International Series in
Operations Research & Management Science

Eduardo Souza de Cursi

Uncertainty Quantification with R

Bayesian Methods

Springer

International Series in Operations Research & Management Science

Volume 352

The book series **International Series in Operations Research and Management Science** encompasses the various areas of operations research and management science. Both theoretical and applied books are included. It describes current advances anywhere in the world that are at the cutting edge of the field. The series is aimed especially at researchers, advanced graduate students, and sophisticated practitioners.

The series features three types of books:

- Advanced expository books that extend and unify our understanding of particular areas.
- Research monographs that make substantial contributions to knowledge.
- Handbooks that define the new state of the art in particular areas. Each handbook will be edited by a leading authority in the area who will organize a team of experts on various aspects of the topic to write individual chapters. A handbook may emphasize expository surveys or completely new advances (either research or applications) or a combination of both.

The series emphasizes the following four areas:

Mathematical Programming: Including linear programming, integer programming, nonlinear programming, interior point methods, game theory, network optimization models, combinatorics, equilibrium programming, complementarity theory, multiobjective optimization, dynamic programming, stochastic programming, complexity theory, etc.

Applied Probability: Including queuing theory, simulation, renewal theory, Brownian motion and diffusion processes, decision analysis, Markov decision processes, reliability theory, forecasting, other stochastic processes motivated by applications, etc.

Production and Operations Management: Including inventory theory, production scheduling, capacity planning, facility location, supply chain management, distribution systems, materials requirements planning, just-in-time systems, flexible manufacturing systems, design of production lines, logistical planning, strategic issues, etc.

Applications of Operations Research and Management Science: Including telecommunications, health care, capital budgeting and finance, economics, marketing, public policy, military operations research, humanitarian relief and disaster mitigation, service operations, transportation systems, etc.

This book series is indexed in Scopus.

Eduardo Souza de Cursi

Uncertainty Quantification with R

Bayesian Methods

Springer

Eduardo Souza de Cursi
Department Mechanics / Civil Engineering
INSA Rouen Normandie
Saint-Etienne du Rouvray, France

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-031-48207-6 ISBN 978-3-031-48208-3 (eBook)
https://doi.org/10.1007/978-3-031-48208-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Introduction

This book is an independent companion volume to *Uncertainty Quantification with R*, complementing certain of its topics and taking up others from a different angle – the Bayesian Standpoint.

This is a fairly general book, covering a range of methods and techniques, each of which can be further developed in the literature specific to each of the fields covered – and the reader can find scientific articles that develop them in depth.

This book targets the use of R, which is a GNU project to develop a tool for language and environment for statistical computing and graphics. An IDE is proposed by RStudio. The popularity of R and RStudio make that the reader will find on the web many sites and information about it. A wide literature can also be found about this software. The community of the users of R proposes a large choice of packages to extend the possibilities of R. You will find repositories containing them, such as, for instance, https://cran.r-project.org/web/packages/.

The book contains many programs. If you are an expert in R, you will find certainly a large number of improvements to the programs presented. Analogously, the community of the users of R proposes many packages to solve a large number of practical problems and implementing the methods considered in this book. We cite many of them, but – probably, even certainly – not all the existing contributions. The author apologizes in advance to any forgotten contributors, whose works are not cited in the book, but who have made the effort and been kind enough to make their work product available to the community. As mentioned above, we recommend that R users search software repositories such as https://cran.r-project.org/.

If you are not familiarized with Uncertainty Quantification (UQ), you can consider it as a collection of methods for the analysis of numerical data, namely when uncertainty or variability is involved, having as general objective the determination of probability distributions. The general aim of UQ is to characterize the observed variability in a quantity X by using another random variable U, by using the available information about (X, U) to construct an explanation of X by U, into a form which will be useful for use in numerical calculations involving X. The information may be, for instance, an equation, a numerical problem involving both the variables, or samples. UQ applies to a wide range of situations.

In the book *Uncertainty Quantification with R*, we tried to illustrate the practical use of UQ techniques under R. In this book, the philosophy is the same: we focus on practical aspects, and the theoretical arguments are reduced to the strictly minimal amount necessary to the understanding of the practical methods introduced. We ask for your indulgence on this point – as indicated, we are more concerned with the practical aspects and do not deal with the theoretical aspects in this book – the reader should refer to the texts in the literature to study the mathematical arguments underlying the methods discussed.

Contents

1 Basic Bayesian Probabilities 1
1.1 A Historical Perspective 1
1.2 Probabilities 6
1.3 Representing a Probability Space in R 15
1.4 Independent Events and Bayes' Formula 33
1.5 Representation of Product Spaces 36
1.6 Random Variables 48
1.7 The Exponential Family 63
Classical Discrete Distributions of the Exponential Family 65
Classical Continuous Distributions 66
1.8 Distributions That Are Not in the Exponential Family 69
1.9 Random Vectors 71
1.10 Independent Random Variables 99
1.11 Central Limit and Levy's Theorem 104
1.12 Exchangeability 107
1.13 Stochastic Processes 124
References 127

2 Beliefs 133
2.1 Beliefs 133
2.2 Working with Beliefs in R 139
2.3 Commonality 149
2.4 Retrieving Mass Functions from Other Representations 151
2.5 Couples, Product Frames, and Marginality 156
2.6 Contour Functions 174
2.7 Combination of Beliefs 180
2.8 Conditional Beliefs, Plausibilities, and Frames 191
2.9 Connections Between Probabilities and Beliefs 196
References 198

3 Information and Entropy 203
3.1 Entropy 204
3.2 Joint and Conditional Entropy 228
3.3 Mutual Information and Relative Entropy 236
3.4 Surprise Quantification 250
References 259

