Z. Wahrscheinlichkeitstheorie verw. Geb. 24, 1–35 (1972)

Skorohod Embedding of Multivariate RV's, and the Sample DF

J. Kiefer*

The main purpose of this paper is to study certain representations of sums of iid k-vector rv's as embeddings in k-dimensional Brownian motion by vectors of stopping times, in extension of Skorohod's scheme [20], and consequent error estimates for weak and strong invariance principles. In particular, letting $k\to\infty$ we embed the sample df in the Gaussian process with 2-dimensional time to which it has long been known to converge weakly. We discuss previous sample df embeddings, which have yielded related results; while some of our estimates are slight improvements, the emphasis here will be on the naturality of the embedding per se (although it will be indicated why it is probably far from the final word on the subject.

1. Introduction

Skorohod's embedding scheme has been used and extended in a number of directions. In particular, the original use [20] to obtain error estimates for weak convergence of certain functionals of a sequence of summands of iidrv's has been broadened by Rosenkrantz [15] and Sawyer [17, 18, 19]. On the other hand, Strassen [21, 22] used such embeddings to obtain his strong invariance principles for martingales.

Turning to questions involving the sample df, a number of authors have worked on analogous schemes. The first of these was Breiman [3], who used Skorohod embedding for iidrv's Y_i with $P\{Y_i>y\}=e^{-(y+1)^+}$, and the familiar fact that $\left\{\sum_1^i (Y_j+1)\Big/\sum_1^{n+1}(Y_j+1),\ 1\leqq i\leqq n\right\}$ have the same joint df as the order statistics $\{X_{n,i},\ 1\leqq i\leqq n\}$ from n iidrv's with uniform density on [0, 1], to approximate the sample df deviations by a Brownian bridge. Then Brillinger [4] independently used the same scheme and also gave an upper bound wp 1 on the error. Rosenkrantz [16, Section III] gave essentially the same bound, although his emphasis was on certain weak convergence problems, e.g., for the v. Mises statistic.

The disadvantage of this representation was evidently pointed out by Pyke (according to [4]) and has been discussed further in [12]: while it yields a satisfactory approximation for a single large n, it does not yield the right joint distribution for several large n's at once. Related weak convergence results have been obtained by Bickel, Billingsley, Pyke and Root, among others.

Subsequently Müller [14] gave a proof of the convergence in law of the sample df process to the Gaussian process with two-dimensional time and the proper covariances, that is, with independent "Brownian bridge" increments, and in a striking analysis he gave the first estimate of the error for certain functionals of the *sequence* of sample df's, analogous to the results of [20, 15, 17, 18, 19]. This estimate is based on an embedding of the sample df which uses a well-known

* Research supported by NSF Grant GP 9297.

representation of the $X_{n,i}$ in terms of exponential rv's, different from Breiman's: $\{(n-i)\log(X_{n,n-i+1}/X_{n,n-i}), 0 \leqq i \leqq n-1\}$ $(X_{n,n+1}=1)$ may be taken to be standard iid exponential rv's like Y_i+1 above. This representation, like Breiman's, depends on n in such a way that it also cannot be used for joint (in n) distributions; however, Müller cleverly adds roughly $t N^{1/3-\varepsilon}$ of these, each approximating the sample df for a different set of $n=N^{2/3+\varepsilon}$ observations, to obtain (using further estimates) an approximation of the joint law of the sample df for each integral number tN of observations.

We shall discuss the embeddings of [3, 4, 14] in more detail in Section 5, where some results concerning them will be stated and proved.

The main development of the present paper is the representation of the sample df by a Skorohod-type embedding in the appropriate two-dimensional Gaussian process. Denote the unit interval by I, the reals by R, the nonnegative reals by R^+, the positive integers by Z^+. Let $\xi^*(\cdot,\cdot)$ be a Gaussian process on $I \times R^+$ with continuous sample functions, zero expectation, and

$$E\,\xi^*(s_1, t_1)\,\xi^*(s_2, t_2) = \min(t_1, t_2)\,[\min(s_1, s_2) - s_1 s_2] \tag{1.1}$$

(so that there are independent increments in t and a Brownian bridge in s for fixed t). For future reference we define the closely related

$$\xi(z, t) = (z+1)\,\xi^*\big(z/(z+1), t\big),$$

a continuous Gaussian process on $R^+ \times R^+$ with zero expectations and independent increments in both directions:

$$E\,\xi(z_1, t_1)\,\xi(z_2, t_2) = \min(t_1, t_2)\min(z_1, z_2). \tag{1.2}$$

(We shall always use ξ^* and ξ as in (1.1)-(1.2); univariate Brownian motions, will be denoted ξ_i, ξ_i', etc.) Let S_n be the sample df based on the first n of the iidrv's $\{X_i, i \geqq 1\}$, uniformly distributed on I, with tS_t defined as usual by linear interpolation from nS_n and $(n+1)S_{n+1}$ if $n<t<n+1$, and $S_0=0$. We take the left-continuous version of S_n to conform with the embeddings of Section 2. Let

$$\hat{\xi}(s, t) = t\,[S_t(s) - s], \qquad s \in I,\ t \in R^+. \tag{1.3}$$

A main consequence of our embedding is

Theorem 1. *ξ^* can be defined on a probability space on which there is defined a random function T: $I \times R^+ \to R^+$ such that*

$$\xi^*(s, T(s,t)) \text{ has the same joint law as } \hat{\xi} \text{ of } (1.3); \tag{1.4}$$

and, as $t \to \infty$,

$$t^{-1/2} \sup_{0 \leqq s \leqq 1} \left|\xi^*(s, T(s,t)) - \xi^*(s,t)\right| = O\big(t^{-1/6}(\log t)^{2/3}\big) \text{ wp 1}. \tag{1.5}$$

(It will become clear from the use made of (3.23) that all our results such as (1.5) and (1.6) can be stated in terms of either continuous t or discrete n.)

A corresponding weak law, essentially Müller's Theorem 3 with obvious changes in some of his assumptions (described just above our Lemma 6) and replacement of n^ε by $(\log n)^\lambda$, is

Corollary 1 (to Theorem 2). *For* G_i *continuous on* $I \times R^*$ *where* R^* *is a subinterval of* $(0, +\infty)$ *of positive (possibly infinite) length and* γ *is given by* (3.34),

$$\begin{aligned} P\left\{G_1\left(s, \frac{k}{n}\right) < n^{-1/2} k[S_k(s) - s] < G_2\left(s, \frac{k}{n}\right) \text{ for } s \in I,\ k \in Z^+,\ k/n \in R^*\right\} \\ - P\{G_1(s,t) < \xi^*(s,t) < G_2(s,t) \text{ for } s \in I,\ t \in R^*\} \\ = O\left(n^{-1/6} (\log n)^\gamma\right). \end{aligned} \tag{1.6}$$

The result also holds if only one G_i is present, of course. (Müller does not state an analogue of (1.5), but one can be obtained from his work.) As indicated above, one reason for giving (1.6) is that the embedding is somewhat more natural, the computations simpler, and the source of error perhaps more transparent, than in Müller's ingenious development. Also, the proof of Theorem 2 helps one to understand the more intricate Theorem 1. In Section 6 we will discuss the possibility of improving the estimate from either approach.

The developments of Lemmas 1′ and 4′ allow explicit constants to be obtained for the bounds of (1.5), (1.6), etc. In an already long paper we have not taken the extra space to do this.

The result (1.6) is one of a spectrum of possible results in terms of the degree of fineness (in s) at which the sample df is considered. At the other extreme we have

Corollary 2 (to Lemma 4′). *Under assumption* (3.34) *with* $\alpha > 1/4$ (*but ignoring conditions in* $|s - s'|$ *there*), *for any finite subset* I' *of* I *and some* λ',

$$[\text{left side of (1.6) with } I' \text{ for } I] = O\left(n^{-1/4} (\log n)^{\lambda'}\right) \tag{1.7}$$

and (1.4) *holds with*

$$[\text{left side of (1.5) with } I' \text{ for } I] = O\left(t^{-1/4} (\log t)^{\lambda'}\right) \ \text{wp}\,1. \tag{1.8}$$

When one considers the error term given in multivariate Berry-Esseen results, or in the considerations of Section 7.3 of [20], or of [15, 17, 18] when I' is a single point and G_i = constant and $R^* = (0, 1]$, it is impossible to believe that the bound in (1.7) cannot be replaced by $n^{-1/2} (\log n)^{\lambda'}$. In Section 6 the inability of the Skorohod technique, as used here, to achieve such a better rate, will be analyzed.

In an attempt to extend the results of Section 7.4 of [20] or of [19], one considers $f(t, x, s)$ continuously differentiable in t, x with partial derivatives slowly varying in x (of order $O(|x|^A)$ for some A). Denote the cardinality of I' by $|I'|$. It is then a routine application (which we shall omit) of Lemma 4′ to prove

Corollary 3. *If the second probability of* (1.9) *has a bounded derivative in* λ, *then there is a value* β *such that, uniformly in* λ,

$$\begin{aligned} P\left\{\frac{1}{n|I'|} \sum_{k=1}^{n} \sum_{s \in I'} f\left(\frac{k}{n}, n^{-1/2} k[S_k(s) - s], s\right) \le \lambda\right\} \\ - P\left\{\frac{1}{n|I'|} \sum_{s \in I'} \int_0^1 f(t, \xi^*(s,t), s)\, dt \le \lambda\right\} = O\left(n^{-1/4} (\log n)^\beta\right). \end{aligned} \tag{1.9}$$

When I' consists of a single element, a very special case of [19] yields (1.9) with the exponent 1/4 replaced by 1/2. (The form of the corresponding statement in [20] is slightly different, but the appropriate value is again 1/2.) Thus, again, the Skorohod technique fails to yield what we believe to be the correct result in (1.9), for reasons described in Section 6. Moreover, it is surely tempting to conjecture that, extending the differentiability assumption to s, one can obtain an error estimate $O(n^{-1/2}(\log n)^\beta)$ with $\sum_s$ replaced by $\int ds$ in (1.9). Using the embedding of the present paper and second derivatives slowly varying, the author can at present only prove

Theorem 3. *Under the above assumptions, if also the second probability below has bounded derivative in λ, then there is a β such that, uniformly in λ,*

$$\begin{aligned} P\left\{\int_0^1 n^{-1}\sum_{k=1}^n f\left(\frac{k}{n}, n^{-1/2}k[S_k(s)-s], s\right) ds \leqq \lambda\right\} \\ -P\left\{\int_0^1\int_0^1 f(t, \xi^*(s,t), s)\, dt\, ds \leqq \lambda\right\} = O(n^{-1/5}(\log n)^\beta). \end{aligned} \tag{1.10}$$

This is slightly better than what one would get directly from Theorem 1 or [14]; the additional strength comes from the possibility of using $B_n = O(n^{1/5}(\log n)^{\beta'})$ in place of the $O([n/\log n]^{1/3})$ which will be used in proving Theorem 2, and this in turn is made possible by using Lemma 1′ and two derivatives in s to estimate the difference of the rv's obtained by integrating over s for fixed k (or t) and corresponding sums over B_n points. The harder part of the proof is essentially in Sawyer's work [19], and we will not include the details in this already long paper, especially since the chief novelty of our development is present in the embedding and proof of Theorem 2, and the conclusion (1.10) seems far from definitive.

The above can be viewed as an extension of Rosenkrantz's result for fixed $k=n$ (or t); while stated in [16] for the von Mises statistic, the result is more general:

Theorem 4 (Rosenkrantz). *If $f \in C^2(R \times I)$ with partial derivatives of slow growth in the unbounded variable, and if the second probability below has bounded derivative in λ, then there is a value $\beta > 0$ such that*

$$\begin{aligned} P\left\{\int_0^1 f(n^{1/2}[S_n(s)-s], s)\, ds \leqq \lambda\right\} - P\left\{\int_0^1 f(\xi^*(s,1), s)\, ds \leqq \lambda\right\} \\ = O(n^{-1/4}(\log n)^\beta). \end{aligned} \tag{1.11}$$

This can be proved using either Breiman's embedding as in [16], or Müller's. Either method also yields the order of (1.11) for the modification of (1.6) obtained by restricting k to the single value n; this conclusion is contained in [14]. A defect of the embedding of the present paper is that, although it yielded a better result than Theorem 1 or direct application of [14] in Theorem 3, it yields the same order as in (1.10) if used in the context of Theorem 4 where the other embeddings do better. As Rosenkrantz points out, the result of Chan-Li-Tsian [7] for approximating the Kolmogorov distribution suggests that $n^{-1/2}(\log n)^\beta$ is again the desired result, and none of the methods yields that at present. This will be discussed further in Section 6.

Remarks on Theorems 3 and 4. While we have not obtained definitive results in this domain, the subject seems important enough to deserve certain comments.

(1) Values of β can be discerned from [19] and the proofs described above. The condition that $f(t, x, s)$ has partials in t and s that are $O(|x|^A)$ can be much weakened without affecting the conclusion. Moreover, the assumption on $\partial f/\partial x$ in (1.9) and (1.10) can be weakened greatly at the expense of obtaining the slightly weaker conclusion $O(n^{-1/5+o(1)})$.

(2) Theorems 3 and 4 are easily modified to allow a finite measure, $\mu(ds)$, to replace ds (and similarly for t). We shall discuss elsewhere the analogue of (1.11) when S_n is replaced by the sample quantile process (essentially S_n^{-1}); (1.11) and a result of [11] immediately yield $O(n^{-1/4}(\log n)^\beta)$ as an estimate of error in this case. In particular, Chernoff-Savage (linear rank) statistics and corresponding location parameter estimators linear in the sample quantiles, for $1-$ and $2-$ sample problems, can be treated in this way. One would hope for better bounds, as illustrated in the easily obtained $O(n^{-1/2})$ for linear combinations of a fixed number of sample quantiles.

(3) We forego the statement and proof of almost sure analogues of (1.9)–(1.11), which are obvious via Borel-Cantelli. As with (1.5), these are statements about imperfect methods, more than anything of intrinsic meaning. See also [19], Corollary 2, for the more definitive result obtained in the case studied there, and [10] regarding limitations of the Skorohod technique discussed elsewhere herein.

(4) One can often obtain the required boundedness of the limiting density function in (1.9)–(1.11) by well known Fourier-analytic techniques we shall not discuss. Lemma 6 treats the corresponding problem for the results of (1.6)–(1.7).

(5) Just as (1.9) with I' a single point is a very special case of Section 7.4 of [20] or of [19], so (1.10) and (1.11) have extensions to cases where $\{S_1(s), s\in I\}$ is replaced by another continuous time process, and $n S_n$ by the sum of n such iid processes. As mentioned in the next section, the crucial thing is that the martingale $\{(z+1)\, S_1(z/(z+1))-z, z\in R^+\}$ be replaced by another martingale, the bounds obtainable depending on continuity properties of the latter. We treat this further in [13]. If, instead, one considers a martingale whose time domain is a discrete set of h points (possibly $h=\infty$), the embedding of Section 2.2 yields approximation theorems for partial sums of the corresponding iid h-vectors with finite fourth moments. Unfortunately, the error term obtained for the normal approximation by this method is again limited to $O(n^{-1/4}(\log n)^\beta)$, compared with the $n^{-1/2}$ of the multivariate Berry-Esseen bound. Section 2.2 also treats k-dimensional time.

The proof of (1.6), once T has been defined, is not too difficult, involving only a slightly delicate balance of several error terms, which are estimated by adopting techniques used in [8] and [9]. The main difficulty is in defining T properly. This involves a Skorohod-type embedding in which we consider simultaneously several stopping times in order to get a vector rv (Section 2.2); in the sample df case, we are interested in the infinite-dimensional rv $\{S_1(s)-s, s\in I\}$. This technique is not in itself so surprising, and we shall elsewhere treat other such multivariate embeddings [13]. The difficulty here is not so much in defining stopping times which yield a representation of S_n, as in identifying where $\xi^*(s,t)$ sits in the resulting picture. We shall discuss the embedding in Section 2.3; an alternate one can be obtained corresponding to the T_p^* of Section 2.1. Theorem 2 will be proved in Section 3; Theorem 1, whose development requires slightly more technical complication, is treated in Section 4. We have tried to spare the reader as much pain as possible by omitting straightforward but long arithmetic in proofs of Lemmas 1′ and 4′, Theorem 1, etc. whenever the ideas are present in earlier proofs.

