^-) \\
0 & \text{if } F_0 \text{ is continuous in } k_\alpha.
\end{cases}
\]
Now we show that \( \Phi \) is UMP among all tests of level \( \alpha \). Take another test \( \Phi^* \) such that \( E_{p_0} [\Phi^*(x)] \leq \alpha \). The goal is to show that \( E_{p_1} [\Phi(x)] \geq E_{p_0} [\Phi^*(x)] \).
\[
\int_{\chi} \left( \Phi(x) - \Phi^*(x) \right) \left( \frac{p_1(x)}{p_0(x)} - k_\alpha \right) d\mu(x) =
\]
\[
= \int_{L} \left( \Phi(x) - \Phi^*(x) \right) \left( \frac{p_1(x)}{p_0(x)} - k_\alpha \right) d\mu(x) + \int_{M} \left( \Phi(x) - \Phi^*(x) \right) \left( \frac{p_1(x)}{p_0(x)} - k_\alpha \right) d\mu(x)
\]
\[
= \int_{L} \left( 1 - \Phi^*(x) \right) \left( \frac{p_1(x)}{p_0(x)} - k_\alpha \right) d\mu(x) + \int_{M} \left( -\Phi^*(x) \right) \left( \frac{p_1(x)}{p_0(x)} - k_\alpha \right) d\mu(x) \geq 0,
\]
where \( L := \{ x : p_1(x) > k_\alpha p_0(x) \} \) and \( M := \{ x : p_1(x) < k_\alpha p_0(x) \} \). Hence, \( \int_{\chi} \left( \Phi(x) - \Phi^*(x) \right) \left( \frac{p_1(x)}{p_0(x)} - k_\alpha \right) d\mu(x) \geq 0 \) and thus we have
\[
E_{p_1} [\Phi(x)] - E_{p_0} [\Phi^*(x)] = \left( \frac{k_\alpha \left( E_{p_0} [\Phi(x)] - E_{p_0} [\Phi^*(x)] \right) \right) = k_\alpha (\alpha - E_{p_0} [\Phi^*(x)]).
\]
Therefore \( E_{p_1} [\Phi(x)] \geq E_{p_0} [\Phi^*(x)] \).

We now show ii) (uniqueness). Take another test \( \Phi^* \) of level \( \alpha \) \((E_{p_0} [\Phi^*(x)] \leq \alpha)\) and such that \( \Phi^* \) is UMP among all
tests of level $\alpha$. Let us consider the following set $S = \{x \in \chi : \Phi^*(x) \neq \Phi(x)\} \cap \{x \in \chi : p_1(x) \neq k_\alpha p_0(x)\}$. We want to show that $\mu(S) = 0$. Assume $\mu(S) > 0$. Consider $f(x) = (\Phi(x) - \Phi^*(x))(p_1(x) - k_\alpha p_0(x))$, $x \in \chi$. Note that $f(x) > 0$ for all $x \in S$. Using lemma we conclude that $\int_S f(x) d\mu(x) > 0$. Now,

$$
\int f(x) d\mu(x) = \int_{S^c} f(x) d\mu(x) + \int_{S^c} f(x) d\mu(x)
$$

where $f(x) = 0$ on $S^c$. This implies that

$$
0 < \int f(x) d\mu(x) = \int (\Phi(x) - \Phi^*(x))(p_1(x) - k_\alpha p_0(x)) d\mu(x)
$$

$$
= (E_{p_1}[\Phi(x)] - E_{p_1}[\Phi^*(x)]) - k_\alpha \left(\alpha - E_{p_0}[\Phi^*(x)]\right)
$$

which means that $E_{p_1}[\Phi(x)] - E_{p_1}[\Phi^*(x)] > k_\alpha (\alpha - E_{p_0}[\Phi^*(x)]) \geq 0$ It follows that $E_{p_1}[\Phi(x)] > E_{p_1}[\Phi^*(x)]$ but this is impossible since by assumption $\Phi^*$ is UMP. We conclude that $\mu(S) = 0$ and that $\mu$–a.e.

$$
\Phi^*(x) = \begin{cases} 
1 & \text{if } \frac{p_1(x)}{p_0(x)} > k_\alpha \\
0 & \text{if } \frac{p_1(x)}{p_0(x)} < k_\alpha.
\end{cases}
$$

\[\square\]

**Corollary 2.4.** Let $\alpha \in (0, 1)$ and $\beta = E_{p_1}[\Phi(x)]$, the power of the Neyman-Pearson test of level $\alpha$. Then $\alpha \leq \beta$ (we say that $\Phi$ is unbiased).

**Proof.** Consider the constant test $\Phi^*(x) = \alpha$ for all $x \in \chi$. $\Phi^*$ is a test of level $\alpha$ and hence

$$
\beta = E_{p_1}[\Phi(x)] \geq E_{p_1}[\Phi^*(x)] = \alpha \Leftrightarrow \alpha \leq \beta.
$$

\[\square\]

**Remark:** We can even show that $\alpha < \beta (\Phi$ is strictly unbiased).

**Remark:** The arguments used to prove the Neyman-Pearson lemma can be used to show that for any pair $(k, \gamma) \in (0, \infty) \times [0, 1]$, the test

$$
\Phi(x) = \begin{cases} 
1 & \text{if } \frac{p_1(x)}{p_0(x)} > k \\
\gamma & \text{if } \frac{p_1(x)}{p_0(x)} = k \\
0 & \text{if } \frac{p_1(x)}{p_0(x)} < k
\end{cases}
$$

(2)

is UMP of level $E_{p_0}[\Phi(x)] = P_{p_0}\left(\frac{p_1(x)}{p_0(x)} > k\right) + \gamma P_{p_0}\left(\frac{p_1(x)}{p_0(x)} = k\right)$.

**Example:** (Quality control) We have a batch of items whose (unknown) proportion of defectiveness is $\theta \in (0, 1)$. To perform a quality control, $n$ items are sampled from this batch to check whether they are defective or not. We want to test $H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$ at some level $\alpha \in (0, 1)$. For $i \in \{1, \ldots, n\}$ define the random variable $X_i := \begin{cases} 
1 & \text{if the } i\text{-th sampled item is defective} \\
0 & \text{otherwise}.
\end{cases}$

We have a random sample $(X_1, \ldots, X_n)$ of iid $\text{Ber}(\theta)$, i.e. $\chi = \{0, 1\}^n = \{0, 1\} \times \cdots \times \{0, 1\}$. We want to apply the Neyman-Pearson lemma to this testing problem. The joint density of $(X_1, \ldots, X_n)$ is

$$
p_\theta(x_1, \ldots, x_n) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}
$$

$$
= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}.
$$

Under $H_0$ we have

$$
p_{\theta_0}(x_1, \ldots, x_n) = \theta_0^{\sum_{i=1}^n x_i}(1-\theta_0)^{n-\sum_{i=1}^n x_i}
$$

$$
= \left(\frac{\theta_0}{1-\theta_0}\right)^{\sum_{i=1}^n x_i}(1-\theta_0)^n,
$$

Under $H_1$ we have

$$
p_{\theta}(x_1, \ldots, x_n) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}
$$

$$
= \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i}(1-\theta)^n,
$$

Under $H_1$ we have
and under $H_1$ we have

$$p_{01}(x_1, \ldots, x_n) = \theta_1^{\sum_{i=1}^n x_i} (1 - \theta_1)^{n - \sum_{i=1}^n x_i} = \left( \frac{\theta_1}{1 - \theta_1} \right)^{\sum_{i=1}^n x_i} (1 - \theta_1)^n.$$  

By applying the Neyman-Pearson lemma we know that the test $\Phi$ given by

$$\Phi(x_1, \ldots, x_n) := \begin{cases} 1 & \text{if } \left( \frac{\theta_1}{1 - \theta_1} \right)^{\sum_{i=1}^n x_i} (1 - \theta_1)^n > k_\alpha \\ \gamma_\alpha & \text{if } \left( \frac{\theta_1}{1 - \theta_1} \right)^{\sum_{i=1}^n x_i} (1 - \theta_1)^n = k_\alpha \\ 0 & \text{if } \left( \frac{\theta_1}{1 - \theta_1} \right)^{\sum_{i=1}^n x_i} (1 - \theta_1)^n < k_\alpha. \end{cases}$$

Such that $\gamma_\alpha$ satisfies $E_{00}[\Phi(X_1, \ldots, X_n)] = \alpha$. Note that $\frac{\theta_1}{1 - \theta_1} > 1$ implies $\frac{\theta_1}{1 - \theta_1} > 1$ which means that the function $t \mapsto \left( \frac{\theta_1}{1 - \theta_1} \right)^n (1 - \theta_1)^n$ is strictly increasing and continuous. Then the test $\Phi$ can also be rewritten as

$$\Phi(x_1, \ldots, x_n) := \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i > t_\alpha \\ \gamma_\alpha & \text{if } \sum_{i=1}^n x_i = t_\alpha \\ 0 & \text{if } \sum_{i=1}^n x_i < t_\alpha \end{cases}$$

where $t_\alpha$ is the $(1 - \alpha)$-quantile of $\sum_{i=1}^n X_i$ under $H_0$ and $\gamma_\alpha$ satisfies $E_{00}[\Phi(x)] = \alpha$. Note that $\sum_{i=1}^n X_i \sim \text{Bin}(n, \theta_0)$ under $H_0$. Let $F_{00}$ be the cdf of Bin($n, \theta_0$):

$$F_{00}(y) := \begin{cases} 0 & \text{if } y < 0 \\ (1 - \theta_0)^y & \text{if } 0 \leq y < 1 \\ (1 - \theta_0)^y + n \theta_0 (1 - \theta_0)^{n-1} & \text{if } 1 \leq y < 2 \\ \vdots & \vdots \\ \sum_{j=0}^{n-1} \binom{n}{j} \theta_0^j (1 - \theta_0)^{n-j} & \text{if } n - 1 \leq y < n \\ 1 & \text{if } y \geq n. \end{cases}$$

$$\gamma_\alpha = \frac{F_{00}(k_\alpha) - (1 - \alpha)}{F_{00}(k_\alpha) - F_{00}(k_\alpha - 1)} = \frac{\sum_{j=0}^{k_\alpha} \binom{n}{j} \theta_0^j (1 - \theta_0)^{n-j} - (1 - \alpha)}{\binom{n}{k_\alpha} \theta_0^{k_\alpha} (1 - \theta_0)^{n-k_\alpha}}.$$  

Graphical illustration:

---

**A numerical illustration:** $\theta_0 = 0.2$ and $\theta_1 = 0.4$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$n = 10$</th>
<th>$n = 20$</th>
<th>$n = 30$</th>
<th>$n = 40$</th>
<th>$n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>0.01</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>17</td>
</tr>
</tbody>
</table>

Values of $t_\alpha$ as a function of $\alpha$ and $n$.

$H_0 : \theta = 0.2$ vs. $H_1 : \theta = 0.4$
<table>
<thead>
<tr>
<th>$n$</th>
<th>$n = 10$</th>
<th>$n = 20$</th>
<th>$n = 30$</th>
<th>$n = 40$</th>
<th>$n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.05$</td>
<td>$0.41$</td>
<td>$0.63$</td>
<td>$0.78$</td>
<td>$0.88$</td>
<td>$0.93$</td>
</tr>
<tr>
<td>$0.01$</td>
<td>$0.19$</td>
<td>$0.40$</td>
<td>$0.57$</td>
<td>$0.70$</td>
<td>$0.80$</td>
</tr>
</tbody>
</table>

Power of $\Phi$ as a function of $n$ and $\alpha$. $E_{\theta_0}[\Phi(X_1, \ldots, X_n)] = P_{\theta_0}\left(\sum_{i=1}^n X_i > t_0\right) + \gamma_\alpha P_{\theta_0}\left(\sum_{i=1}^n X_i = t_0\right)$.

3. COMPOSITE HYPOTHESES FOR TESTING $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$

3.1. Karlin-Rubin Theorem. We will start this section with two examples.

Example 1: (Number of e-mails) The total number of e-mails that I received over a period of two weeks is

1, 0, 10, 11, 7, 8, 2, 0, 3, 7, 9, 13, 6, 5, 0.

Let $X_i$ denote the number of daily e-mails received at day $i$, and denote by $\theta = E[X]$. Is it true that $\theta > 5$?

Example 2: (Airplane noise) The law requires that the noise caused by airplanes take-off should not exceed a certain threshold $\mu_0$. From a sample of size $n$ the noise intensity of airplanes was recorded. We want to test $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, where $\mu$ is the true expectation of noise intensity.

Definition 3.1. MLR Consider the parametric model $\{p_\theta : \theta \in \Theta\}$ and let $\Theta \subseteq \mathbb{R}$ be a parametric family of densities defined on $(\chi, \mathcal{B})$. This family is said to have a monotone likelihood ratio (MLR) if there exists a statistic $T$, and for any parameters $\theta_1 < \theta_2$ there exists a continuous and strictly increasing function $g$ such that $\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)} = g(T(x))$ for all $x \in \chi$ such that $\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)} \in (0, +\infty)$.

