557: MATHEMATICAL STATISTICS II
HYPOTHESIS TESTING

A statistical hypothesis test is a **decision rule** that takes as an input observed sample data and returns an action relating to two mutually exclusive **hypotheses** that reflect two competing hypothetical states of nature. The decision rule partitions the sample space $\mathcal{X}$ into two regions that respectively reflect support for the two hypotheses. The following terminology is used:

- Two **hypotheses** characterize the two possible states of nature. The **null hypothesis** is denoted H_0, the **alternative hypothesis** is denoted H_1.
- In parametric models, the null and alternative hypotheses define a partition of the (effective) parameter space Θ. Suppose that disjoint subsets Θ_0, Θ_1 correspond to H_0 and H_1 respectively. We write

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

- A **test**, $\mathcal{T}$, of H_0 versus H_1 defines a partition of sample space $\mathcal{X}$ into two regions. The hypothesis H_0 is **rejected in favour of** H_1 in the test depending on where the data $\underset{\sim}{x}$ (or a suitably chosen statistic $T(\underset{\sim}{x})$) fall within $\mathcal{X}$.
- A **test statistic**, $T(\underset{\sim}{x})$, is the function of data $\underset{\sim}{x}$ used in a statistical hypothesis test.
- The **critical region**, $\mathcal{R}$, is the region within which $T(\underset{\sim}{x})$ must lie in order for hypothesis H_0 to be rejected in favour of H_1 The complement of $\mathcal{R}$ will be written $\mathcal{R}'$.
- A **test function**, $\phi_{\mathcal{R}}(T(\underset{\sim}{x}))$, is an indicator function that reports the result of the test,

$$\phi_{\mathcal{R}}(T(\underset{\sim}{x})) = \begin{cases} 1 & T(\underset{\sim}{x}) \in \mathcal{R} \\ 0 & T(\underset{\sim}{x}) \in \mathcal{R}' \end{cases}$$

- A **Type I error** occurs when the null hypothesis H_0 is **rejected** when it is in fact **true**.
- A **Type II error** occurs when the null hypothesis H_0 is **accepted** when it is in fact **false**.
- For test with test statistic T and critical region $\mathcal{R} \subset \mathcal{X}$, and $\theta \in \Theta_0$, define the **Type I error probability** $\xi(\theta)$ by

$$\xi(\theta) = \Pr\left[T \in \mathcal{R} | \theta\right] \qquad \theta \in \Theta_0 \tag{1}$$

  If Θ_0 comprises a single value, then

$$\xi = \Pr\left[T \in \mathcal{R} | \theta = \theta_0\right]$$

- The **size** of a statistical test is

$$\overline{\alpha} = \sup_{\theta \in \Theta_0} \xi(\theta)$$

  which is equal to ξ if Θ_0 comprises a single value.
- Suppose $\alpha \geq \overline{\alpha}$. If $T(\underset{\sim}{x}) \in \mathcal{R}$, then H_0 is rejected at **level** α, and rejected at level $\alpha + \epsilon$ for $\epsilon > 0$.
- The **power function**, $\beta(\theta)$, is defined by

$$\beta(\theta) = \Pr\left[T \in \mathcal{R} | \theta\right] \qquad \theta \in \Theta$$

  so that $\beta(\theta) = \xi(\theta)$ for $\theta \in \Theta_0$.

  *Note that this notation is not universally used; commonly the **power** of a statistical test is denoted $1 - \beta(\theta)$ and computed for $\theta \in \Theta_1$, whereas the **Type II error probability** is $\beta(\theta)$ for $\theta \in \Theta_1$.*

Most Powerful Tests: The Neyman-Pearson Lemma

To construct and assess the quality of a statistical test, we consider the power function $\beta(\theta)$. Consider a family tests $\mathcal{C}$ for testing H_0 and H_1 with corresponding subsets Θ_0 and Θ_1.

