On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.

Boomsma, Anne

6. *MONTE CARLO SAMPLING FROM NON-NORMAL DISTRIBUTIONS*

In the previous chapter it was decided to characterize non-normality by the skewness of the observed variables and by categorizing these variables. In this chapter a numerical solution is presented how to generate the sample covariance matrices $\underset{\sim}{S}$ of such discrete and skew variables, while at the same time the true population covariance matrix $\underset{\sim}{\Sigma}$ is known and has a specified structure. Before going into the technical details of such procedures, a theoretical discussion of two possible strategies in treating the non-normal case is given in section 6.1. It will be seen that the true structural model can either hold for unobserved latent variables or for the observed variables and that a choice has to be made between these two approaches. In section 6.4 an overview of the procedure for non-normal observations is given.

It should be stressed that throughout this chapter the concept "*latent variable*" denotes an unobserved normally distributed variable, which has a *non-linear* relation to the observed variable due to a specific categorization of that latent variable. In section 6.2 the relationship between the latent variables and the observed variables is mathematically specified. At this place the reader should note the difference between this new concept and the latent variables in the LISREL-model, which are tied to the observed variables by *linear* relationships of the type $\underset{\sim}{y} = \underset{\sim}{\Lambda}_y \underset{\sim}{\eta} + \underset{\sim}{\varepsilon}$ or $\underset{\sim}{x} = \underset{\sim}{\Lambda}_x \underset{\sim}{\xi} + \underset{\sim}{\delta}$. [cf (2.2) and (2.3)]

The idea of an underlying normal variable which corresponds to a classified, discrete observed variable can be traced back to the work of Karl Pearson (1901). He suggested the tetrachoric correlation coefficient as a measure of bivariate normal correlation. Although Pearson liked the mathematical niceties of bivariate normal distributions he was worried from the start about the real life situation, where skew and discrete variables frequently occurred(Pearson, 1895, 1913). Generalization of the tetrachoric correlation coefficient to more than two classes in the observed variables leads to the polychoric correlation coefficient. Recent work on polychoric correlation is discussed by Olsson (1979a),who also derived asymptotic covariance matrices for estimates of polychoric correlations.

6.1 Considering the true model

Some of the ideas which led to our solution of the non-normal sampling problem arose from Olsson's (1978) dissertation, where a study on the robustness of factor analysis against crude classification of the observations is reported (see also Olsson, 1979b). Crucial in that work is the hypothesis that the observed variables are obtained through a classification of some true, latent variables which have a multivariate normal distribution and for which the (factor analysis) model holds. Olsson numerically demonstrates the relationships between the true latent model and the results obtained from classified observations. A preliminary reflection on his study and its relations to our work can be found in Boomsma (1980a).

The present section serves to clarify that basically *two approaches* can be distinguished in handling non-normal observations.

case a: an approach in which the structural model holds for the observed non-normal variables, and
case b: an approach in which the structural model is true with respect to the latent variables.

The differences between both approaches will be treated now.

Let us assume that there are functional relationships G_i between latent continuous variables ζ_i^* which cannot be observed directly, and observed variables $z_i (i = 1, ..., k)$. It is furthermore assumed that the vector $\underset{\sim}{\zeta}^*$ has a multivariate standard normal distribution, while $\underset{\sim}{z}$ has some non-normal multivariate distribution. The non-normal distributions of z_i might be *any* non-normal distributions, continuous as well as discrete; variables z_i might be numerically discrete, ordinal as well as nominal.

Next, since $\underset{\sim}{\zeta}^*$ is standardized, let us denote the *correlation matrix* of the latent variables $\underset{\sim}{\zeta}^*$ by $\underset{\sim}{\Sigma}^*$ and that of the observed variables $\underset{\sim}{z}$ by $\underset{\sim}{\Sigma}$, while the sample covariance matrix (which might be a correlation matrix) based on the observed scores is $\underset{\sim}{S}$. The basis for the two different approaches thus lies in the answer to the question whether $\underset{\sim}{\Sigma}^*$ or $\underset{\sim}{\Sigma}$ satisfies the structural equation model under study. Each approach has consequences for a possible structural analysis and for the design in a robustness study; these will be discussed in more detail now.

Case a

If *the true model is satisfied by* $\underset{\sim}{\Sigma}$, and a LISREL-analysis is performed on the basis of $\underset{\sim}{S}$, a wrong procedure is used, because the observed variables have a non-normal distribution while the maximum likelihood procedure expects a multivariate normal distribution. Here the latent variables do not play any role of importance. In effect, if the multivariate distribution function of the true observed variables would be known (at least for numerical variables) maximum likelihood estimation procedures might be derived. The only *problem* that remains for a robustness study, if this approach would be chosen, is how to sample from which multivariate distribution, given $\underset{\sim}{\Sigma}$. In sections 6.2 and 6.3 it will be seen that specific functional relationships between $\underset{\sim}{\zeta}$ and $\underset{\sim}{z}$ can be chosen to solve the sampling problem if $\underset{\sim}{\Sigma}$ is the true model.

