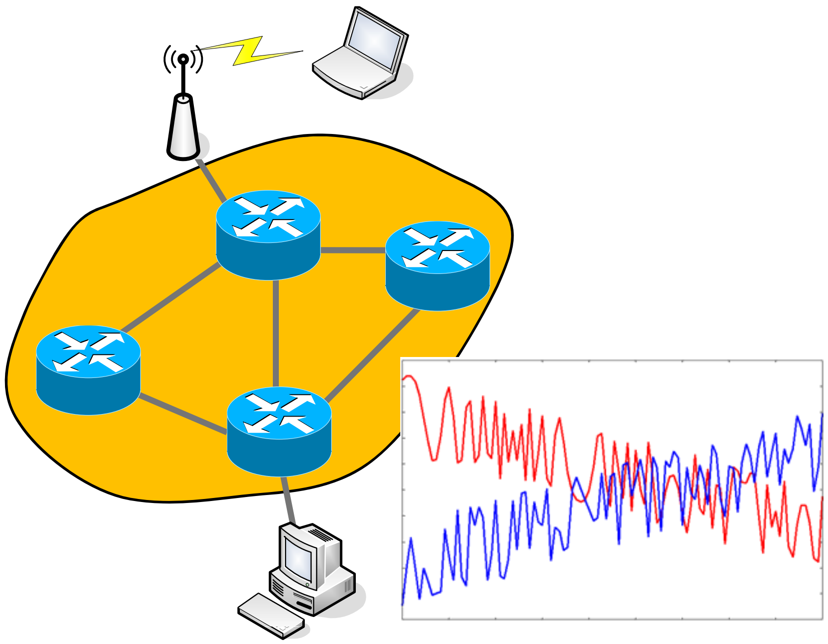
Chapter 11

Output Analysis for a Single Model



Contents

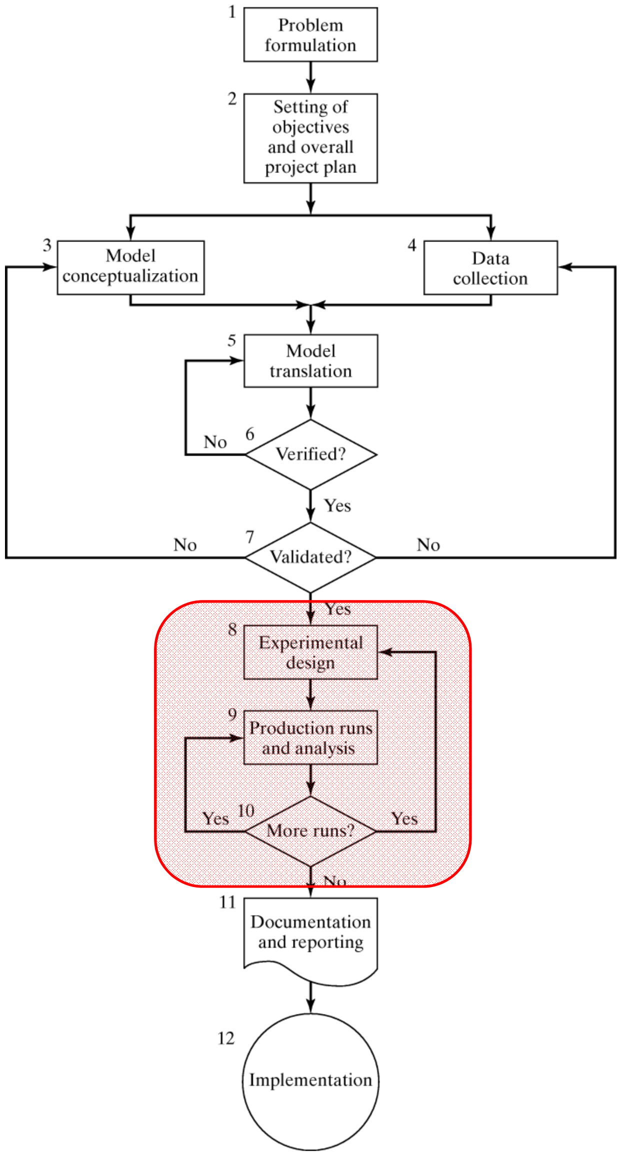
Purpose

- Output analysis: examination of the data generated by a simulation

- Objective:
  - Predict performance of system
  - Compare performance of two (or more) systems

- If θ is the system performance, the result of a simulation is an estimator $\hat{\theta}$

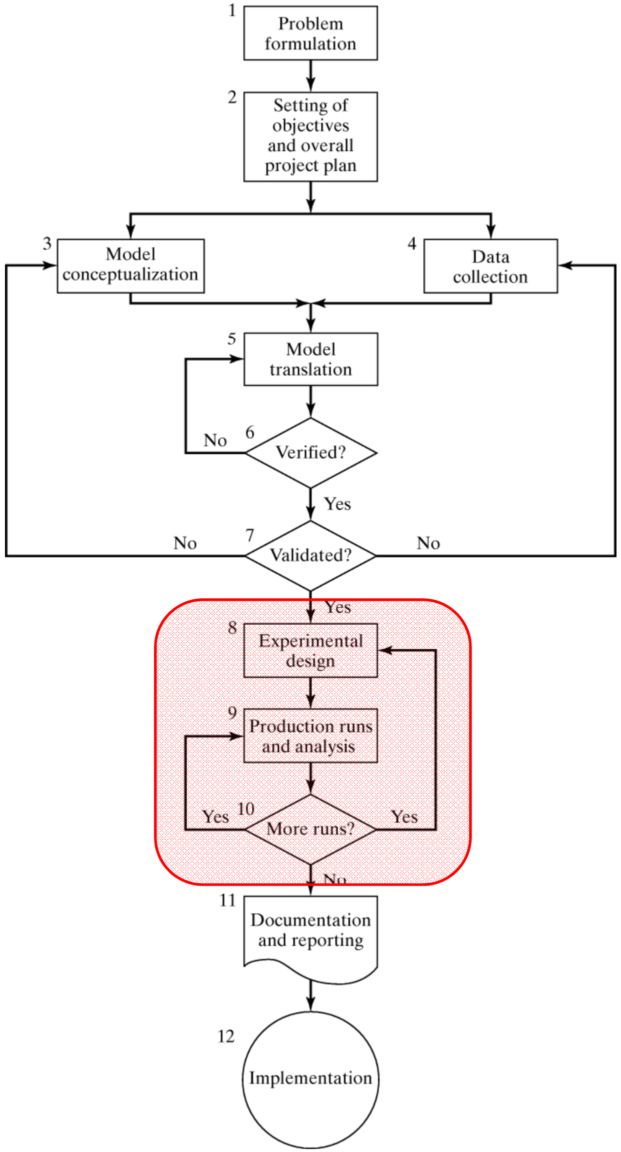
- The precision of the estimator $\hat{\theta}$ can be measured by:
  - The standard error of $\hat{\theta}$
  - The width of a confidence interval (CI) for θ



Purpose

- Purpose of statistical analysis:
  - To estimate the **standard error** and/or **confidence interval**
  - To figure out the **number of observations** required to achieve a desired error or confidence interval

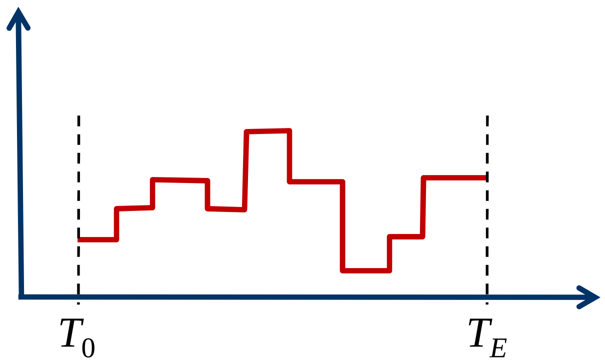
- Potential issues to overcome:
  - **Autocorrelation**, e.g., arrival of subsequent packets may lack statistical independence.
  - **Initial conditions**, e.g., the number of packets in a router at time 0 would most likely influence the performance/delay of packets arriving later.

