Statistica Sinica **17**(2007), 839-840

Highlights

Thirty Years of EM and Much More

Once again we are featuring a self-organized theme, by grouping papers from the backlog. And this time the theme coincides with an anniversary – it has been 30 years since the publication of the seminal paper on the EM algorithm by Dempster, Laird and Rubin in the *Journal of the Royal Statistical Society, Series B*. The last 30 years certainly have witnessed the tremendous growth of statistics, both as a scientific discipline and a practical profession. Perhaps the phrase "modernization" is not too trite here because the recent advances are often contrasted with "classic statistics," however defined, or not defined.

A significant part of our modernization is undoubtedly due to the rapid development of powerful computational methods such as the EM algorithm and Markov chain Monte Carlo (MCMC) algorithm. Actually, neither of these is an "algorithm" in its traditional sense. Rather, they are principles, or very general recipes, for constructing effective algorithms for solving many specific classes of problems. The adjective "effective", however, does not come for free. Much of the research over the past 30 years on EM, and the past 15 years or so on MCMC, is about efficiency in both the narrow and broad sense. Efficiency in the narrow sense refers to pure algorithmic and computational considerations, such as rate of convergence or CPU time, things that often are not a statistician's "cup of tea". In contrast, we statisticians are particularly good at achieving efficiency in a broader sense, that is, achieving a good balance among computational, statistical, and even human efficiency. This balance is critical to maintain our discipline's central role in quantitative sciences, for which a theoretically optimal, but practically inferior, approach is often ignored.

Indeed, the popularity of the EM algorithm is a good testament to this statement. If one only considers the theoretical rate of convergence, then there should be little place for EM-type algorithms because they are known to have only a linear rate, or even a sub-linear rate in some cases. In contrast, the Newton-Raphson (NR) algorithm has the mathematically proven quadratic convergence rate. So why would anyone prefer EM over NR? Here I have a confession to make, a confession that could potentially get my Harvard degree revoked (but I have to do this in the name of science!). I was asked to find an MLE via NR on a qualifying exam. Out of desperation, as I just could not get NR to converge, I cheated – I programmed an EM-type linear iteration. It indeed ran slowly in terms of the CPU time but it converged.

I then perturbed the value found by my "slow" algorithm by an epsilon, and used it as the starting point for my NR algorithm. This time my NR iteration converged immediately, to the same solution as found by the much slower linear iteration! Or should I say "much faster" if we take into account the total amount of human time, not just the computer time, spent on each algorithm?

The MCMC revolution in statistics since 1990 is another testament to our professional strength of staying at the core of quantitative sciences. As we know, MCMC was invented by physicists more than half a century ago. To this day, physicists continuously stand at the frontier of MCMC methods, inventing repeatedly powerful yet simple methods; recent examples include the Wang-Landau algorithm. However, we statisticians can proudly claim credit for fostering the general use of MCMC, because we are the ones who demonstrated the versatility and power of MCMC, and showed how it can be used for fitting realistic statistical models in biological, medical, and social sciences, engineering, and even humanities.

The nine papers in the theme are only a tiny sample of the vast literature on EM and MCMC – we grouped them together because the two "recipes" are intimately related, with EM taking care of deterministic mode search (e.g., MLE; posterior mode), and MCMC handling distribution sampling, including stochastic mode localization. Technically, both share many similar convergence features – much of my own research on MCMC was directly motivated by transferring the methods developed for speeding up the EM-type algorithms to that for MCMC algorithms. And undoubtedly both have been workhorses for statistical computing, and indeed more generally for scientific computing.

If you are disappointed by the articles here because they do not cover topics you are interested in, you can help to remedy the deficiency by submitting your paper(s) to *Statistica Sinica*. In return, Michelle and I can assure you of the efficiency of the journal. That is, with 95% confidence or credibility (see the Highlights of the April issue this year), we will send you our decision on your submission within four months of receiving it, and often much earlier if the decision is negative (so "no news" is good news!). If you are skeptical about our promise, perhaps because you have heard too many empty ones, please talk to a colleague who has had some recent "Sinica experience." If you are convinced, send us your submissions. If not, write us via "Feedback and Suggestions" at **http://www.stat.sinica.edu.tw/statistica/** and tell us why your colleague's experience was not positive. Either way, we hope to hear from you, and we thank you for your support to *Statistica Sinica*, as an author, a reviewer, or a reader.

— Xiao-Li Meng