Submitted to the Annals of Applied Probability

CLUSTERING OF LARGE DEVIATIONS IN MOVING AVERAGE PROCESSES: THE SHORT MEMORY REGIME

By Arijit Chakrabarty and Gennady Samorodnitsky *

Indian Statistical Institute, Kolkata and Cornell University

We describe the cluster of large deviations events that arise when one such large deviations event occurs. We work in the framework of an infinite moving average process with a noise that has finite exponential moments.

1. Introduction.

A very crude classification of how we analyse random systems might split the work into distributional analysis and large deviations analysis. The distributional analysis deals with the "usual" deviations of a system from its "average" state, while the large deviations analysis deals the "unusually large" deviations, that are, by necessity, rare (but may have a major impact). Separately, the idea of clustering is also a major idea in how we look at random systems. Clustering typically means that certain related events occur "in proximity to each other" and, when it happens, the impact of the events may be magnified. Clustering is interesting because it may shed light on certain structural elements in a random system. Clustering is most frequently studied in distributional analysis; an important example is clustering of extreme values; see e.g. Embrechts et al. (2003).

In this work we are interested in clustering of large deviations events. From a different point of view, we would like to understand whether or not a (rare) large deviations event is likely to cause a cascade of additional large deviations events and, if so, what does this cascade look like. Literature on large deviations analysis is vast, and the nature of large deviations turns out to be different in stochastic systems with "light tails" and with "heavy tails". The texts such as Dembo and Zeitouni (1998) or Deuschel and Stroock (1989) describe large deviations of light-tailed systems, while Mikosch and Nagaev (1998) will give the reader an idea how large deviations occur in heavy-tailed systems. Large deviations are affected not only by the "tails" in a random system, but also the "memory" in that system, in particular by whether the memory is "short" or "long". The change from

*The corresponding author. Research partially supported by NSF grant DMS-2015242 and AFOSR grant FA9550-22-1-0091 at Cornell University.

AMS 2000 subject classifications: Primary 60F10

Keywords and phrases: large deviations, clustering, infinite moving average

short to long memory in a system can be viewed as a phase transition (see Samorodnitsky (2016)), and it affects large deviations as well. In this work we study clustering of large deviations in a light-tailed system, but we will consider both short memory and long memory situations.

Let us now be more specific about the class of stochastic models we will consider. We will consider centred infinite moving average processes

$$X_n = \sum_{i=0}^{\infty} a_i Z_{n-i}\,,\ n \geq 0\,, \tag{1.1}$$

where $(Z_n : n \in \mathbb{Z})$ is a collection of i.i.d. non-degenerate random variables (*the noise*) with distribution F_Z satisfying

$$\int_{\mathbb{R}} e^{tz}\, F_Z(dz) < \infty \text{ for all } t \in \mathbb{R}\,, \tag{1.2}$$

and

$$\int_{\mathbb{R}} z\, F_Z(dz) = 0\,. \tag{1.3}$$

For future reference we denote

$$\sigma_Z^2 = \int_{\mathbb{R}} z^2\, F_Z(dz). \tag{1.4}$$

Let $a_0, a_1, a_2 \ldots$ be real numbers satisfying

$$\sum_{j=0}^{\infty} a_j^2 < \infty\,. \tag{1.5}$$

Since the assumption (1.2) implies that the noise variables have a finite second moment, the zero mean property assumed in (1.3) and the square integrability of the coefficients (1.5) imply that the infinite sum in the right hand side of (1.1) converges in L^2 and a.s. and defines a zero mean stationary ergodic process. Therefore, for $\varepsilon > 0$ the event

$$E_0(n, \varepsilon) = \left\{ \frac{1}{n} \sum_{i=0}^{n-1} X_i \geq \varepsilon \right\}$$

is, for large n, a rare, large deviations, event. We would like to understand whether occurrence of this event may cause a cascade of related events. Specifically, for $j \geq 0$ we denote

$$E_j(n, \varepsilon) = \left\{ \frac{1}{n} \sum_{i=j}^{n+j-1} X_i \geq \varepsilon \right\}\,, \tag{1.6}$$

so that each event $E_j(n,\varepsilon)$ is equally rare, and we would like to know how many of the events for j "reasonably close to $j=0$" occur if $E_0(n,\varepsilon)$ occurs (the reason for the qualifier "reasonably close to $j=0$" is that by ergodicity, the events $E_j(n,\varepsilon)$ will keep recurring eventually, regardless of the structure of the system).

The difference between short memory infinite moving average processes and long memory infinite moving average processes lies in the rate the coefficients (a_n) converge to zero (subject to the square summability, of course). This will lead to markedly different cascading of the events $E_j(n,\varepsilon)$, conditionally on the event $E_0(n,\varepsilon)$ occurring. In this paper we consider the short memory memory processes, while leaving the long memory case to another paper. Specifically, we describe the limiting distribution of the large deviation cluster caused by the rare event $E_0(n,\varepsilon)$ as well as the behaviour of the size of that cluster as the overshoot ε becomes small. It turns out that for such ε the size of the cluster is of the order ε^{-2}.

Our main results on the cluster of large deviations for short memory infinite moving average processes are in Section 2. The concluding Section 3 contains a discussion of the results we obtain and connects them to what one may expect when the memory becomes long.

2. Short memory moving average processes. We follow the common terminology and say that the infinite moving average process (1.1) has short memory if

$$\sum_{n=0}^{\infty} |a_n| < \infty \quad \text{and} \quad \sum_{n=0}^{\infty} a_n \neq 0. \tag{2.7}$$

We investigate the clustering of the rare events $(E_j(n,\varepsilon))$ in the following way. We will show that the conditional law of the process of occurrences of the large deviation events,

$$\big(\mathbf{1}(E_j(n,\varepsilon),\, j=1,2,\ldots)\big), \tag{2.8}$$

given $E_0(n,\varepsilon)$ has a non-degenerate weak limit and describe that limit. This will show, in particular, that for a fixed $K\in\mathbb{Z}$ the conditional law of the total number of occurrences among the first K of the events $(E_j(n,\varepsilon))$

$$\nu_n(K,\varepsilon)(\cdot) = P\left(\sum_{j=1}^{K} \mathbf{1}(E_j(n,\varepsilon)) \in \cdot \,\middle|\, E_0(n,\varepsilon)\right) \tag{2.9}$$

has a weak limit. That weak limit itself converges weakly, as $K\to\infty$, to an a.s. finite random variable that we interpret as the size of the cluster of large deviation events caused by a large deviation event at time zero. An interesting regime is that of a small $\varepsilon > 0$, and we show a properly normalized size of the cluster of large deviation events converges weakly, as $\varepsilon \to 0$, to an (interesting) limit. As we have explained above, the limits should be taken in this specific order.

To state the main results of this section we need to introduce some notation first. Denote

$$A = \sum_{n=0}^{\infty} a_n. \tag{2.10}$$

In the sequel we will assume that $A > 0$. Note that, in view of (2.7), this introduces no real loss of generality because, if the sum is negative, we simply multiply both (Z_n) and (a_n) by -1 and reduce the situation to the case $A > 0$ we are considering. Further, for $n = 0, 1, 2, \dots$ we write

$$A_n = \sum_{j=0}^{n} a_j \,. \tag{2.11}$$

Next, we let

$$\varphi_Z(t) = \log\left(\int_{\mathbb{R}} e^{tz}\, F_Z(dz)\right),\ t \in \mathbb{R} \tag{2.12}$$

to be the log-Laplace transform of a noise variable. For $\theta \in \mathbb{R}$ we denote by G_θ the probability measure on $\mathbb{R}$ obtained by exponentially tilting F_Z as follows:

$$G_\theta(dx) = e^{-\varphi_Z(\theta)+\theta x} F_Z(dx),\ x \in \mathbb{R}\,. \tag{2.13}$$