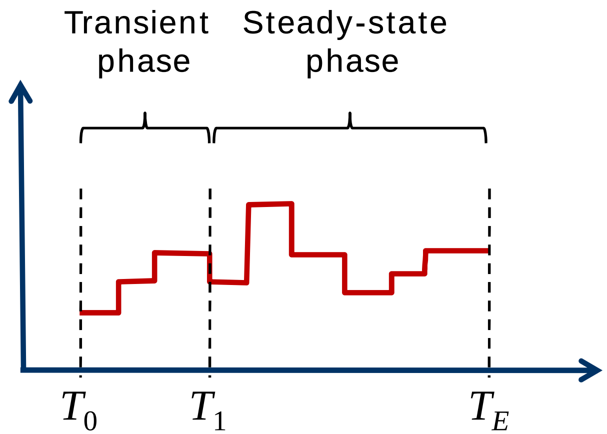


Types of Simulations

Types of Simulations

- Two types of simulation:
  - Terminating (transient)
  - Non-terminating (steady state)





Types of Simulations:
Terminating Simulations

- Terminating (transient) simulation:
  - Runs for some duration of time T_E, where E is a specified event that stops the simulation.
  - Starts at time 0 under well-specified initial conditions.
  - Ends at the stopping time T_E.
  - Bank example: Opens at 8:30 am (time 0) with no customers present and 8 of the 11 teller working (initial conditions), and closes at 4:30 pm (Time T_E = 480 minutes).
    - The simulation analyst chooses to consider it a terminating system because the object of interest is one day's operation.
  - T_E may be known from the beginning or it may not
  - Several runs may result in $T^1{}_E$, $T^2{}_E$, $T^3{}_E$,…
  - Goal may be to estimate $E(T_E)$

Types of Simulations:
Non-terminating Simulations

- Non-terminating simulation:
  - Runs continuously or at least over a very long period of time.
  - Examples: assembly lines that shut down infrequently, hospital emergency rooms, telephone systems, network of routers, Internet.
  - Initial conditions defined by the analyst.
  - Runs for some analyst-specified period of time T_E.
  - Objective is to study the **steady-state** (long-run) properties of the system, properties that are not influenced by the initial conditions of the model.

- Whether a simulation is considered to be terminating or non-terminating depends on both
  - The objectives of the simulation study and
  - The nature of the system

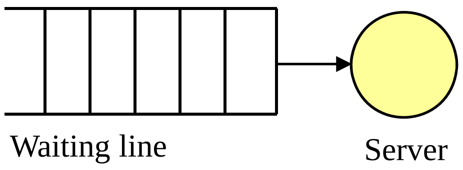
Types of Simulations

- Whether a simulation is considered to be terminating or non-terminating depends on both
  - The objectives of the simulation study and
  - The nature of the system

Stochastic Nature of Output Data

Stochastic Nature of Output Data

- Model output consist of one or more random variables because the model is an input-output transformation and the input variables are random variables.
- M/G/1 queueing example:
  - Poisson arrival rate = 0.1 per time unit and service time ~ $N(\mu = 9.5, \sigma^2 = 1.75^2)$.
  - System performance: long-run mean queue length, $L_Q(t)$.
  - Suppose we run a single simulation for a total of 5000 time units
    - Divide the time interval [0, 5000) into 5 equal subintervals of 1000 time units.
    - Average number of customers in queue from time $(j\text{-}1)1000$ to $j(1000)$ is Y_j .



$$L_Q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{\rho^2}{1-\rho}$$

Stochastic Nature of Output Data

- M/G/1 queueing example (cont.):
  - Batched average queue length for 3 independent replications:

Batching Interval	Batch *j*	Replication Y_{1j}	Y_{2j}	Y_{3j}
[0, 1000)	1	3,61	2,91	7,67
[1000, 2000)	2	3,21	9,00	19,53
[2000, 3000)	3	2,18	16,15	20,36
[3000, 4000)	4	6,92	24,53	8,11
[4000, 5000)	5	2,82	25,19	12,62
[0, 5000)		3,75	15,56	13,66

Across replication

Within replication

  - Inherent variability in stochastic simulation both within a single replication and across different replications.
  - The average across 3 replications, $\bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \bar{Y}_{3\cdot}$, can be regarded as independent observations, but averages within a replication, $Y_{11}, \ldots, Y_{15}$, are not.

Stochastic Nature of Output Data

Measures of performance

Measures of performance

- Consider the estimation of a performance parameter, θ (or ϕ), of a simulated system.
  - Discrete time data: $\{Y_1, Y_2, \ldots, Y_n\}$, with ordinary mean: θ
  - Continuous-time data: $\{Y(t), 0 \le t \le T_E\}$ with time-weighted mean: ϕ

- Point estimation for discrete time data.
  - The point estimator:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

  - Is unbiased if its expected value is θ, that is if: $E(\hat{\theta}) = \theta$ **Desired**
  - Is biased if: $E(\hat{\theta}) \neq \theta$ and $E(\hat{\theta}) - \theta$ is called **bias** of $\hat{\theta}$ **Reality**

Measures of performance:
Point Estimator

- Point estimation for continuous-time data.
  - The point estimator:

$$\hat{\phi} = \frac{1}{T_E}\int_0^{T_E} Y(t)dt$$

  - Is biased in general where: $E(\hat{\phi}) \neq \phi$
  - An unbiased or low-bias estimator is desired.

Measures of performance:
Point Estimator

- Usually, system performance measures can be put into the common framework of θ or ϕ:
  - Example: The proportion of days on which sales are lost through an out-of-stock situation, let:

$$Y(i) = \begin{cases} 1, & \text{if out of stock on day } i \\ 0, & \text{otherwise} \end{cases}$$

  - Example: Proportion of time that the queue length is larger than k_0

$$Y(t) = \begin{cases} 1, & \text{if } L_Q(\mathrm{t}) > \mathrm{k}_0 \\ 0, & \text{otherwise} \end{cases}$$

Measures of performance:
Point Estimator

- Performance measure that does not fit:
  quantile or percentile: $P(Y \le \theta) = p$

  - Estimating quantiles: the inverse of the problem of estimating a proportion or probability.

  - Consider a histogram of the observed values Y:
    - Find $\hat{\theta}$ such that 100p% of the histogram is to the left of (smaller than) $\hat{\theta}$.

  - A widely used performance measure is the median, which is the 0.5 quantile or 50-th percentile.

Measures of performance:
Confidence-Interval Estimation

- Suppose $X_1, X_2, \ldots, X_n$ are an independent sample from a normally distributed population with mean μ and variance σ^2.
- Given the sample mean and sample variance as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

- Then $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ has Student's t-distribution with n-1 degrees of freedom

- If c is the p-th quantile of this distribution, then $P(-c < T < c) = p$
- Consequently

$$P\left(\overline{X} - c\frac{S}{\sqrt{n}} < \mu < \overline{X} + c\frac{S}{\sqrt{n}}\right) = p$$

Measures of performance:
Confidence-Interval Estimation

- To understand confidence intervals fully, distinguish between measures of error and measures of risk:
  - confidence interval versus
  - prediction interval
- Suppose the model is the normal distribution with mean θ, variance σ^2 (both unknown).
  - Let $Y_{i\bullet}$ be the average cycle time for parts produced on the i-th replication of the simulation (its mathematical expectation is θ).
  - Average cycle time will vary from day to day, but over the long-run the average of the averages will be close to θ.
  - Sample variance across R replications:

$$S^2 = \frac{1}{R-1}\sum_{i=1}^{R}(Y_{i\bullet} - Y_{\bullet\bullet})^2$$

Measures of performance:
Confidence-Interval Estimation

- Confidence Interval (CI):
  - A measure of **error**.
  - Where Y_i are normally distributed.

Quantile of the t distribution with R-1 degrees of freedom.

$$\bar{Y}_{\bullet\bullet} \pm t_{\frac{\alpha}{2},R-1} \frac{S}{\sqrt{R}}$$

  - We cannot know for certain how far $\bar{Y}_{\bullet\bullet}$ is from θ but CI attempts to bound that error.
  - A CI, such as 95%, tells us how much we can trust the interval to actually bound the error between $\bar{Y}_{\bullet\bullet}$ and θ.
  - The more replications we make, the less error there is in $\bar{Y}_{\bullet\bullet}$ (converging to 0 as R goes to infinity).

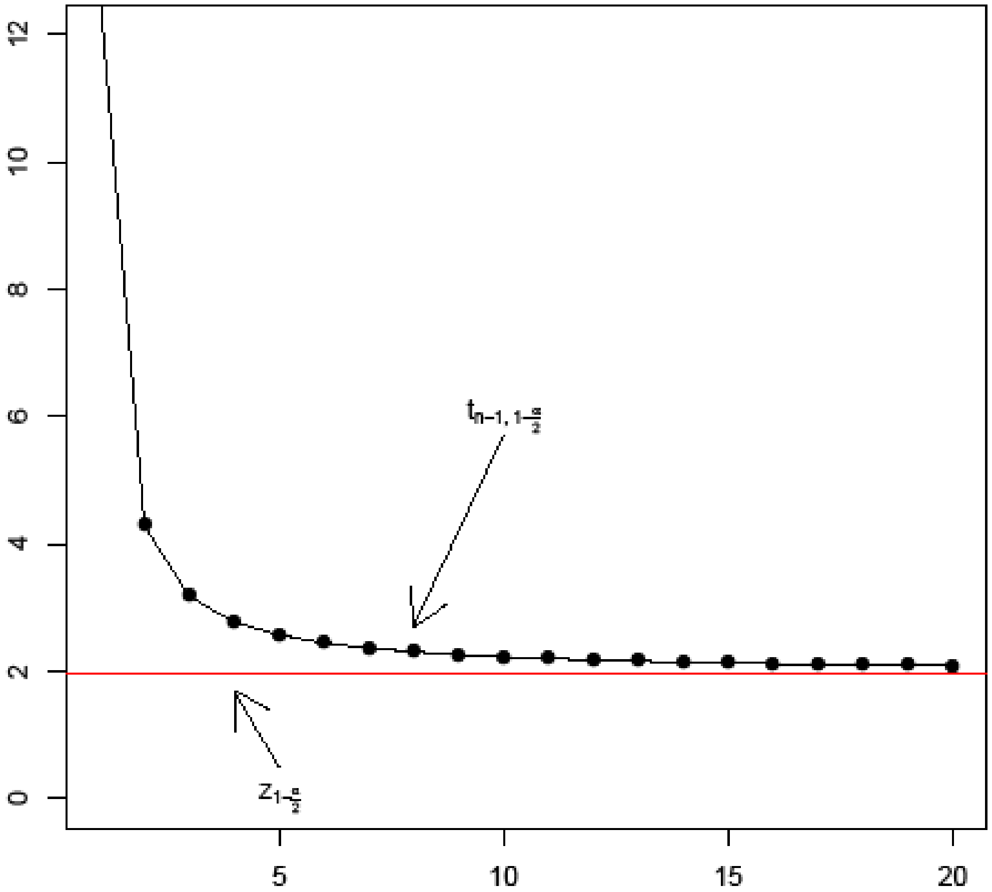
Measures of performance:
Confidence-Interval Estimation

