ECE 830 Fall 2011 Statistical Signal Processing

instructor: R. Nowak

Lecture 11: The Generalized Likelihood Ratio

The generalized likelihood ratio test 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

$$\widehat{\Lambda}(x) \ = \ \frac{\max_{\theta_1 \in \Theta_1} p_1(x|\theta_1)}{\max_{\theta_0 \in \Theta_0} p_0(x|\theta_0)} \ \underset{H_0}{\overset{H_1}{\gtrless}} \ \gamma \ ,$$

or equivalently

$$\log \widehat{\Lambda}(x) \ \underset{H_0}{\overset{H_1}{\gtrless}} \ \gamma \ .$$

1 Example - Signal Detection

Consider two hypotheses

$$\begin{aligned} H_0 &\ : \ X \sim \mathcal{N}(0, \sigma^2 I_n) \\ H_1 &\ : \ X \sim \mathcal{N}(H\theta, \sigma^2 I_n) \end{aligned}$$

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

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

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

Log LR

$$\begin{aligned} \log \Lambda(x) &= -\frac{1}{2\sigma^2}(x - H\theta)^T (x - H\theta) + \frac{1}{2\sigma^2} x^T x \\ &= \frac{1}{\sigma^2}(\theta^T H^T x - \frac{1}{2}\theta^T H^T H\theta) \ . \end{aligned}$$

Since θ is unknown we can't go further, instead we find θ that makes x most likely:

$$\begin{aligned} \hat{\theta} = \arg \ \max_{\theta} \ p(x|H_1, \theta) & \\ &= \arg \ \max_{\theta} \ \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} e^{-\frac{1}{2\sigma^2}(x-H\theta)^T(x-H\theta)} \\ &= \arg \ \max_{\theta} \ -\frac{1}{2\sigma^2}(x - H\theta)^T (x - H\theta) \\ &= \arg \ \min_{\theta} \ (x - H\theta)^T (x - H\theta) \\ &= \arg \ \min_{\theta} \ (x^T x - \theta^T H^T x + \theta^T H^T H\theta) \end{aligned}$$

Taking the derivative with respect to θ

$$\frac{\partial}{\partial\theta}(x^T x - \theta^T H^T x + \theta^T H^T H\theta) = 0$$
$$\Rightarrow 0 - 2H^T x + 2H^T H\theta = 0$$
$$\Rightarrow \hat{\theta} = (H^T H)^{-1} H^T x$$

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

$$\log\hat{\Lambda}(x) := \frac{1}{\sigma^2}(x^T H(H^T H)^{-1}H^T x - \frac{1}{2}x^T H(H^T H)^{-1}H^T H(H^T H)^{-1}H^T x)$$
$$= \frac{1}{2\sigma^2}x^T H(H^T H)^{-1}H^T x$$

Recall that the projection matrix onto the subspace is defined as $P_H := H(H^T H)^{-1}H^T$

$$\Rightarrow \frac{1}{2\sigma^2}x^T P_H x \tag{1}$$
$$= \frac{1}{2\sigma^2}\|P_H x\|_2^2$$

Observe it is simply an energy detector in H, we are taking the projection of x into 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. From Equation (1) we can choose a γ for the desired P_{FA}:

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

What is the distributions of $x^T P_H x$ under H_0?

$$P_H = UU^T\ ,\ \textit{where } U_{nxk} \textit{ with orthonormal columns spanning columns of } H.$$
$$x^T P_H x = x^T UU^T x = y^T y\ ,\ y_{kx1} = U^T x$$
$$\frac{1}{\sigma^2}x^T P_H x = \frac{y^T y}{\sigma^2}$$
$$y \sim \mathcal{N}(0, \sigma^2 U^T U) \equiv \mathcal{N}(0, \sigma^2 I_{kxk})$$
$$y_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)\ ,\ i = 1, \cdots, k$$
$$\Rightarrow \frac{y}{\sigma} \sim \mathcal{N}(0, I_{kxk})$$
$$\Rightarrow \frac{y^T y}{\sigma^2} \sim \chi_k^2\ ,\ \text{chi-square with k-degrees of freedom}$$

GLRT and P_{FA}

$$\frac{1}{\sigma^2}x^T P_H x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$
$$\textit{under } H_0,\ \frac{1}{\sigma^2}x^T P_H x \sim \chi_k^2\ ,\ \textit{i.e., under } H_0:\ 2\log\hat{\Lambda} \sim \chi_k^2$$
$$P_{FA} = \mathbb{P}(\chi_k^2 > \gamma)$$

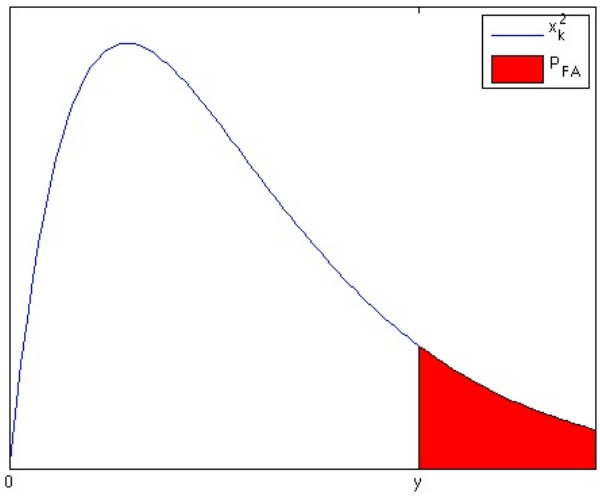

Figure 1: The P_{FA} of a χ_k^2 distribution.

