Package: GPRMortality (via r-universe)

August 25, 2024

Type Package

Title Gaussian Process Regression for Mortality Rates

Version 0.1.0

Author Parinaz Mehdipour <mehdipoor.p@gmail.com> [aut], Ali Ghanbari
<a.ghanbari541@gmail.com> [cre,aut] , Iman Navidi
<iman.navidi.1988@gmail.com> [aut], Farshad Farzadfar
<farzadfar3@yahoo.com> [cph]

Maintainer Ali Ghanbari <a.ghanbari541@gmail.com>

Description A Bayesian statistical model for estimating child
(under-five age group) and adult (15-60 age group) mortality.
The main challenge is how to combine and integrate these
different time series and how to produce unified estimates of
mortality rates during a specified time span. GPR is a Bayesian
statistical model for estimating child and adult mortality
rates which its data likelihood is mortality rates from
different data sources such as: Death Registration System,
Censuses or surveys. There are also various hyper-parameters
for completeness of DRS, mean, covariance functions and
variances as priors. This function produces estimations and
uncertainty (95% or any desirable percentiles) based on
sampling and non-sampling errors due to variation in data
sources. The GP model utilizes Bayesian inference to update
predicted mortality rates as a posterior in Bayes rule by
combining data and a prior probability distribution over
parameters in mean, covariance function, and the regression
model. This package uses Markov Chain Monte Carlo (MCMC) to
sample from posterior probability distribution by 'rstan'
package in R. Details are given in Wang H, Dwyer-Lindgren L,
Lofgren KT, et al. (2012) <doi:10.1016/S0140-6736(12)61719-X>,
Wang H, Liddell CA, Coates MM, et al. (2014)
<doi:10.1016/S0140-6736(14)60497-9> and Mohammadi, Parsaeian,
Mehdipour et al. (2017) <doi:10.1016/S2214-109X(17)30105-5>.

License GPL-2 | GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.0.0), stats, utils, rstan

RoxygenNote 6.1.0

NeedsCompilation no

Date/Publication 2018-08-26 17:14:27 UTC

Repository https://alighanbari26.r-universe.dev

RemoteUrl https://github.com/alighanbari26/gprmortality

RemoteRef HEAD

RemoteSha c63737f867b2b6a566f606b77eeaba5542ff049d

Contents

---

data.mean   *mean of adult mortality rates*

Description

A data frame for GPR mean function and must contain an estimate for each year.

This data must contain year, sex, age_cat and location of each mortality rate. It can be different estimation models of mortality rates by year.

Usage

```
data("data.mean")
```

Format

A data frame with 45942 observations on the following 5 variables.

location a numeric vector

age_cat a numeric vector

sex a numeric vector

year a numeric vector

mean a numeric vector

Details

This data set is the result of Spatio-temporal Model that used in age-sex specific mortality rates as a part of NASBOD project in Iran from 1990 to 2015 for 31 provinces.

Source

This data set was gathered in Non-Communicable Diseases Research Center affiliated to Endocrinology and Metabolism Research Institute,

Tehran University of Medical Sciences. ("http://www.ncdrc.info/")

References

1- Mohammadi Y, Parsaeian M, Farzadfar F, et al. Levels and trends of child and adult mortality rates in the Islamic Republic of Iran, 1990-2013; protocol of the NASBOD study. Arch Iran Med 2014; 17: 176–81.

2-Parsaeian M, Farzadfar F, Zeraati H, et al. Application of spatiotemporal model to estimate burden of diseases, injuries and risk factors in Iran 1990–2013. Arch Iran Med 2014; 17: 28–33.

Examples

```
data(data.mean)
## maybe str(data.mean) ; plot(data.mean) ...
```

---

data.mean.child   *mean of child mortality rates*

Description

A data frame for GPR mean function and must contain an estimate for each year.

This data must contain year and location of each mortality rate. It can be different models for mortality rate estimations by year.

