A Bayesian Method for Robust Tolerance Control and Parameter Design

Ramkumar Rajagopal
Enrique del Castillo
Dept. of Industrial & Manufacturing Engineering
The Pennsylvania State University
University Park, PA 16802

March, 2004

Abstract

This paper proposes a Bayesian method to set tolerance or specification limits on one or more responses and obtain optimal values for a set of controllable factors. The existence of such controllable factors (or parameters) that can be manipulated by the process engineer and that affect the responses is assumed. The dependence between the controllable factors and the responses is assumed to be captured by a regression model fit from experimental data, where the data is assumed to be available. The proposed method finds the optimal setting of the control factors (parameter design) and the corresponding specification limits for the responses (tolerance control) in order to achieve a desired posterior probability of conformance of the responses to their specifications. Contrary to standard approaches in this area, the proposed Bayesian approach uses the complete posterior predictive distribution of the responses, thus the tolerances and settings obtained consider implicitly both the mean and variance of the responses and the uncertainty in the regression model parameters.

1 Introduction: Tolerance Control

In engineering design, the limits defining the acceptable quality of a product are called tolerance or specification limits. The problem of setting these limits based on different criteria is known as tolerance control. This paper addresses the problem of setting tolerance limits on one or more quality characteristics that depend on controllable factors, $(x_1...x_k)$. The approach presented is Bayesian. Bayesian methods have been used in the literature [Peterson (2004), Miro-Quesada *et al.*(2004)] to identify the settings of $(x_1...x_k)$ that maximize the probability of conformance of the responses or quality characteristics to a pre-defined given tolerance region. However, in this paper, one of the problems we address is the inverse problem: that of identifying a tolerance region such that the probability of conformance of the response(s) to the region is at least equal to a user-defined value ϕ.

In tolerance control problems, it is also of interest to identify the setting of the control factors that gives the smallest such tolerance region. For example, suppose the design engineers in a company design a part that should be machined with thickness between *3mm* and *5mm*, and suppose that there are two controllable factors in the machine, cutting speed and pressure, that the operator can adjust to get the required thickness. Using the Bayesian optimization approaches in the literature, it is possible to find the settings of these two controllable factors that maximize the posterior probability that the part will have a thickness that is within the specification (tolerance) limits [Peterson (2004)]. However, it is possible that at these settings the value of the posterior probability is quite low, say 0.6, which means that even at the best operating setting only 60% of the manufactured parts will meet the tolerances that have been set. This is a common problem in tolerancing and can be overcome if the designer transfers tolerance requirements to the manufacturing plant, while keeping in mind the limitations of the machinery and making the most of the design flexibility [Hong *et al.* (2002)]. Thus in the previous example, the designer could prefer to adjust the design such that it is possible to set a different tolerance limit on the part that gives a higher posterior probability of obtaining conforming parts. It may be of interest, for example, to determine if there is another setting of the controllable factors where there is a high posterior probability of conformance to thickness between *5mm* and *6mm*, say a 90% probability. If this is true, at this new setting not only is the probability of conformance higher, but the tolerance region is also smaller, thus giving a lower variation in the conforming parts.

In the standard approaches to statistical tolerancing [Guttman (1988)], if Y is a quality characteristic with probability distribution P_Y^{θ}, where θ are the parameters, and if a sample $(y_1...y_n)$ of n independent observations is available, two common statistical methods have been used for constructing a tolerance region A. These are defined as,

1. the α-expectation tolerance region, given by:

$$E[C(A)] = \alpha, \tag{1}$$

2. the α-content tolerance region at confidence level γ, given by:

$$p[C(A) \geq \alpha] = \gamma, \tag{2}$$

where $C(A)$ is the coverage of the region A. These definitions are applicable for both classical or frequentist and Bayesian approaches. Both approaches are discussed in Guttman [Guttman (1981), Guttman (1988)]. However, they do not consider the problem of tolerance control in conjunction with regression, where the response depends on the settings of control factors. Also, the methods in the literature do not address the problem of finding the smallest tolerance region that satisfies one of the two criteria shown in equations (1) and (2).

The idea of a robust tolerance and parameter design was originally proposed by Taguchi [Taguchi (1986)]. Taguchi recommended tolerance design as the stage in quality control that follows parameter design. Parameter design is used to fit regression models to data and identify levels of the controllable factors that give the required mean and variation of the fitted response models. Taguchi's robust tolerance design is used to adjust the tolerances of the controllable factors that have a large influence on the response(s). However, this does not address the problem of setting tolerance or specification limits on the responses themselves. Taguchi's idea is related to what is called "transmission of errors", where variation in the controllable factors causes additional variation in the responses. In this paper, we first address the problem of setting tolerances on the responses assuming that the controllable factors can be set to fixed desired settings. The effect of transmission of errors because of variation in the controllable factors is addressed in the discussion in section 4

The remainder of the paper is organized as follows. The next section describes the proposed method for constructing tolerance limits for systems with a single response or quality characteristic. Section 3 discusses the multiple response case. A summary of the approach is given in the discussion section.

2 Single Response Systems

In this section, we consider a process with a single response or quality characteristic of interest, y. It is assumed that this response depends on k controllable factors, $x_1 ... x_k$. It is also assumed that we have data from an experiment with n runs from which we can fit a model to the response of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{3}$$

where $\mathbf{x}$ is the $(p \times 1)$ vector of regressors that are functions of the k controllable factors, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of model parameters and ϵ is the error term which is assumed to be normally distributed, $N(0, \sigma^2)$. Denote the design matrix from the experiment by an $(n \times p)$ matrix $\mathbf{X}$ and the vector of observed responses from the experiment by an $(n \times 1)$ vector $\mathbf{y}$.

2.1 Bayesian Predictive Density

The posterior predictive density of a future response vector y^* at a given setting of the model regressors $\mathbf{x}^*$ for the given data $\mathbf{y}$ is defined as [Press (1982)]:

$$p(y^*|\mathbf{x}^*, \mathbf{y}) = \int_{\sigma^2} \int_{\boldsymbol{\beta}} p(y^*|\mathbf{x}^*, \mathbf{y}, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \, d\boldsymbol{\beta} \, d\sigma^2, \tag{4}$$

where $p(y^*|\mathbf{x}^*, \mathbf{y}, \boldsymbol{\beta}, \sigma^2)$ is the likelihood function, and $p(\boldsymbol{\beta}, \sigma^2|\mathbf{y})$ is the posterior distribution of the model parameters. It is noted that the uncertainty in the model parameters is naturally accounted for by considering $\boldsymbol{\beta}$ and σ^2 to be random variables and evaluating their posterior distributions using Bayes' theorem:

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2), \tag{5}$$

where $p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)$ is the likelihood function, and $p(\boldsymbol{\beta}, \sigma^2)$ is the joint prior distribution of the model parameters. For the system described earlier, under a diffuse prior given by,

$$p(\boldsymbol{\beta}) \propto constant, \tag{6}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{7}$$

and

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta}) p(\sigma^2), \tag{8}$$

the posterior predictive density is given by a t-distribution [Press (1982)]. That is,

$$y^*|\mathbf{x}^*, \mathbf{y} \sim t_\nu(\mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \hat{\sigma}^2(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*)), \tag{9}$$

where $\nu = n - p$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{10}$$

and

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{n - p}. \tag{11}$$

As the posterior distribution of the response is a t-distribution, the posterior mean of the response given $\mathbf{x}^*$ is

$$E[y^*|\mathbf{x}^*, \mathbf{y}] = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \tag{12}$$

and the posterior variance of the response given $\mathbf{x}^*$ is

$$Var[y^*|\mathbf{x}^*, \mathbf{y}] = \frac{\nu}{\nu - 2}\hat{\sigma}^2(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*). \tag{13}$$