We turn now to a simple result about approximating S_n, which has nothing to do with embedding but is used in proving the theorems of this paper. (Indeed, we have previously used such estimates elsewhere.) Let B be a positive integer and define $S_{B,n}$ and ξ_B^* by

$$S_{B,n}(x)=\begin{cases} S_n(x) & \text{if } \; x=i/B, \quad 0\leqq i\leqq B; \\ \text{linear for} & i/B\leqq x\leqq (i+1)/B, \quad 0\leqq i<B. \end{cases} \tag{1.12}$$

$$\xi_B^*(x,t)=\text{same with } \xi^*(x,t) \quad \text{for } S_n(x).$$

Lemma 1. *Suppose* $0<\varepsilon_n<1$, $B_n\in Z^+$, $c>1/2$. *There are positive constants* C_1, C_2 *(independent of n) such that, if* $B_n^{-1}<C_1(c-\frac{1}{2})$ *and* $n^{-1}B_n\log(B_n/\varepsilon_n)<C_1(c-\frac{1}{2})^2/c$, *then*

$$P\{\sup_{x\in I} n^{1/2}|S_n(x)-S_{B_n,n}(x)|\geqq [cB_n^{-1}\log(B_n/\varepsilon_n)]^{1/2}\}\leqq C_2\varepsilon_n. \tag{1.13}$$

In particular, (1.13) *holds if* $B_n\to\infty$ *and* $n^{-1}B_n\log(B_n/\varepsilon_n)\to 0$. *If* $n^{1/2}[S_n(x)-x]$ *is replaced by* $\xi^*(x,1)$ *in* (1.12) *and* (1.13), *then* (1.13) *remains valid.*

Proof. Given that $nS_n(B^{-1})=m>0$, the rv

$$\bar{D}_n=\sup_{0\leqq x\leqq B^{-1}} nm^{-1}|S_n(x)-S_{B,n}(x)| \tag{1.14}$$

is clearly distributed as $\sup_{x\in I}|S_m(x)-x|$. Hence [6], for all positive m and d, and some constant c',

$$P\{\bar{D}_n>m^{-1/2}d\,|\,nS_n(B^{-1})=m\}\leqq c'e^{-2d^2}. \tag{1.15}$$

On the other hand, the standard Markov exponential bound for the binomial case yields, for $\delta>0$,

$$\begin{aligned} P\{S_n(B^{-1})\geqq(1+\delta)B^{-1}\} &\leqq e^{-nB^{-1}\delta(1+\delta)}\,E\,e^{\delta nS_n(B^{-1})} \\ &\leqq e^{-nB^{-1}\delta^2[1+O(\delta+B^{-1})]/2} \end{aligned} \tag{1.16}$$

as $\delta, B^{-1}\to 0$. Suppose $0<\varepsilon<1\leqq B$. Set $d=[cnm^{-1}B^{-1}\log(B/\varepsilon)]^{1/2}$ in (1.15) and $\delta=2[cn^{-1}B\log(B/\varepsilon)]^{1/2}$ in (1.16). We then obtain

$$\begin{aligned} &P\{n^{1/2}\sup_{0\leqq x\leqq B^{-1}}|S_n(x)-S_{B,n}(x)|\geqq[cB^{-1}\log(B\varepsilon)]^{1/2}\} \\ &\quad\leqq c'e^{-[2c/(1+\delta)]\log(B/\varepsilon)}+e^{-2c[1+O(\delta+B^{-1})]\log(B/\varepsilon)}. \end{aligned} \tag{1.17}$$

Substituting B_n, ε_n for B, ε, we note that our hypothesis implies that $\delta<2C_1^{1/2}(c-\frac{1}{2})$. Also, the term $O(\delta+B^{-1})$ in (1.16) and (1.17) depends only on δ and B^{-1}, not on n. Thus, for $2c>1$ and C_1 fixed at a suitably small value, the right side of (1.17) is bounded by $(1+c')(\varepsilon_n/B_n)$. Exactly the same development holds if $\{0\leqq x\leqq B_n^{-1}\}$ is replaced in (1.17) by $\{B_n^{-1}i\leqq x\leqq B_n^{-1}(i+1)\}$, $1\leqq i<B_n^{-1}$. Since there are B_n such regions making up I, (1.13) is proved. The remainder of the Lemma is now obvious.

Remarks on Lemma 1. (1) If B_n is constant, it is clear how to modify (1.16) to obtain (1.13) once more, with a change in c there. Also, the uniform spacing of the B_n intervals can be modified. (2) For the purpose of obtaining numerical results for approximating S_n by $S_{B,n}$, the constants C_i can be made explicit with slight additional effort. (3) Finer estimates can be obtained upon replacing c by $2^{-1}+a_n$, with slowly varying $a_n \downarrow 0$; this and corresponding lower and upper class characterizations for the almost sure (in n) analogue of (1.13) are of no concern to us here, although the gross first order result for the latter is essentially present in the proof of Theorem 1. (4) Finally, it is simple to see that the methods of [8] can be used to obtain bounds of the same type for $\sup_{k \leq n,\, x \in I} n^{-1/2} k |S_k(x) - S_{B_n,k}(x)|$ (or, for probability $\varepsilon_n = n^{-r}, r>0$, as we shall require, the result can even be obtained by summing probability bounds obtained for fixed k, $1 \leq k \leq n$). Moreover, if we replace the domain $k \leq n$ here by $k \leq \bar{c}\, n(\log n)^{\beta}$ for fixed positive $\bar{c}$ and β, it is easy to see that $[B_n^{-1} n \log n]^{1/2}$ need only be multiplied by $\bar{c}^{1/2}(\log n)^{\beta/2}$ to yield the probability bound n^{-1}. Thus, in the form we shall require in Section 3, we state

Lemma 1′. *There are positive values ε', C_1', and c such that, for $\beta \geq 0$,*

$$C_1' \leq B_n \leq \varepsilon' n/\log n \Rightarrow P\{\sup_{0 \leq t \leq \bar{c} n (\log n)^{\beta},\, x \in I} t |S_t(x) - S_{B_n,t}(x)| \geq c' [\bar{c} B_n^{-1} n (\log n)^{\beta+1}]^{1/2}\} \leq n^{-1}. \tag{1.18}$$

Again, the same results hold if $t[S_t(x) - x]$ is replaced by $\xi^*(x, t)$ in (1.12) *and* (1.18).

We have used here our repeated notational convention of piecewise linear interpolation in t to obtain such functions as $t S_{B_n,t}$ (from $n S_{B_n,n}$). As we have remarked, (3.23) implies the equivalence of results we use, for continuous t and discrete n.

2. Embeddings

2.1. Some Possibilities

We will depart from the sound principle of not wasting the reader's time with the author's false tries and negative results, for a simple reason: as mentioned in Section 1, the methods finally used herein do not, in their present form, yield the desired (conjectured) results, and it may save time for future workers to see what some of the alternative possibilities are and are not.

The process $\{S_1(p) - p,\ p \in I\}$ is not a martingale, but it is easily checked that

$$L_1(p) = (1-p)^{-1}[S_1(p) - p], \qquad 0 \leq p < 1, \tag{2.1}$$

is (by our choice above (1.3)) a left-continuous martingale of a very simple form, and hence is well known to have a Skorohod embedding in a standard Brownian motion ξ. In fact, there are several ways of generating S_1 from ξ, as described in the next two paragraphs. Since trouble with joint distributions (when marginals behave satisfactorily) will be seen to be a difficulty throughout this work, as in fact has already been mentioned in Section 1 in connection with [3–4], and since the construction of the next paragraph is not the standard Skorohod martingale embedding, let us illustrate by two brief examples the pitfalls one must avoid if one does not mechanically follow one of the standard constructions. For simplicity, let p_1 and p_2 be fixed values, $0<p_1<p_2<1$, and consider the problem of generating rv's with the same distribution as $(S_1(p_1), S_1(p_2))$, from Brownian motion. If we stop a standard Brownian motion $\bar{\xi}$ at the first time T_1' that $\bar{\xi}(T_1')=1-p_1$ or $-p_1$, and then at the first subsequent time T_2' that $\bar{\xi}(T_1'+T_2')-\bar{\xi}(T_1')=1-p_2$ or $-p_2$, we see that $(\bar{\xi}(T_1')+p_1, \bar{\xi}(T_1'+T_2')-\bar{\xi}(T_1')+p_2)$ has the right marginal distributions but the wrong joint distribution. Similarly, if T_2'' is the first time that $\bar{\xi}(T_2'')=p_2^{-1}$ or $-(1-p_2)^{-1}$, then $\bar{\xi}(T_1')+p_1$ and $p_2(1-p_2)\,\bar{\xi}(T_2'')+p_2$ have the right marginals but the wrong joint law since, with positive probability, the first is 1 and the second is 0.

It is not hard to guess natural schemes that work. Motivated by the above, let $T_p^*=\inf\{t\colon \bar{\xi}(t)=1-p \text{ or } -p\}$. Then $\{\bar{\xi}(T_p^*)+p,\ p=p_1 \text{ or } p_2\}$ has the correct joint law of $(S_1(p_1), S_1(p_2))$ precisely because $\bar{\xi}(T_{p_1}^*)=1-p_1$ implies $\bar{\xi}(T_{p_2}^*)=1-p_2$ if $p_1<p_2$. Similarly, we see that $\{\bar{\xi}(T_p^*)+p, p\in I\}$ has exactly the law of $\{S_1(p), p\in I\}$ (or $\{(1-p)^{-1}\,\bar{\xi}(T_p^*), 0\leqq p\leqq 1\}$ has the same law as (2.1)); note that $T_0^*=T_1^*=0$ wp 1.

The details of the remainder of this section and the next two sections can be carried out using $\{T_p^*\}$. However, it is technically somewhat simpler to work in terms of a different infinite set $\{T_z, z\in R_1^+\}$ of stopping times, because of the fact that $T_{p_1}^*$ can be either greater or less than $T_{p_2}^*$ (which is why $\{T_p^*, p\in I\}$ are not the stopping times usually encountered in Skorohod embeddings), while we shall have T_z nondecreasing in z wp 1. We define

$$T_z=\inf\{t\colon \bar{\xi}(t)=1 \text{ or } -z\}. \tag{2.2}$$

Then it is easily checked that $\{(1-p)\,\bar{\xi}(T_{p/(1-p)})+p, p\in I\}$ ($=1$ if $p=1$) has the same law as $\{S_1(p), p\in I\}$. In fact, this is just the simplest Skorohod embedding of the martingale L_1. Working with T_z rather than T_p^*, we will find it convenient to replace $I-\{1\}=\{p\}$ by $R_1^+=\{z\}$, and we shall repeatedly use z, p for variables related, as above, by $p=z/(z+1)$ and $z=p/(1-p)$. Thus, $\{\bar{\xi}(T_z), z\in R_1^+\}$ has the same law as the martingale $L_1(z/(z+1))$, which we hereafter denote by

$$Q_1(z)=(z+1)\,S_1(z/(z+1))-z, \qquad z\in R_1^+. \tag{2.3}$$

While the almost sure limit, 1, of $Q(z)$ or $\bar{\xi}(T_z)$ as $z\to+\infty$ cannot be adjoined while maintaining a martingale, this causes absolutely no trouble at $p=1$, since what matters there is that $\lim_{z\to\infty}(z+1)^{-1}Q_1(z)=0=S_1(1)-1$ wp. 1. We shall hereafter usually work in terms of Q_1 rather than S_1 or L_1, and correspondingly in terms of the ξ of (1.2) rather than the ξ^* of (1.1). We shall write Q_n for (2.3) with S_1 replaced by S_n, and $t\,Q_t$ is obtained from $n\,Q_n$ by piecewise linear interpolation.

A natural way to continue the development is now this: Let $\xi_1', \xi_2', \ldots$ be independent standard Brownian motions. Let $T_{i,z}'$ be the first time that $\xi_i'(T_{i,z}')=1$ or $-z$, so that $\{\xi_i'(T_{i,z}'), z\in R_1^+\}$ are iid (in i) processes, each distributed as Q_1. Then $\left\{\sum_1^n \xi_i'(T_{i,z}'), z\in R_1^+\right\}$ has exactly the distribution of $\{n\,Q_n(z), z\in R_1^+\}$, from which we obtain at once an exact representation of $n\,S_n$.

But where is $\xi(z,t)$? Nothing so obvious as $\xi(z,n)=\sum_1^n \xi_i(z)$ will be useful, since by the central limit theorem that could not even yield $o(1)$ in (1.5)–(1.6). A natural attempt (at least to me) is to let $U'_{m,z}=\sum_1^m T'_{i,z}$ (with $U'_{0,z}=0$) and define

$$\xi'(z,t)=\sum_1^m \xi'_i(T'_{i,z})+\xi'_{m+1}(t-U'_{m,z}) \quad \text{whenever } U'_{m,z}\leqq t\leqq U'_{m+1,z}. \tag{2.4}$$

For then, for fixed $z>0$, by the strong Markov property of the $T'_{i,z}$, we have that $\{\xi'(z,t), t\geqq 0\}$ is a standard Brownian motion and thus $\{\xi'(z,zt), t\geqq 0\}$ has the desired law of $\{\xi(z,t), t\geqq 0\}$ for each z. Unfortunately, the strong Markov property does not yield the right joint distribution: for $0<z_1<z_2$ and $\varepsilon>0$, choose t_1 so small that $P\{T'_{1,z_1}>t_1\}>1-\varepsilon$, and note that $T'_{1,z_1}>t_1$ implies $\xi'(z_1,t_1)=\xi'(z_2,t_1)$. In fact, we even lose joint normality. (A reason for mentioning this "failure" ξ' is that, as discussed in Section 6, it offers some promise for eventual success.)

However, realizing that the joint behavior of the processes $\xi(z_i,\cdot)$ is reflected in the independence, for fixed $z_1<z_2\leqq z'_1<z'_2$, of the processes $\eta(z_1,z_2;\cdot)$ and $\eta(z'_1,z'_2;\cdot)$ defined by

$$\eta(z_1,z_2;t)=\xi(z_2,t)-\xi(z_1,t), \tag{2.5}$$

we are motivated to define rv's distributed like the $\xi'_i(T'_{i,z_j})$ above in terms of the $\eta(z_{j-1},z_j;\cdot)$ rather than in terms of the ξ'_i. This can be done in a more general context, to which we now temporarily turn.

2.2. Skorohod Vector Embeddings

A number of workers have been concerned with the possibility of a Skorohod representation of a vector rv. One obvious piece of wishful thinking must be dispelled at the outset: we cannot hope to succeed with a single stopping time. This is apparent if one tries to embed, in 2-dimensional Brownian motion, a 2-vector taking on the four equiprobable values $(\pm 1, \pm 1)$. On the other hand, if one looks at each of the two coordinates of the Brownian motion at its obvious stopping time to yield the values ± 1, we have a representation.

Let $A_m=(A_{m1}, A_{m2}, \ldots, A_{mh})$ be iid (in m) h-vectors such that $A_{11}, A_{12}, \ldots, A_{1h}$ is a 0-expectation martingale, $E\{A_{1j}|F_{j-1}\}=A_{1,j-1}$ wp 1 for $1\leqq j\leqq h$ where $A_{10}=0$ and F_j is the σ-field generated by $\{A_{11},\ldots,A_{1j}\}$. Without further mention we can and will use conditional probability measures

$$P\{A_{1,j+1}\leqq u|A_{11}=a_1,\ldots,A_{1j}=a_j\}.$$

Assume also, for the moment, that A_1 is bounded wp 1. Let $\{\eta(i-1,i;t), t\geqq 0\}$, $1\leqq i\leqq h$, be h independent standard Brownian motions. (This notation is motivated by the sample df results and (2.5); think of $\eta(i-1,i;t)=\xi(i,t)-\xi(i-1,t)$.)

Now represent $A_{11},\ldots,A_{1h}$ by Skorohod embedding, *except* that instead of using a single Brownian motion for all $A_{1,j+1}-A_{1j}$ we use $\eta(0,1;\cdot)$ for A_{11}, $\eta(1,2;\cdot)$ for $A_{12}-A_{11}$, and so on up to $\eta(h-1,h;\cdot)$ for $A_{1h}-A_{1,h-1}$. Formally, $T^{(1)}_{i-1,i}$ is a stopping time defined inductively on i for $1\leqq i\leqq h$ (and possibly using additional randomization as in Skorohod's original method) such that, if $B^{(1)}_i=\{\eta(j-1,j;T^{(1)}_{j-1,j})=a_j-a_{j-1}, 1\leqq j<i\}$ $(a_0=0)$, then

$$P\{\eta(i-1,i;T^{(1)}_{i-1,i})\leqq u|B^{(1)}_i\}=P\{A_{1i}-A_{1,i-1}\leqq u|A_{1j}=a_j, j<i\}. \tag{2.6}$$

This yields $\{\eta(i-1, i; T^{(1)}_{i-1,i}), 1 \leqq i \leqq h\}$ distributed as $\{A_{1i} - A_{1,i-1}, 1 \leqq i \leqq h\}$.

Then, given $T^{(1)}_{i-1,i} = t_i$ (say), we obtain an h-vector independent of that of the previous sentence, but with the same distribution, by writing $\eta(j-1, j; T^{(2)}_{j-1,j} + t_j) - \eta(j-1, j; t_j)$ in place of $\eta(j-1, j; T^{(1)}_{j-1,j})$ on the left side of (2.6) (including $B^{(1)}_i$). That is, we observe the $\eta(j-1, j; \cdot)$ process from time $T^{(1)}_{j-1,j}$ to $T^{(1)}_{j-1,j} + T^{(2)}_{j-1,j}$ to get a rv which represents $A_{2j} - A_{2,j-1}$, successively for $j = 1, 2, \ldots, h$.

We continue in this fashion. Thus, define $U^{(0)}_{i,i+1} = T^{(0)}_{i,i+1} = 0$ and then $U^{(m)}_{i,i+1} = \sum_{q=1}^{m} T^{(q)}_{i,i+1}$ inductively for $m \geqq 0$ and $0 \leqq i < h$ by letting the vectors $(T^{(q)}_{0,1}, \ldots, T^{(q)}_{h-1,h})$ be iid in q and such that, with

$$B^{(q)}_i = \{\eta(j-1, j; U^{(q)}_{j-1,j}) - \eta(j-1, j; U^{(q-1)}_{j-1,j}) = a_j - a_{j-1}, 1 \leqq j < i\} \qquad (a_0 = 0), \quad (2.7)$$

$T^{(q)}_{i,i+1}$ is any stopping time satisfying

$$\begin{aligned} &P\{\eta(i-1, i; U^{(q)}_{i-1,i}) - \eta(i-1, i; U^{(q-1)}_{i-1,i}) \leqq u \mid B^{(q)}_i\} \\ &\quad = P\{A_{q,i} - A_{q,i-1} \leqq u \mid A_{qj} = a_j, j < i\}. \end{aligned} \qquad (2.8)$$

Again, we can assume our probability structure allows external randomization if needed.

