Improving Efficiency with a Relevant Subset Design

Adam Lane

June 15, 2026

Motivation

① The use of ancillary statistics in post-data inference has a long history with significant interest and support.

② Little work has been done to investigate the use of ancillary statistics in the design.

Ancillary Statistics

① A random variable, $\mathcal{A}$, is *ancillary* if its distribution is independent of the model parameters.

② $\mathcal{A}$ is an *ancillary complement* if $(\hat{\theta}, \mathcal{A})$ is a minimal sufficient statistic, where $\hat{\theta}$ is the maximum likelihood estimate (MLE) of the p-dimensional parameter of interest θ.

③ Fisher (1934) noted that the using only the distribution of $\hat{\theta}$ results in a loss of information and argued that this information can be recovered by conditioning on the observed value of the ancillary complement.

Model

Outcomes are observed sequentially from model:

$$y(j) = \eta[\boldsymbol{x}(j), \boldsymbol{\theta}] + \varepsilon(j), \quad j = 1, \ldots, n,$$

where

① (j) indicates the jth sequential observation,

② $\eta(\boldsymbol{x}, \boldsymbol{\theta})$ is a, potentially nonlinear, function,

③ $\boldsymbol{\theta}$ is a p-dimensional parameter,

④ $\boldsymbol{x}$ is and s-dimensional vector of covariates, and

⑤ the errors follow a known non-Gaussian distribution in the location family with mean 0 and variance σ^2.

Design

① Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)^T$ represent the unique design points.

② Let $t_i(j) = 1$ if observation j is allocated to design point i; that is, if $\boldsymbol{x}(j) = \boldsymbol{x}_i$, and 0 otherwise.

③ Let $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d)$ represent the allocation matrix, where $\boldsymbol{t}_i = [t_i(1), \ldots, t_i(n)]^T$.

④ A fixed design, ξ, is completely determined by the support points and vector of corresponding allocation weights, $\boldsymbol{w} = (w_1, \ldots, w_d)^T$, where $w_i = n_i/n$ and $n_i = \sum_{j=1}^{n} t_i(j)$.

Fisher Information

① Let $l_y = \log f(y|\eta)$ with first and second derivatives with respect to η denoted $\dot{l}_y$ and $\ddot{l}_y$.

② Similarly, let $\dot{l}_{\boldsymbol{y},\boldsymbol{t}_i} = \sum_{j=1}^{n} t_i(j)\dot{l}_{y(j)}$ and $\ddot{l}_{\boldsymbol{y},\boldsymbol{t}_i} = \sum_{j=1}^{n} t_i(j)\ddot{l}_{y(j)}$, be the cumulaitve first and second derivatives for all observations from the ith design point.

③ The observed information can be written as

$$\boldsymbol{I}_{\boldsymbol{y},\boldsymbol{T}}(\boldsymbol{\theta}) = -\sum_{j=1}^{n}(\ddot{l}_{\boldsymbol{y},\boldsymbol{t}_i}\dot{\eta}_i\dot{\eta}_i^T + \dot{l}_{\boldsymbol{y},\boldsymbol{t}_i}\ddot{\eta}_i),$$

where $\dot{\eta}_i$ and $\ddot{\eta}_i$ are the first and second derivatives of $\eta(\boldsymbol{x}_i, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

④ In this context the Fisher information in the sample is

$$\boldsymbol{F}_{\xi}(\boldsymbol{\theta}) = \mathrm{E}\left[\boldsymbol{I}_{\boldsymbol{\mathcal{Y}},\boldsymbol{\tau}}(\boldsymbol{\theta})\right] = \mu\sum_{i=1}^{d} w_i\dot{\eta}_i\dot{\eta}_i^T,$$

where $\mu = -\mathrm{E}[\ddot{l}_y]$.

Information in the Relevant Subset

① When $(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{A}})$ is a minimal sufficient statistic, $\hat{\boldsymbol{\theta}}$ is a sufficient statistic on the subset of the sample space where $\boldsymbol{\mathcal{A}} = A$.

② This subset will be referred to as the *relevant subset*

③ The Fisher information in $\hat{\boldsymbol{\theta}}|(\boldsymbol{\mathcal{A}} = A)$ is equivalent to the Fisher information in $\boldsymbol{\mathcal{Y}}|(\boldsymbol{\mathcal{A}} = A)$ defined as

$$\boldsymbol{H}_{\boldsymbol{\mathcal{A}}}(\boldsymbol{\theta}) = \mathrm{E}\left[\boldsymbol{I}_{\boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{T}}}(\boldsymbol{\theta}) \middle| \boldsymbol{\mathcal{A}}\right] = \sum_{i=1}^{d} h_{\boldsymbol{\mathcal{A}}_i} \dot{\boldsymbol{\eta}}_i \dot{\boldsymbol{\eta}}_i^T,$$

where $h_{\boldsymbol{\mathcal{A}}_i} = -\mathrm{E}[\ddot{l}_{\boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{T}}_i} | \boldsymbol{\mathcal{A}}_i]$.

