An improved optimisation procedure for Desirability Indices *

Dipl.-Stat. Detlef Steuer
Universität Dortmund
LS Computergestützte Statistik
steuer@statistik.uni-dortmund.de

April 2000

Abstract

As will be shown the current use of Desirability Indices for optimisation purposes in experimental design gives biased results in general. Researchers were satisfied with approximative solutions as unbiased results would have required analytical expressions for the distributions of Desirability Indices. These expressions are unavailable. Todays computing power allows to use Monte-Carlo estimators for estimating exact solutions instead of analytical solutions and therefore to improve the estimation process for Desirabilities.

`Keywords:` MCO, MCD, MCDA, Desirability Function, Desirability Index, numerical optimisation, bias, Monte-Carlo estimation, computer intensive procedures.

1 Introduction

Desirability Functions (DFs) are a popular tool to perform Multi-Criteria-Optimisations (MCOs). As is well known the key problem in MCO are so called competing responses, which can not be simultaneously optimised. Any improvement in response, say, Y_1 diminishes response Y_2. DFs translate all responses individually in a user defined way onto a unitless scale. The

*Work was partly supported by the DFG (SFB 475, "Reduction of complexity in multivariate data structures")

individual Desirabilities then are combined to a single measure, the so called Desirability Index (DI) to generate "best compromise" solutions.

Throughout this paper we assume a multivariate response $Y \in I\!R^Z$, i.e. Z individual responses $Y_1, \ldots, Y_Z$ to be measured and F controllable factors $X_1, \ldots, X_F$. A set of values $X_1, X_2, \ldots, X_F$ is called a factor level setting X. The relationship between responses Y and factor level settings X is given by a function f with individual components $f_i, i = 1, 2, \ldots, Z$:

$$
\begin{aligned}
Y &= f(X) + \epsilon && (1)\\
&= (f_1(X), f_2(X), \ldots, f_Z(X))' + \epsilon && (2)\\
&= (f_i(X_1, \ldots, X_F) + \epsilon_i)'_{i=1,2,\ldots,Z}, \quad \epsilon_i \sim N(0, \sigma_i), i = 1, 2, \ldots Z && (3)\\
&= (Y_1, Y_2, \ldots, Y_Z)'. && (4)
\end{aligned}
$$

A brief introduction to the concept of Desirability was given in [Ste99]. An overview and a nice comparison of Desirability and other approaches to MCO can be found in [HdSD92].

Within the wide range of possible DFs those of Derringer/Suich type can be considered as a *de facto* standard in real world applications.

What makes them standard is the flexibility in representing different schemes for weighting deviances from a target value. An experimenter can handle symmetric and asymmetric, target value and maximisation problems using the same class of functions.

The aim of this work is to show some shortcomings of the current way optimisations of DIs are performed and to remedy these by computer intensive procedures. Most important is the correction of the inherent bias for the estimated optimum factor level setting $\widehat{X_{opt}}$ in the classical optimisation procedure. This procedure up to now solves a simplified problem, therefore failing to identify the looked for solution. An improved procedure will be presented.

All simulations were done using the data analysis computer language R [Iha95].

In Section 2 the Derringer/Suich functions are introduced very briefly, together with the formalism needed in this paper. Section 3 describes the standard procedure for Desirability Index optimisation. An improved alternative is presented in Section 4. Section 5 gives an example of both procedures "at work" using a data "classic". The example shows the superior performance of the improved procedure presented here. A short conclusion with directions of further work is given in Section 6.

2 The Desirability Function of Derringer/Suich type

In 1980 Derringer and Suich defined their class of desirability functions in [DS80] as a more flexible alternative to Harrington type functions [Har65]. A detailed discussion of their differences can be found in [Ste99].

The formal definition of a DF of Derringer/Suich type for a target value problem is as follows:

$$d(Y_i) := \begin{cases} 0, & \text{for} \quad Y_i < LSL_i \\ \left(\frac{Y_i - LSL_i}{T_i - LSL_i}\right)^{\beta_{l,i}}, & \text{for} \quad LSL_i \le Y_i \le T_i \\ \left(\frac{USL_i - Y_i}{USL_i - T_i}\right)^{\beta_{r,i}}, & \text{for} \quad T_i < Y_i \le USL_i \\ 0, & \text{for} \quad USL_i < Y_i \end{cases}$$

The parameters LSL_i, USL_i, T_i give lower specification limit, upper specification limit and target value for response Y_i; parameters $\beta_{l,i}$ and $\beta_{r,i}$ are weights for deviations to the left respectively to the right from the target. Exponents near zero indicate unimportant deviations, whereas large exponents stand for very important targets. In Figure 1 characteristic desirability functions have been plotted, to show the flexibility of this function class.

