Splus-functions for likelihood ratio tests in the linear mixed model

Thomas Kneib

Department of Statistics
Ludwig Maximilians-University Munich

1 Introduction

A common approach to deal with several testing problems is the likelihood ratio test. The theory of such tests and the asymptotic distribution of the test statistic are well studied under certain regularity conditions, including the iid-assumption for the given data (compare for example Cox & Hinkley (1974)). Though these regularity conditions are fulfilled in many practical situations, this is not the case for the linear mixed model, where data are correlated. Moreover testing variance components in the linear mixed model is nonstandard, since the variance parameter may be on the boundary of its parameter space under H_0.

A more general theory to deal with these problems has been developed in Crainiceanu, Ruppert & Vogelsang (2002) and Crainiceanu & Ruppert (2002). This theory will be reviewed briefly in the following section. Section 3 describes three different situations in which the proposed theory may be used. They include certain nonparametric and spatial regression models and allow us to test nonparametric functions against polynomial alternatives or to perform a test on the presence of a spatial effect.

Since the results presented in section 2 are nontrivial, in the sense that the asymptotic distribution is no well known distribution, functions to simulate the distribution and to compute different quantities that characterize it have been implemented in Splus. Section 4 contains the description of these functions.

2 Theoretical background

Consider a linear mixed model

$$y = X\beta + Zb + \varepsilon, \tag{1}$$

where β is a $(p+1)$-dimensional vector corresponding to fixed effects, b is a q-dimensional vector corresponding to random effects and ε is the usual measurement error. For ε and b we assume

$$\begin{pmatrix} \varepsilon \\ b \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \tau^2 I_q \end{pmatrix} \right),$$

with $\sigma^2 > 0$ and $\tau^2 \geq 0$ being the variances of the measurement error and the random effects respectively. Obviously, this assumption results in a restriction to models with independent and identically distributed random effects b. This includes a restriction to models with only a single variance component for the random effects.

For the marginal distribution of the response vector y, one obtains

$$\mathbb{E}(y) = X\beta \text{ and } \operatorname{Var}(y) = \sigma^2 I_n + \tau^2 ZZ' = \sigma^2 V_\lambda$$

with $V_\lambda = I_n + \lambda ZZ'$ and $\lambda = \tau^2/\sigma^2$. The parameter λ is the ratio of the variance of the random effects b and the variance of the error term ε and may therefore be interpreted as a signal to noise ratio.

The difference between the usual linear model and the linear mixed model (1) is the additional presence of random effects. Naturally the question arises, whether these additional model components are needed to describe the underlying structure in a given data set in an adequate way. This question may be analyzed through testing whether the variance component of the random effects is zero:

$$H_0 : \tau^2 = 0 \text{ vs. } H_1 : \tau^2 > 0. \tag{2}$$

Equivalently, one may test whether the signal to noise ratio λ is zero:

$$H_0 : \lambda = 0 \text{ vs. } H_1 : \lambda > 0. \tag{3}$$

The latter formulation has the advantage of eliminating the nuisance parameter σ^2 from the testing problem.

Inference in the linear mixed model may either be based on the log-likelihood

$$l(\beta, \lambda, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|V_\lambda|) - \frac{1}{2\sigma^2}(y - X\beta)' V_\lambda^{-1}(y - X\beta) \tag{4}$$

or on the restricted log-likelihood

$$\begin{aligned} l^*(\beta, \lambda, \sigma^2) &= -\frac{n-p-1}{2}\log(\sigma^2) - \frac{1}{2}\log(|V_\lambda|) - \frac{1}{2}\log(|X'V_\lambda^{-1}X|) \\ &\quad -\frac{1}{2\sigma^2}(y - X\beta)'V_\lambda^{-1}(y - X\beta). \end{aligned} \tag{5}$$

Maximization of (4) and (5) with respect to β and σ^2 gives the (restricted) maximum likelihood estimators

$$\begin{aligned} \hat{\beta}(\lambda) &= (X'V_\lambda^{-1}X)^{-1}X'V_\lambda^{-1}y, \\ \hat{\sigma}^2_{ML}(\lambda) &= \frac{1}{n}(y - X\hat{\beta}(\lambda))'V_\lambda^{-1}(y - X\hat{\beta}(\lambda)), \\ \hat{\sigma}^2_{REML}(\lambda) &= \frac{1}{n-p-1}(y - X\hat{\beta}(\lambda))'V_\lambda^{-1}(y - X\hat{\beta}(\lambda)). \end{aligned}$$

