Package 'hdbayes'

July 22, 2025

Title Bayesian Analysis of Generalized Linear Models with Historical Data

Version 0.1.1

Description User-friendly functions for leveraging (multiple) historical data set(s) for generalized linear models. The package contains functions for sampling from the posterior distribution of a generalized linear model using the prior induced by the Bayesian hierarchical model, power prior by Ibrahim and Chen (2000) <doi:10.1214/ss/1009212673>, normalized power prior by Duan et al. (2006) <doi:10.1002/env.752>, normalized asymptotic power prior by Ibrahim et al. (2015) <doi:10.1002/sim.6728>, commensurate prior by Hobbs et al. (2011) <doi:10.1111/j.1541-0420.2011.01564.x>, robust meta-analytic-predictive prior by Schmidli et al. (2014) <doi:10.1111/biom.12242>, the latent exchangeability prior by Alt et al. (2023) <doi:10.48550/arXiv.2303.05223>, and a normal (or half-normal) prior. Functions for computing the marginal log-likelihood under each of the implemented priors are also included. The package compiles all the 'CmdStan' models once during installation using the 'instantiate' package.

License MIT + file LICENSE

URL https://github.com/ethan-alt/hdbayes

BugReports https://github.com/ethan-alt/hdbayes/issues

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.2.0)

Imports instantiate (>= 0.1.0), callr, fs, formula.tools, stats, posterior, enrichwith, mclust, bridgesampling, mvtnorm

Suggests cmdstanr (>= 0.6.0), ggplot2, knitr, parallel, rmarkdown, tibble, dplyr

Additional_repositories https://mc-stan.org/r-packages/

SystemRequirements CmdStan (https://mc-stan.org/users/interfaces/cmdstan)

LazyData true

Collate 'E1684-data.R' 'E1690-data.R' 'E1694-data.R' 'E2696-data.R'
'IBCSG_curr-data.R' 'IBCSG_hist-data.R' 'actg019-data.R'
'actg036-data.R' 'data_checks.R' 'expfam_loglik.R'
'get_stan_data.R' 'glm_bhm.R' 'glm_bhm_lognc.R'
'glm_commensurate.R' 'mixture_loglik.R'
'glm_commensurate_lognc.R' 'glm_leap.R' 'glm_leap_lognc.R'
'glm_logml_commensurate.R' 'glm_logml_leap.R' 'glm_logml_map.R'
'glm_logml_napp.R' 'glm_logml_npp.R' 'glm_logml_post.R'
'glm_pp_lognc.R' 'glm_logml_pp.R' 'glm_napp.R'
'glm_npp_lognc.R' 'glm_npp.R' 'glm_post.R' 'glm_pp.R'
'glm_rmap.R' 'hdbayes-package.R' 'lm_npp.R' 'zzz.R'

VignetteBuilder knitr

NeedsCompilation yes

Author Ethan M. Alt [aut, cre, cph] (ORCID:
<https://orcid.org/0000-0002-6112-9030>),
Xinxin Chen [aut],
Luiz M. Carvalho [aut],
Joseph G. Ibrahim [aut],
Xiuya Chang [ctb]

Maintainer Ethan M. Alt <ethanalt@live.unc.edu>

Repository CRAN

Date/Publication 2024-08-22 23:50:02 UTC

Contents

`actg019` *AIDS Clinical Trial ACTG019*

Description

A data set from the AIDS clinical trial ACTG019 (https://clinicaltrials.gov/ct2/show/NCT00000736) comparing zidovudine (AZT) with a placebo in adults with asymptomatic HIV infection. The study results were described in Volberding et al. (1990) doi:10.1056/NEJM199004053221401.

Usage

```
actg019
```

Format

A data frame with 822 rows and 5 variables:

outcome outcome variable with 1 indicating death, development of AIDS or AIDS-related complex (ARC) and 0 otherwise

age patient age in years

treatment treatment indicator, 0 = placebo, 1 = AZT

race race indicator, 0 = non-white, 1 = white

cd4 CD4 cell count

References

Volberding, P. A., Lagakos, S. W., Koch, M. A., Pettinelli, C., Myers, M. W., Booth, D. K., Balfour, H. H., Reichman, R. C., Bartlett, J. A., Hirsch, M. S., Murphy, R. L., Hardy, W. D., Soeiro, R., Fischl, M. A., Bartlett, J. G., Merigan, T. C., Hyslop, N. E., Richman, D. D., Valentine, F. T., Corey, L., and the AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases (1990). Zidovudine in asymptomatic human immunodeficiency virus infection. New England Journal of Medicine, 322(14), 941–949.

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999). Prior elicitation, Variable Selection and Bayesian computation for Logistic Regression Models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 61(1), 223–242.

actg036 *AIDS Clinical Trial ACTG036*

Description

A data set from the AIDS clinical trial ACTG036 (https://clinicaltrials.gov/study/NCT00001104) comparing zidovudine (AZT) with a placebo in patients with hereditary coagulation disorders and HIV infection. The study results were described in Merigan et al. (1991) doi:10.1182/blood.V78.4.900.900. This data set has the same variables as the actg019 data set. We can use the actg019 data as the historical data and the actg036 data as the current data.

Usage

```
actg036
```

Format

A data frame with 183 rows and 5 variables:

outcome outcome variable with 1 indicating death, development of AIDS or AIDS-related complex (ARC) and 0 otherwise

age patient age in years

treatment treatment indicator, 0 = placebo, 1 = AZT

race race indicator, 0 = non-white, 1 = white

cd4 CD4 cell count

References

Merigan, T., Amato, D., Balsley, J., Power, M., Price, W., Benoit, S., Perez-Michael, A., Brownstein, A., Kramer, A., and Brettler, D. (1991). Placebo-controlled trial to evaluate zidovudine in treatment of human immunodeficiency virus infection in asymptomatic patients with hemophilia. NHF-ACTG 036 Study Group. Blood, 78(4), 900–906.

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999). Prior elicitation, Variable Selection and Bayesian computation for Logistic Regression Models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 61(1), 223–242.

E1684 *ECOG E1684 Trial*

Description

A data set from the ECOG E1684 trial comparing the high-dose interferon alfa-2b (IFN) therapy with the observation in resected high-risk melanoma patients. The study results were described in Kirkwood et al. (1996) doi:10.1200/JCO.1996.14.1.7.

Usage

```
E1684
```

Format

A data frame with 262 rows and 8 variables:

failtime time to relapse in years

failcens censoring indicator for time to relapse, 0 = did not relapse, 1 = relapsed

survtime time to death in years

survcens censoring indicator for time to death, 0 = alive, 1 = dead

treatment treatment indicator, 0 = observation, 1 = high-dose IFN

sex gender indicator, 0 = male, 1 = female

age patient age in years

node_bin indicator for having more than one cancerous lymph node, 0 = with one or no cancerous lymph nodes, 1 = with more than one cancerous lymph node

References

Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C., and Blum, R. H. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: The Eastern Cooperative Oncology Group trial EST 1684. Journal of Clinical Oncology, 14(1), 7–17.

E1690 *ECOG E1690 Trial*

Description

A data set from the ECOG E1690 trial evaluating the effectiveness of the interferon alfa-2b (IFN) therapy compared to the observation in high-risk melanoma patients. There were three arms in the trial: high-dose IFN, low-dose IFN, and observation. The study results were described in Kirkwood et al. (2000) doi:10.1200/JCO.2000.18.12.2444. Here, we only consider the high-dose IFN arm and the observation arm so that this data set has the same variables as the E1684 data set. We can use the E1684 data as the historical data and the E1690 data as the current data.

