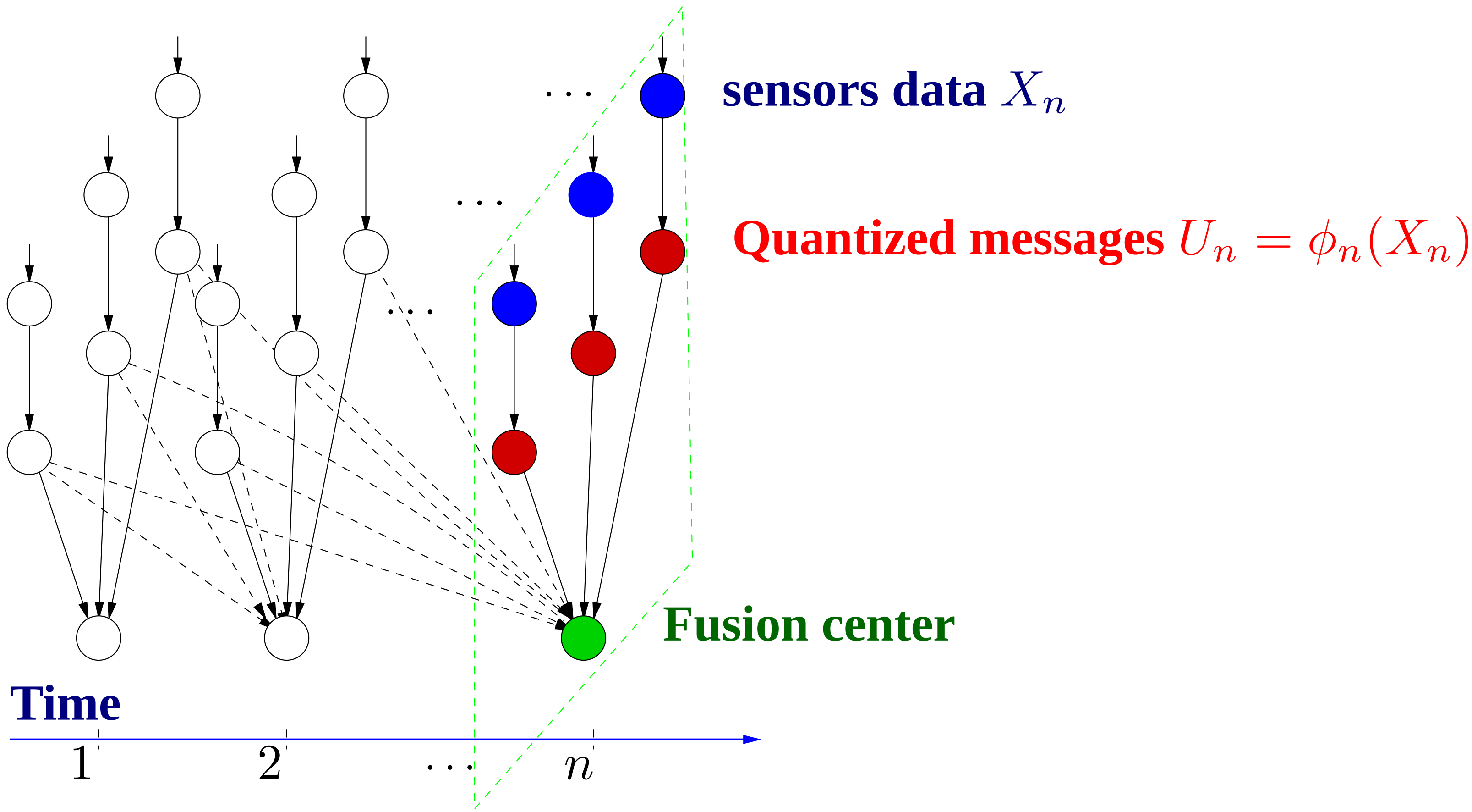
On optimal quantization rules in some sequential decision problems

