Research Report
Department of Statistics
Göteborg University
Sweden

Sequential probability ratio tests when using randomized play-the-winner allocation

Anna Ekman

Research Report 1997:1
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Department of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se
Göteborg University	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
S-411 80 Göteborg			
Sweden			

Sequential probability ratio tests when using randomized play-the-winner allocation.

Anna Ekman

Abstract

In many clinical experiments there is a conflict between ethical demands to provide the best possible medical care for the patients and the statisticians desire to obtain an efficient experiment.

Play-the-winner allocations is a group of designs that, during the experiment, tends to place more patients on the treatment that seems to be better. Using a randomized play-the-winner allocation and making a suitable inference for the design, is a suggestion to perform a reasonable experiment for the above mentioned considerations.

In this paper we will concentrate on sequential inference, for the case of simple hypotheses and for the case with simple hypotheses with a nuisance parameter.

The response to treatment is assumed to be dichotomous. We proceed from Wald's sequential probability ratio test, SPRT, and Cox's maximum likelihood SPRT, for the two hypothesis cases above.

Contents

1. Introduction

In many clinical trials, patients enter the study sequentially. The outcome can be examined repeatedly, so that early termination of the study can be considered, when sufficient information is obtained from the experiment.

Consider an experiment where two treatments, say A and B, are compared. The response to the treatment is dichotomous, namely success or failure. The question arises how to allocate patients to the two treatments. A simple allocation rule is to randomly choose a treatment for each patient. Because of ethical considerations the randomized play-the-winner, RPW, allocation has been developed. Play-the-winner allocations was first suggested by Zelen (1969).A characteristic of the Play-the-winner allocations is that the probability for a treatment increases as the number of successes for the treatment increases. Therefore, we can assume that less patients are allocated to the inferior treatment than to the superior one. Three Play-the-winner allocation rules, Play-the-winner (PW) allocation, modified Play-the-winner (MPW) allocation and randomized Play-the-winner (RPW) allocation, are summarized in Section 2.

The RPW will be examined for comparing two treatments, A and B. In Sections 3, 4 and 5 we will examine the case of simple hypotheses,

$$H_0 : p_A = p_B = p_{A0} = p_{B0} \text{ vs. } H_1 : p_A = p_{A1}, p_B = p_{B1} ,$$

using Wald's sequential probability ratio test, SPRT. It will be assumed that $p_{A0} \neq p_{A1}$ and $p_{B0} \neq p_{B1}$.

Wald's SPRT is originally presented in the case of independent identically distributed random variables. If a play-the-winner allocation is used the variables are not independent and identically distributed: the treatment of the current patient depends on the treatment,and the response to the treatment, of one or more of the previous patients. We will present a generalization of the test which is suitable for this more complicated situation. We will assume immediate responses, to simplify the calculations. In Section 3 we recall Wald's SPRT and some of its basic properties. In Section 4 the test is generalized to our experimental situation, where the random variables are not independent and identically distributed. As main result, of the first part of the paper, we will show that the error probabilities have a certain bound, for inference based on Wald's SPRT, when the RPW allocation rule is used in the experimental setting of allocating patients to the two treatments. In Section 5 we compare two Play-the-winner allocation rules, the RPW and the MPW, and total randomization (TR), when using the SPRT, by means of expected sample size and expected number of patients allocated to the inferior treatment. TR assigns treatment A with probability $\frac{1}{2}$ and treatment B with probability $\frac{1}{2}$, each time a patient

arrives. Hence, TR represents an allocation rule for which the random variables are independent and identically distributed. The expectations are calculated through simulations.

In Sections 6, 7 and 8 we will discuss a test of simple hypotheses,

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1 \ ,$$

in the presence of a nuisance parameter, μ. The parametrization used is

$$\theta = \ln \frac{p_A}{1-p_A} - \ln \frac{p_B}{1-p_B}$$

$$\mu = \frac{\ln \frac{p_A}{1-p_A} + \ln \frac{p_B}{1-p_B}}{2} \ .$$

This can easily be extended to testing a composite hypothesis in the presence of a nuisance parameter. The test, to be investigated, is the maximum likelihood SPRT proposed by Cox (1963). Bartlett (1946) also proposed a maximum likelihood SPRT. These two test are closely related. In both tests the maximum likelihood estimation of the nuisance parameter is used in the likelihood ratio. In Bartlett's test both the maximum likelihood estimation under the null hypothesis and the maximum likelihood estimation under the alternative hypothesis are used in the likelihood ratio, and in Cox's test the maximum likelihood estimation obtained by solving the normal equations for both the parameter of interest and the nuisance parameter is used in the likelihood ratio. The tests are asymptotically equal. For a further discussion of this and of the asymptotic behavior of these tests, a paper by Holm (1985) is recommended.

In Section 6 we recall Cox's maximum likelihood SPRT. We give a quite detailed derivation of it and describe the connection with the theory of Wald's SPRT. Cox's test in the RPW-case is derived in Section 7. The properties of Cox's test, when using RPW allocation, are discussed in Section 8. To investigate the behavior of the error probabilities, the expected sample size and the expected number of patients assigned to the inferior treatment, a simulation study is performed. The result and a discussion can also be found in Section 8.

2. Play-the-winner allocations

The first Play-the-winner allocation was introduced by Zelen (1969). Zelen's allocation, denoted by PW, is best described by an urn. For each successful treatment, we put a ball in the urn representing this treatment, and for each failure we put a ball representing the other treatment. When a patient is to be allocated to a treatment we draw a ball from the urn, without replacement. If the urn is empty, as it is at the start, each treatment has probability $\frac{1}{2}$. This

method allows the responses to be delayed, but if many responses are delayed for substantial time treatments will mostly be assigned with probability $\frac{1}{2}$ each.

In the same paper Zelen introduced another allocation rule, where the responses were assumed to be immediate. When allocating the first patient, the treatments have probability $\frac{1}{2}$ each to be assigned to the patient. For the following allocations one keeps on assigning the same treatment until it gives a failure, then switches to the other treatment and keeps assigning this one until it gives a failure. The allocation rule is denoted by MPW, modified play-the-winner. Note that with immediate responses the PW allocation is identical to the MPW allocation.

Later, a randomized play-the-winner allocation was introduced by Wei and Durham (1978). Now we can think of an urn with ω_A balls representing treatment A and ω_B balls representing treatment B, at the start. When a patient is to be assigned to a treatment a ball is drawn from the urn, with replacement. If the response "success" is received, from a patient assigned to treatment A (or B respectively), we add ρ A-balls (or ρ B-balls) to the urn. If the response is a "failure" we add ρ B-balls (or ρ A-balls) to the urn. Note that here, the history of successes and failures will affect every allocation, even if the responses are delayed. This last allocation rule is denoted by RPW$(\omega_A, \omega_B, \rho)$. A special case of the RPW$(\omega_A, \omega_B, \rho)$ is when $\omega_A = \omega_B = \omega$, which is denoted by RPW(ω, ρ).

The statistical analysis of results from experiments where RPW(ω, ρ) allocation rules have been used has been discussed by several authors. Wei and Durham (1978) suggested an inverse stopping rule for deciding which of two treatments is the better one. They also compared the RPW$(0, 1)$ with PW, with respect to the expected number of patients treated by the inferior treatment, the average sample size and the estimated probabilities of correct selection of the inferior treatment. The comparison of the average sample sizes was done when the inverse stopping rule was used. Wei and Durham concluded that the RPW$(0, 1)$ seemed to be approximately equal to the PW for practical use.

Wei (1988) used the inverse stopping rule to stop the experiment, and proposed a fixed sample permutation test for the analysis. For comments on the work of Wei (1988), and general comments on difficulties with the inference after using the RPW(ω, ρ) allocation, the discussion by Begg (1990) is recommended. In Wei, Smythe, Lin and Park (1990) exact conditional, exact unconditional and approximate confidence intervals were studied. One of their conclusions was that the design should not be ignored in the analysis. However, in many suggested analyses the authors have chosen either not to include the stopping rule (see the articles mentioned above) and therefore, by our opinion, leaving part of the design out or concentrated on fixed sample sizes (see Rosenberger (1996) for a summary). An interesting collection of reports, on adaptive designs and inference in combination with these, was published after a 1992 joint AMS-IMS-SIAM Summer Conference, Flournoy and Rosenberger (1995).

In the present paper we will use sequential analysis as a suggestion to include

a stopping rule and to handle the inference problem that arises. We have chosen significance tests since these are commonly used and asked for in clinical trials. Where nothing else is mentioned we concentrate on the RPW(1, 1) allocation.

3. Wald's Sequential Probability Ratio Test

In this section Wald's sequential probability ratio test, SPRT, and some of its basic properties are described. The theory below follows unpublished lecture notes, Holm (1990). A good introduction to the theory of sequential analysis can be found, for example, in Govindaraluju (1981), Ghosh (1970) or Siegmund (1985).

3.1. The test

We are interested in discriminating between two simple hypotheses

$$H_0 : \theta = \theta_0,\ H_1 : \theta = \theta_1 \text{ , where } \theta_0 \neq \theta_1 \text{ ,}$$

with a sequential probability ratio test, with desired significance level α and desired power in the alternative $1 - \beta$. The test is constructed as follows.

Assume that $X_1, \ldots, X_n$ are independent identically distributed random variables with probability distribution function $f_\theta(.)$, and with joint probability distribution function $f_{n,\theta}(.)$. Then the likelihoods ratio λ_n can be written as

$$\lambda_n = \frac{f_{n,\theta_1}(\mathbf{x}_n)}{f_{n,\theta_0}(\mathbf{x}_n)}$$

where $\mathbf{x}_n = (x_1, \ldots, x_n)$.

Let A and B be absorbing barriers. Then we have three possibilities:

$$B < \lambda_n < A \text{ continue with an additional observation}$$

$$\lambda_n \leq B \text{ stop the experiment and accept } H_0$$

$$\lambda_n \geq A \text{ stop the experiment and reject } H_0$$

Often it is more practical to work with the log likelihood ration $\ln \lambda_n$, $a = \ln A$ and $b = \ln B$.

3.2. Some important properties

The following results enable us to choose the bounds A and B, so that the true significance level, α^*, and the true power under the alternative, $1 - \beta^*$, will be close to the desired ones. Equations 3.2.1 and 3.2.2 give us the bounds of the true α^* and $1 - \beta^*$, while α and $1 - \beta$ are desired.

The proofs of the propositions below are included to illustrate that the proof of Proposition 3.2 is the only one that requires the assumption of independent and identically distributed random variables.

Proposition 3.1 :Assume that $P(N < \infty) = 1$. Then for given A and B

$$A \leq \frac{1-\beta^*}{\alpha^*} \text{ and } B \geq \frac{\beta^*}{1-\alpha^*} ,$$

where α^* and β^* are the true type-I and type-II errors.

Proof. Let

$$R_n = \{\mathbf{x}_n; N = n, \lambda_N \geq A\} .$$

Therefore, the $R_n's$ are mutually disjoint. Now we can write

$$\alpha^* = P_{\theta_0}(\lambda_N \geq A) = \sum_{n=1}^{\infty} P_{\theta_0}(R_n) = \sum_{n=1}^{\infty} \int_{R_n} f_{n,\theta_0}(\mathbf{x}_n)\, d\mathbf{x}_n .$$

Remembering that

$$\frac{f_{n,\theta_1}}{f_{n,\theta_0}} \geq A$$

on R_n we obtain

$$\alpha^* = \sum_{n=1}^{\infty} \int_{R_n} f_{n,\theta_0}(\mathbf{x}_n)\, d\mathbf{x}_n \leq \frac{1}{A} \sum_{n=1}^{\infty} \int_{R_n} f_{n,\theta_1}(\mathbf{x}_n)\, d\mathbf{x}_n = \frac{1}{A} P_{\theta_1}(\lambda_N \geq A) = \frac{1}{A}(1-\beta^*) .$$

Hence,

$$A \leq \frac{1-\beta^*}{\alpha^*} .$$

Similarly we can show that

$$B \geq \frac{\beta^*}{1-\alpha^*}.$$

□

Proposition 3.2 : If $X_1, \ldots, X_n$ are independent and identically distributed random variables and

$$P\left(\ln \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)} = 0\right) < 1$$

then

$$P(N < \infty) = 1 .$$

Proof. Let $\Delta > 0$. Then

$$P\left(\ln\frac{f_{n,\theta_1}(X_i)}{f_{n,\theta_0}(X_i)}=0\right)=\lim_{\Delta\to 0}P\left(-\Delta\le\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}\le\Delta\right)<1$$

implies that

$$\exists\Delta\ ;\ P\left(\left|\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}\right|\le\Delta\right)=\gamma<1 \text{ for some } \gamma>0\ .$$

Furthermore,

$$P\left(\left|\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}\right|>\Delta\right)=P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta\right)+P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta\right)$$

$$\le 2\max\left\{P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta\right),P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta\right)\right\}\ ,$$

and hence,

$$\max\left\{P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta\right),P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta\right)\right\}\ge\frac{1-\gamma}{2}\ ,$$

Let

$$n_0=\left[\frac{a-b}{\Delta}\right]\ .$$

Then

$$P\left(\sum_{i=1}^{n_0}\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}\notin(b,a)\right)\ge P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta,\ \forall i\right)+P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta,\ \forall i\right)$$

$$=\prod_{i=1}^{n_0}P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta\right)+\prod_{i=1}^{n_0}P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta\right)$$

$$=\left(P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta\right)\right)^{n_0}+\left(P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta\right)\right)^{n_0}$$

$$\ge\left(\max\left\{P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}<-\Delta\right),P\left(\ln\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}>\Delta\right)\right\}\right)^{n_0}\ge\left(\frac{1-\gamma}{2}\right)^{n_0}>0\ .$$

Furthermore, let us denote

$$\varepsilon=\left(\frac{1-\gamma}{2}\right)^{n_0}\ .$$

We now have that

$$P\left(N>kn_0\right)\le(1-\varepsilon)^k\ ,$$

that implies

$$\lim_{k\to\infty} P(N > kn_0) = 0 .$$

Therefore

$$P(N < \infty) = 1 .$$

□

Proposition 3.3 : Assume that $P(N < \infty) = 1$. If A and B are chosen to be

$$A = \frac{1-\beta}{\alpha} \text{ and } B = \frac{\beta}{1-\alpha} ,$$

where α and β are the desired levels of error probabilities, then the true level of significance, α^*, and the true power under the alternative, $1-\beta^*$, satisfy

$$\alpha^* + \beta^* \leq \alpha + \beta .$$

Proof. As $P(N < \infty) = 1$ is assumed, the requirement of Proposition 3.1 is satisfied. Hence

$$\begin{cases} A\alpha^* \leq 1-\beta^* \\ \beta^* \leq B(1-\alpha^*) \end{cases} .$$

Now it follows that

$$\begin{cases} \frac{1-\beta}{\alpha}\alpha^* \leq 1-\beta^* \\ \beta^* \leq \frac{\beta}{1-\alpha}(1-\alpha^*) \end{cases} ,$$

which implies that

$$\begin{cases} \alpha^*(1-\beta) \leq \alpha(1-\beta^*) \\ \beta^*(1-\alpha) \leq \beta(1-\alpha^*) \end{cases} .$$

Thus,

$$\alpha^*(1-\beta) + \beta^*(1-\alpha) \leq \alpha(1-\beta^*) + \beta(1-\alpha^*)$$

and finally

$$\alpha^* + \beta^* \leq \alpha + \beta .$$

□

Note that Proposition 3.2 implies Proposition 3.1, which implies Proposition 3.3. The allocation rule TR, where the two treatments are allocated with probability $\frac{1}{2}$ each, satisfy the requirement for Propositions 3.1-3.3, as the random variables in this case are independent and identically distributed.

