(An Introduction to) Classical Composite Hypothesis Testing

First, recall that, in the composite testing case, we have Θ_0 and Θ_1 that form a *partition* of the parameter space Θ:

$$\Theta_0 \cup \Theta_1 = \Theta, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

and we wish to identify *which* of the following two hypotheses is true:

$$\begin{aligned} \mathcal{H}_0 &: \quad \theta \in \Theta_0, \quad \text{null hypothesis} \\ \mathcal{H}_1 &: \quad \theta \in \Theta_1, \quad \text{alternative hypothesis}. \end{aligned}$$

Here, we adopt the classical Neyman-Pearson approach — maximize the detection probability for a specified false-alarm rate.

Example: Suppose that we wish to detect an unknown positive DC level A ($A > 0$):

$$\begin{aligned} \mathcal{H}_0 : & \qquad x[n] = w[n], \quad n = 1, 2, \ldots, N \\ \mathcal{H}_1 : & \qquad x[n] = A + w[n], \quad n = 1, 2, \ldots, N \end{aligned}$$

where $w[n]$ is zero-mean white Gaussian noise with known variance σ^2. Here is an alternative formulation: Consider this family of probability density functions (pdfs):

$$p(\boldsymbol{x}\,;\,\theta) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \cdot \exp\Big[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-\theta)^2\Big] \qquad (1)$$

and the following (equivalent) hypotheses:

$$\begin{aligned} &\mathcal{H}_0: \quad \theta = 0 \quad \text{(signal absent)}, \ \Theta_0 = \{0\} \quad \text{versus} \\ &\mathcal{H}_1: \quad \theta = A > 0 \quad \text{(signal present)}, \ \Theta_1 = (0, \infty) \end{aligned}$$

where A is unknown, except for its sign. Let us try the classical Neyman-Pearson approach (which required the exact knowledge of A since we considered only simple hypotheses under the classical setting, until now):

$$\Lambda(\boldsymbol{x}) = \frac{1/(2\pi\sigma^2)^{N/2} \cdot \exp[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-A)^2]}{1/(2\pi\sigma^2)^{N/2} \cdot \exp(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}x[n]^2)} > \lambda.$$

Taking log etc. leads to

$$A\sum_{n=1}^{N} x[n] > \sigma^2 \log\lambda + N\,A^2/2.$$

Since we know that the DC level under $\mathcal{H}_1$ is positive (i.e. $A > 0$), we can divide both sides of the above expression by

NA and accept $\mathcal{H}_1$ if

$$T(\boldsymbol{x}) = \overline{x} = \frac{1}{N}\sum_{n=1}^{N} x[n] > \lambda'.$$

How to determine the threshold λ'? Under $\mathcal{H}_0 :\ T(\boldsymbol{X})\,|\,\theta = 0 \sim \mathcal{N}(0, \sigma^2/N)$ and hence

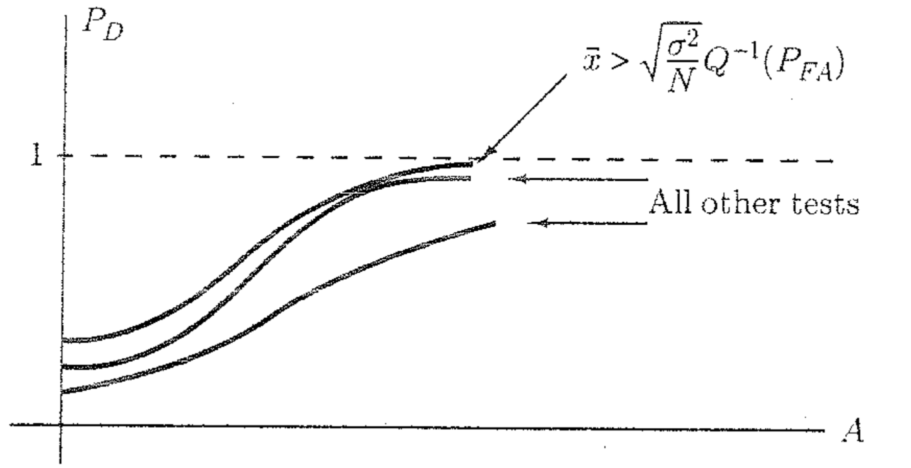
$$P_{\mathrm{FA}} \quad = \quad Q\Big(\frac{\lambda'}{\sqrt{\sigma^2/N}}\Big)$$

