The Annals of Probability
1991, Vol. 19, No. 4, 1756–1767

STRONG LIMIT THEOREMS OF EMPIRICAL DISTRIBUTIONS FOR LARGE SEGMENTAL EXCEEDANCES OF PARTIAL SUMS OF MARKOV VARIABLES

BY AMIR DEMBO AND SAMUEL KARLIN[1]

Stanford University

Let $A_1, A_2, \dots, A_n$ be generated governed by an r-state irreducible Markov chain and suppose (X_i, U_i) are real valued independently distributed given the sequence $A_1, A_2, \dots, A_n$, where the joint distribution of (X_i, U_i) depends only on the values of A_{i-1} and A_i and is of bounded support.

Where A_0 is started with its stationary distribution, $E[X_1] < 0$ and the existence of a finite cycle $C = \{A_0 = i_0, \dots, A_k = i_k = i_0\}$ such that $\Pr\{\sum_{i=1}^m X_i > 0,\ m = 1, \dots, k;\ C\} > 0$ is assumed. For the partial sum realizations where $\sum_{i=k}^l X_i \to \infty$, strong laws are derived for the sums $\sum_{i=k}^l U_i$. Examples with $r = 2$, $X \in \{-1, 1\}$ and the cases of Brownian motion and Poisson process with negative drift are worked out.

1. Introduction. In the accompanying paper [Dembo and Karlin (1991)], we characterized the composition of high scoring segments among partial sums of i.i.d. random variables. For biological motivations and applications, see the introduction and Section 4 of Dembo and Karlin (1991), Karlin and Altschul (1990) and Karlin, Dembo and Kawabata (1990). In the present paper the sequence consists of letters $A_1, A_2, \dots, A_n$ assuming values from a finite alphabet $\{a_i\}_1^r$ generated under Markov dependence. Scores (reflecting on letter attributes) are associated with these letters and can depend on diletter occurrences such that $X_i = s_{\alpha\beta}$ for $A_{i-1} = a_\alpha$ and $A_i = a_\beta$. Thus $\{A_i, X_i\}$ are jointly Markov of order 1. Let $S_{n\alpha} = \sum_{i=1}^n X_i$ be the partial sum process of scores induced by the letter sequence with initial letter $A_0 = \alpha$. A limit distribution for the maximal segment score $M_\alpha(n) = \max_{0 \le k \le l \le n}[S_{l\alpha} - S_{k\alpha}]$ is described in Karlin and Dembo (1992).

The focus of this paper concerns the empirical distribution function of the random variables $U_1, U_2, \dots, U_n, \dots$ (U_i defined with respect to X_i) during a high scoring segment of $\{S_{m\alpha}\}$. (We suppress the designation of the initial state if no ambiguity is likely.) For example, where $U = 1$ iff $X = s_{ij}$, the partial sum $\sum_{k=1}^m U_k$ effectively assesses the aggregate count of a specific score count over the letter segment $(A_0, A_1, \dots, A_m)$.

Formally, the model is as follows: Let the letter sequence $A_1, A_2, \dots, A_n, \dots$ generated be governed by an r-state irreducible Markov chain with transition probabilities p_{ij}, and suppose (X_m, U_m) are independently distributed given

Received December 1989; revised November 1990.
[1]Research supported in part by NIH Grants GM 39907-02, GM 10452-27 and NSF Grant DMS-86-06244.
AMS 1980 *subject classifications*. Primary 60F10, 60F15, 60J10, 60K15.
Key words and phrases. Strong laws, Markov additive processes, large segmental sums.

the sequence $A_1, A_2, \dots, A_n, \dots,$ where the joint distribution of (X_m, U_m) depends only on the values of A_{m-1} and A_m and is of bounded support.

Where A_0 is started with its stationary frequencies, $E[X_1] < 0$ is assumed, so $\{S_m\}$ entails a negative drift. For $\{S_m\}$ to have positive probability of achieving early any positive score, the existence of a finite cycle of letters $\mathscr{C} = \{A_0 = i_0, \dots, A_k = i_k = i_0\}$ such that $\Pr\{\min_{m=1,\dots,k} \sum_{i=1}^m X_i > 0; \mathscr{C}\} > 0$ is assumed. The quantity

$$M_\alpha(n) = \sup_{0 \le k \le l \le n} (S_l - S_k) \tag{1}$$

corresponds to a segment of the sequence $\{S_m\}_0^n$ with maximal score starting in state $A_0 = \alpha$. For the sample path $(A_1, X_1), (A_2, X_2), \dots$ with some prescribed initial state we define sequentially the stopping times

(2)
$$K_0 = 0, \qquad K_\nu = \min\{k \ge K_{\nu-1} + 1,\, S_k - S_{K_{\nu-1}} \le 0\}, \qquad \nu = 1, 2 \dots,$$

which are finite valued by virtue of the negative drift of $\{S_m\}$. For any fixed i, $K_\nu - K_{\nu-1}$, conditioned on $A_{K_{\nu-1}} = i$, $\nu = 1, 2, \dots,$ are i.i.d. integer valued random variables. Their distribution function has tails of exponential decay as

$$\Pr\{K_\nu - K_{\nu-1} > l | A_{K_{\nu-1}} = i\} \le \Pr\{S_l > 0 | A_0 = i\}.$$

