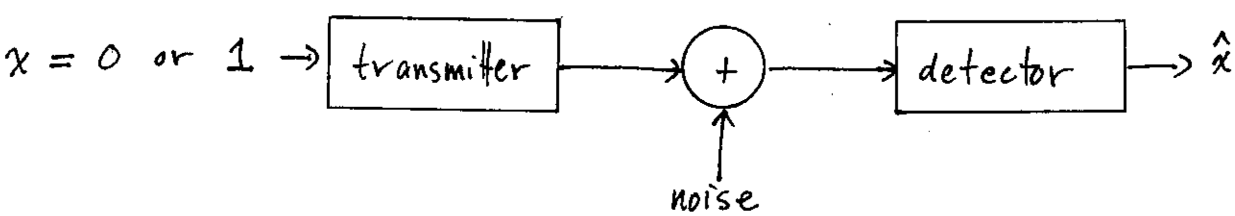
NEYMAN-PEARSON DETECTION

In deriving the Bayes detector we assumed $\pi_i = P(H_i)$ to be known. Some times this is a reasonable assumption, other times it isn't.

Examples

1. Binary communication channel

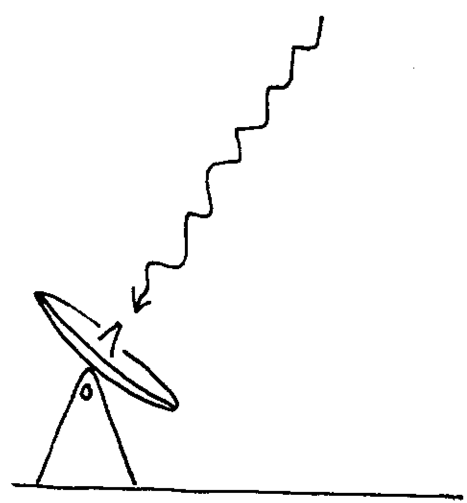


$$\pi_0 = \pi_1 = \frac{1}{2}$$

2. Search for extra-terrestrial life

H_0: $\underline{X} \sim$ cosmic radiation

H_1: $\underline{X} \sim$ cosmic radiation + intelligent signal



$P(H_1) = ?$

In a binary hypothesis testing problem

$$H_0 : \underline{X} \sim f_0(\underline{x})$$

$$H_1 : \underline{X} \sim f_1(\underline{x})$$

we assign names to the four possible outcomes

		truth H_0	truth H_1
decision	H_0	rejection	miss
	H_1	false alarm	detection

$P_D = P(H_1 \mid H_1)$ "detection probability"

$P_M = P(H_0 / H_1)$ " miss " "

$P_F = P(H_1 \mid H_0)$ "false alarm " "

$P_R = P(H_0 / H_0)$ " rejection " "

Note that

$$P_D = 1 - P_M$$

$$P_F = 1 - P_R$$

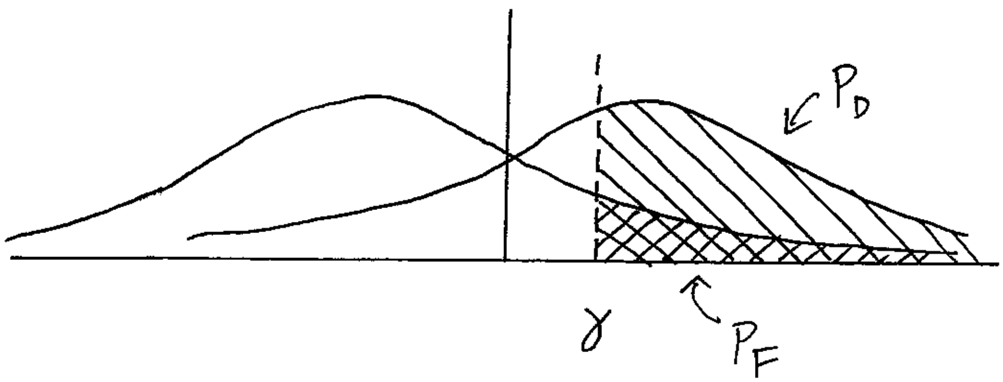
so there are only <u>two degrees of freedom</u> for evaluating a hypothesis test.

Also note: P_D and P_F do not involve <u>prior probabilities</u> on $\mathcal{H}_0, \mathcal{H}_1$.

Idea: formulate a detection criterion in terms of P_D, P_F.

