DDP ANOVA Model

May 9, 2007

R topics documented:

`ddpanova-package` *DDP ANOVA with univariate response*

Description

The package implements posterior simulation and posterior predictive inference for the DDP ANOVA model, assuming a univariate response.

Details

Package:	ddpanova
Type:	Package
Version:	1.0
Date:	2007-04-23
License:	GNU?

Use `ddpanova` to initialize and run the posterior MCMC simulation. Use `post.pred` to get posterior predictive densities for the observable outcomes under assumed design vectors for future patients. Use `post.sim` to get posterior simulations for the ANOVA parameters.

The model is as described in De Iorio et al. (2004). The same model is used in Mueller et al. (2005) as the random effects distribution in a repeated measurements model. Let y_i denote the i-th observed response for an individual (experimental unit) with design vector d_i. For example, if the model is used to jointly analyze data from two related clinical trials, with 3 treatment levels in each trial, the design vector for a patient in study 2, at treatment level 3 might be $d_i = (1, -1, 0, 1)$, including an intercept, an offset for the 2nd study, no offset for treatment level 2 and an offset for treatment level 3.

Let $N(x; m, s), Ga(x; a, b)$ and $Wish(f, A)$ denote a normal p.d.f. with moments (m, s), a Gamma p.d.f. with shape a and expectation a/b, and a Wishart distribution with degrees of freedom f and expectation fA.

We assume a mixture of ANOVA sampling model

$$y_i \sim \int N(y_i; d_i'\mu, V) dG(\mu, V),$$

with a Dirichlet process (DP) prior for the random mixing measure, $G \sim DP(G_0, \alpha)$. Here, $G_0(\mu, V)$ is the base measure and α is the total mass parameter.

Let $H_d(y) = \int N(y; d'\mu, V) dG(\mu, V)$ denote the sampling model for an observation with design vector d. The ANOVA DDP defines a joint probability model on the set of random probability measures

$$\{H_d(\cdot),\ d \in D\}$$

for designs from a design space D. The model is completed with conditioally conjugate hyperpriors. We assume $G_0(\mu, V) = Ga(1/V; s/2, sS/2)N(\mu; m, B)$ and hyperpriors $S \sim Ga(0.5q, 0.5q/R), m \sim N(a, A)$

Author(s)

Peter Mueller

Maintainer: Peter Mueller <pm@wotan.mdacc.tmc.edu>

References

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.

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.

`ddpanova` *DDP ANOVA*

Description

Initialize and run posterior MCMC for DDP ANOVA

Usage

```
ddpanova(Y = NULL, D = NULL,
         n.iter = 1000, n.discard = 100, n.reinit = NULL,
         n.batch = 50, n.predupate = 100, n.printallpars = NULL,
         verbose = 3,
         m.prior = 1, B.prior = 1, S.prior=1, alpha.prior=2,
         s = NULL, S.init = NULL,
         q = NULL, R = NULL,
         m.init = NULL, B.init = NULL,
         cc = NULL, CC = NULL,
         a = NULL, A = NULL,
         alpha = 1, a0 = 1, b0 = 1,
         k0 = 1, member0 = 0,
         px = 1,
         d0 = NULL, resid = NULL,
         nx = 100, ygrid = NULL,
         xlist = 0, header = 0)
```

Arguments

Y	(n by 1) data vector (or file name)
D	(n by p) design matrix (or file name)
n.iter	number MCMC iterations
n.discard	initial burn-in
n.reinit	reinitialize every so many iterations.
n.batch	save posterior simulations every so many iterations
n.predupate	update posterior predictive every so many iterations
n.printallpars	print all parameters every so many iterations
verbose	0=silent
m.prior	indicator for resampling (vs. fixing) m.
B.prior	indicator for resampling (vs. fixing) B.
S.prior	indicator for resampling (vs. fixing) S.
alpha.prior	indicator for resampling (vs. fixing) α. Use alpha.prior=2 for fixing over the first n.discard/2 iterations, and imputing thereafter.
s	d.f. and scale in Gamma(s/2, 1/2*s*S) base measure for precision (1/variance)
S.init	initial value for S
q	d.f. and scale in Gamma(q,q/R) prior for S
R	
m.init	mean and variance in N(m,B) base measure for location parameter
B.init	
cc	d.f. and matrix paramter in Wishart(cc,1/ccCC^{-1})$prior for$B^{-1}
mC	mean and covariance matrix in N(a,A) prior for m
A	
alpha	initial value for the total mass parameter
a0	shape and scale of Gamma prior for alpha
b0	
k0	number of initial clusters
member0	(n by 1) vector of cluster membership indicators (only necessary if k0 is not 1 or n)
px	indicator for carrying out posterior predictive inference
d0	(nd times p) design matrix for nd future observations.
resid	binary vector of length nd, of indicators for including residual variance in the posterior predictive.
nx	grid size for plotting posterior predictive
ygrid	lower and upper bound of the grid for posterior predictive inference.
xlist	indicator for evaluating posterior fitted values for the observed data points.
header	indicator for the data files including a header line (only meaningful if Y and D are file names).

