Three Faces of Factor Analysis

David J Bartholomew

London School of Economics and Political Science

Abstract

Spearman's[1] basic idea was that the mutual correlation of variables (such as test scores) might be explained by their common dependence on a latent variable, which he called a factor. Such factors were to be uncovered, therefore, by appropriate numerical analysis of the correlation matrix. Spearman was able to show that the presence of an underlying factor would be revealed by a particular pattern in the correlation matrix. Although others extended this idea to several factors, the analysis of the correlation matrix on these lines dominated factor analysis for the next half century.

The second face is most easily seen in Lawley and Maxwell's *Factor Analysis as a Statistical Method* which first appeared in 1963 but it would be difficult to identify its precise origin. It was part of a general shift in statistics to the use of probability models which became obvious around the 1950s. The starting point was now a linear model in which the observed variables were expressed as linear functions of the factors and random 'errors'. This enabled the theory of statistical inference to be brought to bear on estimation and hypothesis testing using the likelihood function. A subtle change of focus from the correlation matrix to the covariance matrix accompanied this move and paved the way for covariance structure analysis.

The third face went unrecognized for nearly half a century but it was already present in Lazarsfeld's work on latent structure analysis. Spearman's original insight here re-surfaced in the observation that association between categorical variables might be explained by latent

[1]Spearman's contribution to the origins of factor analysis was described in Bartholomew (1995). That paper has much in common with this but here the focus is not primarily on Spearman

variations in continuous or categorical 'factors'. In fact, the difference between factor and latent structure analysis is more apparent than real. The two can be brought together in a factor analysis framework by supposing that the normal distribution of classical factor analysis is replaced by a member of the one-parameter exponential family in which the canonical parameter, rather than the mean, is linear in the factors and errors.

Factor analysis, in its maturity, thus emerges as one method of studying the inter-dependence structure of random variables. It differs from recent work on graphical models, multivariate dependencies etc., principally by introducing latent variables. Spearman was handicapped by the limits of the statistical and computational technology a century ago but, even more, by his narrow focus on measuring intelligence. Such is the fate of truly original pioneers. Genuinely new ideas are rare in any field and Spearman should be honoured for one which virtually created psychometrics and, eventually, had much wider effects.

1 Origins

1.1 Spearman's idea

The origin of factor analysis is to be found in Spearman [9] but the details are tantalizingly brief. The paper is a very long one—91 pages—but only a small part, notably the table on page 276, is concerned with something recognizable today as factor analysis and then almost as a side issue. The focus is more on the question of whether the common factor underlying a variety of branches of intellectual activity is the same as that which underlies tests of sensory discrimination. Thus, on page 272, Spearman reaches the conclusion:'On the whole then, we reach the profoundly important conclusion that *there really exists a something that we may provisionally term "General Sensory Discrimination" and similarly a "General Intelligence", and further that the functional correspondence between these two is not appreciably less than absolute.'*

Of more immediate relevance to factor analysis, he states what he calls "our general theorem" which is *Whatever branches of intellectual activity are at all dissimilar, then their correlations with one another appear wholly due to their being all variants wholly saturated with some common fundamental*

Function (or group of Functions). He distinguishes this central Function from "the specific function (which) seems in every instance new and wholly different from that in all the others". These ideas are translated into numbers in the table on page 271 with only the sketchiest justification for the calculations involved. Nevertheless it is a table of factor loadings (correlations of the test scores with the factor, general intelligence) and communalities. It is, then, the first example of a factor analysis.

The simplest way to justify this analysis, (according to the remark at the bottom of page (ii) of the Appendix in Spearman 1927) is to appeal to the theory of partial correlation. This had been introduced by Yule [13] in some detail. The essence of what Spearman needed is contained in the formula for the partial correlation between two variables, i and j say, given a third variable which following Spearman we call G. Thus

$$r_{ij.G} = \frac{r_{ij} - r_{iG}r_{jG}}{\sqrt{(1 - r_{iG}^2)(1 - r_{jG}^2)}}. \tag{1}$$

If the correlation between i and j is wholly explained by their common dependence on G then $r_{ij.G}$ must be zero. This implies that

$$r_{ij} = r_{iG}\, r_{jG} \qquad (i,\, j = 1,\, 2,\, \ldots,\, p). \tag{2}$$

If the correlation matrix $\mathbf{R} = \{r_{ij}\}$ can be represented in this way then we have evidence for an underlying common factor. Spearman found this to be so and gave estimates of $\{r_{iG}\}$ in his table on page 276.

