Discussion of "Bayesian Model Averaging for Disclosure Risk Assessment"

Stephen E. Fienberg
CREST, INSEE, Paris, France
And
Department of Statistics
Center for Automated Learning & Discovery
Center for Computer & Communications Security
Carnegie Mellon University
Pittsburgh, PA, U.S.A.

Overview

- **Key Elements of Paper**
- **Role of Model Averaging**
- **The Examples**
- **What to Report?**

Key Elements of Paper

- **Neat paper!**
- **Built around combining three of my favorite topics:**
  - **Bayesian framework**
  - **Log-linear models**
  - **Model averaging (even though I don't like it in general!)**

Model Averaging

- **Two strategies for working with statistical models, both of which require specification of class of relevant models:**
  - **Model search and estimation:**
    - **Focuses in on single model and allows for estimation of posterior given prior for that model.**
    - **Ignores variability due to model search.**
  - **Model averaging:**
    - **Uses information from all models.**
    - **Can produce nonsensical results for parameters that change meaning across models.**

When Does MA Work?

- **When we focus in on potential observables:**
  - **Predicting results for new observations.**
- **When differing model parameters are incidental:**
  - **Estimating size of closed population using multiple recapture methods (and possibly log-linear models).**
  - **For integrating across models in assessing disclosure risk (e.g., Dobra, Fienberg, & Trottini, 2003).**

From Observables to Rest of Population

- **Bayesians can reason from sample to remainder of population without directly taking all features of sample design into account, i.e., via $P(F\text{-}f \mid f)$.**
  - **MA computes $P(f \mid m)$ and then weights according to the models $m \in M$.**
  - **By treating those not in sample as predicted values we get $P(F\text{-}f \mid f)$ or $P(F \mid f)$.**

Decomposable Log-Linear Models

- **Decomposable models are wonderful to work with analytically.**
- **As I showed in my presentation showed, for reporting purposes we often need to go beyond this class of models.**
- **Issue: Does model averaging finesse this shortcoming?**
  - **Do decomposable models span sufficient spectrum of relevant model space to get posterior distribution approximately correct?**

The Examples

- **Example 1 (2×3×3 table), is close in spirit to analyses in Dobra-Fienberg-Trottini (2003, Valencia):**
  - **DFT look at posterior inferences of intruder for 4 small cells assuming this is population table--they pose serious disclosure risk.**
  - **Here we see value of sampling! As N↑, the risk goes down.**
  - **Data actually come from 1 in 1000 sample, and so release appears to be relatively safe.**

What to Report?

- **Suppose we use these risk measures and the Bayesian model averaging methods and risk is "high".**
- **What do we do next?**
- **Does MA help here or do we need to revert to individual models?**