CONFIDENCE REGIONS FOR MULTIDIMENSIONAL SCALING ANALYSIS

J. O. RAMSAY
MCGILL UNIVERSITY

Techniques are developed for surrounding each of the points in a multidimensional scaling solution with a region which will contain the population point with some level of confidence. Bayesian credibility regions are also discussed. A general theorem is proven which describes the asymptotic distribution of maximum likelihood estimates subject to identifiability constraints. This theorem is applied to a number of models to display asymptotic variance-covariance matrices for coordinate estimates under different rotational constraints. A technique is described for displaying Bayesian conditional credibility regions for any sample size.

Keywords: conditional credibility regions, Bayesian analysis, asymptotic distribution, maximum likelihood estimation, variance-covariance of configuration.

1. Introduction

The primary goal in multidimensional scaling is to represent each of $n$ objects or stimuli as a point in a space (usually Euclidean) of dimensionality $k$. This point estimation problem leads naturally to the companion problem of estimating a region in this space surrounding a given point such that one can assert with some level of probability that the “true” or population point lies within this region. The Bayesian statistician might rephrase this problem as one of defining a region about the point estimate for which the volume under the marginal posterior density for this point is a certain value.

Figure 1 indicates graphically what is desired. The ellipse surrounding each point describes a confidence region with probability content .95. Thus the classical statistician would say that the probability that some particular ellipse contains the population point is .95 while the Bayesian would say that the volume under the marginal posterior density function over this ellipse is .95.

A number of arguments point to the importance of confidence region estimation as a secondary goal in multidimensional scaling. Confidence regions in a multidimensional scaling solution can indicate graphically the precision with which a point is located. Moreover, they can also indicate the directions in which the location of the point is relatively tightly defined. A knowledge of the way in which the volume of a confidence region interacts with sample size can assist the experimenter in determining the number of replications that he will employ. The same might be said of the number of stimuli. These regions can also indicate whether two points are significantly separated or the extent to which some other relationship between subsets of points is satisfied.

The reader must be warned at the outset that the adjective “confidence” used throughout this paper is rather broadly defined. In fact, a Bayesian perspective, which subsumes classical inference as a special case, would require the term “credibility region” as used in Novick and Jackson [1974] or “highest posterior density (HPD) region” as used in Box and Tiao [1973]. Indeed, the computation of credibility regions is only one...
technique for examining posterior distributions and is no substitute for a complete plot of
a distribution where feasible. Some of the techniques described in this paper lend them-
selves to the production of such plots.

Finally, it is worth pointing out that there are various ways in which a confidence or
credibility region can be defined in a multiparameter situation. A particular region is
typically defined by a closed curve determined by a function \( g(\theta) \) of the vector of
parameters \( \theta \) and the equation

\[
g(\theta) = 0.
\]

If \( R \) is the region in parameter space \( \Omega \) included within this closed curve, one requires at
least that

\[
\text{Prob}\{\theta \in R\} \geq 1 - \alpha
\]

for some smallish constant \( \alpha \).

Some secondary constraint on \( R \) is as a rule also required. One possibility is the
volume of this region \( R \) be the smallest possible among the whole class of regions defined
by condition (1.2). Another possible secondary constraint is that the posterior density
function will be constant on curve (1.1). This constraint implies that one is seeking a
contour or level curve for the posterior density function (or likelihood function) which is a
base for a certain substantial percentage of the volume under this function. This constraint
is easier to implement in practice, although as a rule both constraints are asymptotically
equivalent since, for large samples, posterior densities (and likelihoods) typically become
progressively more multinormal in shape [Walker, 1969].

In the following section four multidimensional scaling models are defined. The
estimation of the parameters for these models by maximum likelihood has been described
by Ramsay [1977, Note 1]. In Section 3, a result on the asymptotic distribution of
constrained maximum likelihood estimates is discussed. Its applications to multi-
dimensional scaling are taken up in Section 4. The next section deals with a different type
of confidence region—the conditional confidence region, and discusses some numerical
procedures for the computation of such regions.

2. The Models to be Considered

Let \( d_{ijr} \) be the \( r \)th replication of an observation of the distance between points \( i \) and \( j \)
(\( i, j = 1, \ldots, n; r = 1, \ldots, N \)). Let \( d_{ijr}^* \) be the corresponding errorless distance. The
distribution of \( d_{ijr} \) conditional on \( d_{ijr}^* \) and a dispersion parameter \( \sigma^2 \) is defined by a
probability density function \( f(d_{ijr}|d_{ijr}^*, \sigma^2) \). It will be assumed that all observations are
independently distributed, so that the log likelihood function is given by

\[
\log L = \sum_i \sum_j \sum_r \log f(d_{ijr}|d_{ijr}^*, \sigma^2).
\]

The summations should be understood to be over the observations actually made, without
any presupposition that all possible pairs of \( i \) and \( j \) have been observed.

In this paper three definitions of \( d_{ijr}^* \) will be considered and referred to by the abbreviations \( M1, M2, \) and \( M3 \). These are:

\[
M1: d_{ijr}^* = \left[ \frac{\sum_m (x_{im} - x_{jm})^2}{m} \right]^{1/2},
\]

\[
M2: d_{ijr}^* = \frac{1}{v_r} \left[ \frac{\sum_m (x_{im} - x_{jm})^2}{m} \right]^{p_r/2}, \text{ and}
\]

\[
M3: d_{ijr}^* = \frac{1}{v_r} \left[ \frac{\sum m w_{rm}(x_{im} - x_{jm})^2}{m} \right]^{p_r/2}.
\]