4 Maximum Entropy 265
4.1 The Principle of Maximum Entropy 265
4.2 Karhunen-Loève Expansions for the Numerical Generation of Stochastic Processes 282
4.3 Numerical Construction of Karhunen-Loève Expansions 291
References 313

5 Bayesian Inference 321
5.1 Bayesian Estimation 321
5.2 Jeffreys' Priors 331
5.3 Using UQ in Bayesian Estimation 340
5.4 Bayesian Regression 366
5.5 Model Selection and Hypothesis Testing 375
5.6 The Expectation-Maximization Algorithm 391
References 402

6 Sequential Bayesian Estimation 413
6.1 Monte-Carlo Approach 414
6.2 Metropolis-Hastings Algorithm 422
6.3 Kalman Filtering 429
6.4 Particle Filtering 441
6.5 Determination of the Distribution of the Noise 447
6.6 Bayesian Optimization 459
References 468

Index 481

Chapter 1
Basic Bayesian Probabilities

Bayesian probabilities can be seen as a particular point of view on probabilities. In this chapter, we recall the basic elements of probability theory that will be used in the book, and we give some elements about the Bayesian point of view about probabilities.

1.1 A Historical Perspective

The fundamental text for Bayesian probabilities is the article "An Essay Towards Solving a Problem in the Doctrine of Chances," conceived by Thomas Bayes and published posthumously by his friend Richard Price in *Philosophical Transactions of the Royal Society* (Bayes and Price 1763). In his text, Price says that Bayes formulates the following problem:

> PROBLEM.
>
> *Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a ſingle trial lies ſomewhere between any two degrees of probability that can be named.

E. Souza de Cursi, *Uncertainty Quantification with R*, International Series in Operations Research & Management Science 352,
https://doi.org/10.1007/978-3-031-48208-3_1

In terms of modern probability, Bayes' problem reads as

Problem 1.1 Let be given a sample $\mathcal{X} = (X_1, \ldots, X_n)$ of n variates from a Bernoulli distribution $B(p)$ $(P(X_i = 1) = p, P(X_i = 0) = 1 - p)$. Let be given $0 \leq a < b \leq 1$. Assuming that $\sum_{i=1}^{n} X_i = k$, determine the probability of the event $p \in (a, b)$, conditionally to the observed data.

A complete analysis of Bayes' solution can be found in Stigler (1982). It is based on two propositions: on the one hand,

> P R O P. 3.
>
> The probability that two ſubſequent events will both happen is a ratio compounded of the probability of the 1ſt, and the probability of the 2d on ſuppoſition the 1ſt happens.

i.e., in terms of modern probability:

$$P(E_1 \cap E_2) = P(E_2 | E_1) P(E_1). \tag{1.1}$$

On the other hand,

> P R O P. 5.
>
> If there be two ſubſequent events, the probability of the 2d $\frac{b}{N}$ and the probability of both together $\frac{P}{N}$, and it being 1ſt diſcovered that the 2d event has happened, from hence I gueſs that the 1ſt event has alſo happened, the probability I am in the right is $\frac{P}{b}$†.

i.e., in terms of modern probability (see Sect. 1.3):

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}. \tag{1.2}$$

Combining these two expressions, we obtain the classical equality known as ***Bayes' formula*** (see Sect. 1.3):

$$P(E_1 | E_2) = \frac{P(E_2 | E_1)}{P(E_2)} P(E_1). \tag{1.3}$$

In modern probability, Bayes' formula is used to show that the solution for Problem 1.1 verifies

$$P(p \in (a,b)|\boldsymbol{\mathcal{X}}) = \int_a^b P(p|\boldsymbol{\mathcal{X}})dp, \tag{1.4}$$

with

$$P(p|\boldsymbol{\mathcal{X}}) = \frac{P(\boldsymbol{\mathcal{X}}|p)P(p)}{P(\boldsymbol{\mathcal{X}})}, \quad P(\boldsymbol{\mathcal{X}}) = \int_{\mathbf{0}}^{\mathbf{1}} P(\boldsymbol{\mathcal{X}}|p)P(p)dp. \tag{1.5}$$

The practical use of Eq. (1.5) requires the knowledge of $P(\boldsymbol{\mathcal{X}}|p)$ (usually referred to as ***Likelihood***) and $P(p)$ (usually referred to as ***Prior Distribution***) . In practice, Likelihood yields from a model, while the Prior must be chosen by the user, according to his assumptions or beliefs about p – except in particular situations. In the case studied by Bayes, an analogy – often called *Bayes' billiard* – is used to conclude that $P(p) = 1$, which corresponds to a uniform distribution, while the model for the Likelihood is binomial (Bayes and Price 1763). As a consequence, we have, in modern probabilistic language,

$$P(p|\boldsymbol{\mathcal{X}}) = \frac{\binom{k}{n} p^k (1-p)^{n-k}}{\int_0^1 \binom{k}{n} p^k (1-p)^{n-k} dp}. \tag{1.6}$$

This expression can be written in a simplified form as

$$P(p|\boldsymbol{\mathcal{X}}) = A\, p^k (1-p)^{n-k} \quad , \tag{1.7}$$

where A is a constant such that

$$A \int_0^1 p^k (1-p)^{n-k} dp = 1 \Rightarrow \mathrm{A} = \frac{(n+1)!}{k!(n-k)!}. \tag{1.8}$$

The work of Bayes remained ignored for a long time (a complete history can be found in (Dale 1999)). In 1774, Pierre-Simon Laplace published a memoir entitled *Mémoire sur la probabilité des causes par les événements* (Laplace 1774, 1891). In this work, Laplace considers the following problem:

Dans le Problème ſuivant, *une urne étant ſuppoſée renfermer un nombre donné de billets blancs & noirs dans un rapport inconnu, ſi l'on tire un billet & qu'il ſoit blanc, déterminer la probabilité que le rapport des billets blancs aux noirs eſt celui de* p *à q;* l'évènement eſt connu & la cauſe inconnue.

