Computers & Industrial Engineering 113 (2017) 206–220

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Robust design under normal model departure

Chanseok Park [a,*], Linhan Ouyang [b], Jai-Hyun Byun [c], Mark Leeds [d]

[a] Department of Industrial Engineering, Pusan National University, Busan 46241, Republic of Korea
[b] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China
[c] Department of Industrial and Systems Engineering, Gyeongsang National University, Jinju, Gyeongnam 52828, Republic of Korea
[d] Statematics Consulting, 4 Sylvan Terrace, New York, NY 10032, USA

ARTICLE INFO

Article history:
Received 4 October 2016
Received in revised form 19 July 2017
Accepted 11 September 2017
Available online 15 September 2017

Keywords:
Quality control
Robust design
Response surface methodology
Monte Carlo method
Model departure
Relative efficiency

ABSTRACT

The basic underlying assumption in robust design is that the experimental data have a normal distribution. However, in many practical cases, the experimental data may actually have an underlying distribution that is not normal. The existence of model departure can have a significant effect on the optimal operating condition estimates of the control factors obtained in the robust design framework.

In this article, the effect of normal model departure on the optimal operating condition estimates is investigated and a methodology is constructed to deal with the effect of normal model departure. We provide simulation results which indicate that the sample mean and sample variance should not be used as estimators if one suspects that the underlying distribution of the sample is not normal. Extensive Monte Carlo simulations indicate that there exist attractive alternative estimators to the sample mean and sample variance. These estimators exhibit solid performance when the data are normally distributed and at the same time are quite insensitive to normal model departure.

1. Introduction

One of the most important design methodologies for quality improvement purposes is robust design. Taguchi's robust design methods often include orthogonal arrays and signal-to-noise (SN) ratios. Although Taguchi's methodology is a useful and convenient tool in quality engineering, some of his work has drawn criticism. For example, interaction effects are ignored and the SN ratio cannot distinguish between inputs affecting the process mean from those affecting the variance (Nair, 1992). Also, some of Taguchi's methodologies are intuitively sound but introduce statistical biases at the same time. For more details, see Box (1985, 1988), Vining and Myers (1990), Myers, Khuri, and Vining (1992), Tiao, Bisgaard, Hill, Peña, and Stigler (2000), Gauri and Pal (2014), Myers, Montgomery, and Anderson-Cook (2016) among others.

A great deal of work has been done to improve the various pitfalls of Taguchi's designs. Response surface methods for robust design were eventually developed as a more effective alternative to Taguchi's methods. Response surface methods are excellent tools for robust design and are particularly useful for process optimization. Often, in practical applications, the objective of a response surface approach is to select the levels of controllable variables in order to minimize the variability in a response variable, while keeping the mean of the response variable close to a target value. In other applications, the goal of response surface methods might be to minimize the variability while maximizing or minimizing the mean response. In order to satisfy dual goals such as these, the dual response surface approach was developed and popularized (Vining & Myers, 1990). We should point out that the dual response surface approach has been further studied by several authors, including Copeland and Nelson (1996), Kim and Lin (1998), Oyeyemi (2004), Lee and Park (2006), and Ouyang, Ma, Byun, Wang, and Tu (2016) among others.

Other studies focus on the notions of optimization formulation and multi-component objective functions. Ardakani and Wulff (2013) and Ardakani (2016) develop a Pareto optimal methodology in order to deal with the fact that the optimization formulation can often contain multiple components that conflict with one another and therefore need to be traded off. Rather than obtaining one optimal point in the optimization step, the Pareto optimal framework constructs an efficient frontier of optimal points. Each of the points on the frontier is optimal in the sense that, at each of the points on the frontier, one cannot improve on one component of the objective function without degrading it with respect to another component. In most cases, the two conflicting variables are the estimated mean response and estimated variance response.

* Corresponding author.
E-mail address: cpark2@gmail.com (C. Park).

http://dx.doi.org/10.1016/j.cie.2017.09.010
0360-8352/

There has also been a great deal of research on using more effective loss functions in the dual response framework. Jin, Liu, and Wang (2015) analyzed the optimal tolerance design problem based on asymmetric quadratic quality loss assuming non-normal distributions such as the triangular and trapezoidal distributions. Hazrati-Marangaloo and Shahriari (2017) introduce an improved quality loss function along with the incorporation of multivariate ANOVA concepts. Furthermore, Rathod, Yadav, Rathore, and Jain (2013) constructed a hybrid loss function that performs well even if the response distribution is skewed and does not have a normal distribution. Wan and Birch (2011) developed a robust regression technique in order to improve the quality of model estimation. They then applied this modelling technique to the process optimization problem. Lastly, there has been some research related to the design of experiments when there is non-normal response. Biswas, Das, and Mandal (2015) considered the robustness of designs. They investigate the robustness of the *F*-test for testing the equality of treatment effects when the response has a non-normal distribution. In addition, Woods, Overstall, Adamou, and Waite (2017) illustrated that generalized linear models can be a useful class of models in robust design when the response cannot be modelled using the classical linear regression model formulation.

