<table>
<thead>
<tr>
<th>Title</th>
<th>Some Thoughts in Sequential Two Sample Problems with Data Dependent Allocation Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Takahashi, Hajime</td>
</tr>
<tr>
<td>Citation</td>
<td>Hitotsubashi Journal of Economics, 28(2): 173-181</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1987-12</td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Text Version</td>
<td>publisher</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://doi.org/10.15057/7851">http://doi.org/10.15057/7851</a></td>
</tr>
</tbody>
</table>
SOME THOUGHTS IN SEQUENTIAL TWO SAMPLE PROBLEMS WITH DATA DEPENDENT ALLOCATION RULE

HAJIME TAKAHASHI

I. Introduction

Let \( x_1, x_2, \ldots, x_m, \ldots \) and \( y_1, y_2, \ldots, y_n, \ldots \) be independent normally distributed random variables with means \( \mu_1, \mu_2 \) and variances \( \sigma_1^2, \sigma_2^2 \) (\( m, n = 1, 2, \ldots \)). A sequential decision problem of testing \( H_1: \delta > 0 \) against \( H_2: \delta < 0 \), \( \delta = \mu_1 - \mu_2 \) with data dependent allocation rule has been considered by many authors including Robbins and Siegmund (1974), Louis (1975), Hayre (1979) and Hayre and Gittins (1981). The problem is invariant with respect to the common change of location for both \( x \) and \( y \), it is reasonable to restrict to the invariant procedure. So after having observed \( x_1, \ldots, x_m, \) and \( y_1, \ldots, y_n, \) we shall consider

\[
Z_{m,n} = \frac{mn}{\sigma_2^2 m + \sigma_1^2 n} (\bar{x}_m - \bar{y}_n), \quad \bar{x}_m = \frac{1}{m} \sum_{i=1}^{m} x_i, \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i
\]

[Robbins and Siegmund (1974), Hayre and Gittins (1981)]. By simple algebra it follows that

\[
E \{ Z_{m,n} \} = \frac{mn(\sigma_2^2 m + \sigma_1^2 n)}{\sigma_2^2 m + \sigma_1^2 n} \delta
\]

and

\[
\text{Var}_\delta \{ Z_{m,n} \} = \frac{mn(\sigma_2^2 m + \sigma_1^2 n)}{\sigma_2^2 m + \sigma_1^2 n}.
\]

Now by applying Wald’s SPRT to the statistic \( Z_{m,n} \), Robbins and Siegmund (1974) proved that (to the extent that we disregard the excess over the boundaries) the error probabilities and \( E \{ MN/\sigma_2^2 M + \sigma_1^2 N \} \) are independent of the allocation rules used (here \( (M,N) \) denotes 2-dimensional stopping time), provided they depend on previous \( x \)'s and \( y \)'s only through their differences. Related to this result, they have proved the following technical lemma

**Lemma 1.** Let \( W(t) \) be a Brownian motion with drift \( \delta \) and variance \( 1 \) per unit time. Let \( \tau_{m,n} = mn(\sigma_2^2 m + \sigma_1^2 n) \). Then for any allocation rules which depend on \( x \)'s and \( y \)'s only through the \( Z_{i,j} \)'s, the joint distribution of \( \{ Z_{m,n}, m,n = 1, 2, \ldots \} \) is equal to that of \( \{ W(\tau_{m,n}), m,n = 1, 2, \ldots \} \).

Since Robbins and Siegmund (1974) considers the two-population analogue of Wald’s SPRT, their result shares the same drawback as SPRT; the test is open ended. To remedy this, Hayre (1979) considered a sequential decision problem of choosing one of three hypo-
theses, \( H_0: \delta = 0 \), \( H_1: \delta > 0 \) and \( H_2: \delta < 0 \). He proposed to use the boundary suggested by Armitage (1957), which is a version of truncated SPRT. Continuing in this direction, Siegmund (1985, Ch. 6) applied the boundary for the repeated significance test of Armitage (1975). To be more specific, let \( b > 0 \) and \( 0 \leq K_0 < K \) are given constant. We shall let

\[
(M, N) = \text{first } \{(m, n): \tau_{m,n} > K_0, \quad |Z_{m,n}| > b \sqrt{\tau_{m,n}} \}.
\]

We stop sampling at \( \min \{\tau_{M,N}, K\} \) and

1. Accept \( H_1 \) if \( Z_{M,N} > b \sqrt{\tau_{M,N}} \) and \( \tau_{M,N} \leq K \)
2. Accept \( H_2 \) if \( Z_{M,N} < -b \sqrt{\tau_{M,N}} \) and \( \tau_{M,N} \leq K \)
3. Accept \( H_0 \) if \( Z_{M,N} > K \).