We thus obtain

Lemma 2. *Skorohod vector embedding:*

$$\left\{\left\{\sum_{j=1}^{i} \eta(j-1, j; U^{(n)}_{j-1,j}), 1 \leqq i \leqq h\right\}, \; n \in Z^+\right\} \qquad (2.9)$$

has the same joint law as $\left\{\sum_{m=1}^{n} A_m, n \in Z^+\right\}$.

Remarks on the Embedding (2.9). 1. It is now obvious how to study, for each j, the difference between $\eta(j-1, j; U^{(n)}_{j-1,j})$ and $\eta(j-1, j; E U^{(n)}_{j-1,j})$, and thus obtain precise almost sure results on this difference as in [10], and also the k-dimensional analogues of Strassen's results [21, 22]. (See Section 6A for further comments.) Also, the analogues of the results of (1.7) and (1.9) (described there for the special case $A_{1i} = L_1(z_i)$) follow at once from using our developments with the techniques of [20] and [19] for more general summands. (See, however, Section 6C regarding inapplicability of some of these techniques.) Moreover, the representation (2.9) does not require boundedness of A_1, and the error term one obtains in these approximations will depend on (conditional) moments in a manner made clear in [2, 10, 19].

Thus, (1.7), (1.8), (1.9) *carry over, with appropriate modification of their right sides as described in* [2, 10, 19] *under various moment conditions, to general* $\{A_m\}$ *with* $\{A_{1j}\}$ *a martingale, upon replacing*

$$\{k(S_k(s) - s) \text{ and } \xi^*(s, t); \; 1 \leqq k \leqq n, t \in I, s \in I'\} \qquad (2.10)$$

by

$$\left\{\sum_{m=1}^{k} \sum_{j=1}^{i} A_{mj} \text{ and } \sum_{j=1}^{i} \eta(j-1, j; t E(A_{1j} - A_{1,j-1})^2), \; 1 \leqq k \leqq n, t \in I, 1 \leqq i \leqq h\right\}. \quad (2.11)$$

(As remarked earlier and as is explained in Section 6, the exponent 1/4 in (1.7) and (1.9) cannot be replaced by 1/2 with the present scheme, even if $h=2$.)

2. The iid structure in m of $\{A_m\}$ is inessential. In fact, there is no difficulty in embedding processes $\left\{\sum_{m=1}^{n} A_{mj}; j, m\in Z^+\right\}$ which are martingales in n as well as in j. Moreover, there are embeddings for which, as with the T_p^* of the previous section, the stopping times for $A_{1j}-A_{1,j-1}$ need not be ordered in j.

3. Of great interest is the possibility of replacing j (and, possibly, n) by a continuous time parameter in (2.9). Some circumstances where this is possible will be treated in [13]. In Section 2.3, below, we will see what goes wrong with this attempt in the sample df example.

4. Another direction of extension is to the case where the index set of j in A_{mj} is not 1-dimensional, e.g., to the sample df of chance k-vectors. The arithmetic of [8] shows the kind of modifications that are needed in Lemmas 1′ and 4′. One obtains somewhat worse orders than in (1.5), (1.6), (1.10), (1.11), in terms of a process $\xi^{*(k)}(s, t)$ with k-dimensional s. Results can be obtained for general $\{A_{1j}\}$ with j in a lattice of R^k, a setting considered more extensively by J. Zinn, a student of J. Kuelbs.

2.3. *Back to the Sample* df

Let I' be any subset of I, as in Corollary 2, (1.7). In view of the nature of $S_n(1)$, we can limit our treatment to the case $1\notin I'$. Let $R'=\{z\colon z=p/(1-p), p\in I'\}$ be the corresponding set of non-negative values under the correspondence described above (2.3). Adjoining $z_0=0$ (unless $0\in R'$ already), we write $R'=\{z_0, z_1, \ldots, z_k\}$ with $z_i<z_{i+1}$. Now, with ξ the process of (1.2), replace the $\eta(i-1, i; t)$ of Section 2.2 by the $\eta(z_{i-1}, z_i; (z_i-z_{i-1})^{-1}t)$ of (2.5), which have the same distribution. (The convenience of this normalization of the η of (2.5) will be seen below to be that the analogue here of $T_{i-1,i}^{(1)}$ of (2.6) will have expectation z_i-z_{i-1}.) The rv's A_{1i} are the $Q_1(z_i)$ of (2.3). Thus, for the embedding of Section 2.2, (2.7)–(2.8) with $\eta(j-1, j; U_{j-1,j}^{(q)})$ there replaced by $\eta(z_{j-1}, z_j; (z_j-z_{j-1})^{-1}U_{z_{j-1},z_j}^{(q)})$ here, can be realized by treating $\eta(z_{j-1}, z_j; (z_j-z_{j-1})^{-1}T_{z_{j-1},z_j}^{(1)})$ exactly as $\xi(T_{z_j})-\xi(T_{z_{j-1}})$ of (2.2) or $\xi_1'(T_{1,z_j}')-\xi_1'(T_{1,z_{j-1}}')$ below (2.3):

$$T_{z_{j-1},z_j}^{(1)}=\inf\left\{t\colon t\geqq 0, \sum_{i=1}^{j-1}\eta(z_{i-1}, z_i; (z_i-z_{i-1})^{-1}T_{z_{i-1},z_i}^{(1)}) + \eta(z_{j-1}, z_j; (z_j-z_{j-1})^{-1}t)=1 \text{ or } -z_j\right\} \tag{2.12}$$

and, similarly,

$$\begin{aligned} T_{z_{j-1},z_j}^{(q)}=\inf\Big\{t\colon t\geqq 0, &\sum_{i=1}^{j-1}\left[\eta(z_{i-1}, z_i; (z_i-z_{i-1})^{-1}U_{z_{i-1},z_i}^{(q)})\right.\\ &\left.-\eta(z_{i-1}, z_i; (z_i-z_{i-1})^{-1}U_{z_{i-1},z_i}^{(q-1)})\right]\\ &+\left[\eta(z_{j-1}, z_j; (z_j-z_{j-1})^{-1}(U_{z_{j-1},z_j}^{(q)}+t))\right.\\ &\left.-\eta(z_{j-1}, z_j; (z_j-z_{j-1})^{-1}U_{z_{j-1},z_j}^{(q-1)})\right]=1 \text{ or } -z_j\Big\}. \end{aligned} \tag{2.13}$$

Since $\{\eta(z_{j-1}, z_j; (z_j-z_{j-1})^{-1}t), t\geqq 0\}$ has variance 1 per unit of time t, we see that $T_{z_{j-1},z_j}^{(q)}$ has the same expectation as the $T_{z_j}-T_{z_{j-1}}$ of (2.2) or $T_{1,z_j}'-T_{1,z_{j-1}}'$

below (2.3), namely

$$E\,T^{(q)}_{z_{j-1},z_j} = z_j - z_{j-1}; \tag{2.14}$$

thus, our normalization assures that the *expectation* of the third argument of $\eta\big(z_{i-1}, z_i; (z_i - z_{i-1})^{-1}\, U^{(q)}_{z_{i-1},z_i}\big)$ is q.

We have eliminated the difficulty of not finding the ξ of (1.2) in the ξ' of (2.4) obtained from the $T'_{i,z}$, by now embedding $\{Q_n(z), z \in R', n \in Z^+\}$ in ξ. As we shall indicate in the next paragraph, all of $\{Q_n(z), z \leqq z_k\}$ has in fact been embedded in ξ, but $Q_n(z)$ for all z is not as evident as one might like in the above embedding using R'. (The R' above can be modified to include, for example, an infinite sequence approaching 0 or ∞, but this will not help achieve what we want here.) For simplicity, consider $\{Q_1(z), z \leqq 1\}$. We would like to replace the above embedding of $\{Q_1(z), z \in R'\}$ by $\{Q_1(z), z \leqq 1\}$, by piecing together (roughly) stopped parts of a continuum of differential processes $\xi(z, z+dz; t)$. One might try to get at this by considering a sequence of nested finite sets whose union is dense, e.g., $R'_L = \{z: z = i/2^L, 0 \leqq i \leqq 2^L\}$, hoping somehow to take a limit with L of the previous embeddings to obtain $\{Q_1(z), z \leqq 1\}$. We do not see how to make this work. Roughly, although $2^L T^{(1)}_{z_j, z_{j+1}}$ has expectation 1 for all L and $z_{j+1} - z_j = 2^{-L}$, it follows from (2.19) that it has variance 2^L and must itself be very large for some (chance) j. The latter is certainly the case for an interval (z_j, z_{j+1}) containing the z for which Q_1 jumps, and it will also be true for (chance) values $\bar{z}$ for which, in terms of (2.2), ξ has a local minimum at $t = T_{\bar{z}}$ so that $\inf_{z > \bar{z}} (T_z - T_{\bar{z}}) > 0$.

This failure reflects our feeling that the present method of embedding, while it has proved a useful tool, is not an ultimate one. Actually, all of $\{Q_n(z), 0 \leqq z \leqq z_k\}$ (and, with a slight modification, all of $\{Q_n(z), z \in R_1^+\}$) *does* sit in ξ in the above embedding with R', but not in a form which yields the desired Skorohod-type estimates without further calculations. For example, consider $R'\{0, 1, 2\}$. It is convenient to think of ξ as being defined on the domain of planar Borel sets, with, as usual, $\xi(A)$ having variance A and being independent over disjoint A's; thus, we simply write $\xi([0, z] \times [0, t])$ for $\xi(z, t)$. In Fig. 1 is shown a possible realization of the set

$$A = \{[0, 1] \times [0, U^{(2)}_{0,1}]\} \cup \{[1, 2] \times [0, U^{(2)}_{1,2}]\}$$

for which $\xi(A) = 2Q_2(2)$. The shaded subset B of A, for which $\xi(B) = 2Q_2(1.6)$, is obtained by letting

$$\begin{aligned} \tau^{(i)} = \inf \Big\{ t:\ & t \geqq 0, [\eta(0, 1; U^{(i)}_{0,1}) - \eta(0, 1; U^{(i-1)}_{0,1})] \\ & + \Big[\eta\Big(1, 2; \sum_{j=1}^{i-1} \tau^{(j)} + t\Big) - \eta\Big(1, 2; \sum_{j=1}^{i-1} \tau^{(j)}\Big)\Big] = 1 \text{ or } -1.6 \Big\}, \end{aligned} \tag{2.15}$$

and $2Q_2(z)$ is obtained similarly for other values z.

In similar fashion, by using an unbounded sequence $\{z_i\}$ one obtains an explicit embedding of $\{Q_n(z), z \in R^+\}$; however, the large z_j's play no role in the proofs and will simply be omitted. Thus, in Section 3 one can define z_j's for $j \geqq B_n$, to exhibit all of Q_n, but it is unnecessary to consider them in proving Theorem 2.

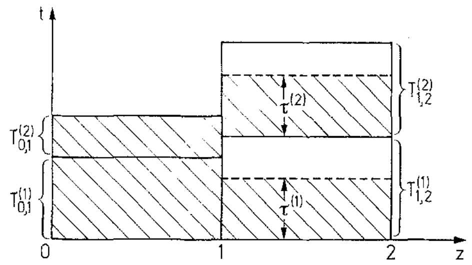

Fig. 1. Examples of $2Q_2(2)$ and $2Q_2(1.6)$

A similar remark will apply in Section 4 (Theorem 1), where for technical simplicity the values $z > B_n$ are treated slightly differently.

In terms of the discussion, above, of the difficulties accompanying a process of subdivision, the reader might find it instructive to see what happens to this already complicated picture when one subdivides the simple R' of Fig. 1!

What we shall do, then, is to use Lemma 1′ to approximate S_n by $S_{B_n,n}$ for an appropriate B_n, and then use the embedding of Sections 2.2–2.3 for $S_{B_n,n}$ in ξ, to prove Theorem 2.

We conclude this section by computing some elementary properties of the $T^{(i)}$ and $U^{(i)}$. It is well known that, for $\alpha < \pi^2/2(1+z)^2$ and $z > 0$,

$$E\exp\{\alpha T^{(1)}_{0,z}\} = \frac{\cos[(\alpha/2)^{1/2}(1-z)]}{\cos[(\alpha/2)^{1/2}(1+z)]}. \tag{2.16}$$

Hence, as $\alpha \to 0$,

$$E\exp\{\alpha T^{(1)}_{0,z}\} = 1 + \alpha z + \alpha^2(z + 3z^2 + z^3)/6 + O(\alpha^3(z+1)^4 z). \tag{2.17}$$

From the above or Wald's equation, as already noted in (2.14), $E T^{(1)}_{z_1,z_2} = z_2 - z_1$. It will simplify notation and yield the same result if we work in terms of the $\bar{\xi}$ of (2.2) rather than the η of (2.12)–(2.13). We note now that $P\{T^{(1)}_{z_1,z_2} = 0 \mid \bar{\xi}(T^{(1)}_{0,z_1}) = 1\} = 1$; and from looking at stopping boundaries we see that, given that $\bar{\xi}(T^{(1)}_{0,z_1}) = -z_1$, the conditional distribution of

$$\{(1+z_1)^{-1}[\bar{\xi}([1+z_1]^2 t + T^{(1)}_{0,z_1}) + z_1],\ 0 \leqq [1+z_1]^2 t \leqq T^{(1)}_{z_1,z_2}\}$$

is that of the unconditional distribution of $\{\bar{\xi}(t), 0 \leqq t \leqq T^{(1)}_{0,(1+z_1)^{-1}(z_2-z_1)}\}$. Since $P\{T^{(1)}_{0,z_1} = 1\} = z_1/(1+z_1)$, we obtain from (2.17),

$$\begin{aligned} E\exp\{\alpha T^{(1)}_{z_1,z_2}\} &= (1+z_1)^{-1}[z_1 + E\exp\{\alpha(1+z_1)^2\, T^{(1)}_{0,(1+z_1)^{-1}(z_2-z_1)}\}] \\ &= 1 + \alpha(z_2 - z_1) + \alpha^2\{(z_2-z_1)(1+z_1)^2 + 3(z_2-z_1)^2(1+z_1) \\ &\quad + (z_2-z_1)^3\}/6 + O(\alpha^3(1+z_2)^4(z_2-z_1)). \end{aligned} \tag{2.18}$$

Consequently, routine calculation shows that

$$\begin{aligned} &\log E\exp\{\alpha[T^{(1)}_{z_1,z_2} - (z_2-z_1)]\} \\ &\sim \alpha^2\{(z_2-z_1)(1+z_1)^2 + 3(z_2-z_1)^2 z_1 + (z_2-z_1)^3\}/6 \\ &= \alpha^2 h(z_1,z_2) \text{ (say)}, \end{aligned} \tag{2.19}$$

provided that this last expression is $o(1)$ and that the ratio of $\alpha^3(1+z_2)^4(z_2-z_1)$ to this expression is $o(1)$.

Elementary estimates on the h of (2.19) yield

$$\tfrac{1}{12}(z_2-z_1)(1+z_2)^2<h(z_1,z_2)<\tfrac{5}{24}(z_2-z_1)(1+z_2)^2. \tag{2.20}$$

We now compute ordinary Markov exponential bounds (as in (1.16) for Bernoulli rv's). Putting $\alpha=q_N/2hN$ in (2.19) ($q_N>0$) yields

$$\begin{aligned}P\{U^{(N)}_{z_1,z_2}-N(z_2-z_1)>q_N\}&\leqq E\exp\{\alpha[U^{(N)}_{z_1,z_2}-N(z_2-z_1)-q_N]\}\\&=\exp\{-(q_N^2/4hN)(1+o(1))\}.\end{aligned} \tag{2.21}$$

The negative deviations of $U^{(N)}$ are treated similarly. Moreover, considering the conditions just below (2.19), from (2.20) we have $\alpha^2 h$ of exactly order

$$N^{-2}q_N^2(z_2-z_1)^{-1}(1+z_2)^{-2}$$

and the ratio of $\alpha^3(1+z_2)^4(z_2-z_1)$ to $\alpha^2 h$ of exactly order $N^{-1}q_N(z_2-z_1)^{-1}$. If the latter is $o(1)$, so is the former. Thus, we have

$$\begin{aligned}&q_N/N(z_2-z_1)=o(1)\\&\quad\Rightarrow P\{|U^{(N)}_{z_1,z_2}-N(z_2-z_1)|>q_N\}\leqq\exp\{-(q_N^2/4hN)(1+o(1))\}.\end{aligned} \tag{2.22}$$

Finally, at slight expense in sharpness we will put the estimate (2.22) in a form useful in the proof of Theorem 2 of the next section. We will use (2.22) with z_1, z_2 replaced by z_{j-1}, z_j where $z_j>z_{j-1}$ and $1\leqq j<B_n$ ($z_0=0$). From (2.22) and the right side of (2.20) we obtain

Lemma 3. *There is a positive constant C_3 such that*

$$\begin{aligned}q_N/N\leqq C_3(z_j-z_{j-1})&\Rightarrow P\{|U^{(N)}_{z_{j-1},z_j}-N(z_j-z_{j-1})|>q_n\}\\&\leqq\exp\{-q_N^2/N(z_j-z_{j-1})(1+z_j)^2\}.\end{aligned} \tag{2.23}$$

In fact, in applying this in Section 3 we shall use the particular z_j of (3.2) ($1\leqq j<B_N$) with

$$q_N=2(z_j-z_{j-1})[NB_n\log(nB_n)]^{1/2} \tag{2.24}$$

for various large integers N. From (3.2), $B_n(z_j-z_{j-1})/(1+z_j)^2=(B_n-j)/(B_n-j+1)\geqq 1/2$. Hence, (2.23) becomes

$$\begin{aligned}&(3.2)\text{ with } N^{-1}B_n\log(nB_n)\leqq C_3^2/4\\&\quad\Rightarrow P\{|U^{(N)}_{z_{j-1},z_j}-N(z_j-z_{j-1})|>q_N\text{ of (2.24)}\}\leqq(nB_n)^{-2}.\end{aligned} \tag{2.25}$$

3. Statement and Proof of Theorem 2

It is convenient to divide the considerations into two parts – bounding in probability the difference between ξ and the embedded process, and then bounding the limiting density of the functional.