Plugging these expressions back into (4) and (5) yields the profile log likelihood

$$l(\lambda) = -\frac{1}{2}\log(|V_\lambda|) - \frac{n}{2}\log(y'R_\lambda V_\lambda^{-1}R_\lambda y)$$

and the profile restricted log likelihood

$$l^*(\lambda) = -\frac{1}{2}\log(|V_\lambda|) - \frac{1}{2}\log(|X'V_\lambda^{-1}X|) - \frac{n-p-1}{2}\log(y'R_\lambda V_\lambda^{-1}R_\lambda y)$$

with the residual matrix

$$R_\lambda = I - X(X'V_\lambda^{-1}X)^{-1}X'V_\lambda^{-1}. \tag{6}$$

Since we are interested in testing (3) through a likelihood ratio test, we define the likelihood ratio and the restricted likelihood ratio through

$$LR_n(\lambda) = 2l(\lambda) - 2l(0)$$

and

$$RLR_n(\lambda) = 2l^*(\lambda) - 2l^*(0).$$

One then obtains the likelihood ratio test statistics

$$LR_n = \sup_{\lambda \geq 0} LR_n(\lambda) = LR_n(\hat{\lambda}_{ML})$$

for the profile log likelihood and

$$RLR_n = \sup_{\lambda \geq 0} RLR_n(\lambda) = RLR_n(\hat{\lambda}_{REML})$$

for the profile restricted log likelihood.

Now, testing (3) is non-standard for two reasons: First, the signal to noise ratio λ is on the boundary of its parameter space $[0, \infty)$ under H_0, and second, the usual iid assumption for the vector of observed data y is not matched (at least under the alternative). The first problem has been dealt with by Self & Liang (1987), who proved, that under the iid assumption, the asymptotic distribution both for the likelihood ratio test statistic LR_n and the restricted likelihood ratio test statistic RLR_n is a $0.5\chi^2_0 + 0.5\chi^2_1$-mixture. However, since data are correlated in the presented setting, more general results have to be used. These general results are provided in Crainiceanu et al. (2002) and Crainiceanu & Ruppert (2002) and will now be described briefly.

Note, that the asymptotic distributions of LR_n and RLR_n under the iid assumption have a positive point mass in zero. This is also the case for the more general models discussed here. To analyze the point mass in zero Crainiceanu et al. (2002) developed formulae that allow an approximation to this quantity even for finite samples. By definition, the point mass in zero of the distribution of the likelihood ratio test statistic equals the probability of a global maximum of the profile log likelihood in zero. A required condition for such a global maximum is

$$\left.\frac{\partial l(\lambda)}{\partial \lambda}\right|_{\lambda=0} \leq 0. \tag{7}$$

More precisely, condition (7) corresponds to a local maximum of the profile log likelihood in zero. Therefore the probability

$$\mathbb{P}_\lambda\left(\left.\frac{\partial l(\lambda)}{\partial \lambda}\right|_{\lambda=0} \leq 0 \right) \tag{8}$$

forms an upper bound for the point mass in zero of the finite sample distribution of the likelihood ratio test statistic. As Crainiceanu et al. (2002) have shown, (8) also gives a good approximation to this point mass.

While the point mass in zero can not be computed directly, probability (8) may be computed as

$$\begin{aligned}
\mathbb{P}_\lambda\left(\left.\frac{\partial l(\lambda)}{\partial \lambda}\right|_{\lambda=0} \leq 0 \right) &= \mathbb{P}\left(\frac{u' V_\lambda^{\frac{1}{2}} R_0 Z Z' R_0 V_\lambda^{\frac{1}{2}} u}{u' V_\lambda^{\frac{1}{2}} R_0 V_\lambda^{\frac{1}{2}} u} \leq \frac{1}{n} \operatorname{tr}(Z'Z) \right) &(9)\\
&= \mathbb{P}(u'(A - \bar{d}B)u \leq 0) &(10)\\
&= \mathbb{P}\left(\sum_{i=1}^{n} \psi_i v_i \right), &(11)
\end{aligned}$$

with

$$\begin{aligned}
A &= V_\lambda^{\frac{1}{2}} R_0 Z Z' R_0 V_\lambda^{\frac{1}{2}},\\
B &= V_\lambda^{\frac{1}{2}} R_0 V_\lambda^{\frac{1}{2}},\\
\bar{d} &= \frac{1}{n} \operatorname{tr}(Z'Z)\\
u &\sim \mathrm{N}(0, \sigma^2 I_n)\\
v_1, \ldots, v_n &\ \text{iid}\ \chi_1^2
\end{aligned}$$

and $\psi_1, \ldots, \psi_n$ as the eigenvalues of the matrix $A - \bar{d}B$. Formulae (10) or (11) may then be used to evaluate the probability using exact algorithms (Farebrother (1990) or Davies (1980)) or simulation.