Knowing that the error probabilities and the expected sample size \( E\{\tau_{M,N} \land K\} \) are approximately the same for all invariant allocation rules, Siegmund (1985) has calculated these characteristics under the pairwise allocation rule.

Now the extension to the regression model may be of interest. Let \( x_m = \mu_1 + \beta \xi_m + \varepsilon_{x_m}, m=1, 2, \ldots \) and \( y_n = \mu_2 + \beta \eta_n + \varepsilon_{y_n}, n=1, 2, \ldots \), where \( \mu_1 \) and \( \mu_2 \) are unknown intercepts and \( \beta \) denotes an unknown common slope, \( \{\xi_m, m \geq 1\} \) and \( \{\eta_n, n \geq 1\} \) are sequences of known constants and \( \{\varepsilon_{x_m}, m \geq 1\} \) and \( \{\varepsilon_{y_n}, n \geq 1\} \) are independent and normally distributed random variables with means 0 and known variances \( \text{Var} \varepsilon_{x_m} = \sigma_1^2 \) and \( \text{Var} \varepsilon_{y_n} = \sigma_2^2 \). By long and tedious algebra, after having observed \((x_1, \xi_1), \ldots, (x_m, \xi_m)\) and \((y_1, \eta_1), \ldots, (y_n, \eta_n)\), our invariant test statistic is

\[
\hat{Z}_{m,n} = \frac{mn}{\sigma_2^2 + \nu_1 \sigma_1^2} [\bar{x}_m - \bar{y}_n - b_{m,n}(\bar{\xi}_m - \bar{\eta}_n)]
\]

where

\[
\bar{x}_m = \frac{1}{m} \sum_{i=1}^{m} x_i, \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \bar{\xi}_m = \frac{1}{m} \sum_{i=1}^{m} \xi_i, \quad \bar{\eta}_n = \frac{1}{n} \sum_{i=1}^{n} \eta_i
\]

and

\[
b_{m,n} = \left( \sigma_2^2 \sum_{i=1}^{m} \xi_i^2 + \sigma_1^2 \sum_{i=1}^{n} \eta_i^2 \right) \left( m \sigma_2^2 + n \sigma_1^2 \right) - \left( m \sigma_2^2 \bar{\xi}_m + n \sigma_1^2 \bar{\eta}_n \right) \left( m \sigma_2^2 \bar{\xi}_m + n \sigma_1^2 \bar{\eta}_n \right)
\]

\[
\div \left( \left( \sigma_2^2 \sum_{i=1}^{m} \xi_i^2 + \sigma_1^2 \sum_{i=1}^{n} \eta_i^2 \right) \left( m \sigma_2^2 + n \sigma_1^2 \right) - \left( m \sigma_2^2 \bar{\xi}_m + n \sigma_1^2 \bar{\eta}_n \right)^2 \right)
\]

(See Takahashi (1977) for the special case \( \sigma_1^2 = \sigma_2^2 = 1 \).

