Stochastic Processes and their Applications 45 (1993) 259-271
North-Holland

Central limit theorems of partial sums for large segmental values

Amir Dembo and Samuel Karlin*

Department of Mathematics, Stanford University, Stanford, CA, USA

Received 15 August 1991
Revised 11 March 1992

Let (X_i, U_i) be i.i.d., X_i real valued and U_i vector valued, bounded random variables or governed by a finite state Markov chain. Assuming that $E[X]<0$ and $P(X>0)>0$, central limit theorems are derived for $\sum_i U_i$ on segments conditioned that $\sum_i X_i$ is increasingly high, going to $+\infty$. While these segments are exponentially rare, they are of importance in many models of stochastic analysis including queueing systems and molecular sequence comparisons. Particular applications give central limit theorems for the empirical frequencies over such segments and for their length.

AMS 1991 Subject Classifications: Primary 60F05, 60F10.

large segmental sums * conditioned central limit theorem * large deviations

Introduction

Many random structures of theoretical and practical importance are associated with sequences of real random variables of high aggregate values having small probability, often exponentially small. In this context we set forth a class of Gaussian distributional limit theorems conditioned on rare events. The results can be construed as a central limit theorem in the context of large deviation theory. Examples relate to the maximal waiting time in stable queueing systems, and with reference to maximal riskless insurance epochs, see Iglehart (1972, 1974), Siegmund (1975), Kao (1978), Asmussen (1982), Arantharam (1988). Our motivation stems from biomolecular sequence comparisons, Karlin and Altschul (1990), Karlin et al. (1990).

In the simplest formulation let $X_1, X_2, \ldots, X_n$ be i.i.d. real valued bounded random variables obeying the conditions

$$E[X]=\mu<0 \quad \text{and} \quad \Pr\{X>0\}>0 \tag{1}$$

so that unrestricted the partial sum process $\{S_0=0,\ S_m=\sum_{i=1}^m X_i\}$ entails a negative drift. Let $U_1, U_2, \ldots, U_n$ be i.i.d. bounded random vectors, U_k generally dependent

Correspondence to: Dr. Samuel Karlin, Department of Mathematics, Stanford University, Stanford, CA 94305-2125, USA.

* Research supported in part by NIH grants GM 39907-02, GM 10452-27 and NSF grant DMS 86-06244.

on X_k. Define $T(0, y)$ to be the first passage time (index) when the process $\{S_m\}_{m=1}^{\infty}$ first departs the open interval $(0, y)$, $y>0$ and designate $\mathscr{E}_{0,y}^{y}$ as the event $S_{T(0,y)} \geq y$.

On the basis of (1) it is simple to see that

$$\Pr\{\mathscr{E}_{0,y}^{y}\} \to 0 \quad \text{as } y \to \infty. \tag{2}$$

For the realizations of $\mathscr{E}_{0,y}^{y}$ (increasingly rare events as $y \to \infty$) we establish a central limit theorem for the partial sums $W(y) = \sum_{k=1}^{T(0,y)} \boldsymbol{U}_k$. When properly normalized, the precise statement for the i.i.d. model is stated in Theorem 1 and for the case of Markov dependence in Theorem 2. To this end we determine the following quantities. Let

$$\rho(\theta, \boldsymbol{t}) = \log E[\exp\{\theta X_1 + \langle \boldsymbol{t}, \boldsymbol{U}_1 \rangle\}] \tag{3}$$

where $\langle \boldsymbol{t}, \boldsymbol{U} \rangle$ is the standard inner product of the indicated vectors. Designate by $\theta^* > 0$ the unique positive root of the equation $\rho(\theta^*, \boldsymbol{0}) = 0$ (existence and uniqueness ensue on account of (1)). Set

$$w^* = \frac{\partial \rho}{\partial \theta}(\theta^*, \boldsymbol{0}) = E[X_1 \exp(\theta^* X_1)] \quad \text{and} \quad \boldsymbol{u}^* = E[\boldsymbol{U}_1 \exp(\theta^* X_1)]. \tag{4}$$

The value w^* is positive.

Theorem 1. *Conditioned on* $\mathscr{E}_{0,y}^{y}$,

$$\sqrt{\frac{w^*}{y}} \sum_{k=1}^{T(0,y)} (\boldsymbol{U}_k - \boldsymbol{u}^*) \tag{5}$$

converges in distribution $(y \to \infty)$ *to a multivariate normal distribution of zero mean vector and covariance matrix* $\boldsymbol{C} = \|c_{ij}\|$ *where*

$$c_{ij} = \frac{\partial^2}{\partial t_i \, \partial t_j} \rho(\theta^*, \boldsymbol{0}) = E((\boldsymbol{U}_1 - \boldsymbol{u}^*)_i (\boldsymbol{U}_1 - \boldsymbol{u}^*)_j \exp(\theta^* X_1)). \tag{6}$$

For applications, specifying $U_k = X_k$ (so $u^* = w^*$) and since conditioned on $\mathscr{E}_{0,y}^{y}$, $S_{T(0,y)} - y$ is bounded (because X_i are bounded) we get:

Corollary 1. *Conditioned on* $\mathscr{E}_{0,y}^{y}$, $\sqrt{w^*/y}\{y - w^* T(0, y)\}$ *converges in distribution as* $y \to \infty$ *to a normal random variable of zero mean and variance* $\sigma^2 = E[(X - w^*)^2 \exp\{\theta^* X\}]$. □

This corollary was first obtained by Siegmund (1975). The method of proof is different from ours.

