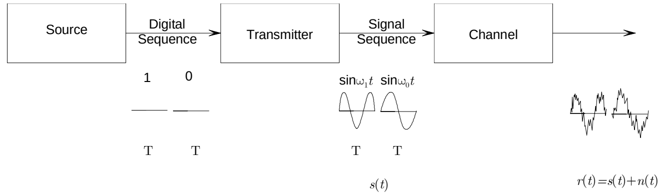
Bayesian Hypothesis testing

Statistical Decision Theory I.
Simple Hypothesis testing.
  Binary Hypothesis testing
  Bayesian Hypothesis testing.
Minimax Hypothesis testing.
Neyman-Pearson criterion.
M-Hypotheses.
Receiver Operating Characteristics.
Composite Hypothesis testing.
Composite Hypothesis testing approaches.
Performance of GLRT for large data records.
Nuisance parameters.

Classical detection and estimation theory.

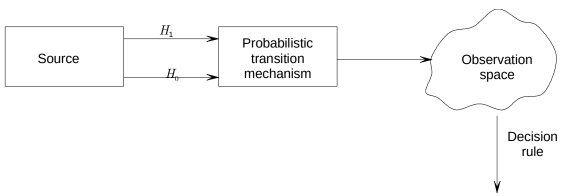
What is detection?

- Signal detection and estimation is the area of study that deals with the processing of information-bearing signals for the purpose of extracting information from them.



A simple digital communication system.

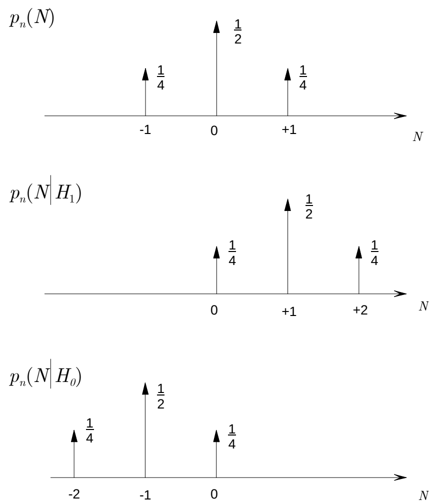
Components of a decision theory problem.

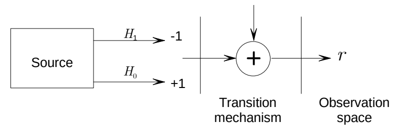


Components of a decision theory problem.

1. Source - that generates an output.
2. Probabilistic transition mechanism - a device that knows which hypothesis is true. It generates a point in the observation space accordingly to some probability law.
3. Observation space – describes all the outcomes of the transition mechanism.
4. Decision - to each point in observation space is assigned one of the hypotheses

Example:





- When H_1 is true the source generates +1.
- When H_0 is true the source generates -1.
- An independent discrete random variable n whose probability density is added to the source output.

- The sum of the source output and n is observed variable r.
- Observation space has finite dimension, i.e. observation consists of a set of N numbers and can be represented as a point in N dimensional space.
- Under the two hypotheses, we have

$H_1 : r = 1 + n$

$H_0 : r = -1 + n$

- After observing the outcome in the observation space we shall guess which hypothesis is true.
- We use a decision rule that assigns each point to one of the hypotheses.

- Detection and estimation applications involve making inferences from observations that are distorted or corrupted in some unknown manner.

Simple binary hypothesis testing.

- The decision problem in which each of two source outputs corresponds to a hypothesis.
- Each hypothesis maps into a point in the observation space.
- We assume that the observation space is a set of N observations: $r_1, r_2, \ldots, r_N$.
- Each set can be represented as a vector r:

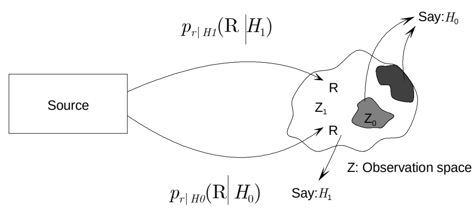
$$\mathbf{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}$$

- The probabilistic transition mechanism generates points in accord with the two known conditional densities $p_{\mathbf{r}|H_1}(\mathbf{R} \mid H_1), p_{\mathbf{r}|H_0}(\mathbf{R} \mid H_0)$.
- The objective is to use this information to develop a decision rule.

Decision criteria.