By using Proposition 3.1 and Proposition 3.3 we obtain

$$\frac{1-B}{A^+-B} \leq \alpha^* \leq \frac{1-B^-}{A-B^-} \qquad (3.2.1)$$

and

$$A^+\frac{1-B}{A^+-B} \leq 1-\beta^* \leq A\frac{1-B^-}{A-B^-} \qquad (3.2.2)$$

where A^+ is the maximum value of the likelihood ratio, when stopping and rejecting H_0, and B^- is the minimum value of the likelihood ratio, when stopping and accepting H_0.

4. SPRT for a response dependent allocation

We investigate an allocation rule that creates dependence between the random variables. Therefore we need to generalize the theory of Wald's SPRT to this situation. We do this for our specific experimental setting, where we want to compare two treatments that both have two possible responses, success and failure.

In Section 4.1 we derive the log likelihood ratio, $\ln \lambda_n$, and in Section 4.2 we show that the properties, discussed in Section 3.2, are true also for the generalized SPRT. These properties enable us to construct a generalized test in the same way as the ordinary Wald's SPRT is constructed.

4.1. The log likelihood ratio

To simplify the calculations immediate responses are assumed. From now on we use the following notations :

$$\begin{aligned} p_T &= \text{ the probability of success for treatment T} \\ S_{T(i-1)} &= \text{ number of successes of treatment T among the } (i-1) \text{ first patients} \\ F_{T(i-1)} &= \text{ number of failure of treatment T among the } (i-1) \text{ first patients} \end{aligned}$$

where T is either A and B. Furthermore, it is assumed that $0 < p_A < 1$, $0 < p_B < 1$ and that $p_A \geq p_B$.

Simple hypotheses of the following kind are considered

$$H_0 : p_A = p_B = p_{A0} = p_{B0} \text{ versus } H_1 : p_A = p_{A1} \text{ , } p_B = p_{B1} \text{ .}$$

It is assumed that

$$p_{A0} \neq p_{A1} \text{ and } p_{B0} \neq p_{B1} \text{ .}$$

Our response variables are

$$Y_i = \begin{cases} 1 \text{ if patient i was allocated to treatment A} \\ 0 \text{ if patient i was allocated to treatment B} \end{cases}$$

$$X_i = \begin{cases} 1 \text{ if the treatment on patient i resulted in a success} \\ 0 \text{ if the treatment on patient i resulted in a failure} \end{cases}$$

In the following we will use the notation $\{X\}_i^j = \{X_i, \ldots, X_j\}$.

The likelihood function can be determined as

$$L_N = P\left(Y_1 = y_1, X_1 = x_1\right)$$

$$* \prod_{i=2}^{N} \left\{ P\left(Y_i = y_i \left| \{Y\}_1^{i-1} = \{y\}_1^{i-1}, \{X\}_1^{i-1} = \{x\}_1^{i-1} \right.\right) P\left(X_i = x_i \left| Y_i = y_i \right.\right) \right\}$$

$$= P(Y_1 = y_1) * \prod_{i=2}^{N} p_A^{y_i x_i} (1-p_A)^{y_i(1-x_i)} p_B^{(1-y_i)x_i} (1-p_B)^{(1-y_i)(1-x_i)}$$

$$* \left(P\left(Y_i = 1 \middle| \{Y\}_1^{i-1} = \{y\}_1^{i-1}, \{X\}_1^{i-1} = \{x\}_1^{i-1} \right) \right)^{y_i}$$

$$* \left(1 - P\left(Y_i = 1 \middle| \{Y\}_1^{i-1} = \{y\}_1^{i-1}, \{X\}_1^{i-1} = \{x\}_1^{i-1} \right) \right)^{(1-y_i)}$$

Hence, the likelihood ratio is

$$\lambda_N = \prod_{i=2}^{N} \left(\frac{p_{A1}}{p_{A0}} \right)^{y_i x_i} \left(\frac{1-p_{A1}}{1-p_{A0}} \right)^{y_i(1-x_i)} \left(\frac{p_{B1}}{p_{B0}} \right)^{(1-y_i)x_i} \left(\frac{1-p_{B1}}{1-p_{B0}} \right)^{(1-y_i)(1-x_i)}$$

The test statistic we use is the log likelihood ratio, the same as in Wald's SPRT. It can be written as

$$\ln \lambda_N = \sum_{i=2}^{N} \left[y_i x_i \ln \left(\frac{p_{A1}}{p_{A0}} \right) + y_i (1-x_i) \ln \left(\frac{1-p_{A1}}{1-p_{A0}} \right) \right.$$
$$\left. + (1-y_i) x_i \ln \left(\frac{p_{B1}}{p_{B0}} \right) + (1-y_i)(1-x_i) \ln \left(\frac{1-p_{B1}}{1-p_{B0}} \right) \right]$$

$$= S_{A(N)} \ln \left(\frac{p_{A1}}{p_{A0}} \right) + F_{A(N)} \ln \left(\frac{1-p_{A1}}{1-p_{A0}} \right) + S_{B(N)} \ln \left(\frac{p_{B1}}{p_{B0}} \right) + F_{B(N)} \ln \left(\frac{1-p_{B1}}{1-p_{B0}} \right)$$

4.2. Some properties

Assuming that Condition 1 below holds we will show Proposition 4.1, the correspondence to Proposition 3.2. This implies Proposition 3.1, which implies Proposition 3.3.

Condition 1 :

$$P(Y_{i+1} = 1 \,|N > i) \leq P(Y_{i+2} = 1 \,|N > i, Y_{i+1} = 1, X_{i+1} = 1) \leq \ldots$$
$$\leq P\left(Y_{i+m} = 1 \middle| N > i, \{Y\}_{i+1}^{i+m-1} = \{1\}_{i+1}^{i+m-1}, \{X\}_{i+1}^{i+m-1} = \{1\}_{i+1}^{i+m-1} \right)$$

and

$$P(Y_{i+1} = 0 \,|N > i) \leq P(Y_{i+2} = 0 \,|N > i, Y_{i+1} = 0, X_{i+1} = 1) \leq \ldots$$
$$\leq P\left(Y_{i+m} = 0 \middle| N > i, \{Y\}_{i+1}^{i+m-1} = \{0\}_{i+1}^{i+m-1}, \{X\}_{i+1}^{i+m-1} = \{1\}_{i+1}^{i+m-1} \right) ,$$
$$\forall \text{ m}$$

That is, we will assume that, given that the process has not stopped at stage i, the probability to allocate treatment A to a patient will increase, given that we allocate treatment A to every patient from stage i and on and that the responses all turn out to be successes.

Proposition 4.1 : If Condition 1 is satisfied, then $P(N < \infty) = 1$.

Proof. The following is true for all values of i.

Let $m = \left[\frac{a-b}{t}\right]+1$, where $t = \left[\begin{array}{c}\text{the absolute value of the possible} \\ \text{increments of the log likelihood, in one step}\end{array}\right]$.

By the assumptions that $p_{A0} \neq p_{A1}$ and $p_{B0} \neq p_{B1}$, we have that

$$t = \min\left[\left|\ln\left(\frac{p_{A1}}{p_{A0}}\right)\right|, \left|\ln\left(\frac{1-p_{A1}}{1-p_{A0}}\right)\right|, \left|\ln\left(\frac{p_{B1}}{p_{B0}}\right)\right|, \left|\ln\left(\frac{1-p_{B1}}{1-p_{B0}}\right)\right|\right] > 0 \ .$$

Let us denote

$$C_{A,i+1} = \{(Y_{i+1}=1, X_{i+1}=1), \ldots, (Y_{i+m}=1, X_{i+m}=1)\}$$

and

$$C_{B,i+1} = \{(Y_{i+1}=0, X_{i+1}=1), \ldots, (Y_{i+m}=0, X_{i+m}=1)\} \ .$$

That is, we look at two events such that one treatment is allocated for m steps in a row and the responses from all these allocations are successes.

The probability of the union of these two events, given that the process has not stopped at or before stage i, is then

$$P(C_{A,i+1} \cup C_{B,i+1} \,|N > i) = P(C_{A,i+1} \,|N > i) + P(C_{B,i+1} \,|N > i)$$

$$= P(Y_{i+1}=1 \,|N > i)\, P(X_{i+1}=1 \,|Y_{i+1}=1)$$

$$* \prod_{j=2}^{m} P\left(Y_{i+j}=1 \,\middle|\, N > i, \{Y\}_{i+1}^{i+j-1} = \{1\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1} = \{1\}_{i+1}^{i+j-1}\right) P(X_{i+j}=1 \,|Y_{i+j}=1)$$

$$+P(Y_{i+1}=0 \,|N > i)\, P(X_{i+1}=1 \,|Y_{i+1}=0)$$

$$* \prod_{j=2}^{m} P\left(Y_{i+j}=0 \,\middle|\, N > i, \{Y\}_{i+1}^{i+j-1} = \{0\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1} = \{1\}_{i+1}^{i+j-1}\right) P(X_{i+j}=1 \,|Y_{i+j}=0) \ .$$

As a result of condition 1 we obtain

$$P(C_{A,i+1} \cup C_{B,i+1} \,|N > i) \geq \begin{cases} \left(\frac{1}{2}p_A\right)^m & \text{if } P(Y_{i+1}=1 \,|N > i) \geq \frac{1}{2} \\ \left(\frac{1}{2}p_B\right)^m & \text{if } P(Y_{i+1}=0 \,|N > i) \geq \frac{1}{2} \end{cases}$$

and therefore

$$P(C_{A,i+1} \cup C_{B,i+1} \,|N > i) \geq \min\left(\left(\frac{1}{2}p_A\right)^m, \left(\frac{1}{2}p_B\right)^m\right) \ .$$

Note that $C_{A,i+1}$ and $C_{B,i+1}$ are two possible events, but not the only ones, for hitting the boundary in m steps. Hence,

$$P(N=\infty) \leq P(N > n_0+m, N > n_0+2m, \ldots)$$

$$= \lim_{r\to\infty} P(N > n_0 + m) \prod_{j=2}^{r} P(N > n_0 + jm \,|N > n_0 + (j-1)m)$$

$$= \lim_{r\to\infty} (1 - P(N \leq n_0 + m)) \prod_{j=2}^{r} (1 - P(N \leq n_0 + jm \,|N > n_0 + (j-1)m))$$

$$\leq \lim_{r\to\infty} \prod_{j=2}^{r} \left(1 - P\left(C_{A,n_0+(j-1)m+1} \cup C_{B,n_0+(j-1)m+1} \,|N > n_0 + (j-1)m\right)\right)$$

$$\leq \lim_{r\to\infty} \left(1 - \min\left[\left(\frac{1}{2}p_A\right)^m, \left(\frac{1}{2}p_B\right)^m\right]\right)^{r-1} = 0 .$$

We then have that

$$P(N = \infty) = 0 ,$$

and hence

$$P(N < \infty) = 1 .$$

□

This implies Proposition 3.1. We can now construct the SPRT in the same way as the original Wald's SPRT.

4.3. Properties when using RPW and MPW

In Sections 4.3.1 and 4.3.2 we show that Condition 1 is satisfied for RPW and MPW, and for PW, as we assume immediate responses.

4.3.1. RPW

At the start of the experiment we have ω balls representing each treatment in the urn. When receiving a response ρ balls are added to the urn : balls of type A if we received a success for treatment A or a failure for treatment B and balls of type B if we received a success for treatment B or a failure of treatment A.

In the RPW case the probability of allocating a patient to treatment A is

$$P(Y_i = 1) = \frac{\omega + \rho\left(S_{A(i-1)} + F_{B(i-1)}\right)}{2\omega + \rho(i-1)} .$$

To see that Condition 1 is satisfied for the RPW allocation we need to check both inequalities. To show that the first one holds we first write

$$P(Y_{i+1} = 1 \,|N > i) = \frac{P(Y_{i+1} = 1 \,|N > i)(2\omega + \rho i) + \rho P(Y_{i+1} = 1 \,|N > i)}{(2\omega + \rho(i+1))}$$

$$\leq \frac{P\left(Y_{i+1}=1\,|N>i\right)(2\omega+\rho i)+\rho}{(2\omega+\rho\,(i+1))}$$

$$= P\left(Y_{i+2}=1\,|N>i, Y_{i+1}=1, X_{i+1}=1\right)\ ,$$

and, for j = 2,3,. . . ,

$$P\left(Y_{i+j}=1\,\middle|N>i, \{Y\}_{i+1}^{i+j-1}=\{1\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1}=\{1\}_{i+1}^{i+j-1}\right)$$

$$\leq \frac{P\left(Y_{i+j}=1\,\middle|N>i, \{Y\}_{i+1}^{i+j-1}=\{1\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1}=\{1\}_{i+1}^{i+j-1}\right)(2\omega+\rho\,(i+j-1))+\rho}{(2\omega+\rho\,(i+j))}$$

$$P\left(Y_{i+j+1}=1\,\middle|N>i, \{Y\}_{i+1}^{i+j}=\{1\}_{i+1}^{i+j}, \{X\}_{i+1}^{i+j}=\{1\}_{i+1}^{i+j}\right)\ .$$

Hence,

$$P\left(Y_{i+1}=1\,|N>i\right) \leq P\left(Y_{i+2}=1\,|N>i, Y_{i+1}=1, X_{i+1}=1\right) \leq \ldots$$

$$\leq P\left(Y_{i+m}=1\,\middle|N>i, \{Y\}_{i+1}^{i+m-1}=\{1\}_{i+1}^{i+m-1}, \{X\}_{i+1}^{i+m-1}=\{1\}_{i+1}^{i+m-1}\right)\ .$$

Now we want to show the second part of Condition 1. Note that

$$P\left(Y_{i+1}=1\,|N>i\right)=\frac{P\left(Y_{i+1}=1\,|N>i\right)(2\omega+\rho i)+\rho P\left(Y_{i+1}=1\,|N>i\right)}{(2\omega+\rho\,(i+1))}$$

$$\geq \frac{P\left(Y_{i+1}=1\,|N>i\right)(2\omega+\rho i)}{(2\omega+\rho\,(i+1))}$$

$$= P\left(Y_{i+2}=1\,|N>i, Y_{i+1}=0, X_{i+1}=1\right)\ .$$

Note also that, for j = 2, 3, . . .

$$P\left(Y_{i+j}=1\,\middle|N>i, \{Y\}_{i+1}^{i+j-1}=\{0\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1}=\{1\}_{i+1}^{i+j-1}\right)$$

$$\geq \frac{P\left(Y_{i+j}=1\,\middle|N>i, \{Y\}_{i+1}^{i+j-1}=\{0\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1}=\{1\}_{i+1}^{i+j-1}\right)(2\omega+\rho\,(i+j-1))}{(2\omega+\rho\,(i+j))}$$

$$P\left(Y_{i+j+1}=1\,\middle|N>i, \{Y\}_{i+1}^{i+j}=\{0\}_{i+1}^{i+j}, \{X\}_{i+1}^{i+j}=\{1\}_{i+1}^{i+j}\right)\ .$$

Hence,

$$1 - P\left(Y_{i+1} = 1 \,|N > i\right) \leq 1 - P\left(Y_{i+2} = 1 \,|N > i, Y_{i+1} = 0, X_{i+1} = 1\right) \leq \ldots$$

$$\leq 1 - P\left(Y_{i+m} = 1 \,\middle|N > i, \{Y\}_{i+1}^{i+m-1} = \{0\}_{i+1}^{i+m-1}, \{X\}_{i+1}^{i+m-1} = \{1\}_{i+1}^{i+m-1}\right) ,$$

which is equivalent to

$$P\left(Y_{i+1} = 0 \,|N > i\right) \leq P\left(Y_{i+2} = 0 \,|N > i, Y_{i+1} = 0, X_{i+1} = 1\right) \leq \ldots$$

$$\leq P\left(Y_{i+m} = 0 \,\middle|N > i, \{Y\}_{i+1}^{i+m-1} = \{0\}_{i+1}^{i+m-1}, \{X\}_{i+1}^{i+m-1} = \{1\}_{i+1}^{i+m-1}\right) .$$

Hence, it is proved that Condition 1, and therefore Proposition 4.1, 3.1 and 3.3, are satisfied for the RPW allocation.

