D3: Statistical Dessision Theory I

Consider a measurement $x[0]$ that is either $w[0]$ or $s[0]+w[0]$, where $s[0]=1$ and $w[0] \sim \mathcal{N}(0,1)$. Given $x[0]$, deciding weather we have $s[0]=1$ or the other case is a simple hypothesis testing problem.

$$\mathcal{H}_0 : x[0] = w[0]$$
$$\mathcal{H}_1 : x[0] = s[0] + w[0]$$

In this example we have a simple "signal detection in noise" situation.

Miss: Deciding $\mathcal{H}_0$ when $\mathcal{H}_1$ is true.

False alarm: Deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true.

Detection: Deciding $\mathcal{H}_1$ when $\mathcal{H}_1$ is true.

Let $P_{FA} = P(\mathcal{H}_1 ; \mathcal{H}_0)$, $P_D = P(\mathcal{H}_1 ; \mathcal{H}_1)$. We have

$$P_{FA} = Pr\{x[0] > \gamma ; \mathcal{H}_0\} = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = Q(\gamma)$$

$$P_D = Pr\{x[0] > \gamma ; \mathcal{H}_1\} = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(t-1)^2/2} dt = Q(\gamma-1)$$

In general a sequence $\{x[0], x[1], --, x[N-1]\}$ is observed. Let $\mathbf{x} = [x[0], --, x[N-1]]^T$ be the data vector. Then the critical region (region for x where decision will be $\mathcal{H}_1$) is

$$R_1 = \{x : \text{decide } \mathcal{H}_1 \text{ or reject } \mathcal{H}_0\}.$$

Here $R_1 \subset \mathbb{R}^N$ and $R_0 \cup R_1 = \mathbb{R}^n$ with $R_0 = \mathbb{R}^n - R_1$ denoting the region in which we decide $\mathcal{H}_0$. Clearly,

$$P_{FA} = \int_{R_1} p(x;\mathcal{H}_0)\,dx = \alpha \quad \text{and} \quad P_D = \int_{R_1} p(x;\mathcal{H}_1)\,dx = 1-\beta$$

and α is called the significance level (or size) of the test, $(1-\beta)$ is called the power of the test.

<u>Thm 3.1</u>: (Neyman-Pearson)

To maximize P_D for a given $P_{FA} = \alpha$, decide $\mathcal{H}_1$ if

$$L(x) = \frac{p(x;\mathcal{H}_1)}{p(x;\mathcal{H}_0)} > \gamma, \text{ where } \gamma \text{ is found from } P_{FA} = \int_{\{x : L(x) > \gamma\}} p(x;\mathcal{H}_0)\,dx = \alpha.$$

Here, $L(x)$ is called the likelihood ratio and the decision process in the theorem is the likelihood ratio test (LRT).

Proof: $F = P_D + \lambda(P_{FA} - \alpha) = \int_{R_1} p(x;\mathcal{H}_1)\,dx + \lambda\left(\int_{R_1} p(x;\mathcal{H}_0)\,dx - \alpha\right)$

Lagrange multiplier

$$= \int_{R_1} \left[p(x;\mathcal{H}_1) + \lambda p(x;\mathcal{H}_0)\right] dx - \lambda\alpha$$

To maximize F, we should include x in R_1 if the integrand is positive, i.e. if $p(x;\mathcal{H}_1) + \lambda p(x;\mathcal{H}_0) > 0$. This yields $\frac{p(x;\mathcal{H}_1)}{p(x;\mathcal{H}_0)} > -\lambda$ should lead to deciding $\mathcal{H}_1$ for x. The Lagrange multiplier is found using the constraint (and by letting $\gamma = -\lambda$ we get the formula in the theorem). We must have $\lambda < 0$ or equivalently $\underline{\underline{\gamma > 0}}$.

Ex] DC Level in WGN

$\mathcal{H}_0: \; x[n] = w[n] \quad n = 0, 1, \dots, (N-1)$ Here, $s[n] = A$ for

$\mathcal{H}_1: \; x[n] = A + w[n] \quad n = 0, 1, \dots, (N-1)$ $A > 0$ and $w[n] \sim$ WGN with variance σ^2.

