ECE 830 Fall 2010 Statistical Signal Processing

instructor: R. Nowak , **scribe:** C. Hall

Lecture 10: The Generalized Likelihood Ratio

1 Composite Hypothesis Tests

Recall the composite hypothesis testing problem:

$$H_0 : X \sim \mathcal{N}(0, 1)$$
$$H_1 : X \sim \mathcal{N}(\mu, 1) \ , \ \mu > 0 \ \textit{unknown}$$

The densities look like:

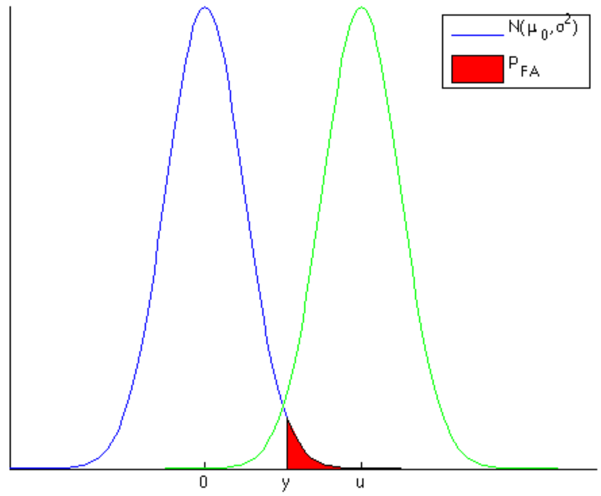


Figure 1: P_{FA} given $\mu > 0$, γ.

The Probability of False Alarm (P_{FA}) is the shaded area to the right of the threshold γ. It is easy to see that the Likelihood Ratio Test (LRT) at threshold γ is the most powerful test (by Neyman-Pearson (NP) Lemma) for every $\mu > 0$, for a given P_{FA}. In otherwords, the test is Uniformly Most Powerful (UMP, Karlin-Rubin Theorem).

2 Wald Test

Now consider:

$$H_0 : X \sim \mathcal{N}(0, 1)$$
$$H_1 : X \sim \mathcal{N}(\mu, 1) \ , \ \mu \neq 0$$

We considered the Wald Test test which is of the form: Wald Test

$$|x| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{1}$$

We can set γ for a desired P_{FA}, but it isn't UMP for all $\mu \neq 0$ for example, if $\mu > 0$ then the one-sided threshold test $x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$ is more powerful, see Figure **??**.

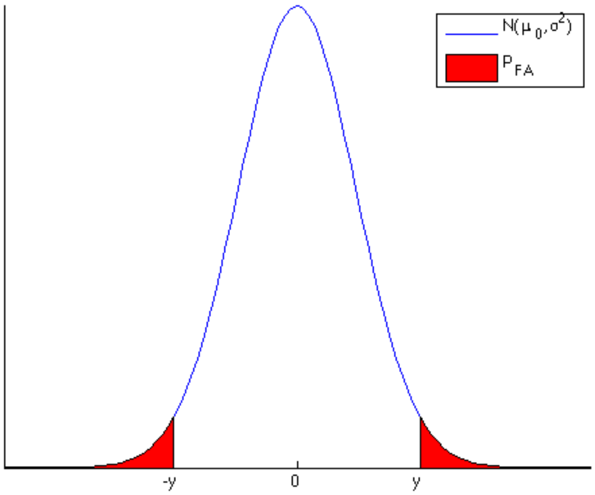
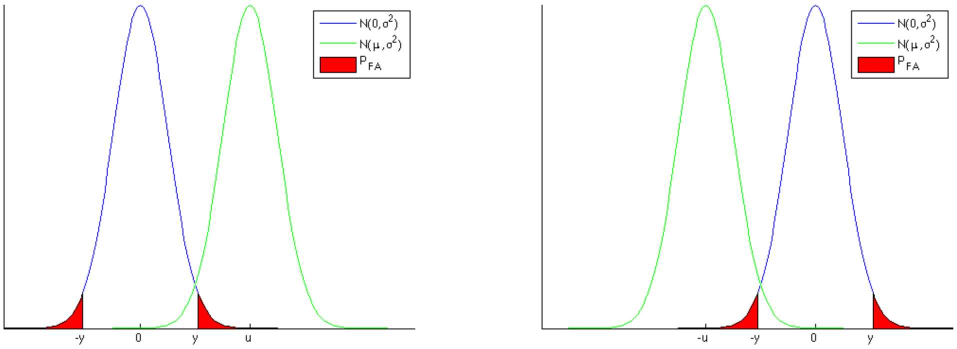


Figure 2: P_{FA} given $\mu \neq 0$, γ (The Wald Test).



