Comparative analysis of expected utility theory versus prospect theory and critique of their recent developments

Sassan Sadeghi

Comparative Analysis of Expected Utility Theory versus Prospect Theory and Critique of their Recent Developments

---

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Sassan Sadeghi

June 2011

COMPARATIVE ANALYSIS OF EXPECTED UTILITY THEORY VERSUS PROSPECT THEORY AND CRITIQUE OF THEIR RECENT DEVELOPMENTS

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Sassan Sadeghi

June 2011

Approved by:

[redacted]

Dr. Joseph Chavez, Committee Chair

6/9/11

Date

[redacted]

Dr. Zahid Hasan, Committee Member

[redacted]

Dr. Chetan Prakash, Committee Member

[redacted]

Dr. Peter Williams, Chair,
Department of Mathematics

[redacted]

Dr. Charles Stanton
Graduate Coordinator,
Department of Mathematics

Abstract

This paper is an investigation of the decision making theories, their developments, and especially, their applications. After locating the two rivals, the Expected Utility Theory (EUT) and the Prospect Theory (PT), within the general context of decision making situations, it compares their main features and examines the PT extensions. EUT and PT as descriptive models are then shown coexisting within the same grand likelihood function, thus providing a detailed example for Finite Mixture Models. Other illustrative application examples in the fields of astronomy and medical diagnosis expose some of the technical difficulties when constructing maximum likelihood functions.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank Dr. Chavez, for all his support and encouragement throughout this project. More so, CSUSB has been the first American university that I have attended, again thanks to Dr. Chavez who received, encouraged and guided me from the very beginning of my studies. I would also like to thank the rest of my master's committee, Dr. Hasan and Dr. Prakash, for agreeing to be a part of this final stage of my masters degree. And I need to thank all my professors who taught me mathematics so patiently. Finally, I would like to thank my family and friends for their encouragement and understanding. Especially, I am grateful to my wife, who had to disappointedly watch me cry like a baby over Latex.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

There is a lot written in the literature comparing Expected Utility Theory and Prospect Theory. For one, EU theory is very old, dating back to Bernouilli himself, and Prospect Theory grew out of its shortcomings.

The second chapter of this study reviews the important basic concepts such as decision making under certainty, decision making under risk, utility functions, decision under uncertainty and a few others that are used in this document. It will also review the desirable properties any theory of human decision might want to have.

The third chapter presents examples of utility theory and details its specific failures. This section will show, for instance, that Expected Utility Theory being true for any utility function is equivalent to a set of axioms: the Von-Neumann Morgenstern axioms. In fact, Prospect Theory initially noted that some of these axioms are not empirically true: people just don't make decisions like that. The relation between the curvature of the utility function and risk aversion is analyzed at this point. This section then shows the remedies suggested by the Prospect Theory before assessing this same theory's limitations and extensions. For instance, a short section will show that important properties of the Prospect Theory such as first order stochastic dominance, and transitivity come into play only when there are more than 2 outcomes. Additionally, Cumulative Prospect Theory, which is the most theoretically sound formulation of PT because it respects stochastic dominance and transitivity, is mostly useful when there is a continuum of outcomes.

The fourth chapter analyzes the more recent models developed by Glenn W. Harrison, E. Elisabet Rutstrom in 2008 [GH08]. These authors suggest that a finite mixture model can be used to estimate the parameters of each decision process while simultaneously estimating the probability that each theory applies to the sample. This section compares EUT versus PT and the degree to which they satisfy the properties of Von-Neumann Morgenstern properties, stochastic dominance, transitivity, certainty, reflection and isolation effects. In addition to these properties, we will look at scale independence of these models: do the behavioral predictions of these models depend on the denomination/units of the rewards/losses. It could be that people's decisions are influenced by the denominations or units in which rewards are counted; however this might seem unlikely. What are the conditions on the families of utility functions and EUT and PT frameworks for the decisions predicted to be unchanged by rescaling the rewards/losses, once the model parameters have been inferred?

The fifth and last chapter will review some real world applications of the finite mixture models developed by Glenn W. Harrison, E. Elisabet Rutstrom. One of the main objectives of this study has been, all along, to construct the background concepts and understanding necessary for a full appreciation of the application examples presented in this chapter. This is done through four application cases. The first two cases are practical overviews with the intension of introducing how the finite mixture models are used for identifying subpopulations, and fitting them into their respective appropriate mathematical models. The two cases presented are in the area of blood cell sampling for diagnosis purposes, and in the field of astronomy investigation. Then, the more technically involved section of this chapter starts with another application example in the field of medical diagnosis. This application example exposes the technical aspects and difficulties involved in the mathematical construction of the finite mixture models. It lays out the mathematical heart of the finite mixture modeling. As such, through this application example we see how the engine of the finite mixture model is actually put together. Finally, the fourth application example is in the field of consumer spending. This application example addresses additional technical difficulties involving cases where we don't even know in advance how many subpopulations exist out there to be modeled. Building on the previous application example in the medical diagnosis field, and by comparison to it, this consumer spending modeling application also illustrates other important components of the finite mixture modeling process such as the procedure of "Expectation Maximization", commonly referred to as EM in the specialized literature

Chapter 2

Decision Making Environments

This chapter reviews the important basic concepts such as decision making under certainty, decision under uncertainty, decision making under risk, and utility functions, used in this document. It will also review the desirable properties any theory of human decision might want to have.

A general reference for this chapter will be [RT75], and [FH05].

Decision problems usually involve a finite number of alternatives. The tools used to solve these problems depend basically on the type of context and their specific data. The main families of context are:

1. Deterministic, or under certainty, involving mainly rational thinking,
2. Uncertain, involving personal attitudes and feelings towards uncertainty,
3. Probabilistic, or under risk, involving theoretical modeling.

These three broad families of decision making situations have their own corresponding decision making tools. Yet, in the real world, any combination of these three situations can be present and interacting among them. The next few pages of this document therefore introduces the decision making tools for each of these three situations.

2.1 Analytic Hierarchy Process, AHP.

The analytic hierarchy process involves rational thinking, and is used in deterministic situations where decision making takes place under certainty. The process of rational thinking needs to be constructed. Such construction converts the context of subjective judgment into its quantified translation, and therefore, made ready for rational decision making. The first step is usually done through determination of weights in a logical manner [FH05].

2.2 Determination of the Weights.

The objective of this step is to determine relative weights in order to obtain a ranking among decision alternatives. For instance, dealing with n criteria arranged under a given hierarchy, the procedure constructs an $n \times n$ *pairwise* comparison matrix, A. This matrix quantifies the decision maker's judgment regarding the relative importance of the different criteria. This pairwise comparison ranks each criterion in row i ($i = 1, 2, ..., n$) relative to every other criterion as follows: Letting a_{ij} define the element (i, j) of A, AHP uses a discrete scale from 1 to 9 in which $a_{ij} = 1$ means that i and j are of *equal importance*, $a_{ij} = 5$ indicates that i is *strongly more important* than j, and $a_{ij} = 9$ indicates that i is *extremely more important* than j. Other intermediate values between 1 and 9 are interpreted correspondingly. Obviously, $a_{ij} = k$ would automatically imply that $a_{ji} = 1/k$. Also, all the diagonal elements a_{ii} of A must equal 1, because they rank a criterion against itself.

Example: Let us imagine a context where Location and Reputation of a town play an important role in choosing it for an investment decision. Further, suppose that we have a ranking analysis comparing three candidate towns of X, Y and Z as follows:

Reputation is strongly more important than the location, hence $a_{12} = 5$. And the corresponding comparison matrix is:

$$A = \begin{array}{c} \\ L \\ R \end{array} \!\!\! \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} 1 & \frac{1}{5} \\ 5 & 1 \end{pmatrix} \end{array}.$$

Normalizing A would give us the relative weights of R and L into a new matrix N. This can be obtained by dividing the elements of each column by the sum of the elements of the same column. Thus, to compute N, we divide the elements of columns 1 by $(5 + 1 = 6)$ and those of column 2 by $(1 + \frac{1}{5} = 1.2)$.

$$N = \begin{array}{c c} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{pmatrix} .17 & .17 \\ .83 & .83 \end{pmatrix} \end{array}.$$

The relative weights , w_R and w_L, are then computed as the row average: $w_L = \frac{.17+.17}{2} = .17$ $w_R = \frac{.83+.83}{2} = .83$

The columns of N are identical, since the decision maker exhibits perfect consistency in specifying the entries of the comparison matrix A. These relative weights with respect to Location versus Reputation will be used for the final Decision Mating calculation. At this point, we need to compare the relative importance of Location among the three towns, as well as the relative importance of Reputation among them. Matrices A_L and A_R compare these two sets of relative importance.

Within each of the two criteria Location and Reputation, the relative importance among the three towns, are given by the decision makers best judgment as follows: The relative importance of location for the towns X, Y, and Z are respectively 1,2 and 5. For their Reputation, these relative numbers are respectively 1, 1/2, and 1/3.

Expressed into A_L and A_R we obtain:

$$A_L = \begin{array}{c c} & \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{array}{c} X \\ Y \\ Z \end{array} & \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 2 & 1 & \frac{1}{2} \\ 5 & 2 & 1 \end{pmatrix} \end{array}$$

$$A_R = \begin{array}{c c} & \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{array}{c} X \\ Y \\ Z \end{array} & \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{2} & 1 & \frac{3}{2} \\ \frac{1}{3} & \frac{2}{3} & 1 \end{pmatrix} \end{array}.$$

Adding the columns, we get: A_L-column sum $= (8,\ 3.5,\ 1.7)$ A_R-column sum $= (1.83,\ 3.67,\ 5.5)$ Then normalizing through division of all the entries by the respective column-sums, we obtain:

$$N_L = \begin{array}{c c} & \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{array}{c} X \\ Y \\ Z \end{array} & \begin{pmatrix} .125 & .143 & .118 \\ .250 & .286 & .294 \\ .625 & .571 & .588 \end{pmatrix} \end{array}$$

$$N_R = \begin{array}{c c} & \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{array}{c} X \\ Y \\ Z \end{array} & \begin{pmatrix} .545 & .545 & .545 \\ .273 & .273 & .273 \\ .182 & .182 & .182 \end{pmatrix} \end{array}.$$

Calculating the corresponding row averages, we get:

$$\begin{array}{ll} w_{LA} = \frac{.125+.143+.118}{3} = .129 & w_{RA} = \frac{.545+.545+.545}{3} = .545 \\ w_{LB} = \frac{.250+.286+.294}{3} = .277 & w_{RB} = \frac{.273+.273+.273}{3} = .273 \\ w_{LC} = \frac{.625+.571+.588}{3} = .594 & w_{RC} = \frac{.182+.182+.182}{3} = .182 \end{array}$$

Thus, the Location weights, and Reputation weights for the towns X, Y, and Z can be summarized as follows:

Criterion	Town X	Town Y	Town Z	Sum
Location	.129	.277	.594	1
Reputation	.545	.273	.182	1

And finally, using the normalized relative weights W_L= .17 and W_R= .83 calculated at the beginning, the final ranking among the three candidate towns can now be obtained by:

$$\begin{array}{lllllllllll} \text{Rank} & = & \text{Weight} & \times & \text{Location} & + & \text{Weight} & \times & \text{Reputation} & & \\ X & = & .17 & \times & .129 & + & .83 & \times & .545 & & = .4743 \\ Y & = & .17 & \times & .277 & + & .83 & \times & .273 & & = .2737 \\ Z & = & .17 & \times & .594 & + & .83 & \times & .182 & & = .2520 \end{array}$$

Town X wins the investment decision.

2.3 Decision under Uncertainty

Under uncertainty, the payoffs depend on the random states of nature [FH05]. The payoff matrix of a decision problem with m alternative actions and n states of nature can be represented as:

$$\begin{array}{c c}
 & \begin{array}{ccc} s_1 & s_2 & s_n \end{array} \\
\begin{array}{c} a_1 \\ a_2 \\ \\ a_m \end{array} & \begin{pmatrix} v(a_1, s_1) & v(a_1, s_2)\ldots & v(a_1, s_n) \\ v(a_2, s_1) & v(a_2, s_2)\ldots & v(a_2, s_n) \\ \vdots & \vdots & \vdots \\ v(a_m, s_1) & v(a_m, s_2)\ldots & v(a_m, s_n) \end{pmatrix},
\end{array}$$

where the elements a_i stand for action i, and the element s_j for state of nature j. The payoff associated with action a_i and state s_j is $v(a_i, s_j)$. These payoff value elements $v(a_i, s_j)$ are known in advance for every state of the nature.

Making a decision under risk involves probability, whereas in the case of uncertainty, the probability distribution associated with the states $s_j, j = 1, 2, ..., n$, is either unknown or cannot be determined.

This lack of information has led to the development of the following alternative strategies for analyzing the decision problem [RT75]:

1. Laplace,
2. Minimax,
3. Savage regret,
4. Hurwicz.

These strategies differ in how conservative the decision maker is when facing uncertainty.

