Theory and Applications of A New Methodology of Modified Sequential Probability Ratio Test

N. Mukhopadhyay* and Basil M de Silva†

Mathematics and Statistics Research Report 2006/04

July 2006

Department of Mathematics and Statistics

RMIT University

GPO Box 2476V, Melbourne VIC 3001

AUSTRALIA

2000 AMS Classification: Primary 62L10; Secondary 62F03.

ISBN 10 : 1-921166-39-8

ISBN 13 : 978-1-921166-39-6

*Department of Statistics, UBox 4120 CLAS Building, University of Connecticut, 215 Glenbrook Road, Storrs CT 06269-4120 U.S.A., email: nitis.mukhopadhyay@uconn.edu

†email: desilva@rmit.edu.au

Abstract

Abraham Wald's original sequential probability ratio test (SPRT) and its generalizations have been utilized successfully in designing clinical trials, reliability studies, and others during the past sixty plus years. Sequential methodologies are now being used in solving many kinds of contemporary statistical problems including those in data mining, gene sequencing, and detection of computer network intrusions. The manuscript's Introduction includes some specific references.

Wald's SPRT methodology works well when we gather one (or m) observations at-a-time, and then we determine when to stop gathering more observations in order to make a decision. But, we often face practical situations where one may only observe a random number of observations at every step. We have cited a number of practical problems where this will be the scenario. Now, Wald's SPRT has to be reworked from ground zero. Actually, one first needs to examine first what test could serve as a parallel to the customary best fixed-sample-size or UMP test.

We have formulated these ideas clearly and then proposed a parallel of Wald's original SPRT in the present situation. Important characteristics of our modified SPRT (MSPRT) and their approximations are given. Computer simulations are used heavily to check out the roles of those approximations in the performance of MSPRT. Toward the end, we have described two real problems with data analysis .

Key Words and Phrases: Applications, average number of steps, average total sample size, data analysis, most powerful test, number of steps, preset level, preset power, quality control, random group sizes, sequential probability ratio test, simulations, real data.

1. Introduction.

Let $M_1, M_2, \ldots, M_k, \ldots$ be a sequence of non-negative integer valued random variables. For example, a local bank in a university campus may be interested in monitoring the number of predominantly student-customers accessing savings or checking accounts and the amount of withdrawal (X dollars) per hit through an automatic teller machine (ATM) located in the student union. Students are known to be more prone to withdraw funds before "lunch time" but then there is no fixed set time for lunch either. The management fixed a two-hour window starting from 11:00 AM through 1:00 PM on a typical Monday, M_i being the number of hits in the i^{th} fifteen-minute interval and $X_{i1}, X_{i2}, \ldots, X_{iM_i}$ are the associated amounts of withdrawals, $i = 1, \ldots, 8$. Each student stays at the ATM for couple of seconds only and hence from practical considerations, we disregard the minuscule possibility of a customer's arrival just prior to endpoint of one time-interval and crossing over to the next time-interval. One may surely think of more such time-intervals as one may continue collecting data on the following weekdays between 11:00 AM and 1:00 PM as needed. We need to emphasize certain points. First, this kind of data are readily available on a real-time basis to a bank's local manager and it is updated instantaneously as soon as a new customer accesses his/her account through ATM. Secondly, this kind of monitoring is important to estimate the number of customers standing in line before lunch-time waiting to access the ATM. Otherwise how would the management know whether or not there is any pressing need for installing another ATM elsewhere on campus? It is also crucial that the ATM carries "enough" cash in five, ten, and twenty dollar bills so that a requested amount of money can be disbursed hassle-free to each customer. With all likelihood, the ATM must not run out of cash during the lunch-time rush! That is the reason why one should also closely monitor the X variable.

One may also face such a situation especially in a context of monitoring "inventory" and/or "serving queues". It is well-known, for example, that if a car's brake lights do not work properly, then this can be very hazardous. A driver following a car without brake lights will not be forewarned if the car in front suddenly decides to brake. Brake-light violation does cause rear-ending cars as well as serious accidents especially at high speed. The department of motor vehicles (DMV) may publicize this issue with important bulletins on roadways and decide to monitor stopped cars at large intersections. In one such intersection, officials from the department of Transportation (DOT) and DMV may record the number of stopped cars (M_i) and the number of cars ($\Sigma_{j=1}^{M_i} X_{ij}$) *without brake lights* whenever the traffic signal changes from green to red. Here, X_{ij}'s are thought of as independent Bernoulli variables and the data $\Sigma_{j=1}^{M_i} X_{ij}$ is easily recorded since every working brake-light will glow bright red. It may be quite reasonable to assume that M_i's themselves are random variables too! In a busy intersection, it may not make much sense to try and record data on the status of brake lights (X) for car #1, car #2, and so on when a group of cars are stopped at a red light! Yet, it is important to monitor the percentage (θ) of cars without brake lights on roadways. In certain areas, if it is felt that rear-end collisions are on the rise, then the DMV and DOT may be keen on adding appropriate resources, for example, increase surveillance, educate drivers through commercials on TV, issue more warnings or tickets to violators to ultimately help in lowering the value of θ.

In one single pass, unfortunately, we cannot address every single issue raised in the two paragraphs above. But, it is true that the present setup would certainly demand new kinds of formulations of customary sequential testing and estimation problems. In the present article, we only aim at developing the theory and some applications of a new kind of *sequential probability ratio test* (SPRT) by upgrading Wald's (1947) methodologies. For brevity, we refer to this new methodology as a *modified sequential probability ratio test* (MSPRT).

One is referred to Wald (1947), Bechhofer et al. (1968), Ghosh (1970), Govindarajulu (1981,2004), Sen (1981), Siegmund (1985), and Whitehead (1997) for a broad range of review in the area of SPRT and other sequential tests. While Govindarajulu (1981,2004), Sen (1981), Woodroofe (1982), and Siegmund (1985) gave valuable introductions on many important aspects of sequential estimation, Ghosh et al. (1997) could serve as a comprehensive resource.

The wonderful paper of Lai (2004a) followed by a lengthy set of discussions on that piece have provided invaluable links of Wald identities with important areas of statistical research and practice, including clinical trials, financial mathematics, and security breach of computer networks.

For a glimpse on specific applications of sequential methodologies in some of the contemporary areas of statistics, one may also take a look at the following papers: Abraham (2004), Baron (2004), Chang and Martinsek (2004), Lai (2004b), Li and Solanky (2004), Mukhopadhyay et al. (2004), Tartakovsky and Veeravalli (2004), Young (2004), and Zacks and Rogatko (2004). The breadth of the areas of applications included passive acoustic detection of marine mammals, multistate processes, multichannel distributed systems, data mining, clinical trial, target tracking, change-point detection, integrated pest management, horticulture, and gene ordering.

2. Basic structure and the most powerful test. In general, we think of the following independent observations

$$\{M_i, X_{i1}, X_{i2}, \ldots, X_{iM_i}; i = 1, 2, \ldots, k, \ldots\} \tag{2.1}$$

where we also assume that:

(A1) the M_i's constitute a sequence of independent observations with its probability mass function (p.m.f.)

$$g_i(m_i) = \mathcal{P}(M_i = m_i), m_i = 0, 1, 2, \ldots, i = 1, 2, \ldots. \tag{2.2}$$

One immediately notes that we do not rule out the possibility that in some time-intervals, there may be no observations available. That is, no customers may arrive to use the ATM in our specific example.

(A2) the X's constitute a sequence of conditionally independent and identically distributed (i.i.d.) observations, given the M's, with the common probability density function (p.d.f.) or p.m.f. $f(x;\theta)$.

After observing five consecutive time-intervals with $k=5$, for example, we may have observed the number of arriving customers

$$m_1 = 3, m_2 = 6, m_3 = 0, m_4 = 8, \text{ and } m_5 = 4,$$

with the likelihood function

$$\Pi_{i=1}^{5} \Pi_{j=1}^{m_i} f(x_{ij};\theta). \tag{2.3}$$

Here, $\Pi_{j=1}^{m_3} f(x_{3j};\theta)$ is interpreted as one. Since there are no recorded X observations from the third interval, it seems reasonable to assume that the likelihood function should correspond to the product of the p.d.f.'s (or p.m.f.'s) of all available remaining observations!

In general, let us denote $\mathbf{M}^{(k)} = (M_1, \ldots, M_k)$ for the associated data on the number of arriving customers in k consecutive time-intervals where k has been fixed in advance. Then, the likelihood function of the observations $\{M_i, X_{i1}, X_{i2}, \ldots, X_{iM_i};\ i = 1, 2, \ldots, k\}$ is expressed as $\Pi_{i=1}^{k} \Pi_{j=1}^{M_i} f(X_{ij};\theta)$. Now, in order to test a simple null hypothesis $H_0 : \theta = \theta_0$ against a simple alternative hypothesis $H_1 : \theta = \theta_1 (\neq \theta_0)$ based on the data $\{M_i, X_{i1}, X_{i2}, \ldots, X_{iM_i}; i = 1, 2, \ldots, k\}$, the following theorem provides a way to obtain the *most powerful* (MP) level α test.

Theorem 2.1. *Consider the test described by the following critical function*:

$$\psi(X_{11}, \ldots, X_{1M_1}, \ldots, X_{k1}, \ldots, X_{kM_k}) = \begin{cases} 1 & \text{if } \dfrac{\Pi_{i=1}^{k} \Pi_{j=1}^{M_i} f(X_{ij};\theta_1)}{\Pi_{i=1}^{k} \Pi_{j=1}^{M_i} f(X_{ij};\theta_0)} > c \\ 0 & \text{if } \dfrac{\Pi_{i=1}^{k} \Pi_{j=1}^{M_i} f(X_{ij};\theta_1)}{\Pi_{i=1}^{k} \Pi_{j=1}^{M_i} f(X_{ij};\theta_0)} < c \end{cases} \tag{2.4}$$

where the cut-off point $c \equiv c_{\mathbf{M}^{(k)}}$ *is chosen in such a way that*

$$\mathbf{E}_{\theta_0}\left[\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k})\right] = \alpha. \tag{2.5}$$

Then, under the assumptions (A1)-(A2), the associated test is level α *most powerful for deciding between the two hypothesis* H_0 *and* H_1.

This theorem is proved in Appendix A and it shows that the traditional Neyman-Pearson Lemma (see, for example, Lehmann (Section 3.2, 1986)) remains applicable in constructing the MP level α test even when the *total sample size* (TSS) $M_{(k)}(=\Sigma_{i=1}^{k} M_i)$ is a random variable when the assumptions (A1)-(A2) hold. Observe that in the case when the X's have a continuous p.d.f., the test function given by (2.4) may be rewritten as follows:

$$\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k}) = 0 \text{ or } 1 \text{ according as}$$
$$\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_1)/\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_0) \geq \text{ or } < c \text{ respectively.}$$

But, when the X's have a discrete p.m.f., the test function given by (2.4) may be rewritten as follows:

$$\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k}) = 0 \text{ or } \gamma \text{ or } 1 \text{ according as}$$
$$\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_1)/\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_0) > \text{ or } = \text{ or } < c \text{ respectively.}$$

Here, $c > 0$ and $0 < \gamma < 1$, as needed, are to be determined in such a way that

$$\mathbf{E}_{\theta_0}\left[\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k})\right] = \alpha.$$

This clearly extends the reign of the traditional optimality property associated with the MP fixed-sample-size level α test. We may denote the power function associated with the MP test (2.4)-(2.5):

$$\Delta(\theta_1) \equiv \mathbf{E}_{\theta_1}\left[\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k})\right]. \tag{2.6}$$

Example 2.1. Suppose that $f(x;\theta)$ belongs to a one-parameter exponential family, namely

$$f(x;\theta) = q(\theta)p(x)\exp\{\theta R(x)\}, \text{ for } x \in \mathcal{X} \text{ and } \theta \in \Theta, \tag{2.7}$$

and we want to test a simple null hypothesis $H_0 : \theta = \theta_0$ against a simple alternative hypothesis $H_1 : \theta = \theta_1 (> \theta_0)$. Now, with $a(\theta) = \{q(\theta)\}^{M_{(k)}}$ and $b(\mathbf{d}) = \Pi_{i=1}^{k} \Pi_{j=1}^{M_i} p(X_{ij})$, the likelihood function of the observations $\mathbf{d} = \{M_i, X_{i1}, X_{i2}, \ldots, X_{iM_i}; i = 1, 2, \ldots, k\}$ can be expressed as

$$\begin{gathered} \Pi_{i=1}^{k} \Pi_{j=1}^{M_i} f(X_{ij}; \theta) = a(\theta) b(\mathbf{d}) \exp\{\theta \textstyle\sum_{i=1}^{k} \sum_{j=1}^{M_i} R(X_{ij})\} \\ \text{with } M_{(k)} = \textstyle\sum_{i=1}^{k} M_i, \end{gathered} \tag{2.8}$$

so that the MP level α test function described in Theorem 2.1 can be alternatively written as

$$\psi(X_{11}, \ldots, X_{1M_1}, \ldots, X_{k1}, \ldots, X_{kM_k}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{k} \sum_{j=1}^{M_i} R(X_{ij}) > c \\ 0 & \text{if } \sum_{i=1}^{k} \sum_{j=1}^{M_i} R(X_{ij}) < c \end{cases} \tag{2.9}$$

where the cut-off point $c \equiv c_{\mathbf{M}^{(k)}}$ would be chosen in such a way that

$$\mathbf{E}_{\theta_0} \left[\psi(X_{11}, \ldots, X_{1M_1}, \ldots, X_{k1}, \ldots, X_{kM_k}) \right] = \alpha.$$

