Optimal Sequential Experimental Design for Stochastic Root-finding in Drug Development

Peter I. Frazier[1] Zachary Owen[1] Rodrigo C. Bicalho[2]
Thiago M.A. Santos[2] André G.V. Teixera [2]

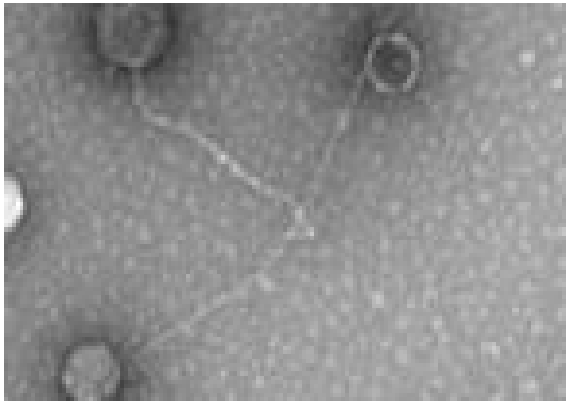
[1]School of Operations Research & Information Engineering, Cornell University

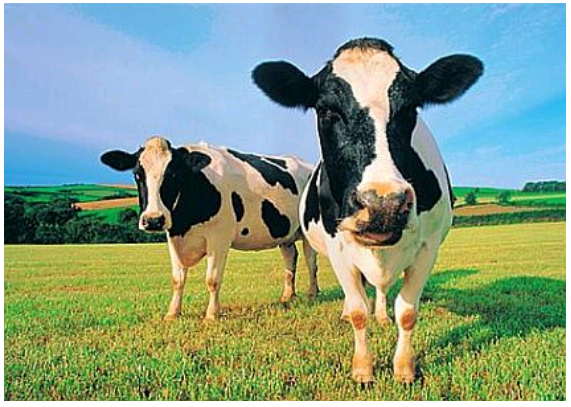
[2]College of Veterinary Medicine, Cornell University

Tuesday June 21, 2011
INFORMS Healthcare Conference
Montreal, Quebec

Research supported by the Atkinson Center for a Sustainable Future

The Disease: Metritis



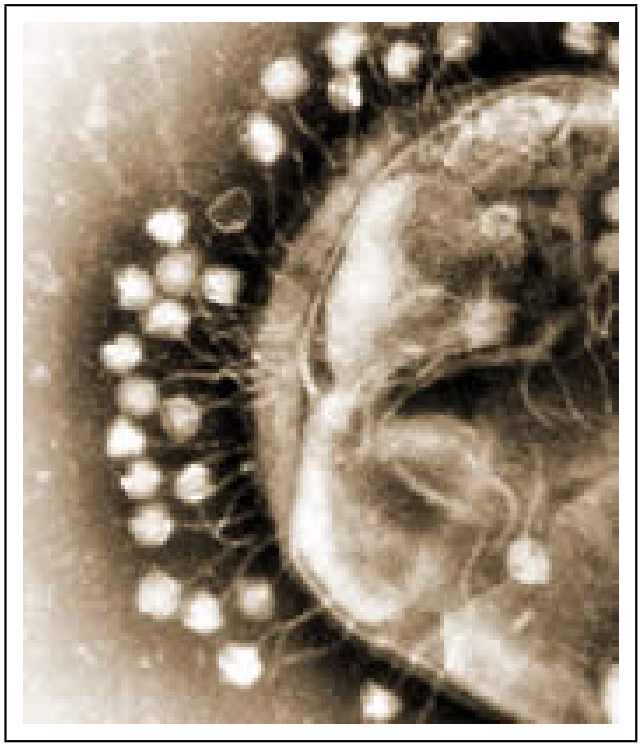


- Metritis is a bacterial infection of the uterus.
  - It is a leading cause of loss of milk production and fertility in dairy cows.
  - It also occurs in other animals and people.
- The treatments developed may also be useful for other problematic bacterial infections, in both animals and people (e.g. MRSA).

Current Treatment: Broad-Spectrum Antibiotics

- Metritis is caused primarily by *E. coli* and *A. pyogenes* bacteria.
- Antibiotics are used to treat metritis in sick cows.
- Antibiotics are also given to well cows as a preventative measure.
- Causes to be concerned about the indiscriminate use of broad-spectrum antibiotics:
  1. Bacteria develop **resistance to antibiotics** if they are used too widley.
  2. Releasing large quantities of antibiotics into the environment via farm effluent may have negative **environmental effects**.

An Alternative Treatment: Bacteriophages



- Bacteriophages (abbreviated phages) are **viruses that kill bacteria**.
- Can we treat bacterial infections with **phages instead of antibiotics**?

Bacteriophage-based Treatments: Advantages

- Reduced risk of **bacterial resistance**
  - Phages are a new type of treatment.
  - Each phage would be used against a few bacteria, limiting its use.
  - Increasing the number of available treatments, and limiting their use, mitigates the problem of bacterial resistance.
- Reduced **environmental impact**:
  - Each phage kills a few very specific strains of bacteria, and nothing else. In contrast, each type of antibiotic kills a wide variety of bacteria.
  - Phages already exist naturally at dairy farms.

