Stochastic Signals and Systems

Contents

4 Signal Detection

4.1 Introduction to Hypothesis Testing

Introductory Example

$$X_t = \eta s_t + Z_t \quad (t = 1,\dots,n) \text{ or } \mathbf{X} = \eta \mathbf{s} + \mathbf{Z}$$

with $\mathbf{X} = (X_1,\dots,X_n)^T$, $\mathbf{s} = (s_1,\dots,s_n)^T$ and $\mathbf{Z} = (Z_1,\dots,Z_n)^T$,

where $X_1,\dots,X_n$ are assumed to be stochastically independent. Furthermore

s_t : denotes the known wave form

$\eta \geq 0$: indicates the unknown amplitude

Z_t : represents white noise with $Z_t \sim \mathcal{N}(0, \sigma_Z^2)$

Let $x_1,\ldots,x_n$ denote the observations of $X_1,\ldots,X_n$. Then we are faced with the two following Problems.

Problem 1:

Parameter Estimation, i.e. we have to estimate η, e.g. by means of the least squares approach, cf. Chapter 3.

$$\hat{\eta} = (\mathbf{s}^T\mathbf{s})^{-1}\mathbf{s}^T\mathbf{x} = \frac{\mathbf{s}^T\mathbf{x}}{\mathbf{s}^T\mathbf{s}}$$

Problem 2:

Signal Detection, i.e. we have to decide whether

$$\mathbf{X} = \eta\,\mathbf{s} + \mathbf{Z} \;\text{ or }\; \mathbf{X} = \mathbf{Z}, \;\text{ i.e. }\; \eta > 0 \;\text{ or }\; \eta = 0.$$

Procedure to construct a hypotheses test

1) Setting up of a hypothesis H_0
   X does not contain the signal waveform **s**, i.e.
   $\mathbf{X} = \mathbf{Z}$ or $\eta = 0$ and $\mathbf{X} \sim \mathcal{N}_n(0, \sigma_Z^2 \mathbf{I})$.

2) Setting up of an alternative H_1
   X contains the signal waveform **s**, i.e.
   $\mathbf{X} = \eta \mathbf{s} + \mathbf{Z}$ and $\mathbf{X} \sim \mathcal{N}_n(\eta \mathbf{s}, \sigma_Z^2 \mathbf{I})$,
   where $\eta > 0$ is an unknown parameter.

3) Selection of a favorable statistic $t(\mathbf{x})$ of the observation **x** for testing the hypothesis H_0.

4) Determination of the distribution of $T = t(\mathbf{X})$ under H_0.
   $T = t(\mathbf{X}) = \mathbf{s}^T \mathbf{X}$ with $\mathbf{X} \sim \mathcal{N}_n(\mathbf{0}, \sigma_Z^2 \mathbf{I}) \Rightarrow T \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{s}^T \mathbf{s})$

since $\mathrm{E}(T)=\mathbf{s}^T\mathrm{E}\mathbf{X}=0$, $\mathrm{Var}(T)=\mathrm{E}(\mathbf{s}^T\mathbf{X}\mathbf{X}^T\mathbf{s})=\sigma_Z^2\mathbf{s}^T\mathbf{s}=\sigma_T^2$.

5) Calculation of the critical region of the observations $\mathbf{x}$ for discarding hypothesis H_0.

For a given size α we can derive via

$$\alpha = P\left(T > \kappa \mid H_0\right) = P\left(T/\sigma_T > \kappa/\sigma_T \mid H_0\right)$$

$$= 1 - P\left(T/\sigma_T \le \kappa/\sigma_T \mid H_0\right) = 1 - \Phi(\kappa/\sigma_T)$$

the critical region $R_C = \left\{\mathbf{x}\colon\ t(\mathbf{x}) \in I = (\kappa, \infty)\right\}$.

6) If $t(\mathbf{x}) > \kappa$ one decides for H_1, i.e. $\mathbf{x}$ contains the wave-form $\mathbf{s}$, with the probability of error α. If $t(\mathbf{x}) \le \kappa$ one de-cides for H_0, i.e. $\mathbf{x}$ does not contain the waveform $\mathbf{s}$.

Classification of hypotheses tests

The density $f_{\mathbf{X}}(\mathbf{x})$ of $\mathbf{X} = (X_1,\ldots,X_n)^T$ is element of the known set $\{f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta}): \boldsymbol{\theta} \in \Omega\}$, where $\boldsymbol{\theta}$ denotes an unknown parameter vector and $\boldsymbol{\Omega}$ the parameter space.

Binary Hypotheses Tests

Let hypothesis H_0 (called null hypothesis) and the alternative hypothesis H_1 (also called one hypothesis) divide the parameter space $\boldsymbol{\Omega}$ into the disjoint subset $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$, respectively. Then the test

$$H_0: \quad \boldsymbol{\theta} \in \Omega_0 \quad \text{versus} \quad H_1: \quad \boldsymbol{\theta} \in \Omega_1$$

is said to be a binary test of hypotheses and one is testing which of the two subsets contains the unknown $\boldsymbol{\theta}$.

Simple and Composite Hypotheses

If Ω_m ($m = 0,1$) contains only a single element $\boldsymbol{\theta}_m$ the corresponding hypothesis H_m is said to be simple. Otherwise it is composite.

Multiple or *M*-ary Hypotheses Tests

Let $\Omega = \Omega_0 \cup \Omega_1 \cup \ldots \cup \Omega_{M-1}$ represents a disjoint covering of the parameter space and let H_m denote the hypothesis that $\boldsymbol{\theta} \in \Omega_m$. The test

$$H_0 : \boldsymbol{\theta} \in \Omega_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \in \Omega_1 \quad \text{versus} \quad \ldots$$

$$\ldots \quad \text{versus} \quad H_{M-1} : \boldsymbol{\theta} \in \Omega_{M-1}$$

is called multiple, or *M*-ary, hypothesis test.

Exercise 4.1-1:
Introductory example

Exercise 4.1-2:
M-ary communication

Testing of binary hypotheses

A test of H_0 versus H_1 is described by a test function ϕ (function of the observation $\mathbf{x}$) that satisfies $0 \leq \phi(\mathbf{x}) \leq 1$.

The value $\phi(\mathbf{x})$ obtained for a given observation $\mathbf{x}$ means that with $P = 1 - \phi(\mathbf{x})$ the hypothesis H_0 and with $P = \phi(\mathbf{x})$ the hypothesis H_1 should be selected.

If $\phi(\mathbf{x})$ can take only the values zero and one, i.e.

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in R_C \\ 0 & \text{elsewhere} \end{cases},$$

the test is termed deterministic, where R_C denotes the so-called critical region.

It is desirable to construct a test function ϕ such that the probability to deciding for

- H_1 even though H_0 is correct (probability of type 1 error)
- H_0 even though H_1 is correct (probability of type 2 error)

is minimized.

Unfortunately both probabilities can not be controlled simultaneously. Therefore one assigns a bound to the probability of the type 1 error by imposing the constraint

$$\alpha \geq \mathrm{E}\big(\phi(\mathbf{X})\big) = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x} \,|\, \boldsymbol{\theta}) d\mathbf{x} \quad \forall \boldsymbol{\theta} \in \Omega_0$$

which simplifies in case of a deterministic test to

$$\alpha \geq \mathrm{E}\big(1_{R_C}(\mathbf{X})\big) = \int_{R_c} f_{\mathbf{X}}(\mathbf{x} \,|\, \boldsymbol{\theta}) d\mathbf{x} \quad \forall \boldsymbol{\theta} \in \Omega_0 .$$

Subject to this constraint it is then desired to minimize the probability of the type 2 error or equivalently to maximize the so-called power function

$$\beta_\phi(\boldsymbol{\theta}) = \mathrm{E}\left(\phi(\mathbf{X})\right) = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})\, d\mathbf{x} \quad \forall \boldsymbol{\theta} \in \Omega_1$$

which in case of a deterministic test can be expressed by

$$\beta_\phi(\boldsymbol{\theta}) = \mathrm{E}\left(1_{R_C}(\mathbf{X})\right) = \int_{R_C} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})\, d\mathbf{x} \quad \forall \boldsymbol{\theta} \in \Omega_1.$$

The level of significance α with $0 < \alpha < 1$ is called the size of the test. The probability $\beta_\phi(\boldsymbol{\theta})$ of correctly accepting H_1 is said to be the power of the test.

$\phi(\mathbf{x})$ is said to be a uniformly most powerful (UMP) test of the test problem (α, H_0, H_1) if the inequality

$$\beta_\phi(\boldsymbol{\theta}) \geq \beta_{\tilde{\phi}}(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Omega_1$$

holds for any test $\tilde{\phi}(\mathbf{x})$ of (α, H_0, H_1).

Cross reference of statistical terms

Statisticians	**Engineers**
Observations $\mathbf{x}$	Receiver output data
Null hypothesis H_0	Noise only hypothesis
Alternative hypothesis H_1	Signal + Noise hypothesis
Test function $\phi(\mathbf{x})$	Detector
Size of the test α Probability of type 1 error	Probability of false alarm (P_{FA})
Power of the test $\beta_\phi(\boldsymbol{\theta}) \ \forall \boldsymbol{\theta} \in \Omega_1$	Probability of detection (P_D)
Probability of type 2 error	Probability of miss ($P_M = 1 - P_D$)

4.2 Simple Hypothesis Testing

Let H_0 and H_1 be simple hypotheses, i.e.

$$\Omega_0 = \{\boldsymbol{\theta}_0\} \text{ and } \Omega_1 = \{\boldsymbol{\theta}_1\}.$$

In this case one only has to distinguish whether

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}_0) = f_{\mathbf{X}}(\mathbf{x}|0) \text{ or } f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}_1) = f_{\mathbf{X}}(\mathbf{x}|1).$$

4.2.1 Neyman-Pearson Hypothesis Testing

For binary hypotheses tests with simple hypotheses the following theorem, known as the fundamental lemma of Neyman-Pearson, holds.

Theorem:

Let H_0 and H_1 be simple hypotheses with corresponding densities $f_{\mathbf{X}}(\mathbf{x}|0)$ and $f_{\mathbf{X}}(\mathbf{x}|1)$.

1) For testing H_0 against H_1 there exists a test ϕ and a constant κ such that

   a) $\beta_\phi(0) = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x} \mid 0) d\mathbf{x} = \alpha$,

   b) $$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{\mathbf{X}}(\mathbf{x} \mid 1) > \kappa f_{\mathbf{X}}(\mathbf{x} \mid 0) \\ 0 & \text{if } f_{\mathbf{X}}(\mathbf{x} \mid 1) < \kappa f_{\mathbf{X}}(\mathbf{x} \mid 0) \end{cases}.$$

2) If ϕ satisfies 1a) and 1b) for some κ then ϕ is most powerful for (α, H_0, H_1).

3) If ϕ is most powerful for (α, H_0, H_1) then for some κ it satisfies 1b) almost surely and it satisfies also 1a) unless there exists a test of size $< \alpha$ and with power 1.

Proof of 1):

Let $g(k) = P\{f_{\mathbf{X}}(\mathbf{X}|1) > k\, f_{\mathbf{X}}(\mathbf{X}|0) \,|\, H_0\}$

$$= P\left\{T = \frac{f_{\mathbf{X}}(\mathbf{X}|1)}{f_{\mathbf{X}}(\mathbf{X}|0)} > k \,|\, f_{\mathbf{X}}(\mathbf{X}|0) > 0, H_0\right\} = 1 - F_T(k).$$

Thus $g(k)$ is non-increasing, right continuous, $g(-\infty) = 1$ and $g(\infty) = 0$. Given any α $(0 < \alpha < 1)$ and let $k = \kappa$ be such that $g(\kappa) \leq \alpha \leq g(\kappa - 0)$ the test ϕ can be defined by

$$\phi(\mathbf{x})=\begin{cases}1 & \text{if } f_{\mathbf{X}}(\mathbf{x}|1)>\kappa f_{\mathbf{X}}(\mathbf{x}|0)\\ 0 & \text{if } f_{\mathbf{X}}(\mathbf{x}|1)<\kappa f_{\mathbf{X}}(\mathbf{x}|0)\\ \dfrac{\alpha-g(\kappa)}{g(\kappa-0)-g(\kappa)} & \text{if } f_{\mathbf{X}}(\mathbf{x}|1)=\kappa f_{\mathbf{X}}(\mathbf{x}|0),\ g(\kappa-0)-g(\kappa)>0\\ \text{arbitrary} & \text{if } f_{\mathbf{X}}(\mathbf{x}|1)=\kappa f_{\mathbf{X}}(\mathbf{x}|0),\ g(\kappa-0)-g(\kappa)=0\end{cases}$$

with

$$\begin{aligned}\beta_\phi(0) &= \int_{\mathbb{R}^n}\phi(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}\,|\,0)\,d\mathbf{x}\\ &= 1\cdot P\{f_{\mathbf{X}}(\mathbf{x}|1)>\kappa f_{\mathbf{X}}(\mathbf{x}|0)|H_0\}+0\cdot P\{f_{\mathbf{X}}(\mathbf{x}|1)<\kappa f_{\mathbf{X}}(\mathbf{x}|0)|H_0\}+\\ &\quad+\frac{\alpha-g(\kappa)}{g(\kappa-0)-g(\kappa)}P\{f_{\mathbf{X}}(\mathbf{x}|1)=\kappa f_{\mathbf{X}}(\mathbf{x}|0)|H_0\}\\ &= g(\kappa)+\frac{\alpha-g(\kappa)}{g(\kappa-0)-g(\kappa)}\big(g(\kappa-0)-g(\kappa)\big)=\alpha.\end{aligned}$$

