Quantile Search: A Distance-Penalized Active Learning Algorithm for Spatial Sampling

John Lipor[1], Laura Balzano[1], Branko Kerkez[2], and Don Scavia[3]

[1]Department of Electrical and Computer Engineering
[2]Department of Civil and Environmental Engineering
[3]School of Natural Resources and Environment
University of Michigan, Ann Arbor

September 30, 2015

Motivation: Sampling hypoxia in Lake Erie

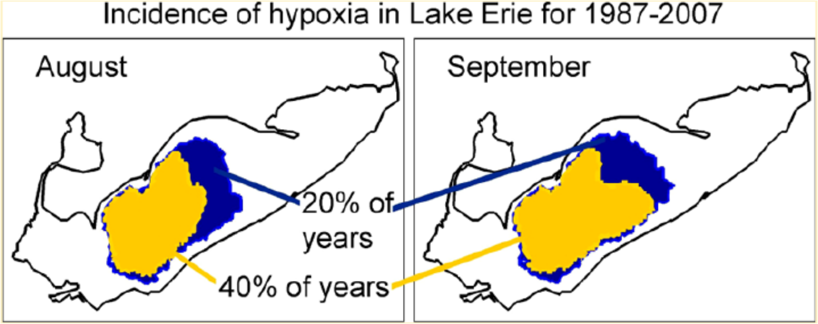
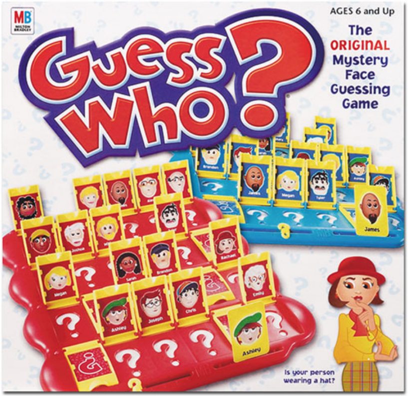
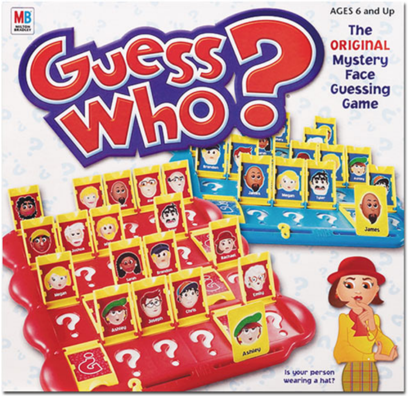


Figure: Figure from "Spatial and Temporal Trends in Lake Erie Hypoxia, 1987-2007". Zhou, Obenour, Scavia, Johengen, Michalak. ACS Journal of Environmental Science and Technology, 2012.

Active Learning/Adaptive Sampling

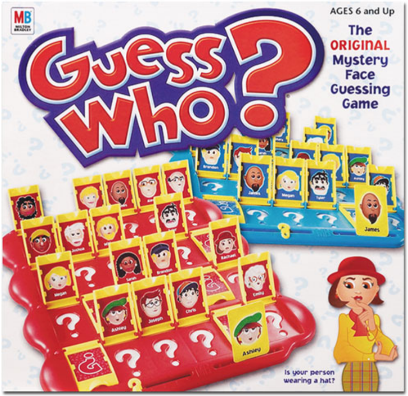


Active Learning/Adaptive Sampling



"Is the person male/female?"

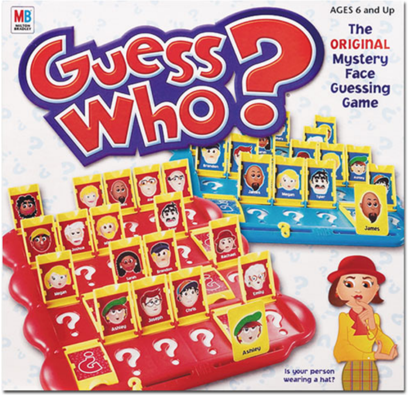
Active Learning/Adaptive Sampling



"Is the person male/female?"

"Is he wearing a hat?"

Active Learning/Adaptive Sampling

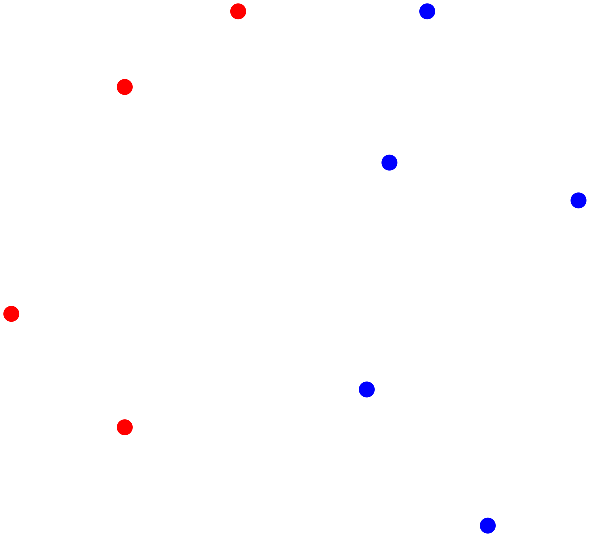


"Is the person male/female?"

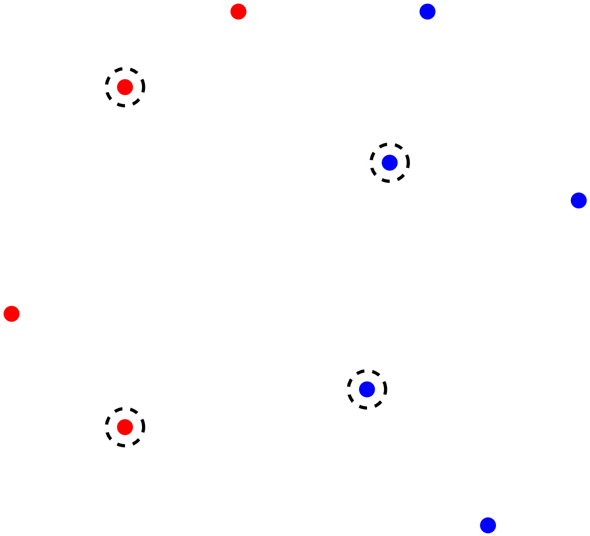
"Is she wearing a hat?"

