Statistica Sinica **7**(1997), 109-126

A PARTICULAR APPLICATION OF BROWNIAN MOTION TO SEQUENTIAL ANALYSIS

Steven L. Hubert and Ronald Pyke

University of Washington

Dedicated to Herbert E. Robbins on the Occasion of his 80th Birthday

Abstract: This paper studies uniform distributional properties of particular stopping times for Brownian motion that are determined by a family of stopping curves indexed by $p \in [0, 1]$. These curves derive from the stopping curve for a sequential estimation problem in which the goal is to estimate a function of the Binomial parameter p that diverges as p approaches zero. The almost sure convergence and asymptotic normality of the stopping times for the Brownian analogue of this problem are obtained straightforwardly. The main result is the derivation of exponential bounds for the tail probabilities of a weighted mean square loss function expressible in terms of these stopping times. This result suffices to establish the uniform integrability of these loss functions in this continuous model, providing more importantly the methodology to prove the more difficult consistency result for the discrete Binomial problem. Brief historical comments about Brownian motion are included, as well as several open problems related to Brownian processes and sequential methods.

Key words and phrases: Bachelier, Binomial inference, Brownian motion, sequential estimation, stopping times, uniform integrability.

1. Introduction

The purpose of this paper is to study the asymptotic behavior of a particular family of stopping times for Brownian motion. The motivation for considering these stopping times comes from a particular sequential estimation problem in the context of a sequence of independent Bernoulli-p observations. The specific problem is discussed in Section 2 below, but for now, consider the prototypical example in which one wishes to estimate sequentially the rate $1/p$ by $1/\bar{X}_N$, the reciprocal of the stopped sample mean, in such a way as to maintain a specified limiting risk as $p \searrow 0$ under the weighted mean-square loss function, $p(\bar{X}_N^{-1} - p^{-1})^2$. The stopping time, N, is chosen to be the smallest sample size n that exceeds the (approximate) best fixed sample size for the Binomial problem when $p = \bar{X}_n$. General problems of this type are studied in detail in Hubert and Pyke (1995), where the asymptotic results include the uniform integrability of the loss function which is needed to insure the desired limit for the risk function when $p \to 0$. We expand a little on this in Section 2.

What is the role of Brownian motion in this particular sequential Binomial problem? Clearly, Brownian motion and sequential analysis have had many interconnections over the years. In fact, the stopping time aspect of sequential analysis can be found in the very first paper in which the mathematical construct now known as Brownian motion is introduced, namely the 1900 paper by Louis Bachelier. This lengthy treatise is an excellent example of stochastic modeling in which the real-life phenomenon being modeled is the fluctuation of prices on the French bond market. Both the Binomial random walk and the approximating Brownian motion are presented in Bachelier (1900). The latter is of course not done rigorously, but certainly the ideas are discussed in a surprisingly up-to-date way by Bachelier, including the Chapman-Kolmogorov equations (page 29) the second-order heat equations (page 46) and the reflection principle (page 75). Though other processes may be preferred today for modeling market price fluctuations, Bachelier, for his part, included many examples of actual data in order to indicate the good level of fit for his model of these particular prices.

In particular, the idea of a stopping time occurs already in this first mathematical paper on Brownian motion. For among the many specific problems considered by Bachelier is that of a person who buys a simple option (to buy some bonds at a specified price) with the intention of selling a futures option on these bonds when their price attains a predetermined gain. Thus, Bachelier (1900), p. 84 studies the one-sided barrier problem for Brownian motion. Specifically, if T_t denotes the (truncated by t) first time for standard Brownian motion Z to hit level $x > 0$ during $(0, t]$, his results include the probability distribution of T_t, its mean and the conditional law of $Z(t)$ given that the barrier had not been reached prior to it.

Though in recent decades there has been considerable activity in the use of Brownian-related processes in several areas of finance, the original paper of Bachelier did not receive its due recognition. Instead, during the first half of this century, the tremendous developments in the theory and applications of Brownian motion followed not from Bachelier's work but the independent later work of Albert Einstein. In Einstein (1905), this mathematical construct known as Brownian motion was introduced to model the fluctuations of molecules within the molecular theory of heat. Indeed, this particular physical model provided a theoretical justification for the movement of particles observed much earlier by Sir Robert Brown in 1827 (see Brown (1828)) from whom the process takes its name. The great developments in the probabilistic theory of Brownian motion followed from Einstein through the work of Norbert Wiener and Paul Lévy, rather than from the potentially superior but overlooked beginnings of Bachelier. (A fuller discussion on the history of Brownian motion will appear in Pyke (1995).)

A poignant reflection upon the lack of attention given to Bachelier's wealth of ideas is given in Mandelbrot (1989) where it is stated that

> "He invented efficient markets in 1900, sixty years before the idea came into vogue. He described the random walk model of prices, ordinary diffusion of probability—also called Brownian motion—and martingales, which are the mathematical expression of efficient markets. He even attempted an empirical verification. But he remained a shadowy presence until 1960 or so, when his major work was revived in English translation."

In any event, the subject of Brownian motion has flourished greatly this century, and represents today a major cornerstone of probability. In particular, for this paper, we emphasize the significant roles played by Brownian motion within sequential analysis. Perhaps most importantly is its natural role in providing central-limit-type approximations to the random walks of discrete models, approximations that have been around since the beginnings of Brownian motion; note for example, Lord Rayleigh's response to Karl Pearson's question about the 'Random Walker' (cf. Rayleigh (1905), Pearson (1905)), in the same year as Einstein's construction (Einstein (1905)). Some early sequential analysis examples of this role are Vogel (1960), Anderson (1960) and Chernoff (1961). In the first of these, Brownian motion is used to approximate a random walk constrained between two-sided barriers, $\pm b$. After computing $E(T)$ for the Brownian motion case, it is then used as a heuristic limit for the actual mean stopping time. In Anderson (1960) the Brownian motion approximation to normal random walks is used to simplify calculations in numerical comparisons of sequential procedures. The third paper is the first of a series of four papers (see Chernoff (1965) for references) in which the Brownian approximation to a random walk of normal increments is studied within the context of sequentially testing for the sign of the drift or mean. These papers provide explicit characterizations and expansions for the optimal stopping regions, along with rates of approximation.

