Use of a side effect as a covariate in a problem of sequential analysis

Martha Bilotti-Aliaga

Department of Statistics, University of Michigan, Ann Arbor, MI, USA

Received 14 April 1989; revised manuscript received 20 September 1991
Recommended by T.L. Lai

Abstract: A drug is administered sequentially to incoming patients. A response Y to treatment and a covariate X is measured (X might be a side effect). The experiment is stopped when the covariate falls outside some acceptable region. We study the effect that this optional stopping has on the significance level of the test and we found that this effect is surprisingly small in the examples considered. An approximation to the problem is found. This approximation does not depend on the distribution of the variable X but only on the correlation coefficient between X and Y.

AMS Subject Classification: Primary 62L10.

Key words and phrases: Estimation; clinical trials; backward induction; dynamic programming; binary binomial case; testing without batches; testing with batches; Wiener process.

1. Introduction and summary

Suppose that there is an infinite sequence \((X_i, Y_i), i=1,2,\ldots\), of covariate-response pairs which are independent with common unknown distribution \(H\), and that the parameter \(E(Y) = \theta\) is of primary interest. Let \(F_n = \sigma\{X_1, \ldots, X_n\}\) denote the smallest \(\sigma\)-algebra with respect to which \(X_1, \ldots, X_n\) are measurable. We sample sequentially according to a stopping rule \(f\) depending on the \(X_i\)’s.

Suppose that a one-sided hypothesis about \(\theta\) is of interest, say \(H_0: \theta \leq \theta_0\), and that \(H_0\) is to be rejected if

\[
\sqrt{n}(Y_n - \theta_0) / \hat{\sigma}_n > c
\]

at the time \(t=n\) of stopping, where \(Y_n\) is the sample mean, \(\hat{\sigma}_n^2, n \geq 1\), denotes a consistent sequence of estimates of the variance of \(Y\), and \(c\) is a constant. If we stop at time \(t\), then the actual significance level is
\[
\alpha_t = P_0 \left\{ \frac{\sqrt{n}(\bar{Y}_t - \theta_0)}{\hat{\sigma}_t} > c \right\},
\]

where \( P_0 \) denotes probability when \( \theta = \theta_0 \) for fixed values of any nuisance parameters.

We may determine the maximum possible effect of such optional stopping on the attained significance level as follows. Let

\[
Z_n = P_0 \left\{ \frac{\sqrt{n}(\bar{Y}_n - \theta_0)}{\hat{\sigma}_n} > c \mid X_1, \ldots, X_n \right\}.
\]

Then

\[
\alpha_t = E_0(Z_t),
\]

so we may determine the maximal effect of optional stopping on \( \alpha_t \) by attempting to maximize \( E_0(Z_t) \) by choice of \( t \).

We present numerical solutions to the optimal stopping problem in Section 2 for some special cases. The effect of optional stopping on the significance level is surprisingly small in the examples considered, an increase by less than a factor of two.

Patients may be measured individually, but treated in batches of different sizes. In many situations treating in batches could be a more economical way of treatment. When we consider this complication, from the data we obtain it seems that 'the size of the batch' has a modest effect.

In Section 3, we find an approximation to the hypothesis testing problem which does not depend on the distribution of the variable \( X \) but only on the correlation coefficient (\( \rho \)) between \( X \) and \( Y \). The approximation is thus found when we formulate the analogous problem in continuous time.

2. Numerical solutions

2.1. Introduction

We consider a numerical solution using the method of backward induction. The examples contain a number of important special cases and provide motivation for the general theoretical approach in Section 3.

We calculate the maximal effect on the significance level for testing the mean of a response variable \( Y \) (\( E(Y) = \theta \)) for a one sided null hypothesis \( H_0: \theta \leq \theta_0 \). The stopping time considered is defined with respect to a covariate \( X \).

2.2. Hypothesis testing

Let \( H \) be the unknown common joint distribution of the pairs of observable covariate-response random variables \((X_i, Y_i), i =1, 2, \ldots \). We suppose that \( Y \) has an
arbitrary distribution with unknown mean \( \theta = E(Y) \), and finite variance \( \sigma_0^2 = V(Y) \) which is known under \( H_0 \). We want to perform a one side sequential test for \( \theta \), say \( H_0: \theta \leq \theta_0 \).

