Information in a Two-Stage Adaptive Design

Adam Lane

University of Missouri

Isaac Newton Institute for the Mathematical Sciences
November 24, 2011

Motivation

- Does considering the dependence of stages in a sequential design effect information?
  - Can an information measure be used to make design parameter decisions.
  - Is there an advantage to considering the dependent information measure in place of the independent measure as an estimate of variance.

A Regression Model with General Mean Function

Suppose n_i subjects are treated at x_i, with weights $w_i = \frac{n_1}{n}$ at each stage, $i = 1, 2$ and responses

$$y_{ij} = \eta(x_i, \theta) + \varepsilon_{ij},\ \varepsilon_{ij} \sim \mathcal{N}(0, 1)$$

with likelihood, regardless of the dependence of stages,

$$\mathcal{L}(\theta|\mathbf{y_1}, \mathbf{y_2}) \propto \exp\left\{-\frac{w_1}{2}\left(\bar{y}_1 - \eta(x_1, \theta)\right)^2 - \frac{w_2}{2}\left(\bar{y}_2 - \eta(x_2, \theta)\right)^2\right\}$$

Adaptive Procedure

Score function per stage

$$S_i = n_i \left(\bar{y}_1 - \eta(x_i, \theta)\right) \frac{d\eta(x_i, \theta)}{d\theta}$$

Choose the second stage treatment by finding

$$x_2 = \arg\max_x \mathrm{Var}_{\bar{y}_2|x} \left[S_2\right]_{\theta=\hat{\theta}_1}$$

where $\hat{\theta}_1$ is the MLE of the first stage data. x_2 is an estimate of the locally optimal design point $x^* = \arg\max_x \mathrm{Var}_{\bar{y}_2|x} \left[S_2\right]$ and provided certain regularity conditions $x_2 \to x^*$ as $n_1 \to \infty$.

Information

Let $S = \sum_{i=1}^{2} S_i$. Then the expected information is

$$M_{dep}(\xi, \theta) = \frac{1}{n}\text{Var}(S) = \frac{1}{n}\left(\text{Var}(S_1) + \text{Var}(S_2) + \text{Cov}(S_1, S_2)\right)$$

$$= \text{E}\left[\sum_{i=1}^{2} w_i \left(\left(\frac{d\eta(x_i,\theta)}{d\theta}\right)^2 - (\bar{y}_i - \eta(x_i,\theta))\frac{d^2\eta(x_i,\theta)}{d\theta^2}\right)\right]$$

$$= w_1\left(\frac{d\eta(x_1,\theta)}{d\theta}\right)^2 + w_2 \text{E}_{\hat{\theta}_1}\left(\frac{d\eta(x_2(\hat{\theta}_1),\theta)}{d\theta}\right)^2,$$

since $\text{Cov}(S_1, S_2) = 0$.

Independent versus Dependent Information in Adaptive Design

- If x_1 and x_2 are considered fixed then

$$M_{ind}(\xi, \theta, \hat{\theta}_1) = w_1 \left(\frac{d\eta(x_1, \theta)}{d\theta} \right)^2 + w_2 \left(\frac{d\eta(x_2(\hat{\theta}_1), \theta)}{d\theta} \right)^2$$

- For a specific function η $M_{ind}(\xi, \theta, \hat{\theta}_1)$ and $M_{dep}(\xi, \theta)$ may have different asymptotic limits.

One Parameter Exponential Mean Function

- Let $\eta(x,\theta) = \exp(-\theta x)$, $0 < \theta < \infty$, and $x_i \in (\overline{\theta}^{-1}, \underline{\theta}^{-1})$, $0 < \underline{\theta} < \overline{\theta} < \infty$. Then the constrained MLE given $\theta \in (\underline{\theta}, \overline{\theta})$

$$\hat{\theta}_1 = \begin{cases} \dfrac{-\ln\ \bar{y}_1}{x_1}, & \text{if } \bar{y}_1 \in \left(e^{-\overline{\theta}x_1}, e^{-\underline{\theta}x_1}\right), \\ \underline{\theta} & \text{if } \bar{y}_1 \geq e^{-\underline{\theta}x_1}, \\ \overline{\theta} & \text{if } \bar{y}_1 \leq e^{-\overline{\theta}x_1}. \end{cases}$$

- The adaptively selected second stage treatment

$$x_2 = \arg\max_x \left[x^2 e^{-2\hat{\theta}_1 x}\right] = \hat{\theta}_1^{-1}$$

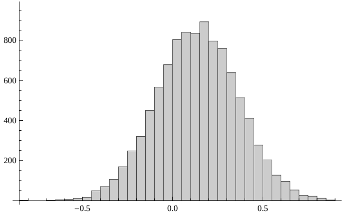
Illustration: $\theta = 1$, $n = 100$, $x_1 = 2$, $\underline{\theta} = 0.10$, and $\overline{\theta} = 4$

