Optimizing Sequential Design of Experiments

by

Michael Baron

University of Texas at Dallas

IBM Research Division

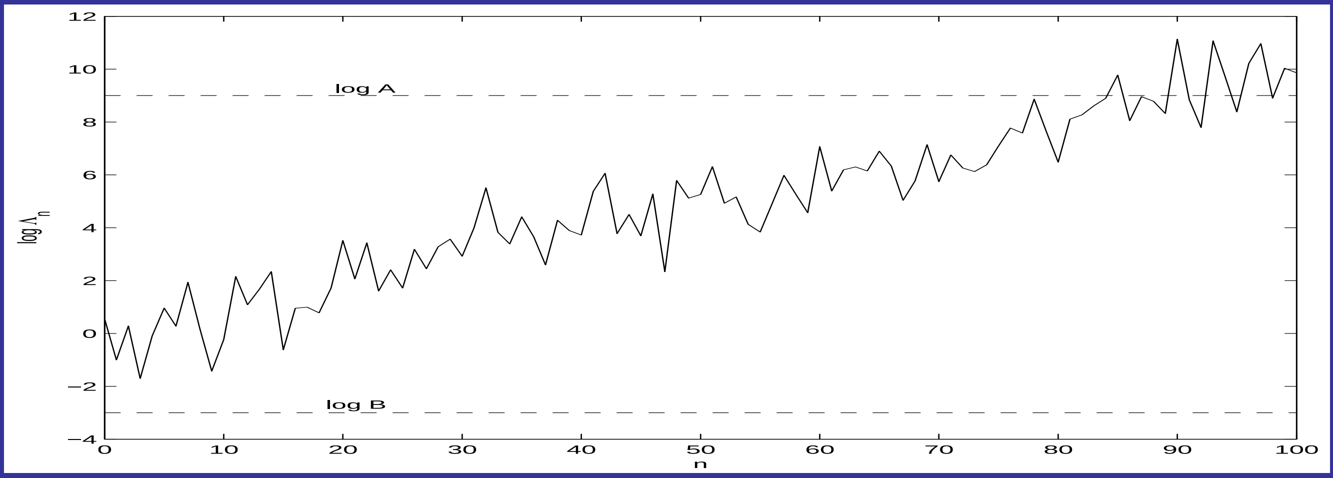
Claudia Schmegner

University of Texas at Dallas

IBM T. J. Watson Research Ctr.,
Yorktown Heights, NY May 22, 2003

Abstract

It is often impractical or expensive to sample according to the classical sequential scheme, that is, one observation at a time. Sequential planning extends and generalizes the "pure" sequential procedures by allowing to sample observations in groups. At any moment, all the collected data are used to determine the size of the next group and to decide whether or not sampling should be terminated. We discuss optimality of sequential designs taking into account both the variable and the fixed cost of experiments. Some general guidelines for optimal sequential planning are established. It is shown that the total cost of standard sequential procedures can be reduced significantly without increasing the loss. Specific types of sequential plans are introduced and compared, some existing plans are modified and improved.



Main principle

Based on the collected data, we ...

- Non-sequential (retrospective) analysis

  ... *decide what to report*

- Sequential (pure sequential) analysis

  Sample one observation at a time

  ... *decide* $\left\{ \begin{array}{l} \textit{when to stop sampling} \\ \textit{what to report} \end{array} \right.$

- Sequential planning

  Sample a group of observations at a time

  ... *decide* $\left\{ \begin{array}{l} \textit{how to sample} \\ \textit{when to stop sampling} \\ \textit{what to report} \end{array} \right.$

For example...

Hypothesis Testing

Based on $X_1, \ldots, X_n$*, we ...*

- Non-sequential (retrospective) analysis

  ... accept or reject

- Sequential (pure sequential) analysis

  ... accept, reject, or collect X_{n+1}

- Sequential planning

  ... accept,
  reject,
  collect X_{n+1}*,*
  collect (X_{n+1}, X_{n+2})*,*
  collect $(X_{n+1}, X_{n+2}, X_{n+3})$*,*
  collect $(X_{n+1}, X_{n+2}, X_{n+3}, X_{n+4})$*, etc.*

Risk Function

- Non-sequential (retrospective) analysis

  $$Risk = E(loss)$$

- Sequential (pure sequential) analysis

  $$Risk = E(loss + cost\ of\ observations)$$

- Sequential planning

  $$Risk = E(loss + cost\ of\ observations + cost\ of\ sampled\ groups)$$

Formally ...