Case b

If, however, *the true model is satisfied by* $\underset{\sim}{\Sigma}^*$, and again a LISREL-analysis is carried out on the basis of the sample covariance matrix $\underset{\sim}{S}$ of non-normal variables, the main handicap is that the model is true for the latent variables and *not* for the observed non-normal variables.

The difficulties involved are illustrated for a factor analysis model. Suppose for example that a researcher doing a robustness study knows what the true model is (given $\underset{\sim}{\Sigma}^*$, multivariate normal variables), and let us assume that it is a factor analysis model. This means that if the correlation matrix $\underset{\sim}{\Sigma}^* = \underset{\sim}{\Lambda}^*\underset{\sim}{\Phi}^*\underset{\sim}{\Lambda}^{*\prime} + \underset{\sim}{\Theta}^*$ specifies such a model, all elements of the matrices on the right-hand side of this equation are known. However, $\underset{\sim}{\Sigma}^*$ is *not* the correct model for the observed variables (linked to the latent variables by some functional form), because that would be $\underset{\sim}{\Sigma}$.

If normal variables would be categorized to discrete and possibly skew variables, in a way described in section 6.2, the difference between the absolute value of the normal product-moment correlation and the corresponding so-called multinomial product-moment correlation is always non-negative: $|\rho^*_{ij}| - |\rho_{ij}| \geqslant 0$ (see Lord & Novick, 1968, p. 346 for dichotomized variables). So, if LISREL would be used to estimate $\underset{\sim}{\Sigma}^*$ on the basis of $\underset{\sim}{S}$ the fit would be imperfect, because of the discrepancies between $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\Sigma}^*$. For example, if $\underset{\sim}{\Sigma}^* = \underset{\sim}{\Lambda}^*\underset{\sim}{\Phi}^*\underset{\sim}{\Lambda}^{*\prime} + \underset{\sim}{\Theta}^*$ specifies a one-factor model, and $\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\ \underset{\sim}{\Phi}\ \underset{\sim}{\Lambda}' + \underset{\sim}{\Theta}$ is the analogous decomposition of $\underset{\sim}{\Sigma}$, the corresponding order of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Phi}$, let alone the values of the elements of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Theta}$, are *unknown*. It is very unlikely that, given $\underset{\sim}{\Sigma}^*$, a one-factor model also fits $\underset{\sim}{\Sigma}$, and very probable that more than one factor is needed (o.c., p. 382).

Let us summarize what can and what cannot be done in case b. Given that $\underset{\sim}{\Sigma}^* = \underset{\sim}{\Lambda}^*\underset{\sim}{\Phi}^*\underset{\sim}{\Lambda}^{*\prime} + \underset{\sim}{\Theta}^*$ specifies the true factor analysis model, and given the estimate $\hat{\underset{\sim}{\Sigma}}^*$ based on $\underset{\sim}{S}$ (multivariate non-normal variables), by using the LISREL-procedure

(i) the parameter estimates and their corresponding standard errors can be compared with the true values, which are known,

(ii) the approximate chi-square goodness of fit statistic can be compared with the known true chi-square distribution with a specified number of degrees of freedom.

But in both cases, (i) and (ii), it is clear from the start that there will be discrepancies in the comparisons because of the differences between $\underset{\sim}{\Sigma}^*$ and $\underset{\sim}{\Sigma}$!

Olsson's (1978) computations exactly show what happens in such a situation. In his study the true model for the latent variables is a one-factor model. A maximum likelihood estimation procedure M was used, which can be summarized by $M(\underset{\sim}{S} = \underset{\sim}{\Sigma}|\underset{\sim}{\Sigma}^*, MND) = \hat{\underset{\sim}{\Sigma}}^*$, meaning that on the basis of the perfect covariance matrix $\underset{\sim}{S}$, given the true population model $\underset{\sim}{\Sigma}^*$ and an estimation procedure assuming that the latent variables have a multivariate normal distribution (MND), an estimation of the population covariance matrix, $\hat{\underset{\sim}{\Sigma}}^*$, is made. His results for goodness of fit and for the bias of computed factor loadings very well show the discrepancies which might occur, depending on the skewnesses and the number of categories of the observed variables.

What *cannot* be done in case b is to evaluate parameter estimates, corresponding standard errors and the chi-square statistic for goodness of fit under the model $\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\ \underset{\sim}{\Phi}\ \underset{\sim}{\Lambda}' + \underset{\sim}{\Theta}$, because this model is simply *unknown*.