Goal is to cut remaining possibilities in half with each question.

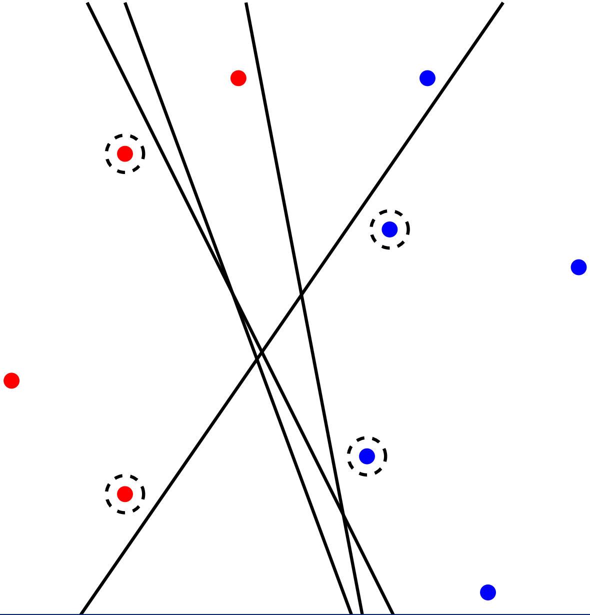
An Active Learning Example



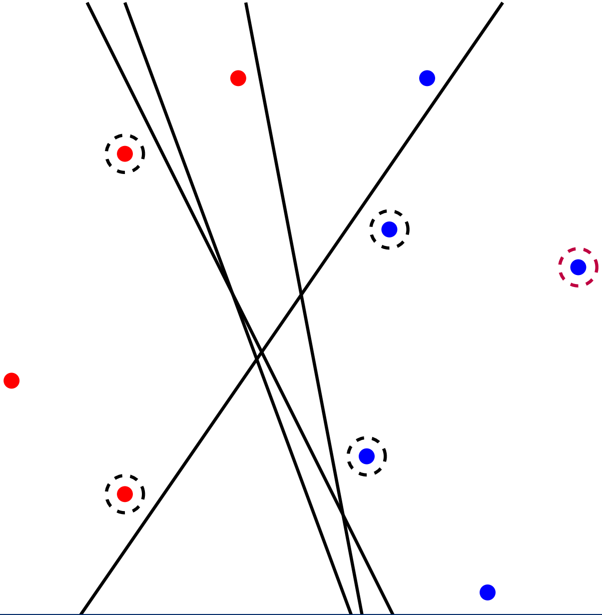
An Active Learning Example



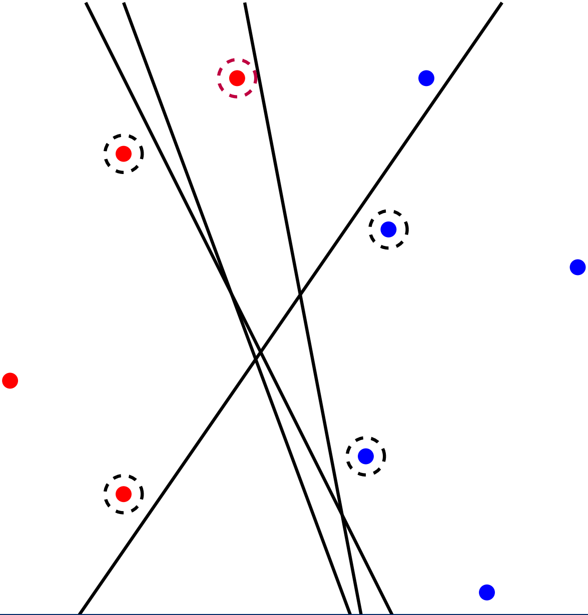
An Active Learning Example



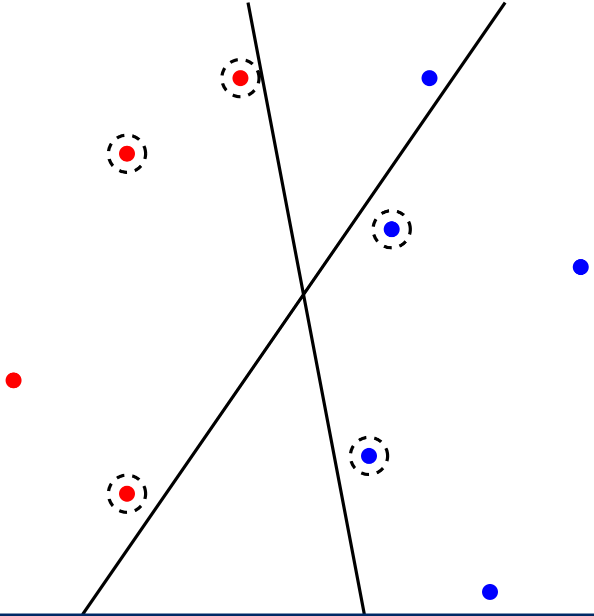
An Active Learning Example



An Active Learning Example



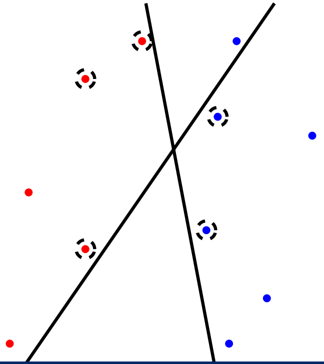
An Active Learning Example



Active Learning/Adaptive Sampling

In learning setting, binary search-like methods are nearly optimal (Dasgupta, “Analysis of a Greedy Active Learning Strategy,” NIPS 2005).

Similar results for *compressive binary search* in compressed sensing (Davenport & Arias-Castro, “Compressive Binary Search,” ISIT 2012).



Sampling the Great Lakes

Most existing algorithms minimize only the number of samples. This does not necessarily reduce the total sampling cost (Settles, *Active Learning*, Morgan & Claypool, 2012).

