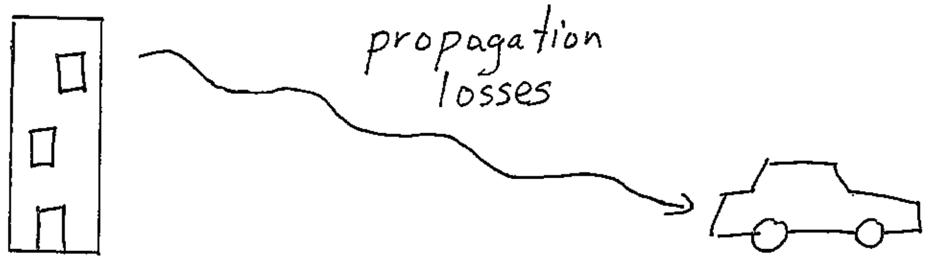
# UMP TESTS & THE KARLIN-RUBIN THEOREM

## Signal Detection in the Presence of Unknowns

In many real world detection problems, the characteristics of the signal and/or noise are not perfectly known:

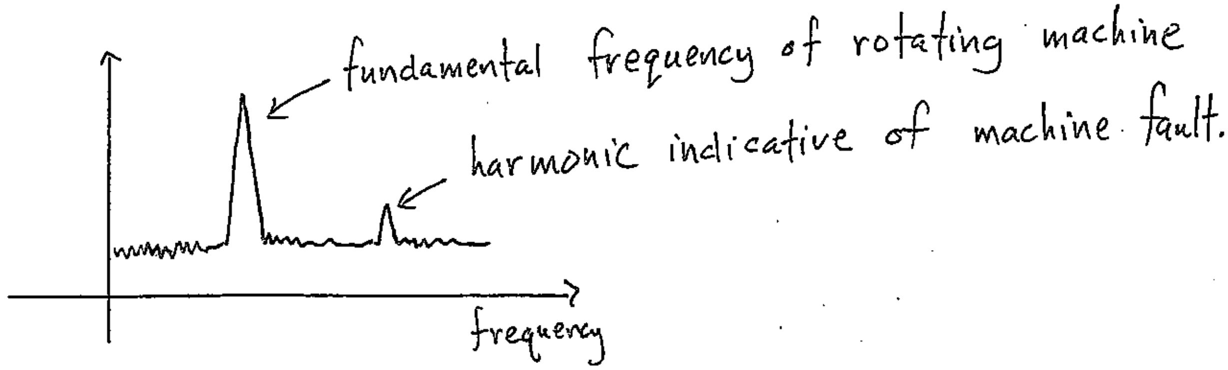
**Ex 1** Unknown signal amplitude

(a) Wireless comm:




received signal attenuated by unknown factor

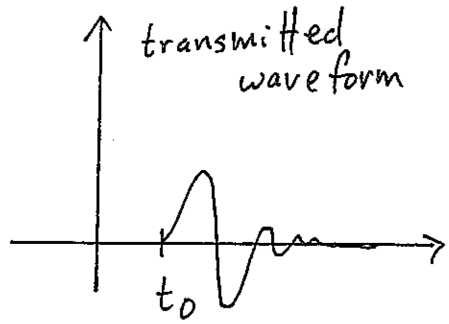
(b) machine fault detection

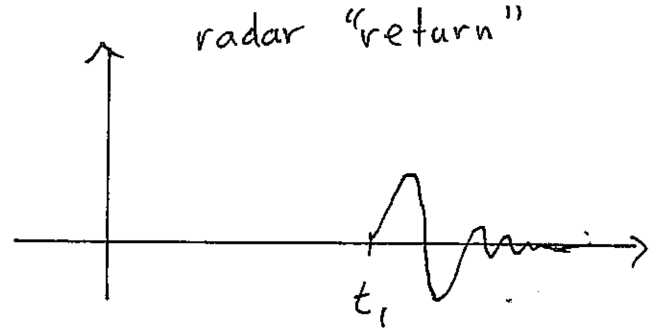



strength of harmonic distortion is uncertain

## Ex 2] Unknown location / delay

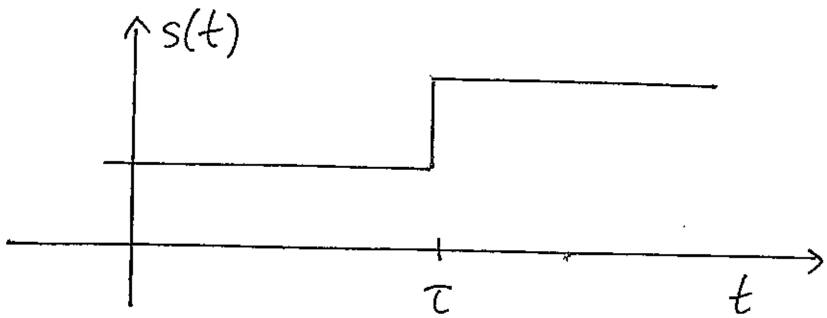
(a) Radar:




