Multimedia Security and Privacy

TP3: Elements of Detection Theory

Prof. Sviatoslav Voloshynovskiy,
Denis Ullmann <denis.ullmann@unige.ch>.

Stochastic Information Processing Group

April 4, 2019

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'informatique

SIP stochastic information processing

Submission

Please archive your report and codes in "Name_Surname.zip" (replace "Name" and "Surname" with your real name), and upload to "`Assignments/TP3: Elements of Detection Theory`" on `https://chamilo.unige.ch` before Wednesday, April 10 2018, 23:59 PM. Note, **the assessment is mainly based on your report, which should include your answers to all questions and the experimental results**.

1 Elements of Detection Theory

Pattern classification is the process of assigning a class label to a physical object, process, or event based on measurement data. Bayesian decision theory is a statistical approach to pattern classification. Using probability it allows to quantify the performance (and costs) that are associated with various classification decisions.

Gaussian Random Variables

Definition 1.1. The probability density function (PDF) $f_X(x)$ of a Gaussian random variable X with (μ, σ) is defined as:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Theorem 1.1. For a Gaussian random variable X also denoted as $X \sim \mathcal{N}(\mu, \sigma)$:

$$E[X] = \mu, \quad Var[X] = \sigma^2$$

Theorem 1.2. If $X \sim \mathcal{N}(\mu, \sigma)$, and $Y = \alpha X + b$, then $Y \sim \mathcal{N}(\alpha\mu + b, \alpha\sigma)$

Definition 1.2. The standard normal cumulative density function (CDF) of N, where $N \sim \mathcal{N}(0, 1)$:

$$\Phi(n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{n} e^{\frac{-u^2}{2}} du$$

Theorem 1.3. If $X \sim \mathcal{N}(\mu, \sigma)$, the CDF of X is:

$$F_X(x) = \Phi\Big(\frac{x-\mu}{\sigma}\Big)$$

and the probability that X is in the interval $(a, b]$ is:

$$P[a < X \leq b] = \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big)$$

This means that we can use this theorem to transform values of some Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma)$ to the standard normal $N \sim \mathcal{N}(0, 1)$ for which the $\Phi(n)$ values can be looked up. E.g, for a sample x of X the corresponding value n of N is:

$$n = \frac{x-\mu}{\sigma}$$

Theorem 1.4.

$$\Phi(-n) = 1 - \Phi(n)$$

Example 1.1. Let $X \sim \mathcal{N}(61, 10)$, what is $P[X \leq 46]$?

$$P[X \leq 46] = F_X(46) = \Phi\Big(\frac{46-61}{10}\Big) = \Phi(-1.5) = 1 - \Phi(1.5) \approx 1 - 0.933 = 0.067$$

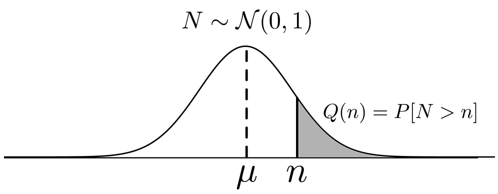


Figure 1 – Q-function illustration for $Q(n)$.

Example 1.2. Let $X \sim \mathcal{N}(61, 10)$, what is $P[51 < X \leq 71]$?

Noting that the standard normal $N = \frac{X-61}{10}$ and the event $\{51 < X \leq 71\}$ corresponds to the event $\{(51-61)/10 < N \leq (71-61)/10\}$ or $\{-1 < N \leq 1\}$:

$$P[-1 < N \leq 1] = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 \approx 0.683$$

Definition 1.3. The standard normal complementary CDF:

$$Q(N) = P[N > n] = \frac{1}{\sqrt{2\pi}} \int_n^{\infty} e^{\frac{-u^2}{2}} du = 1 - \Phi(n)$$

See Figure 1.

Exercise 1

Let $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 3)$. What is:

- $P[-1 < Y \leq 1]$
- $P[Y > 3.5]$
- $P[-2 < X \leq 2]$
- $P[X > 2.5]$

Exercise 2

Let X denote the peak temperature in Geneva, in June, as measured in Celsius, for which holds: $X \sim \mathcal{N}(30, 11)$. What is:

- $P[X > 40]$
- $P[X \leq 15]$
- $P[20 < X \leq 40]$

Exercise 3

Let X be a Gaussian random variable, for which $E[X] = 0$ and $P[|X| \leq 10] = 0.2$. What is σ_X? Note that tables for Φ can be found online, or can be determined by using the (inverse) Cumulative Distribution Function, present Matlab.

