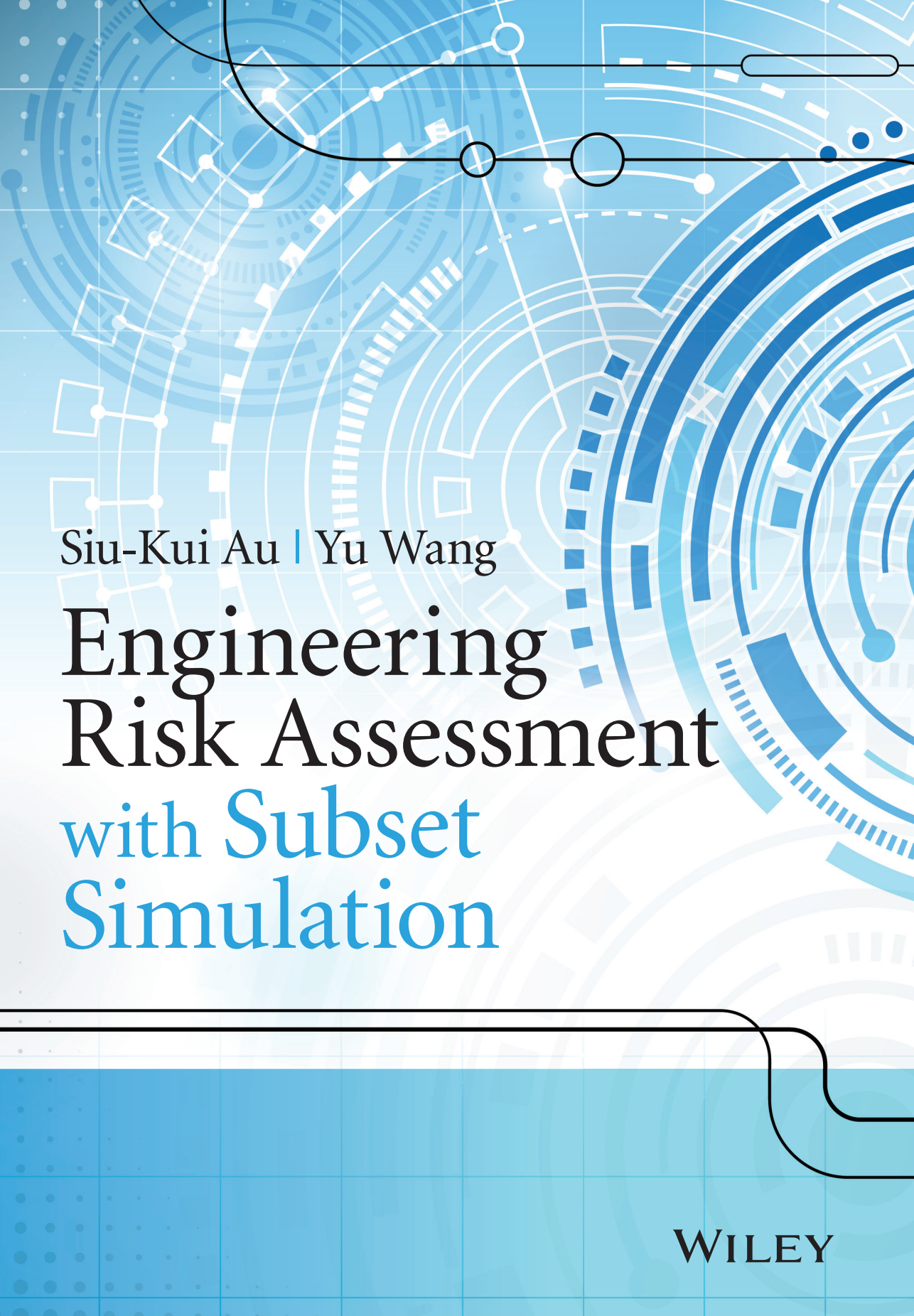




Siu-Kui Au | Yu Wang

Engineering Risk Assessment with Subset Simulation

WILEY

ENGINEERING RISK ASSESSMENT WITH SUBSET SIMULATION

ENGINEERING RISK ASSESSMENT WITH SUBSET SIMULATION

Siu-Kui Au

University of Liverpool, UK

Yu Wang

City University of Hong Kong, China

WILEY

This edition first published 2014
© 2014 John Wiley & Sons Singapore Pte. Ltd.

Registered office

John Wiley & Sons Singapore Pte. Ltd., 1 Fusionopolis Walk, #07-01 Solaris South Tower, Singapore 138628.

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as expressly permitted by law, without either the prior written permission of the Publisher, or authorization through payment of the appropriate photocopy fee to the Copyright Clearance Center. Requests for permission should be addressed to the Publisher, John Wiley & Sons Singapore Pte. Ltd., 1 Fusionopolis Walk, #07-01 Solaris South Tower, Singapore 138628, tel: 65-66438000, fax: 65-66438008, email: enquiry@wiley.com.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data has been applied for.

A catalogue record for this book is available from the British Library.

