Distributed Detection and Data Fusion

Peter Willett
University of Connecticut
2016

Outline

- Detection basics
- Distributed detection and decision fusion
- Some fun pathologies
  - Identical sensors can be different
  - Dependence
- New structures
  - Censoring sensors
  - Feedback of decisions
  - Sequential networks
  - Learning decision-makers' biases
- Other topics
  - Byzantine sensors
  - Secrecy and malicious sensors
  - Sequential and quickest tests

Detection and Decision-Making

- Problem is to choose between

$$H_0 \quad : \quad X_i \sim P_{0i}$$
$$H_1 \quad : \quad X_i \sim P_{1i}$$

- Optimal decision uses likelihood ratio (LRT) to compute probability of deciding H_1:

$$\phi(x) = \begin{cases} 1 & \text{if } L(x) > \tau \\ \gamma & \text{if } L(x) = \tau \\ 0 & \text{if } L(x) < \tau \end{cases} \Leftarrow L(x) = \prod_{i=1}^{n} \frac{p_1(x_i)}{p_0(x_i)} \rightarrow \lambda(x) = \sum_{i=1}^{n} \log\left(\frac{p_1(x_i)}{p_0(x_i)}\right)$$

- For example in Gaussian shift-in-mean case:

$$\lambda(x) = \log\left(\prod_{i=1}^{n} \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-x_i^2/2\sigma^2}}\right) = \frac{\mu}{\sigma^2} \sum_{i=1}^{n} \left[x_i - \frac{1}{2}\mu\right] \Rightarrow T(x) = \sum_{i=1}^{n} x_i$$

Optimality?

- Neyman-Pearson
  - Maximize P_d = Pr(decide H_1 | H_1 true) for given specified level of P_{fa} = Pr(decide H_1 | H_0 true)
  - LRT is optimal, threshold τ determined by P_{fa}
- Bayes
  - Assume "costs" c_{ij} = Cost of deciding H_i when H_j is true
  - Assume prior probabilities Pr(H_i)
  - LRT is optimal, threshold

$$\tau = \frac{Pr(H_0)[c_{10} - c_{00}]}{Pr(H_1)[c_{01} - c_{11}]}$$ and if P(e) minimized: $$\tau = \frac{Pr(H_0)}{Pr(H_1)}$$

Performance

- In some cases you can compute the performance
  - T(x) is Gaussian:

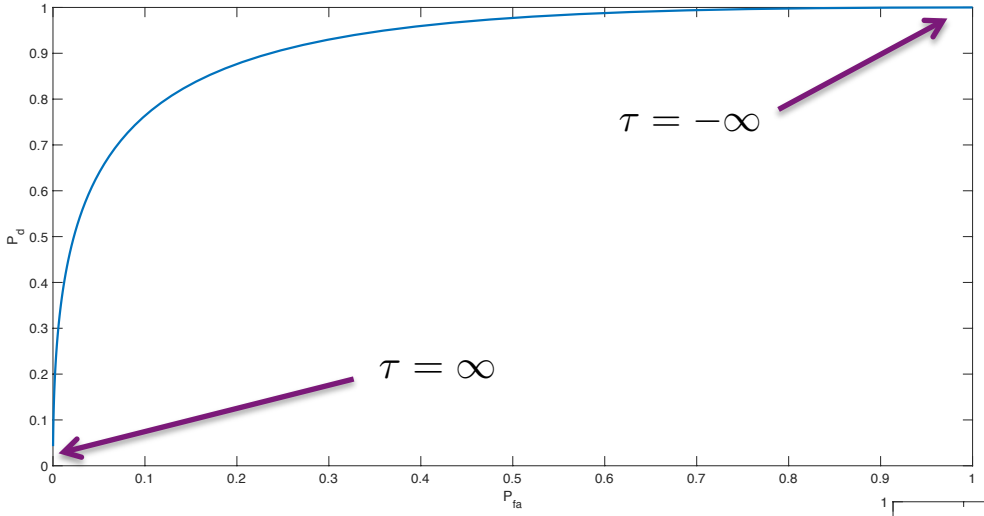
$$P_{fa} = \Pr\left(T(x) > \tau \mid H_0\right) = Q\left[\frac{\tau - \mu_0}{\sigma_0}\right]$$

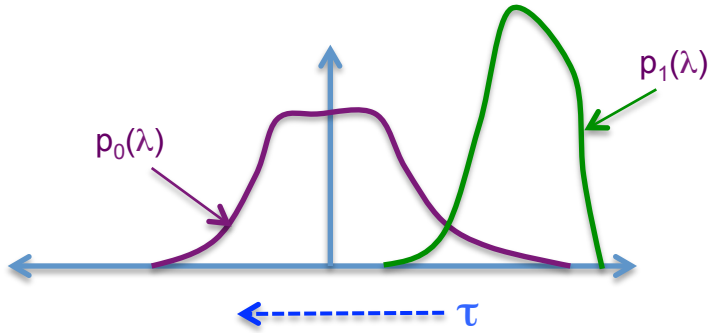
$$P_d = \Pr\left(T(x) > \tau \mid H_1\right) = Q\left[\frac{\tau - \mu_1}{\sigma_1}\right]$$

$$= Q\left[\frac{\sigma_0 Q^{-1}(P_{fa}) - (\mu_1 - \mu_0)}{\sigma_1}\right]$$

  - The Q-function is the unit-Gaussian tail probability.
  - The importance of the SNR (difference in means divided by the standard deviation) is obvious.
  - In many cases – especially asymptotic ones – the test statistic is approximated as Gaussian, at least under H_1.

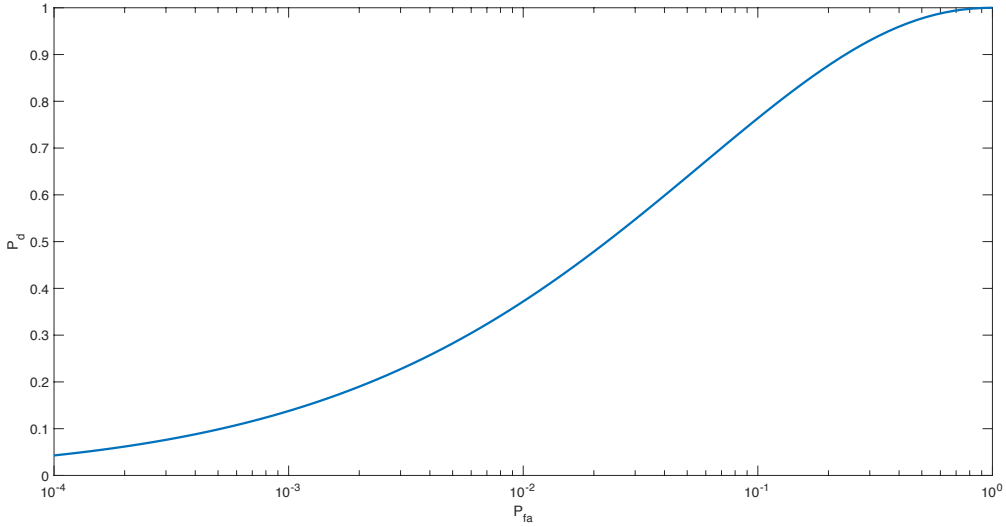
Receiver Operating Characteristic





ROC: Plot of probability of detection versus probability of false alarm.

Sometimes a plot of probability of detection versus signal to noise ratio for a fixed false alarm rate is also called an "ROC."



Discrete Data

- Suppose our testing problem is

$$H_0: \; Pr(x_i = 1) \;=\; 1 - Pr(x_i = 0) \equiv q_0$$

$$H_1: \; Pr(x_i = 1) \;=\; 1 - Pr(x_i = 0) \equiv q_1$$

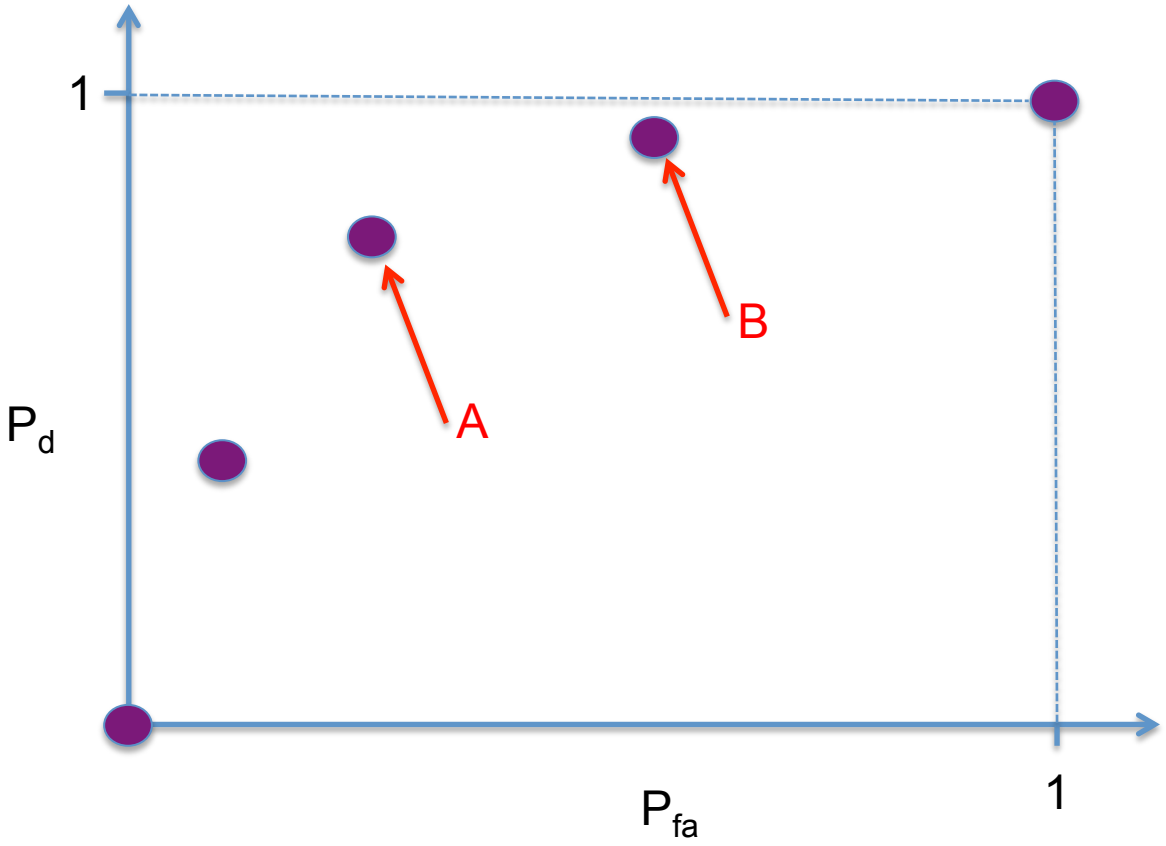
- Then our test statistic is

$$\lambda(x) \;=\; \log\left(\prod_{i=1}^{n} \frac{q_1^{x_i}(1-q_1)^{1-x_i}}{q_0^{x_i}(1-q_0)^{1-x_i}}\right) \;=\; n\log\left(\frac{1-q_1}{1-q_0}\right) + \log\left(\frac{q_1(1-q_0)}{(1-q_1)q_0}\right)\sum_{i=1}^{n} x_i \quad \Longrightarrow \quad T(x) \;=\; \sum_{i=1}^{n} x_i$$

- Now we can only reach certain probabilities of false alarm:

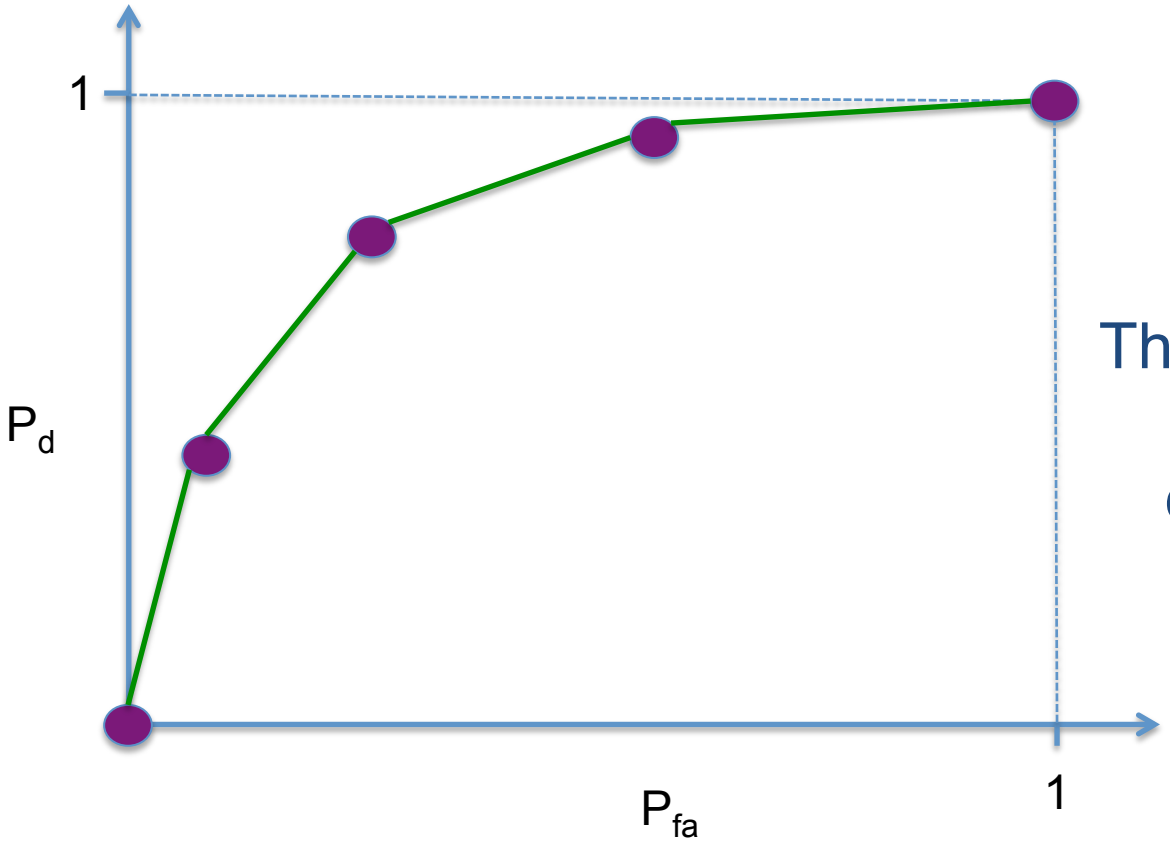
$$P_{fa} \;=\; 0, \quad q_0^n, \quad n(1-q_0)q_0^{n-1}, \; \ldots$$

Discrete ROC

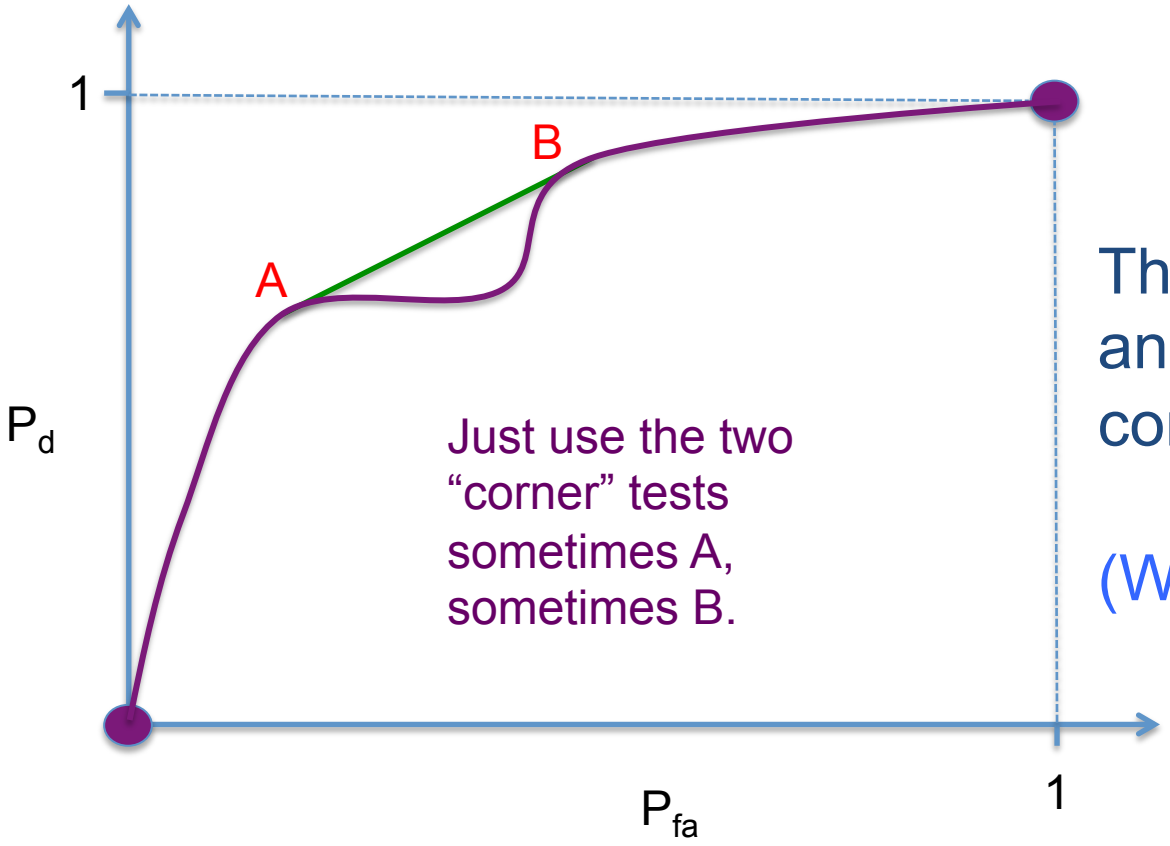


Suppose we used test A 50% of the time and test B 50% of the time?