XuanLong Nguyen, Martin J. Wainwright & Michael I. Jordan

Department of EECS and Department of Statistics
U.C. Berkeley

Sequential detection in a distributed system



sensors have no memory and no feedback

Problem statement

- **setting:** given sequence of (multivariate) sensor data $X_1, X_2, \ldots$, all of which are generated by either $\mathbb{P}_0$ or $\mathbb{P}_1$ (i.e., $H = 0$ or 1)
- prior π^0 and π^1
- cost c for each time delay (or each test sample)
- decision rules consists of
  - local quantization rules $Z_n = \phi_n(X_n)$
  - stopping time N wrt sigma field $\sigma(Z_1, \ldots, Z_N)$
  - global decision rule $\hat{Y} = \gamma(Z_1, \ldots, Z_N)$
- **problem**: find decision rules (ϕ, N, γ) that minimize (Bayesian cost):

$$J(\phi, N, \gamma) := \mathbb{E}\{cN + \mathbb{I}[\gamma(Z_1, \ldots, Z_N) \neq H]\},$$

Related work

- Sequential hypothesis testing

  (Wald, Wald & Wolfowitz, Arrow, Blackwell & Girshick, etc)

  (cf. Siegmund (1985), Shiryaev (1978), Lai (2001))

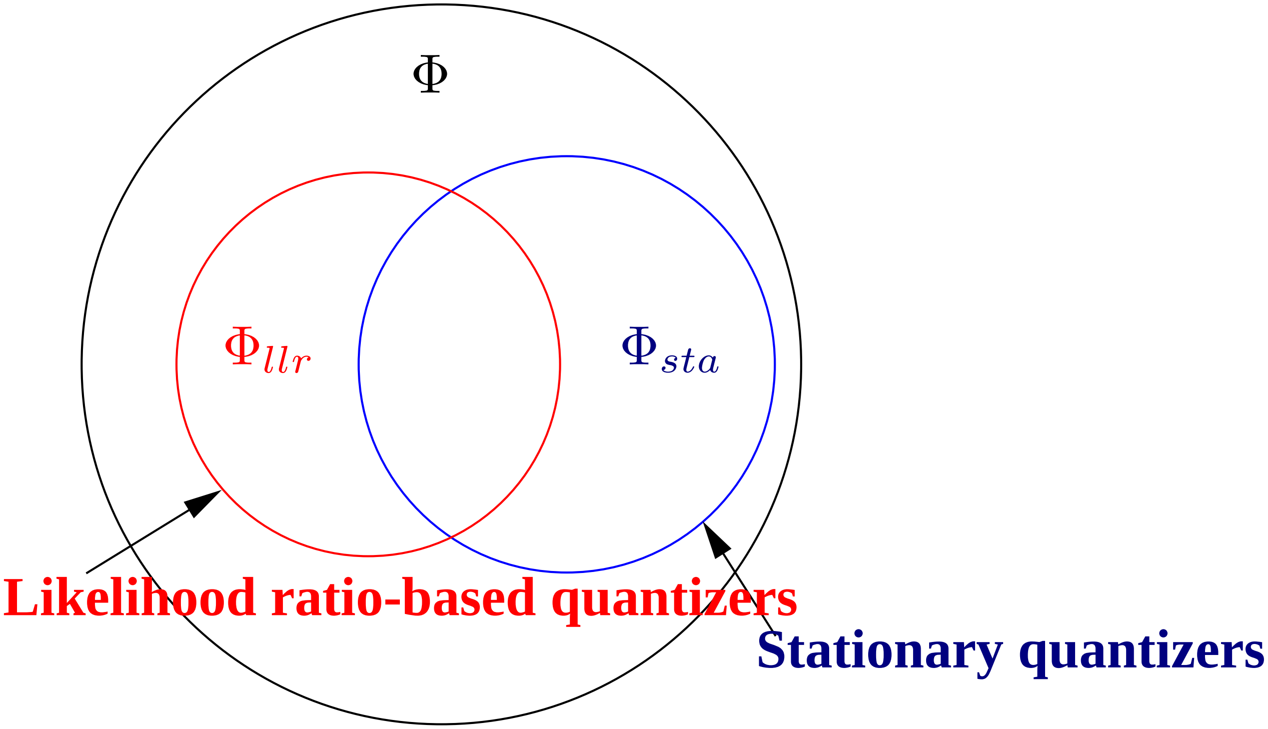
- Decentralized detection

  (cf. Tenney & Sandell (1981), Tsitsiklis (1993), Chamberland & Veeravalli (2004), Nguyen et al (2005), etc)