$d = t_1 - t_0$ is unknown

(b) Step-change detection:



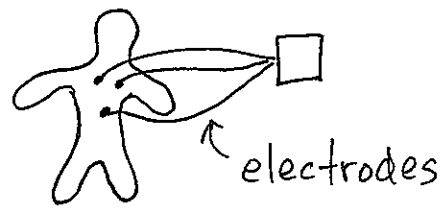

$\tau$ unknown

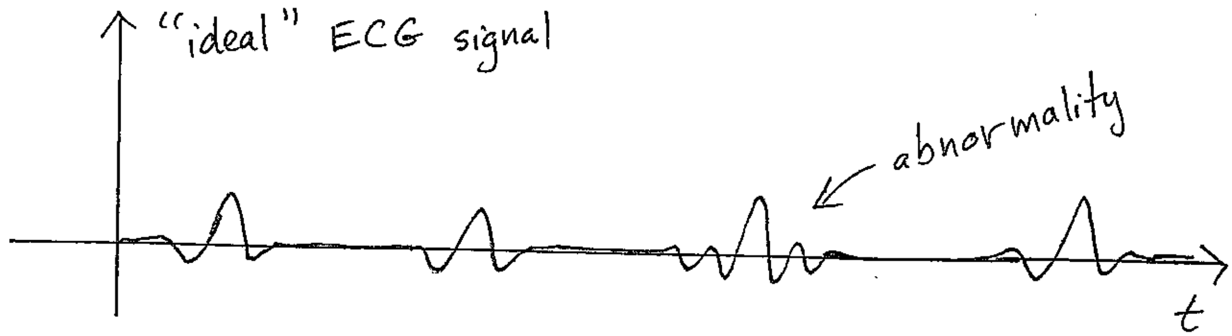
In 2-d, this amounts to <u>edge-detection</u>: where do edges occur?

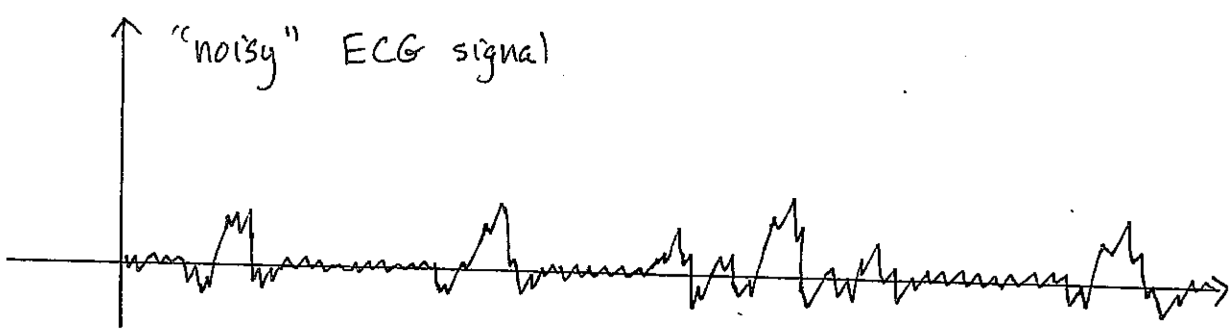
## Ex 3 Unknown noise power

We may know noise is white, but what is $\sigma^2$?

Heart monitoring:




<u>Goal</u>: detect abnormal heartbeat.

Noise level depends on patient, electrode placement, and ambient noise from environment, all of which may be unknown.

# Modeling Uncertainty

① Parametric uncertainty: unknown parameters in pdf or pmf of observation.

② Nonparametric uncertainty: we don't even know the functional form of the pdf or pmf

Non parametric uncertainty is much more challenging. We will focus on parametric uncertainty in this course.

# Composite Hypothesis Testing

$$H_0: \underline{X} \sim f(\underline{x}; \underline{\theta}_0), \qquad \underline{\theta}_0 \in \Theta_0$$

$$H_1: \underline{X} \sim f(\underline{x}; \underline{\theta}_1), \qquad \underline{\theta}_1 \in \Theta_1$$

So far, we have only considered <u>simple</u> hypotheses: $\Theta_0 = \{\underline{\theta}_0\}$, $\Theta_1 = \{\underline{\theta}_1\}$. When $|\Theta_k| > 1$, $H_k$ is a <u>composite</u> hypothesis.

