Determining Statistically Significant Parameter Regions for Bounded Specification

Arvind Rangarajan
Sponsored by CODEF

Abstract— Experiments are performed to observe the influence or quantify the effect a process choice has on the process outcome. Utilization of results from controlled experiments for production plan design using conventional data analysis techniques can lead to inefficient use of available information. The scheme described in this report helps overcome this pitfall by modifying the objective function used to compute the regression parameters. The report also presents a discussion on the various ways of applying this scheme to generate the envelope.

Keywords: process planning, precision manufacturing, surface finish, bounded specification, statistical.

1. Introduction

A specification in the manufacturing sense implies a measurable physical characterestic like surface finish or edge quality. Experiments are performed to characeterize these specifications. They are generally the response variable in a statistical experiment. There is prior knowledge on the nature of effect like linear, quadratic etc., with respect to the variables that affect the response. These are tunable parameters in the physical domain. The aim of experiments conducted in industry is to identify the range of parameters that would satisfy the designer's criteria. These are generally observed in the form of bounds on a response. Parameter estimation is predominantly limited to fitting the best surface. This provides enormous information in terms of identifying the basic underlying relationships [1,2]. Experiments performed for production planning is to ensure that a specification is below the designers choice of maxiumum, minmum or specified bounds. It can be clearly seen that such an approach, even though provides information, does not directly address the issue at hand. The objective of this paper is to estimate the regression parameters with the view that the results would be tested against a bound either upper or lower. The first section defines the choice of objective function and its utility for this objective. The second section describes the methodology for estimating the parameter. The scheme is applied to a standard dataset from Rpackage to demonstrate its utility. Finally, the asymptotic properties of this response surface is discussed to gain better insight into the data fitting scheme. Testing the envelopes for the family of estimated parameters is discussed.

2. Objective Function

The least squares estimation procedure is actually minimizing the residual sum square error in the regression. The objective function is symmetric about the errors, i.e. the positive and negative errors are weighted evenly, Figure 1. The estimate fits the line so that sum of errors is zero. In contrast, the new objective function is designed to weight the errors depending on the bound that is specified. Upper bound specification is used an example to demonstrate the concept. If the objective is to maintain the response below a value, the fit surface should be erring on the negative side or the positive errors in the objective function should be weighted less than the negative error promoting positive error by the design of objective function.

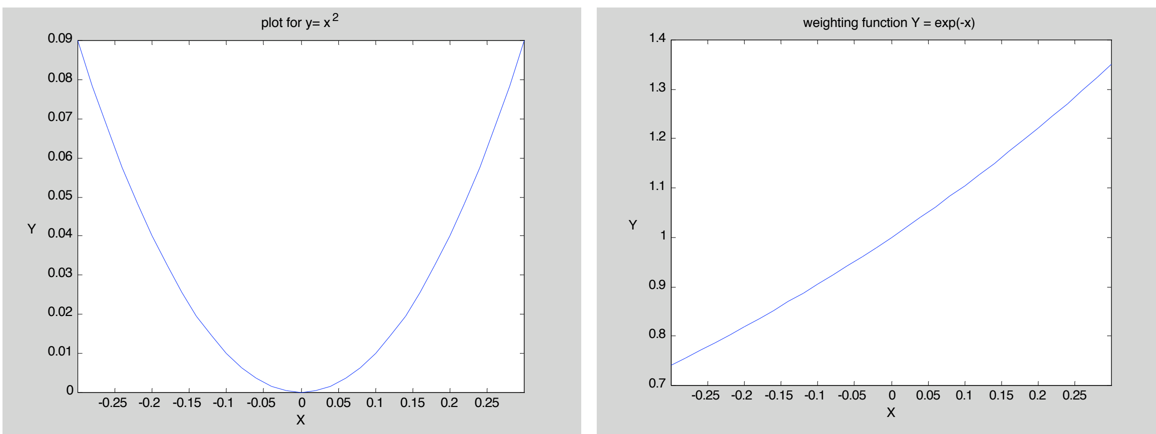

Figure 1. Plots of component functions in the objective.

After investigating various functions *exp(-ax)*, where *a* is a constant, is the best solution as it has a minimum at zero, implying the fact that a perfect fit if exists will not be lost.The beauty of this function is that it monotically decreases from $+\infty \, to -\infty$. Behavior of the newly created objective function can be demonstrated by observing the functions x^2, x^2e^{-x}, x^2e^{-2x}, x^2e^{-5x}. In Figure 2, *a* refers to the coefficient of the exponential function. LSE can be viewed as the case when $a=0$. The assymmetry implies that for any value of *x* various exponentially weighted curves are above the RSE on the positive side and lies below the curve for negative values of *x*.

