10.0 Hypothesis Testing and Detection Theory

10.1 Hypothesis Testing and the Neyman-Pearson Proposal

Suppose we have n samples of the output of a detector. We need to decide on two hypothesis:

$$H_0 : \text{The signal is not present}$$
$$H_1 : \text{The signal is present}$$

We can view this as rejecting or failing to reject H_0.

Definition: A *type I error* is the error that occurs if we reject H_0 when we should have not rejected it.

Definition: A *type II error* is the error that occurs if we fail to reject H_0 when we should have rejected it.

It is the usual convention in statistics to label the hypotheis so as to make the type I error the more serious of the two errors if we have some control over the hypothesis setup.

In our detection example a type I error occurs when we think a signal is present when it is not. This is also called a *false alarm*. We can design our detection criteria (e.g., choose thresholds) to make the probability of a type I error as small as desired (at the cost of usually increasing the probability of a type II error).

The concept of controlling the type I error is the Neyman-Pearson proposal.

Definition: The *power* of a test is the probability of rejecting H_0 when H_1 is true.

So the power is $1 - P(\text{type II error})$ and is the probability that the test will "detect" when H_1 is true.

10.2 Detection Theory

Let $\mathbf{X} \in \mathbf{R}^n$ be a random vector. We have two hypothesis:

$$\begin{array}{l} H_0 : \mathbf{X} \text{ has distribution } F_0(\mathbf{x}) \\ H_1 : \mathbf{X} \text{ has distribution } F_1(\mathbf{x}) \end{array}$$

Our decision rule is

$$S : \ \mathbf{R}^n \to \{0, 1\}$$

where

$$S(\mathbf{X}) = \begin{cases} 0, & \mathbf{X} \in S_0 \\ 1, & \mathbf{X} \in S_1 \end{cases}$$

We have two error probabilities:

$$P_{e0} = P(S(\mathbf{X}) = 1 | H_0 \text{ is true}) \rightarrow \text{ type I error}$$

$$P_{e1} = P(S(\mathbf{X}) = 0 | H_1 \text{ is true}) \rightarrow \text{ type II error}$$

Define

$$P_e = P_{e0} + P_{e1}.$$

Using a ratio of density functions we would reject H_0 if

$$L(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} \geq 1.$$

$L(\mathbf{x})$ is called the *likelihood ratio.*

Our decision rule then is to partition $\mathbf{R}^n$ as

$$S_0 = \{\mathbf{x} : \ f_0(\mathbf{x}) > f_1(\mathbf{x})\}$$

$$S_1 = \{\mathbf{x} : \ f_0(\mathbf{x}) \leq f_1(\mathbf{x})\}$$

Thus

$$P_{e0} = \int \cdots \int_{S_1} f_0(\mathbf{x}) d\mathbf{x},$$

$$P_{e1} = \int \cdots \int_{S_0} f_1(\mathbf{x}) d\mathbf{x},$$

$$P_s = \int \cdots \int \min \{f_0(\mathbf{x}), \ f_1(\mathbf{x})\} \, d\mathbf{x}.$$

<u>Neyman-Pearson Lemma:</u> Say we have two hypothesis:

$$H_0 : \mathbf{X} \text{ has distribution } F_0(\mathbf{x})$$
$$H_1 : \mathbf{X} \text{ has distribution } F_1(\mathbf{x})$$

We test for H_0 so we either reject or fail to reject H_0. Let

$$A = \{\mathbf{x} : \; g(\mathbf{x}) \geq k\}$$

where k is chosen so that $P_0(A) = \alpha$ (type I error). Then for all B with $P_0(B) \leq \alpha$

$$\int_B g dF_0 \leq \int_A g dF_0.$$

Proof:

$$P_0(B) = P_0(B \cap A) + P_0(B \cap A^c) \leq P_0(A) = P_0(A \cap B) + P_0(A \cap B^c)$$

so

$$P_0(B \cap A^c) \leq P_0(A \cap B^c)$$

thus

$$\int_{B \cap A^c} g dF_0 \leq k P_0(B \cap A^c) \leq k P_0(A \cap B^c) \leq \int_{A \cap B^c} g dF_0.$$

We now add $\int_{A \cap B} g dF_0$ to both sides of this last equation to get

$$\int_{B \cap A} g dF_0 + \int_{B \cap A^c} g dF_0 \leq \int_{A \cap B} g dF_0 + \int_{A \cap B^c} g dF_0$$

which implies

$$\int_B g dF_0 \leq \int_A g dF_0$$

which completes the proof.