Proof of 2):

Suppose $\phi(\mathbf{x})$ satisfies 1a) and 1b) and $\tilde{\phi}(\mathbf{x})$ denotes any test of (α, H_0, H_1) we can argue as follows.

$$\beta_{\tilde{\phi}}(0) = \int_{\mathbb{R}^n} \tilde{\phi}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x} \,|\, 0)\, d\mathbf{x} \le \alpha = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x} \,|\, 0)\, d\mathbf{x}$$

With the sets

$$S_> = \left\{\mathbf{x} : \ \phi(\mathbf{x}) > \tilde{\phi}(\mathbf{x})\right\} \text{ and } S_< = \left\{\mathbf{x} : \ \phi(\mathbf{x}) < \tilde{\phi}(\mathbf{x})\right\}$$

we can state

$$\text{if } \mathbf{x} \in S_> \text{ then } \phi(\mathbf{x}) > 0, \text{ i.e. } f_{\mathbf{x}}(\mathbf{x} \,|\, 1) \ge \kappa f_{\mathbf{x}}(\mathbf{x} \,|\, 0)$$

$$\text{if } \mathbf{x} \in S_< \text{ then } \phi(\mathbf{x}) < 1, \text{ i.e. } f_{\mathbf{x}}(\mathbf{x} \,|\, 1) \le \kappa f_{\mathbf{x}}(\mathbf{x} \,|\, 0).$$

Hence,

$$\int_{\mathbb{R}^n} \left(\phi(\mathbf{x}) - \tilde{\phi}(\mathbf{x})\right)\left(f_{\mathbf{X}}(\mathbf{x} \mid 1) - \kappa f_{\mathbf{X}}(\mathbf{x} \mid 0)\right) d\mathbf{x} =$$

$$= \int_{S_> \cup S_<} \left(\phi(\mathbf{x}) - \tilde{\phi}(\mathbf{x})\right)\left(f_{\mathbf{X}}(\mathbf{x} \mid 1) - \kappa f_{\mathbf{X}}(\mathbf{x} \mid 0)\right) d\mathbf{x} \geq 0$$

and the difference in power between $\phi(\mathbf{x})$ and any $\tilde{\phi}(\mathbf{x})$ of (α, H_0, H_1) therefore satisfies

$$\int_{\mathbb{R}^n} \left(\phi(\mathbf{x}) - \tilde{\phi}(\mathbf{x})\right) f_{\mathbf{X}}(\mathbf{x}|1)\, d\mathbf{x} = \beta_\phi(1) - \beta_{\tilde{\phi}}(1) \geq$$

$$\geq \kappa \int_{\mathbb{R}^n} \left(\phi(\mathbf{x}) - \tilde{\phi}(\mathbf{x})\right) f_{\mathbf{X}}(\mathbf{x}|0)\, d\mathbf{x} = \beta_\phi(0) - \beta_{\tilde{\phi}}(0) \geq 0$$

as was to be proved.

Thus, the fundamental lemma of Neyman-Pearson provides an approach for constructing a most powerful test for (α, H_0, H_1).

Examples to the Neyman-Pearson Approach

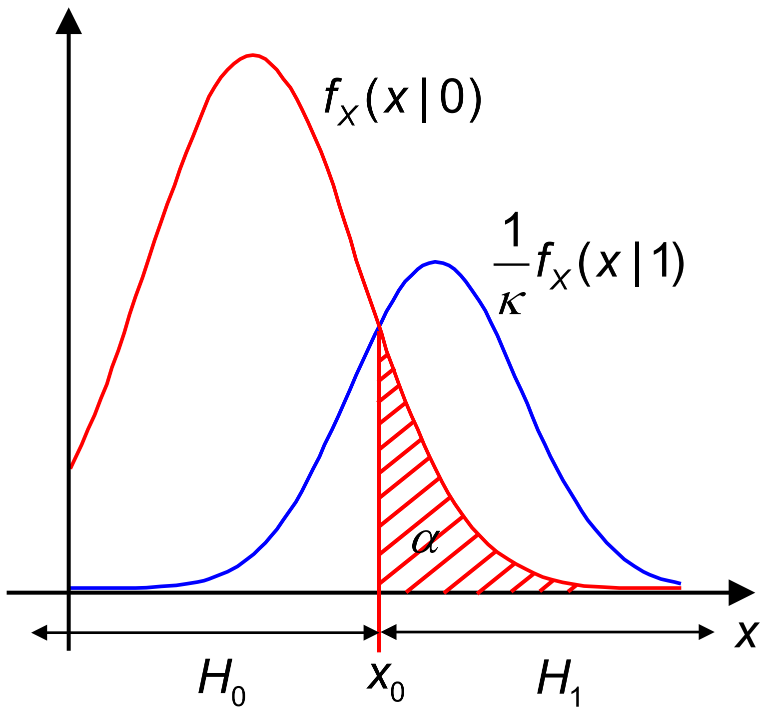
In the following we assume $n=1$, i.e. $\mathbf{x}=x$.

1) For $P\{f_X(x|1)=\kappa f_X(x|0)\,|\,H_0\}=0$ the test function is given by

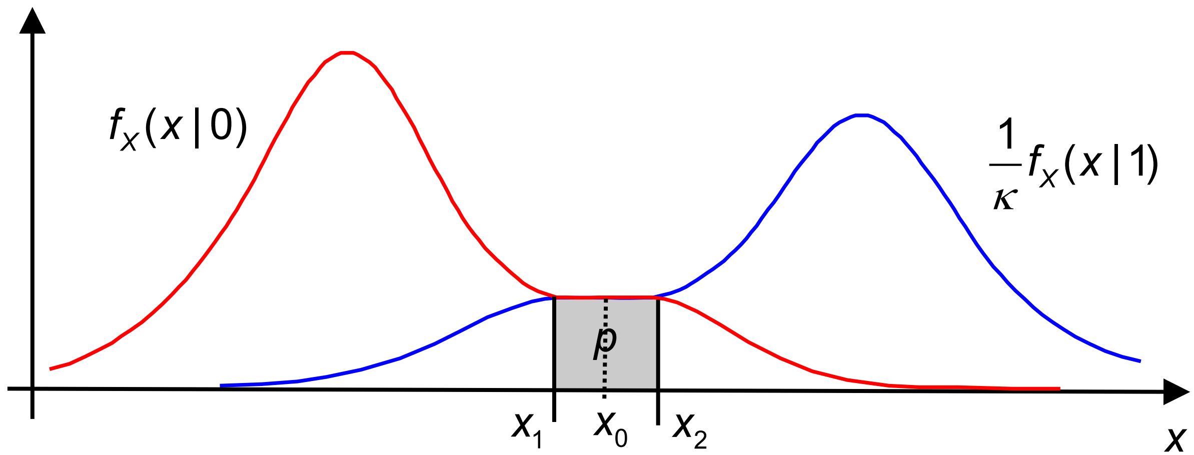
$$\phi(x)=\begin{cases}1 & \text{if } \; x>x_0\\ 0 & \text{if } \; x<x_0\\ \text{e.g. } 0 & \text{if } \; x=x_0\end{cases}$$

with

$$\alpha=\int_{x_0}^{\infty} f_X(x\,|\,0)\,dx.$$



2) For $P\left\{ f_X(x|1) = \kappa f_X(x|0) \,|\, H_0 \right\} = p > 0$, cf. figure below,



the test function can be defined by

$$\phi(x) = \begin{cases} 1 & \text{if } \; x > x_2 \\ \dfrac{\alpha - g(\kappa)}{g(\kappa - 0) - g(\kappa)} & \text{if } \; x_1 \le x \le x_2 \\ 0 & \text{if } \; x < x_1 \end{cases}$$

with

$$g(\kappa) = P\{f_X(x|1) > \kappa f_X(x|0) \,|\, H_0\} = \int_{x_2}^{\infty} f_X(x\,|\,0)\,dx$$

and

$$p = \int_{x_1}^{x_2} f_X(x\,|\,0)\,dx.$$

Moreover, for this case another most powerful test $\tilde{\phi}(\mathbf{x})$ can be constructed by

$$\tilde{\phi}(\mathbf{x}) = \begin{cases} 1 & \text{if } x > x_0 \\ 0 & \text{if } x \le x_0 \end{cases} \quad \text{with} \quad \begin{aligned} \alpha &= \int_{x_0}^{\infty} f_X(x|0)\,dx \\ &= \int_{x_0}^{x_2} f_X(x|0)\,dx + \int_{x_2}^{\infty} f_X(x|0)\,dx. \end{aligned}$$

For $f_X(x|0) = \text{const.}$ in (x_1, x_2) the boundary x_0 is given by

$$x_0 = x_2 - \frac{x_2 - x_1}{p} \cdot \left(\alpha - \int_{x_2}^{\infty} f_X(x|0)\,dx \right).$$

Exercise 4.2-1:
Simple Hypothesis testing between two normal distributions with unequal mean and variance

General description of the signal detection problem

H_0 : $\mathbf{X}=\mathbf{U}$, i.e. noise only, $f_{\mathbf{X}}(\mathbf{x})=f_{\mathbf{U}}(\mathbf{x})=f_{\mathbf{X}}(\mathbf{x}|0)$

H_1 : $\mathbf{X}=\mathbf{S}+\mathbf{U}$, i.e. signal + noise, $f_{\mathbf{X}}(\mathbf{x})=f_{\mathbf{S}+\mathbf{U}}(\mathbf{x})=f_{\mathbf{X}}(\mathbf{x}|1)$

Now, we are looking for a detector $\phi(\mathbf{x})$ of the problem (α,H_0,H_1), whereby the level for the probability of false alarm $P_{FA}=\beta_\phi(0)=\alpha$ is given.

For the detector ϕ the probability of detection $P_D=\beta_\phi(1)$ is determined, where P_D is a function of $\alpha=P_{FA}$.

The probability of detection P_D as a function of the probability of false alarm P_{FA} is termed receiver operating characteristic (ROC).

Properties of $P_D(P_{FA})$

If $\phi(\mathbf{x})$ is a most powerful test for (α, H_0, H_1) we can state:

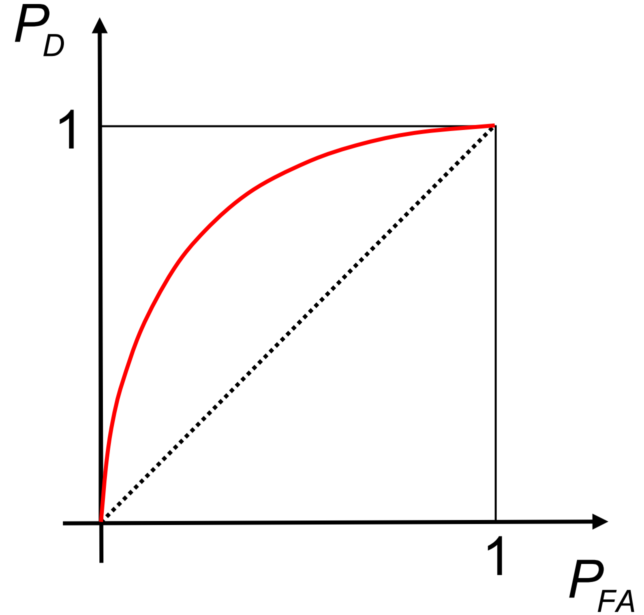
1) $P_D(P_{FA})$ is non-decreasing and concave in $0 < P_{FA} < 1$,

2) $P_D(P_{FA})$ is continuous in $0 < P_{FA} < 1$,

3) $P_D(P_{FA}) / P_{FA}$ is non-increasing in $0 < P_{FA} < 1$,

4) $\lim_{P_{FA} \to 1} P_D(P_{FA}) / P_{FA} = 1$,

5) $0 < P_D(P_{FA}) / P_{FA} < 1 / P_{FA}$.