- Decentralized sequential detection

  (cf. Veeravalli et al (1993, 1998), Mei (2003, 2006))

Characterization of optimal quantizers



- ϕ^* denotes the optimal quantizer
- Tsitsiklis (1986) showed that $\phi^* \in \Phi_{llr}$
- Veeravalli et al (1993) **conjectured that** $\phi^* \in \Phi_{sta}$

Main results

- resolve Veeravalli's stationarity conjecture by providing counterexamples in both exact and asymptotic cases such that $\phi^* \notin \Phi_{sta}$
- show that when restricting quantizers to stationary class Φ_{sta}

$$\phi^* \in \Phi_{llr}$$

- simple characterization of optimal stationary quantizer provides an useful objective function for quantizer design algorithms

Sequential probability ratio test (SPRT) (background)

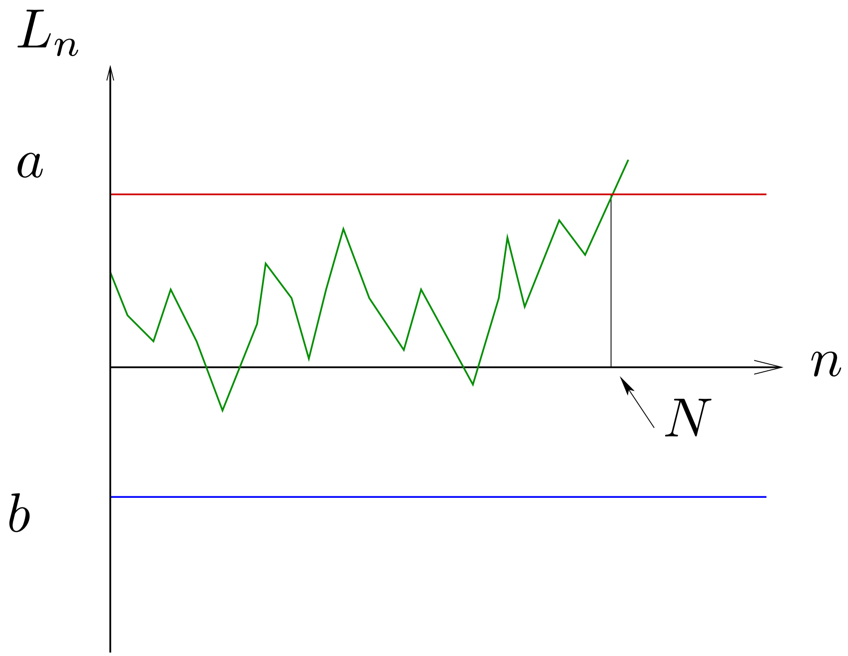
- **setting:** given sequence of (multivariate) sensor data $X_1, X_2, \ldots$, all of which are generated by either $\mathbb{P}_0$ or $\mathbb{P}_1$
- the optimal stopping rule is a sequential probability ratio test:

$$N = \inf \left\{ n \geq 1 \mid L_n := \sum_{i=1}^{n} \log \frac{f^1(X_i)}{f^0(X_i)} \notin (a, b) \right\},$$

  for some real numbers $a < b$.
- given this stopping rule, the optimal decision function has the form

$$\gamma(L_N) = \begin{cases} 1 & \text{if } L_N \geq b, \\ 0 & \text{if } L_N \leq a. \end{cases}$$

Wald's approximation



- cost of a sequential test by ignoring the **overshoot**:

$$G(\alpha,\beta) := c\pi^0 \frac{D(\alpha, 1-\beta)}{\mu^0} + c\pi^1 \frac{D(1-\beta, \alpha)}{\mu^1} + \pi^0\alpha + \pi^1\beta,$$

- optimal sequential cost

$$\inf_{a,b} J(a,b) \approx \inf_{\alpha,\beta} G(\alpha,\beta)$$

Distributions induced by quantizer rules ϕ

- quantizer ϕ_n yields a sequence of compressed data $U_n = \phi_n(X_n) \in \mathcal{U}$