- In the binary hypothesis problem either H_0 or H_1 is true.
- We are seeking decision rules for making a choice.
- Each time the experiment is conducted one of four things can happen:
  1. H_0 true; choose $H_0 \rightarrow$ correct
  2. H_0 true; choose H_1
  3. H_1 true; choose $H_1 \rightarrow$ correct
  4. H_1 true; choose H_0
- The purpose of a decision criterion is to attach some relative importance to the four possible courses of action.
- The method for processing the received data depends on the decision criterion we select.

Bayesian criterion.



Source generates two outputs with given (*a priori*) probabilities P_1, P_0. These represent the observer information before the experiment is conducted.

- The cost is assigned to each course of actions. $C_{00}, C_{10}, C_{01}, C_{11}$.
- Each time the experiment is conducted a certain cost will be incurred.
- The decision rule is designed so that on the average the cost will be as small as possible.
- Two probabilities are averaged over: the *a priori* probability and probability that a particular course of action will be taken.

- The expected value of the cost is

$$\mathsf{R} = C_{00}P_0 \Pr\left(\text{ say } H_0 \mid H_0 \text{ is true}\right)$$
$$+C_{10}P_0 \Pr\left(\text{ say } H_1 \mid H_0 \text{ is true}\right)$$
$$+C_{11}P_1 \Pr\left(\text{ say } H_1 \mid H_1 \text{ is true}\right)$$
$$+C_{01}P_1 \Pr\left(\text{ say } H_0 \mid H_1 \text{ is true}\right)$$

- The binary observation rule divides the total observation space Z into two parts: Z_0, Z_1.
- Each point in observation space is assigned to one of these sets.
- The expression of the risk in terms of transition probabilities and the decision regions:

$$\mathsf{R} = C_{00}P_0\int_{Z_0} p_{\mathrm{r}|H_0}\left(\mathrm{R} \mid H_0\right)d\mathrm{R} + C_{10}P_0\int_{Z_1} p_{\mathrm{r}|H_0}\left(\mathrm{R} \mid H_0\right)d\mathrm{R}$$
$$+C_{11}P_1\int_{Z_1} p_{\mathrm{r}|H_1}\left(\mathrm{R} \mid H_1\right)d\mathrm{R} + C_{01}P_1\int_{Z_0} p_{\mathrm{r}|H_1}\left(\mathrm{R} \mid H_1\right)d\mathrm{R}$$

- Z_0, Z_1 cover the observation space (the integrals integrate to one).
- We assume that the cost of a wrong decision is higher than the cost of a correct decision.

  $C_{10} > C_{00}$

  $C_{01} > C_{11}$
- For Bayesian test the regions Z_0 and Z_1 are chosen such that the risk will be minimized.

- We assume that the decision is to be made for each point in observation space. $\left(Z = Z_0 + Z_1\right)$
- The decision regions are defined by the statement:

$$\mathsf{R} = C_{00}P_0\int_{Z_0} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right)d\mathbf{R} + C_{10}P_0\int_{Z-Z_0} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right)d\mathbf{R}$$
$$+C_{11}P_1\int_{Z-Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right)d\mathbf{R} + C_{01}P_1\int_{Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right)d\mathbf{R}$$

Observing that

$$\int_Z p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right)d\mathbf{R} = \int_Z p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right)d\mathbf{R} = 1$$

$$\mathsf{R} = P_0C_{10} + P_1C_{11} + \int_{Z_0}\left\{\begin{bmatrix} P_1\left(C_{01} - C_{11}\right)p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) \\ -P_0\left(C_{10} - C_{00}\right)p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right)\end{bmatrix}\right\}d\mathbf{R}$$

- The integral represents the cost controlled by those points $\mathbf{R}$ that we assign to Z_0.
- The value of $\mathbf{R}$ where the second term is larger than the first contribute to the negative amount to the integral and should be included in Z_0.
- The value of $\mathbf{R}$ where two terms are equal has no effect.
- The decision regions are defined by the statement:
  If $P_1\left(C_{01} - C_{11}\right)p_{\mathrm{r}|H_1}\left(\mathrm{R} \mid H_1\right) \geq P_0\left(C_{10} - C_{11}\right)p_{\mathrm{r}|H_0}\left(\mathrm{R} \mid H_0\right)$,
  assign $\mathbf{R}$ to Z_1 and say that H_1 is true. Otherwise assign $\mathbf{R}$ to Z_0 and say that H_0 is true.