- Prediction Interval (PI):
  - A measure of **risk**.
  - A good guess for the average cycle time on a particular day is our estimator but it is unlikely to be exactly right.
  - PI is designed to be wide enough to contain the actual average cycle time on any particular day with high probability.
  - Normal-theory prediction interval:

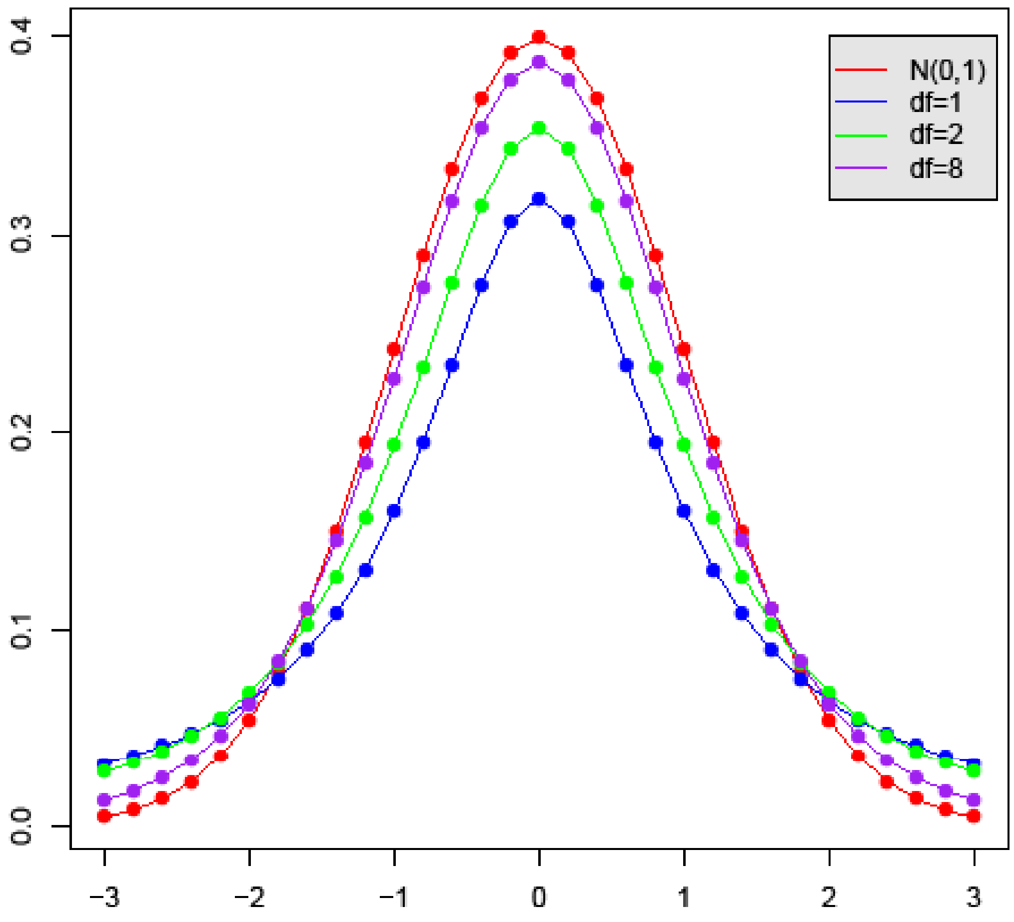
$$\bar{Y}_{\bullet\bullet} \pm t_{\frac{\alpha}{2},R-1} S \sqrt{1+\frac{1}{R}}$$

  - The length of PI will not go to 0 as R increases because we can never simulate away risk.
  - Prediction Intervals limit is: $\theta \pm z_{\frac{\alpha}{2}} \sigma$

Measures of performance:
Confidence-Interval Estimation



Measures of performance:
Confidence-Interval Estimation



Output Analysis for Terminating Simulations

Output Analysis for Terminating Simulations

- A terminating simulation: runs over a simulated time interval $[0, T_E]$.
- A common goal is to estimate:

$$\theta = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right), \qquad \text{for discrete output}$$

$$\phi = E\left(\frac{1}{T_E}\int_0^{T_E} Y(t)dt\right), \quad \text{for continuous output} \quad Y(t), \quad 0 \le t \le T_E$$

- In general, independent replications are used, each run using a different random number stream and independently chosen initial conditions.

Statistical Background

- Important to distinguish within-replication data from across-replication data.
- For example, simulation of a manufacturing system
  - Two performance measures of that system: cycle time for parts and work in process (WIP).
  - Let Y_{ij} be the cycle time for the j-th part produced in the i-th replication.
  - Across-replication data are formed by summarizing within-replication data $\bar{Y}_{i\bullet}$

	Within-Replication Data				Across-Rep. Data			
1	Y_{11}	Y_{12}	$\cdots$	Y_{1n_1}	$\bar{Y}_{1\bullet}$,	S_1^2,	H_1	Within replication performance measure
2	Y_{21}	Y_{22}	$\cdots$	Y_{2n_2}	$\bar{Y}_{2\bullet}$,	S_2^2,	H_2	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		
R	Y_{R1}	Y_{R2}	$\cdots$	Y_{Rn_R}	$\bar{Y}_{R\bullet}$,	S_R^2,	H_R	
					$\bar{Y}_{\bullet\bullet}$,	S^2,	H	Across replication performance measure

Statistical Background

- Across Replication:
  - For example: the daily cycle time averages (discrete time data)
    - The average: $\bar{Y}_{\bullet\bullet} = \frac{1}{R}\sum_{i=1}^{R} Y_{i\bullet}$
    - The sample variance: $S^2 = \frac{1}{R-1}\sum_{i=1}^{R}(Y_{i\bullet} - \bar{Y}_{\bullet\bullet})^2$
    - The confidence-interval half-width: $H = t_{\frac{\alpha}{2}, R-1} \frac{S}{\sqrt{R}}$
- Within replication:
  - For example: the WIP (a continuous time data)
    - The average: $\bar{Y}_{i\bullet} = \frac{1}{T_{E_i}} \int_0^{T_{E_i}} Y_i(t)\,dt$
    - The sample variance: $S_i^2 = \frac{1}{T_{Ei}} \int_0^{T_{Ei}} \left(Y_i(t) - \bar{Y}_{i\bullet}\right)^2 dt$

Statistical Background

- Overall sample average, $\bar{Y}_{\bullet\bullet}$, and the interval replication sample averages, $\bar{Y}_{i\bullet}$, are always unbiased estimators of the expected daily average cycle time or daily average WIP.

- **Across-replication** data are **independent** and **identically distributed**
  - Same model
  - Different random numbers for each replications

- **Within-replication** data are **not independent** and not identically distributed
  - One random number stream is used within a replication

Output Analysis for Terminating Simulations

Confidence Intervals with Specified Precision

Confidence Intervals with Specified Precision

- The half-length H of a $100(1 - \alpha)\%$ confidence interval for a mean θ, based on the t distribution, is given by:

$$H = t_{\frac{\alpha}{2}, R-1} \frac{S}{\sqrt{R}}$$

S^2 is the sample variance

R is the number of replications

- Suppose that an error criterion ε is specified with probability 1-α, a sufficiently large sample size should satisfy:

$$P\left(\left|\overline{Y}_{\bullet\bullet} - \theta\right| < \varepsilon\right) \geq 1 - \alpha$$

Confidence Intervals with Specified Precision

- Assume that an initial sample of size R_0 (independent) replications has been observed.
- Obtain an initial estimate S_0^2 of the population variance σ^2.

$$H = t_{\frac{\alpha}{2},R-1} \frac{S_0}{\sqrt{R}} \le \varepsilon$$

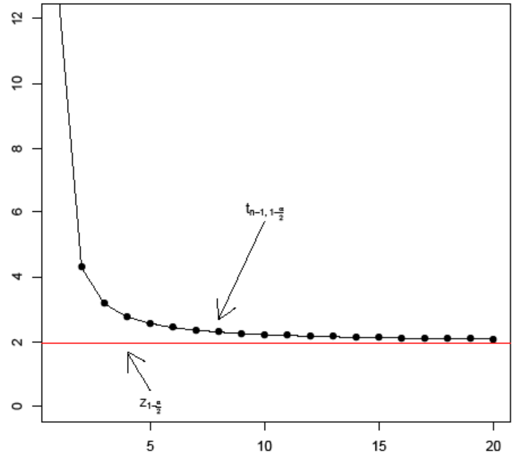
- Then, choose sample size R such that $R \ge R_0$
- Solving for R

$$R \ge \left(\frac{t_{\alpha/2,R-1} S_0}{\varepsilon} \right)^2$$

Confidence Intervals with Specified Precision

- Since $t_{\alpha/2,R-1} \geq z_{\alpha/2}$, an initial estimate for R is given by

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\varepsilon} \right)^2, \quad z_{\alpha/2} \text{ is the standard normal distribution.}$$



  - For large R $\quad t_{\alpha/2,R-1} \approx z_{\alpha/2}$
  - R is the smallest integer satisfying $R \geq R_0$

- Collect $R - R_0$ additional observations.

