MEM6810 Engineering Systems Modeling and Simulation
工程系统建模与仿真

Theory Analysis

Lecture 6: Verification and Validation of Simulation Models

SHEN Haihui 沈海辉

Sino-US Global Logistics Institute
Shanghai Jiao Tong University

shenhaihui.github.io/teaching/mem6810f
shenhaihui@sjtu.edu.cn

Spring 2024 (full-time)

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY | 董浩云智能制造与服务管理研究院 CY TUNG Institute of Intelligent Manufacturing and Service Management (中美物流研究院) (Sino-US Global Logistics Institute)

Contents

1 Introduction

2 Verification of Simulation Models

3 Validation of Simulation Models
- Build High Face Validity
- Validate Model Assumptions
- Compare Input-Output Transformation

- One of the most important and difficult tasks facing a model developer is the *verification* and *validation* of the simulation model.
- **Verification** (检验): building the model correctly.
  - Is the desired model implemented correctly in computer (e.g., simulation software, programming language)?
- **Validation** (验证): building the correct model.
  - Is the simulation model an accurate representation of the real system?

Introduction

- Repeated steps in model building:
  - Observe the real system.
    – collecting data
    – consult domain experts
  - Construct a conceptual model.
    – assumptions about the system
    – hypotheses about the parameter values
  - Implement an operational model.
    – using simulation software or programming language

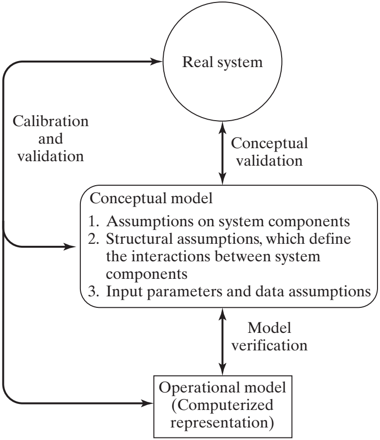


Figure: Model Building Process (*from Banks et al. (2010)*)

Introduction

- Verification and validation are integral parts of model development.
    - To produce a model that represents true behavior closely enough for decision-making purposes.
    - To increase the model's credibility to an acceptable level.
- It is the job of the model developer to work closely with the end users throughout the period of development.

Verification of Simulation Models

- The purpose is to ensure that the conceptual model is reflected accurately in the computerized representation.
- Many common-sense suggestions, for example:
  - Have someone else check the model.
  - Make a flow diagram that includes each logically possible action a system can take when an event occurs.
  - Closely examine the model output for reasonableness under a variety of input parameter settings. (Often overlooked!)
    – The model should be run, when possible, under simplifying assumptions for which its true measures can be analytically solved.
  - Debug a discrete-event simulation program using a "trace".
    – In a trace, the **state** of the simulated system is displayed just after each event occurs and is compared with hand calculation.
    – Desirable to observe the simulated system under "extreme" conditions.
  - If the operational model is animated, verify that what is seen in the animation imitates the actual system.

Verification of Simulation Models

- Example of examining the model output for reasonableness: A model of a complex queueing network.
  - Suppose the interested quantity is the sojourn time of customers.
  - However, it is important and helpful to collect many other quantities, for example:
    - the current number of customers in each station;
    - the current server utilization of each station.
  - If the utilization of one station is unreasonably low (or high), it may indicate some possible mistakes in parameters and/or model logic.
  - It may be possible to set the input and/or logic (e.g., Poisson arrival, exponential service time, FCFS) so that the complex queueing network becomes a Jackson queueing network.
    - Compare the analytically solved limiting distribution and L, W, L_Q, W_Q with the simulation results.

Validation of Simulation Models

- Validation is the process of determining whether a simulation model is an accurate representation of the actual system, *for the particular purpose of the study*.
  - *Subjective tests* usually involve people, who are knowledgeable about the system and making judgments about the model.
  - *Objective tests* require to compare the data on the system's behavior with the corresponding data produced by the model.
  - A model that is valid for one purpose may not be for another.

