Introduction to Frequentist Detection, Simple Hypotheses

Aleksandar Dogandžić

April 15, 2017

Contents

READING: [Hero 2015, §7], [Van Trees et al. 2013, §2], [Johnson 2013, §5].

Introduction

CRITICISM of Bayesian approach to detection.

- Requires assigning prior to the parameter θ,
- Only ensures best average performance with respect to selected prior,

- Provides no guaranteed protection against false alarm or miss.

The frequentist approach assumes no priors on $\mathbb{H}_0$ or $\mathbb{H}_1$; hence, we cannot define an average probability of error or risk to minimize. We adopt an alternative criterion: constrain the probability of false alarm and minimize the probability of miss. To find an optimal test that satisfies such a constraint, we need to extend our previous definition of a test function ϕ to allow for randomized decisions:

$$\phi(x) = \begin{cases} 1, & \text{say } \mathbb{H}_1 \\ q, & \text{flip a coin w/ Pr(heads)} = q \\ 0, & \text{say } \mathbb{H}_0 \end{cases}$$

interpreted as

$$\phi(x) = \Pr(\text{say } \mathbb{H}_1 \,|\, \text{observe } x).$$

The probabilities of false alarm and detection are functions of θ:

$$\begin{aligned} \mathrm{E}[\phi(X) \,|\, \theta] &= \int_{\mathcal{X}} \phi(x) f(x \,|\, \theta) \,\mathrm{d}x \\ &= \begin{cases} P_{\mathrm{FA}}(\theta), & \theta \in \mathrm{sp}_{\Theta}(0) \\ P_{\mathrm{D}}(\theta), & \theta \in \mathrm{sp}_{\Theta}(1) \end{cases}. \end{aligned} \tag{1}$$

$f(x \,|\, \theta)$ denotes the likelihood function of θ for measurements x.

Definition 1. A test ϕ is said to be of (false-alarm) level $\alpha \in [0, 1]$ if its probabilities of false alarm are upper-bounded by α:

$$\max_{\theta \in \mathrm{sp}_{\Theta}(0)} P_{\mathrm{FA}}(\theta) \le \alpha. \tag{2}$$

Definition 2. The *power function* of a test ϕ is

$$P_{\mathrm{D}}(\theta) = \beta(\theta) = 1 - P_{\mathrm{M}}(\theta).$$

$\theta \in \mathrm{sp}_{\Theta}(1)$

☞ To increase the probability of detection P_{D} (and reduce the probability of miss P_{M}), we must also allow for the probability of false alarm P_{FA} to increase. This behavior

- represents the fundamental tradeoff in hypothesis testing and detection theory and
- motivates us to introduce the frequentist approach to testing simple hypotheses, pioneered by Neyman and Pearson.

Why upper-bound the false-alarm probability?

Why upper-bound the false-alarm probability?

We give two examples.

For a radar target detection system to process all declared targets, false alarms must be controlled at a level of $\alpha = 10^{-6}$.

The false-alarm level may be specified by a regulatory agency. For example, before a medical diagnostic instrument is allowed to be marketed, the Food and Drug Administration (FDA) may require that any reports of diagnostic effectiveness, i.e., miss rate less than 10 %, have a level of significance of at least $\alpha = 0.01$, i.e., false alarm rate less than 1 %.

Example: Range-gated radar

We listen for a radar return from a perfect reflecting target at a known range from the radar platform. In many radar processing systems, a range gate is applied to filter out all radar returns except for those reflected from an object at a specified distance from the platform. When the attenuation coefficient due to free space electromagnetic propagation is known, then both the amplitude and delay of the return signal are known. In this case, we wish to detect an exactly known signal and can reduce the signal detection problem to a simple hypothesis test of the form:

$$\begin{aligned} \mathbb{H}_0 &: X = W \qquad \text{versus} \\ \mathbb{H}_1 &: X = s + W \end{aligned} \tag{3a}$$

where X is a time sample of the gated radar return, s is the known target return amplitude that would be measured if a target were present and there was no noise, and W is random variable that models measurement noise, which is assumed to have a probability distribution that is also known exactly. Under the assumptions we have made, the hypotheses are simple as the distribution of X is completely known under both $\mathbb{H}_0$ and $\mathbb{H}_1$.

It is useful to write the above problem as testing the value of a parameter θ. Rewrite the radar return measurement as

$$X = \theta s + W$$

where $\theta \in \{0, 1\}$ is a binary unknown constant. The hypotheses (3a) can then be written more directly as

$$\begin{aligned} \mathbb{H}_0 &: \theta = 0 \qquad \text{versus} \\ \mathbb{H}_1 &: \theta = 1. \end{aligned} \tag{3b}$$

Simple Hypotheses

CONSIDER simple hypotheses

$$\mathbb{H}_0 : \quad \theta = \theta_0$$

versus

$$\mathbb{H}_1 : \quad \theta = \theta_1$$

and Neyman-Pearson strategy: find the most powerful (MP) test $\phi^\star$ of level α:

$$\mathrm{E}[\phi^\star(X) \,|\, \theta_1] \geq \mathrm{E}[\phi(X) \,|\, \theta_1]$$

where $\phi(X)$ is any other test of level α.

Neyman-Pearson lemma

THE MP test[1] of level $\alpha \in [0, 1]$ is the randomized likelihood-ratio test:

$$\phi^\star(x) = \begin{cases} 1, & f(x \,|\, \theta_1) > \eta f(x \,|\, \theta_0) \\ q, & f(x \,|\, \theta_1) = \eta f(x \,|\, \theta_0) \\ 0, & f(x \,|\, \theta_1) < \eta f(x \,|\, \theta_0) \end{cases} \tag{4}$$

[1] also known as the Neyman-Pearson test

where the threshold $\eta \geq 0$ and randomization probability q are selected to satisfy

$$\mathrm{E}[\phi^\star(X) \,|\, \theta_0] = \alpha.$$

Proof: We need to show that, for any decision rule $\phi(X)$ satisfying

$$\mathrm{E}[\phi(X) \,|\, \theta_0] \leq \alpha$$

we have

$$\mathrm{E}[\phi^\star(X) \,|\, \theta_1] \geq \mathrm{E}[\phi(X) \,|\, \theta_1]$$

where the threshold η in $\phi^\star$ is chosen to satisfy $\mathrm{E}[\phi^\star(X) \,|\, \theta_0]$.

From (4), we have, for an arbitrary $\eta \geq 0$,

$$\phi^\star(x) \big[f(x \,|\, \theta_1) - \eta f(x \,|\, \theta_0) \big] \geq \phi(x) \big[f(x \,|\, \theta_1) - \eta f(x \,|\, \theta_0) \big] \tag{5}$$

HW: check this

and, consequently,

$$\int_{\mathcal{X}} \phi^\star(x) \big[f(x \,|\, \theta_1) - \eta f(x \,|\, \theta_0) \big] \,\mathrm{d}x \geq \int_{\mathcal{X}} \phi(x) \big[f(x \,|\, \theta_1) - \eta f(x \,|\, \theta_0) \big] \,\mathrm{d}x$$

i.e.,

$$\underbrace{\int_{\mathcal{X}} \phi^\star(x) f(x \,|\, \theta_1) \,\mathrm{d}x}_{\mathrm{E}[\phi^\star(X) \,|\, \theta_1]} - \eta \underbrace{\int_{\mathcal{X}} f(x \,|\, \theta_0) \,\mathrm{d}x}_{\mathrm{E}[\phi^\star(X) \,|\, \theta_0]} \geq \underbrace{\int_{\mathcal{X}} \phi(x) f(x \,|\, \theta_1) \,\mathrm{d}x}_{\mathrm{E}[\phi(X) \,|\, \theta_1]} - \eta \underbrace{\int_{\mathcal{X}} f(x \,|\, \theta_0)] \,\mathrm{d}x}_{\mathrm{E}[\phi(X) \,|\, \theta_0]}$$

rewritten as

$$\mathrm{E}[\phi^\star(X)\,|\,\theta_1] - \mathrm{E}[\phi(X)\,|\,\theta_1] \geq \eta\big(\underbrace{\mathrm{E}[\phi^\star(X)\,|\,\theta_0]}_{\alpha} - \underbrace{\mathrm{E}[\phi(X)\,|\,\theta_0]}_{\leq\alpha}\big)$$
$$\geq 0. \qquad \square$$

