Strong Approximations under Dependence

by
Wei Biao Wu
The University of Chicago
5734 S. University Avenue
Chicago, IL 60637

Based on the joint work Berkes, Liu and Wu (2010) and Wu (2007).

Gaussian Approximations

- $(\mathbf{X}_i)_{i \in Z}$: a centered d-dimensional random processes
- The partial sum $\mathbf{S}_i = \sum_{j=1}^{i} \mathbf{X}_i$
- Approximate $\mathbf{S}_i$ by Gaussian processes
- On a richer probability space, there exists a Gaussian process $\mathbf{G}_i$ and a process $\check{\mathbf{S}}_i$ such that
  - $(\check{\mathbf{S}}_i)_{i \in N}$ and $(\mathbf{S}_i)_{i \in N}$ are identically distributed;
  - The super distance between $\check{\mathbf{S}}_i$ and $\mathbf{G}_i$:

$$\max_{1 \le i \le n} |\check{\mathbf{S}}_i - \mathbf{G}_i| = O(r_n). \tag{1}$$

Here r_n is the rate of approximation and $O(r_n)$ in (1) can be $O_P(r_n)$ or the almost sure rate $O_{\text{a.s.}}(r_n)$.

Gaussian Approximations

- $(\mathbf{X}_i)_{i \in Z}$: a centered d-dimensional random processes
- The partial sum $\mathbf{S}_i = \sum_{j=1}^{i} \mathbf{X}_i$
- Approximate $\mathbf{S}_i$ by Gaussian processes
- On a richer probability space, there exists a Gaussian process $\mathbf{G}_i$ and a process $\check{\mathbf{S}}_i$ such that
  - $(\check{\mathbf{S}}_i)_{i \in N}$ and $(\mathbf{S}_i)_{i \in N}$ are identically distributed;
  - The super distance between $\check{\mathbf{S}}_i$ and $\mathbf{G}_i$:

$$\max_{1 \le i \le n} |\check{\mathbf{S}}_i - \mathbf{G}_i| = O(r_n). \tag{1}$$

  Here r_n is the rate of approximation and $O(r_n)$ in (1) can be $O_P(r_n)$ or the almost sure rate $O_{\mathrm{a.s.}}(r_n)$.

Gaussian Approximations

- With the approximation (1), if r_n is sufficiently small, then statistics involve the partial sum process $(\mathbf{S}_i)_{i=1}^n$ can be approximated by functionals of the Gaussian process $(\check{\mathbf{G}}_i)_{i=1}^n$
- Gaussian processes have many nice and tractable properties
- A substantial generalization of the central limit theorem

Gaussian Approximations

A partial list of contributors:

- DOOB, J. L.
- STRASSEN, V.: i.i.d. random variables and martingale differences
- CSORGO, M. and REVESZ, P.
- KOMLÓS, J., MAJOR, P., AND TUSNÁDY, G: i.i.d. random variables, partial sum process and empirical processes
- BERKES, I., DEHLING, H., PHILIPP, W., STOUT, W.: strong mixing processes, vector-valued random variables
- BRADLEY, R. C.
- EBERLEIN, E.
- SHAO, Q., Einmahl, U.

Gaussian Approximations

A partial list of contributors:

- Sakhanenko, A.
- RIO, E.
- LIN and LU
- DEDECKER, J. and PRIEUR, C.
- VOLNÝ, D.: ergodic processes, martingales

Theorem (Komlós, Major and Tusnády, KMT, 1975, 1976). Assuming that

- e_i are iid with mean 0 and $\sigma^2 = E(e_i^2)$
- $E(|e_i|^p) < \infty$ for some $p > 2$
- $S_n = e_1 + \ldots + e_n$

Then on a richer probability space, there exists a Brownian motion B such that

$$\max_{i \le n} |S_i - \sigma B(i)| = o_{\text{a.s.}}(n^{1/p}). \tag{2}$$

The bound $o_{\text{a.s.}}(n^{1/p})$ is optimal.