Randomization



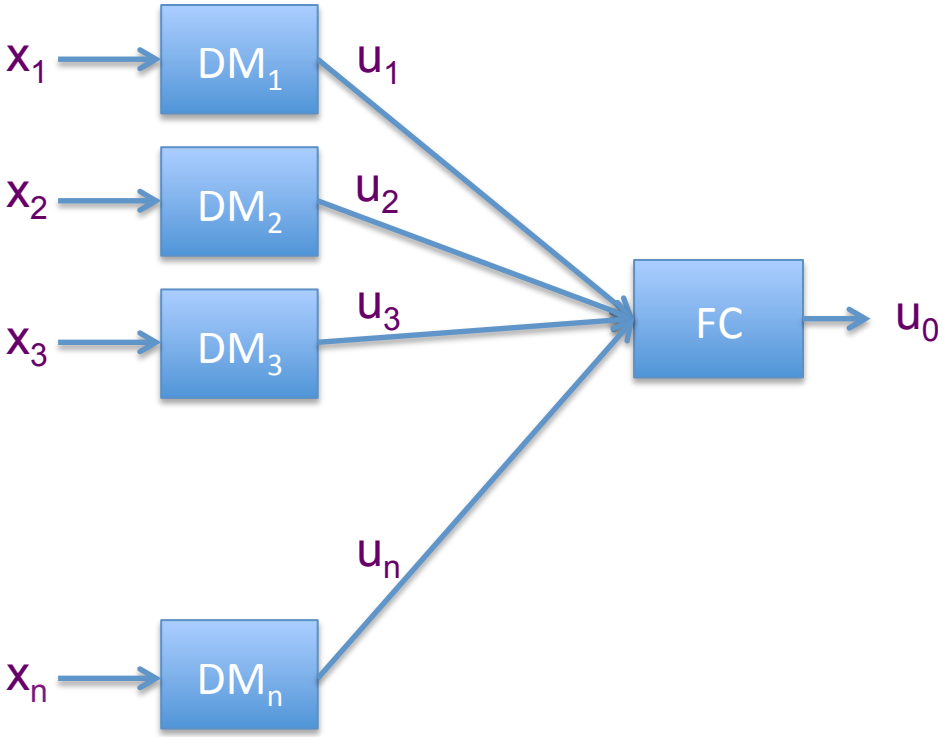
This is one reason why an ROC has to be concave – always!



This is one reason why an ROC has to be concave – always!

(Well, almost always.)

Decentralized Detection



Individual decision-makers (DMs) ingest local observations (x_i) and provide a quantized version (u_i) of that data to the fusion center (FC) who makes the final decision u_0.

For now assume the x_i's are independent conditioned on the hypothesis.

For now also assume the quantization is binary – that is, the u_i's are "local decisions".

- Tenney & Sandell, "Detection with distributed sensors," TAES 1981.
- Chair & Varshney "Optimal data fusion in multiple sensor detection systems," TAES 1986.

The Chair/Varshney Result for the FC

- In a sense, we've already seen it: it's the case with binary observations, which is a "counting rule" for observations that are *iid (k*-out-of-*n)*:

$$H_0: \; Pr(x_i = 1) = 1 - Pr(x_i = 0) \equiv q_{0i}$$
$$H_1: \; Pr(x_i = 1) = 1 - Pr(x_i = 0) \equiv q_{1i}$$

$$\Rightarrow \quad \lambda(x) = \sum_{i=1}^{n} \log\left(\frac{q_{1i}(1-q_{0i})}{(1-q_{1i})q_{0i}}\right) x_i = \sum_{i=1}^{n} w_i x_i$$

- The cases k=1 and k=n are interesting: OR and AND.
  - Naturally an OR rule needs much more selective decision-makers, since the low threshold at the fusion center requires high thresholds at the local DM level.
  - It is not clear in general which is better.

What about the DMs?

- It's fairly easy to show that the DMs are likelihood ratio tests in the case of binary observations:
  - The FC benefits from the best q_0 and q_1.
  - That means a LRT at the local DM level.
- In the more general case of multi-level observations the situation is not so clear.
  - But is turns out to be the natural extension.
  - The DMs quantize their local likelihood ratios.
- Hence both DM and FC are likelihood ratio tests.
  - The FC is fairly simple given the DM, but optimizing the DMs' quantization rules is not straightforward.
- This is true even when the channels are error prone.

- Tsitsiklis "Decentralized Detection," (in Advances in Statistical Signal Processing, vol. 2), 1990.
- Chen & Willett, "On the Optimality of the Likelihood Ratio Test for Local Sensor Decision Rules in the Presence of Non-Ideal Channels" T-IT 2005.

Calculation of DM rules

- Using Tang, Pattipati & Kleinman's idea, an optimization can proceed using Gauss-Seidel:
  - Fix the FC rule.
  - Guess the quantizations at the DMs.
  - Optimize DM_1's rule given all others fixed.
  - Proceed to DM_2 and continue.
- It does converge, although no proof to an optimum.
- Requires independent DM's.
- Must be done for all FC rules.
  - Fortunately there are only a finite number.

- Tang, Pattipati & Kleinman, "An algorithm for the detection thresholds in a distributed detection problem," SMC-A 1991.

More Comprehensive DM Rule

- Blum provided the following iteration:
  - Define "everyone else": $\tilde{u}_k \equiv \{u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n\}$
  - Define:

$$D_{jk}(x_k) \equiv p_j(x_k) \left[\sum_{\tilde{u}_k} p(u_0 = 1 | \tilde{u}_k, u_k = 1) - p(u_0 = 1 | \tilde{u}_k, u_k = 0) \right] p(\tilde{u}_k | x_k, H_j)$$

  - Then the DM rule must be a "likelihood ratio" test on:

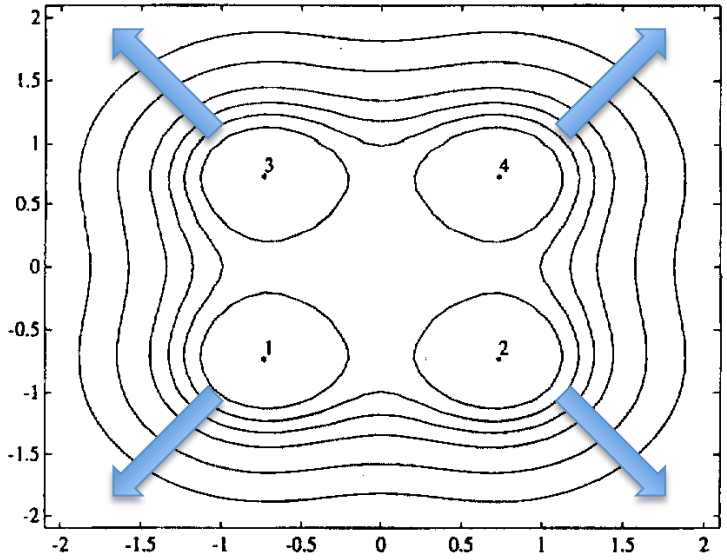
$$D_{1k}(x_k) / D_{0k}(x_k)$$

  - The recursion is similar, except that the FC rule changes in each iteration
- This is more general in the sense that the FC rule is implicitly optimized and independence is not necessary.
- However there is no general rule for *M*-ary quantizers.

- Blum, "Necessary Conditions for Optimum Distributed Detectors Under the Neyman-Pearson Criterion, T-IT 1996.

An Example

- Consider we have 2, 3 or 4 sensors with an inverse-square power law and Swerling targets. We optimize over both DM and FC rules for each position:

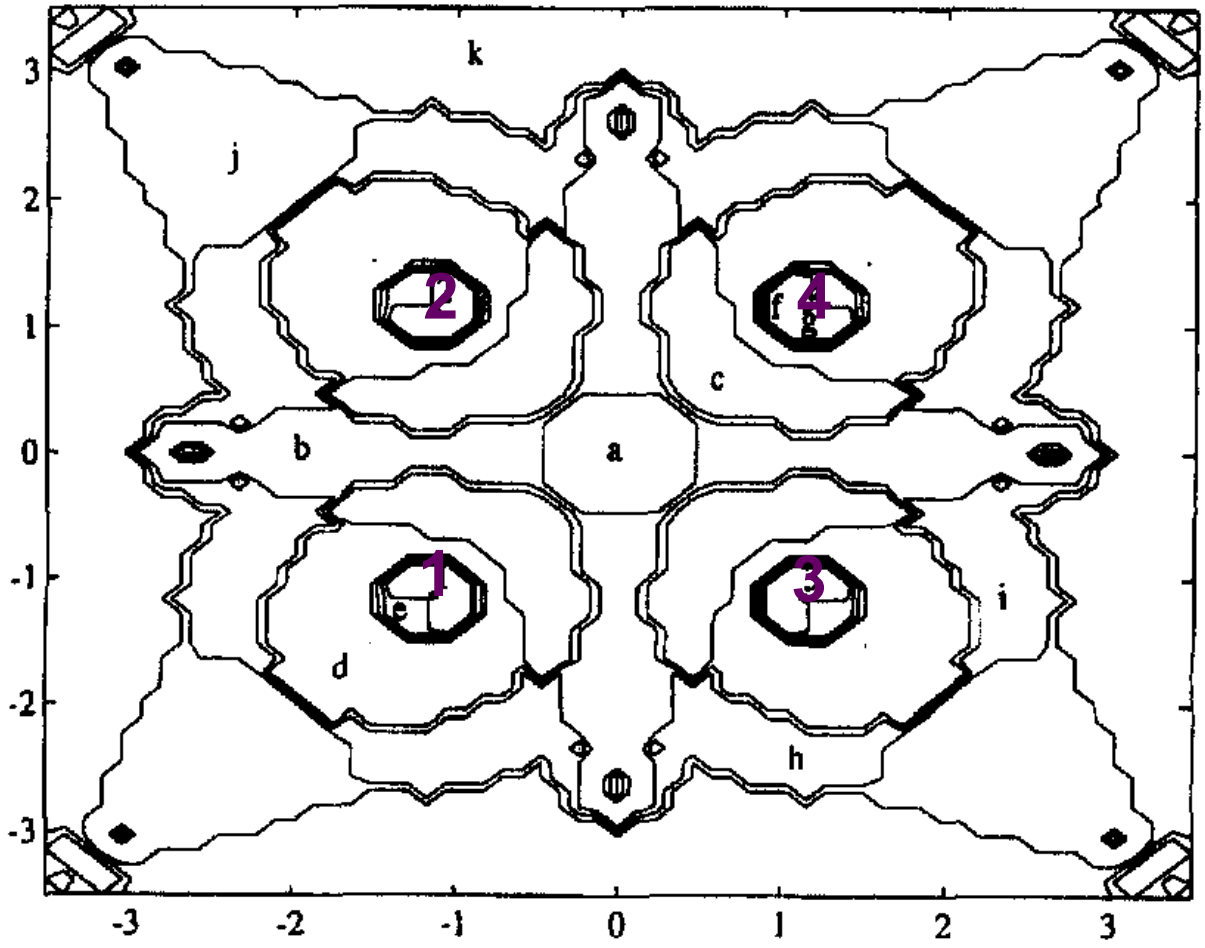


This plot shows contours of constant P_d with $P_{fa} = 10^{-5}$.

The sensors are moved apart to see the various fusion rules and whether the coverage area is improved.

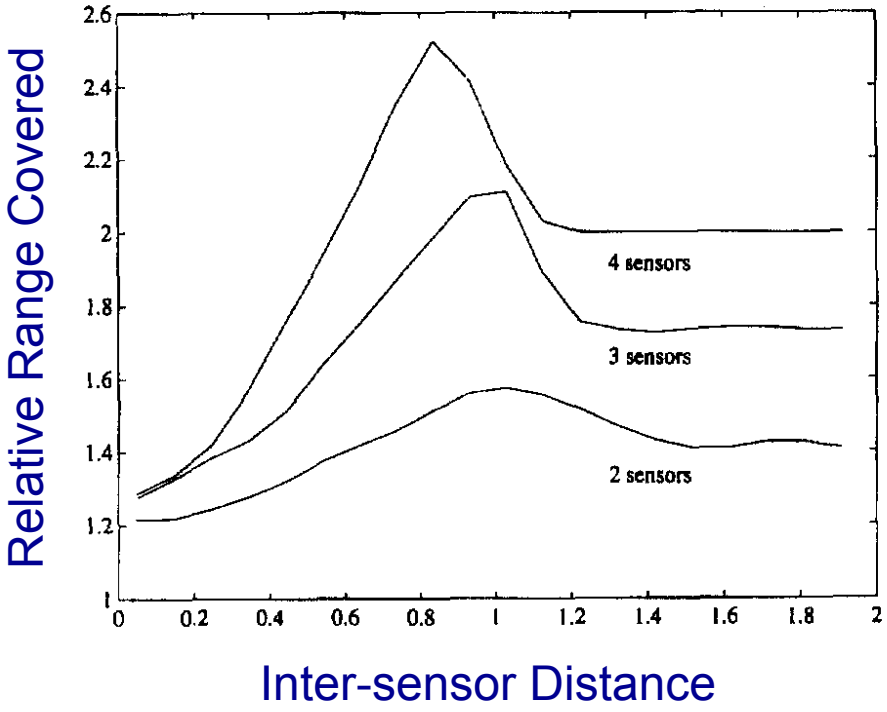
- Willett, Alford & Vannicola, "The Case for Like-Sensor Pre-Detection Fusion," TAES 1994.

Some Optimal Fusion Rules



Location	Fusion Rule	Cardinality
a	*at least three of* $\{u_1, u_2, u_3, u_4\}$	5
b	$u_1.u_2 + (u_1 + u_2).u_3.u_4$	6
c	$u_4.(u_1 + u_2 + u_3) + u_1.u_2.u_3$	8
d	$u_1.(u_2 + u_3 + u_4)$	7
e	$u_1 + u_2.u_3$	10
f	$u_4 + u_2.(u_1 + u_3)$	11
g	$u_4 + u_3.(u_1 + u_2)$	11
h	$u_3.(u_1 + u_2.u_4)$	5
i	$u_3.(u_1.u_2 + u_4)$	5
j	$u_2.$(*at least two of* $\{u_1, u_3, u_4\}$)	4
k	$u_2.u_4.(u_1 + u_3)$	3

The Advantage



The message here is that distributed detection *can* help in a certain "sweet spot" where cooperation between the DMs is effective.

How Much Do You Lose?