Further, let

$$s_0 = \sup\{x \in \mathbb{R} : F_Z(x) < 1\} \in (0, \infty]$$

be the right endpoint of the support of a noise variable. We will consider the events $(E_j(n, \varepsilon))$ for ε satisfying

$$0 < \varepsilon/A < s_0\,. \tag{2.14}$$

The function φ_Z is infinitely differentiable, and its first derivative φ_Z' strictly increases from 0 at $t = 0$ to s_0 as $t \to \infty$. Therefore, for ε satisfying (2.14), we can unambiguously define $\tau(\varepsilon) > 0$ by

$$\varphi_Z'\left(\tau(\varepsilon)\right) = \varepsilon/A\,. \tag{2.15}$$

We now introduce a collection $\{Z_j^u : j \in \mathbb{Z},\, u = + \text{ or } -\}$ of independent random variables with the following laws:

$$
\begin{aligned}
&Z_{-j}^- \sim G_{(1-A^{-1}A_{j-1})\tau(\varepsilon)}\,, && j \geq 1\,,\\
&Z_j^- \sim G_{\tau(\varepsilon)}\,, && j \geq 0\,,\\
&Z_{-j}^+ \sim G_{A^{-1}A_{j-1}\tau(\varepsilon)}\,, && j \geq 1\,,\\
&Z_j^+ \sim F_Z\,, && j \geq 0\,.
\end{aligned}
\tag{2.16}
$$

Finally, let T^* be an exponential random variable with parameter $\tau(\varepsilon)/A$, independent of the family (2.16).

It is elementary to check that for any $j \geq 0$ and $u = +$ or $-$,

$$E\left(|Z_j^u|\right) \leq \int_{\mathbb{R}} |x|\, G_{A^{-1}\bar{A}\tau(\varepsilon)}(dx) \, + \int_{\mathbb{R}} |x|\, G_{-A^{-1}\bar{A}\tau(\varepsilon)}(dx) < \infty\,,$$

where

$$\bar{A} = \sum_{n=0}^{\infty} |a_n|.$$

Therefore, the infinite series

$$
\begin{aligned}
U_n^- &= \sum_{i=0}^{\infty} a_i Z_{n-i}^-,\ n \geq 0\,,\\
U_n^+ &= \sum_{i=0}^{\infty} a_i Z_{n-i}^+,\ n \geq 0
\end{aligned}
\tag{2.17}
$$

converge in L^1 and a.s. Finally, we define

$$V_j(\varepsilon) = \mathbf{1}\left(T^* \geq \sum_{i=0}^{j-1}(U_i^- - U_i^+)\right),\ j \geq 1. \tag{2.18}$$

We are now ready to state the main theorems of this section. They rely on a technical assumption, excluding the case of a lattice-valued noise. We assume that

$$\left|\int_{\mathbb{R}} e^{itz}\, F_Z(dz)\right| < 1 \ \text{ for any } t \neq 0 \text{ where } i = \sqrt{-1}. \tag{2.19}$$

Our first result describes the behaviour of the sequence of conditional laws of the process of occurrences of the large deviation events (2.8) and of the sequence (2.9) of the total number of occurrences among the first K of the events $(E_j(n,\varepsilon))$.

THEOREM 1. *Assume that* (2.7) *holds and* $A > 0$ *in* (2.10). *Assume, further, that the characteristic function of the noise variables satisfies* (2.19). *Let* ε *be as in* (2.14). *Then, as* $n \to \infty$,

$$P\Big(\big(\mathbf{1}(E_j(n,\varepsilon),\, j=1,2,\ldots)\big) \in \cdot \Big| E_0(n,\varepsilon)\Big) \Rightarrow P\Big(\big(V_j(\varepsilon),\, j=1,2,\ldots\big) \in \cdot\Big) \tag{2.20}$$

in $\{0,1\}^\infty$. *In particular, for every fixed* $K \geq 1$, *the conditional laws* $(\nu_n(K,\varepsilon))$ *in* (2.9) *satisfy*

$$\nu_n(K,\varepsilon)(\cdot) \Rightarrow P\left(\sum_{j=1}^{K} V_j(\varepsilon) \in \cdot\right) \quad \text{as } n \to \infty\,. \tag{2.21}$$

It is natural to interpret the statement of Theorem 1 as saying that a large deviation event $E_0(n,\varepsilon)$, upon occurring, leads to a random cluster of large deviation events, and the limiting (as $n \to \infty$) total size of this cluster has the law of

$$D_\varepsilon = \sum_{j=1}^{\infty} V_j(\varepsilon),\ \varepsilon > 0\,. \tag{2.22}$$

Our second result of this section shows that this total cluster size is a.s. finite and describes its limiting behaviour as the overshoot ε becomes small.

THEOREM 2. *Under the assumptions of Theorem 1, the total cluster size* D_ε *is a.s. finite. Further, as* $\varepsilon \to 0$,

$$\varepsilon^2 D_\varepsilon \Rightarrow A^2\sigma_Z^2 \int_0^\infty \mathbf{1}\left(T_0 \geq (\sqrt{2}B_t + t)\right) dt,$$

where A *is the sum of the coefficients* (2.10) *and* σ_Z^2 *is the noise variance* (1.4). *Furthermore,* T_0 *is a standard exponential random variable independent of a standard Brownian motion* $(B_t : t \geq 0)$.

We prove Theorem 1 first, and so $\varepsilon > 0$ (satisfying (2.14)) is for now fixed. The proof is via several lemmas. To simplify the notation we will write E_j instead of $E_j(n,\varepsilon)$ throughout.

LEMMA 2.1. *Denote*

$$S_n = \sum_{i=0}^{n-1} X_i,\ n \geq 1\,, \tag{2.23}$$

and let

$$\psi_n(t) = n^{-1} \log E\left(e^{tS_n}\right),\ t \in \mathbb{R},\ n \geq 1\,. \tag{2.24}$$

Then for all large enough n there exists a unique $\theta_n > 0$ such that

$$\psi_n'(\theta_n) = \varepsilon\,.$$

Furthermore,

$$P(E_0) \sim \frac{C}{\sqrt{n}} \exp\Big(-n\big(\theta_n\varepsilon - \psi_n(\theta_n)\big)\Big),\ n \to \infty\,,$$

with

$$C = \frac{1}{\tau(\varepsilon)\sqrt{2\pi\varphi_Z''(\tau(\varepsilon))}}\,,$$

and φ_Z and $\tau(\varepsilon)$ defined, respectively, in (2.12) *and* (2.15).

Proof. We write

$$S_n = \sum_{j=0}^{n-1} A_{n-1-j} Z_j + \sum_{j=1}^{\infty} (A_{j+n-1} - A_{j-1}) Z_{-j}\,, \tag{2.25}$$

with A_n defined by (2.11) and check the conditions of Theorem 3 in the appendix. As a first step we show that

$$\lim_{n\to\infty} \psi_n''(t) = A^2 \varphi_Z''(At) \tag{2.26}$$

locally uniformly in $t \in \mathbb{R}$. Indeed, by (2.25),

$$\psi_n(t) = \frac{1}{n}\left[\sum_{j=0}^{n-1} \varphi_Z(A_j t) + \sum_{j=1}^{\infty} \varphi_Z\left((A_{j+n-1} - A_{j-1})t\right)\right],$$

and taking the Cesaro limits shows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} A_j^2 \varphi_Z''(A_j t) = A^2 \varphi_Z''(At)\,.$$

Furthermore, since φ_Z'' is locally bounded,

$$\begin{aligned}\sum_{j=1}^{\infty}(A_{j+n-1}-A_{j-1})^2\varphi_Z''\left((A_{j+n-1}-A_{j-1})t\right) &= O\left(\sum_{j=1}^{\infty}(A_{j+n-1}-A_{j-1})^2\right)\\ &= O\left(\sum_{j=1}^{\infty}\sum_{i=j}^{j+n-1}|a_i|\right)\\ &= O\left(\sum_{i=1}^{n}\sum_{j=i}^{\infty}|a_j|\right) = o(n),\ n\to\infty\,.\end{aligned}$$

Since all these steps are locally uniform in $t \in \mathbb{R}$, (2.26) follows. Since this argument also shows that ψ_n'' is, uniformly in n, locally bounded, and the values of ψ_n, φ_Z and their respective first derivatives at 0 are 0, we also conclude that ψ_n' is, uniformly in n, locally bounded and for every $t \geq 0$,

$$\lim_{n\to\infty}\psi_n'(t) = A\varphi_Z'(At)\,, \tag{2.27}$$

$$\lim_{n\to\infty}\psi_n(t) = \varphi_Z(At)\,. \tag{2.28}$$

The assumption (2.14) together with (2.27) implies that for large n there exists a unique $\theta_n > 0$ such that $\psi_n'(\theta_n) = \varepsilon$, and that

$$\lim_{n\to\infty}\theta_n = A^{-1}\tau(\varepsilon)\,. \tag{2.29}$$

We choose n_1 so large that for $n \geq n_1$, θ_n is well defined, $A/2 \leq A_n \leq \sqrt{2}A$ and $\theta_n \leq \sqrt{2}\tau(\varepsilon)/A$.

