Jrl Syst Sci & Complexity (2007) 20: 243–250

AN EXPONENTIAL INEQUALITY FOR AUTOREGRESSIVE PROCESSES IN ADAPTIVE TRACKING*

Bernard BERCU

Received: 10 January 2007

Abstract A wide range of literature concerning classical asymptotic properties for linear models with adaptive control is available, such as strong laws of large numbers or central limit theorems. Unfortunately, in contrast with the situation without control, it appears to be impossible to find sharp asymptotic or nonasymptotic properties such as large deviation principles or exponential inequalities. Our purpose is to provide a first step towards that direction by proving a very simple exponential inequality for the standard least squares estimator of the unknown parameter of Gaussian autoregressive process in adaptive tracking.

Key words Adaptive tracking, autoregressive process, exponential inequalities, least squares, martingales.

1 Introduction

First of all, we recall the celebrated Hoeffding, Bennett and Bernstein exponential inequalities for sums of bounded independent random variables. We refer the reader to [1], [2] and the excellent survey of McDiarmid[3].

Theorem 1 (Hoeffding's inequality) *Let (X_n) be a sequence of independent random variables such that, for each $1 \leq k \leq n$, $a_k \leq X_k \leq b_k$ a.s. for some constants $a_k < b_k$. If $S_n = X_1 + X_2 + \cdots + X_n$, then for all $x \geq 0$,*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) \leq 2\exp\left(-\frac{2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right). \tag{1}$$

Theorem 2 (Bennett's inequality) *Let (X_n) be a sequence of independent square integrable random variables such that, for each $1 \leq k \leq n$, $X_k \leq c$ a.s. for some constant $c > 0$. If $S_n = X_1 + X_2 + \cdots + X_n$ and V_n is the variance of S_n, then for all $x \geq 0$,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{V_n}{c^2} h\left(\frac{xc}{V_n}\right)\right), \tag{2}$$

where $h(x) = (1+x)\log(1+x) - x$.

Bernard BERCU
Institut de Mathematiques de Bordeaux, Universite Bordeaux 1, *UMR*5251, 351, *cours de la liberation,* 33405 *Talence cedex, France.* Email: Bernard.Bercu@math.u-bordeaux1.fr.
*This paper is dedicated to Professor Han-Fu Chen for his 70th birthday.

Theorem 3 (Bernstein's inequality) *Under the assumptions of Theorem* 2, *we have for all* $x \geq 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{x^2}{2(V_n + \frac{xc}{3})}\right). \tag{3}$$

Remark 1 One can observe that Bernstein's inequality immediately follows from (2) as for all $x \geq 0$,

$$h(x) \geq \frac{3x^2}{2(3+x)}.$$

Moreover, if $V_n = \sigma^2 n$ with $\sigma^2 > 0$ and $x = o(n)$, then the upper bound in (3) behaves like $\exp(-\frac{x^2}{2\sigma^2 n})$ which coincides with the Gaussian upper bound since by the central limit theorem,

$$\frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

2 Exponential Inequalities for Martingales

We shall now focus our attention on exponential inequalities for martingales. Let (M_n) be a locally square integrable real martingale adapted to a filtration $\mathbb{F} = (\mathcal{F}_n)$ with initial value $M_0 = 0$. The predictable quadratic variation and the total quadratic variation of (M_n) are respectively given by

$$\langle M\rangle_n = \sum_{k=1}^{n} \mathbb{E}[\Delta M_k^2 | \mathcal{F}_{k-1}] \quad \text{and} \quad [M]_n = \sum_{k=1}^{n} \Delta M_k^2,$$

where $\Delta M_n = M_n - M_{n-1}$. Hoeffding's inequality holds for bounded martingale difference sequences by the so-called Azuma-Hoeffding's inequality[4].