4.3.2. MPW

In the MPW case the probability of allocating a patient to treatment A equals 1 given that the previous treatment was A and the response was a success or if the previous treatment was B and the response was a failure, and for the probability of allocating a patient to treatment B, respectively. Therefore it is easy to see that the MPW, and therefore also the PW, satisfies Condition 1, as shown below.

To show the first inequality of Condition 1 note that

$$P\left(Y_{i+1} = 1 \,|N > i\right) = 1 \Rightarrow P\left(Y_{i+2} = 1 \,|N > i, Y_{i+1} = 1, X_{i+1} = 1\right) = 1$$

and, for j = 2,3,. . . ,

$$P\left(Y_{i+j} = 1 \,\middle|N > i, \{Y\}_{i+1}^{i+j-1} = \{1\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1} = \{1\}_{i+1}^{i+j-1}\right)$$

$$= P\left(Y_{i+j+1} = 1 \,|N > i, Y_{i+j-1} = 1, X_{i+j-1} = 1\right) = 1 .$$

Hence, the first part of Condition 1 is satisfied. To show the second part of Condition 1 note that

$$P\left(Y_{i+1} = 0 \,|N > i\right) \leq 1 = P\left(Y_{i+2} = 0 \,|N > i, Y_{i+1} = 0, X_{i+1} = 1\right)$$

and for j = 2,3,. . . ,

$$P\left(Y_{i+j} = 0 \,\middle|N > i, \{Y\}_{i+1}^{i+j-1} = \{0\}_{i+1}^{i+j-1}, \{X\}_{i+1}^{i+j-1} = \{1\}_{i+1}^{i+j-1}\right)$$

$$= P\left(Y_{i+j+1} = 0 \,|N > i, Y_{i+j-1} = 0, X_{i+j-1} = 1\right) = 1 .$$

Therefore, Condition 1 is satisfied also for the MPW allocation.

5. Comparison of three allocation rules

The RPW(ω, ρ) is constructed to allocate more patients to the treatment that, during the experiment, seems to be better. We would like to know if the RPW(ω, ρ) allocates fewer patients to the inferior treatment, than to the treatment superior in reality.

Each allocation to the treatment inferior in reality can be viewed as a loss. It is therefore of interest to minimize the number of patients allocated to that treatment.

Another important question is how the expected sample size behaves when one uses a RPW(ω, ρ) allocation, compared to other allocation rules.

We compare the RPW(ω, ρ) with the MPW and with total randomization, TR. TR is the simplest randomized allocation rule used and satisfies the requirements for Wald's SPRT (independent identically distributed random variables). MPW, on the other hand, is one of the simplest play-the-winner allocation rules, but given that the response from the previous patient is known, the next allocation is deterministic. RPW(ω, ρ) is response dependent, but not deterministic.

We study differences between the allocation rules by means of the expected sample size, $E[N]$, and the expected number of patients allocated to the inferior treatment, $E[N_B]$. These expectations are complicated to compute for the MPW and RPW, since we do not have independent identically distributed random variables. The probability that patient i is allocated to treatment A depends on all the earlier allocations and responses. Simulations were therefore conducted to investigate the behavior of the expected values $E[N_B]$ and $E[N]$.

5.1. Description of the simulation study

The RPW was simulated with five different combinations of the parameters ω and ρ, namely RPW(100.000, 1), RPW(10, 1), RPW(1, 1), RPW(1, 10) and RPW(1, 100.000).

At the start of the experiment the allocation probability is $\frac{1}{2}$ for each treatment, for every value on ω and ρ. The larger the ratio $\frac{\rho}{\omega}$ is, the faster the response affects the allocation probability.

Large response dependency, in the early states of the experiment, could be hard to get accepted when the allocation rule is used in a real practical setting, but to get a good understanding of how the RPW allocation behaves, extreme response dependency is included.

As ω increases the RPW gets closer to total randomization. High values of ω let the play-the-winner quality come slower into the experiment. Correspondingly for ρ, the RPW gets closer to the MPW as ρ increases.

For each of the three allocation rules two different hypothesis cases were tested, namely

$$1.\ H_0 : p_A = p_B = 0.7 \text{ vs. } H_1 : p_A = 0.8\ ,\ p_B = 0.6$$

and

$$2.\ H_0 : p_A = p_B = 0.6 \text{ vs. } H_1 : p_A = 0.8\ ,\ p_B = 0.4$$

Hypothesis case 1. represents a small treatment difference, while 2. represents a larger treatment difference. Simulations were done in these two cases both under the assumption that H_0 is false and H_1 is true, and under the assumption that H_0 is true and H_1 is false.

We used Wald's sequential probability ratio test, with significance level $\alpha = 0.05$ and power under the alternative $1 - \beta = 0.95$.

The true significance level, α^*, and the true power under the alternative, $1 - \beta^*$, are bounded as below (see equations 3.2.1 and 3.2.2).

Hypothesis case 1 :

$$0.0375 \approx \tfrac{54}{1441} \leq \alpha^* \leq \tfrac{55}{1081} \approx 0.0509$$
$$0.9493 \approx \tfrac{4104}{4323} \leq 1 - \beta^* \leq \tfrac{1045}{1081} \approx 0.9667$$

Hypothesis case 2 :

$$0.0333 \approx \tfrac{36}{1081} \leq \alpha^* \leq \tfrac{37}{721} \approx 0.0513$$
$$0.9491 \approx \tfrac{1026}{1081} \leq 1 - \beta^* \leq \tfrac{703}{721} \approx 0.9750$$

For each of the four cases, two hypothesis under two different assumptions, 500.000 independent experiments were simulated.

5.2. Results of the simulations

For hypothesis case 1. (see figure 5.2.1), the sample sizes were about the same for total randomization and RPW(100.000, 1). The sample size seems to decrease with decreasing ω. Values of ρ seem not to affect the sample size in a major way: when ω is held constant, $\omega = 1$, and ρ is increasing there is no difference in sample size for the chosen values of ρ.

The results for Hypothesis case 2. were similar, see figure 5.2.2, but the differences in the average sample sizes were small, and might therefore not be of importance in practical settings.

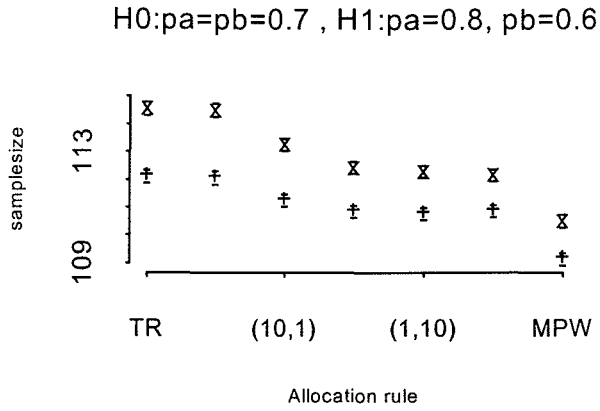

FIGURE 5.2.1 : The mean, $\pm$ two times the standard error. On the x-axis: TR, RPW(100.000, 1), RPW(10, 1), RPW(1, 1), RPW(1, 10), RPW(1, 100.000), MPW. $\mathbf{x} = H_1$ is true , $+ = H_0$ is true.

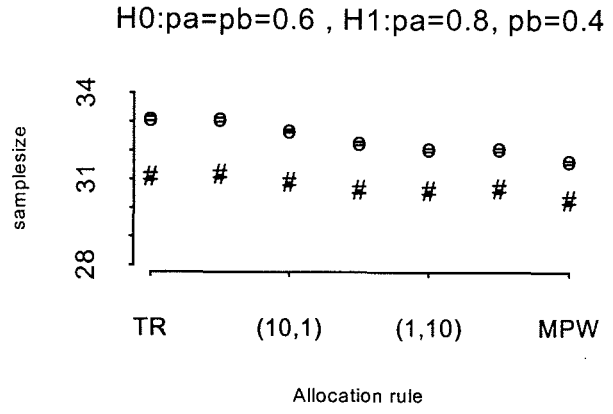

FIGURE 5.2.2 : The mean, $\pm$ two times the standard error. On the x-axis: TR, RPW(100.000, 1), RPW(10, 1), RPW(1, 1), RPW(1, 10), RPW(1, 100.000), MPW. $\mathbf{a} = H_1$ is true , $\mathbf{b} = H_0$ is true.

Note that the sample size follows approximately the same pattern both under the null hypothesis and under the alternative hypothesis, but that it is slightly smaller under the null hypothesis. That is, when the treatments are as good the

sample size is smaller than when there is a true difference between them. Note that treatment B is the inferior treatment when there is a treatment difference.

The behavior of the number of patients allocated to treatment B follows a different pattern than the sample size, see figures 5.2.3 and 5.2.4, except for total randomization and RPW(100.000, 1). These two allocation rules behave quite similarly, and they are the allocation rules that allocate more patients to treatment B than the others, both when B is inferior to A and when A and B are equal.

When there is a treatment difference MPW seems to allocate the least number of patients to treatment B, the inferior treatment, and the number of patients allocated to B decreases with increasing $\frac{\rho}{\omega}$, which was expected.

When there is no difference between the treatments one could think that the number of allocations to the two treatments would be the same. This is true for total randomization and for RPW(100.000, 1). On the other hand, the RPW allocations and the MPW allocation still allocates fewer patients to treatment B than to A. This can be understood by looking at the Wald statistic for the RPW and the MPW. Note that the increments of the Wald statistic are not symmetric, but differs for the four possible events in one step.

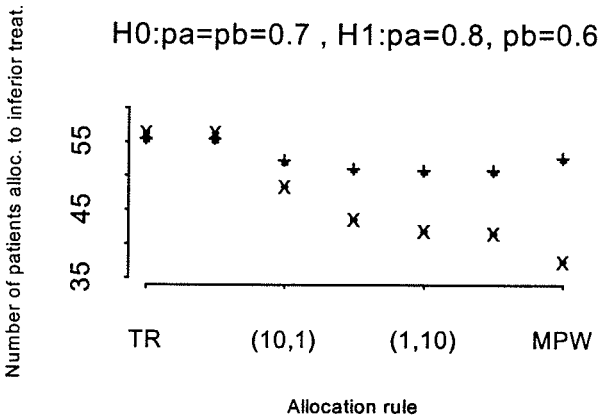

FIGURE 5.2.3 : The mean, ± two times the standard error. On the x-axis: TR, RPW(100.000, 1), RPW(10, 1), RPW(1, 1), RPW(1, 10), RPW(1, 100.000), MPW. **o** = H_1 is true , □ = H_0 is true.

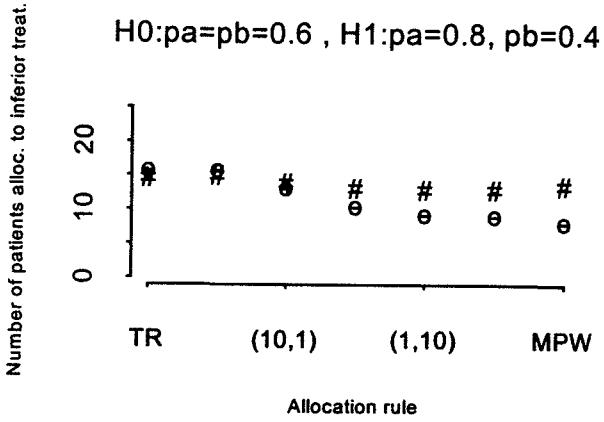

FIGURE 5.2.4 : The mean, $\pm$ two times the standard error. On the x-axis: TR, RPW(100.000, 1), RPW(10, 1), RPW(1, 1), RPW(1, 10), RPW(1, 100.000), MPW. **a** = H_1 is true , **b** = H_0 is true.

For all allocation rules the true significance level and the true power under the alternative are close to the intended quantities (see the tables in the appendix). Note also that for all allocation rules the standard errors of the means seems to be small (see the figures and the tables in the appendix).

Some interesting remarks are that the RPW gets closer to total randomization as ω increases, and for the smaller treatment difference the sample size is slightly larger for total randomization than for the allocations with response dependency, even under H_0. It means that even if there is no treatment difference we obtain a slightly smaller sample size by using a response dependent allocation rule rather than the total randomization. For the large treatment difference the sample sizes are about the same for all allocation rules compared, but there is, as expected, a large difference in the number of patients allocated to the inferior treatment.

For the case with less treatment difference, $p_A = 0.8$ and $p_B = 0.6$, the actual number of patients allocated to the inferior treatment differs slightly more, between the allocation rules (note especially between RPW(1, 100.000) and MPW), than in the case with the greater treatment difference, $p_A = 0.8$ or $p_B = 0.4$.

5.3. Discussion

Wald's sequential probability ratio test is originally presented for independent identically distributed random variables. As shown in Section 4 some important properties of the test, holds for a certain class of response dependent allocation rules, like RPW and MPW.

We compare the RPW allocation rule to the MPW allocation rule and to total randomization. Two important quantities are the expected sample size and the number of patients allocated to the inferior treatment. However, both expectations are hard to derive theoretically, and thus they were estimated by simulations.

Regarding the expected sample size and the number of patients allocated to the inferior treatment, the MPW is slightly better than the others, for the cases studied. After the MPW comes the RPW, in this aspect. There is, however, a negative characteristic of the MPW to consider. The MPW allocation is non-random in the following sense :

Given that the response from the previous patient is known, the next allocation is deterministic. In addition, the MPW requires immediate responses.

Non-randomness could, for example, lead to selection bias. By selection bias we mean that when the experimenter knows, for certain, which treatment will be assigned to the next patient he may, consciously or unconsciously, bias the experiment by letting this knowledge influence the decision of who is or is not a suitable experimental subject.

One could argue that the disadvantages, non-randomness and immediate responses, could be reduced by using Zelen's PW allocation. It allocates the treatments with probability $\frac{1}{2}$ if there are a lot of delayed responses, and it tends to be close to the MPW allocation if there are few delayed responses.

The RPW(ω, ρ) allocation, for the cases studied, was quite good, both in terms of the expected sample size and in terms of minimizing the number of patients allocated to the inferior treatment, and it does not have the disadvantage of Zelen's MPW allocation, mentioned above.

A general remark on the comparisons, in the cases with simple hypotheses, indicated here, is that if the responses are allowed to affect the allocation enough the play-the-winner rules decreases the number of patients allocated to the inferior treatment, but also decreases the sample size, compared to total randomization. For the RPW(1,1), and for the rules with even more response dependence, the simulations indicated the statement above.

6. Maximum Likelihood SPRT procedures

6.1. The parametrization

Two treatments A and B are of interest to compare. The probability of success is denoted, p_i , $i = A, B$. The aim is to see if the two treatments differ or if they are equally good. We will use the parametrization

$$\theta = \ln \frac{p_A}{1 - p_A} - \ln \frac{p_B}{1 - p_B}$$

$$\mu = \frac{\ln \frac{p_A}{1-p_A} + \ln \frac{p_B}{1-p_B}}{2}$$

where μ is the nuisance parameter. The parameter θ of interest, then is the log odds ratio. In this way we will get a parameter of interest with suitable properties and a good covering of the whole space of possible values (p_A, p_B).

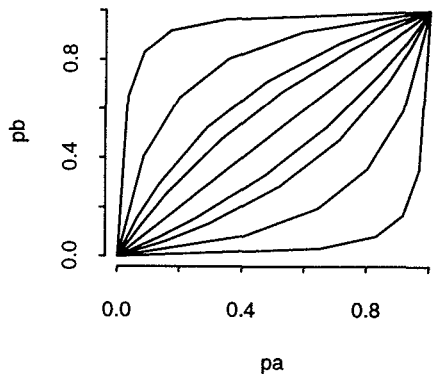

FIGURE 6.1.1 Each arc represents a specific value on θ, and along each arc μ takes values between $-\infty$ and ∞.

Se also Lehman (1991) for discussion of the odds ratios.

