Chapter 8
Testing

Hypothesis Testing

• A statistical hypothesis test is a method of making decisions using experimental data. A result is called statistically significant if it is unlikely to have occurred by chance.

• These decisions are made using (null) hypothesis tests. A hypothesis can specify a particular value for a population parameter, say \( \theta = \theta_0 \). Then, the test can be used to answer a question like: Assuming \( \theta_0 \) is true, what is the probability of observing a value for the test statistic that is at least as big as the value that was actually observed?

• Uses of hypothesis testing:
  - Check the validity of theories or models.
  - Check if new data can cast doubt on established facts.
Hypothesis Testing

• We will emphasize statistical hypothesis testing under the classical approach (frequentist school).

• There is a Bayesian approach to hypothesis testing. The decisions regarding the parameter $\theta$ are based on the posterior probability —i.e., the conditional probability that is computed after the relevant evidence (the data, $X$) is taken into account. Based on the posterior probabilities associated with different hypothetical values for $\theta$, we assess which hypothesis about $\theta$ is more likely.

\[
p(\theta | X) \propto p(\theta) p(X|\theta). \quad (\propto: \text{proportional})
\]

$p(\theta)$: Prior.

$p(X | \theta)$: Likelihood

Hypothesis Testing

• In general, there are two kinds of hypotheses:
  (1) About the form of the probability distribution

   \textbf{Example:} Is the random variable normally distributed?

  (2) About the parameters of a distribution function

   \textbf{Example:} Is the mean of a distribution equal to 0?

• The second class is the traditional material of econometrics. We may test whether the effect of income on consumption is greater than one, or whether the size coefficient on a CAPM regression is equal to zero.
Hypothesis Testing

- Hypothesis testing involves the comparison between two competing hypothesis (sometimes, they represent partitions of the world).
  - The null hypothesis, denoted $H_0$, is sometimes referred to as the maintained hypothesis.
  - The alternative hypothesis, denoted $H_1$, is the hypothesis that will be considered if the null hypothesis is “rejected.”

- Idea: We collect a sample of data $X_1,\ldots,X_n$. This sample is a multivariate random variable, $E_n$ (an element of an Euclidean space). Then, based on this sample, we follow a decision rule:
  - If the multivariate random variable is contained in space $R$, we reject the null hypothesis.
  - Alternatively, if the random variable is in the complement of the space $R (R^c)$ we fail to reject the null hypothesis.

Hypothesis Testing

- Decision rule:
  
  \[
  \begin{align*}
  \text{if } X \in R, & \quad \Rightarrow \text{ reject } H_0 \\
  \text{if } X \notin R \text{ or } X \in R^c, & \quad \Rightarrow \text{ fail to reject } H_0
  \end{align*}
  \]

  The set $R$ is called the region of rejection or the critical region of the test.

- The rejection region is defined in terms of a statistics $T(X)$, called the test statistic. Note that like any other statistic, $T(X)$ is a random variable. Given this test statistic, the decision rule can then be written as:

  \[
  \begin{align*}
  T(X) \in R & \Rightarrow \text{ reject } H_0 \\
  T(X) \in R^c & \Rightarrow \text{ fail to reject } H_0
  \end{align*}
  \]
Hypothesis Testing: A brief comment

• What we present as classical approach is a synthesized approach.
• Ronald Fisher defined only \( H_0 \). Under his approach we:
  1. Identify \( H_0 \).
  2. Determine the appropriate \( T(X) \) and its distribution under the assumption that \( H_0 \) is true.
  3. Calculate \( T(X) \) from the data.
  4. Determine the achieved significance level that corresponds to the \( T(X) \) using the distribution under the assumption that \( H_0 \) is true.
  5. Reject \( H_0 \) if the achieved significance level is sufficiently small. Otherwise, reach no conclusion.

• This construct leads to the question of what \( p\text{-value} \) is sufficiently small as to warrant rejection of \( H_0 \). Fisher favored 5% or 1%.