- Calibration is the iterative process of making adjustments (or even major changes) to the model until the outputs from simulation model agree closely with those from actual system.
  - Use some sets of data for calibration and other independent sets for "final" validation.

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Validation of Simulation Models

- No model is ever a perfect representation of the actual system.
  - The modeler must weigh the possible, but not guaranteed, increase in model accuracy versus the cost of increased validation effort.
  - The level of model detail should be consistent with the type of data available.
- Widely used approaches in the validation process:
  - Build a model that has high face validity (表面效度).
  - Validate model assumptions.
  - Compare the model input-output transformations with the actual system's data.

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

- The first goal of simulation modeler: To construct a model that appears reasonable on its face to the users who are knowledgeable about the real system.
- The potential users of a model should be involved:
  - to ensure that a high degree of realism is built into the model;
  - to check a model's face validity e.g., whether the model behaves in the expected way when one or more input variables is changed?

    – Example: In most queueing systems, if the arrival rate of customers increases, it would be expected that server utilization, queue length, and waiting time will increase (although by how much might be unknown).

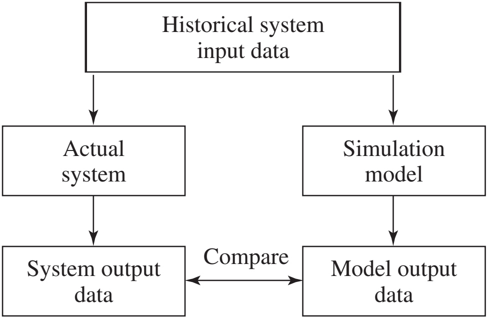
- General classes of model assumptions:
  - Structural assumptions: how the system operates.
  - Data assumptions: reliability of data and their statistical analysis.
- Example: Simulation of customer queueing and service facility in a bank.
  - Structural assumptions: customers wait in one line vs. many lines, FCFS vs. priority.
  - Data assumptions: interarrival times of customers (peak or slack periods), service times (commercial or personal accounts).
    - Verify data reliability with bank managers.
    - Test homogeneity (when combining data) and correlation.
    - Test the goodness of fit for hypothesized distribution.

- Validate the model's ability to predict future behavior.
  - The model is viewed as an input-output transformation.
  - A necessary condition is that some version of the system under study exists.

- Use historical data that have been reserved for validation purposes **only**.
  - if some data sets have been used to develop and calibrate the model, use separate data sets as the final validation test.

- Use the main responses of interest as the primary criteria for validating a model.
  - If the model is used later for a different purpose, it should be revalidated in terms of the new responses of interest.

- Input-output validation using historical input data.



- Example: Simulation of a store with 2 parallel servers and 2 waiting lines.
  - It opens at 9 AM and closes at 5 PM every working day.†
  - The primary interested response (i.e., measure) is the average waiting time of customers during peak hour (11 AM to 1 PM).
  - The question we want to ask is, how much waiting time can be reduced by adding one more staff during peak hour.
  - For K similar working days, we record every customer's time when he arrives, time when he gets service, and time when he leaves.
  - Simulation model is built based on some assumption and simplification.
    - In actual system, customers may switch waiting lines spontaneously.
    - In simulation model, we assume that customers choose the shortest line when they come and will not change later.
  - Before using the simulation model to answer our question, we first need to validate the model.

† It is known as a terminating simulation (see more details in Lec 7).

- For each day $i = 1, \dots, K$, we can compute every customer's arrival time and service time (A_{ij}, S_{ij}), and his waiting time W_{ij}, $j = 1, 2, \dots$. We can also get the average waiting time during peak hour, W_i.
- Suppose the data are reliable, and we can model the arrival (perhaps by a nonhomogeneous Poisson process) and service time with satisfying goodness of fit.
- The remaining issue is on the validity of the structural assumption, i.e., the queue behavior.

