Journal of Statistical Planning and Inference 32 (1992) 311-323
North-Holland

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 σ-algebra with respect to which $X_1,\ldots,X_n$ are measurable. We sample sequentially according to a stopping rule t depending on the X_i's.

Suppose that a one-sided hypothesis about θ is of interest, say H_0: $\theta\leqslant\theta_0$, and that H_0 is to be rejected if

$$\sqrt{n}(\bar{Y}_n-\theta_0)/\hat{\sigma}_n>c$$

at the time $t=n$ of stopping, where $\bar{Y}_n$ is the sample mean, $\hat{\sigma}_n^2$, $n\geqslant 1$, 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

Correspondence to: Dr. Martha Bilotti-Aliaga, Dept. of Statistics, 1444 Mason Hall, University of Michigan, Ann Arbor, MI 48109-1027, USA.

$$\alpha_t = P_0\left\{\frac{\sqrt{t}(\bar{Y}_t - \theta_0)}{\hat{\sigma}_t} > c\right\},$$

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 α_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 (ϱ) 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: $\theta \leqslant \theta_0$. 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 θ, say H_0: $\theta \leqslant \theta_0$.

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

$$E(\bar{Y}_n - \theta \mid X_1, \ldots, X_n) = \delta(\bar{X}_n - \mu) \tag{2.1}$$

and, if the conditional variance σ^2 of Y given X is independent of X, then

$$V\{\bar{Y}_n \mid X_1, \ldots, X_n\} = \frac{\sigma^2}{n}, \tag{2.2}$$

where δ 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

$$Z_n^* = 1 - \Phi\left\{\frac{c\sigma_0 - \delta\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right\} = 1 - \Phi\left\{\frac{c - \varrho\sqrt{n}\sigma_x^{-1}(\bar{X}_n - \mu)}{\sqrt{1-\varrho^2}}\right\}.$$

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

$$V^* = \max_{t \in C} E_0(Z_t^*).$$

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 $E(Y) = \theta$, $\operatorname{corr}(X, Y) = \varrho$ 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

$$Z_n^* = 1 - \Phi\left\{\frac{c - 2\sqrt{n}\varrho(\bar{X}_n - \frac{1}{2})}{\sqrt{1-\varrho^2}}\right\}.$$

Table 1

	$X=0$	$X=1$	
$Y=0$	$(1-\theta)^2 + \varrho\theta(1-\theta)$	$\theta(1-\theta) - \varrho\theta(1-\theta)$	$1-\theta$
$Y=1$	$\theta(1-\theta) - \varrho\theta(1-\theta)$	$\theta^2 + \varrho\theta(1-\theta)$	θ
	$1-\theta$	θ	

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

$$Z_n^* = 1 - \Phi\left\{\frac{c - 2\sqrt{n}\varrho(\bar{X}_n - \frac{1}{2})}{\sqrt{1-\varrho^2}}\right\} = \psi_n(S_n), \quad \text{say,} \tag{2.3}$$

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)

m	$N=25$, $c=1.96$		
	$\varrho=0.2$	$\varrho=0.4$	$\varrho=0.6$
1	0.0302563	0.0347522	0.0388983
2	0.0298616	0.0347522	0.0388983
3	0.0291243	0.0336786	0.0338983
4	0.0286438	0.0325705	0.0374607
5	0.0282968	0.0318539	0.0360676
6	0.0279349	0.0311908	0.0351405
7	0.02769	0.0305473	0.0342162
8	0.02742	0.0300974	0.033292
9	0.027203	0.029593	0.032616
10	0.0269886	0.0291692	0.0318769
11	0.0267989	0.0287644	0.0312053
12	0.0266197	0.0283809	0.0306324
13	0.0264521	0.0280448	0.0300252
14	0.0262974	0.0277005	0.0295269
15	0.0261482	0.0274142	0.0289993
16	0.0260105	0.0271086	0.0285209
17	0.0258765	0.0268451	0.0280584
18	0.0257513	0.0265709	0.0276085
19	0.0256298	0.0263258	0.0271958
20	0.0255143	0.026076	0.0267767
21	0.0254034	0.025848	0.0264001
22	0.0252955	0.0256176	0.0260112
23	0.0251935	0.0254053	0.0256582
24	0.0250925	0.0251917	0.0252976
25	0.0249974	0.0249896	0.0249551
Nominal	0.0249979	0.0249979	0.0249979
Actual/nominal	1.21	1.39	1.56

where

$$C_m^N = \{t; t \text{ is a stopping rule, } m \leqslant t \leqslant N\}.$$

To find the optimal solution for V_m^N 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 γ_N^N, $\gamma_{N-1}^N, \dots, \gamma_m^N$ by setting

$$\begin{aligned} \gamma_N^N &= Z_N^* = \psi_N(S_N), \\ \gamma_n^N &= \max(Z_n^*, E_0(\gamma_{n+1}^N \mid F_n)) \quad (n = N-1, \dots, m). \end{aligned} \tag{2.4}$$