In the beginning, factor analysis consisted in seeing whether the correlation matrix had the required structure. In particular, whether the 'tetrad differences' given by $r_{ij}r_{hk} - r_{ik}r_{jh}$ were all zero. These are equivalent to (2) and also to

$$\frac{r_{ij}}{r_{ih}} = \frac{r_{iG}r_{jG}}{r_{iG}r_{hG}} = \frac{r_{jG}}{r_{hG}} \qquad (i \neq j \neq h). \tag{3}$$

This ratio does not depend on i and hence is the same for all rows of the table (excluding the diagonal elements). A similar argument applies to the columns because the matrix is symmetrical.

At this basic level, we see that the correlation matrix is the key to unearthing a common factor. Later it was shown that if there were two or more (independent) underlying factors, the correlation matrix would have a structure of the form

$$r_{ij} = \sum_{h=1}^{q} \lambda_{ih}\, \lambda_{jh} \qquad (i \neq j). \tag{4}$$

The first face of factor analysis thus starts from the correlation matrix. What can be learnt about the factor structure is, therefore, contained in that matrix. Early factor analysis was dominated by the study of correlation matrices and mathematicians with skills in this field were enlisted to help (see, for example, references to Ledermann on p. 386 of Thomson 1950).

There is more to this than meets the eye because a correlation coefficient is a measure of *linear* correlation. The decision to use product moment correlation, therefore, implies an assumption that item test scores are linearly related to any underlying factors. Although implicit from the beginning it only became the central idea in the second phase of factor analysis' history—its second face. Before moving on to that we digress to notice that the structure (3) does not imply that the observed correlations were generated from a common dependence on a single factor. There is at least one other explanation, associated with the name of Godfrey Thomson.

1.2 Thomson's Alternative Model

It is common in statistics to find that more than one model makes exactly the same observational predictions. We may describe this as a lack of 'model identification'. Spearman's one-factor model, which provided a good fit to many data sets, supposed that individuals varied along a scale of what we will continue to call G. This led, naturally, to the supposition that this underlying factor was 'real'. Thomson[12] pointed out that there was another model, capable of describing the data equally well, which did not involve such a common factor. At most, only one of these models could describe physical reality and hence there was no certain empirical evidence for the reality of Spearman's factor.

This matter was much debated in the psychological literature in the 1920s and, for good reasons, Spearman's model came out on top. Thomson's model is largely forgotten though it is worth noting that Mackintosh (1998) has recently pointed out that it corresponds, in part at least, with contemporary ideas on brain function. It is, therefore, worth taking another look at Thomson's 'sampling model'. The original debate was somewhat clouded by the lack of a clear notion of a random variable. Thomson, himself, used simple examples based on dice and such like to get the idea across but this seems to have engendered misunderstanding and comprehension in equal measure! The nearest and clearest exposition seems to be due to Dodd [3]with whom Thomson [12] (p.43) said he agreed on a great deal if not on everything.

The following account, in modern dress, is very similar to Dodd's treatment though it is not the most general form possible. However, it is sufficient to make the point that it is, empirically, indistinguishable from Spearman's one-factor model.

Suppose the brain contains N 'bonds' which may be called into play when test items are attempted. (N is thought of as 'large' but this is not necessary for the argument.) Some items will be more demanding than others and so more bonds will be used in their solution. The correlations between two test scores are supposed to result from the use of some bonds being common to both items. The term 'sampling theory model' arises from the fact that the bonds used by any item are supposed to be selected at random from the N available.

