Detection and Estimation
Chapter 1. Hypothesis Testing

Husheng Li

Min Kao Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville

Spring, 2015
Syllabus

- **Homework**: 4 problems each week; hand in your homework after one week. 20%

- **Midterm and Final Exams**: 50%

- **Project**: 30%, choose one topic such as large scale hypothesis testing, quickest detection, robust detection, distributed detection, distributed estimation, large covariance matrix estimation, et al; the topic should be beyond your own research; send me your topic right after the midterm; read the most important books and papers in that topic; apply it in an application you choose by your own; provide a presentation of the survey and your own results before the final.
The book by H. V. Poor will be the textbook. Other books are for references.
The First Paper on Hypothesis Testing

In 1900, K. Pearson published the first paper on hypothesis testing with the long title “On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling”.

In the paper, he discussed the criterion to judge whether the deviations of a set of samples can be explained by the randomness of sampling.
The First Paper on Hypothesis Testing: Continued

- Pearson made a mistake in this paper, which was pointed out by an agriculture statistician R. Fisher (Pearson was 65 and Fisher was 32).
- People in early 1900s considered Fisher as an infant claiming to be taller than his father.
Bayesian Hypothesis Testing

- We assume that two hypotheses $H_0$ and $H_1$ corresponds to two distributions. The set of observations is denoted by $\Gamma$. The goal is to determine whether a set of observations are generated by $H_0$ or $H_1$.

- A decision rule $\delta$ is to divide $\Gamma$ into $\Gamma_0$ and $\Gamma_1$ such that

$$
\delta(y) = \begin{cases} 
1, & \text{if } y \in \Gamma_1 \\
0, & \text{if } y \in \Gamma_1^c
\end{cases}
$$

- $C_{ij}$ means the cost incurred by choosing $H_i$ while $H_j$ is true. The conditional risk is defined as

$$
R_j(\delta) = C_{1j}P_j(\Gamma_1) + C_{0j}P_j(\Gamma_0), \quad j = 0, 1.
$$
Bayesian Risk and Decision Rule

- We assume that $H_0$ and $H_1$ has prior probabilities $\pi_0$ and $\pi_1$. Then, the Bayes risk is defined as

$$r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta).$$

Thus the optimal decision rule is to minimize the Bayes risk.

- Theorem: The decision rule minimizing the Bayes risk is to calculate the likelihood ratio $L(y) = \frac{p_1(y)}{p_0(y)}$ and make the decision:

$$\delta_B(y) = \begin{cases} 
1, & \text{if } L(y) \geq r \\
0, & \text{if } L(y) < r 
\end{cases}$$

where $r$ is a threshold.
The difference between the Bayesian and Frequentist schools is whether the unknown is fixed or random.

In early 20th century, Bayesian statistics were disliked by famous statisticians Neyman and Fisher. But it flourished in late 20th century.

The main challenge of the Bayesian school is how to determine the prior distribution.

Empirical Bayesian method is kind of a bridge between the two schools.
The Bayesian risk based testing requires the prior distributions of the hypotheses. What if we do not know them?

If the prior distributions are not known, then we should minimize the worst case, i.e., the maximum risk:

$$\delta^* = \arg \min_{\delta} \max_{\pi_0} \pi_0 R_0(\delta) + (1 - \pi_0) R_1(\delta).$$

Again, we can prove that the likelihood ratio based testing is optimal for the MiniMax hypotheses testing.
Illustration of the MiniMax Rule
Randomized Decision Rule

When if $V$ is not differentiable at the least favorable prior probability $\pi_L$. 
Neyman-Pearson Hypothesis Testing

- Sometimes, the costs are defined directly on the error probabilities.
- Type I error (false alarm): $H_0$ is falsely rejected, probability $P_F$.
- Type II error (missed detection): $H_1$ is falsely rejected, probability $P_M$.
- Detection probability $P_D = 1 - P_M$.
- Neyman-Pearson design criterion:

$$\max_{\delta} P_D(\delta), \quad \text{s.t.} \quad P_F(\delta) \leq \alpha.$$
Neyman-Pearson Lemma

Theorem: Consider the hypothesis with densities $p_1$ and $p_0$. Assume $\alpha > 0$. Then, the following statements are true:

- Let $\tilde{\delta}$ be any decision rule satisfying $P_F(\tilde{\delta}) \leq \alpha$ and $\tilde{\delta}'$ be any decision rule of the form

$$
\tilde{\delta}'(y) = \begin{cases} 
1, & p_1(y) > \eta p_0(y) \\
\gamma(y), & p_1(y) = \eta p_0(y) \\
0, & p_1(y) < \eta p_0(y)
\end{cases}
$$

where $\eta$ and $\gamma(y)$ are such that $P_F(\tilde{\delta}') = \alpha$. Then, $P_D(\tilde{\delta}') \geq P_D(\tilde{\delta})$.