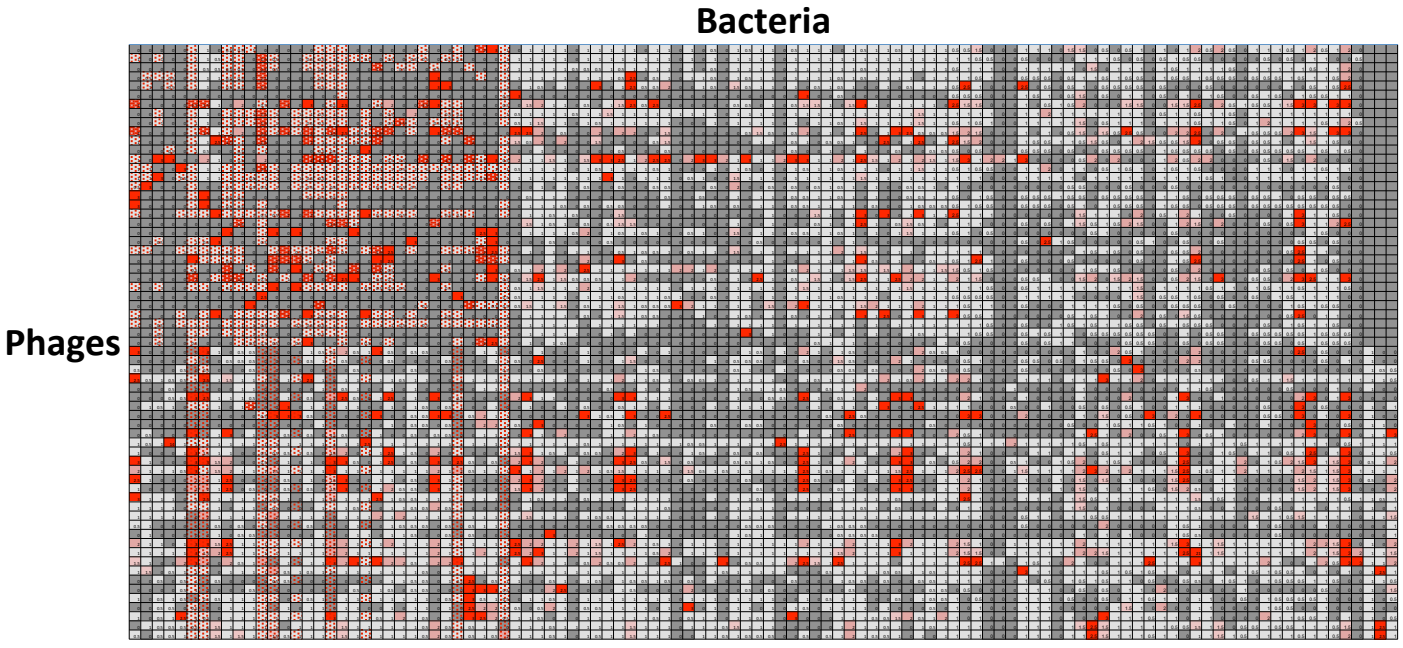
Specificity and Cocktails

- Specificity:
  - Each phage kills only a few very specific strains of bacteria.
  - An infection could be caused by any one of a number of strains of bacteria.
  - An effective treatment must then be a cocktail of phages that will be effective against each common metritis-causing strains of bacteria.
- A **cocktail** is a collection of phages.
  - We also specify a concentration (# phages/mL) for each phage in the cocktail.

Production Costs

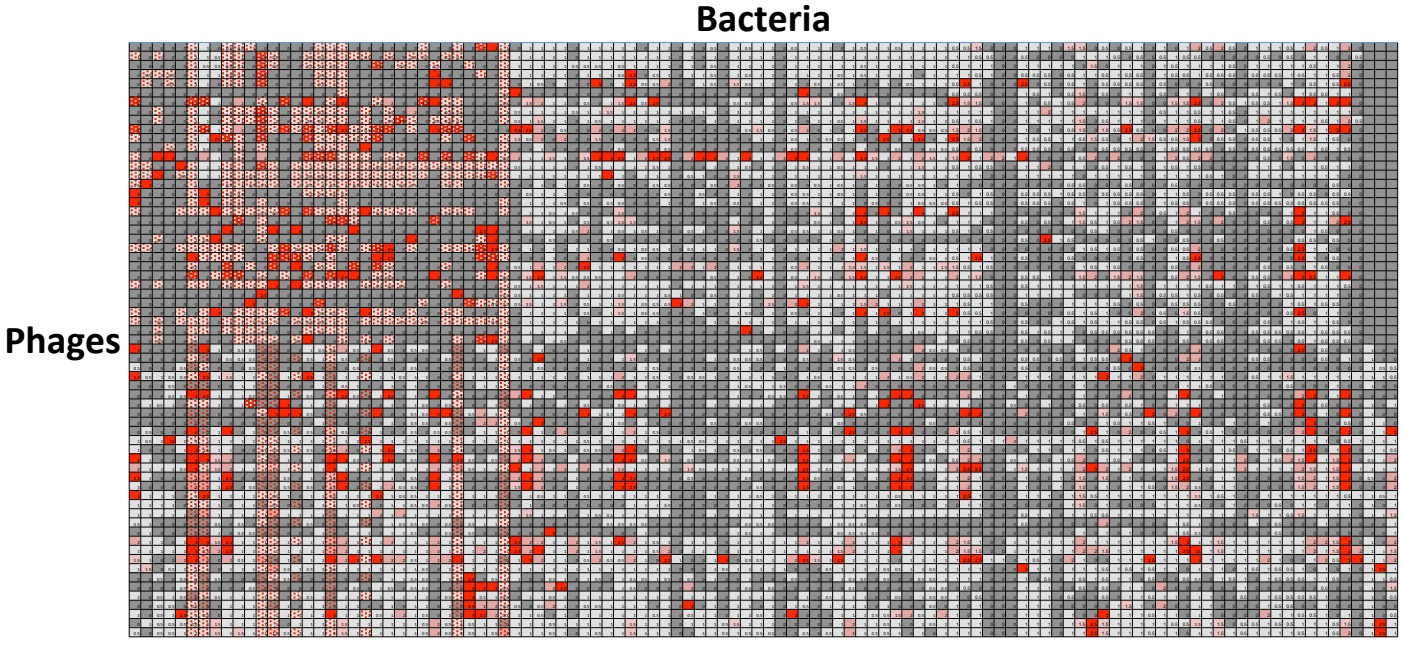
- A cocktail is a collection of phages,
  with a concentration specified for each phage.
- Production costs
  - A cocktail is cheaper to produce if it contains few phages.
  - A cocktail is cheaper to produce if the concentrations are small.
- **What is the cheapest cocktail that kills all targeted bacteria?**

The Matrix



- Given many lab experiments (**600 days of experiments**), we could create a matrix like the one above, but with accurate values.
- The color in each cell would give the minimum concentration of that phage needed to kill that bacteria.
  - Red cells: the phage kills the bacteria at very low concentrations.
  - White cells: the phage kills the bacteria at high concentrations.
  - Gray cells: the phage does not kill the bacteria at any concentration.

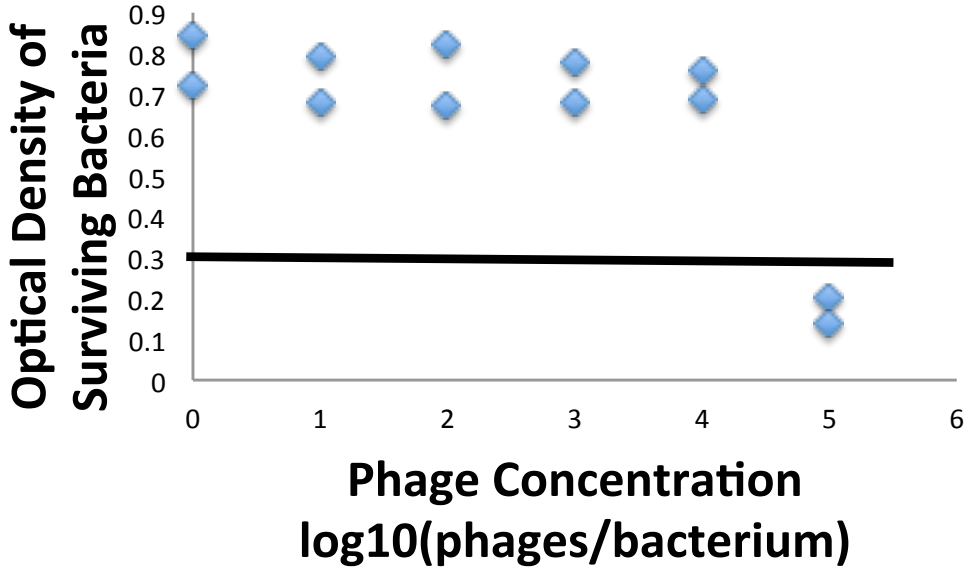
The Matrix



- With this matrix, we could find the cheapest cocktail that kills every bacteria.
  - (Actually, a partially filled matrix would be sufficient — This is a subject of ongoing work)
  - The matrix is also of independent scientific interest.
- The goal of this talk is to develop an efficient method to create this matrix.

