BIOMETRICS 48, 665–677
September 1992

Regression Models for an Angular Response

N. I. Fisher and A. J. Lee

CSIRO Division of Mathematics and Statistics, Lindfield, N.S.W., Australia

and

Department of Mathematics and Statistics, University of Auckland,
Auckland, New Zealand

SUMMARY

This paper considers the problem of regressing an angular response variate on a set of linear explanatory variables. A general class of models is proposed in which the mean direction and dispersion of a von Mises variate are related to the explanatory variables by general link functions. Appropriate regression diagnostics, estimation and testing procedures are developed for fitting the models. The meaning of "correlation" between an angular and a linear variable is clarified, and leads to a general notion of multiple correlation associated with the regression model. The methods are applied to a set of data arising from a study of movements of intertidal gastropods.

1. Introduction

The theory of regression models when the response variable is angular is a somewhat neglected area of statistics, although problems of this sort are not uncommon in a number of areas of application, particularly in biology, geology, and meteorology. Examples include the dependence of the direction an animal moves on the distance moved (as studied later in this paper), the dependence of the strike of a fault plane on displacement, and the dependence of wind direction on wind speed.

Such work as exists relates largely to the study of potentially interesting families of joint distributions of an angular variate Θ and a continuous linear variate X. Gould (1969), Laycock (1975), and Mardia (1972, pp. 127–128, 167) have considered various forms of the so-called "barber's pole" model, in which the mean response of Θ, conditional on $X = x$, is a curve winding in an infinite number of spirals up the surface of an infinite cylinder. Johnson and Wehrly (1978) proposed a different class of models in which the response completes just a single spiral as x increases through its range.

In Section 2 of this paper we review this work and in Section 3 propose some extensions to the Johnson and Wehrly models, and introduce a multiple correlation coefficient that measures the type of association described by one of the models. This section also sheds light on the curious fact that the "correlation" between an angular and a linear variable should be defined in the context of the nature of the regression model being contemplated—Θ regressed on X or X regressed on Θ—with quite different types of correlation coefficient being appropriate to the two cases. Finally, in Section 4 we discuss the application of the methods to a set of data arising from a study of movements of intertidal gastropods.

Key words: Angular-linear correlation; Angular response; Iteratively reweighted least squares; Maximum likelihood; von Mises distribution.

2. Some Models for Angular Regression

There are a number of ways in which we may want to model Θ in terms of several explanatory variables. For example, we may wish to:

A: Model the mean direction of Θ in terms of a vector of covariates $\mathbf{x} = (x_1, \ldots, x_k)'$.

B: Model the dispersion of Θ in terms of $x_1, \ldots, x_k$.

C: Model both the mean direction and the dispersion simultaneously, in terms of $x_1, \ldots, x_k$.

A peculiarity of regression models with a circular response variable is that a special null model is available, namely the uniform distribution, for which there is no mean direction. This is one reason for considering regression models based on dispersion as well as on mean direction; another is the fact that many data sets encountered in practice (e.g., wind direction and wind speed measurements) exhibit features such as increased variability of Θ for small values of X.

Modelling dispersion for an angular deviate presents certain difficulties, because there is no natural measure of scale for circular distributions. For this reason, it is convenient to work with the von Mises family of distributions, which has an in-built measure of dispersion, and shares many of the key properties for statistical inference that the normal distribution has for linear data. The density of a von Mises VM(μ, κ) distribution is given by

$$f(\theta;\, \mu, \kappa) = [2\pi I_0(\kappa)]^{-1}\exp[\kappa \cos(\theta - \mu)], \qquad -\pi < \theta, \mu \leqslant \pi, \quad \kappa \geqslant 0,$$

where μ is the mean direction and κ the concentration parameter. The value $\kappa = 0$ corresponds to the circular uniform distribution; for $\kappa > 0$, the density is unimodal and symmetrical about μ, and increasingly concentrated as κ increases. For $\kappa \geqslant 2$ the density at the antipode $\mu + \pi$ is effectively zero, and f is well approximated by a normal distribution with variance $1/\kappa$. In the sequel, it will be assumed that Θ is a von Mises variate.

A regression model of type A was proposed by Gould (1969), who considered the structure

$$\mu = \mu_0 + \sum \beta_j x_j$$

for the mean direction, resulting in a "barber's pole" form. Gould gave an iterative method for calculating the maximum likelihood estimates (MLEs) of the model parameters, and provided some approximate methods of inference for the case when κ can be considered large. Laycock (1975) also discussed this model, and noted that maximum likelihood (ML) is equivalent to least squares. Johnson and Wehrly (1978) pointed out that the likelihood function in the Gould model has infinitely many equally high peaks, leading to ambiguously defined MLEs. As an alternative, for a single explanatory variable, they suggested an approach via a specific model for the joint distribution of Θ and a linear variate X with a completely specified marginal distribution function $F(x)$. Their conditional distribution of Θ given $X = x$ is VM($\mu + 2\pi F(x)$, κ), a model which allows direct estimation of μ and κ. We will refer to this as the Johnson–Wehrly type A model.

For type B models with a single explanatory variable, they suggested modelling the conditional distribution of Θ given $X = x$ as VM(μ, κx), which also allows the direct estimation of the parameters, and also includes the null case of zero concentration. In the next section we consider extensions of these Johnson–Wehrly models for cases A and B, which we then combine to give a model of type C.

