Uncertainty Analysis with High Dimensional Dependence Modelling

Dorota Kurowicka and Roger Cooke
Delft University of Technology, The Netherlands

John Wiley & Sons, Ltd
Uncertainty Analysis with High Dimensional Dependence Modelling
WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS


A complete list of the titles in this series appears at the end of this volume.
Uncertainty Analysis with High Dimensional Dependence Modelling

Dorota Kurowicka and Roger Cooke
Delft University of Technology, The Netherlands
## Contents

<table>
<thead>
<tr>
<th>Preface</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Wags and Bogsats</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Uncertainty analysis and decision support: a recent example</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Outline of the book</td>
<td>9</td>
</tr>
<tr>
<td><strong>2 Assessing Uncertainty on Model Input</strong></td>
<td>13</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Structured expert judgment in outline</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Assessing distributions of continuous univariate uncertain quantities</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Assessing dependencies</td>
<td>16</td>
</tr>
<tr>
<td>2.5 Unicorn</td>
<td>20</td>
</tr>
<tr>
<td>2.6 Unicorn projects</td>
<td>20</td>
</tr>
<tr>
<td><strong>3 Bivariate Dependence</strong></td>
<td>25</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>25</td>
</tr>
<tr>
<td>3.2 Measures of dependence</td>
<td>26</td>
</tr>
<tr>
<td>3.2.1 Product moment correlation</td>
<td>26</td>
</tr>
<tr>
<td>3.2.2 Rank correlation</td>
<td>30</td>
</tr>
<tr>
<td>3.2.3 Kendall’s tau</td>
<td>32</td>
</tr>
<tr>
<td>3.3 Partial, conditional and multiple correlations</td>
<td>32</td>
</tr>
<tr>
<td>3.4 Copulae</td>
<td>34</td>
</tr>
<tr>
<td>3.4.1 Fréchet copula</td>
<td>36</td>
</tr>
<tr>
<td>3.4.2 Diagonal band copula</td>
<td>37</td>
</tr>
<tr>
<td>3.4.3 Generalized diagonal band copula</td>
<td>41</td>
</tr>
<tr>
<td>3.4.4 Elliptical copula</td>
<td>42</td>
</tr>
<tr>
<td>3.4.5 Archimedean copulae</td>
<td>45</td>
</tr>
<tr>
<td>3.4.6 Minimum information copula</td>
<td>47</td>
</tr>
<tr>
<td>3.4.7 Comparison of copulae</td>
<td>49</td>
</tr>
<tr>
<td>3.5 Bivariate normal distribution</td>
<td>50</td>
</tr>
<tr>
<td>3.5.1 Basic properties</td>
<td>50</td>
</tr>
<tr>
<td>3.6 Multivariate extensions</td>
<td>51</td>
</tr>
<tr>
<td>3.6.1 Multivariate dependence measures</td>
<td>51</td>
</tr>
</tbody>
</table>
### CONTENTS

3.6.2 Multivariate copulae ........................................ 53
3.6.3 Multivariate normal distribution ............................ 53
3.7 Conclusions ....................................................... 54
3.8 Unicorn projects .................................................. 55
3.9 Exercises .......................................................... 61
3.10 Supplement ........................................................ 67

