The Annals of Statistics
1985, Vol. 13, No. 3, 958–969

CONSISTENT ESTIMATION IN PARTIALLY OBSERVED RANDOM WALKS[1]

BY PETER GUTTORP AND ANDREW F. SIEGEL

University of Washington at Seattle

If a discrete-time random walk, consisting of the sum of independent increments from an unknown underlying distribution, is observed at every time instant then it is clear that the underlying distribution can be consistently estimated. If, however, we are restricted to observing only a subset of the times, and if this subset is too sparse, then a central limit effect takes over and only two moments can be consistently estimated. We show that divergence of the sum of the reciprocals of the observation time-intervals is a necessary and sufficient condition to permit consistent estimation of the third moment. Corresponding conditions permit consistent estimation of moments of higher order. An explicit consistent estimator for the distribution itself is presented when all moments can be consistently estimated and the distribution is determined by its moments.

1. Introduction. Let $(S_n)_{n\geq 0}$ be a random walk, i.e., $S_n = X_1 + \cdots + X_n$ where the X_i are independent and identically distributed according to a distribution function F, and $S_0 = 0$. Let the moments

$$(1.1)\qquad m = \int_{-\infty}^{\infty} x\, dF(x), \quad \text{and} \quad \mu_k = \int_{-\infty}^{\infty} (x-m)^k\, dF(x), \quad k = 2, 3, \cdots$$

all be assumed to exist and be finite. We will consider the problem of consistent estimation of the moments of F when only a partial sequence S_{N_i}, $(i = 0, 1, \cdots)$ is available where $(N_i)_{i\geq 0}$ is a strictly increasing sequence of nonnegative integers.

If the sequence N_i grows too fast, one would expect a central limit effect to take place, resulting in the ability to estimate only two moments consistently. Our purpose here is to provide precise conditions on the speed of growth of the N_i while still permitting consistent estimation of moments of higher order and, ultimately, of F itself within the class of distributions that are determined by their moments.

Consistent estimation of the first two moments is reviewed in Section 2. A necessary and sufficient condition for consistent estimation of the third moment is given in Section 3; necessity is established using a counterexample which shows that the gamma distribution cannot be distinguished from the limiting normal distribution if the observation times are too sparse. Corresponding conditions for moments of higher order are presented in Section 4. For the case

Received July 1983; revised March 1985.

[1] This work was supported in part by the Graduate School Research Fund at the University of Washington, by the National Science Foundation under Grant MCS-8302573, and by U.S.A.R.O. Contract DAAG29-82-K-0178 to Princeton University.

AMS 1980 *subject classifications.* Primary 62G99.

Key words and phrases. Sparse sampling, moment estimation, non-uniform convergence in distribution.

when the distribution is determined by its moments, and the observation times permit consistent estimation of these moments, we provide an explicit consistent estimate of the distribution F itself. The relationship to the stochastic geyser problem (Bartfai, 1966; Komlas, Major and Tusnady, 1975) is explored. In Section 5 some extensions to continuous-time processes, random stopping times, and Galton-Watson processes are considered.

2. Estimating the mean and variance. Let $Y_i = S_{N_i} - S_{N_{i-1}}$ and $n_i = N_i - N_{i-1}$ represent, respectively, the independent increments of the observed sequence and the number of terms from F summed in each increment, where we define $N_0 = 0$. In this section we establish that one can always obtain strongly consistent estimates of the first two moments, m and μ_2 in a natural way.

LEMMA 2.1. *There is a consistent estimator for m.*

PROOF. By the strong law of large numbers,

$$\tilde{m} = (1/N_i)S_{N_i} \to m \quad \text{a.s.} \quad \square \tag{2.1}$$

Without loss of generality, we may now assume that the mean m is known to be zero (cf. Section 4).

LEMMA 2.2. *There is a consistent estimator for μ_2.*

PROOF. Define

$$\tilde{\mu}_2 = (1/I) \textstyle\sum_{i=1}^{I} Y_i^2/n_i. \tag{2.2}$$

Since $E(Y_i^2/n_i) = \mu_2$, it follows that $\tilde{\mu}_2$ is an unbiased estimate of μ_2. Furthermore, $E(Y_i^4/n_i^2) = 3\sigma^4 + (\mu_4 - 3\sigma^4)/n_i$. Thus, since the Y_i are independent,

$$\operatorname{Var}(\tilde{\mu}_2) = 2\sigma^4/I + [(\mu_4 - 3\sigma^4)/I^2] \textstyle\sum_{i=1}^{I} 1/n_i = O(I^{-1}), \tag{2.3}$$

because $n_i \geq 1$ for all i. Hence $\tilde{\mu}_2$ is weakly consistent. To show the strong consistency, we use Theorem 2.7.5 of Révész (1968). We need only verify that a sum of variances is finite, which follows from

$$\textstyle\sum_{i=1}^{\infty} \operatorname{Var}(Y_i^2/n_i)i^{-2} = 2\sigma^4 \sum_{i=1}^{\infty} i^{-2} + (\mu_4 - 3\sigma^4) \sum_{i=1}^{\infty} 1/(i^2 n_i) < \infty. \quad \square \tag{2.4}$$

Without loss of generality, we may now assume that the second moment μ_2 is known to be 1.

