# Analysis of Predictive Uncertainty Measures of Regression and Bayesian Methods


Dan Lu[1] (dl07f@fsu.edu), Ming Ye[1] and Mary C. Hill[2]

[1]Department of Scientific Computing, Florida State University, Tallahassee, FL

[2]U.S. Geological Survey, Boulder, CO


## Abstract


Predictive uncertainty is always quantified using *confidence/probability* intervals constructed around predictions. The confidence intervals are evaluated using regression inferential statistics, while the probability intervals are obtained using Bayesian methods such as Markov Chain Monte Carlo techniques. These two approaches are conceptually different and only mathematically equivalent under certain conditions. Using simple test cases, we found that, for linear models, the two types of intervals are mathematically equivalent with proper choice of prior probability. However, for nonlinear models, regardless of choice of prior probability, the two types of intervals are always different; the discrepancy depends on the model total nonlinearity. This work was then extended to a controlled numerical experiment of groundwater flow modeling developed based on Hill et al. (1998). For the complex synthetic groundwater problem, the MCMC probability intervals are always narrower than the confidence intervals obtained using linear/nonlinear regression methods. The results of this study provides theoretical basis of quantifying predictive uncertainty.


## Confidence Intervals and Probability Intervals

**Confidence interval:** The Frequentist confidence interval is interpreted in the context of a large number of different data sets. The data sets differ in the random error realization. If each data set was used to produce one confidence interval, 95% of the calculated confidence intervals would include the true value of parameter. In this philosophy, 95% is not a probability, but percent of the time in repeated sampling that the confidence intervals contain true parameter.

**Probability interval:** In Bayesian statistics, a parameter is thought of as a random variable with its own distribution rather than as a constant. The posterior distribution summarized the state of knowledge about unknown parameters conditional on the prior and current data. The amount is measured by a probability interval which is a probabilistic region around posterior statistics. In this philosophy, 95% is the posterior probability that parameter lies in the interval.

For a linear model: $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ with parameters $\boldsymbol{\beta}$, true errors $\boldsymbol{\varepsilon}\sim N_n(\mathbf{0},\mathbf{C}_\varepsilon)$, where $\mathbf{C}_\varepsilon=\sigma^2\boldsymbol{\omega}^{-1}$,
The $(1-\alpha)\times 100\%$ linear confidence interval for prediction $g(\boldsymbol{\beta})=\mathbf{Z}\boldsymbol{\beta}$ is

$$g(\hat{\boldsymbol{\beta}})\pm t_{1-\alpha/2}(n-p)[s^2\mathbf{Z}^T(\mathbf{X}^T\boldsymbol{\omega}\mathbf{X})^{-1}\mathbf{Z}]^{1/2} \tag{1}$$

when $\sigma^2$ is unknown; and

$$g(\hat{\boldsymbol{\beta}})\pm z_{1-\alpha/2}[\mathbf{Z}^T(\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{Z}]^{1/2} \tag{2}$$

when $\sigma^2$ is known.

For noninformative priors, $p(\boldsymbol{\beta})\propto \text{constant}$ and $p(\sigma)\propto 1/\sigma$, the posterior distribution of $g(\boldsymbol{\beta})$ is multivariate *t*-distribution. Thus, its $(1-\alpha)\times 100\%$ probability interval is:

$$g(\hat{\boldsymbol{\beta}})\pm t_{1-\alpha/2}(n-p)[s^2\mathbf{Z}^T(\mathbf{X}^T\boldsymbol{\omega}\mathbf{X})^{-1}\mathbf{Z}]^{1/2} \tag{3}$$

same with the linear confidence interval in equation (1).

For informative conjugate prior with $p(\boldsymbol{\beta})\sim N_p(\boldsymbol{\beta}_p,\mathbf{C}_p)$, and assume $\sigma^2$ is known, the posterior distribution of $g(\boldsymbol{\beta})$ is multivariate normal. Thus, its $(1-\alpha)\times 100\%$ probability interval is:

$$g(\boldsymbol{\beta}_p')\pm z_{1-\alpha/2}[\mathbf{Z}^T(\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X}+\mathbf{C}_p^{-1})^{-1}\mathbf{Z}]^{1/2} \tag{4}$$

As $\mathbf{C}_p\to\infty\mathbf{I}$, the probability interval reduces to the form of equation (2).

In general, for a model: $\mathbf{y}=f(\boldsymbol{\beta})+\boldsymbol{\varepsilon}$ with parameters $\boldsymbol{\beta}$, true errors $\boldsymbol{\varepsilon}\sim N_n(\mathbf{0},\mathbf{C}_\varepsilon)$ with known $\mathbf{C}_\varepsilon$
Based on Bayesian theorem, with noninformative prior, the posterior density of parameter $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta}\mid\mathbf{y})=\frac{\exp[\log p(\mathbf{y}\mid\boldsymbol{\beta})]}{\int\exp[\log p(\mathbf{y})]d\boldsymbol{\beta}} \tag{5}$$