Shorthand notation for the likelihood ratio test:

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}\,|\,\theta_1)}{f(\boldsymbol{x}\,|\,\theta_0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \eta. \tag{6}$$

Randomization

THE probability of false alarm of the MP test is

Λ denotes $\Lambda(X)$

iterated expectations

$$\begin{aligned}\mathrm{E}[\phi^\star(X)\,|\,\theta_0] &= \underbrace{\mathrm{E}[\phi^\star(X)\,|\,\Lambda > \eta, \theta_0]}_{1} \Pr\{\Lambda > \eta\,|\,\theta_0\} \\ &\quad + \underbrace{\mathrm{E}\big[\phi^\star(X)\,\big|\,\Lambda = \eta, \theta_0\big]}_{q} \Pr\{\Lambda = \eta\,|\,\theta_0\} \\ &\quad + \underbrace{\mathrm{E}[\phi^\star(X)\,|\,\Lambda < \eta, \theta_0]}_{0} \Pr\{\Lambda < \eta\,|\,\theta_0\} \\ &= \underbrace{\Pr\{\Lambda > \eta\,|\,\theta_0\}}_{1 - F_\Lambda(\eta\,|\,\theta_0)} + q \Pr\{\Lambda = \eta\,|\,\theta_0\}\end{aligned} \tag{7}$$

where $F_\Lambda(\lambda\,|\,\theta) = \Pr\big\{\Lambda \leq \lambda\,|\,\theta\big\}$ is the cumulative distribution function (cdf) of the likelihood ratio Λ.

Randomization must be performed only if it is impossible to find a threshold η that satisfies

$$\Pr\big\{\Lambda > \eta\,|\,\theta_0\big\} = \alpha.$$

This can only occur if the cdf $F_\Lambda(\lambda\,|\,\theta_0)$ has jump discontinuities, i.e., there exist values of Λ with nonzero probability mass:

$$\Pr\big\{\Lambda = t\,|\,\theta_0\big\} > 0$$

implying that $\Lambda = \Lambda(X)$ is not a continuous random variable.

☞ OTHERWISE, if $\Lambda = \Lambda(X)$ given $\theta = \theta_0$ is a continuous random variable, q can be set to zero and randomization is not necessary.

If there is no threshold η that gives $\Pr\big\{\Lambda > \eta\,|\,\theta_0\big\} = \alpha$, select the randomization probability q as follows:

1. Set the threshold η to the smallest value of t for which $\Pr\big\{\Lambda > t\,|\,\theta_0\big\} < \alpha$ and define[2]

$$\alpha^- \triangleq \Pr(\Lambda > \eta\,|\,\theta_0).$$

[2] When there is a jump discontinuity in the cdf, this always exists because all cdfs are right-continuous.

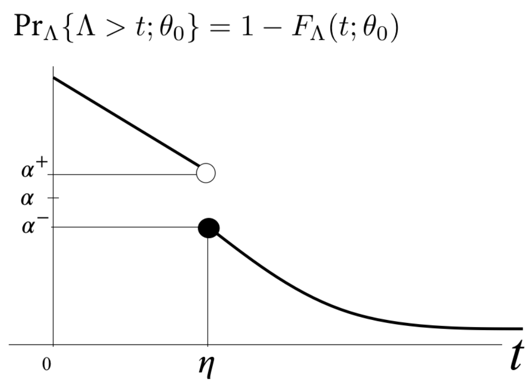

Figure 1: Complementary cdf $\Pr(\Lambda > t \mid \theta_0) = 1 - F_\Lambda(t \mid \theta_0)$ as a function of t.

2. Define

$$\alpha^+ \triangleq \alpha^- + \Pr\{\Lambda = \eta \mid \theta_0\}.$$

Then, from (7), for any value $q \in [0, 1]$, the test (4) will have probability of false alarm

$$\begin{aligned} P_{\text{FA}} &= \mathrm{E}[\phi^\star(X) \mid \theta_0] \\ &= \Pr\{\Lambda > \eta \mid \theta_0\} + q \Pr\{\Lambda = \eta \mid \theta_0\} \\ &= \alpha^- + q(\alpha^+ - \alpha^-). \end{aligned}$$

see (7)

Set $P_{\text{FA}} = \alpha$ and solve for q:

$$q = \frac{\alpha - \alpha^-}{\alpha^+ - \alpha^-}.$$

✻ COMMENTS.

- If we select threshold η^- (slightly lower than η), we will have false-alarm probability $\Pr(\Lambda > \eta^- \mid \theta_0) = \alpha^+$.
- If we select threshold η^+ (slightly lower than η), $\Pr(\Lambda > \eta^+ \mid \theta_0) = \alpha^-$.

See Fig. 1.

The likelihood-ratio test statistic

$$\frac{f(x \mid \theta_1)}{f(x \mid \theta_0)}$$

is also the Bayes' decision rule test statistic for simple hypotheses.

If $T = T(X)$ is a sufficient statistic for θ, the likelihood-ratio test statistic depends on the measurements X only through $T(X)$; indeed, by the factorization theorem,

recall the factorization theorem from handout `sufficiency`

$$f(x \mid \theta) = g(T(x), \theta) h(x).$$

$$\frac{f(x \mid \theta_1)}{f(x \mid \theta_0)} = \frac{g\big(T(x), \theta_1\big)}{g\big(T(x), \theta_0\big)} \triangleq \Lambda(\boldsymbol{T}).$$

likelihood-ratio test statistic is a function of sufficient statistic T

☞ WE can formulate the likelihood-ratio test based on probability density function (pdf) of T instead of the pdf of the entire measurement set X.

The receiver operating characteristic (ROC) allows us to visualize achievable $P_{\mathrm{FA}}(\theta_0)$ and $P_{\mathrm{D}}(\theta_1)$. A point $(P_{\mathrm{FA}}, P_{\mathrm{D}})$ is in the shaded region if we can find a rule $\phi(X)$ such that its probabilities of false alarm and detection are $P_{\mathrm{FA}}(\theta_0) = P_{\mathrm{FA}}$ and $P_{\mathrm{D}}(\theta_1) = P_{\mathrm{D}}$, see Fig. 2.

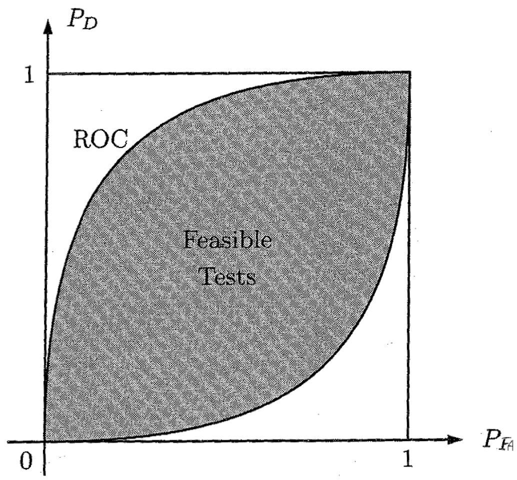

Figure 2: Receiver operating characteristic.