The time frame $K_{\nu-1} + 1$ to K_ν encompasses the νth excursion epoch, that is, the νth segment of the process $\{S_m\}$ starting from zero until hitting a nonpositive value. For each $y > 0$ and within the νth excursion epoch, define the stopping time

$$\begin{aligned} T_\nu(y) = \min\{m\colon m > K_{\nu-1} \text{ and either } & S_m - S_{K_{\nu-1}} \le 0 \\ \text{or } & S_m - S_{K_{\nu-1}} \ge y\}. \end{aligned} \tag{3}$$

Note that the distribution of $T_\nu(y) - K_{\nu-1}$ depends upon the state of $A_{K_{\nu-1}}$. The quantities

$$L_\nu(y) = T_\nu(y) - K_{\nu-1} \quad \text{and} \quad W_\nu(y) = \sum_{m=K_{\nu-1}+1}^{T_\nu(y)} U_m \tag{4}$$

assess the elapsed time and associated sum of U's, respectively, until the first departure of $\{S_m - S_{K_{\nu-1}}\}$ from the open interval $(0, y)$.

The realizations in (3) are of two kinds:

$$I_\nu(y) = 1 \text{ or } 0 \text{ if } S_{T_\nu(y)} - S_{K_{\nu-1}} \ge y \text{ or } S_{T_{\nu(y)}} - S_{K_{\nu-1}} \le 0, \text{ respectively}. \tag{5}$$

Let $y > 0$ be given. Because of irreducibility and the presence of a finite cycle of states of positive increase, $\mathscr{C} = \{A_0 = i_0, \dots, A_{k_0} = i_{k_0} = i_0\}$, where

$$\Pr\left\{\sum_{i=1}^m X_i > 0,\, m = 1, \dots, k_0; \mathscr{C}\right\} > 0,$$

it follows that within each successive $r - 1$ excursions (recall r is the number of states) there is a uniform (with respect to the initial state) positive probability that a level exceeding y is traversed during one or more of these excursions. Consequently, with probability 1 some excursion (actually infinitely many excursions) of the partial sum process reaches level y or greater. In particular, for each $y > 0$, there exists a first excursion epoch with the properties $I_\nu(y) = 1$, while $I_1(y) = \cdots = I_{\nu-1}(y) = 0$.

We can now state our main results (resolving the determination of the parameters w^*, u^* in Lemma 1).

THEOREM 1. *Let* $I_\nu(y) = 1$ *while* $I_1(y) = \cdots = I_{\nu-1}(y) = 0$ [*see* (5)]. *Then*

$$\frac{L_\nu(y)}{y} \to \frac{1}{w^*} \quad a.s.\ as\ y \to \infty, \tag{6}$$

for any value of A_0.

THEOREM 2. *With the index* ν *determined as in Theorem* 1,

$$\frac{W_\nu(y)}{L_\nu(y)} \to u^* \quad a.s.\ as\ y \to \infty, \tag{7}$$

for any value of A_0.

Since $M_\alpha(n) \to \infty$ a.s. as $n \to \infty$ [see Karlin and Dembo (1991)], it follows from Theorems 1 and 2 that also

$$\frac{L(M_\alpha(n))}{M_\alpha(n)} \to \frac{1}{w^*} \quad \text{and} \quad \frac{W(M_\alpha(n))}{L(M_\alpha(n))} \to u^* \quad \text{a.s. as } n \to \infty,$$

with $L(M_\alpha(n)), W(M_\alpha(n))$ defined on the segment of maximal score over the time frame $\{1, \dots, n\}$.

Let $\rho(\theta)$ be the maximal eigenvalue of the matrix of elements

$$(P_\theta)_{ij} = p_{ij} E\left[e^{\theta X_1} | A_0 = i,\ A_1 = j\right]. \tag{8}$$

and $\pi_\theta \gg 0$ and $\psi_\theta \gg 0$ the corresponding right and left eigenvectors normalized so that $\langle \pi_\theta, \psi_\theta \rangle = 1$ and $\langle \pi_\theta, \mathbf{e} \rangle = \sum_{j=1}^{r} \pi_\theta(j) = 1$.

LEMMA 1. *The equation* $\rho(\theta) = 1$ *has a unique positive root* θ^* *and*

$$w^* = \sum_{i,j} p_{ij} E\left(X_1 e^{\theta^* X_1} | A_0 = i,\ A_1 = j\right) \pi_{\theta^*}(j) \psi_{\theta^*}(i) > 0, \tag{9}$$

$$u^* = \sum_{i,j} p_{ij} E\left(U_1 e^{\theta^* X_1} | A_0 = i,\ A_1 = j\right) \pi_{\theta^*}(j) \psi_{\theta^*}(i). \tag{10}$$

For the i.i.d. case cf. Dembo and Karlin [(1990a), Equations (21) and (23)]. Proofs are given in Sections 2 and 3 and examples are given in Sections 4 and 5.

2. Proof of Theorems 1 and 2. The proofs of Theorems 1 and 2 in several respects parallel the i.i.d. case [Dembo and Karlin (1991)]. In particular an analogous sequence of nine lemmas is used here.

