Package 'aggutils'

August 22, 2023

Title Utilities for Aggregating Probabilistic Forecasts

Version 1.0.2

URL https://github.com/forecastingresearch/aggutils

BugReports https://github.com/forecastingresearch/aggutils/issues

Description Provides several methods for aggregating probabilistic forecasts. You have a group of people who have made probabilistic forecasts for the same event. You want to take advantage of the ``wisdom of the crowd'' and combine these forecasts in some sensible way. This package provides implementations of several strategies, including geometric mean of odds, an extremized aggregate (Neyman, Roughgarden (2021) <doi:10.1145/3490486.3538243>), and ``high-density trimmed mean'' (Powell et al. (2022) <doi:10.1037/dec0000191>).

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports stats, docstring

Suggests testthat

NeedsCompilation no

Author Molly Hickman [aut, cre] (<https://orcid.org/0009-0007-5144-0080>),
Zach Jacobs [aut]

Maintainer Molly Hickman <molly@forecastingresearch.org>

Repository CRAN

Date/Publication 2023-08-22 18:30:12 UTC

R topics documented:

---

geoMeanCalc   *Geometric Mean*

Description

Calculate the geometric mean of a vector of forecasts. We handle 0s by replacing them with the qth quantile of the non-zero forecasts.

Usage

```
geoMeanCalc(x, q = 0.05)
```

Arguments

x	Vector of forecasts in 0 to 100 range (%)
q	The quantile to use for replacing 0s (between 0 and 1)

Value

(numeric) The geometric mean of the vector

Note

agg(a) + agg(not a) does not sum to 1 for this aggregation method.

---

geoMeanOfOddsCalc   *Geometric Mean of Odds*

Description

Convert probabilities to odds, and calculate the geometric mean of the odds. We handle 0s by replacing them with the qth quantile of the non-zero forecasts, before converting.

Usage

```
geoMeanOfOddsCalc(x, q = 0.05, odds = FALSE)
```

Arguments

x	A vector of forecasts (probabilities! unless odds = TRUE)
q	The quantile to use for replacing 0s (between 0 and 1)
odds	Whether x is already in odds form (TRUE) or probabilities

Value

(numeric) The geometric mean of the odds

Note

agg(a) + agg(not a) does not sum to 1 for this aggregation method.

---

hd_trim   *Highest-Density Trimmed Mean*

Description

From Powell et al. (2022) doi:10.1037/dec0000191. You find the shortest interval containing (1-p) * 100% of the data and take the mean of the forecasts within that interval.

Usage

```
hd_trim(x, p = 0.1)
```

Arguments

x	Vector of forecasts in 0 to 100 range (%)
p	The proportion of forecasts to trim (between 0 and 1)

Value

(numeric) The highest-density trimmed mean of the vector

Note

As p gets bigger this acts like a mode in a similar way to the symmetrically-trimmed mean acting like a median.

---

neymanAggCalc   *Neyman Aggregation (Extremized)*

Description

Takes the arithmetic mean of the log odds of the forecasts, then extremizes the mean by a factor d, where d is

$(n*(sqrt((3n^2) - (3n) + 1) - 2))/(n^2 - n - 1)$

where n is the number of forecasts.

Usage

```
neymanAggCalc(x)
```

Arguments

x   Vector of forecasts in 0 to 100 range (%)

Value

(numeric) The extremized mean of the vector

References

Neyman, E. and Roughgarden, T. (2021). Are you smarter than a random expert? The robust aggregation of substitutable signals: doi:10.1145/3490486.3538243. Also Jaime Sevilla's EAF post "Principled extremizing of aggregated forecasts."

---

preprocess   *Preprocessing function for agg methods*

Description

This does the preprocessing steps that all the agg methods have in common.

Usage

```
preprocess(x, q = 0)
```

Arguments

x   A vector of forecasts

q   The quantile to use for replacing 0s and 1s (between 0 and 1)

Value

A vector of forecasts with 0s are replaced by the qth quantile and 100s are replaced by the (1 - q)th quantile.

Note

Assumes forecasts are in the range 0 to 100, inclusive.

---

soften_mean *Soften the mean.*

Description

If the mean is > .5, trim the top trim%; if < .5, the bottom trim%. Return the new mean (i.e. soften the mean).

Usage

```
soften_mean(x, p = 0.1)
```

Arguments

x	Vector of forecasts in 0 to 100 range (%)
p	The proportion of forecasts to trim from each end (between 0 and 1)

Value

(numeric) The softened mean of the vector

Note

This goes against usual wisdom of extremizing the mean, but performs well when the crowd has some overconfident forecasters in it.

---

trim *Trimmed mean*

Description

Trim the top and bottom (p*100)% of forecasts

Usage

```
trim(x, p = 0.1)
```

Arguments

x	Vector of forecasts in 0 to 100 range (%)
p	The proportion of forecasts to trim from each end (between 0 and 1)

Value

(numeric) The trimmed mean of the vector

Index