The Selection Bayesian Polynomial Regression with INLA using DIC, WAIC and CPO.

**Hery Tri Sutanto1, Henny Pramoedyo2 , Wayan Surya Wardhani2, Suci Astutik2**

1Student Mathematics Department, Brawijaya University, Jl. Veteran, Malang, Indonesia

2Statistics Department, Brawijaya University, Jl. Veteran, Malang, Indonesia

herysutanto@student.ub.ac.id

Abstract

The polynomial regression model is extended the multiple linear regression.The selection bayesian polynomial regression model with INLA required Criterion . Criterion is using the measure fit model with the avalaible data. There criterio : DIC, WAIC and CPO. The smaller criterion value from DIC, WAIC and CPO on a model show the best bayesian polynomial regression model with INLA.

Key word:

 Bayesian INLA, polynomial regression , DIC, WAIC, CPO, smaller value..

1. **Introduction**
	1. The background

 The regression analysis is the statistics model which use for study relationship between the one dependent variable with the one or more independent variable (*Draper* and *Smith*, 1998). The construct regression model fit for statistic data required “plot diagram” for the interest variable. The using regression model have two objectif:

1. Predict relationship between one or more independent varaiable with one dependent variable.
2. Predict the expected value for the dependent variable when the independent variable is given (Agung, 1988)

 The one from the regression model using for study relationship between the dependent variable with independent variable in the polynomial function form is the polynomial regression. Polynomial regression is the multiple linear regression which is constructed with add the impact independent variable from the little degree to the highest degree(Johnson, 1987).

 The difficult on the polynomial regression model is determine the degree polynomial regression(Guttman, 2005).The polynomial degree with the classic statistics(*frequentist*), yaitu (1) Thompson on 1978 is using the variable selection with *foreward selection* or *backward elemination* and (2) Akaike on 1973 the model selection the polynomial degree with *Akaike Information Criterion.*

On 2009 Rue development Integrated Nested Laplace Approximation(INLA) as alternative the Bayesian method. INLA is a approximation which have fast and exact compare MCMC. INLA is constructed use the *Latent Gaussian Model with product output fast and exact compare*  MCMC. The compute Bayesian INLA fast than MCMC, because:a) the computation parameters joint posterior distribution model can be instead the each parameter posterior marginal distribution on model model(Gomes, 2020), b) the computation GMRF(Gaussian Markov Random Fields) will decrease time for running,c) Rue(2009) compute posterior marginal distribution for the parameter model with Laplace Aproximation.

 The selection best bayesian polynomial regression model from Bayesian polynomial regression model candidat use the DIC, WAIC and CPO criterion. This criterion is used for measure the fit model with the given data.The first Bayes Information Criterion proposed by Schwarz on 1978, then the revised Kass and Raftery (1995) develop WBIC (Watanabe, 2010). There are other criterion popular DIC(Deviance Information Criterion) by Spiegelhalter et all (2002). The first Akaike Information Criterion(AIC) proposed by Akaike on 1973, become WAIC1 (Watanabe, 2010) and WAIC2(Watanabe, 2010). Then CPO(Conditional Predictive Ordinate) criterion proposed by Geisser and Eddy( 1979).

Regard to the background above, here the following problem in this study:1) how determine the best bayesian polynomial regression model with INLA using DIC method, 2).how determine the best bayesian polynomial regression model with INLA using WAIC method and 3).how determine the best bayesian polynomial regression model with INLA using CPO method

**2. Study Literature**

## 2.1 Polynomial Regression

## If Y is the respon variable and *x* the independent variable, the degree j polynomial regression model :

 (2.1)

If take sample with n number *n*, than model for each observation:

 (2.2)

With asumption ,

### 2..2 Approximation Bayesian Inference with INLA

If defining a *Latent Gaussian Model* for Bayesian inference in INLA on all latent parameters and hyperparameter with compute the marginal posterior distribution for each element of the parameter vector.