3. Extending the Johnson–Wehrly Models

3.1 *The Basic Models*

The Johnson–Wehrly models of Section 2 may be extended by assuming that we have angular observations $\Theta_1, \ldots, \Theta_n$, which follow von Mises distributions with mean directions $\mu_1, \ldots, \mu_n$ and concentration parameters $\kappa_1, \ldots, \kappa_n$. Our first model, which generalizes the Johnson–Wehrly type A model, assumes that all the concentration parameters are equal to κ say, and that the μ's are related to the covariates by means of a link function g so that

$$\mu_i = \mu + g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i),$$

where $\mathbf{x}_i$ is the vector of explanatory variables for the ith case and $\boldsymbol{\beta}$ is a k-vector of regression coefficients. Parallels with the theory of generalized linear models are obvious. The purpose of the function g is to map the real line to the circle, and we shall consider only monotone link functions having the property that as x ranges from $-\infty$ to ∞, $g(x)$ ranges from $-\pi$ to π, thus avoiding the problem of nonidentifiability of MLEs that afflicts the Gould model. In order that μ have an interpretation as an origin, we will also assume that $g(0) = 0$. In the sequel, we refer to such links as "angular-monotone" or briefly, "A-monotone."

In the case of a single explanatory variable X, the regression function $\theta(x) = \mathrm{E}(\Theta \mid x)$ (i.e., the conditional expectation of Θ given x) can be regarded as a curve on the surface of an infinite cylinder, which rotates once around the cylinder as x varies from $-\infty$ to ∞. Other forms of regression function are of course possible—for example, functions that rotate twice, three times, . . . , around the cylinder. The Gould model corresponds to a "barber's pole" regression function that spirals infinitely many times around the cylinder. We do not consider such possibilities further in the present paper.

The link function g can be assumed known, or to be one of a parametric family of suitable links, in a manner similar to the situation in the theory of generalized linear models. One possibility is to use the function

$$g(x) = 2 \tan^{-1}(\operatorname{sgn}(x) \mid x \mid^{\lambda}),$$

where $\lambda = 0$ corresponds to a log transformation. The parameter λ can then be estimated from the data, analogously to the estimation of Box–Cox transformations. Another possibility is to regard the covariates as lying in a bounded region, which after scaling we may take to be $[0, 1]^k$. The μ_i's may then be modelled by

$$\mu_i = \mu \pm 2\pi g(\mathbf{x}_i),$$

where g is now a member of some flexible parametric family of k-dimensional distribution functions concentrated on $[0, 1]^k$. For example, in the case of a single covariate, we could take $g(x) = I_x(\alpha, \beta)$, the incomplete beta function

$$I_x(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^x t^{\alpha-1}(1 - t)^{\beta-1}\, dt.$$

To extend the Johnson and Wehrly type B model, we can imagine that the mean directions μ_i are all equal to μ say, and the κ's are modelled by $\kappa = h(\mathbf{x})$, where h is some function mapping $\mathscr{R}_k$ to $[0, \infty)$. One possibility is

$$h(\mathbf{x}) = A \exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}),$$

where $\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x} = \gamma_1 x_1 + \cdots + \gamma_k x_k$. The null case of no concentration then corresponds to $A = 0$. Alternatively, the constant A can be absorbed into the exponential, and we can consider the function

$$h(\mathbf{x}) = \exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}),$$

where now $\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x} = \alpha + \gamma_1 x_1 + \cdots + \gamma_k x_k$. This eliminates the need for a separate type of parameter A, but no longer includes the null case, except in the form $\alpha = -\infty$. However, we can perform a preliminary test of uniformity (see, e.g., Mardia, 1972, p. 173) and proceed to the fitting of the model only if the uniformity test is rejected.

To conclude this section, we present a few simulated data sets to indicate how data following one of these models might appear in practice. Figure 1 shows four synthetic

