OUTPUT ANALYSIS : CHAPTER 8

Motivation : when we program a simulation of a system, often we have in mind to estimate given performance quantities under different scenarios, for example in order to test various health policies (vaccination, sanitation, etc). Other goals of a simulation include finding optimal control values, as in simulated annealing and random search methods.

The outcome of the simulation will be a vector of performance measures. Consider one such performance and call it X.

The simulation model is of the form of a stochastic process $\{\xi_n, n=0,1,..\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The performance is usually a functional X of the trajectory, so that X is a well defined rv on $(\Omega, \mathcal{F}, \mathbb{P})$. The goal of the simulation is to estimate:

$$\theta = \mathbb{E}(X).$$

The rv X may be very complex, and sometimes we can only hope to generate approximations of its distribution (as in the case of random search or MCMC, or stationary problems).

Example:

$$X = \lim_{n\to\infty} f(\xi_n)$$

may be a stationary revenue (longterm), in which case if $\{\xi_n\}$ is ergodic, then

$$\mathbb{E}(X) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(\xi_k).$$

Main questions:

- what statistical estimators are suitable? (usually running or sample averages)
- how can we estimate the error (or the "precision") in the approximation?
- when should we stop an infinite-horizon simulation?

Def: A sequence of random variables $X_1, X_2, ..$ with common mean $\theta = \mathbb{E}(X_i)$, $i \in \mathbb{N}$ is said to satisfy:

- the Weak Law of Large Numbers if the sample average converges to θ in probability, that is, $\forall \epsilon > 0$
$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n}\sum_{k=1}^{n} X_k - \theta \right| > \epsilon \right) = 0,$$
- the Strong Law of Large Numbers if the sample average converges a.s. to θ, that is:
$$\mathbb{P}\left(\omega : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k(\omega) = \theta\right) = 1.$$

(often referred to as SLLN).

WLLN : the sample avg is "likely" to be near μ, but $|\bar{X}_n - \theta| > \epsilon$ is an event that can happen infinitely often although very infrequently. In contrast, this is impossible if SLLN holds.

Thm: If $\{X_i\}_{i\in\mathbb{N}}$ are iid, $\theta = \mathbb{E}(X_i)$ and $\sigma^2 = Var(X_i) < \infty$, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} \theta.$$

Thm: Let $\{X_n\}$ be an ergodic Markov chain. Then it satisfies the SLLN for θ the stationary avg, whenever $\sup_n Var(X_n) < \infty$.

Central Limit Theorem: If $\{X_n\}$ is a sequence of iid rv's with $\theta = \mathbb{E}X_n$, $\sigma^2 = Var(X_n) < \infty$, then:

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{satisfies}$$

$$\sqrt{n}\,\frac{(\bar{X}_n - \theta)}{\sigma} \Rightarrow \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

Proposition: Let $\{X_n\}$ iid rv's with $\theta = \mathbb{E}X_n$, $\sigma^2 = Var(X_n) < \infty$, ~~th~~ and define the sample variance as:

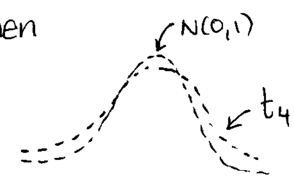
$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} (X_n - \bar{X}_n)^2.$$

Then $\mathbb{E}(S_n^2) = \sigma^2$ (is an unbiased estimator of σ^2).

(Slutski)

Thm: Let $\{X_n\}$ iid rv's with $\theta = \mathbb{E}X_n$, $\sigma^2 = Var(X_n) < \infty$, ~~then~~ if $X_n \sim N(\theta, \sigma^2)$ are normal rv's, then

$$\frac{\sqrt{n}\,(\bar{X}_n - \theta)}{\sqrt{S_n^2}} \sim t_{(n-1)}$$



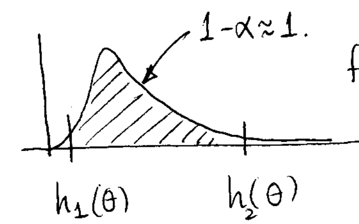
Student-t distribution with n-1 degrees of freedom, $t_n \Rightarrow N(0,1)$.