Maximisation- (minimisation-) problems can be handled consistently within this function class using only the right (left) branch of a target value DF.

In the following a Derringer/Suich DF will be identified by its five parameters and called a DF of type $(LSL, T, USL, \beta_l, \beta r)$.

Maxi- and minimisation can be written as DFs of type $(LSL, T, \infty, \beta_l, 0)$ (maximisation) resp. type $(-\infty, T, USL, 0, \beta_r)$ (minimisation).

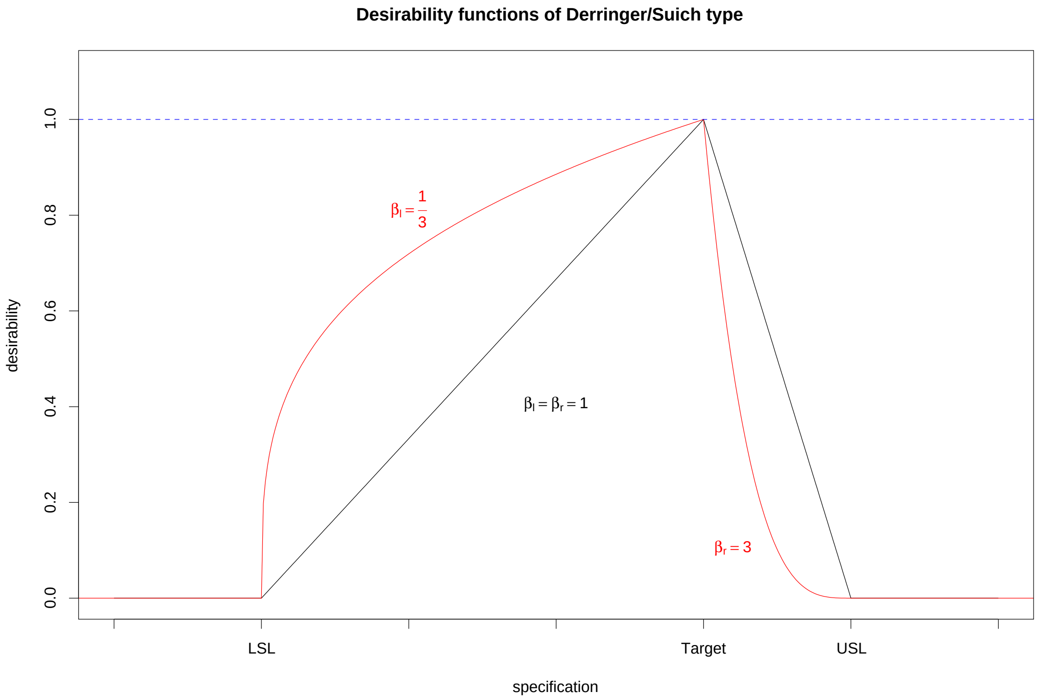
To combine these functions to a *Multi*-Criteria tool, Derringer/Suich use the geometric mean of multiple DFs as Desirability Index D of a multivariate response Y. The formula for $D(Y)$ gives the formal definition:

$$D(Y) := \sqrt[z]{\prod_{i=1}^{Z} d_i(Y_i)} \tag{5}$$

$$= \sqrt[z]{\prod_{i=1}^{Z} d_i(f_i(X_1, X_2, \ldots, X_F) + \epsilon_i)}. \tag{6}$$

Implementations of Derringer/Suich desirabilities are found in a number of software packages, mostly for experimental design. Two name just two

Figure 1: Desirability functions of Derringer/Suich type for two different values of β_l and β_r, asymmetric case.



there are Design Expert [Inc00] from Stat-Ease Inc. und STAVEX [AG00], developed at CIBA-GEIGY.

The following discussion is based on an experimental design. Each f_i represents a linear or quadratic relationship between factors $X_1, X_2, \ldots, X_F$ and $Y_i, i = 1, 2, \ldots, Z$. For each target Y_i a DF d_i of Derringer/Suich type is specified.