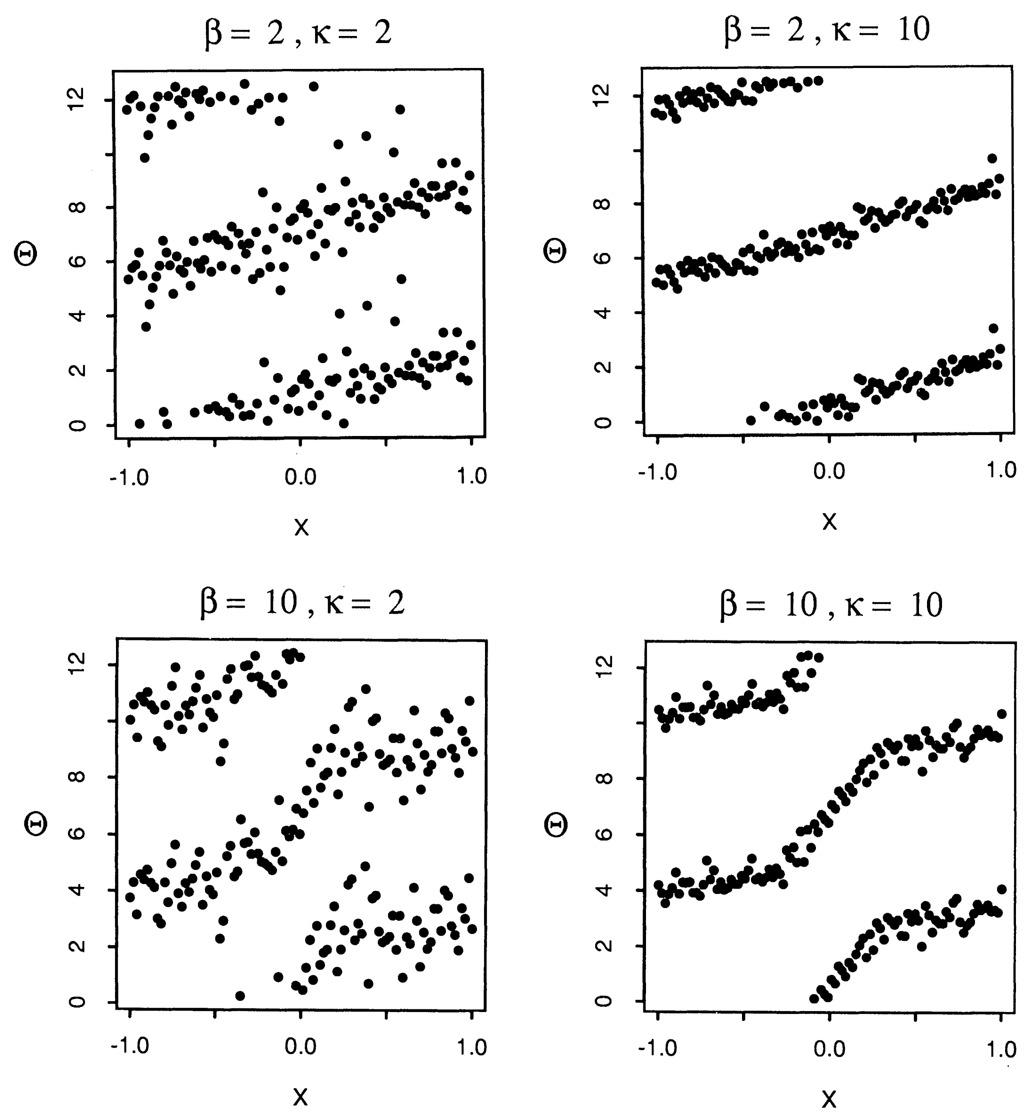

Figure 1. Four simulated data sets, showing typical behaviour of realizations of A-monotone regression of Θ on X, when the conditional distribution of Θ_i given $X_i = x_i$ is von Mises $\mathrm{VM}(\mu_i, \kappa)$ with $\mu_i = \mu + g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)$ and link function $g(x) = 2\tan^{-1}(x)$ and the (one-dimensional) $\mathbf{X}$'s equally spaced on $(-1, 1)$. Each data set is of size 100.

samples, each of size 100 simulated from the model

$$\Theta_i \sim \mathrm{VM}(\mu_i, \kappa),$$

where

$$\mu_i = \mu + g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$$

with link function

$$g(x) = 2 \tan^{-1}(x)$$

and the (one-dimensional) $\mathbf{x}$'s equally spaced on $(-1, 1)$.

The method of display has been to plot the 100 points (x, Θ) in Cartesian coordinates, and then to plot the additional points $(x, \Theta + 2\pi)$ so that the relationship (if any) is fully displayed.

3.2 *Inference for the Mean Model*

We first consider in more detail the model where the Θ_i's are independently $\mathrm{VM}(\mu_i, \kappa)$, where $\mu_i = \mu + g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$ and g is assumed known. The extension to the case where g is a member of a parametric family is in principle straightforward but will not be considered explicitly. The log likelihood is

$$-n \log I_0(\kappa) + \kappa \sum_{i=1}^{n} \cos(\Theta_i - \mu - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)),$$

where $I_p(\kappa)$ is the modified Bessel function of the first kind and order p. Defining

$$\begin{aligned}
u_i &= \sin(\Theta_i - \mu - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)),\\
\mathbf{u}^{\mathrm{T}} &= (u_1, \ldots, u_n),\\
\mathbf{X} &= \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix},\\
\mathbf{G} &= \mathrm{diag}(g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_1), \ldots, g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_n)),\\
S &= \sum_{i=1}^{n} \sin(\Theta_i - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i))/n,\\
C &= \sum_{i=1}^{n} \cos(\Theta_i - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i))/n,\\
R &= (S^2 + C^2)^{1/2}, \qquad (1)
\end{aligned}$$

we see that the MLEs are solutions to the equations

$$\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{u} = 0, \qquad (2)$$

$$R \sin \hat{\mu} = S, \qquad (3)$$

$$R \cos \hat{\mu} = C, \qquad (4)$$

$$A(\hat{\kappa}) = R, \qquad (5)$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. An iterative scheme for the solution of the equations (2)–(5) suggests itself. We can start with an initial value for $\hat{\beta}$, calculate S, C, and R and hence $\hat{\mu}$

and $\hat{\kappa}$ by (3)–(5). These estimates can then be used to solve (2) for an updated version $\hat{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$. The solution of (2) is easily accomplished by means of the iteratively reweighted least squares (IRLS) algorithm of Green (1984). The updating equations are

$$\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}) = \mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{y}, \tag{6}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ and $y_i = u_i/[A(\kappa)g'(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)]$. We update $\hat{\boldsymbol{\beta}}$ by solving (6) and then update $\hat{\mu}$ and $\hat{\kappa}$ by (3)–(5), which are analogous to the ML equations for an independent, identically distributed (i.i.d.) von Mises sample. This cycle is repeated until convergence.

From standard likelihood theory, the asymptotic standard errors of $\boldsymbol{\beta}$ are obtainable from

$$\operatorname{var} \hat{\boldsymbol{\beta}} = \frac{1}{\kappa A(\kappa)}\left\{(\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X})^{-1} + \frac{(\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{g}\mathbf{g}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X})^{-1}}{(n - \mathbf{g}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{g})}\right\}, \tag{7}$$

where $\mathbf{g}$ is a vector whose elements are the diagonal elements of $\mathbf{G}$. The asymptotic variance of $\hat{\kappa}$ and the circular variance of $\hat{\mu}$ are given respectively by

$$\operatorname{var}(\hat{\kappa}) = 1/(nA'(\kappa)) \tag{8}$$

and

$$\text{circ. var}(\hat{\mu}) = [2(n - k)\kappa A(\kappa)]^{-1}$$

whence, from Fisher and Lewis (1983, Example 1), the estimated circular standard error of $\hat{\mu}$ is

$$\hat{\sigma}_{\hat{\mu}} = [(n - k)\hat{\kappa}A(\hat{\kappa})]^{-1/2},$$

from which a large-sample confidence interval for μ can be calculated.