Exercise 4

Next to the Q-function, the communications field uses the *complementary error function* (ERFC) for the tail probabilities of Gaussian random variables. It is defined as:

$$\operatorname{erfc}(n) = \frac{2}{\sqrt{\pi}} \int_n^\infty e^{-x^2} dx$$

Prove that:

$$Q(n) = \frac{1}{2}\operatorname{erfc}\Big(\frac{n}{\sqrt{2}}\Big) \tag{1}$$

Hint

- Formulate for yourself, what the difference is between the Matlab `erf`, `erfc`, `normcdf` and the $Q(.)$ functions.

Definition 1.4. Integral substitution.

The substitution method for integrals is defined as follows:

$$\int_{\varphi(b)}^{\varphi(a)} f(x)dx = \int_b^a f\{\varphi(z)\}\varphi'(z)dz. \tag{2}$$

Where φ is a bijective differentiable function, $x = \varphi(z)$ and $dx = \varphi'(z)dz$

Example

For example let:

$$\int (2x^7 - 5)^3 \cdot 28x^6 dx \qquad \text{Define:} \tag{3}$$

$$z = 2x^7 - 5$$
$$\frac{dz}{dx} = 14x^6$$
$$x^6 dx = \frac{1}{14}dz$$

Hence,

$$\int (2x^7 - 5)^3 \cdot 28x^6 dx = \frac{28}{14}\int z^3 dz = 2 \cdot \frac{z^4}{4} = \frac{1}{2}(2x^7 - 5)^4 + C \tag{4}$$

For example let:

$$\int_0^2 x\cos(x^2 + 1)dx \qquad \text{Define:} \tag{5}$$

$$u = x^2 + 1$$
$$du = 2xdx$$
$$dx = \frac{1}{2}\frac{du}{x}$$

Hence,

$$\int_{x=0}^{x=2} x\cos(x^2 + 1)dx = \int_{u=1}^{u=5} x\cos(u)\frac{1}{2}\frac{du}{x} \tag{6}$$
$$= \frac{1}{2}\int_{u=1}^{u=5} \cos(u)du \tag{7}$$

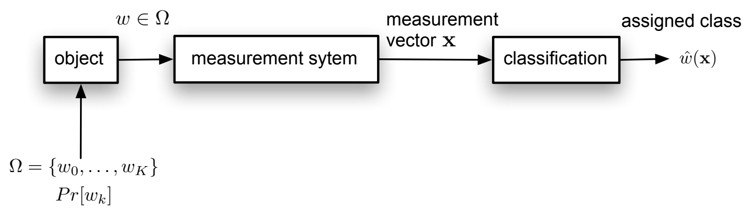


Figure 2 – Statistical pattern classification framework

2 Bayesian Two-class Classification

The starting point is a stochastic experiment as seen in Figure 2 where the objects are categorised by a set $\Omega = \{w_0, \cdots, w_K\}$ of K classes. The probability of having an object of class k in the real world, $P(w_k)$, is called the *prior probability.* It represents the gathered knowledge we have about this class of object prior to taking any measurements. Further more, we stipulate that classes are exclusive and that no other classes are in play:

$$\sum_{k=0}^{K} P(w_k) = 1 \tag{8}$$

The measurement system produces a *measurement vector* $\mathbf{x}$ with some dimension. Objects from different classes should produce different measurements, but those from identical classes will also differ. The reason is two-fold. Imagine you are classifying between cats and dogs based on weight. Obviously not all cats weigh the same. In addition, the measurements can be subject to noise due to the measurement device itself. All these variations are modelled by the probability density function $p(\mathbf{x})$. The conditional probability mass function $p(\mathbf{x}|w_k)$ models the density of the measurements giving that they originate from objects of class w_k. Since we assume that the classes are mutual exclusive, the unconditional density $p(\mathbf{x})$ becomes:

$$p(\mathbf{x}) = \sum_{k=0}^{K} p(\mathbf{x} \mid w_k) P(w_k) \tag{9}$$

This leaves one last probability, the *posterior probability* $P(w_k \mid \mathbf{x})$. It denotes the probability that an object belongs to class w_k given measurement vector $\mathbf{x}$.