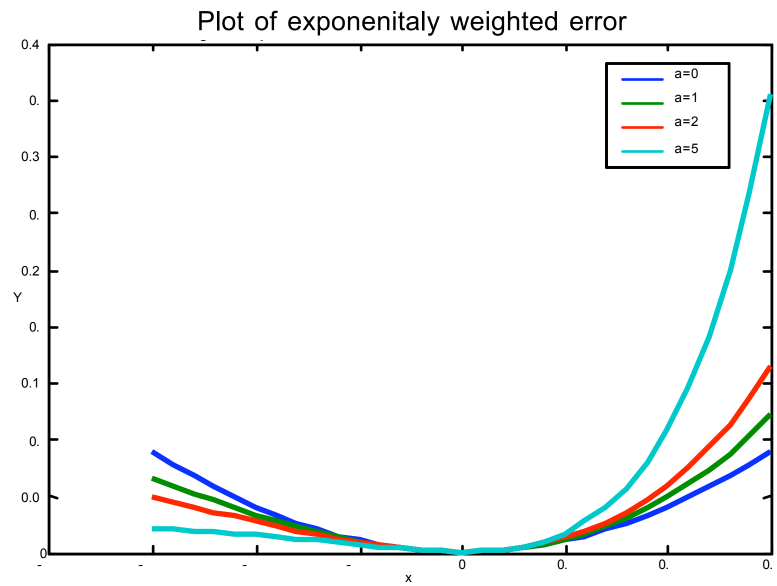

Figure 2. Objective functions utilized by the scheme.

3. Coefficient Estimation

Method for estimating the coefficients β_i's:

$$\hat{Y} = \beta_0 + \sum_j \beta j X j$$

$$\varepsilon_i = Y_i - \hat{Y}_i$$

$$Objective(obj) = \sum \varepsilon_i^2 e^{a\varepsilon_i}$$

Minimizing this objective function for constant values of a with respect to beta we get the following set of equations:

Equation for β_o:

$$\frac{\partial(obj)}{\partial \beta_0} = \sum_i e^{a\varepsilon_i}(a\varepsilon_i^2 - 2\varepsilon_i) = 0$$

Equation for the i^{th} coefficient where i != 0:

$$\frac{\partial(obj)}{\partial \beta_i} = \sum_i X_i e^{a\varepsilon_i}(a\varepsilon_i^2 - 2\varepsilon_i) = 0$$

Solving this set of equations for various simple data sets, shown in Table 1, for various values of '*a*' we obtain a series of solutions with the solution moving from least fit towards the upper bound, Figure 3. If the relationship between reponse and control variables is in product form, the coefficients on the powers of products can be determined by taking log and solving the problem in the previously described fashion.

Table 1. Sample dataset used for estimating upperbound coefficients.

X	Y input
5	12
7	15
8	16

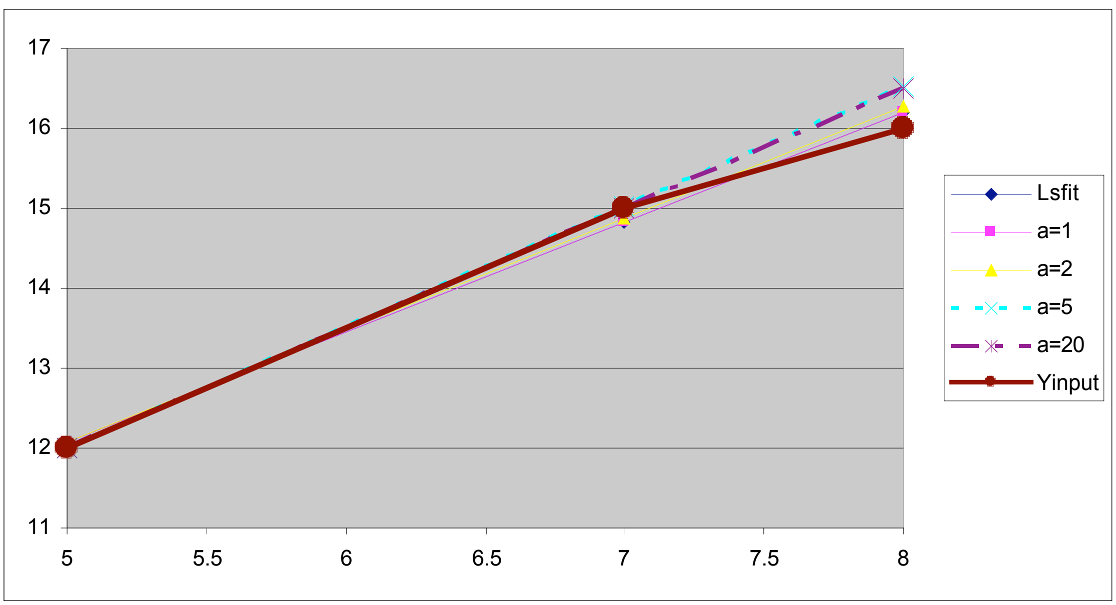

Figure 3. Upperbound fits for varying values of parameter 'a'.

