

# IMAPPNIO STSM REPORT

**Topic:** Drift analysis for surrogate-assisted evolutionary optimisation

**Time Frame:** 18/02/2017-18/03/2017

**Host:** Michael Emmerich, Leiden University, NL

**Guest:** Vanessa Volz, TU Dortmund University, GER

---

## 1 Purpose of the STSM

The purpose of the STSM was to address two research directions that are important for the analysis of an algorithm (SAPEO) that my co-authors Günter Rudolph, Boris Naujoks and I recently proposed. The algorithm originally intended for application in game balancing was published and accepted as both a single- as well as a multi-objective optimisation algorithm [1, 2].

However, in benchmarks<sup>1</sup> the algorithm did not perform as well as expected. During the STSM, we therefore wanted to find an explanation for this behaviour by identifying theoretical performance limits.

## 2 Work carried out and main results

At the start of the STSM, I reviewed three methods for stochastic convergence analysis suggested by my host and investigated their applicability. Starting from there, we set out to find theoretical bounds for different metrics in order to eventually find limits for the progress rate of SAPEO.

SAPEO is based on the observation that ranking solutions based on the prediction of a surrogate-model is possible without actual evaluation if you allow for a certain risk of ranking errors, i.e. individuals that are sorted into a wrong order. In [2] we already gave some approximation of the risk accepted by applying several different dominance relations we suggested. However, in the paper we were assuming independence of the error distributions of the surrogate model. More precise bounds can be found by taking the correlation assumption into account that is inherent to the Gaussian process model (Kriging). During the STSM, we were able to achieve a more general formulation of the probability that can be used given a general positive definite covariance matrix. We were able to identify theoretical upper and lower bounds for the probability of ranking errors for a (1+1)-EA and to verify them empirically. We were also able to transfer the concept to larger populations. A summary of the work can be found in appendix A. However, these are just preliminary results and require thorough checks before publication.

The second approach we tried during the STSM was to formalise SAPEO performance as a Markov process as done in standard convergence analysis. However, the surrogate models used in SAPEO are local and learned based on the neighbours of the point to predict. The samples that are selected for the learning process are the k-nearest neighbours from the archive of evaluated points. However, because of the growing archive the Markovian kernel would not be bounded in size. It was then investigated whether convergence bounds can be obtained making the assumption of a bounded size archive and, instead of the statistically involved Gaussian process model, assuming a linear Lipschitz model. We were able to come up with a theoretical argument of why

---

<sup>1</sup>BBOB single- and bi-objective <http://coco.gforge.inria.fr/>

restricting the archive to a size of 1 or even doing a worst-case analysis with a larger archive is not possible. This is because with the Lipschitz model we used for analysis, the individuals in a population are never distinguishable just based on model predictions in these cases. We thus switched to a different worst-case analysis: assuming a constant (worst-case) uncertainty for all predictions. With this approach, we were able to find theoretical bounds for the probability of acceptance, rejection and evaluation that are all necessary for a Markov formulation. We were also able to verify the results empirically. A draft of the results can be found in appendix B. The plan here is to identify an asymptotic bound for the uncertainty threshold experimentally (based on the results obtained from running the BBOB benchmarks). With this input, a better, more praxis-related approximation of the probabilities addressed should be possible.

What is still missing is the complete formalisation of the Markov process since we were not able to find the correct limits for the progress rate yet. However, we have three promising approaches based on the probabilities we computed before. But since we have no empirically verifiable results yet, these are not included in this report.

During the STSM and because of the discussion we had, we also were able to develop a new dominance relation that could be used for SAPEO. This would be based on a controllable threshold for the probability of ranking errors, and it could replace the assumption that the true function value is inside a given interval. We intend to test this approach in the future on the same BBOB benchmarks.

### **3 Future collaboration and publications**

Future collaboration and publications are definitely planned. One publication is intended to be on a benchmark of the new dominance relation for SAPEO. Additionally, once the remaining issues regarding the progress rate are fixed, we also intend to do and publish a complete convergence analysis on SAPEO.

Additionally, with the obtained results there were also some insights into possible shortcomings of SAPEO in terms of the adaptation of the uncertainty threshold that could potentially allow negative progress rate. This issue we will definitely investigate further both empirically as well as theoretically. We were thus able to develop useful rules for a practical application guided by theoretical insights.

## **References**

- [1] Vanessa Volz, Günter Rudolph, and Boris Naujoks. Investigating uncertainty propagation in surrogate-assisted evolutionary algorithms. In *Genetic and Evolutionary Computation Conference*, page Accepted, New York, NY, 2017. ACM Press.
- [2] Vanessa Volz, Günter Rudolph, and Boris Naujoks. Surrogate-assisted partial order-based evolutionary optimisation. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 639–653. Springer, 2017.

# A Probability of Ranking Errors under Different Strategies

## A.1 General Approach

Let  $i_X, i_Y$  be individuals in a population with predicted values  $\mu_X, \mu_Y$  and uncertainties  $\sigma_X, \sigma_Y$ , respectively. We now want to compute the probability that different ranking strategies sorted the individuals correctly. More formally, we compute  $P(i_X \geq_f i_Y)$  given a certain ranking by one of the other dominance relations.

Since all of the dominance relations trust the prediction of the surrogate model, we will assume (Assumption 1) that the error distributions are estimated correctly.

Now, let  $X, Y$  be random variables distributed according to the prediction of a surrogate model with  $X \sim N(\mu_X, \sigma_X^2)$  and  $Y \sim N(\mu_Y, \sigma_Y^2)$ . We can now express the probability we are looking for in terms of random variables, namely  $P(i_X \geq_f i_Y) = P(X \leq Y)$

Given Assumption 1, what is the probability that  $X \leq Y$ ? We can express the probability as in equation 1.

$$P(X \leq Y) = P(X - Y \leq 0) =: P(Z \leq 0) \quad (1)$$

$$Z = \vec{a}^T X = \sum_{j=1}^d a_j X_j, \text{ with } d = 2, a_1 = 1, a_2 = -1 \quad (2)$$

We can then identify the distribution of  $X - Y$  by expressing the distribution with a characteristic function (not assuming independence).

