Input Analysis for Stochastic Simulations

David Fernando Muñoz
Instituto Tecnológico Autónomo de México, México

INTRODUCTION

Simulation has been recognized as a powerful tool for forecasting, risk assessment, animation and illustration of the evolution of a system in many areas (see, e.g., Kelton et al., 2011), including business analytics. When uncertainty on the behavior of some components of the simulation model is present, these *random components* of a stochastic simulation are modeled through the use of probability distributions and/or stochastic processes that drive the simulation experiments.

In order to illustrate how the concept of a random component was introduced in stochastic simulations, let us suppose that we are interested in simulating the congestion at an automatic teller machine (ATM) during a particular lapse of time, say during the time interval $[0, t]$. To that end, we may assume that the ATM is empty and idle at time 0. Let A_1 be the arrival time of customer 1, and A_i be the inter-arrival time between customer $i - 1$ and customer i (for $i = 2, 3, \ldots$). Further, let S_i be the service-time requirement of customer i at the ATM (for $i = 1, 2, \ldots$). Then, the waiting time in queue of customer 1 is $W_q(1) = 0$, and the waiting times for customers $i = 2, 3, \ldots$ can be obtained using Lindley's recurrence (Lindley, 1952):

$$W_q(i) = \max \left\{ W_q(i-1) + S_{i-1} - A_i, 0 \right\}$$

(1)

The congestion at the ATM can be simulated by using Equation (1) (e.g., on a spreadsheet) to produce clients' waiting times until the last client's departure exceeds time t. However, in order to produce the simulation output given by Equation

(1), we need to produce two streams of random inputs: $A_1, A_2, \ldots$, and $S_1, S_2, \ldots$. These streams are usually (although not always) produced by assuming that the inter-arrival times are independent and identically distributed (i.i.d.) random variables from a density function $f_A(x, \theta_A)$, and the service times are i.i.d. random variables from a density function $f_S(x, \theta_S)$, where θ_A and θ_S are known parameters. Then, standard methods for random variate generation (see, e.g., Law 2007) can be applied to obtain the required streams $A_1, A_2, \ldots$, and $S_1, S_2, \ldots$.

We remark that a random component (also called random input) of a stochastic simulation is a sequence $U_1, U_2, \ldots$ of random quantities (may be multivariate) that are needed as input to the simulation. When the U_i's are assumed to be i.i.d., a random component is identified by the corresponding probability distribution, which is typically assumed to be a member of a parametric family. Input analysis for stochastic simulations is concerned with the appropriate modeling of the random components that are considered in a simulation model, and is particularly relevant when data from the system under study is available. For the case of our ATM simulation, it will be worthwhile to obtain observations from the real system, for instance, a sample $x_1, x_2, \ldots, x_n$ of observations from the customers' inter-arrival times, and then choose the input distribution $f_A(x, \theta_A)$ that best fits the data $x_1, x_2, \ldots, x_n$.

Concepts of probability and statistics, including random variables and related definitions such as cumulative distribution function (c.d.f.), probability mass function (p.m.f.) for discrete random variables and probability density function (p.d.f.)

for continuous random variables, conditional probability and independence, point estimation and hypothesis testing are explained, e.g., in Walpole et al. (2012). At this point, it will be convenient to introduce the convention that a "family of distributions $f(x,\theta)$" will refer to the corresponding p.m.f. when the distribution is discrete, and to the corresponding p.d.f., when the distribution is continuous.

In this chapter we discuss the main techniques for input analysis that are proposed for stochastic simulation. In the next Section we discuss the main families of distributions that that are useful to model a univariate random component as well as the use of software for input analysis that is distributed along with this chapter. In Section *Advanced Input Modeling,* we discuss input modeling techniques that are not considered in typical commercial software for input analysis, including the fitting of multimodal, correlated and time-dependent input from available data, and the incorporation of parameter uncertainty that is induced from the estimation of the input model's parameters. In the last section, we present our conclusions.