Usage

```
E1690
```

Format

A data frame with 426 rows and 8 variables:

failtime time to relapse in years

failcens censoring indicator for time to relapse, 0 = did not relapse, 1 = relapsed

survtime time to death in years

survcens censoring indicator for time to death, 0 = alive, 1 = dead

treatment treatment indicator, 0 = observation, 1 = high-dose IFN

sex gender indicator, 0 = male, 1 = female

age patient age in years

node_bin indicator for having more than one cancerous lymph node, 0 = with one or no cancerous lymph nodes, 1 = with more than one cancerous lymph node

References

Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., Ernstoff, M. S., Smith, T. J., Rao, U., Steele, M., and Blum, R. H. (2000). High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190. Journal of Clinical Oncology, 18(12), 2444–2458.

E1694 *ECOG E1694 Trial*

Description

A data set from the ECOG E1694 trial comparing the GM2-KLH/QS-21 (GMK) vaccine with high-dose interferon alfa-2b (IFN) therapy in resected high-risk melanoma patients. The study results were described in Kirkwood et al. (2001) doi:10.1200/JCO.2001.19.9.2370. This data set only includes patients without nodal metastasis and has the same variables as the E2696 data set. We can use the E2696 data as the historical data and the E1694 data as the current data.

Usage

```
E1694
```

Format

A data frame with 200 rows and 6 variables:

failtime relapse-free survival (RFS) times (in months)

failind relapse indicator, 0 = right censored, 1 = relapse

treatment treatment indicator, 0 = GMK, 1 = IFN

sex gender indicator, 0 = male, 1 = female

age patient age in years

perform ECOG performance status indicator, 0 = fully active patient, able to carry on all pre-disease performance without restriction, 1 = restricted in physically strenuous activity, but are ambulatory and able to carry out work of a light or sedentary nature

References

Kirkwood, J. M., Ibrahim, J. G., Sosman, J. A., Sondak, V. K., Agarwala, S. S., Ernstoff, M. S., and Rao, U. (2001). High-dose interferon alfa-2b significantly prolongs relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: Results of intergroup trial E1694/S9512/C509801. Journal of Clinical Oncology, 19(9), 2370–2380.

`E2696` *ECOG E2696 Trial*

Description

A data set from the ECOG E2696 trial comparing the combination of the GM2-KLH/QS-21 (GMK) vaccine and high-dose interferon alfa-2b (IFN) therapy with the GMK vaccine alone in resected high-risk melanoma patients. The study results were described in Kirkwood et al. (2001) doi:10.1200/JCO.2001.19.5.1430. This data set only includes patients without nodal metastasis.

Usage

```
E2696
```

Format

A data frame with 105 rows and 6 variables:

failtime relapse-free survival (RFS) times (in months)

failind relapse indicator, 0 = right censored, 1 = relapse

treatment treatment indicator, 0 = GMK, 1 = GMK and IFN

sex gender indicator, 0 = male, 1 = female

age patient age in years

perform ECOG performance status indicator, 0 = fully active patient, able to carry on all pre-disease performance without restriction, 1 = restricted in physically strenuous activity, but are ambulatory and able to carry out work of a light or sedentary nature

References

Kirkwood, J. M., Ibrahim, J., Lawson, D. H., Atkins, M. B., Agarwala, S. S., Collins, K., Mascari, R., Morrissey, D. M., and Chapman, P. B. (2001). High-dose interferon alfa-2b does not diminish antibody response to GM2 vaccination in patients with resected melanoma: Results of the multicenter eastern cooperative oncology group phase II trial E2696. Journal of Clinical Oncology, 19(5), 1430–1436.

`glm.bhm` *Posterior of Bayesian hierarchical model (BHM)*

Description

Sample from the posterior distribution of a GLM using the Bayesian hierarchical model (BHM).

Usage

```
glm.bhm(
  formula,
  family,
  data.list,
  offset.list = NULL,
  meta.mean.mean = NULL,
  meta.mean.sd = NULL,
  meta.sd.mean = NULL,
  meta.sd.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`formula`	a two-sided formula giving the relationship between the response variable and covariates.
`family`	an object of class `family`. See `?stats::family`.
`data.list`	a list of `data.frames`. The first element in the list is the current data, and the rest are the historical data sets.
`offset.list`	a list of vectors giving the offsets for each data. The length of `offset.list` is equal to the length of `data.list`. The length of each element of `offset.list` is equal to the number of rows in the corresponding element of `data.list`. Defaults to a list of vectors of 0s.

`meta.mean.mean` a scalar or a vector whose dimension is equal to the number of regression coefficients giving the means for the normal hyperpriors on the mean hyperparameters of regression coefficients. If a scalar is provided, `meta.mean.mean` will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.

`meta.mean.sd` a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sds for the normal hyperpriors on the mean hyperparameters of regression coefficients. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 10s.

`meta.sd.mean` a scalar or a vector whose dimension is equal to the number of regression coefficients giving the means for the half-normal hyperpriors on the sd hyperparameters of regression coefficients. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 0s.

`meta.sd.sd` a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sds for the half-normal hyperpriors on the sd hyperparameters of regression coefficients. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 1s.

`disp.mean` a scalar or a vector whose dimension is equal to the number of data sets (including the current data) giving the location parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 0s.

`disp.sd` a scalar or a vector whose dimension is equal to the number of data sets (including the current data) giving the scale parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 10s.

`iter_warmup` number of warmup iterations to run per chain. Defaults to 1000. See the argument `iter_warmup` in `sample()` method in cmdstanr package.

`iter_sampling` number of post-warmup iterations to run per chain. Defaults to 1000. See the argument `iter_sampling` in `sample()` method in cmdstanr package.

`chains` number of Markov chains to run. Defaults to 4. See the argument `chains` in `sample()` method in cmdstanr package.

`...` arguments passed to `sample()` method in cmdstanr package (e.g., `seed`, `refresh`, `init`).

Details

The Bayesian hierarchical model (BHM) assumes that the regression coefficients for the historical and current data are different, but are correlated through a common distribution, whose hyperparameters (i.e., mean and standard deviation (sd) (the covariance matrix is assumed to have a diagonal structure)) are treated as random. The number of regression coefficients for the current data is assumed to be the same as that for the historical data.

The hyperpriors on the mean and the sd hyperparameters are independent normal and independent half-normal distributions, respectively. The priors on the dispersion parameters (if applicable) for the current and historical data sets are independent half-normal distributions.

Value

The function returns an object of class `draws_df` giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  data_list = list(currdata = actg019, histdata = actg036)
  glm.bhm(
    formula = outcome ~ scale(age) + race + treatment + scale(cd4),
    family = binomial('logit'),
    data.list = data_list,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

`glm.commensurate` *Posterior of commensurate prior (CP)*

Description

Sample from the posterior distribution of a GLM using the commensurate prior (CP) by Hobbs et al. (2011) doi:10.1111/j.1541-0420.2011.01564.x.

Usage

```
glm.commensurate(
  formula,
  family,
  data.list,
  offset.list = NULL,
  beta0.mean = NULL,
  beta0.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  p.spike = 0.1,
  spike.mean = 200,
  spike.sd = 0.1,
  slab.mean = 0,
  slab.sd = 5,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
```