2.2 Optimization for Tolerance Control

The objective of the tolerance control problem is to find the setting of the controllable factors, $x_1^*...x_k^*$, that gives the smallest interval $[l, u]$ such that the posterior probability of conformance, $p(y^* \in [l, u]|\mathbf{x}^*, \mathbf{y})$, is at least ϕ, where ϕ is decided by the process engineer or the designer. In addition, there may be constraints imposed on the ranges of $\mathbf{x}^*$, l, and u. In mathematical notation, the problem is formulated as:

$$\begin{aligned}
&\min_{x_1^*...x_k^*} \; u - l \\
&s.t., \\
&p(l \leq y^* \leq u|\mathbf{x}^*, \mathbf{y}) \;\geq\; \phi \\
&\qquad\qquad\qquad\quad l \;\geq\; B_l \\
&\qquad\qquad\qquad\quad u \;\leq\; B_u \\
&\qquad\qquad\quad x_1^*...x_k^* \;\in\; \Re,
\end{aligned}$$

where B_l and B_u are, respectively, a lower bound on l and an upper bound on u, determined by the user. For a given $\mathbf{x}^*$, we have from equation (9) that $y^*|\mathbf{x}^*, \mathbf{y} \sim t_\nu(\mu_{y^*}, \sigma^2_{y^*})$, where $\mu_{y^*} = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}$, and $\sigma^2_{y^*} = \hat{\sigma}^2(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*)$. Let F_t and f_t be, respectively, the c.d.f. and the p.d.f. of this posterior t-distribution (note that F_t and f_t depend on $\mathbf{x}^*$). Based on the constraints on the bounds, B_l and B_u, a given setting $\mathbf{x}^*$ is infeasible if

$$F_t(B_u) < \phi, \tag{14}$$

or if

$$F_t(B_l) > 1 - \phi. \tag{15}$$

If the above inequalities are not true, then $P(l \leq y^* \leq u|\mathbf{x}^*, \mathbf{y}) \geq \phi$ can be satisfied, and from figure 1, it is evident that the smallest interval $[l, u]$ that encloses an area at least equal to ϕ should be centered about the mean μ_{y^*} and can be computed by

$$l = F_t^{-1}\left(\frac{1-\phi}{2}\right), \tag{16}$$

$$u = 2\mu_{y^*} - l. \tag{17}$$

However, the values of l and u computed using equations (16) and (17) need to satisfy also the constraints $l \geq B_l$ and $u \leq B_u$. Thus, when using equations (16) and (17) to obtain l and u, there are four cases as shown in figure 2. These are:

1. $\underline{l \geq B_l, u \leq B_u}$: This case satisfies all the constraints. Here, the values of l and U from equations (16) and (17) give the smallest interval $[l, u]$ for the given $\mathbf{x}^*$.

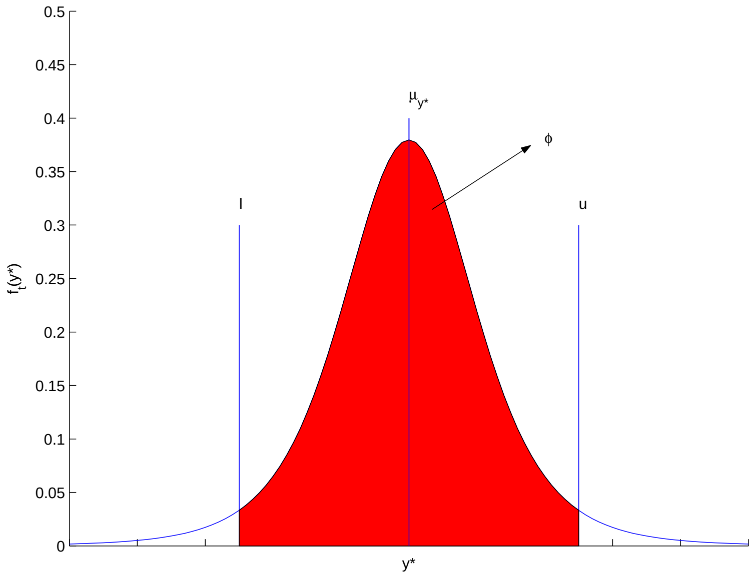

Figure 1: Sample t-distribution

2. $\underline{l \geq B_l, u > B_u}$: Here, the value of u from (17) does not satisfy the constraint $u \leq B_u$. So to find the smallest interval, we set $u = B_u$ and $l = F_t^{-1}[F_t(B_u) - \phi]$. Note that because of the feasibility check in equation (14), $F_t(B_u) - \phi$ is never less than zero. Now if the new value of $l \geq B_l$, then $[l = F_t^{-1}(F(B_u) - \phi), u = B_u]$ is the smallest interval satisfying all constraints. Otherwise, if the new $l < B_l$, then the problem for the given $\mathbf{x}^*$ is infeasible.

3. $\underline{l < B_l, u \leq B_u}$: Here, the value of l from (16) does not satisfy the constraint $l \geq B_l$. So, to get the smallest interval, we set $l = B_l$ and $u = F_t^{-1}[F_t(B_l) + \phi]$. Here, because of the feasibility check in equation (15), $F_t(B_l) + \phi$ is never greater than one. Now if the new value of $u \leq B_u$, then $[l = B_l, u = F_t^{-1}(F_t(B_l) + \phi)]$ is the smallest interval satisfying all constraints. Otherwise, if the new $u > B_u$, then the problem for the given $\mathbf{x}^*$ is infeasible.

4. $\underline{l < B_l, u > B_u}$: In this case, the problem for the given $\mathbf{x}^*$ is infeasible, as no interval $[l, u]$ that satisfies the constraints $l \geq B_l$ and $u \leq B_u$ will contain an area under the p.d.f. curve at least equal to ϕ.

The analysis of these cases provides a solution algorithm for the tolerance control problem. This is summarized in the algorithm in figure 3 to compute the values of l and u that give the minimum bound, $u - l$, for a given $\mathbf{x}^*$. The algorithm is used within a nonlinear optimization program that searches in the space of $(x_1^* ... x_k^*)$ to find the setting that gives the smallest