Hypothesis Testing: A brief comment

• Neyman and Pearson in their approach added \( H_1 \). Steps:
  1. Identify \( H_0 \) and a complementary hypothesis, \( H_1 \).
  2. Determine the appropriate \( T(X) \) and its distribution under the assumption that \( H_1 \) is true.
  3. Specify a significance level (\( \alpha \)), and determine the corresponding critical value of \( T(X) \) under the assumption that \( H_1 \) is true.
  4. Calculate \( T(X) \) from the data.
  5. Reject \( H_1 \) and accept \( H_0 \) if the \( T(X) \) is further than the critical value from \( E[T(X) | H_0 \text{ true}] \).

• The Neyman-Pearson approach is important in decision theory. The final step is assigned a risk function computed as the expected loss from making an error.
Hypothesis Testing

• There are two types of hypothesis regarding parameters:
  (1) A simple hypothesis. Under this scenario, we test the value of a parameter against a single alternative.
    **Example:** $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$.
  
  (2) A composite hypothesis. Under this scenario, we test whether the effect of income on consumption is greater than one. Implicit in this test is several alternative values.
    **Example:** $H_0: \theta > \theta_0$ against $H_1: \theta < \theta_1$.

• **Definition:** Simple and composite hypotheses
  A hypothesis is called *simple* if it specifies the values of all the parameters of a probability distribution, say $\theta = \theta_0$. Otherwise, it is called *composite*.

Type I and Type II Errors

• **Definition:** Type I and Type II errors
  A *Type I error* is the error of rejecting $H_0$ when it is true. A *Type II error* is the error of accepting $H_0$ when it is false (that is when $H_1$ is true).

• **Notation:**
  - Probability of Type I error: $\alpha = P[X \in R \mid H_0]$
  - Probability of Type II error: $\beta = P[X \in R^c \mid H_1]$

• **Definition:** Power of the test
  The probability of rejecting $H_0$ based on a test procedure is called the *power of the test*. It is a function of the value of the parameters tested, $\theta$:
  $$\pi = \pi(\theta) = P[X \in R].$$

  **Note:** when $\theta \in H_1 \implies \pi(\theta) = 1 - \beta(\theta).$
Type I and Type II Errors

- We want \( \pi(\theta) \) to be near 0 for \( \theta \in H_0 \) and \( \pi(\theta) \) to be near 1 for \( \theta \in H_1 \).

- **Definition**: Level of significance

  When \( \theta \in H_0 \), \( \pi(\theta) \) gives you the probability of Type I error. This probability depends on \( \theta \). The maximum value of this when \( \theta \in H_0 \) is called *level of significance (significance level)* of a test, denoted by \( \alpha \). Thus,

  \[
  \alpha = \sup_{\theta \in H_0} \mathbb{P}[X \in R | H_0] = \sup_{\theta \in H_0} \pi(\theta)
  \]

Define a level \( \alpha \) test to be a test with \( \sup_{\theta \in H_0} \pi(\theta) \leq \alpha \).

Sometimes, \( \alpha \) is called the *size* of a test.

---

<table>
<thead>
<tr>
<th>Decision</th>
<th>State of World</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cannot reject</strong> (&quot;\text{accept}) H_0**</td>
<td><strong>Correct decision</strong></td>
</tr>
<tr>
<td><strong>Reject H_0</strong></td>
<td><strong>Type I error</strong></td>
</tr>
</tbody>
</table>

Need to control both types of error:

\[
\alpha = \mathbb{P}(\text{rejecting } H_0 | H_0) \\
\beta = \mathbb{P}(\text{not rejecting } H_0 | H_1)
\]
Type I and Type II Errors

\[ \beta = \text{Type II error} \]
\[ \alpha = \text{Type I error} \]
\[ \pi = \text{Power of test} \]

Type I and Type II Errors: Example

Example. Let \( X \) have the density
\[
    f(x) = \begin{cases} 
        1 - \theta + x & \text{for } \theta - 1 \leq x < \theta \\
        1 + \theta - x & \text{for } \theta \leq x \leq \theta + 1 
    \end{cases}
\]
This is a triangular probability density function.

We test \( H_0: \theta = 0 \) against \( H_1: \theta = 1 \), using a single observation of \( X \).
Type I and Type II Errors: Example

Type I and Type II errors –i.e., the areas of the isosceles triangles- are then defined by the choice of $t$, the cut off region:

$$\alpha = \frac{1}{2} \left( 1 - t \right)^2$$

$$\beta = \frac{1}{2} t^2$$

Deriving $\beta$ in terms of $\alpha$ yields:  

$$\beta = \frac{1}{2} \left( 1 - \sqrt{2\alpha} \right)^2$$

- The choice of any $t$ yields an admissible test. However, any randomized test is inadmissible.

