

# Model selection and comparison: Bayesian models

*Miao Cai*

*2019-09-12*

## 1 Theory

### 1.1 DIC, WAIC, LOOIC, and PSIS-LOO

This is copied from Stan online forum

1. Bayesian LOO estimates the predictive performance given the predictions are made using the posterior predictive distributions (ie integrating over the posterior of parameters) while DIC estimates the predictive performance if the predictions are made using plug-in predictive distributions (ie using a point estimate like the posterior mean of the parameters),
2. Bayesian LOO is applicable also for singular models where the posterior converges to an analytic set instead of a point,
3. DIC works well only if the posterior is close to Gaussian (not easy to say when it's close enough) and thus it's also not invariant to parameterization,
4. Bayesian LOO computed with PSIS-LOO is almost as fast to compute as DIC, but has good diagnostic to tell when the computation fails, while DIC doesn't have good diagnostic.

## 2 Practice

### 2.1 `loo::compare()`

`loo::compare()` says (cite from [here](#)):

The difference will be positive if the expected predictive accuracy for the second model is higher.

and

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.

First, instead of SE, it's better to consider something like 2SE or more cautious 4SE, where 4 comes from the fact that SE for LOO can be underestimated for small n or under bad model misspecification. Second, the models can be very different and the predictions can be very different, it's just that the average predictive accuracies are close to each other. Third, SE describe uncertainty, so if SE is large then it's likely that the models do have big difference in predictive accuracy, but we don't know whether the difference is negative or positive.

### 2.2 `p_loo` and Pareto k values

This a quick note on different cases why Pareto k values can be large. I'll try to extend this to a longer explanation and a case study from [Stan forum by Aki Vehtari](#).

- If all Pareto  $k$  small, model is likely to be ok (although there can be better models)
- If high Pareto  $k$  values

- If  $p_{loo} \ll$  the number of parameters  $p$ , then the model is likely to be misspecified. PPC is likely to detect the problem, too. Try using overdispersed model, or add more structural information (nonlinearity, mixture model, etc.).
- If  $p_{loo} >$  the number of parameters  $p$ , then the model is likely to be badly misspecified. If the number of parameters  $p \ll n$ , then PPC is likely to detect the problem, too. Case example [https://rawgit.com/avehtari/modelselection\\_tutorial/master/roaches.html](https://rawgit.com/avehtari/modelselection_tutorial/master/roaches.html) 41
- If  $p_{loo} >$  the number of parameters  $p$ , then the model is likely to be badly misspecified. If the number of parameters  $p$  is relatively large compared to the number of observations  $p > n/5$  (more accurately we should count number of observations influencing each parameter as in hierarchical models some groups may have small  $n$  and some groups large  $n$ ), it is possible that PPC doesn't detect the problem. Case example [Recommendations for what to do when k exceeds 0.5 in the loo package?](#) 58
- If  $p_{loo} <$  the number of parameters  $p$  and the number of parameters  $p$  is relatively large compared to the number of observations  $p > n/5$ , it is likely that model is so flexible or population prior is so weak that it's difficult to predict for left out observation even if the model is true one. Case example is the simulated 8 schools in <https://arxiv.org/abs/1507.04544> 24 and Gaussian processes and spatial models with short correlation lengths.

### 3 Useful papers and case studies

- [Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC](#),
- [R-squared for Bayesian regression models](#) on JASA, and an associated case study [Bayesian R2 and LOO-R2](#),
- [Bayesian data analysis - roaches cross-validation demo](#), which includes a case study using a **negative binomial model**
- [Model assesment, selection and inference after selection](#)