Thm: Let $\{X_n\}$ iid $\sim$ Bernoulli(θ), $\sigma^2 = \theta(1-\theta)$, then as $n \to \infty$

$$\frac{\sqrt{n}\,(\bar{X}_n - \theta)}{\sqrt{\bar{X}_n(1 - X_n)}} \Rightarrow N(0,1).$$

Confidence Intervals

Let $\hat{\theta}_n$ be a random variable (the estimator of θ, that may or may not have the form of a sample average), and let $F_n(\cdot)$ be its distribution (that depends usually on θ).



for given θ, $h_1(\theta)$, $h_2(\theta)$ are values such that $(1-\alpha)\%$ of the distribution is inside.

Def: a confidence interval at significance level α for θ is a random interval $I(\hat{\theta}_n)$ satisfying: $\mathbb{P}(\theta \in I(\hat{\theta}_n)) = 1 - \alpha$.

To build the interval, one may proceed by letting $h_1(\theta)$ and $h_2(\theta)$ be defined by the quantiles:

$$F_n(h_1(\theta)) = \frac{\alpha}{2}, \quad F_n(h_2(\theta)) = 1 - \frac{\alpha}{2}$$

Using this (implicit function) definition, if h_1, h_2 have well defined inverses as functions of θ, then:

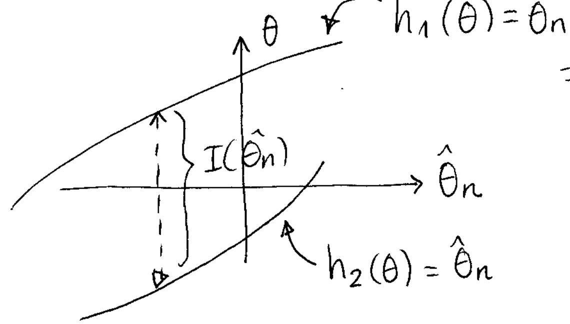
$$I(\hat{\theta}_n) \equiv [\,h_2^{-1}(\hat{\theta}_n),\, h_1^{-1}(\hat{\theta}_n)\,]$$

will be a CI at significance level α. To show this, use:

$$\mathbb{P}(\theta \in I(\hat{\theta}_n)) = \mathbb{P}(h_2^{-1}(\hat{\theta}_n) \le \theta,\, h_1^{-1}(\hat{\theta}_n) \geq \theta) =$$

$$= \mathbb{P}(\hat{\theta}_n \le h_2(\theta),\, \hat{\theta}_n \geq h_1(\theta)) =$$

$$\mathbb{P}(h_1(\theta) \le \hat{\theta}_n \le h_2(\theta)) = F_n(h_2(\theta)) - F_n(h_1(\theta)) = 1 - \alpha.$$



For the various possible observed values of the estimator $\hat{\theta}_n$, the plot shows the corresponding CI $I(\hat{\theta}_n)$.

Example: If $\{X_n\} \sim$ iid $N(\theta, \sigma^2)$, the corresponding confidence interval @ level α is

$$\hat{\theta}_n \pm t_{n-1,1-\alpha/2} \sqrt{\frac{S_n^2}{n}},$$

where $t_{n-1,1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the Student-t distribution with $n-1$ degrees of freedom.

In most cases, however, the exact distribution $F_n(\cdot)$ of the estimator $\hat{\theta}_n$ is not known analytically, which is the reason why CLT's are very important: a CLT provides an approximate CI for sample means.

Def: Given an approximate CI (based on an asymptotic result or other means) of the form $I(\hat{\theta}_n)$ at significance level α, let

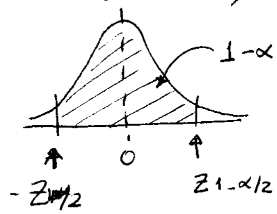
$$p_{n,\alpha} = \mathbb{P}(\theta \in I(\hat{\theta}_n)).$$

We call this probability the coverage of the CI.

