2.4 Examples of the SPRT

Let N be the stopping time for the SPRT. Then, from Theorem 2, $P_i(N < \infty) = 1$ for $i = 0, 1$.

Example. Let $X_1, X_2, \ldots$ be independent normal random variables with mean μ and unit variance. Consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$, where $\mu_0 < \mu_1$. The likelihood ratio is

$$\begin{aligned} \ell_n &= \prod_{k=1}^{n} \frac{\phi(x_k - \mu_1)}{\phi(x_k - \mu_0)} \\ &= \exp\left\{(\mu_1 - \mu_0)S_n - \frac{n}{2}(\mu_1^2 - \mu_0^2)\right\}, \end{aligned}$$

where $\phi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$ and $S_n = \sum_{k=1}^{n} x_k$. Hence, the stopping rule for the SPRT is given by

$$N = \text{first } n \geq 1 \text{ such that } S_n - \tfrac{n}{2}(\mu_1 + \mu_0) \notin (a, b),$$

where $a = \log A/(\mu_1 - \mu_0)$ and $b = \log B/(\mu_1 - \mu_0)$. Note that, in the **symmetric case**, $\mu_1 = -\mu_0$ and $b = -a$, and we have that

$$N = \text{first } n \geq 1 \text{ such that } |S_n| \geq b.$$

Example. Let $X_1, X_2, \ldots$ be independent random variables with $P_p(X = 1) = p$ and $P_p(X = -1) = q$, where $p + q = 1$. Consider testing $H_0 : p = p_0$ against $H_1 : p = p_1$, where $p_0 < p_1$. The likelihood ratio is

$$\begin{aligned} \ell_n &= \left(\frac{p_1}{p_0}\right)^{\frac{n+S_n}{2}} \left(\frac{q_1}{q_0}\right)^{\frac{n-S_n}{2}} \\ &= \left(\frac{p_1 q_0}{p_0 q_1}\right)^{\frac{S_n}{2}} \left(\frac{p_1 q_1}{p_0 q_0}\right)^{\frac{n}{2}}, \end{aligned}$$

where $S_n = \sum_{k=1}^{n} x_k$. Note that, in the symmetric case, $p_0 = q_1$ and $B = A^{-1}$, the stopping rule for the SPRT is given by

$$N = \text{first } n \geq 1 \text{ such that } |S_n| \geq b,$$

where $b = \log B/\log(q_0/p_0)$.

3 Group sequential tests

3.1 Analysing the data in groups

The SPRT is an example of a fully sequential test, since a test is performed after every observation. In practice, especially in the context of clinical trials, it is more convenient

to analyse the data after groups of observations. In fact, a group sequential approach can often achieve most of the efficiency gains of an analogous fully sequential one. In this course, group sequential tests are described in the context of two-treatment clinical trials.

Let $X_{A1}, X_{A2}, \ldots$ and $X_{B1}, X_{B2}, \ldots$ denote the responses of subjects assigned to two treatments, A and B. Interest lies in testing the null hypothesis of no treatment difference $H_0 : \theta = 0$ against the two-sided alternative $H_1 : \theta \neq 0$ that there is a treatment difference with type I error probability α and power $1 - \beta$ when $\theta = \pm\delta$. Suppose that there are a maximum of K groups and that m denotes the **group size**.

For $k = 1, 2, \ldots, K$, a **standardised test statistic** Z_k is calculated from the first k groups of observations and H_0 is rejected if $Z_k \notin (a_k, b_k)$, where a_k and b_k denote the critical values for the kth analysis. If the test continues to the Kth analysis and $Z_K \in (a_K, b_K)$, it terminates and H_0 is accepted. The critical values are chosen to achieve the required type I error probability and the power condition determines the group size.

3.2 Designing a group sequential test

Suppose that a group sequential test with a maximum of K analyses yields the sequence of test statistics $\{Z_1, \ldots, Z_K\}$. These statistics are said to have the **canonical joint distribution** with information levels $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$ for the parameter θ if (i) $(Z_1, \ldots, Z_K)$ is multivariate normal, (ii) $E(Z_k) = \theta\sqrt{\mathcal{I}_k}$, $k = 1, 2, \ldots, K$, and (iii) $\mathrm{cov}(Z_{k_1}, Z_{k_2}) = \sqrt{\mathcal{I}_{k_1}/\mathcal{I}_{k_2}}$, $1 \leq k_1 \leq k_2 \leq K$. The fact that this implies that $\{Z_1, \ldots, Z_K\}$ is a **Markov sequence** simplifies the calculations. Let $\Delta_k = \mathcal{I}_k - \mathcal{I}_{k-1}$ for $k = 2, 3, \ldots, K$. Then $Z_1 \sim \mathrm{N}(\theta\sqrt{\mathcal{I}_1}, 1)$, and, for each $k = 2, 3, \ldots, K$,

$$Z_k\sqrt{\mathcal{I}_k} - Z_{k-1}\sqrt{\mathcal{I}_{k-1}} \sim \mathrm{N}(\theta\Delta_k, \Delta_k)$$

independently of $Z_1, \ldots, Z_{k-1}$.