Specifically, we have $\mathcal{H}_0: x \sim \mathcal{N}(0, \sigma^2 I)$ and $\mathcal{H}_1: x \sim \mathcal{N}(A\mathbf{1}, \sigma^2 I)$.

Equivalently, $\mathcal{H}_0: \mu = 0$ and $\mathcal{H}_1: \mu = A\mathbf{1}$.

According to NP, we should decide if

$$\frac{(2\pi\sigma^2)^{-N/2} \, e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2}}{(2\pi\sigma^2)^{-N/2} \, e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]}} > \gamma$$

or equivalently, taking the ln of both sides,

$$-\frac{1}{2\sigma^2}\left[-2A\sum_{n=0}^{N-1} x[n] + NA^2\right] > \ln\gamma$$

$$\Leftrightarrow \quad \frac{A}{\sigma^2}\sum_{n=0}^{N-1} x[n] > \ln\gamma + \frac{NA^2}{2\sigma^2}$$

(since $A>0$) $\Leftrightarrow \quad \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\ln\gamma + \frac{A}{2} \triangleq \gamma'$ (compare the sample mean to a threshold γ')

By changing γ', we can control the trade off between P_{FA} and P_D; as γ' increases P_{FA} decreases but P_D also decreases.

Notice that the test statistic $T(x) = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ is Gaussian under each hypothesis. Specifically, $E[T(x); \mathcal{H}_k] = A\overset{\swarrow \text{impul}}{\delta_k}$ and $\operatorname{var}[T(x); \mathcal{H}_k] = \frac{\sigma^2}{N}$ for $k = 0, 1$. Hence

$$T(x) \sim \begin{cases} \mathcal{N}(0, \sigma^2/N) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(A, \sigma^2/N) & \text{under } \mathcal{H}_1 \end{cases}$$

Then $P_{FA} = Pr\{T(x) > \gamma' ; \mathcal{H}_0\} = Q\left(\frac{\gamma'}{\sigma/\sqrt{N}}\right)$

$$P_D = Pr\{T(x) > \gamma' ; \mathcal{H}_1\} = Q\left(\frac{\gamma' - A}{\sigma/\sqrt{N}}\right)$$

Since Q is monotonically strictly decreasing, it is invertible.

Therefore, $\gamma' = \frac{\sigma}{\sqrt{N}} Q^{-1}(P_{FA})$ and

$$P_D = Q\left(Q^{-1}(P_{FA}) - \frac{A}{\sigma}\sqrt{N}\right)$$

In general, if we have $T \sim \begin{cases} \mathcal{N}(\mu_0, \sigma^2) \text{ under } \mathcal{H}_0 \\ \mathcal{N}(\mu_1, \sigma^2) \text{ under } \mathcal{H}_1 \end{cases}$, we will have for $d^2 = \frac{[E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0)]^2}{var(T;\mathcal{H}_0)} = \frac{(\mu_1 - \mu_0)^2}{\sigma^2}$

and $P_{FA} = Q\left(\frac{\gamma' - \mu_0}{\sigma}\right)$, $P_D = Q\left(Q^{-1}(P_{FA}) - d\right)$. This is a mean-shifted Gauss-Gauss problem.

Ex] Change in Variance.

We observe $x[n]$ for $n = 0, \dots, (N-1)$ which are iid with $x[n] \sim \mathcal{N}(0, \sigma_0^2)$ under $\mathcal{H}_0$ and $x[n] \sim \mathcal{N}(0, \sigma_1^2)$ under $\mathcal{H}_1$. Assume $\sigma_1^2 > \sigma_0^2$. Then the NP test to decide $\mathcal{H}_1$ is

$$\frac{(2\pi\sigma_1^2)^{-N/2} e^{-\frac{1}{2\sigma_1^2}\sum_{n=0}^{N-1} x^2[n]}}{(2\pi\sigma_0^2)^{-N/2} e^{-\frac{1}{2\sigma_0^2}\sum_{n=0}^{N-1} x^2[n]}} > \gamma$$ } taking ln of both sides

$$-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{n=0}^{N-1} x^2[n] > \ln\gamma + \frac{N}{2}\ln\frac{\sigma_1^2}{\sigma_0^2}$$

$$\Leftrightarrow \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] > \gamma' = \frac{(2/N)\ln\gamma + \ln(\sigma_1^2/\sigma_0^2)}{(1/\sigma_0^2)-(1/\sigma_1^2)} .$$

We decide $\mathcal{H}_1$ if the sample estimate of power is sufficiently large.