Example

$$H_0: X \sim \mathcal{N}(-1, 1)$$
$$H_1: X \sim \mathcal{N}(1, 1)$$



Consider the decision rule

$$x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

As γ increases,

P_F decreases (good)

P_D decreases (bad)

More generally, P_F and P_D are indirectly related through the <u>decision region</u> R_1:

$$P_D = \int_{R_1} f_1(\underline{x})\,d\underline{x}$$

$$P_F = \int_{R_1} f_0(\underline{x})\,d\underline{x}$$

As R_1 <u>expands</u>, P_D and P_F increase

As R_1 <u>shrinks</u>, P_D and P_F decrease

Ideally, we would like

$$P_D = 1 \quad , \quad P_F = 0$$

but this is only possible when

(a)

So what is the best way to choose R_1?

The Neyman-Pearson Criterion

The Neyman-Pearson (NP) detector solves the following optimization problem:

$$\max \; P_D \quad \text{s.t. } P_F \leq \alpha$$

In words, the NP detector has the largest detection probability among all detectors with false alarm probability no greater than α.

Terminology

P_D = <u>power</u>

P_F = <u>size</u>

So the NP detector is the <u>most powerful test of size</u> (not exceeding) <u>α</u>.

The Neyman-Pearson Lemma: A First Look

Let $\alpha \in [0,1]$. The NP detector is

$$\Lambda(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where $\Lambda(\underline{x}) = \dfrac{f_1(\underline{x})}{f_0(\underline{x})}$ and η is

chosen such that

$$P_F = \int_{\Lambda(\underline{x})>\eta} f_0(\underline{x})\,d\underline{x} = \alpha$$

Note/ It may not always be possible to set η such that $P_F = \alpha$, such as in the case of discrete data. We will return to this case later. For now, assume $P_F = \alpha$ is achievable.

So the optimal detector is once again a <u>likelihood ratio test.</u>

The Bayes detector and NP detector lead to the same test. The difference is in how we select the <u>threshold</u> η.

- For the Bayes detector:

$$\eta = \frac{\pi_0}{\pi_1} \cdot \frac{(c_{10} - c_{00})}{(c_{01} - c_{11})}$$

- For the NP detector

$$\eta = P_F^{-1}(\alpha)$$

where

$$P_F(\eta) = \int_{\Lambda > \eta} f_0(\underline{x})\, d\underline{x}$$

We can prove the NP Lemma using the theory of constrained optimization (Lagrange multiplier theory)

Constrained Optimization

Consider the problem

$$\max_{x \in \Omega} h(x) \quad \text{subject to } g(x) \le C,$$

where $h, g : \Omega \to \mathbb{R}$, $C \in \mathbb{R}$, and Ω is an arbitrary set.

Theorem Let $\lambda \geq 0$ and suppose $x_0(\lambda) \in \Omega$ maximizes

$$L(x, \lambda) \equiv h(x) - \lambda g(x)$$

for each λ. Then $x_0(\lambda)$ maximizes h over all x such that $g(x) \le g(x_0(\lambda))$.

__Corollary:__ If λ^* is such that $g(x_0(\lambda^*)) = C$, then $x_0(\lambda^*)$ maximizes $h(x)$ over all x such that $g(x) \le C$.

__Proof of Theorem:__

By assumption, $x_0 = x_0(\lambda)$ satisfies

$$h(x_0) - \lambda g(x_0) \ge h(x) - \lambda g(x)$$

for all $x \in \Omega$. Equivalently,

$$h(x_0) - h(x) \ge \lambda(g(x_0) - g(x))$$

for all $x \in \Omega$. Define

$$S = \{x \in \Omega \mid g(x) \le g(x_0)\}.$$

Then for all $x \in S$,

$$h(x_0) - h(x) \ge \lambda(g(x_0) - g(x)) \ge 0$$

so $h(x_0) \ge h(x) \quad \forall x \in S.$ ∎

Proof of Neyman-Pearson Lemma

Apply constrained optimization theorem with

$$h = P_D$$
$$g = P_F$$
$$c = \alpha$$
$$x = R_1$$
$$\Omega = \text{all possible decision regions}$$

To do this, we must find $R_1 = R_1(\lambda)$ that maximizes

$$L = P_D - \lambda P_F$$
$$= \int_{R_1} f_1(\underline{x})\,dx - \lambda \int_{R_1} f_0(\underline{x})\,d\underline{x}$$
$$= \int_{R_1} \left[f_1(\underline{x}) - \lambda f_0(\underline{x}) \right] d\underline{x}$$

So choose

$$R_1 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_0(\underline{x})} > \lambda \right\}.$$

Now, to maximize P_D over all R_1 such that $P_F \le \alpha$, we take λ such that

$$P_F = \int_{\Lambda > \lambda} f_0(\underline{x})\,d\underline{x} = \alpha$$

$\blacksquare$

Setting the threshold

Computing η such that $P_F = \alpha$ is not always easy. It usually requires the use of monotonic transformations and test statistics as the following example demonstrates.

