Information in a Two-Stage Adaptive Optimal Design

Nancy Flournoy

Department of Statistics, University of Missouri

Designed Experiments:
Recent Advances in Methods and Applications
DEMA 2011
Isaac Newton Institute for the Mathematical Sciences
Stanford University, June 14-16, 2011

Motivating Question

For adaptive designs,

- How does the selection of sequential treatments affect the properties of estimators?
- Even if the design is ancillary to the experiment, can it be ignored?

Heuristics Behind Adaptive Optimal Designs

- Optimal designs (e.g., designs that minimize the variance of best dose) are functions of the unknown parameters for nonlinear response functions. So they need to be estimated.
- If MLEs are consistent, in the limit MLEs of the optimal designs will be consistent.
- Hence estimating the optimal design with accruing data from sequential cohorts of subjects will provide increasing efficient designs, and a reasonable overall strategy for treatment allocation.
- This strategy has been proposed frequently in the optimal design literature starting with (before?) Box and Hunter (1963).

Outline: Information in a Two-stage model

1. One Parameter Regression Model with Exponential Mean Function
2. Basic Review for Independent Observations
3. A Two-Stage Design
4. Illustration with Exponential Mean Function
5. Conclusions

Notation

treatments/stages $x_i, i = 1, 2;$

total sample size $n = \sum n_i;$

sample weights $w_i = n_i/n; \ \sum w_i = 1$

design $\{w_i, x_i\}$, n fixed;

responses $\mathbf{y_i} = (y_{i1}, \ldots, y_{in_i});$

expected response $\eta_i = \eta(x_i, \theta);$

mean response $\bar{y}_i = n_i^{-1} \sum_{j=i}^{n_i} y_{ij}$

A Regression Model with Exponential Mean Function

$$y = \eta(x,\theta) + \epsilon, \qquad \epsilon \sim \mathcal{N}(0,1)$$
$$\eta(x,\theta) = \exp(-\theta x), \qquad \theta \in (-\infty,\infty),\ 0 < x \le \overline{b} < \infty$$

- Observe responses $\mathbf{y_i} = (y_{i1}, \ldots, y_{in_i})$ at x_i.
- For two treatments, in canonical exponential family form:

$$\mathcal{L}(\theta, \mathbf{y_1}, \mathbf{y_2} | x_1, x_2) = \prod_{i=1}^{2} f(\theta, \mathbf{y_i}|x_i) \propto \prod_{i=1}^{2} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_i}(y_{ij} - \eta_i)^2\right\}$$
$$\propto exp\left\{\sum_{i=1}^{2} nw_i\left(\eta_i \bar{y}_i - \frac{1}{2}\eta_i^2(x_i,\theta)\right)\right\}$$

A Regression Model

- The probabilities of estimates on the boundaries goes to zero as $n \to \infty$,
- so I refer just to the interior for clarity of exposition.

Notation and Basic Elements: jth subject in ith stage

single unit score function $s_{ij} = s_{ij}\left(y_{ij}|x_i,\theta\right) = \frac{d}{d\theta} ln\ f\left(\theta, y_{ij}|x_i\right) = \left(y_{ij} - \eta_i\right)\frac{d\eta_i}{d\theta} = \left(y_{ij} - \eta_i\right)x_i\eta_i$

within-stage scores $S_i = \sum_{j=1}^{n_i} s_{ij}$;

total score $S = \sum_{i=1}^{2} S_i = \sum_{i=1}^{2} n_i\left(\bar{y}_i - \eta_i\right)\frac{d\eta_i}{d\theta}$

expected unit information

$$\mu_i = \mu(x_i,\theta) = \mathrm{Var}_{y_{ij}|x_i}[s_{ij}] = \mathrm{E}_{y_{ij}|x_i}\left[-\frac{d}{d\theta}s_{ij}\middle| x_i\right] = \mathrm{E}_{y_{ij}|x_i}\left[\left(\frac{d\eta_i}{d\theta}\right)^2 - \left(y_{ij} - \eta_i\right)\frac{d^2\eta_i}{d\theta^2}\middle| x_i\right] = \left(\frac{d\eta_i}{d\theta}\right)^2 = x_i^2\eta_i^2$$

per unit expected information

$$\mathcal{M}(\xi,\theta) = \frac{1}{n}\mathrm{Var}\left[S\right] = \sum_{i=1}^{2} w_i\mu_i = \sum_{i=1}^{2} w_i x_i^2\eta_i^2.$$