Theorem 4 (Azuma-Hoeffding's inequality) *Let* (M_n) *be a locally square integrable real martingale such that, for each* $1 \leq k \leq n$, $a_k \leq \Delta M_k \leq b_k$ *a.s. for some constants* $a_k < b_k$. *Then, for all* $x \geq 0$,

$$\mathbb{P}(|M_n| \geq x) \leq 2\exp\left(-\frac{2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right). \tag{4}$$

Another result which involves the predictable quadratic variation $(\langle M\rangle_n)$ is the famous Freedman's inequality[5].

Theorem 5 (Freedman's inequality) *Let* (M_n) *be a locally square integrable real martingale such that, for each* $1 \leq k \leq n$, $\Delta M_k \leq c$ *a.s. for some constant* $c > 0$. *Then, for all* $x, y > 0$,

$$\mathbb{P}(M_n \geq x, \langle M\rangle_n \leq y) \leq \exp\left(-\frac{x^2}{2(y + cx)}\right). \tag{5}$$

Over the last decade, extensive study has been made to establish exponential inequalities for self-normalized martingales relaxing the boundedness assumption on the increments. The most important advance is probably De la Peña's contribution[6]. We shall say that a real martingale (M_n) adapted to $\mathbb{F} = (\mathcal{F}_n)$ is conditionally symmetric if, for all $n \geq 1$, the conditional distribution of ΔM_n given $\mathcal{F}_{n-1}$ is symmetric.

Theorem 6 (De la Peña's inequality) *Let* (M_n) *be a locally square integrable and conditionally symmetric real martingale. Then, for all* $x, y > 0$,

$$\mathbb{P}(M_n \geq x, [M]_n \leq y) \leq \exp\left(-\frac{x^2}{2y}\right). \tag{6}$$

In addition, for all $a \geq 0$, $b > 0$,

$$\mathbb{P}\Big(\frac{M_n}{a+b[M]_n} \geq x\Big) \leq \sqrt{\mathbb{E}\Big[\exp\Big(-x^2\Big(ab+\frac{b^2}{2}[M]_n\Big)\Big)\Big]},$$

$$\mathbb{P}\Big(\frac{M_n}{a+b[M]_n} \geq x, [M]_n \geq \frac{1}{y}\Big) \leq \exp\Big(-x^2\Big(ab+\frac{b^2}{2y}\Big)\Big).$$

Our goal is to propose similar exponential inequalities for (M_n) self-normalized by its predictable quadratic variation $\langle M\rangle_n$. We shall say that a locally square integrable real martingale (M_n) is conditionally Gaussian if, for all $n \geq 1$, the conditional distribution of ΔM_n given $\mathcal{F}_{n-1}$ is $\mathcal{N}(0, \Delta\langle M\rangle_n)$ where $\Delta\langle M\rangle_n = \langle M\rangle_n - \langle M\rangle_{n-1}$.

Theorem 7 *Let (M_n) be a locally square integrable and conditionally Gaussian real martingale. Then, all the results of Theorem 6 are true replacing $[M]_n$ by $\langle M\rangle_n$ everywhere. In addition, for all $x > 0$, $a \geq 0$, and $b > 0$, we also have*

$$\mathbb{P}\Big(\frac{M_n}{a+b\langle M\rangle_n} \geq x\Big) \leq \inf_{p>1}\Big(\mathbb{E}\Big[\exp\Big(-(p-1)x^2\Big(ab+\frac{b^2}{2}\langle M\rangle_n\Big)\Big)\Big]\Big)^{\frac{1}{p}}. \tag{7}$$

The proof is given in Appendix A.