1.0.1 χ_k^2 Distributions

To calculate the tails on χ_k^2 distributions (as in Figure 1 and 2) you can look it up in the back of a good book or use Matlab (chi2cdf(x,k), chi2inv(γ,k), chi2cdf(x,k)). Remember the mean of a χ_k^2 distribution is k, so we want to choose a γ bigger than k to produce a small P_{FA}.

2 Wilks' Theorem

Wilk's Theorem was established in 1938 [2].

Theorem 1 *Consider a composite hypothesis testing problem*

$$
\begin{aligned}
H_0 \quad &: \quad X_1, X_2, ..., X_n \overset{iid}{\sim} p(x|\theta_0)\ , \\
& \qquad \textit{where } \theta_{0,1}, \ldots, \theta_{0,\ell} \in \mathbb{R} \textit{ are free parameters and} \\
& \qquad \theta_{0,\ell+1} = a_{\ell+1}, \ldots, \theta_k = a_k \textit{ are fixed at the values } a_{\ell+1}, \ldots, a_k \\
H_1 \quad &: \quad X_1, X_2, ..., X_n \overset{iid}{\sim} p(x|\theta_1)\ ,\ \theta_1 \in \mathbb{R}^k \textit{ are all free parameters}
\end{aligned}
$$

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

$$\hat{\Lambda}_n(X) = \frac{\max\limits_{\theta_1} p(x|\theta_1)}{\max\limits_{\theta_0} p(x|\theta_0)} \tag{2}$$

has the following asymptotic distribution when $X \sim p(x|\theta_0$ (a model in hypothesis H_0):

$$2 \log \hat{\Lambda}(x) \overset{n \to \infty}{\sim} \chi^2_{k-\ell}$$

$$\textit{i.e. } 2 \log \hat{\Lambda}(x) \xrightarrow{D} \chi^2_{k-\ell}$$

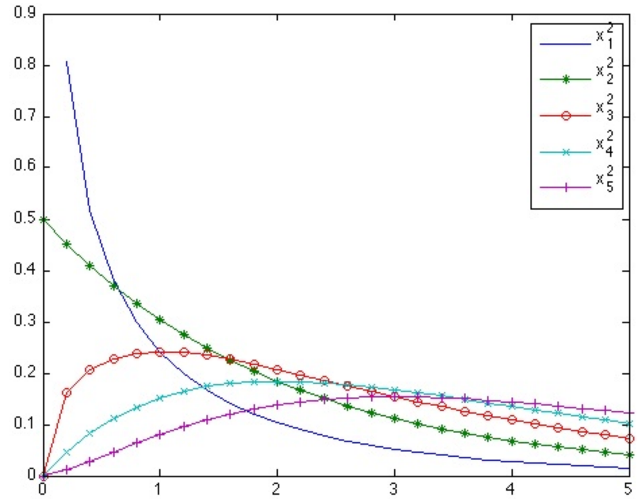

Figure 2: χ_k^2 distributions, for $k > 2$ they all take on the same general form.

Proof: (Sketch) under the conditions of the theorem, the log GLRT tends to log GLRT in Gaussian setting (aka the Central Limit Theorem (CLT)).

2.1 Example of a Nested Condition

$$H_0 : x_i \overset{iid}{\sim} \mathcal{N}(0,1)$$

$$H_1 : x_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)\ ,\ i = 1,\ 2,\ \cdots,\ n\ \sigma^2 > 0\ unknown$$

log LR:

$$\sum \left(-\frac{1}{2} log_e \sigma^2 - x_i^2 \left(\frac{1}{2\sigma^2} - \frac{1}{2} \right) \right)$$

MLE of σ^2:

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

log GLRT:

$$2 \left(\sum -\frac{1}{2} log_e \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 \right) - \frac{x_i^2}{2} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - 1 \right) \right) \overset{n \to \infty}{\sim} \chi_1^2\ ,\ under\ H_0$$

2.2 Example Multiple Source Internet Tomography

Wilk's theorem does have real world application, it was used in a computer network to determine the network topology. It worked well in simulation as well as in practice [1].

References

[1] M.G. Rabbat, M.J. Coates, and R.D. Nowak. Multiple-Source internet tomography. *Selected Areas in Communications, IEEE Journal on*, 24(12):2221–2234, 2006.

[2] S. S. Wilks. The Large-Sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, March 1938. ArticleType: research-article / Full publication date: Mar., 1938 / Copyright 1938 Institute of Mathematical Statistics.