The University of Chicago
Department of Statistics
TECHNICAL REPORT SERIES

MODERATE DEVIATIONS FOR STATIONARY PROCESSES

Wei Biao Wu and Zhibiao Zhao

TECHNICAL REPORT NO. 571

Departments of Statistics
The University of Chicago
Chicago, Illinois 60637

October 2, 2006

Moderate deviations for stationary processes

By Wei Biao Wu and Zhibiao Zhao

Department of Statistics, The University of Chicago

Abstract: We obtain asymptotic expansions for probabilities of moderate deviations for stationary causal processes. The imposed dependence conditions are easily verifiable and they are directly related to the data-generating mechanism of the underlying processes. The results are applied to functionals of linear processes and nonlinear time series. We carry out a simulation study and investigate the relationship between accuracy of tail probabilities and the strength of dependence.

Keywords: Moderate deviations, nonlinear time series.

1 Introduction

Let $(X_i)_{i\in\mathbb{Z}}$ be a mean zero strictly stationary process. Define

$$S_n = \sum_{i=1}^{n} X_i.$$

We are interested in the asymptotic behavior of $\mathbb{P}(S_n \geq \sqrt{n}r)$, where $r = r_n$ is a sequence of positive numbers and r_n diverges to ∞ at an appropriate rate. The central limit theorem (CLT) asserts that, for a fixed r, $\mathbb{P}(S_n/\sigma \geq \sqrt{n}r) \to 1 - \Phi(r)$ as $n \to \infty$, where $\sigma = \lim_{n\to\infty} \|S_n\|_2/\sqrt{n}$. By allowing $r \to \infty$, the moderate deviation principle (MDP) provides a tail bound associated with CLT. For the special case in which X_i are independent and identically distributed (iid), one has the following classical result. Let $c > 0$. Assume that $\mathbb{E}(|X_1|^q) < \infty$ for some $q > c^2 + 2$ and let $\sigma > 0$ be the standard deviation of X_1. Then

$$\frac{\mathbb{P}(S_n/\sigma \geq c\sqrt{n \log n})}{1 - \Phi(c\sqrt{\log n})} = 1 + o(1), \tag{1}$$

where Φ is the standard normal distribution function. The moderate deviation principle of type (1) has been investigated by Osipov (1972), Michel (1976), Amosova (1982) and

Frolov (1998, Theorem 6) for sums of iid random variables and by Rubin and Sethuraman (1965), Amosova (1972), Petrov (2002) and Frolov (2005) for arrays of independent random variables. De Acosta (1992) considered MDP of iid Banach space valued random vectors.

It is a challenging problem to establish moderate deviation results for dependent random variables. Ghosh (1974) obtained an MDP for sums of m-dependent sequences. Several researchers studied MDP for mixing processes; see Ghosh and Babu (1977), Babu and Singh (1978) and Gao (1996) among others. For MDP for Markov processes, see Chen (2001) and references therein. Recently, Dong, Tan and Yang (2005) considered moving average processes. For martingales deep results are obtained in Bose (1986), Dembo (1996), Gao (1996), Grama (1997) and Grama and Haeusler (2006). The latter two papers develop asymptotic expansions of the probabilities $\mathbb{P}(S_n/\sigma \geq \sqrt{n}r_n)$. Such asymptotic expansions appear more accurate than the results based on a logarithmic scale and they may require different sets of conditions.

In this paper we shall study asymptotic properties of the probability $\mathbb{P}(S_n/\sigma \geq \sqrt{n}r_n)$ itself instead of the one on the logarithmic scale. In particular, we shall obtain an asymptotic expansion for $\mathbb{P}(S_n/\sigma \geq \sqrt{n}r_n)$ for stationary processes. We assume that (X_i) is a stationary causal process with the form

$$X_i = g(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{2}$$

where $\varepsilon_i, i \in \mathbb{Z}$, are iid random variables and g is a measurable function such that X_i is well-defined. The framework (2) does not seem to be overly restrictive. The Wiener-Rosenblatt conjecture states that, for every stationary and ergodic process (X_i), there exists iid random variables ε_i and a measurable function g such that the processes $(X_i)_{i\in\mathbb{Z}}$ and $(g(\ldots, \varepsilon_{i-1}, \varepsilon_i))_{i\in\mathbb{Z}}$ are identically distributed. See Wiener (1958), Rosenblatt (1971), Kallianpur (1981) and Borkar (1993).

We introduce the following notation. Let the shift process $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. For a random variable Z write $Z \in \mathcal{L}^p, p > 0$, if $\|Z\|_p = [\mathbb{E}(|Z|^p)]^{1/p} < \infty$, and $\|Z\| = \|Z\|_2$. For two real sequences $\{a_n\}$ and $\{b_n\}$, write $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \leq C|b_n|$ holds for large n and $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. The main result on asymptotic expansions is presented in Section 2 and proved in Section 5. Applications to nonlinear transforms of linear processes and nonlinear time series are given in Section 3. In

Section 4, we perform a simulation study and show that the accuracy of tail probabilities decreases as the dependence gets stronger.