 (2.3)

And for each element of the hyperparameter vector

 (2.4)

With = latent parameter number

 s = hyperparameter number

 This integral-integral are not the analytic solution than the marginal posterior distribution are then approximated by

 (2.5)

and

 (2.6)

 The marginal posterior distribution approximation (2.5) and (2.6) then the hyperparameter joint posterior distribution with the Laplace approximation using equation (2.4) and the numerical integration on hyperparameter for

Solution the marginal posterior distribution process through INLA algoritm as follows:

Exploration for ̃( |𝒚) : for the parameter joint posterior distribution, INLA search the parameter space and in selection to detect good points . Rue eat al (2009) propose two different schemes: the grid strategy and the central composite design,strategy,

2.3 Model Selection

 If the best model selection for several model possibility required a criterion is used for measure the model fit with the avalaible data(Faraway, 2016). Need criterion which is used for maeasure consistent with the model. There several criterion which is used for the best model selection.

1. The Bayes Information Criterion (BIC)

The Bayes Information Criterion(BIC) proposed by Schwarz(1978) with definition:

 -log (2.7)

Where : n is sample size

 is the maksimum likelihood estimator from parameter

 is dimention for parameter

 On 1995, BIC criterion is revised by Kass and Wasserman and Kass and Raftery with definition

 + (2.8)

 is the deviance measure from model with definition

 Then : (2.9)

where

 : the initial model which interest

 the other model

Marginal Likelihood with definition with is the prior density on parameter . Then BIC= -2 log + k log n (2.10)

WBIC is the free energy estimator F(D) for the singular model (Watanabe, 2010), with is a the posterior distribution under inverse temperature .

 is derive from BIC for the singular model with definition

 (2.11)

with

1. Akaike Information Criterion(AIC)

Distance Kullback-Leibler is measure from the interest model to the true model, which have density g(y) with definition

 (2.12)

 Akaike Information Criterion with definition

 (2.13)

With k is number parameter

 is the parameter estimator obtained from maksimum likelihood method

 Value AIC smaller is show the selection best model from model-model possibility.

WAIC is derived from the singular learn theory (Watanabe, 2009) as the new information Criterion for the singular model. Watanabe is introduced 4 quantity:

 (Bayes generalization), (Bayes training loss), (Gibbs generalization loss), (Gibbs training loss)

 (2.14)

 (2.15)

 (2.16)

 (2.17)

 where:

 = the true distribution

 =posterior distribution

 = predictive distribution

 WAIC1 is estimator from derive AIC for the regular model (Watanabe, 2010) with (2.18)

 WAIC2 is estimator from (Watanabe, 2010) with

 (2.19)

1. Deviance Information Criterion

Kriteria deviance Information Criterion is be proposed by Spiegelhalter et all (2002) with definition:

 The Bayesian model for random variable is used compute the expected deviance under posterior distribution as the fit model measure. This measure is show the effective number parameter

then DIC=

1. CPO

CPO(Conditional Predictive Ordinate) is the cross-validation method for estimation the leave-one-out predictive distribution(Geisser and Eddy, 1979). The posterior predictive distribution fot observation i with definition :

where

 is observation to i from y

 is observation to i out y

 Then CPO is the posterior probability observation when model without .

The biggest value from CPO is show a the best adjustment . For the complete model sthe each time CPO approximation is obstained with the sampling posterior distribution :

 Where is a sample from and T is number sample posterior. Inverse from mean posterior for inverse likelihood. This method is efficients ini effisen but required the numerical computation (Held, Schrodle and Rue, 2010)

If INLA is computed CPO the efficient less.

The smallest value from log CPO is show the best model.

3.Discussion

* 1. Data and scatter diagram

If data on table 3.1 about machine setting and energy consumption number

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3.1

|  |  |  |
| --- | --- | --- |
| Number | Machine setting | Energy consumption |
| 1 | 11.15 | 21.6 |
| 2 | 15.17 | 4 |
| 3 | 18.9 | 1.8 |
| 4 | 19.4 | 1 |
| 5 | 21.4 | 1 |
| 6 | 21.7 | 0.8 |
| 7 | 25.3 | 3.8 |
| 8 | 26.4 | 7.8 |
| 9 | 26.7 | 4.3 |
| 10 | 29.1 | 36.2 |

 |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Data on table3.1 is plotted obstained scatter diagram figure3.1 |  |  |



Figure 3.1

The plot data from scatter diagram figure 3.1 is obstained several the polynomial regression possibility : the linear regression model , the square regression model, the degree 3 polynomial regression model and the degree 4 polynomial regression model .