- This may be expressed as:

$$\frac{p_{\mathrm{r}|H_1}\left(\mathrm{R} \mid H_1\right)}{p_{\mathrm{r}|H_0}\left(\mathrm{R} \mid H_0\right)} \underset{H_1}{\overset{H_0}{\lesseqgtr}} \frac{P_0\left(C_{10} - C_{00}\right)}{P_1\left(C_{01} - C_{11}\right)}$$

- $\Lambda\left(\mathrm{R}\right) = \dfrac{p_{\mathrm{r}|H_1}\left(\mathrm{R} \mid H_1\right)}{p_{\mathrm{r}|H_0}\left(\mathrm{R} \mid H_0\right)}$ is called likelihood ratio.
- Regardless of the dimension of $\mathbf{R}$, $\Lambda\left(\mathbf{R}\right)$ is one-dimensional variable.
- Data processing is involved in computing $\Lambda\left(\mathrm{R}\right)$ and is not affected by the prior probabilities and cost assignments.
- The quantity $\eta \triangleq \dfrac{P_0\left(C_{10} - C_{00}\right)}{P_1\left(C_{01} - C_{11}\right)}$ is the threshold of the test.

- The η can be left as a variable threshold and may be changed if our a priori knowledge or costs are changed.
- Bayes criterion has led us to a Likelihood Ratio Test (LRT)

$$\Lambda\left(\mathrm{R}\right) \underset{H_1}{\overset{H_0}{\lesseqgtr}} \eta$$

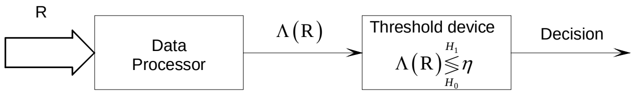
- An equivalent test is $\ln\Lambda\left(\mathrm{R}\right) \underset{H_1}{\overset{H_0}{\lesseqgtr}} \ln\eta$

Summary of the Bayesian test:

- The Bayesian test can be conducted simply by calculating the likelihood ratio $\Lambda(\mathrm{R})$ and comparing it to the threshold.

Test design:

- Assign a-priori probabilities to the source outputs.
- Assign costs for each action.
- Assume distribution for $p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right), p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right)$.
- calculate and simplify the $\Lambda(\mathrm{R})$



Likelihood ratio processor

Special case.

$$C_{00} = C_{11} = 0 \quad C_{01} = C_{10} = 1$$

$$\mathsf{R} = P_0 \int_{Z-Z_0} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right) d\mathbf{R} + P_1 \int_{Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) d\mathbf{R}$$

$$\ln \Lambda(\mathrm{R}) \underset{H_1}{\overset{H_0}{\lessgtr}} \ln \frac{P_0}{P_1} = \ln P_0 - \ln\left(1 - P_1\right)$$

- When the two hypotheses are equally likely, the threshold is zero.

Sufficient statistics.

- Sufficient statistics is a function T that transfers the initial data set to the new data set $T(\mathbf{R})$ that still contains all necessary information contained in $\mathbf{R}$ regarding the problem under investigation.
- The set that contains a minimal amount of elements is called minimal sufficient statistics.
- When making a decision knowing the value of the sufficient statistic is just as good as knowing $\mathbf{R}$.

The integrals in the Bayes test.

False alarm:

$$P_F = \int_{Z_1} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right) d\mathbf{R}.$$

We say that target is present when it is not.

Probability of detection:

$$P_D = \int_{Z_1} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) d\mathbf{R}.$$

Probability of miss:

$$P_M = \int_{Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) d\mathbf{R}.$$

We say target is absent when it is present.

Special case: the prior probabilities unknown.

Minimax test.

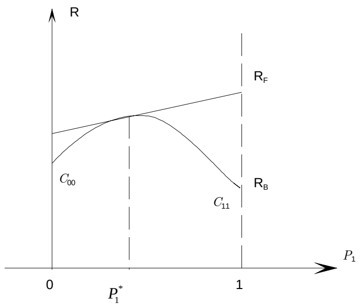
$$\mathsf{R} = C_{00}P_0\int_{Z_0} p_{\mathbf{r}|H_0}\left(\mathbf{R}\mid H_0\right)d\mathbf{R} + C_{10}P_0\int_{Z-Z_0} p_{\mathbf{r}|H_0}\left(\mathbf{R}\mid H_0\right)d\mathbf{R}$$
$$+C_{11}P_1\int_{Z-Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R}\mid H_1\right)d\mathbf{R} + C_{01}P_1\int_{Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R}\mid H_1\right)d\mathbf{R}$$

