Statistica Sinica **18**(2008), 769-782

MODERATE DEVIATIONS FOR STATIONARY PROCESSES

Wei Biao Wu and Zhibiao Zhao

University of Chicago and Pennsylvania State University

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.

Key words and phrases: Martingale, moderate deviation, 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 the CLT. For the special case in which the X_i are independent and identically distributed (iid), one has the following classical result. Let $c > 0$. Assume that $\mathrm{E}(|X_1|^q) < \infty$ for some $q > c^2 + 2$ and let $\sigma > 0$ be the standard deviation of X_1. Then

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

where Φ is the standard normal distribution function. The moderate deviation principle of type (1) has been investigated by Osipov (1972), Michel (1976) and Amosova (1982) for iid random variables, and by Rubin and Sethuraman (1965), Amosova (1972), Petrov (2002), and Frolov (2005) for arrays of independent random variables.

It is a challenging problem to establish moderate deviation results for dependent random variables. Ghosh (1974) obtained an MDP for 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 averages. 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.

In this paper we study asymptotic properties of the probability $\mathbb{P}(S_n/\sigma \geq \sqrt{n}r_n)$ itself instead of the one based on the logarithmic scale. In particular, we obtain an asymptotic expansion for $\mathbb{P}(S_n/\sigma \geq \sqrt{n}r_n)$ for stationary causal processes of 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) is quite general and it includes many linear processes and nonlinear time series models; see Section 3.2 and Wu and Shao (2004).

We now introduce some notation. Let $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. For a random variable Z write $Z \in \mathcal{L}^p$, $p > 0$, if $\|Z\|_p := [\mathrm{E}(|Z|^p)]^{1/p} < \infty$, and $\|Z\| = \|Z\|_2$. For $a, b \in \mathbb{R}$, let $a \wedge b = \min(a, b)$. For two real sequences $\{a_n\}$ and $\{b_n\}$, write $a_n = O(b_n)$ if $\limsup_{n\to\infty} |a_n/b_n| < \infty$, 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. Section 3 provides applications to linear processes and nonlinear time series. 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 (2005), 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 $\mathcal{F}'_i = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_i)$ the coupled version of $\mathcal{F}_i$. Assume $X_i \in \mathcal{L}^q$, $q > 0$, and define

$$\theta_q(i) = \|X_i - X'_i\|_q, \text{ where } X'_i = g(\mathcal{F}'_i). \tag{3}$$

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 (2005) 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 the 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 (2005), it is called the strong stability condition. If (4) is violated, then S_n may have a non-Gaussian limiting distribution with a non-$\sqrt{n}$ convergence rate; see for example Ho and Hsing (1997).

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

$$x = (1 + r_x)^{\nu(p)} \exp\left(\frac{r_x^2}{2}\right), \text{ where } \nu(p) = \begin{cases} 2p + 1 & \text{if } p \in (1, \frac{3}{2}]; \\ 6p - 3 & \text{if } p \in (\frac{3}{2}, 2]. \end{cases} \tag{5}$$

We also write $x_r = (1 + r)^{\nu(p)} \exp(r^2/2)$. The function $\nu(p)$ results from martingale moderate deviations; see Theorem 2 and Remark 5. 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 MDP with rate* τ_n *and exponent* $p > 0$ if, for every $a > 0$, there exists a constant $C = C_{a,p}$, independent of x and n, such that

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

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

$$\mathbb{P}(U_n \geq r_x) = [1 - \Phi(r_x)]\Big\{1 + O\Big[\Big(\frac{x}{\tau_n}\Big)^{\frac{1}{(1+2p)}}\Big]\Big\} = \frac{\exp\left(\frac{-r_x^2}{2}\right)}{r_x\sqrt{2\pi}}\{1 + o(1)\}$$

as $x \to \infty$, with $x = o(\tau_n)$. Following Remark 1 in Grama and Haeusler (2006), as $x \to \infty$, r_x has the asymptotic expansion $r_x^2 = 2\log x - [2\nu(p) + o(1)]\log(1 + \sqrt{2\log x})$.