Ideally, we want $p_{n,\alpha} \approx 1-\alpha$, but in many cases this requires prohibitive large sample sizes n. The actual coverage is rarely easy to calculate: only repeating the simulation many times can one have an idea, using specific examples where θ is known.

Suppose that a CLT holds for $\hat{\theta}_n$, so that as $n \to \infty$,

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma} \Rightarrow N(0,1)$$



then for large n, $\hat{\theta}_n$ has an approximate normal distribution $N(\theta, \sigma^2/n)$. Then

$$\mathbb{P}\left(\theta \in \left[\hat{\theta}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\theta}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]\right) \approx 1-\alpha.$$

~~Two~~ Three scenarios are possible:

- if n is fixed, then one proceeds with the simulation and estimates the CI that provides an error estimation,
- if n is not fixed, but a desired precision (tolerance) $\epsilon > 0$ is required, then it is necessary to estimate the required sample size n such that the CI has half-width ϵ.
- if a relative error is specified, namely $\epsilon = \delta\theta$, $\delta \in (0,1)$.

Remark: in many situations σ^2 is not known, so that it is necessary to establish a CLT replacing it by S_n^2, its estimator.

Fixed-width Intervals (fixed tolerance)

When $\epsilon > 0$ is given, the problem is that the simulation length becomes a random stopping time. p. 11-12 (A-S)

Let $\epsilon > 0$ and define:

$$\tau = \min\left(n \geq 2 : z_{1-\alpha/2} \frac{\sqrt{\widehat{Var}(\hat{\theta}_n)}}{n} \leq \epsilon\right), \qquad (1)$$

where $\widehat{Var}(\hat{\theta}_n)$ is an estimator of σ^2, assumed to be a consistent estimator, that is,

$$\lim_{n\to\infty} \mathbb{E}(\widehat{Var}(\hat{\theta}_n)) = \sigma^2.$$

In the examples where $\{X_n\}$ are iid and $\hat{\theta}_n$ is the sample mean, S_n^2 is unbiased (thus, consistent).

Then there is no guarantee (because τ is random) that

$$\frac{\sqrt{\tau}(\hat{\theta}_\tau - \theta)}{\widehat{Var}(\hat{\theta}_\tau)}$$ is anywhere close to $N(0,1)$ in distribution!

Thm: [Chow & Robbins]. Let $\{X_n\}$ be a sequence of iid rv's with $\mathbb{E}(X_n) = \theta$, $Var(X_n) = \sigma^2 < \infty$, and $\epsilon > 0$ a pre-specified error tolerance. Define the random stopping time τ as:

$$(2) \quad \tau = \min\left(n \geq 2 : \delta(n,\alpha) \leq \sqrt{\frac{n}{n-1}\epsilon^2 - \frac{t^2_{n-1,1-\alpha/2}}{n(n-1)}}\right),$$

where $\delta(n,\alpha) = t_{n-1,1-\alpha/2} \frac{\sqrt{S_n^2}}{n}$ is the estimated half-width of the CI using the Student-t approximation. Then

$$\lim_{\epsilon\to 0} \mathbb{P}(\theta \in \hat{\theta}_\tau \pm \epsilon) = 1 - \alpha.$$

The above result provides the basis for sequential estimation of the confidence interval using (1), because of the following facts: when n is large, $t_{n-1,q} \to z_q$ for any quantile q (and in practice they differ very little). In addition, the bound for $\delta(n,\alpha)$ in (2) converges to ϵ.

