Lecture 10 Likelihood Ratio Tests

In this lecture, a basic framework of hypothesis testing will be presented, followed by an introduction to likelihood ratio tests.

10.1 Basic elements in hypothesis testing

- Data model: $X \sim f(x|\theta)$, $x \in \mathcal{X}$, $\theta \in \Theta$.
- Parameter space decomposition: $\Theta = \Theta_0 \cup \Theta_1$ with $\Theta_0 \cap \Theta_1 = \emptyset$.
- A two-decision problem: $\mathcal{A} = \{a_0, a_1\}$, where the action a_i claims $H_i : \theta \in \Theta_i$, $i = 0, 1$, with a null hypothesis H_0 and an alternative hypothesis H_1.
- A *test function* $d : \mathcal{X} \to \mathcal{A}$, or simply $d : \mathcal{X} \to \{0, 1\}$.
- The 0-1 loss $L(\theta, a_i) = I_{\{\theta \notin \Theta_i\}}$ corresponds to the type I error $L(\theta, a_1)$ and type II error $L(\theta, a_0)$ respectively.
- "risk function = error probability": $R(\theta, d) = P_\theta(d(X) \neq i)$, $\theta \in \Theta_i,\ i = 0, 1$.
- *Power function* of test d: $\beta(\theta) = \beta(\theta; d) = P_\theta(d(X) \neq 0) = P_\theta(\text{reject } H_0)$; in particular,

$$\beta(\theta) = \begin{cases} \text{type I error probability}, & \text{for } \theta \in \Theta_0, \\ 1 - \text{type II error probability}, & \text{for } \theta \in \Theta_1,\ (\text{power of the test } d). \end{cases}$$

- $\alpha \in (0, 1)$ is called the *significance level* of test d if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$; α is called the *size* of test d if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$. Apparently, a size-$\alpha$ test is also a level-α test, but not vice versa.
- A hypothesis H (null or alternative) is called a *simple* (resp. *composite*) hypothesis if and only if H is a singleton, i.e. H contains only a single parameter value.
- *Critical region* (*rejection region*): $R = \{x \in \mathcal{X} :\ d(x) \neq 0\}$, and *acceptance region*: $A = \{x \in \mathcal{X} :\ d(x) = 0\}$; having observed $x \in R$, a size-α test rejects H_0 with type I error probability $P_\theta(R) \leq \alpha\ \forall\ \theta \in \Theta_0$.
- *Randomized test* δ: given observation x, a probability distribution is defined over $\mathcal{A}$ with $\delta(x, a_0) + \delta(x, a_1) = 1$; i.e. Having observed x, we will reject H_0 with probability $\delta(x, a_1)$ and accept H_0 with probability $\delta(x, a_0)$. (Think about how to realize that.) The power function for a randomized test δ is $\beta(\theta; \delta) = E_\theta\ \delta(X, a_1)$. Randomized decision rules can be useful in certain applications, e.g. when we study Neyman-Pearson tests in Lecture 11.

- p-value (observed size) for test d: a test statistic $p(X) \in [0,1]$, the smallest significance level α such that d rejects H_0.

10.2 Likelihood ratio tests (LR tests)

Just like MLEs in point estimation, LR tests enjoy great popularity in hypothesis testing due to their generality. The idea applies to almost all parametric models, and some nonparametric extensions of LR tests have also been developed.

First, for given data x, the ratio $R(x) \triangleq \frac{\sup_{\theta\in\Theta_1} L(\theta|x)}{\sup_{\theta\in\Theta_0} L(\theta|x)}$ tends to get large if H_0 is false. Second, it is more convenient to work with the ratio $\Lambda(x) \triangleq \frac{\sup_{\theta\in\Theta} L(\theta|x)}{\sup_{\theta\in\Theta_0} L(\theta|x)}$, called a *likelihood ratio* statistic.

Note: $\Lambda(x) = \max\{R(x),\ 1\}$ hence the two ratios are equivalent. $\lambda(x) = 1/[\Lambda(x)]$ is used as a LR statistic in Casella and Berger.

A recipe for LR tests

Step 1: Find MLE $\hat{\theta}$.

Step 2: Find the restricted MLE $\hat{\theta}_0$ on Θ_0.

Step 3: Convert the LR statistic $\Lambda(x) = \frac{L(\hat{\theta}|x)}{L(\hat{\theta}_0|x)}$ to an operating characteristic whose distribution can be found or approximated.

Example 10.1 (a two-sample t test) Let $X_1, ..., X_m$ be iid samples from $N(\mu_x, \sigma^2)$, and $Y_1, ..., Y_n$ be iid samples from $N(\mu_y, \sigma^2)$. Assume X^m and Y^n are independent, $\theta = (\mu_x, \mu_y, \sigma^2)$. Test $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$.

Step 1: The MLEs are $\hat{\mu}_x = \bar{X}$, $\hat{\mu}_y = \bar{Y}$, and (a pooled estimate) $\widehat{\sigma^2} = \frac{1}{m+n}[\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2]$.

Step 2: Under H_0, $\mu_x = \mu_y \triangleq \mu$. The restricted MLEs on Θ_0 are $\hat{\mu} = \frac{1}{m+n}(\sum_{i=1}^m X_i + \sum_{j=1}^n Y_j)$ and $\widehat{\sigma_0^2} = \frac{1}{m+n}[\sum_{i=1}^m (X_i - \hat{\mu})^2 + \sum_{j=1}^m (Y_j - \hat{\mu})^2]$.

Step 3: $\Lambda_{m,n} \triangleq \Lambda(X^m, Y^n) = \left(\frac{\widehat{\sigma_0^2}}{\widehat{\sigma^2}}\right)^{\frac{m+n}{2}}$. To simplify $\Lambda_{m,n}$, note that $\frac{\widehat{\sigma_0^2}}{\widehat{\sigma^2}} = 1 + (m+n-2)^{-1}T^2$ where under H_0, $T = \frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\frac{m+n}{m+n-2}\,\widehat{\sigma^2}}}$ follows a central t distribution with degree of freedom $m+n-2$. Since $\Lambda_{m,n}$ is increasing in $|T|$, we reject H_0 if $|T| > c$, where c is determined by the size α.

The calculation of the power function involves a non-central t distribution with non-centrality parameter $\eta = \frac{\mu_x - \mu_y}{\sqrt{(\frac{1}{m}+\frac{1}{n})\sigma^2}}$ for $\theta \in \Theta_1$.

Note: More examples of finite-sample LR tests will be dealt with in homework problems. Later, we will study large-sample LR tests.