**Ex 1** Unknown mean:

$$H_0: \quad X \sim \mathcal{N}(0,1)$$

$$H_1: \quad X \sim \mathcal{N}(\mu, 1), \qquad \mu > 0$$

$\Longrightarrow$ $H_1$ is composite

**Ex 2** Unknown noise power

$$H_0: \quad x(n) = s_0(n) + w(n)$$

$$H_1: \quad x(n) = s_1(n) + w(n)$$

$w(n) \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, $\sigma^2$ unknown $\Longrightarrow$ $H_0, H_1$ composite

Whether we want a Bayes Risk or Neyman-Pearson detector, the optimal decision rule is the LRT:

$$\Lambda(\underline{x}) = \frac{f(\underline{x}; \underline{\theta}_1)}{f(\underline{x}; \underline{\theta}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

In this form, the LRT requires knowledge of the unknown parameter, and hence is not useful.

Sometimes, however, if we write the test in a different form, the dependence on the unknown parameter goes away.

$\boxed{\text{Ex}}$ Signal Detection in AWGN, $\sigma^2$ unknown:

$$H_0 : \underline{x} = \underline{w} \qquad\qquad w(n) \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$H_1 : \underline{x} = \underline{s} + \underline{w} \qquad\qquad \sigma^2 \ \underline{\text{unknown}}$$

LRT reduces to

$$\underline{s}^T \underline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \log(\eta) + \frac{\underline{s}^T \underline{s}}{2}$$

If the hypotheses are equally probable a priori, then $\eta = 1$ and we have

$$\underline{s}^T \underline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\underline{s}^T \underline{s}}{2} \longleftarrow \boxed{\text{independent of } \sigma^2}$$

For $M > 2$ hypotheses

$$H_k : \underline{x} = \underline{s}_k + \underline{w}$$

$$\Rightarrow \text{minimize } \frac{1}{\sigma^2} \| \underline{x} - \underline{s}_k \|^2$$

This situation is rare in practice.

A more common occurance is when the LRT can be reduced to a test statistic that does not depend on the unknown parameter under $H_0$. This allows us to set the false alarm rate.

Ex| Unknown signal amplitude

$$H_0 : \quad x(n) = w(n)$$
$$H_1 : \quad x(n) = As(n) + w(n)$$

$w(n) \overset{IID}{\sim} N(0, \sigma^2)$, $\sigma^2$ known, A unknown

The LRT reduces to

$$\underbrace{A \underline{s}^T \underline{x}}_{} \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \log(\eta) + A^2 \frac{\underline{s}^T \underline{s}}{2}$$

↳ test statistic depends on unknown amplitude ⇒ don't know distribution under $H_0$ ⇒ can't set $P_F$

What if we know $A > 0$? Then, dividing by $A$, we have

$$\underline{s}^T \underline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2}{A} \log(\eta) + A \frac{\underline{s}^T \underline{s}}{2} \equiv \gamma$$

Under $H_0$

$$\underline{s}^T \underline{x} \sim$$

independent of $A$!

ⓐ

We can now set the threshold $\gamma$ to achieve a certain $P_F$:

$$P_F =$$

$$\Longrightarrow \gamma =$$

So we can set $\gamma$ to achieve $P_F$. So what? Is this detector optimal in any sense? Is $P_D$ maximized?

Since the LRT is equivalent to

$$\underline{s}^T \underline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