In general, assume that we observe $x = [x[0], \ldots, x[N-1]]^T$ from a pdf parameterized by θ; $p(x;\theta)$. We wish to test for the value of θ as

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{and} \quad \mathcal{H}_1 : \theta = \theta_1 .$$

If a sufficient statistic exists for θ, then by the Neyman-Fisher factorization theorem, we can express the pdf as $p(x;\theta) = g(T(x), \theta)\, h(x)$, where $T(x)$ is a sufficient statistic for θ. Then the NP test is

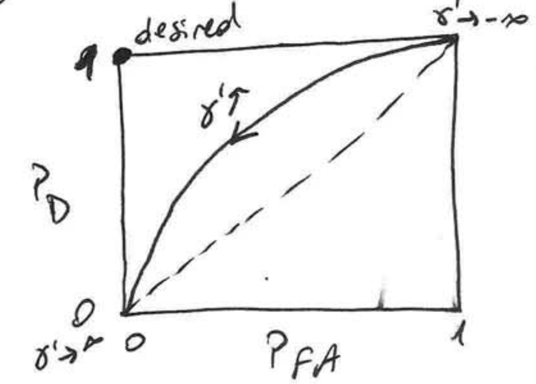
$$\frac{p(x;\theta_1)}{p(x;\theta_0)} = \frac{g(T(x), \theta_1)}{g(T(x), \theta_0)} > \gamma .$$

$\therefore$ The test will depend on the data only through $T(x)$.

Receiver Operating Characteristics (ROC)

Recall that in the two examples above we had decision rules in the form $T(x) > \gamma' \Rightarrow$ decide $\mathcal{H}_1$ and $P_{FA}(\gamma')$ and $P_D(\gamma')$ presented a trade-off to choose γ'.

The curve $P_D(\gamma')$ vs $P_{FA}(\gamma')$ traced by taking γ' from $-\infty$ to ∞ is called the ROC curve.



We want $P_{FA}(\gamma')$ to be 0 and $P_D(\gamma')$ to be 1. We can only choose a solution on the ROC curve by setting γ' to a value.

Irrelevant Data: If we observe two data vectors x_1 and x_2, then the LRT from the NP theorem is

$$L(x_1, x_2) = \frac{p(x_1, x_2; \mathcal{H}_1)}{p(x_1, x_2; \mathcal{H}_0)} = \frac{p(x_2 | x_1; \mathcal{H}_1)\, p(x_1; \mathcal{H}_1)}{p(x_2 | x_1; \mathcal{H}_0)\, p(x_1; \mathcal{H}_0)}$$

Then, if $p(x_2|x_1; \mathcal{H}_1) = p(x_2 | x_1; \mathcal{H}_0)$, then $L(x_1, x_2) = L(x_1)$ and x_2 is irrelevant to the detection problem. The condition here means the distribution of x_2 given x_1 does not depend on the hypothesis (e.g. for signal detection in noise, x_1 is for a segment with signal and x_2 is for a segment without; just noise ~~)~~ and noise is not correlated in time).

As a special case, if $x_1 \perp\!\!\!\perp x_2$ under either hypothesis then $p(x_2; \mathcal{H}_1) = p(x_2; \mathcal{H}_0)$ makes the cancellation above occur.

Minimum Probability of Error

In some cases one can assign prior probabilities to each hypothesis. Then, in the Bayesian paradigm, we can define the probability of error as

$$P_e = P(\mathcal{H}_0|\mathcal{H}_1)P(\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)P(\mathcal{H}_0)$$

The Bayesian detector with minimum Pe is given by

$$\frac{p(x|\mathcal{H}_1)p(\mathcal{H}_1)}{p(x|\mathcal{H}_0)p(\mathcal{H}_0)} > 1 \iff \frac{p(x|\mathcal{H}_1)}{p(x|\mathcal{H}_0)} > \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \triangleq \gamma .$$

Bayes Risk: Let C_{ij} be the cost if we decide $\mathcal{H}_i$ but $\mathcal{H}_j$ is true. For example, we would probably want $C_{10} < C_{01}$ in many applications (though not always), where too many false alarms would render the detector "useless".