Mathematica is used to solve the nonlinear simultaneous equations. The equations get more cumbersome to solve with number of data. To accelerate convergence we start with coefficients obtained for the 'a' parameter that is immediately lower than one we try to estimate. Error from the actual dataset can be captured using the residual sum square error. The asymptotic error behavior of the curve exhibits trends contrary to the expectations. The upper bound curve converges for large values of a with reducing resdiual sum square error, Figure 4.

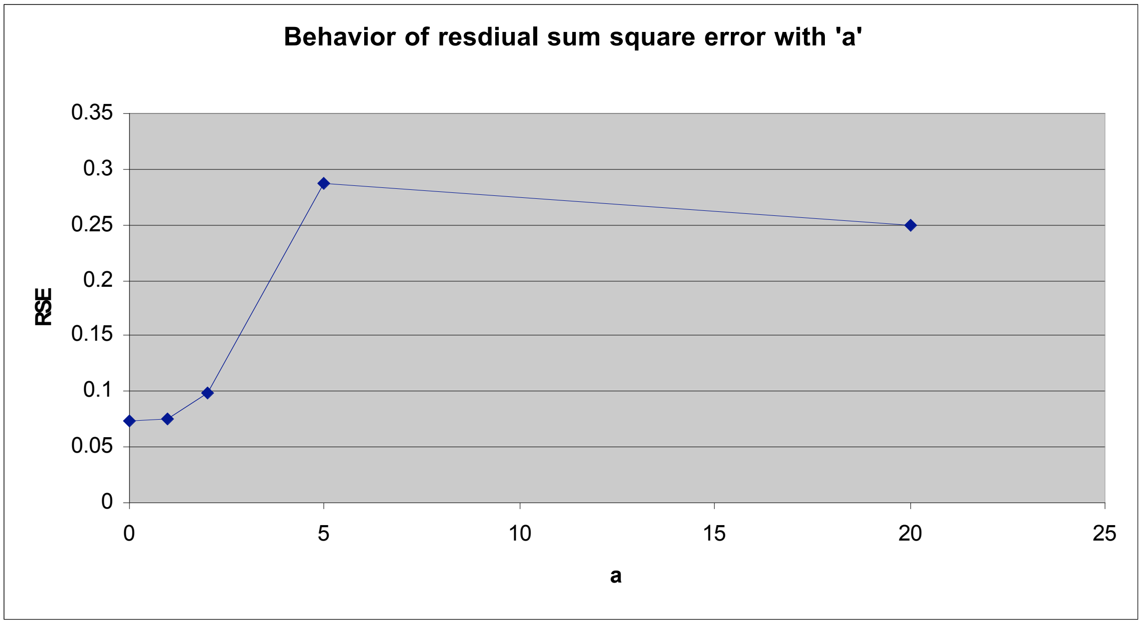

Figure 4. Residual sum square error plot for various 'a' values.

The convergence was also observed on other data sets tested. We can immediately see that by using function e^{-ax} the error can be weighted to observe lower bound. The given set of data along with a known relationship can be used to bound data from either sides.

4. Application for manufacturing process optimization.

The significance of the devised procedure is demonstrated by applying to manufacturing examples. Surface finish has been empirically modeled for ages to understand the basic relationships. Least square fit is commonly used to identify the relationship. This information is very relevant for analyzing the underlying physical process. Nevertheless, this data with its scatter cannot ably guarantee a surface finish. A design engineer generally tends to specify an upperbound for the surface finish which has to be met by every part during processing. The most common approach is to use a large factor of safety on the experimentally observed data. With high speed machining gaining prominence, and the direct relation that exist between most of these parameters and productivity, it is possible that pockets of high uncertainty might occur, so a high factor of safety might exclude potential high benefit regions. The previously presented results can be utlized much more effectively by employing one of the three procedures outlined below. These normally incorporate both the mean and variance during experiments.