3 Application to Adaptive Tracking

Consider the autoregressive process given, for all $n \geq 0$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}, \tag{8}$$

where X_n, U_n and ε_n are the observation, control and driven noise, respectively. We assume that (ε_n) is a sequence of independent and identically distributed random variables with standard $\mathcal{N}(0,\sigma^2)$ distribution where $\sigma^2 > 0$. For the sake of simplicity, we also assume that the initial state X_0 is independent of (ε_n) with $\mathcal{N}(0,\tau^2)$ distribution, where $\tau^2 \geq \sigma^2$. A common way to estimate the unknown parameter θ is to make use of the standard least-squares estimator defined, for all $n \geq 1$, by

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}(X_k - U_{k-1})}{\sum_{k=1}^n X_{k-1}^2}. \tag{9}$$

One can also choose the Yule-Walker estimator

$$\widetilde{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}(X_k - U_{k-1})}{\sum_{k=0}^n X_k^2}. \tag{10}$$

The only difference between $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ is the additional term X_n^2 in the denominator of (10). In the situation without control, it is well-known that $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ both converge a.s. to θ and central limit theorems may be founded in [7]. It is only recently that sharp asymptotic properties such as large deviation principles were established for the Yule-Walker estimator $\widetilde{\theta}_n$ and much work remains to be done for the least-squares estimator $\widehat{\theta}_n$[8,9]. In the situation under control, $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ both converge a.s. to θ and central limit theorems or laws of iterated logarithm may be founded in [10], [11]. Moreover, it is extremely difficult to prove large deviation results as the adaptive control U_n completely erases the independence structure behind the autoregressive process (X_n). However, we shall establish a very simple exponential inequality for both $\widehat{\theta}_n$ and $\widetilde{\theta}_n$.

The role of the adaptive control U_n is to regulate the dynamic of the process (X_n) by forcing X_n to track, as closed as possible, a given predictable reference trajectory (x_n). For all $n \geq 0$, we have

$$\begin{cases} X_{n+1} = Y_n + \varepsilon_{n+1}, \\ \;\;Y_n \;\;= \theta X_n + U_n. \end{cases} \tag{11}$$

Since Y_n has to be as closed as possible to x_{n+1}, the natural choice for the adaptive control U_n is given by $U_0 = 0$ and, for all $n \geq 1$,

$$U_n = x_{n+1} - \widehat{\theta}_n X_n.$$

Our main result is as follows.

Theorem 8 *For all $n \geq 1$ and $x > 0$, we have*

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \exp\left(- \frac{nx^2}{2(1+y_x)} \right), \tag{12}$$

where y_x is the unique positive solution of the equation $h(y_x) = x^2$ and the function $h(x) = (1+x)\log(1+x) - x$. Moreover, for all $n \geq 1$ and $x > 0$, we also have

$$\mathbb{P}(|\widetilde{\theta}_n - \theta| \geq x + |\theta|) \leq 2 \exp\left(- \frac{nx^2}{2(1+y_x)} \right). \tag{13}$$

The proof is given in Appendix B.

Remark 2 First of all, it is important to see that inequality (12) is true for all values $n \geq 1$, it isn't an asymptotic result. Moreover, inequality (12) can be very simple if x is small enough. More precisely, one can easily see that for all $0 < x < 1$, $h(x) < \frac{x^2}{4}$. Consequently, it immediately follows from (12) that, for all $0 < x < \frac{1}{2}$,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \exp\left(- \frac{nx^2}{2(1+2x)} \right).$$

Finally, if $\theta > 0$, we can deduce from (10) that, for all $x > 0$,

$$\mathbb{P}(\widetilde{\theta}_n - \theta \geq x) \leq \exp\left(- \frac{nx^2}{2(1+y_x)} \right).$$

4 Conclusion

In this paper, we propose a simple exponential inequality for the least squares estimator as well as for the Yule-Walker estimator of the unknown parameter θ of a Gaussian autoregressive process in adaptive tracking. It would be a very attractive challenge for the control community to establish large or moderate deviation principles for these estimators in adaptive tracking. More precisely, let (a_n) be a sequence of positive real numbers increasing to infinity, such that $a_n = o(n)$, and denote

$$V_n = \sqrt{\frac{n}{a_n}}(\widehat{\theta}_n - \theta).$$

One can conjecture that the sequence (V_n) satisfies a large deviation principle with speed a_n and rate function $I(x) = \frac{x^2}{2}$.