$$\phi_Z(t) = E[\exp(itZ)] = E[\exp(it\vec{a}^T X)] = \phi_X(t\vec{a}) \quad (3)$$

$$= \exp(it \sum_{j=1}^2 a_j \mu_j - \frac{1}{2} t^2 \sum_{j=1}^2 \sum_{k=1}^2 a_j a_k \Sigma_{jk}) \quad (4)$$

$$\stackrel{2}{=} \exp(it(\mu_X - \mu_Y) - \frac{1}{2} t^2 (\Sigma_{11} - \Sigma_{12} - \Sigma_{21} + \Sigma_{22})) \quad (5)$$

$$\text{with } \Sigma_{11} = \sigma_X^2, \Sigma_{22} = \sigma_Y^2, \Sigma_{12} = \Sigma_{21} = \sigma_{XY} \quad (6)$$

$$\Rightarrow Z = X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}) \quad (7)$$

It follows that

$$P(X - Y \leq a) \stackrel{7}{=} \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{a - (\mu_X - \mu_Y)}{\sqrt{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) \quad (8)$$

$$\stackrel{1}{\Rightarrow} P(X \leq Y) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{-\mu_X + \mu_Y}{\sqrt{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) \quad (9)$$

Now we analyse  $P(X \leq Y)$  for interesting cases:

## A.2 Case I: Dominance Relation $p$

$X \leq_p Y \Leftrightarrow \mu_X \leq \mu_Y$  This represents domination relation  $\leq_p$ , where individuals are ranked according to their predicted value.

Let  $\mu_X + d = \mu_Y$  with  $d \geq 0$ . Then, because the error function is strictly monotonous increasing (e), we can do the following estimate.

$$P(X \leq Y) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{-\mu_X + \mu_X + d}{\sqrt{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{d}{\sqrt{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) \quad (10)$$

$$\stackrel{d \geq 0, e}{\geq} \frac{1}{2} (1 + \operatorname{erf}(0)) = \frac{1}{2} \quad (11)$$

It follows that  $P(i_X \leq_f i_Y | i_X \leq_p i_Y) \geq \frac{1}{2}$

## A.3 Case II: Dominance Relation $u$

$X \leq_u Y \Leftrightarrow \mu_X + \sigma_X \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \leq \mu_Y - \sigma_Y \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)$  This represents domination relation  $\leq_u$ , where individuals are only considered dominated if the confidence intervals do not overlap.

Let

$$\mu_X + \sigma_X \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) + d = \mu_Y - \sigma_Y \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \quad (12)$$

$$\mu_Y = \mu_X + \sigma_X \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) + \sigma_Y \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) + d \quad (13)$$

$$\mu_Y = \mu_X + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) (\sigma_X + \sigma_Y + d'), \text{ with } d' = \frac{d}{\Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)} \quad (14)$$

Therefore:

$$P(X \leq Y) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{-\mu_X + \mu_X + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) (\sigma_X + \sigma_Y + d')}{\sqrt{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) \quad (15)$$

$$= \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{\Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) (\sigma_X + \sigma_Y + d')}{\sqrt{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) \quad (16)$$

$$= \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{\Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)}{\sqrt{2}} \sqrt{\frac{(\sigma_X + \sigma_Y + d')^2}{(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}} \right) \right) \quad (17)$$

Further:

$$\Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) = \sqrt{2} \operatorname{erf}^{-1} \left( \left( 1 - \frac{\alpha}{2} \right) - 1 \right) = \sqrt{2} \operatorname{erf}^{-1}(1 - \alpha) \quad (18)$$

$$\frac{(\sigma_X + \sigma_Y + d')^2}{(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})} \geq \frac{(\sigma_X + \sigma_Y)^2}{(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})} \geq 1 + \frac{2\sigma_X \sigma_Y}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{XY}} + \frac{2\sigma_{XY}}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{XY}} \quad (19)$$

Now, without loss of generality, let  $\sigma_Y = \beta \sigma_X$ , with  $\beta > 0 \in \mathbb{R}$ .  $\beta > 0$  since both  $\sigma_X, \sigma_Y > 0$ . Then:

$$\frac{2(\sigma_X \sigma_Y)}{\sigma_x^2 + \sigma_y^2} = \frac{2\beta \sigma_Y^2}{(1 + \beta^2) \sigma_Y^2} = \frac{2\beta}{(1 + \beta^2)} \text{ and} \quad (20)$$

$$\frac{2\beta}{(1 + \beta^2)} \in [0, 1] \text{ since } \beta > 0 \quad (21)$$

$\alpha$	$[0, 1]$	confidence level
$\rho$	$\{0, 0.5, 0.75, 1\}$	correlation of error distributions
$\beta$	$\{0.1, 0.6, 1.1, 1.6, 2.1, 2.6\}$	relation between uncertainties
$d$	0	distance between confidence interval $\alpha$ -level bounds
$\sigma_X$	1	uncertainty X
$\mu_X$	0	prediction X
$\sigma_Y$	$\beta\sigma_X$	uncertainty Y
$\mu_Y$	$\mu_X + \sigma_X\Phi^{-1}(1 - \frac{\alpha}{2}) + \sigma_Y\Phi^{-1}(1 - \frac{\alpha}{2}) + d$	prediction Y
$\Sigma$	$\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$	error covariance

Table 1: Parameter settings for experiments

Also:

$$\sigma_{XY} = \rho_{XY}\sigma_X\sigma_Y \geq 0, \text{ since } \rho_{XY} \in [0, 1] \quad (22)$$

$$\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} > 0 \quad (23)$$

Proof for equation 23:

- Case 1:  $\rho_{XY} = 1$ .

$$- \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y \stackrel{\rho_{XY}=1}{=} (\sigma_X - \sigma_Y)^2 \geq 0$$

- Case 1a:  $\sigma_X = \sigma_Y$

$$* \Rightarrow Z \sim N(\mu_X, \mu_Y, 0) \text{ TODO?}$$

- Case 1a:  $\sigma_X \neq \sigma_Y$

$$* \Rightarrow (\sigma_X - \sigma_Y)^2 > 0$$

- Case 2:  $\rho_{XY} < 1$ .

$$- \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} > (\sigma_X - \sigma_Y)^2 \geq 0$$

Thus, since the error function is strictly monotonous increasing:

$$P(X \leq Y) \geq \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{\sqrt{2} \operatorname{erf}^{-1}(1 - \alpha)}{\sqrt{2}} \sqrt{1 + 0 + 0} \right) \right) \quad (24)$$

$$= \frac{1}{2} (1 + \operatorname{erf}(\operatorname{erf}^{-1}(1 - \alpha))) \quad (25)$$

$$= 1 - \frac{1}{2}\alpha \quad (26)$$

Let  $\gamma = \frac{d}{\Phi^{-1}(1 - \frac{\alpha}{2})\sigma_Y}$ . Figure 1 visualises the lower bound (black) and the upper bound for different  $\gamma$  for the probability  $P(i_X \leq_f i_Y | i_X \leq_u i_Y)$  and different confidence levels  $\alpha$  in cases of independent error distributions, i.e.  $\sigma_{XY} = 0$ .