④ According to Fisher, this measure correctly reflects the information about $\boldsymbol{\theta}$ in the MLE, $\hat{\boldsymbol{\theta}}$, for the observed value of A.

Opportunity for Improvement

1. Prior to data collection it is reasonable to target the Fisher information in the sample.
2. The observed value of the ancillary complement, $\boldsymbol{\mathcal{A}} = \boldsymbol{A}$ is random.
3. However, if observations occur sequentially, in a series of runs, the relevant subset from the preceding runs is known and can be used in the design assignment of the current run.
4. The objective of this work is to propose a framework which allows the relevant subset to be incorporated into an adaptive design to improve confidence regions (inference) without losing information in the sample.

Information Relation

① The difference between the information measures can be described as

$$n^2(\mathrm{E}[\boldsymbol{H}_{\boldsymbol{\mathcal{A}}}^{-1}] - \boldsymbol{F}_{\xi}^{-1}) \approx \mathrm{Var}[\boldsymbol{u}_{\boldsymbol{\mathcal{A}}}] + \mathrm{Var}[\boldsymbol{v}_{\boldsymbol{\mathcal{A}}}] \tag{1}$$

where $\boldsymbol{u}_A = (u_{\boldsymbol{a}_1}, \ldots, u_{\boldsymbol{a}_d})^T$, $u_{\boldsymbol{a}_i} = n^{-1/2}[h_{\boldsymbol{a}_i} - w_i \sum_{i=1}^{d} h_{\boldsymbol{a}_i}]$, $\boldsymbol{v}_A = (v_{\boldsymbol{a}_1}, \ldots, v_{\boldsymbol{a}_d})^T$, $v_{\boldsymbol{a}_i} = n^{-1/2}[w_i(\sum_{i=1}^{d} h_{\boldsymbol{a}_i} - n)]$.

② Both of the variance terms on the RHS of (1), converge to a constant; however, $\mathrm{Var}[\boldsymbol{v}_{\boldsymbol{\mathcal{A}}}]$ cannot be altered in a design with fixed sample size.

③ **The main conceptual point of the current work is that $\mathrm{Var}[\boldsymbol{u}_{\boldsymbol{\mathcal{A}}}]$ can be reduced through adaptation in a sequential experiemnt.**

Adaptive Experiment

① In this work a method is proposed that uses the relevant subset from the first r runs to determine the probability of assigning observation in run $r+1$ to the ith support point.

② Let $\mathcal{P}_i\{r\}$ represent the probability of assigning an observation in run r to support point i

③ Let $\mathcal{N}\{r\}$ represent the (random) sample size of run r.

④ The notation $\{r\}$ is used to denote the rth run.

A Two-Stage Illustration

① To begin suppose a fixed design, ξ, with support points $\boldsymbol{X}$ and allocations w has the properties desired for the proposed experiment.

② Assign $n\{1\}$ samples in the first stage according to design ξ.

③ Let $u_{\boldsymbol{\mathcal{A}}_i\{1;2\}}$ represent the value of $u_{\boldsymbol{\mathcal{A}}}$ [in equation (1)] from a two-stage experiment.

④ The expectation of $u_{\boldsymbol{\mathcal{A}}_i\{1;2\}}$ conditional on $\boldsymbol{\mathcal{A}}\{1\}$

$$\mathrm{E}\left[u_{\boldsymbol{\mathcal{A}}_i\{1;2\}} | \boldsymbol{\mathcal{A}}\{1\}\right] = u_{\boldsymbol{\mathcal{A}}_i\{1\}} + \mathcal{N}\{2\}(\mathcal{P}_i\{2\} - w_i).$$

⑤ The notation $\{1;2\}$ is used to indicate quantities that include the data from run 1 to run 2.

A Two-Stage Illustration (cont.)

① Set the right-hand side of the preceding equation equal to 0 and solve for $\mathcal{P}_i\{2\}$ to obtain

$$\mathcal{P}_i\{2\} = w_i - u_{\boldsymbol{\mathcal{A}}_i\{1\}}/\mathcal{N}\{2\}$$

for $i = 1, \ldots, d$.

② Defining $\mathcal{N}_i\{2\} = u_{\boldsymbol{\mathcal{A}}_i\{1\}}/w_i$ and

$$\mathcal{N}\{2\} = \lceil \max\{\mathcal{N}_1\{2\}, \ldots, \mathcal{N}_d\{2\}\} \rceil,$$

ensures that all $\mathcal{P}_i\{2\} \geq 0$. These selection immediately yield $\mathrm{E}[u_{\boldsymbol{\mathcal{A}}_i\{1;2\}} | \boldsymbol{\mathcal{A}}\{1\}] = 0$.

A Two-Stage Illustration (cont.)

① In stage 2, allocate $\mathcal{N}\{2\}$ observations with each observation having probability $\mathcal{P}_i\{2\}$ of being assigned to support point i.

