Bayes-optimal Policies for Multiple Comparisons with a Known Standard

Jing Xie and Peter I. Frazier

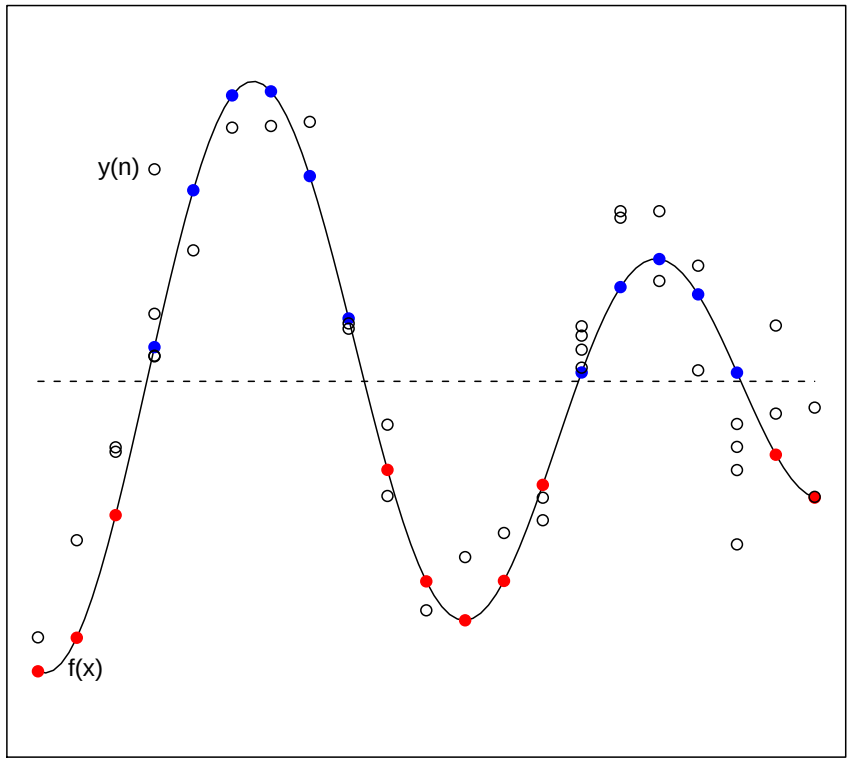
Operations Research & Information Engineering, Cornell University

Monday January 7, 2013
INFORMS Computing Society Conference
Sante Fe, NM

Supported by AFOSR YIP #FA9550-11-1-0083

What is Multiple Comparisons with a Known Standard?

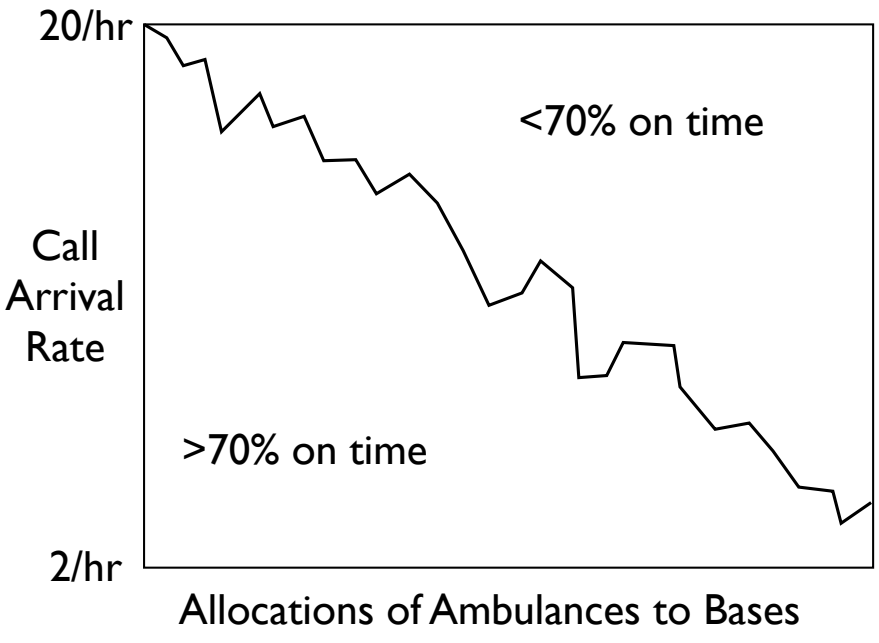
- We have a stochastic simulator.
- Given a set of input parameters x, it provides a random sample $y(x)$.
- For which inputs x is $E[y(x)] > 0$?

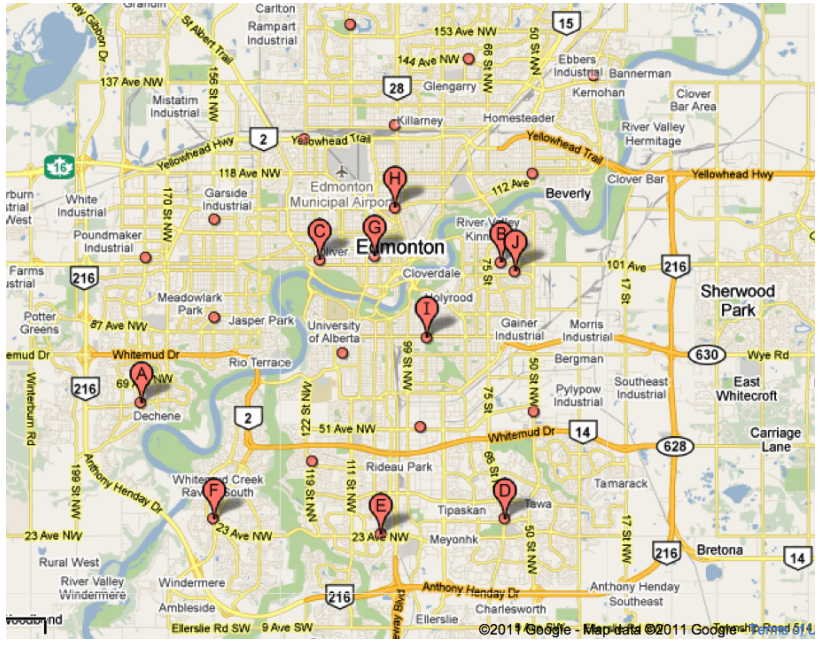


In other words, find the level set of $x \mapsto \mathbb{E}[f(x)]$.