The aim is to find a factor level setting X_{opt} which gives the best possible response Y_{opt} measured on the DI scale.

In this final "calibration" step the difference between usage of DFs in observational studies and usage of DFs in experimental design can be found. While in observational studies the (multivariate) observations are ranked using desirabilities only, in experimental design optimisation an additional step is considered. The estimated functional relationship $\hat{f}$ is used to estimate the best setting for the controllable factors $\widehat{X_{opt}}$. The optimisation takes place in factor space, not in response space.

3 Standard optimisation procedure

Before finding X_{opt} the corresponding Y_{opt} has to be found which maximises $D(Y)$. It must be noted that neither Y_{opt} is necessarily unique, nor X_{opt}. Nevertheless for the paper presented here it is enough to find *one* Y_{opt}.

The procedure adapted in practise to find a solution first estimates the functions f_i using Least-Squares estimates $\hat{f}_i(X) = E(Y_i|X), i = 1, \dots, Z$.

As the optimisation takes place in factor space the definition of Desirability in response space has to be transferred to factor space in a proper way. In the classical procedure the following analogy is used to define a Desirability Index for factor level settings X:

$$D(X) := D(f(X)). \tag{7}$$

Obviously the error term in (1) was replaced by its expected value 0. Therefore DFs in factor space such defined are called *idealised* DFs.

For estimating the DI all estimated $\hat{f}_i$ are inserted into the user defined expression for the Desirability Functions and the Desirability Index:

$$\widehat{D^{ideal}}(X) := \sqrt[Z]{\prod_{i=1}^{Z} d_i(\hat{Y}_i)} \tag{8}$$

$$= \sqrt[Z]{\prod_{i=1}^{Z} d_i(\hat{f}_i(X_1, X_2, \dots, X_F)}. \tag{9}$$

This function D^{ideal} is maximised in factor space, resulting in an estimation $\widehat{X_{opt}^{ideal}}$ for the optimum factor level setting X_{opt}.

For optimisation practise it has to be noted that neither DFs nor DI functions are unimodal in general. Therefore gradient based optimisation techniques may only be used after further considerations.

3.1 Estimating a single *idealised* DF

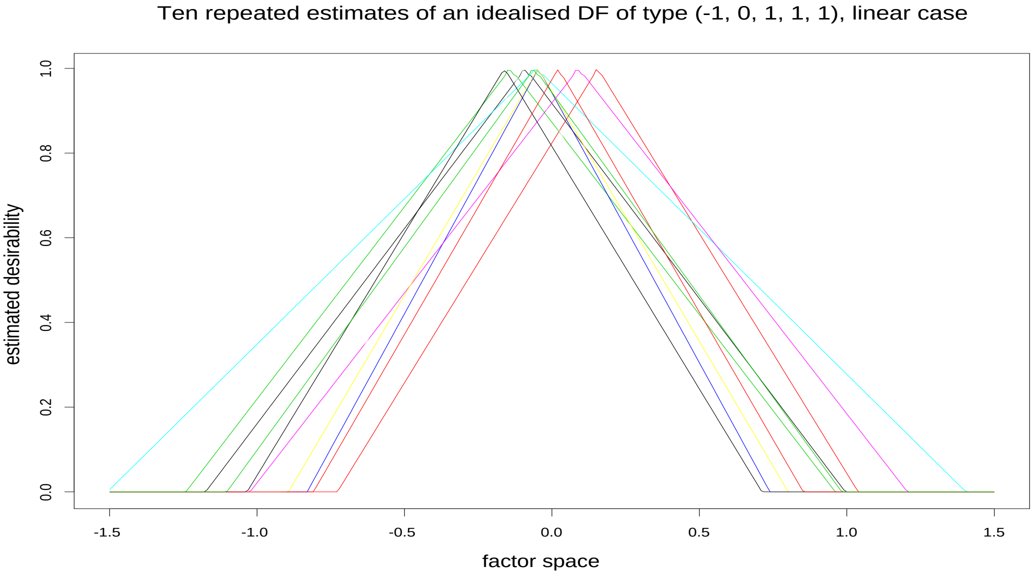
As seen above the estimation of Desirabilities essentially takes place by estimating the individual *idealised* DFs. In this section the implications of ignoring the error in the classical procedure will be investigated.

A simulation study was performed to give a first impression of estimating *idealised* DFs.

