Multivariate distributions with correlation matrices for nonlinear repeated measurements.

J.K. Lindsey[†]
P.J. Lindsey[‡]

October 3, 2002

Abstract: The polynomial growth curve model based on the multivariate normal distribution has dominated the analysis of continuous longitudinal repeated measurements for the last 50 years. The main reasons include the ease of modelling dependence because of the availability of the correlation matrix and the linearity of the regression coefficients. However, a variety of other useful distributions also involve a correlation matrix: the multivariate Student t, multivariate power-exponential, and multivariate skew Laplace distributions, as well as Gaussian copulas with arbitrarily chosen marginal distributions. As well, with modern computing power and software, nonlinear regression functions can be fitted as easily as linear ones.
By a number of examples, we show that these distributions, combined with nonlinear regression functions, generally yield an improved fit, as compared to the standard polynomial growth curve model, and can provide different conclusions.

Keywords: Copula, Laplace distribution, power-exponential distribution, Student t distribution.

[†]Biostatistics, Limburgs Universitair Centrum, Diepenbeek
(Email: jlindsey@luc.ac.be and Web: www.luc.ac.be/~jlindsey)
[‡]Eurandom, Eindhoven

1 Introduction

Box (1950) introduced the polynomial growth curve model for longitudinal repeated measurements. Elston and Grizzle (1962) extended it to random time coefficients (now often unfairly called the Laird and Ware model) in order to provide more flexible correlation matrices. Since then, the analysis of such data has relied primarily upon polynomials in time and the multivariate normal distribution. Reasons include

- wide familiarity with the normal distribution,
- linearity of the parameters in polynomial functions,
- belief in 'robustness' of the results,
- availability of standard analysis of variance procedures if the regression function is linear and the data are balanced,
- relative ease of manipulation of the likelihood function,
- availability of the correlation matrix to model dependence among responses.

A number of these reasons have become less compelling with the availability of modern computers and statistical software.

With the development of a variety of multivariate distributions involving correlation matrices, the last reason also disappears. As emphasized by Taylor and Law (1998), the covariance structure can be extremely important in longitudinal modelling. Random coefficient models based on polynomials in time are an *ad hoc* construction, generally with no plausible interpretation (see for example Elston, 1964; Lindsey, 1999, pp. 100–102; Davis, 2002, pp. 149–151).

The classical competitor of the normal distribution has been the multivariate Student t distribution, especially useful if the data are overdispersed with respect to the normal distribution, that is, have outliers with respect to it. However, other multivariate distributions have also recently become available: power-exponential (Gómez *et al.*, 1998) and skew Laplace (Kotz *et al.*, 2001) distributions, as well as Gaussian copulas (Song, 2000) with arbitrarily chosen marginal distributions. Many of these are skewed and/or have heavier tails than the multivariate normal, which can be useful in providing robustness against 'outliers'.

These multivariate distributions are not members of the exponential family. Thus, expanding the choice of distributions in this way implies the loss of the ease of estimation provided by sufficient statistics so that a nonlinear optimizer is required. But this has the compensating advantage that nonlinear regression functions can be fitted as easily as linear ones.

If open software for modelling with the multivariate normal distribution is available, it can very easily be modified to accommodate all of these distributions. Almost all of the required components are already being computed and one has only to change the formula for the final likelihood being calculated. Depending on the distribution used, one will also require the gamma

function (Student t and power-exponential), a modified Bessel function (skew Laplace), and the normal quantile function as well as cumulative distribution functions of any desired marginal distributions (Gaussian copulas).

Models based on all of these multivariate distributions are now available in two functions for the R software system (Ihaka and Gentleman, 1996) and can be obtained from the first author. This is accompanied by a nonlinear function interpreter which allows any nonlinear regression function easily to be specified and fitted. An illustration is provided in the appendix below.

2 Multivariate Distributions

We first briefly describe the various multivariate distributions available. For further details, the reader is referred to the literature cited.

2.1 Normal distribution

First let us recall the form of the multivariate normal distribution for R repeated measurements,

$$f(\mathbf{y};\boldsymbol{\mu},\Sigma) = \frac{\mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}}{(2\pi)^{\frac{R}{2}}\sqrt{|\boldsymbol{\Sigma}|}} \tag{1}$$

with covariance matrix $\boldsymbol{\Sigma}$. However, a number of other families of multivariate distributions also have such a matrix. Like the multivariate normal distribution, many, but not all, of them are elliptically-contoured. Thus, skewed distributions are also available. The normal distribution is often skewed by taking logarithms of the response variable, yielding the log normal distribution; this can also be done with the other symmetric multivariate distributions presented below.