3. Estimating the third moment. Having established in the previous section that we may assume that the mean of F is zero and the variance is 1, we will estimate the third moment μ_3 by a weighted average of Y_i^3/n_i. The weights w_i should be chosen to be inversely proportional to the variance of each term, which is

$$\operatorname{Var}(Y_i^3/n_i) = 15n_i + 15\mu_4 + 9\mu_3^2 - 45 + (\mu_6 - 15\mu_4 - 10\mu_3^2 - 30)/n_i.$$

However, since this contains unknown parameters, we will use only the leading term, setting

$$w_i = 1/(15n_i). \tag{3.1}$$

PROPOSITION 3.1. *If* $\sum_{i=1}^{\infty} 1/n_i = \infty$, *then a consistent estimator of* μ_3 *is given by*

$$\tilde{\mu}_3 = [\textstyle\sum_{i=1}^{I} w_i]^{-1} \sum_{i=1}^{I} w_i Y_i^3/n_i. \tag{3.2}$$

PROOF. Weak consistency follows because

$$\operatorname{Var}(\tilde{\mu}_3) = O([\textstyle\sum_{i=1}^{I} w_i]^{-1}) = O([\sum_{i=1}^{I} 1/n_i]^{-1}). \tag{3.3}$$

Strong consistency follows from Theorem 2.10.1 of Révész (1968) upon noting that $\sum_{i=1}^{\infty} w_i = \infty$. □

In order to establish that the growth condition $\sum_{i=1}^{\infty} 1/n_i = \infty$ is necessary, we will show that if this sum converges, then there exist two distributions with different third moments that cannot be distinguished. We will exhibit two distributions F with different third moments such that the likelihood ratio of two infinite sequences, constructed in the manner of Section 1 from these distributions, converges almost surely to a finite random variable. In other words, the probability measures corresponding to the two sequences will be mutually absolutely continuous, provided the sum of the reciprocals of the observation times converges. Consequently, if P^{∞} and Q^{∞} are measures corresponding to sequences based on distributions P and Q, and T_n is a consistent estimator sequence for Θ, we have $P^{\infty}\{\lim T_n = \Theta(P)\} = 1$. Thus if $\Theta(P) \neq \Theta(Q)$, $P^{\infty}\{\lim T_n = \Theta(Q)\} = 0$ so by absolute continuity $Q^{\infty}\{\lim T_n = \Theta(Q)\} = 0$, contradicting the consistency of T_n.

THEOREM 3.1. *A necessary and sufficient condition for consistent estimation of* μ_3 *is* $\sum_{i=1}^{\infty} 1/n_i = \infty$.

PROOF. Sufficiency was established by Proposition 3.1. For necessity, let X_1 be a random variable with a gamma distribution with mean 0, variance 1 and skewness $\alpha^{-1/2}$; and let X_2 be a random variable with the standard normal distribution. Construct sequences $Y_{n,1}$ and $Y_{n,2}$ from independent and identically distributed observations of X_1 and X_2 in the manner of Section 1. The likelihood ratio of $(Y_{i,1})_{i=1}^{I}$ to $(Y_{i,2})_{i=1}^{I}$ is

$$\Lambda_I = \prod_{i=1}^{I} \frac{(2\pi n_i)^{1/2}}{\Gamma(n_i\alpha)} \alpha^{n_i\alpha/2}(y_i + n_i\alpha^{1/2})^{n_i\alpha-1}\exp\left\{-(y_i + n_i\alpha^{1/2})\alpha^{1/2} + \frac{y_i^2}{2n_i}\right\}. \tag{3.4}$$

Taking logarithms, using Stirling's formula and a McLaurin expansion of $\log(1 + x)$ we obtain

$$L_I = \log(\Lambda_I) = \sum_{i=1}^{I}\left[-\frac{y_i}{n_i\alpha^{1/2}} + \frac{y_i^3}{3n_i^2\alpha^{1/2}} + O\left(\frac{y_i^2}{n_i^2}\right)\right]. \tag{3.5}$$

Next, define

(3.6) $$M_{I,j} = \sum_{i=1}^{I} Y_{i,j}/n_i.$$

These are martingales, and satisfy

(3.7) $$E\{M_{I,j}^2\} = \sum_{i=1}^{I} 1/n_i.$$

Hence, if $\sum_{i=1}^{\infty} 1/n_i < \infty$, then the $M_{I,j}$ are L_2-bounded martingales and therefore converge almost surely (Feller, 1966, Theorem VII.8.1). We next show that $S_{I,j} = \sum_{i=1}^{I} Y_{i,j}^3/n_i^2$ is bounded in probability. Straightforward algebra shows that