A second role for Brownian motion has been in direct modeling. For example, De Groot (1960) studies sequential tests for a one-sided hypothesis-testing problem involving directly the parameter μ in the context of Brownian motion with drift, $\{Z(t) + \mu t : t \geq 0\}$.

The subject of this paper represents a third role, namely, the role of Brownian motion in suggesting methodology for more difficult discrete problems. The main asymptotic challenge for the Binomial sequential estimation problems of Hubert and Pyke (1995) involved uniform integrability of loss functions and these remained intractable until the problem was translated into the Brownian framework. The self-similarity of Brownian motion then enabled one to transform these problems into ones involving a family of stopping curves for a single process, a significant simplification of the original single stopping curve for a family of processes. (By the way, self-similarity for Brownian motion is another of the central concepts for Brownian motion that appears already in Bachelier (1900); see e.g. page 76 where in the context of applications to his just derived distribution of the first passage time, he emphasizes that the probability is unchanged if the height of the barrier is changed proportionally to the square-root of time.)

In Section 2 we state briefly the motivating Binomial problem, give the appropriate continuous problem in terms of Brownian motion, and prove the strong law and asymptotic normality of the resulting stopping times. The main theoretical result, the uniform integrability of these stopping times is presented in Section 3. A brief addendum follows in Section 4 in which the second author summarizes some further comments made during the oral presentation and includes statements of open problems that were outlined briefly at that time.

2. The Problem

The Binomial inference problem studied in Hubert and Pyke (1995) is as follows: Let S_n be the partial sums of independent Bernoulli random variables X_i with success probability p, so that $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$, with $P(X_i = 1) = p = P(X_i \neq 0)$. Set $S_0 = 0$. Given constants a, b, and c with $a \neq 0$, $2a+b < 0$, and $c > 0$, we wish to estimate the power of p, $g(p) = p^a$, in such a way as to keep the *risk* close to the specified constant c where risk $= E(L_n(t_n, p))$ and the loss function is $L_n(t_n, p) = h(p)(t_n - g(p))^2$ with weight function $h(p) = p^b$. We consider the estimator $t_n = g(\overline{X}_n)$ where $\overline{X}_n = S_n/n$. Since the approximate ideal fixed-sample size for a specific (but unknown) p can be shown to be

$$\xi_p := (a^2/c)p^{2a+b-1} = (a^2/c)p^{-1/(1-\rho)}, \tag{2.1}$$

where

$$0 < \rho := \frac{2a+b}{2a+b-1} < 1, \tag{2.2}$$

we consider in Hubert and Pyke (1995) the stopping time N defined by $N = \inf\{n > 0 : n \geq \xi_{\overline{X}_n}\}$. By the definition of ξ_p, this stopping time is expressible directly in terms of the partial sums as

$$N = \inf\{n > 0 : S_n \geq C(n)\}, \tag{2.3}$$

where

$$C(n) := (a^2/c)^{1-\rho} n^{\rho}. \tag{2.4}$$

The stopping rule is to stop as soon as one of the partial sums, S_n, is greater than the curve, $C(n)$. Since $0 < \rho < 1$ the curve is increasing and concave.

We use this definition to motivate the definition of a similar stopping time for a family of Brownian motions. Since it becomes necessary to restrict N to be no less than a minimum positive size, m say, replace N with

$$N' = \inf\{n \geq m : S_n \geq C(n)\} \tag{2.5}$$
$$= \inf\{n \geq m : \frac{S_n - np}{(pq)^{1/2}} \geq \frac{C(n) - np}{(pq)^{1/2}}\}.$$

If $Z = \{Z(t) : t \geq 0\}$ is standard Brownian motion the analogous continuous version of the stopping time N' would be

$$\widetilde{T}_p = \inf\{t \geq m : Z(t) \geq \widetilde{C}_p(t)\}, \tag{2.6}$$

where

$$\widetilde{C}_p(t) = \frac{C(t) - tp}{(pq)^{1/2}}. \tag{2.7}$$

Note that ξ_p is a root of $\widetilde{C}_p$.

A simplifying feature of this Brownian motion formulation of the stopping problem is the fact that Brownian motion is self-similar. This allows one to rescale the time axis so that the point ξ_p is always at 1. To this end, set

$$C_p(t) = \xi_p^{-1/2} \widetilde{C}_p(t\xi_p) = (a^2/cq)^{1/2} p^{-\rho/2(1-\rho)}(t^\rho - t) \tag{2.8}$$

and since $\xi_p^{-1/2} Z(t\xi_p)$ is still standard Brownian motion, the appropriate stopping time is expressible as

$$T_p = \inf\{t \geq m_p : Z(t) \geq C_p(t)\}, \tag{2.9}$$

where

$$m_p = \frac{m}{\xi_p} = m/(a^2/c)p^{-1/(1-\rho)}. \tag{2.10}$$

Observe that T_p has the same distribution as $\widetilde{T}_p/\xi_p$. We note here that $\xi_p \to \infty$ as $p \to 0$ so our minimum stopping point, m_p, is converging to zero. The important advantage of this formulation is that all of the stopping problems are now defined on a single random process Z with the parameter p entering only through the curves, $\{C_p\}$. This is not easily possible in the above discrete problem where there is a single curve C, and hence a single stopping time N, but a family of different probability models indexed by p.

In this paper we focus on the Brownian motion problem alone, and therefore we delete the inconsequential factor of $q^{-1/2}$ from the definition of C_p. Note that this factor approaches 1 as $p \to 0$. Thus for the remainder of the paper, let $A = (a^2/c)^{1/2}$ and assume the family of stopping curves is given by

$$C_p(t) := Ap^{-r}(t^\rho - t); t \geq 0 \tag{2.11}$$

for $A > 0$ and parameters $0 < p \le 1$, $0 < \rho < 1$ and $r = r(\rho) = \rho/2(1-\rho) = -a - b/2$. Note that r is increasing in ρ with range $(0, \infty)$, and $r(2/3) = 1$. Also, C_p is increasing on $(0, \rho^{1/(1-\rho)})$ and decreasing on $(\rho^{1/(1-\rho)}, \infty)$.

