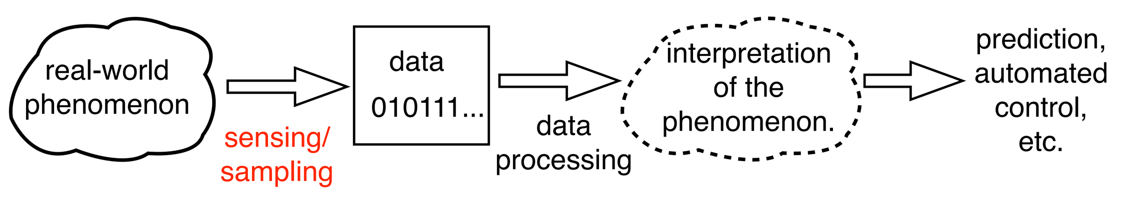
Unequal Error Protection Querying Policies for the Noisy 20 Questions Problem

Hye Won Chung*, Brian M. Sadler†, Lizhong Zheng††, and Alfred O. Hero**

* KAIST, † US Army Research Laboratory, †† MIT, ** University of Michigan.

Motivation

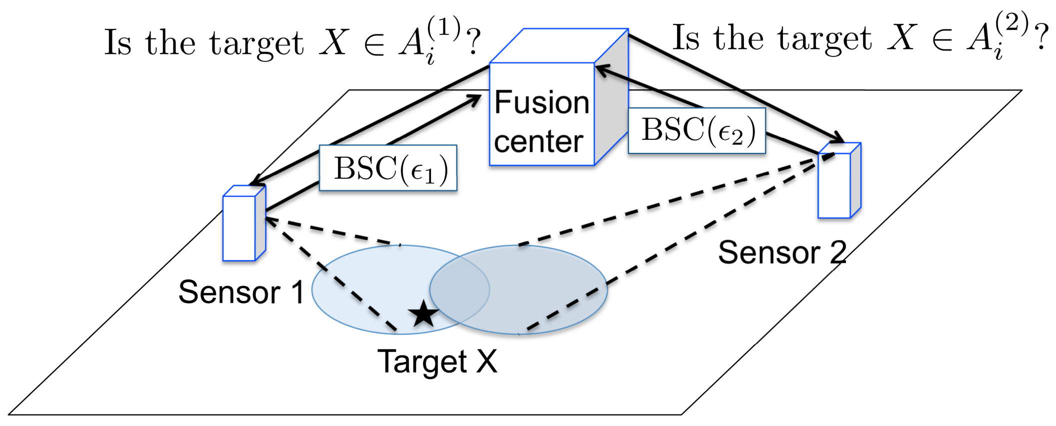
Data acquisition:



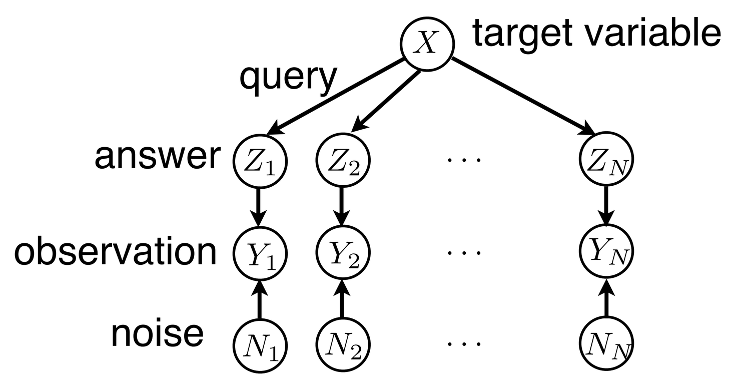
Quality of inference highly depends on the sensing/sampling methods.

Examples of data acquisition:

1. Medical diagnosis: Design sequence of tests that generate most informative data about the patient's condition.
2. Target localization: Choose the most informative querying regions to localize a target.