<u>Example</u> DC signal in AWGN

$$H_0 : \underline{X} \sim \mathcal{N}(\underline{0}, \sigma^2 \underline{I})$$

$$H_1 : \underline{X} \sim \mathcal{N}(A\underline{1}, \sigma^2 \underline{I}) \quad , \quad A > 0$$

Let's design a NP detector

From the last lecture we saw

$$\Lambda(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

$$\Updownarrow$$

$$\frac{1}{N}\sum_{n=1}^{N} x_n \equiv t \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \equiv \frac{\sigma^2}{NA}\ln(\eta) + \frac{A}{2}$$

Recall $t \sim \mathcal{N}(0, \frac{\sigma^2}{N})$ under H_0

$t \sim \mathcal{N}(A, \frac{\sigma^2}{N})$ under H_1

Exercise (a) Use the Q function to express P_F, P_D in terms of γ and known quantities (b) Find γ for the NP detector of size α (c) Express P_D in terms of P_F and SNR.

Solution

$$P_F = \text{Prob}(t > \gamma \mid H_0)$$

$$= Q\left(\frac{\gamma}{\sigma/\sqrt{N}}\right) \leq \alpha$$

$$P_D = \text{Prob}(t > \gamma \mid H_1)$$

$$= Q\left(\frac{\gamma - A}{\sigma/\sqrt{N}}\right)$$

To set the threshold, we take

$$P_F = \alpha$$

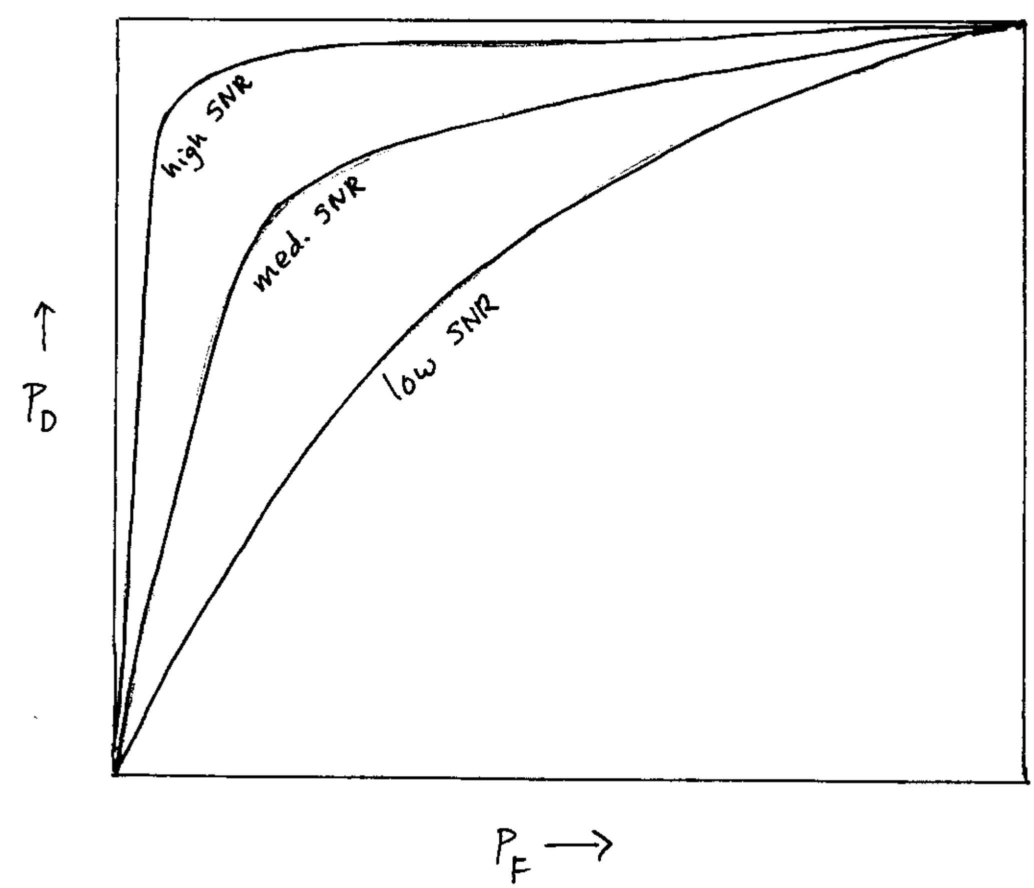
$$\Rightarrow \quad \gamma = \frac{\sigma}{\sqrt{N}} Q^{-1}(\alpha)$$

$$\Rightarrow \quad P_D = Q\left(Q^{-1}(P_F) - \frac{A\sqrt{N}}{\sigma}\right)$$

$$= Q\left(Q^{-1}(P_F) - \sqrt{SNR}\right)$$

The Receiver Operating Characteristic

The ROC of a detector is a plot of P_D vs. P_F.