Assume that item i requires Np_i (assumed to be an integer) bonds. The contribution which bond i makes to the score on that item, x_i, is a random variable e_i. The score may thus be expressed as

$$x_i = a_{i1}e_i + a_{i2}e_2 + \ \ldots\ + a_{iN}e_N \qquad\qquad (i = 1,\, 2,\, \ldots\, n) \qquad (5)$$

or

$$\mathbf{x} = \mathbf{Ae}$$

where n is the number of variables and the coefficients $\{a_{ij}\}$ are indicator variables taking the value 1 if the bond is selected and 0 otherwise. The as are, therefore, also random variables with joint distribution determined by the method of sampling, It is convenient to suppose that the es are mutually independent and have zero means but we allow their variances to differ with $\mathrm{var}(e_i) = \sigma_j^2$ $(j = 1,\, 2,\, \ldots,\, N)$. We now have to find the structure of the covariance (or correlation) matrix. Given these assumptions

$$\begin{aligned} E(x_i) &= 0 \qquad\qquad (i = 1,\, 2,\, \ldots,\, n) \\ E(x_i x_j) &= E\, E(x_i x_j | a_i,\, a_j) \\ &= E \sum_{h=1}^{N} \sum_{k=1}^{N} a_{ih} a_{jk} E(e_h\, e_k) \\ &= E \sum_{h=1}^{N} a_{ih} a_{jh} \sigma_h^2 \qquad (i,\, j = 1,\, 2,\, \ldots,\, N \qquad i \neq j). \end{aligned}$$

If we assume that successive samplings are independent and that all bonds are equally likely to be selected then

$$E(a_{ih} a_{jh}) = Np_i\, Np_j.$$

Hence,

$$E(x_i\, x_j) = N^2 p_i\, p_j \sum_{h=1}^{N} \sigma_h^2 = \mathrm{cov}(x_i,\, x_j) \qquad\qquad (i,\, j = 1, 2,\, \ldots,\, n).$$

When $i = j$,

$$\mathrm{var}(x_i) = \sum_{h=1}^{N} E a_{ih}^2 \sigma_h^2 = \sum_{h=1}^{N} E a_{ih} \sigma_h^2 = N p_i \sum_{h=1}^{N} \sigma_h^2,\ (i = 1,\, 2,\, \ldots,\, n).$$

Hence

$$\mathrm{corr}(x_i,\, x_j) = \sqrt{p_i\, p_j} \qquad\qquad (i,\, j = 1,\, 2,\, \ldots,\, n;\, i \neq j).$$

This has exactly the same form as the correlation matrix of Spearman's one-factor model. Hence the two are not empirically distinguishable.

The parameters $\{\sqrt{p_i}\}$ in Thomson's model correspond to the factor loadings $\{r_{iG}\}$ in Spearman's model. Estimates of the factor loadings can, therefore, be translated into estimates of the proportions of bonds which are selected in attempting an item, by the fact that $p_i = r_{iG}^2$ $(i = 1,\, 2,\, \ldots,\, n)$. Typical values of r_{iG} are in the range (0.5-1) so this would imply quite a high proportion of bonds being used.

Spearman criticized the sampling theory model on the grounds that it allowed no individual differences. Thomson denied this but he and Spearman may have been at cross purposes. Thomson pointed out that the sampling of bonds would be a separate operation for each individual and thus that the selection would not usually be the same for any two individuals. However, to *estimate* the correlations it would be necessary to suppose that they had the same expectation for all individuals. One cannot estimate a correlation from a single pair of test scores from one individual.Only by assuming that the correlations are the same for every individual does one have the replication necessary to make an estimate. One has to assume, therefore, that the method of sampling and the parameters $\{p_i\}$ (and N) are the same for all individuals.

There is an inherent implausibility about assuming homogeneity in any human population even if one does not wish to attribute any heterogeneity to differences in innate ability. Whether or not the assumption of a fixed N—the number of bonds—is plausible or whether it is sensible to suppose that the *number* of bonds called into play by a given item is a fixed number is also

questionable. Perhaps the fact that Spearman's model could be extended to several factors and, in that form, successfully fitted to a very wide range of data gave it the advantage.