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* $0 < \alpha \le \beta \le \alpha + 1/2$ *such that the following conditions hold:*

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

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

Let $\eta = \alpha\beta/(1+\alpha)$. *Then* $S_n/(\sigma\sqrt{n})$ *satisfies MDP 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, and exponent* p.

Remark 1. Throughout the paper we assume $\sigma > 0$. If $\sigma = 0$, then $S_n/\sqrt{n} \to 0$ in probability and has a degenerate limiting distribution. One way out is to consider $S_n/\|S_n\|$. It is unclear how to establish an MDP for $S_n/\|S_n\|$.

Remark 2. Clearly (7) implies (8) if $\alpha \ge \beta$. On the other hand, if $\beta \ge \alpha + 1/2$, then (8) implies (7). To see this, by Schwarz's inequality for $k \in \mathbb{N}$, $\sum_{i=k}^{2k-1} \theta_{2p}(i) \le [k \sum_{i=k}^{2k-1} \theta_{2p}^2(i)]^{1/2} = O(k^{1/2-\beta})$. So (7) follows by summing up the latter inequality over $k = 2^r m$, $r = 0, 1, \ldots$. Hence the condition $\alpha \le \beta \le \alpha + 1/2$ 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 either* [i] (7) *holds for some*

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

or [ii] (8) *holds for some*

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

Then $S_n/(\sigma\sqrt{n})$ *satisfies MDP with rate* $\tau_n = n^{p-1}$ *and exponent* p.

Proof. Let η be as in Theorem 1. [i] If (7) holds, then (8) holds with $\beta = \alpha$. Observe that (9) implies $\eta = \alpha^2/(1+\alpha) > 1 - 1/p$. [ii] 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 $\eta = \beta(\beta - 1/2)/(\beta + 1/2) > 1 - 1/p$. By Theorem 1, Corollary 1 follows.

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 calculate $\theta_q(i)$ for functionals of linear processes and some 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 $\mathrm{E}(\varepsilon_0) = 0$ if $q \geq 1$. Let a_i be real numbers satisfying $\sum_{i=0}^{\infty} |a_i|^{q\wedge 2} < \infty$. By Kolmogorov's Three Series Theorem, the linear process

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

is well defined and strictly stationary (cf., Corollary 5.1.3 in Chow and Teicher (2003)). Let $0 < \varsigma \leq 1$ and $\upsilon \geq 0$; let $\mathcal{H}(\varsigma, \upsilon)$ be the collection of functions h such that

$$|h(x) - h(x')| \leq c|x - x'|^{\varsigma}(1 + |x| + |x'|)^{\upsilon}, \quad x, x' \in \mathbb{R}, \tag{12}$$

where $c = c_{h,\varsigma,\upsilon}$ is a constant independent of x and x'. Clearly, $\mathcal{H}(\varsigma, 0)$ corresponds to globally Hölder-continuous functions with index ς. If $h(x) = |x|^b$, $b > 1$, then $h \in \mathcal{H}(1, b-1)$. Let $h \in \mathcal{H}(\varsigma, \upsilon)$ and consider

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

Assume $h(Y_0) \in \mathcal{L}^{2p}$, where $p \in (1, 2]$ satisfies $2p(\varsigma + \upsilon) \leq q$. Then either $2p\varsigma < q$ or $2p\upsilon < q$. If $2p\varsigma < q$, let $\varrho = q/(2p\varsigma)$ and $\varrho' = \varrho/(\varrho - 1)$, then $2p\upsilon\varrho' \leq q$. Recall $\varepsilon_0, Y_0 \in \mathcal{L}^q$. Define $Y_i' = Y_i + a_i(\varepsilon_0' - \varepsilon_0)$. Then $\theta_{2p}(i) = O(|a_i|^{\varsigma})$ since, by Hölder's inequality,

$$\begin{aligned} \theta_{2p}^{2p}(i) = \|h(Y_i) - h(Y_i')\|_{2p}^{2p} &\leq c^{2p}\Big\||Y_i - Y_i'|^{2p\varsigma}(1 + |Y_i| + |Y_i'|)^{2p\upsilon}\Big\|_1 \\ &\leq c^{2p}\Big\||Y_i - Y_i'|^{2p\varsigma}\Big\|_{\varrho} \times \Big\|(1 + |Y_i| + |Y_i'|)^{2p\upsilon}\Big\|_{\varrho'} = O(|a_i|^{2p\varsigma}). \end{aligned} \tag{13}$$