The Laplace criterion uses the principle of insufficient reason. Since the probability distribution for the states of nature is not known, the alternatives are simply evaluated using the optimistic assumption that all states are equally likely to occur, meaning: $P\{s_1\} = P\{s_2\} = ... = P\{s_n\} = \frac{1}{n}$. Given $v(a_i, s_j) > 0$, the best alternative is the one that yields

$$\max_i \left\{ \frac{1}{n} \sum_{j=1}^{n} v\left(a_i, s_j\right) \right\}.$$

If $v(a_i, s_j) < 0$, then minimization replaces maximization.

The minimax (maximin) criterion is a conservative attitude of making the best of the worst possible conditions. Therefore the decision maker uses:

$$\begin{cases} \min_i \left[\max_j v\left(a_i, s_j\right) \right], \text{ when } v\left(a_i, s_j\right) < 0, \\ \max_i \left[\min_j v\left(a_i, s_j\right) \right], \text{ when } v\left(a_i, s_j\right) > 0. \end{cases}$$

The Savage regret criterion tries to moderate the conservatism in the minimax (maximin) criterion by replacing the (gain or loss) payoff matrix $v\left(a_i, s_j\right)$ with a loss (or regret) $r\left(a_i, s_j\right)$ matrix, using the following transformation:

$$r\left(a_i, s_j\right) = \begin{cases} v\left(a_i, s_j\right) - \min_k \left\{ v\left(a_k, s_j\right) \right\}, \text{ if } v \text{ is loss} \\ \max_k \left\{ v\left(a_k, s_j\right) \right\} - v\left(a_i, s_j\right), \text{ if } v \text{ is gain} \end{cases}$$

For instance, consider the minimax criterion in the following loss matrix, where the unsigned elements of the matrix represent losses. Without the regret matrix, in this case, one would select a_2, but when applied to the regret matrix, a_1 will be selected instead.

In fact, in the case of minimax we have:

$$\begin{array}{c c} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} a_1 \\ a_2 \end{array} & \begin{pmatrix} \$11,000 & \$90 \\ \$10,000 & \$10,000 \end{pmatrix} \end{array}.$$

But, converting to its corresponding regret matrix, we get:

$$\begin{array}{c c} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} a_1 \\ a_2 \end{array} & \begin{pmatrix} \$1,000 & \$0 \\ \$0 & \$9,910 \end{pmatrix} \end{array}.$$

The Hurwicz criterion is designed to let a parameter α fit the decision-making attitudes into ranges going from the most optimistic to the most pessimistic (or conservative). Assuming $v\left(a_i, s_j\right)$ represents gain, define $0 \leq \alpha \leq 1$. Then the selected action must be associated with

$$\max_i \left\{ \alpha \max_j v\left(a_i, s_j\right) + (1-\alpha) \min_j v\left(a_i, s_j\right) \right\}.$$

The parameter α is called the index of optimism. When $\alpha = 0$, the criterion is conservative because it applies the regular minimax criterion. If $\alpha = 1$, the criterion produces optimistic results because it seeks the best of the best conditions. The proper selection of the value of α would therefore indicate the degree or the index of optimism desired. In the absence of strong feeling regarding optimism and pessimism, $\alpha = .5$ may be an appropriate choice. If $v\left(a_i, s_j\right)$ represents loss, then the criterion is changed to

$$\min_i \left\{ \alpha \min_j v\left(a_i, s_j\right) + (1-\alpha) \max_j v\left(a_i, s_j\right) \right\}.$$

Finally, another variation of decision making environment is situations in which two intelligent opponents with conflicting objectives are trying to outdo one another. Game theory best models these situations where aiming at the best out of the worst conditions available is the main concern. Typical examples include advertising campaigns for competing products and planning strategies for warring armies.

The so called players of the game will each have a finite or infinite number of alternatives or strategies. Such games are known as two-person zero-sum games because a gain by one player signifies an equal loss to the other.

A payoff is associated with each pair of strategies, which is what one player receives from the other. It suffices, then, to summarize the game in terms of the payoff to one player. Designating the two players as A and B with m and n strategies, respectively, the game is usually represented by the payoff matrix to player A as:

$$\begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_m \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 \dots & B_n \end{array} \\ \begin{pmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} \dots & a_{mn} \end{pmatrix} \end{array}.$$

This representation indicates that if A uses strategy i and B uses strategy j, the payoff to A is a_{ij}, which means that the payoff to B is $-a_{ij}$.

2.4 Probabilistic, or Decision Making under Risk Situations.

Decision making under risk is based on the expected value criterion because in this situation each decision alternative is described by a probability distribution. Maximizing the expected profit or minimizing the expected loss can be achieved here through the formation of decision trees, themselves subjected to utility functions.

Under conditions of risk, the payoffs associated with each decision alternative are described by probability distributions. For this reason, decision making under risk can be based on the expected value criterion, in which decision alternatives are compared based on the maximization of expected profit or the minimization of expected cost.

The probabilities used in the expected value criterion are usually determined from historical data. These probabilities may need to be adjusted using additional sampling or experimentation. The resulting probabilities are referred to as posterior (or Bayes) probabilities, as opposed to the prior probabilities determined from raw data. In practice, there usually are cases where the utility rather than the real value should be used in the analysis. This is done through utility functions.

2.4.1 Utility Functions.

The determination of the utility is subjective [Cen06]. Suppose there is a 50-50 chance that a \$20,000 investment will produce a net profit of \$40,000 or be lost completely. The associated expected profit is $40{,}000 \times .5 - 20{,}000 \times .5 = \$10{,}000$. Although there is a net expected profit of \$10,000, an investor who is willing to accept risk may undertake the investment for a 50% chance to make a \$40,000 profit. Conversely, a conservative investor may not be willing to risk losing \$20,000. Thus, we say that different individuals exhibit different attitudes towards risk, meaning that individuals exhibit different utility regarding risk.

The determination of utility therefore depends on one's attitude toward accepting risk. Let's see how a utility function can take the place of real money: In the preceding investment example, the best payoff is \$40,000, and the worst is -\$20,000. We thus establish an arbitrary, but logical, utility scale, U, from 0 to 100, in which $U(-\$20,000) = 0$ and $U(\$40,000) = 100$. We can define the Utility Function for this example as follows: If the decision makers attitude is indifferent toward risk, then the resulting utility function will be a straight line joining (0,-\$20,000) and (100, \$40,000).

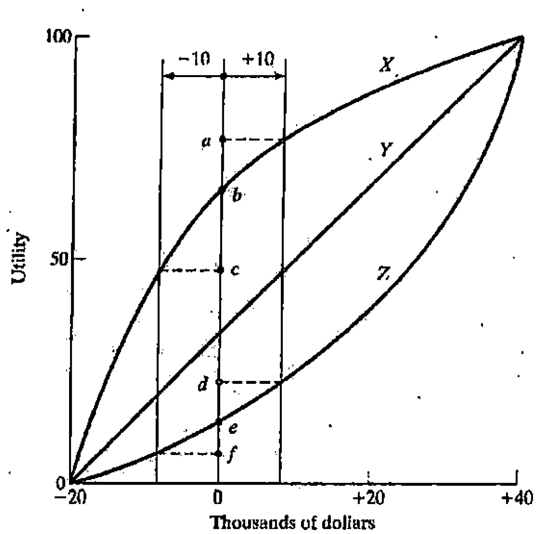

Figure 2.1: Utility Function for Risk Averse (X), Indifferent (Y), and Risk Seeker (Z) Decision Makers

In this case, both the real money and its utility will produce the same decisions. More realistically, the utility function takes over forms that reflect the attitude of the decision maker toward risk. Figure 2.1 illustrates the cases of individuals X, Y, and Z. Individual X is risk-averse because of exhibiting higher sensitivity to loss than to profit. Individual Z is the opposite, and hence is a risk-seeker. The figure demonstrates that for the risk averse individual, X, the drop in utility bc corresponding to a loss of \$10,000 is larger than the increase ab associated with a gain of \$10,000. For the same ± \$10,000 changes, the risk seeker, Z, exhibits an opposite behavior because $de > ef$. Further, individual Y is risk neutral because the suggested changes yield equal changes in utility.

In general, an individual may be both risk averse and risk seeking, in which case the associated utility curve will follow an elongated S-shape.

Utility curves similar to the ones demonstrated in the above figure are determined by quantifying the decision maker's attitude toward risk for different levels of cash money. In our example, the desired range is -\$20,000 to \$40,000, and the corresponding utility range is 0 to 100. What we would like to do is specify the utility associated with intermediate cash values, such as -\$10,000, \$0, \$10,000, \$20,000, and \$30,000. The procedure starts by establishing a lottery for a cash amount x whose expected utility is given as:

$$\begin{aligned} U(x) &= pU(-20,000) + (1-p)\,U(\$40,000), \quad 0 \le p \le 1 \\ &= 0p + 100(1-p) \\ &= 100 - 100p. \end{aligned}$$

To determine $U(x)$, we ask the decision maker to state a preference between a guaranteed cash amount x and the chance to play a lottery in which a loss of $20,000 occurs with probability p, and a profit of $40,000 is realized with probability $1-p$. The decision maker translates the preference by specifying the value of p that will render him indifferent between the two choices. For example, if $x = \$20,000$, the decision maker may say that a guaranteed $20,000 cash and the lottery are equally attractive if $p = .8$. In this case, we can compute the utility of $x = \$20,000$ as $U(\$20,000) = 100 - 100 \times .8 = 20$.

We continue in this manner until we generate enough points [x , $U(x)$] to identify the shape of the utility function. We may then determine the desired utility function using regression analysis or simply by using a piecewise-linear function.

Although we are using a quantitative procedure to determine the utility function, the approach is far from being scientific. The fact that the procedure is entirely driven by the contributed opinion of the decision maker casts doubt on the reliability of the process. In particular, the procedure implicitly assumes that the decision maker is rational. But, this requirement that cannot always be reconciled with the wide changes in behavior and mood that typify human beings. In this regard, decision makers should take the concept of utility in the broad sense that monetary values should not be the only critical factor in decision making (see [FH05], Chapter 13).

Chapter 3

Expected Utility Theory Versus Prospect Theory

The two major decision theories that have undertaken to model human decision making processes are the Expected Utility Theory, and the Prospect Theory [DK79].

In Chapter 2 we saw the three major situations where decision making takes place. Also, we have analyzed the utility function as the basic tool used in the third situation, that of decision making under risk. The Expected Utility Theory, EUT, and the Prospect Theory, PT, have both used the concept of utility function [Cen06]. Although the Expected Utility Theory, EUT, has used this basic approach presented in the previous chapter as its backbone, it has considerably enriched the concept of utility function and has constructed specific expected utility functions that can be used in modeling data pertaining to different real world situations. While doing so, the expected utility has also raised important questions, especially in the areas where it has failed to realistically explain the actual human decision making behaviors. As we will see in the second half of this chapter, the Prospect Theory has been trying to identify and remedy EUT's shortcomings. At this point let's begin by asking the following question: What are the desirable properties any theory of human decision making might want to have? In the following section we illustrate the approach adopted by the Expected Utility Theory.

3.1 Expected Utility Theory

This theory essentially assumes that human decisions are taken based on rational analyses. The ingredients of this rationality are formalized by the Expected Utility Theory. These assumptions constitute the heart of EUT. Interestingly, we will see how EUT's assumptions do not fit into the moods and attitudes of real human beings. In fact, most of EUT's assumptions will end up either being adjusted or plainly replaced by the components of the prospect theory's decision making model. The Prospect Theory will then be capable of better explaining and modeling the process of human decision making in many situations.

This third chapter therefore includes examples of utility theory and details its specific failures. In this section we will show, for instance, that the Expected Utility Theory being true for any utility function is equivalent to a set of axioms: the Von-Neumann Morgenstern axioms. In fact, as we will see, the Prospect Theory initially noted that some of these Von-Neumann Morgenstern axioms are not empirically true: people just don't make decisions like that. The relation between the curvature of the utility function and risk aversion will be further analyzed at this point.

The next section then shows the remedies suggested by the Prospect Theory before assessing this same theory's limitations and extensions. For instance, a short section will show that important properties of the Prospect Theory such as first order stochastic dominance, and transitivity come into play only when there are more than 2 outcomes. Additionally, cumulative Prospect Theory, which is the most theoretically sound formulation of PT because it respects stochastic dominance and transitivity, is mostly useful when there is a continuum of outcomes.

3.1.1 The Expected Utility Property

A utility function u is seen as having the expected utility property if, for a gamble g with prospects $a_1, a_2, ..., a_n$, with effective probabilities $p_1, p_2, ..., p_n$ respectively, we have: $u(g) = p_1u(a_1) + p_2u(a_2) + ... + p_nu(a_n)$ where $u(a_i)$ is the decision-maker's utility for prospect a_i. Definition: An individual who chooses one gamble over another if and only if its expected utility is higher is called an expected utility maximizer.

The main contribution of Von-Neumann and Morgenstern is to prove that, in order for a utility function to exist, and to fulfill the expected utility property, all the preference axioms must be respected (adapted from [Cen06]).

The Preference Axioms Before we construct examples of utility functions over lotteries, or gambles, we need to make the following assumptions on decision makers' preferences. In these examples $\geq$ designates the binary preference relation "is weakly preferred to", which would include both "strictly preferred to", and "indifferent to".

Completeness For 2 given gambles g and g' in G, either $g \leq g'$ or $g \geq g'$. Meaning, people have preferences over all lotteries, and rank them all.