(3.8) $$E(Y_{i,j}^6/n_i^4) = O(n_i^{-1}),$$

and

(3.9) $$E\left(\frac{Y_{i,j}^3}{n_i^2}\right) = \begin{cases} O(n_i^{-1}) & \text{if } j = 1 \\ 0 & \text{if } j = 2. \end{cases}$$

Let C be arbitrary. Then $P\{|S_{I,j}| > C\} \le \sup_i \mathrm{Var}(S_{i,j}/C^2)$, which can be made arbitrarily small. Similar computations show that the error term in (3.5) goes in probability to zero under both measures. Thus the likelihood ratio is bounded in probability. It follows that the process constructed from the gamma variables is absolutely continuous with respect to that constructed from the normal variables, and hence μ_3 cannot be consistently estimated. □

4. Estimating the distribution function. If the n_i grow slowly enough, we can estimate higher order moments μ_k consistently. We no longer assume that m and μ_2 are known. The proof of the following proposition also verifies that this assumption entailed no loss of generality.

PROPOSITION 4.1. *If $\sum_i n_i^{2-k} = \infty$, then we can compute consistent estimates of $m, \mu_2, \cdots, \mu_k$.*

PROOF. First notice that for $k \ge 2$,

(4.1) $$E[(Y_i - n_i m)^k] = n_i \mu_k + \phi(n_i; \mu_2, \cdots, \mu_{k-2}),$$

where

$$\phi(n_i; \mu_2, \cdots, \mu_{k-2}) = \sum \binom{k}{2 \cdots 2, \cdots, (k-2) \cdots (k-2)} \binom{n_i}{\alpha_2, \alpha_3, \cdots, \alpha_{k-2}} \mu_2^{\alpha_2} \mu_3^{\alpha_3}, \cdots, \mu_{k-2}^{\alpha_{k-2}},$$

where there are α_2 "2"s, α_3 "3"s, $\cdots$, and α_{k-2} "$k-2$"s in the first multinomial coefficient on the right-hand side, and the sum ranges over all $(\alpha_2, \cdots, \alpha_{k-2})$ with nonnegative integer components such that $2\alpha_2 + 3\alpha_3 + \cdots + (k-2)\alpha_{k-2} = k$. A related formula can be found in Kendall and Stuart (1977, Volume I, page 70).

From (4.1) it appears reasonable to use (changing the notation somewhat from the previous sections) the estimate

$$\tilde{\mu}_k = \tilde{\mu}_{k,I} = [\sum_{i=1}^I w_i]^{-1} \sum_{i=1}^I (w_i/n_i)[(Y_i - n_i\tilde{m})^k - \phi(n_i; \tilde{\mu}_2, \cdots, \tilde{\mu}_{k-2})], \tag{4.2}$$

where the weights w_i ($= w_{i,k}$) are chosen inversely proportional to the leading term of the variance of $(Y_i - n_i m)^k/n_i$, so that $w_i = n_i^{2-k}$, as can be seen by using (4.1) to find this higher moment. Since the estimate (4.2) is recursively defined, we will verify its consistency by induction. First let $k = 2$, so that

$$\begin{aligned}\tilde{\mu}_2 &= \frac{1}{I} \sum_{i=1}^{i} \frac{1}{n_i} (Y_i - n_i\tilde{m})^2 \\ &= \frac{1}{I} \sum_{i=1}^I \frac{(Y_i - n_i m)^2}{n_i} + 2\,\frac{m - \tilde{m}}{I} \sum_{i=1}^I (Y_i - n_i m) + (m - \tilde{m})^2 \frac{\sum n_i}{I} \\ &= A + B + C.\end{aligned}$$

The first term A has mean μ_2 and variance μ_2/I, so by Chebyshev's inequality $P\{|A - \mu_2| > \varepsilon\} \le \mu_2/(I\varepsilon^2)$. To deal with the second term, B, estimate

$$\begin{aligned}&P\{|(m - \tilde{m})(1/I) \textstyle\sum_{i=1}^I (Y_i - n_i m)| > \varepsilon\} \\ &\quad \le P\{|m - \tilde{m}| > (I/\textstyle\sum n_i)^{1/2}\} + P\{|(1/I)\textstyle\sum_{i=1}^I (Y_i - n_i m)| > \varepsilon(\textstyle\sum n_i/I)^{1/2}\} \\ &\quad \le C'/I.\end{aligned}$$