First the most simple case of estimating DFs is explored. The known "true" model was chosen as $Y_1 = f_1(X) + \epsilon = X + \epsilon, \epsilon \sim N(0, 0.1)$, with a type

(-1, 0, 1, 1, 1) DF d_1. Repeatedly five data points $Y_{1,i}$ for the design points $X_i \in \{-1, -0.5, 0, 0.5, 1\}, i = 1, 2, \ldots, 5$, were generated. These points were used to estimate *idealised* desirabilities for the linear model $\hat{f}_1 = \hat{\alpha} + \hat{\beta}x$. In Figure 2 ten estimated *idealised* desirabilities are shown.

Figure 2: Ten estimated *idealised* DFs, $Y = X + \epsilon \sim N(0, 0.1)$, d of type (-1, 0, 1, 1, 1), $\hat{f}_1 = \hat{\alpha} + \hat{\beta}x$.



Most important in Figure 2 (and in Figure 4) is the seeming precision the results are reported with. Each replication of the experiment gives a seemingly exact estimation of the *idealised* desirability. Especially no uncertainty may be assigned to an $\widehat{X_{opt}^{ideal}}$ after the calibration step.

Nevertheless $\widehat{X_{opt}^{ideal}}$ is a random variate. To get an idea about its distribution the estimation process was repeated 1000 times. A line search was performed to find $\widehat{X_{opt}^{ideal}}$ for each replication. Figure 3 shows a histogram for $\widehat{X_{opt}^{ideal}}$. Its distribution is symmetric about the known "true" x_{opt}^{ideal} zero. The calibration gives an unbiased result for a response assumed being error free.

In a second step analogous simulations were performed for the quadratic case.

The known "true" model was chosen as $Y_2 = f_2(X) + \epsilon = X^2 + \epsilon, \epsilon \sim N(0, 0.1)$, with a type (-1, 0, 1, 0.1, 1) DF d_2. The same design was used to generate data and subsequently estimate a linear model $\hat{f}_2 = \hat{\alpha} + \hat{\beta}x + \hat{\gamma}x^2$. In

Figure 3: Histogram of estimated $\widehat{X_{opt}^{ideal}}$, 1000 replications, linear case; for exact specifications see Figure 2

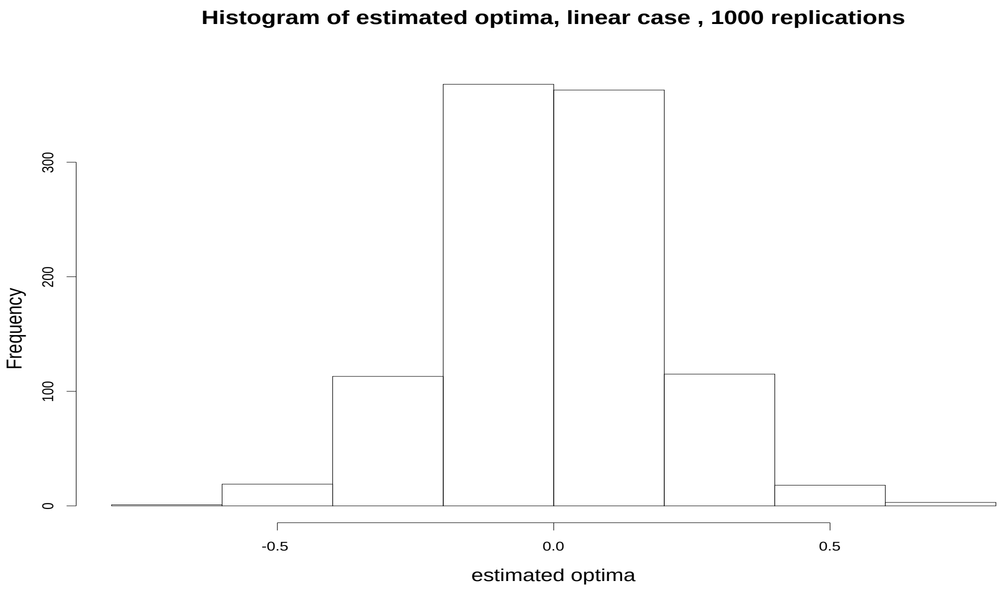


Figure 4 ten estimated *idealised* desirabilities for this simulation are shown.