**Lemma 2.1.** If the conditional expectation of \( Y \) given \( X \) is a linear function of \( X \), then

\[
E(Y_n - \theta \mid X_1, \ldots, X_n) = \delta(X_n - \mu)
\]

(2.1)

and, if the conditional variance \( \sigma^2 \) of \( Y \) given \( X \) is independent of \( X \), then

\[
V\{Y_n \mid X_1, \ldots, X_n\} = \frac{\sigma^2}{n}
\]

(2.2)

where \( \delta \) is a constant and \( \mu = E(X) \).

In the remainder of this section, the covariate variable \( X \) is assumed to be a dichotomous random variable taking the values 0 and 1 with probability \( \frac{1}{2} \) each. Then, the conditional expectation of \( Y \) given \( X \) is linear, say \( E(Y \mid X) = \delta X + \beta \). The conditional variance of \( Y \) given \( X \) is assumed to be independent of \( X \), say \( V(Y \mid X) = \sigma^2 < \infty \). By Lemma 2.1, the Central Limit Theorem suggests approximating \( Z_n \) by

\[
Z_n^* = 1 - \Phi \left( \frac{c\sigma_0 - \delta \sqrt{n}(X_n - \mu)}{\sigma} \right) = 1 - \Phi \left( \frac{c - \varrho \sqrt{n\sigma^{-1}_x}(X_n - \mu)}{\sqrt{1 - \varrho^2}} \right).
\]

When \( \varrho = 0 \) this reduces to the nominal value \( 1 - \Phi(c) \). We solve the approximate problem, in which \( Z_n \) is replaced by \( Z_n^* \). That is, we find

\[
V^* = \max_{i \in C} E_0(Z_i^*).
\]

Observe that \( Z_n^* = Z_n \) if the conditional distribution of \( Y \) given \( X \) is normal.

**2.3. The binomial example**

Consider the symmetric case in Table 1. Then \( E(Y) = \theta \), \( \text{corr}(X, Y) = \varrho \) and \( E(Y \mid X) = \theta + \varrho(X - \theta) \). For \( \theta = \frac{1}{2} \), \( V(Y \mid X) = \frac{1}{4}(1 - \varrho^2) \) that is, the conditional variance of \( Y \) given \( X \) is independent of \( X \), and the conditions of Lemma 2.1 are satisfied.

We may write the approximate payoff as

\[
Z_n^* = 1 - \Phi \left( \frac{c - 2 \sqrt{n\varrho(X_n - \frac{1}{2})}}{\sqrt{1 - \varrho^2}} \right).
\]

**Table 1**

<table>
<thead>
<tr>
<th>( X = 0 )</th>
<th>( X = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 0 )</td>
<td>((1 - \theta)^2 + \varrho\theta(1 - \theta))</td>
</tr>
<tr>
<td>( Y = 1 )</td>
<td>( \theta(1 - \theta) - \varrho\theta(1 - \theta) )</td>
</tr>
<tr>
<td></td>
<td>( 1 - \theta )</td>
</tr>
</tbody>
</table>
2.4. Dynamic programming

In this section we suppose that $X$ takes the values 0 and 1 with probability $\frac{1}{2}$ each and that the conditional variance of $Y$ given $X$ is independent of $X$. Then

$$Z_n^* = 1 - \phi \left\{ c - 2 \sqrt{n} \sigma (X_n - \frac{1}{2}) \right\} = \psi_n(S_n), \quad \text{say,}$$  

(2.3)

where $S_n = \sum_{i=1}^{n} X_i$.

Consider the restricted problem in which $t$ is required to be at most some specified $N$, and at least some specified $m$. The value for this restricted problem is

$$V_m^N = \sup_{t \in C_m^N} E_0(Z_t^*)$$

Table 2
Actual values for testing hypothesis (without batches)

