UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXV

Multivariate Preference Models for Scoring or Ranking Units

Direttore della Scuola: Prof. Monica Chiogna

Supervisore: Prof. Luigi Fabbris

Co-supervisore: Dott. Giovanna Boccuzzo

Dottoranda: Manuela Scioni

31/01/2013

Acknowledgements

First of all, I would like to thank my supervisor Professor Luigi Fabbris for the advices, the patience and all the opportunities he gave me during the last three years.
I am very grateful also to my co-supervisor Dott.ssa Giovanna Boccuzzo for her continuous help and all her suggestions.
A special thank goes to my family, for always believing in me and for supporting always my decisions. I would have never had the possibility to write this thesis without them so I would like to dedicate this work to them.
I would also like to thank my fiancè Federico for supporting me in the last 6 years, for his love and his precious suggestions. A special thank also to his parents, Franco and Rosella, for always make me feel at home.
Thanks to my colleagues, now special friends, of the XXV cycle Elena, Marina, Clovis, Enrico, Federico and Ricardo. It has been a pleasure to share this great experience with them and I will always remember the great fun we had during the first year.
I would like to thank all of the other PhD students and post-doctoral guys for trying to keep me calm before each due date.
Thanks to all my friends living in Genova, Padova, Pinerolo and Valdagno, for their support, remote and not.
Finally, a loving thougth to my godchildren Giulia and Lorenzo, for the smiles they always give me.

Abstract

Measuring individuals' preferences for goods and services has recently obtained considerable attention in both public and private contexts. Individuals' judgments are used for many different purposes, including setting social policies and forecating the acceptance of a new product in the market. While preference modeling is a long-studying problem, modern applications, related to the web, make it an actual topic. Respondents are called to express their preferences among a set of alternatives and collected data can be represented in various kinds of matrices. This thesis is focused on some popular methods to estimate either scores or ranks of a set of alternatives by analyzing a generalized tournament matrix. The proposed methods are compared via simulation and some special situations are investigated to detect their reliability. Our aim is to compare methods that assume parametric hypotheses on data distribution with methods that do not require such hypotheses. When respondents do not compare directly two alternatives, the matrix representing their preferences may show one or more missing values. We propose a method to estimate the missing entries of a generalized tournament matrix based on the minimization of the sum of its singular values, i.e. the nuclear norm. We perform some simulation studies to investigate the nuclear norm minimization effectiveness.

Sommario

Conoscere le opinioni e le preferenze degli individui su beni o servizi, ha da sempre rivestito notevole importanza, in contesti sia pubblici che privati. Le preferenze delle persone vengono, infatti, rilevate per diversi scopi, come il definire nuove politiche sociali o il valutare se un nuovo prodotto potrà essere recepito sul mercato. Le applicazioni moderne dell'analisi delle preferenze, connesse al web, lo rendono un argomento attuale. Il punto di partenza l'espressione da parte di un campione di individui delle proprie preferenze in merito alle possibili alternative di un insieme. I criteri per rilevare le preferenze sono numerosi. In questa tesi si presentano alcuni metodi per stimare i punteggi o i ranghi delle alternative partendo da un matrice con struttura di rilevazione a torneo generalizzata. Vengono realizzate alcune simulazioni allo scopo di confrontare i metodi proposti e di investigare alcune situazioni particolari utili a verificarne la affidabilità. L'obiettivo della tesi è di confrontare metodi che assumono ipotesi parametriche sulla distribuzione dei dati e metodi che non richiedono ipotesi. Nella tesi, inoltre, si propone un metodo per stimare un dato non validamente espresso in una matrice di torneo generalizzata. Il metodo si basa sulla minimizzazione della somma dei valori singolari, vale a dire la norma nucleare, della stessa matrice. Inoltre, sono effettuate simulazioni allo scopo di analizzare l'efficacia del metodo di stima basato sulla minimizzazione della norma nucleare.

Contents

Chapter 1

Introduction

1.1 Overview

Preference elicitation is a common issue in many different disciplines, such as economics, sociology, political science and psychology. To elicit preferences, groups of individuals are called to express their preferences among a set of alternatives. While preference modeling is a long studying problem, modern applications, related to the web, make it an actual topic.

There are several methods to elicit preferences: ordering the alternatives, picking up one or more out of them, rating the alternatives, allocating a budget among them and comparing them in pairs (Fabbris, 2011). According to the latter method, alternatives are presented in pairs to one or more judges, in a random sequence (Kendall & Babington-Smith, 1940). The judge can choose either one, declare a tie, or express his or her preference on some scale. The method of paired comparisons presents some advantages with respect the other tecniques, since most people cannot evaluate many alternatives at a time. Moreover, in case of very similar alternatives, compare them in pairs helps to express a preference.

The data, collected through all methods can be represented, sometimes after

a pre-treatment, in a generalized tournament matrix $\boldsymbol{P}$, defined as follows (Moon & Pullman, 1970).
Let $A = \{A_1, A_2, \ldots, A_k\}$ be a set of k alternatives, $\boldsymbol{P} = [\pi_{ij}]$ is a $(k \times k)$ matrix that satisfies

$$\boldsymbol{P} + \boldsymbol{P^T} = \boldsymbol{J} - \boldsymbol{I}, \qquad \boldsymbol{P} \geq 0.$$

In this case π_{ij} represents the probability that A_i is preferred to A_j and it can be estimated through p_{ij}, the proportion of comparisons in which A_i is preferred to A_j, provided there are no ties. Hence, $p_{ij} = 1/2$ indicates indifference between A_i and A_j, $p_{ij} = 1$ indicates that A_i is always preferred to A_j and $p_{ij} > 1/2$ indicates that A_i is preferred to A_j.
Another kind of matrix, widely used in literature (Saaty, 1977, 2008), is the multiplicative paired comparison matrix $\boldsymbol{M} = [m_{ij}]$ in which $m_{ij} \in]0, c[$ represents the preference ratio of A_i over A_j: $m_{ij} > 1$ implies that A_i is strictly preferred to A_j, whereas $m_{ij} < 1$ expresses the opposite preference and $m_{ij} = 1$ means indifference.
Starting from a generalized tournament matrix $\boldsymbol{P}$ it is possible to construct a ranking of the alternatives or to define a set of weights that reflects their relative importance. There are several methods in literature for these purposes such as the score-vector (Kendall, 1955; David, 1987; Thurstone, 1927b), the eigenvector method (Kendall, 1955; Wei, 1952) and the linear models (David, 1988).
The score vector $\boldsymbol{\omega^{(1)}}$ is the vector of row sums obtained by

$$\boldsymbol{\omega^{(1)}} = \boldsymbol{P} \cdot \boldsymbol{e}$$

where $\boldsymbol{e}$ is a column vector of all 1's. According to this method the alternatives are ranked in the corresponding order. For a generalized tournament matrix the i-th row sum can be interpreted as the expected score of A_i.
The eigenvector method is based on the eigenvalues and eigenvectors associated to the generalized tournament matrix. Suppose $\boldsymbol{P}$ is a positive irreducible matrix, then the Perron-Frobenius theorem guarantees the existence

of an unique large real eigenvalue λ_1 as solution of the characteristic equation

$$\boldsymbol{P}\boldsymbol{\omega} = \lambda_1 \boldsymbol{\omega}$$

whose eigenvector $\boldsymbol{\omega}$ has strictly positive components (Keener, 1993). $\boldsymbol{\omega}$ is a unique solution of the characteristic equation provided $\|\boldsymbol{\omega}\|_2 = 1$. Its i-th entry is assumed to represent the relative importance of the i-th alternative with respect to the others.

The linear paired comparison model (David, 1988) assumes that the k alternatives can be represented along a linear continuum. Each of them has true rating V_i which determines its position in such representation. The rating of the i-th item will vary from respondent to respondent and it may be represented by Y_i a continuous variable with mean V_i and distribution called "sensation distribution".

The Y_i's $(i = 1, \ldots, k)$ are independent and identically distributed variables with the same variance. Defining $Z_i = Y_i - V_i$, we have that $Z_i - Z_j$ is a symmetric variable with mean 0 and distribution, named defining distribution, that depends on the distribution of Y_i.

Under these assumptions, the probability of preferring A_i to A_j can be expressed as follows

$$\pi_{ij} = F_{Z_i - Z_j}(V_i - V_j) \qquad (i, j = 1, \ldots, k)$$

Hence, to estimate the ratings V_i's, we have to solve the linear system

$$d_{ij} = \hat{V}_i - \hat{V}_j = F^{-1}_{Z_i - Z_j}(p_{ij})$$

with k unknown parameters and $k(k-1)$ equations. If $k > 3$ the system results overdetermined and we need to impose some contraints to solve it.

Among the others, the Thurstone & Mosteller model (Thurstone, 1927a; Mosteller, 1951), assumes that variables $Z_i - Z_j$ have a normal distribution and the Bradley & Terry model, assumes a logistic distrubution (Bradley & Terry, 1952).

Suppose now that the generalized tournament matrix $\boldsymbol{P}$ is incomplete, that is some of its entries are unknown. This happens for different reasons: for

example, if the number of alternatives is large and it is not possible to submit all the possible pairs to the respondents. The topic of missing values in a preference matrix has been extensively discussed in literature (Harker, 1987; Carmone et al., 1997; Kwiesielewicz & Van Uden, 2003; Fedrizzi & Giove, 2007; Dittrich et al., 2012).

When a respondent does not compare directly two alternatives, A_i and A_j, the linear models permit to estimate the preference relation between them using the information achieved from their comparison with a shared alternative A_l. In practice, linear models allow to obtain p_{ij} through p_{il} and p_{lj} for $i, j, l = 1, \ldots, k$ and $i \neq j \neq l$. In particular, if we assume that $F^{-1}(p)$ exists and is unique for $p \in (0, 1)$ then (Latta, 1979)

$$\begin{aligned} p_{ij} &= Pr\left(Y_i \geq Y_j\right) = F(V_i - V_j) \\ &= F(V_i - V_l + V_l - V_j) \\ &= F[F^{-1}(p_{il}) + F^{-1}(p_{lj})]. \end{aligned}$$

1.2 Main Contributions of the Thesis

In this thesis we compare the above-introduced methods proposed for ranking and scoring a set of alternatives. Starting from a generalized tournament matrix $\boldsymbol{P}$, the comparison is performed via simulation and some special situations are investigated to detect the reliability of the methods. The aim is to compare methods based on distributional parametric hypotheses with methods that do not require such hypotheses.

When respondents do not compare directly two alternatives, the matrix designated to represent their preferences is characterized by one or more missing values. In this thesis, we propose a method to estimate the missing probabilities of a generalized tournament matrix $\boldsymbol{P}$ based only on its valid entries. The main assumption is that only a few dimensions contribute to individual preferences, which corresponds to state that $\boldsymbol{P}$ can be well approximated by

a lower-rank matrix. So, we search for the matrix completion that minimizes its rank; intuitively, we seek for the simplest completion values that fit the observed data. Since rank minimization problems result to be unsolvable, it is possible to use a recent heuristic that minimizes the sum of the singular values, i.e. the nuclear norm (Fazel et al., 2001).
In literature there are articles on matrix completion based on the minimization of the nuclear norm (Candès & Recht, 2009; Gleich & Lim, 2011), especially for lower-than-full rank matrices, whereas generalized tournament matrices are of full rank.
Finally, we perform some simulation studies to compare nuclear norm minimization with linears model composition rules.

This thesis consists of six chapters. In Chapter 2 we present a general background on preference modeling, in particular on elicitation preference methods. The main focus is on the paired comparison method and on the matrices used to represent paired comparison preference data. Chapter 3 describes some methods for ranking or scoring alternatives. In Chapter 4 we present the results of the simulation studies to compare the proposed ranking and scoring methods. Chapter 5 is focused on the estimation of missing values within a generalized tournament matrix. Thus composition rules associated to linear models are presented and a procedure based on the nuclear norm minimization is investigated. In Chapter 6 some concluding remarks are drawn.

Chapter 2

Background

Measuring individuals' preferences for goods and services has recently obtained considerable importance for both public and private contexts as a key element in the decision-making process. Individuals' judgments are used for many different purposes, including setting social policies and evaluating the acceptance of a new product in the market. To give some examples, eliciting public preferences for healthcare results to be fundamental in allocating resources across competing services. Indeed, given the limited availability of resources, public opinions are recognized to be a fundamental criterion for their allocation (Kassirer, 1994; Ryan et al., 2001). Another application comes from electronic commerce. To increase business success and customer loyalty it is necessary to offer consumers personalized services. To this end, it is essential to understand individual customers' preferences for products and recommend them the most appropriate ones (Schafer et al., 2001; Devaraj et al., 2002). Therefore, politicians want to know voters' opinions; companies want to know consumers' preferences and people in general wants to know what others think about political, social, health and other issues.

There are two main paths to eliciting preferences: either from stated or from revealed preference analysis (Train, 2003). Stated preferences are choices that individuals would make from a hypothetical choice set. So, if we want respondents to state their preferences, we can simply ask them. There is

a slight difference between choice and preference. Indeed, a choice applies to expressions such *"I choose that option"* while a preference applies to *"I prefer this instead of that"* (Fabbris, 2013). However, in this thesis the two terms are used without distinctions because our analyses apply to both.
Revealed preferences, on the other hand, are implicit in individuals' choice actions. To reveal respondents' preferences, we have to observe their actual choice behaviors. For example, if in a survey we ask a respondent which car he would buy among a set of three cars, we are eliciting his stated preferences. On the other hand, if we ask him which car he bought when he last bought a car, we are revealing his preferences.
Economists generally prefer to analyze revealed preferences. Psychologists and other social scientists, instead, use mainly stated preference data. In the following, we refer to stated preferences.
Questionnaires represent the principal mean to collect individuals' opinions and preferences. Before expressing their preferences among a fixed set of k alternatives individuals have to do some mental exercises. First of all they are implicitely required to create in their own mind a measurement scale based on their values and social rules. Then, they try to put the alternatives, also called items or stimuli, on the derived ordinal or cardinal scale and, finally, they express their choices. During this process respondents may encounter difficulties. Sometimes, for example, they cannot express a preference among two alternatives, as they are too similar, or they may not be able to construct a measurement scale because the question put has no real meaning to them.
Preference data collected from n sample units can be used to estimate either scores or ranks for the alternatives in the choice set. Ranking procedures simply return the alternatives in order of importance. Scoring techniques, instead, assign values to the alternatives, according to a convenient scale. Ranks can be adopted to define priorities among the choice set, to show hierarchies among items, or to point out if an alternative improves its position with respect to a previous ranking. Scores may be of interest when quantitative and precise values are needed, for instance, to allot resources among

the units of the choice set (Fabbris, 2013).

2.1 Methods for Eliciting the Preferences

The questions used to elicit preferences are usually closed questions, with predetermined answers, and respondents express their preferences within the given set. There are several techniques apt to elicit preferences among a fixed set of alternatives. We will examine the following techniques, trying to emphasize their strengths and weaknesses: ranking, picking the best alternative, rating, partitioning a constant sum among alternatives and comparing alternatives in pairs.

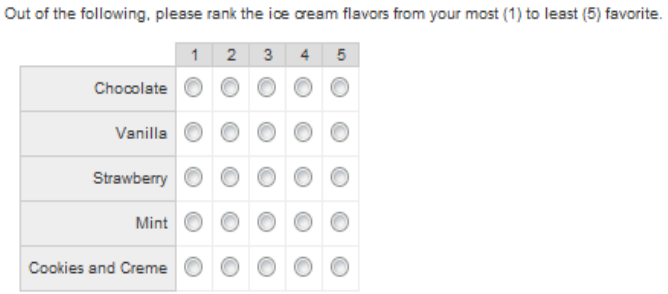
2.1.1 Ranking

According to this method, respondents are asked to rank the alternatives in order of importance. The problem with this method, however, is that the more numerous the alternatives in the choice set, the more difficult it is for the respondent to answer (Inglehart & Abramson, 1993). Indeed, this procedure requires that each respondent evaluates all the items of the set simultaneously. Ties may be allowed if the researcher perceives that the task required is too difficult for respondents (Fabbris, 2013).
A typical ranking question asks the respondent to attribute to each of the k alternatives a number from 1, for the most preferred alternative, to k, for the least preferred one (See Fig. 2.1).

2.1.2 Pick Up one or more Alternatives

Respondents can be asked to select their one or two most preferred alternatives out of the list. This method requires an effort similar but not as extensive as the ranking technique. This is the simplest way for respondents to express their preference concerning a set of alternatives. However, no

Figure 2.1: Ranking question example



information regarding the relationship among the non-selected alternatives is derived (Sato, 2004). This method is simple and fast, even in presence of many alternatives. Among the ones we present, picking up one or more alternatives is the procedure most similar to choice.
Fig. 2.2 shows a possible question of this type, always referred to the example presented in Fig. 2.1.

Figure 2.2: Picking question example

2.1.3 Rating

According to this method, respondents are asked to rate each alternative using a vote from 1 to a fixed number. Ideally, a rating scale should consist of enough points to extract the necessary information. Odd numbers of points have generally been preferred to even numbers because they allow the middle category to be interpreted as a neutral point. In a literature

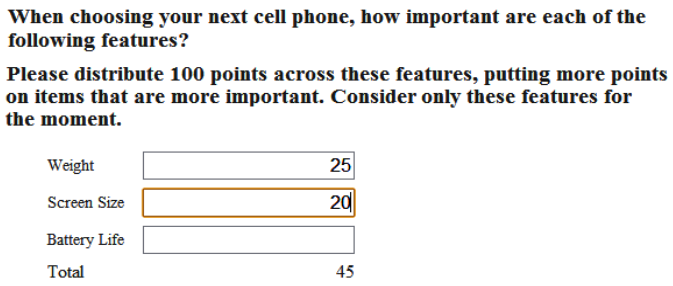
Figure 2.3: Amazon rating system

review, Cox (1980) concluded that there is no single number of points for a rating scale that is appropriate for all situations. Some researchers claim that scales consisting of three points are sufficient (Matell & Jacoby, 1971). However, a 5 option system results often better than a 3 option system, since it makes possible more accurate predictions (Churchill et al., 1984). For instance, the five points scale is the most used method in web applications like Amazon (See Fig. 2.3), Youtube or Netflix. Friedman and Friedman (1986) concluded that researchers should consider using anywhere from 5- to 11-point scales. A rating scale can be either unipolar or bipolar. Rating alternatives method is not particularly affected by the increasing size of the choice set and it presents the advantage that the scores can be treated roughly like cardinal measurements. It is characterized by low discriminatory power, indeed it leads to less differentiation among items, with the possibility that a respondent rates every item identically.

