BIOMETRICS 55, 1277–1280
December 1999

Multivariate Elliptically Contoured Distributions for Repeated Measurements

J. K. Lindsey

Biostatistics, Limburgs Universitair Centrum, Diepenbeek, Belgium
email: jlindsey@luc.ac.be

SUMMARY. The multivariate power exponential distribution, a member of the multivariate elliptically contoured family, provides a useful generalization of the multivariate normal distribution for the modeling of repeated measurements. Both light and heavy tailed distributions are included. The covariance matrix retains its interpretation so that it can easily be structured for serial dependence and several levels of variance components. A crossover trial on insulin applied to rabbits, with a series of repeated measurements within each period, is analyzed by means of this distribution using autocorrelation and two levels of variance components.

KEY WORDS: Akaike information criterion; Autoregression; Crossover trial; Multivariate elliptically contoured distribution; Multivariate normal distribution; Nonconstant variance; Nonlinear regression; Power exponential distribution; Variance components.

1. Introduction

The analysis of repeated measurements has traditionally been handicapped by the lack of choice among realistic multivariate models to describe data-generating mechanisms. For continuously distributed data, statisticians have relied on the multivariate-normal distribution using software such as BMDP5V, SAS Proc Mixed, or S-Plus/R `lme`. However, in many empirical contexts, the method of analysis often has no other justification than the availability of software. The major advantage of the normal models is that serial correlation and heterogeneity among individuals (variance components or random effects) can easily be modeled simultaneously, simply by appropriately structuring the covariance matrix.

The choice among models has recently been increasing, especially with the development of Kalman filtering techniques for non-normal data (Lambert, 1996). However, one drawback of such an approach is that it does not seem easy to combine serial dependence and heterogeneity among individuals in the same model.

Gómez, Gómez-Villegas, and Marin (1998) have recently introduced a multivariate generalization of the power exponential distribution that can help remedy this situation. This is a subfamily of the elliptically contoured distributions and includes the multivariate normal distribution as a special case. The importance of this distribution lies in that it only requires two simple modifications to the multivariate normal distribution that can be easily programmed: the quadratic form in the exponential is raised to a power, and the additional normalizing constant only contains this power and the number of observations. The covariance matrix retains its interpretation (at least up to a constant).

Thus, this distribution has several attractive features:

(1) it covers both thick- and thin-tailed distributions, so that models can be checked for robustness;
(2) the location parameter is the mean, so that linear and nonlinear models can easily be constructed;
(3) the covariance matrix can be structured in appropriate ways for uniform and/or serial dependence, retaining its correlation interpretation.

After a brief description of the essential features of the model, I shall illustrate its use by applying it to a crossover trial of insulin tested on rabbits with a series of repeated measurements within each period.

2. Multivariate Power Exponential Distributions

The univariate power exponential distribution is given by

$$f(y;\mu,\sigma,\beta) = \frac{1}{\sigma\Gamma\left(1+\frac{1}{2\beta}\right)2^{1+\frac{1}{2\beta}}}\exp\left[-\frac{1}{2}\left|\frac{y-\mu}{\sigma}\right|^{2\beta}\right],$$
$$-\infty<\mu<\infty,\ 0<\sigma,\ 0<\beta\leq\infty. \quad (1)$$

Its multivariate extension is

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

where the mean and variance are

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu},$$

$$\text{var}(\mathbf{Y}) = \frac{2^{\frac{1}{\beta}}\Gamma\left(\frac{n+2}{2\beta}\right)}{n\Gamma\left(\frac{n}{2\beta}\right)}\boldsymbol{\Sigma},$$

and β determines kurtosis (Gómez et al., 1998). Thus, the correlation structure can be obtained directly from $\boldsymbol{\Sigma}$ in the usual way. However, when $\beta \neq 1$, the even cumulants are nonzero, in contrast to the multivariate normal distribution.

