

Analytical approximations for leave one out utility standard errors

Leevi Lindgren *

October 5, 2022

*I thank Aki Vehtari and Nikolas Siccha for their comments.

Note

The approach in this report was motivated by the work by Hastings (1970) who proposes an approximation for the variance of the ratio of two random variables. However, the formula provided by Hastings (1970) contains two typos: on page 8, in the formula for the variance of the ratio, \bar{Z} is missing the "bar" symbol and computed variance $s_{\bar{Z}}$ is missing the second power. The missing second power caused the author of this report to mistake it for the standard deviation instead of variance, which was followed by weird simulation results. Unfortunately, research papers are not like software: you cannot make a pull request to fix a bug.

Contents

1	Introduction	4
2	Predictive methods for Bayesian model assessment	4
2.1	Estimating the predictive performance	4
2.2	Model comparison	5
2.3	Assessing the uncertainty of the estimators	6
3	Taylor series approximations	7
3.1	Taylor approximation for the expected value	7
3.2	Taylor approximation for variance	8
4	Approximations for standard errors of LOO utilities	8
4.1	Mean squared error	8
4.2	Root mean squared error	9
4.3	R-squared	11
5	Simulation experiments	12
5.1	RMSE	12
5.2	LOO-R2	12

1 Introduction

This report presents analytical approximations for standard errors of different predictive metrics in the context of Bayesian predictive model assessment.

Usefulness of a statistical model is often evaluated by assessing the predictive performance of the model. Significant discrepancies between the predicted values and observations suggest that the model is not useful, even though the ultimate goal is to make inference in some parameter, such as treatment effect. There is a major challenge when evaluating the predictive power: future observations are not available.

Predictive performance of a statistical model is hence estimated as the expected predictive performance (or utility).

Introduce LOO

Discuss comparison using LOO utilities

Discuss the uncertainty evaluation of the estimation. Why this analytical approximation approach would be useful (compared to e.g. BB)?

The report is structured as follows: The second section follows closely the ideas presented in Vehtari and Ojanen (2012) and introduces briefly some theory on model assessment and comparison based on predictive performance. Third section presents Taylor approximation for functions of two variables, and shows how expectations and variances can be approximated when the variables are considered to be random. Fourth section discusses X different metrics and the analytical formulas for the variance of the estimator. Moreover, approximation of variance of the difference between two estimators is also represented. In section five, we verify the validity of the approximations with simulation studies. Finally, the last section provides concluding remarks.

2 Predictive methods for Bayesian model assessment

This section reviews three concepts central to the later discussion in the paper: 1) estimating the predictive performance of a statistical model 2) comparing statistical models and 3) quantifying the uncertainty in the performance estimates. We follow closely the ideas from Vehtari and Ojanen (2012). Notation through out this report, and specifically this section, is summarized in Table 1.

2.1 Estimating the predictive performance

As noted in the first section, the usefulness of a statistical model is often evaluated on the basis of predictive performance. Performance is measured by some utility function u . Formally, the expected predictive performance for utility u can be written as

$$u(M_k|y) = \int u(M_k, \hat{a}_k, \tilde{y}|y) p_t(\tilde{y}) d\tilde{y}, \quad (1)$$

where \hat{a}_k denotes the optimal decision under M_k . For example, if the defined utility is mean squared error (MSE), the optimal decision, or prediction, is using the posterior predictive expectation of \tilde{y} , i.e., $E[\tilde{y}|y, M_k]$. Then Equation (1) for MSE can be written as

$$u_{\text{MSE}}(M_k|y) = \int (\tilde{y} - E[\tilde{y}|y, M_k])^2 p_t(\tilde{y}) d\tilde{y}$$

Computing (1) poses a challenge as we rarely know the true mechanism which generates the data, $p_t(y)$. In optimal case we would have independent test data $\tilde{y}_i, i = 1, \dots, \tilde{n}$ and (1) would

be easy to compute using

$$\hat{u}_{\text{test}}(M_k|y) = \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} u(M_k, \hat{a}_k, \tilde{y}_i|y). \quad (2)$$

However, typically we don't have access to such data.

To tackle this, naive approach would be to re-use the observed data which was used to train the model, and compute "training" version of (1):

$$\hat{u}_{\text{train}}(M_k|y) = \frac{1}{n} \sum_{i=1}^n u(M_k, \hat{a}_k, y_i|y). \quad (3)$$

Training utility in (3) is a biased estimate of (1), and typically is over optimistic about the predictive performance of the model.

One way to improve the quality of the training utility estimate is to introduce a sample re-use strategy in which data is split into folds I_1, \dots, I_K and each dataset y_{I_k} is used in turn as a validation set, while model is trained with y_{-I_k} . This approach is called cross-validation (CV). Special case of CV is leave-one-out corr-validation (LOO-CV), where data of size n is split into n folds. LOO-CV utility estimate is then given as

$$\hat{u}_{\text{LOO}}(M_k|y) = \frac{1}{n} \sum_{i=1}^n u(M_k, \hat{a}_k, y_i|y_{-i}). \quad (4)$$

2.2 Model comparison

Typically in Bayesian workflow, the modeller has multiple candidate models. If model averaging is not possible or the set of candidate models need to be thinned, e.g. due to the cost of data gathering, she would like to choose one model from the set of candidates. The goal is to choose model having the best expected utility.

