On the Validity of Long-Run Estimation Methods for Discrete-Event Systems

Peter J. Haas
IBM Research Division
Almaden Research Center
San Jose, CA
peterh@almaden.ibm.com

Peter W. Glynn
Dept. of Manag. Sci. & Engrg.
Stanford University
Stanford, CA
glynn@stanford.edu

MAMA 2002

Motivation: Time-Average Limits

- Often want to study performance measures of the form

$$r(f) = \lim_{t \to \infty} \frac{1}{t} \int_0^t f\big(X(u)\big)\, du$$

  - f is real-valued reward/cost function
  - $\{ X(t) : t \geq 0 \}$: underlying process of discrete-event stochastic system
  - Ex: long-run utilization, reliability, operating cost
  - System modeled as generalized semi-Markov process (GSMP)
  - Also want discrete-time averages over state-transition epochs

- For complex systems, must estimate $r(f)$ using simulation

  - Want point estimates and confidence intervals

Some Common Interval Estimation Methods

- Assume $\{ f\big(X(t)\big) : t \geq 0 \}$ obeys CLT with variance constant σ^2
- Methods based on consistent estimation of σ^2
  - Regenerative methods
  - Variable batch means
  - Spectral methods
- Cancellation methods
  - Require FCLT
  - Rest on limit theorem in which σ^2 "cancels out"
  - Ex: classical batch means
  - Ex: original standardized time series (STS) methods

When are Estimation Methods Guaranteed to Work?

- When are time-average limits well defined?

- When are estimation methods valid?

- Typical conditions on output process $\{ f(X(t)) : t \geq 0 \}$ hard to verify
  - Output process is regenerative
  - Output process is stationary (usually false!) and ϕ-mixing
  - Output process obeys an FCLT
  - Output process obeys a strong invariance principle

- We want conditions on model building blocks

Road Map for Remainder of Talk

- Review of GSMPs and FCLTs
- Cancellation methods
  - New limit theorems for GSMPs (SLLNs and FCLTS)
  - Ensures validity of cancellation methods
  - Weakest moment conditions possible (vs Haas 1999)
- Conditions for (weak) consistency of estimators of σ^2
  - Coupling approach for “quadratic form” estimators of σ^2
  - Exploit existing results for stationary processes
  - Obtain conditions for variable batch means & spectral estmators

GSMPs

- $\{ X(t) : t \geq 0 \}$ makes state transitions when events occur
  - Discrete set S of states and finite set E of events
  - $E(s)$ = set of active events in state $s \in S$
  - State-transition probabilities $p(s'; s, e^*)$

- Occurrence of events governed by clocks
  - Clock for $e \in E$ records time until scheduled occurrence of e
  - Set according to distribution function $F(x; e)$
  - Runs down to 0 at speed $r(s, e)$

- Defined in terms of Markov chain $\{ (S_n, C_n) : n \geq 0 \}$
  - S_n = state and C_n = clock-reading vector (after nth state transition)
  - Denote state space of chain by Σ