Stochastic Root-finding

- Problem: Given a phage and a bacteria, find the **minimal** concentration of the phage that kills the bacteria.



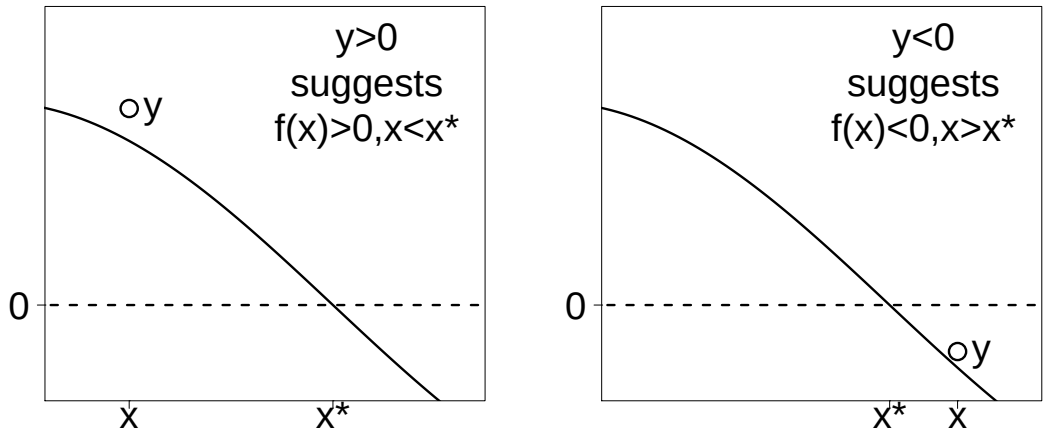
- Each experiment is time consuming (1 day), so we should **use as few experiments as possible** (< 10 experiments).
- We act **sequentially**, basing each new experiment on previous results. (Experiments are done in batches of 96 per day. To allow sequential decision-making, we consider 96 phage-bacteria pairs simultaneously.)

Stochastic Root-Finding

- The standard approach is stochastic approximation [Robbins and Monro, 1951].
- This approach requires many (20 or more) samples to be reliable.
- We can afford < 10 samples per phage-bacteria pair.
- We use methods from sequential Bayesian experimental design.

Stochastic Root-Finding: Motivation

Motivation for our approach:



- We think of $z = \mathrm{sgn}(y)$ as a noisy bit indicating whether x_* is left or right of x_n.
- When $|y|$ is big, this bit is more likely to be correct.

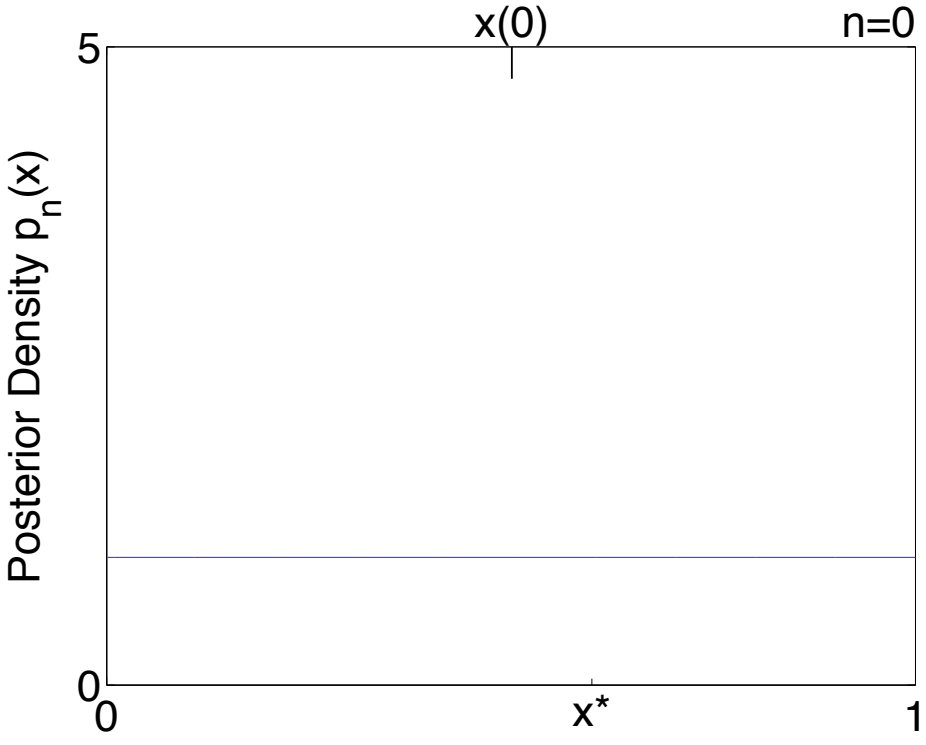
Stochastic Root-Finding: Mathematical Model

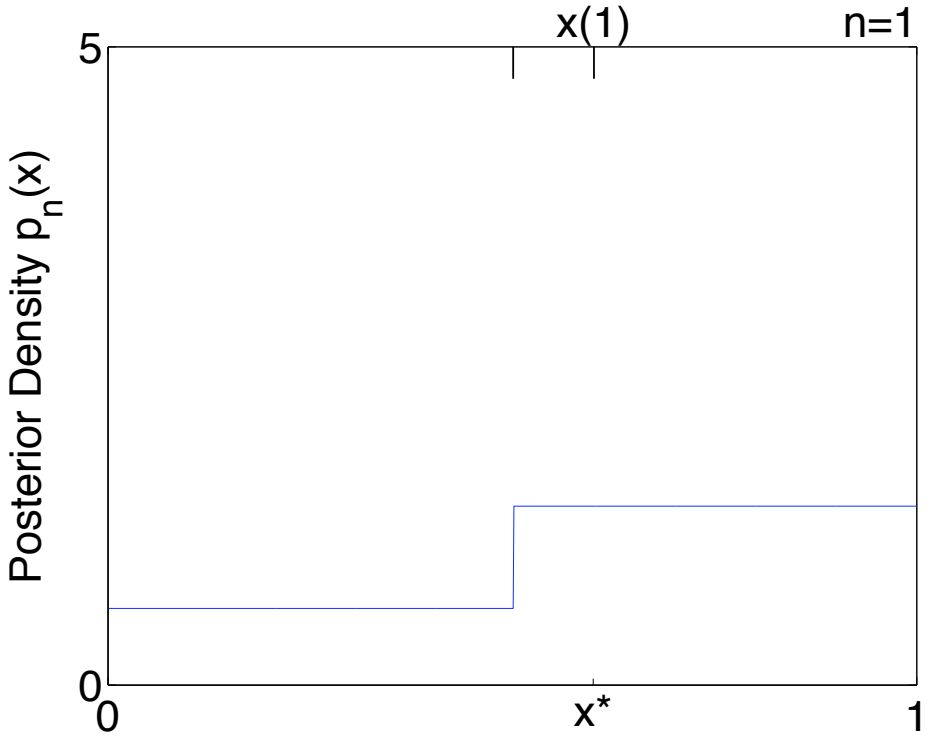
- $f : \mathbb{R} \mapsto \mathbb{R}$ is a decreasing function, with root x_*, so $f(x_*) = 0$.
- When we measure x_n we see a direction $z_n \in \{-1, +1\}$ and a probability r_n.
- The true direction, $\mathrm{sgn}(x_n - x_*)$, is equal to z_n with probability r_n. (Nature gives the right direction with probability r_n).
- We choose each x_n given the available information, $(x_m, z_m, r_m), m < n$.
- Central Question: Given a budget of N measurements, $x_1, \ldots, x_N$, how should we place them to find x_* as accurately as possible?

