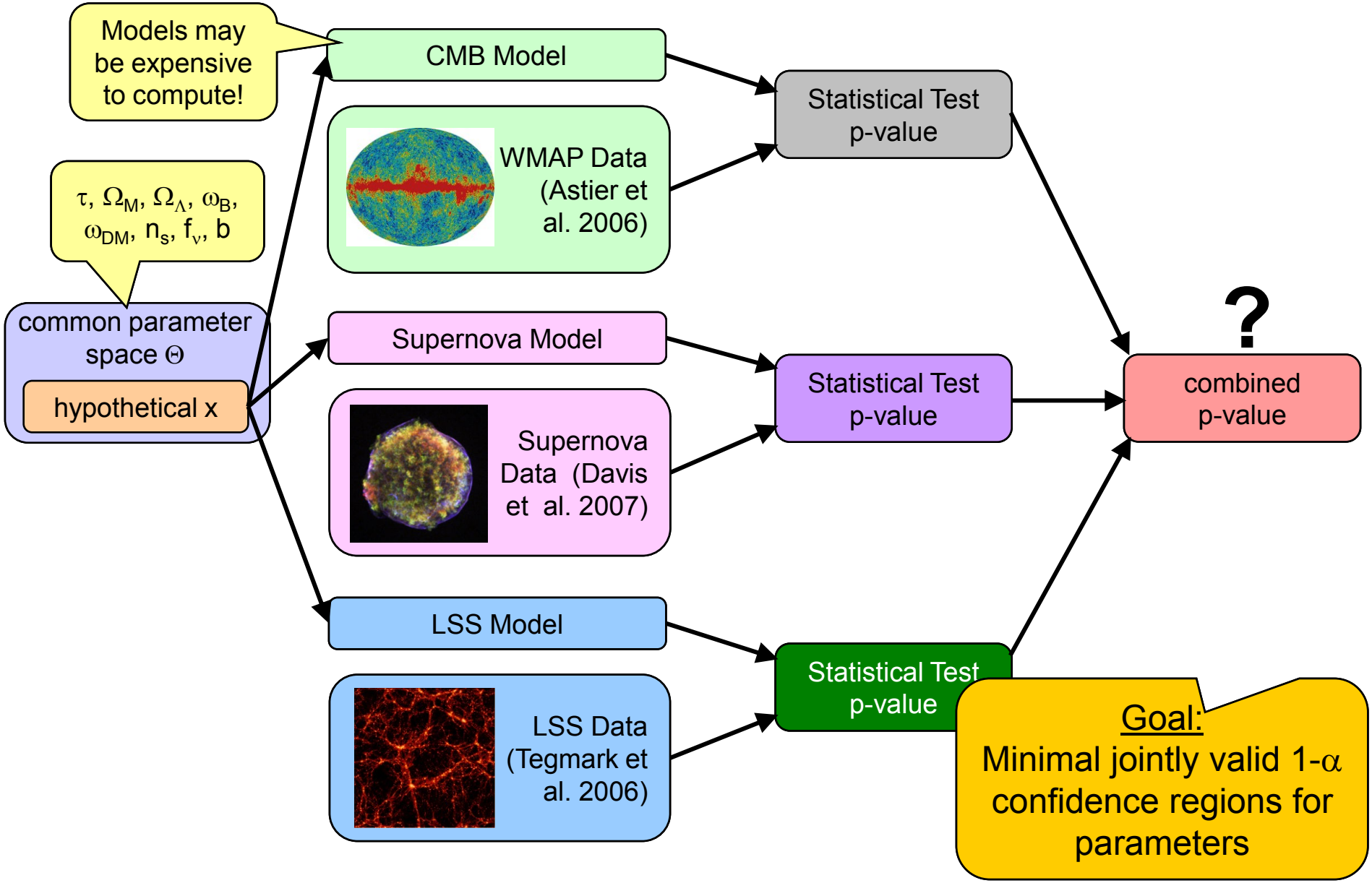
Actively Learning Level-Sets of Composite Functions

Brent Bryan, Jeff Schneider

Motivation: Statistical Analysis



Motivation: Statistical Analysis

- Many ways to combine p-values:
  - Bonferronoi's method, inverse normal, inverse logit

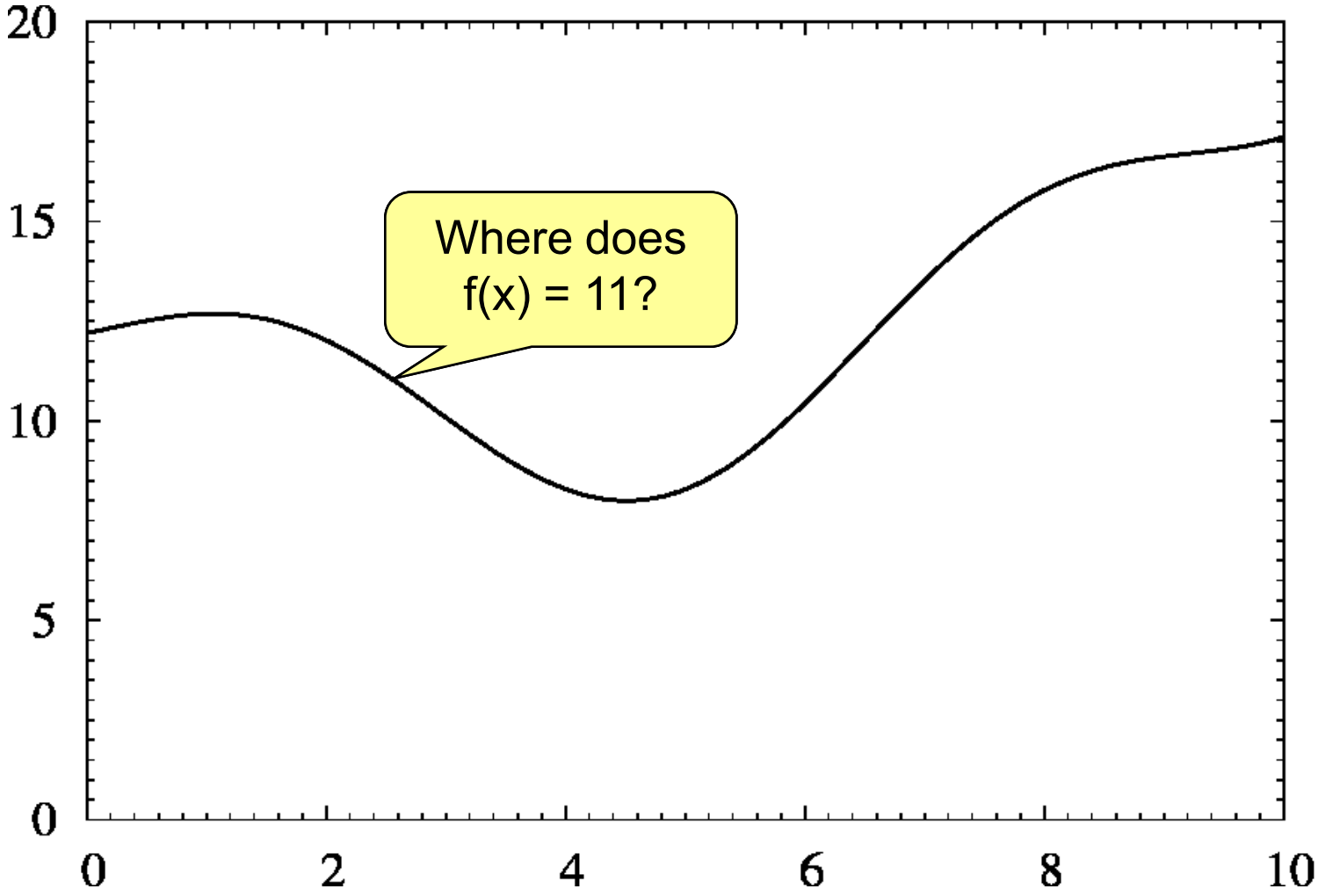
- Fisher's Method (Fisher 1932):