The expected cost (Bayes risk) is defined as

$$R = E(c) = \sum_{i=0}^{1}\sum_{j=0}^{1} c_{ij}\, p(\mathcal{H}_i|\mathcal{H}_j)\,p(\mathcal{H}_j) .$$

While mostly we might choose $C_{ii} = 0$ (no error $\Rightarrow$ no cost), this may not be the case always. For instance the cost of the appropriate action if $\mathcal{H}_i$ is detected (correctly) might be a consideration in our strategy when deciding $\mathcal{H}_i$ or not.

Then, the optimal decision rule is to choose $\mathcal{H}_1$ when

$$(C_{10} - C_{00}) P(\mathcal{H}_0) p(x|\mathcal{H}_0) < (C_{01} - C_{11}) P(\mathcal{H}_1) p(x|\mathcal{H}_1)$$

Proof: Let $R_1 = \{x : \text{decide } \mathcal{H}_1\}$ and R_0 is its complement.

Then
$$R = C_{00} P(\mathcal{H}_0) \int_{R_0} p(x|\mathcal{H}_0)\,dx + C_{01} P(\mathcal{H}_1) \int_{R_0} p(x|\mathcal{H}_1)\,dx$$
$$+ C_{10} P(\mathcal{H}_0) \int_{R_1} p(x|\mathcal{H}_0)\,dx + C_{11} P(\mathcal{H}_1) \int_{R_1} p(x|\mathcal{H}_1)\,dx$$

Since $\int_{R_0} p(x|\mathcal{H}_i)\,dx = 1 - \int_{R_1} p(x|\mathcal{H}_i)\,dx$, we have

$$R = C_{01} P(\mathcal{H}_1) + C_{00} P(\mathcal{H}_0) + \int_{R_1} \Big[\big(C_{10} P(\mathcal{H}_0) - C_{00} P(\mathcal{H}_0)\big) p(x|\mathcal{H}_0) + \big(C_{11} P(\mathcal{H}_1) - C_{01} P(\mathcal{H}_1)\big) p(x|\mathcal{H}_1) \Big] dx$$

We include x in R_1 only if the integrand is ~~positive~~ negative.

Assuming that $C_{10} > C_{00}$ and $C_{01} > C_{11}$, we could express the optimal Bayesian (minimum risk) decision rule as

$$\frac{p(x|\mathcal{H}_1)}{p(x|\mathcal{H}_0)} > \frac{(C_{10} - C_{00}) P(\mathcal{H}_0)}{(C_{01} - C_{11}) P(\mathcal{H}_1)} = \gamma$$

Also note that if $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 1$, then $R = P_e$; so probability of error is a special case.

Once again, the conditional LRT is employed with a suitable threshold to make a decision.

Multiple Hypothesis Testing

Assume that we want to decide between M possible hypotheses $\{\mathcal{H}_0, \dots, \mathcal{H}_{M-1}\}$ and C_{ij} is the cost of choosing $\mathcal{H}_i$ when $\mathcal{H}_j$ is true. The expected cost (Bayes risk) is

$$R = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_i | \mathcal{H}_j) P(\mathcal{H}_j).$$

Note that $C_{ij} = \delta_{ij} \Rightarrow R = P_e$. We can write

$$R = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} \int_{R_i} p(x|\mathcal{H}_j) p(\mathcal{H}_j)\, dx$$

$$= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} C_{ij} p(x|\mathcal{H}_j) P(\mathcal{H}_j)\, dx$$

$$= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j | x) p(x)\, dx$$

$$= \sum_{i=0}^{M-1} \int_{R_i} C_i(x) p(x)\, dx$$

$C_i(x) \triangleq \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|x)$ is the average cost of deciding $\mathcal{H}_i$ given x.

Each x must be assigned to one and only one R_i partition. Assigning x to R_k when $C_k(x)$ is the minimum average cost among all $\{C_0(x), \dots, C_{M-1}(x)\}$ ensures R is minimized.

