Identifiability and Underdetermined Inverse Problems

Nathan Stein

Department of Statistics, The Wharton School, University of Pennsylvania

ISSI Workshop, Bern, Switzerland, May 2015

Motivating Problem and Some Notation

- R_{cT} = discretized response at temperature T in channel c
- We observe counts $y = (y_1, \ldots, y_{N_c})$
- For simplicity, suppose $y_c \sim \text{Poisson}(\mu_c)$, $c = 1, \ldots, N_c$, where

$$\mu_c = \sum_T R_{cT}\theta_T$$

- More compactly,

$$\mu = R\theta,$$

  where $\mu = (\mu_1, \ldots, \mu_{N_c})$ and R is the $N_c \times N_T$ response matrix
- $\theta = (\theta_1, \ldots, \theta_{N_T})$ is the DEM
- Underdetermined: $N_T > N_c$

Some Definitions

In general:

- Data y
- Parameter θ
- Likelihood $p(y|\theta)$

A **sufficient** parameter is a function $f(\theta)$ such that

$$f(\theta) = f(\theta') \text{ implies that } p(y|\theta) = p(y|\theta') \quad \forall \theta, \theta'$$

An **identifiable** parameter is a function $f(\theta)$ such that

$$f(\theta) \neq f(\theta') \text{ implies that } p(y|\theta) \neq p(y|\theta') \quad \forall \theta, \theta'$$

Definitions from Barankin (1960)

If $f(\theta)$ is sufficient and identifiable, then it makes sense to write the likelihood as

$$p(y|\theta) = p(y \mid f(\theta))$$

Sufficiency and Identifiability

Intuitively:

- sufficiency ensures that the parameter $f(\theta)$ is **rich enough** to "use" all of the information in the data, and
- identifiability ensures that $f(\theta)$ is **not too rich** for the data to be informative about it.

Examples:

$f(\theta) = \mu = R\theta$	**sufficient** and **identifiable**
$f(\theta) = \mu_c = \sum_T R_{cT}\theta_T$	**identifiable** but **not sufficient**
$f(\theta) = \theta$	**sufficient** but **not identifiable**

Partially Identified Models

- Existing literature on **partially identified models**, primarily in econometrics and causal inference
- Example of a partially identified model: Conditional on an identifiable parameter μ, the parameter of interest θ is known to lie in the set $\Theta(\mu)$ (e.g., Moon and Schorfheide, 2012)
- The key question: How should we perform (Bayesian or frequentist) inference on the identified set $\Theta(\mu)$?
- Connection to the DEM problem:

$$\Theta(\mu) = \{\theta : \mu = R\theta\}$$

- Separates inference about $\Theta(\mu)$ (accounting for, e.g., Poisson noise) from the solution to the underdetermined/ill-posed inverse problem

The Simplest Case

- Suppose we permit ourselves to place a prior $p(\mu)$ on the sufficient and identifiable parameter μ
  - (Because μ is identifiable, as we collect more data, information about μ will accumulate via the likelihood.)
- This implies a prior for the identified set $\Theta(\mu)$
- ... but *not* a full prior for θ
- We can obtain the posterior distribution

$$p(\mu|y) \propto p(y|\mu)\ p(\mu)$$

- Suppose we sample S values from this posterior:

$$\mu^{(1)}, \ldots, \mu^{(S)} \sim p(\mu|y)$$

- Then we can calculate the collection of sets

$$\Theta(\mu^{(1)}), \ldots, \Theta(\mu^{(S)})$$

- The variability in these sets characterizes the uncertainty due to, e.g., Poisson noise, without relying on a full prior for θ. This preserves (most of) the separation between the statistical uncertainty and the ill-posedness.

Complication: Additional Levels of Uncertainty

- Now suppose we want to incorporate additional uncertainties
- We encapsulate these in a prior $p(R)$ on the response R
- Now the identified set is a function of **both** μ and R:

$$\Theta(\mu, R) = \{\theta : \mu = R\theta\}$$

- Naive strategy: alternate between updating μ given R, and R given μ:

$$\mu^{(s+1)} \sim p(\mu \mid R^{(s)}, y)$$
$$R^{(s+1)} \sim p(R \mid \mu^{(s+1)}, y)$$

- But it is not clear how to obtain the required conditional distributions **without first specifying a full prior on θ**
  - This sacrifices the separation from the underdetermined/ill-posed inverse problem.
- How can we specify a sensible **conditional** prior $p(\mu|R)$?
  - Is there a feasible way to do this coherently without first specifying a prior on θ?