* 1. The Bayesian Linear Regression Model with INLA

The Bayesian Linear Regression Model with INLA with program.

summary(m.gaussianpol1)

Call:

 c("inla(formula = formula1, family = \"gaussian\", data = datakonsmes, ", "

 control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))" )

Time used:

 Pre = 1.28, Running = 0.33, Post = 0.249, Total = 1.86

Fixed effects:

 mean sd 0.025quant 0.5quant 0.975quant mode kld

(Intercept) 1.089 16.443 -31.817 1.088 33.945 1.088 0

mes 0.329 0.740 -1.153 0.329 1.808 0.329 0

The model has no random effects

Model hyperparameter

 mean sd 0.025quant 0.5quant 0.975quant mode

Precision for 0.008 0.004 0.003 0.008 0.017 0.007

the Gaussian observations

Expected number of effective parameters(stdev): 2.00(0.00)

Number of equivalent replicates : 5.00

Deviance Information Criterion (DIC) ...............: 82.41

Deviance Information Criterion (DIC, saturated) ....: 16.28

Effective number of parameters .....................: 3.14

Watanabe-Akaike information criterion (WAIC) ...: 85.71

Effective number of parameters .................: 4.84

Marginal log-Likelihood: -54.35

CPO and PIT are computed

Posterior marginals for the linear predictor and

 the fitted values are computed

* 1. The Bayesian Square Regression Model with INLA

The Bayesian Square Regression Model with INLA with program.

summary(m.gaussianpol2)

Call:

 c("inla(formula = formula2, family = \"gaussian\", data = datakonsmes, ", "

 control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))" )

Time used:

 Pre = 1.22, Running = 0.526, Post = 0.454, Total = 2.2

Fixed effects:

 mean sd 0.025quant 0.5quant 0.975quant mode kld

(Intercept) 130.056 26.574 76.449 130.163 182.947 130.323 0

mes -13.207 2.683 -18.550 -13.218 -7.802 -13.234 0

meskw 0.331 0.065 0.200 0.331 0.460 0.332 0

The model has no random effects

Model hyperparameters:

 mean sd 0.025quant 0.5quant 0.975quant mode

Precision for 0.036 0.017 0.011 0.033 0.076 0.028

the Gaussian observations

Expected number of effective parameters(stdev): 2.99(0.005)

Number of equivalent replicates : 3.34

Deviance Information Criterion (DIC) ...............: 69.15

Deviance Information Criterion (DIC, saturated) ....: 17.54

Effective number of parameters .....................: 4.27

Watanabe-Akaike information criterion (WAIC) ...: 71.37

Effective number of parameters .................: 4.88

Marginal log-Likelihood: -52.94

CPO and PIT are computed

Posterior marginals for the linear predictor and

 the fitted values are computed

 Value DIC (69.15072) and WAIC(71.37171) on the bayesian square polynomial regression model with INLA less than value DIC(82.40976) and WAIC(85.70997) the Bayesian linear regression model with INLA then the best Bayesian square polynomial regression model with INLA.

* 1. The Bayesian degree 3 polynomial regression Model with INLA

The Bayesian degree 3 polynomial regression Model with INLA with program.

summary(m.gaussianpol3)

Call:

 c("inla(formula = formula2, family = \"gaussian\", data = datakonsmes, ", "

 control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))" )

Time used:

 Pre = 1.54, Running = 0.187, Post = 0.0939, Total = 1.82

Fixed effects:

 mean sd 0.025quant 0.5quant 0.975quant mode kld

(Intercept) -38.113 73.092 -179.873 -39.528 112.280 -41.815 0

mes 15.075 11.977 -9.537 15.309 38.332 15.685 0

meskw -1.157 0.622 -2.364 -1.169 0.124 -1.189 0

mesp3 0.025 0.010 0.004 0.025 0.045 0.025 0

The model has no random effects

Model hyperparameters:

 mean sd 0.025quant 0.5quant 0.975quant mode

Precision for 0.062 0.031 0.018 0.057 0.136 0.047

the Gaussian observations

Expected number of effective parameters(stdev): 3.85(0.074)