Example 2.2. As a special case, let us suppose that $f(x; \theta)$ is a normal p.d.f. so that

$$f(x; \theta) = \{2\pi\sigma^2\}^{-1/2} \exp\{-\tfrac{1}{2}(x - \theta)^2/\sigma^2\} \text{ for } -\infty < \theta, x < \infty \tag{2.10}$$

with $0 < \sigma^2 < \infty$ known, and we want to test a simple null hypothesis $H_0 : \theta = \theta_0$ against a simple alternative hypothesis $H_1 : \theta = \theta_1 (> \theta_0)$. Now, this p.d.f. can be easily rewritten as that in (2.7) where one obviously has

$$q(\theta) = \{2\pi\sigma^2\}^{-1/2} \exp\{-\tfrac{1}{2}\theta^2/\sigma^2\}, p(x) = \exp\{-\tfrac{1}{2}x^2/\sigma^2\}, \text{ and } R(x) = x/\sigma^2.$$

In this situation, the MP level α test (2.9) simplifies further to the following form:

$$\psi(X_{11}, \ldots, X_{1M_1}, \ldots, X_{k1}, \ldots, X_{kM_k}) = \begin{cases} 1 & \text{if } M_{(k)}^{1/2} \left\{ \overline{X}_{\mathbf{M}^{(k)}} - \theta_0 \right\} \geq \sigma z_\alpha \\ 0 & \text{if } M_{(k)}^{1/2} \left\{ \overline{X}_{\mathbf{M}^{(k)}} - \theta_0 \right\} < \sigma z_\alpha \end{cases} \tag{2.11}$$

where $M_{(k)} = \sum_{i=1}^{k} M_i, \overline{X}_{\mathbf{M}^{(k)}} = M_{(k)}^{-1} \sum_{i=1}^{k} \sum_{j=1}^{M_i} X_{ij}$, and z_α is the upper $100\alpha\%$ point of the standard normal distribution. This follows easily since the random variable

$$M_{(k)}^{1/2} \left\{ \overline{X}_{\mathbf{M}^{(k)}} - \theta \right\} \text{ is distributed as } N(0, \sigma^2) \text{ when } \theta \text{ obtains.} \tag{2.12}$$

2.1. *The MP test with preassigned power.* Even though we have treated k as fixed, the MP level α test given by (2.4)-(2.5) cannot be called a fixed-sample-size best test in the usual sense. Yet, we may still set out to determine the minimum value of k, the fixed number of steps or groups, such that the associated power $\Delta(\theta_1)$ defined by (2.6) is at least $1 - \beta$ where $0 < \beta < 1$ is preassigned and "small".

Example 2.2 (Continued). The Type II error probability associated with the test (2.11) can be expressed as

$$\begin{aligned}
&\mathcal{P} \left\{ \text{Accept } H_0 \text{ when } H_1 \text{ is true} \right\} \\
&= \mathcal{P}_{\theta_1} \left\{ M_{(k)}^{1/2} \left\{ \overline{X}_{\mathbf{M}^{(k)}} - \theta_0 \right\} < \sigma z_\alpha \right\} \\
&= \mathcal{P}_{\theta_1} \left\{ M_{(k)}^{1/2} \left\{ \overline{X}_{\mathbf{M}^{(k)}} - \theta_1 \right\} \sigma^{-1} < z_\alpha - M_{(k)}^{1/2} (\theta_1 - \theta_0) \sigma^{-1} \right\} \\
&= \mathbf{E}_{\theta_1} \left[\Phi \left(z_\alpha - M_{(k)}^{1/2} (\theta_1 - \theta_0) \sigma^{-1} \right) \right].
\end{aligned} \tag{2.13}$$

Now, then the basic problem amounts to this:

$$\begin{gathered}
\text{We must determine minimum } k \equiv k_{\min} \text{ such that} \\
\mathbf{E}_{\theta_1} \left[\Phi \left(z_\alpha - M_{(k)}^{1/2} (\theta_1 - \theta_0) \sigma^{-1} \right) \right] \leq \beta.
\end{gathered} \tag{2.14}$$

At this point, we do realize that we can only hope to solve (2.14) numerically on a case by case basis. The nature of the associated solution is bound to depend upon how the process generates the random variables $M_1, \ldots, M_k$. In what follows, we especially consider three particular circumstances in the context of Example 2.2.

Example 2.3. Suppose that M_i is distributed as binomial with parameters r_i and p ($\text{Bin}(r_i, p)$) where r_i, p are known, $0 < p < 1, i = 1, \ldots, k$. Obviously then

$M_{(k)}$ is distributed as $\text{Bin}(r, p)$ with $r = \sum_{i=1}^{k} r_i$. In this situation, (2.14) can be equivalently written as follows:

$$\begin{gathered}\text{We must determine minimum } k \equiv k_{\min} \text{ such that} \\ \sum_{y=0}^{r} \Phi\left(z_\alpha - y^{1/2}(\theta_1 - \theta_0)\sigma^{-1}\right) \binom{r}{y} p^y (1-p)^{r-y} \leq \beta.\end{gathered} \tag{2.15}$$

In Section 2.2, we consider r_i's equal.

Example 2.4. Suppose that M_i is distributed as Poisson with mean λ_i ($\text{Poi}(\lambda_i)$) where λ_i is known, $0 < \lambda_i < \infty, i = 1, \ldots, k$. Obviously then $M_{(k)}$ is distributed as $\text{Poi}(\lambda)$ where $\lambda = \sum_{i=1}^{k} \lambda_i$. In this situation, (2.14) can be equivalently written as follows:

$$\begin{gathered}\text{We must determine minimum } k \equiv k_{\min} \text{ such that} \\ \sum_{y=0}^{\infty} \Phi\left(z_\alpha - y^{1/2}(\theta_1 - \theta_0)\sigma^{-1}\right) \exp(-\lambda)\lambda^y (y!)^{-1} \leq \beta.\end{gathered} \tag{2.16}$$

In Section 2.2, we consider λ_i's equal.

Example 2.5. Suppose that M_i's are assumed i.i.d. Geometric ($\text{Geo}(p)$) with the p.m.f. $(1-p)^{y-1}p, y = 1, 2, \ldots$ where p is known, $0 < p < 1, i = 1, \ldots, k$. Obviously then $M_{(k)}$ has a Negative Binomial distribution. In this situation, (2.14) can be equivalently written as follows:

$$\begin{gathered}\text{We must determine minimum } k \equiv k_{\min} \text{ such that} \\ \sum_{y=k}^{\infty} \Phi\left(z_\alpha - y^{1/2}(\theta_1 - \theta_0)\sigma^{-1}\right) \binom{y-1}{k-1} p^k (1-p)^{y-k} \leq \beta.\end{gathered} \tag{2.17}$$

Some tables of such $k_{\min}$ values satisfying (2.15)-(2.17) are provided for a few selected sets of values of r, λ and p, as the case may be, for immediate reference. One may visualize other analogous examples for non-normal X's too. In Section 3, Wald's (1947) classical *sequential probability ratio test* (SPRT) is suitably modified in the present context and its theory is developed. Some verifications of the stated results are deferred to an Appendix. We also compare the newly proposed sequential methodology with the MP test based on the data $\{X_{i1}, \ldots, X_{iM_i}\}, i = 1, 2, \ldots, k_{\min}$

for a number of distributions, including the normal example, via computer simulations (Section 4). Some real applications are presented in Section 5.

2.2 *Performances of the MP test with preassigned power: Normal mean.* Consider the MP test given by (2.11) in the case of testing a normal mean that was described by Example 2.2. We wish to test, for example, a null hypothesis $H_0 : \theta = 0$ against an alternative hypothesis $H_1 : \theta = 1$ with $\alpha = \beta = 0.05$ where θ is the mean of a normal population with standard deviation 5. We intend to revisit the Examples 2.3-2.5 with data X_{ij}'s that are generated from $N(\theta, 25)$. Here, the aim is to determine the minimum value of k, that is, $k_{\min}$ successively for the three distributions of M_i's given in those examples so that the power of the MP test (2.11) is at least $1 - \beta$.

Table 2.1. $k_{\min}$ **for the Example 2.3:** M_i**'s Are I.I.D. Bin**(r_0, p)

	p								
r_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	544	272	182	136	109	91	78	68	61
10	272	136	91	68	55	46	39	34	31
20	136	68	46	34	28	23	20	17	16
40	68	34	23	17	14	12	10	9	8

First, we let M_i's be i.i.d. $\text{Bin}(r_0, p)$ as in Example 2.3. Then, $k_{\min}$ is given by (2.15) where $r = kr_0, z_{0.05} = 1.645$ and $\sigma^{-1} = 0.2$. By numerically solving the inequality in (2.15) we obtain $k_{\min}$ associated with the preassigned power at least 0.95. Table 2.1 gives the computed values of $k_{\min}$ for $r_0 = 5, 10, 20, 40$ and $p = 0.1, 0.2, \ldots, 0.9$. Since $E(M_i) = r_0 p$, the values of $k_{\min}$ decreases with increasing r_0 and p.

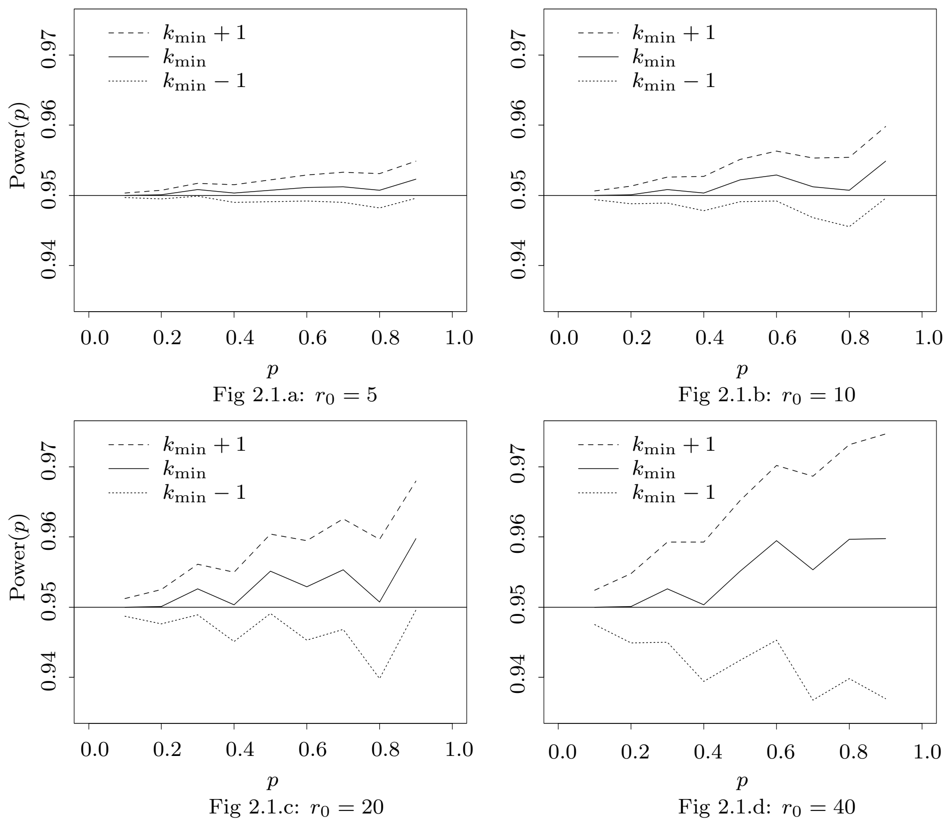

Figure 2.1. Power(p) *versus* p *for* $k = k_{\min} - 1, k_{\min}, k_{\min} + 1$ *When* M_i*'s Are I.I.D.* $Bin(r_0, p)$.

Since $k_{\min}$ is necessarily a positive integer, the power curve obtained by using $k = k_{\min}$ could be higher than the preassigned power, namely 0.95. To examine this potential discrepancy, we plot the power curve given by

$$\text{Power}(p) = 1 - \sum_{y=0}^{r} \Phi\left(1.645 - 0.2y^{1/2}\right) \binom{r}{y} p^y (1-p)^{r-y} \tag{2.18}$$

with $k = k_{\min}-1, k_{\min}, k_{\min}+1$. We evaluated the power curves from (2.18) with the p values given in Table 2.1 and $r_0 = 5, 10, 20, 40$. Figure 2.1 shows these power curves. As expected, the power falls under the preassigned value 0.95 when $k = k_{\min} - 1$ but it stays higher than 0.95 when $k = k_{\min} + 1$. Further, the difference between any two power curves increases as r_0 or p increases. Also, the difference between any two power curves increases as $E(M_i) = r_0 p$ increases.

