

Stan, HMC, etc.

10 Apr 2023

Table of contents

MCMC, review	1
HMC	2
autodiff (“algorithmic”)	2
diagnostics	3
divergences	3
centered and non-centered parameters	4
challenges	4

MCMC, review

- detailed balance: $\pi_i p_{ij} = \pi_j p_{ji}$
 - MCMC mapping is $\int \pi_x p_{xy} dy$
 - integrate LHS wrt i , RHS wrt j (p. 328 of Tierney’s notes)
- implies that π is the stationary distribution
- also need aperiodicity to get to a **unique** stationary distribution
- technical conditions for “fast enough” convergence, CLT applying, etc.

Tierney’s notes

- standard tricks
 - sampling from conjugate priors

- sequential (Gibbs) sampling/conditional distributions
- Metropolis-Hastings
- **data augmentation:** like E-M but stochastic at both steps:
 - sample expected values of missing data/latent variables from their *conditional posterior distributions* (instead of taking expectation)
 - sample parameter values from *their* conditional posterior distribution (instead of maximizing)
- e.g. impute missing values on the fly

HMC

- Radford Neal’s 1995 thesis is [here](#) (Wayback Machine); also published by Springer (Neal 2012)
- augment position (current parameter values) with “momentum”; randomly perturb momentum at each step, integrate dynamics
- leapfrog integration: have to pick stepsize/number of steps
- No-U-Turn sampler: integrate until trajectory starts to turn back
- [animations](#)

Neal, Radford M. 2012. *Bayesian Learning for Neural Networks*. Vol. 118. Springer Science & Business Media.

autodiff (“algorithmic”)

- magic technology: “the evaluation of a gradient requires never more than five times the effort of evaluating the underlying function by itself”
- operator overloading
- reverse mode (best when we have a mapping from $R^n \rightarrow R$)

([Wikipedia](#)):

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial w_{n-1}} \frac{\partial w_{n-1}}{\partial x} \quad (1)$$

$$= \frac{\partial y}{\partial w_{n-1}} \left(\frac{\partial w_{n-1}}{\partial w_{n-2}} \frac{\partial w_{n-2}}{\partial x} \right) \quad (2)$$

$$= \frac{\partial y}{\partial w_{n-1}} \left(\frac{\partial w_{n-1}}{\partial w_{n-2}} \left(\frac{\partial w_{n-2}}{\partial w_{n-3}} \frac{\partial w_{n-3}}{\partial x} \right) \right) \quad (3)$$

$$= \dots \quad (4)$$

- lots of other engines (PyTorch, JAX, ...)

diagnostics

- assuming an AR1 model,

$$\text{SD}(\hat{\beta}) = \frac{\text{SD}(\beta|z)}{\sqrt{N}} \sqrt{\frac{1 + \rho_\beta}{1 - \rho_\beta}}$$

- effective sample size = $N(1 - \rho)/(1 + \rho)$ (AR1), $N(\sum \rho_k)^{-1}$ more generally
- efficiency is ESS/N
- \hat{R} (Gelman-Rubin statistic: potential scale-reduction factor), improved \hat{R} (Vehtari et al. 2021; Lambert and Vehtari 2022): R code [here](#)

- sensitivity to chains with different variances, infinite means
- compare within- and between-chain variances
- at least 4 chains
- threshold of 1.01
- improved ESS

divergences

- energy changes too much

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. “Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC (with Discussion).” *Bayesian Analysis* 16 (2): 667–718. <https://doi.org/10.1214/20-BA1221>.

Lambert, Ben, and Aki Vehtari. 2022. “ \hat{R}_* : A Robust MCMC Convergence Diagnostic with Uncertainty Using Decision Tree Classifiers.” *Bayesian Analysis* 17 (2): 353–79. <https://doi.org/10.1214/20-BA1252>.

centered and non-centered parameters

- funnels
- centered is better when groups are well characterized (“informative data”, large N per group), non-centered is better when joint prior contributes a lot (“noninformative data”, small N per group)
- ??performance of JAGS/Gibbs on 8-schools problem?

challenges

- high dimensionality (*always* hard)
- documentation
- debugging!
- resolving divergences
- discrete latent variables (“Rao-Blackwellization”, Robert and Roberts (2021)); marginalize/find conditional expectation
 - e.g. discrete mixture models
- speed

Robert, Christian P., and Gareth O. Roberts. 2021. “Rao-Blackwellization in the MCMC Era.” arXiv. <https://doi.org/10.48550/arXiv.2101.01011>.