We conclude this section by establishing the strong law and asymptotic normality for the stopping times, T_p.

Theorem 2.1. (i) $T_p \to 1$ *almost surely as* $p \to 0$.

(ii) *As* $p \to 0$, $\tau_p \equiv \frac{T_p - 1}{\sigma_p} \xrightarrow{L} N(0,1)$ *where* ξ_p *is defined in* (2.1) *and*

$$\sigma_p = A^{-1}(1-\rho)^{-1}p^r. \tag{2.12}$$

Proof. For (i), the simple form of the curves C_p makes it clear that for any $t \in (0,1)$, $C_p(t)$ increases to $+\infty$ as $p \searrow 0$ while for any $t > 1$, $C_p(t)$ decreases to $-\infty$. If we ignore for the moment the possibility of Z hitting C_p near the origin, we can almost surely force T_p to be arbitrarily close to 1; given any interval around 1, one may take p so small that the curve C_p is steep enough to insure that Z first hits the curve within that interval. Therefore, the proof will be complete if we show that with probability 1, $Z(m_p) < C_p(m_p)$ for all sufficiently small p, thereby preventing stops near the origin. Set $L(t) = t^{1/2}\ln(1/t)$ for $0 < t < 1$. By the LIL, there exists a (random) τ such that $Z(t) < L(t)$ for all $t \le \tau$. Hence

$$Z(m_p) < L(m_p) \tag{2.13}$$

for all p such that $m_p \le \tau$, say for all $p \le p^*(\tau)$. It is clear from the simple form of C_p that if $\rho < 1/2$ then the C_p curve falls completely above the L curve for p sufficiently small. For $\rho \ge 1/2$, consider the ratio

$$\frac{C_p(t)}{L(t)} = Ap^{-r}t^{\rho-1/2}(1-t^{1-\rho})/\ln(1/t). \tag{2.14}$$

At $t = m_p = m/\xi_p = mA^{-1}p^{1/(1-\rho)}$, the factor

$$p^{-r}t^{\rho-\frac{1}{2}} = \text{const.} \times p^{-r}p^{(\rho-\frac{1}{2})/(1-\rho)} = \text{const.} \times p^{-1/2}$$

and so $C_p(m_p)/L(m_p) \to +\infty$ as $p \to 0$. Therefore,

$$L(m_p) < \frac{1}{2}C_p(m_p) \tag{2.15}$$

for sufficiently small p. Combining (2.13) and (2.15) gives $Z(m_p) < C_p(m_p)$ for all sufficiently small p, which together with the monotoniety of C_p completes the proof of (i).

We note from (2.14) that if $\rho \ge 1/2$, $C_p(t)/L(t) \to 0$ as $t \to 0$ so that the restriction $T_p \ge m_p$ is necessary in this case. We also note in passing that in

Theorem 2.1 the minimum stopping time, m_p, could have been somewhat smaller without invalidating the result. For $\rho = 1/2$ any positive power of p would be sufficient and for $\rho > 1/2$ any power of p, say p^α, with $\alpha < \rho/\{(1-\rho)(2\rho-1)\}$ would be sufficient.

The proof of (ii) follows directly from (i) by utilizing a Taylor's expansion of $C_p(t)$ around $t = 1$. By (2.11)

$$\begin{aligned} C_p'(t) &= Ap^{-r}(\rho t^{\rho-1} - 1) \\ &= \sigma_p^{-1}\left\{1 + \frac{\rho}{\rho-1}(t^{\rho-1} - 1)\right\} = \sigma_p^{-1}(1 + o(1)). \end{aligned}$$

Hence $Z(T_p) = C_p(T_p) = (T_p - 1)\sigma_p^{-1}(1 + o(1))$, where by (i) the random remainder term in this last expression converges a.s. to zero. Moreover, by (i) and the continuity of Z, $Z(T_p) \to Z(1)$ a.s. The proof is complete.

3. Uniform Integrability of the Stopping Times

In the motivating discrete problem described at the start of Section 2, the weighted loss function of interest becomes $p^b((\overline{X}_N)^a - p^a)^2$ when estimating p^a by $(\overline{X}_N)^a$. However, since S_N differs from $C(S_N)$ only by the amount of the "jump over the boundary", this loss is approximately

$$L_n(t_n, p) = p^b(\{C(S_N)/N\}^a - p^a)^2. \tag{3.1}$$

To translate this into the continuous model, use (2.7) and (2.11) to get

$$\frac{C_p(T_p)}{T_p} = \frac{\widetilde{C}_p(\widetilde{T}_p)\sqrt{\xi_p}}{\widetilde{T}_p} = \left\{\frac{C(\widetilde{T}_p)}{\widetilde{T}_p} - p\right\}(\xi_p/p)^{1/2} \tag{3.2}$$

noting that the factor $q^{1/2}$ is being omitted. Thus together with (2.1) and (2.11), relation (3.2) implies

$$\frac{C(\widetilde{T}_p)}{\widetilde{T}_p} = (p/\xi_p)^{1/2}\frac{C_p(T_p)}{T_p} + p = pT_p^{-(1-\rho)}. \tag{3.3}$$

The loss function (3.1) then translates into

$$L_p \equiv L(T_p, p) := p^b(p^a T_p^{-a(1-\rho)} - p^a)^2 = p^{-2r}(T_p^{-a(1-\rho)} - 1)^2. \tag{3.4}$$

The stopping curve C, in the discrete Binomial case, is chosen to insure that the approximate risk for each fixed $p \in (0,1)$ is equal to a specified constant, c say. In Hubert and Pyke (1995) it is shown that the exact risk converges to c as $p \searrow 0$. To prove this result, it was necessary to establish first that the losses are uniformly integrable in p, and this required a lengthy argument for which the following somewhat simpler situation provided the successful outline.