Table 2.2. $k_{\min}$ **for the Example 2.4:** M_i**'s Are I.I.D. Poi**(λ_0)

λ_0	0.5	1.0	2.0	3.0	4.0	5.0	7.0	10.0	15.0	20.0	25.0
$k_{\min}$	545	273	137	91	69	55	39	28	19	14	11

Table 2.3. $k_{\min}$ **for the Example 2.5:** M_i**'s Are I.I.D. Geo**(p)

p	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$1/p$	1.11	1.25	1.43	1.67	2.0	2.5	3.33	5	10
$k_{\min}$	244	217	190	163	137	110	83	56	29

Next, we considered the Example 2.4 with M_i's distributed as i.i.d. Poi(λ_0). Then, $k_{\min}$ was given by (2.16) where $\lambda = k\lambda_0, z_{0.05} = 1.645$ and $\sigma^{-1} = 0.2$. We also looked at the Example 2.5 with M_i's distributed as i.i.d. Geo(p) and fixed $\lambda_0 = 0.5, 1(1)5, 7, 10, 15, 20, 25$, $p = 0.1(0.1)0.9$. Then, $k_{\min}$ is given by (2.16) and (2.17) respectively where $\lambda = k\lambda_0, z_{0.05} = 1.645$ and $\sigma^{-1} = 0.2$. By numerically solving the inequalities in (2.16) and (2.17), we obtained $k_{\min}$ associated with the preassigned power at least 0.95. Tables 2.2 and 2.3 respectively provide the $k_{\min}$ values corresponding (2.16) and (2.17). In the Example 2.4, since $E(M_i) = \lambda_0$, clearly $k_{\min}$ decreases with increasing λ_0 in Table 2.2. But, in the case of the Example 2.5, note that $E(M_i) = 1/p$ and hence $k_{\min}$ increases with increasing p which is validated by those entries in Table 2.3. To be consistent with the behavior found in the two previous tables, Table 2.3 supplies the $k_{\min}$ values as $1/p$ increases. Figures 2.2 and 2.3 respectively show the power curves when the M_i's are Poi(λ_0) and Geo(p) respectively. Here, the notable features are quite similar in principle to those shown in Figure 2.1.

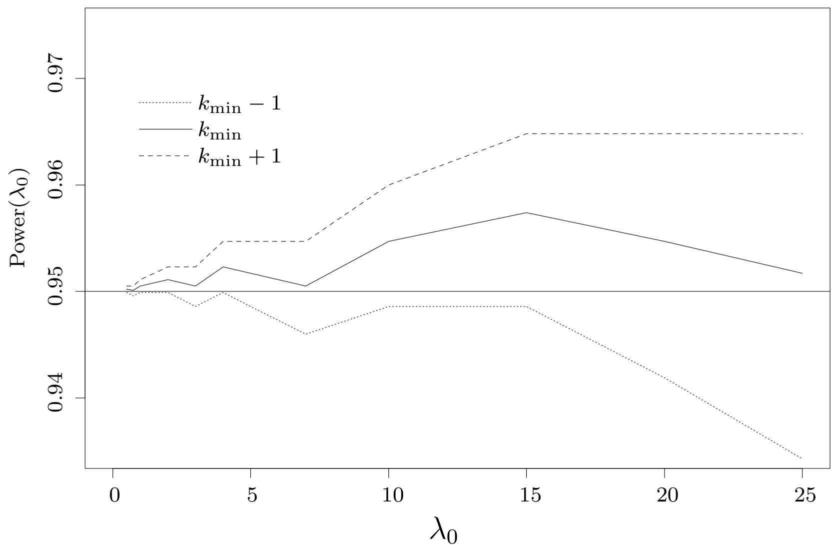

Figure 2.2. Power(λ_0) *versus* λ_0 *for* $k = k_{\min} - 1, k_{\min}, k_{\min} + 1$ *When* M_i *'s Are I.I.D.* $Poi(\lambda_0)$.

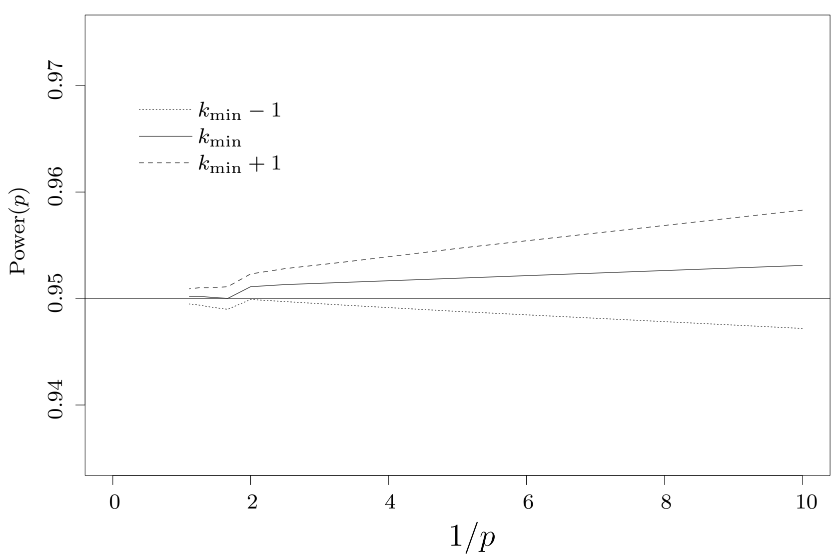

Figure 2.3. Power(p) *versus* $1/p$ *for* $k = k_{\min} - 1, k_{\min}, k_{\min} + 1$ *When* M_i*'s Are I.I.D.* $Geo(p)$.

3. A modified sequential probability ratio test. The total sample size associated with the best level α test with the preassigned power $1-\beta$ is then given by $M_{(k_{\min})} = \sum_{i=1}^{k_{\min}} M_i$. A natural question that arises is whether we should observe the available sequence of data $\{X_{i1}, \ldots, X_{iM_i}\}, i = 1, 2, \ldots$ successively based on "as needed basis" in order to simultaneously control both Type I and Type II errors at preassigned levels α, β respectively. We will show that a properly executed sequential strategy with approximately same α, β will require fewer than $k_{\min}$ groups of observations on the average.

We proceed by suitably modifying Wald's (1947) SPRT in order to come up with a well-defined sequential sampling strategy that would determine when to stop with a random number (K) of groups of random number (M_i) of random observations $\{X_{i1}, \ldots, X_{iM_i}\}$, $i = 1, 2, \ldots, K$. A terminal decision rule based on the data $\{X_{i1}, \ldots, X_{iM_i}\}, i = 1, 2, \ldots, K$ that would let one go in favor of either H_0 or H_1 as an integral part of this methodology. We set up the machinery in this section.

Along the line of Wald's original SPRT, let us initially choose and fix two numbers $0 < A < 1 < B < \infty$. A likelihood ratio that falls under A or above B would be deemed "too small" or "too large" respectively. We start with the first group of observations $\{X_{11}, \ldots, X_{1M_1}\}$ and then continue taking one additional group of observations at a time as long as

$$A < \Lambda_{l,\mathbf{M}^{(l)}} \equiv \frac{\Pi_{i=1}^{l} \Pi_{j=1}^{M_i} f(X_{ij}; \theta_1)}{\Pi_{i=1}^{l} \Pi_{j=1}^{M_i} f(X_{ij}; \theta_0)} < B, l = 1, 2, \ldots. \tag{3.1}$$

In other words, (3.1) defines the "continuation" region through the evaluation of the likelihood ratio at stage $l = 1, 2, \ldots$. We stop observing further additional groups of observations as soon as the likelihood ratio at some stage becomes too small $(\leq A)$ or too large $(\geq B)$. In other words, we continually watch for the likelihood ratio to go out of the interval (A, B) for the first time. At termination, if the likelihood ratio appears too small (too large), then we decide in favor of the hypothesis $H_0 (H_1)$.

The stopping time is defined as follows:

$$\begin{gathered}\text{Stop sampling with } K \text{ groups of observations where} \\ K \text{ is the first integer } k(\geq 1) \text{ such that } \Lambda_{k,\mathbf{M}^{(k)}} \notin (A,B).\end{gathered} \tag{3.2}$$

That is, we can express our sequential methodology as follows:

$$\begin{gathered}K = \inf\left\{k \geq 1 : \Lambda_{k,\mathbf{M}^{(k)}} \notin (A,B)\right\} \text{ whereas we accept } H_0 \ (H_1) \\ \text{if and only if } \Lambda_{K,\mathbf{M}^{(K)}} \leq A \ (\Lambda_{K,\mathbf{M}^{(K)}} \geq B) \text{ when sampling terminates.}\end{gathered} \tag{3.3}$$

Equivalently, with

$$\begin{gathered}Z_{ij} = \log \frac{f(X_{ij};\theta_1)}{f(X_{ij};\theta_0)}, a = \log(A), b = \log(B), \text{ and} \\ S_{i,M_i} = \textstyle\sum_{j=1}^{M_i} Z_{ij}, i = 1, 2, \ldots,\end{gathered} \tag{3.4}$$

we can rewrite the stopping time from (3.3) as

$$\begin{gathered}K = \inf\left\{k \geq 1 : \textstyle\sum_{i=1}^{k} S_{i,M_i} \notin (a,b)\right\} \text{ whereas we accept } H_0(H_1) \\ \text{if and only if } \textstyle\sum_{i=1}^{K} S_{i,M_i} \leq a \left(\textstyle\sum_{i=1}^{K} S_{i,M_i} \geq b\right) \text{ when sampling terminates.}\end{gathered} \tag{3.5}$$

3.1. *Termination with probability one.* Along the line of Stein's (1946) exponential rate of convergence for the tail probability of a stopping time based on CUSUM, we formulate the following result. Its proof is constructed along the line of Stein (1946) with some modifications which are briefly provided in the Appendix B.

Theorem 3.1. Exponential Rate. *Let* $\{M_i, Z_{i1}, \ldots, Z_{iM_i}, i = 1, 2, \ldots\}$ *be random variables defined on a probability* ($\mathcal{P}$) *space satisfying the conditions* **(A1)**-**(A2)** *from (2.2)-(2.3) such that* $\mathcal{P}(Z = 0) < 1$. *Let* a, b *be real numbers,* $a < b$, *and* K *be the smallest positive integer* k *for which the inequality* $a < S_{k,\mathbf{M}^{(k)}} < b$ *is first violated. Then, there exist constants* $c > 0$ *and* $0 < r < 1$ *such that* $\mathcal{P}(K > k) \leq cr^k$ *for all fixed* $k = 1, 2, \ldots$.

From this theorem, we can immediately summarize the following conclusions regarding the stopping variable K that was defined equivalently via (3.3) or (3.5).

Under the assumptions $\mathcal{P}(Z=0)<1$ and (A1)-(A2), we conclude:

(i) $\mathcal{P}(K<\infty)=1$, that is the modified SPRT (3.2) or equivalently (3.3) terminates with probability one;

(ii) $\mathbf{E}_{\mathcal{P}}[K^s]<\infty$ for all $s>0$, that is all positive moments of K are finite, in particular the mean and variance of K are finite; and

(iii) $\mathbf{E}_{\mathcal{P}}[\exp\{tK\}]<\infty$ for some $t, 0<t<t_0$, that is the moment generating function (m.g.f.) of K exists.

(3.6)

In (3.6), obviously the conclusion (iii) is the strongest which implies conclusion (ii) and then conclusion (i) follows. In view of (i) in (3.6), we can easily verify that

$$A \geq \beta/(1-\alpha) \text{ and } B \leq (1-\beta)/\alpha, \tag{3.7}$$

essentially by rehashing Wald's orginal derivation which leads to Wald's approximate boundaries

$$A \approx \beta/(1-\alpha) \text{ and } B \approx (1-\beta)/\alpha. \tag{3.8}$$

In practice, with preassigned α, β, one would first determine approximate A, B using (3.8) and then implement the sequential sampling methodology (3.3) or equivalently (3.5) for termination and decision making.

Example 2.1 (Continued). If $f(x;\theta)$ comes from a one-parameter exponential family (2.7), then it easily checks out that

$$\begin{aligned} &Z_{ij} = \log\{q(\theta_1)/q(\theta_0)\} + (\theta_1-\theta_0)R(X_{ij}), M_{(k)} = \textstyle\sum_{i=1}^{k} M_i, \text{ and} \\ &S_{k,\mathbf{M}^{(k)}} = M_{(k)}\log\{q(\theta_1)/q(\theta_0)\} + (\theta_1-\theta_0)\textstyle\sum_{i=1}^{k}\sum_{j=1}^{M_i} R(X_{ij}), \\ &\qquad\qquad k=1,2,\ldots. \end{aligned} \tag{3.9}$$

Without any loss of generality, let us suppose that $\theta_1>\theta_0$ so that the *continuation region* will then correspond to

$$\begin{aligned} &a^{\prime} - c^{\prime} M_{(k)} < \textstyle\sum_{i=1}^{k}\sum_{j=1}^{M_i} R(X_{ij}) < b^{\prime} - c^{\prime} M_{(k)} \text{ where } a^{\prime} = \frac{a}{\theta_1-\theta_0}, \\ &b^{\prime} = \frac{b}{\theta_1-\theta_0}, \text{ and } c^{\prime} = \frac{\log\{q(\theta_1)/q(\theta_0)\}}{\theta_1-\theta_0}, k=1,2,\ldots. \end{aligned} \tag{3.10}$$

The stopping time K is defined as the first integer k such that either $\sum_{i=1}^{k}\sum_{j=1}^{M_i} R(X_{ij}) \leq a^{'} - c^{'}M_{(k)}$ or $\geq b^{'} - c^{'}M_{(k)}$. In the former (latter) situation, we would decide in favor of $H_0(H_1)$.

In the case of a specific one-parameter exponential family $f(x;\theta)$, given fixed values of α, β and $\theta_0(<)\theta_1$, one can first determine $a^{'}, b^{'}$ and $c^{'}$, and then implement the modified sequential test (3.10).

3.2. *The Wald type approximations for the OC and ATSS functions.* A natural question is to ask how the proposed sequential test (3.3) would perform compared with the MP test given by Theorem 2.1. But, before we can embark upon addressing such queries, we need to obtain approximate expressions for the *operations characteristic* (OC) function and the *average total sample size* (ATSS) in the spirit of Wald. First, we state the associated version of Wald's fundamental identity of sequential analysis.