Multiple Comparisons Appears In Ambulance Positioning

We must allocate ambulances across 11 bases in the city of Edmonton. Which allocations satisfy mandated minimums for percentage of calls answered in time, under a variety of different possible call arrival rates?





[Thanks to Shane Henderson and Matt Maxwell for providing the ambulance simulation]

We Provide an Optimal Way to Choose Where to Sample

- If our simulator is complex and takes a long time to run, the number of samples we can take is limited.
- This makes accurate MCC more difficult.
- **Where should we place our limited samples to estimate the level set as accurately as possible?**
- Our contribution: We provide an answer to this question with Bayes-optimal performance.

Literature Review

- Procedures with frequentist indifference-style guarantees on solution quality: [Paulson, 1962, Kim, 2005, Bechhofer and Turnbull, 1978, Nelson and Goldsman, 2001, Andradóttir et al., 2005, Andradóttir and Kim, 2010, Batur and Kim, 2010, Healey et al., 2012]
- Procedures based on large deviations analysis, which use allocations that minimize the rate of decay of the error probability as the sample size grows to infinity: [Szechtman and Yücesan, 2008, Hunter and Pasupathy, 2010, Hunter and Pasupathy, 2012].
- All of this literature is in the frequentist setting. We focus on sampling sequentially in a Bayes-optimal way, and use a non-asymptotic analysis.

Mathematical Model

- We have alternatives $x = 1, \ldots, k$.
- Samples from alternative x are $\mathrm{Normal}(\mu_x, \sigma_x^2)$.
- μ_x is unknown, while σ_x^2 is assumed known.
- We have an independent normal Bayesian prior on each μ_x.
- At each point in time, I either choose one alternative to sample, or I choose to stop sampling. This decision is adaptive.
- We model a limited ability to sample in two ways:
  - An externally imposed deadline T may require us to stop. For analytic tractability, this deadline is random, and is geometric with mean $1/(1-\alpha)$. More on the fixed deadline case later.
  - Each sample carries a cost $c \geq 0$.
  - We can include just one of these limits, or both.
    (To ignore deadline, set $\alpha = 1$, $T = \infty$. To ignore cost, set $c = 0$.)
- When sampling stops, we estimate the level set $\{x : \mu_x > 0\}$ based on the samples. The reward is the number of alternatives correctly classified.

Finding the Optimal Policy Means Solving a Dynamic Program

- Let $\tau \leq T$ be the number of samples taken (either equal to the external deadline, or can be strictly less if we choose to stop early).
- Let R be the number of alternatives classified correctly when sampling stops.
- We wish to find the policy π that solves:

$$\sup_{\pi} \mathbb{E}^{\pi}[R - c\tau],$$

  where a policy π is a rule for choosing whether and where to sample next, based on previous observations.
- The solution is characterized via dynamic programming...
- ...But, the **curse of dimensionality** usually makes computing the solution to such dynamic programs intractable.

We Rewrite the Problem as a Bandit Problem

- Let's analyze the case where we have a geometric external deadline ($\alpha < 1$), and no sampling costs ($c = 0$). Then $\tau = T$ is optimal, and we choose π to maximize $\mathbb{E}^{\pi}[R]$.
- The expected reward is the expected number of alternatives correctly classified at the end.
- We decompose this expected reward into an infinite sum of discounted expected one-step rewards

$$\mathbb{E}^{\pi}[R] = R_0 + \mathbb{E}^{\pi}\left[\sum_{n=1}^{\infty} \alpha^{n-1} R_n\right].$$

Here,

- α is the parameter of the geometric distribution of the deadline.
- R_0 is the expected reward if we stop after taking no samples.
- R_n is the expected one-step improvement, due to sampling, of the probability of correctly classifying the alternative sampled.

We Can Compute the Optimal Policy

- Written in this way, the problem becomes a **multi-armed bandit** problem.
- [Gittins and Jones, 1974] shows the optimal solution is

$$\arg\max_x \nu_x(S_{nx}),$$

where S_{nx} is a parameterization of the Bayesian posterior on μ_x.

- The Gittins index $\nu_x(\cdot)$ is defined in terms of a single-alternative version of the problem

$$\nu_x(s) = \sup_{\rho>0} \frac{\mathbb{E}\left[\sum_{n=1}^{\rho} \alpha^{n-1} R_n | S_{0x} = s, x_1 = \cdots = x_\rho = x\right]}{\mathbb{E}\left[\sum_{n=1}^{\rho} \alpha^{n-1} | S_{0x} = s, x_1 = \cdots = x_\rho = x\right]}.$$

- We can compute Gittins indices efficiently because the single-alternative problem is much smaller than the full DP.