Exercise 4.2-2:
Particular cases and most powerful tests

4.2.2 Most Powerful Tests for Normal Variates

According to the Neyman-Pearson Lemma a most powerful test asks whether

$$f_{\mathbf{X}}(\mathbf{x}\,|\,1) \gtrless \kappa f_{\mathbf{X}}(\mathbf{x}\,|\,0).$$

Since density functions of normally distributed random vectors $\mathbf{X}$ satisfy $f_{\mathbf{X}}(\mathbf{x}) > 0$ one can also ask whether

$$\Lambda_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}\,|\,1) / f_{\mathbf{X}}(\mathbf{x}\,|\,0) \gtrless \kappa$$

or

$$\ln\big(\Lambda_{\mathbf{X}}(\mathbf{x})\big) = \ln\big(f_{\mathbf{X}}(\mathbf{x}\,|\,1)\big) - \ln\big(f_{\mathbf{X}}(\mathbf{x}\,|\,0)\big) \gtrless \ln(\kappa).$$

The function $\Lambda_{\mathbf{X}}(\mathbf{x})$ and $\ln(\Lambda_{\mathbf{X}}(\mathbf{x}))$ are termed likelihood ratio and log-likelihood ratio, respectively.

Let $f_{\mathbf{X}}(\mathbf{x}|0)$ be the density of $\mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{X},0},\boldsymbol{\Sigma}_{\mathbf{XX},0})$ under H_0 and $f_{\mathbf{X}}(\mathbf{x}|1)$ be the density of $\mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{X},1},\boldsymbol{\Sigma}_{\mathbf{XX},1})$ under H_1 with $\boldsymbol{\mu}_{\mathbf{X},0} \neq \boldsymbol{\mu}_{\mathbf{X},1}$ and/or $\boldsymbol{\Sigma}_{\mathbf{XX},0} \neq \boldsymbol{\Sigma}_{\mathbf{XX},1}$. Hence, the log-likelihood ratio

$$\ln\left(\Lambda_{\mathbf{X}}(\mathbf{x})\right) = \frac{1}{2}\Big\{\ln(\det \boldsymbol{\Sigma}_{\mathbf{XX},0}) - \ln(\det \boldsymbol{\Sigma}_{\mathbf{XX},1}) + \boldsymbol{\mu}_{\mathbf{X},0}^T \boldsymbol{\Sigma}_{\mathbf{XX},0}^{-1} \boldsymbol{\mu}_{\mathbf{X},0} - \boldsymbol{\mu}_{\mathbf{X},1}^T \boldsymbol{\Sigma}_{\mathbf{XX},1}^{-1} \boldsymbol{\mu}_{\mathbf{X},1} + \mathbf{x}^T\left(\boldsymbol{\Sigma}_{\mathbf{XX},0}^{-1} - \boldsymbol{\Sigma}_{\mathbf{XX},1}^{-1}\right)\mathbf{x}\Big\} + \left(\boldsymbol{\mu}_{\mathbf{X},1}^T \boldsymbol{\Sigma}_{\mathbf{XX},1}^{-1} - \boldsymbol{\mu}_{\mathbf{X},0}^T \boldsymbol{\Sigma}_{\mathbf{XX},0}^{-1}\right)\mathbf{x}.$$

provides a most powerful test

$$\phi_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \ln\left(\Lambda_{\mathbf{X}}(\mathbf{x})\right) > \ln(\kappa) \text{ resp. } t_{\mathbf{X}}(\mathbf{x}) > \tilde{\kappa} \\ 0 & \text{elsewhere} \end{cases}$$

for (α, H_0, H_1), where

$$t_{\mathbf{X}}(\mathbf{x}) = \left(\boldsymbol{\mu}_{\mathbf{X},1}^T \boldsymbol{\Sigma}_{\mathbf{XX},1}^{-1} - \boldsymbol{\mu}_{\mathbf{X},0}^T \boldsymbol{\Sigma}_{\mathbf{XX},0}^{-1}\right)\mathbf{x} + \frac{1}{2}\mathbf{x}^T\left(\boldsymbol{\Sigma}_{\mathbf{XX},0}^{-1} - \boldsymbol{\Sigma}_{\mathbf{XX},1}^{-1}\right)\mathbf{x}$$

and

$$\tilde{\kappa} = \ln(\kappa) - \frac{1}{2}\Big(\ln(\det \boldsymbol{\Sigma}_{\mathbf{XX},0}) - \ln(\det \boldsymbol{\Sigma}_{\mathbf{XX},1}) + \\ + \boldsymbol{\mu}_{\mathbf{X},0}^T \boldsymbol{\Sigma}_{\mathbf{XX},0}^{-1} \boldsymbol{\mu}_{\mathbf{X},0} - \boldsymbol{\mu}_{\mathbf{X},1}^T \boldsymbol{\Sigma}_{\mathbf{XX},1}^{-1} \boldsymbol{\mu}_{\mathbf{X},1}\Big).$$

Detection of a known deterministic signal in Gaussian noise with known distribution

$$H_0: \ \mathbf{X} = \mathbf{U}, \ f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{U}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|0), \ \mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{U}}, \boldsymbol{\Sigma}_{\mathbf{UU}})$$

$$H_1: \ \mathbf{X} = \mathbf{U} + \mathbf{s}, \ f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{U}+\mathbf{s}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|1), \ \mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{U}} + \mathbf{s}, \boldsymbol{\Sigma}_{\mathbf{UU}})$$

Thus, the log-likelihood ratio can be expressed by

$$\ln\left(\Lambda_{\mathbf{X}}(\mathbf{x})\right) = \frac{1}{2}\left(\boldsymbol{\mu}_{\mathbf{U}}^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} \boldsymbol{\mu}_{\mathbf{U}} - (\boldsymbol{\mu}_{\mathbf{U}} + \mathbf{s})^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} (\boldsymbol{\mu}_{\mathbf{U}} + \mathbf{s})\right) + \mathbf{s}^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} \mathbf{x}$$

$$= -\frac{1}{2}\mathbf{s}^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} \mathbf{s} - \mathbf{s}^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} \boldsymbol{\Sigma}_{\mathbf{U}} + t_{\mathbf{X}}(\mathbf{x})$$

with

$$t_{\mathbf{X}}(\mathbf{x}) = \mathbf{s}^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} \mathbf{x}.$$

For the particular case that **U** is a sequence of white noise Z_i with $\mathrm{Var}(Z_i) = 1$, i.e. $\mathbf{U} = \mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, we obtain

$$t_{\mathbf{X}}(\mathbf{x}) = \mathbf{s}^T \mathbf{x} \gtrless \tilde{\kappa} = \ln(\kappa) + \frac{1}{2}\mathbf{s}^T \mathbf{s}.$$

The critical region $R_C = \left\{\mathbf{x} : t_{\mathbf{X}}(\mathbf{x}) > \tilde{\kappa}\right\}$ is then the set of all points that are lying above the plane $t_{\mathbf{X}}(\mathbf{x}) = \mathbf{s}^T \mathbf{x} = \tilde{\kappa}$.

Prewhitening Interpretation

An invertible linear transformation of normally distributed data does not change the properties of the test problem.

As an particular example we consider

$$\mathbf{Z} = \mathbf{C}^{-1}(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})$$

with

$$\boldsymbol{\Sigma}_{\mathbf{UU}} = \mathbf{C}\mathbf{C}^T \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} = (\mathbf{C}^{-1})^T \mathbf{C}^{-1},$$

where $\mathbf{Z}$ is white noise with $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0},\mathbf{I})$ and the inversion of the transformation is given by

$$\mathbf{U} = \mathbf{C}\mathbf{Z} + \boldsymbol{\mu}_{\mathbf{U}}.$$

Hence, due to the transformation

$$\mathbf{Y} = \mathbf{C}^{-1}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{U}})$$

the test problem can be reformulated as follows.

$$H_0: \ \mathbf{Y} = \mathbf{C}^{-1}(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}}) = \mathbf{Z}, \ \mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$$

$$H_1: \ \mathbf{Y} = \mathbf{C}^{-1}(\mathbf{U} + \mathbf{s} - \boldsymbol{\mu}_{\mathbf{U}}) = \mathbf{Z} + \mathbf{C}^{-1}\mathbf{s} = \mathbf{Z} + \tilde{\mathbf{s}}, \ \mathbf{Y} \sim \mathcal{N}_n(\tilde{\mathbf{s}}, \mathbf{I})$$

Since the noise model $\mathbf{U} \sim \mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{U}}, \boldsymbol{\Sigma}_{\mathbf{UU}})$ is mapped into the white noise model $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$ the transformation is called prewhitening.

The log-likelihood ratio is given by

$$\ln\left(\Lambda_{\mathbf{Y}}(\mathbf{y})\right) = -\frac{1}{2}\tilde{\mathbf{s}}^T\tilde{\mathbf{s}} + \tilde{\mathbf{s}}^T\mathbf{y} = -\frac{1}{2}\tilde{\mathbf{s}}^T\tilde{\mathbf{s}} + t_{\mathbf{Y}}(\mathbf{y}) = \ln\left(\Lambda_{\mathbf{X}}(\mathbf{x})\right),$$

whereby the transformation does not alter the value of the log-likelihood ratio.

If $\mathbf{C}^{-1}$ is selected as a lower triangle matrix (Cholesky decomposition of $\mathbf{\Sigma}_{\mathbf{UU}}$) the transformation can be formulated as a causal, digital, time-variant filtering procedure.

$$Y_t = \sum_{\tau=1}^{t} (X_\tau - \mu_{U_\tau}) h_{t,\tau}, \quad \text{where} \quad \mathbf{C}^{-1} = \left(h_{t,\tau}\right) \text{ with } h_{t,\tau} = 0 \text{ for } t < \tau$$

Matched Filtering Interpretation

Suppose **Y** represents a sequence of a time discrete stochastic process Y_t $(t = 1, \ldots, n)$. Then the time-invariant filter with impulse response

$$\tilde{h}_\tau = \begin{cases} \tilde{s}_{n-\tau} & \tau = 1, \ldots, n \\ 0 & \text{elsewhere} \end{cases}$$

and filter output

$$v_t = \sum_{\tau=1}^{n} \tilde{h}_\tau y_{t-\tau} \;\Rightarrow\; t_{\mathbf{Y}}(\mathbf{y}) = \tilde{\mathbf{s}}^T \mathbf{y} = v_n$$

is called matched filter for $\tilde{\mathbf{s}} = (\tilde{s}_1, \ldots, \tilde{s}_n)^T$ in white noise $\mathbf{Z}$.
The filter with impulse response

$$h_\tau = \begin{cases} g_{n-\tau} & \tau = 1, \ldots, n \\ 0 & \text{elsewhere} \end{cases}, \quad \text{where} \quad \mathbf{g}^T = \mathbf{s}^T \mathbf{\Sigma}_{\mathbf{UU}}^{-1}$$

and filter output

$$w_t = \sum_{\tau=1}^{n} h_\tau x_{t-\tau} \;\Rightarrow\; t_{\mathbf{X}}(\mathbf{x}) = \mathbf{g}^T \mathbf{x} = w_n$$

is called matched filter for $\mathbf{s} = (s_1, \ldots, s_n)^T$ in colored noise $\mathbf{U}$ with covariance matrix

$$\mathbf{\Sigma}_{\mathbf{UU}} = \text{Toeplitz}\big(c_{UU}(0), \ldots, c_{UU}(n-1)\big).$$

Calculation of $\tilde{\kappa} = \tilde{\kappa}(P_{FA})$ *and* $P_D = P_D(P_{FA})$

The test function

$$\phi_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} 1 & \text{if } t_{\mathbf{Y}}(\mathbf{y}) = \tilde{\mathbf{s}}^T \mathbf{y} > \ln(\kappa) + \tilde{\mathbf{s}}^T \tilde{\mathbf{s}}/2 = \tilde{\kappa} \\ 0 & \text{elsewhere} \end{cases}$$

is most powerful for (α, H_0, H_1), where $\tilde{\kappa}$ has to be determined such that the P_{FA} equals the predefined α.

Now, employing the statistic $T_{\mathbf{Y}} = t_{\mathbf{Y}}(\mathbf{Y}) = \tilde{\mathbf{s}}^T \mathbf{Y}$ we can write

$$H_0: \; T_{\mathbf{Y}} \sim \mathcal{N}(0, \tilde{\mathbf{s}}^T \tilde{\mathbf{s}}) \quad \text{and} \quad H_1: \; T_{\mathbf{Y}} \sim \mathcal{N}(\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}, \tilde{\mathbf{s}}^T \tilde{\mathbf{s}})$$

such that the probability of false alarm is given by

$$P_{FA} = \alpha = P(T_{\mathbf{Y}} > \tilde{\kappa} \mid H_0) = 1 - P(T_{\mathbf{Y}} \le \tilde{\kappa} \mid H_0)$$
$$= 1 - \Phi\left(\tilde{\kappa} \big/ \sqrt{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}}\right) = 1 - \Phi(N_\alpha),$$

where $N_\alpha = \tilde{\kappa} \big/ \sqrt{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}}$ denotes the value, which is exceeded by a standardized normally distributed random variable with probability $P_{FA} = \alpha$.