Theorem 3.2. Generalized Fundamental Identity of Sequential Analysis. *Let us consider the random variables* $\{M_i, Z_{i1}, \ldots, Z_{iM_i}, i = 1, 2, \ldots\}$ *defined on a probability* ($\mathcal{P}$) *space satisfying the conditions* **(A1)-(A2)** *from (2.2)-(2.3) such that* $\mathcal{P}(Z = 0) < 1$. *Let* a, b *be real numbers,* $a < b$, *and* K *be the smallest positive integer* k *for which the inequality* $a < S_{k,\mathbf{M}^{(k)}} < b$ *is first violated. Let* $\psi(t) \equiv \mathbf{E}_{\mathcal{P}}[\exp(tZ)]$, *the m.g.f. of* Z *with* t *real. Then,*

$$\mathbf{E}_{\mathcal{P}}\left[\exp\left\{tS_{K,\mathbf{M}^{(K)}}\right\}\{\psi(t)\}^{M_{(K)}}\right] = 1 \tag{3.11}$$

for all real t *such that* $\psi(t)$ *is finite.*

One can easily verify this identity along the line of Wald (1947). A prettier proof can be constructed by exploiting the Theorem 3.1 along the line of Bahadur (1958). We omit its proof for brevity. Now, in a wide variety of problems, one can (Wald (1947, p. 158)) find a unique real number $t_0 \equiv t_0(\mathcal{P})$ such that $\psi(t_0) = 1$. Then, (3.11) can be restated in general as

$$\mathbf{E}_{\mathcal{P}}\left[\exp\left\{t_0 S_{K,\mathbf{M}^{(K)}}\right\}\right] = 1 \tag{3.12}$$

In a standard parametric testing scenario, let us write

- (i) $L(\theta)$ for the probability of accepting H_0 when θ obtains, that is the customary OC function, and
- (ii) $\mathbf{E}_\theta\left[M_{(k)}\right]$ for the customary ATSS function where $M_{(k)} = \sum_{i=1}^{k} M_i$.

Now, in the light of (3.12), Wald's classical approximation for the OC function would continue to hold for the sequential test (3.3) and equivalently for (3.5) if we assume that the excess over boundaries at the point of termination is negligible. Henceforth, we will continue to enforce this assumption and that $\psi(t)$ is finite. We can immediately express the following approximation:

$$L(\theta) \approx \begin{cases} \{\exp(t_0 b) - 1\}\{\exp(t_0 b) - \exp(t_0 a)\}^{-1} \\ \qquad\qquad \text{for all } \theta \text{ such that } \mathbf{E}_\theta[Z] \neq 0 \\ b(b-a)^{-1} \text{ for } \theta \text{ such that } \mathbf{E}_\theta[Z] = 0. \end{cases} \tag{3.13}$$

Next, we turn our attention to Wald's first and second equations or lemmas (Wald (1947), Ghosh et al. (1997)) and we can write

$$\begin{aligned} \mathbf{E}_\theta\left[S_{K,\mathbf{M}^{(K)}}\right] &= \mathbf{E}_\theta\left[M_{(K)}\right]\mathbf{E}_\theta[Z], \\ \mathbf{E}_\theta\left[\left\{S_{K,\mathbf{M}^{(K)}} - M_{(K)}\mathbf{E}_\theta[Z]\right\}^2\right] &= \mathbf{E}_\theta\left[M_{(K)}\right]V_\theta[Z]. \end{aligned} \tag{3.14}$$

These lead to the following customary approximations for the ATSS function:

$$\mathbf{E}_\theta\left[M_{(K)}\right] \approx \begin{cases} \{aL(\theta) + b(1 - L(\theta))\}/\mathbf{E}_\theta[Z] \\ \qquad\qquad \text{for all } \theta \text{ such that } \mathbf{E}_\theta[Z] \neq 0 \\ -ab/\mathbf{E}_\theta[Z^2] \text{ for } \theta \text{ such that } \mathbf{E}_\theta[Z] = 0. \end{cases} \tag{3.15}$$

In addition to **(A1)**, if we also assume that $M_1, M_2, \ldots$ are i.i.d. with $\mathbf{E}[M_1]$ finite, then we note that $\mathbf{E}_\theta\left[M_{(K)}\right] = \mathbf{E}_\theta[K]\,\mathbf{E}[M_1]$. Thus, an approximation for the *average number of steps* or *stages* (ANS) function is immediately obtained from (3.15):

$$\mathbf{E}_\theta[K] \approx \begin{cases} \{aL(\theta)+b(1-L(\theta))\} / \{\mathbf{E}_\theta[Z]\,\mathbf{E}[M_1]\} \\ \qquad\qquad \text{for all } \theta \text{ such that } \mathbf{E}_\theta[Z] \neq 0 \\ -ab/\{\mathbf{E}_\theta[Z^2]\,\mathbf{E}[M_1]\} \text{ for } \theta \text{ such that } \mathbf{E}_\theta[Z]=0. \end{cases} \tag{3.16}$$

3.3. *Truncation of the MSPRT.* In some practical scenarios, one may not like to proceed further once one reaches certain number of steps, say l. A positive integer l may be fixed in advance that is guided by one or more natural restrictions on resources (for example, time, money) available to an experimenter. In a situation such as this, one will continue with the stopping rule (3.3) on a step by step basis. If the rule (3.3) stops sampling at the stage l or earlier, that is if $K \leq l$, then a decision to accept the hypothesis $H_0(H_1)$ in the usual fashion. On the other hand, if K reaches the level l but the rule (3.3) *demands* to move forward by observing the next batch of X's, then one must terminate the sampling procedure artificially right then and there. We define a truncated MSPRT that is associated with the stopping time given by

$$K_l \equiv \min\{K, l\}. \tag{3.17}$$

To be more specific, one would continually watch the successive likelihood ratios with the hope that at the l^{th} step or earlier, the likelihood ratio would go out of the interval (A, B) for the first time. The truncated stopping time is defined as follows:

Stop sampling with K groups of observations where K is the first integer $1 \le k \le l$ such that $\Lambda_{k,\mathbf{M}^{(k)}} \notin (A, B)$. If there is such an integer k, then we accept H_0 (H_1) if and only if $\Lambda_{K,\mathbf{M}^{(K)}} \le A$ ($\Lambda_{K,\mathbf{M}^{(K)}} \ge B$). But, if $\Lambda_{k,\mathbf{M}^{(k)}} \in (A, B)$ for all $k = 1, \ldots, l$, then the truncated MSPRT will forcibly terminate sampling at this point by accepting H_0 (H_1) if and only if $A < \Lambda_{l,\mathbf{M}^{(l)}} \le 1$ ($1 < \Lambda_{l,\mathbf{M}^{(l)}} < B$). In other words, the truncated number of steps is $K_l \equiv \min\{K, l\}$ whereas a terminal decision is made as described. (3.18)

In a practical scenario, it is reasonable to expect that $\alpha+\beta < 1$ so that this proposed truncated version (3.18) of the original MSPRT (3.3) would be implementable with the Wald approximations for A, B given by (3.8).

4. Computer simulations and data analysis. We reconsider the MP test given in Section 2.2 and the corresponding MSPRT described in Section 3. In this section, we set out to compare the MP test and MSPRT in the case of the earlier normal testing problem with $\theta_0 = 0$ and $\theta_1 = 1$, $\alpha = \beta = 0.05$ with the computer-generated data X_{ij}'s are from $N(\theta, 5^2)$. We again suppose that the M_i's are either i.i.d. Bin(r, p), Poi(ξ) or Geo(p) respectively as in the case of Examples 2.3, 2.4 and 2.5. We have performed the MSPRT using randomly generated data consisting of the X_{ij}'s and M_i's from their respective distributions by repeating the simulations $10,000$ times for each configuration. The simulation study was carried out using the parameters values listed in Table 4.1. For brevity, however, we present summaries of our findings for some of the selected configurations only.

Table 4.2 displays some typical simulated K and $M_{(K)}$ for MSPRT together with computed values of $k_{\min}$ and $\mathbf{E}[M_{k_{\min}}](= k_{\min}\mathbf{E}[M_1])$ for the corresponding MP test. In these cases, the parameters for the distribution of M_i's are selected in such a way that $\mathbf{E}[M_{k_{\min}}]$ values are approximately equal, so that we can compare the simulated $M_{(K)}$ values across examples. Clearly $k_{\min}$ and $\mathbf{E}[M_{k_{\min}}]$ are within their respective simulated minimum and maximum values but the average simulated values of K are much smaller than the corresponding $k_{\min}$ values.

Table 4.1. Parameters of the Distribution of M_i

1. Binomial Distribution: $\text{Bin}(s, p)$ $s = 5, 10, 15, 20, 25, 30, 35, 40$ and $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$
2. Poisson Distribution: $\text{Poi}(\xi)$ $\xi = 0.50, 0.75, 1, 2, 4, 5, 7, 10, 15, 20, 25$
3. Geometric Distribution: $\text{Geo}(p)$ $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

Consider a particular configuration, let K_j be the number of steps required for stopping in the j^{th} replication. Define

$$u_j = \begin{cases} 1 & \text{if } K_j \leq k_{min} \\ 0 & \text{if } K_j > k_{min} \end{cases}$$

and

$$v_j = \begin{cases} 1 & \text{if } M_{(K_j)} \leq \mathbf{E}\left[M_{(k_{min})}\right] \\ 0 & \text{if } M_{(K_j)} > \mathbf{E}\left[M_{(k_{min})}\right] \end{cases}$$

Then, clearly $\sum_j u_j$ gives the number of times the observed K was less or equal to k_{min}. Further, $\sum_j v_j$ gives the number of times the observed $M_{(K)}$ was less or equal to $M_{(k_{min})}$. The simulated mean and standard errors of these two variables using 10,000 replications are also given in Table 4.2. We find that more than 86% replications required fewer steps and observations to stop the procedure.

Table 4.2. Minimum K for Examples 2.3, 2.4 and 2.5 Together with Corresponding Simulated K and $M_{(K)}$ for MSPRT

Example 2.3: $M_i \sim \text{Bin}(20, 0.8)$

$$\Rightarrow \; k_{\min} = 17 \;\&\; \mathbf{E}[M_{(k_{\min})}] = 272$$

	Mean	**Std Error**	**Minimum**	**Maximum**
Simulated K	10.19	0.07	1	66
Simulated $M_{(K)}$	163.13	1.10	14	1064
u_j	0.8699	0.0033		
v_j	0.8593	0.0035		

$\#\{\text{Simulated } K \le k_{min}\} = 8699$; $\#\left\{\text{Simulated } M_{(K)} \le \mathbf{E}[M_{(k_{\min})}]\right\} = 8593$

Example 2.4: $M_i \sim \text{Poi}(2)$

$$\Rightarrow \; k_{\min} = 137 \;\&\; \mathbf{E}[M_{(k_{\min})}] = 274$$

	Mean	**Std Error**	**Minimum**	**Maximum**
Simulated K	62.34	0.45	5	466
Simulated $M_{(K)}$	144.19	1.02	16	1033
u_j	0.9333	0.0025		
v_j	0.8945	0.0031		

$\#\{\text{Simulated } K \le k_{min}\} = 9333$; $\#\left\{\text{Simulated } M_{(K)} \le \mathbf{E}[M_{(k_{\min})}]\right\} = 8945$

Example 2.5: $M_i \sim \text{Geo}(0.8)$

$$\Rightarrow \; k_{\min} = 217 \;\&\; \mathbf{E}[M_{(k_{\min})}] = 271.25$$

	Mean	**Std Error**	**Minimum**	**Maximum**
Simulated K	113.01	0.79	11	706
Simulated $M_{(K)}$	141.28	0.99	12	858
u_j	0.8973	0.0030		
v_j	0.8970	0.0030		

$\#\{\text{Simulated } K \le k_{min}\} = 8973$; $\#\left\{\text{Simulated } M_{(K)} \le \mathbf{E}[M_{(k_{\min})}]\right\} = 8970$

Our simulations show that, in general MSPRT test requires on an average fewer sampling steps (or stages) and also fewer observations on an average than the MP test. Thus, we feel tempted to claim that the proposed MSPRT is more efficient then the MP test having comparable values of α and β.

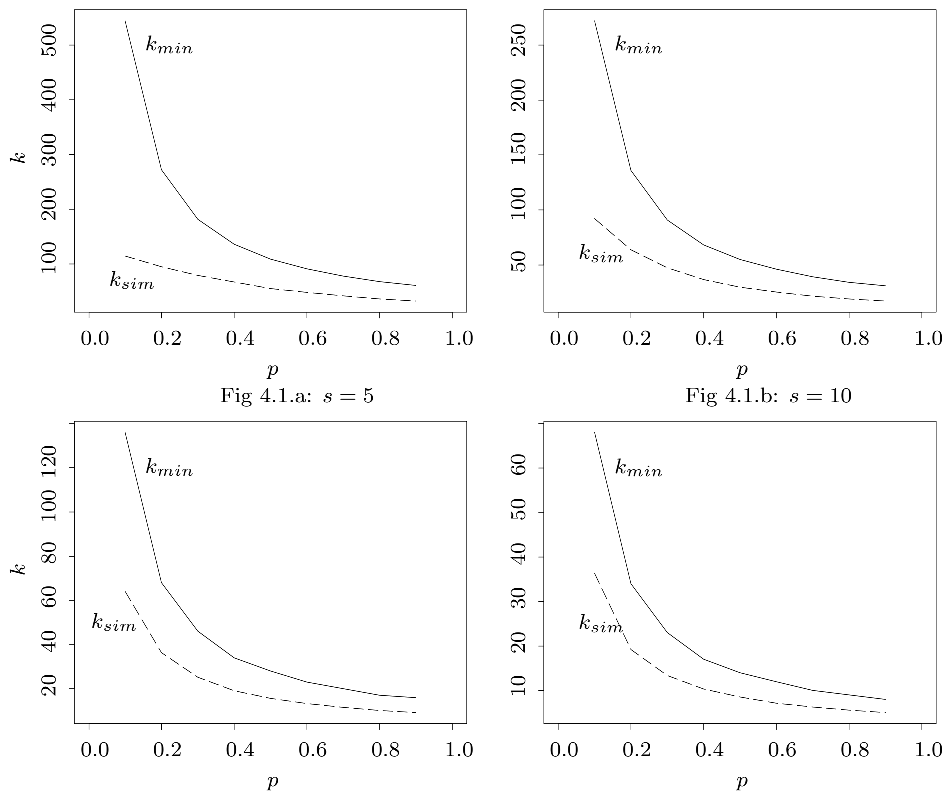

Fig 4.1.a: $s = 5$
Fig 4.1.b: $s = 10$
Fig 4.1.c: $s = 20$
Fig 4.1.d: $s = 40$

Figure 4.1. $k_{\min}$ *and* K_{sim} *When* $M_i \sim Bin(s, p)$.