- Theorem.

The set of admissible characteristics plotted on the $\alpha, \beta$ plane is a continuous, monotonically decreasing, convex function which starts at a point with $[0,1]$ on the $\beta$ axis and ends at a point within the $[0,1]$ on the $\alpha$ axis.
Type I and Type II Errors

- There is a natural trade-off between Type I and Type II errors. It is impossible to minimize both.

- Q: How do we select a test?
Assume that we want to compare two critical regions $R_1$ and $R_2$. Assume that we choose either confidence region $R_1$ or $R_2$ randomly with probabilities $\delta$ and $1-\delta$, respectively. This is called a randomized test.

If the probabilities of the two types of error for $R_1$ and $R_2$ are $(\alpha_1, \beta_1)$ and $(\alpha_2, \beta_2)$ respectively. The probability of each type of error becomes:

$$\alpha = \delta \alpha_1 + (1 - \delta) \alpha_2$$
$$\beta = \delta \beta_1 + (1 - \delta) \beta_2$$

The values $(\alpha, \beta)$ are the characteristics of the test.

More Powerful Test

- **Definition**: More Powerful Test
Let $(\alpha_1, \beta_1)$ and $(\alpha_2, \beta_2)$ be the characteristics of two tests. The first test is more powerful (better) than the second test if $\alpha_1 \leq \alpha_2$, and $\beta_1 \leq \beta_2$ with a strict inequality holding for at least one point.

If we cannot determine that one test is better by the definition, we could consider the relative cost of each type of error. Classical statisticians typically do not consider the relative cost of the two errors because of the subjective nature of this comparison.

**Note**: Bayesian statisticians compare the relative cost of the two errors using a loss function.
Most Powerful Test

- **Definition**: Admissible test
  A test is *inadmissible* if there exits another test, which is better. Otherwise, it is called *admissible*.

- **Definition**: Most powerful test of size $\alpha$
  $R$ is the *most powerful test of size* $\alpha$ if $\alpha(R) = \alpha$ and for any test $R_1$ of size $\alpha$, $\beta(R) \leq \beta(R_1)$.

- **Definition**: Most powerful test of level $\alpha$
  $R$ is the *most powerful test of level* $\alpha$ (that is, such that $\alpha(R) \leq \alpha$) and for any test $R_1$ of level $\alpha$ (that is, $\alpha(R_1) \leq \alpha$), if $\beta(R) \leq \beta(R_1)$.

UMP Test

- **Definition**: Uniformly most powerful (UMP) test
  $R$ is the *uniformly most powerful test of level* $\alpha$ (that is, such that $\alpha(R) \leq \alpha$) and for every test $R_1$ of level $\alpha$ (that is, $\alpha(R_1) \leq \alpha$), if $\pi(R) \leq \pi(R_1)$.

  “For every test”: for alternative values of $\theta_1$ in $H_1; \theta = \theta_1$.

  - Choosing between admissible test statistics in the $(\alpha, \beta)$ plane is similar to the choice of a consumer choosing a consumption point in utility theory. Similarly, the tradeoff problem between $\alpha$ and $\beta$ can be characterized as a ratio.

  - This idea is the basis of the Neyman-Pearson Lemma to construct a test of a hypothesis about $\theta$: $H_0; \theta = \theta_0$ against $H_1; \theta = \theta_1$. 

Neyman-Pearson Lemma

- Neyman-Pearson Lemma provides a procedure for selecting the best test of a simple hypothesis about \( \theta \): \( H_0: \theta = \theta_0 \) against \( H_1: \theta = \theta_1 \).

- Let \( L(x | \theta) \) be the joint density function of \( X \). We determine \( R \) based on the ratio \( L(x | \theta_1) / L(x | \theta_0) \). (This ratio is called the *likelihood ratio.*) The bigger this ratio, the more likely the rejection of \( H_0 \).