Ideas from the previous section can be extended into the context of model comparison. Consider two models, M_A and M_B . With a slight use of notation, we write the difference on the expected utility between models A and B as

$$\begin{aligned} u(M_A, M_B|y) &= u(M_A|y) - u(M_B|y) \\ &= \int u(M_A, \hat{a}_k, \tilde{y}|y) p_t(\tilde{y}) d\tilde{y} - \int u(M_B, \hat{a}_k, \tilde{y}|y) p_t(\tilde{y}) d\tilde{y} \\ &= \int (u(M_A, \hat{a}_k, \tilde{y}|y) - u(M_B, \hat{a}_k, \tilde{y}|y)) p_t(\tilde{y}) d\tilde{y} \end{aligned} \quad (5)$$

Following the ideas from the previous section, we can compute the estimators for the difference in utilities given e.g. independent test data, training data or LOO re-use of data. As we will focus on LOO estimators in the following sections, we only state LOO estimator for the difference. We use the definitions by Vehtari et al. (2016) and Sivula et al. (2022):

$$\begin{aligned} \hat{u}_{\text{LOO}}(M_A, M_B|y) &= \frac{1}{n} \sum_{i=1}^n u(M_A, \hat{a}_k, y_i|y_{-i}) - \frac{1}{n} \sum_{i=1}^n u(M_B, \hat{a}_k, y_i|y_{-i}) \\ &:= \frac{1}{n} \sum_{i=1}^n u(M_A, M_B, y_i|y_{-i}). \end{aligned} \quad (6)$$

2.3 Assessing the uncertainty of the estimators

If we knew $p_t(y)$, we could compute the quantity (1) with arbitrary accuracy. Alternatively, if we had an independent test dataset, then the uncertainty could be quantified with the Monte Carlo standard error of (2). Moreover, if the variance of the individual components $\hat{u}_{\text{test}}(M_k, \hat{a}_k, \tilde{y}_i|y)$ is finite, the error distribution will converge towards normal distribution.

Following Vehtari et al. (2016) and Sivula et al. (2022), we can compute Monte Carlo error of (4) and (6) in a similar way, by considering the individual terms in the corresponding sums. These Monte Carlo error estimates describe the *epistemic* uncertainty about the true values (1) and (5), arising due to the fact that we are working with finite data. (NOTE: check if this is correct)

Variance of the estimator is

$$\text{var}(\hat{u}_{\text{LOO}}(M_k|y)) = \frac{1}{n-1} \sum_{i=1}^n \left(\hat{u}_{\text{LOO}}(M_k, y_i|y_{-i}) - \frac{1}{n} \sum_{i=1}^n \hat{u}_{\text{LOO}}(M_k, y_i|y_{-i}) \right)^2, \quad (7)$$

and the variance of the difference (6):

$$\text{var}(\hat{u}_{\text{LOO}}(M_A, M_B|y)) = \frac{1}{n-1} \sum_{i=1}^n \left(\hat{u}_{\text{LOO}}(M_A, M_B, y_i|y_{-i}) - \frac{1}{n} \sum_{i=1}^n \hat{u}_{\text{LOO}}(M_A, M_B, y_i|y_{-i}) \right)^2. \quad (8)$$

When estimating LOO utilities, Sivula et al. (2022) note that these naive variance estimators are typically biased since the individual LOO terms are not independent in general.

For particular utilities, such root mean squared error (RMSE) and (Bayesian) R-squared, we "lose" the information of individual terms. Take the RMSE as an example. It is defined as the square root of the mean squared error, so we don't have the individual utility terms $u(M_k, \hat{a}_k, y_i|\cdot)$ to work with. In these cases, the uncertainty could be quantified by methods such as Bayesian bootstrap (BB) (Rubin (1981), Vehtari and Lampinen (2002), Vehtari and Ojanen (2012)). We propose an analytical approximation making use of Taylor expansions of functions of random variables.

In the upcoming sections, we demonstrate how Taylor expansions can be used to obtain analytical approximations of the variance estimators in cases when it is not possible to compute (7) or (8) directly.

notation	meaning
$\hat{\cdot}$	A hat symbol is used when the underlying quantity is an estimator.
$p_t(\tilde{y})$	The "true" data generating distribution, which is typically unknown.
$y_i y_{-i}$	y_i given all other observations than y_i . E.g. $p(y_i y_{-i})$ denotes the leave one out predictive distribution of y_i .
$u(M_k y)$	Utility for model M_k with respect to the true data generating process given the observed data y_i .
$u(M_k, y_i y)$	Utility for a single data point, if applicable.
$u(M_A, M_B y)$	Difference in utilities, that is, $u(M_A y) - u(M_B y)$.
$u(M_A, M_B, y_i y)$	Point-wise difference in utilities, if applicable.
$\hat{e}_i^{(k)}$	Squared residual of data point i when predicted with the conditional mean by model k .
$\hat{e}_{i,loo}^{(k)}$	Squared leave-one-out residual of data point i when predicted with $E[y_i y_{-1}, M_k]$

Table 1: Notation used throughout the paper.

3 Taylor series approximations

Taylor series of a function is a infinite sum in which each term is a partial derivative of increasing order. Function value at a particular point and the value of the Taylor series sum coincide at that point. By truncating the Taylor sum, one obtains approximations of a function around some point. Perhaps the most common approximations are the first order (linear) and second order approximations.