Hence, the detection threshold is

$$\tilde{\kappa} = N_\alpha \sqrt{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}} = N_\alpha d \quad \left(\kappa = e^{(\tilde{\kappa} - \tilde{\mathbf{s}}^T \tilde{\mathbf{s}}/2)} = e^{(N_\alpha d - d^2/2)} \right),$$

where

$$d^2 = \tilde{\mathbf{s}}^T \tilde{\mathbf{s}} = \mathbf{s}^T \boldsymbol{\Sigma}_{\mathbf{UU}}^{-1} \mathbf{s} = n \cdot (S/N)_{\text{in}} = (S/N)_{\text{out}}$$

is called deflection coefficient. Finally, the detection probability can be determined by

$$P_D = P(T_{\mathbf{Y}} > \tilde{\kappa} \mid H_1) = 1 - P(T_{\mathbf{Y}} \le \tilde{\kappa} \mid H_1)$$

$$= 1 - \Phi\left((\tilde{\kappa} - \tilde{\mathbf{s}}^T \tilde{\mathbf{s}}) \big/ \sqrt{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}} \right) = 1 - \Phi(N_\alpha - d).$$

Receiver Operating Characteristic

$$P_{FA} = 1 - \Phi(N_\alpha)$$

and

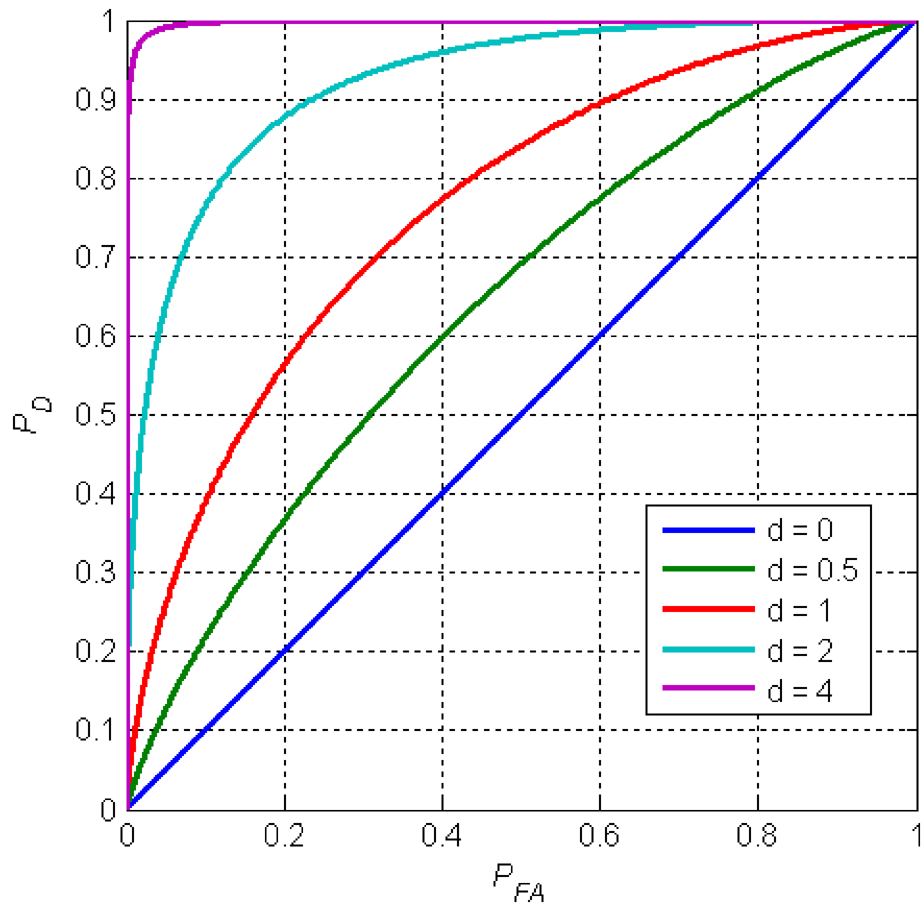
$$P_D = 1 - \Phi(N_\alpha - d)$$

with

$$d = \sqrt{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}}$$

and

$$N_\alpha \in (-\infty, \infty)$$



Exercise 4.2-3:

Detection of a Gaussian signal with known distribution in Gaussian noise with known distribution, signal and noise are stochastically independent

Bounds for P_{FA} and P_D

$\Lambda_{\mathbf{X}}(\mathbf{X}) = f_{\mathbf{X}}(\mathbf{X}\,|\,1)/f_{\mathbf{X}}(\mathbf{X}\,|\,0)$ represents a random variable, where $f_{\mathbf{X}}(\mathbf{X}|0) > 0$ is supposed for all $\mathbf{X}$.

For the probability of false alarm

$$P_{FA} = \int_{\Lambda_{\mathbf{X}}(\mathbf{x})>\kappa} f_{\mathbf{X}}(\mathbf{x}\,|\,0)\,d\mathbf{x}$$

$$= \int_{\kappa}^{\infty} f_{\Lambda_{\mathbf{X}}}(\lambda\,|\,0)\,d\lambda = \int_{\kappa}^{\infty} e^{\nu(c)-\nu(c)+c\ln\lambda-c\ln\lambda} f_{\Lambda_{\mathbf{X}}}(\lambda\,|\,0)\,d\lambda$$

with

$$\nu(c) = \ln\left(\int_0^{\infty} e^{c\ln\lambda} f_{\Lambda_{\mathbf{X}}}(\lambda\,|\,0)\,d\lambda\right) = \ln\left(\int_0^{\infty} \lambda^c f_{\Lambda_{\mathbf{X}}}(\lambda\,|\,0)\,d\lambda\right)$$

$$= \ln\left(\int_{\mathbb{R}^n} \left(f_{\mathbf{X}}(\mathbf{x}\,|\,1)/f_{\mathbf{X}}(\mathbf{x}\,|\,0)\right)^c f_{\mathbf{X}}(\mathbf{x}\,|\,0)\,d\mathbf{x}\right)$$

and $c \geq 0$ the so-called Chernoff bound

$$P_{FA} \leq e^{\nu(c) - c\ln\kappa} \int_{\kappa}^{\infty} e^{c\ln\lambda} f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda \Big/ e^{\nu(c)} \leq$$

$$\leq e^{\nu(c) - c\ln\kappa} \int_{0}^{\infty} e^{c\ln\lambda} f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda \Big/ e^{\nu(c)} = e^{\nu(c) - c\ln\kappa},$$

can be derived.

Moreover, one can show that $d\nu/dc$ is monotonic increasing and that $\nu(c)$ is convex. Hence, the Chernoff bound is minimized by a c_0 with $0 \leq c_0 \leq \infty$ that satisfies

$$c_0 = \begin{cases} \text{solution of } \nu'(c) = \ln\kappa & \text{if } \nu'(0) \leq \ln\kappa \leq \nu'(\infty) \\ 0 & \text{if } \ln\kappa < \nu'(0) \\ \infty & \text{elsewhere} \end{cases}.$$

Similarly, for the probability of detection

$$P_D = 1 - \int_{\Lambda_{\mathbf{x}}(\mathbf{x})<\kappa} f_{\mathbf{x}}(\mathbf{x}\,|\,1)\,d\mathbf{x} = 1 - \int_{\Lambda_{\mathbf{x}}(\mathbf{x})<\kappa} \Lambda_{\mathbf{x}}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}\,|\,0)\,d\mathbf{x}$$

$$= 1 - \int_0^{\kappa} \lambda f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda = 1 - \int_0^{\kappa} e^{\ln\lambda} f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda$$

$$= 1 - \int_0^{\kappa} e^{\nu(c)-\nu(c)+c\ln\lambda+(1-c)\ln\lambda} f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda$$

and $c \leq 1$ a lower bound can be found by

$$P_D \geq 1 - e^{\nu(c)+(1-c)\ln\kappa} \int_0^{\kappa} e^{c\ln\lambda} f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda \Big/ e^{\nu(c)} \geq$$

$$\geq 1 - e^{\nu(c)+(1-c)\ln\kappa} \int_0^{\infty} e^{c\ln\lambda} f_{\Lambda_{\mathbf{x}}}(\lambda\,|\,0)\,d\lambda \Big/ e^{\nu(c)} =$$

$$= 1 - e^{\nu(c)+(1-c)\ln\kappa}.$$

The bound for the probability of detection is maximized by a c_1 with $-\infty \le c_1 \le 1$ that satisfies

$$c_0 = \begin{cases} \text{solution of } \nu'(c) = \ln\kappa & \text{if } \nu'(-\infty) \le \ln\kappa \le \nu'(1) \\ 1 & \text{if } \ln\kappa > \nu'(1) \\ -\infty & \text{elsewhere} \end{cases}.$$

For normal distributions $\mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{x},i}, \boldsymbol{\Sigma}_{\mathbf{xx},i})$, $i = 0,1$ we obtain

$$\begin{aligned} \nu(c) &= \frac{c}{2}\ln(\det \boldsymbol{\Sigma}_{\mathbf{xx},0}) + \frac{1-c}{2}\ln(\det \boldsymbol{\Sigma}_{\mathbf{xx},1}) \\ &- \frac{1}{2}\ln\left(\det\left((1-c)\boldsymbol{\Sigma}_{\mathbf{xx},1} + c\,\boldsymbol{\Sigma}_{\mathbf{xx},0}\right)\right) - \frac{1}{2}(c - c^2)\times \\ &\times (\boldsymbol{\mu}_{\mathbf{x},0} - \boldsymbol{\mu}_{\mathbf{x},1})^T \left((1-c)\boldsymbol{\Sigma}_{\mathbf{xx},1} + c\,\boldsymbol{\Sigma}_{\mathbf{xx},0}\right)^{-1} (\boldsymbol{\mu}_{\mathbf{x},0} - \boldsymbol{\mu}_{\mathbf{x},1}) \end{aligned}$$

which simplifies to

$$\nu(c) = -\frac{c(1-c)}{2}\mathbf{s}^T\boldsymbol{\Sigma}_{\mathbf{UU}}^{-1}\mathbf{s} = -\frac{c(1-c)}{2}d^2$$

for the detection problem introduced on p. 30. Hence,

$$\nu'(\tilde{c}) = \left(\tilde{c} - \frac{1}{2}\right)d^2 = \ln\kappa \;\Rightarrow\; \tilde{c} = \frac{1}{2} + \frac{\ln\kappa}{d^2}$$

together with the constraints on c_0 and c_1 provides

$$c_0 = \tilde{c},\ \text{if } \tilde{c} > 1,\ c_1 = \tilde{c},\ \text{if } \tilde{c} < 0 \ \text{and}\ c_0 = c_1 = \tilde{c},\ \text{if } 0 \le \tilde{c} \le 1.$$

Finally, for $\kappa = 1$ we obtain $c_0 = c_1 = .5$ and the inequalities

$$P_{FA} \le e^{-d^2/8} \quad \text{and} \quad P_D \ge 1 - e^{-d^2/8}.$$

4.2.3 Bayes Hypothesis Testing

Average Error Probability

Suppose the occurrence of the hypotheses can also be considered as a result of a random experiment, where

$$p_0 = P(H_0) \quad \text{and} \quad p_1 = P(H_1) = 1 - p_0$$

denote the a priori probabilities for the occurrence of the null and alternative hypothesis. Now, we are interested in finding an optimal test $\psi(\mathbf{x})$ that minimizes the probability of a wrong decision, i.e. the average error probability

$$P_E(\psi) = P(H_0) \underbrace{P(H_1 \mid H_0)}_{\substack{P_{FA} = \beta_\psi(0) \\ \text{type 1 error probability}}} + P(H_1) \underbrace{P(H_0 \mid H_1)}_{\substack{P_M = (1 - \beta_\psi(1)) \\ \text{type 2 error probability}}} .$$

Since

$$P_E(\psi) = p_0 \int_{\mathbb{R}^n} \psi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}\,|\,0)\,d\mathbf{x} + p_1 \int_{\mathbb{R}^n} \big(1-\psi(\mathbf{x})\big) f_{\mathbf{X}}(\mathbf{x}\,|\,1)\,d\mathbf{x}$$

$$= p_1 + \int_{\mathbb{R}^n} \psi(\mathbf{x}) \big(p_0\, f_{\mathbf{X}}(\mathbf{x}\,|\,0) - p_1\, f_{\mathbf{X}}(\mathbf{x}\,|\,1)\big)\,d\mathbf{x}$$

the test function that minimizes P_E is given by

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{if } p_1\, f_{\mathbf{X}}(\mathbf{x}\,|\,1) > p_0\, f_{\mathbf{X}}(\mathbf{x}\,|\,0) \\ 0 & \text{if } p_1\, f_{\mathbf{X}}(\mathbf{x}\,|\,1) < p_0\, f_{\mathbf{X}}(\mathbf{x}\,|\,0) \end{cases}.$$

This test is a special case of the more general Bayesian approach. It is for $p_1 > 0$ obviously a most powerful test for (α_ψ, H_0, H_1) with

$$\alpha_\psi = \int_{\mathbb{R}^n} \psi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}\,|\,0)\,d\mathbf{x} \quad \text{and} \quad \kappa = p_0 / p_1 .$$

If the prior probabilities are equal, i.e.