$\mathbf{X}_k = (X_1, \ldots, X_k)$ = data

N_0 = size of the 1^{st} group, const or random

$N_{j-1} = N_{j-1}(\mathbf{X}_{M_{j-1}})$ = size of the j^{th} group, after observing $\mathbf{X}_{M_{j-1}}$

$M_k = \sum_{j=1}^{k} N_j$

<u>Non-randomized sequential plan</u>

$$N = \left\{ N^{(k)} : \mathcal{X}^k \to \{0, 1, 2, ...\} \right\}$$

data → size of the next group

<u>Randomized sequential plan</u>

$$N = \left\{ \mathcal{P}^{(k)} : \mathcal{X}^k \to \{p_0, p_1, p_2, ...\}, \sum p_j = 1 \right\}$$

data → distribution on $\{0, 1, 2, ...\}$

$$\begin{aligned} T &= \min\{j \geq 1 | N_j = 0\} \\ &= \text{total number of sampled groups} \\ &= \text{“stopping time”} \end{aligned}$$

Decision theory

$L(\theta, \delta)$ = loss function

$$\left.\begin{array}{lcl} c & = & \text{cost of each observation} \\ a & = & \text{fixed cost of each group} \end{array}\right\} \begin{array}{c} \text{standard} \\ \text{simplification} \end{array}$$

$$\begin{aligned} R(\theta, N) &= E^X \left\{ L(\theta, \delta) + c \sum_{j=1}^{T} N_j + aT \right\} \\ &= E^X \left\{ L(\theta, \delta) + \sum_{j=1}^{T} (cN_j + a) \right\} \\ &= \text{risk function} \end{aligned}$$

$\pi(\cdot)$ - the prior distribution of θ

$r(\pi, N) = E^\pi R(\theta, N)$ = Bayes risk

For a randomized plan,

$$R(\theta, N) = E^X E^{\{\mathcal{P}\}} \left\{ L(\theta, \delta) + \textstyle\sum_1^T (cN_j + a) \right\}$$

Rather standard ...

A plan N is $R-better$ than a plan $\tilde{N}$ if

$$\begin{cases} R(\theta, N) \leq R(\theta, \tilde{N}) \text{ for any } \theta \in \Theta, \\ R(\theta, N) < R(\theta, \tilde{N}) \text{ for some } \theta \in \Theta. \end{cases}$$

A plan N is $admissible$, if there is no plan $\tilde{N}$ that is R-better than N.

A plan N is $minimax$ if

$$\sup_{\theta} R(\theta, N) \leq \sup_{\theta} R(\theta, \tilde{N})$$

for any plan $\tilde{N}$.

A plan N is $preferred$ to a plan $\tilde{N}$ if

$$r(\pi, N) < r(\pi, \tilde{N})$$

A plan N is $Bayes$, if

$$r(\pi, N) \leq r(\pi, \tilde{N})$$

for any plan $\tilde{N}$.

The question is ...

How to choose $N_j(X_{M_j})$?

Examples

R. Lewis, D. Berry (JASA, 1994) and many others -

$$N_j \equiv N_0$$

L. Hayre (JRSS-B, 1985) -

$$N_j \approx pE\left(\begin{array}{c}\text{number of observations}\\ \text{needed to cross the}\\ \text{stopping boundary}\end{array}\right)$$

[expectation-based plan]

$$N_j \approx \text{ such that } P(\text{cross boundary}) \approx p$$

[quantile-based plan]

Examples, continued

D. Assaf (Annals of Statistics, 1988)

N_j minimize *P(false alarm)* under the fixed average sampling rate ("dynamic sampling" for change detection)

N. Schmitz (Springer-Verlag, 1993)

Existence of Bayes sequential plans

M. Roters (Sequential Analysis, 2002)

Extension to continuous time

Examples ... the least that can be done

Conservative plan

$$N_j^{(0)} = \min \left(\begin{array}{c} \text{number of observations} \\ \text{needed to cross the} \\ \text{stopping boundary} \end{array} \right)$$

$$M_{conserv} = M_{pure\ seq}$$

where T is the final sample size

m-conservative plan

Choose $N_j^{(m)} \in [N_j^{(0)}, N_j^{(0)} + m]$

$$M_{m-conserv} \leq M_{pure\ seq} + m$$

They solve a variational problem

Example: Bernoulli(θ); test θ_0 vs $1-\theta_0$. There are $2^{\frac{m(m+1)}{2}-1}$ m-conservative plans. One can find the optimal plan.