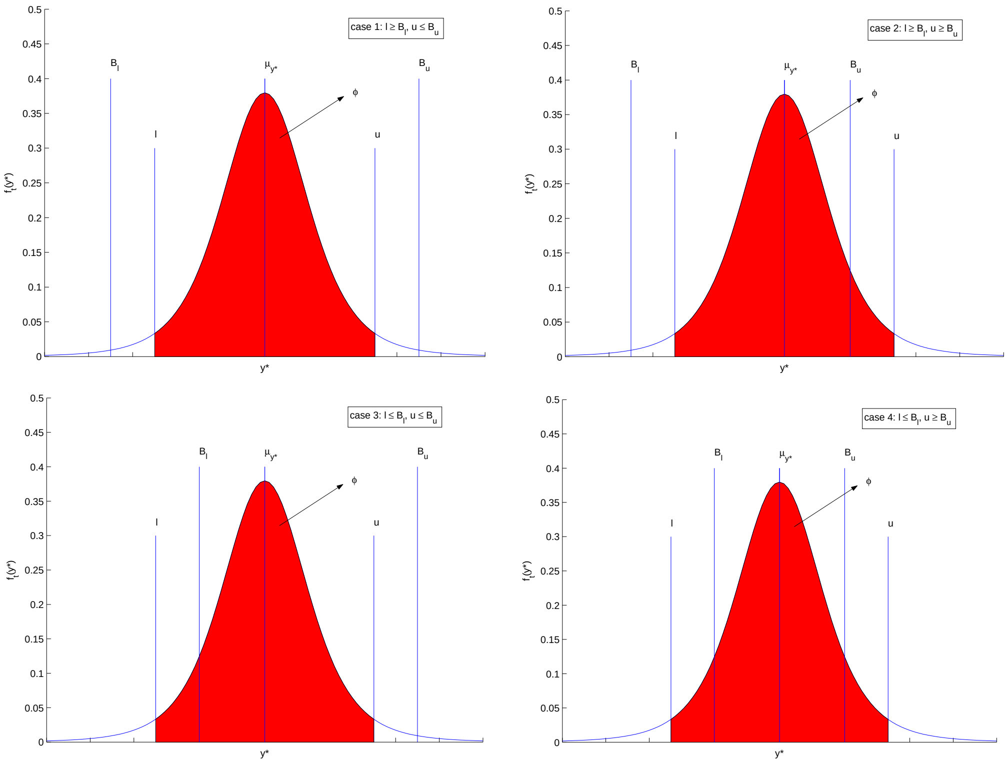

Figure 2: Four possible cases for constraints on the bounds

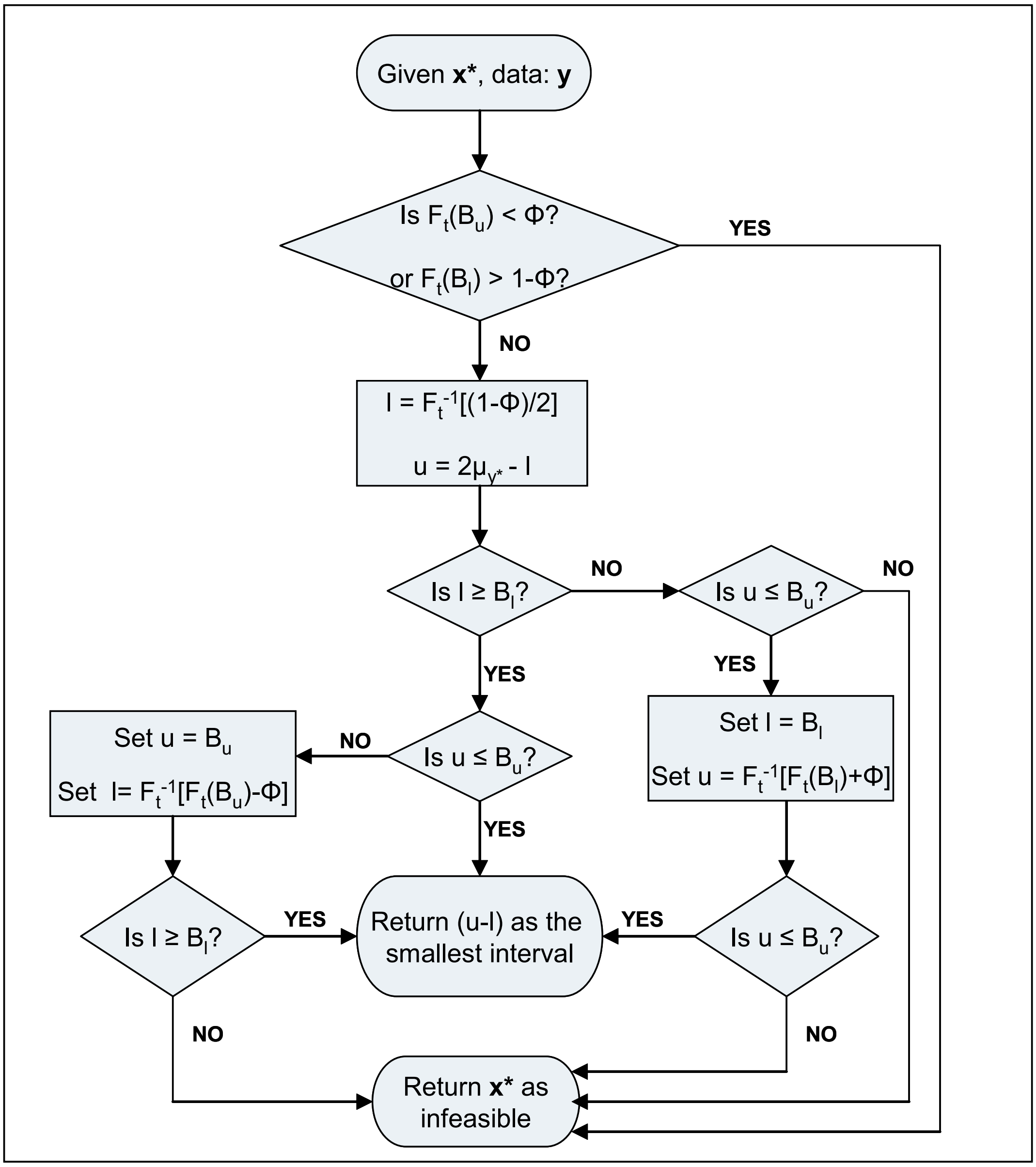

Figure 3: Algorithm to determine smallest interval $[l, u]$ for a given $\mathbf{x}^*$ under given constraints

Control Factors		Response	
x_1	x_2	Replicate 1	Replicate 2
-1	-1	7.52	8.12
1	-1	12.37	11.84
-1	1	13.55	12.35
1	1	16.48	15.32
-1.414	0	8.63	9.44
1.414	0	14.22	12.57
0	-1.414	7.90	7.33
0	1.414	16.49	17.40
0	0	15.73	17.00

Table 1: Design and experimental data [Khuri *et al.* (1987)] for example in section 2.3

$(u-l)$ under the given constraints. Note that if there are no constraints on l and u, then the optimization problem reduces to finding the value of $\mathbf{x}^*$ that gives the minimum posterior variance. The optimization problem then can be reformulated as:

$$\min_{x_1^* \dots x_k^*} \sigma^2_{y^*} \qquad s.t., \qquad x_1^* \dots x_k^* \in \Re.$$

In this case, the smallest interval $[l, u]$ can be found at any $\mathbf{x}^*$ using equations (16) and (17). We note that it is frequently the case where several settings $(x_1^* \dots x_k^*)$ satisfy the constraints of the tolerance problem, and this set of feasible solutions need not be convex or even connected. As the feasible regions maybe non-convex and even disconnected, it is recommended to run the nonlinear search algorithm in the space of $(x_1^* \dots x_k^*)$ using multiple starting points to avoid local optimums.

The examples below illustrate the method. In the first example, we illustrate how the set of feasible solutions $(x_1^* \dots x_k^*)$ may be formed by disconnected subsets. Our second example in section 2.4 presents a real manufacturing experiment.

2.3 Example 1: Two controllable factors

In this example, the data we use is taken from Khuri and Cornell [Khuri *et al.* (1987)], and is shown in table 1. The goal of the experiment was to investigate the effect of two controllable factors on a single response. As the response in this example is yield, a higher value of the response is desired. In the table, the factors are given in coded variables determined by a rotatable central composite design (CCD). The data shows observed responses for two replicates of each treatment combination. Based on the two replicates of the experiment, the model fitted to the data using the parameter estimates shown in equation (10) is given by,

$$\hat{y} = 16.3647 + 1.6753x_1 + 2.7651x_2 - 0.3337x_1x_2 - 2.4637x_1^2 - 1.9310x_2^2. \tag{18}$$

Note that the models are different when only one of the replicates is used. The model using only the first replicate is given by

$$\hat{y} = 15.7296 + 1.9608x_1 + 2.7862x_2 - 0.4800x_1x_2 - 1.9851x_1^2 - 1.6000x_2^2, \tag{19}$$

while the model using only the second replicate is given by

$$\hat{y} = 16.9999 + 1.3897x_1 + 2.7440x_2 - 0.1875x_1x_2 - 2.9423x_1^2 - 2.2621x_2^2. \tag{20}$$

As the treatment combinations for both the replicates are identical, it is noted that the coefficients in equation (18) are the average of the respective coefficients in equations (19) and (20). It is assumed that the goal of the experiment is to find the value of the controllable factors in the interval $[-1, 1]$, while at the same time setting tolerance or specification limits $[l, u]$ on the response with the desired probability of conformance. As higher values of the response are desirable in this example, it is important to set a lower bound B_l on the value of l. In addition, since the optimization finds the smallest interval $u - l$, operating at the optimal set point $(x_1^* ... x_k^*)$ not only gives the desired probability of conformance, but also the least variation in the response under the given constraints.

The dots in figure 4 represent all feasible $\mathbf{x}^*$ computed at points on a grid spaced 0.05 apart in the region $\{x_1^* \in [-1, 1], x_2^* \in [-1, 1]\}$ using the algorithm shown in figure 3. Figure 4 shows four cases based on different values of B_l, B_u and ϕ. The figure is plotted using the data from both replicates and the corresponding model in equation (18). Figure 5 shows the feasible $\mathbf{x}^*$ for two of those cases using just the data from replicate 1 and the corresponding model in equation (19), and figure 6 shows the feasible $\mathbf{x}^*$ for the same two cases as in figure 5 using just the data from replicate 2 and the corresponding model in equation (20). It can be seen from these figures that:

1. The feasible region need not be convex as shown in figure 4 for the case where $B_l = 13$, $B_u = 20$ and $\phi = 0.99$. It is noted that depending on the data and the constraints, the feasible region may even be disconnected.

2. For the case where $B_l = 13$, $B_u = 20$ and $\phi = 0.99$, the feasible set is empty when only replicate 1 (figure 5) is used or when only replicate 2 (figure 6) is used. However, when both replicates are used, although the posterior mean of y^* at a given $\mathbf{x}^*$ is the average of the posterior means of the two replicates, the posterior standard deviation of y^* is less than that of either of the replicates. Therefore, the feasible region in figure 4 is not empty when both replicates are used for $B_l = 13$, $B_u = 20$ and $\phi = 0.99$.

