Open Journal of Statistics, 2016, 6, 545-553
Published Online June 2016 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2016.63046

Frequentist Model Averaging and Applications to Bernoulli Trials

Georges Nguefack-Tsague[1], Walter Zucchini[2], Siméon Fotso[3]

[1]Biostatistics Unit, Department of Public Health, Faculty of Medicine and Biomedical Sciences, University of Yaounde 1, Yaounde, Cameroon
[2]Institute for Statistics and Econometrics, University of Goettingen, Goettingen, Germany
[3]Department of Mathematics, Higher Teachers' Training College, University of Yaoundé I, Yaoundé, Cameroon
Email: nguefacktsague@yahoo.fr, walter.zucchini@wiwi.uni-goettingen.de, simeonfotso@yahoo.fr

Received 26 April 2016; accepted 25 June 2016; published 28 June 2016

Abstract

In several instances of statistical practice, it is not uncommon to use the same data for both model selection and inference, without taking account of the variability induced by model selection step. This is usually referred to as post-model selection inference. The shortcomings of such practice are widely recognized, finding a general solution is extremely challenging. We propose a model averaging alternative consisting on taking into account model selection probability and the likelihood in assigning the weights. The approach is applied to Bernoulli trials and outperforms Akaike weights model averaging and post-model selection estimators.

Keywords

Model Selection, Post-Model Selection Estimator, Frequentist Model Averaging, Bernoulli Trials

1. Introduction

In statistical modeling practice, it is typical to ignore the variability of the model selection step, which can result in inaccurate post-selection inference (Berk *et al*. ([1] [2]), Belloni *et al*. ([3] [4]), Tibshirani *et al*. [5], and Chernozhukov *et al*. [6]). The model selection step is often a complex decision process and can involve collecting expert opinions, preprocessing, applying a variable selection rule, data-driven choice of one or more tuning parameters, among others. Except in simple cases, it is hard to explicitly characterize the form of the post-selection of interest while incorporating the variability of model selection. References for model selection include e.g. Zucchini [7] and Zucchini *et al*. [8]. An alternative to selecting a single model for estimation purposes is to use a weighted average of the estimates resulting from each of the models under consideration. This leads to the class of model averaging estimators. Model averaging can be done either in Bayesian and

How to cite this paper: Nguefack-Tsague, G., Zucchini, W. and Fotso, S. (2016) Frequentist Model Averaging and Applications to Bernoulli Trials. *Open Journal of Statistics*, **6**, 545-553. http://dx.doi.org/10.4236/ojs.2016.63046

frequentist approaches. The most common Bayesian approach is Bayesian model averaging (BMA) and its variants, using Bayesian information criterion (BIC) as approximation (Schwarz [9]). The seminal paper of Hoeting *et al*. [10] fully describes the basic of BMA. BMA and its applications can be found in Nguefack-Tsague ([11] [12]), Nguefack-Tsague and Ingo [13], Nguefack-Tsague and Zucchini ([14] [15]). Several options are available for specifying the weights in frequentist approaches; references on least squares regression types and like include Hansen ([16]-[21]), Hansen and Racine [22], Cheng and Hansen [23], Charkhi *et al*. [24], and Wan *et al*. [25]. The aforementioned weighting schemes perform model averaging on a set of nested candidate models with the weights vector chosen such that a specific criterion is minimized.

References using Akaike's information criterion, AIC (Akaike [26]) include Burnham and Anderson [27], Nguefack-Zucchini [28], Nguefack-Tsague ([29]-[32]). The R package [33] MuMIn is used to perform model averaging based on Burnham and Anderson [27]. Schomaker and Heumann [34], and Schomaker [35] developes model averaging schemes based on multiple imputation and shrinkage; the R package MAMI is used for practical implementations. This paper is organized as follows: In Section 2, we develop model averaging based on information criterion while, in Section 3, we propose a new approach for computing the weights for the competing models, one that takes both account the selection probability and the likelihood of each model. Section 4 illustrates with applications to Bernoulli trials. The paper ends with concluding remarks.