We also tested the result empirically. We chose the error distribution parameters of two random variables so that  $\leq_u$  can just distinguish between, i.e. the condition for case II. We set some parameters for testing purposes and set the others accordingly (see table 1).

We can thus define a binormal distribution  $N((\mu_X, \mu_Y)^T, \Sigma)$  that fulfills the conditions of case II to simulate the error distributions of two individuals. We then sample from the distribution 1000000 times and record the frequency of simulated ranking errors, i.e. where  $x > y : (x, y) \sim$

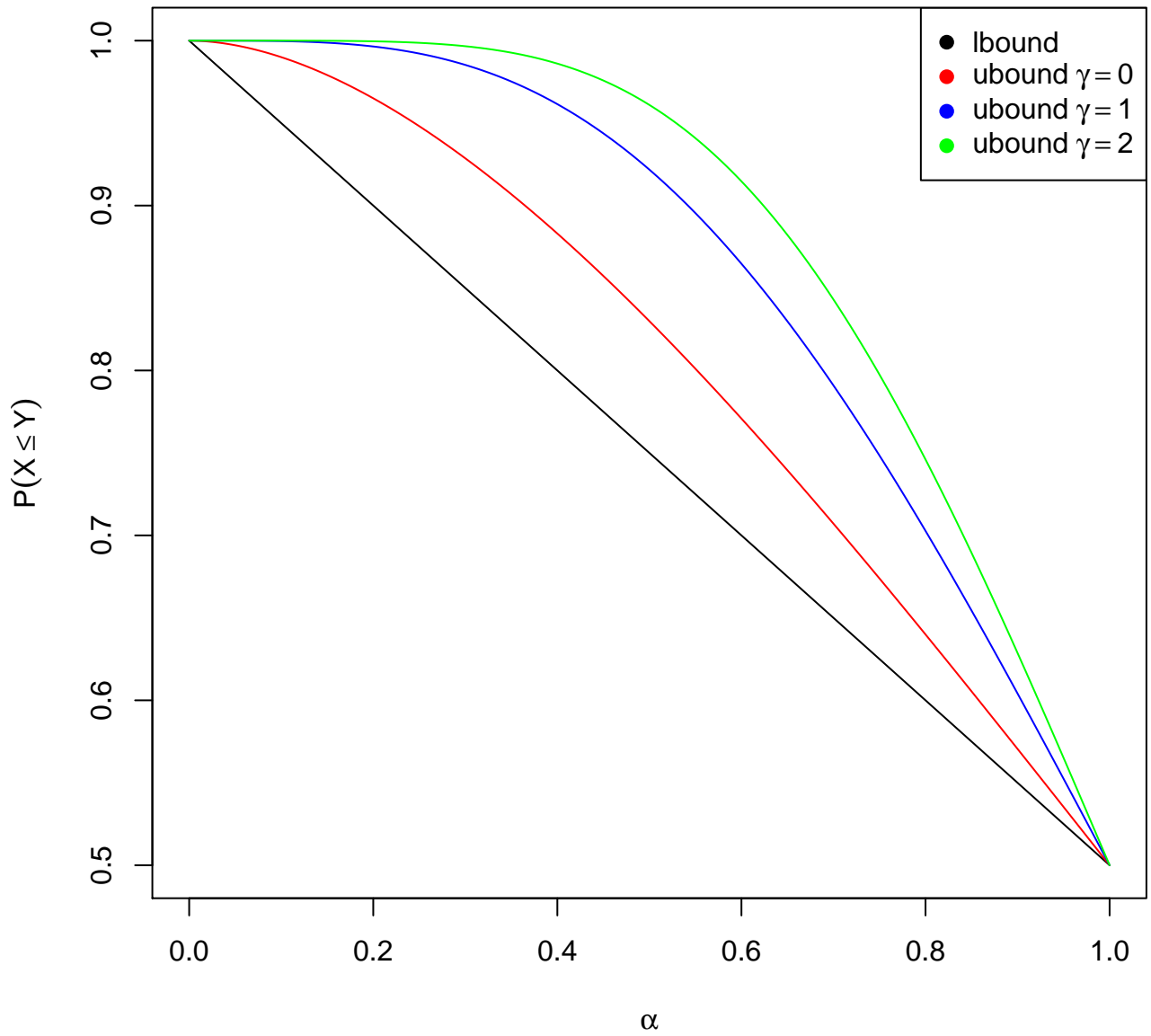


Figure 1: Probability bounds of  $P(X \leq Y)$ , given a ranking of  $i_X \leq_u i_Y$  for different confidence levels  $\alpha$  and different gap ratios  $\gamma$ .

$N((\mu_X, \mu_Y)^T, \Sigma)$ . The results for different settings of  $\alpha, \rho, \beta$  as listed in 1 are visualised in figure 2.