Usage

```
data("data.mean.child")
```

Format

A data frame with 1728 observations on the following 3 variables.

year a numeric vector

location a numeric vector

mean a numeric vector

Details

This data set is the result of Spatio-temporal Model that used in child mortality project in Iran from 1990 to 2015 for 31 provinces.

Source

This data set was gathered in Non-Communicable Diseases Research Center affiliated to Endocrinology and Metabolism Research Institute,

Tehran University of Medical Sciences. ("http://www.ncdrc.info/")

the result of this data published in "http://dx.doi.org/10.1016/S2214-109X(17)30105-5" and is available on "https://data.mendeley.com/datasets/9z3pzd6rmd/1" .

References

1- Mohammadi Y, Parsaeian M, Farzadfar F, et al. Levels and trends of child and adult mortality rates in the Islamic Republic of Iran, 1990-2013; protocol of the NASBOD study. Arch Iran Med 2014; 17: 176–81.

2-Parsaeian M, Farzadfar F, Zeraati H, et al. Application of spatiotemporal model to estimate burden of diseases, injuries and risk factors in Iran 1990–2013. Arch Iran Med 2014; 17: 28–33.

Examples

```
data(mean.rates.child)
## maybe str(mean.rates.child) ; plot(mean.rates.child) ...
```

---

  data.mortality   *age-sex specific mortality rates data set*

Description

A data frame of mortality rates using as data likelihood of the GPR model.

This data set for child mortality rates should contain the year, sex, age_cat, location, completeness of each mortality rate and population of each raw data.

Usage

```
data("data.mortality")
```

Format

A data frame with 395 observations on the following 7 variables.

sex a numeric vector

age_cat a numeric vector

location a numeric vector

year a numeric vector

pop a numeric vector

mortality a numeric vector

completeness a numeric vector

Details

This is a real data gathered in Iran including 31 provinces and 26 years for age-sex specific and adult mortality.

Source

This data set was gathered in Non-Communicable Diseases Research Center affiliated to Endocrinology and Metabolism Research Institute,

Tehran University of Medical Sciences. ("http://www.ncdrc.info/")

References

1-Mehdipour P, Navidi I, Parsaeian M, Mohammadi Y, Moradi Lakeh M, Rezaei Darzi E, Nourijelyani K, Farzadfar F. . Application of Gaussian Process Regression (GPR) in estimating under-five mortality levels and trends in Iran 1990-2013, study protocol. Archives of Iranian medicine. 2014;17(3):189.

2-Mohammadi Y, Parsaeian M, Mehdipour P, Khosravi A, Larijani B, Sheidaei A, et al. Measuring Iran's success in achieving Millennium Development Goal 4: a systematic analysis of under-5 mortality at national and subnational levels from 1990 to 2015. The Lancet Global Health. 2017;5(5):e537-e44.

Examples

```
data(data.mortality)
## maybe str(data.mortality) ; plot(data.mortality) ...
```

---

data.mortality.child   *child mortality rates data*

Description

A data frame of mortality rates using as data likelihood of the GPR model.

This data set for child mortality rates should contain the year, location, name of each source of data, the population of each raw data and whether it comes from a Death Registration System or not.

Usage

```
data("data.mortality.child")
```

Format

A data frame with 4107 observations on the following 6 variables.

year a numeric vector

location a numeric vector

type a character vector

mortality a numeric vector

isDR a numeric vector

pop a numeric vector

Details

This is a real data gathered in Iran including 31 provinces and 52 years for child mortality.

There are three types of data that include the summary (SBH) and complete birth history (CBH) data,

Census data and the Demographic and Health Survey (DHS) contain SBH questions only; the DHS contains CBH questions as well.

Subsequently, the census data of 1986, 1996, 2006, and 2011, the DHS data of 2000 and 2010 were determined as the sources of data.

Source

This data set was gathered in Non-Communicable Diseases Research Center affiliated to Endocrinology and Metabolism Research Institute,

Tehran University of Medical Sciences. ("http://www.ncdrc.info/")

the result of this data published in "http://dx.doi.org/10.1016/S2214-109X(17)30105-5" and is available on "https://data.mendeley.com/datasets/9z3pzd6rmd/1" .