A Classical FCLT: Donsker's theorem

- $\{ X_n : n \geq 0 \}$ i.i.d. mean 0 random variables

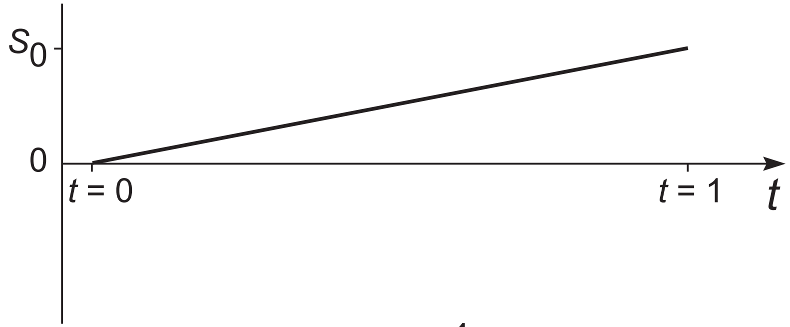
- $S_n = X_0 + X_1 + \cdots + X_n$

- $U_n(t) = \frac{1}{\sqrt{n}} \int_0^{nt} X_{\lfloor u \rfloor} \, du$

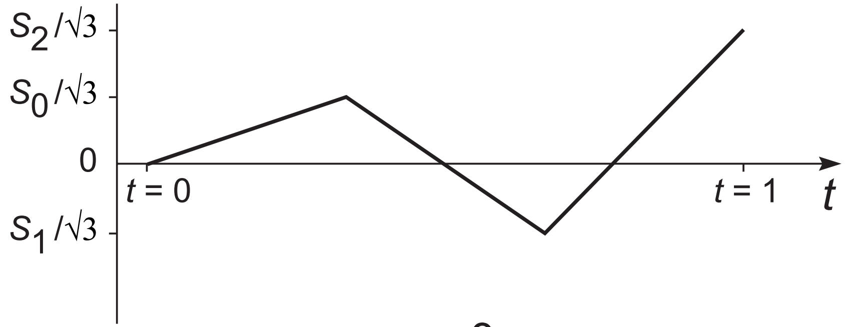
- *Theorem*: if $\sigma^2 = \mathrm{Var}\,[X_0] < \infty$, then $U_n \Rightarrow \sigma W$
  - the symbol $\Rightarrow$ denotes weak convergence on $C[0,1]$
  - W is a standard Brownian motion on $[0,1]$

- By continuous mapping theorem: $U_n(1) = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j \Rightarrow \sigma W(1)$

