Predictive performance of Bayesian diagnoses

Isogawa, Naoki
Pfizer Japan Inc., Clinical Statistics
3-22-7, Yoyogi, Shibuya-ku
Tokyo, 151-8589, Japan
E-mail: Naoki.Isogawa@pfizer.com

Masashi Goto
Nonprofit Organization Biostatistical Research Association
2-22-10-A411, Kami-shinden, Toyonaka-shi
Osaka, 560-0085, Japan
E-mail: gotoo@bra.or.jp

1 Introduction

In a process of Bayes inference which formulate iterative procedure of scientific research, we select a prior distribution based on cumulative experiences, experiments and knowledge, and compose a probability model under the prior distribution. If the model is correct, we can make proper inferences about parameter using a posterior distribution which is combination of prior information with data information. However, the posterior distribution is composed using only a pair of data that has actually occurred, so it is important to make diagnosis and checking for the model. Then, we can diagnose the model in the following three terms at least: (1) Sensitivity analysis for changes of prior distribution and likelihood, (2) Appropriateness of posterior inference for the model in the context of the actual situation, (3) Fitness between the model and the data. In this paper, we notice on a model diagnosis in terms of (3). In the framework of traditional model selection such as Bayesian information criteria (BIC) (Schwarz, 1978) and Bayes factor (BF), a model with highest posterior model probability is selected. However, in fact, we are often interested to data which will be gained in the future. Therefore, we consider two diagnostic methods that focus on prediction: the Bayesian predictive information criterion (BPIC) (Ando, 2007) and the prior and posterior predictive checking approach (Prior- and Post-PCA) (Box, 1980; Rubin, 1984; Gelman, Meng and Stern, 1996; Daimon and Goto, 2007).

BIC which is most familiar information criterion in Bayesian approach is a criterion of model evaluation based on a posterior probability and select a best model with the highest posterior probability among several model candidates. Bayes factor, extended Bayesian information criteria (Konishi et al., 2004) are well-known as other model evaluation criteria in the same position. Recently, BPIC has been proposed as new diagnosis method which evaluates model fitness from a position of the prediction. BPIC selects a model with the highest expected log likelihood.

Prior-PCA provides checking models or indices by comparing data to the prior predictive distribution. This approach contrasts the prior information and the data information, and checks their compatibility. Post-PCA replaces the role of the prior distribution in Prior-PCA with it of the posterior distribution. Main feature of Prior- and Post-PCA is to be able to check not only a model itself in themselves but also interesting indices or statistics by setting proper predictive checking functions. Therefore, we can judge whether the model is suitable for taking the specific occasion or not. It is considered that this feature is quite effective because we don't always have to focus on the model itself and can select a proper model which serves the purpose of the research.

Though Bayesian approach has the advantage that it is possible to select a prior distribution according to an individual situation, therefore, there exists many models which should be evaluated.

So we consider that a specified situation suitable for each model diagnosis may exist and it's also true in BPIC and PCA. But the profiles about BPIC and PCA haven't been clarified enough yet. In this paper, our purpose is to make clear the properties of BPIC and PCA and propose the effective diagnosis situations.

2 Bayesian predictive diagnosis

2.1 Bayesian predictive information criterion

Ando(2007) proposed BPIC as an estimator of the posterior mean of the expected loglikelihood of the predictive distribution. In this criteria, we can evaluate the predictive distributions of hierarchical and empirical Bayes model even when the assumed family of probability distributions doesn't always contain the true model. We set a prior probability, a likelihood function of sampling distribution and a posterior probability as $p(\theta)$, $p(y|\theta)$ and $p(\theta|y)$. Under the weak regular conditions (Unimodal of the posterior distribution, consistency of the posterior mode, asymptotic normality), BPIC is defined as follows:

$$\mathrm{BPIC} = -2\int \log\{p(y|\theta)\}p(\theta|y)d\theta + 2n\hat{b}(\hat{G})$$

where $\hat{b}(\hat{G})$ is

$$\hat{b}(\hat{G}) = \frac{1}{n}\int p(y|\theta)p(\theta)p(\theta|y)d\theta - \frac{1}{n}\log p(y|\hat{\theta}_n)p(\hat{\theta}_n) + \frac{1}{n}\mathrm{tr}S_n^{-1}(\hat{\theta}_n)Q_n(\hat{\theta}_n) + \frac{p}{2n}$$

and

$$\begin{aligned} Q_n(\hat{\theta}_n) &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial\{\log p(y_i|\theta)+\log\pi(\theta)/n\}}{\partial\theta}\cdot\frac{\partial\{\log f(y_i|\theta)+\log\pi(\theta)/n\}}{\partial\theta}\bigg|_{\theta=\hat{\theta}_n}\right] \\ S_n(\hat{\theta}_n) &= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2\{\log p(y_i|\theta)+\log p(\theta)/n\}}{\partial\theta\partial\theta^{\mathrm{T}}}\bigg|_{\theta=\hat{\theta}_n}\right], \end{aligned}$$

and $\hat{\theta}_n$ is the posterior mode, p is the number of parameter, n is the sample size. We select a lowest model of BPIC as well as other information criteria such as AIC.