Following Bayes' theorem:

$$P(w_k \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid w_k) P(w_k)}{p(\mathbf{x})} \tag{10}$$

Informally, one can read this as:

$$\text{posterior} = \frac{\text{likelihood} \times \text{priors}}{\text{evidence or scale factor}} \tag{11}$$

We will take a small side-step and have a look at the probability of error while classifying between to classes, w_0 and w_1 based on measurement vector $\mathbf{x}$:

The probability of an error can be simply expressed:

$$p(error \mid \mathbf{x}) = \begin{cases} p(w_0 \mid \mathbf{x}) & \text{if we choose class } w_1 \\ p(w_1 \mid \mathbf{x}) & \text{if we choose class } w_0 \end{cases} \tag{12}$$

Now, there will be scenario's in which some errors might be modelled as more serious than others, but for now we will use what is known as a *uniform cost function.* This means that there is a unit cost for an error, and zero cost in the case there is no error.

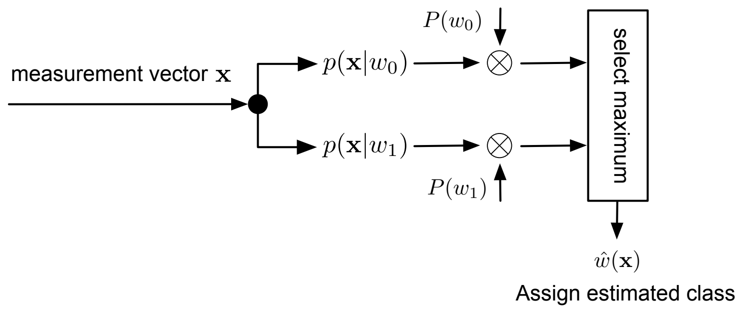


Figure 3 – Bayes decision function with uniform cost function or MAP classification.

Minimizing the *uniform cost function* is equivalent to maximizing the *posterior probability* $P(w_k \mid \mathbf{x})$. Using Bayes' rule:

$$P(w_k \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid w_k)P(w_k)}{p(\mathbf{x})} \tag{13}$$

becomes the so called *Maximum A Posteriori* (MAP) classifier when:

$$\begin{aligned} \hat{w}_{MAP}(\mathbf{x}) &= \operatorname{argmax} P(w_k \mid \mathbf{x}) \\ &= \operatorname{argmax} p(\mathbf{x} \mid w_k)P(w_k) \end{aligned} \tag{14}$$

Visually, one can see how this method works in Figure 3. For every measurement vector $\mathbf{x}$ and all classes k, one evaluates $p(\mathbf{x}|w_k) \times P(w_k)$. The classifier than selects the maximum result and returns the estimated class $\hat{w}(\mathbf{x})$.

In the special 2-class case, the MAP classifier simply tests:

$$p(\mathbf{x} \mid w_0)P(w_0) > p(\mathbf{x} \mid w_1)P(w_1) \begin{cases} true & \rightarrow \text{decide } w_0 \\ false & \rightarrow \text{decide } w_1 \end{cases} \tag{15}$$

Refactoring:

$$p(\mathbf{x} \mid w_0)P(w_0) \gtrless p(\mathbf{x} \mid w_1)P(w_1) \tag{16}$$

$$\frac{p(\mathbf{x} \mid w_0)}{p(\mathbf{x} \mid w_1)} \gtrless \frac{P(w_1)}{P(w_0)} \tag{17}$$

Now, if we look at $p(\mathbf{x} \mid w_k)P(w_k)$ as a function of w_k:

$$\mathcal{L}(\mathbf{x}) = \frac{p(\mathbf{x} \mid w_0)}{p(\mathbf{x} \mid w_1)} \tag{18}$$

$\mathcal{L}(\mathbf{x})$ is called the *likelihood function* of w_k or the *likelihood ratio*.