Define $V_p = p^{-r}(T_p^{-a(1-\rho)} - 1)$ so that $L_p = V_p^2$ and note that by part (ii) of Theorem 2.1, it follows directly that $V_p \xrightarrow{L} (-a/A)Z$ where Z denotes a $N(0,1)$ r.v. Consequently, $L_p \xrightarrow{L} (a^2/A^2)Z^2 = cZ^2$ since we introduced the constant A to stand for $\sqrt{a^2/c}$. It is the purpose of this section to prove that the "risk" $E(L_p) \to E(cZ^2) \equiv c$ as $p \to 0$. To do this, we will obtain rather precise tail estimates of the distributions of V_p, from which their uniform integrability may be deduced. The results to be proved are as follows.

Theorem 3.1. *The family* $\{|V_p|^s : 0 < p \leq p_0\}$ *is uniformly integrable for any* $p_0 \in (0,1)$ *and* $s > 0$.

Theorem 3.2. *There exist positive constants* K, K_1, K_2, K_3 *and* $0 < \varepsilon < 1$ *such that for all* $u \geq 0$

(i) $P(|V_p| > u, T_p < 1) \leq e^{-K_1 p^{-(1-\varepsilon)/2}} + 4e^{-K_2 p^{-2\varepsilon r}} + 4e^{-K_3 u^2}$
for p *sufficiently small, and*

(ii) $P(|V_p| > u, T_p \geq 1) \leq 2e^{-Ku^{2\gamma}}$ *for all* $p \in (0,1)$, *where*

$$\gamma = \begin{cases} 1, & \text{if } a > 0, \\ \min(1, -1/2a(1-\rho)), & \text{if } a < 0. \end{cases}$$

The first theorem follows directly from the bounds obtained in the second, as we now show. For notational convenience let $Z_p = |V_p|$.

Proof of Theorem 3.1. It suffices to consider the uniform boundedness of EZ_o^{s+1}. Write

$$\begin{aligned} EZ_p^{s+1} &= \int_0^\infty P(Z_p > u)\,du^{s+1} \\ &= \int_0^\infty P(Z_p > u, T_p < 1)\,du^{s+1} + \int_0^\infty P(Z_p > u, T_p \geq 1)\,du^{s+1}. \end{aligned} \tag{3.5}$$

It is clear from the bounds of Theorem 3.2 that the second integral is uniformly bounded in p. For the first integral the boundedness comes from showing that the range of integration is bounded. For $a < 0$, observe that whenever $T_p < 1$, $Z_p = p^{-r}|T_p^{-a(1-\rho)} - 1| \leq p^{-r}$. Hence the integrand is zero unless $u \leq p^{-r}$. For $a > 0$, the event in the first integral in (3.5) is

$$\begin{aligned} [\,p^{-r}|T_p^{-a(1-\rho)} - 1| > u, T_p < 1] &= [\,p^{-r}(T_p^{-a(1-\rho)} - 1) > u] \\ &= [T_p < (1 + up^r)^{-1/a(1-\rho)}] \end{aligned} \tag{3.6}$$

since $(1 + up^r)^{-1/a(1-\rho)}$ decreases as $u \to \infty$. Also, recall that there is a lower limit of m_p on the stopping time T_p. From (2.1) and (2.10), the inequality

$(1+up^r)^{-1/a(1-\rho)} \geq m_p$ is equivalent to $u \leq (m_p^{-a(1-\rho)} - 1)p^{-r}$. The latter shows that the range of integration is covered by

$$0 \leq u \leq m_p^{-a(1-\rho)} p^{-r} \equiv (m/A^2)^{-a(1-\rho)} p^{b/2}.$$

We use K, with or without subscripts, as generic positive constants which may have different values at different times. Thus in either case, the range of integration for the first integral may be taken as $0 \leq u \leq Kp^{-\beta}$ for suitable positive constants K and β. It then follows directly from the form of the bounds in Theorem 3.2 that $\{EZ_p^{s+1}; 0 \leq p \leq p_0\}$ are bounded where p_0 is implicit in Theorem 3.2 (i).

The particular consequence of this result that is of methodological interest for the sequential analysis context is the following:

Corollary 3.1. *The family of loss functions* $\{L_p\}$*, as defined in* (3.4)*, is uniformly integrable with*

$$\lim_{p \to 0} E(L_p) = c.$$

Proof. Since L_p is just the square of Z_p the family $\{L_p\}$ is uniformly integrable by Theorem 3.1. The result is then a consequence of the convergence in law of $\{L_p\}$ given in Part (ii) of Theorem 2.1 and the construction of the stopping curves.

In the remainder of this section, we give the derivation of the probability bounds used in the proof of uniform integrability.

Proof of Theorem 3.2. There are small differences that arise in the proof depending on the sign of a so it is necessary to split the proof into two cases.

Case 1. $a > 0$. From (3.6)

$$[Z_p > u, T_p < 1] = [T_p < t_{p^-}(u)], \tag{3.7}$$

where

$$t_{p^-}(u) \equiv (1 + up^r)^{-1/a(1-\rho)}.$$

Different bounds turn out to be necessary for different ranges of u. Recall that the stopping curves C_p rise steeply from zero at $t = 1$ to arbitrarily great heights (as $p \to 0$) before returning steeply to pass through zero again at $t = 1$. Thus it is very unlikely that the Brownian motion will hit the curve anywhere but near the origin or near 1. Note that if the Brownian motion Z is above the stopping curve C_p for any t between m_p and $t_{p^-}(M)$ then by definition Z_p is larger than M and this is the event we are integrating over.

To get suitable bounds we first split the range of u's into three pieces. It is easier and more useful to think of splitting the range of t's into three intervals. First of all, fix $\varepsilon \in (0,1)$ and set

$$v_p = kp^{\frac{1-\varepsilon}{2(1-\rho)}} = kp^{(1-\varepsilon)(r+1/2)}, \tag{3.8}$$

where k is a positive constant. Note that the maximum value of C_p occurs at $t_{\max} = \rho^{1/(1-\rho)}$ so that v_p is to the left of $t_{\max}$ for p sufficiently small. Let w_p be the unique value greater than $t_{\max}$ which satisfies $C_p(v_p) = C_p(w_p)$. The three regions into which we split the t interval are delineated for p sufficiently small by the points $0, m_p, v_p, w_p$, and 1. In order to talk about the u range being split into three intervals, define u_1 and u_2 as the u's which satisfy the equations: $t_{p^-}(u_1) = v_p,$ and $t_{p^-}(u_2) = w_p$.