or

$$\lambda' = \sqrt{\frac{\sigma^2}{N}} \cdot Q^{-1}(P_{\mathrm{FA}}).$$

Thus, in this case, only the sign of the DC level A under $\mathcal{H}_1$ is needed to find the Neyman-Pearson test, because the pdf of $T(\boldsymbol{X})$ under $\mathcal{H}_0$ does not depend on A. Furthermore, under $\mathcal{H}_1 : T(\boldsymbol{X})\,|\,\theta = A \sim \mathcal{N}(A, \sigma^2/N) \Rightarrow$ clearly, P_{D} depends on A.

Since the Neyman-Pearson test is optimal in terms of maximizing the detection probability subject to a specified false-alarm probability, all other tests are poorer with respect to this criterion. Hence, Neyman-Pearson tests are said to be *uniformly most powerful (UMP)*.



UMP tests seldom exist. For example, if A can be negative as well (in the DC-level detection setting), the "Neyman-Pearson test" becomes

$$\begin{aligned} \overline{x} \quad &> \quad \sqrt{\frac{\sigma^2}{N}} \cdot Q^{-1}(P_{\mathrm{FA}}) \quad (A > 0) \\ &< \quad -\sqrt{\frac{\sigma^2}{N}} \cdot Q^{-1}(P_{\mathrm{FA}}) \quad (A < 0). \end{aligned}$$

If A is *completely unknown* (i.e. we do not know its sign), no single test is optimal and, therefore, no UMP test exists. However, for

$$\begin{aligned} \mathcal{H}_0 : \quad & \theta = 0 \quad \text{versus} \\ \mathcal{H}_1 : \quad & \theta = \underbrace{A}_{\text{unknown}} > 0 \end{aligned}$$

a (one-sided) UMP test does exist (as shown in our previous discussion). Clearly, for

$$\begin{aligned}
\mathcal{H}_0: &\quad \theta = 0 \quad \text{versus} \\
\mathcal{H}_1: &\quad \theta = \underbrace{A}_{\text{unknown}} < 0
\end{aligned}$$

a UMP test exists as well (by symmetry). For the two-sided case,

$$\begin{aligned}
\mathcal{H}_0: &\quad \theta = 0 \quad (\Theta_0 = \{0\}) \quad \text{versus} \\
\mathcal{H}_1: &\quad \theta = \underbrace{A}_{\text{unknown}} \neq 0 \quad (\Theta_1 = (-\infty, \infty)\backslash\{0\})
\end{aligned}$$

no UMP test exists. Here, we may choose

$$|\overline{x}| > \lambda''$$

but no optimality properties can be claimed.

Generalized Likelihood Ratio (GLR) Test

In the "classical" spirit, let us replace the unknown parameters by their maximum likelihood (ML) estimates under the two hypotheses. Hence, we accept $\mathcal{H}_1$ if

$$\Lambda_{\mathrm{GLR}}(\boldsymbol{x}) = \frac{\max_{\theta \in \Theta_1} p(\boldsymbol{x}\,;\,\theta)}{\max_{\theta \in \Theta_0} p(\boldsymbol{x}\,;\,\theta)} > \gamma.$$

This test has no UMP optimality properties, but often works well in practice.