The first four lemmas furnish the essential estimates for a typical excursion extended to the Markov case; their proofs are given in the next section. The revised Lemma 1 is stated above, while the revised Lemmas 2, 3 and 4 are stated here where

$$\Omega_+ = \{i: \text{for all } y > 0, \Pr\{I_1(y) = 1 | A_0 = i\} > 0\}.$$

LEMMA 2. *For any $i \in \Omega_+$ (θ^* determined as in Lemma 1),*

$$0 < \delta \leq \Pr\{I_1(y) = 1 | A_0 = i\} e^{\theta^* y} \leq K,$$

where K and δ do not depend on i.

LEMMA 3. *For $y \to \infty$ [see (4) for definitions],*

$$E\left[\left|\frac{L_1(y)}{y} - \frac{1}{w^*}\right|^4 \middle| I_1(y) = 1, A_0 = i\right] = O\left(\frac{1}{y^2}\right) \tag{11}$$

uniformly with respect to $i \in \Omega_+$.

LEMMA 4. *For $y \to \infty$ [see (4)]*

$$E\left[\left|\frac{W_1(y)}{L_1(y)} - u^*\right|^4 \middle| I(y) = 1, A_0 = i\right] = O\left(\frac{1}{y^2}\right) \tag{12}$$

uniformly with respect to $i \in \Omega_+$.

Since the Markov chain is finite, for all $y \geq y_0$, any integer ν and any initial state, $I_\nu(y) = 1$ imply $A_{K_{\nu-1}} = i \in \Omega_+$, where y_0 is independent of both ν and A_0. Thus the estimates of Lemmas 3 and 4 establish the analogs of Lemmas 5, 8 and 9 paraphrasing the arguments of the i.i.d. case with averaging over $\Pr\{A_{K_{\nu-1}} = i | A_{K_{\nu-1}} \in \Omega_+\}$.

The index random variable J, where the global maximum of the partial sums sequence $\{S_m\}$ is first attained, is well defined and finite valued owing to $E[X] < 0$. Moreover, as in the i.i.d. case, from large deviation theory J has at least an exponential decay tail probability, even conditioning on $A_0 = i$. We can conclude that $\Pr\{J = k | A_0 = i\} \leq ce^{-bk}$ with c and b both positive constants independent of i and the proof of Lemma 7 follows the i.i.d. case, mutatis mutandis.

The excursions characterized by $\{\kappa_j(n), \tau_j(n), \sigma_j(n)\}$ (the first $[A \log n]$ excursions with $I_\nu(n) = 1$) are no longer independent, and their initial states $A_{\tau_j(n)}$ may belong to Ω_-, the complement of Ω_+ in $\{1, \dots, r\}$. Nevertheless, the key step of Lemma 6, namely $\Pr\{\mathscr{E}_n\} \leq (1 - a)^{[A \log n]}$, follows from

$$\Pr\{-n_0 < S_1, S_2, \dots, S_{m_0 - 1} < 1 < S_{m_0} | A_0 = i\} = a > 0, \tag{13}$$

with $m_0, n_0 < \infty$ and $a > 0$ all independent of i. We establish (13) as follows. Since the chain is irreducible there exists $m_1 < \infty$ and for $n_0 > m_1 K$, such that $\Pr\{A_{m_1} = i_0 | A_0 = i\} > 0$, $\min\{S_1, \dots, S_{m_1}\} > -n_0$. Here i_0 is the initial state of a cycle $\mathscr{C}$ of length k with $\Pr\{\Sigma_{i=1}^m X_i > 0,\ m = 1, \dots, k;\ \mathscr{C}\} > 0$. As $\mathscr{C}$ is of finite length and $\Pr\{\Sigma_{i=1}^m X_i \geq \delta,\ m = 1, \dots, k;\ \mathscr{C}\} > 0$ for some $\delta > 0$, (13) follows by cycling through $\mathscr{C}$ at most $(1 + n_0)/\delta$ times.

3. Proofs of Lemmas 1–4.

PROOF OF LEMMA 1. Since P is irreducible, so are P_θ, and the Frobenius theory guarantees that their spectral radii $\rho(\theta) > 0$ are also simple eigenvalues having unique (up to scaling) strictly positive $\boldsymbol{\pi}(\theta)$ and $\boldsymbol{\psi}(\theta)$ principal eigenvectors. Without loss of generality we normalize $\boldsymbol{\pi}(\theta)$ and $\boldsymbol{\psi}(\theta)$ so that $\langle \boldsymbol{\psi}(\theta), \boldsymbol{\pi}(\theta) \rangle = 1$ and $\langle \mathbf{e}, \boldsymbol{\pi}(\theta) \rangle = 1$.