- The $100(1-\alpha)\%$ confidence interval for θ:

$$\bar{Y}_{\bullet\bullet} \pm t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$$

Confidence Intervals with Specified Precision

- Call Center Example: estimate the agent's utilization ρ over the first 2 hours of the workday.
  - Initial sample of size $R_0 = 4$ is taken and an initial estimate of the population variance is $S_0^{\,2} = (0.072)^2 = 0.00518$.
  - The error criterion is $\varepsilon = 0.04$ and confidence coefficient is 1-$\alpha = 0.95$, hence, the **final sample size** must be at least:

$$\left(\frac{z_{0.025} S_0}{\varepsilon}\right)^2 = \frac{1.96^2 \times 0.00518}{0.04^2} = 12.44$$

  - For the final sample size:

R	**13**	**14**	**15**
$t_{0.025,\, R-1}$	2,18	2,16	2,14
$\left(t_{\alpha/2,R-1} S_0 / \varepsilon\right)^2$	15,39	15,1	14,83

  - $R = 15$ is the smallest integer satisfying the error criterion so R - $R_0 = 11$ additional replications are needed. $R \geq \left(\frac{t_{\alpha/2,R-1} S_0}{\varepsilon}\right)^2$
  - After obtaining additional outputs, half-width should be checked.

Output Analysis for Terminating Simulations

Quantiles

Quantiles

- Here, a proportion or probability is treated as a special case of a mean.
- When the number of independent replications $Y_1, \ldots, Y_R$ is large enough that $t_{\alpha/2,R-1} \approx z_{\alpha/2}$, the confidence interval for a probability p is often written as:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{R-1}}$$

The sample proportion

- A quantile is the inverse of the probability estimation problem:

p is given

$$\boxed{\text{Find } \theta \text{ such that } P(Y \leq \theta) = p}$$

Quantiles

- The best way is to sort the outputs and use the $(R \times p)$-th smallest value, i.e., find θ such that $100p\%$ of the data in a histogram of Y is to the left of θ.
  - Example: If we have R=10 replications and we want the $p = 0.8$ quantile, first sort, then estimate θ by the $(10)(0.8) = 8$-th smallest value (round if necessary).

5.6 ←sorted data
7.1
8.8
8.9
9.5
9.7
10.1
12.2 ←this is our point estimate
12.5
12.9

Quantiles

- **Confidence Interval of Quantiles:** An approximate $(1-\alpha)100\%$ confidence interval for θ can be obtained by finding two values θ_l and θ_u.
  - θ_l cuts off $100p_l\%$ of the histogram (the $R\times p_l$ smallest value of the sorted data).
  - θ_u cuts off $100p_u\%$ of the histogram (the $R\times p_u$ smallest value of the sorted data).

$$\text{where} \quad p_\ell = p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{R-1}}$$

$$p_u = p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{R-1}}$$

Quantiles

- Example: Suppose $R = 1000$ replications, to estimate the $p = 0.8$ quantile with a 95% confidence interval.
  - First, sort the data from smallest to largest.
  - Then estimate of θ by the (1000)(0.8) = 800-th smallest value, and the point estimate is 212.03.
  - And find the confidence interval:

$$p_\ell = 0.8 - 1.96\sqrt{\frac{0.8(1-0.8)}{1000-1}} = 0.78$$

$$p_u = 0.8 + 1.96\sqrt{\frac{0.8(1-0.8)}{1000-1}} = 0.82$$

$$\text{The CI is the } 780^{\text{th}} \text{ and } 820^{\text{th}} \text{ smallest values}$$

A portion of the 1000 sorted values:

	Output	Rank	
	180.92	779	
p_l	**188.96**	**780**	
	190.55	781	
	208.58	799	
	212.03	**800**	
	216.99	801	
	250.32	819	
p_u	**256.79**	**820**	
	256.99	821	

  - The point estimate is 212.03
  - The 95% CI is [188.96, 256.79]

Output Analysis for Steady-State Simulation

Output Analysis for Steady-State Simulation

- Consider a single run of a simulation model to estimate a steady-state or long-run characteristics of the system.
  - The single run produces observations $Y_1, Y_2, \ldots$ (generally the samples of an autocorrelated time series).
  - Performance measure:

$$\theta = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_i, \qquad \text{for discrete measure} \qquad \text{(with probability 1)}$$

$$\phi = \lim_{T_E\to\infty} \frac{1}{T_E} \int_0^{T_E} Y(t)\,dt, \qquad \text{for continuous measure} \qquad \text{(with probability 1)}$$

  - Independent of the initial conditions.

Output Analysis for Steady-State Simulation

- The sample size is a design choice, with several considerations in mind:
  - Any bias in the point estimator that is due to artificial or arbitrary initial conditions (bias can be severe if run length is too short).
  - Desired precision of the point estimator.
  - Budget constraints on computer resources.

- Notation: the estimation of θ from a discrete-time output process.
  - One replication (or run), the output data: $Y_1, Y_2, Y_3, \ldots$
  - With several replications, the output data for replication r: $Y_{r1}, Y_{r2}, Y_{r3}, \ldots$

Output Analysis for Steady-State Simulation

Initialization Bias

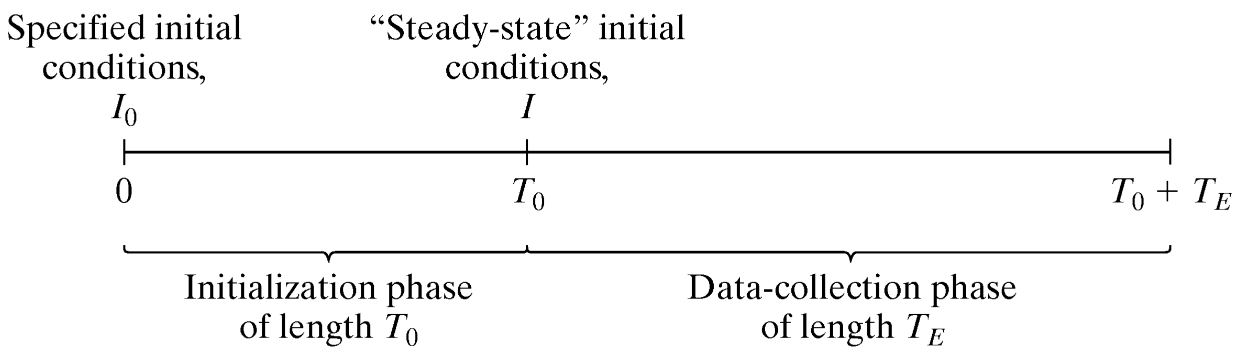
Initialization Bias

- Methods to reduce the point-estimator bias caused by using artificial and unrealistic initial conditions:
  - Intelligent initialization.
  - Divide simulation into an initialization phase and data-collection phase.

- Intelligent initialization
  - Initialize the simulation in a state that is more representative of long-run conditions.
  - If the system exists, collect data on it and use these data to specify more nearly typical initial conditions.
  - If the system can be simplified enough to make it mathematically solvable, e.g., queueing models, solve the simplified model to find long-run expected or most likely conditions, use that to initialize the simulation.

Initialization Bias

- Divide each simulation into two phases:
  - An initialization phase, from time 0 to time T_0.
  - A data-collection phase, from T_0 to the stopping time T_0+T_E.
  - The choice of T_0 is important:
    - After T_0, system should be more nearly representative of steady-state behavior.
  - System has reached steady state: the probability distribution of the system state is close to the steady-state probability distribution (bias of response variable is negligible).