S.M. Stiegler proposes the following translation (Laplace 1986):

In the following problem, *an urn is supposed to contain a given number of white and black tickets in an unknown ratio; if one draws a ticket and finds it white, determine the probability that the ratio of white to black tickets is that of p to q.* The event is known and the cause is unknown.

The problem considered by Laplace is analogous to the problem introduced by Bayes. Nevertheless, Laplace did not make any reference to Bayes – we can suppose that he was not aware of Bayes' work. Laplace states a principle for the solution:

Si un évènement peut être produit par un nombre *n* de cauſes différentes, les probabilités de l'exiſtence de ces cauſes priſes de l'évènement, ſont entre elles comme les probabilités de l'évènement priſes de ces cauſes, & la probabilité de l'exiſtence de chacune d'elles, eſt égale à la probabilité de l'évènement priſe de cette cauſe, diviſée par la ſomme de toutes les probabilités de l'évènement priſes de chacune de ces cauſes.

S.M. Stiegler proposes the following translation (Laplace 1986):

If an event can be produced by a number *n* of different causes, the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes.

In modern probability writing: Laplace considers n possible distinct causes C_i, $1 \leq i \leq n$ and an event : $C_i \cap C_j = \emptyset$, for $i \neq j$ and $\sum_{j=1}^{n} P(C_j) = 1$. Then, for an event E,

$$\frac{P(C_i|\mathrm{E})}{P(C_j|\mathrm{E})} = \frac{P(E|C_i)}{P(E|C_j)} \tag{1.9}$$

and (see Sect. 1.3)

$$P(C_i|\mathrm{E}) = \frac{P(E|C_i)}{\sum_{j=1}^{n} P\big(E|C_j\big)}. \tag{1.10}$$

Notice that $P(E) = \sum_{j=1}^{n} P\big(E|C_j\big)P\big(C_j\big)$ and $P(E \cap C_i) = P(E|C_i)P(C_i)$, so that Eq. (1.10) is equivalent to Eq. (1.2) when all the causes are equiprobable, i.e., $P(C_j) = 1/n, \forall j$. Indeed, Laplace assumes the equiprobability of the causes implicitly. A complete analysis of Laplace's work can be found in Dale (1982) or Dale (1999). A generalization to the actual form (1.10) appears in the works of Augustus de Morgan, William Fishburn Donkin, William Allen Whitworth, and Mathieu Paul Hermann Lauren (see Dale (1999)).

By 1920, the work of Bayes was rediscovered, namely, by Karl Pearson and Ronald Aylmer Fisher. The rediscovery generated a wide and polemical controversy about the use of a uniform distribution for p, justified by Bayes using the analogy of the billard.

One of the main criticisms was that such a choice is not invariant by transformation, so that it may lead to contradictions, if applied without precaution. Indeed, the criticism was not so much directed at Bayes' work as at the formulation that, in the absence of any information except the possible values of a quantity, one must consider a uniform distribution on these values. If such a principle is applied without adequate care, it can lead to contradictions, illustrated by a simple example: if a quantity X is uncertain and we do not have any information other than $X \in \left(1/_a, a\right)$, with $a > 1$, then the principle leads to

$$P(X \le 1) = \frac{1 - \frac{1}{a}}{a - \frac{1}{a}} = \frac{1}{1 + a}.$$

However, $Y = \frac{1}{X}$ is also uncertain and $Y \in \left(1/_a, a\right)$. Thus, the same principle leads to

$$P(Y \ge 1) = \frac{a - 1}{a - \frac{1}{a}} = \frac{a}{1 + a} > \frac{1}{1 + a} = P(X \le 1).$$

These results may appear to be contradictory, since $X \le 1 \Leftrightarrow Y \ge 1$. Such a difficulty generated a controversy, which turned into a philosophical dispute about the meaning of probability, with two extreme standpoints: on the one hand, the "frequentist" one, where probabilities are the limits of relative frequencies or the result of enumerations; on the other hand, the "subjectivist" one, where probabilities are measurements degrees of belief. A highlight in this debate is the sentence of Bruno de Finetti (see (De Finetti 2017), p. xv) :

My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this:

PROBABILITY DOES NOT EXIST

The abandonment of superstitious beliefs about the existence of Phlogiston, the Cosmic Ether, Absolute Space and Time, . . ., or Fairies and Witches, was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs.

Nowadays, many researchers (but not all!) regard the controversy as mainly obsolete and consider that they have at their disposal tools having as origin Bayes, Fisher, Pearson, etc. For these researchers, these tools form a toolbox from which they can choose what is necessary for their studies according to the circumstances.

Such a point de view is not universal, and the discussion goes on about the philosophical aspects and interpretations of probability – see, for instance, Howson and Urbach (2006), Jeffreys (1939), Savchuk and Tsokos (2011), Press and Tanur (2001), and Jaynes (1989, 2003).

The reader will certainly have his own point of view or will build it from his readings and his own thinking. We will not discuss these aspects in the following, and we will concentrate on the methods and their practical use, namely, with R.

1.2 Probabilities

Uncertainty can be modeled by different ways. For instance, we can use fuzzy variables to classify objects or individuals according to classes having imprecise limits. We can also model uncertain numerical variables as intervals and use interval analysis to manipulate them.