ISBN: 978-1-118-39804-3

Set in 10/12pt Times by Aptara Inc., New Delhi, India

To our families
To Professor Gerhart I. Schuëller

Contents

About the Authors	xiii
Preface	xv
Acknowledgements	xvii
Nomenclature	xix
1 Introduction	1
1.1 Formulation	2
1.2 Context	5
1.3 Extreme Value Theory	5
1.4 Exclusion	6
1.5 Organization of this Book	7
1.6 Remarks on the Use of Risk Analysis	7
1.7 Conventions	8
References	8
2 A Line of Thought	9
2.1 Numerical Integration	10
2.2 Perturbation	10
2.3 Gaussian Approximation	12
2.3.1 <i>Single Design Point</i>	12
2.3.2 <i>Multiple Design Points</i>	14
2.4 First/Second-Order Reliability Method	14
2.4.1 <i>Context</i>	15
2.4.2 <i>Design Point</i>	16
2.4.3 <i>FORM</i>	17
2.4.4 <i>SORM</i>	18
2.4.5 <i>Connection with Gaussian Approximation</i>	22
2.5 Direct Monte Carlo	24
2.5.1 <i>Unbiasedness</i>	25
2.5.2 <i>Mean-Square Convergence</i>	25
2.5.3 <i>Asymptotic Distribution (Central Limit Theorem)</i>	28
2.5.4 <i>Almost Sure Convergence (Strong Law of Large Numbers)</i>	31
2.5.5 <i>Failure Probability Estimation</i>	32

2.5.6	<i>CCDF Perspective</i>	34
2.5.7	<i>Rare Event Problems</i>	38
2.5.8	<i>Variance Reduction by Conditioning</i>	41
2.6	Importance Sampling	44
2.6.1	<i>Optimal Sampling Density</i>	45
2.6.2	<i>Failure Probability Estimation</i>	45
2.6.3	<i>Shifting Distribution</i>	46
2.6.4	<i>Benefits and Side-Effects</i>	48
2.6.5	<i>Bias</i>	50
2.6.6	<i>Curse of Dimension</i>	53
2.6.7	<i>CCDF Perspective</i>	56
2.7	Subset Simulation	58
2.8	Remarks on Reliability Methods	60
2A.1	Appendix: Laplace Type Integrals	61
	References	62
3	Simulation of Standard Random Variable and Process	65
3.1	Pseudo-Random Number	65
3.2	Inversion Principle	66
3.2.1	<i>Continuous Random Variable</i>	67
3.2.2	<i>Discrete Random Variables</i>	67
3.3	Mixing Principle	68
3.4	Rejection Principle	69
3.4.1	<i>Acceptance Probability</i>	71
3.5	Samples of Standard Distribution	72
3.6	Dependent Gaussian Variables	78
3.6.1	<i>Cholesky Factorization</i>	78
3.6.2	<i>Eigenvector Factorization</i>	81
3.7	Dependent Non-Gaussian Variables	83
3.7.1	<i>Nataf Transformation</i>	83
3.7.2	<i>Copula</i>	87
3.8	Correlation through Constraint	89
3.8.1	<i>Uniform in Sphere</i>	89
3.8.2	<i>Gaussian on Hyper-plane</i>	92
3.9	Stationary Gaussian Process	95
3.9.1	<i>Autocorrelation Function and Power Spectral Density</i>	95
3.9.2	<i>Discrete-Time Process</i>	99
3.9.3	<i>Sample Autocorrelation Function and Periodogram</i>	100
3.9.4	<i>Time Domain Representation</i>	101
3.9.5	<i>The ARMA Process</i>	103
3.9.6	<i>Frequency Domain Representation</i>	108
3.9.7	<i>Remarks</i>	115
3A.1	Appendix: Variance of Linear System Driven by White Noise	115
3A.2	Appendix: Verification of Spectral Formula	117
	References	118

4	Markov Chain Monte Carlo	119
4.1	Problem Context	119
4.2	Metropolis Algorithm	122
4.2.1	<i>Proposal PDF</i>	123
4.2.2	<i>Statistical Properties</i>	123
4.2.3	<i>Detailed Balance</i>	128
4.2.4	<i>Biased Rejection</i>	132
4.2.5	<i>Reversible Chain</i>	134
4.3	Metropolis–Hastings Algorithm	134
4.3.1	<i>Detailed Balance</i>	135
4.3.2	<i>Independent Proposal and Importance Sampling</i>	135
4.4	Statistical Estimation	137
4.4.1	<i>Properties of Estimator</i>	137
4.4.2	<i>Chain Correlation</i>	139
4.4.3	<i>Ergodicity</i>	143
4.5	Generation of Conditional Samples	148
4.5.1	<i>Curse of Dimension</i>	149
4.5.2	<i>Independent Component MCMC</i>	152
	References	155
5	Subset Simulation	157
5.1	Standard Algorithm	157
5.1.1	<i>Simulation Level 0 (Direct Monte Carlo)</i>	158
5.1.2	<i>Simulation Level $i = 1, \dots, m - 1$ (MCMC)</i>	159
5.2	Understanding the Algorithm	160
5.2.1	<i>Direct Monte Carlo Indispensible</i>	160
5.2.2	<i>Rare Regime Explored by MCMC</i>	161
5.2.3	<i>Stationary Markov Chain from the Start</i>	161
5.2.4	<i>Multiple Chains</i>	161
5.2.5	<i>Seeds Discarded</i>	162
5.2.6	<i>CCDF Perspective</i>	162
5.2.7	<i>Repeated Samples</i>	162
5.2.8	<i>Uniform Conditional Probabilities</i>	163
5.3	Error Assessment in a Single Run	166
5.3.1	<i>Heuristic Argument</i>	167
5.3.2	<i>Efficiency Over Direct Monte Carlo</i>	169
5.4	Implementation Issues	173
5.4.1	<i>Proposal Distribution</i>	173
5.4.2	<i>Ergodicity</i>	173
5.4.3	<i>Generalizations</i>	174
5.4.4	<i>Level Probability</i>	175
5.5	Analysis of Statistical Properties	179
5.5.1	<i>Random Intervals</i>	180
5.5.2	<i>Random CCDF Values</i>	181
5.5.3	<i>Summary of Results</i>	182
5.5.4	<i>Expectation</i>	183