<table>
<thead>
<tr>
<th>$m$</th>
<th>$N = 25, \ c = 1.96$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = 0.2$</td>
</tr>
<tr>
<td>1</td>
<td>0.0302563</td>
</tr>
<tr>
<td>2</td>
<td>0.0298616</td>
</tr>
<tr>
<td>3</td>
<td>0.0291243</td>
</tr>
<tr>
<td>4</td>
<td>0.0286438</td>
</tr>
<tr>
<td>5</td>
<td>0.0282968</td>
</tr>
<tr>
<td>6</td>
<td>0.0279349</td>
</tr>
<tr>
<td>7</td>
<td>0.02769</td>
</tr>
<tr>
<td>8</td>
<td>0.02742</td>
</tr>
<tr>
<td>9</td>
<td>0.027203</td>
</tr>
<tr>
<td>10</td>
<td>0.0269886</td>
</tr>
<tr>
<td>11</td>
<td>0.0267989</td>
</tr>
<tr>
<td>12</td>
<td>0.0266197</td>
</tr>
<tr>
<td>13</td>
<td>0.0264521</td>
</tr>
<tr>
<td>14</td>
<td>0.0262974</td>
</tr>
<tr>
<td>15</td>
<td>0.0261482</td>
</tr>
<tr>
<td>16</td>
<td>0.0260105</td>
</tr>
<tr>
<td>17</td>
<td>0.0258765</td>
</tr>
<tr>
<td>18</td>
<td>0.0257513</td>
</tr>
<tr>
<td>19</td>
<td>0.0256298</td>
</tr>
<tr>
<td>20</td>
<td>0.0255143</td>
</tr>
<tr>
<td>21</td>
<td>0.0254034</td>
</tr>
<tr>
<td>22</td>
<td>0.025295</td>
</tr>
<tr>
<td>23</td>
<td>0.0251935</td>
</tr>
<tr>
<td>24</td>
<td>0.0250925</td>
</tr>
<tr>
<td>25</td>
<td>0.0249974</td>
</tr>
</tbody>
</table>

Nominal | 0.0249979 | 0.0249979 | 0.0249979
Actual/nominal | 1.21 | 1.39 | 1.56
where

\[ C^N_m = \{ t; t \text{ is a stopping rule, } m \leq t \leq N \}. \]

To find the optimal solution for \( V^N_m \) we use the theorem which can be found in Chow, Robbins and Siegmund (1971, p. 50).

Let \( N \) be a fixed positive integer. Define successively \( \gamma^N_1, \gamma^N_2, \ldots, \gamma^N_m \) by setting

\[
\gamma^N_1 = Z^N_1 = \psi_N(S_N), \\
\gamma^N_n = \max \left( Z^N_n E_0(\gamma^N_{n+1} | F_n) \right) (n = N - 1, \ldots, m).
\]

For each \( n = m, m+1, \ldots, N \) let

\[
s^N_n = \inf \{ i > n; Z^N_i = \gamma^N_i \}.
\]

It follows from that theorem that \( s^N_m \) is optimal in \( C^N_m \).

In our problem, the recursion has a very simple form. Let \( k \) denote a typical value of \( S_n \), i.e., \( k = 0, 1, \ldots, n \), and let \( \gamma^N(N, k) = \psi_N(k) \) as defined in (2.5). Then

\[
\gamma^N(n, k) = \max \left\{ \psi_N(k), \frac{\gamma^N(n+1, k) + \gamma^N(n+1, k+1)}{2} \right\}
\]

for \( m \leq n < N \).

It is straightforward to compute \( \gamma^N(n, k) \) numerically for fixed \( N \) and \( m \).

Table 2 presents the actual values and the nominal values for different values of \( \varrho, c, m \) and \( N \). The effect of optimal stopping appears to be quite modest for the values of \( N \) and \( \varrho \) considered. For example, when \( N = 25 \), \( c = 1.96 \) and \( \varrho = 0.6 \), the actual (worst case) value is 3.89\%, less than 1.5\% larger than the nominal value of 2.49\%, and this is the largest increase reported in any of the tables. For \( N = 10 \), \( c = 1.96 \), and \( \varrho = 0.2 \), the actual (worst case) value is 2.91\%, less than 0.5\% larger than the nominal value of 2.49\%.

As expected, the effect increases with \( N \) and \( \varrho \). By the law of the iterated logarithm, the value must approach one as \( N \to \infty \). The tables indicate that the limit is approached very slowly.

One possible explanation for the small increase is that the \( Z_n \) are conditional probabilities while the nominal alpha is unconditional. Comparisons with a nominal alpha computed conditionally given the covariates awaits future work.

### 2.5. Hypothesis testing: observations taken in batches

It might not be feasible to analyze the data one at a time; instead we might need to see them in groups of size \( M \), say. We will study the effect of batch size in this section.

Denote the reward by \( G(k, j) \) where the sample size \( n = Mk \) and \( j \) is a typical value of \( S_n \), i.e. \( j = 0, 1, \ldots, Mk \). Then, the recursion takes the form

\[
G(K, j) = v(K, j) = \psi_{MK}(j) \quad (j = 0, 1, \ldots, MK),
\]
\[ G(k, j) = \max \left\{ v(k, j); \sum_{i=0}^{M} \left( \begin{array}{c} M \\ i \end{array} \right) \left( \frac{1}{2} \right)^i G(k+1, j+i) \right\} \] 

for \( j = 0, \ldots, Mk \) and \( k = 1, \ldots, K-1 \).