2. Frequentist Model Averaging Based on Information Criterion

Let $\mathcal{M}=\{M_1,\cdots,M_K\}$ be a set of K plausible models to estimate μ, the quantity of interest. Denote by $\hat{\mu}_k$ the estimator of μ obtained when using model M_k. Model averaging involves finding non-negative weights, $w_1,\cdots,w_K$, that sum to one, and then estimating μ by

$$\hat{\mu}=\sum_{k=1}^{K} w_k\hat{\mu}_k. \tag{1}$$

In model selection, the model selection criterion determines which model is to be assigned weight one, *i.e*. which model is selected and subsequently used to estimate the parameter of interest. We note that, since the value of the selection criterion depends on the data, the index, $\hat{k}$, of the selected model is a random variable. We therefore denote the selected model by $M_{\hat{k}}$, and the corresponding estimator of the quantity of interest, μ, by $\hat{\mu}_{\hat{k}}$. In terms of the notation introduced above, we may write

$$M_{\hat{k}}=\sum_{k=1}^{K} I\left(\text{model } k \text{ is selected}\right)M_k,\qquad \hat{\mu}_{\hat{k}}=\sum_{k=1}^{K} I\left(\text{model } k \text{ is selected}\right)\hat{\mu}_k.$$

Clearly, the selected model depends on the set of candidate models, $\mathcal{M}$, and on the selection procedure, which we denote by S. However, it is important to realize that, even if the same $\mathcal{M}$ and S, are used, different samples can lead to different models being selected; $M_{\hat{k}}$ is a "randomly selected model". In this section we focus attention on post-model selection estimators (PMSEs), which is the special case of model averaging estimators with zero/one weights only.

Some classical model averaging weights base the weights on penalized likelihood values. Let IC_k denote an "information criterion"of the form

$$IC_k=-2\log L_k+s_k, \tag{2}$$

where s_k is a penalty term, and L_k is the maximized likelihood value for the model M_k. The Akaike information criterion (AIC, Akaike [26]) is the special case with $s_k=2q_k$, where q_k is the number of parameters of model M_k. Buckland *et al*. [36] proposed using weights of the form:

$$w_k=\frac{\exp\left(-s_k/2\right)L_k}{\sum_{l=1}^{K}\exp\left(-s_l/2\right)L_l}=\frac{\exp\left(-IC_k/2\right)}{\sum_{l=1}^{K}\exp\left(-IC_l/2\right)}. \tag{3}$$

"Akaike weights" (denoted by $w_{aic,k}$) refer to the case with $IC_k=AIC_k$. Numerous applications of Akaike weights are given in Burnham and Anderson [27].

3. Likelihood and Selection Probability in Assigning the Weights

Since the selection procedure (S) and likelihood are important for model selection, we therefore suggest estimating μ by a weighted average of the $\hat{\mu}_k$ in which the weights take account of *S*, specifically where they depend on estimators $p(M_k | S) = \widehat{\Pr}(M_k \text{ is selected} | S),\ k = 1, \cdots, K$.

$$w_{al,k} = \frac{p(M_k | S) L_k}{\sum_{i=1}^{K} p(M_i | S) L_i}, \quad k = 1, 2, \cdots, K. \tag{4}$$

The likelihoods are taken into account because they quantify the relative plausibility of the data under each competing model; the estimated selection probability $p(M_k | S)$ adjusts the weights for the selection procedure. Both of these components are required. If one were to use only the likelihoods to determine the weights then complex models (*i.e*. models having many parameters) would automatically be assigned larger weights. The weights $w_{al,k}$ are similar to the weights w_k defined in (3) but they differ in the way the likelihood is adjusted. With the proposed method a "bad" model will be penalized by any reasonable selection procedure through the probability $p(M_k | S)$, even if it is complex in terms of the number of parameters. We let the selection procedure determine in how far a model is penalized.

If the selection probabilities depend on some parameter for which a closed form expression exists, and if one can find an estimator of the parameter, then it is possible to obtain estimators for these probabilities.

4. Applications to Bernoulli Trials

Let $X_1, \cdots, X_n$ be *n* independent Bernouilli trials, that is $X_i \sim Be(\theta)$, $Y = \sum_{i=1}^{n} X_i$ is the number of successes; *Y*-binomial (*n*, θ), θ unknown. Inference will be based on *Y*, since the likelihood function of the X_i's is $\theta^Y (1-\theta)^{n-Y}$ and involves the sufficient statistic *Y*. $f(y | \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$, $y = 0, 1, \cdots, n$, is the probability mass function (PMF) of *Y*; the quantity of interest is $\Delta = \theta$. Sensitivity analyses showed that the finding obtained here are insensitive to parameter choice, irrespective of the sample size *n*.

