Chapter 11
Output Analysis for a Single Model

Introduction

- Output analysis is the examination of data generated by a simulation. OA is used to:
  - predict the performance of a system or
  - compare the performance of 2 or more alternative system designs.
- Issues of output analysis:
  - The sequence of output variables $Y_1, Y_2, \ldots, Y_n$ may be autocorrelated. Therefore, the classical methods of statistics which assume independence are not directly application to output analysis.
  - Initial conditions of the system may influence the output data.

Types of Simulation wrt Output Analysis

- When analyzing simulation output data, we distinguish:
  - Terminating or transient simulations:
    - Runs for some duration of time, T_E, until a specified event (or set of events) happens. (Examples 11.1-11.3)
    - Depends on both the objectives of the simulation and the nature of the system.
  - Nonterminating or steady-state simulations
    - Runs continuously or over a long period of time.
    - Study steady-state, or long-run, properties of the system (not influenced by the initial conditions).
    - T_E is not determined by the nature of the problem (e.g. end of day or a certain period of time), rather it is set as one parameter in the simulation design.
    -

Stochastic Nature of Output Data

- Some of the model input variables are random variables and the model is an input-output transformation, it follows that the model output variables are random variables. Cases to discuss:

1. Output data from various runs are independent and identically distributed. Classical methods of statistics may be used because the data constitute a random sample.
   - point estimate and
   - estimation of error in the point estimate

Stochastic Nature of Output Data

2. The effects of correlation and initial conditions on the estimation of long-run mean measures of performance.
   - See Table 11.2. The sequence of batches is autocorrelated because all of the data are obtained from within one replication.
   - Example 11.9: Avoid direct statistical analysis of the within-replication output $\{D_i, i = 1, 2, ..\}$, because the sequence is, in general, a nonstationary autocorrelated stochastic process. In other words, $D_1, D_2, \ldots$, are not identically distributed. If D_i is small, D_{i+1} will tend to be relatively small.

Measures of Performance and Their Estimation

- It's desired to have a point estimate and an interval estimate. The later is a measure of the error in the point estimate.
- Simulation output data of the form $\{Y_1, Y_2, \ldots\}$ for estimating θ: discrete-time data.
- Simulation output data of the form $\{Y(t), 0 \le t \le T_E\}$ for estimating ϕ: continuous-time data, cause t is continuous valued.
- The parameter θ is an ordinary mean; ϕ is time-weighted mean.

Point Estimation

- θ hat = $1/n\ \Sigma^n Y_i$: sample mean based on a sample of size n.
- The point estimator θ hat is unbiased for θ hat if E(θ hat) = θ. But in general, E(θ hat) $\neq \theta$.
- E(θ hat) - θ is called the bias in the point estimator θ hat.
  - Examples of estimator include w and w_Q hat discussed in the queuing modeling, Y_i is the time spent in the system by customer i.
- ϕ hat = $\int^{TE} Y(t)dt$: time average of Y(t) over [0, T_E]
  - Examples include L hat and L_Q hat.

Interval Estimation

- Valid interval estimation requires a method of estimating the variance of the point estimator, θ hat or ϕ hat.
- σ^2 (θ hat) = var (θ hat) : true variance of a point estimator.
- σ^2 hat (θ hat): estimator of σ^2 (θ hat) based on the data {Y_1, Y_2, ..., Y_n}.
- t = (θ hat - θ) / σ^2 hat (θ hat) is approximately t-distributed with some degrees of freedom, f, if σ^2 hat (θ hat) is approximately unbiased.
- Confidence interval, 100(1-α)%, for θ is given by

  θ hat $\pm\ t_{\alpha/2,f}$ σ hat (θ hat) or

  θ hat - $t_{\alpha/2,f}$ σ hat (θ hat) $\leq \theta \leq$ θ hat + $t_{\alpha/2,f}$ σ hat (θ hat)
- One of the main problems in simulation output analysis is obtaining approximately unbiased estimate σ^2(θ hat), the variance of the point estimator.

Output Analysis for Terminating Simulations

- The goal is to estimate θ.
- The method used is called method of independent replications.
  - The simulation is repeated R times, each run using a different random-number stream and independently chosen initial conditions.
  - Y_{ri}: *i*th observation within replication r.
  - For fixed replication r, Y_{r1}, Y_{r2}, ..., is an autocorrelated sequence within replication r; but across different replications, Y_{ri} and Y_{si} are statistically independent.
  - θ_r hat = $1/n_r \Sigma^{nr} Y_{ri}$, r = 1, 2, ..., R
  - The R sample means are statistically independent and identically distributed. Thus, classical methods of confidence interval estimation can be applied.
  - Similarly, ϕ is defined by (11.6)

Statistics Review (again?)