Because all the entries of P_θ are analytic and log convex in θ, it is well known that $\log \rho(\theta)$ is convex and the components of $\boldsymbol{\pi}(\theta)$ and $\boldsymbol{\psi}(\theta)$ under the normalizations prescribed are analytic in θ. For the case $P_\theta = \| p_{ij} e^{\theta s_{ij}} \|$ we apply Karlin and Ost (1985), Theorem 1, to conclude that $\log \rho(\theta)$ is *strictly convex*. In fact, setting

$$A = \|\sqrt{p_{ij}} \exp(\theta_1 s_{ij}/2)\|, \qquad B = \|\sqrt{p_{ij}} \exp(\theta_2 s_{ij}/2)\|,$$

in terms of the Schur product matrices we have

$$\rho\left(\frac{\theta_1 + \theta_2}{2}\right) = \rho(A \circ B) \leq \sqrt{\rho(A \circ A)}\sqrt{\rho(B \circ B)} = \sqrt{\rho(\theta_1)}\sqrt{\rho(\theta_2)}$$

with equality (since A and B are irreducible) iff $A = D^{-1}BD$ for some positive definite diagonal matrix D. But this is not possible for $\theta_1 \neq \theta_2$ under our assumptions on the X process. In the more general case of $(P_\theta)_{ij} = p_{ij} E[e^{\theta X} | A_0 = i,\ A_1 = j]$, by the Schwarz inequality,

$$\begin{aligned} &E\left[\exp(((\theta_1 + \theta_2)/2)X) | A_0 = i,\ A_1 = j\right] \\ &\quad \leq \left(E\left[e^{\theta_1 X} | A_0 = i,\ A_1 = j\right]\right)^{1/2} \left(E\left[e^{\theta_2 X} | A_0 = i,\ A_1 = j\right]\right)^{1/2} \end{aligned}$$

with equality only if X conditioned on $A_0 = i$ and $A_1 = j$ is constant. The proof continues as above where the strict convexity of $\log \rho(\theta)$ follows by the assumptions on the X process; trivially so if $p_{ij} > 0$ for all i and j and with some further technical arguments if P is merely irreducible. □

The proofs of Lemmas 2–4 rely on the martingale family (with respect to $A_0, A_1, \dots$)

$$P_m = \frac{\exp(\theta S_m + t W_m)}{\rho(\theta, t)^m} \frac{\pi_{\theta,t}(A_m)}{\pi_{\theta,t}(A_0)}, \qquad m = 0, 1, \dots, \tag{14}$$

which constitute the natural Markov chain extension of the Wald martingale prominent in the i.i.d. case. Here $W_m = \sum_{i=1}^{m} U_i$, $\rho(\theta, t)$ is the spectral radius of the matrix $(P_{\theta,t})_{ij} = p_{ij}E[\exp(\theta X_1 + tU_1)|A_0 = i,\ A_1 = j]$, and the strictly positive vector $\boldsymbol{\pi}_{\theta,t}$ is the corresponding right eigenvector, normalized so that $\sum_i \pi_{\theta,t}(i) = 1$. Analyses and extensions of Markov chain martingale families as in (14) are presented in Ney and Nummelin (1987a, b) and Nummelin (1983).

We apply the optional sampling theorem to the first exit time variable of the sum process S_m from the interval $(0, y)$, $L = L(y) \le K_1$. By large deviation estimates the distribution of $L(y)$ tails down exponentially fast as $\Pr\{K_1 > l\} \le \Pr\{S_l > 0\}$ and S_m has negative drift. Therefore, for every i and for $|t|$ sufficiently small and θ close to θ^*,

$$E\left[\exp(\theta S_L + tW_L - L\zeta(\theta, t))\frac{\pi_{\theta,t}(A_L)}{\pi_{\theta,t}(A_0)}\middle| A_0 = i\right] \equiv 1 \tag{15}$$

with $\zeta(\theta, t) = \log \rho(\theta, t)$.

Proof of Lemma 2. The monotonicity of $I(y)$ implies the existence of y_0 such that for all $y \ge y_0$, $\Pr\{I(y) = 1|A_0 \in \Omega_-\} = 0$.

Take $t = 0$, $\theta = \theta^*$ in (15) so that $\zeta(\theta^*, 0) = 0$. Then for any $i \in \Omega_+$,

$$\begin{aligned} 1 &= \Pr\{I(y) = 1|A_0 = i\}e^{\theta^* y} \\ &\quad \times E\left[\exp\big(\theta^*(S_{L(y)} - y)\big)\frac{\pi_{\theta^*,0}(A_{L(y)})}{\pi_{\theta^*,0}(A_0)}\middle| A_0 = i,\ I(y) = 1\right] \\ &\quad + \Pr\{I(y) = 0|A_0 = i\}E\left[\exp(\theta^* S_{L(y)})\frac{\pi_{\theta^*,0}(A_{L(y)})}{\pi_{\theta^*,0}(A_0)}\middle| A_0 = i,\ I(y) = 0\right]. \end{aligned} \tag{16}$$

Since $S_{L(y)} \ge y$ when $I(y) = 1$ denoting $c(\theta^*, 0) = \min_i\{\pi_{\theta^*,0}(i)\} > 0$, and by the normalization, $\pi_{\theta^*,0}(i) \le 1$, the upper bound

$$1 \ge \Pr\{I(y) = 1|A_0 = i\}e^{\theta^* y}c(\theta^*, 0)$$

ensues by discarding the second term of (16). For any $a_+ > 0 > a_-$, let $I(a_+, a_-) = 1$ if and only if there exists m such that $S_1, \dots, S_{m-1} \in (a_-, a_+)$ and $S_m \ge a_+$. As $I(a_+, a_-)$ decreases in both a_- and a_+ for any $y_0 \in (0, y - K)$,

$$\begin{aligned} &\Pr\{I(y) = 1|A_0 = i\} \\ &\qquad \ge \Pr\{I(y_0) = 1|A_0 = i\}\min_j \Pr\{I(y - y_0, -y_0) = 1|A_0 = j\}. \end{aligned} \tag{17}$$