4 High-dimensional Dependence Modelling .......................... 81
4.1 Introduction ......................................................... 81
4.2 Joint normal transform ............................................ 82
4.3 Dependence trees ................................................... 86
  4.3.1 Trees .......................................................... 86
  4.3.2 Dependence trees with copulae ............................... 86
  4.3.3 Example: Investment .......................................... 90
4.4 Dependence vines .................................................. 92
  4.4.1 Vines ........................................................ 92
  4.4.2 Bivariate- and copula-vine specifications .................... 96
  4.4.3 Example: Investment continued ............................... 98
  4.4.4 Partial correlation vines ..................................... 99
  4.4.5 Normal vines ................................................ 101
  4.4.6 Relationship between conditional rank and partial correla-
        tions on a regular vine ....................................... 101
4.5 Vines and positive definiteness ................................ 105
  4.5.1 Checking positive definiteness ............................... 105
  4.5.2 Repairing violations of positive definiteness ............... 107
  4.5.3 The completion problem ..................................... 109
4.6 Conclusions ......................................................... 111
4.7 Unicorn projects .................................................. 111
4.8 Exercises .......................................................... 115
4.9 Supplement ........................................................ 116
  4.9.1 Proofs ....................................................... 116
  4.9.2 Results for Section 4.4.6 ................................... 127
  4.9.3 Example of fourivariate correlation matrices ............... 129
  4.9.4 Results for Section 4.5.2 ................................... 130

5 Other Graphical Models ............................................. 131
5.1 Introduction ......................................................... 131
5.2 Bayesian belief nets .............................................. 131
  5.2.1 Discrete bbn’s .............................................. 132
  5.2.2 Continuous bbn’s .......................................... 133
5.3 Independence graphs .............................................. 141
5.4 Model inference .................................................... 142
  5.4.1 Inference for bbn’s ......................................... 143
  5.4.2 Inference for independence graphs ............................ 144
  5.4.3 Inference for vines ......................................... 145
## Contents

5.5 Conclusions .......................................................... 150  
5.6 Unicorn projects ...................................................... 150  
5.7 Supplement .......................................................... 157  

6 Sampling Methods .................................................. 159  
6.1 Introduction .......................................................... 159  
6.2 (Pseudo-) random sampling ....................................... 160  
6.3 Reduced variance sampling ......................................... 161  
6.3.1 Quasi-random sampling ......................................... 161  
6.3.2 Stratified sampling ............................................... 164  
6.3.3 Latin hypercube sampling ....................................... 166  
6.4 Sampling trees, vines and continuous bbn’s ................. 168  
6.4.1 Sampling a tree .................................................. 168  
6.4.2 Sampling a regular vine ......................................... 169  
6.4.3 Density approach to sampling regular vine ................. 174  
6.4.4 Sampling a continuous bbn ..................................... 174  
6.5 Conclusions .......................................................... 180  
6.6 Unicorn projects ...................................................... 180  
6.7 Exercise .............................................................. 184  

7 Visualization ......................................................... 185  
7.1 Introduction .......................................................... 185  
7.2 A simple problem .................................................... 186  
7.3 Tornado graphs ....................................................... 186  
7.4 Radar graphs .......................................................... 187  
7.5 Scatter plots, matrix and overlay scatter plots ............... 188  
7.6 Cobweb plots ........................................................ 191  
7.7 Cobweb plots local sensitivity: dike ring reliability ....... 195  
7.8 Radar plots for importance; internal dosimetry .............. 199  
7.9 Conclusions .......................................................... 201  
7.10 Unicorn projects ..................................................... 201  
7.11 Exercises ........................................................... 203  