Details

See ddpanova-package for an explaination of all parameters. The function ddpanova initializes and carries out MCMC posterior simulation. Simulation output is written to auxilary files (with .mdp extension) in the working directory. Use post.pred and post.sim to read in summaries.

Value

Posterior simulations are written into working files in the current working directory. Use `setwd` to change it if desired.

The function has no return values.

Author(s)

Peter Mueller

References

See `ddpanova-package`

Examples

```
## Not run:
  ## simulate data from a mixture of two normal densities
    nobs <- 500
    y1   <-rnorm(nobs, 3,.8)
    ## y2 = 0.6
    y21 <- rnorm(nobs,2.0, 0.8)
    y22 <- rnorm(nobs,4.0, 1.0)
    u <- runif(nobs)
    y2 <- ifelse(u<0.6,y21,y22)
    y    <- c(y1,y2)

    ## design matrix with main effect mu and +/-offset
    d    <- cbind(rep(1,2*nobs),c(rep(-1,nobs),rep(1,nobs)))
    p    <- ncol(d)
    nobs <- 2*nobs

  ## run MCMC
    d0 <- rbind(c(1,-1),c(1,1))  # design matrix for post predictive
    ddpanova(y,d,n.iter=1000,d0=d0,ygrid=c(0,6))

  ## get posterior pred density for future observations
  ## with design matrix d=(1,-1)
  ## rows of d0.
    pp <- post.pred()
    plot(pp$ygrid[1,],pp$py[1,],type="l",xlab="Y",ylab="p",bty="l")
  ## overlay truth:
    p1 <- dnorm(pp$ygrid[1,], 3.0, 0.8)
    lines(pp$ygrid[1,],p1, lty=2)

  ## ... and d=(1,1)
    plot(pp$ygrid[2,],pp$py[2,],xlab="Y",ylab="p",bty="l",type="l")
    p2 <- 0.6*dnorm(pp$ygrid[2,], 2.0, 0.8) +
          0.4*dnorm(pp$ygrid[2,], 4.0, 1.0)
    lines(pp$ygrid[2,],p2, lty=2)
## End(Not run)
```

`ddpsurvival` *DDP SURVIVAL*

Description

Initialize and run posterior MCMC for DDP ANOVA with right censored survival data.

Usage

```
ddpsurvival(Y = NULL, D = NULL,
         n.iter = 1000, n.discard = 100, n.reinit = NULL,
         n.batch = 50, n.predupate = 100, n.printallpars = NULL,
         verbose = 3,
         m.prior = 1, B.prior = 1, S.prior=1, alpha.prior=2,
         s = NULL, S.init = NULL,
         q = NULL, R = NULL,
         m.init = NULL, B.init = NULL,
         cc = NULL, CC = NULL,
         a = NULL, A = NULL,
         alpha = 1, a0 = 1, b0 = 1,
         k0 = 1, member0 = 0,
         px = 1,
         d0 = NULL, resid = NULL,
         nx = 100, ygrid = NULL,
         xlist = 0, header = 0)
```

Arguments

All arguments are the same as in `ddpanova`, except the following

`Y`	(`n` by 3) data matrix (or file name). The first column is the event time. For rows with censored observations (2nd column=0), the first column is an initial value for an imputed event time. The second column is an indicator for an event, i.e., 0 for censored observations and 1 for event times. The third column is the last follow-up time (equal the 1st column for events).

Details

See `ddpanova-package` for an explaination of all parameters. The function `ddpsurvival` initializes and carries out MCMC posterior simulation. Simulation output is written to auxilary files (with .mdp extension) in the working directory. Use `post.pred` and `post.sim` to read in summaries.