Posterior Distributions

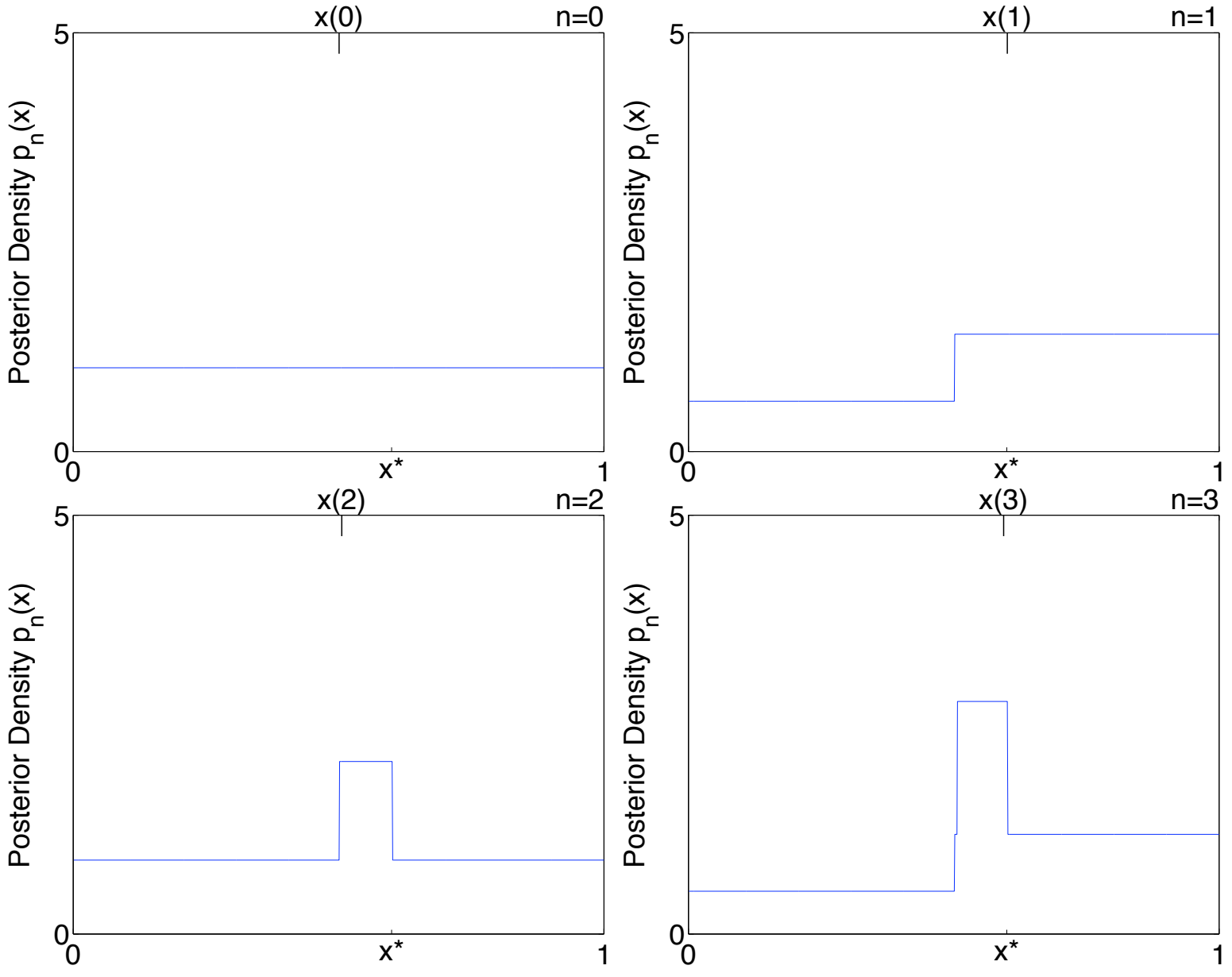
- Place a prior density p_0 on the root x_*, e.g., uniform on $[0,1]$.
- Each measurement x_n produces a new posterior density p_n on x_*:

$$p_n(x) = \mathbb{P}\{X_* \in dx \mid x_{1:n}, z_{1:n}, r_{1:n}\}$$





Posterior Distributions



Optimality and Dynamic Programming

- One measure of success is the entropy of the final posterior distribution,
$$H(p_N) = -\int p_N(x) \log p_N(x)\, dx.$$
- A policy π is a method for choosing the next experiment based on the results so far.
- The optimal policy π^* is the solution to the stochastic optimization problem
$$\inf_{\pi} \mathbb{E}^{\pi} \left[H(p_N) \right],$$
- Given a very large computer, we could compute π^* by **dynamic programming**.

Bayes Optimality

Theorem

If r_n is chosen independently, and its distribution does not depend on x_n, then **the policy that chooses x_n at the median of p_n is the Bayes optimal fully sequential policy** *for the entropy loss function.*

- The assumption on r_n is not met in our application, but the policy can still be used here.
- This policy was introduced by [Horstein, 1963], and is called probabilistic bisection.
- For r_n constant, the proof is due to [Waeber, Frazier, Henderson 2011]. For more general r_n, the proof follows [Jedynak, Frazier, Sznitman 2011].