When $\beta = 1$, we have a multivariate normal distribution; when $\beta = 1/2$, a form of multivariate Laplace (double exponential) distribution; and when $\beta \to \infty$, a multivariate uniform distribution. Thus, for $\beta < 1$, the distribution has heavier tails than the multivariate normal distribution and can be useful in providing robustness against "outliers."

The marginal and conditional distributions are more complex elliptically contoured distributions, and not of the power exponential type.

In constructing multivariate normal models for repeated measurements, the covariance matrix can be written simultaneously as one large matrix for all observations on all individuals by setting the appropriate elements to zero so that observations on different individuals are independent. In other words, this is equivalent to taking a separate multivariate normal distribution for each individual with a suitable covariance structure, and multiplying them. This is no longer true once $\beta \neq 1$. Thus, for example, when $\boldsymbol{\Sigma}$ is diagonal in equation (2) so that the correlation among observations is zero, this distribution cannot be written as a product of independent univariate distributions from equation (1), unless $\beta = 1$. For $\beta \neq 1$, the multivariate distribution retains a dependence structure among the observations on an individual even though the correlation among them may be zero. Of course, correlations among observations on different individuals can be set to zero, as is usual in this type of study, by multiplying the multivariate power exponential distributions for the observations on each subject.

I have a function in R (Ihaka and Gentleman, 1996) for fitting this distribution to repeated measurements in continuous time (i.e., unequally spaced), with the possibility of nonlinear regression equations both for the mean and for the variance, a first-order autoregression, and two levels of variance components. (Modification of my existing function for the multivariate normal distribution took me a few minutes.) Parameter estimates are obtained by building up the complete likelihood function (in a dynamically loaded Fortran program) and maximizing it with a nonlinear optimizer (Dennis and Schnabel, 1983) built into R that calculates derivatives numerically. I have not encountered any problems of numerical instability in the many data sets to which I have applied this model. The function, called `elliptic`, used to fit models in this family is available in the public R library, called `growth`, on CRAN (ftp.stat.math.ethz.ch).

This ease of programming and use contrasts with generalized linear mixed models that require elaborate methods or approximations to perform numerical integration. Although serial dependence can be included, this is through direct conditioning on the previous response, not through autocorrelated residuals as here and in standard normal autoregression models. Such direct dependence is much more difficult to interpret. Nevertheless, generalized linear mixed models will describe data-generating mechanisms different from the family presented here so that the two are complementary. In contrast, generalized estimating equations may also be considered as a competitor, but they provide no model upon which likelihood inferences can be based and hence no insight into the data-generating mechanism; many other difficulties are also associated with them (Lindsey and Lambert, 1998). Both approaches are based on linear regression, whereas the family presented here can have both the mean and the variance depending nonlinearly on covariates.

In the following example, the inference criterion for comparing the models under consideration will be their ability to predict the observed data, i.e., how probable they make the data. In other words, they will be compared directly through the minimized-log likelihood. Because the number of parameters in the models differ, this will be penalized by adding the number of estimated parameters, a form of the Akaike information criterion (AIC; see Akaike, 1973), although the interpretation here is not asymptotic. Smaller values indicate relatively more preferable models.

3. Example

In the analysis of crossover data, modeling of the dependence structure is especially critical because treatments are compared within subjects, so that ignoring or poorly modeling the dependence will generally lead to the significance of treatment effects being underestimated.

Ciminera and Wolfe (1953) give data on the comparison of neutral protamine Hagedorn insulin mixtures. Two mixtures of insulin, the standard (A) and one containing 5% less protamine (B), were tested on rabbits in a crossover design. Two groups of 11 female rabbits were injected with the insulin at weekly intervals in the orders ABAB (sequence 1) and BABA (sequence 2). For each treatment, the blood sugar level was measured on injection and at four equally spaced postinjection times over 6 hours. The blood sugar level fell and then rose close to the pretreatment level. A glance at the individual response profiles in Figure 1 reveals what appear to be several outliers, especially at midtime (3 hours), when several responses rose and then fell.