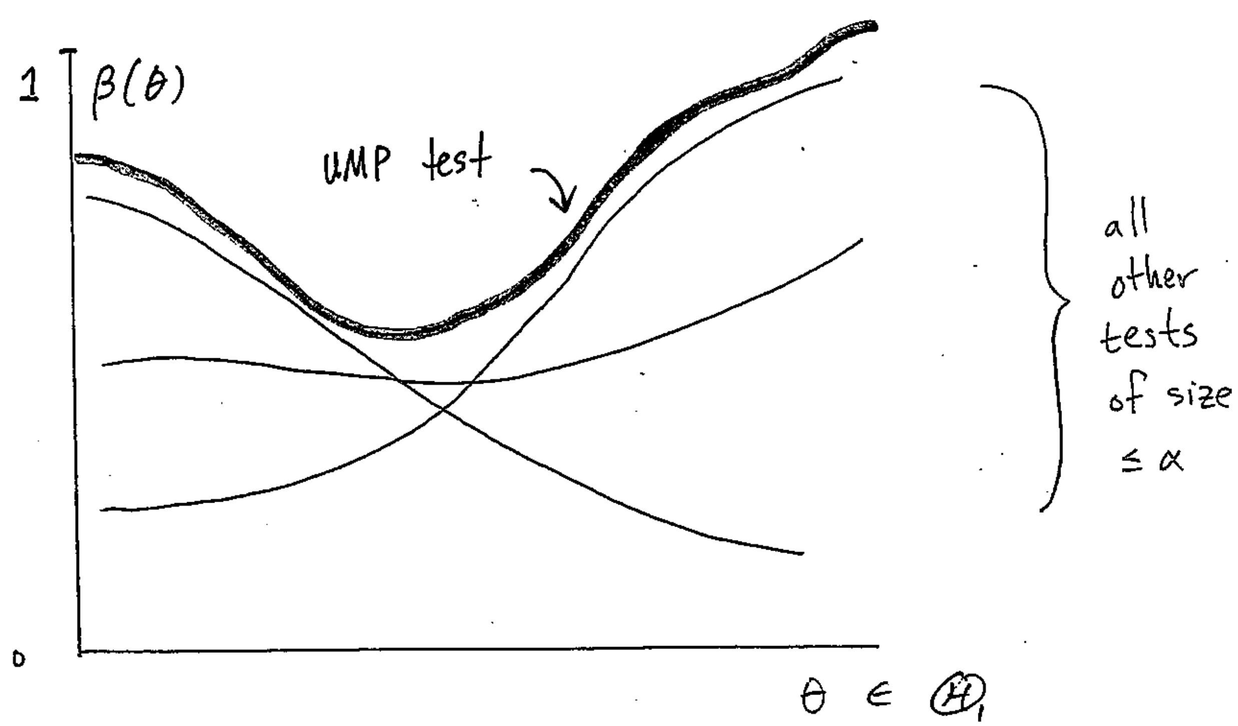
and $\gamma$ can be selected to ensure $P_F = \alpha$, the NP lemma implies that this test is optimal regardless of the value of $A$!

## $\boxed{\text{UMP Tests}}$

A test/detector is a uniformly most powerful (UMP) test of size $\alpha$ if it has the largest <u>power</u>

$$\beta(\underline{\theta}) := P(\text{declare } H_1 ; \underline{\theta}), \quad \underline{\theta} \in \Theta_1$$

among all tests of size $\leq \alpha$, $\forall \underline{\theta} \in \Theta_1$



Unfortunately, UMP tests rarely exist. However, there is a certain class of problems for which they do.

## Monotone Likelihood Ratios

Suppose a measurement $\underline{x}$ has pdf/pmf determined by a scalar parameter $\theta$ and let $\theta_0$ be fixed. Suppose we are interested in the one-sided problem

$$H_0 : \ \theta = \theta_0$$
$$H_1 : \ \theta > \theta_0$$

Further suppose that $t$ is a scalar sufficient statistic for $\theta$.

Observe that for any $\theta_1 > \theta_0$, the likelihood ratio is a function of $t$:

$$\Lambda(\underline{x}) = \frac{f(\underline{x};\theta_1)}{f(\underline{x};\theta_0)} = \frac{a(\underline{x})\, b_{\theta_1}(t)}{a(\underline{x})\, b_{\theta_0}(t)} = \frac{b_{\theta_1}(t)}{b_{\theta_0}(t)} = \tilde{\Lambda}(t)$$

by the Fisher-Neyman factorization.

**Proposition** If $\tilde{\Lambda}(t)$ is monotone increasing for all $\theta_1 > \theta_0$, then

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

is UMP of size $\alpha$, where $\gamma$ is determined by

$$P(T > \gamma \,;\, \theta = \theta_0) = \alpha.$$

A similar result holds if $\tilde{\Lambda}(t)$ is monotone decreasing for all $\theta_1 > \theta_0$.

**Proof** Suppose $\theta = \theta_1 > \theta_0$. We need to show that

$$P(T > \gamma \,;\, \theta = \theta_1)$$

is as large as possible among all tests with size $\alpha$. By the NP Lemma, the most powerful test for $\theta = \theta_1$ vs. $\theta = \theta_0$ is

$$\tilde{\Lambda}(t) \gtrless \eta$$

where $\eta$ is such that

$$P(\tilde{\Lambda}(T) > \eta \,;\, \theta = \theta_0) = \alpha.$$

Since $\tilde{\Lambda}(t)$ is monotone increasing, so is its inverse, and the most powerful test simplifies to

$$t \underset{H_0}{\overset{H_1}{\gtrless}} \tilde{\Lambda}^{-1}(\eta) \equiv \gamma.$$

Since the distribution of $T; \theta = \theta_0$ is independent of $\theta_1$, we can set the value of $\gamma$ without knowledge. In other words, our test has greatest power for all $\theta_1 > \theta_0$. $\blacksquare$

---

In short, the monotone LR property implies that we can eliminate $\theta_1$ from the test statistic, and since $H_0$ is simple, we can set the threshold to ensure the desired size.