- If the regions Z_0 and Z_1 fixed the integrals are determined.
  $\mathsf{R} = P_0C_{10} + P_1C_{11} + P_1\left(C_{01} - C_{11}\right)P_M - P_0\left(C_{10} - C_{00}\right)\left(1 - P_F\right)$
  $P_0 = 1 - P_1$
- The Bayesian risk will be function of P_1.

$$\mathsf{R}\left(P_1\right) = C_{00}\left(1 - P_F\right) + C_{10}P_F$$
$$+P_1\left[\left(C_{11} - C_{00}\right) + \left(C_{01} - C_{11}\right)P_M - \left(C_{10} - C_{00}\right)P_F\right]$$

- Bayesian test can be found if all the costs and a priori probabilities are known.
- If we know all the probabilities we can calculate the Bayesian cost.
- Assume that we do not know P_1 and just assume a certain one P_1^* and design a corresponding test.
- If P_1 changes the regions Z_0, Z_1 changes and with these also P_F and P_D.

- The test is designed for P_1^* but the actual a priori probability is P_1.

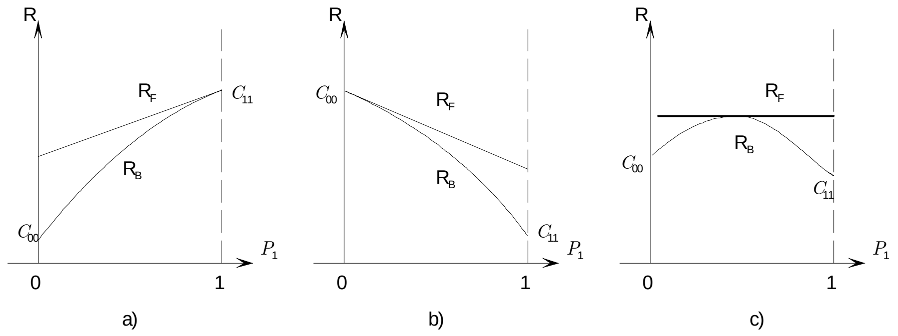


A function of P_1 if P_1^* is fixed

- By assuming P_1^* we fix P_F and P_D.
- Cost for different P_1 is given by a function $\mathsf{R}\left(P_1^*, P_1\right)$.
- Because the threshold η is fixed the cost $\mathsf{R}\left(P_1^*, P_1\right)$ is a linear function of P_1^*.
- Bayesian test minimizes the risk for P_1^*.
  For other values of P_1
  $\mathsf{R}\left(P_1^*, P_1\right) \geq \mathsf{R}\left(P_1\right)$
- $\mathsf{R}\left(P_1\right)$ is strictly concave. (If $\Lambda(\mathrm{R})$ is a continuos random variable with strictly monotonic probability distribution function, the change of η always change the risk.)

Minimax test.

- The Bayesian test designed to minimize the maximum possible risk is called a minimax test.
- P_1 is chosen to maximize our risk $\mathsf{R}\left(P_1^*, P_1\right)$.
- To minimize the maximum risk we select the P_1^* for which $\mathsf{R}\left(P_1\right)$ is maximum.
- If the maximum occurs inside the interval $[0,1]$, the $\mathsf{R}\left(P_1^*, P_1\right)$will become a horizontal line. Coefficient of P_1 must be zero.
- $\left(C_{11} - C_{00}\right) + \left(C_{01} - C_{11}\right)P_M - \left(C_{10} - C_{00}\right)P_F = 0$ This equation is the minimax equation.



Risk curves: maximum value of R at a) $P_1 = 1$ b) $P_1 = 0$ c) $0 \le P_1 \le 1$

Special case.

Cost function is

$C_{00} = C_{11} = 0$

$C_{01} = C_M, C_{10} = C_F$.

The risk is

$\mathsf{R}(P_1) = C_F P_F + P_1 (C_M P_M - C_F P_F) = P_0 C_F P_F + P_1 C_M P_M$.

The minimax equation is

$C_M P_M = C_F P_F$.

<u>Neyman-Pearson test.</u>

- Often it is difficult to assign realistic costs of a priori probabilities. This can be bypassed if to work with the conditional probabilities P_F and P_D.
- We have two conflicting objectives to make P_F as small as possible and P_D as large as possible.

Neyman-Pearson criterion.

Constrain $P_F = \alpha' \le \alpha$ and design a test to maximize P_D (or minimize P_M) under this constraint.