- Y_1, ..., Y_n are statistically independent observations.
- Sample mean & sample variance S^2?
- The variance of θ hat is given by
  s^2 (θ hat) = s^2 / n
- An unbiased estimator of s^2 (θ hat), with f = n – 1 degrees of freedom, is provided by
  s^2 hat (θ hat) = S^2 / n
- Confidence interval: see (11.10)
- s^2 hat (θ) = S / √n, denoted by s.e.(θ hat), is called standard error of the input parameter θ hat.
  - A measure of the precision of a point estimator, or
  - The average deviation to be expected between the point estimator θ hat and the true mean θ.

Confidence-Interval Estimation for a Fixed Number of Replications

- If R independent replications are made,
  - The overall point estimate, θ hat, is computed by θ hat = 1/R ?^{R} θ_r hat
  - Estimate of the variance of θ hat is computed by s^2 hat (θ hat) = S^2 / n = 1/[(R-1)R] ?^{R} (θ_r hat - θ hat)^2
  - A 100(1-α)% confidence interval is also defined by (11.10)
  - As R increases, the standard error s hat (θ hat) tends to become smaller and approach zero.
- Similarly, if the output data are of the form $\{Y_r(t), 0 <= t <= T_E\}$, we use (11.17, 11.18)
- Example 11.10

Confidence Intervals with Specified Precision

- Half-length (h.l.) of a 100(1-α)% confidence interval for a mean θ, based on the t-distribution, is h.l. = $t_{\alpha/2, R-1}$ s hat (θ hat), where s hat (θ hat) = S / vR.
- If an error ε is specified, i.e., |θ - θ hat| < ε with high probability 1 - α., then we need a large sample size, R, to satisfy P(|θ - θ hat| < ε) = 1 - α. Steps:
  - Make R_0 independent replications, R_0 = 2, least 4 or 5 is recommended; 10 or more is desirable.
  - Obtain an initial estimate S_0^2 of the population variance s^2 based on R_0 replications. Thus,
    h.l. = $t_{\alpha/2, R-1} S_0$ / vR = $\varepsilon \Rightarrow$ R = $(t_{\alpha/2, R-1} S_0 / \varepsilon)^2$
  - $t_{\alpha/2, R-1} = z_{\alpha/2}$, an initial estimate for R is R = $(z_{\alpha/2} S_0 / \varepsilon)^2$.
    $t_{\alpha/2, R-1} \sim z_{\alpha/2}$ for large R.
  - After determining R, collect R – R_0 additional observations and form the 100(1-α)% confidence interval. See (11.23)
  - If the confidence interval is too large, the procedure may be repeated to determine an even larger sample size.
- Example 11.12.

Output Analysis for Steady-State Simulations

- The goal is to estimate a steady-state or long-run characteristics of the system. For steady-state, the measure of performance, θ, to be estimated is defined by $\theta = \lim_{n \to \infty} 1/n \,?^n Y_i$
- The value of θ is independent of the initial conditions.
- In practice, simulation stops after some number of observations, n, have been collected; or for some length of time T_E. The sample size n (or T_E) is a design choice; not determined by the nature of the problem. To choose simulation run length, we consider:
  - The bias in the point estimator due to initial conditions.
  - The desired precision of the input estimator.
  - Budget constraints.

Initialization Bias in Steady-State Simulations

- Initial conditions may be artificial or unrealistic.
- Methods to reduce the point-estimator bias include:
  - 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 them.
    - If the system doesn't exist, use any data on similar systems or build a simplified model that is mathematically solvable and collect data from it.
  - Divide each simulation run into two phases:
    - Initialization phase from time 0 to T_0; and
    - Data-collection phase from T_0 to $T_0 + T_E$.
    - See Figure 11.3.
    - Choice of T_0 (representative of steady-state behavior) and T_E (long enough) is important.
    - Example 11.14.

How Much Data to Delete?

- For the second method, unfortunately, there is no widely accepted, objective and proven technique to determine how much data to delete to reduce initialization bias to a negligible level.
- Some points to consider:
  - Ensemble averages will reveal a smoother trend as R increases. Since each ensemble average is the sample mean of i.i.d. observations, a confidence interval based on the t-distribution can be placed around each point. See figure 11.6.
  - Use a moving average rather than the original ensemble averages. In a moving average, each plotted point is actually the average of several adjacent ensemble averages.
  - Cumulative averages should only be used if it's not feasible to computer ensemble averages.
  - The more correlation present in simulation data, the longer it takes to $Y_{.j}$ bar to approach steady state.
  - Different performance measures may approach steady-state at different rates. It is important to examine each performance measure individually.
  - No silver bullet.

