Raw run-time data (each spike one run)
RTD Graphs
Protocol for obtaining the empirical RTD for an LVA $A$ applied to a given instance $\pi$ of a decision problem:

- Perform $k$ independent runs of $A$ on $\pi$ with cutoff time $t'$. (For most purposes, $k$ should be at least 50–100, and $t'$ should be high enough to obtain at least a large fraction of successful runs.)

- Record number $k'$ of successful runs, and for each run, record its run-time in a list $L$.

- Sort $L$ according to increasing run-time; let $rt(j)$ denote the run-time from entry $j$ of the sorted list ($j = 1, \ldots, k'$).

- Plot the graph $(rt(j), j/k)$, i.e., the cumulative empirical RTD of $A$ on $\pi$. 
Distribution of median search cost for WalkSAT/ SKC over set of 1000 randomly generated, hard 3-SAT instances:
RTDs for WalkSAT/SKC, a prominent SLS algorithm for SAT, on three hard 3-SAT instances:
Approximation of an empirical RTD with an exponential
distribution $\text{ed}[m](x) := 1 - 2^{-x/m}$:
Approximation of an empirical RTD with an exponential distribution $ed[m](x) := 1 - 2^{-x/m}$.

The optimal fit exponential distribution obtained from the Marquardt-Levenberg algorithm passes the $\chi^2$ goodness-of-fit test at $\alpha = 0.05$. 
RTD Approximation with Mixture of Exponential Distributions

CP1(#815,#74)

0.49*ed[7000]+0.51*ed[10^7]
Example of an empirical RTD of an SLS algorithm on a problem instance for which static restarting is effective:

‘ed[18]’ is the CDF of an exponential distribution with median 18; the arrows mark the optimal cutoff-time for static restarting.
Performance differences detectable by the Mann-Whitney U-test for various sample sizes (sign. level 0.05, power 0.95):

<table>
<thead>
<tr>
<th>sample size</th>
<th>$m_1/m_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3010</td>
<td>1.1</td>
</tr>
<tr>
<td>1000</td>
<td>1.18</td>
</tr>
<tr>
<td>122</td>
<td>1.5</td>
</tr>
<tr>
<td>100</td>
<td>1.6</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

$m_1/m_2$ is the ratio between the medians of the two empirical distributions.
Example of crossing RTDs for two SLS algorithms for the TSP applied to a standard benchmark instance (1000 runs/RTD):
Correlation between median run-time for two SLS algorithms for the TSP over a set of 100 randomly generated instances:

10 runs per instance.