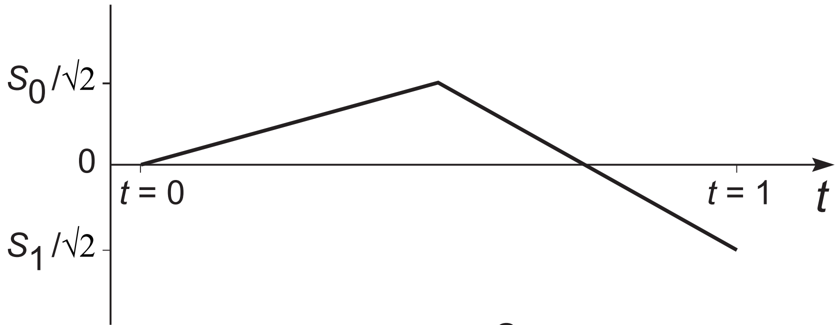
The Function $U_n(t)$



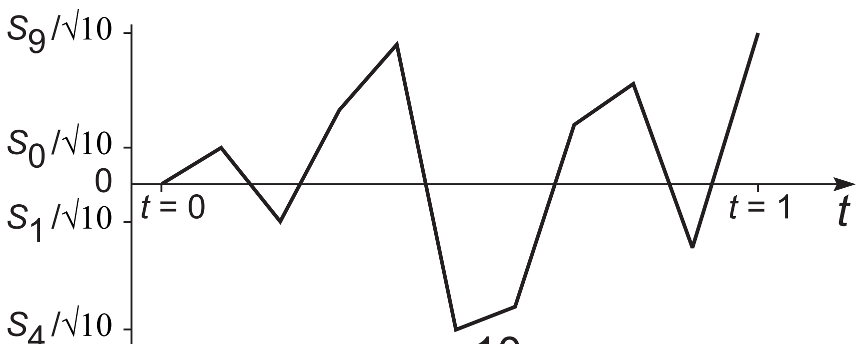
$n = 1$



$n = 3$

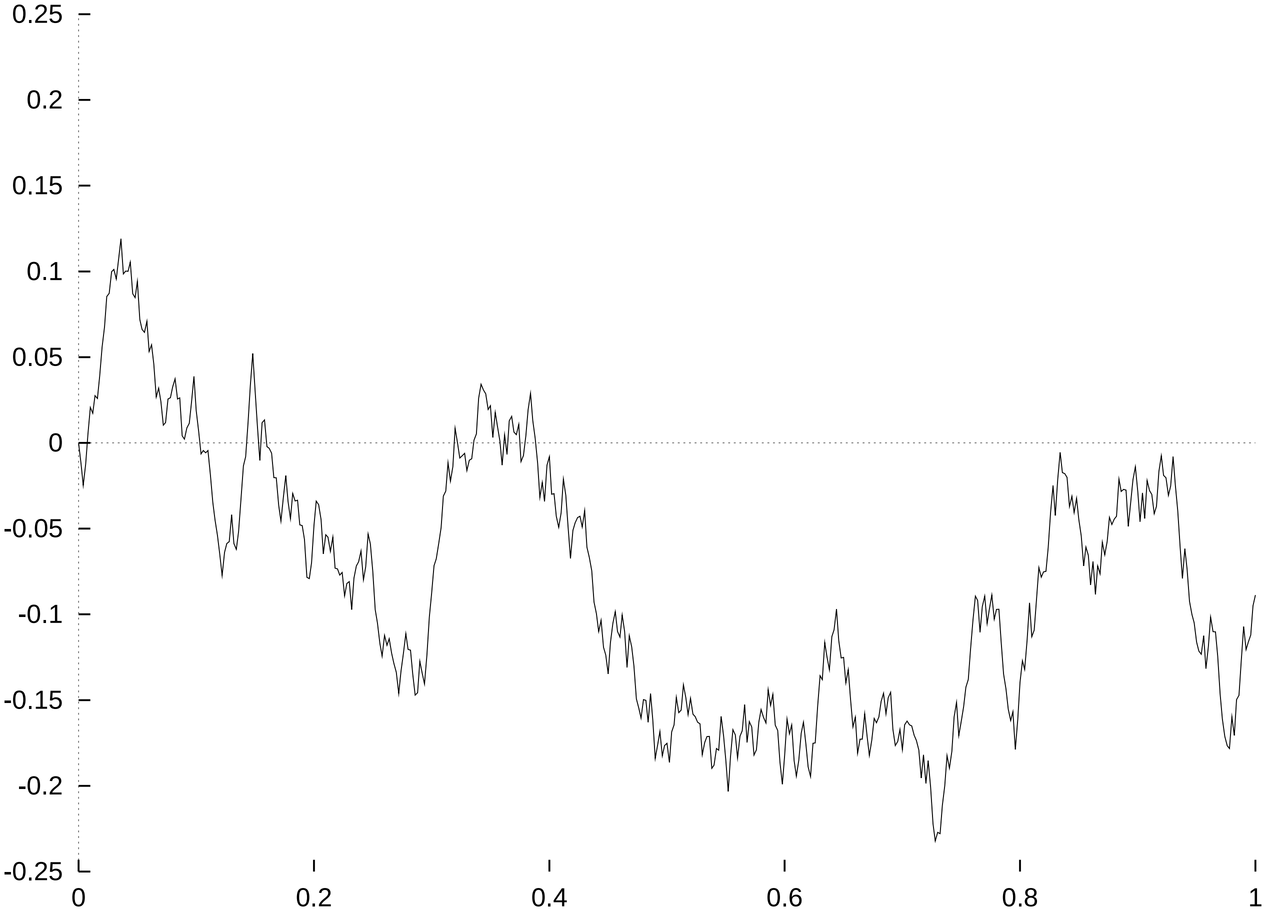


$n = 2$



$n = 10$

The Function $U_n(t)$ with $n = 500$



Consequences of FCLT (Glynn and Iglehart 1990)

- Estimate $r = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(X_j)$ by $\hat{r}_n = \frac{1}{n}\sum_{j=0}^{n-1} f(X_j)$

- Suppose that $U_n \Rightarrow \sigma W$, where $U_n(t) = \frac{1}{\sqrt{n}}\int_0^{nt}\big(f(X_{\lfloor u \rfloor}) - r\big)\, du$

- Suppose $\xi : C[0,1] \mapsto \Re$ is "continuous" and, for $x \in C[0,1]$ and $a \in \Re$,

$$\xi(ax) = a\xi(x) \text{ and } \xi(x - ae) = \xi(x), \text{ where } e(t) = t$$

- Set $\xi_n = \xi(\overline{Y}_n)$, where $\overline{Y}_n(t) = \frac{1}{n}\int_0^{nt} f(X_{\lfloor u \rfloor})\, du$