A similar formula may be derived for the probability of a local maximum of the profile restricted log likelihood in zero:

$$\begin{aligned}
&\mathbb{P}_\lambda\left(\left.\frac{\partial l^*(\lambda)}{\partial \lambda}\right|_{\lambda=0} \leq 0 \right)\\
&= \mathbb{P}\left(\frac{u' V_\lambda^{\frac{1}{2}} R_0 Z Z' R_0 V_\lambda^{\frac{1}{2}} u}{u' V_\lambda^{\frac{1}{2}} R_0 V_\lambda^{\frac{1}{2}} u} \leq \frac{1}{n-p-1} \operatorname{tr}(Z' R_0 Z) \right) &(12)\\
&= \mathbb{P}\left(u' \left(A - \frac{1}{n-p-1} \operatorname{tr}(Z' R_0 Z) B \right) u \leq 0 \right). &(13)
\end{aligned}$$

Again this probability forms an upper bound for the point mass in zero for the finite sample distribution of the restricted likelihood ratio test statistic.

Now we will focus on the asymptotic distributions of the likelihood ratio test statistic and the restricted likelihood ratio test statistic. Therefore we first express the profile log-likelihood and the restricted profile log-likelihood in terms of the eigenvalues $\kappa_{s,n}$, $s = 1,\dots,q$ of the matrix $Z'R_0Z$ and the eigenvalues $\pi_{s,n}$, $s = 1,\dots,q$ of the matrix $Z'Z$ (Crainiceanu & Ruppert 2002):

$$l(\lambda) \stackrel{\mathcal{D}}{=} -\frac{1}{2}\sum_{s=1}^{q}\log(1+\lambda\pi_{s,n}) - \frac{n}{2}\log\left[\sigma^2\left(\sum_{s=1}^{q}\frac{1}{1+\lambda\kappa_{s,n}}v_s + \sum_{s=q+1}^{n-p-1} v_s\right)\right]$$

and

$$l^*(\lambda) \stackrel{\mathcal{D}}{=} -\frac{1}{2}\sum_{s=1}^{q}\log(1+\lambda\kappa_{s,n}) - \frac{n-p-1}{2}\log\left[\sigma^2\left(\sum_{s=1}^{q}\frac{1}{1+\lambda\kappa_{s,n}}v_s + \sum_{s=q+1}^{n-p-1} v_s\right)\right],$$

where v_s, $s = 1,\dots,n-p-1$ are assumed to be independent and identically χ^2_1-distributed and $\stackrel{\mathcal{D}}{=}$ denotes equality of distributions.

Crainiceanu & Ruppert (2002) then establish the following general result: Assume that H_0 in (3) is true. If there exists an $a \geq 0$, so that the eigenvalues $\kappa_{s,n}$, $s = 1,\dots,q$ of the matrix $Z'R_0Z$ meet the condition $\lim_{n\to\infty} n^{-a}\kappa_{s,n} = \kappa_s$, with at least one κ_s being positive, one obtains for $n \to \infty$

$$RLR_n(n^{-a}\lambda) \xrightarrow{\mathcal{D}} RLR_\infty(\lambda) \tag{14}$$

with

$$RLR_\infty(\lambda) = \sum_{s=1}^{q}\frac{\lambda\kappa_s}{1+\lambda\kappa_s}v_s - \sum_{s=1}^{q}\log(1+\lambda\kappa_s). \tag{15}$$

If in addition the eigenvalues $\pi_{s,n}$, $s = 1,\dots,q$ of the matrix $Z'Z$ satisfy $\lim_{n\to\infty} n^{-a}\pi_{s,n} = \pi_s$, again with at least one π_s being positive, one obtains

$$LR_n(n^{-a}\lambda) \xrightarrow{\mathcal{D}} LR_\infty(\lambda) \tag{16}$$

with

$$LR_\infty(\lambda) = \sum_{s=1}^{q}\frac{\lambda\kappa_s}{1+\lambda\kappa_s}v_s - \sum_{s=1}^{q}\log(1+\lambda\pi_s). \tag{17}$$

Here $\xrightarrow{\mathcal{D}}$ denotes weak convergence.

