Handout 16: Maximum Likelihood Estimators

The method of maximum likelihood estimation has a somewhat more complex history than that of the method of moments. The basic idea of maximum likelihood appeared in the writings of the German mathematician Carl Friedrich Gauss (1777− 1855), the French mathematician Pierre-Simon, Marquis de Laplace (1749− 1827), Thorvald Nicolai Thiele (1838− 1910), a Danish astronomer, actuary, and mathematician and Francis Ysidro Edgeworth (1845− 1926), the Irish philosopher and political economist who made significant contributions to statistical theory, even though, as a mathematician, he was largely self-taught. Thus, the approach to estimation has traceable roots as far back as the 1700s and the main idea of the approach was reasonably well known by the early 1900s. But it is the British statistician and geneticist Sir Ronald A. Fisher (1890− 1962) who named the method, made it the centerpiece of his approach to statistical estimation, and led to its widespread adoption as the preferred method of statistical estimation on the basis of its asymptotic behavior.

The basic idea behind maximum likelihood estimation is extraordinarily simple. The term "maximum likelihood" is a perfect description for the meaning and intent of the approach. It simply answers the question: "What value of the parameter would make my experimental outcome most likely to occur?" It seems quite reasonable to expect that your guess at an unknown parameter should be as compatible as possible with the data on which your guess is based. Formally, we have:

$$\widehat{\theta} \in \underset{\theta \in \Theta}{\arg\min} \left\{ f_\theta(x_1, \ldots, x_n) \right\}$$

We will call the function inside of the argmax the likelihood of the data, and use the following notation that stresses its dependence on the value of θ:

$$\mathcal{L}(\theta; x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n)$$

As it is often useful to maximize the logarithm of the likelihood function, we have a special notation of this logarithm using a lower case l:

$$l(\theta; x_1, \ldots, x_n) = ln\left(\mathcal{L}\right)$$

Maximum likelihood estimators have stronger theoretical properties than method of moment estimators.[1] What makes them slightly more difficult to calculate is the need to work with the joint density of the data.

[1] Take MATH330 next Fall if you're interested!