- The **uniformly most powerful (UMP)** test $\mathcal{T}$ is the test whose power function $\beta(\theta)$ dominates the power function, $\beta^\dagger(\theta)$, of any other test $\mathcal{T}^\dagger \in \mathcal{C}$ at all $\theta \in \Theta_1$,

$$\beta(\theta) \geq \beta^\dagger(\theta) \qquad \forall\, \theta \in \Theta_1.$$

- A test with power function $\beta(\theta)$ is **unbiased** if

$$\beta(\theta_1) \geq \beta(\theta_0) \qquad \text{for all } \theta_0 \in \Theta_0, \theta_1 \in \Theta_1$$

- A **simple hypothesis** is one which specifies the distribution of the data completely. Consider a parametric model $f_{X|\theta}(x|\theta)$ with parameter space $\Theta = \{\theta_0, \theta_1\}$, and the test of

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

  Then both H_0 and H_1 are simple hypotheses.

- A parametric model $f_{X|\theta}(x|\theta)$ for $\theta \in \Theta$ is **identifiable** if

$$f_{X|\theta}(x|\theta_0) = f_{X|\theta}(x|\theta_1) \quad \text{for all } x \in \mathbb{R} \qquad \Longleftrightarrow \qquad \theta_0 = \theta_1.$$

Theorem (The Neyman-Pearson Lemma)
Consider a parametric model $f_{X|\theta}(x|\theta)$ with parameter space $\Theta = \{\theta_0, \theta_1\}$. A test of

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

is required. Consider a test $\mathcal{T}$ with rejection region $\mathcal{R}$ that satisfies

$$\begin{aligned} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) > k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0) &\implies \underset{\sim}{x} \in \mathcal{R} \\ f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) < k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0) &\implies \underset{\sim}{x} \in \mathcal{R}' \end{aligned}$$

for some $k \geq 0$, and $\Pr[\underset{\sim}{X} \in \mathcal{R}|\theta = \theta_0] = \alpha$. Then $\mathcal{T}$ is UMP in the class, $\mathcal{C}_\alpha$, of tests at level α. Further, if such a test exists with $k > 0$, then **all** tests at level α also have size α (that is, α is the least upper bound of the power function $\beta(\theta)$), and have rejection region identical to that of $\mathcal{T}$, except perhaps if $\underset{\sim}{x} \in A$ and

$$\Pr[\underset{\sim}{X} \in A|\theta = \theta_0] = \Pr[\underset{\sim}{X} \in A|\theta = \theta_1] = 0.$$

Proof As $\Pr[\underset{\sim}{X} \in \mathcal{R}|\theta = \theta_0] = \alpha$, the test $\mathcal{T}$ has size and level α. Consider the test function $\phi_\mathcal{R}(\underset{\sim}{x})$ for this test, and $\phi_{\mathcal{R}^\dagger}(\underset{\sim}{x})$ be the test function for any other α level test, $\mathcal{T}^\dagger$. Denote by $\beta(\theta)$ and $\beta^\dagger(\theta)$ be the power functions for these two tests. Now

$$g(\underset{\sim}{x}) = (\phi_\mathcal{R}(\underset{\sim}{x}) - \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}))(f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) - k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0)) \geq 0$$

as

$$\begin{aligned} \underset{\sim}{x} \in \mathcal{R} \cap \mathcal{R}^\dagger &\implies \phi_\mathcal{R}(\underset{\sim}{x}) = \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}) = 1 \ \therefore\ g(\underset{\sim}{x}) = 0 \\ \underset{\sim}{x} \in \mathcal{R} \cap \mathcal{R}^{\dagger\prime} &\implies \phi_\mathcal{R}(\underset{\sim}{x}) = 1, \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}) = 0, f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) > k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0) \ \therefore\ g(\underset{\sim}{x}) > 0 \\ \underset{\sim}{x} \in \mathcal{R}' \cap \mathcal{R}^\dagger &\implies \phi_\mathcal{R}(\underset{\sim}{x}) = 0, \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}) = 1, f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) < k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0) \ \therefore\ g(\underset{\sim}{x}) > 0 \\ \underset{\sim}{x} \in \mathcal{R}' \cap \mathcal{R}^{\dagger\prime} &\implies \phi_\mathcal{R}(\underset{\sim}{x}) = \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}) = 0 \ \therefore\ g(\underset{\sim}{x}) = 0. \end{aligned}$$