A key quantity to calculate for a group sequential test is the probability of crossing a specific stopping boundary at a particular analysis. For each $k = 1, 2, \ldots, K$, let

$$\psi_k(a_1, b_1, \ldots, a_k, b_k; \theta) = P_\theta(a_1 < Z_1 < b_1, \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k \geq b_k)$$

and

$$\xi_k(a_1, b_1, \ldots, a_k, b_k; \theta) = P_\theta(a_1 < Z_1 < b_1, \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k \leq a_k).$$

Now, Z_1 has density

$$f_1(z_1; \theta) = \phi(z_1 - \theta\sqrt{\mathcal{I}_1}).$$

Further, the conditional density of Z_k given $Z_1 = z_1, \ldots, Z_{k-1} = z_{k-1}$ depends only on z_{k-1} and is

$$f_k(z_{k-1}, z_k; \theta) = \frac{\sqrt{\mathcal{I}_k}}{\sqrt{\Delta_k}}\phi\left(\frac{z_k\sqrt{\mathcal{I}_k} - z_{k-1}\sqrt{\mathcal{I}_{k-1}} - \theta\Delta_k}{\sqrt{\Delta_k}}\right).$$

Hence, for each $k = 2, 3, \ldots, K$,

$$\begin{aligned}\psi_k(a_1, b_1, \ldots, a_k, b_k; \theta) &= \int_{a_1}^{b_1} \ldots \int_{a_{k-1}}^{b_{k-1}} \int_{b_k}^{\infty} f_1(z_1; \theta) f_2(z_1, z_2; \theta) \ldots f_k(z_{k-1}, z_k; \theta) dz_k \ldots dz_1 \\ &= \int_{a_1}^{b_1} \ldots \int_{a_{k-1}}^{b_{k-1}} f_1(z_1; \theta) f_2(z_1, z_2; \theta) \ldots f_{k-1}(z_{k-2}, z_{k-1}; \theta) \\ &\quad \times e_{k-1}(z_{k-1}, b_k; \theta) dz_{k-1} \ldots dz_1,\end{aligned}$$

where

$$e_{k-1}(z_{k-1}, b_k; \theta) = \Phi\left(\frac{z_{k-1}\sqrt{\mathcal{I}_{k-1}} + \theta\Delta_k - b_k\sqrt{\mathcal{I}_k}}{\sqrt{\Delta_k}} \right).$$

A similar expression can be obtained for $\xi_k(a_1, b_1, \ldots, a_k, b_k; \theta)$.

Although the above multiple integral appears difficult, it can be rewritten to simplify the computation. To see this, let $g_k(z_k; \theta)$, $k = 1, 2, \ldots, K$, denote the **sub-densities** of $Z_1, \ldots, Z_K$, their integrals being less than unity for $k > 1$ due to early stopping at stages $1, 2, \ldots, k-1$. In other words,

$$g_1(z_1; \theta) = f_1(z_1; \theta)$$

and

$$g_k(z_k; \theta) = \int_{a_{k-1}}^{b_{k-1}} g_{k-1}(z_{k-1}; \theta) f_k(z_{k-1}, z_k; \theta) dz_{k-1}$$

for $k = 2, 3, \ldots, K$. It then follows that we can write

$$\begin{aligned}\psi_k(a_1, b_1, \ldots, a_k, b_k; \theta) &= \int_{b_k}^{\infty} g_k(z_k; \theta) dz_k \\ &= \int_{a_{k-1}}^{b_{k-1}} \int_{b_k}^{\infty} g_{k-1}(z_{k-1}; \theta) f_k(z_{k-1}, z_k; \theta) dz_k dz_{k-1} \\ &= \int_{a_{k-1}}^{b_{k-1}} g_{k-1}(z_{k-1}; \theta) e_{k-1}(z_{k-1}, b_k; \theta) dz_{k-1}.\end{aligned}$$

Thus, the computation only requires a succession of univariate integrations.

The values of $\psi_k(a_1, b_1, \ldots, a_k, b_k; \theta)$ and $\xi_k(a_1, b_1, \ldots, a_k, b_k; \theta)$ for $k = 1, 2, \ldots, K$ determine the distribution of the stopping time and associated decision for a group sequential test. From these, we can obtain the test's error probabilities for any θ. For example, the test's type I error probability is

$$P_{\theta=0}(\text{Reject } H_0) = \sum_{k=1}^{K} \{\psi_k(a_1, b_1, \ldots, a_k, b_k; 0) + \xi_k(a_1, b_1, \ldots, a_k, b_k; 0)\}.$$

Similarly, the test's power when $\theta = \delta$ is

$$\begin{aligned}P_{\theta=\delta}(\text{Reject } H_0) &= \sum_{k=1}^{K} \{\psi_k(a_1, b_1, \ldots, a_k, b_k; \delta) + \xi_k(a_1, b_1, \ldots, a_k, b_k; \delta)\} \\ &\simeq \sum_{k=1}^{K} \psi_k(a_1, b_1, \ldots, a_k, b_k; \delta)\end{aligned}$$

if $\delta > 0$ is large. The approximate power when $\theta = -\delta$ has the same form, but with ψ_k replaced with ξ_k and δ with $-\delta$. For specified values of K and the type I error probability α, a **numerical search** can be used to find the a_k and b_k.