Given these two approaches a decision must be made what our point of view *in this robustness study* will be. Should the true model hold for the normal latent variables or for the non-normal observed variables? Clearly, *two aspects* of robustness must be distinguished if the LISREL-procedure is used. *First*, the robustness against non-normal observations (including discrete classification) in a pure sense, where the observed non-normal variables are confronted with a maximum likelihood procedure which assumes a multivariate normal distribution (relevant in case a as well as in case b). *Secondly*, the robustness against the wrong model fit, e.g. $M(\underset{\sim}{S}|\underset{\sim}{\Sigma}^*,MND) = \hat{\underset{\sim}{\Sigma}}^*$ (relevant in case b, not in case a).

This creates a rather complex situation. The theoretical idea of continuous latent variables is attractive and sometimes unavoidable. If the model would hold for these latent variables, in our robustness study we could simply sample from a multivariate normal distribution and observe what the effect of a certain categorization (with or without skewness side effects) of the normal variables is. That is what Olsson (1978) did. Or we might want to transform the normal variables to non-normal continuous variables. But for sure we would fit the wrong model. And it would be very hard, if not impossible, to disentangle the two aspects of robustness just mentioned.

On the other hand, for our purposes it would be very nice to know what the true model is for the observed variables, what the values of parameters and standard errors are in the non-normal population, like we knew it all the time in the normal population for the small sample case. Such an approach would have the advantage that only one main aspect of robustness, that of the use of non-normal observations, is dealt with. As mentioned on page 136 there is only one restriction: a solution must be found for the problem how to obtain random samples from a discrete multivariate distribution with a specified $\underset{\sim}{\Sigma}$. In the next sections it will be seen that there is a conditional solution to this problem.

It was therefore decided that in our robustness study the true model would hold for the observed non-normal variables; the structural model would be satisfied by $\underset{\sim}{\Sigma}$, not by $\underset{\sim}{\Sigma}^*$. This decision can be *summarized* by saying that it was decided to choose $M(\underset{\sim}{S}|\underset{\sim}{\Sigma},MND) = \hat{\underset{\sim}{\Sigma}}$ as the estimation procedure for the non-normal case (in fact case a).

So far the discussion was rather general. We did neither specify the functional relationship between $\underset{\sim}{\zeta}^*$ and $\underset{\sim}{z}$, nor the non-normality of $\underset{\sim}{z}$ in distributional terms. This was postponed in an attempt to illustrate that the problems discussed are indeed of a general kind, not restricted to discrete observations alone. Our choice for discrete skew variables has empirical grounds, which were treated in chapter 5 . This choice now has to be made explicit in an operational procedure how to sample from a non-normal distribution with specified $\underset{\sim}{\Sigma}$.

Before that problem is dealt with in the following sections, two final points should be made here. *First*, the observed variables which are chosen are assumed to be *discrete* (or discretized), *numerical variables* with sample covariance matrix $\underset{\sim}{S}$. In section 6.2 it will be seen that for the non-normal case the observed values have been assigned discrete values, which happen to be non-negative integers (see page 141). It would also be possible to assume that the variables are ordinal, or nominal, but there is one reason why we did not do so. Since the scales of ordinal and nominal variables are arbitrary, there is no unique true population model. Such a model depends on what numbers are assigned to the categories. On the other hand, a unique true population model for $\underset{\sim}{\Sigma}$ may exist if the variables are discrete and numerical. *Secondly*, in this section the pro's and contra's of case a and case b have been discussed primarily within the framework of our robustness study, apart from considerations with regard to the practical application of structural equation models; the consequences of both approaches were compared from the viewpoint of a researcher doing Monte Carlo work. As such, the function of introducing the concept of latent variables and their relationships with observed variables served mainly technical purposes of a specified sampling design: the exposition was a theoretical discourse with technical, Monte Carlo purposes. If the applied statistician would be restricted to the LISREL-program, he of course has no choice from a practical point of view: he had to use

$M(S|\Sigma^*,MND) = \hat{\Sigma}^*$, whether his observed variables are multivariate normal, or whether they are not, or leave it. In applied situations one can undoubtedly think of different analytical methods M which can be used, depending on the theoretical and observed characteristics of variables, of the availability of programs which may or may not be suited for non-normal variables, and which may or may not assume specific distributional properties. It is up to the user with his theoretical knowledge of the variables under study, whether latent normal variables are assumed or whether they are not.

6.2 The relationship between latent and observed variables

For the univariate case there are relatively easy methods for generating non-normal distributions with a controlled degree of skewness and kurtosis (Fleishman, 1978). However, the introduction of non-normality into a multivariate distribution makes it difficult, especially when the number of variables is larger than two and a specified covariance structure is desired.

