11. The Generalized Likelihood Ratio
ECE 830, Spring 2014
The **generalized likelihood ratio test (GLRT)** is a general procedure for composite testing problems. The basic idea is to compare the best model in class \( H_1 \) to the best in \( H_0 \), which is formalized as follows. We have two composite hypotheses of the form:

\[
H_i : X \sim p_i(x|\theta_i), \theta_i \in \Theta_i, i = 0, 1.
\]

The parametric densities \( p_0 \) and \( p_1 \) need not have the same form. The GLRT based on an observation \( x \) of \( X \) is

\[
\hat{\Lambda}(x) = \frac{\max_{\theta_1 \in \Theta_1} p_1(x|\theta_1)}{\max_{\theta_0 \in \Theta_0} p_0(x|\theta_0)} \begin{array}{c} \sim \ H_1 \\ \geq H_0 \end{array} \gamma,
\]

or equivalently

\[
\log \hat{\Lambda}(x) \begin{array}{c} \sim \ H_1 \\ \geq H_0 \end{array} \gamma.
\]
Example: Signal Detection

Consider two hypotheses

\[ H_0 : X \sim \mathcal{N}(0, \sigma^2 I_n) \]
\[ H_1 : X \sim \mathcal{N}(H\theta, \sigma^2 I_n) \]

where \( \sigma^2 > 0 \) is known, \( H \) is a known \( n \times k \) matrix, and \( \theta \in \mathbb{R}^k \) is unknown. The mean vector \( H\theta \) is a model for a signal that lies in the \( k \)-dimensional subspace spanned by the columns of \( H \) (e.g., a narrowband subspace, polynomial subspace, etc.). In other words, the signal has the representation

\[ s = \sum_{i=1}^{k} \theta_i h_i, \quad H = [h_1, \ldots, h_k]. \]

The null hypothesis is that no signal is present (noise only).
Log LR

\[
\log \Lambda(x) = -\frac{1}{2\sigma^2} (x - H\theta)^\top (x - H\theta) + \frac{1}{2\sigma^2} x^\top x
\]

\[
= \frac{1}{\sigma^2} (\theta^\top H^\top x - \frac{1}{2} \theta^\top H^\top H\theta).
\]

Since \( \theta \) is unknown we can’t go further, instead we find \( \theta \) that makes \( x \) most likely:

\[
\hat{\theta} = \arg \max_{\theta} p(x|H_1, \theta)
\]

\[
= \arg \max_{\theta} \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-\frac{1}{2\sigma^2} (x-H\theta)^\top (x-H\theta)}
\]

\[
= \arg \max_{\theta} -\frac{1}{2\sigma^2} (x - H\theta)^\top (x - H\theta)
\]

\[
= \arg \min_{\theta} (x - H\theta)^\top (x - H\theta)
\]

\[
= \arg \min_{\theta} (x^\top x - 2\theta^\top H^\top x + \theta^\top H^\top H\theta)
\]
Example: (cont.)

Taking the derivative with respect to $\theta$

\[
\frac{\partial}{\partial \theta} (x^\top x - 2\theta^\top H^\top x + \theta^\top H^\top H\theta) = 0
\]

\[
\Rightarrow 0 - 2H^\top x + 2H^\top H\theta = 0
\]

\[
\Rightarrow \hat{\theta} = (H^\top H)^{-1} H^\top x
\]

Now we plug $\hat{\theta}$ into the GLRT: $\theta \leftarrow \hat{\theta}$

\[
\log \hat{\Lambda}(x) := \frac{1}{\sigma^2} \left[ x^\top H(H^\top H)^{-1} H^\top x \\
- \frac{1}{2} x^\top H(H^\top H)^{-1} H^\top H(H^\top H)^{-1} H^\top x \right]
\]

\[
= \frac{1}{2\sigma^2} x^\top H(H^\top H)^{-1} H^\top x
\]
Example: (cont.)

