Topics in the Design of Experiments
Part 2: Sequential Design

Steve Coad*

March 2020

1 Introduction

1.1 Topics to be covered

Part 1 of this course showed how an experiment may be designed optimally when the number of observations is **fixed** in advance. The main focus was the application of this theory to **non-linear models** and the construction of D-optimum designs. In this part, we are concerned with sequential designs, where the number of observations to be taken is not fixed in advance or the design points are chosen sequentially depending on the current data.

We first introduce the sequential probability ratio test for testing two simple hypotheses and study several of its properties. Since this is a fully sequential design, in that a test is performed after every observation, we then introduce group sequential designs and show how these may be carried out in practice. Finally, we study some adaptive treatment allocation rules, where the treatment allocation probabilities are functions of the current data.

1.2 Examples of sequential designs

A sequential design is often **more efficient** than an equivalent fixed-sample one. The examples below demonstrate the wide range of applications of a sequential approach.

Example. Curtailed test.
Suppose that a machine produces items which may be judged good or defective, and that the true proportion of defectives in a large batch is p. Let S_m denote the number of defectives in a random sample of size m. Consider testing $H_0 : p \leq p_0$ against $H_1 : p > p_0$. A reasonable rule is to reject H_0 if $S_m \geq r$ for some constant r.

*School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, United Kingdom

Now let T be the smallest value of k for which $S_k = r$ and put $T' = \min(T, m)$. Consider the procedure which stops sampling at the **random time** T' and decides that $p > p_0$ if and only if $T \leq m$. If one considers these two procedures as tests of H_0 against H_1, their rejection regions, namely, $\{T \leq m\}$ and $\{S_m \geq r\}$, are the same events, and hence the two tests have the same **power function**. Clearly, the test which stops at the random time T' has a reasonable claim to be regarded as more efficient.

Example. Repeated significance test.
Let $X_1, X_2, \ldots$ be independent normal random variables with unknown mean μ and unit variance. Consider testing $H_0 : \mu = 0$ against $H_0 : \mu \neq 0$. Then the standard fixed-sample 0.05 level significance test rejects H_0 if and only if $|S_n| \geq 1.96\sqrt{n}$, where $S_n = \sum_{k=1}^{n} x_k$.

Now suppose that, if H_1 is true, a minimum amount of experimentation is desired, but no similar constraint exists under H_0. Let $b > 0$ and let m be a maximum sample size. Sample sequentially, stopping with rejection of H_0 at the first $n \leq m$, if one exists, such that $|S_n| \geq b\sqrt{n}$. Otherwise, stop sampling at m and accept H_0. The significance level of this procedure is

$$\alpha = \alpha(b, m) = P_0(|S_n| \geq b\sqrt{n} \text{ for some } n \leq m),$$

where P_0 denotes probability under H_0. Clearly, b must be somewhat larger than 1.96, depending on m, in order that $\alpha(b, m) = 0.05$.

2 The sequential probability ratio test (SPRT)

2.1 Definitions

Let $X_1, X_2, \ldots$ be a sequence of random variables with joint probability density functions

$$P(X_1 \in d\xi_1, \ldots, X_n \in d\xi_n) = f_n(\xi_1, \ldots, \xi_n) d\xi_1 \ldots d\xi_n$$

for $n = 1, 2, \ldots$. Consider testing the **simple hypotheses** $H_0 : f_n = f_{0n}$ for all n against $H_1 : f_n = f_{1n}$ for all n. The **likelihood ratio** is

$$\ell_n = \ell_n(x_1, \ldots, x_n) = \frac{f_{1n}(x_1, \ldots, x_n)}{f_{0n}(x_1, \ldots, x_n)}.$$

The **sequential probability ratio test (SPRT)** chooses constants $0 < A < B < \infty$, usually $A < 1 < B$, and samples $x_1, x_2, \ldots$ sequentially until the random time

$$\begin{aligned} N &= \text{first } n \geq 1 \text{ such that } \ell_n \notin (A, B) \\ &= \infty \text{ if } \ell_n \in (A, B) \text{ for all } n \geq 1. \end{aligned}$$

The test stops sampling at time N, and, if $N < \infty$, rejects H_0 if $\ell_N \geq B$ and accepts H_0 if $\ell_N \leq A$.