Ideally, we would generate random multinomial deviates from a multivariate multinomial distribution with population covariance structure Σ. To our knowledge there are no subroutines available in statistical computer libraries which can handle this, like the IMSL-routines used for the normal, small sample case. So we must either make such a program of our own, which might be tedious, or find another solution, which we did.

The problem was attacked by specifying the relationship between latent variables ζ^* and observed variables z. Once more, the only role of ζ^* in what follows is that it is of help in finding a solution for the non-normal sampling problem. First, together with the basic concepts some notation will be introduced, partly following Olsson (1978) for reasons of analogy and efficiency.

In the following it is assumed that there is a relationship between a latent continuous variable ζ_i^*, which cannot be observed directly, and an observed discrete variable z_i $(i=1, \ldots, k)$. It is furthermore assumed that each ζ_i^* has a standard normal distribution. From now on, for reasons of simplicity in most notations the index i is omitted as long as the relationship for one variable is discussed. If τ_s (s=1, ..., r) are thresholds in the standard normal distribution for variable i, which are unknown in practice, the monotone relationship between the latent variable and the observed variable is given by the following prescriptions:

$$
\begin{array}{llll}
\text{if} & \zeta^* < \tau_1 & \text{then } z = 0 ; \\
\text{if} & \tau_1 \leq \zeta^* < \tau_2 & \text{then } z = 1 ; \\
\text{if} & \tau_2 \leq \zeta^* < \tau_3 & \text{then } z = 2 ; \\
 & \vdots & \quad\;\, \vdots \quad\; \vdots \\
\text{if} & \tau_r \leq \zeta^* & \text{then } z = r .
\end{array}
\tag{6.1}
$$

Such a relationship between ζ^* and z is illustrated in Figure 6.1, where the observed variable has five categories and where the skewness of the discrete distribution as depicted by a symmetric histogram equals zero.

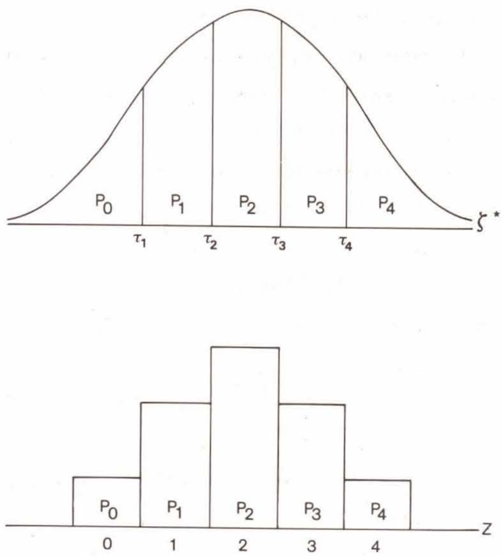

Figure 6.1 *The relationship between the latent variable* ζ^* *and the observed variable z, with skewness* $\gamma = 0$ *and a number of categories* $c = 5$; P_s $(s = 0,\ldots,4)$ *denotes the probability under the distribution functions.*

As has been stated previously, the main characteristics of the observed variables we want to use are the skewness γ and the number of categories c (or its number of thresholds $r = c-1$). It can be understood that we want to specify γ and c for each variable under study. Thus, first of all we need a family of distributions which can be manipulated on c categories and in which the skewness γ can be efficiently modified.

More or less arbitrarily it was chosen to characterize the distribution of z as a *binomial distribution*. That is the reason why non-negative integers have been assigned to z; the value of an observation depends on the category in which it falls.

It may be clarifying to see how the skewness and the kurtosis fluctuate as a function of the parameters r and p of a binomial distribution. These relations are exemplified in Table 6.1 for a number of categories ranging from two to seven, which is the range used in our robustness design (see chapter 7). For $r > 1$ the values for $p > .5$ are omitted because the kurtosis and the absolute value of the skewness are symmetrical about .5, as may be seen from the values for $r = 1$. By varying the number of categories and the skewness [cf. formula (6.3)] in a binomial distribution the kurtosis, $\{3+[1-6p(1-p)]/[rp(1-p)]\}$ is changed at the same time. Note how close the range of skewness and kurtosis in the empirical data sets (Tables 5.1 and 5.2) is reflected by the range of values in Table 6.1.

In Figure 6.1 the probabilities P_s are thus defined by the binomial probability distribution:

$$P_s = P(z=s|r,p) = \binom{r}{s} p^s(1-p)^{r-s}, \; (s=0,1, \ldots, r), \tag{6.2}$$

where r and p are the parameters of the binomial B(r,p) distribution.

In order for z to have the *same skewness* with a varying number of categories, the parameter p was adapted to the number of categories. Results are then comparable for different values of c.