Note that

\[
E[\hat{Z}_{m,n}|\xi, \eta] = \frac{mn}{\sigma_2^2 + \nu_1 \sigma_1^2} C_{m,n} \delta, \quad \text{and}
\]

\[
\text{Var}[\hat{Z}_{m,n}|\xi, \eta] = \frac{mn}{\sigma_2^2 + \nu_1 \sigma_1^2} C_{m,n}
\]

where \( \delta = \mu_1 - \mu_2 \) and

\[
C_{m,n} = \sigma_2^2 \sum_{i=1}^{m} \xi_i^2 + \sigma_1^2 \sum_{i=1}^{n} \eta_i^2 - \left( m \sigma_2^2 \bar{\xi}_m + n \sigma_1^2 \bar{\eta}_n \right)^2
\]

\[
\div \left( \sigma_2^2 \sum_{i=1}^{m} \xi_i^2 + \sigma_1^2 \sum_{i=1}^{n} \eta_i^2 - \left( m \sigma_2^2 \bar{\xi}_m + n \sigma_1^2 \bar{\eta}_n \right)^2 \right) \left( m \sigma_2^2 + n \sigma_1^2 \right)
\]
Since $C_{m,n}$ is increasing in $m$ and $n$ [Lai and Robbins (1977)], we have the analogous result of Lemma 1.

**Lemma 2.** Given $\xi$’s and $\eta$’s, the joint distribution of $\{Z_{m,n}, m,n=1,2,\ldots\}$ is the same as that of $\{W(\tau_{m,n}), m,n=1,2,\ldots\}$, where $\tau_{m,n}=\tau_{m,n}C_{m,n}$.

By the simple algebra $0 \leq C_{m,n} \leq 1$ and $C_{m,n}=1$ if $\xi_m=\eta_m$ and they are not degenerated. It follows that we can get the full efficiency (with respect to the simple two sample problem) by taking the mean of $\xi$’s and $\eta$’s about the same but not degenerated to the same point.

Hence the sequential decision problem of choosing which of these three hypotheses $H_0: \delta = 0$, $H_1: \delta > 0$ and $H_2: \delta < 0$ may be carried out exactly in the same manner as above up to the Brownian approximation.

### II. Allocation Rule

Numerous allocation rules have been proposed since the pioneering work by Louis (1972). Among them the one proposed by Hayre (1979) is of interest. Given $\delta$ the cost of taking one observation on $x$ is $g(\delta) > 0$ and the cost of taking $y$ is $h(\delta) > 0$ for which $h(\delta) - g(\delta)$ has the same sign of $\delta$. A specific example considered below is

$$g(\delta) = h(-\delta) = \begin{cases} 1 & \delta > 0 \\ 1 + d|\delta| & \delta < 0 \end{cases}$$

The optimal allocation rule is defined to be one which minimizes the risk

$$R(\delta) = g(\delta)E_\delta(M) + h(\delta)E_\delta(N).$$

The answer to this question is in *Lemma 3.* [Hayre (1979)]. Let $w(\delta) = [h(\delta)/g(\delta)]^{1/2}$. For the test (3), for any invariant allocation rule

$$R(\delta) \geq (\sigma_1 g^{1/2} + \sigma_2 h^{1/2})^2 E_\delta \{\tau_{m,n}\}$$

and equality holds if and only if

$$P_\delta \{M/N = (\sigma_1/\sigma_2)w(\delta)\} = 1.$$ 

Since $\delta$ is unknown to us, the suggested allocation rule would be:

**Allocation Rule [HA].** After having observed $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$, if $|Z_{m,n}| < b \sqrt{\tau_{m,n}}$, then we choose $x_{m+1}$ next if and only if

$$m/n \leq (\sigma_1/\sigma_2)w(\hat{\delta}_{m,n}), \quad \hat{\delta}_{m,n} = Z_{m,n}/\tau_{m,n}.$$ 

The law of large number implies that (9) will hold approximately in large samples, then it is not difficult to see

$$E_\delta \{M\} \approx \sigma_1 (\sigma_1 + \sigma_2 w(\hat{\delta})) E_\delta \{\tau_{M,N}\}$$

and equality holds if and only if (9) holds. Hence, we may use the same sampling...
TABLE 1
Simple two sample problems with known variance $\sigma_1^2=\sigma_2^2=1$
$b=2.95, K=49, K_0=0$
Pairwise sampling = 1st raw, Hayre's rule = 2nd raw ($d=20$)

