Sequential Experimental Designs for Generalized Linear Models

Hovav A. Dror and David M. Steinberg, JASA (2008)

Bob A. Salim

April 23, 2013

Background (Motivation)

- Designing an experiment or a study
- Restrictions on sample size (due to cost, time, etc)
- How do we design an efficient experiment with valid estimate of the parameter?

Efficient Design

- Efficient experimental designs for GLMs depend on the unknown coefficients.
- Sequential design: next design point chosen based on current data.

What is Efficient?

- Optimality criterion: D-optimal
- D-optimality criterion: maximize the determinant of the information matrix $\mathbf{I}(\boldsymbol{\beta};\mathrm{d})$
  $\boldsymbol{\beta}$ = parameters in the model; d = design
- Bayesian D-optimality criterion (Chaloner and Larntz, 1989):

$$\phi(d) = \int log(\mathbf{I}(\boldsymbol{\beta}; d)) d\pi(\boldsymbol{\beta}) \tag{1}$$

where $\pi(\boldsymbol{\beta})$ = prior distribution on $\boldsymbol{\beta}$.

Previous Works

- Chaudhuri and Mykland (1993): Sequential designs in GLMs could lead to fully efficient designs and asymptotically efficient MLEs
- Dixon and Mood (1948); Haines, Perevozskaya, and Rosenberger (2003); Ivanova and Wang (2004); Biedermann, Dette, and Zhu (2006); Karvanen, Vartiainen, Timofeev, and Pekola (2007): Sequential designs for binary data, focused on single-factor experiments.

Limitations of Previous Works

- Performance in small samples unknown
- Quite a few initial observations are required for MLE to exist
- MLE might be biased in small samples, resulting in inefficient choice of experimental sites

Overview of Proposed Method: Algorithm

- Define augmentation horizon
- Generate an m-point augmentation to the current design
- Select an augmentation site(s) from the points found in the previous step, or their median.

Overview of Proposed Method: Approximate Design Criteria

- Discretized posterior

$$\phi_1(d) = \sum_{u=1}^{n} r_u log(\mathbf{I}(\boldsymbol{\beta}_u; d)) \quad (2)$$

where $r_u = L(\boldsymbol{\beta}_u) / \sum_{u=1}^{n} L(\boldsymbol{\beta}_u)$

- A faster approximation

$$\phi_2(d) = log(\mathbf{I}(\tilde{\beta}; d)) \quad (3)$$

where $\tilde{\beta} = \beta_{(g)}$ such that $\sum_{u=1}^{g} r_{(u)} \geq 0.5$ and $\sum_{u=g}^{n} r_{(u)} \geq 0.5$.

- Another approximation for determining augmentation horizon

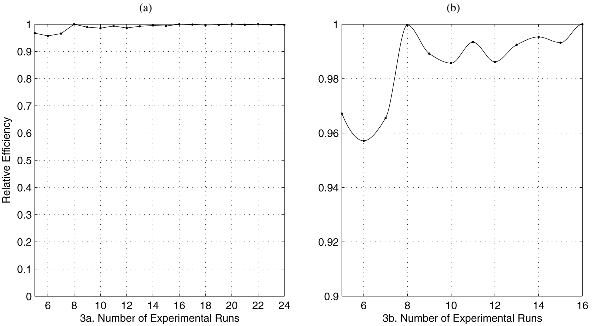
$$\phi_3(d) = (1/p)\phi_2(d) - log(n) \quad (4)$$

Overview of Proposed Method: Augmentation Horizon

- Proposed to avoid problems, such as singular information matrix
- It is the number of observations (m) needed for highly efficient D-optimal design at prior median
- Determined at the start of the experiment

Overview of Proposed Method: Augmentation Horizon

- Find locally D-optimal designs at prior median for $n = p, ..., P$
- Define the efficiency of $d_n = \exp[\phi_3(d_n) - \phi_3(d*)]$
  where $d*$ is the design among $d_p, ..., d_P$ that maximizes $\phi_3(d)$
- Augmentation horizon m is the smallest value of n for which the efficiency is at least 99%



Next Steps

- Detailed study on the augmentation horizon
- Simulations on performance with small samples
- Comparison with existing methods

References

Dror, H.A. and Steinberg, D.M. (2008). Sequential Experimental Designs for Generalized Linear Models. Journal of American Statistical Association 103:481, 288-298

Abdelbasit, K.M. and Plackett, R.L. (1983). Experimental Designs for Binary Data. Journal of American Statistical Association 101, 747-759

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian Design Applied to Logistic Regression Experiments. Journal of Statistical Planning and Inference, 21, 191-208

Dixon, J.W. and Mood, A.M. (1948) A Method for Obtaining and Analyzing Sensitivity Data. *Journal of American Statistical Association* 43, 109-126

Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. Annals of Mathematical Statistics 29, 400-407