- Can show: $\dfrac{\hat{r}_n - r}{\xi_n} = \dfrac{\sqrt{n}(\hat{r}_n - r)}{\sqrt{n}\,\xi_n} = \dfrac{U_n(1)}{\xi(U_n)} \Rightarrow \dfrac{\sigma W(1)}{\sigma\xi(W)} = \dfrac{W(1)}{\xi(W)}$

- Asymptotic $100p\%$ confidence interval: $[\hat{r}_n - \xi_n z_p,\ \hat{r}_n + \xi_n z_p]$

Limit Theorems for GSMPs

- Assumption PD(q)
  - GSMP has finite state space S and positive speeds
  - GSMP is irreducible
  - Each $F(\,\cdot\,;e)$ has finite qth moment and density positive on $[0,\bar{x}]$

- Assumption PDE: $\int_0^\infty v^x \, dF(x;e) < \infty$ for some $v > 1$

- $\mathcal{H}_u$: Set of functions $\tilde{h}\colon \Sigma \mapsto \Re$ such that

$$|\tilde{h}(s,c)| \le a + b\left(t^*(s,c)\right)^u$$

where $t^*(s,c)$ is holding time starting from s with clock-reading vector c

Strong Law of Large Numbers

- *Theorem*: If Assumption PD(1) holds, then for any function $f : S \mapsto \Re$ there exists finite $r(f)$—indep. of initial dist'n—such that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f\big(X(u)\big)\,du = r(f) \text{ a.s.}$$

- *Theorem*: If Assumption PD$(u \vee 1)$ holds $(u \geq 0)$, then for any $\tilde{f} \in \mathcal{H}_u$ there exists finite $\tilde{r}(\tilde{f})$—indep. of initial dist'n—such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \tilde{f}(S_j, C_j) = \tilde{r}(\tilde{f}) \text{ a.s.}$$

FCLT

- Assume time-average limits $r(f)$, $\tilde{r}(\tilde{f})$ exist and set
  - $U_\nu(f)(t) = \frac{1}{\sqrt{\nu}} \int_0^{\nu t} \Big(f\big(X(u)\big) - r(f) \Big)\, du$
  - $\tilde{U}_n(\tilde{f})(t) = \frac{1}{\sqrt{n}} \int_0^{nt} \big(\tilde{f}(S_{\lfloor u \rfloor}, C_{\lfloor u \rfloor}) - \tilde{r}(\tilde{f}) \big)\, du$
- *Theorem*: If Assumption PD(2) holds, then for any function $f : S \mapsto \Re$ there exists $\sigma(f)$ such that $U_\nu(f) \Rightarrow \sigma(f)W$ for any initial dist'n
- *Theorem*: If Assumption PD$\big(2(u \vee 1)\big)$ holds, then for any $\tilde{f} \in \mathcal{H}_u$ there exists $\tilde{\sigma}(\tilde{f})$ such that $\tilde{U}_n(\tilde{f}) \Rightarrow \tilde{\sigma}(\tilde{f})W$ for any initial dist'n

Proof of SLLN and FCLT

- (Haas 1999): Under Assumption PD(q), with $h_q(s,c) = 1 + \max_i c_i^q$,

$$\sup_{(s,c)\in\Sigma-H} E_{(s,c)}[h_q(S_m, C_m) - h_q(S_0, C_0)] \le -\beta_q h_{q-1}(s,c)$$

  for some $m \ge 1$, $\beta_q > 0$, and compact set H

- Thus $\{\,(S_n, C_n) : n \ge 0\,\}$ is positive Harris recurrent
  $\Rightarrow$ wide-sense regenerative (e.g., Meyn and Tweedie 1993)

- Show: cycle length η_1 and $\sum_{j=0}^{\eta_1 - 1} \tilde{f}(S_j, C_j)$ have finite qth moments
  - Show that $E_{(s,c)}[T_B^q] \le \beta_q h_q(s,c)$ for $(s,c) \in \Sigma$ (induction on q)
  - Show that $E_\mu[\eta_1^q] < \infty$ (treat as random sum of T_B's)
  - Show that cycle sum has finite qth moment (use bound in Gut 1988)