```
  ...
)
```

Arguments

formula	a two-sided formula giving the relationship between the response variable and covariates
family	an object of class family. See ?stats::family
data.list	a list of data.frames. The first element in the list is the current data, and the rest are the historical data sets.
offset.list	a list of vectors giving the offsets for each data. The length of offset.list is equal to the length of data.list. The length of each element of offset.list is equal to the number of rows in the corresponding element of data.list. Defaults to a list of vectors of 0s.
beta0.mean	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the mean parameters for the prior on the historical data regression coefficients. If a scalar is provided, beta0.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
beta0.sd	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sd parameters for the prior on the historical data regression coefficients. If a scalar is provided, same as for beta0.mean. Defaults to a vector of 10s.
disp.mean	a scalar or a vector whose dimension is equal to the number of data sets (including the current data) giving the location parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for beta0.mean. Defaults to a vector of 0s.
disp.sd	a scalar or a vector whose dimension is equal to the number of data sets (including the current data) giving the scale parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for beta0.mean. Defaults to a vector of 10s.
p.spike	a scalar between 0 and 1 giving the probability of the spike component in spike-and-slab prior on commensurability parameter τ. Defaults to 0.1.
spike.mean	a scalar giving the location parameter for the half-normal prior (spike component) on τ. Defaults to 200.
spike.sd	a scalar giving the scale parameter for the half-normal prior (spike component) on τ. Defaults to 0.1.
slab.mean	a scalar giving the location parameter for the half-normal prior (slab component) on τ. Defaults to 0.
slab.sd	a scalar giving the scale parameter for the half-normal prior (slab component) on τ. Defaults to 5.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.

chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g., seed, refresh, init).

Details

The commensurate prior (CP) assumes that the regression coefficients for the current data conditional on those for the historical data are independent normal distributions with mean equal to the corresponding regression coefficients for the historical data and variance equal to the inverse of the corresponding elements of a vector of precision parameters (referred to as the commensurability parameter τ). We regard τ as random and elicit a spike-and-slab prior, which is specified as a mixture of two half-normal priors, on τ.

The number of current data regression coefficients is assumed to be the same as that of historical data regression coefficients. The priors on the dispersion parameters (if applicable) for the current and historical data sets are independent half-normal distributions.

Value

The function returns an object of class draws_df giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

References

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047–1056.

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  data_list = list(currdata = actg019, histdata = actg036)
  glm.commensurate(
    formula = cd4 ~ treatment + age + race,
    family = poisson(), data.list = data_list,
    p.spike = 0.1,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

glm.leap *Posterior of latent exchangeability prior (LEAP)*

Description

Sample from the posterior distribution of a GLM using the latent exchangeability prior (LEAP) by Alt et al. (2023).

Usage

```
glm.leap(
  formula,
  family,
  data.list,
  K = 2,
  prob.conc = NULL,
  offset.list = NULL,
  beta.mean = NULL,
  beta.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

formula	a two-sided formula giving the relationship between the response variable and covariates.
family	an object of class family. See ?stats::family.
data.list	a list of data.frames. The first element in the list is the current data, and the rest are the historical data sets. For LEAP implementation, all historical data sets will be stacked into one historical data set.
K	the desired number of classes to identify. Defaults to 2.
prob.conc	a scalar or a vector of length K giving the concentration parameters for Dirichlet prior. If length == 2, a Beta(prob.conc[1], prob.conc[2]) prior is used. If a scalar is provided, prob.conc will be a vector of repeated elements of the given scalar. Defaults to a vector of 1s.
offset.list	a list of matrices giving the offset for current data followed by historical data. For each matrix, the number of rows corresponds to observations and columns correspond to classes. Defaults to a list of matrices of 0s. Note that the first element of offset.list (corresponding to the offset for current data) should be a matrix of repeated columns if offset.list is not NULL.

beta.mean	a scalar or a p x K matrix of mean parameters for initial prior on regression coefficients, where p is the number of regression coefficients (including intercept). If a scalar is provided, beta.mean will be a matrix of repeated elements of the given scalar. Defaults to a matrix of 0s.
beta.sd	a scalar or a p x K matrix of sd parameters for the initial prior on regression coefficients, where p is the number of regression coefficients (including intercept). If a scalar is provided, same as for beta.mean. Defaults to a matrix of 10s.
disp.mean	a scalar or a vector whose dimension is equal to the number of classes (K) giving the location parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, disp.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
disp.sd	a scalar or a vector whose dimension is equal to the number of classes (K) giving the scale parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for disp.mean. Defaults to a vector of 10s.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g., seed, refresh, init).

Details

The latent exchangeability prior (LEAP) discounts the historical data by identifying the most relevant individuals from the historical data. It is equivalent to a prior induced by the posterior of a finite mixture model for the historical data set.

Value

The function returns an object of class draws_df giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

References

Alt, E. M., Chang, X., Jiang, X., Liu, Q., Mo, M., Xia, H. M., and Ibrahim, J. G. (2023). LEAP: The latent exchangeability prior for borrowing information from historical data. arXiv preprint.

Examples

```
data(actg019)
data(actg036)
# take subset for speed purposes
actg019 = actg019[1:100, ]
actg036 = actg036[1:50, ]
if (instantiate::stan_cmdstan_exists()) {
  glm.leap(
```

```
    formula = outcome ~ scale(age) + race + treatment + scale(cd4),
    family = binomial('logit'),
    data.list = list(actg019, actg036),
    K = 2,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

`glm.logml.commensurate` *Log marginal likelihood of a GLM under commensurate prior (CP)*

Description

Uses Markov chain Monte Carlo (MCMC) and bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the commensurate prior (CP).

The arguments related to MCMC sampling are utilized to draw samples from the commensurate prior. These samples are then used to compute the logarithm of the normalizing constant of the commensurate prior using historical data sets.

Usage

```
glm.logml.commensurate(
  post.samples,
  bridge.args = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`post.samples`	output from `glm.commensurate()` giving posterior samples of a GLM under the commensurate prior (CP), with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
`bridge.args`	a `list` giving arguments (other than `samples`, `log_posterior`, `data`, `lb`, and `ub`) to pass onto `bridgesampling::bridge_sampler()`.
`iter_warmup`	number of warmup iterations to run per chain. Defaults to 1000. See the argument `iter_warmup` in `sample()` method in cmdstanr package.
`iter_sampling`	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument `iter_sampling` in `sample()` method in cmdstanr package.
`chains`	number of Markov chains to run. Defaults to 4. See the argument `chains` in `sample()` method in cmdstanr package.
`...`	arguments passed to `sample()` method in cmdstanr package (e.g., `seed`, `refresh`, `init`).

Value

The function returns a `list` with the following objects

model "Commensurate"

logml the estimated logarithm of the marginal likelihood

bs an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the commensurate prior (CP) using all data sets

bs.hist an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the CP using historical data sets

min_ess_bulk the minimum estimated bulk effective sample size of the MCMC sampling

max_Rhat the maximum Rhat

References

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047–1056.

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019   = actg019[1:100, ]
  actg036   = actg036[1:50, ]
  formula   = cd4 ~ treatment + age + race
  family    = poisson()
  data_list = list(currdata = actg019, histdata = actg036)
  d.cp      = glm.commensurate(
    formula = formula,
    family = family,
    data.list = data_list,
    p.spike = 0.1,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
  glm.logml.commensurate(
    post.samples = d.cp,
    bridge.args = list(silent = TRUE),
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

`glm.logml.leap` *Log marginal likelihood of a GLM under latent exchangeability prior (LEAP)*

Description

Uses Markov chain Monte Carlo (MCMC) and bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the latent exchangeability prior (LEAP).

The arguments related to MCMC sampling are utilized to draw samples from the LEAP. These samples are then used to compute the logarithm of the normalizing constant of the LEAP using historical data sets.