If $2p\upsilon < q$, then (13) holds with $\varrho' = q/(2p\upsilon)$ and $\varrho = \varrho'/(\varrho' - 1)$. Corollary 1 [ii] entails

Corollary 2. *Let $\varepsilon_0 \in \mathcal{L}^q$, $q > 0$, and $a_i = O(i^{-\gamma})$ for some $\gamma > 0$. Assume $h \in \mathcal{H}(\varsigma, \upsilon)$, $\varsigma \in (0, 1]$, $\upsilon \geq 0$ and $h(Y_0) \in \mathcal{L}^{2p}$ for some $p \in (1, 2]$. Further assume that*

$$2p(\varsigma + \upsilon) \leq q \text{ and } \gamma\varsigma > \frac{5p - 2 + \sqrt{17p^2 - 20p + 4}}{4p}. \tag{14}$$

Then $S_n/(\sigma\sqrt{n})$ *satisfies MDP with rate* $\tau_n = n^{p-1}$ *and exponent* p.

If $h(x) = x$, then $\varsigma = 1$ and $\upsilon = 0$. Let $\varepsilon_0 \in \mathcal{L}^{2p}$, $p \in (1, 2]$, and $\mathrm{E}(\varepsilon_0) = 0$. Then σ has a closed form: $\sigma = |\sum_{i=0}^{\infty} a_i| \|\varepsilon_0\|$ (cf., (24)). Assume $\sigma > 0$ and that (14) holds with $q = 2p$. Then Corollary 2 is applicable.

Example 1. Let $q \geq 4$ and $a_i = O(i^{-\gamma})$, $\gamma > 1 + \sqrt{2}/2$. Let $h(Y_0) \in \mathcal{L}^4$ and h be Lipschitz continuous. Then $S_n/(\sigma\sqrt{n})$ satisfies MDP with rate $\tau_n = n$ and exponent 2.

Example 2. Consider the AR(r) model $Y_n = b_1 Y_{n-1} + b_2 Y_{n-2} + \cdots + b_r Y_{n-r} + \varepsilon_n$. Assume that $1 - b_1 x - b_2 x^2 - \cdots - b_r x^r \neq 0$ for all $|x| \leq 1$. Then Y_n is of the form (11) with the coefficients $a_i = O(\lambda^i)$ for some $|\lambda| < 1$. Let $\varepsilon_0 \in \mathcal{L}^q$ for some $q > 0$ and $h \in \mathcal{H}(\varsigma, \upsilon)$ such that $h(Y_0) \in \mathcal{L}^4$. If $q \geq 4(\varsigma + \upsilon)$, then $S_n/(\sigma\sqrt{n})$ satisfies MDP with rate $\tau_n = n$ and exponent 2. A similar example is considered in Grama and Haeusler (2006). Comparing with their method, our approach is simpler and it allows for functionals of AR(r) processes. 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 $h_x(\cdot) = \mathbf{1}_{\cdot \leq x}$. Stronger conditions on γ and ε_i are needed.

Corollary 3. *Let* $a_i = O(i^{-\gamma})$, $\gamma > 0$. *Assume either* [i] $\varepsilon_0 \in \mathcal{L}^1$, $\gamma > 4 + 2\sqrt{2}$ *and* ε_0 *has a bounded density, or* [ii] $\varepsilon_0 \in \mathcal{L}^q$, $\gamma(q \wedge 4) > 20 + 10\sqrt{2}$ *and* Y_0 *has a Lipschitz-continuous distribution function. Then* $S_n/(\sigma\sqrt{n})$ *satisfies MDP with rate* $\tau_n = n$ *and exponent* 2.