Consistent Variance Estimation

- Let Assumption PDE hold
- Let $\tilde{f}$ be polynomially dominated: $\tilde{f} \in \mathcal{H}_u$ for some u
- For $\bar{r}(n; \tilde{f}) = \frac{1}{n} \sum_{j=0}^{n-1} \tilde{f}(S_j, C_j)$, previous results show:
  - $\lim_{n\to\infty} \bar{r}(n; \tilde{f}) = \tilde{r}(\tilde{f})$ a.s. for some finite $\tilde{r}(\tilde{f})$
  - $\left(\tilde{\sigma}^2(\tilde{f})/n\right)^{-1/2}\left(\bar{r}(n; \tilde{f}) - \tilde{r}(\tilde{f})\right) \Rightarrow N(0,1)$ for some $\tilde{\sigma}^2(\tilde{f})$
- If $V_n \Rightarrow \tilde{\sigma}^2(\tilde{f})$, then $\left(V_n/n\right)^{-1/2}\left(\bar{r}(n; \tilde{f}) - \tilde{r}(\tilde{f})\right) \Rightarrow N(0,1)$
- Asymptotic $100p\%$ confidence interval:

$$[\bar{r}(n; \tilde{f}) - z_p (V_n/n)^{1/2}, \bar{r}(n; \tilde{f}) + z_p (V_n/n)^{1/2}]$$

Batch-Means Variance Estimator

- Given process $\{ Z_n : n \geq 0 \}$ with variance constant $\tilde{\sigma}^2$
- Break $n = mb$ observations into b batches of length m
  - For consistency, need $b = b(n)$ and $m = m(n)$
- $\overline{X}_n(j)$ is jth batch mean: $\overline{X}_n(j) = \frac{1}{m} \sum_{i=(j-1)m}^{jm-1} Z_i$
- $\overline{X}_n =$ average of batch means
- Estimator of $\tilde{\sigma}^2$:

$$V_n^{(B)} = \frac{m}{b-1} \sum_{j=1}^{b} \left(\overline{X}_n(j) - \overline{X}_n \right)^2$$

Spectral Estimators

- Estimator of $\tilde{\sigma}^2$:

$$V_n^{(S)} = \sum_{h=-(m-1)}^{m-1} \lambda(h/m)\hat{R}_h$$

- $\hat{R}_h$ is estimated lag-h autocovariance:

$$\hat{R}_h = \tfrac{1}{n-|h|}\textstyle\sum_{i=0}^{n-|h|-1}(Z_i - \overline{Z}_n)(Z_{i+|h|} - \overline{Z}_n)$$

- λ is lag window (window length $= 2m$):
  - finite and continuous on $[-1,1]$ with $\lambda(x) = \lambda(-x)$ and $\lambda(0) = 1$
  - $\lambda(x) = 0$ for $x \notin [-1,1]$
  - $\lim_{x\to 0}\big(1-\lambda(x)\big)/|x|^q = \alpha$ for some $q, \alpha \in (0,\infty)$

Quadratic-Form Estimators

- A QF estimator is of the form

$$V_n = V_n(\tilde{f}) = \sum_{i=0}^{n} \sum_{j=0}^{n} Z_i Z_j q_{i,j}^{(n)}$$

where each $q_{i,j}^{(n)}$ is finite with $q_{i,j}^{(n)} = q_{j,i}^{(n)}$

- Localized QF estimators (includes batch-means, spectral estimators)

$$|q_{i,j}^{(n)}| \leq \begin{cases} a_1/n & \text{if } |i-j| \leq m(n) \\ a_2(n)/n & \text{if } |i-j| > m(n) \end{cases}$$

where $a_1 \in [0, \infty)$, $a_2(n) \to 0$, and $m(n)/n \to 0$

Aperiodic GSMPs

- d-cycle: S can be partitioned into disjoint sets $S_1, \ldots, S_d$ such that $s' \in S_{i+1}$ whenever $s \in S_i$ and $p(s'; s, e^*) > 0$ for some e^*