By the optional sampling theorem for the stopping time $L = L(y - y_0, -y_0)$ [the first exit of S_m from the open interval $(-y_0, y - y_0)$, associated with $I(y - y_0, -y_0)$] applied to the martingale $\exp(\theta^* S_n)(\pi_{\theta^*,0}(A_n))/(\pi_{\theta^*,0}(A_0))$, we have

$$\begin{aligned}
1 &= E\left[e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\middle| I(y-y_0,-y_0)=1, A_0=j\right]\\
&\quad\times \Pr\{I(y-y_0,-y_0)=1|A_0=j\}\\
&\quad+E\left[\exp(\theta^* S_L)\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\middle| I(y-y_0,-y_0)=0, A_0=j\right]\\
&\quad\times \Pr\{I(y-y_0,-y_0)=0|A_0=j\}\\
&\le \left[\frac{\exp(\theta^*(y-y_0+K))}{c(\theta^*,0)}\right]\Pr\{I(y-y_0,-y_0)=1|A_0=j\}\\
&\quad+\left[\frac{\exp(-\theta^* y_0)}{c(\theta^*,0)}\right]\left(1-\Pr\{I(y-y_0,-y_0)=1|A_0=j\}\right).
\end{aligned}$$

Combining, we have

$$e^{\theta^* y}\Pr\{I(y)=1|A_0=i\}\ge \Pr\{I(y_0)=1|A_0=i\}\frac{c(\theta^*,0)e^{\theta^* y_0}-1}{e^{\theta^* K}-e^{-\theta^* y}}.$$

Since $i\in\Omega_+$, $\Pr\{I(y_0)=1|A_0=i\}\ge\varepsilon>0$ for $y_0=-(1/\theta^*)\log[c(\theta^*,0)/2]$. So for any y

$$e^{\theta^* y}\Pr\{I(y)=1|A_0=i\}\ge\frac{\varepsilon}{e^{\theta^* K}-1}=\delta>0,$$

where the condition of $y\le y_0$ is clear by the monotonicity of $I(y)$. □

PROOF OF LEMMA 3. Set

$$G_n(j,i)=\frac{d^n}{d\theta^n}\log\left\{\frac{\pi_{\theta,0}(i)}{\pi_{\theta,0}(j)}\right\}\bigg|_{\theta=\theta^*},$$

$\hat{S}_L=S_L-G_1(A_L,i)$ and $k(\theta)=(d^2/d\theta^2)\log\rho(\theta,0)$.

Following the recipe in Dembo and Karlin [(1991), Equations (35)–(36)], differentiating successively (15) in θ leads to

$$E\left[(\hat{S}_L-w^*L)e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right]=0, \tag{18a}$$

$$\begin{aligned}
&E\left[(\hat{S}_L-w^*L)^2e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right]\\
&\quad=E\left[(k(\theta^*)L+G_2(A_L,i))e^{\theta^* S_L}\pi_{\theta,0}(A_L)|A_0=i\right],
\end{aligned}\tag{18b}$$

$$\begin{aligned}
&E\left[(\hat{S}_L-w^*L)^4e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right]\\
&\quad=6E\left[(\hat{S}_L-w^*L)^2(k(\theta^*)L+G_2(A_L,i))e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right]\\
&\qquad+4E\left[(\hat{S}_L-w^*L)(k'(\theta^*)L+G_3(A_L,i))e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right]\\
&\qquad+E\left[(k''(\theta^*)L+G_4(A_L,i))e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right]\\
&\qquad-3E\left[(k(\theta^*)L+G_2(A_L,i))^2e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i\right].
\end{aligned}\tag{19}$$

Since $\pi_{\theta,0}$ is positive and analytic in θ, $\max_{j,i}|G_n(j,i)| \le C$, $n = 1,2,4$. This yields [cf. Dembo and Karlin (1991), Equations (38)–(43)]

$$E\Big[\big(\hat{S}_L - w^*L\big)^4 e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i\Big] = O(y^2). \tag{20}$$

Applying the lower bound of Lemma 2 for $i \in \Omega_+$ results in

$$\begin{aligned} &E\Big[\big(\hat{S}_L - w^*L\big)^4 \exp(\theta^*(S_L - y))\pi_{\theta^*,0}(A_L)|I(y) = 1,\, A_0 = i\Big] \\ &\quad = O(y^2). \end{aligned} \tag{21}$$