Coherent Detection in Gaussian Noise

we considered this testing problem in handout `Bayesdetex` from Bayesian detection perspective

THE measurement vector $\boldsymbol{X}$ given $\boldsymbol{\mu}$ is modeled as

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, C) = \frac{1}{\sqrt{\det(2\pi C)}} \exp\left[-0.5(\boldsymbol{x}-\boldsymbol{\mu})^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

where C is a known positive definite covariance matrix. The space of the parameter $\boldsymbol{\mu}$ and its partitions are

$$\mathrm{sp}_{\boldsymbol{\mu}} = \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_1\}, \qquad \mathrm{sp}_{\boldsymbol{\mu}}(0) = \{\boldsymbol{\mu}_0\}, \qquad \mathrm{sp}_{\boldsymbol{\mu}}(1) = \{\boldsymbol{\mu}_1\}.$$

Our likelihood ratio test is

$$\frac{f(\boldsymbol{x} \mid \boldsymbol{\mu}_1)}{f(\boldsymbol{x} \mid \boldsymbol{\mu}_0)} = \frac{\exp\left[-0.5(\boldsymbol{x}-\boldsymbol{\mu}_1)^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right]}{\exp\left[-0.5(\boldsymbol{x}-\boldsymbol{\mu}_0)^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right]} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \eta. \tag{8}$$

Therefore,

$$-0.5(\boldsymbol{x}-\boldsymbol{\mu}_1)^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + 0.5(\boldsymbol{x}-\boldsymbol{\mu}_0)^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0) \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \ln \eta$$

i.e., $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T C^{-1}\left[\boldsymbol{x} - 0.5(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)\right] \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \ln \eta$ and, finally,

$$T(\boldsymbol{x}) = \boldsymbol{s}^T C^{-1} \boldsymbol{x} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \ln \eta + 0.5(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T C^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) \triangleq \gamma$$

where we have defined

$$\boldsymbol{s} \triangleq \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0.$$

Probabilities of false alarm and detection/miss

GIVEN $\boldsymbol{\mu}$, $T(\boldsymbol{X})$ is a linear combination of Gaussian random variables, implying that it is also Gaussian, with mean and variance:

$$\begin{aligned} \mathrm{E}\big[T(\boldsymbol{X}) \,|\, \boldsymbol{\mu}\big] &= \boldsymbol{s}^T C^{-1} \boldsymbol{\mu} \\ \mathrm{var}\big[T(\boldsymbol{x}) \,|\, \boldsymbol{\mu}\big] &= \boldsymbol{s}^T C^{-1} \boldsymbol{s}. \end{aligned}$$

not a function of $\boldsymbol{\mu}$

* PROBABILITY of false alarm.

$$\begin{aligned} P_{\mathrm{FA}} &= \Pr_{\boldsymbol{X}|\boldsymbol{\mu}} \{T(\boldsymbol{X}) > \gamma \,|\, \boldsymbol{\mu}_0\} \\ &= \Pr_{\boldsymbol{X}|\boldsymbol{\mu}} \left\{ \frac{T(\boldsymbol{X}) - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} > \frac{\gamma - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} \,\middle|\, \boldsymbol{\mu}_0 \right\} \\ &= Q\left(\frac{\gamma - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} \right) \end{aligned} \tag{9}$$

$\left\{ \frac{T(\boldsymbol{X}) - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} \,\middle|\, \boldsymbol{\mu}_0 \right\}$ is a standard normal random variable

* PROBABILITY of detection.

$$\begin{aligned} P_{\mathrm{D}} &= 1 - P_{\mathrm{M}} \\ &= \Pr_{\boldsymbol{X}|\boldsymbol{\mu}} \{T(\boldsymbol{X}) > \gamma \,|\, \boldsymbol{\mu}_1\} \\ &= \Pr_{\boldsymbol{X}|\boldsymbol{\mu}} \left\{ \frac{T(\boldsymbol{X}) - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} > \frac{\gamma - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} \,\middle|\, \boldsymbol{\mu}_1 \right\} \\ &= Q\left(\frac{\gamma - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} \right). \end{aligned} \tag{10}$$

$\left\{ \frac{T(\boldsymbol{X}) - \boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} \,\middle|\, \boldsymbol{\mu}_1 \right\}$ is a standard normal random variable

We use (9) to obtain the threshold γ that satisfies the specified P_{FA}:

$$\frac{\gamma}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}} = Q^{-1}(P_{\mathrm{FA}}) + \frac{\boldsymbol{s}^T C^{-1} \boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}}.$$

Then,

$$P_{\mathrm{D}} = Q\Big(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{\boldsymbol{s}^T C^{-1} \boldsymbol{s}}\Big) = Q\Big(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{d^2}\Big) \tag{11}$$

where

$$d^2 = \boldsymbol{s}^T C^{-1} \boldsymbol{s} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T C^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

is the deflection coefficient, see Fig. 3.

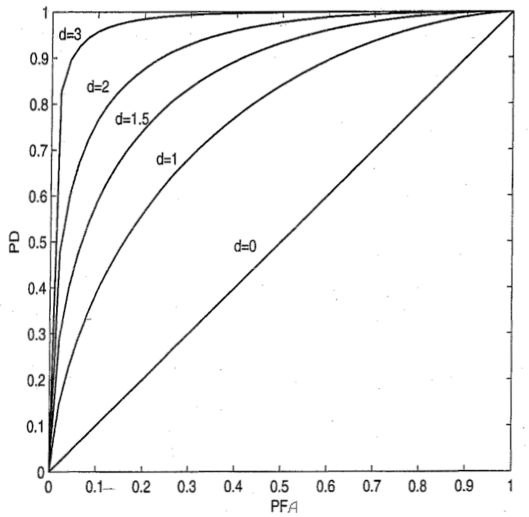

Figure 3: ROC curves for different deflections.

Decentralized Detection for Simple Hypotheses

see [Varshney 1996]

CONSIDER the decentralized detection scenario depicted in Fig. 4.

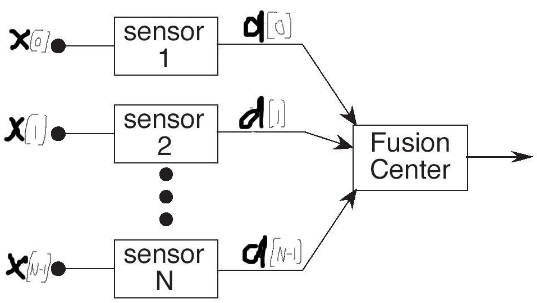

Figure 4: A decentralized detection scheme.

ASSUMPTIONS:

- The observations $\left(X[n]\right)_{n=0}^{N-1}$ made at N spatially distributed sensors (nodes) follow the same probabilistic model:

$$f(x[n] \mid \theta) \tag{12}$$

  and are independent given θ, which may not always be reasonable, but leads to an easy solution.

- We wish to test:

$$\mathbb{H}_0 : \theta = \theta_0 \quad \text{versus}$$
$$\mathbb{H}_1 : \theta = \theta_1.$$

- Each node n makes a *hard local decision* $d[n]$ based on its local observation $x[n]$ and sends it to the *headquarters* (fusion center), which collects all the local decisions and makes the final *global decision* $\mathbb{H}_0$ versus $\mathbb{H}_1$. This structure is clearly suboptimal: it is easy to construct a better decision strategy in which each node sends its[3] likelihood ratio to the fusion center, rather than the decision only. However, such a strategy would have a higher communication (energy) cost.

[3] quantized, in practice

- The probabilities of false alarm $\left(P_{\mathrm{FA},n}\right)_{n=0}^{N-1}$ and detection $\left(P_{\mathrm{D},n}\right)_{n=0}^{N-1}$ of each node's local decision rules can be computed using (12).