The range of integration is then split into three parts corresponding to the labels I, II, and III as follows

$$\text{(I): } u_1 < u; \quad \text{(II): } u_2 < u \le u_1; \quad \text{(III): } 0 \le u \le u_2.$$

It will be seen that the ε in the definition of v_p is needed only for interval II. It could just as well be omitted for intervals I and III.

First of all, consider u such that $t_{p^-}(u) < m_p$. Since T_p is bounded below by m_p, $P(T_p < t_{p^-}(u)) = 0$, yielding a trivial bound in this case. Next, consider a u in interval I. In this case, $m_p \le t_{p^-}(u) < v_p$. If $T_p < t_{p^-}(u)$ then $Z(t) \ge C_p(t)$ for some $m_p \le t < t_{p^-}(u)$, which implies that $Z(t) \ge L_1(t)$ for some $m_p \le t < t_{p^-}(u)$ where L_1 is the line through the points $(m_p, C_p(m_p))$ and $(v_p, C_p(v_p))$. The probability of Z being above the line L_1 somewhere between m_p and $t_{p^-}(u)$ is less than or equal to the probability of Z crossing above the line somewhere from 0 to $+\infty$. A well known expression of Doob (1949) shows that the probability that Z ever crosses a line with positive slope, S, and positive intercept, I, is

$$P(Z(t) \ge L(t), \text{ for some } t > 0) = e^{-2SI}. \tag{3.9}$$

For the particular line L_1, the slope is

$$S = \frac{C_p(v_p) - C_p(m_p)}{v_p - m_p} \ge \frac{C_p(v_p) - C_p(m_p)}{v_p}. \tag{3.10}$$

The numerator of (3.10) satisfies

$$\begin{aligned} C_p(v_p) - C_p(m_p) &= Ap^{-r}\{(v_p^\rho - v_p) - (m_p^\rho - m_p)\} \\ &\ge Ap^{-r}(v_p^\rho/2 - m_p^\rho) \end{aligned}$$

for sufficiently small p. This last expression, for constants K_1, K_2 and K, equals

$$Ap^{-r}\{K_1 p^{(1-\varepsilon)r} - K_2 p^{2r}\} = Ap^{-\varepsilon r}\{K_1 - K_2 p^{(1+\varepsilon)r}\} \ge Kp^{-\varepsilon r}$$

for sufficiently small p. Thus from (3.10), the slope of L_1 satisfies

$$S \geq \frac{Kp^{-\varepsilon r}}{v_p} = Kp^{\frac{-1+\varepsilon(1-\rho)}{2(1-\rho)}} = Kp^{-r-\frac{1}{2}+\frac{\varepsilon}{2}}, \tag{3.11}$$

where K is used generically. By solving the equation of the line for the intercept we get

$$I = C_p(m_p) - m_p S. \tag{3.12}$$

To give a lower bound to this, we need an upper bound on the slope in addition to the above lower bound. To do this, observe that

$$S = \frac{C_p(v_p) - C_p(m_p)}{v_p - m_p} \leq \frac{C_p(v_p)}{v_p - m_p}.$$

But then,

$$\begin{aligned} v_p - m_p &= Kp^{(1-\varepsilon)/2(1-\rho)} - K_1 p^{1/(1-\rho)} \\ &\geq K_2 p^{(1-\varepsilon)/2(1-\rho)} = Kp^{(1-\varepsilon)(r+\frac{1}{2})} \end{aligned}$$

for p sufficiently small. Moreover,

$$C_p(v_p) = Ap^{-r} v_p^{\rho}(1 - v_p^{1-\rho}) \leq Kp^{-r} p^{(1-\varepsilon)r} = Kp^{-\varepsilon r}. \tag{3.13}$$

Consequently, $S \leq Kp^{-r-\frac{1}{2}+\frac{\varepsilon}{2}}$. Substitution of this upper bound into (3.12) yields

$$\begin{aligned} I &\geq C_p(m_p) - m_p K p^{\frac{-1}{2(1-\rho)}+\frac{\varepsilon}{2}} = Ap^{-r} m_p^{\rho}(1 - m_p^{1-\rho}) - m_p K p^{\frac{-1}{2(1-\rho)}+\frac{\varepsilon}{2}} \\ &\geq Kp^r - K_1 p^{\frac{1}{2(1-\rho)}+\frac{\varepsilon}{2}} \geq \frac{1}{2}Kp^r \end{aligned} \tag{3.14}$$

for sufficiently small p. The combination of (3.11) and (3.14) gives $SI \geq Kp^{-(1-\varepsilon)/2}$. Hence, by (3.9)

$$P(Z \text{ crosses line } L_1) \leq e^{-Kp^{-(1-\varepsilon)/2}} \tag{3.15}$$

and thus

$$P(T_p < t_{p^-}(u)) \leq e^{-Kp^{-(1-\varepsilon)/2}}, \text{ for } u > u_1. \tag{3.16}$$

Next, consider $t_{p^-}(u)$ in interval II, i.e. $u_2 < u \leq u_1$. Similar methods for this region yield a bound that is also constant in u and depends only on p. For this interval we are able to use a piece-wise linear lower bound to C_p that involves two lines. The first line is the same one used on interval I and the second is just the horizontal line of height $C_p(v_p)$.

