Some remarks about maximization of the likelihood function

Adelchi Azzalini
Università di Padova, Italia

CFE-CMStatistics Meeting, King's College London
17–19 December 2022

Maximum likelihood estimation

- MLE: maximization of the log-likelihood (logL), in practice
- In most cases, only feasible via numerical computation
- Various issues exist (anomalies, multiple maxima, etc.)
- Here we focus on alternative computational strategies:
  which is the 'best method' to search for the MLE?

Two most popular approaches

- direct numerical maximization: DNM (in many forms)
- expectation-maximization: EM and its variants

Direct numerical maximization (DNM)

A range of possibilities:

- using function values only (notably Nelder and Mead)
- using also gradient (several quasi-Newton methods)
- using also Hessian matrix

Tools in the R environment:
optim, nlminb, nlm, constrOptim, ... et cetera

This was the primary approach up to about 1980. Then ...

EM and its progeny

- Enormous impact of EM on statistics
- 'EM family' has grown a lot
  (as for variants/enhancements and applications)
- "many statisticians are deeply in love with the EM algorithm "
  (Speed, 2008)
- Being in love is a *great* feeling!
- . . . unless it makes you blind

Too much EM?

- Some authors focus entirely on EM-type algorithms
- For instance, there are cases where the authors obtain explicit expression of logL (and possibly its 1st and even 2nd order derivatives), but do not consider them for DNM
  . . . develop an EM algorithm instead, no matter how complex
- MacDonald (2014, 2021) argues against blind usage of EM, providing numerous examples where DNM performs generally better than EM in term of computing speed and simplicity of development

EM is great, but. . .

- EM is a fundamental tool
- . . . but it cannot be the *only* tool.
- According to one of its 'inventors',
  EM has "probably been over-touted"
  (as quoted by MacDonald, 2021)
- Sometimes non-EM methods are too quickly discarded
- A point to consider:
  EM may require much time, notably human time!
- Human time is precious. Use it responsibly.
- . . . especially when a model is at the development stage.

Miscellaneous remarks

- In many cases, DNM techniques return also che Hessian matrix of logL, leading to standard errors of the estimates.
- In various cases, solution of logL equations admits an explicit solution for a subset of the parameters, leading to reduction of dimensionality of the search.
- Constraints on parameter space can often be handled by simple parameter transformations.
- "Although we have no intention of detracting from EM algorithms, their dominance over MM algorithms is a historical accident." (Hunter and Lange, 2004)
- IRLS is best known for GLM's fitting, but its potential usage is much wider. (e.g. Holland and Welsch, 1977; Green, 1984)

No BEM

- The presented view is *not* against EM usage
- it is against BEM (Blind EM) usage

References

- P.J. Green (1984). JRSS-B, 46, 149–182
- P.W. Holland & R.E. Welsch (1977). Comms. Stat TM, 6, 813–827
- D.R. Hunter and K. Lange (2004). Amer.Statistician, 58, 30–37
- I.L. MacDonald (2014). Int. Stat. Rev., 82, 296–308
- I.L. MacDonald (2021). AStA Adv. Stat. Analysis, 105, 629–647
- T. Speed (2008). Bull. IMS, 37(9), p.14