(Back to) Example. DC level in additive white Gaussian noise (AWGN) with completely unknown A ($-\infty < A < \infty$, i.e. unknown sign as well) and known σ^2:

$$\begin{aligned} \mathcal{H}_0 :& \quad a = 0 \quad \text{versus} \\ \mathcal{H}_1 :& \quad a = \underbrace{A}_{\text{unknown}} \neq 0. \end{aligned}$$

Then, our GLR test is

$$\Lambda_{\mathrm{GLR}}(\boldsymbol{x}) = \frac{\overbrace{\max_a p(\boldsymbol{x}\,;\,a)}^{\text{effectively the same as } \max_{a \neq 0} p(\boldsymbol{x}\,;\,a)}}{p(\boldsymbol{x}\,;\,a = 0)} > \gamma$$

see also (1). Therefore,

$$\log \Lambda_{\mathrm{GLR}}(\boldsymbol{x}) = -\frac{1}{2\,\sigma^2}\Big\{\sum_{n=1}^{N}(x[n]-\overline{x})^2 - \sum_{n=1}^{N} x^2[n]\Big\} = \frac{N\,\overline{x}^2}{2\,\sigma^2}$$

or we accept $\mathcal{H}_1$ if

$$(\overline{x})^2 > \gamma'$$

or $|\overline{x}| > \lambda''$. This is the detector that we had guessed earlier. Let us compare this detector with the (unrealizable, also called *clairvoyant*) Neyman-Pearson detector that assumes the knowledge of the sign of A, the DC-level signal to be detected. Assuming that the sign of A is *known*, we can construct the UMP/Neyman-Pearson/clairvoyant detector, whose performance is described by

$$P_{\mathrm{D}} = Q\Big(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{d^2}\Big)$$

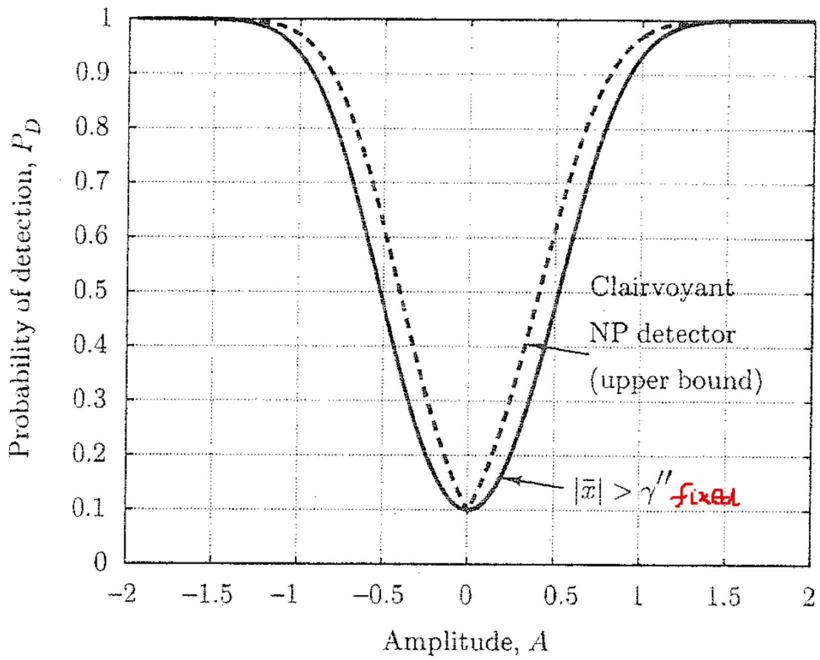
where $d^2 = N A^2/\sigma^2$, see handout # 5. All other detectors have P_{D} below this upper bound.

GLR test: Decide $\mathcal{H}_1$ if $|\overline{x}| > \gamma''$. To make sure that the GLR test is implementable, we must be able to specify a threshold γ'' independent of A. This is possible here, since $p(\boldsymbol{x}\,;\,\theta = 0)$

is *not* a function of A, the DC level under $\mathcal{H}_1$.