- Let's compare $T_l(x) = \sum_{i=1}^{n} x_i$ to $T_s(x) = \sum_{i=1}^{n} \text{sign}(x_i)$ in $\mathcal{N}(\pm\mu, \sigma^2)$

- SNR for linear detector is $\frac{(\mathcal{E}\{x|H_1\} - \mathcal{E}\{x|H_0\})^2}{\mathcal{V}\{x|H_1\}} = \frac{(2\mu)^2}{\sigma^2} = \frac{4\mu^2}{\sigma^2}$
- SNR for sign detector is

$$\frac{(\mathcal{E}\{\text{sign}(x)|H_1\} - \mathcal{E}\{\text{sign}(x_i)|H_0\})^2}{\mathcal{V}\{x|H_1\}}$$

$$= \frac{\left(\left[\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} - \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] - \left[\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x+\mu)^2}{2\sigma^2}} - \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x+\mu)^2}{2\sigma^2}}\right]\right)^2}{1}$$

$$= \left(\int_{-\mu}^{\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\right)^2 \approx \left(4\mu \frac{1}{\sqrt{2\pi\sigma^2}}\right)^2 = \frac{8\mu^2}{\pi\sigma^2}$$

- This is why "you lose 2dB" ($2/\pi$).

Alternative Ways to Optimize

- Traditional criteria:
  - Neyman-Pearson maximizes P_d for fixed P_{fa}.
  - Bayes minimizes average cost (needs priors).
  - Can show that Bayes optimal is optimal for its (P_d, P_{fa}).
- Can also design DM to optimize mutual information:

$$\mathcal{I}(u_0; H) = \sum_{H \in \{H_0, H_1\}} \sum_{u_0 \in \{0,1\}} Pr(u_0, H) \log \left(\frac{Pr(u_0, H)}{Pr(u_0) Pr(H)} \right)$$

Unfortunately this turns out to be no simpler than optimizing under Neyman-Pearson or Bayes.

  - The reason is that u_0 requires the fusion rule.

Simpler Suboptimal Criteria

- These simply try to pass good DM data to the FC
  - mutual information:

$$\mathcal{I}(\{u_i\}; H) = \sum_{i=1}^{n} \sum_{H \in \{H_0, H_1\}} \sum_{u_i} Pr(u_i, H) \log \left(\frac{Pr(u_i, H)}{Pr(u_i) Pr(H)} \right)$$

  - J-divergence:

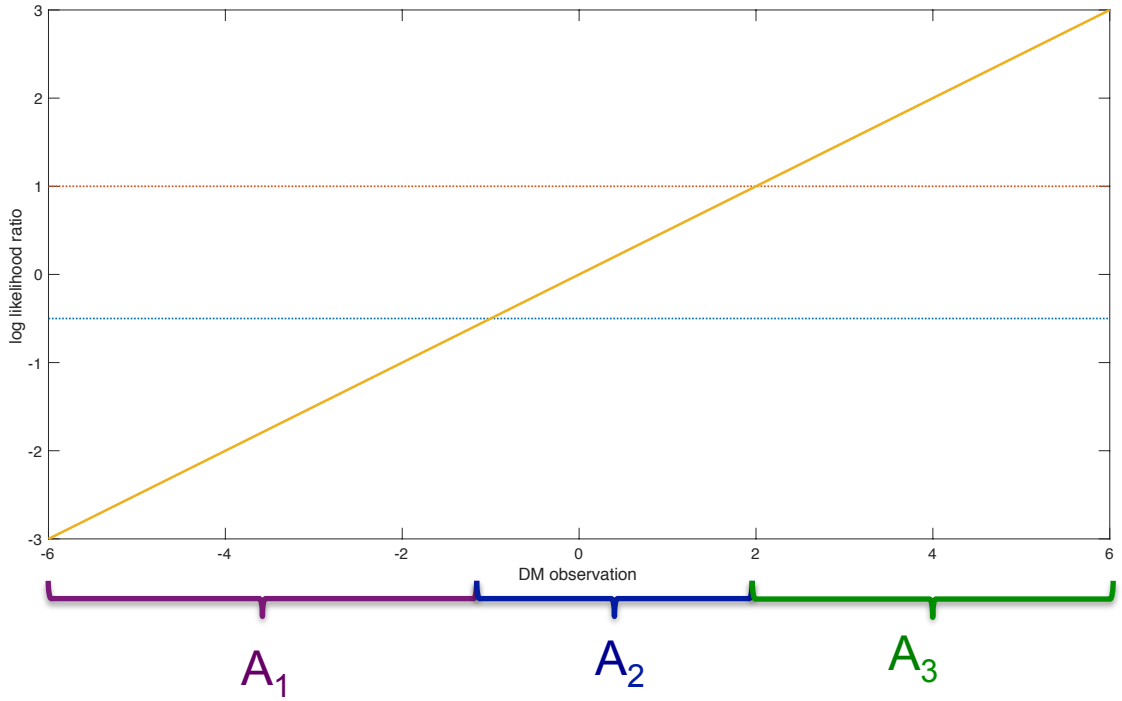
$$J(\{u_i\}) = \sum_{i=1}^{n} \left[\mathcal{E} \left\{ \log \left(\frac{Pr(u_i|H_1)}{Pr(u_i|H_0)} \right) \mid H_1 \right\} - \mathcal{E} \left\{ \log \left(\frac{Pr(u_i|H_1)}{Pr(u_i|H_0)} \right) \mid H_0 \right\} \right]$$

  - Bhattacharyya affinity:

$$B(\{u_i\}) = \sum_{i=1}^{n} \sum_{u_i} \sqrt{Pr(u_i|H_1) Pr(u_i|H_0)}$$

  - Also efficacy, KL-divergence … any Ali-Silvey distance.
- It can be shown that all these result in likelihood ratio quantizations at the sensors.

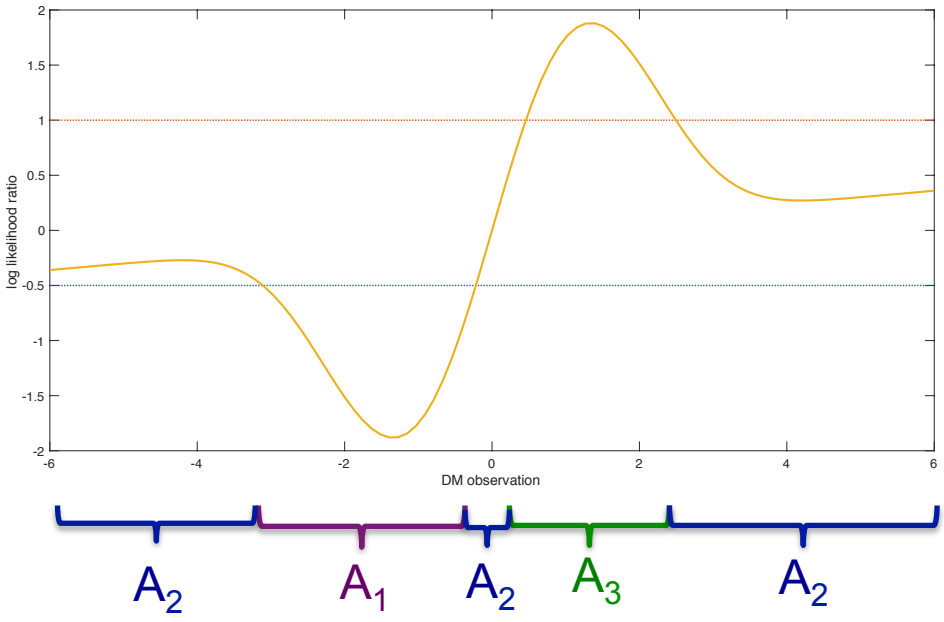
Why Care About The Quantization?

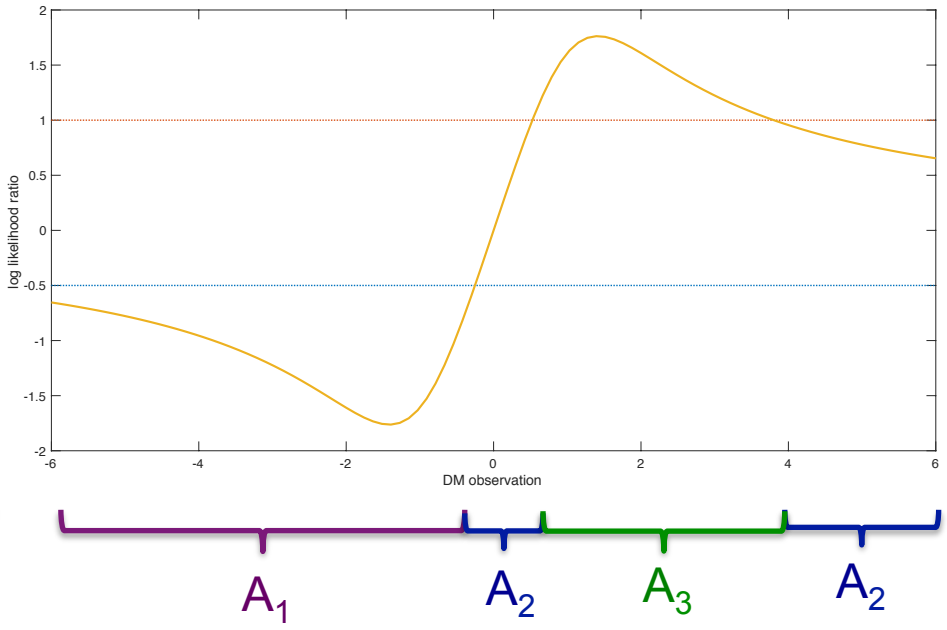


The Gaussian case (additive signal: +/-1) with M=3 levels of quantization corresponding to LLR thresholds -0.5 and +1.0.

- Willett & Warren, "Optimum Quantization for Detector Fusion: Some Proofs, Examples, and Pathology," JFI 1999.

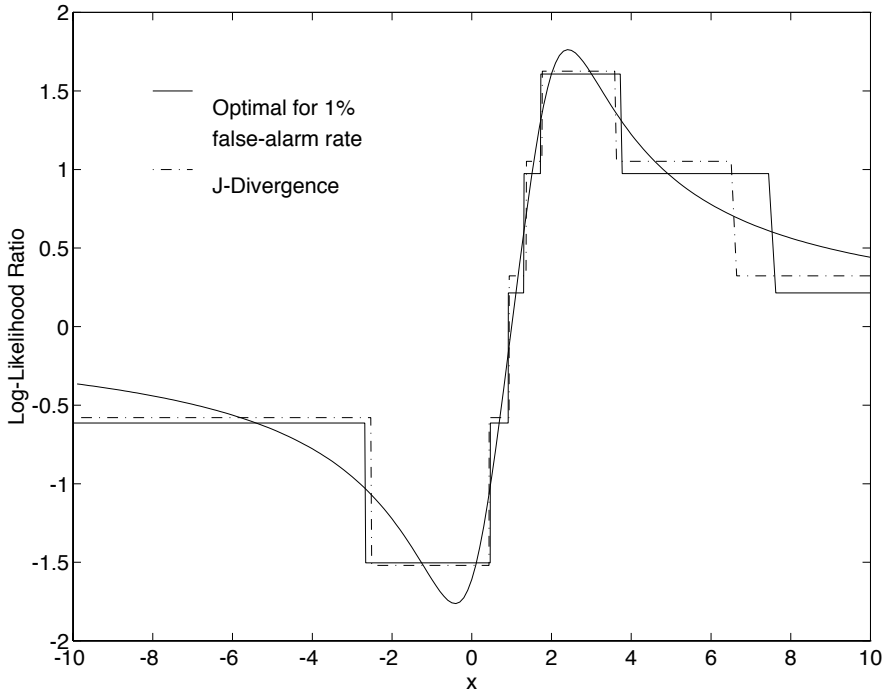
Gauss-mixture and Cauchy Cases





As can be seen, the quantization regions are no longer simply-connected in the observation space.

Use of Distance Proxies



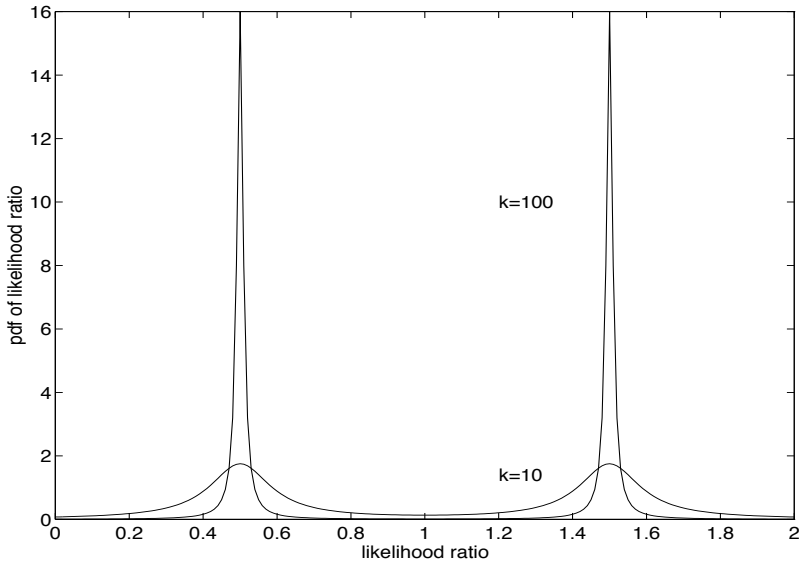
This shows the result of optimizing J-divergence for a 4-level quantization in Cauchy noise. Note the similarity of the optimized quantizer to the proxy one. Note also that the proxy quantizer does not depend on the desired false alarm rate, while the N-P optimal quantizer does.

Some Strange Things Happen

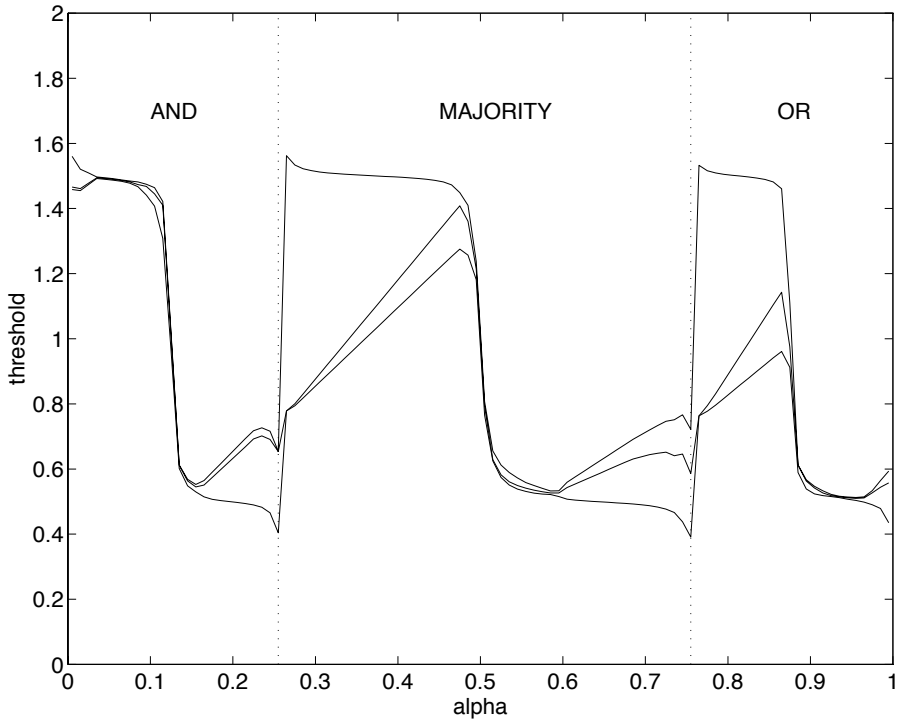
- Suppose we have the likelihood ratio H_0 pdf

$$f_{H_0}(x) = \begin{cases} \frac{c}{1+\left[k(x-\frac{1}{2})\right]^2} + \frac{c}{1+\left[k(x-\frac{3}{2})\right]^2} & 0 \le x \le 2 \\ 0 & else \end{cases}$$

- We need to have this have unity mean for validity.



We Optimize for n=3 "Identical" DMs



The optimal fusion rule changes from AND to Majority-logic to OR as P_{fa} increases.

In the case k=100 (very nearly point masses) the optimal thresholds turn out to be different at the various DMs.