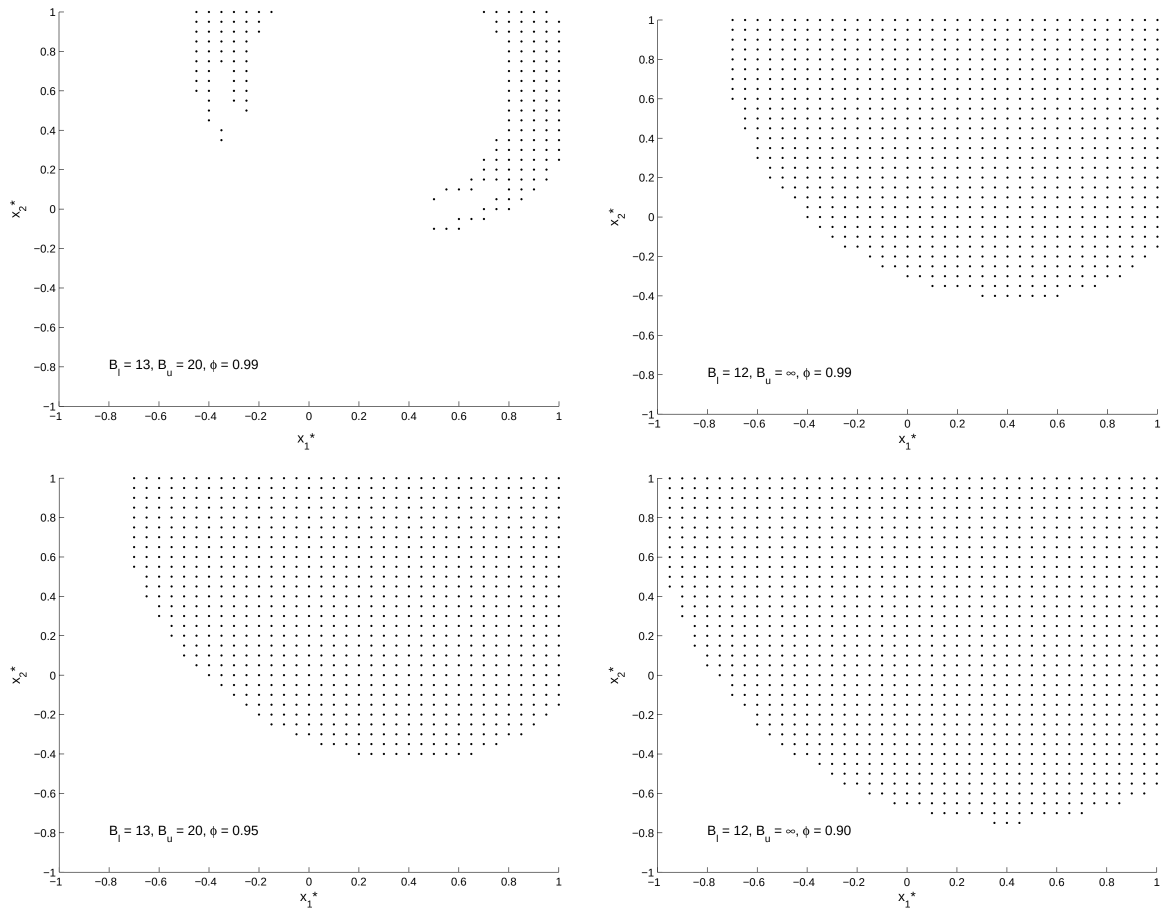

Figure 4: Feasible region for different constraints using both replicates

3. As expected in all cases, for the same values of B_l and B_u the feasible region is larger as the constraint on the value of the probability of conformance ϕ decreases.

The optimization is performed using both replicates of the data and is presented in table 2 for different combinations of B_l, B_u and ϕ. Note that the value of $\mathbf{x}^*$ that gives the minimum posterior variance for the response is obtained by solving the optimization problem without any constraints on l and u, i.e., by setting $B_l = -\infty$ and $B_u = \infty$. In this example, when the posterior variance of the response is minimized, the smallest interval $(u - l)$ obtained is 6.8967 for $\phi = 0.99$, 4.9194 for $\phi = 0.95$, and 4.0241 for $\phi = 0.90$. It can be seen from table 2 that the smallest interval obtained in all the cases is equal to or very close to the smallest possible interval without any constraints on l and u, for the respective values of ϕ. This is because the region with the desired high posterior mean also has the lowest posterior variance for the response, as can be seen in the plots shown in figure 7. However, this need not always be the case as demonstrated in our next example.

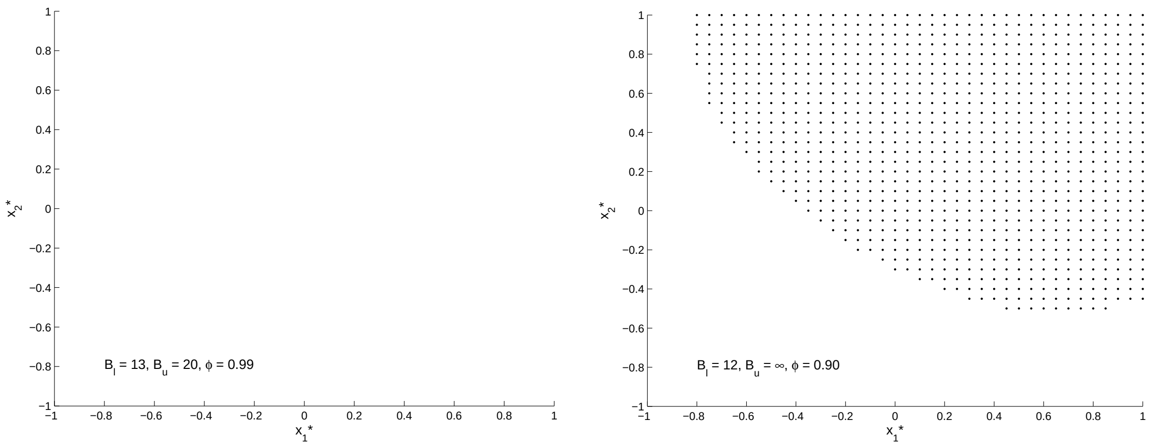

Figure 5: Feasible region under different constraints using replicate 1

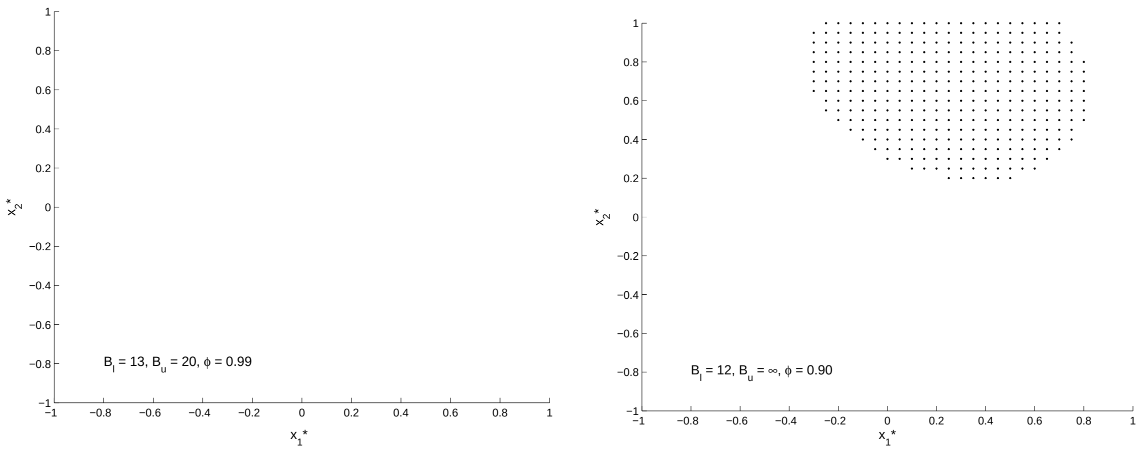

Figure 6: Feasible region under different constraints using replicate 2

B_l	B_u	ϕ	optimal x_1^*	optimal x_2^*	optimal $(u-l)$	l	u
13	20	0.99	0.9365	0.6291	6.8967	13.1032	19.9999
12	∞	0.99	0.7661	0.8281	6.8967	13.5077	20.4044
14	22	0.99	0.3596	0.9878	6.9218	14.0000	20.9218
13	20	0.95	0.7722	0.8224	4.9194	14.4857	19.4051
12	∞	0.95	-0.8022	0.7931	4.9194	12.1663	17.0857
14	22	0.95	0.8308	0.7631	4.9194	14.3702	19.2896
13	20	0.90	0.8269	0.7674	4.0241	14.8264	18.8505
12	∞	0.90	-0.8048	0.7904	4.0241	12.6002	16.6243
14	22	0.90	0.6204	0.9422	4.0241	15.1397	19.1639

Table 2: Optimization results for constructing tolerances for example in section 2.3