Transitivity For 3 gambles g, g', and g'' in G, if $g \geq g'$ and $g' \geq g''$, then $g \geq g''$. In English, if g is preferred (or indifferent) to g', and g' is preferred (or indifferent) to g'', then g is preferred (or indifferent) to g''.

Continuity Mathematically, this assumption claims that the upper and lower contour sets of a preference relation over lotteries or gamble are closed. In conjunction with the other axioms, continuity is needed in order to ensure that for any gamble in G, there exists some probability where the decision-maker is indifferent between the "best" versus the "worst" outcome. This might seem irrational if the best outcome was, for instance, \$1,000, and the worst outcome was to be run over by a truck. However, one could expect that most rational people would be willing to travel across town to collect a \$1,000 prize, even if this might involve some probability of being run over by a truck.

Monotonicity This ugly word simply means that a lottery which assigns a higher probability to a preferred prospect will be preferred to one that assigns a lower probability to a preferred prospect, as long as the other prospects in the lottery remain unchanged. In this case, we are referring to a strict preference over prospects, and do not consider the case where the decision-maker would be indifferent between possible outcomes.

Substitution If a decision-maker is already indifferent between two possible prospects, then they will be indifferent between two gambles which offer them equal probabilities, should the gambles be identical in every other way, meaning the outcomes can, therefore, be substituted. Thus, if outcomes x and y are indifferent, then one would be indifferent between a lottery giving x with the probability p, and z with the probability $(1-p)$, and a lottery yielding y with the probability p, and z with the probability $(1-p)$. Much the same way, if x is preferred to y, then a lottery yielding x with the probability p, and z with the probability $(1-p)$, is preferred to a lottery yielding y with the probability p, and z with the probability $(1-p)$. This last axiom is usually referred to as the Independence axiom, because it refers to the Independence of Irrelevant Alternatives (IIA). This last axiom allows us to reduce compound prospects to simple prospects, since one can also be indifferent between a simple lottery yielding an outcome x with a probability p, and a compound lottery where the prize might yet be another lottery ticket, allowing one to take part in a lottery with x as a possible reward, such that the effective probability of obtaining x would be p.

3.1.2 Human Decision Making Behaviors Inconsistent with the Expected Utility Theory

In 1979, Daniel Kahneman and Amos Tversky conducted a series of experiments testing the Allais Paradox in Israel, at the University of Stockholm, and at the University of Michigan [GH08]. Everywhere the results produced the same pattern. The problem was even framed in many different manners, with prizes involving money, travels, vacation, and so on. In every case, the substitution axiom was violated in exactly the same pattern. Kahnemann and Tversky called this pattern the certainty effect. This would mean that people overweight outcomes which are certain, compared to outcomes which are merely probable. Using the term "prospect" to designate a set of outcomes with a probability distribution over them, Kahnemann and Tversky also state that whenever winning is possible but not probable, meaning when probabilities are very low, most people choose the prospect which offers the larger gain. This fact is illustrated by the second decision stage of the Allais Paradox. Generalizing, if x and y were outcomes with $0 < p, q, r < 1$, where p, q, and r would refer to probabilities, then they state that:

$(y, pq) \sim (x, p) \Rightarrow (y, pqr) > (x, pr)$; where the term (outcome, probability) means a prospect.

The Reflection Effect

Kahnemann and Tversky also discovered strong evidence of what they called the reflection effect. To illustrate the reflection effect: Imagine a typical Allais Paradox problem, framed in the following manner. You first have to choose one of the two gambles, or prospects: Gamble A: A 100% probability of losing \$3000. Gamble B: An 80% chance of losing \$4000, and a 20% probability of losing nothing. Next, you will have to choose between: Gamble C: A 100% probability of receiving \$3000. Gamble D: An 80% probability of receiving \$4000, and a 20% probability of receiving nothing. Kahnemann and Tversky discovered that 20% of people actually chose D; while 92% would chose B. A similar pattern was observed not only for varying positive, but also for negative prizes. This led them to conclude that for decision problems involving possible losses, people's preferences over negative prospects are often a mirror image of their preferences over positive prospects. In other words, the same way they are risk-verse over prospects involving gains, people often become risk-loving when prospects involve losses [NB07].

Combination of the Certainty and Reflection Effects

In case of positive prospects, the certainty effect results in a risk-averse preference for a sure gain, rather than a gain which might be larger but merely probable. However, in case of negative prospects, symmetrically, people adopt risk-loving preferences for larger losses which remain probable, over smaller but certain losses. At this point, one would imagine that if this observation held universally then one would never see people buying insurance. Yet, what this really means, as we will see again in the section on probability transformations, is that when losses are involved with moderate or high probabilities, then risk seeking is often predicted. Prospect theory does, in fact, predict risk-aversion behavior for small-probability losses, which is normally the case in insurance decisions.

The Isolation Effect

Imagine now another lottery problem. Having to choice between the following, which one would you choose? Gamble A: A 25% chance of winning \$3000. Gamble B: A 20% chance of winning \$4000, and an 80% chance of winning nothing. Now imagine having to make a decision in a two-stage problem. The first stage involves a probability of

0.75 for ending the game without losing nor winning anything, and a probability of 0.25 for moving to the second stage, where you are to face with the following choice: Gamble C: A 100% probability of winning \$3000. Gamble D: An 80% probability of winning \$4000, and a 20% probability of winning nothing. 65% of people would chose B, while 78% would chose C. Why would this seem surprising? In fact, the true probabilities involved in the second choice can be rewritten as: $0.25 \times 1 = 0.25$ probability of winning \$3000, and $0.25 \times 0.8 = 0.2$ probability of winning \$4000. Kahnemann and Tversky interpreted this discovery in the following terms: When simplifying the choice between alternatives, people usually disregard components that are shared among, and only focus on those which distinguish them. Because different choice problems can often be decomposed in different ways, this could lead to inconsistent preferences, as seen above. Kahnemann and Tversky called this phenomenon the isolation effect.

The above discoveries about the human decision making behavior are inconsistent with the purely rational assumptions made by the Expected Utility Theory. Therefore, one would need another explicative model which would better fit the contours of the human soul. This is precisely what The Prospect Theory has the ambition to accomplish.

3.2 The Prospect Theory

Given the effects presented above, Kahneman and Tversky suggested a new theory of decision-making under risk, which they called Prospect Theory (see [Wik09], and [DK79]). Prospect Theory differs from Expected Utility Theory in many fundamental ways. Firstly, it distinguishes two phases in the decision-making process: An editing phase, which represents a preliminary analysis of the offered prospects, followed by an evaluation phase, where the prospect perceived as the highest value is chosen among the edited prospects.

3.2.1 The Editing Phase

In the editing phase, a decision-maker reorganizes and reformulates the available options, in an effort to simplify the choice. It consists of the following operations (adapted from [Cen06]):

Coding People perceive prospects as gains or losses, as seen in the above examples, as opposed to final states of wealth or welfare. A gain or loss is, therefore, defined in comparison with some reference point. The location of the reference point affects whether the outcomes are coded as gains or as losses.

Combination Prospects are simplified by combining the probabilities associated with identical outcomes. For instance, the prospect (200, 0.25; 200, 0.25), meaning having two consecutive 25% chances of winning 200 dollars, will be reduced to (200, 0.5). In this example the subject is expected to choose between a prospect where a 25% chance of winning 200 dollars is offered twice in a row, versus some other prospect not mentioned here. Therefore, before even considering the other prospect, the Combination component of the Editing Phase would add the two consecutive 25% chances of winning 200 dollars, by a 50% chance of winning 200 dollars.

Segregation The riskless part of any prospect is separated from its risky part. For instance, the prospect (300, 0.8; 200, 0.2) is decomposed into a sure component of 200 and a risky prospect (100, 0.8). A similar process is applied for losses.

The above editing operations are applied to each prospect separately. Whereas, the following are applied to combinations of two or more prospects:

Simplification Prospects are exposed to be rounded off. For example, a prospect of (101, 0.49) could be seen as a 50-50 chance to win 100 dollars. Also, extremely unlikely outcomes could just be discarded.

Identification of Dominance Outcomes that are strictly dominated are identified and rejected without further evaluation.

Note that some editing operations would allow or prevent others from being carried out. The sequence of editing operations could often vary with the offered setting and the format of the display. As we can imagine, many preference anomalies can arise from the act of editing. For instance, inconsistencies described by the isolation effect, would result from the cancellation of common components. Also, intransitivity cases can result from a simplification that would eliminate small differences between prospects.

3.2.2 The Evaluation Phase

During the evaluation phase, an individual examines all the available prospects and chooses the one perceived to represent the highest value. The overall value of a prospect, V, is expressed in terms of two scales, π and v. The first scale, π, associates some decision weight $\pi(p)$ with each probability p. This would reflect the impact of p on the global value of the prospect. It is important to mention that π is not a probability measure. Also, Kahneman and Tversky prove that $\pi(p) + \pi(1-p)$ is frequently less than 1. The second scale, v, assigns a number $v(x)$ to each outcome x. This would reflect the subjective value of that outcome. Recalling that outcomes are defined relative to a reference point, used as a zero point, v measures deviations from that reference point. Considering a simple prospect of the form $(x, p; y, q)$, where the subject would win x with probability p, y with probability q, and nothing with probability $1 - p - q$, and where $p + q \leq 1$, we say: An offered prospect is strictly positive if its outcomes are all positive. Meaning $(x, p; y, q)$ where $x, y > 0$ and $p + q = 1$. It is strictly negative when all its outcomes are negative. It is regular when it is neither strictly positive, nor strictly negative. Therefore, for a regular prospect, where either $p + q < 1$, or, $x \leq 0 \leq y$, we would have: $V(x, p; y, q) = \pi(p)v(x) + \pi(q)v(y)$, where $v(0) = 0, \pi(0) = 0$, and $\pi(1) = 1$. V is defined on prospects, whereas v is defined on outcomes. The evaluation of strictly positive or strictly negative prospects would follow a different rule, described below: If $p + q = 1$, where either $x > y > 0$, or $x < y < 0$, then:

$V(x, p; y, q) = v(y) + \pi(p)[v(x) - v(y)]$, so that the value of a strictly positive or strictly negative prospect will equal the value of the riskless component augmented by the difference between the values of the two outcomes, multiplied by the weight assigned to the more extreme outcome. Note that a decision weight is only applied to the risky component, not to the riskless one.

For example: $V(400, 0.25; 100, 0.75) = v(100) + \pi(0.25)[v(400) - v(100)]$ Meaning that a decision weight is only applied to the difference in value $v(x) - v(y)$, but not to the riskless component, $v(100)$. Also note that the right side of the equation above is simplified to $\pi(p)v(x) + [1 - \pi(p)]v(y)$. This, in turn, reduces to $\pi(p)v(x) + \pi(q)v(y)$, the equation for a regular prospect, when $\pi(p) + \pi(1-p) = 1$. However, this is usually not satisfied.

While the Prospect Theory equations might appear to resemble those of the Expected Utility Theory, their crucial differences are:

1. Values are attached to changes with respect to reference point, rather than to final states, and
2. The decision weights do not need to coincide with probabilities.

3.3 Additional Observations on Value Functions

The focus on changes as the carriers of value shall not mean that the value of a particular change is totally independent of the initial position. In fact, value functions can become more linear with increases in assets. A change going from \$100 to \$200 would likewise have a much higher value than one from \$1100 to \$1200. The value function is, in fact, concave above the reference point ($v''(x) < 0$ for $x > 0$), and convex below it ($v''(x) > 0$ for $x < 0$). Meaning, it is concave for gains but convex for losses. For instance, most people do not like symmetric gambles of the type (50,0.5; -50,0.5). Thus, if $x > y \geq 0$, then $(y, 0.5; -y, 0.5)$ will be preferred to $(x, 0.5; -x, 0.5)$. This would mean that $v(y) + v(-y) > v(x) + v(-x)$. Letting $y = 0$ gives us $v(x) < -v(-x)$, while letting y approach x gives $v'(x) < v'(-x)$, as long as v remains differentiable. Therefore the value function for losses would be steeper than that for gains.

Thus, for instance, choosing between the prospect of winning or losing \$100 with a 50% chance in either case, expressed as (100,0.5 ; -100,0.5) against another prospect of win or lose, but this time on \$10,000 and with the same 50% chance in either case, expressed as (10000,0.5 ; -10000,0.5), most people would choose the first prospect where the exposure is only a loss of \$100. Yet, from the strict viewpoint of the EUT assumptions the two prospects are equivalent with their respective Utilities equal to 0.

Remark on the weighting function: The weighting function π, which associates decision weights to given probabilities, is an increasing function of p, where $\pi(0) = 0$ and $\pi(1) = 1$. However, people usually overweight very small probabilities, like 0.001, so that $\pi(p) > p$ in cases of very small p.

Definition 1. *In probability and statistics theory, a stochastic order quantifies the concept of a random variable being "larger" than another. These are often partial orders, so that one random variable A might be neither stochastically larger than, less than, nor equal to another random variable B. Many different orders exist, and have different applications.*

A real random variable A is said to be less than a real random variable B in the "usual stochastic order" if $p(A > x) \leq p(B > x)$ for all $x \in (-\infty, \infty)$.