We now discuss the decentralized detection problem. The distributions of local decisions $D[n]$ under the two hypotheses are Bernoulli probability mass functions (pmfs):

$p(d[n];\theta)$ denotes the likelihood function of θ for measurement $d[n]$.

$$\begin{aligned} p(d[n] \mid \theta_1) &= \mathrm{Bin}\left(d[n] \mid 1, P_{\mathrm{D},n}\right) \\ &= P_{\mathrm{D},n}^{d[n]}\left(1 - P_{\mathrm{D},n}\right)^{1-d[n]} \end{aligned}$$

and, similarly,

$$\begin{aligned} p(d[n] \mid \theta_0) &= \mathrm{Bin}\left(d[n] \mid 1, P_{\mathrm{FA},n}\right) \\ &= P_{\mathrm{FA},n}^{d[n]}\left(1 - P_{\mathrm{FA},n}\right)^{1-d[n]}. \end{aligned}$$

Use independence of local decisions $D[n]$ and treat them as measurements to obtain the global decision as the following likelihood-ratio test:

$$\begin{aligned} \ln \Lambda(\boldsymbol{d}) &= \sum_{n=0}^{N-1} \ln\left[\frac{p(d[n] \mid \theta_1)}{p(d[n] \mid \theta_0)}\right] \\ &= \sum_{n=0}^{N-1} \ln\left[\frac{P_{\mathrm{D},n}^{d[n]}(1 - P_{\mathrm{D},n})^{1-d[n]}}{P_{\mathrm{FA},n}^{d[n]}(1 - P_{\mathrm{FA},n})^{1-d[n]}}\right] \\ &\underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \ln \tau \end{aligned}$$

where $\boldsymbol{d} = \left(d[n]\right)_{n=0}^{N-1}$, which can be simplified to

$$\sum_{n=0}^{N-1} d[n] \, \underbrace{\ln\left[\frac{P_{\mathrm{D},n}\left(1 - P_{\mathrm{FA},n}\right)}{P_{\mathrm{FA},n}\left(1 - P_{\mathrm{D},n}\right)}\right]}_{\text{weight of the decision } d[n]} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \gamma.$$

If $P_{\mathrm{D},n} > P_{\mathrm{FA},n}$, then the corresponding weight multiplying $d[n]$ is positive. If $P_{\mathrm{D},n} < P_{\mathrm{FA},n}$, then the weight that multiplies $d[n]$ is negative and we are better off inverting the decision $d[n]$. If $P_{\mathrm{D},n} = P_{\mathrm{FA},n}$, the weight multiplying $d[n]$ is zero, i.e., the decision $d[n]$ is irrelevant.

✻ IDENTICAL sensor performances. We now focus on the case where all sensors have identical performance:

$$P_{\mathrm{D},n} = P_{\mathrm{D}}, \qquad P_{\mathrm{FA},n} = P_{\mathrm{FA}}$$

i.e., all all sensors have independent, identically distributed (i.i.d.) measurements and identical local decision thresholds. Define the number of sensors deciding locally to say $\mathbb{H}_1$:

$$u_1 \triangleq \sum_{n=0}^{N-1} d[n].$$

Then,

$$\begin{aligned} \ln \Lambda(\boldsymbol{d}) &= u_1 \ln\left(\frac{P_{\mathrm{D}}}{P_{\mathrm{FA}}}\right) + (N - u_1) \ln\left(\frac{1-P_{\mathrm{D}}}{1-P_{\mathrm{FA}}}\right) \\ &\underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \ln \eta \end{aligned}$$

or

$$u_1 \ln\left[\frac{P_{\mathrm{D}}(1-P_{\mathrm{FA}})}{P_{\mathrm{FA}}(1-P_{\mathrm{D}})}\right] \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \ln \eta + N \ln\left(\frac{1-P_{\mathrm{FA}}}{1-P_{\mathrm{D}}}\right). \tag{13}$$

Assume $P_{\mathrm{D}} > P_{\mathrm{FA}}$, which implies

$$\frac{P_{\mathrm{D}}(1-P_{\mathrm{FA}})}{P_{\mathrm{FA}}(1-P_{\mathrm{D}})} > 1. \tag{14}$$

Hence, the logarithm of (14) is positive, and the decision rule (13) further simplifies to

$$u_1 \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \gamma. \tag{15}$$

The performance analysis of this detector is easy: the random variable U_1 is binomial given θ and, therefore,

$$\Pr\{U_1 = u_1 \mid \theta\} = \binom{N}{u_1} p^{u_1} (1-p)^{N-u_1}$$

where $p = P_{\mathrm{FA}}$ for $\theta = \theta_0$ and $p = P_{\mathrm{D}}$ for $\theta = \theta_1$. Hence, the probabilities of false alarm and detection for the global test are

$$\begin{aligned} P_{\mathrm{FA,global}} &= \Pr\{U_1 > \gamma \mid \theta_0\} \\ &= \sum_{u_1=\lceil\gamma\rceil}^{N} \binom{N}{u_1} P_{\mathrm{FA}}^{u_1} (1-P_{\mathrm{FA}})^{N-u_1} \\ P_{\mathrm{D,global}} &= \Pr\{U_1 > \gamma \mid \theta_1\} \\ &= \sum_{u_1=\lceil\gamma\rceil}^{N} \binom{N}{u_1} P_{\mathrm{D}}^{u_1} (1-P_{\mathrm{D}})^{N-u_1}. \end{aligned}$$

To design an MP test for a specific global test level α and, we may need to randomize the likelihood ratio test in (15).

ROC Curves

All threshold tests have probabilities of false alarm and detection indexed by a threshold parameter η.

ROC is the plot of P_D as a function of P_{FA}, see Fig. 5.

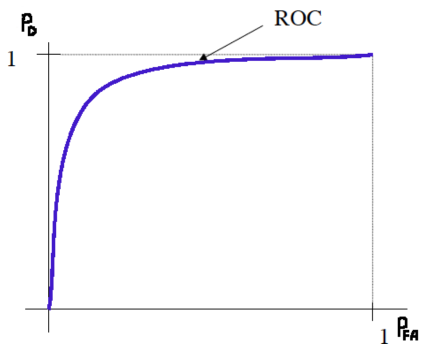

Figure 5: A typical ROC curve.

Properties

We describe basic properties of ROCs.

⊛ Coin-flip detector: For the coin flip detector with $\phi(x) = q$ (independent of data), the ROC is a diagonal line with unit slope, see Fig. 6.

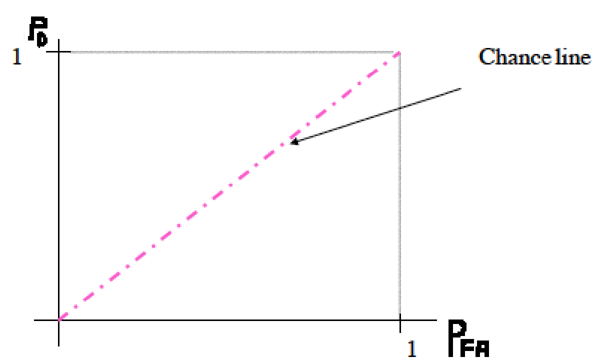

Figure 6: ROC curve for coin flip detector.

❊ THE ROC of any MP test always lies above the diagonal, see Fig. 7.

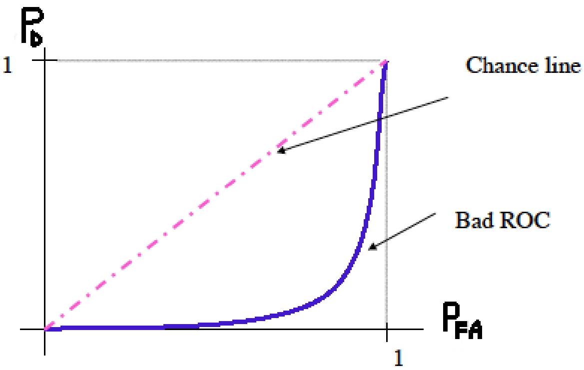

Figure 7: MP tests are unbiased.