Since both $\hat{S}_L - y$ and $\exp(\theta^*(S_L - y))\pi_{\theta^*,0}(A_L)$ are bounded when $I(y) = 1$, we can reduce (21) to

$$E\left[\left(\frac{L(y)}{y} - \frac{1}{w^*}\right)^4 \middle| I(y) = 1,\, A_0 = i\right] \le \frac{C}{y^2} \tag{22}$$

for some constant C independent of i and all y large enough. □

Proof of Lemma 4. In a similar manner we derive the analog of Dembo and Karlin [(1991), Equation (46)], namely

$$\begin{aligned} &E\Big[\big(\hat{W}_L - u^*L\big)^4 \exp(\theta^*(S_L - y))\pi_{\theta^*,0}(A_L)|I(y) = 1,\, A_0 = i\Big] \\ &\quad = O(y^2), \end{aligned} \tag{23}$$

where

$$u^* = \frac{d}{dt}\log\rho(\theta^*, t)\Big|_{t=0},$$

$$\hat{W}_L = W_L - H_1(A_L, i),$$

$$H_n(j,i) = \frac{d^n}{dt^n}\log\left\{\frac{\pi_{\theta^*,t}(i)}{\pi_{\theta^*,t}(j)}\right\}\Bigg|_{t=0}$$

[so $\max_{j,i}|H_n(j,i)| \le C$, $n = 1,2,4$]. We can now convert (23) [as done with (21)] into

$$E\left[\left(\frac{W(y)}{L(y)} - u^*\right)^4\left(\frac{L(y)}{y}\right)^4 \middle| I(y) = 1,\, A_0 = i\right] = O\left(\frac{1}{y^2}\right). \tag{24}$$

As $L(y) \ge y/K$ conditional on $I(y) = 1$, the proof of the lemma is complete. □

4. Examples. A complete accounting of the frequencies of the different transitions for large scoring segments is given for two letter sequences generated as a Markov chain with scoring values of the form $s_{\alpha\beta} = +1$ or -1. In this context consider a two-state Markov chain with scores $X_i = 1$ or -1 and letter transition probability matrix

$$\begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix},$$

where $\alpha = \Pr\{A_1 = 0 | A_0 = 0\}$ and $\beta = \Pr\{A_1 = 1 | A_0 = 1\}$. The transition matrix that governs the score values has the form

$$p_{ij}^* = p_{ij} e^{\theta^* s_{ij}},$$

where θ^* is determined as the unique positive root of the equation $\rho(\theta) = 1$ [$\rho(\theta)$ being the spectral radius of the matrix $\|p_{ij} e^{\theta s_{ij}}\|$].

The feasible scoring arrays (carrying negative mean, with some positive probability for high scores) divide into three categories.

CATEGORY 1. X_1 depends only on one state, for example, $X_1 = (-1)^{A_1}$ ($s_{01} = s_{11} = -1$, $s_{10} = s_{00} = 1$). Here $\beta > \alpha$ for negative mean.

CATEGORY 2. The score measures state transitions, for example, $X_1 = (-1)^{A_0 + A_1}$, so $s_{01} = s_{10} = -1$, $s_{00} = s_{11} = 1$; here $2 > \alpha/(1-\alpha) + \beta/(1-\beta)$ for negative mean.

CATEGORY 3. The score looks for long runs of the same state, for example, $X_1 = (-1)^{\sup(A_0, A_1)}$ with $s_{00} = 1$ and -1 otherwise. Here, $\beta/(1-\beta) + 2 > \alpha/(1-\alpha)$ for negative mean.

In all three categories (as well as in the general case), the strong law of empirical distributions on high score segments is determined by the conjugate Markov chain. This chain has transition matrix

$$\begin{bmatrix} \gamma & 1-\gamma \\ 1-\delta & \delta \end{bmatrix},$$

with $\gamma = \alpha e^{\theta^* s_{00}}$, $\delta = \beta e^{\theta^* s_{11}}$ and where w^* and u^* are first moments of its stationary distribution.

Therefore, in order to specify the parameters of Theorems 1 and 2, one needs to solve for θ^*, the root of the equation $\rho(\theta) = 1$.

For the scoring examples mentioned above,

$$P_\theta^{(1)} = \begin{bmatrix} \alpha e^{\theta} & (1-\alpha)e^{-\theta} \\ (1-\beta)e^{\theta} & \beta e^{-\theta} \end{bmatrix} \quad \text{in Category 1,}$$

$$P_\theta^{(2)} = \begin{bmatrix} \alpha e^{\theta} & (1-\alpha)e^{-\theta} \\ (1-\beta)e^{-\theta} & \beta e^{\theta} \end{bmatrix} \quad \text{in Category 2,}$$

$$P_\theta^{(3)} = \begin{bmatrix} \alpha e^{\theta} & (1-\alpha)e^{-\theta} \\ (1-\beta)e^{-\theta} & \beta e^{-\theta} \end{bmatrix} \quad \text{in Category 3.}$$

Letting $\xi = e^{\theta}$ solves the following equations

(25a) $$1 = \alpha\xi + \beta\xi^{-1} - \alpha\beta + 1 - \alpha - \beta + \alpha\beta \quad \text{for } P_\theta^{(1)},$$

(25b) $$1 = (\alpha + \beta)\xi - \alpha\beta\xi^2 + (1 - \alpha)(1 - \beta)\xi^{-2} \quad \text{for } P_\theta^{(2)},$$

(25c) $$1 = \alpha\xi + \beta\xi^{-1} - \alpha\beta + (1 - \alpha)(1 - \beta)\xi^{-2} \quad \text{for } P_\theta^{(3)}.$$

Equation (25a), apart from the root $\xi = 1$, has the relevant root $\xi^* = \sqrt{\beta/\alpha} > 1$, so that the conjugate transition matrix ($\beta > \alpha$) in this case is

$$\begin{bmatrix} \beta & (1 - \beta) \\ (1 - \alpha) & \alpha \end{bmatrix},$$

where the α and β parameters are interchanged compared to the original transition matrix.