2 Main results

It is necessary to have an appropriate dependence measure to quantify the dependence of the process (X_i). Following Wu (2005b), we can view (2) as a physical system with $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$ being the input, X_i being the output and g being a filter or transform. We then interpret the dependence as the degree of dependence of output on input. To this end, we adopt the idea of coupling. Let $(\varepsilon'_i)_{i\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$ and $\mathcal{F}'_i = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_i)$ be the coupled version of $\mathcal{F}_i$. Define the coupled process $X'_i = g(\mathcal{F}'_i)$ of X_i. Assume $X_1 \in \mathcal{L}^q, q > 0$. Let

$$\theta_q(i) = \|X_i - X'_i\|_q. \tag{3}$$

Roughly speaking, $\theta_q(i)$ measures the degree of dependence of $X_i = g(\mathcal{F}_i)$ on ε_0 and it is directly related to the data-generating mechanism of the underlying process. Wu (2005b) called $\theta_q(i)$ the *physical dependence measure*. Throughout the paper we assume

$$\Theta_q(k) := \sum_{i=k}^{\infty} \theta_q(i) < \infty, \quad k = 0, 1, \ldots. \tag{4}$$

The quantity $\Theta_q(0)$ can be interpreted as cumulative impact of ε_0 on all future values $(X_i)_{i\geq 0}$. In this sense the condition $\Theta_q(0) < \infty$ suggests short-range dependence since the cumulative impact of ε_0 on future outputs is finite. In Wu (2005b), it is called strong stability condition.

Fix $p \in (1, 2]$. For $x > 1$ let $r_x > 0$ be the solution to the equation

$$x = (1 + r_x)^{6p-3} \exp(r_x^2/2). \tag{5}$$

We also write $x_r = (1 + r)^{6p-3} \exp(r^2/2)$. Let $\tau_n \to \infty$ be a positive sequence and U_n a sequence of random variables such that the CLT $U_n \Rightarrow \Phi$ holds. We say that U_n satisfies a *moderate deviation principle with rate* τ_n if, for every $a > 0$, there exists a constant

$C = C_{a,p}$, independent of x and n, such that

$$\left|\frac{\mathbb{P}(U_n \geq r_x)}{1-\Phi(r_x)} - 1\right| \leq C(x/\tau_n)^{1/(1+2p)} \text{ and } \left|\frac{\mathbb{P}(U_n \leq -r_x)}{\Phi(-r_x)} - 1\right| \leq C(x/\tau_n)^{1/(1+2p)} \tag{6}$$

hold uniformly in $x \in [1, a\tau_n]$. The quantity τ_n gives a range that the MDP is applicable and larger τ_n is preferred for wider applicability. The MDP (6) implies the expansion

$$\begin{aligned} \mathbb{P}(U_n \geq r_x) &= [1-\Phi(r_x)]\{1 + O[(x/\tau_n)^{1/(1+2p)}]\} \\ &= \frac{\exp(-r_x^2/2)}{r_x\sqrt{2\pi}}\{1 + o(1)\} \end{aligned}$$

as $x \to \infty$ with $x = o(\tau_n)$. By Remark 1 in Grama and Haeusler (2006), as $x \to \infty$, r_x has the asymptotic expansion

$$r_x^2 = 2\log x - [2(6p-3) + o(1)]\log(1 + \sqrt{2\log x}).$$

Remark 1. As Grama (1997) pointed out, the exponent $6p - 3$ in (5) is not optimal. In Grama and Haeusler (2006), they obtained the optimal exponent $2p + 1$. However, the latter paper assumed that $p \in (1, 3/2]$, which is not needed in Grama's paper. In this paper, we adopt Grama's framework to allow a wider range of p. Clearly, the argument in our proof easily carries over to Grama and Haeusler's framework with optimal exponent by assuming $p \in (1, 3/2]$.

Theorem 1. *Let $X_0 \in \mathcal{L}^{2p}$, $p \in (1, 2]$ and assume $\Theta_{2p}(0) < \infty$. Then the limit $\sigma = \lim_{n\to\infty} \|S_n\|/\sqrt{n}$ exists and is finite. Assume $\sigma > 0$ and that there exist $0 < \alpha \leq \beta \leq \alpha + 1/2$ such that*

$$\Theta_{2p}(m) = O(m^{-\alpha}) \tag{7}$$

and

$$\psi_{2p}(m) := \sum_{i=m}^{\infty} \theta_{2p}^2(i) = O(m^{-2\beta}). \tag{8}$$

Let $\eta = \alpha\beta/(1+\alpha)$. Then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n^{p-1}$ or $\tau_n =$

$n^{p-1}/\log^p n$ *or* $\tau_n = n^{p\eta}$ *under* $\eta > 1-1/p$ *or* $\eta = 1-1/p$ *or* $\eta < 1-1/p$*, respectively.*

Remark 2. It is easily seen that, if $\alpha \geq \beta$, then (7) implies (8). On the other hand, if $\beta \geq \alpha + 1/2$, simple calculations show that (8) implies (7). To see this, by the Cauchy-Schwarz inequality, for all $k \in \mathbb{N}$,

$$\sum_{i=k}^{2k-1} \theta_{2p}(i) \leq \left[k \sum_{i=k}^{2k-1} \theta_{2p}^2(i) \right]^{1/2} = O(k^{-(\beta-1/2)}).$$

So (7) follows by summing up the above inequality over $k = 2^r m, r = 0, 1, \ldots$. Hence the condition $\alpha \leq \beta \leq \alpha + 1/2$ in Theorem 1 is needed to avoid redundancy of either conditions.