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$H_2$</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>$E(M)$</th>
<th>$E(N)$</th>
<th>$E(M+N)$</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.02</td>
<td>0.97</td>
<td>0.01</td>
<td>97.5</td>
<td>97.5</td>
<td>195.0</td>
<td>195.0</td>
</tr>
<tr>
<td>0.2</td>
<td>0.00</td>
<td>0.95</td>
<td>0.02</td>
<td>100.5</td>
<td>105.0</td>
<td>205.5</td>
<td>205.5</td>
</tr>
<tr>
<td>0.4</td>
<td>0.00</td>
<td>0.87</td>
<td>0.13</td>
<td>93.4</td>
<td>93.4</td>
<td>186.8</td>
<td>560.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.00</td>
<td>0.84</td>
<td>0.16</td>
<td>143.4</td>
<td>72.2</td>
<td>215.6</td>
<td>504.4</td>
</tr>
<tr>
<td>0.8</td>
<td>0.00</td>
<td>0.42</td>
<td>0.58</td>
<td>71.6</td>
<td>71.5</td>
<td>143.1</td>
<td>714.5</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>0.45</td>
<td>0.55</td>
<td>138.8</td>
<td>50.0</td>
<td>188.8</td>
<td>588.8</td>
</tr>
</tbody>
</table>

In Tables 1–6 all the values are obtained from the average of 400 repetition. We have used subroutine RANN2 in SSL II and the system FACOM M-360 at Hitotsubashi University.

TABLE 2
Regression case with $\beta=0.5$, known variance $\sigma_1^2=\sigma_2^2=1$
$b=2.95, K=49, K_0=0$
Pairwise sampling = 1st raw, Hayre's rule = 2nd raw ($d=20$)

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$H_2$</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>$E(M)$</th>
<th>$E(N)$</th>
<th>$E(M+N)$</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.02</td>
<td>0.97</td>
<td>0.01</td>
<td>97.5</td>
<td>97.5</td>
<td>195.0</td>
<td>195.0</td>
</tr>
<tr>
<td>0.2</td>
<td>0.00</td>
<td>0.95</td>
<td>0.02</td>
<td>103.1</td>
<td>106.2</td>
<td>209.3</td>
<td>209.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.00</td>
<td>0.87</td>
<td>0.13</td>
<td>93.4</td>
<td>93.4</td>
<td>186.8</td>
<td>560.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.00</td>
<td>0.86</td>
<td>0.14</td>
<td>143.9</td>
<td>73.2</td>
<td>217.1</td>
<td>509.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.00</td>
<td>0.42</td>
<td>0.58</td>
<td>71.5</td>
<td>71.5</td>
<td>143.1</td>
<td>715.4</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>0.46</td>
<td>0.54</td>
<td>140.0</td>
<td>50.5</td>
<td>190.5</td>
<td>594.4</td>
</tr>
</tbody>
</table>

rule [HA] here with $\delta_{m,n}$ replaced by $\tilde{Z}_{m,n}/\tilde{\tau}_{m,n}$. If we may choose $\xi$'s and $\gamma$'s for which $C_{x,\tau,\alpha}(0,1)$, we can show

\[ E_{\alpha} [M] \approx \alpha^{-1} \sigma_1 (s_1 + s_2 w(\delta)) E_{\alpha} [\tilde{\tau}_{x,\alpha}] \]
\[ E_{\alpha} [N] \approx \alpha^{-1} \sigma_2 (s_2 + s_1 w^{-1}(\delta)) E_{\alpha} [\tilde{\tau}_{x,\alpha}] \]
as $b \to \infty$. Some of the simulation results are presented in Table 1 and 2.

III. Unknown Variance

When variances are not known to us, we may replace all $\sigma_i^2$'s by their unbiased esti-
SOME THOUGHTS IN SEQUENTIAL TWO SAMPLE PROBLEMS

Table 3
Simple two sample problems with unknown variance
b=2.95, K=49 (M,N≥10)
Pairwise sampling = 1st raw, Heyre's rule = 2nd raw (d=20)

