Handout journal club : Fast Mixing Markov Chains for Strongly Rayleigh Measures, DPPs, and Constrained Sampling

Presenter: Nicolas Brosse

13 April 2017

This handout is a brief introduction to the main ideas developped in [LSJ16]. Monte Carlo Markov Chains (MCMC) are algorithms that enable to sample from a probability distribution whose normalizing constant is unknown. Their interest is thus widely acknowledged in the Bayesian setting, in combinatorics and statistical physics. The basic idea behind MCMC methods is to create a Markov chain whose stationary distribution is the target distribution. It is fairly easy to create such a Markov chain; besides, its simulation does not need the knowledge of the normalizing constant (see Metropolis-Hastings algorithm on Wikipedia). Of course, numerous other Markov chains are eligible, in the sense that they have the target distribution as stationary distribution. One of the most famous one, besides the Metropolis-Hastings algorithm, is the Gibbs sampler (see the webpage on Wikipedia). The principal problematic with MCMC methods is the analysis of the number of samples necessary to reach the equilibrium (that is, the stationary distribution).

[LSJ16] answers this question in a particular framework. The state space is finite of the form 2^V, where $V = \{0, \ldots, N\}$. The target distribution is $\pi_{\mathcal{C}} \propto \exp(\beta F(S)) \mathbb{1}\{S \in \mathcal{C}\}$ where $S \in 2^V$, $F : 2^V \to \mathbb{R}$ and $\mathcal{C}$ is a set of constraints. Denote by π the unconstrained distribution. The cases tackled by the authors are:

1. π is a strongly Rayleigh distribution.
2. $\mathcal{C}$ is a set of bases of a special matroid, i.e. $\mathcal{C} = \left\{S \in 2^V \,|\, |S| = k\right\}$ for $k \in \{0, \ldots, N\}$ and $|S|$ is the cardinal of S, or S obeys a partition constraint.
3. $|S| \leq k$.

For applications, see the introduction of [LSJ16]. The first case covers Determinantal Point Processes (DPP) used in machine learning (see the references in the article). The constraints for the two other cases are understandable. A matroid is like a "group" structure, that guarantees that the Markov chain (here a Gibbs sampler) stays in the constraint set.

To prove convergence of a Markov chain, numerous techniques exist. In this article, spectral methods are used and we refer the reader to [LPW09, chapters 12,13,14] for an introduction. The book is available online. The aim is to lower bound the spectral gap for the two first cases and a coupling argument is developped for the third case. For all cases, they rely on the Gibbs sampler. The interested reader should look at the detailed proofs of [LSJ16] in the supplementary material. More precisely, the strategies used by the authors are

1. direct application of [AGR16] using a symmetrisation procedure on π that preserves its strongly Rayleigh property. Clever but it is a "trick" and the curious reader should better read [AGR16]. Strongly Rayleigh distributions exhibit the strongest form of negative dependence and they enjoy nice properties. Their definition is quite recent and we refer to [BBL09] for their main properties. To know more about negative dependence, one can also read [Pem00], [PP14].

2. relies on a multicommodity flow argument to lower bound the spectral gap. See [Sin92] and [GHK15]. The convergence properties of the Gibbs sampler are "translated" in a problem of flows and capacities on a graph. The proof is based on combinatorics and is quite elegant. See also [LPW09, Section 13.4].

3. relies on a coupling argument, see [LPW09, Chapter 14]. The coupling is quite intricate.

Except for the strongly Rayleigh distribution, the bound on the number of samples to reach equilibrium depends on unknown quantities related to F. It could be useful to study cases where these quantities are explicitly known (under additional assumptions if necessary).

References

[AGR16] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. "Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes". In: *Proceedings of the 29th Conference on Learning Theory, COLT*. 2016, pp. 23–26.

[BBL09] Julius Borcea, Petter Brändén, and Thomas Liggett. "Negative dependence and the geometry of polynomials". In: *Journal of the American Mathematical Society* 22.2 (2009), pp. 521–567.

[GHK15] Alkis Gotovos, Hamed Hassani, and Andreas Krause. "Sampling from probabilistic submodular models". In: *Advances in Neural Information Processing Systems*. 2015, pp. 1945–1953.

[LPW09] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[LSJ16] Chengtao Li, Suvrit Sra, and Stefanie Jegelka. "Fast Mixing Markov Chains for Strongly Rayleigh Measures, DPPs, and Constrained Sampling". In: *Advances in Neural Information Processing Systems*. 2016, pp. 4188–4196.

[Pem00] Robin Pemantle. "Towards a theory of negative dependence". In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1371–1390.

[PP14] Robin Pemantle and Yuval Peres. "Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures". In: *Combinatorics, Probability and Computing* 23.01 (2014), pp. 140–160.

[Sin92] Alistair Sinclair. "Improved bounds for mixing rates of Markov chains and multicommodity flow". In: *Latin American Symposium on Theoretical Informatics*. Springer. 1992, pp. 474–487.