Initialization Bias

- M/G/1 queueing example: A total of 10 independent replications were made.
  - Each replication begins in the empty and idle state.
  - Simulation run length on each replication: $T_0+T_E = 15000$ time units.
  - Response variable: queue length, $L_Q(t,r)$ (at time t of the r-th replication).
  - Batching intervals of 1000 minutes, batch means

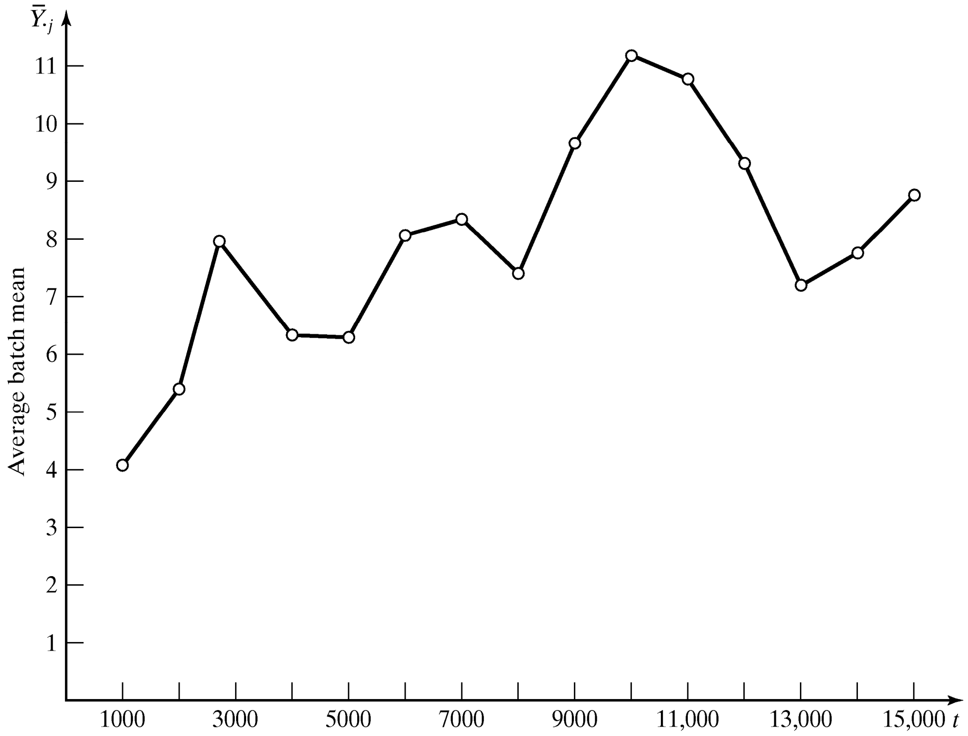
$$Y_{rj} = \int_{(j-1)1000}^{j1000} L_Q(t,r)dt$$

- Ensemble averages:
  - To identify trend in the data due to initialization bias
  - The average corresponding batch means across replications:

$$\bar{Y}_{.j} = \frac{1}{R}\sum_{r=1}^{R} Y_{rj}$$

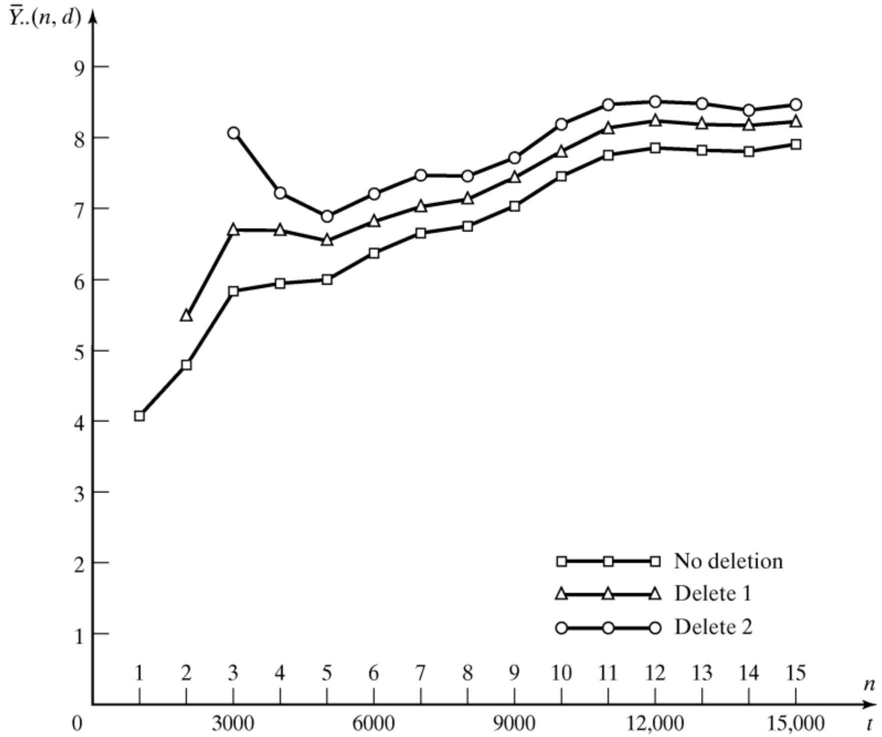
Initialization Bias

- A plot of the ensemble averages, $\bar{Y}_{\bullet j}$, versus $1000j$, for j = 1,2, …,15.



Initialization Bias

- Cumulative average sample mean (after deleting d observations):



$$\overline{Y}_{\bullet\bullet}(n,d) = \frac{1}{n-d} \sum_{j=d+1}^{n} \overline{Y}_{\bullet j}$$

  - Not recommended to determine the initialization phase.
  - It is apparent that downward bias is present and this bias can be reduced by deletion of one or more observations.

Initialization Bias

- No widely accepted, objective and proven technique to guide how much data to delete to reduce initialization bias to a negligible level.
- Plots can, at times, be misleading but they are still recommended.
  - Ensemble averages reveal a smoother and more precise trend as the number of replications, R, increases.
  - Ensemble averages can be smoothed further by plotting a moving average.
  - Cumulative average becomes less variable as more data are averaged.
  - The more correlation present, the longer it takes for $\bar{Y}_{\bullet j}$ to approach steady state.
  - Different performance measures could approach steady state at different rates.

Output Analysis for Steady-State Simulation

Error Estimation

Error Estimation

- If $\{Y_1, \ldots, Y_n\}$ are not statistically independent, then S^2/n is a biased estimator of the true variance.
  - Almost always the case when $\{Y_1, \ldots, Y_n\}$ is a sequence of output observations from within a single replication (autocorrelated sequence, time-series).

- Suppose the point estimator θ is the sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad V(\bar{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}(Y_i, Y_j)$$

  - Variance of $\bar{Y}$ is very hard to estimate.
  - For systems with steady state, produce an output process that is approximately covariance stationary (after passing the transient phase).
    - The covariance between two random variables in the time series depends only on the lag, i.e., the number of observations between them.

Error Estimation

- For a covariance stationary time series, $\{Y_1, \ldots, Y_n\}$:
  - Lag-k autocovariance is: $\gamma_k = \operatorname{cov}(Y_1, Y_{1+k}) = \operatorname{cov}(Y_i, Y_{i+k})$
  - Lag-k autocorrelation is: $\rho_k = \dfrac{\gamma_k}{\sigma^2}\ , \quad -1 \le \rho_k \le 1$
- If a time series is covariance stationary, then the variance of $\bar{Y}$ is:

$$V(\bar{Y}) = \frac{\sigma^2}{n} \underbrace{\left[1 + 2\sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \rho_k \right]}_{c}$$

- The expected value of the variance estimator is:

$$E\left(\frac{S^2}{n} \right) = B \cdot V(\bar{Y}), \qquad \text{where } B = \frac{n/c - 1}{n - 1}$$

Error Estimation

(a) $\rho_k > 0 \quad \text{for most} \quad k$

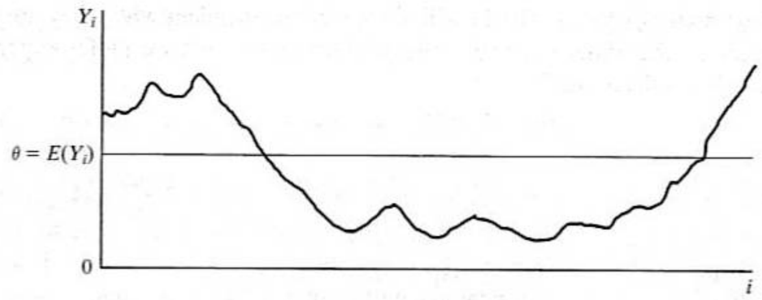
Stationary time series Y_i exhibiting positive autocorrelation.

- Series slowly drifts above and then below the mean.

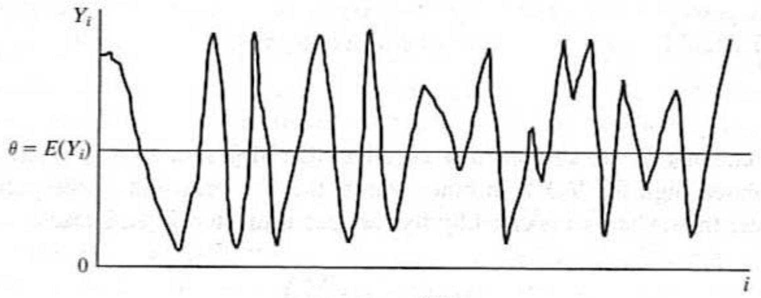
(b) $\rho_k < 0 \quad \text{for most} \quad k$

Stationary time series Y_i exhibiting negative autocorrelation.