Proof. By Corollary 1[ii], it suffices to verify (8) for some $\beta > (1 + \sqrt{2})/2$ and $p = 2$.

[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. Clearly, $\theta_4(0) \leq 1$. Let $i \in \mathbb{N}$. If $a_i = 0$, then $Y_i' = Y_i$ and $\theta_4(i) = 0$. If $a_i \neq 0$, since $\varepsilon_0 \in \mathcal{L}^1$, we have

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

Observe that

$$\mathrm{E}\left|\mathbf{1}_{\varepsilon_i + a_i\varepsilon_0 \leq x} - \mathrm{E}\left(\mathbf{1}_{\varepsilon_i + a_i\varepsilon_0 \leq x} | \varepsilon_i\right)\right| = 2\mathrm{E}\left\{F_\varepsilon\Big(\frac{x - \varepsilon_i}{a_i}\Big)\Big[1 - F_\varepsilon\Big(\frac{x - \varepsilon_i}{a_i}\Big)\Big]\right\}$$

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

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

uniformly in x. By independence, the preceding relation implies $\sup_x \mathrm{E}\,|\mathbf{1}_{Y_i\le x} - \mathbf{1}_{Y_i'\le x}| = O(|a_i|)$. Hence $\theta_4(i) = O(|a_i|^{1/4})$ and (8) is satisfied with $\beta = \gamma/4-1/2$.

[ii] Let i be fixed. Define $\omega(u) = \mathbf{1}_{u\le x} + \mathbf{1}_{x<u<x+\lambda}(x+\lambda-u)/\lambda$. Then $\omega(\cdot)$ is bounded and Lipschitz continuous with Lipschitz constant $1/\lambda$. By the triangle inequality,

$$\begin{aligned}\|\mathbf{1}_{Y_i\le x} - \mathbf{1}_{Y_i'\le x}\|_4 &\le \|\omega(Y_i) - \omega(Y_i')\|_4 + 2\|\mathbf{1}_{Y_i\le x} - \omega(Y_i)\|_4 \\ &= O\left(\frac{|a_i|^{\frac{(q\wedge 4)}{4}}}{\lambda} + \lambda^{\frac{1}{4}}\right)\end{aligned}$$

in view of the Lipschitz continuity of the distribution function of Y_i. Let $\lambda = i^{-\gamma(q\wedge 4)/5}$. Then $\theta_4(i) = O(i^{-\gamma(q\wedge 4)/20})$ and (8) holds with $\beta = \gamma(q\wedge 4)/20-1/2$.

Remark 3. Let $\varepsilon_0 \in \mathcal{L}^q$. If $q \ge 1$, 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 has uniform(0, 1) as its 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 3 are satisfied and the moderate deviation principle (6) holds for $U_n = S_n/(\sigma\sqrt{n})$ with rate $\tau_n = n$ and exponent 2.

3.2. Nonlinear time series

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

$$X_n = R(X_{n-1}, \varepsilon_n), \tag{15}$$

where $R(\cdot, \varepsilon)$ is a measurable random map. Many popular nonlinear time series models are of the form (15), including the TAR(1) model $X_n = aX_{n-1}^+ + bX_{n-1}^- + \varepsilon_n$, the ARCH model $X_n = \varepsilon_n(a^2 + b^2X_{n-1}^2)^{1/2}$, and the EAR model $X_n = [a + b\exp(-cX_{n-1}^2)]X_{n-1} + \varepsilon_n$, among others. Assume that there exist x_0 and $\alpha > 0$ such that $R(x_0, \varepsilon_0) \in \mathcal{L}^\alpha$ and

$$\rho := \sup_{x\ne x'} \frac{\|R(x,\varepsilon_0) - R(x',\varepsilon_0)\|_\alpha}{|x-x'|} < 1. \tag{16}$$

