The Generalized Likelihood Ratio Test

We focus here on a binary composite hypothesis testing problem with $H_0 : x \in X_0$ versus $H_1 : x \in X \setminus X_0$.

The main idea of the GLRT is to

- get an observation $y$
- estimate the most likely value of $x$ under $H_0$ (call this $\hat{x}_0$)
- estimate the most likely value of $x$ under $H_1$ (call this $\hat{x}_1$)

and then use those estimates as “truth” so that we have a simple binary hypothesis testing problem $H_0 : x = \hat{x}_0$ versus $H_1 : x = \hat{x}_1$.

You can then specify the decision rule via the standard N-P lemma for simple binary hypothesis testing.
Connection to Bayesian Composite Hypothesis Testing

Let $p_i(y; x)$ be the family of densities parameterized by $x$ under hypothesis $\mathcal{H}_i$. Often we have $p_0(y; x) = p_1(y; x)$, but these densities don’t have to have the same form.

With the GLRT, we decide $\mathcal{H}_1$ if

$$\frac{\max_{x \in X \setminus X_0} p_1(y; x)}{\max_{x \in X_0} p_0(y; x)} > v$$

In the case of Bayesian binary hypothesis testing, we can show that we decide $\mathcal{H}_1$ if

$$\frac{\int_{x \in X \setminus X_0} p_1(y|x) \, dx}{\int_{x \in X_0} p_0(y|x) \, dx} > v$$

Intuition: The GLRT decision rule compares the most likely model in $\mathcal{H}_1$ to the most likely model in $\mathcal{H}_0$. The Bayesian decision rule compares the average model in $\mathcal{H}_1$ to the average model in $\mathcal{H}_0$, using the prior probability distribution on the unknown state.
GLRT Example (part 1 of 4)

Suppose we get a vector observation \( Y \sim \mathcal{N}(Hx, \sigma^2 I) \) in \( \mathbb{R}^n \) with \( \sigma^2 \) known and have two hypotheses: \( \mathcal{H}_0 : x = 0 \) versus \( \mathcal{H}_1 : x \neq 0 \) for \( x \in \mathbb{R}^N \). We assume \( H^\top H \) is invertible.

Given \( Y = y \), we want to find the most likely \( x \) under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \).

For the denominator of the GLRT, we compute \( \max_{x \in \mathcal{X}_0} p_0(y; x) \). But \( \mathcal{X}_0 = \{0\} \), so the maximization is trivial. The denominator of the GLRT is

\[
\max_{x \in \mathcal{X}_0} p_0(y; x) = p_0(y; x = 0) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left( -\frac{y^\top y}{2\sigma^2} \right)
\]

For the numerator, we compute \( \max_{x \neq 0} p_1(y; x) \). Since this is a linear Gaussian model, we can use the known results for the MLE to write

\[
\hat{x}_1 = (H^\top H)^{-1} H^\top y.
\]

Hence

\[
\max_{x \in \mathcal{X} \setminus \mathcal{X}_0} p_1(y; x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left( -\frac{(y - H\hat{x}_1)^\top (y - H\hat{x}_1)}{2\sigma^2} \right)
\]
The GLRT is then

\[
\frac{\max_{x \in \mathcal{X} \setminus \mathcal{X}_0} p_1(y; x)}{\max_{x \in \mathcal{X}_0} p_0(y; x)} = \frac{\exp \left( -\frac{(y-H\hat{x}_1)\top(y-H\hat{x}_1)}{2\sigma^2} \right)}{\exp \left( -\frac{y\top y}{2\sigma^2} \right)} > v
\]

with \( \hat{x}_1 = (H\top H)^{-1}H\top y = Py \). Simplifying and taking the log of both sides, we have

\[
\frac{-1}{2\sigma^2} \left( y\top y - 2y\top H\hat{x}_1 + \hat{x}_1\top H\top H\hat{x}_1 - y\top y \right) > v'
\]

\[
\Leftrightarrow 2y\top H(H\top H)^{-1}H\top y - y\top H(H\top H)^{-1}H\top H(H\top H)^{-1}H\top y > v''
\]

\[
\Leftrightarrow y\top H(H\top H)^{-1}H\top y > v''
\]

\[
\Leftrightarrow y\top Py > v''
\]

where we choose \( v'' \) to satisfy the false positive probability constraint.
GLRT Example (part 3 of 4)

Note that we can write \( H = QR \) where \( Q \in \mathbb{R}^{N \times n} \) is a matrix with orthonormal columns and \( R \in \mathbb{R}^{n \times n} \) is an invertible upper triangular matrix. This is called the (reduced) QR factorization.

Then

\[
P = H (H^\top H)^{-1} H^\top
\]

\[
= QR (R^\top Q^\top QR)^{-1} R^\top Q^\top
\]

\[
= QR (R^\top IR)^{-1} R^\top Q^\top
\]

\[
= QRR^{-1} (R^\top)^{-1} R^\top Q^\top
\]

\[
= QQ^\top
\]

Hence our decision statistic is

\[
Y^\top PY = Y^\top QQ^\top Y = Z^\top Z.
\]

What is the distribution of \( Z \) under \( \mathcal{H}_0 \)?
GLRT Example (part 4 of 4)

We have $Z = Q^\top Y$ with $Y \sim \mathcal{N}(0, \sigma^2 I)$ under $\mathcal{H}_0$.

Clearly $Z$ is Gaussian with $E[Z] = E[Q^\top Y] = 0$.

We can also compute

$$E[ZZ^\top] = E[Q^\top YY^\top Q] = Q^\top (\sigma^2 I_{N \times N}) Q = \sigma^2 I_{n \times n}$$

So $Z \sim \mathcal{N}(0, \sigma^2 I)$ in $\mathbb{R}^n$ and $\frac{Z^\top Z}{\sigma^2} \sim \chi^2_n$.

Given a false positive probability constraint $\alpha$, you can use the inverse CDF of the Chi-squared distribution with $n$ degrees of freedom to find the optimum decision threshold.

For example, set $\alpha = 0.01$ and $n = 10$. In Matlab, you can use $v = \text{chi2inv}(0.99, 10)$ to get $v = 23.2093$. Then we decide $\mathcal{H}_1$ if $Z^\top Z > v\sigma^2$. 