From these general results it follows, that under the same conditions as in (14) and (16)

$$LR_n \xrightarrow{\mathcal{D}} \sup_{\lambda\geq 0} LR_\infty(\lambda) = LR_\infty, \tag{18}$$

and

$$RLR_n \xrightarrow{\mathcal{D}} \sup_{\lambda\geq 0} RLR_\infty(\lambda) = RLR_\infty. \tag{19}$$

Again we are interested in the point mass in zero of these asymptotic distributions. However, they can not be computed directly, but an upper bound for the quantities is available. Just like in the finite sample case, we treat the probabilities

$$\mathbb{P}_0\left(\frac{\partial}{\partial\lambda}LR_\infty(\lambda)\bigg|_{\lambda=0}\leq 0\right)$$

and

$$\mathbb{P}_0\left(\frac{\partial}{\partial\lambda}RLR_\infty(\lambda)\bigg|_{\lambda=0}\leq 0\right)$$

as approximations to the point masses in zero. They may be computed exact or by simulation from the expressions

$$\mathbb{P}_0\left(\frac{\partial}{\partial\lambda}LR_\infty(\lambda)\bigg|_{\lambda=0}\leq 0\right)=\mathbb{P}\left(\sum_{s=1}^{q}\kappa_s v_s\leq\sum_{s=1}^{q}\pi_s\right)\tag{20}$$

and

$$\mathbb{P}_0\left(\frac{\partial}{\partial\lambda}RLR_\infty(\lambda)\bigg|_{\lambda=0}\leq 0\right)=\mathbb{P}\left(\sum_{s=1}^{q}\kappa_s v_s\leq\sum_{s=1}^{q}\kappa_s\right),\tag{21}$$

again with $v_1,\ldots,v_q$ independent and identically χ^2_1-distributed.

Usually there is no closed form expression for the asymptotic distributions in (18) and (19) but the distributions may easily be simulated, if the asymptotic eigenvalues $\kappa_1,\ldots,\kappa_q$ and $\pi_1,\ldots,\pi_q$ are available. First consider a grid $0=d_1<\ldots<d_{max}$ of values in $[0,\infty)$. To obtain a random number from the distributions of LR_∞ or RLR_∞ one samples q independent χ^2_1-distributed random variables and evaluates $LR_\infty(\lambda)$ or $RLR_\infty(\lambda)$ at the different points of the grid. The maximum of these function evaluations may then be used as a random number from the distribution of LR_∞ or RLR_∞ if the grid is chosen in an appropriate way. One way to do this, is to compute an equidistant grid $t_2<\ldots<t_{max}$ and then to define $d_1=0, d_2=10^{t_2},\ldots,d_{max}=10^{t_{max}}$. The whole procedure is summarized in the following algorithm:

(i) Choose equidistant $t_2<\ldots<t_{max}$, e.g. $t_2=-3$, $t_{max}=2$ and $max=100$ and compute $d_1=0, d_2=10^{t_2},\ldots,d_{max}=10^{t_{max}}$.

(ii) Choose the number K of random numbers that are to be simulated and set $k=1$.

(iii) Sample q independent, χ^2_1-distributed random numbers $v_1,\ldots,v_q$ and evaluate the functions

$$LR_\infty(\lambda)=\sum_{s=1}^{q}\frac{\lambda\kappa_s}{1+\lambda\kappa_s}v_s-\sum_{s=1}^{q}\log(1+\lambda\pi_s)$$

and

$$RLR_\infty(\lambda)=\sum_{s=1}^{q}\frac{\lambda\kappa_s}{1+\lambda\kappa_s}v_s-\sum_{s=1}^{q}\log(1+\lambda\kappa_s)$$

based on this random numbers at the different values of the grid $d_1,\ldots,d_{max}$. Realizations from the asymptotic distributions of LR_n and RLR_n may then be obtained as

$$LR_\infty^{(k)}=\max\{LR_\infty(d_1),\ldots,LR_\infty(d_{max})\}$$

and

$$RLR_\infty^{(k)} = \max\{RLR_\infty(d_1), \dots, RLR_\infty(d_{max})\}.$$

(iv) If $k < K$, set $k = k+1$ and go back to (iii). If $k = K$, $LR_\infty^{(1)}, \dots, LR_\infty^{(K)}$ and $RLR_\infty^{(1)}, \dots, RLR_\infty^{(K)}$ are samples from the asymptotic distributions of LR_n and RLR_n.

3 Examples

3.1 ANOVA

As a first example we consider the balanced one-way ANOVA model with N levels and J observations per level

$$y_{ij} = \beta_0 + b_i + \varepsilon_{ij} \tag{22}$$

with $i = 1, \dots, N$, $j = 1, \dots, J$. We therefore obtain a total number of $n = NJ$ observations. For the random effects b_i we assume b_i iid $\mathrm{N}(0, \tau^2)$, $i = 1, \dots, N$ and for the error variables ε_{ij} iid $\mathrm{N}(0, \sigma^2)$, $i = 1, \dots, N$, $j = 1, \dots, J$.