Recall that a fixed-sample test of $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ with type I error probability α and power $1 - \beta$ at $\theta = \pm\delta$ has information

$$\mathcal{I}_{f,2} = \frac{\{\Phi^{-1}(1-\alpha/2) + \Phi^{-1}(1-\beta)\}^2}{\delta^2}.$$

Then a group sequential test requires a larger maximum sample size and we set a **maximum information level**, $\mathcal{I}_{max} = R\mathcal{I}_{f,2}$, where $R > 1$ and depends on K, α, β and the type of group sequential boundary being used. With equally-spaced information levels, we have

$$\mathcal{I}_k = \frac{k}{K}\mathcal{I}_{max}, \quad k = 1, 2, \ldots, K.$$

By finding the value $\mathcal{I}_{max}$ such that the test's power is $1 - \beta$ for this sequence of information levels, we can obtain R.

The above calculations can be used to design group sequential tests with specific properties. For example, the **Wang and Tsiatis family** of two-sided tests are indexed by a parameter Δ, which gives boundaries of different shapes. Members of this family include the Pocock test with constant critical values and the O'Brien and Fleming test with converging critical values. The test with parameter Δ has boundaries of the form

$$a_k = -c\left(\frac{k}{K}\right)^{\Delta-\frac{1}{2}} \quad \text{and} \quad b_k = c\left(\frac{k}{K}\right)^{\Delta-\frac{1}{2}}.$$

Taking $\Delta = 1/2$ gives Pocock's test and $\Delta = 0$ gives the O'Brien and Fleming test.

3.3 Inference following a group sequential test

The group sequential test stops at stage

$$T = \min\{\text{first } k \geq 1 \text{ such that } Z_k \notin (a_k, b_k), K\}.$$

Now, the sequence of test statistics $\{Z_1, \ldots, Z_K\}$ has the same joint distribution as the sequence $\{(Y_1 + \ldots + Y_k)/\sqrt{\mathcal{I}_k}; k = 1, 2, \ldots, K\}$, where the Y_k are independent such that $Y_k \sim \mathrm{N}(\Delta_k\theta, \Delta_k)$. The structure of the joint density of $(Z_1, \ldots, Z_K)$ shows that (T, Z_T) is a pair of **sufficient statistics** for θ and that the maximum likelihood estimator of θ is given by $\hat{\theta} = Z_T/\sqrt{\mathcal{I}_T}$. Although the form of the maximum likelihood estimator is the same as for a fixed-sample test, its sampling distribution is more complicated.

The sampling density of $\hat{\theta}$ at $\hat{\theta} = y$ is given by

$$\sum_{k=1}^{K} g_k(y\sqrt{\mathcal{I}_k}; \theta)\sqrt{\mathcal{I}_k}1_{y\sqrt{\mathcal{I}_k} \notin (a_k, b_k)},$$

where the contributions from the K sub-densities yield a multi-modal density with a peak for each value of $T \in \{1, 2, \ldots, K\}$. This means that the density is completely different to a normal with mean θ and variance $\mathcal{I}_T^{-1}$, which it would be for a fixed-sample test. As a result, $\hat{\theta}$ is now a biased estimator of θ. More specifically, we can write

$$E_\theta(\hat{\theta}) = \sum_{k=1}^{K} \left\{ \int_{-\infty}^{a_k} g_k(z_k;\theta) \frac{z_k}{\sqrt{\mathcal{I}_k}} dz_k + \int_{b_k}^{\infty} g_k(z_k;\theta) \frac{z_k}{\sqrt{\mathcal{I}_k}} \right\} dz_k.$$

An expression can also be obtained for the variance of $\hat{\theta}$.

As before, the above integrals can be rewritten to simplify the computation. For example, the first integral can be written as

$$\begin{aligned} \int_{-\infty}^{a_k} g_k(z_k;\theta) \frac{z_k}{\sqrt{\mathcal{I}_k}} dz_k &= \int_{a_{k-1}}^{b_{k-1}} \int_{-\infty}^{a_k} g_{k-1}(z_{k-1};\theta) f_k(z_{k-1}, z_k;\theta) \frac{z_k}{\sqrt{\mathcal{I}_k}} dz_k dz_{k-1} \\ &= \int_{a_{k-1}}^{b_{k-1}} g_{k-1}(z_{k-1};\theta) r_{k-1}(z_{k-1}, a_k;\theta) dz_{k-1}, \end{aligned}$$

where

$$\begin{aligned} r_{k-1}(z_{k-1}, a_k;\theta) &= -\frac{\sqrt{\Delta_k}}{\mathcal{I}_k} \phi\left(\frac{a_k\sqrt{\mathcal{I}_k} - z_{k-1}\sqrt{\mathcal{I}_{k-1}} - \theta\Delta_k}{\sqrt{\Delta_k}} \right) \\ &\quad + \frac{(z_{k-1}\sqrt{\mathcal{I}_{k-1}} + \theta\Delta_k)}{\mathcal{I}_k} \Phi\left(\frac{a_k\sqrt{\mathcal{I}_k} - z_{k-1}\sqrt{\mathcal{I}_{k-1}} - \theta\Delta_k}{\sqrt{\Delta_k}} \right). \end{aligned}$$

The second integral can be computed in a similar way.