A correlation or covariance matrix can have any desirable structure. For simplicity here, we shall restrict attention to two of the basic possibilities: constant correlation among all observations on an individual (sometimes called compound symmetry) and the simplest way in which correlation decreases with distance in time (stationary autocorrelation).

Consider the covariance matrix. With N individuals and R repeated measurements on individual i, a variance component for intra-individual correlation can be modelled as

$$\begin{aligned}\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma^2+\delta & \delta & \cdots & \delta \\ \delta & \sigma^2+\delta & \cdots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \cdots & \sigma^2+\delta \end{pmatrix} \\ &= \sigma^2\mathbf{I}_R + \mathbf{J}_R^T\delta\mathbf{J}_R\end{aligned}$$

where $\mathbf{I}_R$ is an $R \times R$ identity matrix and $\mathbf{J}_R$ a vector of ones of length R. Here, σ^2 is the intra-individual variance and δ is both the extra component of variance across individuals and the common covariance among responses on the same individual. In the context of *linear* normal models, this corresponds to a random intercept. (For those familiar with the Laird and Ware

notation, we have $\mathbf{Z} = \mathbf{J}_R$.) Random coefficient models for polynomials in time can also be written in this way, but have a more complex form; for examples, see Lindsey (1999a, pp. 93–102).

If serial correlation is present, the covariance matrix might be structured in the following way:

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^{R-1} \\ \rho & 1 & \cdots & \rho^{R-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{R-1} & \rho^{R-2} & \cdots & 1 \end{pmatrix}$$

This is a first-order dependence structure for equally-spaced times, a stationary AR(1); the extension to unequally-spaced times is straightforward. Dependencies involving other functions of time, perhaps nonstationary, (see, for example, Lindsey, 1999a, pp. 124–127) can also be used, as well as higher order dependencies, although these are not often necessary for the short series usually available in repeated measurements. Variance components and serial correlation can be combined in the same covariance matrix but recall that *any* suitable structure is possible. In the examples below, the dispersion parameter is often not constant in time, inducing a different type of nonstationarity.

In constructing models based on the multivariate normal distributions, the covariance matrix can be written simultaneously as one large matrix for all observations on all individuals: for example, if observations are at the same time points for all individuals, $\boldsymbol{\Sigma}_N = \mathbf{I}_N \otimes \boldsymbol{\Sigma}$. The appropriate elements are zero so that observations on different individuals are independent. This is equivalent to taking a separate multivariate normal distribution for each individual, with a suitable covariance structure, and multiplying them together.

2.2 Student t distribution

A well known alternative to the multivariate normal distribution is the multivariate Student t distribution, defined by

$$f(\mathbf{y}; \boldsymbol{\mu}, \Sigma, \kappa) = \frac{\Gamma\left(\frac{\kappa+R}{2}\right)}{\pi^{\frac{R}{2}} \sqrt{|\Sigma|} \Gamma\left(\frac{\kappa}{2}\right) \kappa^{\frac{R}{2}} \left[1 + \frac{1}{\kappa}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})\right]^{\frac{\kappa+R}{2}}}$$

where κ is the number of degrees of freedom. This is a member of the elliptically-contoured family of distributions. The mean is $\boldsymbol{\mu}$ for $\kappa > 1$ and the covariance matrix is

$$\mathrm{cov}(\mathbf{Y}) \quad = \quad \frac{\kappa}{\kappa - 2} \boldsymbol{\Sigma}, \qquad \kappa > 2$$

Note that zero correlation does not imply independence. Here, $\kappa \to \infty$ yields a multivariate normal distribution, whereas $\kappa = 1$ is a multivariate Cauchy distribution.

The procedure for constructing one large covariance matrix for all individuals, described above for the multivariate normal distribution, no longer works here if $\kappa < \infty$. When $\boldsymbol{\Sigma}$ is diagonal so that the correlation among observations is zero, this distribution cannot be written as a product of independent univariate distributions. The multivariate distribution retains a dependence structure

among the observations on an individual even though the correlation among them may be zero. Of course, responses on different individuals can be made independent by multiplying together the multivariate distributions on individuals in the usual way. These remarks hold for the next two multivariate distributions as well.

2.3 Power-exponential distribution

The multivariate power-exponential distribution given by

$$f(\mathbf{y};\boldsymbol{\mu},\Sigma,\kappa) = \frac{R\Gamma\left(\frac{R}{2}\right)}{\pi^{\frac{R}{2}}\sqrt{|\boldsymbol{\Sigma}|}\,\Gamma\left(1+\frac{R}{2\kappa}\right)2^{1+\frac{R}{2\kappa}}}\exp\left\{-\frac{1}{2}\left[(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]^{\kappa}\right\} \quad (2)$$

is also a member of the elliptically-contoured family (Gómez *et al.*, 1998; Lindsey, 1999b). It is a special case of what has been called a Kotz-type distribution (Kotz and Nadarajah, 2001).