We claim next that for fixed $\delta, \lambda > 0$ there exists $\eta \in (0,1)$ such that

$$\sup_{\delta\leq|t|\leq\lambda\theta_n}\left|\frac{1}{E(e^{\theta_n S_n})}E\left(e^{(\theta_n+it)S_n}\right)\right| = O\left(\eta^n\right),\ n\to\infty, \tag{2.30}$$

with the convention that the supremum of the empty set is zero. To see this, note that $\phi : \mathbb{R}^2 \to \mathbb{C}$ defined by

$$\phi(\theta,t) = \frac{1}{E(e^{\theta Z})}E\left(e^{(\theta+it)Z}\right)\,,$$

is continuous. For a fixed $\theta \in \mathbb{R}$, $\phi(\theta,\cdot)$ is the characteristic function of the distribution G_θ in (2.13). By (2.19), G_θ is not a lattice distribution and, hence, for any fixed $\lambda, \delta > 0$ and θ,

$$\sup_{A\delta/2\leq|t|\leq 2\lambda\tau(\varepsilon)}|\phi(\theta,t)| < 1\,.$$

A standard compactness argument and (2.29) imply that

$$\eta := \sup_{n\geq n_1,\, A\delta/2\leq |t|\leq 2\lambda\tau(\varepsilon)} |\phi(\theta_n, t)| < 1\,,$$

while the choice of n_1 implies that for $n > j \geq n_1$ and $\delta \leq |t| \leq \lambda\theta_n$,

$$A\delta/2 \leq |A_j t| \leq 2\lambda\tau(\varepsilon)\,.$$

Therefore, by (2.25) and the triangle inequality,

$$\frac{1}{E(e^{\theta_n S_n})}\left|E(e^{(\theta_n+it)S_n})\right| \leq \prod_{j=n_1}^{n-1} |\phi(\theta_n, A_j t)| \leq \eta^{n-n_1}\,, \tag{2.31}$$

establishing (2.30).

We have now verified all conditions of Theorem 3 for $T_n = S_n$, $a_n = n$, $m_n = \varepsilon$ and $\tau_n = \theta_n$, and (1.63) gives us the statement of the lemma. □

We proceed with showing uniform boundedness of conditional moments of all noise variables.

Lemma 2.2. *We have*

$$\sup_{n\geq 1,\, j\in\mathbb{Z}} E\left(|Z_j|\big|E_0\right) < \infty\,. \tag{2.32}$$

Proof. Fix $j \in \mathbb{Z}$ and define

$$S_{n,j} = S_n - \beta_{n,j} Z_j,\ n \geq 1,$$

where

$$\beta_{n,j} = \begin{cases} 0, & 1 \leq n \leq j \\ A_{n-1-j}, & n \geq j+1 \end{cases}$$

if $j \geq 0$ and

$$\beta_{n,j} = A_{n-1-j} - A_{-j-1},\ n \geq 1$$

if $j \leq -1$, with S_n is as in (2.23). It follows from (2.25) that $S_{n,j}$ and Z_j are independent. We define

$$\tilde{\psi}_{n,j}(t) = n^{-1}\log E\left(e^{tS_{n,j}}\right),\ t \in \mathbb{R}\,. \tag{2.33}$$

Since the numbers $(\beta_{n,j})$ are bounded uniformly in j and n, it follows that the functions in (2.24) and (2.33) satisfy

$$\psi_n'(\theta) = \tilde{\psi}_{n,j}'(\theta) + O(1/n)\,,$$

with $O(1/n)$ uniform over j and θ in a compact interval. The same argument as in Lemma 2.1 shows that for large n there exists a unique $\tilde{\theta}_{n,j} > 0$ such that

$$\tilde{\psi}'_{n,j}(\tilde{\theta}_{n,j}) = \varepsilon\,.$$

Since φ'_Z is locally bounded away from zero, it follows from (2.27) that

$$\tilde{\theta}_{n,j} = \theta_n + O(1/n) \tag{2.34}$$

with $O(1/n)$ uniform over j. This also implies that

$$\psi_n(\tilde{\theta}_{n,j}) = \psi_n(\theta_n) + O(1/n)\,. \tag{2.35}$$

For large n we can write

$$\begin{aligned} E\left(|Z_j|\mathbf{1}(E_0)\right) &= \int_{\mathbb{R}} |z| P(Z_j \in dz) \int_{\mathbb{R}} \mathbf{1}(s + \beta_{n,j} z \geq n\varepsilon) P(S_{n,j} \in ds) \\ &= \exp\left\{-n\left(\tilde{\theta}_{n,j}\varepsilon - \psi_n(\tilde{\theta}_{n,j})\right)\right\} \int_{\mathbb{R}} |z| e^{-\tilde{\theta}_{n,j}\beta_{n,j} z} P(Z^*_{n,j} \in dz) \\ &\quad \int_{[n\varepsilon - \beta_{n,j}z,\infty)} \exp\left(-\tilde{\theta}_{n,j}(s - n\varepsilon)\right) P(S^*_{n,j} \in ds)\,, \end{aligned}$$

where $S^*_{n,j}$ and $Z^*_{n,j}$ are independent random variables with $Z^*_{n,j}$ having distribution $G_{\tilde{\theta}_{n,j}\beta_{n,j}}$ and

$$P(S^*_{n,j} \in ds) = \frac{1}{E\left(\exp(\tilde{\theta}_{n,j} S_{n,j})\right)} e^{\tilde{\theta}_{n,j} s} P(S_{n,j} \in ds)\,.$$

It follows from (2.34) and (2.35) that, uniformly over j,

$$\begin{aligned} \exp\left\{-n\left(\tilde{\theta}_{n,j}\varepsilon - \psi_n(\tilde{\theta}_{n,j})\right)\right\} &= O\left(\exp\left(-n\big(\theta_n\varepsilon - \psi_n(\theta_n)\big)\right)\right) \\ &= O\left(\sqrt{n} P(E_0)\right)\,, \end{aligned}$$

with the second line implied by Lemma 2.1. Therefore, to complete the proof it suffices to show that, uniformly in j,

$$\begin{aligned} &\int_{\mathbb{R}} |z| e^{-\tilde{\theta}_{n,j}\beta_{n,j} z} P(Z^*_{n,j} \in dz) \int_{[n\varepsilon - \beta_{n,j}z,\infty)} \exp\left(-\tilde{\theta}_{n,j}(s - n\varepsilon)\right) P(S^*_{n,j} \in ds) \\ &= O\left(n^{-1/2}\right)\,. \end{aligned} \tag{2.36}$$

This will follow from the following claim: there is $C' > 0$ such that for all n large,

$$P(y \le S^*_{n,j} \le y+1) \le C' n^{-1/2}\,,\ y \in \mathbb{R}\,, \tag{2.37}$$

uniformly in j. Indeed, suppose that this is the case. Then for large n and every $z \in \mathbb{R}$,

$$\begin{aligned}
&\int_{[n\varepsilon-\beta_{n,j}z,\infty)} \exp\left(-\tilde\theta_{n,j}(s-n\varepsilon)\right) P(S^*_{n,j} \in ds) \\
&\le \sum_{j=1}^{\infty} e^{-\tilde\theta_{n,j}(j-1-\beta_{n,j}z)} P\left(S^*_{n,j} - n\varepsilon - \beta_{n,j}z \in [j-1,j)\right) \\
&\le C' n^{-1/2} e^{\tilde\theta_{n,j}\beta_{n,j}z}\left(1-e^{-\tilde\theta_{n,j}}\right)^{-1}\,,
\end{aligned}$$

which shows (2.36).

