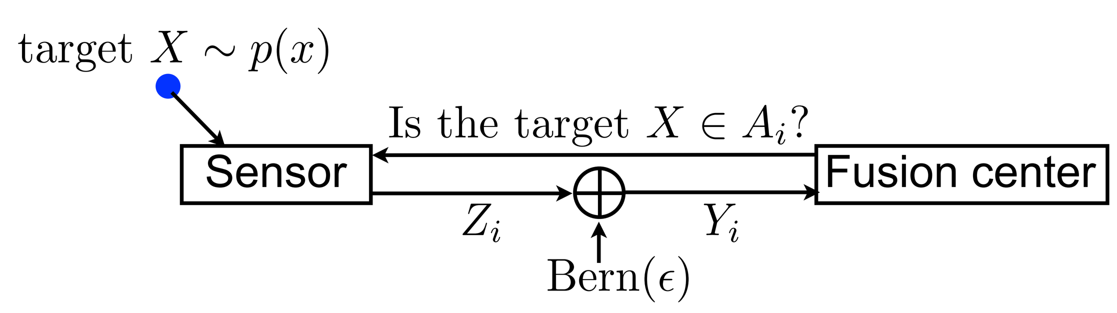
Information Surrogates for State Estimation

Hye Won Chung*, Brian M. Sadler† and Alfred O. Hero*
* University of Michigan, Ann Arbor, MI, USA
† US Army Research Laboratory, Adelphi, MD, USA

Introduction

Aim of the Project
1. Quantify the value of information for inference tasks such as state estimation
2. Find the right information surrogates for estimation errors

Problem Formulation
Noisy 20 questions game to estimate a target location $X \in \mathbb{R}$ in the setting of stochastic search
1. Single player setting:



Model: A fusion center asks queries regarding the target location X to a sensor and receives binary answers from it, which pass through BSC(ϵ)
Goal: To minimize $\mathbb{E}[|X - \hat{X}(Y_1^n)|^2]$
Aspects:

- The value of adaptivity in designing A_i's to reduce the mean squared error (MSE)
- Information surrogate: conditional entropy $h(X|Y_1^n) = \mathbb{E}_{Y_1^n}\left[-\int p(x|y_1^n)\log p(x|y_1^n)dx\right]$

2. Multiple player setting: multiple players with different reliabilities

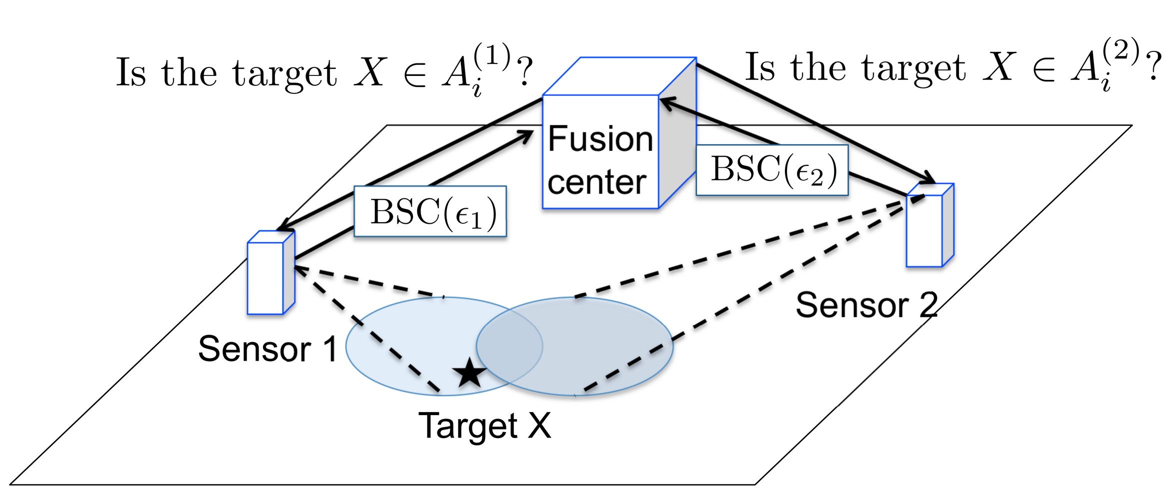


Image source: Tsiligkaridis [1]