Under (16), Wu and Shao (2004) showed that, by iterating (15), X_n is of the form (2) for some function g. Furthermore, X_n satisfies the following property: 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

$$\|X_n - g(\mathcal{F}_n^*)\|_\alpha = O(\rho^n). \tag{17}$$

Following Wu (2005), (17) implies $\theta_\alpha(n) = O(\rho^n)$. By Corollary 1, we have

Corollary 4. *Assume* (X_n) *satisfies* (17) *for some* $\alpha > 2$. *Let* $p = (\alpha \wedge 4)/2$. *Then* $[S_n - \mathrm{E}(S_n)]/(\sigma\sqrt{n})$ *satisfies MDP with rate* $\tau_n = n^{p-1}$ *and exponent* p.

4. A Simulation Study

In this section we carry out a simulation study to investigate the relationship between the accuracy of tail probabilities and the strength of dependence. Given observations $(X_i)_{1\le i\le n}$ of a stationary process, the population mean $\mu = \mathrm{E}(X_i)$ can be estimated by the sample mean $\bar{X}_n = S_n/n$. For $\alpha \in (0,1)$, a $100(1-\alpha)\%$ 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 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 $\mu = \mathrm{E}(X_i) = \int xf(x)dx = \theta\sqrt{2/\pi}$. 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 (2007)).

Larger values of θ indicate higher skewness and stronger dependence. For 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.

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.

α	0.005	0.01	0.025	0.05	0.05	0.025	0.01	0.005
θ=0.9	0.258	0.423	0.667	0.857	1.374	1.679	2.239	2.877
θ=0.7	0.546	0.664	0.807	0.899	1.118	1.228	1.413	1.577
θ=0.5	0.714	0.777	0.867	0.923	1.048	1.087	1.167	1.248
θ=0.3	0.786	0.843	0.905	0.937	1.008	1.021	1.047	1.080
θ=0.1	0.906	0.936	0.948	0.964	0.995	0.994	1.001	1.004

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 Theorem 1. If the dependence is stronger, then the martingale approximation (cf., (25) and (26) 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 MDP 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 ratio 0.258 is far below 1. In comparison, the lower tail probability is $0.258 \times 0.005 = 0.00129$ and the ratio in the logarithmic scale is $\log(0.00129)/\log(0.005) = 1.2557$, which is relatively closer to 1, and it does not seem to imply that the approximation is unsatisfactory.

5. Proofs

To prove Theorem 1, we need the following Theorems 2 and 3. The former is adapted from Grama (1997) and Grama and Haeusler (2006), hereafter abbreviated as G97 and GH06, respectively. The latter is a variant of the maximal inequality given in Peligrad, Utev and Wu (2007), where the case $p \geq 2$ is dealt with.

Theorem 2. *Let $\xi_{n,k} \in \mathcal{L}^{2p}$, $1 < p \leq 2$, be martingale differences with respect to the filtration $\mathcal{F}_{n,k}$, $1 \leq k \leq n$, and let $\Xi_n = \sum_{k=1}^n \xi_{n,k}$. Define*

$$L_p^n = \sum_{k=1}^n \mathrm{E}\,|\xi_{n,k}|^{2p} \quad \text{and} \quad N_p^n = \mathrm{E}\left|\sum_{k=1}^n \mathrm{E}\,(\xi_{n,k}^2|\mathcal{F}_{n,k}) - 1\right|^p. \tag{19}$$

Let r_x be the solution to the equation (5). Then for every $a > 0$ there exists a constant $C_{a,p}$, depending only on a and p, such that uniformly over $x \in [1, a(L_p^n +$

$N_p^n)^{-1}]$,

$$\max\left\{\left|\frac{\mathbb{P}(\Xi_n \geq r_x)}{1-\Phi(r_x)}-1\right|,\ \left|\frac{\mathbb{P}(\Xi_n \leq -r_x)}{\Phi(-r_x)}-1\right|\right\} \leq C_{a,p}[x(L_p^n+N_p^n)]^{\frac{1}{(2p+1)}}. \quad (20)$$