MLE approximation

1. $\ln\{\mathcal{L}_n\}$ is twice differentiable in the neighborhood of the true parameter θ_t, so a Taylor expansion of $\ln\{\mathcal{L}_n\}$ yields

$$\ln\{\mathcal{L}_n\} = \ln\{\mathcal{L}_n\}\,|_{\theta=\theta_t} + (\theta-\theta_t)\left(S_{\theta=\theta_t}\right) + \frac{1}{2}(\theta-\theta_t)^2\left(\left.\frac{dS}{d\theta}\right|_{\theta=\tilde{\theta}}\right),$$

   where $\tilde{\theta} \in (\theta_t, \hat{\theta}_n)$.
2. $\mathrm{Max}_\theta\{\ln\{\mathcal{L}_n\}\}$ occurs where $S + (\theta-\theta_t)\frac{d\ S}{d\theta} = 0$.
3. Taking the derivative of $\ln\{\mathcal{L}_n\}$ and rearranging terms, for $\theta = \hat{\theta}$ in the neighborhood of θ_t,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_t\right) \approx \left(-\frac{1}{n}\frac{d\ S}{d\theta}\right)^{-1}\frac{1}{\sqrt{n}}S.$$

Asymptotic Normality of the MLE - Given x_1 and x_2

$$\frac{1}{\sqrt{n}}S = \frac{1}{\sqrt{n}}\left(\sqrt{w}_1\frac{\sum_{j=1}^{n_1} s_{1j}}{\sqrt{n_1}} + \sqrt{w}_2\frac{\sum_{j=1}^{n_2} s_{2j}}{\sqrt{n_2}}\right)$$
$$\sim \mathcal{N}(0, w_1\mu_1 + w_2\mu_2).$$

$$\left(-\frac{1}{n}\frac{d}{d\theta}\frac{S}{\,}\right) = w_1\frac{\sum_{j=1}^{n_1}\frac{d}{d\theta}s_{1j}}{n_1} + w_2\frac{\sum_{j=1}^{n_2}\frac{d}{d\theta}s_{2j}}{n_2} \qquad \text{LLN}$$
$$\xrightarrow[n\to\infty]{as} w_1\mu_1 + w_2\mu_2.$$

By Slutsky's theorem,

$$\left(-\frac{1}{n}\frac{d}{d\theta}\frac{S}{\,}\right)^{-1}\frac{1}{\sqrt{n}}S \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, [w_1\mu_1 + w_2\mu_2]^{-1}\right).$$

Adaptively Selecting the Stage 2 Design Point

Observe $\mathbf{y}_1$ at fixed x_1. Then select the stage 2 design point as

$$x_2 = \arg\max_x \mathrm{Var}_{y_{2j}|x_2}[s_{2j}]\Big|_{\theta=\hat{\theta}_1}$$
$$= \arg\max_x \left(x^2 exp\left\{-2\hat{\theta}_1 x\right\}\right) = min\left\{\hat{\theta}_1^{-1}, \bar{b}\right\}.$$

The MLE from the stage 1 data is

$$\hat{\theta}_1 = -ln\ \bar{y}_1/x_1,$$

if $0 < \bar{y}_1 < 1$; at bounds else

The Adaptive Likelihood

Assuming responses given the treatment are independent of the past, i.e., $f(\mathbf{y}_2|x_2, x_1, \mathbf{y}_1, \theta) = f(\mathbf{y}_2|x_2, \theta)$,
the total likelihood after stage 2 is

$$\mathcal{L}(x_1, x_2, \mathbf{y}_1, \mathbf{y}_2, \theta) = f(\mathbf{y}_2|x_2, \theta) f(x_2|x_1, \mathbf{y}_1, \theta) f(\mathbf{y}_1|x_1, \theta).$$