Thomson, himself, conceded that Spearman went a long way to meet his objections but his alternative model is worth recalling as a reminder that finding a good fitting model is not the same as finding the real mechanism underlying the data.

2 Factor Analysis: Linear Models

The second face of factor analysis starts, not with a correlation matrix, but with a *model*. The use of models in statistics, on a regular basis, seems to date from the 1950s. A statistical model is a statement about the distribution of a set of random variables. A simple linear regression model, for example, says that the dependent variable y is normally distributed with mean $a + bx$ and variance σ^2, where a and b are unknown constants and x is an observable variable. In the case of factor analysis the move to a model-based approach was gradual. The rudimentary idea was contained in the idea that an observed test score was composed of a common part and a specific part. This is made explicit, for example, in Spearman and Jones [11] (p.37) but was implicit, as we have already noted, in the use of product moment correlations. Hotelling's [4] introduction of principal components analysis, which was concerned with expressing a set of variables $(x_1, x_2, \ldots, x_p)$ as linear function of p orthogonal variables $(y_1, y_2, \ldots, y_p)$ doubtless encouraged factor analysts to think of factor analysis in similar terms. However, it was in Lawley and Maxwell [5] that a linear model was made the starting point for developing the theory in a systematic way. Actually, this formulation was incomplete but, in its essentials, it still holds sway today. In modern notation Lawley and Maxwell supposed that

$$x_i = \lambda_{i1}y_1 + \lambda_{i2}y_2 + \ldots + \lambda_{iq}y_q + e_i \qquad (i = 1, 2, \ldots, p) \qquad (6)$$

In this equation the λs are constants and the xs, ys and es are *random variables*. Thus if one imagines that the y-values for item i are drawn at random from some distribution and the es are drawn similarly, then the model postulates that, if they are combined according to (6), the resulting random variable will be x_i. It is usually assumed that the ys are independent with normal distribution and (without loss of generality) unit variances; e_i is

assumed to be normal with variance ψ_i. Lawley and Maxwell [5] state, like many since, that all the random variables in (6) may be assumed to have zero means without any loss of generality. This is not so. It is tantamount to assuming that the mean of x_i is known. In practice this is seldom the case. It is more accurate to insert an unknown mean μ_i on the right hand side of (6). We then have the *standard linear normal factor model* in use today. The omission of μ_i does not have serious consequences but its inclusion makes for clarity.

Equations in random variables, like (6), need to be thought about very carefully. Much of the confusion which has surrounded the topic of *factor scores* stems from failing to distinguish between random variables and mathematical variables (see later).

An alternative way of writing the model, which is less prone to misunderstanding is to write it in terms of probability distributions as in Bartholomew and Knott [1]. thus we suppose

$$x_i|y_1y_2\ \ldots,\ y_q \ \frown\ N\left(\mu_i + \sum_{j=1}^{q} \lambda_{ij}y_i,\ \psi_i\right) \qquad (i = 1,\ 2,\ \ldots,\ p) \quad (7)$$

$$y_j \ \frown\ N(0,\ 1) \qquad (j = 1,\ 2,\ \ldots,\ q) \quad (8)$$

where y's are mutually independent. Equation(7) is a regression model for x_i in terms of regressor variables $y_1,\ y_2,\ \ldots,\ y_q$. If the ys were known, factor analysis could be handled with the framework of regression theory. As it is, the ys are random variables and so their joint distribution must be specified, as in equation(8).

Once we have a probability model, as specified in (7), the whole field of inferential statistics is available to us. For example, the parameters $\{\mu_i\}$, $\{\lambda_{ij}\}$ and $\{\psi_i\}$ can be estimated by maximum likelihood. Hypotheses can be tested about the values of those parameters and so on. What can be known about any y, once the xs have been observed, is contained in the posterior distribution of y given $x_1,\ x_2,\ \ldots,\ x_p$.