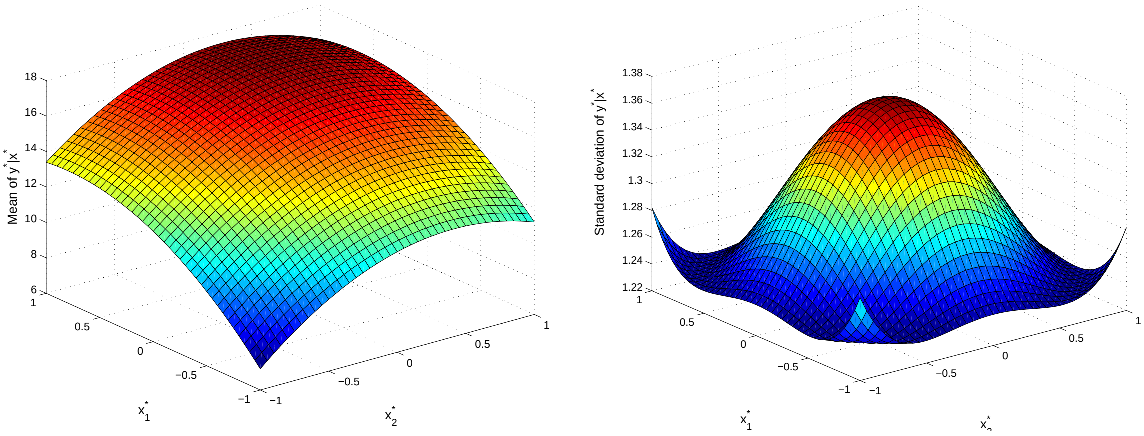

Figure 7: Posterior mean and standard deviation of the response for example in section 2.3

2.4 Example 2: Three controllable factors

This example uses machining data from Taraman (1974) and is presented in table 3. There are 3 controllable factors - cutting speed (x_1), feed (x_2) and depth of cut (x_3), and three responses - surface roughness (R), tool life (T) and cutting force (F). Table 3 gives the values of the controllable factors in $[-1, 1]$ coded form based on a central composite design. The table also shows the logarithm of the observed responses, which are used for modelling because it is expected, based on prior knowledge of the process [Taraman (1974)], that the log scaled response is better suited to fit a linear statistical model of the form shown in equation (3). In this section the proposed approach for tolerance control is demonstrated for the response, tool life (T). Based on the given data, the model fit to the log of the response, $y = log(T)$, is

$$\hat{y} = 3.5009 - 0.3031x_1 - 0.0922x_2 - 0.0915x_3 + 0.0483x_1^2 + 0.0416x_2^2 + 0.0682x_3^2 \qquad (21)$$

Based on the above model, the optimal tolerance interval can be computed for any given value of B_l , B_u and ϕ. In this example, when the posterior variance of the response is minimized, the smallest interval $(u - l)$ obtained is 0.8297 for $\phi = 0.99$, 0.6040 for $\phi = 0.95$, and 0.4980 for $\phi = 0.90$. The optimization results for a few combinations of these constraints are given in table 4. As the optimization is performed using the log of the tool life as the response, the table also shows the results transformed back into the original variable. Thus, for example, with a constraint on l of $B_l = 40$ and on u of $B_u = 100$, the smallest tolerance interval that can be set is a tool life of $[40, 73.46]$ minutes with a 95% conformance. It can also be seen from table 4 that not all combinations are feasible. For example, with a value of $B_l = 45$ and no constraint on u for tool life, there is no feasible solution at 99% conformance,

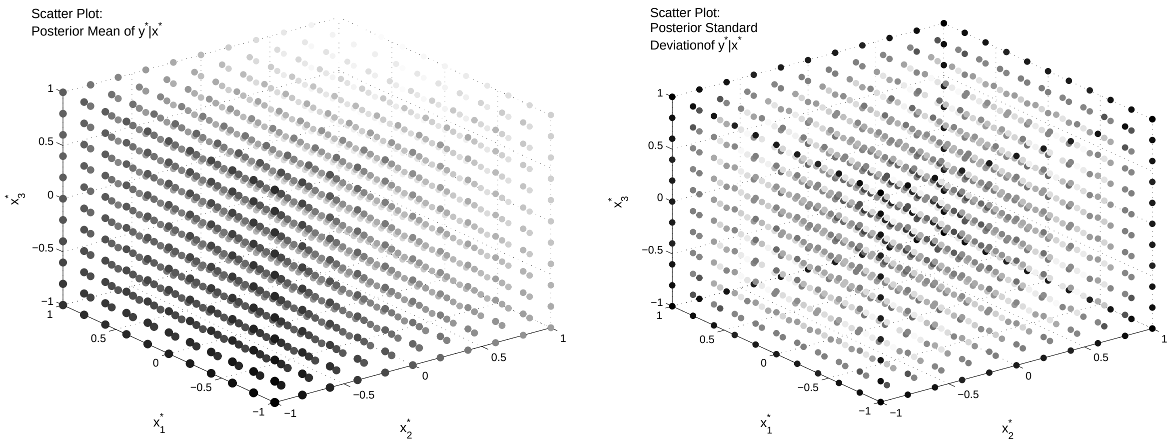

Figure 8: Posterior mean and standard deviation of the response for example in section 2.4

but there are feasible solutions at 95% or lower conformance. It is noted that unlike in the previous example the smallest interval $(u-l)$ here is different for different values of B_l and B_u, for the same value of ϕ. This is because the region with the desired high posterior mean does not provide low posterior variance. Hence, as the constraints on l and u get tighter, the optimal interval $(u-l)$ is larger. This can be seen in the scatter plots of the posterior mean and variance shown in figure 8. In the figure, the darker circles show regions with high magnitudes for both the posterior mean and standard deviation. In the plot the posterior mean is high at the corner point $(-1,-1,-1)$, but the posterior standard deviation is also high in this region. Thus as the constraint B_l on the lower bound is increased, the optimal solution moves closer to the corner point, causing an increase in the posterior standard deviation and consequently resulting in a wider tolerance interval. Using the results from this example, it is possible to choose a setting (x_1^*, x_2^*, x_3^*) where the tool life is obtained within the optimal tolerance intervals at the required probability of conformance. However, the performance of the other two responses, surface finish (R) and cutting force (F), at this setting is not known. The next section extends the Bayesian method for constructing simultaneous tolerance intervals to multiple response systems.

3 Multiple Response Systems

Here, it is assumed that there are q responses or quality characteristics of interest that depend on one or more of k controllable factors. It is assumed that each of the q responses is of the form

$$y_j = \mathbf{x}_j' \boldsymbol{\beta}_j + \epsilon_j, \tag{22}$$

x_1	x_2	x_3	$R(CLA\,\mu\,in.)$	$T(min)$	$F(lbs)$	$log(R)$	$log(T)$	$log(F)$
-1	-1	-1	88	70	53	4.4773	4.2485	3.9703
1	-1	-1	76	29	48	4.3307	3.3673	3.8712
-1	1	-1	259	60	100	5.5568	4.0943	4.6052
1	1	-1	194	28	92	5.2679	3.3322	4.5218
-1	-1	1	105	64	76	4.6540	4.1589	4.3307
1	-1	1	82	32	74	4.4067	3.4657	4.3041
-1	1	1	270	44	155	5.5984	3.7842	5.0434
1	1	1	250	24	150	5.5215	3.1781	5.0106
0	0	0	123	35	82	4.8122	3.5553	4.4067
0	0	0	136	31	85	4.9127	3.4340	4.4427
0	0	0	130	38	83	4.8675	3.6376	4.4188
0	0	0	121	35	85	4.7958	3.5553	4.4427
-1.414	0	0	159	52	88	5.0689	3.9512	4.4773
1.414	0	0	115	23	80	4.7449	3.1355	4.3820
0	-1.414	0	77	40	50	4.3438	3.6889	3.9120
0	1.414	0	324	28	129	5.7807	3.3322	4.8598
0	0	-1.414	114	46	68	4.7362	3.8286	4.2195
0	0	1.414	215	33	124	5.3706	3.4965	4.8203
-1.414	0	0	139	46	87	4.9345	3.8286	4.4659
1.414	0	0	111	27	78	4.7095	3.2958	4.3567
0	-1.414	0	61	37	49	4.1109	3.6109	3.8918
0	1.414	0	340	34	130	5.8289	3.5264	4.8675
0	0	-1.414	128	41	71	4.8520	3.7136	4.2627
0	0	1.414	232	28	123	5.4467	3.3322	4.8122

Table 3: Design and experimental data for machining example in section 2.4 [Taraman (1974)]