Now, deciding between class w_0 and w_1 in the 2-class MAP case becomes:

$$\mathcal{L}(\mathbf{x}) \underset{w_1}{\overset{w_0}{\gtrless}} \frac{P(w_1)}{P(w_0)} \tag{19}$$

If the measurement vectors are Gaussian, e.g. $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma\mathbf{I})$ it is custom to use the *log likelihood function*:

$$\ln(\mathcal{L}(\mathbf{x})) \underset{w_1}{\overset{w_0}{\gtrless}} \ln\left(\frac{P(w_1)}{P(w_0)}\right) \tag{20}$$

$$\Lambda(\mathbf{x}) \underset{w_1}{\overset{w_0}{\gtrless}} T, \quad T = \ln\left(\frac{P(w_1)}{P(w_0)}\right) \tag{21}$$

Detection

In the case of Bayes two-class classification, with classes w_0 and w_1 imagine the following scenario. A burglar alarm, based on an infrared measuring device, classifies between two options: w_0 : no event and w_1 : burglar present. We will call these two options hypothesis:

$$\begin{cases} H_0: & \text{no event} \\ H_1: & \text{event} \end{cases} \tag{22}$$

Clearly, there are four different outcomes:

	$w = w_0$	$w = w_1$
$\hat{w}(\mathbf{x}) = w_0$	True negative	Mis or False negative
$\hat{w}(\mathbf{x}) = w_1$	False Alarm	Hit or True positive

Then the probabilities of *miss*, p_m and *false alarm*, p_f are defined as:

$$p_m = P(\hat{w}_0 \mid w_1) \tag{23}$$

$$p_f = P(\hat{w}_1 \mid w_0) \tag{24}$$

Then, given T as threshold:

$$p_m = P(\Lambda(\mathbf{x}) > T \mid w_1) = \int_T^{\infty} p(\Lambda \mid w_1) d\Lambda \tag{25}$$

$$p_f = P(\Lambda(\mathbf{x}) < T \mid w_0) = \int_{-\infty}^{T} p(\Lambda \mid w_0) d\Lambda \tag{26}$$

Definition 2.1. The total probability of error, p_{err}

The maximum a posteriori probability (MAP) test **minimizes** the total probability of error of a binary hypothesis test, p_{err}.

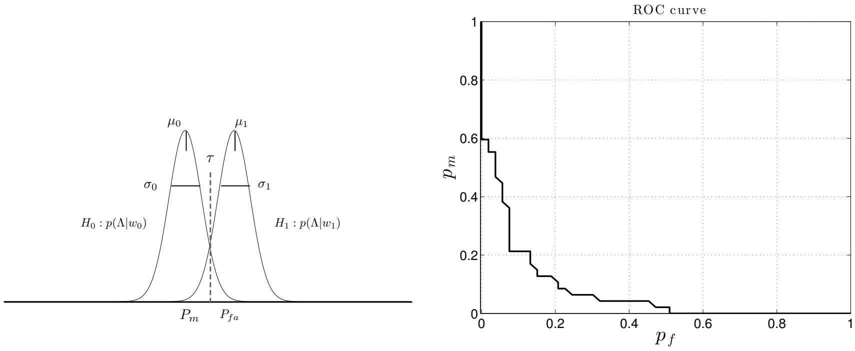
$$\boxed{p_{err} = P(\Lambda(\mathbf{x}) > T \mid w_1)P(w_1) + P(\Lambda(\mathbf{x}) < T \mid w_0)P(w_0)} \tag{27}$$

It should be clear that pending the probabilistic nature of hypothesises H_0 and H_1 there is no value T for which error-less classification is possible, as shown in Figure 4a. For any T the system will thus exhibit false positives and misses. The standard tool to show how T, p_m, and p_f interact is the *Receiver Operator Characteristic* or ROC curve, which is shown in Figure 4b.

Reformulation and Examples

We will now slightly reformulate the MAP framework to a signal processing context, and present a number of examples. The principles remain identical, except that instead of classifying between classes w_0 and w_1 we will let a detector decide between two hypothesises H_0 and H_1.

In this binary hypothesis test, there are two hypothetical probability models, H_0 and H_1, and two possible conclusions. The detector either decides that H_0 is the true model, or H_1. The *a priori* probabilities for these hypothesises are $P[H_0]$ and $P[H_1]$. The binary hypothesis test divides the sample space $S = \{H_0, H_1\}$ into two mutually exclusive, collectively exhaustive sets A_0 and A_1. If the detector outputs $s \in A_0$, it will accept H_0, otherwise, it will accept H_1. If the detector output is based on some stochastic random variable X, the corresponding probability models are the conditional mass functions (discrete case) $P_{X|H_0}(x)$ and $P_{X|H_1}(x)$ and in the continuous case they are the probability density functions $f_{X|H_0}(x)$ and $f_{X|H_1}(x)$.



(a) The log-likelihood ratio in the case where hypothesis H_0 and H_1 are Gaussian with identical σ.

