Statistical properties of MCO using desirabilities

Detlef Steuer
Helmut-Schmidt-Universität
Universität der Bundeswehr Hamburg

steuer@hsu-hh.de

http://fawn.unibw-hamburg.de/steuer.html

Bommerholz, 1.4.2005

HELMUT SCHMIDT
UNIVERSITÄT
Universität der Bundeswehr Hamburg

Contents

The MCO-problem

- Example: design of a new cookie.
- Targets: not too hard to eat, not too crumbly for transport.
  1. Problem: antagonists must be optimised simultaneously.
  2. Problem: different units must be compared.
- Natural aim: optimise **quality** over all targets!
- Need to compromise, expert knowledge necessary.
- Generally: find best object $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$. Each component y_i is of type (TV) or (LB).

Some approaches to MCO (besides desirabilities)

- Mathematical approaches
  - Dominance: assumes existence of a coordinate-wise best object.
  - Pareto-optimality: an object is optimum, if it can not be improved in all directions.
- Approaches in statistics and OR
  - Graphical procedures, i.e. *overlay plots*.
  - Utility functions, see i.e. Jessenberger.
  - Outranking procedures, i.e. Prometee.

General requirements to MCO procedures

- **Pareto-optimality:** a proposed solution must be pareto-optimum, else you can hardly talk of a 'optimum'.
- **Scalability:** it should be possible to assess the usability of a procedure for high dimensions.
- **'Practicability':** possibility to incorporate expert knowlegde and simple interpretation of results.

Here: desirabilities

- Introduced by Harrington 1965.
- Idea: transform all quality measures on a unitless [0, 1] scale.
- 'Oranges and apples' now can be compared. Especially: a mean value of the different quality measures can be calculated.
- Transformations of the targets are defined using expert knowlegde, the averaging allows the usage of weights for the different targets.
- Details: see Trautmann.

Derringer-Suich desirability functions

Derringer-Suich (1980) improved on Harrington by using a more flexible class of functions. Their approach is quasi-standard (NIST).

DS-desirabilites for (TV)

$$d_{DS}^{TV}(y) := \begin{cases} 0, & \text{for} \quad y < l \\ \left(\frac{y-l}{t-l}\right)^{\beta_l}, & \text{for} \quad l \leq y \leq t \\ \left(\frac{u-y}{u-t}\right)^{\beta_r}, & \text{for} \quad t < z \leq u \\ 0, & \text{for} \quad u < y \end{cases}$$

DS-desirabilities for (LB)

$$d_{DS}^{LB}(y) := \begin{cases} 0, & \text{for} \quad y < l \\ \left(\frac{y-l}{t-l}\right)^{\beta_l}, & \text{for} \quad l \leq y \leq t \\ 1, & \text{for} \quad t < y \end{cases}$$

Unified notation

A Quintupel $(l, t, u, \beta_l, \beta_r)$ defines a DS-desirability:

$$\text{If } l < t < u \in \mathbb{R}, \beta_l, \beta_r \in \mathbb{R}^+, \quad \text{then} \quad (l, t, u, \beta_l, \beta_r)(y) := d_{DS}^{TV}(y).$$

$$\text{If } u = \infty, \text{then also } \beta_r = 1 \quad \text{and} \quad (l, t, \infty, \beta_l, 1)(y) := d_{DS}^{LB}(y).$$

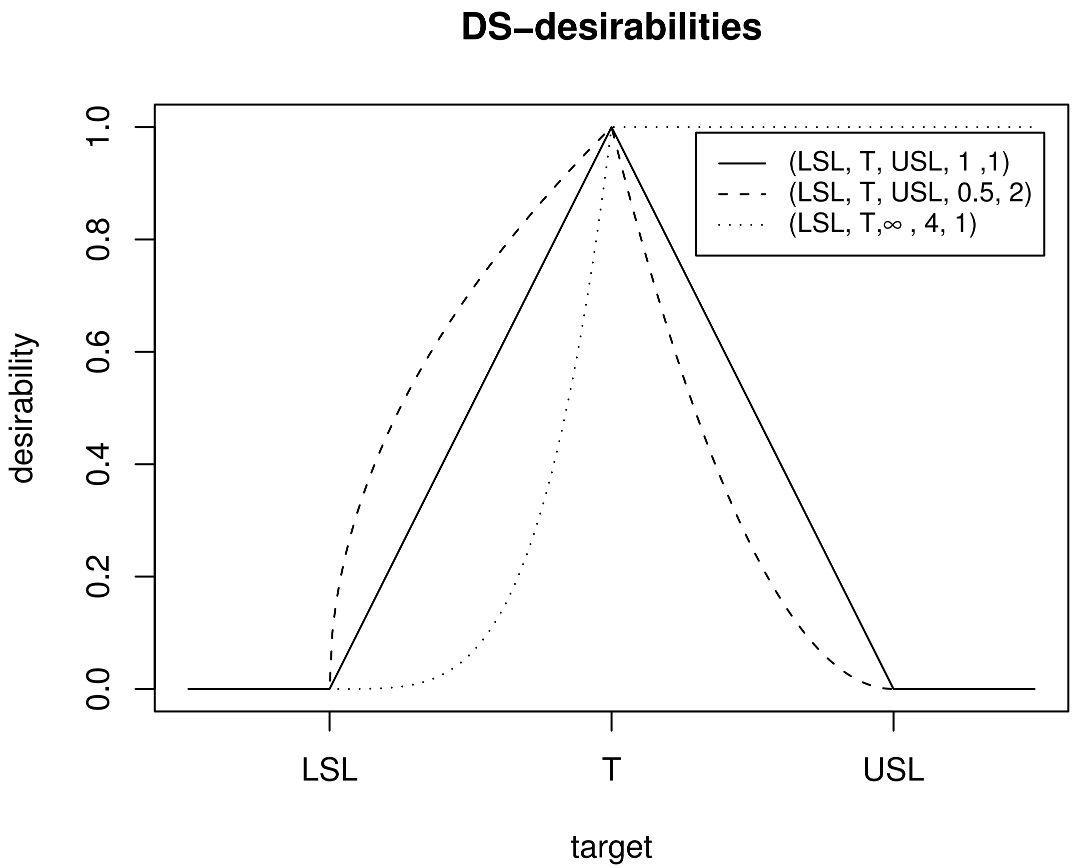
Desirability Index

Usage of the mean to assure comparability of results.

$$\text{(geometric mean) } q(y) := \left(\prod_{i=1}^{m} d_i(y_i) \right)^{\frac{1}{m}} \quad \text{or}$$

$$\text{(maximin) } q(y) := \min_{i=1,\ldots,m} d_i(y_i)$$

DS-desirabilities for some parameter settings



Generalised DS-desirabilities

Definition: Given an increasingly ordered set of pairwise different nodes $y_i \in \mathbb{R}$, i=0,. . . ,n, a set of desirability values $d_i \in [0,1], i = 0, \ldots, n$, for nodes y_i, and a set of weights $\beta_i \in \mathbb{R}^+, i = 1, \ldots, n$, one for each interval $[y_{i-1}, y_i]$. Additionally $y_n = \infty$ together with $d_n = \beta_n = 1$ is allowed for the (LB) case. Then a function with

$$d_{DS}^v(y) := \begin{cases} 0, & \text{for } y < y_0 \\ d_{i-1} + (d_i - d_{i-1})(\frac{y-y_{i-1}}{y_i-y_{i-1}})^{\beta_i}, & \text{for } y \in [y_{i-1}, y_i] \text{ and } d_{i-1} \leq d_i \\ d_i + (d_{i-1} - d_i)(\frac{y-y_i}{y_{i-1}-y_i})^{\beta_i}, & \text{for } y \in [y_{i-1}, y_i] \text{ and } d_i < d_{i-1} \\ 0, & \text{for } y > y_n \end{cases}$$

is called „generalised Derringer-Suich desirability“.