Equation (25b) has four roots, $\xi_0 = 1$, the trivial root [corresponding to $\rho(0) = 1$], a negative irrelevant root and two positive roots. Rearranging (25b) and factoring out $(\xi - 1)$ we obtain

(26) $$\xi\left(\frac{\xi}{\hat{\alpha}} - 1\right)\left(\frac{\xi}{\hat{\beta}} - 1\right) + \left(1 - \frac{\xi}{\sqrt{\hat{\alpha}\hat{\beta}}}\right)\left(1 + \frac{\xi}{\sqrt{\hat{\alpha}\hat{\beta}}}\right) = 0,$$

where $\hat{\alpha} = 1/\alpha - 1$, $\hat{\beta} = 1/\beta - 1$. From the two positive roots, the larger root lies in the interval $(\max(\hat{\alpha}, \hat{\beta}) + 1, \infty)$ and the smaller root is located in the interval $(\min(\hat{\alpha}, \hat{\beta}), \sqrt{\hat{\alpha}\hat{\beta}})$. The larger root entails either $\alpha e^{\theta} > 1$ or $\beta e^{\theta} > 1$, and accordingly this outcome is not relevant. The smaller positive root of (26) produces the correct conjugate transition matrix:

$$\begin{bmatrix} \xi^*\alpha & 1 - \xi^*\alpha \\ 1 - \xi^*\beta & \xi^*\beta \end{bmatrix}.$$

Equation (25c) has three roots, one of them $\xi = 1$. Of the remaining two roots, one is negative and therefore irrelevant. The relevant root is

$$\xi^* = \frac{1}{2\alpha}\left(1 - \alpha + \alpha\beta + \sqrt{(1 - \alpha + \alpha\beta)^2 + 4\alpha(1 - \alpha)(1 - \beta)}\right),$$

yielding the conjugate transition matrix

$$\begin{bmatrix} \alpha\xi^* & 1 - \alpha\xi^* \\ 1 - \beta/\xi^* & \beta/\xi^* \end{bmatrix}.$$

5. Brownian motion and Poisson process with drift. Consider the diffusion $X(t) = \sigma B(t) - \mu t$, $\mu > 0$, with $\{B(t)\}_{t \geq 0}$ being the standard Brownian motion on $\mathbb{R}$. Clearly $X(t) \to -\infty$ a.s. as $t \to \infty$, but the maximal segmental exceedance $M(t) = \max_{0 \leq u \leq v \leq t}(X(v) - X(u)) \to \infty$ a.s. as $t \to \infty$ at the rate of $O(\log t)$; see, for example, Karlin and Dembo (1992). The first passage times $T(y) = \inf\{t\colon M(t) \geq y\}$ are therefore a.s. finite. The composition during the maximal segmental exceedance $\{u^*$ to $v^*\}$ such that $M(t) = X(v^*) - X(u^*)$ is the same as in the first y exceedance $\{K(y)$ to $T(y)\}$ with $X(T(y)) - X(K(y)) = y$.

The latter event can also be characterized as the limit as $\varepsilon \to 0$ of the law of $\{X(t)\}_{t=0}^{L_\varepsilon(y)}$ conditioned on $X(L_\varepsilon(y)) = y$ and $X(t) \in (-\varepsilon, y)$ for $0 < t \le L_\varepsilon(y)$.

The conditioned process, designated $X^*(t)$, is also a diffusion with drift $\mu(x) = \mu \coth(\mu(x+\varepsilon)/\sigma^2)$ and the same variance σ^2. This is an application of the general formula of Karlin and Taylor [(1981), page 263, Equation (9.5)]. The diffusion process $X^*(t)$ on the state space $(-\varepsilon, y)$ has an entrance boundary at $-\varepsilon$ and exit boundary at y.

Consider now the random variables

$$W_\varepsilon(y) = \int_0^{L_\varepsilon(y)} g\left(\frac{X^*(t)}{y}\right) dt,$$

with $g(\cdot)$ bounded and properly continuous. Form $\tilde{X}^*(ty)/y$, a diffusion process derived from $X^*(t)$ by a time and scale change. The process $\tilde{X}^*(t)$ has drift $\mu \coth(\mu(yx+\varepsilon)/\sigma^2)$ and variance σ^2/y, so it converges to a deterministic process with drift μ, uniformly for $x \in [\delta, \infty)$ as $y \to \infty$ (for any $\delta > 0$).