So long as x_2 is a completely determined by x_1 and $\mathbf{y}_1$, $f(x_2|x_1, \mathbf{y_1}, \theta)$ is a delta function; the design is ancillary.
Note density is no longer member of exponential family:

$$\mathcal{L}(x_1, x_2, \mathbf{y}_1, \mathbf{y}_2, \theta) = f(\mathbf{y}_2|x_2, \theta) f(\mathbf{y}_1|x_1, \theta)$$
$$\propto exp\left\{ nw_1 \left(\eta_1 \bar{y}_1 - \frac{1}{2}\eta_1^2 \right) + nw_2 \left(\eta_2 \left(\bar{y}_1, x_1 \right) \bar{y}_2 - \frac{1}{2}\eta_2^2 \left(\bar{y}_1, x_1 \right) \right) \right\}.$$

Adaptive Expected Information: $\mathrm{Var}\left[s_{ij}\right] = \mathrm{E}\left[-\frac{1}{n_i}\frac{d}{d\theta}s_{ij}\right]$

.

$$\mathrm{E}_{y_{ij}|x_i}\left[-\frac{1}{n_i}\frac{d}{d\theta}s_{ij}\right] = \mathrm{E}_{y_{ij}|x_i}\left[\left(\frac{d\eta_i}{d\theta}\right)^2 - (y_{ij}-\eta_i)\frac{d^2\eta_i}{d\theta^2}\middle| x_i\right]$$
$$= x_i^2\eta_i^2 = \mu\left(x_i,\theta\right).$$

$$\mu(x_2,\theta) = x_2^2\exp^{-2\theta x_2} = \left(\frac{-x_1}{ln\bar{y}_1}\right)^2 exp\left\{-2\theta\left(\frac{-x_1}{ln\bar{y}_1}\right)\right\}.$$

$$\mathrm{E}\left[-\frac{1}{n_i}\frac{d}{d\theta}s_{ij}\right] = \begin{cases}\mu\left(x_1,\theta\right) & \text{if } i=1\\ \mathrm{E}_{\bar{y}_1}\left[\mu\left(\bar{y}_1,\theta\right)\right] & \text{if } i=2.\end{cases}$$

Second stage information

NOTE:

- $\mu(x_2, \theta)$ is random function of $\bar{y}_1$!
- $\mu(x_2, \theta)$ will only converge to a constant only as $\bar{y}_1$ converges to a constant.
- Conditioning on x_2 is equivalent to conditioning of stage 1 responses!!!!

$\sqrt{n}\left(\hat{\theta}_n - \theta_t\right) \approx \left(-\frac{1}{n}\frac{d}{d\theta}S\right)^{-1}\frac{1}{\sqrt{n}}S$

$$f(S) = \int f(S|\bar{y}_1)f(\bar{y}_1)d\bar{y}_1.$$

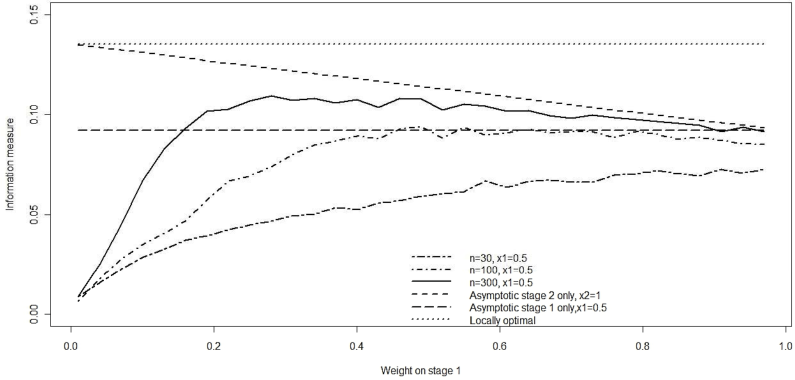
$$\begin{aligned}\frac{1}{\sqrt{n}}S\Big|\,\bar{y}_1 &= \frac{1}{\sqrt{n}}\left(\sqrt{w_1}\frac{\sum_{j=1}^{n_1}s_{1j}}{\sqrt{n_1}} + \sqrt{w_2}\frac{\sum_{j=1}^{n_2}s_{2j}}{\sqrt{n_2}}\right)\Bigg|\,\bar{y}_1 \\ &= \frac{1}{\sqrt{n}}\sqrt{w_1}\frac{\sum_{j=1}^{n_1}s_{1j}}{\sqrt{n_1}} + \mathcal{N}(0, w_2\mu_2)\end{aligned}$$