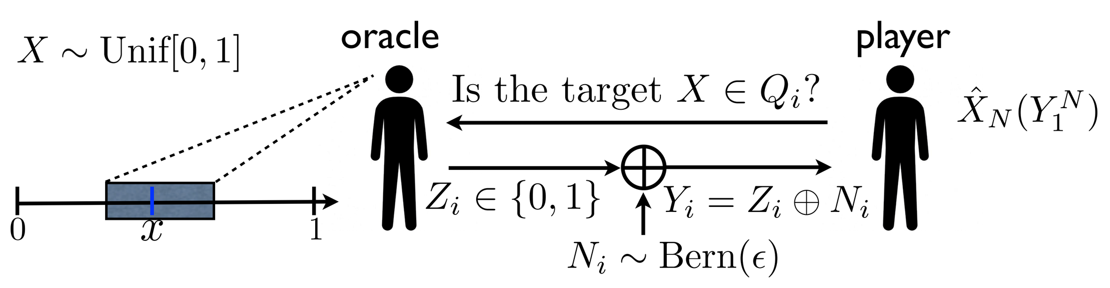
General model:



Goal: Design the sensing or querying strategies to selectively acquire the most useful samples or data given limited sensing resources.

Noisy 20 Questions Problem

Model: Noisy 20 questions problem to estimate the value of X



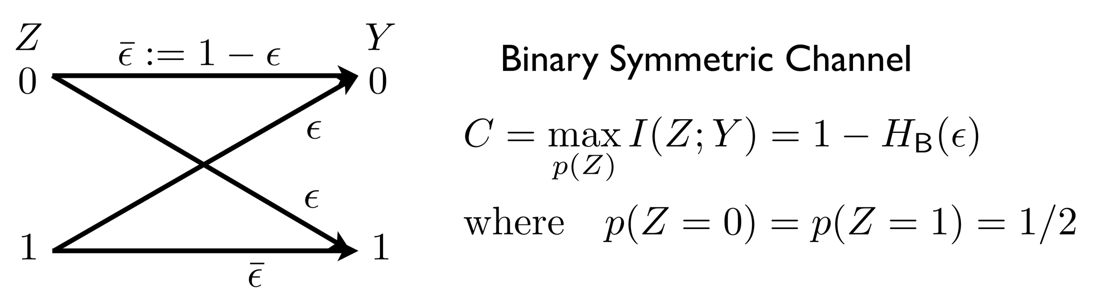
Goal: Design $(Q_1, \ldots, Q_N)$ to minimize estimation error $\mathbb{E}[c(X, \hat{X}_N)]$, e.g. mean squared error (MSE) $c(X, \hat{X}_N) = |X - \hat{X}_N|^2$.

Adaptive vs. non-adaptive policies:

- Adaptive sequential querying: Q_i depends on Y_1^{i-1}
- Non-adaptive block querying: $(Q_1, \ldots, Q_N)$ is fixed in advance

Previous Approaches

Many previous works focused on designing querying strategies that extract the maximum amount of information about X