For each $n = m, m+1, \dots, N$ let

$$s_n^N = \inf\{i \geqslant n, Z_i^* = \gamma_i^N\}.$$

It follows from that theorem that s_m^N is optimal in C_m^N.

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

$$\gamma^N(n,k) = \max\left\{\psi_n(k), \frac{\gamma^N(n+1,k) + \gamma^N(n+1,k+1)}{2}\right\} \tag{2.5}$$

for $m \leqslant n < N$.

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 ϱ, c, m and N. The effect of optimal stopping appears to be quite modest for the values of N and ϱ 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 ϱ. 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 is $n = Mk$ and j is a typical value of S_n, i.e. $j = 0, 1, \dots, Mk$. Then, the recursion takes the form

$$G(K, j) = v(K, j) = \psi_{MK}(j) \quad (j = 0, 1, \dots, MK),$$

$$G(k,j) = \max\left\{ \upsilon(k,j);\ \sum_{i=0}^{M} \binom{M}{i} \left(\frac{1}{2}\right)^{M} G(k+1, j+i) \right\}$$

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

The stopping rule which maximizes this approximation is

$$\tau = \inf\{n;\ M \leqslant n \leqslant N,\ \upsilon(k,j) = G(k,j)\}.$$

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 3
Actual values for testing hypothesis (with batches)

n	$K=25,\ M=5,\ c=1.96$		
	$\varrho=0.2$	$\varrho=0.4$	$\varrho=0.6$
5	0.0310776	0.0373151	0.0447215
10	0.0297829	0.0352482	0.0421819
15	0.0291408	0.0339142	0.0399647
20	0.0286684	0.0328846	0.0382302
25	0.0282947	0.0320360	0.0368202
30	0.027962	0.0313182	0.0356413
35	0.0276746	0.0307106	0.0346493
40	0.0274260	0.0301735	0.0337392
45	0.0271943	0.0296778	0.0329047
50	0.0269879	0.0292301	0.0321442
55	0.0267951	0.0288250	0.0314633
60	0.0266171	0.0284414	0.0308259
65	0.0264512	0.0280886	0.0302222
70	0.0262946	0.0277563	0.0296635
75	0.0261483	0.0274406	0.0291385
80	0.0260081	0.0271473	0.0286339
85	0.0258768	0.0268639	0.0281617
90	0.0257500	0.0265977	0.0277078
95	0.0256298	0.0263399	0.0272713
100	0.0255141	0.0260954	0.0268590
105	0.0254022	0.0258576	0.0264553
110	0.0252960	0.0256319	0.0260730
115	0.0251925	0.0254116	0.0256980
120	0.0250937	0.0252011	0.0253400
125	0.0249977	0.0249961	0.0249893
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)

n	$k=10,\ M=10,\ c=1.96$		
	$\varrho=0.2$	$\varrho=0.4$	$\varrho=0.6$
10	0.0292559	0.034177	0.0397833
20	0.0282157	0.0318416	0.0363462
30	0.0274905	0.0303149	0.0339215
40	0.0269492	0.0291543	0.0319715
50	0.0265096	0.0282175	0.0303981
60	0.0261324	0.0274116	0.0290396
70	0.025802	0.0267076	0.027881
80	0.0255074	0.0260816	0.0268291
90	0.0252417	0.0255155	0.0258703
100	0.0249977	0.0249958	0.0249874
Nominal	0.0249979	0.0249979	0.0249979
Actual/nominal ($n=M$)	1.17	1.37	1.59

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 (ϱ) 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 $Z_n = P_0\{\sqrt{n}(\bar{Y}_n - \theta_c)/\sigma_0 > c \mid X_1, \ldots, X_n\}$ 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

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

$$\varrho = \operatorname{corr}(X, Y), \qquad \sigma_x^2 = V(X), \qquad s_n^2 = \sum_{i=1}^{n} \sigma^2(X_i).$$