2.1.4 Constant Sum Question

Another possibility to elicit preferences is to ask respondents to allocate a budget across the given set of alternatives, up to a cumulative maximum number of points, usually 100. These points are distributed so as to reflect the relative importance of the alternatives, revealing the relative difference between them. Figure 2.4 is a scheme of a question of this kind.
A constant sum question may be useful to differentiate the preference on the alternatives. Moreover, it allows respondents to assign 0 importance to

Figure 2.4: Constant sum question



items, which is not possible in other preference elicitation methods. This method is not reccomended in case of a large choice set, since it requires long time and large mental energy from respondents. A computer-assisted interviewing system allows to check the sum of points cumulated after each assignment (Fabbris, 2013). Conrad et al. (2005) found that running totals improve the likelihood to reach the desired sum and take the respondents less time.

2.1.5 Paired Comparisons

In the method of paired comparisons, the k alternatives are presented in pairs to respondents (Kendall & Babington-Smith, 1940). To control the order effect, it is important to randomize the presentation order of the pairs as well as the order of items within each pair (Bock & Jones, 1968). This is the reason why computer-assisted interviewing systems promoted the use of paired comparisons methods.

According to this procedure, a judge can choose either one, declare a tie, or express his or her preference on some scale. The situation in which every respondent performs every possible paired comparison is called "balanced paired-comparison experiment" and corresponds, in sports terms, to a Round Robin Tournament (David, 1988).

Thanks to the paired comparisons method, it is possible to elicit preferences based on an attribute that can only be subjective as taste. The judgement process is simplified, since respondents have to judge a pair of objects at a

time. Especially when differences between choice alternatives are small, these methods provide more information with respect to rating methods, since the paired comparison between the alternatives is as free as possible from extraneous influence. Moreover, it is possible to identify respondents with not well-defined preferences, through internal consistency checks. If respondents are consistent in their judgments, the researcher can have much greater confidence in them also for further applications.

The main problem of paired comparison method is that every respondent has to compare, for k alternatives, $k(k-1)/2$ distinct pairs of items. For this reason, if k is large, reduced forms of paired comparisons are available, as the incomplete tournament technique proposed by Fabbris and Fabris (2003). This procedure involves ordering the alternatives according to a criterion, submitting for choice firstly the $k/2$ pairs of adjacent alternatives, then the $k/4$ pairs of alternatives preferred at the first level and so on until the most preferred alternative is sorted out.

The literature on paired comparison modeling is vast and spans various fields. In the following we give some references on the origins of this popular method. Paired comparisons techniques have been widely used by psychometricians. The method was introduced by Fechner (1860; 1966), a German experimental psychologist, who firstly described and implemented the concept of paired comparisons in his book "Elemente der Psychophysik". Fechner suggested that choice behavior can be considered as a probabilistic phenomenon, since there exists a probability that a person makes a choice rather than another. Moreover, he stated that this probability may not equal 1 or 0, as it is often observed that a person repeatedly presented with the same pair of alternatives will not always make the same choice.

Paired comparisons method was made popular by L.L. Thurstone (1927a), a prominent psychometrician, who first introduced a scientific approach to using pairwise comparisons for measurement. In its paper "The law of comparative judgment" he provided a method for ordering alternatives along a continuum (Edwards, 1983). As Fechner, Thurstone (1927c) considered

choice behavior to be a probabilistic rather than a deterministic process; he used the normal distribution to estimate the model parameters. A judge may provide different judgements about the same object on different occasions; that is, the judgment process does not always provide the same value on a psychological continuum (Luce, 1959). Although Thurstone focused initially on paired comparisons, he recognized later that many other types of choice data, including rankings, could be modeled in a similar way.
Mosteller (1951) considered and extended Thurstone's model, focusing on a particular case, i.e. the fifth law of comparative judgments of Thurstone's list. The Bradley & Terry model (Bradley & Terry, 1952) was obtained from the Thurstone & Mosteller model substituting the normal distribution with the logistic function.
The literature on the method of paired comparisons has grown over the years and several extensive reviews on method of paired comparisons are now available (Coombs, 1958; Torgerson, 1958; Bock & Jones, 1968; Kendall, 1970). Several applications have been in sensory testing, consumer testing, personnel rating (David, 1988), chess ranking (Joe, 1990), sports tournaments (Chan, 2011) and image quality assessment (Handley, 2001).

2.2 Utilities and Preferences

To describe consumers choices, marketing researchers extended to an econometric context paired comparison models introduced by Thurstone (McFadden, 1980; Manski & McFadden, 1981; Berry, 1994). McFadden was one of the researchers that made these models popular. In his original work on the San Francisco transit system he used a probabilistic choice model to analyze individual's decisions to use various modes of transport such as car, train, carpooling and bus (McFadden, 1974).
Econometric models of probabilistic choice, also called random utility models, assume that choice is a discrete event. Consumers, indeed, cannot leave the supermarket with one half of Coke or one half of Pepsi but they will

tend to leave with a full can of the chosen brand. Moreover, models assume that the utility of a brand varies across individuals as a random variable. Finally, individuals are supposed to be rational agents who make the choice that possesses the highest utility. So, considering the case of a choice among two alternatives A_i and A_j, an individual will choose object A_i over object A_j if the utility of A_i is greater than utility of A_j, that is if $U_i > U_j$. Discussion here is in terms of individual utility but similar reasoning applies to collective choices. The choice between two alternatives reflects paired comparison models assumptions but random utility models have been extended to choice among many possibilities: the preferred alternative is the one with the highest utility.

Random utility models can be used to describe the relationship between the outcome, the choice, and some explanatory variables, such as tastes or personal characteristics.

The utility, for individual h, of alternative A_j, U_{hj}, is composed from an observed part, as some known attribute for example its monetary cost, and an unobserved one, that is everything else that is not specified in the utility. Formally, utility is given by

$$U_{hj} = V_{hj} + E_{hj}$$

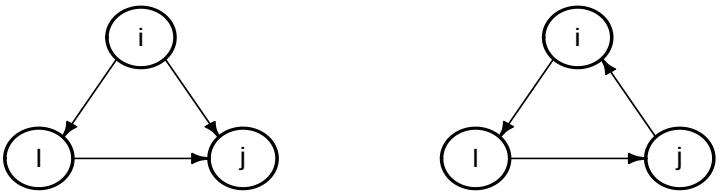
where V_{hj} is the observed part and E_{hj} is the error term. The assumptions about the distribution of the error allow the researcher to specify a density function to estimate the otherwise hidden term. Depending on the assumed distribution on the utilities, we have different classes of random utility models.

2.3 Pairwise Comparisons: Inconsistencies

Let us consider the simplest case in which three alternatives A_i, A_j and A_l are compared in pairs: if the respondent creates a continuum in his mind internally consistent then he expresses transitive preferences over those al-

ternatives. Assuming ties are not allowed, if the respondent prefers A_i to A_l and A_l to A_j then transitivity implies that A_i is preferred to A_j (Fig. 2.5 a). In this case the three alternatives can be ranked from first to third as follows: A_i, A_l and A_j. Conversely, intransitive preferences occur when, for instance, the respondent prefers A_i to A_l and A_l to A_j but A_j to A_i (Fig. 2.5 b).
Transitive preferences allow to position the alternatives along a linear continuum which expresses the dominance relations among them. A condition of complete transitivity is difficult to obtain in practice, especially measuring preferences on a set with a large number of alternatives. On the other hand, intransitivities occur when the items cannot be listed in a strict hierarchy, as it happens when alternatives are preferred the same proportion of times. Intransitivities are called also inconsistencies, or circular triads, since they produce a loop on preferences in which each alternative is preferred to each other including itself (Kendall & Babington-Smith, 1940).

Figure 2.5: Preferences relationships among three alternatives



(a) Transitive relationships (b) Intransitive relationships

Intransitivities may stem from either the incompetence of the respondent or the high similarity among the alternatives, or both. Another possible explanation is that the respondent evaluates the alternatives based on more than one dimension and he cannot order them on a linear scale (David, 1988).
The concept of consistency presented in this section is referred to as internal consistency, which is different from that of agreement between a set of respondents or between a respondent and a true ranking, which are named

rater agreement or external consistency.
Paired comparison experiments, unlike the other preference elicitation methods, allow researchers to identify respondents with inconsistent preferences. Researchers may decide to consider whether or not preferences elicited from these respondents.

2.4 Data Structure

Given a finite set of alternatives $A = (A_1, A_2, \ldots, A_k)$ with $k \geq 2$, respondents are called to express their preferences comparing them in pairs. Consider the case that each respondent makes all the $k(k-1)/2$ possible comparisons. The data collected this way can be represented in several kinds of matrices introduced in the following sections. For each matrix, a condition of reciprocity is assumed in such a way that the preference of A_i over A_j can be derived from the preference of A_j over A_i. Moreover, some condition of consistency is assumed which depends on the different meaning of the elements of the preference matrix.

2.4.1 Tournament Matrix

A simple structure is the so-called tournament matrix (Moon, 1968), a $k \times k$ zero-one matrix $\boldsymbol{T} = [t_{ij}]$ with the following property

$$\boldsymbol{T} + \boldsymbol{T}^{\boldsymbol{T}} = \boldsymbol{J} - \boldsymbol{I}$$

where $\boldsymbol{I}$ is the identity matrix of order k and $\boldsymbol{J}$ is a $k \times k$ matrix of all ones. Alternatives are compared each other once and t_{ij} equals 1 if and only if the i-th alternative is preferred to the j-th one, and 0 otherwise. In this case, $t_{ij} = 1 - t_{ji}$. Preferences can be read easily through the rows of a tournament matrix; for example looking at the following matrix $\boldsymbol{T}$ we can easily conclude that the first alternative is preferred to the other two, the second one is preferred to the third and the third one is never preferred.

$$\boldsymbol{T} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Matrix $\boldsymbol{T}$ is fully transitive. In literature, several measure of transitivity have been analyzed. Kendall and Babington–Smith (1940) proposed a consistency coefficient based on the number of circular triads detected in matrix $\boldsymbol{T}$. A set of judgments may be regarded as more consistent than another if it includes fewer circular triads. The number of circular triads does not provide a complete description of the circularity of preferences. It would be necessary to consider cycles of greater amplitude, as 4- or 5- cycles, but no simple formula is known for their individuation. Slater (1961) proposed as a measure of inconsistency the minimum number s of preference reversals, needed to reach an unambiguous ranking of the alternatives.
Also the maximum eigenvalue, λ_1, provides a measure of how much transitive are the represented preferences. More precisely, defining $\mu = 2\lambda_1/(k-1)$, we have that $0 \leq \mu \leq 1$ for every tournament matrix, since $0 \leq \lambda_1 \leq (k-1)/2$ (Moon & Pullman, 1970). In particular, in case of perfect transitivity λ_1 equals 0 and so does μ. Then, small values of μ correspond to preferences that are nearly transitive. Large values of μ, on the other side, correspond to the opposite situation. In the special case of equal row sums, achievable only for odd k, $\lambda_1(\boldsymbol{T})$ equals $(k-1)/2$ and μ equals 1. In graph theory language, in this case, $\boldsymbol{T}$ is called regular tournament. If k is even and half of the row sums of $\boldsymbol{T}$ equal $k/2$ and the rest equal $(k-2)/2$, then $\boldsymbol{T}$ is called almost regular.

2.4.2 Generalized Tournament Matrix

If every respondent compares all the alternatives in pairs more than once or else, if more than one respondent makes all the comparisons independently, collective preferences can be represented in a generalized tournament matrix

$\boldsymbol{P} = [\pi_{ij}]$, a $k \times k$ matrix that satisfies

$$\boldsymbol{P} + \boldsymbol{P'} = \boldsymbol{J} - \boldsymbol{I}, \qquad \boldsymbol{P} \geq 0$$

where $\boldsymbol{I}$ is the identity matrix of order k and $\boldsymbol{J}$ is a $k \times k$ matrix of all ones (Moon & Pullman, 1970).
In this case π_{ij} represents the probability that A_i is preferred to A_j and it can be estimated through p_{ij}, the number of comparisons in which A_i is preferred to A_j divided by the total number of performed comparisons, provided there are no ties. Hence:

- $p_{ij} = 1/2$ indicates indifference between A_i and A_j;
- $p_{ij} = 1$ indicates that A_i is always preferred to A_j;
- $p_{ij} > 1/2$ indicates that A_i is preferred to A_j.

Also in this case, it holds that $p_{ij} = 1 - p_{ji}$. Given n respondents, a generalized tournament matrix can be interpreted as the mean of n tournament matrices representing individual preferences.
For a generalized tournament matrix we can define different degrees of transitivity. The stochastic transitivity condition holds if, for every triad of alternatives A_i, A_l and A_j, $p_{il} \geq 0.5$ and $p_{lj} \geq 0.5$, imply that p_{ij} is greater than 0.5.
The "strong" stochastic transitivity condition is more stringent. It provides that, if $p_{il} \geq 0.5$ and $p_{lj} \geq 0.5$, then p_{ij} is greater than the maximum of p_{il} and p_{lj}. An intermediate condition is given by the "moderate" stochastic transitivity which provides that, under the same hypotheses, p_{ij} is greater than the minimum of p_{il} and p_{lj} (Coombs, 1958).
Fedrizzi et al. (2007) define a generalized tournament matrix perfectly transitive if and only if $p_{il} = p_{ij} + p_{jl} - 0.5$, $\forall i \neq l \neq j = 1, \ldots, k$. They proposed a consistency index which takes into account whenever this condition is violated.
As for tournament matrices, also in this case the maximum eigenvalue, λ_1, is

proportional to internal consistency. Defining $\mu = 2\lambda_1/(k-1)$ $(0 \leq \mu \leq 1)$, small values of μ correspond to preferences that are nearly transitive and large values of μ to the opposite situation (Moon & Pullman, 1970).

2.4.3 Multiplicative Paired Comparison Matrix

Another kind of matrix, widely used in literature (Saaty, 1977, 2008), is the multiplicative paired comparison matrix $\boldsymbol{M} = [m_{ij}]$ in which $m_{ij} \in]0, +c[$ where $c \in]0, +\infty[$ and

- $m_{ij} = 1/m_{ji}$ for $i, j = 1, \ldots, k$ and $i \neq j$;
- $m_{ii} = 1$ for $i = 1, \ldots, k$.

A matrix entry m_{ij} represents the ratio between A_i and A_j preferences: $m_{ij} > 1$ implies that A_i is preferred to A_j, whereas $m_{ij} < 1$ expresses the opposite preference and $m_{ij} = 1$ means indifference.
Saaty proposed the use of the multiplicative paired comparison in the Analytic Hierarchy Process (AHP), a tool to solve multi-criteria decision problems. The structure of the typical decision problem considered in AHP contains k alternatives and c decision criteria. Each alternative can be evaluated in terms of the decision criteria and the relative importance of each criterion can be elicited through pairwise comparisons. The decision-maker has to express her/his opinion about the value of one single pairwise comparison at a time, choosing a linguistic option such as A_i is x times more important than A_j, or A_i is of the same importance as A_j.
A one-to-one mapping between the set of discrete linguistic choices and a discrete set of numbers representing the importance is defined. According to the scale introduced by Saaty, the available values for the pairwise comparisons are members of the set {9, 8, 7, 6, 5, 4, 3, 2, 1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9}.
A multiplicative paired comparison matrix is perfectly transitive if it respects the multiplicative consistency, that is if $m_{il} = m_{ij}m_{jl} \quad \forall i \neq j \neq l = 1, \ldots, k$.
The assumption of the Saaty scale restricts the respondent's possibility to

be consistent, in fact if for example $m_{ij} = 5$ and $m_{jl} = 3$, clearly m_{il} cannot equals 15. However, Saaty provided a measure of closeness to the consistency (CI) again in terms of the maximum eigenvalue $CI = (\lambda_1 - k)/(k-1)$. Small values of the index indicate a good level of transitivity. $\boldsymbol{M}$ is fully transitive if $\lambda_1 = k$ and $CI = 0$. To determine the goodness of the consistency index, Saaty proposed to compute the consistency ratio CR given by $CR = CI/RI$, where RI is the mean consistency index computed on a sample of randomly generated matrices using the same matrix dimensionality and scale.

2.4.4 Additive Paired Comparison Matrix

The additive paired comparison matrix $\boldsymbol{D} = [d_{ij}]$ is defined as a $k \times k$ matrix in which $d_{ij} \in]-\infty, \infty[$ with the following characteristics:

- $d_{ij} = -d_{ji}$ for $i = 1, \ldots, k$ and $i \neq j$;
- $d_{ii} = 0$.

Matrix $\boldsymbol{D}$ is such that $\boldsymbol{D} = -\boldsymbol{D}^T$. In this case each element d_{ij} represents the difference of preference between A_i and A_j. Hence, $d_{ij} > 0$ implies that A_i is preferred to A_j, $d_{ij} < 0$ implies the opposite preference and $d_{ij} = 0$ implies indifference. The additive paired comparison matrix is known in algebraic literature as skew symmetric matrix. An additive paired comparison matrix is perfectly consistent if and only if $d_{il} = d_{ij} + d_{jl}$ $\forall i, j, l = 1, \ldots, k$ and $i \neq j \neq l$. The additive paired comparison matrices are sometimes derived from large-scale modern internet and e-commerce databases.

Chapter 3

Ranking and Scoring Methods

In this chapter we will describe some popular methods suggested in literature for ranking and scoring the alternatives starting from a generalized tournament matrix $\boldsymbol{P}$. Comparing two alternatives, A_i and A_j, a respondent prefers A_i to A_j with theoretical probability π_{ij}, where $0 \leq \pi_{ij} \leq 1$ for $i, j = 1, \ldots, k$ and $i \neq j$. Supposing that n respondents compare all the $k(k-1)/2$ possible pairs independently, π_{ij} can be estimated through p_{ij}, the proportion of comparisons in which A_i is preferred to A_j. Each respondent is assumed to be equally informative. From the quantitative estimates of the preference relations we want to construct a ranking of the alternatives or define a set of weights that reflect their relative importance.
For each of the proposed methods we give a brief description and some examples to better understand how it works. Moreover we consider their reliability in the cases p_{ij} equals either 0 or 1.

3.1 Score Vector

The score vector method (Thurstone, 1927b; Kendall, 1955; David, 1987) is the simplest way to score the alternatives.

The score vector $\boldsymbol{\omega}^{(1)}$ is the $(k \times 1)$ vector of row sums obtained as

$$\boldsymbol{\omega}^{(1)} = \boldsymbol{P} \cdot \boldsymbol{e}$$

where $\boldsymbol{e}$ is a column vector of 1's. The score vector method is practicable even if p_{ij} equals either 0 or 1.

For a tournament matrix the i-th row sum represents the number of times the i-th alternative is preferred to the others, while for a generalized tournament matrix it can be interpreted as the expected score of A_i, that is the expected number of alternatives the i-th one is preferred to the others.