Sequential method: $\{X_k\}$ iid $\mathbb{E}X_n = \theta$ $Var(X_n) = \sigma^2 < \infty$.

$$\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^n X_k, \quad S_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \hat{\theta}_n)^2$$

Because the stopping time τ (also called the number of replications or the sample size) is not known in advance when using (1) or (2), we can use:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \frac{X_{n+1} - \hat{\theta}_n}{n+1},$$

$$S^2_{n+1} = \left(1 - \frac{1}{n}\right) S_n^2 + (n+1)(\hat{\theta}_{n+1} - \hat{\theta}_n)^2,$$

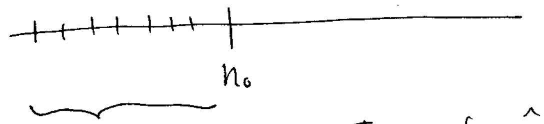
If $\delta(n,\alpha) = z_{1-\alpha/2}\frac{\sqrt{S_n^2}}{n} \leq \epsilon$, then stop the simulation.

DISCUSS CPU TIME, ADDED COMPUTATIONS.

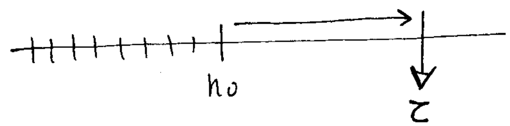
Two-stage method: Let $n_0 > 2$ be given (typically $n_0 \approx 100$), and use $S^2_{n_0}$ as a "good" estimate of σ^2 in order to calculate how many more samples would be required to attain the desired precision:

$$\tau_1 = \min\left(n \geq n_0 : z_{1-\alpha/2}\frac{\sqrt{S^2_{n_0}}}{n} \leq \epsilon\right)$$

or: $\tau_1 = \left\lfloor z_{1-\alpha/2}\frac{\sqrt{S^2_{n_0}}}{\epsilon} \right\rfloor.$



use historical observations for $\hat{\sigma}^2 = S^2_{n_0}$



Then perform the $\tau - n_0$ remaining replications.

[See A-S for examples and treatment of relative error CI's].

Experimental Design for Variance Estimation in Simulation

In our simulations, we may have problems of the following types:

- Finite or random (a.s.-finite) problems.
- Stationary processes
- Infinite horizon problems (with ergodic properties).

The three types require different approaches to establish appropriate CLT results $\rightsquigarrow$ CI estimators.

FINITE HORIZON

Model: $\{\xi_n, n \geq 0\}$ a stochastic process that we will generate via simulation. Let τ be a (possibly random) stopping time adapted to the natural filtration. The goal is to estimate

$$\theta = \mathbb{E}\left(\phi(\xi_1, \dots \xi_\tau)\right),$$

where $\phi(\cdot)$ is a well defined functional of the process.

Examples: ruin probabilities (absorbing problems), ~~time~~ algorithmic complexity (time to find solutions), estimation of direct costs or profits within deterministic time frames.

Let $X_n = \phi(\xi_1^{(n)}, \dots \xi_{\tau_n}^{(n)})$ be the corresponding performance of the n-th (independent) replication of the simulation. Then $\{X_n\}$ satisfies the assumptions of iid sampling and the methods described above are applicable.

STATIONARY PROCESSES

Def: A process $\{\xi_n\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called:

- weakly stationary if $\mathbb{E}\xi_n = \theta$, $Var(\xi_n) = \sigma^2$ are constant, and the autocovariance function satisfies:

$$Cov(\xi_n, \xi_{n+m}) = \overset{C(m)}{\cancel{\rho(m)}} \quad \forall n, m.$$

  the autocorrelation is defined by:

$$\rho(m) = \frac{C(m)}{\sigma^2}, \quad m \in \mathbb{N}.$$

- strongly stationary if the joint distribution of

$$(\xi_{n+m_1}, \xi_{n+m_2}, \dots, \xi_{n+m_k}) \overset{d}{=} (\xi_{m_1}, \dots \xi_{m_k})$$

  for any $k \in \mathbb{N}$, $n \in \mathbb{N}$ and any sequence $(m_1, \dots m_k)$.

  [stationarity: we cannot distinguish the time origin].

If $\{\xi_n\}$ is strongly stationary then it is also weakly stationary.

Exercise: Show that if $\{\xi_n; n \in \mathbb{N}\}$ is weakly stationary, and $\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} \xi_k$, then

$$Var(\hat{\theta}_n) = \frac{\sigma^2}{n}\left(1+\gamma_n\right) \qquad (3)$$

where $\gamma_n = 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k$.