General guidelines

Let N be a randomized plan. Then there exists a **non-randomized plan** whose Bayes risk does not exceed the Bayes risk of N.

(There exist non-randomized Bayes plans)

• • • • • •

For any sequential plan N, there exists a sequential plan $N^\star$ such that

- $N^\star(X_1, \ldots, X_k) = N^\star(k, S_k(X_1, \ldots, X_k))$, (depends on the collected data only through the **sufficient statistic**)
- $r(\pi, N^\star) \leq r(\pi, N)$

(There exists a Bayes sequential plan that is based on the sufficient statistic)

Therefore ... (corollaries)

- In optimal sequential planning with iid observations, the choice of each group size is independent of the order in which the collected data were obtained.

- (Irrelevance of past costs) In optimal sequential planning, the choice of each group size is independent of the amount already paid in sampling costs for the collected groups of observations.

Following the guidelines...

Sequentially Planned Probability Ratio Test (SPPRT)

Test $H_0 : \theta = \theta_0$ vs $H_A : \theta = \theta_1$.

Having observed $X_1, \ldots, X_n$, compute

$$\Lambda_n = \frac{f(X_1, \ldots, X_n|\theta_1)}{f(X_1, \ldots, X_n|\theta_0)}$$

A, B - given constants such that $B < 1 < A$

If $\Lambda_n \geq A$, stop and reject H_0

If $\Lambda_n \leq B$, stop and do not reject H_0

If $\Lambda_n \in (B, A)$, take another sample,

$$X_{n+1}, \ldots, X_{n+N(\Lambda_n)},$$

where $N(\Lambda_n)$ is the size of the next sample, having observed $X_1, \ldots, X_n$,

$$\begin{aligned} N(\boldsymbol{X}_n) &= N(\Lambda_n) \\ &= \begin{cases} 0 & \text{if } \Lambda_n \notin (B, A) \\ > 0 & \text{otherwise} \end{cases} \end{aligned}$$

Agreement with the Sufficiency Principle

$$
\begin{aligned}
\Lambda_n &= \frac{f(X_1, \ldots, X_n | \theta_1)}{f(X_1, \ldots, X_n | \theta_0)} \\
&= \frac{g(S_n(\mathbf{X}_n) | \theta_1) h(\mathbf{X}_n)}{g(S_n(\mathbf{X}_n) | \theta_0) h(\mathbf{X}_n)} \\
&= \frac{g(S_n(\mathbf{X}_n) | \theta_1)}{g(S_n(\mathbf{X}_n) | \theta_0)}
\end{aligned}
$$

Optimizing ...

Risk = E(loss + cost of observations + cost of sampled groups)

First risk component (loss).
Error Probabilities and Stopping Boundaries

$\alpha = P_{\theta_0}\left\{\Lambda_{M_T} \geq A\right\}$ = prob. of Type I error

$\beta = P_{\theta_1}\left\{\Lambda_{M_T} \leq B\right\}$ = prob. of Type II error

---

Result:

$$A \leq \frac{1-\beta}{\alpha} \quad \text{and} \quad B \geq \frac{\beta}{1-\alpha}$$

Hence:

$$\alpha \leq e^{-A}, \quad \beta \leq e^{B}$$

Second risk component
Total sample size, $E(M_T)$

- *Pure sequential plan*

$$\boldsymbol{E}_\theta(M_T) = ARL(\theta)$$

$$\approx \frac{(1 - OC(\theta) \ln A + OC(\theta) \ln B}{\boldsymbol{E}_\theta \ln \dfrac{f(x, \theta_1)}{f(x, \theta_0)}}$$

- *Conservative plan*

$$\boldsymbol{E}_\theta(M_T) = \text{ the same, of course}$$

- *m-conservative plan*

$$\boldsymbol{E}_\theta(M_T) = \text{ differs by at most } m$$

Third risk component
Expected number of groups, $E(T)$
(That's where we gain!)