Let us consider the following fully transitive matrix, $\boldsymbol{T}$, which represents the preferences expressed by a respondent who compares every pair of alternatives once.

$$\boldsymbol{T} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

We have that $\boldsymbol{\omega}^{(1)} = (2, 1, 0)^T$ and the ranking of the alternatives is obviously A_1, A_2 and A_3, with A_1 scoring most.

This ranking agrees with that obtained by the vector of the average probability estimates $\mathbf{p} = (p_{1.}, \ldots, p_{k.})$, whose elements are defined as follows (David, 1988)

$$p_{i.} = \frac{1}{k-1} \sum_{j \neq i} p_{ij} \text{ for } i = 1, \ldots, k.$$

In other words, $p_{i.}$ represents the average proportion of comparisons for which A_i is preferred. In the previous example we have $\mathbf{p} = (1, 0.5, 0)$. The sum of the average probabilities equals $(k-1)/2$.

To estimate the score s_i of the i-th alternative it is possible to use both the score vector entries and the average probabilities estimates. After being normalized, the estimates $\hat{s_i}$ obtained in both cases coincide.

Consider also the case of intransitive preferences, represented by the following matrix $\boldsymbol{T_1}$. It can be associated to the Chinese game of rock (R)-paper (P)-scissors (S), in which rock breaks scissors, scissors cut paper but paper covers rock.

$$\boldsymbol{T_1} = \begin{array}{c} \\ R \\ S \\ P \end{array} \begin{array}{c} \begin{array}{ccc} R & S & P \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

The score vector of $\boldsymbol{T_1}$ is given by $\boldsymbol{\omega}^{(1)} = (1, 1, 1)$. In this case, it is not possible to order the alternatives since each of them is preferred the same number of times. This matrix describes the so called regular tournament in which, for k odd, all of the row sums equal $(k-1)/2$.
To see one more application consider the following generalized tournament matrix

$$\boldsymbol{P_1} = \begin{bmatrix} 0 & 0.9 & 0.9 & 0.6 & 0.4 \\ 0.1 & 0 & 0.8 & 0.8 & 0.8 \\ 0.1 & 0.2 & 0 & 0.7 & 0.7 \\ 0.4 & 0.2 & 0.3 & 0 & 0.6 \\ 0.6 & 0.2 & 0.3 & 0.4 & 0 \end{bmatrix}$$

that represents the situation in which n respondents compare independently 5 alternatives in pairs.
Let us remember that every p_{ij} is the proportion of comparisons in which A_i is preferred to A_j. Looking at this matrix, for example, we can say that, since $p_{12} = 0.9$, 9 out of 10 respondents prefer alternative A_1 to A_2.
The maximum eigenvalue $\lambda_1(\boldsymbol{P_1})$ equals 1.848 and the inconsistency index $\mu = 2\lambda_1/(k-1)$ equals 0.924 (Moon & Pullman, 1970). Let us remember that $0 \leq \mu \leq 1$ and that small values of μ correspond to near transitive preferences while large values to the opposite situation. Then, $\boldsymbol{P_1}$ seems

characterized by low consistency. Considering also the number of 3-cycles, we have three over the ten possible triads that result circular.
Table 3.1 shows the row-sums, the average probabilities and the ranking referred to this specific example.

Table 3.1: Row-sums, average probabilities and ranking derived from $\boldsymbol{P_1}$

Alternatives	Row-sums	Average probabilities	Ranking
1	2.8	0.56	1
2	2.5	0.50	2
3	1.7	0.34	3
4	1.5	0.30	4
5	1.5	0.30	4

It may be noticed that, the last two alternatives have the same row sum and, consequently, the same position in the alternatives' ranking.

3.2 Eigenvector Method

Another method apt to scoring a set of alternatives is based on the eigen-decomposition of the generalized tournament matrix $\boldsymbol{P}$ obtained from a complete paired comparisons experiment in which n respondents compare k alternatives in pairs. Since $\boldsymbol{P}$ is a positive irreducible matrix, the Perron-Frobenius theorem guarantees that a unique large real eigenvalue λ_1 exists as solution of the characteristic equation (Keener, 1993)

$$\boldsymbol{P\omega} = \lambda_1 \boldsymbol{\omega}$$

and that the corresponding eigenvector $\boldsymbol{\omega}$ has strictly positive components. The principal eigenvector $\boldsymbol{\omega}$ is uniquely determined imposing that $\|\boldsymbol{\omega}\|_2 = 1$, that is the squared entries sum up to 1. The i-th entry of ω is assumed to represent the relative importance of the i-th alternative and it can be used to estimate its score. Moreover, alternatives can be ranked according to these estimates (Horn & Johnson, 1985).

The idea of using the eigenvector to score is due to Kendall and Wei (Wei, 1952; Kendall, 1955) and the method has acquired new currency today thanks to web applications. The web search engine Google (www.google.com) uses a variant of this idea to measure the importance of a large number of web sites (Brin & Page, 1998).
As we have seen, the i-th entry of the score vector represents the number of alternatives which the i-th one is preferred to. The score vector ranking, then, takes into account only the number of preferred alternatives. Kendall proposed, as suggested by Wei, to use as measures of strength the row sums of the power of the matrix $\boldsymbol{P}$. In fact, the i-th entry of $\boldsymbol{\omega}^{(2)}$, defined as

$$\boldsymbol{\omega}^{(2)} = \boldsymbol{P}^2\boldsymbol{e}$$

where $\boldsymbol{e}$ is a vector of all 1's, is the sum of the scores of all the alternatives preferred to the i-th one and so it can be considered as another measure of relative strength. The more an alternative is preferred to other alternatives with high row sum, the higher the final score of that alternative.
Iterating the reasoning, we can compute

$$\boldsymbol{\omega}^{(3)} = \boldsymbol{P}^3\boldsymbol{e}$$

and consider this vector as a further measure of relative strength of the alternatives since it contains the sums of the sums of the scores of the preferred alternatives. According to the power method, for $m \to \infty$ we have that

$$\boldsymbol{\omega}^{(m)} = \boldsymbol{P}^m\boldsymbol{e}$$

converges to the principal eigenvector $\boldsymbol{\omega}$ for $\boldsymbol{P}$ (Mises, 1929).
Let us consider the matrix analyzed in section 3.1 which represents perfectly transitive preference probabilities

$$\boldsymbol{T} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

As we have seen, if a matrix is completely transitive, the maximum eigenvalue λ_1 is equal to 0. Moreover, for $k \geq 2$, the matrix representing completely transitive preferences, having $t_{ij} = 1$ if and only if $i < j$, is reducible and so it has all the k eigenvalues equal to 0 (De Caen et al., 1992). In this case, then, the eigenvector method does not provide a meaningful ranking and it may be more convenient to consider the ranking obtainable from the score vector.
In the opposite case of completely intransitive preferences represented by the following matrix

$$\boldsymbol{T_1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

the eigen-decomposition returns the maximum eigenvalue $\lambda_1 = 1$ with associated eigenvector $\boldsymbol{\omega} = (0.577, 0.577, 0.577)$. Also this method returns tied rankings. For any value of k, in case of regular tournament matrix, the maximum eigenvalue λ_1 equals to $(k-1)/2$ and $\omega_i = 1/\sqrt{k}$ for $i = 1, \ldots, k$. The eigenvector method may be applied also if p_{ij} equals either 0 or 1.
Let us see in detail what happens applying the eigenvector method to the matrix $\boldsymbol{P}$ considered in Section 3.1

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0.9 & 0.9 & 0.6 & 0.4 \\ 0.1 & 0 & 0.8 & 0.8 & 0.8 \\ 0.1 & 0.2 & 0 & 0.7 & 0.7 \\ 0.4 & 0.2 & 0.3 & 0 & 0.6 \\ 0.6 & 0.2 & 0.3 & 0.4 & 0 \end{bmatrix}$$

In Table 3.2 we report the score vectors of the power of $\boldsymbol{P}$, each divided by its Frobenius norm, so as to obtain the same normalization of the eigenvector. For completeness, we report, in the last column, also the values of the principal eigenvector $\boldsymbol{\omega}$ of $\boldsymbol{P}$.

Table 3.2: Normalized score vectors after each iteration of the power method applied on $\boldsymbol{P}$

Alternatives	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\omega^{(5)}$	$\omega^{(6)}$	ω
A_1	0.604	0.620	0.589	0.602	0.601	0.599	0.600
A_2	0.539	0.474	0.496	0.497	0.493	0.495	0.494
A_3	0.367	0.338	0.362	0.357	0.355	0.3571	0.357
A_4	0.324	0.356	0.362	0.354	0.357	0.3568	0.356
A_5	0.324	0.386	0.381	0.372	0.377	0.376	0.376

The last two alternatives have the same row sum but they differ from the second iteration on. The rankings according to the score vector and the eigenvector methods agree only for the first two positions. As expected, the eigenvector method weights more A_5 than A_4, because, despite having the same row sum, the last alternative has probability 0.6 to be preferred to A_1, the high ranked alternative.

3.3 Linear Models

Another method, well known in the literature, for ranking and scoring the alternatives is the linear model. The linear paired comparison model (David, 1988) assumes that any of the k alternatives has true rating V_i which can be used to determine their ordering. The scores of the k alternatives can be represented on a continuum. Given the ratings of two competing alternatives, linear models yield the exact probability that one alternative is preferred to the other. Such probability could, in theory, be verified if the alternatives were compared a large number of times.

Every respondent evaluates each alternative more than once and her/his judgment can vary in each replication. The judgment process may be represented by a continuous variable Y_i defined on the real line, with mean V_i, whose distribution is called "sensation distribution".

In a pairwise comparison between two alternatives A_i and A_j, A_i is preferred

to A_j if and only if $Y_i \geq Y_j$. Let us define Z_i and $Z_i - Z_j$ as follows

- $Z_i = Y_i - V_i$ is a continuous variable with 0 mean and the same distribution as Y_i;
- $Z_i - Z_j$ is a symmetric variable with 0 mean, whose distribution, called *defining distribution*, depends on the *sensation distribution* of Y_i (Latta, 1979).

Under linear models hypotheses it is possible to express the preference probability π_{ij} as

$$\begin{aligned}\pi_{ij} &= Pr[Y_i - Y_j > 0] \\ &= Pr[Z_i - Z_j > -(V_i - V_j)] \\ &= F_{Z_i - Z_j}(V_i - V_j).\end{aligned}$$

Starting from an estimated generalized tournament matrix, obtained through a paired comparison experiment in which n respondents compare all the possible pairs, the aim is to estimate the true ratings V_i's (Noether, 1960). Denoting $\delta_{ij} = V_i - V_j$ we have that $\pi_{ij} = F(\delta_{ij})$, where F is the distribution function of a symmetric continuous real variable. Denoting with d_{ij} the estimate of δ_{ij}, it is possible to estimate V_i's, solving the following linear system

$$d_{ij} = \hat{V}_i - \hat{V}_j = F^{-1}_{Z_i - Z_j}(p_{ij})$$

where p_{ij}, the estimate of π_{ij}, is the proportion of comparisons in which the respondent prefers A_i to A_j.
This linear system has k unknown parameters, the k ratings, and $k(k-1)$ equations. So, for $k > 3$, the system results overdetermined and we need to impose some contraints to solve it. Since the origin of the linear scale is arbitrary, David (1988) suggested to impose

$$\sum_{i=1}^{k} \hat{V}_i = 0.$$

To obtain each $\hat{V}_i$ it is sufficient to sum d_{ij} over all $j \neq i$

$$\begin{aligned}\sum_{j\neq i} d_{ij} &= (k-1)\hat{V}_i - \sum_{j\neq i}\hat{V}_j \\ &= (k-1)\hat{V}_i + \hat{V}_i - \sum_{j=1}^{k}\hat{V}_j \\ &= (k\hat{V}_i)\end{aligned}$$

and from this equation it follows that

$$\hat{V}_i = \frac{\sum_{j\neq i} d_{ij}}{k}.$$

$\hat{V}_i$'s result to be the unweighted least squares estimates of the V_i's, since they minimize the following quantity

$$S = \sum_{i=1}^{k}\sum_{j\neq i}(d_{ij} - (V_i - V_j))^2.$$

Indeed, deriving S with respect V_i we have

$$\begin{aligned}\frac{\delta S}{\delta V_i} &= -2\sum_{j\neq i}(d_{ij} - V_i + V_j) + 2\sum_{j\neq i}(d_{ji} - V_j + V_i) \\ &= -4\sum_{j\neq i}(d_{ij} - \delta_{ij}) \\ &= -4(\sum_{j\neq i} d_{ij} - kV_i)\end{aligned}$$

that, if the derivative is set to 0, gives $\hat{V}_i = \frac{\sum_{j\neq i} d_{ij}}{k}$.

3.3.1 Thurstone & Mosteller Model

In the first linear model for paired comparisons by Thurstone (1927c), decision behaviors are accounted for in probabilistic terms. Consequently, also apparent inconsistencies are explained in the same terms. According to Thurstone's model, every respondent, when compares two alternatives, chooses the

one with the greater rating.
The variables Y_i's for $i = 1, \ldots, k$, which model judgments on the i-th alternative, are normally distributed.
Thurstone stated different versions of its approach listed below

- CASE I. Only one respondent compares the alternatives more than once. Y_i's for $i = 1, \ldots, k$ are equicorrelated normal variables with mean V_i, variance σ_i^2 and common correlation coefficient ρ;
- CASE II. The assumption in case I of more than one comparison by the same respondent is generalized such that more respondents compare the alternatives indipendently. Y_i's $(i = 1, \ldots, k)$ are ever equicorrelated normal variables with mean V_i, variance σ_i^2 and common correlation coefficient ρ;
- CASE III. Thurstone simplified its model assuming that the common correlation coefficient ρ equals 0. So, the Y_i's are independent variables with mean V_i and variance σ_i^2;
- CASE IV. The model is further simplified assuming that the Y_i's variances are related through the following $\sigma_i = \sigma_j + d$ for $i, j = 1, \ldots, k$ and $i \neq j$. Furthermore, d is assumed small;
- CASE V. This corresponds to the most famous version of the Thurstone's model, subsequently reconsidered by Mosteller (Mosteller, 1951), in which the Y_i's are assumed to be independent normal variables with mean V_i and common variance σ^2. Under this assumption, the overlap of the distributions of the two alternatives provides a measure for the distance between the mean of the alternatives.

In the following, we refer to the Thurstone & Mosteller model meaning the Thurstone's model CASE V. So in this case, $Z_i - Z_j$ has a normal distribution with 0 mean and variance $2\sigma^2$ and

$$\pi_{ij} = \Phi\left(\frac{V_i - V_j}{\sqrt{2}\sigma}\right)$$

where Φ is the distribution function of a standard normal variable. From the previous relation we can estimates the ratings using the relations

$$\frac{d_{ij}}{\sqrt{2}\sigma} = \frac{\hat{V}_i - \hat{V}_j}{\sqrt{2}\sigma} = \Phi^{-1}(p_{ij})$$

for $i, j = 1, \ldots, k$ and $i \neq j$.
If p_{ij} equals 0 or 1 it is not possible to estimate d_{ij} since the inverse of the standard normal distribution in these points equals, respectively, $-\infty$ and $+\infty$.

3.3.2 Bradley & Terry Model

Another method for paired comparisons was proposed by Bradley & Terry (1952) whose main assumption is that

$$\pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$$

where π_i, for $i = 1, \ldots, k$, is a reparameterization of the true rating V_i of alternative A_i, such that $\sum_{i=1}^{k} \pi_i = 1$. Imposing that $V_i = \log(\pi_i)$, the Bradley & Terry model can be expressed as a linear model.
Even in this case we assume that n respondents compare the k alternatives in a complete paired comparison experiment. We replace the normal distribution, assumed in Thurstone & Mosteller model, with the logistic one. Following the steps specified for the previous method, assuming $i, j = 1, \ldots, k$ and $i \neq j$, we have to specify the following assumptions:

- The judgment processes can be represented by continuous independent random variables Y_i's with Gumbel distribution and parameters $(V_i, 1)$ which correspond to a mean equalling $V_i + \gamma$ where γ is the Eulero-Mascherano constant, and a variance equal to $\pi^2/6$;

- $Z_i = Y_i - V_i$ has a Gumbel standard distribution with mean γ and variance $\pi^2/6$;

- $Z_i - Z_j$ results to be a standard logistic variable with mean 0 and variance $\pi^2/3$.

So, we have that

$$\pi_{ij} = F_{Z_i - Z_j}(V_i - V_j) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{V_i - V_j}{2}\right)$$

and inverting the relation

$$V_i - V_j = 2\,\mathrm{arctanh}(2\pi_{ij} - 1).$$

Using the formula $\mathrm{atanh}(x) = \frac{1}{2}\ln\left(\frac{x+1}{x-1}\right)$ we obtain

$$\delta_{ij} = V_i - V_j = \ln\left(\frac{\pi_{ij}}{\pi_{ji}}\right).$$

$\hat{V}_i$'s can be obtained through the following algorithm

$$\hat{V}_i = \sum_{j \neq i} \ln\left(\frac{p_{ij}}{p_{ji}}\right) = \ln \prod_{j \neq i}\left(\frac{p_{ij}}{p_{ji}}\right).$$

The $\hat{V}_i$'s estimated parameters are such that $\sum_{i=1}^{k} \hat{V}_i = 0$ by construction. Under the Bradley & Terry reparameterization $V_i = \log(\pi_i)$, we get that $\sum_{i=1}^{k} \log \pi_i = 0$ and $\prod_{i=1}^{k} \pi_i = 1$, and not the Bradley & Terry restriction $\sum_{i=1}^{k} \pi_i = 1$. For this reason, it is necessary to make a scale reparametrization

$$\pi_i' = \frac{\pi_i}{\sum \pi_i}.$$

Also in this case, if p_{ij} equals either 0 or 1, it is not possible to estimate V_i, since $\log\left(\frac{0}{1}\right) = \infty$ and $\log\left(\frac{1}{0}\right)$ is undefined.

Likelihood estimation of the Bradley-Terry Model

To estimate the ratings π_i of the k alternatives, under the Bradley & Terry assumptions $\pi_i \geq 0$ and $\sum \pi_i = 1$, it is possible to use a maximum likelihood approach.

Because of the indipendence of all comparisons, the probability of observing

α_{ij} preferences for A_i in comparisons with A_j is

$$\binom{n}{\alpha_{ij}} \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{\alpha_{ij}} \left(\frac{\pi_j}{\pi_i + \pi_j}\right)^{n-\alpha_{ij}}.$$

The likelihood function L is the product of such probabilities for any $\binom{k}{2}$ independent pairings and, after some computations, it may be expressed in terms of the row sums a_i as

$$L = C \frac{\prod_{i=1}^{k} \pi_i^{a_i}}{\prod_{i<j} (\pi_i + \pi_j)^n}$$

where $C = \prod_{i<j} \binom{n}{\alpha_{ij}}$ and $a_i = \sum_j \alpha_{ij}$. Since C is independent of π_i's, we have that a_i's are sufficient statistics for the π_i's. For any generalized tournament matrix with $\pi_{ij} > 0$, with $i \neq j$ all the scores are sufficient statistics under the Bradley & Terry model.