6.2. Cox's maximum likelihood SPRT

Suppose we are interested in testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ and that μ a nuisance parameter.

Maximum likelihood SPRT's have been proposed by Bartlett (1946) and Cox (1963).Whether Bartlett's or Cox's test is the most suitable in a specific situation mainly is determined by the maximum likelihood estimations. In our case the maximum likelihood estimation for the nuisance parameter has a quite complex

expression for Bartlett's test. Below Cox's procedure will be introduced. First the log likelihood ratio will be approximated by Taylor's expansion.

Let $f(x,y;\theta,\mu)$ be the common probability density of a sequence of independent identically distributed random variables X_i, where $\theta \in \Theta$ and $\mu \in \Delta$. Let

$$I_{\theta\theta} = Var(\frac{\partial}{\partial\theta}\ln f(x,y;\theta,\mu)) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,y;\theta,\mu)\right]$$

$$I_{\theta\mu} = Cov(\frac{\partial}{\partial\theta}\ln f(x,y;\theta,\mu), \frac{\partial}{\partial\mu}\ln f(x,y;\theta,\mu)) = -E\left[\frac{\partial^2}{\partial\theta\partial\mu}\ln f(x,y;\theta,\mu)\right]$$

Let $\hat{\theta}_n$ and $\hat{\mu}_n$ denote the maximum likelihood estimates of θ and μ based on $(\mathbf{x}_n, \mathbf{y}_n)$. The Taylor's expansions for $\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta_i, \hat{\mu}_n)$ for $i = 0, 1$ about the true (θ, μ) yield

$$\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta_i, \hat{\mu}_n)$$

$$= \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + (\theta_i - \theta)\,\tfrac{\partial}{\partial\theta}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + \left(\hat{\mu}_n - \mu\right)\tfrac{\partial}{\partial\mu}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu)$$

$$+\frac{1}{2}\left[(\theta_i - \theta)^2\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + 2(\theta_i - \theta)\left(\hat{\mu}_n - \mu\right)\frac{\partial^2}{\partial\theta\partial\mu}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu)\right]$$

$$+\left(\hat{\mu}_n - \mu\right)^2\frac{\partial^2}{\partial\mu^2}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + R_2\left((\theta_i - \theta), \left(\hat{\mu}_n - \mu\right)\right)$$

where $R_2\left((\theta_i - \theta), \left(\hat{\mu}_n - \mu\right)\right) = \left((\theta_i - \theta)^2 + \left(\hat{\mu}_n - \mu\right)^2\right)^{3/2} H\left((\theta_i - \theta), \left(\hat{\mu}_n - \mu\right)\right)$.
The function H is bounded in a neighbourhood of $(0,0)$.

Now remember that Wald's SPRT is based on the difference of the log likelihood function under the alternative hypotheses and under the null hypotheses. Using the Taylor expansion above

$$\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta_1, \hat{\mu}_n) - \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta_0, \hat{\mu}_n)$$

$$= (\theta_1 - \theta_0)\frac{\partial}{\partial\theta}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + \frac{1}{2}\left[\left(\theta_1^2 - \theta_0^2 + 2\theta(\theta_0 - \theta_1)\right)\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu)\right]$$

$$+(\theta_1 - \theta_0)\left(\hat{\mu}_n - \mu\right)\frac{\partial^2}{\partial\theta\partial\mu}\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + R_n^*((\mathbf{x}_n, \mathbf{y}_n))$$

$$= (\theta_1 - \theta_0) \frac{\partial}{\partial \theta} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + \frac{1}{2} \left[\left(\theta_1^2 - \theta_0^2 + 2\theta (\theta_0 - \theta_1) \right) \frac{\partial^2}{\partial \theta^2} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) \right]$$

$$+ R_n^* ((\mathbf{x}_n, \mathbf{y}_n))$$

$$= \theta_1 \frac{\partial}{\partial \theta} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + \frac{1}{2} \left(\theta_1^2 - 2\theta\theta_1 \right) \frac{\partial^2}{\partial \theta^2} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu)$$

$$+ \theta_1 \left(\hat{\mu}_n - \mu \right) \frac{\partial^2}{\partial \theta \partial \mu} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + R_n^* ((\mathbf{x}_n, \mathbf{y}_n))$$

where $R_n^* ((\mathbf{x}_n, \mathbf{y}_n))$involves the differences of the second order derivatives, and it converges to zero in probability when $|\theta_i - \theta|$, i = 0,1 are sufficiently small and the second derivatives are smooth (see remark 3.7.1 in Govindaraluju, page182). Next expanding $\frac{\partial}{\partial \hat{\theta}_n} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \hat{\theta}_n, \mu) = 0$ about the true (θ, μ)

$$0 = \frac{\partial}{\partial \theta} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + \left(\hat{\theta}_n - \theta \right) \frac{\partial^2}{\partial \theta^2} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu)$$

$$+ \left(\hat{\mu}_n - \mu \right) \frac{\partial^2}{\partial \theta \partial \mu} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) + R_2 \left(\left(\hat{\theta}_n - \theta \right), \left(\hat{\mu}_n - \mu \right) \right)$$

This gives

$$\frac{\partial}{\partial \theta} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) \approx - \left(\hat{\theta}_n - \theta \right) \frac{\partial^2}{\partial \theta^2} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu) - \left(\hat{\mu}_n - \mu \right) \frac{\partial^2}{\partial \theta \partial \mu} \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta, \mu)$$

Substituting the equation above into the expansion of
$\ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta_1, \hat{\mu}_n) - \ln f_n(\mathbf{x}_n, \mathbf{y}_n; \theta_0, \hat{\mu}_n)$ gives Cox's test statistic, which will be denoted by C, and the following Theorem.

Theorem : When $|\theta_i - \theta|$, $i = 0, 1$, is sufficiently small and the second partial derivatives are smooth, then we have

$$C \approx (\theta_1 - \theta_0) I_{\theta\theta} n \left(\hat{\theta}_n - \frac{(\theta_0 + \theta_1)}{2} \right)$$

where $n\hat{\theta}_n$ is asymptotically normal with mean $n\theta$ and variance $nI^{\theta\theta}$, where $I^{\theta\theta} = I_{\theta\theta} - I_{\theta\mu}^2 / I_{\mu\mu}$.

For large n, $n\hat{\theta}_n$ is the sum of i.i.d. random variables $\{Y_i\}$, $i = 1, \ldots, n$, where

$$Y_i - \theta = \frac{I_{\mu\mu} \frac{\partial}{\partial \theta} \ln f(x_i, y_i; \theta, \mu) - I_{\theta\mu} \frac{\partial}{\partial \mu} \ln f(x_i, y_i; \theta, \mu)}{I_{\theta\theta} I_{\mu\mu} - I_{\theta\mu}^2}$$

see Govindaraluju (1981, p. 182-185).

Now, we can use Wald's approximations for the boundary values in terms of error probabilities, see Govindaraluju (1981, p.185). That is

$$\ln B = \frac{I_{\theta\theta}}{I^{\theta\theta}} \ln \frac{\beta}{1-\alpha} \approx C_n \ln \frac{\beta}{1-\alpha}$$

$$\ln A = \frac{I_{\theta\theta}}{I^{\theta\theta}} \ln \frac{1-\beta}{\alpha} \approx C_n \ln \frac{1-\beta}{\alpha}$$

and

$$C_n = \frac{I_{\hat{\theta}_n \hat{\theta}_n}}{I^{\hat{\theta}_n \hat{\theta}_n}} = \left(1 - \frac{I^2_{\hat{\theta}_n \hat{\mu}_n}}{I_{\hat{\theta}_n \hat{\theta}_n} I_{\hat{\mu}_n \hat{\mu}_n}}\right)^{-1}$$

Or equivalently would be to use the test statistic

$$C' = (\theta_1 - \theta_0)\, I^{\theta\theta} n \left(\hat{\theta}_n - \frac{(\theta_0 + \theta_1)}{2}\right)$$

and the limits

$$\ln B' \approx \ln \frac{\beta}{1-\alpha}$$

$$\ln A' \approx \ln \frac{1-\beta}{\alpha}$$

Rigorous asymptotic treatment can be found in Holm (1985).

7. Cox's SPRT in the RPW case

To derive Cox's SPRT first the test statistic

$$C = (\theta_1 - \theta_0)\, n I_{\theta\theta} \left(\hat{\theta}_n - \frac{(\theta_0 + \theta_1)}{2}\right)$$

need to be derived, which includes to estimate $I_{\theta\theta}$ with the sample, by substituting θ with the maximum likelihood-estimate. Note that Cox assumed the random variables to be independent and identically distributed, which gives that $I_{\theta\theta}$ is the same for all random variables in the sample. In the RPW case the random variables are dependent and have different probability distributions, call them $f_{(i)}(x_i, y_i; \theta, \mu)$. Note that $f_n(x_n, y_n; \theta, \mu)$ is the joint probability function. Hence, let

$$I_{\theta\theta(i)} = Var(\frac{\partial}{\partial\theta} \ln f_{(i)}(x_i, y_i; \theta, \mu)) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f_{(i)}(x_i, y_i; \theta, \mu)\right]$$

$$I_{\theta\mu(i)} = Cov(\frac{\partial}{\partial\theta} \ln f_{(i)}(x_i, y_i; \theta, \mu), \frac{\partial}{\partial\mu} \ln f_{(i)}(x_i, y_i; \theta, \mu)) = -E\left[\frac{\partial^2}{\partial\theta\partial\mu} \ln f_{(i)}(x_i, y_i; \theta, \mu)\right]$$

and let

$$I_{\theta\theta n} = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f_n(x_n, y_n; \theta, \mu)\right]$$

$$I_{\theta\mu n} = -E\left[\frac{\partial^2}{\partial\theta\partial\mu}\ln f_n(x_n, y_n; \theta, \mu)\right]$$

Therefore the test statistic, in the RPW case, will be defined as

$$C = (\theta_1 - \theta_0)\, I_{\theta\theta n}\left(\hat{\theta}_n - \frac{(\theta_0 + \theta_1)}{2}\right)$$

The limits for continuing sampling needs to be derived. These will be defined as

$$\ln B \approx \frac{\ln\frac{\beta}{1-\alpha}}{\left(1 - \frac{I^2_{\theta\mu n}}{I_{\theta\theta n} I_{\mu\mu n}}\right)}$$

$$\ln A \approx \frac{\ln\frac{1-\beta}{\alpha}}{\left(1 - \frac{I^2_{\theta\mu n}}{I_{\theta\theta n} I_{\mu\mu n}}\right)}$$

by the same reason as above.

For simplifying the coming calculations the following rewriting of the expression will be used. Let

$$C'_{(n)} = (\theta_1 - \theta_0)\, I_n^{\theta\theta}\left(\hat{\theta}_n - \frac{(\theta_0 + \theta_1)}{2}\right)$$

where $I_n^{\theta\theta} = \left(I_{\theta\theta n} I_{\mu\mu nn} - I^2_{\theta\mu n}\right)/I_{\mu\mu n}$.

Then sampling is continued as long as

$$\ln\frac{\beta}{1-\alpha} < (\theta_1 - \theta_0)\frac{I_{\theta\theta n} I_{\mu\mu n} - I^2_{\theta\mu n}}{I_{\mu\mu n}}\left(\hat{\theta}_n - \frac{(\theta_0 + \theta_1)}{2}\right) < \ln\frac{1-\beta}{\alpha}$$

and the limits are

$$b^* = \ln\frac{\beta}{1-\alpha}$$

$$a^* = \ln\frac{1-\beta}{\alpha}$$

7.1. The maximum likelihood estimates

To get the maximum likelihood estimate for (θ, μ)solve the equation system

$$\begin{cases} \frac{\partial}{\partial\mu}\ln f_n(x_n, y_n; \theta, \mu) = 0 \\ \frac{\partial}{\partial\theta}\ln f_n(x_n, y_n; \theta, \mu) = 0 \end{cases}$$

where

$$\ln f_n(x_n, y_n; \theta, \mu) = S_{A(n)} \ln \frac{1}{1+e^{-\mu-\frac{\theta}{2}}} + F_{A(n)} \ln \frac{1}{1+e^{\mu+\frac{\theta}{2}}} + S_{B(n)} \frac{1}{1+e^{-\mu+\frac{\theta}{2}}} + F_{B(n)} \frac{1}{1+e^{\mu-\frac{\theta}{2}}}$$

$$= -S_{A(n)} \ln \left(1+e^{-\mu-\frac{\theta}{2}}\right) - F_{A(n)} \ln \left(1+e^{\mu+\frac{\theta}{2}}\right) - S_{B(n)} \ln \left(1+e^{-\mu+\frac{\theta}{2}}\right) - F_{B(n)} \ln \left(1+e^{\mu-\frac{\theta}{2}}\right)$$

That is to solve

$$\begin{cases} \frac{S_{A(n)}}{1+e^{\mu+\frac{\theta}{2}}} - \frac{F_{A(n)}}{1+e^{-\mu-\frac{\theta}{2}}} + \frac{S_{B(n)}}{1+e^{\mu-\frac{\theta}{2}}} - \frac{F_{B(n)}}{1+e^{-\mu+\frac{\theta}{2}}} = 0 \\ \frac{S_{A(n)}}{2\left(1+e^{\mu+\frac{\theta}{2}}\right)} - \frac{F_{A(n)}}{2\left(1+e^{-\mu-\frac{\theta}{2}}\right)} - \frac{S_{B(n)}}{2\left(1+e^{\mu-\frac{\theta}{2}}\right)} + \frac{F_{B(n)}}{2\left(1+e^{-\mu+\frac{\theta}{2}}\right)} = 0 \end{cases}$$

$$\Longleftrightarrow$$

$$\begin{cases} \frac{S_{A(n)}}{1+e^{\mu+\frac{\theta}{2}}} - \frac{F_{A(n)}}{1+e^{-\mu-\frac{\theta}{2}}} = 0 \\ \frac{S_{B(n)}}{1+e^{\mu-\frac{\theta}{2}}} - \frac{F_{B(n)}}{1+e^{-\mu+\frac{\theta}{2}}} = 0 \end{cases}$$

$$\Longleftrightarrow$$

$$\begin{cases} S_{A(n)} - F_{A(n)} e^{\mu+\frac{\theta}{2}} = 0 \\ S_{B(n)} - F_{B(n)} e^{\mu-\frac{\theta}{2}} = 0 \end{cases}$$

$$\Longleftrightarrow$$

$$\begin{cases} \mu + \frac{\theta}{2} = \ln \frac{S_{A(n)}}{F_{A(n)}} \\ \mu - \frac{\theta}{2} = \ln \frac{S_{B(n)}}{F_{B(n)}} \end{cases}$$

which yields

$$\hat{\theta}_n = \ln \frac{S_{A(n)}}{F_{A(n)}} - \ln \frac{S_{B(n)}}{F_{B(n)}}$$

$$\hat{\mu}_n = \frac{\ln \frac{S_{A(n)}}{F_{A(n)}} + \ln \frac{S_{B(n)}}{F_{B(n)}}}{2}$$

Hence

$$\hat{p}_{A(n)} = \frac{S_{A(n)}}{N_{A(n)}}$$

$$\hat{p}_{B(n)} = \frac{S_{B(n)}}{N_{B(n)}}$$

7.2. The test statistic

From Section 7.1 we have that the test statistic, in the RPW case, is as follows.

$$C'_{(n)} = (\theta_1 - \theta_0) + \frac{I_{\theta\theta n} I_{\mu\mu n} - I^2_{\theta\mu n}}{I_{\mu\mu n}} \left(\ln \frac{S_{A(n)} F_{B(n)}}{F_{A(n)} S_{B(n)}} - \frac{(\theta_0 + \theta_1)}{2} \right)$$