- We conduct input-output validation using historical input data.
  - Run the simulation model for K replications, and for each replication $i = 1, \ldots, K$, historical data $\{(A_{ij}, S_{ij}) : j = 1, 2, \ldots\}$ were fed into the model as input.
  - The simulated average waiting time during peak hour is Y_i, $i = 1, \ldots, K$.
  - Compare system output data $\{W_1, \ldots, W_K\}$ with model output data $\{Y_1, \ldots, Y_K\}$.
  - We are interested in knowing if the two sets of output data are equal statistically.

Input Data Set	*System Output* W_i	*Model Output* Y_i	*Observed Difference* d_i	*Squared Deviation from Mean* $(d_i - \bar{d})^2$
1	W_1	Y_1	$d_1 = Y_1 - W_1$	$(d_1 - \bar{d})^2$
2	W_2	Y_2	$d_2 = Y_2 - W_2$	$(d_2 - \bar{d})^2$
3	W_3	Y_3	$d_3 = Y_3 - W_3$	$(d_3 - \bar{d})^2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
K	W_K	Y_K	$d_K = Y_K - W_K$	$(d_K - \bar{d})^2$
			$\bar{d} = \frac{1}{K}\sum_{i=1}^{K} d_i$	$S_d^2 = \frac{1}{K-1}\sum_{i=1}^{K}(d_i - \bar{d})^2$

- Let $D := Y - W$ be a random variable of the difference.
  - If the K input data sets are fairly homogeneous, then $d_1, \ldots, d_K$ can be considered as iid realizations of D.
  - It is reasonable to assume that D follows normal distribution.
    - Each Y and W is an average over customers.
    - By the Central Limit Theorem, Y, W, and D are approximately normally distributed when the customer number is large.

- $D \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are unknown. Our goal is to check if $\mu = 0$.
- $d_1, \ldots, d_K$ are iid observations of D, from which we get the sample mean $\bar{d}$ and sample variance S_d^2.
- $\bar{D} \sim \mathcal{N}(\mu, \sigma^2/K)$, and $\frac{\bar{D}-\mu}{S_d/\sqrt{K}}$ follows t distribution with $K-1$ degrees of freedom.
- Consider hypothesis test: H_0: $\mu = 0$ vs H_1: $\mu \neq 0$.
  - Use the t test statistic $T = \frac{\bar{D}-0}{S_d/\sqrt{K}}$, whose observed value is $t = \frac{\bar{d}}{S_d/\sqrt{K}}$.
  - The p-value is $\mathbb{P}(|T| \geq |t|) = 2\,\mathbb{P}(T \geq |t|)$.
  - Under significance level α, reject H_0 if p-value $< \alpha$, or $|t| > t_{K-1,1-\alpha/2}$, where $t_{K-1,1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of t distribution with $K-1$ degrees of freedom.

- To validate the structural assumptions of the simulation model, it is usually more preferable to compare the model and actual system under the same historical input.
  - We could have chosen to generate the arrival times and service times from the fitted input models in the simulation. Let Y' denote the random variable of average waiting time.
  - However, $\mathrm{Var}(Y' - W) > \mathrm{Var}(Y - W)$, since $\mathrm{Cov}(Y', W) = 0$ and $\mathrm{Cov}(Y, W) > 0$.
- Using historical input in simulation is only recommended for model validation, not for making *production* run.
- One may also consider to construct a confidence interval for μ to test H_0.
- When only one data set (instead of K) is available,† one needs to use some time-series approaches to compare model output data with system output data.

† A common situation in nonterminating simulation (see more details in Lec 7).

- A Turing test can be used in addition to statistical test, or when no statistical test is readily applicable.
- It utilizes personnel knowledgeable about system behavior.
- For example:
  - Present 10 system performance reports to a manager of the system. Five of them are from the real system and the rest are "fake" reports based on simulation output data.
  - If the manager identifies a substantial number of the fake reports, interview the manager to get information for model improvement.
  - If the manager cannot distinguish between fake and real reports with consistency, conclude that the test gives no evidence of model inadequacy.