The most commonly identified influence parameters for finish are speed, feed and depth of cut. Experiments were perfomed on an engine block of a certain material for 12 different conditions: 3 speeds * 2 feeds * 2 depths of cut, and the observed values are shown in the plot by the thick line (Figure 3). The basic relationship is assumed to be additive and the finish is proportional to the square of the feed and linear with respect to speed and depth of cut. These are used as *X1*, *X2*, and *X3* in the scheme, and the linear coefficients are estimated. It should be noted that, even though rugosity is proportional to the square of the feed, the equation is still linear in coefficients; hence, the above scheme can be applied without modification. This scheme was applied to estimate the values of β for the case of surface finish and the plots are shown for various values of the '*a*' parameter, Figure 5. We can see that as *a* increases, the curve moves away, enveloping the observed values at those conditioins. This is shown by the bold line corresponding to *a*=40.

The upperbound of the surface finish or other manufacturing experimental data sets can be obtained starting with two different assumptions:

1. The observed data can be assumed to be representative of the data to be observed in production, based on a sufficient experiment condition, and find the upper bound for the existing data set. This is an aggressive approach.

2. The results of repeated experiments at a given condition can be assumed as gaussian and the observed data as a small sample. A 99% limit for one sided t-distribution can be used as the value to be bounded by the scheme. This is a less conservative approach, but it can yield more statistically confident results.

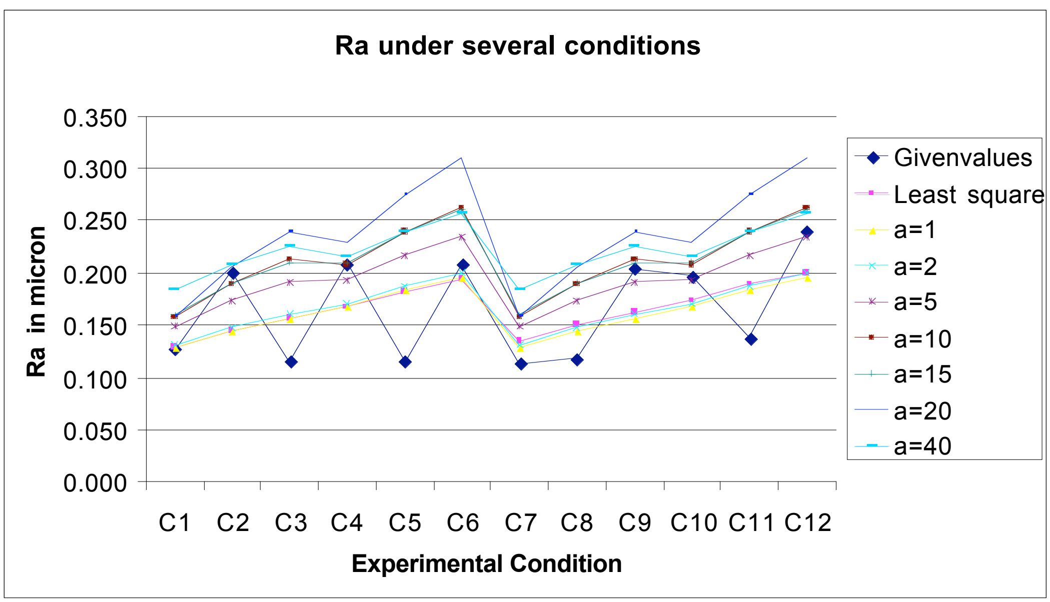

Figure 5. Upperbound Ra plots.

The graph shown in Figure 5 is based on the type 2 approach. Either scheme can be chosen, depending on the manufacturing philosophy. By applying the bound, any point in the solution space provided can be used for further optimization:

$$Solution\,space: \beta_0 + \beta_1 speed + \beta_2 feed^2 + \beta_3 depthofcut < Raspecified\ .$$

5. Conclusion

A procedure for estimating parameters that can bound experimental data from either side has been developed. The scheme was also successfully applied to estimating surface finish upperbounds for various cutting conditions. The surprising behavior of upper bound curves in limit is observed. Future work would involve applying this to more practical scenarios and demonstrate the ability of the scheme to capture high productivity pockets effecitively. Bootstrapping and Crossvalidation techniques will be used to further strengthen the claim on the equations' asymptotic behavior.

References

[1] Hines, W .H. and Montgomery, C.D., *Probability and Statistics in Engineering and Management Science*, John Wiley and Sons, 1980.

[2] Montgomery, C.D., *Response Surface Methodology*, Wiley Series in Probability and Statistics, 1995.