$$-2\sum_{i=1}^{m}\log(p_i(x)) \geq C$$

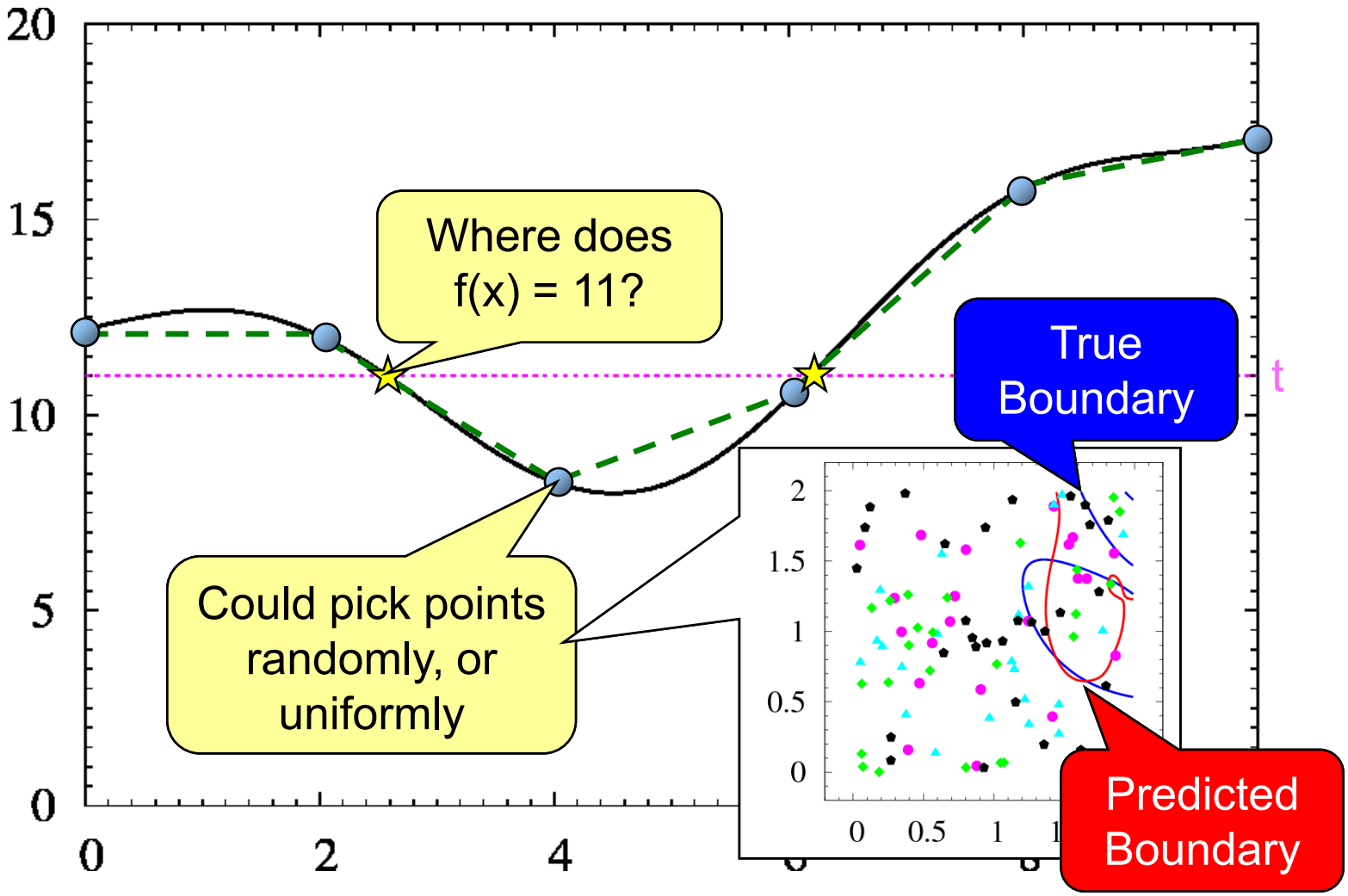
where C is the critical value of a χ^2_{2m} distribution