It is worth pausing to ask the justification for regarding this as another way of looking at Spearman's factor analysis. The answer, of course, is that the correlation structure to which it leads is the same. If, for example, one writes down the likelihood function it turns out that the sample means, $\bar{x_1},\ \bar{x_2},\ \ldots,\ \bar{x_p}$ and the sample covariance matrix $\hat{\mathbf{\Sigma}}$ are sufficient for the parameters. In particular, the covariance matrix $\mathbf{\Sigma}$ has the form

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi} \quad (9)$$

which is precisely what the correlation approach (4) leads to. If there is one factor, for example, the theoretical covariance between x_i and x_j is

$$\text{cov}(x_i\, x_j) = \lambda_i\, \lambda_j \qquad\qquad (i = j)$$

and it can easily be shown that λ_i is the correlation between y and x_i. The correlation between x_i and x_j is of the same form.

As in standard linear regression, much can be achieved without the assumptions of normality. The covariance structure has the same form if the distributional assumptions are dropped.

Once factor analysis is viewed in this way it is easier to pose the *factor scores* problem. This is concerned with predictions, or estimating the unobserved values of $\mathbf{y}$ to be associated with and $\mathbf{x}$. The problem had been recognized from the beginning, of course. Spearman had proposed an *ad hoc* method (Spearman [**?**]; Spearman and Jones [11] 1950). This involved obtaining p estimates of y by putting $e_i = 0$ in (6) and then taking an average. Thomson [12] found the best linear regression of y on $\mathbf{x}$, noting that this did not require knowledge of the individual values of the ys. For obvious reasons, these became known as regression scores. Other types of scores were proposed using various criteria but the subject became confused by attempts to interpret the linear equations of (6 or 7), by analogy with principal components analysis, as though they were equations in mathematical variables. The argument would usually go as follows.

If $q = p$ and if $e_i = 0$ for all i the equations of (6) are formally the same as in principal components analysis. The xs and ys are not then random variables but represent real numbers. The equations may then be invoked to give $\mathbf{y}$ in terms of $\mathbf{x}$. If $q < p$, the argument goes, there are more unknowns than equations: p es and q ys gives $p + q$ unknowns and only p equations. The factors are then said to be indeterminate. Thomson's regression method can the be regarded as one possible way of getting a 'best fit' solution to the equations. This operation changes the random variables into mathematical variables and thus changes the question being asked.

It is clear that, once the model is formulated in terms of probability distributions, as in (7), that the question which the model is capable of answering is :what is the distribution of $\mathbf{y}$ *given* $\mathbf{x}$? The answer follows inexorably from the laws of probability and is given by the posterior distribution of $\mathbf{y}$ given $\mathbf{x}$. Point predictions of $\mathbf{y}$ can then be found as measures of location of that distribution. Posterior measures of spread then give an indication of the imprecision of those predictions.

The model-based approach thus enables us to provide rigorous methods for answering the traditional questions addressed by factor analysis.

3 The Third Face of Factor Analysis

It is a curious fact of statistical history that there has been a strong focus on methods for continuous data. Regression and correlation analysis and then the analysis of variance have, for the most part, pre-supposed that the variables involved were continuous. Other multivariate methods, introduced along the way, such as discriminant analysis, principal components analysis and canonical correlation fall into the same mould. It is interesting to speculate how far this can be attributed to the fact that the data on crop yields which confronted Sir Ronald Fisher at Rothamsted were continuous. It is unsurprising that factor analysis should have started from the same supposition and concentrated on correlation.

Of course, these methods have been widely used on data which were not continuous. Coarsely grouped variables, ordered categorical variables—even binary variables—have been grist to the analysts' mill. Indeed, much ingenuity has been exercised to treat categorical data as if it were continuous by introducing, for example, pseudo-correlation coefficients of one sort or another.

In practice, and especially in the social sciences, much of the data we encounter is not continuous but categorical. Sometimes the categories are ordered and sometimes not. Often they are binary, being derived from true/false or yes/no questions in sample surveys. In fact, sample surveys are a common source of data for which continuous methods are not appropriate. Matters are often made more difficult by the fact that a survey is likely to lead to a mixture of types of variable thus calling for hybrid methods capable of coping with all sorts of variable.