**Remarks** ① In the case of discrete data, the thresholding test may have the form

$$t \underset{H_0}{\overset{H_1}{\gtreqless}} \gamma$$

where if $t = \gamma$ we flip a "$\rho$-coin" such that

$$P\{T > \gamma ; \theta_0\} + \rho P\{T = \gamma ; \theta_0\} = \alpha.$$

② If $\tilde{\Lambda}(t)$ is monotone decreasing, then the inequalities in the thresholding test are reversed.

## The Karlin-Rubin Theorem

The preceeding result can be generalized to the case where $\Theta_0$ is also composite.

Now suppose $\theta_0$ is fixed and consider the one-sided problem

$$H_0: \quad \theta \le \theta_0$$

$$H_1: \quad \theta > \theta_0$$

To state the general result, we need to generalize our notion of "size."

Let $\phi(\underline{x}) \in [0,1]$ denote an arbitrary test. We define the <u>size</u> of $\phi$ by

$$\text{size}(\phi) = \sup_{\underline{\theta} \in \Theta_0} P\{\phi(\underline{X}) = 1 \mid \underline{\theta}\}$$

$$= \sup_{\underline{\theta} \in \Theta_0} E_{\underline{\theta}}\{\phi(\underline{X})\}.$$

This is essentially the maximum false alarm rate over all possible null hypotheses.

**Theorem** Suppose $t$ is a scalar suff. stat. for $\theta$ and that

$$\tilde{\Lambda}_{\theta_1,\theta_0}(t) = \frac{b_{\theta_1}(t)}{b_{\theta_0}(t)}$$

is monotone increasing for each $\theta_1 > \theta_0$.

Then a UMP test of size $\alpha$ is given by

$$\phi(t) = \begin{cases} 1 & \text{if } t > \gamma \\ \rho & \text{if } t = \gamma \\ 0 & \text{if } t < \gamma \end{cases}$$

where $\gamma, \rho$ are chosen such that

$$P\{T > \gamma \mid \theta = \theta_0\} + \rho P\{T = \gamma \mid \theta = \theta_0\} = \alpha.$$

**Remarks** ① See Scharf, p 124, for a proof.

② Similar result holds if $\Lambda(t)$ is monotone decreasing.

③ Similar result applies to the problem

$$H_0 : \quad \theta_{min} \leq,< \theta \leq \theta_0$$

$$H_1 : \quad \theta_0 < \theta <,\leq \theta_{max}$$

where $\theta_{min}, \theta_{max}$ are fixed and possibly unknown.

When is $\tilde{\lambda}(t)$ monotone increasing? When the LRT can be reduced to

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

by a series of monotone increasing transformations.

$\boxed{\text{Exercise}}$ Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $i = 1, \ldots, N$ and $\mu$ known, and consider testing

$$H_0 : \quad 0 < \sigma^2 \leq \sigma_0^2$$

$$H_1 : \quad \sigma^2 \geq \sigma_0^2$$

with $\sigma_0^2$ fixed, known. Find a UMP test.

Solution Let $\sigma_1^2 > \sigma_0^2$. Then

$$\Lambda(\underset{\sim}{x}) = \frac{(2\pi\sigma_1^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{N}(x_i-\mu)^2\right\}}{(2\pi\sigma_0^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}(x_i-\mu)^2\right\}}$$

$$= \left(\frac{\sigma_0}{\sigma_1}\right)^N \exp\left\{\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\underbrace{\sum_{i=1}^{N}(x_i-\mu)^2}_{=:\, t}\right\}$$

Then

$$\Lambda(t) \gtrless \eta \iff \exp\left\{\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)t\right\} \gtrless \eta\left(\frac{\sigma_1}{\sigma_0}\right)^N$$

$$\iff \left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)t \gtrless 2\log\left[\eta\left(\frac{\sigma_1}{\sigma_0}\right)^N\right]$$

$$\iff t \gtrless \frac{2\log\left[\eta\left(\frac{\sigma_1}{\sigma_0}\right)^N\right]}{\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)} =: \gamma$$

Since $\sigma_1^2 > \sigma_0^2$, all steps are monotone increasing $\Longrightarrow$ UMP test exists.