Then

$$\frac{1}{y} W_\varepsilon(y) = \int_0^{L_\varepsilon(y)/y} g\big(\tilde{X}^*(t)\big)\, dt.$$

As $y \to \infty$ and $\varepsilon \to 0$, $-\varepsilon/y \to 0$ is the entrance boundary of $\tilde{X}^*(t)$. Therefore, as $y \to \infty$, $\tilde{X}^*(t) \ge \delta$ for all $t \ge \tau$ and some small $\delta, \tau > 0$. In particular, for $g = 1$, $W_\varepsilon(y) = L_\varepsilon(y)$ and as $y \to \infty$, $\tilde{X}^*(t) \to \mu t$. The i.i.d. theory readily implies that $L_\varepsilon(y)/y \to 1/\mu$.

It follows that

$$\begin{aligned}\lim_{\varepsilon \downarrow 0} \lim_{y \to \infty} \frac{1}{L_\varepsilon(y)} W_\varepsilon(y) &= \lim_{\varepsilon \downarrow 0} \lim_{y \to \infty} \mu \int_0^{L_\varepsilon(y)/y} g\big(\tilde{X}^*(t)\big)\, dt \\ &= \mu \int_0^{1/\mu} g(\xi\mu)\, d\xi = \int_0^1 g(\xi)\, d\xi.\end{aligned}$$

Consider $S(t) = N(t) - at$ where $N(t)$ is a Poisson process of rate parameter λ and $a > \lambda$ is assumed. Let

$$M(u) = \max_{\substack{0 \le \tau \le u \\ 0 \le t \le u - \tau}} [S(t+\tau) - S(t)]$$

be the maximal segmental value of the process $S(t)$ in the time interval $[0, u]$. Let $[K_u, T_u]$ be any segment of maximal value, that is, $T_u \ge K_u$ and

$$S(T_u) - S(K_u) = M(u).$$

We ascertain strong laws for the length of such segments $L_u = T_u - K_u$ and therefore also for the number of jumps occurring in them. For this objective let θ^* be the unique positive root of the equation

$$\theta a - \lambda(e^\theta - 1) = 0.$$

The strong law of $L_u/M(u)$ as $u \to \infty$ is the same as the strong law for $L_\varepsilon(y)/y$ (as $\varepsilon \to 0$ and $y \to \infty$), $L_\varepsilon(y)$ being the first exit time of $S(t)$ from

$(-\varepsilon, y)$ conditioned on the event $\mathscr{E}(y)$ $\{S(L_\varepsilon(y)) \geq y\}$. Using the Wald analog exponential martingale family $P_\theta(t) = \exp(\theta N(t) - \lambda t(e^\theta - 1))$, it can be shown that

$$\frac{(\lambda e^{\theta^*} - a)}{y} L_\varepsilon(y) \to 1 \quad \text{a.s. as } \varepsilon \downarrow 0 \text{ and } y \uparrow \infty,$$

reflecting the limiting behavior of $\{N(t)\}_{t=0}^{L_\varepsilon(y)}$, conditional on $\mathscr{E}(y)$ as a Poisson process of parameter λe^{θ^*}.

The proof may either be directly implemented by exploiting the martingale character of $P_\theta(t)$, or via discretization of $N(t)$ as the sum of tn i.i.d. Poisson random variables each with rate λ/n and letting $n \to \infty$.

An alternative method is to exploit the i.i.d. exponential waiting times between unit increases. Since the maximal segment will start and end at a positive jump, this segment score is based on partial sums of the i.i.d. variables $X_i = 1 - aW_i$, where W_i are exponential with parameter λ and the 1 reflects the Poisson jump. We can apply the results of Dembo and Karlin (1991) for X_i with $E[e^{\theta^* X_i}] = 1$ equivalent to $e^{\theta^*}E[\exp(-a\theta^* W_i)]$ or $\lambda e^{\theta^*} = \theta^* a + \lambda$. Moreover, multiplying by $e^{\theta^* X_i}$ changes the original Poisson process to the conjugate Poisson process of rate λe^{θ^*}.

The foregoing method can also be used to analyze count renewal processes subject to linear negative drift.

REFERENCES

DEMBO, A. and KARLIN, S. (1991). Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables. *Ann. Probab.* **19** 1737–1755.

KARLIN, S. and ALTSCHUL, S. F. (1990). New methods for assessing the statistical significance of molecular sequence features using general scoring schemes. *Proc. Nat. Acad. Sci. USA* **87** 2264–2268.

KARLIN, S. and DEMBO, A. (1992). Limit distributions of maximal segmental score among Markov dependent partial sums. *Adv. in Appl. Probab.* To appear.

KARLIN, S., DEMBO, A. and KAWABATA, T. (1990). Statistical composition of high scoring segments from molecular sequences. *Ann. Statist.* **18** 571–581.

KARLIN, S. and OST, F. (1985). Some monotonicity properties of Schur powers of matrices and related inequalities. *Linear Algebra Appl.* **68** 47–65.

KARLIN, S. and TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic, New York.

NEY, P. and NUMMELIN, E. (1987a). Markov additive processes. I. Eigenvalues properties and limit theorems. *Ann. Probab.* **15** 561–592.

NEY, P. and NUMMELIN, E. (1987b). Markov additive processes. II. Large deviations. *Ann. Probab.* **15** 593–609.

NUMMELIN, E. (1983). *General Irreducible Markov Chains and Non Negative Operators*. Cambridge Univ. Press.

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305