In turning to the third face of factor analysis we are looking below the surface to identify the essential questions which factor analysis is intended to answer.

In factor analysis we are asking whether the dependencies among a set of variables can be explained by their common dependence on one, or more, unobserved latent variables (or factors). There is nothing in this statement which refers to the level of measurement of the variables involved. If, therefore, we formulate the problem in sufficiently general terms we should have a

general enough framework to include variables of all sorts. The essential elements of the problem are the inter-dependence of a set of observable variables and the notion of conditional independence.

Suppose we have p observable random variables $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$ with joint probability distribution $f(\mathbf{x})$. This may be a joint probability density if the xs are all continuous, a joint probability function if they are all discrete; otherwise it is a mixed function.

The question is: Do there exist factors $y_1, y_2, \ldots, y_q\,(= \mathbf{y}')$ such that the xs are conditionally independent? That is, can we find a q and variables $y_1, y_2, \ldots, y_q$ such that

$$f(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^{p} f(x_i|\mathbf{y}). \tag{10}$$

In the case of the normal linear model this question is answered by finding a q such that an adequate fit is obtained with the linear model. With other kinds of variable (non-normal as well as as non-continuous) different methods will be required but, conceptually, the problem is the same.

This insight seems to have escaped factor analysts until quite recently. In fact, this more general way of looking at the problem had an independent origin in what seemed to be a quite different problem—in sociology, not psychology. This disciplinary divide probably accentuated the gulf and prolonged the separate existence of a distinct body of theory.

Lazarsfeld was the pioneer of this distinct kind of 'factor analysis' and he called it *latent structure analysis*. See, for example, [6]. Essentially, he allowed one or both of the sets of variables $\mathbf{x}$ and $\mathbf{y}$ to be categorical. Factor analysis was thus excluded because both $\mathbf{x}$ and $\mathbf{y}$ are then continuous. In retrospect, this seems to be a curious development but, although Lazarsfeld recognized that there were similarities with factor analysis he thought that the differences were more significant. The differences were, in fact, in matters of computation and in the appearance of the formulae. These things are not fundamental. The similarities were in the nature of the questions asked. This family likeness becomes more apparent when we adopt a sufficiently general notation as in (10) above.

The closeness of latent structure analysis and factor analysis is even more obvious when we discover that all the variants in common use can be subsumed under a linear model called, in Bartholomew and Knott (1999), *The General Linear Latent Variable Model* (GLLVM). This may be seen as a

generalization of the normal linear factor model in almost exactly the same way as the generalized linear model generalizes the general linear model of statistics. In the normal linear factor model it is assumed that each x_i has a linear regression on the factors with normal residuals. All other standard latent structure models emerge as special cases if we suppose that each x_i comes from a one-parameter exponential family distribution. In this case, it is the canonical parameter which is a linear function of the factors rather than the mean. In the standard linear model there is nothing which requires the regression variables (the factors in a factor model) to be continuous. Categorical variables can be represented by indicator variables, or vectors. Viewed in this way, factor analysis is simply the attempt to explain the interdependence of a set of variables in terms of their dependence on a small set of underlying variables which may be categorical or continuous.

There is one important class of problem which does not fit nearly into this framework.This occurs when some of the variables, the manifest variables in particular, are *ordered* categorical. These can be accommodated in one of two ways. One is to put order constraints on the parameters of the model in a way calculated to reflect the fact that the 'higher' effect on the response probability increases monotonically as we move through the categories. The other way is to regard the ordered categories as a grouped form of a continuous variables. The latter is often the most realistic but the choice, ideally, should be guided by what gives rise to the ordering in the first place. For a fuller discussion of this matter see Bartholomew, Steele, Moustaki and Galbraith [2] especially Chapter 8.