Sampling the Great Lakes: Problem Setup

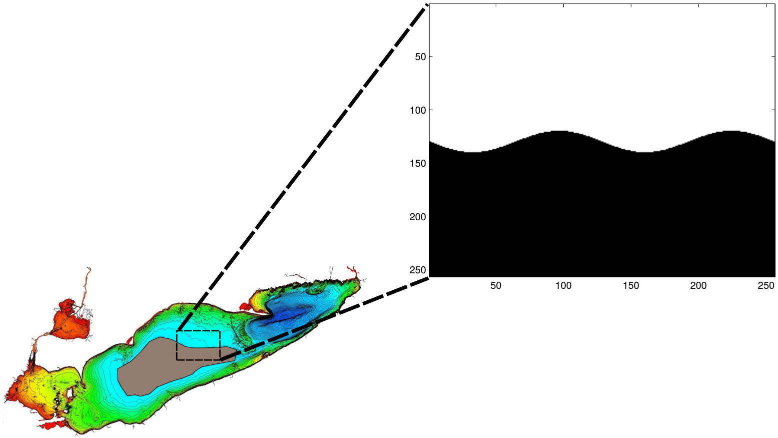


Figure: Measurements are thresholded and can be viewed as binary step functions along any path that crosses the boundary.

Problem Formulation

Our function comes from the class of step functions

$$\mathcal{F} = \{f : [0,1] \to \{0,1\} \,|\, f(x) = \mathbf{1}_{\{[0,\theta)\}}(x)\}$$

Goal: Find θ while minimizing the *total sampling time*.

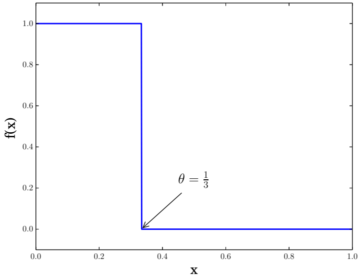
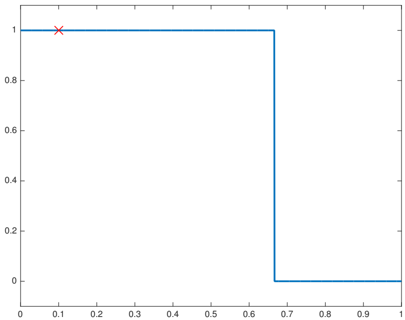
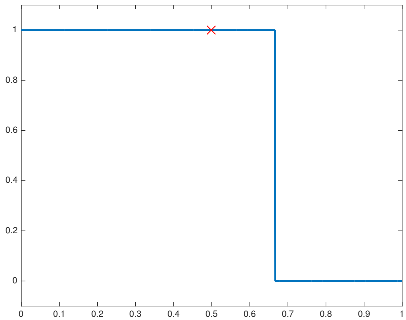


Figure: Measurements are thresholded and can be viewed as binary step functions along any path that crosses the boundary.