Figure 4.1 shows a typical set of curves for K_{sim} (the average simulated values of K for MSPRT) and $k_{\min}$ (minimum K for MP test). The curves in this figure correspond to Example 2.3 from Section 2.2, that is, the distribution of M_i is $\text{Bin}(s, p)$. The k versus p curves are plotted for $p = 0.1, 0.2, \ldots, 0.9$ in Figures 4.1.a, 4.1.b, 4.1.c and 4.1.d for a fixed s value with $s = 5, 10, 20$ and 40 respectively. Similar plots were also obtained in all simulated exercises. The curves in Figure 4.1 show that $k_{\min}$ - $K_{sim} > 0$ for all p and s under consideration, however, the difference between K_{sim} and $k_{\min}$ reduces when p or s increases. That is, as $p \to 1$ and/or as $s \to \infty$, we noted that the difference $k_{\min}$ - K_{sim} became incredibly small. When M_i's were generated from a Poi(ξ) or Geo(p) distribution, we noticed similar features too. Further, in the Poi(ξ) (or Geo(p)) case, we found that as $\xi \to \infty$ (or as $p \to 0$), the difference $k_{\min}$ - K_{sim} became incredibly small.

4.1. *Evaluation of the OC functions for the MSPRT.* We continue with MSPRT for testing $H_0 : \theta = 0$ against $H_1 : \theta = 1$ at $\alpha = \beta = 0.05$ when the data X_{ij}'s followed $N(\theta, 5^2)$ distribution. Then from (3.4) or (3.9), we have

$$Z_{ij} = (2X_{ij} - 1)/50. \tag{4.1}$$

Obviously, $\mathbf{E}_\theta[Z] = (2\theta - 1)/50$ and thus from (3.13), an approximate OC function, $L(\theta)$, for the MSPRT is given by

$$L(\theta) \approx \left(19^{(2\theta-1)} + 1\right)^{-1}, \ \theta \in (-\infty, \infty). \tag{4.2}$$

Next, we set out to compare the approximate expression for $L(\theta)$ given above with the corresponding simulated values. Here, we again simulated the data X_{ij}'s that was generated from $N(\theta, 25)$ while the size of each batch was determined by the i.i.d. distribution of M_i. The simulations were carried out for all the distributions listed in Table 4.1.

Figure 4.2.a and Figure 4.2.b provide the plots for the approximate expression of $L(\theta)$ $(-0.5 \le \theta \le 1.5)$ from (4.2) as well as the simulated operations characteristic curve, namely OC(θ) ($\theta = -0.50, -0.45, \ldots, 1.50$) versus θ for the above MSPRT test. Here we assumed that the M_i's were distributed as Poi(λ_0) with $\lambda_0 = 1, 5$. Clearly the results show that the approximate values of $L(\theta)$ given in (4.2) are very close to the simulated values. However, the simulated values were slightly higher than the approximate values at or near $\theta = 0$ (that is, when H_0 is true) and slightly

lower around $\theta = 1$ (that is, when H_1 is true). Similar features were also noted for other values of λ_0 and also for other distributions of the M_i's given in Table 4.1.

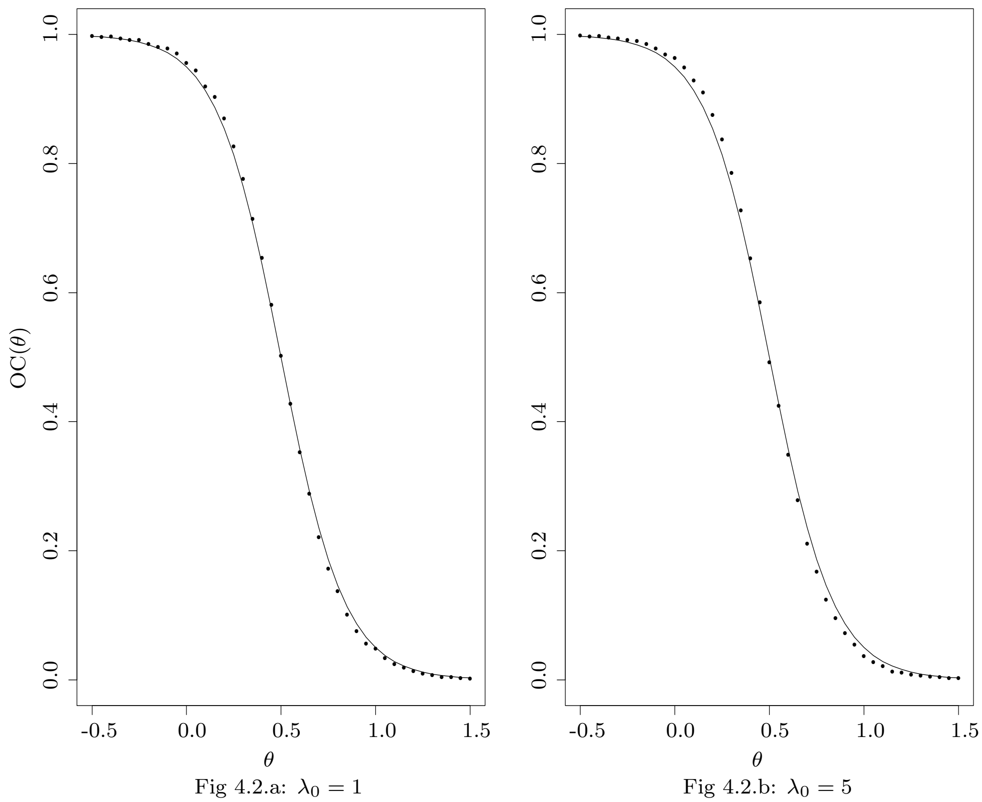

Figure 4.2. *Approximate $L(\theta)$ and Simulated OC(θ) When M_i's Are I.I.D. Poi(λ_0)* (*Legend*: '____' = *Approximate* $L(\theta)$ *and* '**....**' = *Simulated* $OC(\theta)$) .

4.2. *ATSS and ANS functions for Examples 2.3-2.5.* Since the data X_{ij}'s was generated from $N(\theta, 25)$ from (4.1),

$$\mathbf{E}[Z^2] = \frac{1}{25} \quad \text{when } \theta = \frac{1}{2}.$$

Therefore from (3.15) the Average Total Sample Size (ATSS) function, $\mathbf{E}_\theta[M_{(K)}]$ for the MSPRT test in Section 4.1 is given by

$$\mathbf{E}_\theta[M_{(K)}] \approx \begin{cases} 50\log(19)\left\{1 - 2L(\theta)\right\}(2\theta - 1)^{-1} & \text{if } \theta \neq \frac{1}{2} \\ 25\left\{\log(19)\right\}^2 & \text{if } \theta = \frac{1}{2}. \end{cases} \tag{4.3}$$

Also because in the examples considered, $M_1, M_2, \ldots$ are i.i.d. random variables, $\mathbf{E}_\theta[M_{(K)}] = \mathbf{E}_\theta[K]\mathbf{E}_\theta[M_1]$ and thus the Average Number of Steps (ANS) is given by

$$\mathbf{E}_\theta[K] \approx \begin{cases} \mathbf{E}_\theta[M_{(K)}]/\,(r_0 p) & \text{if } M_i \sim \text{Bin}(r_0, p) \\ \mathbf{E}_\theta[M_{(K)}]/\lambda_0 & \text{if } M_i \sim \text{Poi}(\lambda_0) \\ p\mathbf{E}_\theta[M_{(K)}] & \text{if } M_i \sim \text{Geo}(p). \end{cases} \tag{4.4}$$

Table 4.3. Expected and Average Simulated Values of K and $M_{(K)}$ When M_i's Are I.I.D. Bin$(20, p)$

		$p = 0.1$			$p = 0.5$			$p = 0.9$		
θ	$\mathbf{E}[M_{(K)}]$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$
-1.00	49.1	24.5	23.2	52.7	4.9	5.8	58.0	2.7	3.5	62.8
-0.75	58.8	29.4	28.0	63.6	5.9	6.9	69.4	3.3	4.1	73.8
-0.50	73.2	36.6	34.7	79.0	7.3	8.4	84.3	4.1	5.1	91.3
-0.25	95.8	47.9	45.2	102.8	9.6	11.2	111.9	5.3	6.5	117.7
0.00	132.5	66.2	63.4	144.1	13.2	15.8	158.1	7.4	9.4	168.7
0.25	184.6	92.3	90.0	204.8	18.5	22.6	226.3	10.3	13.5	244.0
0.50	216.7	108.4	107.7	245.1	21.7	27.7	277.1	12.0	16.9	303.8
0.75	184.6	92.3	91.3	207.8	18.5	22.5	224.8	10.3	13.6	244.6
1.00	132.5	66.2	62.7	142.7	13.2	15.8	158.7	7.4	9.3	167.3
1.25	95.8	47.9	45.0	102.5	9.6	11.1	111.3	5.3	6.6	119.0
1.50	73.2	36.6	34.5	78.5	7.3	8.6	86.3	4.1	5.1	91.0
1.75	58.8	29.4	28.1	64.0	5.9	7.0	69.6	3.3	4.1	73.9
2.00	49.1	24.5	23.3	53.1	4.9	5.9	58.7	2.7	3.5	62.9

Table 4.4. Expected and Average Simulated Values of K and $M_{(K)}$ When M_i's Are I.I.D. Bin$(r_0, 0.5)$

		r_0= 10			r_0= 20			r_0= 40		
θ	$\mathbf{E}[M_{(K)}]$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$
-1.00	49.1	9.8	11.0	55.1	4.9	5.8	58.0	2.5	3.2	64.0
-0.75	58.8	11.8	13.1	65.8	5.9	6.9	69.4	2.9	3.8	75.3
-0.50	73.2	14.6	16.3	81.4	7.3	8.4	84.3	3.7	4.6	92.4
-0.25	95.8	19.2	21.2	106.1	9.6	11.2	111.9	4.8	6.1	121.1
0.00	132.5	26.5	30.0	150.0	13.2	15.8	158.1	6.6	8.4	168.6
0.25	184.6	36.9	42.6	213.1	18.5	22.6	226.3	9.2	12.5	250.1
0.50	216.7	43.3	51.2	256.4	21.7	27.7	277.1	10.8	15.3	306.4
0.75	184.6	36.9	43.3	216.5	18.5	22.5	224.8	9.2	12.3	246.7
1.00	132.5	26.5	29.7	148.4	13.2	15.8	158.7	6.6	8.4	168.2
1.25	95.8	19.2	21.1	105.6	9.6	11.1	111.3	4.8	6.0	119.1
1.50	73.2	14.6	16.3	81.3	7.3	8.6	86.3	3.7	4.6	91.8
1.75	58.8	11.8	13.0	65.3	5.9	7.0	69.6	2.9	3.8	75.5
2.00	49.1	9.8	11.0	54.9	4.9	5.9	58.7	2.5	3.2	63.5

Table 4.5. Expected and Average Simulated Values of K and $M_{(K)}$ When M_i's Are I.I.D. Poi(λ_0)

		$\lambda_0 = 0.5$			$\lambda_0 = 1.0$			$\lambda_0 = 5.0$		
θ	$\mathbf{E}[M_{(K)}]$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$
-1.00	49.1	98.1	40.8	51.8	49.1	33.2	52.6	9.8	11.0	55.2
-0.75	58.8	117.6	48.8	62.0	58.8	39.6	62.5	11.8	13.1	66.0
-0.50	73.2	146.4	60.9	77.4	73.2	49.3	78.0	14.6	16.4	82.3
-0.25	95.8	191.6	79.1	100.5	95.8	64.8	102.4	19.2	21.2	106.5
0.00	132.5	265.0	110.5	140.4	132.5	88.4	139.7	26.5	29.7	149.7
0.25	184.6	369.1	157.5	200.2	184.6	129.9	205.4	36.9	42.6	214.6
0.50	216.7	433.5	189.0	240.1	216.7	153.4	242.8	43.3	51.3	258.2
0.75	184.6	369.1	158.3	201.1	184.6	128.3	203.0	36.9	42.9	215.8
1.00	132.5	265.0	109.8	139.6	132.5	90.4	143.0	26.5	29.7	149.0
1.25	95.8	191.6	79.0	100.4	95.8	64.1	101.5	19.2	21.3	107.1
1.50	73.2	146.4	61.1	77.7	73.2	49.1	77.7	14.6	16.3	81.9
1.75	58.8	117.6	48.8	62.0	58.8	39.2	62.0	11.8	13.1	65.7
2.00	49.1	98.1	41.0	52.0	49.1	33.1	52.2	9.8	11.1	55.8

