Package ‘VirtualPop’

March 18, 2024

Type Package

Title Simulation of Populations by Sampling Waiting-Time Distributions

Version 2.0.2

Imports msm,HMDHFDplus

Suggests knitr, kableExtra, ggplot2, foreign, lubridate, xml2, eha,
        survival, survminer, rmarkdown

BuildResaveData best

VignetteBuilder knitr

LazyData true

Date 2024-03-18

Maintainer Frans Willekens <willekens@nidi.nl>

Description Constructs a virtual population from fertility and mortality rates for any country,
calendar year and birth cohort in the Human Mortality Database <https://www.mortality.org>
and the Human Fertility Database <https://www.humanfertility.org>. Fertility histories are simulated for every individual and their offspring, producing a multi-
generation virtual population.

License GPL-2

NeedsCompilation no

Depends R (>= 4.3.0),

Encoding UTF-8

BugReports https://github.com/willekens/VirtualPop/issues

RoxygenNote 7.2.3

Author Frans Willekens [aut, cre] (<https://orcid.org/0000-0001-6125-0212>),
        Tim Riffe [ctb] (<https://orcid.org/0000-0002-2673-4622>)

Repository CRAN

Date/Publication 2024-03-18 10:30:02 UTC
**BuildViP**

**Builds a Virtual Population in a Single Step**

**Description**

Builds a virtual population from mortality and fertility rates retrieved from the Human Mortality Database (HMD) and the Human Fertility Database (HFD) in a single step.

**Usage**

```r
BuildViP(
  user = NULL,
  pw_HMD = NULL,
  pw_HFD = NULL,
  countrycode,
  cohort = NULL,
  refyear = NULL,
  ncohort,
  ngen,
  mort = TRUE
)
```

**Arguments**

- **user**: User name (e-mail address)
- **pw_HMD**: Password Human Mortality Database
- **pw_HFD**: Password Human Fertility Database
Children

Generates Individual Fertility Histories

Description

Builds individual fertility histories from conditional fertility rates. Children() uses the function Sim_bio().

Usage

Children(dat0, rates, mort = NULL)
Arguments

- dat0: Data frame with data on individual members of the virtual population (dLH format)
- rates: Mortality and fertility rates. The object 'rates' is produced by the function Getrates().
- mort: Presence or absence of mortality (optional). Default: mortality is present (mort=TRUE). If mortality is absent, set mort=FALSE.

Value

List object with two components:
- data: Data frame with updated information on members of the virtual population
- dch: Data frame with information on children

Examples

```r
# The example generates data on children of the first 10 female members of
# the first generation of the virtual population.
utils::data(dLH, package="VirtualPop")
utils::data(rates, package="VirtualPop")
dat0 <- dLH[dLH$sex=="Female" & dLH$gen==1,][1:10,]
out <- VirtualPop::Children(dat0=dat0, rates=rates)
```

---

**dLH**

*Individual fertility histories based on period data and in the presence of mortality (USA 2021)*

---

**Description**

Fertility histories based on period data and in the presence of mortality. The histories are simulated from age-specific death rates and conditional fertility rates of USA 2021.

**Usage**

```r
data(dLH, package="VirtualPop")
```

**Format**

A data frame with data about 7,000 individuals (2000 in initial cohort).

- **ID**: Identification number
- **gen**: Generation
- **cohort**: Birth cohort (year of birth)
- **sex**: Sex. A factor with levels Males and Females
\textbf{bdated} Date of birth (decimal date)
\textbf{ddated} Date of death (decimal date)
\textbf{x\_D} Age at death (decimal number)
\textbf{IDmother} ID of mother
\textbf{IDfather} ID of father
\textbf{jch} Child’s line number in the nuclear family (household)
\textbf{IDpartner} ID of partner
\textbf{udated} Date of union formation
\textbf{nch} Number of children ever born to the individual

The object has four attributes:

- Country
- type: Type of data used to produce the histories (period data or cohort data)
- refyear: Calendar year for which period data are used. If cohort data are used, refyear is missing (NA)
- cohort: Year of birth of cohort for which the data are used. If period data are used, cohort is missing (NA)

\textbf{Source}

The virtual population is produced from period mortality rates by age and period fertility rates by age and parity from the United States 2021. The data are from the Human Mortality Database (HMD) and the Human Fertility Database (HFD).

