Maximum Likelihood Estimation

Prepared by Rebecca E. Curtis

Definition: Maximum likelihood estimation (MLE) is a popular statistical method used to make inferences about parameters of the underlying probability distribution of a given data set. It looks at a large class of distributions and then chooses the "best" distribution. For each distribution, a likelihood is computed, and the best distribution is the one that maximizes this likelihood. The method was pioneered by geneticist and statistician Sir Ronald A. Fisher between 1912 and 1922.

Given a discrete probability distribution D with known probability mass function f_D and distributional parameter θ, a sample may be drawn $X_1, X_2, \ldots X_n$ of n values from this distribution and then using the mass function the probability associated with the observed data is computed:

$$\mathbb{P}(\text{we sample values } x_1, x_2, \ldots, x_n) = f_D(x_1, \ldots, x_n \mid \theta)$$

However, it may be that the value of the parameter θ is unknown despite knowing (or believing) that the data comes from the distribution D. How should θ be estimated? It is a sensible idea to draw a sample of n values $X_1, X_2, \ldots X_n$ and use this data to help make an estimate.

Once there is a sample $X_1, X_2, \ldots X_n$, an estimate of the value of θ from that sample can be sought. MLE seeks the most likely value of the parameter θ (i.e. we maximize the *likelihood* of the observed data set over all possible values of θ). This is in contrast to seeking other estimators, such as an unbiased estimator of θ, which may not necessarily yield the most likely value of θ but which will yield a value that (on average) will neither tend to over-estimate nor under-estimate the true value of θ.

To implement the MLE method mathematically, the *likelihood* is defined:

$$\operatorname{lik}(\theta) = f_D(x_1, \ldots, x_n \mid \theta)$$

and maximize this function over all possible values of the parameter θ. The value $\hat{\theta}$ which maximizes the likelihood is known as the maximum likelihood estimator (MLE) for θ.

There are only two drawbacks to Maximum Likelihood Estimation, but they are important ones:

- With small numbers of failures (less than 5, and sometimes less than 10 is small), MLE can be heavily biased and the large sample optimality properties do not apply
- Calculating MLE often requires specialized software for solving complex non-linear equations. This is less of a problem as time goes by, as more statistical packages are upgrading to contain MLE analysis capability every year.

References/Sources:

Aldrich, J. 1997. R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3): 162–176.

Eliason, S.R. 1993. *Maximum Likelihood Estimation: Logic and Practice*. Sage Publications.

Harris, J. W. and Stocker, H. 1998. Maximum Likelihood Method. *Handbook of Mathematics and Computational Science*. New York: Springer-Verlag.

Hoel, P.G. 1962. *Introduction to Mathematical Statistic - 3rd ed.* New York: Wiley.

Press, W.H. et. al. 1992. Least Squares as a Maximum Likelihood Estimator. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge: Cambridge University Press.