Session 18

18.1

The Generalized Likelihood Ratio Test

- Sometimes we can't find a UMP test
  - Doesn't exist
  - Difficult to construct
- We then need an alternative:
  - Locally Most Powerful (LMP) tests
  - Generalized Likelihood ratio test (GLRT)

The Generalized Likelihood Ratio Test

The *generalized likelihood ratio test* (GLRT) gets around the composite hypothesis testing problem by effectively turning it into a test between two simple hypotheses.

These simple hypotheses are selected to be the most likely value of

$$\theta_0 \in \Theta_0 \text{ under } H_0$$

and

$$\theta_1 \in \Theta_1 \text{ under } H_1$$

given the observed data.

The Generalized Likelihood Ratio Test (Cont.)

- Using these two simple hypotheses, a likelihood ratio test is implemented to test between them.
- The composite hypothesis corresponding to the simple hypothesis declared by the simple likelihood test is the composite hypothesis declared by the GLRT.
- While the GLRT is not optimal in any particular sense, it seems like a reasonable approach to dealing with the composite hypothesis testing problem.
- In many cases where a UMP test does exist, the GLRT exhibits nearly optimal behavior.

Generalized Likelihood Ratio Test (GLRT)

Consider two composite hypotheses $H_0 : \underline{\theta} \in \Theta_0$ and $H_1 : \underline{\theta} \in \Theta_1$. The *Generalized Likelihood Ratio Test* (GLRT) consists of the following procedure:

1. Assume H_0 is true and estimate the value of θ from the observed data using a *maximum likelihood estimate* (MLE):
$$\hat{\underline{\theta}}_0 = \arg\max_{\underline{\theta} \in \Theta_0} f_{\underline{\theta}}(\underline{X}).$$
2. Assume H_1 is true and estimate the value of θ from the observed data using a (MLE):
$$\hat{\underline{\theta}}_1 = \arg\max_{\underline{\theta} \in \Theta_1} f_{\underline{\theta}}(\underline{X}).$$
3. Replace the original problem of testing the composite hypotheses H_0 versus H_1 with the problem of testing the simple hypotheses $\hat{H}_0 : \hat{\Theta}_0 = \{\hat{\underline{\theta}}_0\}$ versus $\hat{H}_1 : \hat{\Theta}_1 = \{\hat{\underline{\theta}}_1\}$. If $\hat{H}_0$ is decided in the simple hypothesis problem, then H_0 is decided as the composite hypothesis. If $\hat{H}_1$ is decided in the simple hypothesis problem, then H_1 is decided as the composite hypothesis.

GLRT-continued

When we carry out this procedure, we get a GLRT of the form

$$L_g(\underline{X}) = \frac{\max_{\underline{\theta} \in \Theta_1} f_{\underline{\theta}}(\underline{X})}{\max_{\underline{\theta} \in \Theta_0} f_{\underline{\theta}}(\underline{X})} \underset{H_0}{\overset{H_1}{\gtrless}} L_0,$$

which yields a statistical test of the form

$$\phi(\underline{X}) = \begin{cases} 1, & \text{for } L_g(\underline{X}) > L_0, \\ \gamma, & \text{for } L_g(\underline{X}) = L_0, \\ 0, & \text{for } L_g(\underline{X}) < L_0, \end{cases}$$

where, in principle, L_0 and γ are selected to yield a size α test.

18.6

L_0 and γ

$$\phi(\underline{X}) = \begin{cases} 1, & \text{for } L_g(\underline{X}) > L_0, \\ \gamma, & \text{for } L_g(\underline{X}) = L_0, \\ 0, & \text{for } L_g(\underline{X}) < L_0. \end{cases}$$

Selecting L_0 and γ to yield a size α test is difficult.

The reason is that the size of the test is still defined as

$$\alpha = \sup_{\underline{\theta} \in \Theta_0} \mathrm{E}_{\underline{\theta}} \left[\phi(\underline{X})\right].$$

We do not use $\mathrm{E}_{\hat{\underline{\theta}}_0} \left[\phi(\underline{X})\right]$ as the size of the test.

18.7

Example: *Suppose we wish to test the hypotheses H_0 versus H_1 that the random sample $\underline{X} = (X_1, \ldots, X_N)$ comes from a density*

$$f_\theta(x_n) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-(x_n - \theta)^2}{2}\right\},$$

where under $H_0 : \Theta_0 = [-1, 1]$, and under $H_1 : \Theta_0 = \{\theta \in \mathbf{R} : |\theta| > 1\}$.