Numerical Results: Ambulance Positioning

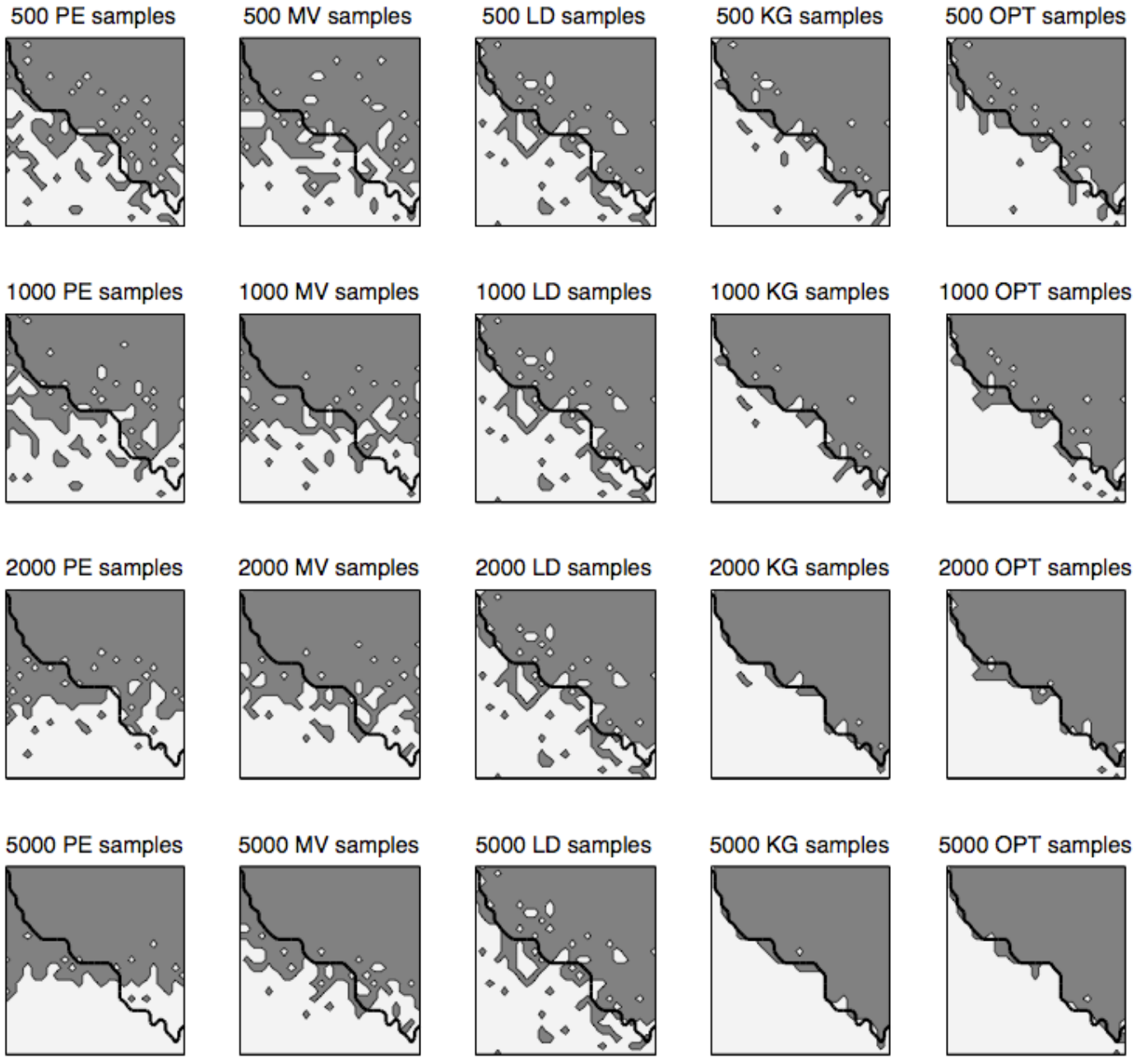
PE=pure exploration (sample at random);

MV=max variance (equal allocation);

LD=large deviations.

KG=knowledge-gradient.

OPT=optimal policy.



Samples ↓

What About Other Ways of Modeling A Limited Ability to Sample?

- Recall that we were considering the case with $\alpha < 1$ and $c = 0$ (geometric deadline and no sampling costs).
- If $\alpha < 1$ and $c > 0$, we can rewrite the problem as a slightly different bandit problem and also compute the optimal policy.
- If $\alpha = 1$ (no external deadline) and $c > 0$, then we can decompose across alternatives in a different way (not a bandit problem), and again compute the optimal policy.

What About A Fixed Deadline?

If our external deadline T is fixed, then we cannot compute the optimal policy, but we can compute an upper bound using a Lagrangian relaxation:

- Relax the constraint "take no more than T samples" to "take no more than T samples on average", and allow adaptive stopping.
- Put the constraint in the relaxed problem into the objective with a Lagrange multiplier λ,

$$\sup_{\pi} \mathbb{E}^{\pi}[R - \lambda(T - \tau)]$$

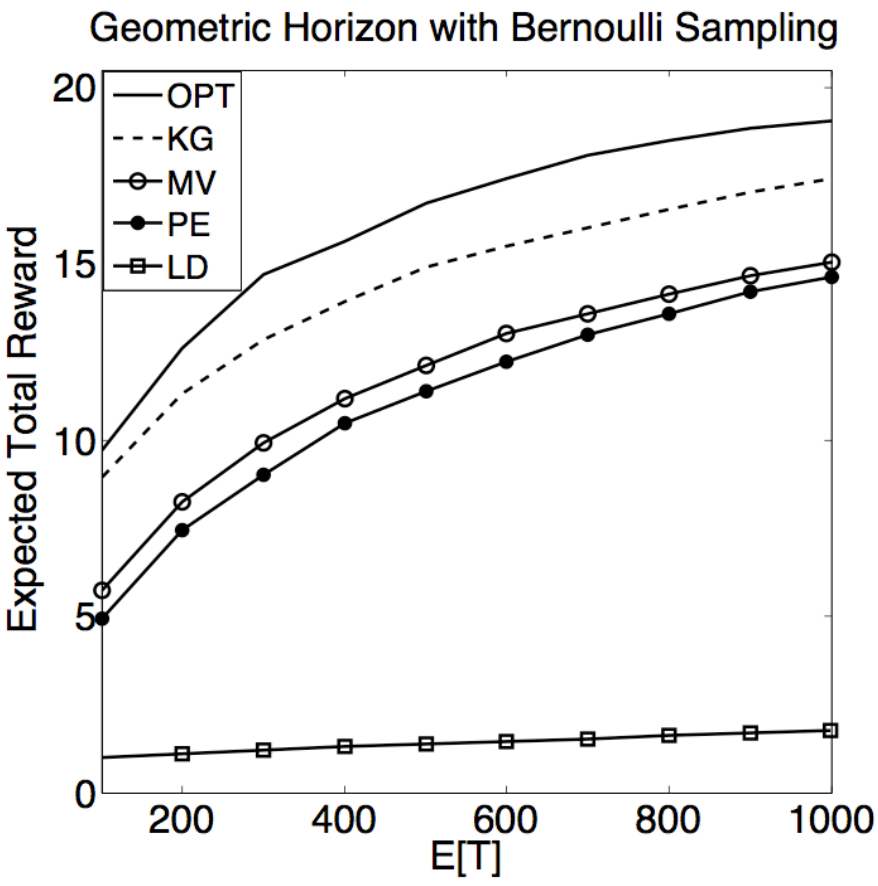
- The sup over π can be solved using the same decomposition technique as $\alpha = 1$. λ plays the role of c.
- Find λ that causes $\mathbb{E}^{\pi}[\tau] = T$ under the π attaing the supremum.
- The resulting solution gives an upper bound on the un-relaxed (fixed deadline) problem, and a heuristic policy.