Appendix A

This appendix is devoted to the proof of Theorem 7 inspired by the original work of De la Peña[6]. First of all, for all $n \geq 0$ and $t \in \mathbb{R}$, let

$$Z_n(t) = \exp\left(tM_n - \frac{t^2}{2}\langle M\rangle_n\right).$$

If (M_n) is conditionally Gaussian then it is easy to see that $(Z_n(t))$ is a martingale with expectation $\mathbb{E}[Z_n(t)] = 1$. For all $x, y > 0$, let

$$A_n = \left\{M_n \geq x, \langle M\rangle_n \leq y\right\}.$$

By Markov's inequality, we have for all $t > 0$,

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{E}\left[\exp\left(\frac{t}{2}M_n - \frac{tx}{2}\right)\mathbb{I}_{A_n}\right], \\
&\leq \mathbb{E}\left[\sqrt{Z_n(t)}\exp\left(\frac{t^2}{4}\langle M\rangle_n - \frac{tx}{2}\right)\mathbb{I}_{A_n}\right], \\
&\leq \exp\left(\frac{t^2 y}{4} - \frac{tx}{2}\right)\sqrt{\mathbb{P}(A_n)} \qquad \text{(A.1)}
\end{aligned}$$

because $\mathbb{E}[Z_n(t)] = 1$. Therefore, dividing both sides of (A.1) by $\sqrt{\mathbb{P}(A_n)}$ and taking the value $t = \frac{x}{y}$, we obtain that

$$\mathbb{P}(A_n) \leq \exp\left(-\frac{x^2}{2y}\right).$$

We next carry out the proof in the special case $a = 0$ and $b = 1$ because the proof for the general case follows exactly the same lines. For all $x > 0$, let $B_n = \{M_n \geq x\langle M\rangle_n\}$. By Markov's inequality together with Holder's inequality, we have for all $t > 0$ and $q > 1$,

$$\begin{aligned}
\mathbb{P}(B_n) &\leq \mathbb{E}\left[\exp\left(\frac{t}{q}M_n - \frac{tx}{q}\langle M\rangle_n\right)\mathbb{I}_{B_n}\right], \\
&\leq \mathbb{E}\left[(Z_n(t))^{\frac{1}{q}}\exp\left(\frac{t}{2q}(t-2x)\langle M\rangle_n\right)\mathbb{I}_{B_n}\right], \\
&\leq \left(\mathbb{E}\left[\exp\left(\frac{tp}{2q}(t-2x)\langle M\rangle_n\right)\right]\right)^{\frac{1}{p}} \qquad \text{(A.2)}
\end{aligned}$$

because once again $\mathbb{E}[Z_n(t)] = 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. As $\frac{p}{q} = p - 1$, we can deduce from (A.2) and the particular choice $t = x$ that

$$\mathbb{P}(B_n) \leq \inf_{p>1}\left(\mathbb{E}\left[\exp\left(-(p-1)\frac{x^2}{2}\langle M\rangle_n\right)\right]\right)^{\frac{1}{p}},$$

which immediately implies (7). In the case $p=2$, we find that

$$\mathbb{P}(B_n) \leq \sqrt{\mathbb{E}\Big[\exp\Big(-\frac{x^2}{2}\langle M\rangle_n\Big)\Big]}.$$

Furthermore, for all $x, y > 0$, let

$$C_n = \Big\{M_n \geq x\langle M\rangle_n, \langle M\rangle_n \geq \frac{1}{y}\Big\}.$$

As in the proof of (A.1), we obtain that for all $0 < t < 2x$,

$$\begin{aligned}\mathbb{P}(C_n) &\leq \mathbb{E}\Big[\sqrt{Z_n(t)}\exp\Big(\frac{t}{4}(t-2x)\langle M\rangle_n\Big)\mathbb{1}_{C_n}\Big],\\ &\leq \exp\Big(\frac{t}{4y}(t-2x)\Big)\mathbb{E}\Big[\sqrt{Z_n(t)}\mathbb{1}_{C_n}\Big],\\ &\leq \exp\Big(\frac{t}{4y}(t-2x)\Big)\sqrt{\mathbb{P}(C_n)}.\end{aligned} \tag{A.3}$$

Therefore, dividing both sides of (A.3) by $\sqrt{\mathbb{P}(C_n)}$ and taking the value $t = x$, we obtain that

$$\mathbb{P}(C_n) \leq \exp\Big(-\frac{x^2}{2y}\Big),$$

which completes the proof of Theorem 7.