Suppose we receive a vector $\mathbf{X}$ with distribution function F($\mathbf{x}$). Then

$$H_0 : F = F_0$$
$$H_1 : F = F_1$$

Consider

$$L(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}.$$

We decide to reject H_0 if $L(\mathbf{x}) \geq k$. Let

$$A = \{\mathbf{x}: \ L(\mathbf{x}) \geq k\}\,, \ P_0(A) = \alpha, \quad L = \frac{dF1}{dF0} \text{ (Radon Nikodyn Derivative).}$$

The power of A is

$$P_1(A) = \int_A \left(\frac{dF_1}{dF_0}\right) dF_0 = \int_A L dF_0 \geq \int_B L dF_0 = \int_B \left(\frac{dF_1}{dF_0}\right) dF_0 = P_1(B).$$

So this detection scheme gives us the greatest power. This is the Neyman-Pearson Lemma. Thus, if we choose a threshold to limit a type I error, then using a likelihood ratio test for detection maximizes the power of the test (the probability of rejecting H_0 when H_1 is true) over any other detection scheme with the same or lower type I error.

10.2.1 Correlation Detection in White Noise

Suppose the received vector is

$$\mathbf{x} = \mathbf{s}_i + \mathbf{W}(u), \ i = 0, 1$$

where $E[\mathbf{W}(u)] = \mathbf{0}$ and $\mathbf{K_W} = \sigma^2\mathbf{I}$. Here $\mathbf{s}_i$ is known. Then we fail to reject H_0 if and only if

$$||\mathbf{x} - \mathbf{s}_0||^2 \leq ||\mathbf{x} - \mathbf{s}_1||^2$$

or

$$||\mathbf{x}||^2 - 2\text{Re} < \mathbf{x}, \ \mathbf{s}_0 > + ||\mathbf{s}_0||^2 \leq ||\mathbf{x}||^2 - 2\text{Re} < \mathbf{x}, \ \mathbf{s}_1 > + ||\mathbf{s}_1||^2$$

or

$$-2\text{Re} < \mathbf{x}, \ \mathbf{s}_0 > + ||\mathbf{s}_0||^2 \leq -2\text{Re} < \mathbf{x}, \ \mathbf{s}_1 > + ||\mathbf{s}_1||^2$$

or

$$2\text{Re} < \mathbf{x}, \ \mathbf{s}_1 > -2\text{Re} < \mathbf{x}, \ \mathbf{s}_0 > \leq ||\mathbf{s}_1||^2 - ||\mathbf{s}_0||^2$$

or

$$\text{Re} < \mathbf{x}, \ \mathbf{s}_1 - \mathbf{s}_0 > \leq \frac{||\mathbf{s}_1||^2 - ||\mathbf{s}_0||^2}{2}.$$

If the vectors are real then we fail to reject H_0 if and only if

$$< \mathbf{x}, \ \mathbf{s}_1 - \mathbf{s}_0 > \leq \frac{||\mathbf{s}_1||^2 - ||\mathbf{s}_0||^2}{2}.$$

Recall, P_{e0} is the probability of rejecting H_0 when H_0 is true. So

$$P_{e0} = P\left(< \mathbf{x},\ \mathbf{s}_1 - \mathbf{s}_0 > > \frac{||\mathbf{s}_1||^2 - ||\mathbf{s}_0||^2}{2} \Big| \mathbf{x} = \mathbf{s}_0 + \mathbf{w}\right)$$

$$= P\left(< \mathbf{s}_0 + \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 > > \frac{||\mathbf{s}_1||^2 - ||\mathbf{s}_0||^2}{2}\right)$$

$$= P\left(< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 > > \frac{||\mathbf{s}_1||^2 - ||\mathbf{s}_0||^2 - 2 < \mathbf{s}_0,\ \mathbf{s}_1 - \mathbf{s}_0 >}{2}\right)$$

$$= P\left(< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 > > \frac{||\mathbf{s}_1||^2 + ||\mathbf{s}_0||^2 - 2 < \mathbf{s}_0,\ \mathbf{s}_1 >}{2}\right)$$

$$= P\left(< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 > > \frac{||\mathbf{s}_1 - \mathbf{s}_0||^2}{2}\right)$$

$$\leq P\left(|< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 >| \ > \frac{||\mathbf{s}_1 - \mathbf{s}_0||^2}{2}\right)$$

$$\leq \frac{Var(< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 >)}{\left(\frac{1}{2}||\mathbf{s}_1 - \mathbf{s}_0||^2\right)^2}.$$