Probability of a Boundary Value of $\hat{\theta}_1$ for $w_1 = .2$

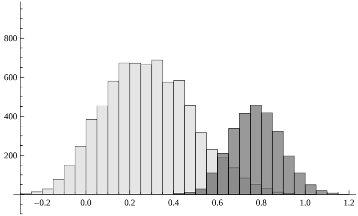
$$\pi_1 = P(\hat{\theta}_1 = \overline{\theta}) = P\left(\bar{y}_1 \leq e^{-\overline{\theta} x_1}\right) = .27$$

$$\pi_3 = P(\hat{\theta}_1 = \underline{\theta}) = P\left(\bar{y}_1 \geq e^{-\underline{\theta} x_1}\right) = .001$$

Mean Responses from 10,000 simulations for $n = 100$, $\theta = 1.0$, $\underline{\theta} = 0.10$, $\overline{\theta} = 4.0$, $x_1 = 2$ and $w_1 = 0.2$.

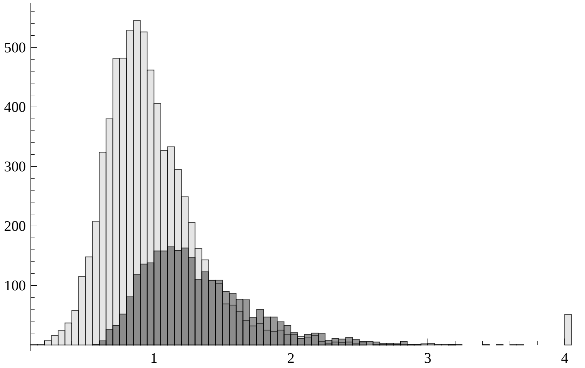


(a) Histogram of $\bar{y}_1$ from Stage 1

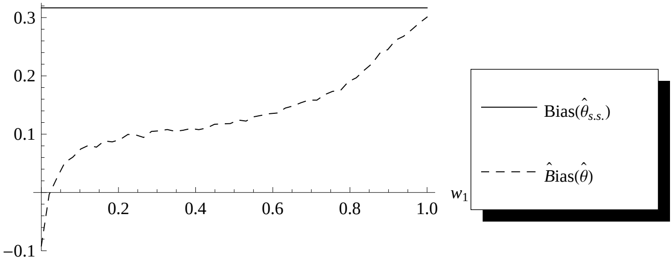


(b) Histogram of $\bar{y}_2$ from Stage 2

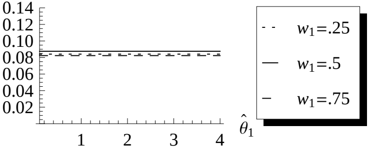
Histogram of the Final MLE $\hat{\theta}$



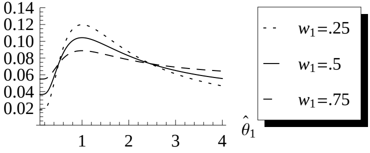
Bias of $\hat{\theta}_{s.s.}$ and Mean of Simulated Bias of $\hat{\theta}$



Information Measures Plotted by $\hat{\theta}_1$ for $w_1 = (.25, .5, .75)$



(d) $M_{dep}(\xi, \theta)$



(e) $M_{ind}(\xi, \theta, \hat{\theta}_1)$

Rao-Cramér Lower Bound

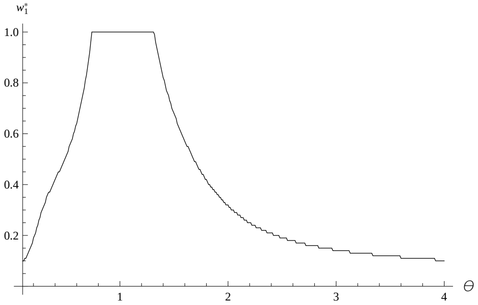
- If $\mathrm{E}(\tilde{\theta}_n) = \theta + b(\theta)$, then

$$\mathrm{Var}\left(\tilde{\theta}_n\right) \geq \frac{Cov(\tilde{\theta}_n, S)}{nM_{dep}(\xi, \theta)} = \frac{(1 + b'(\theta))^2}{nM_{dep}(\xi, \theta)}$$

- This suggest there is a potential proportional relationship between $M_{dep}(\xi, \theta)$ and $\mathrm{Var}^{-1}\left(\hat{\theta}_n\right)$ as a function of w_1 for moderately large n