Finally, for the third term, C, we estimate

$$P\{(\textstyle\sum n_i/I)(m - \tilde{m})^2 > \varepsilon\} \le 1/(\varepsilon I)$$

using Markov's inequality. Thus, for some constant C'',

$$P\{|\tilde{\mu}_2 - \mu_2| > \varepsilon\} \le C''/I.$$

Now for general k, we make the induction hypothesis that

$$P\{|\tilde{\mu}_j - \mu_j| > \varepsilon\} \le C_j/\textstyle\sum_{i=1}^I n_i^{2-j}, \quad j = 2, \cdots, k-1,$$

for some constants C_j. Then we can break $\tilde{\mu}_k - \mu_k$ up into three parts:

$$\tilde{\mu}_k - \mu_k = A + B + C,$$

where

$$\begin{aligned}A &= (-\mu_k + \textstyle\sum_{i=1}^I (w_i/n_i)\{(Y_i - n_i m)^k - \phi(n_i; \mu_2, \cdots, \mu_{k-2})\})/\textstyle\sum_{i=1}^I w_i, \\ B &= (\textstyle\sum_{i=1}^I (w_i/n_i)\{\phi(n_i; \mu_2, \cdots, \mu_{k-2}) - \phi(n_i; \tilde{\mu}_2, \cdots, \tilde{\mu}_{k-2})\})/\textstyle\sum_{i=1}^I w_i, \\ C &= (\textstyle\sum_{j=1}^k \sum_{i=1}^I \binom{k}{j} w_i n_i^{j-1}(m - \tilde{m})^j (Y_i - n_i m)^{k-j})/\textstyle\sum_{i=1}^I w_i.\end{aligned}$$

Term A has zero mean and has variance $O[(\sum_{i=1}^I w_i)^{-1}]$ by construction. Term B is the sum of a bounded number of terms, a typical term being proportional to

$$T_\alpha = (\mu_2^{\alpha_2} \cdot \cdots \cdot \mu_{k-2}^{\alpha_{k-2}} - \tilde{\mu}_2^{\alpha_2} \cdot \cdots \cdot \tilde{\mu}_{k-2}^{\alpha_{k-2}}) \sum_{i=1}^I \frac{w_i}{n_i} \binom{n_i}{\alpha_2 \cdots \alpha_{k-2}}.$$

By a Taylor expansion, we see that

$$\mu_2^{\alpha_2} \cdot \cdots \cdot \mu_{k-2}^{\alpha_{k-2}} - \tilde{\mu}_2^{\alpha_2} \cdot \cdots \cdot \tilde{\mu}_{k-2}^{\alpha_{k-2}} = \sum_{\tau=2}^{k-2} \alpha_\tau(\mu_\tau - \tilde{\mu}_\tau)(\partial/\partial\mu_\tau) \prod_{\eta=2}^{k-2} \mu_\eta^{\alpha_\eta} \,|_{\mu_\tau=\mu_\tau^*}$$

for some $\mu_\tau^* = \mu_\tau + \delta(\tilde{\mu}_\tau - \mu_\tau)$, $0 \le \delta \le 1$. Let $f_\tau(\mu) = (\partial/\partial\mu_\tau) \prod_{\eta=2}^{k-2} \mu_\eta^{\alpha_\eta}$, and denote by μ^* the vector μ with μ_τ replaced by μ_τ^*. Write

$$\begin{aligned} T_\alpha &= \sum_\tau \alpha_\tau(\mu_\tau - \tilde{\mu}_\tau) f_\tau(\mu) \sum_i \frac{w_i}{n_i} \binom{n_i}{\alpha_2 \cdots \alpha_{k-2}} \Big/ \sum_i w_i \\ &\quad + \sum_\tau \alpha_\tau(\mu_\tau - \tilde{\mu}_\tau)[f_\tau(\mu^*) - f_\tau(\mu)] \sum_i \frac{w_i}{n_i} \binom{n_i}{\alpha_2 \cdots \alpha_{k-2}} \Big/ \sum_i w_i \\ &= T_\alpha^0 + R_\alpha. \end{aligned}$$

For a typical term in T_α^0, we estimate

$$\begin{aligned} P\left\{ \left| \alpha_\tau(\mu_\tau - \tilde{\mu}_\tau) f_\tau(\mu) \sum_i \frac{w_i}{n_i} \binom{n_i}{\alpha_2 \cdots \alpha_{k-2}} \Big/ \sum_i w_i \right| > \varepsilon \right\} \\ \le C_\tau \frac{[\sum_{i=1}^I n_i^{1-k+\alpha_2+\cdots+\alpha_{k-2}}]^2}{[\sum_{i=1}^I n_i^{2-\tau}][\sum_{i=1}^I n_i^{2-k}]^2} = O[[\textstyle\sum_{i=1}^I n_i^{2-k}]^{-1}] \end{aligned}$$