It remains to prove (2.37). We start by observing that for $M > 0$

$$P(S^*_{n,j} > nM) \le \exp\left\{n\left[\tilde\psi_{n,j}(\tilde\theta_{n,j}+1) - \tilde\psi_{n,j}(\tilde\theta_{n,j}) - M\tilde\theta_{n,j}\right]\right\}.$$

Since the values of both $\tilde\psi_{n,j}(\tilde\theta_{n,j}+1)$ and $\tilde\psi_{n,j}(\tilde\theta_{n,j})$ remain with a compact set independent of n and j, while $\theta_{n,j}$ converges, uniformly in j, to $A^{-1}\tau(\varepsilon) > 0$, we see that by choosing M large enough we can ensure that there is $c > 0$ such that for all n large enough,

$$P(S^*_{n,j} > nM) \le e^{-cn} \ \text{ for all } j.$$

An identical argument shows that, if $M > 0$ is large enough, then here is $c > 0$ such that for all n large enough,

$$P(S^*_{n,j} < -nM) \le e^{-cn} \ \text{ for all } j.$$

That means that it suffices to prove that (2.37) holds for all $|y| \le nM$, uniformly in j.

Notice that by part (b) of Theorem 3, for any $h > 0$ there is $C'_h > 0$ such that

$$P(y \le S^*_n \le y+h) \le C'_h n^{-1/2}\,,\ y \in \mathbb{R}, \tag{2.38}$$

where S^*_n is a random variable with the law

$$P(S^*_n \in ds) = \frac{1}{E\left(\exp(\theta_n S_n)\right)} e^{\theta_n s} P(S_n \in ds)\,.$$

Write

$$P(y \le S_n^* \le y+h) = \frac{E\left(\exp(\tilde\theta_{n,j} S_{n,j})\right)}{E\left(\exp(\theta_n S_n)\right)} \frac{1}{E\left(\exp(\tilde\theta_{n,j} S_{n,j})\right)} \int_{[y,y+h]} \exp\bigl\{(\theta_n - \tilde\theta_{n,j})s\bigr\} e^{\tilde\theta_{n,j}s} P(S_n \in ds).$$

By (2.34), the factor $\exp\{(\theta_n - \tilde\theta_{n,j})s\}$ above is uniformly bounded away from zero over $s \in [y, y+h]$, $|y| \le nM$ and j. Furthermore,

$$\frac{E\left(\exp(\tilde\theta_{n,j} S_{n,j})\right)}{E\left(\exp(\theta_n S_n)\right)} = exp\left\{-n\left[\psi_n(\theta_n) - \psi_n(\tilde\theta_{n,j})\right] - \phi_Z(\tilde\theta_{n,j}\beta_{n,j})\right\},$$

and it follows from (2.35) and uniform boundedness of the argument of ϕ_Z that the ratio above is bounded away from zero over n and j. We conclude that for some $c > 0$, for all n large enough and $|y| \le nM$,

$$\begin{aligned} P(y \le S_n^* \le y+h) \ge & c\frac{1}{E\left(\exp(\tilde\theta_{n,j} S_{n,j})\right)} \int_{[y,y+h]} e^{\tilde\theta_{n,j}s} P(S_n \in ds) \\ \ge & cP(0 \le \beta_{n,j} Z \le h-1) P(y \le S_{n,j}^* \le y+1). \end{aligned}$$

Since $\beta_{n,j}$ is uniformly bounded, we can choose h large enough such that $P(0 \le \beta_{n,j} Z \le h-1)$ is uniformly bounded away from zero, and (2.37) follows from (2.38). □

The next, final, lemma is a major ingredient in the proof of Theorem 1.

LEMMA 2.3. *For a fixed $k \ge 1$, the conditional law of $(S_n - n\varepsilon, Z_{-k}, \dots, Z_k, Z_{n-k}, \dots, Z_{n+k})$ given E_0 converges weakly, as $n \to \infty$, to the law of*

$$\left(T^*, Z_{-k}^-, \dots, Z_k^-, Z_{-k}^+, \dots, Z_k^+\right).$$

PROOF. Since $Z_n, \dots, Z_{n+k}$ are independent both of E_0 and of the rest of the components of the vector whose weak convergence we need to prove, it suffices to show that as $n \to \infty$,

$$(S_n - n\varepsilon, Z_{-k}, \dots, Z_k, Z_{n-k}, \dots, Z_{n-1}) \Rightarrow (T^*, Z_{-k}^-, \dots, Z_k^-, Z_{-k}^+, \dots, Z_{-1}^+)\,, \tag{2.39}$$

with the law in the left hand side being conditional on E_0. Consider the following truncated version of S_n:

$$\bar S_n = \sum_{j=k+1}^{n-k-1} A_{n-1-j} Z_j + \sum_{j=k+1}^{\infty} (A_{j+n-1} - A_{j-1}) Z_{-j},\ n \ge 2(k+1)\,.$$

We claim that there exists $c_n > 0$ such that for any $x \in \mathbb{R}$ any any sequence $x_n \to x$,

$$P(\bar{S}_n \geq n\varepsilon + x_n) \sim c_n e^{-x\tau(\varepsilon)/A} . \tag{2.40}$$

To show this we proceed as in the proof of Lemma 2.1. Let

$$\bar{\psi}_n(t) = n^{-1} E\left(e^{t\bar{S}_n}\right), \ t \in \mathbb{R} .$$

Repeating the argument in Lemma 2.1 shows that

$$\lim_{n\to\infty} \bar{\psi}_n''(t) = A^2 \varphi_Z''(At) \tag{2.41}$$

locally uniformly in $t \in \mathbb{R}$, and that for large n there exists $\bar{\theta}_n > 0$ such that

$$\begin{aligned} \bar{\psi}_n'(\bar{\theta}_n) &= \varepsilon , \\ \lim_{n\to\infty} \bar{\theta}_n &= A^{-1}\tau(\varepsilon) , \end{aligned} \tag{2.42}$$

and

$$P(\bar{S}_n \geq n\varepsilon) \sim \frac{C}{\sqrt{n}} \exp\left(-n(\bar{\theta}_n \varepsilon - \bar{\psi}_n(\bar{\theta}_n))\right), \ n \to \infty , \tag{2.43}$$

with C as in Lemma 2.1. The same argument shows that, if $x_n \to x$, then for large n there exists $\bar{\theta}_{n,x} > 0$ such that

$$\begin{aligned} \bar{\psi}_n'(\bar{\theta}_{n,x}) &= \varepsilon + n^{-1}x_n , \\ \lim_{n\to\infty} \bar{\theta}_{n,x} &= A^{-1}\tau(\varepsilon) , \end{aligned} \tag{2.44}$$

and

$$P(\bar{S}_n \geq n\varepsilon + x_n) \sim \frac{C}{\sqrt{n}} \exp\left(-n(\bar{\theta}_{n,x}(\varepsilon + n^{-1}x_n) - \bar{\psi}_n(\bar{\theta}_{n,x}))\right), \ n \to \infty . \tag{2.45}$$

The mean value theorem applied to (2.42) and (2.44), together with (2.41) implies that

$$A^2 \varphi_Z''(\tau(\varepsilon))(\bar{\theta}_{n,x} - \bar{\theta}_n) = n^{-1}x_n + o(n^{-1}) = n^{-1}x + o(n^{-1}) .$$

This, another application of the mean value theorem, together with (2.42) and locally uniform boundedness of the second derivative implied by (2.41), shows that

$$\bar{\psi}_n(\bar{\theta}_{n,x}) - \bar{\psi}_n(\bar{\theta}_n) = \frac{1}{n} \frac{\varepsilon x}{A^2 \varphi_Z''(\tau(\varepsilon))} + o(n^{-1}) .$$

Putting together the above two displays, we see that

$$\left(\bar\theta_{n,x}-\bar\theta_n\right)\varepsilon-\bar\psi_n(\bar\theta_{n,x})+\bar\psi_n(\bar\theta_n)=o(n^{-1})\,,$$

which in conjunction with (2.45) establishes (2.40) with c_n given by the right-hand side of (2.43).