References

1-Mehdipour P, Navidi I, Parsaeian M, Mohammadi Y, Moradi Lakeh M, Rezaei Darzi E, Nourijelyani K, Farzadfar F. . Application of Gaussian Process Regression (GPR) in estimating under-five mortality levels and trends in Iran 1990-2013, study protocol. Archives of Iranian medicine. 2014;17(3):189.

2-Mohammadi Y, Parsaeian M, Mehdipour P, Khosravi A, Larijani B, Sheidaei A, et al. Measuring Iran's success in achieving Millennium Development Goal 4: a systematic analysis of under-5 mortality at national and subnational levels from 1990 to 2015. The Lancet Global Health. 2017;5(5):e537-e44.

Examples

```
data(mortality.rate.child)
## maybe str(mortality.rate.child) ; plot(mortality.rate.child) ...
```

GPRMortality	*Gaussian Process Regression for estimation and projection each age-sex combination of mortality rates*

Description

Searching for the latest methods of estimating mortality rates is a major concern. The need to have accurate and valid estimation of age and sex mortality rate led to apply more powerful and reliable methods. The main challenge is how to combine and integrate these different time series and how to produce unified estimates of mortality rates during a specified time span.

GPR is a Bayesian statistical model for estimating adult (15-60 age group or age-specific) mortality which its data likelihood could be mortality rates from different data sources such as: Death Registration System, Censuses or surveys. This function produces a unique estimation and 95% (or any desirable percentiles) uncertainty based on sampling and non-sampling errors due to variation in data sources.

The GP model utilizes Bayesian inference to update predicted mortality rates as a posterior in Bayes rule by combining data and a prior probability distribution over parameters in mean, covariance function, and the regression model. This package uses Markov Chain Monte Carlo (MCMC) to sample from posterior probability distribution by rstan package in R.

Usage

```
GPRMortality(data.mortality,data.mean,minYear,maxYear,
             nu,rho_,product , n.itr=4000,n.warm=3000,verbose=FALSE)
```

Arguments

data.mortality	a data frame on mortality rates, which needs seven compulsory basic variables: year, mortality rates, the completeness of corresponding mortality rates and their corresponding population; different levels of data such as age_cat, sex, location(countries and so on) can be changed .
data.mean	a data frame of mean data, with the same levels including sex, age_cat, year and location.
minYear	min year will predicted
maxYear	max year will predicted
nu	the degree of differentiability parameter that is ranged from 0.2 to 2 and it controls the smoothness of samples driven from GP model.
rho_	a scale parameter that ranges from 0 to 1 and it defines the amount of correlation between years of data.
product	the value for calculating variance of Registration System, the default is 0.1 for high quality data.
n.itr	the number of iterations to use for running model in Rstan. If not specified, then the default is 4000.

| n.warm | the number of samples for warm-up in MCMC. If not specified, the default is 3000. |
| verbose | TRUE or FALSE: flag indicating whether to print intermediate output from Stan on the console, which might be helpful for model debugging. |

Details

This package intra and extrapolates the mortality rates excluded from different data sources and various demographic models and produces uncertainty.

The algorithm for GPR package was developed in rstan.

The mortality rates will change to per 100,000 population.

non-sampling variance is defined as 0.1 of completeness for countries/locations with high quality of registration more than 90%.

Value

a matrix of GPR result.

Note

Each data source must contin at least two data points in the time span between min and max years to be predicted.

This package need Rtools to be installed. The version should match with R version.

min and max years determined the number of years in GPR results.

Author(s)

Parinaz Mehdipour, Ali Ghanbari, Iman Navidi

References

1. Wang H, Dwyer-Lindgren L, Lofgren KT, et al. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012; 380: 2071–94.

2. Rajaratnam JK, Marcus JR, Levin-Rector A, et al. Worldwide mortality in men and women aged 15–59 years from 1970 to 2010: a systematic analysis. Lancet 2010; published online April 30. DOI:10.1016/S0140-6736(10)60517-X.