② For an experiment conducted as described, it can be shown that

$$\lim_{n\to\infty} \text{Var}[u_{\mathcal{A}\{1;2\}}] = 0$$

This is a significant improvement to the expansion in (1) where $\text{Var}[u_{\mathcal{A}\{1;2\}}]$ converges to a constant.

Randomized Relevant Subset Design (RRSD)

① Allocate $n\{1\}$, observations according to the predetermine design with support $\boldsymbol{X}$ and allocations $\boldsymbol{w}$.

② Compute $\mathcal{N}\{r+1\} = \lceil \max\{\mathcal{N}_1\{r+1\}, \ldots, \mathcal{N}_d\{r+1\}\} \rceil$ and $\mathcal{P}_i\{r+1\} = w_i - u_{\boldsymbol{\mathcal{A}}_i\{1;r\}}/\mathcal{N}\{r+1\}$, where $\mathcal{N}_i\{r+1\} = u_{\boldsymbol{\mathcal{A}}_i\{1;r\}}/w_i$.

③ Allocate observations in run r to the ith support point with probability $\mathcal{P}_i\{r+1\}$.

④ Repeat steps 3-6 until $\mathcal{N}\{1; r+1\} = n$.

Deterministic Relevant Subset Design (DRSD)

The preceding design is a "proof of concept" design. For the RRSD it is possible to show that the variance of $u_{\mathcal{A}}$ is controlled. However, it may not be the most efficient adaptive design. A deterministic approach is also worth considering.

① Repeat steps 1 of the RRSD algorithm.

② For $r = n\{1\} + 1, \ldots, n$ let $\mathcal{P}_i\{r+1\} = w_i - u_{\mathcal{A}_i\{1;r\}}/\mathcal{N}\{r+1\}$.

③ Set $i\{r+1\} = \arg\max_{i=1,\ldots,d} \mathcal{P}_i\{r+1\}$.

④ Allocate the $\{r+1\}$th observation to the support point to $x_{i\{r+1\}}$.

Preliminaries

① Consider an experiment with two treatment factors $\boldsymbol{x} = (x_1, x_2)^T$, where $x_k = \{0, 1\}$, for k=1,2.

② Let $\eta(\boldsymbol{x}, \boldsymbol{\theta}) = \mathbf{f}^T(\boldsymbol{x})\boldsymbol{\theta}$, where $\mathbf{f}(\boldsymbol{x}) = (1, x_1, x_2, x_1x_2)$ and $\boldsymbol{\theta} = (1, 1, 1, 1)$.

③ A factorial design, denoted ξ_F, places equal weight on the points $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$and allows for the efficient estimation of the main and interaction effects.

④ For $n\{1\} = 8$ the first run places 2 observations on each point in ξ_F.

⑤ It is assumed that the errors are Cauchy distributed.

Results

Variance matrices from 2000 simulations conducted according to the described experiment for the a Fixed, RRSD and DRSD. The inverse of Fisher's information matrix is given for a benchmark comparison. All matrices are scaled by $n = 60$.

$$n\mathrm{Var}[\hat{\boldsymbol{\theta}}_{\mathrm{Fixed}}]$$

$$\begin{pmatrix} 10.1 & -10.2 & -10.2 & 10.6 \\ -10.2 & 20.9 & 10.4 & -21.2 \\ -10.2 & 10.4 & 19.7 & -19.8 \\ 10.6 & -21.1 & -19.9 & 40.7 \end{pmatrix}$$

$$n\mathrm{Var}[\hat{\boldsymbol{\theta}}_{\mathrm{RRSD}}]$$

$$\begin{pmatrix} 9.4 & -9.3 & -9.4 & 9.6 \\ -9.3 & 18.7 & 9.4 & -18.8 \\ -9.4 & 9.4 & 18.8 & -19.3 \\ 9.6 & -18.8 & -19.3 & 38.2 \end{pmatrix}$$

$$n\boldsymbol{F}_{\xi_F}^{-1}$$

$$\begin{pmatrix} 8.0 & -8.0 & -8.0 & 8.0 \\ -8.0 & 16.0 & 8.0 & -16.0 \\ -8.0 & 8.0 & 16.0 & -16.0 \\ 8.0 & -16.0 & -16.0 & 32.0 \end{pmatrix}$$

$$n\mathrm{Var}[\hat{\boldsymbol{\theta}}_{\mathrm{DRSD}}]$$

$$\begin{pmatrix} 8.5 & -8.4 & -8.8 & 8.9 \\ -8.4 & 17.0 & 8.8 & -17.3 \\ -8.8 & 8.8 & 17.7 & -17.6 \\ 8.9 & -17.3 & -17.6 & 34.1 \end{pmatrix}$$

Summary

① Ancillary statistics, specifically, those that constitute a relevant subset, can be incorporated into the design to improve efficiency.

② This was illustrated by developing randomized and deterministic designs that improve the efficiency relative to a fixed design.

Fisher, R., 1934. Two new properties of mathematical likelihood. Proc. Roy. Soc. 144, 285–307.