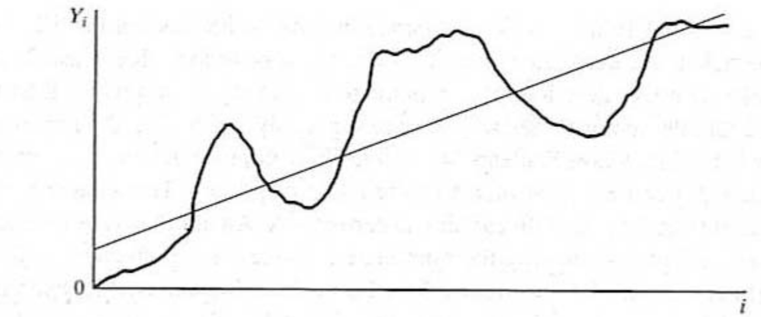
(c) Non-stationary time series with an upward trend



(a)



(b)



(c)

Error Estimation

- The expected value of the variance estimator is:

$$E\left(\frac{S^2}{n}\right) = B \cdot V(\bar{Y}), \qquad \text{where } B = \frac{n/c-1}{n-1} \text{ and } V(\bar{Y}) \text{ is the variance of } \bar{Y}$$

  - If Y_i are independent ➡ $\rho_k=0$, then S^2/n is an unbiased estimator of $V(\bar{Y})$
  - If the autocorrelation ρ_k are primarily positive, then S^2/n is biased low as an estimator of $V(\bar{Y})$.
  - If the autocorrelation ρ_k are primarily negative, then S^2/n is biased high as an estimator of $V(\bar{Y})$.

Output Analysis for Steady-State Simulation

Replication Method

Replication Method

- Use to estimate point-estimator variability and to construct a confidence interval.
- Approach: make R replications, initializing and deleting from each one the same way.
- Important to do a thorough job of investigating the initial-condition bias:
  - Bias is not affected by the number of replications, instead, it is affected only by deleting more data (i.e., increasing T_0) or extending the length of each run (i.e. increasing T_E).
- Basic raw output data $\{Y_{rj}, r = 1, ..., R; j = 1, \ldots, n\}$ is derived by:
  - Individual observation from within replication r.
  - Batch mean from within replication r of some number of discrete-time observations.
  - Batch mean of a continuous-time process over time interval j.

Replication	Observations 1	$\cdots$	d	$d+1$	$\cdots$	n	Replication Averages
1	$Y_{1,1}$	$\cdots$	$Y_{1,d}$	$Y_{1,d+1}$	$\cdots$	$Y_{1,n}$	$\bar{Y}_{1\bullet}(n,d)$
2	$Y_{2,1}$	$\cdots$	$Y_{2,d}$	$Y_{2,d+1}$	$\cdots$	$Y_{2,n}$	$\bar{Y}_{2\bullet}(n,d)$
$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$
R	$Y_{R,1}$	$\cdots$	$Y_{R,d}$	$Y_{R,d+1}$	$\cdots$	$Y_{R,n}$	$\bar{Y}_{R\bullet}(n,d)$
	$\bar{Y}_{\bullet 1}$	$\cdots$	$\bar{Y}_{\bullet d}$	$\bar{Y}_{\bullet (d+1)}$	$\cdots$	$\bar{Y}_{\bullet n}$	$\bar{Y}_{\bullet\bullet}(n,d)$

Replication Method

- Each replication is regarded as a single sample for estimating θ.
  For replication r:

$$\bar{Y}_{r\bullet}(n,d) = \frac{1}{n-d}\sum_{j=d+1}^{n} Y_{rj}$$

- The overall point estimator:

$$\bar{Y}_{\bullet\bullet}(n,d) = \frac{1}{R}\sum_{r=1}^{R} \bar{Y}_{r\bullet}(n,d) \quad \text{and} \quad \mathrm{E}[\bar{Y}_{\bullet\bullet}(n,d)] = \theta_{n,d}$$

- If d and n are chosen sufficiently large:
  - $\theta_{n,d} \sim \theta$.
  - $\bar{Y}_{\bullet\bullet}(n,d)$ is an approximately unbiased estimator of θ.

Replication Method

- To estimate the standard error of $\bar{Y}_{\bullet\bullet}$, compute the sample variance and standard error:

$$S^2 = \frac{1}{R-1}\sum_{r=1}^{R}(\bar{Y}_{r\bullet} - \bar{Y}_{\bullet\bullet})^2 = \frac{1}{R-1}\left(\sum_{r=1}^{R}\bar{Y}_{r\bullet}^{\;2} - R\bar{Y}_{\bullet\bullet}^2\right) \quad \text{and} \quad s.e.(\bar{Y}_{\bullet\bullet}) = \frac{S}{\sqrt{R}}$$

Mean of the undeleted observations from the r-th replication.

Mean of $\bar{Y}_{1\bullet}(n,d),\ldots,\bar{Y}_{R\bullet}(n,d)$

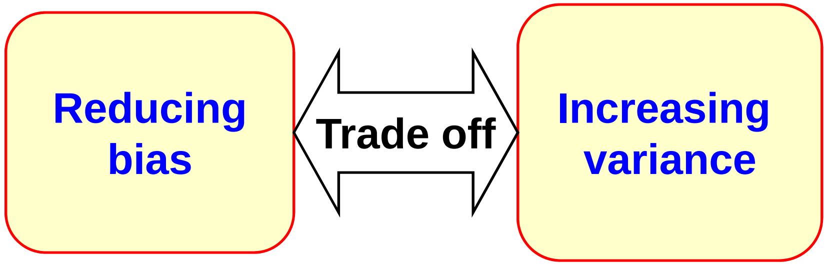
Standard error

Replication Method

- Length of each replication (n) beyond deletion point (d):

$$(n-d) > 10d \quad \text{or} \quad T_E > 10T_0$$

- Number of replications (R) should be as many as time permits, up to about 25 replications.
- For a fixed total sample size (n), as fewer data are deleted ($\downarrow d$):
  - Confidence interval shifts: greater bias.
  - Standard error of $\bar{Y}_{\bullet\bullet}(n,d)$ decreases: decrease variance.



Replication Method

- M/G/1 queueing example:
  - Suppose R=10, each of length T_E =15000 time units, starting at time 0 in the empty and idle state, initialized for T_0 = 2000 time units before data collection begins.
  - Each batch means is the average number of customers in queue for a 1000-time-unit interval.
  - The 1-st two batch means are deleted (d=2).

Replication, r	Sample Mean for Replication r (No Deletion) $\bar{Y}_{r\cdot}(15,0)$	(Delete 1) $\bar{Y}_{r\cdot}(15,1)$	(Delete 2) $\bar{Y}_{r\cdot}(15,2)$
1	3.27	3.24	3.25
2	16.25	17.20	17.83
3	15.19	15.72	15.43
4	7.24	7.28	7.71
5	2.93	2.98	3.11
6	4.56	4.82	4.91
7	8.44	8.96	9.45
8	5.06	5.32	5.27
9	6.33	6.14	6.24
10	10.10	10.48	11.07
$\bar{Y}_{\cdot\cdot}(15,d)$	7.94	8.21	8.43
$\sum_{r=1}^{R} \bar{Y}_{r\cdot}^2$	826.20	894.68	938.34
S^2	21.75	24.52	25.30
S	4.66	4.95	5.03
$S/\sqrt{10} = \text{s.e.}(\bar{Y}_{\cdot\cdot})$	1.47	1.57	1.59

- The point estimator and standard error are:

$$\bar{Y}_{\bullet\bullet}(15,2) = 8.43 \qquad \text{and} \qquad s.e.\left(\bar{Y}_{\bullet\bullet}(15,2)\right) = 1.59$$

- The 95 % CI for long-run mean queue length is:

$$\bar{Y}_{\bullet\bullet} - t_{\alpha/2,R-1}\frac{S}{\sqrt{R}} \le \theta \le \bar{Y}_{\bullet\bullet} + t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$$

$$8.43 - 2.26(1.59) \le L_Q \le 8.43 + 2.26(1.59)$$

- A high degree of confidence that the long-run mean queue length is between 4.84 and 12.02 (if d and n are "large" enough).

