Actively Learning Level-Sets of Composite Functions

Brent Bryan, Jeff Schneider
Motivation: Statistical Analysis

- Models may be expensive to compute!
- common parameter space $\Theta$
  - $\tau$, $\Omega_M$, $\Omega_\Lambda$, $\omega_B$, $\omega_{DM}$, $n_s$, $f_\gamma$, $b$
- hypothetical $x$

Goal:
- Minimal jointly valid $1-\alpha$ confidence regions for parameters

CMB Model
- WMAP Data (Astier et al. 2006)
  - Statistical Test $p$-value

Supernova Model
- Supernova Data (Davis et al. 2007)
  - Statistical Test $p$-value

LSS Model
- LSS Data (Tegmark et al. 2006)
  - Statistical Test $p$-value

combined $p$-value
Many ways to combine p-values:
- Bonferroni’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 \(\chi^2_{2m}\) distribution

\[f_i(x) = -2 \log(p_i(x))\]
The Level-Set Problem

Where does $f(x) = 11$?
The Level-Set Problem

Where does $f(x) = 11$?

Could pick points randomly, or uniformly

True Boundary

Predicted Boundary
The Level-Set Problem

Where does \( f(x) = 11 \)?

Use samples to estimate \( f \): \( \hat{f}, \hat{\sigma}^2(x) \)

Could try:
- entropy,
- variance,
- misclassification probability,
- etc.

“Straddle” heuristic works best (Bryan et al. 2005)
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.

$t$
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:

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

"Combined Straddle" heuristic

Pick \( x \) which maximizes:

\[
1.96 \sum_{i=1}^{m} \hat{\sigma}_i^2(x) - \sum_{i=1}^{m} f_i(x) - t
\]

But, we want to minimize samples…
The Level-Set Problem

"Combined Straddle" heuristic

Pick x which maximizes:

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

But, we want to minimize samples...

How can we take advantage of this intuition?

This sample gives full information!

Is \(f(x) \leq t\)?
Level-Set Problem Summary

**Single Function Case:**
- Use straddle heuristic to balance exploration and exploitation
- \[ 1.96 \hat{\sigma}^2(x) - |\hat{f}(x) - t| \]
- Mimics information gain

**Multiple Function Case:**
- Only sample one \( f_i \)
- Don’t expect to reduce the variance by \( \hat{\sigma}^2(x) \) but \( \hat{\sigma}^2_i(x) \)
- A better estimate of the knowledge gained is:

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

Parameter space, \( \Theta \) → Generate candidates

Datasets: \( \{x, f_i(x)\} \) → Regression models

One for each \( f_i \)

Choose \( x, f_i \) → Compute \( f_i(x) \)

Possible Heuristics:
- random
- variance
- combined-straddle
- Var-MaxVarStraddle

Gaussian process
Gaussian process regression models
One for each \( f_i \)
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

\[
\hat{i} = \arg\max \hat{\sigma}_i^2(x)
\]

select \( x \) which maximizes

\[
\max_i 1.96\hat{\sigma}_i^2(x) - \sum_{i=1}^{m} \hat{f}_i(x) - t
\]
Experimental Results

Target function is the composite of 2 observable functions

Target function is the composite of 4 observable functions
Application: Cosmology

Models may be expensive to compute!

τ, Ω_M, Ω_Λ, ω_B, ω_DM, n_s, f_ν, b

common parameter space Θ

hypothetical x

CMB Model

WMAP Data (Astier et al. 2006)

Supernova Model

Supernova Data (Davis et al. 2007)

LSS Model

LSS Data (Tegmark et al. 2006)

Statistical Test p-value

combined p-value

Goal: Minimal jointly valid 1-α confidence regions for parameters
Application: Cosmology

Supernova

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

$95\% \chi^2$ confidence regions from supernova based on Davis et al. (2007) data
Conservative estimate

\( \forall x \in X: \exists x \text{ s.t. } x \text{ in square and } p(x) \geq \alpha \)

Square included if any cell has \( x \) such that \( p(x) \geq \alpha \)

\[ x = \{65, 0.23, \cdot \} \]
Application: Cosmology

- **CMB**
  - $\Omega_M$ vs $\Omega_\Lambda$
  - 1.2 billion samples on uniform grid

- **Supernova**
  - $\Omega_M$ vs $\Omega_\Lambda$
  - Color Key:
    - $\frac{1}{2}\sigma$: 38%
    - $1\sigma$: 68%
    - $1\frac{1}{2}\sigma$: 86%
    - $2\sigma$: 95%

- **Large Scale Structure**
  - $\Omega_M$ vs $\Omega_\Lambda$

- **Combined**
  - $\Omega_M$ vs $\Omega_\Lambda$
  - 3 million samples using Var-MaxVar Straddle
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