A g-prior extension for $p > n$

Yuzo Maruyama* Edward, I. George†

Abstract

For the normal linear model regression setup, Zellner's g-prior is extended for the case where the number of predictors p exceeds the number of observations n. Exact analytical calculation of the marginal density under this prior is seen to lead to a new closed form variable selection criterion. This results are also applicable to the multivariate regression setup.
AMS subject classification: Primary 62F07, 62F15, secondary 62C10, 62J07
Keywords: Bayesian variable selection, Bayesian model selection, Bayes factor, ridge regression, singular value decomposition, model selection consistency, full Bayes method, grouping effect

1 Introduction

Suppose the normal linear regression model is used to relate Y to the potential predictors $X_1, \ldots, X_p$,

$$Y \sim N_n(\alpha 1_n + X\beta, \sigma^2 I) \tag{1}$$

where α is an intercept parameter, 1_n is an $n \times 1$ vector each component of which is one, $X = (X_1, \ldots, X_p)$ is an $n \times p$ design matrix, β is a $p \times 1$ vector of unknown regression coefficients, and σ^2 is an unknown positive scalar. We assume that rank $X = \min(n, p)$ and also that X is in advance standardized so that $(X_i - \bar{X}_i 1_n)'(X_i - \bar{X}_i 1_n)$ for $1 \le i \le p$ is equal to each other. The variable selection problem is important because there is usually

*University of Tokyo e-mail: maruyama@csis.u-tokyo.ac.jp
†University of Pennsylvania e-mail: edgeorge@wharton.upenn.edu