STANDARD INPUT MODELING

The conventional approach for input modeling of a (univariate) random component consists of identifying a standard family of distributions that best fits our needs, under the assumption that the required random component consists of i.i.d. observations from the identified family of distributions. A good selection of a standard family of distributions is made when the properties of the corresponding family are appropriate for the experiment that we intend to simulate, and/or the family of distributions is the one that best fits a sample $X_1,\ldots,X_n$ of (real) observations from the random component. For the particular case where no observations are available, a common practice is to assume a triangular distribution (see, e.g., Law, 2007) and define the parameters based on

the opinion of an expert. For the case where a sample $X_1,\ldots,X_n$ is available, the family of distributions $f(x,\theta)$ is identified by applying the following three steps to each candidate family: (i) find a good estimator for parameter θ, (ii) group the observations and (visually) compare a plot of the relative frequencies versus a plot of the probabilities corresponding to $f(x,\theta)$, and (iii) compute goodness-of-fit statistics. The software *Simple Analyzer* that is distributed along with this chapter can be used to accomplish these three steps for the most common families of distributions.

The standard families of continuous distributions that are considered in the Simple Analyzer are: Uniform, Triangular, Exponential, Weibull, Gamma, Normal, Lognormal, and Beta. The first two (Uniform and Triangular) are suitable for the case where no sample data are available, and the rest can be used when observations $x_1,\ldots,x_n$ from the random component are available. For a given family of continuous distributions, the parameters that identify the specific family member can usually be classified as being one of three basic types: location, scale or shape parameters (see Law, 2007 for details).

The standard families of discrete distributions that are considered in the Input Analyzer are: Bernoulli, Binomial, Negative Binomial and Poisson.

Using the Simple Analyzer

The reader can download an Excel file named *SimpleAnalyzer.xls* from the Web page http://ciep. itam.mx/~davidm/sofdop.htm. In order to load the libraries that are required to run the macros in this file, the User must have previously installed the "Random Number Generators" (from the same Web page). In the worksheets named *Continuous* and *Discrete*, we may introduce observations $x_1,\ldots,x_n$ from a random component, and the option buttons included in each sheet allow us to fit each of the standard family of distributions discussed in this chapter; worksheet *Continuous*

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/input-analysis-for-stochastic-simulations/107320

Related Content

Data Mining for Health Care Professionals: MBA Course Projects Resulting in Hospital Improvements
Alan Olinskyand Phyllis A. Schumacher (2010). *International Journal of Business Intelligence Research (pp. 30-41).*
www.irma-international.org/article/data-mining-health-care-professionals/43680

EA Knowledge for ACE Deployment
Jay Ramanathanand Rajiv Ramnath (2009). *Co-Engineering Applications and Adaptive Business Technologies in Practice: Enterprise Service Ontologies, Models, and Frameworks (pp. 159-219).*
www.irma-international.org/chapter/knowledge-ace-deployment/6594

QoS Oriented Enhancement based on the Analysis of Dynamic Job Scheduling in HPC
Reshmi Raveendranand D. Shanthi Saravanan (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1800-1817).*
www.irma-international.org/chapter/qos-oriented-enhancement-based-on-the-analysis-of-dynamic-job-scheduling-in-hpc/142702

A Timeline Optimization Approach of Green Requirement Engineering Framework for Efficient Categorized Natural Language Documents in Non-Functional Requirements
K. Mahalakshmi, Udayakumar Allimuthu, L Jayakumarand Ankur Dumka (2021). *International Journal of Business Analytics (pp. 21-37).*
www.irma-international.org/article/a-timeline-optimization-approach-of-green-requirement-engineering-framework-for-efficient-categorized-natural-language-documents-in-non-functional-requirements/269485

Utility Functions and Risk Attitudes in Decision Analysis
Candice H. Huynhand Wenting Pan (2014). *Encyclopedia of Business Analytics and Optimization (pp. 2616-2629).*
www.irma-international.org/chapter/utility-functions-and-risk-attitudes-in-decision-analysis/107441