 Remark: Note that $g$ can depend on $\theta_1$ or $\theta_2$.

Example: (Quality Control with one sample) Let $X \sim \text{Bin}(n, \theta)$, $\theta \in \Theta = (0, 1)$. For $\theta_1 < \theta_2$, we have

$$\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)} = \frac{C_n^x \theta_2^n (1-\theta_2)^{n-x}}{C_n^x \theta_1^n (1-\theta_1)^{n-x}} = \left(\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right)^x \left(\frac{1-\theta_2}{1-\theta_1}\right)$$

for $x \in \chi = \{1, \ldots, n\}$. Put $T(x) = x$ and $g(t) = \left(\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right)^t \left(\frac{1-\theta_2}{1-\theta_1}\right)^n$. Note that $g(t)$ is continuous strictly increasing since $\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)} > 1$.

Example: (Airplane noise with one sample) Suppose $X \sim \mathcal{N}(\mu, \sigma^2_0)$, $\sigma^2_0$ known and $\mu \in \Theta = \mathbb{R}$. We know that $p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2_0}} \exp\left(-\frac{1}{2\sigma^2_0}(x-\mu)^2\right)$. Let $\mu_1 \leq \mu_2$:

$$\frac{p_{\mu_2}(x)}{p_{\mu_1}(x)} = \exp\left\{-\frac{1}{2\sigma^2_0}\left((x-\mu_2)^2 - (x-\mu_1)^2\right)\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2_0}\left(\mu_2^2 - 2\mu_2 x + x^2 - x^2 + 2x\mu_1 - \mu_1^2\right)\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2_0}\left(2x(\mu_2 - \mu_1) + \mu_2^2 - \mu_1^2\right)\right\}$$

$$= \exp\left\{\frac{x(\mu_2 - \mu_1)}{\sigma^2_0} - \frac{\mu_2^2 - \mu_1^2}{2\sigma^2_0}\right\}$$

Put $T(x) = x$ and $g(t) = \exp\left(\frac{t(\mu_2 - \mu_1)}{\sigma^2_0} - \frac{\mu_2^2 - \mu_1^2}{2\sigma^2_0}\right)$. Note that $g(t)$ is continuous and strictly increasing.

Theorem 3.2. Karlin-Rubin Consider the testing problem $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ and fix $\alpha \in (0, 1)$. Suppose that $\{p_\theta : \theta \in \Theta\}$ admits the MLR property and let us denote by $F_{\theta_0}$ the cdf of $T(x)$ under $\theta = \theta_0$. 


Let us now show a test \( \Phi \) for any test \( \Phi \) and level \( \alpha \). This in turn implies that the supremum is admitted at \( \theta \) (we say that \( \Phi \) is UMP of level \( \alpha \)). Furthermore, using the remark after the proof of the Neyman-Pearson lemma, we conclude that \( \Phi \) is the test you get for the hypothesis \( H' : \theta = \theta_1 \) versus \( H_1 : \theta > \theta_0 \). By the Neyman-Pearson lemma, we know that the test

\[
\Phi(x) := \begin{cases} 
1 & \text{if } T(x) > t_{\alpha} \\
\gamma_{\alpha} & \text{if } T(x) = t_{\alpha} \\
0 & \text{if } T(x) < t_{\alpha},
\end{cases}
\]

where \( t_{\alpha} \) is the \( (1 - \alpha) \)-quantile of \( F_{\theta_0} \) and \( \gamma_{\alpha} \) satisfies

\[
E_{\theta_0}[\Phi(X)] = P_{\theta_0}(T(X) > t_{\alpha}) + \gamma_{\alpha} P_{\theta_0}(T(X) = t_{\alpha}) + 0P_{\theta_0}(T(X) < t_{\alpha}) = \alpha
\]

Proof. i) and ii) Consider first the testing problem \( H : \theta = \theta_0 \) versus \( K : \theta = \theta_1 \) with \( \theta_1 > \theta_0 \). By the Neyman-Pearson lemma, we know that the test

\[
\Phi(x) := \begin{cases} 
1 & \text{if } \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} > k_{\alpha} \\
\gamma_{\alpha} & \text{if } \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = k_{\alpha} \\
0 & \text{if } \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} < k_{\alpha},
\end{cases}
\]

where \( k_{\alpha} \) is the \( (1 - \alpha) \)-quantile of \( \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} \) under \( \theta_0 \) and \( \gamma_{\alpha} \) is such that \( E_{\theta_0}[\Phi(X)] = \alpha \). Since \( \Phi \) does not involve \( \theta_1 \), we conclude that \( \Phi \) must be UMP of level \( \alpha \) for testing \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta > \theta_0 \).

Let us now show ii). Pick arbitrary \( \theta' \) and \( \theta'' \) such that \( \theta' < \theta'' \). The test \( \Phi \) is the test you get for the hypothesis \( H' : \theta = \theta' \) versus \( H'' : \theta = \theta'' \) by applying the Neyman-Pearson lemma and thus \( \frac{p_{\theta''}(x)}{p_{\theta'}(x)} = \tilde{g}(T(x)) \) where \( \tilde{g} \) is continuous and strictly increasing (and may depend \( \theta' \) and \( \theta'' \)). This implies that

\[
\Phi(x) := \begin{cases} 
1 & \text{if } \frac{p_{\theta''}(x)}{p_{\theta'}(x)} > k'_{\alpha} \\
\gamma_{\alpha} & \text{if } \frac{p_{\theta''}(x)}{p_{\theta'}(x)} = k'_{\alpha} \\
0 & \text{if } \frac{p_{\theta''}(x)}{p_{\theta'}(x)} < k'_{\alpha},
\end{cases}
\]

Furthermore, using the remark after the proof of the Neyman-Pearson lemma, we conclude that \( \Phi \) must be UMP of level \( \alpha'' = E_{\theta'[\Phi(X)]} \). Using Corollary 2.1, we have that

\[
\alpha' \leq E_{\theta''}[\Phi(X)] \Rightarrow E_{\theta''}[\Phi(X)] \leq E_{\theta'[\Phi(X)]}
\]

We say that \( \Phi \) is unbiased. Since \( \theta' \) and \( \theta'' \) were chosen arbitrarily it follows that \( \theta \mapsto E_{\theta}[\Phi(X)] \) is non-decreasing. This in turn implies that the supremum is admitted at \( \theta_0 \) i.e. \( \sup_{\theta \in \Theta_0} E_{\theta}[\Phi(X)] = E_{\theta_0}[\Phi(X)] = \alpha \) (recall that the level of a test \( \Phi \) for testing \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \in \Theta_1 \) is \( \sup_{\theta \in \Theta_0} E_{\theta}[\Phi(X)] \)). This concludes the proof that \( \Phi \) is UMP of level \( \alpha \) for testing \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta > \theta_0 \).

iv) Fix \( \theta < \theta_0 \). By the MLR property, we know that there exists a strictly increasing and continuous function \( g \) such that \( \frac{p_{\theta_0}(x)}{p_{\theta}(x)} = g(T(x)) \). Thus the Karlin-Rubin test can be also given by

\[
\Phi(x) := \begin{cases} 
1 & \text{if } \frac{p_{\theta_0}(x)}{p_{\theta}(x)} > k_{\alpha} \\
\gamma_{\alpha} & \text{if } \frac{p_{\theta_0}(x)}{p_{\theta}(x)} = k_{\alpha} \\
0 & \text{if } \frac{p_{\theta_0}(x)}{p_{\theta}(x)} < k_{\alpha},
\end{cases}
\]

where \( k_{\alpha} \) is linked to \( t_{\alpha} \) through \( k_{\alpha} = g(t_{\alpha}) \). Now

\[
\int \left( \Phi(x) - \Phi^*(x) \right) \left( p_{\theta_0}(x) - k_{\alpha} p_{\theta}(x) \right) d\mu(x) \geq 0
\]

for any test \( \Phi^* \). Thus, \( E_{\theta_0}(\Phi(X)) - E_{\theta_0}(\Phi^*(X)) \geq k_{\alpha} E_{\theta_0}(\Phi(X)) - E_{\theta_0}(\Phi^*(X)) \) and \( E_{\theta_0}(\Phi(X)) - E_{\theta_0}(\Phi^*(X)) = 0 \) if \( E_{\theta_0}(\Phi^*(X)) = 0 \). Thus \( E_{\theta_0}(\Phi(X)) \leq E_{\theta_0}(\Phi^*(X)) \).
Corollary 3.3. application to exponential families Suppose that \( p_\theta(x) = c(\theta)h(x)\exp(Q(\theta)T(x)) \) with \( \theta \in \Theta \subseteq \mathbb{R} \) (one dimensional parameter space). If \( \theta \mapsto Q(\theta) \) is continuous and strictly increasing, then \( \{p_\theta : \theta \in \Theta\} \) admits the MLR property.

We now go back to the introductory examples.

Example 1: (Number of e-mails) We want to test \( H_0 : \theta \leq 5 \) versus \( H_1 : \theta > 5 \). Here we assume that \( X_1, \ldots, X_n \) \( \text{iid} \sim \text{Pois}(\theta) \) with \( n = 15 \). Hence we have density

\[
 p_\theta(x) = \frac{e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \exp \left( \log(\theta) \sum_{i=1}^{n} x_i \right) = c(\theta)h(x_1, \ldots, x_n)\exp(Q(\theta)T(x_1, \ldots, x_n))
\]

with \( Q(\theta) = \log(\theta) \), \( \theta \in \Theta \) and \( T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \). Hence at a given level \( \alpha \)

\[
 \Phi(x) := \begin{cases} 
 1 & \text{if } \sum_{i=1}^{n} x_i > t_\alpha \\
 \gamma_\alpha & \text{if } \sum_{i=1}^{n} x_i = t_\alpha \\
 0 & \text{if } \sum_{i=1}^{n} x_i < t_\alpha,
\end{cases}
\]

with \( t_\alpha \) being the \((1-\alpha)\)-quantile of \( \sum_{i=1}^{n} x_i \) under \( \theta = \theta_0 = 5 \) and \( \gamma_\alpha \) such that \( E_{\theta_0}[\Phi(x)] = \alpha \), is UMP at level \( \alpha \). We know that if \( X_1, \ldots, X_n \) \( \text{iid} \sim \text{Pois}(\theta_0) \), then \( \sum_{i=1}^{n} X_i \) \( \text{iid} \sim \text{Pois}(n\theta_0) \). \( t_\alpha \) is the \((1-\alpha)\)-quantile of \( \text{Pois}(n\theta_0) \) \( n=15, \theta_0=5, \alpha=0.05 \sim \).

\[
 \gamma_\alpha = \frac{F_{\theta_0}(t_\alpha)-(1-\alpha)}{p_{\theta_0}(\sum_{i=1}^{n} X_i = t_\alpha)} = \frac{0.960076-0.95}{0.0102} \approx 0.98.
\]

\[
 \Phi(x_1, \ldots, x_{15}) := \begin{cases} 
 1 & \text{if } \sum_{i=1}^{15} x_i > 90 \\
 0.98 & \text{if } \sum_{i=1}^{15} x_i = 90 \\
 0 & \text{if } \sum_{i=1}^{15} x_i < 90,
\end{cases}
\]

We have that \( \sum_{i=1}^{15} X_i = 82 \) and thus we accept \( H_0 : \theta \leq 5 \).

Example 2: (Take-off noise) If we assume that the noise intensity follows \( \mathcal{N}(\mu, \sigma_0^2) \), \( \sigma_0 > 0 \) known, then

\[
 p_\mu(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left( -\frac{1}{2\sigma_0^2}(x_i - \mu)^2 \right)
\]

\[
 = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)
\]

\[
 = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp \left( -\frac{1}{2\sigma_0^2} \left( \sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2 \right) \right)
\]

\[
 = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp \left( -\frac{n\mu^2}{2\sigma_0^2} \right) \exp \left( -\frac{\sum_{i=1}^{n} x_i^2}{2\sigma_0^2} \right) \exp(Q(\mu)T(x_1, \ldots, x_n))
\]

with \( T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \), \( Q(\mu) = \frac{\mu}{\sigma_0^2} \) continuous and strictly increasing. A UMP test of level \( \alpha \) for testing \( H_0 : \mu \leq \mu_0 \) versus \( H_1 : \mu > \mu_0 \) is given by

\[
 \Phi(x_1, \ldots, x_n) := \begin{cases} 
 1 & \text{if } \sum_{i=1}^{n} x_i > t_\alpha \\
 0 & \text{if } \sum_{i=1}^{n} x_i \leq t_\alpha
\end{cases}
\]

with \( E_{\mu_0}[\Phi(X_1, \ldots, X_n)] = \alpha \) if and only if \( P_{\mu_0}\left( \sum_{i=1}^{n} X_i > t_\alpha \right) = \alpha \).
\[ P_{\mu_0} \left( \sum_{i=1}^n X_i > t_\alpha \right) = \alpha \iff P_{\mu_0} \left( \overline{X}_n > t_\alpha / n \right) = \alpha \]

\[ \iff P_{\mu_0} \left( \overline{X}_n - \mu_0 > t_\alpha / n - \mu_0 \right) = \alpha \]

\[ \iff P_{\mu_0} \left( \frac{\overline{X}_n - \mu_0}{\sqrt{\sigma_0^2/n}} > \frac{t_\alpha/n - \mu_0}{\sqrt{\sigma_0^2/n}} \right) = \alpha \]

\[ \iff P \left( Z > \frac{t_\alpha/n - \mu_0}{\sqrt{\sigma_0^2/n}} \right) = \alpha \]

where \( Z \sim N(0, 1) \). Hence \( \frac{\sqrt{n}(t_\alpha/n - \mu_0)}{\sigma_0} = \zeta_\alpha \) the \((1 - \alpha)\)-quantile of \( N(0, 1) \).

\[ \Phi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}(x - \mu_0)}{\sigma_0} > \zeta_\alpha \\ 0 & \text{otherwise.} \end{cases} \]

Now chose \( \alpha = 0.05 \) then (you can compute with software) \( \zeta_\alpha \approx 1.64 \). Let \( n = 100, \sigma_0 = 7 \) and \( \mu_0 = 78 \). Then, again using software, we compute \( \mu_0 + \frac{\sigma_0}{\sqrt{n}} \zeta_\alpha \approx 79.15 \). We observe \( \overline{X}_n = 82 > 79.15 \) and hence decide to reject \( H_0 \).