Sample Complexity

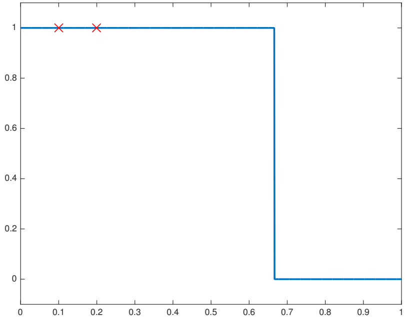


(a) Uniform sampling

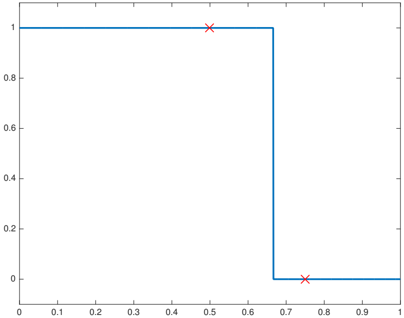


(b) Binary bisection

Sample Complexity

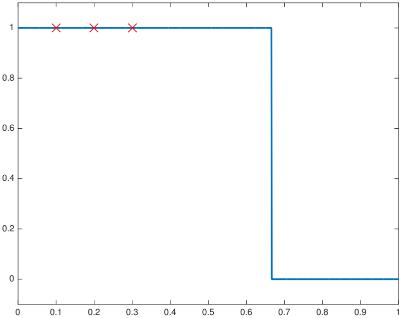


(a) Uniform sampling

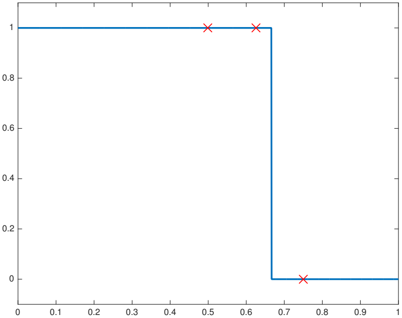


(b) Binary bisection

Sample Complexity

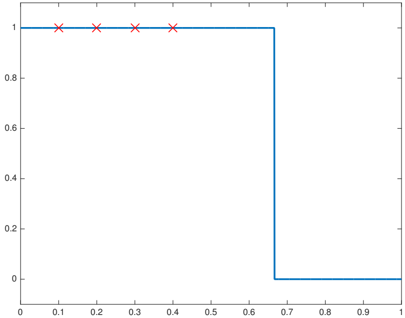


(a) Uniform sampling

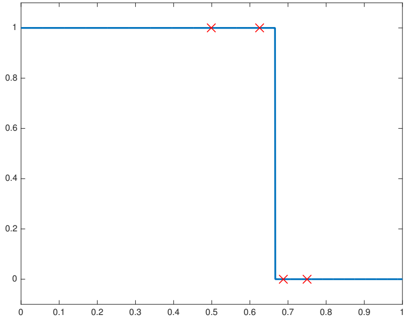


(b) Binary bisection

Sample Complexity

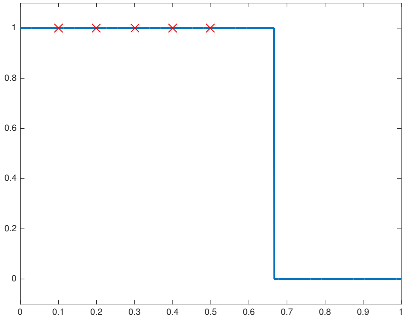


(a) Uniform sampling

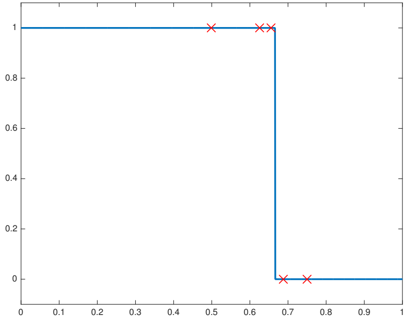


(b) Binary bisection

Sample Complexity

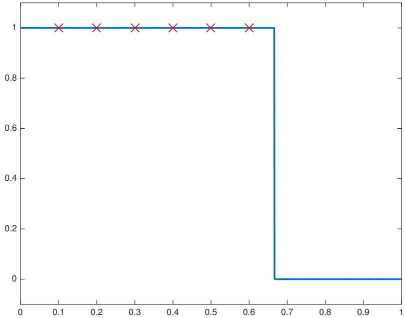


(a) Uniform sampling

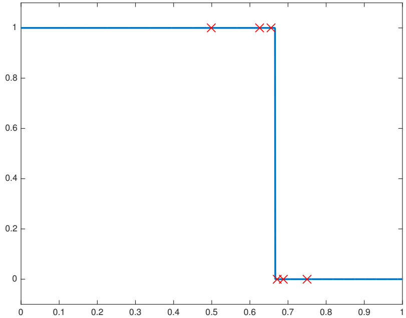


(b) Binary bisection

Sample Complexity

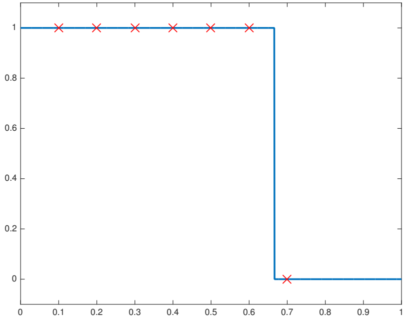


(a) Uniform sampling

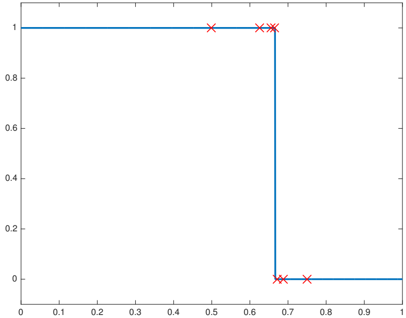


(b) Binary bisection

Sample Complexity



(a) Uniform sampling



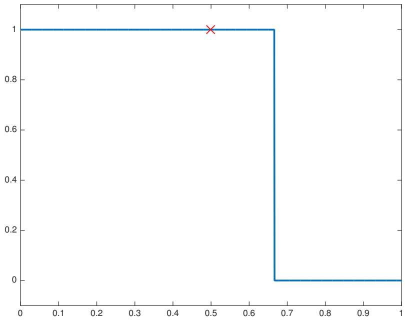
(b) Binary bisection

Complexity and Distance of Binary Search

	worst case	expected (with uniform prior on θ)
non-adaptive sample complexity	$\sup_{\theta\in[0,1]} \left\|\hat{\theta}_n - \theta\right\| \leq \frac{1}{2}\frac{1}{n+1}$	$\mathbb{E}\left[\left\|\hat{\theta}_n - \theta\right\|\right] = \frac{1}{4}\frac{1}{n+1}$
bisection sample complexity	$\sup_{\theta\in[0,1]} \left\|\hat{\theta}_n - \theta\right\| \leq \frac{1}{2}\frac{1}{2^n}$	$\mathbb{E}\left[\left\|\hat{\theta}_n - \theta\right\|\right] = \frac{1}{4}\frac{1}{2^n}$
bisection distance traveled (fix desired error ε, start at $X_1 = 0$)	$1 - \varepsilon$	$1 - 2\varepsilon$

Quantile Search

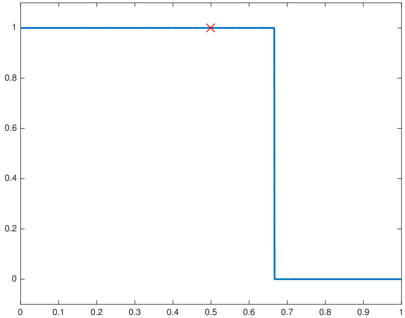
Quantile search samples at $1/m$ $(m \geq 2)$ into the feasible interval at each step.