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{\mathbf{x}}(\mathbf{x}\,|\,1) > f_{\mathbf{x}}(\mathbf{x}\,|\,0) \\ 0 & \text{if } f_{\mathbf{x}}(\mathbf{x}\,|\,1) < f_{\mathbf{x}}(\mathbf{x}\,|\,0) \end{cases},$$

the hypothesis with the larger conditional likelihood is chosen and the test is termed (conditional) maximum likelihood test.

Moreover, exploiting Bayes rule

$$P(H_i\,|\,\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}\,|\,H_i)P(H_i)}{f_{\mathbf{x}}(\mathbf{x})} = \frac{f_{\mathbf{x}}(\mathbf{x}\,|\,i)\,p_i}{f_{\mathbf{x}}(\mathbf{x})} \qquad i = 0,1,$$

where

$$f_{\mathbf{x}}(\mathbf{x}) = p_0\, f_{\mathbf{x}}(\mathbf{x}\,|\,0) + p_1\, f_{\mathbf{x}}(\mathbf{x}\,|\,1)$$

does not depend on the true hypothesis, the test can be expressed by

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{if } P(H_1 \mid \mathbf{x}) > P(H_0 \mid \mathbf{x}) \\ 0 & \text{if } P(H_1 \mid \mathbf{x}) < P(H_0 \mid \mathbf{x}) \end{cases}$$

and is therefore called maximum a posteriori (MAP) test.

Bayes risk

Now, we are going to generalize the former approach by assigning losses to each type of error, i.e.

L_{00} = loss of a correct rejection,
L_{01} = loss of a miss,
L_{10} = loss of a false alarm,
L_{11} = loss of a correct detection,

introducing the risk

$$R(\psi\,|\,\boldsymbol{\theta}) = \begin{cases} L_{00}\underbrace{P(H_0\,|\,H_0)}_{(1-\beta_\psi(0))} + L_{10}\underbrace{P(H_1\,|\,H_0)}_{P_{FA}=\beta_\psi(0)} & \text{if } \boldsymbol{\theta}=\boldsymbol{\theta}_0 \\ L_{01}\underbrace{P(H_0\,|\,H_1)}_{P_M=(1-\beta_\psi(1))} + L_{11}\underbrace{P(H_1\,|\,H_1)}_{P_D=\beta_\psi(1)} & \text{if } \boldsymbol{\theta}=\boldsymbol{\theta}_1 \end{cases}$$

and defining the expected risk

$$R(\psi) = \mathrm{E}\big(R(\psi\,|\,\boldsymbol{\theta})\big) = \sum_{i=0}^{1}\sum_{j=0}^{1} L_{ij}\,P(H_i\,|\,H_j)P(H_j)$$

as Bayes risk which can also be expressed by

$$R(\psi) = L_{00}\big(1-\beta_\psi(0)\big)p_0 + L_{01}\big(1-\beta_\psi(1)\big)p_1 + \\ + L_{10}\,\beta_\psi(0)\,p_0 + L_{11}\,\beta_\psi(1)\,p_1.$$

Since

$$R(\psi) = L_{00}\,p_0 + L_{01}\,p_1 + \underbrace{(L_{10}-L_{00})\,p_0\,\beta_\psi(0) - (L_{01}-L_{11})\,p_1\,\beta_\psi(1)}_{\int_{\mathbb{R}^n}\psi(\mathbf{x})\left[(L_{10}-L_{00})\,p_0\,f_{\mathbf{X}}(\mathbf{x}|0)-(L_{01}-L_{11})\,p_1\,f_{\mathbf{X}}(\mathbf{x}|1)\right]d\mathbf{x}}$$

the test function $\psi(\mathbf{x})$ should be one only if the expression in rectangular brackets is negative, i.e.

$$(L_{10}-L_{00})\,p_0\,f_{\mathbf{X}}(\mathbf{x}\,|\,0) < (L_{01}-L_{11})\,p_1\,f_{\mathbf{X}}(\mathbf{x}\,|\,1)$$

Thus, assuming $L_{10} > L_{00}$ and $L_{01} > L_{11}$ the test function that minimizes $R(\psi)$ is given by

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{\mathbf{X}}(\mathbf{x}|1) > \kappa\, f_{\mathbf{X}}(\mathbf{x}|0) \\ 0 & \text{if } f_{\mathbf{X}}(\mathbf{x}|1) < \kappa\, f_{\mathbf{X}}(\mathbf{x}|0) \end{cases}, \quad \text{where } \kappa = \frac{(L_{10}-L_{00})\,p_0}{(L_{01}-L_{11})\,p_1}.$$

Exercise 4.2-4:
Binary Signal Detection, MAP hypothesis test

Multiple Hypothesis Testing

Here, we wish to decide among M possible hypotheses.

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ versus } H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1 \text{ versus } \ldots$$
$$\ldots \text{ versus } H_{M-1} : \boldsymbol{\theta} = \boldsymbol{\theta}_{M-1}.$$

The loss assigned to the decision to choose H_i when H_j is true is denoted by L_{ij}. Hence, the expected/Bayes risk can be expressed by

$$R(\psi) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} L_{ij} P(H_i \mid H_j) P(H_j).$$

Our goal is now to construct a test function $\psi(\mathbf{x})$ that takes on values in the set $\{0, 1, \ldots, M-1\}$, where $\psi(\mathbf{x}) = m$ corresponds to selecting H_m.

Let $R_0,\dots,R_{M-1}$ denote the partitioning of the observation space for deciding $H_0,\dots,H_{M-1}$ respectively, so that

$$R(\psi) = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} L_{ij} \int_{R_i} f_{\mathbf{x}}(\mathbf{x} \mid H_j) P(H_j) d\mathbf{x}$$

$$= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} L_{ij} f_{\mathbf{x}}(\mathbf{x} \mid H_j) P(H_j) d\mathbf{x}$$

$$= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} L_{ij} P(H_j \mid \mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \sum_{i=0}^{M-1} \int_{R_i} L_i(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x},$$

where

$$L_i(\mathbf{x}) = \sum_{j=0}^{M-1} L_{ij} P(H_j \mid \mathbf{x}), \quad i = 0,\dots,M-1$$

describes the average loss of deciding H_i if $\mathbf{x}$ is observed.

Now, each observation **x** has to be assigned to one and only one R_i. The assignment of **x** to R_i contributes to the Bayes risk with $L_i(\mathbf{x})f_{\mathbf{x}}(\mathbf{x})d\mathbf{x}$.

To minimize the Bayes risk we should assign **x** to R_k if

$$k = \underset{i\in\{0,\ldots,M-1\}}{\operatorname{argmin}} \{L_i(\mathbf{x})\}.$$

Hence, the test function is given by

$$\psi(\mathbf{x}) = \begin{cases} 0 & \text{if } L_0(\mathbf{x}) < \min\limits_{i\neq 0}\{L_i(\mathbf{x})\} \\ 1 & \text{if } L_1(\mathbf{x}) < \min\limits_{i\neq 1}\{L_i(\mathbf{x})\} \\ \vdots & \vdots \\ M-1 & \text{if } L_{M-1}(\mathbf{x}) < \min\limits_{i\neq M-1}\{L_i(\mathbf{x})\} \end{cases}.$$

For the loss function defined by

$$L_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases},$$

i.e. $R(\psi) = P_E$, we have to minimize

$$L_i(\mathbf{x}) = \sum_{j=0, j\neq i}^{M-1} P(H_j|\mathbf{x}) = \sum_{j=0}^{M-1} P(H_j|\mathbf{x}) - P(H_i|\mathbf{x}), \quad i = 0,\ldots,M-1$$

or equivalently to maximize $P(H_i|\mathbf{x})$ with respect to i. Since one seeks to maximize the a posteriori probability

$$\psi(\mathbf{x}) = \begin{cases} 0 & \text{if } P(H_0|\mathbf{x}) > \max\limits_{i\neq 0}\{P(H_i|\mathbf{x})\} \\ \vdots & \vdots \\ M-1 & \text{if } P(H_{M-1}|\mathbf{x}) > \max\limits_{i\neq M-1}\{P(H_i|\mathbf{x})\} \end{cases}$$

is called *M*-ary maximum a posteriori (MAP) test.

If the prior probabilities are equal and therefore

$$P(H_i|\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}|H_i)P(H_i)}{f_{\mathbf{x}}(\mathbf{x})} = \frac{f_{\mathbf{x}}(\mathbf{x}|i)}{\sum_{i=0}^{M-1} f_{\mathbf{x}}(\mathbf{x}|i)}, \quad i = 0,\ldots,M-1$$

we can conclude that to maximize $P(H_i|\mathbf{x})$ we only have to maximize $f_{\mathbf{x}}(\mathbf{x}|i)$. Hence,

$$\psi(\mathbf{x}) = \begin{cases} 0 & \text{if } f_{\mathbf{x}}(\mathbf{x}|0) > \max\limits_{i\neq 0}\{f_{\mathbf{x}}(\mathbf{x}|i)\} \\ \vdots & \vdots \\ M-1 & \text{if } f_{\mathbf{x}}(\mathbf{x}|M-1) > \max\limits_{i\neq M-1}\{f_{\mathbf{x}}(\mathbf{x}|i)\} \end{cases}$$

which is known as *M*-ary maximum likelihood (ML) test.

Exercise 4.2-5:
Multiple Signal Detection, MAP hypotheses test

4.3 Composite Hypothesis Testing

4.3.1 Sufficient Statistic

A function $\mathbf{T}=\mathbf{t}(\mathbf{X})$ that is only depending on the observation model $\mathbf{X}$ is called a statistic.

Definition: A statistic $\mathbf{T}=\mathbf{t}(\mathbf{X})$ is called sufficient for the parameter $\boldsymbol{\theta}$ if the conditional distribution of $\mathbf{X}$ given $\mathbf{T}=\mathbf{t}(\mathbf{x})$ is independent of $\boldsymbol{\theta}$ for all $\mathbf{t}$, i.e.

$$F_{\mathbf{X}}\left(\mathbf{x}\,|\,\mathbf{T}=\mathbf{t}(\mathbf{x});\boldsymbol{\theta}\right)=F_{\mathbf{X}}\left(\mathbf{x}\,|\,\mathbf{T}=\mathbf{t}(\mathbf{x})\right).$$

Because the conditional distribution has to be determined a direct evaluation of sufficiency is usually difficult.

Fortunately, the following theorem exists whose conditions can be verified easily.

<u>Theorem:</u> (factorization theorem for densities)
A necessary and sufficient condition for a statistic $\mathbf{T} = \mathbf{t}(\mathbf{X})$ to be sufficient is that there exist non-negative functions $g(\mathbf{t} \mid \theta)$ and $h(\mathbf{x})$ such that $f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$ satisfies

$$f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta}) = g\left(\mathbf{t}(\mathbf{x}) \mid \boldsymbol{\theta}\right) \cdot h(\mathbf{x}).$$

Minimal Sufficient Statistic
If for any sufficient statistic $\mathbf{T}'$ there exists a function s such that $\mathbf{T} = s(\mathbf{T}')$.

Complete Sufficient Statistic
A sufficient statistic $\mathbf{T}$ is said to be complete if condition $E_{\boldsymbol{\theta}}\left(\mathbf{f}(\mathbf{T})\right) = \mathbf{0}$ for all $\boldsymbol{\theta} \in \Omega$ implies $\mathbf{f}(\mathbf{T}) = \mathbf{0}$ with probability 1 for all $\boldsymbol{\theta}$.

Exponential Families

A family $\{F_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})\}$ of distributions is forming a *k*-dimensional exponential family if the distributions $F_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})$ have densities of the form

$$f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta}) = h(\mathbf{x}) \cdot \exp\left(\sum_{i=1}^{k} \xi_i(\boldsymbol{\theta}) t_i(\mathbf{x}) - B(\boldsymbol{\theta})\right).$$

Frequently, it is more convenient to use the ξ_i as the parameters and write the density in the canonical form

$$f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\xi}) = h(\mathbf{x}) \cdot \exp\left(\sum_{i=1}^{k} \xi_i\, t_i(\mathbf{x}) - A(\boldsymbol{\xi})\right).$$

Applying the factorization theorem for densities one can easily observe that $\mathbf{T} = (T_1,\ldots,T_k)^T = (t_1(\mathbf{X}),\ldots,t_k(\mathbf{X}))^T$ constitutes a sufficient statistic for the exponential family.

Exercise 4.3-1:
Introductory example

4.3.2 Bayesian Approach

The unknown parameter vector of the density function $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$ is supposed to be a realization $\boldsymbol{\theta}\in\Omega$ of the random vector $\boldsymbol{\Theta}$ that possesses the densities $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|0)$ and $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|1)$ under H_0 and H_1, respectively.