Again $\boldsymbol{\mu}$ is the mean and the covariance matrix is

$$\mathrm{cov}(\mathbf{Y}) = \frac{2^{\frac{1}{\kappa}}\Gamma\left(\frac{R+2}{2\kappa}\right)}{R\Gamma\left(\frac{R}{2\kappa}\right)}\boldsymbol{\Sigma}$$

κ determines kurtosis; when $\kappa \neq 1$, the even cumulants are nonzero, in contrast to the multivariate normal distribution.

When $\kappa = 1$, this is a multivariate normal distribution, when $\kappa = 0.5$, a form of multivariate Laplace (double exponential) distribution discussed by Kotz *et al.* (2001, pp. 312–313), and when $\kappa \to \infty$, a multivariate uniform distribution. The marginal and conditional distributions are more complex, elliptically-contoured distributions, not of the power-exponential type.

2.4 Skew Laplace distribution

The multivariate power-exponential distribution contains, as a special case, one possible definition of a multivariate Laplace distribution. Kotz *et al.* (2001, p. 250) give another possibility:

$$f(\mathbf{y};\boldsymbol{\Sigma},\boldsymbol{\kappa}) = \frac{2e^{\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}}}{(2\pi)^{R/2}|\boldsymbol{\Sigma}|^{1/2}}\left(\frac{\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}}{2+\boldsymbol{\kappa}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}}\right)^{(2-R)/4} K_{(2-R)/2}\left(\sqrt{(2+\boldsymbol{\kappa}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa})\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}}\right) \quad (3)$$

where $K_u(\cdot)$ is the modified Bessel function of the third kind and $\boldsymbol{\kappa}$ is a vector of skew parameters, one for each observation. For $\boldsymbol{\kappa} = \mathbf{0}$, this is a symmetric multivariate Laplace distribution which is elliptically-contoured; otherwise, it is not a member of that family.

Here, the mean is $\boldsymbol{\kappa}$ and the covariance matrix

$$\mathrm{cov}(\mathbf{Y}) = \boldsymbol{\Sigma} + \boldsymbol{\kappa}\boldsymbol{\kappa}^{\mathrm{T}}$$

As with the previous two distributions, zero correlation does not imply independence.

A regression function can be introduced in at least two distinct ways. If $\mathbf{Y}$ has an asymmetric multivariate Laplace distribution and we want the mean to be determined by the regression function, we can set $\boldsymbol{\kappa}$ equal to that function. This, however, means that the amount of symmetry is also changing with the regression function.

A second possibility, suggested by Kotz *et al.* (2001, pp. 261–268), is to introduce the regression function as a location shift; they propose this for linear regression with zero correlation ($\boldsymbol{\Sigma} = \mathbf{I}_R$). Here, we extend it to nonlinear regression functions and model the correlation matrix. In this situation, $\mathbf{y}$ in Equation (3) is replaced by $\mathbf{y} - g(\boldsymbol{\beta}, \mathbf{X})$ where $g(\cdot)$ is some linear or nonlinear regression function. Note that $\boldsymbol{\kappa}$ is not a shift parameter so that $\mathbf{Y} - g(\boldsymbol{\beta}, \mathbf{X})$, but not $\mathbf{Y}$, will have an asymmetric multivariate Laplace distribution (Kotz *et al.*, 2001, p. 244). When this distribution is applied to repeated measurements, it seems most sensible to fix all elements of $\boldsymbol{\kappa}$ to be identical so that the distribution of all repeated responses has the same asymmetry and only one additional parameter need be estimated. In the examples to follow, this second model consistently fitted better so that it is the only one for which results are presented.

2.5 Copulas

One approach to the construction of multivariate distributions is by the specification of the univariate marginals. This is not sufficient to define completely the distribution. Many supplementary conditions have been proposed, usually for specific marginals.

Suppose that the desired univariate marginal distributions are given by their cumulative distribution functions, $F(y_i)$. Then, a multivariate distribution with these marginals can be formed as

$$F_R(y_1, \ldots, y_R) = c[F(y_1), \ldots, F(y_R)]$$

where $c(\cdot)$ is a function from $[0,1]^R$ to $[0,1]$. Such a construction is called a copula. Although all of the margins are shown to have the same form here, they can, in principle, be different.

The equation above can be inverted to give

$$c(u_1, \ldots, u_R) = F_R[F^{-1}(u_1), \ldots, F^{-1}(u_R)]$$

where $F^{-1}(\cdot)$ is the corresponding quantile function and u_i is a uniform $[0,1]$ variable.