To emphasize the essentials of the proof, we postpone the statement of the required Lemma 4′ and first prove Lemma 4, which is of no interest in itself in view of the existence of embeddings for fixed n mentioned in Section 1, but which contains the main ideas of Lemma 4′ with simpler arithmetic. We use the notation of Lemma 1 and of Section 2.3, and write $\varepsilon=\min(C_3^2, 1/2)$ where C_3 is as described in (2.25).

Lemma 4. *With the embedding of nQ_n in ξ described in Section 2.3, with the p_j equally spaced, and with ε as defined just above, there is a positive constant C_4 such that, for n sufficiently large,*

$$\begin{aligned} 2 \leqq B_n \leqq \varepsilon n/8 \log n \Rightarrow P\{ &\max_{1\leqq j<B_n} (z_j+1)^{-1} |nQ_n(z_j) - \xi(z_j, n)| \\ &\leqq C_4[(nB_n)^{1/4}(\log n)^{3/4} + B_n \log n]\} \geqq 1-n^{-1}. \end{aligned} \tag{3.1}$$

Proof. Corresponding to equally spaced $p_j = B_n^{-1} j$ $(1 \leqq j < B_n)$, we let $z_j = p_j/(1-p_j) = j/(B_n - j)$ and thus

$$\begin{aligned} z_j - z_{j-1} &= B_n/(B_n - j)(B_n - j + 1), \\ z_j + 1 &= B_n/(B_n - j). \end{aligned} \tag{3.2}$$

Fixing n and B_n and dropping the subscript on the latter, for positive integral N we define

$$f(N) = 2[NB \log(nB)]^{1/2}. \tag{3.3}$$

Let ε be as defined above. We define inductively

$$\begin{aligned} &m_0 = n_0 = n, \\ &m_{i+1} = f(m_i), \\ &n_{i+1} = m_i - m_{i+1} \quad \left(\text{and hence } \sum_1^i n_r = n - m_i\right). \end{aligned} \tag{3.4}$$

This yields

$$\begin{aligned} m_i &= n[4n^{-1} B \log(nB)]^{1-2^{-i}} \quad &&\text{for } i \geqq 0, \\ m_i/m_{i-1} &= [4Bn^{-1} \log(nB)]^{2^{-i}} &&\text{for } i \geqq 1. \end{aligned} \tag{3.5}$$

Let

$$K = \begin{cases} 0 & \text{if } m_1/m_0 > \varepsilon, \\ \max\{i: m_i/m_{i-1} \leqq \varepsilon\} & \text{otherwise}. \end{cases} \tag{3.6}$$

Then

$$m_i < m_{i-1} \quad \text{and} \quad n_i \geqq (1-\varepsilon) m_{i-1} \quad \text{for } 1 \leqq i \leqq K \tag{3.7}$$

and, from (3.5) with $B > 1$, if $n \geqq 3$,

$$K \leqq [\log^+ (\log n / \log \varepsilon^{-1})]/\log 2. \tag{3.8}$$

In the remainder of this proof we will simplify notation by writing m_i and n_i, rather than integers close to them, as numbers of observations. It can be seen that proper reinterpretation of these symbols yields (3.1) without difficulty when n is large. Also for brevity, we shall write

$$\begin{aligned} &\bar\eta_j(t) = \eta(z_{j-1}, z_j; t), \\ &V_j^{(i)} = \begin{cases} 0 & \text{if } i = 0, \\ (z_j - z_{j-1})^{-1} U_{z_{j-1}, z_j}^{(n_1 + \cdots + n_i)} & \text{if } 1 \leqq i \leqq K, \\ (z_j - z_{j-1})^{-1} U_{z_{j-1}, z_j}^{(n)} & \text{if } i = K+1. \end{cases} \end{aligned} \tag{3.9}$$

(In particular, if $K=0$ we have $V_j^{(1)}=(z_j-z_{j-1})^{-1}U^{(n)}$; we drop obvious subscripts on $U^{(n)}$.)

To estimate $\bar\eta_j((z_j-z_{j-1})^{-1}U^{(n)})-\bar\eta((z_j-z_{j-1})^{-1}EU^{(n)})=\bar\eta_j(V_j^{(K+1)})-\bar\eta_j(n)$ (recall (2.14): $(z_j-z_{j-1})^{-1}EU^{(n)}=n$), we write the telescoping sum

$$\begin{aligned}\bar\eta_j(V_j^{(K+1)})-\bar\eta_j(n)=\bar\eta_j(V_j^{(K)}+[V_j^{(K+1)}-V_j^{(K)}])-\bar\eta_j(V_j^{(K)}+m_K)\\ +\sum_{i=1}^{K}\{\bar\eta_j(V_j^{(i)}+m_i)-\bar\eta_j(V_j^{(i)}+m_i-[V_j^{(i)}-V_j^{(i-1)}-n_i])\}.\end{aligned}\tag{3.10}$$

We first consider the first two terms on the right side of (3.10); in the next paragraph it will be seen why this difference requires a separate treatment. We see that, whether or not $K>0$,

$$4\varepsilon^{-1}B\log(nB)\leqq m_K<4\varepsilon^{-2}B\log(nB);\tag{3.11}$$

the right hand inequality follows from substituting $m_{K+1}/m_K>\varepsilon$ and the second line of (3.5) into the first line; the left hand inequality follows similarly from $m_K/m_{K-1}\leqq\varepsilon$ if $K>0$, and from the condition on B_n in (3.1) if $K=0$. Thus, conditional on $V_j^{(K)}$, the $V_j^{(K+1)}-V_j^{(K)}=(z_j-z_{j-1})^{-1}[U^{(n)}-U^{(n-m_K)}]$ of (3.10) has the same distribution as $(z_j-z_{j-1})^{-1}U^{(N)}$ of (2.25) with $N=m_K$ satisfying the condition of (2.25). We now recall (2.24) and let j vary, and consider the sum of probabilities (2.25) over $1\leqq j<B_n$. We obtain, wp 1,

$$P\{|(V_j^{(K+1)}-V_j^{(K)})-m_K|\leqq 2[m_KB\log(nB)]^{1/2},1\leqq j<B|\{V_j^{(K)}\}\}\geqq 1-n^{-2}B^{-1}.\tag{3.12}$$

Let A be the event of (3.12). Under the same conditioning as in (3.12), $\eta_j'(\tau)=\bar\eta_j((z_j-z_{j-1})^{-1}\tau+V_j^{(K)})$ is a standard Brownian motion for $\tau\geqq 0$. Write

$$\begin{aligned}\Gamma_j&=\{\tau:|\tau|\leqq 2(z_j-z_{j-1})[m_KB\log(nB)]^{1/2},\ \tau+(z_j-z_{j-1})m_K\geqq 0\},\\ q_j'&=3(z_j-z_{j-1})^{1/2}(m_KB)^{1/4}[\log(nB)]^{3/4}.\end{aligned}\tag{3.13}$$

By the standard inequality for the tail probabilities of the maximum of a Brownian motion, we have for $n\geqq 10$,

$$\begin{aligned}&P\{|\bar\eta_j(V_j^{(K)}+[V_j^{(K+1)}-V_j^{(K)}])-\bar\eta_j(V_j^{(K)}+m_K)|>q_j'|A,\{V_j^{(K)}\}\}\\ &\leqq P\{\sup_{\tau\in\Gamma_j}|\eta_j'(\tau+(z_j-z_{j-1})m_K)-\eta_j'((z_j-z_{j-1})m_K)|>q_j'|\{V_j^{(K)}\}\}<(nB)^{-2}.\end{aligned}\tag{3.14}$$

From (3.13), (3.11), and (3.2) we have

$$\begin{aligned}\sum_{r=1}^{j}q_r'&<5B^{1/2}\varepsilon^{1/2}\log(nB)\sum_{1}^{j}(z_r-z_{r-1})^{1/2}\\ &<5B\varepsilon^{-1/2}\log(nB)\log(2[z_j+1]).\end{aligned}\tag{3.15}$$

Thus, from (3.12), (3.13), (3.14), and (3.15), for n sufficiently large,

$$\begin{aligned}P\Big\{&\sum_{r=1}^{j}|\bar\eta(V_r^{(K+1)})-\bar\eta(V_r^{(K)}+m_K)|\\ &\leqq 5B\varepsilon^{-1/2}\log(nB)\log(2[z_j+1]),1\leqq j<B\Big\}\geqq 1-n^{-2}.\end{aligned}\tag{3.16}$$

We next consider the i-th term of the sum in (3.10). First note that, by (3.6)–(3.7), $n_i \geqq (1-\varepsilon)\, m_{i-1} \geqq \varepsilon^{-1}(1-\varepsilon)\, m_i \geqq m_K$; thus, since m_K satisfied the condition for N in (2.25), so does n_i. Hence, from (2.25) with $N=n_i$, we obtain, as in (3.12),

$$P\{|V_j^{(i)} - V_j^{(i-1)} - n_i| \leqq 2[n_i B \log(nB)]^{1/2},\ 1 \leqq j < B\} \geqq 1 - n^{-2} B^{-1}. \qquad (3.17)$$

Since $n_i < m_{i-1}$, the product of the two equations of (3.5) yields

$$2[n_i B \log nB]^{1/2} < m_{i-1}^{1/2} [4B \log nB]^{1/2} = m_{i-1}^{1/2} \left[m_i \frac{m_i}{m_{i-1}}\right]^{1/2} = m_i. \qquad (3.18)$$

This means that, if the event A' (say) of (3.17) occurs, then, conditional on $\{V_j^{(i)}\}$ and $\{V_j^{(i-1)}\}$, the arguments of the two $\bar{\eta}_j$ terms in the i-th difference of the sum of (3.10) *are both* $> V_j^{(i)}$. That difference is thus conditionally of the form $\bar{\eta}_j(t_1) - \bar{\eta}_j(t_2)$ where t_1 and t_2 are determined by $\{\bar{\eta}_j(t),\ t \leqq \min(t_1, t_2)\}$. (The absence of this property is what entailed separate treatment of the difference studied in the previous paragraph; that treatment used here would yield a bound inferior to (3.1).) The event of (3.17) consequently implies a corresponding change in the argument of the η'_j process (defined below (3.12)), of $(z_j - z_{j-1})[V_j^{(i)} - V_j^{(i-1)} - n_i]$ time units. Since the η'_j are standard, we obtain in terms of a standard η'_0,

$$\begin{aligned} P&\left\{\left|\sum_{r=1}^{j} \{\bar{\eta}_r(V_r^{(i)} + m_i) - \bar{\eta}_r(V_r^{(i-1)} + m_i - n_i)\}\right| > \bar{q}_j \,|\, A'_j, \{V_j^{(i)}\}, (V_j^{(i-1)}\}\right\} \\ &= P\{|\eta'_0(\tau_j)| > \bar{q}_j\} \leqq 2\tau_j^{1/2}\, \bar{q}_j^{-1}\, e^{-\bar{q}_j^2/2\tau_j}, \qquad (3.19) \\ \tau_j &= \sum_{r=1}^{j} (z_r - z_{r-1})\, |V_r^{(i)} - V_r^{(i-1)} - n_i| \leqq z_j m_i. \end{aligned}$$

Thus, putting $\bar{q}_j = 2[z_j m_i \log(nB)]^{1/2}$, (3.17) and (3.19) yield for n sufficiently large,

$$\begin{aligned} P&\left\{\left|\sum_{r=1}^{j} \{\bar{\eta}_r(V_r^{(i)} + m_i) - \bar{\eta}(V_r^{(i-1)} + m_i - n_i)\}\right| \right. \\ &\qquad \left. \leqq 2[z_j m_i \log(nB)]^{1/2},\ 1 \leqq j < B\right\} \geqq 1 - n^{-2}. \end{aligned} \qquad (3.20)$$

By (3.6), $\sum_1^K m_i^{1/2} < m_i^{1/2}/[1-\varepsilon^{1/2}] < 8[nB \log(nB)]^{1/4}$. Thus, finally, from (3.8), (3.10), (3.16), (3.20), and the condition on B in (3.1), we have for n sufficiently large and some positive constant C_4,

$$\begin{aligned} P&\left\{\left|\sum_{r=1}^{j} \{\bar{\eta}_r(V_r^{(K+1)}) - \bar{\eta}_r(n)\}\right| \right. \\ &\left. \leqq C_4 \{z_j^{1/2} (nB)^{1/4} (\log n)^{3/4} + B(\log n) \log[2[z_j + 1]]\},\ 1 \leqq j < B\right\} > 1 - n^{-1}. \end{aligned} \qquad (3.21)$$

The fact that $(z+1)^{-1} \max[z^{1/2}, \log 2(z+1)] < 1$, with (3.21), yields (3.1).

Remarks on Lemma 4. (1) This is our crucial estimate concerning the embedding of Section 2.3. The martingale structure of the $U_z^{(n)} - nz$, in both n and z, can be used to give an alternate proof, but that approach is also long due to the complication of $nQ_n(z) - \xi(z, n)$ being determined by stopping times dependent on ξ. This last effect is minimized in the present proof by the device which enables the sum in (3.10) to be treated in terms of Brownian motion deviations for fixed epochs which can be summed on j as independent normal rv's, leaving only the first difference on the right side of (3.10) for grosser path-dependent estimation. (2) We have been somewhat cavalier in the choice of constants in the limits entering into the proof, but a more careful choice would only alter C_4, not the order of the deviation in (3.1). However, the behavior of the embeddings as $p \to 0$ or 1 can obviously be sharpened; as one might expect, the behavior for $|p-1/2|<\varepsilon'$ accounts for most of the deviation. (3) We have given the proof for equally spaced p_j, for the sake of Corollary 1. The changes in estimates for other choices of the p_j should be clear. In particular, if B_n is bounded in n and the p_j are fixed, the bounds stated in (3.1) are valid; this is used in Corollary 2. (4) It is easily seen that, with a slight change in the values of ε and C_4 and an accompanying increase in the coefficient 2 in the definition of q_N in (2.24), one obtains (3.1) with $n Q_n(z_j)-\xi(z_j, n)$, replaced by $m Q_m(z_j)-\xi(z_j, m)$, for each $m \leqq n$, and with the probability bound $1-n^{-1}$ replaced by $1-n^{-3}$. As in the statement of Lemma 1', we require such a result for $m \leqq \bar{c}\, n(\log n)^{\beta}$, and it is easily verified that this essentially involves substituting this larger value for n, while leaving B_n unchanged, in the upper bound on $|t Q_t-\xi|$ of (3.21). With a slight additional argument this last comment yields

Lemma 4'. *There is an $\varepsilon'>0$ and $C_4'>0$ such that, for $\beta \geqq 0$, $\bar{c}>0$, and n sufficiently large,*

$$2 \leqq B_n \leqq \varepsilon' n/\log n$$
$$\Rightarrow P\{(z_j+1)^{-1}|t Q_t(z_j)-\xi(z_j, t)| \leqq 2 C_4'[(\bar{c}\, n[\log n]^{\beta+3} B_n)^{1/4}+B_n \log n] \quad (3.22)$$
$$\textit{for } 0 \leqq t \leqq \bar{c}\, n(\log n)^{\beta} \textit{ and } 1 \leqq j<B_n\} \geqq 1-n^{-1}.$$

The required additional argument is in fact given by

$$\begin{aligned} &0 \leqq \sup_{m \leqq t \leqq m+1,\, 0 \leqq z<\infty} (z+1)^{-1}|t Q_t(z)-m Q_m(z)| \leqq 1 \text{ wp } 1, \\ &P\{\sup_{m \leqq t \leqq m+1,\, 0 \leqq z<\infty} (z+1)^{-1}|\xi(z, t)-\xi(z, m)| \geqq 2 \log n\} \leqq C_5 n^{-3} \end{aligned} \quad (3.23)$$

for some constant C_5; the latter is easily proved by the methods of [8].

We can think of $\{t(z_j+1)^{-1} Q_t(z_j/(z_j+1))\}$ of (3.22) as an embedding of $\{t S_{B_n, t}(j/B_n)-t j/B_n,\ 1 \leqq j<B_n,\ 0 \leqq t \leqq \bar{c}\, n(\log n)^{\beta}\}$ of (1.18). It is easily seen that the maximum of the rates in (1.18) and (3.22) has its minimum when the two rates have the common value $n^{1/3}(\log n)^{(\beta+2)/3}$, at $B_n=n^{1/3}(\log n)^{(\beta-1)/3}$. Recalling how, in connection with Fig. 1, we could consider all of $k[S_k(x)-x]$, $x \in I$, to be embedded in ξ^*, we conclude easily from Lemma 1' and Lemma 4' (with, in fact, any power of n in place of n^{-1} on the right side of (3.24)),

Theorem 2. *For $\beta \geqq 0$ and $\bar{c}>0$ there is an embedding of $\{k[S_k(x)-x],\ k \leqq \bar{c}\, n(\log n)^{\beta},\ x \in I\}$ in ξ^* such that, for some constant c^*,*

$$\begin{aligned} P\{&\sup_{t \leqq \bar{c} n(\log n)^{\beta},\, x \in I} n^{-1/2}|t[S_t(x)-x]-\xi^*(x, t)| \\ &\geqq c^* n^{-1/6}(\log n)^{(\beta+2)/3}\}=O(n^{-1}). \end{aligned} \quad (3.24)$$

We have stated (3.24) with continuous t, for comparison with Theorem 1. As in Corollary 1, the main interest is of course in integral t for S_t and continuous t for ξ^*, although Corollary 1 remains essentially unchanged for any of the four possible combinations. (See Remark 3 on Lemma 6.) In Lemma 5 we consider only the combination stated in Corollary 1.