Upon termination of the group sequential test, rather than just concluding that we accept or reject H_0, we can report the p-value of the observed data for testing H_0. Now, the **sample space** Ω defined by the group sequential design is the set of all pairs (k, z), where $z \notin (a_k, b_k)$, so that the test can terminate with $(T, Z_T) = (k, z)$. Let the observed value of (T, Z_T) be denoted by (k^*, z^*). Then the p-value is

$$P_{\theta=0}\{\text{Obtain } (k, z) \text{ as extreme or more extreme than } (k^*, z^*)\}.$$

In order to calculate this, we need to specify the ordering of Ω. We write $(k', z') \succ (k, z)$ to denote that (k', z') is above (k, z) in a given ordering.

There are a number of orderings available. In **stage-wise ordering**, $(k', z') \succ (k, z)$ if any of the following conditions hold: (i) $k' = k$ and $z' \geq z$; (ii) $k' < k$ and $z' \geq b_{k'}$; (iii) $k' > k$ and $z \leq a_k$. As an example, suppose that the test terminates after crossing the upper boundary. Then the one-sided upper p-value is

$$P_{\theta=0}\{(T, Z_T) \succeq (k^*, z^*)\} = \sum_{j=1}^{k^*-1} \psi_j(a_1, b_1, \ldots, a_j, b_j; 0) + \psi_{k^*}(a_1, b_1, \ldots, a_{k^*-1}, b_{k^*-1}, a_{k^*}, z^*; 0).$$

One-sided lower p-values are found in the same manner and the two-sided p-value is twice the smaller of these two quantities.

Equal-tailed $100(1-\alpha)\%$ confidence intervals for θ can be obtained by inverting a family of hypothesis tests with two-sided type I error probability α. For any given value θ_0, we can find pairs $(k_u(\theta_0), z_u(\theta_0))$ and $(k_\ell(\theta_0), z_\ell(\theta_0))$ such that

$$P_{\theta=\theta_0}\{(T, Z_T) \succeq (k_u(\theta_0), z_u(\theta_0))\} = \frac{\alpha}{2}$$

and

$$P_{\theta=\theta_0}\{(T, Z_T) \preceq (k_\ell(\theta_0), z_\ell(\theta_0))\} = \frac{\alpha}{2}.$$

It follows that the acceptance region

$$A(\theta_0) = \{(k, z) : (k_\ell(\theta_0), z_\ell(\theta_0)) \prec (k, z) \prec (k_u(\theta_0), z_u(\theta_0))\}$$

defines a two-sided hypothesis test of $\theta = \theta_0$ with type I error probability α. This implies that the set $\{\theta : (T, Z_T) \in A(\theta)\}$ obtained by inverting this family of tests is a $100(1-\alpha)\%$ equal-tailed **confidence set** for θ. If $P_\theta\{(T, Z_T) \succeq (k, z)\}$ is an increasing function of θ for each $(k, z) \in \Omega$, then this set is an interval.

3.4 Examples of group sequential designs

Two examples are now given to show how group sequential designs may be constructed in practice.

Example. Let $X_{A1}, X_{A2}, \ldots$ be independent normal random variables with mean μ_A and unit variance and let $X_{B1}, X_{B2}, \ldots$ be independent normal random variables with mean μ_B and unit variance. For $k = 1, 2, \ldots, K$, let n_{Ak} and n_{Bk} denote the cumulative numbers of observations on treatments A and B, respectively, at the time of the kth analysis. Then the parameter of interest is $\theta = \mu_A - \mu_B$ and its natural estimator is

$$\overline{X}_A^{(k)} - \overline{X}_B^{(k)} = \frac{1}{n_{Ak}} \sum_{i=1}^{n_{Ak}} X_{Ai} - \frac{1}{n_{Bk}} \sum_{i=1}^{n_{Bk}} X_{Bi} \sim \mathrm{N}(\theta, \mathcal{I}_k^{-1}),$$

where

$$\mathcal{I}_k = \left(\frac{1}{n_{Ak}} + \frac{1}{n_{Bk}}\right)^{-1}$$

is the information for θ. So the standardised statistic at analysis k for testing $H_0 : \theta = 0$ is

$$Z_k = \left\{\overline{X}_A^{(k)} - \overline{X}_B^{(k)}\right\} \sqrt{\mathcal{I}_k}$$

for $k = 1, 2, \ldots, K$.

It is easily verified that the above statistics have the canonical joint distribution with information levels $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$ for θ. Firstly, $(Z_1, \ldots, Z_K)$ is multivariate normal, since each

Z_k is a linear combination of the independent normal random variables X_{Ai} and X_{Bi} for $i = 1, 2, \ldots$. Secondly, we know that $Z_k \sim \mathrm{N}(\theta\sqrt{\mathcal{I}_k}, 1)$. Lastly, for $k_1 \leq k_2$,

$$\begin{aligned}
\mathrm{cov}(Z_{k_1}, Z_{k_2}) &= \mathrm{cov}\left[\left\{\overline{X}_A^{(k_1)} - \overline{X}_B^{(k_1)}\right\}\sqrt{\mathcal{I}_{k_1}}, \left\{\overline{X}_A^{(k_2)} - \overline{X}_B^{(k_2)}\right\}\sqrt{\mathcal{I}_{k_2}}\right] \\
&= \left(\frac{1}{n_{Ak_1}}\frac{1}{n_{Bk_2}}n_{Ak_1} + \frac{1}{n_{Bk_1}}\frac{1}{n_{Ak_2}}n_{Bk_1}\right)\sqrt{\mathcal{I}_{k_1}}\sqrt{\mathcal{I}_{k_2}} \\
&= \mathcal{I}_{k_2}^{-1}\sqrt{\mathcal{I}_{k_1}}\sqrt{\mathcal{I}_{k_2}} = \sqrt{\mathcal{I}_{k_1}/\mathcal{I}_{k_2}},
\end{aligned}$$

as required.