5.5.5	<i>Variance</i>	185
5.6	Auxiliary Response	190
5.6.1	<i>Statistical Properties</i>	192
5.6.2	<i>Design of Driving Response</i>	194
5.7	Black Swan Events	195
5.7.1	<i>Diagnosis</i>	197
5.8	Applications	199
5.9	Variants	201
	References	202
6	Analysis Using Conditional Failure Samples	205
6.1	Probabilistic Failure Analysis	206
6.2	Uncertain Parameter Sensitivity	207
6.3	Conditional Samples from Direct Monte Carlo	208
6.3.1	<i>Conditional Expectation</i>	208
6.3.2	<i>Parameter Sensitivity</i>	210
6.4	Conditional Samples from Subset Simulation	216
6.4.1	<i>Sample Partitioning</i>	217
6.4.2	<i>Conditioning Structure</i>	219
6.4.3	<i>Conditional Expectation</i>	220
6.4.4	<i>Parameter Sensitivity</i>	224
	References	231
7	Spreadsheet Implementation	233
7.1	Microsoft Excel and VBA	233
7.1.1	<i>Excel Spreadsheet</i>	234
7.1.2	<i>Illustrative Example – Polynomial Function</i>	236
7.1.3	<i>Visual Basic for Applications (VBA)</i>	242
7.1.4	<i>VBA User-Defined Functions</i>	245
7.1.5	<i>VBA Subroutines</i>	247
7.1.6	<i>Macro Recorder</i>	251
7.2	Software Package UPSS	255
7.2.1	<i>Installation in Excel 2003</i>	255
7.2.2	<i>Installation in Excel 2010</i>	258
7.2.3	<i>Software Context</i>	260
7.2.4	<i>Deterministic System Modeling</i>	261
7.2.5	<i>Uncertainty Modeling</i>	262
7.2.6	<i>Uncertainty Propagation</i>	262
7.2.7	<i>Pre-Processing Tools</i>	265
7.2.8	<i>Post-Processing Tools</i>	268
7.3	Tutorial Example – Polynomial Function	269
7.3.1	<i>Deterministic System Modeling</i>	270
7.3.2	<i>Uncertainty Modeling</i>	270
7.3.3	<i>Uncertainty Propagation</i>	272
7.3.4	<i>Direct Monte Carlo</i>	274
7.3.5	<i>Subset Simulation</i>	275

7.4	Tutorial Example – Slope Stability	278
7.4.1	<i>Problem Context*</i>	278
7.4.2	<i>Deterministic System Modeling</i>	279
7.4.3	<i>Uncertainty Modeling</i>	279
7.4.4	<i>Histogram Tool</i>	281
7.4.5	<i>Uncertainty Propagation</i>	282
7.4.6	<i>CCDF of Driving Variable</i>	286
7.4.7	<i>Auxiliary Variable</i>	286
7.5	Tutorial Example – Portal Frame	288
7.5.1	<i>Problem Context*</i>	289
7.5.2	<i>Deterministic System Modeling</i>	290
7.5.3	<i>Uncertainty Modeling</i>	291
7.5.4	<i>Uncertainty Propagation</i>	294
7.5.5	<i>Transforming Standard Normal Random Variables</i>	295
7.5.6	<i>Introducing Correlation</i>	299
	References	302
A	Appendix: Mathematical Tools	303
A.1	Calculus	303
A.1.1	<i>Lagrange Multiplier Method</i>	303
A.1.2	<i>Asymptotics</i>	304
A.2	Linear Algebra	304
A.2.1	<i>Linear Independence, Span, Basis</i>	304
A.2.2	<i>Orthogonality and Norm</i>	305
A.2.3	<i>Gram–Schmidt Procedure</i>	306
A.2.4	<i>Eigenvalue Problem</i>	307
A.2.5	<i>Real Symmetric Matrices</i>	307
A.2.6	<i>Function of Real Symmetric Matrices</i>	308
A.3	Probability Theory	309
A.3.1	<i>Conditional Expectation</i>	309
A.3.2	<i>Conditional Variance Formula</i>	310
A.3.3	<i>Chebyshev's Inequality</i>	310
A.3.4	<i>Jensen's Inequality</i>	310
A.3.5	<i>Modes of Stochastic Convergence</i>	311
	Index	313

*May be skipped by advanced readers.

About the Authors

Dr Au is Chair of Uncertainty, Reliability and Risk with the Center for Engineering Dynamics and the Institute for Risk and Uncertainty at the University of Liverpool. He obtained his PhD in civil engineering from the California Institute of Technology. Dr Au specializes in both fundamental and applied research in engineering reliability analysis and structural health monitoring. He is experienced in full-scale dynamic testing of structures and has consulted on structural vibration projects on long-span pedestrian bridges, large-span floors, super-tall buildings, and microtremors. He is a member of the Hong Kong Institution of Engineers, Institution of Engineers Singapore, American Society of Civil Engineers, and the Earthquake Engineering Research Institute. He is a recipient of the IASSAR Junior Research Prize and the Nishino Prize.