Usage

```
glm.logml.leap(
  post.samples,
  bridge.args = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`post.samples`	output from `glm.leap()` giving posterior samples of a GLM under the latent exchangeability prior (LEAP), with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
`bridge.args`	a `list` giving arguments (other than `samples`, `log_posterior`, `data`, `lb`, and `ub`) to pass onto `bridgesampling::bridge_sampler()`.
`iter_warmup`	number of warmup iterations to run per chain. Defaults to 1000. See the argument `iter_warmup` in `sample()` method in cmdstanr package.
`iter_sampling`	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument `iter_sampling` in `sample()` method in cmdstanr package.
`chains`	number of Markov chains to run. Defaults to 4. See the argument `chains` in `sample()` method in cmdstanr package.
`...`	arguments passed to `sample()` method in cmdstanr package (e.g., `seed`, `refresh`, `init`).

Value

The function returns a `list` with the following objects

model "LEAP"

logml the estimated logarithm of the marginal likelihood

bs an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the latent exchangeability prior (LEAP) using all data sets

bs.hist an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the LEAP using historical data sets

min_ess_bulk the minimum estimated bulk effective sample size of the MCMC sampling

max_Rhat the maximum Rhat

References

Alt, E. M., Chang, X., Jiang, X., Liu, Q., Mo, M., Xia, H. M., and Ibrahim, J. G. (2023). LEAP: The latent exchangeability prior for borrowing information from historical data. arXiv preprint.

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019   = actg019[1:100, ]
  actg036   = actg036[1:50, ]
  formula   = outcome ~ scale(age) + race + treatment + scale(cd4)
  family    = binomial('logit')
  data_list = list(currdata = actg019, histdata = actg036)
  d.leap    = glm.leap(
    formula = formula,
    family = family,
    data.list = data_list,
    K = 2,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
  glm.logml.leap(
    post.samples = d.leap,
    bridge.args = list(silent = TRUE),
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

`glm.logml.map` *Log marginal likelihood of a GLM under meta-analytic predictive (MAP) prior*

Description

Uses Markov chain Monte Carlo (MCMC) and bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the meta-analytic predictive (MAP) prior. The MAP prior is equivalent to the prior induced by the Bayesian hierarchical model (BHM).

The arguments related to MCMC sampling are utilized to draw samples from the MAP prior. These samples are then used to compute the logarithm of the normalizing constant of the BHM using only historical data sets.

Usage

```
glm.logml.map(
  post.samples,
  bridge.args = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`post.samples`	output from `glm.bhm()` giving posterior samples of a GLM under the Bayesian hierarchical model (BHM), with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
`bridge.args`	a `list` giving arguments (other than `samples`, `log_posterior`, `data`, `lb`, and `ub`) to pass onto `bridgesampling::bridge_sampler()`.
`iter_warmup`	number of warmup iterations to run per chain. Defaults to 1000. See the argument `iter_warmup` in `sample()` method in cmdstanr package.
`iter_sampling`	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument `iter_sampling` in `sample()` method in cmdstanr package.
`chains`	number of Markov chains to run. Defaults to 4. See the argument `chains` in `sample()` method in cmdstanr package.
`...`	arguments passed to `sample()` method in cmdstanr package (e.g., `seed`, `refresh`, `init`).

Value

The function returns a `list` with the following objects

model "MAP"

logml the estimated logarithm of the marginal likelihood of the meta-analytic predictive (MAP) prior

bs an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the Bayesian hierarchical model (BHM) using all data sets

bs.hist an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the BHM using historical data sets

min_ess_bulk the minimum estimated bulk effective sample size of the MCMC sampling

max_Rhat the maximum Rhat

References

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  formula = outcome ~ scale(age) + race + treatment + scale(cd4)
  family = binomial('logit')
  data_list = list(currdata = actg019, histdata = actg036)
  d.bhm = glm.bhm(
    formula = formula,
    family = family,
    data.list = data_list,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
  glm.logml.map(
    post.samples = d.bhm,
    bridge.args = list(silent = TRUE),
    chains = 1, iter_warmup = 1000, iter_sampling = 2000
  )
}
```

---

glm.logml.napp *Log marginal likelihood of a GLM under normalized asymptotic power prior (NAPP)*

Description

Uses bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the normalized asymptotic power prior (NAPP).

Usage

```
glm.logml.napp(post.samples, bridge.args = NULL)
```

Arguments

post.samples	output from glm.napp() giving posterior samples of a GLM under the normalized asymptotic power prior (NAPP), with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
bridge.args	a list giving arguments (other than samples, log_posterior, data, lb, and ub) to pass onto bridgesampling::bridge_sampler().

Value

The function returns a list with the following objects

model "NAPP"

logml the estimated logarithm of the marginal likelihood

bs an object of class bridge or bridge_list containing the output from using bridgesampling::bridge_sampler() to compute the logarithm of the marginal likelihood of the normalized asymptotic power prior (NAPP)

References

Ibrahim, J. G., Chen, M., Gwon, Y., and Chen, F. (2015). The power prior: Theory and applications. Statistics in Medicine, 34(28), 3724–3749.

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  data_list = list(currdata = actg019, histdata = actg036)
  formula = cd4 ~ treatment + age + race
  family = poisson('log')
  d.napp = glm.napp(
    formula = formula, family = family,
    data.list = data_list,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
  glm.logml.napp(
    post.samples = d.napp,
    bridge.args = list(silent = TRUE)
  )
}
```

`glm.logml.npp` *Log marginal likelihood of a GLM under normalized power prior (NPP)*

Description

Uses bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the normalized power prior (NPP).

Usage

```
glm.logml.npp(post.samples, bridge.args = NULL)
```

Arguments

`post.samples`	output from `glm.npp()` giving posterior samples of a GLM under the normalized power prior (NPP), with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
`bridge.args`	a `list` giving arguments (other than samples, log_posterior, data, lb, ub) to pass onto `bridgesampling::bridge_sampler()`.

Value

The function returns a `list` with the following objects

model "NPP"

logml the estimated logarithm of the marginal likelihood

bs an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the marginal likelihood of the normalized power prior (NPP)

References

Duan, Y., Ye, K., and Smith, E. P. (2005). Evaluating water quality using power priors to incorporate historical information. Environmetrics, 17(1), 95–106.

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if(requireNamespace("parallel")){
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]

  library(parallel)
```

```
  ncores    = 2
  data.list = list(data = actg019, histdata = actg036)
  formula   = cd4 ~ treatment + age + race
  family    = poisson()
  a0        = seq(0, 1, length.out = 11)
  if (instantiate::stan_cmdstan_exists()) {
    ## call created function
    ## wrapper to obtain log normalizing constant in parallel package
    logncfun = function(a0, ...){
      hdbayes::glm.npp.lognc(
        formula = formula, family = family, a0 = a0, histdata = data.list[[2]],
        ...
      )
    }

    cl = makeCluster(ncores)
    clusterSetRNGStream(cl, 123)
    clusterExport(cl, varlist = c('formula', 'family', 'data.list'))
    a0.lognc = parLapply(
      cl = cl, X = a0, fun = logncfun, iter_warmup = 500,
      iter_sampling = 1000, chains = 1, refresh = 0
    )
    stopCluster(cl)
    a0.lognc = data.frame( do.call(rbind, a0.lognc) )

    ## sample from normalized power prior
    d.npp = glm.npp(
      formula = formula,
      family = family,
      data.list = data.list,
      a0.lognc = a0.lognc$a0,
      lognc = matrix(a0.lognc$lognc, ncol = 1),
      chains = 1, iter_warmup = 500, iter_sampling = 1000,
      refresh = 0
    )
    glm.logml.npp(
      post.samples = d.npp,
      bridge.args = list(silent = TRUE)
    )
  }
}
```

`glm.logml.post` *Log marginal likelihood of a GLM under a normal/half-normal prior*

Description

Uses bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the normal/half-normal prior.

Usage

```
glm.logml.post(post.samples, bridge.args = NULL)
```

Arguments

post.samples	output from glm.post() giving posterior samples of a GLM under the normal/half-normal prior, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
bridge.args	a list giving arguments (other than samples, log_posterior, data, lb, and ub) to pass onto bridgesampling::bridge_sampler().

Value

The function returns a list with the following objects

model "Normal/Half-Normal"

logml the estimated logarithm of the marginal likelihood

bs an object of class bridge or bridge_list containing the output from using bridgesampling::bridge_sampler() to compute the logarithm of the marginal likelihood of the normal/half-normal prior

References

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  actg019 = actg019[1:100, ]
  data.list = list(currdata = actg019)
  formula = cd4 ~ treatment + age + race
  family = poisson('log')
  d.post = glm.post(
    formula = formula, family = family,
    data.list = data.list,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
  glm.logml.post(
    post.samples = d.post,
    bridge.args = list(silent = TRUE)
  )
}
```

`glm.logml.pp` *Log marginal likelihood of a GLM under power prior (PP)*

Description

Uses Markov chain Monte Carlo (MCMC) and bridge sampling to estimate the logarithm of the marginal likelihood of a GLM under the power prior (PP).