Definition 3. A test ϕ is *unbiased* if its power (probability of detection) P_{D} is at least as great as its significance level (probability of false alarm) P_{FA}:

$$P_{\mathrm{D}} \geq P_{\mathrm{FA}}.$$

❊ THE ROC of any MP test is concave, see Fig. 8.

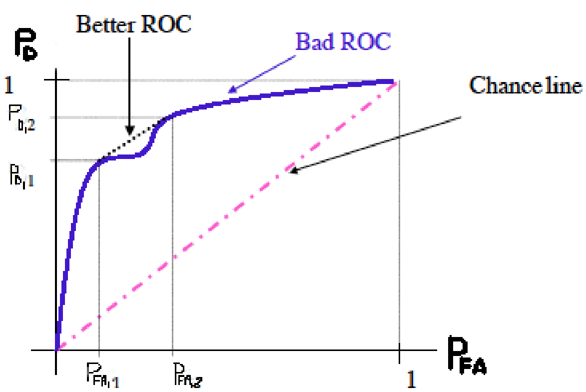

Figure 8: ROC of any MP test is concave. A test with non-convex ROC (thick line) can always be improved by randomization which has effect of connecting two endpoints.

To see concavity, define the significance level and power of tests ϕ_1 and ϕ_2 by $(P_{\mathrm{FA},1}, P_{\mathrm{D},1})$ and $(P_{\mathrm{FA},2}, P_{\mathrm{D},2})$, respectively. Define the test

$$\phi_{12} = p\phi_1 + (1-p)\phi_2$$

implemented by selecting ϕ_1 and ϕ_2 at random with probabilities p and $1-p$, respectively. The significance level and power of this test are

$$\begin{aligned}
P_{\mathrm{FA},12} &= \mathrm{E}\big[\phi_{12}(\boldsymbol{X}) \,|\, \theta_0\big] \\
&= p\,\mathrm{E}\big[\phi_1(\boldsymbol{X}) \,|\, \theta_0\big] + (1-p)\,\mathrm{E}\big[\phi_2(\boldsymbol{X}) \,|\, \theta_0\big] \\
&= pP_{\mathrm{FA},1} + (1-p)P_{\mathrm{FA},2} \\
P_{\mathrm{D},12} &= \mathrm{E}\big[\phi_{12}(\boldsymbol{X}) \,|\, \theta_1\big] \\
&= p\,\mathrm{E}\big[\phi_1(\boldsymbol{X}) \,|\, \theta_1\big] + (1-p)\,\mathrm{E}\big[\phi_2(\boldsymbol{X}) \,|\, \theta_1\big] \\
&= pP_{\mathrm{D},1} + (1-p)P_{\mathrm{D},2}
\end{aligned}$$

see (1). Thus, as p varies between 0 and 1, the performance $(P_{\text{FA},12}, P_{\text{D},12})$ of ϕ_{12} varies along the straight line connecting the points $(P_{\text{FA},1}, P_{\text{D},1})$ and $(P_{\text{FA},2}, P_{\text{D},2})$.

⁎ If ROC curve is differentiable, the MP likelihood-ratio test threshold needed to attain any pair $(P_{\text{FA}}, P_{\text{D}})$ on its ROC can be found graphically as the slope of ROC at the point P_{FA}:

$$\eta = \frac{\mathrm{d}\, P_{\text{D}}(P_{\text{FA}})}{\mathrm{d}\, P_{\text{FA}}}.$$

See [Van Trees et al. 2013, prop. 3 in §2.2.2].

Proof. Indeed, from (6), we have

$$P_{\text{FA}}(\eta) = \int_{\eta}^{+\infty} f_{\Lambda}(\lambda \,|\, \theta_0)\, \mathrm{d}\lambda, \qquad P_{\text{D}}(\eta) = \int_{\eta}^{+\infty} f_{\Lambda}(\lambda \,|\, \theta_1)\, \mathrm{d}\lambda.$$

Differentiate these expressions with respect to η:

$$\frac{\mathrm{d}\, P_{\text{FA}}(\eta)}{\mathrm{d}\eta} = -f_{\Lambda}(\eta \,|\, \theta_0), \qquad \frac{\mathrm{d}\, P_{\text{D}}(\eta)}{\mathrm{d}\eta} = -f_{\Lambda}(\eta \,|\, \theta_1).$$

Note that

$$\begin{aligned} P_{\text{D}}(\eta) &= \int_{\mathcal{X}_1^\star} \Lambda(x) f(x \,|\, \theta_0)\, \mathrm{d}x \\ &= \int_{\eta}^{+\infty} \lambda f(\lambda \,|\, \theta_0)\, \mathrm{d}\eta \end{aligned} \tag{16}$$

$\mathcal{X}_1^\star = \{x : \phi^\star(x) = 1\}$, see (4)

and differentiate:

$$\frac{\mathrm{d}\, P_{\text{D}}(\eta)}{\mathrm{d}\eta} = -\eta f(\lambda \,|\, \theta_0).$$

Hence,

$$\frac{\mathrm{d}\, P_{\text{D}}(\eta)}{\mathrm{d}\eta} = -\eta f(\lambda \,|\, \theta_0) = -f_{\Lambda}(\eta \,|\, \theta_1)$$

implying

$$\frac{f_{\Lambda}(\eta \,|\, \theta_1)}{f_{\Lambda}(\eta \,|\, \theta_0)} = \eta.$$

Finally,

$$\frac{\frac{\mathrm{d}\, P_{\text{D}}(\eta)}{\mathrm{d}\eta}}{\frac{\mathrm{d}\, P_{\text{FA}}(\eta)}{\mathrm{d}\eta}} = \frac{f_{\Lambda}(\eta \,|\, \theta_1)}{f_{\Lambda}(\eta \,|\, \theta_0)} = \eta$$

i.e.,

$$\frac{\mathrm{d}\, P_{\text{D}}}{\mathrm{d}\, P_{\text{FA}}} = \eta$$

and (16) follows. □

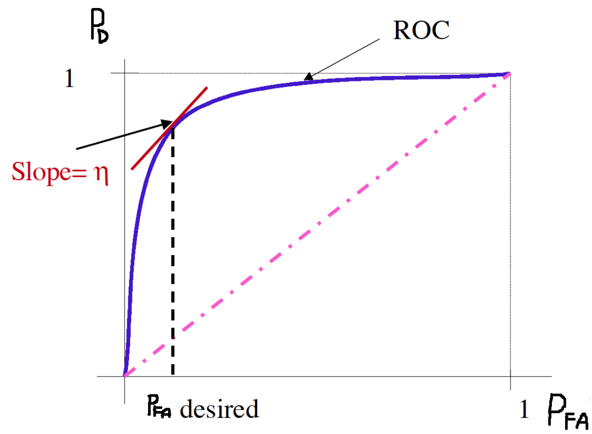

Figure 9: Threshold of MP likelihood ratio test can be found by differentiating the ROC curve.

Example: Test against uniform density

CONSIDER testing between $\mathbb{H}_0$ and $\mathbb{H}_1$ where $f(x \mid \mathbb{H}_0)$ and $f(x \mid \mathbb{H}_1)$ are two densities shown in Fig. 10. Find the MP test.

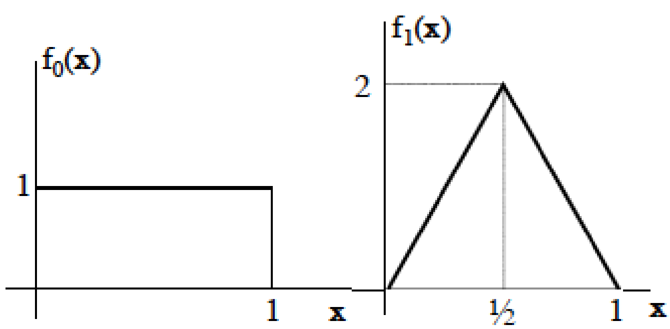

Figure 10: Two densities to be tested.