Remark:
As \( n \to \infty \), the power of \( \Phi \) increases to 1 for any fixed alternative. Indeed let \( \mu \in \Theta_1 = (\mu_0, +\infty) \)

\[ \beta(\mu) = E_{\mu}[\Phi(X_1, \ldots, X_n)] \]

\[ = P_{\mu}(\overline{X}_n > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \zeta_\alpha) \]

\[ = P_{\mu} \left( \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma_0} > \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \zeta_\alpha \right) \]

\[ = P \left( Z > \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \zeta_\alpha \right) \]

\[ = 1 - P \left( Z \leq \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \zeta_\alpha \right) \]

\[ = 1 - F_Z \left( \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \zeta_\alpha \right). \]

But since \( \lim_{n \to \infty} - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \zeta_\alpha = -\infty \) we conclude that \( \lim_{n \to \infty} 1 - F_Z \left( -\frac{\sqrt{n}(\mu - \mu_0)}{\sigma_0} + \zeta_\alpha \right) = 1 \). We say that the test \( \Phi \) is consistent.

4. P-VALUES

Suppose we have an observation \( \theta \) and want to make a decision whether \( \theta \in \Theta_0 \) or \( \theta \in \Theta_1 \). To do so we use a statistical procedure (a test) which we either accept or reject. Let us revisit Example 2 and suppose that we observed a mean \( \overline{X}_n = 100 \). This would not change our initial decision of rejecting \( H_0 \) but this somehow looks 'more convincing' or may seem like we have 'more' evidence against \( H_0 : \mu \leq \mu_0 \). This leads to the notion of p-values. Assume we are in a simple setting: \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \in \Theta_1 \) (which may be composite but \( \theta_0 \notin \Theta_1 \)). Consider a test function

\[ \Phi(x) = \begin{cases} 1 & \text{if } T(x) > t_\alpha \\ 0 & \text{otherwise,} \end{cases} \]

where \( t_\alpha \) denotes the \((1 - \alpha)\)-quantile of \( T(X) \) under \( H_0 : \theta = \theta_0 \). Assume that \( F_{\theta_0} \), the cdf of \( T(X) \) under \( \theta = \theta_0 \), is continuous and strictly increasing, that is bijective.

**Definition 4.1. p-value** Let \( \mathcal{R}_\alpha = \{ x' \in X : T(x') > t_\alpha \} \) be a rejection region for some fixed \( \alpha \). We define the p-value of an observation \( x \in X \) with respect to \( \Phi \) by \( p_\Phi(x) = \inf \{ \alpha : x \in \mathcal{R}_\alpha \} \).

**Lemma 4.2.** For the test \( \Phi \) given above, it holds that \( p_\Phi(x) = P_{\theta_0}(T(X) \geq T(x)) \).
Proof. Recall that \( \Phi(x) = \begin{cases} 1 & \text{if } T(x) \geq t_x, \\ 0 & \text{otherwise} \end{cases} \) with \( t_x = F_{\theta_0}^{-1}(1 - \alpha) \) (we have assumed that \( F_{\theta_0} \) is bijective).

\[
p_{\Phi}(x) = \inf\{\alpha : x \in \mathcal{R}_{\alpha}\}
= \inf\{\alpha : T(x) > F_{\theta_0}^{-1}(1 - \alpha)\}
= \inf\{\alpha : F_{\theta_0}(T(x)) > (1 - \alpha)\}
= \inf\{\alpha > 1 - F_{\theta_0}(T(x))\}
= \inf\{1 - F_{\theta_0}(T(x)) \}, \infty\)
= 1 - F_{\theta_0}(T(x))
= P_{\theta_0}(T(X) > T(x))
\]

whereas the last equality holds because \( F_{\theta_0} \) is the cdf of \( T(X) \) under \( \theta = \theta_0 \). \( \square \)

Lemma 4.3. \( p_{\Phi}(X) \sim \mathcal{U}([0, 1]) \) under \( H_0 : \theta = \theta_0 \).

Proof. We know that \( p_{\Phi}(X) = 1 - F_{\theta_0}(T(X)) \). Recall that if \( Y \) is some random variable with cdf equal to \( F \), and \( F \) is bijective, then \( U = F(Y) \sim \mathcal{U}([0, 1]) \). Indeed, since \( F(Y) \leq u \) if and only if \( Y \leq F^{-1}(u) \), we see that the cdf of \( U \) is

\[
P(U \leq u) = \begin{cases} 0 & \text{if } u < 0 \\ u & \text{if } 0 \leq u < 1, \text{ because } u = F(F^{-1}(u)) = P(Y \leq F^{-1}(u)) \text{ and thus } F(Y) \sim \mathcal{U}([0, 1]) \text{. Thus } F_{\theta_0}(T(X)) \sim \mathcal{U}([0, 1]) \text{ and therefore } 1 - F_{\theta_0}(T(X)) \sim \mathcal{U}([0, 1]). \end{cases}
\]

Recall that we have considered a simple setting. P-values can also be defined through the following definition.

Definition 4.4. proper p-value Consider testing \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \in \Theta_1 \) such that \( \Theta_0 \cap \Theta_1 = \emptyset \). A p-value \( p(X) \) is said to be valid (or proper) if for all \( \theta \in \Theta_0 \) and for all \( t \in [0, 1] \) we have \( P_{\theta}(p(X) \leq t) \leq t \). This means that \( p(X) \) is a valid p-value if it is stochastically larger than \( U \sim \mathcal{U}([0, 1]) \) under any \( \theta \in \Theta_0 \).

Remark: Note that Definition (in the simple setting) gives a p-value that is stochastically equal to \( U \sim \mathcal{U}([0, 1]) \).

Example: Let \( T \) be some statistic used for testing \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \in \Theta_1 \). Define \( p(x) = \sup_{\theta \in \Theta_0} P_{\theta}(T(X) \geq T(x)) \). We want to check that this defines a valid p-value. For that, we will need the following result.

Lemma 4.5. Let \( Z \) be any random variable with distribution function \( F \) (not necessarily continuous or strictly increasing). Then \( U = F(Z) \) satisfies \( P(U \leq u) \leq u \) for all \( u \in [0, 1] \).

Proof. We either have

\[
F(\zeta) \leq u \Leftrightarrow \zeta \leq \zeta_u
\]
or

\[
F(\zeta) \leq u \Leftrightarrow \zeta < \zeta_u.
\]

\[
P(F(Z) \leq u) = \begin{cases} P(Z \leq \zeta_u) & \text{if } F(\zeta_u) = u \\ P(Z < \zeta_u) & \text{if } F(\zeta_u) > u \end{cases} = \begin{cases} F(\zeta_u) = u \\ F(\zeta_u^-) \leq u \end{cases}
\]

In any case we arrive at \( P(F(Z) \leq u) = P(U \leq u) \leq u \). \( \square \)
Remark: This is saying for any distribution function $F$, $F(x)$ is stochastically larger than $U \sim \mathcal{U}([0, 1])$ with $Z \sim F$. Now let us return to $p(x) = \sup_{\theta \in \Theta_0} P_\theta(T(X) \geq T(x))$. We will check that this defines a valid p-value.

Proof. Fix $\theta \in \Theta_0$ and denote by $F_\theta$ the cdf of $-T(X)$. Define

$$p_\theta(x) = P_\theta(T(X) \geq T(x))$$

$$= P_\theta(-T(X) \leq -T(x)) = F_\theta(-T(x)).$$

Using Lemma we know that $p_\theta(X)$ is stochastically larger than $\mathcal{U}([0, 1])$.

For $\tilde{\theta} \in \Theta_0$:

$$P_{\tilde{\theta}}(p(X) \leq t) = P_{\tilde{\theta}}(\sup_{\theta \in \Theta_0} F_\theta(-T(X)) \leq t)$$

$$= P_{\tilde{\theta}}(\forall \theta \in \Theta_0, F_\theta(-T(X)) \leq t)$$

$$\leq P_{\tilde{\theta}}(F_{\tilde{\theta}}(-T(X)) \leq t)$$

$$= P_{\tilde{\theta}}(p_{\tilde{\theta}}(X) \leq t) \leq t.$$

In conclusion: $\forall t \in [0, 1], \forall \tilde{\theta} \in \Theta_0$: $P_{\tilde{\theta}}(p(X) \leq t) \leq t \Leftrightarrow \sup_{\theta \in \Theta_0} P_\theta(p(X) \leq t) \leq t$ which means that $p(X)$ is indeed a valid p-value. \qed

What is the link between a valid p-value and testing? Given any valid p-value, we can construct the following test $\Phi$ at a given level $\alpha$: $\Phi(x) = 1$ if and only if $p(x) \leq \alpha$.

Type-I error $\sup_{\theta \in \Theta_0} E_\theta[\Phi(x)] = \sup_{\theta \in \Theta_0} P_\theta(\Phi(x) = 1) = \sup_{\theta \in \Theta_0} P_\theta(p(x) \leq \alpha) \leq \alpha$.

5. Brief look at multiple testing

Consider multiple hypothesis that we want to test at the same time. Call these (null) hypotheses $H_0^{(1)}, H_0^{(2)}, \ldots, H_0^{(m)}$ for some integer $m \geq 2$. Suppose for all $i \in \{1, 2, \ldots, m\}$ we have a test $\Phi_i$ for testing $H_0^{(i)}$ versus $H_1^{(i)}$ (some alternative). Consider the combined test $\Phi$ which rejects/accepts $H_0^{(i)}$ if $\Phi_i$ does. Let us suppose $\Phi_i$ has level $\alpha$ and that these tests are independent.

$$H_0 = H_0^{(1)} \cap H_0^{(2)} \cap \ldots \cap H_0^{(m)}$$

The Type-I error of

$$\Phi = P_{H_0}(\text{rejecting at least one } H_0^{(i)} \text{ for some } i \in \{1, \ldots, m\})$$

$$= 1 - P_{H_0}(\text{accepting } H_0^{(1)} \text{ and } H_0^{(2)} \text{ and } \ldots \text{ and } H_0^{(m)})$$

$$= 1 - \prod_{i=1}^{m} P_{H_0}(\text{accepting } H_0)$$

$$= 1 - \prod_{i=1}^{m} P_{H_0}(\text{accepts } H_0^{(i)})$$

$$= 1 - \prod_{i=1}^{m} P_{H_0}(\Phi_i \text{accepts } H_0^{(i)})$$

$$= 1 - (1 - \alpha)^m$$

Numerical illustration:

$$m = 10 \quad \alpha = 0.05 \quad \text{Type-I error} = 0.4$$

$$m = 50 \quad \alpha = 0.01 \quad \text{Type-I error} = 0.39$$

This means that we need to be more strict when choosing the levels of the individual tests.
5.1. **Bonferroni’s correction.** gives a solution to this problem. Here we are not going to assume that tests \( \Phi_i \) are independent.

\[
P_{H_0} \left( \text{rejecting at least } H_0^{(i)} \text{ for some } i \in \{1, \ldots, m\} \right) = P_{H_0} \left( \exists i \in \{1, \ldots, m\} : \Phi \text{ rejects } H_0^{(i)} \right)
\]

\[
= P_{H_0} \left( \bigcup_{i=1}^{m} \{ \Phi \text{ rejects } H_0^{(i)} \} \right)
\]

\[
\leq \sum_{i=1}^{m} P_{H_0} \left( \Phi \text{ rejects } H_0^{(i)} \right)
\]

\[
= \sum_{i=1}^{m} P_{H_0} \left( \Phi_i \text{ rejects } H_0^{(i)} \right)
\]

\[
= \sum_{i=1}^{m} P_{H_0}^{(i)} \left( \Phi \text{ rejects } H_0^{(i)} \right)
\]

If we chose the level of each test \( \Phi_i \) to be \( \frac{\alpha}{m} \), then the Type-I error of \( \Phi \leq m \frac{\alpha}{m} = \alpha \). Alternatively, we can require in this correction to have \( \alpha_i \) (the level of \( \Phi_i \)) satisfy \( \sum_{i=1}^{m} \alpha_i \leq \alpha \) (this will imply that the Type-I error of \( \Phi \leq \sum_{i=1}^{m} \alpha_i \leq \alpha \)).