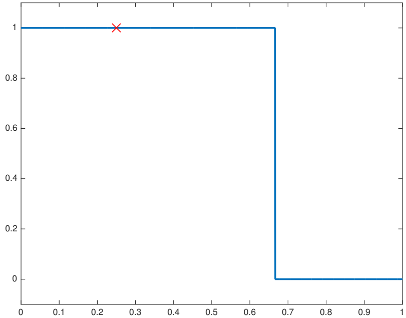
(a) Binary bisection

(b) Quantile search, $m = 4$

Quantile Search

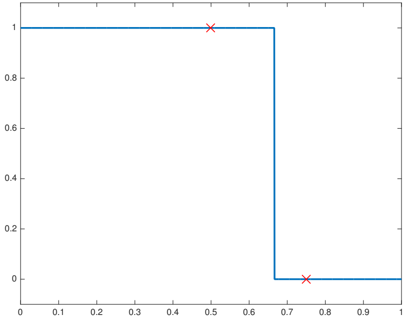


(a) Binary bisection

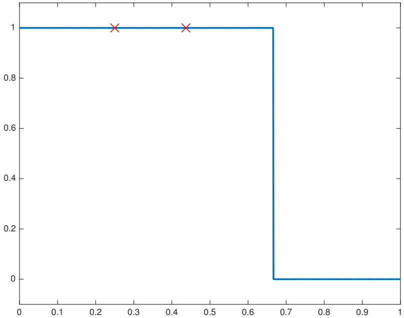


(b) Quantile search, $m = 4$

Quantile Search

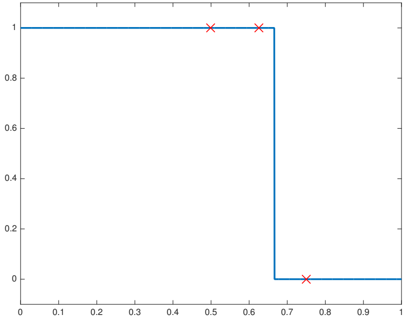


(a) Binary bisection

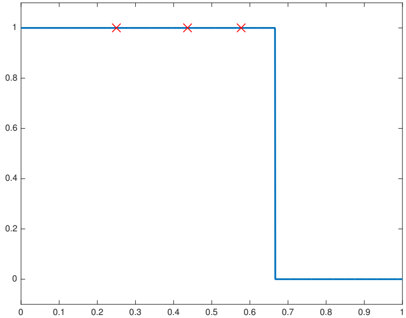


(b) Quantile search, $m = 4$

Quantile Search

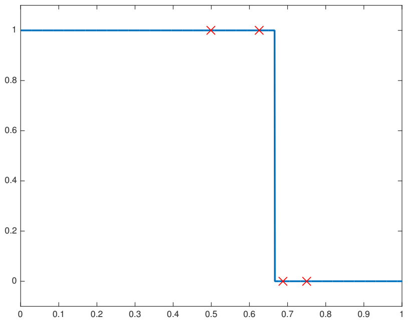


(a) Binary bisection

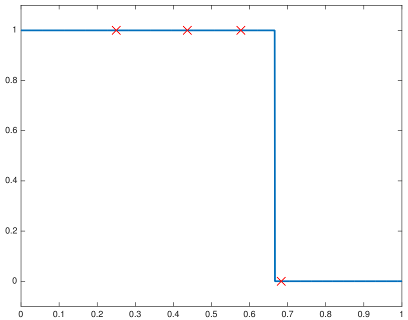


(b) Quantile search, $m = 4$

Quantile Search

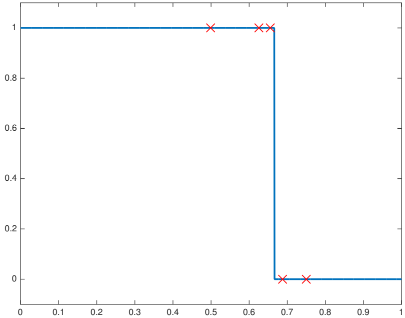


(a) Binary bisection

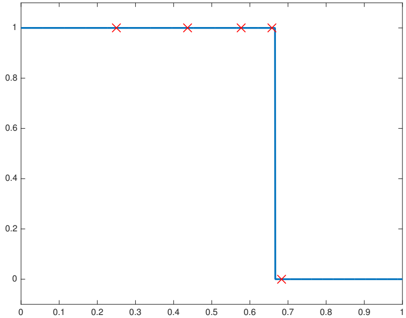


(b) Quantile search, $m = 4$

Quantile Search

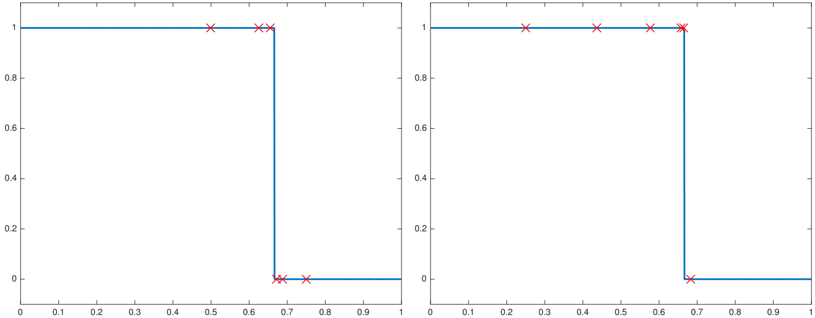


(a) Binary bisection



(b) Quantile search, $m = 4$

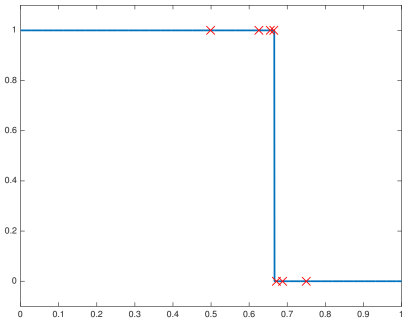
Quantile Search



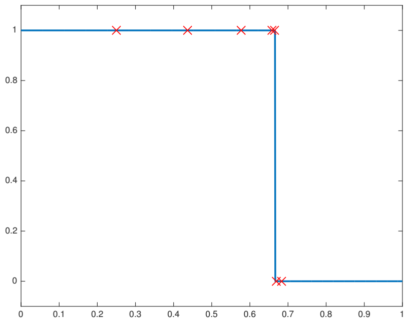
(a) Binary bisection

(b) Quantile search, $m = 4$

Quantile Search



(a) Binary bisection

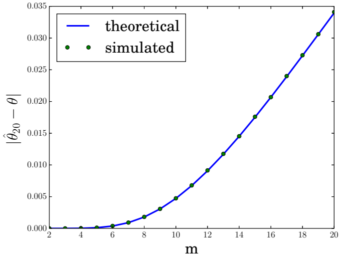


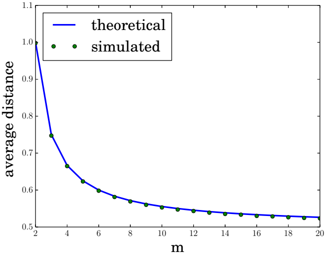
(b) Quantile search, $m = 4$

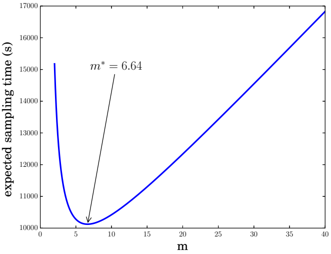
Quantile Search Cost

	worst case	expected (with uniform prior on θ)
bisection sample complexity	$\sup_{\theta\in[0,1]} \left\|\hat{\theta}_n - \theta\right\| \leq \frac{1}{2}\frac{1}{2^n}$	$\mathbb{E}\left[\left\|\hat{\theta}_n - \theta\right\|\right] = \frac{1}{4}\frac{1}{2^n}$
m-quantile search sample complexity $\rho = \frac{m-1}{m}$	$\sup_{\theta\in[0,1]} \left\|\hat{\theta}_n - \theta\right\| \leq \frac{1}{2}\rho^n$	$\mathbb{E}\left[\left\|\hat{\theta}_n - \theta\right\|\right] = \frac{1}{4}\left(\rho^2 + (1-\rho)^2\right)^n$
bisection distance traveled (as desired error $\varepsilon \to 0$)	1	1
m-quantile search distance traveled (as desired error $\varepsilon \to 0$)	1	$\frac{m}{2m-2}$