4.1. A Two-Model Selection Problem

(a) Consider the choice between the 2 models: $M_1 : \theta = \theta_1$ and $M_2 : \theta = \theta_2$. The true model may not belong to these 2 models. Suppose that the selection procedure chooses the model with smaller AIC. In this case, this entails to choosing the model with higher likelihood, since there is no parameter to be estimated for each model. M_1 will be chosen if $f(y | \theta_1) > f(y | \theta_2)$ or equivalently if $R = \log(f(y | \theta_1)) - \log(f(y | \theta_2)) > 0$.

$$\begin{aligned} R &= \log\left(\binom{n}{y} \theta_1^y (1-\theta_1)^{n-y}\right) - \log\left(\binom{n}{y} \theta_2^y (1-\theta_2)^{n-y}\right) \\ &= \log\binom{n}{y} + y \log\theta_1 + (n-y)\log(1-\theta_1) - \log\left(\binom{n}{y}\right) - y\log\theta_2 - (n-y)\log(1-\theta_2) \\ &= y \log\frac{\theta_1}{\theta_2} + (n-y)\log\left[\frac{1-\theta_1}{1-\theta_2}\right] = y\left[\log\frac{\theta_1}{\theta_2} - \log\left[\frac{1-\theta_1}{1-\theta_2}\right]\right] + n\log\left[\frac{1-\theta_1}{1-\theta_2}\right] \end{aligned}$$

$$R > 0 \Leftrightarrow y > \frac{-n\log\left[\dfrac{1-\theta_1}{1-\theta_2}\right]}{\log\left[\dfrac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)}\right]} = a_n(\theta_1, \theta_2).$$

Let $P_\theta(M_1 | AIC, \mathcal{M})$ and $P_\theta(M_2 | AIC, \mathcal{M}) = 1 - P_\theta(M_1 | AIC, \mathcal{M})$ be the probabilities of choosing models 1 and 2, respectively.

$$P_\theta(M_1 | AIC, \mathcal{M}) = P_\theta(Y > a_n(\theta_1, \theta_2)) = 1 - P_\theta(Y \le a_n(\theta_1, \theta_2)) = 1 - F_{B(n,\theta)}(a_n(\theta_1, \theta_2)),$$

where $F_{B(n,\theta)}$ is the cumulative distribution function of binomial (*n*, θ).

The estimated probabilities are given by $p(M_1 | AIC) = 1 - F_{B(n,\hat{\theta})}(a_n(\theta_1, \theta_2))$, where $\hat{\theta} = y/n$ and

$p(M_1 \mid AIC) = 1 - p(M_1 \mid AIC)$. The PMSE $\tilde{\theta} = \theta_1$ if $y > a_n(\theta_1, \theta_2)$ and θ_2 otherwise. The properties of $\tilde{\theta}$ are given by

$$\begin{aligned} \mathrm{E}_\theta(\tilde{\theta}) &= \sum_{y > a_n(\theta_1,\theta_2)} \theta_1 f(y \mid \theta) + \sum_{y \le a_n(\theta_1,\theta_2)} \theta_2 f(y \mid \theta) \\ &= \theta_1 \sum_{y > a_n(\theta_1,\theta_2)} f(y \mid \theta) + \theta_2 \sum_{y \le a_n(\theta_1,\theta_2)} f(y \mid \theta) = \theta_1 p_1 + \theta_2 p_2. \end{aligned}$$

$$\begin{aligned} \mathrm{Var}_\theta(\tilde{\theta}) &= \sum_{y > a_n(\theta_1,\theta_2)} \left(\theta_1 - \mathrm{E}_\theta(\tilde{\theta})\right)^2 f(y \mid \theta) + \sum_{y \le a_n(\theta_1,\theta_2)} \left(\theta_2 - \mathrm{E}_\theta(\tilde{\theta})\right)^2 f(y \mid \theta) \\ &= \mathrm{E}_\theta(\tilde{\theta})^2 - \mathrm{E}_\theta^2(\tilde{\theta}) = \theta_1^2 p_1 + \theta_2^2 p_2 - (\theta_1 p_1 + \theta_2 p_2)^2. \end{aligned}$$

$$\mathrm{Bias}_\theta(\tilde{\theta}) = \mathrm{E}_\theta(\tilde{\theta}) - \theta.$$

$$\mathrm{MSE}_\theta(\tilde{\theta}) = \mathrm{Var}_\theta(\tilde{\theta}) + \mathrm{Bias}_\theta^2(\tilde{\theta}).$$

