Evaluation of Frequentist and Bayesian Inferences by Relevant Simulation

Yuntae Kim
Korea Institute for Defence Analyses
Cheong Ryang P.O. Box 250
Seoul, Korea
ytkim@kida.re.kr

Overview

A reasonable statistical inference procedure would be justified by some validity and performance proved under some basic assumptions. But in many real situations for statistical procedures, the validity and performance are affected by further assumptions and decisions inevitably involved in the procedures. Here are some examples:

- The classical strategy in the regression model having a suite of highly correlated covariates is to select a subset model, and then make the inference assuming the subset model. We call this kind of method a selection and estimation (S/E) procedure. All the justification of the inferences coming from the subset model can not be guaranteed since it ignores the uncertainty in the selected subset model; and
- In a parametric model with nuisance parameters, an estimate of the nuisance parameter is often substituted for true value in the inference procedure. Ignoring the uncertainty in the estimated nuisance values could invalidate the inference.

When we get an inference result from such statistical procedures, we may want to check whether or not the additional assumptions and decisions distort validity and performance of the inference. However, the situation can be too complicated to get an analytical evaluation for the inference in many cases. For example, the complicated selection step in the S/E procedure for regression modeling, makes it impossible to get an analytical evaluation tool. Simulation study could be an alternative way of evaluating such a complicated inference procedure, but the evaluations coming from general simulation study would only apply to the specific simulated circumstances. For the simulation in the parametric model, the simulation result may depend on the chosen parameter value.

We suggest an evaluation methodology relying on an observation-based simulation for general frequentist and Bayesian inferences on parametric model. The suggested methodology can be applied to any inference procedure, no matter how complicated, as long as it can be codified for repetition on the computer. Unlike general simulation, the suggested methodology provides the evaluation result which does not depend on the parameter value.

Generally speaking, evaluation of an object has three main components: Which factor of the object is interesting for evaluation? How does one measure the factor? How does one evaluate the object based on the measurement for the factor? Here are our suggestions for these components for the evaluation of the general inference. We also provide a method for fixing up the inference methods found to be invalid under the suggested evaluation methodology.

Our suggestion for the factors of interest are the ideal sampling properties of the inference. For examples, a valid confidence interval guarantees the coverage probability of including true parameter in the sampling space is larger than the nominal value. The justification for the P-value of the frequentist test of a simple null hypothesis is that repeated samplings of P-value under the null follow uniform distribution over (0,1). We will adopt this frequentist criterion for both frequentist and Bayesian inferences. It seems natural to expect the frequentist inferences to have the ideal sampling properties. Though Bayesian inference procedures are not designed for ideal sampling properties, it is also expected to satisfy them well in order to be a reasonable statistical framework.

All the interesting ideal sampling properties of inferences could be expressed as the conditions for parameter dependent frequentist risks over a portion of parameter space. For example, achievement of the nominal coverage probability of a ν-level confidence interval $CI_{\nu}(X)$ for some interesting function $f(q)$ of the parameter with parameter space Θ can be expressed as

$$E(I[CI_{n}(X) \ni f(q)] \mid q = q^{*}) \geq n, \forall q^{*} \in \Theta .$$

Relevant simulation methodology (RSM) is suggested for parameter free measurement for the risks of interesting ideal sampling properties. The main idea of RSM is to integrate out the parameter in the risk with a relevant measure of parameter to common sense and observations. Our choice for the relevant measure is based on posterior probability of parameter with non-informative prior. The basic idea of RSM, emphasizing the relevant region of parameter to the common sense and observation is not a new one. There are many simulation studies emphasizing the simulation results on the parameter region the researcher feels relevant. Bootstrap is also a well-known observation-based simulation methodology. The Bayesian procedures integrating over the parameter space with posterior distribution of parameter seem to be closer to the our RSM. The simulation-based RSM could provide not only an estimate for the risk, but also a measure of precision, such as standard error. Using these quantities, formal inference procedures like confidence intervals and hypothesis tests could be provided to determine whether the risk condition is satisfied. Additionally, we provide the adjustment methods for the inferences founded to be invalid with respect to an ideal sampling property by the suggested evaluation scheme.

RESUME

Yuntae Kim received his Ph. D. in Statistics from North Carolina State University in 2000. He has worked as a researcher for military analysis in the Department of Force Development at Korea Institute for Defense Analysis since 1990.