In the case of a single predictor, (7) reduces to

$$\operatorname{var} \hat{\beta} = \frac{1}{\kappa A(\kappa) \sum_{i=1}^{n} (v_i - \bar{v})^2},$$

where $v_i = g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)x_i$.

One difficulty exists in the maximisation of this log likelihood. If $\boldsymbol{\beta} = \mathbf{0}$, we have a null model where all observations cluster around a common mean direction μ. However, as the elements of $\boldsymbol{\beta}$ become infinite, use of any monotone link function also leads to a null model with mean direction $\mu \pm \pi$. In such cases, the likelihood will have a peak in the vicinity of zero, and attain a similar magnitude as the elements of $\boldsymbol{\beta}$ tend to $\pm\infty$. The estimate can arbitrarily be taken to be the location of the peak around zero in this case, even though this may be only a local maximum. Note that the optimisation process is simplified if the X's are centred at their means.

An alternative procedure, particularly useful in the case of one or two explanatory variables, is as follows. We can write

$$\sum_{i=1}^{n} \cos(\Theta_i - \mu - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)) = R\cos(\mu - \hat{\mu}), \tag{9}$$

where R is a function of $\boldsymbol{\beta}$ but not of κ or μ. From the form of the log likelihood, the value of $\boldsymbol{\beta}$ that maximises R is also the value of $\boldsymbol{\beta}$ that maximises the log likelihood. Once the maximising $\boldsymbol{\beta}$ is found, the values of $\hat{\mu}$ and $\hat{\kappa}$ can be found from (3)–(5). Inspecting the graph of R as a function of $\boldsymbol{\beta}$ will produce a good starting value for the numerical maximisation of R. Simulations show that local maxima can exist quite close to the global maximum so that graphical exploration of the likelihood surface is advisable anyway.

Asymptotic tests and confidence intervals for the parameters based on the asymptotic normality of the ML estimators can be carried out in the usual way. The behaviour of the estimators in small samples was investigated in a small simulation study, using the model $\mu_i = g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$ and the link

$$g(x) = 2 \tan^{-1}(x),$$

with a single explanatory variable. For sample sizes n of 10, 20, 50, von Mises data were generated using the algorithm of Best and Fisher (1979) for values of x equally spaced from -1 to 1, $\kappa = 1, 2, 5$, and $\beta = 0, 2, 6$. The ML estimates were computed using the iterative scheme proposed above, and their global (as opposed to local) maximality checked by evaluating the likelihood on a grid of points. The simulations showed that there are no difficulties with the estimation of μ. There is no appreciable bias in the ML estimators, and formula (9) provides a good approximation to the mean resultant length. The circular standard error of $\hat{\mu}$ can be computed using the methods of Fisher and Lewis (1983, Example 1); this point is pursued in the next subsection.

The simulations also showed that there is some degree of bias in the ML estimator of β in the case of small dispersions and small sample sizes. This bias is negligible when the sample size exceeds 20 or when κ exceeds unity, for the ranges of parameter values considered. For the same range of parameter values that involve some bias, there were some problems with convergence, and formula (7) tends to underestimate the standard errors. The likelihoods tend to be rather badly behaved for these parameter values, with local minima close to the global maximum, so that a careful choice of starting values is essential.

Finally, there may be substantial bias in the estimation of κ that persists even at large sample sizes, which should not be surprising in view of the large biases that occur in the estimation of κ from i.i.d. von Mises samples (see Best and Fisher, 1981). As in the i.i.d. case, there may well be a case for estimating κ by some resampling method such as the parametric bootstrap or jackknifing.

3.3 *Inference for the Dispersion Model*

As mentioned in Section 2, there may be occasions when it is desired to model the dispersion rather than the mean direction in terms of the covariates $\mathbf{x}$. Specifically, we may assume that the observed angles $\Theta_1, \ldots, \Theta_n$ follow von Mises distributions with common mean direction μ and dispersions κ_i given by $\kappa_i = h(\mathbf{x}_i)$, where h is a link function mapping $\mathscr{R}_k$ to $[0, \infty)$. Assuming a link of the form $h(\mathbf{x}) = h(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x})$, where $\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x} = \alpha + \gamma_1 x_1 + \cdots + \gamma_k x_k$, the log likelihood for this model is

$$-\sum_{i=1}^{n} \log I_0(\kappa_i) + \sum_{i=1}^{n} \kappa_i \cos(\Theta_i - \mu),$$

where $\kappa_i = h(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}_i)$. The ML estimates are obtained by solving the equations

$$\sum_{i=1}^{n} \{\cos(\Theta_i - \mu) - A(\kappa_i)\} h'(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}_i) x_{ij} = 0, \tag{10}$$

$$R \sin \hat{\mu} = S, \tag{11}$$

$$R \cos \hat{\mu} = C, \tag{12}$$

where for this model

$$S = \sum_{i=1}^{n} \kappa_i \sin(\Theta_i),$$

$$C = \sum_{i=1}^{n} \kappa_i \cos(\Theta_i),$$

$$R = (S^2 + C^2)^{1/2}.$$

Assuming the value of $\hat{\mu}$ is known, the equations (10) may be solved by IRLS: Given a starting value $\hat{\boldsymbol{\gamma}}$, an updated $\hat{\boldsymbol{\gamma}}^*$ is obtained from the updating equation

