QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2005; **21**:669–675
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/qre.707

Special Issue

A Large-sample Confidence Band for a Multi-response Ridge Path

Rui Ding[1], Dennis K. J. Lin[2,*,†] and John J. Peterson[3]
[1]*Amgen, 40 Technology Way, ARI-7-2A, West Greenwich, RI 02817, U.S.A.*
[2]*Department of Supply Chain and Information Systems, The Pennsylvania State University, University Park, PA 16802, U.S.A.*
[3]*GlaxoSmithKline Pharmaceuticals, R&D, 709 Swedeland Road (UW218A), King of Prussia, PA 19406, U.S.A.*

Ridge analysis in response surface methodology has received extensive discussion in the literature, while little is known for ridge analysis in the multi-response case. In this paper, the ridge path is investigated for multi-response surfaces and a large-sample simultaneous confidence interval (confidence band) for the ridge path is developed. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: confidence interval; constrained optimization; desirability function; response surface methodology

INTRODUCTION

An important tool in response surface methodology (RSM) is the ridge analysis, originally introduced by Hoerl[1] and Draper[2]. See also Hoerl[3] for a nice review. Suppose the true response function is $f(\mathbf{x}, \boldsymbol{\theta})$, where $\mathbf{x}$ is the vector of input variables and $\boldsymbol{\theta}$ is the vector of model parameters, and suppose we are interested in maximizing the response. Let $g(\boldsymbol{\theta}, r) = \max_{\mathbf{x}'\mathbf{x}=r^2} f(\mathbf{x}, \boldsymbol{\theta})$ be the constrained maximal mean response, where r is the distance from the center of the experiment region. A *ridge path* is defined as the plot of $g(\boldsymbol{\theta}, r)$ versus r. This is typically used to locate the optimal operating conditions.

In practice, we usually do not know the value of model parameters $\boldsymbol{\theta}$ in advance and we have to use the estimated value $\hat{\boldsymbol{\theta}}$. The plot of $g(\hat{\boldsymbol{\theta}}, r)$ versus r is not the true ridge path, merely an estimator. It is thus important to find the confidence interval associated with the relevant estimator, to address the issue of the accuracy of the estimation. Carter *et al.*[4] and Peterson[5] discussed such an issue for the single-response ridge analysis, but little is known for the multiple-response case. Experimenters, however, often face the simultaneous optimization of several response variables. In this paper, we construct an estimated ridge path for multiple response surfaces and develop the corresponding confidence band.

This paper is organized as follows. We first define the ridge path in the case of multi-responses, based on a desirability function approach. The confidence interval and confidence band about the ridge path are then developed. Next, a real-life example is used for illustration and the empirical coverage rate of the confidence band is validated via simulation. The conclusion and discussion are presented at the end. For simplicity of the presentation, the theoretical derivation is given in Appendix A.

*Correspondence to: Dennis K. J. Lin, Department of Supply Chain and Information Systems, The Pennsylvania State University, U.S.A.
†E-mail: dennislin@psu.edu

Received 18 April 2004
Revised 23 August 2004

RIDGE PATH WITH CONFIDENCE INTERVALS FOR MULTI-RESPONSE SURFACES

Suppose there are m response variables with the ith true response function being

$$E(Y_i) = f(\mathbf{x}, \boldsymbol{\theta}_i) \tag{1}$$

for $i = 1, \ldots, m$, where Y_i is the ith response, and $\mathbf{x}$ and $\boldsymbol{\theta}_i$ are the vectors of input variables and model parameters. Here, each $f(x, \theta_i)$ is assumed to be a linear function of θ_i. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m)'$ be the vector which includes all of the model parameters. One popular method in multiple response optimization is the desirability function approach suggested by Harrington[6] and later modified by Derringer and Suich[7], and Del Castillo *et al.*[8].

The basic idea is to use a group of functions $d_i(\cdot)$, called desirability functions, to transform the expected responses into values in [0, 1], showing how 'desirable' the optimization results are, the larger the better. The overall desirability function can thus be defined as the geometric mean

$$D(\mathbf{x}, \boldsymbol{\theta}) = \left(\prod_{i=1}^{m} d_i(E(y_i))\right)^{1/m} \tag{2}$$

A natural definition for a multi-response ridge path is thus the plot of $g(\theta, r)$ versus radius r, where

$$g(\boldsymbol{\theta}, r) = \max_{\mathbf{x}'\mathbf{x}=r^2} D(\mathbf{x}, \boldsymbol{\theta}) \tag{3}$$

An estimator of the ridge path is the plot of $g(\hat{\boldsymbol{\theta}}, r)$ versus r, where the estimator of the parameters $\hat{\boldsymbol{\theta}}$ is obtained by fitting multivariate linear regression models.