- $f_i(x) = -2\log(p_i(x))$

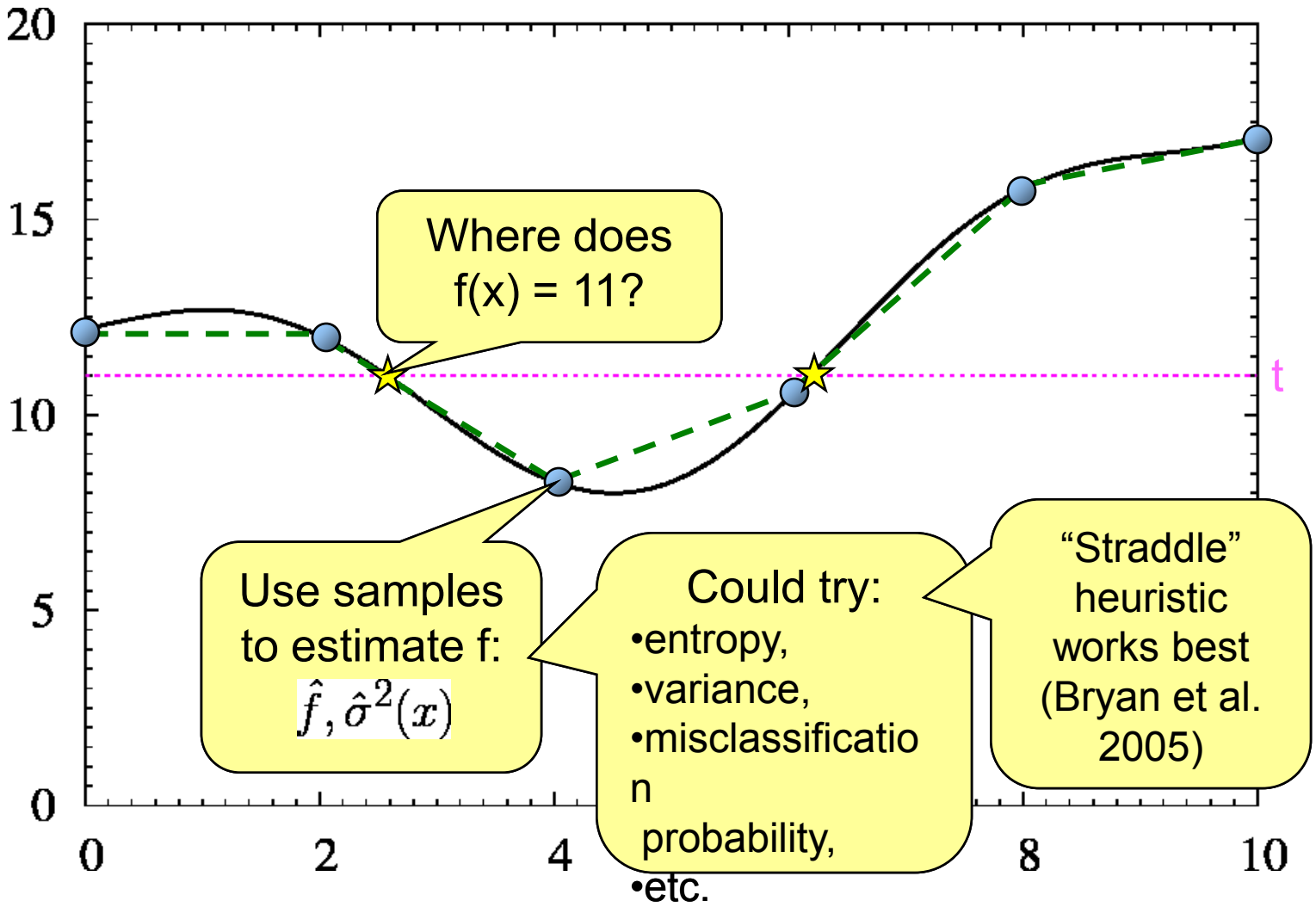
The Level-Set Problem



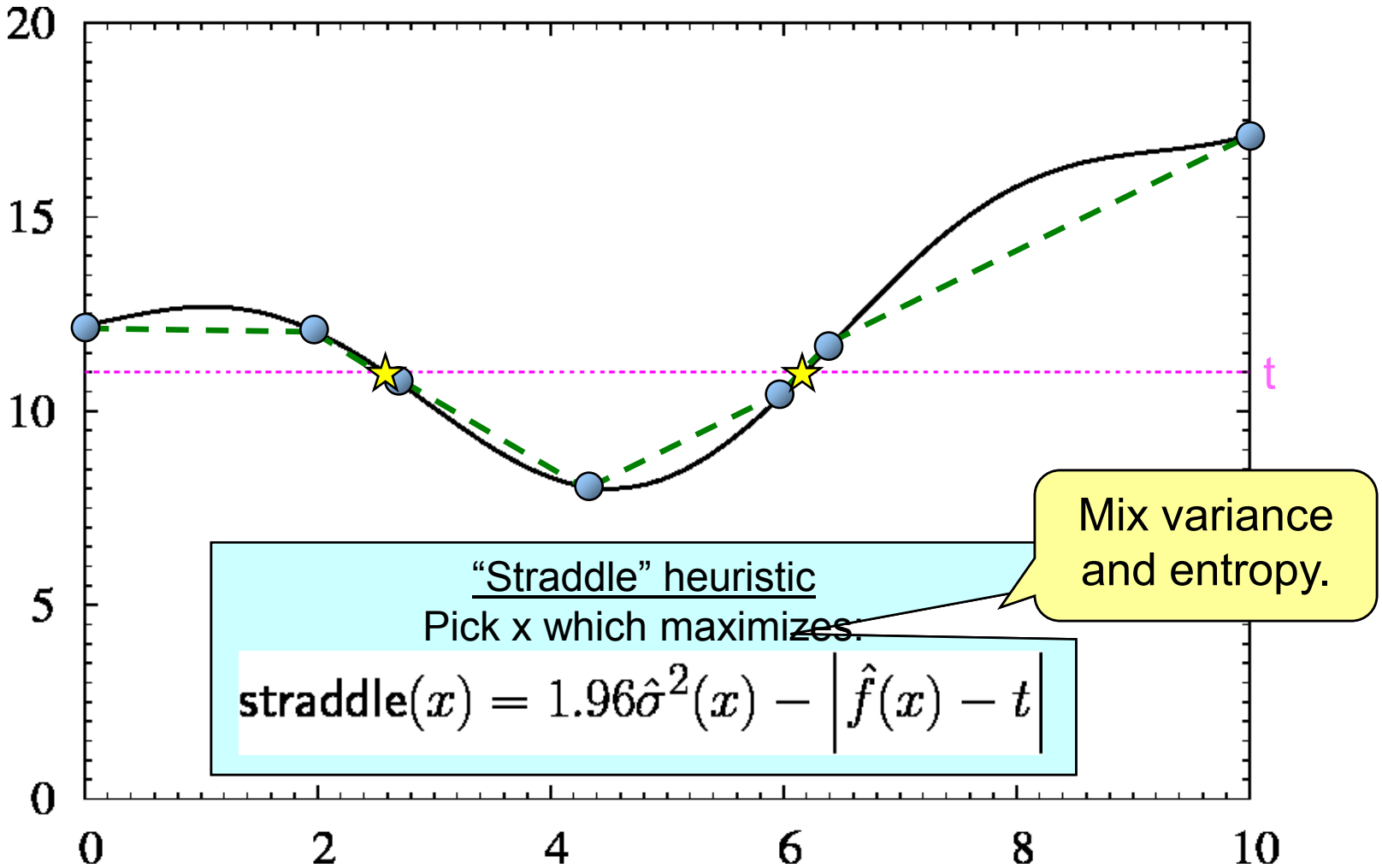
The Level-Set Problem



The Level-Set Problem



The Level-Set Problem



"Straddle" heuristic

Pick x which maximizes:

$$\text{straddle}(x) = 1.96\hat{\sigma}^2(x) - \left|\hat{f}(x) - t\right|$$

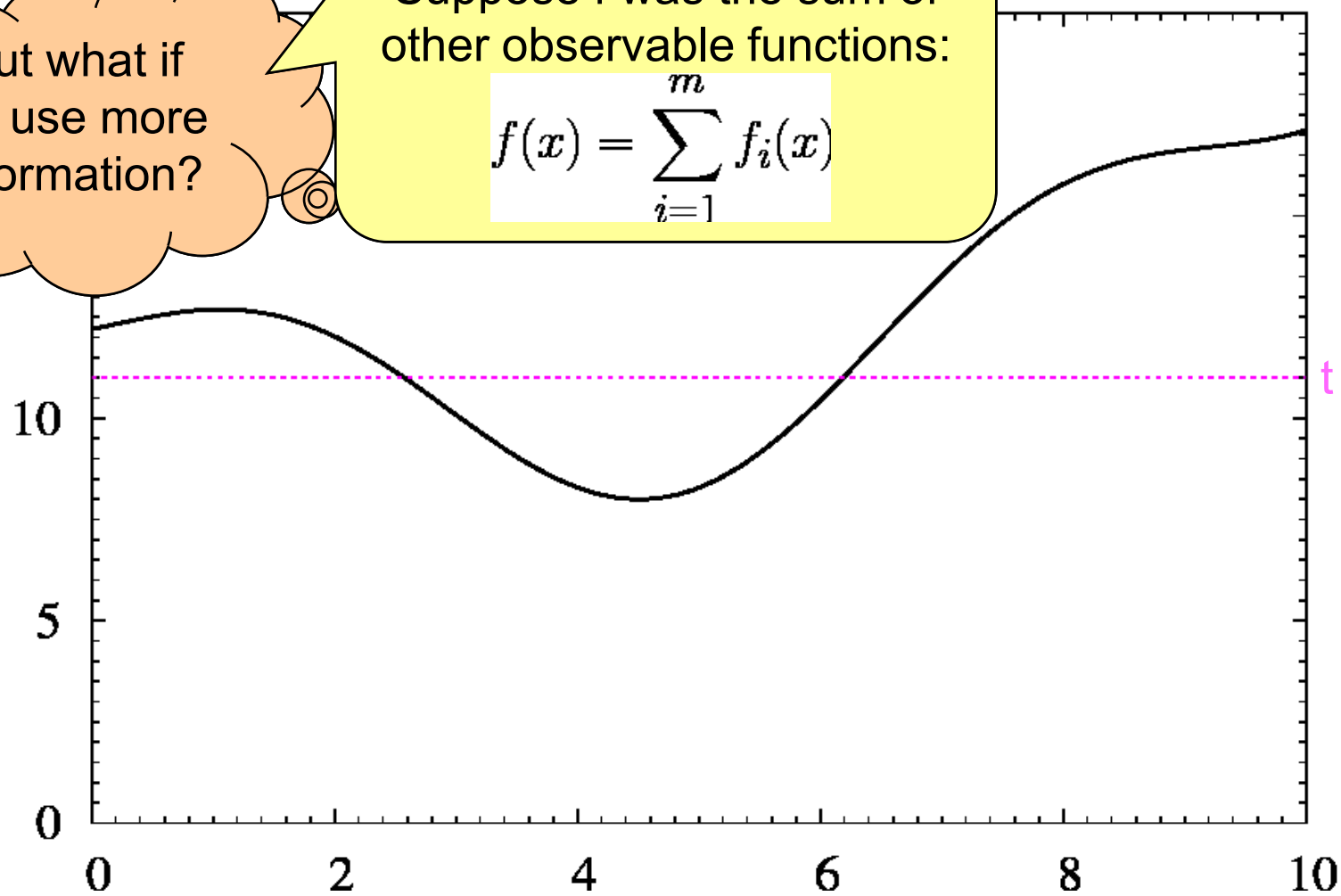
Mix variance and entropy.