It follows from this result that if $\lim_{n\to\infty} \gamma_n = 0$, then $Var(\hat{\theta}_n) \to 0 \Rightarrow \hat{\theta}_n \to \theta$ a.s. (the analogous to SLLN for stationary processes). This condition requires that the correlations decrease very rapidly: $\lim_{n\to\infty}\sum_{k=1}^{n} \rho_k < \infty$.

For stationary processes $\{\xi_n; n \geq 0\}$ with $\mathbb{E}\xi_n = \theta$, $Var(\xi_n) = \sigma^2$, the sample average $\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$ is an unbiased estimator.

However, we must find:

- estimator of $Var(\hat{\theta}_n) \equiv \hat{V}_n$. Notice that $nVar(\hat{\theta}_n) \to \underbrace{\sigma^2(1+\gamma_\infty)}_{\neq \sigma^2}$
- a result establishing a CLT for the standardized estimator

$$\frac{\sqrt{n}\,(\hat{\theta}_n - \theta)}{\sqrt{\hat{V}_n}} \Rightarrow N(0,1).$$

(*)

Approaches:

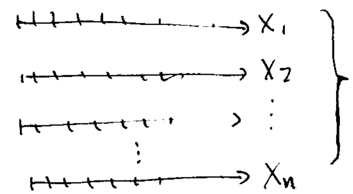
Assume that $\gamma_{n_0} \to 0$. Use n_0 "large" and set:

$$X_n = \frac{1}{n_0}\sum_{k=1}^{n_0} \xi_k^{(n)}$$

where n denotes the number of the replication. $\{X_n\}$ will approximate the stationary problem with a finite horizon approach. We use the results:

(a) For large n_0, $\{X_n\}$ are approximately normally distributed with mean θ, so that we can use the CI's for FH,

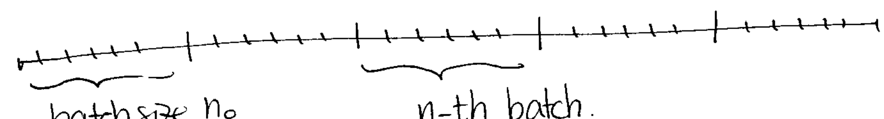
(b) $Var(X_n) \approx \frac{\sigma^2}{n_0}$, disregarding the correlation factor.



$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_n - \hat{\theta}_n)^2$$

$$\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} X_n = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{n_0}\sum_{i=1}^{n_0}\xi_i^{(n)}\right).$$

The second approach is known as the "batch-means" method:



In the context of stationary processes, when the initial distribution of $\xi_0^{(n)}$ is always the same, both approaches are equivalent.

CLT's for Markov Chains

Let $\{\xi_n\}$ be a Markov chain and assume that it is ergodic. The question is for which functionals $f: S \to \mathbb{R}^1$ can we establish that; if $\theta = \mathbb{E}_\pi(f(\xi_n))$, then:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(\xi_i) - \theta\right) \Rightarrow N(0, \sigma_f^2),$$

and what is σ_f^2 (the asymptotic variance).

Def: Let $\{\xi_n\}$ be an irreducible pos. recurrent Markov chain on state-space $\mathcal{S}$ (not necessarily finite or discrete). Assume that the stationary measure π exists and is unique, and consider a revenue function $f : \mathcal{S} \to \mathbb{R}$.

- The MC is called uniformly (geometrically) ergodic if there are constants $0<\rho<1$, $\kappa<\infty$ such that $\forall \xi_0 \in \mathcal{S}$:

$$\sup_{y\in\mathcal{S}} \| \mathbb{P}(\xi_n = y \mid \xi_0) - \pi(y) \| \le \kappa \rho^n .$$

- We say that the Markov reward process satisfies the weak approximation asumption (AWA) if $\exists\, \sigma_\infty < \infty$:

$$\frac{n(\hat{\theta}_n - \theta)}{\sigma_\infty} \Rightarrow W(n),$$

where $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i)$, $\theta = \mathbb{E}_\pi[f(\xi_n)]$, and $\{W(t); t\ge 0\}$ is the standard Brownian motion.

