Package 'prnsamplr'

April 17, 2025

Type Package

Title Permanent Random Number Sampling

Version 1.1.0

Maintainer Kira Coder Gylling <kira.gylling@gmail.com>

Description Survey sampling using permanent random numbers (PRN's). A solution to the problem of unknown overlap between survey samples, which leads to a low precision in estimates when the survey is repeated or combined with other surveys. The PRN solution is to supply the U(0, 1) random numbers to the sampling procedure, instead of having the sampling procedure generate them. In Lindblom (2014) <doi:10.2478/jos-2014-0047>, and therein cited papers, it is shown how this is carried out and how it improves the estimates. This package supports two common fixed-size sampling procedures (simple random sampling and probability-proportional-to-size sampling) and includes a function for transforming the PRN's in order to control the sample overlap.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 4.2)

Imports rlang (>= 1.1.0), stats (>= 4.2)

Suggests data.table (>= 1.0.0), stringr (>= 1.4.0), testthat (>= 3.0.0), tibble (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Kira Coder Gylling [aut, cre] (<https://orcid.org/0000-0002-9070-8381>)

Repository CRAN

Date/Publication 2025-04-17 11:50:02 UTC

Contents

`prnsamplr-package` *Permanent Random Number Sampling in R*

Description

This package provides two functions for drawing stratified PRN-assisted samples: `srs` and `pps`. The former – simple random sampling – assumes that each unit k in a given stratum h is equally likely to be sampled, with inclusion probability

$$\pi_k = \frac{n_h}{N_h}$$

for each stratum h. The function then samples the n_h elements with the smallest PRN's, for each stratum h.

The latter – Pareto πps sampling – assumes that large units are more likely to be sampled than small units. The function approximates this unknown inclusion probability as

$$\lambda_k = n_h \frac{x_k}{\sum_{i=1}^{n_h} x_i},$$

where x_k is a size measure, and samples the n_h elements with the smallest values of

$$Q_k = \frac{PRN_k(1-\lambda_k)}{\lambda_k(1-PRN_k)},$$

for each stratum h.

These two functions can be run standalone or via the wrapper function `samp`. Input to the functions is the sampling frame, stratification information and PRN's given as variables on the frame, and in the case for `pps` also a size measure given as variable on the frame. Output is a copy of the sampling frame containing sampling information, and in the case for `pps` also containing λ and Q.

Provided is also a function `transformprn` via which it is possible to select where to start counting and in which direction when enumerating the PRN's in the sampling routines. This is done by specifying starting point and direction to `transformprn` and then calling `srs` or `pps` on its output.

Finally, an example dataset is provided that can be used to illustrate the functionality of the package.

Author(s)

Maintainer: Kira Coder Gylling <kira.gylling@gmail.com> (ORCID)

References

Lindblom, A. (2014). "On Precision in Estimates of Change over Time where Samples are Positively Coordinated by Permanent Random Numbers." *Journal of Official Statistics*, vol.30, no.4, 2014, pp.773-785. https://doi.org/10.2478/jos-2014-0047.

See Also

srs, pps, samp, transformprn, ExampleData

Examples

```
dfSRS <- srs(
  frame = ExampleData,
  nsamp = ~nsample,
  stratid = ~stratum,
  prn = ~rands
)

dfPPS <- pps(
  frame = ExampleData,
  nsamp = ~nsample,
  stratid = ~stratum,
  prn = ~rands,
  size = ~sizeM
)

dfPRN <- transformprn(
  frame = ExampleData,
  prn = ~rands,
  direction = "U",
  start = 0.2
)
```

`ExampleData` *ExampleData*

Description

Artificial dataset to be used with `samp` and `transformprn`.

Usage

```
ExampleData
```

Format

'ExampleData'

A data frame with 40,000 rows and 6 columns:

`stratum` a character vector

`id` a numeric vector

`npopul` a numeric vector

`nsample` a numeric vector

`rands` a numeric vector

`sizeM` a numeric vector

Source

Ad-hoc simulation in base R.

See Also

prnsamplr, samp, srs, pps, transformprn

`pps` *Stratified probability-proportional-to-size sampling*

Description

Stratified probability-proportional-to-size (Pareto PiPS) sampling using permanent random numbers. Can also be used for non-stratified Pareto PiPS using a dummy stratum taking the same value for each object.