The arguments related to MCMC sampling are utilized to draw samples from the power prior (PP). These samples are then used to compute the logarithm of the normalizing constant of the PP using only historical data sets.

Usage

```
glm.logml.pp(
  post.samples,
  bridge.args = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`post.samples`	output from `glm.pp()` giving posterior samples of a GLM under the power prior (PP), with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.
`bridge.args`	a `list` giving arguments (other than `samples`, `log_posterior`, `data`, `lb`, and `ub`) to pass onto `bridgesampling::bridge_sampler()`.
`iter_warmup`	number of warmup iterations to run per chain. Defaults to 1000. See the argument `iter_warmup` in `sample()` method in cmdstanr package.
`iter_sampling`	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument `iter_sampling` in `sample()` method in cmdstanr package.
`chains`	number of Markov chains to run. Defaults to 4. See the argument `chains` in `sample()` method in cmdstanr package.
`...`	arguments passed to `sample()` method in cmdstanr package (e.g., `seed`, `refresh`, `init`).

Value

If all of the power prior parameters (a_0's) are equal to zero, or if the posterior samples are obtained from using only one data set (the current data), then the function will return the same result as the output from `glm.logml.post()`.

If at least one of the power prior parameters (a_0's) is non-zero, the function will return a `list` with the following objects

model "PP"

logml the estimated logarithm of the marginal likelihood

bs an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the power prior (PP) using all data sets

bs.hist an object of class `bridge` or `bridge_list` containing the output from using `bridgesampling::bridge_sampler()` to compute the logarithm of the normalizing constant of the PP using historical data sets

min_ess_bulk the minimum estimated bulk effective sample size of the MCMC sampling

max_Rhat the maximum Rhat

References

Chen, M.-H. and Ibrahim, J. G. (2000). Power prior distributions for Regression Models. Statistical Science, 15(1).

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  data_list = list(currdata = actg019, histdata = actg036)
  formula = cd4 ~ treatment + age + race
  family = poisson('log')
  a0 = 0.5
  d.pp = glm.pp(
    formula = formula, family = family,
    data.list = data_list,
    a0.vals = a0,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
  glm.logml.pp(
    post.samples = d.pp,
    bridge.args = list(silent = TRUE),
    chains = 1, iter_warmup = 1000, iter_sampling = 2000
  )
}
```

`glm.napp` *Posterior of normalized asymptotic power prior (NAPP)*

Description

Sample from the posterior distribution of a GLM using the normalized asymptotic power prior (NAPP) by Ibrahim et al. (2015) doi:10.1002/sim.6728.

Usage

```
glm.napp(
  formula,
  family,
  data.list,
  offset.list = NULL,
  a0.shape1 = 1,
  a0.shape2 = 1,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

formula	a two-sided formula giving the relationship between the response variable and covariates.
family	an object of class family. See ?stats::family.
data.list	a list of data.frames. The first element in the list is the current data, and the rest are the historical datasets.
offset.list	a list of vectors giving the offsets for each data. The length of offset.list is equal to the length of data.list. The length of each element of offset.list is equal to the number of rows in the corresponding element of data.list. Defaults to a list of vectors of 0s.
a0.shape1	first shape parameter for the i.i.d. beta prior on a0 vector. When a0.shape1 == 1 and a0.shape2 == 1, a uniform prior is used.
a0.shape2	second shape parameter for the i.i.d. beta prior on a0 vector. When a0.shape1 == 1 and a0.shape2 == 1, a uniform prior is used.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g., seed, refresh, init).

Details

The normalized asymptotic power prior (NAPP) assumes that the regression coefficients and logarithm of the dispersion parameter are a multivariate normal distribution with mean equal to the maximum likelihood estimate of the historical data and covariance matrix equal to a_0^{-1} multiplied by the inverse Fisher information matrix of the historical data, where a_0 is the power prior parameter (treated as random).

Value

The function returns an object of class `draws_df` giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

References

Ibrahim, J. G., Chen, M., Gwon, Y., and Chen, F. (2015). The power prior: Theory and applications. Statistics in Medicine, 34(28), 3724–3749.

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  data_list = list(currdata = actg019, histdata = actg036)
  glm.napp(
    formula = cd4 ~ treatment + age + race,
    family = poisson('log'),
    data.list = data_list,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

`glm.npp` *Posterior of normalized power prior (NPP)*

Description

Sample from the posterior distribution of a GLM using the normalized power prior (NPP) by Duan et al. (2006) doi:10.1002/env.752.

Usage

```
glm.npp(
  formula,
  family,
  data.list,
```

```
  a0.lognc,
  lognc,
  offset.list = NULL,
  beta.mean = NULL,
  beta.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  a0.shape1 = 1,
  a0.shape2 = 1,
  a0.lower = NULL,
  a0.upper = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`formula`	a two-sided formula giving the relationship between the response variable and covariates.
`family`	an object of class `family`. See `?stats::family`.
`data.list`	a list of `data.frames`. The first element in the list is the current data, and the rest are the historical data sets.
`a0.lognc`	a vector giving values of the power prior parameter for which the logarithm of the normalizing constant has been evaluated.
`lognc`	an S by T matrix where S is the length of `a0.lognc`, T is the number of historical data sets, and the j-th column, j = 1, ..., T, is a vector giving the logarithm of the normalizing constant (as estimated by `glm.npp.lognc()` for `a0.lognc` using the j-th historical data set.
`offset.list`	a list of vectors giving the offsets for each data. The length of `offset.list` is equal to the length of `data.list`. The length of each element of `offset.list` is equal to the number of rows in the corresponding element of `data.list`. Defaults to a list of vectors of 0s.
`beta.mean`	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the mean parameters for the initial prior on regression coefficients. If a scalar is provided, `beta.mean` will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
`beta.sd`	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sd parameters for the initial prior on regression coefficients. If a scalar is provided, same as for `beta.mean`. Defaults to a vector of 10s.
`disp.mean`	location parameter for the half-normal prior on dispersion parameter. Defaults to 0.
`disp.sd`	scale parameter for the half-normal prior on dispersion parameter. Defaults to 10.

a0.shape1	first shape parameter for the i.i.d. beta prior on a_0 vector. When a0.shape1 == 1 and a0.shape2 == 1, a uniform prior is used.
a0.shape2	second shape parameter for the i.i.d. beta prior on a_0 vector. When a0.shape1 == 1 and a0.shape2 == 1, a uniform prior is used.
a0.lower	a scalar or a vector whose dimension is equal to the number of historical data sets giving the lower bounds for each element of the a_0 vector. If a scalar is provided, a0.lower will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
a0.upper	a scalar or a vector whose dimension is equal to the number of historical data sets giving the upper bounds for each element of the a_0 vector. If a scalar is provided, same as for a0.lower. Defaults to a vector of 1s.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g., seed, refresh, init).