- We say that it satisfies the strong approximation (ASA) if it satisfies ASA and there is a constant $\lambda \in (0, 1/2]$ and a finite random variable C such that, w.p.1:

$$|n(\hat{\theta}_n - \theta) - \sigma_\infty W(n)| \le C n^{1/2-\lambda} , \text{ as } n\to\infty.$$

Thm: (Jones, 2004) Let $\{\xi_n\}$ be a uniformly geometrically ergodic MC with stationary distribution π, and assume that $\mathbb{E}_\pi[f^2(\xi_n)]<\infty$. Then for any initial distribution, as $n\to\infty$

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \sigma_\infty^2), \text{ where}$$

$$\sigma_\infty^2 = Var_\mu[f(\xi_n)] + 2\sum_{n=1}^{\infty} Cov_\pi[f(\xi_0), f(\xi_n)]$$

is the asymptotic variance.

REMARK: Most texts deal with $f(\xi)=\xi$, the identity function. However most real problems require estimation of revenue functions (instantaneous payoffs), and there is no reason why $\{f(\xi_n)\}$ should be a MC itself. In the case $f(\xi)=\xi$, the above result reproduces the asymptotic value of (3), when $n\to\infty$.

Establishing the CLT allows us to use asymptotic normality in order to provide approximate confidence intervals.

The main problem is HOW TO ESTIMATE $Var(\hat{\theta}_n)$ or σ_∞^2 directly?

[Goto: APPROACHES]

Notation for batch-means:

$$X_n = \frac{1}{n_0}\sum_{i=1}^{n_0} f(\xi_i^{(n)}) \quad \text{for } n\text{-th batch}$$

$$\hat{\theta}_{n,n_0} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{(overall mean)}$$

$$\hat{V}_{n,n_0} = \frac{1}{n-1}\sum_{k=1}^{n} (X_k - \hat{\theta}_n)^2 \quad \text{sample variance.}$$

Idea: if n_0 is "large enough", then $n\hat{V}_n$ should be a "good" estimator of $nVar(X_n) = \sigma_\infty^2 \frac{(1+r_{n_0})}{(1+r_{n_0})} \approx \sigma_\infty^2$

The rigorous justification for the method is given by the following results:

<u>Thm 1</u> [Glynn, Idlehart, 1990] If $\{\xi_n, f\}$ satisfy AWA then $\hat{\theta}_{n,n_0} \xrightarrow{P} \theta$, and the CLT holds. Furthermore if the total number of batches N is constant, and $\{b_n, n \geq 1\}$ is an increasing sequence of batch sizes as $n \to \infty$, with $b_n \to \infty$, then

$$\frac{\hat{\theta}_{N,b(n)} - \theta}{\sqrt{\hat{V}_{N,b(n)}/N}} \Rightarrow t_{N-1} \quad \text{as } n \to \infty.$$

<u>Thm 2</u> [Dameraji, 1994]. If $\{\xi_n, f\}$ satisfy ASA then $\hat{\theta}_{n,n_0} \to \theta$ a.s. and the CLT holds. Let $(b(n), N(n))$ be increasing sequences of batch sizes and number of batches such that $\frac{1}{b(n)} N_n^{1-2\lambda} \ln N_n \to 0$ as $n \to \infty$ and assume that for some integer m $\sum_{n=1}^{\infty} \left(\frac{b(n)}{n}\right)^m < \infty$. Then, as $n \to \infty$

$$b(n)\hat{V}_{N(n),b(n)} \xrightarrow{a.s.} \sigma_\infty^2 \text{ and}$$

$$\frac{\hat{\theta}_{N(n),b(n)} - \theta}{\sqrt{\hat{V}_{N(n),b(n)}/N(n)}} \Rightarrow N(0,1).$$

Both theorems justify the appropriate construction of CI's for stationary Markov chains.

[see AIS for examples].

