

Package ‘loo’

March 27, 2017

Type Package

Title Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models

Version 1.1.0

Date 2016-03-27

Maintainer Jonah Gabry <jsg2201@columbia.edu>

URL <http://mc-stan.org/>,
<https://groups.google.com/forum/#!forum/stan-users>

BugReports <https://github.com/stan-dev/loo/issues>

Description Efficient approximate leave-one-out cross-validation (LOO) using Pareto smoothed importance sampling (PSIS), a new procedure for regularizing importance weights. As a byproduct of the calculations, we also obtain approximate standard errors for estimated predictive errors and for the comparison of predictive errors between models. We also compute the widely applicable information criterion (WAIC).

License GPL (>= 3)

LazyData TRUE

Depends R (>= 3.1.2)

Imports graphics, matrixStats (>= 0.50.0), parallel, stats

Suggests knitr, rmarkdown, rstan, rstanarm, testthat

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Author Aki Vehtari [aut],
Andrew Gelman [aut],
Jonah Gabry [cre, aut],
Juho Piironen [ctb],
Ben Goodrich [ctb]

Repository CRAN

Date/Publication 2017-03-27 15:21:52 UTC

R topics documented:

loo-package	2
compare	5
extract_log_lik	6
E_loo	7
loo	9
pareto-k-diagnostic	11
print.loo	13
psislw	14
waic	16
Index	18

loo-package	<i>Efficient LOO and WAIC for Bayesian models</i>
-------------	---

Description

This package implements the methods described in Vehtari, Gelman, and Gabry (2016a). The package documentation is largely based on excerpts from the paper.

Summary

Leave-one-out cross-validation (LOO) and the widely applicable information criterion (WAIC) are methods for estimating pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values. LOO and WAIC have various advantages over simpler estimates of predictive error such as AIC and DIC but are less used in practice because they involve additional computational steps. This package implements the fast and stable computations for LOO and WAIC laid out in Vehtari, Gelman, and Gabry (2016a). From existing posterior simulation draws, we compute LOO using Pareto Smoothed Importance Sampling (PSIS; Vehtari, Gelman, and Gabry, 2016b), a new procedure for regularizing importance weights. As a byproduct of our calculations, we also obtain approximate standard errors for estimated predictive errors and for comparing of predictive errors between two models.

Details

After fitting a Bayesian model we often want to measure its predictive accuracy, for its own sake or for purposes of model comparison, selection, or averaging. Cross-validation and information criteria are two approaches for estimating out-of-sample predictive accuracy using within-sample fits.

Exact cross-validation requires re-fitting the model with different training sets. Approximate leave-one-out cross-validation (LOO) can be computed easily using importance sampling (Gelfand, Dey, and Chang, 1992, Gelfand, 1996) but the resulting estimate is noisy, as the variance of the importance weights can be large or even infinite (Peruggia, 1997, Epifani et al., 2008). Here we propose a novel approach that provides a more accurate and reliable estimate using importance weights that are smoothed by fitting a generalized Pareto distribution to the upper tail of the distribution of the importance weights.

WAIC (the widely applicable or Watanabe-Akaike information criterion; Watanabe, 2010) can be viewed as an improvement on the deviance information criterion (DIC) for Bayesian models. DIC has gained popularity in recent years in part through its implementation in the graphical modeling package BUGS (Spiegelhalter, Best, et al., 2002; Spiegelhalter, Thomas, et al., 1994, 2003), but it is known to have some problems, arising in part from it not being fully Bayesian in that it is based on a point estimate (van der Linde, 2005, Plummer, 2008). For example, DIC can produce negative estimates of the effective number of parameters in a model and it is not defined for singular models. WAIC is fully Bayesian and closely approximates Bayesian cross-validation. Unlike DIC, WAIC is invariant to parametrization and also works for singular models. WAIC is asymptotically equal to LOO, and can thus be used as an approximation to LOO. In the finite case, WAIC and LOO often give very similar estimates, but for influential observations WAIC underestimates the effect of leaving out one observation.

One advantage of AIC and DIC has been their computational simplicity. In this package we present fast and stable computations for LOO and WAIC that can be performed directly on posterior simulations, thus allowing these newer tools to enter routine statistical practice. As a byproduct of our calculations, we also obtain approximate standard errors for estimated predictive errors and for the comparison of predictive errors between two models.