### Part 2. Further methods for constructing tests

1. **Likelihood Ratio Tests**

**Definition 1.1. likelihood** Let \( X_1, \ldots, X_n \) be iid random variables admitting a density assumed to belong to the parametric family \( \{ p_\theta, \theta \in \Theta \} \)

- We call likelihood the function

\[
\Theta \to [0, \infty)
\]

\[
\theta \mapsto L_n(\theta) = \prod_{i=1}^{n} p_\theta(X_i)
\]

- We call log-likelihood the function

\[
\Theta \to \mathbb{R}
\]

\[
\theta \mapsto l_n(\theta) = \log \left( L_n(\theta) \right)
\]

**Definition 1.2. MLE** The maximum likelihood estimator (MLE) is any \( \hat{\theta}_n \) satisfying \( L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta) \) and since the logarithm is continuous and increasing \( l_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} l_n(\theta) \)

Remarks:
- The MLE does not have to exist.
- If the MLE exists it is not necessarily unique.
- For any subset \( \Theta' \subset \Theta \) we can define the restricted MLE which maximises \( \theta \mapsto L_n(\theta) \) (or \( \theta \mapsto l_n(\theta) \)) over \( \Theta' \).

**Definition 1.3. likelihood ratio statistic** Let \( \Theta_0 \) and \( \Theta_1 \) be two subsets of \( \Theta \) such that \( \Theta_0 \cap \Theta_1 = \emptyset \) (\( \Theta_0 \cup \Theta_1 = \Theta \)) and consider the testing problem \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \in \Theta_1 \). The likelihood ratio statistic is defined as \( \Lambda_n = \frac{\sup_{\Theta_0} L_n(\theta)}{\sup_{\Theta_1} L_n(\theta)} \).

**Definition 1.4. LRT** The likelihood ratio test for a given level \( \alpha \) is given by

\[
\Phi(X_1, \ldots, X_n) = \begin{cases} 
1 & \text{if } \Lambda_n > \lambda_\alpha \\
\gamma_\alpha & \text{if } \Lambda_n = \lambda_\alpha \\
0 & \text{if } \Lambda_n < \lambda_\alpha
\end{cases}
\]

where \( \gamma_\alpha \) and \( \lambda_\alpha \) are such that \( \sup_{\theta \in \Theta} E_\theta [\Phi(X_1, \ldots, X_n)] \leq \alpha \).

Remark: The idea behind the definition of LRT is to reject \( H_0 : \theta \in \Theta_0 \) when \( \frac{\sup_{\Theta_0} L_n(\theta)}{\sup_{\Theta_1} L_n(\theta)} \) is large. (see exercise)
2. Gaussian vectors and related distributions

2.1. Multivariate Gaussian distribution.

- Let \( X = (X_1, \ldots, X_d) \in \mathbb{R}^d \). We say that \( X \) is Gaussian if any linear combination of components, \( X_j, 1 \leq j \leq d \), has a Gaussian distribution: For all \( a_i \in \mathbb{R} \) for \( j \in \{1, \ldots, d\} \), \( \sum_{i=1}^{d} a_i X_i \) is a normal random variable.
- Two Gaussian vectors \( X = (X_1, \ldots, X_d) \) and \( Y = (Y_1, \ldots, Y_m) \) are independent if and only if \( \text{Cov}(X, Y) = 0 \) for all \( (i, j) \in \{1, \ldots, d\} \times \{1, \ldots, m\} \).
- If \( X \sim \mathcal{N}(\mu, \Sigma) \) with \( \mu \in \mathbb{R}^d \) and \( \Sigma \in \text{Mat}(\mathbb{R}^d \times \mathbb{R}^d) \) then for any matrix \( A \in \mathbb{R}^{mxd} \) \((m \geq 1)\) we have \( AX \sim \mathcal{N}(A\mu, A\Sigma A^\top) \).
- If \( X \sim \mathcal{N}(\mu, \Sigma) \) and \( \Sigma \) is invertible, then \( X \) admits density \( f_x(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right) \).

2.2. Gamma function. The gamma function is defined for all complex numbers except the non-positive integers. For complex numbers with a positive real part, it is defined via a convergent improper integral \( \Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx \). Note that if \( n \in \mathbb{Z}_{>0} \) then \( \Gamma(n) = (n-1)! \), \( \Gamma(1) = 1 \) and \( n\Gamma(n) = \Gamma(n+1) \).

2.3. \( \chi^2_k \): Chi-square distribution with \( k \) degrees of freedom. We say that \( Y \sim \chi^2_k \) if we can find \( X = (X_1, \ldots, X_k) \sim \mathcal{N}(0, \mathbb{I}_k) \) such that \( Y = \sum_{j=1}^k X_j^2 = ||X||_2^2 \) (the square of the euclidean norm of \( X \)). \( Y \) admits a density

\[
 f_Y(y) = \frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2 - 1} \exp(-y/2) 1_{y \geq 0}.
\]  

We recognize that \( Y \sim \text{Gamma} \left( \frac{k}{2}, \frac{1}{2} \right) \). Moreover if \( X \sim \mathcal{N}(\mu, \Sigma) \) and \( \Sigma \) is invertible then \( (x - \mu)^\top \Sigma^{-1} (x - \mu) \sim \chi^2_k \) (see exercise).

2.4. Distribution of Student(t-) of \( k \) degrees of freedom. We say that \( T \) follows a \( t \)-distribution with \( k \) degrees of freedom if we can find independent random variables \( X \) and \( Y \) with \( X \sim \mathcal{N}(0,1) \) and \( Y \sim \chi^2_k \) such that \( T = \frac{X}{\sqrt{Y/k}} \). We write \( T \sim T_k \). \( T \) admits density given by

\[
 f_T(t) = \frac{\Gamma \left( \frac{k+1}{2} \right)}{\sqrt{k\pi} \Gamma \left( \frac{k}{2} \right) \left( 1 + \frac{t^2}{k} \right)^{(k+1)/2}}, \quad t \in \mathbb{R}.
\]  

Note that \( T_{(1)} \) is the Cauchy distribution.

2.5. F-distribution. We say that \( Y \) admits an F-distribution with \( (p, q) \) degrees of freedom if we can find two random variables \( U \) and \( V \) such that \( U \) and \( V \) are independent, \( U \sim \chi^2_{p} \), \( V \sim \chi^2_{q} \) and \( Y \sim \frac{U/p}{V/q} \). We will write \( Y \sim F_{p,q} \). \( Y \) admits density given by

\[
 f_Y(y) = \frac{\Gamma \left( \frac{p+q}{2} \right)}{\Gamma(p/2)\Gamma(q/2) y^{p/2} (q/y)^{q/2} (p+q/(p+q)y)^{-(p+q)/2} 1_{y > 0}}.
\]  

3. Example for LRT

3.1. Example a. Let \( X_1, \ldots, X_n \) \( \sim \mathcal{N}(\theta, \sigma^2_0) \), where \( \theta \in \mathbb{R} \) and \( \sigma^2_0 > 0 \) is known. We want to test \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \).

Hence we have \( \Theta_0 = \{\theta_0\} \) (a simple hypothesis) and \( \Theta_1 = \mathbb{R} \setminus \{\theta_0\} \) (a composite hypothesis) such as \( \Theta = \Theta_0 \cup \Theta_1 = \mathbb{R} \). Recall that \( L_n = \sup_{\Theta_0} \frac{L_n(\theta)}{L_n(\theta_0)} = \sup_{\Theta_0} \frac{L_n(\theta)}{L_n(\theta_0)} \).

\[
 L_n(\theta) = \prod_{i=1}^{n} p_\theta(X_i)
 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2_0}} \exp \left( -\frac{1}{2\sigma^2_0} (X_i - \theta)^2 \right)
 = \frac{1}{(2\pi\sigma^2_0)^{n/2}} \exp \left( -\frac{1}{2\sigma^2_0} \sum_{i=1}^{n} (X_i - \theta)^2 \right).
\]

\[
 l_n(\theta) = \log(L_n(\theta)) = \text{constant} - \frac{1}{2\sigma^2_0} \sum_{i=1}^{n} (X_i - \theta)^2
\]
We want to show that $\text{argmax}_{\theta \in \mathbb{R}} L_n(\theta) = \overline{X}_n$. Our goal is to maximize $\theta \mapsto \exp \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \theta)^2 \right)$ over $\mathbb{R}$ or equivalently maximize $-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \theta)^2$ over $\mathbb{R}$.

\[
\frac{d}{d\theta} \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \theta)^2 \right) = -2 \sum_{i=1}^{n} (X_i - \theta) = 0 \iff \theta = \overline{X}_n
\]

and

\[
\frac{d^2}{d\theta^2} \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \theta)^2 \right) = 2n > 0
\]

which means that the function is convex on $\mathbb{R}$ and hence $\overline{X}_n$ gives the global maximum of $L_n$.

\[
\Lambda_n = \frac{L_n(\overline{X}_n)}{L_n(\theta_0)} = \frac{\exp \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \right)}{\exp \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \theta_0)^2 \right)} = \frac{\exp \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 + \frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \theta_0)^2 \right)}
\]

Recall that the event $\{\Lambda_n > \lambda_\alpha\}$ happens with probability equal to zero and hence the LRT is given by $\Phi(X_1, \ldots, X_n) = \begin{cases} 
1 & \text{if } \Lambda_n > \lambda_\alpha \text{ almost surely and we are going to find } \lambda_\alpha \text{ such that } E_{\theta_0}(\Phi(X_1, \ldots, X_n)) = \alpha. \text{ Note that} \\
0 & \text{if } \Lambda_n \leq \lambda_\alpha
\end{cases}$

$\Lambda_n$ is 'large' $\iff$ $\sum_{i=1}^{n} (X_i - \theta_0)^2 - \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ is 'large'

$\iff$ $\sum_{i=1}^{n} (X_i - \overline{X}_n + \overline{X}_n - \theta_0)^2 - \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ is 'large'

$\iff$ $\sum_{i=1}^{n} (X_i - \overline{X}_n)^2 + 2 \left( \sum_{i=1}^{n} (X_i - \overline{X}_n) \cdot (\overline{X}_n - \theta_0) + n(\overline{X}_n - \theta_0)^2 - \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \right)$ is 'large'

$\iff$ $n(\overline{X}_n - \theta_0)^2$ is 'large'

$\iff$ $\frac{n(\overline{X}_n - \theta_0)^2}{\sigma_0^2}$ is 'large'

$\iff$ $\frac{\sqrt{n} |\overline{X}_n - \theta_0|}{\sigma_0}$ is 'large'

$\Phi(X_1, \ldots, X_n) = \begin{cases} 
1 & \text{if } \frac{\sqrt{n} |\overline{X}_n - \theta_0|}{\sigma_0} > q_\alpha \text{ such that } E_{\theta_0}(\Phi(X_1, \ldots, X_n)) = P_{\theta_0} \left( \frac{\sqrt{n} |\overline{X}_n - \theta_0|}{\sigma_0} > q_\alpha \right) = \alpha. \text{ We need to determine} \\
0 & \text{otherwise}
\end{cases}$

the quantile $q_\alpha$. Recall $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta_0, \sigma_0^2)$ under $H_0$ which means that $\overline{X}_n \sim \mathcal{N}(\theta_0, \sigma_0^2/n)$ $\iff$ $\frac{\sqrt{n} |\overline{X}_n - \theta_0|}{\sigma_0} \overset{d}{=} Z \sim \mathcal{N}(0, 1)$.

$P_{\theta_0} \left( \frac{\sqrt{n} |\overline{X}_n - \theta_0|}{\sigma_0} > q_\alpha \right) = P(|Z| > q_\alpha)$

$= P(Z > q_\alpha) + P(Z < -q_\alpha)$

$= P(Z > q_\alpha) + P(-Z > q_\alpha)$

$= 2P(Z > q_\alpha)$
by symmetry around zero of the Z distribution. Hence,

\[ a = P_\alpha (\Phi \text{ rejects } H_0) = 2P(Z > q_\alpha) = P(Z > q_\alpha) = \alpha/2 \]

\[ \Rightarrow F_2(q_\alpha) = 1 - \alpha/2 \]

therefore \( \Phi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}X_{n-60}}{\sigma_0} > \xi_{1-\alpha/2} \\ 0 & \text{otherwise} \end{cases} \) where \( \xi_{1-\alpha/2} = q_\alpha = (1 - \alpha/2)\)-quantile of \( N(0, 1) \) and \( F_2(\xi) = \int_{-\infty}^\xi \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx. \)

3.2. Cochran’s Theorem.

**Theorem 3.1.** Let \((X_1, \ldots, X_d) = X \sim N_d(0, \mathbb{1})\) be a Gaussian vector. Let \(A_1, \ldots, A_J\) be \(d \times d\) matrices such that \(\sum_{i=1}^J \text{rank}(A_i) \leq d\) and for all \(i \in \{1, \ldots, J\}\)

(i) \(A_i\) is symmetric and \(A_i^2 = A_i\).