Finally, for $t>0$ and a compact rectangle $R\subset\mathbb{R}^{3k+1}$,

$$\begin{aligned}
&P\left([S_n-n\varepsilon_n\geq t,\,(Z_{-k},\dots,Z_k,Z_{n-k},\dots,Z_{n-1})\in R]\cap E_0\right)\\
=&P\left(S_n-n\varepsilon\geq t,\,(Z_{-k},\dots,Z_k,Z_{n-k},\dots,Z_{n-1})\in R\right)\\
=&\int_{(x_{-k},\dots,x_k,y_{-k},\dots,y_{-1})\in R}P\Bigg(\bar S_n\geq n\varepsilon+t-\sum_{j=0}^k A_{n-1-j}x_j-\sum_{j=1}^k(A_{j+n-1}-A_{j-1})x_{-j}\\
&\qquad-\sum_{j=1}^k A_{j-1}y_{-j}\Bigg)F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_{-1}) \qquad (2.46)\\
\sim&c_ne^{-\tau(\varepsilon)t/A}\int_{(x_{-k},\dots,x_k,y_{-k},\dots,y_{-1})\in R}\exp\Bigg\{\frac{\tau(\varepsilon)}{A}\Bigg(A\sum_{j=0}^k x_j+\sum_{j=1}^k(A-A_{j-1})x_{-j}\\
&\qquad+\sum_{j=1}^k A_{j-1}y_{-j}\Bigg)\Bigg\}F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_{-1})
\end{aligned}$$

as $n\to\infty$. In order to justify the asymptotic equivalence above, note that for each fixed $x_{-k},\dots,x_k,y_{-k},\dots,y_{-1}$, c_n^{-1} times the integrand of (2.46) converges, by (2.40), to

$$\exp\left\{\frac{\tau(\varepsilon)}{A}\left(-t+A\sum_{j=0}^k x_j+\sum_{j=1}^k(A-A_{j-1})x_{-j}+\sum_{j=1}^k A_{j-1}y_{-j}\right)\right\}.$$

Moreover, replacing each of the variables $x_{-k},\dots,x_k,y_{-k},\dots,y_{-1}$ by their upper bounds implies by the rectangle R and using (2.40) once again, provides a bound to use in the dominated convergence theorem.

We claim that, as $n\to\infty$,

$$P(E_0)\sim c_n\int_{\mathbb{R}^{3k+1}}\exp\left\{\frac{\tau(\varepsilon)}{A}\left(A\sum_{j=0}^k x_j+\sum_{j=1}^k(A-A_{j-1})x_{-j}+\sum_{j=1}^k A_{j-1}y_{-j}\right)\right\}$$
$$F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_{-1}). \qquad (2.47)$$

Once this has been established, (2.39) will follow from (2.46) and (2.47), completing the argument. To prove (2.47), we notice that by (2.40) and Fatou's lemma,

(2.48)
$$\liminf_{n\to\infty} c_n^{-1} P(E_0) \geq \int_{\mathbb{R}^{3k+1}} \exp\left\{ \frac{\tau(\varepsilon)}{A} \left(A \sum_{j=0}^{k} x_j + \sum_{j=1}^{k} (A - A_{j-1}) x_{-j} + \sum_{j=1}^{k} A_{j-1} y_{-j} \right) \right\}$$
$$F_Z(dx_{-k}) \ldots F_Z(dx_k) F_Z(dy_{-k}) \ldots . F_Z(dy_{-1}).$$

By Lemma 2.2 the sequence of the conditional laws of $(Z_{-k}, \ldots, Z_k, Z_{n-k}, \ldots, Z_{n-1})$ given E_0 is tight in $\mathbb{R}^{3k+1}$. Let ν be a subsequential limit of this sequence. It follows from the above inequality and (2.46) with $t = 0$ that

$$\nu(R) \leq P\big((Z^-_{-k}, \ldots, Z^-_k, Z^+_{-k}, \ldots, Z^+_{-1}) \in R\big)$$

for any compact rectangle R in $\mathbb{R}^{3k+1}$, which can only happen if ν is, in fact, the law of the random vector $(Z^-_{-k}, \ldots, Z^-_k, Z^+_{-k}, \ldots, Z^+_{-1})$. Therefore, (2.48) must hold as an equality. □

We are now ready to prove the first of our main theorems.

Proof of Theorem 1. We start with showing that for every fixed $k \geq 1$, conditionally on E_0 as $n \to \infty$,

(2.49)
$$(S_n - n\varepsilon, X_0, \ldots, X_k, X_n, \ldots, X_{n+k}) \Rightarrow (T^*, U^-_0, \ldots, U^-_k, U^+_0, \ldots, U^+_k)\,,$$

with (U^-_k) and (U^+_k) defined in (2.17). For all $i \geq 1$ let

$$X_n^{(i)} = \sum_{j=0}^{i} a_j Z_{n-j},\ n \geq 0\,.$$

Lemma 2.3 implies that for a fixed i,

$$\Big(S_n - n\varepsilon, X_0^{(i)}, \ldots, X_k^{(i)}, X_n^{(i)}, \ldots, X_{n+k}^{(i)}\Big)$$

converges weakly as $n \to \infty$, conditionally on E_0, to

$$(T^*, U_0^{-(i)}, \ldots, U_k^{-(i)}, U_0^{+(i)}, \ldots, U_k^{+(i)})\,,$$

where

$$U_m^{\pm(i)} = \sum_{j=0}^{i} a_j Z^{\pm}_{m-j},\ m \geq 0\,.$$

Note that by Lemma 2.2, for every $\delta > 0$,

$$\sup_{n\geq 1}\sup_{m\in\mathbb{Z}} P\left(\left|X_m^{(i)} - X_m\right| > \delta|E_0\right) \leq \frac{1}{\delta}\left[\sup_{n\geq 1,\, m\in\mathbb{Z}} E(|Z_m||E_0)\right]\sum_{j=i+1}^{\infty}|a_j| \to 0$$

as $i\to\infty$. Since the two series in (2.17) converge in probability, in probability, the claim (2.49) follows from Theorem 3.2 in Billingsley (1999).

Notice that for any $j \geq 1$,

$$\begin{aligned}
\mathbf{1}(E_j) &= \mathbf{1}\left(\sum_{i=j}^{j+n-1} X_i \geq n\varepsilon\right) \\
&= \mathbf{1}\left(\sum_{i=0}^{n-1} X_i - n\varepsilon \geq \sum_{i=0}^{j-1} X_i - \sum_{i=n}^{n+j-1} X_i\right) \\
&= \mathbf{1}\left(S_n - n\varepsilon - \sum_{i=0}^{j-1} X_i + \sum_{i=n}^{n+j-1} X_i \geq 0\right).
\end{aligned}$$

We conclude by (2.49) and the continuous mapping theorem that for $K \geq 1$,

$$\begin{aligned}
\big(\mathbf{1}(E_j),\, j=1,\ldots,K\big) &\Rightarrow \left(\mathbf{1}\left(T^* - \sum_{i=0}^{j-1} U_j^- + \sum_{i=0}^{j-1} U_j^+ \geq 0\right),\, j=1,\ldots,K\right) \\
&= \big(V_j(\varepsilon),\, j=1,\ldots,K\big)
\end{aligned}$$

as $n\to\infty$, where the law of the vector in the left hand side is computed conditionally on E_0. Indeed, the continuity of the exponential random variable T^* means that the boundary of the K-dimensional set above has limiting probability zero. This proves (2.20). □

Finally, we prove our second main result.