The unifying effect of adopting this way of looking at factor analysis has some interesting consequences for dealing with the problem of factor scores. (A first look at this problem from the point of view of the linear model was given in Section 2.) The problem, we recall, was to locate an individual in the factor space on the basis of the observed value of their $\mathbf{x}$. Within the general framework, the way we do this is obvious. If $\mathbf{x}$ and $\mathbf{y}$ are random variables then, when $\mathbf{x}$ is known, all of the information about $\mathbf{y}$ is conveyed by its posterior distribution $f(\mathbf{y}|\mathbf{x})$. This simple fact shows that there is no single value of $\mathbf{y}$ to be associated with any $\mathbf{x}$. There is a probability distribution over the $\mathbf{y}$-space and any 'score' must, therefore, be a summary measure of the distribution. Measures of location are the natural measures to use. A posterior measure of dispersion is then appropriate to show how imprecise the score is—how reliable, in other language.

Is is curious how, in the latent structure tradition, this is exactly the

approach which has been used. To take the simplest case, suppose that we have fitted a model in which a single y is supposed to be binary—meaning that there are just two latent classes. The posterior distribution of y is thus a two-point distribution. If, without loss of generality, we take the two values of y to be 0 and 1, the expectation, for example, is $E(y|\mathbf{x}) = \Pr\{y = 1 \,|\, \mathbf{x}\}$. *A posteriori*, therefore, we calculate the probability that the individual falls into category 1.

In the factor model, by contrast, this route has not been followed, though Thomson's 'regression' estimate is a distribution-free implementation of the same idea.

4 A Broader Perspective

The unified approach seen in the third face of factor analysis does more than simplify our understanding and make for economy of thought It also gives a deeper insight into the nature of certain familiar features of individual techniques.

We have seen that the existence of q continuous factors $y_1, y_2, \ldots, y_q$ means that the joint distribution can be expressed in the form

$$f(\mathbf{x}) = \int_y \prod_{i=1}^{p} f(x_i|\mathbf{y})\, f(\mathbf{y})\, dy. \tag{11}$$

The xs may be continuous or categorical. It is immediately clear that this has the form of a mixture with $f(\mathbf{y})$ as the mixing distribution. Mixtures occur in many branches of statistics and a great deal is known about them and their properties.

A second important feature is that any transformation $\mathbf{y} \rightarrow \mathbf{z}$ in (11) leaves $f(\mathbf{x})$ unchanged because if is merely a change of variable in the integral. There are thus infinitely many pairs $\{f(\mathbf{y}),\ f(x_i|\mathbf{y})\}$ leading to the same $f(\mathbf{x})$. Since the only distribution we can directly learn about is $f(\mathbf{x})$ there is no empirical way of distinguishing among this infinite set of possible models. In practice, of course, we narrow the set down by fixing $f(\mathbf{y})$ or requiring $f(x_i|\mathbf{y})$ to belong to some convenient family but these choices are, essentially , arbitrary. There is thus an inevitable indeterminacy in all factor models.

A special case of this indeterminacy is very familiar in the linear factor model where it lies behind the concept of rotation. In that case, the transformation to $\mathbf{z} = \mathbf{My}$ where $\mathbf{M}$ is an orthogonal matrix leads to the same

covariance matrix and hence, (under the usual normal assumptions) to the same joint distribution. The general formulation shows this to be a special case of a more fundamental indeterminacy.

Rotation by a linear transformation is not peculiar to the linear factor model. All members of the class of GLLVM's with continuous y's have the canonical parameter as a linear combination of the factors.Thus if

$$\theta_i = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \ \ldots \ + \alpha_{iq}y_q \tag{12}$$

or

$$\theta = \mathbf{A}\,\mathbf{y}, \text{ say,}$$

then

$$\theta = \mathbf{A}\mathbf{M}^{-\mathbf{1}}\mathbf{z} \tag{13}$$

where $\mathbf{z} = \mathbf{M}\mathbf{y}$. The zs have the same independent standard normal distributions as the ys but their coefficients are transformed from $\mathbf{A}$ to $\mathbf{A}\mathbf{M}^{-\mathbf{1}}$. The two versions of the model are thus indistinguishable because the distribution $f(\mathbf{x})$ is unaffected. This is the usual account of rotation but it is now revealed as characteristic of a much wider class of models.