A classical model for uncertain quantities is the model of random variables, based on the classical definition of probabilities:

Definition 1.1 Let Ω be a nonempty set and $\mathcal{P}(\Omega) = \{E : E \subset \Omega\}$ be the set of the parts of Ω. A ***probability*** defined on Ω is an application $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ such that

1. $P(E) \geq 0, \forall E \in \mathcal{P}(\Omega)$;
2. $P(\Omega) = 1$;
3. $P\left(\bigcup_{n \in \mathbb{N}} E_n\right) = \sum_{n \in \mathbb{N}} P(E_n), \forall \{E_n : n \in \mathbb{N}\}$ *such that* $E_i \cap E_j = \emptyset$, *if* $i \neq j$. ∎

In the context of probability, Ω is the ***universe*****,** and (Ω, P) is referred to as a ***probability space***. A subset $E \subset \Omega$ is an ***event***. From this definition, we can state the standard properties of probabilities, namely,

$$P(\varnothing) = 0; \tag{1.11}$$

$$P(E - F) = P(E) - P(E \cap F); \tag{1.12}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B); \tag{1.13}$$

$$A \subset B \Longrightarrow P(A) \leq P(B). \tag{1.14}$$

Definition 1.2 We say that $E \subset \Omega$ is ***almost sure (a.s.)*** if and only if $P(E) = 1$. We say that $E \subset \Omega$ is ***almost impossible (a.i.)*** or ***negligible*** if and only if $P(E) = 0$. ∎

Probabilities can be defined by **probability mass functions** on finite or enumerable universes, while **probability mass densities** are used on intervals of real numbers:

Definition 1.3

1. Let Ω be finite or enumerable. A **probability mass function** $\boldsymbol{\mu}$ on Ω is an application $\mu: \Omega \longrightarrow \mathbb{R}$ such that $\forall \omega \in \Omega: \mu(\omega) \geq 0$ and $\sum_{\omega \in \Omega} \mu(\omega) = 1$. The probability generated by μ is given by $P(\{\omega\}) = \mu(\omega)$.
2. Let $\Omega \subset \mathbb{R}$. A **probability mass density** $\boldsymbol{\mu}$ on Ω is an application $\mu: \Omega \longrightarrow \mathbb{R}$ such that $\mu(\omega) \geq 0$ on Ω and $\int_{\Omega} \mu(\omega) d\omega = 1$. We define a probability on Ω by considering $P((a, b)) = \int_a^b \mu(\omega) d\omega, \forall (a, b) \subset \Omega$. ∎

The extension to other subsets of Ω is made using condition 3 in Definition 1.1, the classical properties of probabilities, namely, Eqs. (1.12 and 1.13). For instance, when $\Omega = \{\omega_1, \ldots, \omega_n\}$ and

$$\mu(\omega_i) = p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1. \tag{1.15}$$

Then,

$$P(\{\omega_{i_1}, .., \omega_{i_k}\}) = \sum_{j=1}^{k} \mu\left(\omega_{i_j}\right). \tag{1.16}$$

R proposes functions for the manipulation of mass functions and mass densities. For instance, sum and integrate. The package pracma emulates Matlab® and proposes functions quad, romberg for the evaluation of integrals.

Example 1.1 Let us consider $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $\mu(\omega_i) = \beta e^{-\alpha i}$. Let us determine the possible values of α and β. We have

$$\mu(\omega_i) \geq 0 \Leftrightarrow \beta \geq 0$$

and

$$\sum_{i=1}^{n} \mu(\omega_i) = 1 \Leftrightarrow \beta \sum_{i=1}^{n} e^{-\alpha i} = 1.$$

Thus,

$$\beta = \frac{1}{\sum_{i=1}^{n} e^{-\alpha i}}$$

We can plot the possible values as follows:

```
n = 6 # number of elements of Omega
Omega <- c(seq(1,n)) # universe
alfa = seq(-5,5,length.out = 1001)  # values of alpha
betta = rep(0, length(alfa))
for (i in seq(1,length(alfa))){betta[[i]]= 1/sum(exp(-alfa
[[i]]*Omega))}

str2 = substitute(paste("possible values of ( ", alpha,",
",beta," ), for n = ",v),list(v = n))
plot(alfa, betta,type = 'n', xlab = expression(alpha),ylab
= expression(beta), main = str2)
lines(alfa,betta,col='blue',lwd=2)
```

(continued)

Example 1.1 (continued)

The results are shown in Fig. 1.1. ∎

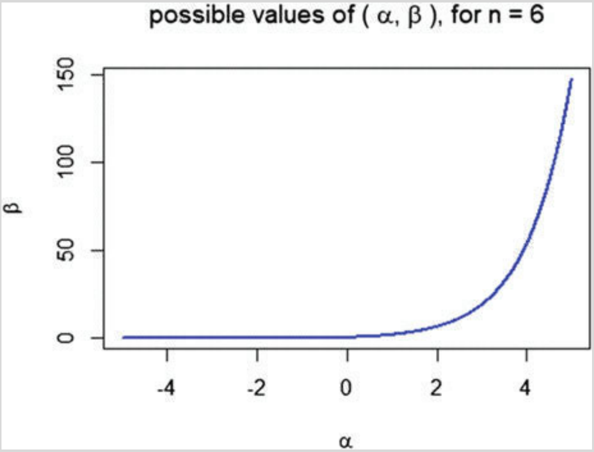

Fig. 1.1 Determination of the coefficients in the mass function

Example 1.2 Let us consider $\Omega = (0, 1) \subset \mathbb{R}$ and $\mu(\omega) = \beta e^{-\alpha\omega}$. Let us determine the possible values of α and β. We have

$$\mu(\omega) \geq 0 \Leftrightarrow \beta \geq 0$$

And, for $\alpha \neq 0$:

$$\int_0^1 \mu(\omega) d\omega = 1 \Leftrightarrow \beta \frac{1 - e^{-\alpha}}{\alpha} = 1.$$