The MP likelihood-ratio test is

$$\Lambda(x) = \frac{f(x \mid \mathbb{H}_1)}{f(x \mid \mathbb{H}_0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \eta.$$

For a given threshold η, the $\mathbb{H}_1$ decision region is

$$\mathcal{X}_1 = \begin{cases} x \in [\eta/4, 1-\eta/4], & 0 \le \eta \le 2 \\ \text{empty}, & \eta > 2 \end{cases}$$

see Fig. 11.

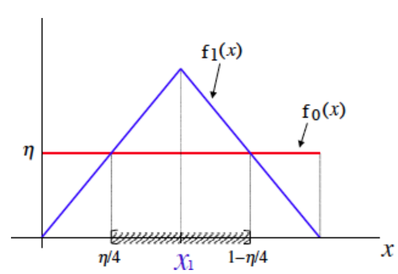

Figure 11: Region $\mathcal{X}_1$ for which MP likelihood ratio test decides $\mathbb{H}_1$ is the set of values x for which triangle exceeds horizontal line of height η.

✻ SETTING threshold. Select η to meet the constraint $P_{\text{FA}} = \alpha$. Assume $0 \le \eta \le 2$.

$$\begin{aligned} P_{\text{FA}} &= \alpha \\ &= \Pr\{X \in \mathcal{X}_1 \mid \mathbb{H}_0\} \\ &= \int_{\eta/4}^{1-\eta/4} f(x \mid \mathbb{H}_0)\, \mathrm{d}x \\ &= 1 - \eta/2 \end{aligned}$$

implying

$$\eta = 2(1-\alpha) \tag{17}$$

and no randomization is required.

⊛ Power of the MP likelihood ratio test:

$$\begin{aligned} P_{\mathrm{D}} &= \beta \\ &= \Pr\{X \in \mathcal{X}_1 \mid \mathbb{H}_1\} \\ &= \int_{\eta/4}^{1-\eta/4} f(x \mid \mathbb{H}_1)\,\mathrm{d}x \\ &= 2\int_{\eta/4}^{0.5} f(x \mid \mathbb{H}_1)\,\mathrm{d}x \\ &= 2\int_{\eta/4}^{0.5} 4x\,\mathrm{d}x \\ &= 1-\eta^2/4 \end{aligned} \tag{18}$$

Plug in the level-α threshold (17) to the power expression (18) to obtain the ROC curve

$$\begin{aligned} P_{\mathrm{D}} = \beta &= 1-(1-\alpha)^2 \\ &= 1-(1-P_{\mathrm{FA}})^2. \end{aligned}$$

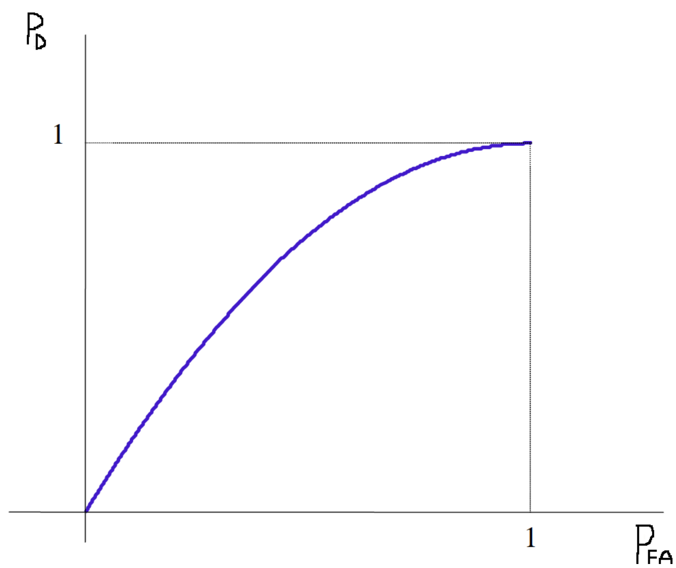

Figure 12: ROC curve for uniform versus triangle pdf example.

Detecting an Increase in Poisson Rate

We measure the number of photons X collected by a charge-coupled-device (CCD) array over a certain period of time. We adopt the common Poisson measurement model for X:

$$\{X[n] \mid \lambda\} \sim \text{Poisson}(\lambda).$$

In ambient conditions, the average number of photons $\lambda_0 > 0$ incident on the array is fixed and known.[4]

[4] λ_0 is sometimes called the dark-current rate.

When a known source of photons is present, the photon rate increases to a known value λ_1, where $\lambda_1 > \lambda_0$. The goal of the photodetector is to detect the presence of the source based on measuring $X = x$. Hence, we wish to test

$$\begin{aligned} \mathbb{H}_0 : &\qquad \lambda = \lambda_0 \\ \text{versus} & \\ \mathbb{H}_1 : &\qquad \lambda = \lambda_1. \end{aligned}$$

We design an MP test of prescribed level $\alpha \in [0, 1]$. We know that the MP test is a likelihood ratio test:

$$\begin{aligned} \frac{f(x \mid \lambda_1)}{f(x \mid \lambda_0)} &= \frac{\text{Poisson}(x \mid \lambda_1)}{\text{Poisson}(x \mid \lambda_0)} \\ &= \frac{\frac{\lambda_1^x}{x!} e^{-\lambda_1}}{\frac{\lambda_0^x}{x!} e^{-\lambda_0}} \\ &= \left(\frac{\lambda_1}{\lambda_0}\right)^x e^{\lambda_0 - \lambda_1} \\ &\underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \eta \end{aligned}$$

which simplifies to

$$x \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \gamma.$$

take log and use the fact that $\lambda_1 > \lambda_0$

Setting the decision threshold

FIRST try to set the threshold γ without randomization:

$$\begin{aligned} P_{\text{FA}} &= \alpha \\ &= \Pr\{X > \gamma \mid \lambda_0\} \\ &= 1 - F_X(\gamma \mid \lambda_0) \end{aligned}$$

where $F_X(x \mid \lambda)$ denotes the cdf of a Poisson random variable with rate λ, see Fig. 13.

As the Poisson cdf is not continuous, only a discrete set of test levels is attainable by the nonrandomized likelihood ratio test: $\alpha_1, \alpha_2, \ldots$, see Fig. 13.

Suppose that the desired test level is $\alpha \in (\alpha_i, \alpha_{i+1})$. Then, we need to randomize the likelihood ratio test by selecting the threshold γ and randomization probability q to satisfy:

see (7)

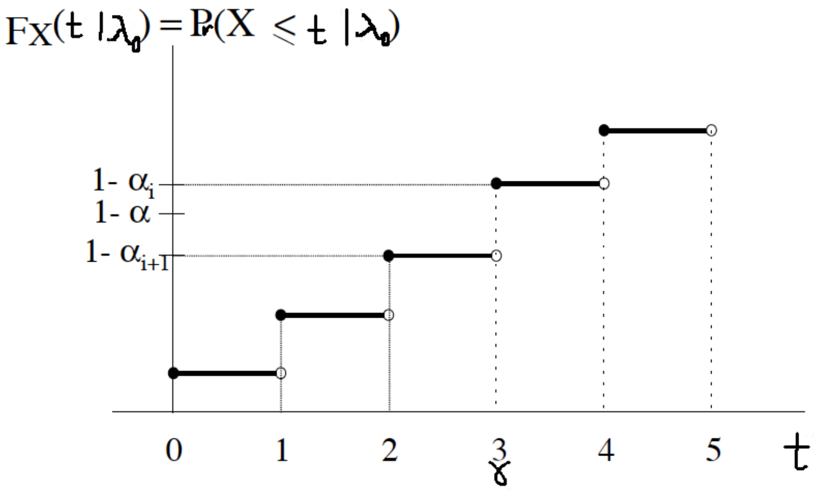

Figure 13: Cdf of a Poisson random variable that we need to set the decision threshold is a staircase function.