- For every $\alpha \in (0, 1)$, there exists a decision rule $\tilde{\delta}_{NP}$ satisfying (1) with $\gamma(y) = \gamma_0$ and $P_F(\tilde{\delta}_{NP}) = \alpha$.

- Suppose that $\tilde{\delta}''$ is any $\alpha$-level Neyman-Pearson decision rule. Then, it must be of the form (1) except possibly on a subset of $\Gamma$ of zero probability measure.
Neyman (in Poland) and Pearson (in UK) collaborated between 1926 and 1936. Their paper (98 pages) on the likelihood ratio test was published in 1928.

But their collaboration did not survive...

Some Gossips

From 1926-36 we were working together in excited co-operation. My clumsily defined ideas, sharpened by your mathematical formulation, and we went on and on together, until by about 1936-37 we had solved between us what seemed the basic problems, and so found a statistical philosophy. But the time came when to find new mountains to scale, you were forced to tackle more and more mathematically complex problems – tests of “Type” $A_1$ or $B_2$, etc., etc., and I began to lose interest because I was always aiming at attacking types of problems with probability tools which seemed to get fairly simply into gear with the way which the human reason worked. And “Types” $A_y$, $B_y$, etc., etc., seemed to me stepping out of this field. Then you rightly went to the U.S. and war came. And I suppose we both turned our statistically trained minds to different kinds of jobs: bombs, A.A. shells and what not.
Composite Hypothesis Testing

\[ \delta_B(y) = \begin{cases} 
1 & \text{if } E\{C[1, \Theta]|Y = y\} < E\{C[0, \Theta]|Y = y\} \\
0 \text{ or } 1 & \text{if } E\{C[1, \Theta]|Y = y\} = E\{C[0, \Theta]|Y = y\} \\
0 & \text{if } E\{C[1, \Theta]|Y = y\} > E\{C[0, \Theta]|Y = y\}. 
\end{cases} \]

- In simple hypothesis testing, each hypothesis corresponds to only one distribution.
- In composite hypothesis testing, each hypothesis may correspond to multiple distributions. Usually we consider a family of distributions indexed by a parameter \( \theta \).
- In the Bayesian setup, we assume \( \theta \) is random and the corresponding prior probability is known.
Composite Hypothesis Testing: Decomposition

In many situations, the parameter space can be decomposed to two disjoint sets $\Lambda_0$ and $\Lambda_1$, representing $H_0$ and $H_1$, respectively.

The decision rule is still based on likelihood ratio test, where the likelihood is calculated by

$$\delta_B(y) = \begin{cases} 
1 & \text{if } L(y) > \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \\
0 \text{ or } 1 & < \\
0 & < 
\end{cases}$$

$$p(y|\Theta \in \Lambda_j) = \int_{\Lambda_j} p_\theta(y)w(\theta)d\theta.$$
Example II.E.1: Testing on the Radius of a Point in the Plane

Suppose that $\Gamma = \mathbb{R}^2$ [i.e., $Y = (Y_1, Y_2)^T$] and our hypotheses are as follows:

\[
H_0 : Y_1 = \epsilon_1, \quad Y_2 = \epsilon_2
\]

versus

\[
H_1 : Y_1 = A \cos \Psi + \epsilon_1, \quad Y_2 = A \sin \Psi + \epsilon_2,
\]  

(II.E.13)
UMP Test

- For composite hypothesis testing problems in which we do not have a prior distribution for the parameter, the development of hypothesis testing that satisfy precise analytical definitions of optimality is very often an illusive task. We can generalize the Neyman-Pearson test by defining the false alarm and detection probabilities to

\[ P_F(\tilde{\delta}, \theta) = E_{\theta} \left[ \tilde{\delta}(Y) \right], \theta \in \Lambda_0 \]

\[ P_F(\tilde{\delta}, \theta) = E_{\theta} \left[ \tilde{\delta}(Y) \right], \theta \in \Lambda_1. \]

- The ideal case is that the test maximizes \( P_D \) for every \( \theta \) while satisfying \( P_F \leq \alpha \) for every \( \theta \). This is called the uniformly most powerful (UMP) test.
Example II.E.2: UMP Testing of Location

Consider the parametric family of distributions \( \{P_{\theta}; \theta \in \Lambda\} \), where \( P_{\theta} \) is the \( \mathcal{N}(\theta, \sigma^2) \) distribution and \( \Lambda \) is a subset of \( \mathbb{R} \). Suppose that we have the hypothesis pair

\[
H_0 : \theta = \mu_0
\]

versus

\[
H_1 : \theta > \mu_0
\]  

(II.E.21)
What If UMP Test Does Not Exist?

In many situations, UMP test does not exist. This can be overcome by applying other constraints to eliminate unreasonable tests from consideration.

One such condition is unbiasedness.

We can try to optimize the derivative of $P_D(\theta_0)$ for the locally optimal test. Or we can try generalized likelihood ratio test:

$$\frac{\max_{\theta \in \Lambda_1} p_\theta(y)}{\max_{\theta \in \Lambda_0} p_\theta(y)}.$$