The stopping rule which maximizes this approximation is

\[ \tau = \inf\{ n; M \leq n \leq N, v(k, j) = G(k, j) \}. \]

Tables 3 and 4 present the actual values and the nominal values for different values of batches' size and total amount of observations. The size of a batch appears to have very little influence.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \varrho = 0.2 )</th>
<th>( \varrho = 0.4 )</th>
<th>( \varrho = 0.6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0310776</td>
<td>0.0373151</td>
<td>0.0447215</td>
</tr>
<tr>
<td>10</td>
<td>0.0297829</td>
<td>0.0352482</td>
<td>0.0421819</td>
</tr>
<tr>
<td>15</td>
<td>0.0291408</td>
<td>0.0339142</td>
<td>0.039647</td>
</tr>
<tr>
<td>20</td>
<td>0.0286684</td>
<td>0.0328846</td>
<td>0.0382302</td>
</tr>
<tr>
<td>25</td>
<td>0.0282947</td>
<td>0.0320360</td>
<td>0.0368202</td>
</tr>
<tr>
<td>30</td>
<td>0.0279620</td>
<td>0.0313182</td>
<td>0.0356413</td>
</tr>
<tr>
<td>35</td>
<td>0.0276746</td>
<td>0.0307106</td>
<td>0.0346493</td>
</tr>
<tr>
<td>40</td>
<td>0.0274260</td>
<td>0.0301735</td>
<td>0.0337392</td>
</tr>
<tr>
<td>45</td>
<td>0.0271943</td>
<td>0.0296778</td>
<td>0.0329047</td>
</tr>
<tr>
<td>50</td>
<td>0.0269879</td>
<td>0.0292301</td>
<td>0.0321442</td>
</tr>
<tr>
<td>55</td>
<td>0.0267951</td>
<td>0.0288250</td>
<td>0.0314633</td>
</tr>
<tr>
<td>60</td>
<td>0.0266171</td>
<td>0.0284414</td>
<td>0.0308259</td>
</tr>
<tr>
<td>65</td>
<td>0.0264512</td>
<td>0.0280886</td>
<td>0.0302222</td>
</tr>
<tr>
<td>70</td>
<td>0.0262946</td>
<td>0.0277563</td>
<td>0.0296635</td>
</tr>
<tr>
<td>75</td>
<td>0.0261483</td>
<td>0.0274406</td>
<td>0.0291385</td>
</tr>
<tr>
<td>80</td>
<td>0.0260081</td>
<td>0.0271473</td>
<td>0.0286339</td>
</tr>
<tr>
<td>85</td>
<td>0.0258768</td>
<td>0.0268639</td>
<td>0.0281617</td>
</tr>
<tr>
<td>90</td>
<td>0.0257500</td>
<td>0.0265977</td>
<td>0.0277078</td>
</tr>
<tr>
<td>95</td>
<td>0.0256298</td>
<td>0.0263399</td>
<td>0.0272713</td>
</tr>
<tr>
<td>100</td>
<td>0.0255141</td>
<td>0.0260954</td>
<td>0.0268590</td>
</tr>
<tr>
<td>105</td>
<td>0.0254022</td>
<td>0.0258576</td>
<td>0.0264553</td>
</tr>
<tr>
<td>110</td>
<td>0.0252960</td>
<td>0.0256119</td>
<td>0.0260730</td>
</tr>
<tr>
<td>115</td>
<td>0.0251925</td>
<td>0.0254116</td>
<td>0.0256980</td>
</tr>
<tr>
<td>120</td>
<td>0.0250937</td>
<td>0.0252011</td>
<td>0.0253400</td>
</tr>
<tr>
<td>125</td>
<td>0.0249977</td>
<td>0.0249961</td>
<td>0.0249893</td>
</tr>
</tbody>
</table>

Nominal: 0.0249979 0.0249979 0.0249979
Actual/nominal (\( n = M \)) 1.24 1.49 1.789
Table 4
Actual values for testing hypothesis (with batches)