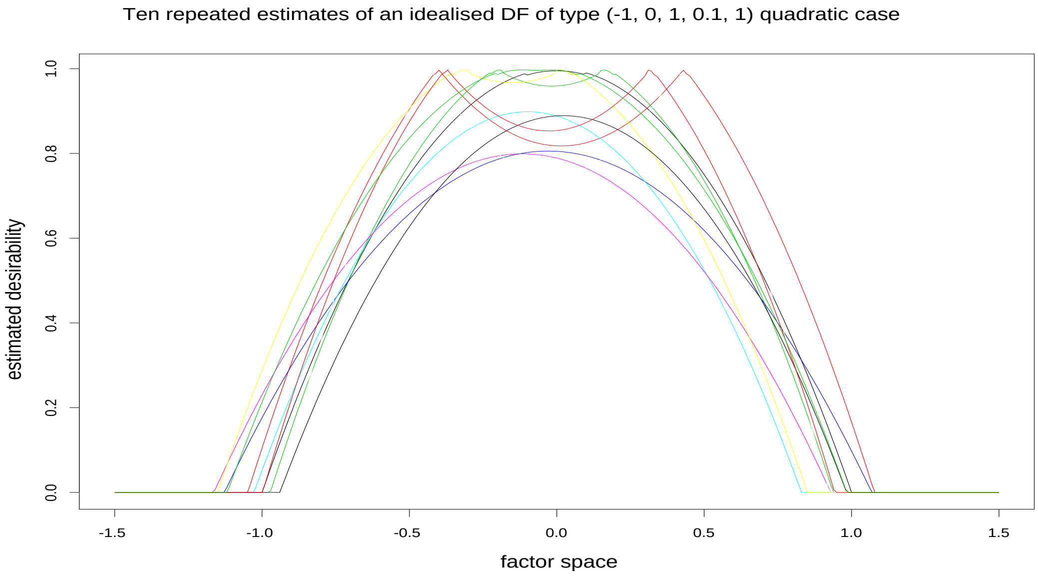
Figure 4 is somewhat more interesting than Figure 2. The specified example shows a target value problem modelled quadratically. The response reaches $T = 1$ exactly in $X_{opt}^{ideal} = 0$.

The estimated DFs fall into two classes. On one hand there are parabolas of the same type as the specification of the example. Their vertex gives a response below T. For this class $\widehat{X_{opt}^{ideal}}$ is distributed symmetrically around the known optimum factor level setting. On the other hand there are parabolas with maximum response higher than T. Those result in DFs with two local extrema in the points giving exactly a response T. Again the calibration was performed using line search. In case of multiple optima one of those was chosen at random.

As the distribution of $\widehat{X_{opt}^{ideal}}$ is a mixture of the two distributions for the two classes, tri-modality may be expected. Indeed Figure 5 shows this behaviour for 1000 replications of the estimation process.

To stress the point of these simulations: In an experimental design context there are no repetitions. The experimenter has no hint about the reliability of the estimated optimum using the standard procedures. One has to believe in the resulting factor level settings. Certainly the estimated errors $\hat{\sigma}_i$ for

Figure 4: Ten estimated *idealised* DFs, $Y = X^2 + \epsilon \sim N(0, 0.1)$, d of type (-1, 0, 1, 0.1, 1), $\hat{f}_2 = \hat{\alpha} + \hat{\beta}x + \hat{\gamma}x^2$.



the individual functions $f_i, i = 1, \ldots, Z$ are known, but it is not possible to extrapolate these to a confidence interval of any kind for $\widehat{X_{opt}^{ideal}}$.

The approach used in the standard optimisation procedure deliberately chooses to ignore knowledge about errors in the model.

