Available online at www.sciencedirect.com
ScienceDirect
Statistics and Probability Letters 78 (2008) 2339–2345
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Limit theorems for correlated Bernoulli random variables

Barry James, Kang James, Yongcheng Qi*

Department of Mathematics and Statistics, University of Minnesota Duluth, 1117 University Drive, Duluth, MN 55812, USA

Received 4 April 2006; received in revised form 8 January 2008; accepted 21 January 2008
Available online 26 February 2008

Abstract

In this paper we study limit theorems for a class of correlated Bernoulli processes. We obtain the strong law of large numbers, central limit theorem and the law of the iterated logarithm for the partial sums of the Bernoulli random variables.

1. Introduction

Consider a Bernoulli process $\{X_j, j \geq 1\}$ in which the random variables X_j are correlated in the sense that the success probability of a trial conditional on the previous trials depends on the total number of successes achieved to that point. More precisely, assume that for some $0 < p < 1$,

$$P(X_{j+1} = 1|\mathcal{F}_j) = (1 - \theta_j)p + \theta_j j^{-1} S_j,$$

where $0 \leq \theta_j \leq 1$ are dependence parameters, $S_n = \sum_{j=1}^n X_j$ for $n \geq 1$, and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, the σ-field generated by $X_1, \ldots, X_n$. If X_1 has a Bernoulli distribution with parameter p, it follows that $X_1, X_2, \ldots$ are identically distributed Bernoulli random variables. This process was introduced by Drezner and Farnum (1993), and the distribution of S_n is called a generalized binomial distribution. When all θ_j's are zero, this is the classic Bernoulli process. When $\theta_j > 0$, an intuitive interpretation for the model is that the $(j + 1)$th trial is expected to have a larger success probability than p if in the first j trials the average number of successes is larger than p. Many examples are given in Drezner and Farnum (1993).

For the special case when $\theta_n = \theta \in [0, 1)$ for all $n \geq 1$, Heyde (2004) proved that

$$n^{-1/2}(S_n - np) \xrightarrow{d} N\left(0, \frac{p(1-p)}{1-2\theta}\right) \quad \text{if } \theta < 1/2,$$

$$(n \log n)^{-1/2}(S_n - np) \xrightarrow{d} N(0, p(1-p)) \quad \text{if } \theta = 1/2,$$

$$n^{-\theta}(S_n - np) \xrightarrow{as} W \quad \text{if } \theta > 1/2,$$

where W is some non-degenerate mean-zero random variable with a non-normal distribution, and $\xrightarrow{d}$ and $\xrightarrow{as}$ denote distributional convergence and almost sure convergence, respectively.

* Corresponding author. Fax: +1 218 726 8399.
E-mail address: yqi@d.umn.edu (Y. Qi).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2008.01.104

The generalized Bernoulli process offers parameters for overdispersion and thus is more flexible than the classic Binomial model. If $\theta_n = \theta$ for all n, the parameter θ can be consistently estimated by maximum likelihood from the data. See the comments in Heyde (2004).

In this paper we will consider the general case. We will investigate the conditions for the strong law of large numbers, the central limit theorem and the law of the iterated logarithm for the partial sums of the dependent Bernoulli random variables. The main results of the paper will be given in Section 2, and all the proofs are given in Section 3.

2. Main results

Define

$$a_1 = 1, \quad a_n = \prod_{j=1}^{n-1}(1 + j^{-1}\theta_j) \quad \text{for } n \geq 2,$$

and

$$A_n^2 = \sum_{j=1}^{n} a_j^{-2} \quad \text{for } n \geq 1.$$

Theorem 2.1. *If* $\lim_{n\to\infty} a_n/n = 0$, *then*

$$\frac{S_n}{n} \xrightarrow{as} p. \tag{2.1}$$

Conversely, if (2.1) *holds, then* $\lim_{n\to\infty} a_n/n = 0$.

