Reading Group: Recursive Estimation of Time-Average Variance Constants (Wu, 2009)

HEMAN LEUNG

SPRING, 2020

Introduction

SECTION 1

Time-average variance constant (p.1)

Let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary and ergodic process with mean $\mu = E(X_0)$ and finite variance

- Denote covariance function by $\gamma_k = Cov(X_0, X_k)\ \forall k \in \mathbb{Z}$

Sample mean: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- Asymptotic normality under suitable conditions: $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$
- σ^2 here is called the time-average variance constant (TAVC) or long-run variance
  - Note that $Var(X_i) = \gamma_0 \neq \sigma^2$ in time series setting

Estimation of σ^2 is important for inference of time series

- Representation under suitable conditions: $\sigma^2 = \sum_{k\in\mathbb{Z}} \gamma_k$
  - Check previous reading group meeting (slide p.20, also check Keith's note) for the conditions

Overlapping batch means (p.2)

Overlapping batch means (OBM): $\hat{\sigma}^2_{obm}(n) = \frac{l_n}{n-l_n+1} \sum_{j=1}^{n-l_n+1} \left(\frac{1}{l_n} \sum_{i=j}^{j+l_n-1} X_i - \bar{X}_n \right)^2$

- First proposed by Meketon and Schmeiser (1984)
- Closely related to lag window estimator using Bartlett kernel (Newey & West, 1987)
  - An illustration assuming $\mu = 0$
  - Same AMSE if bandwidth l_n are both chosen optimally
- Nonoverlapping (NBM) version is also possible, but with worse properties
  - Song (2018) suggested an optimal linear combination of OBM and NBM would be better than solely using OBM
  - I discussed with Keith and we thought that her evidence was not solid enough (e.g. no theoretical properties shown)

Recursive estimation

Recursive formula for sample mean: $\bar{X}_n = \frac{n-1}{n}\bar{X}_{n-1} + \frac{1}{n}X_n$

Recursive formula for sample variance: $S_n^2 = \frac{n-2}{n-1}S_{n-1}^2 + \frac{1}{n}(X_n - \bar{X}_{n-1})^2$

- This is Welford's (1962) online algorithm

Recursive formula for TAVC: did not exist

- Note that $\hat{\sigma}_{obm}^2(n)$ has both $O(n)$ computational and memory complexity
  - When $l_n \neq l_{n-1}$, all batch means need to be updated
- However it is important for
  - Convergence diagnostics of MCMC
  - Sequential monitoring and testing

Notations (p.3)

$\mathcal{L}^p$ norm: $\|X\|_p \stackrel{\text{def}}{=} (E|X|^p)^{\frac{1}{p}}$, $X \in \mathcal{L}^p$ if $\|X\|_p < \infty$

- Write $\|X\| = \|X\|_2$

Same order: $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$

- $a_n \asymp b_n$ if $\exists c > 0$ such that $\frac{1}{c} \leq \left|\frac{a_n}{b_n}\right| \leq c$ for all large n

Let $S_n = \sum_{i=1}^{n} X_i - n\mu$ and $S_n^* = \max_{i \leq n} |S_i|$

Recursive TAVC estimates

SECTION 2

Algorithm when $\mu = 0$

Start of each block: $\{a_k\}_{k\in\mathbb{N}}$ is a strictly increasing integer sequence such that

- $a_1 = 1$ and $a_{k+1} - a_k \to \infty$ as $k \to \infty$
- Start of each batch: $t_i = a_k$ if $a_k \le i < a_{k+1}$

Component: $V_n = \sum_{i=1}^{n} W_i^2$ where $W_i = X_{t_i} + X_{t_i+1} + \cdots + X_i$

- $v_n = \sum_{i=1}^{n} l_i$ where $l_i = i - t_i + 1$
- Observe that W_i is the batch sum and l_i is the batch size

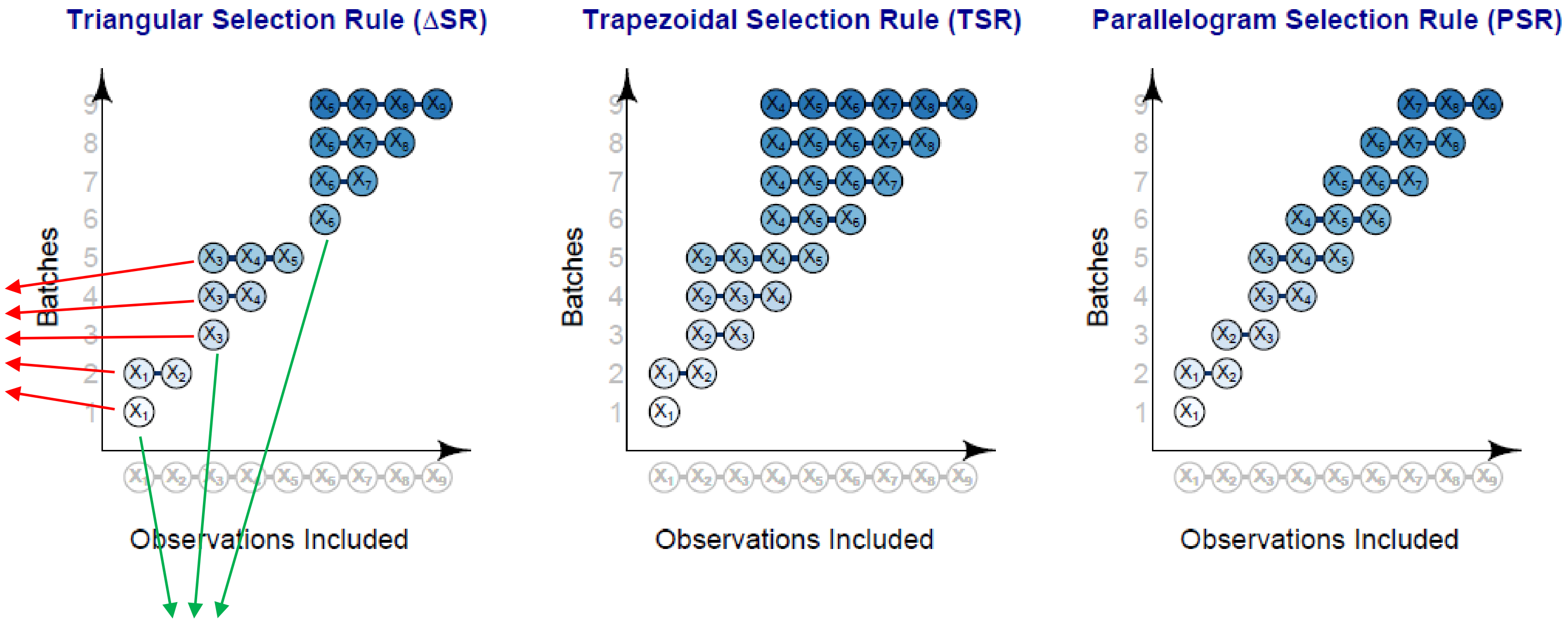
Algorithm: at stage n, we store $\left(n, k_n, a_{k_n}, v_n, V_n, W_n\right)$. At stage $n+1$,

- If $n + 1 = a_{k_n+1}$, set $k_{n+1} = k_n + 1$ and $W_{n+1} = X_{n+1}$. Otherwise set $k_{n+1} = k_n$ and $W_{n+1} = W_n + X_{n+1}$
- Set $V_{n+1} = V_n + W_{n+1}^2$ and $v_{n+1} = v_n + \left(n + 2 - a_{k_{n+1}}\right)$ since $t_{n+1} = a_{k_{n+1}}$
- The estimate is $\hat{\sigma}_{\Delta SR}^2(n+1) = \frac{V_{n+1}}{v_{n+1}}$

Graphical illustration (Chan and Yau, 2017)

Intuitions

Start of each batch $= t_i$



Start of each block $= a_k$; thus a block B_k contains $\{a_k, a_k + 1, \dots, a_{k+1} - 1\}$

Choice of a_k and t_n (p.3-4)

A simple choice is $a_k = \lfloor ck^p \rfloor$ where $c > 0$ and $p > 1$ are constants

- Optimal choice of functional is not known
  - I discussed with Keith and we need to resort to variational calculus for this problem
  - However it seems to be unsolvable without proper boundary conditions (tried on SymPy)

Note that t_n is implicitly determined by choice of a_k

- Since $a_k \leq n < a_{k+1}$, choosing $a_k = \lfloor ck^p \rfloor$ means $ck^p - 1 < n < c(k+1)^p - 1$
- Solving $k = k_n$ from the above inequalities, we have
- $t_n = a_{k_n}$ where $k_n = \left\lceil \left(\frac{n+1}{c}\right)^{\frac{1}{p}} \right\rceil - 1$

Modification when $\mu \neq 0$ (p.4-5)

General component: $V_n' = \sum_{i=1}^{n} (W_i')^2$ where $W_i' = X_{t_i} + X_{t_i+1} + \cdots + X_i - l_i \bar{X}_n$

- Observe that $(W_i')^2 = W_i^2 - 2 l_i W_i \bar{X}_n + (l_i \bar{X}_n)^2$
- Let $U_n = \sum_{i=1}^{n} l_i W_i$ and $q_n = \sum_{i=1}^{n} l_i^2$
  - Note that they can also be updated recursively
- Then $V_n' = V_n - 2 U_n \bar{X}_n + q_n (\bar{X}_n)^2$ and $\hat{\sigma}_{\Delta SR}^2(n) = \frac{V_n'}{v_n}$
- Complete algorithm is similar to previous logic so we skip it here

Generalization to spectral density estimation is possible

- Relation between spectral density and TAVC was discussed in previous reading group (slide p.47)

Convergence properties

SECTION 3

Representation of TAVC (p.5-6)

Consider Wu's (2005) nonlinear Wold process

- Weak stability with $p = 2$ (i.e. $\Omega_2 < \infty$) guarantees invariance principle, which entails CLT

Representation of TAVC

- Assume $E(X_i) = 0$ and $\sum_{i=0}^{\infty} \|\mathcal{P}_0 X_i\|_2 < \infty$ where $\mathcal{P}_i \cdot := E(\cdot\,|\mathcal{F}_i) - E(\cdot\,|\mathcal{F}_{i-1})$
  - The later assumption is equivalent to $\Omega_2 < \infty$ (which suggest short-range dependence)
- Then $D_k \overset{\text{def}}{=} \sum_{i=k}^{\infty} \mathcal{P}_k X_i \in \mathcal{L}^2$ and is a stationary martingale difference sequence w.r.t. $\mathcal{F}_k$
  - Proved in previous reading group (slide p.21)
- By theorem 1 in Hannan (1979), we have invariance principle and $\sigma = \|D_k\|_2$
  - Why not $\|D_0\|_2$? Because they have same distribution by stationarity and we cannot observe X_0 in practice
- Let $S_n = \sum_{i=1}^{n} X_i$ and $M_n = \sum_{i=1}^{n} D_i$
- If $\Omega_\alpha < \infty$ for $\alpha > 2$, then $\|S_n - M_n\|_\alpha = o(\sqrt{n})$
  - This partly comes from moment inequality. See previous reading group (slide p.20)