(b) Receiver Operator Characteristic

Figure 4 – Bayes two-class classification.

Definition 2.2. The probability of error, p_{err}

The maximum a posteriori probability (MAP) test **minimizes** the total probability of error of a binary hypothesis test, p_{err}:

$$\boxed{p_{err} = P(A_1 \mid H_0)P(H_0) + P(A_0 \mid H_1)P(H_1)} \tag{28}$$

Theorem 2.1. The Maximum a posteriori probability (MAP) binary hypothesis test:

Given a binary hypothesis test, with outcome s, the following rule leads to the lowest p_{err} value:

$$s \in A_0, \text{if } P[H_0 \mid s] \geq P[H_1 \mid s] \tag{29}$$

$$s \in A_1, \text{if } P[H_1 \mid s] > P[H_0 \mid s] \tag{30}$$

Theorem 2.2. For an experiment that produces a discrete random variable X, the MAP hypothesis test is:

$$x_i \in A_0, \text{if } \frac{P_{X|H_0}(x_i)}{P_{X|H_1}(x_i)} \geq \frac{P[H_1]}{P[H_0]} \tag{31}$$

Theorem 2.3. For an experiment that produces a continuous random variable X, the MAP hypothesis test is:

$$x_i \in A_0, \text{if } \frac{f_{X|H_0}(x_i)}{f_{X|H_1}(x_i)} \geq \frac{P[H_1]}{P[H_0]} \tag{32}$$

Examples binary hypothesis testing

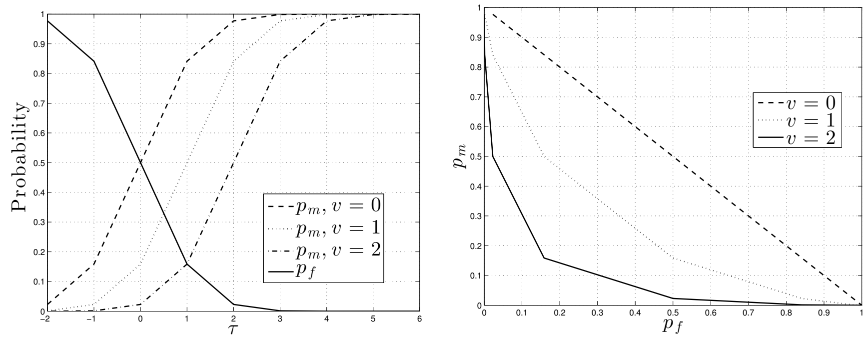
We will now go through the Bayesian two-class classification or binary hypothesis testing from a signal processing (radar) viewpoint, with a number of examples.

Example 2.1. Let X denote the received signal for a radar detection system. This system continuously checks the following hypothesis:

$$\begin{cases} H_0 : & \text{No target} \\ H_1 : & \text{Target} \end{cases} \tag{33}$$

These two hypotheses are based on X, which is also called the *sufficient statistic* as follows:

$$\begin{cases} H_0 : & X = N \\ H_1 : & X = v + N \end{cases} \tag{34}$$



(a) **(b)**

Figure 5 – **(a)** shows the p_m and p_f versus the value of τ for example 2.1. **(b)** shows the corresponding ROC graph.

where, $N \sim \mathcal{N}(0, 1)$ denotes the noise of the system and $v \geq 0$. Note the absence of any information about the a priori probabilities $P(H_0)$ and $P(H_1)$! We can now define the acceptance sets in which a result must lie to accept one of the hypotheses. A result in set $A_0 = \{X \leq \tau\}$ lets the detector decide for H_0, or no target present. Alternative, results in set $A_1 = \{X > \tau\}$ let the system chose H_1.

It should be obvious that the probability of miss, p_m, and probability of false alarm, p_f, of this system are dependent on v and the chosen threshold τ.

Observe that under H_0, X behaves as $X \sim \mathcal{N}(0, 1)$ yet under H_1, $X \sim \mathcal{N}(v, 1)$. Therefore:

$$p_m = P[A_0 \mid H_1] = P[X \leq \tau \mid H_1] = \Phi(\tau - v) \tag{35}$$

$$p_f = P[A_1 \mid H_0] = P[X > \tau \mid H_0] = 1 - \Phi(\tau) \tag{36}$$

Shown in Figure 5a and 5b are p_m and p_f as functions of τ for different values $v \in \{0, 1, 2\}$. Note that p_f does not depend on v. The ROC curve shows that for $v = 0$ the received signal is identical whether or not a target is present, and that as v increases in value, the detector starts performing better.