3. Mohammadi Y, Parsaeian M, Farzadfar F, et al. Levels and trends of child and adult mortality rates in the Islamic Republic of Iran, 1990-2013; protocol of the NASBOD study. Arch Iran Med 2014; 17: 176–81.

4. Mehdipour P, Navidi I, Parsaeian M, Mohammadi Y, Moradi Lakeh M, Rezaei Darzi E, Nouri-jelyani K, Farzadfar F. . Application of Gaussian Process Regression (GPR) in estimating under-five mortality levels and trends in Iran 1990-2013, study protocol. Archives of Iranian medicine. 2014;17(3):189.

5. Williams CK, Rasmussen CE. Gaussian processes for machine learning. the MIT Press. 2006;2(3):4.

See Also

GPRMortalityChild,GPRMortalitySummary

Examples

```
library("rstan")
library("GPRMortality")
head(data.mortality)
head(data.mean)
mortality <- data.mortality[data.mortality$location%in%c(0,5) &
         data.mortality$age_cat%in%c(1,10) & data.mortality$sex%in%c(0,1),]
mean <- data.mean[data.mean$location%in%c(0,5) &
     data.mean$age_cat%in%c(1,10) & data.mean$sex%in%c(0,1),]
# WARNING: The following code will take a long time to run

fit = GPRMortality(mortality,mean,minYear=1990,maxYear=2015,
                 nu = 2,rho_ =0.4 ,product =  0.1 ,verbose=TRUE)
fit_sum = GPRMortalitySummary(fit)
fit_sum
```

GPRMortalityChild	*Gaussian Process Regression for estimation and projection of child (under-5) mortality rates*

Description

Searching for the latest methods of estimating mortality rates is a major concern. The need to have accurate and valid estimation of under-5 mortality rate led to apply more powerful and reliable methods. The main challenge is how to combine and integrate these different time series and how to produce unified estimates of mortality rates during a specified time span.

GPR is a Bayesian statistical model for estimating child (under-five age group) mortality which its data likelihood could be mortality rates from different data sources such as: Death Registration System, Censuses or surveys. This function produces estimations and uncertainty (95% or any desirable percentiles) based on sampling and non-sampling errors due to variation in data sources.

The GP model utilizes Bayesian inference to update predicted mortality rates as a posterior in Bayes rule by combining data and a prior probability distribution over parameters in mean, covariance function, and the regression model. This package uses Markov Chain Monte Carlo (MCMC) to sample from posterior probability distribution by rstan package in R.

Usage

```
GPRMortalityChild( data.mortality , data.mean ,minYear,maxYear,
                 nu,rho_,n.itr=4000,n.warm=3000,verbose=FALSE)
```

Arguments

`data.mortality`	a data frame on mortality rates and population. It needs five compulsory basic variables: year, type of data sources, mortality rates and their corresponding population, location (countries and so on).
`data.mean`	a data frame of data mean, with the same levels (year and location) as data.mortality.
`minYear`	min year will predicted
`maxYear`	max year will predicted
`nu`	the degree of differentiability parameter that is ranged from 0.2 to 2 and it controls the smoothness of samples driven from GP model.
`rho_`	a scale parameter that ranges from 0 to 1 and it defines the amount of correlation between years of data.
`n.itr`	the number of iterations to use for running model in rstan. If not specified, then the default is 4000.
`n.warm`	the number of samples for warm-up in MCMC. If not specified, the default is 3000.
`verbose`	TRUE or FALSE: flag indicating whether to print intermediate output from Stan on the console, which might be helpful for model debugging.

Details

This package intra and extrapolates the mortality rates excluded from different data sources and various demographic models and produces uncertainty.

The algorithm for GPR package was developed in rstan.

The mortality rates will change to per 1000 population.

Value

a list including two parts:

1. a matrix of GPR result.