Replication Method for Steady-State Simulations

- The method of independent replications can be used to estimate point-estimator variability and to construct a confidence interval.
- How? Make R replications, initializing and deleting from each one the same way.
- NOTE: bias is not affected by R; it is affected only by deleting more data (i.e., increasing T_0) or extending the length of each run (increasing T_E). If significant bias remains in the point estimator and a large # of replications are used to reduce point estimator variability, the result can be misleading.

Replication Method for Steady-State Simulations (2)

- When using the replication method, each replication is regarded as a single sample for the purpose of estimating θ.
- For replication r, define

  $Yr.(n, d)$ bar = $1 / (n - d) \sum_{j=d+1}^{n} Y_{rj}$

  as the sample mean of all nondeleted observations in replication r.
- See Table 11.7 for $Y_{.i}$, $Y_{r.}$ (n, d), and Y..(n,d). Also $Y_{r.}$ bar and Y.. bar for abbreviation.
- E[Y..(n, d) bar] = $\theta_{n,d}$
- To estimate the standard error of Y.. Bar, first compute the sample variance S^2 as shown in (11.40)
- Standard error is given by s.e.(Y.. Bar) = $S / \sqrt{R}$.
- A $100(1-\alpha)\%$ confidence interval for θ, based on the t-distribution is given by (11.42)
- Examples 11.15 & 11.16

Sample Size in Steady-State Simulations

- In a steady-state simulation, a specified precision may be achieved either by increasing the # of replications (R) or by increasing the run length (T_E).
- Example 11.17 shows the first approach (by increasing R)
- An alternative to increasing R is to increase total run length $T_0 + T_E$ within each replication.
  - If $(R - R_0)$ additional replications are needed, then an alternative is to increase the run length $(T_0 + T_E)$ in the same proportion (R/R_0) to a new run length $(R/R_0)(T_0+T_E)$
  - Additional data will be deleted, from time 0 to time $(R/R_0)\ T_0$, and more data will be used to compute the point estimates.
  - Figure 11.8 illustrates the approach.
  - Advantage: bias may be further reduced
  - Disadvantage: it may need to save the state of the model at time)(T_0+T_E), and to be able to restart the model and run it for the additional required time.
  - Example 11.18

Batch Means for Interval Estimation in Steady-State Simulations

- One disadvantage of the replication method is that data must be deleted on each replication.
- To reduce the issue, we may use an experiment design that is based on a single, long replication. But one disadvantage of this is the problem to compute the standard error of the sample mean. Also, the data are dependent and usual estimator is biased.
- Batch means is a method developed to solve this problem by dividing the output data from one replication (after deletion) into a few large batches, and then treating the means of these batches as if they were independent.
- Compute the batch means (Y_j bar) based on the form of the raw output data:
  - Continuous:
  - Discrete:
  - The variance of the sample mean: (11.43)
  - Y_1 bar, Y_2 bar, ..., Y_k bar are not independent; but if the batch size is sufficiently large, successive batch means will be approximately independent.

Batch Means and Batch Size

- What is an acceptable batch size m?
  - No widely accepted method.
- General guidelines:
  - For a fixed sample size, 10 <= m <= 30 should be used.
  - Lag-1 autocorrelation ρ_1 = corr (Y_j bar, Y_{j+1} bar) is usually studied to assess the dependence between batch means.
  - Estimate lag-1 autocorrelation from a large number of batch means based on a smaller batch size (100 <= k <= 400). (Rebatching into a larger batch size is to get smaller autocorrelation).
  - If the total sample size is to be chosen sequentially, then it is helpful to allow the batch size and # of batches to grow as the run length increases.

Summary

- Before sound conclusions can be drawn on the basis of the simulation-generated output data, a proper statistical analysis is required.
- The simulation experiment: estimate the performance measures of the system.
- The statistical analysis: acquire some assurance that these estimates are sufficiently precise for the proposed use of the model.
- Terminating simulations vs. steady-state simulations.
  - Initial conditions and the choice of run length.
- Standard error or a confidence interval can be used to measure the precision of point estimators.
  - Use the method of independent replications to generate statistically independent observations and apply standard statistically methods.
  - Batch means.

Batch Means and Batch Size

- General strategy:
  1. Obtain output data from a single replication and delete as appropriate.
  2. Compute the batch means (100 <= k <= 400) and estimate the sample lag-1 autocorrelation:
  3. Check the correlation to see if it's sufficiently small.
     a. If <= 0.2, the rebatch the data into 30 <= k <= 40 batches and form a confidence interval using k – 1 degrees of freedom for t-test and estimate the variance.
     b. If > 0.2, then extend the replication and repeat step 2.
  4. Check on the confidence interval, examine the batch means for independence using : C = ??
- Example 11.19