After taking a logistic transformation, as shown in Appendix A, we can construct a $100(1-\alpha)\%$ asymptotic confidence interval of $\text{logit}(g(\boldsymbol{\theta}, r))$ by the *Delta* method (e.g. Serfling[9] (p. 118)) in the form of

$$[L, U] = \text{logit}(g(\hat{\boldsymbol{\theta}}, r)) \pm z_{\alpha/2} \cdot \hat{c}(r) \tag{4}$$

where $\hat{c}(r)$ is the estimated standard error of $\text{logit}(g(\hat{\boldsymbol{\theta}}, r))$ (see Appendix A for the formula) and $z_{\alpha/2}$ is the upper $\alpha/2$ critical value of $\mathbf{N}(0, 1)$. Thus a $100(1-\alpha)\%$ large-sample confidence interval for $g(\boldsymbol{\theta}, r)$, at each fixed r, may take the form of

$$\left[\frac{\exp^L}{1+\exp^L}, \frac{\exp^U}{1+\exp^U}\right] \tag{5}$$

Since our ridge path plot involves all of the different radii, we need to construct a $100(1-\alpha)\%$ simultaneous confidence band for $g(\boldsymbol{\theta}, r)$ or $\text{logit}(g(\boldsymbol{\theta}, r))$. Namely,

$$\mathbf{P}\left(\frac{|\text{logit}(g(\hat{\boldsymbol{\theta}}, r)) - \text{logit}(g(\boldsymbol{\theta}, r))|}{\hat{c}(r)} \le \rho_\alpha, \forall r > 0\right) \doteq 1 - \alpha \tag{6}$$

where ρ_α is some critical value (e.g. Miller[10] (p. 49)). Note that if $\rho_\alpha = z_{\alpha/2}$, the above equation will be equivalent to (4).

If the experimenter is only interested in several (say q, instead of infinitely many) discrete values of r, the critical value of $z_{\alpha/2}$ can be replaced by $z_{\alpha/2q}$ in (4), based on *Bonferroni's* adjustment (e.g. Miller[10] (p. 8)). For the case when r can take any values in an interval, Scheffé[11] provided a simultaneous confidence band for a linear response surface using the critical value $(pF^{\alpha}_{p,n-p})^{1/2}$, where n is the number of observations and p is the number of model parameters. Such a critical value can be very conservative. Peterson's[5] simulations showed that $(2F^{\alpha}_{2,n-p})^{1/2}$ can be a good approximate critical value for a confidence band about a ridge trace

Table I. Specified values used in the SOVRING example

Response	Objective	$y_i^{\min}$	$y_i^{\max}$	a_i	b_i	γ_i
y_1	Maximize	200	600	400	54.592	0.025
y_2	Maximize	400	1200	800	109.183	0.025

for a mean response surface. Peterson[12] stated that, heuristically, this should be the case for a ridge trace and associated x-coordinate plot having shallow curvature. This may also be true when the response curvature is not too significant for a nonlinear response surface. The empirical simultaneous coverage rates using these two critical values, $z_{\alpha/2q}$ and $(\chi^2_{2,\alpha})^{1/2}$, as the limit of $(2F^{\alpha}_{2,n-p})^{1/2}$, can be obtained by simulations.

SOVRING EXAMPLE

Eriksson *et al.*[13] (p. 335) described the following SOVRING example. An experiment was conducted to test the effect of three input variables, raw material input Tonin (x_1), magnetic grinder speed variables HS1 (x_2) and HS2 (x_3), on the two response variables of product output in a mine, PAR (y_1) and FAR (y_2). Both responses are larger-the-better. The desirability functions proposed by Gibb *et al.*[14] is used here for illustration:

$$d_i = \left[1 + \exp\left(-\frac{E(y_i) - a_i}{b_i}\right)\right]^{-1}$$

where $a_i = (y_i^{\min} + y_i^{\max})/2$ and

$$b_i = \frac{(y_i^{\max} - y_i^{\min})}{2\ln((1-\gamma_i)/\gamma_i)}, \quad y_i^{\min} < y_i^{\max}, \quad \gamma_i \in (0, 1)$$

Specific values involved are listed in Table I.

