Package ‘predfinitepop’

April 24, 2014

Type Package

Title Predictive Inference on Totals and Averages of Finite Populations Segmented in Planned and Unplanned Domains.

Version 1.0

Date 2014-04-24

Author Juan Carlos Martinez-Ovando, Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez.

Maintainer Juan Carlos Martinez-Ovando <juan.martinez@banxico.org.mx>

Description This package computes Bayesian predictive inference on totals and averages of finite populations segmented in planned and unplanned domains. Inference is based on Bayesian nonparametric methods using species-sampling models (specifically the Dirichlet process).

Acknowledgement The authors appreciate advice from Pedro Gonzalez-Alegria and Claudia Velazquez-Villegas of Banco de Mexico.

R topics documented:

`conteo` *Counts the number of individuals in $\mathcal{S}\_j$ and $\widetilde{\mathcal{S}}\_j$.*

Description

This function counts the number of individuals in $\mathcal{S}_j$ and $\widetilde{\mathcal{S}}_j$, with respect to the total number of individuals in a given planned domain $\mathcal{P}_j$.

Usage

```
conteo(datos_j, N_j)
```

Arguments

`datos_j`	Data matrix with individual measurements for $\mathcal{S}_j$
`N_j`	Number of individuals in $\mathcal{P}_j$ (the jth planned domain) (N_j is assumed known)

Details

"datos_j" - This object should contain two columns labeled "n_i" and "domplan".

"domplan" - Column vector with categories for the planned domains.

"n_i"- Number of individuals in the ith group of individuals, and

"N_j" - This object should include two colums labeled "N_j" and "domplan".

"domplan" - Column vector with categories for the planned domains.

"N_j" - Number of individuals in the jth planned domain, and

Value

The function 'conteo' produces an output list with three elements:

`M`	Number of groups in "$\mathcal{S}_j$"
`CardS`	Number of individuals in "$\mathcal{S}_j$" (if "n_i"= 1, for any "j", then $Cards = M = nrow(datos_j)$)
`CardNoS`	Number of individuals out of sample

Author(s)

Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

`dataref` *Reference data set of a simulated population segemented in planned and unplanned domains.*

Description

This data set corresponds to a simulated population of 500 individuals grouped in two planned domains. Each planned domain of the population has been segmented, as well, into the three unplanned domains. Within the planned domain A, the segmentation is given as domain I (200), domain II (400) and domain III (200), whereas within planned domanin B, the segmentation is given as domain I (780), domain II (400) and domain III (20). Individual measurements were simulated from log-normal and Weibull distributions with different parameterizations for each combination of planned and unplanned domains. See Martinez-Ovando et al. (2014) for further explanations.

Usage

```
data(dataref)
```

Format

A data frame with 500 observations on the following 3 variables.

`id` Identification code for simulated data.

`y_i` Simulated individual ourcomes.

`domplan` Identification for planned domains (domain A and B).

Source

Simulated data. See Martinez-Ovando et al. (2014) for further details.

References

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

Examples

```
data(dataref)
```

`datasample` *Data set of a simulated population segemented in planned and unplanned domains.*

Description

This is an object list with 20 simulated samples from the simulated population of 2,000 individuals grouped in two planned domains. Planned domain A is formed of 800 individuals, and domain B is composed of 1.2 thousand individuals. Each planned domain of the population has been segmented, as well, into the three unplanned domains.

Usage

```
data(datasample)
```

Format

List of lenght 20, with entries:

l1. 'data.frame': 100 obs. of 8 variables with a sample of size 5 percent

l2. 'data.frame': 200 obs. of 8 variables with a sample of size 10 percent

l3. 'data.frame': 300 obs. of 8 variables with a sample of size 15 percent

l4. 'data.frame': 400 obs. of 8 variables with a sample of size 20 percent

l5. 'data.frame': 500 obs. of 8 variables with a sample of size 25 percent

l6. 'data.frame': 600 obs. of 8 variables with a sample of size 30 percent

l7. 'data.frame': 700 obs. of 8 variables with a sample of size 35 percent

l8. 'data.frame': 800 obs. of 8 variables with a sample of size 40 percent

l9. 'data.frame': 900 obs. of 8 variables with a sample of size 45 percent

l10. 'data.frame': 1000 obs. of 8 variables with a sample of size 50 percent

l11. 'data.frame': 1100 obs. of 8 variables with a sample of size 55 percent

l12. 'data.frame': 1200 obs. of 8 variables with a sample of size 60 percent

l13. 'data.frame': 1300 obs. of 8 variables with a sample of size 65 percent

l14. 'data.frame': 1400 obs. of 8 variables with a sample of size 70 percent

l15. 'data.frame': 1500 obs. of 8 variables with a sample of size 75 percent

l16. 'data.frame': 1600 obs. of 8 variables with a sample of size 80 percent

l17. 'data.frame': 1700 obs. of 8 variables with a sample of size 85 percent

l18. 'data.frame': 1800 obs. of 8 variables with a sample of size 90 percent

l19. 'data.frame': 1900 obs. of 8 variables with a sample of size 95 percent

l20. 'data.frame': 2000 obs. of 8 variables with a sample of size 100 percent

List of variables:

id Individual identification code

y_i Group measurenment

domnoplan Identification code for unplanned domains

n_i Group size

domplan Identification code for planned domains

1 Indicator variable for unplanned domain I

2 Indicator variable for unplanned domain II

3 Indicator variable for unplanned domain III

Details

Within the planned domain A, the segmentation is given as domain I (200), domain II (400) and domain III (200), whereas within planned domanin B, the segmentation is given as domain I (780), domain II (400) and domain III (20). Individual measurements were simulated from log-normal and Weibull distributions with different parameterizations for each combination of planned and unplanned domains. The original sample is of size 5 percent. The upcoming samples were extrated increasing the sample size in 5 percent at a time. See Martinez-Ovando et al. (2014) for further explanations.

Source

Simulated data. See Martinez-Ovando et al. (2014) for further details.

References

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

Examples

```
data(datasample)
```

`dinvgauss` *Inverse Gaussian distribution - Density.*

Description

Computes de p.d.f. function for the inverse Gaussian distribution.

Usage

```
dinvgauss(y, m, s, logcond=FALSE)
```

Arguments

`y`	p-dimensional vector with positive entries
`m`	Mean parameter
`s`	Scale parameter
`logcond`	Logic parameter for log-scale form

Author(s)

Jim Lindsey <jlindsey@luc.ac.be>

References

- "Utilities for Nonlinear Regression and Repeated Measurements Models." Package "rmutil" version 1.0.

`domnoplan` *Finite population inference on unplanned domains using reference data.*

Description

Generates Monte Carlo samples from the predictive distribution of totals of a finite population segmented in planned and unplanned domains, along with simulations of the predictive distribution for the composition of the population between the unplanned domains.

Usage

```
domnoplan(datos,datos_ant,domplan_N,alphaDP,colid_D,alpha_D,inter,part,nSim)
```

Arguments

`datos`	$(M$ X $p)$-dimensional array with positive entries for $\mathcal{S}_j$
`datos_ant`	$(M$ X $p)$-dimensional reference array for calibration of G_0
`domplan_N`	Matrix array with counts of individuals in each planned domain
`alphaDP`	J-dimensional array with positive entries for the parameters of the Dirichlet process for F_j (with J being the number of planned domains)
`colid_D`	D-dimensional matrix array with the columns in datos that correspond to the indicator variables of the planned domains (those indicator variables represent a partion of "datos")
`alpha_D`	$(J$ X $D)$-dimensional array with positive entires for the parameters of the multinomial-Dirichlet component for the composition across unplanned domains. Note: Each one of the J rows is a vector of composition for $\mathcal{P}_j$ divided across the D unplanned domains
`inter`	Tuning parameter for model comparison and selection (related to calibration of G_0)
`part`	Number of partitions for predictive cross-validation (related to calibration of G_0)
`nSim`	Number of Monte Carlo simulated replicates of the predictive distribution

Details

- datos : Represent the data sample of the target population, unplanned domains labeled.

It should contain the following columns:

"domplan" - Categories for planned domains.

"y_i" - Actual individual measurements (for the moment, they must be positive and continuous) for the group of observation.

"n_i" - Number of individuals in the group (if the unit of observation in the sample are individuals, then "n_i" must be equal to 1)

- datos_ant : Represents the data reference used to calibrate G_{j0}

The data must be labeled by planned domains. It should contain the following columns:

"domplan" - Categories of planned domains

"y_i" - Positive real and individual measurements of each group in the sample (when the units of observation are the groups, "y_i" should be per capita measurement)

- domplan_N : Represents counts (or reference population) of the target population, divided by the planned domains.

Tagged data must be labeled by domains planned. It should contain the following columns:

"domplan" - Categories fot planned domains.

"N_j" - Number of individuals in each planned domain.

Value

The function 'domnoplan' produces:

`total_domnoplan_sim`

: Matrix array of dimension "J * (3 +2 D)* nSim" with predictions for the planned domains.
Column 1 - Indicator of the planned domains
Column 2 - T_j (totals of the planned domains)
Column 3 - N_j (composition of the planned domains)
Column 4 - $(4 + D - 1) - T_{d_j}$ (totals of unplanned domains, such that $T_j = \sum_{j=1}^{J} T_{d_j}$ (across d))
Column 5 - $(4 + D)$ to $(3 + 2 * D) - N_{d_j}$ (composition of unplanned domains, such that $N_j = \sum_{j=1}^{J} N_{d_j}$ (across d))

Author(s)

Setgio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domnoplan_composicion, domnoplan_totalcomp.