Consider testing a simple hypothesis \( H_0: \theta = \theta_0 \) vs. \( H_1: \theta = \theta_1 \), where the pdf corresponding to \( \theta_i \) is \( L(x | \theta_i) \), \( i=0,1 \), using a test with rejection region \( R \) that satisfies

\[
\begin{align*}
(1) \quad & x \in R \text{ if } L(x | \theta_1) > k L(x | \theta_0) \\
& x \in R^c \text{ if } L(x | \theta_1) < k L(x | \theta_0),
\end{align*}
\]

for some \( k \geq 0 \), and

\[
(2) \quad \alpha = P[X \in R | H_0]
\]

Then,

(a) Any test that satisfies (1) and (2) is a UMP level \( \alpha \) test.
(b) If there exists a test satisfying (1) and (2) with \( k > 0 \), then every UMP level \( \alpha \) test satisfies (2) and every UMP level \( \alpha \) test satisfies (1) except perhaps on a set \( A \) satisfying \( P[X \in A | H_0] = P[X \in A | H_1] = 0 \).
Neyman-Pearson Lemma

Note that, if $\alpha = P[X \in R \mid H_0]$, we have a size $\alpha$ test and hence a level $\alpha$ test because $\sup_{\theta \in \Theta_0} P[X \in R \mid H_0] = P[X \in R \mid H_0] = \alpha$, since $\Theta_0$ has only one point.

Define the test function $\phi$ (maps data into chosen hypothesis (1 or 0) as:
\[
\phi(x) = \begin{cases} 
1 & \text{if } x \in R, \\
0 & \text{if } x \in R^c.
\end{cases}
\]

Let $\phi(x)$ be the test function of a test satisfying (1) and (2) and $\phi'(x)$ be the test function for any other level $\alpha$ test, where the corresponding power functions are $\pi(\theta)$ and $\pi'(\theta)$.

Since $0 \leq \phi'(x) \leq 1$, $(\phi(x) - \phi'(x))(L(x \mid \theta_1) - k(L(x \mid \theta_0)) \geq 0$, for every $x$. Thus,
\[
0 \leq \int [\phi(x) - \phi'(x)][L(x \mid \theta_1) - k(L(x \mid \theta_0)]dx = \pi(\theta_1) - \pi'(\theta_1) - k(\pi(\theta_0) - \pi'(\theta_0)).
\]

Neyman-Pearson Lemma

Proof of (a)

(a) is proved by noting $\pi(\theta_0) - \pi'(\theta_0) = \alpha - \pi'(\theta_0) \geq 0$.

Thus with $k \geq 0$ and (3),
\[
0 \leq \pi(\theta_1) - \pi'(\theta_1) - k(\pi(\theta_0) - \pi'(\theta_0)) \leq \pi(\theta_1) - \pi'(\theta_1)
\]

showing $\pi(\theta_1) \geq \pi'(\theta_1)$. Since $\phi'$ is arbitrary and $\theta_1$ is the only point in $\Theta_0$, $\phi$ is an UMP test.
Neyman-Pearson Lemma

Proof of (b)
Now, let φ′ be the test function for any UMP level α test.

By (a), φ, the test satisfying (1) and (2), is also a UMP level α test. Thus, π(θ) = π′(θ). Using this result, (3), and k ≥ 0,
\[ \alpha - \pi′(\theta_0) = \pi(\theta_0) - \pi′(\theta_0) \leq 0. \]

Since φ′ is a level α test, π′(θ_0) ≤ α, that is, φ′ is a size α test implying that (3) is an equality. But the nonnegative integrand in (3) will be 0 only if φ′ satisfies (1) except, perhaps, where
\[ \int \limits A L(x | \theta) \, dx = 0 \text{ on a set A.} \]

Neyman-Pearson Lemma: Example

Let X_1,...,X_n be a random sample from a N(θ, 1) population. Test : H_0 : θ = θ_0 vs H_1 : θ = θ_1.
The Neyman - Pearson lemma is based on the ratio \( \lambda(x) \):
\[ \lambda(x) = \frac{L(\hat{\theta}_0 | x)}{L(\hat{\theta}_l | x)} = \frac{(2\pi)^{n/2} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_0)^2}}{(2\pi)^{n/2} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_1)^2}} = e^{\frac{-\sum_{i=1}^{n} (x_i - \theta_0)^2 + \sum_{i=1}^{n} (x_i - \theta_1)^2}{2}} \]