Improvements and Generalizations: Shao, Sakhanenko, among others.

Gaussian Approximations under Independence

Theorem (Komlós, Major and Tusnády, KMT, 1975, 1976). Assuming that

- e_i are iid with mean 0 and $\sigma^2 = E(e_i^2)$
- $E(|e_i|^p) < \infty$ for some $p > 2$
- $S_n = e_1 + \ldots + e_n$

Then on a richer probability space, there exists a Brownian motion B such that

$$\max_{i \le n} |S_i - \sigma B(i)| = o_{\mathrm{a.s.}}(n^{1/p}). \tag{2}$$

The bound $o_{\mathrm{a.s.}}(n^{1/p})$ is optimal.
Improvements and Generalizations: Shao, Sakhanenko, among others.

Gaussian Approximations under Dependence

Many of previous results on strong approximations require strong mixing types of conditions, and establish results of the following type:

$$\max_{i \le n} |S_i - \sigma B(i)| = o_{\text{a.s.}}(n^{1/2-\epsilon}), \quad (3)$$

where ϵ is a small positive number.

- Not (nearly) optimal
- Of limited use in certain statistical inference.
- Wu and Zhao (2007) considered statistical inference of trends in time series and constructed simultaneous confidence bands for mean trends with asymptotically correct nominal coverage probabilities. It requires error bound of type $o_{\text{a.s.}}(n^{1/p})$.

Gaussian Approximations under Dependence

Many of previous results on strong approximations require strong mixing types of conditions, and establish results of the following type:

$$\max_{i \le n} |S_i - \sigma B(i)| = o_{\mathrm{a.s.}}(n^{1/2-\epsilon}), \qquad (3)$$

where ϵ is a small positive number.

- Not (nearly) optimal
- Of limited use in certain statistical inference.
- Wu and Zhao (2007) considered statistical inference of trends in time series and constructed simultaneous confidence bands for mean trends with asymptotically correct nominal coverage probabilities. It requires error bound of type $o_{\mathrm{a.s.}}(n^{1/p})$.

Gaussian Approximations under Dependence

Can we establish a strong invariance principle which can have optimal or nearly optimal bounds, as the one in Komlós, Major and Tusnády (1975,76)?

Yes, we can. However, ...

Can we establish a strong invariance principle which can have optimal or nearly optimal bounds, as the one in Komlós, Major and Tusnády (1975,76)?
Yes, we can. However, ...

Dependence Measures

In the recent ten plus years, there have been a substantial account on dependence measures:
Dedecker, Doukhan, Louhichi, Merlevéde, Prieur, etc. See

- Dedecker, Jérôme and Doukhan, Paul and Lang, Gabriel and León R., José Rafael and Louhichi, Sana and Prieur, Clémentine, *Weak dependence: with examples and applications*, Lecture Notes in Statistics, 190, Springer, New York,
- Patrice Bertail, Paul Doukhan and Philippe Soulier. (ed.) *Dependence in probability and statistics.* Lecture Notes in Statistics, 187. Springer, New York, 2006.
- Dehling, Herold and Mikosch, Thomas and Sørensen, Michael (2002) *Empirical process techniques for dependent data.* Birkhäuser Boston Inc., Boston, MA, 2002

Dependence Measures

Stationary and causal processes; let

$$X_n = g(\ldots, \varepsilon_{n-1}, \varepsilon_n), \text{ where} \tag{1}$$

- Input: ε_i are iid innovations or shocks that drive the system (Box and Jenkins, 1970). Let $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$
- g: a filter or a transform
- $X_n = g(\mathcal{F}_n)$: output or response

Interpret the dependence as *"How output depends on input"*

- Wiener, N. (1958). *Nonlinear Problems in Random Theory.* MIT Press, Cambridge, MA
- Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*
- Howell Tong (1990) *Nonlinear time series. A dynamical system approach.*
- Rosenblatt (2009, 1958)