Sample-Distance Tradeoff



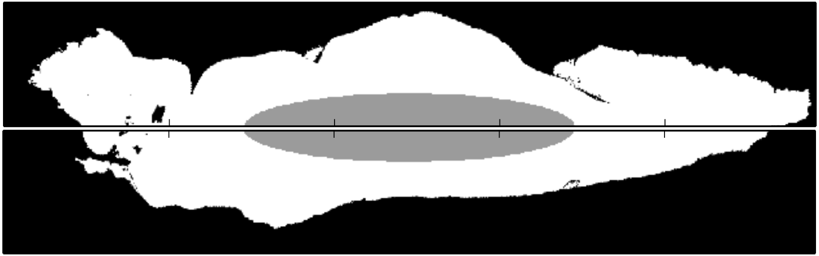




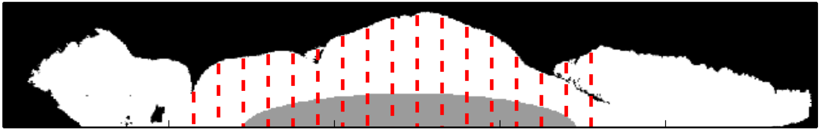
(Top left) Simulated and theoretical expected error after 20 samples as a function of m, (Top right) Simulated and theoretical distance traveled as a function of m, and (Bottom left) Optimal m for $\gamma = 60\frac{s}{\text{sample}}$ and $\eta = \frac{1}{4}\frac{s}{m}$.

Solving the Lake Problem

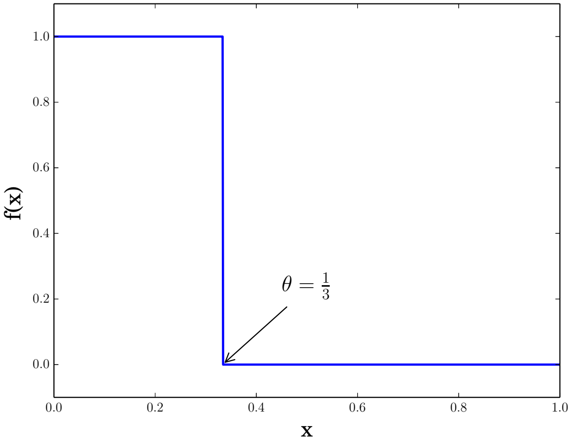
Step 1: Split the lake



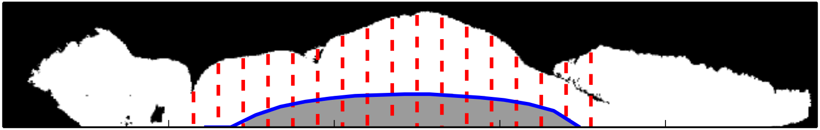
Step 2: Divide each half into strips



Step 3: Quantile Search



Step 4: Piecewise Linear Estimate



This approach is shown to be optimal to within a logarithmic factor for a version of binary search (Castro & Nowak, "Minimax Bounds for Active Learning," IEEE Trans. Inf. Theory, 2008).

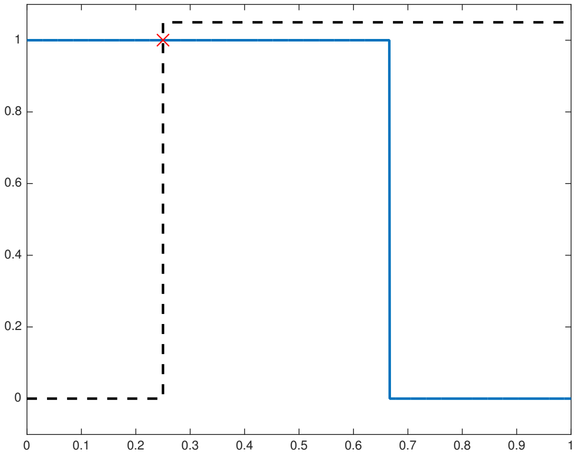
Lake Erie Results (no noise)

Sampling Time (s)	Speed (m/s)	m	Total Time (hrs)
60	4	2	62
60	4	6.64	43
60	2	2	123
60	2	8.92	81
10	4	2	61
10	4	14.63	35
10	2	2	122
10	2	20.26	64

We split the lake in half and then into 16 strips and perform DQS. In most cases, we can sample the entire boundary in 2-3 days.

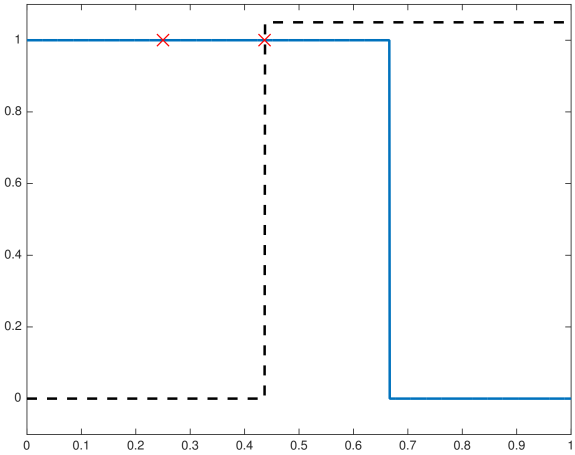
Quantile Search: An Alternative Perspective

Quantile search samples at the first m-quantile of the posterior distribution of θ.