Definition 2. *Stochastic dominance is a type of stochastic ordering. The term is used in decision theory and decision analysis in order to refer to situations where one prospect, can be ranked as superior to another prospect. It is based on preferences regarding the outcomes. A preference may be a simple ranking of outcomes from the most to the least favored, or it may also use a value measure, meaning a number associated with each outcome which allows comparison of multiples between one outcome and another outcome, such as two instances of winning a dollar versus one instance of winning two dollars. Only limited knowledge of preferences is needed for determining dominance. Risk aversion, using this definition, is therefore a factor in second order stochastic dominance only.*

In our comparison of Expected Utility Theory versus the Prospect Theory, the first order stochastic order and stochastic dominance are only mathematically modeled by the EUT. Whereas the PT identifies, defines, and uses human interpretations of the otherwise pure quantitative values, in order to re-establish stochastic order and dominance that better fit and explain human decision making behaviors.

Definition 3. *Statewise dominance. The simplest case of stochastic dominance is statewise dominance, also referred to as state-by-state dominance, defined as follows: Prospect A is statewise dominant over prospect B if A yields a better outcome compared to B in every possible state. More precisely, A would yield at least as good an outcome in every state, with strict inequality in at least one state. For instance, if a dollar is added to one or more prizes in a lottery, then the new lottery would statewise dominate the old one. Similarly, if a risk insurance policy offers a lower premium and a better coverage than another policy, then with or without damage, the outcome will be better. In simple terms, anyone who prefers more to less, or said in formal terminology, anyone having monotonically increasing preferences, will always prefer the statewise dominant prospect.*

In our comparison of EUT versus PT, the feature of statewise dominance is assumed to be true and operational by the EUT whereas the PT may easily accept to ignore it through the Editing step of its Utility Function.

3.3.1 First-Order Stochastic Dominance

Statewise dominance is a special case of the first-order stochastic dominance. First order dominance is defined as follows: Prospect A has first-order stochastic dominance over prospect B if for any outcome x, A yields at least as high a probability of winning at least x as does B, and for some x, A yields a higher probability of winning at least x. In notation form, $p(A \geq x) \geq p(B \geq x)$ for all x, and, $p(A \geq x) > p(B \geq x)$, for some x [Wik09]. Now, let us define F_A and F_B as the cumulative distribution functions of the prospects. In terms of the cumulative distribution functions of the two prospects, A dominates B, means that $F_A(x) \geq F_B(x)$ for all x, with strict inequality for some x. For instance, consider a die toss where 1 through 3 would win \$1 and 4 through 6 would win \$2 in prospect B. This is dominated by a prospect A that yields \$3 for 4 through 6 and \$1 for 1 through 3, and also dominated by a prospect C which yields \$2 for 3 through 6 and \$1 for 1 and 2. Prospect A would, in this case, have statewise dominance over B if we re-order the values won by the die toss outcomes, whereas prospect C will keep first-order stochastic dominance over B, and without statewise dominance, no matter what the order of prospects. Further, although whenever A dominates B, the expected value of the payoff in A is greater than the expected value of the payoff in B, this is not a sufficient condition for dominance, and thus, one cannot order lotteries using the concept of stochastic dominance simply by comparing their probability distribution means.

Every expected utility maximizer with an increasing utility function will prefer gamble A over gamble B if and only if A first-order stochastically dominates B. But then again, the concept of expected utility maximizer is a necessary assumption for EUT, whereas the PT at the cost of loosing some coherence provides for deviation even from the first-order stochastic dominance [NB07]. In fact, we will see in the following chapter the extent to which the human behavior in decision making for a subpopulation can deviate from the EUT model, in which case EUT would then apply only to the other portion of the population. Modeling these disparities is exactly the motivation behind the mixture theories that we are analyzing in the following chapter.

3.3.2 Second-Order Stochastic Dominance

The other commonly used type of stochastic dominance is second-order stochastic dominance. Roughly speaking, for two gambles A and B, gamble A has second-order stochastic dominance over gamble B if A is more predictable, meaning involves less risk, and has at least as high a mean. All risk-averse expected-utility maximizers (that is, those with increasing and concave utility functions) prefer a second-order stochastically dominant gamble to a dominated gamble. The same is true for non-expected utility maximizers with utility functions that are locally concave.

For an example, let's compare the die toss gamble A as seen previously, with the gamble B respectively as follows:

Gamble A	\$1	\$1	\$1	\$2	\$2	\$2	
Probability	1/6	1/6	1/6	1/6	1/6	1/6	$\sum = 1.0$
Exp. Utility	\$1/6	\$1/6	\$1/6	\$2/6	\$2/6	\$2/6	$\sum = 1.5$
Gamble B	-\$1	\$1	\$2	\$2	\$2	\$3	
Probability	1/6	1/6	1/6	1/6	1/6	1/6	$\sum = 1.0$
Exp. Utility	-\$1/6	\$1/6	\$2/6	\$2/6	\$2/6	\$3/6	$\sum = 1.5$

Note that the Expected Utility of prospect $A = \$1.50$, while that of prospect B also carries a the total Expected Utility of \$1.50.
Yet, gamble B will be dominated by the gamble A, because, as we have seen previously, the \$-1 which represents a possible loss of 1 dollar, is perceived as involvement in additional risk involved in the gamble B, compared to the gamble A.

Using cumulative distribution functions F_A and F_B, A second-order stochastically dominates over B if and only if the area under F_A from minus infinity to x is greater than or equal to the area under F_B from minus infinity to x for all real numbers x, with strict inequality for some x.

For an example, let us consider the gamble C, compared to the gamble D both laid out below [Wik09]. In this example we are using discrete measures, but the example illustrates the cumulative distribution functions (CDF) of F_C and F_D, and shows the dominance of D over C.

Gamble C	-\$2	-\$1	\$1	\$2	\$3	\$4	
Probability	1/12	0/12	5/12	4/12	1/12	1/12	$\sum = 1.0$
Exp. Utility	-\$2/12	\$0	\$5/12	\$8/12	\$3/12	\$4/12	$\sum = 1.5$
CDF	-\$2/12	-\$2/12	\$3/12	\$11/12	\$14/12	\$18/12	$\sum = 1.5$

Gamble D will dominate gamble C if and only if its Cumulative Distribution Function, meaning its cumulative densities provided by the last line CDF dominates that of the gamble C in at least one area.

Gamble D	-\$2	-\$1	\$1.2	\$7/4	\$3	\$4	
Probability	1/12	0/12	5/12	4/12	1/12	1/12	$\sum = 1.0$
Exp. Utility	-\$2/12	\$0	\$6/12	\$7/12	\$3/12	\$4/12	$\sum = 1.5$
CDF	-\$2/12	-\$2/12	\$4/12	\$11/12	\$14/12	\$18/12	$\sum = 1.5$

Thus gamble D dominates gamble B since F_D, the CDF for D is greater or equal to F_C for all x, and strictly greater at $x = \$1.2$.

3.3.3 Cumulative Prospect Theory

Cumulative Prospect Theory (CPT) is a model to describe decisions under risk which has been introduced by Amos Tversky and Daniel Kahneman in 1992 [DK79]. CPT is therefore a further development and a variant of Prospect Theory. Its main difference compared to the original version of Prospect Theory is that the weighting is applied to the cumulative probability distribution function, as it also is the case in rank-dependent Expected Utility Theory, rather than to the probability distributions of individual outcomes. For his contributions to behavioral economics, and specially for his development of the Cumulative Prospect Theory (CPT), Daniel Kahneman has received the prize in Economic Sciences in 2002.

Outline of the Model

(Adapted from [Wik09]) The value function both in Prospect Theory and in Cumulative Prospect Theory, is based on a reference point. This is in contrast with the Expected Utility Theory which deals with final outcomes. The reference point in PT and CPT corresponds to what the subjects perceive as the breakpoint between losing and winning. In Figure 3.1, we can see that the reference point is placed at the origin. This figure also shows that the slope on the negative side is sharper than on the positive side, corresponding to the loss prospects looming more than prospects of gain.

Figure 3.1 shows a typical value function in Prospect Theory and Cumulative Prospect Theory. It assigns values to possible outcomes of a lottery. On the other hand,

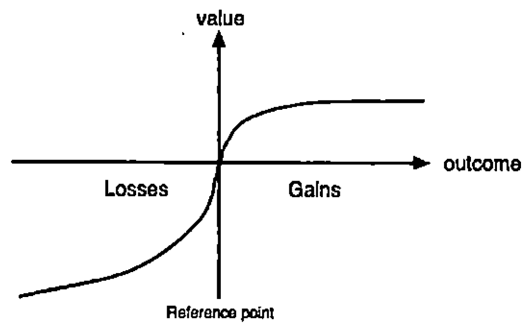

Figure 3.1: A Typical Value Function for Prospect Theory and Cumulative Prospect Theory

a typical weighting function in the Cumulative Prospect Theory, represented graphically by Figure 3.2, overweights both ends of the probability distribution, not due to the over weighing of very small and very large probabilities, but instead, due to the relative values that usually coincide with those probabilities.

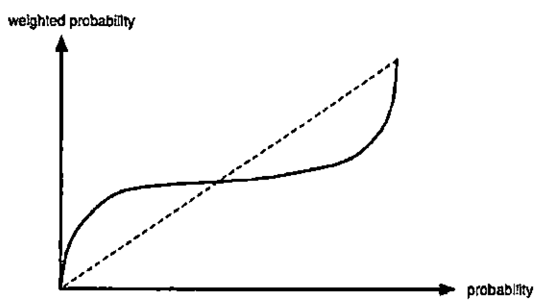

Figure 3.2: Typical Weighting Function in Cumulative Prospect Theory

Figure 3.2 is a typical weighting function in Cumulative Prospect Theory [MA07]. It transforms objective cumulative probabilities into subjective cumulative probabilities. To summarize, the main observation of CPT, and its predecessor Prospect Theory, is that people usually think of possible outcomes relative to a certain reference point, often the status quo, rather than based on to the final status. This phenomenon is called the framing effect. Moreover, people have different risk attitudes towards gains, meaning simply towards outcomes above the reference point, compared to their attitudes towards losses, or outcomes below the reference point. In essence, people feel generally more strongly about potential losses than they do about potential gains. And finally, people in CPT tend to overweight extreme, and unlikely, prospects, while they underweight "average" prospects. This would mean, for example, that both small amounts in a lottery prospect as well as the very large amounts, at the other end will be overweighted, while most people would underweight the intermediary amounts. This last point is in contrast to Prospect Theory which assumes that people overweight unlikely events, without regard to where in the spectrum of amounts they are located. Thus, CPT incorporates these facts in a modification of Expected Utility Theory by replacing final wealth with outcomes relative to a reference point, replacing the utility function by a value function that depends on relative outcomes, and replacing cumulative probability distributions with weighted cumulative probability distributions. In the general case, this can be represented by the following formula for the subjective utility of a risky outcome described by the probability measure p:

$$U(p) := \int_{-\infty}^{0} v(x) \frac{d}{dx} \left(w \left(F(x) \right) \right) dx + \int_{0}^{+\infty} v(x) \frac{d}{dx} (-w(1 - F(x))) dx,$$

where v is the value function, whose typical form was shown in Figure 3.1, and w is the weighting function, as graphed in Figure 3.2. In this formula, $F(x) := \int_{-\infty}^{x} dp$, represents the integral of the probability densities over all values up to x, thus it is the cumulative probability. It is what we have been referring to as cumulative probability densities. The function w, therefore, represents the twist that the decision maker's subjective perceptions inflict to $F(x)$ rather than to the individual probabilities. This formula generalizes the original formulation by Tversky and Kahneman from finitely many distinct outcomes to infinite, and therefore continuous, outcomes.

The main modification to Prospect Theory is that, as in rank-dependent Expected Utility Theory, cumulative probability distributions are transformed, rather than the probabilities themselves. This takes us to the overweighting of extreme events which occur only at both ends, carrying for instance very small or very large outcomes, with small probabilities, rather than to an overweighting of every small probability regardless of outcome values. The modification helps to avoid the violation of first order stochastic dominance by the PT, and makes the generalization to arbitrary outcome distributions easier. CPT is therefore from a theoretical standpoint an improvement over Prospect Theory (See [Wik09]).

Chapter 4

Finite Mixture Models

This chapter analyzes the more recent models developed by Glenn W. Harrison, E. Elisabet Rutstrom in 2008. These authors suggest that a finite mixture model can be used to estimate the parameters of each decision process while simultaneously estimating the probability that each theory applies to the sample. We will then look at scale independence of these models: Do behavioral predictions of these models depend on the denomination/units of the rewards/losses? It could be that peoples' decisions are influenced by the denominations or units in which rewards are counted; however this might seem unlikely. What are the conditions on the families of utility functions and EUT and PT frameworks for the decisions predicted to be unchanged by rescaling the rewards/losses, once the model parameters have been inferred?

