Package 'h.likelihood'

September 20, 2010

Type Package

Title Statistical Modeling and Inference via Hierarchical Likelihood

Version 2010.9.20

Date 2010-09-20

Author Xia Shen, Marek Molas and Il Do Ha

Maintainer Xia Shen <xia.shen@lcb.uu.se>

Description The package provides a top interface of hierarchical likelihood (h-likelihood) based models. It currently covers the estimation of hierarchical generalized linear models (HGLMs) and frailty models.

License GPL

LazyLoad yes

Depends hglm, HGLMMM, lattice, Matrix, numDeriv

R topics documented:

`h.likelihood-package`

Statistical Modeling and Inference via Hierarchical Likelihood

---

Description

The package provides a top interface of hierarchical likelihood (h-likelihood) based models. It currently covers the estimation of hierarchical generalized linear models (HGLMs) and frailty models.

Details

Package:	h.likelihood
Type:	Package
Version:	2010.9.20
Date:	2010-00-20
License:	GPL
LazyLoad:	yes
Depends:	hglm, HGLMMM

Author(s)

Xia Shen, Marek Molas and Il Do Ha

Maintainer: Xia Shen <xia.shen@lcb.uu.se>

References

Ha, I.D. and Lee, Y. (2003). Estimating frailty models via Poisson Hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, **12**, 663-681.

Ha, I.D. and Lee, Y. (2005). Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models. *Biometrika*, **92**, 717-723.

Ha, I.D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**, 233-243.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. Series B (Methological)* **58**, 619-678.

Lee, Y. and Nelder, J.A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88**, 987-1006.

Lee, Y., Nelder, J.A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects*. Boca Raton: Chapman & Hall/CRC.

Molas, M. and Lesaffre, E. (2010). Hierarchical Generalized Linear Models: the R Package HGLMMM. **Submitted**.

Noh, M. and Lee, Y. (2007). REML estimation for binary data in GLMMs. *Journal of Multivariate Analysis* **98**, 896-915.

Ronnegard, L., Shen, X. and Alam, M. (2010). hglm: A Package for Fitting Hierarchical Generalized Linear Models. *The R Journal.* ***(to appear)***

See Also

hglm-package, HGLMMM-package

cgd *Chronic Granulomatous Disease (CGD) Infection Data*

Description

The CGD data set in Fleming and Harrington (1991) consists of a placebo-controlled randomized trial of gamma interferon in chronic granulomatous disease.

Usage

```
data(cgd)
```

Format

A data frame with 203 observations on the following 16 variables.

`OBS` There were 203 observations.

`HOSPITAL` Hospitial ID: There were 13 hospotials where trials were conducted.

`PATIENT` Patient ID: There were 128 patients.

`TIME` The recurrent infection times of each pateient from the different hospitals.

`DEL` Censoring indicator(1 = uncensored, 0 = censored).

`TRT` Treatment Code(1 = gamma interferon, 0 = placebo).

`INHERIT` Pattern of inheritance (0 = autosomal recessive, 1 = X-linked).

`AGE` Age of each patient, years.

`HEIGHT` Height of each patient, in cm.

`WEIGHT` Weight of each patient, in kg.

`STEROIDS` Using corticosteroids at times of study centry(1 = Yes, 0 = No).

`PROPYLAC` Using prophylactic antibiotics at time of study entry(1 = Yes, 0 = No).

`SEX` Sex of each patient(0 = male, 1 = female).

`H` Hospital region(0 = U.S.A., 1 = Europe).

`LONGI` A longitudinal variable representing the accumulated time from the firts infection(in years).

`ENUM` Sequence number. For each patient, the infection records are in sequnce number.

References

Fleming and Harrington (1991). *Counting processes and survival analysis*. Wiley: New York.

Examples

```
data(cgd)
```

`Frailty.h` *Fitting Frailty Models using H-likelihood Approach*

Description

Frailty.h is used to fit frailty models using h-likelihood approach. The package fits Cox's proportional hazards models with random effects (or frailties). Here, for the frailty distribution lognormal or gamma is allowed. The h-likelihood obviates the need for marginalization over the frailty distribution, leading to a statistically efficient procedure for various random-effect models including frailty models.

Usage

```
Frailty.h(formulaMain, censor, DataMain, RandDist = "Normal", mord = 0,
          dord = 1, Maxiter = 200, convergence = 1e-7, contrasts = NULL)
```

Arguments

`formulaMain`	Specify the formula for the mean structure of the model. e.g. `y ~ x + (1|id)`, `y`: response, `x`: fixed covariate, `id`: random effect.
`censor`	One vector for censoring indicator (1 = uncensored, 0 = censored).
`DataMain`	Dataframe for formulaMain.
`RandDist`	Distribution for random effect ("Normal" or "Gamma").
`mord`	Order for the mean model (0 or 1); default = 0.
`dord`	Order for the dispersion components (1 or 2); default = 1.
`Maxiter`	Maximum number of iterations; default = 1.
`convergence`	Specify the convergence criterion, the default is 1e-7.
`contrasts`	Caution as it is currently not fully developed.