Generalised Derringer-Suich desirabilities

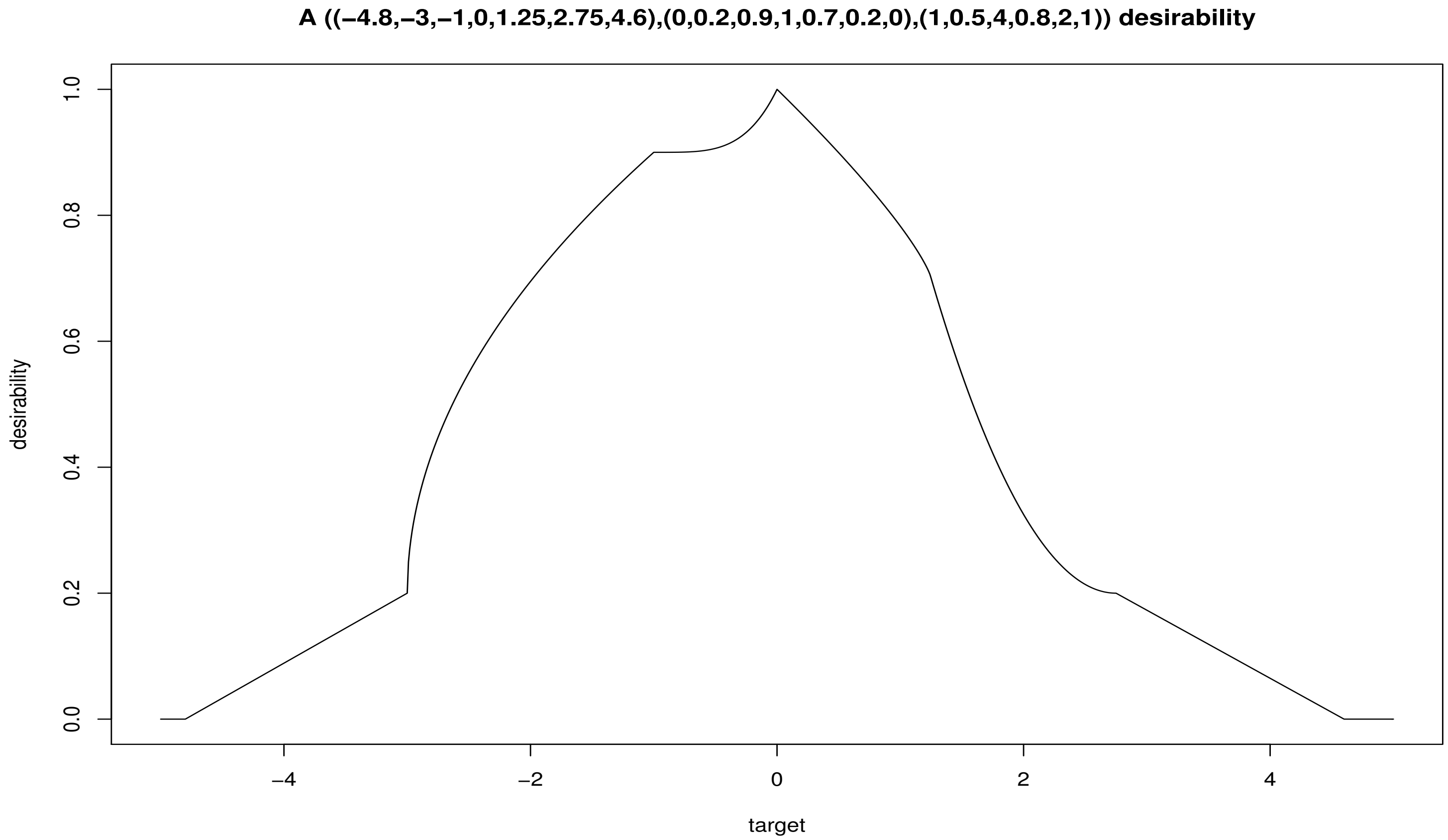
Notation:
Function and the list of parameters can be identified:

$$d_{DS}^{v}(y) =: ((y_0, \ldots, y_n), (d_0, \ldots, d_n), (\beta_1, \ldots, \beta_n)).$$

Special case:
Normal DS-desirabilities are special cases of the generalised ones:

$$(l, t, u, \beta_l, \beta_r) = ((l, t, u), (0, 1, 0), (\beta_l, \beta_r)) .$$

Example for a generalised DS-desirability

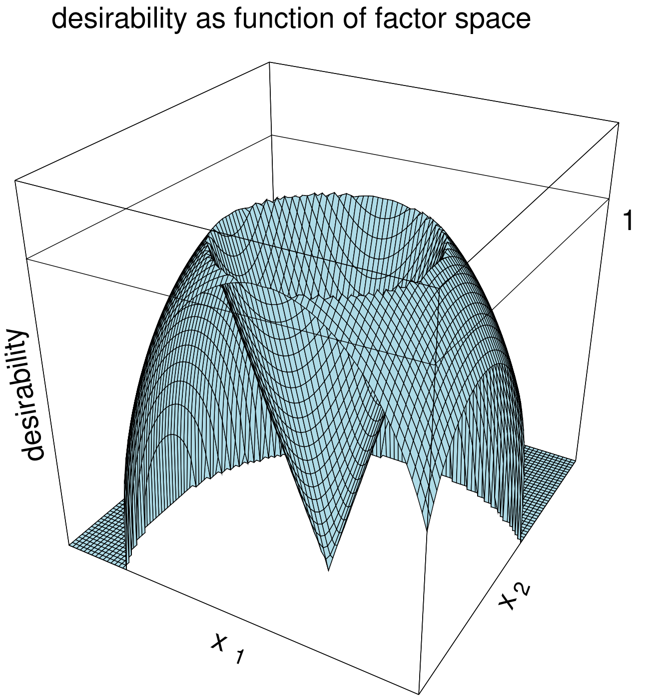


Analysis of common practice

- To optimise $Y = (Y_1, \ldots, Y_m)$, Y depends on factor settins $x = (x_1, \ldots, x_k)$, $Y = f(x) + \epsilon$, ϵ multivariate normal with diagonale covariance.
- Define desirabilities $d_i(y_i)$ for individual components.
- Perform experiments according to a DOE.
- Fit linear or quadratic response surfaces $\hat{f}_i, i = 1, \ldots, m$ for the components.
- Perform numerical optimisation for $q(\widehat{f(x)})$ over the region of operability $\mathbb{O}$ to estimate the optimum factor settings $\widehat{x_{opt}}$. (Calibration!)