Examples

```
#---------------------------------------------------------
# Uses of library:  "predfinitepop"
#---------------------------------------------------------

#             Read files with functions in R
rm(list=ls())

library("predfinitepop")

set.seed(12345)

#--------------------------------------------------------------
# DATA
#--------------------------------------------------------------
data(dataref)
data(datasample)
data(popcomp)

# Sample (with 10% of the population)

datos <- datasample[[1]]
datos_ant <- dataref
domplan_N <- popcomp[[1]]

#--------------------------------------------------------------
# MODEL PARAMETERS
#--------------------------------------------------------------

# a. Parameter "alpha" of the Dirichlet process
# (this is for J02 planned domains)
alphaDP <- matrix(0.03,2,1)

# b. vector with indicator variables for unplanned domains
colid_D <- t(as.matrix(c(6:8)))

# c. Matrix with parameters for unplanned domains
# (this is for J=2 planned domains)
alpha_D <- matrix(0.3,2,length(colid_D))

#--------------------------------------------------------------
# CODE PARAMETERS
#--------------------------------------------------------------

# A. Interval for calibration of G_0
inter <- 3

# B. Number of groups for predictive validation and model comparison
part <- 20

# C. Number of Monte Carlo replicates of the predictive distributions
nSim <- 10
```

```
#--------------------------------------------------------------
# TESTING "domnoplan.R"
#--------------------------------------------------------------
date.ini <- date()
domnoplan_sim <- domnoplan(datos,datos_ant,domplan_N,alphaDP,colid_D,alpha_D,inter,part,nSim)
date.fin <- date()

#
# --END --
```

---

domnoplan_composicion *Generates Monte Carlo samples from the predictive distribution for the vector N_j across the D unplanned domains within a given planned domain.*

Description

This function simulates Monte Carlo samples from the predictive distribution of the vector N_j across the D unplanned domains within a given planned domain j.

Usage

```
domnoplan_composicion(datos_j,N_j,N_S_j,N_Stil_j,colid_D,alpha_D,nSim)
```

Arguments

datos_j	Data matrix with features and number of individuals in the sample $\mathcal{S}_j$
N_j	Number of individuals in $\mathcal{P}_j$ (jth planned domain)
N_S_j	Number of individuals in the sample $\mathcal{S}_j$
N_Stil_j	Number of individuals out the sample $\mathcal{S}_j$
colid_D	D-dimensional matrix array with the columns in 'datos' that correspond to the indicator variables of the planned domains (those indicator variables represent a partion of 'datos')
alpha_D	D-dimensional array with positive entires for the parameters of the multinomial-Dirichlet component for the composition across unplanned domains (NOTE: this one makes reference to a single planned domain)
nSim	Number of Monte Carlo simulated replicates of the predictive distribution

Value

This function 'domnoplan_composicion' produces:

N_S_domnoplan	Composition of the number of individuals in saample $\mathcal{S}_j$ ($N_j^{\mathcal{S}}$) across the D unplanned domains
domnoplan_composicion_sim	(1 X D X nSim) matrix with Monte Carlo samples of the predictive distribution of the composition of $\tilde{\mathcal{S}}_j$

Author(s)

Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domnoplan

---

domnoplan_g0 — *Finite population inference on unplanned domains using subjective G_j0.*

Description

Generates Monte Carlo samples of the predictive distribution of total baseline distributions of a finite population segmented in planned and unplanned domains, using a predefined set of $(G_{j0})_{j=1}^{J}$, along with simulations of the predictive distribution for the composition of the population between the unplanned domains.

Usage

```
domnoplan_g0(datos,domplan_N,alphaDP,colid_D,alpha_D,nSim,g0_licitacion_sal)
```

Arguments

`datos`	(M X p)-dimensional array with positive entries for $\mathcal{S}_j$
`domplan_N`	Matrix array with counts of individuals in each planned domain
`alphaDP`	J-dimensional array with positive entries for the parameters of the Dirichlet process for F_j (with J being the number of planned domains)
`colid_D`	D-dimensional matrix array with the columns in datos that correspond to the indicator variables of the planned domains (those indicator variables represent a partion of datos)
`alpha_D`	(J X D)-dimensional array with positive entires for the parameters of the multinomial-Dirichlet component for the composition across unplanned domains. Note: Each one of the J rows is a vector of composition for $\mathcal{P}_j$ segmented across D unplanned domains
`nSim`	Number of Monte Carlo simulated replicates of the predictive distribution
`g0_licitacion_sal`	(J X 1) object list, each entry is and the object list itself associated with each G_{j0} for the J planned domains. The first element for arch G_{j0} should be the name of the chosen distribution (see details below for alternatives), the second element should be a vector object with the parameters associated with distribution, and the third element should be its associated expectation

Details

- datos : Represent the data sample of the target population, unplanned domains labeled.

It should contain the following columns:

"domplan" - Categories fot planned domains.

"y_i" - Actual individual measurements/putcomes (for the moment, they must be positive) for the group of observation.

"n_i" - Number of individuals in the group (if the unit of observation in the sample are individuals, then "n_i" must be equal to 1)

- domplan_N : Represents counts (or reference population) of the target population, divided by the planned domains.

Tagged data must be labeled by domains planned. It should contain the following columns:

"domplan" - Categories fot planned domains.

"N_j" - Number of individuals in each planned domain.