Aspects:

- Joint query design vs. sequential query design
- Information surrogate: conditional entropy

Information Surrogate

Conditional Entropy

- Why conditional entropy?

$$\min_{\{A_1,\ldots,A_n\}} h(X|Y_1^n) \Leftrightarrow \max_{\{A_1,\ldots,A_n\}} I(X;Y_1^n)$$

- Necessary condition: estimation counterpart to the Fano's inequality

$$\mathbb{E}[|X - \hat{X}(Y_1^n)|^2] \geq \frac{1}{2\pi e}e^{2\cdot h(X|Y_1^n)}$$

$\Rightarrow$ A small entropy is necessary to achieve a small mean squared error

Two Questions
1. Under the entropy proxy, what is the value of adaptivity?
2. Is minimizing conditional entropy sufficient to reduce the mean squared error (MSE)?

Main Contributions

1. Under entropy proxy, both adaptive and non-adaptive polices can achieve the optimal performance
2. For unimodal distributions with exponential tails, MSE decreases exponentially with the conditional entropy

Single Player Problem

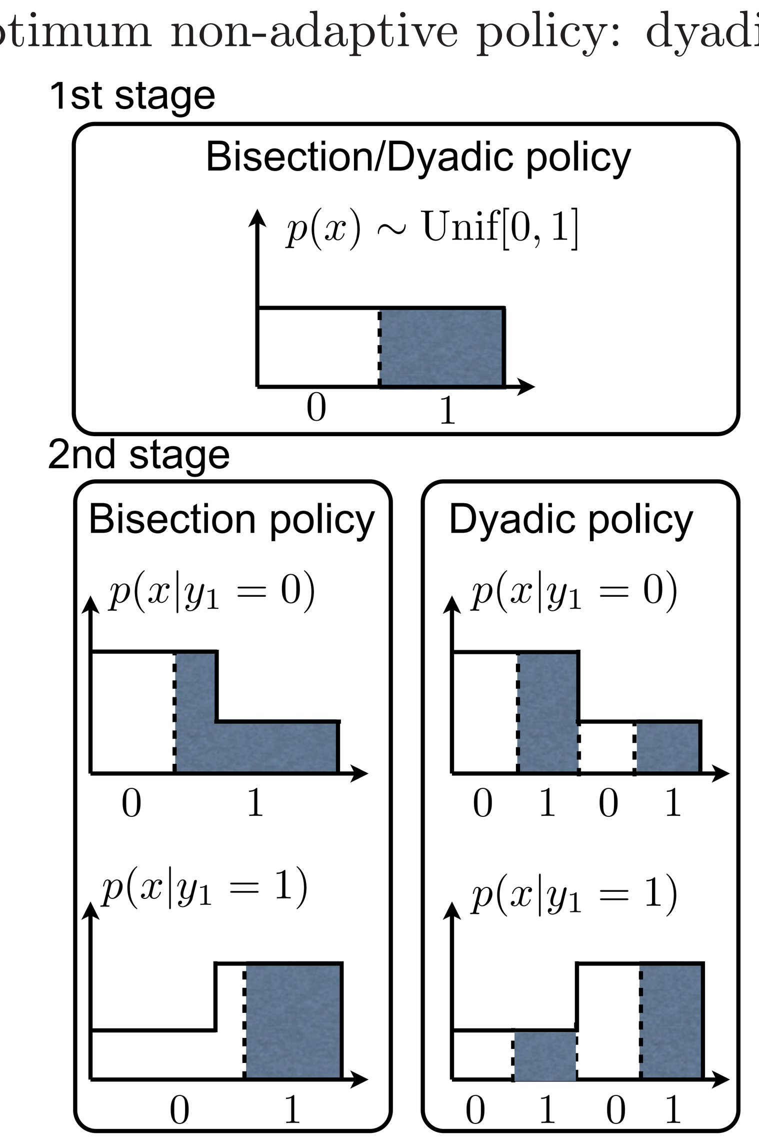
A Lower Bound on Conditional Entropy

$$h(X|Y_1^n) \geq h(X) - nC \text{ where } C = 1 - H_B(\epsilon)$$

Equality can be achieved by designing A_i's such that $\{\Pr(X \in A_i), \Pr(X \notin A_i)\}$ to be the optimum input distribution to the BSC(ϵ), which is $\{1/2, 1/2\}$.

Adaptive vs. Non-Adaptive Policy

- Optimum adaptive policy: bisection rule
- Optimum non-adaptive policy: dyadic rule [2]



$\Rightarrow$ both bisection and dyadic policies achieve the optimum input distribution

$$\int_{A_i} p(x|y_1^{i-1})dx = 1/2, \quad \forall y_1^{i-1} \in \{0,1\}^{i-1},$$

and thus the maximum reduction of entropy.

Multiple Player Problem

Sequential vs. Joint Query Design

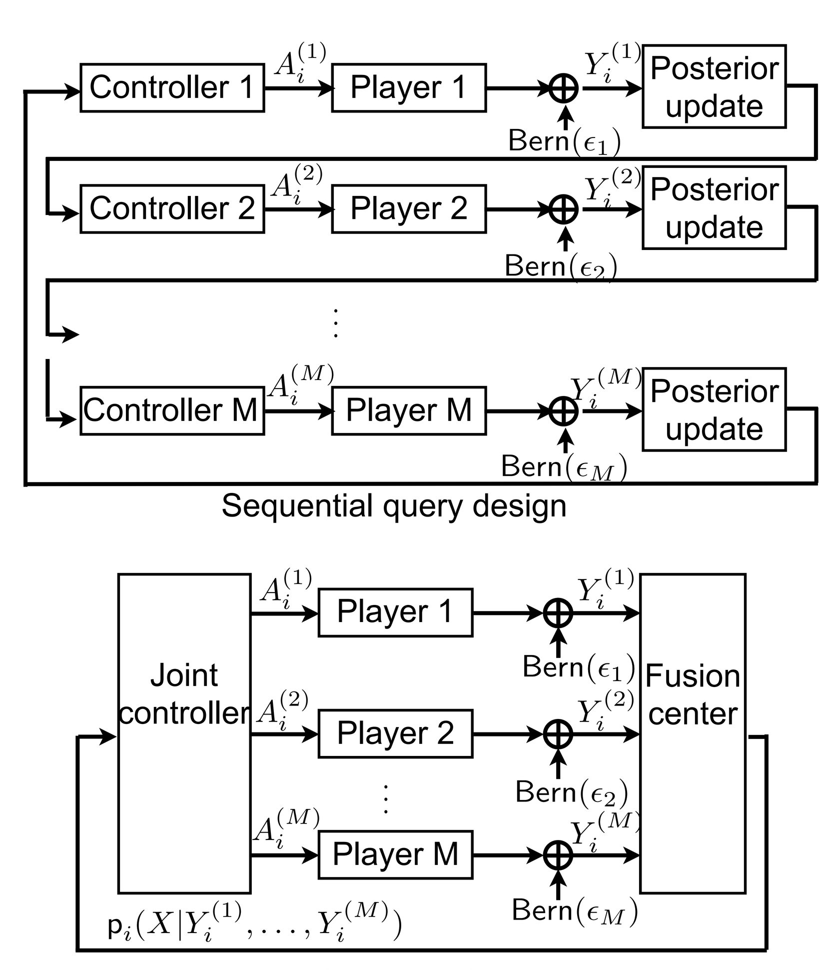


Image source: Tsiligkaridis [1]