$q(y)$ may have complicated structure

Model: $Y = x_1^2 + x_2^2, \quad d = (-1, 0, 4, 1/2, 1/2)$



Is current practice adequate for the MCO-problem?

No!

Common practice ignores the error terms and the non-linearity of the desirabilities.

Today's practice:

$$\widehat{x_{opt}} := \max_{x \in \mathbb{O}} q(E(Y|x)).$$

Simplified and in general **wrong** Solution!

Actually looking for:

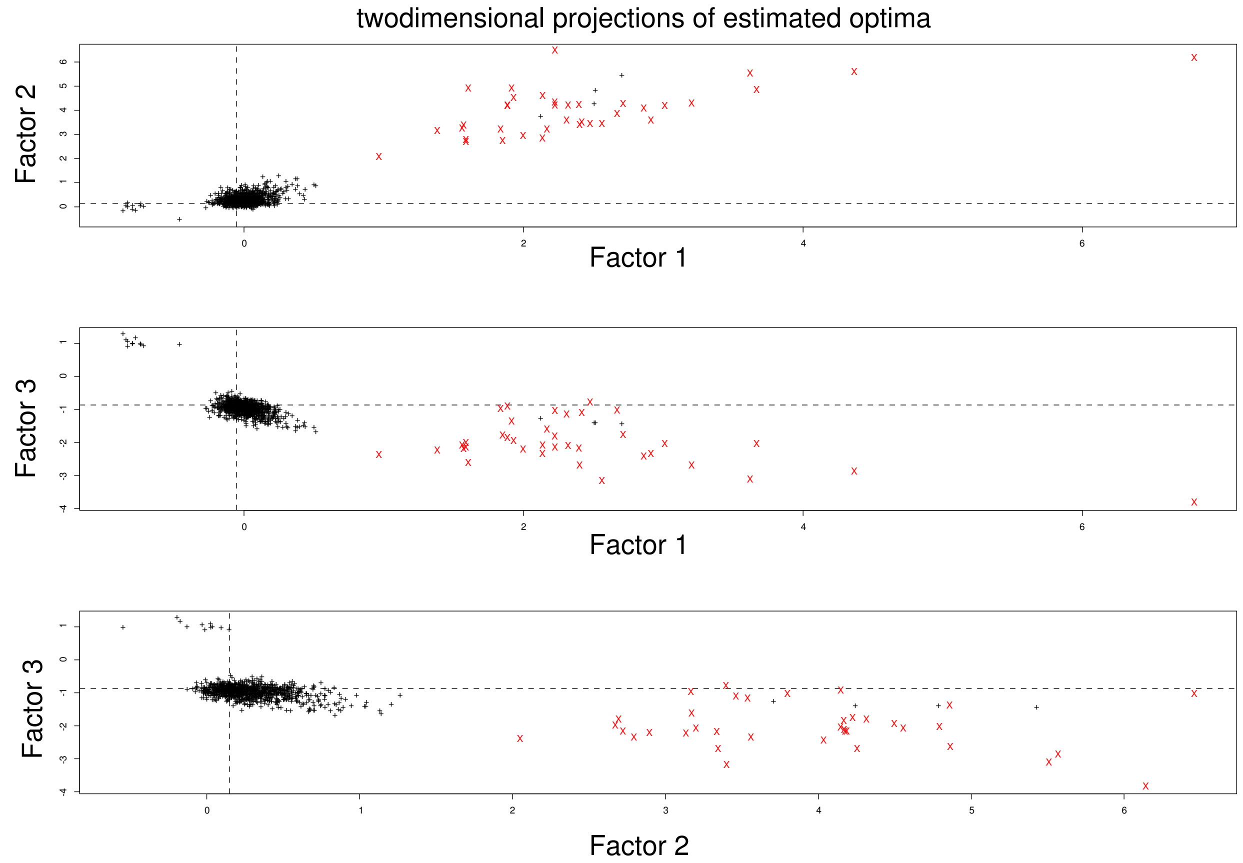
$$\widehat{x_{opt}} := \max_{x \in \mathbb{O}} E(q(Y|x))!$$

What are the costs of ignorance?

Simulation: repeat optimisation of Derringer-Suich.

- Four targets y_1 bis y_4, two of type (TV), two of type (LB), three factors x_1 bis x_3.
- Central-composite design with 20 experiments.
- Quadratic models $\hat{f}_i$ including all interactions.
- Repeated data generation according to the estimated model.
- Find $\widehat{x_{opt}}$ for each repetition. What can be found?

What are the costs of ingnorance?



Around 5% of the estimated $\widehat{x_{opt}}$ have true desirability **0**!

Incorporation of error terms
Desirability as random variate

Each d_i and q are random variates! The problem has turned into ordering of random variates.

First concentrate on the expected value as best analogy to the classical approach.

Model: $Y = f(x) + \epsilon, \epsilon \sim N(0, \sigma^2)$.

Random desirability defined as:

$$d(x, \epsilon) = \begin{cases} \dfrac{f(x) + \epsilon - l}{t - l} & \text{for } l \leq f(x) + \epsilon < t; \\ \dfrac{u - f(x) - \epsilon}{u - t} & \text{for } t \leq f(x) + \epsilon < u; \\ 0 & \text{else.} \end{cases}$$

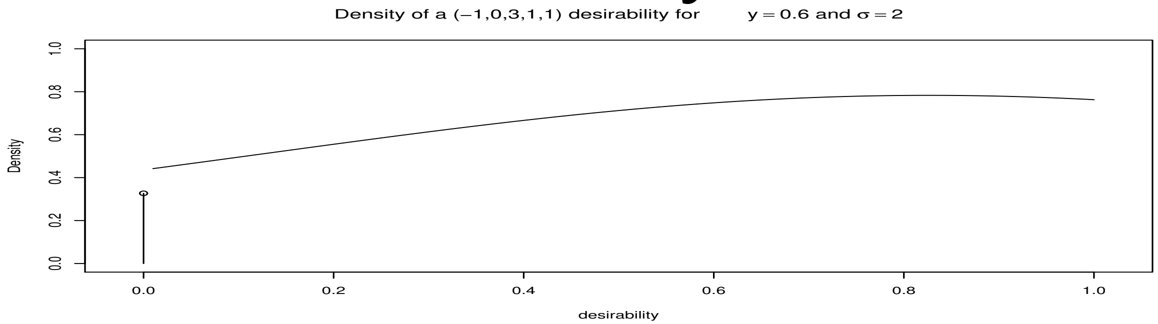
Distribution $F_{d(x,\epsilon)}$ for d of type $(l, t, u, 1, 1)$ for fixed x