$$\alpha = \Pr\{X > \gamma \,|\, \lambda_0\} + q \Pr\{X = \gamma \,|\, \lambda_0\}.$$

Pick γ as smallest t such that $\Pr\{X > t \,|\, \lambda_0\} = \alpha_i$, i.e., the smallest t such that $F_X(\gamma \,|\, \lambda_0) = 1 - \alpha_i$, see Fig. 13. Then,

$$\alpha = \underbrace{\Pr\{X > \gamma \,|\, \lambda_0\}}_{\alpha_i} + q \underbrace{\Pr\{X = \gamma \,|\, \lambda_0\}}_{\alpha_{i+1} - \alpha_i}.$$

and solve for q:

$$q = \frac{\alpha - \alpha_i}{\alpha_{i+1} - \alpha_i}.$$

With these settings, the power of the randomized MP likelihood-ratio test is simply

$$\begin{aligned} P_{\mathrm{D}} = \beta &= \Pr\{X > \gamma \,|\, \lambda_1\} + q \Pr\{X = \gamma \,|\, \lambda_1\} \\ &= \Pr\{X > \gamma \,|\, \lambda_1\} + \frac{\alpha - \alpha_i}{\alpha_{i+1} - \alpha_i} \Pr\{X = \gamma \,|\, \lambda_1\} \end{aligned}$$

use iterated expectations or total probability, similar to (7)

plotted as an ROC curve in Fig. 14.

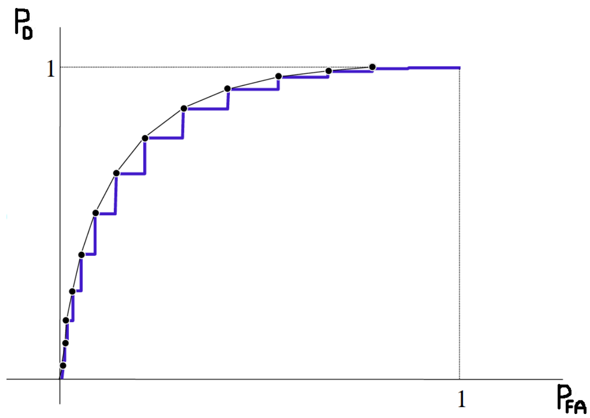

Figure 14: ROCs of randomized (smooth curve) and non-randomized (staircase curve) likelihood ratio tests for detecting an increase in rate of a Poisson random variable.

On-off Keying in Gaussian Noise

THE measurement X is modeled as

$$X = \theta + W$$

where θ is either zero (if the transmitted bit is zero) or one (if the transmitted bit is one). Hence, the decoder decides

$$\mathbb{H}_0\colon \theta = 0 \qquad \text{versus}$$
$$\mathbb{H}_1\colon \theta = 1.$$

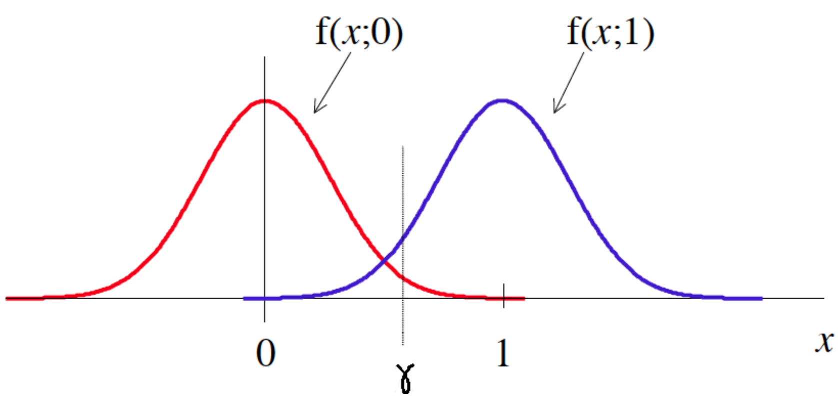

Figure 15: Densities under $\mathbb{H}_0$ and $\mathbb{H}_1$ for on-off keying detection.

Assume $W \sim \mathcal{N}(0, 1)$. The likelihood-ratio test statistic is

$$\begin{aligned}\Lambda(x) &= \frac{\mathcal{N}(x \mid 1, 1)}{\mathcal{N}(x \mid 0, 1)} \\ &= \frac{\frac{1}{\sqrt{2\pi}} \exp\left[-0.5(x-1)^2\right]}{\frac{1}{\sqrt{2\pi}} \exp(-0.5x^2)} = \exp(x - 0.5) \qquad (19)\end{aligned}$$

and the likelihood-ratio test is

$$x \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \gamma$$

with

$$\begin{aligned}\Pr(X > \gamma \mid \theta) &= \Pr(X - \theta > \gamma - \theta \mid \theta) \\ &= Q(\gamma - \theta)\end{aligned}$$

$X - \theta$ is a standard normal random variable

which implies

$$P_{\mathrm{FA}} = Q(\gamma) \tag{19a}$$

$$\begin{aligned}P_{\mathrm{D}} &= Q(\gamma - 1) \\ &= Q\big(Q^{-1}(P_{\mathrm{FA}}) - 1\big).\end{aligned} \tag{19b}$$

The curve (19b) is shown in Fig. 16 along with operating points for the maximum-likelihood (ML) Bayes' decision rule[5]

$$x \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} 0.5$$

[5] assuming 0-1 loss and equiprobable hypotheses, leading to $\Lambda(x) \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} 1$, see (19)

and MP likelihood ratio test with level $\alpha = 0.001$:

$$x \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} 2.329.$$

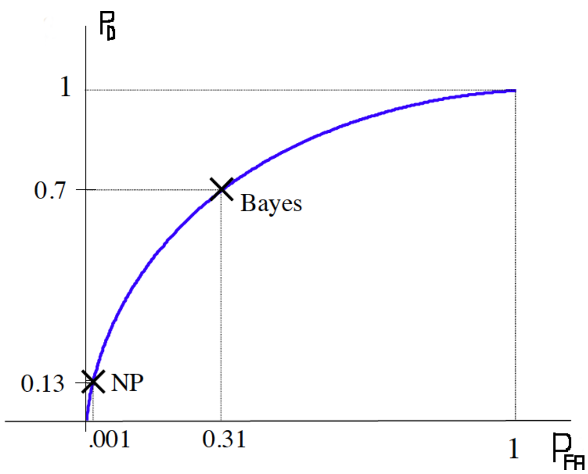

Figure 16: ROC curve for the on-off keying example.

Assuming equiprobable hypotheses, the ML test achieves the minimum average error probability

$$\begin{aligned}\text{av. error probability for ML test} &= 0.5P_{\mathrm{FA}} + 0.5(\underbrace{1 - P_{\mathrm{D}}}_{P_{\mathrm{M}}}) \\ &\approx 0.31\end{aligned}$$

whereas the MP likelihood ratio test with level $\alpha = 0.001$ achieves the average error probability

$$\begin{aligned}\text{av. error probability for MP test} &= 0.5P_{\text{FA}} + 0.5P_{\text{M}} \\ &\approx 0.45.\end{aligned}$$

P-values

Definition 4. Size of a hypothesis test $\phi(\boldsymbol{X})$ is defined as its largest possible false-alarm probability:

$$\max_{\theta \in \text{sp}_{\Theta}(0)} \mathrm{E}(\phi(\boldsymbol{X}) \mid \theta). \tag{20}$$

☞ A hypothesis test is *said to have level α* if its size is less than or equal to α, see Definition 1. Therefore, a level-α test is guaranteed to have a false-alarm probability less than or equal to α.