- g0_licitacion_sal: Chose one and only one of the distribution,

i) Gamma

ii) Weibull

iii) Lognormal

iv) Inverse-Gaussian

Parameterizations for distribution:

i) `Gamma` distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto x^{\alpha-1}\exp\{-x/\beta\}$,

where α is the shape parameter, and β is the scale parameter.

ii) `Weibull` distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto (x/\beta)^{\alpha-1}\exp\{-(x/\beta)^{\alpha}\}$,

where α is the shape parameter, and β is the scale parameter.

iii) `Lognormal` distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto \exp\{-(\log x - \alpha)^2/2\beta^2\}$,

where α is the mean, and β is the standard deviation of the logarithm.

iv) $\texttt{Inverse} - \texttt{Gaussian}$ distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto x^{-3/2}\exp\{-\alpha(x-\beta)^2/2x\beta^2\}$,

where α is the shape parameter, and β is the mean parameter.

Value

The function 'domnoplan' produces:

`total_domnoplan_sim`

Matrix array of dimension "J * (3 +2 D)* nSim" with predictions for the planned domains.

Column 1 - Indicator of the planned domains

Column 2 - T_j (totals of the planned domains)

Column 3 - N_j (composition of the planned domains)

Column 4 - $(4 + D - 1) - T_{d_j}$ (totals of unplanned domains, such that $T_j = \sum_{j=1}^{J} T_{d_j}$ (across d))

Column 5 - $(4 + D)$ to $(3 + 2 * D) - N_{d_j}$ (composition of unplanned domains, such that $N_j = \sum_{j=1}^{J} N_{d_j}$ (across d))

Author(s)

Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domnoplan_composicion, domnoplan_totalcomp.

Examples

```
#--------------------------------------------------------
# Uses of library:  "predfinitepop"
#--------------------------------------------------------

#             Read files with functions in R
rm(list=ls())

library("predfinitepop")

set.seed(12345)

#-------------------------------------------------------------
# DATA
#-------------------------------------------------------------
data(datasample)
data(popcomp)

# Sample (with 10% of the population)

datos <- datasample[[1]]
domplan_N <- popcomp[[1]]

#-------------------------------------------------------------
# MODEL PARAMETERS
#-------------------------------------------------------------

# a. Parameter "alpha" of the Dirichlet process
# (this is for J02 planned domains)
```

```
alphaDP <- matrix(0.03,2,1)

# b. vector with indicator variables for unplanned domains
colid_D <- t(as.matrix(c(6:8)))

# c. Matrix with parameters for unplanned domains
# (this is for J=2 planned domains)
alpha_D <- matrix(0.3,2,length(colid_D))

#------------------------------------------------------------
# CODE PARAMETERS AND INCORPORATION OF THE DISTRIBUTION
#------------------------------------------------------------
#  Incorporate each element ("g0_licitacion_in_J") according to the number of domains planned
#    g0_licitacion_in_1 <- list("distribution",teta1,esp(teta1))
#
# In this case we have two domains planned
g0_licitacion_in_1 <- list("Lognormal",c(6.198,0.840),.5)
g0_licitacion_in_2 <- list("Gamma",c(6.198,0.840),.5)

g0_licitacion_sal <- list(g0_licitacion_in_1,g0_licitacion_in_2)

# B. Number of Monte Carlo replicates of the predictive distributions
nSim <- 10

#------------------------------------------------------------
# TESTING "domnoplan.R"
#------------------------------------------------------------
date.ini <- date()
domnoplan_sim <- domnoplan_g0(datos,domplan_N,alphaDP,colid_D,alpha_D,nSim,g0_licitacion_sal)
date.fin <- date()

#
# --END --
```

`domnoplan_totalcomp` *Simulates Monte Carlo samples from the predictive distribution of the vector T_j across the D unplanned domains within a given planned domain.*

Description

This function simulates Monte Carlo samples from the predictive distribution of the vector T_j across the D unplanned domains in a given planned domain j.

Usage

```
domnoplan_totalcomp(datos_j,rho,ystar,phi,g0_licitacion_sal,N_Stil_domnoplan_j_sim,
nSim,colid_D)
```

Arguments

`datos_j` Data matrix with features and number of individuals in the sample $\mathcal{S}_j$

rho — (U X 2)-dimensional vector with weights associated with the sample ties 'ystar'

ystar — Sample ties y_i^* in the sample $\mathcal{S}_j$

phi — Probability weight asociated with G_{j0} (the continuos part of $\hat{G}_j$)

g0_licitacion_sal — Object list with the continuous component in $\hat{G}_j$

N_Stil_domnoplan_j_sim — (1 X D X nSim) matrix with Monte Carlo samples of the predictive distribution of the composition of $\tilde{\mathcal{S}}_j$

nSim — Number of Monte Carlo simulated replicates of the predictive distribution

colid_D — D-dimensional matrix array with the columns in 'datos' that correspond to the indicator variables of the planned domains (those indicator variables represent a partion of 'datos')

Value

The function 'domnoplan_totalcomp' produces an object list with three entries:

T_S_domnoplan — Composition of T_j for the planned domain j in sample $\mathcal{S}_j$

T_Stil_domnoplan_j — (1 X D X nSim)-dimensional array with simulated samples of the convolution for $T^{\tilde{\mathcal{S}}_j}$

domnoplan_totalcomp_sim — (1 X D X nSim)-dimensional array with samples from the predictive distribution of the composition of $\tilde{\mathcal{S}}_j$

Author(s)

Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domnoplan.

domplan — *Finite population inference on unplanned domains using reference data.*

Description

Generates Monte Carlo samples of the predictive distribution of totals of a finite population segmented in planned domains.