Remark 5. Under $p \in (1,2]$, G97 proved (20) for r_x satisfying $x = (1+r_x)^{6p-3}\exp(r_x^2/2)$. If $p \in (1,3/2]$, GH06 obtained (20) for r_x satisfying $x = (1+r_x)^{2p+1}\exp(r_x^2/2)$, thus improving G97's result by allowing a wider range of r_x since $6p-3 > 2p+1$. GH06 argued that the exponent $2p+1$ is optimal if $p \in (1,3/2]$. However, if $p \in (3/2,2]$, then in certain applications the G97 result allows a wider range of r_x, while for GH06 the higher moment property $p \in (3/2,2]$ is not advantageous since one can only use $p = 3/2$. For example, in the application to nonlinear time series in Section 3.2, if (17) holds with $\alpha = 4$, then Corollary 4 asserts an MDP with $\tau_n = n^{p-1} = n$ since $p = 2$. In comparison, using GH06, one can only obtain the narrower range with $\tau_n = n^{3/2-1} = n^{1/2}$.

Theorem 3. *Let ξ_i, $i \in \mathbb{Z}$, be a stationary Markov chain; let $Z_i = h(\xi_i)$ be a stationary process with zero mean and $Z_i \in \mathcal{L}^p$, $1 < p \leq 2$. Write $T_i = Z_1 + \cdots + Z_i$ and $T_n^* = \max_{i \leq n}|T_i|$. Then for every non-negative integer d, we have*

$$\|T_{2^d}^*\|_p \leq C_p 2^{\frac{d}{p}} \sum_{r=0}^{d} 2^{-\frac{r}{p}} \|\mathrm{E}\,(T_{2^r}|\xi_0)\|_p + B_p 2^{\frac{d}{p}}\|Z_1\|_p, \quad (21)$$

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

Proof. We apply an induction argument. Clearly (21) holds if $d = 0$. Assume that it holds for $d-1$. Let $Y_i = \mathrm{E}\,(Z_{2i-1}|\xi_{2i-2}) + \mathrm{E}\,(Z_{2i}|\xi_{2i-1})$, $W_i = Y_1 + \cdots + Y_i$ and $W_n^* = \max_{i \leq n}|W_i|$. By the induction hypothesis,

$$\|W_{2^{d-1}}^*\|_p \leq C_p 2^{\frac{(d-1)}{p}} \sum_{r=0}^{d-1} 2^{-\frac{r}{p}} \|\mathrm{E}\,(W_{2^r}|\xi_0)\|_p + B_p 2^{\frac{(d-1)}{p}}\|Y_1\|_p. \quad (22)$$

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

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

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

$$\|T_{2^d}^*\|_p \leq B_p 2^{\frac{d}{p}}\|L_1\|_p + \|W_{2^{d-1}}^*\|_p + 2^{\frac{(d-1)}{p}}\|\mathrm{E}\,(Z_1|\xi_0)\|_p. \quad (23)$$

Note that $\mathrm{E}\,(W_{2^r}|\xi_0) = \mathrm{E}\,(T_{2^{1+r}}|\xi_0)$. Elementary calculations show that (21) follows from (22) and (23) in view of $\|L_1\|_p \leq \|\mathrm{E}\,(Z_1|\xi_0)\|_p + \|Z_1\|_p$ and $\|\mathrm{E}\,(Z_1|\xi_0)\|_p \leq \|Z_1\|_p$.