The density functions of $\mathbf{X}$ and $\boldsymbol{\Theta}$ are supposed to be known. Hence, with

$$\beta_\phi(\boldsymbol{\theta}) = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})\, d\mathbf{x}$$

the expected power of the test $\phi(\mathbf{x})$ can be defined by

$$\begin{aligned}\overline{\beta}_\phi(i) &= \int_{\mathbb{R}^p} \beta_\phi(\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|i)\, d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^n} \phi(\mathbf{x}) \left(\int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|i)\, d\boldsymbol{\theta} \right) d\mathbf{x} = \int_{\mathbb{R}^n} \phi(\mathbf{x}) \overline{f}_{\mathbf{X}}(\mathbf{x}|i)\, d\mathbf{x},\end{aligned}$$

where $\bar{f}_{\mathbf{X}}(\mathbf{x}|i)$ can be interpreted as the density function of $\mathbf{X}$ under H_i which is independent of $\boldsymbol{\theta}$.

Thus, the test problem with composite hypotheses could be transformed into one with simple hypotheses, so that the Neyman-Pearson Lemma and the likelihood ratio

$$\Lambda_{\mathbf{X}}(\mathbf{x}) = \frac{\bar{f}_{\mathbf{X}}(\mathbf{x}\,|\,1)}{\bar{f}_{\mathbf{X}}(\mathbf{x}\,|\,0)} = \frac{\int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta}) f_{\Theta}(\boldsymbol{\theta}\,|\,1)\,d\boldsymbol{\theta}}{\int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta}) f_{\Theta}(\boldsymbol{\theta}\,|\,0)\,d\boldsymbol{\theta}} \quad \text{if } \bar{f}_{\mathbf{X}}(\mathbf{x}\,|\,0) > 0$$

provide a most powerful test

$$\phi_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \Lambda_{\mathbf{X}}(\mathbf{x}) = \dfrac{\bar{f}_{\mathbf{X}}(\mathbf{x}\,|\,1)}{\bar{f}_{\mathbf{X}}(\mathbf{x}\,|\,0)} > \kappa \\ 0 & \text{elsewhere} \end{cases}.$$

Theorem: Suppose, there exits a sufficient statistic $\mathbf{t}(\mathbf{x})$ for $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ denotes a realization of the random parameter vector $\boldsymbol{\Theta}$. Then, a most powerful test $\phi(\mathbf{x})$ does only depend over $\mathbf{t}(\mathbf{x})$ on $\mathbf{x}$.

Proof:

In accordance with the Neyman-Pearson Lemma a most powerful test asks whether

$$\int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}\,|\,1)\,d\boldsymbol{\theta} \gtrless \kappa \int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}\,|\,0)\,d\boldsymbol{\theta}.$$

Using $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) = g\left(\mathbf{t}(\mathbf{x})|\boldsymbol{\theta}\right) h(\mathbf{x})$ with $h(\mathbf{x}) \neq 0$ we obtain

$$\int_{\mathbb{R}^p} g\left(\mathbf{t}(\mathbf{x})\,|\,\boldsymbol{\theta}\right) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}\,|\,1)\,d\boldsymbol{\theta} \gtrless \kappa \int_{\mathbb{R}^p} g\left(\mathbf{t}(\mathbf{x})\,|\,\boldsymbol{\theta}\right) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}\,|\,0)\,d\boldsymbol{\theta}$$

which was to be proven.

Exercise 4.3-2:
Detection of a sinusoid with unknown phase in white Gaussian noise with known variance

4.3.3 Monotone Likelihood Ratio and UMP Tests

Definition: A real-parameter family of densities $\{f_{\mathbf{x}}(\mathbf{x}\,|\,\theta)\}$ is said to have monotone likelihood ratio if there exists a real-valued function $t(\mathbf{x})$ such that for any $\theta > \tilde{\theta}$ the densities $f_{\mathbf{x}}(\mathbf{x}|\theta)$ and $f_{\mathbf{x}}(\mathbf{x}|\tilde{\theta})$ are distinct and the ratio

$$\frac{f_{\mathbf{x}}(\mathbf{x}\,|\,\theta)}{f_{\mathbf{x}}(\mathbf{x}\,|\,\tilde{\theta})} = g\left(t(\mathbf{x})\,|\,\tilde{\theta},\theta\right)$$

is a non-decreasing function of $t(\mathbf{x})$.

Example: (introductory example, cf. Exercise 4.1-1,4.3-1)

$$f_{\mathbf{x}}(\mathbf{x}\,|\,\eta) = (2\pi\sigma_Z^2)^{-\frac{n}{2}} \exp\left(\frac{-\mathbf{x}^T\mathbf{x}}{2\sigma_Z^2}\right) \exp\left(\frac{\eta\,\mathbf{s}^T\mathbf{x}}{\sigma_Z^2}\right) \exp\left(\frac{-\eta^2\mathbf{s}^T\mathbf{s}}{2\sigma_Z^2}\right)$$

Hence, the likelihood ratio given by

$$\frac{f_{\mathbf{X}}(\mathbf{x}\,|\,\eta)}{f_{\mathbf{X}}(\mathbf{x}\,|\,0)} = \exp\left(\frac{-\eta^2\mathbf{s}^T\mathbf{s}}{2\sigma_Z^2}\right)\exp\left(\frac{\eta\,\mathbf{s}^T\mathbf{x}}{\sigma_Z^2}\right)$$

$$= g\big(t(\mathbf{x})\,|\,0,\eta\big) \quad \text{with} \quad t(\mathbf{x}) = \mathbf{s}^T\mathbf{x}$$

increases monotonic in $t(\mathbf{x})$, i.e. the family $\{f_{\mathbf{X}}(\mathbf{x}\,|\,\eta)\}$ has monotone likelihood ratio.

Theorem:
Let the random vector **X** have a real-parameter density $f_{\mathbf{X}}(\mathbf{x}|\theta)$ with monotone likelihood ratio in $t(\mathbf{x})$. For testing

$$H_0 : \Omega_0 = \{\theta : \theta \le \theta_0\} \ \text{ against } \ H_1 : \Omega_1 = \{\theta : \theta > \theta_0\}$$

there exists a uniformly most powerful test

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } t(\mathbf{x}) > \tilde{\kappa} \\ 0 & \text{if } t(\mathbf{x}) < \tilde{\kappa} \\ \gamma & \text{if } t(\mathbf{x}) = \tilde{\kappa} \end{cases},$$

where γ and $\tilde{\kappa}$ are determined by

$$\beta_\phi(\theta_0) = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x} \mid \theta_0)\, d\mathbf{x} = \alpha$$

and the power function

$$\beta_\phi(\theta) = \mathrm{E}\big(\phi(\mathbf{X})\big) = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x} \mid \theta)\, d\mathbf{x}$$

is strictly monotonic increasing for all θ for which

$$0 < \beta_\phi(\theta) < 1.$$

Theorem:

One-parameter exponential families which possess density functions of the form

$$f_{\mathbf{X}}(\mathbf{x} \mid \theta) = h(\mathbf{x}) \cdot \exp\big(\xi(\theta) t(\mathbf{x}) - B(\theta)\big)$$

have a monotone likelihood ratio in the sufficient statistic $t(\mathbf{x})$, provided $\xi(\theta)$ is strictly monotonic increasing in θ.

Hence, if the distribution of **X** (model of the observation vector) belongs to a one-parameter exponential family with strictly monotonic increasing $\xi(\theta)$, then there exits a uniformly most powerful test $\phi(\mathbf{x})$ for testing

$$H_0 : \Omega_0 = \{\theta : \theta \le \theta_0\} \text{ against } H_1 : \Omega_1 = \{\theta : \theta > \theta_0\}.$$

Exercise 4.3-3:

Detection of a deterministic signal with unknown amplitude in Gaussian noise with known distribution

Exercise 4.3-4:

Detection of a Gaussian signal with unknown amplitude in Gaussian noise with known distribution, signal and noise are stochastically independent

4.3.4 Invariance Principle and UMP Invariant Tests

Invariant and uniformly most powerful invariant tests

Definition: (group of transformations)

A set G of transformations of some set $\mathcal{X}$ onto itself is called group of transformations if the following holds.

1) The identity transformation belongs to G, i.e.

$$id \in G \quad \text{with} \quad id(x) = x$$

2) G is closed with respect to inversion, i.e.

$$g \in G \Rightarrow g^{-1} \in G \quad \text{with} \quad g^{-1}\big(g(x)\big) = x$$

3) G is closed with respect to compositions, i.e.

$$g_1 \in G, g_2 \in G \Rightarrow g_1 \circ g_2 \in G$$

Example: (scaling and translation)

Let $\mathcal{X} = \mathbb{R}$ and

$$G = \left\{ g_{a,b} : \; g_{a,b}(x) = ax + b, \; a, b \in \mathbb{R}, \; a \neq 0 \right\}$$

for $x \in \mathbb{R}$. Since

1) $g_{1,0} \in G$, identity transformation

2) $g_{1/a,-b/a} \in G$, inverse transformation of $g_{a,b}$

3) $g_{a_1 a_2, a_1 b_2 + b_1} \in G$, composition of g_{a_1,b_1} and g_{a_2,b_2}

we can conclude that G is a group of transformations of $\mathcal{X}$ onto itself.

Let G denote a group of one-to-one transformations of the sample space $\mathcal{X}$ onto itself.

<u>Definition:</u> *(Invariance of a hypotheses testing problem)*
A hypotheses testing problem is said to remain invariant under transformations $\mathbf{g}(\mathbf{x}) \in G$ if $\mathbf{g}(\mathbf{x})$ leaves the distribution invariant in form, i.e.

$$P\left(\mathbf{Y} = \mathbf{g}(\mathbf{X}) \leq \mathbf{y} \,|\, \boldsymbol{\theta}\right) = F_{\mathbf{Y}}(\mathbf{y} \,|\, \boldsymbol{\theta}) = F_{\mathbf{X}}\left(\mathbf{y} \,|\, \bar{\boldsymbol{\theta}} = \bar{\mathbf{g}}(\boldsymbol{\theta})\right)$$

$$= P\left(\mathbf{X} \leq \mathbf{y} \,|\, \bar{\boldsymbol{\theta}} = \bar{\mathbf{g}}(\boldsymbol{\theta})\right) \quad \text{with} \quad \bar{\mathbf{g}}(\Omega) = \Omega,$$

and if the corresponding $\bar{\mathbf{g}}(\boldsymbol{\theta})$ preserves both Ω_0 and Ω_1, so that $\bar{\mathbf{g}}(\Omega_0) = \Omega_0$ holds in addition to $\bar{\mathbf{g}}(\Omega) = \Omega$.

Exercise 4.3-5:
Invariant distribution, invariant parameter space

Example: (invariant test problem)
The problem of detecting a signal with unknown amplitude in Gaussian noise with unknown variance should not depend on the amplification of the receiver system.

$$H_0: \mathbf{X} = \mathbf{Z}, \qquad \mathbf{X} \sim \mathcal{N}_n(\mathbf{0}, \sigma_Z^2 \mathbf{I}), \quad \Omega_0 = \left\{(0, \sigma_Z^2)^T : \sigma_Z^2 > 0\right\}$$

$$H_1: \mathbf{X} = \mathbf{Z} + \eta \mathbf{s}, \ \mathbf{X} \sim \mathcal{N}_n(\eta \mathbf{s}, \sigma_Z^2 \mathbf{I}), \ \Omega_1 = \left\{(\eta, \sigma_Z^2)^T : \eta > 0, \sigma_Z^2 > 0\right\}$$

with density

$$f_{\mathbf{X}}(\mathbf{x} \,|\, \eta, \sigma_Z^2) = (2\pi\sigma_Z^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_Z^2}(\mathbf{x} - \eta \mathbf{s})^T (\mathbf{x} - \eta \mathbf{s})\right\},$$

where η and σ_Z^2 are unknown.

The transformation

$$\mathbf{Y} = \nu\,\mathbf{X} \quad \text{with} \quad \nu > 0$$

leads to the detection problem

$$H_0: \ \mathbf{Y} = \nu\mathbf{Z}, \qquad \mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \nu^2\sigma_Z^2\mathbf{I})$$

$$H_1: \ \mathbf{Y} = \nu\mathbf{Z} + \nu\eta\mathbf{s}, \ \mathbf{Y} \sim \mathcal{N}_n(\nu\eta\mathbf{s}, \nu^2\sigma_Z^2\mathbf{I}).$$

The detection problem remains invariant under the transformation (positive scaling changes) since it preserves the distribution type

$$f_{\mathbf{Y}}(\mathbf{y}|\nu\eta, \nu^2\sigma_Z^2) = f_{\mathbf{X}}(\mathbf{y}/\nu|\eta, \sigma_Z^2)\nu^{-n}$$

$$= (2\pi\sigma_Z^2)^{-n/2}\nu^{-n}\exp\left\{-\frac{(\mathbf{y}/\nu - \eta\mathbf{s})^T(\mathbf{y}/\nu - \eta\mathbf{s})}{2\sigma_Z^2}\right\}$$

$$f_{\mathbf{Y}}(\mathbf{y}|\nu\eta,\nu^2\sigma_Z^2)=(2\pi\nu^2\sigma_Z^2)^{-n/2}\exp\left\{-\frac{(\mathbf{y}-\nu\eta\mathbf{s})^T(\mathbf{y}-\nu\eta\mathbf{s})}{2\nu^2\sigma_Z^2}\right\}$$

$$=f_{\mathbf{X}}(\mathbf{y}|\bar{\eta},\bar{\sigma}_Z^2)\quad\text{with}\quad\bar{\eta}=\nu\eta\quad\text{and}\quad\bar{\sigma}_Z^2=\nu^2\sigma_Z^2$$

and the parameter set

$$\Omega_0=\left\{(0,\bar{\sigma}_Z^2)^T:\ \bar{\sigma}_Z^2>0\right\},\quad\Omega_1=\left\{(\bar{\eta},\bar{\sigma}_Z^2)^T:\ \bar{\eta}>0,\bar{\sigma}_Z^2>0\right\}.$$

as required by the definition above.