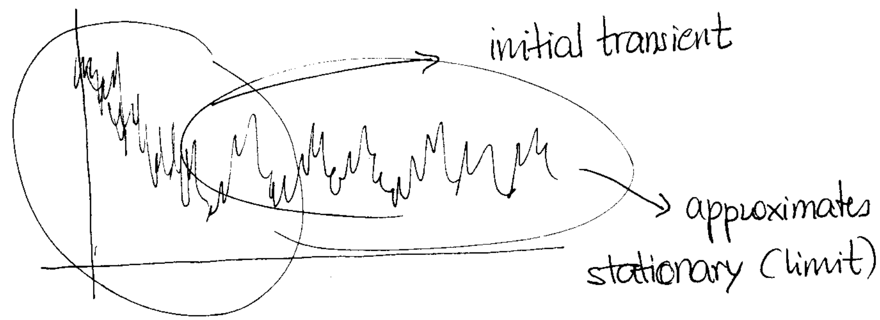
INFINITE HORIZON

$\{\xi_n\}$ a stochastic process that has a unique stationary measure π (ergodic). In many applications, because we do not know the stationary distribution, we cannot use $\xi_0 \stackrel{d}{=} \pi$. This implies that there can be an initial bias in the estimation.

Solutions:

- Reduce initial bias using "warm-up".
- Regenerative simulation approach.
- Exact (or perfect) sampling for ξ_0.

Examples: M/M/1 queue, inventory problems, autorregresive processes

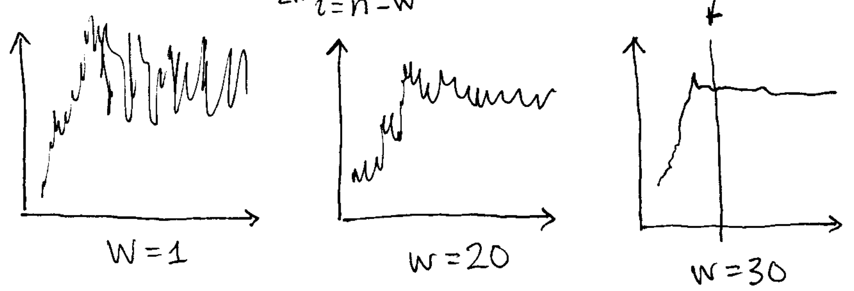


<u>Reduction of initial (transient) bias</u>: Welch statistics is a visual "smoothing" technique to determine a sufficiently large number N_0 after which we are "reasonably" confident that $\mathbb{P}(X_0 = i) \approx \pi(i)$.

<u>Idea</u>: make K replications of the simulation $\{\xi_1^{(k)}, \ldots \xi_m^{(k)}\}$, $k = 1, \ldots K$, and evaluate the "mean" process $\bar{\xi}_1, \ldots \bar{\xi}_m$; $\bar{\xi}_i = \frac{1}{K}\sum_{k=1}^{K} \xi_i^{(k)}$

Next, consider a window of size w and calculate the averages at $n = 1, \ldots m$ as follows:

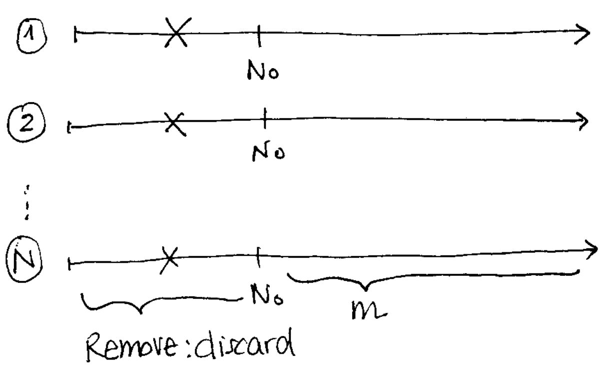
$$\bar{\xi}_n^{(w)} = \frac{1}{2w} \sum_{i=n-w}^{n+w} \bar{\xi}_i$$



This helps to provide a visual test to choose an appropriate N_0.