Dr Wang is an Assistant Professor at the Department of Civil and Architectural Engineering, City University of Hong Kong. He obtained his PhD in geotechnical engineering from Cornell University. His research focuses on geotechnical risk and reliability (e.g., reliability-based design of foundations, development of Monte Carlo simulation-based methods for probabilistic analysis in geotechnical engineering, and probabilistic site characterization), seismic risk assessment of lifeline systems, soil–structure interaction, and geotechnical laboratory and in situ testing. Dr Wang was the President of the American Society of Civil Engineers – Hong Kong Section in 2012–2013. He is a recipient of the inaugural “Editor’s Choice” Paper Award by the *Canadian Geotechnical Journal* and the inaugural Wilson Tang Best Paper Award.

Preface

Modern engineering systems are designed with increasing complexity and higher expectation on their reliable performance. Assessing the effects of uncertainties on system performance and design implications is assuming greater importance. With the rapid development of computer technology, there is also an increasing trend of assessing risk and design via computer simulation, such as Monte Carlo methods. Failure is by design intended to be a rare event, but this makes its assessment by Direct Monte Carlo method computationally prohibitive.

This book introduces the reader to a simulation method called “Subset Simulation” for efficient engineering risk assessment involving rare failure events. Rare events (small probabilities) and high dimensions (a large number of random variables) are two main themes. The book is intended to provide an easy access to the necessary theories and computational tools for setting up and solving a risk assessment problem by Subset Simulation. It is targeted at graduate students, academics, researchers, and engineers interested in assessing the effects of uncertainties on system predictions. Undergraduate background in probability and statistics is assumed. Mathematical tools are provided in the Appendix for reference if necessary.

The book starts with basic theories in uncertainty propagation using Monte Carlo methods and the generation of random variables and stochastic processes for some common distributions encountered in engineering applications. It then introduces a powerful simulation tool called Markov Chain Monte Carlo method (MCMC), a pivotal machinery behind Subset Simulation that allows one to generate samples for investigating rare scenarios in a probabilistically consistent manner. The theory of Subset Simulation is then presented, addressing related practical issues encountered in the actual implementation. The book also discusses how to investigate scenarios when failure occurs, using the samples generated in Direct Monte Carlo or Subset Simulation.

A unique feature of this book is that it is supplemented with a VBA (Visual Basic for Applications) code that implements Direct Monte Carlo and Subset Simulation in the Excel spreadsheet environment. It can be downloaded at the following web site:

<https://sites.google.com/site/upssvba/>

The VBA code allows the reader to experiment with the examples in the book and get hands-on experience with simulation. One chapter of the book is devoted to a software framework that allows a practical solution by resolving the risk assessment problem into three uncoupled procedures, namely, deterministic modeling, uncertainty modeling and uncertainty propagation.

Acknowledgements

The first author was introduced to structural reliability research by Professor Lambros Katafygiotis (Hong Kong University of Science and Technology, HKUST) and Professor Costas Papadimitriou (University of Thessaly) while he was pursuing master research at HKUST. Professor James Beck (California Institute of Technology) posed a challenging but then “discouraging” problem of performing advanced Monte Carlo for reliability analysis with a large (possibly infinite) number of random variables, which subsequently led to the invention of Subset Simulation. The wonderful vision, excellent education, and unfailing support from these teachers are gratefully acknowledged.

The authors’ research in engineering reliability has been supported by the Pacific Earthquake Engineering Research Center (USA), Ministry of Education (Singapore), Defense Science Office (Singapore), Hong Kong Research Grant Council, and National Natural Science Foundation of China.

The manuscript was drafted while the first author was on sabbatical visit at the Tokyo City University hosted by Professor Ikumasa Yoshida, whose warm hospitality is gratefully acknowledged. Dr Zijun Cao (Wuhan University) assisted in the literature review of Subset Simulation in Chapter 5 and provided valuable comments on the manuscript. Dr Hongshuang Li (Nanjing University of Aeronautics and Astronautics) and Dr Konstantin Zuev (University of Liverpool) provided critical review of the manuscript during preparation. Dr Yan-Chun Ni (City University of Hong Kong) assisted in word-processing of the manuscript.

This book is dedicated to the life-long distinguished achievement of the late Professor Gerhart I. Schuëller in computational stochastic mechanics and reliability analysis of complex engineering systems. The first author would like to express his deepest gratitude to Professor Schuëller for his continuing encouragement and unfailing support, dating back to the days when the first author was pursuing a PhD on engineering reliability methods. To the first author, Professor Schuëller is a figure of wisdom and a caring mentor. The first author benefits greatly from Professor Schuëller’s vision and persistence in stochastic research especially related to complex engineering systems, whatever the challenge may be.