The probability of Z crossing above the piece-wise boundary formed by L_1 and L_2 between m_p and $t_{p^-}(u)$ is less than or equal to the probability of crossing

L_1 anywhere from 0 to $+\infty$, plus the probability of crossing L_2 between 0 and 1. The first probability is the one for which we just computed a bound in (3.15). The second is just the probability of the Brownian motion Z crossing a horizontal line. It is well known that this probability equals twice the probability of being above the line at 1. The height of the horizontal line is $C_p(v_p)$ so the probability of crossing above L_2 equals

$$2P(Z(1) \geq C_p(v_p)) \leq 4e^{-C_p^2(v_p)/2}. \tag{3.17}$$

But, similarly to (3.13), one obtains $C_p(v_p) = Ap^{-r}v_p^{\rho}(1 - v_p^{1-\rho}) \geq Kp^{-\varepsilon r}$ for p sufficiently small. Thus for the second interval we have

$$P(T_p < t_{p^-}(u)) \leq e^{-K_1 p^{-(1-\varepsilon)/2}} + 4e^{-K_2 p^{-2\varepsilon r}}, \text{ for } u_2 < u \leq u_1. \tag{3.18}$$

As this shows, it was for this second interval that we had to introduce ε.

Now consider $t_{p^-}(u)$ in interval III, namely, $0 \leq u \leq u_2$. For this interval the bound depends on the specific value of u. The picture is similar to the last one. We once again use a piece-wise linear lower bound to C_p that involves two lines, namely, the same sloping line L_1 as before together with the horizontal line L_3 of height $C_p(t_{p^-}(u))$. Note that L_3 is necessarily lower than L_2. Use the same bound as before for the probability of crossing L_1. The probability of exceeding L_3, as for L_2, can be shown to be bounded by $4e^{-C_p^2(t_{p^-}(u))/2}$ (see (3.17)). However, $C_p(t_{p^-}(u))$ is rather complex here, and some additional work is required to find a suitable lower bound for it.

Observe first that since $v_p \to 0$ as $p \to 0$, the structure of the curves C_p implies that $w_p \to 1$ as $p \to 0$. Though it is not possible to solve for w_p explicitly in the defining equation $C_p(v_p) = C_p(w_p)$, we only need an approximate value, $\overline{w}_p$, that is less than w_p. If one ignores higher order terms in the defining equation, the value

$$\overline{w}_p := (1 - 2v_p^{\rho})^{1/(1-\rho)} \tag{3.19}$$

is suggested. To see that $\overline{w}_p < w_p$, note firstly that $\overline{w}_p > t_{\max}$ when p is small, so that it suffices to check that $C_p(\overline{w}_p) > C_p(w_p)$. But

$$\frac{C_p(\overline{w}_p)}{C_p(w_p)} = \frac{\overline{w}_p^{\rho}}{1 - v_p^{1-\rho}} \cdot \frac{1 - \overline{w}_p^{1-\rho}}{v_p^{\rho}} = 2\frac{\overline{w}_p^{\rho}}{1 - v_p^{1-\rho}} > 1$$

for p sufficiently small since $\overline{w}_p \to 1$ and $v_p \to 0$. Define $\overline{u}_p$ by $t_{p^-}(\overline{u}_p) = \overline{w}_p$. Then by the definition of t_{p^-} in (3.7), $\overline{u}_p = o(p^{-r})$ since $\overline{w}_p \to 1$. Also, since $u < \overline{u}_p$ for all u's in interval III, it follows that up^r converges to zero uniformly over $0 \leq u \leq u_2$; recall that $u_2 = u_2(p)$ depends on p, and could be diverging.

Consequently, $t_{p^-}(u) \to 1$ uniformly for $0 \le u \le u_2$. We can now get the desired bound, since

$$\begin{aligned} C_p(t_{p^-}(u)) &\equiv Ap^{-r} t_{p^-}^{\rho}(u)(1 - (t_{p^-}(u))^{1-\rho}) \\ &\ge \frac{1}{2} Ap^{-r}\{1 - (1 + up^r)^{-1/a}\} \end{aligned}$$

and the uniform convergence to zero of up^r then gives

$$C_p(t_{p^-}(u)) \ge Ku, \qquad 0 \le u \le u_2, \tag{3.20}$$

for sufficiently small p and a suitable constant K. (Note that we could have used $A - \delta$ for any $\delta \in (0,1)$ in place of $A/2$ above, so that $A/a = c^{-1/2}$ could be used for K in (3.20).) In view of (3.15) and (3.20), this proves

$$P(T_p < t_{p^-}(u)) \le e^{-K_1 p^{-(1-\varepsilon)/2}} + 4e^{-K_2 u^2}; \qquad 0 \le u \le u_2. \tag{3.21}$$

A combination of the bounds for the three cases then yields the desired bound (i) of the theorem.

To verify the bound in (ii) for Case 1 of $a > 0$, observe that for this other tail,

$$[Z_p > u, T_p \ge 1] = [\,p^{-r}(1 - T_p^{-a(1-\rho)}) > u] = [T_p^{-a(1-\rho)} < 1 - up^r].$$

Note that this is the empty event if $up^r \ge 1$. Assume therefore that $0 < up^r < 1$ and write

$$[Z_p > u, T_p \ge 1] = [T_p > t_{p^+}(u)], \tag{3.22}$$

where

$$t_{p^+}(u) = (1 - up^r)^{-1/a(1-\rho)}. \tag{3.23}$$

Clearly, $[T_p > t \ge 1] \subset [Z(t) \le C_p(t)]$, so that

$$\begin{aligned} P[T_p > t_{p^+}(u)] &\le P[Z(t_{p^+}(u)) \le C_p(t_{p^+}(u))] = \Phi\Big(C_p(t_{p^+}(u))/\sqrt{t_{p^+}(u)}\,\Big) \\ &\le 2\exp\{-C_p^2(t_{p^+}(u))/2t_{p^+}(u)\}. \end{aligned} \tag{3.24}$$

By setting $x = up^r$, direct substitution yields

$$\begin{aligned} C_p(t_{p^+}(u))/\sqrt{t_{p^+}(u)} &= Aux^{-1}(1-x)^{(-2\rho+1)/2a(1-\rho)}\{1 - (1-x)^{-1/a}\} \\ &\equiv -AuJ(x), \text{ say.} \end{aligned} \tag{3.25}$$

It is easy to check that $J(1-) = +\infty$, $J(0+) = 1/a$, and J is continuous and never zero on $0 < x < 1$. Thus J is bounded away from zero, implying from (3.22), (3.24) and (3.25) that for all $p \in (0,1)$

$$P[Z_p > u, T_p \ge 1] \le 2e^{-Ku^2}, \qquad 0 < u < p^{-r}.$$

Recall that a bound of zero applies for $u \geq p^{-r}$, and this completes the proof when $a > 0$.

