Package 'braggR'

October 12, 2022

Type Package

Title Calculate the Revealed Aggregator of Probability Predictions

Version 0.1.1

Description Forecasters predicting the chances of a future event may disagree due to
differing evidence or noise. To harness the collective evidence of the crowd,
Ville Satopää (2021) ``Regularized Aggregation of One-off Probability Predictions''
<https:
//papers.ssrn.com/sol3/papers.cfm?abstract_id=3769945> proposes a Bayesian
aggregator that is regularized by analyzing the forecasters' disagreement and ascribing
over-dispersion to noise. This aggregator requires no user intervention and can be computed
efficiently even for a large numbers of predictions. The author evaluates
the aggregator on subjective probability predictions collected during
a four-year forecasting tournament sponsored by the US intelligence community.
The aggregator improves the accuracy of simple averaging by around 20% and
other state-of-the-art aggregators by 10-25%. The advantage stems almost
exclusively from improved calibration. This aggregator -- know as ``the revealed
aggregator'' -- inputs a) forecasters' probability predictions (p) of a future binary event
and b) the forecasters' common prior (p0) of the future event. In this R-package,
the function sample_aggregator(p,p0,...) allows the user to calculate the revealed
aggregator. Its use is illustrated with a simple example.

License GPL-2

Copyright (c) Ville Satopaa

Encoding UTF-8

RoxygenNote 7.1.1

LinkingTo Rcpp

Imports Rcpp

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation yes

Author Ville Satopää [aut, cre, cph]

Maintainer Ville Satopää <ville.satopaa@gmail.com>

Repository CRAN

Date/Publication 2021-05-29 13:10:02 UTC

R topics documented:

sample_aggregator	*Revealed Aggregator*

Description

This function allows the user to compute the revealed aggregator from *Satopää, V.A. (2021): Regularized Aggregation of One-off Probability Predictions*. The current version of the paper is available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3769945.

Usage

```
sample_aggregator(
  p,
  p0 = NULL,
  alpha = NULL,
  beta = NULL,
  a = 1/2,
  b = 1/2,
  num_sample = 1e+06,
  burnin = num_sample/2,
  thin = 1,
  seed = 1
)
```

Arguments

p	Vector of $K \geq 2$ forecasters' probability estimates of a future binary event. These values represent probability predictions and must be strictly between 0 and 1.
p0	The forecasters' common prior. This represents a probability prediction based on some of the forecasters' common evidence and must be strictly between 0 and 1.
alpha, beta	The shape and scale parameters of the prior beta distribution of the common prior. If omitted, the sampler uses the fixed common prior given by p0. However, if alpha and beta are provided, they must be strictly positive. In this case, the common prior p0 will be treated as a random variable and sampled along with the other model parameters.
a, b	The parameters for the prior distribution of (ρ, γ, δ) in *Satopää, V.A. (2021)*. The default choice a = 1/2 and b = 1/2 gives the Jeffreys' independence prior. If p_0 is not equal to 0.5, then a = 1 and b = 1/2 give the Jeffreys' prior.

num_sample	The number of posterior samples to be drawn. This does not take into account burnin and thinning.
burnin	The number of the initial `num_sample` posterior draws that are discarded for burnin. This value cannot exceed `num_sample`.
thin	After `burnin` draws have been discarded, the final sample is formed by keeping every `thin`'th value. To ensure that the final sample holds at least two draws, `thin` can be at most `(num_sample-burnin)/2`.
seed	The seed value for random value generation.

Value

A data frame with rows representing posterior draws of $(p*, \rho, \gamma, \delta, p0)$. The columns are:

- `aggregate`: The posterior samples of the oracle aggregator $p*$. The average of these values gives the revealed aggregator p''. The 95% interval of these values gives the 95% credible interval of the oracle aggregator.
- `rho`: The posterior samples of the forecasters' shared evidence, ρ.
- `gamma`: The posterior samples of the forecasters' total evidence, γ. The difference `gamma-rho` gives the posterior samples of the forecasters' rational disagreement.
- `delta`: The posterior samples of the forecasters' total evidence plus noise, δ. The difference `delta-gamma` gives the posterior samples of the forecasters' irrational disagreement.
- `p0`: The posterior samples of the forecasters' common prior. If a beta prior distribution is not specified via the arguments `alpha` and `beta`, then all elements of this column are equal to the fixed common prior given by the `p0` argument.

Examples

```
# Illustration on Scenario B in Satopää, V.A. (2021).
# Forecasters' probability predictions:
p = c(1/2, 5/16, 1/8, 1/4, 1/2)

# Aggregate with a fixed common prior of 0.5.
# Sample the posterior distribution:
post_sample = sample_aggregator(p, p0 = 0.5, num_sample = 10^6, seed = 1)
# The posterior means of the model parameters:
colMeans(post_sample[,-1])
# The posterior mean of the oracle aggregator, a.k.a., the revealed aggregator:
mean(post_sample[,1])
# The 95% credible interval for the oracle aggregator:
quantile(post_sample[,1], c(0.025, 0.975))


# Aggregate based a uniform distribution on the common prior
# Recall that Beta(1,1) corresponds to the uniform distribution.
# Sample the posterior distribution:
post_sample = sample_aggregator(p, alpha = 1, beta = 1, num_sample = 10^6, seed = 1)
# The posterior means of the oracle aggregate and the model parameters:
colMeans(post_sample)
```

Index

sample_aggregator,