Details

Before using this function, users must estimate the logarithm of the normalizing constant across a range of different values for the power prior parameter (a_0), possibly smoothing techniques over a find grid. The power prior parameters (a_0's) are treated as random with independent beta priors. The initial priors on the regression coefficients are independent normal priors. The current and historical data sets are assumed to have a common dispersion parameter with a half-normal prior (if applicable). For normal linear models, the exact normalizing constants for NPP can be computed. See the implementation in lm.npp().

Value

The function returns an object of class draws_df giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

References

Duan, Y., Ye, K., and Smith, E. P. (2005). Evaluating water quality using power priors to incorporate historical information. Environmetrics, 17(1), 95–106.

See Also

glm.npp.lognc()

Examples

```
if(requireNamespace("parallel")){
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]

  library(parallel)
  ncores    = 2
  data.list = list(data = actg019, histdata = actg036)
  formula   = cd4 ~ treatment + age + race
  family    = poisson()
  a0        = seq(0, 1, length.out = 11)
  if (instantiate::stan_cmdstan_exists()) {
    ## call created function
    ## wrapper to obtain log normalizing constant in parallel package
    logncfun = function(a0, ...){
      hdbayes::glm.npp.lognc(
        formula = formula, family = family, a0 = a0, histdata = data.list[[2]],
        ...
      )
    }

    cl = makeCluster(ncores)
    clusterSetRNGStream(cl, 123)
    clusterExport(cl, varlist = c('formula', 'family', 'data.list'))
    a0.lognc = parLapply(
      cl = cl, X = a0, fun = logncfun, iter_warmup = 500,
      iter_sampling = 1000, chains = 1, refresh = 0
    )
    stopCluster(cl)
    a0.lognc = data.frame( do.call(rbind, a0.lognc) )

    ## sample from normalized power prior
    glm.npp(
      formula = formula,
      family = family,
      data.list = data.list,
      a0.lognc = a0.lognc$a0,
      lognc = matrix(a0.lognc$lognc, ncol = 1),
      chains = 1, iter_warmup = 500, iter_sampling = 1000,
      refresh = 0
    )
  }
}
```

`glm.npp.lognc` *Estimate the logarithm of the normalizing constant for normalized power prior (NPP) for one data set*

Description

Uses Markov chain Monte Carlo (MCMC) and bridge sampling to estimate the logarithm of the normalizing constant for the NPP for a fixed value of the power prior parameter $a_0 \in (0, 1)$ for one data set. The initial priors are independent normal priors on the regression coefficients and a half-normal prior on the dispersion parameter (if applicable).

Usage

```
glm.npp.lognc(
  formula,
  family,
  histdata,
  a0,
  offset0 = NULL,
  beta.mean = NULL,
  beta.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  bridge.args = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`formula`	a two-sided formula giving the relationship between the response variable and covariates.
`family`	an object of class `family`. See `?stats::family`.
`histdata`	a `data.frame` giving the historical data.
`a0`	the power prior parameter (a scalar between 0 and 1).
`offset0`	vector whose dimension is equal to the rows of the historical data set giving an offset for the historical data. Defaults to a vector of 0s.
`beta.mean`	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the mean parameters for the normal initial prior on regression coefficients given the dispersion parameter. If a scalar is provided, beta.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
`beta.sd`	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sd parameters for the initial prior on regression coefficients. The sd used is `sqrt(dispersion) * beta.sd`. If a scalar is provided, same as for beta.mean. Defaults to a vector of 10s.
`disp.mean`	location parameter for the half-normal prior on dispersion parameter. Defaults to 0.

disp.sd	scale parameter for the half-normal prior on dispersion parameter. Defaults to 10.
bridge.args	a list giving arguments (other than samples, log_posterior, data, lb, ub) to pass onto bridgesampling::bridge_sampler().
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g. seed, refresh, init).

Value

The function returns a vector giving the value of a0, the estimated logarithm of the normalizing constant, the minimum estimated bulk effective sample size of the MCMC sampling, and the maximum Rhat.

References

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg036)
  ## take subset for speed purposes
  actg036 = actg036[1:50, ]
  glm.npp.lognc(
    cd4 ~ treatment + age + race,
    family = poisson(), histdata = actg036, a0 = 0.5,
    chains = 1, iter_warmup = 500, iter_sampling = 5000
  )
}
```

---

glm.post *Posterior of a normal/half-normal prior*

Description

Sample from the posterior distribution of a GLM using a normal/half-normal prior.

Usage

```
glm.post(
  formula,
  family,
  data.list,
  offset.list = NULL,
  beta.mean = NULL,
  beta.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

formula	a two-sided formula giving the relationship between the response variable and covariates.
family	an object of class family. See ?stats::family.
data.list	a list consisting of one data.frame giving the current data. If data.list has more than one data.frame, only the first element will be used as the current data.
offset.list	a list consisting of one vector giving the offset for the current data. The length of the vector is equal to the number of rows in the current data. The vector has all values set to 0 by default. If offset.list has more than one vector, same as for data.list.
beta.mean	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the mean parameters for the normal prior on regression coefficients. If a scalar is provided, beta.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
beta.sd	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sd parameters for the normal prior on regression coefficients. If a scalar is provided, same as for beta.mean. Defaults to a vector of 10s.
disp.mean	location parameter for the half-normal prior on dispersion parameter. Defaults to 0. If disp.mean is a vector with length > 1, only the first element will be used as disp.mean.
disp.sd	scale parameter for the half-normal prior on dispersion parameter. Defaults to 10. If disp.sd is a vector with length > 1, same as for disp.mean.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.

... arguments passed to `sample()` method in cmdstanr package (e.g., `seed`, `refresh`, `init`).

Details

The priors on the regression coefficients are independent normal distributions. When the normal priors are elicited with large variances, the prior is also referred to as the reference or vague prior. The dispersion parameter is assumed to be independent of the regression coefficients with a half-normal prior (if applicable).

Value

The function returns an object of class `draws_df` giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  data.list = list(currdata = actg019)
  glm.post(
    formula = cd4 ~ treatment + age + race,
    family = poisson('log'),
    data.list = data.list,
    beta.sd   = 10,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

---

glm.pp *Posterior of power prior (PP) with fixed a_0*

Description

Sample from the posterior distribution of a GLM using the power prior (PP) by Ibrahim and Chen (2000) doi:10.1214/ss/1009212673.

Usage

```
glm.pp(
  formula,
  family,
  data.list,
  a0.vals,
  offset.list = NULL,
  beta.mean = NULL,
  beta.sd = NULL,
```

```
  disp.mean = NULL,
  disp.sd = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

formula	a two-sided formula giving the relationship between the response variable and covariates.
family	an object of class family. See ?stats::family.
data.list	a list of data.frames. The first element in the list is the current data, and the rest are the historical data sets.
a0.vals	a scalar between 0 and 1 or a vector whose dimension is equal to the number of historical data sets giving the (fixed) power prior parameter for each historical data set. Each element of vector should be between 0 and 1. If a scalar is provided, same as for beta.mean.
offset.list	a list of vectors giving the offsets for each data. The length of offset.list is equal to the length of data.list. The length of each element of offset.list is equal to the number of rows in the corresponding element of data.list. Defaults to a list of vectors of 0s.
beta.mean	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the mean parameters for the initial prior on regression coefficients. If a scalar is provided, beta.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
beta.sd	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sd parameters for the initial prior on regression coefficients. If a scalar is provided, same as for beta.mean. Defaults to a vector of 10s.
disp.mean	location parameter for the half-normal prior on dispersion parameter. Defaults to 0.
disp.sd	scale parameter for the half-normal prior on dispersion parameter. Defaults to 10.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g., seed, refresh, init).