Note:

$$E[< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 >] = 0$$

and

$$E\left[(< \mathbf{w},\ \mathbf{s}_1 - \mathbf{s}_0 >)^2\right] = E\left[\left(\mathbf{w}^T(\mathbf{s}_1 - \mathbf{s}_0)\right)^T \left(\mathbf{w}^T(\mathbf{s}_1 - \mathbf{s}_0)\right)\right] = \sigma^2||\mathbf{s}_1 - \mathbf{s}_0||^2$$

10.2.2 Correlation Detection in Colored Noise

Here

$$H_0: \ \mathbf{x} = \mathbf{s}_0 + \mathbf{N}$$
$$H_1: \ \mathbf{x} = \mathbf{s}_1 + \mathbf{N}$$

$\mathbf{N}$ has mean zero and covariance $\mathbf{K_N}$. If $\mathbf{K_N}^{-1}$ exists then there exists $\mathbf{G}$ such that $\mathbf{G}^\dagger\mathbf{G} = \mathbf{K_N}^{-1}$ [$\mathbf{K_N} = \mathbf{H}\mathbf{H}^\dagger$ so set $\mathbf{G} = \mathbf{H}^{-1}$].

Our new hypothesis is

$$H_0': \ \mathbf{G}\mathbf{x} = \mathbf{G}\mathbf{s}_0 + \mathbf{G}\mathbf{N}$$
$$H_1': \ \mathbf{G}\mathbf{x} = \mathbf{G}\mathbf{s}_1 + \mathbf{G}\mathbf{N}$$

Note that $\mathbf{W} = \mathbf{G}\mathbf{N}$ is a white vector:

$$\mathbf{K_W} = \mathbf{G}\mathbf{K_N}\mathbf{G}^\dagger = \mathbf{I}.$$

$$S_0 = \{\mathbf{x} : \ ||\mathbf{G}(\mathbf{x} - \mathbf{s}_0)|| \le ||\mathbf{G}\mathbf{x} - \mathbf{s}_1||\}$$

$$S_1 = \{\mathbf{x} : \ ||\mathbf{G}(\mathbf{x} - \mathbf{s}_0)|| > ||\mathbf{G}\mathbf{x} - \mathbf{s}_1||\}$$

The correlation detector becomes

$$< \mathbf{G}\mathbf{x}, \ \mathbf{G}(\mathbf{s}_1 - \mathbf{s}_0) > \le \frac{||\mathbf{G}\mathbf{s}_1||^2 - ||\mathbf{G}\mathbf{s}_0||^2}{2}.$$

Now

$$\begin{aligned} < \mathbf{G}\mathbf{x}, \ \mathbf{G}(\mathbf{s}_1 - \mathbf{s}_0) > &= \left(\mathbf{G}(\mathbf{s}_1 - \mathbf{s}_0)\right)^\dagger \mathbf{G}\mathbf{x} \\ &= (\mathbf{s}_1 - \mathbf{s}_0)^\dagger \mathbf{G}^\dagger \mathbf{G}\mathbf{x} \\ &= (\mathbf{s}_1 - \mathbf{s}_0)^\dagger \mathbf{K}_\mathbf{N}^{-1}\mathbf{x} \\ &=< \mathbf{x}, \mathbf{K}_\mathbf{N}^{-1}(\mathbf{s}_1 - \mathbf{s}_0) > . \end{aligned}$$

The threshold T is

$$\begin{aligned} T &= \frac{||\mathbf{G}\mathbf{s}_1||^2 - ||\mathbf{G}\mathbf{s}_0||^2}{2} \\ &= \frac{(\mathbf{G}\mathbf{s}_1)^\dagger (\mathbf{G}\mathbf{s}_1) - (\mathbf{G}\mathbf{s}_0)^\dagger (\mathbf{G}\mathbf{s}_0)}{2} \\ &= \frac{\mathbf{s}_1^\dagger \mathbf{G}^\dagger \mathbf{G}\mathbf{s}_1 - \mathbf{s}_0^\dagger \mathbf{G}^\dagger \mathbf{G}\mathbf{s}_0}{2} \\ &= \frac{\mathbf{s}_1^\dagger \mathbf{K}_\mathbf{N}^{-1}\mathbf{s}_1 - \mathbf{s}_0^\dagger \mathbf{K}_\mathbf{N}^{-1}\mathbf{s}_0}{2}. \end{aligned}$$