Usage

```
domplan(datos,datos_ant,domplan_N,alphaDP,inter,part,nSim)
```

Arguments

datos	(M X p)-dimensional array with positive entries for $\mathcal{S}_j$
datos_ant	(M X p)-dimensional refenrence array for calibration of G_0
domplan_N	Matrix array with counts of individuals in each planned domain
alphaDP	J-dimensional array with positive entries for the parameters of the Dirichlet process for F_j (with J being the number of planned domains)
inter	Tunning parameter for model comparison and selection (related to calibration of G_0)
part	Number of partitions for predictive cross-validation (related to calibration of G_0)
nSim	Number of Monte Carlo simulated replicates of the predictive distribution

Details

- datos : Represents the data sample of the target population, unplanned domains labeled.

It should contain the following columns:

"domplan" - Categories fot planned domains.

"y_i" - Actual individual measurements(for the moment, they must be positive) for the group of observation.

"n_i" - Number of individuals in the group (if the unit of observation in the sample are individuals, then "n_i" must be equal to 1)

- datos_ant : Represents the data of reference used to calibrate G_0.

The data must be labeled by domains planned. It should contain the following columns:

"domplan" - Planned categories of domains

"y_i" - Positive real and individual measurements of each group in the sample (when the units of observation are groups, "y_i" should be a per capita measurement)

- domplan_N : Represents counts (or reference population) of the target population, divided by the planned domains.

Tagged data must be labeled by domains planned. It should contain the following columns:

"domplan" - Categories fot planned domains.

"N_j" - Number of individuals in each planned domain.

Value

The function 'domplan' produces:

`total_domnoplan_sim`
: Matrix array of dimension "J * (3 +2 D)* nSim" with predictions for the planned domains.
Column 1 - Indicator of the planned domains
Column 2 - T_j (totals of the planned domains)
Column 3 - N_j (composition of the planned domains)

Author(s)

Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domplan_total

Examples

```
#--------------------------------------------------------
# Uses of library:  "predfinitepop"
#--------------------------------------------------------

#             Read files with functions in R
rm(list=ls())

library("predfinitepop")

set.seed(12345)

#-------------------------------------------------------------
# DATA
#-------------------------------------------------------------
data(dataref)
data(datasample)
data(popcomp)

# Sample (with 10% of the population)

datos <- datasample[[1]]
datos_ant <- dataref
domplan_N <- popcomp[[1]]

#-------------------------------------------------------------
```

```
# MODEL PARAMETERS
#-------------------------------------------------------------

# a. Parameter "alpha" of the Dirichlet process
# (this is for J02 planned domains)
alphaDP <- matrix(0.03,2,1)

# b. vector with indicator variables for unplanned domains
colid_D <- t(as.matrix(c(6:8)))

# c. Matrix with parameters for unplanned domains
# (this is for J=2 planned domains)
alpha_D <- matrix(0.3,2,length(colid_D))

#-------------------------------------------------------------
# CODE PARAMETERS
#-------------------------------------------------------------

# A. Interval for calibration of G_0
inter <- 3

# B. Number of groups for predictive validation and model comparison
part <- 20

# C. Number of Monte Carlo replicates of the predictive distributions
nSim <- 10

#-------------------------------------------------------------
# TESTING "domplan.R"
#-------------------------------------------------------------
date.ini <- date()
domplan_sim <- domplan(datos,datos_ant,domplan_N,alphaDP,inter,part,nSim)
date.fin <- date()

#
# --END --
```

`domplan_g0` *Finite population inference on unplanned domains using subjective G_j0.*

Description

Generates Monte Carlo samples of the predictive distribution of totals of a finite population segmented in planned domains, using a predefined set of baseline distributions $(G_{j0})_{j=1}^{J}$.

Usage

```
domplan_g0(datos,domplan_N,alphaDP,nSim,g0_licitacion_sal)
```

Arguments

datos	$(M$ X $p)$-dimensional array with positive entries for $\mathcal{S}_j$
domplan_N	Matrix array with counts of individuals in each planned domain

alphaDP J-dimensional array with positive entries for the parameters of the Dirichlet process for F_j (with J being the number of planned domains)

nSim Number of Monte Carlo simulated replicates of the predictive distribution

g0_licitacion_sal
(J X 1) object list, each entry is itself an object list associated with each G_{j0} for the J planned domains. The first element for arch G_{j0} should be the name of the chosen distribution (see details below for the admissible alternatives), the second element should be a vector object with the parameters associated with "distribution", and the third element should be its associated expectation

Details

- datos : Represents the data sample of the target population, unplanned domains labeled.