To obtain Corollary 1 from Theorem 2, we need only (3.23) and the familiar device used in such weak convergence results, e.g. by Skorohod, Müller, Sawyer, of using boundedness near 0 of the density of an appropriate functional of ξ^*. Let R^* be a subinterval of R^+ and write

$$\begin{aligned} R_n^* &= \{t\colon t\in R^* \text{ and } nt\in Z^+\}, \\ H_2 &= \sup_{s\in I,\, t\in R^*} [\xi^*(s,t)-G_2(s,t)], \\ H_1 &= -\inf_{s\in I,\, t\in R^*} [\xi^*(s,t)-G_1(s,t)]. \end{aligned} \tag{3.25}$$

The random function $\bar{\xi}(s,t)$ in (3.26) below can be any function defined on the same probability space as ξ^*, although in proving Corollary 1 we take $\bar{\xi}(s,t)$ to be our embedding of $n^{1/2}\,t\,[S_{nt}(s)-s]$.

Lemma 5. *Suppose $P\{0\leqq H_i<\delta+\delta'\}<C(\delta+\delta')$ for $i=1$ or 2, that R_n^* is non-empty, and that $|G_i(s,t_1)-G_i(s,t_2)|<C'|t_1-t_2|^\alpha$ whenever $C'|t_1-t_2|^\alpha<\delta'$, for $s\in I$, $t_j\in R^*$. Then*

$$\begin{aligned} 2n^{-1/2}\log n + C'n^{-\alpha}<\delta' \Rightarrow |&P\{G_1(s,t)<\xi^*(s,t)<G_2(s,t),\,(s,t)\in I\times R^*\} \\ &-P\{G_1(s,t)<\bar{\xi}(s,t)<G_2(s,t),\,(s,t)\in I\times R_n^*\}| \\ &<2C_5|R_n^*|\,n^{-3}+2C(\delta+\delta')+P\{\sup_{I\times R_n^*}|\xi^*(s,t)-\bar{\xi}(s,t)|\geqq\delta\}. \end{aligned} \tag{3.26}$$

Proof. A quarter of the demonstration is that, if $|\xi^*(s,t)-\bar{\xi}(s,t)|<\delta$ on $I\times R_n^*$ and $\bar{\xi}(s,t)<G_2(s,t)$ on that set, then $\xi^*(s,t)<G_2(s,t)+\delta$ thereon and hence (with $\xi^*(s,t)$ here $=n^{-1/2}(1-s)\,\xi(s(1-s)^{-1},nt)$ of (3.23)), except on a set of probability $2C_5|R_n^*|\,n^{-3}$, we have $\xi^*(s,t)<G_2(s,t)+\delta+2n^{-1/2}\log n+C'n^{-\alpha}$ on $I\times R^*$; consequently, $\xi^*(s,t)$ can be $\geqq G_2(s,t)$ somewhere on $I\times R^*$ with probability at most $C(\delta+\delta')$. For $\xi^*(s,t)<G_2(s,t)\leqq\bar{\xi}(s,t)$ on R_n^* the only contribution is from the last term of (3.26).

What remains to be verified in order that Theorem 2 and Lemma 5 can be applied in particular cases is of course the boundedness near 0 of the densities of the H_i. This has been proved in particular cases by Skorohod (for I replaced by a point and $R^*=(0,A]$) and by Müller (for $R^*=[1,\infty)$, essentially with $b_\infty\geqq 1/2$ in (3.34) below while assuming no Lipschitz condition but certain other restrictions described in Remark (1) on Lemma 6). A simple condition and proof (more elementary than Müller's) are given in

Lemma 6. *Suppose the G_i are continuous with $|G_i(s,t)|<G(t)$ and $|G_i(s,t)-G_i(s',t)|<\bar{G}(t)\,|s-s'|$ for $s,s'\in I$ and $t\in R^*$. Also (if $0\in\bar{R}^*$) suppose* $\inf\{t\colon \xi^*(s,t)=G_i(s,t)$ *for some i and s*$\}>0$ wp 1. *Then, for $\bar{\delta}>0$,*

$$P\{0\leqq H_i<\bar{\delta}\}\leqq 6\bar{\delta}\inf_{t\in R^*}\{t^{-1/2}+2t^{-1}[G(t)+\bar{G}(t)]\}. \tag{3.27}$$

Proof. We shall give the proof for H_2. Let

$$\begin{aligned}\gamma_1 &= \inf\{s\colon \sup_{t\in R^*}[\xi^*(s,t)-G_2(s,t)]=0\},\\ 1-\gamma_2 &= \inf\{s\colon \sup_{t\in R^*}[\xi^*(1-s,t)-G_2(1-s,t)]=0\}.\end{aligned} \tag{3.28}$$

The γ_i are defined on the same set Γ (say) of sample paths, on which $\gamma_1 \leqq \gamma_2$. On Γ, define

$$\tau_i = \inf_{t\in R^*}\{t\colon [\xi^*(\gamma_i,\tau_i)-G_2(\gamma_i,\tau_i)]=0\}. \tag{3.29}$$

Then, given Γ and $\gamma_1 = s_1 \leqq 1/2$, $\tau_1 = t_1 > 0$, the process $\{t_1^{-1/2}\xi^*(s,t_1),\ s_1 \leqq s \leqq 1\}$ has the same conditional law as $\{\xi^*(s,1),\ s_1 \leqq s < 1\}$ given that $\xi^*(s_1,1) = t_1^{-1/2} G_2(s_1,t_1)$. The obvious analogue holds for time variable $1-s$ given $\gamma_2 = s_2 \geqq 1/2$, $\tau_2 = t_2 > 0$. Since always $\gamma_1 \leqq 1/2$ or $\gamma_2 \geqq 1/2$ on Γ,

$$\begin{aligned}P\{0\leqq H_2<\bar\delta\} &\leqq P\{0\leqq H_2<\bar\delta|\Gamma\}\\ &\leqq P\{0\leqq H_2<\bar\delta,\gamma_1\leqq 1/2|\Gamma\}+P\{0\leqq H_2<\bar\delta,\gamma_2\geqq 1/2|\Gamma\}.\end{aligned} \tag{3.30}$$

We treat in detail only the first term on the right side of (3.30), which is no greater than the infimum over $s_1 \leqq 1/2$ and t_1 of

$$P\{0\leqq \sup_{s_1\leqq s\leqq 2/3}[\xi^*(s,t_1)-G_2(s,t_1)]<\bar\delta|\Gamma,\gamma_1=s_1,\tau_1=t_1\}. \tag{3.31}$$

Writing, as earlier, $(z+1)\,\xi^*(s,t)=\xi(z,t)$ and $z_1 = s_1(1-s_1)^{-1}$, and using

$$|G_2(s,t)-G_2(s_1,t)| < \bar G(t)\,|z-z_1|/(1+z)(1+z_1),$$

we have

$$\begin{aligned}\xi^*(s,t_1)-G_2(s,t_1) &= (z+1)^{-1}\{\xi(z,t_1)-(z-z_1)\,G_2(s,t_1)\\ &\qquad -(z_1+1)[G_2(s,t_1)-G_2(s_1,t_1)]-\xi(z_1,t_1)\}\\ &\geqq (z+1)^{-1}\{\xi(z,t_1)-\xi(z_1,t_1)-(z-z_1)[G(t)+\bar G(t)]\}.\end{aligned} \tag{3.32}$$

Substituting $z_1 \leqq 1$, $z \leqq 2$, we see that the probability of (3.31) is no greater than

$$\begin{aligned}P\{&\sup_{0\leqq z-z_1\leqq 1} t_1^{-1/2}\{\xi(z,t_1)-\xi(z_1,t_1)-(z-z_1)[G(t)+\bar G(t)]\}<3\bar\delta t_1^{-1/2}\}\\ &\leqq 3\bar\delta t_1^{-1/2}\{1+2t_1^{-1/2}[G(t)+\bar G(t)]\},\end{aligned} \tag{3.33}$$

the last by a Brownian motion estimate (e.g., [20], p. 173).

Remarks on Lemma 6. (1) The dependence on s can be weakened, both in allowing a weaker Lipschitz condition and also in allowing G_i and its differences to vanish, but sufficiently slowly, near $s=0$ or 1 (in particular, Müller's assumption that $\liminf_{s(1-s)\downarrow 0} G_i > 0$ is unnecessary); we forego the altered statement of Lemma 6 and Corollary 1. (2) Somewhat different conditions can be given using the "first t" in place of (3.28), but the present conditions and proof seem, on the whole, more expeditious. (3) If also $\inf_{I\times R^*} G_2 > \bar\delta > 0$, the analogous result is obtained for $P\{-\bar\delta < H_2 \leqq 0\}$, from the above with G_2 replaced by $G_2-\bar\delta$; this would be needed only for the somewhat artificial combination (in the paragraph following (3.24)) of continuous t for S_t and discrete t for ξ^*.

Assumptions and Proof of Corollary 1. We adopt the following notation to describe the behavior of the G_i:

For positive constants $C_0, L, a_\infty, b_\infty, b_0$, and real a_0,

$$|G_i(s,t)|+|s-s'|^{-1}|G_i(s,t)-G_i(s',t)|<\begin{cases} C_0\, t^{-a_0}, & t\leqq 1,\\ C_0\, t^{a_\infty}, & t\geqq 1;\end{cases}$$

$$|G_i(s,t_1)-G_i(s,t_2)|<C_0(t_1^L+t_1^{-L})\,|t_2-t_1|^\alpha \tag{3.34}$$

for $0<t_2-t_1<1$ and some $\alpha>1/6$;

$$(-1)^i\, G_i(s,t)>\begin{cases} C_0^{-1}\, t^{1/2-b_0}, & t\to 0,\\ C_0^{-1}\, t^{1/2+b_\infty}, & t\to\infty;\end{cases}$$

$$\gamma=2/3+1/6b_\infty+\max\left[(1+a_0)/2b_0,\ 1/4b_\infty,\ (a_\infty-1)/2b_\infty\right],$$

where the terms with subscript 0 (resp., ∞) are omitted if $0\notin\overline{R^*}$ (resp., if R^* is bounded).

The last assumption on the G_i implies (from standard estimates for $\sup_{s\in I,\,0\leqq t\leqq T}[\xi^*(s,t)$ or $n^{1/2}t|S_t(s)-s|]$, as in [6, 8]), that for some $c'>0$

$$\begin{aligned} P\{G_1(s,t)<\xi^*(s,t),\ n^{1/2}\,t[S_{nt}(s)-s]<G_2(s,t)\text{ for } s\in I\\ \text{and } t<c'(\log n)^{-1/2b_0}\text{ or } t>c'(\log n)^{1/2b_\infty}\}>1-n^{-1}.\end{aligned} \tag{3.35}$$

It suffices, therefore, to prove (1.6) for $R^*\cap[c'(\log n)^{-1/2b_0}, c'(\log n)^{1/2b_\infty}]$. Consequently, the coefficient of $\bar\delta$ in (3.27) (and hence the C of Lemma 5) is of order $(\log n)^\gamma$ where $\gamma=\max[1/4b_0,\ (1+a_0)/2b_0,\ 1/4b_\infty,\ (a_\infty-1)/2b_\infty]$. Note that $1/2-b_0\geqq -a_0$, so that $(1+a_0)/2b_0\geqq 1/4b_0$. The second assumption of (3.34) allows C' to be taken to be a power of $\log n$ in Lemma 5, and thus, since $\alpha>1/6$, δ' can be taken as $n^{-1/6}$. From Theorem 2 with $\beta=1/2b_\infty$, we thus obtain that the right side of (3.26) (with $\delta=c^*\,n^{-1/6}(\log n)^{(\beta+2)/3}$) is of order

$$\delta C=n^{-1/6}(\log n)^{\gamma+(\beta+2)/3},$$

which yields (1.6).

Remark. The modifications for dependence in s (near 0 and 1) and bounds of orders other than monomials in (3.34), are straightforward. The third condition, where we have used positive powers b_0 and b_∞ rather than suitable slowly varying functions, is but the simplest condition to yield a result like (3.35) and avoid the trivial case where the second probability of (1.6) is 1; in the latter degenerate case one can also obtain estimates like (1.6). Of course, $a_\infty\geqq 1/2+b_\infty$.

Proof of Corollary 2. (Of course, the $|s-s'|$ term in (3.34) can be omitted.) B_n is now fixed in (3.22), and the proof of Lemma 4′ obviously holds even though the p_j need not be uniformly spaced in I'. The analogue of Lemma 5 is valid, and Lemma 6 can be replaced by an analogous result following the lines of Skorohod's corresponding result for I' a point. The remainder of the proof of (1.7) is as for

Corollary 1, and (1.8) follows from the fact that probabilities corresponding to the n^{-1} of (3.35) can be made n^{-2} by merely multiplying the deviations considered in Lemma 1′, 4′, etc. by suitable constants.

4. Proof of Theorem 1

We shall give details of the features not present in the proof of Theorem 2, but will merely sketch the differences when parts of the argument are similar to those of the proofs of Lemma 1′ and 4′, etc.

We must first alter the embedding by letting B_n and the z_j vary with the n-th observation. (The use of Section 3, with n varying, would only yield the existence of $\{T_n(s,t), n\in Z^+\}$ so that (1.5) held with T_n for T there whenever $t\leqq n$.) We shall make a single choice of the parameters of the embedding from the outset, to yield the best order obtainable in (1.5) from the analogues of Lemmas 1′ and 4′. As in Section 3, we shall save space by not distinguishing notationally between large real values and their integral parts.

We shall keep even less track of constants than in Section 3, and will denote by $\Omega(1)$ any positive function (whose domain will depend on the context) which is bounded away from 0 and ∞ and whose definition may change from one usage to the next. $O(1)$ is used similarly.

Define

$$\begin{aligned} M_r &= r\,8^r \quad \text{for } r\geqq 0,\\ B_n &= 2^r \quad \text{for } M_{r-1}<n\leqq M_r, \end{aligned} \tag{4.1}$$

so that

$$B_n=\Omega(1)\,(n/\log n)^{1/3}. \tag{4.2}$$

Whenever we are using 2^r subdivision of I, they are of equal length, and the corresponding z_j, here termed $z_j^{(r)}$, are given by (3.2) with $B_n=2^r$.

For $0<n\leqq M_1$ we use the embedding of Section 2 at the single value $z_1^{(1)}=1(p_1=1/2)$, and denote the embedding by $\eta(0,z_1^{(1)};\, U_{[1,1]}^{(n)})$; the subscript $[r;j]$ abbreviates $(z_{j-1}^{(r)}, z_j^{(r)})$. Since the "last $z_j^{(r)}$", which by (4.1) is $B_n-1=2^r-1$, is going to vary with n, it is convenient (although it is not strictly necessary), here and below, to adjoin $\{\eta(z',z'';n),\, z',z''\geqq B_n-1\}$ to the $\{\eta(z_{j-1}^{(r)}, z_j^{(r)};\, W_{[r,j]}^{(n)}),\, j<B_n\}$ (with $W=U$ above and as defined generally below), so as to have the expected value of the last argument of η be n regardless of the first two arguments. We call $A=\{(p,n)\colon p>1-B_n^{-1}\}$ the "adjoined region". We then define, for $n\leqq M_1$,

$$n\,\tilde{Q}_n(z)=\begin{cases}\eta(0,z;\,U_{[1,1]}^{(n)}) & \text{for } 0<z\leqq z_1^{(1)},\\ \eta(0,z_1^{(1)};\,U_{[1,1]}^{(n)})+\eta(z_1^{(1)},z;\,n) & \text{for } z\geqq z_1^{(1)}.\end{cases} \tag{4.3}$$

(An alternative to adjoining for $z\geqq B_n-1$ as above is in fact to define $Q_n(z)=Q_n(B_n-1)$ for $z\geqq B_n-1$. This means deleting the rectangle $[1,3]\times[0,M_1]$ in Fig. 2 and lowering what is above it. A subsequent modification five paragraphs below must then be altered accordingly.)

We write $j'=b$ if $j=2b$ or $2b-1$. This means the two intervals $[z_{2j'-2}^{(r)}, z_{2j'-1}^{(r)}]$ and $[z_{2j'-1}^{(r)}, z_{2j'}^{(r)}]$ make up $[z_{j'-1}^{(r-1)}, z_{j'}^{(r-1)}]$.