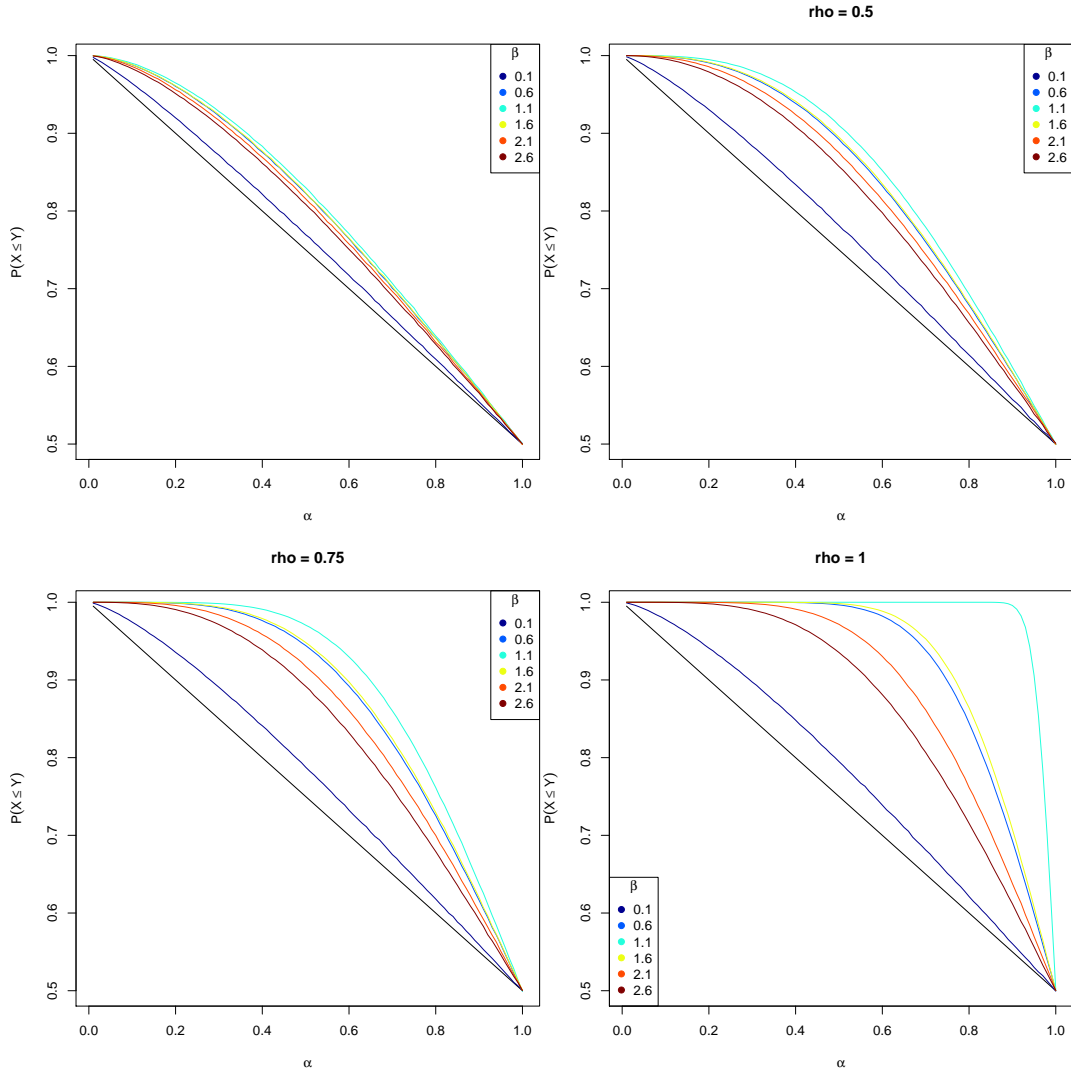


Figure 2: Empirical frequency of ranking errors for different  $\alpha$  (confidence levels),  $\beta$  (relation between uncertainties) and  $\rho$  (correlation between error distributions) as well as the theoretical lower bound (black).

#### A.4 Larger Population

We now look at what changes if we increase the population size. Let  $R_i$  be the set of individuals of rank  $i$  according to  $\leq_c$  and  $r$  be the highest rank. The ranking is correct iff

$$\forall x \in R_i : x \leq_f y \forall y \in R_{i+1}, i \in \{0, \dots, r-1\} \quad (27)$$

The necessary comparisons can be encoded in a similar form as above, i.e. as the difference of two random variables. Thus, let

$$Z_{k,l}^{(i)} = R_i^k - R_{i+1}^l, i \in \{0, \dots, r-1\}, k \in \{1, \dots, |R_i|\}, l \in \{1, \dots, |R_{i+1}|\} \quad (28)$$

, where  $R_i^k$  is the  $k$ th individual at rank  $i$ . Thus, the probability of no ranking errors can be described as  $P(Y_{1,1}^{(0)} \leq 0, \dots)$ . In order to express this probability, we can use an affine transformation

with  $y = c + Bx$  and  $c = \vec{0}$  where

$$x \sim N(m, \Sigma), m = (\hat{f}(x_1), \dots, \hat{f}(x_\lambda))^T, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,\lambda} \\ \dots & & & \\ \sigma_{\lambda,1} & \dots & \dots & \sigma_\lambda^2 \end{pmatrix} \quad (29)$$

$$y \sim N(Bm, B\Sigma B^{-1}) \quad (30)$$

and where  $B$  encodes the comparisons, i.e.  $B$  contains  $\lambda$  columns and a row for each  $Z_{k,l}^{(i)}$  where the  $k$ th element is 1, the  $l$ th  $-1$  and the rest 0. We can thus compute the probability of no ranking errors (denoted as  $p$  in the following) with the cdf of the multivariate normal distribution for  $y$ . We can then define a distribution of the worst-case selection errors with

$$\begin{cases} 0 & p \\ \mu & 1-p \end{cases} \quad (31)$$

. With this very rough estimate we already have an upper bound of the expected value for the number of selection errors with  $\mu(1-p)$ . In order to make a better estimate, we would also need to find the probabilities of  $1 \dots \mu$  sorting or even better selection errors.

## B Multi-Fidelity Models

In the following, we look at the progress rate and similar metrics using a Multi-Fidelity Models Approach in order to express SAPEO. We denote  $a_1, \dots, a_A$  as the points in search space that have been sampled previously. Together, they form the archive of size  $A$  from which values are selected to build the local surrogate models for SAPEO. For this analysis, however, we assume a constant size for the archive. This allows us to express SAPEO as a Markovian process despite the growing state space in the original algorithm. We further denote  $d_X^{(i)}$  as the distance from individual  $X$  to archive point  $i$  in search space.  $\varepsilon_g$  is the uncertainty that is allowed for the surrogate model prediction at generation  $g$ , which we assume to be constant throughout the runtime of the algorithm. In practice,  $\varepsilon$  would decrease with runtime in order to force better predictions and an approximation of the optimisation target.

SAPEO can be interpreted as a strategy that uses multi-fidelity models:

- The surrogate model (Kriging or Lipschitz) used up to a specific uncertainty  $\varepsilon$  of the prediction and certainty of selection decisions
- The actual fitness evaluation function

### B.1 Approach 1a: Exact model Lipschitz model with archive size 1

Now let us first assume an archive size of  $A = 1$ , so let  $d_X^{(1)} = d_X$  be the distance from individual  $X$  to the only point in the archive. In the Lipschitz model, the bounds of the confidence interval with only one sample are  $f(a_1) [-Ld_X\Phi^{-1}(1 - \frac{\alpha}{2}), Ld_X\Phi^{-1}(1 - \frac{\alpha}{2})]$  with some constant  $L > 0 \in \mathbb{R}$ . This analysis would be very similar for a Kriging Model since only the confidence interval bounds are relevant here.