Example 2.2. Let a communications system transmit either a 0 with probability p or a 1 with probability $1 - p$.

$$\begin{cases} H_0 : & 0 \text{ received} \\ H_1 : & 1 \text{ received} \end{cases} \tag{37}$$

The detector receives a signal modelled with random variable X:

$$\begin{cases} H_0 : & X = -v + N \\ H_1 : & X = v + N \end{cases} \tag{38}$$

where, $N \sim \mathcal{N}(0, 1)$ denotes the noise of the system and voltage $v \geq 0$ is the information component. Given X what is the minimum probability of error rule for deciding whether a 0 or a 1 was sent?

Observe that when a 0 is transmitted, $X \sim \mathcal{N}(-v, \sigma)$ and when a 1 is transmitted, $X \sim \mathcal{N}(v, \sigma)$. With hypothesis H_i denoting that the send bit was i the following likelihood functions can be derived:

$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x+v)^2}{2\sigma^2}} \tag{39}$$

$$f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-v)^2}{2\sigma^2}} \tag{40}$$

Recall from the beginning that the probability that a 1 is send is defined as $1 - p$ and hence $P[H_0] = p$ and $P[H_1] = 1 - p$. Defining the region A_0 a result must lie in to accept H_0 and A_1 for H_1, the continuous likelihood ratio test, or MAP hypothesis test becomes:

$$x \in A_0, \text{ if } \frac{f_{x|H_0}(x)}{f_{x|H_1}(x)} \geq \frac{P[H_1]}{P[H_0]} \tag{41}$$

Thus, in this example:

$$x \in A_0, \text{ if } \frac{e^{\frac{-(x+v)^2}{2\sigma^2}}}{e^{\frac{-(x-v)^2}{2\sigma^2}}} \geq \frac{1-p}{p} \tag{42}$$

$$x \in A_1, \text{ otherwise} \tag{43}$$

Taking the logarithm and refractoring yields:

$$x \in A_0, \text{ if } x \leq \frac{\sigma^2}{2v} \ln\left(\frac{1-p}{p}\right) \tag{44}$$

$$x \in A_0, \text{ if } x \leq \tau \tag{45}$$

Note that if $p = 0.5$, the threshold $\tau = 0$ and the detector decision is only based on evidence in the signal and not on prior information, which is now information-less. In this case, the detector reverts to a *Maximum Likelihood* classifier using only the information in the signal.

Once $p \neq 0.5$, the prior information as modelled with $P[H_0]$ and $P[H_1]$ will cause the decision threshold for the minimum probability of error to shift. The influence of the priors also depends on the signal-to-noise voltage ratio $\frac{\sigma^2}{2v}$. The higher this ratio, the more reliable the information in the signal and thus it will have more influence on the detector in comparison to the priors.

3 Neyman-Pearson Test

As stated, the MAP test minimizes the the probability of accepting a wrong hypothesis. However, it requires knowledge about the a priori probabilities $P[H_i]$. There are situations in which these a priori probabilities are hard or impossible to obtain. An alternate approach is to first specify an acceptable level for either the probability of miss or false alarm. This idea is the basis for the Neyman-Pearson test.

Neyman-Pearson minimizes p_m subject to the constraint $p_f = \alpha$. As

$$p_f = P[A_1 \mid H_0] \tag{46}$$

$$p_m = P[A_0 \mid H_1] \tag{47}$$

are conditional probabilities the test does not require the a priori probabilities $P[H_0]$ and $P[H_1]$.

Definition 3.1. The Continuous Neyman-Pearson binary hypothesis test. Let sufficient statistic $\mathbf{X}$ a continuous random vector, the rule that minimizes p_m constraint to $p_f = \alpha$ is:

$$\mathbf{x} \in A_0, \text{ if } \mathcal{L}(\mathbf{x}) = \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \gamma \tag{48}$$

$$\mathbf{x} \in A_1, \text{ otherwise,} \tag{49}$$

where for γ holds: $\int_{\mathcal{L}(\mathbf{x})<\gamma} f_{\mathbf{X}|H_0}(\mathbf{x}) d\mathbf{x} = \alpha$.