(a) Case 1: $\mu > 0$

(b) Case 2: $\mu < 0$

Figure 3: The Wald test is not UMP for a given P_{FA}, because if it is found $\mu > 0$ then the LRT is the most powerful test

Another way to arrive at the Wald test is to use the data to estimate μ and plug the estimate, $\hat{\mu}$, into the LRT. The logLRT has the form

$$\mu x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \ (Wald\ Test)$$

and the natural estimate for μ in H_1 is $\hat{\mu} = x$ (if given one data point). Substituting $\hat{\mu}$ for μ for μ gives:

$$x^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \ (which\ is\ equivalent\ to\ the\ Wald\ Test.)$$

3 Generalized Likelihood Ratio Test

Suppose we have the following composite hypothesis testing problem

$$H_0 : \ X \sim \mathcal{N}(0,1)$$

$$H_1 : \ X \sim \mathcal{N}(\mu,1) \ \ \mu \neq 0 \ , \ \theta_0, \ \theta_1 \ unknown$$

The Generalized Likelihood Ratio Test (GLRT) is:

$$\frac{\max\limits_{\theta_1} p(x|H_1,\theta_1)}{\max\limits_{\theta_0} p(x|H_0,\theta_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \ , \ (GLRT) \tag{2}$$

We pick the density that is the highest probability for x. (ie $\mu \hat{=} x$ where $x \sim \mathcal{N}(\mu,1)$).
In other words, each hypothesis is composite, meaning that $x \sim p_i$ with

$$p_i \in \{p(x|H_i,\theta_i)\}_{\theta_i \in \Theta_i}$$

and we simply select the density from this collection that places the largest possible probability on the data x, see Figure 4.

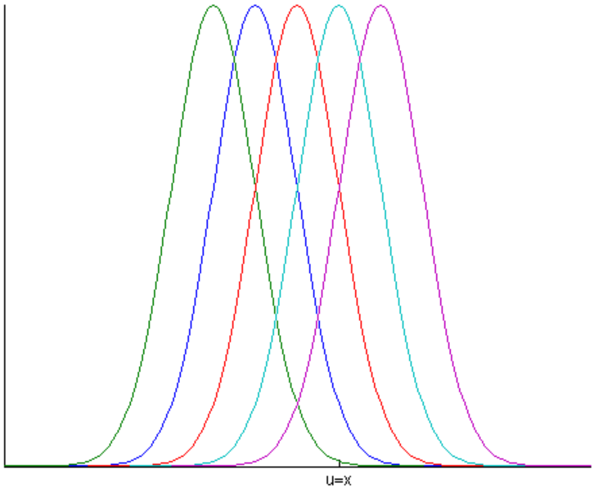


Figure 4: Maximum Likelihood Estimate (MLE), given data choose/estimate the parameter that fits the data best.

With a fixed observation of x we view $p(x|H_i,\theta_i)$ as a function of θ_i. This function is called the likelihood function of θ_i given x. The value of $p(x|H_i,\theta_i)$, for a specific θ_i, is called the likelihood of θ_i given x.

$$\hat{\theta}_i := arg \max_{\theta_i} p(x|H_i, \theta_i)$$

is called the Maximum Likelihood Estimate of θ_i.

Log LR

$$log_e \Lambda(x) := log_e \frac{p(x|H_1, \theta_1)}{p(x|H_0, \theta_0)} \tag{3}$$

Log GLRT

$$log_e \hat{\Lambda}(x) := log_e \frac{p(x|H_1, \hat{\theta}_1)}{p(x|H_0, \hat{\theta}_0)} \tag{4}$$

We have considered the case where the parameters are unknown and deterministic, we will ignore the case where the parameters themselves are random variables, for which the MLE may not be ideal. If the parameters were not deterministic, it may fit a Gaussian mixed fit better, see Figure 5 .

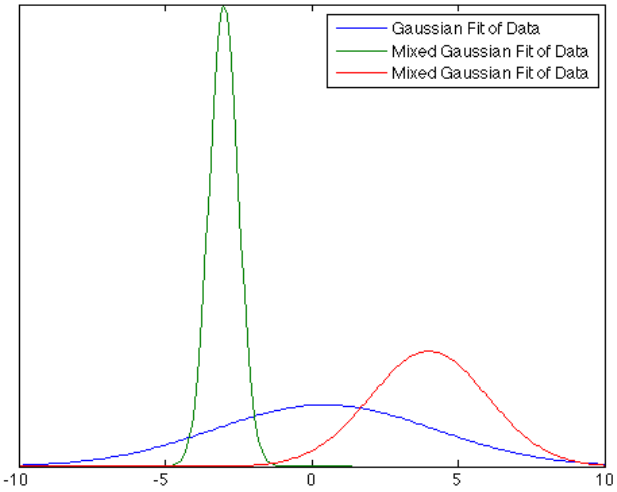


Figure 5: Raw data was used to estimate the Gaussian parameters for the Gaussian Fit, the data was generated using two distinct Gaussian models, the Gaussian Mixed Fit.