One particularly useful copula is the Gaussian (Song, 2000). The general form of the corresponding density can be obtained from the standardized multivariate normal density of Equation (1). We transform the response using a vector of univariate normal quantile functions, $\boldsymbol{\Phi}^{-1}(\mathbf{u})$ (without forgetting the Jacobian):

$$\frac{\partial^R c(\mathbf{u}; \boldsymbol{\Sigma})}{\partial u_1 \ldots \partial u_R} = \frac{\mathrm{e}^{-\frac{1}{2}[\boldsymbol{\Phi}^{-1}(\mathbf{u})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Phi}^{-1}(\mathbf{u}) + \frac{1}{2}[\boldsymbol{\Phi}^{-1}(\mathbf{u})]^{\mathrm{T}} \boldsymbol{\Phi}^{-1}(\mathbf{u})}}{\sqrt{|\boldsymbol{\Sigma}|}}$$

where $\boldsymbol{\Sigma}$ is a (Spearman-type) correlation matrix (not a covariance matrix as in the previous families).

When we substitute in the chosen univariate marginal distributions for $\mathbf{u}$, we obtain the multivariate density

$$f_R(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{\mathrm{e}^{-\frac{1}{2}\{\boldsymbol{\Phi}^{-1}[F(\mathbf{y};\boldsymbol{\theta})]\}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1} - \mathbf{I}_R]\boldsymbol{\Phi}^{-1}(F(\mathbf{y};\boldsymbol{\theta}))}}{\sqrt{|\boldsymbol{\Sigma}|}} \prod_{i=1}^{R} f(y_i; \boldsymbol{\theta})$$

where $F(\mathbf{y};\boldsymbol{\theta})$ is the vector of chosen univariate marginal cumulative distributions and $f(y_i;\boldsymbol{\theta})$ is the corresponding individual univariate density. If the margins are chosen to be normal, this reduces to the standard multivariate normal distribution of Equation (1) and if they are Laplace, we have a third type of multivariate Laplace distribution, not covered by Kotz *et al.* (2001). Then, as with the other distributions above, $\boldsymbol{\Sigma}$ can be structured in any desired way to create dependencies among the responses, here indirectly through $F(\mathbf{y};\boldsymbol{\theta})$.

3 Examples

We have applied the models described above to a large number of data sets. We have chosen to present here the analyses for several data sets that are available in the literature so that the reader can compare our approach to the standard ones. For lack of space, we do not provide complete details of the studies, nor of how the final model was obtained, but instead refer the reader to the relevant literature.

We concentrate on finding appropriate nonlinear regression functions for these data. One good check on such functions is also to fit a saturated mean or location model, but we shall not present results for this here. For simplicity and clarity, we restrict covariance modelling to the two basic structures described above. However, we emphasize that *any* appropriate form of matrix can be used and that better covariance structures may exist for these data.

In the analysis of cross-over data, modelling of the dependence structure is especially critical because treatments are compared within subjects so that ignoring or poorly modelling the dependence will generally lead to the significance of treatment effects being *under*estimated. Thus, we shall look at several examples involving such trials. But first we consider a more classical longitudinal repeated measurements study.

For inferences, we shall use the AIC: minus the log likelihood plus the number of estimated parameters. Thus, smaller values indicate more preferable models. This allows us to compare non-nested models. If the reader wishes to apply classical likelihood ratio tests, where applicable, all log likelihoods must be multiplied by two.

3.1 Coronary sinus potassium trial

Grizzle and Allen (1969) give measurements of coronary sinus potassium (mil equivalents per litre) in 36 nondescript dogs. Observations were made at two-minute intervals after coronary occlusion under four treatment conditions: A, control; B, bilateral thoracic sympathectomy and stellectomy three weeks prior to occlusion; C, extrinsic cardiac denervation immediately prior to occlusion; and D, extrinsic cardiac denervation three weeks prior to occlusion.

Analysis shows that the control differs from three treatments which are fairly similar. We have found empirically that the following sum of exponentials curve in time, with a total of nine

regression parameters, fits well:

$$\begin{aligned}\mu_t &= \alpha_1 + \alpha_2(x_B + x_C + x_D) + (1 - x_B)\alpha_3(\alpha_5 - t)\mathrm{e}^{-\mathrm{e}^{\alpha_4(\alpha_5 - t)}} \\ &\quad + \alpha_6(\alpha_8 - t)\left(x_C\mathrm{e}^{-\mathrm{e}^{\alpha_7(\alpha_8 - t)}} + x_D\mathrm{e}^{-\mathrm{e}^{\alpha_9(\alpha_8 - t)}}\right) \qquad (4)\end{aligned}$$

where x_j is an indicator variable for treatment j. Notice that, in this function, the mean or location parameter of the response to treatment B is constant over time.