PSIS-LOO

The distribution of the importance weights used in LOO may have a long right tail. We use the empirical Bayes estimate of Zhang and Stephens (2009) to fit a generalized Pareto distribution (gPd) to the tail (20% largest importance ratios). By examining the shape parameter k of the fitted gPd, we are able to obtain sample based estimates of the existence of the moments (Koopman et al, 2009). This extends the diagnostic approach of Peruggia (1997) and Epifani et al. (2008) to be used routinely with importance-sampling LOO for any model with a factorizing likelihood.

Epifani et al. (2008) show that when estimating the leave-one-out predictive density, the central limit theorem holds if the variance of the weight distribution is finite. These results can be extended using the generalized central limit theorem for stable distributions. Thus, even if the variance of the importance weight distribution is infinite, if the mean exists the estimate's accuracy improves when additional draws are obtained. When the tail of the weight distribution is long, a direct use of importance sampling is sensitive to one or few largest values. By fitting a gPd to the upper tail of the importance weights we smooth these values. The procedure (implemented in the `psislw` function) goes as follows:

1. Fit the gPd to the 20% largest importance ratios r_s . The computation is done separately for each held-out data point i . In simulation experiments with thousands and tens of thousands of draws, we have found that the fit is not sensitive to the specific cutoff value (for a consistent estimation the proportion of the samples above the cutoff should get smaller when the number of draws increases).
2. Stabilize the importance ratios by replacing the M largest ratios by the expected values of the order statistics of the fitted gPd

$$G((z - 0.5)/M), z = 1, \dots, M,$$

where M is the number of simulation draws used to fit the Pareto (in this case, $M = 0.2 * S$) and G is the inverse-CDF of the gPd.

3. To guarantee finite variance of the estimate, truncate the smoothed ratios with

$$S^{3/4}\bar{w},$$

where \bar{w} is the average of the smoothed weights.

The above steps must be performed for each data point i . The result is a vector of weights w_i^s , $s = 1, \dots, S$, for each i , which in general should be better behaved than the raw importance ratios r_i^s from which they were constructed.

The results can be then combined to compute the desired LOO estimates.

Pareto k diagnostic

The reliability of the PSIS-based estimates can be assessed using the estimates for the shape parameter k of the generalized Pareto distribution.

- If $k < 1/2$ the variance of the raw importance ratios is finite, the central limit theorem holds, and the estimate converges quickly.
- If k is between $1/2$ and 1 the variance of the raw importance ratios is infinite but the mean exists, the generalized central limit theorem for stable distributions holds, and the convergence of the estimate is slower. The variance of the PSIS estimate is finite but may be large.
- If $k > 1$ the variance and the mean of the raw ratios distribution do not exist. The variance of the PSIS estimate is finite but may be large.

If the estimated tail shape parameter k exceeds 0.5, the user should be warned, although in practice we have observed good performance for values of k up to 0.7. Even if the PSIS estimate has a finite variance, the user should consider sampling directly from $p(\theta^s|y_{-i})$ for the problematic i , use k -fold cross-validation, or use a more robust model.

Importance sampling is likely to work less well if the marginal posterior $p(\theta^s|y)$ and LOO posterior $p(\theta^s|y_{-i})$ are much different, which is more likely to happen with a non-robust model and highly influential observations. A robust model may reduce the sensitivity to highly influential observations.

References

- Vehtari, A., Gelman, A., and Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Advance online publication. doi:10.1007/s11222-016-9696-4. (published version, [arXiv preprint](#)).
- Vehtari, A., Gelman, A., and Gabry, J. (2016b). Pareto smoothed importance sampling. [arXiv preprint: http://arxiv.org/abs/1507.02646/](http://arxiv.org/abs/1507.02646/)
- Epifani, I., MacEachern, S. N., and Peruggia, M. (2008). Case-deletion importance sampling estimators: Central limit theorems and related results. *Electronic Journal of Statistics* **2**, 774-806.
- Gelfand, A. E. (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson, D. J. Spiegelhalter, 145-162. London: Chapman and Hall.
- Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 147-167. Oxford University Press.

- Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **24**, 997-1016.
- Ionides, E. L. (2008). Truncated importance sampling. *Journal of Computational and Graphical Statistics* **17**, 295-311.
- Koopman, S. J., Shephard, N., and Creal, D. (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics* **149**, 2-11.
- Peruggia, M. (1997). On the variability of case-deletion importance sampling weights in the Bayesian linear model. *Journal of the American Statistical Association* **92**, 199-207.
- Stan Development Team (2016). The Stan C++ Library, Version 2.10.0. <http://mc-stan.org/documentation/>.
- Stan Development Team (2016). RStan: the R interface to Stan, Version 2.10.1 <http://mc-stan.org/interfaces/rstan.html>.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely application information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571-3594.
- Zhang, J., and Stephens, M. A. (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics* **51**, 316-325.

 compare

Model comparison

Description

Compare fitted models on LOO or WAIC

Usage

```
compare(..., x)
```

Arguments

- ... At least two objects returned by `loo` (or `waic`).
- x A list of at least two objects returned by `loo` (or `waic`). This argument can be used as an alternative to specifying the models in ...

Details

When comparing two fitted models, we can estimate the difference in their expected predictive accuracy by the difference in `elpd_loo` or `elpd_waic` (multiplied by -2 , if desired, to be on the deviance scale). To compute the standard error of this difference we can use a paired estimate to take advantage of the fact that the same set of N data points was used to fit both models. These calculations should be most useful when N is large, because then non-normality of the distribution is not such an issue when estimating the uncertainty in these sums. These standard errors, for all their flaws, should give a better sense of uncertainty than what is obtained using the current standard approach of comparing differences of deviances to a Chi-squared distribution, a practice derived for Gaussian linear models or asymptotically, and which only applies to nested models in any case.

Value

A vector or matrix with class 'compare.loo' that has its own print method. If exactly two objects are provided in `...` or `x`, then the difference in expected predictive accuracy and the standard error of the difference are returned (see Details). *The difference will be positive if the expected predictive accuracy for the second model is higher.* If more than two objects are provided then a matrix of summary information is returned.

References

Vehtari, A., Gelman, A., and Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Advance online publication. doi:10.1007/s11222-016-9696-4. ([published version](#), [arXiv preprint](#)).

Vehtari, A., Gelman, A., and Gabry, J. (2016b). Pareto smoothed importance sampling. arXiv preprint: <http://arxiv.org/abs/1507.02646/>

Examples

```
## Not run:
loo1 <- loo(log_lik1)
loo2 <- loo(log_lik2)
print(compare(loo1, loo2), digits = 3)

waic1 <- waic(log_lik1)
waic2 <- waic(log_lik2)
compare(waic1, waic2)

## End(Not run)
```

extract_log_lik

Extract pointwise log-likelihood matrix from a Stan model

Description

Convenience function for extracting the pointwise log-likelihood from a fitted Stan model.

Usage

```
extract_log_lik(stanfit, parameter_name = "log_lik")
```

Arguments

`stanfit` A stanfit object (**rstan** package).

`parameter_name` A character string naming the parameter (or generated quantity) in the Stan model corresponding to the log-likelihood.

Details

Stan does not automatically compute and store the log-likelihood. It is up to the user to incorporate it into the Stan program if it is to be extracted after fitting the model. In a Stan model, the pointwise log likelihood can be coded as a vector in the transformed parameters block (and then summed up in the model block) or it can be coded entirely in the generated quantities block. We recommend using the generated quantities block so that the computations are carried out only once per iteration rather than once per HMC leapfrog step.

For example, the following is the generated quantities block for computing and saving the log-likelihood for a linear regression model with N data points, outcome y , predictor matrix X , coefficients β , and standard deviation σ :

```
vector[N] log_lik;
for (n in 1:N) log_lik[n] = normal_lpdf(y[n] | X[n] * beta, sigma);
```

Value

An S by N matrix of (post-warmup) extracted draws, where S is the size of the posterior sample and N is the number of data points.

References

Stan Development Team (2016). The Stan C++ Library, Version 2.10.0. <http://mc-stan.org/documentation/>

Stan Development Team (2016). RStan: the R interface to Stan, Version 2.10.1. <http://mc-stan.org/interfaces/rstan.html>

E_loo

Compute weighted expectations

Description

The `E_loo` function computes weighted expectations (means, variances, quantiles) using the smoothed importance weights obtained from the [PSIS](#) procedure. The expectations estimated by the `E_loo` function assume that the PSIS approximation is working well. **A small Pareto k estimate is necessary, but not sufficient, for `E_loo` to give reliable estimates.** Additional diagnostic checks for gauging the reliability of the estimates are in development and will be added in a future release.