$\therefore$ We decide $\mathcal{H}_i$ for which $\sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|x)$ is minimum

In the special case of P_e, $C_i(x) = \sum_{j=0}^{M-1} P(\mathcal{H}_j|x) - P(\mathcal{H}_i|x)$.

Here minimizing $C_i(x)$ is equivalent to maximizing $P(\mathcal{H}_i|x)$.

The minimum Pe decision rule is

Decide $\mathcal{H}_k$ where $k = \arg\max_i P(\mathcal{H}_i | \mathbf{x})$

which is analogous to MAP estimation. If the prior probabilities are equal ($P(\mathcal{H}_i) = \frac{1}{M}$), then we get

$$p(\mathcal{H}_i | x) = \frac{p(x|\mathcal{H}_i)\, p(\mathcal{H}_i)}{p(x)} = \frac{p(x|\mathcal{H}_i)}{M p(x)}$$

So $k = \arg\max_i p(x|\mathcal{H}_i)$, which is maximum likelihood.

An equivalent MAP decision rule is:

$$k = \arg\max_i \ln P(x|\mathcal{H}_i) + \ln p(\mathcal{H}_i) .$$

Ex) Multiple DC levels in WGN

$\mathcal{H}_0 : x[n] = -A + w[n]$ we observe $x[n]$ for $n = 0, \dots, (N-1)$.

$\mathcal{H}_1 : x[n] = 0 + w[n]$ $A > 0$, $w[n] \sim$ WGN with var. σ^2.

$\mathcal{H}_2 : x[n] = A + w[n]$

Assume that priors are equal ($P(\mathcal{H}_i) = 1/3$). Then the ML decision rule is optimal in terms of minimum expected cost.

$$p(x|\mathcal{H}_i) = (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A_i)^2}$$

where $A_0 = -A$, $A_1 = 0$, $A_2 = A$. Maximizing $p(x|\mathcal{H}_i)$ is equivalent to minimizing $D_i^2 = \sum_{n=0}^{N-1}(x[n]-A_i)^2$.

Note that $D_i^2 = \sum_{n=0}^{N-1}(x[n]-\bar{x})^2 + N(\bar{x}-A_i)^2$ after some manipulations.

To minimize D_i^2, we choose $\mathcal{H}_k$ for which A_k is closest to $\bar{x}$. $\therefore$ we decide $\mathcal{H}_0$ if $\bar{x} < -A/2$; $\mathcal{H}_1$ if $-\frac{A}{2} < \bar{x} < \frac{A}{2}$; and $\mathcal{H}_2$ if $\bar{x} > A/2$.

Let $P_c = 1 - P_e$ be the probability of correct decision.

$$P_c = \sum_{i=0}^{2} P(\mathcal{H}_i | \mathcal{H}_i) P(\mathcal{H}_i) = \frac{1}{3} \sum_{i=0}^{2} P(\mathcal{H}_i | \mathcal{H}_i)$$

$$= \frac{1}{3} \left[\Pr\left\{ \bar{x} < -\frac{A}{2} \middle| \mathcal{H}_0 \right\} + \Pr\left\{ -\frac{A}{2} < \bar{x} < \frac{A}{2} \middle| \mathcal{H}_1 \right\} + \Pr\left\{ \bar{x} > \frac{A}{2} \middle| \mathcal{H}_2 \right\} \right]$$

Since $\bar{x} \sim \mathcal{N}(A_i, \sigma^2/N)$ conditioned on $\mathcal{H}_i$, we have

$$P_c = \frac{1}{3} \left[1 - Q\left(\frac{-\frac{A}{2} + A}{\sigma/\sqrt{N}} \right) + Q\left(\frac{-A/2}{\sigma/\sqrt{N}} \right) - Q\left(\frac{A/2}{\sigma/\sqrt{N}} \right) + Q\left(\frac{A/2 - A}{\sigma/\sqrt{N}} \right) \right]$$

$$= 1 - \frac{4}{3} Q\left(\frac{A\sqrt{N}}{2\sigma} \right)$$

and $P_e = \frac{4}{3} Q\left(\frac{A\sqrt{N}}{2\sigma} \right)$.

Suggested Problems: 3, 7, 11, 13, 17, 20 (if you are into communications)