For these data, the dispersion parameter depends log linearly on time. An AR(1) is also required, but no variance component, although this will be reconsidered below for the final model. The fitted models are compared in the first column of Table 1. Without transforming the data, we see that the Gaussian copula with gamma margins fits best, followed by the multivariate skew Laplace distribution.

The fact that the best models are based on skewed distributions seems to indicate that a log transformation might be preferable in the multivariate normal distribution, although this does not seem to have been used in the literature for these data. We keep the same regression model but apply an exponential link to μ_t so that the log response will follow the logarithm of the complete nonlinear relationship on the right hand side of Equation (4). Indeed, these models fit better for all of the distributions, even the Gaussian copulas with gamma, Weibull, and inverse Gauss margins. The (log) power-exponential now replaces the skew Laplace as the closest rival of the Gaussian copula with (log) gamma margins, as can also be seen in the first column of Table 1.

Each of these models has one more parameter than the multivariate log normal so that the difference in log likelihood is almost 4, or the deviance about 7.8, with one degree of freedom, clearly rejecting the latter by frequentist criteria as well as by the AIC. The log power-exponential distribution has $\hat{\kappa} = 0.48$, indicating a distribution close to a multivariate log Laplace distribution of Equation (2), not that of Equation (3) nor a Gaussian copula with Laplace margins.

The autocorrelation is estimated to be about 0.9 for the two best models. A variance component is not necessary for the copula but reduces the AIC to 113.5 for the log power-exponential. The latter now has an autocorrelation estimated to be $\hat{\rho} = 0.81$ and power parameter $\hat{\kappa} = 0.44$. On the other hand, the former can be simplified by keeping the dispersion parameter constant (AIC 113.9).

The final location function with the log power-exponential distribution is estimated to be

$$\begin{aligned}\mathrm{e}^{\mu_t} &= 4.16 - 0.655(x_B + x_C + x_D) + (1 - x_B)1.927(14.1 - t)\mathrm{e}^{-\mathrm{e}^{0.151(14.1 - t)}} \\ &\quad - 1.783(13.4 - t)\left(x_C\mathrm{e}^{-\mathrm{e}^{0.245(13.4 - t)}} + x_D\mathrm{e}^{-\mathrm{e}^{0.188(13.4 - t)}}\right) \qquad (5)\end{aligned}$$

with

$$\log(\sigma_t^2) = -9.28 + 0.0546t$$

for the dispersion parameter.

Standard analysis (Lindsey, 1999a, pp. 127–131) with the multivariate normal polynomial growth curve and an AR(1) requires a third degree polynomial; the AIC is 143.4. This is reduced to 138.4 if the log variance is allowed to depend on time and to 123.7 if a log transformation is also applied. Replacing the normal distribution with this polynomial by the log Student t gives 120.5 and by the log power-exponential 120.0. Thus, non-normality, nonconstant dispersion, and nonlinearity are all important for these data.

3.2 Insulin cross-over trial

Two mixtures of neutral protamine Hagedorn (NPH) insulin, the standard (A) and one containing 5% less protamine (B) were tested on rabbits in a cross-over design (Ciminera and Wolfe, 1953). Two groups of eleven female rabbits were injected with the insulin at weekly intervals in the orders ABAB (sequence 1) and BABA (sequence 2). For each treatment, blood sugar level was measured at injection and at four equally-spaced post-injection times over six hours. Lindsey (1999b) reanalyzed these data using the multivariate power-exponential distribution and a quadratic regression function. Here, we compare the distributions discussed above using nonlinear regression functions.

Two levels of variance components are possible: the rabbit and the period nested within rabbit; a first-order autocorrelation may also be necessary over the six hours. The explanatory variables are the period, the treatment, and the group, as well as a sum of exponential time effect. We have found the following regression function to fit well:

$$\mu_t = (\alpha_1 + \alpha_2 x_{14}) \left(e^{1 - \frac{t}{\alpha_3 + \alpha_4 x_{22} + \alpha_5 x_{3B}}} + e^{\frac{t}{\alpha_6 + \alpha_7 x_{14} + \alpha_8 x_{22}}} \right)$$

where x_{1j} is an indicator for period j, x_{2j} for group j, and x_{3j} for treatment j. Again, the log variance depends linearly on time. Only the rabbit-level variance component, and not the period-level one, is required, not surprisingly because period 4 is used in the regression function and is by far the most significant.

The resulting fits for the final model are compared for each distribution in the third column of Table 1. We see that the Student t and power-exponential distributions fit best. The parameters are estimated to be $\hat{\kappa} = 22.0$ for the Student t d.f. and $\hat{\kappa} = 0.50$ for the power-exponential. Both indicate a distribution with somewhat heavier tails than the multivariate normal. The Gaussian copulas with skewed marginal distributions fit less well than the other distributions. These results confirm that a symmetric distribution is the best choice for these data, at least among those tried.

