Generalization of the Likelihood Ratio Test

The Neyman-Pearson Lemma says that the likelihood ratio test is optimal for simple hypotheses.

Goal: to develop a generalization of this test for use in situations in which the hypotheses are not simple.

- Generalized likelihood ratio tests are not generally optimal, but they perform reasonably well.
  - Often there are no optimal tests at all.
- Generalized likelihood ratio tests have wide utility.
  - They play the same role in testing as MLEs do in estimation.
Generalized Likelihood Ratio Test

Let $X = (X_1, \ldots, X_n)$ be data and let $\pi(x|\theta)$ be the joint density of the data. The likelihood function is then

$$L(\theta) = \pi(X|\theta)$$

Suppose we wish to test

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1$$

where $\Theta_0$ and $\Theta_1$ are two disjoint sets of the parameter space $\Theta$, $\Theta = \Theta_0 \sqcup \Theta_1$.

- Based on the data, a measure of relative plausibility of the hypotheses is the ratio of their likelihoods.
- If the hypotheses are composite, each likelihood is evaluated at that value of $\theta$ that maximizes it.

This yields the **generalized likelihood ratio**:

$$\Lambda^* = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta_1} L(\theta)}$$

Small values of $\Lambda^*$ tend to discredit $H_0$. 
Generalized Likelihood Ratio Test

For technical reasons, it is preferable to use the following statistic instead of $\Lambda^*$:

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)}$$

- $\Lambda$ is called the likelihood ratio statistic.
- Note that

$$\Lambda = \min\{\Lambda^*, 1\}$$

Thus, small values of $\Lambda^*$ correspond to small values of $\Lambda$.

The rejection region $R$ for a generalized likelihood test has the following form:

$$\text{reject } H_0 \iff X \in R = \{X : \Lambda(X) < \lambda\}$$

The threshold $\lambda$ is chosen so that

$$\mathbb{P}(\Lambda(X) < \lambda | H_0) = \alpha,$$

where $\alpha$ is the desired significance level of the test.
Example

Let $X_1, \ldots, X_n$ be i.i.d. from $\mathcal{N}(\mu, \sigma^2)$, where variance $\sigma^2$ is known. Consider testing the following hypothesis:

$$H_0 : \mu = \mu_0 \quad \text{and} \quad H_1 : \mu \neq \mu_0$$

Construct the generalized likelihood test with significance level $\alpha$.

Answer:

$$\text{Reject } H_0 \iff \frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\sigma} > z_{\frac{\alpha}{2}}$$
Distribution of $\Lambda(X)$

In order for the generalized likelihood ratio test to have the significance level $\alpha$, the threshold $\lambda$ must be chosen so that

$$\mathbb{P}(\Lambda(X) < \lambda | H_0) = \alpha$$

If the distribution of $\Lambda(X)$ under $H_0$ is known, then we can determine $\lambda$.

- In the Example, $-2 \log \Lambda(X) \sim \chi^2_1$.

Generally, the distribution of $\Lambda$ is not of a simple form, but in many situations the following theorem provides the basis for an approximation of the distribution.

**Theorem**

*Under smoothness conditions on $\pi(x|\theta)$, the null distribution of $-2 \log \Lambda(X)$ (i.e. distribution under $H_0$) tends to a $\chi^2_d$ as the sample size $n \to \infty$, where*

$$d = \dim \Theta - \dim \Theta_0,$$

*where $\dim \Theta$ and $\dim \Theta_0$ are the numbers of free parameters in $\Theta$ and $\Theta_0$.*

- In the Example, $\dim \Theta = 1$ and $\dim \Theta_0 = 0$. 
Summary

- **Generalized likelihood ratio tests** are used when the hypothesis are composite
  - They are not generally **optimal**, but perform reasonably well.
  - They play the same role in testing as MLEs do in estimation.

- The **rejection region** $\mathcal{R}$ for a generalized likelihood test has the following form:

  \[
  \text{reject } H_0 \iff X \in \mathcal{R} = \{X : \Lambda(X) < \lambda\}
  \]

  - $\Lambda$ is the likelihood ratio statistic,
    \[
    \Lambda = \frac{\max_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\max_{\theta \in \Theta} \mathcal{L}(\theta)}
    \]

  - The **threshold** $\lambda$ is chosen so that
    \[
    \mathbb{P}(\Lambda(X) < \lambda | H_0) = \alpha,
    \]
    where $\alpha$ is the desired significance level of the test.

- As sample size $n \to \infty$, the null distribution of $-2 \log \Lambda(X)$ tends to a $\chi^2_d$, where
  \[
  d = \text{dim } \Theta - \text{dim } \Theta_0
  \]