2.2 Predictive checking approach

2.2.1 Prior predictive checking approach

Box(1980) proposed Prior-PCA as the diagnostic method that focused on prediction. In the prior predictive checking approach, a model is evaluated by comparing the prior predictive distributions of future observations to the data that have actually occurred and calculating the prior predictive checking probability (Prior-PCP).

In Pre-PCA, a Bayes model is given by

$$p(y,\theta) = p(y|\theta)p(\theta).$$

Then, a prior predictive distribution for the observations of the future, y, is obtained by

$$p(y) = \int_{\theta\in\Theta} p(y,\theta)d\theta.$$

Given the actual data y_d, Prior-PCA for the model itself is calculated by comparing the prior density at y_d, $p(y_d)$ to the density function $p(y)$ as the below:

$$\Pr[p(y) < p(y_d)]$$

Tab.1: BPIC and PCP for triglyceride concentration data (Wood, 1973)

		Prior distribution $\mu \sim \mathrm{N}[\mu_0, \sigma_0^2]$			
		$\mathrm{N}[125, 20]$	$\mathrm{N}[200, 20]$	$\mathrm{N}[125, 4000]$	$\mathrm{N}[200, 4000]$
BPIC		40.35	41.36	41.98	41.98
PCP	Prior-model	0.527	0.000	0.528	0.461
	Post-model	0.527	0.003	0.528	0.527
	Prior-mean	0.887	0.334	0.978	0.255
	Post-mean	0.900	0.214	0.997	0.884

We can assess not only the model itself but also interesting indices such as sample mean, sample variance or on which the decision making is based. When we express an interesting indice as $g(y)$, Pre-PCA for it is as the below:

$$\Pr[p\{g(y)\} < p\{g(y_d)\}]$$

2.2.2 Posterior Predictive checking approach

Rubin(1984) proposed Post-PCA, as an alternative method of the prior predictive checking approach. In the posterior predictive checking approach, a model is evaluated by comparing the posterior predictive distributions of future observations to the data that have actually occurred and calculating the posterior predictive checking probability (Post-PCA). In order to evaluate a model under the prior predictive distribution in Pre-PCA and the posterior predictive distribution in Post-PCA, the large difference in the prior and posterior predictive checking probabilities indicates that the prior distribution is wrong.

3 Examination on some literature example

We applied these Bayesian predictive diagnoses which were introduced in Section 2. to data of triglyceride concentration in the plasma (Wood, 1973), and evaluated the appropriate of several models. These data (sample mean 126.8, sample variance 3973) were measured to examine whether improvement in lifestyles impact on the measurements by a team in Stanford University, and we used the pre-treatment data here. The sample size was 30. We assumed that the data followed $\mathrm{N}[\mu, \sigma^2]$ where the variance σ^2 was known and the mean μ was given the normal prior distributions $\mathrm{N}[\mu_0, \sigma_0^2]$. We set the prior means μ_0 as 125 (close to sample mean) or 200 (not close to sample mean) and the prior variance σ_0^2 as 20 (strong prior information) or 4000 (weak prior information). Then, we calculated BPIC and PCP for model and sample mean and represented the results in Tab.1. In the table, we gained the results that the model with the prior distribution $\mathrm{N}[125, 20]$ (strong prior information and close to sample mean) had the lowest BPIC, but the model with the prior distribution $\mathrm{N}[125, 4000]$ (weak prior information and close to sample mean) indicated higher Prior- and Post- PCP for sample mean than the model with the prior distribution $\mathrm{N}[125, 20]$ in PCA. However, Prior- and Post-PCP for model were almost the same probabilities together. Moreover, BPIC for the model with weak prior information ($\mathrm{N}[125, 4000]$ and $\mathrm{N}[200, 4000]$) had much the same value and there wasn't difference between them.

4 Simulation

In this section, we conduct some simulations for evaluating some relative performances of Bayesian predictive diagnoses to BPIC. It is important to consider the amount of information of the prior distribution in advance because it is useful to interpret the simulation results. So we express the amount of information of the prior distribution (prior information) as "the number of observation which is required for obtaining the same estimate accuracy as Bayes estimator" (Mori, 2010) and describe it as "prior samples".