Corollary 2. *Let A be a Borel set in the range of X. Define* $U_k = I_A(X_k)$, *where* $I_A(X) = 1$ *for* $X \in A$, 0 *for* $X \notin A$. *Conditioned on the event* $\mathscr{E}_{0,y}^{y}$ *the empirical distribution of samples* $\mu(A; y)$ *satisfies* $\sqrt{T(0, y)}[\mu(A; y) - \mu^*(A)] \to \mathrm{N}(0, c^*)$, *where* $\mu^*(A) = E[I_A(X) \exp\{\theta^* X\}]$ *and* $c^* = \mu^*(A) - \mu^*(A)^2$. □

For the applications in molcular biology we need to adapt Theorem 1 with the conditioning put in the following form. Let $\sigma(y)$ be the one-sided first passage time that some segment $(S_l - S_k)$ reaches a level $\geq y$. This event occurs with probability 1 (Dembo and Karlin, 1991a). Let $\kappa(y)$ be the largest index $< \sigma(y)$ with the property that $S_k - S_{\kappa(y)} > 0$ for $k = \kappa(y)+1, \ldots, \sigma(y)$ and $S_{\sigma(y)} - S_{\kappa(y)} \geq y$. Set $L(y) = \sigma(y) - \kappa(y)$. Then

$$\frac{1}{\sqrt{L(y)}} \sum_{k=\kappa(y)+1}^{\sigma(y)} [\boldsymbol{U}_k - \boldsymbol{u}^*] \tag{7}$$

converges in distribution to a normal random vector of mean $\mathbf{0}$ and convariance matrix as displayed in (6).

For the Markov development of Theorem 1 we consider an underlying irreducible Markov chain (M.C.) governed by the transition probability matrix $\boldsymbol{P} = \| p_{\alpha\beta} \|$ on r states with stationary frequency vector $\boldsymbol{\pi} = (\pi_\alpha)$. Consider, with each transition α to β, i.i.d. real valued bounded random variables $\{X_{\alpha\beta}\}$. We focus on those realizations $\{\alpha_i\}$ of the M.C. and corresponding sample paths of $S_{m,\alpha} = \sum_{i=0}^{m-1} X_{\alpha_i \alpha_{i+1}}$ $(\alpha_0 = \alpha)$ with the property that $S_{m,\alpha}$ first departs the open interval $(0, y)$ through the upper level. Denote this first passage time by $T_\alpha(0, y)$ where α is the initial state of the M.C. (We suppress α when no ambiguity is possible.) Again we denote the event $S_{T(0,y)} \geq y$ by $\mathscr{E}_{0,y}^{y}(\alpha)$. The assumptions corresponding to the requirements (1) are

$$\sum_{\alpha,\beta} \pi_\alpha p_{\alpha\beta} E[X_{\alpha\beta}] < 0 \tag{8}$$

and for some cycle of states for every initial state α the condition

$$\Pr\left\{ S_{k,\alpha} = \sum_{i=0}^{k-1} X_{\alpha_i \alpha_{i+1}} > 0,\ k = 1, \ldots, m-1 \,\middle|\, \alpha_0 = \alpha_m = \alpha \right\} > 0. \tag{9}$$

holds.

To achieve the analog of Theorem 1 in the M.C. case we construct the irreducible non-negative matrix family

$$q_{ij}(\theta; \boldsymbol{t}) = p_{ij} E[\exp\{\theta X_{\alpha_0, \alpha_1} + \langle \boldsymbol{t}, \boldsymbol{U} \rangle\} \mid \alpha_0 = i,\ \alpha_1 = j]. \tag{10}$$

Let $\lambda(\theta, \boldsymbol{t})$ be the spectral radius of the matrix $\boldsymbol{Q}(\theta, \boldsymbol{t}) = \| q_{ij}(\theta, \boldsymbol{t}) \|$ which is strictly log convex, see Dembo and Karlin (1991b) and form

$$\rho(\theta, \boldsymbol{t}) = \log \lambda(\theta, \boldsymbol{t}). \tag{11}$$

Because of (8) and (9), there is a unique positive θ^* satisfying $\rho(\theta^*, \mathbf{0}) = 0$ (Dembo and Karlin, 1991b). The computations of (4) in the M.C. case have the form

$$w^* = \frac{\partial}{\partial \theta} \rho(\theta^*, \mathbf{0}) \quad \text{and} \quad \boldsymbol{u}^* = \nabla_{\boldsymbol{t}} \rho(\theta^*, \mathbf{0}) \tag{12}$$

where $\nabla_{\boldsymbol{t}}$ signifies the gradient vector extracted from $\rho(\theta, \boldsymbol{t})$ in the $\boldsymbol{t}$ variable.

We are now prepared to state our main result of the conditioned central limit law in the M.C. setting. For each realization of the M.C. let $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots$ be a bounded vector sequence $U_k = U_{\alpha_k, \alpha_{k+1}}$ depending on the M.C. transition α_k to α_{k+1} and $X_{\alpha_k, \alpha_{k+1}}$ and form the partial sums $W_m = \sum_{k=1}^{m} U_k$.

Theorem 2. *Conditioned on $\mathscr{E}_{0,y}^{y}(\alpha)$ for any initial state α,*

$$\frac{\sqrt{w^*}}{\sqrt{y}} \sum_{k=1}^{T(0,y)} (U_k - u^*) \xrightarrow{\text{law}} \mathrm{N}(\mathbf{0}, C) \tag{13}$$

where the covariance matrix $C = \|c_{ij}\|$ has $c_{ij} = (\partial^2/\partial t_i\, \partial t_j)\rho(\theta^*, \mathbf{0})$.

