An Adaptive Optimal Design with a Small Fixed Stage One Sample Size.

Adam Lane and Nancy Flournoy

Universities of Cincinnati and Missouri

Design and Analysis of Experiments in Healthcare
Issac Newton Institute of Mathematical Sciences
Cambrige, UK

July 6, 2015

Outline

Outline

Adaptive Optimal Designs

1. The first experiment is the optimal design assuming prior estimates of model parameters are true.

2. Each subsequent experiment is the optimal design assuming updated estimates of model parameters are true.

****

Maximum likelihood estimates are normally distributed under standard regularity conditions *under the assumption that the sample sizes at each stage go to infinity*.

****

Adaptive Optimal Designs with A First Stage having a Fixed Sample Size (like Pilot Data)

Motivation: Pilot studies (stage 1) with a small sample size preceeding a larger experiment (stage 2) are common.

The assumption that $n_1 \to \infty$ and $n_2 \to \infty$ as $n \to \infty$, (the traditional basis for inference in adaptive optimal design) is **not valid here.**

We approximate the distribution of MLEs by keeping the sample size of the small first stage fixed and letting $n_2 \to \infty$.

Outline

Normal Regression Model with a General Mean Function

Suppose n_i subjects are treated at x_i, with weights $w_i = n_i/n$ at each stage, $i = 1, 2$ and responses

$$y_{ij} = \eta(x_i, \theta) + \varepsilon_{ij},\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2). \tag{1}$$

Assuming x_2 is a many to one, onto function of $(y_1, \ldots, y_{n_1}, x_1)$,

$$f_{\mathbf{y_1},\mathbf{y_2}|x_1}(\mathbf{y_1}, \mathbf{y_2}|x_1, \theta) = f_{\mathbf{y_2}|\mathbf{y_1},x_1}(\mathbf{y_2}|\mathbf{y_1}, x_1, \theta) f_{\mathbf{y_1}|x_1}(\mathbf{y_1}|x_1, \theta)$$

$$\propto \exp\left\{-\frac{w_1}{2\sigma^2}\left(\bar{y}_1 - \eta(x_1, \theta)\right)^2 - \frac{w_2}{2\sigma^2}\left(\bar{y}_2 - \eta(x_2, \theta)\right)^2\right\}$$

is not a member of the exponential family.

Rationale for Using Fisher Information, $M = \mathrm{Var}[S]$, in Design and Inference

Modivation is provided by the *Cramèr-Rao Lower Bound*:

Suppose that $\tilde{\theta}_n$ is an estimator of θ based on n trials with finite expectation $\mathrm{E}\left[\tilde{\theta}_n\right] = \theta + b(x_1, \theta)$.

The Cramèr-Rao Lower Bound is

$$\mathrm{Var}\left[\tilde{\theta}_n\right] \geq \frac{\left[\mathrm{Cov}\left[\tilde{\theta}_n, S\right]\right]^2}{\mathrm{Var}[S]},$$

where $S = \partial \ln \mathscr{L}\left(\theta|\mathbf{y_1}, \mathbf{y_2}\right)/\partial\theta$ and $\mathrm{Cov}\left[\tilde{\theta}_n, S\right] = 1 + \frac{\partial}{\partial\theta} b(x_1, \theta)$.
Equality is achievable only for exponential family members.

Adaptive Optimal Second Stage Treatment x_2

The total score function $S = \partial \ln \mathscr{L}(\theta|\mathbf{y_1}, \mathbf{y_2}) / \partial\theta = \sum\sum s_{ij}$;

$$s_{ij} = \frac{1}{\sigma^2}(y_{ij} - \eta(x_i, \theta)) \frac{\partial \eta(x_i, \theta)}{\partial \theta}.$$

is the score function for the j^{th} subject in the i^{th} stage.

- Let the *optimal design point for stage 2* be $x^*(\theta) = \arg\max_x \mathrm{Var}_{\bar{y}_2|x}[S_2] = \arg\max_x \mathrm{Var}_{y_{2j}|x}[s_{2j}]$;
- $\hat{\theta}_{n_1}$ denotes the MLE of θ based on the first stage data.
- $x_2\left(\hat{\theta}_{n_1}\right)$ is the MLE of the locally optimal design point.
- Design Rule: Select the second stage treatment x_2 to be

$$x_2\left(\hat{\theta}_{n_1}\right) = \arg\max_x \left[\mathrm{Var}_{\bar{y}_2|x}[s_{2j}]\right]_{\theta=\hat{\theta}_{n_1}}$$

Outline

Fisher's Information for an adaptive design $\xi_A = \{x_i, w_i\}_{i=1}^2$

USES of M: (1) Design (2) Estimate Var$[\widehat{\theta}]$

$$\mathrm{M}(\xi_A, \theta) = \frac{1}{n}\mathrm{Var}[S] = \frac{1}{n}\left(\mathrm{Var}[S_1] + \mathrm{Var}[S_2] + 2\mathrm{Cov}[S_1, S_2]\right)$$

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

$$= \frac{w_1}{\sigma^2}\left(\frac{\partial\eta(x_1,\theta)}{\partial\theta}\right)^2 + \frac{w_2}{\sigma^2}\mathrm{E}_{\hat{\theta}_{n_1}}\left[\frac{\partial\eta(x_2(\hat{\theta}_{n_1}),\theta)}{\partial\theta}\right]^2,$$

since $\mathrm{Cov}[S_1, S_2] = 0$.