The Akaike weights are defined by

$$W_{a_1} = \frac{f(y \mid \theta_1)}{f(y \mid \theta_1) + f(y \mid \theta_2)}, \quad W_{aka_2} = \frac{f(y \mid \theta_2)}{f(y \mid \theta_1) + f(y \mid \theta_2)}.$$

The adjusted likelihood weights are defined by

$$W_{al_1} = \frac{p(M_1 \mid AIC) f(y \mid \theta_1)}{p(M_1 \mid AIC) f(y \mid \theta_1) + p(M_2 \mid AIC) f(y \mid \theta_2)},$$

$$W_{al_2} = \frac{p(M_2 \mid AIC) f(y \mid \theta_2)}{p(M_1 \mid AIC) f(y \mid \theta_1) + p(M_2 \mid AIC) f(y \mid \theta_2)}.$$

The weighted estimators are

$$\hat{\theta}_a = \theta_1 W_{a_1} + \theta_2 W_{a_2}.$$

$$\hat{\theta}_{al} = \theta_1 W_{al_1} + \theta_2 W_{al_2}.$$

$$\mathrm{MSE}_\theta(\hat{\theta}_a) = \sum_{y=0}^{n} \left(\hat{\theta}_a - \theta\right)^2 f(y \mid \theta).$$

$$\mathrm{MSE}_\theta(\hat{\theta}_{al}) = \sum_{y=0}^{n} \left(\hat{\theta}_{al} - \theta\right)^2 f(y \mid \theta).$$

Figure 1 shows model selection probabilities for $\theta_1 = 0.6$, $\theta_2 = 0.4$ and $n = 41$ for the range of parameter space. The two curves cross at $\theta = 0.5$ showing different values of the parameters space used for weighting.

Figure 2 compares PMSE to estimators based on Akaike weights and adjusted weights using true model selection probabilities. It can be seen that adjusted likelihood is always better than PMSE and Akaike weights estimators. However, for some values of the true parameter, the risk of Akaike weight tends to be slightly bigger than that of PMSEs. Maxima occur at $\theta = 0.5$ while minima occur at 0.4 and 0.6.

(b) Consider now a choice between the following two models:

$M_1 : Y \sim \mathrm{binomial}(\theta_1, n)$ and $M_2 : Y \sim \mathrm{binomial}(\theta, n)$.

AIC is used to select a model, $\hat{\theta}_2 = y/n$, for illustration, we choose $\theta_1 = 0.5$.

$$AIC_{M_1} = -2\log\left(f(y \mid \theta_1)\right), \quad AIC_{M_2} = -2\log\left(f\left(y \mid \hat{\theta}_2\right)\right) + 2.$$

Model 1 is chosen if

$$AIC_{M_1} > AIC_{M_2}, \quad P(M_1 \mid AIC, \mathcal{M}) = P_\theta\left(AIC_{M_1} > AIC_{M_2}\right), \quad P(M_2 \mid AIC, \mathcal{M}) = P_\theta\left(AIC_{M_1} \le AIC_{M_2}\right).$$

$p(M_1 \mid AIC)$ and $p(M_2 \mid AIC)$ are obtained by replacing θ by $\hat{\theta}_2 = y/n$.

The PMSE $\tilde{\theta} = \theta_1$ if $AIC_{M_1} > AIC_{M_2}$ and $\hat{\theta}_2$ otherwise.

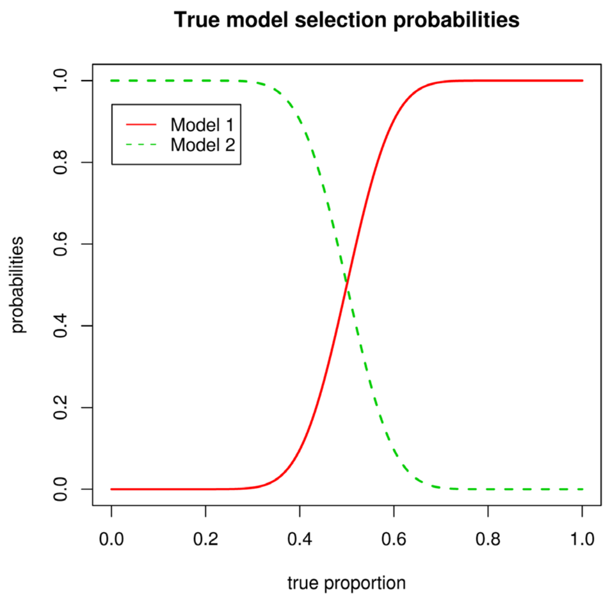

Figure 1. Model selection probabilities as a function θ, $\theta_1 = 0.6$ and $\theta_2 = 0.4$.