Differentiating $\log L$ with respect to π_i we obtain the maximum likelihood estimates p_i. They are functions of the a_i and do not involve the individual α_{ij} in contrast with what happens fot the estimates of the V_i's.

We can write the likelihood equations in the form

$$p_i = \frac{a_i}{n \sum_{i<j} (p_i + p_j)^{-1}}$$

and we find the solutions using an iterative process. Starting with a set of trial solutions $(p_1^{(0)}, p_2^{(0)}, \ldots, p_k^{(0)})$ we can obtain $p_i^{(1)}$, $(i = 1, 2, \ldots, k)$ from

$$p_i^{(1)} = \frac{a_i}{n} \left[\frac{1}{p_i^{(0)} + p_1^{(1)}} + \cdots + \frac{1}{p_i^{(0)} + p_{i-1}^{(1)}} + \frac{1}{p_i^{(0)} + p_{i+1}^{(0)}} + \cdots + \frac{1}{p_i^{(0)} + p_k^{(0)}} \right]$$

and continuing the iterative process until the deviation between $p_i^{(r+1)}$ and p_i^r is sufficiently thin. This procedure has a slow convergence.

The p_i's obtained with the maximum likelihood estimates do not add to one and so they need to be reparameterized. Zermelo (1929) pointed out that the

ranking achieved by the maximum likelihood estimates is always the same as that obtained with the score vector method.

3.3.3 Cauchy Model

The Cauchy model assumes that

- Y_i's for $i = 1, \ldots, k$ are independent continuous variables with Cauchy distribution and location and scale parameters, respectively, V_i and σ^2;
- $Y_i - V_i$ for $i = 1, \ldots, k$ is an independent continuous variable with Cauchy distribution and location and scale parameters, respectively, equal to 0 and σ^2;
- $Z_i - Z_j$ for $i = 1, \ldots, k$ and $i \neq j$ is an independent continuous variable with Cauchy distribution and location and scale parameters 0 and $2\sigma^2$ respectively. Changing scale, we can say that $Z_i - Z_j$ has a Student's t distribution with 1 degree of freedom.

By these hypotheses we have

$$\pi_{ij} = F_{Z_i - Z_j}(V_i - V_j) = \frac{1}{2} + \frac{1}{\pi} \arctan(V_i - V_j)$$

from which we obtain

$$V_i - V_j = \tan\left[\pi(\pi_{ij} - 1/2)\right]$$

where π is the pi greek constant.
It is possible to obtain $\hat{V}_i$, for $i = 1, \ldots, k$, substituting the observed preference proportions p_{ij} to the theoretical probabilities π_{ij} and solving the linear system under the usual restriction that $\sum \hat{V}_i = 0$.
It has to be noted that the event $p_{ij} = 0$ leads to $d_{ij} = \tan\left(-\frac{\pi}{2}\right) = -\infty$, while $p_{ij} = 1$ leads to $d_{ij} = \tan\left(\frac{\pi}{2}\right) = \infty$.

3.3.4 Uniform Model

The uniform model assumes that $Z_i - Z_j$ for $i = 1, \ldots, k$ and $i \neq j$ are independent variables with uniform distribution on the interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$. Then, we have that

$$\pi_{ij} = F(V_i - V_j) = V_i - V_j + \frac{1}{2}$$

from which we can easily obtain

$$V_i - V_j = \pi_{ij} - \frac{1}{2}.$$

Finally, the estimates of V_i are given by

$$\begin{aligned} \hat{V_i} &= \frac{1}{k} \sum_{j \neq i} (p_{ij} - \frac{1}{2}) \\ &= \left[a_i - \frac{1}{2} n(k-1) \right] \end{aligned}$$

where a_i is the row sum associated to the i-th alternative. David (1988) points out that, from the point of view of estimating the ratings V_i, using this distribution to model π_{ij} is equivalent to using the row sums of $\boldsymbol{P}$.

3.3.5 Exponential Model

The exponential model assumes that $Z_i - Z_j$ for $i, j = 1, \ldots, k$ and $i \neq j$ are indipendent variables with standard Laplace distribution, so that

$$\pi_{ij} = F_{Z_i - Z_j}(V_i - V_j) = \frac{1}{2} \left[1 + \text{sign}(V_i - V_j)(1 - \exp(-|V_i - V_j|)) \right]$$

from which, supposing without loss of generality that $V_i \geq V_j$, we have

$$V_i - V_j = \ln \left(-2(\pi_{ij} - 1) \right)^{-1}.$$

It is possible to obtain the estimate of the V_i's in the usual way solving the linear system

$$d_{ij} = \hat{V_i} - \hat{V_j} = \frac{1}{2} \left[1 + \text{sign}(V_i - V_j)(1 - \exp(-|V_i - V_j|)) \right]$$

under the constraint that $\sum_i \hat{V}_i = 0$.

The sensation and defining distributions corresponding to the linear models just described are summarized in the synoptic Table 3.3.

Table 3.3: Sensation and defining distributions of the examined linear models

Model	Sensation Distribution	Defining Distribution
Thurstone&Mosteller	Normal	Normal
Bradley& Terry	Gumbel	Logistic
Cauchy	Cauchy	Cauchy
Uniform	Uniform	Triangular
Exponential	Exponential	Laplace

Chapter 4

Ranking the alternatives: A simulation study

In the following, we will analyze the methods presented in Chapter 3 for ranking or scoring the alternatives, starting from a generalized tournament matrix $\boldsymbol{P} = [p_{ij}]$, obtained from a paired comparison experiment in which n respondents compare the k alternatives in pairs. We assume that every respondent performs all the $k(k-1)/2$ possible comparisons.

The analysis of matrix $\boldsymbol{P}$ allows to estimate the relative position of the alternatives on a continuum, as well as their ranks. Scoring methods attribute a numerical value to each alternative, which quantifies their relative importance. Ranking methods intend to capture the ordinal aspects by assigning a rank to the alternatives, ignoring quantitative information on how much an alternative is important relatively to the others.

To perform the comparison between the methods, we will rely on properties of ranking. For this purpose, we have performed some simulation studies using various data generation systems. In particular, we used two different approaches: simulating directly the matrices (Section 4.1) and starting from the random variables that model the preference relations under the linear model hypotheses (Section 4.2).

4.1 Simulating Typical Matrices

4.1.1 Random Matrices

First of all, we generated randomly some generalized tournament matrices. For every $k \times k$ matrix $\boldsymbol{P}$ we obtained $k(k-1)/2$ observations $\mathbf{x} = (x_1, \ldots, x_{k(k-1)/2})$ from a continuous uniform variable $X \sim \mathcal{U}(0,1)$. These values compose the upper triangular block of $\boldsymbol{P}$, whose entries p_{ij} are such that $i < j$ and $j = 2, \ldots, k$. The lower triangular block is computed using the relation $p_{ji} = 1 - p_{ij}$. So, each matrix is obtained according to the scheme represented in $\boldsymbol{P}$.

$$\boldsymbol{P} = \begin{bmatrix} 0 & x_1 & x_2 & x_3 & \ldots & \ldots & x_{k-1} \\ 1-x_1 & 0 & x_k & x_{k+1} & \ldots & \ldots & \ldots \\ 1-x_2 & 1-x_k & 0 & \ldots & \ldots & \ldots & \ldots \\ 1-x_3 & 1-x_{k+1} & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & 0 & x_{k(k-1)/2} \\ 1-x_{k-1} & \ldots & \ldots & \ldots & \ldots & 1-x_{k(k-1)/2} & 0 \end{bmatrix}$$

We computed the rankings, respectively, with

- the score vector method (SV);
- the eigenvector method (EV);
- the linear models: Thurstone & Mosteller (T&M), Bradley & Terry (B&T), Cauchy (CAU), Uniform (UNI), Exponential (EXP)

and we studied the association between the rankings obtained with the different methods by means of the Spearman's Rank-Order Correlation ρ.
Defining y_{mi} the rank of the i-th alternative according to the method m, for every couple of methods m and n we have

$$\rho_{mn} = \frac{\sum_i (y_{mi} - \bar{y}_m)(y_{ni} - \bar{y}_n)}{\sqrt{\sum_i (y_{mi} - \bar{y}_m) \sum_i (y_{ni} - \bar{y}_n)}} \text{ with } i = 1, \ldots, k.$$

The Rank-Order Correlation coefficient ρ can take values between -1 and 1; ρ equals 1 if the rankings are in perfect agreement, $\rho = -1$ if they are in perfect disagreement and $\rho = 0$ signifies that there is no relationship (Spearman, 1904).
This step is performed just for exploratory purposes, so to understand which methods give similar results. However, the method least correlated with the others could be the more reliable method.
In the following we present the mean values of Spearman's correlation coefficients between the rankings obtained with the different methods over $n = 500$ replications for some values of k, in particular $k = (8, 16, 24)$.

Table 4.1: Mean Spearman's correlation coefficient, $\bar{\rho}$, for $n = 500$ replications and $k = 8$ (s.e. in parenthesis)

	SV	EV	T&M	B&T	CAU	UNI	EXP
SV	–	0.954	0.945	0.917	0.667	1.000	0.925
	–	(0.047)	(0.066)	(0.088)	(0.241)	(0.000)	(0.082)
EV	–	–	0.920	0.893	0.647	0.954	0.902
	–	–	(0.080)	(0.098)	(0.252)	(0.047)	(0.092)
T&M	–	–	–	0.983	0.782	0.945	0.955
	–	–	–	(0.028)	(0.176)	(0.066)	(0.055)
B&T	–	–	–	–	0.823	0.917	0.951
	–	–	–	–	(0.154)	(0.088)	(0.055)
CAU	–	–	–	–	–	0.667	0.773
	–	–	–	–	–	(0.241)	(0.170)
UNI	–	–	–	–	–	–	0.925
	–	–	–	–	–	–	(0.082)

As expected (See Table 4.1), the ranking obtained with the score-vector method corresponds to that obtained with the uniform linear model. The method least correlated with the others is the Cauchy linear model. The

Table 4.2: Mean Spearman's correlation coefficient, $\bar{\rho}$, for $n = 500$ replications and $k = 16$ (s.e. in parenthesis)

	SV	EV	T&M	B&T	CAU	UNI	EXP
SV	–	0.977	0.960	0.933	0.590	1.000	0.945
	–	(0.016)	(0.029)	(0.049)	(0.193)	(0.000)	(0.038)
EV	–	–	0.947	0.921	0.583	0.977	0.932
	–	–	(0.032)	(0.051)	(0.190)	(0.016)	(0.041)
T&M	–	–	–	0.988	0.717	0.960	0.969
	–	–	–	(0.012)	(0.151)	(0.029)	(0.022)
B&T	–	–	–	–	0.766	0.933	0.965
	–	–	–	–	(0.129)	(0.049)	(0.024)
CAU	–	–	–	–	–	0.590	0.719
	–	–	–	–	–	(0.193)	(0.142)
UNI	–	–	–	–	–	–	0.945
	–	–	–	–	–	–	(0.038)

highest mean value of the Spearman's correlation coefficients, $\bar{\rho}$, is between the Thurstone & Mosteller and the Bradley & Terry methods ($\bar{\rho} = 0.983$). Also the eigenvector and the score vector methods are highly correlated ($\bar{\rho} = 0.954$).

We highlight that, while increasing k, results substantially do not change (See Tab. 4.2). The correlations between the Cauchy and the other rankings further decrease, while all the correlations among the rankings obtained with the other methods increase. The highest averages of Spearman's correlation coefficients are between the Thurstone & Mosteller and the Bradley & Terry methods ($\bar{\rho} = 0.988$) and between the eigenvector and the score vector methods ($\bar{\rho} = 0.977$).

Table 4.3: Mean Spearman's correlation coefficient, $\bar{\rho}$, for $n = 500$ replications and $k = 24$ (s.e. in parenthesis)

	SV	EV	T&M	B&T	CAU	UNI	EXP
SV	–	0.986	0.963	0.939	0.524	1.000	0.950
	–	(0.009)	(0.020)	(0.033)	(0.170)	(0.000)	(0.029)
EM	–	–	0.956	0.932	0.520	0.986	0.943
	–	–	(0.023)	(0.035)	(0.174)	(0.009)	(0.031)
T&M	–	–	–	0.991	0.660	0.963	0.976
	–	–	–	(0.007)	(0.138)	(0.020)	(0.013)
B&T	–	–	–	–	0.711	0.939	0.973
	–	–	–	–	(0.123)	(0.033)	(0.014)
CAU	–	–	–	–	–	0.524	0.668
	–	–	–	–	–	(0.170)	(0.131)
UNI	–	–	–	–	–	–	0.950
	–	–	–	–	–	–	(0.029)

Table 4.3 shows the mean values of Spearman's correlation coefficients between the rankings obtained with the different methods over $n = 500$ replications for $k = 24$. The results confirm the findings obtained with the other values of k considered.

4.1.2 First Raw Dominant Matrices

Consider the case in which the set of k alternatives can be partitioned in two classes I and J such that each alternative in I is preferred to every alternative in J by more than 50% of the respondents ($p_{ij} > 0.5\ \forall i \in I,\ j \in J$). In order for the resulting ranking to be consistent, each member of I should also be preferred to any member of J.
In the particular case in which only the first alternative belongs to the class I while the others belong to J, we should expect that a good ranking method

would place the earliest alternative at the first place, since it is always preferred to the others. In many practical contexts the first objective in eliciting preferences from a sample of respondents is to identify the best alternative. For this reason the examination of this particular situation is crucial.

To analyze the behavior of the proposed ranking methods, we simulated some generalized tournament matrices imposing that the elements of the first row be greater than 0.5. We imposed the first alternative as the always preferred, just for representation convenience. We could bring any matrix with an alternative dominant to this just permuting the rows and columns of the matrix itself.

So, for each matrix, we obtained $k-1$ observations $\mathbf{u} = (u_1, \ldots, u_{k-1})$ from a continuous uniform variable $U \sim \mathcal{U}(0.5, 1)$ that constitute the first dominant row. The other elements $\mathbf{x} = (x_1, \ldots, x_{(k-1)(k-2)/2})$ of the upper triangular block of the matrix are obtained from a continuous uniform variable $X \sim \mathcal{U}(0, 1)$. The lower triangular matrix can be filled using the relation $p_{ji} = 1 - p_{ij}$. $\boldsymbol{P}$ shows the fill pattern of each simulated matrix

$$\boldsymbol{P} = \begin{bmatrix} 0 & u_1 & u_2 & \ldots & \ldots & \ldots & u_{k-1} \\ 1-u_1 & 0 & x_1 & x_2 & \ldots & \ldots & \ldots \\ 1-u_2 & 1-x_1 & 0 & \ldots & \ldots & \ldots & \ldots \\ \ldots & 1-x_2 & \ldots & 0 & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & 0 & x_t \\ 1-u_{k-1} & \ldots & \ldots & \ldots & \ldots & 1-x_t & 0 \end{bmatrix}$$

whith t equals $(k-1)(k-2)/2$.

As an example, let us consider the following matrix $\boldsymbol{P_1}$

$$\boldsymbol{P_1} = \begin{bmatrix} 0.00 & 0.75 & 0.85 & 0.82 & 0.86 & 0.66 & 0.83 & 0.93 \\ 0.25 & 0.00 & 0.18 & 0.99 & 0.32 & 0.38 & 0.58 & 0.58 \\ 0.15 & 0.82 & 0.01 & 0.72 & 0.55 & 0.16 & 0.43 & 0.05 \\ 0.18 & 0.00 & 0.28 & 0.00 & 0.26 & 0.46 & 0.68 & 0.88 \\ 0.14 & 0.68 & 0.45 & 0.74 & 0.00 & 0.17 & 0.66 & 1.00 \\ 0.34 & 0.62 & 0.84 & 0.54 & 0.83 & 0.00 & 0.03 & 0.04 \\ 0.17 & 0.42 & 0.57 & 0.32 & 0.34 & 0.97 & 0.00 & 0.62 \\ 0.07 & 0.42 & 0.95 & 0.12 & 0.00 & 0.96 & 0.38 & 0.00 \end{bmatrix}$$

The first alternative results dominant since $p_{1j} \geq 0.5 \ \forall j = 2, \ldots, 8$, while the other rows are obtained completely at random. Rankings achieved from matrix $\boldsymbol{P_1}$ are presented in Table 4.4.

Table 4.4: Ranking obtained with the different methods applied to matrix $\boldsymbol{P_1}$

Alternatives	SV	EV	T&M	B&T	CAU	UNI	EXP
A_1	1	1	1	1	3	1	1
A_2	4	5	3	3	2	4	3
A_3	7	6	6	6	5	7	6
A_4	8	8	8	8	8	8	8
A_5	2	2	2	2	1	2	2
A_6	5	3	5	5	6	5	5
A_7	3	4	4	4	4	3	4
A_8	6	7	7	7	7	6	7

In this specific example all the methods but the Cauchy one classify alternative A_1 as first.