\textbf{\textit{e0}} \hspace{1cm} \textit{Mean Ages at Death and Probabilities of Surviving to Selected Ages, by Sex}

\textbf{Description}

Computes (a) Life expectancy at birth, (b) Probability of surviving at age 65, and (c) Probability of surviving at age 85

\textbf{Usage}

\texttt{e0(d)}

\textbf{Arguments}

\texttt{d} \hspace{1cm} The name of the database. If missing, dLH is used if it exists.
Value

e0          Mean ages at death
Prob65      Probability of surviving at age 65
Prob85      Probability of surviving at age 85

Examples

utils::data(dLH, package="VirtualPop")
e0(d=dLH)

Description

Reads data from the HMD and HFD into R. The function uses the readHMDweb() and the readHFDweb() functions of the HMDHFDplus package.

Usage

GetData(country, user, pw_HMD, pw_HFD)

Arguments

country      Code of the selected country. The code must be one of the country codes of HMD and HFD.
user         email address of the user, used at registration with the HMD and HFD. It is assumed that the same email address is used for both HMD and HFD.
pw_HMD       Password to access HMD, provided at registration.
pw_HFD       Password to access HFD, provided at registration

Value

data_raw      A list object with four elements:
country      Country
LTf           Life table for female population for all years available in the HMD
LTm           Life table for male population for all years available in the HMD
fert_rates   Conditional fertility rates for all years available in the HFD
### Description

Builds a virtual population from mortality rates by age and sex, and fertility rates by age of mother and parity.

### Usage

```r
GetGenerations(rates, ncohort = NULL, ngen = NULL, mort = NULL)
```

### Arguments

- **rates**: List object with death rates (ASDR) and birth rates (ASFR). Produced by function `VirtualPop::GetRates()`. Rates of USA 2021 are distributed with the VirtualPop package.
- **ncohort**: Size of hypothetical birth cohort (first generation)
- **ngen**: Number of generations to be simulated. No upper limit.
- **mort**: Presence or absence of mortality. This parameter is optional. Default is TRUE. If mortality is absent, mort=FALSE.

### Value

- **dataAllgen**: The database of simulated individual lifespans and fertility histories (all generations).

The object `dataAllgen` has four attributes:

- **country**: The country
- **type**: The type of data (period data or cohort data).
- **refyear**: The calendar year for which the period data are used (reference year).
- **cohort**: The birth cohort (if applicable).
Examples

```r
utils::data(rates, package = "VirtualPop")
dLH <- VirtualPop::GetGenerations (rates=rates, ncohort=1000, ngen=4)
```

---

**GetRates**

*Retrieves Period Mortality and Fertility Rates from HMD and HFD for a Selected Country and Selected Year*

---

**Description**

The rates are retrieved from the life tables and fertility tables included in the raw data downloaded from the HMD and HFD.

**Usage**

```r
GetRates(data, refyear)
```

**Arguments**

- `data` data (the object `data_raw`, produced by the `GetData()` function.)
- `refyear` Reference year, which is the year of period data

**Value**

A list object with three elements:

- `ASDR` Age-specific death rates, by sex for reference year
- `ASFR` Age-specific birth rates by birth order for reference year
- `ratesM` Matrix of transition rates in format required for multistate modelling

The object returned by the function has three attributes:

- `country` Country
- `type` Type of data (period data or cohort data)
- `year` Calendar year for which period death rates are used to complete cohort experience in case of incomplete mortality experience (reference year).

**Examples**

```r
# Not run: # Not run because passwords needed
# Input data: data_raw produced by GetData().
rates <- GetRates(data=data_raw, refyear=2021)

# End(Not run)
```
GetRatesC

Retrieves Cohort Data from the HMD and HFD and Obtains Cohort Rates

Description

Retrieves cohort data from the HMD and HFD and produces cohort rates (death rates by age and sex and conditional fertility rates by age and parity). The function combines the steps of (a) data retrieval and (b) extraction of mortality and fertility rates.