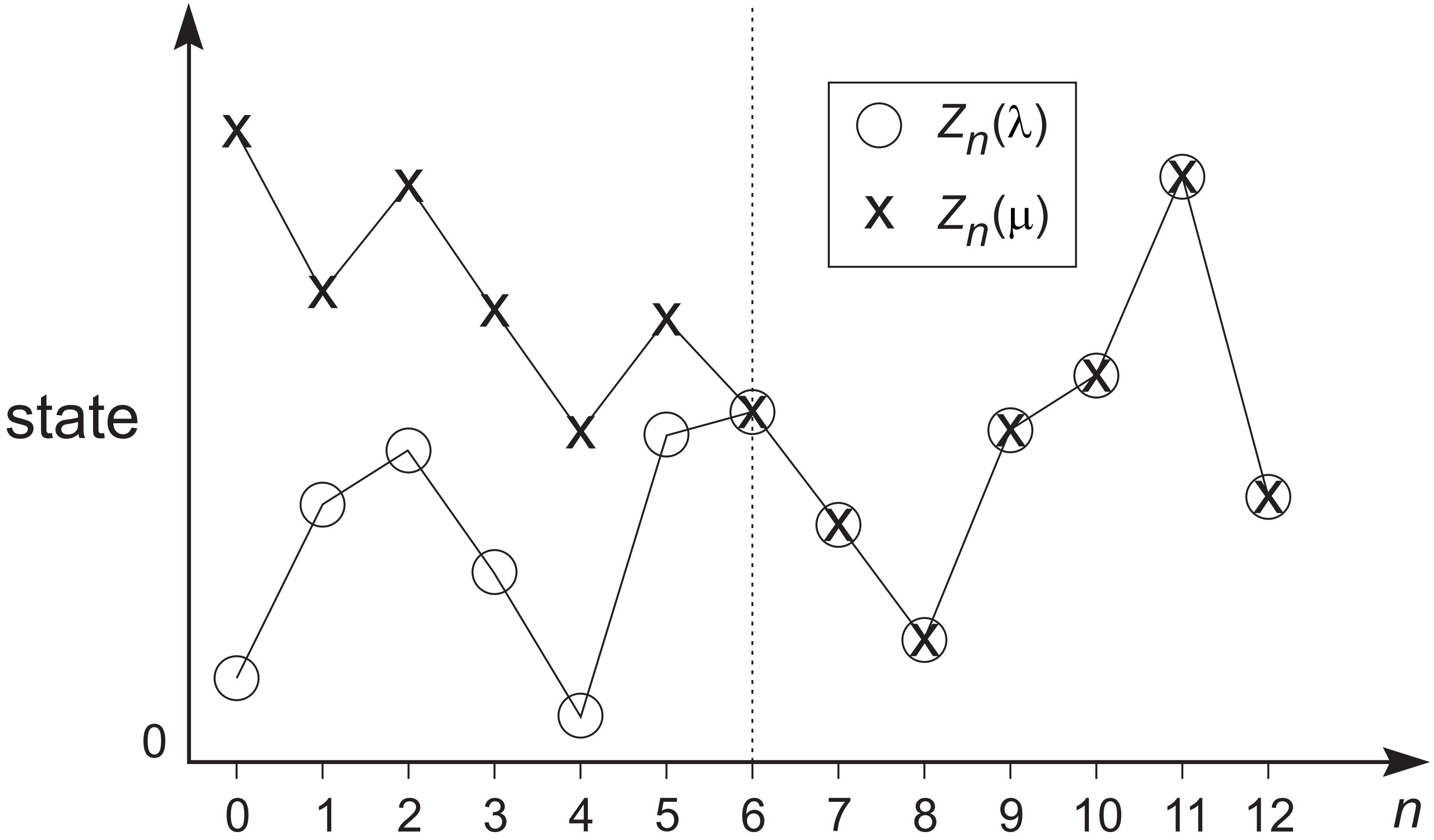
- Period of GSMP: largest d for which d-cycle exists

- GSMP with period 1 is aperiodic

- *Lemma*: If Assumption PDE holds for an aperiodic GSMP, then underlying chain $\{ (S_n, C_n) : n \geq 0 \}$ is aperiodic, hence Harris ergodic