We simulated 500 matrices for different values of $k = (8, 16, 24)$ and we computed the rankings obtained with the different methods. Then, for each

Table 4.5: Proportion of first alternative well classified for each method for $n = 500$ replications and $k = 8, 16, 24$

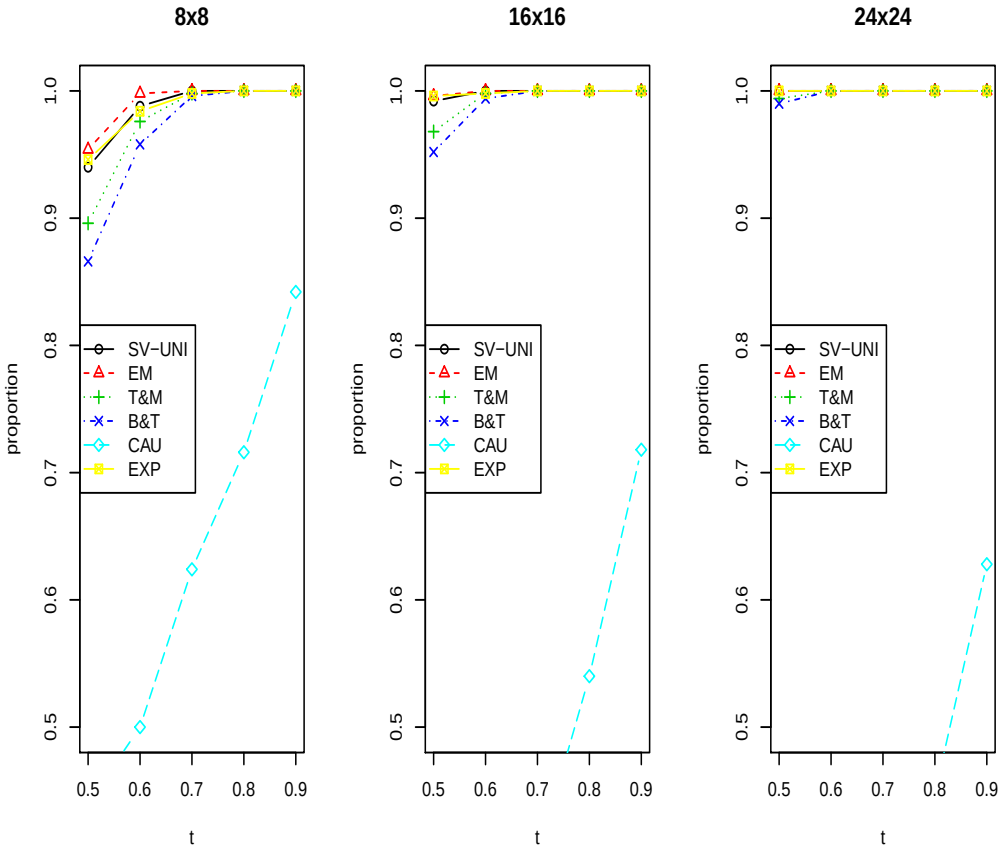
	SV	EV	T&M	B&T	CAU	UNI	EXP
k=8	0.962	0.978	0.914	0.882	0.444	0.962	0.968
k=16	0.984	0.988	0.968	0.954	0.208	0.984	0.994
k=24	1	1	0.994	0.986	0.170	1	1

method, we computed the proportion of trials in which alternative A_1 ranked first (Table 4.5).

The Cauchy method has again the worst performance with very low proportions of always preferred alternative classified as first and the eigenvector method is the best performer. Increasing k, the matrix dimensionality, all the methods improve their performance except the Cauchy one that aggravates its results. Among the linear models, the exponential one shows the best performance. For $k = 24$ some methods recognize, in all the performed replications, the always preferred alternative at the first position. In particular, we are referring to the score vector and the eigenvector methods and the uniform and exponential linear models.

In order to better detecting the ability of the considered methods in recognizing the dominant alternative, we allowed the probability of preferring the first alternative to vary. We simulated some matrices imposing that the elements of the first row are greater than a treshold t with t=(0.5,0.6,0.7,0.8,0.9). So, for each matrix we obtained $k - 1$ observations from the continuous uniform variable $U_2 \sim \mathcal{U}(t, 1)$ that constituted the first dominant row, while the other elements of the upper triangular block of the matrix were obtained at random as before. The lower triangular matrix is filled using the relation $p_{ij} = 1 - p_{ji}$. We simulated $n = 500$ matrices for each value of t. In Figure 4.1 we can observe the resulting proportion of well classified first alternative, on the y-axis, for different values of the treshold t (on the x-axis). Three values of k are considered $(8, 16, 24)$.

Figure 4.1: Proportions of first alternative well classified for each method for $n = 500$ replications and different values of the treshold t



Increasing the treshold probability t all the methods improve their performance. For example, for $k = 8$ and $t \geq 0.7$ all the methods but the Cauchy one classify the first alternative A_1 as the best alternative in all the replications. Cauchy method deserves a separate discussion, since it is the only method whose performance deteriorates as k, the matrix dimensionality, increases and that for high values of t does not achieve satisfactory results. Among the others, eigenvector method seems to have the best performance and the Bradley & Terry approach the worst.

4.1.3 Strictly Ordered Matrices

Now we consider the case in which alternative A_1 is preferred to all the others, alternative A_2 is preferred to all the alternatives but A_1, alternative A_3 is preferred to all the alternatives but A_1 and A_2, and so on. In this situation, the consistent ranking should be from first to last $A_1, A_2, \ldots, A_k$. Indeed, for a ranking to be consistent, if p_{ij} is greater than 0.5, alternative A_i should be ranked higher than A_j.

For each matrix we obtain $k(k-1)/2$ observations $\mathbf{u} = (u_1, \ldots, u_{k(k-1)/2})$ from a continuous uniform variable $U \sim \mathcal{U}(0.5, 1)$. These elements will constitute the upper triangular block of the matrix while the corresponding lower triangular matrix is filled using the relation $p_{ji} = 1 - p_{ij}$. $\boldsymbol{P}$ shows how such a matrix is constructed.

$$\boldsymbol{P} = \begin{bmatrix} 0 & u_1 & u_2 & \ldots & \ldots & \ldots & u_{k-1} \\ 1-u_1 & 0 & u_k & u_{k+1} & \ldots & \ldots & \ldots \\ 1-u_2 & 1-u_k & 0 & \ldots & \ldots & \ldots & \ldots \\ \ldots & 1-u_{k+1} & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & 0 & u_{k(k-1)/2} \\ 1-u_{k-1} & \ldots & \ldots & \ldots & \ldots & 1-u_{k(k-1)/2} & 0 \end{bmatrix}$$

Just as an example, consider the following matrix $\boldsymbol{P_1}$

$$\boldsymbol{P_1} = \begin{bmatrix} 0.000 & 0.508 & 0.645 & 0.900 & 0.898 & 0.780 & 0.656 & 0.842 \\ 0.492 & 0.000 & 0.937 & 0.656 & 0.867 & 0.616 & 0.742 & 0.746 \\ 0.355 & 0.063 & 0.000 & 0.736 & 0.670 & 0.775 & 0.696 & 0.723 \\ 0.100 & 0.344 & 0.264 & 0.000 & 0.835 & 0.504 & 0.831 & 0.640 \\ 0.102 & 0.133 & 0.330 & 0.165 & 0.000 & 0.910 & 0.870 & 0.907 \\ 0.220 & 0.384 & 0.225 & 0.496 & 0.090 & 0.000 & 0.973 & 0.996 \\ 0.344 & 0.258 & 0.304 & 0.169 & 0.130 & 0.027 & 0.000 & 0.894 \\ 0.158 & 0.254 & 0.277 & 0.360 & 0.093 & 0.004 & 0.106 & 0.000 \end{bmatrix}$$

The upper triangular matrix elements are all greater than 0.5. Table 4.6 shows the corresponding rankings.

Table 4.6: Ranking obtained with the different methods applied to matrix $\boldsymbol{P_1}$

	SV	EV	T&M	B&T	CAU	UNI	EXP
A_1	1	1	1	1	2	1	1
A_2	2	2	2	2	3	2	2
A_3	3	3	4	4	6	3	3
A_4	4	4	6	6	5	4	5
A_5	5	6	5	5	4	5	6
A_6	6	5	3	3	1	6	4
A_7	7	7	7	7	7	7	7
A_8	8	8	8	8	8	8	8

As we can see, in this example, only the Cauchy ranking method does not put A_1 in the first position. More in general only the score vector and the uniform linear model recognize the consistent ranking $A_1, A_2, \ldots, A_k$. We compute the Spearman correlation coefficients between the rankings obtained with the different methods and the consistent ranking. Spearman's ρ treats all ranks equally, to give more importance to top ranks than lower ones, we computed also a weighted version of the Spearman's ρ correlation coefficient, named r_w (Pinto da Costa & Soares, 2005) given by

$$r_w = 1 - \frac{6 \sum_i (y_{mi} - y_{ni})(2k - y_{mi} - y_{ni} + 2)}{k^4 + k^3 - k^2 - k}$$

where y_{mi} and y_{ni} are the i-th ranks obtained respectively with method m and n. This correlation coefficient weights ranks in proportion to how high they are, assuming that the higher rank is 1, and that it corresponds to the best element in the ranking. The weighted correlation coefficient r_w can take values in the range $[-1, 1]$; r_w equals 1 if the rankings are same,

$r_w = -1$ if they are in perfect disagreement and $r_w = 0$ indicates that there is no relationship. Table 4.7 shows the weighted and unweighted correlation coefficients referred to this example.

Table 4.7: Spearman's ρ and weighted correlation coefficient r_w between ranking methods and "true" ranking obtained for matrix $\boldsymbol{P_1}$

	SV	EV	T&M	B&T	CAU	UNI	EXP
ρ	1	0.976	0.833	0.833	0.548	1	0.929
r_w	1	0.981	0.836	0.836	0.468	1	0.937

The score vector method and the uniform linear model are perfectly correlated. Among the others, the method most correlated with the consistent ranking is the eigenvector ($\rho = 0.976, r_w = 0.981$) and the least correlated method is once again the Cauchy model ($\rho = 0.548, r_w = 0.468$).
To understand better the behavior of the ranking methods in the hypothized situation, we performed 500 simulations for different values of k, in particular $k = 8, 16, 24$. For each matrix, we computed the rankings with the different methods and we measured the correlation between these rankings and the "consistent" ranking (Tables 4.8, 4.9, 4.10).

Table 4.8: Mean Spearman's $\bar{\rho}$ and mean weighted correlation coefficient, $\bar{r}_w$, between ranking methods and true ranking over 500 replications and $k = 8$ (s.e. in parenthesis)

	SV	EV	T&M	B&T	CAU	UNI	EXP
$\bar{\rho}$	0.954	0.942	0.926	0.914	0.738	0.954	0.957
	(0.037)	(0.046)	(0.054)	(0.061)	(0.152)	(0.037)	(0.037)
$\bar{r}_w$	0.954	0.952	0.926	0.914	0.740	0.954	0.958
	(0.040)	(0.038)	(0.060)	(0.067)	(0.158)	(0.040)	(0.040)

All the methods but Cauchy provide good results, with mean correlation and mean weighted correlation coefficients larger than 0.9. Also in this case, then,

Cauchy model has the worst performance. Among the more consistent methods, the exponential linear model is the most correlated with the consistent ranking. The weighted correlation coefficient does not change substantially the results.

Table 4.9: Mean Spearman's $\bar{\rho}$ and mean weighted correlation coefficient, $\bar{r}_w$, between ranking methods and true ranking over 500 replications and $k = 16$ (s.e. in parenthesis)

	SV	EV	T&M	B&T	CAU	UNI	EXP
$\bar{\rho}$	0.974	0.966	0.957	0.947	0.733	0.974	0.974
	(0.012)	(0.016)	(0.020)	(0.023)	(0.110)	(0.012)	(0.014)
$\bar{r}_w$	0.974	0.974	0.957	0.947	0.734	0.974	0.974
	(0.013)	(0.013)	(0.021)	(0.025)	(0.121)	(0.013)	(0.015)

Increasing k, all the methods improve their results, except the Cauchy linear model which aggravate its results. The methods most correlated with the consistent ranking are the score vector and the eigenvector methods and the uniform and exponential linear models.

Table 4.10: Mean Spearman's $\bar{\rho}$ and mean weighted correlation coefficient $\bar{r}_w$, between ranking methods and true ranking over 500 replications and $k = 24$ (s.e. in parenthesis)

	SV	EV	T&M	B&T	CAU	UNI	EXP
$\bar{\rho}$	0.982	0.976	0.970	0.964	0.741	0.982	0.982
	(0.007)	(0.009)	(0.011)	(0.013)	(0.087)	(0.007)	(0.007)
$\bar{r}_w$	0.982	0.982	0.970	0.964	0.741	0.982	0.982
	(0.008)	(0.007)	(0.012)	(0.015)	(0.093)	(0.008)	(0.008)

Considering all the k values, only for the eigenvector method, the weighted

correlation coefficient registers a slight improvement compared to the Spearman's ρ.

4.2 Simulating the Y_i's

Linear models assume that every alternative A_i $(i = 1, \ldots, k)$ has a true merit V_i which determines its position in their ranking. In a complete balanced paired comparison experiment n respondents perform their comparisons indipendently. So, for each alternative, the judgment process can be represented by a continuous variable Y_i with mean V_i. We would like to check whether the ranking methods are affected by the distribution of the Y_i's, for $i = 1, \ldots, k$ or by the closeness of the alternatives.
For this purposes, we simulated V_i's from a uniform distribution $U_t \sim \mathcal{U}(0, t)$ where $t = (4, 3, 2, 1)$, so that the larger t is, the more the alternatives should be spaced. Then we obtained Y_i's from different distributions (Normal, Gumbel, Cauchy, Uniform and Exponential) and assuming as centrality parameters the V_i's. We describe more in detail the used procedure in the following:

- We simulated $k = 8$ values from a uniform variable $U \sim \mathcal{U}(0, t)$. The values, sorted in decreasing order, constitute the true ratings V_i's of the k alternatives. Thus V_1 is the highest rating and the associated alternative A_1 should be ranked first;

- For each alternative A_i, $i = 1, \ldots, k$, we obtained $n = 100$ replications from a continuous variable Y_i with normal distribution, mean V_i and unit variance;

- Each matrix element p_{ij} is obtained as a_{ij}/n where a_{ij} is the number of times Y_i is greater than Y_j in the simulated sample;

- We get $m = 500$ matrices following the above procedure. For each of them we computed the rankings based on the selected methods;

- For each method, we computed $\bar{\rho}$ and $\bar{r}_w$, the mean values of Spearman's and the mean weighted correlation coefficients, between the corresponding ranking and the "true ranking", by construction, $A_1, \ldots, A_k$;
- This procedure is reiterated changing the distribution of Y_i at point two of the list.

4.2.1 Normal Distribution

Table 4.11 shows the mean values of Spearman's correlation, $\bar{\rho}$, and weighted correlation coefficients, $\bar{r}_w$, computed over 500 replications using the procedure already explained. V_i's values for $i = 1, \ldots, 8$ are obtained from a uniform variable $U_t \sim \mathcal{U}(0, t)$. Y_i's are normally distributed with mean V_i and unit variance.

Table 4.11: Mean Spearman's, $\bar{\rho}$, and mean weighted correlation coefficient, $\bar{r}_w$, for $Y_i \sim \mathcal{N}(V_i, 1)$ (s.e. in parenthesis)

t		SV	EV	T&M	B&T	CAU	UNI	EXP
4	$\bar{\rho}$	0.978	0.976	0.974	0.972	0.957	0.978	0.973
		(0.030)	(0.031)	(0.031)	(0.032)	(0.048)	(0.030)	(0.033)
	$\bar{r}_w$	0.978	0.977	0.974	0.973	0.958	0.978	0.973
		(0.030)	(0.031)	(0.031)	(0.033)	(0.054)	(0.030)	(0.034)
3	$\bar{\rho}$	0.968	0.967	0.967	0.966	0.959	0.968	0.964
		(0.037)	(0.038)	(0.038)	(0.039)	(0.049)	(0.037)	(0.039)
	$\bar{r}_w$	0.968	0.967	0.966	0.966	0.959	0.968	0.964
		(0.041)	(0.042)	(0.043)	(0.044)	(0.055)	(0.041)	(0.043)
2	$\bar{\rho}$	0.947	0.946	0.946	0.946	0.944	0.947	0.944
		(0.058)	(0.058)	(0.058)	(0.058)	(0.058)	(0.057)	(0.060)
	$\bar{r}_w$	0.948	0.947	0.947	0.947	0.945	0.948	0.946
		(0.060)	(0.061)	(0.061)	(0.061)	(0.062)	(0.060)	(0.061)
1	$\bar{\rho}$	0.882	0.881	0.880	0.880	0.880	0.882	0.871
		(0.108)	(0.108)	(0.108)	(0.108)	(0.109)	(0.108)	(0.116)
	$\bar{r}_w$	0.886	0.885	0.884	0.884	0.883	0.886	0.875
		(0.106)	(0.107)	(0.107)	(0.107)	(0.109)	(0.106)	(0.115)

The closer the alternatives, decreasing t, the larger the difficulties of meth-

ods in identifying the true ranking. The weighted version of the correlation coefficient gives substantially the same results. Anyhow, all methods provide rankings highly correlated with the true ranking and results which are very similar.

4.2.2 Gumbel Distribution

Let us assume Y_i's are continuous variables with a Gumbel distribution, with parameters $(V_i, 1)$ which correspond to a mean equalling $V_i+\gamma$, where γ is the Eulero-Mascherano constant, and a variance equal to $\pi^2/6$. The mean values of Spearman's correlation coefficients, $\bar{\rho}$, and the mean values of weighted correlation coefficients, $\bar{r}_w$, between the rankings obtained with the examined methods and the true ranking are presented in Table 4.12. As previously, we consider different values of $t = 4, 3, 2, 1$ and $k = 8$.

Table 4.12: Mean Spearman's $\bar{\rho}$ and mean weighted correlation coefficient $\bar{r}_w$ for $Y_i \sim Gum(V_i, 1)$ (s.e. in parenthesis)

t		SV	EV	T&M	B&T	CAU	UNI	EXP
4	$\bar{\rho}$	0.974	0.970	0.971	0.970	0.959	0.974	0.972
		(0.031)	(0.036)	(0.035)	(0.036)	(0.046)	(0.031)	(0.034)
	$\bar{r}_w$	0.975	0.972	0.973	0.972	0.963	0.975	0.973
		(0.033)	(0.036)	(0.036)	(0.036)	(0.044)	(0.033)	(0.036)
3	$\bar{\rho}$	0.964	0.962	0.963	0.962	0.957	0.964	0.960
		(0.042)	(0.043)	(0.042)	(0.043)	(0.045)	(0.042)	(0.043)
	$\bar{r}_w$	0.965	0.963	0.964	0.963	0.959	0.965	0.960
		(0.042)	(0.043)	(0.042)	(0.043)	(0.047)	(0.042)	(0.044)
2	$\bar{\rho}$	0.943	0.941	0.942	0.942	0.941	0.943	0.941
		(0.061)	(0.062)	(0.062)	(0.061)	(0.063)	(0.061)	(0.058)
	$\bar{r}_w$	0.944	0.943	0.943	0.943	0.942	0.944	0.941
		(0.067)	(0.067)	(0.067)	(0.067)	(0.068)	(0.067)	(0.064)
1	$\bar{\rho}$	0.867	0.866	0.867	0.867	0.867	0.867	0.861
		(0.112)	(0.112)	(0.112)	(0.112)	(0.111)	(0.112)	(0.110)
	$\bar{r}_w$	0.869	0.867	0.868	0.868	0.868	0.869	0.861
		(0.114)	(0.114)	(0.114)	(0.114)	(0.113)	(0.114)	(0.115)

As the closeness among the alternatives increases, all the methods worsen their performance. We do not observe substantial differences in the results from the normal assumption case (Table 4.11), but a slight deterioration appears for all methods.