<table>
<thead>
<tr>
<th>n</th>
<th>k=10, M=10, c=1.96</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ϕ = 0.2</td>
<td>ϕ = 0.4</td>
<td>ϕ = 0.6</td>
</tr>
<tr>
<td>10</td>
<td>0.0292559</td>
<td>0.034177</td>
<td>0.0397833</td>
</tr>
<tr>
<td>20</td>
<td>0.0282157</td>
<td>0.0318416</td>
<td>0.0363462</td>
</tr>
<tr>
<td>30</td>
<td>0.0274905</td>
<td>0.0303149</td>
<td>0.0339215</td>
</tr>
<tr>
<td>40</td>
<td>0.0269492</td>
<td>0.0291543</td>
<td>0.0319715</td>
</tr>
<tr>
<td>50</td>
<td>0.0265096</td>
<td>0.0282175</td>
<td>0.0303981</td>
</tr>
<tr>
<td>60</td>
<td>0.0261324</td>
<td>0.0274116</td>
<td>0.0290396</td>
</tr>
<tr>
<td>70</td>
<td>0.025802</td>
<td>0.0267076</td>
<td>0.027881</td>
</tr>
<tr>
<td>80</td>
<td>0.0255074</td>
<td>0.0260816</td>
<td>0.0268291</td>
</tr>
<tr>
<td>90</td>
<td>0.0252417</td>
<td>0.0255155</td>
<td>0.0258703</td>
</tr>
<tr>
<td>100</td>
<td>0.0249977</td>
<td>0.0249958</td>
<td>0.0249874</td>
</tr>
</tbody>
</table>

Nominal Actual/nominal (n=M)

<table>
<thead>
<tr>
<th></th>
<th>0.0249979</th>
<th>0.0249979</th>
<th>0.0249979</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.17</td>
<td>1.37</td>
<td>1.59</td>
</tr>
</tbody>
</table>

3. Theoretical solution to the hypothesis testing problem

In Section 3.1 we prove that $Z_n - Z_n^* \to 0$ with probability one. In Section 3.3 we approximate the discrete problem with the corresponding continuous problem. In doing so, we find that the approximate solution depends only on the correlation coefficient ($\varphi$) between $X$ and $Y$, and not on the distribution of $X$.

3.1. Convergence of $Z_n - Z_n^*$ to zero with probability one

Since $Z_n = P_0\{\sqrt{n}(Y_n - \theta_0)/\sigma_0 > c \mid X_1, \ldots, X_n\}$ we see that $1 - Z_n$ is a random variable. In addition to that, we know that $Y_1, \ldots, Y_n$ are conditionally independent variables given $X_1, \ldots, X_n$ and that the conditional distribution of $Y_i$ given $X_1, \ldots, X_n$ is the same as the conditional distribution of $Y_i$ given $X_i$.

Let

$$\theta(X_i) = E(Y_i \mid X_i), \quad \sigma^2(X_i) = V(Y_i \mid X_i),$$

$$\varphi = \text{corr}(X, Y), \quad \sigma^2_X = V(X), \quad s_n^2 = \sum_{i=1}^{n} \sigma^2(X_i).$$

Then, by the law of large numbers,

$$\frac{s_n^2}{n} \to \sigma^2 = \int V(Y \mid X) \, dF(x) \quad \text{and} \quad \sigma^2 = \sigma_0^2(1 - \varphi^2).$$

We see that

$$\sum (Y_i - \theta(X_i)) = n(\bar{Y}_n - \theta_0) - \varphi \frac{\sigma_0}{\sigma_X} n(\bar{X}_n - \mu)$$
and
\[ Z_n = P_0 \left\{ \frac{\sum_{i=1}^{n} (Y_i - \theta(X_i))}{s_n} > \frac{\sigma_0 \sqrt{n}}{s_n} \left( c - \frac{\rho}{\sigma_x} \sqrt{n}(\bar{X}_n - \mu) \right) \mid X_1, \ldots, X_n \right\} . \]

If we define
\[ G_n(z) = G_n(z; X_1, \ldots, X_n) = P_0 \left\{ \frac{\sum_{i=1}^{n} (Y_i - \theta(X_i))}{s_n} \leq z \mid X_1, \ldots, X_n \right\} , \]
then
\[ Z_n = 1 - G_n \left\{ \frac{\sigma_0 \sqrt{n}}{s_n} \left( c - \frac{\rho}{\sigma_x} \sqrt{n}(\bar{X}_n - \mu) \right) \right\} . \]