using the Cauchy-Schwarz inequality and properties of the α_j. Similar estimates yield the same order bound on R_α. Finally, the typical term in C is handled much as in arguing the case $k = 2$. It now follows that

$$P\{|\tilde{\mu}_k - \mu_k| > \varepsilon\} \le C_k / \textstyle\sum_{i=1}^I n_i^{2-k}$$

from which weak consistency follows. In order to obtain strongly consistent estimates, we may use a subsequence of the $\tilde{\mu}_{k,I}$. Choose I_ξ such that $\sum_{i=1}^{I_\xi} n_i^{2-k} \ge \xi^{3/2}$. Then, by Borel-Cantelli, since $n_i^{2-j} > n_i^{2-k}$,

$$\tilde{\mu}_{j,I_\xi} \to \mu_j, \quad \text{a.s.} \quad j = 2, \cdots, k. \quad \square$$

REMARKS.

(1) Whenever $n_i \le ci^\gamma$ for some $\gamma \le 1/(k-2)$ one can verify that $(\tilde{\mu}_k)$ are strongly consistent estimates, without needing to take subsequences.

(2) Term A in the proof always converges strongly to 0, by Révész's Theorem 2.10.1. One can construct sequences (n_i) such that term C does not converge a.s. Thus one cannot, in general, avoid taking subsequences of the $(\tilde{\mu}_k)$ to get strongly consistent estimates.

From Remark 1 above, we see that if $n_i = o(i^\alpha)$ for all $\alpha > 0$, then we can consistently estimate all moments. If the distribution is determined by its moments, it should follow that we can consistently estimate the distribution itself. Below we present an explicit estimate of the distribution function and show its consistency. The idea is to construct a distribution having, as much as possible, the same moments as we have estimated. To implement this idea, we need some conditions on when a sequence of numbers are the moments of some distribution function. This is the Hamburger moment problem and is discussed in Shohat and Tamarkin (1943). Let $(\mu_k)_0^\infty$ be a sequence of numbers, and let

$$\Delta_n = \Delta_n(\mu) = \det(\mu_{i+j})_{i,j=0}^n, \quad n = 1, 2, \cdots.$$

Then $\mu_k \int t^k \, dF(t)$ for some cumulative distribution function F if and only if $\Delta_n \geq 0$ for all positive integer values of n. If for some value of N we have $\Delta_{N-1} > 0$ but $\Delta_N = 0$, then it follows that $\Delta_{N+1} = \Delta_{N+2} = \cdots = 0$, and F has a distribution with exactly N points of support. An algorithm for the computational solution of a given moment problem was given by Mammana (1954). A solution at the nth stage (when $\Delta_1, \cdots, \Delta_{n-1} > 0$) has support $\{x_1, \cdots, x_n\}$ given by the roots of

$$\begin{vmatrix} 1 & x & x^2 & \cdots & x^n \\ \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{n-1} & \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} \end{vmatrix} = 0$$

with probabilities $p_1, \cdots, p_n$ given by the solution to

$$\sum p_i x_i^k = \mu_k, \quad k = 0, \cdots, n-1.$$

For a sequence of empirical moments from a finite set of data, this algorithm will produce the empirical distribution function.

Based on I observations $Y_1, \cdots, Y_I$ with moment estimates $\tilde{\mu}_{i,I}$, $i = 1, \cdots, I$, we now compute an estimate F_I of the distribution function. Let $k = [I/2]$ and set $\tilde{\mu}_{0,I} = 1$. Compute $\Delta_j(\tilde{\mu}_I)$ for $j = 1, \cdots, k-1$. Let $\tilde{k}_I$ be the smallest value of j for which $\Delta_j(\tilde{\mu}_I) \leq 0$, or else set $\tilde{k}_I = k$ if no such j exists. Solve the moment problem for a distribution F_I with $\tilde{k}_I$ points of support $\{x_1, \cdots, x_{\tilde{k}_I}\}$ and probabilities $\{p_1, \cdots, p_{\tilde{k}_I}\}$, so that the first $2\tilde{k}_i - 1$ moments of F_I are the estimates $\tilde{\mu}_{1,I}, \cdots, \tilde{\mu}_{2\tilde{k}_I-1,I}$. We will now prove a technical fact about weak convergence to a distribution with finite support.

LEMMA 4.1. *Let Z be a discrete random variable with j points of support. Let Z_n be a sequence of random variables whose first $2j$ moments converge to those of Z, that is*

$$E(Z_n^k) \to E(Z^k) \quad k = 0, 1, \cdots, 2j.$$

Then Z_n converges weakly to Z.