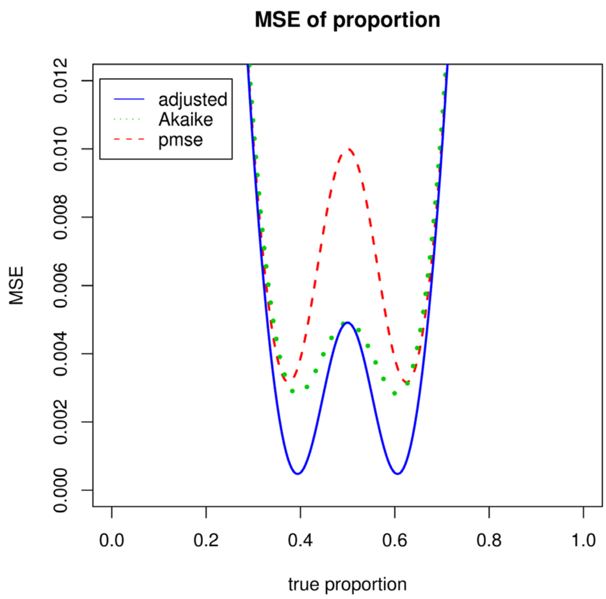

Figure 2. Risk of two simple proportions comparing PMSEs, Akaike weights estimators and adjusted estimators as a function of θ.

$$\text{MSE}_\theta\left(\tilde{\theta}\right) = \sum_{AIC_{M_1} > AIC_{M_2}} \theta_1 f\left(y \mid \theta\right) + \sum_{AIC_{M_1} \le AIC_{M_2}} \hat{\theta}_2 f\left(y \mid \theta\right).$$

The Akaike weights are defined by

$$W_{a_1} = \frac{f\left(y \mid \theta_1\right)}{f\left(y \mid \theta_1\right) + f\left(y \mid \hat{\theta}_2\right)}, \quad W_{a_2} = \frac{f\left(y \mid \hat{\theta}_2\right)}{f\left(y \mid \theta_1\right) + f\left(y \mid \hat{\theta}_2\right)}$$

and the adjusted weights is defined by

$$W_{al_1} = \frac{p(M_1 \mid AIC) f(y \mid \theta_1)}{p(M_1 \mid AIC) f(y \mid \theta_1) + p(M_2 \mid AIC) f(y \mid \hat{\theta}_2)},$$

$$W_{al_2} = \frac{p(M_2 \mid AIC) f(y \mid \hat{\theta}_2)}{p(M_1 \mid AIC) f(y \mid \theta_1) + p(M_2 \mid AIC) f(y \mid \hat{\theta}_2)}.$$

Figure 3 displays model selection probabilities with both curves crossing at 0.6 and 0.4. At 0.5, while Model 2 is at the minimum, Model 1 is at maximum. **Figure 4** displays risks performance of estimators. It can be seen that Akaike weighting does not perform better than PMSEs when the true parameter is between $(0, 0.3)$ and between $(0.7, 1)$. However, the adjusted weights perform better than both.

4.2. Multi-Model Choice

Consider also a choice between the following models: $M_k : Y \sim \text{binomial}(\theta_k, n)$ for arbitrary K models; θ_k known. For a choice using AIC criterion, since there is no unknown parameter, this is the same as selecting the model with higher likelihood. Model $M_{\max}$ is chosen if $L_{\max} \geq L_k, \forall k \in \{1, \cdots, K\}$.

PMSE $\tilde{\theta} = \theta_k$ if M_k is selected.

$\tilde{\theta} = \sum_{k=1}^{K} I_k \left(f(y \mid \theta_k) = L_{\max} \right) \theta_k$, $I_k = 1$ if M_k is chosen and 0 otherwise. Model selection probability for model M_k is given by: $P_\theta(M_k \mid AIC, \mathcal{M}) = P_\theta\left(f(y \mid \theta_k) = L_{\max} \right)$.