Now suppose that we wish to test $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ with type I error probability $\alpha = 0.05$ and power $1-\beta = 0.90$ when $\theta = \pm 0.5$. We will use a Pocock test with a maximum of $K = 5$ analyses. The information for θ required by a fixed-sample test with these error probabilities is

$$\mathcal{I}_{f,2} = \frac{\{\Phi^{-1}(0.975) + \Phi^{-1}(0.9)\}^2}{0.5^2} = 42.032$$

and the maximum information level for the group sequential test can be shown to be

$$R_P(5, 0.05, 0.1) \times \mathcal{I}_{f,2} = 1.207 \times 42.032 = 50.7.$$

Assuming that $n_{Ak} = n_{Bk}$ for each $k = 1, 2, \ldots, 5$, we see that $\mathcal{I}_5 = n_5/2$, where n_5 denotes the common value of n_{A5} and n_{B5}. Thus, solving $\mathcal{I}_5 = 50.7$ yields $n_5 = 101.4$, which we round to 110 to obtain a multiple of 10. This means that five groups of 11 observations per treatment should be planned. It may also be shown that H_0 is rejected at analysis k if $|Z_k| \geq 2.413$, $k = 1, 2, \ldots, 5$.

Example. Let $X_{A1}, X_{A2}, \ldots$ be independent Bernoulli random variables with parameter p_A and let $X_{B1}, X_{B2}, \ldots$ be independent Bernoulli random variables with parameter p_B. Then the parameter of interest is $\theta = p_A - p_B$ and its natural estimator is $\hat{p}_A^{(k)} - \hat{p}_B^{(k)} = \overline{X}_A^{(k)} - \overline{X}_B^{(k)}$. Let $\overline{p} = (p_A + p_B)/2$. Then, under H_0, $p_A = p_B = \overline{p}$, and the information for θ is

$$\mathcal{I}_k = \left\{\overline{p}(1-\overline{p})\left(\frac{1}{n_{Ak}} + \frac{1}{n_{Bk}}\right)\right\}^{-1}.$$

Estimating $\overline{p}$ by

$$\tilde{p}_k = \frac{\sum_{i=1}^{n_{Ak}} X_{Ai} + \sum_{i=1}^{n_{Bk}} X_{Bi}}{n_{Ak} + n_{Bk}},$$

we obtain the estimated information level at analysis k given by

$$\hat{\mathcal{I}}_k = \left\{\tilde{p}_k(1-\tilde{p}_k)\left(\frac{1}{n_{Ak}} + \frac{1}{n_{Bk}}\right)\right\}^{-1}$$

for $k = 1, 2, \ldots, K$. So the standardised statistics for testing H_0 are

$$Z_k = \left\{\hat{p}_A^{(k)} - \hat{p}_B^{(k)}\right\}\sqrt{\hat{\mathcal{I}}_k}$$

for $k = 1, 2, \ldots, K$. If θ is small, these statistics can be shown to follow approximately the canonical joint distribution with information levels $\{\hat{\mathcal{I}}_1, \ldots, \hat{\mathcal{I}}_K\}$ for θ.

Now suppose that we wish to test $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ with type I error probability $\alpha = 0.05$ and power $1 - \beta = 0.8$ when $\theta = \pm 0.2$. We will use an O'Brien and Fleming test with a maximum of $K = 10$ analyses. The information for θ required by a fixed-sample test with these error probabilities is $\mathcal{I}_{f,2} = 196.224$ and the maximum information level for the group sequential test can be shown to be

$$R_B(10, 0.05, 0.2) \times \mathcal{I}_{f,2} = 1.040 \times 196.224 = 204.1.$$

Taking $\overline{p} = 0.5$ and assuming that $n_{Ak} = n_{Bk}$ for $k = 1, 2, \ldots, 10$, we see that $\mathcal{I}_{10} = 2n_{10}$. Thus, solving $\mathcal{I}_{10} = 204.1$ yields $n_{10} = 102.1$, which we round to 120 to obtain a multiple of 20. This means that 10 groups of six observations per treatment should be planned. It may also be shown that H_0 is rejected at analysis k if $|Z_k| \geq 2.087\sqrt{10/k} = 6.600/\sqrt{k}$, $k = 1, 2, \ldots, 10$.