Theorem 2.2. *If* $\lim_{n\to\infty} A_n = \infty$, *then*

$$\frac{S_n - np}{a_n A_n} \xrightarrow{d} N(0, p(1-p)); \tag{2.2}$$

and if $\lim_{n\to\infty} A_n < \infty$, *then*

$$\frac{S_n - np}{a_n} \xrightarrow{as} V, \tag{2.3}$$

where V *is some non-degenerate mean-zero random variable.*

Theorem 2.3. *If* $\lim_{n\to\infty} A_n = \infty$, *then*

$$\limsup_{n\to\infty} \pm \frac{S_n - np}{a_n A_n \sqrt{\log\log A_n}} \stackrel{as}{=} \sqrt{2p(1-p)}.$$

A function $\phi(x)$ defined on $(0, \infty)$ is said to be regularly varying at infinity with index α if

$$\lim_{x\to\infty} \frac{\phi(tx)}{\phi(x)} = t^\alpha \quad \text{for all } t > 0.$$

(Notation $\phi \in RV(\alpha)$.)

For convenience set $a_0 = 1$. Define the function $a(x) = a_{\lfloor x \rfloor}$ for $x > 0$, where $\lfloor x \rfloor$ denotes the integer part of x.

Corollary 2.1. *Assume that* $a(x) \in RV(\theta)$, *where* $\theta \in [0, 1/2)$. *Then*

$$\frac{S_n - np}{\sqrt{n}} \xrightarrow{d} N\left(0, \frac{p(1-p)}{1-2\theta}\right)$$

and

$$\limsup_{n\to\infty} \pm \frac{S_n - np}{\sqrt{n \log\log n}} \stackrel{as}{=} \sqrt{\frac{2p(1-p)}{1-2\theta}}.$$

Corollary 2.2. *If* $\lim_{n\to\infty}\theta_n = \theta \in [0, 1/2)$, *then*

$$\limsup_{n\to\infty} \pm \frac{S_n - np}{\sqrt{n \log\log n}} \overset{as}{=} \sqrt{\frac{2p(1-p)}{1-2\theta}}.$$

Corollary 2.3. *Assume that* $\theta_n = \theta \in [0, 1/2]$ *for all* $n \geq 1$.
(i) *If* $\theta \in [0, 1/2)$, *then*

$$\limsup_{n\to\infty} \pm \frac{S_n - np}{\sqrt{n \log\log n}} \overset{as}{=} \sqrt{\frac{2p(1-p)}{1-2\theta}}.$$

(ii) *If* $\theta = 1/2$, *then*

$$\limsup_{n\to\infty} \pm \frac{S_n - np}{\sqrt{n \log n \log\log\log n}} \overset{as}{=} \sqrt{2p(1-p)}.$$

We end this section with some remarks and an example. In Corollary 2.1, if we assume that $a(x) \in RV(\theta)$ for some $\theta \in (1/2, 1]$, then we have $\lim_{n\to\infty} A_n < \infty$ from properties of regular variation. Hence we can apply Theorem 2.2 to get the limit from the almost sure convergence. When $\theta = 1/2$, it is rather complicated. The normalizing constants are not as simple as in Corollary 2.1. They are dependent on the asymptotic behavior of $a(x)$. The following example shows that the normalizing constants in this case may be quite different.

Example. Let $\delta \in (0, 1)$ be fixed. Define $b(x) = x^{1/2}\exp\{-(\log x)^\delta\}$ for $x > 1$. It is easily seen that $\lim_{x\to\infty} x(\frac{b(x+1)}{b(x)} - 1) = \frac{1}{2}$. Let n_0 be an integer such that $x(\frac{b(x+1)}{b(x)} - 1) < 1$ for all $x \geq n_0$, and define $\theta_j = 0$ for $1 \leq j < n_0$, and $\theta_j = j(\frac{b(j+1)}{b(j)} - 1)$ for all $j \geq n_0$. Then we have

$$a_n = \prod_{j=1}^{n-1}(1 + j^{-1}\theta_j) = \begin{cases} 1 & \text{if } n < n_0 \\ \dfrac{b(n)}{b(n_0)} & \text{if } n \geq n_0. \end{cases}$$

Calculations show that

$$A_n^2 = \sum_{j=1}^{n} a_j^{-2} \sim \frac{(b(n_0))^2}{2\delta}(\log n)^{1-\delta}\exp\{2(\log n)^\delta\} \quad \text{as } n \to \infty.$$

Therefore,

$$a_n A_n \sim \frac{n^{1/2}(\log n)^{(1-\delta)/2}}{(2\delta)^{1/2}}$$

and

$$a_n A_n \sqrt{\log\log A_n} \sim \frac{1}{\sqrt{2}} n^{1/2}(\log n)^{(1-\delta)/2}(\log\log n)^{1/2}.$$