Usage

```
E_loo(x, lw, ...)

## Default S3 method:
E_loo(x, lw, ..., type = c("mean", "var", "quantile"),
      probs)

## S3 method for class 'matrix'
E_loo(x, lw, ..., type = c("mean", "var", "quantile"), probs)
```

Arguments

x	A numeric matrix or vector.
lw	A numeric matrix (or vector) of smoothed log-weights with the same dimensions (or length) as x. Typically this will be the "lw_smooth" element of the list returned by <code>psislw</code> .
...	For the generic function, arguments passed to the individual methods.
type	The type of expectation to compute. The options are "mean", "var" (variance), and "quantile".
probs	A vector of probabilities. Ignored unless type is "quantile".

Value

The matrix method returns a vector with `ncol(x)` elements, with one exception: when type is "quantile" and multiple values are specified in `probs` the returned object is a `length(probs)` by `ncol(x)` matrix.

The default/vector method returns a scalar, with one exception: when type is "quantile" and multiple values are specified in `probs` the returned object is a vector with `length(probs)` elements.

Examples

```
# Use rstanarm package to quickly fit a model and get both a log-likelihood
# matrix and draws from the posterior predictive distribution
library("rstanarm")

# data from help("lm")
ctl <- c(4.17, 5.58, 5.18, 6.11, 4.50, 4.61, 5.17, 4.53, 5.33, 5.14)
trt <- c(4.81, 4.17, 4.41, 3.59, 5.87, 3.83, 6.03, 4.89, 4.32, 4.69)
d <- data.frame(
  weight = c(ctl, trt),
  group = gl(2, 10, 20, labels = c("Ctl", "Trt"))
)
fit <- stan_glm(weight ~ group, data = d)
yrep <- posterior_predict(fit)
dim(yrep)

ll <- log_lik(fit)
lw <- psislw(-ll, cores = 2)$lw_smooth
dim(lw)

E_loo(yrep, lw, type = "mean")
E_loo(yrep, lw, type = "var")
E_loo(yrep, lw, type = "quantile", probs = 0.5) # median
E_loo(yrep, lw, type = "quantile", probs = c(0.1, 0.9))
```

loo *Leave-one-out cross-validation (LOO)*

Description

Efficient approximate leave-one-out cross-validation for Bayesian models. See [loo-package](#) and Vehtari, Gelman, and Gabry (2016a, 2016b) for background.

Usage

```
loo(x, ...)

## S3 method for class 'matrix'
loo(x, ...)

## S3 method for class 'function'
loo(x, ..., args)
```

Arguments

x	A log-likelihood matrix or function. See the Methods (by class) section below for a detailed description.
...	Optional arguments to pass to psislw . Possible arguments and their defaults are: wcp = 0.2 The proportion of importance weights to use for the generalized Pareto fit. The 100*wcp% largest weights are used as the sample from which to estimate the parameters k and σ of the generalized Pareto distribution. wtrunc = 3/4 For truncating very large weights to S^{wtrunc} (set to zero for no truncation). cores = <code>getOption("loo.cores", parallel::detectCores())</code> The number of cores to use for parallelization. This can be set for an entire R session by <code>options(loo.cores = NUMBER)</code> . The default is <code>detectCores()</code> . We recommend using the default values for the <code>psislw</code> arguments unless there are problems (e.g. NA or NaN results).
args	Only required if x is a function. A list containing the data required to specify the arguments to the function. See the Methods (by class) section below for how args should be specified.

Value

A named list with class 'loo' and components:

elpd_loo, se_elpd_loo Expected log pointwise predictive density and standard error.

p_loo, se_p_loo Estimated effective number of parameters and standard error.

`looic`, `se_looic` The LOO information criterion ($-2 \times \text{elpd_loo}$, i.e., converted to deviance scale) and standard error.

`pointwise` A matrix containing the pointwise contributions of each of the above measures.

`pareto_k` A vector containing the estimates of the shape parameter k for the generalized Pareto fit to the importance ratios for each leave-one-out distribution. See PSIS-LOO section in [loo-package](#) for details about interpreting k . (By default, the `print` method for 'loo' objects will also provide warnings about problematic values of k .)