Usage

GetRatesC(country, user, pw_HMD, pw_HFD, refcohort)

Arguments

- **country**: Code of the country selected. The code must be one of the country codes of HMD and HFD.
- **user**: Name of the user, used at registration with the HMD and HFD. It is assumed that the same name is used for both HMD and HFD.
- **pw_HMD**: Password to access HMD, provided at registration.
- **pw_HFD**: Password to access HFD, provided at registration.
- **refcohort**: Year of birth of cohort for which the data are used for the simulation.

Value

A list object with three elements:

- **ASDR**: Age-specific death rates by sex for selected birth cohort
- **ASFR**: Age-specific fertility rates by parity for selected birth cohort
- **ratesM**: Matrix of transition rates in format required for multistate modelling

The object returned by the function has five attributes:

- **country**: Country
- **type**: Type of data (period data or cohort data)
- **cohort**: Birth cohort (year of birth)
- **refyear**: Calendar year for which period death rates are used to complete cohort experience in case of incomplete mortality experience (reference year).
- **start_pASDR**: Lowest age for which cohort data are missing. The mortality rates of that age and higher ages are borrowed from period data collected in the reference year.
Examples

## Not run:
ratesC <- GetRatesC(country="USA",user,pw_HMD,pw_HFD,refcohort)

## End(Not run)

---

### H_pw

**Computes Cumulative Hazard at Duration t under a Piecewise Exponential Model**

**Description**

Computes cumulative hazard at duration t from piecewise-constant rates.

**Usage**

```
H_pw(t, breakpoints, rates)
```

**Arguments**

- **t**
  - Duration at which cumulative hazard is required. It may be a vector of durations.
- **breakpoints**
  - Breakpoints: values of time at which piecewise-constant rates change.
- **rates**
  - Piecewise-constant rates

**Value**

Cumulative hazard at duration t

**See Also**

- `pw_root()` and `r_pw_exp()`: Function `H_pw()` is called by `pw_root()`, which is called by `r_pw_exp()`.

**Examples**

# Example 1
breakpoints <- c(0, 10, 20, 30, 60)
rates <- c(0.01, 0.02, 0.04, 0.15)
z <- VirtualPop::H_pw(t=0:40, breakpoints=breakpoints, rates=rates)

# Example 2
utils::data(rates, package="VirtualPop")
ages <- as.numeric(rownames(rates$ASDR))
breakpoints <- c(ages,120)
zz <- VirtualPop::H_pw(t=ages, breakpoints=breakpoints, rates=rates$ASDR[,1])
**Lifespan**

*Generates Individual Lifespan(s)*

**Description**

Uses age-specific death rates to simulate length of life. The function generates age(s) at death and date(s) of death. The function uses the function `rpexp()` of the `msm` package and `uniroot()` of base R.

**Usage**

```r
Lifespan(data, ASDR, mort = NULL)
```

**Arguments**

- `data` Data frame with individual data. If the object "data" includes date of birth (bdated; decimal date), then the date of death is computed.
- `ASDR` Age-specific death rates
- `mort` Presence or absence of mortality. This parameter is optional. Default is TRUE. If mortality is (should be) absent, mort=FALSE.

**Value**

`LS` Data frame with age(s) at death and date(s) of death

**Examples**

```r
utils::data(dLH, package="VirtualPop")
utils::data(rates, package="VirtualPop")
d <- VirtualPop::Lifespan (dLH[1:5,1:5], ASDR=rates$ASDR)
```

---

**PartnerSearch**

*Simple Partner Search Simulation*

**Description**

In this simple model, a partner is an individual of a different sex selected at random among members of the same generation. The function is called by `GetGenerations()`.

**Usage**

```r
PartnerSearch(dLH)
```
pw_root

Arguments

dLH Database

Value

dLH Updated version of database (dLH), which includes, for each individual without
a partner and able to find a partner, the ID of the partner.

Examples

d <- VirtualPop::PartnerSearch(dLH=dLH)

Description

The function pw_root() specifies the mathematical function g(t). The equation to be solved is g(t)=0,
with g(t) the cumulative hazard function of the piecewise exponential distribution + log(u) with u
a random draw from standard uniform distribution (see vignette "Piecewise_exponential", Section
2.2.4).