- choice of ϕ induces distributions of compressed data (wrt $\mathbb{P}_0$ and $\mathbb{P}_1$):

$$f^i_{\phi_n}(u) \quad := \quad \mathbb{P}_i(\phi_n(X_n) = u), \qquad \text{for} \quad i = 0, 1$$

- induced KL divergences: $\mu^1_\phi := D(f^1_\phi || f^0_\phi)$ and $\mu^0_\phi := D(f^0_\phi || f^1_\phi)$.

A characterization lemma

- **key assumption:**

$$\sup_{\phi\in\Phi}\sup_{u\in\mathcal{U}}\log(f_\phi^1(u)/f_\phi^0(u)) \leq M$$

for some constant M over a class Φ of quantizers $\phi : \mathcal{X} \to \mathcal{U}$.

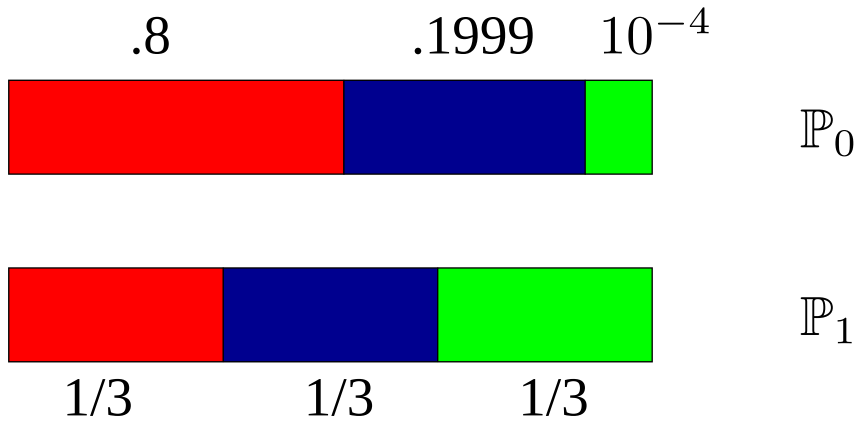
- approximation error by ignoring the overshoot

$$|J_\phi(a,b) - G_\phi(\alpha,\beta)| \leq c\,M\left(\frac{\pi^0}{\mu_\phi^0} + \frac{\pi^1}{\mu_\phi^1}\right).$$

- **optimal cost:** defined as $J_\phi^* = \inf_{a,b} J_\phi(a,b)$. Then as $c \to 0$, we have

$$J_\phi^* = \left(\frac{\pi^0}{\mu_\phi^0} + \frac{\pi^1}{\mu_\phi^1}\right) c\log\frac{1}{c} + O(c).$$

Suboptimality of stationary quantizer design



Setting: prior prob are $\pi^1 = \frac{8}{100}$ and $\pi^0 = \frac{92}{100}$, and sample cost $c = \frac{1}{100}$

Binary quantizers: there are only three possible *stationary* designs:

1. Design A: $\phi_A(X_n) = 0 \iff X_n = 1$.

2. Design B: $\phi_B(X_n) = 0 \iff X_n \in \{1, 2\}$.

3. Design C: $\phi_C(X_n) = 0 \iff X_n \in \{1, 3\}$.

Numerical example

- numerically computed costs for the three stationary designs J_A, J_B and J_C, and for the mixed design J_*
- J_* is obtained by applying A on the first sample, and B on the rest
- table shows $J_* < \min(J_A, J_B, J_C)$!