Case 2. $a < 0$. The main difference from Case 1 is in the definitions of $t_{p^-}(u)$ and $t_{p^+}(u)$, where the change in the sign of a effectively amounts to a substitution of $-u$ for u; namely

$$[Z_p > u, T_p < 1] = [T_p < t_{p^-}(u)] \text{ with } t_{p^-}(u) = (1 - up^r)^{-1/a(1-\rho)} \tag{3.26}$$

when $up^r < 1$, and for all u,

$$[Z_p > u, T_p \geq 1] = [T_p > t_{p^+}(u)] \text{ with } t_{p^+}(u) = (1 + up^r)^{-1/a(1-\rho)}. \tag{3.27}$$

These may be compared with (3.7) and (3.23), respectively. For the proof of (i), the range of u is split as before. The steps for intervals I and II are essentially unchanged since they depend only on the curve C_p and the points m_p, v_p and w_p whose definitions do not change. For interval III, an analogous calculation as before shows that

$$C_p(t_{p^-}(u)) = Ap^{-r}t_{p^-}^{\rho}(u)(1 - t_{p^-}^{1-\rho}(u)) \geq Ku$$

for sufficiently small p. Thus all three terms in the bound remain the same except for the constants, thereby proving (i) when $a < 0$.

For the right hand tail bound of (ii), it follows as before that (3.24) holds, and, moreover, that the term in the exponent can be written as

$$-C_p(t_{p^+}(u))/\sqrt{t_{p^+}(u)} = AuH(x), \tag{3.28}$$

where $x = up^r$,

$$H(x) \equiv x^{-1}(1+x)^{\alpha}\{(1+x)^{-1/a} - 1\} \tag{3.29}$$

and $\alpha = (-2\rho + 1)/2a(1-\rho)$. By (3.25), note that $H(x) = -J(-x)$. Now, however, the range for x is $(0, \infty)$ rather than $(0, 1)$, which forces a significant change in the required lower bound for H. Since $H(0+) = -1/a > 0$ and H is always positive, the lower bound is determined by its behavior as $x \to +\infty$. Clearly,

$$H(x) \geq (1+x)^{\alpha-1/a-1}\{1 - (1+x)^{1/a}\} = 0(x^{\alpha-1/a-1})$$

so that for $\alpha - 1/a \geq 1$, H is bounded below, so that from (3.24) and (3.29),

$$P[T_p > t_{p^+}(u)] \leq 2\exp(-Ku^2) \tag{3.30}$$

when $\alpha - 1/a \geq 1$. For $\alpha - 1/a < 1$, rewrite (3.29) as

$$H(x) = (1+x)^{\alpha-1/a}\{1 - (1+x)^{1/a}\}/x$$

and observe that $H(0+) = -1/a$ and for x large, $H(x) \geq K(1+x)^{\alpha-\frac{1}{a}-1} \geq Kx^{\alpha-\frac{1}{a}-1}$ for suitable constants K. Thus there are constants such that for any $x > 0$, $H(x) \geq \min(Kx^{\alpha-\frac{1}{a}-1}, K')$. Since $\alpha - 1/a = -1/2a(1-\rho)$, it follows that in both cases we have shown that $H(x) \geq Kx^{\gamma-1}$ where $\gamma \equiv \min\{1, -1/2a(1-\rho)\}$ when $a < 0$. Thus from (3.28) applied to (3.24) it follows that

$$P[T_p > t_{p^+}(u)] \leq 2\exp(-Kp^{2(\gamma-1)r}u^{2\gamma}) \tag{3.31}$$

for all $p \in (0,1)$ and $u > 0$. This bound improves as $p \to 0$, but note that since $0 < \gamma \leq 1$, one has $2\exp(-Ku^{2\gamma})$ as a bound independent of p. The proof is complete.

4. Addendum

It was a great privilege for the second author to be able to participate in the conference at Rutgers University to honor Professor Robbins on the occasion of his eightieth birthday. My earliest contacts with Herb would have been through his writings, possibly first (other than through the text he wrote with Courant) while learning about minimum variance estimation in 1954 during graduate school, where the initials in "C-R lower bounds" referred to both Cramér-Rao and Chapman-Robbins. Thereafter, two of the most stimulating early years of my career revolved around the students and faculty of Fayerweather Hall, Columbia University, during 1958–60.

Although the coverage of the Rutgers conference appropriately focused primarily upon statistical topics of central interest to Herb, I hope this paper might serve to emphasize his interest in and contributions to several areas of probability, including martingales, Brownian motion and stopping times. One indication of his contributions to the latter topic were his 1969 IMS Wald Lectures (cf. Robbins (1970) and references therein, including Robbins and Siegmund (1970)). In fact, if one defines a probabilist as one who works in *one* probability space (in contrast to a statistician whose work involves more than one) then I assume that Bayesian statisticians must be viewed as probabilists, thereby expanding greatly Herb's probabilistic contributions.

Throughout his work, Herb shows his special knack for capturing and describing clearly the essence of a new type of problem, often by means of a deceptively simple example. Conversations as well often included the posing of deceptively simple sounding problems whose eventual solutions would frequently require considerable effort and new methodology, sometimes opening up completely new directions for research.

Because of this, I included at the end of the oral presentation of this paper, brief statements of three very simply stated type of problems that to date remain unsolved. The first two of these, involving 2-parameter Brownian-related processes, have been stated elsewhere so are only briefly recalled here with references:

Problem I. *If C is a fixed compact convex subset of $\mathbb{R}^2$ and $Z(t) \equiv Z(C+t)$ is the Brownian* (*white noise*) *measure of this set translated by t, is knowledge of Z over the unit square equivalent almost surely to knowledge of C*? (The answer is yes if C is a finite sided polygon; see Adler and Pyke (1997).)