That is, \( \lambda(x) = e^{\frac{-\sum_{i=1}^{n} (x_i - \theta_0)^2 + \sum_{i=1}^{n} (x_i - \theta_1)^2}{2}} > k \)
\[ \Rightarrow \log \lambda(x) = \frac{-\sum_{i=1}^{n} (x_i - \theta_0)^2 + \sum_{i=1}^{n} (x_i - \theta_1)^2}{2} > \log k \]
\[ \Rightarrow \log \lambda(x) = \frac{-\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i \theta_1 + n(\theta_1^2) + \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i \theta_0 + n(\theta_0^2)}{2} > \log k \]
\[ \Rightarrow \log \lambda(x) = \frac{n \sum_{i=1}^{n} (\theta_1 - \theta_0) + n(\theta_0^2 - \theta_1^2)}{2} = n(\theta_1 - \theta_0) + n \overline{x}(\theta_0^2 - \theta_1^2) / 2 > \log k \]
Neyman-Pearson Lemma: Example

We will reject $H_0$ if $\ln \lambda(x) > \ln k$. But, this reduces to $\bar{x} > d$, where $d$ is selected to give a size $\alpha$ test.

Thus, the critical region is $R = \{ x: \bar{x} > d \}$, and $P[\bar{x} > d | \theta = \theta_0] = \alpha$.

Under $H_{10}$, we have $z = \bar{x} - \theta_0 \sim N(0,1)$

$=>$ $P[\bar{x} > d | \theta = \theta_0] = P[ z > (d - \theta_0) | \theta = \theta_0] = \alpha$.

$=>$ $d = z_{\alpha} + \theta_0$.

$=>$ $R = \{ x: x > z_{\alpha} + \theta_0 \}$.

Note: We reject $H_0$ if the sample mean is greater than $z_{\alpha} + \theta_0$. But, $R$ is independent of $\theta_1$ and it is the same for any $\theta_1 > \theta_0$. Thus, $R$ gives a UMP for $H_{10}: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$.

Monotone Likelihood Ratio

• In general, we have no basis to pick $\theta_1$. We need a procedure to test composite hypothesis, preferably with a UMP.

Definition: Monotone Likelihood Ratio

The model $f(X, \theta)$ has the monotone likelihood ratio property in $u(X)$ if there exists a real valued function $u(X)$ such that the likelihood ratio $\lambda = L(x | \theta) / L(x | \theta_0)$ is a non-decreasing function of $u(X)$ for each choice of $\theta_1$ and $\theta_0$ with $\theta_1 > \theta_0$.

If $L(x | \theta)$ satisfies the MLRP with respect to $L(x | \theta_0)$ the higher the observed value $u(X)$, the more likely it was drawn from distribution $L(x | \theta)$ rather than $L(x | \theta_0)$. 
Monotone Likelihood Ratio

• Consider the exponential family:
  \( L(X; \theta) = \exp\{\sum U(X_i) - A(\theta) \sum T(X_i) + n B(\theta)\} \).
Then, \( \ln \lambda = \sum T(X_i) [A(\theta_1) - A(\theta_0)] + n B(\theta_1) - n B(\theta_0) \).
Let \( u(X) = \sum T(X_i) \).
Then,
  \( \delta \ln \lambda / \delta u = [A(\theta_1) - A(\theta_0)] > 0 \), if \( A(.) \) is monotonic in \( \theta \).
In addition, \( u(X) \) is a sufficient statistic.

• Some distributions with MLRP in \( T(X) = \sum x_i \): normal (with \( \sigma \) known), exponential, binomial, Poisson.

Karlin-Rubin Theorem

Theorem: Karlin-Rubin (KR) Theorem
Suppose we are testing \( H_0: \theta \leq \theta_0 \) vs. \( H_1: \theta > \theta_0 \). Let \( T(X) \) be a sufficient statistic, and the family of distributions \( g(.) \) has the MLRP in \( T(X) \).
Then, for any \( t_0 \) the test with rejection region \( T > t_0 \) is UMP level \( \alpha \), where \( \alpha = \Pr(T > t_0 | \theta_0) \).