The parameter p is thus a function of r and γ. For given r and γ, p can be solved from

$$\gamma = (1-2p)/[rp(1-p)]^{\frac{1}{2}}, \tag{6.3}$$

Table 6.1 *The values of the skewness and the kurtosis in binomial B(r,p) distributions.*

r	p	skewness	kurtosis
1	.1	2.67	8.11
	.2	1.50	3.25
	.3	.87	1.76
	.4	.41	1.17
	.5	.00	1.00
	.6	-.41	1.17
	.7	-.87	1.76
	.8	-1.50	3.25
	.9	-2.67	8.11
2	.1	1.89	5.56
	.2	1.06	3.13
	.3	.62	2.38
	.4	.29	2.08
	.5	.00	2.00
3	.1	1.54	4.70
	.2	.87	3.08
	.3	.50	2.59
	.4	.24	2.39
	.5	.00	2.33
4	.1	1.33	4.28
	.2	.75	3.06
	.3	.44	2.69
	.4	.20	2.54
	.5	.00	2.50
5	.1	1.19	4.02
	.2	.67	3.05
	.3	.39	2.75
	.4	.18	2.63
	.5	.00	2.60
6	.1	1.09	3.85
	.2	.61	3.04
	.3	.36	2.79
	.4	.17	2.69
	.5	.00	2.67

which is the skewness of the B(r,p) distribution.

Now the binomial probabilities P_s can be computed, which clearly depend on the number of categories and the skewness of variable z.

In Figure 6.1 the only unknown quantities are the thresholds τ_s. For given $P_{s-1}(s=1, \ldots, r)$ they can be computed by the inverse standard normal distribution function: $\tau_1 = \Phi^{-1}(P_0)$, $\tau_s = \Phi^{-1}[P_{s-1} + \Phi(\tau_{s-1})]$, for $s = 2, \ldots, r$,

where

$$\Phi(\tau_s) = P(\zeta^* < \tau_s) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\tau_s} e^{-\frac{1}{2}x^2} dx \ , \qquad (6.4)$$

in which $\tau_0 = -\infty$ and $\tau_{r+1} = \infty$.

The functional relationship between a latent variable and an observed variable, as illustrated in Figure 6.1, has now been fully specified: crucial is that for a given number of categories and a given skewness a binomial distribution is chosen, resulting in probabilities P_s which are linked to the standard normal distribution by thresholds τ.

Our next goal is to investigate the relationship between *two variables* ζ_i^* and ζ_j^*, having a standardized bivariate normal distribution with product-moment correlation ρ_{ij}^*, where ρ_{ij}^* is an element of the correlation matrix $\underset{\sim}{\Sigma}^*$, which is of order kxk. (It is only for reasons of simplicity that $\underset{\sim}{\Sigma}^*$ is a correlation matrix and not a general covariance matrix.) As before, corresponding to the latent variables $\underset{\sim}{\zeta}^*$ there are observed, discrete variables $\underset{\sim}{z}$. The bivariate situation is illustrated in Figure 6.2, where z_i has four categories and z_j has three categories.

Given the correlation ρ_{ij}^* between two latent, bivariate normal distributed variables we are ultimately in a position to calculate the correlation ρ_{ij} between the corresponding observed, classified variables z_i and z_j. If, given the skewnesses and the number of categories, the thresholds for ζ_i^* are $\tau_{1,i}, \tau_{2,i}, \ldots, \tau_{r_i,i}$ and

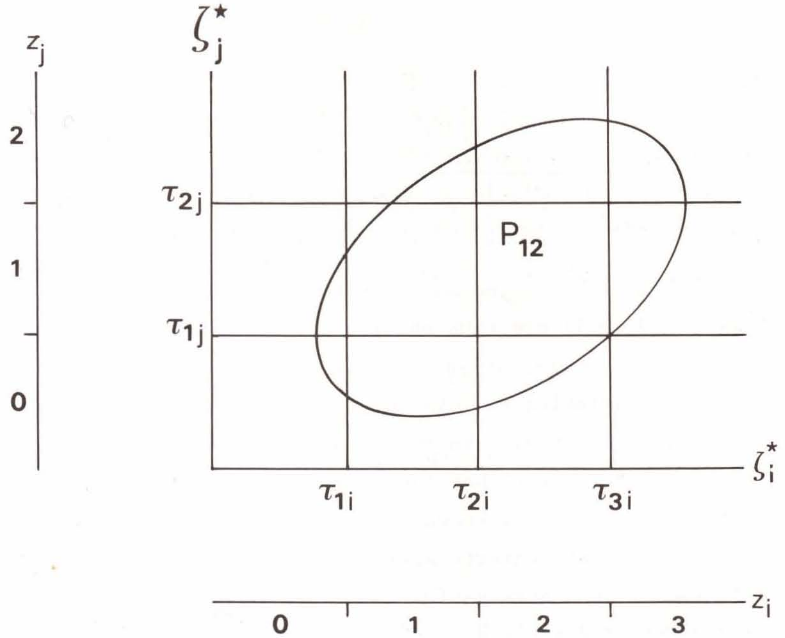

Figure 6.2 *The positive relationship between ζ_i^* and ζ_j^* depicted by an isodensity contour; z_i and z_j are the corresponding observed variables.*