Usage

pw_root(t, breakpoints, rates, uu)

Arguments

t Vector of durations for which the equation g(t)=0 should be solved.
breakpoints Breakpoints
rates Piecewise-constant rates
uu Random draw from standard uniform distribution.

Details

pw_root is an argument of the function uniroot() of base R (argument "f"). It is required by uniroot().
The function uniroot() is called by r.pw_exp(). See also Functions H_pw() and r.pw_exp().

Value

Vector of differences between cumulative hazard and -log(uu) for different values of t.
r.pw_exp

Examples

```r
c(0, 10, 20, 30, 60)
c(0.01, 0.02, 0.04, 0.15)
c(10, 18.3, 23.6, 54.7)
0.43
```

---

Draws Waiting Times from a Piecewise-Exponential Distribution.

Description

The function produces \( n \) realizations of a piecewise-exponentially distributed random waiting time.

Usage

```r
r.pw_exp(n, breakpoints, rates)
```

Arguments

- **\( n \)**: Number of random draws
- **breakpoints**: Breakpoints in piecewise-exponential distribution
- **rates**: Piecewise-constant rates

Value

Vector of waiting times, drawn randomly from a piecewise-exponential survival function.

Examples

```r
c(0, 10, 20, 30, 60)
c(0.01, 0.02, 0.04, 0.15)
0.43
```
**Description**

Data consisting of period rates of mortality by age and sex and fertility by age and parity, USA 2021

**Usage**

data(rates, package="VirtualPop")

**Format**

A list of three objects.

- **ASDR**  Mortality rates
- **ASFR**  Fertility rates
- **ratesM**  Multistate transition rates

The dataset has three attributes:

- Country
- Type of rates: period rates or cohort rates
- Calendar year for which period death rates are used to complete cohort experience in case of incomplete mortality experience (reference year).

**Source**

The data are downloaded from the Human Mortality Database (HMD) and the Human Fertility Database (HFD). Country: USA. Year: 2021

---

**Description**

Cohort rates of mortality by age and sex and fertility by age and parity, USA birth cohort 1964

**Usage**

data(ratesC, package="VirtualPop")
Sim_bio

Format

A list of three objects.

ASDR  Mortality rates
ASFR  Fertility rates
ratesM  Multistate transition rates

The object returned by the function has five attributes:

- Country
- type: Type of data (period data or cohort data)
- cohort: Birth cohort (year of birth)
- year: Calendar year for which period death rates are used to complete cohort experience in case of incomplete mortality experience (reference year).
- start_pASDR: Lowest age for which cohort data are missing. The mortality rates of that age and higher ages are borrowed from period data collected in the reference year.

Source

The data are downloaded from the Human Mortality Database (HMD) and the Human Fertility Database (HFD). Country: USA. Cohort: 1964

Sim_bio

Generic Function to Generate Single Life History

Description

The function generates a single life history from age-specific transition rates (rates$ratesM) and an initial state. RatesM is an object with the rates in the proper format for multistate modelling. The user supplies the starting age and ending age of the simulation.

Usage

Sim_bio(datsim, ratesM)

Arguments

datsim  Dataframe with, for each individual, ID, date of birth, starting and ending times (ages) of the simulation, and the state occupied at the start of the simulation (see vignette "Tutorial").

ratesM  Multistate transition rates in standard (multistate) format

Details

The function is called from the function VirtualPop::Children(). It uses the rpexp() function of the msm package.
Value

- **age_startSim**: Age at start of simulation
- **age_endSim**: Age at end of simulation
- **nstates**: Number of states
- **path**: path: sequence of states occupied
- **ages_trans**: Ages at transition

Examples

```r
# Fertility history is simulated from starting age to ending age
# Individual starts in state "par0"
utils::data(rates, package="VirtualPop")
popsim <- data.frame(ID=1, born=2000.450, start=0, end=80, st_start="par0")
ch <- VirtualPop::Sim_bio (datsim=popsim, ratesM=rates$ratesM)
```
Index

BuildViP, 2
Children, 3
dLH, 4
e0, 5
GetData, 6
GetGenerations, 7
GetRates, 8
GetRatesC, 9
H_pw, 10
Lifespan, 11
PartnerSearch, 11
pw_root, 12
r.pw_exp, 13
rates, 14
ratesC, 14
Sim_bio, 15