<table>
<thead>
<tr>
<th>δ</th>
<th>H₀</th>
<th>H₁</th>
<th>E(M)</th>
<th>E(N)</th>
<th>E(M+N)</th>
<th>Risk</th>
<th>σ² / σ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.01</td>
<td>0.98</td>
<td>0.01</td>
<td>109.5</td>
<td>109.5</td>
<td>219.0</td>
<td>219.0</td>
</tr>
<tr>
<td>0.2</td>
<td>0.01</td>
<td>0.88</td>
<td>0.11</td>
<td>105.5</td>
<td>105.5</td>
<td>211.0</td>
<td>633.0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.00</td>
<td>0.86</td>
<td>0.56</td>
<td>83.9</td>
<td>83.9</td>
<td>167.8</td>
<td>839.0</td>
</tr>
<tr>
<td>0.6</td>
<td>0.00</td>
<td>0.41</td>
<td>0.96</td>
<td>51.2</td>
<td>51.2</td>
<td>102.4</td>
<td>716.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.00</td>
<td>0.04</td>
<td>1.00</td>
<td>34.2</td>
<td>34.2</td>
<td>68.4</td>
<td>615.6</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>49.6</td>
<td>13.0</td>
<td>62.6</td>
<td>323.5</td>
</tr>
</tbody>
</table>

mator δ₀² (i=1, 2). By the strong law of large numbers, δ₀² converges strongly to σ² as b→∞. Hence at least asymptotically, the modified procedure possesses the same properties discussed in the previous section. Probabilistically the problem is too complicated and we do not discuss in detail here. Some of their small sample properties are obtained by Monte Carlo studies. We first use unbiased estimator of σ² (i=1, 2) in both simple and regression cases. In the simple two sample problem, it turns out that usual unbiased estimator seriously underestimate the true σ² for all values of δ. We then used the estimator proposed by Lai et al. (1975). The results are satisfactory as shown in Table 3. As to the regression case, the usual unbiased estimator behaves fairly well (Table 4). We do not know why these phenomena occurs.

IV. Estimation after Testing

In this section we shall suppose that σ²₁=σ²₂=1 and consider testing sequentially H₀:δ=0 against δ≠0 with data dependent allocation rule [HA]. When H₀ is rejected, we are interested in estimating the value of δ based on the data collected so far. The problem of estimating δ after sequential testing has been considered by Siegmund (1978), Takahashi (1987), Woodroofe and Keener (1987), where they have utilized the non-linear renewal theory.
to approximate the excess over the boundary. In this section, however, we shall use the Brownian motion $W(t)$ to approximate the original process $Z_{m,n}$, which is tantamount to disregarding the excess over the boundary. As has been pointed out by several authors, Brownian approximation usually does not give us a good numerical accuracy [Siegmund (1985), Ch. 3], which is the consequence of disregarding the excess over the boundary. But in this case, the situation is much better. After having observed $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$, if we take one more $x$ sample ($y$ sample), then the average increment would be $[a^2/(m+n)(m+n+1)]\delta$. This is substantially smaller than $\delta$ when $m \gg n$ ($n \gg m$), which may explain the good approximation we have below (see Table 5).

As in the previous sections, we let $\{W(t), t \geq 0\}$ denote the Brownian motion with drift $\delta$ and variance 1 per unit time. For $b>0$ we shall define a stopping time

$$T = \inf \{t \geq 0; |W(t)| \geq b \sqrt{t}\}.
$$

Then the continuous time analogue of the repeated significance test for $\delta = 0$ would be

$$\text{Reject } H_0: \delta = 0 \quad \text{iff} \quad T \leq K
$$

for some given constant $K>0$.