PROOF. Let x^* be any fixed point distinct from the support points $x_1, \cdots, x_j$ of Z. Using the arguments in the proof of the Tchebycheff inequalities (Shohat and Tamarkin, 1943, page 43), there exists a polynomial $g(x) = \sum_{i=0}^{2j} a_i x^i$ of degree $2j$ such that

$$g(x_i) = \begin{cases} 1 & \text{if} \quad x_i < x^* \\ 0 & \text{if} \quad x_i > x^*, \end{cases}$$

$$g(x^*) \geq 1,$$

$$g'(x_i) = 0 \qquad \text{for} \quad i = 1, \cdots, j,$$

and

$$g(x) \geq \mathbf{1}(x \leq x^*) \quad \text{for all} \quad x,$$

where $\mathbf{1}$ is used to denote the indicator function. Figure 4.1 illustrates the polynomial g in the case $x^* = 4$, with support points 2, 3, 5 and 7, together with the auxiliary lower bound polynomial h to be defined soon. Note that the cumulative distribution G of Z satisfies

$$G(x^*) = P(Z \leq x^*) = E[\mathbf{1}(Z \leq x^*)] = E[g(Z)],$$

and the distribution function G_n of Z_n satisfies

$$G_n(x^*) = E[\mathbf{1}(Z_n \leq x^*)] \leq E[g(Z_n)].$$

Because the above right-hand expression depends only on the first $2j$ moments of Z_n, as n tends to infinity we have

$$E[g(Z_n)] = \textstyle\sum_{i=0}^{2j} a_i E(Z_n^i) \to \sum_{i=0}^{2j} a_i E(Z^i) = E[g(Z)] = G(x^*),$$

and therefore

$$\limsup_{n\to\infty} G_n(x^*) \leq G(x^*).$$

By similar arguments there also exists a polynomial h of degree $2j$ with $h(x_i) = \mathbf{1}(x_i < x^*)$, $h(x^*) = 0$, $h'(x_i) = 0$, and such that $h(x) \leq \mathbf{1}(x \leq x^*)$ for all x. Proceeding as before, we find that

$$\liminf_{n\to\infty} G_n(x^*) \geq G(x^*).$$

It now follows that G_n tends to G at all x^* distinct from the support points $x_1, \cdots, x_j$ of G, establishing weak convergence. □

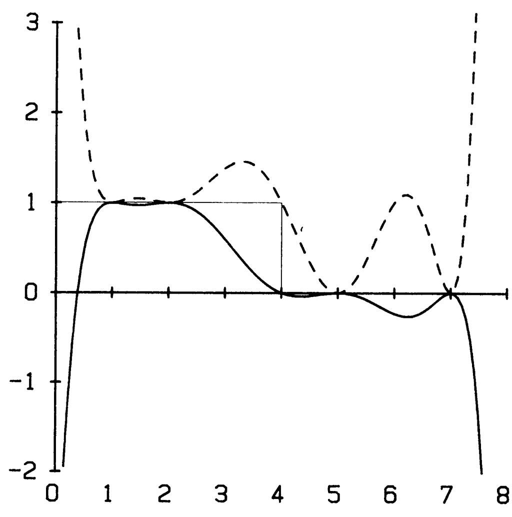

FIG. 4.1. *The polynomials g (dashed curve) and h (solid curve) used to bound the cumulative distributions at $x^* = 4$ above and below, for a distribution with support points at 1, 2, 5 and 7.*

REMARK. Lemma 4.1 exhibits a sufficient condition for weak convergence to a distribution with finite support. It is quite general in that it does not require that the sequence be discrete and does not even require the existence of higher order moments.

THEOREM 4.1. *Suppose that $n_i = o(i^\alpha)$ for all $\alpha > 0$, and that F is determined by its moments. If x is a continuity point of F then*

$$F_I(x) \to F(x) \quad \text{a.s.}$$

PROOF. Fix some j, and let $\mu_F(j, I)$ be the jth moment of F_I and let $I > 2j$. Assume first that F has infinite support. Since $\Delta_j > 0$ and $\Delta_j(\tilde{\mu}_I)$ is a continuous function of $\tilde{\mu}_{1,I}, \cdots, \tilde{\mu}_{2j-1,I}$, we will have $\Delta_j(\tilde{\mu}_I) > 0$ a.s. for I large enough. Hence F_I will then have jth moment

$$\tilde{\mu}_{j,I} \to \mu_j = \int t^j \, dF(t) \quad \text{a.s.}$$

It follows from Moran (1968, Theorem 6.16) that F_I converges weakly to F, almost surely.