Thus

$$\int_{\mathcal{X}} (\phi_{\mathcal{R}}(\underset{\sim}{x}) - \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}))(f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) - k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0))\, d\underset{\sim}{x} \geq 0$$

but this inequality can be written in terms of the power functions as

$$(\beta(\theta_1) - \beta^\dagger(\theta_1)) - k(\beta(\theta_0) - \beta^\dagger(\theta_0)) \geq 0 \tag{2}$$

As $\beta(\theta)$ and $\beta^\dagger(\theta)$ are bounded above by α, and $\beta(\theta_0) = \alpha$ as $\mathcal{T}$ is a size α, we have that

$$\beta(\theta_0) - \beta^\dagger(\theta_0) = \alpha - \beta^\dagger(\theta_0) \geq 0 \qquad \therefore \qquad \beta(\theta_1) - \beta^\dagger(\theta_1) \geq 0$$

Thus $\beta(\theta_1) \geq \beta^\dagger(\theta_1)$, and hence $\mathcal{T}$ is UMP, as θ_1 is the only point in Θ_1, and the test with power function $\beta^\dagger$ is arbitrarily chosen.

Now consider any UMP test $\mathcal{T}^\dagger \in \mathcal{C}_\alpha$. By the result above, $\mathcal{T}$ is UMP at level α, so $\beta(\theta_1) = \beta^\dagger(\theta_1)$. In this case, if $k > 0$, we have from equation (2) that

$$\beta(\theta_0) - \beta^\dagger(\theta_0) = \alpha - \beta^\dagger(\theta_0) \leq 0.$$

But, by assumption, $\mathcal{T}^\dagger$ is a level α test, so we also have

$$\alpha - \beta^\dagger(\theta_0) \geq 0$$

and hence $\beta^\dagger(\theta_0) = \alpha$, that is, $\mathcal{T}^\dagger$ is also a size α test. Therefore

$$\int_{\mathcal{X}} (\phi_{\mathcal{R}}(\underset{\sim}{x}) - \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}))(f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) - k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0))\, d\underset{\sim}{x} = 0 \tag{3}$$

where the integrand in equation (3) is a non-negative function. Let $\mathcal{A}$ be the collection of sets of probability (that is, density) zero under both $f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0)$ and $f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1)$, then

$$\int_{A} (\phi_{\mathcal{R}}(\underset{\sim}{x}) - \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}))(f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1) - k f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0))\, d\underset{\sim}{x} = 0 \qquad A \in \mathcal{A}$$

irrespective of the nature of $\mathcal{R}^\dagger$, so the functions $\phi_{\mathcal{R}}(\underset{\sim}{x})$ and $\phi_{\mathcal{R}^\dagger}(\underset{\sim}{x})$ may not be equal for $\underset{\sim}{x}$ in such a set A. Apart from that specific case, the integral in equation (3) can only be zero if at least one of the two factors is identically zero for all $\underset{\sim}{x}$. The second factor cannot be identically zero for all $\underset{\sim}{x}$, as the densities must integrate to one. Thus, for all $\underset{\sim}{x} \in \mathcal{X} \setminus \mathcal{A}$

$$\phi_{\mathcal{R}}(\underset{\sim}{x}) = \phi_{\mathcal{R}^\dagger}(\underset{\sim}{x}),$$

and hence $\mathcal{R}^\dagger$, satisfies the same conditions as $\mathcal{R}$. ■

- To evaluate the value of constant k that appears in the Theorem, we need to compute $\Pr\left[\underset{\sim}{X} \in \mathcal{R}|\theta_0\right]$ for a fixed level/size α.
- It is possible that, for given alternative hypotheses, no UMP test exists. Also, for discrete data, it may not be possible to solve the equation $\Pr\left[\underset{\sim}{X} \in \mathcal{R}|\theta_0\right] = \alpha$ for every value of α, and hence only specific values of α may be attained.
- The test can be reformulated in terms of the statistic $\lambda(\underset{\sim}{x})$ where

  $$\lambda(\underset{\sim}{x}) = \frac{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1)}{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0)}$$

  where $\underset{\sim}{x} \in \mathcal{R} \iff \lambda(\underset{\sim}{x}) \in \mathcal{R}_\lambda$, where $\mathcal{R}_\lambda \equiv \{t \in \mathbb{R}^+ : t > k\}$