The Level-Set Problem

But what if we use more information?

Suppose f was the sum of other observable functions:

$$f(x) = \sum_{i=1}^{m} f_i(x)$$



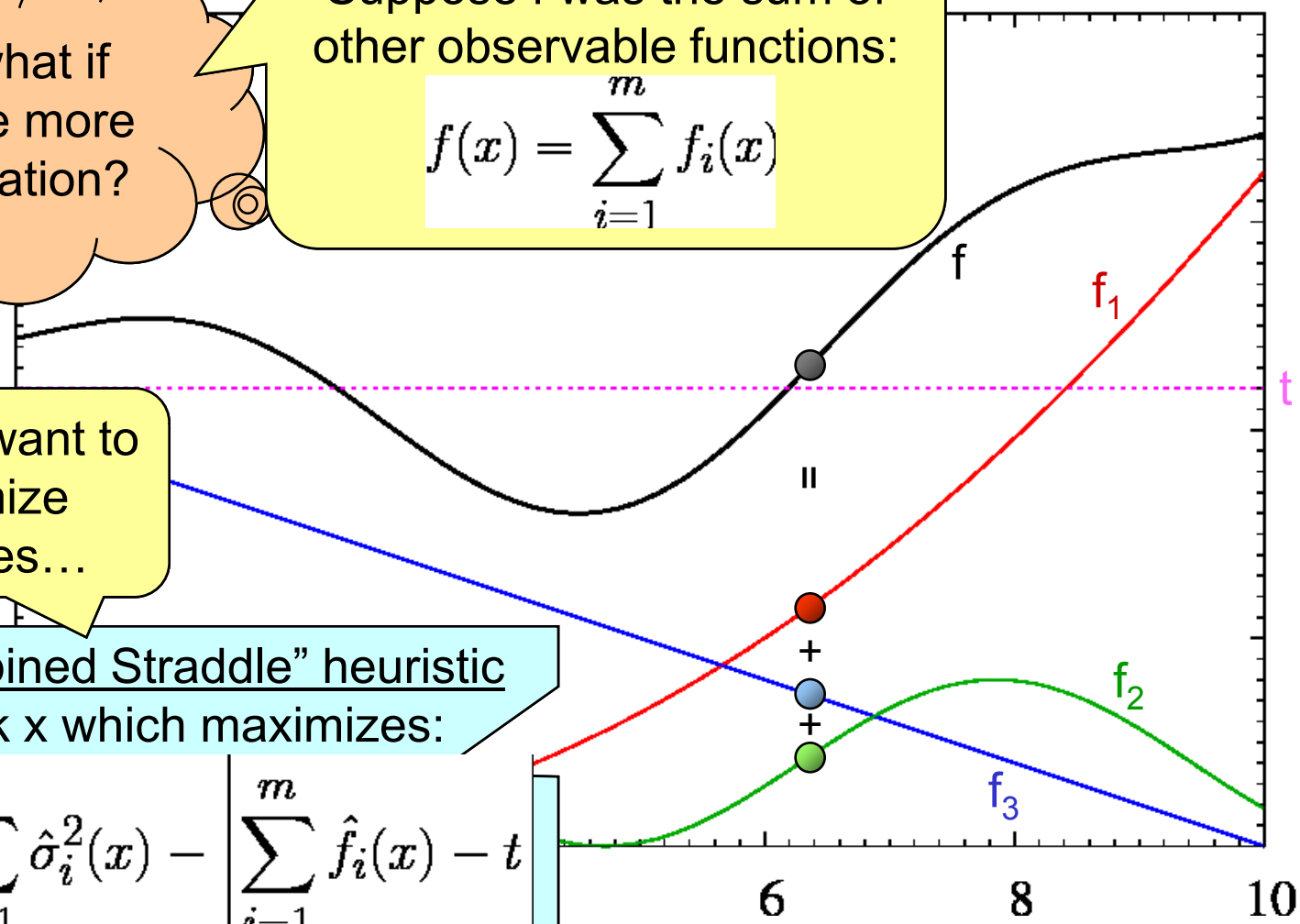
The Level-Set Problem

But what if we use more information?

Suppose f was the sum of other observable functions:

$$f(x) = \sum_{i=1}^{m} f_i(x)$$

But, we want to minimize samples…

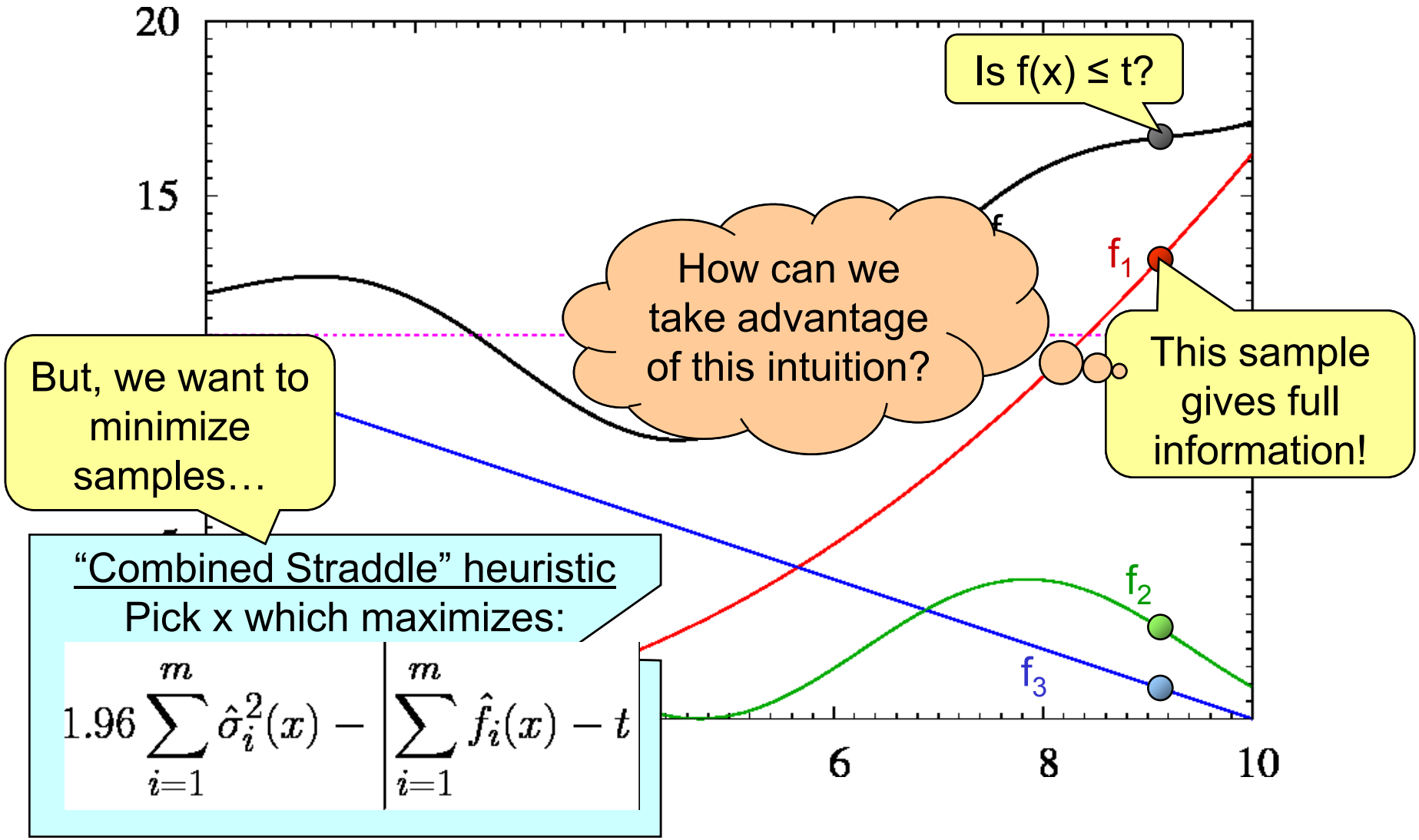


<u>"Combined Straddle" heuristic</u>

Pick x which maximizes:

$$1.96 \sum_{i=1}^{m} \hat{\sigma}_i^2(x) - \left| \sum_{i=1}^{m} \hat{f}_i(x) - t \right|$$