(ii) \(A_iA_j = A_jA_i = 0\) for all \(i \neq j\).

Then,

(i) \(A_iX \sim N(0, A_i)\) for all \(i \in \{1, \ldots, J\}\) and \(A_1X, \ldots, A_JX\) are mutually independent.

(ii) The random variables \(\|A_iX\|^2 \sim \chi^2_{\text{rank}(A_i)}\) and they are mutually independent.

**Proof.** i) We know that \(X \sim N(\mu, \Sigma)\) implies \(AX \sim N(A\mu, A\Sigma A^\top)\). Thus \(A_iX \sim N(0, A_iA_i^\top) = N(0, A_i)\). Then, showing mutual independence of \(A_1X, \ldots, A_JX\) is equivalent to showing \(\text{Cov}(A_iX, A_jX) = 0\) for all \(i \neq j\). Let \(E[X] = \mu\) and recall that

\[
\text{Cov}(AX, BX) = E[A(X - \mu)(B(X - \mu))^\top]
\]

\[
= E[A(X - \mu)(X - \mu)^\top B^\top]
\]

\[
= AE[(X - \mu)(X - \mu)^\top] B^\top
\]

\[
= A\Sigma B^\top.
\]

Hence in our case for \(i \neq j \in \{1, \ldots, J\}\) we have

\[
\text{Cov}(A_iX, A_jX) = A_i\mathbb{I}A_j^\top
\]

\[
= A_iA_j
\]

\[
= 0
\]

by assumption.

ii) \(A_1X, \ldots, A_JX\) mutually independent implies \(f(A_1X), \ldots, f(A_JX)\) mutually independent for some measurable function \(f\). In particular, this is true for \(f(\alpha) = \|\alpha\|^2 (\alpha \in \mathbb{R}^d)\) continuous on \(\mathbb{R}^d\) and hence measurable. We now show that \(\|A_iX\|^2 \sim \chi^2_{\text{rank}(A_i)}\). \(A_i\) is symmetric. We can orthogonalize \(A_i\) in an orthonormal basis. There exists an orthogonal matrix \(P\) so that we can decompose \(A_i = P\Lambda_i P^\top\) where \(\Lambda_i = \text{diag}(\lambda_1, \ldots, \lambda_d)\) and \(\lambda_1, \ldots, \lambda_d\) denote the eigenvalues of \(A_i\).

Using the assumption \(A_i^2 = A_i\), we conclude that \(A_1, \ldots, A_d \in \{0, 1\}\). Further we can decompose \(A_i^2\) in the following
way

\[
A_i^2 = P^T \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 \\ 0 & \ldots & 0 & \lambda_d \end{pmatrix} P^T \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 \\ 0 & \ldots & 0 & \lambda_d \end{pmatrix} P
\]

which means that \( \lambda_i^2 = \lambda_i \) for all \( i \in \{1, \ldots, d\} \) and hence there are only two solutions. We can also write \( A_i = P^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} P \). Then \( \|A_iX\|_2 = (A_iX)^T A_iX \)

\[
= X^T A_i^2 X \\
= X^T P^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} P X \\
= (PX)^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} P X \\
= Y^T \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} Y \\
= \sum_{j=1}^{\text{rank}(A)} Y_j^2.
\]

On the other hand, \( Y = PX \sim N(0, P^T P) \). Hence \( \|A_iX\|_2^2 \) is the norm of a squared vector \( \sim N(0, \mathbb{I}_{\text{rank}(A_i)}) \); in other words \( Y_1, \ldots, Y_{\text{rank}(A)} \) are iid \( \sim N(0, 1) \).

3.3. Example b. Let \( X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2) \) with \( \theta \in \mathbb{R} \) and \( \sigma \in (0, \infty) \) both unknown. Here \( \sigma \) is acting as a nuisance parameter. We want to test

\[
H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0
\]

whereas \( \Theta_0 = \{(\theta_0, \sigma) : \sigma \in (0, \infty)\} = [\theta_0] \times (0, \infty) \) and \( \Theta = \{(\theta, \sigma) : \theta \in \mathbb{R} \text{ and } \sigma \in (0, \infty)\} = \mathbb{R} \times (0, \infty) \). Since \( \sigma \) is unknown, we have

\[
\Lambda_n = \sup_{\theta \in \Theta_0} L_n(\theta) / \sup_{\theta \in \Theta_0} L_n(\theta)
\]

and

\[
L_n(\theta, \sigma) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta)^2 \right).
\]

We need to maximize \( (\theta, \sigma) \mapsto L_n(\theta, \sigma) \) over \( \Theta \). This is equivalent to maximizing

\[
l_n(\theta, \sigma) = -n/2 \log(2\pi) - n \log(\sigma) - 1/(2\sigma^2) \sum_{i=1}^n (X_i - \theta)^2.
\]

3.3.1. Maximisation via profiling: Let us fix \( \sigma \in (0, \infty) \) and define the function \( g_n(\theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta)^2 \) which we are going to maximize over \( \mathbb{R} \). Since \( -\frac{1}{2\sigma^2} \) is a constant here. We can use previous calculations from example a). To
show that the minimum is attained at \( \theta = \bar{X}_n \). \( \sup_{\theta \in \mathbb{R}} L_n(\theta, \sigma) = L_n(\bar{X}_n, \sigma) \) for any fixed \( \sigma \in (0, \infty) \). Now, we go back to the log-likelihood and plug in \( \bar{X}_n \): define the function

\[
h(\sigma) = l_n(\bar{X}_n, \sigma) = -n/2 \log(2\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_n - \bar{X}_n)^2
\]

which we want to maximize over \((0, \infty)\).

\[
h'(\sigma) = -n/\sigma + 1/\sigma^3 \sum_{i=1}^n (X_i - \bar{X}_n)^2 = 0
\]

\[
\Leftrightarrow \sigma^2 = 1/n \sum_{i=1}^n (X_i - \bar{X}_n)^2
\]

\[
\Leftrightarrow \sigma = \hat{\sigma} = \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]^{1/2}
\]

which we want to maximize over \((0, \infty)\).

\[
h''(\sigma) = n/\sigma^2 - 3/\sigma^4 \sum_{i=1}^n (X_i - \bar{X}_n)^2
\]

\[
= n/\sigma^2 - 3/\sigma^4 n \hat{\sigma}^2
\]

\[
= n/\sigma^2 - \frac{3n \hat{\sigma}^2}{\sigma^2}
\]

\[
= n/\sigma^2 (\sigma^2 - 3\hat{\sigma}^2).
\]

The function \( h \) has a local maximum at \((7)\). But, since \( h \) has a unique critical point, the function cannot go up to a larger value (\( > h(\hat{\sigma}) \)) because otherwise \( h \) has to go down to reach another critical point. Therefore, \((7)\) must be the global maximizer of \( h \) over \((0, \infty)\). We need to compute \( \sup_{(\theta, \sigma) \in \Theta} L_n(\theta, \sigma) = \sup_{\theta \in (0, \infty)} L_n(\theta, \sigma) \). Using similar arguments as for showing that \((7)\) is the global maximizer of the function \( \sigma \mapsto l_n(\bar{X}_n, \sigma) \) we can show that \( \sup_{\theta \in (0, \infty)} L_n(\theta, \sigma) = L_n(\theta_0, \hat{\sigma}_0) \) with

\[
\hat{\sigma}_0 = \left( \frac{1}{n} \sum_{i=1}^n (X_i - \theta_0)^2 \right)^{1/2}.
\]

\[
\Lambda_n = \sup_{(\theta, \sigma) \in \Theta} L_n(\theta, \sigma) = \sup_{(\theta, \sigma) \in \Theta} L_n(\theta, \sigma)
\]

\[
= L_n(\bar{X}_n, \hat{\sigma}) = L_n(\theta_0, \hat{\sigma}_0)
\]

\[
= \left[ \frac{1}{(2\pi)^{n/2}} \frac{1}{\hat{\sigma}_0^n} \exp \left( -\frac{1}{2\hat{\sigma}_0^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) \right]
\]

\[
= \frac{1}{\hat{\sigma}_0^n} \exp (-n/2)
\]

\[
= \left( \frac{\hat{\sigma}_0}{\hat{\sigma}} \right)^n \frac{1}{\hat{\sigma}_0^n} \exp (-n/2)
\]

\[
= \left( \frac{\hat{\sigma}_0}{\hat{\sigma}} \right)^n \left( \frac{\hat{\sigma}_0}{\hat{\sigma}} \right)^{n/2}
\]
We reject when $\Lambda_n$ is 'large' but

$$\Lambda_n \text{ is 'large'} \iff \frac{\theta_0^2}{\sigma^2} \text{ is 'large'}$$

$$\iff \frac{1}{n} \sum_{i=1}^{n} (X_i - \theta_0)^2 \text{ is 'large'}$$

$$\iff \frac{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \theta_0)^2}{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2} \text{ is 'large'}$$

$$\iff 1 + \frac{n(\bar{X}_n - \theta_0)^2}{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2} \text{ is 'large'}$$

$$\iff \frac{\sqrt{n|\bar{X}_n - \theta_0|}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}} \text{ is 'large'}$$

$$\iff \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}} \text{ is 'large'}.$$  

We can find the distribution of $T_n := \frac{\sqrt{n|\bar{X}_n - \theta_0|}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}}$ under $H_0 : \theta = \theta_0$ using Cochran's theorem. If $(X_1, \ldots, X_n) = X \sim \mathcal{N}(\theta_0, \sigma^2 \mathbb{I})$ then $(\frac{X_1 - \theta_0}{\sigma_0}, \ldots, \frac{X_n - \theta_0}{\sigma_0}) \sim Y \sim \mathcal{N}(0, \mathbb{I})$. Define $A_1 = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$ and $A_2 = \mathbb{I} - A_1$. We have to check that $A_1$ and $A_2$ fulfil the assumptions of Cochran's theorem.

$$A_1^2 = \frac{1}{n^2} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} = \frac{1}{n^2} \begin{pmatrix} n & \cdots & n \\ \vdots & \ddots & \vdots \\ n & \cdots & n \end{pmatrix} = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} = A_1$$

and $A_2 = \mathbb{I} - A_1 A_1 (\mathbb{I} - A_1) = A_1 - A_1^2 = 0 = (\mathbb{I} - 1) A_1 \text{ rank}(A_1) = 1$ and rank($A_2$) = $n - 1$. Therefore, by Cochran's theorem, we know that $A_1 Y$ is independent of $A_2 Y$ and $||A_2 Y||^2 \sim \chi^2_{n-1}$

$$A_1 Y = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \frac{X_1 - \theta_0}{\sigma_0} \\ \vdots \\ \frac{X_n - \theta_0}{\sigma_0} \end{pmatrix} = \frac{\bar{X}_n - \theta_0}{\sigma_0} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$A_2 Y = (\mathbb{I} - A_1) Y = Y - A_1 Y = \begin{pmatrix} \frac{X_1 - \theta_0}{\sigma_0} \\ \vdots \\ \frac{X_n - \theta_0}{\sigma_0} \end{pmatrix} - \frac{\bar{X}_n - \theta_0}{\sigma_0} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{X_1 - \bar{X}_n}{\sigma_0} \\ \vdots \\ \frac{X_n - \bar{X}_n}{\sigma_0} \end{pmatrix}$$

so that $||A_2 Y||^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$. ... Now $A_1 Y \perp A_2 Y \Rightarrow A_1 Y \perp ||A_2 Y||^2 \Rightarrow \frac{\bar{X}_n - \theta_0}{\sigma_0} \perp \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

$$\Rightarrow \frac{\sqrt{n|\bar{X}_n - \theta_0|}}{\sigma} \sim \mathcal{N}(0,1) \underbrace{\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2}_{\sim \chi^2_{n-1}}$$

and using 

$$\Rightarrow \frac{\sqrt{n|\bar{X}_n - \theta_0|}}{\sigma} \sim \mathcal{T}_{(n-1)} \text{ under } H_0.$$
Note that the obtained statistic $T_n = \frac{\chi^2_n - \chi^2_0}{\sum \chi^2_{i-1}}$. Thus, the LRT is given by $\Phi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{if } |T_n| > q_\alpha \\ 0 & \text{otherwise} \end{cases}$ where $q_\alpha = t_{\nu\alpha - 1, 1-\alpha/2}$ the $(1-\alpha/2)$-quantile of $T(n-1)$.