- If $T(\underset{\sim}{X})$ is a sufficient statistic for θ, then by the Neyman factorization theorem

$$\frac{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_1)}{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta_0)} = \frac{g(T(\underset{\sim}{x})|\theta_1)h(\underset{\sim}{x})}{g(T(\underset{\sim}{x})|\theta_0)h(\underset{\sim}{x})} = \frac{g(T(\underset{\sim}{x})|\theta_1)}{g(T(\underset{\sim}{x})|\theta_0)}$$

so that

$$\lambda(\underset{\sim}{x}) \in \mathcal{R}_\lambda \qquad \Longleftrightarrow \qquad T(\underset{\sim}{x}) \in \mathcal{R}_T$$

say. Thus any test based on $T(\underset{\sim}{x})$ with critical region $\mathcal{R}_T$ is a UMP α level test, and

$$\alpha = \Pr[T(\underset{\sim}{X}) \in \mathcal{R}_T|\theta_0]$$

Composite Null Hypotheses

Often the null and alternative hypotheses do not specify the distribution of the data completely. For example, the specification

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned}$$

could be of interest. If, in general, a UMP test of size α is required, then its power must equal the power of the most powerful test of

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

for all $\theta_1 \in \Theta_1$.

For one class of models, finding UMP tests for composite hypotheses is possible in general. A parametric family $\mathcal{F}$ of probability models indexed by parameter $\theta \in \Theta$ has a **monotone likelihood ratio** if for $\theta_2 > \theta_1$, and for x in the union of the supports of the two densities $f_{X|\theta}(x|\theta_1)$ and $f_{X|\theta}(x|\theta_2)$,

$$\lambda(x) = \frac{f_{X|\theta}(x|\theta_2)}{f_{X|\theta}(x|\theta_1)}$$

is a monotone function of x.

Theorem (Karlin-Rubin Theorem)

Suppose that a test of the hypotheses

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0 \end{aligned}$$

is required. Suppose that $T(\underset{\sim}{X})$ is a sufficient statistic for θ, and that $f_{T|\theta}$ for $\theta \in \Theta$ has a monotone non-decreasing likelihood ratio, that is for $\theta_2 \geq \theta_1$ and $t_2 \geq t_1$

$$\frac{f_{T|\theta}(t_2|\theta_2)}{f_{T|\theta}(t_2|\theta_1)} \geq \frac{f_{T|\theta}(t_1|\theta_2)}{f_{T|\theta}(t_1|\theta_1)}.$$

Then for any t_0, the test $\mathcal{T}$ with critical region $\mathcal{R}_T$ defined by

$$\begin{aligned} T(\underset{\sim}{x}) > t_0 &\implies T(\underset{\sim}{x}) \in \mathcal{R}_T \\ T(\underset{\sim}{x}) \leq t_0 &\implies T(\underset{\sim}{x}) \in \mathcal{R}_T' \end{aligned}$$

is a UMP α level test, where

$$\alpha = \Pr[T > t_0|\theta_0].$$

Proof Let $\beta(\theta)$ be the power function of $\mathcal{T}$. Now, for $t_2 \geq t_1$,

$$\frac{f_{T|\theta}(t_2|\theta_2)}{f_{T|\theta}(t_2|\theta_1)} \geq \frac{f_{T|\theta}(t_1|\theta_2)}{f_{T|\theta}(t_1|\theta_1)} \qquad \Longleftrightarrow \qquad f_{T|\theta}(t_1|\theta_1)f_{T|\theta}(t_2|\theta_2) \geq f_{T|\theta}(t_1|\theta_2)f_{T|\theta}(t_2|\theta_1) \tag{4}$$