Ex

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

$$SNR = \frac{A^2 N}{\sigma^2}$$

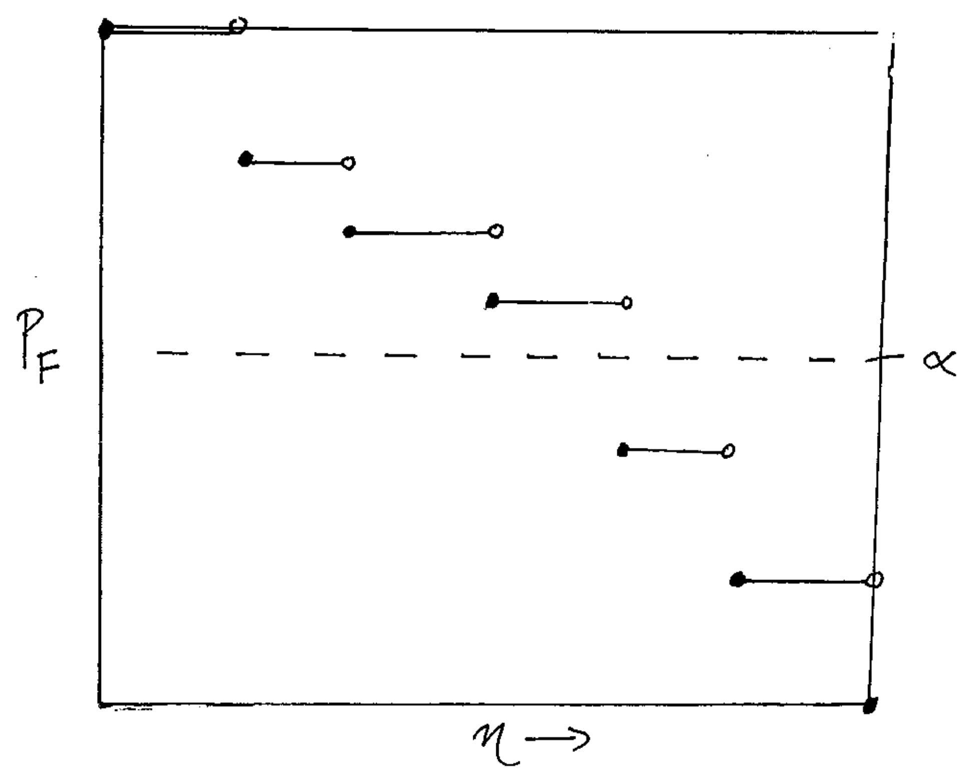
NP Lemma: General Case

Suppose the data $\underline{X}$ is discrete.

Then

$$P_F = \sum_{\underline{x} : \Lambda(\underline{x}) > \eta} f_0(\underline{x})$$

So it may not be possible to have $P_F = \alpha$ for all α in the current setup.



What if we choose η so that P_F is as large as possible? Does the LRT still solve

$$\begin{aligned} &\max \; P_D \\ &\text{s.t. } P_F \leq \alpha \end{aligned} \qquad ?$$

Not quite.