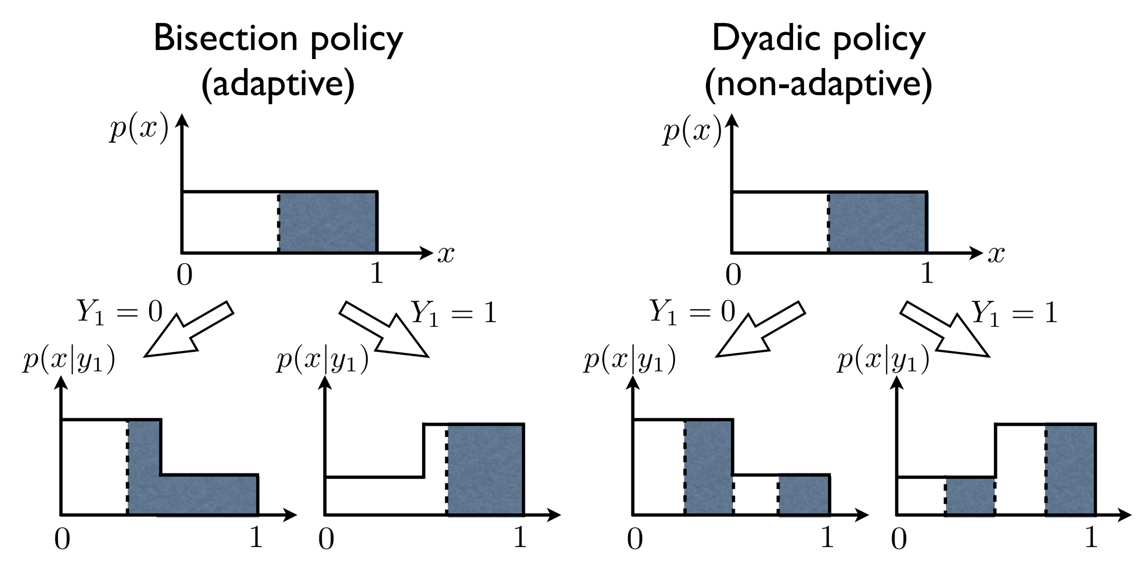


$$C = \max_{p(Z)} I(Z;Y) = 1 - H_B(\epsilon)$$

where $p(Z=0) = p(Z=1) = 1/2$

$$I(X; Y_1^N) = h(X) - h(X|Y_1^N) \le CN$$

Successive Entropy Minimization



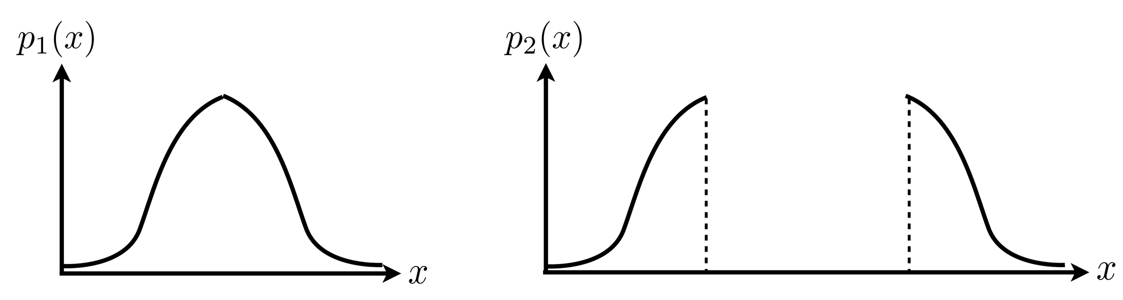
- Querying policies based on successive entropy minimization
  e.g. bisection policy [Horstein 1963], dyadic policy [Jedynak *et al.* 2012]
- Dyadic policy extracts bits in $X \approx 0.B_1B_2\ldots$
- Both achieve the maximum reduction of entropy $h(X|Y_1^N) \ge h(X) - NC$
- But very different performances in estimation
  - Bisection: $\mathbb{E}[|X - \hat{X}_N|^2] \le c_1 e^{-c_2 N}$ for $c_1, c_2 > 0$
  - Dyadic: $\mathbb{E}[|X - \hat{X}_N|^2] \ge \frac{\epsilon}{2} > 0$ even for $N \to \infty$

$\Rightarrow$ In general, maximizing mutual information and minimizing estimation error are very different goals.

Q: When do the two different goals coincide?

Variance vs. Entropy

- The estimation counterpart to the Fano's inequality shows $\frac{1}{2\pi e} e^{2h(p)} \le \mathrm{Var}(X)$.
- In general, no upper bound on variance in terms of differential entropy.