The analogs of Corollary 1 and 2 hold as well in the M.C. framework.

We conclude this introduction and statement of results with some general comments on the organization of the paper and method of proof.

In the i.i.d. univariate case we rely foremostly on properties of the Wald Martingale family

$$\exp[\theta S_n + \langle t, W_n\rangle - n\rho(\theta, t)] \tag{14}$$

and expanding about $(\theta^*, 0)$ coupled to an implicit function argument leading to the following weighted central limit theorem.

Lemma 1. *Assume $u^* = E[U \exp(\theta^* X)] = 0$. Then*

$$\lim_{y\to\infty} \frac{E[\exp\{t\sqrt{w^*/y}\; W_{T(0,y)}\} \exp\{\theta^*(S_{T(0,y)} - y)\} \mid \mathscr{E}_{0,y}^{y}]}{E[\exp\{\theta^*(S_{T(0,y)} - y)\} \mid \mathscr{E}_{0,y}^{y}]} = \exp\{\tfrac{1}{2}t^2 v^*\} \tag{15}$$

where $v^* = E[U^2 \exp(\theta^* X)] = (\partial^2/\partial t^2) \log E[\exp(\theta X + tU)]$ for $\theta = \theta^*$, $t = 0$.

The proof of Theorem 1 is completed by establishing the asymptotic independence of $W_{T(0,y)}$ and $S_{T(0,y)} - y$ conditioned on $\mathscr{E}_{0,y}^{y}$. The multivariate version is reduced to the univariate case by a standard procedure.

The M.C. case of Theorem 1 relies on the generalized Martingale family (prescribing the initial state α_0)

$$R_n(\theta, t) = \exp\{\theta S_n + tW_n - n\rho(\theta, t)\} \frac{\psi_{\alpha_n}(\theta, t)}{\psi_{\alpha_0}(\theta, t)} \tag{16}$$

with $\boldsymbol{\psi} = \{\psi_\alpha(\theta, t)\}$ the unique (normalized so that $\sum_\alpha \psi_\alpha(\theta, t) = 1$) right eigenvector of the matrix $Q(\theta, t)$ defined in (10).

Section 2 is devoted to the proof of Theorem 1. Section 3 elaborates the modifications enabling the proof for the M.C. version (Theorem 2). Asmussen (1982) provides a different proof of Theorem 1 and a functional limit law for the i.i.d. non-lattice case.

2. Proof of Theorem 1

The multivariate version can be reduced to the univariate case (see the discussion at the end of this section). The notation is described in connection with Theorem 1 in the introduction.

Lemma 2. *Let $\{X_i, U_i\}$ be a sequence of i.i.d. pairs of bounded (bound K) real variables such that $E[U\exp(\theta^*X)]=0$, then for all real t,*

$$\lim_{y\to\infty} E\left[\exp\left\{t\sqrt{\frac{w^*}{y}}\sum_{k=1}^{T(0,y)} U_k\right\}\,\middle|\, \mathscr{E}^y_{0,y}\right] = \exp\{\tfrac{1}{2}t^2v^*\}. \tag{17}$$

where v^* is given explicitly in (15).

Let $T(-a, b)$, $-a\leqslant 0<b$, denote the stopping time variable indicating the time (index) where the sum process S_m first exits the open interval $(-a, b)$ and $\mathscr{E}^b_{-a,b}$ signifies the event that $\{S_m\}$ upcrosses the barrier b before downcrossing $-a$. Let $I(-a, b)$ denote the binary variable for the event $\mathscr{E}^b_{-a,b}$ such that $I(-a,b)=1$ or 0 iff $\mathscr{E}^b_{-a,b}$ occurs $(S_{T(-a,b)}\geqslant b)$ or not $(S_{T(-a,b)}\leqslant -a)$, respectively.

Form the Wald martingale sequence

$$R_n(\theta, t)=\exp\{\theta S_n+tW_n-n\rho(\theta,t)\} \tag{18}$$

with

$$S_n=\sum_{i=1}^n X_i,\quad W_n=\sum_{i=1}^n U_i\quad (S_0=W_0=0).$$

The optional sampling theorem for $R_n(\theta^*,0)$ entails, since $\rho(\theta^*,0)=0$, $|S_{T(0,y)}|\leqslant y+K$ and $T(0,y)<\infty$ a.s., that

$$1=E[\exp\{\theta^*S_{T(0,y)}\}]. \tag{19}$$

By the implicit function theorem, since $\rho(\theta,t)$ is analytic for t small enough, and $(\partial\rho/\partial\theta)(\theta^*,0)=w^*>0$, there exists $\theta(t)$ near θ^* and analytic in t satisfying

$$\begin{aligned}0&=\rho(\theta(t),t)\\&=(\theta(t)-\theta^*)w^*+\tfrac{1}{2}v^*t^2+\mathrm{O}[(\theta(t)-\theta^*)^2+t|\theta(t)-\theta^*|+t^3].\end{aligned} \tag{20}$$

It follows that $(\theta(t)-\theta^*)=\mathrm{O}(t^2)$ and (20) reverts to

$$\theta(t)=\theta^*-\tfrac{1}{2}\frac{v^*}{w^*}t^2+\mathrm{O}(t^3). \tag{21}$$

We need the following large deviations estimate:

Lemma 3. *For each $y>0$ and $|t|$ small enough,*

$$E[\exp(tW_{T(0,y)});\ T(0,y)>n]\to 0\quad \textit{uniformly in } y \textit{ as } n\to\infty.$$