Dependence Measures

Stationary and causal processes; let

$$X_n = g(\ldots, \varepsilon_{n-1}, \varepsilon_n), \text{ where} \tag{1}$$

- Input: ε_i are iid innovations or shocks that drive the system (Box and Jenkins, 1970). Let $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$
- g: a filter or a transform
- $X_n = g(\mathcal{F}_n)$: output or response

Interpret the dependence as *"How output depends on input"*

- Wiener, N. (1958). *Nonlinear Problems in Random Theory.* MIT Press, Cambridge, MA
- Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*
- Howell Tong (1990) *Nonlinear time series. A dynamical system approach.*
- Rosenblatt (2009, 1958)

Dependence Measures

- Two problems left:
  - How to define dependence measures based on the above interpretation?
  - How general is the representation (1)?
- Physical (or functional) dependence measure[1]
  - (ε_i'): iid copy of (ε_i)
  - $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$
  - Coupling: $\mathcal{F}_n^* = (\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n)$
  - $\|X\|_p = [E(|X|^p)]^{1/p}$, $p \geq 1$
  - $\delta_p(n) = \|g(\mathcal{F}_n) - g(\mathcal{F}_n^*)\|_p$,
- p-strong stability: $\sum_{n=0}^{\infty} \delta_p(n) < \infty$

[1] Wu (2005) Nonlinear system theory: another look at dependence. *Proc. Natl Acad. Sci. USA, 102, 14150–14154.*

Example: Linear processes

- $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, a_j are real coefficients
- ε_i are iid innovations and $\varepsilon_i \in \mathcal{L}^p$, $p > 0$
- $g(\mathcal{F}_n) - g(\mathcal{F}_n^*) = a_n(\varepsilon_0 - \varepsilon_0')$
- $\delta_p(n) = |a_n| c$, where $c = \|\varepsilon_0 - \varepsilon_0'\|_p$
- $(a_j)_{j \geq 0}$ are called impulse response function of the system (Box and Jenkins, 1970).

Example: Nonlinear Time Series

- Bilinear time series $X_n = (a + b\varepsilon_n)X_{n-1} + c\varepsilon_n$
- ARCH process $X_n = \varepsilon_n \sqrt{a^2 + b^2 X_{n-1}^2}$
- Threshold AR (TAR) process $X_n = \theta_1 X_{n-1}^+ + \theta_2 X_{n-1}^- + \varepsilon_n$
- Exponential autoregressive (EAR) process $X_n = [a + b\exp(-cX_{n-1}^2)]X_{n-1} + \varepsilon_n,\ \ c > 0$
- ...

Example: Nonlinear Time Series

- Let
$$X_n = R(X_{n-1}, \varepsilon_n) \qquad (2),$$
where R is bivariate measurable and ε_n iid innovations.
- Theorem (Diaconis and Freedman, 1999): The recursion (2) has a unique and stationary solution if there exist $\alpha > 0$ and x_0 such that
  - $L_{\varepsilon_0} + |R_{\varepsilon_0}(x_0)| \in \mathcal{L}^\alpha$
  - $E[\log(L_{\varepsilon_0})] < 0$, where
  - $L_\varepsilon = \sup_{x \neq x'} |R_{\varepsilon_0}(x) - R_{\varepsilon_0}(x')|/|x - x'|$.
- Theorem. (Wu and Shao, 2004) The above conditions imply that there exists a $p > 0$ such that the physical dependence measure
$$\delta_p(n) = O(r^n) \text{ for some } r \in (0, 1).$$

Example: Nonlinear Time Series

- Let
$$X_n = R(X_{n-1}, \varepsilon_n) \qquad (2),$$
where R is bivariate measurable and ε_n iid innovations.
- Theorem (Diaconis and Freedman, 1999): The recursion (2) has a unique and stationary solution if there exist $\alpha > 0$ and x_0 such that
    - $L_{\varepsilon_0} + |R_{\varepsilon_0}(x_0)| \in \mathcal{L}^\alpha$
    - $E[\log(L_{\varepsilon_0})] < 0$, where
    - $L_\varepsilon = \sup_{x \neq x'} |R_{\varepsilon_0}(x) - R_{\varepsilon_0}(x')|/|x - x'|$.
- Theorem. (Wu and Shao, 2004) The above conditions imply that there exists a $p > 0$ such that the physical dependence measure
$$\delta_p(n) = O(r^n) \text{ for some } r \in (0, 1).$$