An obvious outlier (the original 62nd observation) has been deleted from the original data set, leading to the final sample size of 229. The Web site http://www.smeal.psu.edu/faculty/dkl5/SOVRING.xls has coded data available.

Assuming a quadratic model for both responses, the multivariate regression analysis results in the fitted models

$$\begin{aligned}\widehat{y_1} &= 283.6516 + 116.1375x_1 + 6.8199x_2 + 22.4611x_3 + 13.7544x_1x_2 + 31.0183x_1x_3 \\ &\quad + 18.6729x_2x_3 + 24.8719x_1^2 - 7.5994x_2^2 + 10.2035x_3^2 \\ \widehat{y_2} &= 688.8352 + 244.0213x_1 + 29.3419x_2 - 47.4965x_3 + 1.1136x_1x_2 - 62.1298x_1x_3 \\ &\quad - 56.7020x_2x_3 + 6.8384x_1^2 - 9.7295x_2^2 - 35.8798x_3^2\end{aligned}$$

and the estimated covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 195.13 & 101.91 \\ 101.91 & 1063.4 \end{pmatrix}$$

Note that all input variables (x_i) have been coded between -1 and 1.

A confidence band of the ridge path can be constructed via (5). We did this in three ways: using the critical value of $z_{\alpha/2}$ (for each individual r), $z_{\alpha/20}$ (for 10 different r) and $(\chi^2_{2,\alpha})^{1/2}$ (for infinitely many r), respectively. Figure 1(a) is the ridge path plot with individual 95% confidence intervals and two 95% simultaneous confidence bands constructed using the critical values of $(\chi^2_{2,\alpha})^{1/2}$ and $z_{\alpha/20}$. The ridge path shows how the predicted maximal desirability changes with the radius r. Although the differences among all the confidence bands are rather limited, it is clear that the Bonferroni $z_{\alpha/20}$ band is wider than the $(\chi^2_{2,\alpha})^{1/2}$ band and the unadjusted $z_{\alpha/2}$ band.

