Bayesian data analysis – reading instructions 13

Aki Vehtari

Chapter 13: Modal and distributional approximations

Chapter 4 presented normal distribution approximation at the mode (aka Laplace approximation). Chapter 13 discusses more about distributional approximations.

Outline of the chapter 13

- 13.1 Finding posterior modes
  - Newton's method is very fast if the distribution is close to normal and the computation of the second derivatives is fast
  - Stan uses limited-memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS) which is a quasi-Newton method which needs only the first derivatives (provided by Stan autodiff). L-BFGS is known for good performance for wide variety of functions.
- 13.2 Boundary-avoiding priors for modal summaries
  - Although full integration is preferred, sometimes optimization of some parameters may be sufficient and faster, and then boundary-avoiding priors maybe useful.
- 13.3 Normal and related mixture approximations
  - Discusses how the normal approximation can be used to approximate integrals of a a smooth function times the posterior.
  - Discusses mixture and t approximations.
- 13.4 Finding marginal posterior modes using EM
  - Expectation maximization is less important in the time of efficient probabilistic programming frameworks, but can be sometimes useful for extra efficiency.
- 13.5 Conditional and marginal posterior approximations
  - Even in the time of efficient probabilistic programming, the methods discussed in this section can produce very big speedups for a big set of commonly used models. The methods discussed are important part of popular INLA software and are coming also to Stan to speedup latent Gaussian variable models.
- 13.6 Example: hierarchical normal model
- 13.7 Variational inference
  - Variational inference (VI) is very popular in machine learning, and this section presents it in terms of BDA. Auto-diff variational inference in Stan was developed after BDA3 was published.
- 13.8 Expectation propagation
  - Practical efficient computation for expectation propagation (EP) is applicable for more limited set of models than post-BDA3 black-box VI, but for those models EP provides better posterior approximation. Variants of EP can be used for parallelization of any Bayesian computation for hierarchical models.
- 13.9 Other approximations

- Just brief mentions of INLA (uses methods discussed in 13.5), CCD (deterministic adaptive quadrature approach) and ABC (inference when you can only sample from the generative model).

13.10 Unknown normalizing factors

- Often the normalizing factor is not needed, but it can be estimated using importance, bridge or path sampling.