Nomenclature

CCDF	Complementary cumulative distribution function
CDF	Cumulative distribution function
c.o.v.	Coefficient of variation = standard deviation/mean
i.i.d.	Independent and identically distributed
LHS	Left hand side
MCMC	Markov Chain Monte Carlo
PDF	Probability density function
PMF	Probability mass function
RHS	Right hand side
$\mathbf{i}$	Purely imaginary number, $\sqrt{-1}$
$\mathcal{R}$	The set of real numbers
$\mathcal{R}^n$	The n -dimensional Euclidean space
$\mathcal{C}$	The set of complex numbers, i.e., $\{x + y\mathbf{i} : x, y \in \mathcal{R}\}$
$\Phi(\cdot)$	Standard Gaussian cumulative distribution function
$ \mathbf{x} $	Euclidean norm of vector $\mathbf{x}$, the square root of the sum of squares
$\det \mathbf{A}$	Determinant of matrix $\mathbf{A}$
$p_X(x)$	Probability density function of random variable X evaluated at argument x
$P(A)$	Probability of event A
$I(A)$	Indicator function of event A , equal to 1 if A is true and zero otherwise
$E[\cdot]$	Expectation (mean) of the argument
$\text{var}[\cdot]$	Variance of the argument
$\text{cov}[X, Y]$	Covariance between X and Y
$O(\epsilon)$	Of the order of ϵ

1

Introduction

Modern engineering has seen a booming demand for analyses of complex systems to unprecedented detail, paralleled with an increasing reliance on numerical models for performance predictions. Systems are designed with an increasing expectation of high performance reliability and robustness in functionality. Assessing the effects of uncertainties and their mitigation in the design decision-making process allows one to make risk-informed decisions even in a state of uncertainty. Uncertainties in engineering may arise from incomplete knowledge about the modeling of system behavior, model parameter values, measurement, environmental loading conditions, and so on. Probability theory allows a rational framework for plausible reasoning and decision-making in the presence of uncertainties. The analysis of the effects of uncertainty includes, but is by no means limited to, the following objectives:

1. Reliability (or risk) analysis – to assess the likelihood of violating specified system performance criteria. It involves assessing the probability distribution or performance margins of some critical system response. This can be used for examining whether the system is likely to pass specified performance criteria in the presence of modeled uncertainties.
2. Failure analysis – to assess the characteristics of failure scenarios, for example, the likely cause and consequence of failure. The former provides insights about system failures and helps devise effective measures for their mitigation. The latter reveals the likely scenarios when failure occurs and provides information for loss estimation, devising contingency measures, or trading-off cost–benefits in design.

Models for complex systems are characterized by a large number of governing state variables, time-varying and response-dependent nonlinear behavior. They are also increasingly governed by multi-physics laws. Although the advent of computer technology has allowed the analysis of complex systems for a given scenario to be performed with affordable computational time, the same is not true for analyzing the effects of uncertainty, since the latter involves information from multiple scenarios and hence repeated system analyses. Even if resources are available, they should be deployed in an effective manner that yields information on failure scenarios of concern with a consistent weight on their likelihood. This motivates the development of efficient yet robust computational algorithms for propagating uncertainties in complex systems.

This book is primarily concerned with performing risk and failure analysis by means of an advanced Monte Carlo method called “Subset Simulation.” The method is based on the simple idea that a small failure probability can be expressed as the product of a number of not-so-small conditional failure probabilities. This idea has led to algorithms that generate random samples gradually propagating towards the failure region in the uncertain parameter space. The samples provide information for estimating the whole distribution of the critical response quantity that governs failure, covering large (central) to small (tail) probability regimes. The method has been found to be efficient for investigating rare failure events, but still retains some robustness to problem complexity in different applications. It treats the system as a black box and hence does not explore any prior information one may have regarding the system behavior, which can possibly be incorporated into the solution process. Thus, for a particular application, it may not be the most efficient method. However, since it can be applied without much knowledge about the system (like Direct Monte Carlo) it may still be a competitive algorithm when robustness is taken into consideration. The possibility of using the generated samples for investigating failure scenarios also makes the method versatile for risk and failure analysis.

1.1 Formulation

Despite the wide variety of problems encountered in engineering applications, a failure event can often be represented as the exceedance of a critical scalar response variable Y over a specified threshold b . The response variable Y is assumed to be completely determined by a set of “input variables” $\mathbf{X} = [X_1, \dots, X_n]$. The relationship is generically represented as

$$Y = h(\mathbf{X}) \quad (1.1)$$

where $h : \mathcal{R}^n \mapsto \mathcal{R}$ is a known deterministic function that represents the computational process, for example, the analytical formula, empirical formula, finite element model, computational dynamics, and so on. Clearly, when $\mathbf{X}$ is uncertain, so is Y . Using a probabilistic approach, $X_1, \dots, X_n$ are modeled as random variables with prescribed joint probability distribution assigned based on the analyst’s knowledge. Induced by the probabilistic modeling on $\mathbf{X}$, Y is also a random variable. However, its probability distribution is not arbitrary and is not up to the analyst to decide. Rather, it is completely determined by the probability distribution of $\mathbf{X}$ and the function h . This is depicted in Figure 1.1.

In order to make decisions related to Y , which is nevertheless uncertain, one needs to have information about its probability distribution. This is generally unknown, however. It must be determined in accordance with the function h and the probability distribution of $\mathbf{X}$. The effort required depends largely on which part of the distribution of Y is relevant. Statistical quantities related to the “frequent” or central part of the distribution, such as the mean or variance, are often easier to obtain than those related to the “rare” or tail part of the distribution, such as the exceedance probability $P(Y > b)$ when b is large. The latter is the primary interest in this book.