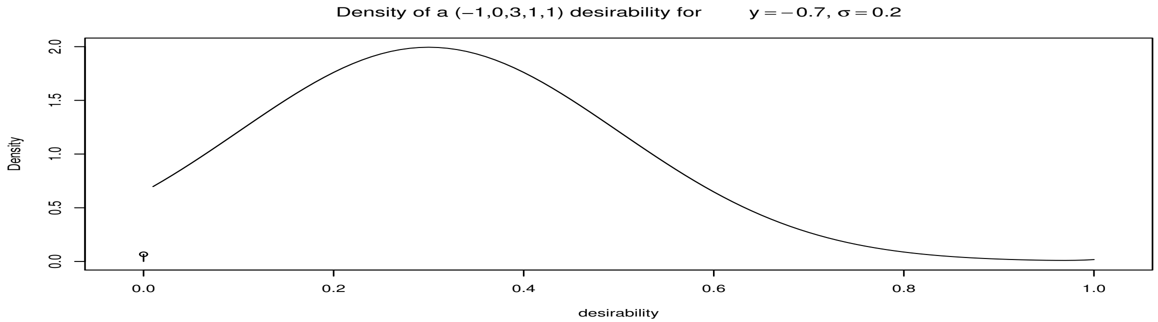
$$F_{d(x,\epsilon)}(d) = \begin{cases} 0 & \text{for } d < 0; \\ \Phi\left(\dfrac{l + d \cdot (t - l) - f(x)}{\sigma}\right) + & \\ \quad 1 - \Phi(\dfrac{u - d \cdot (u - t) - f(x)}{\sigma}) & \text{for } d \in [0, 1]; \\ 1 & \text{for } d > 1. \end{cases}$$

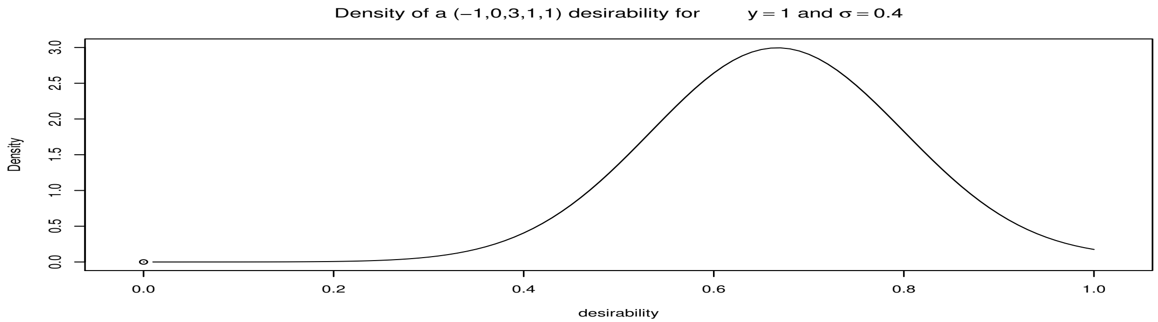
Density $f_{d(x,\epsilon)}$ for d of type $(l, t, r, 1, 1)$ for fixed x

$$f_{d(x,\epsilon)}(d) = \begin{cases} 0 & \text{for } d \notin [0, 1); \\ \Phi(\frac{l - f(x)}{\sigma}) + 1 - \Phi(\frac{u - f(x)}{\sigma}) & \text{for } d = 0 \\ & \text{(failure rate)}; \\ \frac{t-l}{\sigma}\varphi(\frac{l + d\cdot(t-l) - f(x)}{\sigma}) + \frac{u-t}{\sigma}\varphi(\frac{u - d\cdot(u-t) - f(x)}{\sigma}) & \text{for } d \in (0, 1). \end{cases}$$

Some desirability densities







Expected value for fixed x and d of type $(l, t, r, 1, 1)$

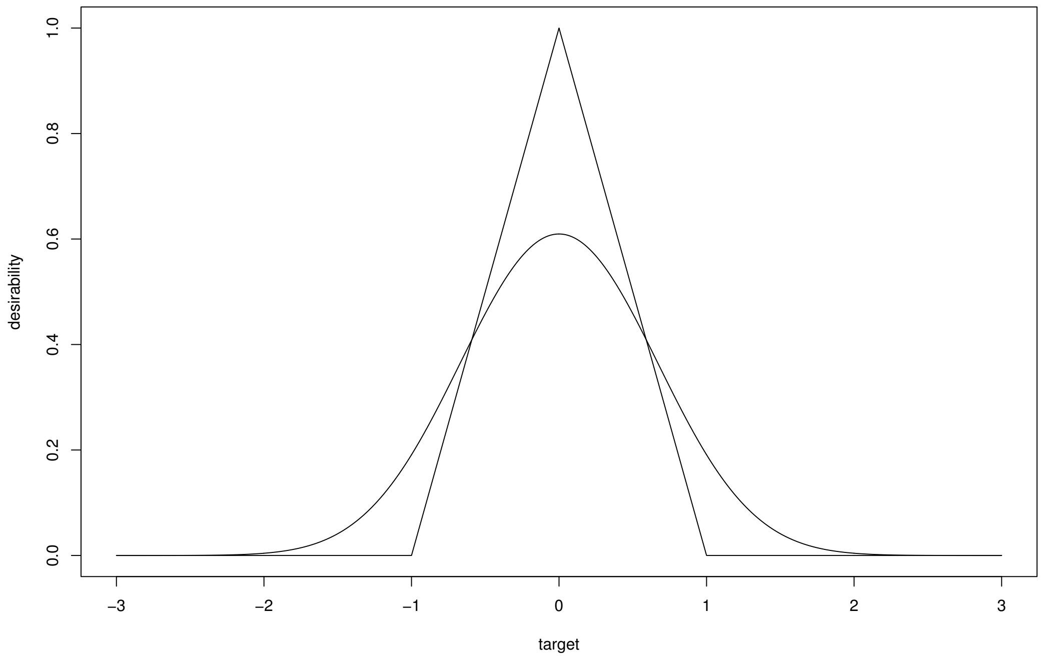
$$E(d(x,\epsilon)) = G_l \frac{f(x)-l}{t-l} + g_l \frac{\sigma}{t-l} + G_r \frac{u-f(x)}{u-t} - g_r \frac{\sigma}{u-t} \text{ with}$$

$$\begin{aligned}
G_l &= \Phi\left(\frac{t-f(x)}{\sigma}\right) - \Phi\left(\frac{l-f(x)}{\sigma}\right), \\
g_l &= \varphi\left(\frac{l-f(x)}{\sigma}\right) - \varphi\left(\frac{t-f(x)}{\sigma}\right), \\
G_r &= \Phi\left(\frac{u-f(x)}{\sigma}\right) - \Phi\left(\frac{t-f(x)}{\sigma}\right), \\
g_r &= \varphi\left(\frac{t-f(x)}{\sigma}\right) - \varphi\left(\frac{u-f(x)}{\sigma}\right).
\end{aligned}$$

Realistic desirabilities

Define the **realistic** desirability for fixed x as $E(d_i(x, \epsilon))$.