The test statistic involves the expression

$$I_{\theta\theta n} - \frac{I^2_{\theta\mu n}}{I_{\mu\mu n}}$$

Below the identities will be derived and expressed both in terms of (θ, μ) and (p_A, p_B). Then approximations of the expectations of $S_{A(n)}$, $F_{A(n)}$, $S_{B(n)}$ and $F_{B(n)}$ will be presented and expressed in terms of p_A and p_B. These approximations will then be used to approximate $I_{\theta\theta n} - \frac{I^2_{\theta\mu n}}{I_{\mu\mu n}}$.

$$I_{\theta\theta n} = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f_n(x_n, y_n; \theta, \mu) \right]$$

$$= E\left[\frac{\partial}{\partial\theta} \left(-\frac{S_{A(n)}}{2\left(1 + e^{\mu + \frac{\theta}{2}}\right)} + \frac{F_{A(n)}}{2\left(1 + e^{-\mu - \frac{\theta}{2}}\right)} + \frac{S_{B(n)}}{2\left(1 + e^{\mu - \frac{\theta}{2}}\right)} - \frac{F_{B(n)}}{2\left(1 + e^{-\mu + \frac{\theta}{2}}\right)} \right) \right]$$

$$= E\left[\frac{S_{A(n)} e^{\mu + \frac{\theta}{2}}}{4\left(1 + e^{\mu + \frac{\theta}{2}}\right)^2} + \frac{F_{A(n)} e^{-\mu - \frac{\theta}{2}}}{4\left(1 + e^{-\mu - \frac{\theta}{2}}\right)^2} + \frac{S_{B(n)} e^{\mu - \frac{\theta}{2}}}{4\left(1 + e^{\mu - \frac{\theta}{2}}\right)^2} + \frac{F_{B(n)} e^{-\mu + \frac{\theta}{2}}}{4\left(1 + e^{-\mu + \frac{\theta}{2}}\right)^2} \right]$$

$$= \frac{1}{4} \left(\frac{e^{\mu + \frac{\theta}{2}}}{\left(1 + e^{\mu + \frac{\theta}{2}}\right)^2} E\left[S_{A(n)}\right] + \frac{e^{\mu + \frac{\theta}{2}}}{\left(1 + e^{\mu + \frac{\theta}{2}}\right)^2} E\left[F_{A(n)}\right] \right.$$

$$\left. + \frac{e^{\mu - \frac{\theta}{2}}}{\left(1 + e^{\mu - \frac{\theta}{2}}\right)^2} E\left[S_{B(n)}\right] + \frac{e^{\mu - \frac{\theta}{2}}}{\left(1 + e^{\mu - \frac{\theta}{2}}\right)^2} E\left[F_{B(n)}\right] \right)$$

$$= \frac{1}{4} \left(\frac{e^{\mu + \frac{\theta}{2}}}{\left(1 + e^{\mu + \frac{\theta}{2}}\right)^2} E\left[N_{A(n)}\right] + \frac{e^{\mu - \frac{\theta}{2}}}{\left(1 + e^{\mu - \frac{\theta}{2}}\right)^2} E\left[N_{B(n)}\right] \right)$$

$$\boxed{I_{\theta\theta n} = \frac{1}{4} \left(\frac{e^{\mu + \frac{\theta}{2}}}{\left(1 + e^{\mu + \frac{\theta}{2}}\right)^2} E\left[N_{A(n)}\right] + \frac{e^{\mu - \frac{\theta}{2}}}{\left(1 + e^{\mu - \frac{\theta}{2}}\right)^2} E\left[N_{B(n)}\right] \right)}$$

$$I_{\theta\mu n} = -E\left[\frac{\partial^2}{\partial\theta\partial\mu} \ln f_n(x_n, y_n; \theta, \mu) \right]$$

$$= E\left[\frac{\partial}{\partial\mu} \left(-\frac{S_{A(n)}}{2\left(1 + e^{\mu + \frac{\theta}{2}}\right)} + \frac{F_{A(n)}}{2\left(1 + e^{-\mu - \frac{\theta}{2}}\right)} + \frac{S_{B(n)}}{2\left(1 + e^{\mu - \frac{\theta}{2}}\right)} - \frac{F_{B(n)}}{2\left(1 + e^{-\mu + \frac{\theta}{2}}\right)} \right) \right]$$

$$= E\left[\frac{S_{A(n)}e^{\mu+\frac{\theta}{2}}}{2\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}+\frac{F_{A(n)}e^{-\mu-\frac{\theta}{2}}}{2\left(1+e^{-\mu-\frac{\theta}{2}}\right)^2}-\frac{S_{B(n)}e^{\mu-\frac{\theta}{2}}}{2\left(1+e^{\mu-\frac{\theta}{2}}\right)}-\frac{F_{B(n)}e^{-\mu+\frac{\theta}{2}}}{2\left(1+e^{-\mu+\frac{\theta}{2}}\right)^2}\right]$$

$$= \frac{1}{2}\left(\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[S_{A(n)}\right]+\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[F_{A(n)}\right]\right.$$

$$\left.-\frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[S_{B(n)}\right]-\frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[F_{B(n)}\right]\right)$$

$$= \frac{1}{2}\left(\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[N_{A(n)}\right]-\frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[N_{B(n)}\right]\right)$$

$$\boxed{I_{\theta\mu n}=\frac{1}{2}\left(\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[N_{A(n)}\right]-\frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[N_{B(n)}\right]\right)}$$

$$I_{\mu\mu n} = -E\left[\frac{\partial^2}{\partial\mu^2}\ln f_n(x_n,y_n;\theta,\mu)\right]$$

$$= E\left[\frac{\partial}{\partial\mu}\left(-S_A\frac{e^{-\mu-\frac{\theta}{2}}}{1+e^{-\mu-\frac{\theta}{2}}}+F_A\frac{e^{\mu+\frac{\theta}{2}}}{1+e^{\mu+\frac{\theta}{2}}}-S_B\frac{e^{-\mu+\frac{\theta}{2}}}{1+e^{-\mu+\frac{\theta}{2}}}+F_B\frac{e^{\mu-\frac{\theta}{2}}}{1+e^{\mu-\frac{\theta}{2}}}\right)\right]$$

$$= E\left[\frac{\partial}{\partial\mu}\left(-S_A\frac{1}{1+e^{\mu+\frac{\theta}{2}}}+F_A\frac{1}{1+e^{-\mu-\frac{\theta}{2}}}-S_B\frac{1}{1+e^{\mu-\frac{\theta}{2}}}+F_B\frac{1}{1+e^{-\mu+\frac{\theta}{2}}}\right)\right]$$

$$= E\left[S_A\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}+F_A\frac{e^{-\mu-\frac{\theta}{2}}}{\left(1+e^{-\mu-\frac{\theta}{2}}\right)^2}+S_B\frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}+F_B\frac{e^{-\mu+\frac{\theta}{2}}}{\left(1+e^{-\mu+\frac{\theta}{2}}\right)^2}\right]$$

$$= \frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}\left(E\left[S_A\right]+E\left[F_A\right]\right)+\frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}\left(E\left[S_B\right]+E\left[F_B\right]\right)$$

$$= 4I_{\theta\theta n}$$

$$\boxed{I_{\mu\mu n} = 4I_{\theta\theta n}}$$

Let us now express the identities above in terms of p_A and p_B. First, note that

$$\frac{e^{\mu+\frac{\theta}{2}}}{1+e^{\mu+\frac{\theta}{2}}} = p_A$$
$$\frac{1}{1+e^{\mu+\frac{\theta}{2}}} = 1-p_A$$
$$\frac{e^{\mu-\frac{\theta}{2}}}{1+e^{\mu-\frac{\theta}{2}}} = p_B$$
$$\frac{1}{1+e^{\mu-\frac{\theta}{2}}} = 1-p_B$$

Hence, we have that

$$\boxed{I_{\theta\theta n} = \tfrac{1}{4}\left(p_A\left(1-p_A\right)E\left[N_{A(n)}\right] + p_B\left(1-p_B\right)E\left[N_{B(n)}\right]\right)}$$

$$\boxed{I_{\theta\mu n} = \tfrac{1}{2}\left(p_A\left(1-p_A\right)E\left[N_{A(n)}\right] - p_B\left(1-p_B\right)E\left[N_{B(n)}\right]\right)}$$

Then

$$I_{\theta\theta n} - \frac{I_{\theta\mu n}^2}{I_{\mu\mu n}} = \frac{4I_{\theta\theta n}^2 - I_{\theta\mu n}^2}{4I_{\theta\theta n}}$$

The numerator is

$$4I_{\theta\theta n}^2 - I_{\theta\mu n}^2$$

$$= \frac{1}{4}\left(\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[N_{A(n)}\right] + \frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[N_{B(n)}\right]\right)^2$$

$$-\frac{1}{4}\left(\frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[N_{A(n)}\right] - \frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[N_{B(n)}\right]\right)^2$$

$$\boxed{4I_{\theta\theta n}^2 - I_{\theta\mu n}^2 = \frac{e^{\mu+\frac{\theta}{2}}e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[N_{A(n)}\right]E\left[N_{B(n)}\right]}$$

and the denominator is

$$\boxed{4I_{\theta\theta n} = \frac{e^{\mu+\frac{\theta}{2}}}{\left(1+e^{\mu+\frac{\theta}{2}}\right)^2}E\left[N_{A(n)}\right] + \frac{e^{\mu-\frac{\theta}{2}}}{\left(1+e^{\mu-\frac{\theta}{2}}\right)^2}E\left[N_{B(n)}\right]}$$

Expressed in terms of p_A and p_B the numerator is

$$4I_{\theta\theta n}^2 - I_{\theta\mu n}^2 = \frac{1}{4}\left(p_A\left(1-p_A\right)E\left[N_{A(n)}\right] + p_B\left(1-p_B\right)E\left[N_{B(n)}\right]\right)^2$$

$$-\frac{1}{4}\left(\left(p_A\left(1-p_A\right)E\left[N_{A(n)}\right] - p_B\left(1-p_B\right)E\left[N_{B(n)}\right]\right)\right)^2$$

And the denominator then is

$$4I_{\theta\theta n} = p_A\left(1-p_A\right)E\left[N_{A(n)}\right] + p_B\left(1-p_B\right)E\left[N_{B(n)}\right]$$

7.3. Approximation of the test statistic

Approximate the earlier expression of $I_{\theta\theta n} - \frac{I_{\theta\mu n}^2}{I_{\mu\mu n}}$ by using the below results for $P\left(Y_i=1\right)$ and $P\left(Y_i=0\right)$.

$$\lim_{i\to\infty} P\left(Y_i=1\right) = \frac{1-p_B}{2-p_A-p_B} \text{ almost surely}$$

$$\lim_{i\to\infty} P\left(Y_i=0\right) = \frac{1-p_A}{2-p_A-p_B} \text{ almost surely}$$

These limit results are discussed in Wei (1979).

$$E\left[S_{A(n)}\right] = \sum_{i=1}^{n} E\left[Y_i X_i\right] = \sum_{i=1}^{n} p_A P\left(Y_i=1\right) \approx n\frac{p_A\left(1-p_B\right)}{2-p_A-p_B}$$

$$E\left[F_{A(n)}\right] = \sum_{i=1}^{n} E\left[Y_i\left(1-X_i\right)\right] = \sum_{i=1}^{n}\left(1-p_A\right)P\left(Y_i=1\right) \approx n\frac{\left(1-p_A\right)\left(1-p_B\right)}{2-p_A-p_B}$$

$$\boxed{E\left[N_{A(n)}\right] \approx n\frac{1-p_B}{2-p_A-p_B}}$$

$$E\left[S_{B(n)}\right] = \sum_{i=1}^{n} E\left[\left(1-Y_i\right)X_i\right] = \sum_{i=1}^{n} p_B P\left(Y_i=0\right) \approx n\frac{\left(1-p_A\right)p_B}{2-p_A-p_B}$$

$$E\left[F_{B(n)}\right] = \sum_{i=1}^{n} E\left[\left(1-Y_i\right)\left(1-X_i\right)\right] = \sum_{i=1}^{n}\left(1-p_B\right)P\left(Y_i=0\right)$$

$$\approx n\frac{\left(1-p_A\right)\left(1-p_B\right)}{2-p_A-p_B}$$

$$\boxed{E\left[N_{B(n)}\right] \approx n\frac{1-p_A}{2-p_A-p_B}}$$

Hence

$$\boxed{I_{\theta\theta n} \approx \frac{n}{4}\frac{\left(p_A+p_B\right)\left(1-p_A\right)\left(1-p_B\right)}{2-p_A-p_B}}$$

$$\boxed{I_{\theta\mu n} \approx \frac{n}{2}\frac{\left(p_A-p_B\right)\left(1-p_A\right)\left(1-p_B\right)}{2-p_A-p_B}}$$

Now let us use these approximations, of the identities, in the expressions obtained earlier. That is

$$4I_{\theta\theta n}^2 - I_{\theta\mu n}^2 \approx \frac{1}{4}n^2\left(\frac{p_A(1-p_A)(1-p_B)}{2-p_A-p_B} + \frac{p_B(1-p_B)(1-p_A)}{2-p_A-p_B}\right)^2$$

$$-\frac{1}{4}n^2\left(\frac{p_A(1-p_A)(1-p_B)}{2-p_A-p_B} - \frac{p_B(1-p_B)(1-p_A)}{2-p_A-p_B}\right)^2$$

$$= \frac{n^2}{4}\frac{(1-p_A)^2(1-p_B)^2(p_A+p_B)^2 - (1-p_A)^2(1-p_B)^2(p_A-p_B)^2}{(2-p_A-p_B)^2}$$

$$= \frac{n^2}{4}\frac{(1-p_A)^2(1-p_B)^2\left((p_A+p_B)^2 - (p_A-p_B)^2\right)}{(2-p_A-p_B)^2}$$

And the denominator then is

$$4I_{\theta\theta n} \approx n\frac{p_A(1-p_A)(1-p_B)}{2-p_A-p_B} + n\frac{p_B(1-p_B)(1-p_A)}{2-p_A-p_B}$$

$$= n\frac{(1-p_A)(1-p_B)(p_A+p_B)}{2-p_A-p_B}$$

Hence

$$\frac{4I_{\theta\theta n}^2 - I_{\theta\mu n}^2}{4I_{\theta\theta n}}$$

$$\approx \frac{n^2}{4}\frac{(1-p_A)^2(1-p_B)^2\left((p_A+p_B)^2 - (p_A-p_B)^2\right)}{(2-p_A-p_B)^2}\frac{2-p_A-p_B}{n(1-p_A)(1-p_B)(p_A+p_B)}$$

$$= \frac{n}{4}\frac{(1-p_A)(1-p_B)\left((p_A+p_B)^2 - (p_A-p_B)^2\right)}{(2-p_A-p_B)(p_A+p_B)}$$

Sampling is continued as long as

$$\ln\frac{\beta}{1-\alpha} < (\theta_1-\theta_0)\frac{4I_{\theta\theta n}^2 - I_{\theta\mu n}^2}{4I_{\theta\theta n}}\left(\ln\frac{S_A F_B}{F_A S_B} - \frac{(\theta_0+\theta_1)}{2}\right) < \ln\frac{1-\beta}{\alpha}$$

where

$$\frac{4I_{\theta\theta n}^2 - I_{\theta\mu n}^2}{4I_{\theta\theta n}} \approx \frac{n}{4}\frac{(1-p_A)(1-p_B)\left((p_A+p_B)^2 - (p_A-p_B)^2\right)}{(2-p_A-p_B)(p_A+p_B)}$$

7.3.1. The approximation of the test statistic with the maximum likelihood estimates

Substituting the maximum likelihood estimates into the approximation of the expression $\frac{4I^2_{\theta\theta n}-I^2_{\theta\mu n}}{4I_{\theta\theta n}}$ yields

$$\frac{4I^2_{\theta\theta n}-I^2_{\theta\mu n}}{4I_{\theta\theta n}}$$

$$\approx \frac{n}{4}\frac{\frac{F_{A(n)}}{N_{A(n)}}\frac{F_{B(n)}}{N_{B(n)}}\left(\left(\frac{S_{A(n)}}{N_{A(n)}}+\frac{S_{B(n)}}{N_{B(n)}}\right)^2-\left(\frac{S_{A(n)}}{N_{A(n)}}-\frac{S_{B(n)}}{N_{B(n)}}\right)^2\right)}{\left(\frac{F_{A(n)}}{N_{A(n)}}+\frac{F_{B(n)}}{N_{B(n)}}\right)\left(\frac{S_{A(n)}}{N_{A(n)}}+\frac{S_{B(n)}}{N_{B(n)}}\right)}$$

This is the expression that is actually used in the test statistic, when the simulations are run.