Proof. Note that $T(0,y)>n$ entails $S_n>0$. Since S_n has negative drift, the simplest large deviation inequality, e.g. Chernoff (1952) or Varadhan (1984), gives

$$\Pr\{T(0,y)>n\}\leqslant \Pr\{S_n\geqslant 0\}\leqslant C\exp(-I(0)n), \tag{22}$$

where $I(0)=-\inf_{\theta>0}\rho(\theta,0)$. But the terms of $\{U_i\}$ are bounded ($|U_i|\le K$) by hypothesis and therefore

$$E[\exp(tW_{T(0,y)});\ T(0,y)>n]\le E[\exp(|t|KT(0,y));\ T(0,y)>n].$$

The right hand side goes to zero exponentially fast when $|t|$ is small enough. The proof of Lemma 3 is complete. □

On account of Lemma 3 we can apply the optional sampling theorem, maintaining $|t|$ small enough, to get

$$\begin{aligned}1&=E[\exp(\theta(t)S_{T(0,y)}+tW_{T(0,y)}-T(0,y)\rho(\theta(t),t))]\\&=E[\exp(\theta(t)S_{T(0,y)}+tW_{T(0,y)})].\end{aligned}\tag{23}$$

Let $t_y=t\sqrt{w^*/y}$ with y specified sufficiently large in (23) and combine with (19) yielding the equation

$$0=E[\exp\{\theta^*S_{T(0,y)}\}-\exp\{\theta(t_y)S_{T(0,y)}+t_yW_{T(0,y)}\}].\tag{24}$$

Lemma 4.

$$\lim_{y\to\infty}E[(\exp\{\theta(t_y)S_{T(0,y)}+t_yW_{T(0,y)}\}-\exp\{\theta^*S_{T(0,y)}\});\ I(0,y)=1]=0.\tag{25}$$

Proof. Partition the equation (24) into the expression of (25) plus the complementary terms

$$\begin{aligned}&E[\exp\{\theta(t_y)S_{T(0,y)}+t_yW_{T(0,y)}\}-\exp\{\theta^*S_{T(0,y)}\};\ I(0,y)=0]\\&\quad=E[\exp\{\theta(t_y)S_{T(0,y)}\}-\exp\{\theta^*S_{T(0,y)}\};\ I(0,y)=0]\\&\qquad+E[\exp\{\theta(t_y)S_{T(0,y)}+t_yW_{T(0,y)}\}(1-\exp\{-t_yW_{T(0,y)}\});\ I(0,y)=0]\\&\quad=\mathrm{II}+\mathrm{III}.\end{aligned}$$

Note that $S_{T(0,y)}$ under the condition $I(0,y)=0$ is bounded allowing the quantity II to be estimated above by $C|\theta(t_y)-\theta^*|$ which tends to zero at a rate $1/y$.

Since $S_{T(0,y)}\le 0$ in the event $I(0,y)=0$ we have $\exp\{\theta(t_y)S_{T(0,y)}\}\le 1$. Now for each fixed N and y large enough we have

$$\begin{aligned}&E[|\exp\{t_yW_{T(0,y)}\}(1-\exp\{-t_yW_{T(0,y)}\})|;\ I(0,y)=0]\\&\quad\le C_N\left(\frac{1}{\sqrt{y}}\right)+\sum_{n=N}^{\infty}\exp\left\{\frac{|t|K\sqrt{w^*}}{\sqrt{y}}(n+1)\right\}\Pr\{T(0,y)=n+1;\ I(0,y)=0\}\end{aligned}\tag{26}$$

and C_N is a constant depending on N and t. But $\Pr\{T(0,y)=n+1\}\le\exp\{-I(0)n\}$, see Lemma 3, and therefore the series converges geometrically fast keeping t bounded and y large enough. Determine N so that the series term is less than a small arbitrary quantity ε. Subsequently, let $y\to\infty$ showing that the quantity III becomes arbitrarily small and the demonstration of (25) is complete. □

Proof of Lemma 1. We rewrite the limit relation (25) in the form

$$\left\{E\left[\exp((\theta(t_y)-\theta^*)S_{T(0,y)})\exp(\theta^*(S_{T(0,y)}-y))\exp\left(t\sqrt{\frac{w^*}{y}}\,W_{T(0,y)}\right)\,\middle|\,I(0,y)=1\right]\right.$$

$$\left.-E[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1]\right\}\exp(\theta^*y)\Pr\{I(0,y)=1\}\to 0 \quad \text{as } y\to\infty. \tag{27}$$

Now substitute from (21), $\theta(t_y)-\theta^*=-(t^2/2y)v^*+\mathrm{O}(1/y^{3/2})$, and recognizing that $S_{T(0,y)}/y\to 1$ boundedly uniformly over sample paths of the realization $I(0,y)=1$ while $\exp(\theta^*y)\Pr\{I(0,y)=1\}$ (see (31) below) and $E[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1]$ are bounded away from zero we obtain the result of Lemma 1. □

To complete the proof of Lemma 2 it remains to show the asymptotic independence of $W_{T(0,y)}$ and $S_{T(0,y)}-y$ conditioned on $I(0,y)=1$. For ease of exposition we assume X_i are continuous random variables. We need the following lemma.