B_l	B_u	ϕ	optimal x_1^*	optimal x_2^*	optimal x_3^*	optimal $(u-l)$	l	u
Transformed Response: log (Tool Life in min.)								
$log(45)$	∞	0.99	Infeasible					
$log(40)$	log(100)	0.99	-1.0000	-0.8471	-0.9385	0.8598	3.6889	4.5486
$log(45)$	∞	0.95	-1.0000	-0.8533	-0.9465	0.6264	3.8067	4.4331
$log(40)$	log(100)	0.95	-0.8687	-0.6983	-0.7361	0.6079	3.6889	4.2968
$log(45)$	∞	0.90	-0.9845	-0.7429	-0.8020	0.5071	3.8067	4.3138
$log(40)$	log(100)	0.90	-0.7669	-0.6668	-0.6874	0.4988	3.6889	4.1876
Original Response: Tool Life in min.								
45	∞	0.99	Infeasible					
40	100	0.99	-1.0000	-0.8471	-0.9385	54.4992	40.0008	94.5000
45	∞	0.95	-1.0000	-0.8533	-0.9465	39.1903	45.0017	84.1920
40	100	0.95	-0.8687	-0.6983	-0.7361	33.4635	40.0008	73.4643
45	∞	0.90	-0.9845	-0.7429	-0.8020	29.7222	45.0017	74.7239
40	100	0.90	-0.7669	-0.6668	-0.6874	25.8637	40.0008	65.8645

Table 4: Optimization results for constructing tolerances for example in section 2.4

	identical regressors	unrelated regressors
uncorrelated errors	All the y_j have the same set of regressors, $\mathbf{x}_j = \mathbf{x}$, and the error terms ϵ_j are uncorrelated between the responses	Different y_j may have a different set of regressors, i.e., $\mathbf{x}_j$ is different for different j, and the error terms ϵ_j are uncorrelated between the responses
correlated errors	All the y_j have the same set of regressors, $\mathbf{x}_j = \mathbf{x}$, and the error terms ϵ_j are correlated between the responses	Different y_j may have a different set of regressors, i.e., $\mathbf{x}_j$ is different for different j, and the error terms ϵ_j are correlated between the responses

Table 5: Categories of models for multiple response systems

where $\mathbf{x}_j$ is a $(p_j \times 1)$ vector of regressors, $\boldsymbol{\beta}_j$ is a $(p_j \times 1)$ vector of model parameters and ϵ_j is the error term for response y_j. Denote by $\mathbf{\Sigma}$ the $(q \times q)$ variance-covariance matrix of the error terms. Note that if all the responses have identical regressors, then $\mathbf{x}_j = \mathbf{x}$ for all j, and if the error terms ϵ_j are uncorrelated between the responses, then $\mathbf{\Sigma}$ is a diagonal matrix. There are four different ways to model multiple response systems based on the regressors present in the models for the individual responses and the correlation of the error terms between the responses, as summarized in table 5. The Bayesian posterior predictive density depends on how the multiple responses are modelled and is discussed below.

3.1 Bayesian Predictive Density

For the cases where the error term ϵ_j is uncorrelated between the responses, i.e., $\mathbf{\Sigma}$ is diagonal, each of the q responses can be modelled independently from the data, regardless of whether the responses have identical regressors or not. As in the single response case, it is assumed that there is data from an experiment with n runs. The observed responses from the experiment are denoted by $(n \times 1)$ vectors $\mathbf{y}_j$, where $j = \{1...q\}$, and the corresponding design matrices are denoted by $\mathbf{X}_j$. It is noted that $\mathbf{X}_j = \mathbf{X}$ for all j if the responses are modelled with identical regressors. When $\mathbf{\Sigma}$ is diagonal, the joint posterior probability of conformance at a new setting $\{x_1^*...x_k^*\}$ for the q responses $\mathbf{y}^* = (y_1^*...y_q^*)$ is simply the product of the marginal posterior probabilities of conformance of the individual responses. Thus, given the data $\mathbf{Y} = (\mathbf{y}_1...\mathbf{y}_q)$,

$$\begin{aligned} p(\mathbf{y}^* \in V|\mathbf{x}_j^* \forall j, \mathbf{Y}) &\equiv p(y_1^* \in [l_1, u_1], y_2^* \in [l_2, u_2]...y_q^* \in [l_q, u_q]|\mathbf{x}_j^* \forall j, \mathbf{Y}) \quad (23) \\ &= \prod_{j=1}^{q} p(y_j^* \in [l_j, u_j]|\mathbf{x}_j^*, \mathbf{y}_j), \quad (24) \end{aligned}$$

where V is the region enclosed by $[l_i, u_i]\,\forall\, i$. Therefore, for the diffuse priors described by equations (6), (7) and (8), each of the $p(y_j \in [l_j, u_j]|\mathbf{x}_j^*, \mathbf{y}_j)$ is obtained from the c.d.f. of the t-distribution shown in equation (9).

Equations (23) and (24) do not hold if the error terms are correlated, i.e., $\mathbf{\Sigma}$ is non-diagonal. In such cases, the responses can be modelled as either Standard Multivariate Regression (SMR) or Seemingly Unrelated Regression (SUR), where the former assumes that all the response models have the same set of regressors, i.e., $\mathbf{X}_j = \mathbf{X}\ \forall j$ and the latter assumes that each response model may have different regressors. For the SMR case, the joint posterior probability distribution is given by a multivariate $\mathbf{T}$-distribution [Press (1982)]. That is, given the $(n \times p)$ design matrix $\mathbf{X}$, and the $(n \times q)$ response data matrix $\mathbf{Y}$, the posterior density at a future set of observations given by $(p \times 1)$ vector $\mathbf{x}^*$ is

$$\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y} \sim T_\nu^q(\mathbf{B}'\mathbf{x}^*, \mathbf{H}^{-1}), \quad (25)$$

where $\nu = n - p - q + 1$,

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (26)$$

$$\mathbf{H} = \frac{\nu \mathbf{S^{-1}}}{1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*}, \quad (27)$$

and

$$\mathbf{S} = (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}). \quad (28)$$

For the SUR case, the posterior predictive density has to be computed by Gibbs sampling. Percy (1992) shows how a sample of the posterior observation $\mathbf{y}^*$ can be obtained from its posterior distribution using Gibbs sampling.

3.2 Optimization

It is assumed that the objective function is to find the setting $(x_1^*...x_k^*)$ that minimizes some given function, A, such that the posterior probability of conformance of each response, $p(y_i^* \in [l_i, u_i]|\mathbf{x}_i^*, \mathbf{y}_i)$ is at least ϕ_i, where the ϕ_i's are decided by the plant engineer or designer. Here also, each of the l_i's and u_i's may be constrained to lie within given bounds. For example, the objective function $A = \prod_i (u_i - l_i)$ finds the smallest (in terms of volume) q-dimensional cuboid that satisfies the given constraints. The function A could also be chosen such that different weights are given to the bounds on different responses. The objective function in this case is formulated as,

$$
\begin{aligned}
&\min_{x_1^*...x_k^*} \ A = \prod_{i=1}^{q} (u_i - l_i) \\
&\quad s.t., \\
&p(l_1 \le y_1^* \le u_1|\mathbf{x}_1^*, \mathbf{y}_1) \ \ge \ \phi_1 \\
&p(l_2 \le y_2^* \le u_2|\mathbf{x}_2^*, \mathbf{y}_2) \ \ge \ \phi_2 \\
&\qquad\qquad\qquad\qquad \vdots \ \ge \ \vdots \\
&p(l_q \le y_q^* \le u_q|\mathbf{x}_q^*, \mathbf{y}_q) \ \ge \ \phi_q \\
&\qquad\qquad\qquad\qquad l_i \ \ge \ B_{l_i}, \ \ \forall i = \{1...q\} \\
&\qquad\qquad\qquad\qquad u_i \ \le \ B_{u_i}, \ \ \forall i = \{1...q\} \\
&\qquad\qquad\qquad x_1^*...x_k^* \ \in \ \Re.
\end{aligned}
$$

For a given $\mathbf{x}_i^*$, if the error terms are uncorrelated, the smallest interval $[l_i, u_i]$ for each response y_i can be found using the marginal posterior distribution $(y_i^*|\mathbf{x}_i^*, \mathbf{y}_i)$ and the algorithm previously shown in figure 3. As all the $(u_i - l_i) > 0$, this also gives the smallest value of $A = \prod_i (u_i - l_i)$ for that $\mathbf{x}^*$. This is true for any A that is an increasing function of each $(u_i - l_i)$. The algorithm for finding the smallest A for a given $\mathbf{x}_i^*$ is used within a nonlinear optimization program that searches within the space of the feasible $(x_1^*...x_k^*)$ to find the setting that gives the smallest value of A. The methodology is illustrated by an example in the next section.