Integrating both sides with respect to t_1 on $(-\infty, t_2)$, we obtain

$$F_{T|\theta}(t_2|\theta_1)f_{T|\theta}(t_2|\theta_2) \geq F_{T|\theta}(t_2|\theta_2)f_{T|\theta}(t_2|\theta_1) \qquad \therefore \qquad \frac{f_{T|\theta}(t_2|\theta_2)}{f_{T|\theta}(t_2|\theta_1)} \geq \frac{F_{T|\theta}(t_2|\theta_2)}{F_{T|\theta}(t_2|\theta_1)}.$$

Alternatively, integrating both sides of equation (4) with respect to t_2 on (t_1, ∞), we similarly obtain

$$\frac{f_{T|\theta}(t_1|\theta_2)}{f_{T|\theta}(t_1|\theta_1)} \leq \frac{1 - F_{T|\theta}(t_1|\theta_2)}{1 - F_{T|\theta}(t_1|\theta_1)}$$

But setting $t_1 = t_2 = t$ in these two inequalities yields

$$\frac{1 - F_{T|\theta}(t|\theta_2)}{1 - F_{T|\theta}(t|\theta_1)} \geq \frac{F_{T|\theta}(t|\theta_2)}{F_{T|\theta}(t|\theta_1)}$$

which, on rearrangement yields

$$\frac{1 - F_{T|\theta}(t|\theta_2)}{F_{T|\theta}(t|\theta_2)} \geq \frac{1 - F_{T|\theta}(t|\theta_1)}{F_{T|\theta}(t|\theta_1)} \qquad \therefore \qquad F_{T|\theta}(t|\theta_2) \leq F_{T|\theta}(t|\theta_1) \tag{5}$$

as $F_{T|\theta}(t|\theta)$ is non-decreasing in t, and the function $g(x) = (1-x)/x$ is non-increasing for $0 < x < 1$. Finally,

$$\beta(\theta_2) - \beta(\theta_1) = \Pr[T > t_0|\theta_2] - \Pr[T > t_0|\theta_1] = (1 - F_{T|\theta}(t|\theta_2)) - (1 - F_{T|\theta}(t|\theta_1)) = F_{T|\theta}(t|\theta_1) - F_{T|\theta}(t|\theta_2) \geq 0$$

so $\beta(\theta)$ is non-decreasing in θ. Hence

$$\sup_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = \Pr[T > t_0|\theta_0] = \alpha$$

so $\mathcal{T}$ is an α level test. Now, let $\theta^\star > \theta_0$, and consider the simple hypotheses

$$\begin{aligned} H_0^\star &: \theta = \theta_0 \\ H_1^\star &: \theta = \theta^\star. \end{aligned}$$

Let $k^\star$ be defined by

$$k^\star = \inf_{t \in \mathcal{T}_0} \frac{f_{T|\theta}(t|\theta^\star)}{f_{T|\theta}(t|\theta_0)}$$

where $\mathcal{T}_0 = \{t : t > t_0, \text{ and } f_{T|\theta}(t|\theta^\star) > 0 \text{ or } f_{T|\theta}(t|\theta_0) > 0\}$. Then

$$T > t_0 \quad \Longleftrightarrow \quad \frac{f_{T|\theta}(t|\theta^\star)}{f_{T|\theta}(t|\theta_0)} > k^\star$$

so that, by the Neyman-Pearson Lemma, $\mathcal{T}$ is UMP for testing $H_0^\star$ versus $H_1^\star$; for **any** other test $\mathcal{T}^\star$ of $H_0^\star$ at level α with power function $\beta^\star$ that satisfies $\beta^\star(\theta_0) \leq \alpha$, we have that $\beta(\theta^\star) \geq \beta^\star(\theta^\star)$. But for any α level test $\mathcal{T}^\dagger$ of H_0, we have $\beta^\dagger(\theta_0) \leq \alpha$. Thus taking $\mathcal{T}^\star \equiv \mathcal{T}^\dagger$, we can conclude that

$$\beta(\theta^\star) \geq \beta^\dagger(\theta^\star).$$

This inequality holds for all $\theta^\star \in \Theta_1$, so $\mathcal{T}$ must be UMP at level α. ∎