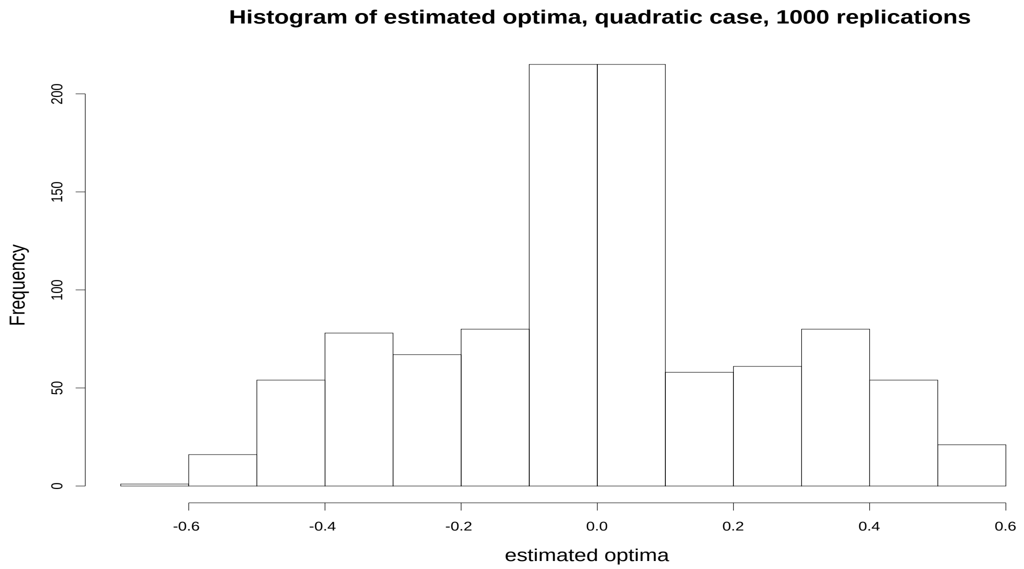
One final note on the analytical difficulties of the calibration step: Calibration is equivalent to estimating roots of polynomials with coefficients one knows the common distribution of. When for example considering the linear case with $Y_{opt} := T$ and $\hat{y} = \hat{\alpha} + \hat{\beta}x$ one finds $\widehat{x_{opt}} = \frac{T - \hat{\alpha}}{\hat{\beta}}$. Even for this linear case the general distribution of $\frac{\hat{\alpha}}{\hat{\beta}}$ is unknown, while the common distribution of $\hat{\alpha}$ and $\hat{\beta}$ is known.

4 Improved optimisation procedure

As seen above all information about the error is neglected in DI optimisation. This exactly is the point where a great improvement can be achieved.

Is an experimenter really looking for an optimum factor setting for $D(f(X))$? Or wouldn't it be much more sensible to take process quality, measured by the size of σ, into account and optimise $E(D(f(X) + \epsilon))$?

Figure 5: Histogram of estimated X_{opt}^{ideal}, 1000 replications, qudratic case; for exact specifications see Figure 4



In analogy to definition (1) another way to define Desirability in factor space obviously is:

$$D^{real}(X) := E(D(f(X) + \epsilon)). \tag{10}$$

Desirabilities defined this way are called *realistic* DFs (resp. DIs).

The maximisation problem to solve alters directly to

$$\max_X E(D(f(X) + \epsilon)). \tag{11}$$

Solutions of this problem are called X_{opt}^{real}.

The two approaches coincide for applications if σ_ϵ is very small, but will give different solutions if σ_ϵ is not negligible.

Furthermore DFs are not linear. Therefore using $d(\hat{f}(X))$ as an estimator for $d(f(X) + \epsilon)$ only is an approximative solution.

This choice of an approximative solution was natural at times when the necessary computing power was unavailable. Analytically the distributions of Desirability Function values or Desirability Index values for a specified X can not be handled efficiently. Using the computational power of today one can afford to approximate the distribution of DF and DI using Monte-Carlo Simulations. That way it becomes possible to optimise $E(D(f(x)+\epsilon))$ directly, using all information from the data, including size of errors!

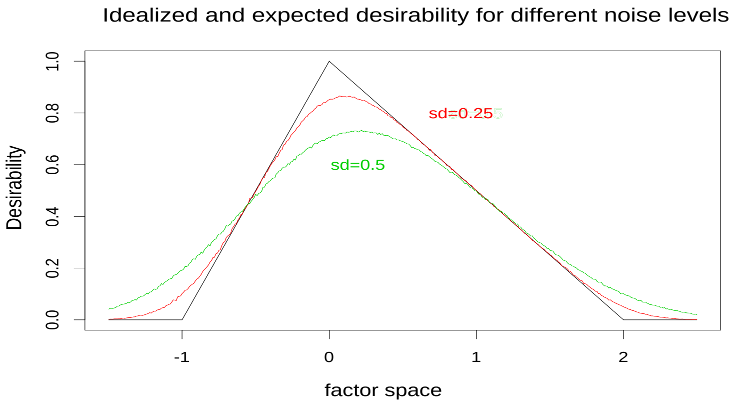
A short remark has to be made concerning the use of the geometric mean: While in the *idealised* approach it was deliberate to take the z-th root, it is obligatory now. Exchanging the expectation and the z-th root may alter the order of two factor level settings X^1, X^2.