The Level-Set Problem



"Combined Straddle" heuristic

Pick x which maximizes:

$$1.96 \sum_{i=1}^{m} \hat{\sigma}_i^2(x) - \left| \sum_{i=1}^{m} \hat{f}_i(x) - t \right|$$

Level-Set Problem Summary

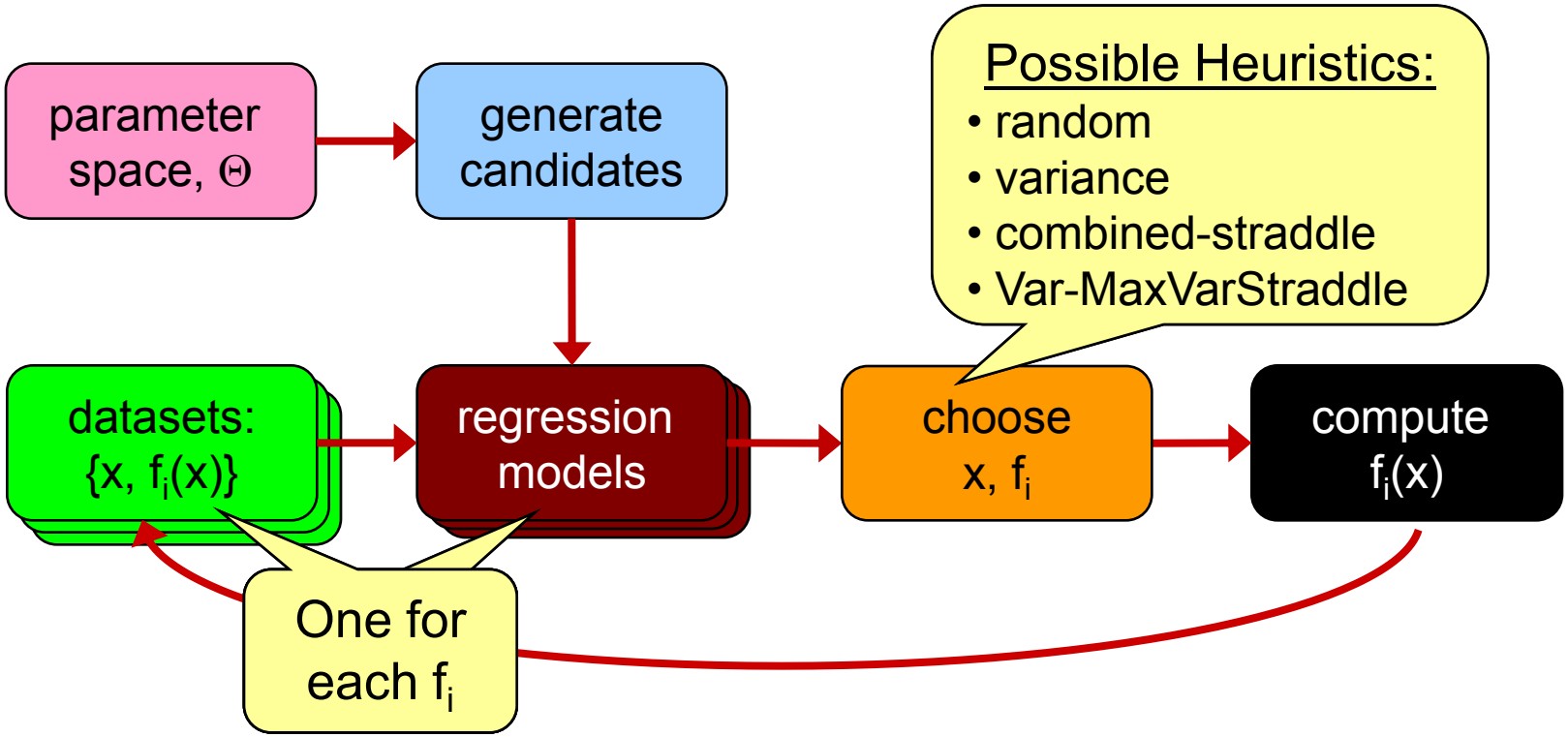
Single Function Case:

- Use straddle heuristic to balance exploration and exploitation
- $1.96\hat{\sigma}^2(x) - \left|\hat{f}(x) - t\right|$
- Mimics information gain

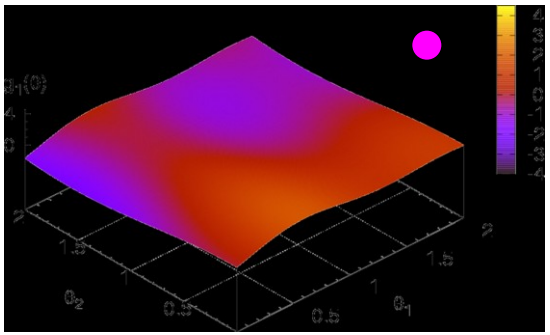
Multiple Function Case:

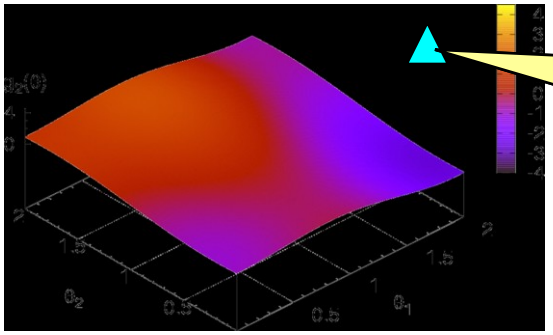
- Only sample one f_i
- Don't expect to reduce the variance by $\hat{\sigma}^2(x)$ but $\hat{\sigma}_i^2(x)$
- A better estimate of the knowledge gained is:
- $\max_i 1.96\hat{\sigma}_i^2(x) - \left|\sum_{i=1}^{m} \hat{f}_i(x) - t\right|$