those for ζ_j^* are $\tau_{1,j}, \tau_{2,j}, \ldots, \tau_{r_j,j}$, then the probability that $z_i = s$ and $z_j = t$ may be computed as

$$P_{st} = \int_{\tau_{s,i}}^{\tau_{s+1,i}} \int_{\tau_{t,j}}^{\tau_{t+1,j}} \phi(z_i, z_j | \rho_{ij}^*) \, dz_i dz_j, \tag{6.5}$$

where

$$\phi(z_i, z_j | \rho_{ij}^*) = [(2\pi)^2(1-\rho_{ij}^{*2})]^{-\frac{1}{2}} \exp\left[-\frac{z_i^2+z_j^2 - 2\rho_{ij}^* z_i z_j}{2(1-\rho_{ij}^{*2})}\right] \tag{6.6}$$

is the density of the standardized bivariate normal distribution with correlation ρ_{ij}^*.

Given the thresholds and ρ_{ij}^*, the bivariate discrete distribution of z_i and z_j is completely specified by P_{st} and their correlation may be computed as

$$\rho_{ij} = \frac{\sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,st - \sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,s \sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,t}{\left[\sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,s^2 - \left(\sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,s\right)^2\right]^{\frac{1}{2}}\left[\sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,t^2 - \left(\sum_{0}^{r_i}\sum_{0}^{r_j} p_{st}\,t\right)^2\right]^{\frac{1}{2}}} . \tag{6.7}$$

Such correlations will be called multinomial correlations.

From (6.5) and (6.7) one thus obtains the relationship between $\underset{\sim}{\Sigma}^*$ and $\underset{\sim}{\Sigma}$, where $\underset{\sim}{\Sigma}^*$ is the correlation matrix of latent variables $\underset{\sim}{\zeta}^*$ and $\underset{\sim}{\Sigma}$ that of observed variables $\underset{\sim}{z}$. This relationship will be basic for the next section, where it will be *reversed*. In our Monte Carlo study we need to compute $\underset{\sim}{\Sigma}^*$, given that the structural equation model is satisfied by $\underset{\sim}{\Sigma}$,which is fully known. In this section it was shown how to compute $\underset{\sim}{\Sigma}$ given $\underset{\sim}{\Sigma}^*$. Unfortunately, it is not possible to calculate $\underset{\sim}{\Sigma}^*$ given $\underset{\sim}{\Sigma}$ in a straightforward way. In section 6.3 an iterative procedure is described by which this problem can be solved.

6.3 The numerical solution

In section 6.1 it was decided that for the non-normal case the true population model should hold for the multivariate multinomial variables. In the previous section a specified relationship between $\underset{\sim}{\Sigma}^*$ and $\underset{\sim}{\Sigma}$ was formulated, taking discreteness and skewness into account. The solution of the problem how to solve the sampling process from a multivariate multinomial distribution with a true model $\underset{\sim}{\Sigma}$ can shortly be described as follows: take random samples from a multivariate normal distribution with $\underset{\sim}{\Sigma}^*$ corresponding to the true population model $\underset{\sim}{\Sigma}$.

The solution looks simple, but as will be seen (page 147) it is not without restrictions. The advantage of our proposal is that the basic generation of random deviates is exactly the same as for the normal, small sample case, making results at least comparable in that respect.

In order to compute $\underset{\sim}{\Sigma}^*$ corresponding to a given $\underset{\sim}{\Sigma}$ a Fortran computer program called NONNOR was written by the author (Boomsma, 1980b). The algorithm, which in its simplified form is shown in

Figure 6.3, can be summarized as follows.