A finite mixture model can be used to estimate the parameters of each decision process while simultaneously estimating the probability that each process applies to the sample [Eve96]. In this chapter we will be using the canonical case of lottery choices in a laboratory experiment. The main focus of this section is the heterogeneity of the subpopulations and its treatment by the mixture model. More precisely, dealing with heterogeneity, we want to identify which people behave according to what theory and where. This would allow for heterogeneous theories to co-exist within a grand likelihood function. As a result, we will no more need to pose the famously extreme, unrealistic, and increasingly criticized assumption known as Representative Agent. In fact, the idea of Representative Agent and his expected decision making behavior in society have long been used in the Economic Theories, before being recently severely criticized and essentially abandoned. Let us first summarize our experimentation context, or frames, and the specifications of the two models EUT and PT that we are intending to nest within one same Grand Likelihood function. This would mean that although we admit that the heterogeneity can involve more than two different families of populations, for our purpose, and without loss of generality, we are using a model that would only contain EUT and PT decision making patterns. The decision making contexts or frames used in this chapter for thé purpose of building a finite mixture model are as follows:

The data gathered by Glenn W. Harrison, E. Elisabet Rutstrom involves 158 subjects who took part in their experiment. They will be making a total of 9311 choices. That is, each individual would be making about 60 decisions. Our gain frame contains 0, 5, 10, 15 dollar prizes, involving 63 subjects. The loss frame starts off with a 15 dollar initial endowment, then 0, -5, -10, -15 dollar prizes, involving 53 subjects. The purpose of the initial endowment for decisions involving possible losses is to make sure the net final outcomes remain positive. The mixed frame offers an $8 initial endowment to each subject, and then -8, -3, 3, 8 dollar prizes, involving 37 subjects. Again, the initial endowment is offered in order to avoid negative net prospects. An initial random endowment of 1 to 10 dollars is affected to all participants in order to further raise the possible outcomes towards the positive side. Probabilities used in the decision making situations are: 0, 0.13, 0.25, 0.37, 0.5, 0.62, 0.75, and 0.87. These probabilities are roughly evenly spread across the interval 0 to 1. In fact, the steps are either +0.12 or +0.13. For every prospect, these probabilities indicate the chances of receiving the indicated prize. The even distribution of the probabilities would help avoid possible biases due to abrupt changes in the chances of winning the prospects. In fact, the experimentation is based on presenting to the decider one pair of prospects on the left side of the screen, and one pair of prospects on the right hand side. The subject would therefore choose the pair of prospects that seems more worthy than the other pair. Thus, given the above probabilities a typical question presented to a given individual would involve two amounts x and y on the left hand side, and two amounts z and w on the right hand side. Each amount would have an associated probability of win. The formalized representation of such questions would look like the following:

Left prospect (x, .13; y, .37) versus Right prospect (z, .25; w, .75).

4.1 Expected Utility Specification

The CRRA, Constant Relative Risk Aversion, called the parameter r, is defined over the final monetary prize [GH08]. As seen above, the monetary prize is forced to be positive due to initial endowments. We assume that the utility function is given by: $U(s,x) = (s+x)^r$ where,

- r is the CRRA parameter to be determined,

- s is the fixed endowment mentioned above, and

- x is the lottery prize.

- The expected utility of every lottery $EU = \sum_{k=1}^{4} P_k.U_k$, where U is the utility function specified above and indexed here from 1 to 4 since 4 outcomes are presented to the subject in each lottery. And, P_k represents the probability associated to U_k. The values used for probabilities P_k are chosen from the evenly distributed values seen above.

- For a given r, $\nabla EU = EU_R - EU_L$ is called index ∇EU. This index represents the difference between the pair of lottery prizes presented to the subject on the right hand side of the screen, and the pair of prize presented on the left.

- Logistic function: $G(\nabla EU) = \frac{e^{(\nabla EU)}}{1+e^{\nabla EU}}$

This Logistic Function can be interpreted as: The probability that the right outcome be chosen is $G(\nabla EU)$. Thus, the probability that the left outcome be chosen is $(1 - G(\nabla EU))$.

Now using the binary notation $y_i = 1$ if right, $y_i = 0$ if left, we can write:

$Y_i = (G^{y_i})(1-G)^{(1-y_i)^i}$. This means $Y_i = G$ when choice of Right, and, $Y_i = 1 - G$ when choice of Left.

Note: Y represents a given person; i, a given bet; Y_i, the probability of observing the response y_i.

Considering that each bet is independent, the probability of one given set of outcomes, or likelihood, is the product $\prod_i Y_i$, which obviously results in an extremely small number. Too small to work with in fact! Therefore we use their Logs, which will of course be negative values.

The function Log-likelihood will thus end up looking like:

$$Ln^{EUT}(r;y,X) = \sum_i LnL_i{}^{EUT} = \sum_i \left[y_i LnG(\nabla EU_i) + (1-y_i)\, Ln(1 - G(\nabla EU_i))\right],$$

where X is a vector of individual characteristics that implicitly conditions ∇EUthrough r.

That is: $\widehat{r} = \widehat{r_0} + (\widehat{r_F} \times F) + (\widehat{r_B} \times B) + (\widehat{r_H} \times H) + \left(\widehat{r_{BUS}} \times BUS\right) + \left(\widehat{r_{GPA}} \times GPA\right) + (\widehat{r_{age}} \times age)$. This function will ultimately be going to be maximized.

4.2 Prospect Theory Specification

Tversky and Kahnman have used a popular parametric specification which is the one used here. The two main components of this specification are the utility function, and the probability weighting function.

The Utility function applies over gains and losses separately and relative to a reference point, as opposed to applying to final outcomes as in EUT.

A probability weighting function humanizes the pure rational probabilities and gives us subjective probabilities, thus more humanized.

Other characteristics of the model are as follows:

Losses loom larger than gains, since the humans would behave as such.

Non linearity in the transformed probabilities accounts for different risk attitudes.

This mixture model also provides for individual characteristics (called vector X) built into r for EUT and into α, β, λ for PT.

The separate functions defined for gains and losses are:

Gains $(x \geq 0) \Rightarrow U(x) = x^{\alpha}$,

Losses $(x \langle 0) \Rightarrow U(x) = -\lambda(-x)^{\beta}$;

where α and β are risk aversion parameters, and λ is the coefficient of loss aversion.

Subjective probabilities which characterize the Prospect Theory are here modeled by:

$\omega(P) = \frac{P^{\gamma}}{[P^{\gamma}+(1-P)^{\gamma}]^{\frac{1}{\gamma}}}$, where ω humanizes the propabilities, or said differently, introduces the decision maker's subjectivities. The rest of procedures are identical to EUT; meaning :

$\nabla PU = PU_R - PU_L$.

The likelihood, therefore, will depend on estimates of $\alpha, \beta, \lambda, \gamma$ and on ingredients of X.

And finally, the function Log-likelihood for the Prospect Theory component of the mixture model at this point is:

$$LnL^{PT}(\alpha, \beta, \lambda, \gamma; y, X) = \sum_i [y_i LnG(\nabla PU_i) + (1 - y_i) Ln(1 - G(\nabla PU_i))]$$
$$= \sum_i LnL_i^{PT}$$

4.3 Nesting of the Expected Utility Theory and Prospect Theory Models Inside the Function Grand-Log-Likelihood

Let $\prod^{EUT}$denote the probability that EUT is the correct model, and, $\prod^{PT}$ denote the probability that PT is the correct model. Then: $\prod^{PT} = \left(1 - \prod^{EUT}\right)$ The function GrandLog-Likelihood can now be written as: $LnL\left(r, \alpha, \beta, \lambda, \gamma, \prod^{EUT}; y, X\right) = \sum_i Ln\left[\left(\prod^{EUT} \times L_i^{EUT}\right) + \left(\prod^{PT} \times L_i^{PT}\right)\right]$

At this point, the right choices of the parameter arguments will be the ones that maximize the Grand-Log-Likelihood.

4.3.1 The Results of the Finite Mixture Model Experimentation

Table 4.1 Table 4.1 shows that EUT and PT probabilities indicate that each is equally likely for the data we had [GH08]. They are 0.550 versus 0.450.

Secondly, we see that the estimates for PT specification are only weekly consistent with a priori predictions of the theory: $\lambda = 1.380$ while$\gamma = 0.911$, are both too close to 1; $\alpha = 0.710$ and $\beta = 0.723$ are almost identical.. $d > 0$

Whereas, When the mixture model is used, the value for $\lambda = 5.781$, while $\gamma = 0.681$ becomes, both more PT like. While, $\alpha = 0.614$ and $\beta = 0.132$, are no more identical. Using the PT model for the subpopulation that has a better probability of fitting

Parameter or Test	Estimates from Conditional Models					Estimates from Mixture Model				
	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
r	0.867	0.029	0.000	0.809	0.924	0.846	0.044	0.000	0.759	0.933
α	0.710	0.046	0.000	0.620	0.801	0.614	0.057	0.000	0.501	0.727
β	0.723	0.065	0.000	0.695	0.851	0.312	0.132	0.019	0.052	0.572
λ	1.380	0.223	0.000	0.940	1.820	5.781	1.612	0.000	2.598	8.965
γ	0.911	0.061	0.000	0.790	1.033	0.681	0.047	0.000	0.587	0.774
π^{EUT}						0.550	0.072	0.000	0.407	0.692
π^{PT}						0.450	0.072	0.000	0.308	0.592
$H_0: \pi^{EUT} = \pi^{PT}$								0.490		
$H_0: \alpha = \beta$			0.861					0.046		
$H_0: \lambda = 1$			0.090					0.003		
$H_0: \gamma = 1$			0.151					0.000		

Table 4.1: Estimates for Parameters in Conditional and Mixture Models

into that model and letting EUT talk for the rest of the population seems therefore to be better modeling of this set of data.

Table 4.2 Table 4.2 shows the optimal estimates calculated under mixture model with individual characteristics included. For example, a 21 year old non Hispanic black female, who did not have a business major and had an average GPA, would have an estimate of r given by:

$$r = r_0 + \hat{r}FEMALE + \hat{r}BLACK + \hat{r}AGE \times 21.$$

Notice that the set of estimated characteristics is reasonably large, allowing considerable heterogeneity for a given subject and among subjects.

Figure 4.1 Figure 4.1 shows the distribution of predicted probabilities of the two competing models. The two panels are by construction the mirror image of each other. In fact, we have had $\prod^{PT} = \left(1 - \prod^{EUT}\right)$, since we only have two competing models.

Figure 4.2 Figure 4.2 indicates the uncertainty of these predicted probabilities. Note that uncertainty is smaller at the end points. This result is consistent with the use of discrimination functions such as the logistic function. Also, EUT has more of its support closer to the upper end-point where some subjects are better characterized by it.