Proof:
Let \( \pi(\theta) \) be the power function for the test mentioned in KR.
\( \pi(\theta) \) is nondecreasing, meaning for any \( \theta_1 > \theta_2 \),
  \( \pi(\theta_1) \geq \pi(\theta_2) \)
\( \Pr(T(X) > t_0 | \theta_1) \geq \Pr(T(X) > t_0 | \theta_2) \).
This implies \( \sup_{\theta \in H_0} \pi(\theta) = \pi(\theta_0) \leq \alpha \), so the test is level.
Karlin-Rubin Theorem

Proof (continuation):
Now, consider testing the simple hypotheses $H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta'$, with $\theta_0 > \theta'$.
Define
$$k' = \inf_{t \in \Psi} g(t|\theta')/g(t|\theta_0).$$
where $\Psi$ is the region where $t > t_0$ and at least one of the densities is nonzero. Then, from the MLRP in $T(X)$ of $g$
$$T(X) > t_0 \equiv g(t|\theta')/g(t|\theta_0) > k'$$
Thus, $\pi(\theta)$ satisfies the definition of the test given in the NP Lemma for testing $H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta'$, thus it is the UMP test for those hypotheses. Since $\theta'$ was arbitrary, the test is simultaneously most powerful for every $\theta' > \theta_0$, thus it is UMP level for the composite alternative hypothesis. ■

KR Theorem: Practical Use

Goal: Find the UMP level $\alpha$ test of $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$ (similar for $H_0: \theta \geq \theta_0$ vs. $H_1: \theta < \theta_0$)
1. If possible, find a univariate sufficient statistic $T(X)$. Verify its density has an MLR (might be non-decreasing or non-increasing, just show it is monotonic).
2. KR states the UMP level $\alpha$ test is either 1) reject if $T > t_0$ or 2) reject if $T < t_0$. Which way depends on the direction of the MLR and the direction of $H_1$.
3. Derive $E[T]$ as a function of $\theta$. Choose the direction to reject ($T > t_0$ or $T < t_0$) based on whether $E[T]$ is higher or lower for $\theta$ in $H_1$. If $E[T]$ is higher for values in $H_1$, reject when $T > t_0$, otherwise reject for $T < t_0$. 
KR Theorem: Practical Use

4. $t_0$ is the appropriate percentile of the distribution of $T$ when $\theta = \theta_0$. This percentile is either the $\alpha$ percentile (if you reject for $T < t_0$) or the $1 - \alpha$ percentile (if you reject for $T > t_0$).

Nonexistence of UMP tests

- For most two-sided hypotheses — i.e., $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ — no UMP level test exists.
- Simply intuition: the test which is UMP for $\theta < \theta_0$ is not the same as the test which is UMP for $\theta > \theta_0$. A UMP test must be most powerful across every value in $H_1$.

- Definition: Unbiased Test
  A test is said to be unbiased when
  $\pi(\theta) \geq \alpha$ for all $\theta \in H_1$ and $P[\text{Type I error}] = P[X \in \mathbb{R} | H_0] = \pi(\theta) \leq \alpha$ for all $\theta \in H_0$.

  Unbiased test $\Rightarrow \pi(\theta_0) < \pi(\theta_1)$ for all $\theta_0$ in $H_0$ and $\theta_1$ in $H_1$.

  Most two-sided tests we use are UMP level $\alpha$ unbiased (UMPU) tests.
Some problems left for students

• So far, we have produced UMP level $\alpha$ tests for simple versus simple hypotheses ($H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta_1$) and one sided tests with MLRP ($H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$).

• There are a lot of unsolved problems. In particular,
  (1) We did not cover unbiased tests in detail, but they are often simply combinations of the UMP tests in each directions
  (2) Karlin-Rubin discussed univariate sufficient statistics, which leaves out every problem with more than one parameter (for example testing the equality of means from two populations).
  (3) Every problem without an MLR is left out.

No UMP test

• Power function (again)
  We define the power function as $\pi(\theta) = P[X \in R]$. Ideally, we want $\pi(\theta)$ to be near 0 for $\theta \in H_0$, and $\pi(\theta)$ to be near 1 for $\theta \in H_1$.

  The classical (frequentist) approach is to look in the class of all level $\alpha$ tests (all tests with $\sup_{\theta \in H_0} \pi(\theta) \leq \alpha$) and find the MP one available.