It follows from Theorem 2.3 that

$$\limsup_{n\to\infty} \pm \frac{S_n - np}{\sqrt{n(\log n)^{1-\delta}\log\log n}} \overset{as}{=} \sqrt{p(1-p)}.$$

3. Proofs

For $\theta > 0$ let

$$a_1(\theta) = 1, \qquad a_n(\theta) = \prod_{j=1}^{n-1}(1 + j^{-1}\theta) \quad \text{for } n \geq 2.$$

Lemma 3.1. *For any* $\theta > 0$, $a_n(\theta) \sim \frac{n^\theta}{\theta \Gamma(\theta)}$ *as* $n \to \infty$.

The conclusion in the lemma can be found in Heyde (2004).

Lemma 3.2. (i) n/a_n *is non-decreasing in* n.
(ii) *If* $\lim_{n\to\infty} A_n = \infty$, *then* $\lim_{n\to\infty} a_n/n = 0$.

Proof. (i) Note that $a_n(1) = n$ for all $n \geq 1$. We have that

$$\frac{n}{a_n} = \frac{a_n(1)}{a_n} = \prod_{j=1}^{n-1} \frac{1 + j^{-1}}{1 + j^{-1}\theta_j}$$

is non-decreasing since all terms in the above product are at least 1.
(ii) Since a_n/n is monotone, its limit exists; denote this limit by ν. The limit ν must be zero. Otherwise we have $\frac{1}{a_n^2} \leq \frac{1}{\nu^2 n^2}$ for all $n \geq 1$, which implies $\lim_{n\to\infty} A_n < \infty$. □

Lemma 3.3. *Let* $\{Z_n, \mathcal{G}_n, n \geq 1\}$ *be a sequence of martingale differences. If* $\sum_{n=1}^\infty E(Z_n^2|\mathcal{G}_{n-1}) < \infty$ *a.s., then* $\sum_{j=1}^n Z_j$ *converges almost surely.*

This lemma is a special case of Theorem 2.17 in Hall and Heyde (1980).

Lemma 3.4. *Let* $\{Z_n, \mathcal{G}_n, n \geq 1\}$ *be a sequence of bounded martingale differences. Assume that there exists a sequence of positive constants* $\{W_n\}$ *such that* $W_n \to \infty$ *as* $n \to \infty$ *and* $\frac{1}{W_n^2}\sum_{j=1}^n E(Z_j^2|\mathcal{G}_{j-1}) \xrightarrow{p} \sigma^2$. *Then* $\frac{\sum_{j=1}^n Z_j}{W_n} \xrightarrow{d} N(0, \sigma^2)$.

Proof. Let $Z_{ni} = Z_i/W_n$ for $1 \leq i \leq n$. By Corollary 3.1 in Hall and Heyde (1980), it suffices to prove the conditional Lindeberg condition

$$\sum_{j=1}^n E(Z_{nj}^2 I(|Z_{nj}| > \varepsilon)|\mathcal{G}_{j-1}) \xrightarrow{p} 0 \quad \text{for all } \varepsilon > 0,$$

which is trivial since, for any given $\varepsilon > 0$, all sets $\{|Z_{nj} > \varepsilon|\}$ are empty for large n. □

Lemma 3.5. *Let* $\{Z_n, \mathcal{G}_n, n \geq 1\}$ *be a sequence of bounded martingale differences. Assume that there exists a sequence of positive constants* $\{W_n\}$ *such that*

$$W_{n+1}^{-1} W_n \to 1, \quad W_n \to \infty, \tag{3.1}$$

and

$$\frac{1}{W_n^2}\sum_{j=1}^n E(Z_j^2|\mathcal{G}_{j-1}) \xrightarrow{as} 1. \tag{3.2}$$

Then

$$\limsup_{n\to\infty} \pm \frac{\sum_{j=1}^n Z_j}{\sqrt{2W_n^2 \log\log W_n^2}} \stackrel{as}{=} 1.$$