To set the threshold, recall

$$\frac{T}{\sigma_0^2} = \sum_{i=1}^{N} \left( \frac{X_i - \mu}{\sigma_0} \right)^2 \sim \chi_N^2 \qquad \text{if } \sigma^2 = \sigma_0^2$$

Define

$$Q_{\chi_N^2}(r) := P\{ \chi_N^2 > r \}.$$

Then the size of our test is

$$P\{ T > \gamma \; ; \sigma^2 = \sigma_0^2 \}$$

$$= P\left\{ \frac{T}{\sigma_0^2} > \frac{\gamma}{\sigma_0^2} \; ; \sigma^2 = \sigma_0^2 \right\}$$

$$= Q_{\chi_N^2}\left( \frac{\gamma}{\sigma_0^2} \right)$$

$$= \alpha$$

$$\Longrightarrow \quad \gamma = \sigma_0^2 \, Q_{\chi_N^2}^{-1}(\alpha).$$

# Two-Sided Problems

What if we are interested in a slightly different problem?

$$H_0 : x(n) = w(n)$$
$$H_1 : x(n) = As(n) + w(n), \qquad A \neq 0$$

That is

$$H_0 : A = 0$$
$$H_1 : A \neq 0$$

This is called a <u>two-sided</u> test, and UMPs <u>never</u> exist for such tests. We must be content with a suboptimal detector.

What can we do?

Consider the <u>scalar</u> case:

$$H_0: \quad x \sim \mathcal{N}(0, \sigma^2) \qquad \sigma^2 \text{ known}$$

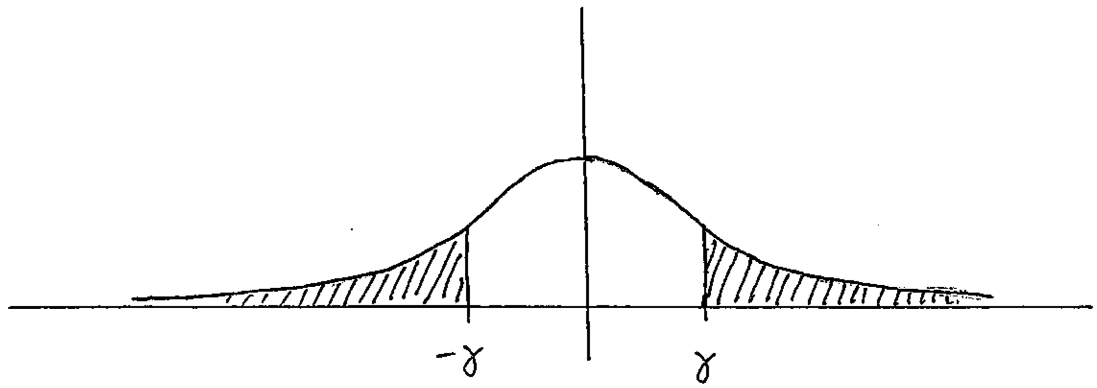
$$H_1: \quad x \sim \mathcal{N}(A, \sigma^2) \qquad A \neq 0, \text{ unknown}$$

Intuitively, the decision rule

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

comes to mind.

Large excursions of the observation $x$ from 0 may indicate the signal is present.



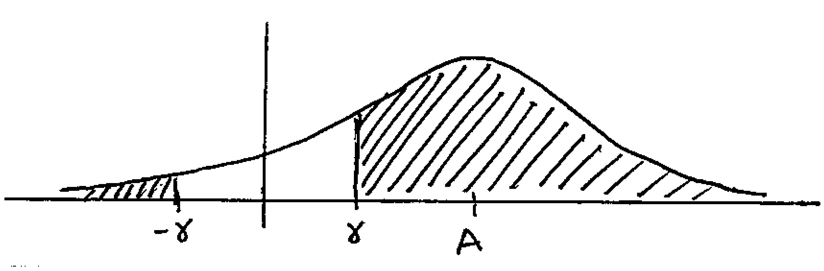

$H_0$ does not depend on $A$, so we may set the threshold $\gamma$ by constraining $P_F$:

$$P_F = 2\int_\gamma^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\,dx = 2Q\left(\frac{\gamma}{\sigma}\right)$$

$$\Rightarrow \quad \gamma = \sigma Q^{-1}\left(\frac{P_F}{2}\right)$$

In terms of $A$, the detection probability is

(b) $P_D =$



To evaluate our suboptimal detector, we can compare it to the <u>clairvoyant</u> detector, which assumes full knowledge of unknowns.

What is the clairvoyant detector for this problem (unknown amplitude)?

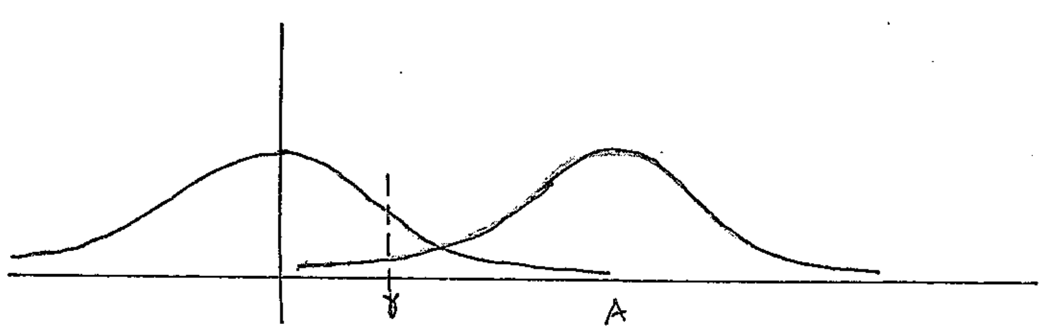
If $A > 0$,

$$x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \equiv \frac{\sigma^2}{A} \log(\eta) + \frac{A}{2}$$