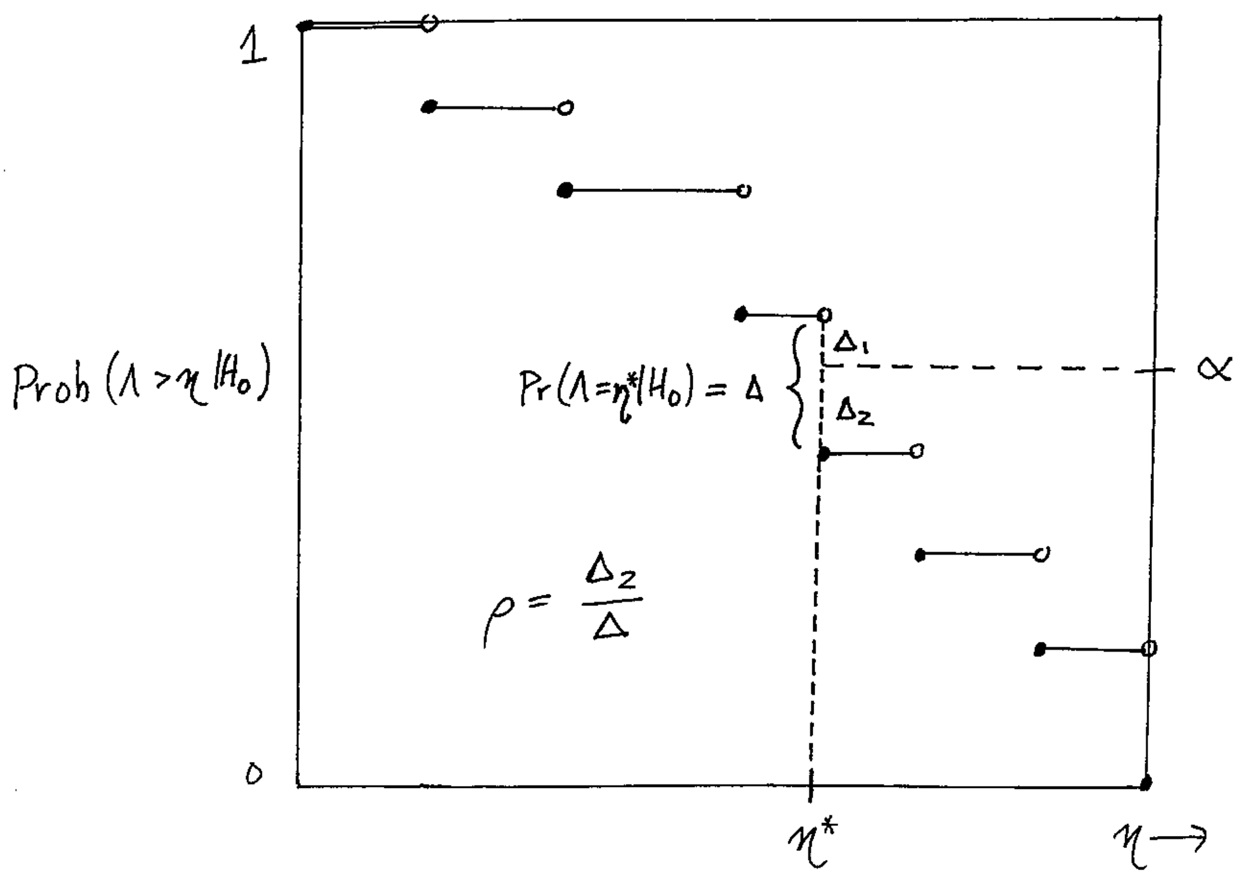
We must now be concerned with the case $\Lambda(\underline{x}) = \eta$, which can occur with <u>nonzero</u> probability.

Let $\alpha \in [0,1]$, and let η^* be as small as possible such that

$$\Pr(\Lambda > \eta^* | H_0) < \alpha$$

Choose $\rho \in [0,1)$ such that

$$\Pr(\Lambda > \eta^* | H_0) + \rho \Pr(\Lambda = \eta^* | H_0) = \alpha$$



Consider the decision rule

$$\begin{cases} \text{declare } H_1 & \text{if } \Lambda(\underline{x}) > \eta^* \\ \text{flip a ``}\rho\text{-coin''} & \text{if } \Lambda(\underline{x}) = \eta^* \\ \text{declare } H_0 & \text{if } \Lambda(\underline{x}) < \eta^* \end{cases}$$

Then $P_F = \alpha$

A "ρ-coin" turns up heads (H_1) with probability ρ.

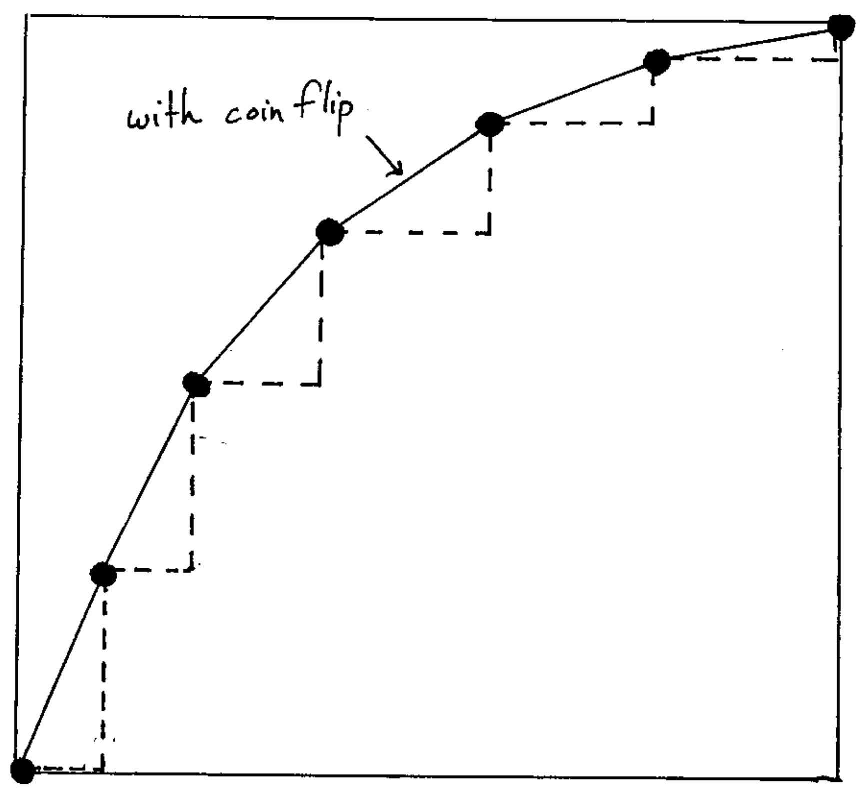
If you think back to the proof of the NP Lemma, we can redistribute

$$\{\underline{x} : \Lambda(\underline{x}) = \eta\}$$

as we see fit and still maximize the Lagrangian. We use just enough probability mass to bring P_F up to α and use the rest to increase P_D.