We note that in general,

$$f_\theta(\underline{X}) = \frac{1}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(x_n - \theta)^2\right\}, \tag{1}$$

from which it follows that the (unconstrained) maximum likelihood estimate $\hat{\theta}_{ML}$ of θ can be found by solving

$$\frac{\partial}{\partial\theta} f_\theta(\underline{X}) = 0$$

for θ, yielding

$$\hat{\theta}_{ML} = \frac{1}{N}\sum_{n=1}^{N} X_n.$$

For this case, we can easily see that

$$\hat{\theta}_0 = \arg\max_{\underline{\theta}\in\Theta_0} f_{\underline{\theta}}(\underline{X}) = \begin{cases} \hat{\theta}_{ML}, & \text{for } \hat{\theta}_{ML} \in [-1,1], \\ -1, & \text{for } \hat{\theta}_{ML} < -1, \\ 1, & \text{for } \hat{\theta}_{ML} > 1, \end{cases}$$

and

$$\hat{\theta}_1 = \arg\max_{\underline{\theta}\in\Theta_1} f_{\underline{\theta}}(\underline{X}) = \begin{cases} \hat{\theta}_{ML}, & \text{for } \hat{\theta}_{ML} \notin [-1,1], \\ -1, & \text{for } \hat{\theta}_{ML} \in [-1,0], \\ 1, & \text{for } \hat{\theta}_{ML} \in (0,1]. \end{cases}$$

Using this $\hat{\theta}_0$ and $\hat{\theta}_1$ we can now construct the GLRT

$$L_g(\underline{X}) = \frac{f_{\hat{\theta}_1}(\underline{X})}{f_{\hat{\theta}_0}(\underline{X})} \underset{H_0}{\overset{H_1}{\gtrless}} L_0$$

with an appropriately chosen threshold L_0.

Recall: $\alpha = \sup_{\underline{\theta}\in\Theta_0} \mathrm{E}_{\underline{\theta}}[\phi(\underline{X})]; \quad \alpha \neq \mathrm{E}_{\underline{\hat{\theta}}_0}[\phi(\underline{X})].$

Bayesian Detection Theory

In *classical* detection, we decide between $H_0 : \underline{\theta} \in \Theta_0$ versus $H_1 : \underline{\theta} \in \Theta_1$ based on an observation $\underline{X}$ governed by the parameterized cdf $F_{\underline{\theta}}(\underline{x})$

Here $\underline{\theta}$ was assumed to be an *unknown* but fixed parameter. The parameter $\underline{\theta}$ was *not* assumed to be random, just unknown.

In the Bayesian detection framework, we again assume that our observation $\underline{X}$ is governed by a distribution $F_{\underline{\theta}}(\underline{X})$,

but we now assume that $\underline{\theta}$ is a random vector taking on one of two possible values: $\underline{\theta}_0$ or $\underline{\theta}_1$. These values are taken on with probabilities

$$\begin{aligned} p_0 &= P(\{\underline{\theta} = \underline{\theta}_0\}), \\ p_1 &= P(\{\underline{\theta} = \underline{\theta}_1\}) = 1 - p_0, \end{aligned}$$

where $p_0 \in [0,1]$.

18.10

We have a random experiment with probability space $(\mathcal{S}, \mathcal{F}, P)$ having two random variables $\underline{\theta}$ and $\underline{X}$ defined on it.

When the random experiment is performed, $\underline{\theta}$ takes on a value from the set $\{\underline{\theta}_0, \underline{\theta}_1\}$ with probabilities p_0 and p_1, respectively.

The observed value of $\underline{X}$ takes on a value consistent with the distribution $F_{\underline{\theta}}(\underline{x})$ for the value $\underline{\theta}$ takes on.

Thus we have a conditional distribution function for $\underline{X}$:

$$F(\underline{x}|\underline{\theta}_j) = P(\{\underline{X} \leq \underline{x}\}|\{\underline{\theta} = \underline{\theta}_j\}) = F_{\underline{\theta}_j}(\underline{x}), \quad j = 0, 1.$$

The joint distribution of $\underline{\theta}$ and $\underline{X}$ is given by

$$F(\underline{\theta}, \underline{x}) = F(\underline{x}|\underline{\theta})P(\underline{\theta}) = \begin{cases} p_0 F(\underline{x}|\underline{\theta}_0), & \text{for } \underline{\theta} = \underline{\theta}_0, \\ p_1 F(\underline{x}|\underline{\theta}_1), & \text{for } \underline{\theta} = \underline{\theta}_1. \end{cases}$$

18.11

It follows that

$$P(\underline{\theta}) = \int_{\mathbf{R}^n} dF(\underline{\theta}, \underline{x}) = \int_{\mathbf{R}^n} f(\underline{\theta}, \underline{x})\, d\underline{x} = \begin{cases} p_0, & \text{for } \underline{\theta} = \underline{\theta}_0, \\ p_1, & \text{for } \underline{\theta} = \underline{\theta}_1, \end{cases}$$

and

$$\begin{aligned} F(\underline{x}) &= F(\underline{x}|\underline{\theta}_0)P(\underline{\theta}_0) + F(\underline{x}|\underline{\theta}_1)P(\underline{\theta}_1) \\ &= p_0 F(\underline{x}|\underline{\theta}_0) + p_1 F(\underline{x}|\underline{\theta}_1), \end{aligned}$$

where $\mathbf{R}^n$ is the observation space of the n-dimensional observation vector.