Algorithm Outline

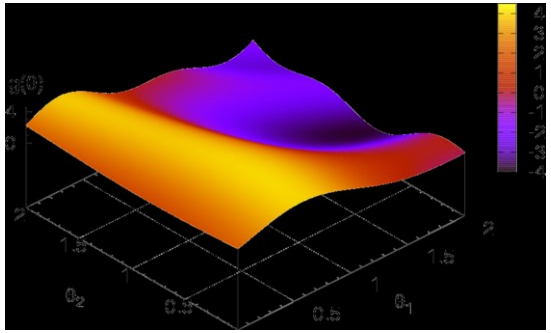


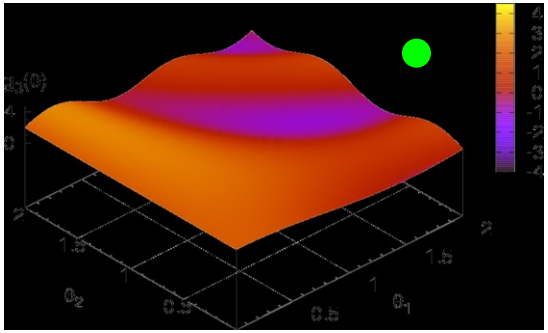
2D Example

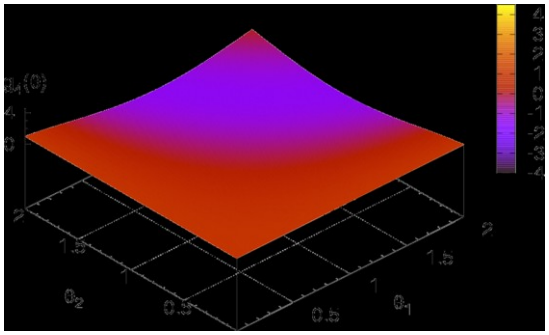




Use colors to denote samples

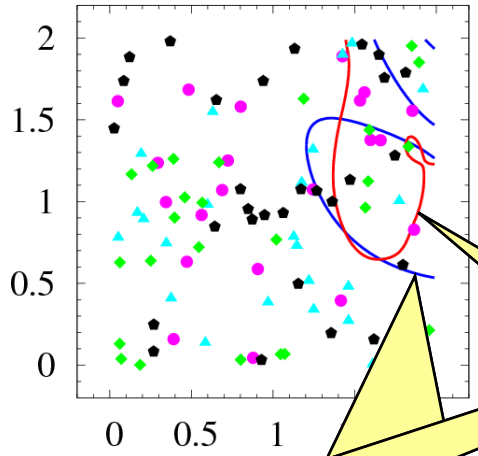




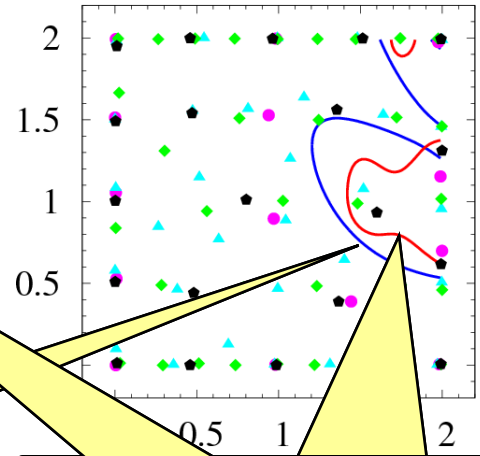


Possible Sampling Heuristics

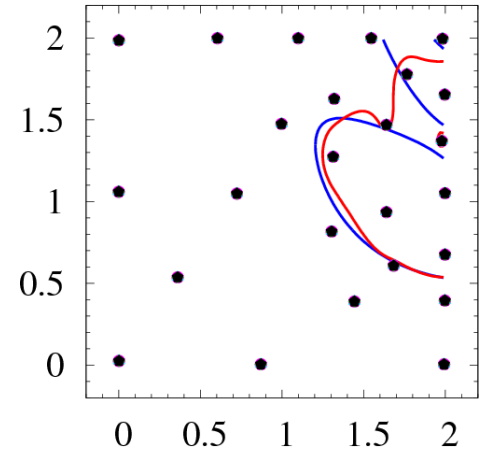
- Random



- Variance



- Combined-Straddle

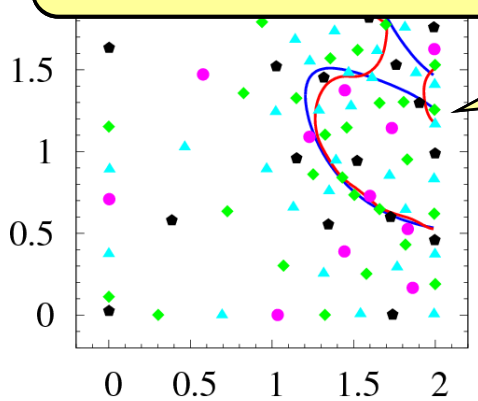


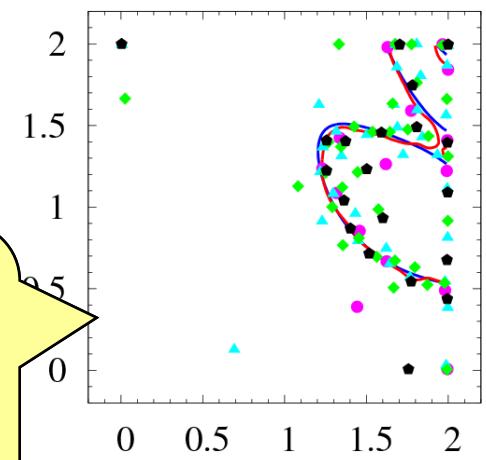
blue lines:
true level-set

Use straddle to select x, then

$$i = \operatorname{argmax} \hat{\sigma}_i^2(x)$$

- Var-MaxVarStraddle

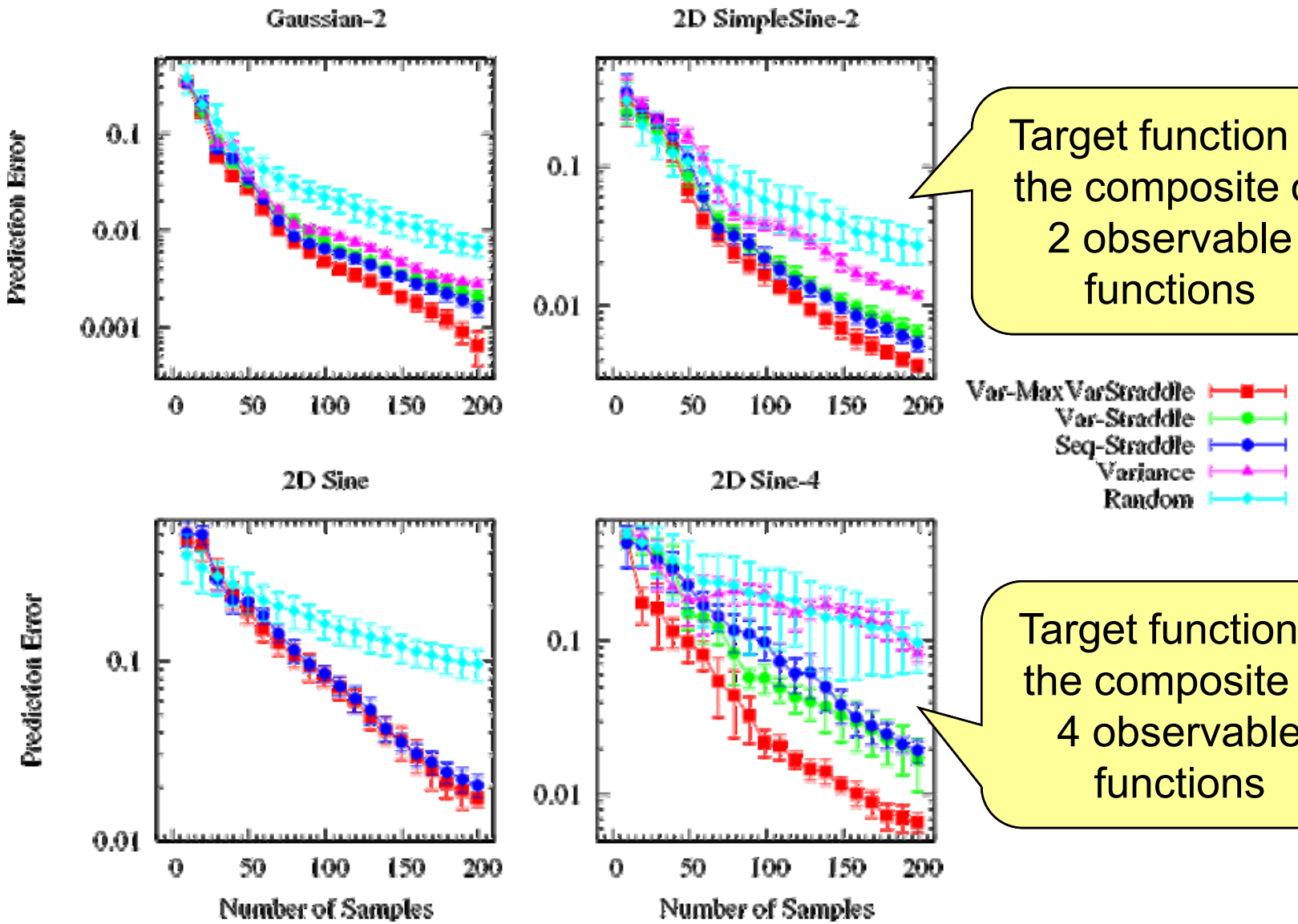




select x which maximizes

$$\max_i 1.96\hat{\sigma}_i^2(x) - \left| \sum_{i=1}^{m} \hat{f}_i(x) - t \right|$$