Example: $d = (-1, 0, 1, 1, 1)$ and $\sigma = 0.5$, idealised and realistic desirability.



Is it worth the effort?

- Optimisation using **realistic** desirabilities $E(d_i(x, \epsilon))$ of the realistic index $q^{real}(y) = \left(\prod_1^m E(d_i(x, \epsilon)) \right)^{\frac{1}{m}}$.

- Repeat again simulation of Derringer-Suich, but using q^{real}.

- Result for realistic desirabilities:
  Different estimations for $\widehat{x_{opt}}$:
  $\widehat{x_{opt}^{ideal}} = (-0.05, 0.145, -0.868)$ (Derringer-Suich),
  $\widehat{x_{opt}^{real}} = (0.13, 0.50, -1.08)$ (realistic desirabilities).

- **Better! True** 10% relative improvement for the values of q:

$$q(\widehat{x_{opt}^{real}}) = 0.44 \quad \text{and} \quad q(\widehat{x_{opt}^{ideal}}) = 0.40.$$

Conclusion (so far)

- For usual DS-desirabilities with linear weights the exact distribution of a desirability for fixed x and known σ can be calculated.

- Improvements can be expected, if using realistic desirabilities.

- Up to here: results only for exponents $\beta_l, \beta_r = 1$ possible.

- Restricted on the expected value, actually a new MCO over important properties of the resp. distribution necessary (median, failure rate, quantiles etc.).

How to handle weights $\beta_l, \beta_r \neq 1$?

Idea: Use specially constructed generalised desirabilities with linear weights to approximate normal DS-desirabilities with non-linear weights.

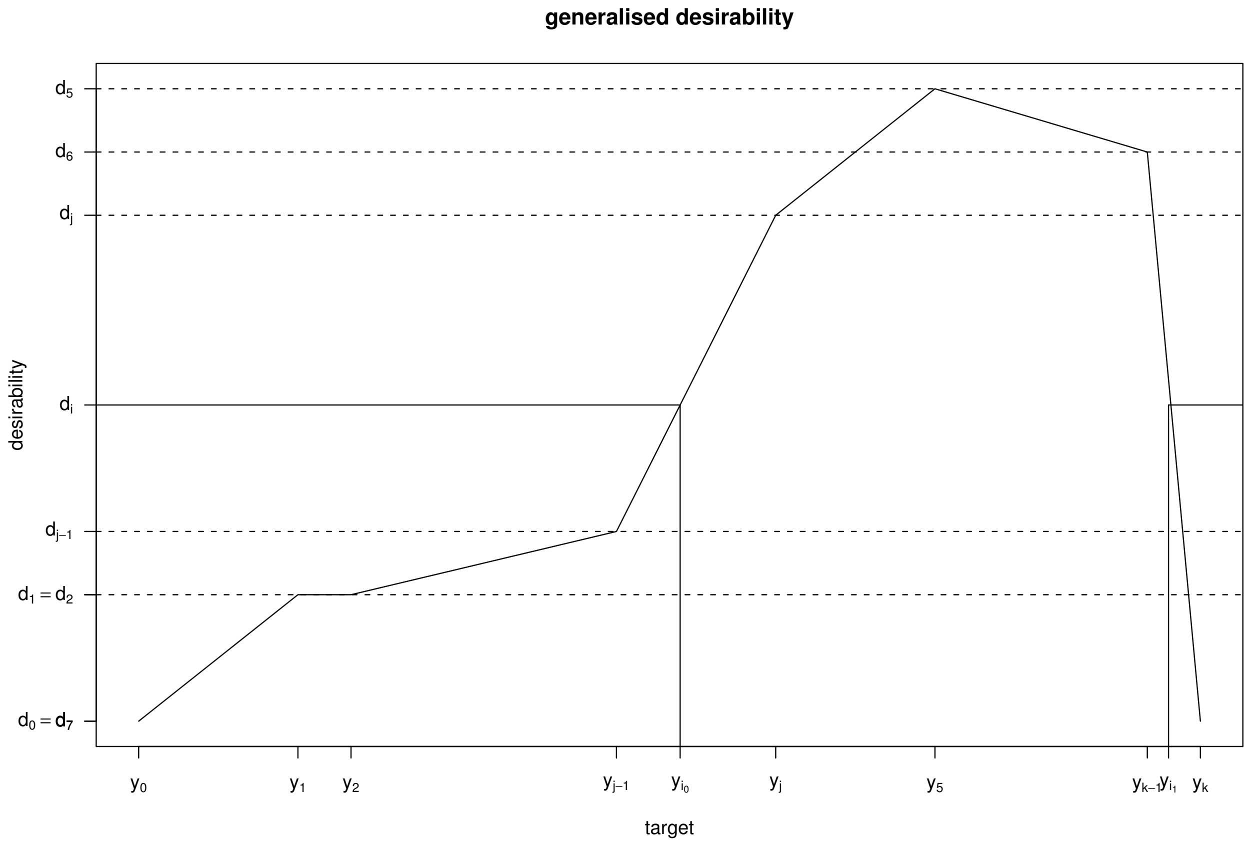
For these generalised desirabilities the distributions can be given explicitly, too!

Distribution of a generalised desirability

Let $d := ((y_0, \dots, y_n), (d_0, \dots, d_n), (\beta_1, \dots, \beta_n))$ with $\beta_i = 1$ for all i, then:

$$F_{d(Y)}(d) = \begin{cases} 0 & \text{for } d < 0; \\ \Phi\left(\frac{y_0 - f(x)}{\sigma}\right) + 1 - \Phi\left(\frac{y_n - f(x)}{\sigma}\right) + \\ \quad \sum_{i: d_i, d_{i+1} \le d} \left[\Phi\left(\frac{y_{i+1} - f(x)}{\sigma}\right) - \Phi\left(\frac{y_i - f(x)}{\sigma}\right)\right] + \\ \quad \sum_{i: d_i < d \le d_{i+1}} \left[\Phi\left(\frac{y_i + \frac{d - d_i}{d_{i+1} - d_i}(y_{i+1} - y_i) - f(x)}{\sigma}\right) - \Phi\left(\frac{y_i - f(x)}{\sigma}\right)\right] + \\ \quad \sum_{i: d_{i+1} < d \le d_i} \left[\Phi\left(\frac{y_{i+1} - f(x)}{\sigma}\right) - \Phi\left(\frac{y_{i+1} - \frac{d - d_{i+1}}{d_i - d_{i+1}}(y_{i+1} - y_i) - f(x)}{\sigma}\right)\right] \\ & \text{for } 0 \le d \le 1; \\ 1 & \text{for } d > 1. \end{cases}$$

