Uncertainty Quantification for Machine Learning and Statistical Models

David J. Stracuzi

Joint work with: Max Chen, Michael Darling, Stephen Dauphin, Matt Peterson, and Chris Young
Inverse Problems

\[ D = M(T) \]
\[ T = M^{-1}(D) \]

- \( M \) has a known form
- Determining \( M^{-1} \) is ill-posed (unstable, non-unique)
- Example: determine Earth’s subsurface density (\( T \)) from gravitational field measurements (\( D \))
  - \( M \) is based on Newton’s law, \( F = Gm/r^2 \)
  - \( D \) is almost certainty insufficient to parameterize \( M^{-1} \)
  - Regularize and optimize to get a “best” solution
- Uncertainty comes from
  - \( D \)
  - Regularization + Optimization (definition of “best”)
  - Model parameters
Machine Learning and Statistical Modeling

D = M(T)
T = M⁻¹(D)

- M has an unknown form
- Substitute statistics for domain knowledge
- Uncertainty comes from
  - D
  - Regularization + Optimization (definition of “best”)
  - Model form (and parameters)
  - Inference
- Lack of domain knowledge impacts selection of model structure, regularization methods, and definition of “best”
Example: Seismic Onset Detection

Given
- Waveform data containing both noise and seismic signals

Produce
- Signal onset time
- Precision is critical to downstream processing

Known
- Wave propagation theory

Unknown
- Point of origin
- Event size or type
- Composition/density of medium (rock)

Simplified: signal is 3D.

Ground truth does not exist.
Seismic Detection Models

Model

- ARMA(p,q): $x_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$

Basic Approach

- Model the noise (M1) and the signal plus noise (M2) separately
- Optimize model parameters $\theta, \phi$ via maximum likelihood
- Akaike information criterion (AIC) to select $p, q$
- Point at which two models meet is the “best guess” signal onset
Seismic Detection Uncertainty

Method 1: Parametric Bootstrap

- Construct a sampling distribution for the input waveform
- Sample $M_1$ & $M_2$ to create new waveforms
- Fit a new model to each sample and record $k$

Method 2: Model Sampling

- $M_1=\text{ARMA}(p_1,q_1); M_2=\text{ARMA}(p_2,q_2)$
- $\langle p_1, q_1, p_2, q_2 \rangle$ determines $k$ and a likelihood
- Sample from $\langle p_1, q_1, p_2, q_2 \rangle$ and fit the data
- Use the likelihoods to construct probability distribution over $k$

Not clear that these produce similar distributions
Key Question: How much do sensor observations really tell us about the world?

- *Stats / ML research communities do not typically frame questions this way.*
- *Most work focuses on building a better statistical model*
- *“What” the model says emphasized over “how well”*
Impact on Downstream Analyses

- Several analyses depend on onset time:
  - Location (hypocenter)
  - Event type (natural or man-made) & size
  - Subsurface tomography
  - Earth model

- Provide additional information to human analysts
  - Distribution over onset times (vs. confidence interval)
  - Relative reliability of data points and sensors

- Rely more on current data, less on historical data and modeling assumptions
  - Improve sensitivity?

- Near term:
  - Model selection (improve fit and uncertainty)
  - Validate against synthetic & known events
  - Compare uncertainty results against analyst picks
Example 2: Multimodal Image Analysis

- **Given**
  - Co-registered imagery from multiple sources

- **Produce**
  - Segmentation with associated class probabilities and uncertainties, or
  - Detection of specific objects

- **Motivation**
  - Trend toward automated image analysis
  - Trend toward ”layered” analysis and combining information sources
  - Open question: How reliable are the results?
  - Open question: Which data is useful?

- **Currently**
  - As many analytic methods as there are tasks
  - Some ignore uncertainty
  - Others evaluate it in their own uniquely sane way
Approach

1. Merge data sources
   - Pixel = R + G + B + Height

2. Sample the data to create a new “image”
   - Non-parametric bootstrap

3. Probabilistically cluster the pixels
   - Gaussian Mixture
   - Nonparametric Mixture

4. Record the class probabilities for each pixel
   - Keep a histogram for each class at each pixel

5. Bootstrap: Repeat 2,3,4 many times
   - Histograms represent the posterior probability distributions

6. Future work: Translate unsupervised classes into semantic labels
Multimodal Image Analysis

Top row: optical image only
Bottom row: combined optical and lidar imagery
Closer Look at the Uncertainty

Comparison of Class probabilities for Optical Only and Combined Imagery.
Value of Information

- Evaluate data sufficiency
- How much does a given data source contribute to an analytic result?
  - Which data do we really need?
  - Which data do we wish we had?
Resource Tasking Under Uncertainty

Goal: Maximize utility of available data and resources

Approach: Quantify impact of data on detection quality
- Use uncertainty as basis for value
- Define value as improved label, geospatial, or temporal discrimination
- Incorporate information values into optimization objective functions

Relax assumption that desired information is actually obtained from a scheduled collect
Summary

- Uncertainty analysis does not “build a better model”
  It indicates how well a given model captures the data

- Research is to bridge the theory – application gap
  - Identifying which information is useful
  - Digging it out from deep within algorithms
  - Propagating uncertainty through layers of analysis

- Important questions
  - What’s the relationship between uncertainties generated by
    - Measurement errors & data sampling (bootstrap)
    - Model selection and induction processes (model sampling)
    - Inference (MCMC)
  - ...and how do we combine them? Or should we?
  - What issues arise when we propagate uncertainties from one statistical inverse problem to another?
Uncertainty Quotes

“We demand rigidly defined areas of doubt and uncertainty!”

“It ain’t what you don’t know that gets you into trouble. It’s what you know for sure that just ain’t so.” – Uncertain Origin

“Uncertainty is an uncomfortable position. But certainty is an absurd one.” – Voltaire

"Ignorance more frequently begets confidence than does knowledge...”
– Charles Darwin

“As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.”
– Albert Einstein

POC: David Stracuzzi
djstrac@sandia.gov