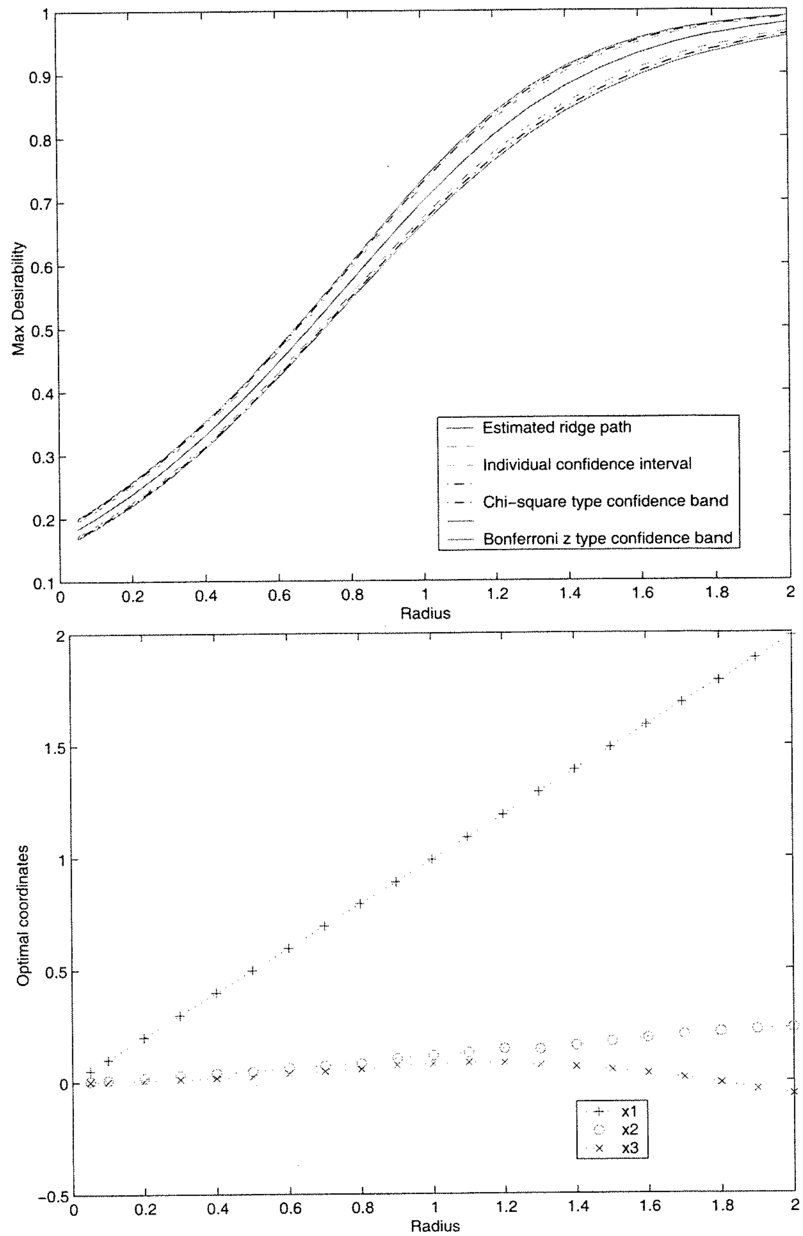

Figure 1. The SOVRING example: (a) ridge path, individual 95% confidence interval and confidence band; (b) constrained optimal coordinate plot

For any specific radius r, the estimated optimal design level can be identified from Figure 1(b). Figure 1(b) shows that the optimal level of x_1 changes dramatically with radius r, while x_2 and x_3 are less variable. This implies that the production outputs (y_1 and y_2) depend heavily on the amount of raw material (x_1). If the speeds of the two grinder machines (x_2 and x_3) are calibrated to generate maximal output from small amounts of raw material, the same calibration should be kept for larger amount of raw material. The graph also shows that when x_1 is large, the optimal level of x_3 goes down, which suggests that the speed of the second grinder should be reduced. Note that all x_i are between -1 and 1, thus the inference for radius r between $\sqrt{3}$ and 2 is considered to be an extrapolation.

To verify the coverage rate of the constructed confidence bands, 1000 data sets are generated based on the estimated models above. The empirical coverage rate is then obtained by checking whether the true maximal desirability values are simultaneously inside the confidence intervals at 10 different r, ranging from 0 to 2. For a nominal 95% confidence band, the empirical coverage rate is 95.2% for the χ^2-type confidence band and 98.4% for the Bonferroni-type confidence band. Both of these are reasonably close to the nominal rate, while the Bonferroni-type confidence band is somewhat conservative, as expected.

CONCLUSIONS

The ridge analysis for the multi-response surfaces is investigated in this paper. A large sample confidence band is developed. Simulation shows a close to nominal coverage rate for the confidence bands when sample size is relatively large. Although we used the desirability functions suggested by Gibb *et al.*[14], the method discussed here can be applied to other desirability functions (e.g. Kim and Lin[15,16]), provided that they are everywhere differentiable. Note that our confidence interval is derived by applying delta method to $\text{logit}(g(\boldsymbol{\theta}, r))$, not directly to $g(\hat{\boldsymbol{\theta}}, r)$. This is because such an approach can guarantee that the derived confidence interval is contained by the interval [0, 1], which is the range of the function $\exp^x/(1 + \exp^x)$. All the calculations and simulations given in this paper were done with Matlab and the computer program is available from the first author upon request.

In addition to the SOVRING example, we have also applied the proposed method to numerous examples: including, texture characteristics example (Khuri and Conlon[17] (Example 1)), whey protein example (Khuri and Conlon[17] (Example 2)), and tire tread example (Derringer and Suich[7]). The empirical coverage rates indicate that the proposed large-sample band may even work well for sample sizes as small as 20.

As pointed out by one referee

> '. . . *D can be defined in many different ways. The use of different D's will certainly result in different ridge paths for the same problem. . . . Perhaps, the (multiple) ridge paths for the individual responses might provide more direct and useful insights, especially to the engineers in practice*'.

We think that this is a very sensible comment. Derringer[18] stressed the importance of having an expert panel construct a desirability function to link various values of the overall desirability function to product quality. However, we also believe that if the model is correctly identified and sufficiently efficient, any reasonable measurement of the desirability D should lead to similar ridge paths. The ridge path for each individual (univariate) response can be a good reference indeed. How to compromise different ridge paths (from each response) remains a non-trivial issue to be addressed, however.