1. Given that $\underset{\sim}{\Sigma}$ satisfies a LISREL-model, start with $\underset{\sim}{\Sigma}^*_0 = \underset{\sim}{\Sigma}$, and given the skewnesses and the number of categories find the corresponding multinomial $\underset{\sim}{\Sigma}_0$ by means of (6.5) and (6.7).
2. Use an iterative procedure resulting in a final $\underset{\sim}{\Sigma}^*_f$ which corresponds to a $\underset{\sim}{\Sigma}_f$ in which the elements $\rho_{ij,f}$ differ from the true values ρ_{ij} only to a small degree (for details see Figure 6.3).
3. Check whether $\underset{\sim}{\Sigma}^*$ is a positive definite matrix by investigating its eigenvalues.
4. If the smallest eigenvalue is positive the algorithm stops.
5. If there are eigenvalues less than zero, the in absolute sense largest negative eigenvalue is looked for; say its absolute value equals d.
6. In order to arrive at a positive definite correlation matrix the non-diagonal elements of $\underset{\sim}{\Sigma}^*_f$ are divided by $(1 + d + \varepsilon)$, where $\varepsilon = 10^{-5}$.Denote this adapted matrix by $\underset{\sim}{\Sigma}^*_{df}$. (Since for the models used in chapter 7 it was never necessary to adapt $\underset{\sim}{\Sigma}^*_f$, we do not comment on this provision).
7. Calculate $\underset{\sim}{\Sigma}_{df}$ corresponding to $\underset{\sim}{\Sigma}^*_{df}$ and see whether it is close enough to $\underset{\sim}{\Sigma}$. If it is, the algorithm stops. If it is not, the number of categories and/or the skewnesses are manipulated and a new attempt is made to find a suitable $\underset{\sim}{\Sigma}^*$ corresponding to $\underset{\sim}{\Sigma}$.

From Figure 6.3 it can be seen that the iterative procedure is first used to find ρ^*_{21} of $\underset{\sim}{\Sigma}^*$; successively all $\frac{1}{2}k(k - 1)$ different elements of $\underset{\sim}{\Sigma}^*$ are computed. In the algorithm provisions were made to prevent $|\rho^*_{ij}| \geq 1$.

It turned out that the algorithm is rather fast. In particular, when the skewnesses and the number of categories are chosen in *harmony* with the correlations ρ_{ij}, hardly ever more than 10 iterations are required for convergence.

This condition of harmony, which is the main *restriction* of our procedure,may well be illustrated. Suppose variable i is a dichotomized variable with $\gamma_i = 1$ and variable j is also a dichotomized variable with $\gamma_j = -2$, while $\rho_{ij} = 0.3$. It is then impossible to find a suitable $\rho^*_{ij} < 1$; if $\rho^*_{ij} = 0.99999$ the corresponding