If we denote the failure event as $F = \{Y > b\}$, then we can write

$$\text{Failure probability} = P(F) = P(Y > b) \quad (1.2)$$

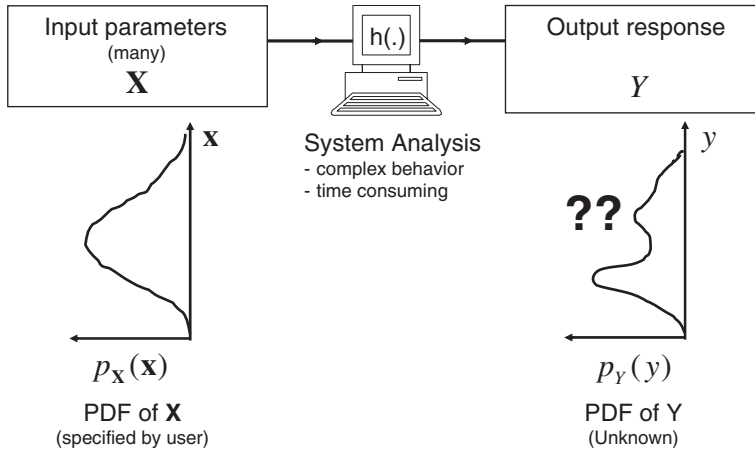


Figure 1.1 Input–output context.

Complementary to the failure probability is the “reliability”:

$$\text{Reliability} = 1 - P(F) = P(Y \leq b) = 1 - P(Y > b) \quad (1.3)$$

Evaluating the failure probability and conditional expectation for failure analysis requires information about the system when failure occurs. Properly designed engineering systems are intended to have high reliability (close to 1) and hence small failure probability (close to zero). Target failure probabilities often needed to be estimated are in the order of $10^{-3} \sim 10^{-6}$, which nevertheless depends on the class of applications.

For complex problems, the relationship between $\mathbf{X}$ and Y is analytically intractable and is often only known implicitly. That is, the value of Y for a given $\mathbf{X}$ can be calculated but no other information (e.g., derivative) is available. The relationship is also difficult to visualize when $\mathbf{X}$ contains a large number of uncertain variables. Analytical or closed-form solutions for the required statistics of Y are rarely available.

The Direct Monte Carlo method provides a robust means for estimating the statistics by averaging over pseudo-random samples generated according to the distribution of $\mathbf{X}$. It has become increasingly popular due to the advent of modern computer technology. When the statistics are related to the tail of the distribution of Y , however, it is not efficient because most of the samples lie in the frequent region. Only those lying at the tail of the distribution of Y provide useful information for estimating the tail statistics, but their occurrence is rare.

The failure probability can be mathematically formulated in several ways that lead to different strategies for its computation. Without loss of generality (see Section 1.2), assume that $\mathbf{X} = [X_1, \dots, X_n]$ is a set of continuous-valued random variables with probability density function (PDF) $q(\mathbf{x})$. The failure probability can be formulated as a “probability integral”:

$$P(Y > b) = \int_F q(\mathbf{x}) d\mathbf{x} \quad (1.4)$$

where

$$F = \{\mathbf{x} : h(\mathbf{x}) > b\} \quad (1.5)$$

denotes the “failure region,” that is, a subset in the parameter space of $\mathbf{X}$ that corresponds to failure. The failure probability can thus be viewed as a sum of the probability content within the failure region. Alternatively, the integral can be written as being over the whole parameter space:

$$P(Y > b) = \int I_F(\mathbf{x})q(\mathbf{x})d\mathbf{x} \quad (1.6)$$

where

$$I_F(\mathbf{x}) = \begin{cases} 1 & \text{if } h(\mathbf{x}) > b \\ 0 & \text{if } h(\mathbf{x}) \leq b \end{cases} \quad (1.7)$$

is the “indicator function” that reveals whether $\mathbf{x}$ lies in the failure region or not. This form is often used for mathematical derivations. Another useful perspective is via the expectation:

$$P(Y > b) = E[I_F(\mathbf{X})] \quad (1.8)$$

where $E[\cdot]$ denotes the mathematical expectation when $\mathbf{X}$ is distributed as q . This leads to the idea of “statistical averaging” and hence Monte Carlo simulation.

Viewing $P(Y > b)$ as a function of b , finding the failure probability is equivalent to finding the “complementary cumulative distribution function” (CCDF) of Y (CCDF = 1 – CDF), especially at the tail where small failure probabilities are the main interest. Of course, finding the whole CDF is much more difficult, or at least computationally more expensive, than finding just the failure probability at a single threshold level. Nevertheless, estimating small failure probabilities is intimately related to estimating the upper tail of the CCDF.

Example 1.1 Definition of response variable

Many system failure events can be expressed in terms of the union or intersection of exceedance events, say, corresponding to system components connected (logically) in series or in parallel. A failure event of this kind can be expressed in terms of the exceedance of a scalar response Y . Clearly Y should be defined such that $P(Y > b)$ corresponds to the failure probability of interest. It is also preferable to define Y in a non-dimensional manner.

Suppose $F = \{C < D\}$, where C and D are the “capacity” and “demand” of a system that can possibly depend on $\mathbf{X}$. Then Y may be defined in a dimensionless manner as $Y = D/C$ so that $P(F) = P(Y > 1)$.

Suppose now $F = \bigcap_{i=1}^{n_1} \{C_i < D_i\}$, where C_i and D_i ($i = 1, \dots, n_1$) can possibly depend on $\mathbf{X}$. This can be interpreted as the failure of a system of components connected in parallel where the system fails only when all the components have failed. In this case the critical response Y may be defined as $Y = \min_{i=1}^{n_1} D_i/C_i$ so that $P(F) = P(Y > 1)$.