3.1 Example - Signal Processing

$$H_0 : \ x \sim \mathcal{N}(0, \sigma^2 I)$$

$$H_1 : \ x \sim \mathcal{N}(H\theta, \sigma^2 I)$$

$$\sigma^2 > 0 \ known \ , \ H_{nxk} \ known \ , \ \theta_{kx1} \ unknown$$

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

Log LR

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

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

$$\begin{aligned}\hat{\theta} &= arg \max_{\theta} p(x|H_1,\theta) \\ &= arg \max_{\theta} \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} e^{-\frac{1}{2\sigma^2}(x-H\theta)^T(x-H\theta)} \\ &= arg \max_{\theta} -\frac{1}{2\sigma^2}(x-H\theta)^T(x-H\theta) \\ &= arg \min_{\theta} (x-H\theta)^T(x-H\theta) \\ &= arg \min_{\theta} (x^Tx - \theta^T H^T x + \theta^T H^T H\theta) \\ &\Rightarrow \frac{\partial}{\partial\theta}(x^Tx - \theta^T H^T x + \theta^T H^T H\theta) = 0\end{aligned}$$

$$\Rightarrow 0 - 2H^Tx + 2H^TH\theta = 0$$
$$\Rightarrow \hat{\theta} = (H^TH)^{-1}H^Tx$$

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

$$log_e\hat{\Lambda}(x) := \frac{1}{\sigma^2}(x^TH(H^TH)^{-1}H^Tx - \frac{1}{2}x^TH(H^TH)^{-1}H^TH(H^TH)^{-1}H^Tx)$$
$$= \frac{1}{2\sigma^2}x^TH(H^TH)^{-1}H^Tx$$

Note: the Projection Matrix is defined as $P_H := H(H^TH)^{-1}H^T$

$$\Rightarrow \frac{1}{2\sigma^2}x^TP_Hx \tag{5}$$
$$= \frac{1}{2\sigma^2}\|P_Hx\|_2^2$$

Observe it is simply an energy detector in H, we are taking the projection of x into H and measuring the energy, see Figure 6.

$$\mathbb{E}_{H_0}\left[\|P_H\|_2^2\right] = I_{kxk}^T I_{kxk} = k\sigma^2$$

3.1.1 Analysis of GLRT performance:

From Equation (5) we can choose a γ for the desired P_{FA}:

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

What is the distributions of x^TP_Hx <u>under H_0</u>?

$$P_H = UU^T \text{ , } where\ U_{nxk}\ with\ orthonormal\ columns\ spanning\ columns\ of\ H.$$

$$x^TP_Hx = x^TUU^Tx = y^Ty \text{ , } y_{kx1} = U^Tx$$
$$\frac{1}{\sigma^2}x^TP_Hx = \frac{y^Ty}{\sigma^2}$$
$$y \sim \mathcal{N}(0, \sigma^2U^TU) \equiv \mathcal{N}(0, \sigma^2 I_{kxk})$$

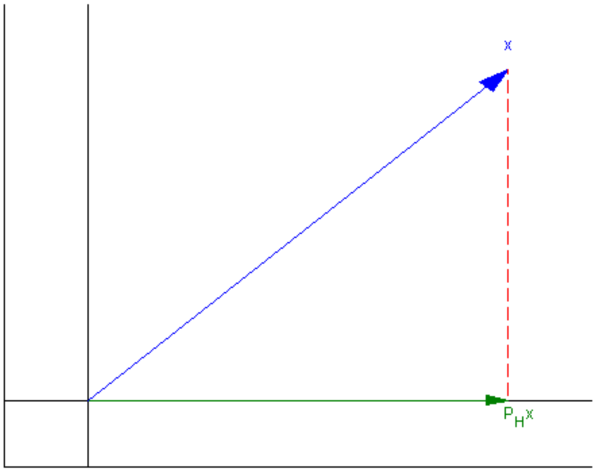


Figure 6: The projection of x onto a subspace H, a simple example.

$$y_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)\ ,\ i=1,\cdots,k$$
$$\Rightarrow \frac{y}{\sigma} \sim \mathcal{N}(0, I_{kxk})$$
$$\Rightarrow \frac{y^Ty}{\sigma^2} \sim \chi_k^2\ ,\ Chi - square\ with\ k - degrees\ of\ freedom$$

<u>GLRT and P_{FA}</u>

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

3.1.2 χ_k^2 Distributions

To calculate the tails on χ_k^2 distributions (as in Figure 7) 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}.

