Recall last lecture

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

with \( \sigma^2 \) known, \( H_{n \times k} \) known, and \( \theta_{k \times 1} \) unknown. GLRT:

\[
2 \log \hat{\Lambda}(x) \frac{x^T P_H x}{2\sigma^2} \overset{H_0}{\gtrless} \gamma
\]

Under \( H_0 \),

\[
\frac{x^T P_H x}{2\sigma^2} \sim \chi^2_k
\]

**Example 1** Generalize the above question:

\[ H_0 : X \sim \mathcal{N}(0, \Sigma) \]
\[ H_1 : X \sim \mathcal{N}(H\theta, \Sigma) \]

with \( \Sigma \) known, \( H_{n \times k} \) known, and \( \theta_{k \times 1} \) unknown.

**1 Wilk’s Theorem**

**Theorem 1** Suppose \( H_0 \) and \( H_1 \) composite with

\[ H_0 \text{ models } l \text{ dofs} \subset H_1 \text{ models, } k > l \text{ dofs} \]

Then under mild regularity assumptions, if

\[
X_1, \ldots, X_n \overset{i.i.d.}{\sim} \chi^2_{k-l}
\]

Then, under \( H_0 \)

\[
\underbrace{2 \log \hat{\Lambda}_n(x)}_{\log \text{ GLRT}} \overset{n \rightarrow \infty}{\sim} \chi^2_{k-l}
\]
2 Unknown Noise Level

Now let’s look at case where noise level is unknown. Suppose

\[ H_0 : X \sim \mathcal{N}(0, \sigma^2 I) \]
\[ H_1 : X \sim \mathcal{N}(s, \sigma^2 I) \]

where \( \sigma^2 > 0 \) is unknown and \( s \) is \( n \times 1 \) and known.

log Likelihood Ratio:

\[ \log \Lambda(x) = -\frac{1}{2\sigma^2} (x - s)^T (x - s) + \frac{1}{2\sigma^2} x^T x \]

So our test is equivalent to

\[ \frac{1}{\sigma^2} s^T x \begin{cases} \geq \frac{H_1}{H_0} \gamma' \\ \geq \gamma \end{cases} \]

or

\[ t(x) := s^T x \begin{cases} \geq \frac{H_1}{H_0} \gamma' \\ \geq \gamma \end{cases}, \text{ since } \sigma^2 > 0 \]

Then what is the distribution of \( t(x) \)?

\[ H_0 : t(x) \sim N(0, \sigma^2 s^T s) \]
\[ H_1 : t(x) \sim N(s^T s, \sigma^2 s^T s) \]

Both distributions depend on unknown \( \sigma^2 \)!

Let’s look at the GLRT. The MLE for \( \sigma^2 \) is

\[ \hat{\sigma}^2_i = \arg \max_{\sigma^2} \mathbb{P}(x|H_i), \ i = 0, 1 \]
For \( H_0 \) we have
\[
\hat{\sigma}_0^2 = \arg \max_{\sigma^2} 1/2\pi\sigma^2)^{n/2} \exp \left( -\frac{1}{2\sigma^2} \right) x^T x
\]
\[
= \arg \max -\frac{n}{2} \left( \log \sigma^2 + \log 2\pi - \frac{1}{2\sigma^2} x^T x \right)
\]
\[
= \arg \min \frac{n}{2} \log \sigma^2 + \frac{1}{\sigma^2} x^T x
\]
Take derivative with respect to \( \sigma^2 \)
\[
\Rightarrow \frac{\partial}{\partial \sigma^2} \mathbb{P}(x|H_0) = \frac{n}{2} \sigma^2 - \frac{1}{2\sigma^4} x^T x = 0
\]
\[
\Rightarrow \hat{\sigma}_0^2 = \frac{1}{n} x^T x
\]
Similarly,
\[
\hat{\sigma}_1^2 = \frac{1}{n} (x - s)^T (x - s)
\]
So the GLRT is
\[
\hat{\Lambda}(x) = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp \left( -\frac{1}{2\sigma_0^2} (x - s)^T (x - s) \right)
\]
the log GLRT is
\[
\log \hat{\Lambda}(x) = \frac{n}{2} \log \left( \frac{x^T x}{(x - s)^T (x - s)} \right)
\]
What is the distribution of \( X \) under \( H_0 ? \)
\( X \sim \mathcal{N}(0, \sigma^2 I) \)
\[
\Rightarrow \log \hat{\Lambda}(x) = \frac{n}{2} \log \left( \frac{\sigma^2 w^T w}{(\sigma w - s)^T (\sigma w - s)} \right), \text{ where } w \sim \mathcal{N}(0, \sigma^2 I)
\]
And, Wilks’ Theorem doesn’t apply since both \( H_0 \) and \( H_1 \) models have one degree of freedom. We cannot set \( \gamma \) to control either error.

3 Unknown Signal and Noise Amplitudes

Let’s look at a slightly different problem. The problem in the previous case is that the unknown noise amplitude affected the variance of both distributions, and the MLE of the noise variance differed in the two hypotheses. Let us now suppose:
\[
H_0 : X \sim \mathcal{N}(0, \sigma^2 I)
\]
\[
H_1 : X \sim \mathcal{N}(\theta s, \sigma^2 I)
\]
with \( \sigma^2 \) unknown, \( \theta \) unknown, and \( s_{n \times 1} \) unknown.
Under \( H_0 \)
\[
X = \sigma w, \quad w \sim \mathcal{N}(0, I)
\]
Under $H_1$

\[ X = \theta s + \sigma w = \sigma(\theta' s + w), \text{ where } \theta' = \frac{\theta}{\sigma} \]

The advantage here is that $\sigma$ can be viewed as a scaling factor for the observation in both cases.