On the other hand, if $F = \bigcup_{i=1}^{n_2} \{C_i < D_i\}$, then it can be interpreted as the failure of a system of components connected in series where the system fails if any one of the components fails. In this case Y may be defined as $Y = \max_{i=1}^{n_2} D_i/C_i$ so that $P(F) = P(Y > 1)$.

In general, if F is defined via $\cap$ and/or $\cup$, then Y can be defined using “min” and/or “max” appearing in the same order. For example, if $F = \bigcap_{i=1}^{n_1} \bigcup_{j=1}^{n_2} \{C_{ij} < D_{ij}\}$ then we can define $Y = \min_{i=1}^{n_1} \max_{j=1}^{n_2} D_{ij}/C_{ij}$ so that $P(F) = P(Y > 1)$. ■

1.2 Context

Unless otherwise stated, the problems that we deal with in this book have the following context:

1. The input random variables $X_1, \dots, X_n$ are continuous-valued.
2. The input random variables $X_1, \dots, X_n$ are mutually independent.
3. The (one-dimensional) PDF of each X_i , denoted by $q_i(x)$ corresponds to some known “standard distribution” (e.g., Gaussian, exponential) so that
 - (a) the value of $q_i(x)$ can be evaluated efficiently for any given x
 - (b) random samples distributed as q_i can be generated efficiently.
4. The relationship between $\mathbf{X}$ and Y is not explicitly known. That is, we can evaluate the value of $Y = h(\mathbf{x})$ for a given $\mathbf{x}$ but generally we are not able to obtain other information such as gradient or Hessian. The latter quantities if needed have to be computed numerically, for example, using finite difference.
5. The computational effort for evaluating $h(\mathbf{x})$ for a given $\mathbf{x}$ is significant. The total computational effort is dominated by the number of function evaluations of $h(\mathbf{x})$.
6. Interest is focused on small failure probabilities or, equivalently, the tail of the CCDF of $Y = h(\mathbf{X})$.
7. The number of random variables in $\mathbf{X}$ can be very large (possibly infinite).

Some comments are in order regarding the above context. Assumption 1 on continuous random variables is introduced primarily for the sake of discussion and elegance in the theory (e.g., integrals instead of sums). It does not introduce much loss of generality in practice, because discrete-valued random variables can be generated by a mapping of continuous-valued random variables. Assumption 2 on mutual independence of input random variables does not generate any loss of generality because, in reality, dependent variables are generated by independent ones. Assumption 3 on standard distributions distinguishes the problems discussed in this book from Bayesian inference problems, in which case the posterior distribution of random variables given data often do not correspond to any standard distribution (see Section 1.4).

1.3 Extreme Value Theory

As mentioned in the beginning of this chapter, Subset Simulation treats the input–output relationship of a system as a black box and so it (often) need not be the most efficient procedure for a particular application. When there is some knowledge about the relationship

between $\mathbf{X}$ and Y it may be possible to take advantage of it to derive useful statements about the distribution of Y . One classical example with profound results is when Y is defined as the maximum over a large number (theoretically infinite) of i.i.d. (independent and identically distributed) random variables in $\mathbf{X}$. This has been studied extensively, leading to “extreme value theory” (Gumbel, 1958; Galambos, 1978; David, 1981). When the problem context fits and the asymptotic distribution of the extreme exists, it is usually more efficient to apply the theory to determine the failure probability $P(Y > b)$. In this case the main task is to identify the type of limiting distribution and then to determine the distribution parameters accordingly. Standard statistical tools are available (Coles, 2001). Although one can still apply Subset Simulation to solve the same problem, it is less efficient because it does not take advantage of the special mathematical structure of the problem.

1.4 Exclusion

This book does not deal with the case when the distribution of $\mathbf{X}$ arises from Bayesian inference problems, which is nevertheless a very important problem with wide application (Cox, 1961; Jaynes, 2003). In this area, the interest is to determine the distribution of $\mathbf{X}$ and update response predictions based on some observed data D . According to Bayes’ Theorem, the “posterior distribution” (i.e., given data) of $\mathbf{X}$ that incorporates the information from the data D is given by

$$p(\mathbf{x} | D) = p(D)^{-1} p(D | \mathbf{x}) p(\mathbf{x}) \quad (1.9)$$

The RHS of this equation should be viewed as a probability distribution of $\mathbf{X}$. The first term $p(D)^{-1}$ does not depend on $\mathbf{x}$ and so, as far as the distribution of $\mathbf{X}$ is concerned, it can be ignored. The middle term $p(D | \mathbf{x})$ is called the “likelihood function,” which must be formulated based on modeling assumptions relating the observed data to $\mathbf{X}$ in a probabilistic manner. The last term $p(\mathbf{x})$ is called the “prior distribution” and it reflects one’s knowledge about $\mathbf{X}$ in the absence of data.

Estimating the posterior statistics of $\mathbf{X}$ or updating system response prediction by means of Monte Carlo simulation requires efficient generation of samples according to the posterior distribution $p(\mathbf{x} | D)$. This is generally a highly non-trivial task, however. Although the prior distribution $p(\mathbf{x})$ is often chosen to follow a standard distribution (like those considered in Chapter 3), the resulting posterior distribution does not necessarily follow a standard distribution because the likelihood function $p(D | \mathbf{x})$ arises from system modeling and is problem-dependent. In many applications the likelihood function is only known implicitly and its dependence on $\mathbf{x}$ is rather complicated.