Lemma 5. *Let Z^- be the first non-positive partial sum among $\{S_m\}$ and $F(y)$ the distribution of* $\max(0,S_1,\ldots,S_\sigma)$ *where σ is the first passage time to the non-positive axis. Then,*

$$\lim_{y\to\infty} E[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1]$$

$$=\frac{1-E[\exp(\theta^*Z^-)]}{\lim_{y\to\infty}(\exp(\theta^*y)[1-F(y)])}=\frac{1}{e^*}, \tag{28}$$

$0<e^*<\infty$. *Moreover for any $a(y)$,*

$$\lim_{\substack{y\to\infty\\ a(y)\to\infty}} E[\exp(\theta^*(S_{T(-a(y),y)}-y))\,|\,I(-a(y),y)=1]=\frac{1}{e^*} \tag{29}$$

where in the lattice case both y and $a(y)$ traverse the lattice array.

Proof. Let $M(x)$ be the probability distribution function of the global maximum $\max(0,S_1,S_2,\ldots)$. (Owing to $E[X]<0$, $M(x)$ is a bona fide distribution.) Iglehart (1972) established (see also Karlin and Dembo, 1992) that

$$\lim_{y\to\infty}\exp(\theta^*y)[1-M(y)]=e^* \tag{30}$$

exists, with $0<e^*<\infty$, and

$$\lim_{y\to\infty}\exp(\theta^*y)[1-F(y)]=e^*[1-E(\exp(\theta^*Z^-))]. \tag{31}$$

Applying the optional sampling theorem for $T(0, y)$ on the Wald martingale (18) gives

$$\begin{aligned}1 &= E[\exp(\theta^* S_{T(0,y)})]\\ &= [1-F(y)]\exp(\theta^* y)E[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1]\\ &\quad + F(y)E[\exp(\theta^* S_{T(0,y)})\,|\,I(0,y)=0]\end{aligned} \tag{32}$$

and thereby we deduce (since $\{S_{T(0,y)}\,|\,I(0,y)=0\}\to Z^-$)

$$\begin{aligned}&\lim_{y\to\infty} E[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1]\\ &= \lim_{y\to\infty}\frac{1-F(y)E[\exp(\theta^* S_{T(0,y)})\,|\,I(0,y)=0]}{\exp(\theta^* y)[1-F(y)]}\\ &= \frac{1-E[\exp(\theta^* Z^-)]}{\lim_{y\to\infty}\exp(\theta^* y)[1-F(y)]}=\frac{1}{e^*}\end{aligned} \tag{33}$$

(for the last equation use (31)). Proceeding similarly to (33) for the stopping time $T(-a(y), y)$, $y\to\infty$, $a(y)\to\infty$ we deduce

$$\begin{aligned}&\lim_{y\to\infty} E[\exp(\theta^*(S_{T(-a(y),y)}-y))\,|\,I(-a(y),y)=1]\\ &= \frac{1}{\lim_{y\to\infty}\exp(\theta^* y)[1-M(y)]}=\frac{1}{e^*}.\end{aligned}$$

The limit relations of (28) and (29) and their identity are hereby established. Lemma 5 is proved.

When the distribution of X_i is lattice of span δ, we may assume that both y and $a(y)$ in (27)-(29) are lattice points. When restricted to the lattice array, both (30) and (31) hold (see Karlin and Dembo, 1992) and therefore (28) and (29) follow for y and $a(y)$ traversing the lattice array. □

Now let $a(y)=y-\log y$ and y large enough such that $\log y > K$. Note in view of Lemma 5 that the denominator of (15) converges to $1/e^*$ $(0<e^*<\infty)$. Consider now the numerator expression in (15) denoted by $\Gamma(y)$. Let

$$\hat{\Gamma}(y)=E\left[\exp\left(t\sqrt{\frac{w^*}{y}}\,W_{T(0,a(y))}\right)\exp(\theta^*(S_{T(0,y)}-y))\,\middle|\,I(0,y)=1\right]. \tag{34}$$

Lemma 6.

$$\lim_{y\to\infty}\Gamma(y)=\lim_{y\to\infty}\hat{\Gamma}(y).$$

Proof. Abbreviating

$$t\sqrt{\frac{w^*}{y}}[W_{T(0,y)}-W_{T(0,a(y))}] \text{ conditioning on } I(0,y)=1 \quad\text{by } W(y),$$

and

$$\exp\left(t\sqrt{\frac{w^*}{y}}\,W_{T(0,a(y))}\right) \quad \text{by } Z(y),$$

note that since all the terms of $\{U_i\}$ and $\{X_i\}$ are bounded $|W(y)| \leq K|t|\sqrt{w^*/y}\,(T(0,y)-T(0,a(y))) = C(y)$, and hence,

$$|\Gamma(y)-\hat{\Gamma}(y)| \leq \exp(\theta^* K) E[Z(y)(\exp(C(y))-1) \mid I(0,y)=1].$$

Conditioning on the sample realization

$$\omega = \{S_1, S_2, \ldots, S_{T(0,a(y))}, W_1, \ldots, W_{T(0,a(y))}, T(0,a(y)) \text{ and } I(0,y)=1\}$$

yield for (34),

$$\begin{aligned} &E[Z(y)(\exp(C(y))-1) \mid I(0,y)=1] \\ &\quad = E[Z(y)E[\exp(C(y))-1 \mid \omega] \mid I(0,y)=1] \\ &\quad = E\left[Z(y)\int_{a(y)}^{a(y)+K} E\left[\exp\left(K|t|\sqrt{\frac{w^*}{y}}\,T(-\xi,y-\xi)\right)-1 \,\Big|\, \mathscr{E}_{-\xi,y-\xi}^{y-\xi}\right]\right. \\ &\qquad\qquad \left.\times G_y(\mathrm{d}\xi) \,\Big|\, I(0,y)=1\right] \end{aligned}$$

where $G_y(\xi)$ is the distribution function of $S_{T(0,a(y))}$ conditioned on $I(0,y)=1$. Let

$$D(y) = \sup_{a(y)\leq\xi\leq a(y)+K} E\left[\exp\left(K|t|\sqrt{\frac{w^*}{y}}\,T(-\xi,y-\xi)\right)-1 \,\Big|\, \mathscr{E}_{-\xi,y-\xi}^{y-\xi}\right].$$