Table 4.6. Expected and Average Simulated Values of K and $M_{(K)}$ When M_i's Are I.I.D. Geo(p)

		$p = 0.9$			$p = 0.5$			$p = 0.1$		
θ	$\mathbf{E}[M_{(K)}]$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	$\mathbf{E}[K]$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$
-1.00	49.1	44.2	46.4	51.5	24.5	26.3	52.6	4.9	6.3	62.9
-0.75	58.8	52.9	55.4	61.6	29.4	31.7	63.2	5.9	7.4	73.9
-0.50	73.2	65.9	68.5	76.1	36.6	39.3	78.7	7.3	9.0	90.3
-0.25	95.8	86.2	91.5	101.6	47.9	51.7	103.5	9.6	11.6	116.5
0.00	132.5	119.2	126.1	140.0	66.2	71.8	143.5	13.2	16.3	162.7
0.25	184.6	166.1	175.8	195.3	92.3	102.1	204.2	18.5	24.0	240.7
0.50	216.7	195.1	210.6	233.9	108.4	122.2	244.5	21.7	29.1	291.5
0.75	184.6	166.1	176.1	195.7	92.3	101.8	203.5	18.5	24.2	240.6
1.00	132.5	119.2	127.5	141.6	66.2	72.5	145.2	13.2	16.5	164.8
1.25	95.8	86.2	90.0	100.0	47.9	51.7	103.2	9.6	11.6	116.6
1.50	73.2	65.9	69.0	76.6	36.6	39.6	79.4	7.3	9.2	90.8
1.75	58.8	52.9	55.4	61.5	29.4	31.7	63.6	5.9	7.3	73.4
2.00	49.1	44.2	46.2	51.4	24.5	26.3	52.7	4.9	6.2	62.5

Tables 4.3 - 4.6 give the computed approximate values of $\mathbf{E}_\theta[K]$ and $\mathbf{E}_\theta[M_{(K)}]$ and their simulated values for θ. The approximation for $\mathbf{E}_\theta[M_{(K)}]$ given in (4.3) is not entirely dependent on the distribution of M_i. However, the tables show that the average simulated values of $M_{(K)}$, ($\overline{M}_{(K)}$) depend on the distribution of M_i and they are increasing with decreasing values of $\mathbf{E}_\theta[K]$. Further, the simulated values of $\overline{M}_{(K)}$ are higher than the corresponding $\mathbf{E}_\theta[M_{(K)}]$ for all the distributions considered in simulations. Figure 4.3 gives the curves $\mathbf{E}_\theta[M_{(K)}]$ versus θ and $\overline{M}_{(K)}$ versus θ for $p = 0.1, 0.5, 0.9$ for the case where the distribution of M_i is Bin$(20, p)$. Even though the approximate formula for $\mathbf{E}_\theta[M_{(K)}]$ given by (4.3) does not entirely depend on the distribution of M_i.

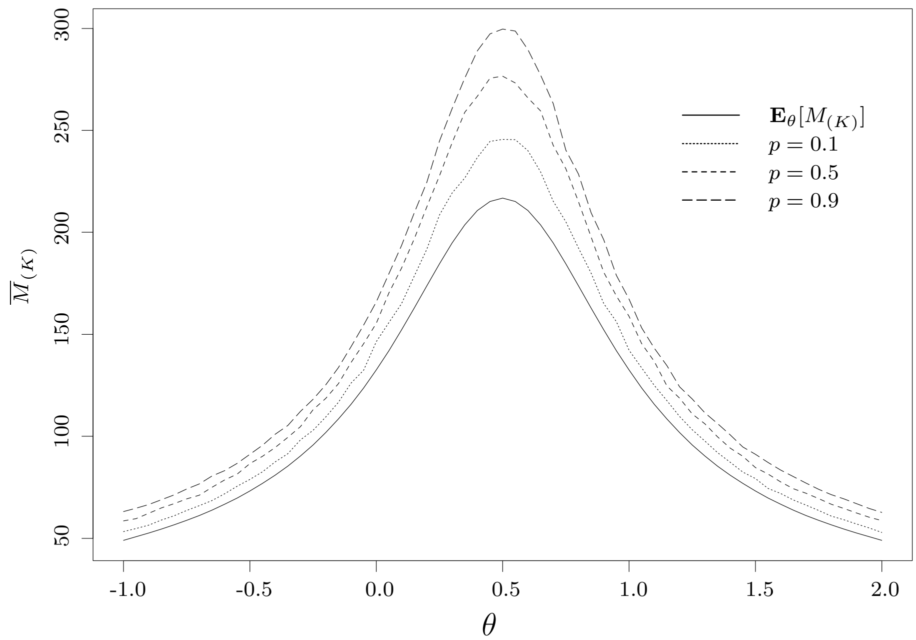

Figure 4.3. $\mathbf{E}_\theta[M_{(K)}]$ *and* $\overline{M}_{(K)}$ *versus* θ *When* M_i*'s Are I.I.D.* $\text{Bin}(20, p)$.

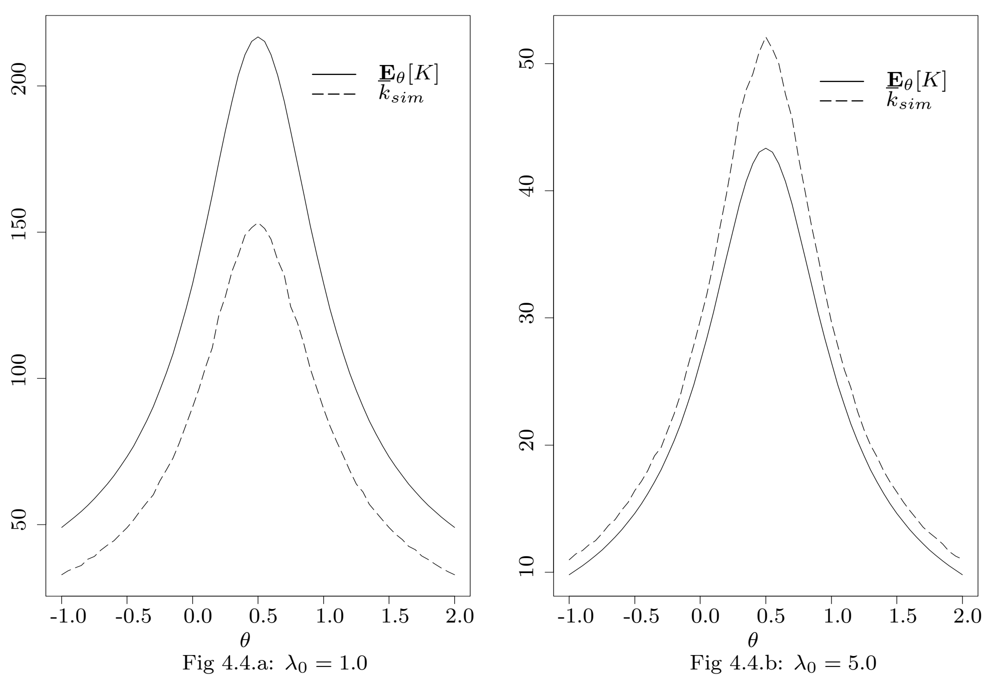

Figure 4.4. $\mathbf{E}_\theta[K]$ *and* $\overline{k}_{sim}$ *versus* θ *When* M_i*'s Are I.I.D.* $\text{Poi}(\lambda_0)$.

Figure 4.4 gives plots of the curves $\mathbf{E}_\theta[K]$ versus θ and $\overline{k}_{sim}$ versus θ when M_i is Poi(λ_0) $(\lambda_0 = 1, 5)$. This figure and Tables 4.3 - 4.6 show that the the difference, $(\overline{k}_{sim} - \mathbf{E}_\theta[K])$ goes from negative to positive as the parameter value (that is, $\mathbf{E}_\theta[M_i]$) increases.

4.3. *Performance of a truncated version of the MSPRT.* Consider simulation of truncated MSPRT as explained Section 3.3. Simulations were performed using data generated from $N(0, 5^2)$ and $N(1, 5^2)$ in order to estimate the power (P) and the probability of type one error (α) of the truncated test respectively. Table 4.7 shows the these estimated values, $\widehat{p}$ and $\widehat{\alpha}$, for four different truncated steps (l) and untruncated case $(l = \infty)$ for three examples considered in Section 2. Simulations were carried out for three examples using the parameters listed in Table 4.1. However, Table 4.7 shows only the results for some selected parameter values (as in Table 4.2) for brevity. Here, N_l is the number of simulated tests that were truncated at the step l among the total of 10,000 replications.

Table 4.7 shows that $\overline{k}_{sim}$ and $\overline{M}_{(K)}$ increase with increasing l and approach their limiting values (untruncated) as $l \to \infty$. Further, in general the estimated probability of Type I error $\widehat{\alpha}$ is higher than the preset value $\alpha = 0.05$. Similarly, the truncated tests gave lower power than the preset value 0.95. However, the simulation results show that reasonably high values of l give results that are close to the ones obtained in the untruncated cases. For Example 2.3 with $M_i \sim$ Bin$(20, 0.8)$ when $l = 20$, the estimated values are $\widehat{\alpha} = 0.049$ and $\widehat{p} = 0.95$. Similarly, truncated tests gave good results when $l = 150$ and $l = 300$ respectively in the case of Example 2.4 with $M_i \sim$ Poi(2) and Example 2.5 with $M_i \sim$ Geo(0.8).

Table 4.7. Estimated Power ($\widehat{p}$) and Probability of Type I Error ($\widehat{\alpha}$) for Truncated MSPRT Together with Corresponding K_{sim} and $\overline{M_K}$ for Examples 2.3, 2.4 and 2.5

		$\theta=0$				$\theta=1$			
	l	N_l	$\widehat{\alpha}$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$	N_l	$\widehat{p}$	$\overline{k}_{sim}$	$\overline{M}_{(K)}$
$M_i \sim \text{Bin}(20, 0.8)$	10	3631	0.107	7.60	121.58	3655	0.897	7.58	121.28
	13	2344	0.074	8.57	137.16	2367	0.922	8.52	136.22
	15	1731	0.067	9.00	143.94	1695	0.934	8.95	143.16
	20	778	0.049	9.53	152.50	814	0.950	9.55	152.82
	∞	0	0.030	10.27	164.41	0	0.968	10.25	164.09
$M_i \sim \text{Poi}(2)$	50	4964	0.145	40.63	94.00	5020	0.856	40.63	93.97
	75	2791	0.104	50.08	115.78	2756	0.900	49.99	115.63
	100	1565	0.068	55.86	129.31	1505	0.927	54.91	127.01
	150	474	0.051	59.96	138.62	486	0.949	60.09	138.99
	∞	0	0.042	62.67	144.88	0	0.959	62.83	145.33
$M_i \sim \text{Geo}(0.8)$	100	4401	0.131	77.95	97.40	4345	0.870	77.54	97.01
	150	2257	0.089	93.80	117.17	2308	0.912	94.01	117.57
	200	1243	0.069	103.39	129.23	1233	0.930	102.76	128.44
	300	362	0.049	110.30	137.87	358	0.948	110.58	138.28
	∞	0	0.047	111.61	139.53	0	0.956	112.79	140.91

5. Two examples with data analyses.

In the introduction, we had mentioned an example of gathering information about p, the fraction of cars without working brake lights on a highway. The first example we have in mind relates to testing for θ in a specific binomial problem of interest and we present this in Section 5.1. The next problem relates to testing for ν in a $N(\nu, \nu)$ population which is presented in Section 5.2. In both problems, we have collected real data with the help of our proposed MSPRT methodology and we present their analysis.

5.1. *Testing the quality of packaged rice.*

Following a disaster, natural or not, refugees often gather around in some remote place and they are provided shelter, security, clothing, medicines, food, and some other absolute necessities by many relief agencies run by the government and private sector. A vendor may supply hundreds and thousands of bags of rice, for example, to various relief agencies for distribution among families in refugee-shelters and bill those agencies. Price hiking (for example, per bag or per pound) happens to be a routine phenomenon in an otherwise chaotic and emergency situation of human suffering. Vendors are obviously out there to make money quickly at the expense of urgency to alleviate day-to-day suffering, but on the other hand money happens to be short in supply for any relief agency. The "quality" of rice in packages, for example, may be quantified by the percentage p of broken, chipped, or discolored grains per pound. Obviously, a higher value of p would indicate a lesser quality for the supplied bags.

A vendor may set a unit price that is on the higher side while pretending to supply rice grains of better quality than what the true quality is. Suppose that a vendor charges \$7.95 per five-pound bag while claiming that $p = 0.15$ for the supplied load. On the other hand, this vendor may be supplying truck-loads of five-pound bags with $p = 0.30$ that should normally sell for \$6.15 per bag, so that the vendor charges \$1.80 extra on each bag. Now, let us throw in some numbers, for example! In a refugee camp, suppose that there are 2300 refugee families and each family supposedly consumes two packages of rice on an average per week simply because there is hardly anything else to eat. So, by price hiking alone, the vendor will overcharge \$8280 per week! Now, this may obviously go on for months or years at a time and across many different items (for example, timber, clothings) along with bags of rice grains. Also, the number of refugees may go much higher! Thus, evaluating the true "quality" of all supplied items is a matter of real importance in situations like these, because otherwise the relief agencies end up making overpayments by hundreds of thousands of dollars to their suppliers within a rather short period of time!