Outline of a proof



Expected value of a generalised desirability

$$E(d^v(y)) = \sum_{i:d_{i-1}=d_i} d_i \cdot \left[\Phi_{f(x),\sigma^2}(y_i) - \Phi_{f(x),\sigma^2}(y_{i-1})\right]$$

$$+ \sum_{j:d_{j-1}<d_j} \left\{ \left(d_{j-1} + \frac{d_j - d_{j-1}}{y_j - y_{j-1}} \left[f(x) + \sigma^2 \frac{\varphi_{f(x),\sigma^2}(y_{j-1}) - \varphi_{f(x),\sigma^2}(y_j)}{\Phi_{f(x),\sigma^2}(y_j) - \Phi_{f(x),\sigma^2}(y_{j-1})} - y_{j-1} \right] \right) \times \left[\Phi_{f(x),\sigma^2}(y_j) - \Phi_{f(x),\sigma^2}(y_{j-1})\right] \right\}$$

$$+ \sum_{k:d_{k-1}>d_k} \left\{ \left(d_k + \frac{d_{k-1} - d_k}{y_k - y_{k-1}} \left[f(x) + \sigma^2 \frac{\varphi_{f(x),\sigma^2}(y_{k-1}) - \varphi_{f(x),\sigma^2}(y_k)}{\Phi_{f(x),\sigma^2}(y_k) - \Phi_{f(x),\sigma^2}(y_{k-1})} + y_k \right] \right) \times \left[\Phi_{f(x),\sigma^2}(y_k) - \Phi_{f(x),\sigma^2}(y_{k-1})\right] \right\}$$

1st idea: once linearised approximation

Given a (LB) desirability $(l, t, \infty, \beta_l, 1)$. Define

$$(l, t, \infty, \beta_l, 1)_{lin} := \begin{cases} (l, t, \infty, 1, 1) & \text{for } \beta_l = 1; \\ (l + c^*, t, \infty, 1, 1) & \text{for } \beta_l > 1; \\ (l, t - c^{**}, \infty, 1, 1) & \text{for } \beta_l < 1. \end{cases}$$

c^* and c^{**} chosen in a way, that:

$$c^*(resp. c^{**}) \text{ solves } \min_{c^*(\text{resp.} c^{**}) \in [0, t-l]} \int_l^t ((l, t, \infty, \beta_l, 1) - (l, t, \infty, \beta_l, 1)_{lin}))^2 dx.$$

$(l, t, \infty, \beta_l, 1)_{lin}$ is called *once linearised (LB) desirability*.

1st idea: once linearised approximation

(TV)-problems $(l, t, u, \beta_l, \beta_r)$ are partitioned in two (LB)-problems:

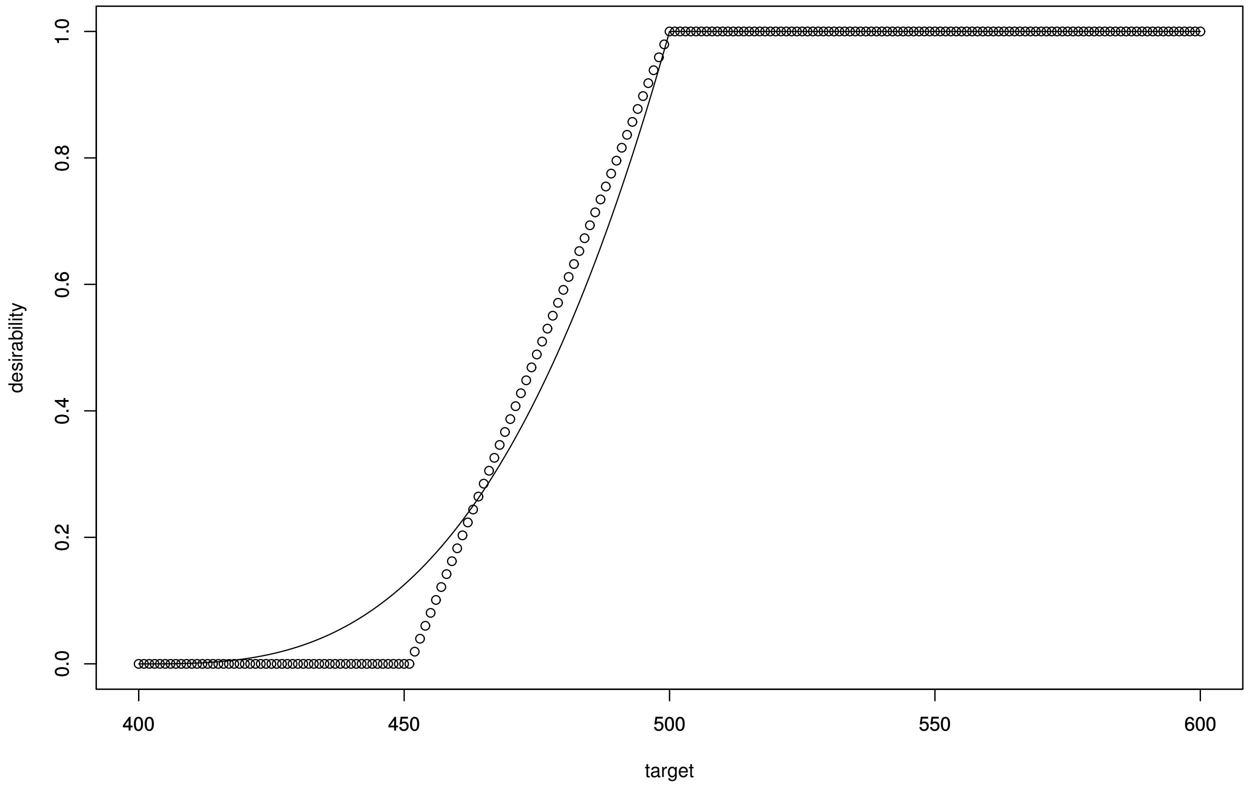
$$(l, t, u, \beta_l, \beta_r) = \begin{cases} (l, t, \infty, \beta_l, 1)(x) & \text{for } x \leq t; \\ (-u, -t, \infty, \beta_r, 1)(-x) & \text{for } t < x. \end{cases}$$

Define now

$$(l, t, u, \beta_l, \beta_r)_{lin} = \begin{cases} (l, t, \infty, \beta_l, 1)_{lin}(x) & \text{for } x \leq t; \\ (-u, -t, \infty, \beta_r, 1)_{lin}(-x) & \text{for } t < x. \end{cases}$$

$(l, t, u, \beta_l, \beta_r)_{lin}$ is called *once linearised (TV) desirability*.