First order Taylor approximation of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ of a single variable around point x_0 is

$$f(x) \approx f(x_0) + \frac{\partial f}{\partial x}(x_0)(x - x_0).$$

Bivariate case is similar. Let $f(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a function of two (random) variables. The first order Taylor approximation of a function of two variables around point x_0, y_0 is given by

$$f(x, y) \approx f(x_0, y_0) + \frac{\partial}{\partial x}f(x_0, y_0)(x - x_0) + \frac{\partial}{\partial y}f(x_0, y_0)(y - y_0)$$

Now, let's take a random variable X and approximate $f : \mathbb{R} \rightarrow \mathbb{R}$ around the expected value of X , μ_X :

$$f(X) \approx f(\mu_X) + \frac{\partial f}{\partial x}(X)(X - \mu_X). \quad (9)$$

With two random variables X and Y and the approximation around the expected values $(\mu_X, \mu_Y) = (E[X], E[Y])$, we get

$$f(X, Y) \approx f(\mu_X, \mu_Y) + \frac{\partial}{\partial x}f(\mu_X, \mu_Y)(X - \mu_X) + \frac{\partial}{\partial y}f(\mu_X, \mu_Y)(Y - \mu_Y). \quad (10)$$

3.1 Taylor approximation for the expected value

Given that terms containing partial derivatives in (10) go to zero under expectation, we get the following, simple, first order approximation for the expected value of $f(X, Y)$:

$$E[f(X, Y)] \approx f(\mu_X, \mu_Y) \quad (11)$$

3.2 Taylor approximation for variance

For variance

$$\begin{aligned}\text{var}(f(X)) &= E \left[(f(X) - E[f(X)])^2 \right] \\ &\approx E \left[(f(X) - f(\mu_X))^2 \right]\end{aligned}$$

and

$$\begin{aligned}\text{var}(f(X, Y)) &= E \left[(f(X, Y) - E[f(X, Y)])^2 \right] \\ &\approx E \left[(f(X, Y) - f(\mu_X, \mu_Y))^2 \right].\end{aligned}$$

Using (9), Taylor approximation for the variance of a function of single random variable is then (we write $\frac{\partial}{\partial x} f(\mu_X) = \frac{\partial f}{\partial x}$ for notational simplicity):

$$\begin{aligned}\text{var}(f(X)) &\approx E \left[(f(X) - f(\mu_X))^2 \right] \\ &\approx E \left[\left(\frac{\partial f}{\partial x} \right)^2 (X - \mu_X)^2 \right] \\ &= \left(\frac{\partial f}{\partial x} \right)^2 \text{var}(X)\end{aligned}\tag{12}$$

Next, we use (10) for $f(X, Y)$ which yields

$$\begin{aligned}\text{var}(f(X, Y)) &\approx E \left[\left(\frac{\partial f}{\partial x} (X - \mu_X) + \frac{\partial f}{\partial y} (Y - \mu_Y) \right)^2 \right] \\ &= E \left[\left(\frac{\partial f}{\partial x} \right)^2 (X - \mu_X)^2 + 2 \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} (X - \mu_X)(Y - \mu_Y) + \left(\frac{\partial f}{\partial y} \right)^2 (Y - \mu_Y)^2 \right] \\ &= \left(\frac{\partial f}{\partial x} \right)^2 \text{var}(X) + 2 \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \text{cov}(X, Y) + \left(\frac{\partial f}{\partial y} \right)^2 \text{var}(Y).\end{aligned}\tag{13}$$

4 Approximations for standard errors of LOO utilities

In what follows, we consider a general setting, where we have n observations (y_i, x_i) in which $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$. We are estimating some utility $u(M_k|y)$ with $\hat{u}(M_k|y)$ computed from the training data, using LOO or some approximation of LOO, e.g. Pareto smoothed importance sampling LOO estimator (Vehtari et al. (2016)).

4.1 Mean squared error

Mean squared error (MSE) is a well known metric from the point estimation literature. It also turns out, that we can express other utilities (root mean squared error and R-squared) and their standard error estimators in terms of mean squared error. This is also case for the variance of the difference. Hence we need to have estimator for the MSE and variance for that estimator.

$u(M y)$	$\widehat{Var}(u)$	$\widehat{Var}(u(M_A, M_B))$	Comment
MSE	(17)		Needed for the Taylor approximation
RMSE	(18)	(19)	RMSE is expressed in terms of MSE
R^2	(28)	??	R^2 is expressed in terms of ratio of two mean squared errors and Taylor approximation is used.

Table 2: Taylor approximations for the standard error estimators of different utilities. The middle column refers to the equation of standard error estimator. The equations for standard error estimator for the difference between two utilities are reported in the right-most column.

For notational simplicity, we define the squared residual $\hat{e}_i^{(k)} := (y_i - E[y_i|y, M_k])^2$ and LOO-residual $\hat{e}_{i,loo}^{(k)} := (y_i - E[y_i|y_{-i}, M_k])^2$. MSE is simply defined as

$$u_{\text{MSE}}(M_k|y) = \int (\tilde{y} - E[\tilde{y}|y, M_k])^2 p_t(\tilde{y}) d\tilde{y}. \quad (14)$$