I shall allow for two levels of variance components, the rabbit and the period nested within the rabbit, and for a first-order autocorrelation over 6 hours. The explanatory variables

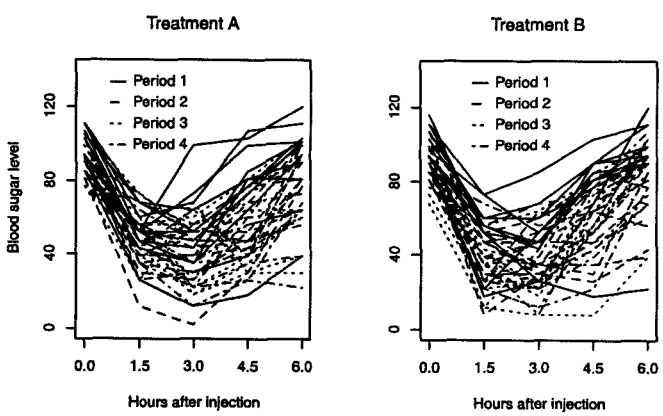

Figure 1. Plots of blood sugar level versus time (in hours) for the two treatments of rabbits.

are the period, the treatment, the group, and the quadratic time effects. Because there are only two treatments that alternate, the only sequence effect comes from the treatment that is administered in the first period; this is covered by the group effect.

Consider first the classical multivariate normal model. I shall begin with a location model having a quadratic polynomial in time, with differences in level for each treatment, period, and group, but without interactions among them. This model, with eight location parameters, has an AIC of 1899.0 under independence. Adding second-order interactions between time and these variables reduces the AIC to 1894.8. This is the most complex linear regression fitted, and it has the form

$$(\texttt{time} + \texttt{time}^2) * (\texttt{group} + \texttt{treat} + \texttt{period}).$$

However, it can be simplified by contrasting the fourth period with the other three and using only the linear interaction of time with this new-period effect and with treatment, yielding an AIC of 1891.7. Adding two levels of variance components and a first-order autoregression lowers the AIC to 1853.8 without changing the variables required in the model. However, only the rabbit-level variance component, and not the period-level one, is required, with an AIC of 1852.8. Letting the log variance depend linearly on time lowers the AIC to 1838.2. The elimination of treatment and its interaction with time further lowers the AIC to 1837.8, which yields the final model and from it the conclusion that treatment appears to have no effect. In contrast, complete elimination of the period effect raises the AIC to 1880.4 and of group to 1846.6. Thus, the period effect is by far the most significant. The parameter estimates, including treatment effect, are given in Table 1.

Without the rabbit-level variance component, the AIC increases to 1843.9, and without the autocorrelation, it increases to 1844.4. The log variance is estimated to be 4.74 with a slope (over time) of 0.24, the autocorrelation is estimated to be 0.37, and the rabbit-level variance component is estimated to be 32.5, so that the intrarabbit correlation varies from 0.22 at time 0 to 0.064 after 6 hours.

Now consider adding the parameter of the multivariate power exponential distribution to this final model (without treatment). The AIC reduces to 1833.7 with $\hat{\beta} = 0.40$, strongly indicating non-normality. Adding the second-level variance component raises the AIC to 1834.7, whereas removing both levels of variance components raises it to 1838.1, and removing the autoregression raises it to 1840.0.

Table 1
Parameter estimates for the linear regression from the two distributions

	Normal	Elliptical
Intercept	37.17	39.29
Time	−2.22	−1.99
Time^2	5.08	4.86
Group	−7.35	−10.43
Group × time	1.85	1.68
Group × time^2	1.23	1.59
Treat	0.79	1.42
Treat × time	1.02	1.23
Period 4	15.52	15.70
Period 4 × time	1.56	1.57

Table 2
AICs for various models. AR: first-order autoregression; VC: variance component; CVT: changing variance with time. All models have a quadratic polynomial in time, along with period, treatment, and group parameters and their interactions with time, as in Table 1.