The Likelihood Ratio Test

The **Likelihood Ratio Test (LRT)** statistic for testing H_0 against H_1

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

is based on the statistic

$$\lambda_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{\sup\limits_{\theta \in \Theta_0} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)}{\sup\limits_{\theta \in \Theta_1} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)} = \frac{L(\widehat{\theta}_0 \mid \underset{\sim}{x})}{L(\widehat{\theta}_1 \mid \underset{\sim}{x})}$$

and H_0 is **rejected** if $\lambda_{\underset{\sim}{X}}(\underset{\sim}{x})$ is **small enough**, that is, $\lambda_{\underset{\sim}{X}}(\underset{\sim}{x}) \leq k$ for some k to be defined.

Theorem If $T(\underset{\sim}{X})$ is a sufficient statistic for θ, then

$$\lambda_{\underset{\sim}{X}}(\underset{\sim}{x}) = \lambda_T(T(\underset{\sim}{x})) = \frac{\sup\limits_{\theta \in \Theta_0} f_{T|\theta}(T(\underset{\sim}{x})|\theta)}{\sup\limits_{\theta \in \Theta_1} f_{T|\theta}(T(\underset{\sim}{x})|\theta)} \qquad \forall\, \underset{\sim}{x} \in \mathcal{X}$$

Proof As $T(\underset{\sim}{X})$ is sufficient, for any θ_0, θ_1,

$$\frac{L(\theta_0 \mid \underset{\sim}{x})}{L(\theta_1 \mid \underset{\sim}{x})} = \frac{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}\,|\theta_0)}{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}\,|\theta_1)} = \frac{g(T(\underset{\sim}{x})|\theta_0)h(\underset{\sim}{x})}{g(T(\underset{\sim}{x});\theta_1)h(\underset{\sim}{x})} = \frac{g(T(\underset{\sim}{x})|\theta_0)}{g(T(\underset{\sim}{x})|\theta_1)} = \frac{f_{T|\theta}(T(\underset{\sim}{x})|\theta_0)}{f_{T|\theta}(T(\underset{\sim}{x})|\theta_1)}$$

by the Neyman factorization theorem, where the last equality follows as the normalizing constants in numerator and denominator are identical. Hence, at the suprema, the LRT statistics are equal. ■

Union and Intersection Tests

- Suppose first that we require a test $\mathcal{T}$ for the null hypothesis expressed as

  $$H_0 \,:\, \theta \in \Theta_0 \equiv \bigcap_{\gamma \in \mathcal{G}} \Theta_\gamma$$

  where $\Theta_\gamma, \gamma \in \mathcal{G}$ are a collection of subsets of Θ. Suppose that $\mathcal{T}_\gamma$ is a test for the hypotheses

  $$\begin{aligned} H_{0\gamma} &: \theta \in \Theta_\gamma \\ H_{1\gamma} &: \theta \in \Theta'_\gamma \end{aligned}$$

  with test statistic $T_\gamma(\underset{\sim}{X})$ and critical region $\mathcal{R}_\gamma$. Then the rejection region for $\mathcal{T}$ is

  $$\mathcal{R}_{\mathcal{G}} \equiv \bigcup_{\gamma \in \mathcal{G}} \mathcal{R}_\gamma \qquad \Longrightarrow \qquad \mathcal{T} \text{ rejects } H_0 \text{ if } \underset{\sim}{x} \in \bigcup_{\gamma \in \mathcal{G}} \{T_\gamma(\underset{\sim}{x}) \in \mathcal{R}_\gamma\}$$

  that is, if **any one** of the $\mathcal{T}_\gamma$ rejects $H_{0\gamma}$. This test is termed a **Union-Intersection Test (UIT)**.
- Suppose now that we require a test $\mathcal{T}$ for the null hypothesis expressed as

  $$H_0 \,:\, \theta \in \Theta_0 \equiv \bigcup_{\gamma \in \mathcal{G}} \Theta_\gamma$$