8. Properties of Cox's SPRT in the RPW case

It would be satisfying if Cox's SPRT worked as well in the RPW-case as we earlier showed that Wald's SPRT did. Unfortunately this is harder to show strictly mathematically in this case.

Cox's test is based on

$$T_n = n\left(\hat{\theta}_n - \frac{1}{2}(\theta_1+\theta_0)\right)$$

Under the assumption of independent identically distributed observations the process T_n is a random walk with independent increments of mean $\theta - \frac{1}{2}(\theta_1+\theta_0)$ and variance $I_{\mu\mu}/(I_{\theta\theta}I\mu\mu) - I^2_{\theta\mu}$. In the RPW case we do not have a constant variance since it depends on which n we have reached. The problem, in this case, is to know how closely related T_n is to a random walk with independent increamants and constant variance. How rough is the assumption of asymptotic normality is as an approximation (the adjustment of the test limits is based on this assumption)?

Figure 8.1 shows three realizations (when $\theta = 1$, $\mu = \ln(3/2)$) of the Cox's statistic under some restrictions on the simulations, described in Section 8.2.1. Briefly described these restrictions are that we take some minimum number of observations before we allow the experiment to stop and we put an upper limit to the sample size. The figure shows a case when there is a treatment difference. The first illustrates an experiment of expected length. The second illustrates an "unexpectedly" long experiment and the third illustrates an "unexpectedly" short experiment. Note that the longest experiment starts at a negative value on the test statistic.

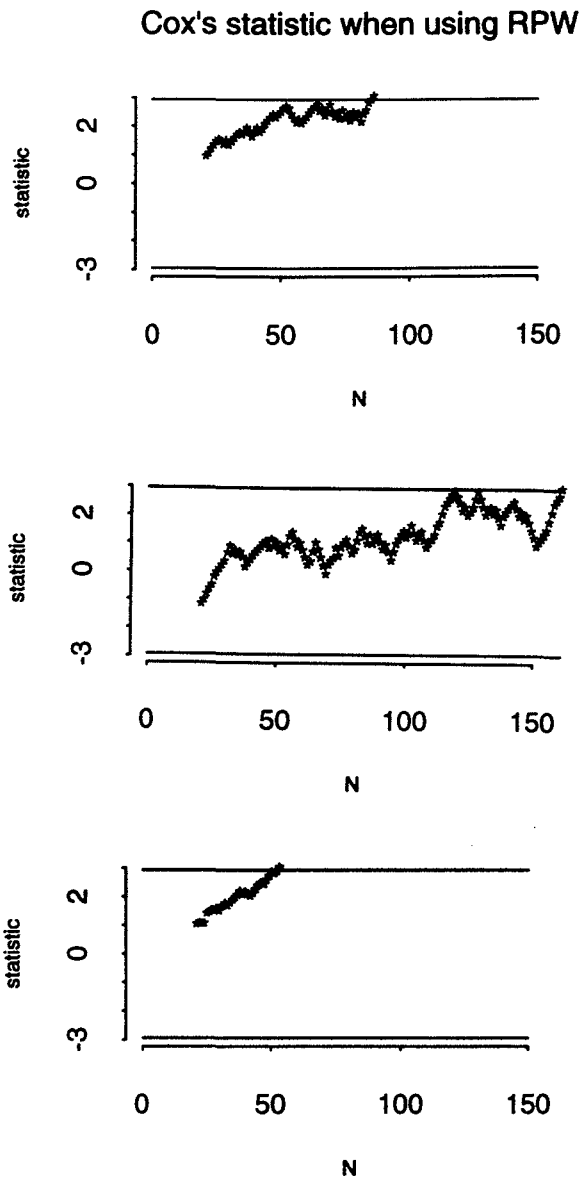

FIGURE 8.1

8.1. A heuristic discussion of properties

As mentioned in Section 6.2, $n\widehat{\theta}_n$ is, for large n, approximately a sum of independent identically distributed random variables. To prove this five equations are used, see Govindaraluju (1981, p. 182-185). Two of the equations are Taylor expansions, but the other three are a collection of limit results, namely

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbf{x}_n,\mathbf{y}_n;\theta,\mu) \underset{n\to\infty}{\longrightarrow} -I_{\theta\theta} \text{ in probability}$$

$$\frac{1}{n}\frac{\partial^2}{\partial\theta\mu}\ln f_n(\mathbf{x}_n,\mathbf{y}_n;\theta,\mu) \underset{n\to\infty}{\longrightarrow} -I_{\theta\mu} \text{ in probability}$$

$$\frac{1}{n}\frac{\partial^2}{\partial\mu^2}\ln f_n(\mathbf{x}_n,\mathbf{y}_n;\theta,\mu) \underset{n\to\infty}{\longrightarrow} -I_{\mu\mu} \text{ in probability}$$

In the RPW case there is no immediate equivalence as the informations in the RPW case are not constant for all n. The correspondence for $I_{\theta\theta}$ is the one dimensional $I_{\theta\theta(n)}$, equivalently for $I_{\theta\mu}$ and $I_{\mu\mu}$.

We think that the properties are still approximately true since the following is true in the RPW case.

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbf{x}_n,\mathbf{y}_n;\theta,\mu) \underset{n\to\infty}{\longrightarrow} -\frac{1}{4}\frac{(p_A+p_B)(1-p_A)(1-p_B)}{2-p_A-p_B} \text{ almost surely}$$

and

$$-\frac{1}{n}I_{\theta\theta n} \underset{n\to\infty}{\longrightarrow} -\frac{1}{4}\frac{(p_A+p_B)(1-p_A)(1-p_B)}{2-p_A-p_B}$$

Correspondingly for $I_{\theta\mu}$ and $I_{\mu\mu}$.
These limit results hold due to the following argumentation.
In Section 7.2 we saw that

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbf{x}_n,\mathbf{y}_n;\theta,\mu) = -\frac{1}{4}\left(p_A(1-p_A)\frac{N_{A(n)}}{n} + p_B(1-p_B)\frac{N_{B(n)}}{n}\right)$$

From Wei (1979) we have that

$$\frac{N_{A(n)}}{n} \underset{n\to\infty}{\longrightarrow} \frac{1-p_B}{2-p_A-p_B} \text{ almost surely}$$

$$\frac{N_{B(n)}}{n} \underset{n\to\infty}{\longrightarrow} \frac{1-p_A}{2-p_A-p_B} \text{ almost surely}$$

for the RPW$(1,1)$ case.
That implies

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbf{x}_n,\mathbf{y}_n;\theta,\mu) \underset{n\to\infty}{\longrightarrow} -\frac{1}{4}\frac{(p_A+p_B)(1-p_A)(1-p_B)}{2-p_A-p_B} \text{ almost surely}$$

The second limit result follows the argumentation below.
Remember that

$$I_{\theta\theta n} = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f_n(x_n, y_n; \theta, \mu)\right]$$

Note that

$$\left|\frac{N_{A(n)}}{n}\right| \leq 1$$

By the Dominated-Convergence Theorem (see for example Williams 1991, p. 54) the below statement follows.

$$\lim_{n\to\infty} E\left[\frac{N_{A(n)}}{n}\right] = \frac{1-p_B}{2-p_A-p_B}$$

A similar reasoning implies that

$$\lim_{n\to\infty} E\left[\frac{N_{B(n)}}{n}\right] = \frac{1-p_A}{2-p_A-p_B}$$

Therefore

$$-\frac{1}{n}I_{\theta\theta n} \underset{n\to\infty}{\longrightarrow} -\frac{1}{4}\frac{(p_A+p_B)(1-p_A)(1-p_B)}{2-p_A-p_B}$$

Correspondingly with $I_{\theta\mu(n)}$ and $I_{\mu\mu(n)}$.
The proportions of A- and B-balls in the urn converge to

$$\lim_{i\to\infty} P(Y_i = 1) = \frac{1-p_B}{2-p_A-p_B} \text{ almost surely}$$

$$\lim_{i\to\infty} P(Y_i = 0) = \frac{1-p_A}{2-p_A-p_B} \text{ almost surely}$$

see Wei (1979).
As $n \to \infty$ observation of the two kinds A and B tend to be taken in those proportions, which is seen above since $\frac{N_{A(n)}}{n} \underset{n\to\infty}{\longrightarrow} \frac{1-p_B}{2-p_A-p_B}$ almost surely and $\frac{N_{B(n)}}{n} \underset{n\to\infty}{\longrightarrow} \frac{1-p_A}{2-p_A-p_B}$ almost surely.
In Section 8.2 some of the properties of the Cox test in the RPW case will be investigated with help of simulations and these results seems to agree with the heuristic discussion above.

8.2. A simulation study of properties

8.2.1. Description

When performing the simulations we hade to decide on some special cases, namely

$$H_0 : \theta_0 = 0 \text{ vs. } H_1 : \theta_1 = 0.6$$

and

$$H_0 : \theta_0 = 0 \text{ vs. } H_1 : \theta_1 = 1$$

In both cases we let the nuisance parameter μ take the values $\ln\frac{1}{9}$,$\ln\frac{1}{4}$, $\ln\frac{2}{3}$, $\ln\frac{3}{2}$, $\ln 4$ and $\ln 9$. Figure 8.2.1 illustrates the simulated cases in the p_A-p_B-space.

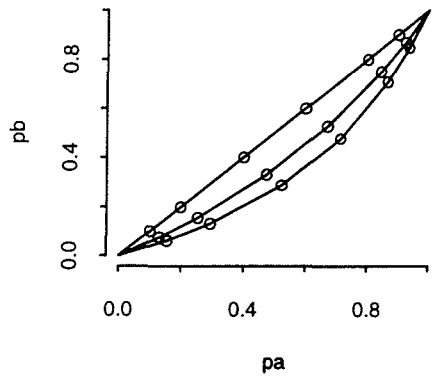

FIGURE 8.2.1

When used in practice, the test statistic $C'_{(n)}$ includes estimation of $\frac{4I^2_{\theta\theta n}-I^2_{\theta\mu n}}{4I_{\theta\theta n}}$. This estimation can be expressed in terms of $\hat{p}_{A(n)}$ and $\hat{p}_{B(n)}$, which are defined as soon as patients have been allocated to both treatments (see Section 7.3.1). This suggests that the experiment should not stop until a fixed number of patients,say n_0, have been allocated to both treatments. The estimation of θ is not defined if one or more of the values $S_{A(n)}$, $F_{A(n)}$, $S_{B(n)}$ or $F_{B(n)}$ is equal to zero. This was solved by substituting each of the values, $S_{A(n)}$, $F_{A(n)}$, $S_{B(n)}$ or $F_{B(n)}$, that were equal to zero with one in $\hat{\theta}_n$.

Hence, we decided to first allocate n_0 patients to each treatment. Denote these first $2n_0$ observations the starting period. During the starting period the allocation were not adaptive. After the starting period the RPW allocation was used, but the first $2n_0$ observations were allowed to affect the urn. We then calculated the values of $\hat{\theta}_n$, $\hat{p}_{A(n)}$ and $\hat{p}_{B(n)}$. Now the sequential procedure started and the use of Cox's statistic.

To select a suitable n_0 we simulated the case when $\theta_1 = 1$ and used a starting period of length 10, $n_0 = 5$, but this number seemed to be too small. Too many experiments (about 10 %) ended at the first possible stage. This behavior can still be observed, with about 1 % of the observations, in the case when $\theta_1 = 0.6$ and $n_0 = 10$. The fact that experiments stop too early leads to a lowered power. In the actual simulations n_0 were set at ten.

An advantage of a starting period of length $n_0 = 10$, is that the parameters are estimated based on 20 observations, at the start of the sequential procedure. A disadvantage is that the RPW allocation will not be used until the 21'st step.

In reality it is often required that the sample size will not exceed a certain fixed number. This cut-of point were 300 in these simulations.

All results in this section are based on 100.000 independently simulated experiments. In the simulations the allocation RPW(1, 1) were used.

8.2.2. Results

We think of studies that are preferred to be of a size of about 150 observations. For an experimenter samples of this sizes could seem large, but one should bear in mind that a smaller sample could be taken if one can accept a lower power. For $\theta_1 = 1$ the sample sizes seems to be reasonable for values of the nuisance parameter, μ, between -1.5 and 1.5, see Figure 8.2.2. For the situation with $\theta_1 = 0.6$ the sample sizes seems to be quite large for all values of the nuisance parameter, μ, investigated, see Figure 8.2.2.

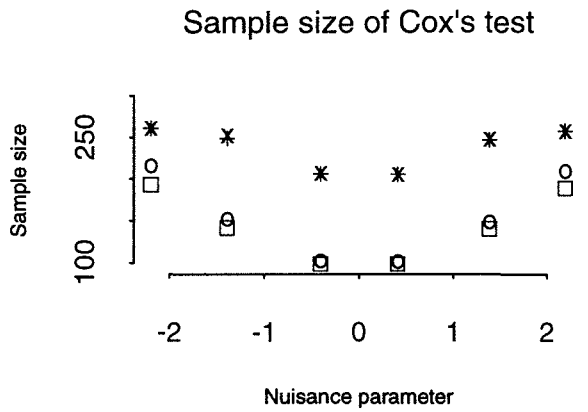

FIGURE 8.2.2 : $\theta_1 = 0.6 \begin{cases} + & H_0 \text{ is true} \\ \text{x} & H_1 \text{ is true} \end{cases}$, $\theta_1 = 1 \begin{cases} \square & H_0 \text{ is true} \\ \text{o} & H_1 \text{ is true} \end{cases}$

For both values on θ_1 the significance level is below the intended level, 0.05, in the area investigated, see Figure 8.2.3.

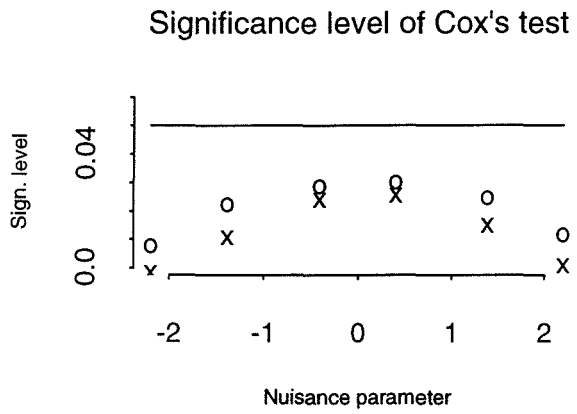

FIGURE 8.2.3: x : $\theta_1 = 0.6$, o : $\theta_1 = 1$.

The power under the alternative is unfortunately below the intended level, 0.95, but for μ around zero it is quite close to the intended level for the case with $\theta_1 = 1$, see Figure 8.2.4. In the case with $\theta_1 = 0.6$ the power is around 0.65 and below, for the investigated values of μ, see Figure 8.2.4.

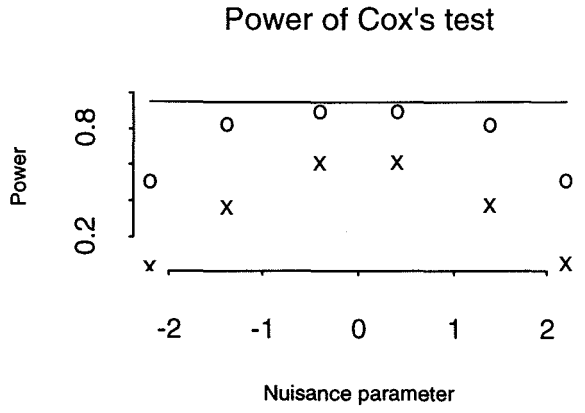

FIGURE 8.2.4 : x : $\theta_1 = 0.6$, o : $\theta_1 = 1$.