First we may rewrite (22) as

$$y_i = X_i \beta_0 + Z_i b_i + \varepsilon_i$$

with

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iJ} \end{pmatrix}, \; X_i = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \; Z_i = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \; \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iJ} \end{pmatrix}$$

and $i = 1, \dots, N$. Note that the y_i are independent and identically distributed, so that the theory of Self & Liang (1987) may be used if one assumes that the number of levels N tends to infinity. However in many applications the number of levels is moderate but the total number of observations is large. So we are interested in the asymptotic distributions of the likelihood ratio test statistics for $J \to \infty$.

To get these distributions we express the model in matrix notation as

$$y = X\beta_0 + Zb + \varepsilon,$$

with

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \; X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \; Z = \begin{pmatrix} Z_1 & 0 & \dots & \dots & 0 \\ 0 & Z_2 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & Z_N \end{pmatrix},$$

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}, \; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

Now we may apply the theory presented in section 2. Therefore we need the asymptotic eigenvalues $\pi_1, \ldots, \pi_N$ and $\kappa_1, \ldots, \kappa_N$. One obtains, that all eigenvalues of the matrix $Z'Z$ are equal to J so that

$$\lim_{n\to\infty} n^{-1}\pi_{s,n} = \pi_s = \frac{1}{N} \text{ for } s = 1, \ldots, N.$$

The first $N-1$ eigenvalues of the matrix $Z'R_0Z$ are also equal to J while the last eigenvalue equals zero. One obtains

$$\lim_{n\to\infty} n^{-1}\kappa_{s,n} = \kappa_s = \frac{1}{N} \text{ for } s = 1, \ldots, N-1 \text{ and } \lim_{n\to\infty} n^{-1}\kappa_{N,n} = \kappa_N = 0.$$

In this simple example we are able to establish the asymptotic distributions of the likelihood ratio test statistics analytically through maximizing the expressions in (15) and (17):

$$LR_\infty \stackrel{\mathcal{D}}{=} \left(X_{N-1} - N - N \log\left(\frac{X_{N-1}}{N}\right)\right) \mathbf{1}_{\{X_{N-1} > N\}} \tag{23}$$

and

$$RLR_\infty \stackrel{\mathcal{D}}{=} \left(X_{N-1} - (N-1) - (N-1) \log\left(\frac{X_{N-1}}{N-1}\right)\right) \mathbf{1}_{\{X_{N-1} > N-1\}}, \tag{24}$$

with $X_{N-1} \sim \chi^2_{N-1}$. So in this case the point masses of the asymptotic distributions in zero are given by $\mathbb{P}(X_{N-1} > N-1)$ and $\mathbb{P}(X_{N-1} > N)$. Note that both probabilities are larger than 0.5 but tend to 0.5 as N increases to infinity. Moreover the distributions (23) and (24) approach the mixture $0.5\chi^2_0 + 0.5\chi^2_1$ for $N \to \infty$ by the central limit theorem.

3.2 P-splines

It is a well known fact, that many penalization approaches used in nonparametric regression may be reexpressed as mixed models (compare for example Lin & Zhang (1999) for smoothing splines or Currie & Durban (2002) for P-splines). Consider the simple model

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{25}$$

with ε_i iid $\mathrm{N}(0, \sigma^2)$. If we use P-splines (Eilers & Marx 1996) to model the nonparametric function f, we may reexpress (25) in matrix notation as

$$y = B\zeta + \varepsilon$$

where the matrix B contains the evaluations of the B-spline basis at r equally spaced knots $(\xi_1, \ldots, \xi_r)'$ and y and ε are the stacked vectors of the response and the error terms respectively. Note, that the total number of knots is determined through the number of inner knots m and the degree of the B-spline basis l via $r = m + l - 1$.

With P-splines the vector of regression parameters ζ is not estimated through ordinary least squares but through minimizing the penalized least squares criterion

$$(y - B\zeta)'(y - B\zeta) + \frac{1}{2\lambda}\zeta' D' D\zeta \tag{26}$$

with the difference matrix D, that is, jumps between adjacent coefficients in ζ are penalized using squared differences. The order of the differences is denoted by k. Note, that for $\lambda \to 0$ the estimated function $\hat{f}$ approaches a polynomial of degree $k - 1$.

If we reparameterize ζ as

$$\zeta = \tilde{X}\beta + \tilde{Z}b$$

with a k-dimensional vector β, the $(r \times k)$-dimensional matrix

$$\tilde{X} = \begin{pmatrix} 1 & \xi_1 & \dots & \xi_1^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_r & \dots & \xi_r^{k-1} \end{pmatrix},$$

an $(r - k)$-dimensional vector b and the $(r \times r - k)$-dimensional matrix $\tilde{Z} = D'(DD')^{-1}$ we may rewrite the penalized least squares criterion (26) as

$$(y - B\tilde{X}\beta - B\tilde{Z}b)'(y - B\tilde{X}\beta - B\tilde{Z}b) + \frac{1}{2\lambda}b'b.$$

In fact, this turns out to be the log likelihood of a mixed model (1) with design matrices $X = B\tilde{X}$ and $Z = B\tilde{Z}$ and random effects $b \sim \mathrm{N}(0, \tau^2 I_{r-k})$, where $\tau^2 = \sigma^2\lambda$.