4 Adaptive treatment allocation rules

4.1 Definitions

So far, we have been concerned with how to construct sequential tests of some null hypothesis H_0 against an alternative H_1 which have certain error probabilities. We now turn our attention to the problem of how to assign patients to treatments in the context of a fixed-sample clinical trial. The incorporation of stopping rules will be addressed later. Suppose initially that there are $t \geq 2$ treatments.

If **complete randomisation** is used, the next patient is equally likely to be assigned to any of the t treatments, so that the treatment allocation probabilities are all $1/t$. So this randomisation rule does not take into account the previous treatment assignments and responses, or any other information. Consequently, complete randomisation is a non-adaptive treatment allocation rule.

Since complete randomisation can lead to treatment group imbalances, a **restricted randomisation** rule can be used to ensure that each treatment group has roughly the same number of patients. For one of the simplest such rules, the treatment which most reduces the imbalance is assigned with probability p, $0.5 < p < 1$, and the other $t - 1$ treatments are assigned with probability $(1 - p)/(t - 1)$.

As the trial progresses, some treatments may look more promising than others and it would be desirable to allocate a higher proportion of patients to these treatments. In such cases, a **response-adaptive randomisation** rule is used. The simplest such rules may be represented as urn models, in which balls of different types are added to or removed from the urn according to the previous assignments and responses.

4.2 Properties of adaptive treatment allocation

Let N_{jn} denote the number of patients on treatment j after n assignments for $j = 1, 2$. First suppose that complete randomisation is used. Then it is easy to see that $N_{1n}/n \to 1/2$ almost surely and

$$\sqrt{n}\left(\frac{N_{1n}}{n} - \frac{1}{2}\right) \to \mathrm{N}\left(0, \frac{1}{4}\right)$$

in distribution as $n \to \infty$. If restricted randomisation is used instead, then we still have $N_{1n}/n \to 1/2$, but, since the aim is now to balance the treatment groups, N_{1n}/n will be **less variable**. This means that any tests will have greater power.

As an example of a restricted randomisation rule, suppose that the responses on the two treatments are independent with variance σ^2. Then, after n assignments, the covariance matrix of the estimated means is $\sigma^2 \mathrm{diag}(1/n_{1n}, 1/n_{2n})$. Since the parameter of interest is the difference between the means, the D_A-optimum design minimises $\sigma^2(1/n_{1n}+1/n_{2n})$. This design assigns the next patient to treatment 1 if $n_{2n}^2 > n_{1n}^2$ and to treatment 2 if $n_{1n}^2 > n_{2n}^2$. From this deterministic design, we can construct a **biased coin design** which assigns the nth patient to treatment 1 with probability

$$\phi_n = \frac{N_{2,n-1}^2}{N_{1,n-1}^2 + N_{2,n-1}^2}.$$

It can be shown that

$$\sqrt{n}\left(\frac{N_{1n}}{n} - \frac{1}{2}\right) \to \mathrm{N}\left(0, \frac{1}{20}\right)$$

in distribution as $n \to \infty$. This means that N_{1n}/n is now asymptotically 80% less variable.

The above biased coin design is a special case of a **generalised biased coin design** which assigns the nth patient to treatment 1 with probability

$$\phi_n = \frac{N_{2,n-1}^\gamma}{N_{1,n-1}^\gamma + N_{2,n-1}^\gamma},$$

where $\gamma \geq 0$. It can be shown that

$$\sqrt{n}\left(\frac{N_{1n}}{n} - \frac{1}{2}\right) \to \mathrm{N}\left\{0, \frac{1}{4(1+2\gamma)}\right\}$$

in distribution as $n \to \infty$. When $\gamma = 0$, we have complete randomisation. The recommended design is $\gamma = 5$, for which the asymptotic variance is $1/44$.

Now suppose that response-adaptive randomisation is used. In order to assess how good an allocation rule is, we need to study the behaviour of N_{1n}/n for large n. Although response-adaptive randomisation will assign a higher proportion of patients to the better treatment, it induces correlation among treatment assignments, so that N_{1n}/n may be **more variable**. This means that any tests may have lower power. If the distribution of the responses on

treatment j depends on the parameter θ_j and $\rho_j(\theta_1, \theta_2)$ denotes the **target allocation** for treatment j for $j = 1, 2$, then we can use the variance of N_{1n}/n to compare rules with the same target allocation.

For a given target allocation, a Cramér-Rao lower bound on the asymptotic variance of the allocation proportions can be obtained. Suppose that

$$\frac{N_{1n}}{n} \to \rho_1(\theta_1, \theta_2)$$

almost surely and

$$\sqrt{n}\left\{\frac{N_{1n}}{n} - \rho_1(\theta_1, \theta_2)\right\} \to \mathrm{N}\{0, V_1(\theta_1, \theta_2)\}$$

in distribution as $n \to \infty$. Then it can be shown that

$$V_1(\theta_1, \theta_2) \geq B(\theta_1, \theta_2) = \frac{\{\frac{\partial \rho_1(\theta_1, \theta_2)}{\partial \theta_1}\}^2}{\rho_1(\theta_1, \theta_2) I_1(\theta_1)} + \frac{\{\frac{\partial \rho_2(\theta_1, \theta_2}{\partial \theta_2}\}^2}{\rho_2(\theta_1, \theta_2) I_2(\theta_2)},$$

where $I_j(\theta_j)$ denotes the Fisher information for a single observation on treatment j for $j = 1, 2$. Any rule that attains this lower bound is called **asymptotically best**.