Appendix B

We shall now focus our attention on the proof of Theorem 8. It immediately follows from (8) together with (9) that, for all $n \geq 1$,

$$\widehat{\theta}_n - \theta = \sigma^2 \frac{M_n}{\langle M\rangle_n}, \tag{B.1}$$

where

$$M_n = \sum_{k=1}^{n} X_{k-1}\varepsilon_k \qquad \text{and} \qquad \langle M\rangle_n = \sigma^2 \sum_{k=1}^{n} X_{k-1}^2.$$

The driven noise (ε_n) is a sequence of independent and identically distributed random variables with $\mathcal{N}(0,\sigma^2)$ distribution. Consequently, for all $n \geq 1$, the conditional distribution of ΔM_n given $\mathcal{F}_{n-1}$ is $\mathcal{N}(0,\sigma^2 X_{n-1}^2)$ which implies that (M_n) is a locally square integrable and conditionally Gaussian real martingale. Therefore, we infer from inequality (7) of Theorem 7 that, for all $n \geq 1$ and $x > 0$,

$$\begin{aligned}\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) = \mathbb{P}\Big(|M_n| \geq \frac{x}{\sigma^2}\langle M\rangle_n\Big) &= 2\mathbb{P}\Big(M_n \geq \frac{x}{\sigma^2}\langle M\rangle_n\Big),\\ &\leq 2\inf_{p>1}\Big(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2\sigma^4}\langle M\rangle_n\Big)\Big]\Big)^{\frac{1}{p}}.\end{aligned} \tag{B.2}$$

Similar result may be found in [12], [13]. We are now halfway to our goal and it remains to find an exponential inequality for $\langle M\rangle_n$. For all $t \in \mathbb{R}$ such that $1-2\sigma^2 t > 0$, if $\alpha = \frac{1}{\sqrt{1-2\sigma^2 t}}$, we deduce from (11) that, for all $n \geq 1$,

$$\begin{aligned}\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] &= \exp(tY_{n-1}^2)\mathbb{E}[\exp(2tY_{n-1}\varepsilon_n + t\varepsilon_n^2)|\mathcal{F}_{n-1}],\\ &= \frac{\exp(tY_{n-1}^2)}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left(-\frac{x^2}{2\alpha^2\sigma^2}\right)\exp(2tY_{n-1}x)dx.\end{aligned}$$

Hence, if $\beta = 2t\alpha\sigma Y_{n-1}$, we find via the change of variables $y = \frac{x}{\alpha\sigma}$ that

$$\begin{aligned}\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] &= \frac{\alpha\exp(tY_{n-1}^2)}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-\frac{y^2}{2}+\beta y)\,dy,\\ &= \alpha\exp\left(tY_{n-1}^2+\frac{\beta^2}{2}\right) = \alpha\exp(t\alpha^2Y_{n-1}^2),\end{aligned}$$

which implies that, for all $t<0$ and $n \geq 1$,

$$\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] \leq \alpha. \tag{B.3}$$

Furthermore, as X_0 is $\mathcal{N}(0,\tau^2)$ distributed with $\tau^2 \geq \sigma^2$, we also have, for all $t<0$,

$$\mathbb{E}[\exp(tX_0^2)] \leq \alpha.$$

It immediately follows from (B.3) together with the tower property of the conditional expectation that, for all $t<0$, $n \geq 0$,

$$\mathbb{E}\left[\exp\left(t\sum_{k=0}^{n} X_k^2\right)\right] \leq \alpha^{n+1},$$

which ensures that

$$\mathbb{E}[\exp(t\langle M\rangle_n)] \leq (1-2\sigma^4 t)^{-\frac{n}{2}}. \tag{B.4}$$