4.1 Estimating a single *realistic* DF

In this section the implications of using a *realistic* DF for the estimating and calibrating process will be shown.

Figure 6 displays the difference between the two presented procedures for a simple example. The triangle function shows a typical idealised DF of type (-1, 0, 2, 1, 1) for a one dimensional Y. As constructed X_{opt}^{ideal} is found to be zero.

Figure 6: Idealised and realistic Desirability for a Type (-1,0,2,1,1) DF, 5000 replications , $Y = X + \epsilon, \sigma\epsilon = 0.25, 0.5$



The other curves show estimated *realistic* DFs for different values of $\sigma_\epsilon = 0.25$ and $\sigma_\epsilon = 0.5$. The value for each point X is calculated performing a Monte-Carlo simulation approximatinging the mean desirability response for each point.

To get a smooth curve a relatively high number of repetitions is necessary in this example ($n = 5000$). The number of repetitions is an important aspect, as it determines the computing time needed by this algorithm. This aspect gets more important as the dimension of factor space increases. The example in Section 5 shows a way to determine a reasonable sample size from the data. Optimisation is performed via grid-search along the x-axis.

What can be seen from Figure 6? Obviously the curves from *realistic* DFs are smoother than from *idealised* DFs. If an experimenter is sure about unimodality of the Desirability Index function gradient based methods may fearlessly be applied. The most important difference is the direct impact the size of σ_ϵ has for the absolute value of DF. If σ_ϵ increases the maximum desirability decreases, while at the same time the area with non-zero desirability is enlarged. This behaviour reflects very nicely the trust an experimenter should have in the process. If there is a high noise level in the data one can not be too sure about the performance in an estimated optimum factor level setting. At the same time one can not be too sure about a real bad performance at some design point.

The most important point however is the difference of the estimations for X_{opt}^{ideal} and X_{opt}^{real}.

It is seen that the realistic optimum X_{opt}^{real} slowly drifts away from X_{opt}^{ideal} when the noise level is increased. The direction of drift is a consequence of the asymmetric specification of the DF, giving a deviation to the right less weight. Table 1 summarises some of the simulation results. In columns three and four the expected desirabilities of the estimated optima are compared. The improvement in Desirability by using *realistic* DFs is obvious. The high number of repetitions (10000) was necessary to make the difference between columns three and four for $\sigma_\epsilon = 0.25$ significant.

This constructed example shows the principle only. The advantage of using *realistic* DFs does not look too impressive, but in the next chapter an example with real data will be given, which shows the worthwhileness of such efforts.

Table 1: Results for the example shown Figure 6

σ_ϵ	$\widehat{X_{opt}^{real}}$	$E(D(\widehat{X_{opt}^{ideal}}))$	$E(D(\widehat{X_{opt}^{real}}))$	repetitions
0.25	0.11	0.85	0.86	10000
0.5	0.25	0.70	0.73	10000

5 Realistic Desirabilities at work

As the classic Derringer/Suich paper [DS80] gives a nice example for optimising a Desirability Index, it is an obvious choice for trying to improve on their result.

In their paper they use a set of chemical data to apply their DFs to. They have four targets Y_1 to Y_4 and three controllable variables X_1 to X_3. The data were generated using a central-composite design with 20 experiments, to fit a second-order model $\hat{f}_i$ including all interactions. They got the following estimates:

$$\begin{aligned}
\hat{f}_1(X) &= 139.1 + 16.5X_1 + 17.9X_2 + 10.9X_3 - 4.0X_1^2 - 3.5X_2^2 - 1.6X_3^2 \\
&\quad +5.1X_1X_2 + 7.1X_1X_3 + 7.9X_2X_3; \quad \sigma_1 = 5.6, \\
\hat{f}_2(X) &= 1261.1 + 268.2X_1 + 246.5X_2 + 139.5X_3 - 83.6X_1^2 - 124.8X_2^2 \\
&\quad +199.2X_3^2 + 69.4X_1X_2 + 94.1X_1X_3 + 104.4X_2X3; \quad \sigma_2 = 328.7, \\
\hat{f}_3(X) &= 400.4 - 99.7X_1 - 31.4X_2 - 73.9X_3 + 7.9X_1^2 + 17.3X_2^2 + 0.4X_3^2 \\
&\quad +8.8X_1X_2 + 6.3X_1X_3 + 1.3X_2X_3; \quad \sigma_3 = 20.6, \\
\hat{f}_4(X) &= 68.9 - 1.4X_1 + 4.3X_2 + 1.6X_3 + 1.6X_1^2 + 0.1X_2^2 - 0.3X_3^2 \\
&\quad -1.6X_1X_2 + 0.1X_1X_3 - 0.3X_2X_3; \quad \sigma_4 = 1.27.
\end{aligned}$$

(More details are found in [Ste99]).