  Then, by the same logic as above, the rejection region for $\mathcal{T}$ is

  $$\mathcal{R}_{\mathcal{G}} \equiv \bigcap_{\gamma \in \mathcal{G}} \mathcal{R}_\gamma \qquad \Longrightarrow \qquad \mathcal{T} \text{ rejects } H_0 \text{ if } \underset{\sim}{x} \in \bigcap_{\gamma \in \mathcal{G}} \{T_\gamma(\underset{\sim}{x}) \in \mathcal{R}_\gamma\}$$

that is, if **all** of the $\mathcal{T}_\gamma$ reject $H_{0\gamma}$. This test is termed an **Intersection-Union Test (IUT)**. Note that if α_γ is the size of the test of $H_{0\gamma}$, then the IUT is a level α test, where

$$\alpha = \sup_{\gamma \in \mathcal{G}} \alpha_\gamma$$

as, for each γ and for any $\theta \in \Theta_0$,

$$\alpha \geq \alpha_\gamma = \Pr[\underset{\sim}{X} \in \mathcal{R}_\gamma|\theta] \geq \Pr[\underset{\sim}{X} \in \mathcal{R}|\theta]$$

Theorem Consider testing

$$\begin{aligned} H_0 \quad &: \quad \theta \in \Theta_0 \equiv \bigcap_{\gamma \in \mathcal{G}} \Theta_\gamma \\ H_1 \quad &: \quad \theta \in \Theta_0' \end{aligned}$$

using the global likelihood ratio statistic

$$\lambda(\underset{\sim}{x}) = \frac{\sup\limits_{\theta \in \Theta_0} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)}{\sup\limits_{\theta \in \Theta_1} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)}$$

equipped with the usual critical region $\mathcal{R} \equiv \{\underset{\sim}{x} : \lambda(\underset{\sim}{x}) < c\}$, and the collection of likelihood ratio statistics $\lambda_\gamma(\underset{\sim}{x})$

$$\lambda_\gamma(\underset{\sim}{x}) = \frac{\sup\limits_{\theta \in \Theta_{0\gamma}} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)}{\sup\limits_{\theta \in \Theta_{1\gamma}} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)}$$

Define statistic $T(\underset{\sim}{x}) = \inf\limits_{\gamma \in \mathcal{G}} \lambda_\gamma(\underset{\sim}{x})$, and consider the critical region

$$\mathcal{R}_\mathcal{G} \equiv \{\underset{\sim}{x} : \lambda_\gamma(\underset{\sim}{x}) < c, \text{some } \gamma \in \mathcal{G}\} \equiv \{\underset{\sim}{x} : T(\underset{\sim}{x}) < c\},$$

Then

(a) $T(\underset{\sim}{x}) \geq \lambda(\underset{\sim}{x})$ for all $\underset{\sim}{x}$.

(b) If β_T and β_λ are the power functions for the tests based on $T(\underset{\sim}{X})$ and $\lambda(\underset{\sim}{X})$ respectively, then

$$\beta_T(\theta) \leq \beta_\lambda(\theta) \qquad \text{for all } \theta \in \Theta$$

(c) If the test based on $\lambda(\underset{\sim}{X})$ is an α level test, then the test based on $T(\underset{\sim}{X})$ is also an α level test.

Proof For (a), as $\Theta_0 \subset \Theta_\gamma$, we have

$$\lambda_\gamma(\underset{\sim}{x}) \geq \lambda(\underset{\sim}{x}) \text{ for each } \gamma \qquad \therefore \qquad T(\underset{\sim}{x}) = \inf_{\gamma \in \mathcal{G}} \lambda_\gamma(\underset{\sim}{x}) \geq \lambda(\underset{\sim}{x})$$

and thus for (b), for any θ,

$$\beta_T(\theta) = \Pr[T(\underset{\sim}{X}) < c|\theta] \leq \Pr[\lambda(\underset{\sim}{X}) < c|\theta] = \beta_\lambda(\theta).$$