SECTION 3.1

Moment convergence

Moment convergence (p.6-7)

Theorem 1: let $E(X_i) = 0$ and $X_i \in \mathcal{L}^{\alpha}$ where $\alpha > 2$

- Assume $\sum_{i=0}^{\infty} \|\mathcal{P}_0 X_i\|_{\alpha} < \infty$
  - Equivalent to $\Omega_{\alpha} < \infty$, which is mild as σ^2 does not always exist for long-range dependent processes
- Further assume as $m \to \infty$, $a_{m+1} - a_m \to \infty$ and $\frac{(a_{m+1}-a_m)^2}{\sum_{k=2}^{m}(a_k-a_{k-1})^2} \to 0$
  - Earlier condition $a_{m+1} - a_m \to \infty$ is needed to account for dependence
  - Later condition is needed so that a_m does not diverge to ∞ so fast
- Then $\left\|\frac{V_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$
  - This implies finite forth moment is not necessary for consistency of $\hat{\sigma}^2_{\Delta SR}(n)$ (e.g. take $\alpha = 3$)
  - Convergence in $\mathcal{L}^{\frac{\alpha}{2}}$ norm where $\alpha > 2$ implies convergence in probability (i.e. consistency)

Corollary 1: under same assumptions of theorem 1, we also have $\left\|\frac{V_n'}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$

Proof of theorem 1: blocking (p.13)

Blocking: for $n \in \mathbb{N}$ choose $m = m_n \in \mathbb{N}$ such that $a_m \leq n < a_{m+1}$

- m represent total number of complete blocks
- Then $v_n = \sum_{j=1}^{n}\left(j - t_j + 1\right) = \sum_{i=2}^{m}\sum_{j=a_{i-1}}^{a_i-1}\left(j - t_j + 1\right) + \sum_{j=a_m}^{n}\left(j - t_j + 1\right)$
- $= \frac{1}{2}\sum_{i=2}^{m}(a_i - a_{i-1})(a_i - a_{i-1} + 1) + \frac{1}{2}(n - a_m)(n - a_m + 1)$
  - $\sim \frac{1}{2}\sum_{i=2}^{m}(a_i - a_{i-1})^2$ by assumption of theorem 1

Note that $1 \leq \liminf\limits_{m\to\infty}\frac{v_n}{v_{a_m}} \leq \limsup\limits_{m\to\infty}\frac{v_{a_{m+1}}}{v_{a_m}}$ since $v_{a_{m+1}} \geq v_n$ (?)

- By assuming $\frac{(a_{m+1}-a_m)^2}{\sum_{k=2}^{m}(a_k-a_{k-1})^2} \to 0$, $\limsup\limits_{m\to\infty}\frac{v_{a_{m+1}}}{v_{a_m}} = 1$
- Hence both limits are 1

Proof of theorem 1: martingale approximation (p.13)

For any fixed $k_0 \in \mathbb{N}$, since $a_{m+1} - a_m$ is increasing to ∞, we have

- $\lim_{m\to\infty} \frac{1}{v_n} \sum_{i=1}^{n} \mathbb{I}(i - t_i + 1 \leq k_0) \leq \lim_{m\to\infty} \frac{1}{v_n} m k_0 = 0$
  - Using $(m+1)k_0$ is better (?)

Martingale approximation: $\sum_{i=0}^{\infty} \|\mathcal{P}_0 X_i\|_\alpha < \infty$ implies $D_k = \sum_{i=k}^{\infty} \mathcal{P}_k X_i \in \mathcal{L}^\alpha$

- Let $M_n = \sum_{i=1}^{n} D_i$. By theorem 1 in Wu (2007), the above condition also implies
- $\|S_n\|_\alpha = O(\sqrt{n}), \|M_n\|_\alpha = O(\sqrt{n})$ and $\|S_n - M_n\|_\alpha = o(\sqrt{n})$
- Hence as $n \to \infty$, $\rho_n \overset{\text{def}}{=} \frac{1}{n} \|S_n^2 - M_n^2\|_{\frac{\alpha}{2}} \leq \frac{1}{n} \|S_n - M_n\|_\alpha \|S_n + M_n\|_\alpha \to 0$
  - Inequality by Cauchy–Schwarz: $\|(S_n - M_n)(S_n + M_n)\|_{\frac{\alpha}{2}} \leq \|S_n - M_n\|_\alpha \|S_n + M_n\|_\alpha$
- Aim to approximate V_n by $Q_n = \sum_{i=1}^{n} R_i^2$ where $R_i = D_{t_i} + D_{t_i+1} + \cdots + D_i$
  - Such that $\|Q_n - V_n\|_{\frac{\alpha}{2}} = o(v_n)$ and show that $\left\|\frac{Q_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$

Proof of theorem 1: $\|Q_n - V_n\|_{\frac{\alpha}{2}} = o(v_n)$ (p.13)

$\limsup_{n\to\infty} \frac{1}{v_n} \|V_n - Q_n\|_{\frac{\alpha}{2}} \le \limsup_{n\to\infty} \frac{1}{v_n} \sum_{i=1}^{n} \left\|R_i^2 - W_i^2\right\|_{\frac{\alpha}{2}}$ (by Minkowski inequality)

- $\le \limsup_{n\to\infty} \frac{1}{v_n} \sum_{i=1}^{n} (i - t_i + 1)\rho_{i-t_i+1}$ (by definition of ρ_n and stationarity)
- $\le \limsup_{n\to\infty} \frac{1}{v_n} \sum_{1\le i\le n: i-t_i+1>k_0} (i - t_i + 1)\rho_{i-t_i+1}$ (by $\lim_{m\to\infty} \frac{1}{v_n} \sum_{i=1}^{n} \mathbb{I}(i - t_i + 1 \le k_0) = 0$)
- $\le \sup_{k\ge k_0} \rho_k$ (by $\sum(i - t_i + 1)\rho_{i-t_i+1} \le \sup_{k\ge k_0} \rho_k \sum(i - t_i + 1)$)
- $\to 0$ (by $\rho_n \to 0$ as $n \to \infty$)

Proof of theorem 1: $\left\|\frac{Q_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$ (p.14)

Recall that $t_i = a_k$ if $a_k \le i \le a_{k+1} - 1$

- Block square of sum: $Y_k = \sum_{i=a_k}^{a_{k+1}-1} \left(D_{t_i} + D_{t_i+1} + \cdots + D_i\right)^2 = \sum_{i=a_k}^{a_{k+1}-1} \left(D_{a_k} + D_{a_k+1} + \cdots + D_i\right)^2$
- Block sum of square: $\tilde{Y}_k = \sum_{i=a_k}^{a_{k+1}-1} \left(D_{a_k}^2 + D_{a_k+1}^2 + \cdots + D_i^2\right)$
- $\|Y_k\|_{\frac{\alpha}{2}} \le \sum_{i=a_k}^{a_{k+1}-1} \left\|\left(D_{a_k} + D_{a_k+1} + \cdots + D_i\right)^2\right\|_{\frac{\alpha}{2}}$ (by Minkowski inequality)
- $= \sum_{i=a_k}^{a_{k+1}-1} \left\|D_{a_k} + D_{a_k+1} + \cdots + D_i\right\|_{\alpha}^2$
- $\le \sum_{i=a_k}^{a_{k+1}-1} c_\alpha (i - a_k + 1) \|D_1\|_\alpha^2$ where c_α is a constant which only depends on α
  - By Burkholder's inequality and $\mathcal{L}^\alpha$ stationarity. See previous reading group (slide p. 21-22)
- On the other hand, $\left\|\tilde{Y}_k\right\|_{\frac{\alpha}{2}} \le \sum_{i=a_k}^{a_{k+1}-1} (i - a_k + 1) \|D_1\|_\alpha^2$ (by Minkowski inequality and $\mathcal{L}^\alpha$ stationarity)

Proof of theorem 1: $\left\|\frac{Q_n}{v_n}-\sigma^2\right\|_{\frac{\alpha}{2}}=o(1)$ (p.14-15)

Since $1<\frac{\alpha}{2}\leq 2$ and $Y_k-E\left(Y_k|\mathcal{F}_{a_k}\right)$ is a MDS, we have

- It seems this impose $\alpha\leq 4$ on theorem 1
- $\left\|\sum_{k=1}^{m}\left[Y_k-E\left(Y_k|\mathcal{F}_{a_k}\right)\right]\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}\leq c_\alpha\sum_{k=1}^{m}\left\|Y_k-E\left(Y_k|\mathcal{F}_{a_k}\right)\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$ (by Burkholder's inequality)
- $\leq c_\alpha\sum_{k=1}^{m}\|Y_k\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$ (by Jensen's inequality, c_α actually changes)
- Similarly, $\left\|\sum_{k=1}^{m}\left[\tilde{Y}_k-E\left(\tilde{Y}_k|\mathcal{F}_{a_k}\right)\right]\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}\leq c_\alpha\sum_{k=1}^{m}\left\|\tilde{Y}_k\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$

Note that D_i are also MDS and $E\left(\tilde{Y}_k|\mathcal{F}_{a_k}\right)=E\left(Y_k|\mathcal{F}_{a_k}\right)$

- Difference between $\tilde{Y}_k$ and Y_k lies in the cross terms, e.g. $D_{a_k}D_{a_k+1}$
- However by property of MDS, $E\left(D_{a_k}D_{a_k+1}\right)=0$

Proof of theorem 1: $\left\|\frac{Q_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$ (p.15)

Note that $\left\|\sum_{k=1}^{m}\left(Y_k - \tilde{Y}_k\right)\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}} = \left\|\sum_{k=1}^{m}\left[Y_k - \tilde{Y}_k - E\left(Y_k \middle| \mathcal{F}_{a_k}\right) + E\left(Y_k \middle| \mathcal{F}_{a_k}\right)\right]\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$

- We do not work on cross-term directly with Minkowski directly as the bound is looser

