A Graphical Approach for Evaluating Experimental Designs under Generalized Linear Models

Dae-Heung Jang

Pukyong National University, Division of Mathematical Sciences

599-1 Daeyeon-dong, Nam-gu

Pusan, Republic of Korea

dhjang@pknu.ac.kr

1. Generalized Linear Models

The generalized linear models provide a way to estimate the link function of the mean response as a linear function of the values of some set of predictors. This is written as:

$$g(E[y(\mathbf{x})]) = g(\mu(\mathbf{x})) = \eta(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta} \tag{1}$$

where g is the link function , $\mathbf{f}'(\mathbf{x})$ is a vector of order $p\times 1$ whose elements consist of powers and products of powers of $\mathbf{x}$ up to degree $d(\geq 1)$, and $\boldsymbol{\beta}$ is a vector of p unknown parameters. The mean response at a point $\mathbf{x}$ in a region of interest R is formulated as a function of the input variables of the form $\mu(\mathbf{x}) = h[\eta(\mathbf{x})] = g^{-1}[\eta(\mathbf{x})]$.

Let $\mathbf{x}_i$ be a vector of input variables associated with the i-th response value, $y_i, i = 1, 2, ..., n$. The distributions in the generalized linear models belong to the exponential family of distributions having a density as following:

$$f(y_i : \theta_i, \phi) = exp[\phi\{y_i\theta_i - b(\theta_i) + c(y_i)\} + d(y_i, \phi)], i = 1, 2, ..., n \tag{2}$$

where θ_i is the canonical parameter and ϕ is the dispersion parameter. We know that $E(y_i) = \mu_i = \frac{db(\theta_i)}{d\theta_i}$ and $Var(y_i) = \frac{d\mu_i}{\phi d\theta_i} = \frac{d^2b(\theta_i)}{\phi d\theta_i^2}$.

An approximate prediction variance of $\hat{\mu}(\mathbf{x})$ is given by

$$Var[\hat{\mu}(\mathbf{x})] \simeq \frac{1}{\phi}(\frac{d\mu}{d\eta})^2 \mathbf{f}'(\mathbf{x})[X'WX]^{-1}\mathbf{f}(\mathbf{x}) \tag{3}$$

where X is a matrix whose i-th row is $\mathbf{f}'(\mathbf{x_i})$ and $W = Diag(w_1, w_2, ..., w_n)$, where $w_i = \frac{\phi(\frac{d\mu_i}{d\eta_i})^2}{\frac{d\mu_i}{d\theta_i}}$ with $\frac{d\mu_i}{d\eta_i}$ denoting the derivative of μ with respect to η evaluated at $\mathbf{x}_i$. The dispersion parameter ϕ is factored out in expression (3).

2. A Method for Evaluating Experimental Designs under Generalized Linear Models

Giovannitti-Jensen and Myers(1989) suggested variance dispersion graph. The variance dispersion graph is a plot which consist of the maxima, minima, and average of the prediction variance on concentric spheres contained inside an the region of interest R. Khuri, Kim, and Um(1996) proposed quantile plot which is a plot of quantiles of the prediction variance on concentric spheres contained inside an the region of interest R. Khuri and Lee(1998) suggested quantile dispersion plot of estimated scaled mean-squared error of prediction for nonlinear models. Robinson(2000) proposed quantile dispersion plot of estimated scaled mean-squared error of prediction for logistic models.

We can extend Robinson's idea to all members in exponential family. We can find the corresponding form of $Var[\hat{\mu}(\mathbf{x})]$ with respect to each distribution in the exponential family. Consider several concentric surfaces R_λ which are located within R and are obtained by reducing the boundary of R using a shrinkage factor λ. For a given design D and $\boldsymbol{\beta} \in B$, let $Q_D(p, \boldsymbol{\beta}, \lambda)$ denote the p-th quantile of the exponential family distribution of values of $Var[\hat{\mu}(\mathbf{x})]$ on R_λ. B is the parameter space of $\boldsymbol{\beta}$, which must be considered in order to assess the dependency of the prediction variance on the unknown parameters. Let $G_D^{max}(p, \lambda) = \max_{\boldsymbol{\beta} \in B} Q_D(p, \boldsymbol{\beta}, \lambda)$, $G_D^{min}(p, \lambda) = \min_{\boldsymbol{\beta} \in B} Q_D(p, \boldsymbol{\beta}, \lambda)$. Plots of $G_D^{max}(p, \lambda)$ and $G_D^{min}(p, \lambda)$ against p result in quantile dispersion graphs of the prediction variance. This quantile dispersion graphs can be applied to the mean-squared error of prediction.

REFERENCES

Giovanniti-Jensen, A. and Myers, R. H.(1989). Graphical assessment of the prediction capacity of response surface designs. *Technometrics*, **31**, 159-171.

Khuri, A. I., Kim, H., and Um, Y.(1996). Quantile plots of the prediction variance for response surface designs. *Computational Statistics and Data Analysis*, **22**, 395-407.

Khuri, A. I. and Lee, J.(1998). A graphical approach for evaluating and comparing designs for nonlinear models. *Computational Statistics and Data Analysis*, **27**, 433-443.

Robinson, K. S.(2000). Quantile dispersion graphs for design comparisons for logistic models and other modeling issues. *PhD Dissertation.*

RESUME

Professor, Division of Mathematical Sciences.