4.2.3 Cauchy Distribution

Table 4.13 shows the mean values of Spearman's correlation coefficients, $\bar{\rho}$, and the mean values of weighted correlation coefficients, $\bar{r}_w$, computed over 500 replications obtained following the usual procedure. In this case Y_i are continuous variables with a Cauchy distribution, location parameter V_i and scale parameter 1.

Table 4.13: Mean Spearman's $\bar{\rho}$ and mean weighted correlation coefficient $\bar{r}_w$ for $Y_i \sim Cau(V_i, 1)$ (s.e. in parenthesis)

t		SV	EV	T&M	B&T	CAU	UNI	EXP
4	$\bar{\rho}$	0.949	0.949	0.947	0.947	0.942	0.949	0.948
		(0.053)	(0.049)	(0.053)	(0.053)	(0.055)	(0.053)	(0.056)
	$\bar{r}_w$	0.949	0.951	0.948	0.947	0.942	0.949	0.949
		(0.057)	(0.053)	(0.057)	(0.057)	(0.059)	(0.057)	(0.057)
3	$\bar{\rho}$	0.936	0.936	0.935	0.934	0.933	0.936	0.937
		(0.056)	(0.056)	(0.058)	(0.058)	(0.058)	(0.056)	(0.058)
	$\bar{r}_w$	0.935	0.937	0.934	0.934	0.933	0.935	0.937
		(0.059)	(0.059)	(0.061)	(0.061)	(0.061)	(0.059)	(0.062)
2	$\bar{\rho}$	0.894	0.892	0.893	0.893	0.892	0.894	0.893
		(0.095)	(0.097)	(0.096)	(0.096)	(0.096)	(0.095)	(0.094)
	$\bar{r}_w$	0.896	0.895	0.895	0.895	0.895	0.896	0.895
		(0.099)	(0.100)	(0.100)	(0.100)	(0.100)	(0.099)	(0.096)
1	$\bar{\rho}$	0.778	0.778	0.777	0.777	0.778	0.778	0.770
		(0.165)	(0.166)	(0.166)	(0.166)	(0.166)	(0.165)	(0.168)
	$\bar{r}_w$	0.779	0.778	0.777	0.777	0.778	0.779	0.770
		(0.173)	(0.175)	(0.175)	(0.175)	(0.175)	(0.173)	(0.177)

As the alternatives get closer, both the unweighted and the weighted correlation coefficients decrease. With respect the other distribution assumptions,

we can globally observe lower values of both the correlation coefficients.

4.2.4 Uniform Distribution

Table 4.14 shows the mean of Spearman's correlation coefficients, $\bar{\rho}$, and the mean values of weighted correlation coefficients, $\bar{r}_w$, obtained over 500 replications for $k = 8$ and different values of t. Y_i's are assumed to be continuous variables with a uniform distribution, mean V_i and unit variance.

Table 4.14: Mean Spearman's, $\bar{\rho}$, and mean weighted correlation coefficient, $\bar{r}_w$, for $Y_i \sim \mathcal{U}(V_i, 1)$ (s.e. in parenthesis)

	t	SV	EV	T&M	B&T	CAU	UNI	EXP
4	$\bar{\rho}$	0.977	0.977	0.976	0.975	0.967	0.977	0.976
		(0.033)	(0.031)	(0.033)	(0.034)	(0.044)	(0.033)	(0.034)
	$\bar{r}_w$	0.977	0.977	0.976	0.975	0.967	0.977	0.976
		(0.037)	(0.035)	(0.038)	(0.039)	(0.049)	(0.037)	(0.037)
3	$\bar{\rho}$	0.968	0.967	0.968	0.967	0.961	0.968	0.965
		(0.040)	(0.038)	(0.038)	(0.039)	(0.039)	(0.040)	(0.040)
	$\bar{r}_w$	0.968	0.967	0.968	0.967	0.961	0.968	0.965
		(0.043)	(0.042)	(0.042)	(0.043)	(0.044)	(0.043)	(0.043)
2	$\bar{\rho}$	0.950	0.949	0.949	0.949	0.947	0.950	0.946
		(0.050)	(0.051)	(0.051)	(0.051)	(0.054)	(0.050)	(0.054)
	$\bar{r}_w$	0.950	0.950	0.950	0.949	0.947	0.950	0.947
		(0.053)	(0.054)	(0.054)	(0.054)	(0.055)	(0.053)	(0.058)
1	$\bar{\rho}$	0.878	0.877	0.877	0.877	0.877	0.878	0.873
		(0.105)	(0.105)	(0.106)	(0.106)	(0.106)	(0.105)	(0.102)
	$\bar{r}_w$	0.877	0.877	0.877	0.877	0.876	0.877	0.873
		(0.107)	(0.107)	(0.108)	(0.108)	(0.108)	(0.107)	(0.106)

Also in this case, as the closeness of the alternatives increases, both correlation coefficients decrease. We obtained for both the mean of Spearman's correlation coefficients, $\bar{\rho}$, and weighted correlation coefficients, $\bar{r}_w$, values close to those obtained in simulations conducted under the assumptions of, respectively, Normal and Gumbel distributions (Tables 4.11, 4.12) and better

than the case of assumed Cauchy distribtion (Table 4.13).

4.2.5 Exponential Distribution

Suppose Y_i has assumed exponential distribution with mean V_i and variance V_i^2. The variance is not constant for all Y_i's, $(i = 1, \ldots, k)$ but it is proportional to the mean, by hypothesis.

Table 4.15: Mean Spearman's $\bar{\rho}$ and mean weighted correlation coefficient $\bar{r}_w$ for $Y_i \sim Exp(1/V_i)$ (s.e. in parenthesis)

t		SV	EV	T&M	B&T	CAU	UNI	EXP
4	$\bar{\rho}$	0.938	0.939	0.933	0.931	0.904	0.938	0.933
		(0.072)	(0.073)	(0.074)	(0.078)	(0.109)	(0.072)	(0.075)
	$\bar{r}_w$	0.925	0.927	0.919	0.916	0.883	0.925	0.921
		(0.088)	(0.088)	(0.091)	(0.095)	(0.133)	(0.088)	(0.090)
3	$\bar{\rho}$	0.945	0.946	0.940	0.937	0.911	0.945	0.940
		(0.062)	(0.063)	(0.068)	(0.071)	(0.106)	(0.062)	(0.064)
	$\bar{r}_w$	0.935	0.937	0.928	0.925	0.892	0.935	0.930
		(0.076)	(0.076)	(0.082)	(0.085)	(0.127)	(0.076)	(0.077)
2	$\bar{\rho}$	0.942	0.943	0.938	0.935	0.909	0.942	0.938
		(0.068)	(0.067)	(0.070)	(0.072)	(0.102)	(0.068)	(0.071)
	$\bar{r}_w$	0.931	0.932	0.925	0.921	0.889	0.931	0.925
		(0.085)	(0.082)	(0.088)	(0.090)	(0.124)	(0.085)	(0.089)
1	$\bar{\rho}$	0.936	0.938	0.933	0.929	0.900	0.936	0.931
		(0.079)	(0.080)	(0.078)	(0.082)	(0.113)	(0.079)	(0.088)
	$\bar{r}_w$	0.924	0.925	0.919	0.915	0.877	0.924	0.918
		(0.097)	(0.097)	(0.095)	(0.100)	(0.138)	(0.096)	(0.106)

Differently from the other assumed distributions, all ranking methods were not affected by the closeness of the alternatives. All methods provided good results with mean values of Spearman's correlation coefficients, $\bar{\rho} \geq 0.9$ for all t values and methods (Table 4.15). The mean values of weighted correlation coefficients, $\bar{r}_w$, provide slight worse results.

Chapter 5

Estimation of Missing Values in a Matrix

In a paired comparison experiment alternatives are presented in pairs to respondents, who have to indicate the one they like most. Given a set of k alternatives $A = \{A_1, A_2, \ldots, A_k\}$, the results of these evaluations can be naturally recorded in a $k \times k$ matrix, in which the (i, j) entry represents the outcome of the comparison between A_i and A_j. As we have seen, there are various kinds of matrices suitable for this purpose.
When respondents do not compare directly two alternatives, the matrix designated to represent their preferences is characterized by one or more missing values. The topic of missing values in a preference matrix has been extensively discussed in literature (Harker, 1987; Carmone et al., 1997; Kwiesielewicz & Van Uden, 2003; Fedrizzi & Giove, 2007).
In the following section we analyze the main reasons for which some comparisons may be missing.

5.1 Possible Causes of Missing Observations

In Dittrich et al (2012) authors identify six specific types of scenarios that can cause missing data in paired comparison experiments. In the following

we will analyze each of the proposed causes.

1. Missing paired comparisons by design. In a completely balanced paired-comparison experiment, respondents have to compare all possible pairs of alternatives. Indeed, for k alternatives, each respondent have to compare $k(k-1)/2$ pairs of alternatives so that, increasing k, the number of required comparisons increase exponentially. For instance, in a full complete design, for $k = 4$, respondents compare 6 pairs while $k = 15$ provides 105 paired comparisons. If k is large, this technique is unapplicable, that is why received criticism from several researchers (Kendall & Babington-Smith, 1940; David, 1988; Fabbris, 2013). Each respondent has to perform an excessive number of comparisons which leads to an increase of the response error because of respondent's fatigue. To contain this problem, it is necessary to reduce the number of required comparisons by design. The researcher will choose which comparisons have to be performed and which other not.

2. Respondents may not complete the paired comparison experiment, since for different reasons, for instance fatigue or information overload, s/he has to interrupt the procedure after completing only a portion of the paired comparison tasks. Also this situation is motivated by the excessive number of comparisons, but in this case the respondent chooses which comparisons avoiding.

3. The researcher prematurely halts the paired comparison experiment for a particular respondent, since he judges s/he not to take the experiment seriously. We state that the actual problem, in this situation, are not the missing values but the given responses, thought of as improper.

4. Respondent fails to answer to a comparison due to his/her insufficient knowledge of the alternatives being compared. This can happen, for example, when highly technical objects are being compared. Also in this case the matter is how to consider the given responses, since the respondent is considered unreliable.

5. Respondents are unable to compare two alternatives since they cannot distinguish among them. This outcome is likely if two objects are equally preferred, and there is no "no preference" possibility in the response set.

6. The respondent has made a preference choice, but s/he knows that the choice can be socially disapproved and s/he is reluctant to express that preference. An example would be preferring a far right political party to a mainstream party.

Missing data can be classified in three different categories (Schafer & Graham, 2002). Data can miss completely at random (MCAR) if a missing observation for an individual cannot depend either on the value of other dected variables nor on any observed or missing response. Missing data are defined at random (MAR) if a missing observation can depend on the value of the other variables and on observed responses made, but not on the value that would have been observed. Finally, missing are not at random (MNAR) if the missing observation also depends on the value that would have been observed.
Under this classification hypotheses, missing paired comparisons by design can be considered completely at random, since the missing data mechanism is determined before the data collection. The second and third cases, of early interruption of the experiment, can be considered missing at random, as missing data depend on the performed comparisons. Also the fourth scenario is missing at random, since missing data depend on respondents' knowledge. Finally, the last two cases are missing not at random as missing observations depend on the responses that would be given.

5.2 Estimating the Missing Values

Consider the case in which n respondents elicit, independently of each other, their preferences on a set of k alternatives using paired comparisons. Data

collected can be naturally represented in a generalized tournament matrix $\boldsymbol{P} = [p_{ij}]$, a $k \times k$ matrix whose elements p_{ij} can be viewed as the proportions of comparisons in which the i-th alternative A_i is preferred to the j-th. Respondents' preferences are collected in order to determine a ranking or to define a system of weights that reflect the relative importance of the alternatives.

Computerized questionnaires can insist on a respondent to answer a question, so as missing data from the last two scenarios can be avoided. Moreover, data collected on unreliable respondents, or with insufficient knowledge about the alternatives, can be at all not considered. The main cause of missing data, then, remains the excessive number of pairwise comparisons, which occur increasing k, the number of alternatives. Mathematical constraints impose a minimum of only $k-1$ comparisons to establish the ranking for k attributes, but paired comparisons methods use $k(k-1)/2$ comparisons. To semplify the procedure, the number of required comparisons may be reduced so as to be between $k-1$ and $k(k-1)/2$.

In the following sections, we propose two methods to estimate the missing cells in a generalized tournament matrix $\boldsymbol{P}$, within which cells are missing by design. First, we analyze an estimation method associated with the linear models. Then, we propose a procedure based on the minimization of the sum of the singular values of $\boldsymbol{P}$.

5.3 Composition Rules for Linear Models

The linear paired comparison model (David, 1988) assumes that each of the k alternatives has true rating V_i which can be used to determine their ordering. We are considering the case in which n respondents evaluate the alternatives independently. The judgment process may be represented by a continuous variable Y_i with mean V_i whose distribution is called "sensation distribution". In a pairwise comparison between two alternatives A_i and A_j, A_i is preferred to A_j if and only if $Y_i \geq Y_j$. So, each preference probability π_{ij} represents the

probability that Y_i is greater than Y_j and it can be estimated through p_{ij}, the proportion of comparisons in which A_i is preferred to A_j. Let us define Z_i and $Z_i - Z_j$ as follows

- $Z_i = Y_i - V_i$ is a continuous variable with zero mean and the same distribution as Y_i;
- $Z_i - Z_j$ is a symmetric variable with zero mean, whose distribution, called *defining distribution*, depends on the *sensation distribution* of Y_i (Latta, 1979).

We have immediately that $\pi_{ij} = F_{Z_i - Z_j}(V_i - V_j)$, where F is the distribution function of $Z_i - Z_j$.
When respondents do not compare directly two objects, A_i and A_j, linear models permit to estimate the preference relation between them using the information achieved from their comparison with a shared alternative A_l. In practice, it is possible to obtain $\hat{p}_{ij}$ through p_{il} and p_{lj} for $i, j, l = 1, \ldots, k$ and $i \neq j \neq l$.
The function which relates $\hat{p}_{ij}$ with p_{il} and p_{lj}, called composition rule, may be written in a general form as

$$\hat{p}_{ij,l} = G(p_{il}, p_{lj}).$$

For values of p_{il} and $p_{lj} \in (0, 1)$, $G(p_{il}, p_{lj}) \in [0, 1]$. So, under linear models assumptions we have (Latta, 1979)

$$\begin{aligned}
\hat{p}_{ij,l} &= Pr\,(Y_i \geq Y_j) = F(V_i - V_j) \\
&= F(V_i - V_l + V_l - V_j) \\
&= F[F^{-1}(p_{il}) + F^{-1}(p_{lj})] \\
&= G(p_{il}, p_{lj}).
\end{aligned}$$

5.3.1 Composition Rules' Properties

In Latta (1979) composition rules are formally defined and several properties are presented. In the following we list the main ones.

- In generalized tournament matrices $p_{ij} = 1 - p_{ji}$. Then, since

$$\hat{p}_{ij,l} = G(p_{il}, p_{lj})$$
$$\hat{p}_{ji,l} = G(p_{jl}, p_{li}) = G(1 - p_{lj}, 1 - p_{il})$$

  we have that $G(p_{il}, p_{lj}) = 1 - G(1 - p_{lj}, 1 - p_{il})$ for $p_{il}, p_{lj} \in (0, 1)$ and $i \neq l \neq j = 1, \ldots, k$.

- From the previous property it follows that, for $p_{il} \in (0, 1)$

$$\begin{aligned} G(p_{il}, 1 - p_{il}) &= 1 - G(1 - p_{li}, 1 - p_{il}) \\ &= 1 - G(p_{il}, 1 - p_{il}) \\ &= \frac{1}{2} \end{aligned}$$

- Consider a tern of probability estimates p_{ij}, p_{jl}, p_{il} with $i, l, j = 1, \ldots, k$ and $i \neq l \neq j$. Each of them can be expressed as function of the others through the composition rules as follows

$$\hat{p}_{ij,l} = G(p_{il}, 1 - p_{jl}) \Leftrightarrow \hat{p}_{il,j} = G(p_{ij}, p_{jl}) \Leftrightarrow \hat{p}_{jl,i} = G(1 - p_{ij}, p_{il})$$

  Proof of the above property is available in Latta (1979).

- From the last two properties we obtain that, for $p_{lj} \in (0, 1)$

$$G\left(\frac{1}{2}, p_{lj}\right) = \hat{p}_{lj}$$
$$G\left(p_{lj}, \frac{1}{2}\right) = \hat{p}_{lj}$$

  Indeed, let us suppose without loss of generality that

$$\hat{p}_{ij,l} = G\left(\frac{1}{2}, p_{lj}\right).$$

  Then, by previous property

$$G(p_{ij}, p_{jl}) = \frac{1}{2}$$

  from which we obtain $p_{ij} = 1 - p_{jl} = p_{lj}$. Second equality is proved similarly.

- Composition rules are symmetric, monotonic and continuous functions.
- Since $G(p_{il}, p_{lj})$ is a monotonic function and $G\left(\frac{1}{2}, p_{lj}\right) = \hat{p}_{lj}$, we have that if p_{il} and p_{lj} are greater than $\frac{1}{2}$ then $G(p_{il}, p_{lj})$ is greater than the maximum of p_{il} and p_{lj}. For the same reason, if p_{il} and p_{lj} are smaller than $\frac{1}{2}$ then $G(p_{il}, p_{lj})$ is smaller than their minimum.
  These last two properties are particularly meaningful, as they show that linear models' composition rules follow the strong transitivity condition. Let us remember that, for a generalized tournament matrix, we can define different degrees of transitivity. We constate that the strong stochastic transitivity condition holds if, for every triad of alternatives A_i, A_j and A_l, if $p_{il} \geq 0.5$ and $p_{lj} \geq 0.5$, then p_{ij} is greater than the maximum of p_{il} and p_{lj} (Coombs, 1958).

Thurstone & Mosteller Model

Under the Thurstone & Mosteller model assumptions, (Thurstone, 1927c; Mosteller, 1951) p_{ij} can be estimated using p_{il} and p_{lj} through the following formula

$$\hat{p}_{ij,l} = \Phi\left[\Phi^{-1}(p_{il}) + \Phi^{-1}(p_{lj})\right]$$

where Φ is the distribution function of a standard normal variable.