Our general approach in this handout has been to set α in advance and to make a hard binary decision ("accept $\mathbb{H}_0$" or "accept $\mathbb{H}_1$") depending on α. Such hard decisions do not convey information about how close we were to the opposite decision.

☞ If $\mathbb{H}_1$ is accepted for a certain specified α, it will be accepted for $\alpha' > \alpha$. Therefore, there exists a smallest α at which $\mathbb{H}_1$ is accepted. This motivates the introduction of the p-value.

Definition 5. Suppose that, for every α, we have a size-α rule ϕ_α:

$$\text{rule } \phi_\alpha: \qquad \mathcal{X}_{0,\alpha} = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 0\}, \qquad \mathcal{X}_{1,\alpha} = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 1\}$$

meaning that,

$$\alpha = \max_{\theta \in \text{sp}_{\Theta}(0)} \Pr_{X \mid \Theta}(X \in \mathcal{X}_{1,\alpha} \mid \theta).$$

Then, the p-value for this test is the smallest size α for which we can declare $\mathbb{H}_1$:

$$p\text{-value} = \inf\{\alpha \mid \boldsymbol{x} \in \mathcal{X}_{1,\alpha}\}.$$

Informally, the p-value is a measure of evidence for supporting $\mathbb{H}_1$. For example, p-values less than 0.01 are considered *strong evidence* supporting $\mathbb{H}_1$.

There are many misconceptions and warnings regarding p-values. Here are the most important ones.

☞ Warning: A large p-value is *not* strong evidence in favor of $\mathbb{H}_0$; a large p-value can occur for two reasons:

i) $\mathbb{H}_0$ is true or

ii) $\mathbb{H}_0$ is false but the test has low detection probability (power).

☞ WARNING: *Do not* confuse the p-value with $\Pr_{\Theta|X}\{\mathbb{H}_0 \mid x\}$ used in Bayesian inference. The p-value is *not* the probability that $\mathbb{H}_0$ is true.

Theorem 1. *Suppose that the size-α test is of the form:*

$$\text{declare } \mathbb{H}_1 \text{ if and only if } T(x) \geq c_\alpha.$$

c_α is a threshold

Then, the p-value for this test is

$$p\text{-value} = \max_{\theta \in \text{sp}_\Theta(0)} \Pr_{X \mid \Theta}\{T(X) \geq T(x) \mid \theta\}$$

where x is the observed value of X.

For simple null hypothesis $\text{sp}_\Theta(0) = \{\theta_0\}$,

$$p\text{-value} = \Pr_{X \mid \Theta}\{T(X) \geq T(x) \mid \theta_0\}.$$

In words, Theorem 1 states that

☞ THE p-value is the probability that, under $\mathbb{H}_0$, a random measurement realization $\boldsymbol{X}$ is observed yielding a value of the test statistic $T(\boldsymbol{X})$ that is greater than or equal to what has actually been observed, which is $T(\boldsymbol{x})$.

This interpretation requires that we allow the experiment to be repeated many times. This is what Bayesians criticize by saying that "data that have never been observed are used for inference."

Theorem 2. *If the test statistic has a continuous distribution, then, under the simple null hypothesis* $\text{sp}_\Theta(0) = \theta_0$, *the p-value has the uniform* $\text{U}(0, 1)$ *distribution. Therefore, if we reject $\mathbb{H}_0$ when the p-value is less than or equal to α, the probability of false alarm is α.*

In other words, if $\mathbb{H}_0$ is true and if the conditions of Theorem 2 hold, the p-value is like a random draw from the uniform $\text{U}(0, 1)$ distribution. If $\mathbb{H}_1$ is true and if we repeat the experiment many times, the random p-values will concentrate closer to zero.

✻ COMMENTS:

- The p-value depends on the observed value of the test statistic $T(X)$. It also requires $T(X)$ to have a known distribution under $\mathbb{H}_0$ and no unknown nuisance parameters. This is always true for simple null hypothesis where $\mathbb{H}_0$ contains a single point θ_0.
- For composite null hypotheses, computation of a p-value requires finding a suitable statistic $T(X)$ whose distribution has no nuisance parameters. Construction of such a statistic is similar to the problem of finding a lilekihood-ratio test for which a suitable false-alarm threshold can be determined. This problem is discussed in handout on composite hypothesis testing.

Radar example

THE nth sample of the output $\boldsymbol{X} = \big(X[n]\big)_{n=0}^{N-1}$ of a range-gated radar is

$$X[n] = \theta s + W[n]$$

The null hypothesis is $\mathbb{H}_0 : \theta = 0$, where $W[n]$ is additive white Gaussian noise (AWGN) with known variance σ^2. Consider the test statistic

$$T_1(\boldsymbol{X}) = \Big|\sum_{n=0}^{N-1} X[n]\Big|.$$

Under $\mathbb{H}_0$, $\sum_{n=0}^{N-1} X[n] \sim \mathcal{N}(0, N\sigma^2)$ and

$$\begin{aligned}
\Pr(T_1 > t_1 \,|\, \mathbb{H}_0) &= 2\Pr(\sum_{n=0}^{N-1} X[n] > t_1) \\
&= 2\Pr\Big(\frac{\sum_{n=0}^{N-1} X[n]}{\sqrt{N\sigma^2}} > \frac{t_1}{\sqrt{N\sigma^2}}\Big) \\
&= 2Q\Big(\frac{t_1}{\sqrt{N\sigma^2}}\Big). \qquad (21)
\end{aligned}$$

$\frac{\sum_{n=0}^{N-1} X[n]}{\sqrt{N\sigma^2}}$ is a standard normal random variable

Hence, the p-value associated with the observation $t_1 = T_1(\boldsymbol{x})$ is $2Q\big(t_1/\sqrt{N\sigma^2}\big)$. This p-value depends on the noise variance σ^2 and gets more significant (smaller) as t_1 becomes larger.

We can construct another test using the sum of signs of radar returns:

$$T_2(\boldsymbol{X}) = \Big|\sum_{n=0}^{N-1} \operatorname{sgn}\big(X[n]\big)\Big|.$$

Since $\operatorname{sgn}(X[n])$ are binary, the null distribution of $0.5[T_2(X) + n)]$ is binomial $\operatorname{Bin}(N, 0.5)$, which, by the CLT, can be approximated by a Gaussian pdf when N is large, yielding

Central Limit Theorem (CLT): If $(Z_n)_{n=1}^N$ are i.i.d. random variables with $\mathrm{E}(Z_1) = \mu$ and $\mathrm{E}[(Z_1-\mu)^2] = \sigma^2$, then $\frac{1}{\sqrt{N}}\sum_{n=1}^N Z_n \xrightarrow{\mathrm{d}} \mathcal{N}(\mu, \sigma^2)$.

$$\Pr(T_2 > t_2 \,|\, \mathbb{H}_0) \approx 2Q\Big(\frac{2t_2}{\sqrt{N}}\Big) \qquad (22)$$

where $t_2 = T_2(\boldsymbol{x})$. This does not depend on σ^2.

Both p-values are valid, but when W is Gaussian, (21) will generally provide better evidence against $\mathbb{H}_0$ (smaller p-values) than (22) *because* (21) accounts for the knowledge of σ^2. However, (22) may be better if W is heavy tailed noise, e.g., following Laplace distribution.

References

Hero, Alfred O. (2015). *Statistical Methods for Signal Processing*. Lecture notes. Univ. Michigan, Ann Arbor, MI (cit. on p. 1).

Johnson, Don H. (2013). *Statistical Signal Processing*. Lecture notes. Rice Univ., Houston, TX (cit. on p. 1).

Van Trees, Harry L., Kristine L. Bell, and Zhi Tian (2013). *Detection, Estimation, and Modulation Theory, Part I*. 2nd ed. New York: Wiley (cit. on pp. 1, 14).
Varshney, Pramod K (1996). *Distributed Detection and Data Fusion*. New York: Springer (cit. on p. 9).