It should contain the following columns:

"domplan" - Categories fot planned domains.

"y_i" - Actual individual measurements/putcomes (for the moment, they must be positive) for the group of observation.

"n_i" - Number of individuals in the group (if the unit of observation in the sample are individuals, then "n_i" must be equal to 1)

- domplan_N : Represents the counts (or number of individuals) in the target population, segmented in the planned domains.

Data must be labeled by the planned domain. It should contain the following columns:

"domplan" - Categories fot planned domains.

"N_j" - Number of individuals in each planned domain.

- g0_licitacion_sal: Chose one and only one of the distribution,

i) Gamma

ii) Weibull

iii) Lognormal

iv) Inverse-Gaussian

Parameterizations for distribution:

i) `Gamma` distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto x^{\alpha-1} \exp\{-x/\beta\}$,

where α is the shape parameter, and β is the scale parameter.

ii) `Weibull` distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto (x/\beta)^{\alpha-1} \exp\{-(x/\beta)^{\alpha}\}$,

where α is the shape parameter, and β is the scale parameter.

iii) `Lognormal` distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto \exp\{-(\log x - \alpha)^2/2\beta^2\}$,

where α is the mean, and β is the standard deviation of the logarithm.

iv) $\texttt{Inverse} - \texttt{Gaussian}$ distribution, with parametetrs $\boldsymbol{\theta} = c(\alpha > 0, \beta > 0)$ and density function

$f(x) \propto x^{-3/2} \exp\{-\alpha(x-\beta)^2/2x\beta^2\}$,

where α is the shape parameter, and β is the mean parameter.

Value

The function 'domplan' produces:

`total_domnoplan_sim`
: Matrix array of dimension "J * (3 +2 D)* nSim" with predictions for the planned domains.
Column 1 - Indicator of the planned domains
Column 2 - T_j (totals of the planned domains)
Column 3 - N_j (composition of the planned domains)

Author(s)

Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domplan_total

Examples

```
#--------------------------------------------------------
# Uses of library:  "predfinitepop"
#--------------------------------------------------------

#             Read files with functions in R
rm(list=ls())

#library("predfinitepop")

set.seed(12345)

#-------------------------------------------------------------
# DATA
#-------------------------------------------------------------
data(datasample)
data(popcomp)

# Sample (with 10% of the population)

datos <- datasample[[1]]
domplan_N <- popcomp[[1]]

#-------------------------------------------------------------
# MODEL PARAMETERS
#-------------------------------------------------------------
```

```
# a. Parameter "alpha" of the Dirichlet process
# (this is for J02 planned domains)
alphaDP <- matrix(0.03,2,1)

# b. vector with indicator variables for unplanned domains
colid_D <- t(as.matrix(c(6:8)))

# c. Matrix with parameters for unplanned domains
# (this is for J=2 planned domains)
alpha_D <- matrix(0.3,2,length(colid_D))

#-------------------------------------------------------------
# CODE PARAMETERS AND INCORPORATION OF THE DISTRIBUTION
#-------------------------------------------------------------
#  Incorporate each element ("g0_licitacion_in_J") according to the number of domains planned
#    g0_licitacion_in_1 <- list("distribution",teta1,esp(teta1))
#
# In this case we have two domains planned
g0_licitacion_in_1 <- list("Lognormal",c(6.198,0.840),.5)
g0_licitacion_in_2 <- list("Gamma",c(6.198,0.840),.5)

g0_licitacion_sal <- list(g0_licitacion_in_1,g0_licitacion_in_2)

# B. Number of Monte Carlo replicates of the predictive distributions
nSim <- 10

#-------------------------------------------------------------
# TESTING "domplan.R"
#-------------------------------------------------------------
date.ini <- date()
domplan_sim <- domplan_g0(datos,domplan_N,alphaDP,nSim,g0_licitacion_sal)
date.fin <- date()

#
# --END --
```

`domplan_total` *Monte Carlo simulation of total for a given planned domain.*

Description

Simulates Monte Carlo samples of size '$nSim$' of the final distribution for the total for the given planned domain $\mathcal{P}_j$.

Usage

```
domplan_total(datos_j,rho,ystar,phi,g0_licitacion_sal,N_Stil_j,nSim)
```

Arguments

datos_j	Data matrix with features and number of individuals in the sample '$\mathcal{S}_j$'
rho	(U X 2)-dimensional vector with weights associated with the sample ties '$ystar$'
ystar	Sample ties y_i^*, in the sample $\mathcal{S}_j$

phi	Probability weight asociated with 'G_{j0}'(the continuos part of '$\hat{G}_j$')
g0_licitacion_sal	Object list with the continuous component in '$\hat{G}_j$'
N_Stil_j	Composition of individuals out of '$\mathcal{S}_j$'
nSim	Number of Monte carlosimulated replicates of the predictive distribution

Value

The function 'domplan_total' produces and object list with two entries:

T_S_j	Composition of 'T_j' of the given planned domains in the sample 'S_j'
domplan_total_sim	(1 * nSim) matrix with samples from the predictive distribution of the total 'T_Stil_j'

Author(s)

Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

domnoplan

---

g0_licitacion — *Prior elicitation of G_j0 using reference data.*

Description

This function computes the prior elicitation of G_{j0} using reference data for the planned domain j. G_{j0} is used by the SSM as the baseline function. The distribution is elicited using a predictive cross-validation procedure for model comparison and selection among the following alternatives: Gamma, Weibull, Lognormal and Inverse-Gaussian. The function also computes the expectation of the chose distribution.