The estimated model selection probabilities $p(M_k \mid AIC)$ are given by replacing θ by the estimated $\hat{\theta} = y/n$. The Akaike weights are defined by $W_{a_k} = \dfrac{f(y \mid \theta_k)}{\sum_{i=1}^{K} f(y \mid \theta_i)}$, and the adjusted weights by

$$W_{al_k} = \frac{p(M_k \mid AIC) f(y \mid \theta_k)}{\sum_{i=1}^{K} p(M_i \mid AIC) f(y \mid \theta_i)}.$$

Numerical computations of the properties for these estimators are for $n = 41$, $K = 30$, models are between 0.1 and 0.9 and are given in **Figure 5**. One can see that Akaike weights are not better than PMSEs for certain

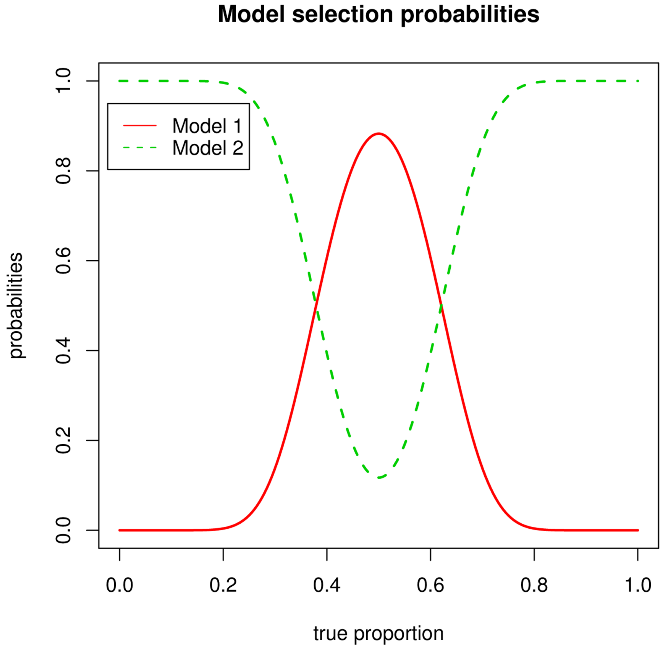

Figure 3. Model selection probabilities as a function θ.

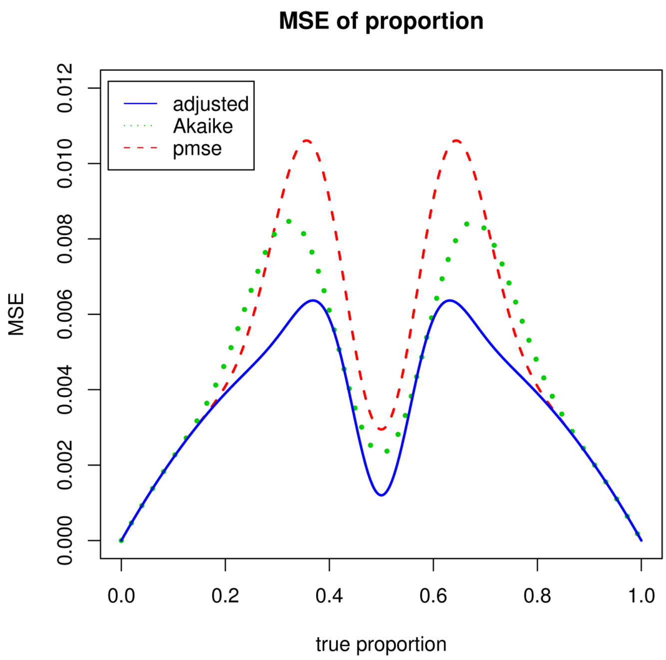

Figure 4. Risk of two proportions comparing PMSEs, Akaike weights estimators and adjusted estimators as a function of θ.