Method	$J_A(0.08)$	$J_B(0.08)$	$J_C(0.08)$	$J_*(0.08)$
Cost	0.0567	0.0532	0.0800	0.0528

Asymptotic suboptimality of stationary design

- recall the characterization lemma

$$J^*_\phi = \left(\frac{\pi^0}{\mu^0_\phi} + \frac{\pi^1}{\mu^1_\phi}\right) c \log \frac{1}{c} + O(c).$$

$\Longrightarrow$ Optimal stationary quantizer ϕ minimizes

$$\frac{\pi^0}{\mu^0_\phi} + \frac{\pi^1}{\mu^1_\phi}$$

- if we interleave design A and B, the induced KL divergences are

$$\mu^0_{\phi_{AB}} = \frac{1}{2}(\mu^0_{\phi_A} + \mu^0_{\phi_B})$$

$$\mu^1_{\phi_{AB}} = \frac{1}{2}(\mu^1_{\phi_A} + \mu^1_{\phi_B})$$

Proposition: Asymptotic suboptimality of stationary design

(based on asymmetry of KL divergences)

given

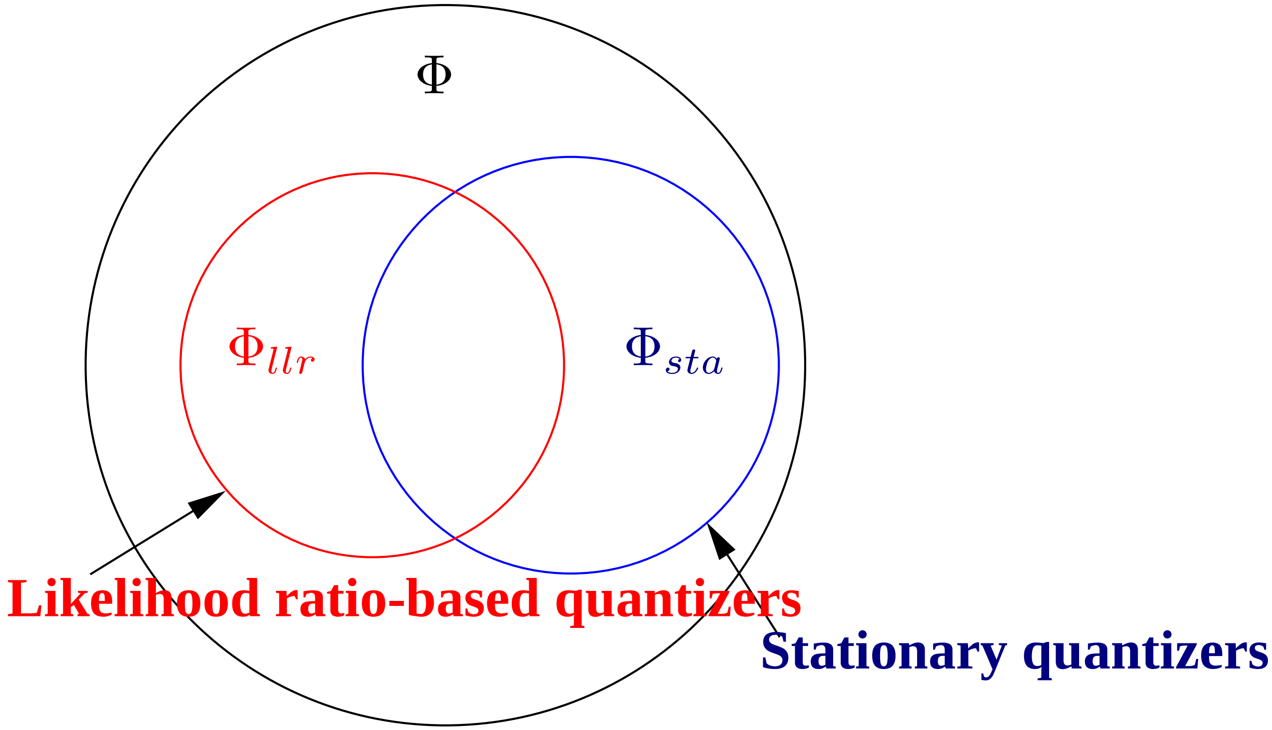
- $J^*_{\phi_i}$, the optimal cost of the stationary design based on ϕ_i
- $J^*_{\phi_A,\phi_B}$, the optimal cost of a sequential test that alternates between using ϕ_A and ϕ_B on odd and even samples respectively
- assume that

$$\mu^0_{\phi_A} < \mu^0_{\phi_B} \text{ and } \mu^1_{\phi_A} > \mu^1_{\phi_B}$$

then there exists a non-empty interval for π^0 such that as $c \longrightarrow 0$,

$$J^*_{\phi_A,\phi_B} < \min\{J^*_{\phi_A}, J^*_{\phi_B}\} - \Theta(c \log c^{-1})$$

Characterization of optimal quantizers



- ϕ^* denotes the optimal quantizer
- Tsitsiklis (1986) showed that $\phi^* \in \Phi_{llr}$
- we just showed that $\phi^* \notin \Phi_{sta}$
- what is ϕ^* when restricting to Φ_{sta}?