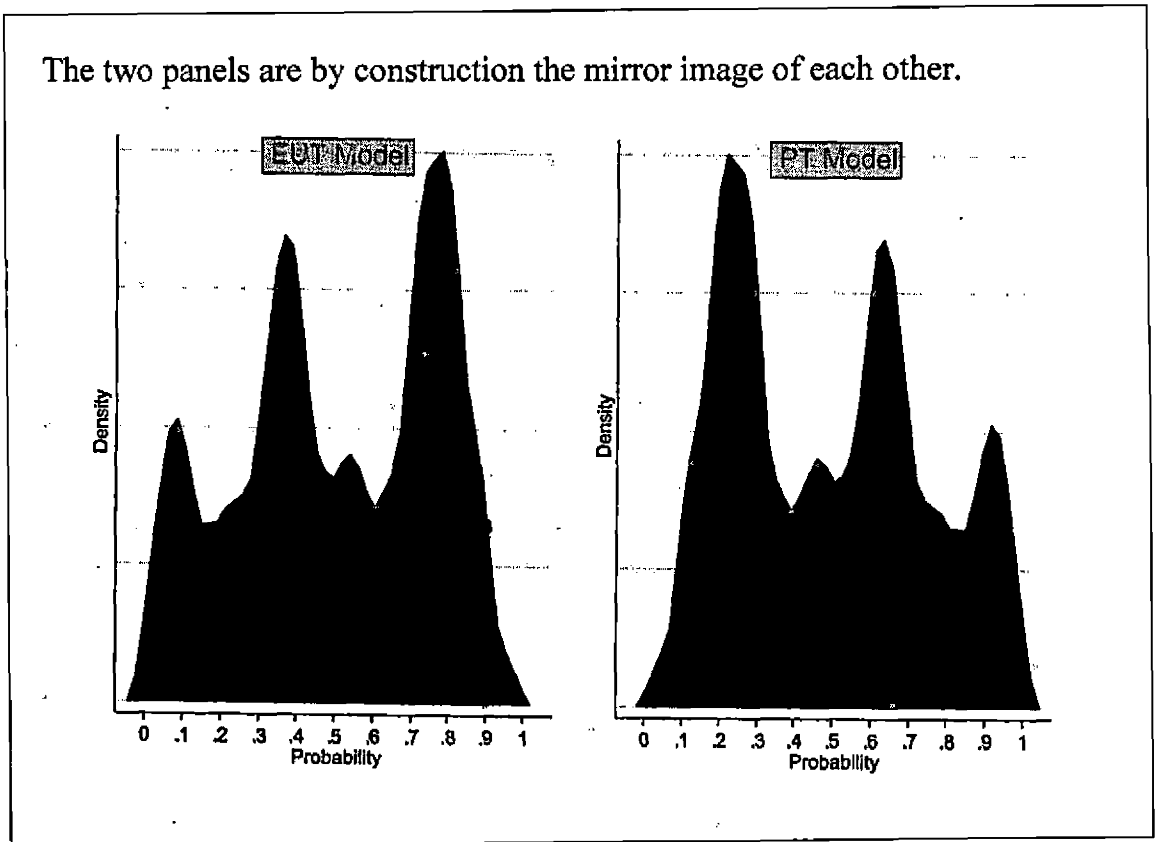

Figure 4.1: Probability of Competing Models

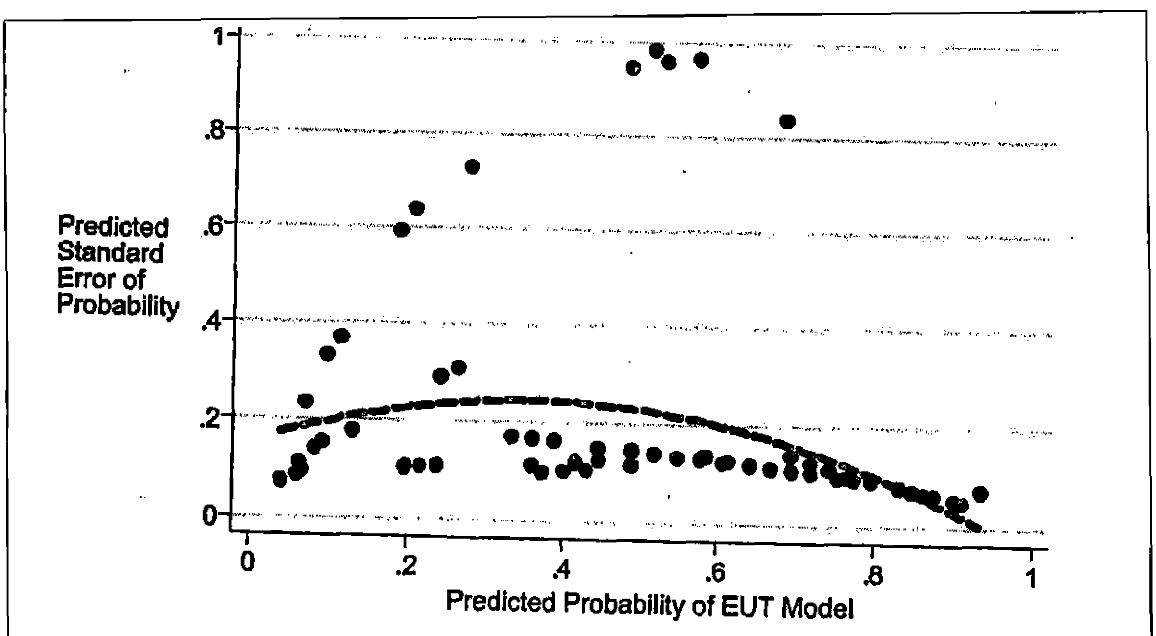

Figure 4.2: Predicted Probability of Expected Utility Theory Model

Figures 4.3 and 4.4 Figures 4.3 and 4.4 clearly indicate that males are more likely to behave according to EUT model. Whereas females density is highest on the smaller $\prod^{EUT}$ probabilities, indicating that their behavior is less likely to be according to EUT.

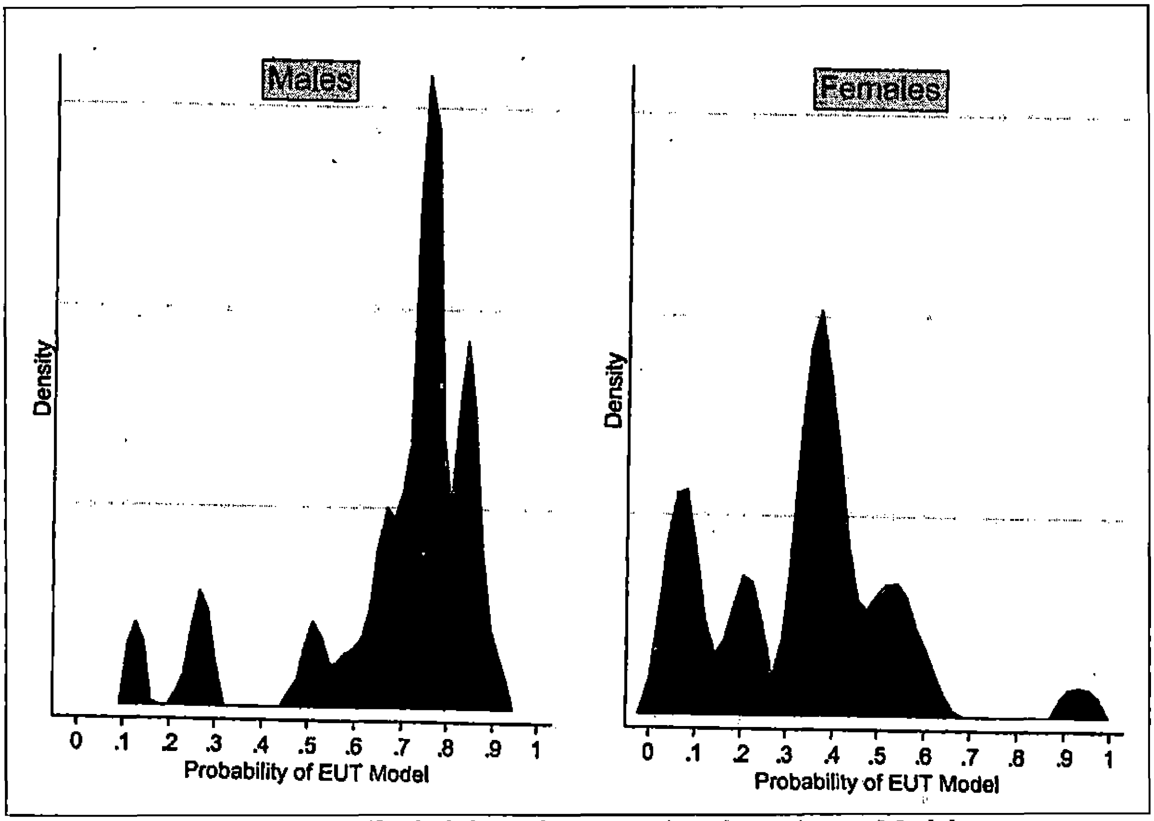

Figure 4.3: Probability of Expected Utility Theory Model

On the other hand, the black almost never make their decisions on the EUT model. For the others, they are mixed, with majority of them on the right side of $\prod^{EUT} = 0.5$.

Figures 4.5 and 4.6 Another benefit of using a mixed model can be seen here: Among those with at least 25% chance of being EUT-consistent, that is 126 subjects out of our total of 158, their CRRA, or risk aversion coefficient average of 0.98.

As expected from an EUT subpopulation.

Whereas, the same CRRA components among those with at least 25% chance of being PT-consistent, that is 114 out of 158 subjects, are as follows: $\alpha < 0.5$ avg

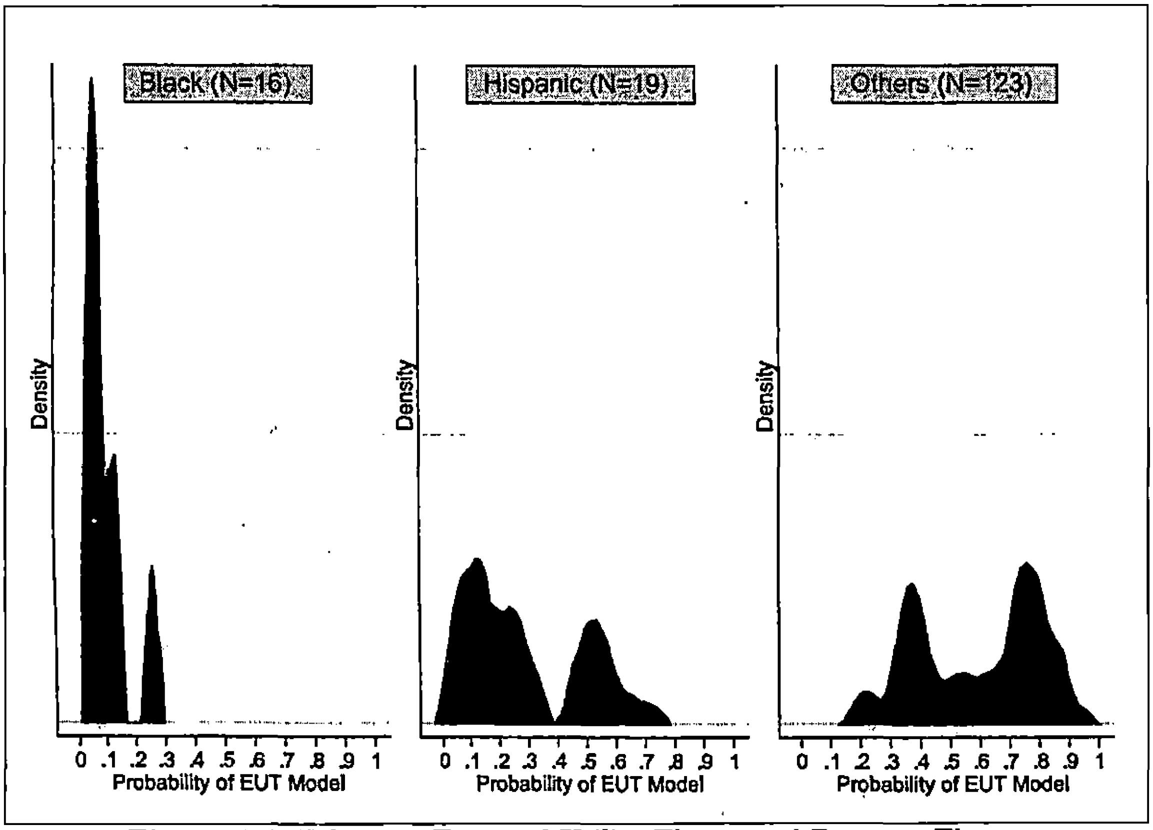

Figure 4.4: Ethnicity, Expected Utility Theory and Prospect Theory

0.44 $\beta > 0.5$ avg 0.51 $\lambda > 1$ avg 5.81 $\gamma < 1$ avg 0.89

As expected from a PT-inclined subpopulation.

4.4 Scale Independence of Finite Mixture Models

A utility function is a function U of some amount of money x that represents the satisfaction of taking possession of x. Utility functions are used both in EU and Prospect Theory. In general, we do not have access to a direct measurement of U, so the utility function is only considered to be known up to some parameters which must be estimated. As a modeler, we postulate that the true utility experienced by people belongs to some family of functions, parameterized lets say by h; $U(x, h)$. EU and Prospect Theory are frameworks for mapping a family of utility functions $U(x, h)$ to the choices that people make when making decisions under uncertainty. These frameworks provide a probability that a person will make a given set of choices for a given h. When data has been acquired

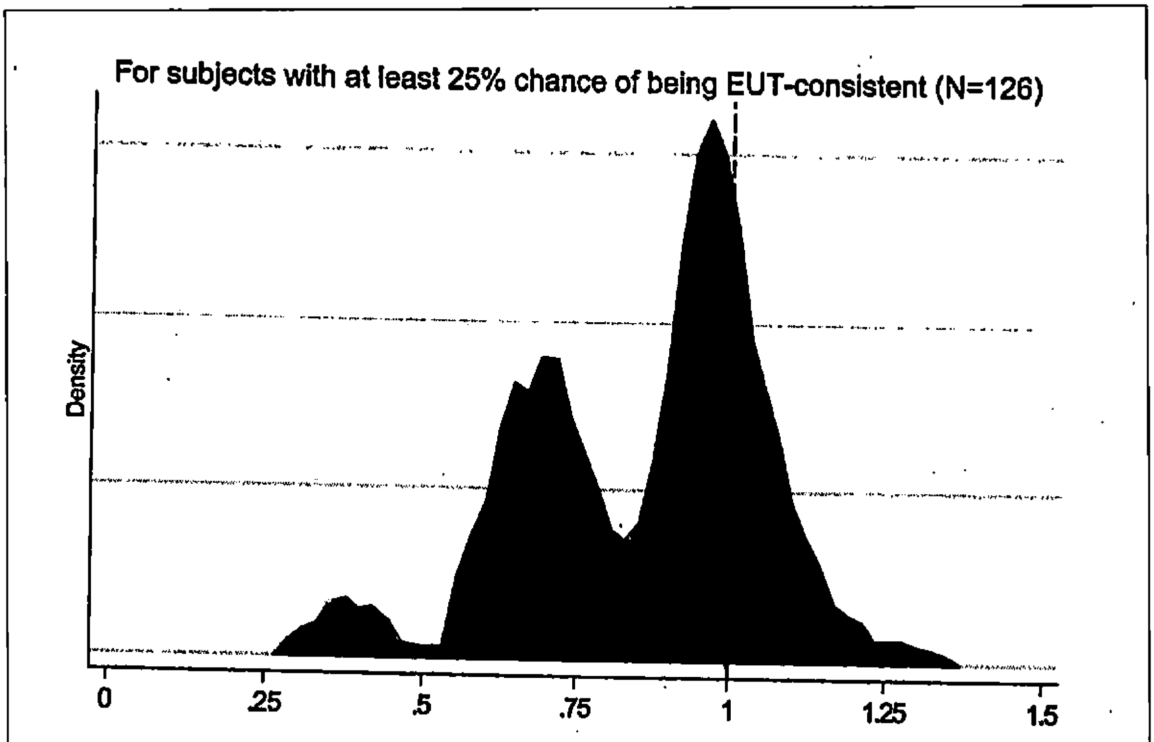

Figure 4.5: Constant Relative Risk Aversion parameter of the Expected Utility Theory Model

on the actual decisions of an actual person, we can then find the value h_0 of h that maximizes the probability that the person made the decisions it actually made. The function $x \to U(x, h)$ for this value of h_0 is then deemed to be the best description of the actual satisfaction U of the person among the family we started off with. The process of using data to find h_0, the h that best explains the data, is called inferring h_0, or (statistical) inference.

