

# HW0: Fitting Distributions with Maximum Likelihood

CS249 — Spring 2016 — D.S. Parker

The basic problem here is to determine, when given an input sequence of real values, which distribution it follows. More specifically, for this assignment you are to develop a notebook that reads in a numeric table, and – for each dataset (i.e., each column in the table) – determines the distribution and parameters that gives the closest match to it.

For example, your notebook could be given an input table like this:

D1	D2	D3	D4	D5	D6
7.1378018	6.8581740	0.2379494	0.1476523	5.3174948	3.1291521
3.3713903	6.2437282	0.2138276	0.1699299	1.5583491	0.6543210
⋮	⋮	⋮	⋮	⋮	⋮
8.7408465	2.7770890	2.7271062	0.5956197	2.3983975	8.8628316

The columns of this table define six datasets. Your notebook should produce a CSV file `HW0_output.csv` giving distributions that (it thinks) best fit the data. A correct output file could then look like this:

```
gamma,9,2,  
normal,4,2,  
lognormal,0,1  
exponential,1,,  
chi-squared,5,,  
logistic,3,2,
```

For simplicity, the parameters used in this assignment will always be integers, so the printed output should always have integer parameter values.

Your notebook can determine the distribution that fits best in any way you like. However, the `fitdistr()` function in the `MASS` library gives a simple way to do this. The notebook describes the assignment in much more detail, and gives orientation about how to do things in R.

In other words: yes, this is a simple assignment. It is intended as a warmup.

After running your notebook on the test input file `HW0_test.csv`, to complete this assignment please upload two files to CCLE:

1. your output CSV file `HW0_output.csv`
2. your notebook file `HW0_Fitting_Distributions.ipynb`

The notebook should have the commands you used to produce the output file. All assignment grading in this course will be automated, so please assume that when uploading files.

See the HW0 notebook for more details.