The usual underlying assumption in the response surface methodology is that the error term in the response function model has a normal distribution. For example, see Sections 1.1.1 and 10.8.8 of Myers et al. (2016). This assumption then implies that the responses, which are the sample observations, are normally distributed with the mean response function as the mean (or location) parameter and the variance response function as the variance (or dispersion) parameter. However, the normality assumption in the response surface framework is often tenuous and may not hold in practice. It is well known that, under the normality assumption, the sample mean and sample variance are efficient and unbiased estimators of the location and dispersion parameters. However, a violation of the normality assumption may have a significant effect on the resulting estimated optimal operating conditions. Thus, a method used in the response surface methodology that is less sensitive to normal model departure could be quite useful in practical applications of robust design.

It is important to distinguish the model departure case from the similar situation where the data contain outliers and are possibly contaminated. In the case of outliers and possible contamination, there may be only one or more observations whose values seem extreme relative to the rest of the observations in the sample. It is unknown whether these unusual observations truly come from the sampled population or if they are due to some kind of error occurring during the data gathering process. The type of error causing the unusual value could range from a typographical error to an environmental condition (for example, wrong temperature setting). Regardless of the type of error, the contamination or outlier phenomenon refers to specific observations in the sample taking on unusual values relative to the majority of the observations in the sample. Conversely, model departure refers to the case where the underlying distribution of the population being sampled is different than that which is being assumed. Therefore, the model departure phenomenon refers to the nature of the sample itself. Despite these differences, the concepts of contamination and model departure are related and we review the relevant robust design research related to these issues as follows.

In the specific case where the data are subject to contamination, Park and Cho (2003) proposed a robust design methodology using outlier-resistant estimators of the mean and variance parameters. Specifically, the median was used to estimate the mean parameter (location) and the median absolute deviation (MAD) and interquartile range (IQR) were used to estimate the standard deviation parameter (scale). These estimators were then incorporated into the robust design model, and their behavior was shown to be efficient in the following sense. When the data were contaminated, the generalized variance of the estimated optimal operating conditions obtained when using the alternative estimators was significantly less than the resulting generalized variance based on the sample mean and sample variance. Recently, Park and Leeds (2016) extended their work by considering a larger set of alternative methods. They investigated the effectiveness of the various methods using simulations and provided conclusive evidence that there exist alternative methods. However, Park and Leeds (2016) considered the quality of the estimators only when outliers and the possibility of contamination was present. The issue of normal model departure was not investigated. Park and Cho (2003) addressed the issue of normal model departure but a much smaller set of alternative estimators were considered. In this paper, we extend the work of Park and Cho (2003) and Park and Leeds (2016) by investigating the performance of various methods in the case of normal model departure.

2. Robust design based on dual response surfaces and the proposed methodology

In what follows, we use the same assumptions, notations and framework as in Park (2013), so only a brief review is given here. We assume a system with a response Y which depends on the levels of the k control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. Thus, Y is a function of $\mathbf{x}$, that is, $Y = F(x_1, x_2, \ldots, x_k)$, where the functional structure of the function $F(\cdot)$ is assumed to be unknown. The levels of each control factor x_i for $i = 1, 2, \ldots, k$ are continuous and quantitative, and can be controlled by the experimenter. The process mean and variance response functions are assumed to have the dual quadratic response surface forms shown below.

Thus, the process mean function can be written as

$$M(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} \beta_{ij} x_i x_j + \epsilon_m \tag{1}$$

and the process variance function can be written as

$$V(\mathbf{x}) = \eta_0 + \sum_{i=1}^{k} \eta_i x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} \eta_{ij} x_i x_j + \epsilon_v. \tag{2}$$

Note that in the standard dual response framework, it is assumed that the error terms ϵ_m and ϵ_v in (1) and (2) are normally distributed with mean zero. Addressing this crucial assumption is the subject of this study. For more details on the normality assumption, one is referred to Sections 1.1.1 and 10.8.8 of Myers et al. (2016).

Let $\widehat{M}(\mathbf{x})$ and $\widehat{V}(\mathbf{x})$ represent the fitted dual response functions for the mean and variance respectively. Then the fitted dual response functions are straightforward:

$$\widehat{M}(\mathbf{x}) = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} \hat{\beta}_{ij} x_i x_j \tag{3}$$

and

$$\widehat{V}(\mathbf{x}) = \hat{\eta}_0 + \sum_{i=1}^{k} \hat{\eta}_i x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} \hat{\eta}_{ij} x_i x_j. \tag{4}$$

It is important to point that the resulting fitted response functions are quite useful because they allow for the calculation of the estimated mean and variance response associated with any new design point rather than only those used in the original experiment.

Given a candidate pair of location and scale estimators, we can use the second-order polynomial regression models to estimate