The <u>modified</u> LRT is now <u>optimal</u> for the case of <u>disrete</u> data.

Intuition:



Old LRT: $\Lambda \gtrless \eta$

operates at discrete set of (P_F, P_D)

New LRT (with coin flip)

ROC = "convex hull" of old ROC

Randomized Tests

To state & prove the NP Lemma in full generality, we need to introduce general randomized tests.

A randomized test is a function $\phi : \mathbb{R}^N \longrightarrow [0,1]$ corresponding to the decision rule

"declare H_1 with probability $\phi(\underline{x})$."

If $\phi(\underline{x}) \in \{0,1\}$ $\forall \underline{x}$, then ϕ is deterministic.

Recall: If A is an event and $\underline{Z}$ is a random variable with density g, then

$$P(A) = \int P(A \mid \underline{Z} = \underline{z})\, g(\underline{z})\, d\underline{z}$$

This is one version of the law of total probability. A similar result holds if $\underline{Z}$ is discrete.

Therefore

$$P_F = \text{Prob}(\text{declare } H_1 \mid H_0 \text{ true})$$

$$=$$

$$P_F = \int \text{Prob}(\text{declare } H_1 \mid \underline{X} = \underline{x}) \cdot f_0(\underline{x})\, d\underline{x}$$

$$= \int \phi(\underline{x})\, f_0(\underline{x})\, d\underline{x}.$$

reduces to $\int_{R_1} f_0(\underline{x})\, dx$ for deterministic ϕ

Similarly,

$$P_D = \int \phi(\underline{x})\, f_1(\underline{x})\, d\underline{x}.$$

We are now ready to state and prove the general NP Lemma:

NP Lemma | Let $\alpha \in [0,1]$. Define

$$\phi(\underline{x}) = \begin{cases} 1 & \text{if } \Lambda(\underline{x}) > \eta \\ \rho & \text{if } \Lambda(\underline{x}) = \eta \\ 0 & \text{if } \Lambda(\underline{x}) < \eta \end{cases}$$

where η and ρ are such that

$$P(\Lambda(\underline{X}) > \eta \mid H_0) + \rho P(\Lambda(\underline{X}) = \eta \mid H_0) = \alpha.$$

Then ϕ is the most powerful test of size α among all randomized tests.

The MP test of size $\alpha = 0$ is

$$\phi(\underline{x}) = \begin{cases} 1 & \text{if } f_0(\underline{x}) = 0 \\ 0 & \text{if } f_1(\underline{x}) = 0 \end{cases}$$

<u>Proof</u> Let ϕ' be any randomized test with size $\le \alpha$.

Denote

$$\beta = \int \phi(\underline{x}) f_1(\underline{x})\, d\underline{x} \quad = \text{power of } \phi$$

$$\beta' = \int \phi'(\underline{x}) f_1(\underline{x})\, d\underline{x} \quad = \quad '' \quad '' \quad \phi'$$

$$\alpha = \int \phi(\underline{x}) f_0(\underline{x})\, d\underline{x} \quad = \text{size of } \alpha$$

$$\alpha' = \int \phi'(\underline{x}) f_0(\underline{x})\, d\underline{x} \quad = \quad '' \quad '' \quad \alpha'$$

By def. of ϕ,

$$0 \le \int (\phi(\underline{x}) - \phi'(\underline{x}))[f_1(\underline{x}) - \eta f_0(\underline{x})]\, d\underline{x}$$

$$= \beta - \beta' - \eta(\alpha - \alpha')$$

which implies

$$\beta - \beta' \geq \eta(\alpha - \alpha') \geq 0.$$

For $\alpha = 0$, any test of size 0 must declare H_0 wherever $f_0(\underline{x}) > 0$. To maximize power, declare H_1 everywhere else.