In SAPEO, an individual is evaluated if either its prediction uncertainty is larger than the uncertainty threshold  $\varepsilon$  or (for some strategies) if it cannot be compared to other individuals. We look at the first case in the following. Thus, the probability of evaluation is:

$$P_\varepsilon = P\left(Ld_X\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \varepsilon\right) = P\left(d_X \leq \frac{\varepsilon}{L\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}\right) \quad (32)$$



Now let  $X$  be the child of  $Y$ . Since child  $X$  is generated from parent  $Y$  with a mutation with

$$X = Y + \sigma_{\text{mut}}Z, Z_i \sim N(0, 1) \quad (33)$$

, it holds that  $X_i$  is an independent, normally distributed random variable with mean  $Y_i$  and variance  $\sigma_{\text{mut}}^2$ . Therefore, the test statistic

$$T = \sqrt{\sum_{i=1}^N \left(\frac{x_i}{\sigma}\right)^2} = \frac{1}{\sigma} \sqrt{\sum_{i=1}^N x_i^2} = \frac{d_X}{\sigma} \quad (34)$$

is distributed according to a noncentral chi-distribution with

$$\lambda = \sqrt{\sum_{i=1}^N \left(\frac{y_i}{\sigma}\right)^2} = \frac{1}{\sigma} \sqrt{\sum_{i=1}^N y_i^2} = \frac{d_Y}{\sigma} \quad (35)$$

. Therefore:

$$f\left(\frac{d_X}{\sigma_{\text{mut}}} \mid d_Y, \sigma_{\text{mut}}\right) = \frac{\exp\left(\frac{-(d_X^2 + d_Y^2)}{2\sigma_{\text{mut}}^2}\right) \left(\frac{d_X}{\sigma_{\text{mut}}}\right)^N \left(\frac{d_Y}{\sigma_{\text{mut}}}\right)}{\left(\frac{d_X d_Y}{\sigma_{\text{mut}}^2}\right)^{\frac{N}{2}}} I_{\frac{N}{2}-1}\left(\frac{d_X d_Y}{\sigma_{\text{mut}}^2}\right) \quad (36)$$

$$= \exp\left(\frac{-(d_X^2 + d_Y^2)}{2\sigma_{\text{mut}}^2}\right) d_X^{\frac{N}{2}} d_Y^{-\frac{N}{2}+1} \sigma_{\text{mut}}^{-1} I_{\frac{N}{2}-1}\left(\frac{d_X d_Y}{\sigma_{\text{mut}}^2}\right) \quad (37)$$

$$I_{\frac{N}{2}-1}(x) = \frac{1}{\pi} \int_0^\pi \cos\left(\left(\frac{N}{2} - 1\right)\tau - x \sin(\tau)\right) d\tau \quad (38)$$

For the two-dimensional case,  $T$  also follows a Rician distribution with the same  $\lambda$ , a special case of the noncentral chi distribution.

Let furthermore  $P(X \leq_c Y)$  be the probability that  $X$  and  $Y$  can be distinguished and that  $X \leq_c Y$  according to the dominance relation. The following transitions all happen when  $X \leq_c Y$

$$P(d_Y \rightarrow 0) = f(0 \mid d_Y, \sigma_{\text{mut}}) P_\epsilon + (1 - P_\epsilon) \quad (39)$$

$$P(d_Y \rightarrow d_X) = 0 \text{ for } d_X > \frac{\epsilon}{L\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)} \quad (40)$$

$$P(d_Y \rightarrow d_X) = f(d_X \mid d_Y, \sigma_{\text{mut}}) P_\epsilon \text{ else} \quad (41)$$

The  $d_X$  equals 0 if child  $X$  is evaluated because of the threshold or the mutation lands on the archive point by chance.  $d_X$  is also limited by  $\epsilon$ , which explains the second line. In other cases,  $d_X$  follows a noncentral chi distribution as described above.

In the remaining situations where  $X \parallel_c Y$  it holds that  $d_Y \rightarrow 0$ , if we assume for simplicity that both  $X$  and  $Y$  have to be evaluated. Furthermore, if  $Y \leq_c X$ , it holds that  $d_Y \rightarrow d_Y$ .

However,  $P(X \parallel_c Y) = 1$  for archive size = 1, since:

$$P(X \leq_c Y) = P\left(f(a_1) + Ld_X \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq f(a_1) - Ld_Y \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \quad (42)$$

$$= P\left(L\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)(d_X + d_Y) \leq 0\right) \quad (43)$$

The latter condition can only be true if  $d_X = d_Y = 0$ , since the remaining variables are positive and larger than 0. However, this would mean that,  $X$ ,  $Y$  and the archive point  $a_1$  are all the same and it also holds that  $X \parallel_c Y$  and  $Y \leq_c X$ . We have thus shown that an analysis with just a single point in the archive is not helpful, since no pair of individuals would be distinguishable according to  $\leq_c$ .

## B.2 Approach 1b: Exact model Lipschitz model with arbitrary archive size

We therefore set  $A = k, k \in \mathbb{N}$  instead. Then each distance  $d_X^{(i)}, i \in 1, \dots, k$  is distributed as  $d_X$  in the previous section. We can then formulate the probability of distinguishability:

$$P(X \leq_c Y) = P\left(\min_i \left(f(a_i) + Ld_X^{(i)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \leq \max_i \left(f(a_i) - Ld_Y^{(i)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)\right) \quad (44)$$

$$= P\left(\min_i \left(f(a_i) + Ld_X^{(i)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \leq -\min_i \left(-f(a_i) + Ld_Y^{(i)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)\right) \quad (45)$$

$$= P\left(\min_{i,j} \left(f(a_i) - f(a_j) + L\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)(d_X^{(i)} + d_Y^{(j)})\right) \leq 0\right) \quad (46)$$

In cases where  $i = j$ , it holds that

$$\min_{i,j} \left(f(a_i) - f(a_j) + L\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)(d_X^{(i)} + d_Y^{(j)})\right) = L\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)(d_X^{(i)} + d_Y^{(j)}) \quad (47)$$

, which results in the same problem as with the single archive point in the previous section. However, while this is an exact expression, it is difficult to simplify further. We therefore take another approach in the following section.

## B.3 Approach 2a: Assuming constant uncertainty

In this approach, we simplify the above analysis by assuming that the uncertainty is constant everywhere. In SAPEO, the uncertainty can not be larger than the threshold  $\epsilon$ . This is why this approach is valid and describes a lower bound of the metrics discussed in the following. However, what is ignored here is that some individuals have to be evaluated to enforce this uncertainty limit. This will be discussed in the following section approach 2b. We can now compute certain probabilities.