Details

`Frailty.h` produces estimates of fixed effects and frailty parameters as well as their standard errors. Also, `Frailty.h` makes it possible to fit models where the frailty distribution is not necessarily normal and estimate variance components when frailty structure is shared or nested.

Author(s)

Il Do Ha

References

Ha, I.D. and Lee, Y. (2003). Estimating frailty models via Poisson Hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, **12**, 663-681.

Ha, I.D. and Lee, Y. (2005). Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models. *Biometrika*, **92**, 717-723.

Ha, I.D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**, 233-243.

Lee, Y., Nelder, J.A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects*. Boca Raton: Chapman & Hall/CRC.

Examples

```
## Not run:
data(cgd)
data(kidney)

#### Frailty model

#### Analysis of Kidney data
kln11 <- Frailty.h(time ~ sex + age + (1|patient)-1, kidney$censor, kidney,
                   RandDist = "Normal", mord = 1, dord = 1)
kln12 <- Frailty.h(time ~ sex + age + (1|patient) - 1, kidney$censor, kidney,
                   RandDist = "Normal", mord = 1, dord = 2)
kg11 <- Frailty.h(time ~ sex + age + (1|patient) - 1, kidney$censor, kidney,
                  RandDist = "Gamma", mord = 1, dord = 1)
kg12 <- Frailty.h(time ~ sex + age + (1|patient) - 1, kidney$censor, kidney,
                  RandDist = "Gamma", mord = 1, dord = 2)

#### Analysis of CGD data
cgd11 <- Frailty.h(TIME ~ TRT + (1|HOSPITAL) + (1|PATIENT) - 1, cgd$DEL, cgd,
                   RandDist = "Normal", mord = 1, dord = 1)
cgd12 <- Frailty.h(TIME ~ TRT + (1|HOSPITAL) + (1|PATIENT) - 1, cgd$DEL, cgd,
                   RandDist = "Normal", mord = 1, dord = 2)


## End(Not run)
```

`HGLM` *Fitting Hierarchical Generalized Linear Models*

Description

This function fits hierarchical generalized linear models (HGLMs) using various approximation methods.

Usage

```
HGLM(y = NULL, X = NULL, Z = NULL, X.disp = NULL,
     family = gaussian(link = identity),
     random.family = gaussian(link = identity), method = "EQL",
     conv = 1e-04, maxit = 20, fixed = NULL, random = NULL,
     disp = NULL, link.disp = "log", disp.random = NULL,
     data = NULL, data.random = NULL, fix.disp = NULL,
     Offset = NULL, Weights = NULL, disp.start = 0, binomial.N = NULL,
     start.fixed = NULL, start.random = NULL, start.disp = NULL,
     start.disp.random = NULL, info = TRUE, debug = FALSE,
     contrasts = NULL)
```

Arguments

`y`	the dependent variable, only available when `method = 'EQL'`.
`X`	a design matrix for the fixed effects, only available when `method = 'EQL'`.
`Z`	an design matrix for the random effects, only available when `method = 'EQL'`.

`X.disp` a design matrix for the fixed effects in the dispersion part of the model, only available when `method = 'EQL'`.

`family` a description of the error distribution and link function to be used in the mean part of the model. (See `family` for details of family functions.)

`random.family` a description of the error distribution and link function to be used in the variance part of the model.

`method` estimation method, which can be `'EQL'`, `'HL01'`, or `'HL11'`, where `'EQL'` can ONLY be used when ONLY ONE random effect term is specified. `'EQL'` is the method of interconnected GLMs presented in Lee et al. (2006), and for `'HL01'` and `'HL11'`, see Lee and Nelder (2001).

`conv` convergence criterion, the default is 1e-4, for models with many random effects could be set less strict.

`maxit` maximum number of iterations in the IWLS algorithm, only available when `method = 'EQL'`.

`fixed` a `formula` specifying the fixed effects part of the model, and the format is `Response ~ Fixed.Effect.1 + ... + Fixed.Effect.p`.

`random` a one-sided `formula` specifying the random effects part of the model, and the format is `~ (Random.Effect.1 | Subject.1) + ... + (Random.Effect.q | Subject.q)`.

`disp` a one-sided `formula` specifying the fixed effects in the dispersion part of the model, and the format is `~ Effect.1 + ... + Effect.N`.

`link.disp` the link function for the dispersion part of the model, only available when `method = 'EQL'`.

`disp.random` a list of one-sided `formula`e for the dispersion strucutre of each random effects, which has the format of `list(one = ~ Effect.1.1 + ..., two = ~ Effect.2.1 + ..., three = ..., ...)`, only available when `method = 'HL01'` or `'HL11'`.

`data` the data frame to be used together with `fixed` and `random`.

`data.random` a list of `data.frame`s for `disp.random`, which has the format of `list(one = data.Random.1, two = data.Random.2,...)`, only available when `method = 'HL01'` or `'HL11'`.

