

# Package: stackr (via r-universe)

November 2, 2024

**Title** Create Mixture Models From Predictive Samples

**Version** 0.1.0

**Description** The `stackr` package provides an easy way to combine predictions from individual time series or panel data models to an ensemble. `stackr` stacks (Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman (2018) <[doi:10.1214/17-BA1091](https://doi.org/10.1214/17-BA1091)>) Models according to the Continuous Ranked Probability Score (CRPS) (Tilmann Gneiting & Adrian E Raftery (2007) <[doi:10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437)>) over k-step ahead predictions. It is therefore especially suited for timeseries and panel data. A function for leave-one-out CRPS may be added in the future. Predictions need to be predictive distributions represented by predictive samples. Usually, these will be sets of posterior predictive simulation draws generated by an MCMC algorithm. Given some training data with true observed values as well as predictive samples generated from different models, `crps_weights` finds the optimal (in the sense of minimizing expected cross-validation predictive error) weights to form an ensemble from these models. Using these weights, `mixture_from_samples` can then provide samples from the optimal model mixture by drawing from the predictive samples of the individual models in the correct proportion. This gives a mixture model solely based on predictive samples and is in this regard superior to other ensembling techniques like Bayesian Model Averaging.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Biarch** true

**Depends** R (>= 3.5.0)

**Imports** data.table, methods, Rcpp (>= 0.12.0), RcppParallel (>= 5.0.1), rstan (>= 2.18.1), rstantools (>= 2.4.0)

**LinkingTo** BH (>= 1.66.0), Rcpp (>= 0.12.0), RcppEigen (>= 0.3.3.3.0),

RcppParallel ( $\geq 5.0.1$ ), rstan ( $\geq 2.18.1$ ), StanHeaders ( $\geq 2.18.0$ )

**SystemRequirements** GNU make

**RoxygenNote** 7.3.2

**Suggests** knitr, rmarkdown, testthat ( $\geq 3.0.0$ ), scoringutils

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**Repository** <https://epiforecasts.r-universe.dev>

**RemoteUrl** <https://github.com/epiforecasts/stackr>

**RemoteRef** v0.1.0

**RemoteSha** 6ac296722faf603d5504f65b564bf92183c0bcdb

## Contents

crps_weights . . . . .	2
mixture_from_samples . . . . .	3
<b>Index</b>	<b>5</b>

---

crps_weights	<i>Obtain CRPS stacking weights</i>
--------------	-------------------------------------

---

## Description

given true values and predictive samples from different models, ‘crps\_weights’ returns the stacking weights which produce the ensemble that minimises the Continuous Ranked Probability Score (CRPS).

## Usage

```
crps_weights(data, lambda = NULL, gamma = NULL, dirichlet_alpha = 1.001)
```

## Arguments

data a data.frame with the following entries:

- observed, the true observed values
- predicted, predicted values corresponding to the true values in observed
- model, the name of the model used to generate the corresponding predictions
- geography (optional), the regions for which predictions are generated. If geography is missing, it will be assumed there are no geographical differences to take into account. Internally, regions will be ordered alphabetically
- date (the date of the corresponding prediction / true value). Also works with numbers to indicate timesteps

lambda	weights given to timepoints. If lambda is NULL, the default gives more weight to recent time points with $\text{lambda}[t] = 2 - (1 - t / T)^2$ . Note that elements of lambda need not necessarily sum up to one as the stan model automatically constraints the final weights to sum to one irrespective of lambda. lambda = "equal" uses equal weights
gamma	weights given to regions. If gamma is NULL the default is equal weights for the regions. Weights are mapped to regions alphabetically, so make sure that the the weights correspond to the regions in alphabetical order.
dirichlet_alpha	prior for the weights. Default is 1.001

**Value**

returns a vector with the model weights

**References**

Strictly Proper Scoring Rules, Prediction, and Estimation, Tilmann Gneiting and Adrian E. Raftery, 2007, Journal of the American Statistical Association, Volume 102, 2007 - Issue 477

Using Stacking to Average Bayesian Predictive Distributions, Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman, 2018, Bayesian Analysis 13, Number 3, pp. 917–1003

**Examples**

```
## Not run:
library("data.table")
splitdate <- as.Date("2020-03-28")
data <- setDT(example_data)

traindata <- data[date <= splitdate]
testdata <- data[date > splitdate]

weights <- crps_weights(traindata)

## End(Not run)
```

---

mixture\_from\_samples *Make mixture model from predictive samples*

---

**Description**

The function takes a data.frame with predictive samples generated from different models as well as weights corresponding to these models as input. It then returns predictive samples from a mixture model generated by stacking the original models using these weights.

**Usage**

```
mixture_from_samples(data, weights = NULL, ...)
```

**Arguments**

<code>data</code>	a data.frame with the following entries: <ul style="list-style-type: none"> <li>• <code>observed</code>, the true observed values (optional)</li> <li>• <code>predicted</code>, predicted values corresponding to the true values in <code>observed</code></li> <li>• <code>model</code>, the name of the model used to generate the corresponding predictions</li> <li>• <code>geography</code> (optional), the regions for which predictions are generated. If <code>geography</code> is missing, it will be assumed there are no geographical differences to take into account. Internally, regions will be ordered alphabetically</li> <li>• <code>date</code> (the date of the corresponding prediction / true value). Also works with numbers to indicate timesteps</li> </ul>
<code>weights</code>	stacking weights used to combine the original model to a mixture model. If NULL (default), weights will first be estimated using <code>[crps_weights()]</code> .
<code>...</code>	any additional parameters to pass to <code>[crps_weights()]</code> if <code>'weights'</code> is NULL.

**Value**

data.frame with samples from the mixture model. The following columns are returned:

- `observed`, the true observed values, if they were given as input
- `predicted`, predicted values corresponding to the true values in `observed`
- `model`, the name of the model used to generate the corresponding predictions
- `geography` (optional), the regions for which predictions are generated. If `geography` is missing, it will be assumed there are no geographical differences to take into account. Internally, regions will be ordered alphabetically
- `date` (the date of the corresponding prediction / true value). Also works with numbers to indicate timesteps

**References**

Using Stacking to Average Bayesian Predictive Distributions, Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman, 2018, *Bayesian Analysis* 13, Number 3, pp. 917–1003

**Examples**

```
## Not run:
library("data.table")
data <- setDT(example_data)
weights <- c(0.2, 0.3, 0.4, 0.1)
mix <- mixture_from_samples(data, weights = weights)

## End(Not run)
```

# Index

`crps_weights`, [2](#)

`mixture_from_samples`, [3](#)