The indistinguishability of models extends to what we might term 'near-indistinguishability' which is just as important practically. The best known example, perhaps, has been known for some time but has been investigated most thoroughly by Molenaar and von Eye [8]. Thus it is known that the covariance structure of the linear factor model is the same as that of a latent profile model with one more latent class than there are factors. On the basis of the covariance matrix alone one cannot distinguish between the two models. Only by looking at other aspects of the joint distribution would it, in principle, be possible to discriminate. The full practical implication of this result for the vast number of factor analyses that have been carried out seems to have been scarcely noticed.

A further example is provided by the latent class model with two classes and the latent trait model. Empirically it is very difficult to distinguish these, yet they say radically different things about the prior distribution of the latent variable. In the former case it is a two-point distribution and, in the latter, it is usually taken as a standard normal distribution. Further discussion and an example will be found in Bartholomew *et al.* (2002, section 9.4)

Another way of characterising results of this kind is to say, as the title of Molenaar and von Eye's paper implies, that latent variables are very poorly

determined. This result has far-reaching implications for all work on the relationships among latent variables.

A final result, with practical implications, takes us back to the factor score question. For the family of models of the GLLVM class, the joint distribution of $\mathbf{x}$ can be expressed in the form

$$f(\mathbf{x}) = \int_{\mathbf{y}} f(\mathbf{X}|\mathbf{y})\, f(\mathbf{y})\, d\mathbf{y} \tag{14}$$

where $\mathbf{X}$ is a q-vector with elements of the form

$$X_j = \sum_{i=1}^{p} a_i x_i \qquad\qquad (j = 1,\, 2,\, \ldots,\, q). \tag{15}$$

This shows that f($\mathbf{x}$) depends on $\mathbf{x}$ only through q linear combinations of the xs. Any factor score, for any member of this family, should therefore be a function of these 'sufficient' statistics (as they are called in Bartholomew and Knott 1999).

Many purely empirical attempts at scaling have proposed to use linear combinations of the observed scores—whether in educational testing or other fields. It is interesting to observe that the third face of factor analysis provides theoretical under- pinning for the use of linear combinations.

In a sense, therefore, we have come full circle to a point where we see that Spearman's original attempt to find a method of constructing a measure of general intelligence eventually leads to the same kind of measure as his more empirically minded successors proposed on intuitive grounds.

References

[1] Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis*, 2nd edition, London: Arnold.

[2] Bartholomew, D. J. , Steele, F., Moustaki, I. and Galbraith, J. I. (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists*, Boca Raton, Florida: Chapman and Hall/CRC.

[3] Dodd, S. C. (1929).The sampling theory of intelligence,*The British Journal of Psychology*,**19**, 306-327.

[4] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, **24**, 417-441 and 498-520.

[5] Lawley, D. N. and Maxwell, A. E. (1963), *Factor Analysis as a Statistical Method*, 2nd edition (1971), London: Butterworth.

[6] Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*, New York: Houghton-Mifflin.

[7] Mackintosh, N. J. (1998), *IQ and Human Intelligence*, Oxford: Oxford University Press.

[8] Molenaar, P. C. M and von Eye, A.(1994) On the arbitrary nature of random variables, in *Latent Variables Analysis*, Thousand Oaks: Sage Publications, 226-242.

[9] Spearman, C. (1904). General intelligence objectivel determined and measured, *American Journal of Psychology*, **15**, 201-293.

[10] Spearman, C. (1927 and 1932). *The Abilities of Man*, London: Macmillan.

[11] Spearman, C. and Jones, L. W. (1950). *Human Abilities: a continuation of the "Abilities of Man"*, London: Macmillan.

[12] Thomson, G. (1950). *The Factorial Analysis of Human Ability*, 4th edition, London: University of London Press.

[13] Yule, G. U. (1897). On the theory of correlation, *Journal of the Royal Statistical Society*,**60**, 812-851. 831-.