Proof of Theorem 2. We start with some variance calculations. For a large m,

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=0}^{m} U_i^-\right) =& \mathrm{Var}\left(\sum_{i=0}^{m}\sum_{k=-\infty}^{-1} a_{i-k} Z_k^-\right) + \mathrm{Var}\left(\sum_{i=0}^{m}\sum_{k=0}^{i} a_{i-k} Z_k^-\right) \\
=& \sum_{k=1}^{\infty}\left(\sum_{i=k}^{m+k} a_i\right)^2 \mathrm{Var}(Z_{-k}^-) + \sum_{k=0}^{m} A_{m-k}^2 \mathrm{Var}(Z_k^-).
\end{aligned}$$

It is elementary that

$$\mathrm{Var}(Z_k^-) \to \sigma_Z^2 \text{ as } \varepsilon \to 0$$

uniformly in $k \in \mathbb{Z}$. Therefore,

$$\sum_{k=0}^{m} A_{m-k}^2 \mathrm{Var}(Z_k^-) \sim \sigma_Z^2 \sum_{k=0}^{m} A_k^2 \sim m\sigma_Z^2 A^2$$

as $\varepsilon \to 0$, $m \to \infty$. Furthermore,

$$\sum_{k=1}^{\infty} \left(\sum_{i=k}^{m+k} a_i \right)^2 \mathrm{Var}(Z_{-k}^-) \sim \sigma_Z^2 \sum_{k=1}^{\infty} \left(\sum_{i=k}^{m+k} a_i \right)^2 = o(m)$$

as $\varepsilon \to 0$, $m \to \infty$, with the last statement an easy consequence of the absolute summability of (a_i). The same argument shows that we also have

$$\sum_{k=0}^{m} A_{m-k}^2 \mathrm{Var}(Z_k^+) \sim \sigma_Z^2 \sum_{k=0}^{m} A_k^2 \sim m\sigma_Z^2 A^2$$

$$\sum_{k=1}^{\infty} \left(\sum_{i=k}^{m+k} a_i \right)^2 \mathrm{Var}(Z_{-k}^+) \sim \sigma_Z^2 \sum_{k=1}^{\infty} \left(\sum_{i=k}^{m+k} a_i \right)^2 = o(m)$$

as $\varepsilon \to 0$, $m \to \infty$.

We define

$$
\begin{aligned}
W_\varepsilon(t) =& \frac{\tau(\varepsilon)}{A} \sum_{i=0}^{[t\varepsilon^{-2}]} (U_i^- - U_i^+) \\
=& \left[\frac{\tau(\varepsilon)}{A} \sum_{i=0}^{[t\varepsilon^{-2}]} \sum_{k=-\infty}^{-1} a_{i-k} Z_k^- + \frac{\tau(\varepsilon)}{A} \sum_{i=0}^{[t\varepsilon^{-2}]} \sum_{k=-\infty}^{-1} a_{i-k} Z_k^+ \right] \\
&+ \left[\frac{\tau(\varepsilon)}{A} \sum_{i=0}^{[t\varepsilon^{-2}]} \sum_{k=0}^{i} a_{i-k} Z_k^- + \frac{\tau(\varepsilon)}{A} \sum_{i=0}^{[t\varepsilon^{-2}]} \sum_{k=0}^{i} a_{i-k} Z_k^+ \right] \\
=:& \quad W_\varepsilon^{(1)}(t) + W_\varepsilon^{(2)}(t), \; \varepsilon > 0, \, t \geq 0.
\end{aligned}
\tag{2.50}
$$

The assumption $EZ = 0$ and (2.15) imply that, as $\varepsilon \to 0$,

$$\varepsilon / A \sim \tau(\varepsilon) \varphi_Z''(0) = \sigma_Z^2 \tau(\varepsilon) \, . \tag{2.51}$$

We have, therefore, verified that $\mathrm{Var}\big(W_\varepsilon^{(1)}(t)\big) \to 0$ as $\varepsilon \to 0$ for every t, so

$$W_\varepsilon^{(1)}(t) - E\big(W_\varepsilon^{(1)}(t)\big) \to 0 \text{ in probability as } \varepsilon \to 0. \tag{2.52}$$

Furthermore, for every t, as $\varepsilon \to 0$,

$$\mathrm{Var}(W_\varepsilon^{(2)}(t)) \to \frac{2t}{A^2\sigma_Z^2}\,.$$

A similar calculation shows that for $0 \le s \le t$,

$$\lim_{\varepsilon\to 0} \mathrm{Cov}(\mathrm{W}_\varepsilon^{(2)}(\mathrm{s}), \mathrm{W}_\varepsilon^{(2)}(\mathrm{t})) = \frac{2\mathrm{s}}{\mathrm{A}^2\sigma_{\mathrm{Z}}^2}\,.$$

Observe next that the third absolute moment of both (Z_k^-) and (Z_k^+) is bounded uniformly in ε and k. Therefore, the Lindeberg condition is satisfied by the triangular array defined by any finite linear combination of the type $\theta_1 W_\varepsilon^{(2)}(t_1)+\ldots+\theta_d W_\varepsilon^{(2)}(t_d)$. Applying the Lindeberg Central Limit Theorem (see e.g. Theorem 27.2 in Billingsley (1995)) and the Cramér-Wold device we conclude that the finite dimensional distributions of $W_\varepsilon^{(2)}(t)-E(W_\varepsilon^{(2)}(t))$ converge to those of $(A\sigma_Z)^{-1}\sqrt{2}B_t$, where B_t is a standard Brownian motion. It follows from (2.52) that the finite dimensional distributions of $W_\varepsilon(t) - E(W_\varepsilon(t))$ converge to the same limit.

Next, let $0 \le s < t$. If $\varepsilon^2 > (t-s)$, then for any $s \le r \le t$ either

$$W_\varepsilon(t) - E(W_\varepsilon(t)) = W_\varepsilon(r) - E(W_\varepsilon(r)) \text{ a.s.}$$

or

$$W_\varepsilon(s) - E(W_\varepsilon(s)) = W_\varepsilon(r) - E(W_\varepsilon(r)) \text{ a.s.},$$

so that

$$E\left[\Big(W_\varepsilon(t) - W_\varepsilon(r) - E\big(W_\varepsilon(t) - W_\varepsilon(r)\big)\Big)^2\Big(W_\varepsilon(r) - W_\varepsilon(s) - E\big(W_\varepsilon(r) - W_\varepsilon(s)\big)\Big)^2\right] = 0.$$

Suppose now that $\varepsilon^2 \le (t-s)$. We have

$$\begin{aligned} &E\left[\Big(W_\varepsilon(t) - W_\varepsilon(s) - E\big(W_\varepsilon(t) - W_\varepsilon(s)\big)\Big)^4\right] \\ &\le 8E\left[\Big(W_\varepsilon^{(1)}(t) - W_\varepsilon^{(1)}(s) - E\big(W_\varepsilon^{(1)}(t) - W_\varepsilon^{(1)}(s)\big)\Big)^4\right] \\ &+8E\left[\Big(W_\varepsilon^{(2)}(t) - W_\varepsilon^{(2)}(s) - E\big(W_\varepsilon^{(2)}(t) - W_\varepsilon^{(2)}(s)\big)\Big)^4\right]. \end{aligned}$$