### B.3.1 Probability of acceptance

TODO: capitals for random variables We derive the probability of acceptance depending on the distance to the target of the child  $\tilde{R}$  and the parent  $R$ , denoted as  $P_a(\tilde{R}, R)$ . Due to the uncertainty dominance relation  $\geq_c$ , a child  $x$  with parent  $y$  is accepted if

$$\hat{f}(x) + \epsilon \leq \hat{f}(y) - \epsilon \quad (48)$$

If we assume that the Kriging model has predicted the correct error distributions, we also know that for any individual  $i$  it holds that

$$\hat{f}(i) - \epsilon \leq f(i) \leq \hat{f}(i) + \epsilon \quad (49)$$

with a probability of at least  $1 - \alpha$ , i.e. if the correct value lies within the confidence bounds. In the following, we assume that the actual value of the individual always does lie in these bounds.

Then we know that

$$P(f(x) + 2\epsilon \leq f(y) - 2\epsilon) \leq P(\hat{f}(x) + \epsilon \leq \hat{f}(y) - \epsilon) \leq P(f(x) \leq f(y)) \quad (50)$$

For the sphere function, it also holds that for any individual  $i$   $f(i) = \alpha \sum_{j=1}^N (i_j - \gamma_j)^2 + \beta, \alpha, \beta \in \mathbb{R}, \gamma \in \mathbb{R}^N$ . Since from now on we will express the individuals only in terms of their relative distance from the target, we set  $\gamma = (0)^N$  w.l.o.g.. In the following, we also set  $\alpha = 1, \beta = 0$ . Therefore,

we now know the distance of any individual  $i$  to the target  $R_i$  to be  $R_i = \sqrt{\sum_{j=1}^N i_j^2} = \sqrt{f(i)}$ . Thus:

$$P(\tilde{R}^2 + 2\epsilon \leq R^2 - 2\epsilon) \leq P_\alpha(\tilde{R}, R) \leq P(\tilde{R}^2 \leq R^2) \quad (51)$$

$$P\left(\frac{\tilde{R}^2}{\sigma^2} \leq \frac{R^2 - 4\epsilon}{\sigma^2}\right) \leq R^2 - 2\epsilon \leq P_\alpha(\tilde{R}, R) \leq P\left(\frac{\tilde{R}^2}{\sigma^2} \leq \frac{R^2}{\sigma^2}\right) \quad (52)$$

Since child  $x$  is generated from parent  $y$  with a mutation with

$$x = y + \sigma z, z_i \sim N(0, 1) \quad (53)$$

, it holds that  $x_i$  is a independent, normally distributed random variable with mean  $y_i$  and variance  $\sigma^2$ . Given that  $\gamma = (0)^T$  we then now that the test statistic

$$T = \sqrt{\sum_{i=1}^N \left(\frac{x_i}{\sigma}\right)^2} = \frac{1}{\sigma} \sqrt{\sum_{i=1}^N x_i^2} = \frac{\tilde{R}}{\sigma} \quad (54)$$

is distributed according to a noncentral chi-distribution with

$$\lambda = \sqrt{\sum_{i=1}^N \left(\frac{y_i}{\sigma}\right)^2} = \frac{1}{\sigma} \sqrt{\sum_{i=1}^N y_i^2} = \frac{R}{\sigma} \quad (55)$$

. Similarly, we also know that the test statistic

$$T = \sum_{i=1}^N \left(\frac{x_i}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^N x_i^2 = \frac{\tilde{R}^2}{\sigma^2} = \frac{f(\tilde{R})}{\sigma^2} \quad (56)$$

is distributed according to a noncentral chi-squared distribution with

$$\lambda = \sum_{i=1}^N \left(\frac{y_i}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^N y_i^2 = \frac{R^2}{\sigma^2} = \frac{f(R)}{\sigma^2} \quad (57)$$

With the definition of the CDF of the noncentral chi-squared distribution we therefore have for the probability of acceptance

$$1 - Q_{\frac{N}{2}}\left(\sqrt{\frac{R^2}{\sigma^2}}, \sqrt{\frac{R^2 - 4\epsilon}{\sigma^2}}\right) = 1 - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{\sqrt{R^2 - 4\epsilon}}{\sigma}\right) \quad (58)$$

$$\leq P_\alpha(\tilde{R}, R) \quad (59)$$

$$\leq 1 - Q_{\frac{N}{2}}\left(\sqrt{\frac{R^2}{\sigma^2}}, \sqrt{\frac{R^2}{\sigma^2}}\right) = 1 - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{R}{\sigma}\right) \quad (60)$$

, with for  $M \in \mathbb{Z}$

$$Q_M(a, b) = \exp\left(-\frac{a^2 + b^2}{2}\right) \sum_{k=1-M}^{\infty} \left(\frac{a}{b}\right)^k I_k(ab) \quad (61)$$

$$I_k(x) = \sum_{r=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2r+k}}{\Gamma(r+k+1)r!} \quad (62)$$

with Marcum-Q function  $Q(a, b)$  and modified Bessel function  $I_k$  of order  $k$  of the first kind.

TODO: M not in Z (uneven N)

### B.3.2 Probability of rejection

Similarly to the above, we reject a child  $x$  from parent  $y$  outright if

$$\hat{f}(x) - \epsilon \geq \hat{f}(y) + \epsilon \quad (63)$$

and

$$P(f(x) - 2\epsilon \geq f(y) + 2\epsilon) \leq P(\hat{f}(x) - \epsilon \geq \hat{f}(y) + \epsilon) \leq P(f(x) \geq f(y)) \quad (64)$$

$$P\left(\frac{\tilde{R}^2}{\sigma^2} \geq \frac{R^2 + 4\epsilon}{\sigma^2}\right) \leq P_r(\tilde{R}, R) \leq P\left(\frac{\tilde{R}^2}{\sigma^2} \geq \frac{R^2}{\sigma^2}\right) \quad (65)$$

$$1 - P\left(\frac{\tilde{R}^2}{\sigma^2} \leq \frac{R^2 + 4\epsilon}{\sigma^2}\right) \leq P_r(\tilde{R}, R) \leq 1 - P\left(\frac{\tilde{R}^2}{\sigma^2} \leq \frac{R^2}{\sigma^2}\right) \quad (66)$$

$$Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{\sqrt{R^2 + 4\epsilon}}{\sigma}\right) \leq P_r(\tilde{R}, R) \leq Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{R}{\sigma}\right) \quad (67)$$