- $\leq c_\alpha \sum_{k=1}^{m}\left(\|Y_k\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}} + \left\|\tilde{Y}_k\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}\right)$ (by Minkowski and inequalities proved in last slide)
- $\leq c_\alpha \|D_1\|_\alpha^\alpha \sum_{k=1}^{m}\left[\sum_{i=a_k}^{a_{k+1}-1}(i - a_k + 1)\right]^{\frac{\alpha}{2}}$ (by inequalities proved in two slides ago)
- $\leq c_\alpha \|D_1\|_\alpha^\alpha \max_{h\leq m}\left[\sum_{i=a_h}^{a_{h+1}-1}(i - a_h + 1)\right]^{\frac{\alpha}{2}-1} \sum_{k=1}^{m}\left[\sum_{i=a_k}^{a_{k+1}-1}(i - a_k + 1)\right]$
  - Recall that $v_{a_m} = \sum_{k=1}^{m}\left[\sum_{i=a_k}^{a_{k+1}-1}(i - a_k + 1)\right]$ by blocking

Proof of theorem 1: $\left\|\frac{Q_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$ (p.15)

Now $v_n^{-\frac{\alpha}{2}} \left\|\sum_{k=1}^{m}\left(Y_k - \tilde{Y}_k\right)\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}} \le v_n^{-\frac{\alpha}{2}+1} c_\alpha \|D_1\|_\alpha^\alpha \max_{h\le m}\left[\sum_{i=a_h}^{a_{h+1}-1}(i - a_h + 1)\right]^{\frac{\alpha}{2}-1}$

- By $1 \le \liminf_{m\to\infty} \frac{v_n}{v_{a_m}} \le \limsup_{m\to\infty} \frac{v_{a_{m+1}}}{v_{a_m}} = 1$
- $\le c_\alpha \|D_1\|_\alpha^\alpha \left[\frac{\max_{h\le m}(a_{h+1}-a_h)^2}{v_n}\right]^{\frac{\alpha}{2}-1}$ $\to 0$ (by $\frac{(a_{m+1}-a_m)^2}{\sum_{k=2}^{m}(a_k-a_{k-1})^2} \to 0$)

Ergodic theorem: since $D_k^2 \in \mathcal{L}^{\frac{\alpha}{2}}$, we have $\left\|D_1^2 + \cdots + D_l^2 - l\sigma^2\right\|_{\frac{\alpha}{2}} = o(l)$

- Therefore $\left\|\tilde{Y}_k - E\left(\tilde{Y}_k\right)\right\|_{\frac{\alpha}{2}} = o[(a_{k+1} - a_k)^2]$
  - Recall that $\tilde{Y}_k = \sum_{i=a_k}^{a_{k+1}-1}\left(D_{a_k}^2 + D_{a_k+1}^2 + \cdots + D_i^2\right)$. The sum is a isosceles triangular shaped
- Then $\lim_{n\to\infty} \frac{1}{v_n} \left\|\sum_{k=1}^{m}\left[\tilde{Y}_k - E\left(\tilde{Y}_k\right)\right]\right\|_{\frac{\alpha}{2}} = \lim_{n\to\infty} \frac{1}{v_n} \sum_{k=1}^{m} o[(a_{k+1} - a_k)^2] = 0$
  - By Minkowski inequality and property of little o

Proof of theorem 1: $\left\|\frac{Q_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = o(1)$ (p.15)

Since $\frac{1}{v_n}\left\|\sum_{k=1}^{m}\left(Y_k - \tilde{Y}_k\right)\right\|_{\frac{\alpha}{2}} \to 0 \Leftrightarrow \left\|\sum_{k=1}^{m}\left(Y_k - \tilde{Y}_k\right)\right\|_{\frac{\alpha}{2}} = o(v_n)$ (first part in last slide)

- And $\lim_{n\to\infty}\frac{1}{v_n}\left\|\sum_{k=1}^{m}\left[\tilde{Y}_k - E\left(\tilde{Y}_k\right)\right]\right\|_{\frac{\alpha}{2}} = 0 \Leftrightarrow \left\|\sum_{k=1}^{m}\left[\tilde{Y}_k - E\left(\tilde{Y}_k\right)\right]\right\|_{\frac{\alpha}{2}} = o(v_n)$ (second part in last slide)
- We have $\left\|\sum_{k=1}^{m}\left[Y_k - E\left(\tilde{Y}_k\right)\right]\right\|_{\frac{\alpha}{2}} = \|\sum_{k=1}^{m}[Y_k - E(Y_k)]\|_{\frac{\alpha}{2}}$ (by $E\left(\tilde{Y}_k \middle| \mathcal{F}_{a_k}\right) = E\left(Y_k \middle| \mathcal{F}_{a_k}\right)$)
- $= \left\|\sum_{k=1}^{m} Y_k - v_{a_m}\sigma^2\right\|_{\frac{\alpha}{2}} = o\left(v_{a_m}\right)$ (by ergodic theorem)

Finally we compare Q_n and $Q_{a_{m+1}-1} = \sum_{k=1}^{m} Y_k$

- $\left\|Q_n - Q_{a_{m+1}-1}\right\|_{\frac{\alpha}{2}} = \left\|\sum_{i=n+1}^{a_{m+1}-1} R_i^2\right\|_{\frac{\alpha}{2}}$ (recall $R_i = D_{t_i} + D_{t_i+1} + \cdots + D_i$)
- $\leq \sum_{i=n+1}^{a_{m+1}-1}\|R_i\|_{\alpha}^2$ (by Minkowski inequality)
- $= \sum_{i=n+1}^{a_{m+1}-1} O(i - t_i + 1) \leq (a_{m+1} - a_m)^2 = o(v_n)$ (by $\frac{(a_{m+1}-a_m)^2}{\sum_{k=2}^{m}(a_k-a_{k-1})^2} \to 0$)

Proof of corollary 1: requirement (p.15)

Note that V_n' remains unchanged if X_i is replaced by $X_i - \mu$

- Hence we can assume $\mu = 0$ wlog
- By $V_n' = V_n - 2U_n\bar{X}_n + q_n(\bar{X}_n)^2$ and theorem 1, it suffices to verify
- $\|U_n\bar{X}_n\|_{\frac{\alpha}{2}} = o(v_n)$ and
- $\|q_n(\bar{X}_n)^2\|_{\frac{\alpha}{2}} = o(v_n)$

By moment inequality, $\|S_n\|_\alpha = O(\sqrt{n}) \Rightarrow \|\bar{X}_n\|_\alpha = O\left(n^{-\frac{1}{2}}\right)$

Proof of corollary 1: $\|q_n(\bar{X}_n)^2\|_{\frac{\alpha}{2}} = o(v_n)$ (p.16)

Choose $m \in \mathbb{N}$ such that $a_m \le n < a_{m+1}$, we have

- $(a_{m+1} - a_m)^2 = o(1) \sum_{k=2}^{m} (a_k - a_{k-1})^2$ (by $\frac{(a_{m+1}-a_m)^2}{\sum_{k=2}^{m}(a_k-a_{k-1})^2} \to 0$)
- $\le o(1)[\sum_{k=2}^{m}(a_k - a_{k-1})]^2 = o(a_m^2)$ (by a_k is positive and telescoping sum)

Since $a_m \to \infty$ and is increasing, $\max_{l \le m}(a_{l+1} - a_l) = o(a_m) = o(n)$ (by result of the above)

- Recall that $q_n = \sum_{i=1}^{n} l_i^2$ and $v_n = \sum_{i=1}^{n} l_i$, we have
- $q_n \le v_n \max_{l \le m}(a_{l+1} - a_l)$ (by blocking)
- $= v_n o(n)$

Hence $\|q_n(\bar{X}_n)^2\|_{\frac{\alpha}{2}} = v_n o(n) O(n^{-1}) = o(v_n)$

- $o(a_n)O(b_n) = o(a_n b_n)$ (little o times big O is little o)

Proof of corollary 1: $\|U_n\bar{X}_n\|_{\frac{\alpha}{2}} = o(v_n)$ (p.16)

If $\|U_n\|_\alpha = O(1)\sqrt{\sum_{l=1}^{m}(a_{l+1}-a_l)^5}$, then we have

- $\|U_n\bar{X}_n\|_{\frac{\alpha}{2}} \leq \|U_n\|_\alpha \|\bar{X}_n\|_\alpha$ (by Cauchy–Schwarz inequality)
- $= O\left(n^{-\frac{1}{2}}\right)\sqrt{\sum_{l=1}^{m}(a_{l+1}-a_l)^5}$ (by moment inequality)
- $\leq O\left(n^{-\frac{1}{2}}\right)\left[\sum_{l=1}^{m}(a_{l+1}-a_l)^2\right]\sqrt{\max_{l\leq m}(a_{l+1}-a_l)}$ (by $\sum_{l=1}^{m}(a_{l+1}-a_l)^4 \leq \left[\sum_{l=1}^{m}(a_{l+1}-a_l)^2\right]^2$)
- $= O\left(n^{-\frac{1}{2}}\right)o\left(n^{\frac{1}{2}}\right)\left[\sum_{l=1}^{m}(a_{l+1}-a_l)^2\right]$ (by $\max_{l\leq m}(a_{l+1}-a_l) = o(n)$)
- $= O\left(n^{-\frac{1}{2}}\right)o\left(n^{\frac{1}{2}}\right)o(v_n)$ (by blocking)
- $= o(v_n)$ (little o times big O is little o)

Now we only need to prove $\|U_n\|_\alpha = O(1)\sqrt{\sum_{l=1}^{m}(a_{l+1}-a_l)^5}$

Proof of corollary 1: $\|U_n \bar{X}_n\|_{\frac{\alpha}{2}} = o(v_n)$ (p.16)

Recall $l_i = i - t_i + 1$ and $U_n = \sum_{i=1}^{n} l_i W_i$ where $W_i = X_{t_i} + X_{t_i+1} + \cdots + X_i$

- Let $h_j = h_{j,n} = \sum_{i=1}^{n} l_i \mathbb{I}(t_i \leq j \leq i)\,,\ j = 1, \ldots, n$
- Then $U_n = \sum_{i=1}^{n} l_i \sum_{j=t_i}^{i} X_j = \sum_{j=1}^{n} X_j h_j$
- Since $X_j = \sum_{k=0}^{\infty} \mathcal{P}_{j-k} X_j$ and $\mathcal{P}_{j-k} X_j$ is MDS, we have
- $\|U_n\|_\alpha \leq \sum_{k=0}^{\infty} \left\| \sum_{j=1}^{n} \mathcal{P}_{j-k} X_j h_j \right\|_\alpha$ (by Minkowski inequality)
- $\leq \sum_{k=0}^{\infty} c_\alpha \sqrt{\sum_{j=1}^{n} \| \mathcal{P}_{j-k} X_j h_j \|_\alpha^2}$ (by Burkholder's inequality, not trivial?)
- $= c_\alpha \sqrt{\sum_{j=1}^{n} h_j^2} \sum_{k=0}^{\infty} \|\mathcal{P}_0 X_k\|_\alpha$ (by $\mathcal{L}^\alpha$ stationarity)
- By blocking, $\sum_{j=1}^{n} h_j^2 \leq \sum_{k=1}^{m} \sum_{j=a_k}^{a_{k+1}-1} h_j^2 \leq \sum_{k=1}^{m} \sum_{j=a_k}^{a_{k+1}-1} (a_{k+1} - a_k)^4 = \sum_{k=1}^{m} (a_{k+1} - a_k)^5$
- Hence $\|U_n\|_\alpha = O(1) \sqrt{\sum_{k=1}^{m} (a_{k+1} - a_k)^5}$ (by $\sum_{i=0}^{\infty} \|\mathcal{P}_0 X_i\|_\alpha < \infty$)