- The solution can be obtained by using Lagrange multipliers.

$$F = P_M + \lambda \left[P_F - \alpha' \right]$$

$$\mathrm{F} = \int_{Z_0} p_{\mathrm{r}|H_1}(\mathrm{R} \mid H_1)\, d\mathrm{R} + \lambda \left[\int_{Z-Z_0} p_{\mathrm{r}|H_0}(\mathrm{R} \mid H_0)\, d\mathrm{R} - \alpha' \right]$$

- If $P_F = \alpha$, minimizing F minimizes P_M.

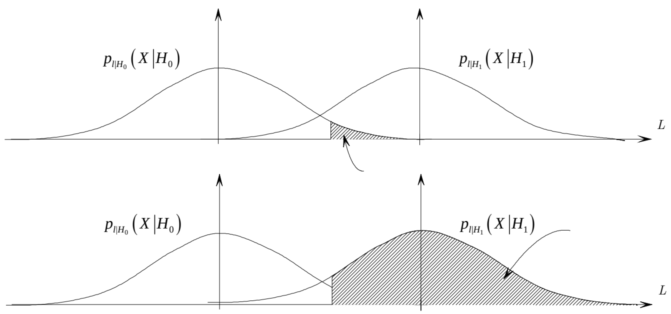
$$\mathrm{F} = \lambda(1-\alpha') + \int_{Z_0} \left[p_{\mathrm{r}|H_1}(\mathrm{R} \mid H_1) - \lambda p_{\mathrm{r}|H_0}(\mathrm{R} \mid H_0) \right] d\mathrm{R}$$

- For any positive value of λ an LRT will minimize F.
- F is minimized if we assign a point R to Z_0 only when the term in the bracket is negative.
- If $\dfrac{p_{\mathrm{r}|H_1}(\mathrm{R} \mid H_1)}{p_{\mathrm{r}|H_0}(\mathrm{R} \mid H_0)} < \lambda$ assign point to Z_0 (say H_0)

- F is minimized by the likelihood ratio test. $\Lambda(\mathrm{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$
- To satisfy the constraint λ is selected so that $P_F = \alpha'$.

$$P_F = \int_{\lambda}^{\infty} p_{\Lambda|H_0}\left(\Lambda \mid H_0\right) d\Lambda = \alpha'$$

- Value of λ will be nonnegative because $p_{\Lambda|H_0}\left(\Lambda \mid H_0\right)$ will be zero for negative values of λ.



Example.

We assume that under H_1 the source output is a constant voltage m. Under H_0 the sourse output is zero. Voltage is corrupted by an additive noise. The out put is sampled with N samples for each second. Each noise sample is a i.i.d. zero mean Gaussian random variable with variance σ^2.

$$H_1 : r_i = m + n_i, \quad i = 1, 2, \ldots, N$$

$$H_0 : r_i = n_i, \quad i = 1, 2, \ldots, N$$

$$p_{n_i}(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{X^2}{2\sigma^2}\right)$$

The probability density of r_i under each hypothesis is:

$$p_{r_i|H_1}\left(R_i \mid H_1\right) = p_{n_i}\left(R_i - m\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(R_i - m\right)^2}{2\sigma^2}\right)$$

$$p_{r_i|H_0}\left(R_i \mid H_0\right) = p_{n_i}\left(R_i\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(R_i\right)^2}{2\sigma^2}\right)$$

- The joint probability of N samples is:

$$p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(R_i - m\right)^2}{2\sigma^2}\right)$$

$$p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{R_i^2}{2\sigma^2}\right)$$

- The likelihood ratio is

$$\Lambda(\mathbf{R}) = \frac{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(R_i - m\right)^2}{2\sigma^2}\right)}{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{R_i^2}{2\sigma^2}\right)}$$

- After cancelling common terms and taking logarithm:

$$\ln \Lambda(\mathbf{R}) = \frac{m}{\sigma^2} \sum_{i=1}^{N} R_i - \frac{Nm^2}{2\sigma^2}.$$

- Likelihood ratio test is:

$$\frac{m}{\sigma^2}\sum_{i=1}^{N} R_i - \frac{Nm^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln\eta,$$

$$\sum_{i=1}^{N} R_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2}{m}\ln\eta + \frac{Nm}{2} \triangleq \gamma.$$

- We use $d \triangleq \frac{\sqrt{N}m}{\sigma}$ for normalisation.

$$l = \frac{1}{\sqrt{N}\sigma}\sum_{i=1}^{N} R_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma}{\sqrt{N}m}\ln\eta + \frac{\sqrt{N}m}{2\sigma}$$

- Under H_0 l is obtained by adding N independent zero mean Gaussian variables with variance σ^2 and then dividing by $\sqrt{N}\sigma$. Therefore l is $N(0,1)$.