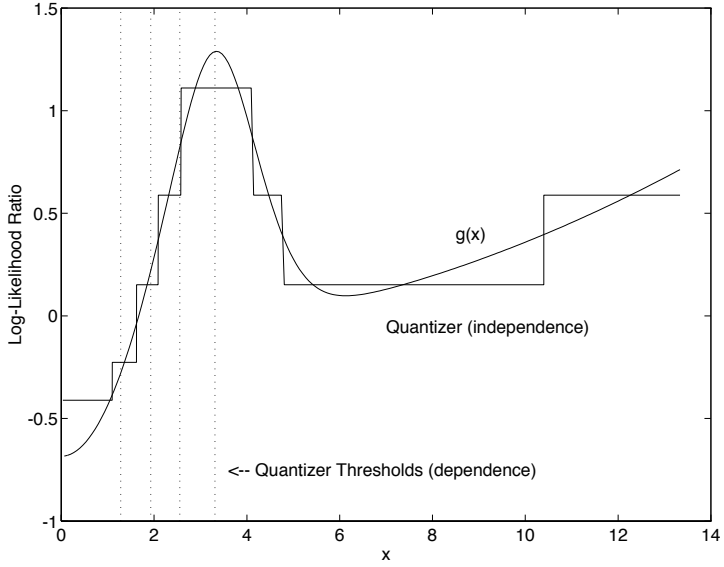
Dependent DMs

- Consider the case n=5 and Gaussian mixture noise

$$H_0: \quad Y_i = (1-Z)N_{1i} + ZN_{2i}$$

$$H_1: \quad Y_i = S_i + (1-Z)N_{1i} + ZN_{2i}$$

- Either all DMs get low noise or all high noise.



Quantizers optimized under Bhattacharyya criterion. Under an (incorrect) assumption of independence the result is a likelihood ratio quantization; correctly assuming dependence results in a *direct data* quantization! This makes sense in that information about Z is contain in the data.

The Good, Bad and Ugly

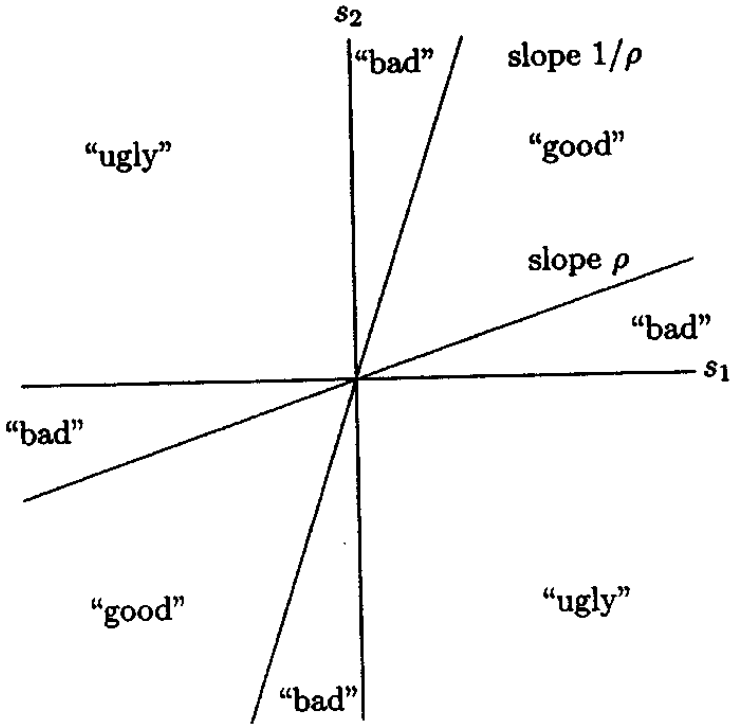
- Perhaps the simplest case of dependent observations we can explore for quantization is that of correlated Gaussian noise, two sensors, binary quantization and an additive signal.
- Mathematically:

$$H\colon\ x_1, x_2 \sim N(0, 0, 1, 1, \rho)$$
$$K\colon\ x_1, x_2 \sim N(s_1, s_2, 1, 1, \rho)$$

- There are clearly only three fusion rules possible: AND, OR and XOR.

- Chen & Papamarcou, "Likelihood ratio partitions for distributed detection in correlated Gaussian noise," ISIT 1996.
- Willett, Swasek & Blum, "The Good, Bad, and Ugly: Distributed Detection of Known Signal in Correlated Gaussian Noise," TSP 2000.

The Regions



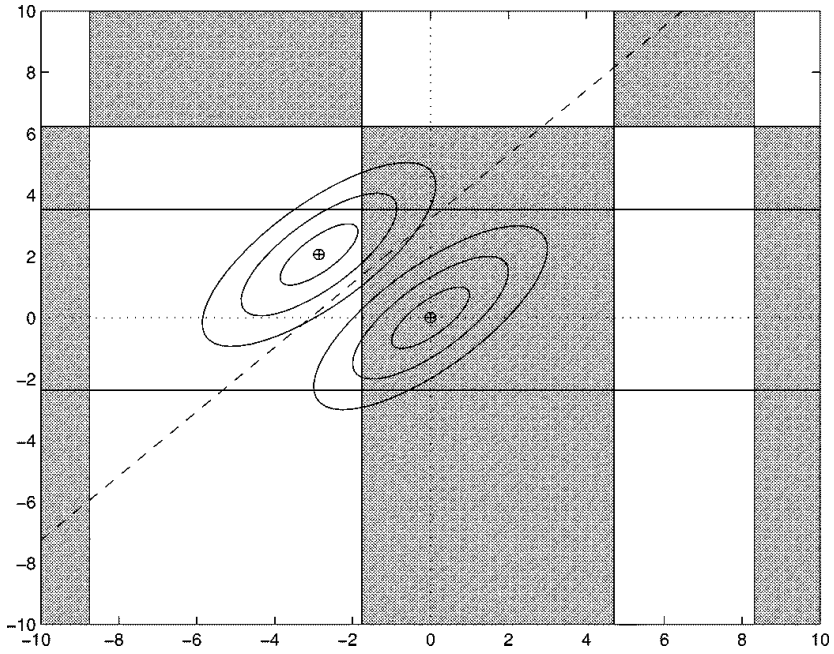
It turns out that we have different behaviors depending on the interplay between signal and correlation.

Good: Quantizers are single-interval and the FC is AND or OR.

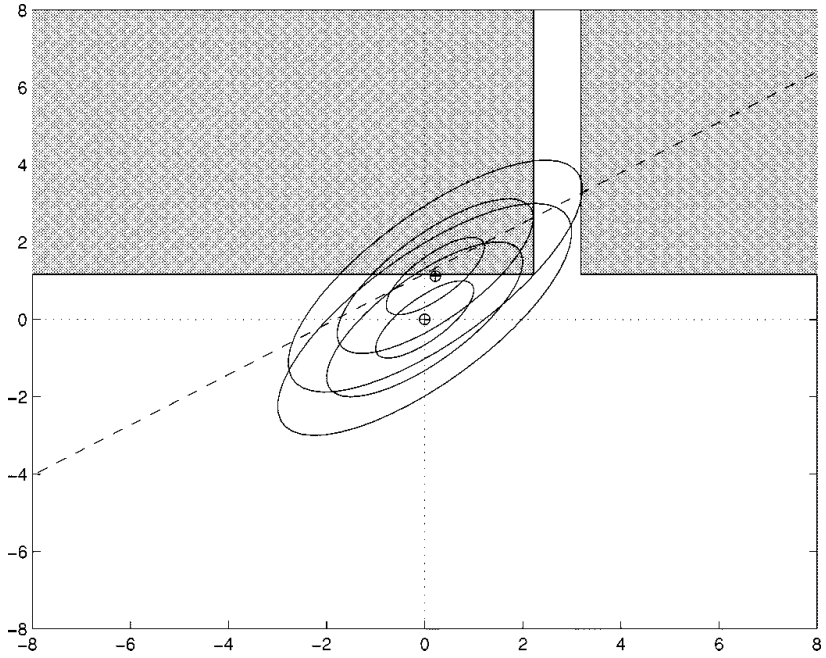
Bad: AND/OR quantizers either **ignore** one sensor or are **non-simply connected**.

Ugly: Little is known, except that AND/OR quantizers are often non-simply connected and that **XOR** rules can be optimal.

XOR is Optimal?



XOR-imputed decision regions.



AND-imputed decision regions.

XOR turns out to be optimal here. Neither is simply-connected.

Censoring Sensors

- In a communications-constrained system it is intuitive that one does not "send" information unless what one has is worth sending:

$$\left\{ \begin{array}{ll} l_i(\mathcal{X}_i) \in R_i & l_i(\mathcal{X}_i) \text{ is sent} \\ l_i(\mathcal{X}_i) \in \overline{R}_i & \text{nothing is sent} \end{array} \right\}$$

- Maximize FC's P_d subject to constraint on P_{fa} and

$$\sum_{i=1}^{N} \Pr(l_i(\mathcal{X}_i) \in R_i \mid H) \leq \kappa_{NP} \leq N$$

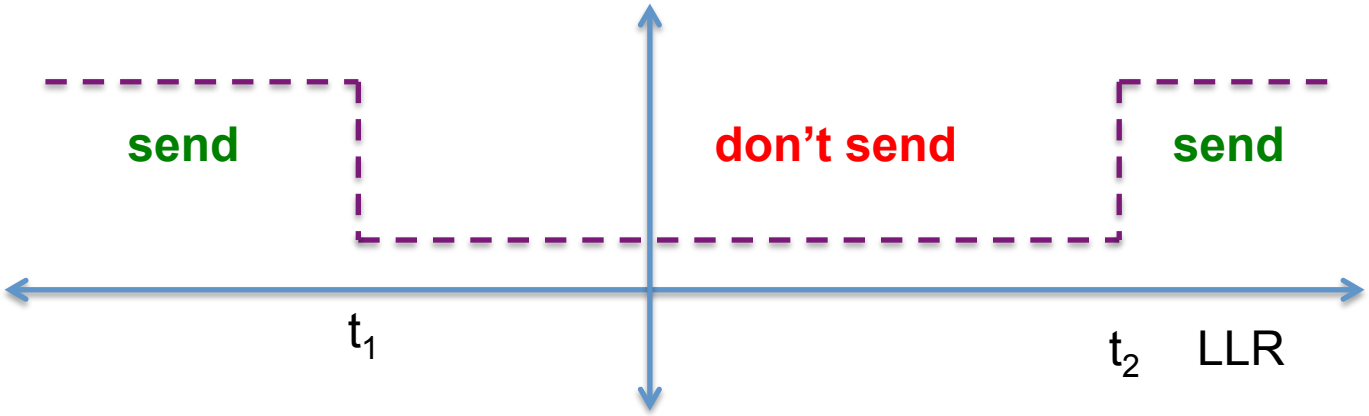
- Naturally there is a Bayesian version of this too.

- Rago, Willett & Bar-Shalom, "Censoring Sensors: A Low Communication Rate Scheme for Distributed Detection," TAES 1996.
- Appadwedula, Veeravalli & Jones, "Decentralized Detection with Censoring Sensors," T-IT 2008.

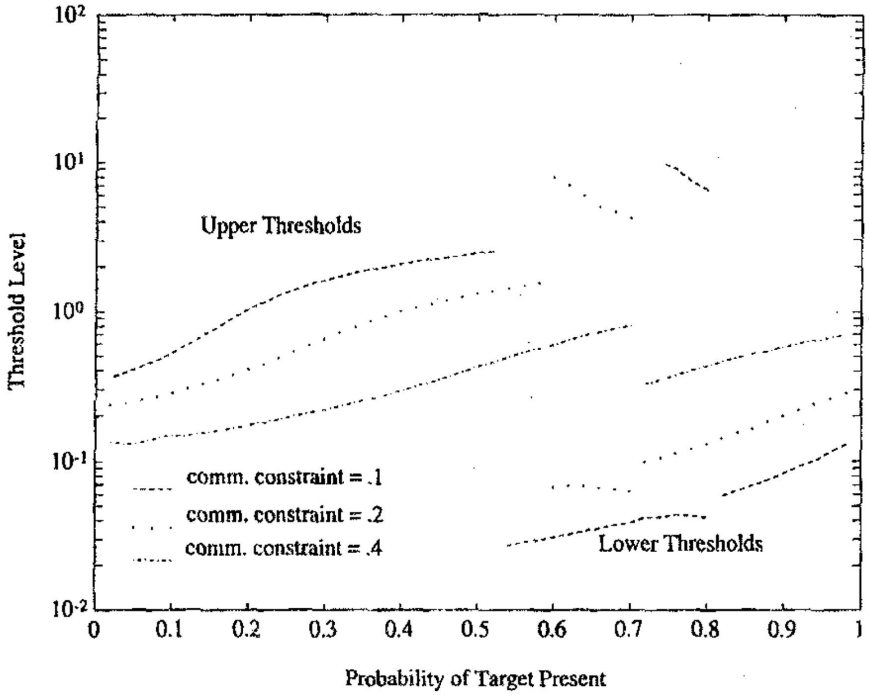
Censoring Region is an Interval

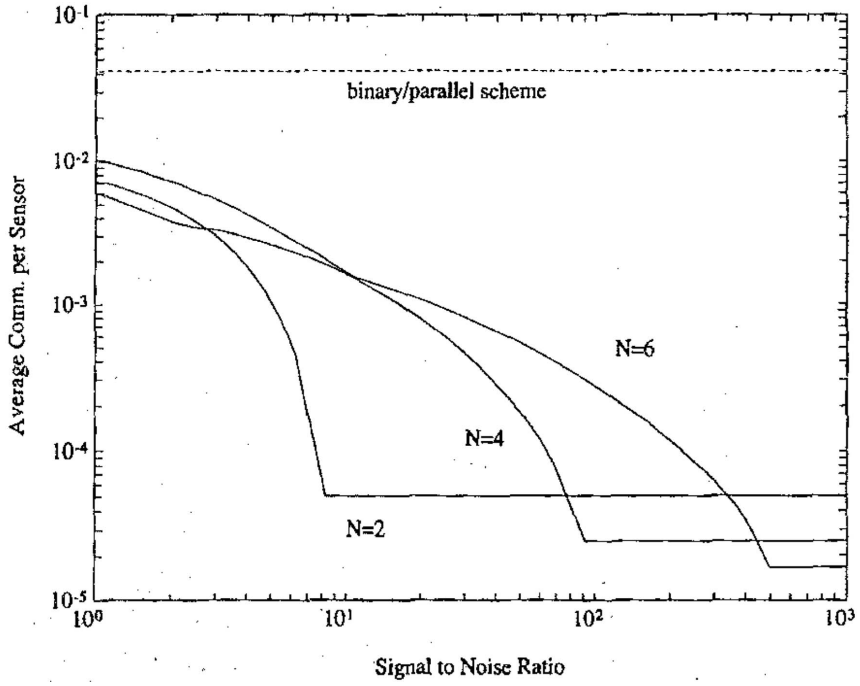
- The result is that the "no-send" LLR region is:

$$\overline{R}_i = \{l_i : t_{1i} \le l_i(\mathcal{X}) \le t_{2i}\}$$



Censoring: CA-CFAR Example



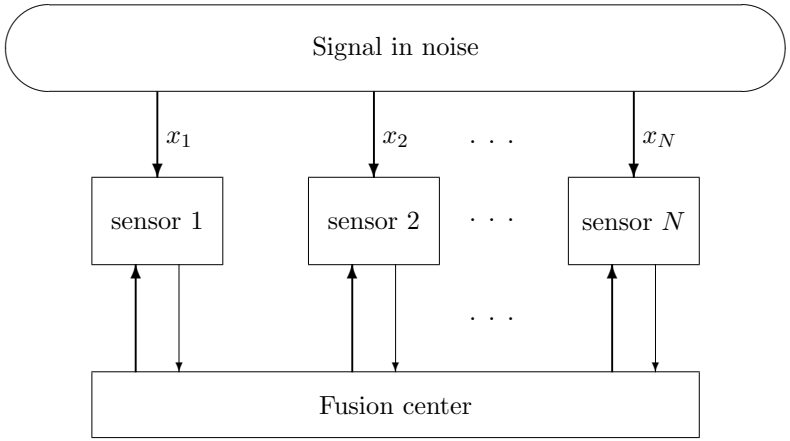


Optimal thresholds for CA-CFAR problem (SNR=10dB, 8 reference cells) and Bayesian case, plotted as a function of the target-presence probability.

Communication rate needed to match performance of uncensored scheme (assumed to use 24 bits). False alarm rate is 0.01%. N is number of sensors.

