ANOVA DDP

April 24, 2007

R topics documented:

---

anovaddp-package   *ANOVADDP package*

Description

An ANOVA Model for Dependent Random Measures.

Details

Package:	anovaddp
Type:	Package
Version:	0.1-0
Date:	2006-09-15
License:	See COPYING for license information

For a complete list of functions, use `library(help="matlab")`. For a high-level summary of the changes for each revision, use `file.show(system.file("NEWS", package="matlab"))`.

Author(s)

P. Roebuck, ⟨roebuck@mdanderson.org⟩

References

:TODO: Literature or other references for background information

---

anovaddp   *ANOVA model for dependent random measures*

Description

The function implements Bayesian inference for the ANOVA DDP model described in De Iorio et al (2004). The ANOVA DDP model is a model for repeated measurements data. The random effects distribution includes a regression on subject-specific covariates. The description of the arguments below includes references to the parameters used in Mueller et al. (2004), henceforth referred to as MRDM.

Usage

```
ans <- anovaddp(t, y, S, theta, C, nki, D, Dpred, a0, mp)
```

Arguments

t	numeric vector specifying all observation times for all patients. In MRDM this is the vector (t[kij], k=1...3, i=1..I[k], j=1..n[ki]).
y	numeric vector specifying the observations for all patients. In MRDM this is the vector (y[kij], k=1...3, i=1..I[k], j=1..n[ki]).
S	numeric (p x p) matrix specifying the prior mean for the covariance matrix of the multivariate normal kernel in the DP mixture. In MRDM this is Φ_0^{-1} and the initial value for S.
theta	
C	numeric (p*q x p*q) matrix specifying the variance-covariance matrix for the base measure of the DP prior. In MRDM this is C in equation (8).
nki	numeric vector of length I. The i-th element specifies the number of observations for the i-th patient. All patients are listed, starting with the first patient in the first study, through the last patient in the last study. In MRDM this is the vector (n[ki], k=1..3, i=1..I[k]).
D	numeric (I x q) matrix specifying the design vectors for the random effects regression. The i-th row specifies the design vector for the i-th patient. In MRDM this is the matrix with rows d[ki], k=1..3, i=1..I[k].
Dpred	numeric (patpred x q) matrix specifyying the design vectors for hypothetical future patients. Future patients, without observed data, are included to allow posterior predictive inference.
a0	numeric (p*q x 1) vector specifying the prior mean for the base measure of the DP. In MRDM this is the vector α in equation (8). Note that a0 includes all random effects, including the (p-p1) random effects that are not subject to the semiparametric prior. In MRDM this is the random effect z[1ki].
mp	named list specifying model parameters with components:

> **N** integer scalar specifying total number of observations regardless of study or individuals
>
> **q** integer scalar specifying number of anova effects (including a common intercept).
>
> **W** integer scalar specifying MCMC iterations
>
> **skip** integer scalar specifying MCMC burn-in
>
> **nu0** integer scalar specifying the degrees of freedom for Wishart prior for S^{-1}. In MRDM this is ν.
>
> **alpha0** numeric scalar. alpha0/2 is the shape parameter for the inverse Gamma prior on the residual variance σ^2. In MRDM this is a_σ.

beta0 numeric scalar. beta0/2 is the scale parameter for the inverse gamma prior on the residual variance σ^2. In MRDM this is ba_σ.

I integer scalar specifying total number of patients. In MRDM, I is the sum of the I[k].

I0 integer scalar specifying the number of hypothetical future patients, i.e., the number of rows in Dpred.

p integer scalar specifying the dimension of the random effects vector. In MRDM this is p, the dimension of (z_{1ki}, θ_{ki}).

p1 integer scalar specifying the dimension of the subvector of the random effects vector that is subject to the semiparametric random effects distribution. In MRDM this is p_1, the dimension of θ_{ki}.

alpM numeric scalar specifying the shape parameter of the Gamma prior on the DP total mass parameter. In MRDM this is a_M.

betM numeric scalar specifying the scale parameter of the Gamma prior on the DP total mass parameter. In MRDM this is b_M.