- *SPRT, $E(T)$ is* ***linear***

$$\lim_{A\to\infty, \ln A = O(\ln B)} \frac{\boldsymbol{E}_{\text{SPRT}}(T \,|H_1)}{\ln A} = \frac{1}{\mathcal{K}(\theta_1, \theta_0)}$$

$$\lim_{B\to 0, \ln B = O(\ln A)} \frac{\boldsymbol{E}_{\text{SPRT}}(T \,|H_0)}{|\ln B|} = \frac{1}{\mathcal{K}(\theta_0, \theta_1)}$$

- *Conservative SPPRT,*

  $E(T)$ is *logarithmic!*

$$\boldsymbol{E}_{SPPRT}(T\,|H_1)$$

$$\leq \frac{1+U/\mathcal{K}}{|\ln(1-\mathcal{K}/\mathcal{U})|}(\ln\ln A - \ln\ln\ln A) + O(1),$$

as $A \to \infty$, $AB = O(1)$,

$$\boldsymbol{E}_{SPPRT}(T\,|H_0)$$

$$\leq \frac{1-L/\mathcal{K}}{|\ln(1+\mathcal{K}/\mathcal{L})|}(\ln\ln|B|-\ln\ln\ln|B|)+O(1),$$

as $B \to 0$, $AB = O(1)$,

where $P\left\{L < \ln\frac{f(x,\theta_1)}{f(x,\theta_1)} < U\right\} = 1.$

- *Expectation-based, quantile-based plans*

  $E(T)$ *is* ***bounded!!!***

If

$$\boldsymbol{P}\left\{T = k + 1 \,|\, \boldsymbol{X}^{M_k}, T > k\right\} \geq p$$

a.s., for some $p > 0$ and all k, then

(i) T is proper, i.e. $P(T = \infty) = 0$;

(ii) $\boldsymbol{E}(T) \leq 1/p$.

Applications:

- quantile-based plans

- expectation-based plans

Sequential Planning on a Lattice

Suppose that $\ln \Lambda_n = k\Delta$, $k \in \boldsymbol{Z}$

Example: $H_0 : \theta = \theta_0$ vs $H_A : \theta = 1 - \theta_0$

Then $\left\{N_j\right\}$ is a stationary Markov chain with a transition probability matrix

$$P = \{P_{kn}\},$$

$$P_{kn} = P(\ln \Lambda_{M_j} = n\Delta | \ln \Lambda_{M_{j-1}} = k\Delta)$$

Computation of the risk

Expected number of groups

$$E(T) = (I - P)^{-1} \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}$$

Expected number of observations

$$E(M_T) = (I - P)^{-1} \begin{pmatrix} N_B \\ N_{Be\Delta} \\ N_{Be^2\Delta} \\ \dots \\ N_A \end{pmatrix}$$

Risk

$$R(\theta, N) = E^X L(\theta, \delta^B) + (I - P)^{-1}(a\boldsymbol{e} + c\boldsymbol{N})$$

Case Study: Pediatric Research for institutional review boards (IRB)

Need to know if a certain medication is efficient for at least 50

Test

$$H_0 : \theta \geq 0.52 \quad vs \quad H_A : \theta \leq 0.48$$
$$\Updownarrow$$
$$H_0 : \theta = 0.52 \quad vs \quad H_A : \theta = 0.48$$

<u>Costs</u>: \$600 per IRB + \$75 per trial

$$a = 75, \quad c = 600$$

$$\ln\left(\frac{\theta_0}{\theta_1}\right) = 0.0800 = -\ln\left(\frac{1-\theta_0}{\theta_1}\right)$$

$\Rightarrow$ SPPRT on a lattice with $\Delta = 0.08$

For $\alpha = \beta = 0.05$, $A = 19$ and $B = 1/19$
$\Rightarrow \ln(A) = 36\Delta$ and $\ln(B) = -36\Delta$

Plan	Pure	Conserv.	5th percentile
E(T)	804	45	20.2
E(M)	804	804	806
aE(T)+cE(M)	543,015	486,041	485,115
Saving ($)	0	56,974	57,900

For $\alpha = \beta = 0.01$, $A = 99$ and $B = 1/99$
$\Rightarrow \ln(A) = 57\Delta$ and $\ln(B) = -57\Delta$

Plan	Pure	Conserv.	5th percentile
E(T)	1396	54	20.2
E(M)	1396	1396	806
aE(T)+cE(M)	942,005	841,417	840,923
Saving ($)	0	100,588	101,082

Case Study: DCP Cooperative Group Treatment Trials

National Cancer Institute:

- cost per patient in 1999 is $c = 3,861$
- average cost of a clinical trial is $a = 31,000$

$\alpha = \beta = 0.05$	Pure	Conserv.	5th percentile
E(T)	804	45	20.2
E(M)	804	804	806
aE(T)+cE(M), mln	28.0	4.5	3.7
Saving, mln	0	23.5	24.3

$\alpha = \beta = 0.01$	Pure	Conserv.	5th percentile
E(T)	1396	54	20.2
E(M)	1396	1396	806
aE(T)+cE(M), mln	48.7	7.1	6.0
Saving, mln	0	41.6	42.7