Proof of Theorem 1. We only prove the first half of (6) since the second half follows similarly. Let C_p be a generic constant which may vary among lines. For notational simplicity we omit the subscript $2p$ and write $\theta(i)$ (resp. $\Theta(i)$ or $\psi(i)$) for $\theta_{2p}(i)$ (resp. $\Theta_{2p}(i)$ or $\psi_{2p}(i)$). Recall $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. For $k \in \mathbb{Z}$ let

$$D_k = \sum_{i=k}^{\infty} \mathcal{P}_k X_i, \quad \text{where} \quad \mathcal{P}_k Z = \mathrm{E}\,(Z|\mathcal{F}_k) - \mathrm{E}\,(Z|\mathcal{F}_{k-1}). \tag{24}$$

Since $\mathcal{P}_0 X_i = \mathrm{E}\,(X_i - X_i'|\mathcal{F}_0)$, Jensen's inequality has $\|\mathcal{P}_0 X_i\|_{2p} \le \theta(i)$ which, by the condition $\Theta_{2p}(0) < \infty$, implies that $D_0 \in \mathcal{L}^{2p}$. Note that the 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 (2007)). Define

$$M_k = \sum_{i=1}^{k} D_i \quad \text{and} \quad R_k = S_k - M_k. \tag{25}$$

By Theorem 1(ii) in Wu (2007), there exists a positive constant C_p such that

$$\|R_n\|_{2p}^2 \le C_p \sum_{i=1}^{n} \Big[\sum_{j=i}^{\infty} \|\mathcal{P}_0 X_j\|_{2p}\Big]^2 \le C_p \sum_{i=1}^{n} \Theta^2(i). \tag{26}$$

Let

$$\Lambda(n) = \Big[n^{-1} \sum_{i=1}^{n} \Theta^2(i)\Big]^p \quad \text{and} \quad \epsilon_x = \frac{[x\Lambda(n)]^{\frac{1}{(1+2p)}}}{1 + r_x}. \tag{27}$$

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

$$\begin{aligned}
\mathbb{P}(M_n \ge \sqrt{n}\sigma(r_x + \epsilon_x)) &\le \mathbb{P}(|R_n| \ge \sqrt{n}\sigma\epsilon_x) + \mathbb{P}(S_n \ge \sqrt{n}\sigma r_x) \\
&\le \frac{\|R_n\|_{2p}^{2p}}{(\sqrt{n}\sigma\epsilon_x)^{2p}} + \mathbb{P}(S_n \ge \sqrt{n}\sigma r_x) \\
&\le \frac{C_p^p \Lambda(n)}{(\sigma\epsilon_x)^{2p}} + \mathbb{P}(S_n \ge \sqrt{n}\sigma r_x).
\end{aligned} \tag{28}$$

Similarly,

$$\mathbb{P}(S_n \ge \sqrt{n}\sigma r_x) \le \mathbb{P}(M_n \ge \sqrt{n}\sigma(r_x - \epsilon_x)) + \frac{C_p^p \Lambda(n)}{(\sigma\epsilon_x)^{2p}}. \tag{29}$$

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 $\alpha \le \beta \le 1/2 + \alpha$, 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 $\epsilon_x(1+r_x) \to 0$ 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

$$\frac{1-\Phi(r_x-\epsilon_x)}{1-\Phi(r_x)} - 1 \leq \frac{\epsilon_x \phi(r_x-\epsilon_x)}{1-\Phi(r_x)} = O(\epsilon_x(1+r_x)e^{\epsilon_x r_x}) = O(\epsilon_x(1+r_x)). \tag{30}$$

To deal with M_n, we now apply Theorem 2. A major difficulty and key step in applying Theorem 2 is to find a bound for

$$I_n = \sum_{i=1}^n \frac{\|D_i\|_{2p}^{2p}}{n^p} + \Big\|\frac{V_n}{n} - \sigma^2\Big\|_p^p = n^{1-p}\|D_0\|_{2p}^{2p} + \Big\|\frac{V_n}{n} - \sigma^2\Big\|_p^p, \tag{31}$$

where V_n is the sum of conditional variances or quadratic characteristic

$$V_n = \sum_{i=1}^n \mathrm{E}\,(D_i^2|\mathcal{F}_{i-1}). \tag{32}$$

Interestingly, with our physical dependence measure (3), a bound with simple and explicit form can be found. To this end, by Proposition 3 in Wu (2007), there exists a constant $C_p > 0$ such that