For a positive constant C independent of ε, s, t, that may change from appearance to appearance, since the fourth moments of (Z_k^-) are bounded uniformly in ε and k, and the coefficients (a_i) are absolutely summable,

$$E\left[\frac{\tau(\varepsilon)}{A}\sum_{i=[s\varepsilon^{-2}]+1}^{[t\varepsilon^{-2}]}\sum_{k=0}^{i}a_{i-k}\big(Z_k^- - E(Z_k^-)\big)\right]^4$$

$$\leq C\varepsilon^4\sum_{k=0}^{[t\varepsilon^{-2}]}E\big(Z_k^- - E(Z_k^-)\big)^4\left(\sum_{i=\max(k,[s\varepsilon^{-2}]+1)}^{[t\varepsilon^{-2}]}a_{i-k}\right)^4$$

$$+\left[C\varepsilon^2\sum_{k=0}^{[t\varepsilon^{-2}]}E\big(Z_k^- - E(Z_k^-)\big)^2\left(\sum_{i=\max(k,[s\varepsilon^{-2}]+1)}^{[t\varepsilon^{-2}]}a_{i-k}\right)^2\right]^2$$

$$\leq C\varepsilon^4\sum_{k=0}^{\infty}\sum_{i=\max(k,[s\varepsilon^{-2}]+1)}^{[t\varepsilon^{-2}]}|a_{i-k}| + \left[C\varepsilon^2\sum_{k=0}^{\infty}\sum_{i=\max(k,[s\varepsilon^{-2}]+1)}^{[t\varepsilon^{-2}]}|a_{i-k}|\right]^2.$$

Since

$$\sum_{k=0}^{\infty}\sum_{i=\max(k,[s\varepsilon^{-2}]+1)}^{[t\varepsilon^{-2}]}|a_{i-k}| \leq \big([t\varepsilon^{-2}]-[s\varepsilon^{-2}]\big)\sum_{i=0}^{\infty}|a_i|$$

$$\leq\big(t\varepsilon^{-2}-(s\varepsilon^{-2}-1)\big)\sum_{i=0}^{\infty}|a_i| \leq 2(t-s)\varepsilon^{-2}\sum_{i=0}^{\infty}|a_i|,$$

we conclude that

$$E\left[\frac{\tau(\varepsilon)}{A}\sum_{i=[s\varepsilon^{-2}]+1}^{[t\varepsilon^{-2}]}\sum_{k=0}^{i}a_{i-k}\big(Z_k^- - E(Z_k^-)\big)\right]^4 \leq C(t-s)^2.$$

A similar argument shows that

$$E\left[\frac{\tau(\varepsilon)}{A}\sum_{i=[s\varepsilon^{-2}]+1}^{[t\varepsilon^{-2}]}\sum_{k=0}^{i}a_{i-k}\big(Z_k^+ - E(Z_k^+)\big)\right]^4 \leq C(t-s)^2,$$

so that

$$E\left[\Big(W_\varepsilon^{(2)}(t)-W_\varepsilon^{(2)}(s)-E\big(W_\varepsilon^{(2)}(t)-W_\varepsilon^{(2)}(s)\big)\Big)^4\right]\leq C(t-s)^2.$$

In the same way we can check that

$$E\left[\Big(W_\varepsilon^{(1)}(t)-W_\varepsilon^{(1)}(s)-E\big(W_\varepsilon^{(1)}(t)-W_\varepsilon^{(1)}(s)\big)\Big)^4\right]\leq C(t-s)^2,$$

so we conclude that

$$E\left[\Big(W_\varepsilon(t) - W_\varepsilon(s) - E\big(W_\varepsilon(t) - W_\varepsilon(s)\big)\Big)^4\right] \le C(t-s)^2 \tag{2.53}$$

if $\varepsilon^2 \le (t-s)$. By the Cauchy-Schwarz inequality, for any $0 \le s \le r \le t$ we have

$$E\left[\Big(W_\varepsilon(t) - W_\varepsilon(r) - E\big(W_\varepsilon(t) - W_\varepsilon(r)\big)\Big)^2\Big(W_\varepsilon(r) - W_\varepsilon(s) - E\big(W_\varepsilon(r) - W_\varepsilon(s)\big)\Big)^2\right]$$
$$\le C(t-s)^2$$

if $\varepsilon^2 \le (t-s)$. Appealing to Theorem 13.5 of Billingsley (1999) we conclude that for any fixed T, the family

$$\{(W_\varepsilon(t) - E(W_\varepsilon(t)) : 0 \le t \le T) : \varepsilon > 0\}$$

is tight in $D[0,T]$ endowed with the Skorohod J_1 topology. Therefore,, as $\varepsilon \to 0$,

$$(W_\varepsilon(t) - E(W_\varepsilon(t)) : 0 \le t \le T) \Rightarrow \Big((A\sigma_Z)^{-1}\sqrt{2}B_t : 0 \le t \le T\Big)\ , \tag{2.54}$$

in $D[0,T]$. Furthermore,

$$E\left(W_\varepsilon(t)\right) = \frac{\tau(\varepsilon)}{A}\Bigg[\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}EZ_k^- + \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{i} a_{i-k}EZ_k^-$$
$$- \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}EZ_k^+ - \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{i} a_{i-k}EZ_k^+\Bigg].$$

The fourth double sum vanishes. Clearly, $|EZ_k^\pm| = O(\tau(\varepsilon))$ uniformly in ε and $k \in \mathbb{Z}$. Therefore,

$$\left|\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}EZ_k^-\right| \le O(\tau(\varepsilon))\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=i+1}^{\infty}|a_k| = o\big(\tau(\varepsilon)\varepsilon^{-2}\big)$$

uniformly in t in a compact set. Similarly,

$$\left|\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}EZ_k^+\right| = o\big(\tau(\varepsilon)\varepsilon^{-2}\big)$$

uniformly in t in a compact set. Finally, $EZ_0^- \sim \tau(\varepsilon)\sigma_Z^2$ as $\varepsilon \to 0$, so

$$\sum_{i=0}^{[t\varepsilon^{-2}]} \sum_{k=0}^{i} a_{i-k} EZ_k^- = EZ_0^- \sum_{i=0}^{[t\varepsilon^{-2}]} A_i \sim \tau(\varepsilon)\sigma_Z^2 t\varepsilon^{-2}$$

uniformly in t in a compact set. We conclude by (2.51) that for all $\varepsilon > 0$ small enough,

$$E\left(W_\varepsilon(t)\right) \geq \frac{t}{2A^2\sigma_Z^2}, \ t \geq 1 \tag{2.55}$$

and

$$E\left(W_\varepsilon(t)\right) \to \frac{t}{A^2\sigma_Z^2}, \ \varepsilon \to 0,$$

uniformly in t in a compact set. Since the addition in $D[0,T]$ is continuous at continuous functions, this along with (2.54) shows that

$$(W_\varepsilon(t) : 0 \leq t \leq T) \Rightarrow \Big((A\sigma_Z)^{-1}\sqrt{2}B_t + (A\sigma_Z)^{-2}t : 0 \leq t \leq T\Big), \ \varepsilon \to 0,$$

in $D[0,T]$.

For any real λ the function $\varphi : \ D[0,T] \to \mathbb{R}$ defined by

$$\varphi(f) = \int_0^T \mathbf{1}\big(\lambda \geq f(t)\big)\, dt$$

is continuous at any continuous f that takes value λ only on a set of measure 0. Therefore, for any such λ, by the continuous mapping theorem,

$$\int_0^T \mathbf{1}\big(\lambda \geq W_\varepsilon(t)\big)\, dt \Rightarrow \int_0^T \mathbf{1}\big(\lambda \geq (A\sigma_Z)^{-1}\sqrt{2}B_t + (A\sigma_Z)^{-2}t\big)\, dt$$

$$\stackrel{d}{=} \int_0^T \mathbf{1}\big(\lambda \geq \sqrt{2}B_{(A\sigma_Z)^{-2}t} + (A\sigma_Z)^{-2}t\big)\, dt = A^2\sigma_Z^2 \int_0^{(A\sigma_Z)^2 T} \mathbf{1}\big(\lambda \geq \sqrt{2}B_t + t\big)\, dt\,.$$