Recall that the projection matrix onto the subspace is defined as
\[ P_H := H (H^\top H)^{-1} H^\top \]

\[ \log \hat{\Lambda}(x) = \frac{1}{2\sigma^2} x^\top P_H x = \frac{1}{2\sigma^2} \|P_H x\|_2^2. \]

Observe that this is simply an energy detector in \( H \): we are taking the projection of \( x \) onto \( H \) and measuring the energy. The expected value of this energy under \( H_0 \) (noise only) is

\[ \mathbb{E}_{H_0} \left[ \|P_H X\|_2^2 \right] = k\sigma^2, \]

since a fraction \( k/n \) of the total noise energy \( n\sigma^2 \) falls into this subspace.
The performance of the subspace energy detector can be quantified as follows. We choose a $\gamma$ for the desired $P_{FA}$:

$$\frac{1}{\sigma^2} x^\top P_H x \overset{H_1}{\underset{H_0}{\geq}} \gamma$$

**What is the distribution of $x^\top P_H x$ under $H_0$?** First use the decomposition

$$P_H = UU^\top$$

where $U \in \mathbb{R}^{n \times k}$ with orthonormal columns spanning columns of $H$, and let $y := U^\top x$. Then

$$\frac{1}{\sigma^2} x^\top P_H x =$$

$$y \sim$$

$$y_i/\sigma \overset{iid}{\sim}$$

$$\Rightarrow \frac{y^\top y}{\sigma^2} \sim$$
GLRT and $P_{FA}$

Example: (cont.)

Under $H_0$,

$$\frac{1}{\sigma^2} x^\top P_H x \sim \chi_k^2 \quad \Rightarrow \quad P_{FA} = \mathbb{P}(\chi_k^2 > \gamma)$$

The $P_{FA}$ of a $\chi_k^2$ distribution.
$\chi^2_k$ Distributions

To calculate the tails on $\chi^2_k$ distributions you can look it up in the back of a good book or use Matlab ($\text{chi2cdf}(x,k)$, $\text{chi2inv}(\gamma,k)$, $\text{chi2cdf}(x,k)$). Remember the mean of a $\chi^2_k$ distribution is $k$, so we want to choose a $\gamma$ bigger than $k$ to produce a small $P_{FA}$.
Wilks’ Theorem

Wilk’s Theorem (1938)

Consider a composite hypothesis testing problem

\[ H_0 : X_1, X_2, \ldots, X_n \sim \text{iid } p(x|\theta_0), \]

where \( \theta_{0,1}, \ldots, \theta_{0,\ell} \in \mathbb{R} \) are free parameters and \( \theta_{0,\ell+1} = a_{\ell+1}, \ldots, \theta_k = a_k \) are fixed at the values \( a_{\ell+1}, \ldots, a_k \)

\[ H_1 : X_1, X_2, \ldots, X_n \sim \text{iid } p(x|\theta_1), \theta_1 \in \mathbb{R}^k \text{ are all free parameters} \]

and the parametric density has the same form in each hypothesis.

In this case family of models in \( H_0 \) is a subset of those in \( H_1 \), and we say that the hypotheses are nested. (This is a key condition that must hold for this theorem.)
Wilk’s Thm (cont.)

If the 1\textsuperscript{st} and 2\textsuperscript{nd} order derivatives of $p(x|\theta_i)$ with respect to $\theta_i$ exist and if $E \left[ \frac{\partial \log p(x|\theta_i)}{\partial \theta_i} \right] = 0$ (which guarantees that the MLE $\hat{\theta}_i \to \theta_i$ as $n \to \infty$), then the generalized likelihood ratio statistic, based on an observation $X = (X_1, \ldots, X_n)$,

$$\hat{\Lambda}_n(X) = \frac{\max_{\theta_1} p(x|\theta_1)}{\max_{\theta_0} p(x|\theta_0)}$$

has the following asymptotic distribution under $H_0$:

$$2 \log \hat{\Lambda}(x) \overset{n \to \infty}{\sim} \chi^2_{k-\ell} \quad \text{i.e.,} \quad 2 \log \hat{\Lambda}(x) \overset{D}{\to} \chi^2_{k-\ell}$$

Proof: (Sketch) under the conditions of the theorem, the log GLRT tends to the log GLRT in a Gaussian setting according to the Central Limit Theorem (CLT).
Example: Nested Condition

\[ H_0 : x_i \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ H_1 : x_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), i = 1, 2, \ldots, n, \sigma^2 > 0 \text{ unknown} \]

\[ \text{log LR:} \]

\[ \text{MLE of } \sigma^2: \]
\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \]

\[ \text{log GLR under } H_0: \]