Methods (by class)

- `matrix`: An S by N matrix, where S is the size of the posterior sample (the number of simulations) and N is the number of data points. For example, the object returned by `extract_log_lik` can be used.
- `function`: A function f that takes arguments `i`, `data`, and `draws` and returns a vector containing the log-likelihood for the i th observation evaluated at each posterior draw. The `args` argument must also be specified and should be a named list with the following components:
 - `draws`: An object containing the posterior draws for any parameters needed to compute the pointwise log-likelihood.
 - `data`: An object containing data (e.g. observed outcome and predictors) needed to compute the pointwise log-likelihood. `data` should be in an appropriate form so that `f(i=i, data=data[i,,drop=FALSE], draws=draws)` returns the S -vector of log-likelihoods for the i th observation.
 - `N`: The number of observations.
 - `S`: The size of the posterior sample.

Note

For models fit to very large datasets we recommend the `loo.function` method, which is much more memory efficient than the `loo.matrix` method. However, the `loo.matrix` method is typically more convenient, so it is usually worth trying `loo.matrix` and then switching to `loo.function` if memory is an issue.

References

Vehtari, A., Gelman, A., and Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Advance online publication. doi:10.1007/s11222-016-9696-4. ([published version](#), [arXiv preprint](#)).

Vehtari, A., Gelman, A., and Gabry, J. (2016b). Pareto smoothed importance sampling. arXiv preprint: <http://arxiv.org/abs/1507.02646/>

See Also

[psislw](#) for the underlying Pareto Smoothed Importance Sampling (PSIS) procedure used for approximating LOO.

[pareto-k-diagnostic](#) for convenience functions for looking at diagnostics.

[compare](#) for model comparison.

`print.loo` for the print method for 'loo' objects.

Examples

```
## Not run:
### Usage with stanfit objects
log_lik1 <- extract_log_lik(stanfit1) # see ?extract_log_lik
loo1 <- loo(log_lik1)
print(loo1, digits = 3)

log_lik2 <- extract_log_lik(stanfit2)
(loo2 <- loo(log_lik2))
compare(loo1, loo2)

## End(Not run)

### Using log-likelihood function instead of matrix
set.seed(024)

# Simulate data and draw from posterior
N <- 50; K <- 10; S <- 100; a0 <- 3; b0 <- 2
p <- rbeta(1, a0, b0)
y <- rbinom(N, size = K, prob = p)
a <- a0 + sum(y); b <- b0 + N * K - sum(y)
draws <- rbeta(S, a, b)
data <- data.frame(y,K)

llfun <- function(i, data, draws) {
  dbinom(data$y, size = data$K, prob = draws, log = TRUE)
}
loo_with_fn <- loo(llfun, args = nlist(data, draws, N, S), cores = 1)

# Check that we get same answer if using log-likelihood matrix
log_lik_mat <- sapply(1:N, function(i) llfun(i, data[i,, drop=FALSE], draws))
loo_with_mat <- loo(log_lik_mat, cores = 1)
all.equal(loo_with_mat, loo_with_fn)
```

Description

Print a diagnostic table summarizing the estimated Pareto shape parameters, find the indexes of observations for which the estimated Pareto shape parameter k is larger than some threshold value, or plot observation indexes vs k estimates.

Usage

```

pareto_k_table(x)

pareto_k_ids(x, threshold = 0.5)

## S3 method for class 'loo'
plot(x, ..., label_points = FALSE)

```

Arguments

x For `pareto_k_ids` and `pareto_k_table`, an object created by `loo` or `psislw`. For `plot`, an object created by `loo`.

threshold The threshold value for k .

label_points, ... For the `plot` method, if `label_points` is `TRUE` the observation numbers corresponding to any values of k greater than 0.5 will be displayed in the plot. Any arguments specified in `...` will be passed to `text` and can be used to control the appearance of the labels.

Value

`pareto_k_table` returns an object of class "pareto_k_table", which is a matrix with columns "Count" and "Proportion" and has its own print method.

`pareto_k_ids` returns an integer vector indicating which observations have Pareto k estimates above threshold.

The `plot` method is called for its side effect and does not return anything. If `x` is the result of a call to `loo`, `plot(x)` produces a plot of the estimates of the Pareto shape parameter k . There is no `plot` method for objects generated by a call to `waic`.

Pareto k diagnostic

The reliability of the PSIS-based estimates can be assessed using the estimates for the shape parameter k of the generalized Pareto distribution.

- If $k < 1/2$ the variance of the raw importance ratios is finite, the central limit theorem holds, and the estimate converges quickly.
- If k is between $1/2$ and 1 the variance of the raw importance ratios is infinite but the mean exists, the generalized central limit theorem for stable distributions holds, and the convergence of the estimate is slower. The variance of the PSIS estimate is finite but may be large.
- If $k > 1$ the variance and the mean of the raw ratios distribution do not exist. The variance of the PSIS estimate is finite but may be large.