If the error terms are correlated, then a nonlinear optimization program that searches in the feasible space of both $(x_1^*...x_k^*)$ and $[l, u]$ must be used to solve the optimization problem for both the SMR and the SUR models. It is noted that depending on the size of the problem, the optimization could be tedious especially in the SUR case.

3.3 Multiple response example

This example uses the machining data from Taraman (1974) shown in table 3 to simultaneously set tolerances on all the three responses. Once again, the log of the responses are used for modelling, where $y_1 = log(R)$, $y_2 = log(T)$, and $y_3 = log(F)$. The response models obtained are,

$$\hat{y}_1 = 4.8773 - 0.0960x_1 + 0.5336x_2 + 0.1429x_3 - 0.0216x_1^2 + 0.0543x_2^2 + 0.0969x_3^2, \quad (29)$$

$$\hat{y}_2 = 3.5009 - 0.3031x_1 - 0.0922x_2 - 0.0915x_3 + 0.0483x_1^2 + 0.0416x_2^2 + 0.0682x_3^2, \quad (30)$$

$$\hat{y}_3 = 4.4260 - 0.0332x_1 + 0.3391x_2 + 0.2092x_3 - 0.0019x_1^2 - 0.0208x_2^2 + 0.0522x_3^2. \quad (31)$$

Figure 9 shows the feasible $\mathbf{x}^*$ plotted on a grid spaced 0.1 apart in the region $\{x_1^* \in [-1,1], x_2^* \in [-1,1], x_3^* \in [-1,1]\}$, assuming a desired probability of conformance $\phi_i = 0.80\,\forall\, i$ for four different cases:

1. All the three responses have constraints on l and u. Here, surface roughness has a constraint $B_u = 120$, tool life has a constraint $B_l = 35$, and cutting force has a constraint $B_u = 60$. Note that all the constraints are one-sided as it is desired that the surface roughness be as low as possible, tool life be as high as possible and cutting force be as low as possible.

2. Only the surface roughness response is constrained with a value of $B_u = 120$.

3. Only the tool life response is constrained with a value of $B_l = 35$.

4. Only the cutting force response is constrained with a value of $B_u = 60$.

As expected, from figure 9, it is seen that the feasible region itself is much smaller when constraints are imposed on all the three responses simultaneously. Table 6 shows the results of the optimization to set simultaneous tolerance limits on all the responses for different combinations of the constraints. Here the objective function used is $A = \prod_i (u_i - l_i)$, where $[l_i, u_i]$ is the tolerance limit on response i in the logarithmic scale. Thus for example, in the table, at a desired value of probability of conformance $\phi_i = 0.9\ \forall i$, and with constraints $B_u = 110$ on surface roughness, $B_l = 45$ on tool life and $B_u = 90$ on cutting force, the optimal setting of the controllable factors obtained is $[-0.9309, -0.8317, -0.8001]$ where the tolerance limit on surface roughness is $[71.8, 110.0]$, on tool life is $[45.0, 74.8]$, and on cutting force is $[53.7, 58.4]$. Thus, using the methodology proposed it is possible to set simultaneous tolerances on multiple responses with a desired probability of conformance.

For the multiple response case, the solution to the optimization problem also depends on the choice of the user-defined function A. For example, suppose there are two responses.

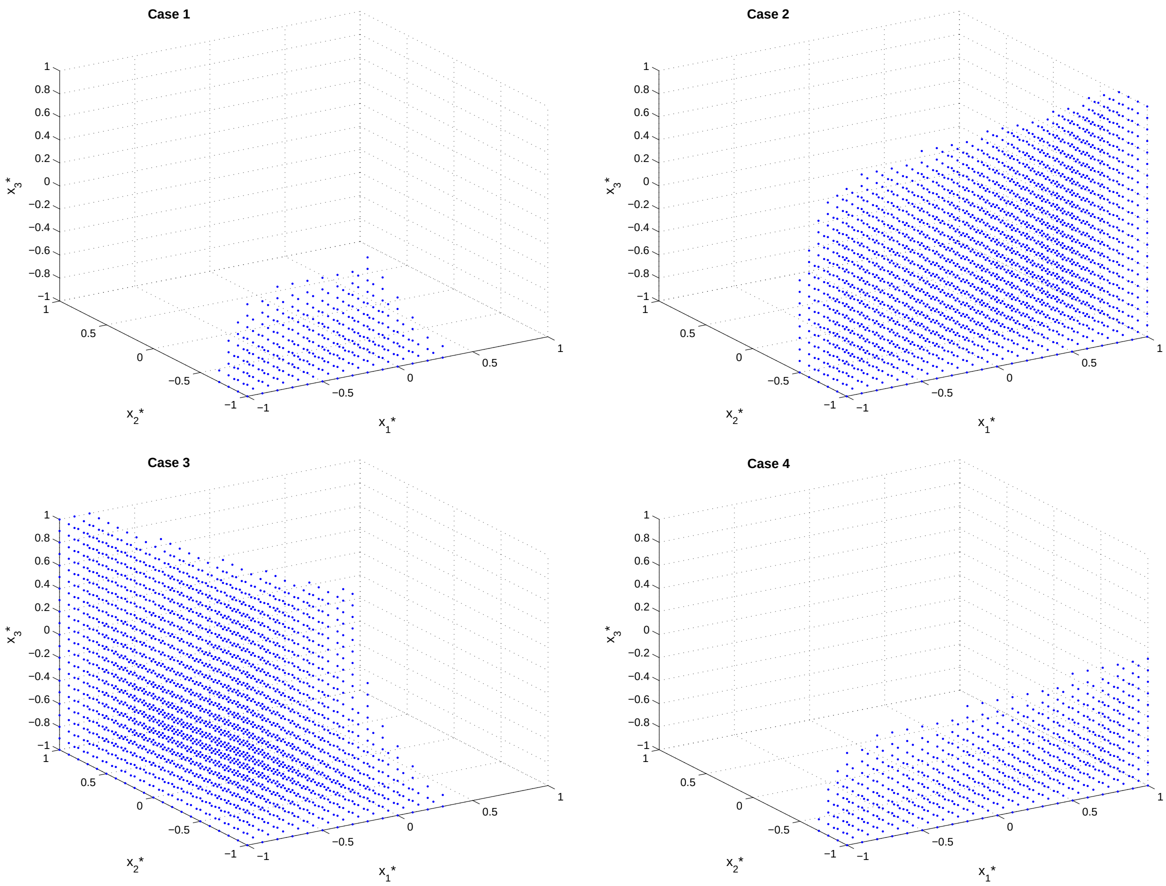

Figure 9: Feasible Regions for $\phi_i = 0.8\ \forall i$ for multiple response example in section 3.3

Suppose response 1 gives the smallest interval $(u_1^a - l_1^a) = 1$ at the setting $\mathbf{x}_a^*$ and the smallest interval $(u_1^b - l_1^b) = 2$ at the setting $\mathbf{x}_b^*$, and response 2 gives the smallest interval $(u_2^a - l_2^a) = 1.2$ at the setting $\mathbf{x}_a^*$ and the smallest interval $(u_2^b - l_2^b) = 0.5$ at the setting $\mathbf{x}_b^*$. Assuming $A = \prod_i (u_i - l_i)$, the value of A at $\mathbf{x}_a^*$ is 1.2 and at $\mathbf{x}_b^*$ is 1.0. In other words, the solution at $\mathbf{x}_a^*$ is sub-optimal given the better value of A at $\mathbf{x}_b^*$. If instead, we assume $A = \sum_i (u_i - l_i)$, then the value of A at $\mathbf{x}_a^*$ is 2.2 and at $\mathbf{x}_b^*$ is 2.5. Therefore, the solution at $\mathbf{x}_b^*$ is now inferior than the solution at $\mathbf{x}_a^*$ for the multiple response problem.