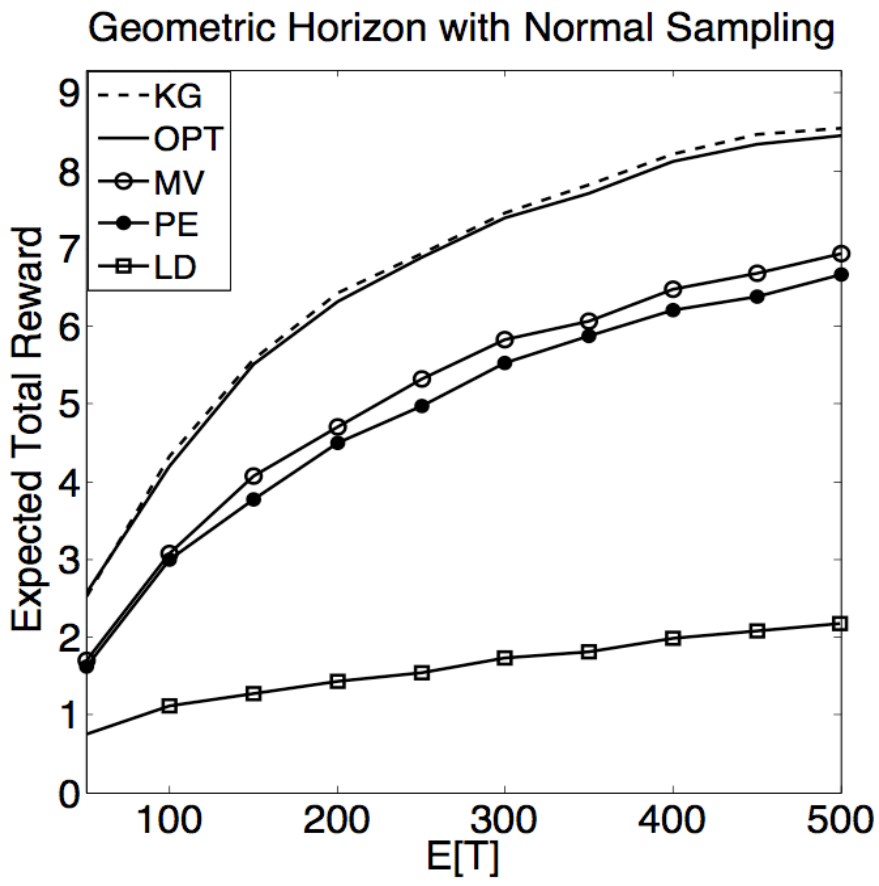
(The above takes $c = 0$. We can do something similar if $c > 0$.)

Other Generalizations are Possible

- Rather $\text{Normal}(\mu_x, \sigma_x^2)$ samples with a normal prior on μ_x and σ_x^2 known, the same method works when the sampling distribution is from an exponential family, and we have the conjugate prior.
  - e.g., normal samples with unknown mean and unknown variance, with a Normal/inverse-gamma prior.
  - e.g., Bernoulli samples with a Beta prior.
- Rather than taking our reward to be the number of correct classifications, we can have more complex functions of the true sampling distribution and the classification decision. Examples:
  - e.g., Reward of θ_x when we classify θ_x as positive, and a reward of 0 when we classify θ_x as negative.
  - e.g., Pay a penalty of a when we incorrectly classify an x with $\theta_x < 0$, and a penalty of b when we incorrectly classify an x with $\theta_x > 0$.