Theorem[2] Assume that (X_i) is a causal process with mean 0 and $E(|X_i|^p) < \infty$ for some $p > 2$. Further assume that

$$\sum_{i=1}^{\infty} i\delta_p(i) < \infty. \qquad (4)$$

Then there exists a standard Brownian motion B such that on a richer probability space

$$\max_{i \le n} |S_i - \sigma B(i)| = o_{\mathrm{a.s.}}(n^{1/p'} \log n), \quad p' = \min(4, p), \qquad (5)$$

where $\sigma^2 = \sum_{k \in \mathbf{Z}} E(X_0 X_k)$ is the long-run variance.

Features:

- The rate is nearly optimal when $2 < p < 4$
- sharp enough for nonparametric inference
- Condition (4) is mild and easily verifiable

[2] Wu (2007), Strong invariance principles for dependent random variables. *Annals of Probability*

Outline of the proof

- Martingale Approximation (Gordin, Woodroofe, Volný, Peligrad, Utev etc):
  Approximate $S_n = \sum_{i=1}^n X_i$ by the martingale

$$M_n = \sum_{i=1}^{n} D_i, \text{ where } D_i = \sum_{j=i}^{\infty} \mathcal{P}_i X_j. \tag{6}$$

  Here $\mathcal{P}_i$ is the projection operator

$$\mathcal{P}_i \cdot = E[\cdot|\mathcal{F}_i] - E[\cdot|\mathcal{F}_{i-1}], \text{ where } \mathcal{F}_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots) \tag{7}$$

- Approximation Bound (Wu, 2007)

$$\|S_n - M_n\|_q^{q'} \le 3B_q^{q'} \sum_{j=1}^{n} \Theta_{j,q}^{q'}, \text{ where } \Theta_{j,q} = \sum_{i=j}^{\infty} \|\mathcal{P}_0 X_i\|_q. \tag{8}$$

Outline of the proof

- For $M_n = \sum_{i=1}^n D_i$, apply Strassen's martingale embedding scheme:
  On a possibly richer probability space, there exist a standard Brownian motion B and nonnegative random variables $\tau_1, \tau_2, \ldots$ with partial sums

$$T_k = \sum_{i=1}^{k} \tau_i \tag{9}$$

such that, for $k \geq 1$,

$$M_k = B(T_k), \quad E(\tau_k | \mathcal{F}_{k-1}) = E(D_k^2 | \mathcal{F}_{k-1}) \tag{10}$$

and

$$E(\tau_k^{q/2} | \mathcal{F}_{k-1}) \leq C_q E(|D_k|^q | \mathcal{F}_{k-1}) \tag{11}$$

almost surely, where C_q is a constant.

Outline of the proof

- Let $Q_n = \sum_{k=1}^{n} [\tau_k - E(\tau_k | \mathcal{F}_{k-1})]$.
- Write

$$T_n - n\sigma^2 = Q_n + \sum_{k=1}^{n} [E(\tau_k | \mathcal{F}_{k-1}) - E(D_k^2 | \mathcal{F}_{k-1})] + J_n$$

Therefore $T_n - n\sigma^2 = o_{\mathrm{a.s.}}(n^{2/q})$, and

$$\begin{aligned} \max_{k \le n} |B(T_k) - B(\sigma^2 k)| &\le \max_{k \le n} \sup_{x:\ |x - \sigma^2 k| \le n^{2/q}} |B(x) - B(\sigma^2 k)| \\ &= o_{\mathrm{a.s.}}[n^{1/q} (\log n)^{1/2}] \qquad (12) \end{aligned}$$

Possible Improvements

- Not good for vector-valued processes. Strassen type martingale embedding usually does not exist for vector-valued martingales (D. Monrad and W. Philipp, 1990).
- Strassen type martingale embedding scheme does not have advantage if $p > 4$.