If $T \leq K$, then the naive estimator for $\delta$ would be

$$\bar{X} = W(T)/T,$$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$H_1$</th>
<th>$H_0$</th>
<th>$H_2$</th>
<th>$E(M)$</th>
<th>$E(N)$</th>
<th>$E(M+N)$</th>
<th>Risk</th>
<th>$\sigma^2 / \sigma^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.02</td>
<td>0.97</td>
<td>0.01</td>
<td>114.3</td>
<td>114.3</td>
<td>228.6</td>
<td>228.6</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.94</td>
<td>0.03</td>
<td>120.4</td>
<td>121.6</td>
<td>242.0</td>
<td>242.0</td>
<td>1.16</td>
</tr>
<tr>
<td>0.2</td>
<td>0.01</td>
<td>0.87</td>
<td>0.12</td>
<td>109.6</td>
<td>109.6</td>
<td>219.2</td>
<td>657.4</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.85</td>
<td>0.15</td>
<td>166.6</td>
<td>83.8</td>
<td>250.4</td>
<td>585.5</td>
<td>1.20</td>
</tr>
<tr>
<td>0.4</td>
<td>0.00</td>
<td>0.44</td>
<td>0.56</td>
<td>84.6</td>
<td>84.6</td>
<td>169.2</td>
<td>845.9</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.44</td>
<td>0.56</td>
<td>164.1</td>
<td>58.4</td>
<td>222.5</td>
<td>689.9</td>
<td>1.18</td>
</tr>
<tr>
<td>0.6</td>
<td>0.00</td>
<td>0.05</td>
<td>0.95</td>
<td>52.4</td>
<td>52.4</td>
<td>104.8</td>
<td>734.1</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.04</td>
<td>0.96</td>
<td>97.0</td>
<td>29.5</td>
<td>126.5</td>
<td>480.0</td>
<td>1.11</td>
</tr>
<tr>
<td>0.8</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>29.4</td>
<td>29.4</td>
<td>58.8</td>
<td>528.8</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>62.6</td>
<td>18.4</td>
<td>81.0</td>
<td>375.6</td>
<td>1.10</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>20.9</td>
<td>20.9</td>
<td>41.8</td>
<td>458.8</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>40.6</td>
<td>13.6</td>
<td>54.2</td>
<td>326.4</td>
<td>1.03</td>
</tr>
</tbody>
</table>

If $T \leq K$, then the naive estimator for $\delta$ would be

$$\bar{X} = W(T)/T,$$
which, however, has a serious upward (downward) bias when \( \delta > 0 \) (\( \delta < 0 \)). In order to obtain the correction term, Siegmund (1978) utilizes our intuitive feelings that small value of the stopping time \( T \) are evidence in favor of large value of \( | \delta | \). On the other hand Takahashi (1987) considers the asymptotic expansion of the distribution of \( (W(T) - \delta T) / \sqrt{T} \) to get the correction terms. Although in the discrete time case these two approaches are based on completely different methods, they depend on the same technical result in this continuous time problem.

**Lemma 4** [Siegmund (1985), Ch. 4]. Let \( \delta > 0 \). Suppose \( b \to \infty \) \( k \to \infty \) such that \( b/\sqrt{k} = \delta_0 \) remain bounded away from 0 and \( \infty \). Then uniformly in \( \delta \) in the closed subintervals of \((0, \infty)\)

\[
P \{ T < k \} = 1 - \Phi[\sqrt{k}(\delta_0 - \delta)] + \left[ \phi[\sqrt{k}(\delta_0 - \delta)] \delta / \sqrt{k} \right] \{ 1 + o(1) \}
\]
as \( b \to \infty \).

**Table 5**
90\% confidence intervals of \( \delta \)
Simple two sample problem with Hayre's rule (\( d = 20 \))
\( b = 2.95, K = 49, K_0 = 0 \)

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>Native estimator</th>
<th>Siegmund's</th>
<th>Takahashi's</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.247</td>
<td>0.077, 0.707</td>
<td>-0.094, 0.398</td>
</tr>
<tr>
<td>0.4</td>
<td>0.097, 0.856</td>
<td>0.392</td>
<td>0.113, 0.634</td>
</tr>
<tr>
<td>0.6</td>
<td>0.125, 1.124</td>
<td>0.477</td>
<td>0.222, 0.934</td>
</tr>
<tr>
<td>0.8</td>
<td>0.155, 1.404</td>
<td>0.625</td>
<td>0.286, 1.195</td>
</tr>
<tr>
<td>1.0</td>
<td>0.189, 1.670</td>
<td>0.779</td>
<td>0.351, 1.473</td>
</tr>
<tr>
<td>1.2</td>
<td>0.223, 1.954</td>
<td>0.922</td>
<td>0.455, 1.777</td>
</tr>
<tr>
<td>1.378</td>
<td>1.089</td>
<td></td>
<td>1.116</td>
</tr>
</tbody>
</table>