**Theorem 3.1.** \( Z_n - Z^*_n \to 0 \) with probability one.

**Proof.** We can write
\[ Z_n - Z^*_n = Z_n - \Phi_n \left\{ -\sqrt{n}(\bar{X}_n - \mu) \right\} - \left\{ Z^*_n - \Phi_n \left\{ -\sqrt{n}(\bar{X}_n - \mu) \right\} \right\} \]
where
\[ \Phi_n(z) = 1 - \Phi \left\{ \frac{\sigma_0 \sqrt{n}}{s_n} \left( c + \frac{\rho}{\sigma_x} z \right) \right\} . \]

**Part 1.** We first prove that
\[ \left| G_n \left\{ \frac{\sigma_0 \sqrt{n}}{s_n} \left( c - \frac{\rho}{\sigma_x} \sqrt{n}(\bar{X}_n - \mu) \right) \right\} - \Phi_n (-\sqrt{n}(\bar{X}_n - \mu)) \right| \to 0 \quad \text{w.p. 1.} \]

If we define
\[ Y_{ni} = \frac{Y_i - \theta(X_i)}{s_n} \]
then
\[ S_n = S_{nn} = \sum_{i=1}^{n} Y_{ni} = \frac{\sum_{i=1}^{n} (Y_i - \theta(X_i))}{s_n} . \]

Thus, we are dealing with a double array of (conditionally) independent random variables:

\[ Y_{11} \]
\[ Y_{21}, Y_{22} ; \]
\[ Y_{31}, Y_{32}, Y_{33} ; \]
\[ \ldots \]
\[ Y_{n1}, Y_{n2}, \ldots, Y_{nn} ; \]
\[ \ldots \]
for which
\[ \sum_{i=1}^{n} \sigma^2(Y_n) = 1. \]

The Central Limit Theorem asserts that \( S_n \) converges in distribution to the unit normal distribution if Lindeberg's condition is satisfied.

Let
\[ LF_n(\varepsilon) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{|y_i - \theta(X_i)| \geq \varepsilon} |y_i - \theta(X_i)|^2 dG_i(y_i). \]

If \( LF_n(\varepsilon) \to 0 \) for all \( \varepsilon \), then \( S_n \) converges in distribution to the standard normal distribution. Thus, we must show that \( LF_n(\varepsilon) \to 0 \) w.p. 1 for all \( \varepsilon > 0 \).

Define the event
\[ B_n = \{ s_n \geq \frac{1}{2} nA \} \quad \text{where} \quad A = \sigma^2_0(1 - \phi^2). \]

Then, the indicator function of the complement of \( B_n \)
\[ I_{B_n} \to 0 \quad \text{w.p. 1.} \]

If \( B_n \) occurs, then for all \( c > 0 \) there exists \( n_0 \) such that for all \( n \geq n_0 \), we have
\[ LF_n(\varepsilon) I_{B_n} \leq \frac{2}{nA} \sum_{i=1}^{n} \int_{|y_i - \theta(X_i)| \geq c} |y_i - \theta(X_i)|^2 dG_i(y_i). \]

If we define
\[ u_c(X_i) = \int_{|y_i - \theta(X_i)| \geq c} |y_i - \theta(X_i)|^2 dG_i(y_i) \]
we see that the random variables \( u_c(X_i) \) are i.i.d.; so, applying the law of large numbers
\[ \frac{1}{n} \sum_{i=1}^{n} u_c(X_i) \to \int_{-\infty}^{\infty} u_c(x) dF(x) \]
where
\[ \int_{-\infty}^{\infty} u_c(x) dF(x) = \int_{|y - \theta(x)| > c} |y - \theta(x)|^2 dH(x, y) \]
and \( H \) denotes the joint distribution of \( X \) and \( Y \). Then, if \( c \to \infty \),
\[ \int_{|y - \theta(x)| \geq c} |y - \theta(x)|^2 dH(x, y) \to 0. \]

The Lindeberg-Feller condition is satisfied w.p. 1, so \( G_n \) converges weakly to \( \Phi \) w.p. 1 Pólya’s theorem asserts: If \( F_n \) and \( F \) are distribution functions, and if \( F_n \) converges weakly to \( F \), and \( F \) is continuous, then the convergence is uniform, i.e.
\[ \sup_x |F_n(x) - F(x)| \to 0. \] Using Pólya’s theorem, we see that \( G_n \) converges uniformly to \( \Phi \). So
\[
Z_n - \Phi_n(-\sqrt{n}(\bar{X}_n - \mu)) = G_n \left( \frac{\sigma_0 \sqrt{n}}{s_n} \left( c - \frac{\rho \sqrt{n}}{\sigma_x}(\bar{X}_n - \mu) \right) \right) - \Phi \left( \frac{\sigma_0 \sqrt{n}}{s_n} \left( c - \frac{\rho \sqrt{n}}{\sigma_x}(\bar{X}_n - \mu) \right) \right) \to 0.
\]