Corollary 1. *Let* $X_0 \in \mathcal{L}^{2p}$*,* $p \in (1, 2]$*. Assume that (7) holds for some*

$$\alpha > \frac{p - 1 + \sqrt{5p^2 - 6p + 1}}{2p}. \tag{9}$$

Then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ *with rate* $\tau_n = n^{p-1}$*.*

Proof. Clearly, if (7) holds, then (8) holds with $\beta = \alpha$. So $\eta \geq \alpha^2/(1+\alpha)$. Simple calculations show that (9) implies $\eta > 1 - 1/p$. By Theorem 1, the Corollary follows. $\diamondsuit$

Corollary 2. *Let* $X_0 \in \mathcal{L}^{2p}$*,* $p \in (1, 2]$*. Assume that (8) holds for some*

$$\beta > \frac{3p - 2 + \sqrt{17p^2 - 20p + 4}}{4p}. \tag{10}$$

Then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ *with rate* $\tau_n = n^{p-1}$*.*

Proof. By Remark 2, if (8) holds for some $\beta > 1/2$, then we have (7) with $\alpha = \beta - 1/2$. Simple calculations show that (10) implies $\eta > 1 - 1/p$. $\diamondsuit$

3 Applications

To apply Theorem 1, one needs to compute the physical dependence measure $\theta_q(i) = \|X_i - X_i'\|_q$. It is usually not difficult to work with $\theta_q(i)$ due to the way it is defined which is directly based on the data-generating mechanism of the underlying process. Here we should consider the calculation of $\theta_q(i)$ for functionals of linear processes and nonlinear time series.

3.1 Functionals of linear processes

Let $\varepsilon_i, i \in \mathbb{Z}$, be iid random variables with $\varepsilon_0 \in \mathcal{L}^q, q > 0$. Assume $\mathbb{E}(\varepsilon_0) = 0$ if $q \geq 1$. Let a_i be real numbers satisfying $\sum_{i=0}^{\infty} |a_i|^{q\wedge 2} < \infty$, where $a \wedge b = \min(a, b)$. Then the linear process

$$Y_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j} \tag{11}$$

is well defined and is strictly stationary. We consider the following functionals of linear processes

$$X_i = h(Y_i) - \mathbb{E}[h(Y_i)]$$

for some measurable function h such that $h(Y_0) \in \mathcal{L}^4$. Assume that $h(\cdot)$ is Lipschitz continuous. Namely, the Lipschitz constant $L_h := \sup_{x\neq x'} |h(x) - h(x')|/|x - x'| < \infty$.

Let $Y_i' = Y_i + a_i(\varepsilon_0' - \varepsilon_0)$. If $q > 2$, by the Lipschitz continuity of $h(\cdot)$,

$$\theta_{q\wedge 4}(i) = \|h(Y_i) - h(Y_i')\|_{q\wedge 4} \leq L_h \|Y_i - Y_i'\|_{q\wedge 4} = O(|a_i|).$$

If $q \leq 2$, we further assume that $B_h := \sup_x |h(x)| < \infty$. Let $C_h = 2B_h + L_h$. Then

$$\begin{aligned} \theta_4(i) &= \|h(Y_i) - h(Y_i')\|_4 \leq C_h \|1 \wedge |Y_i - Y_i'|\|_4 \\ &\leq C_h \||Y_i - Y_i'|^{q/4}\|_4 = O(|a_i|^{q/4}). \end{aligned} \tag{12}$$

By Theorem 1, we have the following Corollary 3.

Corollary 3. *Let $\varepsilon_0 \in \mathcal{L}^q, q > 0$, and $a_i = O(i^{-\gamma})$ for some $\gamma > 1$. Assume that $h(\cdot)$ is a Lipschitz continuous function such that $h(Y_0) \in \mathcal{L}^4$; if $q \leq 2$, we also assume that $h(\cdot)$ is* *bounded. (i) If $q > 2$, let $p = (q \wedge 4)/2$, assume*

$$\gamma > \frac{5p - 2 + \sqrt{17p^2 - 20p + 4}}{4p},$$

then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n^{p-1}$. (ii) If $q \le 2$, assume $q\gamma > 4+2\sqrt{2}$, then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n$.

Example 1. Assume $q \ge 4$ and $a_i = O(i^{-\gamma})$ for some $\gamma > 1 + \sqrt{2}/2$. Let $h(\cdot)$ be a Lipschitz continuous function such that $h(Y_0) \in \mathcal{L}^4$. Then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n$.

Example 2. Consider the AR(r) model

$$Y_n = b_1 Y_{n-1} + b_2 Y_{n-2} + \ldots + b_r Y_{n-r} + \varepsilon_n. \tag{13}$$

Assume that $1 - b_1 x - b_2 x^2 - \ldots b_p x^r \neq 0$ for all $|x| \le 1$. Then Y_n is of the form (11) with the coefficients $a_i = O(\lambda^i)$ with some $|\lambda| < 1$. Let $\varepsilon_0 \in \mathcal{L}^q$ for some $q > 0$ and $h(\cdot)$ be a Lipschitz continuous function such that $h(Y_0) \in \mathcal{L}^4$. If $q > 2$, then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n^{p-1}, p = (q \wedge 4)/2$. If $q \le 2$ and $h(\cdot)$ is bounded, then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n$. A similar example is considered in Grama and Haeusler (2006). Comparing with their method, our approach is simpler and it allows for functionals of the AR(r) process. The latter situation seems difficult to deal with using Grama and Haeusler's method.

It is slightly more complicated to deal with the empirical process in which $h_x(\cdot) = \mathbf{1}_{\cdot \le x}$. Stronger conditions on γ and ε_i are needed.

Corollary 4. *Let $a_i = O(i^{-\gamma})$ for some $\gamma > 0$. Further assume either (i) $\varepsilon_0 \in \mathcal{L}^1, \gamma > 4 + 2\sqrt{2}$ and ε_0 has a bounded density or (ii) $\varepsilon_0 \in \mathcal{L}^q, q\gamma > 4(2 + \sqrt{2})$ and the distribution function of Y_0 is Lipschitz continuous. Then (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n$.*

Proof. Let $p = 2$. By Corollary 2, it suffices to show that (8) holds for some $\beta > (1 + \sqrt{2})/2$ under either (i) or (ii).