Hence

$$\sup_{\theta \in \Theta_0} \beta_T(\theta) \leq \sup_{\theta \in \Theta_0} \beta_\lambda(\theta) \leq \alpha$$

which proves (c). ■

P-values

Consider a test of hypothesis H_0 defined by region Θ_0 of the parameter space. A **p-value**, $p(\underset{\sim}{X})$, is a test statistic such that $0 \leq p(\underset{\sim}{x}) \leq 1$ for each $\underset{\sim}{x}$. A p-value is **valid** if, for every $\theta \in \Theta_0$ and $0 \leq \alpha \leq 1$

$$\Pr[p(\underset{\sim}{X}) \leq \alpha \mid \theta] \leq \alpha.$$

That is, a valid p-value is a test statistic that produces a test at level α of the form

$$\begin{aligned} p(\underset{\sim}{x}) \leq \alpha \quad &\Longrightarrow \quad \underset{\sim}{x} \in \mathcal{R} \\ p(\underset{\sim}{x}) > \alpha \quad &\Longrightarrow \quad \underset{\sim}{x} \in \mathcal{R}' \end{aligned}$$

The most common construction of a valid p-value is given by the following theorem.

Theorem Suppose that $T(\underset{\sim}{X})$ is a test statistic constructed so that a large value of $T(\underset{\sim}{X})$ supports H_1. Then the statistic $p(\underset{\sim}{x})$ given for each $\underset{\sim}{x} \in \mathcal{X}$ by

$$p(\underset{\sim}{x}) = \sup_{\theta \in \Theta_0} \Pr[T(\underset{\sim}{X}) \geq T(\underset{\sim}{x}) \mid \theta] = \sup_{\theta \in \Theta_0} p_\theta(\underset{\sim}{X}) \tag{6}$$

say, is a valid p-value.

Proof For $\theta \in \Theta_0$, we have

$$p_\theta(\underset{\sim}{x}) = \Pr[T(\underset{\sim}{X}) \geq T(\underset{\sim}{x}) \mid \theta] = \Pr[-T(\underset{\sim}{X}) \leq -T(\underset{\sim}{x}) \mid \theta] = F_\theta(-T(\underset{\sim}{x})) \equiv F_S(s)$$

say, defining $F_S \equiv F_\theta$ as the cdf of $S = -T(\underset{\sim}{X})$; clearly $0 \leq p(\underset{\sim}{x}) \leq 1$.

This recalls a result from distribution theory; if $X \sim F_X$, the $U = F_X(X) \sim Uniform(0,1)$. Suppressing the dependence on θ for convenience, define random variable Y by

$$Y = F_\theta(-T(\underset{\sim}{X})) \equiv F_S(S) \qquad (= p_\theta(\underset{\sim}{X}))$$

and let $A_y \equiv \{s : F_S(s) \leq y\}$. If A_y is a half-closed interval $(-\infty, s_y]$, then

$$F_Y(y) = \Pr[Y \leq y] = \Pr[F_S(S) \leq y] = \Pr[S \in A_y] = F_S(s_y) \leq y$$

by definition of A_y, as $s_y \in A_y$. If A_y is a half-open interval $(-\infty, s_y)$

$$F_Y(y) = \Pr[Y \leq y] = \Pr[F_S(S) \leq y] = \Pr[S \in A_y] = \lim_{s \longrightarrow s_y} F_S(s) \leq y$$

by continuity of probability. Putting the components together, for $0 \leq \alpha \leq 1$,

$$\Pr[p_\theta(\underset{\sim}{X}) \leq \alpha \mid \theta] \equiv \Pr[Y \leq \alpha] \leq \alpha$$

But by the definition in equation (6), $p(\underset{\sim}{x}) \geq p_\theta(\underset{\sim}{x})$, so

$$\Pr[p(\underset{\sim}{X}) \leq \alpha \mid \theta] \leq \Pr[p_\theta(\underset{\sim}{X}) \leq \alpha \mid \theta] \leq \alpha$$

and the result follows. ■