- Sequential query design has access to a more refined filtration
- Optimum sequential policy: bisection rule
- Optimum joint policy: generalization of dyadic rule

Theorem 1 ([1]). *Both the optimum sequential policy and the optimum joint policy achieve the maximum entropy reduction of*

$$h(X|\{Y_i^{(1)}, \cdots, Y_i^{(M)}\}_{i=1}^n) \geq h(X) - n \cdot \bar{C} \qquad (1)$$

where $\bar{C} = \sum_{m=1}^M (1 - H_B(\epsilon_m))$ *after n-rounds.*

Mean Squared Error Bounds

Converse Bound
From the estimation counterpart to the Fano's inequality,

$$\mathbb{E}[|X - \hat{X}_n|^2] \geq \frac{e^{2h(p(X))}}{2\pi e}\exp\left(-2n\bar{C}\right) \qquad (2)$$

Achievability Bound of Adaptive Policy
Theorem 2 ([1]). *Using the (adaptive) bisection rule for M-players,*

$$\mathbb{E}[|X - \hat{X}_n|^2] \leq (2^{-2/3} + 2^{1/3})\exp\left(-\frac{2}{3}nC'\right) \qquad (3)$$

where $C' = \sum_{m=1}^M \left(1/2 - \sqrt{\epsilon_m(1-\epsilon_m)}\right)$.

Performance of Non-Adaptive Dyadic Policy
What this policy actually does is sending the binary expansion of the target location $X = 0.b_1 b_2 \ldots b_n$

- This policy cannot correct errors in bits $\Rightarrow$ large estimation error
- Each bit is treated with equal importance

$\Rightarrow$ Not all the policies achieving the $\min h(X|Y_1^n)$ give a good estimation error

What Does Entropy Fail to Capture?
Entropy is invariant to permutations $\Rightarrow$ It fails to capture the *concentration of measure*
Question: Can we find a class of distributions such that variance is a monotonic function of entropy?

Unimodal Distribution

Gaussian Distribution
For Gaussian distributions,

$$\mathbb{E}[|X - \hat{X}(Y_1^n)|^2] = \mathsf{Var}(X|Y_1^n) = \frac{1}{2}e^{2\cdot h(X|Y_1^n)}.$$

$\Rightarrow$ A lower entropy directly implies a lower MSE. Can it be true for more general unimodal distributions?

Unimodal Distribution
Theorem 3. *For unimodal distributions with exponentially decreasing tails, i.e.,* $p(X) = e^{-cx^{|\theta|}}/Z$ *where* $\theta > 1$,

$$\frac{e^{2\cdot h(X)}}{2\pi e} \leq \mathsf{Var}(X) \leq \frac{\alpha}{(\mathcal{I}(X))^\beta} \leq \frac{\alpha \cdot e^{2\beta \cdot h(X)}}{2\pi e} \qquad (4)$$

for some constant $\alpha, \beta > 0$ *and Fisher information* $\mathcal{I}(X) = \mathbb{E}[(\frac{\partial}{\partial x}\log p(x))^2]$.

Conditional Entropy as a Proxy for MSE

- A small entropy is necessary but not sufficient to achieve a small MSE
- Once the posterior distribution satisfies the enhanced unimodality condition, a strategy that minimizes conditional entropy can give an exponentially decreasing MSE
- Open question: what strategy can effectively form and maintain a unimodal posterior distribution?

Acknowledgement: This research was partially supported by MURI grant W911NF-11-1-0391.

Reference:
[1] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "A collaborative 20 questions model for target search with human-machine interaction," in *ICASSP, 2013 IEEE International Conference on.* IEEE, 2013, pp. 6516–6520.
[2] B. Jedynak, P. I. Frazier, R. Sznitman *et al.*, "Twenty questions with noise: Bayes optimal policies for entropy loss," *Journal of Applied Probability*, vol. 49, no. 1, pp. 114–136, 2012.