We can estimate this quantity by in-sample estimator

$$\begin{aligned} \hat{u}_{\text{MSE}}(M_k|y) &= \frac{1}{n} \sum_{i=1}^n \hat{u}_{\text{MSE}}(M_k, y_i|y) \\ &= \frac{1}{n} \sum_{i=1}^n \hat{e}_i^{(k)} \end{aligned} \quad (15)$$

or by leave-one-out estimator

$$\begin{aligned} \hat{u}_{\text{LOO-MSE}}(M_k|y) &= \frac{1}{n} \sum_{i=1}^n \hat{u}_{\text{LOO-MSE}}(M_k, y_i|y_{-i}) \\ &= \frac{1}{n} \sum_{i=1}^n \hat{e}_{i,loo}^{(k)}. \end{aligned} \quad (16)$$

Using (7), we write the variance of the estimator (16) as

$$\widehat{\text{var}}(\hat{u}_{\text{LOO-MSE}}(M_k|y)) = \frac{1}{n-1} \sum_{i=1}^n \left(\hat{e}_{i,loo}^{(k)} - \frac{1}{n} \sum_{i=1}^n \hat{e}_{i,loo}^{(k)} \right)^2. \quad (17)$$

In-sample estimator would be computed analogically.

4.2 Root mean squared error

Root mean squared error is simply defined as the square root of MSE and the estimator is, not surprisingly, the square root of MSE estimator in (16):

$$\hat{u}_{\text{LOO-RMSE}}(M_k|y) = \sqrt{\hat{u}_{\text{LOO-MSE}}(M_k|y)}.$$

As RMSE can be interpreted as a function of MSE, we can use the Taylor method to approximate the variance of the estimator. Following the previous section, the variance of function $f(x)$ can

be expressed in terms of variance of X . First, we need the derivative of the function $f(x) = \sqrt{x}$, which is obviously

$$\frac{\partial f}{\partial x} = \frac{1}{2\sqrt{x}}.$$

Variance of the such function is then approximated by (12):

$$\text{var}(f(X)) \approx \frac{\text{var}(X)}{4\sqrt{\mu_X}}.$$

Since RMSE is a function of MSE, we can plug in estimator and variance of that estimate into μ_X and $\text{var}(X)$ to obtain an approximation of the variance of the RMSE estimator:

$$\widehat{\text{var}}(\hat{u}_{\text{LOO-RMSE}}(M_k|y)) := \frac{\widehat{\text{var}}(\hat{u}_{\text{LOO-MSE}}(M_k|y))}{4\hat{u}_{\text{LOO-MSE}}(M_k|y)} \quad (18)$$

If we want to compare two models based on RMSE we want to quantify the uncertainty in the comparison. Variance of the RMSE difference between two models can be also approximated using the same trick. Now we have a function of two random variables

$$f(x, y) = \sqrt{x} - \sqrt{y}.$$

The partials are

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{1}{2\sqrt{x}} \\ \frac{\partial f}{\partial y} &= -\frac{1}{2\sqrt{y}}, \end{aligned}$$

and plugging these into (13) yields¹

$$\begin{aligned} \text{var}(f(X, Y)) &\approx \frac{\text{var}(X)}{4\sqrt{\mu_X}} - \frac{\text{cov}(X, Y)}{2\sqrt{\mu_X}\sqrt{\mu_Y}} + \frac{\text{var}(Y)}{4\sqrt{\mu_Y}}. \\ \widehat{\text{var}}(\hat{u}_{\text{LOO-RMSE}}(M_A, M_B|y)) &:= \frac{\hat{\sigma}_{\text{MSE}_A}^2}{4\sqrt{\hat{\mu}_{\text{MSE}_A}}} - \frac{\text{cov}(\hat{\mu}_{\text{MSE}_A}, \hat{\mu}_{\text{MSE}_B})}{2\sqrt{\hat{\mu}_{\text{MSE}_A}}\sqrt{\hat{\mu}_{\text{MSE}_B}}} + \frac{\hat{\sigma}_{\text{MSE}_B}^2}{4\sqrt{\hat{\mu}_{\text{MSE}_B}}} \\ \hat{\mu}_{\text{MSE}_A} &= \hat{u}_{\text{LOO-MSE}}(M_A|y) \\ \hat{\mu}_{\text{MSE}_B} &= \hat{u}_{\text{LOO-MSE}}(M_B|y) \\ \hat{\sigma}_{\text{MSE}_A}^2 &= \widehat{\text{var}}(\hat{u}_{\text{LOO-MSE}}(M_A|y)) \\ \hat{\sigma}_{\text{MSE}_B}^2 &= \widehat{\text{var}}(\hat{u}_{\text{LOO-MSE}}(M_B|y)) \\ \text{cov}(\hat{\mu}_{\text{MSE}_A}, \hat{\mu}_{\text{MSE}_B}) &= \frac{1}{n-1} \sum_{i=1}^n \left(\left(\hat{e}_{i,loo}^{(A)} - \frac{1}{n} \sum_{i=1}^n \hat{e}_{i,loo}^{(A)} \right) \left(\hat{e}_{i,loo}^{(B)} - \frac{1}{n} \sum_{i=1}^n \hat{e}_{i,loo}^{(B)} \right) \right) \end{aligned} \quad (19)$$

¹Recall that (13) is the first order Taylor approximation of the variance around the expected values of X and Y

4.3 R-squared

R-squared, or R2, is a typical measure used to assess how well a model fits data. Intuitively speaking is the proportion of the variance in the data a model can explain.

SOME THOUGHTS: In the bayesian version of R2 we propagate the uncertainty from our parameter estimates into the predictions and from there into the Bayesian-R2 estimator. In the PSIS-LOO version of the R2, we "loose" the parameter uncertainty to the importance sampling, and quantify the uncertainty by generating (Bayesian) bootstrap samples of the residuals (y - predictions) and data y . So these uncertainty estimates seem to measure different uncertainty.