		R		T		F		optimal values				R		T		F	
	$\phi_i\ \forall i$	B_l	B_u	B_l	B_u	B_l	B_u	x_1^*	x_2^*	x_3^*	A	l_1	u_1	l_2	u_2	l_3	u_3
Transformed Responses: logarithmic scale																	
1	0.90	$-\infty$	$log(110)$	$log(45)$	∞	$-\infty$	$log(90)$	-0.931	-0.832	-0.800	0.0183	4.274	4.701	3.807	4.315	3.983	4.067
2	0.90	$-\infty$	$log(110)$	$log(45)$	∞	$-\infty$	∞	-0.945	-0.833	-0.773	0.0183	4.274	4.701	3.807	4.315	3.986	4.071
3	0.90	$-\infty$	$log(110)$	$-\infty$	∞	$-\infty$	$log(90)$	0.648	-0.647	-0.646	0.0174	4.222	4.641	3.241	3.739	4.021	4.104
4	0.90	$-\infty$	∞	$log(45)$	∞	$-\infty$	$log(90)$	-0.960	-0.732	-0.820	0.0183	4.323	4.749	3.807	4.315	4.018	4.103
5	0.75	$-\infty$	$log(110)$	$log(45)$	∞	$-\infty$	$log(90)$	-0.777	-0.729	-0.803	0.0057	4.382	4.671	3.807	4.150	4.029	4.087
6	0.75	$-\infty$	$log(100)$	$log(50)$	∞	$-\infty$	$log(60)$	-0.978	-0.900	-0.776	0.0060	4.311	4.605	3.912	4.263	3.975	4.033
7	0.75	$-\infty$	$log(100)$	$log(55)$	∞	$-\infty$	$log(60)$	infeasible									
8	0.75	$-\infty$	$log(90)$	$log(50)$	∞	$-\infty$	$log(60)$	infeasible									
Original Responses																	
1	0.90	0	110	45	∞	0	90	-0.931	-0.832	-0.800	0.0183	71.8	110.0	45.0	74.8	53.7	58.4
2	0.90	0	110	45	∞	0	∞	-0.945	-0.833	-0.773	0.0183	71.8	110.0	45.0	74.8	53.8	58.6
3	0.90	0	110	0	∞	0	90	0.648	-0.647	-0.646	0.0174	68.2	103.6	25.5	42.0	55.7	60.6
4	0.90	0	∞	45	∞	0	90	-0.960	-0.732	-0.820	0.0183	75.4	115.5	45.0	74.8	55.6	60.5
5	0.75	0	110	45	∞	0	90	-0.777	-0.729	-0.803	0.0057	80.0	106.8	45.0	63.4	56.2	59.5
6	0.75	0	100	50	∞	0	60	-0.978	-0.900	-0.776	0.0060	74.5	100.0	50.0	71.0	53.2	56.4
7	0.75	0	100	55	∞	0	60	infeasible									
8	0.75	0	90	50	∞	0	60	infeasible									

Table 6: Optimization results for setting tolerance regions for the multiple response example in section 3.3 where $A = \prod_i (u_i - l_i)$

4 Discussion

A Bayesian method was proposed to set tolerance limits on one or more responses to provide a given desired probability of conformance, and to determine at the same time the optimal settings of the control factors that the response(s) depend on. The method not only gives the tolerance interval that satisfies the constraints on the mean and the probability of conformance, but also finds the smallest such interval in case of single response systems, and the smallest value of a given function of the intervals for multiple response systems. This ensures that the quality characteristics or responses that adhere to the specification also have the smallest variation between themselves. The proposed method was illustrated by two examples for single responses systems and an example for a multiple response system. Some additional comments on the method are given below:

1. As the posterior predictive distribution of the response depends on the observed data, the solution to the tolerance control problem is also dependant on it. As seen in the example in section 2.3, the optimization problem had a non-empty feasible region when data from both the replicates was used, but had empty feasible regions when only one of the replicates was used. Thus, when additional experimental data are used in any of the examples presented, it is possible to get smaller tolerance regions that satisfy the given constraints.

2. In the previous sections, it was assumed that the controllable factors can be set to desired values by the user. However, in practice, there are errors associated with these settings. This error is also transmitted to the response which could result in a lower probability of conformance of the response ϕ_r to the calculated tolerance region, than what was originally constrained by ϕ. The reduced probability of conformance ϕ_r can be estimated if the distribution of the errors in the control factors is known.

   Suppose l^o and u^o are the calculated tolerance limits at the desired setting of the controllable factors $\mathbf{x^o} = (x_1^o ... x_k^o)$ as a result of the optimization, then assuming that the j^{th} controllable factor is normally distributed with a variance of σ_j^2 about the setting x_j^o, it is possible to represent the actual value of the controllable factor as $\tilde{x}_j = x_j^o + z_j$, where $z_j \sim N(0, \sigma_j^2)$. Thus, we have the actual settings of the controllable factors $\tilde{\mathbf{x}} = \mathbf{x^o} + \mathbf{z}$, where $\tilde{\mathbf{x}} = (\tilde{x}_1 ... \tilde{x}_k)$ and $\mathbf{z} = (z_1 ... z_k)$. The posterior distribution of the response for a single response system given the actual setting of the controllable factors $p(y^*|\tilde{\mathbf{x}}, \mathbf{y})$ is given by equation (9). If $\tilde{\mathbf{x}}$ is known then the reduced probability of conformance ϕ_r is given by $p(y^* \in [l^o, u^o]|\tilde{\mathbf{x}}, \mathbf{y})$, and can be computed from the c.d.f. of the t-distribution. However, as the actual value of $\tilde{\mathbf{x}}$ is not known because of the random component $\mathbf{z}$, the value of ϕ_r is computed by taking the expected value with respect to $\mathbf{z}$:

$$\phi_r = E_{\mathbf{z}} \left[p(y^* \in [l^o, u^o] | \tilde{\mathbf{x}}, \mathbf{y}) \right]. \tag{32}$$

In the above equation, the expected value can be estimated by simulation using the steps below:

(a) Choose sufficiently large N. Set $count = 1$.

(b) Generate $\mathbf{z}(count)$ by sampling from the distribution $z_j \sim N(0, \sigma_j^2)\ \forall j$.

(c) Set $\tilde{\mathbf{x}}(count) = \mathbf{x^o} + \mathbf{z}(count)$.

(d) Compute $p(y^* \in [l^o, u^o] | \tilde{\mathbf{x}}(count), \mathbf{y})$.

(e) Set $count = count + 1$. Repeat steps a, b and c until $count > N$.

(f) Estimate the expected value using the Weak Law of Large Numbers (WLLN):

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[p(y^* \in [l^o, u^o] | \tilde{\mathbf{x}}(i), \mathbf{y}) \right] = E_{\mathbf{z}} \left[p(y^* \in [l^o, u^o] | \tilde{\mathbf{x}}, \mathbf{y}) \right]. \tag{33}$$

It is noted that if large variations in the controllable factors are expected, then the reduced probability of conformance ϕ_r may be considerably less than ϕ. In such cases, it is recommended that the user is conservative in choosing ϕ during the optimization to obtain the tolerance limits.

For multiple response systems, the reduced probability of conformance can similarly be obtained for each response by taking the expected value of the marginal posterior probability of conformance $p(y_i^* \in [l_i^o, u_i^o] | \tilde{\mathbf{x}}_i, \mathbf{y}_i)$ with respect to the corresponding $\mathbf{z}_i$.

References

Guttman, I. (1981) *Statistical Tolerance Regions: Classical and Bayesian*, Griffin, London.

Guttman, I. (1988) *Tolerance regions, statistical. Vol. 9. Encyclopedia of Statistical Sciences, Eds.: Kotz, S., and Johnson, N. L.*, John Wiley & Sons, New York.

Hong, Y.S., and Chang, T.C. (2002) "A Comprehensive Review of Tolerancing Research", *International Journal of Production Research*, 40(11), 2425-2459.

Khuri, A.I., and Cornell, J.A. (1987) *Response Surfaces*, Marcel Dekker Inc., New York, NY.

Miro-Quesada, G., Del Castillo, E., and Peterson, J. P. (2004) "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variable", to appear in Journal of Applied Statistics.

Percy, D.F. (1992) "Predictions for Seemingly Unrelated regressions", *Journal of the Royal Statistical Society Series B*, 54, 243-252.

Peterson, J.P. (2004) "A Bayesian Reliability Approach to Multiple Response Surface Optimization", to appear in *Journal of Quality Technology.*

Press, S.J. (1982) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, 2nd ed., Robert E. Krieger Publishing, Malabar, FL.

Press, S.J. (2003) *Subjective and Objective Bayesian Statistics*, 2nd ed., John Wiley & Sons, New York, NY.

Taguchi, G. (1986) *Introduction to Quality Engineering*, Asian Productivity Organization, UNIPUB, White Plains, New York.

Taraman, K. (1974) "Multi Machining Output - Multi independent variable turning research by response surface methodology", *International Journal of Production Research*, 12(2), 233-245.

Zellner, A. (1986) "On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions", *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, eds. Goel, P.K., and Zellner, A., Amsterdam, North-Holland.