A comment on the low power, for $\theta_1 = 0.6$, is that if a larger cut-off point, than $N = 300$, were chosen we would get a better power. The power should also

slightly be raised by choosing a longer starting period as a small proportion of experiments still stops too early. These comments can be summarized by saying that $\theta_1 = 0.6$ requires a larger sample size then $\theta_1 = 1$, to obtain the same power.

Let us observe the number of experiments of different length. Four cases, $\theta_1 = 0.6$ when either $\mu = \ln \frac{3}{2}$ or $\mu = \ln 9$ and $\theta_1 = 1$ and $\mu = \ln \frac{3}{2}$ or $\mu = \ln 9$ are shown in Figure 8.2.5. There it is clear that, when $\theta_1 = 0.6$, a large proportion of the experiments are truncated at $N = 300$. Note that the scales on the y-axes are different for $\theta_1 = 0.6$ and $\theta_1 = 1$.

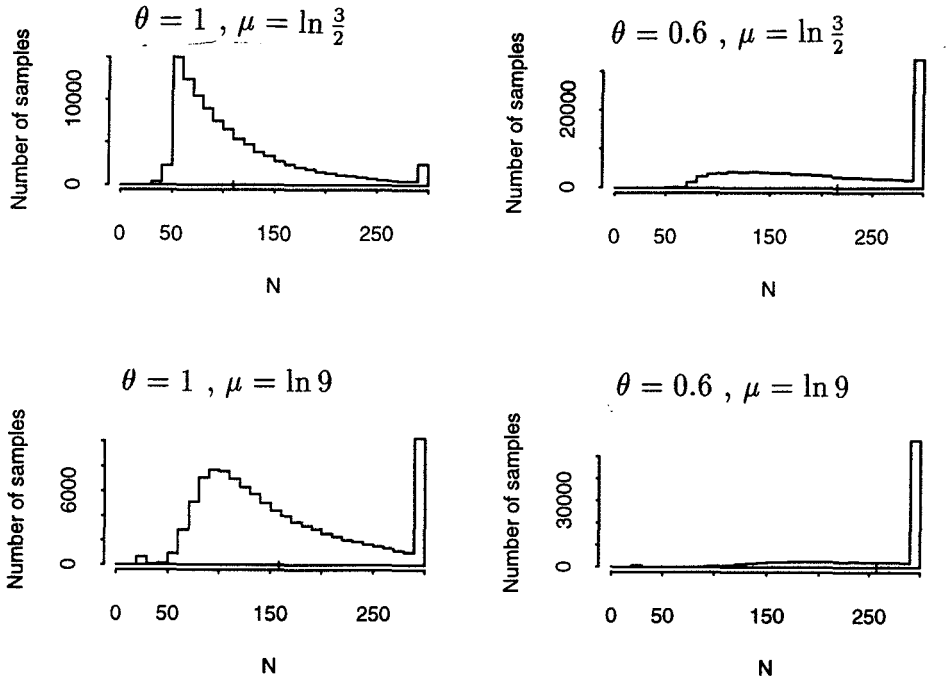

FIGURE 8.2.5

In Figure 8.2.6 we can see that the number of patients assigned to the inferior treatment is fewer in the case of $\theta_1 = 1$ than in the case of $\theta_1 = 0.6$. Note that this is probably due to the smaller sample sizes in the case of $\theta_1 = 1$. When considering the ethical perspective we are both interested in minimizing the actual number of patients receiving the inferior treatment and also to minimize the proportion of patients, in one experiment, receiving the inferior treatment.

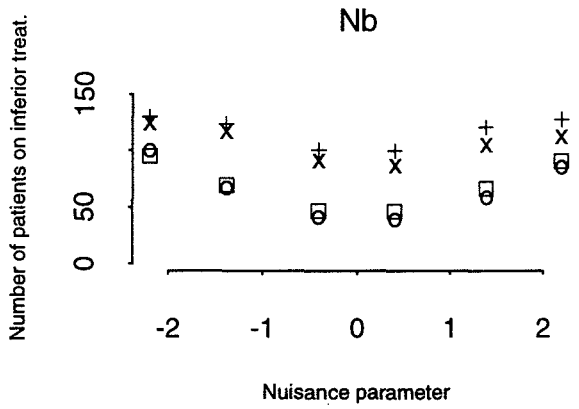

FIGURE 8.2.6 :$\theta_1 = 0.6$ $\begin{cases} + & H_0 \text{ is true} \\ \text{x} & H_1 \text{ is true} \end{cases}$, $\theta_1 = 1$ $\begin{cases} \square & H_0 \text{ is true} \\ \text{o} & H_1 \text{ is true} \end{cases}$

8.2.3. Further results for a special case

It is also of interest to see how the test behaves if the true θ diverge from the value under the null hypothesis and from the value under the alternative. We have chosen to investigate the case where $\theta_1 = 1$ and $\mu = \ln \frac{3}{2} \approx 0.4055$.

Figure 8.2.7 illustrates the power curve for different values of θ, where $\mu = \ln \frac{3}{2}$. When θ is equal to or higher then 0.8 the power is greater then 0.8.

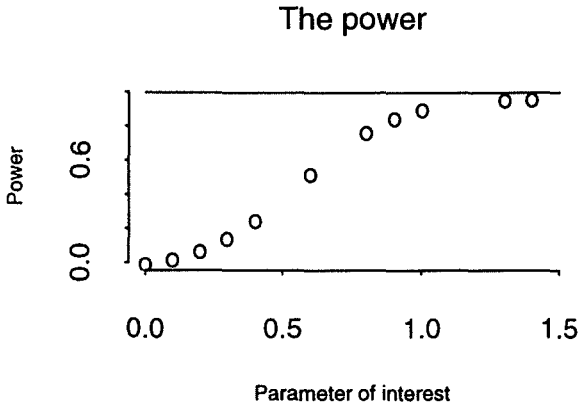

FIGURE 8.2.7

In Figure 8.2.8 below it is illustrated were the chosen values $\mu = \ln \frac{3}{2} \approx 0.4055$ and $0 < \theta < 1.4$ are placed in the p_A -p_B-space.

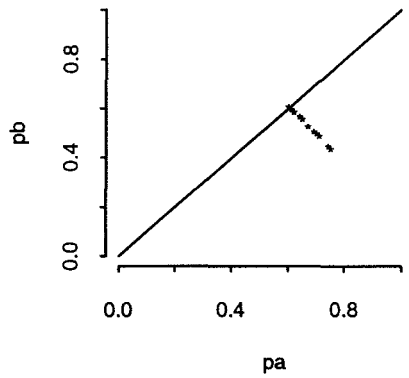

FIGURE 8.2.8 : * symbolizes the simulated points. At the diagonal $\theta = 0$.

In Figure 8.2.9 the sample size and the number of patients on the inferior treatment are shown. It seems to be for values of θ roughly between 0.3 and 1.3 that the number of patients on the inferior treatment is less then half the sample size.

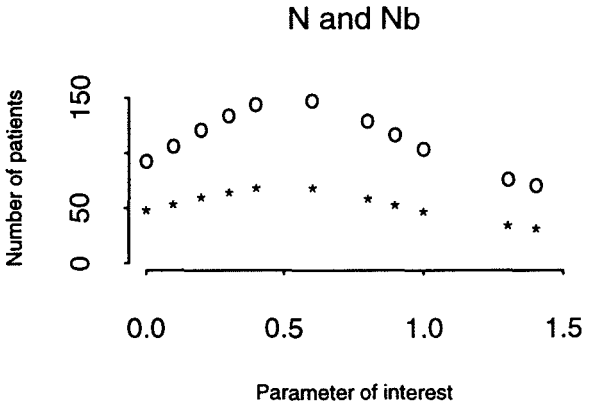

FIGURE 8.2.9 :o represents the sample size, N, and * represents the number of patients on the inferior treatment, N_B.

We would shortly like to comment on the fact that the test is a bit on the conservative side. If the test limits were adjusted, how would this effect the error probabilities and the sample size ?

In a brief investigation, of the cases where $\theta = 1$ and either $\mu = \ln \frac{3}{2}$ or $\mu = \ln 9$, we varied the planned α between 0.05, 0.06 and 0.07. In Figure 8.2.10

we can see how this effected the obtained significance level, α^*, the obtained power, $1-\beta^*$, and the sample size.

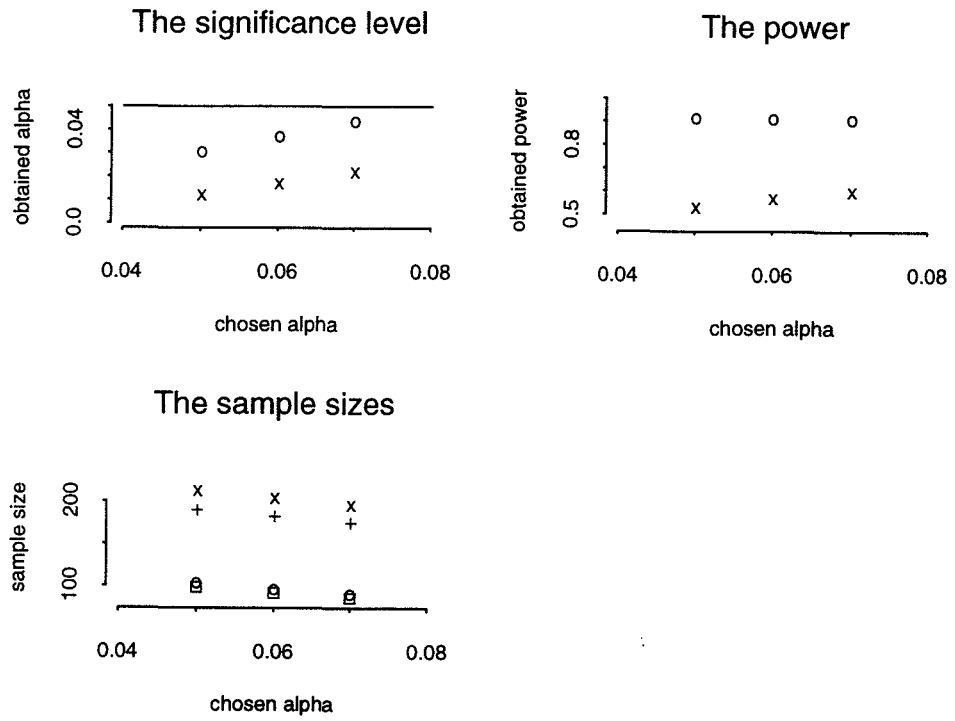

FIGURE 8.2.10 :o (□ when H_1true) when $\theta_1 = 0.6$, **x** (+ when H_1true) when $\theta_1 = 1$.

For $\alpha = 0.07$ as the "intended" level we obtained $\alpha^* \approx 0.046$ as the true level and the sample size decreased from about 99 observations to about 86 observations. This indicates that we could decrease the sample sizes, when Cox's test is used, by working more on the adjustment of the test limits.

8.2.4. Wald's SPRT and Cox's SPRT

To make a brief comparison between Wald's SPRT and Cox's SPRT the two hypothesis cases from Section 5 were simulated when using Cox's SPRT for the analysis. That is when $\theta_1 \approx 0.98$ and $\theta_1 \approx 1.79$.

Figure 8.2.11 summarizes the results.

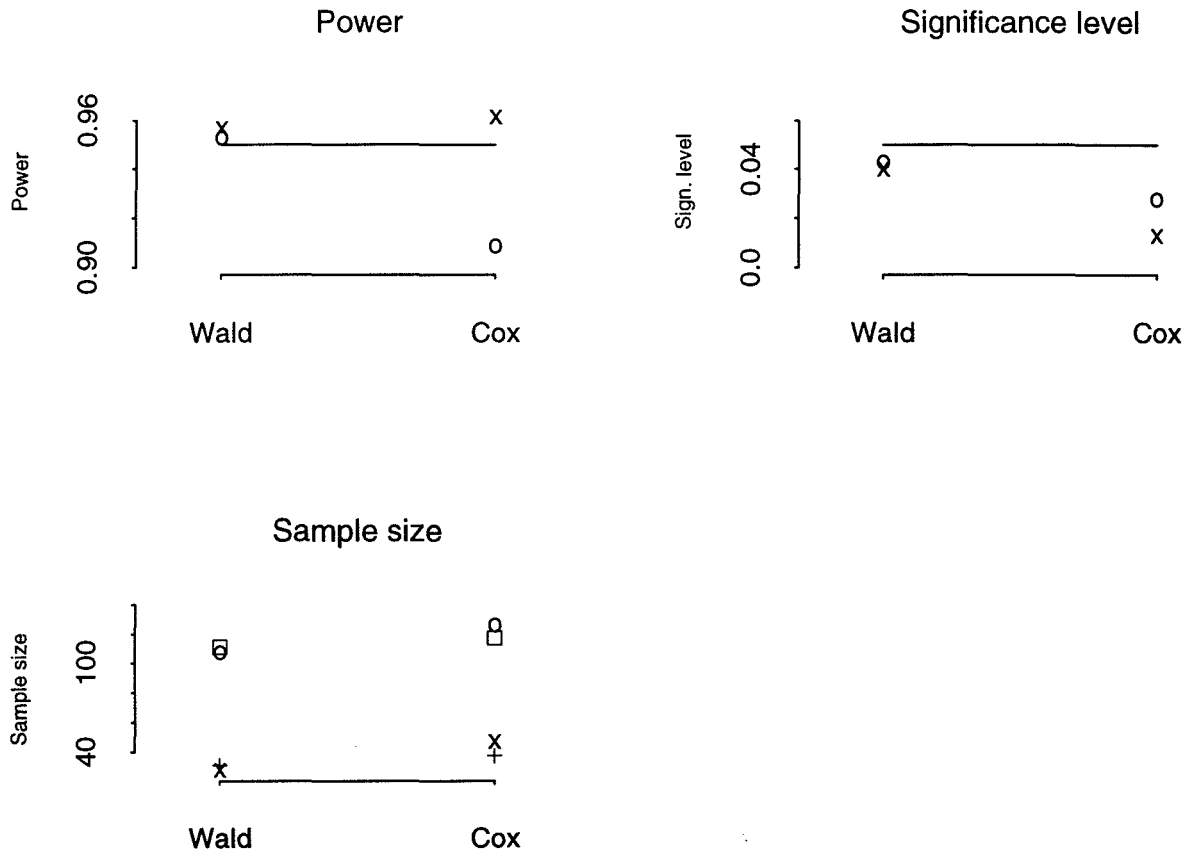

FIGURE 8.2.11 :o (□ when H_0 true) when $\theta_1 = 0.98$ and **x** (+ when H_0 true) when $\theta_1 = 1.79$.

Not surprisingly the loss for Cox's test is regarding the power in the alternative. The test is however constructed for a more complex situation. Note that Cox's test is more conservative then Wald's SPRT, and that is probably the reason why it requires a larger sample.

8.3. Discussion

The power, $1 - \beta$, for different values on μ is more close to the nominal power for the case with $\theta_1 = 1$ than for the case with $\theta_1 = 0.6$. The significance levels were quite close in the two cases and considerably under the planned level. That means that the test is rather conservative, and this makes the sample sizes larger than they need to be. We would have preferred that the test was not this conservative and that the sample sizes had been reduced instead. One way to reduce the sample size could be to choose a higher α then what is actually intended, and use this α in the computations of the test limits. It is, however, necessary to investigate further how much one should depart from the intended α. The most desirably approach would be to learn more about the behavior, for example the convergence speed, of the estimations and approximations used in the test statistic. That is, learn how the process T_n, discussed in Section 8, is

effected by using a RPW. We know that the increments are not independent and that their variances are not constant for all n. From this we should be able to proceed and gain theoretical knowledge which would help us to understand how to adjust the test limits.