Quantizers based on thresholding likelihood ratio

- **Definition:** Quantizer $\phi : \mathcal{X} \rightarrow \mathcal{U}$ is said to be a *likelihood ratio threshold rule* if there are thresholds $d_0 = -\infty < d_1 < \ldots < d_K = +\infty$,
  and a permutation $(u_1, \ldots, u_K)$ of $(0, 1, \ldots, K-1)$
  such that for $l = 1, \ldots, K$, with $\mathbb{P}_0$-probability 1, we have:

$$\phi(X) = u_l \text{ if } d_{l-1} \leq f^1(X)/f^0(X) \leq d_l,$$

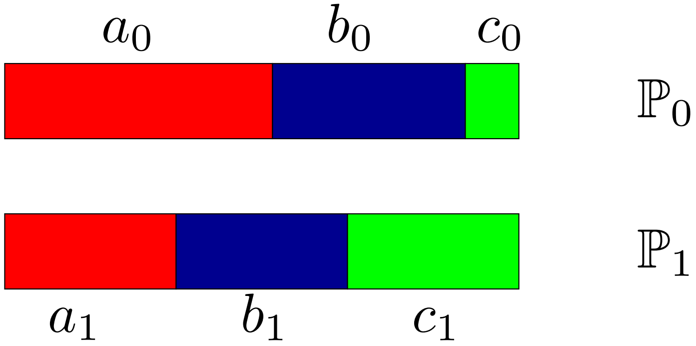
- **Theorem:** Restricting to the class of **stationary** and **deterministic** decision rules, then there exists an asymptotically optimal quantizer ϕ that is a likelihood ratio based threshold rule (LLR).

- **Proof ideas:** based on interplay between two KL divergences

Properties of KL divergences (1)

- given vectors $a = (a_0, a_1)$ and $b = (b_0, b_1)$, define functions $\tilde{D}^0$ and $\tilde{D}^1$

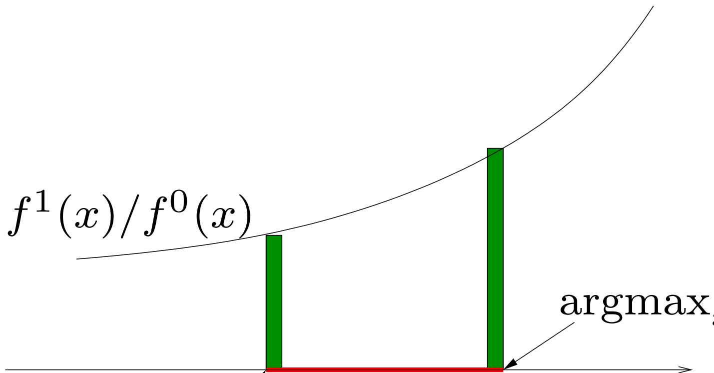
$$\begin{aligned}\tilde{D}^0(a,b) &:= a_0 \log \frac{a_0}{a_1} + b_0 \log \frac{b_0}{b_1} \\ \tilde{D}^1(a,b) &:= a_1 \log \frac{a_1}{a_0} + b_1 \log \frac{b_1}{b_0}.\end{aligned}$$



- for any positive scalars $a_1, b_1, c_1, a_0, b_0, c_0$ such that $\frac{a_1}{a_0} < \frac{b_1}{b_0} < \frac{c_1}{c_0}$, *at least one* of the two following conditions must hold:

$$\tilde{D}^0(a, b+c) > \tilde{D}^0(b, c+a) \quad \text{and} \quad \tilde{D}^1(a, b+c) > \tilde{D}^0(b, c+a), \text{ or}$$
$$\tilde{D}^0(c, a+b) > \tilde{D}^0(b, c+a) \quad \text{and} \quad \tilde{D}^1(c, a+b) > \tilde{D}^0(b, c+a).$$