There are essentially two approaches to response-adaptive randomisation, one based on a class of urn models and the other on a class of adaptive biased coin designs. As an example of the latter, suppose that we wish to minimise a weighted average of the numbers of patients on the two treatments subject to attaining a fixed power, where the weights are functions of the θ_j for $j = 1, 2$. In the binary case, the weights would be the failure probabilities. Then the **optimal** treatment allocation probabilities can be derived and the θ_j replaced by their current maximum likelihood estimates. Thus, we obtain a sequential maximum likelihood estimation rule.

4.3 Examples of adaptive treatment allocation

There are a wide variety of adaptive treatment allocation rules available. Four of the more popular ones are described below.

Example. Efron's biased coin design.
Let the **treatment imbalance** after n assignments be $D_n = N_{1n} - N_{2n}$ and let $0.5 < p < 1$ be a constant. Then the probability that the nth patient is assigned to treatment 1 is $1/2$ if $D_{n-1} = 0$, p if $D_{n-1} < 0$ and $1 - p$ if $D_{n-1} > 0$. It can be shown that

$$\sqrt{n}\left(\frac{N_{1n}}{n} - \frac{1}{2}\right) \to 0$$

in probability as $n \to \infty$. This means that $\mathrm{var}(N_{1n}/n) = o(1/n)$, which shows why this design is so effective in terms of balancing the numbers of patients on the two treatments. In fact, Efron's biased coin design gives a uniformly more powerful Z or t test than complete randomisation.

Example. Adjustable biased coin design.
Let $F(.)$ be a function $F : \mathbf{Z} \to [0, 1]$, where $\mathbf{Z}$ is the set of integers, such that F is non-increasing and $F(-x) = 1 - F(x)$. Then the probability that the nth patient is assigned to treatment 1 is $F(D_{n-1})$. Thus, with this design, the tendency towards balance becomes **stronger** the more we move away from it. It can be shown that

$$\sqrt{n}\left(\frac{N_{1n}}{n} - \frac{1}{2}\right) \to 0$$

almost surely as $n \to \infty$. When $F(x) = p$ for $x < 0$, we have Efron's biased coin design. The adjustable biased coin design yields a uniformly more powerful Z or t test than Efron's biased coin design.

Example. Drop-the-loser rule.
Consider an **urn model** in which there are initially a balls for both treatment types and b immigration balls. The immigration balls are present in order to ensure that the urn does not empty. When a treatment ball is drawn, it is only replaced if the response is a success. If an immigration ball is drawn, it is replaced along with one ball of each treatment type. Assume that the probability of success for treatment j is p_j for $j = 1, 2$. Then it can be shown that

$$\frac{N_{1n}}{n} \to \frac{q_2}{q_1 + q_2}$$

in probability and

$$\sqrt{n}\left(\frac{N_{1n}}{n} - \frac{q_2}{q_1 + q_2}\right) \to \mathrm{N}\left\{0, \frac{q_1 q_2 (p_1 + p_2)}{(q_1 + q_2)^3}\right\}$$

in distribution as $n \to \infty$, where $q_j = 1 - p_j$. Thus, the target allocation for treatment 1 is $\rho_1(p_1, p_2) = q_2/(q_1 + q_2)$. Since $I_j(p_j) = 1/(p_j q_j)$ for $j = 1, 2$, it is easily verified that

$$B(p_1, p_2) = \frac{q_1 q_2 (p_1 + p_2)}{(q_1 + q_2)^3}.$$

Consequently, the drop-the-loser rule is an asymptotically best procedure for the above target allocation.

Example. Sequential maximum likelihood estimation rule.
Suppose that responses are binary and interest lies in minimising the number of treatment failures for a fixed power. Then, if we use the usual large-sample Z test, this means that we need to find the allocation $\rho = \rho_1(p_1, p_2)$ that minimises $q_1 n_1 + q_2 n_2$ subject to

$$\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = C,$$

where C is a constant. Letting $n_1 = \rho n$ and $n_2 = (1 - \rho)n$, we see that

$$n = \frac{p_1 q_1}{\rho C} + \frac{p_2 q_2}{(1 - \rho)C}.$$

Substituting for n in the formula for the number of treatment failures and differentiating with respect to ρ, we obtain the equation

$$\frac{q_1 p_2 q_2}{(1-\rho)^2} - \frac{p_1 q_1 q_2}{\rho^2} = 0.$$

Solving yields

$$\rho = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}}.$$

Since this is a function of the unknown success probabilities, after $n-1$ assignments, we replace p_1 and p_2 by their maximum likelihood estimates based on the first $n-1$ responses. It can be shown that, for this sequential maximum likelihood estimation rule,

$$\frac{N_{1n}}{n} \to \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}}$$

almost surely as $n \to \infty$ and

$$V_1(p_1, p_2) = \frac{p_1^{\frac{3}{2}}(p_2 + q_2/2) + p_2^{\frac{3}{2}}(p_1 + q_1/2)}{(\sqrt{p_1} + \sqrt{p_2})^3 \sqrt{p_1 p_2}}.$$

It is also easily verified that

$$B(p_1, p_2) = \frac{1}{4(\sqrt{p_1} + \sqrt{p_2})^3} \left(\frac{p_2 q_1}{\sqrt{p_1}} + \frac{p_1 q_2}{\sqrt{p_2}} \right).$$

Consequently, this is not an asymptotically best procedure for the above target allocation.