Bradley & Terry Model

Bradley & Terry model (1952) assumes that F is the distribution function of a standard logistic variable, that is

$$F(x) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{2}\right).$$

Inverting this relation we have

$$F^{-1}(p) = \ln\left(\frac{p}{1-p}\right).$$

So the corresponding composition rule is

$$\begin{aligned}
\hat{p}_{ij,l} &= F\left(\ln\left(\frac{p_{il}}{1-p_{il}}\right)+\ln\left(\frac{p_{lj}}{1-p_{lj}}\right)\right)\\
&= \frac{1}{2}+\frac{1}{2}\tanh\left(\frac{1}{2}\ln\left(\frac{p_{il}p_{lj}}{(1-p_{il})(1-p_{lj})}\right)\right)\\
&= \frac{1}{2}+\frac{1}{2}\frac{\exp\left\{\frac{1}{2}\ln\left(\frac{p_{il}p_{lj}}{(1-p_{il})(1-p_{lj})}\right)\right\}-\exp\left\{-\frac{1}{2}\ln\left(\frac{p_{il}p_{lj}}{(1-p_{il})(1-p_{lj})}\right)\right\}}{\exp\left\{\frac{1}{2}\ln\left(\frac{p_{il}p_{lj}}{(1-p_{il})(1-p_{lj})}\right)\right\}+\exp\left\{-\frac{1}{2}\ln\left(\frac{p_{il}p_{lj}}{(1-p_{il})(1-p_{lj})}\right)\right\}}\\
&= \frac{1}{2}+\frac{1}{2}\frac{p_{il}p_{lj}-(1-p_{il})(1-p_{lj})}{p_{il}p_{lj}+(1-p_{il})(1-p_{lj})}\\
&= \frac{p_{il}p_{lj}}{p_{il}p_{lj}+(1-p_{il})(1-p_{lj})}.
\end{aligned}$$

Cauchy Model

Cauchy model assumes that F is the distribution function of a Student's T variable with 1 degree of freedom. Since

$$F(x)=\frac{1}{2}+\frac{1}{\pi}\arctan(x)$$

the corresponding composition rule is given by

$$\begin{aligned}
\hat{p}_{ij,l} &= F(F^{-1}(p_{il})+F^{-1}(p_{lj}))\\
&= F(\tan(\pi(p_{il}-1/2))+\tan(\pi(p_{lj}-1/2))\\
&= \frac{1}{2}+\frac{1}{\pi}\arctan\left[\tan\left(\pi\,p_{il}-\frac{\pi}{2}\right)+\tan\left(\pi\,p_{lj}-\frac{\pi}{2}\right)\right].
\end{aligned}$$

Latta (1979) suggests to equale $\hat{p}_{ij,l}$ to 1 if $\sqrt{2(1-p_{il})}+\sqrt{2(1-p_{lj})}\leq 1$.

Uniform Model

Suppose F is the distribution function of a continuous uniform variable $U\in\left(-\frac{1}{2},\frac{1}{2}\right)$ so that

$$F(x)=x+\frac{1}{2}.$$

The corresponding composition rule results

$$\begin{aligned}\hat{p}_{ij,l} &= F(F^{-1}(p_{il}) + F^{-1}(p_{lj})) \\ &= F(p_{il} - \frac{1}{2} + p_{lj} - \frac{1}{2}) \\ &= p_{il} + p_{lj} - \frac{1}{2}.\end{aligned}$$

Latta (1979) proposed the following composition rule for the uniform model:

$$\begin{aligned}\hat{p}_{ij,l} &= p_{il} + p_{lj} - \frac{3}{2} + \sqrt{2(1-p_{il})} + \sqrt{2(1-p_{lj})} - 2\sqrt{(1-p_{il})(1-p_{lj})} \\ &\qquad \text{if} \qquad p_{il} \geq \frac{1}{2}, p_{lj} \geq \frac{1}{2} \text{ and } \sqrt{2(1-p_{il})} + \sqrt{2(1-p_{lj})} \geq 1 \\ &= p_{il} - p_{lj} - \frac{1}{2} + \sqrt{2p_{lj}} - \sqrt{2(1-p_{il})} + 2\sqrt{p_{lj}(1-p_{il})} \\ &\qquad \text{if} \qquad 1 > p_{il} \geq 1 - p_{lj} \geq \frac{1}{2}.\end{aligned}$$

Latta's composition rule is only partially defined. In the following cases

- $p_{il}, p_{lj} \leq \frac{1}{2}$;
- $1 - p_{lj} \geq p_{il} \geq \frac{1}{2}$;
- $p_{il} \leq 1 - p_{lj} \leq \frac{1}{2}$;
- $1 - p_{lj} \leq p_{il} \leq \frac{1}{2}$

it can be constructed using the following property

$$G(p_{il}, p_{lj}) = 1 - G(1 - p_{lj}, 1 - p_{il}). \tag{5.1}$$

Exponential Model

The Exponential model assumes that F is the distribution function of a standard Laplace variable. Since

$$F(x) = \frac{1}{2}\left(1 + sign(x)(1 - \exp -|x|)\right)$$

we have that

$$F^{-1}(p) = -sign\left(p - \frac{1}{2}\right) \ln\left(1 - 2\left|p - \frac{1}{2}\right|\right).$$

The resulting composition rule is

$$\begin{aligned} \hat{p}_{ij,l} &= 1 - 2(1 - p_{il})(1 - p_{lj}) && \text{if } p_{il} \geq \frac{1}{2}, p_{lj} \geq \frac{1}{2} \\ &= 1 - \frac{1}{2}\left(\frac{1 - p_{il}}{p_{lj}}\right) && \text{if } p_{il} \geq 1 - p_{lj} \geq \frac{1}{2}. \end{aligned}$$

Similarly to the uniform model, the composition rule of the exponential model is only partially defined and can be reconstructed using property 5.1.

5.4 Rank Minimization Problems

Matrix completion problems have received a great deal of attention in the algebraic literature (Laurent, 2001; Lee & Seol, 2001; Cravo, 2009). These kind of problems arise in a variety of applications, such as statistics, chemistry and systems theory. Matrix completion issue tries to answer whether a given partial matrix can be completed according to specified rules. For example, the positive definite completion problem asks which partial Hermitian matrices have a positive definite completion. A variety of matrix properties have been studied. Laurent (2001) considers the following matrix completions: positive (semi)definite matrices, distance matrices, completely positive matrices, contraction matrices and matrices of given rank.

Rank matrix completion problems are concerned with determining whether or not a partial matrix can be completed so that its rank is maximized or minimized. Searching for the completion which minimizes the rank of the matrix corresponds, intuitively, to complete the matrix with values that best fit the observed data.

Our aim is to recover an incomplete generalized tournament matrix $\boldsymbol{P}$ which represent the preference relations among a set of k alternatives. Preference data could be modeled through factor models, which assume that only a very small number of factors influences the preferences. Computationally, factor models are equivalent to low rank approximation of the matrix of observed data. So, assuming that only a few dimensions contribute to individual preferences, our goal is to produce a low-rank matrix that respects the observed

elements of $\boldsymbol{P}$ or at least minimizes the deviation from them.
Unfortunately, the rank minimization problem is unsolvable and all known algorithms that provide an exact solution require a computing time doubly exponential in the dimension k of the matrix in both theory and practice (Chistov & Grigor'ev, 1984).
Let us present a recent heuristic introduced by Fazel et al. (2001), that minimizes the nuclear norm $\|\boldsymbol{P}\|_*$ over the constraint set. The nuclear norm, also known as "trace norm", is defined as

$$\|\boldsymbol{P}\|_* = \sum_{i=1}^{k} \sigma_i.$$

where σ_i for $i = 1, \ldots, k$ are the singular values of $\boldsymbol{P}$, defined in the following section.

5.4.1 Singular Value Decomposition

Given a $k \times k$ generalized tournament matrix $\boldsymbol{P}$, the singular value decomposition is a factorization of the form

$$\boldsymbol{P} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

where $\boldsymbol{U}$ is a $k \times k$ real unitary matrix, such that $\boldsymbol{U}^t\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^t = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{\Sigma}$ is a $k \times k$ diagonal matrix with nonnegative elements on the diagonal and $\boldsymbol{V}^T$ is another $k \times k$ real unitary matrix. We name the diagonal entries σ_i of $\boldsymbol{\Sigma}$ as the singular values of $\boldsymbol{P}$, the k columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are respectively the left-singular and the right-singular vectors of $\boldsymbol{P}$. The singular value decomposition and the eigen-decomposition of a matrix are closely related. Namely:

- The left-singular vectors of $\boldsymbol{P}$ are eigenvectors of $\boldsymbol{P}\boldsymbol{P}^T$;
- The right-singular vectors of $\boldsymbol{P}$ are eigenvectors of $\boldsymbol{P}^T\boldsymbol{P}$;
- The non-zero-singular values of $\boldsymbol{P}$, found on the diagonal entries of Σ are the square roots of the non-zero eigenvalues of both $\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{P}$ and $\boldsymbol{P}\boldsymbol{P}^{\boldsymbol{T}}$.

Nuclear Norm Minimization

Assuming that the generalized tournament matrix $\boldsymbol{P}$ we want to recover can be well approximated by a low-rank matrix, we consider the missing entries in $\boldsymbol{P}$ as variables $x_1, \ldots, x_p$ and we obtain them by minimizing the nuclear norm as a function of these variables. To give an example consider the following 4×4 matrix $\boldsymbol{P_1}$ which represents a paired comparison experiment performed by n respondents on 4 alternatives. Let us suppose that individuals make 11 of the 12 possible comparisons. Supposing, without loss of generality, that the comparison between the first and the fourth alternative was not performed, $\boldsymbol{P}$ has two missing entries which correspond to positions $(1, 4)$ and $(4, 1)$.

$$\boldsymbol{P} = \begin{bmatrix} 0 & p_{12} & p_{13} & x_1 \\ 1 - p_{12} & 0 & p_{23} & p_{24} \\ 1 - p_{13} & 1 - p_{23} & 0 & p_{34} \\ 1 - x_1 & 1 - p_{24} & 1 - p_{34} & 0 \end{bmatrix}$$

Hence, we wish to recover the data matrix by solving the following optimization problem

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{X}\|_* \\ \text{subject to} \quad & x_{ij} = p_{ij} \quad (i,j) \in \Omega \\ & \boldsymbol{X} + \boldsymbol{X^T} = \boldsymbol{J} - \boldsymbol{I} \end{aligned}$$

where $\boldsymbol{P}$ is the matrix expected to recover, $\boldsymbol{X}$ is the matrix after recovering and Ω is the set of positions corresponding to the observed entries, that is $(i, j) \in \Omega$ if p_{ij} is observed. Moreover, $\boldsymbol{J}$ is a square matrix of all 1's and $\boldsymbol{I}$ is the identity matrix and the second constraint sets the structure of a generalized tournament matrix.

The values that minimize the nuclear norm are not necessarily those that minimize the rank, but they are consistent with the hypothesis of a good approximation of a low-rank matrix. Indeed, it can be shown that the rank

of a generalized tournament matrix $\boldsymbol{P}$ of order k is either k or $k-1$ (Brualdi, 2006).

The nuclear norm is a convex function, which can be optimized efficiently. It is the best convex lower approximation of the rank function over the set of matrices with spectral norm less than or equal to one. Intuitively, while rank counts the number of nonvanishing singular values, nuclear norm sums their amplitude. In case of symmetric and positive semidefinite matrix, nuclear norm minimization is equivalent to the trace minimization, the trace of a positive semidefinite matrix being the sum of its singular values.

Nuclear norm minimization is often used to recover a matrix from a sample of its entries, provided the matrix is low-rank or approximately low-rank (Candès & Recht, 2009). Applications include dimensionality reduction (Linial et al., 1995; Weinberger & Saul, 2004), inference with partial information (Rennie & Srebro, 2005). Rank minimization plays a key role in the study of embeddings of discrete metric spaces in Euclidean space (Donoho & Tanner, 2005). In some applications, such as sensor localization, the matrix has exactly low rank, i.e., only the top few of its singular values are nonnull. However, the matrix can be full rank, and well approximated by a low-rank matrix.

Differently from the case in which observed entries are selected at random and no information is available about the matrix to be recovered, we know that diagonal elements of a generalized tournament matrix $\boldsymbol{P}$ are ever equal to 0. Moreover, we have to consider that if p comparisons are missing, then the matrix $\boldsymbol{P}$ contains $2p$ unknown elements, due to the reciprocity condition $p_{ij} = 1 - p_{ji}$.

5.5 The Single Missing Comparison Case

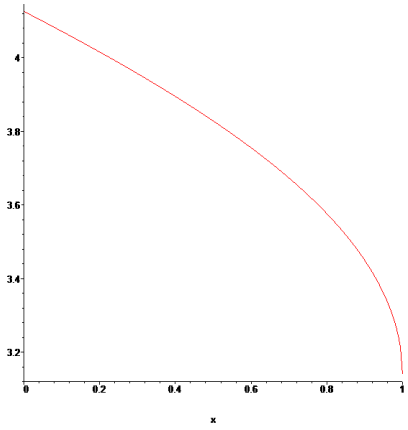
Let us assume that only the entry that defines the preferability between alternatives A_i and A_j, $i \neq j$, is missing. Then p_{ij} and $p_{ji} = 1 - p_{ij}$ are

unknown in the $k \times k$ matrix $\boldsymbol{P}$. Consider the simple case represented in the following matrix according to which A_1 is preferred to A_2 and A_2 is preferred to A_3, while nothing is known about the preferability between A_1 and A_3.

$$\boldsymbol{P_1} = \begin{bmatrix} 0 & 1 & x \\ 0 & 0 & 1 \\ 1-x & 0 & 0 \end{bmatrix}$$

For preferences to be internally consistent, alternative A_1 should be preferred to A_3, and then, x should be equal to 1. In Fig 5.1 it is represented the nuclear norm of $\boldsymbol{P_1}$ for $x \in [0, 1]$. As we can see, the nuclear norm reaches its minimum, as expected, if $x = 1$.

Figure 5.1: Nuclear norm of $\boldsymbol{P_1}$

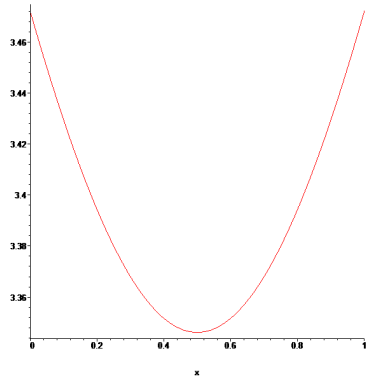


Suppose otherwise that, matrix $\boldsymbol{P_2}$ represents the situation in which both A_1 and A_3 are absolutely preferred to A_2. In this case, it is not possible to express a forecast on the preference relation between A_1 and A_3.

$$\boldsymbol{P_2} = \begin{bmatrix} 0 & 1 & x \\ 0 & 0 & 0 \\ 1-x & 1 & 0 \end{bmatrix}$$

Also in this case, we represent the nuclear norm for $x \in [0, 1]$ (See Fig. 5.2). As we can see, $\|\boldsymbol{P_2}\|_*$ reaches its minimum in $x = 0.5$, which means that it is impossible to express a preference in any direction.

Figure 5.2: Nuclear norm of $\boldsymbol{P_2}$



5.5.1 Comparison between Composition Rules and Nuclear Norm Minimization

To better understand the nuclear norm minimization procedure, we compare, via simulation, its behavior with linear models' composition rules. Just to give an example, consider the following matrix $\boldsymbol{P_3}$, representing a paired comparison among 4 alternatives, obtained assuming Y_i's normally distributed with mean V_i's $\in (0, 1)$ and unit variance. In particular, given $V = (0.46, 0.60, 0.75, 0.88)$, we get

$$\boldsymbol{P_3} = \begin{bmatrix} 0.00 & 0.47 & 0.44 & 0.36 \\ 0.53 & 0.00 & 0.44 & 0.40 \\ 0.56 & 0.56 & 0.00 & 0.44 \\ 0.64 & 0.60 & 0.56 & 0.00 \end{bmatrix}.$$

To verify the different methods reliability, we simulate the darkening of p_{14}, and of its symmetric entry $p_{41} = 1 - p_{14}$, and we try to estimate p_{14} as if they were missing.

Linear models permit to estimate p_{ij} using the information achieved from the comparison of A_i and A_j with a shared alternative A_l with $i \neq j \neq l = 1, \ldots, k$. So, in this case, we can estimate the preference relation between A_1 and A_4 using the information achieved from their comparisons with, respectively, A_2

and A_3. So, it is possible obtain $\hat{p}_{14,l}$ through the following

- $\hat{p}_{14,2} = G(p_{12}, p_{24})$;
- $\hat{p}_{13,4} = G(p_{13}, p_{34})$.

The final estimate of p_{ij} is the mean of all the possible estimates $\hat{p}_{ij,l}$ with $i \neq j \neq l = 1, \ldots, k$

$$\hat{p}_{ij} = \frac{1}{k-2} \sum_{l \neq i,j} \hat{p}_{ij,l}.$$

In this specific example, the estimate of p_{14} is given by

$$\hat{p}_{14} = \frac{1}{2} \left(\hat{p}_{14,2} + \hat{p}_{14,3} \right).$$

Finally, p_{14} is estimated through the nuclear norm minimization procedure (NNM), imposing $p_{14} = x$ and solving the usual optimization problem

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{X}\|_* \\ \text{subject to} & x_{ij} = p_{ij} \quad (i,j) \in \Omega \\ & \boldsymbol{X} + \boldsymbol{X}^{\boldsymbol{T}} = \boldsymbol{J} - \boldsymbol{I} \end{array}$$

where $\boldsymbol{P_3}$ is the matrix we would recover, $\boldsymbol{X}$ is the matrix once recovered and Ω is the set of positions corresponding to the observed entries, that is $(i,j) \in \Omega$ if p_{ij} is observed. The second constraint imposes that $p_{12} = 1 - p_{21}$. Among the linear models, we consider the following

- Thurstone & Mosteller (T & M)
- Bradley & Terry (B & T)
- Cauchy (CAU)
- Uniform (UNI)
- Exponential (EXP)

Table 5.1: p_{14} estimates

	T&M	B&T	CAU	UNI	EXP	NNM
$\hat{p}_{14}$	0.376	0.377	0.379	0.385	0.382	0.375

Table 5.1 shows the $\hat{p}_{14}$ obtained with the examined estimation methods. Once estimated the missing entries, to evaluate their goodness, we can compute the mean square error, defined as

$$MSE = E[(p_{14} - \hat{p}_{14})^2].$$

Table 5.2 shows the mean square error referred to this specific example. As we can see, the nuclear norm minimization provides good results, even better than the Thurstone & Mosteller linear model, although $\boldsymbol{P_3}$ was created under its assumptions.

Table 5.2: Mean square errors of $\hat{p}_{14}$

	T&M	B&T	CAU	UNI	EXP	NNM
MSE	0.00027	0.00028	0.00035	0.00060	0.00047	0.00023

To extend our considerations, we obtained 500 matrices assuming different distributions for the Y_i's and $k = 8$. The simulation scheme is described as follows:

- We simulated $k = 8$ values from a uniform variable $U \sim \mathcal{U}(0, 1)$;
- For each alternative A_i, $i = 1, \ldots, k$, we obtained $n = 100$ replications from a continuous variable Y_i with normal distribution, centrality parameters V_i and unit variance;
- Each matrix element p_{ij} is obtained as a_{ij}/n where a_{ij} is the number of times Y_i is greater than Y_j in the simulated sample;

- We get $m = 500$ matrices following the above procedure;
- The procedure is iterated considering other values of k and different distributions for the Y_i's (Normal, Gumbel, Cauchy, Uniform and Exponential).
- The specific assumptions for each assumed distribution are the same as those described Section 4.2.