Now suppose F has finite support of size k_F. By Lemma 4.1 we need only show that the first $2k_F$ moments of F_N converge almost surely to the corresponding moments of F. For N sufficiently large, by continuity of the determinant, we will have $\Delta_j(\tilde{\mu}_N) > 0$ a.s. for $j = 1, \cdots, k_F - 1$, which implies that $\tilde{k}_N \geq k_F$ (a.s.). This guarantees that the first $2k_F - 1$ moments of F_N will (a.s.) match the estimated moments of F, and will thus converge almost surely to the moments of F. It now remains only to show that the moment of order $2k_F$ also converges. Partition the sequence F_N into two subsequences according to whether $\tilde{k}_N > k_F$ or $\tilde{k}_N = k_F$. When $\tilde{k}_N > k_F$, by construction the moment of F_N of order $2k_F$ will match the estimated moment of F, and will therefore converge almost surely. When $\tilde{k}_N = k_F$ the distribution F_N will have support at exactly k_F points; we may solve for the moment of order $2k_F$ by expanding the determinant $\Delta_{k_F}[\mu(F_N)]$ along its last row and solving for $\mu_{2k_F}(F_N)$. Taking the limit, we see that

$$\mu_{2k_F}(F_N) = \frac{\xi(\tilde{\mu}_{1,N}, \cdots, \tilde{\mu}_{2k_F-1,N})}{\Delta_{k_F-1}(\tilde{\mu}_N)}$$

$$\to \frac{\xi(\mu_1(F), \cdots, \mu_{2k_F-1}(F))}{\Delta_{k_F-1}(\mu(F))} = \mu_{2k}(F) \quad \text{a.s.},$$

where ξ is a polynomial in its arguments, representing the rest of the cofactor expansion of the determinant. Convergence follows by continuity of ξ and Δ and because the denominator of the limit is strictly positive (which follows from F being supported on k_F points). Thus the first $2k_F$ moments of F_N converge almost surely to those of F, and Lemma 4.1 completes the proof. □

REMARKS. (1) It is interesting to note that the convergence is not uniform, except for continuous F. A simple example is provided by taking F to be the Bernoulli distribution with success probability ½, and $n_i = 2$. Some straightforward algebra shows that if $f_1 = ½ + \varepsilon_{1,n}$ and $f_2 = ¼ + \varepsilon_{2,n}$ are the observed frequencies of ones and twos, respectively, we will estimate the support $\tilde{k}$ to be 2 whenever n is large and $\varepsilon_{1,n} < 0$ (by the law of the iterated logarithm this happens infinitely often a.s.), and the coefficient of the constant term in the equation determining the points of support will be nonzero under the same conditions. Thus the support of $\tilde{F}$ will not contain zero for such sample paths, and the convergence cannot be uniform. In fact, the resulting distribution $\tilde{F}$ could not have produced the observed data.

(2) The stochastic geyser problem (see, e.g., Bartfai, 1966) can be applied to this problem. Let $r_n = S_n + \eta_n$, where S_n is a random walk with distribution F, determined by its moments, and η_n is some random noise, independent of the future. Then Bartfai proved that if lim sup$[\eta_n/\log(n)] = 0$ a.s., then F can be estimated consistently from the data although no explicit estimate of F was provided. In our case we would choose

$$\eta_n = -\textstyle\sum_{N_{k(n)}}^{n-1} X_i,$$

where $k(n)$ is the index of the most recent exact observation, so that $N_{k(n)} = \max\{j\colon N_j \le n\}$, and the sum is defined to be zero whenever the lower limit exceeds the upper limit. The sequence (r_n) then looks like n_1 replications of zero, n_2 replications of S_{N_1}, and so on. By the law of the iterated logarithm, one can show that the rate given in Bartfai's theorem translates into $n_k = o[\log^2(k)/\log\log\log(k)]$ for large k. Komlos, Major and Tusnady (1975) show that this rate cannot be improved in the setting of the stochastic geyser problem. Our method gives a stronger result, in a much less general setting.

5. Some extenstions. There is a continuous-time analogue to this problem. Let $(X_t)_{t\ge 0}$ be a process with stationary and independent increments, with $X_0 = 0$. Let $T_1, T_2, \cdots$ be an increasing sequence of fixed times with $T_n \to \infty$, and let $Y_i = X_{T_i} - X_{T_{i-1}}$. Then $Y_i =_d F^{*(t_i/t_1)}$ where $X_1 \sim F$ and $t_i = T_i - T_{i-1}$, and $T_0 = 0$. The Y_i are independent and F is infinitely divisible. As before, $\lim X_{T_i}/T_i = E(X_1)$ a.s. To estimate $\mu_2 = \mathrm{Var}(X_1)$, we need a weighted average of the Y_i^2/t_i. Since $\mathrm{Var}(Y_i^2/t_i) = 2\mu_2^2 + (\mu_4 - 3\mu_2^2)/t_i$ contains unknown parameters, we use weights proportional to $t_i/(1 + t_i)$. This behaves in the same fashion as the optimal weights: observations based on long stretches of time are given high weights, whereas those with only a short observation time are down-weighted. The estimate is

$$\tilde{\mu}_2 = \left(\sum Y_i^2/(1 + t_i)\right)/\left(\sum t_i/(1 + t_i)\right).$$