Estimating the Probability of Correctness, r_n

- In practice we observe y_n, not r_n.
- One can use training data to fit a parametric statistical model,

$$r_n \approx g(y_n, \beta),$$

  where β is one or more parameters to be fit.
- In experimental results to be shown later, we use $g(y, \beta) = \Phi(\beta|y|)$,

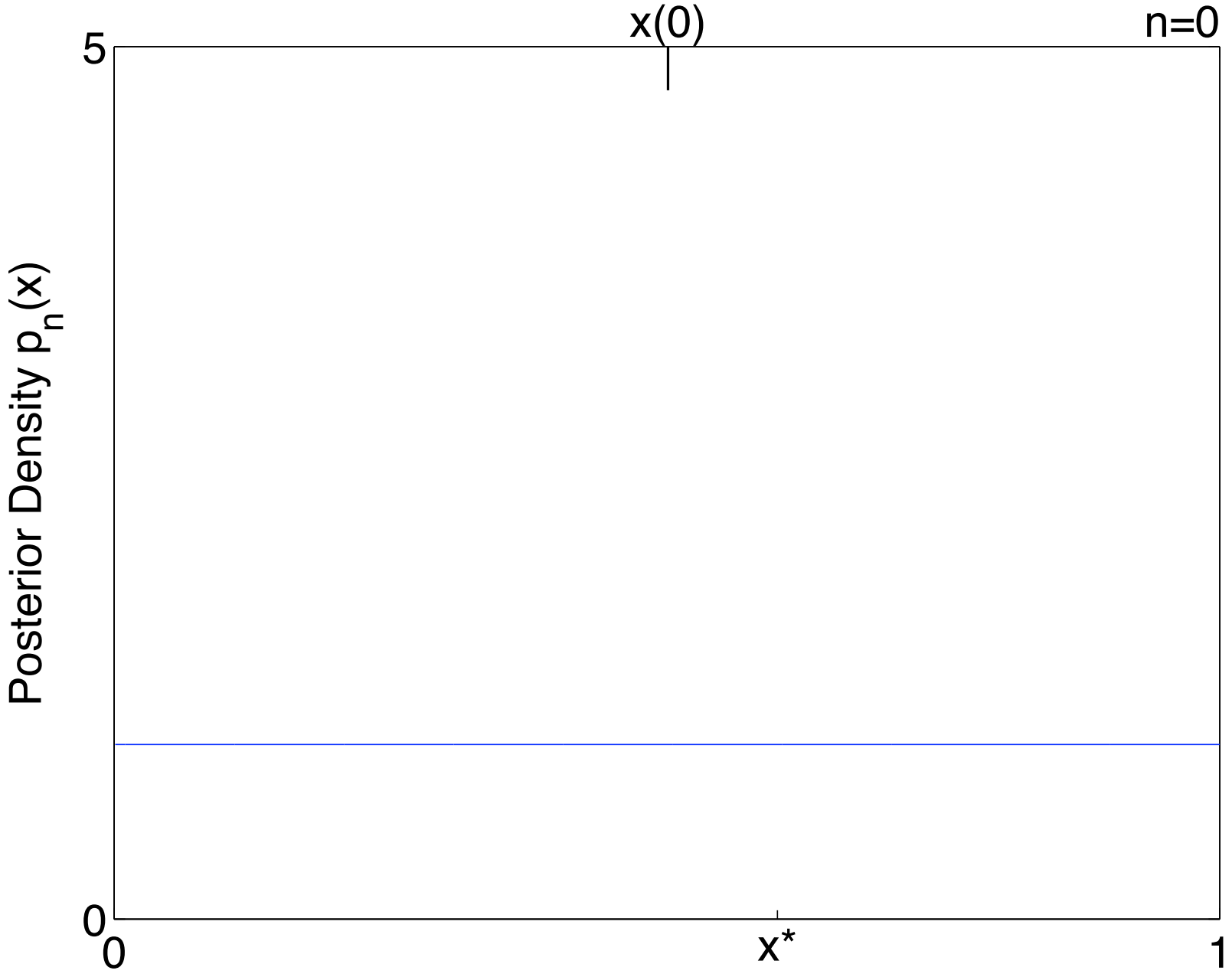
$$r_n \approx \Phi(\beta|y_n|).$$

- For applications allowing more measurements, we could include the whole history $x_{1:n}$, $y_{1:n}$ into the fit to obtain asymptotically consistent estimates of r_n,

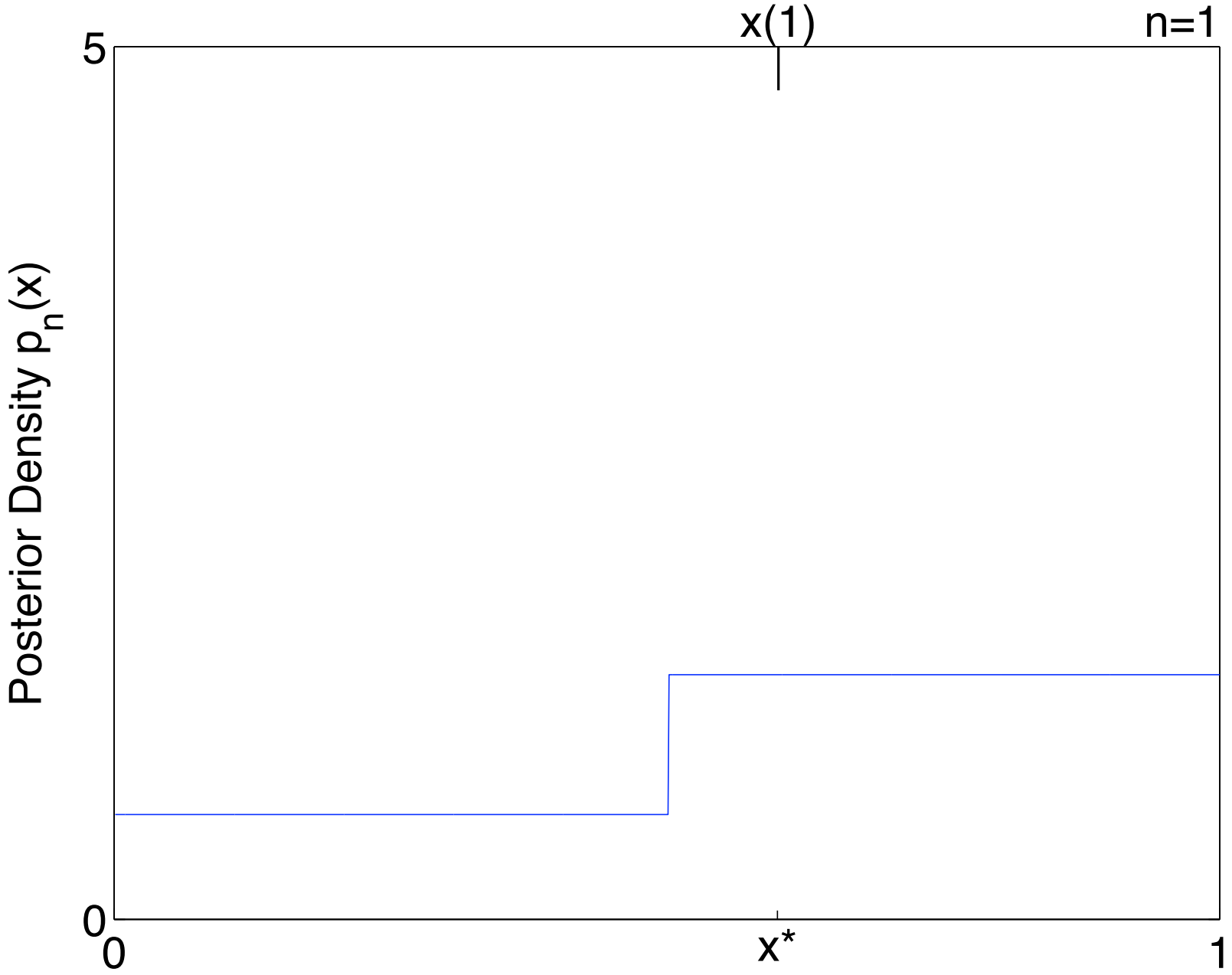
$$r_n \approx g(x_{1:n}, y_{1:n}, \beta),$$

  Open question: Would the resulting procedure provide an asymptotically consistent estimator of x_*?