**Table 6**
90\% confidence intervals of \( \delta \)
Simple two sample problem with Hayre's rule (\( d = 20 \))
\( b = 3.45, K = 148, K_0 = 0 \)

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>Native estimator</th>
<th>Siegmund's</th>
<th>Takahashi's</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.200</td>
<td>0.122, 0.450</td>
<td>0.039, 0.321</td>
</tr>
<tr>
<td>0.4</td>
<td>0.192, 0.705</td>
<td>0.286</td>
<td>0.180</td>
</tr>
<tr>
<td>0.6</td>
<td>0.270, 0.996</td>
<td>0.449</td>
<td>0.260, 0.658</td>
</tr>
<tr>
<td>0.8</td>
<td>0.356, 1.311</td>
<td>0.633</td>
<td>0.376, 0.960</td>
</tr>
<tr>
<td>1.0</td>
<td>0.441, 1.625</td>
<td>0.833</td>
<td>0.506, 1.275</td>
</tr>
<tr>
<td>1.2</td>
<td>0.504, 1.864</td>
<td>1.033</td>
<td>0.608, 1.586</td>
</tr>
<tr>
<td>1.346</td>
<td>1.184</td>
<td></td>
<td>0.709, 1.845</td>
</tr>
<tr>
<td>1.277</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Theorem 1. Under the same asymptotic relations of Lemma 4

\begin{equation}
\Pr \left( \left. \frac{W(T)-\delta T}{\sqrt{T}} \leq x \right\} = \Phi(x) - \frac{1}{b} \phi(x) + o(b^{-1}) \right)
\end{equation}

Proof. Let \( T_+ \) denote the one sided version of \( T \)

\[ T_+ = \inf\{t \geq 0; W(t) \geq b \sqrt{t}\} . \]

It is easily seen that for any \( \delta > 0 \)

\[ \Pr \{ T_+ > T \} = o(b^{-2}) \quad b \to \infty . \]

Hence,

\[ (W(T_+)-\delta T_+)/\sqrt{T_+} = b-\delta \sqrt{T_+} . \]

Thus

\[ \Pr \left( \left. \frac{W(T)-\delta T}{\sqrt{T}} \leq x \right\} = \Pr \{ T > (b-x/\delta)^2 \} + o(b^{-2}) \right) \]

\[ = \Phi(x) - \frac{1}{b} \phi(x) + o(b^{-1}) \quad b \to \infty . \]

Corollary 1.

\begin{equation}
\Pr \left( \left. \frac{W(T)-\delta T}{\sqrt{T}} \leq x + \frac{1}{b} \right\} = \Phi(x) + o(b^{-1}) \right)
\end{equation}

as \( b \to \infty \).

From (18) we may define \((1-\alpha) \times 100\% \) confidence interval of \( \delta \) based on the data \( \{T, W(T)\} \),

\[ \left[ \frac{W(T)-z_{a/2}}{\sqrt{T}}, \frac{W(T)+z_{a/2}}{\sqrt{T}} \right] - 1/(b \sqrt{T}) \quad \text{on} \quad [T \leq K, W(T) \geq 0] \]

\[ \left[ \frac{W(K)-z_{a/2}}{\sqrt{K}}, \frac{W(K)+z_{a/2}}{\sqrt{K}} \right] \quad \text{on} \quad [T > K] \]

\[ \left[ \frac{W(T)-z_{a/2}}{\sqrt{T}}, \frac{W(T)+z_{a/2}}{\sqrt{T}} \right] + 1/(b \sqrt{T}) \quad \text{on} \quad [T \leq K, W(T) \leq 0] \]

where \( \Phi(z_\beta) = 1 - \beta \).

The numerical accuracy of (19) is compared with simulation results, Siegmund’s estimator, and the naive estimator \( W(T)/T \). On a whole Siegmund’s estimator (center of 90\% confidence interval) performs better than (19) \((W(T)/T - 1/(b \sqrt{T})\)). But the width of the intervals of Siegmund’s estimators are usually wider than (19). It suggests us to obtain higher order asymptotic expansion in (17), which will be considered in the next project. [cf. Takahashi (1987), Woodroofe and Keener (1987)].

REFERENCES