Feedback

- Suppose there is a backward path from FC to DM:



- The effect is to tell DM1 (who said “no”) that DM2, DM3 & DM4 said “yes”.
  - Is she is sure about her “no”?
- This amounts to a lowered threshold for DM1.

Feedback: Bayesian Idea

- The idea here is that each DM makes a Bayesian decision at each step to minimize P(e).
- The Bayesian LR test threshold is $\Pr(H_0)/\Pr(H_1)$.
- The feedback information means $\Pr(H_0)$ is modified to be "posterior": given all past information.
- That is, the test is:

$$\Lambda(x_i) \underset{u_{i,m}=0}{\overset{u_{i,m}=1}{\gtrless}} \frac{\pi_0 \Pr(\mathcal{U}_{i,m} \mid H_0)}{\pi_1 \Pr(\mathcal{U}_{i,m} \mid H_1)}$$

where $U_{i,m}$ is all previous data (before time m) from all DMs except DM *i*.

- Swaszek & Willett, "Parley as an Approach to Distributed Detection," TAES 1995.

Feedback: Bayesian Example

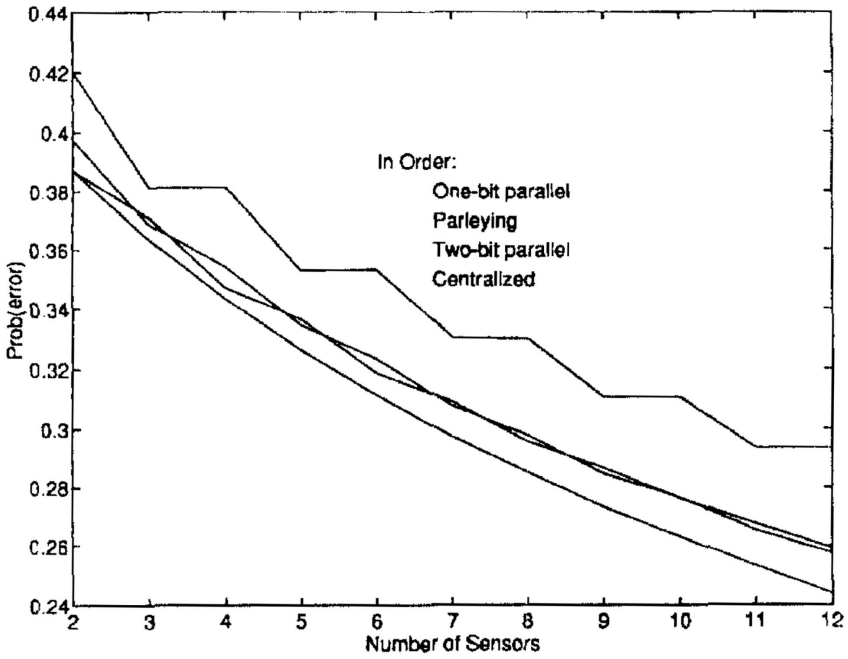
- With $\lambda_{i,m} \equiv \dfrac{\pi_0 \Pr(\mathcal{U}_{i,m} \mid H_0)}{\pi_1 \Pr(\mathcal{U}_{i,m} \mid H_1)}$ we have the example

	Sensor 1	Sensor 2	Sensor 3	Sensor 4
$\Lambda(x_i)$	.75	.92	1.15	1.03
$\lambda_{i,1}$	1.00	1.00	1.00	1.00
$u_{i,1}$	0	0	1	1
$\lambda_{i,2}$	0.85	0.85	1.17	1.17
$u_{i,2}$	0	1	0	0
$\lambda_{i,3}$	0.93	1.13	1.31	1.31
$u_{i,3}$	0	0	0	0

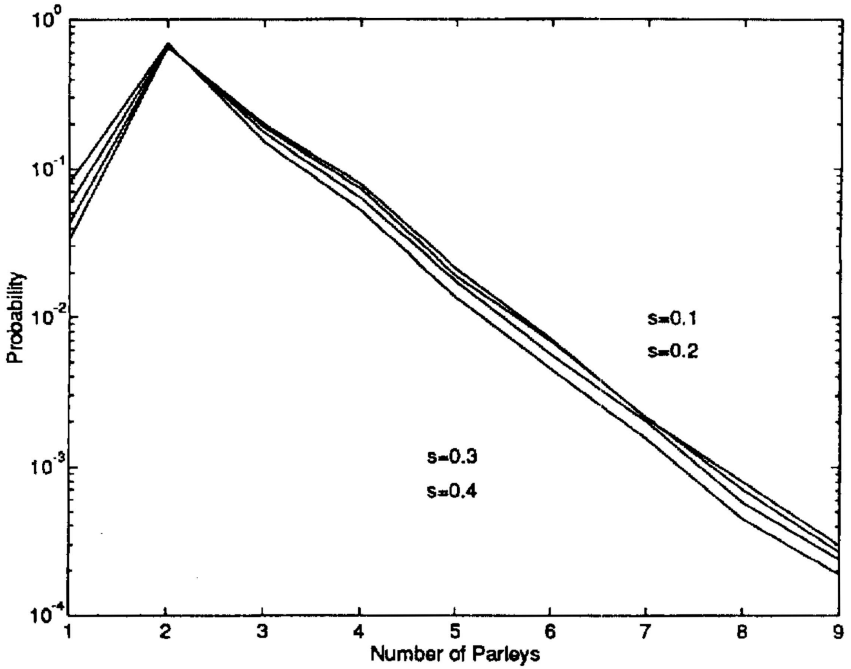
→ product is 0.82

- Apparently DM1 is pretty convincing.
- Convergence to a unanimous decision is guaranteed via a Martingale proof.

Feedback: Bayesian Results



Performance in the Gaussian shift-in-mean problem as function of number of DMs.



Most of these iterations end in a consensus after 2 rounds.

Feedback: The Nth-Root Idea

- If λ is the centralized Bayesian threshold, then we have $\prod_{i=1}^{n} \Lambda(x_i) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$ so why not have DM test according to

$$\Lambda(x_i) \underset{u_{i,m}=0}{\overset{u_{i,m}=0}{\gtrless}} \lambda_{i,m} \quad \text{where} \quad \prod_{i=1}^{n} \lambda_{i,m} = \lambda$$

- It is even simpler if we choose all the same: $\lambda_{i,1} = \lambda^{1/n}$
- We can show that we should use

$$\Lambda_m(x_i) \underset{u_{i,m}=0}{\overset{u_{i,m}=1}{\gtrless}} (\lambda_m^*)^{1/n} \quad \text{where}$$

$$\lambda_m^* = \lambda \prod_{k=1}^{n} \frac{\Pr(s_{k,m} \le \Lambda(x_k) \le t_{k,m} \mid H_0)}{\Pr(s_{k,m} \le \Lambda(x_k) \le t_{k,m} \mid H_1)}$$

$$= \lambda \frac{\Pr(\mathcal{U}_m \mid H_0)}{\Pr(\mathcal{U}_m \mid H_1)}$$

where $\Lambda_m(x_i)$ is the local LLR given its past test outputs.

- Convergence is assured here too.

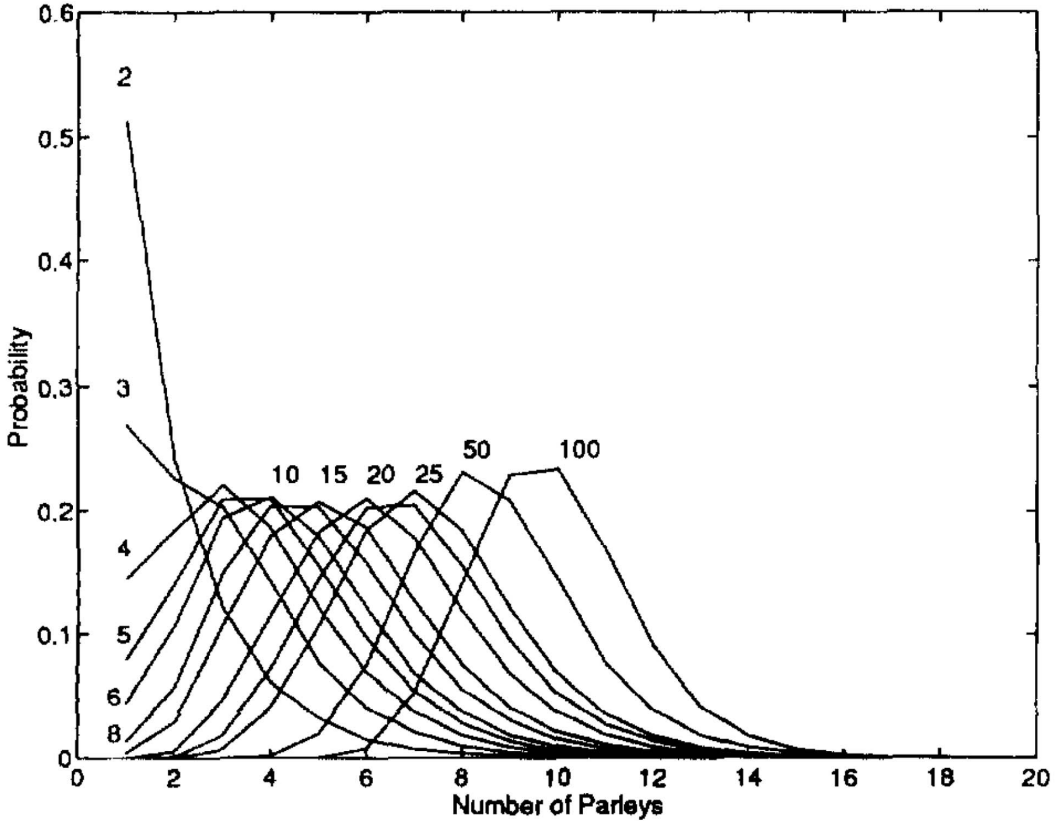
N^{th}-Root Example

	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6
$\Lambda(x_i)$	.99	.95	1.21	.86	.92	1.26
$\lambda_{i,1}$	1.00	1.00	1.00	1.00	1.00	1.00
$u_{i,1}$	0	0	1	0	0	1
$\lambda_{i,2}$	0.90	0.90	1.24	0.90	0.90	1.24
$u_{i,2}$	1	1	0	0	1	1
$\lambda_{i,3}$	0.95	0.95	1.10	0.79	0.95	1.37
$u_{i,3}$	1	1	1	1	0	0
$\lambda_{i,4}$	0.96	0.96	1.14	0.83	0.91	1.27
$u_{i,4}$	1	0	1	1	1	0
$\lambda_{i,5}$	0.96	0.93	0.84	1.15	1.08	0.80
$u_{i,5}$	1	1	1	1	1	1

→ product is 1.13

N^{th}-Root Results

- Since the decision is optimal, there is no need to check performance, just how long it takes.
- Here we plot that for a small-signal Gaussian problem versus number of DMs.



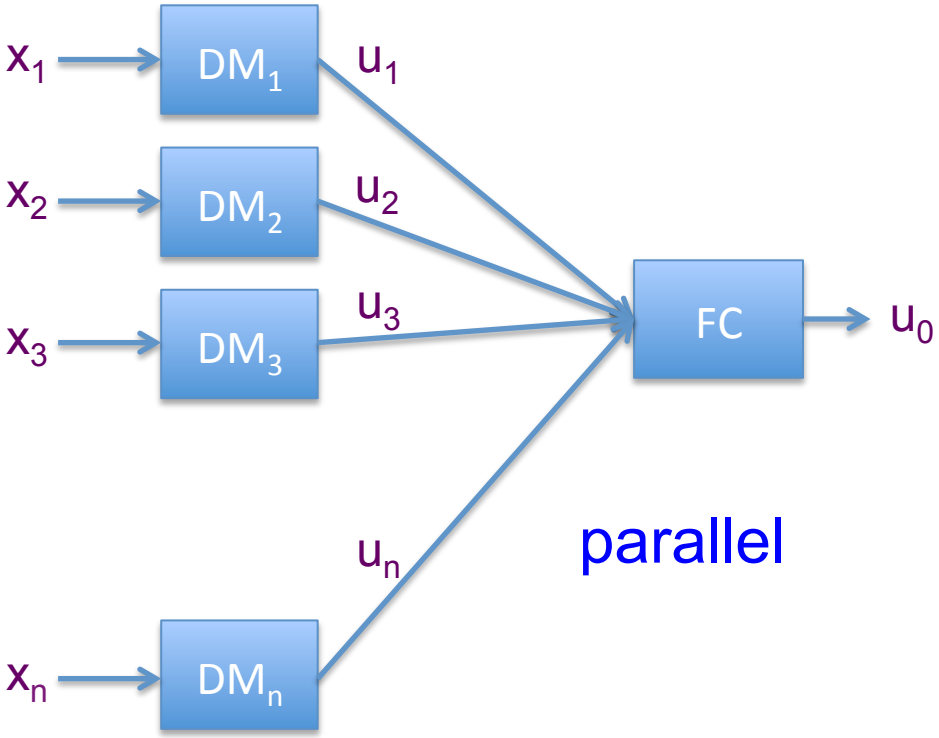
Censoring With Feedback

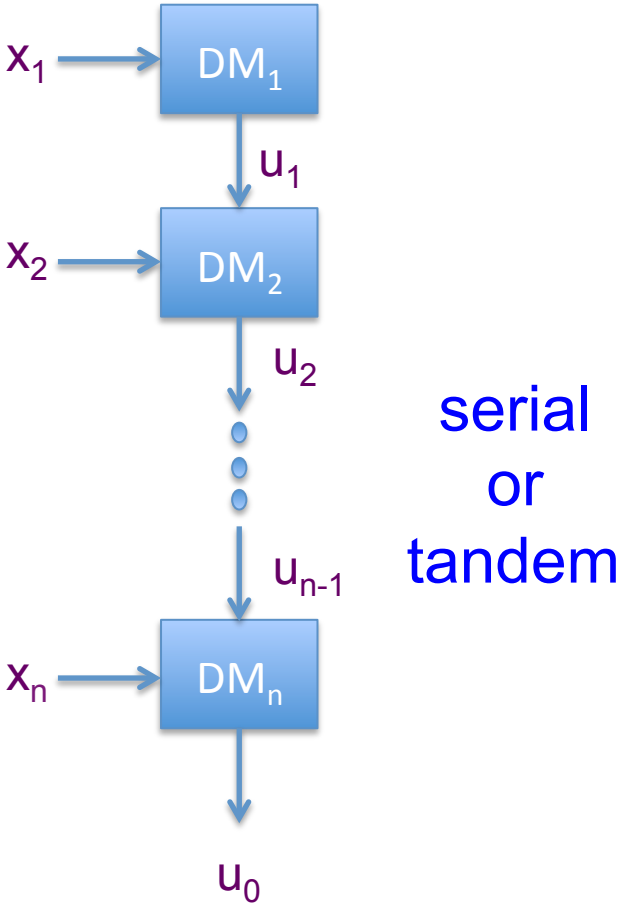
- Suppose the scheme is augmented to the following logical (but suboptimal) scheme:

decide for H	$L_{FC} \leq \tau_{FC}$, no sensor has transmitted
request further information	at least one sensor has not transmitted
decide for K	$L_{FC} > \tau_{FC}$, all sensors have transmitted