If the estimated tail shape parameter k exceeds 0.5, the user should be warned, although in practice we have observed good performance for values of k up to 0.7. Even if the PSIS estimate has a finite variance, the user should consider sampling directly from $p(\theta^s | y_{-i})$ for the problematic i , use k -fold cross-validation, or use a more robust model.

Importance sampling is likely to work less well if the marginal posterior $p(\theta^s|y)$ and LOO posterior $p(\theta^s|y_{-i})$ are much different, which is more likely to happen with a non-robust model and highly influential observations. A robust model may reduce the sensitivity to highly influential observations.

References

Vehtari, A., Gelman, A., and Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Advance online publication. doi:10.1007/s11222-016-9696-4. ([published version](#), [arXiv preprint](#)).

Vehtari, A., Gelman, A., and Gabry, J. (2016b). Pareto smoothed importance sampling. arXiv preprint: <http://arxiv.org/abs/1507.02646/>

See Also

[psislw](#) for the implementation of the PSIS algorithm.

print.loo *Print method for 'loo' objects*

Description

Print method for 'loo' objects

Usage

```
## S3 method for class 'loo'
print(x, digits = 1, plot_k = FALSE, ...)
```

Arguments

x	A list with class 'loo' (as returned by loo or waic).
digits	An integer passed to round .
plot_k	Logical. If TRUE the estimates of the Pareto shape parameter k are plotted. Ignored if x was generated by waic . To just plot k without printing use the plot method .
...	Arguments passed to plot.loo if plot_k is TRUE.

Value

x, invisibly.

See Also

[pareto-k-diagnostic](#)

psislw

*Pareto smoothed importance sampling (PSIS)***Description**

Implementation of Pareto smoothed importance sampling, a method for stabilizing importance weights. For full details about the algorithm see Vehtari, Gelman and Gabry (2016a, 2016b). For diagnostics see the [pareto-k-diagnostic](#) page.

Usage

```
psislw(lw, wcp = 0.2, wtrunc = 3/4, cores = getOption("loo.cores",
  parallel::detectCores()), llfun = NULL, llargs = NULL, ...)
```

Arguments

lw	A matrix or vector of log weights. For computing LOO, $lw = -\log_lik$, the <i>negative</i> of an S (simulations) by N (data points) pointwise log-likelihood matrix.
wcp	The proportion of importance weights to use for the generalized Pareto fit. The $100*wcp\%$ largest weights are used as the sample from which to estimate the parameters of the generalized Pareto distribution.
wtrunc	For truncating very large weights to S^{wtrunc} . Set to zero for no truncation.
cores	The number of cores to use for parallelization. This can be set for an entire R session by <code>options(loo.cores = NUMBER)</code> . The default is detectCores() .
llfun, llargs	See loo.function .
...	Ignored when <code>psislw</code> is called directly. The ... is only used internally when <code>psislw</code> is called by the loo function.

Value

A named list with components `lw_smooth` (modified log weights) and `pareto_k` (estimated generalized Pareto [shape parameter\(s\) k](#)).

PSIS-LOO

The distribution of the importance weights used in LOO may have a long right tail. We use the empirical Bayes estimate of Zhang and Stephens (2009) to fit a generalized Pareto distribution (gPd) to the tail (20% largest importance ratios). By examining the shape parameter k of the fitted gPd, we are able to obtain sample based estimates of the existence of the moments (Koopman et al, 2009). This extends the diagnostic approach of Peruggia (1997) and Epifani et al. (2008) to be used routinely with importance-sampling LOO for any model with a factorizing likelihood.

Epifani et al. (2008) show that when estimating the leave-one-out predictive density, the central limit theorem holds if the variance of the weight distribution is finite. These results can be extended using the generalized central limit theorem for stable distributions. Thus, even if the variance of

the importance weight distribution is infinite, if the mean exists the estimate's accuracy improves when additional draws are obtained. When the tail of the weight distribution is long, a direct use of importance sampling is sensitive to one or few largest values. By fitting a gPd to the upper tail of the importance weights we smooth these values. The procedure (implemented in the `psislw` function) goes as follows:

1. Fit the gPd to the 20% largest importance ratios r_s . The computation is done separately for each held-out data point i . In simulation experiments with thousands and tens of thousands of draws, we have found that the fit is not sensitive to the specific cutoff value (for a consistent estimation the proportion of the samples above the cutoff should get smaller when the number of draws increases).
2. Stabilize the importance ratios by replacing the M largest ratios by the expected values of the order statistics of the fitted gPd

$$G((z - 0.5)/M), z = 1, \dots, M,$$

where M is the number of simulation draws used to fit the Pareto (in this case, $M = 0.2 * S$) and G is the inverse-CDF of the gPd.