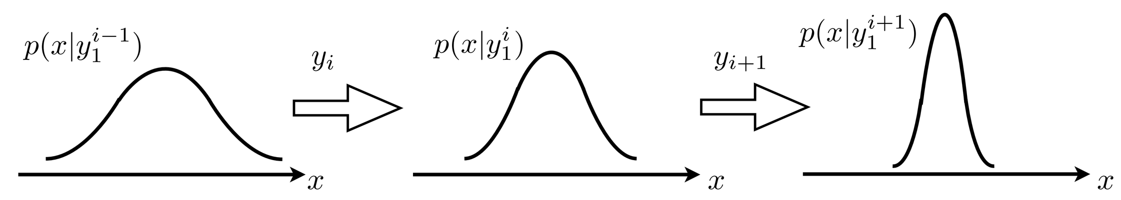
$$h(X) = -\int p_1(x)\log p_1(x)dx = -\int p_2(x)\log p_2(x)dx$$

- For special distributions, monotonic relationship
  Ex: Gaussian, Uniform,
  Log-concave [Bobkov, Madiman 2011]

Theorem 1 (Chung, Sadler and Hero 2017). *For the appropriate subclasses of unimodal distributions,*

$$\frac{1}{2\pi e} e^{2h(X)} \le Var(X) \le c \cdot e^{2h(X)}.$$

- For unimodal distributions, successive entropy minimization can guarantee exponential decrease of MSE



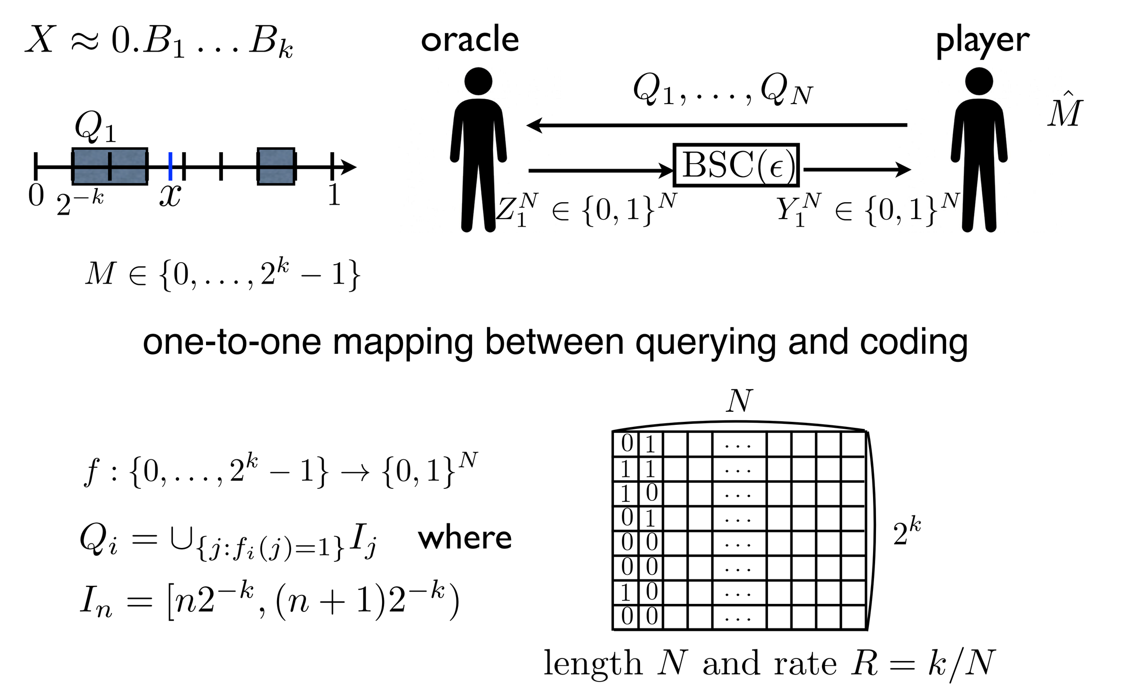
Reference:

[1] H. W. Chung, B. Sadler, and A. Hero, "Bounds on Variance for Symmetric Unimodal Distributions," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 6936–6949, Nov. 2017.