REFERENCES

1. Hoerl AE. Optimum solution of many variables equations. *Chemical Engineering Progress* 1959; **55**:69–78.
2. Draper NR. Ridge analysis of response surfaces. *Technometrics* 1963; **5**:469–479.
3. Hoerl RW. Ridge analysis 25 years later. *The American Statistician* 1985; **39**:186–192.
4. Carter WH, Chinchilli VM, Myers RH, Campebell ED. Confidence intervals and improved ridge analysis of response surface. *Technometrics* 1986; **28**:339–346.
5. Peterson JJ. A general approach to ridge analysis with confidence intervals. *Technometrics* 1993; **35**:204–214.
6. Harrington EC Jr. The desirability function. *Industrial Quality Control* 1965; **21**:494–498.
7. Derringer G, Suich R. Simultaneous optimization of several response variable. *Journal of Quality Technology* 1980; **12**:214–219.
8. DelCastillo E, Montgomery DC, McCarville DR. Modified desirability functions for multiple response optimization. *Journal of Quality Technology* 1996; **28**:337–345.
9. Serfling RJ. *Approximation Theorems of Mathematical Statistics*. Wiley: New York, 1980.
10. Miller RG. *Simultaneous Statistical Inference* (2nd edn). Springer: New York, 1981.
11. Scheffé H. *The Analysis of Variance*. Wiley: New York, 1959.
12. Peterson JJ. First and second order derivatives having applications to estimation of response surface optima. *Statistics and Probability Letters* 1989; **8**:29–34.

13. Eriksson I, Johansson E, Kettaneh-Wold N, Wold S. *Multi- and Megavariate Data Analysis: Principles and Applications*. Umetrics Academy, 2001.
14. Gibb RD, Carter WH, Myers RH. Incorporating experimental variability in the determination of desirable factor levels. *Unpublished Manuscript*, 2001.
15. Kim KJ, Lin DKJ. Optimizing dual response systems. *Journal of Quality Technology* 1988; **30**:1–10.
16. Kim KJ, Lin DKJ. Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions. *Applied Statistics* 2000; **49**:311–325.
17. Khuri AI, Conlon M. Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics* 1981; **23**:363–375.
18. Derringer G. A balancing act: Optimizing a product's properties. *Quality Progress* 1994; **24**:51–58.
19. Arnold SF. *The Theory of Linear Models and Multivariate Analysis*. Wiley: New York, 1981.

APPENDIX A. DERIVATION OF THE CONFIDENCE INTERVALS

By Lemma 2 in Peterson[5], $g(\boldsymbol{\theta}, r)$ is continuously differentiable in $\boldsymbol{\theta}$, as long as $D(\mathbf{x}, \boldsymbol{\theta})$ is continuously differentiable in $(\mathbf{x}, \boldsymbol{\theta})$ and

$$g_{\boldsymbol{\theta}}(\boldsymbol{\theta}, r) = D_{\boldsymbol{\theta}}(\mathbf{x}_0, \boldsymbol{\theta})$$

where $\mathbf{x}_0 = \mathbf{x}_0(\boldsymbol{\theta}, r)$ is assumed to be the unique optimal input value such that

$$D(\mathbf{x}_0, \boldsymbol{\theta}) = \max_{\mathbf{x}'\mathbf{x}=r^2} D(\mathbf{x}, \boldsymbol{\theta})$$

$D_{\boldsymbol{\theta}}(\mathbf{x}_0, \boldsymbol{\theta})$ is the gradient vector of $D(\mathbf{x}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, then evaluated at $(\mathbf{x}_0, \boldsymbol{\theta})$. Note that usually we do not know the explicit form of $\mathbf{x}_0 = \mathbf{x}_0(\boldsymbol{\theta}, r)$, but its value can by found by a nonlinear optimization routine or a grid search method, once $\boldsymbol{\theta}$ is known.

Next, we can estimate the parameters by fitting the standard multivariate regression model in the matrix normal form (see, e.g., Arnold[19] (p. 349)),

$$\mathbf{Y} \sim \mathbf{N}_{n,m}(\mathbf{X\Theta}, \mathbf{I}, \boldsymbol{\Sigma}) \tag{A1}$$

where n is the number of independent experiment runs, m is the number of response variables in each run with a fixed covariance matrix $\boldsymbol{\Sigma}$. The matrices $\mathbf{Y}$, $\mathbf{X}$ and $\boldsymbol{\Theta}$ are the observation matrix $(n \times m)$, design matrix $(n \times p)$ and parameter matrix $(p \times m)$, respectively. This is equivalent to

$$\mathrm{Vec}(\mathbf{Y}) \sim \mathbf{N}_{n,m}(\mathrm{Vec}(\mathbf{X\Theta}), \boldsymbol{\Sigma} \otimes \mathbf{I})$$

where $\otimes$ is the *Kronecker* product.