The two best fitting models give almost identical parameter estimates for the nonlinear regression function:

$$\mu_t = (10.2 + 3.3 x_{14}) \left(e^{1 - \frac{t}{-3.4 + 0.5 x_{22} + 0.34 x_{3B}}} + e^{\frac{t}{-1.5 - 0.15 x_{14} + 0.06 x_{22}}} \right)$$

Of course, the variance component and the dependence of the dispersion parameter on time are quite different, although the latter increases over time in both cases. On the other hand, the autocorrelation is the same: $\hat{\rho} = 0.44$.

Eliminating the treatment effect by setting $\alpha_5 = 0$ raises the AIC to 1805.7 for the Student t and to 1805.8 for the power-exponential. (Thus, the log likelihood increases by about 3, or the deviance by about 6.2, with one degree of freedom.)

Standard analysis with the multivariate normal polynomial growth curve, a variance component for rabbits but not for period, and an AR(1) requires a second degree polynomial; the AIC is 1852.8. This is reduced to 1838.2 if the log variance is allowed to depend on time. Elimination of treatment effect here reduces the AIC to 1837.2. Hence, the standard polynomial growth curve model (even with nonconstant variance) not only fits much more poorly but is not sensitive enough to detect the treatment effect. The power exponential model with a quadratic polynomial has an AIC of 1833.3, decreasing to 1832.8 when the treatment effect is eliminated. Our nonlinear regression function provides a substantial improvement on this and allows clear detection of the treatment effect.

3.3 Pharmacokinetics cross-over trial

There is a strong belief among statisticians analyzing pharmacokinetic data with compartment models that concentrations always follow a log normal distribution. However, we are aware of no studies in which this has been demonstrated for any compound. In all cases in our experience, the log normal distribution has never proven suitable. We now consider one such data set. In contrast to the other examples, here a mechanistic nonlinear regression function is available.

Flosequinan was found, in early trials, to be useful in the management of patients with chronic heart failure. However, a long term study demonstrated that there might be increased mortality in patients taking the drug so that it was withdrawn from the market in the United Kingdom in April, 1993. A Phase I study looked at pharmacokinetic dose proportionality using 50, 100, and 150 mg doses in a cross-over design involving 18 healthy volunteers. Thus, 3 patients were assigned to each of the six possible sequences. Blood samples were taken at 0, 0.5, 1, 1.5, 2, 3, 4, 6, 8, 10, 12, 24, 36, 48, 72, and 96 hours after dosing. Concentrations of flosequinan and of its pharmacologically active metabolite, flosequinoxan, were measured. Time to peak concentration for the drug is less than one hour.

A nonlinear regression function, the open first-order one-compartment model, given by

$$\mu_t = \frac{k_a d}{V(k_a - k_e)} \left(\mathrm{e}^{-k_e t} - \mathrm{e}^{-k_a t}\right) \tag{6}$$

is suitable for these data, where d is the dose and t the time. The volume (V), absorption rate (k_a), and elimination rate (k_e) are parameters to be estimated. The dispersion parameter follows the regression function

$$\sigma_t^2 = \left(\frac{kdt\mathrm{e}^{-kt}}{V'}\right)^{\delta}$$

The expression in the parentheses is obtained as the limit of Equation (6) when $k_a = k_e = k$. Thus, there are three parameters in this function, k, V', and δ, all different from those in Equation (6).

Here, the distribution of the responses is clearly skewed, as is often the case with pharmacokinetic data, so that we require a log transformation if the distribution is not skewed. Lindsey and Jones (2000) analyzed these data using the multivariate log Student t and log power-exponential distributions.

The various distributions are compared in the fourth column of Table 1. Here, the log Student t distribution is clearly superior whereas the multivariate normal and log normal distributions are by far the worst fitting, along with the Gaussian copula with inverse Gauss margins. Again, only one level of variance component is necessary, that between subjects, but not that for periods within subjects. The AR(1) is also required, with $\hat{\rho} = 0.89$ for the best model. The estimate is $\hat{\kappa} = 2.5$ for the log Student t d.f. ($\hat{\kappa} = -0.30$ for the skew log Laplace), pointing to very thick tails. However, the results change rather drastically when the metabolite is modelled jointly with flosequinan and the available covariates are introduced; see Lindsey (2001, pp. 113–131).