4.4 Group-sequential response-adaptive tests

Up to now, we have considered adaptive treatment allocation in the context of a fixed trial size. Since it is often more efficient to conduct a trial group sequentially, it is natural to investigate the consequences of incorporating adaptive treatment allocation. The formulation of such a group sequential test requires the determination of the joint distribution of sequentially computed test statistics. Because of the **dependencies** induced by adaptive treatment allocation, this is difficult in general.

Let $X_{A1}, X_{A2}, \ldots$ be independent normal random variables with mean μ_A and known variance σ_A^2 and let $X_{B1}, X_{B2}, \ldots$ be independent normal random variables with mean μ_B and known variance σ_B^2. Then response-adaptive randomisation can be incorporated into a general family of group sequential tests **without affecting** the error probabilities if the group sizes do not depend on the estimated mean responses at the previous stage in any other way but through their difference.

Let $\theta = \mu_A - \mu_B$ and suppose that we wish to test $H_0 : \theta = 0$. Then the standardised statistic at analysis k for testing H_0 is

$$Z_k = \left\{ \overline{X}_A^{(k)} - \overline{X}_B^{(k)} \right\} \sqrt{\mathcal{I}_k},$$

where

$$\mathcal{I}_k = \left(\frac{\sigma_A^2}{n_{Ak}} + \frac{\sigma_B^2}{n_{Bk}} \right)^{-1}$$

is the information level. Let m_{Ak} and m_{Bk} denote the group sizes on treatments A and B, respectively, at stage k. Then these are allowed to depend on the accumulated data through the current estimate of θ given by $\hat{\theta}^{(k-1)} = \overline{X}_A^{(k-1)} - \overline{X}_B^{(k-1)}$ and chosen to achieve the specified value of $\mathcal{I}_k$. Under such an adaptive sampling scheme, the above statistics have the canonical joint distribution with information levels $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$ for θ.

To see how such a **group-sequential response-adaptive test** is constructed in practice, suppose that we wish to minimise $u(\theta)n_{A\tau} + v(\theta)n_{B\tau}$, where $u(\theta)$ and $v(\theta)$ are specified weights, and τ denotes the stage at which the test terminates. Then the allocation ratio which minimises this weighted average is

$$\frac{n_{A\tau}}{n_{B\tau}} = \frac{\sigma_A}{\sigma_B} w(\theta),$$

where

$$w(\theta) = \sqrt{\frac{v(\theta)}{u(\theta)}}.$$

Since this is a function of the unknown θ, after $k-1$ stages, we replace θ with its maximum likelihood estimate $\hat{\theta}^{(k-1)}$. This means that we choose m_{Ak} and m_{Bk} so that

$$\frac{n_{Ak}}{n_{Bk}} = \frac{\sigma_A}{\sigma_B} w(\hat{\theta}^{(k-1)})$$

for $k = 1, 2, \ldots, K$. If $w(\theta) = 1$, then sampling is non-adaptive and only gives equal group sizes if $\sigma_A = \sigma_B$.

At the first stage, we take

$$m_{A1} = \sigma_A \{\sigma_A + \sigma_B w(\hat{\theta}^{(0)})\} \mathcal{I}_1$$

and

$$m_{B1} = \sigma_B \{\sigma_A + \sigma_B w(\hat{\theta}^{(0)})\} \mathcal{I}_1 / w(\hat{\theta}^{(0)}),$$

where $\hat{\theta}^{(0)}$ is a preliminary estimate of θ and the group sizes are rounded to integers. If no such estimate is available, we can use $w(\hat{\theta}^{(0)}) = 1$. At stage $2 \leq k \leq K$, assuming that the test has not yet terminated, we take

$$m_{Ak} = \sigma_A \{\sigma_A + \sigma_B w(\hat{\theta}^{(k-1)})\} \mathcal{I}_k - n_{A,k-1}$$

and

$$m_{Ak} = \sigma_B \{\sigma_A + \sigma_B w(\hat{\theta}^{(k-1)})\} \mathcal{I}_k / w(\hat{\theta}^{(k-1)}) - n_{B,k-1}.$$

Again, these group sizes are rounded to integers. If either of their values is negative, that group size is set to zero and sufficient observations are taken on the other treatment to achieve the specified information level. The calculation of the information levels and boundaries for a group-sequential response-adaptive test is the **same** as for its non-adaptive analogue.

References

Atkinson, A.C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69**, 61-67. [4]

Baldi Antognini, A. (2008). A theoretical analysis of the power of biased coin designs. *Journal of Statistical Planning and Inference* **138**, 1792-1798. [4]

Ghosh, M., Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation.* Wiley. [2]

Hu, F. and Rosenberger, W.F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials.* Wiley. [4]

Jennison, C. and Turnbull, B.W. (2000). *Group Sequential Methods with Applications to Clinical Trials.* Chapman and Hall. [3,4]

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer. [1,2]