Now, in a testing problem, with $H_0 : p = p_0$ against $H_1 : p = p_1, 0 < p_0 < p_1 < 1$, a consequence of committing the Type II error would normally lead to significant overpayment by the relief agencies to the vendor. Hence, we will be tempted to pick the Type II error probability β (much smaller α) in this specific problem!

From a five-pound bag of packaged (basmati) rice bought in a retail store, we filled 5 separate brown bags each having two-cups of rice grains picked randomly. Let p stand for the fraction of broken, chipped, or discolored rice grains in two cups of the basmati brand that we had bought. We thought of testing a simple null hypothesis $H_0 : p = 0.15$ against a simple alternative hypothesis $H_1 : p = 0.20$ with $\alpha = 0.10$ and $\beta = 0.01$. We choose β much lower than α in this specific problem because committing the Type II error would result in serious overpayment to the retail store.

We have

$$f(x; p) = p^x(1-p)^{1-x} \text{ with } x = 0, 1 \text{ and } p \in (0, 1), \tag{5.1}$$

so that this corresponds to (2.7) where

$$\theta = \log\left(p/(1-p)\right),\ q(\theta) = (1+e^\theta)^{-1},\ p(x) = 1,\ \text{and } R(x) = x,$$

for $\theta \in (-\infty, \infty)$. Now, with

$$\begin{aligned}
\theta_0 &= \log\left(0.15/(1-0.15)\right) = -1.7346, \\
\theta_1 &= \log\left(0.20/(1-0.20)\right) = -1.3863, \\
q(\theta_0) &= (1+e^{-1.7346})^{-1} = 0.85, q(\theta_1) = (1+e^{-1.3863})^{-1} = 0.80,
\end{aligned}$$

and then utilizing (3.9), we can express

$$\begin{aligned}
Z_{ij} = \log \frac{f(X_{ij};\theta_1)}{f(X_{ij};\theta_0)} &= \log(\tfrac{0.80}{0.85}) + (-1.3863 + 1.7346)X_{ij} \\
&= -0.060625 + 0.3483 X_{ij},
\end{aligned}$$

where $X_{ij} = 1$ (0) if the j^{th} observation in the i^{th} group happens to be a rice grain that is (not) broken, chipped, or discolored. We reach inside a brown bag and select a pinch of grains at each stage. For the i^{th} group or stage, M_i stands for the number of selected grains whereas $\sum_{j=1}^{M_i} X_{ij}$ stands for the number of broken, chipped, or discolored grains among M_i selected rice grains, $i = 1, 2, \ldots$. Let us denote

$$T_{k,\mathbf{M}^{(k)}} = -0.060625 M_{(k)} + 0.3483 \textstyle\sum_{i=1}^{k} \sum_{j=1}^{M_i} X_{ij}, k \geq 1. \tag{5.2}$$

Now, with $a = \log(0.01/(1-0.10)) = -4.4998$, $b = \log((1-0.01)/0.10) = 2.2925$ we see that (3.10) reduces to the following stopping variable:

$$\begin{gathered}
K = \inf\left\{k \geq 1 : T_{k,\mathbf{M}^{(k)}} \notin (-4.4998, 2.2925)\right\} \text{ whereas we accept } H_0 \ (H_1) \\
\text{if and only if } T_{k,\mathbf{M}^{(k)}} \leq -4.4998 \ (\geq 2.2925) \text{ when sampling terminates.}
\end{gathered} \tag{5.3}$$

Table 5.1 illustrates how the methodology (5.3) is implemented in practice. In step #1 (that is, with $k = 1$), we took out a pinch of grains from the bag according to *simple random sampling without replacement* (SRSWOR) and we found that there were 41 grains out of which 10 were broken, chipped, or discolored so that

$$M_1 = 41, \ M_{(1)} = M_1 = 41, \ \textstyle\sum_{j=1}^{M_1} X_{1j} = 10, \ \mathbf{M}^{(1)} = (41) \text{ and } T_{1,\mathbf{M}^{(1)}} = 0.99738.$$

Table 5.1. Implementation #1 of the MSPRT (5.2) with $\alpha = 0.01$, $\beta = 0.01, a = -4.4998, b = 2.2925$

Step $\#k$	M_k	$M_{(k)}$	$\sum_{i=1}^{k}\sum_{j=1}^{M_i} X_{ij}$	$T_{k,\mathbf{M}^{(k)}}$	Continue or Terminate?
1	41	41	10	0.99738	Continue Sampling
2	45	86	19	1.40400	Continue Sampling
3	27	113	24	1.50860	Continue Sampling
4	34	147	30	1.53710	Continue Sampling
5	52	199	44	3.26080	Terminate Sampling

Decision: With $k = 5$, we accept H_1, that is, we conclude that the fraction of broken, chipped, or discolored rice grains is 20% in the bag.

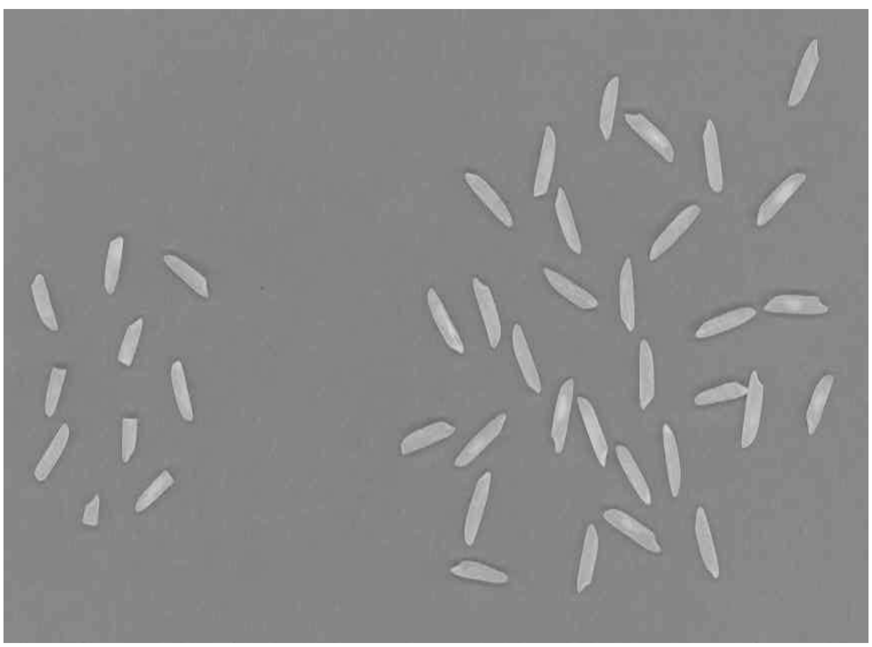

Figure 5.1. *Grain Samples from Step #1 in Table 5.1. Out of 41 Rice Grains, 10 Were Broken, Chipped, or Discolored.*

Table 5.2. Five Independent Implementations of the MSPRT (5.2) with $\alpha = 0.01, \beta = 0.01, a = -4.4998, b = 2.2925$

Implement # K	M_1	M_2	M_3	M_4	M_5	$M_{(K)}$	$\sum_{i=1}^{K}\sum_{j=1}^{M_i} X_{ij}$ $T_{K,\mathbf{M}^{(K)}}$ Terminal Decision
Implement #1 5	41	45	27	34	52	199	44 3.2608 Accept H_1
Implement #2 4	28	59	52	58	-	197	45 3.7304 Accept H_1
Implement #3 2	59	62	-	-	-	121	29 2.7651 Accept H_1
Implement #4 3	41	59	58	-	-	158	36 2.9601 Accept H_1
Implement #5 2	78	38	-	-	-	116	31 3.7648 Accept H_1

In Figure 5.1, we present a picture of the observations obtained only from step #1 given in Table 5.1 to get a feeling of exactly what each step was like and how the experiment was carried out.

The brown bags with 2 cups of rice in each had so many grains that really it would not have mattered if we sampled with replacement (SRSWR) instead. We found that $T_{1,\mathbf{M}^{(1)}}$ was between a and b, and hence we moved to the next step (that is, $k = 2$). The successive steps are shown in Table 5.1. We note that we went up to $k = 5$ where we terminated sampling from the brown bag and decided to accept H_1. At termination, one notes that we had observed $199(= M_{(k)}$, TSS) grains with a total of $44(= \sum_{i=1}^{5}\sum_{j=1}^{M_i} X_{ij})$ broken, chipped, or discolored grains and $T_{5,\mathbf{M}^{(5)}} = 3.2608$.

In Table 5.2, the first row corresponds to the summary statistics obtained from implementation #1 (that is, from Table 5.1). Next, we repeated the same process by sampling from the other four brown bags. Then, we successively required $4, 2, 3$ and 2 steps in order to terminate sampling whereas the decision was to accept H_1 at each termination after observing $197, 121, 158$ and $116(= M_{(k)}$, TSS) grains with a total of $45, 45, 36$ and $31(= \sum_{i=1}^{5}\sum_{j=1}^{M_i} X_{ij})$ broken, chipped, or discolored grains respectively. If we had repeated the process many more times from these same bags, surely we would have seen some instances where we would decide to accept H_0 instead of accepting H_1 at termination!

5.2. *Testing the average number of typographical errors per page.* We considered a sequence of independent observations from a $N(\nu, \nu)$ distribution where $\nu(> 0)$ is an unknown parameter. Such a model would make some good sense to approximate a Poi(ν) distribution, especially when $\nu(> 0)$ is moderately large. It is well-known that the Poisson distribution has been suggested as a way of modeling typographical errors in a manuscript.

We conducted an experiment using an early draft of a textbook on probability and statistics consisting of more than 300 pages that was sent to one of us by a publisher for review. A normal Q-Q plot constructed by Mukhopadhyay and Cicconetti (2004) for a pilot sample of size 36 indicated that the observed data was approximated well by a normal distribution. Also, the mean and variance from the pilot data sample were reported to be 9.305556 and 11.24683 respectively, and Mukhopadhyay and Cicconetti (2004) felt that it was reasonable to assume that the mean and variance were same as they attempted to describe the number of typographical errors per page in this dataset. This scenario led them to consider the $N(\nu,\nu)$ distribution as a possible model for the number of typographical errors per page in this dataset.

We thought of testing a simple null hypothesis $H_0 : \nu = 12$ against a simple alternative hypothesis $H_1 : \nu = 13$ with $\alpha = \beta = 0.05$. We have

$$f(x;\nu) = \frac{1}{\sqrt{2\pi\nu}} e^{-\nu/2} e^{x} e^{-x^2/(2\nu)}, -\infty < x < \infty, \nu > 0, \tag{5.4}$$

so that this corresponds to (2.7) where

$$\theta = (2\nu)^{-1}, q(\theta) = (\theta/\pi)^{1/2} e^{-1/(4\theta)}, p(x) = e^{x}, \text{ and } R(x) = -x^2,$$

for $-\infty < x < \infty$ and $\theta \in (0,\infty)$. Now, with

$$\begin{aligned} \theta_0 &= (2(12))^{-1} = 0.041667, \theta_1 = (2(13))^{-1} = 0.038462, \\ q(\theta_0) &= (0.041667/\pi)^{1/2} e^{-1/(4(0.041667))} = 0.00028548, \\ q(\theta_1) &= (0.038462/\pi)^{1/2} e^{-1/(4(0.038462))} = 0.00016636, \end{aligned}$$

and then utilizing (3.9), we can express

$$\begin{aligned} Z_{ij} = \log \frac{f(X_{ij};\theta_1)}{f(X_{ij};\theta_0)} &= \log(\tfrac{0.00016636}{0.00028548}) + (0.038462 - 0.041667)(-X_{ij}^2) \\ &= -0.54002 + 0.003205 X_{ij}^2, \end{aligned}$$

where X_{ij} is the j^{th} observation in the i^{th} group. Let us denote

$$T_{k,\mathbf{M}^{(k)}} = -0.54002 M_{(k)} + 0.003205 \textstyle\sum_{i=1}^{k} \sum_{j=1}^{M_i} X_{ij}^2, k \geq 1. \tag{5.5}$$

Now, with $a = -b = \log(0.05/(1-0.05)) = -2.9444$ we see that (3.10) reduces to the following stopping variable:

$$\begin{gathered} K = \inf\left\{k \geq 1 : T_{k,\mathbf{M}^{(k)}} \notin (-2.9444, 2.9444)\right\} \\ \text{whereas we accept } H_0 \ (H_1) \text{ if and only if} \\ T_{k,\mathbf{M}^{(k)}} \leq -2.9444 \ (\geq 2.9444) \text{ when sampling terminates.} \end{gathered} \tag{5.6}$$

Table 5.3 illustrates how the methodology (5.6) is implemented in practice. We generated a Poisson random variate with mean 3 and it was 5 so that we selected 5 pages from the 300 page manuscript using SRSWOR and recorded the number of typographic errors. This is step #1 (that is, $k = 1$) given in the first block of the Table 5.3. We provide the number of typos on each page and the sum of squares (474) of these observations and $T_{1,\mathbf{M}^{(1)}} (= -1.1809)$. But, since $T_{1,\mathbf{M}^{(1)}}$ was between a and b, we moved to the next step ($k = 2$). We again generated a Poisson random variate with mean 3 and it happened to be 5 so that we selected 5 new pages from the 300 page manuscript using SRSWOR. This step is presented in the second block of Table 5.3. We note that at the step $k = 2$ and we terminated sampling and accepted H_0.