In this simulation, we assume that independent samples follow a normal distribution with known variance, $y_i \sim \mathrm{N}[\mu, \sigma^2](i = 1, 2, \cdots, n)$, and a prior distribution of a mean μ follows a normal distribution, $\mu \sim \mathrm{N}[\mu_0, \sigma_0^2(= \sigma^2/n_0)]$. When we set a square error as a loss function and calculate Bayes estimator of mean μ (expectation of posterior distribution) and Bayes risk, we can understand that the prior distribution $\mathrm{N}[\mu_0, \sigma^2/n_0(= \sigma_0^2)]$ has a information about n_0 samples. Considered to the information about these prior sample, we plan the following simulations.

4.1 Purpose and Method

1. Because BPIC for the models with the prior distribution which has strong prior information were much the same in the specified sample size, we explored the extent of prior information which was needed for selecting appropriate model with true prior mean in BPIC through the simulation. For $n = 20, 50, 100$, we calculated BPIC when the proportions of prior sample for sample size (n_0/n) were between 0.05 and 0.5, and examined how BPIC changes according to the proportions.

2. We considered whether PCA could be featured in not sample size but the proportion of the prior sample for sample size in this simulation. Because of the larger sample size the larger BPIC, we focused on only PCA which evaluated the models in PCP between 0 and 1. For $n = 20, 50, 100$, we calculated Prior- and Post-PCP for model and sample mean when the proportions for sample size (n_0/n) were 0.005, 0.05, 0.1, 0.5, 1, and compared the percents of these PCP in the same proportions by these sample size.

4.2 Result

1. The results of 25, 50, 75 percents of BPIC were displayed in Fig.6. The dotted line in these figures showed the results of the cases where prior mean was $\mu_0 = 1.5$ and the solid line in these figures showed the results of the cases where prior mean $\mu_0 = 1.5$. Then, BPIC were the same in both cases when the proportions of prior sample for the sample size were less 0.05. However, as the proportions of prior sample for the sample size became larger than 0.1, the difference of BPIC in both cases made increased. Therefore, it indicated that BPIC was suitable for model diagnoses when the proportions of prior sample for the sample size were than 0.1.

2. The results of 25, 50, 75 percents of Prior- and Post-PCP for model and sample mean were displayed in Fig.7. The left side in these figure represented the case of models with prior distribution which has true prior mean and the right side in these figure represented the case models with no true prior means. From the results, we found that Pre- PCP for model indicated almost the same values regardless of sample size when prior mean was true and the proportions were same. However, when prior mean wasn't true, Prior-PCP for model were different by sample size even though the proportions were same, and were smaller as sample size were larger. As seen in Prior-PCP for model, we also found that Post-PCP for model and Prior- and Post-PCP

for sample mean were almost same regardless of sample size when prior mean was true and the proportions were same, but Prior- and Post-PCP for model and sample mean were smaller as sample size increase when prior mean wasn't true.

From these results, we gained the findings that the proportion of prior sample for sample size (not sample size), impacted on Pre- and Post- PCP when prior mean was true.

5 Conclusion

In this paper, we focused on BPIC and PCA which evaluate models from the position of the prediction and conducted some simulations as a purpose of clarifying the features of these model diagnoses in the case of where mean parameter follows normal prior distribution. As the results, we found that BPIC was suitable for model selections in the case with strong prior information but it wasn't really in the case with weak prior information. Also, we gained the findings that models with weak prior information gave higher Prior- and Posterior- PCP for mean sample than those with strong prior information in prior distributions with true prior mean, so PCA was also suitable for model selections in the case with weak prior information contrary to BPIC. Moreover, it was indicated that if prior mean included true value and the proportions of prior sample for sample size were same, these PCP were much the same regardless of sample size. In summary from these findings, we found that an improved model selection could be achieved by using both BPIC and PCA according to the situation.

REFERENCES

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, **94**, 443-458.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J. Roy. Statist. Soc.*, A**153**, 383-430.

Daimon, T. and Goto, M. (2007). Predictive checking approach to Bayesian interim monitoring. *Japanese J.Appl. Statis.*, **36**(2 & 3), 119-137.

Gelman, A., Meng, X. L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, **6**, 733-807.

Rubin, D. B. (1984). Bayesianly justifable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, **12**, 1151-1172.