[2] H. W. Chung, L. Zheng, B. Sadler, and A. Hero, "Unequal Error Protection Coding Approaches to the Noisy 20 Questions Problem," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1105–1131, Feb. 2018.

Non-adaptive Block Querying

Mapping to channel coding:



Unequal error protection (UEP):

- Information bits of different significance

$$\mathbb{E}[|X - \hat{X}_N|^2] \le \sum_{i=1}^{k} \Pr(\hat{B}_i \ne B_i) 2^{-2(i-1)} + 2^{-2k}$$

$\Rightarrow$ Unequal error protection (UEP) is desirable

Superposition Coding

Idea: Partition $M = (B_1, \ldots, B_k)$ into two groups of different priorities

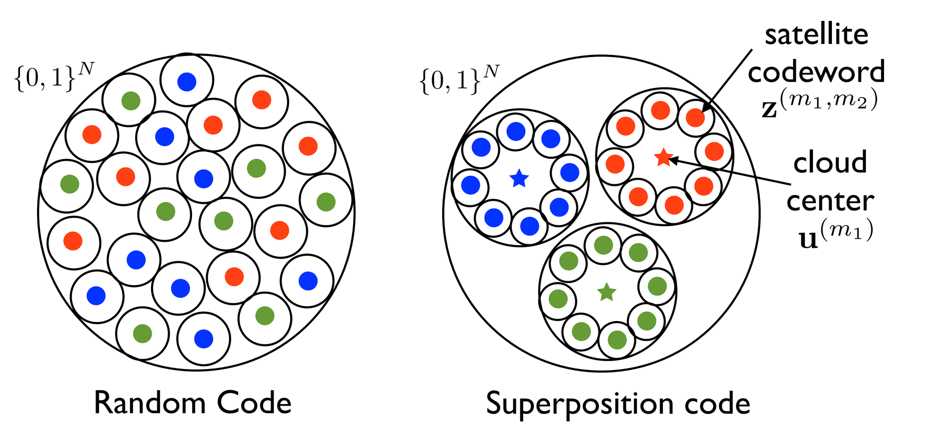
$$M_1 := (B_1, \ldots, B_{k_1}), \quad M_2 := (B_{k_1+1}, \ldots B_k)$$

Different importance of M_1 and M_2:

$$\mathbb{E}[|X - \hat{X}_N|^2] \le \Pr(\hat{M}_1 \ne M_1) + \Pr(\hat{M}_2 \ne M_2 | \hat{M}_1 = M_1) 2^{-2k_1} + 2^{-2k}$$

Goal: Design a block code that provides a better error protection for M_1 (MSBs) than for M_2 (LSBs)

Comparison of codeword distributions:



Superposition codewords $\mathbf{z}^{(m_1,m_2)} = \mathbf{u}^{(m_1)} \oplus \mathbf{v}^{(m_2)}$, for $m_1 \in [1 : e^{NR_1}]$, $m_2 \in [1 : e^{NR_2}]$, where $u_i^{(m_1)} \sim$ Bern(1/2), $v_j^{(m_2)} \sim$ Bern(α), $\alpha \in (0, 1/2)$.

Theorem 2 (Chung, Zheng, Sadler and Hero 2018). *For a very noisy BSC(ϵ), there exists positive gains in the MSE exponent* $\liminf_{N\to\infty} \frac{-\log \mathbb{E}[|X - \hat{X}_N|^2]}{N}$ *from superposition coding in high rate regimes* $R \in (C/6, C)$.

Gain in the MSE exponent: For $k = NR$, $\liminf_{N\to\infty} \frac{-\log \mathbb{E}[|X - \hat{X}_N|^2]}{N} = \min\{E_{\text{policy}}(R), 2R\}$