Proof of moment convergence: summary of techniques

Begin with martingale approximation

- Cater for dependence in time series
  - Projection decomposition available as MDS ($X_j = \sum_{k=0}^{\infty} \mathcal{P}_{j-k} X_j$)
- Enable the use of ergodic theorem for moment convergence
  - WLLN under dependence. Check theorem 7.12 and 7.21 in Keith's STAT4010
- Handle approximation difference with norm and little o (e.g. Y_k and $\tilde{Y}_k$)
  - MDS is uncorrelated

Handle remainder term (e.g. V_n vs V_{a_m})

- By blocking and assumption on growth rate of start of block a_m
  - Suitable for subsampling or even general time series (e.g. m-dependent)
  - Allow sharper bound to be derived. See proof related to $\left\|\sum_{k=1}^{m}\left(Y_k - \tilde{Y}_k\right)\right\|_{\frac{\alpha}{2}}$. Also check lemma 1 in Liu and Wu (2010)
  - Bounding a weighted sum, which may be useful for say SLLN. See proof related to U_n. Also check Kronecker's lemma

Convergence rate, $2 < \alpha \leq 4$

SECTION 3.2.1

Convergence rate (p.8)

Theorem 2: let $a_k = \lfloor ck^p \rfloor, k \geq 1$ where $c > 0$ and $p > 1$ are constants

Theorem 2.1: assume that $X_i \in \mathcal{L}^\alpha$, $E(X_i) = 0$ and $\Delta_\alpha = \sum_{j=0}^{\infty} \delta_\alpha(j) < \infty$ for some $\alpha \in (2,4]$

- Then $\|V_n - E(V_n)\|_{\frac{\alpha}{2}} = O\left(n^{\frac{3}{2} - \frac{3}{2p} + \frac{2}{\alpha}}\right)$

Theorem 2.2: assume that $X_i \in \mathcal{L}^\alpha$, $E(X_i) = 0$ and $\Delta_\alpha = \sum_{j=0}^{\infty} \delta_\alpha(j) < \infty$ for some $\alpha > 4$

- Then $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$

Theorem 2.3: if $X_i \in \mathcal{L}^2$, $E(X_i) = 0$ and $\sum_{j=0}^{\infty} j^q \omega(j) < \infty$ for some $q \in (0,1]$

- Then $E(V_n - v_n\sigma^2) = O\left[n^{1+(1-q)\left(1-\frac{1}{p}\right)}\right]$
- Consequently, if theorem 2.1 also holds, then $\|V_n - v_n\sigma^2\|_{\frac{\alpha}{2}} = O\left(n^{\phi}\right)$
  - $\phi = \max\left[\frac{3}{2} - \frac{3}{2p} + \frac{2}{\alpha}, 1 + (1-q)\left(1 - \frac{1}{p}\right)\right]$
  - $\sum_{j=1}^{\infty} j^q \delta_\alpha(j) < \infty$ is sufficient

Optimal convergence rate (p.8)

To achieve optimal convergence, we should minimize $\phi = \max\left[\frac{3}{2} - \frac{3}{2p} + \frac{2}{\alpha}, 1 + (1-q)\left(1 - \frac{1}{p}\right)\right]$

- Theorem 2 guides us to choose p based on q (dependence condition) and α (moment condition)
- A good p should minimize $n^{\frac{3}{2}-\frac{3}{2p}+\frac{2}{\alpha}} + n^{1+(1-q)\left(1-\frac{1}{p}\right)}$, which also minimize ϕ
- Set $\frac{3}{2} - \frac{3}{2p} + \frac{2}{\alpha} = 1 + (1-q)\left(1 - \frac{1}{p}\right)$ and solve for p
  - The rationale is that the optimal rate should be the same regardless of conditions which are hard to verify?
- We have $p = \frac{\frac{1}{2}+q}{q-\frac{1}{2}+\frac{2}{\alpha}}$ (denominator should be $q - \frac{1}{2} + \frac{2}{\alpha}$, probably typo in the paper)

Corollary 2: Let $p = \frac{\frac{1}{2}+q}{q-\frac{1}{2}+\frac{2}{\alpha}}$. Under conditions of theorem 2, $\left\|\frac{V_n}{v_n} - \sigma^2\right\|_{\frac{\alpha}{2}} = O\left(n^{\frac{2}{\alpha}-\frac{1}{2}-\frac{1}{2p}}\right)$

- In particular, if $\alpha = 4$ and $q = 1$, then $p = \frac{3}{2}$ and $\left\|\frac{V_n}{v_n} - \sigma^2\right\|_2 = O\left(n^{-\frac{1}{3}}\right)$

Convergence rate when $\mu \neq 0$ (p.9)

Note that $v_n \sim v_{a_m} \sim \frac{1}{2}\sum_{i=2}^{m}(a_i - a_{i-1})^2$ (by blocking)

- $\sim \frac{1}{2}\sum_{i=2}^{m} c^2 p^2 i^{2p-2}$ (by considering the differential $a_i - a_{i-1} \sim cpi^{p-1}$)
- $\sim \frac{c^2 p^2 m^{2p-1}}{4p-2}$ (by approximating sum $\Sigma_{x=2}^{m}$ with integral $\int_2^m dx$)
- $\sim \frac{c^{\frac{1}{p}} p^2}{4p-2} n^{2-\frac{1}{p}} = O\left(n^{2-\frac{1}{p}}\right)$ (by $n \sim cm^p \Rightarrow m \sim \left(\frac{n}{c}\right)^{\frac{1}{p}}$)

Corollary 2 also applies to $\frac{V_n'}{v_n}$ since $\frac{1}{v_n}\|V_n - V_n'\|_{\frac{\alpha}{2}} = O\left(n^{-\frac{1}{p}}\right)$ and $-\frac{1}{p} < \frac{2}{\alpha} - \frac{1}{2} - \frac{1}{2p}$

- This implies the difference $V_n - V_n'$ cannot be the dominating term
- See remark 4 in paper for proof of $\frac{1}{v_n}\|V_n - V_n'\|_{\frac{\alpha}{2}}$

Proof of theorem 2.1: $\|V_n - E(V_n)\|_{\frac{\alpha}{2}} = O\left(n^{\frac{3}{2}-\frac{3}{2p}+\frac{2}{\alpha}}\right)$ (p.17-18)

Recall $V_n = \sum_{i=1}^{n} W_i^2$. Note that $\|V_n - E(V_n)\|_{\frac{\alpha}{2}} \le \|\sum_{i=1}^{n} W_i^2\|_{\frac{\alpha}{2}}$ (V_n is non-negative)

- $= \left\|\sum_{i=1}^{n}\sum_{k=0}^{\infty}\mathcal{P}_{i-k}W_i^2\right\|_{\frac{\alpha}{2}}$ (by $W_i^2 = \sum_{k=0}^{\infty}\mathcal{P}_{i-k}W_i^2$)
- $\le \sum_{k=0}^{\infty}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}}$ (by Minkowski inequality)
- It suffices to find the order of $\left\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\right\|_{\frac{\alpha}{2}}$

Blocking: let $b_m = \lfloor (1+c)p2^p m^{p-1} \rfloor$

- It can be shown that $i - t_i \le a_{m+1} - 1 - a_m \le b_m \ \forall m \in \mathbb{N}$
  - Obviously the functional of b_m is chosen by solving this inequality
  - This also means that b_m is the bound of block size and batch size
- $\sum_{k=0}^{\infty}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}} = \sum_{k=2b_m}^{\infty}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}} + \sum_{k=0}^{2b_m-1}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}}$

Proof of theorem 2.1: bound of $\left\|\sum_{i=1}^{n} \mathcal{P}_{i-k} W_i^2\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$ (p.17)

Recall that $W_i = X_{t_i} + X_{t_i+1} + \cdots + X_i$. Let $W_i^* = X'_{t_i} + X'_{t_i+1} + \cdots + X'_i$ (coupled batch sum)

- Since $\epsilon'_0 \perp \epsilon_i,\ i \in \mathbb{Z}$, we have $E(X_i|\mathcal{F}_{-1}) = E(X'_i|\mathcal{F}_{-1}) = E(X'_i|\mathcal{F}_0)$
- Stability assumption $\Delta_\alpha < \infty$ implies weak stability $\Theta_\alpha < \infty$
- By theorem 1 in Wu (2007), $\|W_i\|_\alpha \le c_\alpha \Theta_\alpha \sqrt{i - t_i + 1}$ (moment inequality)
- Now $\left\|\mathcal{P}_0 W_i^2\right\|_{\frac{\alpha}{2}} = \left\|E\left(W_i^2\middle|\mathcal{F}_0\right) - E\left(W_i^2\middle|\mathcal{F}_{-1}\right)\right\|_{\frac{\alpha}{2}}$ (definition of projection)
- $= \left\|E\left(W_i^2\middle|\mathcal{F}_0\right) - E[(W_i^*)^2|\mathcal{F}_0]\right\|_{\frac{\alpha}{2}}$ (property of coupled batch sum)
- $\le \left\|W_i^2 - (W_i^*)^2\right\|_{\frac{\alpha}{2}}$ (by Jensen's inequality and tower property)
- $\le \|W_i + W_i^*\|_\alpha \|W_i - W_i^*\|_\alpha$ (by Cauchy–Schwarz inequality)
- $\le 2\|W_i\|_\alpha \sum_{j=t_i}^{i} \delta_\alpha(j)$ (property of coupled batch sum and definition of physical dependence)
- $\le 2c_\alpha \Theta_\alpha \sqrt{i - t_i + 1} \sum_{j=t_i}^{i} \delta_\alpha(j)$ (by moment inequality)

Proof of theorem 2.1: bound of $\left\|\sum_{i=1}^{n} \mathcal{P}_{i-k} W_i^2\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$ (p.17)

Similarly for $k \geq 0$, $\left\|\mathcal{P}_{i-k} W_i^2\right\|_{\frac{\alpha}{2}} \leq 2 c_\alpha \Theta_\alpha \sqrt{i - t_i + 1} \sum_{j=t_i}^{i} \delta_\alpha(k + t_i - j)$