Consequently, we deduce from the conjunction of (B.2) and (B.4) with the value

$$t = -\frac{(p-1)x^2}{2\sigma^4}$$

and the change of variables $y=(p-1)x^2$ that, for all $x>0$ and $n \geq 1$,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2\inf_{y>0}\exp\left(-\frac{nx^2}{2}\ell(y)\right),$$

where the function ℓ is given by

$$\ell(y) = \frac{\log(1+y)}{x^2+y}.$$

We clearly have

$$\ell'(y) = \frac{x^2-h(y)}{(1+y)(x^2+y)^2},$$

where $h(y)=(1+y)\log(1+y)-y$. One can observe that h is the strictly convex function given in Bennett's inequality. Let y_x be the unique positive solution of the equation $h(y_x)=x^2$. The value y_x maximizes the function ℓ and this natural choice immediately leads to (12). Finally, it follows from (10) and (12) that, for all $x > 0$ and $n \geq 1$,

$$\mathbb{P}(|\widetilde{\theta}_n - \theta + \theta f_n| \geq x) \leq 2\exp\left(-\frac{nx^2}{2(1+y_x)}\right), \tag{B.5}$$

where

$$f_n = \frac{X_n^2}{s_n} \qquad \text{and} \qquad s_n = \sum_{k=0}^{n} X_k^2.$$

As $0 \leq f_n \leq 1$, (B.5) implies (13) achieving the proof of Theorem 8.

References

[1] G. Bennett, Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.*, 1962, **57**: 33–45.
[2] W. J. Hoeffding, Probability inequalities sums of bounded random variables, *J. Amer. Statist. Assoc.*, 1963, **58**: 713–721.
[3] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics, Mathematics Series: Algorithms and Combinatorics* (Ed. by M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed), Vol. 16, Springer-Verlag, Berlin, 1998, 195–248.
[4] K. Azuma, Weighted sums of certain dependent random variables, *Tôkuku Math. J.*, 1967, **19**: 357–367.
[5] D. A. Freedman, On tail probabilities for martingales, *Ann. of Probab.*, 1975, **3**: 100–118.
[6] V. H. De la Peña, A general class of exponential inequalities for martingales and ratios, *Ann. of Probab.*, 1999, **27**: 537–564.
[7] J. S. White, The limit distribution of the serial correlation in the explosive case, *Ann. Math. Statist.*, 1958, **29**: 1188–1197.
[8] B. Bercu, F. Gamboa, and A. Rouault, Large deviations for quadratic forms of stationary Gaussian processes, *Stoch. Proc. and Their Appl.*, 1997, **71**: 75–90.
[9] B. Bercu, On large deviations in the Gaussian autoregressive process: stable, unstable and explosive cases, *Bernoulli*, 2001, **7**: 299–316.
[10] B. Bercu, Central limit theorem and law of iterated logarithm for the least squares algorithms in adaptive tracking, *SIAM J. Control and Optimization*, 1998, **36**: 910–928.
[11] B. Bercu and B. Portier, Adaptive control of parametric nonlinear autoregressive models via a new martingale approach, *IEEE Trans. on Automatic Control*, 2002, **47**: 1524–1528.
[12] R. Liptser and V. Spokoiny, Deviation probability bound for martingales with applications to statistical estimation, *Statist. and Probab. Letters*, 2000, **46**: 347–357.
[13] J. Worms, Large and moderate deviations upper bounds for the Gaussian autoregressive process, *Statist. and Probab. Letters*, 2001, **51**: 235–243.
[14] V. H. De la Peña, M. J. Klass, and T. L. Lai, Self-normalized processes: exponential inequalities, moments bounds and iterated logarithm law, *Ann. of Probab.*, 2004, **32**: 1902–1933.