3.4 Pharmacodynamics cross-over trial

To evaluate the dose of an H_2 receptor antagonist required in the treatment of gastric pH, six patients were given five different doses plus placebo in varying orders, following a cross-over design (Ekholm *et al.*, 1989). The dose levels were 10, 20, 40 80, and 160 mg. Each new dose administration was begun after a washout period of one week to ensure that the previous dose had been completely eliminated. Gastric pH level was measured immediately before administering the drug, and at 2, 4, 6, 8, 10, and 12 hours after. A clinically important response would be one raising the gastric pH level over 3 for two consecutive recordings. No adverse effects were recorded at the lower doses, but at 160 mg all patients except number 3 experienced intolerable diarrhoea or abdominal cramps.

A possible nonlinear regression function that fits well is

$$\mu_t = \exp\left(\frac{kte^{-kt}}{V_d}\right)$$

where V_d is a distinct parameter value for each dose level, t is the time, and k is both the absorption and elimination rate. (This nonlinear regression function can be obtained in a similar way to the dispersion function in Equation (7), by taking the limit when the absorption and elimination rates are set to be identical in Equation (6) and then exponentiating it.) A standard analysis using this regression function is given by Lindsey (2001, pp. 142–146).

The models fitted here are compared in the fifth column of Table 1. Symmetric distributions with non-constant dispersion fit best. The dispersion parameter varies as a quadratic function of time, independently of dose. For these data, variance components and auto-regression are not required. The fit of the Gaussian copula with log logistic margins is slightly superior to that with log normal margins. The log Student t, log power-exponential, and skew log Laplace distributions may fit less well because they impose dependence among the repeated responses on an individual even when the correlation is zero.

Only the two highest dose levels meet the requirement of raising pH above 3. Because the highest level yielded adverse effects, the dose of 80 mg of the H_2 receptor antagonist would be chosen.

4 Conclusions

Our general experience in analyzing a fair number of repeated measurements data sets is that for most continuous responses, a multivariate normal distribution is not ideal. Only in the last example is it competitive with the alternatives presented here. More generally, the major exception that we have found, where the normal assumption is valid, is assay data where only measurement error is present. Similarly, polynomials in time are always an *ad hoc* solution; Sandland and McGilchrist (1979) provide a good discussion of reasons for not using them.

The distributions presented here provide considerable flexibility. One important role is to allow checking goodness of fit of the normality assumptions. However, a number of open questions remain. These include:

- A variance component in the covariance matrix for a non-normal distribution imitates a random intercept. How can the covariance matrix be structured to imitate random nonlinear parameters?
- Is it possible to fit these models dynamically as an extension of the Kalman filtering and smoothing algorithms for the multivariate normal distribution (Jones, 1993) so that long time series can easily be handled?
- How can we efficiently obtain the conditional regression equation, given previous responses or residuals, so that recursive fitted values can be calculated (and plotted)? In contrast to models based on joint multivariate distributions, those constructed as the product of conditional distributions yield such fitted values directly (for example, the algorithm mentioned in the previous point).
- How can we allow for censored data, for example the non-detectable values from the assays used to measure concentrations in pharmacokinetic data? The cost of multidimensional integration of arbitrary margins of a multivariate distribution appears to be prohibitive.

None appear to have easy answers.

Acknowledgements All of the examples were analyzed using the R software with the functions `elliptic` (for the multivariate normal, Student t, power-exponential, and skew Laplace distributions) and `gausscop` (for the Gaussian copulas) in the first author's public libraries, respectively called `growth` and `repeated`, available at `http://www.luc.ac.be/~jlindsey/rcode.html`

5 Appendix

To illustrate the ease of fitting these models, the complete R code for fitting the final two models in the first example for the data from the coronary sinus potassium trial is shown below.

```
# load the elliptic function for MVN, power-exponential, skew Laplace,
Student t
library(growth)
# load the gausscop function for Gaussian copulas
library(repeated)

# read in the data and prepare as a data object
cor <- matrix(scan(``coronary.dat'', skip=5), ncol=7, byrow=T)
reps <- rmna(response=restovec(cor, times=seq(1,13,by=2)),
    ccov=tcctomat(as.factor(c(rep(1,9), rep(2,10), rep(3,8),
rep(4,9))),
        name=''treat'', dataframe=F))
# add the log transform to the object, automatically calculating the
Jacobian
reps <- transform(reps, lcor=log(cor))

# define the nonlinear location function
mu <- ~a1+a2*(treat2+treat3+treat4)+
    a3*(1-treat2)*(a5-times)*exp(-exp(a4*(a5-times)))+
    a6*(a8-times)*(treat3*exp(-exp(a7*(a8-times)))+
        treat4*exp(-exp(a9*(a8-times))))

#fit the two models
elliptic(cor, dist=''power exponential'', pell=0.4, model=mu,
    preg=c(1.4,-0.17,0.4,14.3,0.15,-0.37,13.4,0.26,0.19),
    varfn=~times, pvar=c(-8.3,0.04), par=0.8, pre=0.1,
    transform=''log'', envir=reps)

gausscop(lcor, dist=''gamma'', mu=mu,
    pmu=c(0.3,-0.07,0.24,14.9,0.12,-0.2,13.5,0.26,0.18),
    shape=~times, pshape=c(4.7,-0.026), par=0.9, envir=reps)

# lcor could have been used in the first model instead of
trans=''log''
```

Developing an appropriate nonlinear regression function and finding suitable initial parameter estimates can be much more difficult!