Proof. In view of Theorems 4.7 and 4.8 in Hall and Heyde (1980), it suffices to prove

$$\sum_{j=1}^\infty W_j^{-4} E(Z_j^4|\mathcal{G}_{j-1}) < \infty \quad \text{a.s.},$$

which, by (3.2), is equivalent to

$$\sum_{j=1}^\infty V_j^{-4} E(Z_j^4|\mathcal{G}_{j-1}) < \infty \quad \text{a.s.},$$

where $V_n^2 = \sum_{j=1}^n E(Z_j^2|\mathcal{G}_{j-1})$. For simplicity, assume that $V_1^2 = E(Z_1^2|\mathcal{G}_0) = EZ_1^2 > 0$. Since the Z_n's are bounded by some constant c, we have

$$\begin{aligned}\sum_{j=2}^{\infty} V_j^{-4} E(Z_j^4|\mathcal{G}_{j-1}) &\le c^2 \sum_{j=2}^{\infty} V_j^{-4} E(Z_j^2|\mathcal{G}_{j-1}) \\ &\le c^2 \sum_{j=2}^{\infty} V_j^{-4}(V_j^2 - V_{j-1}^2) \le c^2 \sum_{j=2}^{\infty} \frac{V_j^2 - V_{j-1}^2}{V_j^2 V_{j-1}^2} \\ &= c^2 \sum_{j=2}^{\infty} (V_{j-1}^{-2} - V_j^{-2}) = c^2 V_1^{-2}.\end{aligned}$$

This proves the lemma. □

Proof of Theorem 2.1. Define

$$T_n = \frac{S_n - np}{a_n} \quad \text{for } n \ge 1, \quad \text{and} \quad Y_1 = T_1, \quad Y_n = T_n - T_{n-1} \quad \text{for } n \ge 2.$$

Then it is easily checked that $\{T_n, \mathcal{F}_n, n \ge 1\}$ is a martingale, and hence the $Y_n, n \ge 1$ are the martingale differences from $\{T_n\}$. Some simple calculation shows that for $j > 1$

$$Y_j = \frac{1}{a_j}\left(X_j - p + \theta_{j-1}\left(p - \frac{S_{j-1}}{j-1}\right)\right)$$

and

$$E(Y_j^2|\mathcal{F}_{j-1}) = \frac{1}{a_j^2}\left(p(1-p) + \theta_{j-1}(1-2p)\left(\frac{S_{j-1}}{j-1} - p\right) - \theta_{j-1}^2\left(\frac{S_{j-1}}{j-1} - p\right)^2\right). \tag{3.3}$$

Since $0 \le S_{j-1} \le j - 1$, we have for some constant C

$$E(Y_j^2|\mathcal{F}_{j-1}) \le \frac{C}{a_j^2} \quad \text{for } j > 1. \tag{3.4}$$

Let $Z_n = a_n Y_n / n$ for $n \ge 1$. Then $\{Z_n, \mathcal{F}_n, n \ge 1\}$ is a sequence of bounded martingale differences. From (3.4) we get $\sum_{j=2}^{\infty} E(Z_j^2|\mathcal{F}_{j-1}) \le C \sum_{j=2}^{\infty} \frac{1}{j^2} < \infty$, which together with Lemma 3.3 yields

$$\sum_{j=1}^{n} \frac{a_j Y_j}{j} = \sum_{j=1}^{n} Z_j \quad \text{converges almost surely.}$$

Since n/a_n is non-decreasing and goes to infinity from Lemma 3.2, we have in view of Kronecker's lemma that

$$\lim_{n\to\infty}\left|\frac{S_n}{n} - p\right| = \lim_{n\to\infty}\left|\frac{S_n - np}{n}\right| = \lim_{n\to\infty}\left|\frac{a_n T_n}{n}\right| = \lim_{n\to\infty}\left|\frac{\sum_{j=1}^{n} Y_j}{n/a_n}\right| \overset{as}{=} 0,$$

proving (2.1).

Now assume that (2.1) holds. If $\lim_{n\to\infty} a_n/n = 0$ does not hold, then we have $\lim_{n\to\infty} A_n < \infty$ from Lemma 3.2, and $a_n/n \to \nu$ for some $\nu > 0$ from the monotonicity. Then by Theorem 2.2 we have

$$\frac{S_n}{n} - p \quad \text{converges almost surely to some non-degenerate random variable,}$$

which yields a contradiction. It is worth mentioning that in the proof of Theorem 2.2 below, we have to use Theorem 2.1, but we will only use the sufficiency of the theorem.