$$\begin{aligned}
P_{\rm FA} &= P[|\overline{X}| > \gamma''\,;\,\theta = 0] \qquad [\overline{X}\,|\,\theta = 0 \sim \mathcal{N}(0, \sigma^2/N)]\\
&\overset{\text{symmetry}}{=} 2\,P[\overline{X} > \gamma''\,;\,\theta = 0] = 2\,Q(\gamma''/\sqrt{\sigma^2/N})\\
P_{\rm D} &= P[|\overline{X}| > \gamma''\,;\,\theta = A] \qquad [\overline{X}\,|\,\theta = A \sim \mathcal{N}(A, \sigma^2/N)]\\
&= P[\overline{X} > \gamma''\,;\,\theta = A] + P[\overline{X} < -\gamma''\,;\,\theta = A]\\
&= Q\Big(\frac{\gamma'' - A}{\sqrt{\sigma^2/N}}\Big) + Q\Big(\frac{\gamma'' + A}{\sqrt{\sigma^2/N}}\Big)\\
&= Q\Big(Q^{-1}(P_{\rm FA}/2) - \frac{A}{\sqrt{\sigma^2/N}}\Big)\\
&\quad + Q\Big(Q^{-1}(P_{\rm FA}/2) + \frac{A}{\sqrt{\sigma^2/N}}\Big).
\end{aligned}$$

In this case, the GLR test is only slightly worse than the clairvoyant detector, see Figure 6.4 in Kay-II:



Example: DC level in WGN with A and σ^2 *both* unknown. Recall that σ^2 is called a *nuisance parameter*. Here, the GLR test accepts $\mathcal{H}_1$ if

$$\Lambda_{\text{GLR}}(\boldsymbol{x}) = \frac{\max_{\theta,\sigma^2} p(\boldsymbol{x}\,;\,\theta,\sigma^2)}{\max_{\sigma^2} p(\boldsymbol{x}\,;\,\theta=0,\sigma^2)} > \gamma$$

where

$$p(\boldsymbol{x}\,;\,a,\sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \cdot \exp\Big[-\frac{1}{2\,\sigma^2}\sum_{n=1}^{N}(x[n]-a)^2\Big]. \quad (2)$$

Here,

$$
\begin{aligned}
\max_{\theta,\sigma^2} p(\boldsymbol{x}\,;\,\theta,\sigma^2) &= \frac{1}{(2\pi\widehat{\sigma}_1^2)^{N/2}} \cdot e^{-N/2} \\
\max_{\sigma^2} p(\boldsymbol{x}\,;\,\theta=0,\sigma^2) &= \frac{1}{(2\pi\widehat{\sigma}_0^2)^{N/2}} \cdot e^{-N/2}
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\sigma}_0^2 &= \frac{1}{N}\sum_{n=1}^{N} x^2[n] \\
\widehat{\sigma}_1^2 &= \frac{1}{N}\sum_{n=1}^{N} (x[n]-\overline{x})^2.
\end{aligned}
$$

Hence,

$$
\Lambda_{\mathrm{GLR}}(\boldsymbol{x}) = \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_1^2}\right)^{N/2}
$$

i.e. the GLR test fits data with "best" DC-level signal $\widehat{a}_{\mathrm{ML}} = \overline{x}$, finds the residual variance estimate $\widehat{\sigma}_1^2$, and compares this estimate with the variance estimate $\widehat{\sigma}_0^2$ under the null case (i.e. for $\theta = 0$). When signal is present, $\widehat{\sigma}_1^2 \ll \widehat{\sigma}_0^2 \Rightarrow \Lambda_{\mathrm{GLR}}(\boldsymbol{x}) \gg 1$.

Note that

$$
\begin{aligned}
\widehat{\sigma}_1^2 &= \frac{1}{N}\sum_{n=1}^{N}(\overline{x} - x[n])^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}(x^2[n] - 2\,\overline{x}\,x[n] + \overline{x}^2) \\
&= \Big(\frac{1}{N}\sum_{n=1}^{N} x^2[n]\Big) - 2\,\overline{x}^2 + \overline{x}^2 \\
&= \widehat{\sigma}_0^2 - \overline{x}^2.
\end{aligned}
$$

Hence,

$$
2\log\Lambda_{\mathrm{GLR}}(\boldsymbol{x}) = N\log\Big(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_0^2 - \overline{x}^2}\Big) = N\log\Big(\frac{1}{1 - \overline{x}^2/\widehat{\sigma}_0^2}\Big).
$$

Note that

$$
0 \le \frac{\overline{x}^2}{\widehat{\sigma}_0^2} \le 1
$$

and $1/(1-x)$ is monotonically increasing on $x \in (0,1)$. Therefore, an equivalent test can be constructed as follows:

$$
T(\boldsymbol{x}) = \frac{\overline{x}^2}{\widehat{\sigma}_0^2} > \lambda'.
$$

Under $\mathcal{H}_0$, the pdf of $T(\boldsymbol{x})$ *does not* depend on σ^2 $\Rightarrow$ GLR test can be implemented, i.e. it is CFAR.