As with RMSE, we can interpret R2 as a function of two random variables and utilize the Taylor trick to approximate the variance of the R2 estimator.

LOO-R2 is defined as

$$R2_{loo} = 1 - \frac{\text{var}(\hat{e}_{loo})}{\text{var}(y)} \quad (20)$$

where $\hat{e}_{loo} = y - \hat{y}_{loo}$. Note that the nominator and denominator of the second term in (20) can be interpreted as the mean squared error of the LOO predictions and mean squared error of predicting the data with its mean, respectively. So we write (20) as

$$R2_{loo} = 1 - \frac{\text{MSE}_{\hat{e}}}{\text{MSE}_y} \quad (21)$$

We can then compute the estimator for the variance of both using the same approach as in Sivula et al. (2022) and Vehtari et al. (2016):

$$\text{var}(\text{MSE}_{\hat{e}}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{e}_{loo,i}^2 - \text{MSE}_{\hat{e}})^2 \quad (22)$$

and

$$\text{var}(\text{MSE}_y) = \frac{1}{n(n-1)} \sum_{i=1}^n ((y_i - \bar{y})^2 - \text{MSE}_y)^2, \quad (23)$$

where \bar{y} is the sample mean of observations y .

To utilize the Taylor approximation, we need to compute partial derivatives of function $f(x, y) = 1 - \frac{x}{y}$. With a simple calculus, we get

$$\frac{\partial}{\partial x} f(x, y) = -\frac{1}{y} \quad (24)$$

$$\frac{\partial}{\partial y} f(x, y) = \frac{x}{y^2}. \quad (25)$$

Substituting these into (13) yields

$$\text{var}(f(X, Y)) \approx \frac{1}{\mu_Y^2} \left(\text{var}(X) - 2 \frac{\mu_X}{\mu_Y} \text{cov}(X, Y) + \left(\frac{\mu_X}{\mu_Y} \right)^2 \text{var}(Y) \right). \quad (26)$$

To use this expression for the mean squared errors, we also need the covariance between $\text{MSE}_{\hat{e}}$ and MSE_y . This can be estimated in a similar way as the variances:

$$\text{cov}(\text{MSE}_{\hat{e}}, \text{MSE}_y) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{e}_{loo,i}^2 - \text{MSE}_{\hat{e}}) ((y_i - \hat{y})^2 - \text{MSE}_y). \quad (27)$$

Putting all pieces together, variance, and consequently the standard error, of LOO-R2 estimator can then be approximated by letting $\mu_X = \text{MSE}_{\hat{e}}$ and $\mu_Y = \text{MSE}_y$ and then substituting with (22), (23) and (27) into (26):

$$\text{var}(\text{R2}_{loo}) = \frac{1}{\text{MSE}_y^2} \left(\text{var}(\text{MSE}_{\hat{e}}) - 2 \frac{\text{MSE}_{\hat{e}}}{\text{MSE}_y} \text{cov}(\text{MSE}_{\hat{e}}, \text{MSE}_y) + \left(\frac{\text{MSE}_{\hat{e}}}{\text{MSE}_y} \right)^2 \text{var}(\text{MSE}_y) \right). \quad (28)$$

5 Simulation experiments

We use simple simulation experiments to illustrate the behaviour of the proposed estimators for the variance. These simulation are by no means exhaustive, and more elaborate experiments are needed to verify the validity of the Taylor approximation method. We study the properties of the Taylor approximation of the variance of the estimator, and Taylor approximation of the variance of the difference when the utility is either RMSE or R-squared.

We consider simple data generating mechanism in which we vary the sample size $n \in \{20, 50, 100, 200\}$ and number of predictors $p \in \{2, 3, 5\}$.

$$\begin{aligned} y_i &\sim N(\beta^T x_i, \sigma^2) \\ x_i &\sim N(0, 1) \\ \beta &\sim N(0, 1) \end{aligned}$$

We then with two models M_A and M_B with one and two predictors respectively. Hence for the cases in which $p \neq 2$ both of the models are miss-specified. Then we compute the variance of the estimator for (1) using Bayesian bootstrap and the proposed Taylor approximation. We only report the variance estimator for the larger model M_B . We also compute the variance estimator of the difference in utilities (5). For each pair of (n, p) , we repeat the procedure for 100 times.