Now, the above reparameterisation does not only allow to estimate σ^2 and τ^2, and therefore λ, via (restricted) maximum likelihood, but also to apply the theory presented in section 2. As a result we obtain a possibility to test

$$H_0 : f(x) = \beta_0 + \beta_1 x + \ldots + \beta_{k-1}x^{k-1}$$

versus the alternative

$$H_1 : f(x) \neq \beta_0 + \beta_1 x + \ldots + \beta_{k-1}x^{k-1}.$$

So we are able to test, whether f is a polynomial of degree $k - 1$, under the condition, that the asymptotic eigenvalues $\pi_1, \ldots, \pi_q$ and $\kappa_1, \ldots, \kappa_q$ with $q = r - k$ are available. However, this is only the case for P-splines with a B-spline basis of degree zero. Under this additional assumption we obtain, that $B'B$ is a diagonal matrix and therefore X and Z are orthogonal, since $\tilde{X}'\tilde{Z} = 0$. Now $Z'R_0Z$ simplifies to $Z'Z$ and we have $\kappa_{s,n} = \pi_{s,n}$ for $s = 1, \ldots, q$. From the definition of Z we get

$$Z'Z = (DD')^{-1}DB'BD'(DD')^{-1}.$$

For B-splines of degree zero $B'B$ equals a diagonal matrix N with entries n_j, the number of observations in the interval $[\xi_j, \xi_{j+1})$, $j = 1, \dots, r$. If we assume $\lim_{n\to\infty} \frac{n_j}{n} = h_j$, the asymptotic eigenvalues $\pi_1, \dots, \pi_q$ may be obtained as the eigenvalues of the matrix

$$(DD')^{-1}D\tilde{N}D'(DD')^{-1},$$

with $\tilde{N} = \mathrm{diag}(h_1, \dots, h_r)$.

3.3 Markov random fields

Just like P-splines, Markov random fields may be reexpressed as linear mixed models (see Besag (1974) for an introduction to Markov random fields). Consider the model

$$y_i = f_{spat}(R_i) + \varepsilon_i, \qquad i = 1, \dots, n \tag{27}$$

where R_i is the region individual i belongs to and the error terms ε_i are assumed to be iid $\mathrm{N}(0, \sigma^2)$. If we model $f_{spat}(R_i)$ through a Markov random field, we may reexpress the vector of function evaluations $f_{spat} = (f_{spat}(R_1), \dots, f_{spat}(R_n))'$ as $B\zeta$. Here ζ is an r-dimensional vector containing the different values of the spatial function in the distinct regions and B is an $(n \times r)$-dimensional $0/1$ incidence matrix that connects the i-th observation with the corresponding parameter in ζ. Again ζ may be estimated through minimizing a penalized least squares criterion or equivalently through maximizing a penalized log likelihood with penalty matrix K. The matrix K is an adjacency matrix defined through the neighbourhood structure of the Markov random field (see Clayton (1996), section 16.6.7 for details on the definition of this matrix).

Now we want to rewrite (27) as a linear mixed model. Therefore we define L from the spectral decomposition $K = \Gamma\Omega\Gamma'$, where Ω contains the $r-1$ positive eigenvalues of K, as $L = \Gamma\Omega^{\frac{1}{2}}$. Then, splitting up ζ into the scalar β and the $(r-1)$-dimensional vector b and defining the $(r \times 1)$-matrix $\tilde{X} = (1, \dots, 1)'$ and the $(r \times r-1)$-matrix $\tilde{Z} = L(L'L)^{-1}$ yields a linear mixed model (1) with $X = B\tilde{X}$, $Z = B\tilde{Z}$ and $b \sim \mathrm{N}(0, \tau^2 I_{r-1})$.

The theory presented in section 2 allows us to perform a likelihood ratio test on the presence of a spatial effect. Therefore we have to establish the asymptotic distributions of the corresponding test statistics which are determined through the asymptotic eigenvalues $\pi_1, \dots, \pi_q$ and $\kappa_1, \dots, \kappa_q$ with $q = r-1$. Since $N = B'B$ is a diagonal matrix with entries n_j equal to the number of observations in region j, $j = 1, \dots, r$, again X and Z are orthogonal. Thus we have $\kappa_{s,n} = \pi_{s,n}$ for $s = 1, \dots, q$ and the asymptotic eigenvalues are available as the eigenvalues of the matrix

$$\Omega^{-1}L'\tilde{N}L\Omega^{-1}$$

with $\tilde{N} = \mathrm{diag}(h_1, \dots, h_r)$ and $h_j = \lim_{n\to\infty} \frac{n_j}{n}$.