This estimate is unbiased and has variance

$$\frac{2\mu_2^2}{\sum t_i/(1 + t_i)} + \frac{(\mu_4 - 3\mu_2^2)\sum t_i/(1 + t_i)^2}{[\sum t_i/(1 + t_i)]^2}.$$

Since $\sum t_i/(1 + t_i) = \infty$ if and only if $\sum t_i = \infty$, and

$$\frac{\sum t_i/(1 + t_i)^2}{[\sum t_i/(1 + t_i)]^2} \le \frac{1}{\sum t_i/(1 + t_i)},$$

we see that $\mathrm{Var}(\tilde{\mu}_2) \to 0$ if and only if $\sum t_i \to \infty$, which is equivalent to $T_n \to \infty$. Note that this estimate is consistent when $t_i = 1/i$ while the estimate based on equal weights, used in Section 2, is not.

By Liapounov's theorem,

$$(\mu_2 \sum t_i)^{-1/2} \sum Y_i \to_{\mathrm{d}} N(0, 1)$$

provided $\sum t_i \to \infty$. A consistent estimate of μ_3 is in general only possible if $\sum 1/t_i = \infty$. The construction in Section 3 is easily adapted to this case.

A perhaps more interesting generalization is to the case of random stopping times. A simple example would be independent and identically distributed time intervals n_i independent of the random walk. By Kolmogorov's three series theorem $\sum 1/n_i = \infty$ a.s., since $\sum \Pr\{n_i \le c\} = \infty$ for any c such that $\Pr\{n_1 \le c\} > 0$. Conditional on the sequence $(n_i)_{i \ge 1}$, the argument of Section 3 shows that third moments are estimable in this case.

A more important class of random observation time intervals are times that are determined by the history of the random walk. For example, if $N_n = \sum_{k=0}^{n} Z_k$, where $Z_{k+1} = S_{N_k} - S_{N_{k-1}}$, we have a Galton-Watson branching process. The problem of estimating the offspring distribution of the Galton-Watson process from observing only generation sizes was one of the starting points of this investigation. Lockhart (1981) established that at most three moments of the offspring distribution are estimable, by bounding the variational distance between the measures corresponding to two Galton-Watson processes with the same first three moments and the same lattice. As a lemma, Lockhart showed that a supercritical explosive branching process almost surely grows at least as fast as a partially observed random walk with $n_i = \Theta^i$, where $1 < \Theta < M$ and M is the mean of the offspring distribution. Since $\sum 1/\Theta^i < \infty$, our results would indicate that, indeed, it is impossible to estimate the third moment as well. Subsequently, Lockhart (1982) established this result using different methods.

In continuous time, the analog of the Galton-Watson process (when viewed as a partially observed random walk) is a continuous state space branching process (Kallenberg, 1979). Here the underlying process is a subordinator, i.e., an increasing process with stationary, independent increments and paths that are right-continuous with left limits. The possibility of estimating functionals of such processes may well be worth pursuing.

Acknowledgements. We are grateful to R. Pyke, R. Lockhart, C. Hurvich, and the referees for helpful comments and suggestions.

REFERENCES

BARTFAI, P. (1966). Die Bestimmung der zu einen wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. *Studia Sci. Math. Hungar.* **1** 161–168.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*, 2. Wiley, New York.

KALLENBERG, P. J. M. (1979). Branching processes with continuous state space. *Math. Centre Tracts* **117** Amsterdam.

KENDALL, M. and STUART, A. (1977). *The Advanced Theory of Statistics*, 1, 4th ed. Griffin, London.

KOMLOS, J., MAJOR, P. and TUSNADY, G. (1975). An approximation of partial sums of independent random variables and the sample distribution function. *Z. Warsch. verw. Gebiete* **32** 111–131.

LOCKHART, R. A. (1981). On the nonexistence of consistent estimates in branching processes. Technical report, Simon Fraser University.

LOCKHART, R. A. (1982). On the nonexistence of consistent estimates in Galton-Watson processes. *J. Appl. Probab.* **19,** 842–846.

MAMMANA, C. (1954). Sul problema algebraico dei momenti. *Ann. Scuolo Norm. Sup. Pisa* **8** 133–140.

MORAN, P. A. P. (1968). *An Introduction to Probability Theory.* Clarendon Press, Oxford.

RÉVÉSZ, P. (1968). *The Laws of Large Numbers.* Academic, New York.

SHOHAT, J. A. and TAMARKIN, J. D. (1943). *The Problem of Moments.* Amer. Math. Soc. Waverly Press, Baltimore.

DEPARTMENT OF STATISTICS
GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTMENTS OF STATISTICS,
MANAGEMENT SCIENCE, AND FINANCE
GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195