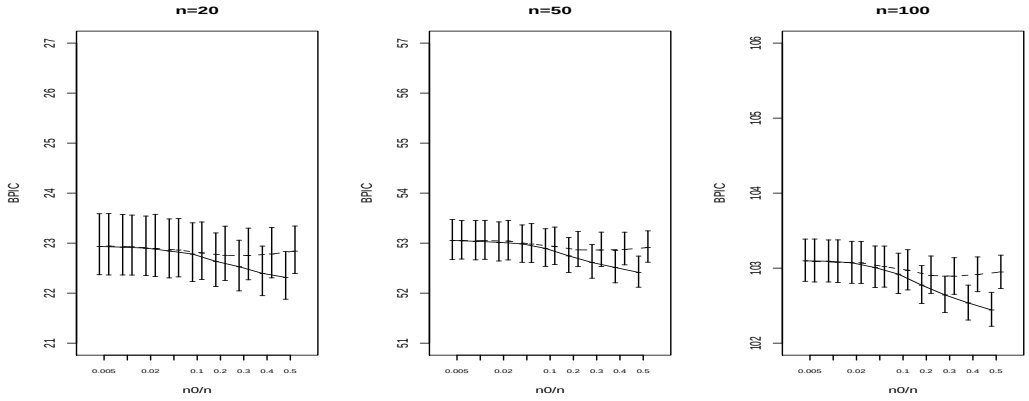

Fig.1: BPIC by the proportions of prior sample for the sample size

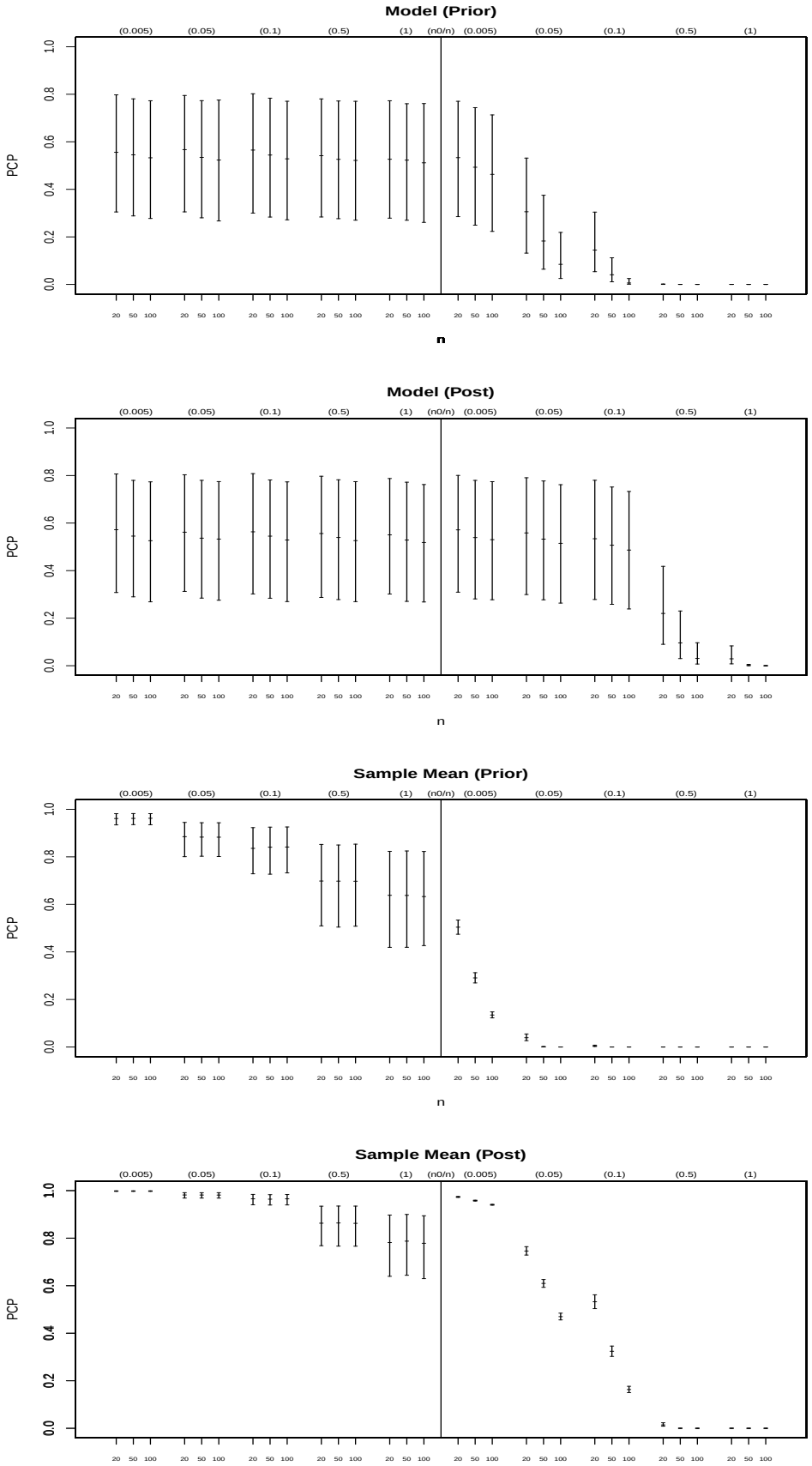

Fig.2: Prior- and Post- PCP by the proportions of prior sample for the sample size