c $\Rightarrow P_F =$

$\gamma =$

$P_D =$



If $A < 0$

$$x \underset{H_1}{\overset{H_0}{\gtrless}} \frac{\sigma^2}{A}\log(\eta) + \frac{A}{2} \equiv \gamma$$

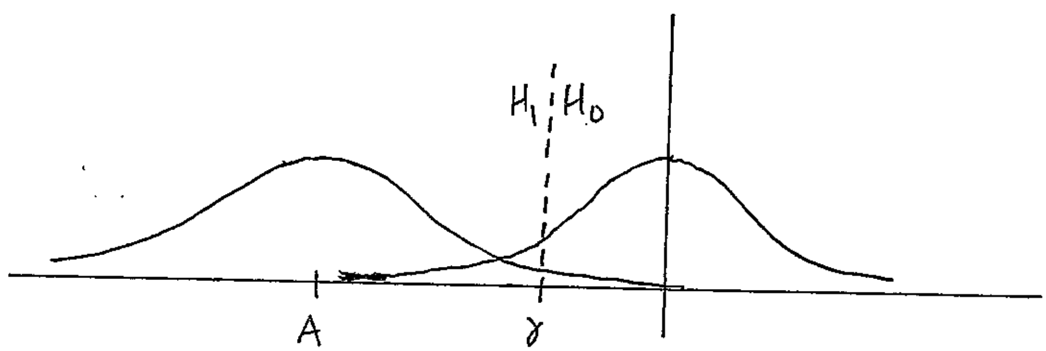
$\boxed{\text{inequalities reversed}}$

$$\Rightarrow \quad P_F = 1 - Q\left(\frac{\gamma}{\sigma}\right) = Q\left(\frac{-\gamma}{\sigma}\right)$$

$$\gamma = -\sigma Q^{-1}(P_F)$$

$$P_D = 1 - Q\left(\frac{\gamma - A}{\sigma}\right) = Q\left(\frac{A-\gamma}{\sigma}\right)$$

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




Summary

$$A > 0: \quad P_D = Q\left(Q^{-1}(P_F) - \frac{A}{\sigma}\right)$$

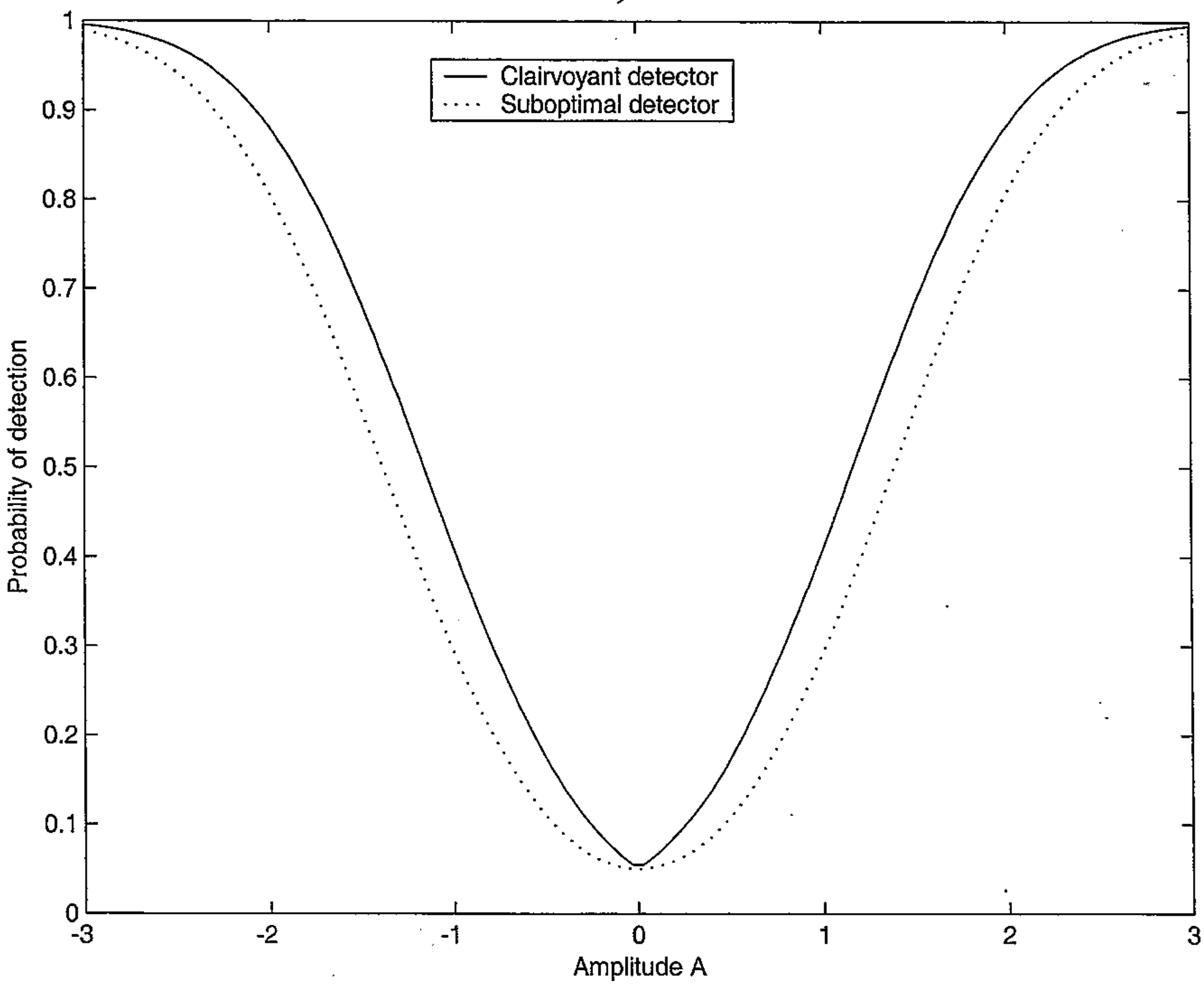
$$A < 0: \quad P_D = Q\left(Q^{-1}(P_F) + \frac{A}{\sigma}\right)$$