$$\begin{aligned}\left(-\frac{1}{n}\frac{d}{d\theta}S\right) &= w_1\frac{\sum_{j=1}^{n_1}\frac{d}{d\theta}s_{1j}}{n_1} + w_2\frac{\sum_{j=1}^{n_2}\frac{d}{d\theta}s_{2j}}{n_2} \\ &\xrightarrow[n\to\infty]{as} w_1\mu_1 + w_2\mu_2\end{aligned}$$

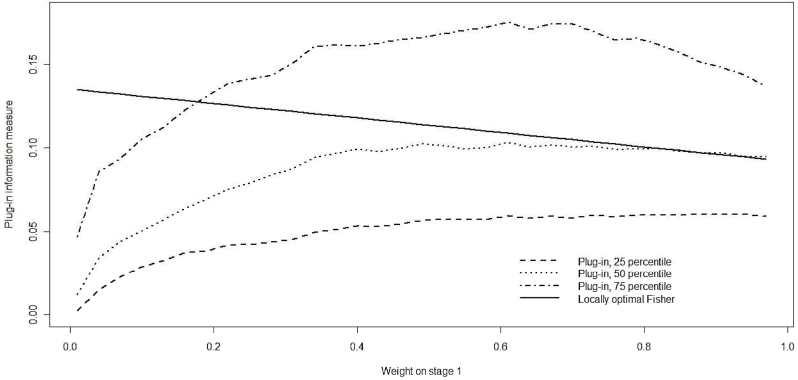
Illustration: $\theta = 1, \quad x \in (.01, 100)$

- $x_1 = 0.5$
- optimal $x_2 = \arg\max_x \mathrm{Var}_{y_{2j}|x_2}[s_{2j}]\Big|_{\theta=1} = 1.0;$
- adaptive $x_2 = \arg\max_x \mathrm{Var}_{y_{2j}|x_2}[s_{2j}]\Big|_{\theta=\hat{\theta}_1}$

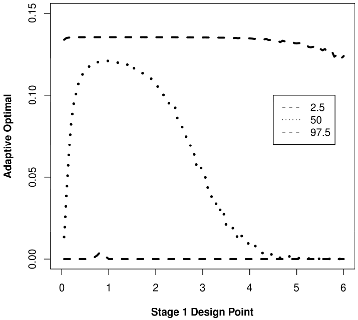
Asymptotic Fisher for $x = .5$ and $x = 1.0$ alone; two-sample locally optimal and median two-stage plug-in estimates for $n = 30, 100, 300$ versus w_1 at $x = .5$.



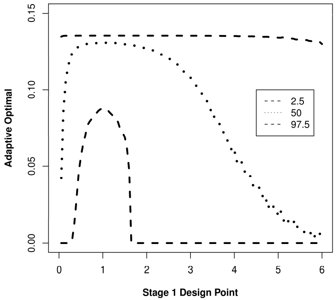
Two-sample locally optimal Fisher and percentiles of two-stage plug in estimates for $n = 1000$ versus w_1 at $x = .5$.