`Weights` prior weights to be specified in weighted regression, only available when `method = 'EQL'`.

`fix.disp` a numeric value if the dispersion parameter of the mean model is known for example 1 for binomial and Poisson models.

`Offset` an offset for the linear predictor of the mean model.

`disp.start` (starting) values for the overdispersion structure - vector of length equal to the number of parameters in the overdispersion structure, only available when `fix.disp = NULL` and `method = 'HL01'` or `'HL11'`.

`binomial.N` the number of trials for each observation for binomial models.

`start.fixed` optional starting values for fixed effects in the mean structure (one vector of numeric values).

`start.random` optional starting values for random effects in the mean structure (one vector of numeric values).

`start.disp` optional starting values for parameters of dispersion components of the residuals (one vector of numeric values).

`start.disp.random`
: optional starting values for parameters of dispersion components of random effects (one vector of numeric values).

`info`
: a request to display of iteration information if `TRUE`, only available when `method = 'HL01'` or `'HL11'`.

`debug`
: a request to display of iteration mechanism progress in detail if `TRUE`, only available when `method = 'HL01'` or `'HL11'`.

`contrasts`
: see `lm`, caution as it is currently not fully developed, only available when `method = 'HL01'` or `'HL11'`.

Details

When `method = 'EQL'`, all the model checking functions in the `hglm-package` are available on the object returned; Otherwise, all the model checking functions in the `HGLMMM-package` are available on the object returned.

Value

When `method = 'EQL'`, an object of class `hglm` is returned, see `hglm`; Otherwise, an object of class `HGLM` is returned, see `HGLMfit`.

Note

The function provides a unified interface to the `hglm-package` developed by Moudud Alam, Lars Ronnegard and Xia Shen, and the `HGLMMM-package` developed by Marek Molas.

Author(s)

Xia Shen and Marek Molas

References

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. Series B (Methological)* **58**, 619-678.

Lee, Y. and Nelder, J.A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88**, 987-1006.

Lee, Y., Nelder, J.A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects*. Boca Raton: Chapman & Hall/CRC.

Noh, M. and Lee, Y. (2007). REML estimation for binary data in GLMMs. *Journal of Multivariate Analysis* **98**, 896-915.

Ronnegard, L., Shen, X. and Alam, M. (2010). hglm: A Package for Fitting Hierarchical Generalized Linear Models. *The R Journal.* ***(to appear)***

Molas, M. and Lesaffre, E. (2010). Hierarchical Generalized Linear Models: the R Package HGLMMM. **Submitted**.

See Also

`hglm-package`, `HGLMMM-package`, `hglm`, `HGLMfit`.

Examples

```
data(semiconductor)

# ----- use 'EQL'

h.gamma.normal <- HGLM(fixed = y ~ x1 + x3 + x5 + x6,
                       random = ~ 1|Device,
                       family = Gamma(link = log),
                       disp = ~ x2 + x3, data = semiconductor)

summary(h.gamma.normal)

plot(h.gamma.normal, cex = .6, pch = 1,
     cex.axis = 1/.6, cex.lab = 1/.6,
     cex.main = 1/.6, mar = c(3, 4.5, 0, 1.5))

# ----- use 'HL(0,1)'

RSC <- data.frame(int = rep(1, 16))
h.gamma.normal <- HGLM(fixed = y ~ x1 + x3 + x5 + x6,
                       random = ~ 1|Device,
                       family = Gamma(link = log),
                       disp = ~ x2 + x3, data = semiconductor,
                       method = 'HL01', disp.start = c(0, 0, 0),
                       disp.random = list(one = ~ 1), data.random = list(RSC))

# ----- use 'HL(1,1)'

RSC <- data.frame(int = rep(1, 16))
h.gamma.normal <- HGLM(fixed = y ~ x1 + x3 + x5 + x6,
                       random = ~ 1|Device,
                       family = Gamma(link = log),
                       disp = ~ x2 + x3, data = semiconductor,
                       method = 'HL11', disp.start = c(0, 0, 0),
                       disp.random = list(one = ~ 1), data.random = list(RSC))
```

`kidney` *Kidney Infection Data*

Description

The data presented by McGilchrist and Aisbett (1991) consist of times to the first and second recurrences of infection in 38 kidney patients using a portable dialysis machine.

Usage

```
data(kidney)
```

Format

A data frame with 76 observations on the following 9 variables.

`time` Survival time: Time to infection since insertion of the catheter

`censor` Censoring indicator(1 = uncensored, 0 = censored).

`obs` There were 76 observations.

`patient` Patient ID: There were 38 patients with two recurrences.

`age` Age of each patient.

`sex` Sex of each patient(1 = male, 2 = female).

`dy0` GN type of disease(1 = Yes, 0 = No).

`dy1` AN type of disease(1 = Yes, 0 = No).

`dy2` PKD type of disease(1 = Yes, 0 = No).

References

McGilchrist and Aisbett(1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461-466.

Examples

```
data(kidney)
```

Index