Benchmark Upper Bound on Information

The maximum amount of information that can be collected is from $\xi^* = \{(x_1, n_1), (x^*(\theta), n_2)\}$:

$$M(\xi^*, \theta) = \frac{1}{\sigma^2} \left[w_1 \left(\frac{\partial \eta(x_1, \theta)}{\partial \theta} \right)^2 + w_2 \left(\frac{\partial \eta(x^*(\theta), \theta)}{\partial \theta} \right)^2 \right]$$

i.e., $M(\xi^*, \theta) \geq \mathrm{M}(\xi_A, \theta)$.

From Cramèr-Rao Lower Bound

$$n\mathrm{Var}\left[\hat{\theta}_n\right] \geq \left[\mathrm{M}(\xi_A, \theta)\right]^{-1} \geq \left[M(\xi^*, \theta)\right]^{-1}.$$

Outline

Approximate Distributions of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$

- If $n_1 \to \infty$ and $n_2 \to \infty$, with $\xi^* = \{(x_1, n_1), (x^*(\theta), n_2)\}$

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathscr{D}} \mathscr{N}\left(0, \left[M\left(\xi^*, \theta\right)\right]^{-1}\right).$$

- If n_1 is fixed, then

$$\mathscr{N}\left(0, \left[M(\xi_A, \theta)\right]^{-1}\right)$$

is a better approximation, because (by CRB) $[M(\xi_A, \theta)]^{-1}$ is closer to $nvar[\widehat{\theta}_n]$ than is $\left[M\left(\xi^*, \theta\right)\right]^{-1}$.

- Conditioning on the design, (i.e., using only the second stage data), the MLE had very poor performance

Lane and Flournoy (2012) get a Better Approximation

They show that as $n_2 \to \infty$,

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathscr{D}} UQ;$$

$Q \sim \mathscr{N}(0, \sigma^2)$;

$U = \left(\dfrac{\partial \eta(x_2, \theta)}{\partial \theta}\right)^{-1}$ is a random function of $\bar{y}_1$.

This is a scale mixture of normal random variables;
a hierarchical model.

U provides a measure of the *cost* of using stages.

Outline

Example: Exponential Mean Function

$$M(\xi^*, \theta) = w_1 x_1^2 e^{-2\theta x_1} + w_2(\theta^{-2} e^{-2})$$

$$M(\xi_A, \theta) = \frac{1}{\sigma^2} \Bigg(w_1 x_1^2 e^{-2\theta x_1} + w_2 \pi_a a^2 e^{-2\theta a} + w_2 \pi_b b^2 e^{-2\theta b}$$
$$+ w_2 \mathrm{E}_{\bar{y}_1} \left[\left(\frac{-x_1}{\log \bar{y}_1} \right)^2 e^{-2\theta\left(\frac{-x_1}{\log \bar{y}_1}\right)} \cdot \mathrm{I}\left(e^{-a^{-1}x_1} < \bar{y}_1 < e^{-b^{-1}x_1} \right) \right] \Bigg)$$

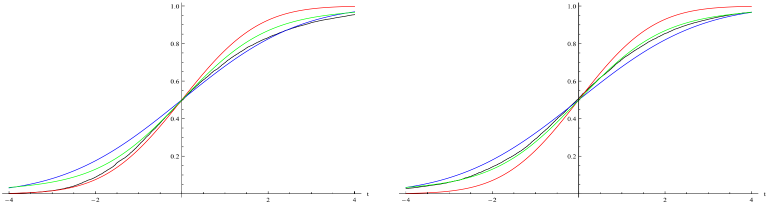
where $\pi_a = \Phi\left(\sqrt{n_1}\left(e^{-a^{-1}x_1} - e^{-\theta x_1}\right)/\sigma\right)$

and $\pi_b = 1 - \Phi\left(\sqrt{n_1}\left(e^{-b^{-1}x_1} - e^{-\theta x_1}\right)/\sigma\right)$.

U was obtained analytically,

so the limiting CDF UQ can be calculated numerically:

CDF Approximations of $\sqrt{n}(\hat{\theta}_n - \theta)$



(a) $n = 30$ (b) $n = 1000$

The 'true' CDF of $\sqrt{n}(\hat{\theta}_n - \theta)$ (black curve) via Monte Carlo.