**Part 2.** We now show that

\[\Phi \left( \frac{\sigma_0 \sqrt{n}}{s_n} \left( c - \frac{\rho \sqrt{n}}{\sigma_x}(\bar{X}_n - \mu) \right) \right) - Z^*_n \to 0.\]

This follows easily from Pólya’s theorem with

\[F_n(z) = \Phi \left( \frac{\sigma_0 \sqrt{n}}{s_n} \left( c + \frac{\rho}{\sigma_x}z \right) \right) \quad \text{and} \quad F(z) = \Phi \left( \frac{1}{\sqrt{1 - \rho^2}} \left( c + \frac{\rho}{\sigma_x}z \right) \right)\]

for \(-\infty < z < \infty\). Then \(F_n\) converges weakly to \(F\) w.p. 1. So

\[|Z_n^* - F_n(-\sqrt{n}(\bar{X}_n - \mu))| = |F(-\sqrt{n}(\bar{X}_n - \mu)) - F_n(-\sqrt{n}(\bar{X}_n - \mu))| \leq \sup_z |F_n(z) - F(z)| \to 0.
\]

Theorem 3.2 will give us an approximation which does not depend on the distribution of \(X\). We use the corresponding continuous problem.

### 3.2. Approximate solution using the corresponding continuous problem

Let \(\{W(t)\}_{t \geq 0}\) be a normalized Brownian motion, then we will show that

\[\sup_{m \leq r \leq N} E(Z_n^*) - \sup_{1 \leq r \leq 1/\epsilon} E \left\{ 1 - \Phi \left( \frac{c - \rho W(t)/\sqrt{\tau}}{\sqrt{1 - \rho^2}} \right) \right\} \to 0\]

as \(m, N \to \infty\) with \(m/N \to \epsilon\), where \(Z^*_n\) was defined in (3.1) and the suprema extend over stopping times.

Let \(W(t)\) be a normalized Wiener process, then

\[W^m(t) = \sqrt{m}W \left( \frac{t}{m} \right)\]

is also a Wiener process. For this \(W^m(t)\) we use Skorokhod’s representation theorem to construct stopping times \(\sigma_{m1}, \sigma_{m2}, \ldots\) such that the partial sums

\[S_n \overset{D}{=} W^m(\sigma_{mn}).\]

There is no loss of generality in supposing that \(S_n\) is equal to \(W^m(\sigma_{mn})\) for all \(n = 1, 2, \ldots\).

We define

\[W_m(t) = \frac{S_{[mt]}}{\sqrt{m}} = \frac{1}{\sqrt{m}} W^m(\sigma_{m[mt]}) = W \left( \frac{\sigma_{m[mt]}}{m} \right).\]  \hspace{1cm} (3.1)

We know that

\[\sup_{0 \leq t \leq c} |W_m(t) - W(t)| \overset{p}{\to} 0 \quad \text{as} \quad m \to \infty \hspace{1cm} (3.2)\]
for any $c$. Now we define

$$D_m(t) = \sigma\{W_m(s) : s \leq t\}, \quad D(t) = \sigma\{W(s) : s \leq t\}$$

and let $C_m$ be the class of all $\tau$ s.t. $\tau \in [1, 1/e]$ w.r.t. $D_m(t)$ and $C$ be the class of all $\tau$ s.t. $\tau \in [1, 1/e]$ w.r.t. $D(t)$. Further, let

$$u(\tau, x) = 1 - \Phi\left(\frac{c - qx/\sqrt{t}}{\sqrt{1 - q^2}}\right).$$

Then

$$Z^*_\tau = u(\tau, W_m(\tau)) \quad \text{and} \quad u(\tau, W(\tau)) = 1 - \Phi\left(\frac{c - qW(\tau)/\sqrt{\tau}}{\sqrt{1 - q^2}}\right).$$

Theorem 3.2.

$$\sup_{\tau \in C_m} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C} E\{u(\tau, W(\tau))\} \to 0$$

as $m, N \to \infty$ with $m/N \to \varepsilon$.