2. a matrix of completeness bias information of Dearth Registration System by location including coefficients, Standard error of coefficient, and p-value of coefficients whether it can assess that DR system is biased in each location.

3. a matrix of different data sources variances.

Note

Each data source must contin at least two data points in the time span between min and max years to be predicted.

min and max years determined the number of years in GPR results.

This package need Rtools to be installed. The version should match with R version.

Author(s)

Parinaz Mehdipour, Ali Ghanbari, Iman Navidi

References

1. Wang H, Liddell CA, Coates MM, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014; published online May 2. http://dx.doi.org/10.1016/S0140-6736(14)60497-9.

2. Wang H, Dwyer-Lindgren L, Lofgren KT, et al. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012; 380: 2071-94.

3. Mohammadi Y, Parsaeian M, Farzadfar F, et al. Levels and trends of child and adult mortality rates in the Islamic Republic of Iran, 1990-2013; protocol of the NASBOD study. Arch Iran Med 2014; 17: 176–81.

4. Parsaeian M, Farzadfar F, Zeraati H, et al. Application of spatiotemporal model to estimate burden of diseases, injuries and risk factors in Iran 1990-2013. Arch Iran Med 2014; 17: 28-33.

5. Mehdipour P, Navidi I, Parsaeian M, Mohammadi Y, Moradi Lakeh M, Rezaei Darzi E, Nourijelyani K, Farzadfar F. . Application of Gaussian Process Regression (GPR) in estimating under-five mortality levels and trends in Iran 1990-2013, study protocol. Archives of Iranian medicine. 2014;17(3):189.

6. Mohammadi Y, Parsaeian M, Mehdipour P, Khosravi A, Larijani B, Sheidaei A, et al. Measuring Iran's success in achieving Millennium Development Goal 4: a systematic analysis of under-5 mortality at national and subnational levels from 1990 to 2015. The Lancet Global Health. 2017;5(5):e537-e44.

7. Williams CK, Rasmussen CE. Gaussian processes for machine learning. the MIT Press. 2006;2(3):4.

See Also

GPRMortality,GPRMortalitySummary

Examples

```
library("rstan")
library("GPRMortality")
head(data.mortality.child)
head(data.mean.child)
mortality <- data.mortality.child[data.mortality.child$location%in%c(0,5) ,]
mean <- data.mean.child[data.mean.child$location%in%c(0,5) ,]
# WARNING: The following code will take a long time to run

fit <- GPRMortalityChild(mortality,mean,minYear=1990,maxYear=2015,
    nu = 2,rho_ = 0.4 , n.itr=2000,n.warm=1000,verbose=TRUE)
fit$simulation
fit$variance
fit_sum = GPRMortalitySummary(fit)
fit_sum
```

GPRMortalitySummary   *summary table for Gaussian Process Regression results*

Description

This function summarize the results by percentiles.

This function works for both child and age-sex specific mortality rates.

Usage

```
GPRMortalitySummary(model,percentile=c(0.025,0.5,0.975))
```

Arguments

model   a model object.

percentile   a vector of uncertainty interval percentile. The default is 0.025, 0.5 and 0.975.

Value

a matrix of GPR result including percentile.

Author(s)

Parinaz Mehdipour, Ali Ghanbari, Iman Navidi

Examples

```
library("rstan")
library("GPRMortality")
head(data.mortality)
head(data.mean)
mortality <- data.mortality[data.mortality$location%in%c(0,5) &
     data.mortality$age_cat%in%c(1,10) & data.mortality$sex%in%c(0,1),]
mean <- data.mean[data.mean$location%in%c(0,5) &
     data.mean$age_cat%in%c(1,10) & data.mean$sex%in%c(0,1),]

# WARNING: The following code will take a long time to run

 fit = GPRMortality(mortality,mean,minYear =  1990,maxYear =  2015,
                    nu = 2,rho_ =0.4 ,product =  0.1 ,verbose=TRUE)
####### summary
 fit_sum = GPRMortalitySummary(fit)
 fit_sum
```

Index