Details

The power prior parameters (a_0's) are treated as fixed. The initial priors on the regression coefficients are independent normal priors. The current and historical data sets are assumed to have a common dispersion parameter with a half-normal prior (if applicable).

Value

The function returns an object of class `draws_df` giving posterior samples, with an attribute called 'data' which includes the list of variables specified in the data block of the Stan program.

References

Chen, M.-H. and Ibrahim, J. G. (2000). Power prior distributions for Regression Models. Statistical Science, 15(1).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  ## take subset for speed purposes
  actg019 = actg019[1:100, ]
  actg036 = actg036[1:50, ]
  data_list = list(currdata = actg019, histdata = actg036)
  glm.pp(
    formula = cd4 ~ treatment + age + race,
    family = poisson('log'),
    data.list = data_list,
    a0.vals = 0.5,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

`glm.rmap` *Posterior of robust meta-analytic predictive prior (RMAP)*

Description

Sample from the posterior distribution of a GLM using the robust meta-analytic predictive prior (RMAP) by Schmidli et al. (2014) doi:10.1111/biom.12242.

Usage

```
glm.rmap(
  formula,
  family,
  data.list,
  offset.list = NULL,
```

```
  w = 0.1,
  meta.mean.mean = NULL,
  meta.mean.sd = NULL,
  meta.sd.mean = NULL,
  meta.sd.sd = NULL,
  disp.mean = NULL,
  disp.sd = NULL,
  norm.vague.mean = NULL,
  norm.vague.sd = NULL,
  bridge.args = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

formula	a two-sided formula giving the relationship between the response variable and covariates.
family	an object of class family. See ?stats::family.
data.list	a list of data.frames. The first element in the list is the current data, and the rest are the historical data sets.
offset.list	a list of vectors giving the offsets for each data. The length of offset.list is equal to the length of data.list. The length of each element of offset.list is equal to the number of rows in the corresponding element of data.list. Defaults to a list of vectors of 0s.
w	a scalar between 0 and 1 giving how much weight to put on the historical data. Defaults to 0.1.
meta.mean.mean	same as meta.mean.mean in glm.bhm(). It is a scalar or a vector whose dimension is equal to the number of regression coefficients giving the means for the normal hyperpriors on the mean hyperparameters of regression coefficients in Bayesian hierarchical model (BHM). If a scalar is provided, meta.mean.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
meta.mean.sd	same as meta.mean.sd in glm.bhm(). It is a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sds for the normal hyperpriors on the mean hyperparameters of regression coefficients in BHM. If a scalar is provided, same as for meta.mean.mean. Defaults to a vector of 10s.
meta.sd.mean	same as meta.sd.mean in glm.bhm(). It is a scalar or a vector whose dimension is equal to the number of regression coefficients giving the means for the half-normal hyperpriors on the sd hyperparameters of regression coefficients in BHM. If a scalar is provided, same as for meta.mean.mean. Defaults to a vector of 0s.
meta.sd.sd	same as meta.sd.sd in glm.bhm(). It is a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sds for the half-normal

hyperpriors on the sd hyperparameters of regression coefficients in BHM. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 1s.

`disp.mean` a scalar or a vector whose dimension is equal to the number of data sets (including the current data) giving the location parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 0s.

`disp.sd` a scalar or a vector whose dimension is equal to the number of data sets (including the current data) giving the scale parameters for the half-normal priors on the dispersion parameters. If a scalar is provided, same as for `meta.mean.mean`. Defaults to a vector of 10s.

`norm.vague.mean` same as `beta.mean` in `glm.post()`. It is a scalar or a vector whose dimension is equal to the number of regression coefficients giving the means for the vague normal prior on regression coefficients. If a scalar is provided, `norm.vague.mean` will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.

`norm.vague.sd` same as `beta.sd` in `glm.post()`. It is a scalar or a vector whose dimension is equal to the number of regression coefficients giving the sds for the vague normal prior on regression coefficients. If a scalar is provided, same as for `norm.vague.mean`. Defaults to a vector of 10s.

`bridge.args` a `list` giving arguments (other than samples, log_posterior, data, lb, ub) to pass onto `bridgesampling::bridge_sampler()`.

`iter_warmup` number of warmup iterations to run per chain. Defaults to 1000. See the argument `iter_warmup` in `sample()` method in cmdstanr package.

`iter_sampling` number of post-warmup iterations to run per chain. Defaults to 1000. See the argument `iter_sampling` in `sample()` method in cmdstanr package.

`chains` number of Markov chains to run. Defaults to 4. See the argument `chains` in `sample()` method in cmdstanr package.

`...` arguments passed to `sample()` method in cmdstanr package (e.g. seed, refresh, init).

Details

The robust meta-analytic predictive prior (RMAP) is a two-part mixture prior consisting of a meta-analytic predictive (MAP) prior (the prior induced by Bayesian hierarchical model (BHM)) and a vague (i.e., non-informative) prior (specifically, the normal/half-normal prior with large variances). Although Schmidli et al. (2014) recommends to use a finite mixture of conjugate priors to approximate the BHM, it can be difficult and time-consuming to come up with an appropriate approximation.

Instead, the approach taken by hdbayes is to use the marginal likelihood of the MAP and vague priors. Specifically, note that the posterior distribution of a GLM under RMAP is also a two-part mixture distribution. The updated mixture weight for posterior density under the MAP prior is

$$\widetilde{w} = \frac{w Z_I(D, D_0)}{w Z_I(D, D_0) + (1 - w) Z_V(D)},$$

where w is the prior mixture weight for the MAP prior in RMAP, $Z_I(D, D_0)$ is the marginal likelihood of the MAP prior, and $Z_V(D)$ is the marginal likelihood of the vague prior.

Value

The function returns a `list` with the following objects

post.samples an object of class `draws_df` giving posterior samples under the robust meta-analytic predictive prior (RMAP)

post.samples.bhm an object of class `draws_df` giving posterior samples under the Bayesian hierarchical model (BHM), obtained from using `glm.bhm()`

post.samples.vague an object of class `draws_df` giving posterior samples under the vague/non-informative prior, obtained from using `glm.post()`

bs.map output from computing log marginal likelihood of the prior induced by the BHM (referred to as the meta-analytic predictive (MAP) prior) via `glm.logml.map()` function

bs.vague output from computing log marginal likelihood of the vague prior via `glm.logml.post()` function

References

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics, 70(4), 1023–1032.

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. Journal of Statistical Software, 92(10).

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019) ## current data
  data(actg036) ## historical data
  ## take subset for speed purposes
  actg019 = actg019[1:150, ]
  actg036 = actg036[1:100, ]
  data.list = list(actg019, actg036)
  glm.rmap(
    formula = outcome ~ scale(age) + race + treatment + scale(cd4),
    family = binomial('logit'),
    data.list = data.list,
    w = 0.1,
    chains = 1, iter_warmup = 1000, iter_sampling = 2000
  )
}
```

IBCSG_curr *International Breast Cancer Study Group (IBCSG) Trial VI Data*

Description

A data set from the IBCSG Trial VI investigating both the duration of adjuvant chemotherapy (3 versus 6 initial cycles of oral cyclophosphamide, methotrexate, and fluorouracil (CMF)) and the reintroduction of single courses of delayed chemotherapy in node-positive premenopausal breast cancer patients. The study results were described by IBCSG (1996) doi:10.1200/JCO.1996.14.6.1885 and Hürny et al. (1992) doi:10.1016/0959-8049(92)90399-m. This data set only includes patients above the age of 40 (i.e., age $\geq$ 40) and treats the measurements of patients' physical well-being on month 18 as the outcome. The IBCSG_hist data set includes patients from the same study but with age < 40. We can use the IBCSG_hist data as the historical data and the IBCSG_curr data as the current data.