8 Probabilistic Sensitivity Measures ................................ 205  
8.1 Introduction .......................................................... 205  
8.2 Screening techniques ................................................. 205  
8.2.1 Morris’ method .................................................. 205  
8.2.2 Design of experiments ........................................... 208  
8.3 Global sensitivity measures ........................................ 214  
8.3.1 Correlation ratio ................................................ 215  
8.3.2 Sobol indices .................................................... 219  
8.4 Local sensitivity measures ......................................... 222  
8.4.1 First order reliability method .................................. 222  
8.4.2 Local probabilistic sensitivity measure ...................... 223  
8.4.3 Computing $\frac{\partial E(X|g_0)}{\partial g_0}$ .......................... 225
9 Probabilistic Inversion .................................................. 239
 9.1 Introduction ......................................................... 239
 9.2 Existing algorithms for probabilistic inversion .................. 240
    9.2.1 Conditional sampling ........................................ 240
    9.2.2 PARFUM ..................................................... 242
    9.2.3 Hora-Young and PREJUDICE algorithms .................... 243
 9.3 Iterative algorithms ............................................. 243
    9.3.1 Iterative proportional fitting ............................. 244
    9.3.2 Iterative PARFUM .......................................... 245
 9.4 Sample re-weighting ............................................ 246
    9.4.1 Notation .................................................... 246
    9.4.2 Optimization approaches .................................. 247
    9.4.3 IPF and PARFUM for sample re-weighting probabilistic
          inversion .................................................... 248
 9.5 Applications .................................................... 249
    9.5.1 Dispersion coefficients .................................... 249
    9.5.2 Chicken processing line .................................... 252
 9.6 Convolution constraints with prescribed margins .............. 253
 9.7 Conclusions ...................................................... 255
 9.8 Unicorn projects ................................................ 256
 9.9 Supplement ....................................................... 258
    9.9.1 Proofs ...................................................... 258
    9.9.2 IPF and PARFUM .......................................... 263

10 Uncertainty and the UN Compensation Commission ............... 269
 10.1 Introduction ..................................................... 269
 10.2 Claims based on uncertainty ................................... 270
 10.3 Who pays for uncertainty ...................................... 272

Bibliography .................................................................. 273

Index ............................................................................ 281
Preface

This book emerges from a course given at the Department of Mathematics of the Delft University of Technology. It forms a part of the program on Risk and Environmental Modelling open to graduate students with the equivalent of a Bachelor’s degree in mathematics. The students are familiar with undergraduate analysis, statistics and probability, but for non-mathematicians this familiarity may be latent. Therefore, most notions are ‘explained in-line’. Readers with a nodding acquaintance with these subjects can follow the thread. To keep this thread visible, proofs are put in supplements of the chapters in which they occur. Exercises are also included in most chapters.

The real source of this book is our experience in applying uncertainty analysis. We have tried to keep the applications orientation in the foreground. Indeed, the whole motivation for developing generic tools for high dimensional dependence modelling is that decision makers and problem owners are becoming increasingly sophisticated in reasoning with uncertainty. They are making demands, which an analyst with the traditional tools of probabilistic modelling cannot meet. Put simply, our point of view is this: a joint distribution is specified by specifying a sampling procedure. We therefore assemble tools and techniques for sampling and analysing high dimensional distributions with dependence. These same tools and techniques form the design requirements for a generic uncertainty analysis program. One such program is UNcertainty analysis wIth CORrelatioNs (UNICORN). A fairly ponderous light version may be downloaded from http://ssor.twi.tudelft.nl/ risk/. UNICORN projects are included in each chapter to give hands on experience in applying uncertainty analysis.

The people who have contributed substantially to this book are too numerous to list, but certainly include Valery Kritchallo, Tim Bedford, Daniel Lewandowski, Belinda Chiera, Du Chao, Bernd Kraan and Jolanta Misiewicz.
Introduction: Uncertainty Analysis and Dependence Modelling

1.1 Wags and Bogsats

‘...whether true or not [it] is at least probable; and he who tells nothing exceeding the bounds of probability has a right to demand that they should believe him who cannot contradict him’. Samuel Johnson, author of the first English dictionary, wrote this in 1735. He is referring to the Jesuit priest Jeronimo Lobo’s account of the unicorns he saw during his visit to Abyssinia in the 17th century (Shepard (1930) p. 200).

Johnson could have been the apologist for much of what passed as decision support in the period after World War II, when think tanks, forecasters and expert judgment burst upon the scientific stage. Most salient in this genre is the book The Year 2000 (Kahn and Wiener (1967)) in which the authors published 25 ‘even money bets’ predicting features of the year 2000, including interplanetary engineering and conversion of humans to fluid breathers. Essentially, these are statements without pedigree or warrant, whose credibility rests on shifting the burden of proof. Their cavalier attitude toward uncertainty in quantitative decision support is representative of the period. Readers interested in how many of these even money bets the authors have won, and in other examples from this period, are referred to (Cooke (1991), Chapter 1).