Conjugate prior distribution is one branch of research that examines the type of prior distribution that should be assumed for some type of likelihood function so that the resulting posterior distribution is also of a standard distribution. The use of conjugate prior distribution is convenient when applicable, but otherwise it limits the type of problem that can be solved. It has become less popular in modern applications due to the advent of computer technology and the development of advanced simulation methods that can efficiently handle arbitrary distributions. The “Markov chain Monte Carlo method” (MCMC) is one popular class of

methods that has been found useful. This method is discussed in Chapter 4 as it is used for generating failure samples in Subset Simulation.

1.5 Organization of this Book

This book is organized into seven chapters. After the introduction (this chapter), Chapter 2 gives an overview of relevant ideas that lead logically to Subset Simulation. These ideas differ in the way they view the failure probability and the way they gather and use information to account for the main contribution to the failure probability. Chapter 3 gives a basic introduction to the digital simulation of random samples according to standard distributions (e.g., Normal, Lognormal, exponential), which is indispensable for uncertainty modeling and performing Monte Carlo simulation. Chapter 4 gives a basic introduction to “Markov Chain Monte Carlo” (MCMC), which is a powerful method for generating random samples according to an arbitrarily given probability distribution. MCMC is not involved in uncertainty modeling in the context of this book, as the uncertain parameters are assumed to have standard distributions. Rather, it is involved in the efficient generation of failure samples in Subset Simulation, which is a highly non-trivial problem. Chapter 4 provides the necessary background where no pre-requisite in Markov Chain theory is needed.

Chapter 5 gives a comprehensive coverage of Subset Simulation for estimating failure probabilities through the CCDF of the critical response governing failure. It covers the basic algorithm, error estimation, choice of parameters, theoretical properties of estimators, and potential problems. Chapter 6 introduces the investigation of failure scenarios using the failure samples in Direct Monte Carlo and Subset Simulation. Chapter 7 presents an Excel spreadsheet package for performing risk assessment by Direct Monte Carlo and Subset Simulation. It contains step-by-step procedures that allow the reader to gain hands-on experience with Monte Carlo simulation. This will hopefully help the reader develop a correct perspective for interpreting and using simulation results. Mathematical tools are contained in the Appendix for reference.

1.6 Remarks on the Use of Risk Analysis

Reliability analysis or probabilistic failure analysis, or any kind of analysis in general, does not itself prevent failure from happening or provide warranty over losses. Nor does it necessarily provide information close to reality, because the underlying assumptions need not do so. These issues should not undermine the value of risk analysis because it is not meant to do so. Risk analysis is only meant to provide the decision-maker with information regarding the effects of uncertainty on the attributes that may affect a decision. The decision-maker is still required to make his or her own judgment on the use of the results. It is just a scientific way of producing relevant information consistent with the assumptions adopted regarding the modeling of uncertainty and system behavior. Having advanced computational tools hopefully allows one to focus on the problem itself, especially the decision-making part. Making assumptions is inevitable and this should be kept in mind. In many cases, an order of magnitude answer on the probability suffices for making decisions, which may also be consistent with the variability of such an answer in view of the assumptions made. Making

assumptions and placing the right confidence into the results is a human art. Practically, it is better to be “approximately right” rather than “precisely wrong.”

1.7 Conventions

Before we leave this chapter, we cover some notations and conventions used in this book. We use $f(x)$ to denote a function of the argument x . When this may be confused with the value of a function at a specific x , we use f or $f(\cdot)$ to denote the function. The notation $f : A \rightarrow B$ is used to denote a function that takes an element in the set A to give a value in the set B . For example, $f : \mathcal{R}^n \rightarrow \mathcal{R}$ denotes a real scalar valued multi-variable function on the n -dimensional Euclidean space.

We reserve $P(\cdot)$ for the probability of the statement in the argument. The notation $p_X(x)$ refers to the PDF of the random variable X evaluated at the value x . When the random variable X is understood in the context it may be omitted for simplicity. Random variables are usually denoted in capital letters and their parameter value in small letters. For example, X is the random variable and $\{X = x\}$ is the event that it is equal to the given parameter value x . Vector-valued quantities are often denoted in bold, for example, $\mathbf{X} = [X_1, \dots, X_n]$ is a vector of random variables. When the limits of summation or domain of integration are understood, they may be omitted for simplicity. An integral sign without the domain indicated is over the whole parameter space on which the integrand is defined. A sequence of quantities may be denoted in an abbreviated manner in curly braces with a running index. For example, $\{X_1, \dots, X_N\}$ may be written as $\{X_k : k = 1, \dots, N\}$ or abbreviated as $\{X_k\}$ when the limits the index runs through are clear. The terms “Gaussian distribution” and “Normal distribution” refer to the same distribution and are used interchangeably. Other notations and abbreviations are contained in the Nomenclature.

References

- Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, Singapore.
- Cox, R.T. (1961) *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore.
- David, H.A. (1981) *Order Statistics*, John Wiley & Sons, Inc., New York.
- Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*, John Wiley & Sons, Inc., New York.
- Gumbel, E.J. (1958) *Statistics of Extremes*, Columbia University Press, New York.
- Jaynes, E.T. (2003) *Probability Theory: The Logic of Science*, Cambridge University Press, UK.