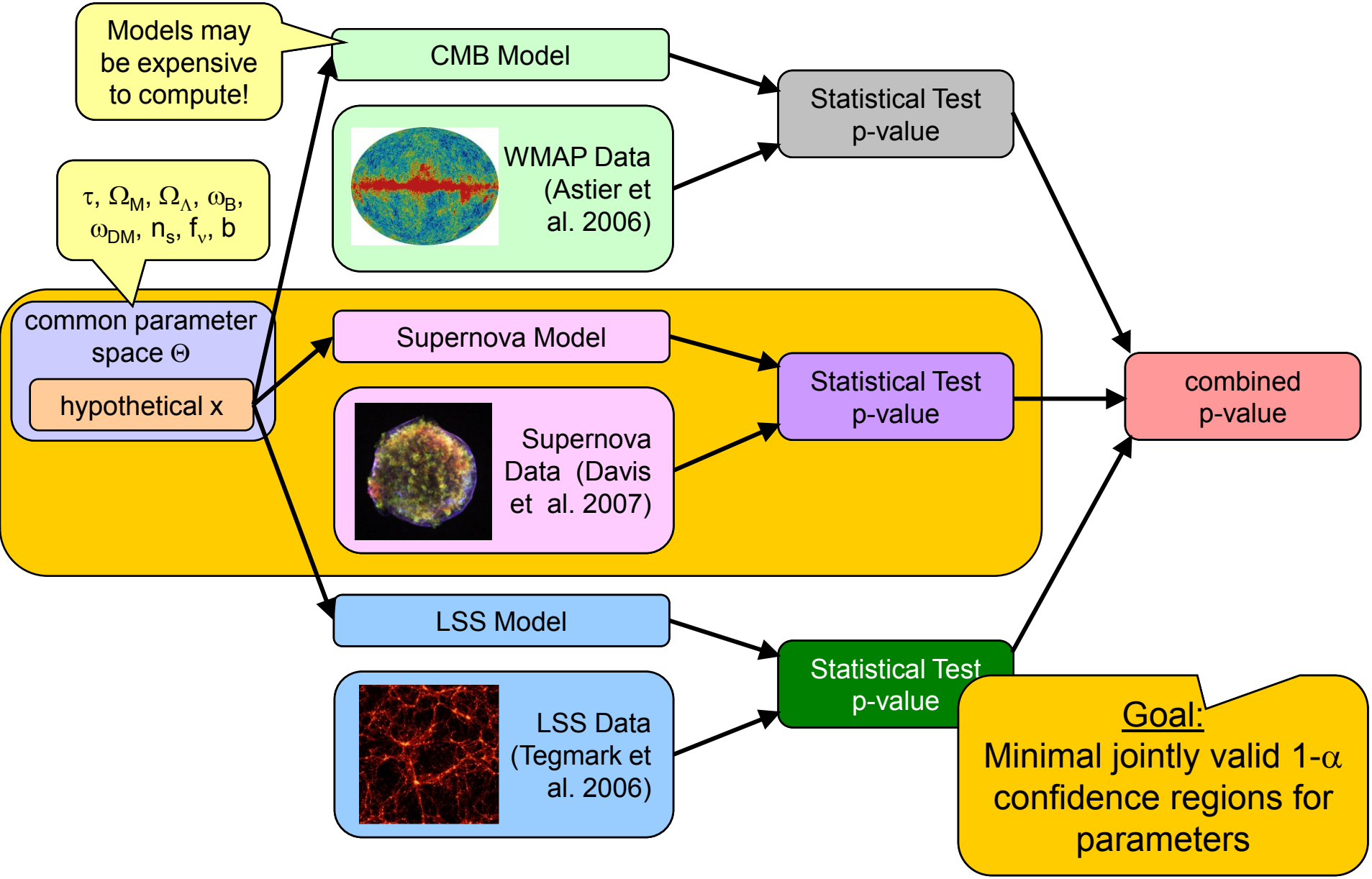
Experimental Results



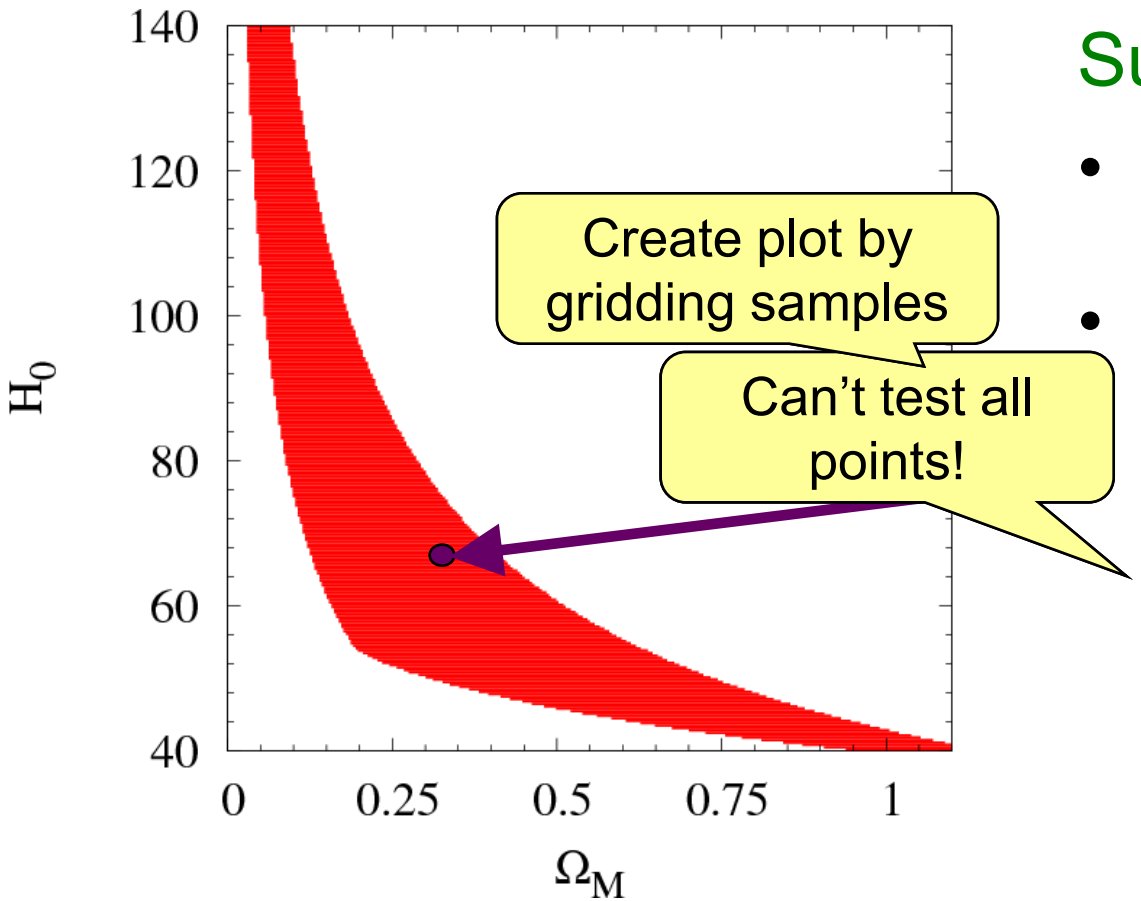
Target function is the composite of 2 observable functions