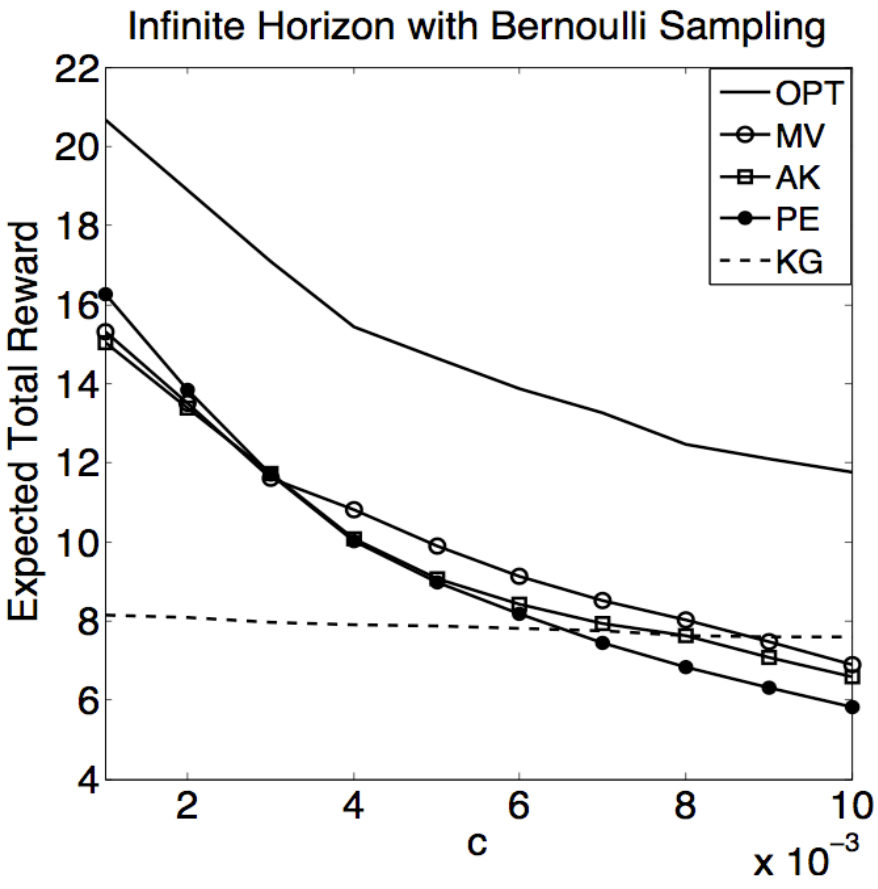
Numerical Results: Idealized Test Problems

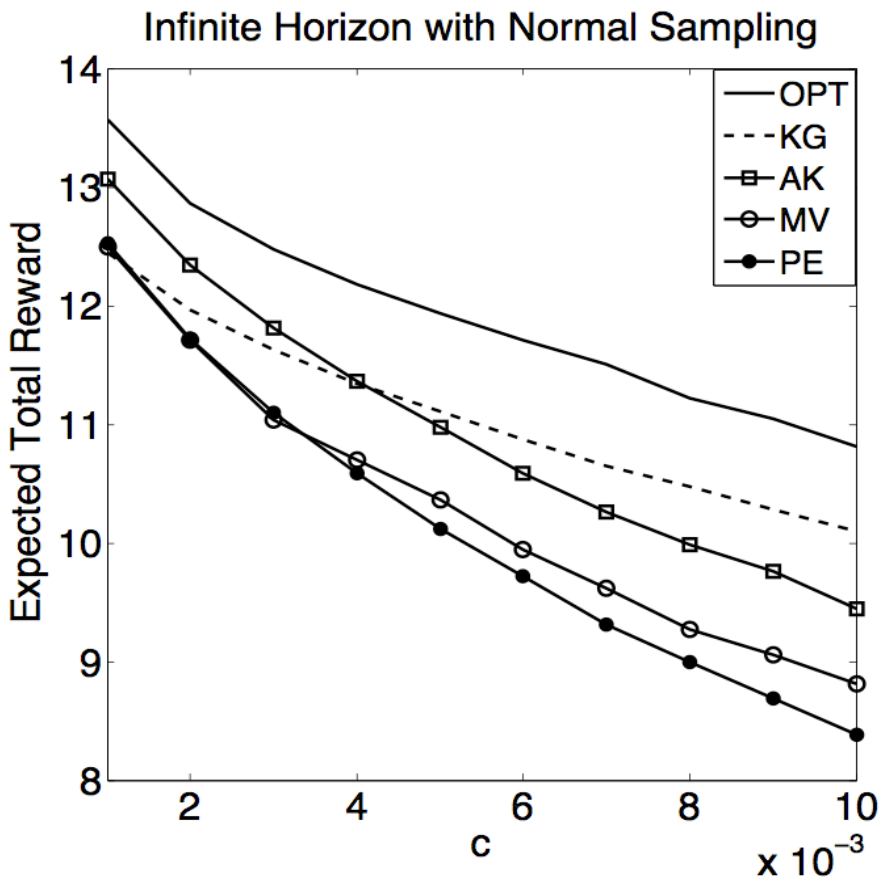




- Here, costs are 0, and we have a geometric external deadline.
- Comparison policies: Pure Exploration (PE), Max Variance (MV), Large Deviations (LD), Knowledge Gradient (KG), and Bayes-Optimal (OPT).

Numerical Results: Idealized Test Problems





- Here, costs are strictly positive, and there is no external deadline.
- Comparison policies: Pure Exploration (PE), Max Variance (MV), Andradóttir and Kim (AK), Knowledge Gradient (KG), and Bayes-Optimal (OPT).

Conclusion: The Optimal Policy Saves Time

- The **Multiple Comparisons with a Control** problem appears in many different simulation applications.
- **We found the optimal method** for deciding where to sample.
- This allows accurately characterizing level sets more quickly and with fewer simulation samples.

Thank You

Any questions?

References I


Andradóttir, S., Goldsman, D., and Kim, S. (2005).
Finding the best in the presence of a stochastic constraint.
In *Simulation Conference, 2005 Proceedings of the Winter*, pages 7–pp. IEEE.


Andradóttir, S. and Kim, S. (2010).
Fully sequential procedures for comparing constrained systems via simulation.
Naval Research Logistics (NRL), 57(5):403–421.


Batur, D. and Kim, S. (2010).
Finding feasible systems in the presence of constraints on multiple performance measures.
ACM Transactions on Modeling and Computer Simulation (TOMACS), 20(3):13.

References II


Bechhofer, R. and Turnbull, B. (1978).
Two (k+1)-decision selection procedures for comparing k normal means with a specified standard.
Journal of the American Statistical Association, pages 385–392.


Gittins, J. C. and Jones, D. M. (1974).
A dynamic allocation index for the sequential design of experiments.
In Gani, J., editor, *Progress in Statistics*, pages 241–266, Amsterdam. North-Holland.