Proof. With $W_m(\tau)$ as defined in (3.2) let $C_{m,k}$ be the class of all $\tau$ s.t. $\tau \in [1, 1/e]$ w.r.t. $D_m(t)$, which are of the form $\tau = j/k, j = k, \ldots, k/e$, and let $C_k$ be the class of all $\tau$ s.t. $\tau \in [1, 1/e]$ w.r.t. $D(t)$, which are of the form $\tau = j/k, j = k, \ldots, k/e$. Then

$$\sup_{\tau \in C_m} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C} E\{u(\tau, W(\tau))\} = \sup_{\tau \in C_m} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C_{m,k}} E\{u(\tau, W_m(\tau))\} + \sup_{\tau \in C_{m,k}} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C_k} E\{u(\tau, W(\tau))\} + \sup_{\tau \in C_k} E\{u(\tau, W(\tau))\} - \sup_{\tau \in C} E\{u(\tau, W(\tau))\} = I + II + III.$$

Next we prove that $III \to 0$. If $\tau$ is a stopping time, let

$$\tau_k = \left[\frac{(k + 1)\tau}{k}\right] \frac{1}{k}.$$

Then $\tau_k \in C_k$ ($\tau_k \leq t$ iff $k\tau_k \leq kt$ iff $(k + 1)\tau \leq [kt]$ which implies that $\tau_k \in C_k$) and $|\tau_k - \tau| \leq 1/k$. It follows easily that $III \to 0$ as $k \to \infty$ and a similar argument shows that sup, $I \to 0$.

For a fixed $k$, we can use the backward induction technique of dynamic programming to obtain

$$\lim_{m \to \infty} II(m, k) = 0 \quad \text{for all } k.$$

To see why, fix $K$, let $K = k/e$, and define functions $u_j^m$ and $u_j, j = k, \ldots, K$, by

$$u_K^m(y) = v_K(y) = u\left(\frac{K}{k}, y\right),$$
\[ u_j^m(y) = \max \left\{ u \left( \frac{j}{k}, y \right), \int_{-\infty}^{\infty} u_{j+1}^m(y+z) \, dF_{m,j}(z) \right\} \]

and

\[ v_j(y) = \max \left\{ u \left( \frac{j}{k}, y \right), \int_{-\infty}^{\infty} v_{j+1}(y+z) \, d\Phi_k(z) \right\} \]

for \(-\infty < y < \infty\), \(j = k, \ldots, K - 1\), where

\[ \Phi_k(z) = \Phi(z \sqrt{k}), \quad F_{m,j}(z) = P(S_{mj} \leq z \sqrt{m}) \]

and

\[ m_j = \left[ \frac{m(j+1)}{k} \right] - \left[ \frac{mj}{k} \right] \]

for \(-\infty < z < \infty\) and \(j = k, \ldots, K - 1\). Then

\[ \sup_{\tau \in \mathcal{C}_{m,k}} E\{u(\tau, W_m(\tau))\} = E\{u_k^m(W_m(1))\} \]

and

\[ \sup_{\tau \in \mathcal{C}_k} E\{u(\tau, W(\tau))\} = E\{v_k(W(1))\} \]

by the dynamic programming relations. Since \(W_m(1)\) converges in distribution to \(W(1)\) as \(m \to \infty\), it suffices to show that

\[ u_k^m(y) \to v_k(y) \quad (3.3) \]

uniformly on compacts (in \(y\)). This may be done by backward induction. If \(j = K\), then \(u_j^m(y) = u(K/k, y) = v_j(y)\) for all \(m\). So the relation is obvious. Now suppose that

\[ u_i^m(y) \to v_i(y) \]

uniformly on compacts in \(y\) as \(m \to \infty\) for all \(i > j\), where \(k \leq j < K\). Then, since \(F_{m,j}\) converges weakly to \(\Phi_k\),

\[ u_j^m(y) = \max \left\{ u \left( \frac{j}{k}, y \right), \int_{-\infty}^{\infty} u_{j+1}^m(y+z) \, dF_{m,j}(z) \right\} \]

\[ \to \max \left\{ u \left( \frac{j}{k}, y \right), \int_{-\infty}^{\infty} v_{j+1}(y+z) \, d\Phi_k(z) \right\} \]

\[ = v_j(z) \]

as \(m \to \infty\). So the induction is complete and (4.4) follows. This completes the proof that \(II \to 0\) and therefore the proof of the theorem.

References