Table 5.3. Implementation #1 of the MSPRT (5.6) with
$\alpha = \beta = 0.05, a = -b = -2.9444$

Step #k	M_k	$M_{(k)}$	$\sum_{i=1}^{k}\sum_{j=1}^{M_i} X_{ij}^2$	$T_{k,\mathbf{M}^{(k)}}$	Continue or Terminate?
1	**Page Numbers (Number of Typos)**				
	207(8), 135(8), 213(9), 95(11), 181(12)				
	5	5	474	-1.1809	Continue Sampling
2	**Page Numbers (Number of Typos)**				
	87(7), 200(8), 162(6), 139(8), 198(8)				
	5	10	751	-2.9932	Terminate Sampling
Decision: With $k = 2$, we accept H_0.					

In Table 5.4, the first row corresponds to the summary statistics obtained from implementation #1 (that is, from Table 5.1). Next, we repeated the same process two more times by sampling from the same manuscript under SRSWOR. We successively required 3 and 5 steps in order to terminate sampling whereas the decision was to accept H_0 at each termination after observing 8 and 24($= M_{(k)}$, TSS) pages with a total of 79 and 254($= \sum_{i=1}^{5}\sum_{j=1}^{M_i} X_{ij}$) typographical errors. If we had repeated the process many more times from the same manuscript, surely we would have seen some instances where we would decide to accept H_1 instead of accepting H_0 at termination!

Table 5.4. Three Independent Implementations of the MSPRT (5.6)

$\alpha = \beta = 0.05, a = -b = -2.9444$

Implement # K	M_1	M_2	M_3	M_4	M_5	# Typos $M_{(K)}$	$\sum_{i=1}^{K}\sum_{j=1}^{M_i} X_{ij}$ $T_{K,\mathbf{M}^{(K)}}$ Terminal Decision
Implement #1 2	5	5	-	-	-	85 10	751 -2.9932 Accept H_0
Implement #2 3	2	5	1	-	-	79 8	887 -3.0974 Accept H_0
Implement #3 5	7	3	5	5	4	254 24	2918 -3.6083 Accept H_0

APPENDIX A: PROOF OF THEOREM 2.1

Let us consider $\psi^*(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k})$, another level α test function. Now, based on the classical Neyman-Pearson Lemma we can claim that

$$\begin{aligned}&\mathbf{E}_{\theta_1}\left[\psi(X_{11},\ldots,X_{1m_1},\ldots,X_{k1},\ldots,X_{km_k})\right]\\ &\qquad\geq \mathbf{E}_{\theta_1}\left[\psi^*(X_{11},\ldots,X_{1m_1},\ldots,X_{k1},\ldots,X_{km_k})\right]\end{aligned}$$

for *all* fixed choices $\mathbf{M}^{(k)} = \mathbf{m}^{(k)} \equiv (m_1,\ldots,m_k)$. Now, the power associated with the test function described in (2.4) can be expressed as

$$\begin{aligned}
&\mathbf{E}_{\theta_1}\left[\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k})\right]\\
&\quad= \mathcal{P}\left[\psi(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k}) = 1 \text{ when } H_1 \text{ is true}\right]\\
&\quad= \mathcal{P}\left[\frac{\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_1)}{\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_0)} > c_{\mathbf{M}^{(k)}} \text{ when } H_1 \text{ is true}\right]\\
&\quad= \sum_{\text{all } \mathbf{m}^{(k)}} \mathcal{P}_{\theta_1}\left[\frac{\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_1)}{\Pi_{i=1}^{k}\Pi_{j=1}^{M_i} f(X_{ij};\theta_0)} > c_{\mathbf{M}^{(k)}} \middle| \mathbf{M}^{(k)} = \mathbf{m}^{(k)}\right] \mathcal{P}_{\theta_1}\left[\mathbf{M}^{(k)} = \mathbf{m}^{(k)}\right]\\
&\quad= \sum_{\text{all } \mathbf{m}^{(k)}} \mathcal{P}_{\theta_1}\left[\frac{\Pi_{i=1}^{k}\Pi_{j=1}^{m_i} f(X_{ij};\theta_1)}{\Pi_{i=1}^{k}\Pi_{j=1}^{m_i} f(X_{ij};\theta_0)} > c_{\mathbf{m}^{(k)}} \middle| \mathbf{M}^{(k)} = \mathbf{m}^{(k)}\right] \mathcal{P}_{\theta_1}\left[\mathbf{M}^{(k)} = \mathbf{m}^{(k)}\right]\\
&\quad= \sum_{\text{all } \mathbf{m}^{(k)}} \mathcal{P}_{\theta_1}\left[\frac{\Pi_{i=1}^{k}\Pi_{j=1}^{m_i} f(X_{ij};\theta_1)}{\Pi_{i=1}^{k}\Pi_{j=1}^{m_i} f(X_{ij};\theta_0)} > c_{\mathbf{m}^{(k)}}\right] \mathcal{P}_{\theta_1}\left[\mathbf{M}^{(k)} = \mathbf{m}^{(k)}\right] \text{ by assumption A.2}\\
&\quad\geq \sum_{\text{all } \mathbf{m}^{(k)}} \mathcal{P}\left[\psi^*(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k}) = 1 \text{ when } H_1 \text{ is true}\right]\\
&\qquad\qquad \times \mathcal{P}_{\theta_1}\left[\mathbf{M}^{(k)} = \mathbf{m}^{(k)}\right]\\
&\quad= \mathbf{E}_{\theta_1}\left[\psi^*(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k})\right],
\end{aligned} \tag{A.1}$$

which is the power associated with the test function

$$\psi^*(X_{11},\ldots,X_{1M_1},\ldots,X_{k1},\ldots,X_{kM_k}).$$

The proof is now complete. ■

APPENDIX B: PROOF OF THEOREM 3.1

The following proof is an adaptation of Stein (1946). Since $\mathcal{P}(Z = 0) < 1$, there exists $\varepsilon > 0$ such that $\mathcal{P}(Z > \varepsilon) > 0$ or $\mathcal{P}(Z < -\varepsilon) > 0$. We assume that $\mathcal{P}(Z > \varepsilon) > 0$ and fix some $\varepsilon > 0$ such that $\mathcal{P}(Z > \varepsilon) \equiv \delta > 0$. Let n be the smallest positive integer such that $n\varepsilon > b - a$. Now, for a given k, we choose that positive integer q for which $qn \leq k < (q+1)n$ and then we have $\mathcal{P}(K > k) \leq \mathcal{P}(K > qn)$ where

$$\begin{aligned}
&\mathcal{P}(K > qn)\\
&\quad= \mathbf{E}_{\mathbf{M}}\left[\mathcal{P}\left\{a < \textstyle\sum_{l=1}^{i} S_{l,m_l} < b \text{ for all } i = 1, \ldots, qn | \mathbf{M} = \mathbf{m}\right\}\right]\\
&\quad\leq \mathbf{E}_{\mathbf{M}}\left[\mathcal{P}\left\{a < \textstyle\sum_{l=1}^{i} S_{l,m_l} < b \text{ for all } i = 1, \ldots, (q-1)n\right.\right.\\
&\qquad\qquad\left.\left.\cap \left|\textstyle\sum_{l=(q-1)n+1}^{qn} S_{l,m_l}\right| < b - a | \mathbf{M} = \mathbf{m}\right\}\right]\\
&\quad= \mathbf{E}_{\mathbf{M}}\left[\mathcal{P}\left\{K > (q-1)n \cap \left|\textstyle\sum_{l=(q-1)n+1}^{qn} S_{l,m_l}\right| < b - a | \mathbf{M} = \mathbf{m}\right\}\right].
\end{aligned} \tag{B.1}$$

Next, proceeding recursively, from (B.1) we can write

$$\begin{aligned}
&\mathcal{P}(K > qn)\\
&\quad\leq \mathbf{E}_{\mathbf{M}}\left[\mathcal{P}\left\{\cap_{l=1}^{q} \left|\textstyle\sum_{l=(q-l)n+1}^{(q-l+1)n} S_{l,M_l}\right| < b - a | \mathbf{M}\right\}\right]\\
&\quad= \mathbf{E}_{\mathbf{M}}\left[\left(\mathcal{P}\left\{\left|\textstyle\sum_{l=1}^{n} S_{l,M_l}\right| < b - a | \mathbf{M}\right\}\right)^q\right]\\
&\quad\leq \left(\mathcal{P}\left\{\textstyle\sum_{l=1}^{n} S_{l,M_l} < b - a\right\}\right)^q.
\end{aligned} \tag{B.2}$$

Now, let us denote $\eta = \mathbf{E}\left[\delta^{M_1}\right], c = (1 - \eta^n)^{-1}, r = (1 - \eta^n)^{1/n}$, and we have

$$\begin{aligned}
\mathcal{P}\left\{\textstyle\sum_{l=1}^{n} S_{l,M_l} \geq b - a\right\} &\geq \mathcal{P}\left\{\textstyle\sum_{l=1}^{n} S_{l,M_l} \geq n\varepsilon\right\} = \Pi_{l=1}^{n} \mathcal{P}\left\{S_{l,M_l} \geq \varepsilon\right\}\\
&= \Pi_{l=1}^{n} \mathbf{E}\left[\delta^{M_l}\right] = \eta^n,
\end{aligned} \tag{B.3}$$

so that (B.2) leads to

$$\begin{aligned}
&\mathcal{P}(K > qn) \leq (1 - \eta^n)^q\\
\Rightarrow\ &\mathcal{P}(K > k) \leq (1 - \eta^n)^q = cr^{(q+1)n} \leq cr^k.
\end{aligned} \tag{B.4}$$

This completes the derivation when $\mathcal{P}(Z > \varepsilon) > 0$ for some $\varepsilon > 0$. In the case when $\mathcal{P}(Z < -\varepsilon) > 0$ for some $\varepsilon > 0$, in the preceding derivation one would simply replace Z with $-Z$. The proof is now complete. ■

REFERENCES

1. Abraham, D. (2004). Passive acoustic detection of marine mammals using Page's test. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds., Marcel Dekker, New York 1-18.

2. Bahadur, R.R. (1958). A note on the fundamental identity of sequential analysis. *Ann. Math. Statis.* **29** 534-543.

3. Baron, M.I. (2004). Sequential methods for multistate processes. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds., Marcel Dekker, New York 53-67.

4. Bechhofer, R.E., Kiefer, J. and Sobel, M. (1968). *Sequential Identification and Ranking Procedures.* Univ. Chicago Press.

5. Chang, Y.C.I. and Martinsek, A.T. (2004). Sequential approaches to data mining. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds., Marcel Dekker, New York 85-103.

6. Ghosh, B.K. (1970). *Sequential Tests of Statistical Hypotheses.* Addison-Welsley, Reading.

7. Ghosh, M., Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation.* Wiley. New York.

8. Govindarajulu, Z. (1981). *The Sequential Statistical Analysis.* American Sciences Press, Columbus.

9. Govindarajulu, Z. (2004). *Sequential Statistics*. World Scientific Publishing, Singapore.

10. Lai, T.L. (2004a). Likelihood ratio identities and their applications to sequential analysis (with discussions). *Sequential Anal.* **23** 467-556.

11. Lai, T.L. (2004b). Interim and terminal analyses of clinical trials with failure-time endpoints and related group sequential designs. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds., Marcel Dekker, New York 193-218.

12. Lehmann, E.L. (1986). *Testing Statistical Hypotheses* 2nd ed. Wiley, New York.

13. Li, X. Rong and Solanky, T.K.S. (2004). Applications of sequential tests to target tracking by multiple models. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds., Marcel Dekker, New York 219-247.

14. Mukhopadhyay, N. and Cicconetti, G. (2004). Applications of sequentially estimating the mean in a normal distribution having equal mean and variance. *Sequential Anal.* **23** 625-665.

15. Mukhopadhyay, N., Son, M.S. and Ko, Y.C. (2004). A new two-stage sampling design for estimating the maximum average time to flower. *J. Agric. Biol. and Environ. Stat.* **9** 479-514.

16. Sen, P.K. (1981). *Sequential Nonparametrics*. Wiley, New York.

17. Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag. New York.

18. STEIN, C. (1946). A note on cumulative sums. *Ann. Math. Statis.* **17** 498-499.

19. TARTAKOVSKY, A.G. and VEERAVALLI, V.V. (2004). Change-point detection in multichannel and distributed Systems. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds., Marcel Dekker, New York 339-370.

20. WALD, A. (1947). *Sequential Analysis.* Wiley, New York.

21. WHITEHEAD, J. (1997). *The Design and Analysis of Sequential Clinical Trials* 2^{nd} ed. Wiley, Chichester.

22. WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* CBMS Lecture Notes #39, SIAM Publications, Philadelphia.

23. YOUNG, L.J. (2004). Sequential testing in agricultural sciences. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds. Marcel Dekker, New York 380-410.

24. ZACKS, S. and ROGATKO, A. (2004). Bayesian sequential procedures for ordering genes. In *Applied Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S., eds. Marcel Dekker, New York 411-432.

NITIS MUKHOPADHYAY
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
215 GLENBROOK ROAD, U 4120
STORRS, CT 06269-4120, U.S.A.
E-MAIL: nitis.mukhopadhyay@uconn.edu

BASIL M. DE SILVA
SCHOOL OF MATHEMATICAL & GEOSPATIAL SCIENCES
RMIT UNIVERSITY, GPO BOX 2476V
MELBOURNE, VIC 3001, AUSTRALIA
E-MAIL: desilva@rmit.edu.au