Definition: *(Invariance of a test)*
A test function $\phi(\mathbf{x})$ satisfying

$$\phi\big(\mathbf{g}(\mathbf{x})\big)=\phi(\mathbf{x})\quad\forall\mathbf{x}\in\mathcal{X}\ \text{ and }\ \forall\mathbf{g}\in G$$

is said to be invariant under G.

Definition: *(Invariance & maximal invariance of a statistic)*
A statistic $\mathbf{t}(\mathbf{x})$ is said to be invariant if

$$\mathbf{t}(\mathbf{x}) = \mathbf{t}\big(\mathbf{g}(\mathbf{x})\big) \quad \forall \mathbf{x} \in \mathcal{X} \text{ and } \forall \mathbf{g} \in G$$

holds and is said to be maximal invariant if in addition

$$\mathbf{t}(\mathbf{x}_1) = \mathbf{t}(\mathbf{x}_2) \text{ implies } \mathbf{x}_2 = \mathbf{g}(\mathbf{x}_1) \text{ for some } \mathbf{g} \in G.$$

Remark:

The distribution of a maximal invariant statistic $\mathbf{T} = \mathbf{t}(\mathbf{X})$ depends only on a parameter vector of the same dimension as $\mathbf{T}$. Thus the invariance principle leads to a reduction of the parameter space.

Theorem:

Let $\mathbf{t}(\mathbf{x})$ be a maximal invariant statistic with respect to G. Then, a test $\phi(\mathbf{x})$ is invariant if and only if it depends on $\mathbf{x}$ only through $\mathbf{t}(\mathbf{x})$, i.e. there exists a function h such that

$$\phi(\mathbf{x}) = h\big(\mathbf{t}(\mathbf{x})\big) \quad \forall \mathbf{x} \in \mathcal{X}.$$

Remark:

For an invariant test problem one would like to find a maximally invariant statistic $\mathbf{t}(\mathbf{x})$. This statistic is then used to construct invariant tests. Within the restricted class of invariant tests, it is often possible to find a uniformly most powerful tests which is termed uniformly most powerful invariant test.

Exercise 4.3-6:
Invariant t-test

Exercise 4.3-7:
Detection of a deterministic signal with unknown amplitude in Gaussian noise with unknown scaling

Linear Hypotheses and Least Squares Estimation

An area for the application of the invariance principle are linear hypotheses, where we assume that the data **X** can be modeled by

$$\mathbf{X} = \mathbf{H}\boldsymbol{\theta} + \mathbf{Z},\ \ \mathbf{H} = \left(h_{ij}\right)_{i=1,\dots,n;\,j=1,\dots,p} \quad \text{and} \quad \mathbf{Z} \sim \mathcal{N}_n(0, \sigma_Z^2\,\mathbf{I}),$$

with $\boldsymbol{\theta} \in \mathbb{R}^p$ and $\sigma_Z^2 > 0$ unknown.

Hence, for a given observation vector **x** the least squares estimate is obtained by minimizing

$$q(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) = \mathbf{x}^T\mathbf{x} - 2\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

with respect to $\boldsymbol{\theta}$. After equating the gradient

$$\nabla_{\boldsymbol{\theta}}\, q(\boldsymbol{\theta}) = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

to zero, i.e. $\nabla_{\boldsymbol{\theta}} q(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, we can derive the estimate

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}, \quad \text{if } \operatorname{rank}(\mathbf{H}) = p.$$

Furthermore, the minimum of the sum of squares

$$\begin{aligned} q(\hat{\boldsymbol{\theta}}) &= \left(\mathbf{x} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right)^T \left(\mathbf{x} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right) \\ &= (\mathbf{x} - \mathbf{P}\mathbf{x})^T(\mathbf{x} - \mathbf{P}\mathbf{x}) = \mathbf{x}^T(\mathbf{I} - \mathbf{P})^T(\mathbf{I} - \mathbf{P})\mathbf{x} \\ &= \mathbf{x}^T\mathbf{P}^{\perp T}\mathbf{P}^{\perp}\mathbf{x} = \mathbf{x}^T\mathbf{P}^{\perp}\mathbf{x} = \operatorname{tr}(\mathbf{P}^{\perp}\mathbf{x}\mathbf{x}^T), \end{aligned}$$

provides by

$$s^2 = q(\hat{\boldsymbol{\theta}})/(n - p)$$

an estimate of σ_Z^2, where

$$\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T \quad \text{and} \quad \mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$$

are projection matrices, which project a vector $\mathbf{a} \in \mathbb{R}^n$ by $\mathbf{Pa}$ and $\mathbf{P}^{\perp}\mathbf{a}$ into $R(\mathbf{H})$ and $N(\mathbf{H}^T)$, respectively.

Statistical properties of the least squares estimator

a) $\hat{\mathbf{\Theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\theta}, \sigma_Z^2 (\mathbf{H}^T\mathbf{H})^{-1}\right)$

b) $(n-p)/\sigma_Z^2 \cdot S^2 \sim \chi^2_{n-p}$, where $S^2 = q(\hat{\boldsymbol{\theta}})/(n-p)$

c) $\hat{\mathbf{\Theta}}$ and S^2 are stochastically independent

d) $(\hat{\mathbf{\Theta}}^T, S^2)^T$ is a sufficient statistic for $(\boldsymbol{\theta}^T, \sigma_Z^2)^T$

e) $(\hat{\mathbf{\Theta}}^T, S^2)^T$ is an efficient estimator for $(\boldsymbol{\theta}^T, \sigma_Z^2)^T$,

i.e. $\operatorname{Cov}\left((\hat{\mathbf{\Theta}}^T, S^2)^T\right) = \mathcal{I}^{-1}\left((\boldsymbol{\theta}^T, \sigma_Z^2)^T\right)$

Geometrical interpretation of least squares estimation

The necessary condition

$$\mathbf{H}^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}})=\mathbf{0}$$

indicates that the vector-valued residual is orthogonal to the linear subspace $V_1 \subset \mathbb{R}^n$ spanned by the columns of the matrix $\mathbf{H}$. If the columns of $\mathbf{H}$ are linear independent, i.e. $\dim(V_1)=p$, the vector

$$\mathbf{H}\hat{\boldsymbol{\theta}}=\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}=\mathbf{P}\mathbf{x}$$

is the orthogonal projection of $\mathbf{x}$ in V_1. Thus, $\mathbf{H}\hat{\boldsymbol{\theta}}$ is independent of the coordinates in V_1, so that one can choose the basis of V_1 such that the columns of $\mathbf{H}$ become orthonormal, i.e. $\mathbf{H}^T\mathbf{H}=\mathbf{I}_p$. Moreover, $\mathrm{E}(\mathbf{X})=\mathbf{H}\boldsymbol{\theta}\in V_1$ holds.

Let V_0 be a linear subspaces of V_1. Then the hypotheses H_0 and H_1 are said to be linear if

$$H_0: \ (\boldsymbol{\theta}^T, \sigma_Z^2)^T \in \Omega_0 = \left\{ (\boldsymbol{\theta}^T, \sigma_Z^2)^T : \ \mathbf{H}\boldsymbol{\theta} \in V_0, \sigma_Z^2 > 0 \right\}$$

and

$$H_1: \ (\boldsymbol{\theta}^T, \sigma_Z^2)^T \in \Omega_1 = \left\{ (\boldsymbol{\theta}^T, \sigma_Z^2)^T : \ \mathbf{H}\boldsymbol{\theta} \in V_1 \setminus V_0, \sigma_Z^2 > 0 \right\}.$$

Now, supposing

$$\mathbf{H}\boldsymbol{\theta} = (\mathbf{H}_1, \mathbf{H}_2) \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} = \mathbf{H}_1\boldsymbol{\theta}_1 + \mathbf{H}_2\boldsymbol{\theta}_2 \quad \text{with} \quad \mathbf{H}_1^T\mathbf{H}_2 = \mathbf{0},$$

where

$$\mathbf{H}_1 = (\mathbf{h}_1, \ldots, \mathbf{h}_{p_1}), \mathbf{H}_2 = (\mathbf{h}_{p_1+1}, \ldots, \mathbf{h}_p) \ \text{ and } \ \boldsymbol{\theta}_1 \in \mathbb{R}^{p_1}, \boldsymbol{\theta}_2 \in \mathbb{R}^{p-p_1}$$

and

$$V_0 = \text{span}(\mathbf{h}_{p_1+1}, \ldots, \mathbf{h}_p), V_1 = \text{span}(\mathbf{h}_1, \ldots, \mathbf{h}_p),$$

one has to test the hypothesis

$$H_0: \ (\boldsymbol{\theta}^T, \sigma_Z^2)^T \in \Omega_0 = \left\{(\boldsymbol{\theta}^T, \sigma_Z^2)^T: \ \boldsymbol{\theta}_1 = \mathbf{0}, \boldsymbol{\theta}_2 \in \mathbb{R}^{p-p_1}, \sigma_Z^2 > 0\right\}$$

against the alternative

$$H_1: \ (\boldsymbol{\theta}^T, \sigma_Z^2)^T \in \Omega_1 = \left\{(\boldsymbol{\theta}^T, \sigma_Z^2)^T: \ \boldsymbol{\theta}_1^T \boldsymbol{\theta}_1 > 0, \boldsymbol{\theta}_2 \in \mathbb{R}^{p-p_1}, \sigma_Z^2 > 0\right\}.$$

Thus, if the hypotheses H_0 and H_1 are assigned to linear spaces appropriate detection problems can be defined.

Assuming that signals and interferences span the space V_1 and interferences only span the subspace V_0, then an element from V_0 is interpreted as interference and one from V_1 as signal plus interference.

If e.g. V_0 contains a dc component, then the dc component of a signal would be interpreted as interference and would therefore be undetectable.

Invariance properties of least squares estimation

The test problem (α, H_0, H_1) is invariant with respect to

a) arbitrary one-to-one transformations in V_0, orthogonal transformations in $V_1^\perp \subset \mathbb{R}^n$, orthogonal transformations in $V_0^\perp \cap V_1$, i.e. in the signal space of interest. Maximal invariant are then

$$W = w(\hat{\boldsymbol{\Theta}}_1) = \hat{\boldsymbol{\Theta}}_1^T \mathbf{D} \hat{\boldsymbol{\Theta}}_1 \quad \text{and} \quad Q = q(\hat{\boldsymbol{\Theta}}) = \mathbf{X}^T \mathbf{X} - \hat{\boldsymbol{\Theta}}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\Theta}},$$

where $\mathbf{D} = (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)$ with $\mathbf{H}^T\mathbf{H} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix}$.

b) positive changes of scaling. Maximal invariant is then

$$V = \frac{W/p_1}{Q/(n-p)}.$$

Statistical properties of the maximal invariant statistics

a) Distributional properties of Q, W and V under H_0

$Q/\sigma_Z^2 \sim \chi_{n-p}^2$ and $W/\sigma_Z^2 \sim \chi_{p_1}^2$,

Q and W are stochastically independent,

$V = (n-p)/p_1 \cdot W/Q \sim F_{p_1,n-p}$.

b) Distributional properties of Q, W and V under H_1

$Q/\sigma_Z^2 \sim \chi_{n-p}^2$ and $W/\sigma_Z^2 \sim$ noncentral $\chi_{p_1}^2$,

with noncentrality parameter $\delta^2 = \mathbf{\Theta}_1^T \mathbf{D} \mathbf{\Theta}_1 / \sigma_Z^2$,

Q and W are stochastically independent,

$V = (n-p)/p_1 \cdot W/Q \sim$ noncentral $F_{p_1,n-p}$

with noncentrality parameter δ^2.