  • In some cases there is a UMP level $\alpha$ test, as given by the Neyman Pearson Lemma (simply hypotheses) and the Karlin Rubin Theorem (one sided alternatives with univariate sufficient statistics with MLRP). But, in many cases, there is no UMP test.

  • When no UMP test exists, we turn to general methods that produce “good” tests.
General Methods

• Likelihood Ratio (LR) Tests
• Bayesian Tests - can be examined for their frequentist properties even if you are not a Bayesian.
• Pivot Tests - Tests based on a function of the parameter and data whose distribution does not depend on unknown parameters. Wald, Score and LR tests are examples of asymptotically pivotal tests.
• Wald Tests - Based on the asymptotic normality of the MLE
• Score tests - Based on the asymptotic normality of the log-likelihood

Pivot Tests

• Pivot Test: A tests whose distribution does not depend on unknown parameters.

• Example: Suppose you draw $X$ from a $N(\mu, \sigma^2)$.

Asymptotic theory implies that $\bar{x}$ is asymptotically $N(\mu, \sigma^2/N)$.
This statistic is not asymptotically pivotal statistic because it depends on an unknown parameter, $\sigma^2$ (even if you specify $\mu_0$ under $H_0$).

On the other hand, the $t$-statistic, $t = (\bar{x} - \mu_0)/s$ is asymptotically $N(0, 1)$.
This is asymptotically pivotal since 0 and 1 are known!

Most statistics are not asymptotically pivotal. Many popular test statistics -for example, Wald, LR- are asymptotically pivotal because they are distributed as $\chi^2$ with known df or follow an $N(0, 1)$ distribution.
**Likelihood Ratio Tests**

• Define the likelihood ratio (LR) statistic

\[ \lambda(X) = \sup_{\theta \in H_0} L(X | \theta) / \sup_{\theta} L(X | \theta) \]

**Note:**
Numerator: maximum of the LF within \( H_0 \)
Denominator: maximum of the LF within the entire parameter space, which occurs at the MLE.

• Reject \( H_0 \) if \( \lambda(X) < k \), where \( k \) is determined by
  \[ \text{Prob}[0 < \lambda(X) < k | \theta \in H_0] = \alpha. \]

**Properties of the LR statistic \( \lambda(X) \)**

• Properties of \( \lambda(X) = \sup_{\theta \in H_0} L(X | \theta) / \sup_{\theta} L(X | \theta) \)

  1. \( 0 \leq \lambda(X) \leq 1 \), with \( \lambda(X) = 1 \) if the supremum of the likelihood occurs within \( H_0 \).

  **Intuition of test:** If the likelihood is much larger outside \( H_0 \) - i.e., in the unrestricted space -, then \( \lambda(X) \) will be small and \( H_0 \) should be rejected.

  2. Under general assumptions, \( -2 \ln \lambda(X) \sim \chi^2_p \), where \( p \) is the difference in df between the \( H_0 \) and the general parameter space.

  3. For simple hypotheses, the numerator and denominator of the LR test are simply the likelihoods under \( H_0 \) and \( H_1 \). The LR test reduces to a test specified by the NP Lemma.
Likelihood Ratio Tests: Example I

Example: $\lambda(x)$ for a $X \sim N(\theta, \sigma^2)$ for $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$. Assume $\sigma^2$ is known.

$$\lambda(x) = \frac{L(\theta_0 | x)}{L(x | x)} = \frac{(2\pi)^{-n/2} e^{-\frac{\sum (x - \theta_0)^2}{2\sigma^2}}}{e^{-\frac{n}{2} \sum (x - \bar{x})^2 / 2\sigma^2}} = e^{\frac{-\sum (x - \theta_0)^2}{2\sigma^2}}$$

Reject $H_0$ if $\lambda(x) < k$ $\Rightarrow$ $\ln \lambda(x) = \frac{-n(\bar{x} - \theta_0)^2}{2\sigma^2} < \ln k$ $\Rightarrow$ $\frac{(\bar{x} - \theta_0)^2}{\sigma^2} > -2\ln k$

Note: Finding $k$ is not needed. Why? We know the left hand side is distributed as a $\chi^2_p$, thus $(-2 \ln k)$ needs to be the 1 - $\alpha$ percentile of a $\chi^2_p$. We need not solve explicitly for $k$, we just need the rejection rule.