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}}) = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}, \tag{13}$$

where

$$y_i = \{\cos(\Theta_i - \mu) - A(\kappa_i)\}/[h'(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}_i)A'(\kappa_i)] \tag{14}$$

and $\mathbf{W}$ is a diagonal matrix with elements

$$w_i = \{h'(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}_i)\}^2 A'(\kappa_i). \tag{15}$$

Then solution of the likelihood equations proceeds as follows: We pick a suitable starting value for $\hat{\boldsymbol{\gamma}}$, compute $\hat{\mu}$ via (11) and (12) and then update $\hat{\boldsymbol{\gamma}}$ using (13). This cycle is then repeated until convergence.

The asymptotic variance of $\hat{\boldsymbol{\gamma}}$ is given by

$$\operatorname{var} \hat{\boldsymbol{\gamma}} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1} \tag{16}$$

and the asymptotic mean resultant length of $\hat{\mu}$ is

$$1 - \frac{1}{2\sum_{i=1}^{n} \kappa_i A(\kappa_i)}. \tag{17}$$

In view of the difficulties involved in estimating κ using ML estimation, it was thought prudent to run another small simulation to check the small-sample behaviour of the estimators, in the case of a single dependent variable using the link $h(x) = \exp(\alpha + \gamma x)$. The details of the simulation were identical to those for the mean model, except that ranges of values selected were $\alpha = 1, 2$ and $\gamma = -1, 0, 1$.

There were no problems with the estimation of μ, with the bias in the estimates being small and the asymptotic formula (17) giving a good approximation to the actual mean resultant length. In view of the difficulties in the ML estimation of κ in the von Mises case, we might expect the situation to be not quite as satisfactory for α and γ. However, for sample sizes of 20 or more, the bias of the estimates is not excessive and their standard errors are well approximated by the asymptotic formula (16).

We conclude this section by presenting a graphical method for diagnosing the possible dependence of dispersion on the covariate. For convenience we assume that there is just a single explanatory variable X. Recalling that the concentration parameter κ of a von Mises distribution is related to the mean resultant length ρ by

$$\rho = A(\kappa),$$

we see that, if the dispersion model is correct, the function $h^{-1}(A^{-1}(\rho))$ is linear in x. Accordingly, we need estimates of $\rho_i = A(\kappa_i)$, $i = 1, \ldots, n$. Without replication, the only precise information about ρ_i is contained in Θ_i, and even if the mean direction μ is known, it is not easy to find a stable estimate of ρ_i from just this one datum. However, a reasonably stable estimate can be obtained by rearranging the Θ's so that their corresponding x_i's are increasing, and then estimating ρ_i from the $2m + 1$ ordered directions $\Theta_{(i-m)}, \ldots, \Theta_{(i+m)}$ for some small value m (e.g., $m = 2$). (Adjustments will clearly be needed if there are a few holes in the range of x-values.) Then plot $(x_i, h^{-1}(A^{-1}(\hat{\rho}_i)))$ for appropriate values of i. An example of this diagnostic using an exponential link function is given in Section 4.

3.4 *The Mixed Model*

The combination of the models A and B into the mixed model C is straightforward. Writing $\kappa_i = h(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}_i)$, the log likelihood is now

$$-\sum_{i=1}^{n} \log I_0(\kappa_i) + \sum_{i=1}^{n} \kappa_i \cos(\Theta_i - \mu - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)) \tag{18}$$

and is maximised by combining the methods of the previous two subsections. Specifically, given starting values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ (a practical method of finding these is discussed in the next section) we maximise (18) by first updating $\boldsymbol{\beta}$ by means of a slightly modified version of (6), the matrix $\mathbf{G}^2$ being replaced by $\mathbf{GKG}$, where $\mathbf{K}$ is a diagonal matrix with elements $\kappa_i A(\kappa_i)$. Also, in the definition of the vector $\mathbf{y}$, $A(\kappa_i)$ replaces $A(\kappa)$. We then update the value of $\boldsymbol{\gamma}$ as in Section 3.3, except that (14) must be changed to

$$y_i = (\cos(\Theta_i - \mu - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})) - A(\kappa_i))/(h'(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}_i)A'(\kappa_i)).$$

The weights (15) are unchanged. These two updating steps are alternated until convergence.

The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\mu}$ are asymptotically uncorrelated with $\hat{\boldsymbol{\gamma}}$. The formula (16) is still correct, and up to $o(n^{-1})$, the asymptotic mean resultant length of $\hat{\mu}$ and the asymptotic variance of $\hat{\boldsymbol{\beta}}$ are given by (17) and

$$\operatorname{var} \hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{GKGX})^{-1},$$

respectively.

3.5 *Measuring Association Between Θ and X*

In this section we introduce a multiple correlation coefficient connected with the type A regression model, and develop a test for zero correlation. As mentioned in the introduction, several authors have derived regression models by considering bivariate distributions for Θ and X. For example, Mardia and Sutton (1978) consider a distribution on the cylinder for which the regression of X on Θ (i.e., the conditional distribution of X given Θ) is of the form

$$X = a \sin(\Theta + b) + \varepsilon, \tag{19}$$

where ε is normally distributed. Measures of the correlation between X and Θ associated with this regression are thus associated with barber's pole models. This is in marked contrast to the sort of correlation between Θ and X appropriate for the angular regression models considered in Section 3.2.