### B.3.3 Probability of evaluation

Since in this subsection, we assume a constant uncertainty, the probability of evaluation is equal to the probability of noncomparability. The probability that two individuals are not comparable according to  $\leq_c (\|\epsilon)$ , denoted as  $P_e(\tilde{R}, R)$  is equal to  $1 - P_a(\tilde{R}, R) - P_r(\tilde{R}, R)$ . Therefore, we have

$$1 - \left(1 - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{R}{\sigma}\right)\right) - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{R}{\sigma}\right) = 0 \quad (68)$$

$$\leq P_e(\tilde{R}, R) \quad (69)$$

$$\leq 1 - \left(1 - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \sqrt{\frac{R^2 - 4\epsilon}{\sigma^2}}\right)\right) - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{\sqrt{R^2 + 4\epsilon}}{\sigma}\right) \quad (70)$$

$$= Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \sqrt{\frac{R^2 - 4\epsilon}{\sigma^2}}\right) - Q_{\frac{N}{2}}\left(\frac{R}{\sigma}, \frac{\sqrt{R^2 + 4\epsilon}}{\sigma}\right) \quad (71)$$

## B.4 Empirical validation

In an experiment very similar to the previous one, we verified the obtained theoretical results as depicted in figures 3 and 4.

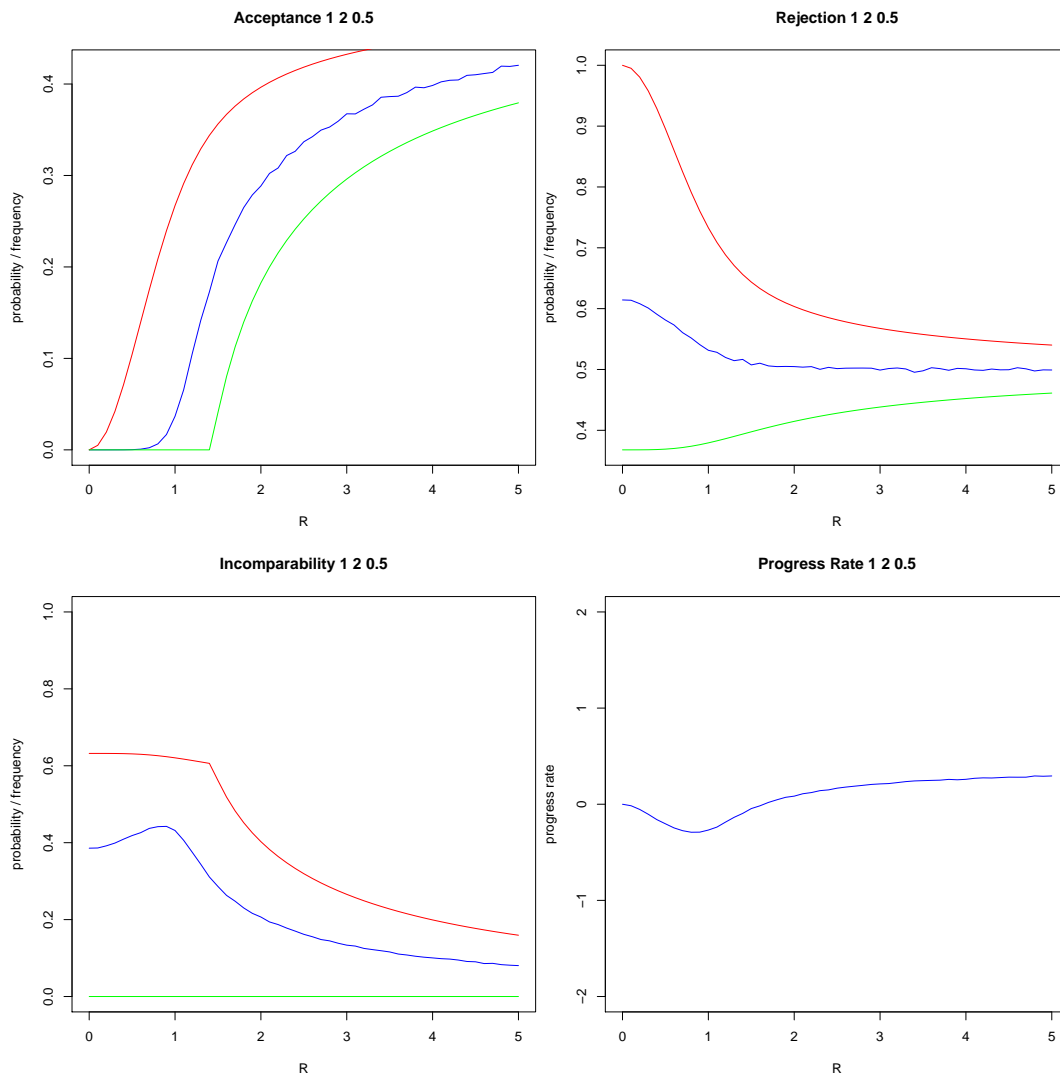


Figure 3: Acceptance, rejection, incomparability and progress rate (blue) for  $\sigma = 1, N = 2, \epsilon = 0.5$  and  $10^5$  experiments per datapoint. The red and green lines represent the discovered theoretical upper and lower bounds, respectively.