- Note that $\mathcal{P}_{i-k} W_i^2, i \in \mathbb{Z}$ form MDS, so $\left\|\sum_{i=1}^{n} \mathcal{P}_{i-k} W_i^2\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$
- $\leq c_\alpha \sum_{i=1}^{n} \left\|\mathcal{P}_{i-k} W_i^2\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$ (by Burkholder's inequality)
- $\leq c_\alpha \Theta_\alpha^{\frac{\alpha}{2}} \sum_{i=1}^{n} \left[\sqrt{i - t_i + 1} \sum_{j=t_i}^{i} \delta_\alpha(k + t_i - j)\right]^{\frac{\alpha}{2}}$ (by moment inequality)

Proof of theorem 2.1: $\|V_n - E(V_n)\|_{\frac{\alpha}{2}} = O\left(n^{\frac{3}{2}-\frac{3}{2p}+\frac{2}{\alpha}}\right)$ (p.18)

Consider first term from blocking $\sum_{k=0}^{\infty}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}}$, $\sum_{k=2b_m}^{\infty}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}}$

- $\leq O(1)\sum_{k=2b_m}^{\infty}\left\{\sum_{i=1}^{n}\left[\sqrt{i-t_i+1}\sum_{j=0}^{b_m}\delta_\alpha(k-j)\right]^{\frac{\alpha}{2}}\right\}^{\frac{2}{\alpha}}$ (by moment inequality in last slide)
  - The summation index can be change since $i-t_i \leq b_m$ and $k-b_m>0$
- $\leq O(1)\left[\sum_{i=1}^{n}(i-t_i+1)^{\frac{\alpha}{4}}\right]^{\frac{2}{\alpha}}\sum_{k=2b_m}^{\infty}\sum_{j=0}^{b_m}\delta_\alpha(k-j)$ (by independence of summation index)
  - The inequality sign in this step should be equal?
- $= O\left(n^{\frac{2}{\alpha}}b_m^{\frac{1}{2}}\right)o(b_m)$ (by $i-t_i \leq b_m$ and $\Delta_\alpha = \sum_{j=0}^{\infty}\delta_\alpha(j) < \infty$)
- $= o\left(n^{\frac{2}{\alpha}}b_m^{\frac{3}{2}}\right)$
- $= o\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)$ (since $b_m = O\left(m^{\frac{1}{p}}\right) = O\left(n^{1-\frac{1}{p}}\right)$)

Proof of theorem 2.1: $\|V_n - E(V_n)\|_{\frac{\alpha}{2}} = O\left(n^{\frac{3}{2}-\frac{3}{2p}+\frac{2}{\alpha}}\right)$ (p.18)

Consider second term from blocking, $\sum_{k=0}^{2b_m-1}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}}$

- $\leq O(1)\left[\sum_{i=1}^{n}(i-t_i+1)^{\frac{\alpha}{4}}\right]^{\frac{2}{\alpha}}\sum_{k=0}^{2b_m-1}\sum_{j=t_i}^{i}\delta_\alpha(k+t_i-j)$ (same steps as last slide)
- $=\left[\sum_{i=1}^{n}(i-t_i+1)^{\frac{\alpha}{4}}\right]^{\frac{2}{\alpha}}O(b_m)$ (use big O because summation index cannot be changed)
- $=O\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)$ (same steps as last slide)

Hence $\sum_{k=0}^{\infty}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}} = o\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)+O\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)$

- $=O\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)+O\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)$ (little o implies big O)
- $=O\left(n^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right)$

Proof of theorem 2.1: summary of techniques

Asymptotic approximation

- Approximate finite difference and sum by differential and integral
  - Be aware of the definition of Riemann sum (e.g. you may need to perform change of variable)
- Identify the dominating term
- Blocking: relate number of blocks m with sample size n

Handle multiple sum

- By blocking and bounding each block size
  - Terms in a double sum may becomes independent. See last two slides
- Break down power into product with maximum
  - E.g. $\sum_{t=1}^{n} t^p \leq \left(\max_{1\leq t\leq n} t\right) \sum_{t=1}^{n} t^{p-1}$

SECTION 3.2.2

Convergence rate, $\alpha > 4$

Proof of theorem 2.2: $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (p.20)

Notice that the condition changes from $\Delta_\alpha < \infty$ for some $\alpha \in (2,4]$ (T2.1) to $\alpha > 4$ (T2.2)

- But the convergence rate is same for $\alpha = 4$ (T2.1) and $\alpha > 4$ (T2.2)
  - This means stronger moment conditions cannot give faster convergence rate. See moment inequality (previous slide p.20)
- Theorem 2.2 gives a close form of asymptotic MSE (AMSE) though
  - $\|V_n - E(V_n)\| = \sqrt{E|V_n - E(V_n)|^2}$, which can give us MSE after some modifications
- Proof of T2.2 requires the use of lemma 1, which we shall prove later

Lemma 1: assume $X_i \in \mathcal{L}^\alpha$, $E(X_i = 0)$ and $\Delta_\alpha < \infty$ for $\alpha > 4$ (conditions of T2.2)

- Let $S_i = \sum_{j=1}^{i} X_j$ (the subscript should be j, probably typo in the paper)
- Then $\left\|\sum_{i=1}^{l}\left[E\left(S_i^2 \middle| \mathcal{F}_1\right) - E\left(S_i^2\right)\right]\right\| = o(l^2)$
- We also have $\lim_{l\to\infty} \frac{1}{l^4}\left\|\sum_{i=1}^{l}\left[S_i^2 - E\left(S_i^2\right)\right]\right\|^2 = \frac{1}{3}\sigma^4$

Proof of theorem 2.2: $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (p.18)

Let block sum of square $G_{h+1} = \sum_{i=a_h}^{a_{h+1}-1} W_i^2$ (target is $V_{a_{m+1}} = \sum_{h=1}^{m} G_{h+1}$)

- It differs from $\tilde{Y}_k$ in the sense that martingale approximation is not used
- By lemma 1, $\lim_{h\to\infty} \frac{1}{(a_{h+1}-a_h)^4} \left\| G_{h+1} - E\left(G_{h+1} \middle| \mathcal{F}_{a_h}\right) \right\|^2 = \frac{1}{3}\sigma^4$
- Since $G_{h+1} - E\left(G_{h+1} \middle| \mathcal{F}_{a_h}\right)$ is MDS wrt $\mathcal{F}_{a_{h+1}}$, we have $\left\| \sum_{h=1}^{m} \left[G_{h+1} - E\left(G_{h+1} \middle| \mathcal{F}_{a_h}\right) \right] \right\|^2$
- $= \sum_{h=1}^{m} E\left| G_{h+1} - E\left(G_{h+1} \middle| \mathcal{F}_{a_h}\right) \right|^2$ (MDS is uncorrelated)
- $\sim \frac{1}{3}\sigma^4 \sum_{h=1}^{m} (a_{h+1} - a_h)^4$ (by lemma 1)
- $\sim \frac{1}{3}\sigma^4 \sum_{h=1}^{m} c^4 p^4 h^{4p-4}$ (by considering the differential $a_h - a_{h-1} \sim cph^{p-1}$)
- $\sim \frac{p^4 c^4}{3(4p-3)} m^{4p-3} \sigma^4$ (by approximating sum $\Sigma_{x=1}^{m}$ with integral $\int_1^m dx$)
- $\sim \frac{p^4 c^{\frac{3}{p}}}{12p-9} n^{4-\frac{3}{p}} \sigma^4$ (by $n \sim cm^p \Rightarrow m \sim \left(\frac{n}{c}\right)^{\frac{1}{p}}$)

Proof of theorem 2.2: $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (p.18-19)

Similarly, $\left\|\sum_{h=1}^{m}\left[E\left(G_{h+1}\middle|\mathcal{F}_{a_h}\right) - E\left(G_{h+1}\middle|\mathcal{F}_{a_{h-1}}\right)\right]\right\|^2$

- $= \sum_{h=1}^{m} E\left|E\left(G_{h+1}\middle|\mathcal{F}_{a_h}\right) - E\left(G_{h+1}\middle|\mathcal{F}_{a_{h-1}}\right)\right|^2$ (MDS is uncorrelated)
- $\leq \sum_{h=1}^{m} E\left|E\left(G_{h+1}\middle|\mathcal{F}_{a_h}\right) - E(G_{h+1})\right|^2$ (property of conditional expectation? Not trivial)
- $= \sum_{h=1}^{m} o[(a_{h+1} - a_h)^4] = o\left(n^{4-\frac{3}{p}}\right)$ (by lemma 1 and result in last slide)

Now deal with $\Xi_m \overset{\text{def}}{=} \sum_{h=1}^{m}\left[E\left(G_{h+1}\middle|\mathcal{F}_{a_{h-1}}\right) - E(G_{h+1})\right]$

- The goal of Ξ_m is to connect everything for $\|\sum_{h=1}^{m}[G_{h+1} - E(G_{h+1})]\| = \left\|V_{a_m} - E\left(V_{a_m}\right)\right\|$
- Since $E\left(W_i^2\middle|\mathcal{F}_{a_{h-1}}\right) - E\left(W_i^2\right) = \sum_{k=0}^{\infty} \mathcal{P}_{i-k} E\left(W_i^2\middle|\mathcal{F}_{a_{h-1}}\right)$ for $a_h \leq i < a_{h+1}$
  - This follows from definition of projection and tower property
- We have $\|\Xi_m\| \leq \sum_{k=0}^{\infty} \left\|\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} \mathcal{P}_{i-k} E\left(W_i^2\middle|\mathcal{F}_{a_{h-1}}\right)\right\|$ (by Minkowski inequality)
- $= \sum_{k=0}^{\infty} \sqrt{\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} E\left|\mathcal{P}_{i-k} E\left(W_i^2\middle|\mathcal{F}_{a_{h-1}}\right)\right|^2}$ (by linearity of expectation and property of MDS)

Proof of theorem 2.2: $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (p.19)

Observe that $\mathcal{P}_{i-k} E\left(W_i^2 \middle| \mathcal{F}_{a_{h-1}}\right) = \begin{cases} 0, & i-k > a_{h-1} \\ \mathcal{P}_{i-k} W_i^2, & i-k \le a_{h-1} \end{cases}$ (by property of projection)

- Hence $\sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} E\left|\mathcal{P}_{i-k} E\left(W_i^2 \middle| \mathcal{F}_{a_{h-1}}\right)\right|^2}$
- $\le O(1) \sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} (i - t_i + 1) \left[\sum_{j=0}^{b_m} \delta_4(j)\right]^2}$ (mimic proof of $\sum_{k=2b_m}^{\infty} \|\sum_{i=1}^{n} \mathcal{P}_{i-k} W_i^2\|_{\frac{\alpha}{2}}$)
- $= O\left(n^{\frac{1}{2}} b_m^{\frac{1}{2}}\right) o(b_m) = o\left(n^{2-\frac{3}{2p}}\right)$ (mimic proof of $\sum_{k=2b_m}^{\infty} \|\sum_{i=1}^{n} \mathcal{P}_{i-k} W_i^2\|_{\frac{\alpha}{2}}$)