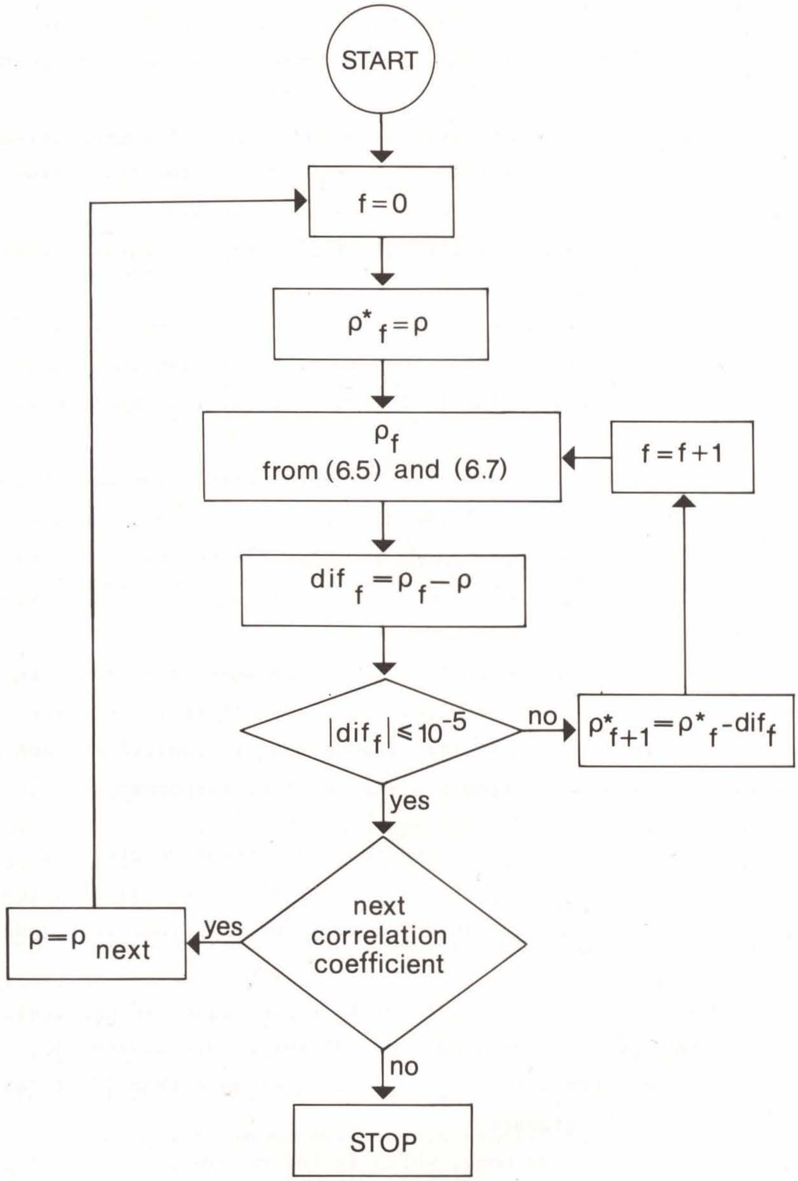

Figure 6.3 *Simplified flow-diagram for the iterative procedure which finds the multinormal $\rho^* = \rho^*_{ij}$ corresponding to the multinomial $\rho = \rho_{ij}$. Starting with ρ_{21} the algorithm treats all lower triangular elements of $\underset{\sim}{\Sigma}$.*

ρ_{ij} = 0.26, which under the circumstances is the largest possible multinomial correlation coefficient.

6.4 Overview of the non-normal procedure

The handling of the non-normal case can briefly be described as follows:

1. For a given structural equation model with $\underset{\sim}{\Sigma}$, the number of categories c_i and the skewnesses γ_i, (i=1, ..., k) are selected. This is done by taking the elements ρ_{ij} of $\underset{\sim}{\Sigma}$ into account, so as to avoid a selection of c_i and γ_i which would be unlikely or impossible, given the size and direction of those elements.
2. With NONNOR an attempt is made to find the multinormal covariance matrix $\underset{\sim}{\Sigma}^*$, given $\underset{\sim}{\Sigma}$, c_i and γ_i, (i=1, ..., k). If such an attempt fails some c_i and/or γ_i are adapted in such a way that a positive definite $\underset{\sim}{\Sigma}^*$ can be computed, where the corresponding multinomial correlation matrix is close enough to $\underset{\sim}{\Sigma}$.
3. With SIMLIS Nxk random normal deviates are generated from a multivariate normal distribution with covariance matrix $\underset{\sim}{\Sigma}^*$ (see section 3.1). These multinormal random deviates are transformed to multinomial random deviates according to (6.1). At this stage a remark should be made about the next step, which is the transformation of the matrix of multinomial random deviates, say $\underset{\sim}{Z}$, to either a sample covariance matrix or a sample correlation matrix. This point was discussed earlier in section 4.1.

   If a *sample correlation matrix* is wanted each multinomial sample covariance matrix $\underset{\sim}{S}$ is standardized by its own *sample* variances: $\underset{\sim}{V}_s^{-\frac{1}{2}}\underset{\sim}{S}\,\underset{\sim}{V}_s^{-\frac{1}{2}}$, where $\underset{\sim}{V}_s^{-\frac{1}{2}}$ is a diagonal matrix of order kxk, and the i-th element of the diagonal of $\underset{\sim}{V}_s^{\frac{1}{2}}$ is the sample standard deviation of variable i.

   If, however, a proper multinomial *sample covariance matrix* is wanted, which can be viewed as a sample from $\underset{\sim}{\Sigma}$, also some standardization has to be performed, because of the difference between multinomial correlations and multinomial covariances. In this chapter the relationship between $\underset{\sim}{\Sigma}^*$ and $\underset{\sim}{\Sigma}$ was treated in terms of correlations. The generated samples $\underset{\sim}{S}$, however, are multinomial covariance samples, with multinomial sample variances on their diagonals. These variances are very restricted because of the specific transformation from continuous multinormal deviates to discrete multinomial deviates. Therefore, each multinomial sample covariance matrix $\underset{\sim}{S}$ is standardized by the multinomial *population* variances: $\underset{\sim}{V}_{\sigma}^{-\frac{1}{2}} \underset{\sim}{S} \underset{\sim}{V}_{\sigma}^{-\frac{1}{2}}$, where $\underset{\sim}{V}_{\sigma}^{-\frac{1}{2}}$ is a kxk diagonal matrix ; the i-th diagonal element of $\underset{\sim}{V}_{\sigma}^{\frac{1}{2}}$ equals $[r_i p_i (1 - p_i)]^{\frac{1}{2}}$, the standard deviation of the multinomial variable i in the population.

The final results after this stage are NRS multinomial random sample covariance (or correlation) matrices $\underset{\sim}{S}$.

4. The following steps are exactly equal to those for the normal, small sample case (see section 3.2 for a detailed description). The sample covariance matrices $\underset{\sim}{S}$ are taken as input for LISREL-analyses. Again, with LISREP a fixed number of replications NR < NRS is analyzed.
5. Finally, the estimates of parameters, standard errors and the goodness of fit statistic, gathered for NR replications, are analyzed by means of DISLIS. The results of these analyses, which is mainly a comparison of theoretical and empirical sampling distributions, are reported in the next chapter for different kinds of models.