(LB) desirability and once linearised approximation



2nd idea: twice linearised approximation

Given a (LB) desirability $(l, t, \infty, \beta_l, 1)$. Define now

$$(l, t, \infty, \beta_l, 1)_{2-lin} := \begin{cases} ((l, y^*, t, \infty), (0, d^*, 1, 1), (1, 1, 1)) & \text{for} \beta_l \neq 1; \\ (l, t, \infty, 1, 1) \text{ for } \beta_l = 1. \end{cases}$$

Here y^* and d^* are chosen such that:

$$(y^*, d^*) \text{ solves } \min_{(y^*, d^*) \in [l,t] \times [0,1]} \int_l^t ((l, t, \infty, \beta_l, 1) - (l, t, \infty, \beta_l, 1)_{2-lin}))^2 dx.$$

$(l, t, \infty, \beta_l, 1)_{2-lin}$ is called *twice linearised (LB) desirability*.

2nd idea: twice linearised approximation

(TV)-problems $(l, t, u, \beta_l, \beta_r)$ again are partitioned in two (LB)-problems:

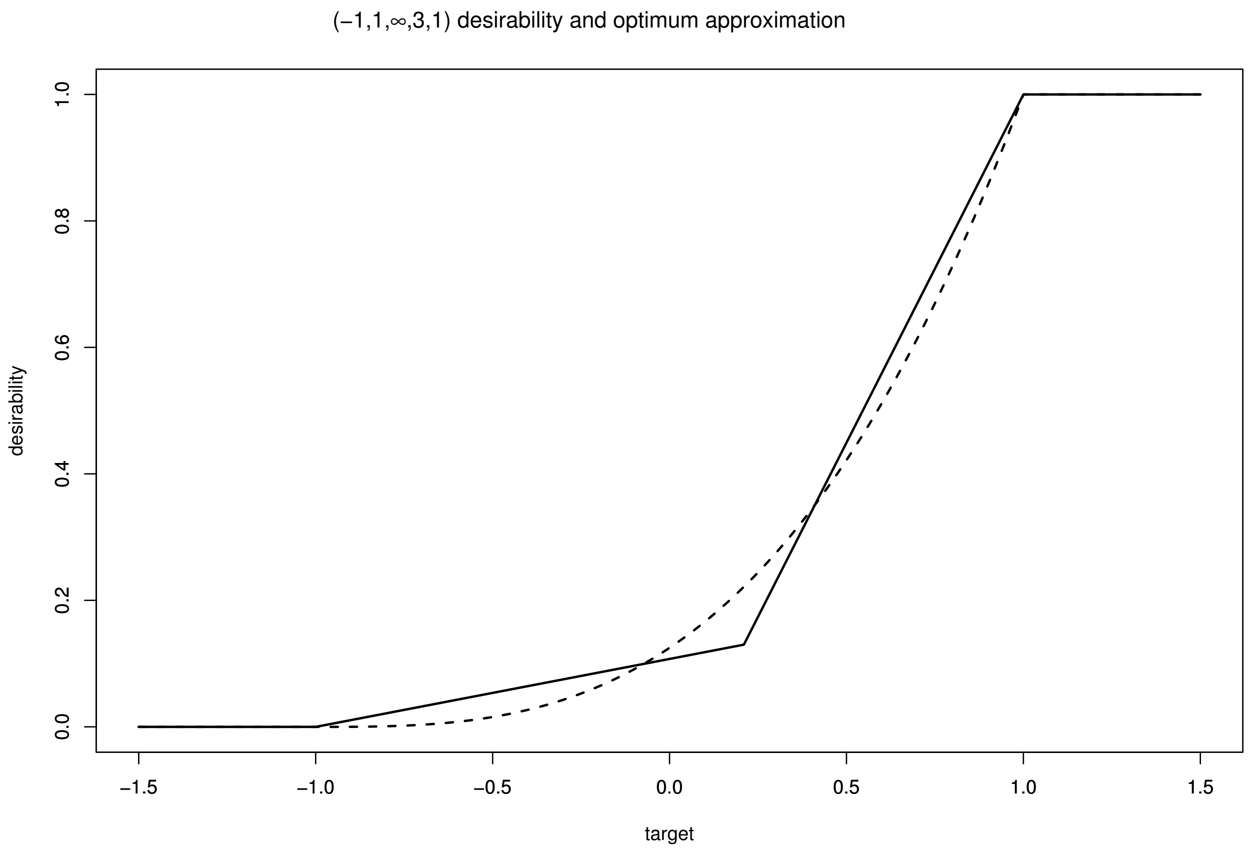
$$(l, t, u, \beta_l, \beta_r) = \begin{cases} (l, t, \infty, \beta_l, 1)(x) & \text{for } x \leq t; \\ (-u, -t, \infty, \beta_r, 1)(-x) & \text{for } t < x. \end{cases}$$

Define now

$$(l, t, u, \beta_l, \beta_r)_{2-lin} = \begin{cases} (l, t, \infty, \beta_l, 1)_{2-lin}(x) & \text{for } x \leq t; \\ (-u, -t, \infty, \beta_r, 1)_{2-lin}(-x) & \text{for } t < x. \end{cases}$$

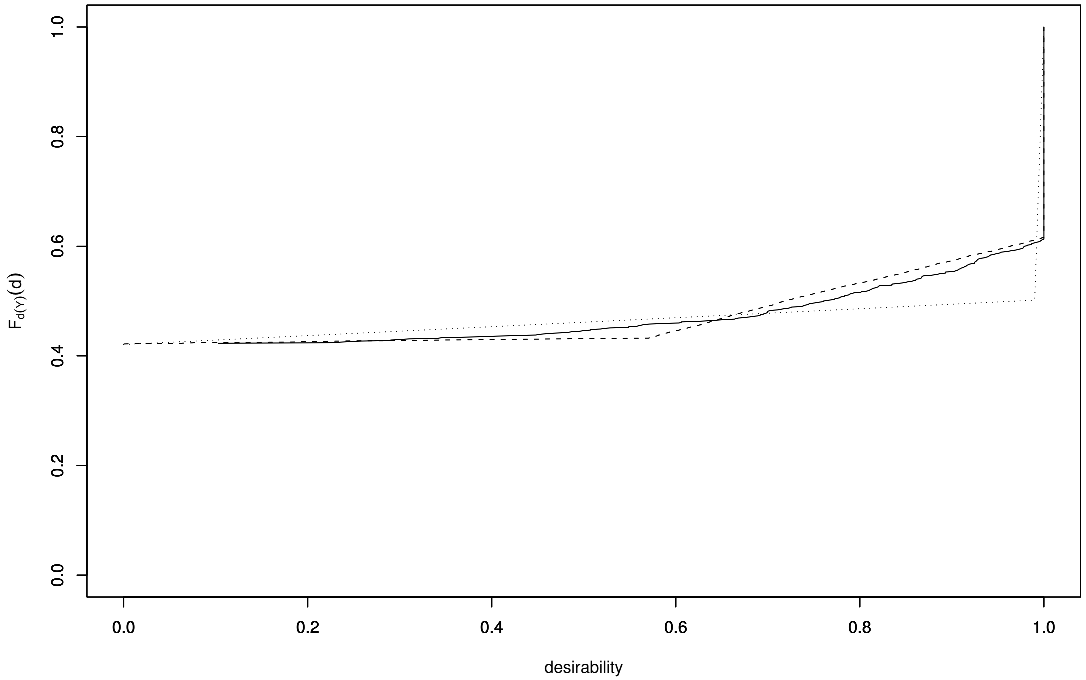
$(l, t, u, \beta_l, \beta_r)_{2-lin}$ is called *twice linearised (TV) desirability*.