4 Splus-functions

As we have seen in section 2, it is possible to compute the probability of local maxima in zero for the likelihood ratios $LR_n(\lambda)$ and $RLR_n(\lambda)$ and for the asymptotic likelihood ratios $LR_\infty(\lambda)$ and $RLR_\infty(\lambda)$. Note again that these probabilities form upper bounds for the point masses in zero of the corresponding distributions of the likelihood ratio test statistics. Moreover it has become clear, how to simulate the asymptotic distributions of the likelihood ratio test statistics LR_n and RLR_n for given eigenvalues $\pi_1, \dots, \pi_q$ and $\kappa_1, \dots, \kappa_q$.

For the three examples discussed in section 3, these tasks have been implemented in Splus. For each of the examples there exist three different functions, that are described in further detail in the following sections. All functions are included in the file `testfunctions.s` and may be installed in Splus through executing the command
`> source("`*path*`\\testfunctions.s")`
where *path* has to be replaced by the correct directory in which `testfunctions.s` is stored.

4.1 Local maxima of the likelihood ratios in zero

Equations (9) and (12) allow us to compute the probabilities for local maxima in zero of the (random) functions $LR_n(\lambda)$ and $RLR_n(\lambda)$ by simulation. To be more specific, equation (11) or an analogue expression based on (13) are used in the simulation. Note, that there exist easier formulae for the ANOVA-example from section 3.1 if the true parameter λ equals zero (compare equations (22) and (23) in Crainiceanu et al. (2002)).

Three functions are available:

- `zero.probability.anova()`,
- `zero.probability.pspline()` and
- `zero.probability.mrf()`.

Each of these functions owns certain arguments which are described in the following:

`zero.probability.anova()`:

`J`	Observations per cluster. *Default:* 5
`N`	Number of clusters. *Default:* 5
`lambda`	True value of λ. *Default:* 0
`method`	How to estimate λ. `"ML"` Maximum likelihood

"REML" Restricted maximum likelihood
Default: "REML"

nsim Number of simulations that are used to evaluate the expression.
Default: 1000

zero.probability.pspline():

x Observed values of the covariate.

nknot Number of inner knots m.
Default: 20

ord Order k of the difference penalty (possible values are $k = 1$ and $k = 2$).
Default: 1

deg Degree l of the B-spline basis.
Default: 0

lambda True value of λ.
Default: 0

method How to estimate λ.
"ML" Maximum likelihood
"REML" Restricted maximum likelihood
Default: "REML"

nsim Number of simulations that are used to evaluate the expression.
Default: 1000

zero.probability.mrf():

K Penalty matrix of the Markov random field.

regions Vector containing the possible (distinct) values of the spatial covariate in the order corresponding to the penalty matrix.

reg Vector containing the realizations of the spatial covariate.

lambda True value of λ.
Default: 0

method How to estimate λ.
"ML" Maximum likelihood
"REML" Restricted maximum likelihood
Default: "REML"

nsim Number of simulations that are used to evaluate the expression.
Default: 1000

The probabilities for local maxima in zero for the asymptotic likelihood ratios $LR_\infty(\lambda)$ and $RLR_\infty(\lambda)$ may (under H_0) be computed from (20) and (21) by simulation. In this simulation we use the asymptotic eigenvalues that have been derived in the sections 3.1 to 3.3. For the ANOVA-example the probabilities for local maxima in zero coincide with the corresponding point masses in zero of the asymptotic distributions (compare Crainiceanu et al. (2002), section 4.1) and may therefore be easier computed from (23) and (24). Again three different functions are available:

- `asy.zero.probability.anova()`,
- `asy.zero.probability.pspline()` and
- `asy.zero.probability.mrf()`.

Note, that in the determination of the asymptotic eigenvalues for the P-splines we assume that the covariate is uniformly distributed over its domain (at least asymptotically). In a similar manner we assume for Markov random fields, that the observations are uniformly distributed over the different regions. Note also, that the asymptotic eigenvalues are only available for P-splines of degree zero. Since the design matrices X and Z are orthogonal for P-splines and Markov random fields, the probabilities do not dependent on the estimation technique in these cases.