Output Analysis for Steady-State Simulation

Sample Size

Sample Size

- To estimate a long-run performance measure, θ, within $\pm\varepsilon$ with confidence $100(1-\alpha)\%$.
- M/G/1 queuing example (cont.):
  - We know: $R_0 = 10$, $d = 2$ deleted and $S_0^2 = 25.30$.
  - To estimate the long-run mean queue length, L_Q, within $\varepsilon = 2$ customers with 90% confidence ($\alpha = 10\%$).
  - Initial estimate:

$$R \geq \left(\frac{z_{0.05} S_0}{\varepsilon}\right)^2 = \frac{1.645^2 \times 25.30}{2^2} = 17.1$$

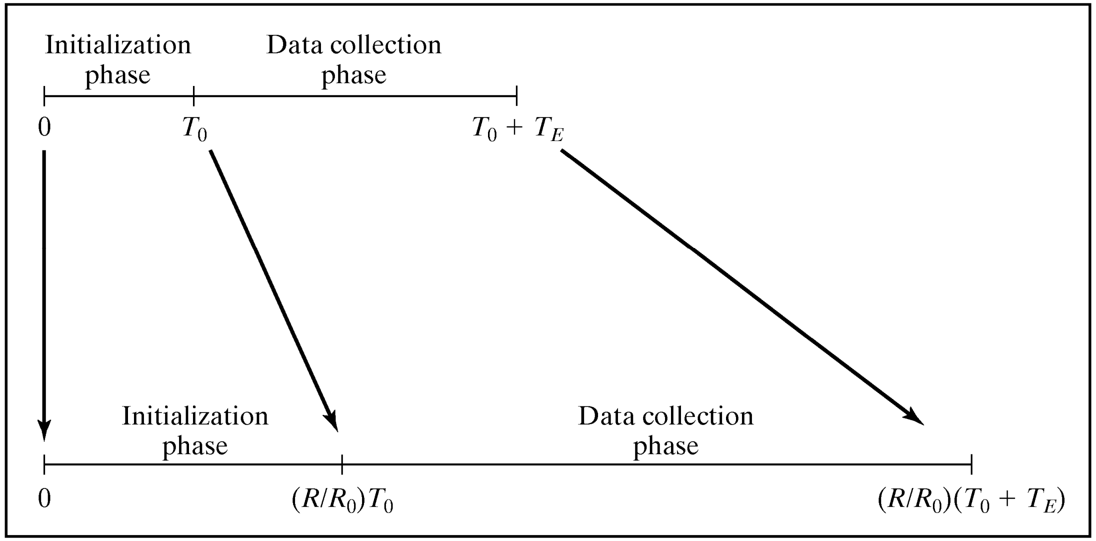
  - Hence, at least 18 replications are needed, next try $R = 18, 19, \ldots$ using $R \geq \left(\frac{t_{0.05,R-1} S_0}{\varepsilon}\right)^2$. We found that:

$$R = 19 \geq \left(\frac{t_{0.05,19-1} S_0}{\varepsilon}\right)^2 = 1.73^2 \times \frac{25.3}{4} = 18.93$$

  - Additional replications needed is $R - R_0 = 19\text{-}10 = 9$.

Sample Size

- An alternative to increasing R is to increase total run length T_0+T_E within each replication.
- Approach:
  - Increase run length from (T_0+T_E) to $(R/R_0)(T_0+T_E)$, and
  - delete additional amount of data, from time 0 to time $(R/R_0)T_0$.



- Advantage: any residual bias in the point estimator should be further reduced.
- However, it is necessary to have saved the state of the model at time T_0+T_E and to be able to restart the model.

Output Analysis for Steady-State Simulation

Batch Means

Batch Means for Interval Estimation

- Using a single, long replication:
  - Problem: data are dependent so the usual estimator is biased.
  - Solution: batch means.
- Batch means: divide the output data from 1 replication (after appropriate deletion) into a few large batches and then treat the means of these batches as if they were independent.
- A continuous-time process, $\{Y(t), T_0 \le t \le T_0+T_E\}$:
  - k batches of size $m = T_E / k$, batch means:

$$\bar{Y}_j = \frac{1}{m}\int_{(j-1)m}^{jm} Y(t+T_0)dt \quad j = 1,2,\ldots,k$$

- A discrete-time process, $\{Y_i, i = d+1, d+2, \ldots, n\}$:
  - k batches of size $m = (n - d)/k$, batch means:

$$\bar{Y}_j = \frac{1}{m}\sum_{i=(j-1)m+1}^{jm} Y_{i+d} \quad j = 1,2,\ldots,k$$

Batch Means for Interval Estimation

$$\underbrace{Y_1, \ldots, Y_d}_{\text{deleted}}, \underbrace{Y_{d+1}, \ldots, Y_{d+m}}_{\bar{Y}_1}, \underbrace{Y_{d+m+1}, \ldots, Y_{d+2m}}_{\bar{Y}_2}, \ \ldots \ , \underbrace{Y_{d+(k-1)m+1}, \ldots, Y_{d+km}}_{\bar{Y}_k}$$

- Starting either with continuous-time or discrete-time data, the variance of the sample mean is estimated by:

$$\frac{S^2}{k} = \frac{1}{k}\sum_{j=1}^{k}\frac{\left(\bar{Y}_j - \bar{Y}\right)^2}{k-1} = \sum_{j=1}^{k}\frac{\bar{Y}_j^{\,2} - k\bar{Y}^{\,2}}{k(k-1)}$$

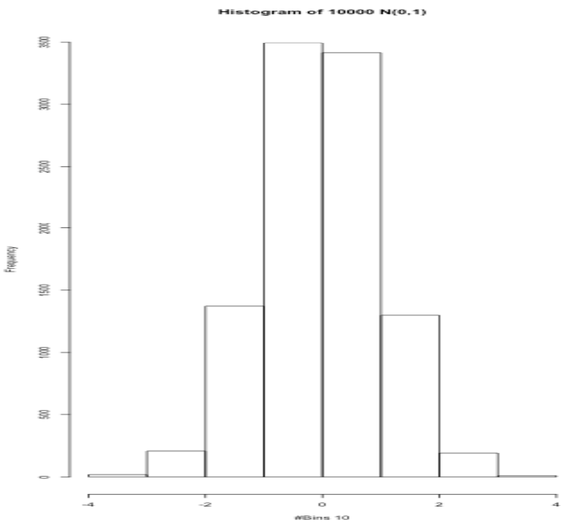
- If the batch size is sufficiently large, successive batch means will be approximately independent, and the variance estimator will be approximately unbiased.
- No widely accepted and relatively simple method for choosing an acceptable batch size m. Some simulation software does it automatically.

The Art of Data Presentation

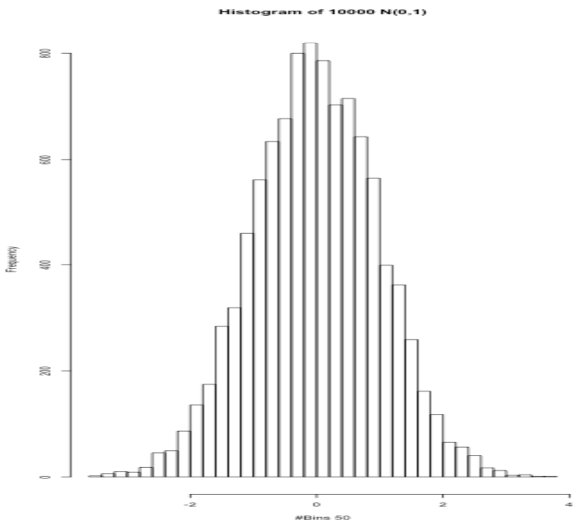
The art of data presentation

- Always get the following statistical sample data
  - Min
  - Max
  - Mean
  - Median
  - Standard deviation
  - CI_low
  - CI_high
  - 1st-quartile
  - 3rd-quartile

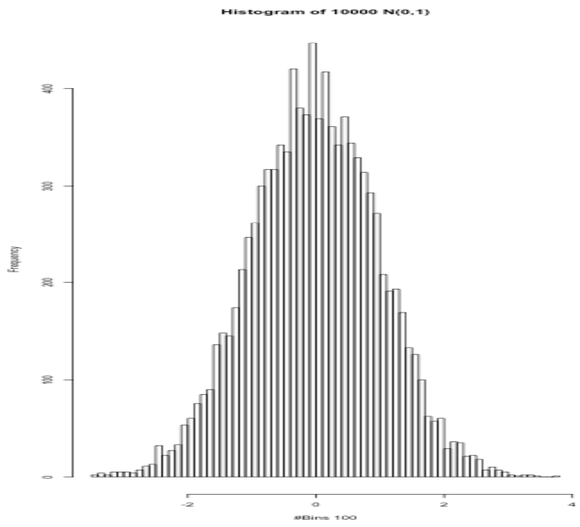
Histograms



Histogram of 10000 N(0,1)
Frequency
#Bins 10



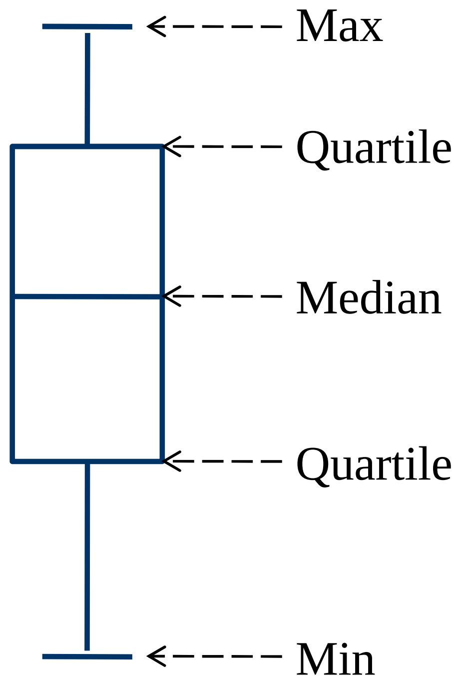
Histogram of 10000 N(0,1)
Frequency
#Bins 50