Then, by the above $|\Gamma(y)-\hat{\Gamma}(y)| \leq \exp(\theta^* K) D(y)\hat{\Gamma}(y)$, and since $\lim_{y\to\infty}\Gamma(y)$ exists and is finite, the proof is complete by showing that $D(y)\to 0$. Clearly,

$$D(y) \leq (\exp(Cy^{-1/4})-1) + \sup_{a(y)\leq\xi\leq a(y)+K} \int_{y^{1/4}}^{\infty} \exp\left(\frac{Cx}{\sqrt{y}}\right) Q_{\xi,y}(\mathrm{d}x)$$

where $C = K|t|\sqrt{w^*}$, $Q_{\xi,y}(x)$ is the distribution function of $T(-\xi, y-\xi)$ conditioned on $\mathscr{E}_{-\xi,y-\xi}^{y-\xi}$. With the first term above converging to zero, as $y\to\infty$, we concentrate on estimating the tail behavior of $[1-Q_{\xi,y}(x)]$. Since $y-a(y)=\log y \geq K$ for all y large enough, we have

$$\{T(-\xi,y-\xi)\geq x \text{ and } \mathscr{E}_{-\xi,y-\xi}^{y-\xi}\} \subset \bigcup_{n\geq x} \{S_n \geq 0\}.$$

Hence, by the union of events bound and the inclusion $\mathscr{E}_{-a,b}^{b} \supset \mathscr{E}_{0,b}^{b}$,

$$1-Q_{\xi,y}(x) = P(T(-\xi,y-\xi)\geq x \mid \mathscr{E}_{-\xi,y-\xi}^{y-\xi}) \leq \frac{\sum_{n=x}^{\infty} P(S_n\geq 0)}{P(I(0,y-\xi)=1)}.$$

Combining the large deviations bound $P(S_n \geq 0) \leq \exp(-nI(0))$, $I(0)>0$ with the basic lower bound $P(I(0,b)=1) \geq \delta\exp(-\theta^* b)$ (see Dembo and Karlin, 1991a), it follows that for some $C_1>0$ independent of y,

$$\sup_{a(y)\leq\xi\leq a(y)+K} [1-Q_{\xi,y}(x)] \leq C_1 \exp(\theta^* \log y)\exp(-I(0)x).$$

Thus, integrating by parts we get for some $C_2, C_3 > 0$ and all y large enough

$$\int_{y^{1/4}}^{\infty} \exp(Cx/\sqrt{y})Q_{\xi,y}(\mathrm{d}x) = \exp(Cy^{-1/4})(1-Q_{\xi,y}(y^{1/4})) + \frac{C}{\sqrt{y}}\int_{y^{1/4}}^{\infty} \exp\left(\frac{Cx}{\sqrt{y}}\right)(1-Q_{\xi,y}(x))\,\mathrm{d}x$$
$$\leq C_2 \exp(Cy^{-1/4})y^{\theta^*}\exp(-I(0)y^{1/4}) \leq C_3 \exp(-\tfrac{1}{2}I(0)y^{1/4}).$$

With the latter bound converging to zero as $y\to\infty$, so does $D(y)$, and the proof of Lemma 6 is complete. □

By a similar analysis,

$$\hat{\Gamma}(y) = E\left[\exp\left(t\sqrt{\frac{w^*}{y}}\, W_{T(0,a(y))}\right)E[\exp(\theta^*(S_{T(0,y)}-y))\,|\,\omega]\,\middle|\, I(0,y)=1\right]$$
$$= E\left[\exp\left(t\sqrt{\frac{w^*}{y}}\, W_{T(0,a(y))}\right) \times \int_{a(y)}^{a(y)+K} (E_\xi[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1])G_y(\mathrm{d}\xi)\,\middle|\, I(0,y)=1\right] \tag{35}$$

where $G_y(\xi)$ is the distribution function of $S_{T(0,a(y))}$ conditional on $I(0,y)=1$ and E_ξ denotes the expectation starting at ξ.

Clearly, $a(y)\leq\xi\leq a(y)+K<y$ since $K<\log y$. Moreover,

$$\int_{a(y)}^{a(y)+K} E_\xi[\exp(\theta^*(S_{T(0,y)}-y))\,|\,I(0,y)=1]G_y(\mathrm{d}\xi)$$
$$= \int_{a(y)}^{a(y)+K} E_0[\exp(\theta^*(S_{T(-\xi,y-\xi)}-(y-\xi)))\,|\,I(-\xi,y-\xi)=1]G_y(\mathrm{d}\xi). \tag{36}$$

As $y\uparrow\infty$, Lemma 5 applies uniformly for all $a(y)<\xi<a(y)+K$ indicating that

$$E_0[\exp(\theta^*(S_{T(-\xi,y-\xi)}-(y-\xi)))\,|\,I(-\xi,y-\xi)=1]\to 1/e^*.$$

Note that when X is a lattice random variable, $G_y(\mathrm{d}\xi)$ concentrates on lattice points and Lemma 5 applies in this context.