Usage

```
pps(frame, stratid, nsamp, prn, size)
```

Arguments

`frame`	Data frame (or data.table or tibble) containing the elements to sample from.
`stratid`	Variable in `frame` containing the strata.
`nsamp`	Variable in `frame` containing the sample sizes.
`prn`	Variable in `frame` containing the permanent random numbers.
`size`	Variable in `frame` containing the size measure.

Value

A copy of the input sampling frame together with the boolean variable `sampled`, indicating sample inclusion, as well as a numeric variable `lambda` containing the estimated first-order inclusion probabilities and the numeric variable

$$Q = \frac{prn(1 - lambda)}{lambda(1 - prn)}$$

that determines which elements are sampled.

See Also

prnsamplr, samp, srs, transformprn, ExampleData

Examples

```
dfOut <- pps(
  frame = ExampleData,
  nsamp = ~nsample,
  stratid = ~stratum,
  prn = ~rands,
  size = ~sizeM
)
```

`samp` *Stratified permanent random number sampling*

Description

Wrapper for stratified simple random sampling (SRS) and probability-proportional-to-size (PPS) sampling using permanent random numbers. Can also be used for non-stratified sampling using a dummy stratum taking the same value for each object.

Usage

```
samp(method, frame, ...)
```

Arguments

`method`	`pps` or `srs`.
`frame`	Data frame (or data.table or tibble) containing the elements to sample from.
`...`	Further method-specific arguments.

Value

A copy of the input data frame together with the boolean variable `sampled`, as well as the numeric variables `lambda` and `Q` when pps is used.

See Also

prnsamplr, srs, pps, transformprn, ExampleData

Examples

```
dfOut <- samp(
  method = pps,
  frame = ExampleData,
  nsamp = ~nsample,
  stratid = ~stratum,
  prn = ~rands,
  size = ~sizeM
)

dfOut <- samp(
  method = srs,
  frame = ExampleData,
  nsamp = ~nsample,
  stratid = ~stratum,
  prn = ~rands
)
```

---

srs *Stratified simple random sampling*

Description

Stratified simple random sampling (SRS) using permanent random numbers. Can also be used for non-stratified SRS using a dummy stratum taking the same value for each object.

Usage

```
srs(frame, stratid, nsamp, prn)
```

Arguments

frame	Data frame (or data.table or tibble) containing the elements to sample from.
stratid	Variable in frame containing the strata.
nsamp	Variable in frame containing the sample sizes.
prn	Variable in frame containing the permanent random numbers.

Value

A copy of the input sampling frame together with the boolean variable sampled, indicating sample inclusion.

See Also

prnsamplr, samp, pps, transformprn, ExampleData

Examples

```
dfOut <- srs(
  frame = ExampleData,
  nsamp = ~nsample,
  stratid = ~stratum,
  prn = ~rands
)
```

`transformprn` *Permanent random number transformation*

Description

Transformation of the permanent random numbers used in the sampling procedure, to control the overlap between samples, and thus control the sample coordination. The method used is specified in Lindblom and Teterukovsky (2007).

Usage

```
transformprn(frame, prn, direction, start)
```

Arguments

`frame`	Data frame (or data.table or tibble) containing the elements to sample from.
`prn`	Variable in `frame` containing the permanent random numbers.
`direction`	Direction for the enumeration. "U" or "R" for upwards, or equivalently to the right on the real-number line. "D" or "L" for downwards, or equivalently to the left on the real-number line.
`start`	Starting point for the transformation. For SRS this corresponds to the point at which one wants to start sampling.

Value

A copy of the input data frame with the permanent random numbers transformed according to specification, along with the numeric variable `prn.old` containing the non-transformed permanent random numbers.

References

Lindblom, A. and Teterukovsky, A. (2007). "Coordination of Stratified Pareto pps Samples and Stratified Simple Random Samples at Statistics Sweden." In *Papers presented at the ICES-III, June 18-21, 2007, Montreal, Quebec, Canada.*

See Also

prnsamplr, samp, srs, pps, ExampleData

Examples

```
dfOut <- transformprn(
  frame = ExampleData,
  prn = ~rands,
  direction = "U",
  start = 0.2
)
```

Index