On the Distribution of EMOA Hypervolumes

Olaf Mersmann\(^1\), Heike Trautmann\(^1\), Boris Naujoks\(^2\), and Claus Weihs\(^1\)

\(^1\) Statistics Faculty, TU Dortmund University, Germany
\{olafm,trautmann,weihs\}@statistik.tu-dortmund.de
\(^2\) Log\!n GmbH, Schwelm, Germany
boris.naujoks@login-online.de

Abstract. In recent years, new approaches for multi-modal and multi-objective stochastic optimisation have been developed. It is a rather normal process that these experimental fields develop independently from other scientific areas. However, the connection between stochastic optimisation and statistics is obvious and highly appreciated. Recent works, such as sequential parameter optimisation (SPO, cf. Bartz-Beielstein [1]) or online convergence detection (OCD, cf. Trautmann et al. [2]), have combined methods from evolutionary computation and statistics.

One important aspect in statistics is the analysis of stochastic outcomes of experiments and optimization methods, respectively. To this end, the optimisation runs of different evolutionary multi-objective optimisation algorithms (EMOA, cf. Deb [3] or Coello Coello et al. [4]) are treated as experiments to analyse the stochastic behavior of the results and to approximate the distribution of the performance of the EMOA.

To combine the outcome of an EMOA and receive a single performance indicator value, the hypervolume (HV) indicator is considered, which is the only known unary quality indicator in this field (cf. Zitzler et al. [5]).

The paper at hand investigates and compares the HV indicator outcome of multiple runs of two EMOA on different mathematical test cases.

1 Experiments

An experiment is conducted to analyze the HV distribution of different EMOA given a fixed test function. NSGA-II [6] and SMS-EMOA [7] are considered. NSGA-II was chosen because it is the industrially most popular EMOA. On the other hand, recent studies have shown, that the hypervolume contribution is a promising selection criterion. The SMS-EMOA is a representative of such algorithms.

Three bi-objective test functions ZDT1, ZDT2, ZDT3 [9] are studied with populations sizes \(\mu = 100\) and \(200\) generations for NSGA-II and \(20,000\) function evaluations (FE) for the SMS-EMOA. The respective parameter settings are summarized in Table I.

For each combination of test function and algorithm 500 independent runs are performed, and the HV of the non-dominated solutions is computed for different numbers of FE using the nadir-point \((f_1, f_2) = (11, 11)\) [12]. The resulting distributions are visualized using boxplots in Fig. I Kernel density estimates [13].
Table 1. Parameter settings of the EMOA

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Implem</th>
<th>p_c</th>
<th>p_m</th>
<th>η_c</th>
<th>η_m</th>
<th>p_swap</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSGA-II</td>
<td>R [10]</td>
<td>0.7</td>
<td>0.2</td>
<td>20</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>SMS-EMOA</td>
<td>R [11]</td>
<td>0.9</td>
<td>1/length(x)</td>
<td>15</td>
<td>20</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Fig. 1. Boxplot of the HV values after different number of FEs for each algorithm and test function.

of the HV distributions at different stages of the EMOA are depicted in Fig. 2 and parallel coordinate plots [14] show the profile of the individual runs over time in Fig. 3.

2 Conclusion

From Fig. 1 it is apparent, that the expected HV value increases monotonically with the number of FEs while the variance of the HV simultaneously decreases. The initial variance of the HV is larger for the SMS-EMOA but as the number of FEs increases, it rapidly becomes smaller than the HV variance of the NSGA-II. This is due to the different selection strategies of the algorithms.

Focusing on NSGA-II, the distributions appear to be symmetrical while for the SMS-EMOA they are skewed. Another difference between the algorithms is the occurrence of outliers. This is especially apparent for the SMS-EMOA, which in some runs gets stuck at suboptimal HV values. This can also be seen in Fig. 3. For ZDT2 and ZDT3 there is a distinct block of runs that above 5 000 FEs is not able to escape from the low ranks. The NSGA-II does not exhibit such a