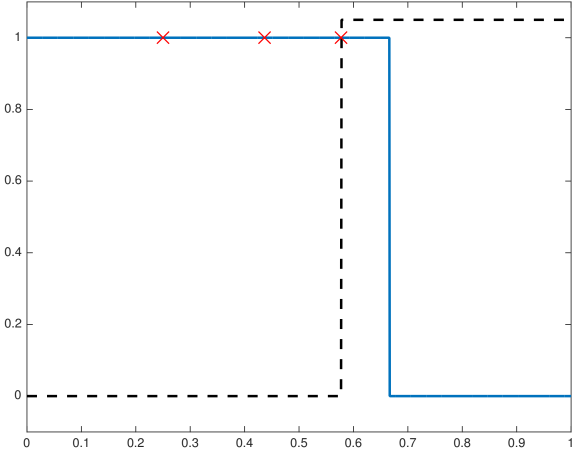
Quantile Search: An Alternative Perspective

Quantile search samples at the first m-quantile of the posterior distribution of θ.



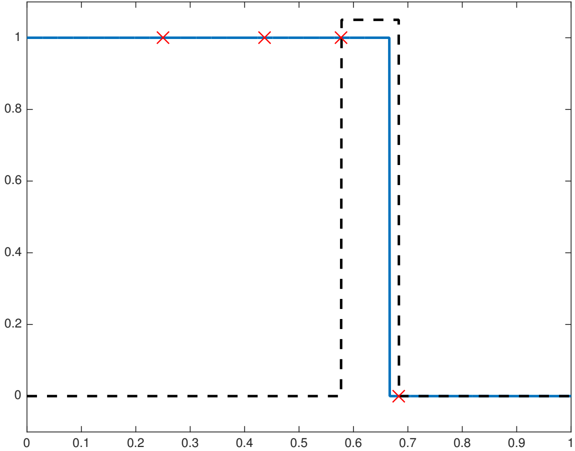
Quantile Search: An Alternative Perspective

Quantile search samples at the first m-quantile of the posterior distribution of θ.



Quantile Search: An Alternative Perspective

Quantile search samples at the first m-quantile of the posterior distribution of θ.



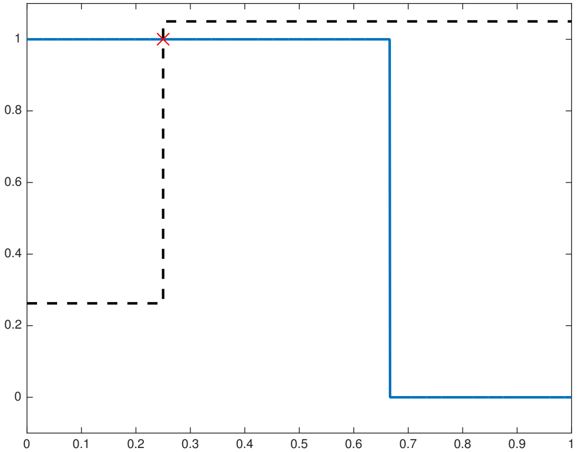
Probabilistic Quantile Search: Handling noise

We deal with noise by maintaining a Bayesian estimate on the posterior distribution of θ, then sample the first m-quantile of this posterior (Burnashev & Zigangirov, "An Interval Estimation Problem for Controlled Observations," Problems in Inf. Transmission, 1974).

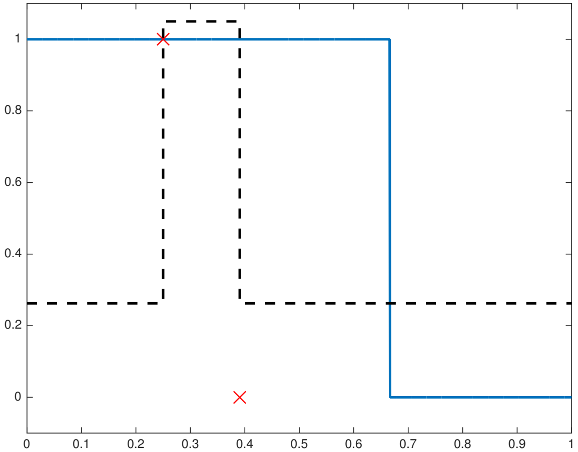
Algorithm 1 Probabilistic Quantile Search (PQS)

1: initialize prior density $\pi_0(\theta = x) = 1$ for all $x \in [0,1]$
2: **while** not converged **do**
3: choose X_n such that $\int_0^{X_n} \pi_n(x)dx = 1/m$
4: $Y_n \leftarrow f(X_n)$
5: perform Bayesian update to obtain $\pi_{n+1}(x)$
6: **end while**
7: **return** $\hat{\theta}_n$ such that $\int_0^{\hat{\theta}_n} \pi_{n+1}(x)dx = 1/2$

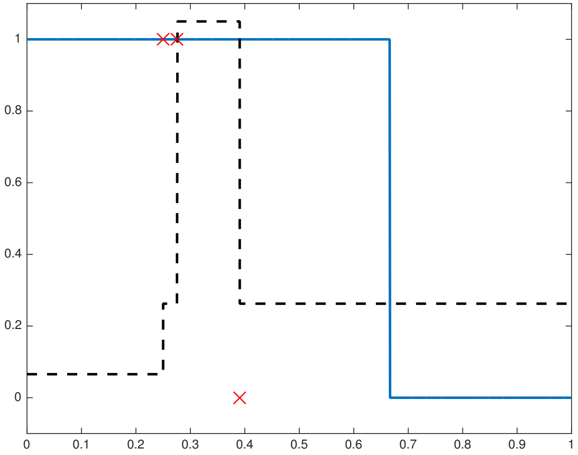
Probabilistic Quantile Search: Handling noise



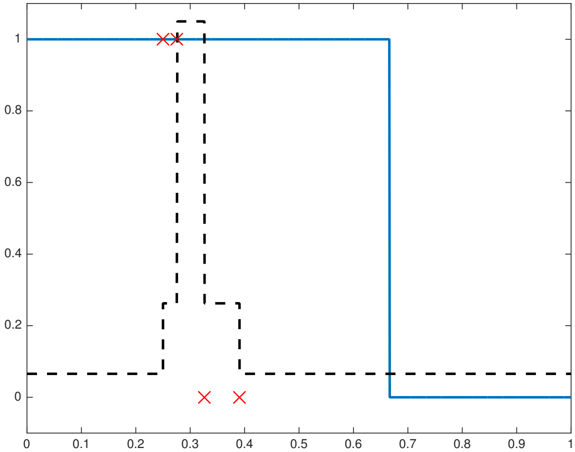
Probabilistic Quantile Search: Handling noise