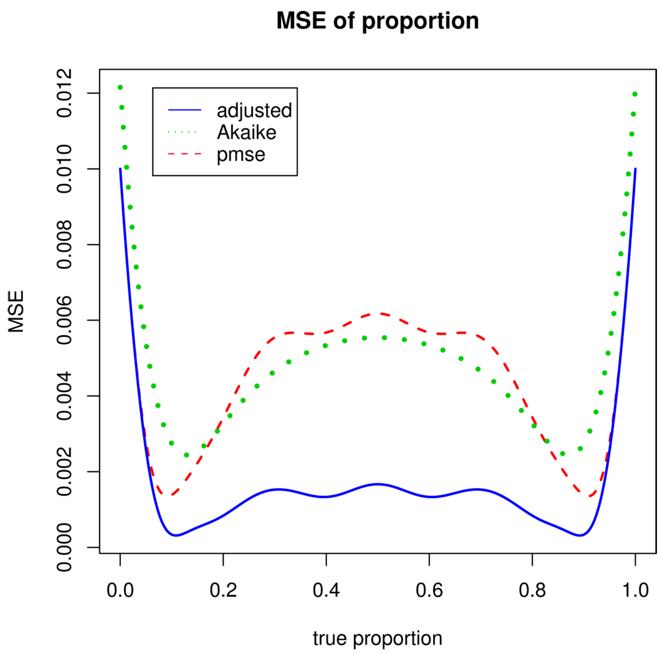

Figure 5. Risk of 30 models comparing PMSEs, Akaike weights estimators and adjusted estimators as a function of θ.

regions of the parameter space, but the adjusted likelihood weights are better than both.

5. Concluding Remarks

In this paper, we have considered model averaging in frequentist perspective; and proposed an approach of assigning weights to competing models taking account model selection probability and likelihood. The method appears to perform well for Bernoulli trials. The method needs to be applied in variety of situations before it can be adopted.

Acknowledgements

We Thank the Editor and the referee for their comments on earlier versions of this paper.

References

[1] Berk, R., Brown, L. and Zhao, L. (2010) Statistical Inference after Model Selection. *Journal of Quantitative Criminology*, **26**, 217-236. http://dx.doi.org/10.1007/s10940-009-9077-7

[2] Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, I. (2013) Valid Post-Selection Inference. *Annals of Statistics*, **41**, 802-837. http://dx.doi.org/10.1214/12-aos1077

[3] Belloni, A., Chernozhukov, V. and Kato, K. (2015) Uniform Post-Selection Inference for Least Absolute Deviation Regression and Other Z-Estimation Problems. *Biometrika*, **102**, 77-94. http://dx.doi.org/10.1093/biomet/asu056

[4] Belloni, A., Chernozhukov, V. and Wei, Y. (2016) Post-Selection Inference for Generalized Linear Models with Many Controls. *Journal of Business and Economic Statistics*. http://dx.doi.org/10.1080/07350015.2016.1166116

[5] Tibshirani, R.J., Taylor, J., Lockhart, R. and Tibshirani, R. (2014) Exact Post-Selection Inference for Sequential Regression Procedures. arXiv:1401.3889.

[6] Chernozhukov, V., Hansen, C. and Spindler, M. (2015) Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach. *Annual Review of Economics*, **7**, 649-688. http://dx.doi.org/10.1146/annurev-economics-012315-015826

[7] Zucchini, W. (2000) An Introduction to Model Selection *Journal of Mathematical Psychology*, **44**, 41-61. http://dx.doi.org/10.1006/jmps.1999.1276

[8] Zucchini, W., Claeskens, G. and Nguefack-Tsague, G. (2011) Model Selection. *International Encyclopedia of Statistical Science*, Springer, Berlin Heidelberg, 830-833. http://dx.doi.org/10.1007/978-3-642-04898-2_373

[9] Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-465. http://dx.doi.org/10.1214/aos/1176344136

[10] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999) Bayesian Model Averaging: A Tutorial (with Discussions). *Statistical Science*, **14**, 382-417.

[11] Nguefack-Tsague, G. (2011) Using Bayesian Networks to Model Hierarchical Relationships in Epidemiological Studies. *Epidemiology and Health*, **33**, e201100633. http://dx.doi.org/10.4178/epih/e2011006

[12] Nguefack-Tsague, G. (2013) Bayesian Estimation of a Multivariate Mean under Model Uncertainty. *International Journal of Mathematics and Statistics*, **13**, 83-92.