Liu and Lin (2009) establish a strong invariance principle for a class of stationary processes.

- Let $X_i = g(\varepsilon_i, \varepsilon_{i-1}, \ldots)$ be a d-dimensional stationary causal process; $S_n = \sum_{i=1}^n X_i$.
- Assume $X_i \in \mathcal{L}^p$, $2 < p < 4$.
- Approximate $S_n = \sum_{i=1}^n X_i$ by $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_i$, where

$$\tilde{X}_i = E(X_i | \varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-m+1}) \tag{13}$$

are m-dependent, $m \in N$.

Liu and Lin (2009) apply the *m*-dependence approximation and block techniques, and prove the following result:
Theorem. If, for some $\tau > 0$, the following holds:

$$\Theta_{n,p} := \sum_{i=n}^{\infty} \delta_{i,p} = O(n^{(1-p/2)/(4-p)-\tau}), \tag{14}$$

then on a richer probability space, there exists an R^d valued Brownian motion $B(t)$ with covariance matrix Γ such that

$$\max_{i \le n} |S_i - B(i)| = o_{\text{a.s.}}(n^{1/p}). \tag{15}$$

How to deal with $p > 4$?

Berkes, Liu and Wu (2010, on progress). Consider

$$X_i = H(\varepsilon_{i-m}, \cdots, \varepsilon_i), \tag{16}$$

where $\varepsilon_k, k \in \mathbf{Z}$, are iid and H is a measurable function with $EX_i = 0$.

Theorem

Assume $EX_1 = 0$ and $E|X_1|^p < \infty$, $p > 2$. Then we can construct a probability space $(\Omega_c, \mathcal{A}_c, P_c)$ on which we can define iid random variables X_k^c, $k \geq 1$, and a standard Brownian motion $B_c(\cdot)$, such that $(X_k^c)_{k\geq 1} =_{\mathcal{D}} (X_k)_{k\geq 1}$ and

$$S_n^c - \sigma B_c(n) = o_{\mathrm{as}}(n^{1/p}) \text{ in } (\Omega_c, \mathcal{A}_c, P_c),$$

where $S_n^c = \sum_{k=1}^n X_k^c$.

Berkes, Liu and Wu (2010, on progress). Let

$$X_i = G(\mathcal{F}_i), \text{ where } \mathcal{F}_i = (\cdots, \varepsilon_{i-1}, \varepsilon_i), \tag{17}$$

To study the SIP for $S_n = \sum_{i=1}^n X_i$, we shall apply the m-dependence approximation. For $k \in \mathbf{N}$ let

$$m_k = \min\{3^b : \; 3^k (3^b)^{p/2-1} \geq (3^{k/2}\Theta_{3^b,p})^p, \; b \in \mathbf{N}\}. \tag{18}$$

How to deal with $p > 4$?

Example

Assume $\Theta_{n,p} = O(n^{-\beta})$, $\beta > 0$. Elementary calculations show that $m_k = O(3^{k\gamma})$, where $\gamma = (1/2 - 1/p)/(1/2 - 1/p + \beta)$. Let

$$\chi_k = 3^{\iota k} k^{1/2}, \text{ where } \iota = 1/p + \gamma(1/2 - 1/p) = \frac{4\beta - 2 + p}{4\beta p + 2p - 4}.$$

Then our Theorem asserts the SIP

$$\max_{1 \le j \le n} \left| \sum_{i=1}^{j} (X_i - \sigma Z_i) \right| = o_{\mathrm{as}}(\chi_{\log n}).$$

Note that $\iota \to 1/p$ if $\beta \to \infty$, and $\iota \to 1/2$ if $\beta \to 0$. The SIP approximation becomes better if β is bigger, indicating that the dependence is weaker. $\square$