Usage

```
g0_licitacion(datos_ant,inter,part)
```

Arguments

`datos_ant`	Reference data for calibration of G_{j0}
`inter`	Tunning parameter for model comparison and selection
`part`	Number of groups for predictive validation. The larger the number of groups the fewer observations are to be removed for calibration

Details

datos_ant: Data collected in a previous time or from another source for each planned domain. It must contain two columns, labeled by 'domplan' and 'y_i' 'domplan' - Labels of planned domains 'y_i' - individual outcome (for the sake of implementation, this must be positive)

Value

The function 'g0_licitacion' produces an object list with three elements:

`distribution`	List arrangement: String object for 'distribution' i) Gamma ii) Weibull iii) Log-normal iv) Inverse Gaussian
`theta`	Vector of parameters associated with 'distribution'
`mu`	Expected value of 'distribution'

Author(s)

Adriana Roldan-Rodriguez

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

`g0_simulacion` *Computes Monte Carlo samples from a given G_j0.*

Description

This function computes Monte Carlo samples from a given continuous G_{j0}.

Usage

```
g0_simulacion(g0_licitacion_sal,nSim)
```

Arguments

`g0_licitacion_sal`
Object list with three elements (produced with 'g0_licitacion'):

a) String object for 'distribution'
i) Gamma
ii) Weibull
iii) Log-normal
iv) Inverse-Gaussian
b) theta - Vector of parameters associated with 'distribution'
c) mu - Expected value of 'distribution'

nSim Number of Monte Carlo simulations

Value

This function 'g0_simulacion' produces:

g0_simulacion_sal
(nSim X 1) dimensional array with simulated data

Author(s)

Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

ghat_simulacion

---

ghat_simulacion *Simulates Monte Carlo samples from $\hat{G}$_j.*

Description

This function simulates Monte Carlo samples from $\hat{G}_j$.

Usage

```
ghat_simulacion(rho,ystar,phi,g0_licitacion_sal,N_Stil_j)
```

Arguments

rho
: (U X 2)-dimensional vector with weights associated with 'ystar'

ystar
: Sample ties y_i^* in $\mathcal{S}_j$

phi
: Probability weight asociated with G_{j0} (the continuos part of $\hat{G}_j$)

g0_licitacion_sal
: Object list with three elements (produced with 'g0_licitacion'):

a)String object for 'distribution'
i) Gamma
ii) Weibull
iii) Log-normal
iv) Inverse-Gaussian
b) theta - Parameters associated with 'distribution'
c) mu - Expected value of 'distribution'

`N_Stil_j` Number of samples to simulate (number of individuals in $\tilde{S}_j$) out of the sample

Value

'ghat_simulacion' produces:

`ghat_simulacion_sal`
(N_Stil_j X 1) dimensional vector with simulated data

Author(s)

Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

`g0_simulacion`, `gstar_simulacion`.

`gstar_simulacion` *Generates Monte Carlo samples from the discrete component of $\hat{G}$_j*

Description

This function generates Monte Carlo samples from the discrete component of $\hat{G}_j$ for the Dirichlet process prior in the planned domain j.

Usage

```
gstar_simulacion(rho,ystar,nSim)
```

Arguments

rho	(U X 2)-dimensional vector with weights associated with 'ystar'
ystar	Sample ties y_i^* in $\mathcal{S}_j$
nSim	Number of Monte Carlo simulations

Value

The function 'gstar_simulacion' produces:

gstar_simulacion_sal
(nSim X 1) dimensional vector with simulated data

Author(s)

Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

See Also

ghat_simulacion

---

pesosDP *Computes the weights associated with the predictive distribution $\hat{G}$_j for a given planned domain j under the Dircihlet processes prior.*

Description

This function computes the weights associated with the predictive distribution $\hat{G}_j$ for a given planned domain j, using the sample ties (y_j^*), under the Dircihlet processes prior.

Usage

```
pesosDP(unicos_sal,alphaDP)
```

Arguments

unicos_sal	(U X 2) matrix with sample ties and asociated frequencies (produced with 'unicos')
alphaDP	Scale parameter of the Dirichlet process

Details

The matrix with sample ties 'unicos_sal' must include a column named 'm_k' for the frequencies of the sample ties.

Value

The function 'pesosDP' produces an object list with two elements:

`rho`	(U X 2) dimensional vector with weights associated with 'ystar' (the sample ties)
`phi`	Probability weight asociated with G_{j0} (the continuos part of $\hat{G}_j$)

Author(s)

Sergio I. Olivares-Guzman, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

`pinvgauss` *Inverse Gaussian distribution - Probability.*

Description

Computes de c.d.f. function for the inverse Gaussian distribution.