One desirable property of a utility function is that it be independent of the denomination of x, whether it be dollars or cents or yens, etc. This makes sense because the denomination shouldn't affect the modeling in any way other than through a re-adjustment of parameters. If a utility function $U(x, h)$ is not scale independent, changing the formulation from dollars to cents would change the family of functions we are considering, and would in general lead to a different value for the function $x \to U(x, h_0)$ once h_0 has been inferred. If the utility function is scale independent, changing the formulation from dollars to cents would change the value of h_0, but the resulting $U(x, h_0)$ would be the same, because the change in inferred h_0 would exactly compensate for the change in

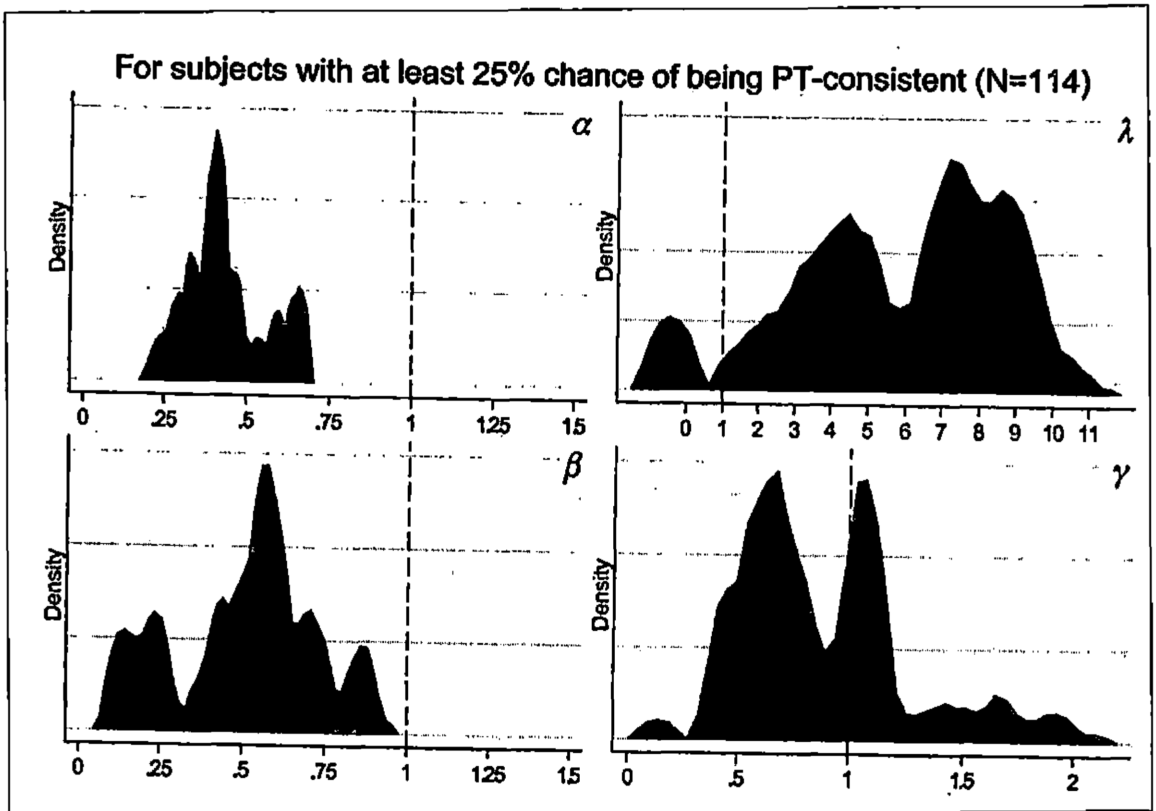

Figure 4.6: Parameters of the Prospect Theory Model

denomination.

In formal terms, a utility function $U(x, h)$ parameterized by h is scale independent if for all h and all re-scalings of the denomination $d > 0$, there exists h_1 such that $U(x, h) = U(d \times x, h_1)$. This is the formalization of exactly what was explained in the previous paragraph.

The utility function developed by Glenn W. Harrison, E. Elisabet Rutstrom in the wedding and funeral paper is not scale independent. They use $U(x, h) = x^r$, where the parameter $h = r$. For a given d different from 1, there exists no way to change r so as to compensate for changing x into $d \times x$. This is bad. A better utility function would include an additional scale variable g: $U(x, \{r, g\})$, so that $h = \{r, g\}$ of the following form: $U(x, \{r, g\}) = (x \times g)^r$. Now all one needs to do to compensate for a change in denomination d is to take $g = 1/d$. Note that this procedure of adding a parameter for scale is general, and can be done for any utility function to make it scale independent.

Parameter	Variable	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
r	Constant	−0.246	0.352	0.485	−0.940	0.449
	Female	−0.273	0.099	0.007	−0.469	−0.076
	Black	−0.043	0.195	0.824	−0.428	0.341
	Hispanic	−0.586	0.174	0.001	−0.929	−0.242
	Age	0.065	0.018	0.001	0.028	0.101
	Business	−0.104	0.099	0.291	−0.299	0.090
	GPAlow	0.042	0.083	0.616	−0.122	0.205
α	Constant	0.549	0.256	0.034	0.043	1.055
	Female	−0.201	0.215	0.350	−0.625	0.223
	Black	−0.101	0.216	0.640	−0.529	0.326
	Hispanic	−0.128	0.355	0.718	−0.829	0.572
	Business	0.053	0.292	0.857	−0.524	0.629
	GPAlow	0.059	0.196	0.765	−0.328	0.446
β	Constant	−0.202	1.199	0.866	−2.571	2.166
	Female	0.453	0.749	0.546	−1.026	1.933
	Black	0.300	0.291	0.303	−0.274	0.875
	Hispanic	0.172	0.504	0.733	−0.822	1.167
	Age	0.010	0.013	0.442	−0.016	0.036
	Business	0.130	0.451	0.774	−0.760	1.020
	GPAlow	0.095	0.175	0.587	−0.250	0.440
λ	Constant	1.592	7.164	0.824	−12.558	15.742
	Female	−4.007	10.037	0.690	−23.832	15.818
	Black	−4.494	2.029	0.028	−8.503	−0.486
	Hispanic	−5.083	2.053	0.014	−9.137	−1.028
	Age	0.523	0.566	0.357	−0.595	1.641
	Business	−2.981	2.226	0.183	−7.378	1.417
	GPAlow	−0.297	1.893	0.875	−4.036	3.441
γ	Constant	0.664	0.257	0.011	0.157	1.171
	Female	0.474	0.106	0.000	0.266	0.683
	Black	0.009	0.123	0.945	−0.234	0.252
	Hispanic	0.971	0.585	0.099	−0.185	2.127
	Age	−0.020	0.010	0.058	−0.041	0.001
	Business	0.333	0.180	0.065	−0.022	0.688
	GPAlow	−0.140	0.199	0.482	−0.533	0.253
$\kappa^{\dagger}$	Constant	0.558	1.268	0.660	−1.946	3.062
	Female	1.638	0.507	0.002	0.637	2.640
	Black	2.387	1.715	0.166	−1.001	5.775
	Hispanic	1.543	3.714	0.678	−5.793	8.880
	Age	−0.120	0.059	0.045	−0.237	−0.003
	Business	0.774	0.592	0.193	−0.395	1.943
	GPAlow	0.477	0.612	0.437	−0.732	1.686

† κ is the log odds of the probabilities of each model, where $\pi^{EUT} = 1/(1+\exp(\kappa))$

Table 4.2: Table of Estimates of all Parameters for Mixture Models with Individual Covariates

Chapter 5

Example Applications of Mixture Models

Mixture models have been used for decades in all sorts of contexts. This last section details some real world applications of finite mixture models in contexts other than the one considered by Glenn W. Harrison and Elisabet Rutstrom , [GH08]. After a general introduction, we review briefly an application used for the diagnosis of blood cells. We then review another application in astronomy. These overviews intend to reveal the general use of the finite mixture models in the real world. They also intend to introduce some basic practical concepts and difficulties only on the surface. We will then continue with two additional applications. These examples will further show how important concepts are used. The third application example will further review the determination of parameters in the making of diagnosis tools. The last application example will show how the important technique of Expectation Maximization (EM) algorithms can be used to obtain optimal values for a model under construction in field of marketing.

5.1 General Introduction and Preview Examples

Finite mixture models have been receiving increasing attention from both a practical and theoretical point of view (see [Sti86] Ch10). Modeling via finite mixture distributions involves identifiability problems, the actual fitting of finite mixtures through use of the Expectation Maximization (EM) algorithm, the properties of maximum likelihood estimators, the assessment of the number of components to be used in the mixture, and the applicability of asymptotic theory in providing a basis for the solutions. Scaling of the EM algorithm allows mixture models to be used in data mining applications involving massively large data bases. Also, recent use of t components in the mixture model provides a robust approach to mixture modeling. A t number of components would mean the number of subpopulations, and therefore the number of likelihood functions that are supposed to model each subpopulation is not known in advance. Some latest developments in finite mixture distributions involve hidden Markov models. Successful application fields of mixture models include astronomy, biology, genetics, medicine, psychiatry, economics, engineering, and marketing [Eve96]. Moreover, finite mixture models, provide cluster and latent class analyses, discriminant analysis, image analysis, and survival analysis. Thus, mixture models can also reveal the existence of previously unrecognized or undefined subpopulations, or substructures, as we will see below in Kriessler and Peers astronomy discoveries (see [GM01], Chapter 6). In fact, in some applications of some mixture models, there is sufficient a priori information for the number of components g in the mixture model to be specified with no uncertainty. However, on many occasions, the number of components has to be inferred from the data, along with the parameters in the component densities. For instance, the red blood cell volume distribution of healthy individuals can be modeled adequately by a single log normal component. However, for patients not completely recovered from anemia, their red blood cells distribution, although unimodal in appearance toward the end of the iron therapy treatment, may still need to be modeled by a two-component log normal mixture due to the presence of a sufficient number of microcytic cells in relation to the normocytic cells. Thus the result of a statistical test on the number of components in the log normal mixture model for a specific patient can be used as an early guide to aid clinicians in making a decision when to suspend iron therapy treatment for a patient. A non-significant test result is consistent with the red blood cell distribution of the patient having returned to a healthy state. For an application of insufficient a priori information for the number of components, let's consider the recent discoveries in astronomy. As explained by Kriessler and Beers (1997), it was once assumed that most clusters of galaxies (subpopulations) were relaxed systems that could be adequately modeled by a simple set of parameters, such as a single-core radius and the velocity dispersion of neighboring galaxies. However, numerous recent studies have concluded that many, perhaps even most, clusterings are far from being in dynamical equilibrium. Evidence cited includes the existence of: (a) clumpy distributions of galaxies seen in the projection on the sky, (b) apparent structure in the distribution of radial velocities for cluster membership, and (c) multiple centers of X-ray-derived temperature profiles, suggestive of ongoing collisions. The desire to identify substructure in clusters of galaxies has led to the bootstrap form of the Likelihood Ratio Test (LRT) for the number of components in a mixture model being applied in studies in astronomy. Bootstrap is a powerful technique that permits the variability in a random quantity to be assessed using just the data in hand. Recent papers in the astronomical literature in which this method has been used for investigating substructure in galaxy clusters include those by Ashman and Bird (1993, 1994), Bird (1994a, 1994b, 1995), Beers and Sommer-Larsen (1995), Bird, Davis, and Beers (1995), Davis et al. (1995), Zepf, Ashman, and Geisler (1995), and Bridges et al. (1997).

For example, Kriessler and Beers (1997) concluded from their use of this statistical test that 57% of the Dressler (1980) morphological-sample clusters have statistically significant substructure. Figure 5.1, which is taken from Kriessler and Beers (1997), gives the contour plots of bivariate normal mixtures fitted to the positions of some of the galaxy clusters (see [GM01] page 197). Figure 5.1 exposes the adaptive-kernel density contour maps of galaxy positions in Dresslers morphological sample. The filled circles in the figure indicate the positions of galaxies identified by Dressler (1980). The crosses mark the average positions identified as significant in the normal mixture fit.

5.2 Medical Diagnosis

Let us continue with an example of using a mixture model to make good use of medical data, in view of predicting the chances that given individuals have certain diseases. Screening for various diseases based on a patient's medical history and data can be hard even for the trained physician because of the number of factors that need to be taken into account [Sti86].