3. To guarantee finite variance of the estimate, truncate the smoothed ratios with

$$S^{3/4} \bar{w},$$

where \bar{w} is the average of the smoothed weights.

The above steps must be performed for each data point i . The result is a vector of weights w_i^s , $s = 1, \dots, S$, for each i , which in general should be better behaved than the raw importance ratios r_i^s from which they were constructed.

The results can be then combined to compute the desired LOO estimates.

References

Vehtari, A., Gelman, A., and Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Advance online publication. doi:10.1007/s11222-016-9696-4. (published version, [arXiv preprint](#)).

Vehtari, A., Gelman, A., and Gabry, J. (2016b). Pareto smoothed importance sampling. [arXiv preprint: http://arxiv.org/abs/1507.02646/](http://arxiv.org/abs/1507.02646/)

See Also

[pareto-k-diagnostic](#) for PSIS diagnostics.

waic *Widely applicable information criterion (WAIC)*

Description

Widely applicable information criterion (WAIC)

Usage

```
waic(x, ...)

## S3 method for class 'matrix'
waic(x, ...)

## S3 method for class 'function'
waic(x, ..., args)
```

Arguments

x	A log-likelihood matrix or function. See the Methods (by class) section below for a detailed description.
...	Other arguments. Currently ignored.
args	Only required if x is a function. A list containing the data required to specify the arguments to the function. See the Methods (by class) section below for how args should be specified.

Value

A named list (of class 'loo') with components:

elpd_waic, se_elpd_waic expected log pointwise predictive density and standard error
 p_waic, se_p_waic estimated effective number of parameters and standard error
 waic, se_waic $-2 * \text{elpd_waic}$ (i.e., converted to the deviance scale) and standard error
 pointwise the pointwise contributions of each of the above measures

Methods (by class)

- **matrix**: An S by N matrix, where S is the size of the posterior sample (the number of simulations) and N is the number of data points. For example, the object returned by [extract_log_lik](#) can be used.
- **function**: A function f that takes arguments i , $data$, and draws and returns a vector containing the log-likelihood for the i th observation evaluated at each posterior draw. The `args` argument must also be specified and should be a named list with the following components:
 - **draws**: An object containing the posterior draws for any parameters needed to compute the pointwise log-likelihood.

- data: An object containing data (e.g. observed outcome and predictors) needed to compute the pointwise log-likelihood. data should be in an appropriate form so that $f(i=i, \text{ data=data}[i, , \text{ drop=FALSE}], \text{ draws=draws})$ returns the S-vector of log-likelihoods for the *i*th observation.
- N: The number of observations.
- S: The size of the posterior sample.

See Also

[compare](#), [print.loo](#), [loo-package](#)

Examples

```
## Not run:
log_lik1 <- extract_log_lik(stanfit1)
log_lik2 <- extract_log_lik(stanfit2)
(waic1 <- waic(log_lik1))
(waic2 <- waic(log_lik2))
print(compare(waic1, waic2), digits = 2)

## End(Not run)
```

Index

compare, [5](#), [10](#), [17](#)

detectCores, [9](#), [14](#)

E_loo, [7](#)

extract_log_lik, [6](#), [10](#), [16](#)

loo, [5](#), [9](#), [12–14](#)

loo-package, [2](#), [9](#)

loo.function, [14](#)

Pareto k, [7](#)

pareto-k-diagnostic, [11](#), [14](#)

pareto_k_ids (pareto-k-diagnostic), [11](#)

pareto_k_table (pareto-k-diagnostic), [11](#)

plot.loo, [13](#)

plot.loo (pareto-k-diagnostic), [11](#)

print, [10](#)

print.loo, [11](#), [13](#), [17](#)

PSIS, [7](#)

psislw, [3](#), [8–10](#), [12](#), [13](#), [14](#), [15](#)

round, [13](#)

shape parameter(s) k, [14](#)

text, [12](#)

waic, [5](#), [12](#), [13](#), [16](#)