Example: (F-test, least squares estimation)

Now, we consider the detection problem

$$H_0 : \mathbf{X} = \mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \sigma_Z^2 \mathbf{I}) \text{ versus } H_1 : \mathbf{X} = \mathbf{H}\boldsymbol{\theta} + \mathbf{Z} \sim \mathcal{N}_n(\mathbf{H}\boldsymbol{\theta}, \sigma_Z^2 \mathbf{I})$$

$$\text{with } \Omega_0 = \left\{ (\mathbf{0}^T, \sigma_Z^2)^T : \sigma_Z^2 > 0 \right\} \text{ and}$$

$$\Omega_1 = \left\{ (\boldsymbol{\theta}^T, \sigma_Z^2)^T : \boldsymbol{\theta}^T \boldsymbol{\theta} > 0, \sigma_Z^2 > 0 \right\}.$$

Application of the former results for $p_1 = p$ provides the F-distributed maximal invariant statistic

$$V = v(\mathbf{X}) = \frac{n-p}{p} \cdot \frac{w(\hat{\boldsymbol{\Theta}})}{q(\hat{\boldsymbol{\Theta}})} = \frac{n-p}{p} \cdot \frac{\hat{\boldsymbol{\Theta}}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\Theta}}}{\mathbf{X}^T \mathbf{X} - \hat{\boldsymbol{\Theta}}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\Theta}}}$$

$$= \frac{n-p}{p} \cdot \frac{\mathbf{X}^T \mathbf{P} \mathbf{X}}{\mathbf{X}^T \mathbf{X} - \mathbf{X}^T \mathbf{P} \mathbf{X}} \quad \text{with} \quad \mathbf{P} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T.$$

Since the noncentral *F*-distribution possesses monotone likelihood ratio the test function

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \ v(\mathbf{x}) > \tilde{\kappa} \\ 0 & \text{elsewhere} \end{cases}$$

defines a uniformly most powerful invariant test. For a given $P_{FA} = \alpha$ the threshold $\tilde{\kappa}$ can be obtained from

$$P_{FA} = P\left(V > \tilde{\kappa} | H_0\right) = 1 - P(V \leq \tilde{\kappa} = F_{p,n-p,\alpha} | H_0),$$

where $\tilde{\kappa} = \tilde{\kappa}(\alpha, p, n)$ does not depend on σ_Z^2, i.e. the test provides a so-called constant false alarm rate (CFAR). Finally, the P_D of the test can be calculated by

$$P_D(\alpha, p, n, \delta^2) = P\left(V > \tilde{\kappa} | H_1\right) = 1 - P(V \leq \tilde{\kappa} = F_{p,n-p,\delta^2,\alpha} | H_1)$$

with $\delta^2 = \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} / \sigma_Z^2$ (non-centrality parameter).

Exercise 4.3-8:
Detection of superimposed sinusoids with unknown amplitudes and phases in white Gaussian noise with unknown variance

4.3.5 Maximum Likelihood Ratio Test

If the methods considered up to now do not allow to construct uniformly most powerful tests, maximum likelihood ratio tests (MLRT) also known as generalized likelihood ratio tests (GLRT) are frequently applied.

Definition:
For the test problem $H_0 : \boldsymbol{\theta} \in \Omega_0$ versus $H_1 : \boldsymbol{\theta} \in \Omega_1$ the maximum likelihood ratio test is defined by

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \ t(\mathbf{x}) > \kappa \\ 0 & \text{elsewhere} \end{cases} \quad \text{with} \quad t(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Omega} f_{\mathbf{X}}(\mathbf{x} \,|\, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Omega_0} f_{\mathbf{X}}(\mathbf{x} \,|\, \boldsymbol{\theta})},$$

where κ has to be determined so that $\beta_\phi(\boldsymbol{\theta}) \le \alpha \ \forall \boldsymbol{\theta} \in \Omega_0$.

Remark:

The maximum likelihood ratio test is sometimes defined in a slightly different form by

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \tilde{t}(\mathbf{x}) > \kappa \\ 0 & \text{elsewhere} \end{cases} \quad \text{with} \quad \tilde{t}(\mathbf{x}) = \frac{\sup\limits_{\boldsymbol{\theta}\in\Omega_1} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})}{\sup\limits_{\boldsymbol{\theta}\in\Omega_0} f_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})}.$$

For $\tilde{t}(\mathbf{x}) \geq 1$ the two definitions coincide due to

$$t(\mathbf{x}) = \frac{\sup\limits_{\boldsymbol{\theta}\in\Omega} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})}{\sup\limits_{\boldsymbol{\theta}\in\Omega_0} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})} = \max\left\{ \frac{\sup\limits_{\boldsymbol{\theta}\in\Omega_1} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})}{\sup\limits_{\boldsymbol{\theta}\in\Omega_0} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})}, 1 \right\} = \frac{\sup\limits_{\boldsymbol{\theta}\in\Omega_1} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})}{\sup\limits_{\boldsymbol{\theta}\in\Omega_0} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})} = \tilde{t}(\mathbf{x}).$$

The computation of the threshold κ can be very complicated. In some cases the following asymptotic result can be utilized.

Theorem:
Let Ω be a *k*-dimensional interval and Ω_0 an *l*-dimensional subinterval of Ω. Furthermore, suppose that certain regularity conditions are satisfied. Then

$$2\ln t(\mathbf{X}) \sim \chi^2_{k-l} \text{ under } H_0$$

holds asymptotically, i.e. for large *n*.

Hence, the threshold κ can be determined by

$$\kappa = \exp\left(\chi^2_{k-l,\alpha}/2\right).$$

Example: (maximum likelihood ratio test, F-test)

Again, we consider the detection problem

$$H_0 : \mathbf{X} = \mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \sigma_Z^2 \mathbf{I}) \text{ versus } H_1 : \mathbf{X} = \mathbf{H}\boldsymbol{\theta} + \mathbf{Z} \sim \mathcal{N}_n(\mathbf{H}\boldsymbol{\theta}, \sigma_Z^2 \mathbf{I})$$

$$\text{with } \Omega_0 = \left\{ (\mathbf{0}^T, \sigma_Z^2)^T : \sigma_Z^2 > 0 \right\} \text{ and}$$

$$\Omega_1 = \left\{ (\boldsymbol{\theta}^T, \sigma_Z^2)^T : \boldsymbol{\theta} \in \mathbb{R}^p \setminus \{\mathbf{0}\}, \sigma_Z^2 > 0 \right\}.$$

The maximization of the log-likelihood function under H_1

$$\ln f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta}, \sigma_Z^2) = -\frac{n}{2} \ln(2\pi\sigma_Z^2) - \frac{1}{2\sigma_Z^2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

provides the estimates, cf. Exercise 3.6-1,

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

and

$$\hat{\sigma}_Z^2 = \mathbf{x}^T \mathbf{P}^{\perp} \mathbf{x} / n \text{ with } \mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P}, \; \mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$$

as well as the supremum

$$\ln f_{\mathbf{x}}(\mathbf{x} \mid \hat{\boldsymbol{\theta}}, \hat{\sigma}_Z^2) = -\frac{n}{2}\left(\ln(2\pi/n) + \ln(\mathbf{x}^T \mathbf{P}^{\perp} \mathbf{x}) + 1\right).$$

The maximization of the log-likelihood function under H_0

$$\ln f_{\mathbf{x}}(\mathbf{x} \mid \mathbf{0}, \sigma_Z^2) = -\frac{n}{2}\ln(2\pi\sigma_Z^2) - \frac{1}{2\sigma_Z^2}\mathbf{x}^T\mathbf{x}$$

leads to the estimate $\hat{\sigma}_Z^2 = \mathbf{x}^T\mathbf{x}/n$ and the supremum

$$\ln f_{\mathbf{x}}(\mathbf{x} \mid \mathbf{0}, \hat{\sigma}_Z^2) = -\frac{n}{2}\left(\ln(2\pi/n) + \ln(\mathbf{x}^T\mathbf{x}) + 1\right).$$

Hence, the logarithm of the maximum likelihood ratio is

$$\ln t(\mathbf{x}) = \frac{n}{2}\ln\left(\frac{\mathbf{x}^T\mathbf{x}}{\mathbf{x}^T\mathbf{P}^{\perp}\mathbf{x}}\right) = \frac{n}{2}\ln\left(1 + \frac{\mathbf{x}^T\mathbf{P}\mathbf{x}}{\mathbf{x}^T\mathbf{P}^{\perp}\mathbf{x}}\right) = \frac{n}{2}\ln\left(1 + \frac{p}{n-p}v(\mathbf{x})\right).$$

The function $\ln t(\mathbf{x})$ increases monotonic with $v(\mathbf{x})$. Thus, a comparison of $\ln t(\mathbf{x})$ with a threshold is equivalent to a comparison of $v(\mathbf{x})$ with a threshold, i.e. the MLRT and F-test for p and $n-p$ degrees of freedom are equivalent in this case.

Moreover, since $\mathbf{x}^T\mathbf{P}^{\perp}\mathbf{x}/n$ provides an accurate estimate of σ_Z^2 and $(1+\alpha/n)^n \approx e^{\alpha}$ for large n the approximation

$$2\ln t(\mathbf{x}) = \ln\left(1+\frac{\mathbf{x}^T\mathbf{P}\mathbf{x}}{\mathbf{x}^T\mathbf{P}^{\perp}\mathbf{x}}\right)^n \approx \ln\left(1+\frac{\mathbf{x}^T\mathbf{P}\mathbf{x}}{n\sigma_Z^2}\right)^n \approx \frac{\mathbf{x}^T\mathbf{P}\mathbf{x}}{\sigma_Z^2}$$

can be deduced and the asymptotic distributional properties of $2\ln t(\mathbf{X})$ stated in the theorem can be established.

Exercise 4.3-9:

Detection of a deterministic signal with unknown amplitude η, unknown arrival time τ and unknown doppler shift in white Gaussian noise with unknown variance

4.3.6 Non Parametric Tests and Invariance

For composite hypotheses testing problems, where the number of unknown parameters is too large or the class of densities cannot be described sufficiently close, it is often inappropriate to try to design parametric tests in the aforementioned way.

Example: (Detection of a deterministic signal with unknown amplitude in independent and identically distributed noise with unknown distribution)

$$H_0 : \; X_k = Z_k, \; k = 1, \ldots, n \; \text{ with } \; \mathrm{E}Z_k = \mu_Z, \text{ where}$$

$$Z_k \text{ are i.i.d. with unknown density.}$$

$$H_1: \; X_k = \eta s_k + Z_k, \; k = 1,\ldots,n, \text{ where}$$

$$s_k \text{ are known with } \sum_{k=1}^{n} s_k = 0, \; \eta > 0 \text{ unknown.}$$

One says that the hypotheses are not parametric. Since the s_i ($i=1,\ldots,n$) are known the data can be reordered by

$$i_k \in \{1,\ldots,n\} \quad k = 1,\ldots,n \text{ with } i_k \neq i_l \text{ for } k \neq l$$

in such a way that

$$s_{i_1} \leq s_{i_2} \leq \ldots \leq s_{i_{n-1}} \leq s_{i_n},$$

where at least once a $<$ sign holds. Hence, X_{i_k} grows under H_1 stochastically with k and under H_0 it does not.

Furthermore, we suppose to know about the receiver system that the amplitude range of the received signal does not lead to saturations, i.e. it possesses an unknown but strictly monotonic characteristic.

The test problem remains invariant under this aspect, if the observations are consistently amplified by a transformation from the group of strictly monotonic growing functions which is denoted by

$$y_i = g(x_i) \quad i = 1,\ldots,n.$$

Maximally invariant are obviously the size relations of the data among themselves, i.e. the ranks r_i of the data. The r_i $(i=1,\ldots,n)$ specify the number of observations x_j $(j=1,$

…,n) that satisfy the inequality

$$x_j \leq x_i .$$

Thus, if e.g. $r_i = 3$ then x_i is the third smallest element of the set $\{x_1,\ldots,x_i,\ldots,x_n\}$.

Invariant tests depend thus only over the ranks r_i on the data and are therefore called rank tests.

A rank test is a non-parametric test that uses for instance the test function

$$t(\mathbf{x}) = \sum\nolimits_{i=1}^{n} s_i \, h(r_i),$$

where $h(i)$ denotes a suitable function of the numbers $i = 1,\ldots,n$.

Under certain regularity conditions such as normality under H_1 one can show that a rank test provides a locally uniformly most powerful invariant test for small η.

Moreover, it can be shown that this rank test applied to the data of Exercise 4.3-5 achieves asymptotically, i.e. for large n, the same efficiency as the t-test utilized there and that the t-test can even fail compared to the rank test in the more general situation described in Exercise 4.3-5.

However, in certain situations, e.g. for dependent data, the design of an appropriate rank test and the evaluation of its efficiency is a difficult and partly still an unsolved problem.

References to Chapter 4

[1] M. Fisz, *Probability Theory and Mathematical Statistics*, Krieger Publishing Company, 1980

[2] C.W. Helstrom, *Elements of Signal Detection and Estimation,* Prentice Hall, 1994

[3] S. Kay, *Fundamentals of Statistical Signal Processing, Vol. 2: Detection Theory*, Prentice Hall, 1998

[4] E.L. Lehmann and J.P. Romano, *Testing Statistical Hypotheses*, Springer, 2006

[5] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, 1994

[6] C.R. Rao, *Linear Statistical Inference and Its Application*, John Wiley, 1973

[7] L.L. Scherf, *Statistical Signal Processing*, Addison Wesley, 1990