Definition. *A test is constant false alarm rate (CFAR) if we can find a threshold that yields a detector with constant (specified) false-alarm rate* P_{FA}.

In other words, we should be able to set the threshold independently of the unknown parameters $\Leftrightarrow$ the distribution of the test statistic under $\mathcal{H}_0$ *does not* depend on the unknown parameters.

Large-data Record Performance of GLR Tests, Section 6.5 in Kay-II

The asymptotic results presented here are valid if

(i) N is large and

(ii) the ML estimate of the parameter vector $\boldsymbol{\vartheta}$ attains the asymptotic pdf $\widehat{\boldsymbol{\vartheta}} \sim \mathcal{N}(\boldsymbol{\vartheta}, \mathcal{I}(\boldsymbol{\vartheta})^{-1})$, see handout # 3.

General result: Consider the parametric model $p(\boldsymbol{x}; \boldsymbol{\vartheta})$ with

$$\boldsymbol{\vartheta} = \left[\begin{array}{c} \boldsymbol{\theta} \\ \boldsymbol{\varphi} \end{array} \right] = \left[\begin{array}{c} r \times 1 \\ s \times 1 \end{array} \right].$$

Here, $\boldsymbol{\theta}$ is to be tested and $\boldsymbol{\varphi}$ is a nuisance parameter vector. We also assume that we wish to test $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (reduced model) versus $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ (full model), where the nuisance parameters $\boldsymbol{\varphi}$ are unknown, but *are the same* under both hypotheses:

$$\begin{aligned} \mathcal{H}_0 \quad &: \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0, \quad \boldsymbol{\varphi} \\ \mathcal{H}_1 \quad &: \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \quad \boldsymbol{\varphi}. \end{aligned}$$

Then, the GLR test is: Decide $\mathcal{H}_1$ if

$$\Lambda_{\text{GLR}}(\boldsymbol{x}) = \frac{\max_{\boldsymbol{\theta},\boldsymbol{\varphi}} p(\boldsymbol{x}\,;\,\boldsymbol{\theta},\boldsymbol{\varphi})}{\max_{\boldsymbol{\varphi}} p(\boldsymbol{x}\,;\,\boldsymbol{\theta} = \boldsymbol{\theta}_0,\boldsymbol{\varphi})} > \lambda.$$

Then, as $N \to \infty$,

$$\begin{aligned} 2 \underbrace{\log}_{\equiv \ln} \Lambda_{\mathrm{GLR}} &\sim \chi_r^2 \qquad \text{under } \mathcal{H}_0 \\ &\sim \underbrace{\chi_r^2(\lambda)}_{\text{noncentral } \chi^2 \text{ pdf}} \qquad \text{under } \mathcal{H}_1 \end{aligned}$$

where the expression for λ is given in eq. (6.24), Ch. 6.5 in Kay-II. Since, under $\mathcal{H}_0$, the asymptotic pdf of the test statistic $\Lambda_{\mathrm{GLR}}(\boldsymbol{x})$ does not depend on any unknown parameters, the threshold required to (approximately) maintain a constant P_{FA} can be found, which is the CFAR property. The approximate CFAR property of GLR tests holds only for large data records (i.e. large N).

