# Automated fitting using Anderson-Darling tests

This document describes an algorithm for automated selection of a distribution from a set of feasible distributions. This document details the theory and justification behind algorithm.

The candidate distributions are those that are defined in HAMISH Reliability Basics:

* Weibull
* Normal
* Exponential
* Rayleigh
* Extreme Value
* Log Normal

and we employ the Anderson-Darling (AD) test to assess the adequacy of the distribution. For a given distribution, , the AD test has the form

If is unknown, it is estimated and the AD test statistic is computed. If the test statistic is such that there is only a probability (typically ) of a more extreme test statistic, is rejected (and so is the distribution).

The algorithm is as follows:

1. **Input** the candidate distributions, and the failure time data to be used for fitting,
2. Compute Anderson-Darling p-values for all .
3. If all p-values then **Return** “none”. Else, Select with highest p-value (low indicates a poor fit)
4. If the selected distribution is Weibull:
   1. If shape parameter confidence interval includes *and* is accepted for the Exponential distribution, set .
   2. If shape parameter confidence bounds include *and* is accepted for the Rayleigh distribution: set
   3. Else set
5. **Return**  as well as a list of all distributions with p-value .

This algorithm was implemented in the MATLAB code selectDistributionAD.m and the code demoAD.m (both of which can be found in the same folder as this document).

In demoAD.m, a series of random “failure times” are generated and then the above algorithm is applied to select the distribution. A few results are demonstrated:

* Fig. 1 shows the result of applying the algorithm to sets of failure time data of size when the true underlying distribution is a Weibull with scale = 200 and shape = 5.
* Fig. 2 shows the result of applying the algorithm to sets of failure time data of size when the true underlying distribution is a Weibull with scale = 200 and shape = 5. We can see that fewer mistakes are made in the selection of the distribution.
* Fig. 3 shows the result of applying the algorithm to sets of failure time data of size when the true underlying distribution is a Weibull with scale = 200 and shape = 2. We can see that the Rayleigh distribution is correctly selected in the majority of the cases.
* Fig. 4 shows the result of applying the algorithm to sets of failure time data of size when the true underlying distribution is uniform. We can see that the algorithm correctly concludes that no distribution fits the data.



Fig. 1



Fig. 2



Fig. 3



Fig. 4