(i) Without loss of generality let $a_0 = 1$. Denote by F_ε and f_ε the distribution and density functions of ε_i, respectively. Let $C_0 = \sup_u f_\varepsilon(u)$. Since $\varepsilon_0 \in \mathcal{L}^1$,

$$\begin{aligned}
\mathbb{E}\{F_\varepsilon(\frac{x-\varepsilon_i}{a_i})[1-F_\varepsilon(\frac{x-\varepsilon_i}{a_i})]\} &= \int_{\mathbb{R}} F_\varepsilon(\frac{x-u}{a_i})[1-F_\varepsilon(\frac{x-u}{a_i})]f_\varepsilon(u)du \\
&= |a_i|\int_{\mathbb{R}} F_\varepsilon(t)[1-F_\varepsilon(t)]f_\varepsilon(x-a_i t)dt \\
&\le |a_i|C_0\int_{\mathbb{R}} F_\varepsilon(t)[1-F_\varepsilon(t)]dt = O(|a_i|).
\end{aligned}$$

Observe that

$$\mathbb{E}|\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0\le x} - \mathbb{E}(\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0\le x}|\varepsilon_i)| = \mathbb{E}\{F_\varepsilon(\frac{x-\varepsilon_i}{a_i})[1-F_\varepsilon(\frac{x-\varepsilon_i}{a_i})]\}$$

and $\mathbb{E}(\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0\le x}|\varepsilon_i) = \mathbb{E}(\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0'\le x}|\varepsilon_i)$, we have by triangle inequality that

$$\mathbb{E}|\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0\le x} - \mathbf{1}_{\varepsilon_i+a_i\varepsilon_0'\le x}| \le 2\mathbb{E}|\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0\le x} - \mathbb{E}(\mathbf{1}_{\varepsilon_i+a_i\varepsilon_0\le x}|\varepsilon_i)| = O(|a_i|)$$

uniformly in x. By independence, the preceding relation implies that $\sup_x \mathbb{E}|\mathbf{1}_{Y_i\le x} - \mathbf{1}_{Y_i'\le x}| = O(|a_i|)$. Hence $\theta_4(i) = O(|a_i|^{1/4})$. By Karamata's theorem, it is easily seen that (8) is satisfied with any $\beta < \gamma/4 - 1/2$.

(ii) Let i be fixed and assume without loss of generality that $a_i \neq 0$. Define

$$\omega(u) = \mathbf{1}_{u\le x-\lambda} - \frac{u-x-\lambda}{2\lambda}\mathbf{1}_{|u-x|<\lambda}.$$

Then $\omega(\cdot)$ is Lipschitz continuous with Lipschitz constant $L_\omega = 1/(2\lambda)$. By triangle inequality,

$$\|\mathbf{1}_{Y_i\le x} - \mathbf{1}_{Y_i'\le x}\|_4 \le \|\omega(Y_i) - \omega(Y_i')\|_4 + 2\|\mathbf{1}_{Y_i\le x} - \omega(Y_i)\|_4 = O(|a_i|^{q/4}/\lambda + \lambda)$$

in view of (12) and the Lipschitz continuity of the distribution function of Y_i. Let $\lambda = |a_i|^{q/8}$. Then $\theta_4(i) = O(|a_i|^{q/8})$. By Karamata's theorem, it is easily seen that (8) holds with any $\beta < q\gamma/8 - 1/2$. $\diamondsuit$

Remark 3. Let $\varepsilon_0 \in \mathcal{L}^q$. If $q \ge 2$, then conditions in (ii) are weaker than those in (i). If $1 \le q < 2$, then (ii) imposes a more restrictive decay rate on a_i while relaxing the assumption on the distribution function of ε_i. If $q < 1$, (i) is not applicable. So (i) and (ii) have different ranges of applicability.

Example 3. Consider the AR(1) process $Y_n = aY_{n-1} + (1-a)\varepsilon_n$, where ε_n are Bernoulli random variables with success probability 1/2. Then $a_n = O(a^n)$. In the particular case of $a = 1/2$, this model takes uniform(0,1) as invariant distribution. Solomyak (1995) showed that for almost all $a \in [1/2, 1)$ (Lebesgue), Y_n has an absolutely continuous invariant measure. Therefore for those a (say, $a = 1/2$) such that the density of Y_n is bounded, conditions (ii) in Corollary 4 are satisfied and the moderate deviation principle (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n$.

3.2 Nonlinear time series

Let $\varepsilon_i, i \in \mathbb{Z}$, be iid random variables and define recursively

$$X_n = R(X_{n-1}, \varepsilon_n), \tag{14}$$

where $R(\cdot, \varepsilon)$ is a measurable random map. Many nonlinear time series are of the form (14). For example, threshold autoregressive models TAR(1) is given by $X_n = aX_{n-1}^+ + bX_{n-1}^- + \varepsilon_n$ (see Tong, 1990), ARCH models (Engle, 1982) has the form $X_n = \varepsilon_n\sqrt{a^2 + b^2X_{n-1}^2}$, random coefficient models (Nicholls and Quinn, 1982) and exponential autoregressive models (Haggan and Ozaki, 1981) among others. Assume that there exists x_0 and $\alpha > 0$ such that $R(x_0, \varepsilon_0) \in \mathcal{L}^\alpha$ and

$$\rho := \sup_{x \neq x'} \frac{\|R(x, \varepsilon_0) - R(x', \varepsilon_0)\|_\alpha}{|x - x'|} < 1. \tag{15}$$