No nuisance parameters: $\boldsymbol{\theta}$ is an $r \times 1$ vector and we test

$$\begin{aligned} \mathcal{H}_0 : \boldsymbol{\theta} &= \boldsymbol{\theta}_0 \\ \mathcal{H}_1 : \boldsymbol{\theta} &\neq \boldsymbol{\theta}_0. \end{aligned}$$

Then, the asymptotic pdfs of $\log \Lambda_{\mathrm{GLR}}(\boldsymbol{x})$ are the same as given above, with

$$\lambda = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathcal{I}(\boldsymbol{\theta}_0) (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

Wald and Rao Tests

Consider testing

$$\mathcal{H}_0\colon\ \boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0} \quad \text{versus} \quad \mathcal{H}_1\colon\ \boldsymbol{h}(\boldsymbol{\theta}) \neq \mathbf{0}$$

where $\boldsymbol{h}$ is an $r \times 1$ $[r \leq \dim(\boldsymbol{\theta})]$ once continuously-differentiable function. The **Wald test** for the above problem is

$$T_{\mathrm{W}}(\boldsymbol{x}) = \boldsymbol{h}(\widehat{\boldsymbol{\theta}})^T \Big[H(\widehat{\boldsymbol{\theta}}) \cdot \mathrm{CRB}(\widehat{\boldsymbol{\theta}}) \cdot H(\widehat{\boldsymbol{\theta}})^T \Big]^{-1} \boldsymbol{h}(\widehat{\boldsymbol{\theta}}) > \lambda,$$

where $H(\boldsymbol{\theta}) = \partial \boldsymbol{h}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T$ (having full rank r), $\mathrm{CRB}(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})^{-1}$, and $\widehat{\boldsymbol{\theta}}$ is an unrestricted ML estimator of $\boldsymbol{\theta}$ (under $\mathcal{H}_1$). Then

$$T_{\mathrm{W}}(\boldsymbol{x}) \quad \sim \quad \chi_r^2 \qquad \text{under } \mathcal{H}_0.$$

Rao test for the above problem:

$$T_{\mathrm{R}}(\boldsymbol{x}) = \boldsymbol{s}(\widetilde{\boldsymbol{\theta}})^T \mathrm{CRB}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{s}(\widetilde{\boldsymbol{\theta}})$$

where

$$\boldsymbol{s}(\boldsymbol{\theta}) = \frac{\partial \log p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

and $\widetilde{\boldsymbol{\theta}}$ is the restricted estimate of $\boldsymbol{\theta}$ (under $\mathcal{H}_0$).

$$T_{\mathrm{R}}(\boldsymbol{x}) \quad \sim \quad \chi_r^2 \qquad \text{under } \mathcal{H}_0.$$

For more on Rao test, see recent article by Bera and Bilias in Journal of Statistical Planning and Inference, vol. 97, Issue 1, 1 August 2001, pp. 9–44.

Important special case: $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\theta}_0$. Then the Wald test becomes: decide $\mathcal{H}_1$ if

$$T_{\mathrm{W}}(\boldsymbol{x}) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \, \mathcal{I}(\widehat{\boldsymbol{\theta}}) \, (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) > \lambda.$$

Important Theorems Used to Derive the Above Results

Mann-Wald Theorem (a part of it):
Assume that we have a sequence of $p \times 1$ vectors $\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(N)$ such that

$$\{\boldsymbol{y}(N)\} \xrightarrow{\mathrm{d}} \boldsymbol{y}, \quad \boldsymbol{y} \sim F.$$

Then, for a $p \times p$ matrix $\boldsymbol{B}$, the following holds:

$$\{\boldsymbol{y}(N)^T \, \boldsymbol{B} \, \boldsymbol{y}(N)\} \xrightarrow{\mathrm{d}} \boldsymbol{y}^T \boldsymbol{B} \boldsymbol{y}.$$

Apply Mann-Wald to

$$\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta} \xrightarrow{\mathrm{d}} \boldsymbol{y}, \quad \boldsymbol{y} \sim \mathcal{N}(0, \mathcal{I}(\boldsymbol{\theta})^{-1}).$$

using $\boldsymbol{B} = \mathcal{I}(\boldsymbol{\theta})$:

$$(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})^T \, \mathcal{I}(\boldsymbol{\theta}) \, (\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} \boldsymbol{y}^T \mathcal{I}(\boldsymbol{\theta}) \boldsymbol{y}. \tag{3}$$