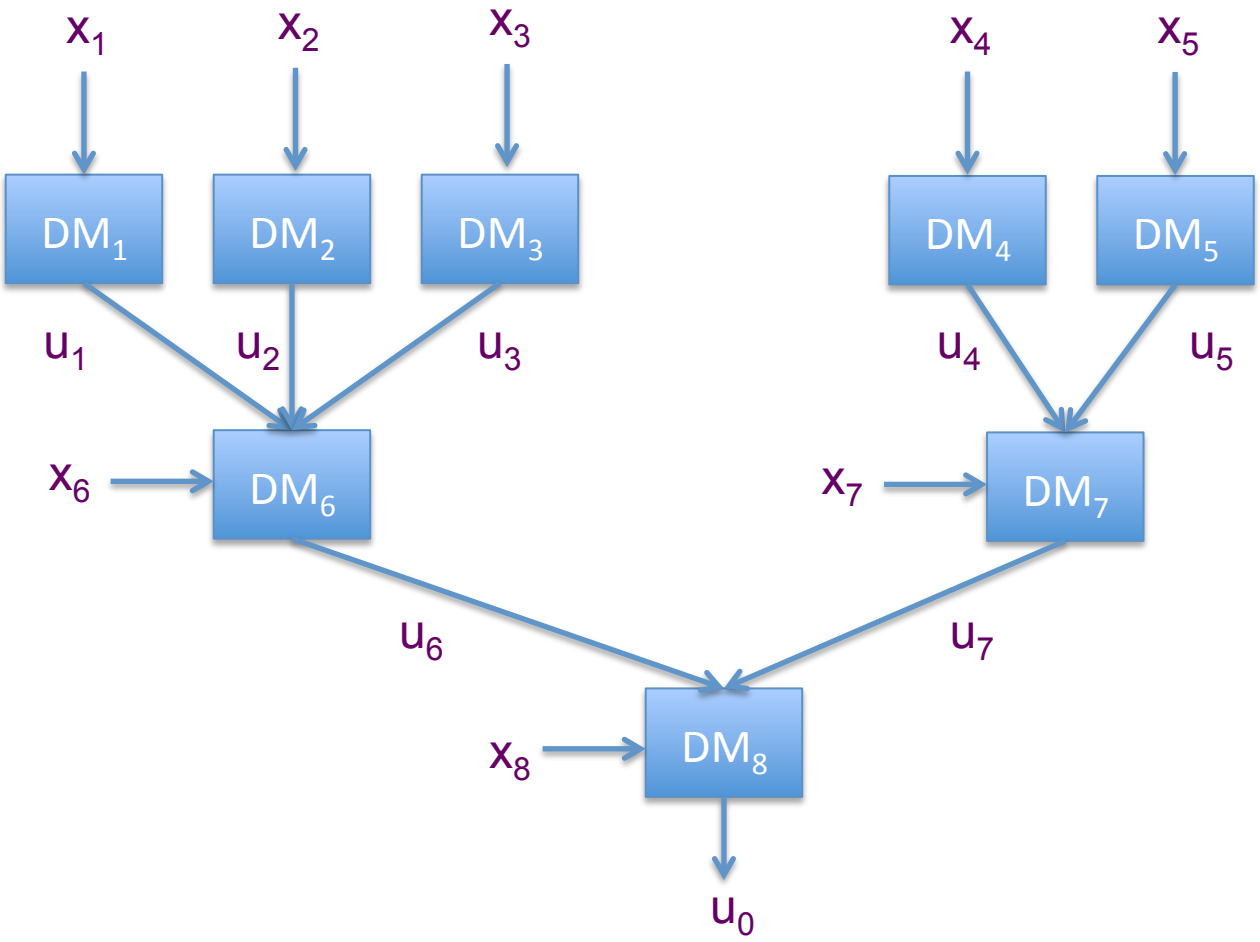
- The request for further information is that all silent sensors transmit.
- We use the "parley" system, where t_2 is the threshold below which a sensor is silent.

Distributed Detection Structures





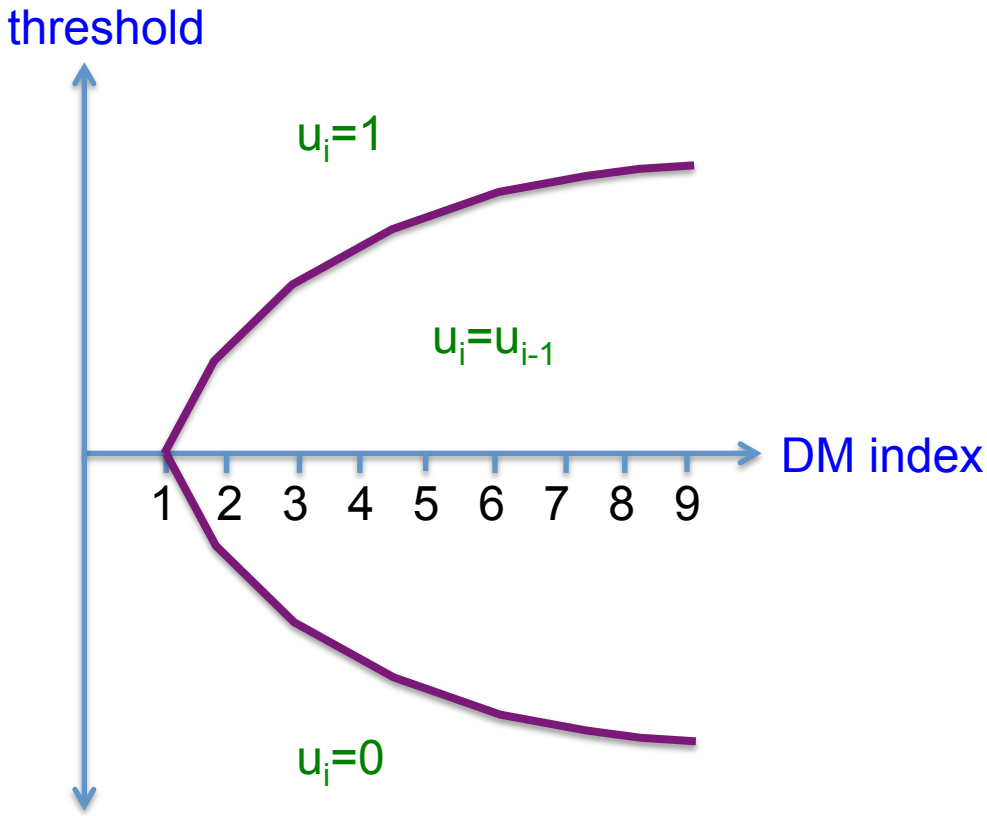
More General Structures



Little is known about the best such structure – or even what "best" might mean.

A human decision-making concept is of "congruence" between task and structure.

Tandem Structure



There is an optimal Bayesian solution:

$$\lambda_i = \log\left(\frac{1-P_e(i-1)}{P_e(i-1)}\right)$$

(symmetry) and provided the LLR is tightly bounded to ±U

$$\lim_{n\to\infty} P_e(n) = \frac{1}{1+U}$$

Interestingly, in the Gaussian case $P_e(n)$ does not go to zero unless

$$\lambda_i \propto \sqrt{\log(i)}$$

- Cover, "Hypothesis Testing with Finite Statistics" Ann Math Stat 1969.
- Swaszek, "On the Performance of Serial Networks in Distributed Detection," TAES 1993.

Decision Networks That Learn

- Consider the problem that the FC (parallel topology) needs to make a decision based on DMs of unknown quality.
- That is, $L(u) = \prod_{i=1}^{n} \frac{[1-P_d(i)]^{1-u_i} P_d(i)^{u_i}}{[1-P_{fa}(i)]^{1-u_i} P_{fa}(i)^{u_i}}$ but $P_d(i)$ and $P_{fa}(i)$ are unknown.
- Now suppose that the FC has access to previous decisions … but does not know ground truth in any of them.
- Assume DMs are conditionally independent.

- Wang, Kaplan, Abdelhazer & Aggarwal, "On Credibility Estimation Tradeoffs in Assured Social Sensing," JSTSP 2013.
- Marano, Matta & Willett, "The Importance of Being Earnest: Social Sensing With Unknown Agent Quality," TSP 2016.

i	Truth	DM 1	DM 2	DM 3	DM 4	DM 5	DM 6
1	H_0	1	0	1	0	0	0
2	H_0	1	0	1	1	0	0
3	H_1	1	1	0	0	1	1
4	H_0	1	0	1	1	0	0
5	H_1	1	1	0	0	1	0
6	H_1	0	1	0	1	0	0
7	H_1	1	1	0	0	1	0
8	H_0	1	0	1	1	0	0
9	H_1	1	1	0	0	1	0
10	H_0	1	0	0	0	0	0

- DM 1 has high P_d and high P_{fa}
- DM 2 has high P_d and low P_{fa}
- DM 3 has low P_d and high P_{fa}
- DM 4 has P_d and P_{fa} near 50%
- DM 5 has high P_d and low P_{fa}
- DM 6 has low P_d and low P_{fa}

How do we find this out?

EM Approach

- The EM approach is a "meta-algorithm" appropriate when there is hidden data that is not desired.
  - Here the hidden data is the truth (hypotheses) at all times.
- Initialize the P_d's and P_{fa}'s – say 80% & 10%.

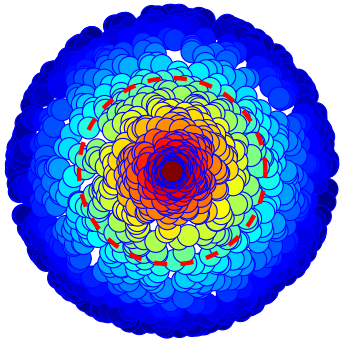
$$\begin{aligned} w_j(t) &= Pr(H_j \text{ true at frame } t | \{U_i(t)\}_{i=1}^n) \\ &= \frac{Pr(H_j) \prod_{i=1}^n \left([1 - P_d(i)]^{1-u_i(t)} P_d(i)^{u_i(t)} \right)}{Pr(H_0) \prod_{i=1}^n \left([1 - P_{fa}(i)]^{1-u_i(t)} P_{fa}(i)^{u_i(t)} \right) + Pr(H_1) \prod_{i=1}^n \left([1 - P_d(i)]^{1-u_i(t)} P_d(i)^{u_i(t)} \right)} \end{aligned}$$

$$\begin{aligned} P_d(i) &\leftarrow \frac{\sum_{t=1}^n w_1(t) u_i(t)}{\sum_{t=1}^n w_1(t)} \\ P_{fa}(i) &\leftarrow \frac{\sum_{t=1}^n w_0(t) u_i(t)}{\sum_{t=1}^n w_0(t)} \\ Pr(H_j) &\leftarrow \frac{\sum_{t=1}^n w_j(t)}{n} \end{aligned}$$

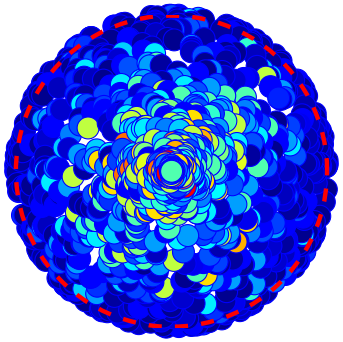
iterate back and forth

How Well Does It Work?

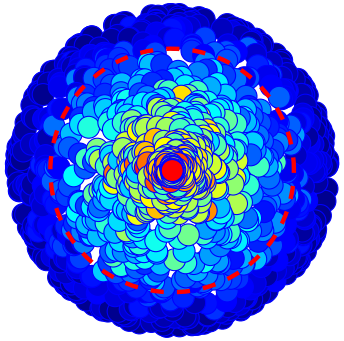
Ground truth



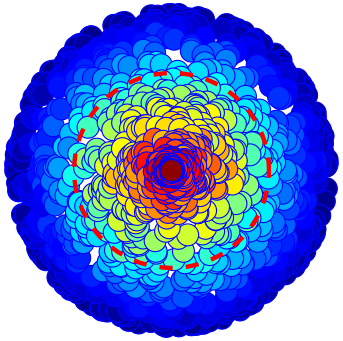
No. of tasks = 30



No. of tasks = 300



No. of tasks = 3000



- Network of 1000 DMs.
- Hot (red) circles denote a high error probability, while cold (blue) circles denote low error probabilities.
- Red dashed circle surrounds low-quality DMs.
- We see how the learning ability of our algorithm progressively increases, so we “unmask” the unreliable DMs.

Byzantine Sensors

- A Byzantine sensor is one that messes with the FC
  - it's tempting (e.g.) for a Byzantine sensor that "knows" H_0 is true to send message that says it knows H_1 is true
  - but this is too obvious, the FC can disregard such messages
  - So must try to send a disguised message
- Assume a fraction α of the sensors are Byzantine
  - report on asymptotic results
  - assume (here) that the Byzantines collude and **know H**
  - non-Byzantine sensors have probabilities of sending message u=k: $q_0(k)$ and $q_1(k)$ respectively under H_0 and H_1
  - Byzantine sensors have corresponding probabilities of sending message u=k: $\theta_0(k)$ and $\theta_1(k)$

- Matta, Marano & Tong, "Distributed Detection in the Presence of Byzantine Attacks," TSP 2009.

- Critical "blinding" fraction of Byzantine sensors

$$\alpha_b = \frac{\sum_k [q_0(k) - q_1(k)]^+}{1 + \sum_k [q_0(k) - q_1(k)]^+} \leq \frac{1}{2}$$

  - if $\alpha > \alpha_b$ then the FC is completely blind
- Otherwise with $\alpha < \alpha_b$

$$\theta_0(k) = \frac{1-\alpha}{\alpha} \sum_k [\gamma_0 q_0(k) - q_1(k)]^+$$

$$\theta_1(k) = \frac{1-\alpha}{\alpha} \sum_k [\gamma_1 q_1(k) - q_0(k)]^+$$

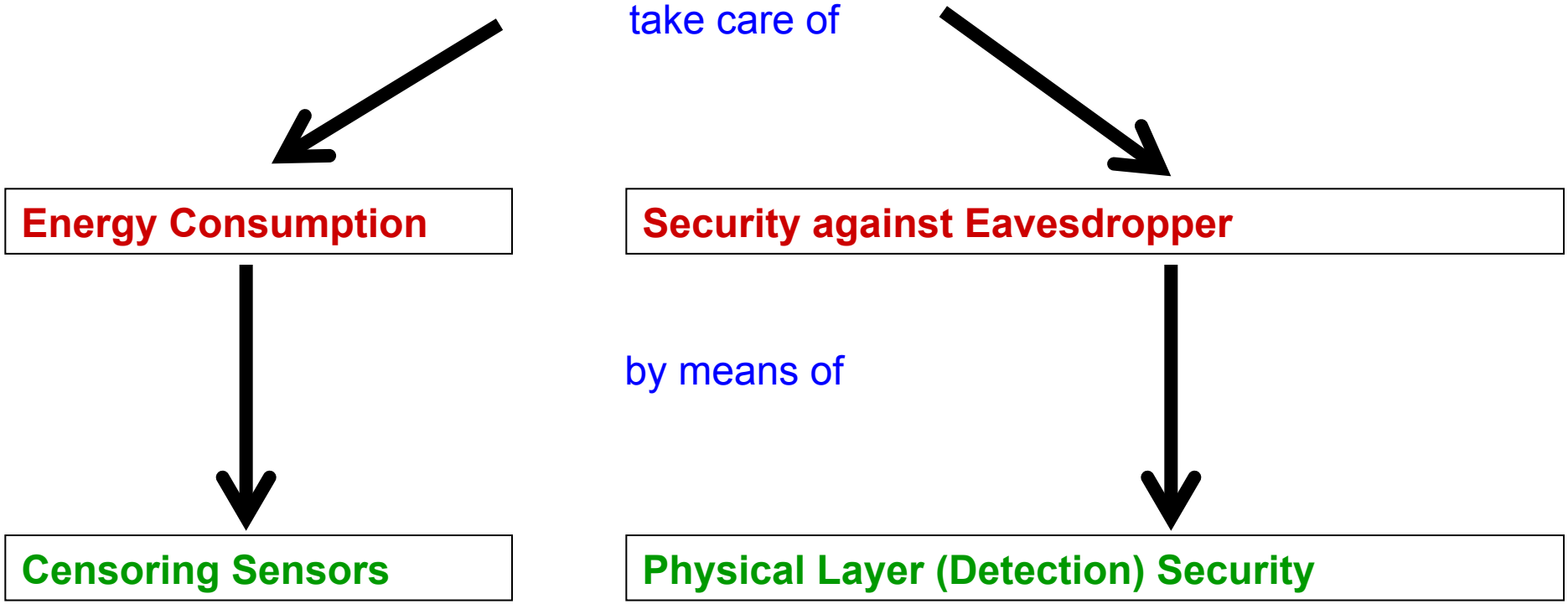
$$\gamma_0 : \quad \sum_k \theta_0(k) = 1$$

$$\gamma_1 : \quad \sum_k \theta_1(k) = 1$$

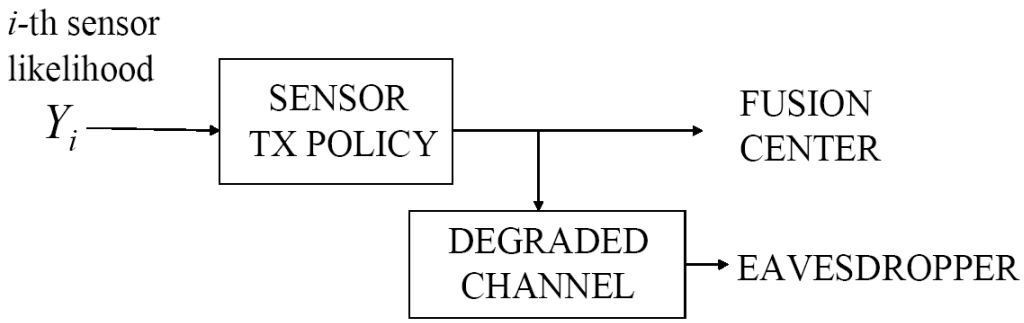
- Other considerations
  - Byzantine sensors do not know the true hypothesis
  - multiple observations from each sensor