Under (15), Wu and Shao (2004) showed that, by iterating (14), X_n is of the form (2) for some function g. Furthermore, X_n satisfies the geometric-moment contraction property. Specifically, let $(\varepsilon_i')_{i\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$ and $\mathcal{F}_n^* = (\ldots, \varepsilon_{-1}', \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n)$ be the coupled processes of $\mathcal{F}_n$. Then

$$\|X_n - g(\mathcal{F}_n^*)\|_\alpha = O(\rho^n). \tag{16}$$

Corollary 5. *Assume (X_n) satisfies (16) for some $\alpha > 2$. Then (6) holds for $U_n = [S_n - \mathbb{E}(S_n)]/(\sigma\sqrt{n})$ with rate $\tau_n = n^{p-1}$, where $p = (\alpha \wedge 4)/2$.*

Proof. Recall the definition of $\theta_\alpha(n)$ and $\mathcal{F}_i$ in Section 2. Since $(|a| + |b|)^\alpha \le 2^\alpha(|a|^\alpha + |b|^\alpha)$ for $\alpha > 0$, by stationarity and (16), we have

$$\theta_\alpha^\alpha(n) \le 2^\alpha \Big[\|g(\mathcal{F}_n) - g(\mathcal{F}_n^*)\|_\alpha^\alpha + \|g(\mathcal{F}_n^*) - g(\mathcal{F}_n')\|_\alpha^\alpha\Big] = O(\rho^{n\alpha}), \tag{17}$$

completing the proof. $\diamondsuit$

4 A Simulation Study

In this section we shall carry out a simulation study to investigate the relationship between the accuracy of tail probabilities and the strength of dependence. Given observations $X_1, \ldots, X_n$ of a stationary process, the population mean $\mu = \mathbb{E}(X_i)$ can be estimated by the sample mean $\bar{X}_n = S_n/n$. For $\alpha \in (0, 1)$, a $1-\alpha$ level confidence interval can be constructed as $\bar{X}_n \pm z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is an estimate of long-run standard deviation σ (see Theorem 1) and $z_{1-\alpha/2}$ is the upper $(1-\alpha/2)$th quantile of a standard normal distribution. In many applications the values of α are usually small and hence it is more desirable to apply results of type (6) which provide asymptotic expansions for tail probabilities. Typical values of α are 0.01 or 0.05.

Consider the nonlinear time series model

$$X_i = \theta|X_{i-1}| + \sqrt{1-\theta^2}\varepsilon_i, \tag{18}$$

where ε_i are iid standard normals and $\theta \in (-1, 1)$. Let $\phi = \Phi'$ be the standard normal density function. The stationary distribution of (18) has a close form density function $f(u) = 2\phi(u)\Phi(\delta u)$ which corresponds to a skew-normal distribution with the skewness parameter $\delta = \theta/\sqrt{1-\theta^2}$ (Andel, Netuka and Svara, 1984). So the mean $\mu = \mathbb{E}(X_i) = \int x f(x)dx = \theta\sqrt{2/\pi}$. For $\theta = 0.1, 0.3, 0.5, 0.7$ and 0.9, the estimated long-run standard deviations $\hat{\sigma}$ are $1.01, 1.04, 1.11, 1.28$ and 1.87 respectively (Wu and Zhao, 2006).

Larger values of θ indicate higher skewness and stronger dependence. In our simulation we choose 4 levels of α: $\alpha = 0.005$, 0.01, 0.025 and 0.05 and calculate the tail probabilities $l(\alpha) = \mathbb{P}[\sqrt{n}(\bar{X}_n - \mu)/\sigma \le z_\alpha]$ and $u(\alpha) = \mathbb{P}[\sqrt{n}(\bar{X}_n - \mu)/\sigma \ge z_{1-\alpha}]$ based on 10^6 realizations of (18). Note that $z_{0.005} = -2.575829$, $z_{0.01} = -2.326348$, $z_{0.025} = -1.959964$ and $z_{0.05} = -1.644854$. The sample size $n = 200$. The ratios of tail probabilities with respect to α are displayed in Table 1. The 2nd-5th columns show the ratios of lower tail probabilities $l(\alpha)$ and α. The last four columns shows the ratios of upper tail probabilities. We say that the approximation is good if the ratio is close to 1.

Table 1 shows that, as θ increases, namely the dependence gets stronger, then the approximation becomes worse, especially when α is small. This phenomenon can be explained by our Theorem 1. If the dependence is stronger, then the martingale approximation (cf (22) and (23) in the proof of Theorem 1) becomes less accurate, and the range of the applicability of MDP is narrower. Consequently the tail probabilities are further away from their nominal levels.

Remark 4. The moderate deviation principle of type (6) provides more accurate information than the one based on the logarithmic scale. For example let $\alpha = 0.005$ and $\theta = 0.9$. Then the lower tail probabilities is $0.258 \times 0.005 = 0.00129$ and the ratio in the logarithmic scale is $\log(0.00129)/\log(0.005) = 1.255703$, which is relatively closer to 1 and it does not seem to imply that the approximation is unsatisfactory. In comparison the ratio 0.258 is far below 1.

α	0.005	0.01	0.025	0.05	0.05	0.025	0.01	0.005
θ=.9	0.258	0.423	0.667	0.857	1.374	1.679	2.239	2.877
θ=.7	0.546	0.664	0.807	0.899	1.118	1.228	1.413	1.577
θ=.5	0.714	0.777	0.867	0.923	1.048	1.087	1.167	1.248
θ=.3	0.786	0.843	0.905	0.937	1.008	1.021	1.047	1.080
θ=.1	0.906	0.936	0.948	0.964	0.995	0.994	1.001	1.004

Table 1. Ratios of tail probabilities with respect to α. The 2nd-5th columns: the ratios of lower tail probabilities $l(\alpha)$ and α. The 6th-9th columns: the ratios of upper tail probabilities $u(\alpha)$ and α.