We continue as follows: Suppose $\tilde{Q}_n$ is defined for $n\leqq M_{r-1}$. Then, for $M_{r-1}<n\leqq M_r$, we define $n\tilde{Q}_n-M_{r-1}\tilde{Q}_{M_{r-1}}$ at the 2^r-1 arguments $z_j^{(r)}$ by using

stopping times $W^{(n)}_{[r,j]}-W^{(M_{r-1})}_{[r-1,j']}$ on $\{\eta(z^{(r)}_{j-1}, z^{(r)}_j;\ t+W^{(M_{r-1})}_{[r-1,j']}),\ t\geqq 0\}$ in the same way $(z^{(r)}_j - z^{(r)}_{j-1})\, U^{(n-M_{r-1})}_{[r,j]}$ was used to obtain $(n-M_{r-1})\, Q_{n-M_{r-1}}$ from the $\eta(z_{j-1}, z_j; t)$ for $B=2^r-1$ in Section 3. Our "adjoining" described above (4.3) makes $W^{(M_{r-1})}_{[r-1,(2^r-1)']}\equiv M_{r-1}$ for all r. Finally, we write, for $M_{r-1}<n\leqq M_r$ and $z_{j-1}\leqq z\leqq z_j$,

$$\begin{aligned} n\,\tilde{Q}_n(z)-M_{r-1}\,\tilde{Q}_{M_{r-1}}(z)= &\sum_{\alpha=1}^{j-1}[\eta(z^{(r)}_{\alpha-1}, z^{(r)}_\alpha;\ W^{(n)}_{[r;\alpha]})-\eta(z^{(r)}_{\alpha-1}, z^{(r)}_\alpha;\ W^{(M_{r-1})}_{[r-1,\alpha']})] \\ &+[\eta(z^{(r)}_{j-1}, z;\ W^{(n)}_{[r,j]})-\eta(z^{(r)}_{j-1}, z^{(r)}_j;\ W^{(M_{r-1})}_{[r-1,\alpha']})], \end{aligned} \tag{4.4}$$

with the last term in square brackets replaced by

$$[\eta(z^{(r)}_{2^r-1}, z;\ n)-\eta(z^{(r)}_{2^r-1}, z;\ M_{r-1})] \quad \text{if} \quad z>z_{2^r-1}=2^r-1.$$

Here is a diagram of $\tilde{Q}_{M_1+1}(2)$, in Fig. 2: it is ξ (shaded area).

Note especially: as r increases, bases of rectangles are piled on top of broader, lower ones; the region $[1,3]\times[0,M_1]$ comes from the "adjoined" definition which makes the *expected* height of the shaded part equal M_1+1 for each z; the value 2 is not a $z^{(2)}_j$, and so we must split $[1,3]\times[M_1, W^{(M_1+1)}_{[2,3]}]$ vertically although this rectangle was determined by stopping the process with argument $[1,3]\times[M_1, M_1+t]$ by moving up its horizontal upper boundary. This last is a technical convenience in defining $\tilde{Q}$, but will necessitate proving in (4.8) that $\tilde{Q}$ is suitably close to a corresponding interpolated process analogous to that defined in (1.12). Thus, the full $\{n[(z+1)\,S_n(z/(z+1))-z],\ z\leqq 1,\ n\in Z^+\}$ (which can be modified in an obvious way from $\{z\leqq 1\}$ to $\{z\leqq z^{R_0}-1\}$ for any fixed R_0) still sits in ξ in the above picture, but it is given in terms of the horizontal cuts of Fig. 1 rather than the vertical cuts used in $\tilde{Q}_n$, of "whole rectangles" like $[0,1]\times[0, W^{(1)}_{[1,1]}]$.

Theorem 1 requires this embedding for $\{z<\infty\}$, not just $\{z\leqq \text{large constant}\}$. The additions we must make to achieve this turn out to be trivial, since it will turn out that we do not need to estimate *differences* between ξ and the embedded $n\,S_n(p)-M_{r-1}\,S_{M_{r-1}}(p)$ for $p>1-B_n^{-1}$ (when $M_{r-1}<n\leqq M_r$) in obtaining our estimates; this will be seen explicitly in connection with the discussion of (4.16). Consequently, our embedding can be completed in $A=\{(p,n)\colon p>1-B_n^{-1}\}$ in any convenient way that exhibits the presence of $S_n(p)$. One possibility is that men-

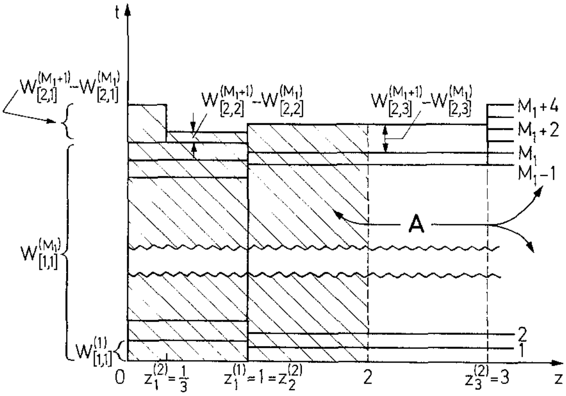

Fig. 2. Example of $Q_{M_1+1}(2)=\xi$ (shaded area)

tioned in connection with Fig. 1, of adjoining any sequence of unbounded $z_j^{(r)} \geqq B_n$. A simpler alternative is to use the standard process $\{\eta(2^r-1, 2^r-1+t; n) - \eta(2^r-1, 2^r-1+t; n-1), t \geqq 0\}$ already exhibited in the "adjoined" region A of our definition of $\tilde{Q}_n$, in the manner the ξ_i' were used in Section 2.1, stopping this process the first time the resulting sample df of the single n-th observation reaches 1. The deficiencies noted for this scheme in Section 2 are now irrelevant for the reason stated earlier in this paragraph.

We define $\tilde{\xi}(s, n) = (1-s)\, n\, \tilde{Q}_n(s/(1-s))$ (with $\tilde{\xi}(1, n) = 0$), including region A in the definition. Recall that $\tilde{\xi}(s, n) - \tilde{\xi}(s, M_{i-1})$ gives an exact embedding of $[n\, S_n(s) - M_{i-1}\, S_{M_{i-1}}(s) - (n - M_{i-1})\, s]$ *for* $s =$ *integral multiples of* 2^{-i}, if $M_{i-1} < n \leqq M_i$. For $M_{i-1} < n \leqq M_i$ define

$$\xi^\#(s, n) - \xi^\#(s, M_{i-1}) = \begin{cases} \tilde{\xi}(s, n) - \tilde{\xi}(s, M_{i-1}) & \text{if} \quad s = 2^{-i} j, j \text{ integral}, \\ \text{linearly interpolated from the above if} & \\ \qquad 2^{-i} j < s < 2^{-i}(j+1). & \end{cases} \tag{4.5}$$

Thus, if $M_{r-1} < n \leqq M_r$, the rv $\xi^\#(2^{-r} j, n)$ is a sum of r independent rv's, one with exactly the distribution of $[n S_n(2^{-r} j) - M_{r-1} S_{M_{r-1}}(2^{-r} j) - (n - M_{r-1})\, 2^{-r} j]$, and the i-th of the others with the distribution of the rv linearly interpolated from $[M_i S_{M_i}(p) - M_{i-1} S_{M_{i-1}}(p) - (M_i - M_{i-1})\, p]$ at the values of $p = 2^{-i} \bar{j}$ closest to $2^{-r} j$. Finally, in conformity with (1.3), we define $\{\hat{\xi}(s, n)\}$ to be the embedded process distributed *exactly* as $\{n S_n(s) - n\, s, s \in I, n \in Z^+\}$; this is obtained from the horizontal cuts described at the end of the second preceding paragraph for the i-th observation when $s/(1-s) \leqq B_i - 1$, and from any exact representation used in the manner of the paragraph just above, for $s/(1-s) > B_i - 1$ (region A).

Our proof is divided into four parts:

$$P\{\sup_{s \in I} |\xi^*(s, n) - \xi^*_{B_n}(s, n)| = O(n^{1/3} (\log n)^{2/3})\} > 1 - n^{-2}, \tag{4.6}$$

$$P\{\max_{0 < j < B_n} |\xi^*(B_n^{-1} j, n) - \tilde{\xi}(B_n^{-1} j, n)| = O(n^{1/3} (\log n)^{2/3})\} > 1 - n^{-2}, \tag{4.7}$$

$$P\{\max_{0 < j < B_n} |\tilde{\xi}(B_n^{-1} j, n) - \xi^\#(B_n^{-1} j, n)| = O(n^{1/3} (\log n)^{2/3})\} > 1 - n^{-2}, \tag{4.8}$$

$$P\{\sup_{s \in I} |\xi^\#(s, n) - \hat{\xi}(s, n)| = O(n^{1/3} (\log n)^{2/3})\} > 1 - n^{-2}. \tag{4.9}$$

These four equations, for $n \in Z^+$, yield (1.5) for integral t; the corresponding result for linearly interpolated (in t) $t S_t$ follows at once from (3.23).

Remark 4 on Lemma 1 yield (4.6), in view of (4.1)–(4.2).

We require an analogue of Lemma 4, to obtain (4.7). Suppose $M_{r-1} < n \leqq M_r$. Temporarily fix $p_{j_r}^{(r)} = 2^{-r} j_r$, $1 \leqq j_r < 2^r$. Denote by $p_{j_i}^{(i)}$ the corresponding right endpoint of the interval of length 2^{-i} (with endpoints multiples of 2^{-i}) containing $[p_{j_r-1}^{(r)}, p_{j_r}^{(r)}]$; thus, $j_{i+1}' = j_i$ and $(j_{i+1} - 1)' = j_i$ for $i < r$. We write $z_{j_i}^{(i)}$ correspondingly. Also, define $\rho \equiv \rho(p_{j_r}^{(r)}) = \min\{i : p_{j_r}^{(r)} \leqq 1 - 2^{-i}\}$. Then $W_{[\rho-1, j_\rho-1]}^{(M_{\rho-1})} = M_{\rho-1}$ (the "adjoined" region A in the definition of $\tilde{Q}_n$), while for $\rho \leqq i \leqq r$ the stopping time $W_{[i, j_i]}^{(M_i)} - W_{[i-1, j_i-1]}^{(M_{i-1})}$ is the sum of $M_i - M_{i-1}$ iidrv's, each satisfying (2.19) with (z_1, z_2) there replaced by $(z_{j_i-1}^{(i)}, z_{j_i}^{(i)})$. (The first $M_{\rho-1}$ summands together yield

$$\log E \exp\{\alpha(W^{(M_{\rho-1})} - M_{\rho-1})\} = 0$$

in place of (2.19).) All of these intervals $[z_{j_i-1}^{(i)}, z_{j_i}^{(i)}]$ are contained in $[z_{j_\rho-1}^{(\rho)}, z_{j_\rho}^{(\rho)}]$, from which it follows that, in the expression h of (2.19)–(2.20), all factors $(1+z_{j_i}^{(i)})$ (replacing $(1+z_2)$ there) are $\Omega(1)(1+z_{j_r}^{(r)})$ for $\rho \leqq i \leqq r$. (The $\Omega(1)$ here is independent of n, r, j_r.) Also, $z_{j_i}-z_{j_i-1}=\Omega(1)(1+z_{j_i})^2 2^{-i}$. Hence, from (2.20), we have as analogue of (2.19) for $M_{i-1}<v \leqq M_i$, in terms of stopping times with unit expectation,

$$\begin{aligned}\log E \exp\{\alpha[W_{[i,j_i]}^{(v)}-W_{[i,j_i]}^{(v-1)}-1]\} &\sim \alpha^2(z_{j_i}^{(i)}-z_{j_i-1}^{(i)})^{-2}\, h(z_{j_i-1}^{(i)}, z_{j_i}^{(i)})\\ &=\alpha^2\,\Omega(1)2^i \leqq \alpha^2\,\Omega(1)B_n.\end{aligned} \tag{4.10}$$

A sum $\overline{W}^{(N)}$ (say) of N such terms (some replaced by 0 if $i<\rho$) yields, in place of (2.21), with $\alpha=q/2\Omega(1)NB_n$,

$$\begin{aligned}P\{\overline{W}^{(N)}>q\} &\leqq E\{\exp[\alpha\,\overline{W}^{(N)}-q]\}\\ &\leqq \exp\{N\alpha^2\,\Omega(1)B_n-\alpha q\}\\ &=\exp\{-q^2/4\Omega(1)NB_n\}.\end{aligned} \tag{4.11}$$

This last is $<n^{-3}$ if $q=\Omega(1)\,(NB_n \log n)^{1/2}$; this and the corresponding result for $\{W^{(N)}<q\}$ make up the analogue of (2.24)–(2.25) (with q here for $(z_j-z_{j-1})^{-1}q_N$ there), and the condition of the first line of (2.25), inherited from those following (2.19), is again seen to be that $N^{-1}B_n \log n <$ some small positive value.

We use the above to prove (4.7) by going through the proof of Lemma 4 with $W_{[r,j_r]}^{(n_1+\cdots+n_i)}$ in place of the $V_j^{(i)}$ of (3.9). We use (3.3)–(3.8) as before (and in view of (4.1) we have $K\sim(\log\log n)/\log 2$). The analogue of (3.12) is valid, since the proof of (3.11) depends only on (3.5). We replace (z_j-z_{j-1}) in (3.13) by $(z_{j_r}^{(r)}-z_{j_r-1}^{(r)})$ and now sum on j_r to obtain the obvious counterparts of (3.14), (3.15), (3.16). These and the analogues of (3.17)–(3.20), with the coefficients 2 and 2^2 altered slightly, yield the analogue of (3.21) with the probability changed to $1-n^{-2}$. Thus, (4.7) follows from the fact that

$$(nB_n)^{1/4}(\log n)^{3/4}+B_n \log n=2n^{1/3}(\log n)^{2/3}. \tag{4.12}$$

We turn to (4.8). We define $\rho(p)$ just as above (4.10). Still supposing $M_{r-1}<n \leqq M_r$, fix $p_{j_r}^{(r)}$ and abbreviate $(p_{j_i-1}^{(i)}, p_{j_r}^{(r)}, p_{j_i}^{(i)})$ by $(p', \bar{p}, p'')$. We first look at that portion Ξ (say) of $\tilde{\xi}(\bar{p},n)-\xi^{\#}(\bar{p},n)$ that comes only from values $t \geqq M_{\rho(p)}$ for each p, $0<p\leqq\bar{p}$. This is the portion from those (s,t) of $I\times R^+$ outside (above and to the left of) the "adjoined" region A. Since, moreover, the (noninterpolated) contributions from values $p \leqq p'$ cancel in the difference Ξ, we are left with

$$\begin{aligned}\Xi=&\{[\tilde{\xi}(\bar{p},n)-\tilde{\xi}(\bar{p},M_{\bar{\rho}-1})]-[\tilde{\xi}(p',n)-\tilde{\xi}(p',M_{\bar{\rho}-1})]\}\\ &-\{[\xi^{\#}(\bar{p},n)-\xi^{\#}(\bar{p},M_{\bar{\rho}-1})]-[\xi^{\#}(p',n)-\xi^{\#}(p',M_{\bar{\rho}-1})]\},\end{aligned} \tag{4.13}$$

where $\bar{\rho}=\rho(\bar{p})$. This difference Ξ may be thought of as a sum of $M_{r-1}-M_{\bar{\rho}-1}$ rv's, corresponding to the $(M_{\bar{\rho}-1}+1)$-th to M_{r-1}-th observation; there is no interpolation, linear or by vertical cuts, from the $(M_{r-1}+1)$-th to n-th observation. For each i, $\bar{\rho}\leqq i \leqq r-1$, the M_i-M_{i-1} rv's corresponding to the observations in that i-th group are iid, and each has a distribution which is the same as that of $\xi^{***}(\pi_i, W^*)$ of (4.14) (below), which we now describe in terms of the notation of Section 2, translating time to begin at $t=0$ for notational ease: For a process ξ^{**} distributed like ξ^*, stop $\xi^{**}(p'',t)-\xi^{**}(p',t)$ at time $t=(z_{j_i}-z_{j_i-1})^{-1}T^{(1)}_{[i,j_i]}=W^*$ (say). The contribution of ξ^{**} from the time rectangle $[p',p'']\times[0,W^*]$ to the $\tilde{\xi}$ part of (4.13) is $\xi^{**}((1-\pi_i)p'+\pi_i p'',W^*)-\xi^{**}(p',W^*)$, where $\pi_i=(\bar{p}-p')/(p''-p')$ is the interpolating proportion. Similarly, the contribution to the $\xi^{\#}$ part of (4.13) is $(1-\pi_i)\xi^{**}(p',W^*)+\pi_i\xi^{**}(p'',W^*)-\xi^{**}(p',W^*)$. The difference between these two contributions can be written as $\xi^{***}(\pi_i,W^*)$, where

$$\begin{aligned}\xi^{***}(\pi,W^*)=\xi^{**}&((1-\pi)p'+\pi_i p'',W^*)\\&-(1-\pi)\xi^{**}(p',W^*)-\pi\,\xi^{**}(p'',W^*).\end{aligned}\tag{4.14}$$

Recall that W^* was defined in terms of the values of $\xi^{**}(p'',t)-\xi^{**}(p',t)$. A simple computation in terms of the Gaussian distribution (more easily done in terms of the corresponding independent increment z-process for which, however, one does not have the simple linear interpolation obtained by putting z', z'' for p', p'' in (4.14)) now shows that the conditional law of $\{\xi^{***}(\pi,w),0\leqq\pi\leqq 1\}$, given $W^*=w_0$ and even the whole path $\{\xi^*(p'',t)-\xi^{**}(p',t),t\leqq W^*\}$, is the same as the unconditional law of $\{\xi^{***}(\pi,w_0),0\leqq\pi\leqq 1\}$, namely, it is that of $(p''-p')^{1/2}w_0^{1/2}$ times a standard Brownian bridge. (This simple dependence on w_0 of course differs strikingly from the conditional distribution of $\hat{\xi}$, obtained from the same horizontal cuts that defined W^*.) Consequently, given $W^{(M_i)}_{[i,j_i]}-W^{(M_{i-1})}_{[i,j_i]}=w_i$, the total contribution to $\tilde{\xi}(\bar{p},n)-\xi^{\#}_{B_n}(\bar{p},n)$ from the $(M_{i-1}+1)$-th to M_i-th observation is that of $w_i^{1/2}2^{-i/2}\xi_i^{**}(\pi_i)$ where ξ_i^{**} is a standard Brownian bridge. If the ξ_i^{**}'s are taken to be independent in i, we thus have, finally, the representation (given the w_i as above for $i<r$)

$$\Xi=\sum_{i=\bar{p}}^{r-1}w_i^{1/2}2^{-i/2}\xi_i^{**}(\pi_i).\tag{4.15}$$

It is now simple to use exponential bounds in the manner of the proof of Lemma 1: $(W^{(M_i)}_{[i,j_i]}-W^{(M_{i-1})}_{[i,j_i]})/(M_i-M_{i-1})$ is close to 1 for all large $i<r$, with high probability; with that probability, the rv's taking on the values w_i are such that

$$\sum_{i=\bar{p}}^{r-1}w_i 2^{-i}=O(n^{2/3}(\log n)^{1/3});$$

also, $E\exp\alpha\,\xi_i^{**}(\pi_i)\leqq e^{\Omega(1)\alpha^2}$. We obtain that the r.v. Ξ of (4.13) and (4.15) is $O(n^{1/3}(\log n)^{2/3})$ for all $\bar{p}=p^{(r)}_{j_r}$ $(1\leqq j_r<B_n)$, with probability $>1-(2n)^{-2}$.

