

## Prediction error criterion for selecting variables in a linear regression model

Yasunori Fujikoshi · Tamio Kan ·  
Shin Takahashi · Tetsuro Sakurai

Received: 12 September 2007 / Revised: 1 September 2008 / Published online: 30 April 2009  
© The Institute of Statistical Mathematics, Tokyo 2009

**Abstract** Several criteria, such as CV,  $C_p$ , AIC, CAIC, and MAIC, are used for selecting variables in linear regression models. It might be noted that  $C_p$  has been proposed as an estimator of the expected standardized prediction error, although the target risk function of CV might be regarded as the expected prediction error  $R_{PE}$ . On the other hand, the target risk function of AIC, CAIC, and MAIC is the expected log-predictive likelihood. In this paper, we propose a prediction error criterion, PE, which is an estimator of the expected prediction error  $R_{PE}$ . Consequently, it is also a competitor of CV. Results of this study show that PE is an unbiased estimator when the true model is contained in the full model. The property is shown without the assumption of normality. In fact, PE is demonstrated as more faithful for its risk function than CV. The prediction error criterion PE is extended to the multivariate case. Furthermore, using simulations, we examine some peculiarities of all these criteria.

**Keywords** Prediction error criterion · Linear regression models · Selection of variables · Risk function · Selection criteria