	Normal	Elliptical
AR + 2 VC	1853.8	1851.1
AR + 1 VC	1852.8	1848.9
AR + 1 VC + CVT	1838.2	1833.3
AR + CVT	1843.9	1838.1
1 VC + CVT	1844.4	1840.0

These results are all similar to those for the multivariate normal distribution. However, adding the treatment effect and a linear interaction of this with time lowers the AIC to 1833.3 with $\hat{\beta} = 0.38$. Thus, the multivariate power exponential distribution not only yields a better fitting model, but it also provides some indication that there may be a treatment effect; the multivariate normal distribution does not, however, show a treatment effect. These results are summarized in Tables 1 and 2. (Standard errors are not shown because it is doubtful if they could provide a meaningful quadratic approximation to the log likelihood in such a thick-tailed distribution.)

The intrarabbit correlation is now estimated to be 0.20 at time 0, decreasing to 0.058 after 6 hours, and the autocorrelation is estimated to be 0.34. Both are fairly similar to those from the multivariate normal distribution. We observe both serial dependence and heterogeneity among rabbits in these data; however, heterogeneity among periods is not observed.

If some or all of the supposed outliers at the mid time measurement are eliminated and the analysis redone, the results vary considerably depending on which values are removed. Thus, the data are sensitive to this type of operation.

4. Discussion

The above example was chosen to illustrate the ability to model autoregression and several levels of variance components outside the multivariate normal distribution because this has been one of the major handicaps in the analysis of repeated measurements. However, it does not show all the possibilities of using multivariate power exponential distribution to model repeated measurements. Among others, these include

(1) observations unequally spaced in time;
(2) nonlinear regression for the mean;
(3) variance depending nonlinearly on covariates, as for example in pharmacokinetic models;
(4) transformation of the observations, for instance, by taking logarithms to obtain skewed distributions to provide, for example, a generalization of the multivariate log–normal distribution.

A major handicap of this distribution is that, when $\beta \neq 1$, equation (2) does not reduce to independence for any specific set of values of the parameters, even when the correlations are

all zero. Note that this is only a lack of independence among observations on an individual; observations on different subjects (or clusters) are independent.

The fact that the multivariate normal distribution is rejected in favor of a more heavily tailed distribution for these data does not imply that this is the most appropriate distribution for them. Further comparisons, for example, with results from models based on the Kalman filtering type of approach, from generalized linear mixed models, or from a multivariate t-distribution, would be required. Nevertheless, it provides a robust extension of the classical multivariate normal distribution that should become an additional indispensable tool for the applied statistician who wishes to construct scientifically reasonable models of the data-generating mechanism for the phenomenon under study.

Acknowledgements

I would like to thank Robert Gentleman and Ross Ihaka for developing the R software, a fast S-Plus clone freely available under the GNU licence.

Résumé

La distribution multivariée exponentielle, qui apparient à la famille des distributions à contours elliptiques, fournit une généralisation de la distribution normale multivariée adaptée à la modélisation des données répétées. On trouve dans la famille des distributions avec des queues plus ou moins importantes. La matrice de covariance y garde son interprétation usuelle et peut donc être structurée en fonction des liens connus entre les observations, ou en fonction d'une structure hiérarchique des variances. Avec cette distribution, pour analyser un essai en cross-over d'insuline administrée à des lapins où les données sont répétées à chaque période, on élabore un modèle avec auto corrélation et deux niveaux de composante de la variance.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Inference Theory*, B. N. Petrov and F. Csàki (eds), 267–281. Budapest: Akadémiai Kiadó.

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.

Dennis, J. E. and Schnabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. New York: Prentice Hall.

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.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational Graphics and Statistics* **5**, 299–314.

Lambert, P. (1996). Modelling irregularly sampled profiles of non-negative dog triglyceride responses under different distributional assumptions. *Statistics in Medicine* **15**, 1695–1708.

Lindsey, J. K. and Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* **17**, 447–469.

Received April 1998. *Revised October* 1998.
Accepted January 1999.