In a similar manner we can construct a joint density in the present case. Consider a joint density of the form

$$f(\mathbf{x}, \theta) = [2\pi I_0(\kappa)]^{-1}\exp(\kappa \cos(\theta - \mu_0 - g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})))f(\mathbf{x}), \tag{20}$$

where $f(\mathbf{x})$ is any density on $\mathscr{R}_k$. The conditional distribution of Θ given $\mathbf{x}$ is von Mises $\mathrm{VM}(\mu_0 + g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}), \kappa)$ and independence of X and Θ corresponds to $\beta = 0$. A test of independence for this model may thus be carried out by testing $\boldsymbol{\beta} = \mathbf{0}$. A multiple correlation coefficient between Θ and X may be defined by introducing a circular variate $\Phi_\mathbf{b}$ given by $\Phi_\mathbf{b} = g(\mathbf{b}^T\mathbf{X})$ and defining a correlation ρ between Θ and X by

$$\rho_{\text{anglin}} = \max_{\mathbf{b}} \rho_A(\mathbf{b}),$$

where $\rho_A(\mathbf{b})$ is some signed angular-angular correlation of Θ and $\Phi_\mathbf{b}$. We shall use the coefficient ρ_{FL} described in Fisher and Lee (1983). Using $\rho_A = \rho_{FL}$, it is immediate from the results of that paper that $\rho_{\text{anglin}} = 0$ when $\beta = 0$ (i.e., under independence of Θ and X). In fact, we have $\rho_{\text{anglin}} = \rho_{FL}(\boldsymbol{\beta})$ since it can be shown that $\rho_{FL}(\mathbf{b})$ attains its maximum at $\mathbf{b} = \boldsymbol{\beta}$. It can also be shown that $\rho_{\text{anglin}} = 1$ if and only if $\Theta = \mu + g(\boldsymbol{\beta}^T\mathbf{x})$, which will happen if and only if κ is infinitely large.

An estimator of ρ_{anglin} is $\hat{\rho}_{FL}(\hat{\boldsymbol{\beta}})$, where $\hat{\rho}_{FL}$ is the estimator of ρ_{FL} discussed in Fisher and Lee (1983, 1986; see latter reference for an efficient computational form), and $\hat{\boldsymbol{\beta}}$ is the ML estimator of Section 3.2. Provided that both $\hat{\rho}_{FL}(\hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\beta}}$ are jointly asymptotically normal (they both are marginally asymptotically normal), the results of Pierce (1982) show that $\hat{\rho}_{FL}(\hat{\boldsymbol{\beta}})$ is also asymptotically normal.

4. An Example of a Mixed Model

It is not uncommon in practice to observe that dispersions of directions about the mean direction tend to be larger for values of explanatory variables that are near zero, in addition to any dependence of the mean direction on these variables. Such is the case for the data in Figure 2(a), which are drawn from a series of experiments by Chapman (1986), Chapman and Underwood (1992), and Underwood and Chapman (1985, 1989, 1992) on distances moved by small blue periwinkles, *Nodilittorina unifasciata*, after they had been transplanted downshore from the height at which they normally live (see Table 1). The results of experiments at two different locations have been combined for the purposes of this example.

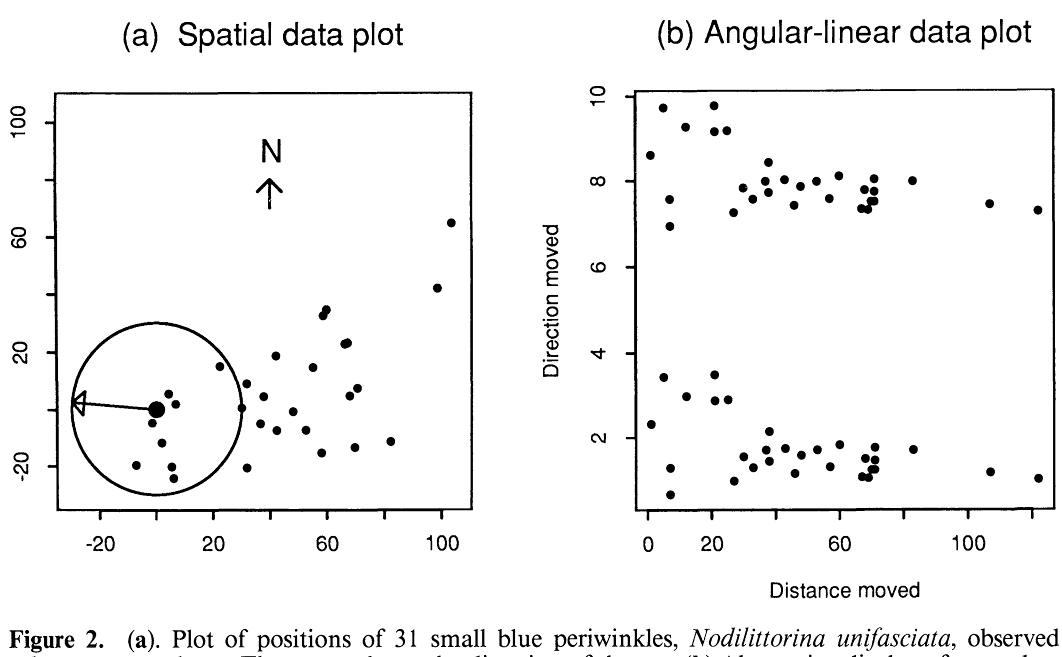

Figure 2. (**a**). Plot of positions of 31 small blue periwinkles, *Nodilittorina unifasciata*, observed subsequent to release. The arrow shows the direction of the sea. (**b**) Alternative display of same data as plot of direction against distance, with each point replicated by having 2π added to its direction as in display in Figure 1.