Target function is the composite of 4 observable functions

Application: Cosmology



Application: Cosmology

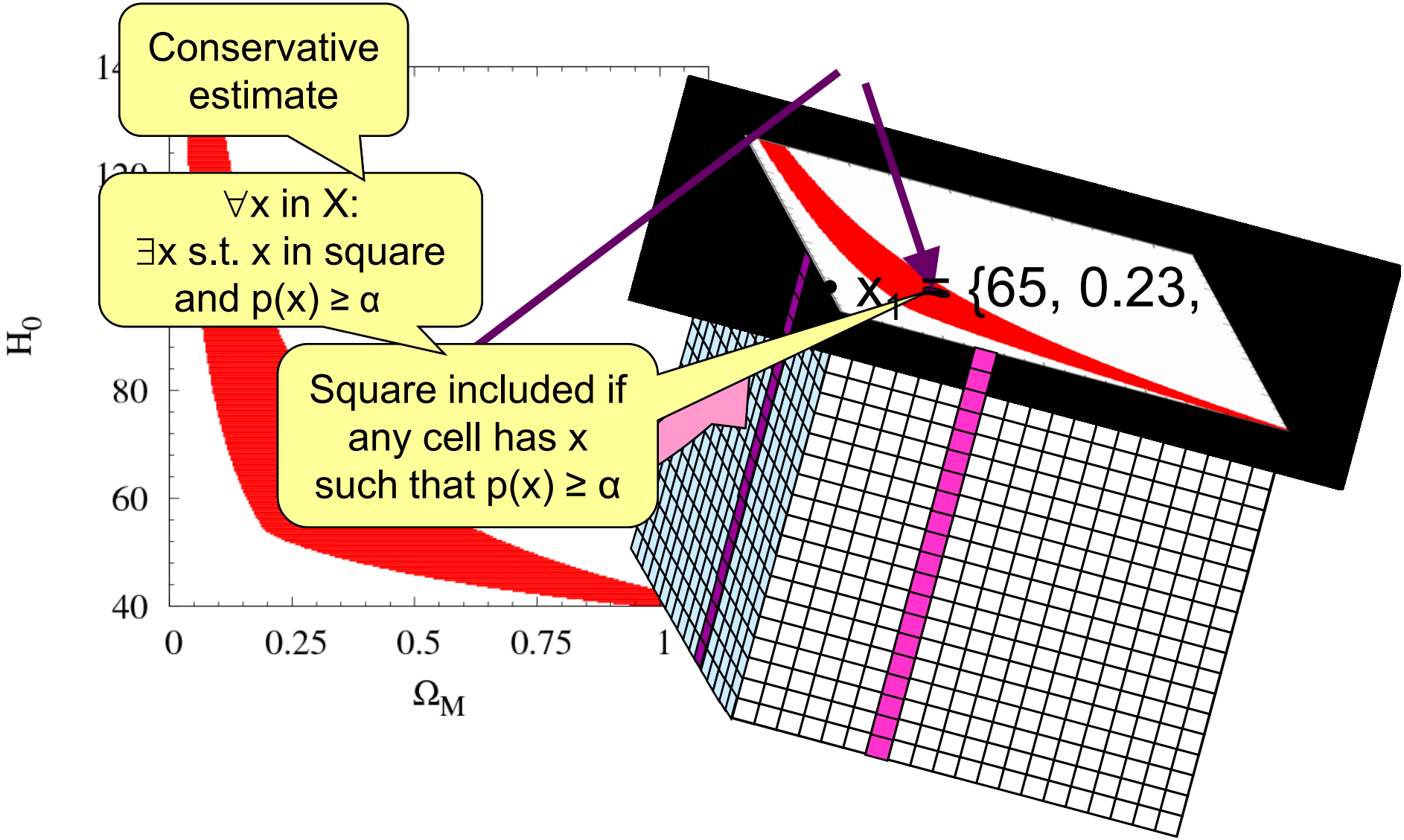


95% χ^2 confidence regions from supernova based on Davis et al. (2007) data

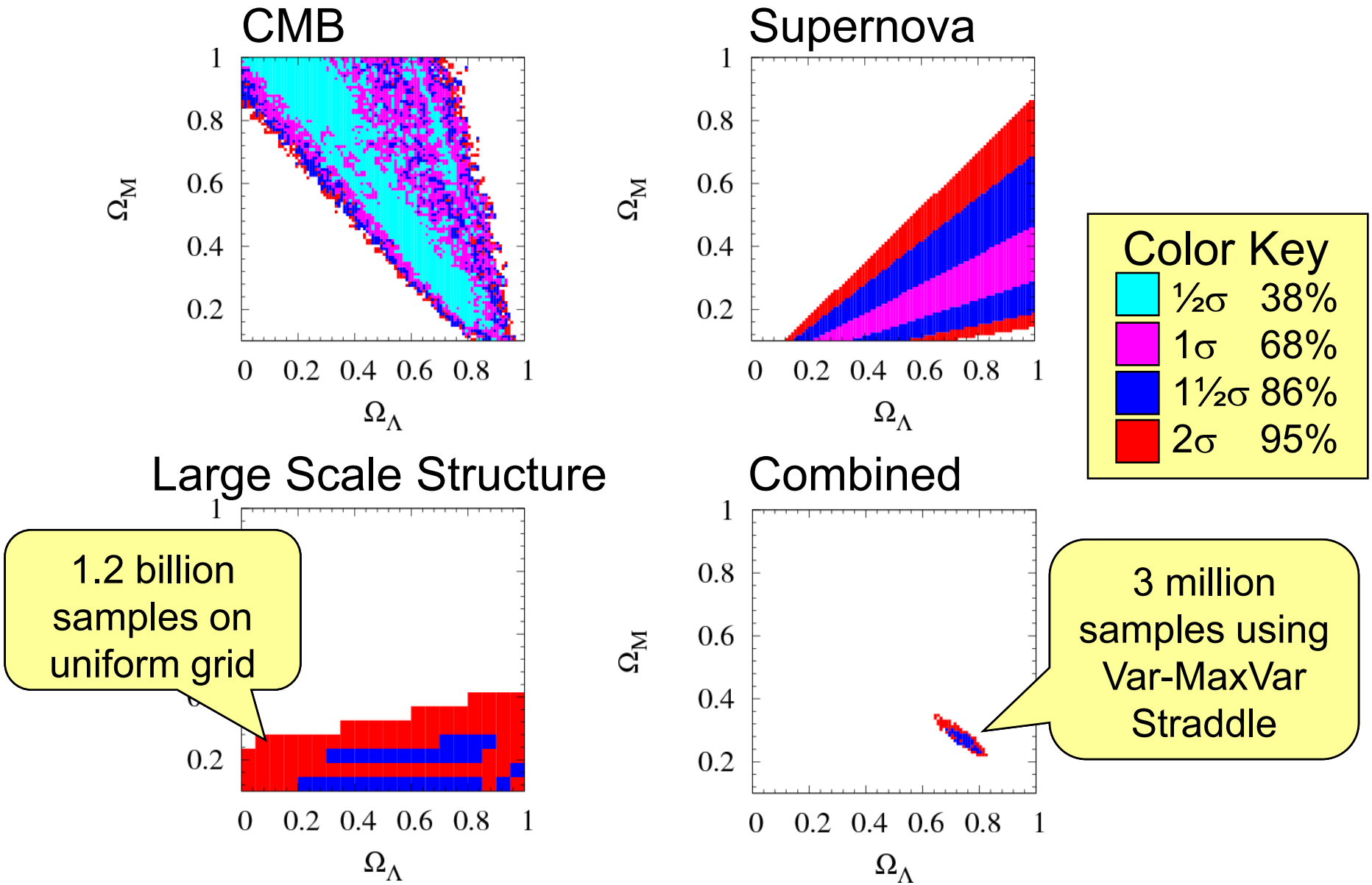
Supernova

- $x = \{H_0, \Omega_M, \Omega_\Lambda\}$
- $x_1 = \{65, 0.23, ?\}$
  - $\exists\ \Omega_\Lambda$ s.t. $p(x) \geq \alpha$

Application: Cosmology



Application: Cosmology



Conclusions

- Extended Straddle algorithm to multiple datasets
- Showed that combining p-values can be written as the sum of observable functions
- Deriving confidence regions this way:
  - Results in smaller regions (than intersection of marginals)
  - Is much more sample efficient than uniform sampling