5 Proofs

Recall that $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Define the projections $\mathcal{P}_i Z = \mathbb{E}(Z|\mathcal{F}_i) - \mathbb{E}(Z|\mathcal{F}_{i-1}), i \in \mathbb{Z}$.

Theorem 2. *Let ξ_i, $i \in \mathbb{Z}$, be a stationary Markov chain; let $Z_i = h(\xi_i)$ be a stationary process with mean* 0 *and $Z_i \in \mathcal{L}^p$, $1 < p \le 2$. Write $T_i = Z_1 + \ldots + Z_i$ and $T_n^* = \max_{i \le n} |T_i|$. Then for every non-negative integer d, we have*

$$\|T_{2^d}^*\|_p \le C_p 2^{d/p} \sum_{r=0}^{d} 2^{-r/p} \|\mathbb{E}(T_{2^r}|\xi_0)\|_p + B_p 2^{d/p} \|Z_1\|_p, \tag{19}$$

where

$$B_p = \frac{18 p^{5/3}}{(p-1)^{3/2}} \text{ and } C_p = B_p + 2^{-1/p} + B_p 2^{1-1/p}.$$

Proof. A similar inequality under $p \ge 2$ is established in Peligrad, Utev and Wu (2006). The proof for the case $1 < p < 2$ is similar. For completeness we provide the proof here. We shall apply an induction argument. Clearly (19) holds if $d = 0$. Assume that it holds for $d-1$. Let $Y_i = \mathbb{E}(Z_{2i-1}|\xi_{2i-2}) + \mathbb{E}(Z_{2i}|\xi_{2i-1})$, $W_i = Y_1 + \ldots + Y_i$ and $W_n^* = \max_{i \le n} |W_i|$. By the induction hypothesis,

$$\|W_{2^{d-1}}^*\|_p \le C_p 2^{(d-1)/p} \sum_{r=0}^{d-1} 2^{-r/p} \|\mathbb{E}(W_{2^r}|\xi_0)\|_p + B_p 2^{(d-1)/p} \|Y_1\|_p, \tag{20}$$

Let $L_j = Z_j - \mathbb{E}(Z_j|\xi_{j-1})$. Then $M_k := \sum_{j=1}^{k} L_j$ is a martingale. Observe that

$$T_{2^d}^* \le \max_{k \le 2^d} |M_k| + W_{2^{d-1}}^* + \max_{k \le 2^{d-1}} |\mathbb{E}(Z_{2k-1}|\xi_{2k-2})|.$$

By Burkholder's inequality, $\| \max_{k \le n} |M_k| \|_p \le B_p n^{1/p} \|L_1\|_p$. Hence

$$\|T_{2^d}^*\|_p \le B_p 2^{d/p} \|L_1\|_p + \|W_{2^{d-1}}^*\|_p + 2^{(d-1)/p} \|\mathbb{E}(Z_1|\xi_0)\|_p. \tag{21}$$

Note that $\mathbb{E}(W_{2^r}|\xi_0) = \mathbb{E}(T_{2^{1+r}}|\xi_0)$. Elementary calculations show that (19) follows from (20) and (21) in view of $\|L_1\|_p \le \|\mathbb{E}(Z_1|\xi_0)\|_p + \|Z_1\|_p$ and $\|\mathbb{E}(Z_1|\xi_0)\|_p \le \|Z_1\|_p$. ♢

Proof of Theorem 1. We only prove the first half of (6) since the second half similar follows. Let C_p be generic constant which may vary among lines. For notational simplicity we shall omit the subscript $2p$ and write $\theta(i)$ (resp. $\Theta(i)$ or $\psi(i)$) for $\theta_{2p}(i)$ (resp. $\Theta_{2p}(i)$ or $\psi_{2p}(i)$).

For $k \in \mathbb{Z}$ let

$$D_k = \sum_{i=k}^{\infty} \mathcal{P}_k X_i.$$

Since $\mathcal{P}_0 X_i = \mathbb{E}(X_i - X_i' | \mathcal{F}_0)$, by Jensen's inequality, $\|\mathcal{P}_0 X_i\|_{2p} \le \theta(i)$, which by the condition $\Theta_{2p}(0) < \infty$ implies that $D_0 \in \mathcal{L}^{2p}$. Note that D_k, $k \in \mathbb{Z}$, are stationary and ergodic martingale differences with respect to $\mathcal{F}_k$, and $\lim_{n\to\infty} \|S_n\|/\sqrt{n} = \sigma = \|D_0\|$ (cf Theorem 1 in Wu (2005a)). Define

$$M_k = \sum_{i=1}^{k} D_i \quad \text{and} \quad R_k = S_k - M_k. \tag{22}$$

By Theorem 1 (ii) in Wu (2005a), there exists a positive constant C_p such that

$$\|R_n\|_{2p}^2 \le C_p \sum_{i=1}^{n} \left[\sum_{j=i}^{\infty} \|\mathcal{P}_0 X_j\|_{2p} \right]^2 \le C_p \sum_{i=1}^{n} \Theta^2(i). \tag{23}$$