Noticing that we can write

$$V_j(\varepsilon) = \mathbf{1}\left(T_0 \geq \frac{\tau(\varepsilon)}{A} \sum_{i=0}^{j-1} (U_i^- - U_i^+)\right), \ j \geq 1,$$

where T_0 is a standard exponential random variable independent of the collection $(Z_j^u : j \in \mathbb{Z}, u = + \text{ or } -)$, we conclude that for any $T > 0$,

$$\varepsilon^2 \sum_{j=0}^{[T\varepsilon^{-2}]} V_j(\varepsilon) = \int_0^T \mathbf{1}\big(T_0 \geq W_\varepsilon(t)\big)\, dt - V_{[T\varepsilon^{-2}]}\big(T - \varepsilon^2[T\varepsilon^{-2}]\big)$$

$$\Rightarrow A^2\sigma_Z^2 \int_0^{(A\sigma_Z)^2 T} \mathbf{1}\big(T_0 \geq \sqrt{2}B_t + t\big)\, dt.$$

It is clear that the latter integral converges a.s. to the integral prescribed in the theorem. Therefore, we can appeal to Theorem 3.2 in Billingsley (1999) and complete the proof once we show that for any $\delta > 0$,

$$\lim_{T\to\infty} \limsup_{\varepsilon\to 0} P\left(\varepsilon^2 \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} V_j(\varepsilon) > \delta\right) = 0\,. \tag{2.56}$$

However, by Markov's inequality

$$P\left(\varepsilon^2 \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} V_j(\varepsilon) > \delta\right) \leq \varepsilon^2\delta^{-1} \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} P\big(T_0 \geq W_\varepsilon((j-1)\varepsilon^2)\big)$$

$$\leq \varepsilon^2\delta^{-1} \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} P\left(W_\varepsilon((j-1)\varepsilon^2) \leq \frac{(j-1)\varepsilon^2}{4A^2\sigma_Z^2}\right) + \varepsilon^2\delta^{-1} \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} \exp\left\{-\frac{(j-1)\varepsilon^2}{4A^2\sigma_Z^2}\right\}.$$

By (2.55) and (2.53) we have for some positive constant C,

$$P\left(W_\varepsilon((j-1)\varepsilon^2) \leq \frac{(j-1)\varepsilon^2}{4A^2\sigma_Z^2}\right) \leq C\varepsilon^{-4}(j-1)^{-2}$$

for all $j > [T\varepsilon^{-2}]$, $T \geq 1$ and $\varepsilon > 0$ small enough. This estimate suffices to establish (2.56) and so the proof is complete. □

3. Discussion. As is usually the case with large deviations, the limiting distributions obtained in (2.20) and (2.21) of Theorem 1 depend largely on the underlying model through the distribution of the noise variables F_Z and the coefficients (a_i). This dependence largely disappears in Theorem 2 where the limiting distribution depends only on the noise variance σ_Z^2 and the sum of the coefficients A. This can be understood by viewing the case of a small overshoot ε as approaching the regime of moderate deviations. Indeed, in the case of moderate deviations one expects that the central limit behaviour becomes visible and leads to a collapse of the model ingredients necessary to describe the limit to a bare minimum consisting of second order information.

This naturally leads to the question of a difference of how large deviations cluster between the short memory moving average processes and long memory moving average processes. It is common to say that the coefficients of the moving average process (1.1) with long memory are square summable but not absolutely summable. Assuming certain regularity of the coefficients (a_i) (e.g. their regular variation), one can show that for any fixed $j \geq 1$ and $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(E_j(n,\varepsilon)\big|E_0(n,\varepsilon)\right) = 1\,,$$

so one expects infinitely many events $(E_j(n,\varepsilon))$ to happen once $E_0(n,\varepsilon)$ does. This necessitates different limiting procedures when studying large deviations clustering of such long memory processes. It is important to note that for these long memory moving average processes, in the notation of (2.23), $n = o(\mathrm{Var}(S_n))$, so one can view the events $(E_j(n,\varepsilon))$ as moderate deviation events and not large deviation events. Indeed, it turns out that a natural limiting procedure leads to a collapse in the amount of information about the model needed to describe the limit, which is similar to the situation with Theorem 2 in the present, short memory case. We will describe the details in a subsequent work.

APPENDIX A: SOME USEFUL FACTS

The following non-logarithmic version of a large deviation statement and a related estimate are from Chaganty and Sethuraman (1983).

THEOREM 3. *Let $\{T_n\}$ be a sequence of random variables with*

$$E\left(e^{zT_n}\right) < \infty \text{ for any } z \in \mathbb{R},\ n \geq 1\,.$$

For a sequence $\{a_n\}$ of positive numbers with

$$\lim_{n\to\infty} a_n = \infty \tag{1.57}$$

we denote

$$\psi_n(z) = a_n^{-1} \log E\left(e^{zT_n}\right),\ z \in \mathbb{R},\ n \geq 1\,.$$

Let $\{m_n\}$ be a bounded sequence of real numbers. Assume that there exists a bounded positive sequence $\{\tau_n\}$ satisfying

$$\psi_n'(\tau_n) = m_n,\ n \geq 1\,, \tag{1.58}$$

$$a_n^{-1/2} = o(\tau_n),\ n \to \infty\,, \tag{1.59}$$

and such that for all fixed $\delta, \lambda > 0$,

$$\sup_{\delta \leq |t| \leq \lambda\tau_n} \left| \frac{1}{E\left(e^{\tau_n T_n}\right)} E\left(e^{(\tau_n + it)T_n}\right) \right| = o\left(a_n^{-1/2}\right),\ n \to \infty\,, \tag{1.60}$$

(with the supremum of the empty set defined as zero). Furthermore, assume that

$$\sup_{n\geq 1,\, z\in[-a,a]} |\psi_n(z)| < \infty \text{ for any } a > 0 \tag{1.61}$$

and that

$$\inf_{n\geq 1} \psi_n''(\tau_n) > 0\,. \tag{1.62}$$

(a) *Under the above assumption,*

(1.63)

$$P\left(a_n^{-1}T_n \geq m_n\right) \sim \frac{1}{\tau_n\sqrt{2\pi a_n \psi_n''(\tau_n)}} \exp\left\{-a_n\left(m_n\tau_n - \psi_n(\tau_n)\right)\right\},\ n\to\infty.$$

(b) *Let*

$$b_n = \tau_n\sqrt{a_n\psi_n''(\tau_n)},$$

and let T_n^* be a random variable with the law

$$P(T_n^* \in\, du) = \frac{1}{E(e^{\tau_n T_n})} e^{u\tau_n} P(T_n \in\, du)\,.$$

Then

$$\sup_{n\geq 1,\, y\in\mathbb{R}} b_n P\left(y \leq \tau_n T_n^* \leq y+1\right) < \infty.$$

Proof. The first part of the theorem is Theorem 3.3 in Chaganty and Sethuraman (1983). Furthermore, Lemmas 3.1 and 3.2 *ibid.* show that the hypotheses (2.7) and (2.8) of Theorem 2.3 therein hold, and the second part of Theorem 3 follows from (2.9) of that paper. □

REFERENCES

P. Billingsley (1995): *Probability and Measure*. Wiley, New York, 3rd edition.

P. Billingsley (1999): *Convergence of Probability Measures*. Wiley, New York, 2nd edition.

N. Chaganty and J. Sethuraman (1983): Strong large deviations and local limit theorems. *Annals of Probability* 21:1671–1690.

A. Dembo and O. Zeitouni (1998): *Large Deviations Techniques and Applications*. Springer-Verlag, New York, 2nd edition.

J.-D. Deuschel and D. Stroock (1989): *Large Deviations*. Academic Press, Boston.

P. Embrechts, C. Klüppelberg and T. Mikosch (2003): *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin. 4th corrected printing.

T. Mikosch and A. Nagaev (1998): Large deviations of heavy-tailed sums with applications to insurance. *Extremes* 1:81–110.

G. Samorodnitsky (2016): *Stochastic Processes and Long Range Dependence*. Springer, Cham, Switzerland.

Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, Kolkata
E-mail: arijit.isi@gmail.com

School of Operations Research
and Information Engineering
Cornell University
E-mail: gs18@cornell.edu