REMARK: Other approaches and statistical hypothesis tests for stationarity also exist. One must explain what criterion is used.

Once an initial value N_0 has been determined, one can do the replication/deletion method:



N independent replications of a stationary process

But this requires $N \times N_0$ wastefull computations. Instead, discard first N_0 and use last point $\xi_m^{(1)}$ as starting point of replication 2, $\xi_m^{(2)}$ as $\xi_0^{(3)}$, etc and we batch means to estimate the variance. [overlapping batch means is also another method]

Regenerative Method

Consider a stochastic process $\{X_t; ~~t \in T~~\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that for certain initial states $X_0 \in S$, there is a random stopping time $\tau_1 = \min(t \geq 0 : X_{t+\tau_1} \stackrel{d}{=} X_t, t \geq 0\}$, that is, at τ_1 the process behaves statistically the same way as if it was started again at X_0, and it is independent of the past.

This is called a "regenerative process". Let

$$\tau_{n+1} = \min(t \geq \tau_n : X_{t+\tau_{n+1}} \stackrel{d}{=} X_{t+\tau_n})$$

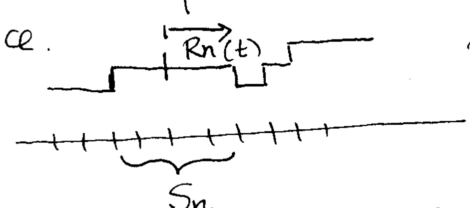
be the sequence of regeneration times. By construction, $\{\tau_n\}$ are iid random variables and the process that counts the number of regenerations, defined by:

$$N(t) = \min(n : \tau_n \geq t)$$

is called a "renewal process".

Example: a positive recurrent MC is regenerative. Any state i that is positive recurrent is a regeneration point.

Example: Consider now a queueing system (in discrete time) where arrivals are Bernouilli but service times are NOT geometrically distributed. Rather, the residual service time distribution depends on how long a customer has been in service.



$R_n(t)$: residual time @ t; for customer n.

The points $X_n = 0$ (queue lenght) are regeneration points.

Renewal Theorem: Let $\{X_n\}$ be a regenerative process with regenerative times $\{\tau_i\}$. Then the limit distribution of the process exists, and it satisfies:

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\sum_{i=1}^{n} f(X_i) = \frac{\mathbb{E}\left(\sum_{i=\tau_1}^{\tau_2-1} f(X_i)\right)}{\mathbb{E}(\tau_1)}.$$

[the continuous version also holds, with the appropriate integral].

This means that within one regenerative cycle there is enough information to build unbiased estimators of the limit distribution.

Remark that cycles are iid "pieces" of the process, so that both numerator and denominator can be estimated using iid method:

$$\hat{F}_n = \frac{1}{n}\sum_{k=1}^{n}\left(\sum_{i=\tau_k}^{\tau_{k+1}-1} f(X_i)\right) \equiv \frac{1}{n}\sum_{k=1}^{n} \underbrace{Y_k}_{\text{cycle revenue}}$$

$$\hat{\tau}_n = \frac{1}{n}\sum_{k=1}^{n} t_k \equiv T_n \quad \text{(length of } n \text{ cycles)}$$

However, notice that

$$\hat{\theta}_n = \frac{\hat{F}_n}{\hat{\tau}} \text{ is BIASED}: \mathbb{E}\hat{\theta}_n \neq \theta = \frac{\mathbb{E}\hat{F}_n}{\mathbb{E}\hat{\tau}}.$$

However, a CLT exists for renewal processes, when $n\to\infty$. It uses that $Z_n = Y_n - \theta\tau_n$ are iid zero-mean rv's, with $Var(Z_n) = \sigma_Y^2 - 2\theta\sigma_{Y\tau} + \theta^2\sigma_\tau^2 \equiv \sigma^2$ to establish that:

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)\hat{\tau}_n}{\sqrt{\hat{\sigma}^2}} \Rightarrow N(0,1). \quad \text{[see A/S for details]}$$