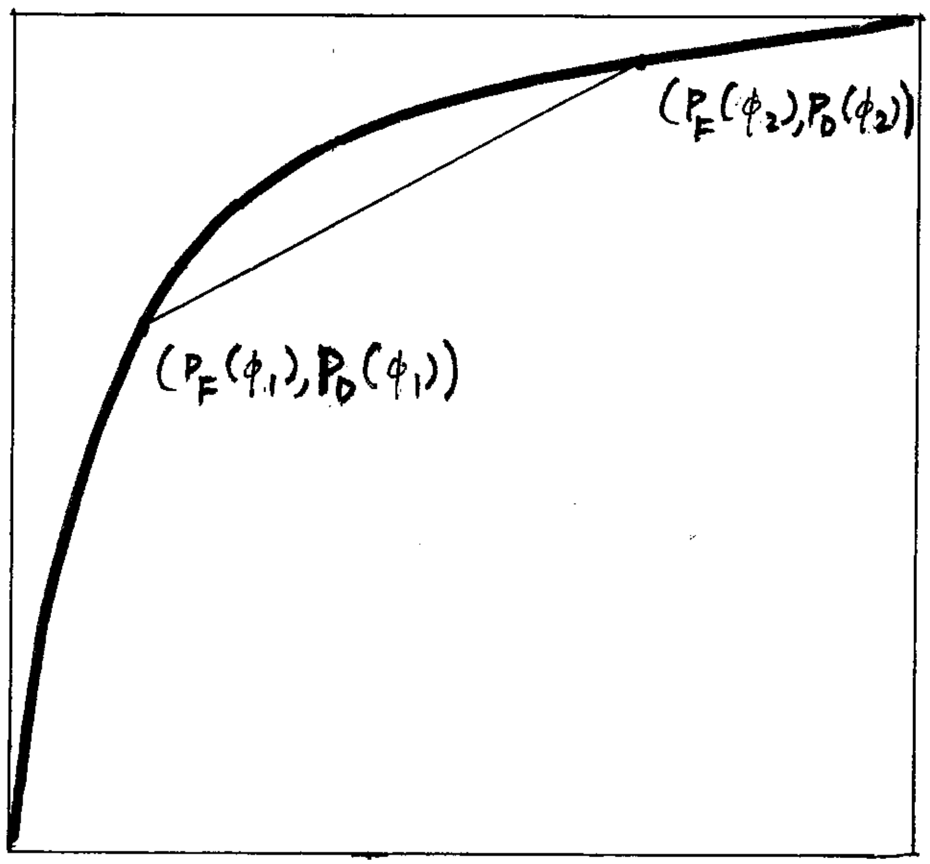
$\blacksquare$

Properties of the ROC of the LRT

Fact 1 The ROC of the LRT is concave. Denote

$$P_D(\phi) = E[\phi(\underline{X}) \mid \underline{X} \sim f_1]$$

$$P_F(\phi) = E[\phi(\underline{X}) \mid \underline{X} \sim f_0]$$



In other words, for any LRTs ϕ_1, ϕ_2, the line segment

$$\left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} P_F(\phi_1) \\ P_D(\phi_1) \end{bmatrix} + (1-\lambda) \begin{bmatrix} P_F(\phi_2) \\ P_D(\phi_2) \end{bmatrix}, \ \lambda \in [0,1] \right\}$$

is <u>below</u> the ROC.

Let's prove this. Suppose it's not true. Then

$\exists$ ϕ_1, ϕ_2 and $\lambda \in (0,1)$ such that

$$\lambda \begin{bmatrix} P_F(\phi_1) \\ P_D(\phi_1) \end{bmatrix} + (1-\lambda) \begin{bmatrix} P_F(\phi_2) \\ P_D(\phi_2) \end{bmatrix}$$

is <u>above</u> the ROC of the LRT. Consider the test

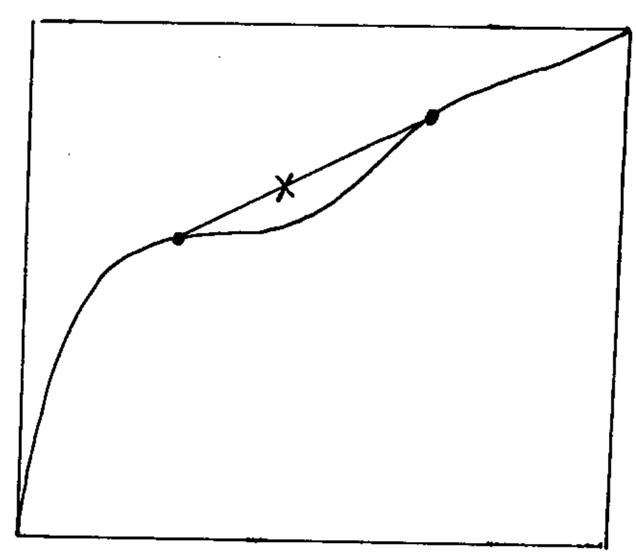
$$\phi(\underline{x}) := \lambda \phi_1(\underline{x}) + (1-\lambda)\phi_2(\underline{x}).$$