Proof of theorem 2.2: $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (p.19)

Now consider $\sum_{k=0}^{2b_m-1} \sqrt{\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} E\left|\mathcal{P}_{i-k} E\left(W_i^2 \middle| \mathcal{F}_{a_{h-1}}\right)\right|^2}$

- $\leq O(1) \sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} (i - t_i + 1)\left[\sum_{j=k+t_i-i}^{i} \delta_4(j)\right]^2 \mathbb{I}(i - k \leq a_{h-1})}$ (mimic proof of $\left\|\mathcal{P}_{i-k} W_i^2\right\|_{\frac{\alpha}{2}}$)
- $= O(1) \sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} \sum_{i=a_h}^{a_{h+1}-1} (i - t_i + 1)\Delta_4^2(a_h - a_{h-1})}$ (by definition of stability)
- $= O(1) \sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} (a_{h+1} - a_h)^2 \Delta_4^2(a_h - a_{h-1})}$ (by blocking)
- $= O(1) \sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} (a_{h+1} - a_h)^2 o(1)}$ (by $\Delta_4^2(a_h - a_{h-1}) \to 0$ as $a_h - a_{h-1} \to \infty$)
- $= O(1) \sum_{k=2b_m}^{\infty} \sqrt{\sum_{h=1}^{m} o(h^{2p-2})}$ (by $a_h - a_{h-1} = O(h^{p-1})$)
- $= o\left(b_m m^{p-\frac{1}{2}}\right) = o\left(n^{2-\frac{3}{2p}}\right)$ (by $b_m = O\left(n^{1-\frac{1}{p}}\right)$ and $m \sim \left(\frac{n}{c}\right)^{\frac{1}{p}}$)

Proof of theorem 2.2: $\lim_{n\to\infty} \frac{\|V_n - E(V_n)\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (p.19)

We have proved $\lim_{n\to\infty} \frac{\left\|\sum_{h=1}^{m}\left[G_{h+1} - E\left(G_{h+1}\middle|\mathcal{F}_{a_h}\right)\right]\right\|}{n^{2-\frac{3}{2p}}} = \frac{\sigma^2 p^2 c^{\frac{3}{2p}}}{\sqrt{12p-9}}$ (four slides ago)

- $\left\|\sum_{h=1}^{m}\left[G_{h+1} - E\left(G_{h+1}\middle|\mathcal{F}_{a_h}\right)\right]\right\| \asymp \|\sum_{h=1}^{m}[G_{h+1} - E(G_{h+1})]\| = \left\|V_{a_{m+1}} - E\left(V_{a_{m+1}}\right)\right\|$ (last three slides)
- It remains to show that $\left\|V_{a_{m+1}} - E\left(V_{a_{m+1}}\right)\right\| \asymp \|V_n - E(V_n)\|$
- Now consider the remainder term $\left\|\sum_{i=n}^{a_{m+1}-1}\left[W_i^2 - E\left(W_i^2\right)\right]\right\|$
- $\leq \sum_{i=n}^{a_{m+1}-1}\left\|W_i^2 - E\left(W_i^2\right)\right\|$ (by Minkowski inequality)
- $\leq \sum_{i=n}^{a_{m+1}-1}\left\|W_i^2\right\|$ (since W_i^2 is non negative)
- $= O(b_m^2)$ (recall the sum is a isosceles triangular shaped)
- $= O\left(n^{2-\frac{2}{p}}\right) \ll o\left(n^{2-\frac{3}{2p}}\right)$ (by $b_m = O\left(n^{1-\frac{1}{p}}\right)$ and $p > 1$)

Proof of lemma 1: $\left\|\sum_{i=1}^{l}\left[E\left(S_i^2\middle|\mathcal{F}_1\right)-E\left(S_i^2\right)\right]\right\| = o(l^2)$ (p.20)

Recall $S_i = \sum_{j=1}^{i} X_j$. Mimicking proof of $\left\|\mathcal{P}_{i-k}W_i^2\right\|_{\frac{\alpha}{2}}$, we have

- $\left\|\mathcal{P}_r S_i^2\right\| \le C\sqrt{i}\sum_{j=1}^{i}\delta_2(j-r)$ for $r \le 1$ where $C = 2c_2\Theta_2$
- Since $\sum_{i=1}^{l}\left[E\left(S_i^2\middle|\mathcal{F}_1\right)-E\left(S_i^2\right)\right] = \sum_{r=-\infty}^{1}\sum_{i=1}^{l}\mathcal{P}_r S_i^2$ (definition of projection), we have
- $\left\|\sum_{i=1}^{l}\left[E\left(S_i^2\middle|\mathcal{F}_1\right)-E\left(S_i^2\right)\right]\right\|^2 = \sum_{r=-\infty}^{1}\left\|\sum_{i=1}^{l}\mathcal{P}_r S_i^2\right\|^2$ (MDS is uncorrelated)
- $\le \sum_{r=-\infty}^{1}\left(\sum_{i=1}^{l}\left\|\mathcal{P}_r S_i^2\right\|\right)^2$ (by Minkowski inequality)
- $\le \sum_{r=-\infty}^{1}\left(C l^{\frac{3}{2}}\sum_{j=1}^{l}\delta_2(j-r)\right)^2$ (by inequality above and bounding $\sum_{j=1}^{i}\delta_2(j-r)$ with $l\delta_2(j-r)$)
  - Is it possible that $\sum_{j=1}^{i}\delta_2(j-r) > l \Rightarrow \sum_{i=1}^{l}\sum_{j=1}^{i}\delta_2(j-r) > l\sum_{j=1}^{l}\delta_2(j-r)$? Then this step do not hold
  - However the result is still correct by considering $\sum_{i=1}^{l}\sum_{j=1}^{i}\delta_2(j-r) \le \left[\sum_{j=1}^{l}\delta_2(j-r)\right]^2$
- $\le C^2 l^3 \Delta_2 \sum_{j=1}^{l}\sum_{r=-\infty}^{1}\delta_2(j-r)$ (by $\left[\sum_{j=1}^{l}\delta_2(j-r)\right]^2 \le \Delta_2\sum_{j=1}^{l}\delta_2(j-r)$)
- $= O(l^3)o(l) = o(l^4)$ (by $\Delta_\alpha < \infty$ for $\alpha > 4$)

Proof of lemma 1: $\lim_{l\to\infty} \frac{1}{l^4} \left\| \sum_{i=1}^{l} [S_i^2 - E(S_i^2)] \right\|^2 = \frac{1}{3}\sigma^4$ (p.21)

Let $A_l = \frac{1}{l^2} \sum_{i=1}^{l} S_i^2$. By invariance principle and continuous mapping theorem,

- $A_l \overset{d}{\to} \sigma^2 \int_0^1 W_t^2 dt$ (continuous mapping changes sum to integral, probably typo for IB)
- By theorem 1 in Wu (2007), $\|S_i\|_\alpha = O(\sqrt{i})$ (moment inequality)
- Hence $\|A_l\|_{\frac{\alpha}{2}} \le \frac{1}{l^2} \sum_{i=1}^{l} \|S_i^2\|_{\frac{\alpha}{2}}$ (by Minkowski inequality)
- $\le \frac{1}{l^2} \sum_{i=1}^{l} \|S_i\|_\alpha^2$ (by definition of norm, should be equal?)
- $= \frac{1}{l^2} \sum_{i=1}^{l} O(i) = O(1)$ (by moment inequality)
- Since $\frac{\alpha}{2} > 2$, $\{[A_l - E(A_l)]^2, l \ge 1\}$ is uniformly integrable (Chow and Teicher, 1988)
- Hence weak convergence of A_l implies the $\mathcal{L}^2$ moment convergence, which is
- $E\{[A_l - E(A_l)]^2\} \to \sigma^4 E\left\{\int_0^1 [W_t^2 - E(W_t^2)] dt\right\}^2 = \frac{1}{3}\sigma^4$ (by stochastic calculus, not trivial…)

Proof of lemma 1: $E\left\{\int_0^1 [W_t^2 - E(W_t^2)]dt\right\}^2 = \frac{1}{3}$

Let $f(t,w) = \frac{1}{6}w^4$. We have $\frac{\partial f}{\partial t} = 0$, $\frac{\partial f}{\partial w} = \frac{2}{3}w^3$ and $\frac{\partial^2 f}{\partial w^2} = 2w^2$. Note that $\mu = 0$ and $\sigma = 1$.

- $df(t,W_t) = \left[\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial W_t} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial W_t^2}\right]dt + \sigma\frac{\partial f}{\partial W_t}dW_t = W_t^2 dt + \frac{2}{3}W_t^3 dW_t$ (by Itô's lemma)
- Rearranging the terms, $\int_0^1 W_t^2 dt = \frac{1}{6}W_1^4 - \frac{2}{3}\int_0^1 W_t^3 dW_t = \frac{1}{2} + \sqrt{\frac{1}{3}}Z$ where $Z \sim N(0,1)$
  - $E\left(\int_0^1 W_t^2 dt\right) = \frac{1}{6}E(W_1^4) = \frac{3!!}{6} = \frac{1}{2}$ (by martingale property and $E(X^{2n}) = \sigma^{2n}(2n-1)!!$ if $X \sim N(0,\sigma^2)$. See this Q&A)
  - $E\left[\left(\int_0^1 W_t^2 dt\right)^2\right] = E\left(\int_0^1\int_0^1 W_t^2 W_s^2 dtds\right) = \int_0^1\int_0^1 E(W_t^2 W_s^2)dtds$ (by Fubini's theorem)
  - $= \int_0^1\int_0^s E[(W_s - W_t)^2 W_t^2 + 2(W_s - W_t)W_t^3 + W_t^4]dtds + \int_0^1\int_s^1 E[(W_t - W_s)^2 W_s^2 + 2(W_t - W_s)W_s^3 + W_s^4]dtds$
  - $= \int_0^1\int_0^s [(s-t)t + 3t^2]dtds + \int_0^1\int_s^1 [(t-s)s + 3s^2]dtds$ (by independent increment and $E(X^{2n+1}) = 0$ if $X \sim N(0,\sigma^2)$)
  - $= \frac{7}{24} + \frac{7}{24} = \frac{7}{12}$, so $Var\left(\int_0^1 W_t^2 dt\right) = \frac{7}{12} - \frac{1}{4} = \frac{1}{3}$
- On the other hand, $\int_0^1 E(W_t^2)dt = \int_0^1 tdt = \frac{1}{2}$ (since $W_t \sim N(0,t)$)
- Hence using representation, $E\left\{\int_0^1 [W_t^2 - E(W_t^2)]dt\right\}^2 = E\left(\frac{1}{3}Z^2\right) = \frac{1}{3}$

Proof of theorem 2.2 and lemma 1: summary of techniques

Stochastic calculus (my RMSC5102 note has a quick summary)

- Useful when we combine invariance principle and continuous mapping theorem
- Break down product of wiener process into sum of independent increment (see last slide)
- Vitali convergence theorem: a sequence of random variables converging in probability also converge in the mean if and only if they are uniformly integrable
  - A class of random variables bounded in $L^p, p > 1$ is uniformly integrable (see two slides ago)

Convergence rate, $\alpha = 2$

SECTION 3.2.3

Proof of theorem 2.3: $E(V_n - v_n\sigma^2) = O\left[n^{1+(1-q)\left(1-\frac{1}{p}\right)}\right]$ (p.20)

We do not have moment inequality when $\alpha = 2$ (i.e. in $\mathcal{L}^1$). Alternative strategy is needed.