3.4. Example c. Let $X_1, \ldots, X_n \sim \mathcal{N}(\theta_0, \sigma^2)$ with $\theta_0 \in \mathbb{R}$ known and $\sigma \in (0, \infty)$ unknown. We want to test $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$

whereas $\Theta_0 = \{\sigma_0\}$ and $\Theta = (0, +\infty)$.

$$\Lambda_n = \sup_{\sigma \in (0, \infty)} \frac{L_n(\theta_0, \sigma)}{L_n(\theta_0, \sigma_0)}$$

$L_n(\theta_0, \sigma) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta_0)^2 \right)$ then

$$l_n(\theta_0, \sigma) = -n/2 \log(2\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta_0)^2.$$ 

$$\frac{d}{d\sigma} (l_n(\theta_0, \sigma)) = -n/\sigma + 1/\sigma^3 \sum_{i=1}^n (X_i - \theta_0)^2 = 0 \Leftrightarrow \sigma^2 = 1/n \sum_{i=1}^n (X_i - \theta_0)^2$$

which implies that there exists a unique critical point

$$\hat{\sigma} = \left(\frac{1}{n} \sum_{i=1}^n (X_i - \theta_0)^2\right)^{1/2}$$

$$\frac{d^2}{d\sigma^2} (l_n(\theta_0, \sigma)) = n/\sigma^2 - 3/\sigma^4 \sum_{i=1}^n (X_i - \theta_0)^2$$

and

$$\frac{d^2}{d\sigma^2} (l_n(\theta_0, \sigma))|_{\sigma = \hat{\sigma}} = n/\hat{\sigma} - 3n\hat{\sigma}^2 / \hat{\sigma}^4 = \frac{2n}{\hat{\sigma}^2} < 0$$

which means that $\hat{\sigma}$ is a local maximizer and hence a global maximizer because otherwise the function $\sigma \mapsto l_n(\theta_0, \sigma)$ will have another critical point. Note that this obtained $\hat{\sigma}$ is equal to (??).

$$\Lambda_n = \frac{L_n(\theta_0, \hat{\sigma})}{L_n(\theta_0, \sigma_0)} = \frac{1}{(2\pi)^{n/2} \hat{\sigma}^n} \exp \left(-\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \theta_0)^2 \right)$$

$$= \frac{1}{(2\pi)^{n/2} \sigma_0^n} \exp \left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (X_i - \theta_0)^2 \right)$$

$$= \frac{1}{\sigma_0^n} \exp \left(-\frac{n\sigma_0^2}{2} \right)$$

$$= \frac{1}{\hat{\sigma}^n} \exp \left(-\frac{n\sigma_0^2}{2\hat{\sigma}^2} \right)$$

$$= \frac{\sigma_0^n}{\hat{\sigma}^n} \exp \left(-\frac{n}{2} + n/\hat{\sigma}^2 / \sigma_0^2 \right)$$

$$= \frac{1}{(\hat{\sigma}/\sigma_0)^n} \exp \left(-\frac{n}{2} \left(\frac{\hat{\sigma}}{\sigma_0}\right)^2 - 1 \right)$$

$$= g \left( \frac{\hat{\sigma}}{\sigma_0} \right)$$

with $g(t) = 1/t^p \exp \left(n/(2t^2 - 1) \right)$ for $t \in (0, +\infty)$.

$$h(t) = \log(g(t))$$

$$= -n \log(t) + n/(2t^2 - 1)$$
\[ h'(t) = -n/t + nt = \frac{t^2 - 1}{t} \]

But we know that, by definition, \( \Lambda_n \geq 1 \) and hence \( \Lambda_n = g(\hat{\sigma}/\sigma_0) \) which implies \( \frac{\hat{\sigma}}{\sigma_0} \in [1, +\infty) \). Since \( g \) is strictly increasing on \([1, +\infty)\),

\[ \Lambda_n \text{ 'large'} \Leftrightarrow \frac{\hat{\sigma}}{\sigma_0} \text{ is 'large'} \]

\[ \Leftrightarrow \frac{\hat{\sigma}^2}{\sigma_0^2} \text{ is 'large'} \]

\[ \Leftrightarrow \frac{1/n \sum_{i=1}^n (X_i - \hat{\theta}_0)^2}{\sigma_0^2} \text{ is 'large'} \]

\[ \Leftrightarrow \sum_{i=1}^n (X_i - \hat{\theta}_0)^2 / \sigma_0^2 \text{ is 'large'} \].

The LRT is given by \( \Phi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{if } \frac{\sum_{i=1}^n (X_i - \hat{\theta}_0)^2}{\sigma_0^2} > q_\alpha \text{ with } P_{\sigma_0}\left(\sum_{i=1}^n \frac{(X_i - \hat{\theta}_0)^2}{\sigma_0^2} > q_\alpha\right) = \alpha. \\ 0 & \text{otherwise} \end{cases} \)

\( X_i - \theta_0 / \sigma_0, \ldots, X_i - \theta_0 / \sigma_0 \overset{iid}{\sim} \mathcal{N}(0, 1) \) under \( H_0 : \sigma = \sigma_0 \) which implies \( \sum_{i=1}^n \frac{(X_i - \hat{\theta}_0)^2}{\sigma_0^2} \sim \chi^2_0 \) and \( q_\alpha \) the \((1 - \alpha)\)-quantile of \( \chi^2_0 \).

3.5. **Example d.** Let \( X_1, \ldots, X_n \sim \mathcal{N}(\theta, \sigma^2) \) with \( \theta \in \mathbb{R} \) and \( \sigma \in (0, \infty) \) both unknown. Here \( \theta \) is acting as a nuisance parameter and we want to test

\[ H_0 : \theta \text{ is something, } \sigma = \sigma_0 \text{ versus } H_1 : \theta \text{ is something, } \sigma \neq \sigma_0 \]

whereas \( \Theta_0 = \{(\theta, \sigma_0) : \theta \in \mathbb{R}\} \) and \( \Theta = \mathbb{R} \times (0, +\infty) \).

\[ \Lambda_n = \sup_{(\theta, \sigma) \in \Theta} \frac{L_n(\theta, \hat{\sigma}^\prime)}{L_n(\theta, \sigma_0)} \]

\[ L_n(\theta, \sigma) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta)^2 \right) \]

We already know from example b that \( \sup_{(\theta, \sigma) \in \Theta} = L_n(\bar{X}_n, \hat{\sigma}) \) with \( \hat{\sigma} = \left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)^{1/2} \) and also

\[ \Lambda_n = \frac{1}{(2\pi)^{n/2} \hat{\sigma}^n} \exp \left( -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) \]

\[ = \frac{1}{1/\hat{\sigma}^n} \exp \left( -n/2 + n/2 \cdot \hat{\sigma}^2 / \hat{\sigma}_0^2 \right) \]

\[ \Lambda_n = g(\hat{\sigma}/\sigma_0) \] where \( g \) is the same function as before. Using similar arguments we show that \( \Lambda_n \) is 'large' if and only if \( \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma_0^2} \) is 'large'. \( \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma_0^2} \sim \chi^2_{(n-1)} \) as a result of Cochran's theorem. The LRT is given by \( \Phi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{if } \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma_0^2} > q_\alpha \text{ with } q_\alpha = (1 - \alpha)\text{-quantile of } \chi^2_{(n-1)} \text{.} \\ 0 & \text{otherwise} \end{cases} \)

4. **F-tests and application in linear regression**

4.1. **Regression model.** A regression model aims at explaining the random behaviour of the response given the explanatory variables also called covariates/predictors. More specifically, a regression model assumes that \( Y = f(\theta, x) + \epsilon \) whereas \( Y \) is the response, \( f \) and \( \theta \) are unknown \( x \) are the covariate(s) and \( \epsilon \) is the noise/error.

There are two settings:

- (1) Random design: the covariate is random and the analysis is done conditionally on \( X \) but in the end randomness is taken into account.
- (2) Fixed design: We observe a realisation \( x \) of \( X \) and we do the analysis conditionally on \( X = x \).

In this course we will place ourselves in the fixed design.
4.2. Linear Regression. When \( f(\theta, x) = \theta^\top x \) with \( \theta, x \in \mathbb{R}^d \), then we talk about linear regression. The model is \( Y = \theta^\top x + \epsilon \) with \( E(\epsilon) = 0 \). If \( \theta_1, \ldots, \theta_d \) are the components of \( \theta \) and \( x_1, \ldots, x_n \) are the components of \( x \) then
\[
Y = x_1 \theta_1 + \ldots + \theta_d x_d + \epsilon.
\]
The main goal is to estimate the unknown regression vector \( \theta \) based on a random sample. We observe independent responses \( Y_1, \ldots, Y_n \) and corresponding covariates \( x_1, \ldots, x_n \in \mathbb{R}^d \). Let
\[
Y_i = \theta^\top x_i + \epsilon_i
\]
with \( x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix} \) for \( i \in \{1, \ldots, n\} \), \( Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \in \mathbb{R}^n \) and \( \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \in \mathbb{R}^n \) and put \( D = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & \cdots & \cdots & x_{nd} \end{pmatrix} \in \mathbb{R}^{n \times d} \). The \( i \)-th row of \( D = x_i^\top = (x_{i1}, \ldots, x_{id}) \). \( D \) is called the design-matrix. We can write the linear regression as
\[
Y = D\theta + \epsilon.
\]

4.3. Least Squares Estimator.

Definition 4.1. LSE Consider the quadratic criterion
\[
Q_n(t) = \sum_{i=1}^n (Y_i - t^\top x_i)^2
\]
for \( t \in \mathbb{R}^d \). \( \hat{\theta}_n = \arg\min_{t \in \mathbb{R}^d} Q_n(t) \) is called (provided it exists) the least squares estimator if it minimizes \( Q_n \) over \( \mathbb{R}^d \).

The rational behind \( \hat{\theta}_n \) is that we can take some random variable \( Z \) with \( \mu = E(Z) < \infty \) and \( \sigma^2 = \text{Var}(Z) < \infty \) then \( \mu = \arg\min_{a \in \mathbb{R}} E[(Z - a)^2] \). Indeed
\[
E[(Z - a)^2] = E[(Z - \mu + \mu - a)^2]
\]
\[
= E[(Z - \mu)^2 + 2(Z - \mu)(\mu - a) + (\mu - a)^2]
\]
\[
= \sigma^2 + 2(\mu - a)E[Z - \mu] + (\mu - a)^2
\]
\[
= \sigma^2 + (\mu - a)^2.
\]

Since \( \arg\min_{a \in \mathbb{R}} (\mu - a)^2 = \mu \) it follows that \( \mu = \arg\min_{a \in \mathbb{R}} E[(Z - a)^2] \). Let us go back to the regression problem and let us also assume that \( \text{Var}(Y_i) < \infty \) for \( i \in \{1, \ldots, n\} \). Since \( E(\epsilon_i) = 0 \) for \( i \in \{1, \ldots, n\} \), this means that \( E(Y_i) = \theta^\top x_i = \mu_i \). We can also show as above that
\[
(\mu_1, \ldots, \mu_n)^\top = \sum_{i=1}^n E[(Y_i - a_i)^2] \Rightarrow \theta = \arg\min_{a \in \mathbb{R}^d} \sum_{i=1}^n E[(Y_i - t^\top x_i)^2] \).
\]
Since we only observe \( Y_1, \ldots, Y_n \) and \( x_1, \ldots, x_n \), we replace this criterion by (10).

Proposition 4.2. Assume that \( D^\top D \) is invertible. Then, \( \hat{\theta}_n \) exists and is unique. Furthermore
\[
\hat{\theta}_n = (D^\top D)^{-1} D^\top Y.
\]

Proof. Recall that for \( v = (v_1, \ldots, v_n) \in \mathbb{R}^n \) the euclidean norm is defined as \( \| v \| = \sqrt{\sum_{i=1}^n v_i^2} \) and \( \| v \|^2 = v^\top v \). Hence
\[
Q_n(t) = \sum_{i=1}^n (Y_i - t^\top x_i)^2
\]
\[
= \| Y - Dt \|^2
\]
\[
= (Y - Dt)^\top (Y - Dt)
\]
\[
= Y^\top Y - Y^\top Dt - t^\top D^\top Y + t^\top D^\top Dt
\]
\[
= Y^\top Y - 2t^\top D^\top Y + t^\top D^\top D t
\]
We look now for a stationary point of \( Q_n : \forall Q_n(t) = -2D^\top Y + 2D^\top Dt \). Recall that for any differentiable function \( g \) defined on \( \mathbb{R}^d \) we have
\[
g(t + h) = g(t) + h^\top \nabla g(t) + o(\|h\|).
\]
Therefore
\[ \nabla Q_n(t) = 0 \Leftrightarrow D^\top Dt = D^\top Y \]
\[ \Leftrightarrow t = (D^\top D)^{-1}D^\top Y. \]

The hessian of \( Q_n(t) \) is \( 2D^\top D \), which is positive definite because for \( a \in \mathbb{R}^d \)

\[ a^\top D^\top Da = (Da)^\top Da \]
\[ = \|Da\|^2 \geq 0 \]

and

\[ a^\top D^\top Da = 0 \Leftrightarrow \|Da\|^2 = 0 \]
\[ \Leftrightarrow Da = 0 \]
\[ \Rightarrow D^\top Da = 0 \]
\[ \Rightarrow a = 0. \]

It follows that \( \hat{\theta}_n = (D^\top D)^{-1}D^\top Y \) is the unique minimizer of (the strictly convex function) \( Q_n \).

\[ \square \]

4.4. Properties of the LSE. In what follows we assume \( E[\epsilon \epsilon^\top] = \sigma^2 I_n \). In other words \( E[\epsilon_i^2] = \sigma^2 \) for \( i \in \{1, \ldots, n\} \) and \( E[\epsilon_i \epsilon_j] = 0 \) for \( i \neq j \in \{1, \ldots, n\} \).