For this test

$$P_F(\phi) = \lambda P_F(\phi_1) + (1-\lambda) P_F(\phi_2)$$

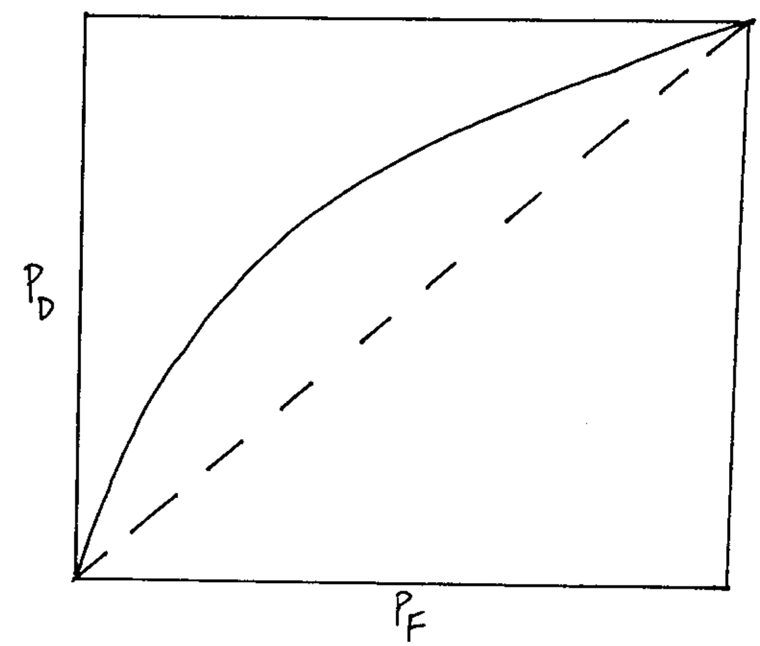
$$P_D(\phi) = \lambda P_D(\phi_1) + (1-\lambda) P_D(\phi_2)$$

This contradicts the optimality of the LRT. $\square$

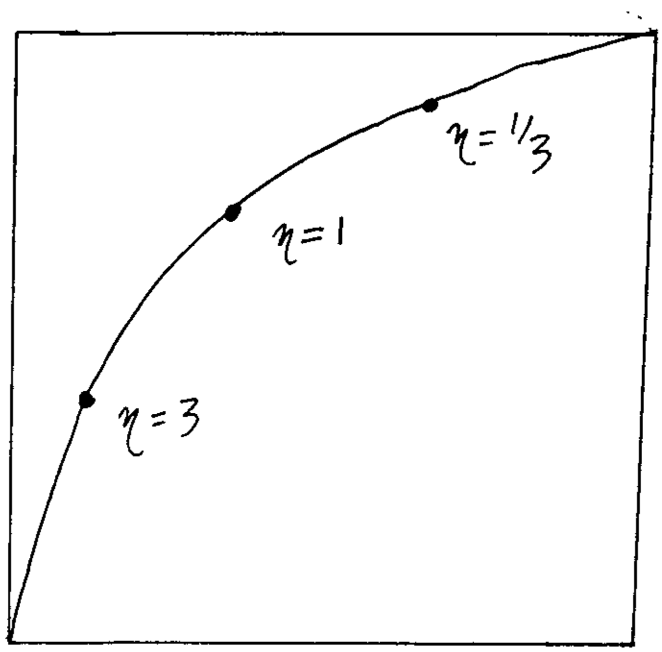


Fact 2 The ROC of the LRT is above the line $P_D = P_F$.

(b) **Proof?**



Fact 3 The <u>slope</u> of the ROC (for the LRT) at a point $(P_F(\eta), P_D(\eta))$ is η.



Proof: see Scharf.

That is,

$$\frac{dP_D}{dP_F} = \eta$$

Fact 3 also implies concavity because it implies that $\frac{\partial P_D}{\partial P_F}$ is non-increasing.

Historical note | These basic properties of the ROC were developed by Theodore Birdsall, a former UM Professor and my academic great-great-granfather. He also developed several other fundamentals of signal detection theory.

Reference: Peterson, Birdsall, and Fox, "The theory of signal detectability," Trans. Inst. Radio Engrs., Professional Group on Information Theory, PGIT-4, 171-212 (1954).

Summary

- Neyman-Pearson detector:

  maximizes P_D s.t. $P_F \leq \alpha$

- NP criterion does not assume knowledge of prior probabilities of each hypothesis (frequentist)

- The Bayes risk detector does (Bayesian)

- Optimal detector for both criteria given by LRT.

Key a. $\text{support}(f_0) \cap \text{support}(f_1) = \phi$

b. The line $P_D = P_F$ corresponds to random guessing, $\phi_\alpha(\mathbf{x}) := \alpha$, and the NP detector does at least as well as that.

Also follows from concavity and values for $\eta = 0, \infty$.