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

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

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

So $\|\mathrm{E}\,(D_m^2|\mathcal{F}_0) - \sigma^2\|_p = O(m^{-\eta})$ and, by the triangle inequality, $\|\mathrm{E}\,(V_m|\mathcal{F}_0) - m\sigma^2\|_p = \sum_{i=1}^m O(i^{-\eta})$. Applying Theorem 3 with $\xi_i = \mathcal{F}_{i-1}$ and $Z_i = V_i - i\sigma^2$, elementary calculations show that $\|V_n - n\sigma^2\|_p = O(n^{1/p})$, $O(n^{1/p}\log n)$, or $O(n^{1-\eta})$ if $\eta > 1-1/p$, $\eta = 1-1/p$, or $\eta < 1-1/p$, respectively. Combining these three cases, we have $I_n = O(\tau_n^{-1})$.

By Theorem 2, since $x_{r_x-\epsilon_x}/x = 1 + O((1+r_x)\epsilon_x)$, there exists a constant C independent of x and n such that

$$\Big|\frac{\mathbb{P}(M_n \geq \sqrt{n}\sigma(r_x-\epsilon_x))}{1-\Phi(r_x-\epsilon_x)} - 1\Big| \leq C(xI_n)^{\frac{1}{(1+2p)}} \tag{33}$$

holds uniformly in $x \in [1, a\tau_n]$. Clearly the above relation also holds with $r_x - \epsilon_x$ replaced by $r_x + \epsilon_x$. By (29), (30) and (33),

$$\frac{\mathbb{P}(S_n \geq \sqrt{n}\sigma r_x)}{1-\Phi(r_x)} - 1 \leq \frac{\mathbb{P}(M_n \geq \sqrt{n}\sigma(r_x-\epsilon_x))}{1-\Phi(r_x)} - 1 + \frac{\Lambda(n)}{(1-\Phi(r_x))(\sigma\epsilon_x)^{2p}}$$

$$= O\Big\{(xI_n)^{\frac{1}{(1+2p)}}\Big\} + O(\epsilon_x(1+r_x)) + \frac{O(\Lambda(n))(1+r_x)}{\phi(r_x)\epsilon_x^{2p}}.$$

A lower bound for $\mathbb{P}(S_n \geq \sqrt{n}\sigma r_x)/[1-\Phi(r_x)]$ can be similarly obtained. So Theorem 1 follows in view of the choice of ϵ_x in (27).

Acknowledgements

The authors are grateful to the referees for their many helpful suggestions. This work is supported in part by NSF grant DMS-0478704.

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.* **20**, 2123-2130.

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

Bose, A. (1986). Certain nonuniform rates of convergence to normality for martingale differences. *J. Statist. Plann. Inference* **14**, 155-167.

Chow, Y. S. and Teicher, H. (2003). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

Chen, X. (2001). Moderate deviations for Markovian occupation times. *Stochastic Process. Appl.* **94**, 51-70.

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

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.

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

Gao, F. Q. (1996). Moderate deviations for martingales and mixing random processes. *Stochastic Process. 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. Theoret. Probab.* **19**, 1-44.

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25**, 1636-1669.

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

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

Peligrad, M., Utev, S. and Wu, W. B. (2007). A maximal L_p-inequality for stationary sequences and application. *Proc. Amer. Math. Soc.* **135**, 541-550.

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

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

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

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

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35**, 2294-2320.

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. (2007). Inference of trends in time series. *J. Roy. Statist. Soc. Ser. B* **69**, 391-410.

Department of Statistics, University of Chicago, 5734 S. University Ave., Chicago, IL 60637, U.S.A.

E-mail: wbwu@galton.uchicago.edu

Department of Statistics, Pennsylvania State University, 326 Thomas Building, University Park, PA 16802, U.S.A.

E-mail: zuz13@stat.psu.edu

(Received October 2006; accepted January 2007)