Then, by the law of large numbers,

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

We see that

$$\sum (Y_i - \theta(X_i)) = n(\bar{Y}_n - \theta_0) - \varrho\,\frac{\sigma_0}{\sigma_x}\,n(\bar{X}_n - \mu)$$

and

$$Z_n = P_0\left\{\frac{\sum_{i=1}^n (Y_i-\theta(X_i))}{s_n} > \frac{\sigma_0\sqrt{n}}{s_n}\left[c-\frac{\varrho}{\sigma_x}\sqrt{n}(\bar{X}_n-\mu)\right] \mid X_1,\dots,X_n\right\}.$$

If we define

$$G_n(z) = G_n(z;X_1,\dots,X_n) = P_0\left\{\frac{\sum_{i=1}^n (Y_i-\theta(X_i))}{s_n} \leqslant z \mid X_1,\dots,X_n\right\},$$

then

$$Z_n = 1 - G_n\left\{\frac{\sigma_0\sqrt{n}}{s_n}\left(c-\frac{\varrho}{\sigma_x}\sqrt{n}(\bar{X}_n-\mu)\right)\right\}.$$

Theorem 3.1. *$Z_n - Z_n^* \to 0$ with probability one.*

Proof. We can write

$$Z_n - Z_n^* = Z_n - \Phi_n\{-\sqrt{n}(\bar{X}_n-\mu)\} - \{Z_n^* - \Phi_n\{-\sqrt{n}(\bar{X}_n-\mu)\}\}$$

where

$$\Phi_n(z) = 1 - \Phi\left\{\frac{\sigma_0\sqrt{n}}{s_n}\left(c+\frac{\varrho}{\sigma_x}z\right)\right\}.$$

Part 1. We first prove that

$$\left|G_n\left\{\frac{\sigma_0\sqrt{n}}{s_n}\left(c-\frac{\varrho\sqrt{n}}{\sigma_x}(\bar{X}_n-\mu)\right)\right\} - \Phi_n(-\sqrt{n}(\bar{X}_n-\mu))\right| \to 0 \quad \text{w.p. 1.}$$

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};

...

$Y_{n1}, Y_{n2}, \dots, Y_{nn}$;

...

for which

$$\sum_{i=1}^{n} \sigma^2(Y_{ni}) = 1.$$

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

Let

$$LF_n(\varepsilon) := \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|y_i - \theta(X_i)| \geqslant \varepsilon s_n} |y_i - \theta(X_i)|^2 \, \mathrm{d}G_i(y_i).$$

If $LF_n(\varepsilon) \to 0$ for all ε, 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

$$B_n = \{s_n^2 \geqslant \tfrac{1}{2} nA\} \quad \text{where } A = \sigma_0^2(1 - \varrho^2).$$

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 $n \geqslant n_0$, we have

$$LF_n(\varepsilon) I_{B_n} \leqslant \frac{2}{nA} \sum_{i=1}^{n} \int_{|y_i - \theta(X_i)| \geqslant c} |y_i - \theta(X_i)|^2 \, \mathrm{d}G_i(y_i).$$

If we define

$$u_c(X_i) = \int_{|y_i - \theta(X_i)| \geqslant c} |y_i - \theta(X_i)|^2 \, \mathrm{d}G_i(y_i)$$

we see that the random variables $u_c(X_i)$ are i.i.d.; so, applying the law of large numbers

$$\frac{\sum_{i=1}^{n} u_c(X_i)}{n} \to \int_{-\infty}^{\infty} u_c(x) \, \mathrm{d}F(x)$$

where

$$\int_{-\infty}^{\infty} u_c(x) \, \mathrm{d}F(x) = \iint_{|y - \theta(x)| > c} |y - \theta(x)|^2 \, \mathrm{d}H(x, y)$$

and H denotes the joint distribution of X and Y. Then, if $c \to \infty$,

$$\iint_{|Y - \theta(x)| \geqslant c} |Y - \theta(x)|^2 \, \mathrm{d}H(x, y) \to 0.$$

The Lindeberg–Feller condition is satisfied w.p. 1, so G_n converges weakly to Φ 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 Φ. So

$$Z_n - \Phi_n(-\sqrt{n}(\bar{X}_n - \mu))$$
$$= G_n\left\{\frac{\sigma_0\sqrt{n}}{s_n}\left(c - \frac{\varrho\sqrt{n}}{\sigma_x}(\bar{X}_n - \mu)\right)\right\} - \Phi\left\{\frac{\sigma_0\sqrt{n}}{s_n}\left(c - \frac{\varrho\sqrt{n}}{\sigma_x}(\bar{X}_n - \mu)\right)\right\} \to 0.$$