- Under H_1 l is $N\left(\frac{\sqrt{N}m}{\sigma},1\right)$.

$$P_F = \int_{(\log\eta)/d+d/2}^{\infty} \frac{1}{\sqrt{2}}\exp\left(-\frac{x^2}{2}\right)dx = erfc\left(\frac{\ln\eta}{d}+\frac{d}{2}\right)$$

where $d \triangleq \frac{\sqrt{N}m}{\sigma}$ is the distance between the means of the two densities.

$$P_D = \int_{(\log\eta)/d+d/2}^{\infty} \frac{1}{\sqrt{2}}\exp\left(-\frac{(x-d)^2}{2}\right)dx$$

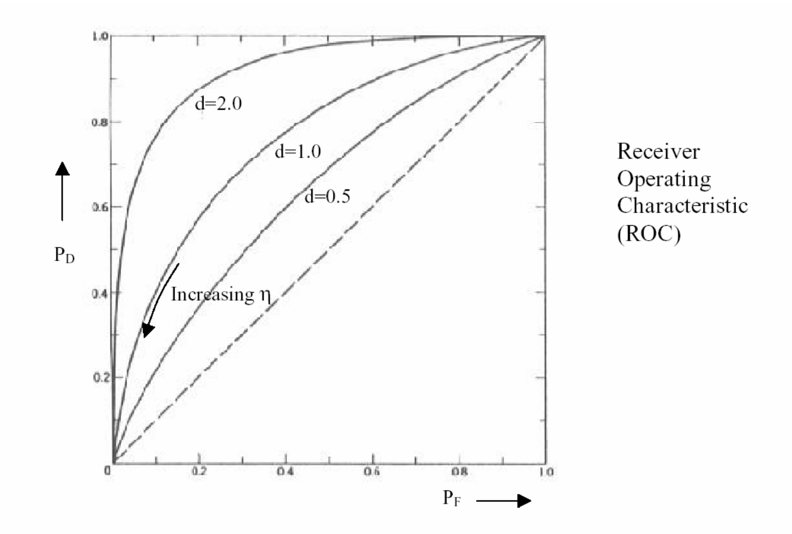
$$= \int_{(\log\eta)/d-d/2}^{\infty} \frac{1}{\sqrt{2}}\exp\left(-\frac{(y)^2}{2}\right)dy = erfc\left(\frac{\log\eta}{d}-\frac{d}{2}\right)$$

- In the communication systems a special case is important $\Pr(\varepsilon) \triangleq P_0 P_F + P_1 P_M$.
- If $P_0 = P_1$ the threshold is one and $\Pr(\varepsilon) \triangleq \frac{1}{2}\left(P_F + P_M\right)$.

Receiver Operating Characteristics (ROC).

- For a Neyman-Pearson test the values of P_F and P_D completely specify the test performance.
- P_D depends on P_F. The function of $P_D\left(P_F\right)$ is defined as the Receiver Operating Characteristic (ROC).
- The Receiver Operating Characteristic (ROC) completely describes the performance of the test as a function of the parameters of interest.

Example.



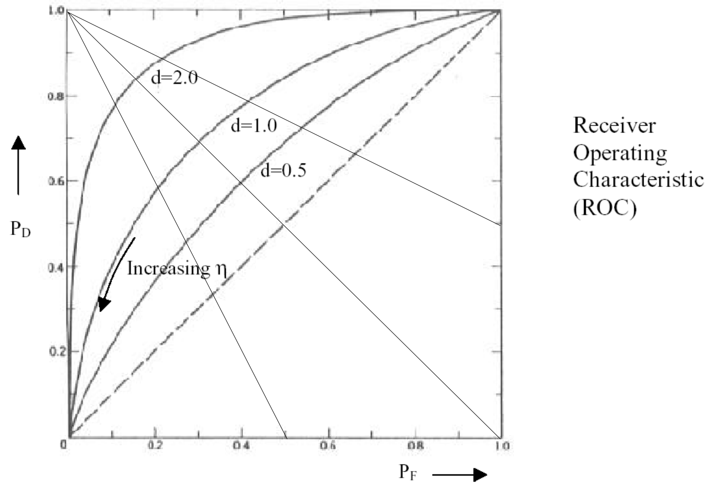
Properties of ROC.