Table 1

Measurements of direction and distance travelled by 31 *blue periwinkles,* Nodilittorina unifasciata, *after they had been transplanted downshore from the height at which they normally live*

i	1	2	3	4	5	6	7	8	9	10	11
x_i (cm)	107	46	33	67	122	69	43	30	12	25	37
Θ_i	67	66	74	61	58	60	100	89	171	166	98

i	12	13	14	15	16	17	18	19	20	21	22
x_i	69	5	83	68	38	21	1	71	60	71	71
Θ_i	60	197	98	86	123	165	133	101	105	71	84

i	23	24	25	26	27	28	29	30	31
x_i	57	53	38	70	7	48	7	21	27
Θ_i	75	98	83	71	74	91	38	200	56

The positions shown in the figure correspond to a total of 31 animals, 15 of which were measured one day after transplantation and the other 16 four days after; plots of the separate sets indicate comparable behaviour. The arrow shows the approximate direction (275°) of the sea.

Figure 2(a), and an alternative view [Figure 2(b)] using the previously described display method of Figure 1, indicate dependence of the mean and the dispersion of direction moved on distance travelled. Further evidence is provided by the sample value of the correlation ρ_{anglin} described in Section 3.5: $\hat{\rho}_{\text{anglin}} = -.316$, which has significance probability .057 using a randomisation test for the hypothesis of no A-monotone dependence of direction on distance.

Fitting of the full mixed model is helped by initial two-stage fitting of mean direction and dispersion models. After fitting a regression model for the mean of Θ, we obtain as initial estimates

$$\hat{\beta} = -.013, \quad \hat{\mu} = 97°.$$

The function R defined in (1) is plotted as a function of β in Figure 3(a). We see that the likelihood has two maxima, one rather more pronounced than the other, illustrating the care that must be taken with starting values as discussed in the previous section. Figure 3(b) shows a plot of residuals from the fitted model, with clear evidence of the dependence of dispersion of the direction on distance travelled. Figure 3(c) is a von Mises Q–Q plot of the residuals, showing that the von Mises assumption is already quite tolerable.

Proceeding to the next stage, we consider modelling the dispersion using an exponential link function. Let $\delta_1, \ldots, \delta_n$ denote the residual deviations (unit vectors) after fitting the first stage of the model. Using the diagnostic plot described at the end of Section 3.3, we get the display in Figure 3(d). Based on this, the exponential form does not seem unreasonable. Initial estimates of the parameters α and γ of the link function can be obtained by fitting a straight line to a scatterplot of points, yielding $\hat{\alpha}_0 = -.36$, $\hat{\gamma}_0 = .04$. A proper fit of the dispersion model to the residuals then yields $\hat{\alpha} = -.0055$, $\hat{\gamma} = .034$.

Using as starting values the estimates obtained from the two-stage process, we now fit the full mixed model, to get $\hat{\mu} = 117.1$, $\hat{\sigma}_{\hat{\mu}} = .0458$ (leading to an asymptotic 95% confidence interval of half-width 5.2°), $\hat{\beta} = -.009$ with standard error (s.e.) .0025, $\hat{\alpha} = 1.78$ (s.e. .25), $\hat{\gamma} = .045$ (s.e. .009), and $\text{cov}(\hat{\alpha}, \hat{\gamma}) = -.00007$. (An alternative to the use of these standard errors would be to calculate confidence intervals based on a parametric bootstrap