To complete the proof of (4.8) we shall show

$$\begin{aligned}P\Big\{\sum_{p=2^{-r}j<\bar{p}}[\tilde{\xi}(p,M_{\rho(p)-1})-\tilde{\xi}(p-2^{-r},M_{\rho(p)-1})]=O(n^{1/3}(\log n)^{2/3})&\\ \text{for }\bar{p}=2^{-r}\bar{j},\ 0<\bar{j}<2^r\Big\}>1-(4n)^{-2}&\end{aligned}\tag{4.16}$$

and the corresponding result for $\xi^{\#}$ replacing $\tilde{\xi}$ in (4.16). (This separate treatment of the $\tilde{\xi}$ and $\xi^{\#}$ parts confirms our earlier assertion that the method of embedding in the "adjoined" region A was irrelevant.) Recalling that $M^{-1/2}\tilde{\xi}(p,M)$ is a standard Brownian bridge in the adjoined region A, and that (p,M_i) and (p,M_{i-1}) are in that region if and only if $p>1-2^{-i}$, we see that the sum in (4.16) has variance no greater than

$$\sum_{i=1}^{r}(M_i-M_{i-1})2^{-i}(1-2^{-i})<\Omega(1)\,n^{2/3}(\log n)^{1/3}. \qquad (4.17)$$

Moreover, if we denote the sum in (4.16) by $\eta_n(1-\bar{p})$, it is easy to see, as in [8] Lemma 2, that $P\{\sup_{0\leqq s\leqq 1/2}|\eta_n(s)|>\lambda\}\leqq O(1)P\{|\eta_n(1/2)|>\lambda/2\}$. Consequently, (4.17) and the usual normal exponential bound yield (4.16). The corresponding result for $\xi^{\#}$ is very similar; the crucial analogue of the abovementioned parallel of [8] now uses not just sample d.f.'s, but sums of random functions *obtained from sample df's linearly interpolated in s at various different spacings.* In the next paragraph we compute bounds of the required type for such sums which arise in treating the more difficult region complementary to A, and the result for $\xi^{\#}$ in (4.16) follows similarly.

We thus turn finally to (4.9). The discussion of the previous paragraph enables us to limit discussion to the complement of A; that is, corresponding to (4.15), to sums of random functions

$$H(p)=\sum_{i=\rho(p)}^{r-1}(M_i-M_{i-1})[S_{(i)}(p)-S^{\#}_{(i)}(p)], \qquad p^{(r)}_{j_r-1}\leqq p\leqq p^{(r)}_{j_r}, \qquad (4.18)$$

where the $S_{(i)}$ are independent sample d.f.'s, $S_{(i)}$ based on M_i-M_{i-1} observations, and $S^{\#}_i$ is the piecewise linear function interpolated from values $S_{(i)}(2^{-i}j)$, j integral. As in (4.13), it is only the contribution from $p^{(i)}_{j_i}$ to $\bar{p}$ that doesn't cancel out in the i-th summand of (4.18). Suppose we set $\varepsilon^{-1}=2^{r+4}$ and can show that, for $\lambda\geqq 8$,

$$\begin{aligned}&P\{\sup_{0\leqq\pi\leqq 1}|H((1-\pi)\,p^{(r)}_{j_r-1}+\pi\,p^{(r)}_{j_r})|\geqq\lambda\}\\&\quad=O(1)\,P\{\max_{0\leqq j\leqq\varepsilon^{-1}}|H(1-j\,\varepsilon)\,p^{(r)}_{j_r-1}+j\,\varepsilon\,p^{(r)}_{j_r}|\geqq\lambda/2\},\end{aligned} \qquad (4.19)$$

with $O(1)$ independent of r. Then, as in [8] and [12a], it is easily seen that the right side of (4.19) can be estimated by exponential bounds of the type we have used repeatedly, to yield (4.9). We suppose the first p in $[p^{(r)}_{j_r-1},p^{(r)}_{j_r}]$ where $|H(p)|\geqq\lambda$, say $p=\bar{p}_0$, is in the left half of this interval; time reversal handles the other half. If $\bar{p}_0+\bar{\varepsilon}_0$ is the least of the ε^{-1} possible arguments of H on the right side of (4.19) for which $\bar{p}_0+\bar{\varepsilon}_0\geqq\bar{p}_0$, we will show

$$P\{|H(p_0+\varepsilon_0)|\geqq\lambda/2\,|\,\bar{p}_0=p_0,\,\bar{\varepsilon}_0=\varepsilon_0,\,|H(p_0)|=\lambda_0>\lambda\}=\Omega(1), \qquad (4.20)$$

which yields (4.19).

Let u_i be the number of observations in $[p^{(i)}_{j_i-1},p^{(i)}_{j_i}]$ among the $(M_{i-1}+1)$-th to M_i-th observation, and let v_i be the number of these in $[p^{(r)}_{j_r-1},p^{(r)}_{j_r}]$. If $p_0=p^{(r)}_{j_r-1}$, (4.20) is trivial; so we assume $H(p^{(r)}_{j_r-1})=h_0$ (say) with $|h_0|<\lambda$. We also treat explicitly only the case $H(p_0)=\lambda_0$ of positive deviations. We obviously have $\sum_i(M_i-M_{i-1})\,[S^{\#}_i(p_0+\varepsilon_0)-S^{\#}_i(p_0)]=\sum_i\varepsilon_0\,2^{-i}\mu_i$. Since also $\lambda_0\geqq\lambda$, we obtain the desired $H(p_0+\varepsilon_0)\geqq\lambda/2$ provided that

$$\sum_i(M_i-M_{i-1})\,[S_i(p_0+\varepsilon_0)-S_i(p_0)]\geqq\sum_i\varepsilon_0\,2^{-i}\mu_i-\lambda/2. \qquad (4.21)$$

The total number of observations in $[p_0, p_{j_r}^{(r)}]$ from the $(M_\rho+1)$-th to M_{r-1}-th observation (where $\rho=\rho(p_{j_r}^{(r)})$) is $N_0=\sum_i v_i 2^{-r}(p_{j_r}^{(r)}-p_0)-(\lambda-h_0)$, and the probability that any specified one of these falls in $[p_0, p_0+\varepsilon_0]$ is $P_0=\varepsilon_0/(p_{j_r}^{(r)}-p_0)$. Thus, conditional on the v_i, $\bar{p}_0=p_0$, $\bar{\varepsilon}_0=\varepsilon_0$, and $H(p_0)=\lambda_0$, the left side of (4.21) is binomial with mean $N_0 P_0$, and $N_0 P_0$ exceeds the right side of (4.21) by

$$N_0 P_0-\sum_i \varepsilon_0 2^{-i}\mu_i+\lambda_0/2=\varepsilon_0\sum_i(2^{-r}v_i-2^{-i}\mu_i)+\left\{\frac{\lambda}{2}-\frac{(\lambda-h_0)\,\varepsilon_0}{p_{j_r}^{(r)}-p_0}\right\}. \quad (4.22)$$

Since $\lambda-h_0<2\lambda$, $\varepsilon_0\leqq 2^{-r-4}$, and $p_{j_r}^{(r)}-p_0\geqq 2^{-r-1}$, the term in braces in (4.22) is $>\lambda/4$. The unconditional probability that the sum in (4.22) (which has expectation 0) is $\geqq -1$ is easily seen to be bounded away from 0, uniformly in r and $p_{j_r}^{(r)}$, by using elementary estimates similar to those used in [12a]. (The latter shows that a binomial rv exceeds its mean by at least -1 with probability bounded away from 0.) Hence, (4.21) is established, as is thus (4.9).

Finally, $T(1,t)$ is defined arbitrarily in (1.4)–(1.5), e.g., as 0.

This completes the proof of Theorem 1.

Müller's method can also apparently be used to give a strong analogue of his weak convergence result.

The technique used in (4.18)–(4.22) to treat sums of non-iid processes has broader usage, whose statement we forego.

In view of our comment in Section 1 that (1.5) is a statement about an imperfect method rather than anything intrinsic about $\hat{\xi}$ and ξ^*, we forego any lower class considerations.

5. Breiman-Brillinger Brownian Bridge

In the second paragraph of Section 1 we mentioned Breiman's representation in terms of iidrv's Y_i with df $1-e^{(y+1)^+}$. In view of the shortcomings we have described for this representation, it will not be worth while to do more than sketch the results alluded to in [12], although some elements of the proofs may warrant mention. Also, as discussed in the next section, certain results about Breiman's scheme, used with Müller's approach, could possibly improve on some of our results; unfortunately, those results will not be found here.

Let $Z_0=0$, $Z_n=\sum_1^n Y_i$, and $Z'_n=Z_n+n$. From our remarks about $\{Z'_i/Z'_{n+1}, 1\leqq i\leqq n\}$ in Section 1, the random function

$$G_n(t)=\begin{cases} i/(n+1) & \text{if } t=Z'_i/Z'_{n+1}, \quad 0\leqq i\leqq n+1,\\ \text{linear for} & t\in[Z'_i/Z'_{n+1}, Z'_{i+1}/Z'_{n+1}], \quad 0\leqq i\leqq n,\end{cases} \quad (5.1)$$

is distributed as the continuous strictly increasing (wp 1) piecewise linear version of the sample df for uniform rv's; it will be obvious that the conclusions contained herein are true for the common discontinuous versions, but the use of the invertible G_n of (5.1) simplifies notation and arithmetic.

We hereafter write $m=n+1$.

Let $\{\xi_1(t), t \geqq 0\}$ be a standard Brownian motion defined on a probability space on which are also defined nonnegative iidrv's $\{T_i\}$ which yield a Skorohod embedding of $\{Z_n\}$ in ξ_1; that is, if $U_n = \sum_1^n T_i (U_0 = 0)$, then $\{\xi_1(U_n), n \geqq 0\}$ is distributed as $\{Z_n\}$. Here $E T_1 = E Z_1^2 = 1$. A number of different embeddings are available, as mentioned in Section 6, with various finite values of $\beta = \mathrm{var}(T_1)$ (see Section 6D). We assume such an embedding and corresponding β are chosen, and hereafter identify Z_n with $\xi_1(U_n)$. We also define Z_t to be linear in $t \in [n, n+1]$, and write $Z_t' = Z_t + t$. Also define $r(t) = Z_t - \xi_1(t)$. As usual, the development can be stated in terms of continuous t or discrete n, and we shall not distinguish (e.g., in the subscript of $U_{n\alpha}$) between a real t and its integral part. From (5.1) and the LIL for Z_m, we have, uniformly for $0 \leqq t \leqq 1$ as $m \to \infty$, wp 1,

$$\begin{aligned} m^{1/2}[G_n^{-1}(t) - t] &= m^{1/2}[Z_{mt}'/Z_m' - t] = m^{1/2}\left\{\frac{Z_{mt} - t Z_m}{Z_m + m}\right\} \\ &= m^{-1/2}(Z_{mt} - t Z_m) + O(n^{-1/2}(\log\log n)^{1/2}) \\ &= m^{-1/2}[\xi_1(m t) - t\,\xi_1(m)] + m^{-1/2}[r(m t) - t\, r(m)] \\ &\quad + O(n^{-1/2}(\log\log n)^{1/2}). \end{aligned} \tag{5.2}$$

It is known from [10] that

$$\limsup_{t \to \infty} \pm [2\beta\, t (\log t)^2 \log\log t]^{-1/4} r(t) = 1 \text{ wp } 1. \tag{5.3}$$

This would yield an estimate of the term $m^{-1/2}[r(m t) - t\, r(m)]$ in (5.2) for fixed t, but not the right constant; nor would it yield an estimate for the supremum over t of this term. However, slight modifications of the proof of (5.3) give the desired result, upon which we shall comment after a sketch of the proof. We apologize in advance for the sketch which follows, which can be made intelligent only by reading it with [10]; this results from our decision to document the assertions in as little additional space as possible[1].

Theorem 5. *For fixed* α, $0 < \alpha < 1$,

$$\begin{aligned} \limsup_{n \to \infty} \pm [r(\alpha n) - \alpha\, r(n)]/[2\beta\, n (\log n)^2 \log\log n]^{1/4} \\ = [\alpha^4 + 2\alpha^3 + \alpha]^{1/4} \text{ wp } 1. \end{aligned} \tag{5.4}$$

Moreover,

$$\limsup_{n \to \infty} \pm \sup_{0 \leqq \alpha \leqq 1} [r(\alpha n) - \alpha\, r(n)]/[2\beta\, n (\log n)^2 \log\log n]^{1/4} = 2 \text{ wp } 1. \tag{5.5}$$

Proof of (5.4). Mainly, one studies the variations in ξ_1 produced by both deviations $U_{\alpha n} - \alpha n$ and $U_n - n$, rather than by just the latter as in (5.3). (Similar considerations have just appeared in [6b]; see also Section 6A.) For the *upper class proof*, we use a LIL for linear combination of partical sums (which has an obvious extension to more than the two summands for which we state and use it):

---

[1] We simplify this discussion by using the same nongeometric n_r for upper class results, as in Theorem 1 of [10]. (In the lower class proofs here and in Theorem 2 of [10], geometric n_r are used.) However, it should be noted that geometric n_r can be used, both in [10] and the present paper, with corresponding changes in the values chosen on the bottom of p. 326 of [10] to apply Lemma 1 there. We also take this opportunity to apologize for the misprints in the Summary of [10], where β was erroneously defined as $E X_i^4$ instead of as $\mathrm{var}(T_1)$.

Lemma 7. *For real constants a_1, a_α and α, with $0<\alpha<1$,*

$$\limsup_{n\to\infty}\,[a_1(U_n-n)+a_\alpha(U_{n\alpha}-n\alpha)]/[2\beta n\log\log n]^{1/2} = [\alpha(a_1+a_\alpha)^2+(1-\alpha)a_1^2]^{1/2}\text{ wp }1. \tag{5.6}$$

This is proved in standard fashion, either by direct treatment of the rv in square brackets, or by approximating the event in question by a union of events which are quadrants in the space of $\{U_n-U_{n\alpha}, U_{n\alpha}\}$; a finite number of such events suffices by the marginal LIL's. For use in the upper class proof of Theorem 5, it is critical also to establish the analogues of the finer conditions used in (2.8) of [10].

We continue with the proof of Theorem 5. Using the $\{n_r\}$ of Theorem 1 of [10], if $U_{hn_r}-h\,n_r=u_h[2\beta n_r\log\log n_r]^{1/2}$ for $h=1,\alpha$, one computes the conditional probability of the event $\{|\xi_1(T_{hn})-\xi_1(h\,n)|>(1+8\varepsilon)\,c_h[2\beta n_r(\log n_r)^2\log\log n_r]^{1/4}$ for hn suitably close to hn_r (made precise in (18)–(21) of [10]), $h=1,\alpha\}$ to be $<\exp\{-(1+2\varepsilon)(\log r)(c_1^2/|u_1|+c_\alpha^2/|u_\alpha|)\}$. A finite number of quadrants of the form $\{(x_1,x_\alpha)\colon \pm x_1>c_1,\ \pm x_\alpha>c_\alpha\}$ with $c_1^2/|u_1|+c_\alpha^2/|u_\alpha|=1$ covers the region $\pm(x_\alpha-\alpha x_1)>(|u_\alpha|+\alpha^2|u_1|)^{1/2}(1+\varepsilon)$ in the (x_1,x_α)-plane. Using (5.6) with $a_1=\pm\alpha^2$ and $a_\alpha=\pm 1$, one obtains the upper class result. The *lower class proof* entails similar modifications of the proof of Theorem 2 of [10]; it is important to understand that one works with u_h which the U_{hn_r} oscillations exceed infinitely often, and then shows that oscillations in ξ_1 produced by these u_h occur for almost all r. One obtains $-(1-\varepsilon)[c_1^2/u_1+c_\alpha^2/u_\alpha](\log n_r)/2$ for the joint probability replacing (49); the detailed changes in definitions (38)–(39) require space which is not warranted here; one crucial change will be alluded to just below, since it reflects a difference between (5.4) and (5.5).