10.2.3 Perfect Decisions

Here we can make decisions with no error. In this case $\mathbf{K_N}$ is singular but it has a zero eigenvalue. Let $\mathbf{e}_0$ be the corresponding eigenvector for $\lambda = 0$. Let $\mathbf{N} = \mathbf{E}\mathbf{W}$ and let $N_{e_0} =< \mathbf{N}, \ \mathbf{e}_0 >= \mathbf{e}_0^\dagger \mathbf{N}$. Then $E[N_{e_0}] = 0$ and $Var[N_{e_0}] = 0$ since we have $\lambda = 0$. So, $< \mathbf{N}, \ \mathbf{e}_0 >= 0$ with probability 1.

$$\begin{aligned} \tilde{H}_0 &: < \mathbf{x}, \ \mathbf{e}_0 >=< \mathbf{s}_0, \ \mathbf{e}_0 > + < \mathbf{N}, \ \mathbf{e}_0 >=< \mathbf{s}_0, \ \mathbf{e}_0 > \\ \tilde{H}_0 &: < \mathbf{x}, \ \mathbf{e}_0 >=< \mathbf{s}_1, \ \mathbf{e}_0 > + < \mathbf{N}, \ \mathbf{e}_0 >=< \mathbf{s}_1, \ \mathbf{e}_0 > \end{aligned}$$

Thus, since we have no noise in the decision rule it is perfect as long as $< \mathbf{s}_0, \ \mathbf{e}_0 > \neq < \mathbf{s}_1, \ \mathbf{e}_0 >$.

10.3 Some Practical Detection Systems

Integration Filtering

Say we have N independent samples of the envelope of a sinusoid with white Gaussian noise. The input to the envelope detector is of the form

$$X(k) = m(k) + W(k)$$

where $m(k)$ represents the signal component and $W(k)$ is the noise component. If no signal is present the output of the envelope detector has Rayleigh statistics. We can find a threshold T_0 so that the probability of exceeding this threshold at the output of the envelope detector is some fixed quantity called the probability of false alarm (P_{fa}).

Often the signal has in-phase and quadrature components, i.e.,

$$X(k) = X_I(k) + iX_Q(k), \ i = \sqrt{-1}$$

so the envelope is

$$|X(k)| = \sqrt{X_I^2(k) + X_Q^2(k)}.$$

The output of the detector, call it R, with noise only present has density (with $n = N$)

$$f_R(r) = \frac{r^{n-1}}{2^{(n-2)/2}\sigma^n\Gamma(n/2)} e^{-r^2/2\sigma^2}, \quad r \geq 0.$$

$\Gamma(p)$ is the gamma function

$$\Gamma(p) = \int_0^\infty t^{p-1}e^{-t}dt, \quad p > 0.$$

If p is an integer, $\Gamma(p) = (p-1)!$. n is the degrees of freedom and σ^2 is the variance of the noise.

If $n = 2$ we get

$$f_R(r) = \frac{r}{\sigma^2}e^{-r^2/2\sigma^2}, \quad r \geq 0.$$

In this case

$$P_{fa} = \int_{T_0}^\infty f_R(r)dr = e^{-T_0^2/2\sigma^2}$$

which implies

$$T_0 = \sqrt{-2\sigma^2 \ln(P_{fa})}.$$

T_0 is thus used to control type I errors.