$P(T_1 \leq t)$, $T_1 \sim \mathcal{N}\left(0, [M(\xi^*, \theta)]^{-1}\right)$.

$P(T_2 \leq t)$, $T_2 \sim \mathcal{N}\left(0, [M(\xi_A, \theta)]^{-1}\right)$.

$P(T_3 \leq t)$, $T_3 \sim UQ$.

$\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = 0.5$, $a = 0.25$ and $b = 4$.

Integrated |difference| of approx. & true CDFs (×100)

$\theta = 0.5$	$n = 50$			$n = 100$			$n = 400$		
n_1	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	19	20	**11**	15	16	**7**	14	15	**5**
10	11	13	**8**	9	11	**6**	7	12	**4**
15	9	10	**8**	6	9	**5**	4	10	**3**

$\theta = 1.0$	$n = 30$			$n = 100$			$n = 400$		
n_1	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	30	33	**25**	39	27	**16**	39	21	**8**
10	40	40	**32**	23	28	**16**	26	20	**8**
15	34	34	**33**	21	23	**15**	18	20	**9**

$\theta = 1.5$	$n = 30$			$n = 100$			$n = 400$		
n_1	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	32	33	**31**	42	**21**	23	42	**17**	21
10	34	33	**22**	32	22	**10**	35	19	**12**
15	35	35	**32**	26	22	**13**	28	21	**7**

Integrated |difference| of approx. & true CDFs($\times$100)

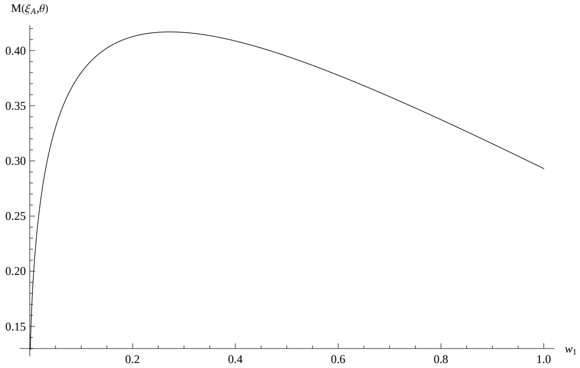
$\theta = 1.0$	$n = 400$		$n = 1000$	
n_1	T_2	T_3	T_2	T_3
5	21	**6**	18	**6**
10	20	**8**	21	**7**
15	20	**9**	20	**8**
50	13	**9**	13	**5**
100	10	**8**	8	**4**
200	**11**	14	**4**	7

CDF differences of $T_2 \sim \mathcal{N}\left(0, [M(\xi_A, \theta)]^{-1}\right)$ and $T_3 \sim UQ$ versus $\sqrt{n}(\hat{\theta}_n - \theta)$.

$x_1 = 2$, $\sigma = 0.5$, $a = 0.25$ and $b = 4$.

Outline

Locally Optimal Sample Allocation



Fisher's information, $M(\xi_A, \theta)$, where $\xi_A = \{x_i, w_i\}_{i=1}^{2}$ plotted as a function of w_1.

Locally Optimal Sample Allocation

The *locally optimal stage one sample size* for fixed n can be written as

$$n_1^*(\theta) = \arg \max_{n_1 \in \{1,\ldots,n\}} \mathrm{M}\left(\xi_A, \theta\right).$$

The table below presents $n_1^*(\theta)$ for the scenarios on the preceding slide

θ	$n = 30$	$n = 100$	$n = 400$
0.5	30	100	400
1.0	12	55	61
1.5	6	16	49

Summary Comments

In adaptive design,

- the design is ancillary but not independent of the responses.
- to *condition on the design* is to condition on the responses from prior stages.
- For a two stage design in which the first stage sample size is fixed
  - The unconditional $(\mathrm{Var}[S])^{-1}$ is closer to $n\mathrm{Var}[\widehat{\theta}]$ than if one fails to take the expectation over the stage 2 design point.

Summary Comments: $\sqrt{n}(\widehat{\theta} - \theta) \sim UN$

The asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ is a scale mixture of normal distributions.

This scale mixture of normals also results from

- several urn designs
  - e.g., the birth and death urn of Ivanova et. al and the randomized reinforcement urns of Li & Flournoy, May & Flournoy and others
  - U comes from the random samples sizes on the individual treatements.
- some interim analysis & sample size recalculation designs (recent work of Sergey Tarima).

Note: standard normality results hold for random walk rules (true up-and-down designs) because sample size proportions on each treatment go to constants.

Summary Comments

- We introduce the notion of *locally optimal sample sizes for stages*.

  This provides a way to study and analyze the effects of sample size choices on design efficiency.

Thank you!