- Let $j > 0$. To bound the autocovariance, we have $|\gamma(j)| = \left|E\left(X_0 X_j\right)\right|$
- $= \left|E\left[\sum_{i\in\mathbb{Z}}(\mathcal{P}_i X_0)\left(\mathcal{P}_i X_j\right)\right]\right|$ (projection decomposition, $X_j = \sum_{i\in\mathbb{Z}} \mathcal{P}_i X_j$)
- $\leq \sum_{i\in\mathbb{Z}} E\left|(\mathcal{P}_i X_0)\left(\mathcal{P}_i X_j\right)\right|$ (by Minkowski inequality)
- $\leq \sum_{i\in\mathbb{Z}}\|(\mathcal{P}_i X_0)\|\left\|\left(\mathcal{P}_i X_j\right)\right\|$ (by Cauchy–Schwarz inequality)
  - Orthogonality of projection gives a equal sign here but it does not affect the result
- $\leq \sum_{i=0}^{\infty} \omega(i)\omega(i+j)$ (by $\|\mathcal{P}_0 X_i\|_p \leq \omega_p(i)$ and $\omega_p(i) = 0$ if $i < 0$)

For $S_l = X_1 + \cdots + X_l$, since $\sum_{j=0}^{\infty} j^q \omega(j) < \infty$ for some $q \in (0,1]$ (by assumption)

- We have $\left|E\left(S_l^2\right) - l\sigma^2\right| = \left|l\gamma(0) + 2\sum_{j=1}^{l}(l-j)\gamma(j) - l\sum_{j\in\mathbb{Z}}\gamma(j)\right|$ (by representation of TAVC)
- $\leq 2\sum_{j=1}^{\infty}\min(j,l)\,|\gamma(j)|$ (by Minkowski inequality)
- $\leq 2\sum_{j=1}^{\infty}\min(j,l)^{1-q}\sum_{i=0}^{\infty}\min(j,l)^q\,\omega(i)\omega(i+j) = O(l^{1-q})$ (by $\sum_{j=0}^{\infty} j^q\omega(j) < \infty$)

Proof of theorem 2.3: $E(V_n - v_n\sigma^2) = o\left[n^{1+(1-q)\left(1-\frac{1}{p}\right)}\right]$ (p.20)

Combining the results, we have $|E(V_n - v_n\sigma^2)|$ (t_n should be v_n, probably typo)

- $\leq \sum_{i=1}^{n} |E(W_i) - (i - t_i + 1)\sigma^2|$ (by Minkowski inequality)
- $= \sum_{i=1}^{n} O[(i - t_i + 1)^{1-q}]$ (by $\left|E\left(S_l^2\right) - l\sigma^2\right| = O(l^{1-q})$)
- $= O\left(nb_m^{1-q}\right)$ (since b_m is the bound of batch size)
- $= o\left[n^{1+(1-q)\left(1-\frac{1}{p}\right)}\right]$ (by $b_m = o\left(n^{1-\frac{1}{p}}\right)$)

Proof of theorem 2.3: summary of techniques

Moment inequality is not available in $\mathcal{L}^1$

- Bound the target using projection decomposition and Wu's dependence measures
  - The polynomial decay rate of stability determines convergence rate

SECTION 3.3

Almost sure convergence

Almost sure convergence (p.9)

Glynn and Whitt (1992) argued that strongly consistent estimate of σ is needed

- For asymptotic validity of sequential confidence intervals
- Hence we need to consider the almost sure convergence behaviour for MCMC application

Corollary 3: Under the conditions in corollary 2,

- i.e. choose $a_k = \lfloor ck^p \rfloor$, $p = \frac{\frac{1}{2}+q}{q-\frac{1}{2}+\frac{2}{\alpha}}$ and assume $X_i \in \mathcal{L}^\alpha$, $E(X_i) = 0$ and $\Delta_\alpha < \infty$ for some $\alpha > 2$
- Or $X_i \in \mathcal{L}^2$, $E(X_i) = 0$ and $\sum_{j=0}^{\infty} j^q \omega(j) < \infty$ for some $q \in (0,1]$

- We have $\left\| \max_{n \le N} |V_n - E(V_n)| \right\|_{\frac{\alpha}{2}} = O(N^\tau \log N)$ where $\tau = \frac{3}{2} - \frac{3}{2p} + \frac{2}{\alpha}$
  - Note that τ is the convergence rate from theorem 2
- Also $V_N - E(V_N) = o_{a.s.}[N^\tau (\log N)^2]$ and $\frac{V_N}{v_N} - \sigma^2 = o_{a.s.}\left[N^{\frac{2}{\alpha}-\frac{1}{2}-\frac{1}{2p}}(\log N)^2\right]$
  - Possible to improve using strong invariance principle in Berkes, Liu and Wu (2014)?

Proof of corollary 3: $\left\|\max_{n\le N}|V_n - E(V_n)|\right\|_{\frac{\alpha}{2}} = O(N^{\tau}\log N)$ (p.21)

Choose $d \in \mathbb{N}$ such that $2^{d-1} < N \le 2^{d}$ (for the use of Borel-Cantelli lemma later?)

- For $1 \le a < b$, $\|V_a - V_b - E(V_b - V_a)\|_{\frac{\alpha}{2}} = \left\|\sum_{i=a+1}^{b}\left[W_i^2 - E\left(W_i^2\right)\right]\right\|_{\frac{\alpha}{2}}$
- $\le \sum_{k=0}^{\infty}\left\|\sum_{i=a+1}^{b}\mathcal{P}_{i-k}W_i^2\right\|_{\frac{\alpha}{2}}$ (by projection decomposition and Minkowski inequality)
- $= \sum_{k=0}^{\infty}\left[\sum_{i=a}^{b}(i - t_i + 1)^{\frac{\alpha}{4}}\right]^{\frac{2}{\alpha}} O\left(b^{1-\frac{1}{p}}\right)$ (mimic proof of $\sum_{k=0}^{2b_m-1}\|\sum_{i=1}^{n}\mathcal{P}_{i-k}W_i^2\|_{\frac{\alpha}{2}}$)
- $= O\left[(b-a)^{\frac{2}{\alpha}}b^{\frac{1}{2}\left(1-\frac{1}{p}\right)}\right]O\left(b^{1-\frac{1}{p}}\right) = O\left[(b-a)^{\frac{2}{\alpha}}b^{\frac{3}{2}\left(1-\frac{1}{p}\right)}\right]$
  - Note that the bound of batch/block size is $b_m = O\left(n^{1-\frac{1}{p}}\right)$ and bound of sample size is b here

Proof of corollary 3: $\left\|\max_{n\le N}|V_n - E(V_n)|\right\|_{\frac{\alpha}{2}} = O(N^\tau \log N)$ (p.21-22)

By proposition 1 in Wu (2007), $\left\|\max_{n\le 2^d}|V_n - E(V_n)|\right\|_{\frac{\alpha}{2}}$ (maximal inequality)

- $\le \sum_{r=0}^{d}\left[\sum_{l=1}^{2^{d-r}}\left\|V_{2^r l} - V_{2^r(l-1)} - E\left[V_{2^r l} - V_{2^r(l-1)}\right]\right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}\right]^{\frac{2}{\alpha}}$
- $= \sum_{r=0}^{d}\left\{\sum_{l=1}^{2^{d-r}} O\left[(2^r)^{\frac{2}{\alpha}}(2^r l)^{\frac{3}{2}\left(1-\frac{1}{p}\right)}\right]^{\frac{\alpha}{2}}\right\}^{\frac{2}{\alpha}}$ (by moment inequality proved in last slide)
- $= \sum_{r=0}^{d}\left\{O\left[(2^r)^{1+\frac{3\alpha}{4}\left(1-\frac{1}{p}\right)}\right]\sum_{l=1}^{2^{d-r}} O\left[l^{\frac{3\alpha}{4}\left(1-\frac{1}{p}\right)}\right]\right\}^{\frac{2}{\alpha}}$ (by independence of summation index)
- $\le \sum_{r=0}^{d}\left\{O\left[(2^d)^{1+\frac{3\alpha}{4}\left(1-\frac{1}{p}\right)}\right]\right\}^{\frac{2}{\alpha}}$ (since $l \le 2^{d-r}$)
- $= O(d+1)O\left[(2^d)^{\frac{2}{\alpha}+\frac{3}{2}-\frac{3}{2p}}\right]$
- $= O(N^\tau \log N)$ (since $\tau = \frac{3}{2} - \frac{3}{2p} + \frac{2}{\alpha}$ and $N \le 2^d \Rightarrow \log N \le d$)

Proof of corollary 3: $V_N - E(V_N) = o_{a.s.}[N^\tau (\log N)^2]$ (p.22)

Note that $\frac{\alpha}{2} > 1$. From $\left\| \max_{n \le N} |V_n - E(V_n)| \right\|_{\frac{\alpha}{2}} = O(N^\tau \log N)$ (proved in last two slides),