**Proposition 4.3.** Assume that \( D^\top D \) is invertible. Then,

(i) \( E[\hat{\theta}_n] = \theta \) and

(ii) \( E[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^\top] = \sigma^2 (D^\top D)^{-1} \).

**Proof.** (i) Use (9) to see that

\[ \hat{\theta}_n = (D^\top D)^{-1}D^\top Y \]
\[ = (D^\top D)^{-1}D^\top (D\theta + \epsilon) \]
\[ = (D^\top D)^{-1}D^\top D\theta + (D^\top D)^{-1}D^\top \epsilon \]
\[ = \theta + (D^\top D)^{-1}D^\top \epsilon \]

\[ \text{(12)} \]

Since \( E[\epsilon] = 0 \) (i) follows.

(ii) Use (12) to see that

\[ E[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^\top] = E[(D^\top D)^{-1}D^\top \epsilon \epsilon^\top D(D^\top D)^{-1}] \]
\[ = (D^\top D)^{-1}D^\top E[\epsilon \epsilon^\top]D(D^\top D)^{-1} \]
\[ = (D^\top D)^{-1}D^\top \sigma^2 I_n D(D^\top D)^{-1} \]
\[ = \sigma^2 (D^\top D)^{-1}D^\top D(D^\top D)^{-1} \]
\[ = \sigma^2 (D^\top D)^{-1} \]

\[ \square \]

**Proposition 4.4.** Let us assume that \( \epsilon \sim N(0, \sigma^2 I_n) \). Then,

(i) \( \hat{\theta}_n \sim N(\theta, \sigma^2 (D^\top D)^{-1}) \).

(ii) \( Y - D\hat{\theta}_n \) and \( D(\hat{\theta}_n - \theta) \) are independent Gaussian vectors.

(iii) \( \frac{\|Y - D\hat{\theta}_n\|^2}{\sigma^2} \sim \chi^2_n \) and \( \frac{\|\hat{\theta}_n - \theta\|^2}{\sigma^2} \sim \chi^2_d \).

**Proof.** (i) Recall that \( D \) is the design matrix and \( Y = D\theta + \epsilon \). Then,

\[ \hat{\theta}_n = (D^\top D)^{-1}D^\top Y = (D^\top D)^{-1}D^\top (D\theta + \epsilon) \]
\[ = \theta + (D^\top D)^{-1}D^\top \epsilon \]

\[ \square \]
whereas \((D^\top D)^{-1}D^\top\) is a matrix and \(\epsilon\) is a gaussian vector. This means that \(\hat{\theta}_n\) is also a gaussian vector with \(E[\hat{\theta}_n] = \theta + 0 = \theta\) and covariance matrix \(E[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^\top] = \sigma^2(D^\top D)^{-1}\) hence \(\hat{\theta}_n \sim N(\theta, \sigma^2(D^\top D)^{-1})\).

(ii) We want to show that \(Y - D\hat{\theta}_n \perp D(\hat{\theta}_n - \theta)\) whereas \(Y - D\hat{\theta}_n\) denotes the estimated residuals.

\[
D(\hat{\theta}_n - \theta) = D((D^\top D)^{-1}D^\top Y - \theta)
\]
\[
= D((D^\top D)^{-1}D^\top(D\theta + \epsilon - \theta))
\]
\[
= A\epsilon.
\]

Note that \(A^\top = A\) and

\[
A^2 = D(D^\top D)^{-1}D^\top D(D^\top D)^{-1}D^\top
\]
\[
= D(D^\top D)^{-1}D^\top
\]
\[
= A.
\]

On the other hand

\[
Y - D\hat{\theta}_n = D\theta + \epsilon - D(D^\top D)^{-1}D^\top(D\theta + \epsilon)
\]
\[
= \epsilon - D(D^\top D)^{-1}D^\top \epsilon
\]
\[
= (I - A)\epsilon.
\]

\(I - A\) is symmetric and satisfies \((I - A)^2 = (I - A)(I - A) = I - A - A + A^2 = I - A\). Furthermore, \((I - A)A = A - A^2 = 0 = A(I - A)\) and rank\((A) = d\) because \(D^\top D\) is invertible (see in the notes on linear algebra) which implies that rank\((I - A) = n - d\). Using Cochran’s theorem, it follows that \(Y - D\hat{\theta}_n \perp D(\hat{\theta}_n - \theta)\) and

\[
\frac{\|D(\hat{\theta}_n - \theta)\|}{\sigma^2}^2 = \frac{\|\epsilon\|}{\sigma^2}^2 \sim \chi^2_{\text{rank}(A)} \frac{d}{D^\top D} \chi^2_d
\]
\[
\frac{\|Y - D\hat{\theta}_n\|}{\sigma^2}^2 = \frac{\|(I_n - A)\epsilon\|}{\sigma^2}^2 \sim \chi^2_{n-d},
\]

which is also proof for (iii).

\[\square\]

**Proposition 4.5.** Consider the linear regression model \(Y = D\theta + \epsilon\) with \(\epsilon \sim N(0, \sigma^2I_n)\). Consider also the testing problem

\[H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.\]

If \(\sigma = \sigma_0\) is known then a test of level \(\alpha\) for this problem is given by

\[
\Phi(X_1, \ldots, X_n) = \begin{cases} 
1 & \text{if} \quad \frac{\|D(\hat{\theta}_n - \theta_0)\|}{\sigma_0^2}^2 > q_{d,1-\alpha} \\
0 & \text{otherwise}
\end{cases}
\]

where \(q_{d,1-\alpha}\) is the \((1 - \alpha)\) quantile of \(\chi^2_d\).

Proof. Under \(H_0\), we know from (iii) that \(\frac{\|D(\hat{\theta}_n - \theta_0)\|}{\sigma_0^2}^2 = \chi^2_d\) so that \(P\left(\frac{\|D(\hat{\theta}_n - \theta_0)\|}{\sigma_0^2}^2 > q_{d,1-\alpha}\right) = \alpha.\)

\[\square\]

**Proposition 4.6.** Let \(Y = D\theta + \epsilon\) with \(\epsilon \sim N(0, \sigma^2I_n)\) and consider the problem (13). Suppose \(\sigma\) is known. Then a test of level \(\alpha\) for this problem is given by

\[
\Phi(X_1, \ldots, X_n) = \begin{cases} 
1 & \text{if} \quad \frac{\|D(\hat{\theta}_n - \theta_0)\|}{\|Y - D\theta_0\|/(n-d)} > q_{d,n-d,1-\alpha} \\
0 & \text{otherwise}
\end{cases}
\]

where \(q_{d,n-d,1-\alpha}\) is the \((1 - \alpha)\) quantile of the F-distribution \((\chi^2_d)\) of \(d\) and \(n - d\) degrees of freedom.

Proof.

\[
\frac{\|D(\hat{\theta}_n - \theta_0)\|}{\sigma_0^2}^2 / d \sim F_{(d,n-d)}
\]

under \(H_0\) because \(\|D(\hat{\theta}_n - \theta_0)\|^2 \perp \|Y - D\hat{\theta}_n\|^2\), (ii) and (iii).

\[\square\]
4.5. \textbf{\( \chi^2 \)- and \( F \)-tests for variable selection.} The question we want to answer is: Which of the covariates are significant (have a non-trivial effect on the response). More formally, the question can be put in the context of testing. We want a test where \( \theta \) is of the form \((\theta_1, \ldots, \theta_{d-m}, 0, \ldots, 0)^T\). Even more formally, we want to test

\[
H_0 : G\theta = 0 \quad \text{versus} \quad H_1 : G\theta \neq 0
\]

where \( G = \begin{bmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \) and \( \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \). Note that \( H_1 \) means that there exists \( j \in \{d-m+1, \ldots, d\} \) \( \theta_j \neq 0 \) and

\[
G\theta = \begin{bmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{d-m} \\ \theta_{d-m+1} \\ \theta_d \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{bmatrix}.
\]

4.5.1. \textbf{LRT for variable selection.} Let us assume that \( \epsilon \sim \mathcal{N}(0, \sigma^2 I_n) \) where \( \sigma^2 \) is known.

\[
\Theta_0 = \{ \theta \in \mathbb{R}^d : G\theta = 0 \} = \{ \theta \in \mathbb{R}^d : \theta_{d-m+1} = \ldots = \theta_d = 0 \}
\]

\[
\Theta = \mathbb{R}^d
\]

\[
L_n(\theta) = \frac{1}{(2\pi)^{d/2} \sigma^n} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \theta^T x_i)^2 \right) = \frac{1}{(2\pi)^{d/2} \sigma^n} \exp \left( -\frac{1}{2\sigma^2} ||Y - D\theta||^2 \right)
\]

\[
l_n(\theta) = -n/2 \log(2\pi) - n \log(\sigma^2) - 1/(2\sigma^2)||Y - D\theta||^2.
\]

Maximizing \( \theta \mapsto l_n(\theta) \) over \( \mathbb{R}^d \) is equivalent to minimizing \( \theta \mapsto ||Y - D\theta||^2 \) over \( \mathbb{R}^d \). We know that the solution is the LSE (??). Hence \( \sup_{\theta \in \Theta} L_n(\theta) = \sup_{\theta \in \Theta} l_n(\theta) \).

Now, we need to maximize \( \theta \mapsto l_n(\theta) \) over \( \Theta_0 \). But this is equivalent to minimize \( \theta \mapsto ||Y - D\theta||^2 \) over \( \Theta_0 \). Under \( H_0 \) we have

\[
D\theta = \begin{bmatrix} x_{11} & \ldots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nd} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{d-m} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} x_{11} & \ldots & x_{1(d-m)} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{n(d-m)} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{d-m} \end{bmatrix} = D\hat{\theta}.
\]

(15)
This problem is equivalent to minimizing $\bar{\theta} \mapsto \|Y - D\bar{\theta}\|^2$. We only need to check that $D^\top D$ is invertible. Note that $D = DG$ with $G = \begin{pmatrix} d_{m-m} \\ 0_m \end{pmatrix}$. Let $a \in \mathbb{R}^{d-m}$. We want to show that $D^\top Da = 0$ implies $a = 0$.

$$D^\top Da = 0 \Rightarrow a^\top D^\top D = 0$$

$$\iff (Da)^\top Da = \|Da\|^2 = 0$$

$$\iff Da = 0$$

$$\iff D\bar{G}a = 0$$

$$\iff Db = 0$$

because $D^\top D$ is invertible if and only if $\text{rank}(D) = d$. Hence $\bar{G}a = 0$ if and only if $a = 0$. $D^\top D$ is invertible and therefore we are in the same setting as in the least squares problem. Hence the minimizer of $\bar{\theta} \mapsto \|Y - \bar{D}\bar{\theta}\|^2$ is given by $(D^\top D)^{-1} \bar{D}^\top Y$ if and only if the minimizer of $\theta \mapsto \|Y - D\theta\|^2$ under $H_0$ is given by $\hat{\theta}_n^0 = \left( (D^\top D)^{-1} \bar{D}^\top Y \right)_0$, $\theta \mapsto l_n(\theta)$ is maximized by $\hat{\theta}_n^0$ under $H_0$ and

$$\Lambda_n = \sup_{\bar{\theta} \in \bar{\Theta}} l_n(\bar{\theta})$$

$$= \frac{1}{(2\pi)^{d/2}\sigma_0^2} \exp\left( -\frac{1}{2\sigma_0^2} \|Y - \hat{D}\hat{\theta}_n\|^2 \right)$$

$$= \exp\left( \frac{1}{2\sigma_0^2} \left( \|Y - \hat{D}\hat{\theta}_n^0\|^2 - \|Y - D\hat{\theta}_n\|^2 \right) \right).$$

We reject if $\Lambda_n$ is 'large' which means that if $\|Y - \hat{D}\hat{\theta}_n^0\|^2 - \|Y - D\hat{\theta}_n\|^2$ is large.

$$\|Y - \hat{D}\hat{\theta}_n^0\|^2 = \|Y - D\hat{\theta}_n + \hat{D}(\hat{\theta}_n - \hat{\theta}_n^0)\|^2$$

$$= \|Y - D\hat{\theta}_n\|^2 + 2(Y - D\hat{\theta}_n)^\top \hat{D}(\hat{\theta}_n - \hat{\theta}_n^0) + \|\hat{D}(\hat{\theta}_n - \hat{\theta}_n^0)\|^2.$$

Now we show that $2(Y - D\hat{\theta}_n)^\top \hat{D}(\hat{\theta}_n - \hat{\theta}_n^0) = 0$. We know that $\hat{\theta}_n$ is a zero of the gradient of the function $Q_n(t) = \|Y - Dt\|^2$, $t \in \mathbb{R}^d$. In other words

$$D^\top D\hat{\theta}_n - D^\top Y = 0 \iff D^\top (D\hat{\theta}_n - Y) = 0$$

$$\iff (Y - D\hat{\theta}_n)^\top D = 0$$

$$\iff (Y - D\hat{\theta}_n)^\top Dv = 0$$

for all $v \in \mathbb{R}^d$. In particular this holds true for $v = \hat{\theta}_n - \hat{\theta}_n^0$. $\Lambda_n$ is 'large' if and only if $\|D(\hat{\theta}_n - \hat{\theta}_n^0)\|^2$ is 'large'. What is the distribution of $\|D(\hat{\theta}_n - \hat{\theta}_n^0)\|^2$ under $H_0$?