T0 integer scalar specifying the size of the time scale grid to report posterior and posterior predictive inference. The grid is all integers 0 through T0.

ns integer scalar specifying the number of studies

Details

All method arguments (except `"mp"`) can also be specified by either connection or filename to a datafile containing the input values.

The ANOVA DDP model described in Mueller et al. (2004) is a model for repeated measurement data. Let y[ij] denote the j-th observation for experimental unit (subject) i at known time t[ij]. The model assumes a nonlinear regression,

$$y[ij] = f(t[ij]; theta[i]) + e[ij],$$

with a normal residual $e[ij]$ $N(0, \sigma^2)$. The (p x 1) random effects vectors theta[i] are assumed to arise from a random effects distribution that is modeled as a mixture of normal kernels

$$\theta[i] \int N(A * d[i], S)dF(A)$$

with a Dirichlet process prior on the mixing measure, F $DP(F0, M)$. The vector d[i] is a known (q x 1) design vector. The (p x q) matrix A is a matrix of ANOVA effects. The base measure F0 is assumed to be multivariate normal. Let $a = vec(A)$ denote the stacked column vectors of A. The model assumes $F0(a) = N(a0, I)$. The model is completed with (conditionally) conjugate priors for S, M and $sigma^2$.

The implementation in MRDM allows to split the random effects vector into (z,theta), where z is assumed to arise from a normal random effects distribution, and only the remaining (p1 x 1) subvector theta is subject to the DP mixture prior. The case study in MRDM indexes subjects i within study k. The function anovaddp(.) does not use the k index. Instead subjects are just indexed with one running index i=1...I.

The non-linear regression function is coded in the C++ function regression(.) in the file regression.cpp. The function takes two arguments, observation time t and a (p x 1) random effects vector theta. The user needs to change this function to the desired mean function and reinstall the package. The package distribution includes an example file that implements the mean funciton used in MRDM.

The arguments Dpred, patpred and TpredMax allow to specify hypothetical future patients without observed responses. The function returns posterior and posterior predictive inference.

Analysis of the dataset described in Mueller(2005) is implemented as a demo. Data for this demo is stored in $R_LIBRARY\anovaddp\data\Rdata.zip(Windows version only). These data files need to be extracted to the $//www.blackwellpublishing.com/rss and has been reformatted for this example and included as files time.txt, dati.tx

Value

Let W0 = W-skip, and let Y denote the observed data. Returns a list with components:

m numeric (W0 x T0) matrix of posterior simulations for the mean function f(t; theta) corresponding to a hypothetical patient with design vector d = (1,0...0). Assuming that the first column in the matrix of ANOVA effects is a common baseline, and the remaining columns are offsets for different ANOVA factors, this provides posterior inference for the mean response of a baseline subject.

A0 numeric (I0 x T0) matrix of the posterior mean E(a0 | Y), for offsets selected by input Dpred.

A02 numeric (I0 x T0) matrix of the posterior mean $E(a0^2|Y)$, for offsets selected by input Dpred. nume $1...patpred, evaluated for a grid t = 0..T0$.

f02 numeric (I0 x T0) matrix specifying the predicted profile $E(\ f(t; theta[i])^2|Y)$ for a future patient i = $1...patpred, evaluated for a grid t = 0..T0$.

Author(s)

P. Roebuck, ⟨roebuck@mdanderson.org⟩ R. Herrick, ⟨rcherric@mdanderson.org⟩

References

Mueller, P., Rosner, G., De Iorio, M., and MacEachern, S. (2005). "A Nonparametric Bayesian Model for Inference in Related Studies." *Applied Statistics*, 54 (3), 611-626.

De Iorio, M., Mueller, P., Rosner, G., and Maceachern, S. (2004). "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99(465), 205–215.

Examples

```
demo(anovaddp_demo)
```