Provided that the Jurekova–Noether (JN) condition holds (namely $(1/n)\mathbf{X}'\mathbf{X} \xrightarrow{n\to\infty} \mathbf{A}$ where matrix $\mathbf{A}$ is positive definite), we have

$$\hat{\boldsymbol{\Theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \xrightarrow{p} \boldsymbol{\Theta}$$
$$\hat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Theta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Theta}})/(n-p) \xrightarrow{p} \boldsymbol{\Sigma}$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{N}_{p,m}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{A}^{-1})$$

where $\boldsymbol{\theta} = \mathrm{Vec}(\boldsymbol{\Theta})$, $\hat{\boldsymbol{\theta}} = \mathrm{Vec}(\hat{\boldsymbol{\Theta}})$. Since $g(\boldsymbol{\theta}, r)$ is continuously differentiable in $\boldsymbol{\theta}$, by the *Delta* method we have

$$\sqrt{n}(\mathrm{logit}(g(\hat{\boldsymbol{\theta}}, r)) - \mathrm{logit}(g(\boldsymbol{\theta}, r))) \xrightarrow{d} \mathbf{N}(0, d(r)^2)$$

where

$$d(r)^2 = g_{\theta}(\theta, r)'(\Sigma \otimes \mathbf{A}^{-1})g_{\theta}(\theta, r)/(D(\mathbf{x}_0, \theta)(1 - D(\mathbf{x}_0, \theta)))^2$$
$$= D_{\theta}(\mathbf{x}_0, \theta)'(\Sigma \otimes \mathbf{A}^{-1})D_{\theta}(\mathbf{x}_0, \theta)/(D(\mathbf{x}_0, \theta)(1 - D(\mathbf{x}_0, \theta)))^2$$

Define

$$\hat{c}(r)^2 = D_{\theta}(\widehat{\mathbf{x}_0}, \hat{\theta})'(\hat{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1})D_{\theta}(\widehat{\mathbf{x}_0}, \hat{\theta})/(D(\widehat{\mathbf{x}_0}, \hat{\theta})(1 - D(\widehat{\mathbf{x}_0}, \hat{\theta})))^2$$

where $\widehat{\mathbf{x}_0} = \mathbf{x}_0(\hat{\theta}, r)$, we have

$$n\hat{c}(r)^2 \xrightarrow{p} d(r)^2$$

since $g_{\theta}(\theta, r) = D_{\theta}(\mathbf{x}_0(\theta, r), \theta)$ is continuous in θ for each fixed r (Lemma 1 of Peterson[12]), the continuous function theorem can be applied here. Thus, by *Slutsky's* theorem,

$$\hat{c}(r)^{-1}(\text{logit}(g(\hat{\theta}, r)) - \text{logit}(g(\theta, r))) \xrightarrow{d} \mathbf{N}(0, 1)$$

So an approximate $100(1 - \alpha)\%$ confidence interval for $\text{logit}(g(\theta, r))$ is

$$\text{logit}(g(\hat{\theta}, r)) \pm z_{\alpha/2} \cdot \hat{c}(r)$$

Authors' biographies

Rui Ding is a statistician at Amgen. He received his PhD in Statistics from Pennsylvania State University in 2004 and BS from Tsinghua University (People's Republic of China) in 1998. His research interests include linear models, experiment design and various statistical issues in aseptic process monitoring and bioassay validation in the biotechnology industry.

Dennis K. J. Lin is a University Distinguished Professor of Supply Chain and Statistics at the Pennsylvania State University, University Park. He received a PhD in Statistics from the University of Wisconsin at Madison. He is an elected Fellow of the American Statistical Association, an elected member of the International Statistical Institute and a senior member of the American Society for Quality. He has published over 100 papers in the areas of design of experiments, response surface methodology, reliability, control charts, data mining and statistical inference.

John J. Peterson is Director, Statistics in the Statistical Sciences Department at GlaxoSmithKline Pharmaceuticals, R&D. He received his BS from S.U.N.Y. at Stony Brook (double major in Applied Mathematics and Computer Sciences) and his PhD in Statistics from The Pennsylvania State University. His research interests are statistical inference for response surface experiments and applications of response surface methods to the pharmaceutical industry. He has published several papers in this area, and is a co-author with Dennis Lin on a book chapter, 'Statistical Inference for Response Surface Optima', to appear in *Response Surface Methodology and Related Topics* published by World Scientific.