Part 2. We now show that

$$\Phi\left\{\frac{\sigma_0\sqrt{n}}{s_n}\left(c - \frac{\varrho\sqrt{n}}{\sigma_x}(\bar{X}_n - \mu)\right)\right\} - Z_n^* \to 0.$$

This follows easily from Pólya's theorem with

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

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

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

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

$$\sup_{m\leqslant\tau\leqslant N} E(Z_\tau^*) - \sup_{1\leqslant\tau\leqslant 1/\varepsilon} E\left\{1 - \Phi\left\{\frac{c - \varrho W(\tau)/\sqrt{\tau}}{\sqrt{1-\varrho^2}}\right\}\right\} \to 0$$

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

$$S_n \overset{\mathrm{D}}{=} W^m(\sigma_{mn}).$$

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

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

We know that

$$\sup_{0\leqslant t\leqslant c} |W_m(t) - W(t)| \xrightarrow{\mathrm{P}} 0 \quad \text{as } m \to \infty \tag{3.2}$$

for any c. Now we define

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

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

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

Then

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

Theorem 3.2.

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

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 τ s.t. $\tau \in [1, 1/\varepsilon]$ w.r.t. $D_m(t)$, which are of the form $\tau = j/k$, $j = k, \dots, k/\varepsilon$, and let C_k be the class of all τ s.t. $\tau \in [1, 1/\varepsilon]$ w.r.t. $D(t)$, which are of the form $\tau = j/k$, $j = k, \dots, k/\varepsilon$. Then

$$\begin{aligned}
&\sup_{\tau \in C_m} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C} E\{u(\tau, W(\tau))\} \\
&\quad = \sup_{\tau \in C_m} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C_{m,k}} E\{u(\tau, W_m(\tau))\} \\
&\qquad + \sup_{\tau \in C_{m,k}} E\{u(\tau, W_m(\tau))\} - \sup_{\tau \in C_k} E\{u(\tau, W(\tau))\} \\
&\qquad + \sup_{\tau \in C_k} E\{u(\tau, W(\tau))\} - \sup_{\tau \in C} E\{u(\tau, W(\tau))\} \\
&\quad = \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}$$

Next we prove that $\mathrm{III} \to 0$. If τ is a stopping time, let

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

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

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

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

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

$$u_K^m(y) = v_K(y) = u\left(\frac{K}{k}, y\right),$$

$$u_j^m(y) = \max\left\{u\left(\frac{j}{k}, y\right),\quad \int u_{j+1}^m(y+z)\,\mathrm{d}F_{m,j}(z)\right\}$$

and

$$v_j(y) = \max\left\{u\left(\frac{j}{k}, y\right),\quad \int v_{j+1}(y+z)\,\mathrm{d}\Phi_k(z)\right\}$$

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

$$\Phi_k(z) = \Phi(z\sqrt{k}), \qquad F_{m,j}(z) = P(S_{mj} \leqslant z\sqrt{m})$$

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

$$\sup_{\tau \in C_{m,k}} E\{u(\tau, W_m(\tau))\} = E\{u_k^m(W_m(1))\}$$

and

$$\sup_{\tau \in C_k} E\{u(\tau, W(\tau))\} = E\{v_k(W(1))\}$$

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

$$u_k^m(y) \to v_k(y) \tag{3.3}$$

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 $k \leqslant j < K$. Then, since $F_{m,j}$ converges weakly to Φ_k,

$$\begin{aligned} u_j^m(y) &= \max\left\{u\left(\frac{j}{k}, y\right), \int_{-\infty}^{\infty} u_{j+1}^m(y+z)\,\mathrm{d}F_{m,j}(z)\right\} \\ &\to \max\left\{u\left(\frac{j}{k}, y\right), \int_{-\infty}^{\infty} v_{j+1}(y+z)\,\mathrm{d}\Phi_k(z)\right\} \\ &= v_j(z) \end{aligned}$$

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

References

Abramowitz, M. and A. Stegun (1964). *Handbook of Mathematical Tables.* National Bureau of Standards, Washington, D.C.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

Breiman, L. (1968). *Probability.* Addison-Wesley, Reading, MA.

Chung, K.L. (1974). *A Course in Probability Theory.* Academic Press, New York.

Chow, Y., H. Robbins and D. Siegmund (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin, Boston, MA.

Freeman, D. (1983). *Brownian Motion and Diffusion.* Springer, New York.

Siegmund, D. (1985). *Sequential Analysis. Test and Confidence Intervals.* Springer, New York.