Quantitative models pervade all aspects of decision making, from failure probabilities of unlaunched rockets, risks of nuclear reactors and effects of pollutants on health and the environment to consequences of economic policies. Such quantitative models generally require values for parameters that cannot be measured or...
INTRODUCTION

assessed with certainty. Engineers and scientists sometimes cover their modesty with churlish acronyms designating the source of ungrounded assessments. ‘Wags’ (wild-ass guesses) and ‘bogsats’ (bunch of guys sitting around a table) are two examples found in published documentation.

Decision makers, especially those in the public arena, increasingly recognize that input to quantitative models is uncertain and demand that this uncertainty be quantified and propagated through the models.

Initially, it was the modellers themselves who provided assessments of uncertainty and did the propagating. Not surprisingly, this activity was considered secondary to the main activity of computing ‘nominal values’ or ‘best estimates’ to be used for forecasting and planning and received cursory attention.

Figure 1.1 shows the result of such an in-house uncertainty analysis performed by the National Radiological Protection Board (NRPB) and The Kernforschungszentrum Karlsruhe (KFK) in the late 1980s (Crick et al. (1988); Fischer et al. (1990)). The models in question predict the dispersion of radioactive material in the atmosphere following an accident in a nuclear reactor. The figure shows predicted lateral dispersion under stable conditions, and also shows wider and narrower plumes, which the modellers are 90% certain will enclose an actual plume under the stated conditions.

It soon became evident that if things were uncertain, then experts might disagree, and using one expert-modeller’s estimates of uncertainty might not be sufficient. Structured expert judgment has since become an accepted method for quantifying models with uncertain input. ‘Structured’ means that the experts are identifiable, the assessments are traceable and the computations are transparent. To appreciate the difference between structured and unstructured expert judgment, Figure 1.2 shows the results of a structured expert judgment quantification of the same uncertainty pictured in Figure 1.1 (Cooke (1997b)). Evidently, the picture of uncertainty emerging from these two figures is quite different.

One of the reasons for the difference between these figures is the following: The lateral spread of a plume as a function of down wind distance $x$ is modelled, per stability class, as

$$\sigma(x) = Ax^B.$$
Both the constants $A$ and $B$ are uncertain as attested by spreads in published values of these coefficients. However, these uncertainties cannot be independent. Obviously if $A$ takes a large value, then $B$ will tend to take smaller values. Recognizing the implausibility of assigning $A$ and $B$ as independent uncertainty distributions, and the difficulty of assessing a joint distribution on $A$ and $B$, the modellers elected to consider $B$ as a constant; that is, as known with certainty. 

The differences between these two figures reflect a change in perception regarding the goal of quantitative modelling. With the first picture, the main effort has gone into constructing a quantitative deterministic model to which uncertainty quantification and propagation are added on. In the second picture, the model is essentially about capturing uncertainty. Quantitative models are useful insofar as they help us resolve and reduce uncertainty. Three major differences in the practice of quantitative decision support follow from this shift of perception.

- First of all, the representation of uncertainty via expert judgment, or some other method is seen as a scientific activity subject to methodological rules every bit as rigorous as those governing the use of measurement or experimental data.

- Second, it is recognized that an essential part of uncertainty analysis is the analysis of dependence. Indeed, if all uncertainties are independent, then their propagation is mathematically trivial (though perhaps computationally

---

1This is certainly not the only reason for the differences between Figures 1.1 and 1.2. There was also ambivalence with regard to what the uncertainty should capture. Should it capture the plume uncertainty in a single accidental release, or the uncertainty in the average plume spread in a large number of accidents? Risk analysts clearly required the former, but meteorologists are more inclined to think in terms of the latter.