Stochastic Gradient
Riemannian Langevin dynamics
on the
Probability Simplex

Sam Patterson (Gatsby, UCL)
Yee Whye Teh (Oxford) (presenter)

Poster: Fri32
Big Data and Bayesian Learning

- Large scale datasets are fast becoming the norm.
- Most current successes in scalable learning are optimization-based and non-Bayesian.

- Stochastic gradient Langevin dynamics:

$$x_t = x_{t-1} + \frac{\epsilon_t}{2} \left( \nabla_X \log P(x_{t-1}) + \frac{N}{n} \sum_{i=1}^{n} \nabla_X \log P(y_{S_i} | x_{t-1}) \right) + \mathcal{N}(0, \epsilon_t)$$

- Step-sizes $\epsilon_t \to 0$ slowly enough.

SGLD on Probability Simplices

- Many models defined over probability simplices, e.g. latent Dirichlet allocation.
- Complications:
  - High dimensional
  - Boundaries and constraints
  - Probability mass located close to boundaries.
  - Riemannian metric structure.
- Contribution:
  - Choice of parameterization
  - Choice of Riemannian metric
Learning LDA from Wikipedia

Test perplexity vs. number of articles processed for different methods: HSVG, OVB, and SGRLD.

OVB, HSVG
[Hoffman et al, Mimno et al]

SGRLD

Stochastic Gradient Riemannian Langevin Dynamics on the Probability Simplex
-- Patterson & Teh

Poster
Fri32