We can add back the terms $\exp(W(y))$ into (35), by an argument paraphrasing the proof of Lemma 6, yielding

$$\lim_{y\to\infty}\Gamma(y) = \lim_{y\to\infty} E\left[\exp\left(t\sqrt{\frac{w^*}{y}}\, W_{T(0,y)}\right)\,\middle|\, I(0,y)=1\right]\frac{1}{e^*}. \tag{37}$$

Combining (37) and (28) we get the assertion of Lemma 2. □

The vector version of Theorem 1 is proved as follows. For each real vector z and vector sequence $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots$ described in the introductory section, we form the sequence of real random variables

$$U_i = \langle z, \boldsymbol{U}_i - \boldsymbol{u}^* \rangle = \sum_k z_k (U_{ik} - u_k^*), \quad i = 1, 2, \ldots . \tag{38}$$

With

$$\rho(\theta, t) = \log E[\exp\{\theta X_i + t\langle z, \boldsymbol{U} - \boldsymbol{u}^* \rangle\}]$$

we see that

$$\frac{\partial \rho}{\partial \theta}(\theta^*, 0) = w^*, \qquad \frac{\partial \rho}{\partial t}(\theta^*, 0) = 0,$$

$$\frac{\partial^2 \rho}{\partial t^2}(\theta^*, 0) = E[(\langle z, \boldsymbol{U} - \boldsymbol{u}^* \rangle)^2 \exp(\theta^* X)].$$

The conditional central limit theorem is for the partial sums of $U_i = \langle z, \boldsymbol{U}_i - \boldsymbol{u}^* \rangle$. This works for each real vector z. Theorem 1 in the vector version follows standardly. □

3. Proof of Theorem 2

To ease the exposition we assume $\boldsymbol{P}$ is a strictly positive matrix and for all pairs of states α, β the strict inequalities $\Pr\{X_{\alpha\beta} > 0\} > 0$ and $\Pr\{X_{\alpha\beta} < 0\} > 0$ hold. The multivariate version can be reduced to the univariate case with $u^* = 0$ following the same argument given for the i.i.d. case. Therefore, Theorem 2 holds provided (17) holds with the left side conditioned on $\alpha_0 = i$, for any i. Here the corresponding Wald martingale family (given α_0) is

$$Q_n(\theta, t) = \exp\{\theta S_n + t W_n - n\rho(\theta, t)\}\psi_{\alpha_n}(\theta, t)/\psi_{\alpha_0}(\theta, t) \tag{39}$$

with $\boldsymbol{\psi} = (\psi_i(\theta, t))$ the unique right eigenvector normalized to $\sum_i \psi_i = 1$, corresponding to the spectral radius of the matrix $\boldsymbol{Q}(\theta, t)$ defined after (10).

The expansion (21) of $\theta(t)$ near θ^* remains valid, as well as Lemma 3 which exploits the corresponding large deviation estimate for the Markov additive process $\{\alpha_n, S_n\}$.

The ratio $\psi_j(\theta, t)/\psi_i(\theta, t)$ is bounded away from zero and infinity uniformly with respect to i, j and (θ, t) around $(\theta^*, 0)$. Further $\boldsymbol{\psi}(\theta, t)$ is analytic in θ and t so that

$$\sup_\alpha \left| \log \frac{\psi_\alpha(\theta(t_y), t_y)}{\psi_\alpha(\theta^*, 0)} \right| \leqslant O(|\theta(t_y) - \theta^*| + |t_y|) \leqslant C/\sqrt{y}. \tag{40}$$

Therefore, applying the optional sampling theorem and paraphrasing the arguments of Lemma 4 we secure the analog of Lemma 1, involving now the additional multiplying factor $\psi_{\alpha_{T(0,y)}}(\theta^*,0)$ and the conditioning on $\alpha_0=i$ in both numerator and denominator of (15).

To establish (17) we rely on the property that $\exp(\theta^*y)\Pr\{I(0,y)=1\,|\,\alpha_0=i\}$ is bounded away from zero for all i; this fact was proved in Dembo and Karlin (1991b).

We restate next Lemma 5 for the M.C. case.

Lemma 5′. *Let* $M_i(x)$ *and* $F_i(x)$ *be the probability distribution functions of* $\max(0,S_1,S_2,\ldots)$ *and* $\max(0,S_1,\ldots,S_\sigma)$ *respectively conditioned in* $\alpha_0=i$ *where* σ *is the first passage time to the non-positive axis and* $Z^-=S_\sigma$. *Then*

$$\lim_{y\to\infty} E[\exp(\theta^*(S_{T(0,y)}-y))\psi_{\alpha_{T(0,y)}}(\theta^*,0)\,|\,I(0,y)=1,\alpha_0=i]$$
$$=\frac{\psi_i(\theta^*,0)-E[\exp(\theta^*Z^-)\psi_{\alpha_\sigma}(\theta^*,0)\,|\,\alpha_0=i]}{\lim_{y\to\infty}(\exp(\theta^*y)[1-F_i(y)])}=\frac{1}{e^*},\tag{41}$$

$0<e^*<\infty$ *and for any* $a(y)$,

$$\lim_{\substack{y\to\infty\\ a(y)\to\infty}} E[\exp(\theta^*(S_{T(-a(y),y)}-y))\psi_{\alpha_{T(-a(y),y)}}(\theta^*,0)\,|\,I(-a(y),y)=1,\ \alpha_0=i]$$
$$=\frac{\psi_i(\theta^*,0)}{\lim_{y\to\infty}\exp(\theta^*y)(1-M_i(y))}=\frac{1}{e^*}.\tag{42}$$

The proof parallels the i.i.d. case. The limits (30) and (31) are replaced by their M.C. analogues, namely

$$\lim_{y\to\infty}\exp(\theta^*y)[1-M_i(y)]=e^*\psi_i(\theta^*,0)\tag{43}$$

and

$$\lim_{y\to\infty}\exp(\theta^*y)[1-F_i(y)]$$
$$=(\psi_i(\theta^*,0)-E[\exp(\theta^*Z^-)\psi_{\alpha_\sigma}(\theta^*,0)\,|\,\alpha_0=i])e^*,\tag{44}$$

which are established in Karlin and Dembo (1992).