- We have $\frac{1}{(2^{d\tau} d^2)^{\frac{\alpha}{2}}} \sum_{d=1}^{\infty} \left\| \max_{n \le 2^d} |V_n - E(V_n)| \right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}} = \sum_{d=1}^{\infty} \frac{O[(d+1)2^{d\tau}]^{\frac{\alpha}{2}}}{(2^{d\tau} d^2)^{\frac{\alpha}{2}}} = \sum_{d=1}^{\infty} O\left(d^{-\frac{\alpha}{2}}\right) < \infty$
- Hence $V_N - E(V_N) = o_{a.s.}[N^\tau (\log N)^2]$ (by Borel-Cantelli lemma)
  - Borel-Cantelli lemma: for a sequence of events $E_1, \ldots$, if $\sum_{n=1}^{\infty} P(E_n) < \infty$, then $P\left(\limsup_{n\to\infty} E_n\right) = 0$
  - $P\left[\frac{\max_{n \le N} |V_n - E(V_n)|}{N^\tau (\log N)^2} > \epsilon\right] = E\left[\mathbb{I}\left(\frac{\max_{n \le N} |V_n - E(V_n)|}{N^\tau (\log N)^2} > \epsilon\right)\right]$ for all $\epsilon > 0$ (write probability as expectation of indicator)
  - $\le E\left|\frac{\max_{n \le N} |V_n - E(V_n)|}{N^\tau \left(\log N\right)^2}\right|$ (not sure if this is true)
  - $\le \frac{1}{N^{\frac{\tau\alpha}{2}} (\log N)^\alpha} \left\| \max_{n \le 2^d} |V_n - E(V_n)| \right\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}}$ (by property of norm and $\frac{\alpha}{2} > 1$)

Proof of corollary 3: $\frac{V_N}{v_N} - \sigma^2 = o_{a.s.}\left[N^{\frac{2}{\alpha}-\frac{1}{2}-\frac{1}{2p}}(\log N)^2\right]$(p.22)

Note that $V_N - E(V_N) = o_{a.s.}[N^\tau(\log N)^2]$ (proved in last slide)

- And $E(V_n - v_n\sigma^2) = O\left[n^{1+(1-q)\left(1-\frac{1}{p}\right)}\right]$ (theorem 2.3, probably typo in t_n)
- By choosing optimal rate $p = \frac{\frac{1}{2}+q}{q-\frac{1}{2}+\frac{2}{\alpha}}$, $E(V_N - v_N\sigma^2) = O(N^\tau) \ll o[N^\tau(\log N)^2]$
- We have $V_N - v_N\sigma^2 = o_{a.s.}\left[N^{\frac{3}{2}-\frac{3}{2p}+\frac{2}{\alpha}}(\log N)^2\right]$
- Finally recall $v_N = O\left(N^{2-\frac{1}{p}}\right)$ (proved in discussion of convergence rate when $\mu \neq 0$)
- Hence $\frac{V_N}{v_N} - \sigma^2 = o_{a.s.}\left[N^{\frac{2}{\alpha}-\frac{1}{2}-\frac{1}{2p}}(\log N)^2\right]$ (little o times big O is little o)

Proof of corollary 3: summary of techniques

Establish almost sure convergence

- Use maximal inequality
- Apply Borel-Cantelli lemma on maximal with expanding samples
  - Cantor’s diagonal argument?

Implementation issues

SECTION 4

Remaining question (p.9)

We can see that choice of block start a_k uniquely determines property of recursive TAVC

- The batch size l_i is determined by the selection rule (e.g. ΔSR, TSR, PSR)
- Under the simple choice $a_k = \lfloor ck^p \rfloor$, we have established the optimal choice of p
- It suffices to find the optimal choice of c in order to minimize AMSE

Assume $\Delta_\alpha < \infty$ for some $\alpha > 4$ and $\sum_{j=0}^{\infty} j^q \omega(j) < \infty$ for $q = 1$

  - Need $\alpha > 4$ for close form of AMSE (T2.2) and $q = 1$ for finite bias
- By corollary 2, optimal choice of $p = \frac{3}{2}$
- Choose data driven estimate of c by procedure in Bühlmann and Künsch (1999)

Close form of AMSE (p.10)

Since $\sum_{j=0}^{\infty} j\omega(j) < \infty$, $\sum_{i=1}^{\infty} i|\gamma(i)| < \infty$ (by bound of autocovariance in proof of T2.3)

- As $l \to \infty$, $E\left(S_l^2\right) - l\sigma^2 = -2\sum_{k=1}^{\infty} \min(k,l)\,\gamma(k) = -2\sum_{k=1}^{\infty} k\gamma(k) + o(1) = \theta + o(1)$
  - Keith (and I) usually denote $v_p \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} k^p\gamma(k)$ and $u_p \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} k^p|\gamma(k)|$
- Thus we have $E(V_n - v_n\sigma^2) = n\theta + o(n)$
- Now we decompose the AMSE in T2.2 into variance and bias^2,
- $\left\|\frac{V_n}{v_n} - \sigma^2\right\|_2^2 = \frac{1}{v_n^2}\left[\|V_n - E(V_n)\|_2^2 + |E(V_n) - v_n\sigma^2|^2\right]$
- $= \frac{(4p-2)^2}{c^{\frac{2}{p}}p^4} n^{\frac{2}{p}-4}\left[\frac{p^4c^{\frac{3}{p}}}{12p-9} n^{4-\frac{3}{p}}\sigma^4 + n^2\theta^2 + o(n^2)\right]$ (by $v_n \sim \frac{c^{\frac{1}{p}}p^2}{4p-2} n^{2-\frac{1}{p}}$, T2.2 and the result above)
- $= \frac{256}{81c^{\frac{4}{3}}} n^{-\frac{8}{3}}\left[\frac{9c^2}{16}\sigma^4 n^2 + \theta^2 n^2 + o(n^2)\right]$
- $= \left(\frac{16}{9}c^{\frac{2}{3}} + \frac{256}{81}c^{-\frac{4}{3}}\kappa^2\right)\sigma^4 n^{-\frac{2}{3}}$ where $\kappa = \frac{|\theta|}{\sigma^2}$

Optimal choice of c (p.10)

The optimal choice of c should minimize $\frac{16}{9}c^{\frac{2}{3}} + \frac{256}{81}c^{-\frac{4}{3}}\kappa^2$

- Illustration with SymPy
  - from sympy import symbols, diff, solve, simplify, Rational, init_printing
  - init_printing() # for printing Latex in console
  - c, kappa = symbols("c, kappa", real=True, positive=True) # kappa = v1/sigma^2
  - # Coefficent of Bias^2 and variance
  - b2 = Rational(256,81) *c**(-Rational(4,3)) *kappa**2
  - v = Rational(16,9) *c**(Rational(2,3)) # use Rational(p, q) if you want solution in fraction
  - mse = b2 +v
  - dMse = diff(mse,c)
  - minC = solve(dMse, c) # optimal c
  - # first root minimize after inspection
  - simplify(minC[0])
  - simplify(mse.subs(c, minC[0]))

```
   ...: minC = solve(dMse, c) # optimal c
   ...: # first root minimize after inspection
   ...: simplify(minC[0])
Out[1]:
```

$4\sqrt{2}\kappa/3$

```
In [2]: simplify(mse.subs(c, minC[0]))
Out[2]:
```

$16\sqrt[3]{12}\kappa^{2/3}/9$

Estimate optimal c (p.10-11)

By prime factorization, we can see that output of SymPy matches with

- Optimal AMSE of $\hat{\sigma}^2_{\Delta SR}(n) = \frac{2^{\frac{14}{3}}}{3^{\frac{5}{3}}}\theta^{\frac{2}{3}}\sigma^{\frac{8}{3}}n^{-\frac{2}{3}}$ with optimal $c = \frac{4\sqrt{2}|\theta|}{3\sigma^2} = \frac{4\sqrt{2}}{3}\kappa$
- Literature shows that optimal AMSE of $\hat{\sigma}^2_{obm}(n) = 2^{\frac{2}{3}}3^{\frac{1}{3}}\theta^{\frac{2}{3}}\sigma^{\frac{8}{3}}n^{-\frac{2}{3}}$
  - With batch size $l_n = \left\lfloor \lambda_* n^{\frac{1}{3}} \right\rfloor$ and optimal $\lambda_*^3 = \frac{3\theta^2}{2\sigma^4} \Rightarrow \kappa = \sqrt{\frac{2}{3}\lambda_*^3}$
    - Recall that we do not know the optimal functional of block start (same for batch size here)
  - This shows $AMSE\left[\hat{\sigma}^2_{\Delta SR}(n)\right] = 1.778 AMSE\left[\hat{\sigma}^2_{obm}(n)\right]$. Chan and Yau's (2017) TSR and PSR dominate it in MSE sense
- Theorem 4.1 in Bühlmann and Künsch (1999) gives $\frac{\hat{l}_n^3}{n} \sim \frac{1}{n\hat{b}^3} \sim \frac{3\theta^2}{2\sigma^4} = \lambda_*^3 \Rightarrow \lambda_* = \hat{l}_n n^{-\frac{1}{3}}$
  - They gives a procedure to estimate $\hat{l}_n$ via pilot simulation. Hence n is the sample size in pilot simulation here
  - Note that this is asymptotic. Can we have better pilot procedure for small sample?
- Using these relationship, we have $\hat{c} = \frac{8}{3\sqrt{3}}\lambda_*^{\frac{3}{2}} = \frac{8}{3\sqrt{3}}\hat{l}_n n^{-\frac{1}{3}}$

Bühlmann and Künsch's (1999) algorithm (p.10-11)

Let the Tukey-Hanning window $w_{TH}(x) = \frac{1}{2}[1 + \cos(\pi x)]\mathbb{I}(|x| \leq 1)$

- Let the splt-cosine window $w_{SC}(x) = \begin{cases} \frac{1}{2}\{1 + \cos[5(x - 0.8)\pi]\}, & 0.8 \leq |x| \leq 1 \\ 1, & |x| < 0.8 \\ 0, & |x| > 1 \end{cases}$
- 1) Compute $\hat{\gamma}(k) = \frac{1}{n}\sum_{i=1}^{n-|k|}(X_i - \bar{X}_n)\left(X_{i+|k|} - \bar{X}_n\right)$ for $k = 1 - n, \dots, n - 1$
- 2) Let $b_0 = \frac{1}{n}$. For $m = 1,2,3,4$, compute $b_m = n^{-\frac{1}{3}}\left[\frac{\sum_{k=1-n}^{n-1}\hat{\gamma}(k)^2}{6\sum_{k=1-n}^{n-1} w_{SC}\left(kb_{m-1}n^{\frac{4}{21}}\right)k^2\hat{\gamma}(k)^2}\right]^{\frac{1}{3}}$
- 3) Let $\hat{l}_n$ be the closest integer of $\hat{b}^{-1}$, where $\hat{b} = n^{-\frac{1}{3}}\left[\frac{2\left(\sum_{k=1-n}^{n-1} w_{TH}\left(kb_4 n^{\frac{4}{21}}\right)\hat{\gamma}(k)\right)^2}{3\left(\sum_{k=1-n}^{n-1} w_{SC}\left(kb_4 n^{\frac{4}{21}}\right)|k|\hat{\gamma}(k)\right)^2}\right]^{\frac{1}{3}}$