Distributed Detection with Secrecy

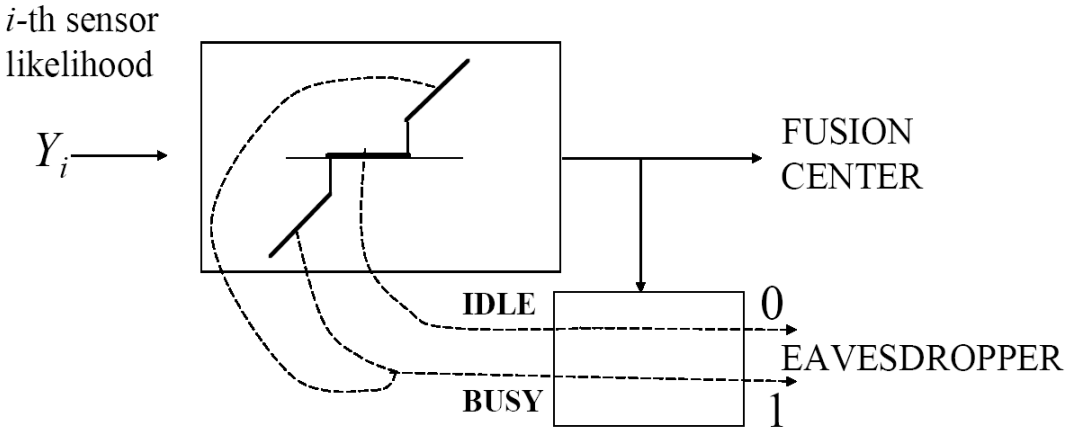
Since detection is possible for an eavesdropper as well as FC,
Then network protection is necessary



General setup



Our setup: the eavesdropper only knows that **something** was sent



- Marano, Matta & Willett, "Distributed Detection with Censoring Sensors under Physical Layer Secrecy," TSP 2013.

divergence @ fusion center (from i-th sensor)

$$D_i = \int_{\mathcal{R}_i} p_0(y) C(y) dy + (1 - \beta_0(i)) C\left(\frac{1 - \beta_1(i)}{1 - \beta_0(i)}\right)$$

overall (additive) divergence

$$\sum_{i=1}^{n} D_i$$

some convex U function, e.g., the two KL numbers

$$C_{10}(y) = y \log y$$

$$C_{01}(y) = \log \frac{1}{y}$$

overall divergence @ eavesdropper

$$D_{eav} = \sum_{i=1}^{n} \beta_0(i) C\left(\frac{\beta_1(i)}{\beta_0(i)}\right) + (1 - \beta_0(i)) C\left(\frac{1 - \beta_1(i)}{1 - \beta_0(i)}\right)$$

Perfect Secrecy

General formulation:

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{n} D_i \\ \text{subject to} & \max\{\beta_1(i), \beta_0(i)\} \le \beta \quad \forall i \\ & D_{eav} \le \Delta. \end{array}$$

Limit to PERFECT SECRECY

$$\Delta = 0 \iff \beta_0(i) = \beta_1(i)$$

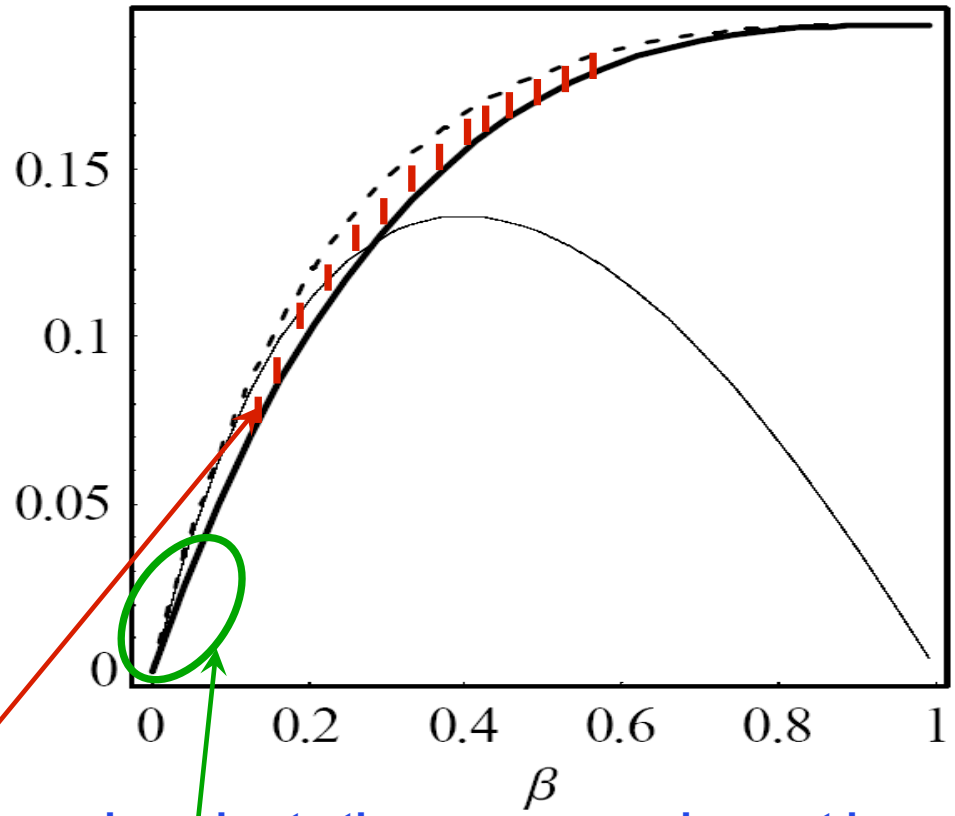
yielding:

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{n} D_i \\ \text{subject to} & \beta_1(i) = \beta_0(i) \le \beta \quad \forall i \end{array}$$

divergence-cost function

or, equivalently, we must compute

$$D(\beta) \stackrel{def}{=} \max_{\beta_1 = \beta_0 \le \beta} \int_{\mathcal{R}} p_0(y) C(y) dy$$



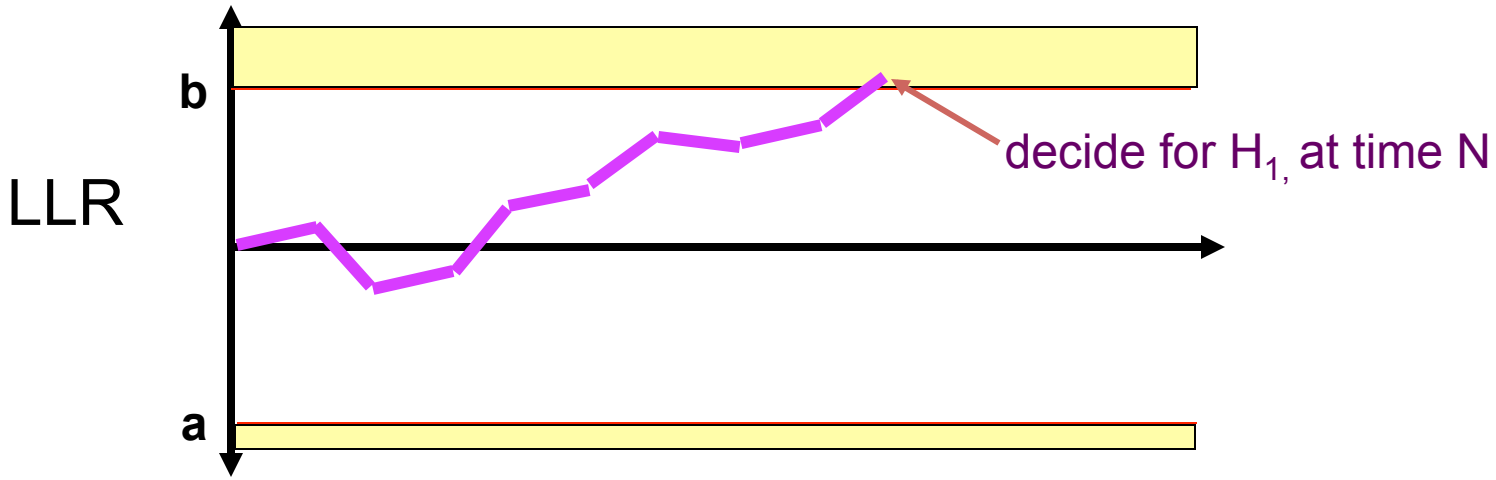
- The divergence loss due to the secrecy requirement is modest.
- With low comms ($\beta \to 0$), the FC and the eavesdropper see the same picture.
- Looking at the transmission activities, rather than the content, is nearly optimum.

More Secrecy

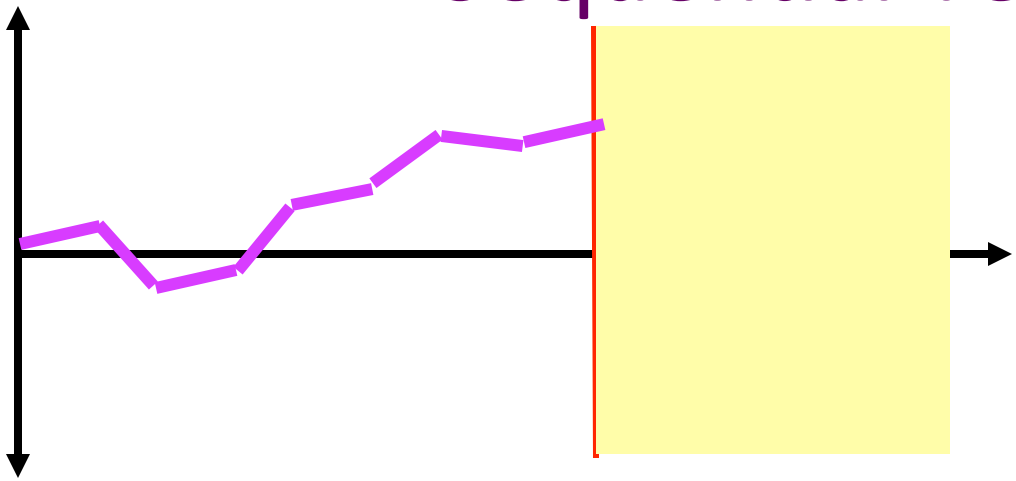
- Encryption
  - eavesdropper cannot decode
- Information theoretic analyses
  - bandwidth / information / secrecy trade-offs
- Homomorphic techniques
  - make a cooperative colleague do work for you while hiding your data from her

Sequential Testing

- suppose simple binary hypothesis test
  - H_0 versus H_1
  - $T_n = \log(f_1(x^n)/f_0(x^n))$ is LLR
  - stopping rule and decision controlled by upper and lower thresholds

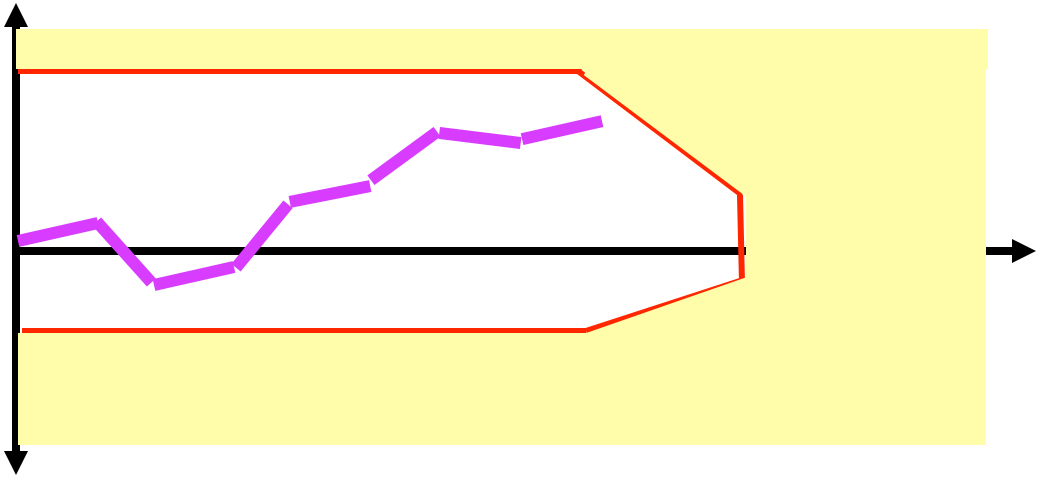


Sequential Testing



fixed sample size (FSS) test

more general test



- define $Q_n=\{x^n$: decide for H_1 at time $N=n\}$
- define $R_n=\{x^n$: decide for H_0 at time $N=n\}$

$$\alpha = \sum_n \int_{Q_n} f_0(x^n) \cong \frac{1}{B}\sum_n \int_{Q_n} f_1(x^n) \cong \frac{\beta}{B}$$

$$1-\beta = \sum_n \int_{R_n} f_1(x^n) \cong A\sum_n \int_{R_n} f_0(x^n) \cong A(1-\alpha)$$

- Wald's approximations for (LR, not LLR) thresholds A & B:
  - $e^b=B=\beta/\alpha$ and $e^a=A=(1-\beta)/(1-\alpha)$
- easier than fixed sample-size (FSS) testing
- Wald's identity can find average sample numbers (ASNs) for iid case
  - $E(N|H_j)=E(T_N|H_j)/\mu_j$ where μ_j is mean of update
  - also easy to figure out
  - Martingale proof, or using moment generating function

Sequential Testing: Analysis

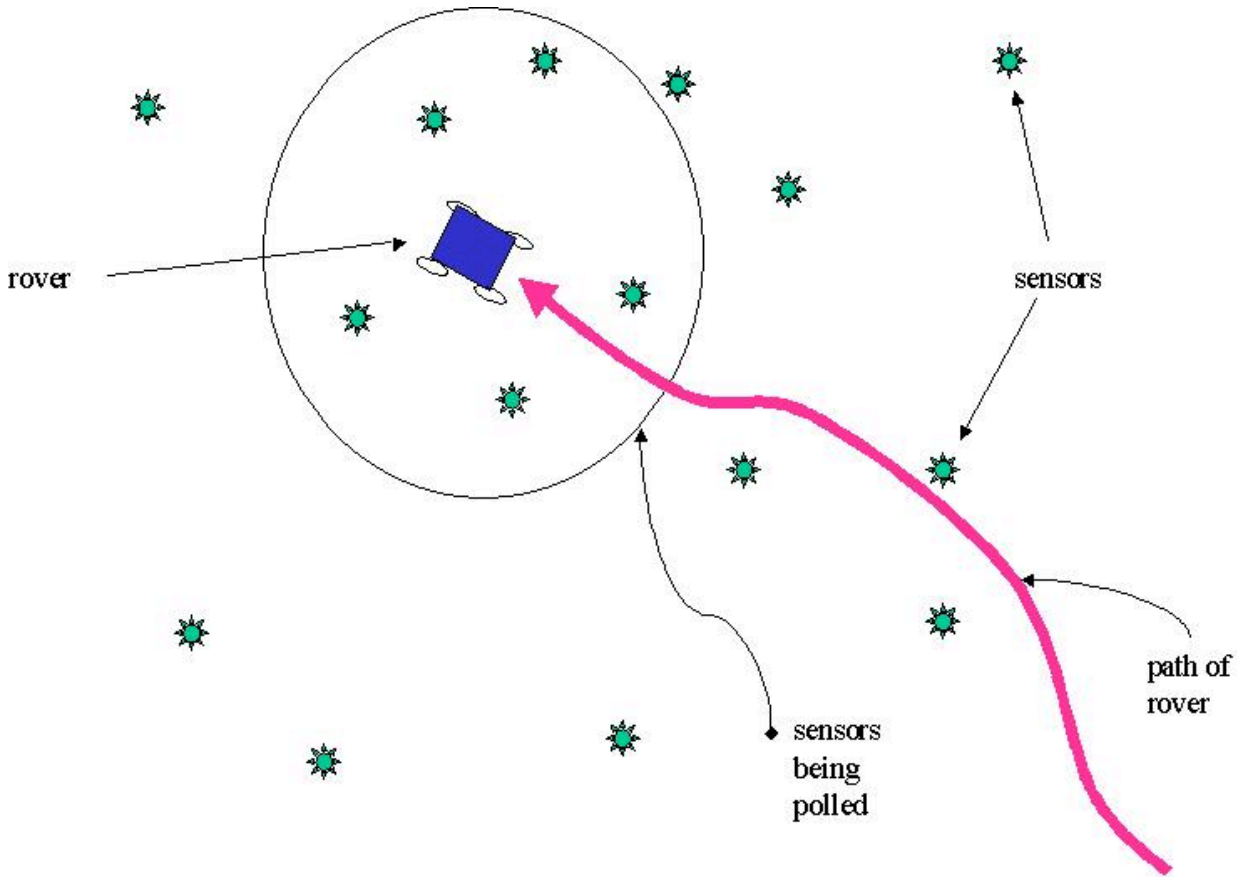
- Wald-Wolfowitz Theorem:
  - any test with better than a given α and β performance must have $E(N|H_j)$ performance no better than the corresponding SPRT
  - no point in trying to play games with decision regions or boundaries
- optimality is slippery
  - FSS test
    - has best error performance & bounded N
    - but worse ASN than SPRT
  - truncated SPRT (Tantaratana & Poor)
    - essentially same ASN as SPRT
    - better ASN than FSS test
- multiple hypotheses:
  - a lot is still open
  - Veeravalli, Tartakovsky