Healey, C., Andradóttir, S., and Kim, S. (2012).
Selection procedures for simulations with multiple constraints.
in review.

References III


Hunter, S. and Pasupathy, R. (2010).
Large-deviation sampling laws for constrained simulation optimization on finite sets.
In *Simulation Conference (WSC), Proceedings of the 2010 Winter*, pages 995–1002. IEEE.


Hunter, S. and Pasupathy, R. (2012).
Optimal sampling laws for stochastically constrained simulation optimization on finite sets.
INFORMS Journal on Computing.


Kim, S. (2005).
Comparison with a standard via fully sequential procedures.
ACM Transactions on Modeling and Computer Simulation (TOMACS), 15(2):155–174.

References IV


Nelson, B. and Goldsman, D. (2001).
Comparisons with a standard in simulation experiments.
Management Science, 47(3):449–463.

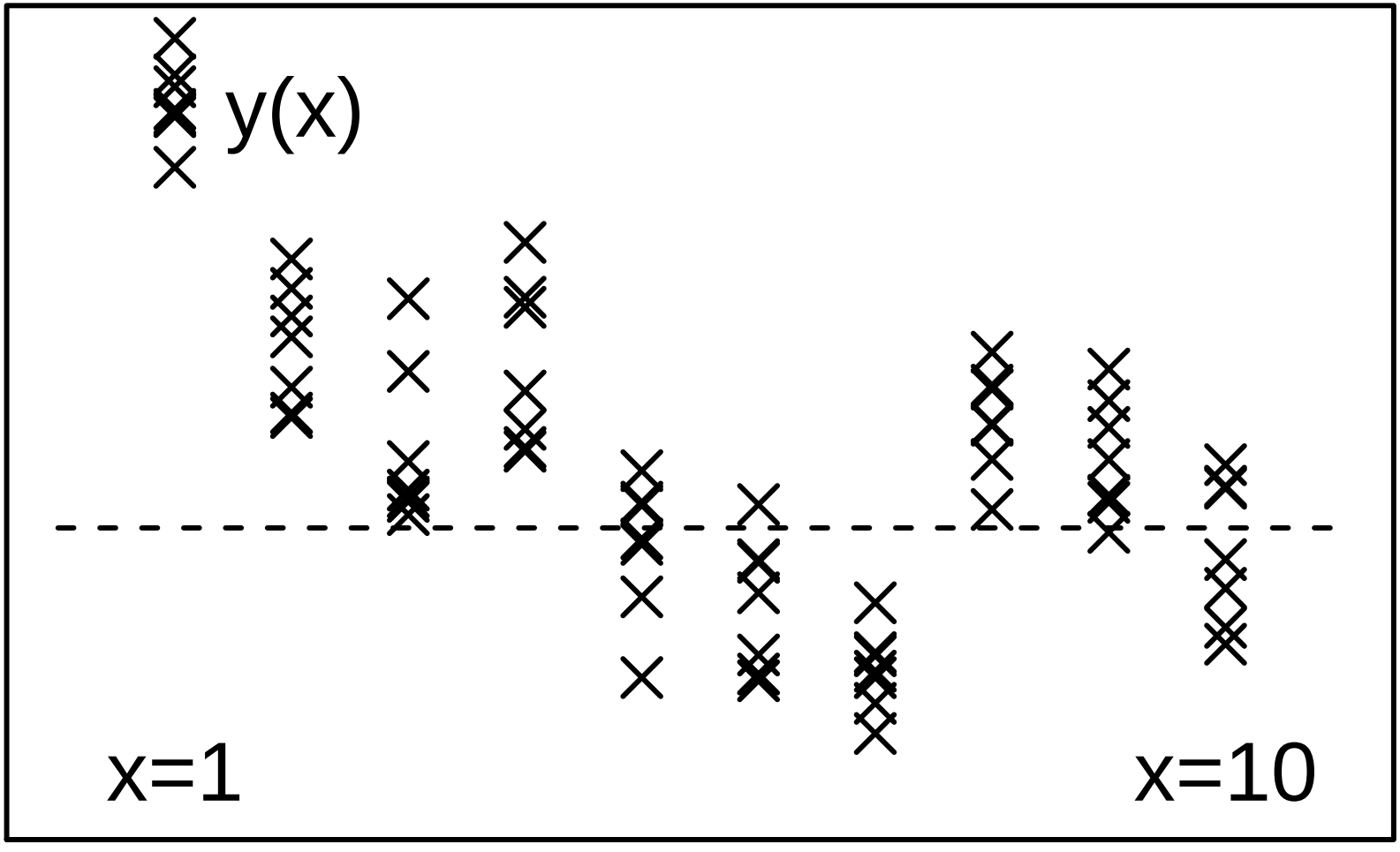

Paulson, E. (1962).
A sequential procedure for comparing several experimental categories with a standard or control.
The Annals of mathematical statistics, pages 438–443.


Szechtman, R. and Yücesan, E. (2008).
A new perspective on feasibility determination.
In *Proceedings of the 40th Conference on Winter Simulation*, pages 273–280. Winter Simulation Conference.