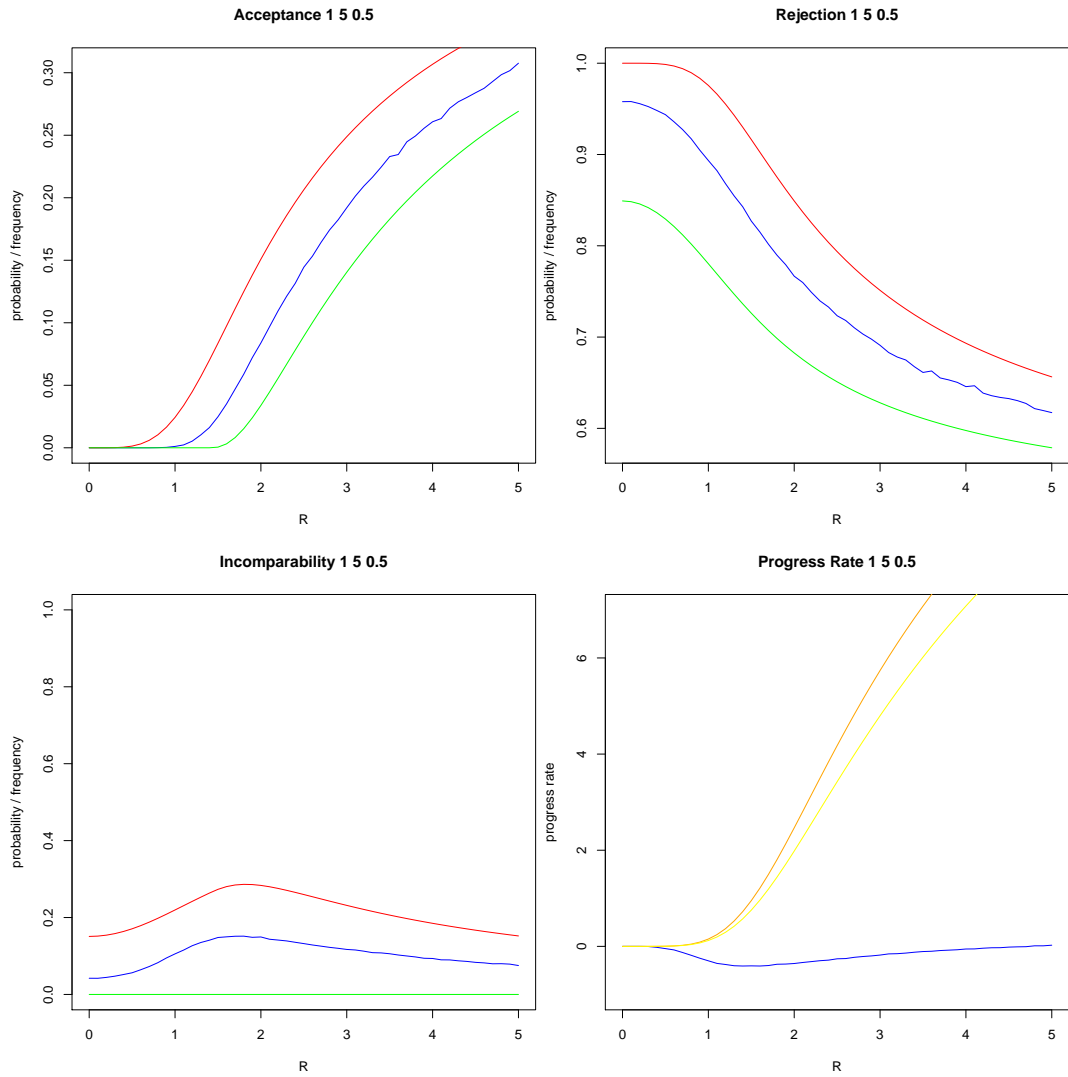


Figure 4: Acceptance, rejection, incomparability and progress rate (blue) for  $\sigma = 1, N = 5, \epsilon = 0.5$  and  $10^5$  experiments per datapoint. The red and green lines represent the discovered theoretical upper and lower bounds, respectively. The yellow and orange line are first estimates of the progress rate.

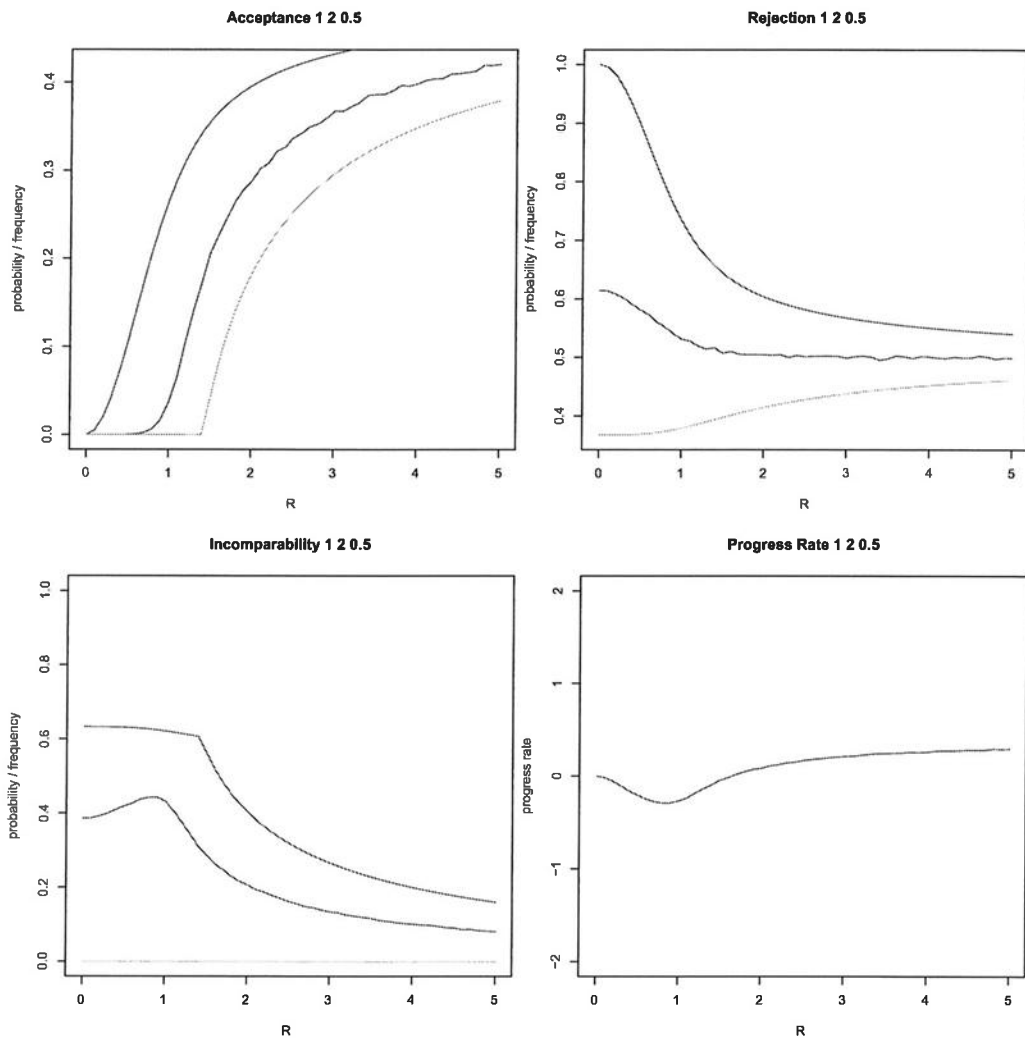


Figure 3: Acceptance, rejection, incomparability and progress rate (blue) for  $\sigma = 1, N = 2, \epsilon = 0.5$  and  $10^5$  experiments per datapoint. The red and green lines represent the discovered theoretical upper and lower bounds, respectively.

3.30.2017 Geide  
M. Epping