- All continuous likelihood tests have ROC's that are concave downward.
- All continuous likelihood ratio tests have ROC's that are above the $P_D = P_F$.
- The slope of a curve in a ROC at a particular point is equal to the value of the threshold η required to achieve the P_F and P_D at that point

Whenever the maximum value of the Bayes risk is interior to the interval (0,1) of the P1 axis the minimax operating point is the intersection of the line

$$\left(C_{11} - C_{00}\right) + \left(C_{01} - C_{11}\right)\left(1 - P_D\right) - \left(C_{10} - C_{00}\right)P_F = 0$$

and the appropriate curve on the ROC.



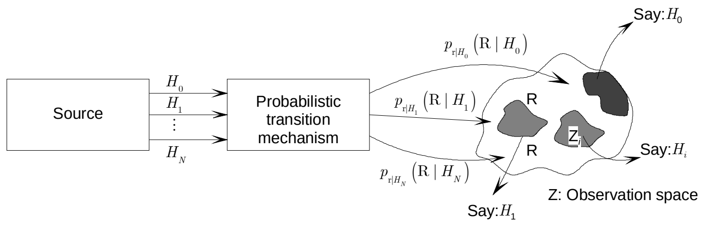
Determination of minimax operating point.

Conclusions.

- Using either the Bayes criterion of a Neyman-Pearson criterion, we find that the optimum test is a likelihood ratio test.
- Regardless of the dimension of the observation space the optimum test consist of comparing a scalar variable $\Lambda(\mathbf{R})$ with the threshold.
- For the binary hypothesis test the decision space is one dimensional.
- The test can be simplified by calculating the sufficient statistics.
- A complete description of the LRT performance was obtained by plotting the conditioning probabilities P_D and P_F as the threshold η was varied.

M Hypotheses.

- We choose one of M hypotheses
- There are M source outputs each of which corresponds to one of M hypotheses.
- We are forsed to make decisions.
- There are M^2 alternatives that may occur each time the experiment is conducted.



Bayes Criterion.

C_{ij} cost of each course of actions.

Z_i region in observation space where we chose H_i

P_i a priori probabilities.

$$\mathsf{R} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_j C_{ij} \int_{Z_i} p_{\mathbf{r}|H_j}\left(\mathbf{R} \mid H_j\right) d\mathbf{R}$$

R is minimized through selecting Z_i.

Example M = 3.

$$Z = Z_0 + Z_1 + Z_2$$

$$\mathsf{R} = P_0 C_{00} \int_{Z-Z_1-Z_2} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right) d\mathbf{R} + P_0 C_{10} \int_{Z_1} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right) d\mathbf{R}$$

$$+ P_0 C_{20} \int_{Z_2} p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0\right) d\mathbf{R} + P_1 C_{11} \int_{Z-Z_0-Z_2} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) d\mathbf{R}$$

$$+ P_1 C_{01} \int_{Z_0} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) d\mathbf{R} + P_1 C_{21} \int_{Z_2} p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right) d\mathbf{R}$$

$$+ P_2 C_{22} \int_{Z-Z_0-Z_1} p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2\right) d\mathbf{R} + P_2 C_{02} \int_{Z_0} p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2\right) d\mathbf{R}$$

$$P_2 C_{12} \int_{Z_1} p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2\right) d\mathbf{R}$$

$$
\begin{aligned}
\mathsf{R} = {} & P_0 C_{00} + P_1 C_{11} + P_2 C_{22} \\
& + \int_{Z_0} \Big[P_2 \left(C_{02} - C_{22} \right) p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2 \right) + P_1 \left(C_{01} - C_{11} \right) p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1 \right) \Big] d\mathbf{R} \\
& + \int_{Z_1} \Big[P_0 \left(C_{10} - C_{00} \right) p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0 \right) + P_2 \left(C_{12} - C_{22} \right) p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2 \right) \Big] d\mathbf{R} \\
& + \int_{Z_2} \Big[P_0 \left(C_{20} - C_{00} \right) p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0 \right) + P_1 \left(C_{21} - C_{11} \right) p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1 \right) \Big] d\mathbf{R}
\end{aligned}
$$

- R is minimized if we assign each $\mathbf{R}$ to the region in which the value of the integrand is the smallest.
- Label the integrals $I_0(\mathbf{R}), I_1(\mathbf{R}), I_2(\mathbf{R})$.