Likelihood Ratio Tests: Example II

Example: $\lambda(x)$ for a $X \sim \text{exponential} (\lambda)$ for $H_0: \lambda = \lambda_0$ vs. $H_1: \lambda \neq \lambda_0$.

$L(X | \lambda) = \lambda^n \exp(-\lambda \sum x) = \lambda^n \exp(-\lambda n \bar{x})$ $\Rightarrow$ $\lambda_{\text{MLE}} = 1/\bar{x}$

$$\lambda(x) = \frac{\lambda_0^n e^{-\lambda_0 \bar{x}}}{(1/\bar{x}) \lambda^n e^{-n}} = (\bar{x} \lambda_0)^n e^{(n(1 - \lambda_0 \bar{x}))}$$

Reject $H_0$ if $\lambda(x) < k$ $\Rightarrow$ $\ln \lambda(x) = n \ln (\bar{x} \lambda_0) + n(1 - \lambda_0 \bar{x}) < \ln k$

We need to find $k$ such that $P[\lambda(x) < k] = \alpha$. Unfortunately, this is not analytically feasible. We know the distribution of $\bar{x}$ is Gamma$(n, \lambda/n)$, but we cannot get further.

It is, however, possible to determine the cutoff point, $k$, by simulation (set $n, \lambda_0$).
Asymptotic Distribution of the LRT – Simple $H_0$

Theorem: Test $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$. Suppose $X_1,...,X_n$ are iid with pdf $f(x|\theta)$.

Let $\hat{\theta}$ be the MLE of $\theta$, and $f(x|\theta)$ satisfies the following regularity conditions:

1. The parameter is identifiable - i.e., if $\theta \neq \theta'$, then $f(x|\theta) \neq f(x|\theta')$.
2. The densities $f(x|\theta)$ have some common support, and $f(x|\theta)$ is differentiable in $\theta$.
3. The parameters space $\Theta$ contains an open set $\omega$ of which the true parameter value $\theta_0$ is an interior point.
4. $\forall x \in X$ the density $f(x|\theta)$ is three times differentiable with respect to $\theta$, the third derivative is continuous in $\theta$, and $\int f(x|\theta)dx$ can be differentiated three times under the integral sign.
5. $\forall \theta \in \Theta, 3c > 0$ and a function $M(x)$ (both depend on $\theta_0$) such that:

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(x|\theta) \right| \leq M(x) \forall x \in X, \theta_0 - c < \theta < \theta_0 + c,$$

with $E_{\theta_0}[M(X)] < \infty$.

Asymptotic Distribution of the LRT – Simple $H_0$

Then under $H_0$, as $n \to \infty$,

$$-2 \log \hat{\lambda}(X) = -2[\log L(X, \theta_0) - \log L(X, \hat{\theta}_n)] \xrightarrow{D} \chi^2_1$$

If $\theta$ is a vector in $\Theta_0$ $\Rightarrow$ $-2 \log \hat{\lambda}(X) \xrightarrow{D} \chi^2_p$, $p : [\text{# of free parameters under } \theta \in \Theta_0] - [\text{# of free parameters under } \theta \in \Theta]$.

Proof: Expand $L(x|\theta)$ around $\hat{\theta}_n$, the MLE.

$$\log L(X, \theta) = \log L(X, \hat{\theta}_n) + nS_n(X, \theta_0)(\theta - \hat{\theta}_n) + \frac{1}{2}(\theta - \hat{\theta}_n)^T nS''_n(X, \hat{\theta}_n)(\theta - \hat{\theta}_n)$$

At $\hat{\theta}_n$, $S_n(X, \hat{\theta}_n) = 0$. Then, at $\theta = \theta_0$ $\Rightarrow$ $\log \hat{\lambda}(X) = \log L(X, \theta_0) - \log L(X, \hat{\theta}_n)$

$$= \frac{1}{2} n(\theta_0 - \hat{\theta}_n)^T nS''_n(X, \hat{\theta}_n)(\theta_0 - \hat{\theta}_n)$$

Since $S''_n(X, \hat{\theta}_n) \xrightarrow{p} I(\theta_0)$ and $n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$. Then, $-2 \log \hat{\lambda}(X) \sim \chi^2_p$