Thus,

$$\beta = \frac{\alpha}{1 - e^{-\alpha}} \quad (\alpha \neq 0).$$

If $\alpha = 0$, then $\mu(\omega) = \beta$, so that we must have $\beta = 1$. We can plot the possible values as follows:

(continued)

Example 1.2 (continued)

```
alfa = seq(-5,5,length.out = 1001)  # values of alpha
betta = rep(1, length(alfa))
for (i in seq(1,length(alfa))){if (alfa[[i]] != 0) {betta[
[i]]= alfa[[i]]/(1 - exp(-alfa[[i]]))}}

str2 = expression(paste("possible values of ( ", alpha,",
",beta," )"))
plot(alfa, betta,type = 'n', xlab = expression(alpha),ylab
= expression(beta), main = str2)
lines(alfa,betta,col='blue',lwd=2)
```

The results are shown in Fig. 1.2. ∎

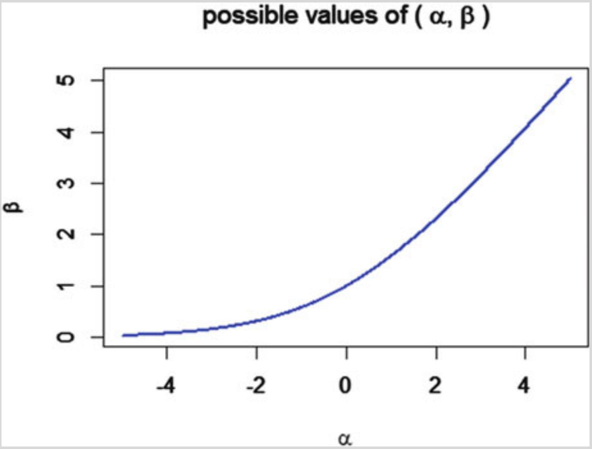

Fig. 1.2 Determination of the coefficients in the mass density

When $\Omega = \{\omega_1, \ldots, \omega_n\}$ is finite, a popular mass function is the uniform mass function: $\mu(\omega) = 1/n, \forall\, \omega \in \Omega$. In such a situation, probabilities are directly given by the number of elements of the subsets: $P(A) = |A|/n$, where $|A|$ is the number of elements of A. Its analogous to $\Omega = (a, b) \subset \mathbb{R}$ is the uniform mass density $\mu(\omega) = \frac{1}{b-a}, \forall \omega \in (a, b)$. Then, for $a \leq \alpha \leq \beta \leq b$, $P((\alpha, \beta)) = \frac{\beta - \alpha}{b - a}$.

Example 1.3 In a 52-card deck of playing cards, you draw simultaneously three at random. Let us evaluate the probability of getting three diamonds: here, the universe $\Omega = \{(C_1, C_2, C_3)$: C_i: *card drawn at* $i\}$. Assuming a

(continued)

Example 1.3 (continued)

uniform mass function on Ω, we must evaluate, on the one hand, the number of elements of Ω, which is $n = 52 \times 51 \times 50$. On the other hand, we must evaluate the number of elements in $A = \{(C_1, C_2, C_3)\colon C_i\colon \textit{card drawn at i is a diamond}\}$, which is $|A| = 13 \times 12 \times 11$, so that

$$P(3\ diamonds\ ) = \frac{13 \times 12 \times 11}{52 \times 51 \times 50.} = \frac{1716}{132600} \approx 0.01294$$

The evaluation can be made by R as follows:

```
library(arrangements)
k = npermutations(13,3)
n = npermutations(52,3)
pA = k/n
print(paste("P(A) = ",pA))
```

The result is

```
## [1] "P(A) =  0.0129411764705882"
```

Now let us evaluate the probability of getting three cards of the same color: in this case, we must evaluate the number of elements in $= \bigcup_{i=1}^{4} A_i\, A_i = \{(C_1,$ $C_2, C_3)\colon C_j\colon \textit{card drawn at j is } c_i\}$, c_i is one of the four colors: clubs, diamonds, hearts, and spades. Thus, $|B| = 4 \times |A| = 4 \times 13 \times 12 \times 11$ and

$$P(3\ of\ \ same\ \ color) = \frac{4 \times 13 \times 12 \times 11}{52 \times 51 \times 50.} = \frac{6864}{132600} \approx 0.05176.$$

Using R:

```
k                    =                    4*npermutations(13,3)
n                     =                     npermutations(52,3)
pB                            =                             k/n
print(paste("P(B) = ",pB))
```

The result is

```
## [1] "P(B) =  0.0517647058823529"
```

Example 1.4 Let $\Omega = (-1, 1)$ with the uniform mass density. Then, $\mu(\omega) = \frac{1}{2}$. Let $A_\alpha = \{\omega \in \Omega\colon \ \log |\omega| \leq \ -\alpha\}$. Since

$$A_\alpha = \{\ \omega \in \Omega : |\omega| \leq e^{-\alpha}\} = \begin{cases} \Omega, & if\ \alpha \leq 0 \\ (-e^{-\alpha}, e^{-\alpha}), & if\ \alpha \geq 0 \end{cases},$$

We have

$$P(\mathrm{A}_\alpha) = \begin{cases} 1, & if\ \alpha \leq 0 \\ e^{-\alpha}, & if\ \alpha \geq 0 \end{cases}.$$

We can evaluate these probabilities using R:

```
a <<- -1   # global variable
b <<- 1   # global variable
mu = function(omega){
  aux = 1/(b-a)
  return(aux)
}
alfa = seq(0,2, length.out = 201)
p_ex = exp(-alfa)
pcalc = numeric(length = length(alfa))
library(pracma)
for (i in seq(1,length(alfa))){
  om = exp(-alfa[[i]])
  pcalc[[i]] = quad(mu,-om,om)
}
str1 = expression(paste('P(A'[alpha],')'))
str2 = expression('probability of A'[alpha])
plot(alfa,pcalc,type='n',xlab = expression(alpha),ylab = s
tr1,
     main = str2)
lines(alfa,p_ex,col='blue',lwd = 2)
lines(alfa,pcalc,lwd=5,lty=2,col='red')
legend(x = "topright",legend = c("exact", "calculated"),
       lty = c(1, 2),col = c('blue', 'red'),lwd = c(2,3))
```

(continued)

Example 1.4 (continued)

The result appears in Fig. 1.3.

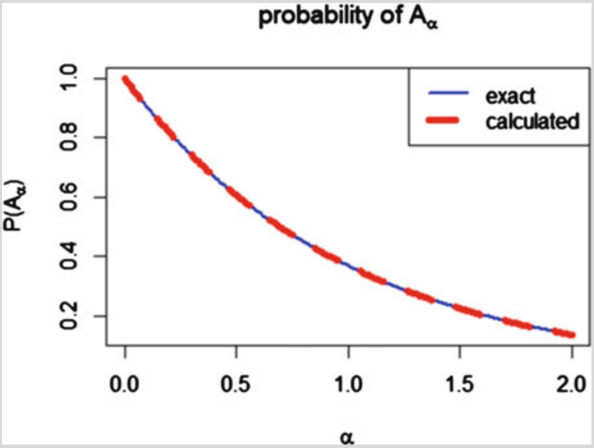

Fig. 1.3 Determination of the probabilities $P(A_\alpha)$

Now, let us consider $B_\alpha = \{\omega \in \Omega\colon \omega^2 - (\alpha + 1)\omega + \alpha \le 0\}$. Since the roots of the polynomial are α and 1, we have

$$B_\alpha = \begin{cases} \Omega, \text{if } \alpha \le -1 \\ (\alpha, 1), \text{if } -1 \le \alpha \le 1, \\ \varnothing, \text{if } \alpha > 1 \end{cases}$$

so that

$$P(B_\alpha) = \begin{cases} 1, \text{if } \alpha \le -1 \\ \frac{1-\alpha}{2}, \text{if } -1 \le \alpha \le 1 \\ 0, \text{if } \alpha > 1 \end{cases}.$$

Again, we can evaluate these probabilities using R:

```
alfa = seq(-1,1, length.out = 201)
p_ex = (1 - alfa)/2
pcalc = numeric(length = length(alfa))
library(pracma)
for (i in seq(1,length(alfa))){
  pcalc[[i]] = quad(mu,alfa[[i]],1)
}
```

(continued)

Example 1.4 (continued)

The result appears in Fig. 1.4.

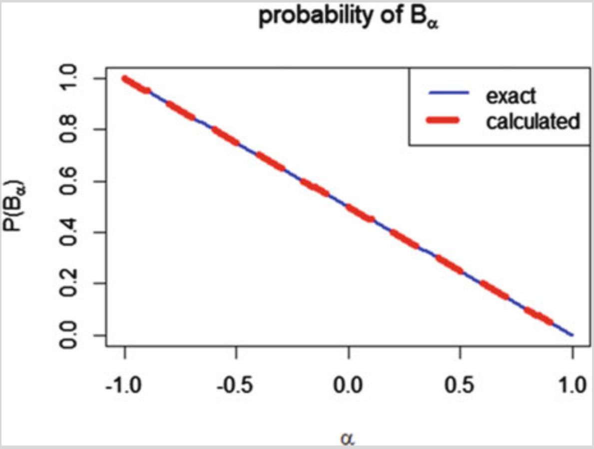

Fig. 1.4 Determination of the probabilities $P(B_\alpha)$

Exercises

1. Consider $\Omega = (0, 1)$ and the mass density $\mu(\omega) = \alpha \ln(2 + \omega) + \beta$. What are the possible values of α and β? Plot the admissible region.
2. Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and the mass function $\mu(\omega_i) = \alpha \ln(1 + i) + \beta$. What are the possible values of α and ? Plot the admissible region.
3. Consider $\Omega = (0, 1)$ and the mass density $\mu(\omega) = \alpha\omega^2 + \beta$. What are the possible values of α and ? Plot the admissible region.
4. Consider $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and the mass function $\mu(\omega_i) = \alpha i + \beta$. What are the possible values of α and ? Plot the admissible region.
5. Consider $\Omega = (0, 1)$ and the mass density $\mu(\omega) = 3\omega^2$. Find the probabilities of the following events:

   (a) $A = \left\{\omega \in \Omega : \omega < \frac{1}{2}\right\}$.
   (b) $B = \left\{\omega \in \Omega : \omega^2 < \frac{1}{2}\right\}$.

(continued)

(c) $C = \{\omega \in \Omega\colon e^{\omega} < 2\}$.
(d) $D = \{\omega \in \Omega\colon 5\omega^2 - 9\omega + 1 \leq 0\}$.
(e) Write programs in R to find these probabilities.

6. Consider $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $\mu(\omega_i) = \frac{2i}{n(n+1)}$.

   (a) Find $P(\{\omega_1, \omega_n\})$.
   (b) Find $P(\{\omega_2, \omega_{n-1}\})$.
   (c) Write programs in R to determine these probabilities.

7. Consider $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and the events $A = \{\omega_1\}$, $B = \{\omega_1, \omega_2\}$, $C = \{\omega_1, \omega_2, \omega_4\}$. Let $P(A) = \frac{1}{5}$, $P(B) = \frac{3}{10}$, $P(C) = \frac{7}{10}$.

   (a) Find $P(\{\omega_2\})$, $P(\{\omega_3\})$, $P(\{\omega_4\})$.
   (b) Find $P(\{\omega_1, \omega_2, \omega_3\})$.

8. Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and the events $A = \{\omega_1\}$, $B = \{\omega_1, \omega_2\}$. Let $P(A) = \frac{1}{4}$ and $P(B) = \frac{1}{2}$.

   (a) Find $P(\{\omega_2\})$.
   (b) Find $P(\{\omega_3\})$.
   (c) Find $P(\{\omega_1, \omega_3\})$.

9. In a 52-card deck of playing cards, you draw sequentially four at random. At each draw, the card is put back in the deck, which is reshuffled. Find the probabilities of getting

   (a) Four spades.
   (b) Four cards of the same color.
   (c) Four aces.
   (d) At least three aces.

10. An urn contains 40 numbered balls of three colors: 12 red, 20 green, and 8 blue. We draw at random two balls simultaneously. Find the probability of getting

   (a) Two green balls.
   (b) Two balls of the same color.
   (c) Two balls of different colors.

1.3 Representing a Probability Space in R

When Ω is finite, we can represent the probability space using the class `dps` (discrete probability space) below:

```
library(R6)
#
dps=R6Class("dps",
            public=list(
              m = NULL,
              names = NULL,
              size = NULL,
              initialize=function(n,mu,nams){
                if (missing(n) == TRUE){
                  self$size = 2
                }else{self$size <<- n}
                if (missing(mu) == TRUE){
                  self$m <<- rep(1/self$size,self$size)
                }else{self$m <<- mu}
                if (missing(nams) == TRUE){
                  self$names <<- as.character(seq(1:self$siz
e))
                }else{self$names <<- nams}
              },
              prob= function(ev,ps){
                if (length(ev) == 0){p = 0}
                else{
                  ev1 = unique(ev)
                  p = 0
                  if (missing(ps) == TRUE){
                    for (i in ev1){p = p + self$m[[i]]}
                  }else{
                    for (i in ev1){p = p + ps$m[[i]]}
                  }
                  return(p)
                }
              },
              print_events_values = function(events,values,v

 arnames, valname){
                nev = length(events)
                mp = matrix(nrow = nev, ncol = 2)
                for (i in 1:nev){
                  aux = events[[i]]
                  ele = length(aux)
                  nam = paste("{",varnames[[aux[[1]]]])
                  if (ele > 1){
                    for (j in 2:ele)
                      nam = paste(nam,",",varnames[[aux[[j]]
]])
```

```
          }
          nam = paste(nam,"}")
          mp[ i,1] = nam
          mp[ , 2] = values[[i]]
        }
        colnames(mp) = c("event",valname)
        mpd = as.data.frame(mp)
        print(mpd)
      },
      show = function(ps){
        if (missing(ps) == TRUE){
          events= seq(1:self$size)
          values = self$m
          varnames = self$names
        }else{
          events= seq(1:ps$size)
          values = ps$m
          varnames = ps$names
        }
        valname = "prob"
        self$print_events_values(events,values,varna
mes, valname)
      }
    )
)
```

These contents must be saved to a file `dps.R`, to be sourced before use. For this class, $\Omega = \{1, 2, \ldots, n\}$, but you can give names to the elements. If the mass function `mu` is not given, the mass is uniformly distributed on the elements of Ω. If the names are not given, they are set to "1," "2," ... R proposes intrinsic functions `intersect`, `union`, `setdiff` which can be used to determine the probabilities associated with intersections, unions, etc.

Example 1.1 Let us consider the result when a coin is tossed once, which can be Head (H) or Tail (T). Then, $\Omega = \{H, T\}$. Let μ be a mass function on Ω and $p = \mu(H)$, $q = \mu(T)$. Then: $p \geq 0$, $q \geq 0$, $p + q = 1$. The coin is fair when $p = q$, i.e., when $p = q = 1/2$.

We can create a probability space for a fair coin as follows:

```
source("dps.R")
ps = dps$new(2)
ps$names = c("H","T")
```

(continued)

Example 1.1 (continued)

We can see the probabilities of the elements of Ω as follows:

```
ps$show()
##    event prob
## 1 { H }  0.5
## 2 { T }  0.5

T = 2
ps$prob(T)
## [1] 0.5
```

```
a = c(1,2)
ps$prob(a)
## [1] 1

ps$prob(intersect(a,T))
## [1] 0.5

ps$prob(union(a,T))
## [1] 1
```
■

Example 1.2 Let us consider the result when a dice having n faces is rolled in a single die. Here, $\Omega = \{1, 2, \ldots, n\}$. Let μ be a mass function on Ω and $p_i = \mu(i)$, $1 \le i \le n$. Then: $p_i \ge 0$, $1 \le i \le n$; $p_1 + \ldots + p_n = 1$. The dice is balanced when all the sides have the same probability – all the p_i are equal, i.e., when $p_i = 1/n$, $1 \le i \le n$. An example of determination of probabilities using R is:

```
source("dps.R")
n = 16 # number of faces

ps = dps$new(n)
```

```
A = c(1,6, 12)
pA = ps$prob(A)

B = c(1, 3, 5, 7)
pB = ps$prob(B)
```

The results are

```
print(paste("P(A) = ",pA," ; P(B) = ", pB))
## [1] "P(A) =  0.1875  ; P(B) =  0.25"
```
■

Alternative representations of probability spaces are the binary and hexadecimal representations.

In the binary representation, the events $E \subset \Omega$ correspond to binary vectors of dimension n having all the components equal to zero, except those corresponding to the elements of E. For instance, let $n = 5$: the event $\{\omega_1, \omega_3\}$ is represented by the binary vector $(1, 0, 1, 0, 0)$, the event $\{\omega_2, \omega_3, \omega_5\}$ is represented by the binary vector $(0, 1, 1, 0, 1)$, and so on. The universe Ω is represented by a vector entirely formed by ones and the empty set ∅ by a vector of zeros. Under the binary representation, the intersection of two events corresponds to the product term by term of the binary vectors or the logical operation "AND." The union of two subsets can be achieved by considering the term-by-term maximum of the vectors or the logical operator "OR."

The difference $A - B$ can be evaluated by subtracting $A \cap B$ from A – the complementary set $\Omega - A$ can be evaluated by subtracting A from the universe vector formed by ones. These operations can be implemented in a class `bds` (*binary discrete set*), defined as follows:

```
library(R6)
#
bds=R6Class("bds",
 public=list(
 initialize=function(){
 },
 binter= function(ev1,ev2){
  result = pmin(ev1,ev2)
  return(result)
 },
 bunion= function(ev1,ev2){
  result = pmax(ev1,ev2)
  return(result)
 },
 bdiff= function(ev1,ev2){
  result = ev1 - pmin(ev1, ev2)
  return(result)
 },
 bcomplement = function(ev1){
  result = rep(1,length(ev1)) - ev1
  return(result)
 },
 bflip = function(ev){
  ind = (length(ev) + 1) - seq(1:length(ev))
  aux = ev[ind]
  return(aux)
 },
 isequal = function(ev1,ev2){
   # TRUE if ev1 is equal to ev2
   aux = sum(abs(ev1-ev2))
   if (aux > 0){r = FALSE}
   else{r = TRUE}
   return(r)
 },
 iscontained = function(ev1,ev2){
   # TRUE if ev1 is part of ev2
   ev = self$binter(ev1,ev2)
   r = self$isequal(ev,ev1)
   return(r)
 },
 contains = function(ev1,ev2){
   # TRUE if ev1 contains ev2
   ev = self$bunion(ev1,ev2)
   r = self$isequal(ev,ev1)
   return(r)
 }
 )
)
```

Example 1.3 Let us consider $n = 5$ and the events

$$A = \{\omega_1, \omega_3\}, B = \{\omega_2, \omega_3, \omega_5\}, C = \{\omega_1, \omega_3, \omega_4\},$$

Their binary representation is

```
A = c(1,0,1,0,0)
B = c(0,1,1,0,0)
```

```
C = c(1,0,1,1,0)
```

Let us evaluate $A \cap B$, $A \cup B$, $A - B$, $B - A$, $\Omega - A$, $\Omega - B$, and verify the inclusions in C.

```
source("bds.R")

bs = bds$new()
A
## [1] 1 0 1 0 0

bs$bflip(A)
## [1] 0 0 1 0 1

B
## [1] 0 1 1 0 0

bs$bflip(B)
## [1] 0 0 1 1 0

bs$binter(A,B)
## [1] 0 0 1 0 0

bs$bunion(A,B)
## [1] 1 1 1 0 0

bs$bdiff(A,B)
```

```
## [1] 1 0 0 0 0

bs$bdiff(B,A)
## [1] 0 1 0 0 0

bs$bcomplement(A)
## [1] 0 1 0 1 1

bs$bcomplement(B)
## [1] 1 0 0 1 1

bs$contains(C,A)
## [1] TRUE

bs$contains(C,B)
## [1] FALSE

bs$iscontained(A,C)
## [1] TRUE

bs$iscontained(B,C)
## [1] FALSE
```
■

In the binary representation, each $\omega_i \in \Omega$ is represented by a one-hot vector, i.e., a binary vector having a single element equal to one. Thus, we can bring the mass function to an Identity Matrix: each line represents one element of Ω. The probability space corresponding to the binary representation is implemented in a class `bps` (binary probability space), defined as follows – notice that class `bps` inherits from class `bds`:

```
library(R6)
source("bds.R")
#
bps=R6Class("bps",
  inherit=bds,
  public=list(
   m = NULL, names = NULL,
   size = NULL,frame = NULL,
   initialize=function(n,mu,nams){
     super$initialize()
     if (missing(n) == TRUE){
       self$size = 2
     }else{self$size <- n}
     if (missing(mu) == TRUE){
       self$m <- rep(1/self$size,self$size)
     }else{self$m <- mu}
    if (missing(nams) == TRUE){
       self$names <- as.character(seq(1:self$size))
     }else{self$names <- nams}
        self$frame <- diag(self$size)
   },
  bprob= function(ev,ps){
   # evaluates the probability of ev
   # in the binary probability space ps
   if (missing(ps) == TRUE){ps = self}
   ind = which(ev > 0)
   if (length(ind) == 0){p = 0}
   else{
     p = 0
     for (i in ind){p = p + ps$m[[i]]}
   }
   return(p)
  },
  bevent = function(nbs,ps){
   # creates an event in the binary probability
   # space ps using the positions given in nbs
   if (missing(ps) == TRUE){ps = self}
   res = rep(0,ps$size)
   if (length(nbs) > 0){res[nbs] = 1}
   return(res)
  },
```