Let

$$\Lambda(n) = \left[n^{-1} \sum_{i=1}^{n} \Theta^2(i) \right]^p$$

and

$$\epsilon = \frac{[x\Lambda(n)]^{1/(1+2p)}}{1 + r_x}. \tag{24}$$

Since $S_n = M_n + R_n$, by the triangle and Markov's inequalities, we have

$$\begin{aligned} \mathbb{P}(M_n \ge \sqrt{n}\sigma(r_x + \epsilon)) &\le \mathbb{P}(|R_n| \ge \sqrt{n}\sigma\epsilon) + \mathbb{P}(S_n \ge \sqrt{n}\sigma r_x) \\ &\le \frac{\|R_n\|_{2p}^{2p}}{(\sqrt{n}\sigma\epsilon)^{2p}} + \mathbb{P}(S_n \ge \sqrt{n}\sigma r_x) \\ &\le \frac{C_p^p \Lambda(n)}{(\sigma\epsilon)^{2p}} + \mathbb{P}(S_n \ge \sqrt{n}\sigma r_x). \end{aligned} \tag{25}$$

Similarly,

$$\mathbb{P}(S_n \geq \sqrt{n}\sigma r_x) \leq \mathbb{P}(M_n \geq \sqrt{n}\sigma(r_x - \epsilon)) + \frac{C_p^p \Lambda(n)}{(\sigma\epsilon)^{2p}}. \tag{26}$$

Observe that $\Lambda(n) = O(n^{-p})$ if $\alpha > 1/2$, $\Lambda(n) = O[(n^{-1}\log n)^p]$ if $\alpha = 1/2$ and $\Lambda(n) = O(n^{-2\alpha p})$ if $\alpha < 1/2$. Since $\alpha \leq \beta \leq 1/2 + \alpha$, simple calculations show that $\tau_n \Lambda(n) \to 0$ for all three cases $\eta > 1 - 1/p$, $\eta = 1 - 1/p$ or $\eta < 1 - 1/p$. Hence $\epsilon(1 + r_x) \to 0$ as $n \to \infty$ uniformly in $x \in (1, \alpha\tau_n]$. Note that $1 - \Phi(t) \geq \phi(t)/(1 + t)$, $t > 0$. Then

$$\frac{1 - \Phi(r_x - \epsilon)}{1 - \Phi(r_x)} - 1 \leq \frac{\epsilon\phi(r_x - \epsilon)}{1 - \Phi(r_x)} = O(\epsilon(1 + r_x)e^{\epsilon r_x}) = O(\epsilon(1 + r_x)). \tag{27}$$

To deal with M_n, we shall now apply the moderate deviation principle for martingales (cf Theorem 2.1 in Grama (1997), see also Theorem 1 in Grama and Haeusler (2006)). A major difficulty and key step in applying their result is to find a bound for

$$I_n = \sum_{i=1}^{n} \frac{\|D_i\|_{2p}^{2p}}{n^p} + \|V_n/n - \sigma^2\|_p^p = n^{1-p}\|D_0\|_{2p}^{2p} + \|V_n/n - \sigma^2\|_p^p, \tag{28}$$

where V_n is the sum of conditional variances or quadratic characteristic

$$V_n = \sum_{i=1}^{n} \mathbb{E}(D_i^2 | \mathcal{F}_{i-1}). \tag{29}$$

Interestingly, with our physical dependence measure (3), a bound with simple and explicit form can be found. To this end, by Propositions 2 and 3 in Wu (2005a), there exists a constant $C_p > 0$ such that

$$\|\mathbb{E}(D_m^2|\mathcal{F}_0) - \sigma^2\|_p \leq \Theta(0)C_p\psi^{1/2}(m) + \Theta(0)C_p\sum_{i=m}^{\infty}\min[\psi^{1/2}(i+1),\, \theta(i-m+1)].$$

Let $m_1 = \lfloor m^{\beta/(1+\alpha)} \rfloor$. By (7) and (8),

$$\sum_{i=m}^{\infty}\min[\psi^{1/2}(i+1),\, \theta(i-m+1)] \leq \sum_{i=m}^{m+m_1}\psi^{1/2}(i+1) + \sum_{i=m+m_1+1}^{\infty}\theta(i-m+1) = O(m^{-\eta}).$$

Therefore $\|\mathbb{E}(D_m^2|\mathcal{F}_0) - \sigma^2\|_p = O(m^{-\eta})$ and by the triangle inequality, $\|\mathbb{E}(V_m|\mathcal{F}_0) - m\sigma^2\|_p = \sum_{i=1}^m O(i^{-\eta})$. Apply Theorem 2 with $\xi_i = \mathcal{F}_{i-1}$ and $Z_i = V_i - i\sigma^2$, elementary manipulations show that $\|V_n - n\sigma^2\|_p = O(n^{1/p})$ if $\eta > 1-1/p$, $\|V_n - n\sigma^2\|_p = O(n^{1/p}\log n)$ if $\eta = 1-1/p$ and $\|V_n - n\sigma^2\|_p = O(n^{1-\eta})$ if $\eta < 1-1/p$. Combining these three cases, we have $I_n = O(\tau_n^{-1})$.

By Theorem 2.1 in Grama (1997), since $x_{r_x-\epsilon}/x = 1 + O((1+r_x)\epsilon)$, there exists a constant C independent of x and n such that

$$\left|\frac{\mathbb{P}(M_n \geq \sqrt{n}\sigma(r_x-\epsilon))}{1-\Phi(r_x-\epsilon)} - 1\right| \leq C(xI_n)^{1/(1+2p)} \tag{30}$$

holds uniformly in $x \in [1, a\tau_n]$. Clearly the above relation also holds with $r_x - \epsilon$ replaced by $r_x + \epsilon$. By (26), (27) and (30),

$$\begin{aligned}\frac{\mathbb{P}(S_n \geq \sqrt{n}\sigma r_x)}{1-\Phi(r_x)} - 1 &\leq \frac{\mathbb{P}(M_n \geq \sqrt{n}\sigma(r_x-\epsilon))}{1-\Phi(r_x)} - 1 + \frac{\Lambda(n)}{(1-\Phi(r_x))(\sigma\epsilon)^{2p}} \\ &= O\{(xI_n)^{1/(1+2p)}\} + O(\epsilon(1+r_x)) + \frac{O(\Lambda(n))(1+r_x)}{\phi(r_x)\epsilon^{2p}}.\end{aligned}$$

A lower bound for $\mathbb{P}(S_n \geq \sqrt{n}\sigma r_x)/[1-\Phi(r_x)]$ can be similarly obtained. Therefore Theorem 1 follows in view of the choice of ϵ in (24). $\diamondsuit$

References

Andel, J., Netuka, I. and Svara, K. (1984) On threshold autoregressive processes. *Kybernetika* **20**, 89–106.