Quadratic Forms:
Define a $p \times p$ matrix $\boldsymbol{B}$. If $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and for $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_p]^T$,

$$\boldsymbol{\eta}^T \boldsymbol{\Sigma} = 0 \;\Rightarrow\; \boldsymbol{\eta}^T \boldsymbol{\mu} = 0,$$

then $\boldsymbol{y}^T \boldsymbol{B} \boldsymbol{y} \sim \chi^2$ with degrees of freedom equal to $\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma})$ and non-centrality parameters $\boldsymbol{\mu}^T \boldsymbol{B} \boldsymbol{\mu}$ iff

$$\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}.$$

Applying the above theorem to $\boldsymbol{y}$ in (3) with $\boldsymbol{B} = \mathcal{I}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma} = \mathcal{I}(\boldsymbol{\theta})^{-1}$, we get

$$(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})^T \, \mathcal{I}(\boldsymbol{\theta}) \, (\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} \chi_p^2.$$

This result continues to hold if $\mathcal{I}(\boldsymbol{\theta})$ is replaced with a consistent estimator (Cramér-Wold device and Slutsky's theorem). Possible choices of consistent estimators of $\mathcal{I}(\boldsymbol{\theta})$:

1. $\mathcal{I}(\widehat{\boldsymbol{\theta}}_N)$,

2. $\mathcal{I}^{\mathrm{obs}}(\widehat{\boldsymbol{\theta}}_N)$, whose (i, k)th element is

$$-\Big[\sum_{n=1}^{N} \frac{\partial^2}{\partial \theta_i \partial \theta_k} \log p(\boldsymbol{x}(n); \boldsymbol{\theta}) \Big] \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_N}.$$

Rao Test Example

DC level in non-Gaussian noise.

$$\begin{aligned}
\mathcal{H}_0 : \quad & x[n] = w[n], \quad n = 1, 2, \ldots, N \\
\mathcal{H}_1 : \quad & x[n] = \underbrace{A}_{\text{unknown}} + w[n], \quad n = 1, 2, \ldots, N.
\end{aligned}$$

$w[1], w[2], \ldots, w[N]$ are i.i.d. with pdf

$$p(w) = c_1 \cdot \exp[-\tfrac{1}{2}\, c_2\, w^4], \quad -\infty < w < \infty,$$

where c_1 and c_2 are constants. To implement the GLR test

$$\Lambda_G(\boldsymbol{x}) = \frac{p(\boldsymbol{x}; \widehat{A}, \mathcal{H}_1)}{p(\boldsymbol{x}; \mathcal{H}_0)} > \lambda,$$

we need ML estimate $\widehat{A}$ of A. But

$$p(\boldsymbol{x}; A, \mathcal{H}_1) = c_1^N \exp[-\tfrac{1}{2} c_2 \sum_{n=1}^{N} (x[n] - A)^4],$$

so we need to minimize $\sum_{n=1}^{N} (x[n] - A)^4$ with respect to A. Hence, $\widehat{A}$ cannot be found in closed form.

Consider now the Rao test, which **can** be computed in closed form:

$$T_{\mathrm{R}}(\boldsymbol{x}) = [s(0)]^2 \cdot \mathrm{CRB}(0)$$

where

$$\begin{aligned} s(0) &= \left.\frac{\partial \log p(\boldsymbol{x}; A)}{\partial A}\right|_{A=0} \\ &= 2c_2 \sum_{n=1}^{N} (x[n] - A)^3 \Big|_{A=0} = 2c_2 \sum_{n=1}^{N} x^3[n]. \end{aligned}$$

Hence

$$\begin{aligned} T_{\mathrm{R}}(\boldsymbol{x}) &= 4c_2^2 \cdot \mathrm{CRB}(0) \cdot \Big(\sum_{n=1}^{N} x^3[n] \Big)^2 \\ &= 4N^2 c_2^2 \cdot \mathrm{CRB}(0) \cdot \Big(\frac{1}{N} \sum_{n=1}^{N} x^3[n] \Big)^2, \end{aligned}$$

see Example 6.9 in Kay-II for the exact expression. Note that, under $\mathcal{H}_0$, $\mathrm{E}\left(x^3[n]\right) = 0$, due to symmetry. Hence,

$$\sum_{n=1}^{N} x^3[n] \approx 0.$$

Under $\mathcal{H}_1$, there will be a signal contribution of A^6.