[13] Nguefack-Tsague, G. and Ingo, B. (2014) A Focused Bayesian Information Criterion. *Advances in Statistics*, **2014**, Article ID: 504325. http://dx.doi.org/10.1155/2014/504325

[14] Nguefack-Tsague, G. and Zucchini W. (2016) A Mixture-Based Bayesian Model Averaging Method. *Open Journal of Statistics*, **6**, 220-228. http://dx.doi.org/10.4236/ojs.2016.62019

[15] Nguefack-Tsague, G. and Zucchini, W. (2016) Effects of Bayesian Model Selection on Frequentist Performances: An Alternative Approach. *Applied Mathematics*, **7**, 1103-1105. http://dx.doi.org/10.4236/am.2016.710098

[16] Hansen, B.E. (2007) Least Squares Model Averaging. *Econometrica*, **75**, 1175-1189. http://dx.doi.org/10.1111/j.1468-0262.2007.00785.x

[17] Hansen, B.E. (2008) Least-Squares Forecast Averaging. *Journal of Econometrics*, **146**, 342-350. http://dx.doi.org/10.1016/j.jeconom.2008.08.022

[18] Hansen, B.E. (2009) Averaging Estimators for Regressions with a Possible Structural Break. *Econometric Theory*, **25**, 1498-1514. http://dx.doi.org/10.1017/S0266466609990235

[19] Hansen, B.E. (2010) Averaging Estimators for Autoregressions with a Near Unit Root. *Journal of Econometrics*, **158**, 142-155. http://dx.doi.org/10.1016/j.jeconom.2010.03.022

[20] Hansen, B.E. (2014) Model Averaging, Asymptotic risk, and Regressor Groups. *Quantitative Economics*, **5**, 495-530. http://dx.doi.org/10.3982/QE332

[21] Hansen, B.E. (2014) Nonparametric Sieve Regression: Least Squares, Averaging Least Squares, and Cross-Validation. In: Racine, J., Su, L.J. and Ullah, A., Eds., *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press, Oxford, 215-248.

[22] Hansen, B.E. and Racine, J.S. (2012) Jackknife Model Averaging. *Journal of Econometrics*, **167**, 38-46.

http://dx.doi.org/10.1016/j.jeconom.2011.06.019

[23] Cheng, X. and Hansen, B.E. (2015) Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach. *Journal of Econometrics*, **186**, 280-293. http://dx.doi.org/10.1016/j.jeconom.2015.02.010

[24] Charkhi, A., Claeskens, G. and Hansen, B.E. (2016) Minimum Mean Squared Error Model Averaging in Likelihood Models. *Statistica Sinica*, **26**, 809-840. http://dx.doi.org/10.5705/ss.202014.0067

[25] Wan, A.T.K., Zhang X. and Zou, G. (2010) Least Squares Model Averaging by Mallows Criterion. *Journal of Econometrics*, **156**, 277-283. http://dx.doi.org/10.1016/j.jeconom.2009.10.030

[26] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. 2*nd International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267-281.

[27] Burnham, K.P. and Anderson, D.R. (2013) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, Cambridge.

[28] Nguefack-Tsague, G. and Zucchini, W. (2011) Post-Model Selection Inference and Model Averaging. *Pakistan Journal of Statistics and Operation Research*, **7**, 347-361. http://dx.doi.org/10.18187/pjsor.v7i2-Sp.292

[29] Nguefack-Tsague, G. (2013) On Bootstrap and Post-Model Selection Inference. *International Journal of Mathematics and Computation*, **21**, 51-64.

[30] Nguefack-Tsague, G. (2013) An Alternative Derivation of Some Commons Distributions Functions: A Post-Model Selection Approach. *International Journal of Applied Mathematics and Statistics*, **42**, 138-147.

[31] Nguefack-Tsague, G. (2014) Estimation of a Multivariate Mean under Model Selection Uncertainty. *Pakistan Journal of Statistics and Operation Research*, **10**, 131-145. http://dx.doi.org/10.18187/pjsor.v10i1.449

[32] Nguefack-Tsague, G. (2014) On Optimal Weighting Scheme in Model Averaging. *American Journal of Applied Mathematics and Statistics*, **2**, 150-156. http://dx.doi.org/10.12691/ajams-2-3-9

[33] R Development Core Team (2016) R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna.

[34] Schomaker, M. and Heumann, C. (2014) Model Selection and Model Averaging after Multiple Imputation. *Computational Statistics and Data Analysis*, **77**, 758-770. http://dx.doi.org/10.1016/j.csda.2013.02.017

[35] Schomaker, M. (2012) Shrinkage Averaging Estimation. *Statistical Papers*, **53**, 1015-1034. http://dx.doi.org/10.1007/s00362-011-0405-2

[36] Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model Selection: An Integral Part of Inference. *Biometrics*, **53**, 603-618. http://dx.doi.org/10.2307/2533961