$I_0(\mathbf{R}) < I_1(\mathbf{R})$ and $I_2(\mathbf{R})$, choose H_0
$I_1(\mathbf{R}) < I_0(\mathbf{R})$ and $I_2(\mathbf{R})$, choose H_1
$I_2(\mathbf{R}) < I_0(\mathbf{R})$ and $I_1(\mathbf{R})$, choose H_2

- If we use likelihood ratios

$$\Lambda_1(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1 \right)}{p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0 \right)}$$

$$\Lambda_2(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2 \right)}{p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0 \right)}$$

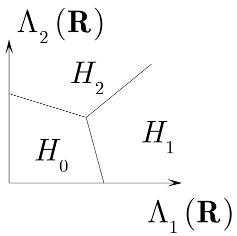
The set of decision equations is:

$$P_1 \left(C_{01} - C_{11} \right) \Lambda_1(\mathbf{R}) \underset{H_1 \text{ or } H_2}{\overset{H_0 \text{ or } H_2}{\gtrless}} P_0 \left(C_{10} - C_{00} \right) + P_2 \left(C_{12} - C_{02} \right) \Lambda_2(\mathbf{R})$$

$$P_2 \left(C_{02} - C_{22} \right) \Lambda_2(\mathbf{R}) \underset{H_2 \text{ or } H_1}{\overset{H_0 \text{ or } H_1}{\gtrless}} P_0 \left(C_{20} - C_{00} \right) + P_1 \left(C_{21} - C_{01} \right) \Lambda_1(\mathbf{R})$$

$$P_2 \left(C_{12} - C_{22} \right) \Lambda_2(\mathbf{R}) \underset{H_0 \text{ or } H_2}{\overset{H_0 \text{ or } H_1}{\gtrless}} P_0 \left(C_{20} - C_{10} \right) + P_1 \left(C_{21} - C_{11} \right) \Lambda_1(\mathbf{R})$$

- M hypotheses always lead to a decision space that has, at most, $M-1$ dimensions.



Decision Space

<u>Special case.</u> (often in communication)

$C_{00} = C_{11} = C_{22} = 0$
$C_{ij} = 1, \; i \neq j$

$$P_1 p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1 \right) \underset{H_0 \text{ or } H_2}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0 p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0 \right)$$

$$P_2 p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2 \right) \underset{H_0 \text{ or } H_1}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0 p_{\mathbf{r}|H_0}\left(\mathbf{R} \mid H_0 \right)$$

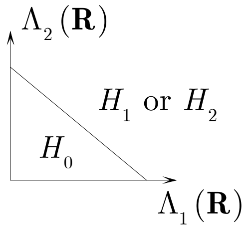
$$P_2 p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2 \right) \underset{H_0 \text{ or } H_1}{\overset{H_0 \text{ or } H_2}{\gtrless}} P_0 p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1 \right)$$

- Compute the posterior probabilities and choose the largest.
- Maximum a posteriori probability computer.

Special case.

Degeneration of hypothesis.

What happens if to combine H_1 and H_2:



$C_{12} = C_{21} = 0.$

For simplicity

$C_{01} = C_{10} = C_{20} = C_{02}$

$C_{00} = C_{11} = C_{22} = 0$

First two equations of the test reduce to

$$P_1\Lambda_1(\mathbf{R}) + P_2\Lambda_2(\mathbf{R}) \underset{H_0}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0$$

Dummy hypothesis.

- Actual problem has two hypothesis H_1 and H_2.
- We introduce a new one H_0 with a priori probability $P_0 = 0$
- Let

  $P_1 + P_2 = 1$ and $C_{12} = C_{02}$, $C_{21} = C_{01}$
- We always choose H_1 or H_2. The test reduces to:

$$P_2\left(C_{12} - C_{22}\right)\Lambda_2(\mathbf{R}) \underset{H_1}{\overset{H_2}{\gtrless}} P_1\left(C_{21} - C_{11}\right)\Lambda_1(\mathbf{R})$$

- Useful if $\dfrac{p_{\mathbf{r}|H_2}\left(\mathbf{R} \mid H_2\right)}{p_{\mathbf{r}|H_1}\left(\mathbf{R} \mid H_1\right)}$ difficult to work with but $\Lambda_1(\mathbf{R})$, $\Lambda_2(\mathbf{R})$ are simple.

Conclusions.

1. The minimum dimension of the decision space is no more that $M-1$. The boundaries of the decision regions are hyperplanes in the $\left(\Lambda_1, \ldots, \Lambda_{m-1}\right)$ plane.

2. The test is simple to find but error probabilities are often difficult to compute.

3. An important test is the minimum total probability of error test. Here we compute the a posteriori probability of each test and choose the largest.