The sample sizes are less in the case of $\theta_1 = 1$. This, and the above discussion, indicate that the test, in the present form, is more suitable for testing the null hypothesis against the alternative $\theta_1 = 1$, rather than $\theta_1 = 0.6$. As commented above in Section 8.2.2, the test, if used for $\theta_1 = 0.6$, needs a larger sample size and therefore the starting period and the cut-off point need to be adjusted if used in this situation.

9. Summary

When starting the work with play-the-winner designs we had two main questions.

$\diamondsuit$ Does the theory of known sequential analysis results hold when play-the-winner designs are used ?

$\diamondsuit$ Is it possible to develop results useful in practice ?

The last question is probably very important to work with to make play-the-winner designs easy of access for conceivable users. It is, for example, of importance to adjust the test procedure to practical requirements and to identify the situations when the specific test procedure obtained is suitable.

We had considerations on how the RPW design worked in general, but one has to specify a more concrete problem. We concentrated on the behavior of the error probabilities, the sample size and the number of unfavorable allocations. We started to investigate the RPW design, and to compare it with two other designs, using Wald's sequential probability ratio test, SPRT, for the analysis. Some important properties of Wald's SPRT were proved to hold for a broad class of play-the-winner designs. The comparison, of the three designs, was made to get a brief indication of how the RPW design worked in terms of the above mentioned quantities. We found that it seemed to work well, when used with Wald's SPRT for the analysis. This is the work mentioned in Sections 3, 4 and 5.

After this we wished to proceed to a more, in practice, useful situation, namely when a nuisance parameter is present. Here Cox's SPRT was used for the analysis and we tried, for some special parameter values, to identify when and how it worked, in terms of the error probabilities, the sample size and the number of unfavorable allocations. The results were basically that when $\theta_1 \geq 1$ the larger θ_1 is the smaller the sample size and the shorter the starting period ($2n_0 \leq 20$) can be. The truncation could also be set at a lower value than 300. If $\theta_1 < 1$ the sample size needed to be increased, but for values of θ roughly from 0.8 and larger the test with $\theta_1 = 1$ worked quite well. Though the larger θ we have the larger we should set θ_1 to get minimal sample and still get good properties. What we saw in this work was also that in order to test $\theta_1 = 0.6$ (or smaller θ_1) a sample

size of , for many applications, unreasonable size is needed. This is discussed in Sections 6, 7 and 8.

Note that we have not let the power, $1 - \beta$, vary. This could of course be done, and we expect that the sample size would decrease with decreasing power. The chosen power level of 0.95 is a common choice in statistical literature, but one should reflect on that this is probably a stronger requirement than what is often true when significance tests are used in practice. That is, it is not fair to compare our sample sizes with sample sizes from commonly used tests, without also remembering to compare the power of the tests.

Let us close this summary by comment on the asked questions above. For the first question we have a positive answer for Wald's SPRT in general and heuristically seen for Cox's SPRT. The second question can also be positively answered, but we here only did the real development for some special cases.

ACKNOWLEDGMENT

I would like to thank my supervisor Sture Holm, for supporting and guiding me in this work. I also would like to thank Göran Broström, for sharing his knowledge about FORTRAN. Aila Särkkä and Kasra Afsarinejad have been most helpful with valuable comments on the draft manuscript.

APPENDIX

The tables to follow contain the arithmetic mean and the standard error of the mean.

Simulation results for Wald's SPRT, presented in figures 5.2.1-4.

$H_0 : p_A = p_B = 0.7$ vs. $H_1 : p_A = 0.8$, $p_B = 0.6$ when H_1 true.

	TR	RPW $(100.000, 1)$	RPW $(10, 1)$	RPW $(1, 1)$	RPW $(1, 10)$	RPW $(1, 100.000)$	MPW
N	114.82	114.76	113.53	112.69	112.55	112.42	110.77
$s.d.$	0.12	0.12	0.11	0.11	0.11	0.11	0.11
N_B	57.40	57.39	49.51	44.64	42.97	42.58	38.46
$s.d.$	0.06	0.06	0.05	0.05	0.05	0.05	0.04
$1-\beta$	0.955872	0.955538	0.955976	0.955760	0.956034	0.955692	0.952918
$s.d.$	0.000290	0.000292	0.000290	0.000291	0.000290	0.000291	0.000299

$H_0 : p_A = p_B = 0.7$ vs. $H_1 : p_A = 0.8$, $p_B = 0.6$ when H_0 true.

	TR	RPW $(100.000, 1)$	RPW $(10, 1)$	RPW $(1, 1)$	RPW $(1, 10)$	RPW $(1, 100.000)$	MPW
N	112.37	112.32	111.51	111.10	111.03	111.12	109.41
$s.d.$	0.12	0.12	0.11	0.11	0.11	0.11	0.11
N_B	56.19	56.15	52.97	51.82	51.53	51.52	53.34
$s.d.$	0.06	0.06	0.05	0.05	0.05	0.05	0.04
α	0.045782	0.045936	0.045710	0.045922	0.046012	0.046292	0.047334
$s.d.$	0.000296	0.000296	0.000295	0.000296	0.000296	0.000297	0.000300

$H_0 : p_A = p_B = 0.6$ vs. $H_1 : p_A = 0.8$, $p_B = 0.4$ when H_1 true.

	TR	RPW $(100.000, 1)$	RPW $(10, 1)$	RPW $(1, 1)$	RPW $(1, 10)$	RPW $(1, 100.000)$	MPW
N	33.34	33.33	32.94	32.52	32.30	32.30	31.88
$s.d.$	0.03	0.03	0.03	0.03	0.03	0.03	0.03
N_B	16.66	16.67	14.15	11.42	10.24	10.03	9.12
$s.d.$	0.02	0.01	0.01	0.01	0.01	0.01	0.01
$1-\beta$	0.959464	0.959676	0.959786	0.959530	0.960056	0.959102	0.957672
$s.d.$	0.000279	0.000278	0.000278	0.000279	0.000277	0.000280	0.000284

$H_0 : p_A = p_B = 0.6$ vs. $H_1 : p_A = 0.8$, $p_B = 0.4$ when H_0 true.

	TR	RPW $(100.000, 1)$	RPW $(10, 1)$	RPW $(1, 1)$	RPW $(1, 10)$	RPW $(1, 100.000)$	MPW
N	31.26	31.34	31.08	30.82	30.79	30.85	30.46
$s.d.$	0.03	0.03	0.03	0.03	0.03	0.03	0.03
N_B	15.29	15.66	14.62	13.84	13.68	13.69	14.27
$s.d.$	0.03	0.02	0.01	0.01	0.01	0.01	0.01
α	0.042368	0.042310	0.042254	0.04288	0.042574	0.04342	0.042822
$s.d.$	0.000285	0.000285	0.000285	0.000286	0.000285	0.000288	0.000286

Simulation results for Cox's SPRT, presented in figures 8.2.2, 8.2.3, 8.2.4 and 8.2.6.

θ_0 0

θ_1 0.6

$\theta = 0.6$, that is H_1 true.

μ	$1-\beta$	*s.d.*	N	*s.d.*	N_B	*s.d.*
$\ln\frac{1}{9}$	0.06658	0.000788	269.51	0.26	130.76	0.13
$\ln\frac{1}{4}$	0.40495	0.001552	261.68	0.18	123.79	0.09
$\ln\frac{2}{3}$	0.65027	0.001508	216.66	0.24	98.05	0.12
$\ln\frac{3}{2}$	0.65289	0.001505	214.59	0.24	93.84	0.12
$\ln 4$	0.41601	0.001559	256.99	0.20	111.68	0.11
$\ln 9$	0.09265	0.000917	266.96	0.27	119.39	0.14

H_0 true

μ	α	*s.d.*	N	*s.d.*	N_B	*s.d.*
$\ln\frac{1}{9}$	0.00093	0.000096	260.90	0.28	130.28	0.14
$\ln\frac{1}{4}$	0.01317	0.000361	249.69	0.21	124.03	0.10
$\ln\frac{2}{3}$	0.02643	0.000507	207.11	0.26	101.63	0.12
$\ln\frac{3}{2}$	0.02783	0.000520	206.22	0.26	100.17	0.12
$\ln 4$	0.17410	0.000414	247.18	0.22	120.92	0.11
$\ln 9$	0.00345	0.000185	258.44	0.28	128.26	0.15

θ_0 0

θ_1 1

H_1 true

μ	$1-\beta$	*s.d.*	N	*s.d.*	N_B	*s.d.*
$\ln\frac{1}{9}$	0.54599	0.001574	224.94	0.26	107.20	0.13
$\ln\frac{1}{4}$	0.86498	0.001081	161.70	0.22	74.07	0.11
$\ln\frac{2}{3}$	0.93685	0.000769	111.36	0.18	47.91	0.09
$\ln\frac{3}{2}$	0.93624	0.000773	110.24	0.19	45.60	0.09
$\ln 4$	0.86141	0.001093	157.97	0.23	64.93	0.11
$\ln 9$	0.55244	0.001572	218.46	0.27	92.69	0.14

H_0 true

μ	α	*s.d.*	N	*s.d.*	N_B	*s.d.*
$\ln\frac{1}{9}$	0.01029	0.000319	193.29	0.30	95.96	0.15
$\ln\frac{1}{4}$	0.02481	0.000492	142.65	0.25	70.04	0.12
$\ln\frac{2}{3}$	0.03100	0.000548	99.01	0.20	47.54	0.09
$\ln\frac{3}{2}$	0.03268	0.000562	98.80	0.20	46.64	0.08
$\ln 4$	0.02702	0.000513	140.76	0.25	66.58	0.11
$\ln 9$	0.01406	0.000372	189.74	0.30	91.61	0.15

Simulation results for the case where $\theta_1 = 1$ and $\mu = \ln \frac{3}{2} \approx 0.4055$ (θ is not constant), presented in figures 8.2.7 and 8.2.9.

θ	$1-\beta(\theta)$	*s.d.*	N	*s.d.*	N_B	*s.d.*
0.0	0.03268	0.000562	98.80	0.20	46.64	0.08
0.1	0.05987	0.000750	112.25	0.22	52.12	0.10
0.2	0.10921	0.000986	127.20	0.25	58.04	0.11
0.3	0.18375	0.001225	140.23	0.26	63.12	0.11
0.4	0.28875	0.001433	150.53	0.27	66.85	0.12
0.6	0.55695	0.001571	153.98	0.27	66.73	0.12
0.8	0.80207	0.001260	135.39	0.24	57.41	0.11
0.9	0.88394	0.001013	122.89	0.21	51.48	0.10
1.0	0.93624	0.000773	110.24	0.19	45.60	0.09
1.3	0.99237	0.000275	82.54	0.11	32.76	0.05
1.4	0.99638	0.000190	76.71	0.09	29.99	0.05

Simulation results for different values on α, presented in Figure 8.2.10.

$p_A = p_B$ 0.6
θ_1 1
θ_0 0

$1-\beta$	α	α^*	*s.d.*	N	*s.d.*	N_B	*s.d.*
0.95	0.06	0.03917	0.000613	92.28	0.19	43.50	0.08
0.95	0.07	0.04565	0.000660	86.45	0.18	40.71	0.08

p_A 0.7121
p_B 0.4764
θ_0 0
θ_1 1

$1-\beta$	α	$1-\beta^*$	*s.d.*	N	*s.d.*	N_B	*s.d.*
0.95	0.06	0.93373	0.000787	102.85	0.18	42.82	0.08
0.95	0.07	0.93062	0.000804	96.32	0.17	40.29	0.08

$p_A = p_B$ 0.9
θ_0 0
θ_1 1

$1-\beta$	α	α^*	*s.d.*	N	*s.d.*	N_B	*s.d.*
0.95	0.06	0.01897	0.000431	181.68	0.3	87.56	0.14
0.95	0.07	0.02399	0.000484	174.91	0.3	84.07	0.14

p_A	0.9369
p_B	0.8452
θ_0	0
θ_1	1

$1-\beta$	α	$1-\beta^*$	$s.d.$	N	$s.d.$	N_B	$s.d.$
0.95	0.06	0.59160	0.001554	209.40	0.27	89.19	0.14
0.95	0.07	0.62062	0.001534	201.16	0.27	85.94	0.14

Comparison between Wald's and Cox's SPRT. The results are presented in figure 8.2.11.

p_A	0.8
p_B	0.6
θ_0	0
θ_1	0.9808

$1-\beta$	$s.d.$	N	$s.d.$	N_B	$s.d.$
0.91174	0.000897	130.88	0.21	53.79	0.10
α	$s.d.$	N	$s.d.$	N_B	$s.d.$
0.03081	0.000546	117.09	0.22	55.20	0.10

p_A	0.8
p_B	0.4
θ_0	0
θ_1	1.7918

$1-\beta$	$s.d.$	N	$s.d.$	N_B	$s.d.$
0.96427	0.000587	51.83	0.07	20.09	0.03
α	$s.d.$	N	$s.d.$	N_B	$s.d.$
0.01600	0.000397	37.40	0.07	17.36	0.03

References

BARTLETT, M. S. (1946). The large sample theory of sequential test. *Proc. Cambridge Philos. Soc.* **42**, 239-44

ZELEN, M. (1969).Play-the-winner rule and controlled clinical trial. *J. Am. Statist. Ass.* **64**,131-4

BEGG, B. (1990). On inference from Wei's baised coin design for clinical trials. *Biometrika.* **77**, 467-84

COX, D. R. (1963). Large sample sequential tests for composite hypotheses. *Sankhya Ser. A.* **25**, 5-12

FLOURNOY,N & ROSENBERGER, W. F. (1995). Adaptive design : Selected Proceedings of a 1992 joint AMS-IMS-SIAM summer conference.

GHOSH. (1970). Sequential tests of statistical hypotheses. *Addison-Wesley.*

GOVINDARALUJU. (1981). The sequential statistical analysis of hypothesis testing, point and interval estimation and desicion theory. *American series in mathematical and management sciences.*

HOLM, S. (1985). Sequential likelihood ratio tests. *Sequential methods in statistics Banach center publications.* **16**, 193-232

HOLM, S. (1990). *Unpublished lecture notes. Department of Mathematics, Chalmers University of Technology and Göteborg University.*

LEHMANN, E. L. (1991). Testing statistical hypotheses. *Wadsworth & Brooks/Cole*

ROSENBERGER,W. F. (1996) New directions in adaptive designs. *Statistical Science.* **11**, 137-149

SIEGMUND, D. (1985). Sequential analysis. Tests and confidence intervals. *Springer-Verlag.*

WEI, L.J. (1979). The generalized Polya's urn design for sequential medical trials. *The Annals of Statistics.* **7**, 291

WEI, L.J. (1988). Exact two-sample permutation test based on the RPW. *Biometrika.* **75**, 603-6

WEI, L.J. & DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Am. Statist. Ass.* **73**, 830-43

WEI, L.J., SMYTHE, R.T., LIN, D.Y. & PARK, T.S. (1990). Statistical Inference With Data-Dependent Allocation Rules. *J. A. Statist. Ass.* **85**, 156-62

WILLIAMS, D. (1991). Probability with martingales. *Cambridge University Press.*

Research Report

1995:1	Arnkelsdottir, H	Surveillance of rare events. On evaluations of the sets method.
1995:2	Sveréus, A	Detection of gradual changes. Statistical methods in post marketing surveillance.
1995:3	Ekman, C	On second order surfaces estimation and rotatability.
1996:1	Ekman, A	Sequential analysis of simple hypotheses when using play-the-winner allocation.
1996:2	Wessman, P	Some principles for surveillance adopted for multivariate processes with a common change point.
1996:3	Frisén, M. & Wessman, P	Evaluations of likelihood ratio methods for surveillance.
1996:4	Wessman, P.	Evaluation of univariate surveillance procedures for some multivariate problems.
1996:5	Särkkä, A.	Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes.