Unimodal density estimation with applications in expert elicitation and decision making under uncertainty

Yacov Salomon

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy
March 2013

Department of Mathematics and Statistics
The University of Melbourne

(archival quality paper)
Abstract

A wide range of applied problems may be reduced to an under-determined inverse problem of the form: find the “best” density $f$ subject to a set of constraints. The problem of choosing the best distribution to represent the data elicited from an expert is one such example. Expert elicitation is a common practice in a range of applications where data regarding a quantity of interest is unavailable or uncertain.

To solve for an optimal density, one of a range of information-theoretic, or norm related, measures is introduced over the feasible set. A particular difficulty arises when some constraints cannot be expressed as moments. Such is the case when the density is known to be unimodal, and its unique mode is given.

In this thesis, we formulate and solve the problem of unimodal density estimation when using the minimum cross entropy measure. In particular, using several previous results by Khinchin, Shapp, and Kempermann we formulate the general nonlinear optimisation problem, and then solve it in two steps.

The first step studies a relaxed form of the general problem. Necessary and sufficient conditions under which solutions of the relaxed problem results in unimodal densities are derived. In the second step, the general optimal solutions are obtained, by establishing convexity properties, and using duality and variational calculus methods.

The problem of choosing the best distribution to represent the data elicited from an expert is formulated as an inverse problem using the minimum cross entropy measure. Throughout the thesis, the analytical results for unimodal density estimation are illustrated in the context of expert elicitation using the 4-step elicitation protocol. This protocol elicits information corresponding to the mode and an inter-percentile of the distribution representing the expert’s opinion.

Finally, the general problem of decision making under conditions of uncertainty is examined. A method is proposed, combining previously suggested safety first and portfolio approaches, with a satisficing objective. Heuristically, the objective advocated is to choose the mixture of actions that maximises the probability of obtaining a favourable outcome. This objective is equivalent to maximising a utility, defined as the indicator function over the favourable outcome. The expert elicitation framework is used to construct subjective probabilities over the possible states of nature (parametric state space). The method is applied in the context of environmental management using two case studies, the management of a koala population, and the conservation of a marine turtle population.

The thesis is concluded by giving future research tasks and directions.
Declaration

This is to certify that

1. The thesis comprises only my original work towards the PhD except where indicated in the Preface.

2. Due acknowledgement has been made in the text to all other material used.

3. The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Yacov Salomon
Preface

The research presented here, unless stated otherwise, is original to the best of my knowledge. Furthermore, it has not been submitted for any other degree.

Chapter 1 gives a brief introduction to decision theory, information theory, expert elicitation and under-determined inverse problems. Relevant and related literature is briefly reviewed. This chapter serves as a motivating force for the rest of the thesis.

Chapter 2 summarises key mathematical techniques and concepts used throughout the thesis. Consequently, more comprehensive references are recommended herein.

Chapter 3 introduces the problem of unimodal density estimation, posed in the context of expert elicitation. The problem is then formulated, and solutions are given under relaxation conditions. The theoretical results are then applied in the context of expert elicitation. The work found here forms the majority of a paper by Y. Salomon and P. Taylor, titled “Using an information-theoretic approach to construct distributions for expert elicited data”, currently being reviewed for publication in the ANZ Journal of Statistics. This work also formed the base for a technical report and an Excel implementation, produced for the Australian Centre of Excellence for Risk Analysis, titled “An Excel implementation of the Minimum Cross-Entropy method for representing expert judgment”. Further collaboration work is currently being undertaken as a result of the report and implementation.

Chapter 4 extends results from Chapter 3 to solve the complete general problem of incorporating unimodality constraints in MCE formulation. This work forms the base of a paper to be submitted to IEEE Transactions on Information Theory as “An Information-theoretic approach to unimodal density estimation – optimality solutions”, by Y. Salomon.

Chapter 5 presents an approach to decisions under uncertainty using expert elicitation. The ideas are borrowed from economic theory and combined with results from previous chapters. The framework proposed is then applied in the context of environmental management decisions. This work is summarised in a paper by Y. Salomon, M. A. McCarthy, P. Taylor, and B. A. Wintle, titled “Incorporating uncertainty of management costs in sensitivity analysis of matrix population models, to appear in Conservation Biology”.

Chapter 6 summarises the results and techniques found throughout the thesis. Future research tasks and directions are given with motivating reasons.
# Contents

1 **Introduction** 1
   1.1 Decision theory, a brief history 1
   1.2 Expert elicitation 4
   1.3 Information theory 6
   1.4 Under-determined inverse problems 7
   1.5 Overview 8
   1.6 Summary of Contributions 9

2 **Mathematical Methods** 11
   2.1 Nonlinear Optimisation 11
      2.1.1 Necessary and sufficient optimality conditions 11
      2.1.2 Duality 14
   2.2 Logarithmic measures of information 16
      2.2.1 Surprise, uncertainty and entropy 16
      2.2.2 Cross entropy 18
      2.2.3 The reference measure 19
      2.2.4 Minimum cross entropy 20

3 **Unimodal density estimation – The relaxation solution** 23
3.1 Background and motivation ........................................... 23
3.2 An Information Theoretic Approach ................................. 25
3.3 A unimodality Constraint in the MCE Formulation ................. 27
   3.3.1 Feasibility ...................................................... 27
   3.3.2 Optimality ...................................................... 31
3.4 Relaxation Approach ................................................... 31
3.5 An application ......................................................... 33
   3.5.1 Case 1: .......................................................... 34
   3.5.2 Case 2: .......................................................... 36
   3.5.3 Case 3: .......................................................... 36
   3.5.4 Resolving discontinuities in the optimal density ............... 37
3.6 Chapter summary ....................................................... 40

4 Unimodal density estimation – General solution ....................... 43
   4.1 Background and Motivation ......................................... 43
   4.2 General optimal solutions .......................................... 44
   4.3 An application ...................................................... 51
   4.4 A note on numerical procedures .................................... 51
   4.5 Chapter summary .................................................... 54

5 Decision making under conditions of uncertainty ...................... 57
   5.1 Background and motivation ......................................... 58
   5.2 Safety first portfolio formulation ................................... 60
      5.2.1 Normally distributed uncertainties ............................. 62
      5.2.2 Uniformly distributed uncertainties ............................. 68
   5.3 Safety first portfolio for perturbation analysis ..................... 77
      5.3.1 koala population on Snake Island .............................. 78
5.3.2 Olive ridley sea turtle population ........................................ 80
5.4 Results .................................................................................. 85
5.5 Testing robustness to the linearity assumption in the portfolio model . . 89
5.6 The general elicited distributions case ..................................... 90
5.7 Chapter summary ................................................................. 95

6 Conclusion .............................................................................. 97
   6.1 Thesis summary ................................................................. 98
   6.2 Further work and future directions .................................... 99

A Appendix .................................................................................. 109
   A.1 Derivation of marginal cost expressions ............................. 109
List of Figures

3.1 MCE densities for uniform reference distribution and 100% confidence.  35
3.2 MCE densities for normal reference distribution and 80% confidence.  37
3.3 MCE densities for beta reference distribution and 80% confidence.  38
3.4 MCE densities for normal reference distribution and the smoothing constraint 39
3.5 MCE densities for beta reference distribution and the smoothing constraint 40
4.1 MCE densities using the relaxation approach ................................. 52
4.2 MCE densities using the general solution ...................................... 53
5.1 The efficiency frontier ................................................................. 81
5.2 Koala case study with normally distributed uncertainties ............... 82
5.3 Koala case study with uniformly distributed uncertainties .............. 83
5.4 Sea turtle case study with normally distributed uncertainties ........... 85
5.5 Sea turtle case study with uniformly distributed uncertainties ......... 86
5.6 Robustness to linearity assumptions ............................................ 91
5.7 MCE elicited densities for the koala case study ............................. 93
5.8 Koala case study with elicited MCE uncertainties .......................... 94
Acknowledgements

An immense amount of gratitude is owed to my supervisors Professor Peter Taylor, A/Professor Michael McCarthy and Dr. Brendan Wintle, for sharing their knowledge and experience when guiding me through my PhD journey. I am particularly in debt to Professor Peter Taylor for patiently supporting me in my attempts to master the art of becoming an applied mathematician.

Financial support was supplied by the Australian Postgraduate Award. A special acknowledgment goes to both the Applied Environmental Decision Analysis research hub, funded through the CERF program, and the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems, for offering further, and much appreciated, financial assistance. I would also like to acknowledge the support offered by the administrative staff at the University of Melbourne.

A very warm acknowledgement goes to my friends and colleagues in the Department of Mathematics and Statistics, and the AEDA and ACERA labs. You have made the overwhelming challenges of undertaking a PhD project less daunting, and I have learned and grown a lot thanks to you.

Finally, I have drawn a great deal of strength and inspiration during my studies from my family and friends both in Israel and Australia. I thank you all for the unconditional support and love you have always offered me.
Chapter 1

Introduction

1.1 Decision theory, a brief history

Decision theory is an interdisciplinary field, with important contributions both from, and in, a plethora of disciplines including mathematics, statistics, economics, psychology, and philosophy. It incorporates both descriptive and prescriptive treatments of the problem of making decisions. Prescriptive, or normative, decision theory explores how decisions should be made. Under the prescriptive heading, decision theory studies the axioms, principles and techniques for making ‘rational’ decisions. The goal of prescriptive decision theory is to produce methodology and tools to support the decision maker.

On the other hand, descriptive decision theory studies how decisions are actually made, and explores possible underlying mechanisms. The two approaches are closely linked, and many descriptive studies use results from prescriptive decision theory as null hypotheses. An example is the seminal work of prospect theory [40] in the field of behavioural economics.

Prescriptive decision theory studies decision making under three general classes of conditions, namely certainty, risk and uncertainty. In decision under certainty, each potential decision leads to only one possible outcome. This class of problems has been extensively studied under the disciplines of operational research and dynamic programming [7].

The classes of decision under risk and under uncertainty make up the core of decision theory. In decision under risk the outcome of each potential decision is described by a random entity following a known probability law. In contrast, in decisions under uncertainty the probability law associated with each random outcome is unknown.

The first rigorous treatment of decision under risk dates back to the 18th century.
In his influential paper, published in 1738, Bernoulli [6] introduced the use of expected utility functions for solving decision under risk type problems. Bernoulli’s approach, known today in its revised form as expected utility, claims that a rational decision maker acting under conditions of risk should maximise the expected value, not necessarily of the profit itself, but of some function of it, the utility function. Equivalently, one can choose to minimise the expected loss, where the loss function is given by the negative of a corresponding utility function.

The exact shape of the utility function cannot often be determined from first principles, but it is generally accepted that it is concave, and reflects the decision maker’s risk preference. Indeed, the lack of a rigorous method for deriving the utility function has been a key criticism of the expected utility approach [23]. Despite such criticisms, expected utility has enjoyed a surge in interest during the 20th century largely due to the works by von Neumann and Wald.

In 1947, John von Neumann and Oskar Morgenstern published an important result that served as a further justification for Bernoulli’s expected utility approach. The Von Neumann-Morgenston utility theorem states that a unique utility function exists (up to a linear transformation) if, and only if, the decision maker follows four basic axioms [60]. The axioms are defined in terms of lotteries (or utilities). They impose four conditions on the set of all lotteries, namely completeness, transitivity, continuity and independence [60].

Completeness assumes that the decision maker has well defined preferences. Given any two lotteries, A and B, either A is preferred, B is preferred, or the decision maker has no preference. Transitivity assumes consistency across any three lotteries. If A is preferred over B, and B is preferred over C, than A must be preferred over C. Continuity assumes that if A is preferred over B, and B is preferred over C, then there exists a probability \( p \in [0, 1] \) such that \( pA + (1 - p)C = B \). In other words, there exists a threshold between preferring and not preferring any intermediate lottery. Finally, independence assumes that a preference between any two lotteries holds, irrespective of another outcome. If A is preferred over B, than for any \( p \in (0, 1] \) and lottery C, \( pA + (1 - p)C \) is preferred over \( pB + (1 - p)C \).

A VNM-rational decision maker is therefore defined as any decision maker that is consistent with these four axioms. More recent work (e.g. [40]) has demonstrated that, in fact, humans are not VNM-rational decision makers, and suggested instead a range of generalised expected utility theorems using various other relaxed sets of axioms.

Building on Bernoulli’s ideas, Wald’s decision theory [80, 81] has synthesised and extended many concepts in statistical analysis and decision theory. Wald viewed both hypothesis testing and parameter estimation as special cases of decision making, and therefore considered decision theory more fundamental than inferential statistics. In its
most basic formulation Wald’s approach includes three components: the set of all possible ‘states of nature’ (more precisely, our knowledge of them); the set of all potential decisions; and a loss function, mapping each combination of a state and a decision to a number representing the subjective loss incurred when making that particular decision and the true state of nature turns out to be that particular state.

In dealing with the problem of decisions under uncertainty, Wald suggested the minimax criterion: for each decision, find the maximum possible loss over all states of nature, then choose the decision that minimises this maximum loss. When dealing with utilities instead of losses one would use the corresponding maximin criterion.

Minimax strategies are fundamental in game theory \cite{60,4} when games are played against intelligent adversaries who can foresee their opponent’s decisions and act accordingly. Criticism of the minimax criterion in the context of decisions under uncertainty argues that such an approach is overly pessimistic for most real world problems. In the decision problems faced by the scientist, engineer or economist, the so called game played is against nature, as a oppose to an intelligent adversary and, by using the minimax approach, one over- emphasises the worst case scenario \cite{35}. Various other minimax type criteria have also been suggested \cite{80,67}, and in more recent times Wald’s minimax has been extended to more general settings in info-gap theory \cite{2,73}.

An alternative to the minimax type approach was offered by the likes of Savage, De Ferrini, Berger and Jaynes, advocates of the Bayesian approach and subjective probabilities. Along with states of nature, decisions, and a loss function, most decision problems also include some prior knowledge, and perhaps also experimental evidence. The question of how to incorporate such information has been the subject of a long standing scientific dispute, dividing the statistical community into two opposing camps, namely, sampling or frequentist statistics, and Bayesian statistics. The Bayesian school advocates the use of posterior probabilities to minimise expected loss. The posterior probabilities are obtained using Bayes law combining a prior probability over the states of nature, and the conditional probability for the evidence given a particular state.

Advocates of the frequentist school strongly oppose such an approach, arguing that the state of nature is not a random entity and therefore cannot have a probability measure, the prior probability, associated with it. Instead, their formulation uses such terms as admissible decision rules and risk functions \cite{4}. We will not dwell on the details of this dispute, but instead note that while philosophically such differences may still need to be resolved, mathematically, both approaches lead to identical conclusions. This last assertion is known as Wald’s complete class theorem \cite{4,35}.

The approach we take in our work is that of the Bayesian school. In particular, we initially follow the formulation of the decision problem as presented in Berger \cite{4} and Jaynes \cite{35}. We then advocate the use of expert elicitation for obtaining subjective prior
probabilities. We pay particular attention to the problem of density estimation in the context of expert elicitation, when it includes eliciting information regarding a unique mode. The process of expert elicitation will be further introduced in section 1.4.

The objective function we use in our formulation of the problem of decisions under uncertainty is a modification of the one suggested by Roy [65] as part of his proposed safety-first principle. The safety-first principle argues that the real objective of a decision maker is to maximise the probability of avoiding an adverse event. This is equivalent to maximising an expected utility where the utility function is given by the indicator function of the complement of the adverse event. A more formal exposition of our approach to decisions under uncertainty will be developed in Chapter 5.

1.2 Expert elicitation

In the context of statistical analysis, elicitation is the process of representing an individual’s opinion regarding some uncertain quantity in a statistically meaningful way, most commonly a probability density [28]. We will refer to such an individual as the expert, and the person conducting the elicitation as the analyst. The probability density elicited is understood in a Bayesian subjective sense to represent the expert’s uncertainty regarding the quantity’s true value.

Elicitation is a common practice in a range of applications where data regarding a quantity of interest is sparse or indeed unavailable. Examples include: Bayesian statistics, where an analyst may use elicitation to construct prior distributions for the model’s unknown parameters to combine later with the likelihood function; decision making under uncertainty, where elicitation may be the only way to obtain probability measures over the state space, used in the derivation of expected utilities; and risk analysis, where elicitation is often used both when data is scarce, as in the case when dealing with rare events, or when it is unavailable, as in the case of future unprecedented events. For some particular examples of applications of elicitation see [29], [61] and references therein.

The elicitation process consists of three components: identifying the expert, obtaining their opinion, and translating that information into a probability distribution. The first two components have been chiefly studied by psychologists. Some key references on the relevant psychological literature are [39], [31], and [75], also see [28] and [44] for a review from a statistical viewpoint. Although the work reported in this thesis is concerned with the translation of elicited information into a probability distribution, it is important to note two key findings emerging from the psychological literature.

First, in any given elicitation process an expert is only expected to make a finite, and usually relatively small, number of statements regarding the uncertain quantity [5]. It is
therefore important to determine the summary statistics that an expert is able to estimate most competently. From extensive research conducted over several decades it turns out that individuals are better in estimating the mode and percentiles of a probability distribution, than estimating any of its moments, or indeed drawing its probability density function [28].

The second key finding is that experts tend to be overconfident in their estimates [74]. Overconfidence is defined as a function of the experts’ confidence in their estimate, how sure they are, and their hit rate, that is how often they are actually correct. Research devoted to identifying the moderators of overconfidence has found some elicitation protocols significantly reduce experts’ overconfidence [75]. In particular, Speirs-Bridge et al. [75] have reported promising results using their own 4-step elicitation protocol.

In the 4-step procedure the expert is first asked for their lowest and highest estimates of the uncertain quantity. Then they are asked for their best guess of its true value. Finally they are asked for their confidence that the true value lies within the interval they estimated. These four pieces of information correspond to percentile and mode summary statistics. The 4-step procedure will be used as a case study throughout this thesis.

We shall refer to the set of distributions satisfying the elicited summary statistics as the feasible set, and the set of summary statistics elicited as the constraint set. Apart from degenerate cases, summary statistics, elicited by any protocol, are insufficient to determine a unique distribution. Herein lies the main class of problems we address in Chapters 3 and 4; how do we single out a probability distribution from the feasible set, to best represent the expert’s opinion, given various classes of elicited data. This class of problems is a subset of a more general class of problems termed under-determined inverse problems [9], which will be further discussed in section 1.4.

Most applications of elicitation to date have used parametric techniques to address this problem [29]. In such cases the analyst assumes the expert’s opinion can be well represented by a member of a family of parametric distributions. The distribution best representing the expert’s opinion is then the member of the family that is the closest, with respect to some distance measure, to the feasible set. Note that the element, minimising such distance, is not necessarily a member of the feasible set. Given the parametric nature of the distribution family, the problem is reduced to identifying the best fit parameters. It is left to be decided, however, whether imposing a particular distributional form on the expert’s opinion is justifiable. For a detailed review of current parametric and non-parametric technics used for elicitation see [28].

In contrast to previous approaches, we argue for the use of an information-theoretic non-parametric approach. We will introduce our approach and discuss its merits in Chapter 3. Next, we include a brief overview of information theory.
1.3 Information theory

Information theory is a branch of the mathematical theory of probability and statistics [43]. As such, it has been applied to a wide variety of fields. Some even consider the notion of information to be more fundamental than that of probability [42]. It was first formally introduced in the context of statistics by Fisher [25] in 1925. Shannon [70] and [70], independently, published works describing logarithmic measures of information for use in communication theory. Information theory has strong ties with the mathematical theory of statistical mechanics through the concepts of disorder and uncertainty, formally quantified by entropy.

Fisher [25] introduced a measure of the amount of information supplied by data about an unknown parameter, as part of his work in the theory of statistical estimation. Wald [80, 81], Savage [67] and others have extended the application of information theory to statistical inference and decision theory. Kullback [43] published a comprehensive book studying the application of logarithmic measure of information to statistical hypothesis testing.

One of the most influential papers in the field of information theory was published by Claude E. Shannon [70] in 1948. In this groundbreaking paper, Shannon posed the problem of communication as a statistical (stochastic) process, developing the qualitative and quantitative model of communication underlying information theory. This work has inspired the application of information theory by communication engineers, psychologists, biologists, physicists, philosophers and others.

One of the fundamental concepts of information theory is that of entropy. There are close parallels between the thermodynamic entropy defined in statistical mechanics by Boltzmann and Gibbs in the 1870s, and that of information entropy. An extensive literature exists devoted to studies of the relation between the two ([43, 32] and references therein). Jaynes [32] integrated the definitions of entropy from the two separate fields under a unified statistical inference framework through the principle of maximum entropy. The principle of maximum entropy states that maximisation of Shannon entropy with respect to a probability distribution, and subject to known data constraints, constructs the most conservative, non-committal distribution consistent with the data. Results by Topsøe [77, 78], drawing parallels between the maximum entropy principle and zero-sum games between a decision maker and nature, act as further support for its generality in the context of decision theory.

In section 2.2 we will formally introduce logarithmic measures of information, and related results relevant for the work presented in this thesis.
1.4 Under-determined inverse problems

Many applied problems in a wide range of applications can often be reduced to finding the "best" $f$ solving the system $A(f) = b$, where $b \in \mathbb{R}^n$, $A$ is either a linear or a non-linear operator, and $f$ lies in an appropriate functional space. This general class of problems is termed under-determined inverse problems [17, 37, 9, 10]. The term inverse stems from the fact that we choose $f$ based on a finite set of summary information loosely termed its moments. Under-determined relates to the fact that the finite set of moments is almost always insufficient to describe a unique $f$. To determine the "best" $f$ we therefore require an additional criterion by which to order the set of feasible solutions. The criterion chosen varies depending on the application. The list of criteria used includes information theoretic measures such as entropy, cross-entropy, Fisher’s information and Burg entropy, as well as norm related measures when available (that is, when $f$ an element of a normed space).

Examples of under-determined inverse problems exist in many applications. Jonathan Borwein, in a recent talk, presented the following partial list of fields where such problems find applications: acoustics, actuarial science, astronomy, biochemistry, constrained spline fitting, engineering, finance, image processing, inverse scattering, optics, option pricing, and statistical moment fitting.

The problem of determining the probability density best representing the information elicited from an expert, introduced in section 1.2, is yet another example of an under-determined inverse problem. In this case, $f$ is a probability density lying in an appropriate functional space, $b$ is the vector of elicited information, and $A$ is a set of linear operators. An important problem arises when information elicited cannot be expressed using a linear operator acting on $f$. Such is the case when eliciting information regarding the most likely value of the uncertain quantity in question, when it corresponds to the unique mode of the probability density representing the expert’s uncertainty. Solving this class of problems is the key mathematical contribution of this thesis.

A common approach to solving for the optimal member of the feasible set is to discretize the problem, reducing it to a finite dimensional (can be very large) setting where $f \in \mathbb{R}^m$ and $A$ an $m \times n$ matrix. Other authors [24, 63, 17, 9] choose to address the problem in its original functional space, and only discretize as necessary for numerical computations.

Solving under-determined inverse problems in the functional space setting for the case when the ordering criterion is a convex functional has attracted reasonable amount of attention since the mid 20th century [24, 63]. Most of the work concentrated on the existence of duality results, necessary and sufficient conditions for optimality, and conditions under which there exist no duality gap [17]. In more recent times these results have been extended to more general settings including relaxing linearity and convexity assumptions.
to some degree [9]. Our work follows the functional settings approach and applies previous duality results to the particular settings of our problem.

1.5 Overview

The work presented in this thesis attempts to answer three key questions. The first question is, how to pick a probability density best representing information elicited from an expert. The second question is, how to incorporate unimodality in density estimation using the minimum cross entropy (MCE) method. Finally, the more general third question we address is, how to make decisions when faced with conditions of uncertainty.

The first two questions are related, the first question being an application of solutions for the second question, in the case when eliciting information regarding a unique mode. As such it offers further support for the use of information theoretic methods in the context of expert elicitation.

It is important to note, however, that these questions also have much broader applications in their own right. The problem of unimodal density estimation arises in various applications unrelated to expert elicitation, including in the fields of engineering and signal processing. Similarly, the use of an information theoretic approach for estimating distributions from expert elicited data is indeed valid under a broad range of classes of elicited data, and not only for the case of elicitation of the unique mode.

The work behind this thesis has been an evolution of ideas and methods originating with the problem of decisions under uncertainty. The order in which we present our results in the thesis does not follow necessarily chronological order. Instead we chose to begin with the problem of unimodal density estimation, using the application to expert elicitation for motivation and illustration. We then combine these results with our proposed framework for decisions under uncertainty, and apply it to two population management case studies. This was done in an attempt to make the exposition as coherent as possible.

The thesis is structured as follows:

- Chapter 2 summarises key mathematical techniques and concepts used throughout the thesis. In particular, we briefly introduce the field of non-linear optimisation concentrating on the KKT conditions and Duality. We then discuss relevant ideas from information theory. We provide an intuitive derivation of the concept of entropy in the context of decision theory. We then discuss various convexity results needed for establishing proofs in the following Chapters.

- In Chapter 3 we introduce the problem of unimodal density estimation, posed in
the context of expert elicitation. We then discuss previous approaches to solving the problem and demonstrate that such efforts lead only to feasible, rather than optimal solutions. Hence, we provide a reformulation of the problem that would lead to optimality, and solve the relaxed problem. We conclude with application of these results to the 4-step elicitation protocol.

- Chapter 4 extends results from Chapter 3. In particular, we solve the complete problem to obtain both the relaxation solution presented in the previous Chapter, as well as results for when the relaxation approach does not hold. We then apply the general solution to a range of elicitation scenarios.

- In Chapter 5 we advocate the use of expert elicitation in the context of decisions under uncertainty. We apply our proposed approach in the context of environmental management. Specifically, we combine our results from Chapters 3 and 4 to estimate probability densities associated with expert elicited information regarding uncertain conservation costs. We then incorporate these densities in a decisions under uncertainty context, and introduce the safety first portfolio decision making formulation.

- We finally conclude with brief summary of our main contributions and a short discussion on directions for future research.

## 1.6 Summary of Contributions

Contributions in Chapter 3 include:

- A formulation of the problem of unimodal density estimation using the MCE method in the context of expert elicitation. This is achieved using an important result due to Khinchin and Shepp that offers a useful representation of unimodal densities.

- A demonstration that previous work results in feasible, but not optimal solutions.

- A derivation of the form of the problem that, if solved, would lead to optimal solutions.

- A derivation of the optimal solution of the relaxation problem, as well as the necessary and sufficient conditions for the relaxation approach to hold.

- An application of the relaxation solution to the 4-step elicitation protocol.

Contributions in Chapter 4 include:
• A proof establishing the equivalence of original and transformed problems.
• A proof of convexity of the transformed objective using Jensen’s inequality.
• A derivation of the expression for the density that minimises the primal problem using the convexity and the Gâteaux variation of the Lagrangian at $f_y$ in the direction $\delta_x$.
• A method for decoupling primal and dual variables.
• A method that uses complimentary slackness to reduce the dual to an $(n + 1)$-dimensional convex problem.

Finally, Chapter 5 includes the following contributions:

• A formulation of the safety first portfolio approach to decision under risk.
• An extension to the safety first portfolio approach for the case of decisions under uncertainty using distributions obtained from expert elicitation.
• A formulation combining the safety first portfolio approach with matrix perturbation analysis to address environmental decision problems in population management.
• An application of our results to two case studies in population management: management of a koala population on Snake Island, Australia; and conservation of a population of Olive-ridely marine turtles.
Chapter 2

Mathematical Methods

2.1 Nonlinear Optimisation

Throughout the thesis we make use of results from the field of nonlinear optimisation. In particular, we will use extensively the Lagrange multiplier theory and the Karush-Kuhn-Tucker (KKT) optimality conditions. We will also make use of results from duality theory and convex programming. In this section we present a brief introduction to nonlinear optimisation, and highlight the main results required for the derivations in the following three chapters. Detailed proofs are often omitted in favour of an intuitive approach. For details we refer the reader to the excellent text by Bertsekas [7].

2.1.1 Necessary and sufficient optimality conditions

Let $f$ and $h_i$ be continuous differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. We consider the following general problem with equality constraints:

$$\begin{align*}
\min & \quad f(x) \\
\text{subject to} & \quad h(x) = 0,
\end{align*}$$

(ECP)

where $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the function

$$h = (h_1, \ldots, h_m).$$

Our aim is to describe necessary and sufficient optimality conditions for (ECP). We begin by defining the Lagrangian function, $L : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, given by

$$L(x, \lambda) = f(x) + \lambda^T h(x),$$
where the entries of the vector $\lambda = (\lambda_1, \ldots, \lambda_m)^T$ are termed Lagrange multipliers. The Lagrange multipliers play a key role in the theory of constrained optimisation, and have different interpretations depending on the applications. From a mathematical viewpoint they provide sensitivity information, quantifying, up to first order, the variation of the optimal solution caused by variation in the associated constraint [7].

**Definition 2.1.** A feasible vector $x$ is called regular if the constraint gradients,

$$\nabla h_1(x), \ldots, \nabla h_m(x)$$

are linearly independent.

**Theorem 2.1** (Lagrange multiplier theorem – necessary conditions). Let $x^*$ be a regular point and a local minimum of (ECP). Then there exist unique Lagrange multipliers $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)^T$, such that

$$\nabla f(x^*) + (\lambda^*)^T \nabla h(x^*) = 0.$$  \hspace{1cm} (2.1)

Equation (2.1) can be interpreted as stating that the objective’s gradient $\nabla f(x^*)$ belongs to the space spanned by the constraints’ gradients $\nabla h(x^*)$ at the optimal point $x^*$. Alternatively it can be viewed as stating that the objective gradient is orthogonal to the space of first order feasible variation

$$V(x^*) = \{ y | \nabla h_i(x^*)^T y = 0, \ i = 1, \ldots, m \}.$$  

$V(x)$ is interpreted as the space of variations $\Delta x$ about a feasible $x$ that are themselves feasible, that is they satisfy $h(x + \Delta x) = 0$. This alternative interpretation is a generalisation of the zero gradient condition $\nabla f(x^*) = 0$ of unconstrained optimisation, to constrained optimisation settings.

The proof of Proposition 2.1 follows one of several approaches, and will not be included. Instead we mention briefly the idea behind one of these approaches, namely the penalty approach. Here, a high penalty for violating the constraints is added to the objective, and the constrains themselves are omitted. This reduces the constrained optimisation problem to an unconstrained one.

The first order necessary condition (2.1) may be satisfied by both local minima and local maxima. The following sufficient conditions guarantee that a given vector is a local minimum.

**Theorem 2.2** (Second order sufficient conditions). Assume $f$ and $h$ are twice continuously differentiable, and let $x^*$ and $\lambda^*$ satisfy

$$\nabla x L(x^*, \lambda^*) = 0, \quad \nabla \lambda L(x^*, \lambda^*) = 0,$$

$$y^T \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \quad \forall \ y \in V(x^*).$$  \hspace{1cm} (2.3)

Then $x^*$ is a strict minimum of problem (ECP).
The second order condition (2.3) requires that the Lagrangian function at $\lambda^*$ is convex on $V(x^*)$, with respect to $x$.

We now extend the results above for equality constraints to the case of inequality constraints. Let $f, h_i, g_j$ be continuous differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. We consider the following general constrained optimisation problem:

$$\min f(x)$$
subject to $h(x) = 0, \ g(x) \leq 0,$

where $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^r$ are the functions

$$h = (h_1, \ldots, h_m), \quad g = (g_1, \ldots, g_r).$$

We first define for any feasible point $x$ of (ICP), the set of active inequality constraints

$$A(x) = \{j | g_j(x) = 0\}.$$

Then $A^c(x) = \{j | g_j(x) < 0\}$ is the set of all inequality constraints that are said to be inactive at $x$. Inactive constraints for a feasible point, in effect, do not matter. Thus, if $x^*$ is a local minimum of problem (ICP), then it is also a local minimum of a problem identical to ICP where the inactive constraints at $x^*$ have been discarded.

**Theorem 2.3** (Karush-Kuhn-Tucker necessary conditions). Let $x^*$ be a local minimum of the problem ICP. Then, there exist unique Lagrange multipliers vectors $\lambda^*$ and $\mu^*$, such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

$$\mu_j^* \geq 0, \quad j = 1, \ldots, r,$$

$$\mu_j^* = 0 \quad \forall j \notin A(x^*).$$

**Theorem 2.4** (Second order sufficient conditions). Assume $f, h$ and $g$ are twice continuously differentiable, and let $x^*$, $\lambda^*$, and $\mu^*$ satisfy,

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \quad h(x^*) = 0 \quad g(x^*) \leq 0,$$

$$\mu_j^* \geq 0, \quad j = 1, \ldots, r,$$

$$\mu_j^* = 0 \quad \forall j \notin A(x^*),$$

$$y^T \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \quad \forall y \in V(x^*),$$

where,

$$V(x^*) = \{y | \nabla h_i(x^*)^T y = 0, \ i = 1, \ldots, m; \ \nabla g_j(x^*)^T y = 0, \ \forall j \in A(x^*)\}.$$

Then $x^*$ is a local minimum of $f$ subject to $h(x) = 0$ and $g(x) \leq 0$. 
The proofs for the inequality constraint optimality conditions above follow almost immediately from the ones for the equality constraint conditions.

The second order sufficient conditions, Theorems 2.3 and 2.4, involve a second derivative and Hessian positive definitiveness assumptions. For the work in this thesis, however, there are other, more useful, assumptions, that also result in sufficient conditions for optimality.

In particular, when both the objective function and constraints are convex, the first order necessary condition (2.4), is also sufficient, and $x^*$ is the global minimum. To see this, consider the problem ICP subject only to inequality constraints, and $x \in X$. Let $x^*$ be a point satisfying Theorem 2.3 together with a vector $\mu^*$. Further assume that

$$x^* = \arg \min_{x \in X} L(x, \mu^*), \quad (2.5)$$

then $x^*$ is the global minimum since,

$$f(x^*) = f(x^*) + \mu^T g(x^*)$$
$$= \min\{f(x) + \mu^T g(x)\}$$
$$\leq \min_{x \in X, g(x) \leq 0} \{f(x) + \mu^T g(x)\}$$
$$\leq \min_{x \in X, g(x) \leq 0} f(x).$$

The first equality holds since, from the assumptions, $\mu^T g(x^*) = 0$. This last assertion is also known as complimentary slackness condition. The last inequality follows from the non-negativity of $\mu^*$. Finally, when $f$ and $g$ are convex $L(x, \mu^*)$ is convex, and the first order necessary condition $\nabla_x L(x^*, \mu^*) = 0$ is exactly the minimisation condition (2.5).

Note that any equality constraint $h_i(x) = 0$ can be transformed into the two inequality constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, each with an associated multiplier. Then, the two multipliers can be combine into a single multiplier, now taking values in $\mathbb{R}$. As a result, for the remainder of the discussion on nonlinear optimisation, and without loss of generality, we study problems subject to inequality constraints only.

### 2.1.2 Duality

The next important concept we discuss is duality. It is a central concept in the field of nonlinear optimisation that leads to many deep insights. We will only touch briefly on a couple of results and key ideas important for the work in subsequent chapters.
We consider the following problem
\[
\min_{x \in X} f(x) \quad \text{subject to} \quad g(x) \leq 0,
\]
where \( f : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^n \to \mathbb{R}^r \), as above. We refer to problem (PP) as the primal problem.

Next we define the dual function
\[ q(\mu) = \inf_{x \in X} L(x, \mu), \quad \mu \in \mathbb{R}^r. \]

If we let \( S = \{(g(x), f(x)) | x \in X\} \), then the dual function may be interpreted as the highest point of intersection with \( \{(0, w) | w \in \mathbb{R}\} \) over all hyperplanes with normal \((\mu, 1)\) that contain the set \( S \) in their positive halfspace. The dual problem is then
\[
\max_{\mu \geq 0} q(\mu) \quad \text{subject to} \quad \mu \geq 0, \tag{DP}
\]
corresponding to finding the maximum point of intersection over all \( \mu \geq 0 \).

**Theorem 2.5** (Weak Duality Theorem). *Assuming the dual problem (DP) has a finite optimal value*
\[ q^* = \max_{\mu \geq 0} q(\mu), \]
*we have*
\[ q^* \leq f^*. \]

**Proof.** For all \( \mu \geq 0 \), and \( x \in X \) with \( g(x) \leq 0 \),
\[ q(\mu) = \inf_{y \in X} L(y, \mu) \leq f(x) + \mu^T g(x) \leq f(x) \]
and
\[ q^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in X} f(x) = f^*. \]

If \( q^* < f^* \) we say that there is a duality gap, conversely, if \( q^* = f^* \) then we say that there is no duality gap. It is most useful to have no duality gap. To guarantee no duality gap one must impose various convexity requirements. The scenario most useful in our case is the following.
Theorem 2.6 (Strong Duality Theorem). Assume the problem (PP) is feasible, and its optimal point $f(x^*)$ is finite. Further assume that $X$ is a convex subset, and $f, g_j$ are convex on $X$. Finally, assume that there exists a vector $y \in X$ such that
\[ g_j(y) < 0, \quad \forall j = 1, \ldots, r. \] (2.6)

Then there is no duality gap.

Condition (2.6) is called the Slater constraint qualification or the interior point condition.

2.2 Logarithmic measures of information

We begin this section by presenting an intuitive introduction to the concept of entropy using the notion of surprise. A more detailed exposition of the approach can be found in Ross [64, p.436-440]. We then formally define the cross entropy measure and derive several related properties and information inequalities relevant for the work presented in the following chapters.

A comment on notation, we use the term densities loosely throughout the thesis to mean the Radon-Nikodym derivatives with respect to a relevant dominant measure $\mu$. In most cases, however, the dominant measure would be the Labesque measure, with $X \subset \mathbb{R}$, and $\mathcal{F}$ the Boreal $\sigma$-algebra on $\mathbb{R}$, as per the most common interpretation of probability densities. We use the notation $\mathcal{F} \prec \mu$ to denote that $\mathcal{F}$ is absolutely continuous with respect to $\mu$.

2.2.1 Surprise, uncertainty and entropy

Consider an event $E$ that may occur when preforming a random experiment. How surprised would you be to discover that $E$, indeed, occurred? Intuitively, it seems reasonable to assume that, the surprise experienced is inversely proportional to the probability of $E$ occurring. For example, consider a game of rolling a pair of dice. Then, we would expect to be somewhat more surprised to learn that the sum of the dice is 12, than if we learned that the sum is even.

We wish to formalise the concept of surprise. Let $S : [0, 1] \to [0, \infty)$ denote the surprise function, $S(p)$ being the surprise evoked by learning of an event with probability $p$ occurring. In order to obtain a functional form for $S$, we state the following four reasonable axioms.
**Axiom 1** There is no surprise associated with an event that is certain to happen,

\[ S(1) = 0. \]

**Axiom 2** The more unlikely an event is, the higher the surprise evoked by learning it occurred. That is, \( S(p) \) is a strictly decreasing function of \( p \),

\[ S(p) > S(q), \quad \text{iff } p < q. \]

**Axiom 3** \( S(p) \) is a continuous function of \( p \).

**Axiom 4** Surprise is additive. Given two independent events \( E, F \) occurring with probabilities \( p, q \) respectively,

\[ S(pq) = S(p) + S(q). \]

The last condition simply states that the surprise evoked by learning of the occurrence of two independent events, is equivalent to learning that one has occurred followed by learning the second has occurred.

**Theorem 2.7.** If \( S(\cdot) \) satisfies Axioms 1–4, then

\[ S(p) = -C \log_2 p \]

where \( C \) is an arbitrary positive integer.

**Proof.** From Axiom 4, \( S(p^2) = S(p) + S(p) = 2S(p) \), and by induction

\[ S(p^m) = mS(p). \]

Also, for any integer \( n \), \( S(p) = nS(p^{1/n}) \), or

\[ S(p^{1/n}) = \frac{1}{n}S(p). \]

Combining we have,

\[ S(p^{m/n}) = \frac{m}{n}S(p). \]

Then, since the rational numbers are dense in the reals, and by Axiom 3,

\[ S(p^x) = xS(p), \quad x \geq 0. \]

Now, for any \( p \in (0,1] \), let \( x = -\log_k p \) \( k > 1 \), or equivalently \( p = k^{-x} \), and we have

\[ S(p) = S(k^{-x}) = xS(\frac{1}{k}) = -C \log_k p, \]

where \( C = S(\frac{1}{k}) > S(1) = 0 \) by Axioms 1 and 2. Common values for the base \( k \) are \( k = 2 \) and \( k = e \). \( \square \)
Consider next a discrete random variable $X$ taking values $x_1, \ldots, x_n$ with respective probabilities $p_1, \ldots, p_n$. Then the expected amount of surprise we shall experience upon learning the value of $X$ is given by,

$$H(X) = \sum_{i=1}^{n} p_i \log p_i. \quad (2.7)$$

The quantity $H(x)$ in (2.7) is termed the entropy of the random variable $X$. It is maximised when $X$ is a uniformly distributed random variable. As well as surprise, we can think of $H$ as representing the expected uncertainty regarding the actual value of $X$, or similarly, the expected information gained by observing the value of $X$.

If we consider next two random variables $X, Y$ taking values $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$, with joint probability $p(x_i, y_j)$, it is relatively easy to show that,

$$H(X, Y) = H(Y) + H_Y(X),$$

where $H_Y(X) = \sum_j H_{Y=y_j}(X) Pr\{Y = y_j\}$, and $H_{Y=y_j}(X) = -\sum_i p(x_i|y_j) \log p(x_i|y_j)$.

2.2.2 Cross entropy

Crossing from the discrete to the continuous case in (2.7) requires considerable more analysis. Indeed, given a continuous random variable $X$ with support $\mathcal{X}$, naively attempting to simply replace summation with integration in (2.7) we obtain the expression,

$$H_d(f) = -\int_{\mathcal{X}} f(x) \log f(x) \, dx, \quad (2.8)$$

termed differential entropy. While the analogy between (2.7) and (2.8) is suggestive, several properties true for the discrete entropy, do not hold for differential entropy. Most importantly (2.8) is not invariant under continuous co-ordinate transformations $x \rightarrow y(x)$, and it may even be negative. It was therefore suggested that differential entropy is not the correct information measure for a continuous distribution.

To obtain an information measure for a continuous distribution we must pass to the limit from the discrete case (2.7). The result is termed the cross entropy (CE) measure, also known as the Khinchin-Kullback-Libelier divergence, or relative entropy. A detailed derivation of the following definition can be found in [35].

**Definition 2.2 (Cross Entropy).** Given measures $F, Q$ and $\mu$ on $(\mathcal{X}, \mathcal{F})$ such that $F \prec \mu$ and $Q \prec \mu$, the Cross Entropy of $F$ with respect to the reference measure $Q$ is defined as

$$H(F\|Q) = \int_{\mathcal{X}} f(x) \log \frac{f(x)}{q(x)} \, \mu(dx), \quad (2.9)$$

where $f, q$ are the Radon-Nikodym derivatives (densities) of $F, Q$ with respect to $\mu$. 

Informally, CE is a measure of the information distance between the density \( f \) and the reference density \( q \). The CE measure is minimal when \( f = q \) almost surely, where it takes the value 0. This last assertion is known as Gibbs’ inequality, and it can be proved using the following result.

**Lemma 2.1** (Jensen’s inequality). Let \( X \) be a real valued \( \mu \)-integrable random variable in \((X,\mathcal{F},\mu)\). Then, for all \( \varphi \) convex on an interval containing the support of \( X \)

\[
\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]).
\]

The proof of Jensen’s inequality follows by induction on the convexity requirement. Using Jensen’s inequality, and the convexity of the function \((- \log)\), we have

\[
\int_X f(x) \log \frac{f(x)}{q(x)} \mu(dx) = - \int_X f(x) \log \frac{q(x)}{f(x)} \mu(dx) \\
\geq - \log \left( \int_X f(x) \frac{q(x)}{f(x)} \mu(dx) \right) \\
= 0,
\]

and the non-negativity of the CE measure is established.

The CE measure is not symmetric, \( H(F||Q) \neq H(Q||F) \), and does not satisfy the triangle inequality, and therefore it does not constitute a proper metric. However, it does induce a topology on the space of generalised probability distributions. Convergence in CE, as well as in other information measures, has proved valuable in establishing several classical results in probability theory, including the Central Limit Theorem [36].

### 2.2.3 The reference measure

For applications of CE to the problem of constructing prior distributions, there remains, however, the question of determining the proper form for the reference density. Intuitively, the reference measure is the distribution representing ‘complete ignorance’.

The notion of ‘complete ignorance’ has been explored in the past. Bayes suggested the use of the uniform distribution to represent complete ignorance. This is also the approach taken by Laplace, coining the term principle of indifference. The textbook examples of applications of the principle of indifference are experiments with discrete and finite sample space, such as flipping of a coin, or rolling of dice. The uniform distribution then reflects the assumption of fairness or symmetry, prior to the experiment.

However, application of the uniform distribution in some cases leads to nonsensical results. For an example see Jaynes [35] treatment of the famous Bertrand’s paradox.
There is also the problem of applying the uniform distribution in cases where the support is unbounded. To address these problems Jaynes introduced the principle of transformation groups. It extends the principle of indifference by stating that the reference distribution must reflect indifference between equivalent forms of the problem.

Jaynes arrived to this conclusion after noticing that similar problems have been solved in Riemannian geometry and General Relativity. There, arbitrary continuous coordinate transformations are performed, while some properties remain invariant. For example, in the theory of continuous groups, the Harr measure is invariant under group transformations. Seen from such perspective, the principle of indifference simply reflects indifference under permutations of a finite group. For example, when flipping a coin, the uniform distribution reflect indifference to permuting the sides of the coin, calling heads tail and tail heads. More elaborated examples can be found in [35], including a consistent argument that resolves Bertrand’s paradox.

### 2.2.4 Minimum cross entropy

The feasible density that minimises the nonlinear CE functional is the most conservative, non-committal density consistent with the information summarised in the constraints and the reference density. We now formally introduce the problem of choosing the density function \( f \) that minimises the CE measure, subject to moment constraints.

Let \( F, Q \) and \( \mu \) be measures on \( (X, \mathcal{F}) \) such that \( F \prec \mu \) and \( Q \prec \mu \). Denote by \( f \) and \( q \) the densities of \( F \) and \( Q \), with respect to \( \mu \). We consider the following problem

\[
\min_f \int_X f(x) \log \frac{f(x)}{q(x)} \mu(dx)
\]

subject to \( \theta_i = \int_X g_i(x)f(x) \mu(dx), \quad i = 0, \ldots, n, \)

\( g_0 = \theta_0 = 1, \quad f \geq 0, \tag{2.10} \)

where \( \theta_0, \ldots, \theta_n \) are given moment values corresponding to the real valued moment functions \( g_0, \ldots, g_n \). The condition \( g_0 = \theta_0 = 1 \) is equivalent to \( \int_X dF = 1 \), which, together with the nonnegativity constraint, ensures that \( f \) is a proper probability density.

We solve (2.10) using the standard method of Lagrange multipliers (see section 2.1.1). We introduce Lagrange multipliers, \( \lambda_0, \ldots, \lambda_n \), and construct the Lagrangian function

\[
L(f, \lambda) = \int_X f(x) \log \frac{f(x)}{q(x)} \mu(dx) + \sum_{i=0}^n \lambda_i \left( \int_X g_i(x)f(x) \mu(dx) - \theta_i \right) . \tag{2.11}
\]
The density $f^*$ that is the optimal solution of (2.10) has the well known form [34, 43]

$$f^*(x) = q(x) \exp\left(-\sum_{i=0}^{n} \lambda_i g_i(x) - 1\right). \quad (2.12)$$

Equation (2.12) can be derived in one of several ways. Here, we present a method based on the Gâteaux differential. We will use the Gâteaux differential again in Chapter 4 to derive key results. The definition below can be found, for example, in [79].

**Definition 2.3 (Gâteaux differential).** Let $Z$ and $W$ be two locally convex topological spaces, $U \subset Z$ open, and $F : Z \to W$. The Gâteaux differential $dF(u, \zeta)$ of $F$ at $u \in U$ in the direction $\zeta \in Z$ is given by

$$dF(u, \zeta) : = \lim_{\epsilon \to 0} \frac{F[u + \epsilon \zeta] - F[u]}{\epsilon} = \frac{\partial}{\partial \epsilon} F[u + \epsilon \zeta] \bigg|_{\epsilon = 0}. \quad (2.13)$$

The second equality requires that the partial derivative exists, and the functional $F$ is continuous about $\epsilon = 0$.

Next, we apply the first order necessary condition (2.4) to (2.11) using the Gâteaux differential. Let $\mathcal{L}^1$ denote the 1-norm functional space. Then note that, for fixed $\lambda$, $L_{\lambda} : \mathcal{L}^1 \to \mathbb{R}$, and both $\mathcal{L}^1$ and $\mathbb{R}$ are locally convex topological spaces. The Gâteaux differential is taken at $f$, in the direction of $\delta_x$, the Dirac delta function at $x$,

$$dL_{\lambda}(f, \delta_x) = \frac{\partial}{\partial \epsilon} L_{\lambda}(f + \epsilon \delta_x) \bigg|_{\epsilon = 0}$$

$$= \frac{\partial}{\partial \epsilon} \left( \int_{\mathcal{X}} (f(t) + \epsilon \delta_x(t)) \log \frac{f(t) + \epsilon \delta_x(t)}{q(x)} \, dt \bigg|_{\epsilon = 0} \right)$$

$$+ \frac{\partial}{\partial \epsilon} \left( \sum_{i=0}^{n} \lambda_i \left( \int_{\mathcal{X}} g_i(t) (f(t) + \epsilon \delta_x(t)) \, dt - \theta_i \right) \right) \bigg|_{\epsilon = 0}$$

$$= \log \frac{f(x)}{q(x)} + 1 + \sum_{i=0}^{n} \lambda_i g_i(x).$$

Equating to zero and rearranging we obtain (2.12).

To complete this approach it is left to show that the necessary condition is also sufficient. This follows since the CE functional is convex, by Jensen’s inequality, as well as the linearity of the constraints.
To solve for the Lagrange multipliers we rewrite (2.12) using the normalising function

\[ Z(\lambda_1, \ldots, \lambda_n) = e^{\lambda_0} = \int_X q(x) \exp \left( - \sum_{i=1}^{n} \lambda_i g_i(x) \right) \mu(dx), \]

defined using the condition \( \theta_0 = g_0 = 1 \), so that

\[ f^*(x) = \frac{q(x)}{Z^*} \exp \left( - \sum_{i=1}^{n} \lambda_i^* g_i(x) \right). \]  

(2.14)

Clearly, we require \( Z(\lambda_1, \ldots, \lambda_n) < \infty \) and the Lagrange multipliers are given by the solutions to

\[ \theta_i = - \frac{\partial \log Z(\lambda_1, \ldots, \lambda_n)}{\partial \lambda_i}, \quad i = 1, \ldots, n. \]

(2.15)

Obtaining explicit expressions for the density (2.14) requires solving for the Lagrange multipliers \( \{\lambda_i\} \). This is a highly non-linear problem as is seen by examining (2.15). Alternatively, [17] demonstrated that these Lagrange multipliers are dual variables in an unconstrained convex problem (see also [11] and [12]). We will introduce the duality approach to solving for the multipliers in Chapter 3.
Chapter 3

Unimodal density estimation –
The relaxation solution

In this chapter we formally introduce the problem of unimodal density estimation using the MCE method. The problem is posed in the context of expert elicitation. In particular, we demonstrate how a previous approach to solving this exact problem leads to feasible, but not optimal solutions. We then formulate the nonlinear problem that would lead to optimality results, and offer relaxation solutions. The relaxation solutions are illustrated using the 4-step elicitation protocol example.

3.1 Background and motivation

The process of expert elicitation was introduced in Section 1.2. We recall that the person from whom the information is elicited is termed the expert, and the person conducting the elicitation is the analyst. We wish to derive a method for representing data elicited from the expert as a probability distribution. Such a method would be applied by the analyst conducting the elicitation. The distribution may then be used for comparing, or combining, different experts’ opinions, or as a prior in a decision process.

We further recall that the set of distributions satisfying the elicited distributional characteristics is termed the feasible set, and the set of distributional characteristics elicited is termed the constraint set. The under-determined inverse problem that we investigate is how to pick the element of the feasible set “best” representing the information summarised in the constraints.

Our approach uses the CE measure [43], defined over the feasible set. In particular,
the distribution chosen to best represent the expert’s opinion is the element of the feasible set that minimises the information distance to some reference distribution set by the analyst. The reference distribution can be thought of as representing the analyst’s ‘complete ignorance’ regarding the shape of the distribution representing the expert’s opinion. Note that, in contrast to a parametric approach [29], the selected distribution is always a member of the feasible set, even when the reference distribution is not.

Shore and Johnson [71] have shown that, given some basic postulates, minimising cross entropy is the only method for choosing a distribution that is self consistent. The resulting distribution is the one that fully utilises the information provided by the set of constraints, while remaining maximally noncommittal with respect to unavailable information [32].

Apart from its strong philosophical and mathematical grounding, the MCE method is also a relatively general approach. It can utilise a variety of distributional characteristics, including moments, percentiles and conditional probabilities, and can be applied to elicitation of both discrete and continuous distributions. The MCE method is a commonly used inference method for constructing prior distributions [33]. Many known univariate [46] and multivariate [21] distributions have an equivalent derivation using MCE. Such derivations use the MCE with respect to a constant density function which is equivalent, up to a constant, to choosing the maximum entropy distribution [35]. MCE has also been used in the context of elicitation for the aggregation of multiple experts’ judgments [58].

Despite its numerous advantages, there remains some difficulty in applying the MCE method in the context of expert elicitation. Distributional characteristics that are not expressible as an expectation over some real valued function cannot be readily used. This poses a particular problem since the mode estimate is an example of such a distributional characteristic, while being a popular choice for elicitation protocols. Brockett et al. [13] used moment transformation techniques, originally developed by [41], in an attempt to incorporate a unimodality constraint in MCE formulation. However, as will be demonstrated below, their work produced only feasible, rather then optimal, distributions, with respect to the CE measure.

The novelty in the work we present here is threefold. First, as far as the authors are aware, this is the first time the MCE approach has been applied in the context of a single expert elicitation procedure. Second, we extend the previous work by Brockett et al. [13] and formulate the general problem of incorporating a unimodality constraint in the MCE formulation that leads to optimal solutions. Finally we demonstrate how to use the MCE approach with a unimodality constraint under a special class of elicited information. This special information class includes the distributional characteristics elicited using the 4-step procedure protocol.
3.2 An Information Theoretic Approach

We begin by formulating the standard MCE approach, posed in the context of expert elicitation. Let $\mathcal{X}$ denote the set of all potential values of a particular uncertain quantity of interest. The analyst’s goal is to determine the density $f$ best representing, in a Bayesian sense, the expert’s opinion regarding the true value of such a quantity.

To construct $f$, the analyst elicits the value of $n$ distributional characteristics, $\theta_i$, $i = 1, \ldots, n$, from the expert. In the case of the standard MCE approach, introduced in Section 2.2.4, each such distributional characteristic is assumed to represent the value of an expectation over some real valued function $g_i$, termed a moment function, defined on $\mathcal{X}$. Such expectations are taken with respect to the yet to be determined density $f$. That is,

$$\theta_i = \mathbb{E}_f[g_i] := \int_{\mathcal{X}} g_i(x)f(x)\mu(dx), \quad i = 0, \ldots, n,$$

(3.1)

where $\mu$ is some dominating measure on $(\mathcal{X}, \mathcal{F})$ (typically the Lebesgue measure), and $g_0 = \theta_0 = 1$, (ensuring that $f$ is a density function corresponding to a probability measure). Equation (3.1) is identical to the constraints in problem (2.10) of Section 2.2.4.

Let $\mathcal{P}$ denote the set of all density functions defined on $\mathcal{X}$. Then (3.1) defines $n + 1$ constraints on $\mathcal{P}$, and we denote the subset of density functions in $\mathcal{P}$ that satisfy all constraints as $\mathcal{P}(\theta)$. This is the feasible set, the collection of all probability density functions that satisfy the information elicited from the expert. In all but degenerate cases, $\mathcal{P}(\theta)$ is neither empty nor does it contain a single element. We therefore require an additional criterion to distinguish between the elements of the feasible set. The MCE principle offers one such criterion, namely, choose the element in $\mathcal{P}(\theta)$ that is the ‘closest’, with respect to its information distance, to a nominated reference density in $\mathcal{P}$ set by the analyst. The MCE method can thus be summarised as a solution of the following nonlinear problem. Find

$$\arg \min_{f \in \mathcal{P}(\theta)} H(f\|q) = \arg \min_{f \in \mathcal{P}(\theta)} \int_{\mathcal{X}} f(x) \log \frac{f(x)}{q(x)} \mu(dx),$$

(3.2)

where $q$ is the reference density set by the analyst and $H(f\|q)$ is the CE measure. Problem (3.2) is identical to (2.10). The density $f^*$ that is the solution to (3.2) was derived in Section 2.2.4 using the standard method of Lagrange multipliers and has the well known form [34, 43]

$$f^*(x) = q(x) \exp \left( -\sum_{i=0}^{n} \lambda_i g_i(x) \right),$$

(3.3)

where $\lambda_0, \ldots, \lambda_n$ are the corresponding Lagrange multipliers.
Obtaining explicit expressions for the density (3.3) requires solving for the Lagrange multipliers \( \{ \lambda_i \} \). This is a highly non-linear problem when solving the primal problem, as was discussed in Section 2.2.4. Alternatively, [17] demonstrated that these Lagrange multipliers are dual variables in an unconstrained convex problem (see also [11] and [12]).

The following Lemma, and its proof, are a slight modification of results presented in [17].

**Lemma 3.1.** The unconstrained non-linear problem

\[
\inf_{\lambda \in \mathbb{R}^{n+1}} \int_{\mathbb{R}} q(s) \exp \left( -\sum_{i=0}^{n} \lambda_i g_i(s) \right) \mu(ds) + \sum_{i=0}^{n} \lambda_i \theta_i - 1 \tag{3.4}
\]

constitutes the dual problem to the negative of the constrained NLP (3.2).

**Proof.** Let \( p, q \) and \( x \) be real variables, and define a real function

\[
K(p, x) = q \exp(-x) + px,
\]

where \( p \geq 0 \) and \( q > 0 \). Then, since \( \nabla^2 K > 0 \) we have

\[
\inf_x K(p, x) = p - p \log \frac{p}{q}.
\]

More generally, let \( p, q \) and \( x \) be real functions, setting \( p(s), q(s) \) and \( x(s) \) for \( s \in \mathbb{R} \) we have

\[
\int_{\mathbb{R}} p(s) \mu(ds) - \int_{\mathbb{R}} p(s) \log \left( \frac{p(s)}{q(s)} \right) \mu(ds) \leq \int_{\mathbb{R}} q(s) \exp(-x(s)) + p(s)x(s) \mu(ds),
\]

where \( \mu \) is some dominant measure. Now, assuming \( p(s) \) is a probability density function, \( \int_{\mathbb{R}} p(s) \mu(ds) = 1 \). Further, suppose that \( x(s) = \sum_i \lambda_i g_i(s) \) and \( \int_{\mathbb{R}} g_i(s) p(s) \mu(ds) = \theta_i \), for all \( i \). Then the dual inequality is given by

\[
-\int_{\mathbb{R}} p(s) \log \left( \frac{p(s)}{q(s)} \right) \mu(ds) \leq \int_{\mathbb{R}} q(s) \exp \left( \sum_i \lambda_i g_i(s) \right) \mu(ds) + \sum_{i=0}^{n} \lambda_i \theta_i - 1. \tag{3.5}
\]

The expression on the left side of the inequality (3.5) is just the negative of the cross entropy measure. Thus the dual non-linear problem (DNLP), to the general NLP (3.2) is given by (3.4), and it is unconstrained as stated.

Charnes et al. [17] showed that if the primal NLP is feasible, then the DNLP (3.4) has a minimum which is equal to the minimum of the original NLP (3.2), so there is no duality gap. The unconstrained DNLP (3.4) is convex and therefore is easily solved numerically, for example using the generalised reduced gradient algorithm [45].
3.3 A unimodality Constraint in the MCE Formulation

The moment functions in (3.1) may be used to generate two classes of constraints. The first class consists of moment or central moment constraints, corresponding to polynomial moment functions. The second class consists of percentile constraints, for which the moment functions are the indicator function of an interval (finite or semi-finite). Whenever the information elicited corresponds to constraints belonging to these two classes, the standard MCE formulation presented above can be used to obtain the required density.

For example, the first of two constraints corresponding to the information elicited using the 4-step elicitation procedure belongs to the percentile constraints class. Let \( \theta \) denote the elicited confidence estimate, and \( l \) and \( u \) denote the elicited lower and upper bounds respectively (\( l \leq u \)). We can then represent this information as the percentile constraint,

\[
\theta = \int_X I_{[l,u]}(x) f(x) \mu(dx),
\]

where \( I_{[l,u]} \) is the indicator function for the interval \([l, u]\) (\( I_{[l,u]}(x) = 1 \) for all \( x \in [l, u] \) and 0 otherwise).

The second constraint obtained from the information elicited using the 4-step procedure consists of the estimate for the most likely value. This may either correspond to the median or the mode of the density representing the expert’s opinion. The median case is straightforward since it can be expressed as an expectation over the appropriate indicator function. However, if it is the mode, it cannot be formulated as an expectation of the form (3.1). This is particularly problematic since, as mentioned in the introduction, the mode is the more intuitive quantity to elicit and one that humans are relatively better in estimating [28]. With this motivation in mind, we turn our attention to the problem of incorporating a mode constraint in the MCE method.

3.3.1 Feasibility

Whenever a mode (best guess) estimate is elicited a common, and often implicit, assumption is that such a mode is in fact unique [28]. In other words, the expert’s opinion is assumed to be represented by a strictly unimodal distribution. For example, the parametric method used for fitting a distribution to data elicited using the 4-step procedure assumed the expert’s belief can be represented by the unimodal PERT (beta) distribution.

Brockett et al. [13] were the first to attempt incorporating a unimodality constraint in an MCE formulation. Their work borrowed from results by Kemperman [41] on moment transformation techniques. Only the key ideas are presented again here.
We state the following definition of unimodality followed by a theorem due to Shepp, both found, for example, in [23]. The theorem is a reformulation of Khinchin’s criterion in terms of random variables instead of characteristic functions. A proof of the theorem can also be found in [23] [ch. 13, p. 158].

**Definition 3.1.** A distribution function $F$ is said to be unimodal with the mode at the origin iff the graph of $F$ is convex in $(-\infty, 0)$, and concave in $(0, \infty)$. The origin may be a point of discontinuity, but apart from this the unimodality requires that there exist a density $f$ which is monotone in $\mathbb{R} \setminus 0$ (intervals of constancy are not excluded).

Note that the above definition of unimodality requires only that a density exists and that the distribution function is convex below the mode and concave above the mode. Hence, a uniform distribution, for example, is a unimodal distribution, the mode being any value in the support.

**Theorem 3.1.** Let $X$ be a random variable with distribution function $F_X$. Then $F_X$ is a unimodal distribution, with a mode at the origin, iff the random variable $X$ can be written as $X = UY$, where $U$ is a random variable distributed uniformly in $[0, 1]$ and $Y$ and $U$ are independent, that is, iff $F_X$ is of the form

$$F_X(x) = \int_0^1 F_Y \left( \frac{x}{u} \right) \, du,$$

for some distribution function $F_Y$.

Note that (3.7) essentially describes a change of measure between the space of unimodal distributions of the random variable $X$ and the space of distributions of the random variable $Y$.

Using this structural relationship between $X$ and $Y$, Kemperman [41] determined constraints for $Y$ corresponding to constraints of the form (3.1) for $X$. Specifically, if $\bar{X}$ is unimodal with mode $m$, then $X = \bar{X} - m$ is unimodal with mode at the origin, and for any moment function $g$, from Theorem 3.1

$$\mathbb{E}_X [g(X)] = \mathbb{E}_U [g(YU + m)]$$

$$= \mathbb{E}_Y \left[ \mathbb{E}_U [g(YU + m) \mid Y] \right]$$

$$= \mathbb{E}_Y \left[ \int_0^1 g(uY + m) f_U(u) \, du \right]$$

$$= \mathbb{E}_Y \left[ \frac{1}{Y} \int_0^Y g(t + m) \, dt \right]$$

$$= \mathbb{E}_Y \left[ \hat{g}(Y + m) \right]$$
where
\[ \hat{g}(y) = \gamma(g, y) = \frac{1}{y} \int_{0}^{y} g(t) \, dt. \] (3.8)

For \( g \) a.s. continuous, rearranging (3.8) we have
\[ g(x) = \gamma^{-1}(\hat{g}, x) = \hat{g}(x) + x \hat{g}'(x). \] (3.9)

For example, consider applying the function transform (3.8) to the moment function \( g_i(y) = y^k \). Transforming \( g_i \) using (3.8) [13] have shown,
\[ \hat{g}(y) = \frac{1}{y} \int_{-m}^{y-m} t^k \, dt, \]
\[ = \frac{(y + m)^{k+1} - m^{k+1}}{(k + 1)y}, \]
\[ = \frac{1}{k + 1} \sum_{j=0}^{k} \binom{k+1}{j} y^{k-j} m^j. \]

So, the polynomial moment function is transformed to a function of the sum of moments up to \( k \). Alternatively, applying (3.8) to a moment function corresponding to a percentile constraint, \( g_i(y) = I_{[l,u]} \), transforms the function into a continuous function defined over disjoint domains,
\[ \hat{g}_i(y) = \frac{1}{y} \int_{0}^{y} I_{[l,u]}(t + m) \, dt, \]
\[ = \frac{1}{y} \int_{0}^{y} I_{[l-m,u-m]}(t) \, dt, \]
\[ = \begin{cases} \frac{m-l}{|y|} & y < l - m, \\ 1 & l - m \leq y \leq u - m, \\ \frac{u-m}{y} & y > u - m. \end{cases} \]

Differentiating (3.7) with respect to \( x \), Brockett et al. [13] obtained the correspondence between the density functions
\[ f_X(x) = \varphi(f_Y,x) = \int_{x}^{\infty \times \text{sgn}(x)} f_Y(y) \frac{1}{y} \, dy = \begin{cases} \int_{x}^{\infty} f_Y(y) [y]^{-1} \, dy & x < 0, \\ \int_{x}^{\infty} f_Y(y) y^{-1} \, dy & x \geq 0. \end{cases} \] (3.10)

Then, for \( f_X(x) \) differentiable a.s., rearranging (3.10) we obtain the reverse correspondence
\[ f_Y(y) = \varphi^{-1}(f_X,y) = -y f_X'(y), \] (3.11)
for all $y$ in the support of $f_X(y)$. The operators $\varphi$ and $\gamma$ are bijections generated by the change of measure implied by Theorem 3.1.

The method of [13], presented in the context of elicitation, can now be stated as follows. Let $\tilde{X}$ be a unimodal random variable following the distribution best representing the expert’s opinion regarding the uncertain quantity of interest. Using (3.8) and (3.11), transform all constraints and the reference distribution defined in terms of $X$, to the auxiliary variable $Y$. Then, defining $\tilde{P}(\theta)$ to be the transformed feasible set, pick the element of $\tilde{P}(\theta)$ that minimises the CE with respect to the transformed reference density. Finally transform the optimal probability density back to the original variable $X$.

That is, Brockett et al. [13] solved the original problem (3.2) subject to the transformed constraints, and with respect to the transformed reference distribution. The optimal density for the auxiliary variable $Y$ is then, from (3.3),

$$f_Y^*(y) = \varphi^{-1}(q, y) \exp \left( - \sum_{i=0}^{k} \lambda_i \hat{g}_i(y) \right),$$

where the $\{\lambda_i\}$ were obtained using the corresponding DNLP.

Transforming back using (3.10) gives the unimodal density,

$$f_X^*(x) = \begin{cases} 
\int_{-\infty}^x \varphi^{-1}(q, y) \exp \left( - \sum_{i=0}^{k} \lambda_i \hat{g}_i(y) \right) dy & x \leq 0, \\
\int_x^\infty \varphi^{-1}(q, y) \exp \left( - \sum_{i=0}^{k} \lambda_i \hat{g}_i(y) \right) y^{-1} dy & x \geq 0,
\end{cases}$$

that satisfies the constraints (3.1) by construction, and is unimodal by Theorem 3.1.

So, the transformation method proposed by [13] indeed produces a feasible probability density. But, is it the probability density that minimises the CE on the original feasible set with respect to the original reference density? In other words, is it optimal?

It turns out that the answer is no. This is because the change of measure prescribed by Theorem 3.1 does not preserve the CE ordering.

Brockett et al. [13] transformed both constraints and prior distribution to the auxiliary space, but they did not perform a change of measure for the original CE objective functional. Instead they used the CE functional with respect to the transformed reference distribution on the set of probability distributions that satisfy the transformed constraints. This is a subtle, but important, distinction. The resulting distribution, while unimodal and satisfying the constraints, is not the one that minimises the information distance to the original reference distribution set by the analyst. Consequently, the resulting distribution is not the one that fully utilises the information provided by the constraint set and reference distribution, while remaining maximally noncommittal with respect to unavailable information.
3.3.2 Optimality

In order to obtain the auxiliary objective functional that is the result of a change of measure, according to Theorem 3.1, we first note the following. The CE functional in (3.2) can be written as an expectation of the random function \( \log \left( \frac{f_X(X)}{q(X)} \right) \). Then (using (3.8)) for all probability densities \( f_X \) that are unimodal we have

\[
\mathbb{E}_X \left[ \log \frac{f_X(X)}{q(X)} \right] = \mathbb{E}_Y \left[ \frac{1}{Y} \int_0^Y \log \frac{\varphi(f_Y, t)}{\varphi(\hat{q}, t)} \, dt \right] = \Psi(f_Y)
\]

Therefore

\[
\arg \min_{f_X \in \mathcal{P}(\theta)} \mathbb{E}_X \left[ \log \frac{f_X(X)}{q(X)} \right] = \varphi \left( \arg \min_{f_Y \in \mathcal{P}(\theta)} \mathbb{E}_Y \left[ \frac{1}{Y} \int_0^Y \log \frac{\varphi(f_Y, t)}{\varphi(\hat{q}, t)} \, dt \right] \right). \tag{3.12}
\]

That is, the optimal density, which is the solution to our original problem (3.2) subject to unimodality constraint, is the transform of the density that minimises \( \Psi(f_Y) \). The auxiliary problem that leads to the optimal solution is then given by the expression in the brackets on the right hand side of (3.12), namely

\[
\arg \min_{f_Y \in \mathcal{P}(\theta)} \mathbb{E}_Y \left[ \frac{1}{Y} \int_0^Y \log \frac{\varphi(f_Y, t)}{\varphi(\hat{q}, t)} \, dt \right]. \tag{3.13}
\]

We consider the direct solution of the auxiliary problem (3.13) in Chapter 4. In the rest of this chapter we examine relaxation solutions to our original problem under a special class of constraints. This approach, as well as providing valuable insight, will prove important for the general solution presented in the next Chapter.

3.4 Relaxation Approach

We turn our attention back to the original MCE problem (3.2), subject to a unimodality constraint, and a set of constraints expressible in the form (3.1). The rational behind the relaxation approach is as follows. Define the relaxed feasible set, \( \mathcal{P}(\theta) \), as the set of all densities satisfying the set of all constraints excluding unimodality. Now, if solving the NLP (3.2) subject to \( \mathcal{P}(\theta) \), results in a density that is, in fact, unimodal, then it must also be the solution to (3.2) over \( \mathcal{P}(\theta) \), the original (not relaxed) feasible set.

The density corresponding to the element of \( \mathcal{P}(\theta) \) that minimises the CE with respect to some reference density \( q \) was given in (3.3)
Now, assume \( q \) is a unimodal density function with mode at \( m \). Then, for \( f^*_X \) to also be a unimodal density with mode at \( m \), it is sufficient to have \( \exp (-\sum_{i=0}^{n} \lambda_i g_i(x)) \) unimodal with mode at \( m \). There are many combinations of moment functions satisfying such requirements.

Consider for example the 4-step procedure. The relaxed constraint set in this case is just

\[
\left\{ \int_{[l,u]}(x) f_X(x) \, dx = \theta, \int f_X(x) \, dx = 1 \right\},
\]

and, from lemma 3.1, the DNLP is given by

\[
\inf_{\lambda \in \mathbb{R}^2} \int_{\mathbb{R}} q(s) \exp \left( -\lambda_0 - \lambda_1 I_{[l,u]}(s) \right) \, ds + \lambda_0 + \lambda_1 \theta.
\]

When the original problem is feasible, from (3.3), the optimal density is

\[
f^*_X(x) = q(x) \exp \left( -\lambda_0^* - \lambda_1^* I_{[l,u]}(x) \right) \tag{3.14}
\]

where \( \lambda_0^*, \lambda_1^* \) are the arguments that minimise the DNLP.

We have the following result.

**Proposition 3.1.** Assume \( q(x) \) is a unimodal probability density function with mode at \( m \in [l, u] \), \( l < u \). Then \( f^*_X(x) \) in (3.14) is unimodal with mode at \( m \) iff \( \int_l^u q(x) \, dx \leq \theta \).

**Proof.** Differentiating the Lagrangian with respect to \( \lambda_0 \) and \( \lambda_1 \), and equating to 0, we have

\[
\exp(-\lambda_0^*) = (1 - \theta) \left( \int_{-\infty}^l q(x) \, dx + \int_u^\infty q(x) \, dx \right)^{-1}, \tag{3.15}
\]

and

\[
\exp(-\lambda_1^*) = \theta \left( \int_{-\infty}^l q(x) \, dx + \int_u^\infty q(x) \, dx \right) \left( 1 - \theta \right) \int_l^u q(x) \, dx^{-1}. \tag{3.16}
\]

Assume the condition \( \theta \geq \int_{l}^{u} q(x) \, dx \) holds. Then, noticing that

\[
\int_{l}^{u} q(x) \, dx = 1 - \left( \int_{-\infty}^{l} q(x) \, dx + \int_{u}^{\infty} q(x) \, dx \right),
\]

by the assumption that \( q(x) \) is a probability density, we have

\[
(1 - \theta) \leq \left( \int_{-\infty}^{l} q(x) \, dx + \int_{u}^{\infty} q(x) \, dx \right).
\]
Hence, from (3.15), $\exp(\lambda_0^*) \geq 1$ and $\lambda_0^* \geq 0$. Furthermore, combining $\exp(\lambda_0^*) \geq 1$ and $\theta \geq \int_l^u q(x) \, dx$ in (3.16), we see that $\lambda_1^* \leq 0$. Now, since $q(x)$ is unimodal with mode $m \in [l, u]$, and $\exp(-\lambda_0^* - \lambda_1^*) \geq \exp(-\lambda_0^*)$, we have $f_X^*(x) = q(x) \exp(-\tilde{\lambda}x_0 - \tilde{\lambda}x_1 I_{[l,u]}(x))$ is non-decreasing (non-increasing) for $x \leq m$ ($x \geq m$), $x \in \text{supp}(f_X^*)$, and therefore $f_X^*$ is unimodal with mode at $m$.

For necessity, assume $\theta < \int_l^u q(x) \, dx$. Then

$$(1 - \theta) > \left(\int_{-\infty}^l q(x) \, dx + \int_u^\infty q(x) \, dx\right),$$

and hence, from (3.15) and (3.16), $\lambda_0^* < 0$ and $\lambda_1^* > 0$, and so $\exp(-\lambda_0^* - \lambda_1^*) < \exp(-\lambda_0^*)$.

We have

$$\lim_{x \to l^-} f_X^*(x) = q(l) \exp(-\lambda_0^*) > q(l) \exp(-\lambda_0^* - \lambda_1^*) = \lim_{x \to l^+} f_X^*(x),$$

and $f_X^*(x)$ is NOT non-decreasing at $x = l < m$, $x \in \text{supp}(f_X^*)$ (a similar argument can be carried out at $x = u$). Therefore $f_X^*$ is not monotonic above and below the mode and thus is not a unimodal distribution, and the proof is completed. \hfill \Box

Note that the proof of Proposition 3.1 only uses monotonicity. Therefore, Proposition 3.1 holds also for the case of the uniform reference density, and the uniform optimal density, as per the Definition 3.1.

### 3.5 An application

To illustrate the use of the relaxation approach we consider elicitation preformed using the 4-step procedure. First, from (3.15) and (3.16) we obtain the following analytical expressions for $\lambda_0^*$ and $\lambda_1^*$,

$$\lambda_0^* = -\log\left(\frac{1 - \theta}{1 - \int_l^u q(x) \, dx}\right),$$

$$\lambda_1^* = -\log\left(\frac{\theta(1 - \int_l^u q(x) \, dx)}{(1 - \theta) \int_l^u q(x) \, dx}\right).$$

The optimal density is then given analytically by substituting (3.17) and (3.18) into (3.3) with $n = 1$, $g_0 = 1$ and $g_1 = I_{[l,u]}$. As an aside, the fact that the Lagrangian multipliers can be obtained analytically is of particular interest to analysts applying the method. This is because it avoids the need for sophisticated numerical procedures and allows the
implementation of the relaxation approach in a simple spreadsheet using common software [66].

We have the following corollaries to Proposition 3.1 for particular choices of reference distributions. Further corollaries can be constructed in the same fashion for alternative reference densities.

**Corollary 3.1.** Assume \( q(x) = I_{[A,B]}(x)/(B - A) \), for some \( A \leq l \) and \( B \geq u \). That is \( q(x) \) is the uniform density. Then \( f_X^* \) is unimodal with mode \( m \in [l,u] \) iff \( (u - l)/(B - A) \leq \theta \).

**Corollary 3.2.** Assume \( q_{\sigma,m}(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\left(x - m\right)^2/(2\sigma^2)\right) \), and let \( N_{m,l,u}(\sigma) = \varphi^m_{\sigma}(u) - \varphi^m_{\sigma}(l) \), where \( \varphi^m_{\sigma} \) is the normal cdf with mean \( m \) and variance \( \sigma^2 \) \((u > l)\). Then there exists a critical value \( \bar{\sigma} = N_{m,l,u}^{-1}(\theta) \) such that \( f_X^* \) is unimodal with mode at \( m \) iff \( \sigma \geq \bar{\sigma} \).

**Corollary 3.3.** Assume \( q_{\beta,m}(x) = x^{t_m(\beta)}(1-x)^{\beta-1}/B(t_m(\beta), \beta) \) where \( \beta > 1 \), \( t_m(\beta) = (m(\beta - 2) + 1)/(1-m) \) and \( B \) is the Beta function. Also let \( B : \mathbb{R} \to [0,1] \) be the real valued function \( B_{m,l,u}(\beta) = \int_l^u q_{\beta,m}(t) \, dt \). Then there exists a critical value \( \bar{\beta} = B_{m,l,u}^{-1}(\theta) \) such that \( f_X^* \) is unimodal with mode at \( m \) iff \( \beta \leq \bar{\beta} \).

We now illustrate the relaxation approach using the following three different elicitation scenarios.

**3.5.1 Case 1:**

In the first scenario, the elicited confidence is assumed to be 100% \((\theta = 1)\). That is, the bounded elicited range corresponds to the entire support of the distribution representing the expert’s opinion. As a result the percentile and normalisation constraints are in fact equivalent and we are left with essentially one constraint.

Figure 3.1 compares the density representing the expert’s opinions obtained using the relaxation approach and that obtained using the transformation approach suggested in [13]. Recall that the method suggested by [13] used the MCE approach on the auxiliary space, and did not lead to optimality.

In particular, since in this scenario the uniform reference density set by the analyst is in fact a member of the feasible set by Corollary 3.1, it must be the density that minimises the information distance, with a CE measure of 0 (a result known as Gibb’s inequality [64]). As illustrated in Figure 3.1, the relaxation approach indeed provides the optimal solution, namely the uniform density over the elicited range.
Also clearly demonstrated in Figure 3.1, the transformation method suggested by Brockett and colleagues, provides a feasible, but not the optimal density. This can also be seen from the corresponding analytical derivation. For this special case, minimising the cross entropy is equivalent, up to a constant, to maximising entropy [13]. From (3.4) the DNLP in this special case is given by

$$\inf_{\lambda \in \mathbb{R}} \int_l^u e^{-\lambda s} \, ds + \lambda_0.$$ 

Differentiating and equating to zero we obtain

$$f^*(y) = \frac{1}{u - l} I_{[l,u]}(y).$$

Finally, substituting into (3.3.1), we have

$$f^*_X(x) = \begin{cases} 
\frac{1}{u - l} \log \left( \frac{m - l}{m - x} \right) & l \leq x \leq m, \\
\frac{1}{u - l} \log \left( \frac{u - m}{x - m} \right) & m \leq x \leq u, \\
0 & \text{elsewhere.}
\end{cases}$$

This is the expression for the logarithmic function plotted in Figure 3.1, and is clearly not optimal, as discussed above.

Figure 3.1: Illustrating the resulting unimodal MCE densities for case 1 scenario. The information elicited is assumed to be: range $[2, 8]$, mode at 7, and the confidence 100%.
### 3.5.2 Case 2:

In the second scenario the confidence elicited is assumed below 100% ($\theta < 1$). It is further assumed that the potential values of the quantity of interest cannot be bounded by any finite interval. To reflect this second assumption, the analyst is required to specify a reference distribution with support over the entire real line.

As an example, assume the analyst has set a normal reference distribution, with mean centred at the estimated mode, and variance to be determined by the analyst. In general the variance is set relatively large to reflect ignorance of the analyst. Larger variance also agrees with the requirement of Corollary 3.2 ensuring unimodality of the optimising density.

The optimising density for the relaxation approach is obtained from (3.14) with $q(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$, and using (3.17) and (3.18). It is given by

$$f^*_X(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \left(\frac{1 - \theta}{1 - N_{m,l,u}(\sigma)} I_{[\mathbb{R}\setminus[l,u]}(x) + \frac{\theta}{N_{m,l,u}(\sigma)} I_{[l,u]}(x)\right), \quad (3.19)$$

where $N_{m,l,u}(\sigma)$ is the normal cdf and we used the notation $N_{m,l,u}(\sigma) = \varphi_{\sigma}(u) - \varphi_{\sigma}(l)$ (as per corollary 3.2).

Figure 3.2 illustrates the optimising density, (3.19), for two different $\sigma$ values. Both variance levels are above the threshold level indicated in Corollary 3.2, resulting in both cases in unimodal densities. As expected, as the variance of the normal reference distribution increases, the shape of the optimising density appears to approach that of a constant density on the elicited range and outside of it.

### 3.5.3 Case 3:

In the third scenario the confidence elicited is still assumed below 100% ($\theta < 1$), but the potential values of the quantity of interest are now assumed to be bounded by a finite interval $[A,B]$. This might be the case, for example, if the quantity of interest is a probability ($A = 0$, $B = 1$).

Further assume that the analyst has set a beta reference distribution with support over the interval of potential values. The analyst is required to set only one of the beta distribution parameters, as the second one is given by the mode estimate and the value of the first one.
Figure 3.2: Illustrating the resulting unimodal MCE densities for case 2 scenario. The elicited information is assumed to be: range [2, 8], mode at 7, and 80% confidence. The standard deviation of the normal reference density was set to either 100 or 5.

This third scenario is equivalent to conditions in Corollary 3.3. We have,

$$q(x) = \frac{x^{t_m(\beta)-1}(1-x)^{\beta-1}}{B(t_m(\beta), \beta)},$$

and using (3.17) and (3.18) the optimising density is given, from (3.3), by

$$f^*_X(x) = \frac{x^{t_m(\beta)-1}(1-x)^{\beta-1}}{B(t_m(\beta), \beta)} \left( \frac{1 - \theta}{1 - B_{m,l,u}(\beta)} I_{[A,B]\setminus[l,u]}(x) + \frac{\theta}{B_{m,l,u}(\beta)} I_{[l,u]}(x) \right). \quad (3.20)$$

Figure 3.3 illustrates (3.20) for two different $\beta$ values. As is clear from Figure 3.3, both values are sufficiently small to ensure unimodality of the optimising density in accordance with Corollary 3.3.

3.5.4 Resolving discontinuities in the optimal density

In all three cases above the optimal density exhibits discontinuities occurring at the boundary of the elicited range. These are due to the discontinuities in the indicator moment function used in the constraint (3.6). Mathematically it is possible to eliminate such discontinuities by replacing the indicator moment function with an appropriate continuous approximation. This can be achieved using sigmoidal function approximations of the
Figure 3.3: Illustrating the resulting unimodal ME densities for case 3 scenario. The finite interval of possible values is assumed to be $[0, 1]$. The elicited information is assumed to be: range $[0.3, 0.8]$, mode at 0.4, and 80% confidence. The reference distribution set by the analyst is assumed to be the beta distribution with either $\beta = 3$ or $\beta = 5$.

indicator function. For example, consider the moment function

$$g_{\delta,l,u}(x) = \begin{cases} 
\frac{1}{2} (1 + \text{erf} \left((x - l + \delta)^{-1} + (x - l - \delta)^{-1}\right) ) & l - \delta \leq x \leq l + \delta, \\
1 & l + \delta \leq x \leq u - \delta, \\
\frac{1}{2} (1 + \text{erf} \left((x - u + \delta)^{-1} + (x - u - \delta)^{-1}\right) ) & u - \delta \leq x \leq u + \delta, \\
0 & \text{elsewhere,}
\end{cases} \tag{3.21}$$

where $\text{erf}$ stands for the error function. The parameter $\delta$ is set by the analyst and defines a region about the lower and upper elicited values. The limit of $g(x, \delta)$ as $\delta \to 0$ is the indicator function, and it is continuous, as well as everywhere differentiable for $\delta > 0$. The value of $\delta$ must clearly be smaller then half the length of the elicited range, and depending on the reference distribution chosen, should be less then the minimum distance from the elicited range to the mode.

An important point to note is that since the indicator moment function has been replaced with (3.21) the corresponding constraint has also changed. In particular, the new constraint no longer describes a percentile requirement, but instead it requires the expectation of the alternative moment function, (3.21), to be equal to the elicited confidence. Nevertheless, the limit of the new constraint as $\delta \to 0$ is clearly the original percentile constraint. In practice, for relatively small values of $\delta$, the deviation of the probability
mass over the elicited range, from the elicited confidence value, is relatively small (see examples below).

Figure 3.4 illustrates the optimising densities when using the relaxation method with the alternative moment function (3.21) for case 2 above. Both densities are everywhere differentiable. As expected, increasing $\delta$ produces a more 'smooth' looking density.

Next we revisit case 3 above. Figure 3.5 illustrates the resulting optimising densities for different values of $\delta$ and $\beta$.

Now, while the alternative moment function (3.21) resolves the discontinuities from a mathematical point of view, its use needs to be justified. One such justification is to consider the case where the analyst wishes to include some uncertainty around the particular lower and upper values elicited. The extent of such uncertainty is determined by the $\delta$ parameter. Of course more parameters, for example a separate $\delta$ value for the lower and upper estimates, can be used. Again, from an information perspective, any addition of assumptions should be justified.

Figure 3.4: Illustrating the resulting unimodal MCE densities for the elicited information illustrated in Figure 3.2 and standard deviation for the normal reference distribution set to 100. Here, instead of the indicator moment function we used (3.21). The two cases illustrated differ in the value of $\delta$ as indicated. The probability mass in the elicited range in each case was: 0.8001 for $\delta = 0.1$ and 0.8097 for $\delta = 1$ (recall that $\theta = 0.8$).
Figure 3.5: Illustrating the resulting unimodal MCE densities for a case 3 scenario, with the elicited information: range \([0.3, 0.8]\), mode at 0.4, and confidence level of 80% \((\theta = 0.8)\). The reference distribution is the Beta distribution, and illustrated are two different values for the \(\beta\) parameter. The probability mass in the elicited range for each of the three examples was: 0.7999 for \(\beta = 3\) and \(\delta = 0.025\), 0.8055 for \(\beta = 3\) and \(\delta = 0.25\), finally 0.8024 for \(\beta = 5\) and \(\delta = 0.25\).

3.6 Chapter summary

We have demonstrated in this chapter the use of CE minimisation for choosing the best probability distribution to represent information elicited from an expert. As well as having a strong mathematical basis, the MCE approach is grounded in information-theoretic philosophical arguments. This is one of its main advantages over other more common methods used in elicitation such as parametric fitting.

Various studies have supported elicitation of a mode, and it is indeed a popular choice in elicitation protocols. Previous attempts to incorporate unimodality requirements in MCE formulation have offered only feasible, but not optimal, solutions. In this chapter we formulated the general problem leading to optimality. We then used a relaxation approach, and demonstrated that there exist combinations of elicited distributional characteristics for which the relaxation approach provides optimal solutions.

Recent studies suggest the 4-step protocol offers significant reduction in expert over-confidence [75]. The MCE relaxation approach was applied to the 4-step protocol in several scenarios, and explicit analytical expressions were obtained. This is particularly of interest as this allows the implementation of the approach using basic spreadsheet software [66].
Applying the MCE approach in the case of the 4-step procedure protocol has resulted in optimising densities exhibiting discontinuities. While, in itself, this fact does not represent a problem, it is sometimes desirable to avoid discontinuities. To that extent, we demonstrated the use of an alternative moment function. Of course, any additional assumption added to the problem by the analyst must ultimately be justified. This highlights one final advantage of the MCE approach, its transparency to all assumptions made. The resulting optimising density is the one that fully utilises the information given while remaining maximally noncommittal otherwise [32].
Chapter 4

Unimodal density estimation – General solution

In this chapter we solve the auxiliary problem (3.13) and derive an expression for the optimal density. The relaxation approach developed in the previous chapter will form part of the general solution. The other part of the solution covers cases for which the relaxation approach fails. As in the previous chapter, we illustrate the results using an example from an expert elicitation scenario.

4.1 Background and Motivation

In Chapter 3 we introduced the problem of unimodal density estimation posed in the context of expert elicitation. Further motivation for the study of unimodal density estimation is gained by highlighting the broader relevance of this problem. In a range of applications, such as engineering and signal processing, it is often required to estimate the probability density associated with some random phenomena under study. In some cases, the density is known to be unimodal. Other information regarding the density, such as some of its moments or percentiles, may also be known. If the explicit parametric form of the density cannot be determined from physical considerations alone then parametric methods may prove unsatisfactory. Assuming that only distributional characteristics, rather than the raw data, is available, common statistical density estimation techniques, such as nonparametric kernel estimation, are not feasible.

A common alternative method for density estimation subject to moment constraints is the MCE method introduced in Section 2.2.4. However, as was demonstrated in Chapter
3, if a standard MCE approach is used, the resulting density may not be unimodal. Using a characterisation of unimodal random variables, Theorem 3.1, we demonstrated in Chapter 3 a method for incorporating unimodality in cross entropy density estimation. The resulting optimization problem (3.13) was solved for a special class of relaxed constraints. In this chapter we derive its general solution.

4.2 General optimal solutions

Let \( \mathcal{X} \) be a subset of \( \mathbb{R} \), \( X \) a random element of \( \mathcal{X} \) with \( F_X \) its continuous a.s. probability distribution defined on \( (\mathcal{X}, \mathcal{F}) \), and \( f_X \) its density (unless otherwise stated, in this chapter density refers to the Radon-Nikodym derivative with respect to the Lebesgue measure). An extension for more general settings will be discussed below. Also, let \( Q \) be some probability distribution defined on \( (\mathcal{X}, \mathcal{F}) \), and \( q \) its density. \( Q \) is termed the reference distribution.

As was discussed in Section 3.3, we are interested in solving (3.2) subject to both moment constraints (3.1), and a unimodality constraint. We can rewrite this problem using the linear expectation operator \( E \) as

\[
\inf_{f_X} \mathbb{E}_X \left[ \log \frac{f_X(X)}{q(X)} \right]
\]

subject to

\[
\mathbb{E}_X [g_i(X)] = \theta_i, \quad i = 0, \ldots, n,
\]

\[
f_X \text{ unimodal, } \quad f_X \geq 0,
\]

where \( g_i : \mathcal{X} \rightarrow \mathbb{R} \) measurable \( \mathcal{F} \), \( g_0 \equiv \theta_0 \equiv 1 \), \( \theta_i \in \mathbb{R} \), unimodality is defined as in Definition 3.1, and \( f_X \) lies in the appropriate functional space. Problem (4.1) is similar to problem (3.2), with \( \mu \) being the Laebesgue measure, and the addition of a unimodality constraint.

Let \( \hat{\mathcal{X}} \) be a subset of \( \mathbb{R} \), \( Y \) a random element of \( \hat{\mathcal{X}} \) with \( F_Y \) its probability distribution defined on \( (\hat{\mathcal{X}}, \hat{\mathcal{F}}) \). Note that \( Y \) could be strictly discrete, continuous, or a mixture of both. Let \( f_Y \) denote its Radon-Nikodym derivative with respect to the appropriate measure. Without loss of generality, we assume \( X \) is unimodal with mode at the origin, then, by Theorem 3.1, \( X = YU \), \( U \) uniform on \([0,1]\) and independent of \( Y \).

Following work by Kemperman [41] we defined in (3.8)–(3.9) and (3.10)–(3.11) the bijective operators \( \gamma \) and \( \varphi \) respectively. We use the shorthand notation \( \gamma(g) \) and \( \varphi(f) \) to denote the respective mapping between the respective functional spaces. We also use the hat notation (\( \hat{\cdot} \)) to denote the resulting function when mapping by \( \gamma \), \( \hat{f}(y) = \gamma(f, y) \).
Proposition 4.1. Problem (4.1) is equivalent to
\[
\inf_{f_Y} \mathbb{E}_Y \left[ \frac{1}{Y} \int_0^Y \log \frac{\varphi(f_Y, t)}{q(t)} \, dt \right]
\]
subject to \( \mathbb{E}_Y [\hat{g}_i(Y)] = \theta_i, \quad i = 0, \ldots, n, \)\( f_Y \geq 0 \),
\[
(4.2)
\]
where \( \hat{g}_i : \mathcal{X} \to \mathbb{R} \) measurable \( \hat{F} \).

Proof. From Theorem 3.1, \( f_X \) unimodal implies the existence of the operators \( \varphi \) and \( \gamma \). Applying \( \gamma \) and \( \varphi \) to the objective, and \( \gamma \) to the constraints, of problem (4.1), we obtain (4.2).

Corollary 4.1. Denoting by \( Z \) a random variable following the probability density \( \varphi(f_Y) \), problems (4.2) and (4.1) are equivalent to
\[
\inf_{f_Y} \mathbb{E}_Z \left[ \log \frac{\varphi(f_Y, Z)}{q(Z)} \right]
\]
subject to \( \mathbb{E}_Z [g_i(Z)] = \theta_i, \quad i = 0, \ldots, n, \)\( f_Y \geq 0 \),
\[
(4.3)
\]
and \( Z = X \) in distribution.

The optimal feasible density that solves problem (4.1) without the unimodality constraint, termed the relaxed problem, was derived in Chapter 3, and is given by
\[
f_X^*(x) = q(x) \exp \left( - \sum_{i=0}^{n} \lambda_i g_i(x) \right),
\]
where, for \( q \) a probability density, \( f_X^* \geq 0 \). For the relaxed problem, \( f_X^* \) is, of course, not necessarily unimodal.

Proposition 4.2. The condition \( f_y \geq 0 \) in (4.2) is equivalent to the constraint \( f_X \) unimodal in (4.1).

Proof. For necessity, let \( X \) be a unimodal random variable with mode at the origin. Then, \( f_X \) is monotone on \( \mathcal{X} \setminus 0 \), and therefore \( \varphi^{-1}(f_X, x) \geq 0 \) for all \( x \in \mathcal{X} \). For sufficiency, without loss of generality, assume \( X \) is a bimodal random variable with one mode at the origin, and another at some positive value of \( x \). Then, \( f_X \) is not monotone on \( x > 0 \), its derivative \( f_X' \) will be positive for some \( x > 0 \), and therefore \( \varphi^{-1}(f_X, x) < 0 \) for some \( x > 0 \).

□
Next we establish the convexity properties of the auxiliary problems (4.2) and (4.3). First we require the following known result [23].

**Lemma 4.1 (Log sum inequality).** For \(\{a_1, \ldots, a_n, b_1, \ldots, b_n\}\) positive real numbers

\[
\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left( \sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.
\] (4.5)

**Proof.** The function \(f(t) = t \log t\) is strictly convex for \(t > 0\) since \(\nabla^2 f > 0\). Therefore, for \((\alpha_i \geq 0, \sum \alpha_i = 1)\) and \(\{t_i\}\) some non-negative sequence, Jensen’s inequality gives

\[
\sum_{i=1}^{n} \alpha_i f(t_i) \geq f \left( \sum_{i=1}^{n} \alpha_i t_i \right).
\] (4.6)

Then, for \(\{a_i\}\) and \(\{b_i\}\) positive sequences, setting \(\alpha_i = b_i / \sum_{i=1}^{n} b_i\) and \(t_i = a_i / b_i\) in (4.6),

\[
\sum_{i=1}^{n} \left( \frac{a_i}{\sum_{i=1}^{n} b_i} \log \frac{a_i}{b_i} \right) \geq \left( \sum_{i=1}^{n} \frac{a_i}{\sum_{i=1}^{n} b_i} \right) \left( \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \right),
\]

and dividing both sides by \(\sum_{i=1}^{n} b_i\) completes the proof.

**Proposition 4.3.** Given a real valued, non-negative function \(f\), the functional

\[
\int_{\mathbb{R}} \varphi(f, x) \log \frac{\varphi(f, x)}{q(x)} \, dx,
\] (4.7)

is convex in \(f\).

**Proof.** Let \(f_1\) and \(f_2\) be two real valued, non-negative functions, and \(\alpha \in [0, 1]\), then, by Lemma 4.1, and the linearity of \(\varphi\),

\[
\varphi(\alpha f_1 + (1 - \alpha)f_2, t) \log \frac{\varphi(\alpha f_1 + (1 - \alpha)f_2, t)}{q(t)}
= \left( \alpha \varphi(f_1, t) + (1 - \alpha)\varphi(f_2, t) \right) \log \frac{\alpha \varphi(f_1, t) + (1 - \alpha)\varphi(f_2, t)}{q(t)}
\leq \alpha \varphi(f_1, t) \log \frac{\alpha \varphi(f_1, t)}{q(t)} + (1 - \alpha)\varphi(f_2, t) \log \frac{(1 - \alpha)\varphi(f_2, t)}{q(t)}.
\]

Then, integrating both sides with respect to \(t\) completes the proof.

**Proposition 4.4.** Problems (4.2) and (4.3) are convex optimisation problems. That is, their respective objective functions and feasible sets are convex.
Proof. The convexity of the feasible sets follows immediately from the linearity of the constraints and \( \varphi \). The convexity of the objective functionals follows from Proposition 4.3 since

\[
\mathbb{E}_Y \left[ \frac{1}{Y} \int_0^Y \log \frac{\varphi(f_Y, x)}{q(x)} \, dx \right] = \int_{\mathbb{R}} f_Y(y) \left( \int_0^y \log \frac{\varphi(f_Y, x)}{q(x)} \, dx \right) \, dy \\
= \int_{\mathbb{R}} \int_{-\infty}^{\infty} sgn(x) f_Y(y) \, dy \log \frac{\varphi(f_Y, x)}{q(x)} \, dx \\
= \int_{\mathbb{R}} \varphi(f, x) \log \frac{\varphi(f_Y, x)}{q(x)} \, dx \\
= \mathbb{E}_Z \left[ \log \frac{\varphi(f, Z)}{q(Z)} \right].
\]

\( \square \)

Introducing multipliers \( \{\lambda_i\} \) and \( \mu(y) \), the Lagrangian for problems (4.2) and (4.3) is

\[
L(f_y, \lambda, \mu) = \int_{\mathbb{R}} f_Y(y) \frac{1}{y} \int_0^y \log \frac{\varphi(f_Y, t)}{q(t)} \, dt \, dy + \sum_{i=0}^n \lambda_i \left( \int_{\mathbb{R}} \hat{g}_i(y) f_Y(y) \, dy - \theta_i \right) \\
- \int_{\mathbb{R}} \hat{\mu}(y) f_Y(y) \, dy \\
= \int_{\mathbb{R}} \varphi(f_Y, y) \log \frac{\varphi(f_Y, y)}{q(y)} \, dy + \sum_{i=0}^n \lambda_i \left( \int_{\mathbb{R}} g_i(y) \varphi(f_Y, y) \, dy - \theta_i \right) \\
- \int_{\mathbb{R}} \mu(y) \varphi(f_Y, y) \, dy. \tag{4.8}
\]

Given the convexity properties of the auxiliary problems (4.2) and (4.3), to check that a feasible density is the global minimum, we only need to establish the first order necessary condition (see section 2.1.1).

**Proposition 4.5.** The

\[
\inf_{f_Y} L(f_y, \lambda, \mu),
\]

is attained at

\[
\varphi(f_Y^*, x) = q(x) \exp \left( - \sum_{i=0}^n \lambda_i g_i(x) + \mu(x) - 1 \right), \tag{4.9}
\]
where,

\[ f^*_Y(x) = \varphi^{-1}(\varphi(f^*_Y, x), x) \]

\[ = -x \exp \left( -\sum_{i=1}^{n} \lambda_i g_i(x) + \mu(x) - 1 \right) \left( q'(x) + q(x) \left( -\sum_{i=1}^{n} \lambda_i g'_i(x) + \mu'(x) \right) \right). \]

(4.10)

**Proof.** The Gâteaux variation of \( L \) at \( f_Y \) in the direction \( \delta_x \) is given by

\[ dL(f_Y, \delta_x) = \frac{\partial}{\partial \epsilon} L(f_Y + \epsilon \delta_x) \bigg|_{\epsilon=0} = \frac{1}{x} \int_{0}^{x} \log \frac{\varphi(f_Y, y)}{q(y)} dy + \sum_{i=0}^{n} \lambda_i \hat{g}_i(x) - \hat{\mu}(x) + 1, \]

(4.11)

where we used

\[ \frac{\partial}{\partial \epsilon} \varphi(f_Y + \epsilon \delta_x, y) = \frac{\partial}{\partial \epsilon} \int_{y}^{\infty \times \text{sgn}(y)} \frac{f_Y(t) + \epsilon \delta(x-t)}{t} dt = \int_{y}^{\infty \times \text{sgn}(y)} \frac{\delta(x-t)}{t} dt = \begin{cases} \frac{1}{x} & y \in [0, x] \\ 0 & \text{otherwise.} \end{cases} \]

Then, the proof is completed by equating to zero, differentiating with respect to \( x \), and using the convexity argument established in Proposition 4.4.

**Theorem 4.1 (Main Result).** Let \( A = \{ x \in \hat{X} \mid \hat{\mu}(x) > 0 \} \), and suppose that \( A \) contains only bounded intervals such that \( \ell(A) = \int_{A} dt < \infty \). Then, the density that is the argument that minimises (4.1) is

\[ \varphi(f^*_Y, x) = \exp \left( c - 1 \right) \mathbb{1}_A(x) + q(x) \exp \left( -\sum_{i=0}^{n} \lambda_i g_i(x) - 1 \right) \mathbb{1}_{A^c}(x), \]

(4.13)

where the constant

\[ c = \frac{1}{\ell(A)} \left( \int_{A} \log q(x) dx - \sum_{i=0}^{n} \lambda_i \int_{A} g_i(x) dx \right). \]

(4.14)
The multipliers are obtained by solving the dual convex problem

\[
\inf_{\lambda \in \mathbb{R}^{n+1}} \ell(A) \exp(c - 1) + \int_{\mathbb{R}\setminus A} q(x) \exp \left( -\sum_{i=0}^{n} \lambda_i g_i(x) - 1 \right) \, dx + \sum_{i=0}^{n} \lambda_i \theta_i
\]

subject to

\[
\sum_{i=0}^{n} \lambda_i \hat{g}_i(x) - \frac{1}{x} \int_{0}^{x} \log q(x) \, dx + c > 0, \quad \forall x \in A,
\]

with dual attainment.

Remark 4.1. For the case when \(A\) is at most a countable collection of points, an analogous result to Theorem 4.1 is obtained by replacing the continuous a.s. measure \(\mu\) by a vector of positive components. The expected value of \(\mu\) will still be equal to zero, and the integral will be taken with respect to the counting measure (summation).

Proof of Theorem 4.1. We examine two cases: \(\hat{\mu} = 0\) (\(A = \emptyset\)), and \(\hat{\mu}(x) > 0\) for some \(x \in \hat{X}\) (\(A \neq \emptyset\)). \(\hat{\mu} = 0\) is equivalent to the relaxation problem, studied in Chapter 3. In particular, since \(\hat{\mu} = 0\) everywhere, we have \(\mu = 0, A = \emptyset\), and from (4.9),

\[
\varphi(f_Y^*, x) = q(x) \exp \left( -\sum_{i=0}^{n} \lambda_i g_i(x) - 1 \right),
\]

where the multipliers are obtained by solving the concave unconstraint dual,

\[
\sup_{\lambda \in \mathbb{R}^{n+1}} - \int_{\mathbb{R}} q(x) \exp \left( -\sum_{i=0}^{n} \lambda_i g_i(x) - 1 \right) \, dx - \sum_{i=0}^{n} \lambda_i \theta_i.
\]

Next, assume that \(\hat{\mu}(x) > 0\) for some \(x \in \hat{X}\), then \(A \neq \emptyset\) and necessarily \(f_Y(x) = 0\) for all \(x \in A\) by complementary slackness. From (4.10)

\[
\left( q'(x) + q(x) \left( -\sum_{i=1}^{n} \lambda_i g'_i(x) + \mu'(x) \right) \right) = 0,
\]

and so

\[
\mu'(x) = -\frac{q'(x)}{q(x)} + \sum_{i=1}^{n} \lambda_i g'_i(x),
\]

which gives

\[
\mu(x) = -\log q(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + c.
\]

Then, substituting (4.16) in (4.9), we obtain (4.13).
To obtain an expression for $c$ note that since $\mathbb{E}_Z[\mu(Z)] = \mathbb{E}_Y[\hat{\mu}(Y)] = 0$, then using (4.13),

$$\int_{\mathbb{R}} \varphi(f_X^*, x) \mu(x) \, dx = \exp(c - 1) \int_A - \log q(x) + \sum_{i=1}^n \lambda_i g_i(x) + c \, dx = 0,$$

and rearranging we obtain (4.14). Finally, the dual constraint $\hat{\mu}(x) > 0$ for all $x \in A$, is equivalent to

$$\hat{\mu}(x) = \gamma(\mu, x)$$

$$= \frac{1}{x} \int_0^x - \log q(x) + \sum_{i=1}^n \lambda_i g_i(x) + c \, dx$$

$$= \sum_{i=1}^n \lambda_i g_i^*(x) - \frac{1}{x} \int_0^x \log q(x) \, dx + c$$

$$> 0.$$ Dual attainment (no duality gap) follows from the convexity of the space of non-negative functions, the convexity of the objective and moment constraint functions, and the satisfying of the Slater constraint qualification condition by the feasibility results obtain from the method proposed by Brockett et al. [13].

The density in equation (4.13) is the optimal solution to problem (4.1). The expression for the optimal density consists of two parts. The first, the solution for all $x \in A$, is a constant in $x$. To understand this result we recall that the set $A$ consists of all values $x \in \hat{\mathcal{X}}$ where the non-negativity constraint is active. Then from complementary slackness condition we must have the auxiliary density equal to zero for all $x \in A$. Then, by the mapping $\varphi$, given in (3.10), the corresponding density in the original problem $f_X$ is a constant.

The second part of the optimal density (4.13), is just the relaxation solution introduced in Chapter 3. That is, for $x \in A^C$, the non-negativity constraint is inactive.
4.3 An application

We now illustrate the general result obtained in Theorem 4.1 by considering the following problem

\[
\inf_{f_X} \mathbb{E}_X \left[ \log \frac{f_X(X)}{q_{\sigma,m}(X)} \right]
\]

subject to \( \mathbb{E}_X [g_{\delta,l,u}(X)] = \theta, \quad f_X \geq 0 \)

\( f_X \) unimodal, mode at \( m \),

(4.17)

where \( g_{\delta,l,u} \) is given by (3.21) with \( \delta > 0 \), and \( q_{\sigma,m}(x) = (2\pi\sigma^2)^{-1/2} \exp \left(-\frac{(x-m)^2}{2\sigma^2}\right) \) the probability density function of a normal random variable with mean \( m \) and variance \( \sigma^2 \).

If we assume \( l = 2, u = 8, m = 3, \theta = 0.8 \) and \( \sigma = 100 \), then, problem (4.17) satisfies the necessary and sufficient conditions of Proposition 3.1. In particular, the set \( A \) is empty in (4.13), and the optimal density is given by the relaxation solution alone. Figure 4.1 illustrates the optimal density for two different \( \delta \) values.

By Corollary 3.2, for a given value of \( \theta \), there exists a value of \( \sigma \) such that the relaxation solution alone is no longer unimodal. In particular, if instead, \( \theta = 0.4 \) and \( \sigma = 4.5 \), then the condition in Proposition 3.1 no longer holds. In this case the set \( A \neq \emptyset \), and the optimal solution is given by expression (4.13). Figure 4.2 illustrates both the relaxed, non unimodal, solution, and the general unimodal solution for those particular values of \( \theta \) and \( \sigma \). Note, as discussed above, that for \( x \in A \) the resulting optimal density is constant.

Problem (4.17) can be thought of as an example in the context of expert elicitation. Consider eliciting information regarding a quantity of interest using the 4-step elicitation. Assume the analyst wishes to incorporate uncertainty regarding the upper and lower elicited range parameters. Then, from Chapter 3, one moment function the analyst may choose to use is (3.21), and, for a normal reference density, (4.17) is the problem that results.

4.4 A note on numerical procedures

To obtain an explicit expression for the optimal density we need to solve the dual problem (4.15) for the Lagrange multipliers. In some cases, these multipliers can be obtained analytically, as was shown for the case of the 4-step protocol with indicator moment function in equations (3.17) and (3.18). However, if the values of the multipliers cannot be solved for analytically, we must use numerical methods. In particular, we have two
Figure 4.1: The assumed elicited information: range [2, 8], mode 3, confidence %80 and normal reference distribution with mean 3 and SD 100.
Figure 4.2: The assumed elicited information: range [2, 8], mode 3, confidence 40, δ = 0.25 and normal reference distribution with mean 3 and SD 4.5.
possible scenarios, the first, when the relaxation solution leads to an optimal unimodal density, and the second, when it does not.

Under the first scenario, the dual problem is given by Lemma 3.1; it is convex in \( \lambda \), and unconstrained. Such problems can be solved using one of several well known methods, for example, the reduced gradient algorithm [45]. In particular, we used the Matlab nonlinear unconstrained optimisation function \textit{fminunc}. When, in addition to the function to be minimised, the gradient is also provided, \textit{fminunc} uses the Trust-Region Method [57]. The standard Trust-Region Method approximates the function to be minimised, \( f \), by a quadratic function \( q \), given by the first two terms in the Taylor expansion of \( f \) at a given starting point \( x \). Informally, the algorithm then progresses by defining an ellipsoid neighbourhood \( N \) about \( x \), minimising \( q \) over \( N \), and updating \( x \) to the new minimum value if \( f \) at \( x \) is greater than \( f \) at the new minimum.

In cases when the relaxation solution does not result in a unimodal density, we must solve the constrained dual (4.15) for the value of the multipliers. Given the nature of the constraint, this is a rather difficult, and computationally expensive, approach. Instead, we opted to reduce the problem to an unconstrained nonlinear optimisation problem and use the Matlab function \textit{fminunc} as in the first scenario above. In particular, our algorithm proceeds as follows. We begin by solving the relaxation approach, we then check the solution for unimodality. If the optimal density is not unimodal, we obtain the corresponding region \( A \) for which the auxiliary density is negative, using the transformation \( \varphi^{-1} \). We then solve (4.15) ignoring the constraints, that is we solve an unconstrained dual problem using \textit{fminunc}. Once we obtain the multipliers, we check again for unimodality, and repeat as required.

The check for unimodality is done using the first derivative at continuity points, and left and right limits at discontinuity points, of the candidate density. This is a relatively simple and fast procedure. In cases when the resulting auxiliary density exhibits negative point masses, the region \( A \) is extended by a predefined \( \epsilon \). The algorithm then proceeds as above, with additional \( \epsilon \) extensions are required. Once the resulting density is unimodal, the algorithm searches on the last \( \epsilon \) increment for a sharper solution, that is, a unimodal solution with a smaller CE value. This last step can be made as accurate as needed, since we are guaranteed an optimal solution by Theorem 4.1.

4.5 Chapter summary

In this chapter we studied the problem of unimodal density estimation using the MCE method. The difficulty with a unimodality requirement stems from the fact that it cannot be represented as an expectation over some real valued function. To address this we used
a reformulation of the Khinchin criterion, due to Shepp, that provides a representation of a unimodal random variable as a product of two independent random variables, one of which is distributed uniformly on the unit interval. Using this result, we transformed the unimodality requirement into an inequality constraint in an auxiliary problem. We then derived again the nonlinear auxiliary optimisation problem that leads to the optimal unimodal density with respect to the CE measure, and subject to a set of moment constraints.

We then discussed briefly the numerical aspects of solving for the Lagrange multipliers, when analytical solutions are not tractable. In particular, when the relaxation solutions result in a unimodal density, the unconstrained dual problem is solved using an implementation in Matlab of the trust-region method. Otherwise, when the relaxation solutions do not lead to a unimodal density, we reduce the constrained dual problem to an unconstrained one, solve for the multipliers, and check for unimodality. Our procedure then updates the region where the unimodality constraint is active, and repeats the process.

We illustrated our results in the context of expert elicitation using the 4-step procedure. Our results in this chapter are significant in a range of applications where it is required to estimate the probability density associated with some random phenomena, and the density is known to be unimodal. Such applications are found in the fields of engineering, signal processing, and expert elicitation, to name a few.
Chapter 5

Decision making under conditions of uncertainty

In this chapter we diverge away from the mathematical problem of unimodal density estimation and turn our attention to the problem of decision making under conditions of uncertainty. In particular, we highlight a Bayesian approach to decision making under uncertainty that combines several previous ideas from economic theory and utility theory. The proposed method utilises expert elicitation for estimating the uncertainty associated with the model parameters.

The context in which we choose to apply our approach is environmental management. The importance of accounting for economic costs, when making environmental management decisions subject to resource constraints, has been increasingly recognised in recent years [62, 51]. In contrast, the uncertainty associated with such costs has often been ignored. The method we propose in this chapter accounts for such uncertainties. In particular, the objective we choose to maximise is the probability of obtaining an outcome above a threshold of acceptability. As outlined in Section 1.1, this is equivalent to maximising an expected utility where the utility function is given by the indicator function of the complement of the adverse event.

To illustrate our approach we revisit a previous study that incorporated cost-efficiency analysis in management decisions based on perturbation analysis of matrix population models. Incorporating their derivation into our framework, we extend their model to address potential uncertainties. We then apply these results to two case studies: management of a koala (*Phascolarctos cinereus*) population, and conservation of an olive-ridley sea turtle (*Lepidochelys olivacea*) population.
5.1 Background and motivation

One of the key problems occupying the conservation community is how to allocate limited resources between multiple possible management actions. This problem applies to a range of conservation questions including the design of reserve networks [62, 56, 50], conservation of threatened species [51, 38], and managing threatened, migratory or invasive species [1, 49].

Several approaches to cost-efficiency analysis have been proposed, depending on the particular circumstances [e.g., 82, 1, 8]. However, these approaches often consider only point estimates of expected costs and benefits, ignoring associated uncertainties. Factors contributing to uncertainty include the lack of information, as well as inherent variability. Indeed the magnitude of such uncertainties may be large, or vary considerably between competing conservation actions [e.g., 56]. Ignoring uncertainties exposes the decision maker to overconfidence and increased likelihood of failure [14].

While rarely considered in the conservation literature, cost uncertainty plays a major role in economic decision theory, where it is closely related to concept of risk. Dating back more than half a century, Roy [65] defined the safety first principle in an attempt to tackle the problem of making economic decisions in the presence of uncertainty. The safety first principle argues that, when faced with uncertainty, one should choose the action that maximises the probability of a satisfactory outcome. This principle is reflected in the prevalent use of Value at Risk (VaR) to measure financial risk [54]. For a given level of probability $\alpha$ (typically 0.95 or 0.99), VaR is defined as the smallest number $\rho$ such that the chance of incurring a loss greater than $\rho$ is no larger than $1 - \alpha$ [54].

Another important contribution to the problem of making economic decisions under conditions of uncertainty is the seminal work by [47, 48] on portfolio selection. Portfolio theory is concerned with the problem of allocating a financial resource (a budget) between various risky investment options termed assets (stocks, bonds etc.). The expected return of an asset is defined as the mean change in its utility value over a given time step. The risk of an asset is defined as the uncertainty associated with its return, commonly measured by its standard deviation.

A portfolio is a linear combination of a set of assets, where the weights associated with each asset represent the fraction of the total resource invested (allocated) in each asset. The portfolio problem is then to determine the value of those weights. Reflecting the notion that individuals prefer higher returns and lower risk, portfolio theory states that the optimal allocation is that which maximises the expected returns of a portfolio given a set level of risk, or equivalently, minimises the risk given a set expected return. The optimal allocation reflects a trade-off between higher expected returns and lower risk. An important result of portfolio theory is that diversifying allocations, meaning distributing
the budget among multiple investment options, offers a way to mitigate risk. The extent of diversification reflects the expected and standard deviation of returns, the degree of correlation between the different assets and the risk appetite of the investor.

Portfolio theory has been applied in the context of environmental management a limited number of times [e.g., 76, 22, 52]. In this context, assets are replaced by different management actions. Just as in the classical portfolio theory, in most environmental management applications, there are uncertainties associated with outcomes, and these have previously been assumed to follow a normal distribution [but see 52]. This assumption is reasonable when the underlying random variables are indeed normally distributed. The normal assumption is also a good approximation when, by the central limit theorem, the portfolio consists of a large enough number of management actions with independent and identically distributed outcomes. In some cases, however, such as when considering only a small number of competing actions, the sum of their outcomes does not follow necessarily a normal distribution. In this paper, motivated by Roy’s safety first principle, we re-formulate the portfolio problem in a more general form where the optimal allocation is defined as the allocation that maximises the probability of obtaining a return above a threshold of acceptability. Determining the optimal allocation in this sense is indeed equivalent to the classical portfolio problem when assets are independent, and the uncertainty is assumed to follow a normal distribution [65, and Lemma 5.2 in section 5.2 below].

We apply the notion of maximising the probability of a favourable outcome, and the portfolio approach, to the problem of conservation of threatened species. Such problems are typically analysed using matrix population models and perturbation analysis [15]. The population matrix model is given by the set of difference equations \( \mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t) \) where \( \mathbf{n}(t) \) is a vector denoting the number of individuals in the population, classified according to age, life stage, or some other characteristics. The elements of the matrix \( \mathbf{A} \) denote transition rates between the different classes. The long term (asymptotic) growth rate of the population is given by the dominant eigenvalue of \( \mathbf{A} \), denoted here by \( \lambda \). A population is said to be asymptotically increasing if the asymptotic growth rate \( \lambda \) is greater than 1, and decreasing otherwise.

Perturbation analysis consists of examining the change (sensitivity) or proportional change (elasticity) of the asymptotic growth rate with respect to changes in the elements of \( \mathbf{A} \), induced by each conservation action. Such analysis, however, only accounts for the biological aspects of the conservation problem.

In order to account for the economic aspects, Baxter et al. [1] combined perturbation analysis with a cost-efficiency analysis. According to such a formulation, the optimal action would be the one offering the greatest change in the asymptotic growth rate of the population per dollar spent. However, the formulation by Baxter et al. [1] does
not account for uncertainty. Applying our suggested approach the optimal allocation is then the one for which a minimally-acceptable change in the asymptotic growth rate of the population, per dollar spent, occurs with the greatest probability. We illustrate our approach numerically by applying it to the conservation of two endangered species; a koala (*Phascolarctos cinereus*) population on Snake Island, and an olive-ridley sea turtle (*Lepidochelys olivacea*) population in Orissa, India.

### 5.2 Safety first portfolio formulation

Consider the case of an environmental manager who must decide how to allocate limited resources between $n$ possible independent management actions in order to achieve some desirable outcome. For simplicity, we can think of the limited resources in terms of a total budget. Next, we define the marginal efficiency of management action $i$ to be the change in the overall management outcome per dollar spent in action $i$. Let the random variable $R_i$ denote the stochastic marginal efficiency of management action $i$. Then, assuming marginal efficiencies are additive and independent of the amount allocated, the stochastic marginal efficiency of a management plan, a mixture of $n$ management actions, is

$$R(x) = \sum_{i=1}^{n} x_i R_i,$$

(5.1)

where $x_i$ is the fraction of the total budget allocated to action $i$ ($\sum_{i=1}^{n} x_i = 1$). Equation (5.1) denotes the overall marginal efficiency that results from a portfolio of $n$ management actions. The manager’s decision is represented by the choice of the amount of the budget allocated to each option, summarised in the vector $x$.

Next we apply the safety first principle. Suppose the manager defines a minimally-acceptable outcome. Then, for a fixed budget, and under the assumptions stated above, this translates to a minimally-acceptable marginal efficiency $\rho$. The optimal allocation is then the one that maximises the probability of obtaining a portfolio marginal efficiency above $\rho$. Mathematically, the optimal allocation is given by the optimisation problem

$$\arg \max_x P\{R(x) \geq \rho\},$$

(5.2)

subject to $0 < x_i < 1$ for $i = 1, \ldots, n$, and $\sum_{i=1}^{n} x_i = 1$. By varying the value of $\rho$ over its entire range, a decision maker is able to explore the trade-off between guarantying a higher minimal marginal efficiency and the probability of achieving it.

Obtaining the optimal allocation in (5.2) requires, in general, first obtaining an expression for the probability distribution of $R(x)$. In the case when the weighted marginal efficiencies, $\{x_i R_i\}$, are conditionally independent (conditioned on $\{x_i\}$), the distribution
of $R(x)$ is given by the convolution of their respective distributions. When assets are independent, and the uncertainty is assumed to follow a normal distribution, the first two moments are sufficient, and there is no need for an explicit expression for the distribution itself. This last assertion is summarised in the following Lemma.

**Lemma 5.1.** Assume the marginal efficiency for each strategy follows a normal distribution, then the portfolio marginal efficiency $R(x)$, defined in Eq. (5.1), also follows a normal distribution with mean $\mathbb{E}[R(x)]$ and variance $\text{Var}[R(x)]$. The following equivalence holds,

$$\arg\min_x P\{R(x) \leq \rho\} = \arg\max_x \frac{(\mathbb{E}[R(x)] - \rho)}{\sqrt{\text{Var}[R(x)]}}.$$  \hspace{1cm} (5.3)

**Proof.** The probability that the marginal efficiency of the portfolio is below a critical value $\rho$ is given by,

$$P\{R(x) \leq \rho\} = P\left\{Z \leq \frac{\rho - \mathbb{E}[R(x)]}{\sqrt{\text{Var}[R(x)]}}\right\} = \Phi\left(\frac{\rho - \mathbb{E}[R(x)]}{\sqrt{\text{Var}[R(x)]}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\rho - \mathbb{E}[R(x)]}{\sqrt{\text{Var}[R(x)]}}} e^{-\frac{u^2}{2}} du,$$  \hspace{1cm} (5.4)

where $Z \sim N(0,1)$ is a standard normal random variable, and $\Phi(z)$ its distribution function. Now, since the integrand in (5.4) is positive, it is clear that $\Phi(z)$ is a monotonically increasing function of $z$, and the equivalence (5.3) follows immediately.

A generalisation of the result from Lemma 5.1 to the case where the marginal efficiency of the portfolio follows other distributions apart from normal, was suggested by [65]. In particular, suppose we are only given the first and second moments of the marginal portfolio efficiency, and no details regarding its explicit distribution. Under such information it is of course impossible to determine precisely the probability of the marginal efficiency of the portfolio being less than or equal to some threshold. However, it is still possible to calculate an upper bound on the probability using Chebyshev’s inequality.

**Lemma 5.2.** Suppose we are only given the values $\mathbb{E}[R(x)]$ and $\text{Var}[R(x)]$, under all possible allocations $x$. The probability of $R(x)$ being less than or equal to some threshold $\rho$ has the following property:

$$P\{R(x) \leq \rho\} \leq \frac{\text{Var}[R(x)]}{(\mathbb{E}[R(x)] - \rho)^2}.$$  \hspace{1cm} (5.5)
Then, obtaining the optimal allocation, $x^*$, given by

$$
\arg \max_x \frac{(E[R(x)] - \rho)}{\sqrt{Var[R(x)]}},
$$

(5.6)

is equivalent to finding the allocation that minimises the Chebyshev upper bound on $P\{R(x) \leq \rho\}$ w.r.t. $x$.

Proof. Using Chebyshev’s inequality we have,

$$
P\{|R(x) - E[R(x)]| \geq (E[R(x)] - \rho)\} \leq \frac{Var[R(x)]}{(E[R(x)] - \rho)^2}.
$$

(5.7)

Then, we can rewrite (5.7) as,

$$
P\{(E[R(x)] - R(x)) \geq (E[R(x)] - \rho)\} = P\{R(x) \leq \rho\} \leq \frac{Var[R(x)]}{(E[R(x)] - \rho)^2};
$$

(5.8)

for $\rho < E[R(x)]$, or

$$
P\{(E[R(x)] - R(x)) \geq (\rho - E[R(x)])\} = P\{R(x) \geq \rho\} \leq \frac{Var[R(x)]}{(E[R(x)] - \rho)^2};
$$

(5.9)

for $\rho \geq E[R(x)]$.

We apply (5.2) in the context of management of threatened species, based on perturbation analysis of matrix population models. In both case studies analysed, the environmental manager is faced with the decision of allocating resources between two independent management actions ($n = 2$). To illustrate our formulation, for each case study, we consider two different models of uncertainty, with either normally or uniformly distributed marginal efficiencies for each action.

Next, we solve the non-linear constrained optimisation problem (5.2) using the KKT conditions (see section 2.1) to obtain analytical expressions for the optimal allocation, and the associated probability of satisfying the threshold of acceptability, for both uncertainty models.

### 5.2.1 Normally distributed uncertainties

For simplicity, we consider the case where the environmental manager must decide how to allocate a budget between only two potential conservation actions. This is, in fact, the
case for both studies used for illustration, and can be readily generalised to an arbitrary number of competing actions.

Let $R_1$ and $R_2$ be the stochastic marginal efficiencies for two separate management strategies. We assume that the marginal efficiency for each strategy follows a normal distribution. Then the marginal efficiency of the portfolio, given by (5.1), also follows a normal distribution with mean and variance,

$$E[R(x)] = xE[R_1] + (1 - x)E[R_2],$$  \hspace{1cm} (5.10)

$$Var[R(x)] = x^2Var[R_1] + (1 - x)^2Var[R_2].$$ \hspace{1cm} (5.11)

We are interested in the optimal allocation ($x^*$) which minimises the probability of obtaining a marginal efficiency below a critical value $\rho$. For the normal distribution, from Lemma 5.1, this is equivalent to maximising the number of standard deviations $\rho$ is below the expected portfolio marginal efficiency. We have the following optimisation problem, find

$$\arg\max_x \Psi(x) = \frac{E[R(x)] - \rho}{\sqrt{Var[R(x)]}} = \frac{xE[R_1] + (1 - x)E[R_2] - \rho}{\sqrt{x^2Var[R_1] + (1 - x)^2Var[R_2]}},$$ \hspace{1cm} (5.12)

s.t $0 \leq x \leq 1$, \hspace{1cm} (5.13)

with a Lagrangian function given by

$$L(x, u) = \frac{xE[R_1] + (1 - x)E[R_2] - \rho}{\sqrt{x^2Var[R_1] + (1 - x)^2Var[R_2]}} + u_1(x - 1) - u_2x,$$ \hspace{1cm} (5.14)

where $u_1$ and $u_2$ are the Lagrangian multipliers of the constraints $g_1(x) = x - 1 \leq 0$ and $g_2(x) = -x \leq 0$, respectively. Problem (5.12) is a non-linear constrained optimisation problem in the single variable $x$. It could be solved using elementary calculus. However, since we wish to describe a general approach applicable to higher dimension problems (that is, portfolios with more than two alternatives), we solve the problem using the KKT conditions, introduced in Section 2.1.1. The first order necessary conditions are (see also (2.4)),

$$\nabla_x L(x^*, u^*) = 0$$ \hspace{1cm} (5.15)

$$u^*_i \geq 0, \hspace{1cm} i \in \{1, 2\}$$ \hspace{1cm} (5.16)

$$u^*_i = 0, \hspace{1cm} \forall i \notin A(x^*),$$ \hspace{1cm} (5.17)

where $A(x^*)$ is the set of active constraints at $x^*$, $A(x^*) = \{i|g_i(x) = 0\}$. We note that any optimal solution to the problem must satisfy conditions ((5.15))–((5.17)), and as such they are necessary (Theorem 2.3). If in addition the objective function, $\Psi$, and constraints $g_i$, are twice continuously differentiable, and $\Psi$ is concave, these conditions are also sufficient, by Theorem 2.4, when the constraints are convex, as is the case here. We examine the following three, mutually exclusive, cases.
**Case 1:** $\rho \leq \min\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}$

In this case the solution to (5.15)–(5.17) is

$$x^* = \frac{\operatorname{Var}[R_2](\mathbb{E}[R_1] - \rho)}{\operatorname{Var}[R_2](\mathbb{E}[R_1] - \rho) + \operatorname{Var}[R_1](\mathbb{E}[R_2] - \rho)},$$

(5.18)

and $u_1^* = u_2^* = 0$, that is both constraints are inactive.

To show that $x^*$ above is indeed the optimal solution, we first obtain an expression for the critical points of the objective function by solving $d\Psi/dx = 0$ where $\Psi(x)$ is given by (5.12). We then have

$$d\Psi/dx = \frac{x \operatorname{Var}[R_1](\rho - \mathbb{E}[R_2]) - (1-x) \operatorname{Var}[R_2](\rho - \mathbb{E}[R_1])}{(x^2 \operatorname{Var}[R_1] + (1-x)^2 \operatorname{Var}[R_2])^{3/2}} = 0.$$  

(5.19)

Solving (5.19) for $x$ we obtain (5.18). We note that from (5.18) there exists only one critical point and that for $\rho \leq \min\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}$, the critical point is indeed in $[0,1]$.

Furthermore, we have

$$\left.\frac{d^2 \Psi}{dx^2}\right|_{x^*} = \frac{\operatorname{Var}[R_1](\rho - \mathbb{E}[R_2]) + \operatorname{Var}[R_2](\rho - \mathbb{E}[R_1])}{(x^*^2 \operatorname{Var}[R_1] + (1-x^*)^2 \operatorname{Var}[R_2])^{3/2}} < 0,$$

(5.20)

since $x^* \in [0,1]$ and the term in the denominator is always positive while the term in the numerator is negative for $\rho \leq \min\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}$. Then, from (5.20), the objective function is concave and conditions (5.15)–(5.17) become sufficient. We conclude that solution (5.18) satisfies both the necessary and sufficient conditions and thus is indeed the global maximum of the objective function (5.12), with the Lagrange multipliers $u_1^* = u_2^* = 0$.

**Case 2:** $\min\{\mathbb{E}[R_1], \mathbb{E}[R_2]\} \leq \rho < \max\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}$

In this case we examine two distinct events, $\mathbb{E}[R_1] < \mathbb{E}[R_2]$ and $\mathbb{E}[R_1] > \mathbb{E}[R_2]$ (note that in the event $\mathbb{E}[R_1] = \mathbb{E}[R_2]$ case 2 is empty). The solutions to conditions (5.15)–(5.17) are

$$x^* = 0, \quad u_1^* = (\rho - \mathbb{E}[R_1]) / \sqrt{\operatorname{Var}[R_2]}, \quad u_2^* = 0 \quad \text{for } \mathbb{E}[R_1] < \mathbb{E}[R_2],$$

(5.21)

$$x^* = 1, \quad u_1^* = 0, \quad u_2^* = (\rho - \mathbb{E}[R_2]) / \sqrt{\operatorname{Var}[R_1]} \quad \text{for } \mathbb{E}[R_1] > \mathbb{E}[R_2].$$

(5.22)
To show that (5.21) and (5.22) indeed satisfy conditions (5.15)–(5.17) we observe that for
\[
\min\{\mathbb{E}[R_1], \mathbb{E}[R_2]\} \leq \rho < \max\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}
\]
and \(x \in [0,1]\), \(d\Psi/dx \neq 0\) and \(d^2\Psi/dx^2|_{x^*} < 0\). We therefore conclude that the function is concave and monotone. Then substituting \(x^* = 0\) and \(x^* = 1\) in (5.15) we have for \(\mathbb{E}[R_1] \leq \rho \leq \mathbb{E}[R_2]\),

\[
u_1^* = (\rho - \mathbb{E}[R_1]) / \sqrt{\text{Var}[R_2]} > 0, \quad \nu_2^* = (\rho - \mathbb{E}[R_2]) / \sqrt{\text{Var}[R_1]} < 0
\]  

(5.23)

and for \(\mathbb{E}[R_2] \leq \rho \leq \mathbb{E}[R_1]\),

\[
u_1^* = (\rho - \mathbb{E}[R_1]) / \sqrt{\text{Var}[R_2]} < 0, \quad \nu_2^* = (\rho - \mathbb{E}[R_2]) / \sqrt{\text{Var}[R_1]} > 0.
\]  

(5.24)

Case 3: \(\rho \geq \max\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}\)

Here, again as in case 2, we examine two distinct events, \((\mathbb{E}[R_2] - \rho) / \sqrt{\text{Var}[R_2]} > (\mathbb{E}[R_1] - \rho) / \sqrt{\text{Var}[R_1]}\) and \((\mathbb{E}[R_2] - \rho) / \sqrt{\text{Var}[R_2]} < (\mathbb{E}[R_1] - \rho) / \sqrt{\text{Var}[R_1]}\). The solutions to conditions (5.15)–(5.17) are

\[
x^* = 0, \quad \nu_1^* = 0, \quad \nu_2^* = (\mathbb{E}[R_1] - \rho) / \sqrt{\text{Var}[R_2]} \quad \text{for} \quad (\mathbb{E}[R_2] - \rho) / \sqrt{\text{Var}[R_2]} > (\mathbb{E}[R_1] - \rho) / \sqrt{\text{Var}[R_1]}
\]  

(5.25)

and,

\[
x^* = 1, \quad \nu_1^* = (\rho - \mathbb{E}[R_2]) / \sqrt{\text{Var}[R_1]}, \quad \nu_2^* = 0 \quad \text{for} \quad (\mathbb{E}[R_2] - \rho) / \sqrt{\text{Var}[R_2]} < (\mathbb{E}[R_1] - \rho) / \sqrt{\text{Var}[R_1]}
\]  

(5.26)

To show that (5.25) and (5.26) indeed satisfy conditions (5.15), (5.16) and (5.17) we first observe that for \(\rho \geq \max\{\mathbb{E}[R_1], \mathbb{E}[R_2]\}\) and \(x^*\) given by (5.18), \(d^2\Psi/dx^2|_{x^*} > 0\). We therefore conclude that the objective function is convex and thus the boundary solutions are local maxima and we proceed by comparing the value of the objective function for each such solution.

So far the analysis of the normal distribution assumed the variance for each of the management programs, \(\text{Var}[R_1]\) and \(\text{Var}[R_2]\), is given. However, in practice, these variances represent some measure of the degree of uncertainty in the estimate of the marginal
efficiency for each management program. To better indicate the dependence of the optimal allocation on the discrepancy in the degree of uncertainty with regards to each program we let

$$\sqrt{Var[R_1]} = \alpha \sqrt{Var[R_2]},$$

(5.27)

where \(\alpha\) represents the proportional difference in degree of uncertainty between the two management programs. Then, summarising the three cases using the relationship in (5.27), the optimal allocation is given by

$$x^* = \begin{cases} 
\frac{(E[R_1] - \rho)}{(E[R_1] - \rho) + \alpha^2 (E[R_2] - \rho)} & \text{for } 0 \leq \rho \leq \min\{E[R_1], E[R_2]\} \\
1 & \text{for } \min\{E[R_1], E[R_2]\} < \rho \leq \max\{E[R_1], E[R_2]\} \\
& \text{and } E[R_1] > E[R_2], \\
& \text{or } \rho \geq \max\{E[R_1], E[R_2]\} \\
& \text{and } (E[R_2] - \rho) \alpha < (E[R_1] - \rho), \\
0 & \text{for } \min\{E[R_1], E[R_2]\} < \rho \leq \max\{E[R_1], E[R_2]\} \\
& \text{and } E[R_1] < E[R_2], \\
& \text{or } \rho \geq \max\{E[R_1], E[R_2]\} \\
& \text{and } (E[R_2] - \rho) \alpha > (E[R_1] - \rho), \\
\end{cases}$$

(5.28)

where the allocation is a function only of the expected marginal efficiency of each management program, the critical value \(\rho\) and \(\alpha\). Also note that the optimal allocation does not depend on \(\sqrt{Var[R_1]}\) or \(\sqrt{Var[R_2]}\), only their relative magnitude \(\alpha\).

As well as the optimal allocation, our approach also provides the probability of satisfying the critical threshold, \(\rho\), for each optimal allocation. From (5.10)

$$\mathbb{E}(R(x^*)) = x^* E[R_1] + (1 - x^*) E[R_2],$$

(5.29)

and from (5.11) and (5.27)

$$\sigma_p^2(x^*) = x^{*2} Var[R_1] + (1 - x^*)^2 Var[R_2] = Var[R_2] \left(\alpha^2 x^{*2} + (1 - x^*)^2\right).$$

(5.30)

The probability is then given by evaluating the normal distribution function for each optimal allocation, \(x^*\), and critical value, \(\rho\),

$$F(x^*, \rho) = \Phi \left( \frac{\rho - \mathbb{E}(R(x^*))}{\sqrt{Var[R(x^*)]}} \right);$$

(5.31)
In order to gain further insight into the characteristics of the optimal allocation, economists resort to a graphical description of the portfolio problem. In the following sections we will present such plots for the particular case studies examined, but for now we just explain the reasoning behind such plots. The graphical interpretation in these plots is obtained by plotting the problem in the standard deviation-mean space. The plot of the set of all possible management portfolios is obtained from (5.29) and (5.30), by solving for \( x \) in both equations and rearranging to give

\[
E_R(\sigma_p) = \frac{\mathbb{E}[R_1] - \mathbb{E}[R_2]}{\text{Var}[R_1] + \text{Var}[R_2]} \left( \text{Var}[R_2] \right) \\
\pm \sqrt{\sigma_p^2 (\text{Var}[R_1] + \text{Var}[R_2]) - \text{Var}[R_1] \text{Var}[R_2]} - \mathbb{E}[R_2],
\]

where \( E_R(\sigma_p) \) is used to emphasise the fact that (5.32) is an expression for \( E(R) \) as a function of \( \sigma_p \).

In particular, economists define the efficiency frontier (EF) as the subset of all feasible portfolios where one cannot simultaneously increase the expected value and decrease the standard deviation. It is given by the graph of (5.32), replacing \( \pm \) with addition only.

Now, by setting the objective function \( \Psi(x) = K \), in (5.12), and rearranging, we obtain the following expression for a family of straight lines in standard deviation-mean space,

\[
E(R) = K\sigma_p + \rho
\]

intersecting the mean axis at \( \rho \). Maximising the objective function can be thus interpreted as choosing the line that intersects the EF, with the maximum slope. For values of \( \rho \leq \min(\mathbb{E}[R_1], \mathbb{E}[R_2]) \) the line with the greatest slope is the one tangent to the EF. The optimal allocation is then given from (5.18), and \( K \) is the number of standard deviations \( \rho \) is below \( \mathbb{E}(R) \). \( K \) is then related to the probability of obtaining a marginal efficiency at least as large as \( \rho \). That is, the greater the slope of the optimal allocation line, the greater the probability of achieving the desired outcome. Also note that when \( \rho > \min(\mathbb{E}[R_1], \mathbb{E}[R_2]) \), there is no tangent to the EF, instead the line intersects the EF at the singular extreme point given by the expected marginal efficiency and standard deviation of the management option that satisfies cases 2 and 3 above, and the corresponding allocation \( (x^* = \{0, 1\}) \). Further increase in \( \rho \) will not change the allocation continuously, instead it will decrease the slope of the objective line, corresponding to decreasing the probability of satisfying the outcome. Also note that the allocation may experience a sharp transition between the two individual options for the case when \( \mathbb{E}[R_1] < \mathbb{E}[R_2] \) and \( \text{Var}[R_1] > \text{Var}[R_2] \) (we shall expand on this below).
5.2.2 Uniformly distributed uncertainties

We consider, again, the case where the environmental manager must decide how to allocate a budget between two potential conservation actions. Let $R_1$ and $R_2$ be the stochastic marginal efficiencies for two separate management strategies. In this section we assume the marginal efficiency for each strategy follows a uniform distribution, $R_1 \sim U(a_1, b_1)$ and $R_2 \sim U(a_2, b_2)$. We have the following results.

**Lemma 5.3.** Let $\tilde{R}_1 \sim U(\tilde{a}_1, \tilde{b}_1)$ and $\tilde{R}_2 \sim U(\tilde{a}_2, \tilde{b}_2)$ be two uniformly distributed random variables, and $\tilde{R}$ be the random variable $\tilde{R} = \tilde{R}_1 + \tilde{R}_2$. Then the distribution function for $\tilde{R}$ is given by,

$$F_{\tilde{R}}(C) = \int_{-\infty}^{C} \frac{1}{(b_2 - \tilde{a}_2)(b_1 - \tilde{a}_1)} [(w - \tilde{a}_1 - \tilde{a}_2)U(w - \tilde{a}_1 - \tilde{a}_2) - (w - \tilde{a}_1 - \tilde{b}_2)U(w - \tilde{a}_1 - \tilde{b}_2)$$

$$- (w - \tilde{b}_1 - \tilde{a}_2)U(w - \tilde{b}_1 - \tilde{a}_2) + (w - \tilde{b}_1 - \tilde{b}_2)U(w - \tilde{b}_1 - \tilde{b}_2)]dw,$$

(5.34)

where $U(t - x)$ is a step function, taking the value 1 when $t \geq x$ and 0 otherwise.

More specifically, probability, $P\{\tilde{R} \leq C\}$ equals:

**Case 1:** $\tilde{a}_1 + \tilde{b}_2 < \tilde{a}_2 + \tilde{b}_1$

$$P\{\tilde{R} \leq C\} = \begin{cases} 
0 & C < \tilde{a}_2 + \tilde{a}_1 \\
\frac{(C - \tilde{a}_2 - \tilde{a}_1)^2}{2(b_2 - \tilde{a}_2)(\tilde{b}_1 - \tilde{a}_1)} & \tilde{a}_2 + \tilde{a}_1 \leq C < \tilde{a}_1 + \tilde{b}_2 \\
\frac{2C - 2\tilde{a}_1 - \tilde{b}_2 - \tilde{a}_2}{2(\tilde{b}_1 - \tilde{a}_1)} & \tilde{a}_1 + \tilde{b}_2 \leq C < \tilde{b}_1 + \tilde{a}_2 \\
1 - \frac{(\tilde{b}_2 - \tilde{b}_1 - C)^2}{2(b_2 - \tilde{a}_2)(\tilde{b}_1 - \tilde{a}_1)} & \tilde{a}_1 + \tilde{b}_2 \leq C < \tilde{b}_1 + \tilde{b}_2 \\
1 & C > \tilde{b}_1 + \tilde{b}_2.
\end{cases}$$
Case 2: \( \tilde{a}_2 + \tilde{b}_1 < \tilde{a}_1 + \tilde{b}_2 \)

\[
P\{ \tilde{R} \leq C \} = \begin{cases} 
0 & \text{if } C < \tilde{a}_2 + \tilde{a}_1 \\
\frac{(C - \tilde{a}_2 - \tilde{a}_1)^2}{2(b_2 - a_2)(b_1 - a_1)} & \text{if } \tilde{a}_2 + \tilde{a}_1 \leq C < \tilde{a}_2 + \tilde{b}_1 \\
\frac{2C - 2\tilde{a}_2 - \tilde{b}_1 - \tilde{a}_1}{2(b_2 - a_2)} & \text{if } \tilde{a}_2 + \tilde{b}_1 \leq C < \tilde{b}_2 + \tilde{a}_1 \\
1 - \frac{(\tilde{b}_2 - \tilde{b}_1 - C)^2}{2(b_2 - a_2)(b_1 - a_1)} & \text{if } \tilde{a}_1 + \tilde{b}_2 \leq C < \tilde{b}_1 + \tilde{b}_2 \\
1 & \text{if } C > \tilde{b}_1 + \tilde{b}_2.
\end{cases}
\]

Case 3: \( \tilde{a}_2 + \tilde{b}_1 = \tilde{a}_1 + \tilde{b}_2 \)

\[
P\{ \tilde{R} \leq C \} = \begin{cases} 
0 & \text{if } C < \tilde{a}_2 + \tilde{a}_1 \\
\frac{(C - \tilde{a}_2 - \tilde{a}_1)^2}{2(b_2 - a_2)(b_1 - a_1)} & \text{if } \tilde{a}_2 + \tilde{a}_1 \leq C < \tilde{a}_2 + \tilde{b}_1 = \tilde{a}_1 + \tilde{b}_2 \\
1 - \frac{(\tilde{b}_2 - \tilde{b}_1 - C)^2}{2(b_2 - a_2)(b_1 - a_1)} & \text{if } \tilde{a}_1 + \tilde{b}_2 = \tilde{a}_2 + \tilde{b}_1 \leq C < \tilde{b}_1 + \tilde{b}_2 \\
1 & \text{if } C > \tilde{b}_1 + \tilde{b}_2.
\end{cases}
\]

**Proof.** The probability density function of the sum of two random variables is given by the convolution of the probability density function of each term.

\[
f_{\tilde{R}} = f_{\tilde{R}_1} \ast f_{\tilde{R}_2}.
\]

(5.35)

From convolution theorem \[64\] we have,

\[
\mathcal{L}\{f_{\tilde{R}_1} \ast f_{\tilde{R}_2}\} = \mathcal{L}\{f_{\tilde{R}_1}\} \mathcal{L}\{f_{\tilde{R}_2}\},
\]

(5.36)

where \( \mathcal{L}\{f\} \) denotes the Laplace transform. Writing the probability density function of \( \tilde{R}_1 \) as,

\[
f_{\tilde{R}_1}(C) = \frac{I(C \in [\tilde{a}_1, \tilde{b}_1])}{\tilde{b}_1 - \tilde{a}_1}
\]

(5.37)

then its Laplace transform is given by,

\[
\mathcal{L}\{f_{\tilde{R}_1}\} = \int_0^\infty \frac{I(C \in [\tilde{a}_1, \tilde{b}_1])}{\tilde{b}_1 - \tilde{a}_1} e^{-sC} dC = \frac{e^{-s\tilde{a}_1} - e^{-s\tilde{b}_1}}{(\tilde{b}_1 - \tilde{a}_1)s}
\]

(5.38)
where we have assumed $\tilde{a}_1 \geq 0$. Similarly we have for $\tilde{R}_2$,

$$\mathcal{L}\{f_{\tilde{R}_2}\} = \frac{e^{-s\tilde{a}_2} - e^{-s\tilde{b}_2}}{(\tilde{b}_2 - \tilde{a}_2)s},$$

(5.39)

and thus,

$$\mathcal{L}\{f_{\tilde{R}}\} = \mathcal{L}\{f_{\tilde{R}_1}\}\mathcal{L}\{f_{\tilde{R}_2}\} = \frac{1}{s^2} \left( \frac{e^{-s(\tilde{a}_1 + \tilde{a}_2)} - e^{-s(\tilde{a}_1 + \tilde{b}_1)} - e^{-s(\tilde{b}_2 + \tilde{a}_1)} + e^{-s(\tilde{b}_2 + \tilde{b}_1)}}{(\tilde{b}_2 - \tilde{a}_2)(\tilde{b}_1 - \tilde{a}_1)} \right).$$

(5.40)

Finally, using the second translation theorem [64] we have,

$$f_{\tilde{R}}(C) = \mathcal{L}^{-1}\{\mathcal{L}\{f_{\tilde{R}}\}\}$$

$$= \frac{1}{(\tilde{b}_2 - \tilde{a}_2)(\tilde{b}_1 - \tilde{a}_1)} [(C - \tilde{a}_1 - \tilde{a}_2) \mathcal{U}(C - \tilde{a}_1 - \tilde{a}_2) - (C - \tilde{a}_1 - \tilde{b}_2) \mathcal{U}(C - \tilde{a}_1 - \tilde{b}_2)$$

$$- (C - \tilde{b}_1 - \tilde{a}_2) \mathcal{U}(C - \tilde{b}_1 - \tilde{a}_2) + (C - \tilde{b}_1 - \tilde{b}_2) \mathcal{U}(C - \tilde{b}_1 - \tilde{b}_2)]$$

(5.41)

where $\mathcal{U}(t - k) = 1$, for $t \geq k$, and 0 otherwise, and,

$$F_{\tilde{R}}(C) = \int_{-\infty}^{C} f_{\tilde{R}}(w)dw$$

$$= \int_{-\infty}^{C} \frac{1}{(\tilde{b}_2 - \tilde{a}_2)(\tilde{b}_1 - \tilde{a}_1)} [(w - \tilde{a}_1 - \tilde{a}_2) \mathcal{U}(w - \tilde{a}_1 - \tilde{a}_2) - (w - \tilde{a}_1 - \tilde{b}_2) \mathcal{U}(w - \tilde{a}_1 - \tilde{b}_2)$$

$$- (w - \tilde{b}_1 - \tilde{a}_2) \mathcal{U}(w - \tilde{b}_1 - \tilde{a}_2) + (w - \tilde{b}_1 - \tilde{b}_2) \mathcal{U}(w - \tilde{b}_1 - \tilde{b}_2)]dw$$

(5.42)

Returning to the problem of portfolio allocation, we are interested in obtaining an expression for the distribution function of the marginal efficiency of the portfolio $R(x)$, as defined in (5.1), for the case when $R_1$ and $R_2$ are uniformly distributed random variables.

**Proposition 5.1.** Let $R_1 \sim U(a_1, b_1)$ and $R_2 \sim U(a_2, b_2)$. Then the distribution function...
of their mixture, $R(x)$, is given by,

\[
F_R(C, x) = \int_{-\infty}^{C} f_R(w, x) dw \\
= \int_{-\infty}^{C} \frac{1}{x(1-x)(b_2-a_2)(b_1-a_1)} \left[ (w - xa_1 - (1-x)a_2) \mathcal{U}(w - xa_1 - (1-x)a_2) \right. \\
\left. - (w - xa_1 - (1-x)b_2) \mathcal{U}(w - xa_1 - (1-x)b_2) \\
- (w - xb_1 - (1-x)a_2) \mathcal{U}(w - xb_1 - (1-x)a_2) \\
+ (w - xb_1 - (1-x)b_2) \mathcal{U}(w - xb_1 - (1-x)b_2) \right] dw 
\] (5.43)

Proof. From Lemma 5.3, substituting \(\{\tilde{a}_1, \tilde{b}_1\} \rightarrow \{xa_1, xb_1\}\) and \(\{\tilde{a}_2, \tilde{b}_2\} \rightarrow \{(1-x)a_1, (1-x)b_1\}\) in (5.34). \(\square\)

**Lemma 5.2.1.** Let \(g : (x, C) \rightarrow \mathbb{R}\) be a continuous function on the domain \(\mathcal{D} = \{(x, C) : x \in (0,1), C \in \mathbb{R}\}\). Then,

\[
u(x, C) := g(x, C) - (C - \alpha x) \mathcal{U}(C - \alpha x), \quad \alpha > C \quad (5.44)
\]
is continuous with respect to both \(x\) and \(C\) for all \((x, C) \in \mathcal{D}\).

**Proof.** We examine the continuity of \(u(x, C)\) about the point \(C = \alpha x\). In particular,

\[
\lim_{C \rightarrow \alpha x^+} u(x, C) = \lim_{C \rightarrow \alpha x^-} u(x, C) = g(x, \alpha x), \quad (5.45)
\]

and,

\[
\lim_{x \rightarrow \frac{C}{\alpha}^+} u(x, C) = \lim_{x \rightarrow \frac{C}{\alpha}^-} u(x, C) = g(C, \alpha), \quad (5.46)
\]

where we used the fact that both \(g(x, C)\) and \((C - \alpha x) \mathcal{U}(C - \alpha x)\) are continuous in \(\mathcal{D}\). \(\square\)

**Lemma 5.4.** The probability distribution function (5.43) is differentiable w.r.t. \(C\) for all \(x \in (0,1)\).

**Proof.** We first note that the probability density function \(f_R(x, C)\) is continuous with respect to \(C\) for \(x \in (0,1)\) using Lemma 5.2.1 since each one of the four accumulation
points of \( f_R(x, C) \) can be analysed locally as in Lemma 5.2.1. Then, using the Fundamental Theorem of calculus, we have,

\[
\frac{\partial}{\partial C} F_R(C, x) = \frac{\partial}{\partial C} \int_{-\infty}^{C} f_R(w, x)dw = f_R(C, x). \tag{5.47}
\]

It is important to note that the distribution function is not differentiable for the special cases \( x = 0 \) or \( x = 1 \). In these cases the density function \( f(x, C) \) collapses to the standard uniform distribution (see Proposition 5.6 below), which is of course discontinuous, and therefore the distribution function is not differentiable.

**Lemma 5.5.** Let \( U : (x, C) \to \mathbb{R} \) be defined by,

\[
U(x, C) := \int_{-\infty}^{C} (w - \alpha x)\mathcal{H}(w - \alpha x) - (w - \beta x)\mathcal{H}(w - \beta x)dw, \quad \beta > \alpha > C \tag{5.48}
\]

over the domain \( \mathcal{D} = \{(x, C) : x \in [0, 1], C \in \mathbb{R}\} \). Then \( U(x, C) \) is differentiable w.r.t. \( x \) for all \( x \in [0, 1] \).

**Proof.** First, \( U(x, C) \) is continuous about \( x = C/\beta \) since,

\[
\lim_{x \to C/\beta^+} U(x, C) = \lim_{x \to C/\beta^-} U(x, C) = \int_{\alpha x}^{C} (w - \alpha x)dw. \tag{5.49}
\]

Then \( U(x, C) \) is differentiable about the point \( x = C/\beta \) since,

\[
\lim_{x \to C/\beta^+} \frac{\partial}{\partial x} U(x, C) = \lim_{x \to C/\beta^-} \frac{\partial}{\partial x} \left( \int_{\alpha x}^{C} (w - \alpha x)dw \right) = \lim_{x \to C/\beta^+} \frac{\partial}{\partial x} \frac{1}{2} (C - \alpha x)^2 = \lim_{x \to C/\beta^+} -\alpha (C - \alpha x) = -\alpha (C - \alpha C/\beta) \tag{5.50}
\]
and,

\[
\lim_{x \to \frac{C}{\beta}} \frac{\partial}{\partial x} U(x, C) = \lim_{x \to \frac{C}{\beta}} \frac{1}{2} \left( 2C(\beta x - \alpha x) + \alpha^2 x^2 - \beta^2 x^2 \right)
\]

\[
= \lim_{x \to \frac{C}{\beta}} (C(\beta - \alpha) + \alpha^2 x - \beta^2 x)
\]

\[
= -\alpha(C - \frac{\alpha C}{\beta}) \quad (5.51)
\]

**Lemma 5.6.** The distribution function, (5.43), is differentiable on \( x \in [0, 1] \).

**Proof.** First, the distribution function (5.43) is continuous on \( x \in [0, 1] \), from Lemma 5.5, and since, using L'Hôpital's rule, we have,

\[
\lim_{x \to 0^+} F_R(C, x) = \frac{C - a_2}{b_2 - a_2}, \tag{5.52}
\]

and,

\[
\lim_{x \to 1} F_R(C, x) = \frac{C - a_1}{b_1 - a_1}. \tag{5.53}
\]

Then, since each one of the four points of the distribution function (5.43) can be analysed locally as in Lemma 5.5 we have our result. \( \square \)

Now, the distribution function for the marginal efficiency of the portfolio is given, from Proposition 5.1, by

\[
P\{R(x) \leq C\} = \int_{-\infty}^{C} \frac{1}{x(1-x)(b_2 - a_2)(b_1 - a_1)}
\]

\[
\times [(w - xa_1 - (1-x)a_2)\mathcal{W}(w - xa_1 - (1-x)a_2)
\]

\[
-(w - xa_1 - (1-x)b_2)\mathcal{W}(w - xa_1 - (1-x)b_2)
\]

\[
-(w - xb_1 - (1-x)a_2)\mathcal{W}(w - xb_1 - (1-x)a_2)
\]

\[
+(w - xb_1 - (1-x)b_2)\mathcal{W}(w - xb_1 - (1-x)b_2)]dw
\]

\[
= \Psi(x). \tag{5.54}
\]
We are interested in the allocation, \( x^* \), that minimises the probability of obtaining a marginal efficiency below a critical value \( \rho \). We have the following non-linear constrained optimisation problem,

\[
\arg \min_x \Psi(x) \\
\text{s.t} \ 0 \leq x \leq 1,
\]

with a Lagrangian given by,

\[
L(x, u) = \Psi(x) + u_1 g_1(x) + u_2 g_2(x) = \Psi(x) + u_1 (x - 1) - u_2 x,
\]

where \( u_1 \) and \( u_2 \) are the Lagrange multipliers of the constraints \( g_1(x) = x - 1 \leq 0 \) and \( g_2(x) = -x \leq 0 \), respectively.

Solving (5.55) using the KKT necessary conditions (5.15)–(5.17) we obtain expressions for the optimal allocation as a function of \( \rho \), the minimal marginal efficiency. First we note that the four step functions in (5.34) divide the space \( \{(x, \rho) | x \in [0, 1], \ \rho \in \mathbb{R}\} \) into six regions. These six regions change depending on the values of the four parameters \((a_1, b_1)\) and \((a_2, b_2)\). In each of these six regions the distribution function, (5.34), and its partial derivative with respect to \( x \), assume a different form. These are,

Region 1: \( xb_1 + (1-x)a_2 \leq \rho \leq x a_1 + (1-x)b_2 \)

\[
F(x, \rho) = \frac{2 \rho - x(b_1 + a_1) - 2(1-x)a_2}{2(1-x)(b_2 - a_2)},
\]

\[
\frac{\partial F}{\partial x} = \frac{2 \rho - b_1 - a_1}{2(1-x)^2(b_2 - a_2)}.
\]

Region 2: \( xa_1 + (1-x)a_2 \leq \rho \leq (xb_1 + (1-x)a_2 \) and \( xa_1 + (1-x)b_2) \)

\[
F(x, \rho) = \frac{(\rho - x(a_1 + a_2) + 2a_2)^2}{2x(1-x)(b_2 - a_2)(b_1 - a_1)},
\]

\[
\frac{\partial F}{\partial x} = \frac{2(\rho - x(a_1 + a_2) + 2a_2) (2a_2 - a_1) x(1-x) - (1-2x) (\rho - x(a_1 + a_2) + 2a_2))}{2x^2 (1-x)^2 (b_2 - a_2)(b_1 - a_1)}.
\]

Region 3: \( xa_1 + (1-x)b_2 \leq \rho \leq xb_1 + (1-x)a_2 \)

\[
F(x, \rho) = \frac{2 \rho - (1-x)(b_2 + a_2) - 2xa_1}{2x(b_1 - a_1)},
\]

\[
\frac{\partial F}{\partial x} = \frac{b_2 + a_2 - 2 \rho}{2x^2(b_1 - a_1)}.
\]
5.2

Region 4: \( (xb_1 + (1-x)a_2 \text{ and } xa_1 + (1-x)b_2) \leq \rho \leq xb_1 + (1-x)b_2 \)

\[
F(x, \rho) = 1 - \frac{(x(b_1 - b_2) + b_2 - \rho)^2}{2x(1-x)(b_2 - a_2)(b_1 - a_1)},
\]

\[
\frac{\partial F}{\partial x} = -\frac{(x(b_1 - b_2) + b_2 - \rho)(2(b_1 - b_2)x(1-x) - (1-2x)(x(b_1 - b_2) + b_2 - \rho))}{2x^2(1-x)^2(b_2 - a_2)(b_1 - a_1)}.
\]

Regions 5: \( \rho < xa_1 + (1-x)a_2 \)

\[
F(x, \rho) = 0,
\]

\[
\frac{\partial F}{\partial x} = 0.
\]

Regions 6: \( \rho > xa_1 + (1-x)a_2 \)

\[
F(x, \rho) = 1,
\]

\[
\frac{\partial F}{\partial x} = 0.
\]

In reporting the expressions for the optimal allocation we distinguish between three different classes of the problem. First, the class when the ranges of marginal efficiency for each action do not intersect, \( a_1 < b_1 < a_2 < b_2 \); the second, when one range is contained within the other, \( a_1 < a_2 < b_2 < b_1 \); and the third, when the ranges overlap, \( a_1 < a_2 < b_1 < b_2 \).

**Class 1: \( a_1 < b_1 < a_2 < b_2 \)**

Within this class, we make a further distinction into different cases based on values of \( \rho \), the minimal marginal efficiency threshold.

Case 1: \( \rho < a_2 \).

In this case the optimal solution is the set of allocations,

\[
x^* \in [0, \min\{1, \frac{a_2 - \rho}{a_2 - a_1}\}],
\]

and \( u_1 = u_2 = 0 \). For all values of \( x \) in (5.69), the probability of the marginal efficiency of the portfolio being below \( \rho \) equals zero, \( P\{R(x) \leq \rho\} = 0 \), which is clearly the minimum.
Case 2: $a_2 \leq \rho < b_2$.

Here the optimal allocation is, $x^* = 0, u_1^* = 0, \text{ and } u_2^* = (2\rho - b_1 - a_1)/2(b_2 - a_2)$.

Case 3: $\rho > b_2$.

In this case, for all $x \in [0, 1]$ the probability that the portfolio marginal efficiency is below the critical threshold $\rho$ is 1. In other words, any allocation will fail to satisfy the threshold.

Class 2: $a_1 < a_2 < b_2 < b_1$

This class represents the case when one range is contained in the other, as mentioned above. It turns out that the position of the contained range within the containing range also has an effect on the solution to the problem. To assist the reporting of the optimal allocation solutions, we fix the contained range within a region of the containing range. In particular, we set the additional assumptions: $a_2 < (b_1 + a_1)/2 < (b_2 + a_2)/2$. We report here only the solutions for this special case, while the solutions for the remaining cases could be obtained in similar fashion. The optimal allocation is,

$$x^* = \begin{cases} [0, \min\{1, \frac{a_2 - \rho}{a_2 - a_1}\}] & u_1^* = u_2^* = 0 \quad \text{for } \rho < a_2 \\ \frac{\rho - a_2}{2\rho - a_2} & u_1^* = u_2^* = 0 \quad \text{for } a_2 \leq \rho \leq \frac{1}{2}(b_1 + a_1) \\ 0 & u_1^* = 0, \ u_2^* = \frac{2\rho - b_1 - a_1}{2(b_2 - a_2)} \quad \text{for } \frac{1}{2}(b_1 + a_1) \leq \rho \leq \frac{a_2 b_1 - a_1 b_2}{(b_1 - b_2) + (a_2 - a_1)} \\ 1 & u_1^* = \frac{b_2 + a_2 - 2\rho}{2(b_1 - a_1)}, \ u_2^* = 0 \quad \text{for } \frac{a_2 b_1 - a_1 b_2}{(b_1 - b_2) + (a_2 - a_1)} \leq \rho < b_1 \\ [0, 1] & u_1^* = u_2^* = 0 \quad b_1 \leq \rho. \end{cases}$$

where, as for class 1, in the last range, $\rho \geq b_1$, the probability that the portfolio’s marginal efficiency is below the threshold is 1.

Class 3: $a_1 < a_2 < b_1 < b_2$

As for class 2, it turns out that the degree of overlap of the two regions affects the solutions to the problem. We therefore again make some additional assumptions, $a_2 < (b_1 + a_1)/2 < (b_2 + a_2)/2 < b_1$, to fix the overlap to a particular region. The optimal allocation is then,

$$x^* = \begin{cases} [0, \min\{1, \frac{a_2 - \rho}{a_2 - a_1}\}] & u_1^* = u_2^* = 0 \quad \text{for } \rho < a_2 \\ \frac{\rho - a_2}{2\rho - a_1 - a_2} & u_1^* = u_2^* = 0 \quad \text{for } a_2 \leq \rho \leq \frac{1}{2}(b_1 + a_1) \\ 0 & u_1^* = 0, \ u_2^* = \frac{2\rho - b_1 - a_1}{2(b_2 - a_2)} \quad \text{for } \frac{1}{2}(b_1 + a_1) \leq \rho < b_2 \\ [0, 1] & u_1^* = u_2^* = 0 \quad b_2 \leq \rho. \end{cases}$$

where, as per the previous 2 classes, in the last range, $\rho \geq b_1$, the probability of actually satisfying the minimum threshold is 0.
This concludes our analysis of the safety first portfolio formulation in the case of the normal and uniform distributed marginal efficiencies. In the next section we apply these results in the context of perturbation analysis. We shall revisit the safety first portfolio formulation in Section 5.6, where we will discuss the case where elicited marginal efficiencies may follow any general distribution.

5.3 Safety first portfolio for perturbation analysis

We are given some, real, nonnegative, age- or stage-classified, population projection matrix $A$. Such matrices are, in general, irreducible and primitive [15]. A matrix is irreducible if, and only if, its life cycle graph contains a path from every node to every other node. An irreducible matrix is said to be primitive if there exists a natural number $n$ such that $A^n > 0$. For irreducible stage-classified matrices, a sufficient condition for primitivity is the existence of at least one self-loop [15].

By Perron-Frobenius theory [15], the projection matrix must have a real and positive dominant eigenvalue, the expected long term population growth rate, which we denote by $\lambda$. Furthermore, the dominant eigenvalue is simple and the right and left eigenvectors ($w$ and $v$ respectively) are also real, strictly positive and unique.

Let $a_{jk}$ be the elements of $A$, and assume that some “lower level” parameters $l_i, i = 1, \ldots, n$, affect one or more elements $a_{jk}$. The sensitivity of $\lambda$ to changes in $l_i$ is defined as

$$\frac{\partial \lambda}{\partial l_i} = \sum_{j,k} \frac{\partial \lambda}{\partial a_{jk}} \frac{\partial a_{jk}}{\partial l_i} = \sum_{j,k} v_j w_k \frac{\partial a_{jk}}{\partial l_i}$$

(5.70)

where $w$ and $v$ are the right and left eigenvectors of $\lambda$ [15].

In the context of environmental management, the lower level parameter ($l_i$) denotes the effect of a particular management action on different characteristics of the population dynamics. For example, consider the following real, non-negative, stage-classified population projection matrix

$$
\begin{pmatrix}
0 & w_2 f & w_3 f & \cdots & w_8 f & w_9 f \\
\gamma_1 q_1 s & (1 - \gamma_2) q_2 s & 0 & \cdots & 0 & 0 \\
0 & \gamma_2 q_2 s & (1 - \gamma_3) q_3 s & \ddots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ddots & (1 - \gamma_8) q_8 s & 0 \\
0 & 0 & 0 & \cdots & \gamma_8 q_8 s & (1 - \gamma_9) q_9 s
\end{pmatrix}
$$

(5.71)

The two lower level parameters of particular interest are the survival, $s$, associated with the diagonal and off diagonal elements of (5.71), and fecundity, $f$, associated with the
first row elements. Substituting \(s\) or \(f\) for \(l_i\), in (5.70), we obtain the sensitivity of \(\lambda\) to changes in the survival and fecundity respectively. The environmental decision is then a question of how much to invest in managing fecundity vs managing survival.

Denote the cost associated with management action \(i\) by \(C_{li}\). In order to incorporate costs in sensitivity analysis, Baxter et al. [1] assumed the marginal costs, \(\partial C_{li}/\partial l_i\), are independent. Such an approximation is valid when the cost function is everywhere differentiable and approximately linear in a small neighbourhood of any value of the parameter. They then defined a deterministic marginal efficiency as the change in the long-term population growth rate for management investment in life-history parameter \(l_i\)

\[
r_i = \frac{\partial \lambda}{\partial C_{li}} = \frac{\partial \lambda/\partial l_i}{\partial C_{li}/\partial l_i};
\]

Using (5.72) to obtain marginal efficiencies for each of a set of competing management actions, [1] suggested that the decision rule should be to choose the management strategy with the highest marginal efficiency. This is equivalent to choosing the strategy that offers the greatest expected effect on the population per dollar spent. Note, however, that the analysis thus far has only considered point estimates; the expression for the marginal efficiencies of each strategy, (5.72), ignores uncertainties.

To incorporate uncertainty we embed the model suggested by Baxter et al. [1] in the safety first portfolio formulation. In particular, we set the costs, and thus the marginal efficiencies, to be random variables in (5.72). We then construct a portfolio of such random marginal efficiencies, as in (5.1), and use the expressions derived to obtain the optimal allocation as well as the probability of satisfying the associated minimally-acceptable marginal efficiency.

5.3.1 koala population on Snake Island

While declining in parts of its range, some koala (Phascolarctos cinereus) populations are in fact over-abundant. The population on Snake Island (Victoria, Australia) is a case of the latter. Therefore the management objective is to reduce the population size, or equivalently decrease the long-term growth rate \(\lambda\) of the population. McLean [53] constructed a \(9 \times 9\) stage-classified population projection matrix for the koala population on Snake Island (Table 5.1). Two alternative management actions were considered [1]. The first, translocation of individuals to the mainland, affected the underlying survival rate \(s\). The second, sub-dermal contraception [55], affected the underlying fecundity rate \(f\). Both these strategies were assumed to target individuals with equal probability across stages 2-9. Matrix (5.71) is in fact a parameterisation of the matrix in Table 5.1 [1].

For the values given in Table 5.1 the dominant eigenvalue of the population matrix
Table 5.1: Stage-based population projection matrix for the koala population on Snake Island, Victoria, Australia [data from 53]. The different stages were determined based on tooth wear.

<table>
<thead>
<tr>
<th>time $t+1$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.3026</td>
<td>0.1663</td>
<td>0.1244</td>
<td>0.0891</td>
<td>0.0556</td>
<td>0.0394</td>
<td>0.0226</td>
<td>0.0118</td>
</tr>
<tr>
<td>2</td>
<td>0.9908</td>
<td>0.5359</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.4580</td>
<td>0.4550</td>
<td>0.0655</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.7272</td>
<td>0.2216</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4617</td>
<td>0.2265</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3538</td>
<td>0.1267</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4693</td>
<td>0.4247</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1762</td>
</tr>
</tbody>
</table>

was calculated to be $\lambda = 1.04074$, and from (5.70) the sensitivities to change in fecundity and survival were $\partial \lambda / \partial f = 21.0844$, and $\partial \lambda / \partial s = 1.3010$ respectively. The expressions for the marginal efficiencies, (5.72), were hence given by [see 1, and Appendix A.1 for detailed derivation]

\[
R_f = \frac{\partial \lambda}{\partial f} \frac{f}{c_f N_{2-9}} = \frac{0.2488}{c_f N_{2-9}},
\]

\[
R_s = \frac{\partial \lambda}{\partial s} \frac{s}{c_s N_{2-9}} = \frac{0.7923}{c_s N_{2-9}},
\]

where $c_f$ and $c_s$ are the management costs per (female) koala for each strategy, and $N_{2-9}$ is the number of individuals in stages 2-9.

Harbutt [30] has estimated the management costs per (female) koala to be between $100 and $120 for contraception $c_f$, and between $300 and $400 for translocation $c_s$. Baxter et al. [1] chose to use the mid point of the reported ranges as the point estimates for obtaining uncertainty-free marginal efficiency values in (5.73) and (5.74). In contrast, as outlined earlier, our approach incorporated the reported uncertainty by considering $c_f$ and $c_s$, and therefore $R_f$ and $R_s$, to be independent random variables. Given the lack of information regarding the nature of the distributions that characterizes the uncertainties associated with this problem, we chose to fit the distributions over the marginal efficiencies rather than the costs since it offered greater mathematical tractability.

In particular, we examined two general types of uncertainty. In the first we assumed that the marginal efficiency of each action follows a normal distribution, where the mean is given by the marginal efficiency corresponding to the mid point of the respective interval of reported cost, and the variance is given by assuming the intervals for the costs represent
probability intervals of varying value of percentiles (20%, 50%, 90%, 95% or 99%). In the second case, we assumed that the marginal efficiencies are distributed uniformly over the interval corresponding to the respective cost interval.

Next we constructed the portfolio. For the special case of two competing management actions, (5.1) is reduced to

$$R(x) = xR_f + (1 - x)R_s,$$  \hspace{1cm} (5.75)

where $x$ is the fraction of total budget allocated to managing fecundity (sub-dermal contraception). We then used our derived analytical expression to obtain the optimal allocation under each of the two uncertainty models, for a given minimally-acceptable marginal efficiency level $\rho$, and the associated probability. Repeating this process for a range of $\rho$ values provided the entire spectrum of interplay between the optimal allocation and the probability of satisfying the critical value (Figures 5.1 and 5.2).

Figure 5.1 illustrates the range of possible portfolios plotted in the standard deviation-mean space. This is the classic efficiency frontier (EF) in economics, describing the trade-off between returns and risk. The dotted curve represents the set of all possible management portfolios (5.32), and the EF is denoted by the darker part of the dotted curve given by the addition only part of (5.32).

5.3.2 Olive ridley sea turtle population

Most sea turtle nesting populations are depleted worldwide. While some populations are increasing as a result of successful conservation efforts, many nesting populations remain very low with little or no signs of recovery [68]. Ongoing threats to sea turtle populations include intentional and accidental harvest and the destruction of nesting habitat.

Two of the main sea turtle conservation management actions practised worldwide are the inclusion on trawling vessels of Turtle Excluding Devices (TEDs), and protection of nesting habitat [19, 20]. Fisheries that have included TEDs have reported significant reduction in by-catch of adult sea turtles [18]. At the same time, protecting nesting beaches increases hatchlings’ survival rate by reducing egg harvest, nest destruction, and hatchling disorientation from artificial lights [69]. Using our safety first portfolio formulation we illustrate how a conservation manager may go about allocating a limited budget between the two potential conservation actions.

To apply the safety first portfolio method one requires both a population matrix and estimates of conservation costs. Unfortunately, as far as we are aware, no sea turtle population exists for which such data has been published. Instead, we used conservation cost estimates for a population of olive ridley sea turtle (*Lepidochelys olivacea*) in Orissa, India (K. Shanker, personal communication), combined with published stage-based population
Figure 5.1: Plot of efficiency frontier (dashed line) in standard deviation-mean space, using the normal distribution for uncertainties and $\alpha = 1.09$. The point of intersection of the straight lines and the EF correspond to the optimal portfolio for a given critical efficiency value $C$. The optimal allocation, $x^*$ is 0.4354, 0.4185, 0.3192, 0.0870, and 0.0010, for critical efficiency values, as a fraction of $E[R_f]$, 0, 0.5, 0.9, 0.99, and 0.9999 respectively.
Figure 5.2: Plot of optimal allocation for the koala population management example (a), and associated probability of satisfying the minimally-acceptable threshold $\rho$ (b), for the case of normally distributed uncertainties. (a) The blue and red lines denote the expected marginal efficiencies for the contraceptive and translocation strategies respectively. The solid line denotes the optimal fraction of the budget allocated to contraceptive; (b) The y-axis denotes the probability of satisfying the minimally-acceptable threshold. Each curve represents this probability for a different assumption regarding the percentile over the interval in marginal efficiency defined by the estimated per individual costs. The proportional difference in degree of uncertainty between the two management options, $\alpha$, is kept constant across all cases.
Figure 5.3: Plot of optimal allocation for the koala population management example (a), and associated probability of satisfying the minimally-acceptable threshold $\rho$ (b), for the case of uniformly distributed uncertainties. (a) The blue and red lines denote the minimum, expected and maximum marginal efficiencies for the contraceptive and translocation strategies respectively. The solid line denotes the optimal fraction of the budget allocated to contraception, and the shaded area denote the set of all equivalently optimal allocations; (b) The y-axis denotes the probability of satisfying the minimally-acceptable threshold.
Table 5.2: Stage-based population projection matrix for loggerhead sea turtles (*Caretta caretta*) (data from Crowder et al., 1994). The five stages are: hatchlings/juveniles, small juveniles, large juveniles, subadults, adults. The different stages were determined based on growth rates, age at first reproduction, and survivorship.

<table>
<thead>
<tr>
<th>time (t+1)</th>
<th>life stage at time (t)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.665</td>
<td>61.896</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.675</td>
<td>0.703</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.047</td>
<td>0.657</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.019</td>
<td>0.682</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.061</td>
<td>0.809</td>
<td></td>
</tr>
</tbody>
</table>

Denoting by \(s\) the underlying adult survival rate affected by inclusion of TEDs, and by \(h\) the hatchling survival rate affected by protection of nesting habitat, we parameterize the stage population matrix (Table 5.2) as follows

\[
\begin{pmatrix}
0 & f_2 & f_3 & f_4 & f_5 \\
0 & (1 - \gamma_2)q_2 & 0 & 0 & 0 \\
0 & \gamma_2q_2 & (1 - \gamma_3)q_3s & 0 & 0 \\
0 & 0 & \gamma_3q_3s & (1 - \gamma_4)q_4s & 0 \\
0 & 0 & 0 & \gamma_4q_4s & q_5s
\end{pmatrix}.
\]

The parameter \(q_i\) relates the stage-specific elements in the matrix (Table 2), to the underlying survivorship \(s\), and \(\gamma_i\) is the proportion of surviving individuals in stage \(i\) that advance to the next stage (\(\gamma_1 = 1\), and \(\gamma_5 = 0\)). For convenience, we set \(q_5 = 1\) and so \(h = 0.675\) and \(s = 0.8091\).

The dominant eigenvalue of the population matrix is \(\lambda = 0.9516\), and from (5.70) the sensitivities are \(\partial \lambda / \partial h = 0.0816\), and \(\partial \lambda / \partial s = 0.7719\). The estimated range of marginal costs for each conservation action, \(\partial C_h / \partial h\) and \(\partial C_s / \partial s\), are, between $3000 and $150000 for habitat protection, and between $48500 and $8240000 for inclusion of TEDs (see Supporting Information). Using these estimates in (5.72) we obtained expressions for the stochastic marginal efficiencies of each conservation action (detailed derivation in Supporting Information).

Two uncertainty models, namely normal and uniform, were considered, as per the koala example. Constructing a portfolio, we solved for the optimal allocation and the cor-
5.4 Results

The optimal allocation is independent of the choice of variance when uncertainties are represented by normal distributions (Figs. 5.2 and 5.4), provided that the ratio of the
Figure 5.5: Plot of optimal allocation for the sea turtle population management example (a) and associated probability of satisfying the minimally-acceptable threshold $\rho$ (b), for the case of uniformly distributed uncertainties. (a) The red and blue lines denote the minimum, and expected marginal efficiencies for the TEDs inclusion, and the minimum marginal efficiency for the habitat protection strategies, respectively. The solid line denotes the optimal fraction of the budget allocated to habitat protection, and the shaded area denotes the set of all equivalently optimal allocations; (b) The y-axis denotes the probability of satisfying the minimally-acceptable threshold.
standard deviations of the two strategies, \( \alpha \), remains constant in (5.28). In contrast, the probability of satisfying a given critical threshold is, indeed, sensitive to the choice of variance (Figs. 5.2 and 5.4). Furthermore, decreasing the standard deviation in this way suggests a limit process whereby allocations for threshold marginal efficiency levels below the lower of the two expected marginal efficiencies are satisfied with probability approaching unity.

Conversely, allocations for threshold values above the higher of the two expected marginal efficiencies are satisfied with probability approaching zero as the standard deviation decreases. This somewhat intuitive result states that as the expected marginal efficiency of each strategy becomes more certain, we would be increasingly certain of obtaining marginal efficiency at the expected value.

Next we note that for positive values of \( \rho \) that are below the expected marginal efficiency of all strategies, the optimal allocation is a diverse one. This is in agreement with the notion that spreading an investment over several uncertain assets mitigates risk, and traces back to the original motivation for the portfolio mean-variance theory [47].

Increasing the value of \( \rho \) beyond the expected marginal efficiency of either action, the optimal allocation transitions to allocating the entire budget to a single strategy, corresponding to \( x^* = 0 \) and \( x^* = 1 \) in (5.28). Although initially this might seem counterintuitive, such a decision could be understood through the analogy of the roulette chance game. For a relatively low requirement for returns, risk could be mitigated in a roulette game by betting on multiple (often negatively correlated) events. As our requirements of return increase above the expected outcome (in a single roulette game the expected outcome is a small proportional loss) our approach shifts towards increasingly placing all our bets on the high risk and high return events (for example, placing all bets on a particular number). Such a strategy maximizes the probability of achieving a big gain in a single game, even though the probability of realizing such an event is small.

When the variance of the action with the least expected marginal efficiency is sufficiently large in comparison to the variance of the competing action, the entire budget allocation can switch between actions as seen in (5.25) and (5.26). Indeed, such a transition occurs for the koala population management example. At a threshold level of around \( 4.25 \times 10^{-6} \), the optimal allocation transitions from allocating the entire budget to managing survival, to allocating the entire budget to managing fecundity (Fig. 5.2). Note that in this case, the probability of satisfying the objective is very small, across all the variance assumptions, and therefore, these allocations would most likely be ineffective. However, in general, such transitions could occur for optimal allocations where the associated probabilities are meaningful.

When the ratio of the standard deviations of the two strategies changes, the optimal allocation, and the associated probability, will both vary. In particular, increasing the
standard deviation of the survival management action will decrease the amount allocated to survival when diversifying allocation in both cases studied. In the case of the management of the koala population, such an increase will also delay the transition to allocating the entire budget to managing fecundity for high levels of the threshold. The change in the optimal allocation is given by the analytical expression (5.28).

When uncertainties are represented by the uniform distributions (Figs. 5.3 and 5.5), all allocations satisfy the objective with probability 1 for values of $\rho$ below the minimum marginal efficiency (that is, the lower bound of the uniform distribution) of both actions. Then, for $\rho$ values between the minimum marginal efficiency of the two actions, a range of optimal allocations satisfy the objective with probability 1. Both of these results are intuitive, if both options guarantee the favourable outcome, then any mixture of them will.

Diversification of allocation occurs under the uniform uncertainties model for $\rho$ values greater than the minimum marginal efficiency of both actions, and below the lesser expected marginal efficiency of either action. In both examples studied, this region is relatively small but could be larger in other cases. For larger values of $\rho$, it is optimal to invest the entire budget in the action offering higher expected marginal efficiency. Finally, for $\rho$ values beyond the greater maximum marginal efficiency of the two actions, all allocations are unsuccessful in satisfying the objective with positive probability.

Some differences in the choice of optimal allocation between the two uncertainty models stem from the fact that the normal distribution has unbounded support, while the uniform distribution has bounded support. That is, in the case of the normal distribution, one can have arbitrarily large and small marginal efficiencies with non-zero probabilities. As a result, under the normal assumption, there always exists a unique optimal allocation. On the other hand, under the uniform assumption, there are regions of $\rho$ values for which there exist multiple equally optimal allocations.

In order to address the assumption of linearly scaling marginal efficiency with respect to the total budget allocated, the decision variables must be changed from fraction allocated to actual budget spent. In accordance, the corresponding constraint would change to ensure the total allocated equals the budget available. This reformulation of the problem would also accommodate stochastic cost functions that are dependent on the actual dollars allocated. This is useful, for example, in the case when potential management actions have fixed initial costs. Such fixed costs can then be represented as constraints on the decision variables denoting dollars allocated to each action.
5.5 Testing robustness to the linearity assumption in the portfolio model

In constructing the portfolio, (5.1), we implicitly make the assumption that the marginal efficiency of the portfolio is a linear combination of the marginal efficiencies of each strategy. Furthermore, by only considering the fraction of the budget allocated to each strategy, and not the actual amount allocated, we make a second implicit assumption that the marginal efficiency scales linearly with respect to the total amount allocated. Both these assumptions are likely to be incorrect in general. The dominant eigenvalue is not in general a linear function of the underlying parameters. In fact, it is not even an everywhere analytic function of such parameters. For example, in the case of the management of the koala population on Snake Island, the dominant eigenvalue is given by the largest magnitude root of the characteristic equation

\[
\phi(\lambda) = f \left( \sum_{j=1}^{8} (-1)^j w_{j+1} \prod_{i=1}^{j} \gamma_i q_i s \prod_{k=j+2}^{9} ((1 - \gamma_k) q_k s - \lambda) \right) - \lambda \prod_{i=2}^{9} ((1 - \gamma_i) q_i s - \lambda) = 0.
\]

Equation (5.77) is a polynomial of degree nine in \( \lambda \). From Perron-Frobenius theory we know that the dominant eigenvalue is real and positive, but we cannot obtain an explicit expression for it. Instead we compute it numerically for each given combination of values for \( f \) and \( s \).

We therefore perform a sensitivity analysis to test robustness of our reported results. We now demonstrate such analysis for the case of management of koala population on Snake Island. In order to test the linearity assumptions we first express \( f \) and \( s \) in terms of the budget allocated to each strategy. Let \( B_f = xB \) and \( B_s = (1 - x)B \) be the amount of the total budget \( B \) allocated to the contraception and relocation strategy respectively. Then, assuming the budget is allocated pro rata across stages, the amount allocated to each stage for each strategy is given by

\[
B_{f,i} = xBN_i/N \quad \text{and} \quad B_{s,i} = (1 - x)BN_i/N.
\]

Adding "\( \sim \)" to any parameter to denote its value after management action, and using (A.1) and (A.3) and the definitions for \( f_i \) and \( s_i \), we have

\[
\tilde{f} = \frac{\tilde{f}_i}{w_i} = f \left( 1 - \frac{xB}{c_f N} \right), \tag{5.79}
\]

\[
\tilde{s} = \frac{\tilde{s}_i}{q_i} = s \left( 1 - \frac{(1 - x)B}{c_s N} \right). \tag{5.80}
\]

We now vary \( x \in [0, 1] \) for a given budget \( B \), and substitute the resulting values for \( \tilde{f} \) and \( \tilde{s} \) from (5.79) and (5.80) in the population matrix \( A \). For each such matrix we
then calculate numerically the dominant eigenvalue $\tilde{\lambda}$. Finally we compare the values for the expected marginal efficiency, $\mathbb{E}[R(x)]$ (identical under both the normal and uniform distributional assumption), with the observed marginal efficiency given by

$$R_{\text{obs}}(x) := \frac{\Delta \lambda}{B} = \frac{\tilde{\lambda} - \lambda}{B},$$

(5.81)

where $\lambda$ is just the original dominant eigenvalue for the population prior to any management action.

Results for a set of four budget values $B \in \{0.00001, 100, 1000, 10000\}$, and a fixed population size, $N = 100$, are plotted in Figure 5.6. First note that the numerical results are indeed approximately linear for budget values sufficiently low. In particular, from (5.79) and (5.80), the change in $f$ and $s$ are a function of $B/N$. From the numerical analysis, the linearity assumptions hold when $B$ and $N$ are of the same order of magnitude. For values of $B$ an order of magnitude larger than $N$ the numerical analysis is no longer linear however the linear approximation is still an average approximation to the overall trend. That is, the linear approximation follows the general trend in the non linear behavior. Beyond that, for values of $B$ more than an order of magnitude larger than $N$ the marginal efficiencies predicted by the linear model are a poor approximation of the actual dynamics of the system. Remembering that the per individual management costs are of the order of 100’s of dollars, this budget translates to treating up to 10% of the population.

### 5.6 The general elicited distributions case

In Section 5.2 we pointed out that obtaining the optimal allocation in (5.2) subject to $0 < x_i < 1$ $i = 1, \ldots, n$, and $\sum_{i=1}^n x_i = 1$, requires, in general, first obtaining an expression for the probability distribution of $R(x)$. When, for all possible $x$, given an $x$, the distributions of the weighted marginal efficiencies $\{x_i R_i\}$ are independent, the distribution of $R(x)$ is given by their convolution. However, obtaining an analytical close form expression for such a convolution, for the case of general marginal efficiency distributions, using for example Fourier transforms, or indeed directly, is not always tractable.

Alternatively, given an $x$, one can always obtain a random sample from $R(x)$, by sampling from each $\{x_i R_i\}$, and summing. To sample from a general distribution one can use inversion sampling, provided that the corresponding cumulative distribution function has a close form that is readily invertable [64]. When these conditions do not hold, other random sampling methods, such as acceptance-rejection sampling, or Markov chain Monte Carlo (MCMC) methods such as the Metropolis-Hastings sampling, Gibbs sampling or Wang-Landau sampling, may be used [27].
Figure 5.6: Plot of marginal efficiencies, 1a–4a, and dominant eigenvalue, 1b–4b, for four total budget values. Black lines are predictions from the model using linear assumptions and blue lines are results from direct numerical calculations (N=100).
In the numerical sampling analysis highlighted below we used the MCMC slice sampling algorithm method [59]. The method is based on the observation that to sample a random variable one can simply sample uniformly from the region under the graph of its density function. The slice sampling algorithm is implemented in Matlab using the \texttt{slicesample()} function.

Once a random sample from $R(x)$ is generated, the probability of obtaining a portfolio marginal efficiency above $\rho$ can be estimated using a Monte Carlo (MC) approach. In particular, the fraction of samples that lie above the minimally-acceptable marginal efficiency $\rho$ approaches the true probability in (5.2), as the sample size increases, by the law of large numbers. The optimal allocation that maximises the probability in (5.2), is chosen by varying $x$ discretely, and picking the $x$ value for which the MC estimated probability is maximised. Then, by repeating this procedure for a range of $\rho$ values, a decision maker is able to explore the trade-off between guarantying a higher minimal marginal efficiency and the probability of achieving it.

To illustrate how pseudo-random sampling and MC simulations can be used in the context of decision making under uncertainty, when applying the safety first method, we revisit the Koala case study from Section 5.3.1. We suppose that instead of assuming either normal or uniform distributions, the four-step elicitation protocol (Section 1.2) was used to elicit the probability density functions over the marginal efficiencies. We further suppose that the reference distribution used is the normal distribution. This is similar to Case 2 in Section 3.5. Using Corollary 3.2 and (3.19) we derive the analytical expressions for the probability density function for each of the marginal efficiencies corresponding to the two alternative environmental management actions. Figure 5.7 illustrates the resulting distributions assuming a particular set of elicited values.

Then, using the slice sampling method implemented in Matlab, and the MC simulation approach highlighted above, we perform repeated simulations on the range of feasible $\rho$ values to produce Figure 5.8. Comparing Figures 5.8 and 5.2, the allocation, as well as the associated probabilities, follow similar patterns. This is of course expected given that in both cases the normal distribution was used, in Figure 5.2 as the only uncertainty model, and in Figure 5.8 as the reference distribution. Some differences nevertheless exist, and are mainly due to the stronger threshold behaviour around limits of the elicited range noticeable in Figure 5.8, and depicted clearly also in Figure 5.7.

The random sampling and simulation approach described above can be applied to any general elicited distributions. This work is very much still in progress and was included here briefly for completeness. Further work is being carried on investigating both numerical and analytical approaches to applying the safety first portfolio method in the general elicited distributions case.
Figure 5.7: Illustrating the resulting unimodal MCE densities over the marginal efficiencies of each of the two potential management actions. The information elicited is assumed to be: *fecundity* range $[2.06975 \times 10^{-6}, 2.48371 \times 10^{-6}]$, mode at $2.40091 \times 10^{-6}$, and the confidence 75%; *survival* range $[2.26379 \times 10^{-6}, 2.64109 \times 10^{-6}]$, mode at $2.37698 \times 10^{-6}$, and the confidence 90%. These numbers were chosen to match the elicit cost range reported, and the assumptions made, in the analysis of the koala case study summarised in Figure 5.2. The prior normal distribution in both cases had mean centred at the respective mode, and variance of 1.
Figure 5.8: Plot of the resulting optimal allocation for the koala population management example (a), and associated probability of satisfying the minimally-acceptable threshold \( \rho \) (b), for marginal efficiencies distributions elicited using the 4-step procedure. (a) The blue and red shades denote the optimal proportion allocated to managing fecundity (contraceptive), and survival (relocation) respectively, for each level of the minimally-acceptable threshold \( \rho \). Dashed and dotted lines denote the expected marginal efficiencies for fecundity and survival management respectively (also in (b)); (b) The curve represents the probability of satisfying the corresponding minimally-acceptable threshold level \( \rho \).
5.7 Chapter summary

Uncertainty is common in environmental management decisions [e.g., 56] and must be addressed when making decisions. Failure to do so runs the risk of over-confidence in the chosen strategy, and exposes the manager to higher rates of failure [14]. Using the case studies for the management of the koala population and the conservation of the olive ridley sea turtle population, we demonstrate a method to account for uncertainty in population management decisions.

Combining perturbation analysis of matrix population models and a safety first portfolio formulation, our results illustrate how a conservation manager is able to explore the entire decision map of trade-offs between the expected effect on the population per dollar spent, and the probability of achieving it.

Baxter et al. [1] noted that their results were highly sensitive to the choice of the expected cost point estimates. The formulation presented in this chapter addresses such sensitivities in two steps. First, portfolios of the actions define the probability of achieving a minimally-acceptable efficiency, determined by the assumptions about the uncertainty. Second, an optimisation problem is solved to find the optimal allocation that maximises the probability of obtaining an outcome above the minimally-acceptable value.

From our analyses, diversification of allocation is optimal only for minimally-acceptable outcomes below the expected marginal efficiencies of both strategies. Once the level of the threshold of acceptability is above the lower of the two expected marginal efficiencies, it is optimal to allocate the entire budget to only one of the strategies. In other words, for low aspirations, bet hedging by diversifying the allocation is the optimal solution. Conversely, for high aspirations, effort is directed solely at management actions that provide the highest potential effect on the population. These results agree with common practice and previous studies [for example 52].

The method we propose necessitates making two important choices. First, they must choose an appropriate uncertainty model. In this chapter, as an illustration, we used the normal and uniform uncertainty models for analytical investigation. We also demonstrated a pseudo-random number sampling approach, using the MCMC algorithm slice sampling technique, for the case of general elicited distributions. Our results show that, while some overarching patterns in the optimal allocation exist across models, the optimal allocation is, indeed, sensitive to the choice of the uncertainty model. This fact highlights the importance of characterising any uncertainty present in the decision process. Characterisation of uncertainty could be done in one of several ways. In some cases, the uncertainty model can be determined from physical considerations or first principles, alone. In other cases, the uncertainty model can be estimated from historical data. Expert elicitation used in the work presented in this chapter, is another common method for estimating uncertainty.
models as was discussed in Chapters 3. Whatever uncertainty model is used, it must be reassessed and updated accordingly once new information becomes available.

The second choice to be considered relates to the value of the minimally-acceptable efficiency. The minimally-acceptable efficiency is chosen by considering either the level of confidence (equivalently, the level of risk) the decision maker is willing to accept, or the minimum level of efficiency required to satisfy a management goal. The choice is made either a priori, or a posteriori. A priori minimally-acceptable efficiency level is determined from either physical consideration, previous data, or set rules and regulations. When an a priori choice is not feasible, the minimally-acceptable efficiency level must be determined a posteriori by examining decision maps such as in Figs. 1-5, and studying the tradeoff between aspiring for higher marginal efficiencies and the probability of achieving them.

The safety first portfolio formulation presented here combines ideas of optimal expected utility [67], robust optimisation [3] and satisfying objective functions [72]. Modified slightly, the safety first portfolio formulation can also represent other robust decision objectives. For example, constraining the probability of satisfying the minimally-acceptable efficiency to unity, and maximising the minimally-acceptable efficiency for uniform uncertainty model is equivalent to minimising worst-case strategy [26]. Furthermore, one can extend the current formulation by considering instead of a single uncertainty model a family of nested models.
Chapter 6

Conclusion

A wide range of applied problems can be reduced to an under determined inverse problem of the form, find ‘best’ $f$ subject to $Af = b$, where $A$ some linear (or nonlinear) operator, $b \in \mathbb{R}^n$ and $f$ an element of an appropriate functional space [9]. This general class of problems finds applications in many fields including: acoustics, actuarial science, astronomy, biochemistry, constrained spline fitting, engineering, finance, image processing, inverse scattering, optics, option pricing, and statistical moment fitting.

To solve for the optimal density one of a range of information-theoretic, or norm related, measures is defined over the feasible set. The resulting, often convex, nonlinear constrained problem, has been shown to exhibit an unconstrained dual, with dual attainment. A particular difficulty in applying under determined inverse problem theory arises when information provided cannot be represented in the form of a moment constraint. Such is the case when the density is known to be unimodal, and its unique mode is given.

Expert elicitation is a common practice in a range of applications where data regarding a quantity of interest is unavailable or uncertain. Expert elicitation has been drawing increasing attention in recent years from both government, industry and also academia. Attempts to introduce more rigour to the process of elicitation resulted in a multitude of papers on the various statistical, psychological, computational and otherwise aspects of elicitation. The problem of choosing the best distribution to represent the data elicited from an expert, understood in a Bayesian sense, is in fact an under determined inverse problem.

In this thesis we studied the problem of unimodal density estimation when using the MCE method. We posed the problem in the context of expert elicitation using the 4-step protocol for illustration. We then examined the general problem of decision making under conditions of uncertainty and the use of expert elicitation to derive prior distributions
over the space of possible states of nature.

6.1 Thesis summary

In Chapter 3 we began by introducing the problem of density estimation using the MCE method, posed in the context of expert elicitation. In particular, we first examined the case when data elicited can be represented as an expectation over some real valued function. The resulting nonlinear problem was trivially solved using the theory of Lagrange multipliers. The optimal density has a well known exponential form and was previously derived in both information theory and statistical mechanics.

We then turned our attention to the problem of incorporating a unimodality constraint in MCE formulation. We revisited work by Brockett et al. [13] that used several previous results by Khinchin, Shapp, and Kempermann to derive an auxiliary MCE problem. We demonstrated that solving the auxiliary problem suggested by Brockett et al. [13] leads to feasible, but not optimal, solutions, with respect to the original constraints and CE objective. Using again Khinchin’s criterion, we derived an alternative auxiliary problem and demonstrated that it does indeed lead to optimal solutions. Instead of solving the new problem we chose, in this Chapter, to solve a relaxed form, with respect to the unimodal constraint, of the original problem. We then provided necessary and sufficient conditions under which solutions of the relaxed problem results in unimodal densities.

The analytical results were illustrated using the 4-step elicitation protocol. This protocol elicits information corresponding to the mode and an inter-percentile of the distribution representing the expert’s opinion. In some applications of the 4-step protocol, it is favourable to avoid discontinuities in the resulting density. We therefore offered an alternative moment function, that ensures differentiability of the optimal density.

In Chapter 4 we began by deriving the general problem of unimodal density estimation using the MCE method. Solving the general problem required first establishing a series of results. First, we established the equivalence of the unimodality requirement of the original problem, and the nonnegative inequality constraint of the auxiliary problem. Using this result we established the equivalence of the original and auxiliary problems.

Establishing next the convexity of the auxiliary problem, we derived, using the Gâteaux differential, the first order necessary conditions. Given the established convexity properties of the objective and the linearity of the constraints, the first order necessary conditions are also sufficient. Using the differential of the Lagrangian function, and by complimentary slackness, we decoupled the primal and dual problems. Finally, we obtained an expression for the optimal density, and the corresponding constrained dual problem. The expression for the optimal density consisted of two parts, the first being the relaxation solution from
Chapter 3. The second, is a constant value defined over elements in the support for which the auxiliary nonnegative constraint is active.

Next we illustrated the analytical results using an example in the context of expert elicitation. We concluded the chapter with a brief discussion on the numerical methods we employed when solving the dual problem to obtain the values of the multipliers. In particular, when the auxiliary non-negative constraint is inactive for all elements in the support, the dual is unconstrained. In such cases we used an implementation of the trust region method in Matlab. When, instead, the auxiliary inequality constraint is active for some elements in the support, the dual is constrained. To solve in this case, we reduced the constrained dual to an unconstrained one, and added a check for unimodality. Our algorithm then proceeds by extending, at each iteration, the set containing the elements in the support of the auxiliary density, where the nonnegative constraint is active, followed by solving the new dual. Given that when the original problem is feasible, the dual problem has an optimal solution, the algorithm will converge to the optimal solutions with any prescribed level of accuracy.

In Chapter 5 we moved to examine the problem of decision making under conditions of uncertainty. We proposed a method that combines the previously suggested safety first and portfolio approaches, with a satisficing objective, and uses subjective probabilities to infer a distribution over the possible states of nature (parametric state space). Heuristically, the objective we advocate for is to choose the mixture of actions that maximises the probability of obtaining a favourable outcome. This objective is equivalent to maximising a utility defined as the indicator function over the favourable outcome. As an example, the resulting nonlinear optimisation problem is solved under two subjective probability assumptions, namely, the normal and uniform distributions.

The general framework is then applied in the context of environmental management. We revisited a previous study that applied cost benefit analysis to perturbation analysis for population management. We then embedded their work in our framework. The result offers a method for incorporating uncertainty for decision making in the context of population management. We illustrated our approach using two case studies, the management of a koala population, and the conservation of a marine turtle population.

6.2 Further work and future directions

While complete solutions to the problem of unimodal density estimation using MCE method were obtained in this thesis, many further generalisations of these results need to be further investigated. For example, [9] has studied the following, relatively general,
problem
\[
\inf_f \int_{\mathcal{X}} \phi(f(x)) \mu(dx)
\]
subject to \(Af = b, \quad f \geq 0, \quad f \in L^p(\mathcal{X}),\)

where \((\mathcal{X}, \mu)\) is a general \(\sigma\)-finite measure space, \(1 \leq p \leq \infty\), \(\phi : \mathbb{R} \to (-\infty, \infty]\) is convex, and \(A : L^p \to \mathbb{R}^n\) is continuous. Solving (6.1) [9] established that the dual problem is finite dimensional and unconstrained, and for a special class of conditions, there is no duality gap. Now, (6.1) shares many similarities with the auxiliary problem (4.3) in our work. However, in our case, the resulting dual problem was, in fact, constrained. Further work is needed to investigate this discrepancy.

Another possible generalisation of our results from Chapter 4 entails investigating their validity under different information theoretic objectives, such as Fisher’s information and Burg entropy. Given that such measures share the convexity property of the cross entropy measure, it is a reasonable assumption to expect the current general results to hold, with some modifications.

The key result that facilitated the incorporation of the unimodal requirement in the MCE method was the representation of a unimodal random variable given by Khinchin’s criterion. Another extension of our work on density estimation could be to other classes of information that cannot be expressed as an expectation over some real valued function. One such example is continuity. Perhaps there are other possible representations of such constraints that would result in transformations converting these constraints to equality or inequality constraints, in some corresponding auxiliary problems.

Using the MCE method in the context of expert elicitation offers many advantages. It is capable of incorporating many different classes of potential elicited information, and, using the work in this thesis, can now also address unimodal information. However, far more work is required in order to facilitate the application of these ideas in real life elicitation procedures.

Some steps in this direction have been taken. Together with colleagues from the Australian Centre of Research for Risk Analysis, we have developed a simple implementation of the relaxation solutions in Excel, documented in a technical paper [66]. Following that, a workshop with expert elicitation practitioners was conducted. The initial feedback was positive, and most attendants found the implementation easy to use and instructive. In the concluding remarks attendants suggested several improvements in particular with regards to the user interface. Future iterations of the implementation are scheduled to address these improvements, as well as incorporate the general results obtained in Chapter 4.
Other, both parametric and non-parametric, approaches to density estimation in the context of expert elicitation, were previously proposed. Future studies need to compare and contrast these methods to learn of the strengths and weaknesses of each approach and devise guidelines for their use.

In Chapter 5 we introduced the safety first portfolio approach to decision making under conditions of uncertainty. The approach borrows from previously suggested ideas in economic theory and Bayesian statistics. In our approach we advocated the use of subjective probabilities. One of the key methods for obtaining informative subjective probabilities is the use of expert elicitation techniques. Uninformative subjective probabilities, on the other hand, may be obtained using the principle of invariance under transformation groups [35].

The specific application of our ideas in the context of perturbation analysis for population management concentrates on uncertainty in cost estimates, long-term asymptotic dynamics and point decisions. Possible future directions include: explicit uncertainty in population matrix parameters, short-term transient dynamics [16], and consecutive decisions. While each of these directions have already been studied, combining them into a coherent framework for informing decision makers, in particular in the context of environmental management, was never done, and is of utmost importance.

Finally, while the approach we advocated in Chapter 5 is applicable under some cases, it is by no means the answer to all decision problems under uncertainty. There exists a large body of literature attempting to grapple with these problems. A vital direction for future research in this area, would be to combine the plethora of ideas and scopes into one coherent overarching theory. Some promising steps in this direction are afforded by insights into links between the principle of maximum entropy and games against nature [35, 77, 78].
Bibliography


6.2


Appendix A

Appendix

A.1 Derivation of marginal cost expressions

Management of Koala population case study

This work was published in Baxter et al. (2006) and is presented here again only for completeness. Let $N_i$ be the number of individuals in stage $i = 2, \ldots, 9$, and $N$ be the total number of individuals in the adult population, stages 2–9. Also let $B$, be the total available budget. We assume that both management actions target any individual in stages 2–9 with equal probability. Then, for fecundity management, if darting one individual costs $c_f$ dollars, investing the entire budget in fecundity management will result in an allocation of $B_i = BN_i/N$ dollars to stage $i$, treating $B_i/c_f$ individuals. Defining the per individual of stage $i$ fecundity as $f_i := w_i f$, the number of young born to stage $i$ after fecundity management is therefore $N_i \tilde{f}_i = w_i f (N_i - B_i/c_f)$. Then, the reduction in fecundity rate of individuals in stage $i$ is given by

$$\Delta f_i = \tilde{f}_i - f_i = w_i f B_i/(c_f N_i).$$  \hfill (A.1)

The marginal cost of contraceptive in stage $i$ is thus $\partial C_{f_i}/\partial f_i \approx B_i/\Delta f_i = c_f N_i/w_i f$, and the marginal cost of contraceptive in all classes is

$$\frac{\partial C_f}{\partial f} = \sum_{i=2}^{9} \frac{\partial C_f}{\partial f_i} \frac{\partial f_i}{\partial f} = \sum_{i=2}^{9} \frac{c_f N_i}{w_i f} p_i = \frac{c_f N}{f},$$  \hfill (A.2)

where we used the fact that $C_f = \sum_i C_{f_i}$.

Following similar logic for translocation management, if removing one individual costs $c_s$ dollars, a translocation budget of $B$ dollars acting pro rata among stages will remove
\( B_i/c_s \) individuals in stage \( i \). The respective changes in number of individuals of stage \( i \) advancing to next stage, and those remaining in the same stage are \( \gamma_i q_i s (N_i - B_i/c_s) \), and \((1 - \gamma_i) q_i s (N_i - B_i/c_s)\). Thus the reduction in the per individual survival rate in stage \( i \), \( s_i = q_i s_i \), is

\[
\Delta s_i = \bar{s}_i - s_i = \gamma_i q_i s B_i/c_s N_i + (1 - \gamma_i) q_i s B_i/c_s N_i = q_i s B_i/c_s N_i = q_i s B_i/c_s N_i
\]  

(A.3)

The marginal cost of translocation in stage \( i \) is therefore \( \partial C_{s_i}/\partial s_i \approx B_i/\Delta s_i = c_s N_i/q_i s \), and the marginal cost of translocation in all stages is

\[
\frac{\partial C_s}{\partial s} = \sum_{i=2}^{9} \frac{\partial C_{s_i}}{\partial s_i} \frac{\partial s_i}{\partial s} = \sum_{i=2}^{9} \frac{c_s N_i}{q_i s} = \frac{c_s N_2}{s}.
\]  

(A.4)

(A.2) and (A.4) are then used in (5.73) and (5.74) to obtain an expression for the marginal efficiency of each management action.

**Conservation of sea turtle population case study**

We obtained the following conservation costs (in USD a year) and population data for a olive ridley sea turtle population in Orissa, India (K. Shanker, personal communication, July, 2010). Nesting habitat protection cost was estimated to be between $1000 and $10000. For these costs the estimated increase in survival of hatchlings was 10% – 50%. Nesting habitat protection included protection of nests and eggs, protection of hatchlings, and elimination of disturbing artificial lighting.

Inclusion of TEDs in the relevant fisheries was estimated to cost between $100000 and $1000000. For these costs the estimated increase in survival of nesting adults was estimated %10 – %25. We also obtained nesting population size, and yearly number of adults stranded in fishing nets, estimates. Using these two estimates we confirmed the increase in adult survival estimate of %10 – %25.

Next, we assumed that the marginal cost of either action is constant, and we let \( h_0 \) and \( s_0 \) denote the initial underlying survival parameter that is effected by conservation action for hatchlings and adults respectively. The minimum and maximum marginal cost estimates were therefore

\[
\begin{align*}
\min \frac{\partial C_h}{\partial h} &= 1000 \quad &\max \frac{\partial C_h}{\partial h} &= 10000 \\
\min \frac{\partial C_s}{\partial s} &= 10^5 \quad &\max \frac{\partial C_s}{\partial s} &= 10^6
\end{align*}
\]

(A.5)  

(A.6)

The marginal costs were then combined with sensitivities to change in the respective conservation action obtained from Eq. (5.70), together with values for \( h_0 \) and \( s_0 \) from the
population matrix (Table 2), to produce expressions for the stochastic marginal efficiency for each conservation action.