Proof of (5.5). This turns out to be somewhat easier. The upper class result follows at once upon noting that the function $[2\beta n(\log n)^2\log\log n]^{1/4}(1+\varepsilon)$ is in the upper class for each of $|\xi_1(U_n)-\xi_1(n)|$ and $\sup_{0<\alpha<1}|\xi_1(U_{\alpha n})-\xi_1(\alpha n)|$. The crucial feature in the lower class proof, which can only be understood by reading [10] in detail, is that when we vary α near 1 we no longer need (as one does in the lower class proof of (5.4)) the same i for the $C'_{r,i,h}$ associated with $h=1$ and α in the analogue of Q'_r of (39). The event $\cap_h\cup_i C'_{r,i,h}$ used instead yields, for $\log P\{C'_{r,i,h}\}$, twice the value obtained just above for (49) as $\alpha\to 1$, and this yields the desired result.

Remarks on Theorem 5. (1) Many applications (e.g., statistics of Kolmogorov-Smirnov type) can be phrased in terms of the deviations of G_n^{-1} rather than of G. For this purpose, Theorem 5 gives the asymptotic maximum deviation (wp 1) of $m^{1/2}[G_n^{-1}(t)-t]-m^{1/2}[\xi_1(mt)-t\,\xi_1(m)]$, for a fixed t or for the maximum over t. (2) The weak law corresponding to the above is not difficult; that corresponding to (5.8) below is harder. The comments made in connection with Theorem 4 indicate why it does not seem worthwhile to spend more space on this. (3) For the deviations of $m^{1/2}[G_n(t)-t]-m^{-1/2}[\xi_1(mt)-t\,\xi_1(m)]$, there remains to be studied the difference

$$R_n(t)=m^{1/2}[G_n(t)-t]-m^{1/2}[t-G_n^{-1}(t)], \tag{5.7}$$

which is the deviation between the sample df and sample quantile process first studied by Bahadur [1], and the exact behavior of whose oscillations was determined for fixed t in [9] and for $\sup_t \pm R_n(t)$ in [11]. The oscillations of the latter are of the same order as those of $n^{-1/2}\sup_\alpha[r(\alpha n)-\alpha r(n)]$, but the constant on the right side of (5.5) gets replaced by $2^{-1/4}$. From this, one obtains the result

$$\limsup_{n\to\infty}\sup_t |m^{1/2}[G_n(t)-t] \\ -m^{-1/2}[t\,\xi_1(n)-\xi_1(n\,t)]|/[n^{-1}(\log n)^2\log\log n]^{1/4}=\Omega(1) \text{ wp } 1. \tag{5.8}$$

The upper bound part of (5.8) was given by Brillinger. His proof is certainly succinct; however, we have given the present analysis, with a different breakup from that of [4] of the components of the difference of (5.8), because it may offer better insight as to the source of deviations. We now sketch very briefly how the somewhat complicated correct constant on the right side of (5.8) can be computed. Roughly, for the more difficult lower class proof, one notes that, with $n_r\sim\gamma^r$ and γ large as in [11], a deviation $G_{n_r}(p_0)-p_0\sim\pm c_{p_0}[2n_r^{-1}p_0(1-p_0)\log\log n_r]^{1/2}$ for infinitely many r produces a deviation of $\sup_{|p-p_0|<\varepsilon}\pm R_{n_r}(p)>(1-\varepsilon)\,c_{p_0}^{1/2}[n_r(\log n_r)^2\log\log n_r]^{1/4}$ for almost all of those r, wp 1. This and the proof of (5.5), which shows that values α near 1 are crucial for $\sup_\alpha[r(\alpha n)-\alpha r(n)]$, shows that the event

$$\{\sup_{|p-p_0|<\varepsilon}\pm R_{n_r}(p)>c_{p_0}^{1/2}[n_r(\log n_r)^2\log\log n_r]^{1/4}, \\ \sup_\alpha \pm[r(\alpha n_r)-\alpha r(n_r)]>2c_1[n_r(\log n_r)^2\log\log n_r]\}$$

occurs infinitely often wp 1 if $\exp\{-(1+\varepsilon)[c_1^2/|u_1|+c_{p_0}^2/|u_{p_0}|]\log r\}$ is not summable, where the u_h's are exceeded infinitely often as normalized deviations of the U_{hn_r}, as in the proof of Theorem 5. Maximizing $c_{p_0}^{1/2}+2c_1$ subject to $c_1^2/|u_1|+c_{p_0}^2/|u_{p_0}|=1-2\varepsilon$ involves solving a cubic, and then p_0 must finally be chosen to give the overall maximum C^*(say) of $c_{p_0}^{1/2}+2c_1$, using Lemma 7 or quadrants for the U_{hn_r}. The geometry of the quadrants for ξ_1 deviations, used as in the proof of (5.4), makes the upper class proof follow for this C^*. For fixed p the conclusion is simpler, since [9] $R_n(p)$ has smaller order than $\sup_\alpha[r(\alpha n)-\alpha r(n)]$. Thus, the upshot of this paragraph is, finally,

Theorem 6. *There is a positive constant* C^* such that

$$\limsup_{n\to\infty}\sup_t |m^{1/2}[G_n(t)-t] \\ -m^{-1/2}[t\,\xi_1(n)-\xi_1(n\,t)]|/[n^{-1}(\log n)^2\log\log n]^{1/4}=C^* \text{ wp } 1; \tag{5.9}$$

moreover,

$$\limsup_{n\to\infty}\pm\{n^{1/2}[G_n(t)-t]-n^{-1/2}[t\,\xi_1(n)-\xi_1(n\,t)]\}/[2\beta\,n(\log n)^2\log\log n]^{1/4} \\ =[t^4+2t^3+t]^{1/4} \text{ wp } 1. \tag{5.10}$$

Remark (4). For application to distribution-free functionals, the results as stated in Theorems 5 and 6 suffice. For other functionals (e.g., linear combinations of sample quantiles, as mentioned in Remark 2 to Theorem 4), corresponding results can be stated by transforming from uniform rv's s as in [9] and [11].

In the next section we shall return to Breiman's representation in conjunction with Müller's proof.

6. Other Results and Possible Directions

A. Theorem 1, or its analogue for the embedding of Lemma 2, implies that various results obtained by Strassen [21, 22] from his strong invariance principle for sums of iid random variables from corresponding Brownian motion results, have analogues for the sample df process (1.3) in terms of ξ^*. Thus, if there were a simple Motoo-type proof of the upper-lower class results for ξ^*, we would have an immediate proof of Chung's corresponding result for S_n. Strassen's elegant characterization of the closure of the functions $\{n^{-1/2}\,\xi_1(n\,t), n\in Z^+, t\in I\}$ where ξ_1 is standard Brownian motion in any number of dimensions, has no startlingly different counterpart for $\{n^{-1/2}\,\xi^*(s, n\,t)\}$: as in [21], after dividing by $(2\log\log n)^{1/2}$, one again obtains that the closure consists of integrals (now in s, t) of functions of L_2 norm $\leqq 1$, and this yields results for $\{n^{1/2}\,\hat{\xi}(s, n\,t)\}$ as mentioned at the beginning of this section. Such calculations can be viewed as extensions of those which appear in the proof of Theorem 5 and in Section 1.8 of [6b]. For fixed t, the corresponding closure and iterated logarithm results for $\{n^{1/2}\,\hat{\xi}(s, n)\}$ have recently been published by Finkelstein [6a]. Recently Wichura [22a] has obtained very general results which include Strassen-type conclusions for the sample df for vector rv's of any dimension r. Wichura's approach uses classical Kolmogorov and Hartman-Wintner approximation techniques rather than Skorohod embedding. While it does not yield explicit error estimates like ours, it appears much more expeditious for its purpose than the embedding techniques of the present paper, which can apparently be extended to higher dimensional cases (and to the independent increment cases of [22a]), but which then require even lengthier calculations than those herein. Bickel has recently applied an estimate with exponent 1/4 (on n) replaced by $1/2(r+1)$ in Theorem 5, in connection with density estimation in dimension r.

My belief is that techniques used in recent work of Nagaev will yield sharper estimates than those obtainable by embedding, especially in higher dimensional analogues of (1.6)–(1.11).

Asymptotic properties of processes with multidimensional time are surveyed in depth in two papers of Pyke [14a, 14b], which also contain new results. The idea of using embeddings like the one arising from (2.2) surely occurred to Pyke and Root and to Brillinger, and probably to others.

B. We have mentioned in Section 1 the limitations of Skorohod embeddings for the problems we consider, and we shall return to this in C below. Nevertheless, it may be worthwhile knowing how far such methods can be pushed. There is certainly no obvious invitation for improvement of the exponent 1/6 in our proof of Theorem 1 and 2. On the other hand, it is conceivable that the exponent could be improved to 1/4 by either of two attacks. Firstly, it is clear from Müller's proof ([14], p. 207) that, if only one had suitable "exponential bounds" for adding his sample df embedding error ([14], p. 199), the exponent could be increased from 1/6 to 1/4. We do not now see how to obtain such bounds; and in fact equation (4) of [14] uses a break-up in s the need for which might not be unrelated to ours.

Perhaps such bounds are more easily obtained for Breiman's representation, which, incidentally, can be used in place of Müller's in the latter's proof outlined in Section 1. Secondly, the "near miss" of (2.4) offers some hope, in that the joint distribution for large m will be very close to normal; this embedding did not evidence ξ explicitly, but it does not have the disadvantageous delicacy of balancing errors from small B_n in Lemma 1′ against those from large B_n in Lemma 4′. Incidentally, Professor Müller has given convincing heuristics against the possibility of improvement beyond $n^{-1/4}$, and in view of the comment below (1.11) this strengthens my belief, mentioned earlier, that Nagaev's approach will yield more than embedding does for weak laws.

C. We have described in Section 1 why the order of error in the results we obtain does not seem sharp. We now describe why we think this is inherent in vector Skorohod embeddings. The idea is explained in a remarkable trick of arithmetic extracted and simplified from Skorohod's proof [20], pp. 177–178. Suppose ξ_1 is a standard Brownian motion and we want to estimate $P_n = P\{\xi_1(t) < g(t),\ 0 < t < 1+\varepsilon_n\}$ in terms of $P\{\xi_1(t) < g(t),\ 0 < t < 1\}$ as $\varepsilon_n \to 0$. One possibility is to begin by estimating the difference between $\xi_1(t)$ and $\xi_1(t(1+\varepsilon_n)) = \xi_1(t')$ (say), and then estimating P_n in terms of $\xi_1(t')$, $0 < t' < 1$. Since $\xi_1(t) - \xi_1(t(1+\varepsilon_n))$ is $\Omega_p(\varepsilon_n^{1/2})$, this last is a lower limit on the result. Skorohod avoids using this. Instead, he writes

$$P_n = P\{(1+\varepsilon_n)^{-1/2}\,\xi_1(t(1+\varepsilon_n)) < (1+\varepsilon_n)^{-1/2}\,g(t(1+\varepsilon_n)),\ 0 < t < 1\},$$

and then estimates the right side of the inequality as $g(t) + O(\varepsilon_n)$, and this $O(\varepsilon_n)$ then yields (with a logarithmic term) the final error. In Skorohod's context the ε_n corresponds to our $n^{-1}(z_2 - z_1)^{-1}\,U^{(n)}_{z_1, z_2} - 1$, which is $O_p(n^{-1/2})$. But in our vector embeddings of Sections 2.2 and 2.3 we do not have a single counterpart of $(1+\varepsilon_n)$ to use in Skorohod's manner. Rather, we have a different such value for each (z_{i-1}, z_i), and must thus use the first, inferior, arithmetical scheme, based on the differences $\eta(z_{i-1}, z_i;\ (z_{i-1} - z_i)^{-1}\,U^{(n)}_{z_{i-1}, z_i}) - \eta(z_{i-1}, z_i;\ n)$, the counterpart of which is what Skorohod avoids. In the absence of the ingenuity to circumvent this difficulty, we cannot improve the exponent 1/4 of (1.7) by the Skorohod technique.

D. If one is using a Skorohod embedding, say for iidrv's, should one use that of Skorohod, Dubins, Breiman, Hall, Root, or some other? If ξ_1 is standard and T_1 is the Skorohod rv such that $\xi_1(T_1)$ has the desired distribution of a given rv X_1 with $EX_1 = 0$, $EX_1^2 = \sigma^2$, $EX_1^4 < \infty$, then all of the above methods have $ET_1 = \sigma^2$ and $\operatorname{var}(T_1) < \infty$, and in view of the results of [10] quoted in Section 5, and analogous weak laws such as [19], it seems desirable to use the method with smallest $\operatorname{var}(T_1)$. Intuitively, Root's nonrandomized stopping boundary (unique according to results of Loynes) would be guessed best, but we do not know how to prove this. However, an easy comparison of two of the methods is sometimes possible. Skorohod's method differs from Breiman's [3] except when X_1 takes on only two values. Since the former chooses at random a pair of functionally related constant stopping bounds for ξ_1, while the latter chooses them with additional randomness so that they are not functionally related, we guessed the latter was inferior. To the contrary, a simple computation shows

Theorem 7. *If X_1 has a symmetric law and $E|X_1|^r = v_r$, then*

$$ET^2_{1,\text{BREIMAN}} = \left[v_4 + \frac{4v_2 v_3}{v_1}\right] \Big/ 3 \leqq 5v_4/3 = ET^2_{1,\text{SKOROHOD}}, \tag{6.1}$$

with equality if and only if $|X_1|$ takes on only one nonzero value.

The superiority of Breiman's method is actually somewhat more general than for symmetric X_1. The supremum of the ratio of the two sides of (6.1) is of course 5. (Sawyer [19] has given bounds on $ET_{1,SK}$ in general.) Should (6.1) shake one's intuition that Root's least randomized T_1 is best?

References

1. Bahadur, R. R.: A note on quantiles in large samples. Ann. math. Statistics **37**, 577–580 (1966).
2. Breiman, L.: On the tail behavior of sums of independent random variables. Z. Wahrscheinlichkeitstheorie verw. Geb. **9**, 20–25 (1967).
3. Breiman, L.: Probability. Reading, Mass. Addison-Wesley 1968.
4. Brillinger, D. R.: An asymptotic representation of the sample df. Bull. Amer. math. Soc. **75**, 545–547 (1969).
5. Chung, K. L.: An estimate concerning the Kolmogoroff limit distribution. Trans. Amer. math. Soc. **67**, 36–50 (1949).
6. Dvoretzky, A., Kiefer, J., and Wolfowitz, J.: Asymptotic minimax character of the sample df and classical multinomial estimator. Ann. Math. Statist. **27**, 642–669 (1956).

6a. Finkelstein, H.: The law of the iterated logarithm for empirical distributions. Ann. math. Statistics **42**, 607–615 (1971).

6b. Freedman, D.: Brownian Motion and Diffusion. San Francisco: Holden-Day 1971.

7. Gnedenko, B., Korolyuk, V. S., and Skorohod, A. V.: Asymptotic expansions in probability theory, Proc. 4th Berkeley Sympos. math. Stat. and Probab. (1960), Vol. II, pp. 153–170. Berkeley: Univ. of Calif. Press 1961.
8. Kiefer, J.: On large deviations of the empiric df of vector chance variables and a law of the iterated logarithm. Pacific J. Math. **11**, 649–660 (1961).
9. Kiefer, J.: On Bahadur's representation of sample quantiles. Ann. math. Statist. **38**, 1323–1342 (1967).
10. Kiefer, J.: On the deviations in the Skorohod-Strassen approximation scheme. Z. Wahrscheinlichkeitstheorie verw. Geb. **13**, 321–332 (1969).
11. Kiefer, J.: Deviations between the sample quantile process and the sample df. Proc. First internat. Conf. Nonpar. Inf. (1969), 299–319, Cambridge Univ. Press 1970.
12. Kiefer, J.: Old and new methods for studying order statistics and sample quantiles. Proc. First theorie verw. Geb. **3**, 211–226 (1964).

12a. Kiefer, J.: Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$, Proc. 6th Berkeley Sympos. math. Statist. Probab. (1970), Vol. 1. Berkeley: Univ. of Calif. Press 1971.

13. Kiefer, J.: Skorohod embedding of multivariate processes: sums of martingales. (To appear.)
14. Müller, D. W.: On Glivenko-Cantelli convergence. Z. Wahrscheinlichkeitstheorie verw. Geb. **16**, 195–210 (1970).

14a. Pyke, R.: Empirical processes. (To appear.)

14b. Pyke, R.: Partial sums of matrix arrays and Brownian sheets. (To appear.)

15. Rosenkrantz, W. A.: On rates of convergence for the invariance principle. Trans Amer. math. Soc. **129**, 542–552 (1967).
16. Rosenkrantz, W. A.: A rate of convergence for the von Mises statistic. Trans. Amer. math. Soc. **139**, 329–337 (1970).
17. Sawyer, S.: A uniform rate of convergence for the maximum absolute value of partial sums in probability. Commun. pure appl. Math. **20**, 647–658 (1967).
18. Sawyer, S.: Uniform limit theorems for the maximum cumulative sum in probability. Trans. Amer. math. Soc. **132**, 363–367 (1968).

19. Sawyer, S.: Rates of convergence for some functionals in probability. (To appear.)
20. Skorohod. A. V.: Studies in the theory of random processes. Reading, Mass.: Addison-Wesley 1965.
21. Strassen, V.: An invariance principle for the law of the iterated logarithm. Z. Wahrscheinlichkeitstheorie verw. Geb. **3**, 211–226 (1964).
22. Strassen, V.: Almost sure behavior of sums of independent random variables and martingales. Proc. 5th Berkeley Sympos. Math. Statist. Probab. (1965), Vol. II (part 1), 315–343. Berkeley: Univ. of Calif. Press 1967.
22a. Wichura, M.: Some Strassen-type laws of the iterated logarithm for multiparameter stochastic processes with independent increments. (To appear.)

J. Kiefer
White Hall
Department of Mathematics
Cornell University
Ithaca, N.Y. 14850
USA

(Received June 1, 1971)