Definition 3.2. The Discrete Neyman-Pearson binary hypothesis test. Let sufficient statistic $\mathbf{X}$ a discrete random vector, the rule that minimizes p_m constraint to $p_f \leq \alpha$ is:

$$\mathbf{x} \in A_0, \text{ if } \mathcal{L}(\mathbf{x}) = \frac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \gamma \tag{50}$$

$$\mathbf{x} \in A_1, \text{ otherwise,} \tag{51}$$

where γ is the largest value for which holds: $\sum_{\mathcal{L}(\mathbf{x})<\gamma} P_{\mathbf{X}|H_0}(\mathbf{x}) \leq \alpha$.

Example 3.1. The previous example 2.1 is also the result of a Neyman-Pearson test. In this example we will show that the Neyman-Pearson test is a threshold test with acceptance set $A_0 = \{X \leq \tau\}$ where τ is related to $p_f = \alpha$ in the following fashion.

Given H_0, $X \sim \mathcal{N}(0,1)$ and given H_1, $X \sim \mathcal{N}(v,1)$. Following definition 3.1, the Neyman-Pearson test is:

$$X \in A_0 \text{ if } \mathcal{L} = \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} = \frac{e^{\frac{-x^2}{2\sigma^2}}}{e^{\frac{-(x-v)^2}{2\sigma^2}}} \geq \gamma \tag{52}$$

$$X \in A_1 \text{ otherwise} \tag{53}$$

Taking the logarithms:

$$X \in A_0 \text{ if } x \leq \tau = \frac{v}{2} - \frac{1}{v}\ln\gamma \tag{54}$$

$$X \in A_1 \text{ otherwise} \tag{55}$$

The choice of *gamma* has a one-to-one correspondence with the choice of τ. Moreover, $\mathcal{L} \geq \gamma$, if and only if, $x \leq \tau$. In terms of τ, the p_f becomes:

$$p_f = P[\mathcal{L} \leq \gamma \mid H_0] = P[X \geq \tau \mid H_0] = Q(\tau). \tag{56}$$

Then, one can choose τ such that $Q(\tau) = \alpha$.

Exercise

Let there be two hypothesizes, H_0 and H_1:

$$\begin{cases} H_0 : & X = Z \\ H_1 : & X = \mu_1 + Z \end{cases}$$

where $Z \sim \mathcal{N}(0,1)$ and $\mu_1 = 1$.

- Determine the separation threshold τ following the MAP hypothesis, or likelihood ratio test.
- Determine the probability of correct detection p_d. $p_d = 1 - p_m$.

Exercise

Let there be two hypothesises, H_0 and H_1:

$$\begin{cases} H_0 : & X = Z \\ H_1 : & X = \mu_1 + Z \end{cases}$$

where $Z \sim \mathcal{N}(0,1)$ and $\mu_1 = \{0,1,2\}$.

- Derive the general formula's for the Probability of Miss, p_m, the Probability of correct detection, p_d, and the Probability of False Alarm, p_f.
- Implement and visualize the Receiver Operating Characteristic (ROC) curve for the binary hypothesis test for the above given hypothesises, H_0 and H_1 for each different value of $\mu_1 = \{0,1,2\}$.
- Formulate the influence of the mean μ_1 on the separation result, p_d and p_f.

Exercise

A binary communication system transmits a signal X, which follows the *Bernoulli* distribution with $p = 0.5$, i.e $X \sim \mathcal{B}(p)$ and:

$$f_X(x) = \begin{cases} p & x = 1 \\ (1-p) & x = 0 \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

The receiver observes Y:

$$Y = VX + W \tag{58}$$

where $V \perp\!\!\!\perp W \perp\!\!\!\perp X$, and V, W are *exponential* random variables for which $\lambda = 1$:

$$f_V(x) = f_W(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

Given the observation Y, the detector must guess whether $X = 0$ or $X = 1$ was transmitted.

- Formulate hypothesises H_0 and H_1 the detector must decide between.
- Use the binary hypothesis likelihood ratio test to determine the rule that minimises p_{err}, the probability of a decoding error.
- Determine p_{err} for the optimum decision rule.

Hint

Theorem 3.1. If $X_1, X_2, \ldots X_n$ are *i.i.d* exponential random variables, then $W = X_1 + X_2 + \ldots + X_n$ has the *Erlang* probability density function:

$$f_W(w) = \begin{cases} \frac{\lambda^n w^{n-1} e^{-\lambda w}}{(n-1)!} & w \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{60}$$