For each matrix, a cell at a time is assumed missing and its estimate is obtained through the linear models' composition rules and the nuclear norm minimization. Then, for each method, we computed the mean square error over all the 500 replications. Tables 5.3, 5.4, 5.5 show the results obtained assuming respectively $k = 8, 16, 24$. In the first column the assumed distribution is specified.

Table 5.3: Mean square errors of $\hat{p}_{ij}$ over 500 replications and $k = 8$ (s.e. in parenthesis)

Distribution	T&M	B&T	CAU	UNI	EXP	NNM
Normal	0.00123 (0.00207)	0.00123 (0.00208)	0.00127 (0.00212)	0.00127 (0.00210)	0.00126 (0.00211)	0.00125 (0.00209)
Gumbel	0.00107 (0.00150)	0.00107 (0.00150)	0.00111 (0.00154)	0.00111 (0.00157)	0.00112 (0.00154)	0.00106 (0.00148)
Cauchy	0.00113 (0.00159)	0.00113 (0.00159)	0.00114 (0.00160)	0.00112 (0.00159)	0.00115 (0.00162)	0.00113 (0.00158)
Uniform	0.00105 (0.00146)	0.00105 (0.00146)	0.0011 (0.00151)	0.00109 (0.00150)	0.00109 (0.00149)	0.00106 (0.00150)
Exponential	0.00125 (0.00196)	0.00136 (0.00227)	0.00288 (0.00475)	0.00306 (0.01447)	0.00176 (0.00295)	0.00254 (0.00363)

Table 5.4: Mean square error of $\hat{p}_{ij}$ over 500 replications and $k = 16$ (s.e. in parenthesis)

Distribution	T&M	B&T	CAU	UNI	EXP	NNM
Normal	0.00086	0.00086	0.00088	0.00092	0.00088	0.00090
	(0.00119)	(0.00119)	(0.00121)	(0.00129)	(0.00121)	(0.00128)
Gumbel	0.00092	0.00092	0.00096	0.00094	0.00096	0.00093
	(0.00140)	(0.00140)	(0.00145)	(0.00143)	(0.00147)	(0.00142)
Cauchy	0.00099	0.00099	0.00099	0.00101	0.00099	0.00100
	(0.00127)	(0.00127)	(0.00125)	(0.00128)	(0.00126)	(0.00128)
Uniform	0.00080	0.00080	0.00085	0.00088	0.00085	0.00081
	(0.00103)	(0.00103)	(0.00108)	(0.00116)	(0.00108)	(0.00105)
Exponential	0.00096	0.00101	0.00226	0.00229	0.00131	0.00216
	(0.00132)	(0.00150)	(0.00368)	(0.01396)	(0.00197)	(0.00338)

Table 5.5: Mean square error of $\hat{p}_{ij}$ over 500 replications and $k = 24$ (s.e. in parenthesis)

Distribution	T&M	B&T	CAU	UNI	EXP	NNM
Normal	0.00083	0.00083	0.00087	0.00091	0.00087	0.00086
	(0.00122)	(0.00122)	(0.00129)	(0.00134)	(0.00129)	(0.00125)
Gumbel	0.00101	0.00101	0.00103	0.00104	0.00104	0.00106
	(0.00132)	(0.00132)	(0.00136)	(0.00139)	(0.00137)	(0.00138)
Cauchy	0.00094	0.00094	0.00094	0.00095	0.00095	0.00094
	(0.00133)	(0.00133)	(0.00135)	(0.00136)	(0.00135)	(0.00133)
Uniform	0.00088	0.00088	0.00090	0.00093	0.00091	0.00092
	(0.00134)	(0.00134)	(0.00135)	(0.00139)	(0.00136)	(0.00138)
Exponential	0.00100	0.00101	0.00210	0.00292	0.00127	0.00148
	(0.00151)	(0.00150)	(0.00339)	(0.01201)	(0.00194)	(0.00257)

The nuclear norm minimization provides very similar results to linear models composition rules. These results are meaningful, especially considering that in these analyses matrices are simulated starting from the same assumptions as linear model. The composition rules associated to the Thurstone &

Mosteller and the Bradley & Terry linear models provide the same results, except for the assumed Exponential distribution. As k increases, we can not observe any regularity in the results. The composition rule associated to the Thurstone & Mosteller linear model seems to be the best method to estimate missing values in a generalized tournament matrix, in case of one missing observation and data generated under linear model assumptions. However it could be interesting to compare linear models' composition rules and nuclear norm minimization in case of two or more missing entries and on real data.

Chapter 6

Conclusions

The first aim of this thesis was to compare methods based on distributional parametric hypotheses with methods that do not require such hypotheses. For this reason, some simulation studies have been performed and the results analyzed. We are now able to state that the Cauchy model is the only ranking method characterized by an unreliable behavior. In fact, in the case of first raw dominant matrices it is characterized by very low proportions of always to-be-preferred alternative classified as first. For strictly ordered matrices, the Cauchy model is again the least correlated with the consistent ranking.

With regard to the other methods, they give similar outcomes. These results are confirmed also by considering the correlations between the examined ranking methods for random matrices. The method least correlated with the others is always the Cauchy linear model while all the remainings are highly correlated. However, we can observe slight differences among other models. The eigenvector method, for instance, is the best performer in identifying the dominant alternative, for all considered values of k and of to-be-preferred probability. The exponential linear model provides reliable estimates in case of strictly ordered matrices.

We notice that the eigenvector and the score-based ranking methods, which do not require any distributional assumption, give results very similar to

linear models, even when matrices are generated under the linear models assumptions.

We can finally affirm that the effectiveness of the examined methods does not depend on the assumed distribution.

Hence, apart from the advice to avoid the Cauchy model, the choice of the ranking method depends essentially on the researcher's objectives. If the goal is to get only a ranking of the alternatives, we advise to use the eigenvector method. On the other hand, linear models can be used to describe the relationship between the preference probabilities and some explainatory variables, such as personal characteristics of the respondents, useful for the estimation of individual data.

It was not possible to compare the different methods for scoring purposes, since scores obtainable with the score vector and the eigenvector methods are expressed on a different scale than linear models.

We proposed a method to estimate missing values within a generalized tournament matrix. This method is based on the minimization of the nuclear norm. It assumes that only a few dimensions contribute to individual preferences and searches for values that appraise and respect the valid elements of the matrix, minimizing the deviation from them.

Some simulations are performed to compare the nuclear norm minimization rule with the composition rules associated to linear models. The nuclear norm minimization provides results similar to composition rules. These results are meaningful, especially considering that in these analyses matrices are simulated starting from the same assumptions as linear model.

For further research, it would be interesting to compare composition rules and the nuclear norm minimization extending the comparison to two or more missing entries and larger matrices. In particular, it would be important to study whether the examined methods are affected by a larger proportion of missing entries within a generalized tournament matrix.

Bibliography

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics,* 25: 242–262.

Bock, R. D. & Jones, J. V. (1968). *The Measurement and Prediction of Judgment and Choice.* Holden-Day, San Francisco.

Bradley, R. A. & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika,* 39: 324–345.

Brauer, A. & Gentry, I. C. (1968). On the characteristic roots of tournament matrices. *Bull. Amer. Math. Soc,* 74: 1133–1135.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems,* 30: 107–117.

Brualdi, R. A. (2006). Combinatorial matrix classes. In *Encyclopedia of Mathematics and its Applications*, vol. 108. Cambridge University Press, Cambridge.

Candès, E. J. & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics,* 9: 717–772.

Carmone, F. J. et al. (1997). A Monte Carlo investigation of incomplete pairwise comparison matrices in AHP. *European Journal of Operational Research,* 102: 538–553.

Chan, V. (2011). Prediction accuracy of linear models for paired comparisons in sports. *Journal of Quantitative Analysis in Sports,* 7: 18–53.

Chistov, A. L. & Grigor'ev, D. Y. (1984). Complexity of quantifier elimination in the theory of algebraically closed fields. In *Mathematical Foundations of Computer Science 1984*, vol. 176: 17–31. Springer, Berlin Heidelberg.

Churchill, J. et al. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research,* 21: 360–375.

Cohen, S. & Orme, B. (2004). What's your preference? Asking survey respondents about their preferences creates new scaling decisions. *Journal of Marketing Research,* 16: 32–37.

Conrad, F. G. et al. (2005). Interactive feedback can improve the quality of responses in web surveys. In *Annual Meeting of the American Association for Public Opinion Research (AAPOR) 60th Annual Conference.*

Coombs, C. H. (1958). On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology,* 55: 1–7.

Coombs, C. H. (1976). *A theory of data.* Mathesis Press, Ann arbor, MI.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research,* 17: 407–442.

Cravo, G. (2009). Matrix completion problems. *Linear Algebra and Its Applications,* 430: 2511–2540.

David, H. A. (1987). Ranking from unbalanced paired-comparison data. *Biometrika,* 74: 432–436.

David, H. A. (1988). *The method of paired comparisons (2nd ed.).* Charles Griffin & Company, London.

De Caen, D. et al. (1992). Algebraic multiplicity of the eigenvalues of a tournament matrix. *Linear Algebra and its Applications,* 169: 179–193.

Devaraj, S. et al. (2002). Antecedents of b2c channel satisfaction and preference: validating e-commerce metrics. *Information Systems Research,* 13: 316–333.

Dittrich, R. et al. (2012). Missing observations in paired comparison data. *Statistical Modelling,* 12: 117–143.

Donoho, D. L. & Tanner, J. (2005). Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America,* 102: 9446–9451.

Edwards, A. L. (1983). *Techniques of Attitude Scale Construction*. Irvington Publishers, New York.

Fabbris, L. (2011). One-dimensional preference data imputation through transition rules. In *Classification and Multivariate Analysis for Complex Data Structures*. 245–252. Springer, Berlin Heidelberg.

Fabbris, L. (2013). Measurement scales for scoring or ranking sets of interrelated items. In *Survey Data Collection and Integration*. 21–43. Springer, Berlin Heidelberg.

Fabbris, L. & Fabris, G. (2003). Sistema di quesiti a torneo per rilevare l'importanza di fattori di customer satisfaction mediante un sistema CATI. In *Laid Out: scoprire i rischi con l'analisi di segmentazione*. 299–322. Cleup, Padova.

Fazel, M. et al. (2001). A rank minimization heuristic with application to minimum order system approximation. In *American Control Conference, 2001. Proceedings of the 2001*, vol. 6. 4734–4739.

Fechner, G. (1860). *Elemente der Psychophysik.* Breitkopf & Härtel, Leipzig.

Fechner, G. (1966). *Elements of psychophysics. Vol. I.* Holt, Rinehart and Winston, New York.

Fedrizzi, M. & Giove, S. (2007). Incomplete pairwise comparison and consistency optimization. *European Journal of Operational Research,* 183: 303–313.

Friedman, H. H. & Friedman, L. W. (1986). On the danger of using too few points in a rating scale: a test of validity. *Journal of Data Collection,* 26: 60–63.

Gleich, D. F. & Lim, L. H. (2011). Rank aggregation via nuclear norm minimization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* 60–68. ACM, San Diego.

Handley, J. C. (2001). Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment. In *PICS 2001: Image Processing, Image Quality, Image Capture Systems.* 108–112. Society for Imaging Science and Technology, Washington.

Harker, P. T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling,* 9: 837–848.

Horn, R. A. & Johnson, C. R. (1985). *Matrix analysis.* Cambridge university press, Cambridge.

Inglehart, R. & Abramson, P. (1993). Values and value change of five continents. In *Annual Meeting of the American Political Science Association.* 1–5. APSA, Washington.

Joe, H. (1990). Extended use of paired comparison models, with application to chess rankings. *Applied statistics,* 39: 85–93.

Kassirer, J. P. (1994). Incorporating patients' preferences into medical decisions. *The New England journal of medicine,* 330: 1895–1896.

Keener, J. P. (1993). The perron-frobenius theorem and the ranking of football teams. *SIAM review,* 35: 80–93.

Kendall, M. (1955). Further contributions to the theory of paired comparisons. *Biometrics,* 11: 43–62.

Kendall, M. (1970). *Rank Correlation Methods.* Charles Griffin and Company, London.

Kendall, M. G. & Babington-Smith, B. (1940). On the method of paired comparisons. *Biometrika,* 31: 324–345.

Kwiesielewicz, M. & Van Uden, E. (2003). Ranking decision variants by subjective paired comparisons in cases with incomplete data. In *Lecture Notes in Computer Science*, vol. 2669. 208–215. Springer-Verlag, Berlin/Heidelberg.

Latta, R. B. (1979). Composition rules for probabilities from paired comparisons. *The Annals of Statistics,* 7: 349–371.

Laurent, M. (2001). Matrix completion problems. *The Encyclopedia of Optimization,* 3: 221–229.

Lee, S. & Seol, H. (2001). A survey on the matrix completion problem. *Trends in Mathematics,* 4: 38–43.

Linial, N. et al. (1995). The geometry of graphs and some of its algorithmic applications. *Combinatorica,* 15: 215–245.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* John Wiley and Sons, New York.

Manski, C. F. & McFadden, D. (1981). *Structural analysis of discrete data with econometric applications*. MIT Press, Cambridge.

Matell, M. S. & Jacoby, J. (1971). Is there an optimal number of alternatives for likert scale items? study 1: Reliability and validity. *Educational and Psychological Measurements,* 31: 657–674.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics,* 3: 303–328.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business,* 53: 13–29.

Mises, P.-G. H., R V. (1929). Praktische verfahren der gleichungsauflösung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik,* 9: 58–77.

Moon, J. W. (1968). *Topics on Tournaments*. Holt, Rinehart and Winston, New York.

Moon, J. W. & Pullman, N. J. (1970). On generalized tournament matrices. *SIAM Review,* 12: 384–399.

Mosteller, F. (1951). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika,* 16: 207–218.

Noether, G. E. (1960). Remarks about a paired comparison model. *Psychometrika,* 25: 357–367.

Pinto da Costa, J. & Soares, C. (2005). A weighted rank measure of correlation. *Australian & New Zealand Journal of Statistics,* 47: 515–529.

Recht, B. et al. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review,* 52: 471–501.

Rennie, J. D. M. & Srebro, N. (2005). Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine Learning*. 713–719. ACM, New York.

Ryan, M. et al. (2001). Eliciting public preferences for healthcare: A systematic review of techniques. *Health Technology Assessment,* 5: 1–186.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology,* 15: 234–281.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences,* 1: 83–98.

Sato, Y. (2004). Comparison between multiple-choice and analytic hierarchy process: Measuring human perception. *International Transactions in Operational Research,* 11: 77–86.

Schafer, J. B. et al. (2001). E-commerce recommendation applications. *Data mining and knowledge discovery,* 5: 115–153.

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods,* 7: 147–177.

Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika,* 48: 303–312.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology,* 15: 72–101.

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review,* 34: 273–286.

Thurstone, L. L. (1927b). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology,* 21: 384–400.

Thurstone, L. L. (1927c). Psychophysical analysis. *The American Journal of Psychology,* 38: 368–389.

THURSTONE, L. L. (1931). Rank order as a psycho-physical method. *Journal of Experimental Psychology,* 14: 187–201.

TORGERSON, W. S. (1958). *Theory and Methods of Scaling.* John Wiley and Sons, New York.

TRAIN, K. E. (2003). *Discrete choice methods with simulation.* Cambridge University Press, Cambridge.

WEI, T. H. (1952). *The Algebraic Foundations of Ranking Theory.* Ph.D. thesis, University of Cambridge.

WEINBERGER, K. Q. & SAUL, L. K. (2004). Unsupervised learning of image manifolds by semidefinite programming. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer vision and pattern recognition.* 988–995. IEEE Computer Society.

ZERMELO, E. (1929). Vortrags-themata fur warschau. *Notes of Six Lectures, Manuscript:* 134–136.

Manuela Scioni

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 8274174
e-mail: scioni@stat.unipd.it

Current Position

Since January 2010; (expected completion: January 2013)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Multivariate preference models for scoring or ranking units
Supervisor: Prof. Luigi Fabbris
Co-supervisor: Prof. Boccuzzo Giovanna.

Research interests

- Stated preference analysis
- Ranking and scoring methods
- Survey research

Education

September 2006 – July 2009

Master (*laurea specialistica*) degree in "Statistics for Demography and Social Sciences".

University of Padova, Faculty of Statistics

Title of dissertation: "La costruzione condivisa di un indicatore di capitale umano dei laureati "

Supervisor: Prof. Luigi Fabbris

Final mark: 104 (out of 110)

September 2002 – July 2006

Bachelor degree (*laurea triennale*) in "Statistics, Mathematics, and Data Management".

University of Genova, Faculty of Mathematical Physical and Natural Science

Title of dissertation: "L'inferenza nell'analisi in componenti principali"

Supervisor: Prof. Maria Piera Rogantin

Final mark: 106 (out of 110)

Work experience

June 2007 – July 2007

Liguria Ricerche S.P.A., Via XX Settembre 42, 16121 Genova.

Research activity on ligurian economy. ASIA data elaboration (ISTAT).

Supervisor: Prof. Luca Beltrametti

October 2009 – December 2009

Università degli Studi di Padova - Dipartimento di Scienze Statistiche Via Cesare Battisti 241 35121 Padova.

Preferences analysis with statistical methods. Elaboration of data on undergraduate and graduate students.

Supervisor: Prof. Luigi Fabbris

Awards and Scholarship

Giugno 2012
Phd Schoolarship, University of Padua, Italy.

Computer skills

- Statistical packages: R, SAS, SPSS
- Mathematical packages: Maple, Matlab
- Other: LaTeX

Language skills

Italian: native; English: good.

Publications

Articles in books
Fabbris, Luigi and Boccuzzo, Giovanna and Martini, MariaCristiana and Scioni, Manuela, (2011). A Participative Process for the Definition of a Human Capital Indicator. *New Perspectives in Statistical Modeling and Data Analysis*, 39–47. Springer Berlin Heidelberg.

References

Prof. Luigi Fabbris
University of Padova
Department of Statistics
Via Cesare Battisti 241-243
049 8274135
fabbris@stat.unipd.it

Dott.ssa Giovanna Boccuzzo
University of Padova
Department of Statistics
Via Cesare Battisti 241-243
049 8274194
boccuzzo@stat.unipd.it