Suppose we have a data set from a large number N of patients who were each found to have one of 5 conditions, one of which is the "healthy" condition, and four of which are pathological; we also assume that patients cannot have multiple conditions, so patient i is associated with a single known condition $J(i)$. For each patient, we have a set of 16 standard medical measurements: blood pressure, age, weight, etc. Each patient i is characterized by a vector x_i of 16 numbers, and by their known condition $J(i)$. We would like to use this data to form a model which will allow us to assess the risk of disease for future patients based on their 16 standard measurements.

Formulated as such, this is a relatively easy problem which maps nicely into the mixture model setting. We can assume that each of the 5 groups of people in the 5 conditions has a distinct distribution of the 16 medical measurements. We will suppose that patients have probability p_j of having condition j, for $j \in \{1,2,3,4,5\}$, and that each condition j gives rise to a Gaussian distribution of medical measurements with mean vector $\mu_j \in \mathbb{R}^{16}$, and covariance matrix $C_j \in \mathbb{R}^{16*16}$, so that the probability density of a patient with condition j having medical measurements x are given by the Gaussian density:

$$p(x_i|J(i)=j) = \frac{1}{\sqrt{|2\pi C_j|}} e^{-\frac{1}{2}(x_i-\mu_j)^T C_j^{-1}(x_i-\mu_j)}$$

It is now possible to infer the parameters $\{(p_j, \mu_j, C_j)\}_{j\in\{1,2,3,4,5\}}$ of this mixture model, because we know in advance which patients belong to which of the 5 mixture components. The parameters p_j are simply obtained from the number of patients N_j in condition j, as the empirical average number of patients having that condition. The parameters μ_j and C_j for patients with condition j are obtained the way parameters of a gaussian are usually estimated: equivalently by maximum likelihood or with the empirical estimators of the mean and covariance of the x_i's:

$$p_j = N_j/N$$

$$\mu_j = \frac{1}{N_j} \sum_{i,J(i)=j} x_i$$

$$C_j = \frac{1}{N_j - 1} \sum_{i,J(i)=j} (x_i - \mu_j)(x_i - \mu_j)^T$$

With these three simple equations, we have estimated all the parameters of the model. Now given a new patient with medical measurements x, we can calculate their probability of having condition j by calculating $p(j|x) = p(x,j)/p(x)$ where $p(x,j) = p(x|j)p(j) = p(x|j)p_j$ and $p(x) = \sum_{j'} p(x,j')$:

$$p(j|x) = \frac{p(x|j)p_j}{\sum_{j'} p(x|j')p_{j'}},$$

$$p(j|x) = \frac{\frac{p_j}{\sqrt{|2\pi C_j|}} \exp\left[-\frac{1}{2}(x_i - \mu_j)^T C_j^{-1}(x_i - \mu_j)\right]}{\sum_{j'} \frac{p_{j'}}{\sqrt{|2\pi C_{j'}|}} \exp\left[-\frac{1}{2}(x_i - \mu_{j'})^T C_{j'}^{-1}(x_i - \mu_{j'})\right]}.$$

This would ultimately mean that, given the medical characteristics of an individual represented by the vector x, we can calculate the probabilities for the individual to be affected by each one of the five health conditions j. Thus the highest probability associated with one of the five health conditions j would be the one the doctor might decide to worry about.

5.3 Consumer Spending

Different consumers choose to spend their money differently. It is reasonable to consider that consumers might fall into groups with similar tastes and spending habits. Having a model of the distribution of consumers, how they cluster into groups, and what the spending habits of each group is could be useful in several business contexts. For example, we could be a business scouting out potential consumer markets, in view of designing a new product targeted for a particular group of consumers. Or we could already have a product which we would like to market to a suitable group of consumers.

Suppose we have data on the spending habits of thousands of people. We know the amount that was spent by each person in each of 12 categories of spending: groceries, clothing, restaurants, automobile, air fares, other transportation, insurance, rent, education, etc. Each person is represented by the 12-dimensional vector of amounts they spent in each category.

If there were only 2 spending categories, we could just make a 2-dimensional scatter plot, with the 2 axes being the amount spent in each of the 2 categories, and one point for each person surveyed. We would then look at this cloud of points and hopefully see that these points cluster into groups, which we could identify as distinct consumer groups. However we have 12 categories instead of 2, so visualizing clusters in this way is impossible.

In this context, mixture models are a good way of inferring clusters of consumers as well as the distribution of their spending habits, even if we cannot do so "by eye". Let's see how this can be formalized, and what the difficulties are in applying mixture model methods.

This new example application of Gaussian mixture models presents a little more difficulty than the previous medical one, because we don't know which cluster each consumer belongs to in advance; we don't even know how many clusters there are. As it turns out, it is still possible to make progress in this situation.

In the medical example, we saw that if we know which cluster each patient belonged to, we could calculate all the model parameters in a straightforward manner. Then given the model parameters, we saw how to calculate the probabilities of each patient belonging to each condition. What we can do in the case where we don't know in advance which cluster each person belongs to is as follows: we can guess how many clusters there should be, and we can start with a random initial guess of the model parameters. Then for these parameters, we can calculate the probabilities of each person belonging to each cluster [GM01]. Using these probabilities, we can estimate the model parameters. Given these new model parameters, we can estimate the probabilities of each person belonging to each cluster once again, and using these new assignments, we can re-estimate the parameters once again. In this way, we can alternate between re-estimating parameters and re-estimating the probabilities of each person belonging to each cluster, and hope that at each iteration, the parameter estimate will get better and better.

As it turns out, this procedure, called Expectation Maximization or the EM algorithm, is guaranteed to improve the model parameter estimates after each iteration, and converge to an estimate of these parameters which locally maximizes the likelihood of the data. Since we are estimating parameters in the maximum likelihood setting, we can see parameter estimation as an optimization problem: we are looking for the parameters which maximize the likelihood. We can view the likelihood function as a landscape with hills and valleys, in the high-dimensional space where parameters live: finding the best parameters consists in finding the highest peak in this landscape. The EM algorithm is guaranteed to find a peak in this landscape. However, it is not guaranteed that the peak it finds is the highest peak in the landscape; this is why we say that EM converges to a local maximum, as opposed to a global maximum. Starting EM with different initial guesses for the parameters will result in converging on different peaks in general.

This is particularly a problem for Gaussian mixture models, because the Gaussian components of the mixture can become unboundedly sharp around a single data point, and this would make the likelihood of the data become infinite: this happens when one of the mixture components j has a mean equal to one of the data points, and the covariance matrix C_j of that component goes to zero. For a Gaussian mixture model, there are in general many peaks, a lot of which are actually infinite, and which we would like to avoid, because they are spurious: they are overfitting single points in the data. There are various ways of dealing with this problem, which we will not get into.

However in practice, given a decent initial guess for the model parameters, applying the EM algorithm will most often converge onto a model which is reasonable and useful [GM01].

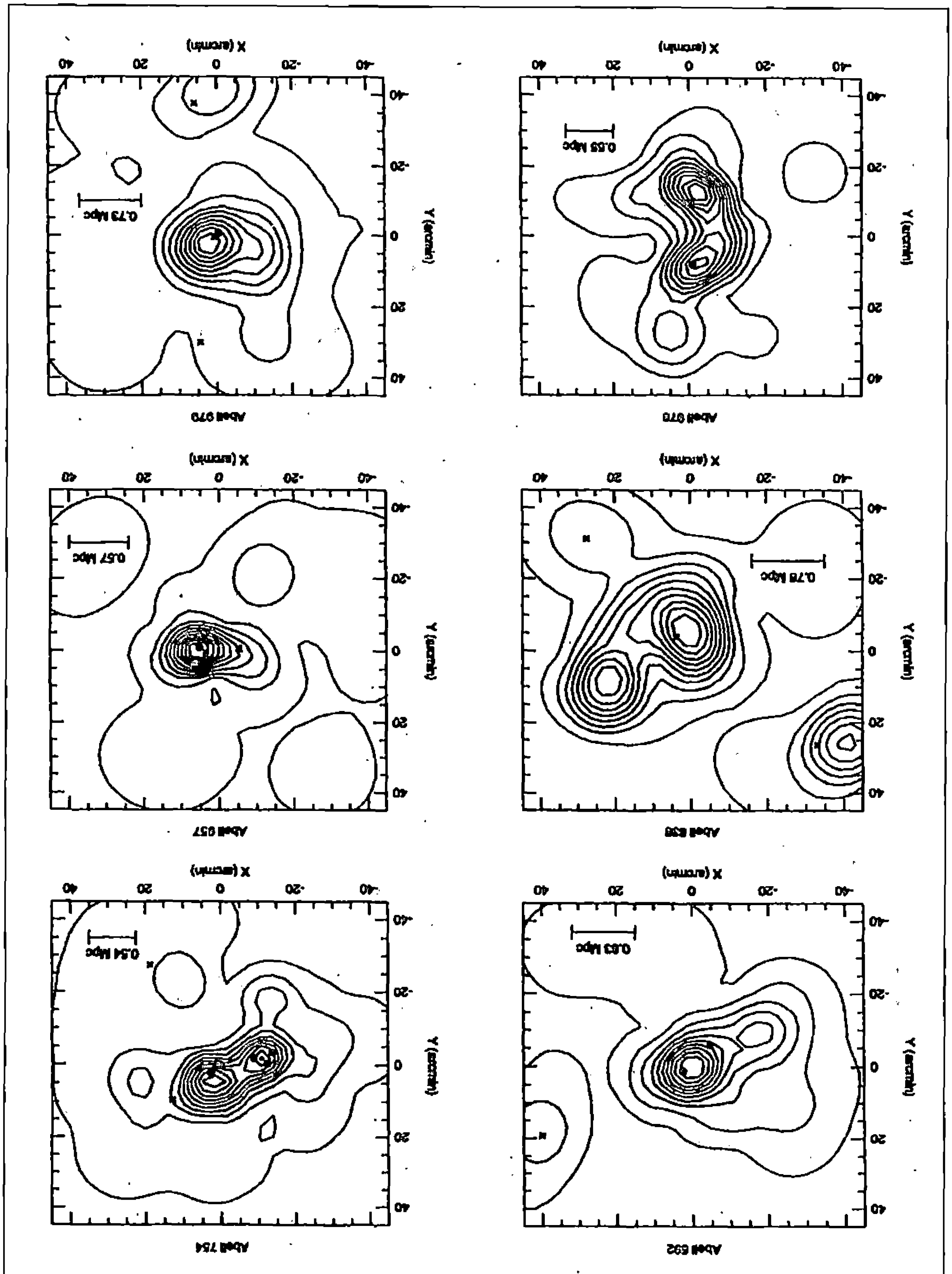

Figure 5.1: Adaptive-Kernel Density Contour Maps of Galaxy Positions in Dressler's Morphological Sample

Chapter 6

Conclusion

In this paper we have introduced the decision making contexts in order to better understand the two major decision making making theories, the Expected Utility Theory, the Prospect Theory, and their applications. We then, investigated the finite mixture models using these two theories within a larger likelihood function. The application example of the finite mixed model containing these two theories as their major ingredients paved the way for approaching more application examples in Chapter 5. These examples included discussions such as identifiability problems, fitting of finite mixture models through the EM algorithm, and construction of maximum likelihood functions. The finite mixture models are increasingly used in all areas of science. This project gave me the opportunity to discover some of the major ways mathematics contribute to improve our lives. The understanding, the curiosity, and the knowledge that I have built through this project add to the excitement that I enjoy as a life time learner. Also, I am so grateful to have had the opportunity to share this enrichment as a graduate student of CSUSB.

Bibliography

[Cen06] Experimental Economics Center. *Decision Making Under Uncertainty.* http://www.econport.org/econport/request?page=man_ru_advanced_prospect, WWW, 2006.

[DK79] A. Tversky D. Kahneman. Prospect theory: An analysis of decision under risk. *The Economic Society, Econometrica, Vol. 47, No. 2*, March 1979.

[Eve96] B. Everitt. *An Introduction to Finite Mixture Distributions.* Biostatistics and Computing Departmen, Institute of Psychiatry, London, 1996.

[FH05] J. Liberman F. Hillier. *Introduction to Operations Research.* Mc Graw Hill, New York, 2005.

[GH08] E. Rutstrom G. Harrison. Expected utility theory and prospect theory: one wedding and a decent funeral. *Economic Science Association*, May 2008.

[GM01] David Peel Geoffrey McLachlan. *Finite Mixture Models.* John Wiley and sons, Inc., New York, NY, 2001.

[MA07] C. Paraschiv M. Abdellaoui, H. Bleichrodt. Loss aversion under prospect theory. *Management Science, Vol. 53, No. 2*, October 2007.

[NB07] J. Pinto-Prades N. Bleichrodt, J. Abellan. Resolving inconsistencies in utility management under risk: Tests of generalizations of expected utility. *Management Science*, March 2007.

[RT75] R. Klekamp R. Thierauf. *Decision Making Through Operations Research.* John Wiley and Sons, NY, 1975.

[Sti86] Stephen M. Stigler. *The History of Statistics.* The Belknap Press of Harvard Press, Cambridge, Massachusetts, 1986.

[Wik09] Wikipedia.org. *Prospect Theory.* http://en.wikipedia.org/wiki/Prospect_theory, WWW, 2009.