Distributed Sequential Detection

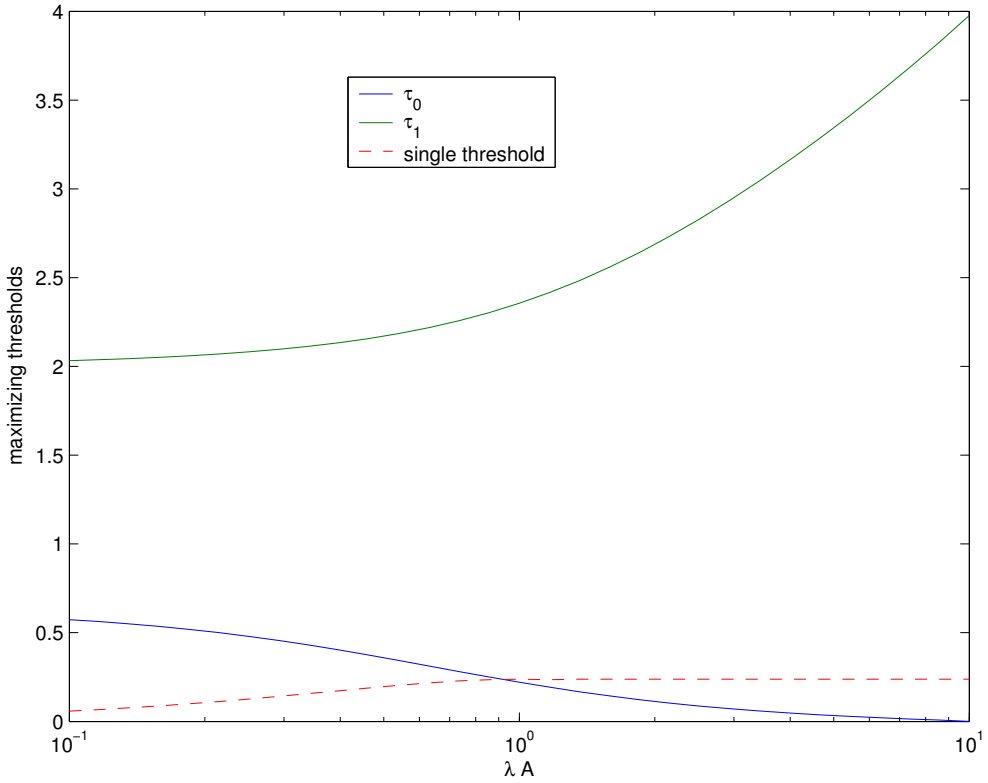
- Could simply have sequential test based on DM-level observations, but that is not so interesting.
- So consider a sequential testing paradigm in the SENMA architecture
  - sensor networks with mobile agents, a.k.a. rover, a.k.a. FC
  - in computer science this is a "data mule"

- Marano, Matta, Willett & Tong, "Cross-Layer Design of Sequential Detectors in Sensor Networks," TSP 2006.

- Poisson field of sensors
- sensors do not know their neighborhood
- FC hears DMs via ALOHA
  - need exactly one transmission for success
  - pure ALOHA efficiency is upper bounded by 18%



- SENMA works best with "censored" observations
- optimization shows Pr(no transmit)=0
  - all "don't transmit" decisions should be based on local threshold
- λ = Poisson density of sensors
- A is area per unit time that rover sees

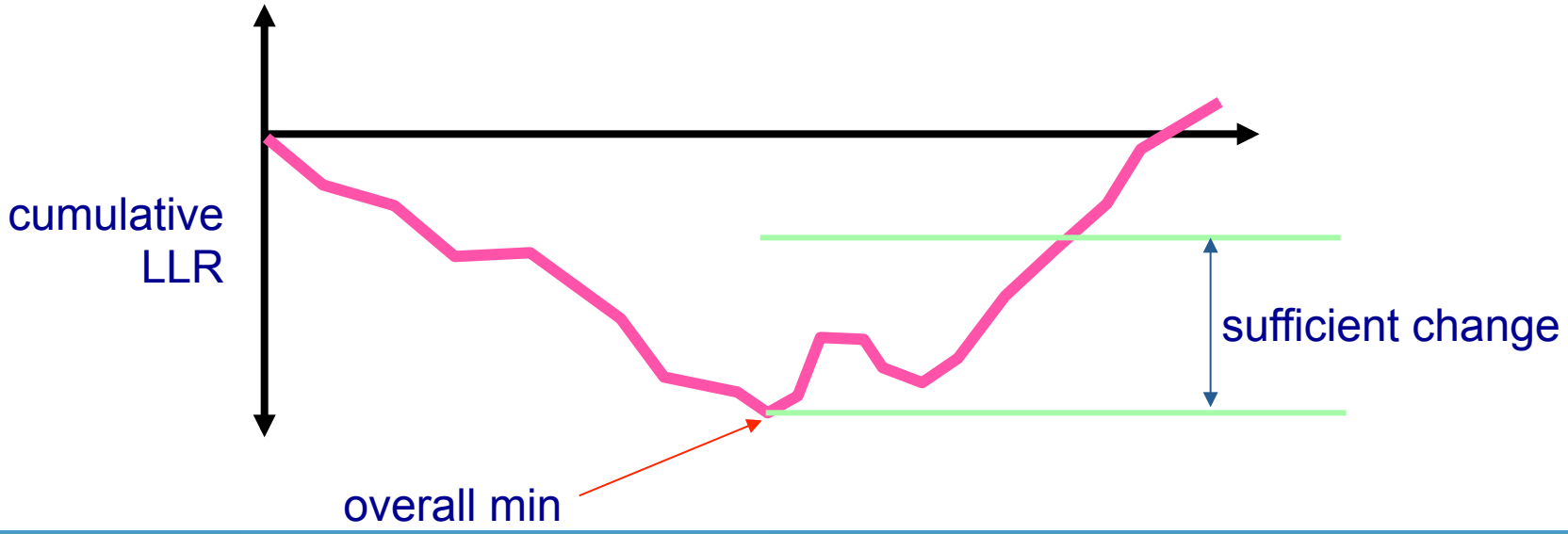


Exponential observation case

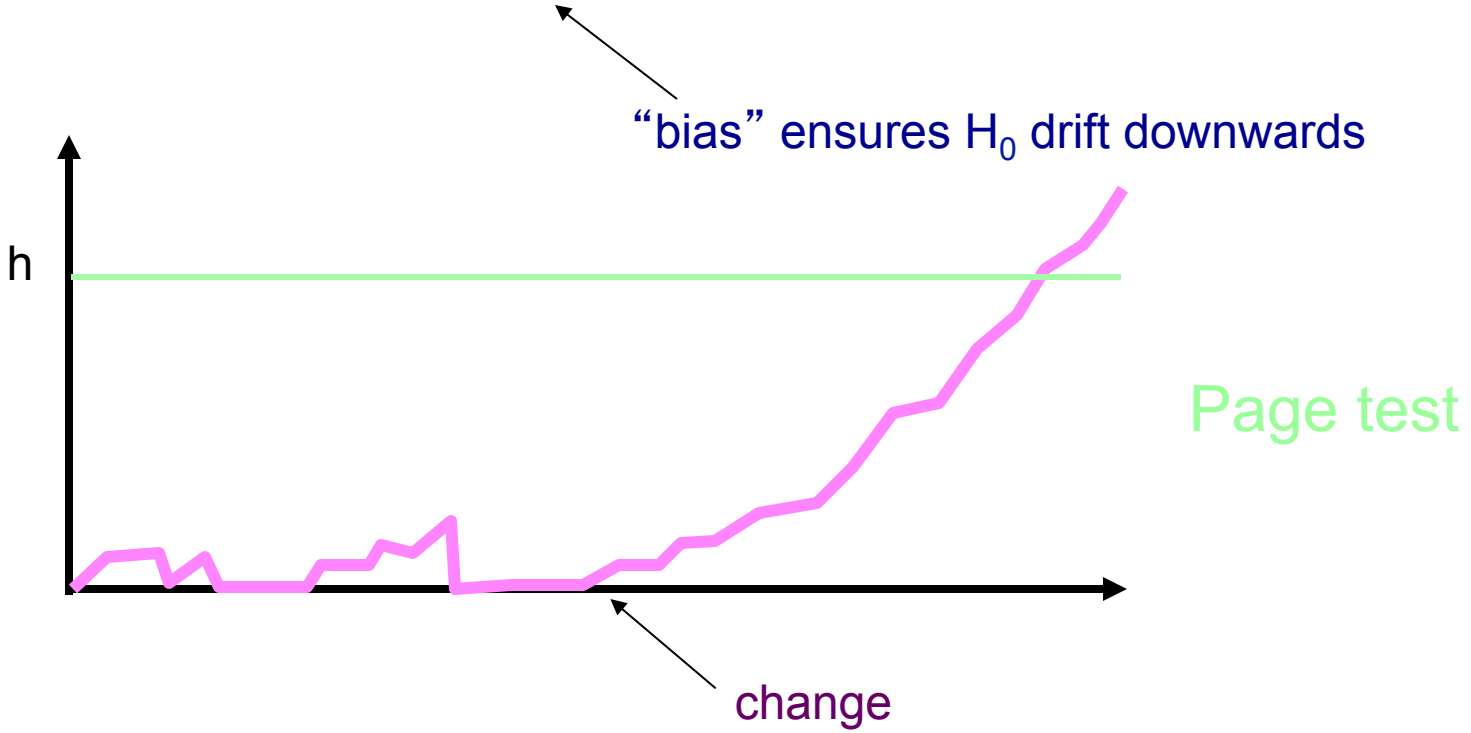
- Now suppose both FC and DMs use sequential tests
  - SPRT is a kind of "censoring"
  - results here for Gaussian shift-in-mean
- $P_e(FC) << P_e(DM)$
  - the DM tests are much less reliable than the FC decision
- $E(T_{FC}) \approx E(T_{DM})$
  - the DMs do not work (much) beyond the final decision
- $E(T_{FC}) \approx E(T_{\text{single sensor}}) \times (SNR/\lambda)^{1/2}$
  - an equivalent single sensor test would take far longer
- $E(N) \approx 0.3 \times \lambda \times E(T)$
  - number of FC reports received is about a third of the nodes encountered, regardless of P_e or SNR

Quickest Detection

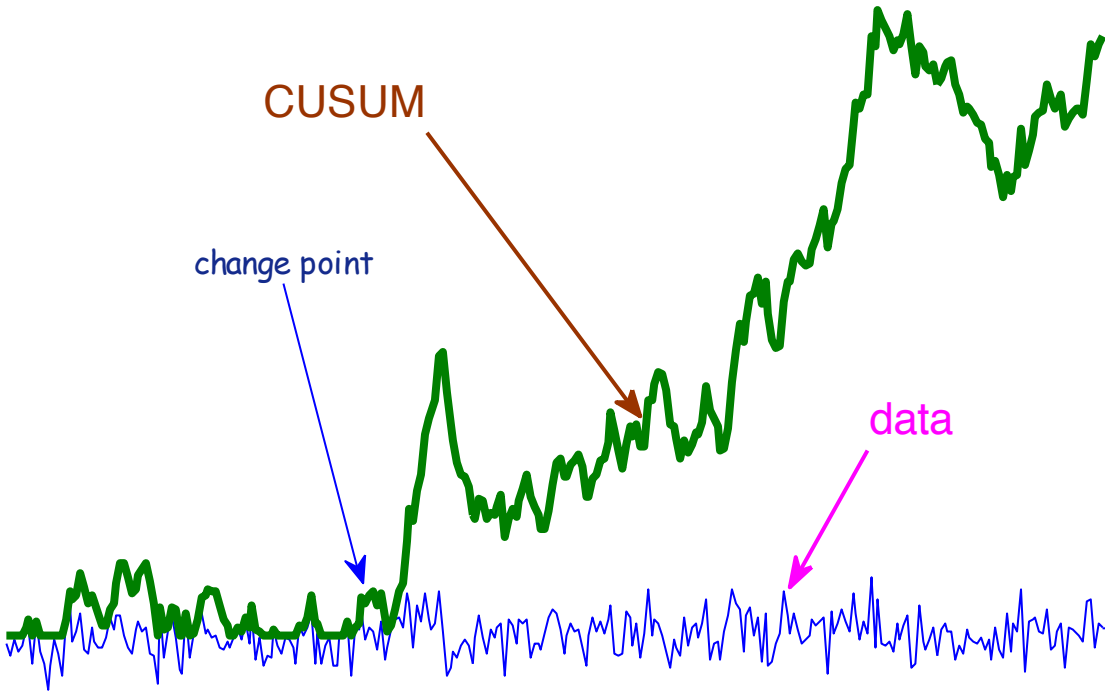
- nature changes from H_0 to H_1
  - want quickest alert of this change
  - assume independent data
- intuition:
  - form cumulative LLR
  - if it jumps enough, declare a change



- same as CUSUM (cumulative sum)
  - $T_n = \max\{0, T_{n-1} + g(x_n)\}$
  - iterated SPRT
  - unit Gaussian shift-in-mean:
    - $T_n = \max\{0, T_{n-1} + (x_n - \mu/2)\}$



- unit Gaussian with mean +/- 0.2



- asymptotically (as *h* gets large) have $T=e^{\eta D}$
  - η is "Page's efficiency"
  - suppose T is 1,000 and D is 10
  - then can have T=1,000,000 with D=20
- otherwise, for performance:
  - Siegmund's correction terms
  - Brownian motion or discrete random walk results
  - "C-matrix" or iterated FFT
- optimality:
  - Page procedure is "quickest" for iid case
    - Lorden, Moustakides
  - for non-iid case it is known to be quickest for only very special cases where updates are conditionally iid
    - some Markov chains (Moustakides)
    - some special HMMs (Fuh)
  - in some dependent cases Page is very suboptimal

Summary

- Detection basics
  - Neyman-Pearson, Bayesian, ROCs, useful alternative metrics
- Distributed detection and decision fusion
  - How to quantize and how to fuse
- Some fun pathologies
  - Identical sensors can be different
  - Dependence: little is known except it's strange
- New structures
  - Censoring sensors: reduced communication
  - Feedback of decisions: "How sure are you?"
  - Sequential networks: tricky to avoid lock-up
  - Learning decision-makers' biases: rich area with human decision-making applications and many extensions
  - Using censoring to hide data – secrecy
  - How Byzantine sensors should work
  - Sequential decentralized detection schemes

Questions?