Classical Linear Model

Recall the classical linear model

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{w}$$

where $\boldsymbol{x}$ is a measured $N \times 1$ vector and $\boldsymbol{H}$ is a known *deterministic* $N \times p$ matrix, where $N \geq p$. Assume $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ and σ^2 known. $\boldsymbol{\theta}$ is deterministic *unknown* parameter vector, to be tested.

In general, we consider

$$\begin{aligned} \mathcal{H}_0 : &\quad \boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b} \quad \text{versus} \\ \mathcal{H}_1 : &\quad \boldsymbol{A}\boldsymbol{\theta} \neq \boldsymbol{b}. \end{aligned}$$

$\boldsymbol{A}$ is $r \times p$ ($r \leq q$) of rank r. Here, GLR test decides $\mathcal{H}_1$ if

$$T(\boldsymbol{x}) = \frac{(\boldsymbol{A}\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{b})^T [\boldsymbol{A}(\boldsymbol{H}^T \boldsymbol{H})^{-1} \boldsymbol{A}^T]^{-1} (\boldsymbol{A}\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{b})}{\sigma^2} > \tau$$

where $\widehat{\boldsymbol{\theta}}_1 = (\boldsymbol{H}^T \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{x}$ is the ML estimator of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ (no restrictions).

(Exact) detection performance is

$$\begin{aligned} P_{\text{FA}} &= Q_{\chi_r^2}(\tau) \\ P_{\text{D}} &= Q_{\chi_r^2(\lambda)}(\tau) \end{aligned}$$

where

$$\lambda = \frac{(\boldsymbol{A}\boldsymbol{\theta}_1 - \boldsymbol{b})^T [\boldsymbol{A}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\boldsymbol{\theta}_1 - \boldsymbol{b})}{\sigma^2}.$$

See Appendix 7B in Kay-II for proof.

Important special case: Test

$$\begin{aligned} \mathcal{H}_0 : &\qquad \boldsymbol{\theta} = \boldsymbol{0}, \\ \mathcal{H}_1 : &\qquad \boldsymbol{\theta} \neq \boldsymbol{0}. \end{aligned}$$

Apply the above result with $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{b} = \boldsymbol{0}$, $r = p$:

$$T(\boldsymbol{x}) = \frac{\widehat{\boldsymbol{\theta}}_1^T \boldsymbol{H}^T \boldsymbol{H} \widehat{\boldsymbol{\theta}}_1}{\sigma^2} = \frac{\boldsymbol{x}^T \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{x}}{\sigma^2} > \tau.$$

Consider GLR test for linear model where σ^2 is unknown, see Theorem 9.1 in Kay-II. GLR test decides $\mathcal{H}_1$ if

$$T(\boldsymbol{x}) = \frac{N-p}{r} \cdot \frac{(\boldsymbol{A}\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{b})^T[\boldsymbol{A}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{b})}{\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T)\boldsymbol{x}}$$

where $\widehat{\boldsymbol{\theta}}_1 = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{x}$ is the ML estimator of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ (no restrictions).

(Exact) detection performance is

$$\begin{aligned} P_{\mathrm{FA}} &= Q_{F_{r,N-p}}(\tau) \\ P_{\mathrm{D}} &= Q_{F_{r,N-p}(\lambda)}(\tau), \end{aligned}$$

where

$$\lambda = \frac{(\boldsymbol{A}\boldsymbol{\theta}_1 - \boldsymbol{b})^T[\boldsymbol{A}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\boldsymbol{\theta}_1 - \boldsymbol{b})}{\sigma^2}.$$

Important special case: Test

$$\begin{aligned} \mathcal{H}_0 : &\quad \boldsymbol{\theta} = \boldsymbol{0}, \\ \mathcal{H}_1 : &\quad \boldsymbol{\theta} \neq \boldsymbol{0}. \end{aligned}$$

Apply the above result with $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{b} = \boldsymbol{0}$, $r = p$:

$$T(\boldsymbol{x}) = \frac{N-p}{p} \cdot \frac{\boldsymbol{x}^T\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{x}}{\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T)\boldsymbol{x}}.$$