Optimal Size of Sample Proportion Allocated to Stage One Versus $\theta \in (\underline{\theta}, \overline{\theta})$ where $x_1 = 1$

$$w_1^* = \arg\max_{w_1} M_{dep}(\xi, \theta)$$

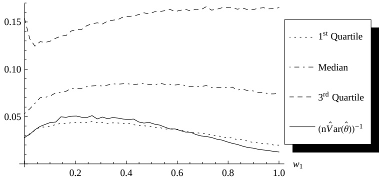


Estimates of Staged Information

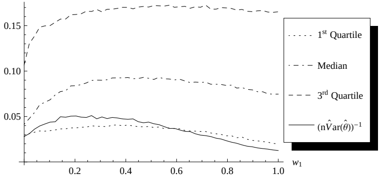
$$M_{ind}\left(\xi,\hat{\theta},\hat{\theta}_1,\theta\right) = w_1 x_1^2 e^{-2\hat{\theta}x_1} + w_2(\hat{\theta}_1^{-2} e^{-2\hat{\theta}\hat{\theta}_1^{-1}})$$

$$M_{dep}\left(\xi,\hat{\theta},\theta\right) = w_1 x_1^2 e^{-2\hat{\theta}x_1} + w_2\left(\hat{\pi}_1 \overline{\theta}^{-2} e^{-2\hat{\theta}\overline{\theta}^{-1}} + \hat{\pi}_3 \underline{\theta}^{-2} e^{-2\hat{\theta}\underline{\theta}^{-1}} + E_{\hat{\theta}_1}\left[\hat{\theta}_1^{-2} e^{-2\theta\hat{\theta}_1^{-1}} | \underline{\theta} < \hat{\theta}_1 < \overline{\theta}\right]_{\theta=\hat{\theta}}\right).$$

Estimates of Information Compared with Simulated $\left[n\widehat{\mathrm{Var}}\left(\hat{\theta}\right)\right]^{-1}$ as Functions of w_1

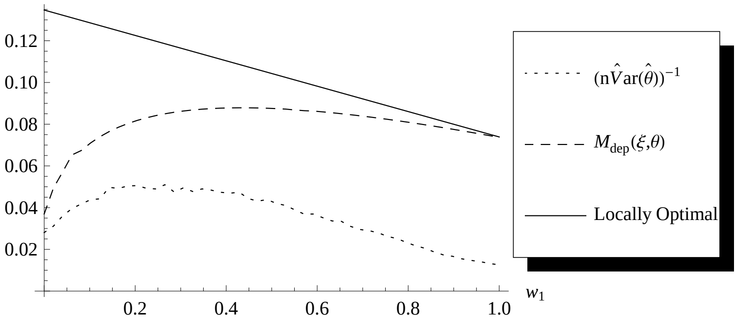


(f) $M_{dep}(\xi, \hat{\theta}, \theta)$

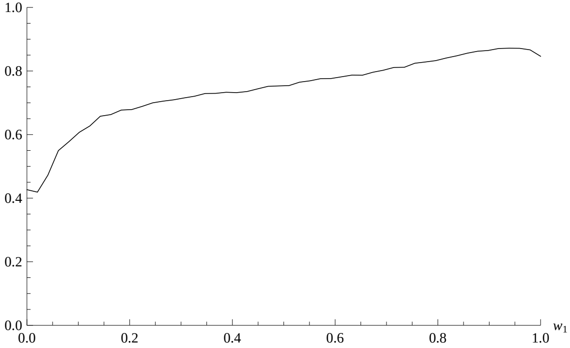


(g) $M_{ind}(\xi, \hat{\theta}, \hat{\theta}_1, \theta)$

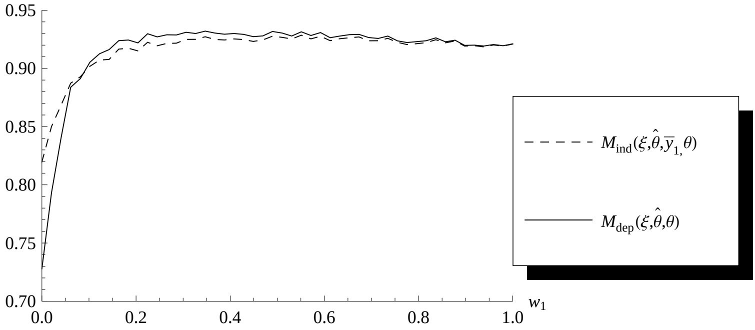
Comparison of $\left[n\widehat{\mathrm{Var}}\left(\hat{\theta}\right)\right]^{-1}$, $M_{dep}(\xi,\theta)$, and Locally Optimal



Percent of Cases Where $M_{dep}(\xi, \hat{\theta}, \theta)$ is Closer to $\left[n\widehat{\mathrm{Var}}\left(\hat{\theta}\right)\right]^{-1}$ than $M_{ind}(\xi, \hat{\theta}, \bar{y}_1, \theta)$



Coverage of $\hat{\theta} \pm 2\hat{M}^{-1/2}$ Using Different Information Estimates, $n = 100$



Conclusions

- There is a potential use for the dependent measure from a design perspective
- From a analysis perspective there does not appear to be a significant difference between the the dependent and independent measure for this specific function

Future Work

- There should exist cases where the dependent and independent measures have different asymptotic limits. Is there a greater effect in terms of analysis between the dependent and independent measure
- Extension to multidimensional parameter space.
- Consider first stage sample as fixed

Thank you!