The arguments of the three functions are listed below:

`asy.zero.probability.anova()`:

`N`	Number of clusters. *Default:* 5
`method`	How to estimate λ. `"ML"` Maximum likelihood `"REML"` Restricted maximum likelihood *Default:* `"REML"`

`asy.zero.probability.pspline()`:

`nknot`	Number of inner knots m. *Default:* 20
`ord`	Order k of the difference penalty (possible values are $k = 1$ and $k = 2$). *Default:* 1
`nsim`	Number of simulations that are used to evaluate the expression. *Default:* 1000

`asy.zero.probability.mrf()`:

`K`	Penalty matrix of the Markov random field.
`nsim`	Number of simulations that are used to evaluate the expression. *Default:* 1000

4.2 Asymptotic distributions of the test statistics

The asymptotic distributions of the likelihood ratio test statistics are given in (18) and (19). Again we use the asymptotic eigenvalues that have been derived in the sections 3.1 to 3.3 and apply the algorithm described in section 2. In the determination of the eigenvalues we assume the same properties as in the previous subsection. For the ANOVA-example the asymptotic distributions are given in (23) and (24), which allow us an easier simulation without the use of the algorithm from section 2. Since the design matrices X and Z are orthogonal for P-splines and Markov random fields, the asymptotic distributions do not dependent on the estimation technique in these cases.

Again three functions are available:

- `sim.distribution.anova()`,
- `sim.distribution.pspline()` and
- `sim.distribution.mrf()`.

The arguments of these functions are:

`sim.distribution.anova()`:

`N`	Number of clusters. *Default:* 5
`nsim`	How many random numbers shall be simulated? *Default:* 1000
`qqplot`	Whether to produce a QQ-plot of the simulated distribution (conditional on $LR_\infty > 0$ or $RLR_\infty > 0$ respectively) against the χ^2_1-distribution or not. *Default:* `T`
`quantiles`	Vector of quantiles that shall be computed from the simulated distribution. *Default:* 0.95

`sim.distribution.pspline()`:

`nknot`	Number of inner knots m. *Default:* 20
`ord`	Order k of the difference penalty. Possible values are $k = 1$ and $k = 2$. *Default:* 1
`nsim`	How many random numbers shall be simulated? *Default:* 1000
`ndgrid`	Number of grid points used in the algorithm. *Default:* 100
`lowert`	Smallest value of the t-grid. *Default:* -3
`uppert`	Largest value of the t-grid. *Default:* 1.2
`qqplot`	Whether to produce a QQ-plot of the simulated distribution (conditional on $LR_\infty > 0$) against the χ^2_1-distribution or not. *Default:* `T`

`quantiles` Vector of quantiles that shall be computed from the simulated distribution.
Default: 0.95

`sim.distribution.mrf()`:

`K` Penalty matrix of the Markov random field.

`nsim` How many random numbers shall be simulated?
Default: 1000

`ndgrid` Number of grid points used in the algorithm.
Default: 100

`lowert` Smallest value of the t-grid.
Default: -3

`uppert` Largest value of the t-grid.
Default: 1.2

`qqplot` Whether to produce a QQ-plot of the simulated distribution (conditional on $LR_\infty > 0$) against the χ^2_1-distribution or not.
Default: `T`

`quantiles` Vector of quantiles that shall be computed from the simulated distribution.
Default: 0.95

Each of the functions returns a list consisting of the following components:

`p0` Point mass of the asymptotic distribution in zero.

`rlr` Realizations from the asymptotic distribution of RLR_n (for ANOVA only).

`lr` Realizations from the asymptotic distribution of LR_n.

`quantiles` Requested quantiles.

For P-splines and Markov random fields an additional error message is generated if the upper value of the grid $LR_\infty(d_{max})$ has been chosen in at least one of the simulation runs. This allows the user to check the quality of the generated random numbers and to change the options (especially `uppert` and `ndgrid`) to appropriate values.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society Series B* **36**: 192–236.

Clayton, D. G. (1996). Generalized linear mixed models, *in* W. R. Gilks, S. Richardson & D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Crainiceanu, C. M. & Ruppert, D. (2002). Asymptotic distribution of likelihood ratio tests in linear mixed models.

Crainiceanu, C. M., Ruppert, D. & Vogelsang, T. J. (2002). Probability that the MLE of a variance component is zero with applications to likelihood ratio tests.

Currie, I. D. & Durban, M. (2002). Flexible smoothing with P-splines: a unified approach, *Statistical Modelling* **4**: 333–349.

Davies, R. B. (1980). The distribution of a linear combination of χ^2 random variables, *Applied Statistics* **29**: 323–333. Algorithm AS 155.

Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties, *Statistical Science* **11**: 89–121.

Farebrother, R. W. (1990). The distribution of a quadratic form in normal variables, *Applied Statistics* **39**: 294–309. Algorithm AS 256.

Lin, X. & Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines, *Journal of the Royal Statistical Society Series B* **61**: 381–400.

Self, S. G. & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association* **82**: 605–610.