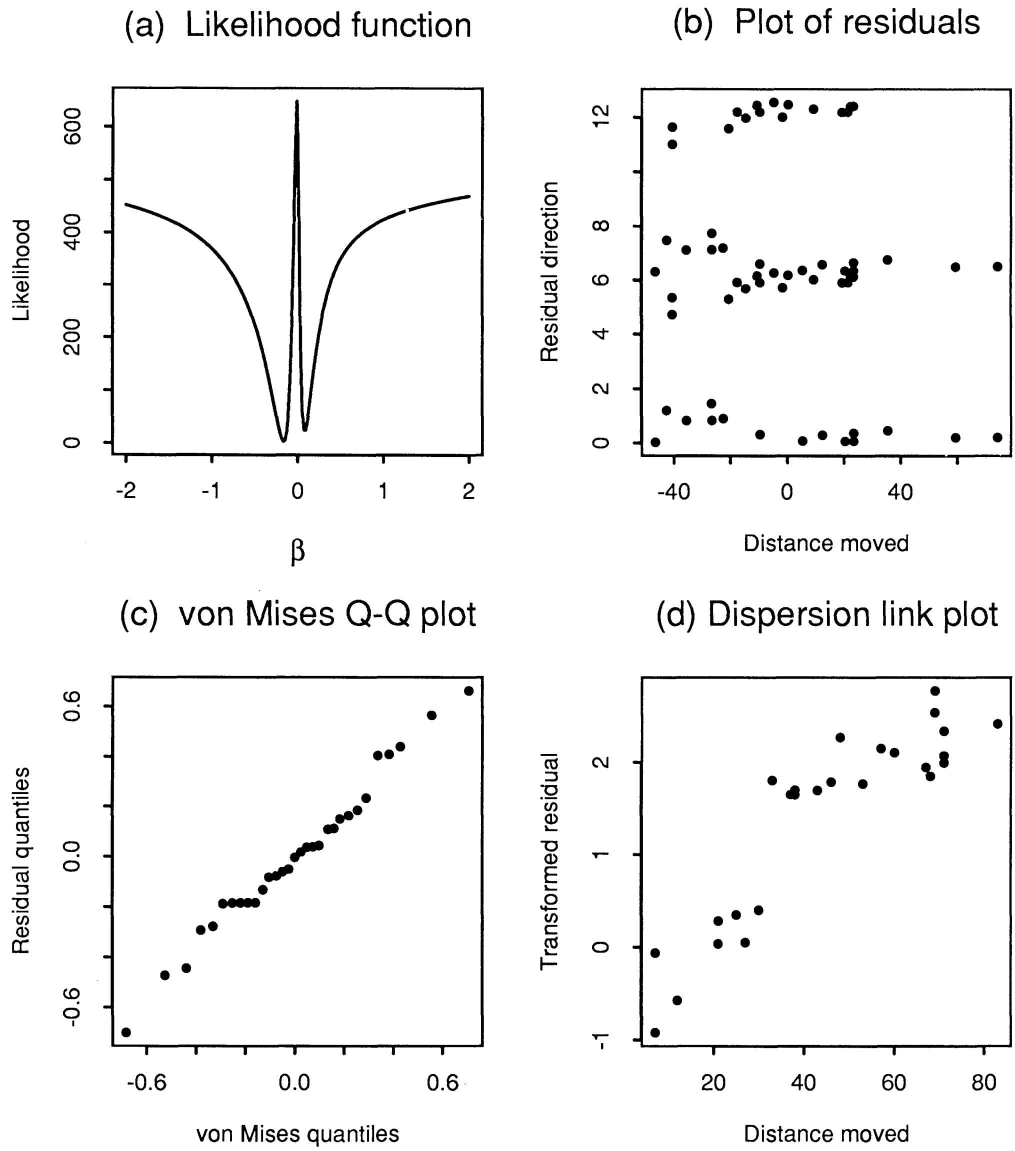

Figure 3. (**a**) Plot of likelihood function for A-monotone regression model fitted to data of Figure 2. (**b**) Residual directions from A-monotone model fitted in (**a**). (**c**) von Mises circular Q-Q plot of residual directions. (**d**) Plot to examine possibility of exponential link function to model dispersions of residual directions.

approach, resampling from the fitted von Mises distribution, but we shall not pursue this here.) On the basis of the circular Q-Q plot and these standard errors of parameter estimates, it seems reasonable to conclude that the data may be satisfactorily described by a model taking account of the dependence of both mean direction and dispersion on distance moved.

Acknowledgements

We thank Dr A. Underwood and Ms. G. Chapman, University of Sydney, for furnishing the data and associated helpful discussions, and the referees for their careful reading of an earlier version of the manuscript.

Résumé

Cet article considère le problème de la régression d'une variable aléatoire sous la forme d'une réponse angulaire sur un ensemble de variables explicatives linéaires. On propose une classe générale de modèles dans lesquels la direction moyenne et la dispersion d'une variable aléatoire de von Mises sont reliées aux variables explicatives par des fonctions de lien générales. Pour ajuster ces modèles, on développe des diagnostics appropriés de régression, des procédures d'estimation et de test. On rend plus claire la signification de la "corrélation" entre une variable angulaire et une variable linéaire, ce qui conduit à la notion générale d'une corrélation multiple associée au modèle de régression. On applique les méthodes à un ensemble de données provenant d'une étude de mouvements de gastropodes.

References

Best, D. J. and Fisher, N. I. (1979). Efficient simulation of the von Mises distribution. *Applied Statistics* **28,** 152–157.

Best, D. J. and Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises–Fisher concentration parameters. *Communications in Statistics—Computation and Simulation* **10,** 493–502.

Chapman, M. G. (1986). Assessment of some controls in experimental transplants of intertidal gastropods. *Journal of Experimental Marine Biology and Ecology* **103,** 181–201.

Chapman, M. G. and Underwood, A. J. (1992). Experimental designs for analyses of movements by molluscs. *Journal of Molluscan Studies,* in press.

Fisher, N. I. and Lee, A. J. (1983). A correlation coefficient for circular data. *Biometrika* **70,** 327–332.

Fisher, N. I. and Lee, A. J. (1986). Correlation coefficients for random variables on the sphere and the hypersphere. *Biometrika* **73,** 159–164.

Fisher, N. I. and Lewis, T. (1983). Estimating the common mean direction of several circular or spherical distributions with different dispersions. *Biometrika* **70,** 333–341.

Gould, A. L. (1969). A regression technique for angular variates. *Biometrics* **25,** 683–700.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society, Series B* **46,** 149–192.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association* **73,** 602–606.

Laycock, P. J. (1975). Optimal design: Regression models for directions. *Biometrika* **62,** 305–311.

Mardia, K. V. (1972). *Statistics of Directional Data.* New York: Academic Press.

Mardia, K. V. and Sutton, T. W. (1978). A model with cylindrical variables with applications. *Journal of the Royal Statistical Society, Series B* **40,** 229–233.

Pierce, D. A. (1982). The asymptotic effects of substituting estimators for parameters in certain types of statistics. *Annals of Statistics* **10,** 475–478.

Underwood, A. J. and Chapman, M. G. (1985). Multifactorial analysis of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology* **91,** 12–43.

Underwood, A. J. and Chapman, M. G. (1989). Experimental analyses of the influences of topography of the substratum on movements and density of an intertidal snail, *Littorina unifasciata. Journal of Experimental Marine Biology and Ecology* **134,** 175–196.

Underwood, A. J. and Chapman, M. G. (1992). Experiments on topographic influences on density and dispersion of *Littorina unifasciata* in New South Wales. *Journal of Molluscan Studies,* in press.

Received April 1990; *revised March and May* 1991; *accepted May* 1991.