4 Wilks' Theorem

Wilk's Theorem was established in 1938 read his paper for the proof. [2] Consider a composite hypothesis testing problem

$$H_0:\ x_1,x_2,...,x_n \overset{iid}{\sim} p(x|H_0,\theta_0)\ for\ \theta_0 \in \mathbb{R}^\ell$$
$$H_1:\ x_1,x_2,...,x_n \overset{iid}{\sim} p(x|H_1,\theta_1)\ for\ \theta_1 \in \mathbb{R}^k\ \ k > \ell$$

<u>Nested</u> where p has same form for H_0, H_1
and θ_0 is fixed for $i = \ell+1,..,k$

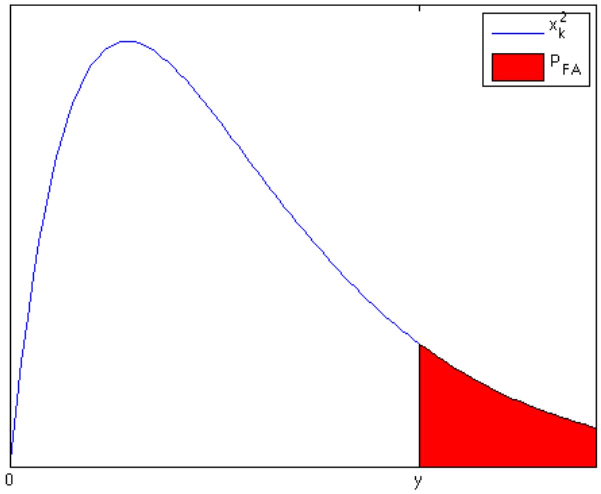


Figure 7: The P_{FA} of a χ_k^2 distribution.

$\theta_1 \in \mathbb{R}^k \ , \ k > \ell$

Otherwise the parameters are unknown.

$$H_1: \ x_1, x_2, ..., x_n \overset{iid}{\sim} p(x|H_1, \theta_1) \ for \ \theta_1 \in \mathbb{R}^k \quad k > \ell$$

Then if the 1^{st} and 2^{nd} order derivatives of $p(x|H_i, \theta_i)$ with respect to θ_i exist and if $\mathbb{E}\left[\frac{\partial log_e p(x|H_i,\theta_i)}{\partial \theta_i}\right] = 0$ (which guarantees that the MLE $\hat{\theta}_i \to \theta_i$ (true) in limit) then the GLRT

$$\hat{\Lambda}_h(x) = \frac{\max\limits_{\theta_1} p(x|H_1, \theta_1)}{\max\limits_{\theta_0} p(x|H_0, \theta_0)} \tag{6}$$

with $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is well-defined and under H_0

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

$$i.e. \ 2log_e\hat{\Lambda}(x) \overset{D}{\to} \chi^2_{k-\ell}$$

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

4.1 Example of a Nested Condition

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

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

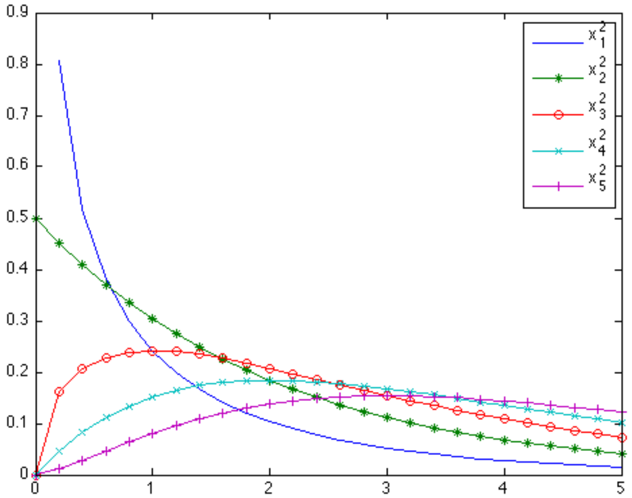


Figure 8: χ_k^2 distributions, for $k > 2$ they all take on the same general form.

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:

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

4.2 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]

5 Random Parameters

$$H_0 : x \sim p(x|H_0,\theta_0),\ \theta_0 \sim p(\theta_0|H_0)$$
$$H_1 : x \sim p(x|H_1,\theta_1),\ \theta_1 \sim p(\theta_1|H_1)$$

LR:

$$\frac{p(x|H_1,\theta_1)}{p(x|H_0,\theta_0)}$$

Bayes Factor:

$$\frac{\int p(x|H_1,\theta_1)\,p(\theta_1|H_1)\,d\theta_1}{\int p(x|H_0,\theta_0)\,p(\theta_0|H_0)\,d\theta_0} = \frac{p(x|H_1)}{p(x|H_0)} \tag{7}$$

$p(x|H_i)$ is called the marginal likelihood of x given H_i.

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.