Usage

```
IBCSG_curr
```

Format

A data frame with 488 rows and 8 variables:

phys18 outcome variable, integer scores between 0 and 100 measuring the patients' physical well-being on month 18, with a higher score indicating a better physical well-being

phys1 physical well-being scores assessed at the start of the study

n_init_cycles number of initial cycles of CMF, equal to 3 or 6

reintroduction indicator of reintroduction of chemotherapy, 0 = no reintroduction, 1 = having reintroduction

age patient age in years

country country, ANZ = New Zealand/Australia, CH = Switzerland, SWED = Sweden

nodegp indicator of number of positive nodes being greater than or equal to 4, 0 = less than 4, 1 = 4+

ER estrogen receptor (ER) status indicator, 0 = negative, 1 = positive

References

International Breast Cancer Study Group. (1996). Duration and reintroduction of adjuvant chemotherapy for node-positive premenopausal breast cancer patients. Journal of Clinical Oncology, 14(6), 1885–1894.

Hürny, C., Bernhard, J., Gelber, R. D., Coates, A., Castiglione, M., Isley, M., Dreher, D., Peterson, H., Goldhirsch, A., and Senn, H.-J. (1992). Quality of life measures for patients receiving adjuvant therapy for breast cancer: An international trial. European Journal of Cancer, 28(1), 118–124.

Chi, Y.-Y. and Ibrahim, J. G. (2005). Joint models for multivariate longitudinal and Multivariate Survival Data. Biometrics, 62(2), 432–445.

IBCSG_hist *International Breast Cancer Study Group (IBCSG) Trial VI Data*

Description

A data set from the IBCSG Trial VI investigating both the duration of adjuvant chemotherapy (3 versus 6 initial cycles of oral cyclophosphamide, methotrexate, and fluorouracil (CMF)) and the reintroduction of single courses of delayed chemotherapy in node-positive premenopausal breast cancer patients. The study results were described by IBCSG (1996) doi:10.1200/JCO.1996.14.6.1885 and Hürny et al. (1992) doi:10.1016/0959-8049(92)90399-m. This data set only includes patients under the age of 40 (i.e., age < 40) and treats the measurements of patients' physical well-being on month 18 as the outcome. The IBCSG_curr data set includes patients from the same study but with age $\geq$ 40. We can use the IBCSG_hist data as the historical data and the IBCSG_curr data as the current data.

Usage

```
IBCSG_hist
```

Format

A data frame with 103 rows and 8 variables:

phys18 outcome variable, integer scores between 0 and 100 measuring the patients' physical well-being on month 18, with a higher score indicating a better physical well-being

phys1 physical well-being scores assessed at the start of the study

n_init_cycles number of initial cycles of CMF, equal to 3 or 6

reintroduction indicator of reintroduction of chemotherapy, 0 = no reintroduction, 1 = having reintroduction

age patient age in years

country country, ANZ = New Zealand/Australia, CH = Switzerland, SWED = Sweden

nodegp indicator of number of positive nodes being greater than or equal to 4, 0 = less than 4, 1 = 4+

ER estrogen receptor (ER) status indicator, 0 = negative, 1 = positive

References

International Breast Cancer Study Group. (1996). Duration and reintroduction of adjuvant chemotherapy for node-positive premenopausal breast cancer patients. Journal of Clinical Oncology, 14(6), 1885–1894.

Hürny, C., Bernhard, J., Gelber, R. D., Coates, A., Castiglione, M., Isley, M., Dreher, D., Peterson, H., Goldhirsch, A., and Senn, H.-J. (1992). Quality of life measures for patients receiving adjuvant therapy for breast cancer: An international trial. European Journal of Cancer, 28(1), 118–124.

Chi, Y.-Y. and Ibrahim, J. G. (2005). Joint models for multivariate longitudinal and Multivariate Survival Data. Biometrics, 62(2), 432–445.

`lm.npp` *Posterior of normalized power prior (NPP) for normal linear models*

Description

Sample from the posterior distribution of a normal linear model using the NPP by Duan et al. (2006) doi:10.1002/env.752. The power prior parameters (a_0's) are treated as random with independent beta priors. The current and historical data sets are assumed to have a common dispersion parameter (σ^2) with an inverse-gamma prior. Conditional on σ^2, the initial priors on the regression coefficients are independent normal distributions with variance $\propto (\sigma^2)^{-1}$. In this case, the normalizing constant for the NPP has a closed form.

Usage

```
lm.npp(
  formula,
  data.list,
  offset.list = NULL,
  beta.mean = NULL,
  beta.sd = NULL,
  sigmasq.shape = 2.1,
  sigmasq.scale = 1.1,
  a0.shape1 = 1,
  a0.shape2 = 1,
  a0.lower = NULL,
  a0.upper = NULL,
  iter_warmup = 1000,
  iter_sampling = 1000,
  chains = 4,
  ...
)
```

Arguments

`formula`	a two-sided formula giving the relationship between the response variable and covariates.
`data.list`	a list of `data.frames`. The first element in the list is the current data, and the rest are the historical data sets.
`offset.list`	a list of vectors giving the offsets for each data. The length of offset.list is equal to the length of data.list. The length of each element of offset.list is equal to the number of rows in the corresponding element of data.list. Defaults to a list of vectors of 0s.
`beta.mean`	a scalar or a vector whose dimension is equal to the number of regression coefficients giving the mean parameters for the initial prior on regression coefficients. If a scalar is provided, beta.mean will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.

beta.sd	a scalar or a vector whose dimension is equal to the number of regression coefficients. Conditional on the variance parameter sigmasq for the outcome, beta.sd * sqrt(sigmasq) gives the sd for the initial prior on regression coefficients. If a scalar is provided, same as for beta.mean. Defaults to a vector of 10s.
sigmasq.shape	shape parameter for inverse-gamma prior on variance parameter. Defaults to 2.1.
sigmasq.scale	scale parameter for inverse-gamma prior on variance parameter. Defaults to 1.1.
a0.shape1	first shape parameter for the i.i.d. beta prior on a0 vector. When a0.shape1 == 1 and a0.shape2 == 1, a uniform prior is used.
a0.shape2	second shape parameter for the i.i.d. beta prior on a0 vector. When a0.shape1 == 1 and a0.shape2 == 1, a uniform prior is used.
a0.lower	a scalar or a vector whose dimension is equal to the number of historical data sets giving the lower bounds for each element of the a0 vector. If a scalar is provided, a0.lower will be a vector of repeated elements of the given scalar. Defaults to a vector of 0s.
a0.upper	a scalar or a vector whose dimension is equal to the number of historical data sets giving the upper bounds for each element of the a0 vector. If a scalar is provided, same as for a0.lower. Defaults to a vector of 1s.
iter_warmup	number of warmup iterations to run per chain. Defaults to 1000. See the argument iter_warmup in sample() method in cmdstanr package.
iter_sampling	number of post-warmup iterations to run per chain. Defaults to 1000. See the argument iter_sampling in sample() method in cmdstanr package.
chains	number of Markov chains to run. Defaults to 4. See the argument chains in sample() method in cmdstanr package.
...	arguments passed to sample() method in cmdstanr package (e.g. seed, refresh, init).

Value

The function returns an object of class draws_df giving posterior samples.

References

Duan, Y., Ye, K., and Smith, E. P. (2005). Evaluating water quality using power priors to incorporate historical information. Environmetrics, 17(1), 95–106.

Examples

```
if (instantiate::stan_cmdstan_exists()) {
  data(actg019)
  data(actg036)
  data_list = list(currdata = actg019, histdata = actg036)
  lm.npp(
    formula = cd4 ~ treatment + age + race,
    data.list = data_list,
    chains = 1, iter_warmup = 500, iter_sampling = 1000
  )
}
```

Index