Evaluation Report on Six Experiments

Abstract. This is the evaluation report on six experiments, namely 1FlipHC, 1FlipHCrs, 2FlipHC, 2FlipHCrs, mFlipHC, and mFlipHCrs on 100 benchmark instances. This report has been generated with the version 0.8.4 of the Evaluator Component of the Optimization Benchmarking Tool Suite.

by Anne Anonymous
on 2015-09-14

1. Instance Information

Experiments were conducted on 100 benchmark instances, which can be distinguished by two features.
The benchmark instances are characterized by two features:

- n (ten values, ranging from 20 to 250)
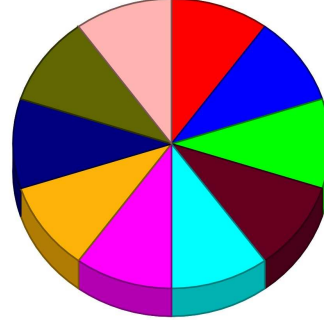- k (ten values, ranging from 91 to 1065)

●	91
●	218
●	325
●	430
●	538
●	645
●	753
●	860
●	960
●	1065

●	20
●	50
●	75
●	100
●	125
●	150
●	175
●	200
●	225
●	250

Fig. 1.1.1. Feature k   **Fig. 1.1.2.** Feature n

Fig. 1.1. The fractions of instances with specific feature values.

In Figure 1.1 we illustrate the relative amount of benchmark instances per feature value over all 100 benchmark instances. The slices in the pie charts are the bigger, the more benchmark instances have the associated feature value in comparison to the other values. The more similar the pie sizes are, the more evenly are the benchmark instances distributed over the benchmark feature values, which may be a good idea for fair experimentation.

2. Performance Comparisons

2.1. Estimated Cumulative Distribution Function

We analyze the estimated cumulative distribution function ($ECDF$) [1], [2], [3] computed based on $\frac{F}{k}$ over $log_{10}FEs$. The $ECDF(FEs, \frac{F}{k} \leq 0)$ represents the fraction of runs which reach a value of $\frac{F}{k}$ less than or equal to 0 for a given ellapsed runtime measured in FEs. The $ECDF$ is always computed over the runs of an experiment for a given benchmark instance. If runs for multiple instances are available, we aggregate the results by computing their arithmetic mean. The x-axis does not represent the values of FEs directly, but instead $log_{10}FEs$. The $ECDF$ is always between 0 and 1 – and the higher it is, the better.