Problem II. *If Z_1 and Z_2 are independent standard Brownian motions over* $[0,1]$, *then $Z_1 \times Z_2$ is very easily defined almost surely* (*by Fubini and stochastic integration*) *for any convex $C \subset [0,1]^2$. Does there exist a* continuous *version of $Z_1 \times Z_2$ as a process indexed by $\mathcal{C}$, the family of closed convex subsets of* $[0,1]^2$? (Consider $\mathcal{C}$ endowed with the Lebesgue measure of symmetric differences as the metric, for example.) This is a separating example in the sense that the answer is known to be yes (no) if the index family is smaller (larger) than $\mathcal{C}$ in terms of exponents of metric entropy (see Pyke (1992), Question 3, page 260).

The third type of problem presented arose from a paper given by Thomas Bruss just five days earlier at a conference on applied probability and time series held in Athens, Greece, to recognize the contributions of Joe Gani (one of the Fayerweather residents during 1959) and Ted Hannan. The paper, joint with Thomas Ferguson, was appropriately titled, "On Robbins' problem of minimizing the expected rank with full information"; see Bruss and Ferguson (1996). A direct extension of their problem would be the following two-sided version.

Problem IIIa. *If P_n denotes the empirical measure based on n independent uniform-*$[0,1]$ *random variables observed sequentially, how does one choose two of these based only on their "pasts" to maximize $EP_n(J)$, where J is the interval determined by one's choices*?

The problem extends also to cubes by using uniform points in $[0,1]^k$ and letting J be the cube having the two chosen points as opposite vertices. However, I chose to modify the optimality criterion from the expected number of trapped points as given above to the expected volume of the enclosed region. This amounts to using $P(J)$ rather than $EP_n(J)$. (The particular terminology I used was probably a consequence of being in the process of selling our farm.)

Problem IIIb. *Suppose a prespecified number, n, of random locations are marked one after another in a rectangular field. If you were invited to place property stakes sequentially at any two of these locations with the promise of being given the rectangular lot with sides parallel to the original field and having your two stakes at its opposite corners, how should you proceed in order to maximize your lot's expected area*?

What if the original area was a disk (sphere) and the randomly selected subregion is the disk (sphere) having the line segment joining your two chosen points as a diameter? Note that both of these formulations reduce in one dimension to the problem of determining a pair of stopping times $1 \leq M \leq N \leq n$ to maximize $E|U_M - M_N|$ where U_1, $U_2, \ldots, U_n$ are i.i.d. Unif$(0, 1)$ observations.

Here is a general formulation of a related problem that reduces to the above only in the 1-dimensional setting:

Problem IIIc. *If* $U_1, U_2, \ldots, U_n$ *are i.i.d. random variables uniformly distributed over a bounded convex set* $C \subset \mathbb{R}^k$, *how would one determine* (*a random number of*) *stopping times* $1 \leq N_1 \leq N_2 < \cdots < N_L \leq n$ *to maximize* $E|$ *Convex Hull of* $\{U_{N_1}, \ldots, U_{N_L}\}|$? (Here, $|\cdots|$ is Lebesgue measure and your sequential procedure is restricted to insure that all points in $\{U_{N_1}, \ldots, U_{N_L}\}$ are extreme points of their convex hull.)

Acknowledgements

We are indebted to Paul Glasserman for the reference to Mandelbrot (1989) which he communicated to us following the Robbins conference. Discussions with Thomas Bruss related to Problem III above are also gratefully acknowledged, as are the careful readings and comments of the referees, including in particular a simplification in the proof of Theorem 3.1.

References

Adler, R. J. and Pyke, R. (1997). Scanning Brownian processes 39 pages, to appear. *J. App. Prob.*

Anderson, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31**, 165-197.

Bachelier, L. (1900). Théorie de la speculation. *Ann. Sci. École Norm. Sup.* **3**, 21-86 (English translation by A. J. Boness: In *The Random Character of Stock Market Prices*), 17-78, P. H. Cootner, Editor, M.I.T. Press, Cambridge, Mass. (1967).

Brown, R. (1828). *Philos. Math.* **4**, 161-173.

Bruss, T. and Ferguson, T. (1996). Half-prophets and Robbins' problem of minimizing the expected rank. Proc. of the Athens Conf. on Applied Prob. and Time Ser. Anal. vol. 1. Applied Prob. *Lecture Notes in Statist.* No. 114, 1-17. Springer, New York.

Chernoff, H. (1961). Sequential tests for the mean of a normal distribution. *Proc. Fourth Berkeley Symp.* **1**, 79-91. Univ. of California.

Chernoff, H. (1965). Sequential tests for the mean of a normal distribution IV (Discrete case). *Ann. Math. Statist.* **36**, 55-68.

Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. *Ann. Phys.* **17**, 549-560.

DeGroot, M. H. (1960). Minimax sequential tests of some composite hypotheses. *Ann. Math. Statist.* **31**, 1193-1200.

Hubert, S. and Pyke, R. (1995). Sequential estimation of functions of a Binomial parameter. In preparation.

Mandelbrot, B. B. (1989). Louis Bachelier. In *The New Palgrave: Finance.* (Edited by J. Eatwell et al.), 86-88, W. W. Norton, New York.

Pearson, K. (1905). The problem of the random walker. *Nature* **72**, 294.

Pyke, R. (1992). Probability in mathematics and statistics: A century's predictor of future directions. *Jber. d. Dt. Math.-Verein.* Jubiläumstagung 1990. S. 239-264.

Pyke, R. (1996). The early history of Brownian motion. In preparation.

Rayleigh, Lord (1905). *Nature* **72**, 318.

Robbins, H. E. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41**, 1397-1409.

Robbins, H. E. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Statist.* **41**, 1410-1429.

Vogel, W. (1960). A sequential design for the two-armed bandit. *Ann. Math. Statist.* **31**, 430-443.

Networks and Distributed Computing, Box 354841, University of Washington, Seattle, WA 98195-4841, U.S.A.

Department of Mathematics, Box 354350, University of Washington, Seattle, WA 98195-4350, U.S.A.

(Received September 1995; accepted August 1996)