Stage 2 2.5th, 50th and 97.5th Percentiles of μ_2; $n_1 = n_2 = .5$



(a) $n_i = 30$



(b) $n_i = 100$

Conclusions

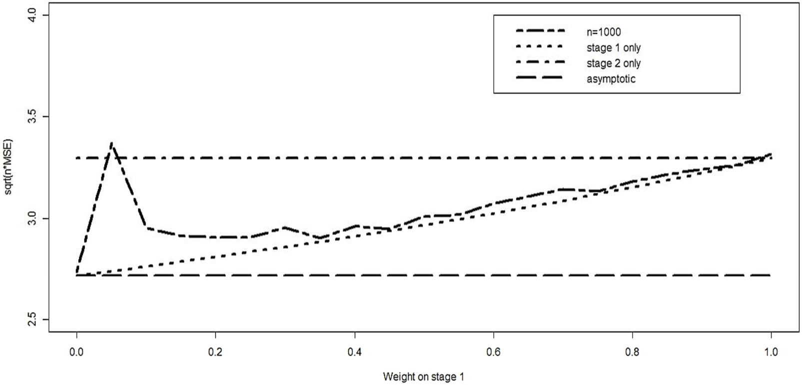
- The locally optimal adaptive design is ancillary, but informative.
- The conditional incremental information after the first stage is a random variable depending on stage one observations.
- The conditional incremental information does not achieve the Cramer-Rao Bound
- MLEs from the locally optimal adaptive design do not have the hoped for optimality, and if the stage one design has a small sample size, their variance is random.

Thank you!

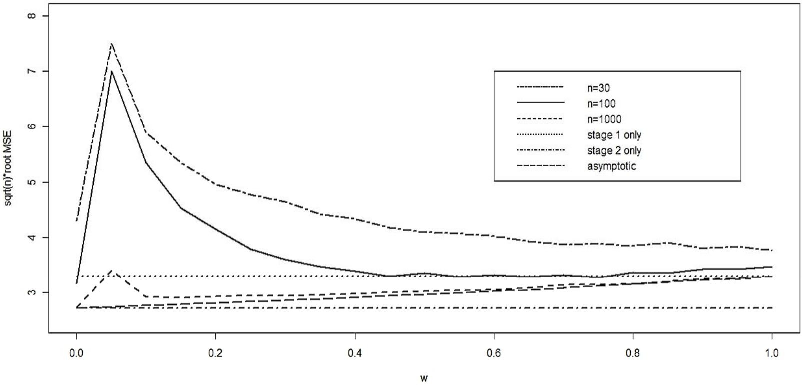
References

Yao, P, Flournoy, N. (2010) Information in a Two-stage Adaptive Optimal Design for Normal Random Variables having a One Parameter Exponential Mean Function. MoDa 9 229-236. Springer (eds. Giovagnoli, A., Atkinson, A.C., Torsney, B., May, C.).

Asymptotic Fisher^{-1} for $x = .5$ and $x = 1.0$ alone; two-sample locally optimal and two-stage $\sqrt{n} \times MSE(\hat{\theta})$, $n = 1000$ versus w_1 at $x = .5$.



Asymptotic Fisher^{-1} for $x = .5$ and $x = 1.0$ alone; two-sample locally optimal and two-stage $\sqrt{n} \times MSE(\hat{\theta})$, $n = 30, 100, 1000$ versus w_1 at $x = .5$.



Remarks

- The $max_{x_1} \{\mu_2\} = 0.135$, which is the asymptotic Fisher's information.
- The 97.5th percentiles of μ_2 attain 0.135 at all but the highest values of x_1 for $n = 100$ and 30.
- In contrast, the 97.5th percentile of $-\frac{d}{d\theta}s_{2j}$ is greater than 0.135 except for values of x_1 somewhat less than one.
- Furthermore, $-\frac{d}{d\theta}s_{2j}$ is negative with high probability.

Remarks

- The median of μ_2, attains its maximum value when $x_1 = 1$ for $n = 100$ and 30.
- The median of μ_2 comes closer to 0.135 at $x_1 = 1$ as the sample size increases.
- Indeed, the median of μ_2 is close to 0.135 for a range of values of x_1 that includes $x_1 = 1$; this range is larger for for $n = 100$ than for $n = 30$.
- For $n = 30$, the 2.5th percentile of μ_2 is zero, except for a very small blip for x_1 just less than one; however, for $n = 100$, the 2.5th percentile of μ_2 is nearly quadratic for $x_1 \in (0.2, 1.8)$ with its maximum approximately 50% of 0.135.