We can combine these to obtain

$$P_D = Q\left(Q^{-1}(P_F) - \sqrt{\frac{A^2}{\sigma^2}}\right)$$

for all $A \neq 0$.

$P_F = 0.5, \quad \sigma^2 = 0.5$




<u>Clairvoyant</u>: $P_D = Q\left(Q^{-1}(P_F) - \sqrt{\frac{A^2}{\sigma^2}}\right)$

<u>Suboptimal</u>: $P_D = Q\left(Q^{-1}\left(\frac{P_F}{2}\right) - \sqrt{\frac{A^2}{\sigma^2}}\right)$

$$+ Q\left(Q^{-1}\left(\frac{P_F}{2}\right) + \sqrt{\frac{A^2}{\sigma^2}}\right)$$

Let's return to the vector case:

$$H_0 : \underline{x} = \underline{w}$$

$$H_1 : \underline{x} = A\underline{s} + \underline{w} \quad , \quad A \neq 0$$

How might we generalize our previous detector?

Recall the LRT reduces to

$$A \underline{s}^T \underline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \log(\eta) + A^2 \frac{\underline{s}^T \underline{s}}{2}$$

For a suboptimal detector we could take

$$|\underline{s}^T \underline{x}| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

<u>Exercise</u> Derive $P_D$ as a function of $A$, $P_F$, $\underline{s}$, and $\sigma^2$, and compare to clairvoyant detector.

$\boxed{\text{Summary}}$

- Most real-world detection problems involve un
- In very special cases: LRT* independent of unk
- One-sided problems: If LR is monotone, UM test exists.
- Two-sided problems: UMP tests never exist, but reasonable suboptimal detectors do.
- Next lecture: General strategies for devising suboptimal detectors when no UMP test exists.

<u>Key</u> a. $\underline{s}^T \underline{x} \sim N(0, \sigma^2 \underline{s}^T \underline{s})$ under $H_0$

$$P_F = P(\underline{s}^T \underline{x} > \gamma \mid H_0) = Q\left(\frac{\gamma}{\sigma\sqrt{\underline{s}^T\underline{s}}}\right) = \alpha$$

$$\Rightarrow \gamma = \sigma\sqrt{\underline{s}^T\underline{s}}\; Q^{-1}(\alpha)$$

b. $P_D = P(|X| > \gamma \mid H_1) = P(X > \gamma \mid H_1) + P(X < -\gamma \mid H_1)$

$$= Q\left(\frac{\gamma - A}{\sigma}\right) + Q\left(\frac{\gamma + A}{\sigma}\right)$$

c. $P_F = Q\left(\frac{\gamma}{\sigma}\right)$, $\gamma = \sigma Q^{-1}(P_F)$,

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