Usage

```
pinvgauss(q, m, s)
```

Arguments

`q`	p-dimensional vector with (0,1)-valued entries
`m`	Mean parameter
`s`	Scale parameter

Author(s)

Jim Lindsey <jlindsey@luc.ac.be>

References

- "Utilities for Nonlinear Regression and Repeated Measurements Models." Package "rmutil" version 1.0.

`popcomp`	*Composition of the simulated population segemented in planned and unplanned domains.*

Description

This object corresponds to a list with the composition of the simulated population across planned and unplanned domains. See Martinez-Ovando et al. (2014) for further details.

Usage

```
data(popcomp)
```

Format

The format is: List of 3 $: num [1:2] 800 1200 $: num [1:3] 980 800 220

Source

Simulated data. See Martinez-Ovando et al. (2014) for further details.

References

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

Examples

```
data(popcomp)
```

`poptot`	*Totals of the simulated population segemented in planned and unplanned domains.*

Description

This object corresponds to a list with the totals of the simulated population across planned and unplanned domains. See Martinez-Ovando et al. (2014) for further details.

Usage

```
data(poptot)
```

Format

The format is:

[[1]] Composition of the population across planned domains (working format).

[[2]] Composition of the population across planned domains.

[[3]] Composition of the population across unplanned domains .

Source

Simulated data. See Martinez-Ovando et al. (2014) for further details.

References

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

Examples

```
data(poptot)
```

`predfinitepop` *Predictive Inference on Totals and Averages of Finite Populations Segmented in Planned and Unplanned Domains.*

Description

This package computes Bayesian predictive inference on totals and averages of finite populations segmented in planned and unplanned domains. Inference is based on Bayesian nonparametric methods using species-sampling models (specifically the Dirichlet process).

Details

Package:	predfinitepop
Type:	Statistical
Version:	v-1
Date:	2014-04-24

Author(s)

Juan Carlos Martinez-Ovando
Sergio I. Olivares-Guzman
Adriana Roldan-Rodriguez

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as

Banco de Mexico Working Paper 2014-04.

`qinvgauss` *Inverse Gaussian distribution - Quantile.*

Description

Computes the quantiles for the inverse Gaussian distribution.

Usage

```
qinvgauss(p, m, s)
```

Arguments

`p`	p-dimensional vector with (0,1)-valued entries
`m`	Mean parameter
`s`	Scale parameter

Author(s)

Jim Lindsey <jlindsey@luc.ac.be>

References

- "Utilities for Nonlinear Regression and Repeated Measurements Models." Package "rmutil" version 1.0.

`randDirichlet` *Simulates samples from the Dirichlet distribution with (px1) vector parameter α.*

Description

This function simulates samples from the Dirichlet distribution with (px1) vector parameter α.

Usage

```
randDirichlet(alpha,n)
```

Arguments

`alpha`	p-dimensional vector with positive entries
`n`	Number of simulated replicates

Value

This function produces:

`randDir`	(n X p) matrix with 'n' simulated replicates

Author(s)

Juan Carlos Martinez-Ovando

References

- "Non-Uniform Random Variate Generation", Luc Devroy (1986), Berlin: Springer-Verlag.

`rinvgauss` *Inverse Gaussian distribution - Simulation.*

Description

Simulates reandom variables from the inverse Gaussian distribution.

Usage

```
rinvgauss(n=1, m, s)
```

Arguments

`n`	Number of simulatios
`m`	Mean parameter
`s`	Scale parameter

Author(s)

Jim Lindsey <jlindsey@luc.ac.be>

References

- "Utilities for Nonlinear Regression and Repeated Measurements Models." Package "rmutil" version 1.0.

`unicstar` *Identifies the sample ties in $\mathcal{S}$_j and computes their associated sample frequencies.*

Description

This function identifies the sample ties (y^*_k) in $\mathcal{S}_j$ and computes their associated sample frequencies.

Usage

```
unicstar(datos_j)
```

Arguments

`datos_j`	Data matrix with individual measurements in "$\mathcal{S}_j$"

Details

The column with individual measurements in "datos_j" should be named "y_i"

Value

The function 'unicstar' produces an object list with two elements:

`U`	Number of sample ties in "$\mathcal{S}_j$"
`unicstar`	Data matrix with sample ties 'y_i' and asociated frequencies

Author(s)

Sergio I. Olivares-Guzman, Adriana Roldan-Rodriguez, Juan Carlos Martinez-Ovando

References

- "A Bayesian nonparametric framework to inference on totals of finite populations," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2013. Contributions of Young Researchers to Bayesian Statistics (eds. E. Lazarone & F. Ieva), Chapter 15 Springer.

- "Predictive inference on finite populations segmented in planned and unplanned domains," Martinez-Ovando, J. C., Olivares-Guzman, S. I., Roldan-Rodriguez, A., 2014. Submitted. Also available as Banco de Mexico Working Paper 2014-04.

Index