Corollary



- if ϕ is an asymptotically optimal quantizer, then for any pairs of $(u_1, u_2) \in \mathcal{U}$, $u_1 \neq u_2$, there holds:

$$\frac{f^1(u_1)}{f^0(u_1)} \notin \left(\inf_{x:\phi(x)=u_2} \frac{f^1(x)}{f^0(x)}, \sup_{x:\phi(x)=u_2} \frac{f^1(x)}{f^0(x)} \right)$$

- i.e., ϕ behaves *almost* like a LLR based rule, but not quite the same.

- hence, need more work

Properties of KL divergences (2): Quasi-concavity

- KL divergence $D(a_0, a_1) := a_0 \log \frac{a_0}{a_1} + (1 - a_0) \log \frac{1-a_0}{1-a_1}$
- Let $F : [0, 1]^2 \to R$ be given by

$$F(a_0, a_1) = \frac{c_0}{D(a_0, a_1) + d_0} + \frac{c_1}{D(a_1, a_0) + d_1}.$$

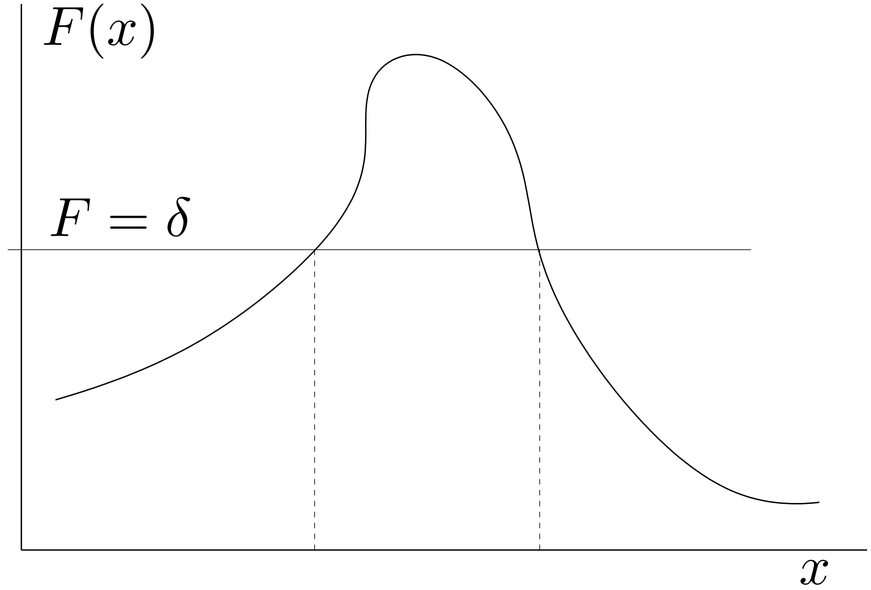
- For any non-negative constants c_0, c_1, d_0, d_1, function F is quasi-concave.

$\Longrightarrow$

Optimal cost function J_ϕ^* is quasi-concave w.r.t. *binary* quantizer ϕ

conjectured to be true for general quantizers

Useful properties of quasi-concavity



- for any δ, $F > \delta$ is a convex set
- minima of F is achieved at the extreme point of its (compact) domain
- extreme points of quantizer class Φ are generally LLR-based rules

$\Longrightarrow$

Restricted to *deterministic* and *stationary* class, there exists an optimal quantizer that is LLR-based.

Discussions

- resolved stationarity conjecture regarding the optimal quantizer designs
- proved the likelihood-ratio characterization of optimal quantizers when restricted to stationary classes
- simple, practical characterization of optimal stationary quantizer via asymptotic identity:

$$J_\phi^* = \left(\frac{\pi^0}{\mu_\phi^0} + \frac{\pi^1}{\mu_\phi^1}\right) c \log \frac{1}{c} + O(c).$$

- future work: nonparametric quantizer design for sequential detection