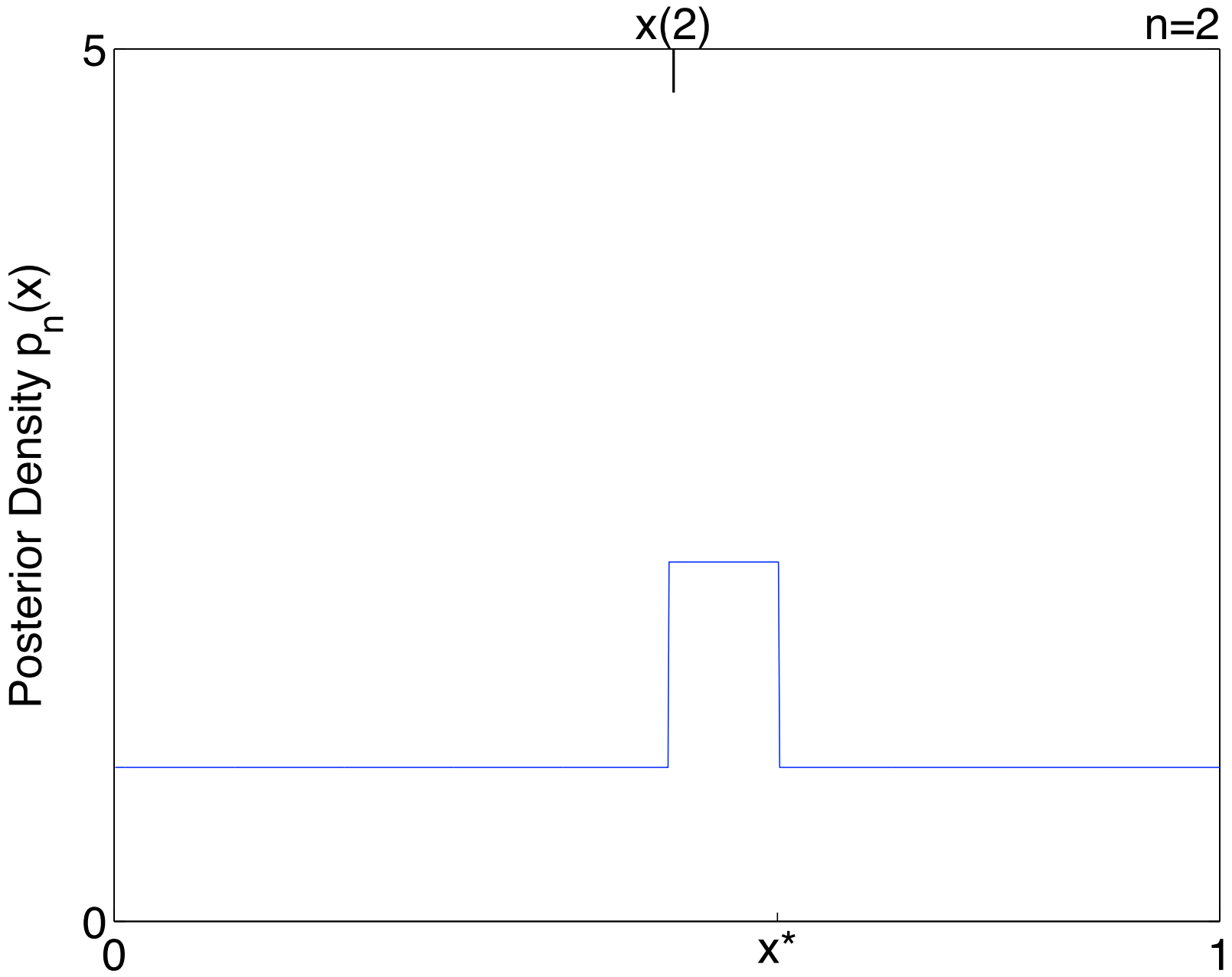
Example



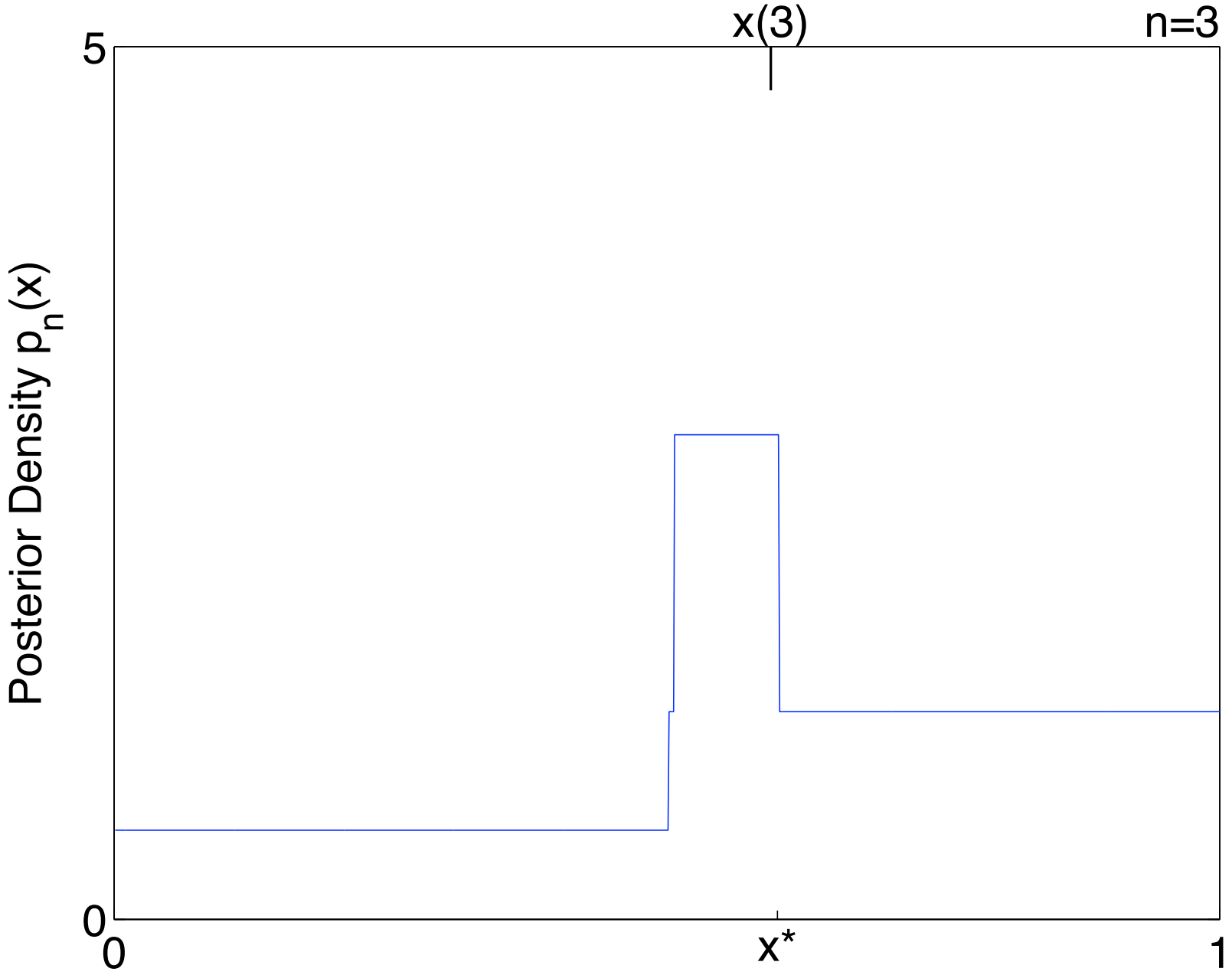
Example



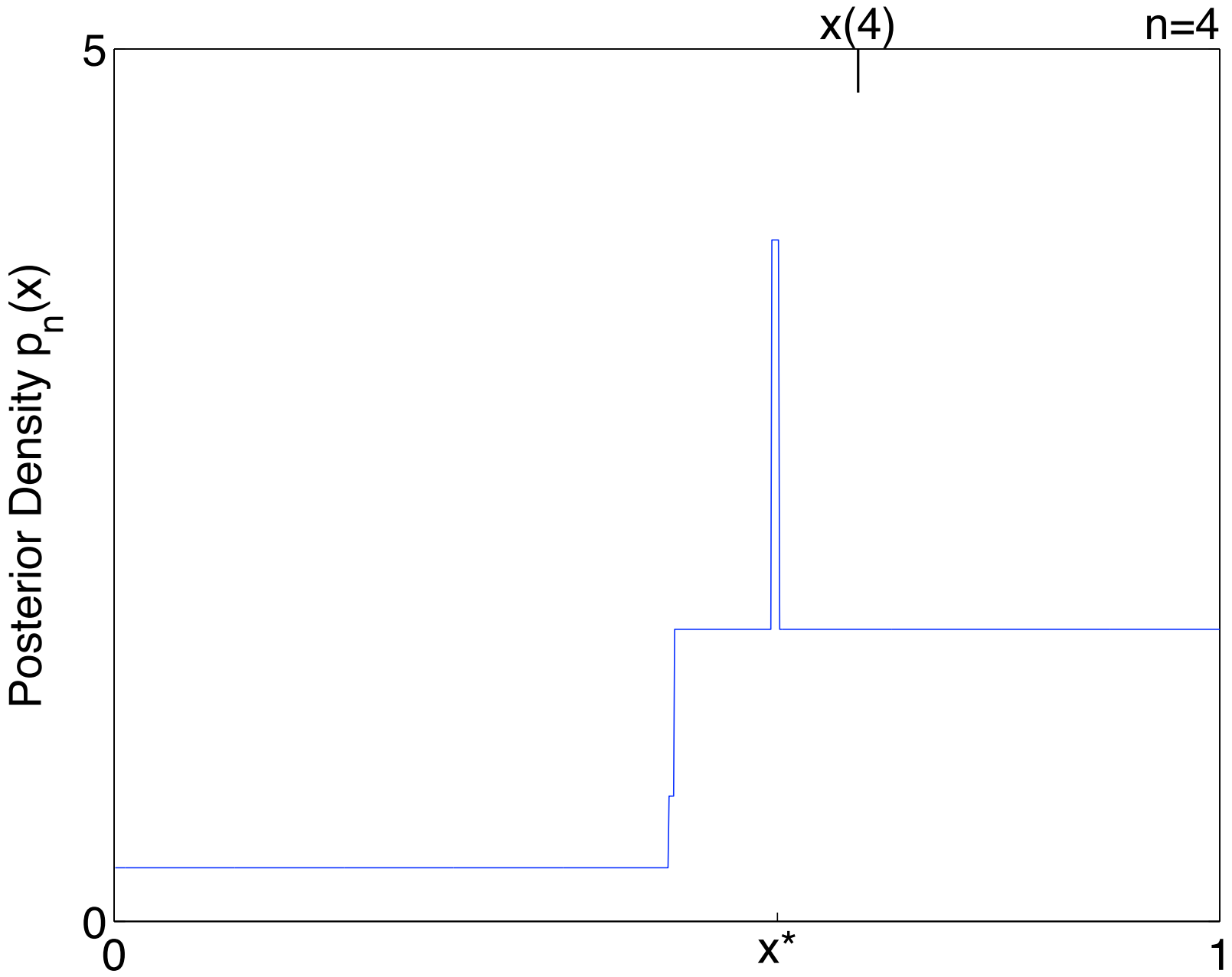
Example



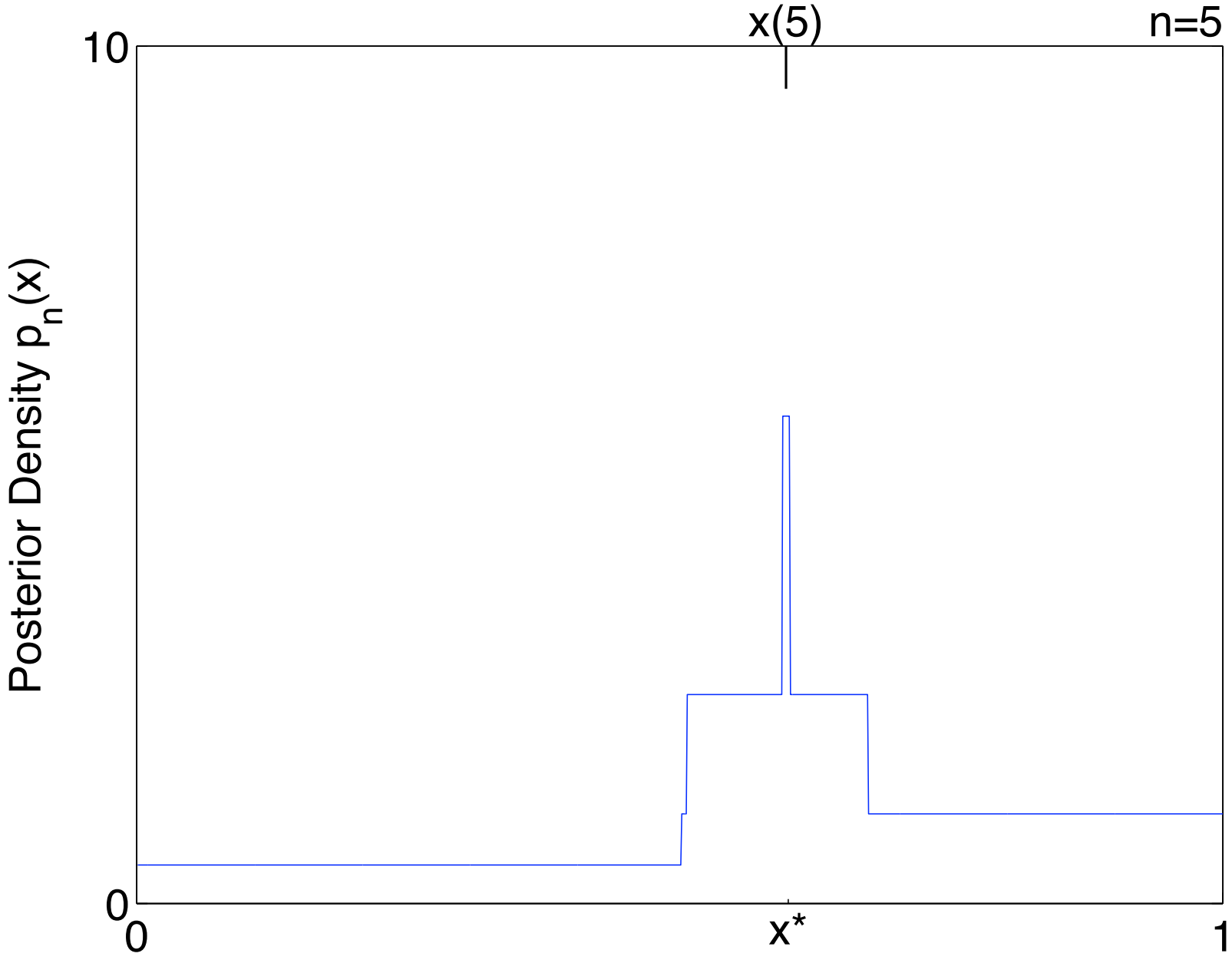
Example



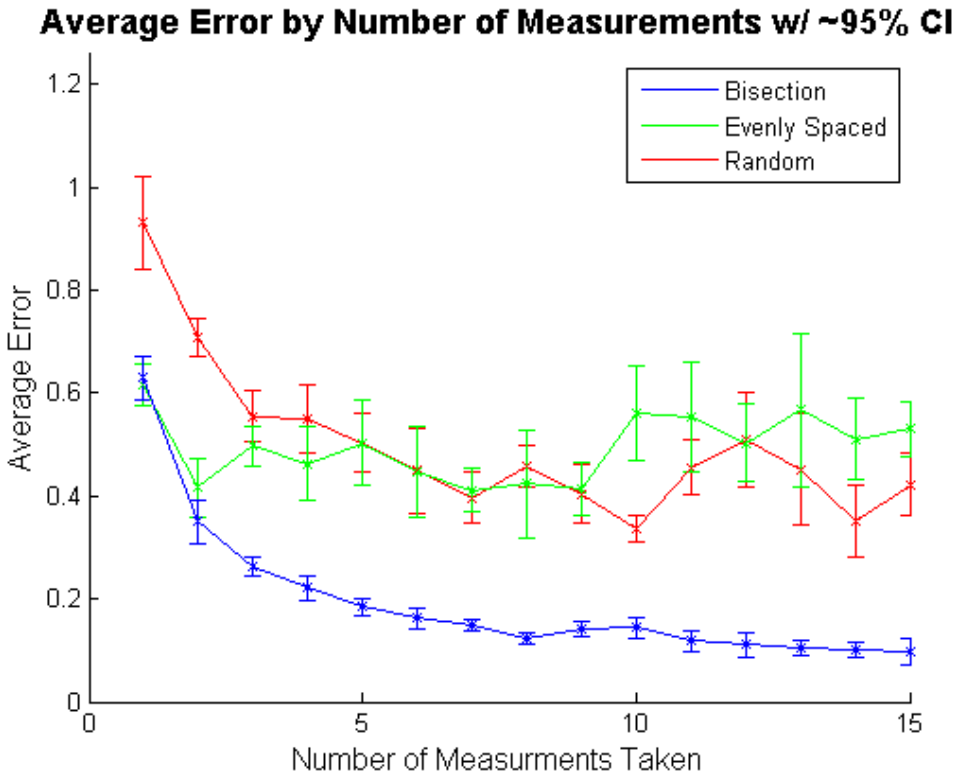
Example



Example



Experimental Results



- Caveat: graph presents initial results based on a small amount of experimental data.
- These results suggest we can use 3 experiments per phage-bacteria pair instead of 6 (currently used), and get better accuracy.
- This would the reduce time required to **300 days** down from **600.**

Ongoing work: Bayesian global optimization

- We can use structure in the phage-bacteria matrix for prediction.
  - We use low rank matrix approximation, as in collaborative filtering.