Value

Posterior simulations are written into working files in the current working directory. Use `setwd` to change it if desired.

The function has no return values.

Author(s)

Peter Mueller

References

See ddpanova-package

Examples

```
## Not run:
  ## simulate data from a mixture of two normal densities
    ni <- 500
    y1   <-rnorm(ni, 3,.8)
    ## y2 = 0.6
    y21 <- rnorm(ni,2.0, 0.8)
    y22 <- rnorm(ni,4.0, 1.0)
    u <- runif(ni)
    y2 <- ifelse(u<0.6,y21,y22)
    y    <- c(y1,y2)
    nobs <- 2*ni

 ## now add censoring times
    yc <- rnorm(nobs,4,1.2)
    event <- ifelse(y<yc,1,0)
    Y <- cbind(y,event,yc)

 ## design matrix with main effect mu and +/-offset
    d    <- cbind(rep(1,nobs),c(rep(-1,ni),rep(1,ni)))
    p    <- ncol(d)

  ## run MCMC
    d0 <- rbind(c(1,-1),c(1,1))  # design matrix for post predictive
    cat("\n Running 1000 iterations next -- be patient...\n")
    ddpsurvival(Y,d,n.iter=1000,d0=d0,ygrid=c(0,6))

  ## get posterior pred density for future observations
  ## with design matrix d=(1,-1)
  ## rows of d0.
    pp <- post.pred()
    plot(pp$ygrid,pp$py[1,],type="l",xlab="Y",ylab="p",bty="l")
  ## overlay truth:
    p1 <- dnorm(pp$ygrid, 3.0, 0.8)
    lines(pp$ygrid, p1, lty=2)

  ## ... and d=(1,1)
    plot(pp$ygrid,pp$py[2,],xlab="Y",ylab="p",bty="l",type="l")
    p2 <- 0.6*dnorm(pp$ygrid, 2.0, 0.8) +
          0.4*dnorm(pp$ygrid, 4.0, 1.0)
    lines(pp$ygrid,p2, lty=2)

  ## and same as survival curves, assuming that the data were log(times)
    Sy  <- pp$Sy # survival curve
    S1  <- 1-pnorm(pp$ygrid, 3.0, 0.8)          # truth 1
    S2 <- 1 - (0.6*pnorm(pp$ygrid, 2.0, 0.8) +  # truth 2
               0.4*pnorm(pp$ygrid, 4.0, 1.0) )

    lygrid <- exp(pp$ygrid)                     # grid on time scale
    plot (lygrid,Sy[1,],type="l",bty="l",           # estimate
          xlab="Y",ylab="px",ylim=c(0,1),xlim=c(0,150))
    lines(lygrid,Sy[2,],type="l",col=2)
```

```
    lines(lygrid,S1,    type="l",col=1,lty=2)       # simuulation truth
    lines(lygrid,S2,    type="l",col=2,lty=2)
## End(Not run)
```

`post.pred` *Posterior predictive inference*

Description

Return posterior predictive densities for future observations.

Usage

```
post.pred()
```

Details

Must have called `ddpanova` or `ddpsurvival` first.

Returns the posterior predictive distribution $p(y^* \mid d^*, data)$ for a future observation $y^* = y_{n+1}$ with design vector d^*. The argument `d0` of `ddpanova` specifies the list of `nd` future designs that are considered.

Value

The function returns a list:

`ygrid`	(`nx` x 1) vector of grid values for the predictive for future units with design vectors given in `d0`
`py`	(`nd` by `nx`) matrix of posterior predictive probabilities $p(y^* \mid d^*, data)$, evaluated on the grid given in `ygrid`.
`Sy`	
`sSy`	
`hy`	
`shy`	
`censoring`	indicator whether call was following `ddpsurvival` (`censoring=1`) or `ddpanova` (`censoring=0`).

Note

Use `post.sim` to obtain a Monte Carlo sample of parameter values.

See `ddpanova` for an example.

Author(s)

Peter Mueller

References

see `ddpanova-package`

post.sim *Posterior Monte Carlo sample*

Description

Retrieves a simulated samples μ from the posterior estiamted mixing measure $\bar{G} = E(G \mid data)$.

Usage

```
post.sim()
```

Value

The function returns a matrix

Z (M by p) matrix with one simuulation in each line.

Note

see ddpanova for an example.

Author(s)

Peter Mueller

References

see ddpanova-package