5.1 RMSE

5.2 LOO-R2

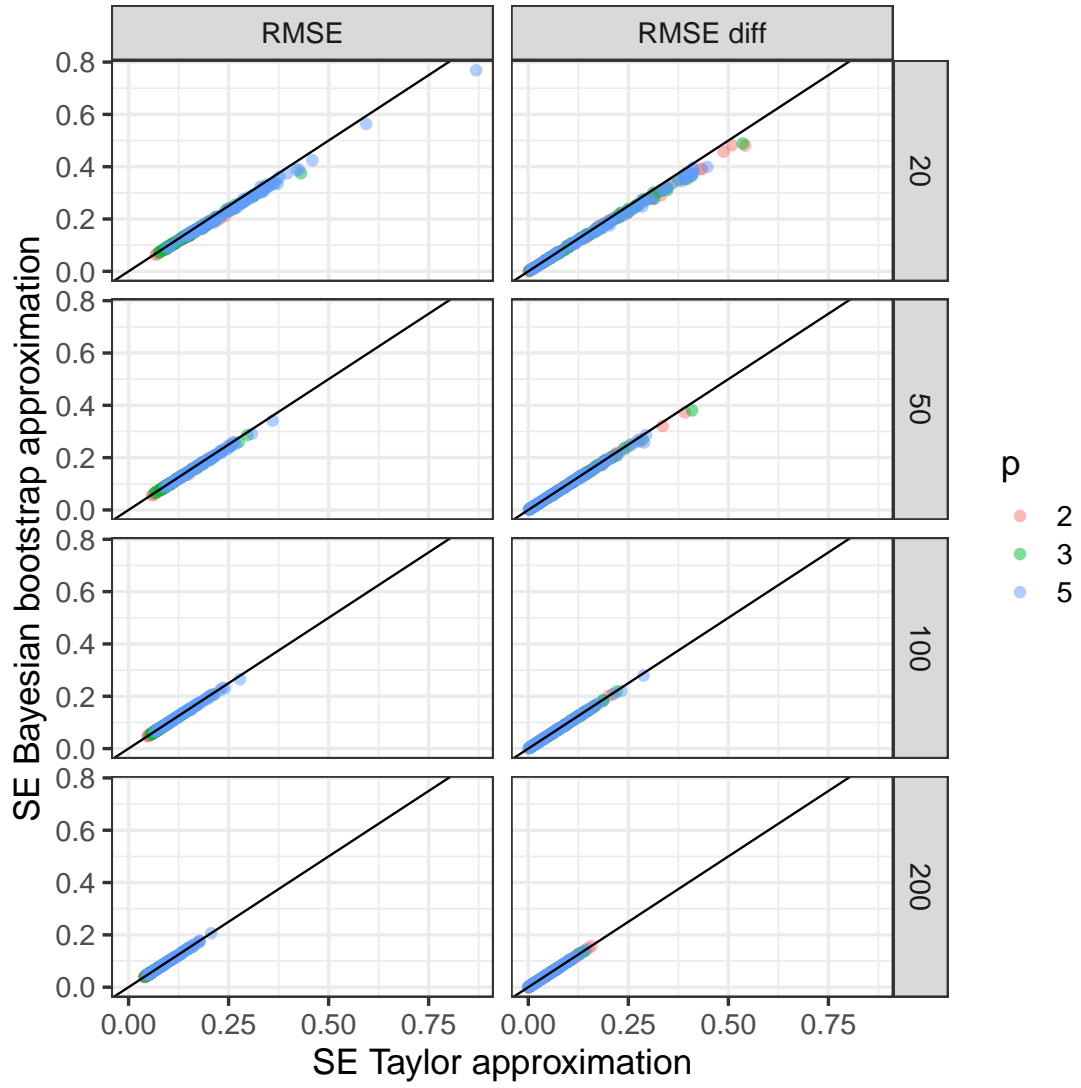


Figure 1: Comparison of Taylor (x-axis) and Bayesian bootstrap (y-axis) approximations. Left panel shows the variance estimator for the RMSE estimator and right panel the variance estimator of the RMSE difference.