References

[1] Box, G.E.P. (1950) Problems in the analysis of growth and wear curves. *Biometrics* **6**, 362–389.

[2] Ciminera, J.L. and Wolfe, E.K. (1953) An example of the use of extended cross-over designs in the comparison of NPH insulin mixtures. *Biometrics* **9**, 431–446.

[3] Davis, C.S. (2002) *Statistical Methods for the Analysis of Repeated Measurements.* Berlin: Springer-Verlag.

[4] Ekholm, B.P., Fox, T.L., and Bolognese, J.A. (1989) Dose-response: relating doses and plasma levels to efficacy and adverse experiences. In Berry, D.A. *Statistical Methodology in the Pharmaceutical Sciences.* Basel: Marcel Dekker, pp. 117–138.

[5] Elston, R.C. (1964) On estimating time-response curves. *Biometrics* **20**, 643–647.

[6] Elston, R.C. and Grizzle, J.F. (1962) Estimation of time response curves and their confidence bands. *Biometrics* **18**, 148–159.

[7] Gómez, E., Gómez-Villegas, M.A., and Marin, J.M. (1998) A multivariate generalization of the power exponential family of distributions. *Communications in Statistics* **A27**, 589–600.

[8] Grizzle, J.E. and Allen, D.M. (1969) Analysis of growth and dose response curves. *Biometrics* **25**, 357–382.

[9] Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314.

[10] Jones, R.H. (1993) *Longitudinal Data Analysis with Serial Correlation: A State-space Approach.* London: Chapman & Hall.

[11] Kotz, S., Kozubowski, T.J., and Podgórski, K. (2001) *The Laplace Distribution and Generalizations. A Revisit with Applications to Communications, Economics, Engineering, and Finance.* Basel: Birkhäuser.

[12] Kotz, S. and Nadarajah, S. (2001) Letter to the editor. *Communications in Statistics* **A30**, 987–992.

[13] Lindsey, J.K. (1999a, 2nd edn.) *Models for Repeated Measurements.* Oxford: Oxford University Press.

[14] Lindsey, J.K. (1999b) Multivariate elliptically-contoured distributions for repeated measurements. *Biometrics* **56**, 1277–1280.

[15] Lindsey, J.K. (2001) *Nonlinear Models in Medical Statistics*. Oxford: Oxford University Press.

[16] Lindsey, J.K. and Jones, B. (2000) Modelling pharmacokinetic data using heavy-tailed multivariate distributions. *Journal of Biopharmaceutical Statistics* **10**, 369–381.

[17] Sandland, R.L. and McGilchrist, C.A. (1979) Stochastic growth curve analysis. *Biometrics* **35**, 255–271.

[18] Song, P.X.K. (2000) Multivariate dispersion models generated from Gaussian copulas. *Scandinavian Journal of Statistics* **27**, 305–320.

[19] Taylor, J.M.G. and Law, N. (1998) Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts? *Statistics in Medicine* **17**, 2381–2394.

Table 1: AICs of the models fitted to the four examples. Only values in the same column are comparable.

Example	1	2	3	4
MVN	130.2	1805.4	2437.4	272.2
Log MVN	117.0	—	2112.7	257.5
Student t	122.7	1803.6	—	—
Log Student t	115.0	—	2013.6	259.2
Power-exponential	121.9	1803.6	—	—
Log power-exponential	114.5	—	2024.3	258.2
Skew Laplace	119.6	1814.2	2167.9	—
Skew log Laplace	117.7	—	2016.9	265.0
Gaussian copulas				
Logistic	126.4	1806.6	—	—
Log logistic	118.5	—	2041.6	257.2
Cauchy	141.3	1834.4	—	—
Log Cauchy	140.4	—	2025.1	290.4
Laplace	126.8	1820.5	—	—
Log Laplace	124.5	—	2039.2	258.3
Gamma	118.7	1899.7	2056.1	261.1
Log gamma	114.1	—	—	—
Weibull	147.3	1865.5	2031.9	277.3
Log Weibull	129.8	—	—	—
Inverse Gauss	121.3	2017.1	2256.3	274.7
Log inverse Gauss	119.6	—	—	—