Assuming temporarily that $P_i(N < \infty) = 1$ for $i = 0, 1$, where P_i denotes probability under H_i, the above test has **significance level** $\alpha = P_0(\ell_N \geq B)$ and **power** $1 - \beta = P_1(\ell_N \geq B)$.

In the fixed-sample case, the **Neyman-Pearson lemma** tells us that, among all tests with the same significance level, the likelihood ratio test has the highest power.

2.2 Properties of the SPRT

Let B_n denote the subset of n-dimensional space in which $A < \ell_k(\xi_1, \ldots, \xi_k) < B$ for $k = 1, 2, \ldots, n-1$ and $\ell_n(\xi_1, \ldots, \xi_n) \geq B$, so that

$$\{N = n, \ell_n \geq B\} = \{(x_1, \ldots, x_n) \in B_n\}.$$

Then

$$\begin{aligned}
\alpha = P_0(\ell_N \geq B) &= \sum_{n=1}^{\infty} P_0(N = n, \ell_n \geq B) \\
&= \sum_{n=1}^{\infty} \int_{B_n} f_{0n} d\xi_1 \ldots d\xi_n \\
&= \sum_{n=1}^{\infty} \int_{B_n} \frac{f_{0n}}{f_{1n}} f_{1n} d\xi_1 \ldots d\xi_n \\
&= \sum_{n=1}^{\infty} E_1\left(\ell_n^{-1}; N = n, \ell_n \geq B\right) \\
&= E_1\left(\ell_N^{-1}; \ell_N \geq B\right) \\
&\leq B^{-1} P_1(\ell_N \geq B) = B^{-1}(1 - \beta).
\end{aligned}$$

Similarly,

$$\beta = P_1(\ell_N \leq A) \leq A P_0(\ell_N \leq A) = A(1 - \alpha).$$

Treating the above inequalities as approximate equalities and solving for α and β leads to the simple approximations

$$\alpha \simeq \frac{1 - A}{B - A} \quad \text{and} \quad \beta \simeq \frac{A(B - 1)}{B - A}.$$

Theorem 1. Wald's equation.
Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with finite mean μ. Let M be any integer-valued random variable such that $\{M = n\}$ is an event determined only by $X_1, \ldots, X_n$ for all $n = 1, 2, \ldots$, and assume that $E(M) < \infty$. Then

$$E\left(\sum_{k=1}^{M} X_k\right) = \mu E(M).$$

Proof. Suppose initially that $X \geq 0$. Write

$$\sum_{k=1}^{M} X_k = \sum_{k=1}^{\infty} 1_{\{M \geq k\}} X_k$$

and note that

$$\{M \geq k\} = \left[\bigcup_{j=1}^{k-1} \{M = j\}\right]^c$$

is independent of $X_k, X_{k+1}, \ldots$. Hence, by the **monotone convergence theorem**, we have that

$$E\left(\sum_{k=1}^{M} X_k\right) = \sum_{k=1}^{\infty} E(X_k; M \geq k) = \mu \sum_{k=1}^{\infty} P(M \geq k) = \mu E(M).$$

For the general case, write

$$\sum_{k=1}^{M} X_k = \sum_{k=1}^{M} X_k^+ - \sum_{k=1}^{M} X_k^-,$$

where $a^+ = \max(a, 0)$ and $a^- = -\min(a, 0)$, and apply the above case to these two terms separately. □

Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables, so that

$$\ell_n = \prod_{k=1}^{n} \frac{f_1(x_k)}{f_0(x_k)},$$

where f_i is the probability density function of X under H_i for $i = 0, 1$. Then

$$\log \ell_n = \sum_{k=1}^{n} \log \left\{\frac{f_1(x_k)}{f_0(x_k)}\right\}$$

is a sum of independent and identically distributed random variables. Further, the **stopping rule** for the SPRT may be written as

$$\begin{aligned} N &= \text{first } n \geq 1 \text{ such that } \log \ell_n \notin (a, b) \\ &= \infty \text{ if } \log \ell_n \in (a, b) \text{ for all } n, \end{aligned}$$

where $a = \log A$ and $b = \log B$. Now, by Theorem 1,

$$E_i(\log \ell_N) = \mu_i E_i(N),$$

where $\mu_i = E_i[\log\{f_1(X)/f_0(X)\}]$ for $i = 0, 1$. Also, we may write

$$E_i(\log \ell_N) \simeq a P_i(\ell_N \leq A) + b P_i(\ell_N \geq B).$$

Combining this approximation with the previous equation yields

$$E_0(N) \simeq \frac{1}{\mu_0}\left\{a \frac{(B-1)}{B-A} + b \frac{(1-A)}{B-A}\right\}$$

and

$$E_1(N) \simeq \frac{1}{\mu_1}\left\{a \frac{A(B-1)}{B-A} + b \frac{B(1-A)}{B-A}\right\}.$$

Alternatively, we may write

$$E_0(N) \simeq \frac{1}{\mu_0}\left\{(1-\alpha)\log\left(\frac{\beta}{1-\alpha}\right)+\alpha\log\left(\frac{1-\beta}{\alpha}\right)\right\}$$

and

$$E_1(N) \simeq \frac{1}{\mu_1}\left\{\beta\log\left(\frac{\beta}{1-\alpha}\right)+(1-\beta)\log\left(\frac{1-\beta}{\alpha}\right)\right\}.$$

The following result, which is stated without proof, shows that the SPRT terminates with probability one.

Theorem 2. Stein's lemma.
Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with the property $P(Y=0)<1$. Let $-\infty<a<b<\infty$ and $S_n=\sum_{k=1}^{n} Y_k$, and define

$$\begin{aligned} M &= \text{first } n \geq 1 \text{ such that } S_n \notin (a,b) \\ &= \infty \text{ if } S_n \in (a,b) \text{ for all } n. \end{aligned}$$

Then there exist constants $C>0$ and $0<\rho<1$ such that $P(M>n) \leq C\rho^n$ for $n=1,2,\ldots$. In particular, $E(M^k)<\infty$ for all $k=1,2,\ldots$ and $E(e^{\lambda M})<\infty$ for $\lambda<\log(\rho^{-1})$.

Theorem 3. Wald's likelihood ratio identity.
Let $X_1, X_2, \ldots$ be an arbitrary sequence of random variables and suppose that there exists a likelihood ratio ℓ_n for $x_1, \ldots, x_n$ under P_1 relative to P_0 such that

$$E_1(Y_n) = E_0(Y_n \ell_n),$$

where $Y_n = g(X_1, \ldots, X_n)$ for some function g. Then, for any stopping time N and non-negative random variable $Y = g(X_1, \ldots, X_N)$, say,

$$E_1(Y; N<\infty) = E_0(Y\ell_N; N<\infty).$$

In particular, if $Y = 1_A$, then

$$P_1[A \cap \{N<\infty\}] = E_0[\ell_N; A \cap \{N<\infty\}].$$

Proof. We have that

$$\begin{aligned} E_1(Y; N<\infty) &= \sum_{n=1}^{\infty} E_1(Y; N=n) \\ &= \sum_{n=1}^{\infty} E_0(Y\ell_n; N=n) \\ &= E_0(Y\ell_N; N<\infty), \end{aligned}$$

as required. $\square$

Note that we have already used Theorem 3 to obtain the approximations for the error probabilities α and β. The following result is a special case of Theorem 3 and is stated without proof.

Corollary. Wald's fundamental identity.
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, and suppose that $\phi(t) = E(e^{tX}) < \infty$ for some $t \neq 0$. Then, if N is a stopping time such that $P(N < \infty) = 1$,

$$E\left[\{\phi(t)\}^{-N} e^{tS_N}\right] = 1,$$

where $S_N = \sum_{k=1}^{N} X_k$.