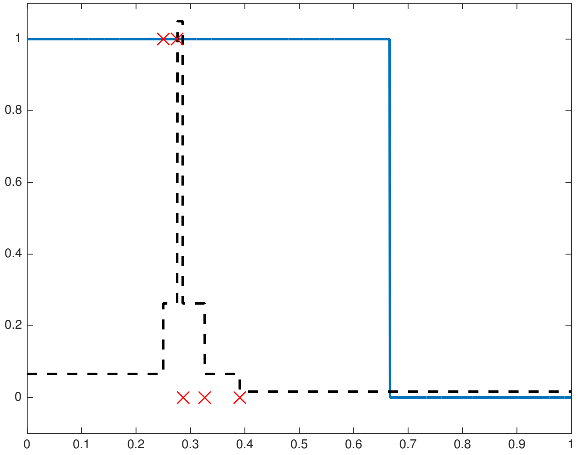
Probabilistic Quantile Search: Handling noise



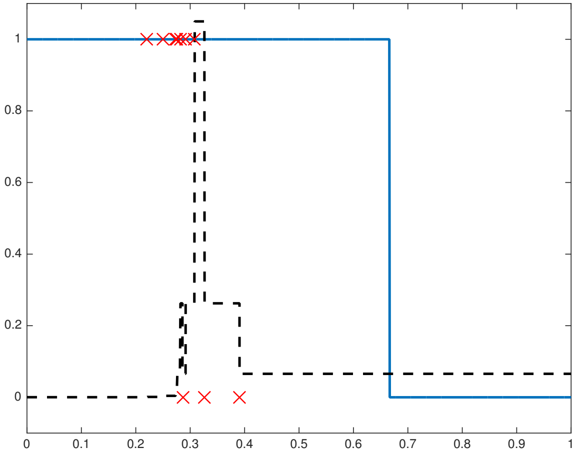
Probabilistic Quantile Search: Handling noise



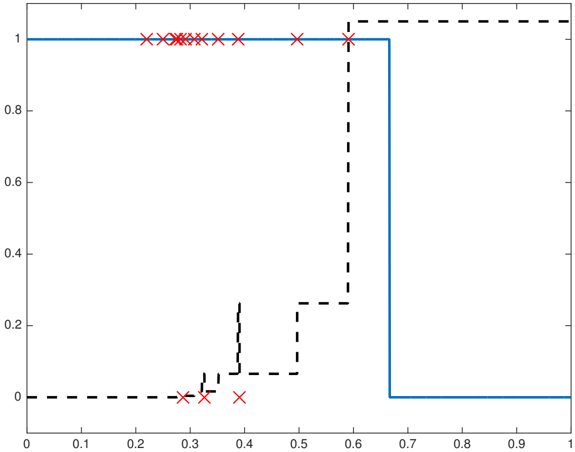
Probabilistic Quantile Search: Handling noise



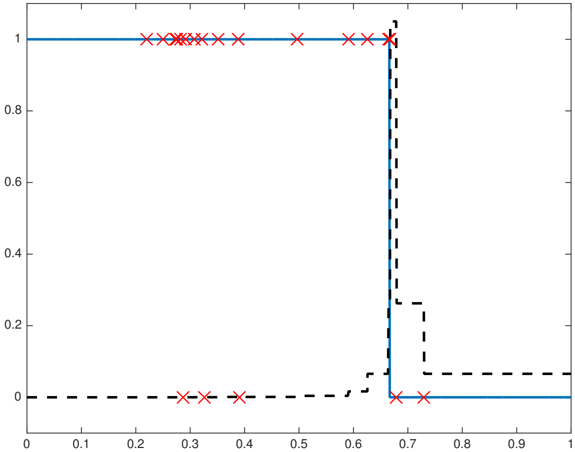
Probabilistic Quantile Search: Handling noise



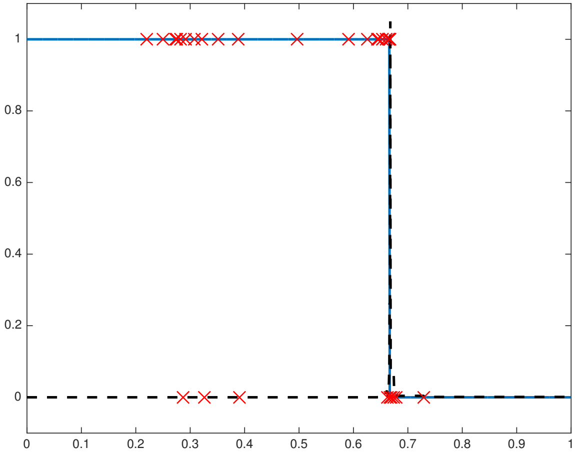
Probabilistic Quantile Search: Handling noise



Probabilistic Quantile Search: Handling noise



Probabilistic Quantile Search: Handling noise

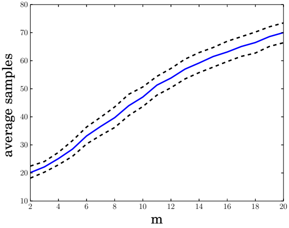


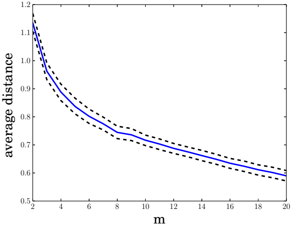
PQS Rate of Convergence

A discretized version of the PQS algorithm satisfies

$$\sup_{\theta \in [0,1]} \mathbb{E}[|\hat{\theta}_n - \theta|] \leq 2 \left(\frac{m-1}{m} + \frac{2\sqrt{p(1-p)}}{m} \right)^{n/2}.$$

The above matches Castro & Nowak for $m = 2$ (binary bisection).





Conclusions and Future Work

Conclusions:

- Quantile search achieves a tradeoff between required samples and distance traveled
- Solve lake problem as series of one-dimensional searches

Future work:

- Optimality?
- Leverage models and smoothness to provide nonuniform prior

Full details found in arXiv paper:

- Lipor, Balzano, Kerkez, and Scavia, "Quantile Search: A Distance-Penalized Active Learning Algorithm for Spatial Sampling," arXiv 1509.08387, 2015.