4.5.2. The LRT for variable selection. $\sigma = \sigma_0$ is known.

$$\Lambda_n \text{ 'is large'} \iff \|D(\hat{\theta}_n - \hat{\theta}_n^0)\|^2 \text{ 'is large'}$$

$$\iff \|D(\hat{\theta}_n - \hat{\theta}_n^0)\|^2 \sigma_0^{-2} \text{ 'is large'}$$

where $\hat{\theta}_n = (D^\top D)^{-1} D^\top Y$ and $\hat{\theta}_n^0 = (\bar{D}^\top \bar{D})^{-1} \bar{D}^\top \bar{Y}$

Question: What is the distribution of $\|D(\hat{\theta}_n - \hat{\theta}_n^0)\|^2$ under $H_0 : G\theta = 0$?

$$D(\hat{\theta}_n - \hat{\theta}_n^0) = D(\hat{\theta}_n - \theta) - D(\hat{\theta}_n^0 - \theta)$$

$$= \left( D(D^\top D)^{-1} D^\top - \bar{D}(\bar{D}^\top \bar{D})^{-1} \bar{D}^\top \right) \epsilon$$
whereas \( Y = D\theta + \epsilon = \hat{D}\theta + \epsilon \) under \( H_0 \) and \( \epsilon \sim N(0, \sigma_0 I_n) \). Recall \([15]\) and observe that
\[
AB = D(D^\top D)^{-1} D^\top \hat{D}(D^\top D)^{-1} \hat{D}^\top
= D(D^\top D)^{-1} D^\top D\hat{G}(D^\top D)^{-1} \hat{D}^\top
= D\hat{G}(\hat{D}^\top \hat{D})^{-1} \hat{D}^\top
= B
\]
and
\[
BA = \hat{D}(\hat{D}^\top \hat{D})^{-1} \hat{D}^\top D(D^\top D)^{-1} D^\top
= \hat{D}(\hat{D}^\top \hat{D})^{-1} \hat{G}^\top D^\top D(D^\top D)^{-1} D^\top
= \hat{D}(\hat{D}^\top \hat{D})^{-1} \hat{G}^\top D^\top
= B
\]
I.e. \( BA = AB \) if and only if \( A \) and \( B \) commute \( (A^\top = A \text{ and } B^\top = B) \). Furthermore, the matrices are projections meaning \( A^2 = A \) and \( B^2 = B \). Hence, we can find an orthogonal matrix \( P \) such that
\[
A = P^\top \begin{pmatrix} I_d & 0 \\ 0 & 0 \end{pmatrix} P \quad \text{and} \quad B = P^\top \begin{pmatrix} I_{d-m} & 0 \\ 0 & 0 \end{pmatrix} P
\]
because \( \text{rank}(A) = \text{rank}(D^\top D) = d \) and \( \text{rank}(B) = \text{rank}(\hat{D}^\top \hat{D}) \) (see notes on linear algebra). Moreover
\[
A - B = P^\top \begin{pmatrix} 0 & 0 \\ 0 & 1_m \end{pmatrix} P
\]
which implies \( \text{rank}(A - B) = m \). Hence we can write
\[
\frac{||D\hat{D} - \theta\|_{\sigma^2}}{\sigma} = ||(A - B) \frac{\hat{\epsilon}}{\sigma^2}||^2 \sim \chi^2_{\text{rank}(A-B)} \text{ under } H_0 \quad \frac{||D\hat{D} - \theta\|_{\sigma^2}}{\sigma} \sim \chi^2_{(m)} \quad \text{with } \hat{\theta} = \left( \hat{\theta}^\top \hat{D}^\top \hat{D} \right)^{-1} \hat{D}^\top \hat{D} \hat{\theta}
\]
The LRT of level \( \alpha \) can be given by
\[
\Phi(Y_1, \ldots, Y_n) = \begin{cases} 1 & \text{if } \frac{||D\hat{D} - \theta\|_{\sigma^2}}{\sigma} > q_{m,1-\alpha} \\ 0 & \text{otherwise} \end{cases}
\]
with \( q_{m,1-\alpha} = (1 - \alpha)\)-quantile of \( \chi^2_{(m)} \).

The likelihood is
\[
L_n = \frac{1}{(2\pi)^{d/2} \sigma^n} \exp \left( -\frac{1}{2\sigma^2} ||Y - D\theta||^2 \right)
\]
with
\[
\Theta = \{ (\theta, \sigma) \in \mathbb{R}^d \times (0, +\infty) \} = \mathbb{R}^d \times (0, +\infty)
\]
and
\[
\Theta_0 = \{ (\theta, \sigma) : G\theta = 0 \text{ and } \sigma \in (0, +\infty) \}
= \{ \theta \in \mathbb{R}^d : \theta_{d-m+1} = \cdots = \theta_d = 0 \} \times (0, +\infty).
\]
The log-likelihood is
\[
l_n(\theta) = -n/2 \log(2\pi) - n \log(\sigma) - 1/(2\sigma^2)||Y - D\theta||^2.
\]
To maximize \( (\theta, \sigma) \mapsto l_n(\theta, \sigma) \) over \( \Theta \) we can use the profiling approach:
- Fix \( \sigma \in (0, +\infty) \) and maximize \( \theta \mapsto l_n(\theta, \sigma) \) over \( \mathbb{R}^d \). It is clear, for a fixed \( \sigma \), the solution \( \hat{\theta}_n \) is the one minimizing \( \theta \mapsto ||Y - D\theta||^2 \) on \( \mathbb{R}^d \), that is \( \hat{\theta}_n \) the LSE.
- We plug the obtained solution \( \hat{\theta}_n \) and maximize the function
\[
\sigma \mapsto l_n(\hat{\theta}_n, \sigma) = -n/2 \log(2\pi) - n \log(\sigma) - 1/(2\sigma^2)||Y - D\hat{\theta}_n||^2.
\]
that ˆ

Using the same arguments as for example b (for testing the mean of a Gaussian with unknown variance) we can show

\[
\frac{d}{d\sigma} l_n(\hat{\theta}_n, \sigma) = -n/\sigma + 1/(\sigma^3) ||Y - D\hat{\theta}_n||^2 = 0
\]

\[\Leftrightarrow \sigma^2 = 1/n ||Y - D\hat{\theta}_n||^2
\]

\[\Leftrightarrow \sigma = 1/\sqrt{n} ||Y - D\hat{\theta}_n||^2
\]

whereas \(\sigma\) is the unique critical point of \(\sigma \mapsto l_n(\hat{\theta}_n, \sigma)\).

\[
\frac{d^2}{d\sigma^2} l_n(\hat{\theta}_n, \sigma)|_{\sigma = \hat{\sigma}_n} = -n/\hat{\sigma}^2 - 3/\hat{\sigma}^4 ||Y - D\hat{\theta}_n||^2
\]

\[= -n/\hat{\sigma}^2 - 3/\hat{\sigma}^4 n\hat{\sigma}_n^2
\]

\[= \frac{2n}{\hat{\sigma}^2} < 0.
\]

Using the same arguments as for example b (for testing the mean of a Gaussian with unknown variance) we can show

that \(\hat{\sigma}_n\) gives the global maximum and also that

\[
\sup_{(\theta, \sigma) \in \Theta} L_n(\theta, \sigma) = L_n(\hat{\theta}_n, \hat{\sigma}_n).
\]

Now we need to find \(\sup_{(\sigma, \theta) \in \Theta} L_n(\sigma, \theta)\). Similar arguments can be used to show that \(\sup_{(\sigma, \theta) \in \Theta} L_n(\sigma, \theta) = L_n(\hat{\theta}_n, \hat{\sigma}_n)\) with \(\sigma_n^0 = \left(\left(D^\top D\right)^{-1} D^\top Y\right)_{0:m}\) and \(\hat{\sigma}_n^0 = \frac{1}{\sqrt{n}} ||Y - D\hat{\theta}_n||^2\).

\[
L_n = \frac{\sup_{(\theta, \sigma) \in \Theta} L_n(\theta, \sigma)}{\sup_{(\theta, \sigma) \in \Theta} L_n(\theta, \sigma)\sup_{(\sigma, \theta) \in \Theta} L_n(\sigma, \theta)}
\]

\[= L_n(\hat{\theta}_n, \hat{\sigma}_n)
\]

\[= \frac{1}{(2\pi)^{r/2} \hat{\sigma}_n^0} \exp \left(-\frac{1}{2\hat{\sigma}_n^0} ||Y - D\hat{\theta}_n||^2\right)
\]

\[= \left(\frac{\hat{\sigma}_n^0}{\hat{\sigma}_n^0}ight)^n / \hat{\sigma}_n^0
\]

\[= \left(\frac{\hat{\sigma}_n^0}{\hat{\sigma}_n^0}\right)^{n/2}
\]

\[\Lambda_n \ 'is\ large' \Leftrightarrow \left(\frac{\hat{\sigma}_n^0}{\hat{\sigma}_n^0}\right)^2 \ 'is\ large'\]

\[\Leftrightarrow \frac{1/n ||Y - D\hat{\theta}_n||^2}{1/n ||Y - D\hat{\theta}_n||^2} \ 'is\ large'.
\]

\[||Y - D\hat{\theta}_n^0||^2 = ||Y - D\hat{\theta}_n||^2 + 2 \left(\frac{Y - D\hat{\theta}_n}{||Y - D\hat{\theta}_n||^2}\right) \left(D\hat{\theta}_n - \hat{\theta}_n^0\right) + ||D(\hat{\theta}_n - \hat{\theta}_n^0)||^2
\]

\[= 0
\]

\[\Lambda_n \ 'is\ large' \Leftrightarrow 1 + \frac{||Y - D\hat{\theta}_n||^2}{||Y - D\hat{\theta}_n||^2} \ 'is\ large'\]

\[\Leftrightarrow \frac{||Y - D\hat{\theta}_n||^2}{||Y - D\hat{\theta}_n||^2} \ 'is\ large'.
\]

We know that \(D(\hat{\theta}_n - \hat{\theta}_n^0) = (A - B)\epsilon\). Also \(Y - D\hat{\theta}_n = D\theta + \epsilon - D(D^\top D)^{-1} D^\top (D\theta + \epsilon) = (I_n - A)\epsilon\).

\[(A - B)(I_n - A) = A - B - (A - B)A
\]

\[= A - B - (A - B) = 0
\]
and similarly \((\mathbb{I}_n - A)(A - B) = 0\). Also
\[
\]
and
\[
(\mathbb{I}_n - A)^2 = (\mathbb{I}_n - A)(\mathbb{I}_n - A) = \mathbb{I}_n - A - A^2 = \mathbb{I}_n - A.
\]
moreover we know \(\text{rank}(A - B) = m\) from previous calculations and \(\text{rank}(\mathbb{I} - A) = n - \text{rank}(A) = n - d\). Using Cochran’s theorem we have
\[
||D(\hat{\theta}_n - \hat{\theta}_0^n)||^2 = ||(A - B)\epsilon\sigma^2||^2 \sim \chi^2_{(m)} \quad \text{and} \quad ||Y - D\hat{\theta}_n||^2 = ||(\mathbb{I}_n - A)\epsilon\sigma^2||^2 \sim \chi^2_{(n-d)}.
\]
Hence, under \(H_0\)
\[
\frac{||D(\hat{\theta}_n - \hat{\theta}_0^n)||^2}{||Y - D\hat{\theta}_n||^2} \sim F_{(m,n-d)}
\]
with \(m\) and \(n - d\) degrees of freedom. The LRT of level \(\alpha\) is given by
\[
\Phi(Y_1, \ldots, Y_n) = \begin{cases} 
1 \quad & \text{if } \frac{||D(\hat{\theta}_n - \hat{\theta}_0^n)||^2}{||Y - D\hat{\theta}_n||^2} > q_{m,n-d,1-\alpha} \\
0 \quad & \text{otherwise},
\end{cases}
\]
whereas \(q_{m,n-d,1-\alpha}\) is the \((1 - \alpha)\)-quantile of \(F_{(m,n-d)}\).

Email address: damark@math.uzh.ch