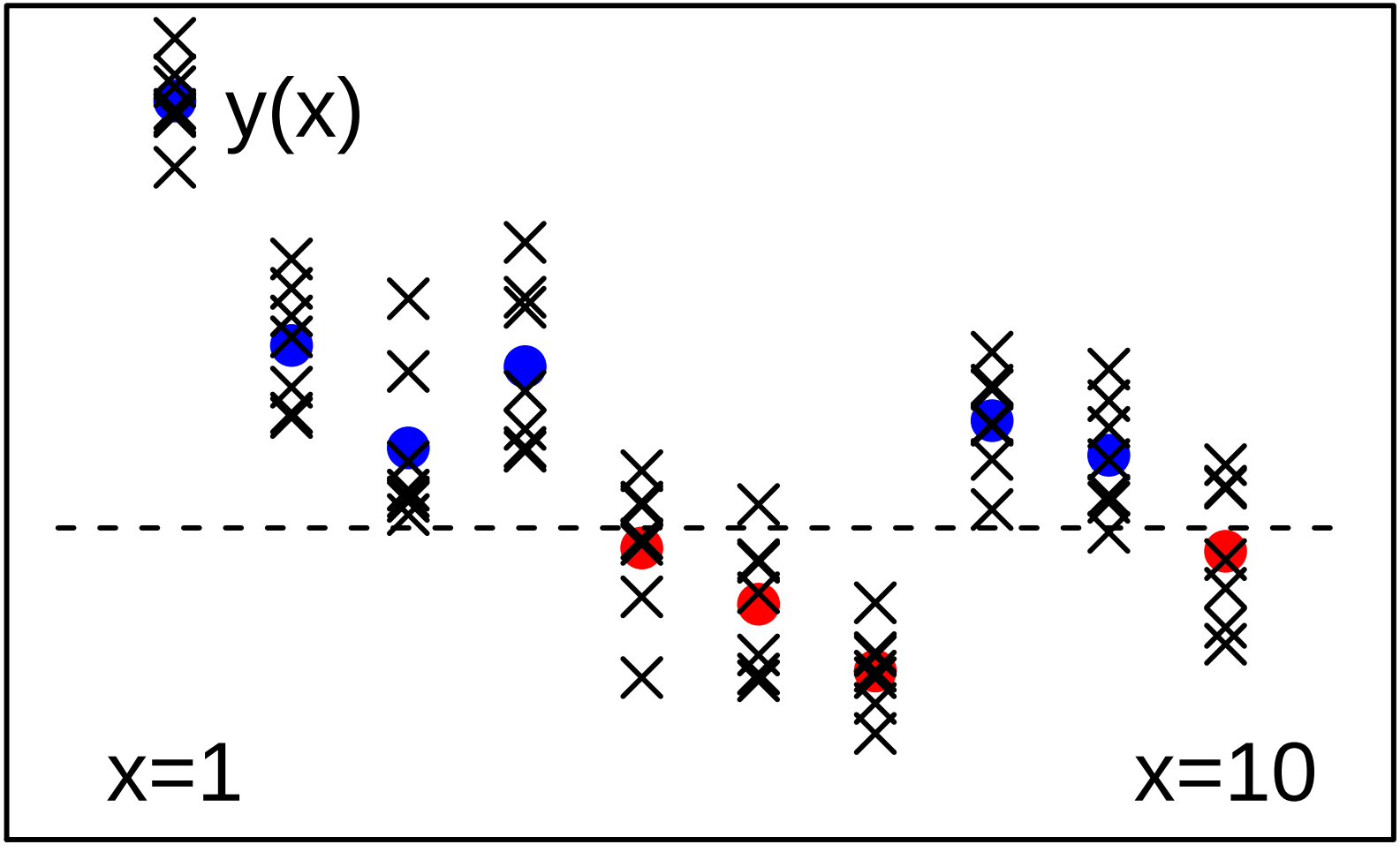
The Optimal Policy Maximizes Average-Case Accuracy

- A policy π is a rule for choosing whether and where to sample next, based on previous observations.
- Let $\tau \leq T$ be the number of samples taken (either equal to the external deadline, or can be strictly less if we choose to stop early).
- Let R be the number of alternatives classified correctly when sampling stops.
- $\mathbb{E}^{\pi}[R - c\tau | \vec{\mu}]$ is the performance under true mean vector $\vec{\mu}$.
- $\int \mathbb{E}^{\pi}[R - c\tau | \vec{\mu}] P(d\vec{\mu}) = \mathbb{E}^{\pi}[R]$ is the Bayes- or average-case performance, where P is the prior.
- We wish to find the policy that maximizes this.

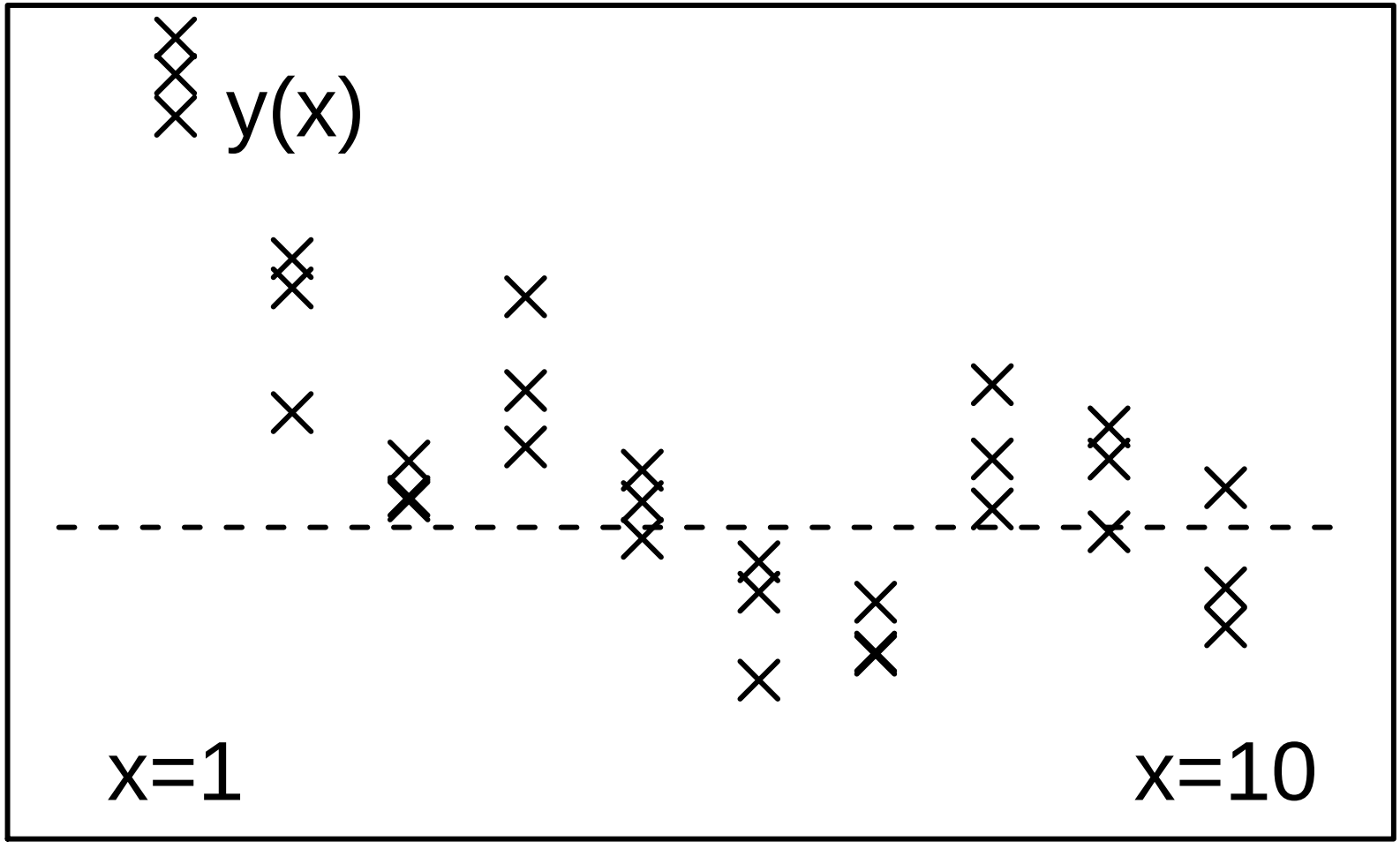
Given Samples, Estimating the Level Set is Well-Understood