Number of equivalent replicates : 2.59

Deviance Information Criterion (DIC) ...............: 64.78

Deviance Information Criterion (DIC, saturated) ....: 18.68

Effective number of parameters .....................: 5.26

Watanabe-Akaike information criterion (WAIC) ...: 65.92

Effective number of parameters .................: 4.88

Marginal log-Likelihood: -58.14

CPO and PIT are computed

Posterior marginals for the linear predictor and

 the fitted values are computed

Value DIC (64.78352) and WAIC(65.92286) on Bayesian degree 3 polynomial regression model less than value DIC(69.15072) dan WAIC(71.37171) on Bayesian square regression model with iNLA the best bayesian degree 3 polynomial regression model with INLA .

1.4 Bayesian Degree 4 Polynomial Regression Model with INLA

 Bayesian Degree 4 Polynomial Regression Model with INLA with program.

Call:

 c("inla(formula = formula3, family = \"gaussian\", data = datakonsmes, ",

 control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))" )

Time used:

 Pre = 1.37, Running = 0.39, Post = 0.109, Total = 1.87

Fixed effects:

 mean sd 0.025quant 0.5quant 0.975quant mode kld

(Intercept) 234.942 152.679 -71.762 237.888 523.976 243.378 0

mes -46.476 32.923 -108.902 -47.056 19.390 -48.127 0

meskw 3.813 2.566 -1.341 3.863 8.671 3.954 0

mesp3 -0.147 0.086 -0.309 -0.149 0.028 -0.153 0

mesp4 0.002 0.001 0.000 0.002 0.004 0.002 0

The model has no random effects

Model hyperparameters:

 mean sd 0.025quant 0.5quant 0.975quant mode

Precision for 0.137 0.091 0.029 0.115 0.374 0.078

the Gaussian observations

Expected number of effective parameters(stdev): 4.30(0.128)

Number of equivalent replicates : 2.33

Deviance Information Criterion (DIC) ...............: 57.51

Deviance Information Criterion (DIC, saturated) ....: 17.81

Effective number of parameters .....................: 5.14

Watanabe-Akaike information criterion (WAIC) ...: 59.56

Effective number of parameters .................: 5.55

Marginal log-Likelihood: -66.41

CPO and PIT are computed

Posterior marginals for the linear predictor and

 the fitted values are computed

Because DIC(57.51199) and WAIC(59.55811) on Bayesian degree 4 polynomial regression model with INLA less than value DIC (64.78352) and WAIC(65.92286) on Bayesian degree 3 polynomial regression with INLA then the best Bayesian degree 4 polynomial regression Model with INLA.

 Conclusion

Fron the problem formulation and discussion can be taken conclusion . The best Bayesian degree 4 polynomial regression model with INLA fit for data relationship between machine setting variable with consumption energy variable Order . Need the four Bayesian polynomial regression model compare joint then determine the Bayesian polynomial regression model fit model with avalaible data.

Reference:

 1. Agung, G., Ng.,(1988),” Garis Patah Paritas Pengembangan Suatu Metode Untuk Memperkirakan Fertilitas”, Pusat Penelitian Kependudukan Universitas Gadjah Mada, Yogyakarta.

 2. Blangiardo,M. et al (2015). Spatial and Spatio-temporal Bayesian Models with R\_INLA,Wiley, UK.

3. Moraga,P. (2019). Geospatial health Data Modeling and Visualization with R\_INLA and Shiny, Chapman & Hall/CRC Biostatistics Series, UK

1. Omez,V. et al (2019). Bayesian Inference With INLA, CRC Press, Spain.
2. Righthetto,R., J. (2017). The impact of choosing different meshes under INLA/SPDE framework for geostatistical modelling , University of Sao Paulo “Luiz de Queiroz” College of Agriculture.
3. Satoh, S. et al(2017). How New Information Criteria WAIC and WBIC Worked for MLP Model Selection, Japan.
4. Wang, X. et al(2018). Bayesian Regression Modeling With INLA, CRC Press, UK.