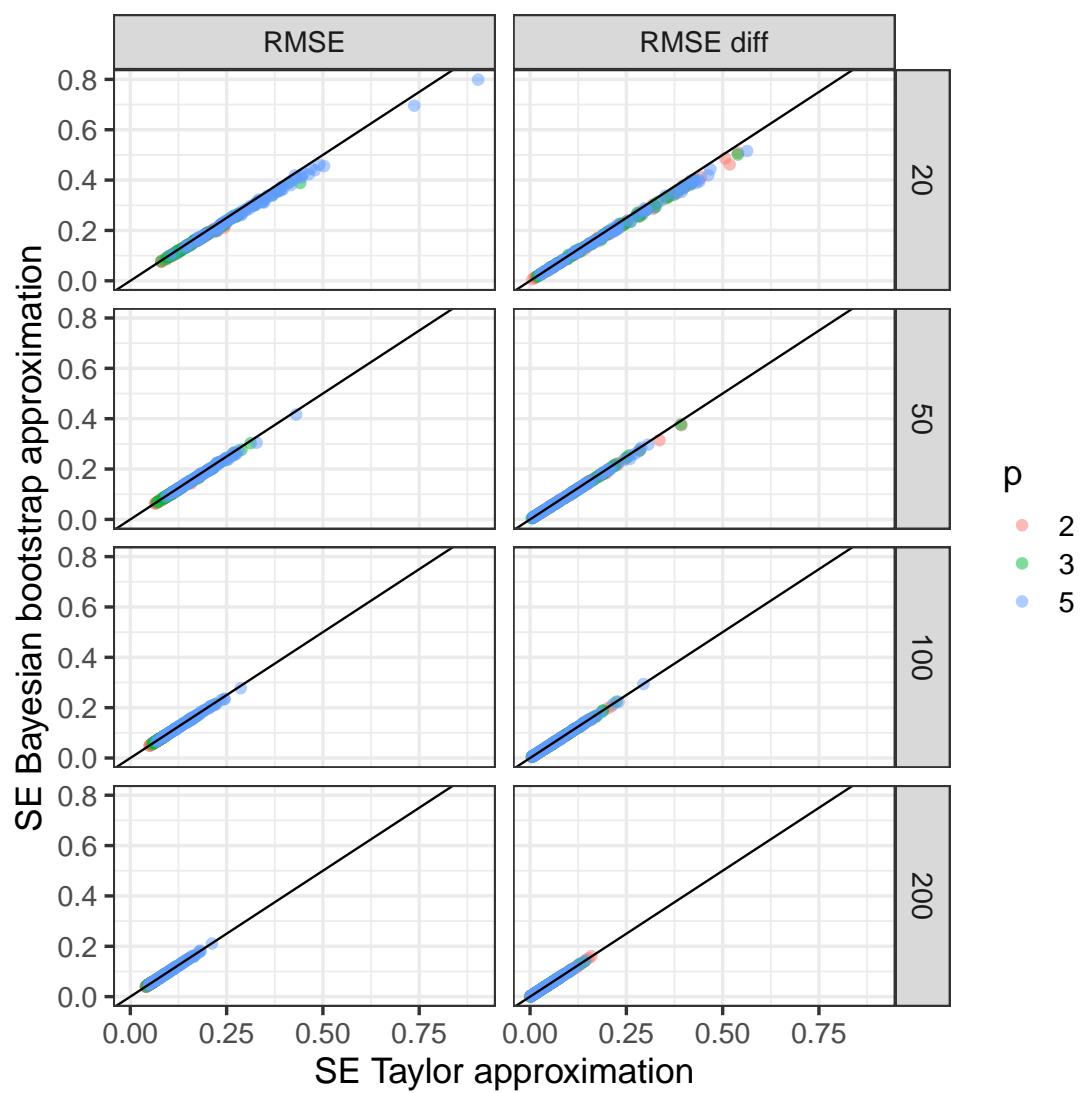


Figure 2: LOO RMSE

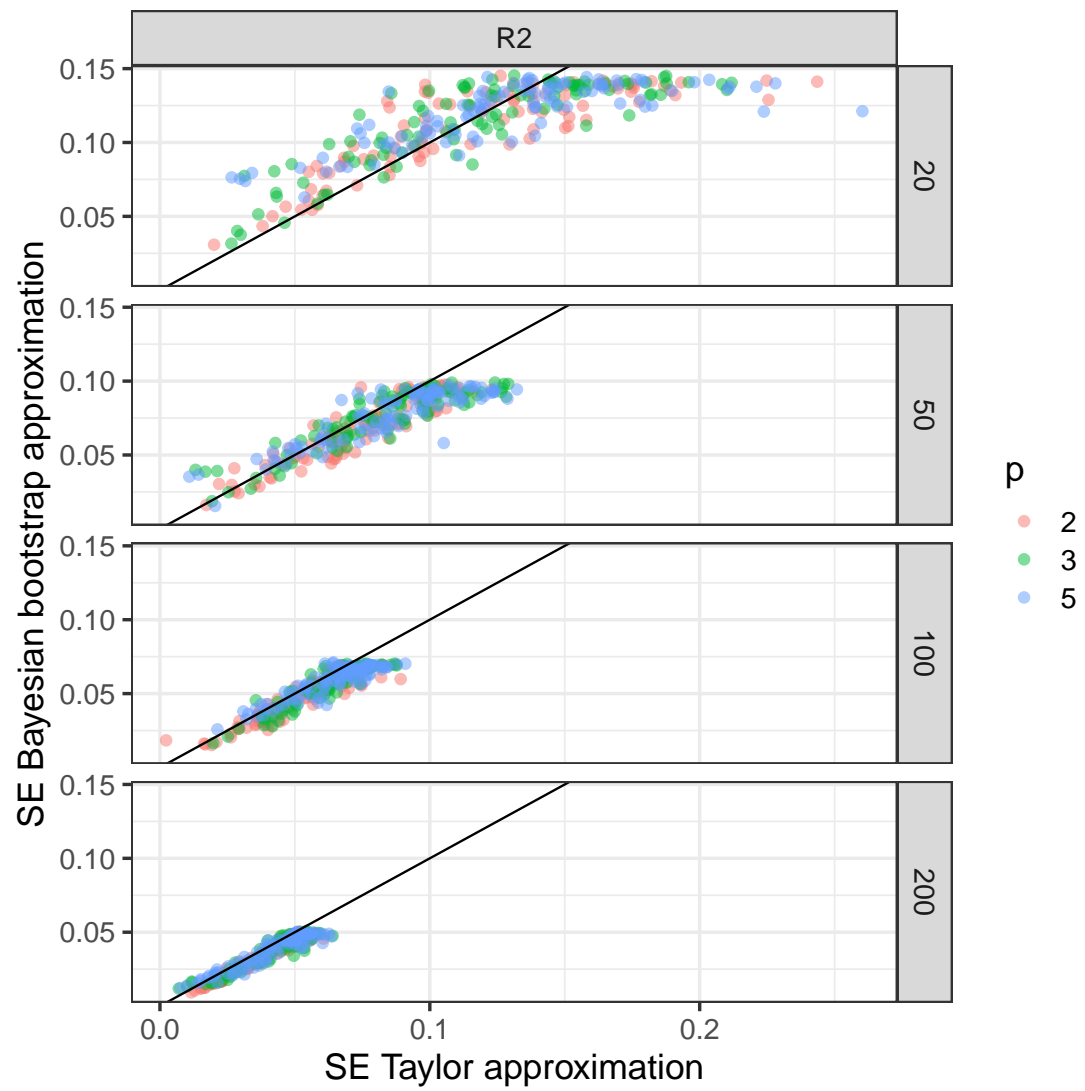


Figure 3: R^2

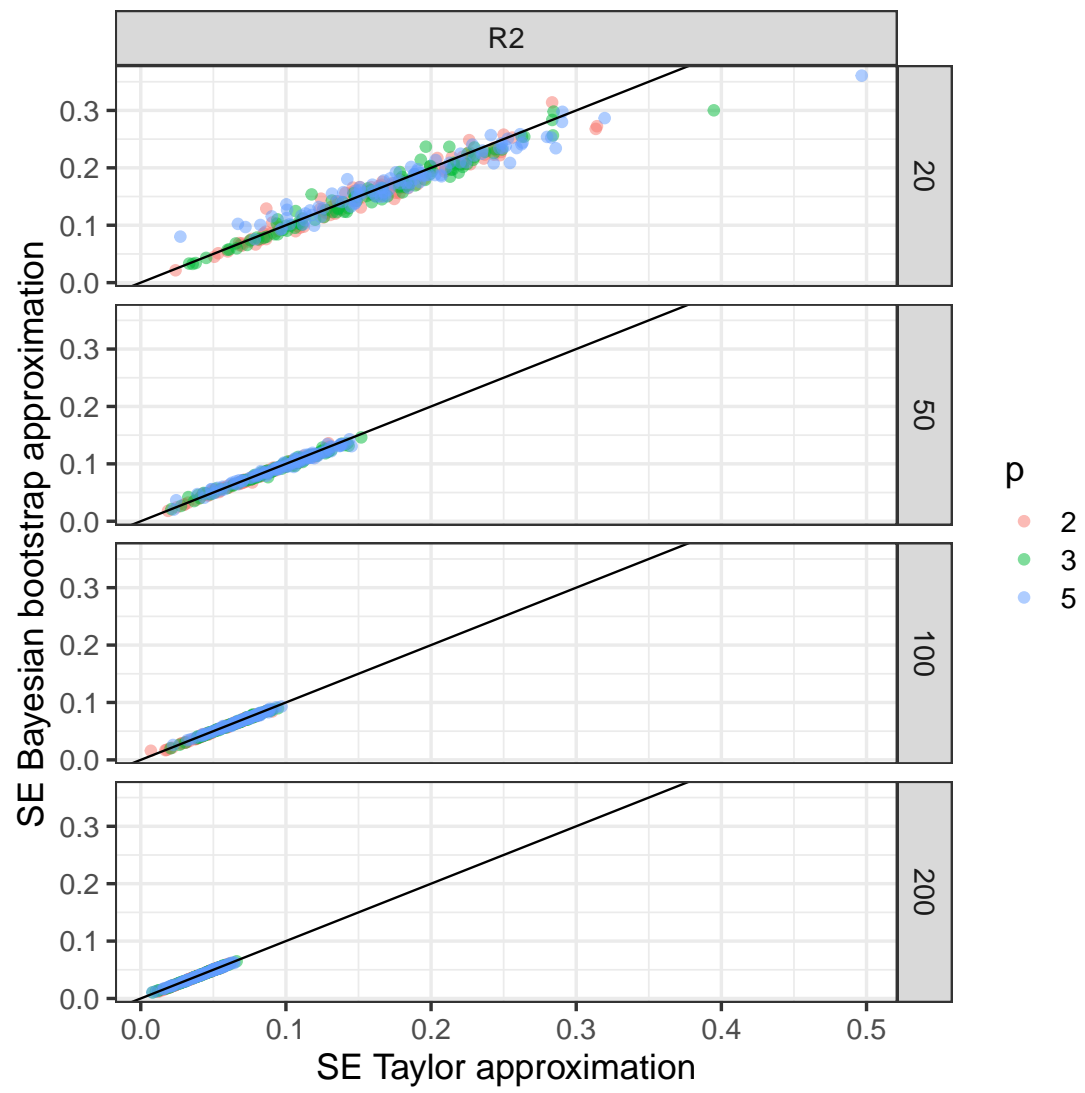


Figure 4: Loo R2

References

- W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 1970.
- Donald B. Rubin. The Bayesian Bootstrap. *The Annals of Statistics*, 9(1):130–134, January 1981. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176345338. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-9/issue-1/The-Bayesian-Bootstrap/10.1214/aos/1176345338.full>. Publisher: Institute of Mathematical Statistics.
- Tuomas Sivula, Måns Magnusson, Asael Alonzo Matamoros, and Aki Vehtari. Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison. *arXiv:2008.10296 [stat]*, March 2022. URL <http://arxiv.org/abs/2008.10296>. arXiv: 2008.10296.
- Aki Vehtari and Jouko Lampinen. Bayesian Model Assessment and Comparison Using Cross-Validation Predictive Densities. *Neural computation*, 14:2439–68, November 2002. doi: 10.1162/08997660260293292.
- Aki Vehtari and Janne Ojanen. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6(none):142–228, January 2012. ISSN 1935-7516. doi: 10.1214/12-SS102. URL <https://projecteuclid.org/journals/statistics-surveys/volume-6/issue-none/A-survey-of-Bayesian-predictive-methods-for-model-assessment-selection/10.1214/12-SS102.full>. Publisher: Amer. Statist. Assoc., the Bernoulli Soc., the Inst. Math. Statist., and the Statist. Soc. Canada.
- Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *arXiv:1507.04544 [stat]*, September 2016. doi: 10.1007/s11222-016-9696-4. URL <http://arxiv.org/abs/1507.04544>. arXiv: 1507.04544.