  - Results so far suggest that bacteria and phages are described by a small number of features.
- Finding a good cocktail with as few measurements as possible is a Bayesian global optimization problem.
  - We do not need to evaluate every cell in the matrix to find a good cocktail.
  - We are planning to use methods based on the value of information, e.g., knowledge-gradient methods, or expected improvement methods.

Conclusion

- We are using fully sequential Bayesian learning methods to make drug development for bacteriophage-based therapies **faster**, **cheaper**, and **more likely to succeed**.
- **Dynamic programming** is an excellent tool for analyzing and developing these methods.
- **Ongoing work** in implementing this approach, analysis of the algorithm, and in Bayesian global optimization.

We gratefully acknowledge support from the Atkinson Center for a Sustainable Future.

References


Horstein, M. (1963).
Sequential decoding using noiseless feedback.
IEEE Trans. Info. Theory, 9(3):136–143.


Jedynak, B., Frazier, P. I., and Sznitman, R. (2011).
Twenty questions with noise: Bayes optimal policies for entropy loss.
submitted.


Robbins, H. and Monro, S. (1951).
A stochastic approximation method.
Annals of Math. Stat., 22:400–407.


Waeber, R., Frazier, P., and Henderson, S. (2011a).
A bayesian approach to stochastic root finding.
In *Proceedings of the Winter Simulation Conference, 2011*. IEEE.


Waeber, R., Frazier, P. I., and Henderson, S. G. (2011b).
Optimal Entropic Bisection Search in a Noisy Environment.
Technical report, Cornell University.

Thank you for coming!