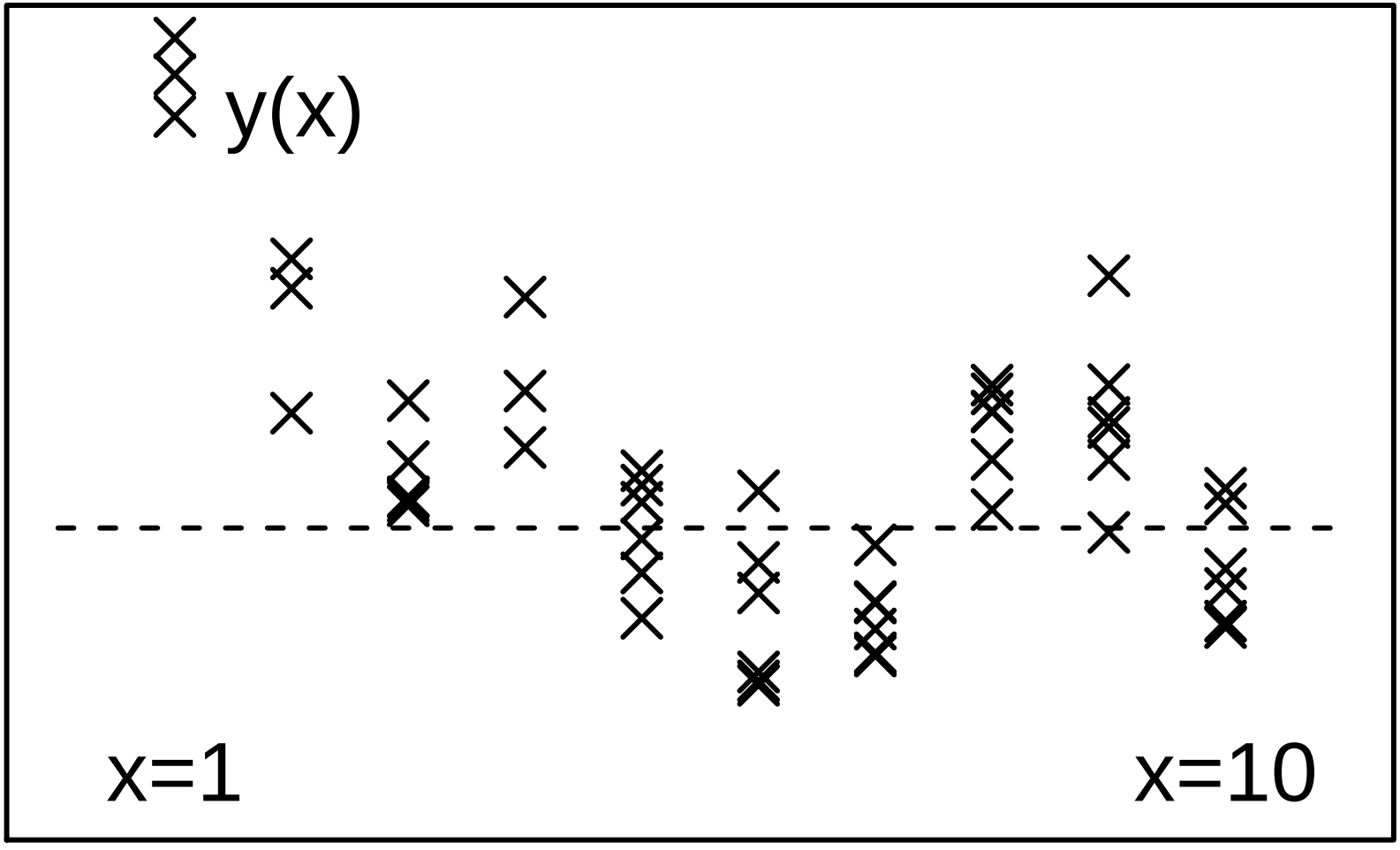
Given Samples, Estimating the Level Set is Well-Understood



The Optimal Policy Puts Samples Where They Help Most



The Optimal Policy Puts Samples Where They Help Most



The Optimal Policy Puts Samples Where They Help Most