If $\phi(t) < \infty$ for $|t| < \delta$, where $\delta > 0$, then the above identity may be differentiated with respect to t at $t = 0$ to reproduce Theorem 1, provided that differentiation under the expectation can be justified.

When testing a simple hypothesis against a simple alternative with independent and identically distributed observations, the **Wald-Wolfowitz theorem** states that the SPRT minimises $E_i(N)$ for $i = 0, 1$, among all tests having no larger error probabilities. For cases where the results for expected sample size are exact, Theorem 5 contains a complete proof.

Theorem 5. Let T be the stopping time of any test of $H_0 : f = f_0$ against $H_1 : f = f_1$ with error probabilities α and β, $0 < \alpha, \beta < 1$. Assume that $E_i(T) < \infty$ for $i = 0, 1$. Then

$$E_0(T) \geq \frac{1}{\mu_0}\left\{(1-\alpha)\log\left(\frac{\beta}{1-\alpha}\right) + \alpha \log\left(\frac{1-\beta}{\alpha}\right)\right\}$$

and

$$E_1(T) \geq \frac{1}{\mu_1}\left\{\beta\log\left(\frac{\beta}{1-\alpha}\right) + (1-\beta) \log\left(\frac{1-\beta}{\alpha}\right)\right\},$$

where

$$\mu_i = E_i\left[\log\left\{\frac{f_1(X)}{f_0(X)}\right\}\right]$$

for $i = 0, 1$.

Proof. Let $R = \{\text{Reject } H_0\}$ and $R^c = \{\text{Accept } H_0\}$. Then, by Theorem 3,

$$\begin{aligned}
\alpha = P_0(R) &= E_1\left(\ell_T^{-1}; R\right) \\
&= E_1\left(e^{-\log \ell_T} | R\right) P_1(R) \\
&\geq \exp\{-E_1(\log \ell_T | R)\}(1-\beta) \\
&= \exp\{-E_1(\log \ell_T; R)/(1-\beta)\}(1-\beta),
\end{aligned}$$

where the penultimate line is due to **Jensen's inequality**. Thus,

$$(1-\beta)\log\left(\frac{\alpha}{1-\beta}\right) \geq -E_1(\log \ell_T; R).$$

Similarly,

$$\beta\log\left(\frac{1-\alpha}{\beta}\right) \geq -E_1(\log \ell_T; R^c).$$

Hence, by Theorem 1,

$$\begin{aligned}(1-\beta)\log\left(\frac{\alpha}{1-\beta}\right) + \beta\log\left(\frac{1-\alpha}{\beta}\right) &\geq -E_1(\log \ell_T) \\ &= -\mu_1 E_1(T).\end{aligned}$$

Since $\mu_1 > 0$, this completes the proof of the first assertion. The second assertion is proved similarly. □

2.3 Estimation following the SPRT

The estimation of a parameter when the data have been obtained from a SPRT is a difficult problem. Even sequentially stopped versions of ordinarily unbiased estimators are biased, and their sampling distributions are often quite complicated. The following result, which is stated without proof, shows that randomly stopped averages are asymptotically normally distributed under quite general conditions.

Theorem 6. Anscombe-Doeblin theorem.
Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with finite mean μ and finite positive variance σ^2. Let $S_n = \sum_{k=1}^n X_k$ and suppose that M_c, $c \geq 0$, are positive integer-valued random variables such that, for some constants $m_c \to \infty$, $M_c/m_c \to 1$ in probability as $c \to \infty$. Then

$$P\left(\frac{S_{M_c} - \mu M_c}{\sigma M_c^{\frac{1}{2}}} \leq x\right) \to \Phi(x)$$

as $c \to \infty$, where Φ denotes the standard normal distribution function.

It follows from Theorem 6 that $N^{-\frac{1}{2}}(S_N - N\mu)/\sigma$ is approximately standard normal. Thus, an approximate 95% confidence interval for μ for large a and b is given by

$$\frac{S_N}{N} \pm 1.96\frac{\sigma}{N^{\frac{1}{2}}}.$$

However, this approximation is very poor for moderate values of a and b.