If $n = 2m$ is even we get

$$F_R(r) = 1 - e^{-r^2/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{r^2}{2\sigma^2} \right)^k, \; r \geq 0.$$

With a signal present we have Rician statistics

$$f_R(r) = \frac{r^{n/2}}{\sigma^2 s^{(n-2)/2}} e^{-(r^2+s^2)/2\sigma^2} I_{n/2-1} \left(\frac{rs}{\sigma^2} \right), \; r \geq 0$$

where s^2 is the envelope squared and $I_{n/2-1}$ is the $(n/2-1)^{th}$-order modified Bessel function of the first kind. With $n = 2$ we get

$$f_R(r) = \frac{r}{\sigma^2} e^{-(r^2+s^2)/2\sigma^2} I_0 \left(\frac{rs}{\sigma^2} \right), \; r \geq 0.$$

So we can find the required signal-to-noise ratio ($SNR = s^2/2\sigma^2$) to give the desired probability of detection, P_d. For $n = 2$

$$P_d = \int_{T_0}^{\infty} \frac{r}{\sigma^2} e^{-(r^2+s^2)/2\sigma^2} I_0 \left(\frac{rs}{\sigma^2} \right) dr$$

which can be evaluated using

$$F_R(r) = 1 - Q_1 \left(\frac{s}{\sigma}, \frac{r}{\sigma} \right)$$

where,

$$\begin{aligned} Q_1(a, b) &= \int_b^{\infty} x e^{-(x^2+a^2)/2} I_0(ax) dx \\ &= e^{-(a^2+b^2)/2} \sum_{k=0}^{\infty} \left(\frac{a}{b} \right)^k I_k(ab). \end{aligned}$$

More generally, we have (with $m = n/2$)

$$P_d = \int_{T_0}^{\infty} \frac{r^{n/2}}{\sigma^2 s^{(n-2)/2}} e^{-(r^2+s^2)/2\sigma^2} I_{n/2-1} \left(\frac{rs}{\sigma^2} \right) dr$$

which can be evaluated using

$$F_R(r) = 1 - Q_m\left(\frac{s}{\sigma}, \frac{r}{\sigma}\right)$$

where,

$$\begin{aligned} Q_m(a,b) &= \int_b^\infty x\left(\frac{x}{a}\right) e^{-(x^2+a^2)/2} I_{m-1}(ax)dx \\ &= Q_1(a,b) + e^{-(a^2+b^2)/2} \sum_{k=1}^{m-1} \left(\frac{b}{a}\right)^k I_k(ab). \end{aligned}$$

The above results can be found in Proakis, *Digital Communications*.

Robertson (1967) produced curves showing the required SNR vs. P_d and P_{fa}. A simple approximation formula exists for calculating the required SNR due to Albersheim (1981):

$$SNR \approx -5\log_{10} N + \left(6.2 + \frac{4.54}{\sqrt{N + 0.44}}\right) \cdot \log_{10}(A + 0.12AB + 1.7B) \quad \text{(dB)}$$

where N is the number of independent post-detection samples that are integrated,

$$A = \ln\left(\frac{0.62}{P_{fa}}\right), \quad B = \ln\left(\frac{P_d}{1 - P_d}\right).$$

The stated accuracy of this formula is within 0.2 dB for $P_d = 0.1$ to 0.9, $P_{fa} = 10^{-3}$ to 10^{-7} and $N = 1$ to 8096. However, I have found this is not the case at the endpoints where P_d is the lowest and P_{fa} is the highest. For example, if $P_{fa} = 10^{-3}$ you need $P_d \geq 0.4$ for the formula to apply. Also, note this SNR is the SNR in the detection bandwidth and not the input SNR.

We will now look at an even simpler way of doing detection. Instead of using the integrator we will use M of N filtering for detection.

M of N Filtering

The idea is to make decisions on a sample basis instead of integrating over all the samples. We need to determine thresholds and probabilities on an independent sample basis. If M or more samples exceed this threshold then we say a signal is present, otherwise we decide only noise is present. Let

$P_{d,s}$ and $P_{fa,s}$ be the probability of detection and probability of false alarm, respectively, on a sample basis. Then

$$P_d = \sum_{K=M}^{N} \binom{N}{K} P_{d,s}^K (1 - P_{d,s})^{N-K},$$

$$P_{fa} = \sum_{K=M}^{N} \binom{N}{K} P_{fa,s}^K (1 - P_{fa,s})^{N-K}.$$

Usually we cannot use Albersheim's equation here since $P_{fa,s}$ can be relatively high on a sample basis and still yield an overall low $P_{fa,s}$. We have to use numerical calculations to compute $P_{d,s}$ and $P_{fa,s}$ to give an overall P_d and P_{fa} and set the thresholds accordingly. For many cases of interest M of N filtering results in about 1 dB loss in performance relative to integration.