Paper: **Stochastic Variational Inference (JMLR 2014)**
URL: http://www.columbia.edu/~jwp2128/Papers/HoffmanBleiWangPaisley2013.pdf
Authors: **Matthew Hoffmann, David Blei, Chong Wang, John Paisley**

1. Summary

This paper proposes SVI (Stochastic Variational Inference), a new algorithm for approximating the posterior distribution in large-scale bayesian inference. Effectiveness of the model is demonstrated by applying it in both the parametric and non-parametric settings to the LDA (Latent Dirichlet Allocation) task on real world datasets. Experiments indicate SVI outperforms the traditional variational inference method.

2. Strong points and main contributions

The key selling point of the paper is the observation authors make that the expectation term in the global variational update step can be decomposed in a way to allow for stochastic noisy estimates to be used. This allows them to develop an algorithm that scales independently with the size of the dataset and enables easy parallelization.

3. Weak points

The paper also has a few weak points - SVI algorithm is not naturally suited to streaming data, i.e.. data that is arriving in an online fashion. Their stochastic algorithm proceeds by uniformly sampling a random document d_i from the pool of documents $D_{1...N}$ and creating a fake corpus consisting of N replicas of d_i. Clearly, this is not possible to do if the number of documents N (i.e., the size of D) is not known in advance.
Another question to ponder about is the central assumption made in the paper that the complete conditionals belong to an exponential family. Although this assumption might be reasonable in a majority of situations, it is worth investigating how to develop an SVI algorithm when this assumption does not hold.

4. Technical soundness of the approaches

The key technical aspects of the paper - approach behind deriving the noisy gradients and using them in the traditional variational inference algorithm are sound and explained well.

5. Experimental Methodology

The authors compare SVI on three datasets - the largest of which is the entire set of documents on wikipedia consisting of 3.8M documents. These do not completely qualify as big data sets as they can be processed on a single machine, at the most they would need shared-memory parallelism. For this scope of this paper, these datasets are reasonably big to demonstrate the effectiveness of their new algorithm, however - it would be good to look at distributed memory implementation of SVI and the challenges associated with that.

6. Organization and Presentation

The paper is organized in a reasonable way to help the reader follow the contents, however the notations used in the paper are very hard to follow along. There are too many symbols being used and the biggest problem seems to be that it is not clear what variables are being conditioned on, as they authors mention that they drop variables occasionally (to keep things simple). It would be much easier to follow if they simplified the notations and either carried forward the derivations without dropping any variables *or* stuck to the same convention on dropping variables throughout the paper.