MLEs That Cannot Be Put In Closed Form

The notion of equations that do not have an analytical “solution" often seems odd when a person first encounters the issue. To motivate this matter we will consider a model useful in estimating the size (*N*) of a closed population (e.g., the 631 marbles in the goldfish bowl). We will illustrate the use of Darroch's (1958) MLE for population size (*N*) under the assumption of closure. Here, one captures (or recaptures) animals on t occasions in a situation mimicking the goldfish bowl containing marbles. The size of the population remains fixed (no births, deaths, none come into the bowl during the study and none leave the bowl during the study (closure)). At time 1 you get a sample of size n_1, then at times 2, 3, 4, ..., t you get samples of size n_2, n_3, n_4, ... n_t. The total number of individuals caught at least once (i.e., the total number of unique animals handled) is denoted M_{t+1} (for reasons that will be a bit obvious latter).

Darroch's model is discussed in Otis et al. (1978: 24-28 and 106-107). The model for this case is multinomial (under the usual assumptions) and Darroch (1958) found the MLE by taking the derivative of the log-likihood function with respect to *N* and setting it equal to 0. After some algebra, the MLE was found to be the equation,

$$\left(1 - \frac{M_{t+1}}{N}\right) = \left(1 - \frac{n_1}{N}\right) \cdot \left(1 - \frac{n_2}{N}\right) \cdot \left(1 - \frac{n_3}{N}\right) \cdot \cdot \cdot \left(1 - \frac{n_t}{N}\right) .$$

It is not possible to “solve" this equation so that only the parameter *N* appears on the left hand side (LHS), while all the other terms (representing data) appear on the RHS. Thus, the estimator cannot be expressed in closed form! Still, it is an equation and does have a solution; however, the solution must be found numerically. That is, numerical, iterative methods must be employed to find the value of *N* that “solves" this equation. That value of *N* is the MLE and denoted as $\hat{N}$.

Program *MARK* has an option to allow you to get more familiar with using simple (read “dumb") trial and error methods to solve this equation. This option is based on the following data:

$\boldsymbol{n_1 = 30, n_2 = 15, n_3 = 22\ n_4 = \mathrm{n}_t = 45,}$ **and** $\boldsymbol{M_{t+1} = 79}$ **.**

Thus, one wants the value of *N* that “solves" the equation,

$$\left(1 - \frac{79}{N}\right) = \left(1 - \frac{30}{N}\right) \cdot \left(1 - \frac{15}{N}\right) \cdot \left(1 - \frac{22}{N}\right) \cdot \left(1 - \frac{45}{N}\right) .$$

One could try to solve this equation by "trial and error." That is, one could plug in a guess for population size and see if the LHS = RHS (not very likely unless you can guess very well). Thinking about the problem a bit, one realizes that $N \geq M_{t+1}$. So, at least, one has a lower bound (in this case, 79 if we restrict the parameter space to integers).

If the first guess for *N* does not satisfy the equation, one could try another guess and see if that either (1) satisfies the equation or (2) is closer than the first guess. The log-likelihood functions are unimodal (for the exponential family), thus, you can usually make a new guess in the right direction. *MARK* allows you do make guesses and shows a plot of your progress. Anyway, one could keep making guesses until a value of *N* (an integer) allows the LHS = RHS, and take this value as the MLE, $\hat{N}$. Clearly, the trial and error method will unravel if there is more than 1 or 2 parameters. Likewise, plotting the log-likelihood function is useful only when 1 or 2 parameters ae involved. We will quickly be dealing with cases where there are 30-40 parameters, thus we must rely on efficient computer routines for finding the maximum point in the multidimensional cases.

Another possibility for the 1-dimensional case would be to start at M_{t+1} as *N* and check to see if the equality was achieved, if not, add one, recompute *N* and check for equality again, and so on. Cleaver search algorithms have been devised for the 1-dimensional case (e.g., the golden search is very good and the best is the Fibonacci search). Computers are great at such routine computation and the MLE in this case can be found very quickly.

You should begin to expect estimators that cannot be put is closed form and rely on computer software to compute MLEs numerically, using efficient methods.