Consider a Taylor series expansion of $\log p(\mathbf{y}\mid\boldsymbol{\beta})$ about $\hat{\boldsymbol{\beta}}$ to the second order term, where $\hat{\boldsymbol{\beta}}$ maximizes $\log p(\mathbf{y}\mid\boldsymbol{\beta})$. Then equation (5) is approximated by:

$$p(\boldsymbol{\beta}\mid\mathbf{y})\cong\frac{\exp\left[\log p(\mathbf{y}\mid\hat{\boldsymbol{\beta}})-\frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T I(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right]}{\int\exp\left[\log p(\mathbf{y}\mid\hat{\boldsymbol{\beta}})-\frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T I(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right]d\boldsymbol{\beta}} \tag{6}$$

$$=\frac{\exp\left[-\frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T I(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right]}{(2\pi)^{p/2}\,|I(\hat{\boldsymbol{\beta}})|^{-1/2}}$$

where $I(\hat{\boldsymbol{\beta}})=-\left[\frac{\partial^2\log p(\mathbf{y}\mid\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$

## Relationship between Two Intervals

When the model is linear with $f(\boldsymbol{\beta})=\mathbf{X}\boldsymbol{\beta}$ the posterior density $p(\boldsymbol{\beta}\mid\mathbf{y})\sim N_n\left(\hat{\boldsymbol{\beta}},[I(\hat{\boldsymbol{\beta}})]^{-1}\right)$ exactly with $[I(\hat{\boldsymbol{\beta}})]^{-1}=\left[\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X}\right]^{-1}$. In this case, the probability interval of $g(\boldsymbol{\beta})$ from posterior distribution is mathematically equivalent with its confidence interval in equation (2) from regression.

When the model is nonlinear, equation (6) approximates posterior density $p(\boldsymbol{\beta}\mid\mathbf{y})$ by ignoring three and higher order terms. If the model is highly nonlinear, as indicated by large total nonlinearity, ignoring the higher order terms will cause significant error. In this case, the confidence intervals and probability intervals may have large discrepancy.

## Simple Test Cases

**Linear test problems**: linear model $\mathbf{y}=a\mathbf{x}+b+\boldsymbol{\varepsilon}$, with parameters a and b; true errors $\varepsilon_i\sim N(0,1)$; we consider conjugate prior of two parameters with $\mathbf{C}_p\to\infty\mathbf{I}$.

**Nonlinear test problem**: nonlinear model $\mathbf{y}=\mathbf{x}/a+\sin(ab\mathbf{x})+\boldsymbol{\varepsilon}$, with parameters a and b; true errors $\varepsilon_i\sim N(0,1)$, we consider conjugate prior of two parameters with $\mathbf{C}_p\to\infty\mathbf{I}$.

**Purpose:** compare confidence interval using regression and probability interval using Markov Chain Monte Carlo of prediction at x=30.

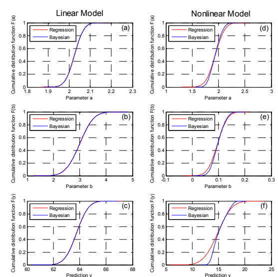


Figure 1: Cumulative distribution functions of parameters and prediction based on regression and Bayesian theory for parameter a, parameter b, and prediction y in both the linear test case (a, b, and c); and the nonlinear test case (d, e, and f).

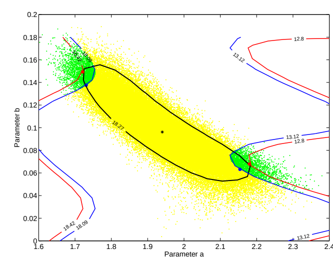


Figure 2: The nonlinear confidence interval limits (red dots), the minimum and maximum values of prediction (red lines), the confidence region of parameter set bounded by the objective function goal (black contour); the probability interval limit (blue dot), where the upper 2.5% and lower 2.5% prediction values include the samples indicated by green dots, and the median 95% prediction values include the samples indicated by yellow dots.

For linear models, confidence intervals and probability intervals are equivalent; but for nonlinear models, the values of these two types of intervals are different. In the simple test case, probability interval is smaller than confidence interval.

## Complex Synthetic Groundwater Problem

**Synthetic groundwater problem:** one true model and three alternative models with different complexity.

**Purpose:** compare confidence interval using regression and probability interval using Markov Chain Monte Carlo of two predictions: drawdown at pumping well P3 and the percent streamflow change at gauge site G2.

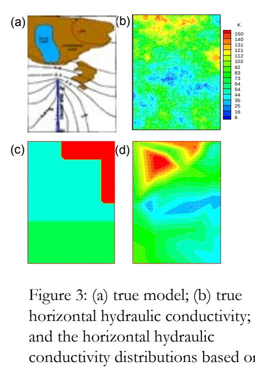

Figure 3: (a) true model; (b) true horizontal hydraulic conductivity; and the horizontal hydraulic conductivity distributions based on (c) model 3Z and (d) model Int.

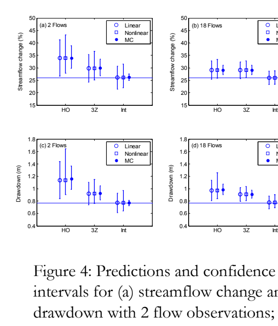


Figure 4: Predictions and confidence intervals for (a) streamflow change and (c) drawdown with 2 flow observations; (b) streamflow change and (d) drawdown with 18 flow observations. The straight lines in each figure represent the true prediction.

For the complex synthetic groundwater problem, the MCMC probability intervals are always narrower than the confidence intervals obtained with regression methods.