After performing the optimisation for *idealised* DFs, an optimal factor level setting $\widehat{X_{opt}^{ideal}} = (-0.05, 0.145, -0.868)$ was identified by Derringer/Suich. In $\widehat{X_{opt}^{ideal}}$ the *idealised* DI is 0.58.

If instead *realistic* DFs are used and the estimated standard deviations $\sigma_i, i = 1, 2, 3, 4$ are taken into account a numerical optimisation (Nelder-Mead) gives $\widehat{X_{opt}^{real}} = (0.13, 0.50, -1.08)$. To get a figure for the sample size needed to get a desired precision a simulaton pre-study was performed. Assuming similarity between the distributions of $D(\widehat{X_{opt}^{ideal}})$ and $D(\widehat{X_{opt}^{real}})$ a Monte-Carlo simulation for estimating the variance of the DI in $\widehat{X_{opt}^{ideal}}$ was performed. This simulation gave $\widehat{Var}(D(\widehat{X_{opt}^{ideal}})) \approx 0.23$.

Using the Central-Limit-Theorem is is easily seen that a sample size of ca. 4800 at each point in factor space is needed to get a standard deviation lower than $\frac{1}{300}$ for the DI near the optimum. Using that sample size gains two usable digits in the estimated DI using 99%-confidence intervals.

Table 2: Estimated optima and estimated DIs; 4800 replications in each factor setting X to reduce the standard deviation of the estimates near the optimum to approximately $\frac{1}{300}$.

	$\widehat{X_{opt}^{real}}$	$\widehat{X_{opt}^{ideal}}$
D^{real}	0.44	0.40
D^{ideal}	0.56	0.58

The idealised DI of $\widehat{X_{opt}^{real}}$ is 0.56, only slightly worse than for $\widehat{X_{opt}^{ideal}}$. Comparing the *realistic* DIs for these two settings gives a promising relative improvement of 10% from 0.4 to 0.44 (see Table 2). A 10% percent improvement in desirability performance certainly is a strong argument towards applying the realistic DFs.

6 Conclusion

The computing power easily available to the statistician nowadays allows to tackle problems which are hard or impossible to solve analytically. Sometimes approximative solutions may be discarded in favour of computer intensive, "exact" procedures. The optimisation of Desirability Indices is one of these.

Only now it is possible to get an answer to the right question! Ignoring the error terms while at the same time having estimated them is insensible. Monte-Carlo simulations give reasonable exact solutions to be used in practical applications. While in this paper normality and independence is assumed for the error, this is not necessary in principle. The procedure does not depend on these assumptions.

Not considered in this paper is the distribution of the realistic Desirability Index itself. Using Mote-Carlo simulations the statistician gets estimates for the whole distribution function. This knowledge might be used to refine and/or flexibilise the optimisation goal.

References

[AG00] AICOS Technologies AG. *STAVEX*. http://www.aicos.com, 4.3 edition, 1997-2000. Software Referenz.

[DS80] G. Derringer and R. Suich. Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12(4):214–219, 1980.

[Har65] E.C. Harrington Jr. The desirability function. *Industrial Quality Control*, 21(10):494–498, 1965.

[HdSD92] M.M.W.B. Hendriks, J. H. de Boer, A.K. Smilde, and D.A. Doornbos. Multicriteria decision making. *Chemometrics and Intelligent Laboratory Sytems*, (16):175–191, 1992.

[Iha95] Gentleman Ihaka. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, (5):299–314, 1995.

[Inc00] Stat-Ease Inc. *Design Expert*. http://www.statease.com, 6 edition, 2000. Software Referenz.

[Ste99] Detlef Steuer. Multi-criteria-optimisation and desirability indices. Technical Report 20, Universität Dortmund, FB Statistik, Vogelpothsweg 87, 44221 Dortmund, 1999.