Let’s consider the GLRT for this problem. Under $H_0$

\[ \hat{\sigma}_0^2 = \frac{1}{n} x^T x, \text{ as before}. \]

Under $H_1$, we must find the MLE of $\theta$ and $\sigma^2$.

\[ P(x|H, \theta, \sigma^2) = \frac{1}{(s\pi \sigma^2)^{n/2}} \exp \left( -\frac{1}{2\sigma^2} (x - \theta s)^T (x - \theta s) \right) \]

Taking the log we have:

\[ -\frac{n}{2} \left( \log \sigma^2 + \log 2\pi \right) - \frac{1}{2\sigma^2} (x^T x - 1\theta s^T x + \theta^2 s^T s) \]

Differentiating with respect to $\sigma^2$:

\[ s^T x = \theta s^T s \]

\[ \Rightarrow \hat{\theta} = \frac{s^T x}{s^T s} \]

Differentiating with respect to $\sigma^2$:

\[ -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^2} (x - \hat{\theta} s)^T (x - \hat{\theta} s) = 0 \]

\[ \Rightarrow \hat{\sigma}_1^2 = \frac{1}{n} (x - \hat{\theta} s)^T (x - \hat{\theta} s) \]

**Note:** Wilks’ Theorem applies:

\[ 2 \log \frac{P(X|H_1, \hat{\theta}, \hat{\sigma}_1^2)}{P(X|H_0, \hat{\sigma}_0^2)} \sim \chi_1^2 \text{ for large } n \]

Let’s look at the GLRT more closely.

\[ \hat{\Lambda}(x) = \frac{P(x|H_1, \hat{\theta}, \hat{\sigma}_1^2)}{P(x|H_0, \hat{\sigma}_0^2)} \]

\[ = \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \right)^{n/2} \exp \left( -\frac{1}{2\hat{\sigma}_1^2} (x - \hat{\theta} s)^T (x - \hat{\theta} s) + \frac{1}{2\hat{\sigma}_0^2 x^T x} \right) \]

\[ = \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \right)^{n/2} \exp \left( -\frac{n}{2} + \frac{n}{2} \right) \]

\[ = \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \right) \]
So the GLRT has the simple form

\[
\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} = \frac{x^T x}{(x - \hat{\theta}_s)^T(x - \hat{\theta}_s)} \quad H_1 \overset{\gtrless}{\sim} H_0 \quad \gamma
\]

or equivalently

\[
\frac{x^T x}{(x - \frac{s s^T x}{s^T s})(x - \frac{s s^T x}{s^T s})} \quad H_1 \overset{\gtrless}{\sim} H_0 \quad \gamma
\]

Under \( H_0 \) we have \( X = \sigma w, w \sim \mathcal{N}(0, I) \), and so the test statistic is

\[
\frac{\sigma^2 w^T w}{\sigma^2 (w - P_s w)^T(w - P_s w)} \rightarrow \text{invariant to } \sigma^2
\]

**Definition 1** \( \gamma \) can be chosen to insure a special \( P_{FA} \) for every value of \( \sigma^2 \). A test like this is said to have a **constant false alarm rate** and is called CFAR detector.

To set \( \gamma \) we need to determine the distribution of \( \frac{w^T w}{(w - P_s w)^T (w - P_s w)} \), \( w \sim \mathcal{N}(0, I) \). Consider the test statistic

\[
t(x) = \frac{x^T x}{(x - P_s x)^T (x - P_s x)} = \frac{x^T (I - P_s) x + x^T P_s x}{x^T (I - P_s) x}
\]

\[
= 1 + \frac{x^T P_s x}{x^T (I - P_s) x}
\]

So equivalently, we can write the GLRT as:

\[
\frac{x^T P_s x}{x^T (I - P_s) x} \quad H_1 \overset{\gtrless}{\sim} H_0 \quad \gamma
\]

Let \( U \) be \( n \times (n - 1) \) matrix whose orthonormal columns span subspace orthogonal to \( s \).

\[
U = [u_1, \ldots, u_{n-1}]
\]

Then \( u_1, \ldots, u_{n-1}, \frac{s}{\|s\|} \) are orthonormal basis for \( \mathbb{R}^n \)

\[
x^T P_s x = \frac{|s^T x|^2}{\|s\|^2} = \sum_{i=1}^{n-1} |u_i^T x|^2
\]

Under \( H_0 \)

\[
\frac{|s^T w|^2}{\|s\|^2} \sim \chi_1^2
\]

\[
\sum_{i=1}^{n-1} (u_i^T w)^2 \sim \chi_{n-1}^2
\]

Moreover, because \( u_1, \ldots, u_{n-1} \) are orthogonal to \( s \), the \( s^T x \) and \( u_i^T x \), \( i = 1, \ldots, n-1 \) are uncorrelated and thus independent.
\[ \Rightarrow \frac{(x^Tw)^2}{\|x\|^2} \] are independent!

The ratio of independent \( \chi^2 \) random variables with degree of freedom \( k \) and \( l \) respectively has been well studied and has a name: F-distributed with \( (k, l) \) degree of freedom

\[ \frac{\chi^2_k}{\chi^2_l} \sim F_{k, l} \]

In our case, under \( H_0 \)

\[ \frac{x^TP_sx}{x^T(I-P_s)x/(n-1)} \sim F_1, n-1 \]

and thus we can use the tail of the F-distribution to set a threshold for a desired \( P_{FA} \).

Note:
If \( X \sim F(\nu_1, \nu_2) \), then \( Y = \lim_{\nu_2 \to \infty} \nu_1 X \) has the chi-square distribution \( \chi^2_{\nu_1} \).

In our case, \( \lim_{n \to \infty} F_{1,n-1} \sim \chi^2_{1} \), which is what Wilks’ Theorem told us.