(LB) desirability and twice linearised approximation



Goodness of approximationen

Simulated distribution of a $(-1, 0, \infty, 0.3, 1)$, respective distributions of the once and twice linearised approximations for $f(x) = -0.6$ and $\sigma = 2$.



Impact of approximationen

- Linearising two times seems to work well for a wide variety of exponents. (reproduces failure rate)
- Both approximations are special cases of the generalised DS-desirabilities and therefore their distributions and derived values are known.
- Simulations necessary to optimise q become **much** faster using the exact expressions for all specified desirabilities.
- As a follow-up some results can be given for the index, if all distributions for the single desirabilities are know at least approximately.

Distribution of the desirability index

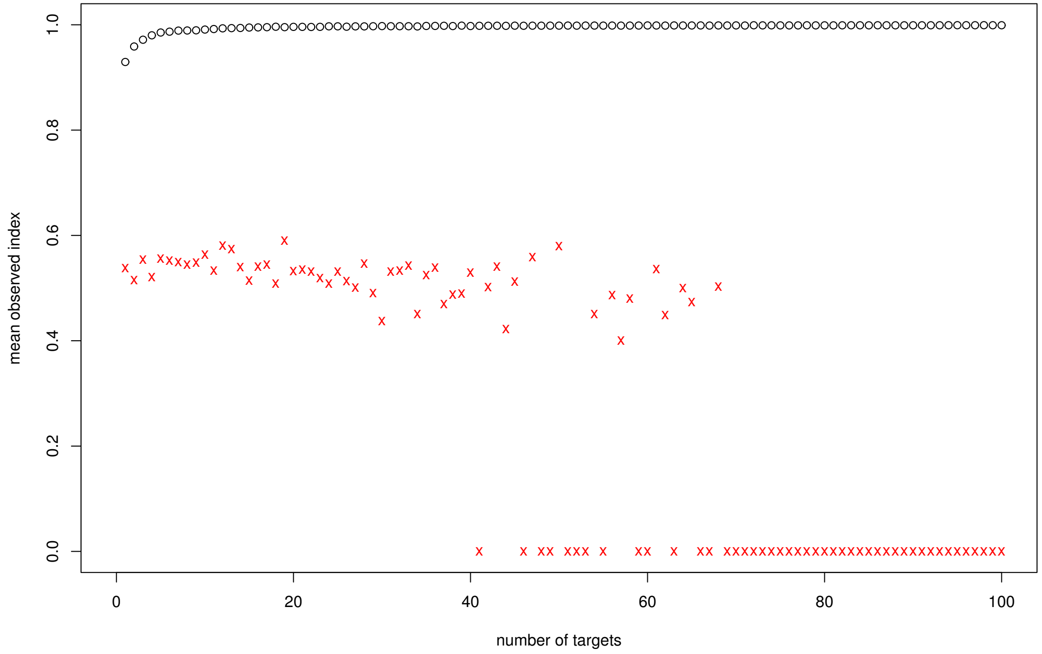
Known: Distributions are not from a class of stable distributions under multiplication.
Therefore: Distribution can not be given explicitly. In the Harrington case this is possible for some special case (Trautmann).
Partial results are possible under assumption of pairwise independence of Y_i:
The **failure rate** for a product with individual desirabilities d_i for a fixed factor setting x is given as:

$$p_0\big(q(Y)\big) = 1 - \prod_{1}^{m} \Big(1 - p_0\big(d_i(Y_i)\big)\Big).$$

Without this assumption at least fast simulations are possible.

Scalability of the desirability procedure I

Simulated desirability indices against the number of targets.



Scalability of the desirability procedure I

The expected value does not depend on the number of targets (under some assumptions). Let $Z := \prod_1^m d_i(Y_i)$, Y_i iid, then

$$\begin{aligned} E(Z) &= P(d_i(Y_i) > 0 \,\forall\, i) \cdot E\left(\prod_1^m d_i(Y_i) \,\middle|\, d_i(Y_i) > 0 \,\forall\, i \right) \\ &= \Big(1 - P\big(d_1(Y_1) = 0\big)\Big)^m \cdot E\Big(d_1(Y_1) \,\Big|\, d_1(Y_1) > 0\Big)^m . \end{aligned}$$

Now it follows

$$E(Z)^{\frac{1}{m}} = \big(1 - P(d_1(Y_1) = 0)\big) \cdot E\Big(d_1(Y_1) \,\Big|\, d_1(Y_1) > 0\Big).$$

Scalability of the desirability procedure II

Let $d_i(Y_i), i = 1, 2 \ldots,$ desirabilities, for which $p_0(d_i(Y_i)) = 0$, $E(\log d_i(Y_i)) < \infty$ for all i and $\sum_1^\infty \frac{Var(\log d_i(Y_i))}{i^2} < \infty$. Then:

$$
\begin{aligned}
E(q_n) &= E\left(\left[\prod_1^n d_i(Y_i)\right]^{\frac{1}{n}}\right) \\
&= E\left(e^{\frac{1}{n}\sum\limits_1^n \log d_i(Y_i)}\right) \\
&\xrightarrow[n\to\infty]{a.s.} E\left(e^{\frac{1}{n}\sum\limits_1^n E(\log d_i(Y_i))}\right) = e^{\frac{1}{n}\sum\limits_1^n E\left(\log d_i(Y_i)\right)}.
\end{aligned}
$$

Conclusion

- Weak points in current practice of using desirabilities were discovered.
- Main failure was ignorance for the random nature of the desirabilities.
- Explicit expressions for the distributions of desirabilities could be given, making simulations much faster.
- It was shown, that approximating these distributions using the normal distribution is not appropriate.
- Realisitic desirabilities fulfill all important requirements for MCO-procedures: pareto-optimum solutions, good scalability, practicability.

Future research

- The problem to put a statistical perspective on the underlying calibration problem remains unsolved. Therefore no uncertainty regions for $\widehat{x_{opt}}$ can be given.

- It would be interesting to try problem specific optimality-criteria for the linearisation step.

Final remarks

Derringer-Suich desirabilities could underline their usefulness as standard approach for MCO problems. To improve on the current practice realistic desirabilities should replace idealised desirabilities.

All in all realistic desirabilities can be recommended as a tool to solve MCO problems.