The proof of Theorem 2 is completed by a similar argument to that elaborated for the i.i.d. case. Here the relevant sample realization for the conditioning operation is $\omega=\{S_1,\ldots,S_{T(0,a(y))},W_1,\ldots,W_{T(0,a(y))},\alpha_0,\alpha_1,\ldots,\alpha_{T(0,a(y))}$ and $T(0,a(y))$ and $I(0,y)=1\}$. So, with $G_{y,i}(\xi,j)$ being the joint distribution of $S_{T(0,a(y))}$ and $\alpha_{T(0,a(y))}$ conditioned on $\alpha_0=i$ and on $I(0,y)=1$, we obtain instead of (35) here

$$\Gamma(y)\approx E\Bigg[\exp\bigg(t\sqrt{\frac{w^*}{y}}\,W_{T(0,a(y))}\bigg)\times\sum_j\int_{a(y)}^{a(y)+K}E[\exp(\theta^*(S_{T(-\xi,y-\xi)}-(y-\xi)))\psi_{\alpha_{T(-\xi,y-\xi)}}(\theta^*,0)\,|\,I(-\xi,y-\xi)=1,\alpha_0=j]G_{y,i}(\mathrm{d}\xi,j)\,\Bigg|\,I(0,y)=1,\alpha_0=i\Bigg].\tag{45}$$

The proof then follows by exploiting (42), (45) and the constancy of e^* independent of the initial state α_0. $\square$

When $\Pr\{X_{i,j} \leq x \mid \alpha_0 = i, \alpha_1 = j\}$ are arithmetic distributions of a common span δ. Lemma 5′ (i.e. (41)-(44)) holds when y and $a(y)$ are restricted to be lattice points.

The assumption (9) can be replaced by the weaker condition that (9) need not hold for all states but does apply to the initial state α which is in the support of positive ladder states of the Markov additive process $\{\alpha_n, S_n\}$. In this context $\alpha_{T(0,a(y))}$ is a positive ladder state.

It is likely that boundedness of $X_{\alpha\beta}$ and $U_{\alpha\beta}$ may be relaxed maintaining Theorem 1 as long as $\rho(\theta, t)$ exists in the neighborhood of $(\theta^*, 0)$.

The finite state character of the M.C. governing $\alpha_1, \ldots, \alpha_n$ was simply relied on to ensure proper boundedness of the eigenvector $\boldsymbol{\psi}(\theta, t)$, for example, for purposes of securing (40) and in the uniformity (with respect to i) of the convergence in (41) and (42). For the general (non-finite) M.C., the results of Iscoe et al. (1985) and Ney and Nummelin (1978a, b) can help to gain the generalization of Theorem 1, postulating $\{\alpha_n, S_n\}$ as a uniformly recurrent Markov Additive Process.

References

V. Anatharam, How large delays build up in a GI/G/1 queue, Queueing Syst. 5 (1988) 345-368.

S. Asmussen, Conditional limit theorems relating the random walk to its associate, with applications to risk processes and the GI/G/1 queue, Adv. Appl. Probab. 14 (1982) 143-170.

H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Statist. 23 (1952) 493-507.

A. Dembo and S. Karlin, Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables, Ann. Probab. 19(4) (1991a) 1737-1755.

A. Dembo and S. Karlin, Strong limit theorems of empirical distributions for large segmental exceedances of partial sums of Markov variables, Ann. Probab. 19(4) (1991b) 1756-1767.

D. Iglehart, Extreme values in the GI/G/1 queue. Ann. Math. Statist. 43 (1972) 627-635.

D. Iglehart, Random walk with negative drift conditioned to stay positive, Ann. Probab. 2 (1974) 608-619.

I. Iscoe, P. Ney and E. Nummelin, Large deviations of uniformly recurrent Markov-additive processes, Adv. Appl. Math. 6 (1985) 373-412.

P. Kao, Limiting diffusion for random walks with drift conditioned to stay positive, J. Appl. Probab. 15 (1978) 280-291.

S. Karlin and S.F. Altschul, New methods for assessing statistical significant molecular sequence features using general scoring schemes, Proc. Nat. Acad. Sci. USA 87 (1990) 2264-2268.

S. Karlin and A. Dembo, Limit distributions of maximal segmental score among Markov dependent partial sums, Adv. Appl. Probab. 24 (1992) 113-140.

S. Karlin, A. Dembo and T. Kawabata, Statistical composition of high-scoring segments from molecular sequences, Ann. Statist. 18 (1990) 568-577.

P. Ney and E. Nummelin, Markov additive processes. I. Eigenvalues, properties and limit theorems, Ann. Probab. 15 (1987a) 561-582.

P. Ney and E. Nummelin, Markov additive processes. II. Large deviations, Ann. Probab. 15 (1987b) 593-609.

D. Siegmund, The time until ruin in collective risk theory, Mitt. Verein. Schweiz. Versicherungsmath. 1 (1975) 157-166.

S.R.S. Varadhan, Large deviations and applications, SIAM J. CBMS Ser. (1984).