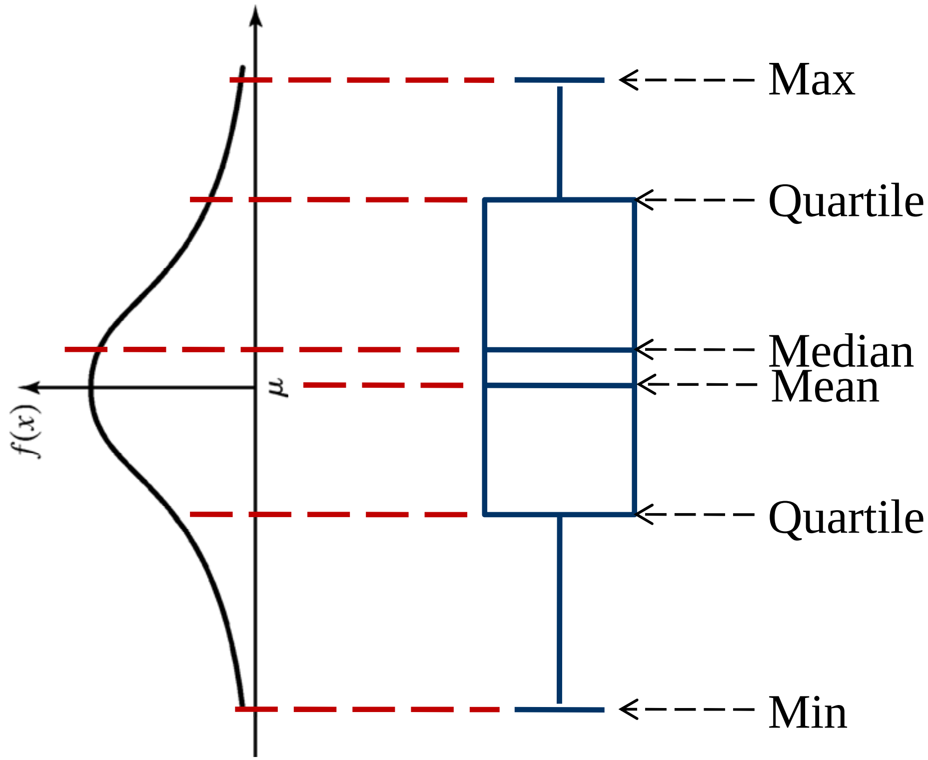
Histogram of 10000 N(0,1)
Frequency
#Bins 100

Box Plot

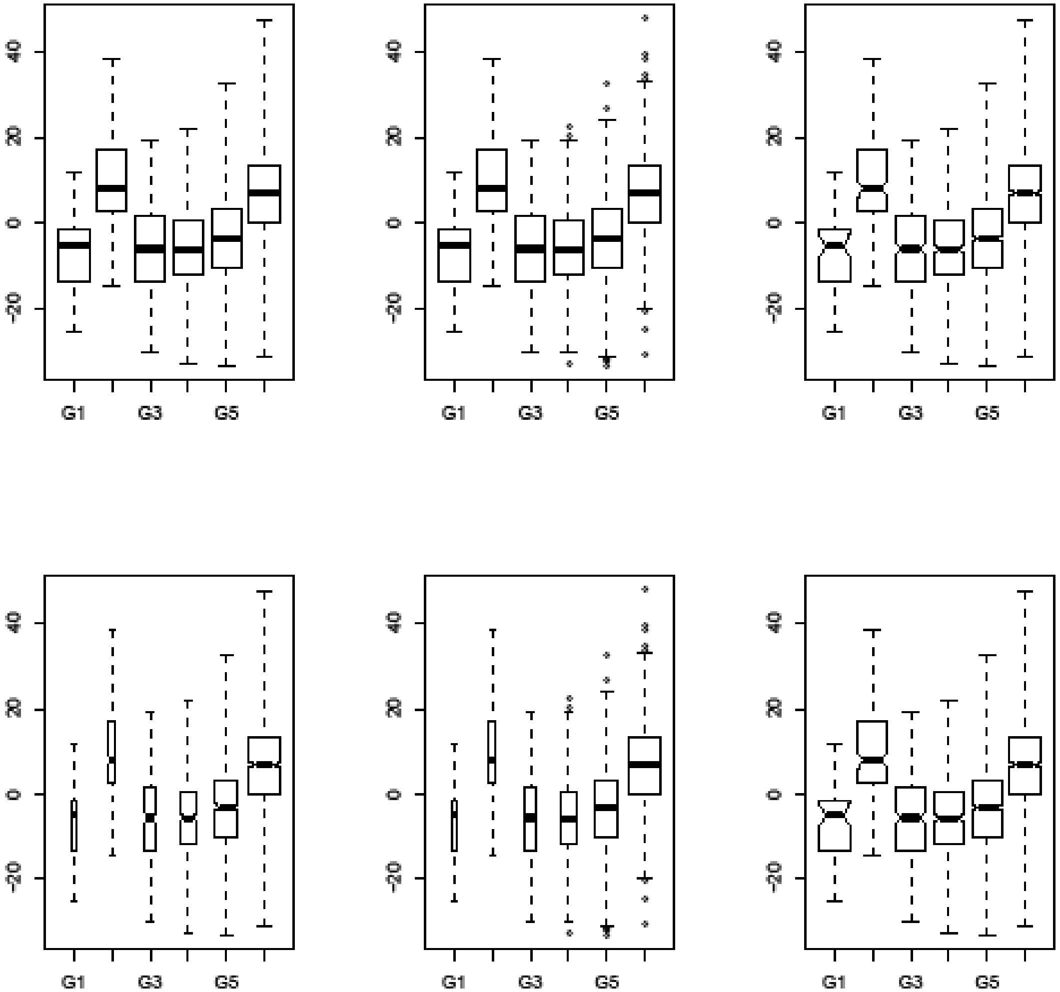
- Various types of Box Plots
  - Standard
  - Variable-width Box Plot
  - Notched Box Plot
  - Variable-width Notched Box Plot



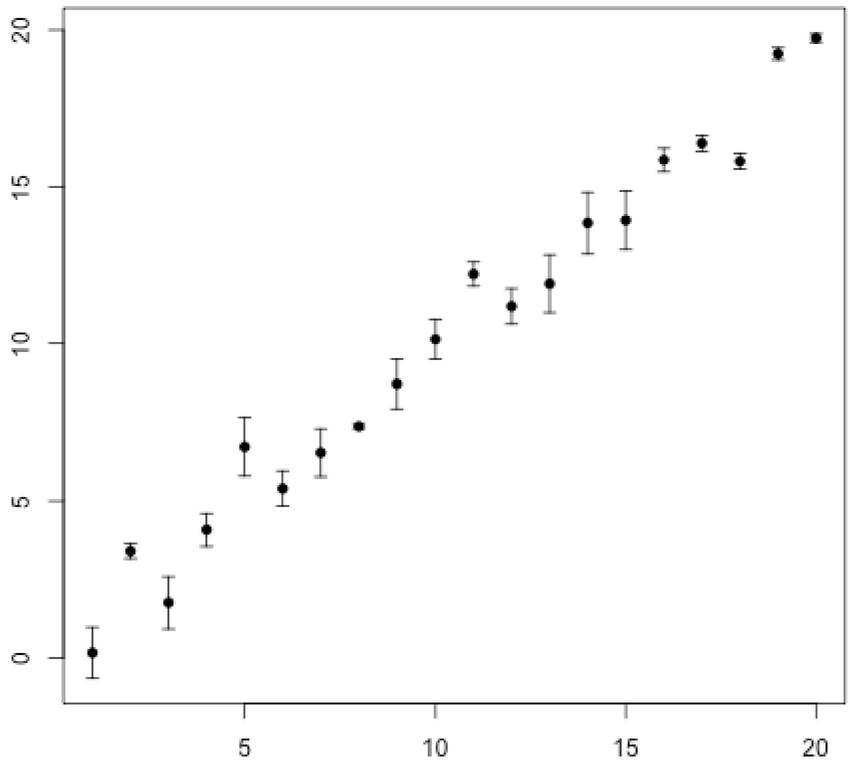
Box Plot



Box Plot



Mean with confidence interval



Summary

- Stochastic discrete-event simulation is a statistical experiment.
  - Purpose of statistical experiment: obtain estimates of the performance measures of the system.
  - Purpose of statistical analysis: acquire some assurance that these estimates are sufficiently precise.
- Distinguish simulation runs with respect to output analysis:
  - Terminating simulations and
  - Steady-state simulations.
- Steady-state output data are more difficult to analyze
  - Decisions: initial conditions and run length
  - Possible solutions to bias: deletion of data and increasing run length
- Statistical precision of point estimators are estimated by standard-error or confidence interval
- Method of independent replications was emphasized.
- Batch mean for a long run replication
- Art of data representation