This finishes the proof of Theorem 2.1. □

Proof of Theorem 2.2. Under the conditions $\lim_{n\to\infty} a_n/n = 0$ and $\lim_{n\to\infty} A_n = \infty$ we have

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} E(Y_j^2|\mathcal{F}_{j-1})}{A_n^2} \stackrel{as}{=} p(1-p). \tag{3.5}$$

In fact, we have from (3.3) and Theorem 2.1 (sufficiency part) that

$$E(Y_j^2|\mathcal{F}_{j-1}) = \frac{p(1-p)}{a_j^2}(1+o(1)) \quad \text{as } j \to \infty$$

with probability one. This together with the assumption $\lim_{n\to\infty} A_n = \infty$ yields (3.5). Hence, by Lemma 3.4 we have

$$\frac{S_n - np}{a_n A_n} = \frac{\sum_{j=1}^{n} Y_j}{A_n} \xrightarrow{d} N(0, p(1-p)),$$

proving (2.2).

If $\lim_{n\to\infty} A_n < \infty$, then from (3.4) we have $\sum_{j=1}^{\infty} E(Y_j^2|\mathcal{F}_{j-1}) < \infty$ a.s. From Lemma 3.3 we obtain

$$\frac{S_n - np}{a_n} = \sum_{j=1}^{n} Y_j \quad \text{converges almost surely to some random variable } V,$$

where V is a well-defined random variable with zero mean and positive variance $\sum_{j=1}^{\infty} E(Y_j^2)$. This proves (2.3), and hence the theorem. □

Proof of Theorem 2.3. Let $Z_n = Y_n$ and $W_n^2 = A_n^2 p(1-p)$. Under the given conditions, (3.2) follows from (3.5). (3.1) holds as well since the Z_n's are bounded and $W_n \to \infty$. Since $\log\log(W_n^2) \sim \log\log A_n$ and

$$\frac{S_n - np}{a_n A_n \sqrt{\log\log A_n}} = \frac{\sum_{j=1}^{n} Z_j}{\sqrt{2W_n^2 \log\log W_n^2}} \sqrt{\frac{\log\log W_n^2}{\log\log A_n}} \sqrt{2p(1-p)},$$

the theorem follows from Lemma 3.5. □

Proof of Corollary 2.1. It is easily seen that $a(x)^{-2} \in RV(-2\theta)$. Since $-2\theta > -1$, from Karamata's Theorem (see, e.g., Theorem 1.5.11 in Bingham et al., 1987) we have

$$A_n^2 = \int_0^n a(x)^{-2} \mathrm{d}x \sim \frac{na(n)^{-2}}{1-2\theta} = \frac{na_n^{-2}}{1-2\theta},$$

which yields $a_n A_n \sim \sqrt{\frac{n}{1-2\theta}}$. Since $xa(x)^{-2} \in RV(1-2\theta)$ and $a(x)/x \in RV(\theta - 1)$ we have from the properties of regular variation that $\log\log A_n \sim \log\log(na_n^{-2}) \sim \log\log n$. Therefore, Corollary 2.1 follows from Theorems 2.2 and 2.3. □

Proof of Corollary 2.2. By Corollary 2.1, it suffices to show that $a(x) \in RV(\theta)$; that is, for every $t \in (0, 1)$, $\lim_{n\to\infty} \frac{a(nt)}{a(n)} = t^{\theta}$. The calculation is straightforward and the details are omitted here. □

Proof of Corollary 2.3. When $\theta < 1/2$, the corollary follows from Corollary 2.2. When $\theta = 1/2$, it follows from Theorem 2.3 and the following estimates that

$$A_n = \sqrt{\sum_{j=1}^{n} a_j^{-2}} \sim \frac{1}{2}\Gamma\left(\frac{1}{2}\right)\sqrt{\sum_{j=1}^{n} j^{-1}} \sim \frac{1}{2}\Gamma\left(\frac{1}{2}\right)\sqrt{\log n}$$

and $a_n A_n \sim \sqrt{n \log n}$ from Lemma 3.1. □

Acknowledgment

Qi's research was supported by NSF grant DMS 0604176.

References

Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. Cambridge University, New York.
Drezner, Z., Farnum, N., 1993. A generalized binomial distribution. Comm. Statist. Theory Methods 22, 3051–3063.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.
Heyde, C.C., 2004. Asymptotics and criticality for a correlated Bernoulli process. Aust. N. Z. J. Statist. 46, 53–57.