- Two consequences of Harris ergodicity
  - There exists invariant distribution π: $P_\pi \{ (S_1, C_1) \in A \} = \pi(A)$
  - Chain admits coupling

Coupling



From Stationary to Nonstationary Consistency Results

- *Theorem*: Given GSMP + polynomially dominated function $\tilde{f}$, suppose
  - GSMP is aperiodic
  - Assumption PDE holds, so that
    - there exists an invariant distribution π for $\{\,(S_n, C_n)\colon n \geq 0\,\}$
    - $\{\,\tilde{f}(S_n, C_n)\colon n \geq 0\,\}$ obeys a CLT with a variance constant $\tilde{\sigma}^2(\tilde{f})$

  If localized QF estimator $V_n(\tilde{f})$ satisfies $V_n(\tilde{f}) \Rightarrow \tilde{\sigma}^2(\tilde{f})$ when initial distribution is π, then $V_n(\tilde{f}) \Rightarrow \tilde{\sigma}^2(\tilde{f})$ for any initial distribution

- Proof: Couple stationary and nonstationary version of underlying chain and show that $|V_n - V_n^{\text{stationary}}| \to 0$ (follows from localization property)

Validity of Variable Batch Means and Spectral Methods

- *Theorem*: Define $V_n^{(B)}$, $V_n^{(S)}$ as before with $Z_n = \tilde{f}(S_n, C_n)$. Suppose
  - GSMP is aperiodic and $\tilde{f}$ is polynomially dominated
  - Assumption PDE holds, so that $\{\, \tilde{f}(S_n, C_n) : n \geq 0 \,\}$ obeys a CLT with a variance constant $\tilde{\sigma}^2(\tilde{f})$

  Then

  - $V_n^{(B)} \Rightarrow \tilde{\sigma}^2(\tilde{f})$ if $b = b(n) \to \infty$ and $m = m(n) \to \infty$
  - $V_n^{(S)} \Rightarrow \tilde{\sigma}^2(\tilde{f})$ if $m = m(n) \to \infty$ with $m(n) = o(n^{1/2})$

- Proof: Use results on consistent estimation in stationary regime (Chien et al. 1997, Anderson 1971) + mixing and moment properties of Harris ergodic chains, then apply coupling result

Consistent Estimation in Continuous Time

- Can extend previous results to quantities of the form

$$g\big(\tilde{r}(\tilde{f}_1), \tilde{r}(\tilde{f}_2), \ldots, \tilde{r}(\tilde{f}_l)\big)$$

where g is nonlinear and differentiable (use Cramér-Wold + delta method)

- But can show that

$$r(f) = \lim_{t\to\infty} \frac{1}{t} \int_0^t f\big(X(u)\big)\, du = \frac{\tilde{r}(ft^*)}{\tilde{r}(t^*)}$$

where t^* is holding-time function and $(ft^*)(s,c) = f(s)t^*(s,c)$

Relation to Results of Damerdji et al.

- Stronger assumption: output process obeys strong invariance principle
- Stronger conclusion: $V_n \to \tilde{\sigma}^2$ with probability 1 (strong consistency)
  - Needed for showing validity of sequential stopping rules
- Hard to establish strong invariance principle
  - E.g., for variable batch means with $b(n) = O(n^{2/3})$
  - Our results establish weak consistency for this method

Summary

- Sufficient building block conditions on GSMPs for validity of a variety of interval estimation methods for time-average limits
- Cancellation methods
  - New SLLNs and FCLTs in discrete and continuous time
- Consistent estimation of the variance
  - General method for proving consistency via coupling
  - Established validity of variable batch means, spectral methods
  - Sufficient conditions for Harris ergodicity of underlying chain

Future Work

- Future work: Cancellation methods
  - Weakening of conditions based on specific model structure
  - Understanding pathological behavior
  - Heavy tails (applicability of current methods vs need for new methods)
- Future work: Consistent estimation methods
  - Refine conditions for specific estimation methods
  - Establish validity of STS variants, Cramér–Von Mises, etc.
  - Complicated sequential procedures
- GSMP extensions: PRI preemption, relocatable clocks, hybrid models
- Automate condition checking in tools (e.g., irreducibility)