Amosova, N.N. (1972) Limit theorems for the probabilities of moderate deviations. (Russian) *Vestnik Leningrad. Univ.* **13** 5–14.

Amosova, N.N. (1982) Probabilities of moderate deviations. *J. Math. Sci.* Springer-Verlag, New York, **20** 2123–2130.

Babu, G.J. and Singh, K. (1978) Probabilities of moderate deviations for some stationary strong-mixing processes. *Sankhya Ser. A* **40** 38–43.

Bose, A. (1986) Certain nonuniform rates of convergence to normality for martingale differences. *J. Statist. Plan. Inf.* **14** 155–167.

Borkar, V.S. (1993) White-noise representations in stochastic realization theory. *SIAM J. Control Optim.* **31** 1093–1102.

Chen, X. (2001) Moderate deviations for Markovian occupation times. *Stoch. Proc. Appl.* **94** 51–70.

Dembo, A. (1996) Moderate deviations for martingales with bounded jumps. *Electron. Comm. Probab.* **1** 11–17.

Dembo, A. and Zeitouni, O. (1993) *Large Deviations Techniques and Applications.* Jones and Barlett, Boston, MA.

Dong, Z.S.; Tan, X.L. and Yang, X.Y. (2005) Moderate deviation principles for moving average processes of real stationary sequences. *Statist. Probab. Lett.* **74** 139–150.

De Acosta, A. (1992) Moderate Deviations and Associated Laplace Approximations for Sums of Independent Random Vectors. *Trans. Amer. Math. Soc.* **329** 357–375.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50** 987–1007.

Frolov, A.N. (1998) On one-sided strong laws for large increments of sums. *Statist. Probab. Lett.* **37** 155–165.

Frolov, A.N. (2005) On probabilities of moderate deviations of sums of independent random variables. *J. Math. Sci.* Springer-Verlag, New York, **127** 1787–1796.

Gao, F.Q (1996) Moderate deviations for martingales and mixing random processes. *Stoch. Proc. Appl.* **61** 263–275.

Ghosh, M. (1974) Probabilities of moderate deviations under m-dependence. *Canad. J. Statist.* **2** 157–168.

Ghosh, M. and Babu, G.J. (1977) Probabilities of moderate deviations for some stationary ϕ-mixing processes. *Ann. Probab.* **5** 222–234.

Grama, I. G. (1997) On moderate deviations for martingales. *Ann. Probab.* **25** 152–183.

Grama, I.G. and Haeusler, E. (2006) An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *J. Theo. Probab.* **19** 1–44.

Haggan, V. and Ozaki, T. (1981) Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, *Biometrika* **68** 189–196.

Kallianpur, G. (1981) Some ramifications of Wieners ideas on nonlinear prediction. In *Norbert Wiener, Collected Works with Commentaries* (MIT Press, MA), pp. 402-424.

Michel, R. (1976) Nonuniform central limit bounds with application to probabilities of deviations. *Ann. Probab.* **4** 102–106.

Nicholls, D.F. and Quinn, B.G. *Random coefficient autoregressive models: An introduction.* Springer-Verlag, New York, 1982.

Osipov, L.V. (1972) On probabilities of large deviations for sums of independent random variables. *Theory. Prob. Appl.* **17** 309-331.

Peligrad, M.; Utev, S. and Wu, W. B. (2005) A maximal L_p-Inequality for stationary sequences and application. *To appear in Proc. AMS.*

Petrov, V.V. (1975) *Sums of independent random variables* Springer, New York.

Petrov, V.V. (2002) On probabilities of moderate deviations. *J. Math. Sci.* **109** 2189–2191.

Račkauskas, A. (1995) Large deviations for martingales with some applications. *Acta Appl. Math.* **38** 109–129.

Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behavior.* Springer, New York.

Rubin, H. and Sethuraman, J. (1965) Probabilities of moderate deviations. *Sankhya Ser. A* **37** 325–346.

Slastnikov, A.D. (1978) Limit theorems for moderate deviation probabilities. *Theory Prob. Appl.* **23** 322–340.

Solomyak, B.M. (1995) On the random series $\sum \pm\lambda^n$ (an Erdös problem). *Ann. of Math.* **142** 611–625.

Tong, H. (1990) *Nonlinear time series Analysis: a dynamic approach.* Oxford University Press, Oxford, 1990.

Wiener, N. (1958) *Nonlinear Problems in Random Theory.* MIT Press, MA.

Wu, W.B. (2005a) On strong convergence for sums of dependent random variables. Preprint.

Wu, W.B. (2005b) Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci.* **102,** 14150-14154.

Wu, W.B. and Shao, X. (2004) Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425-436.

Wu, W.B. and Zhao, Z. (2006) Inference of trends in time series. Preprint.

Department of Statistics
The University of Chicago
5734 S. University Avenue, Chicago, IL, 60637
E-mail: wbwu@galton.uchicago,edu, zhao@galton.uchicago.edu