Reflection - ADVI - Kucukelbir et al.
Nicholas Hoernle   **February 21, 2018**

Kucukelbir extends the VI approximation algorithm aiming to reduce the 'user time' involved in deriving update equations. They introduce a tool for freeing a scientist to pose a generative model and iteratively evaluate the results produced by that model. My concern is that the diagnostics become even more opaque with this approach and the pitfalls introduced by Turner seem more apparent with ADVI (in that a scientist could be less aware of these pitfalls). Kucukelbir doesn't address possible diagnostics for identifying when the model is poorly specified or when the approximation is biased. The sensitivity to the transformation 'T' is also worrying as this transformation directly effects the shape of the approximating distribution.

Gelman (in BDA3) presents a philosophy for Bayesian reasoning where one should always care about the uncertainties of latent parameters (i.e. we should consider the full posterior for any inference task). An application could thus simply be assessing the significance of any treatment effect on a randomly sampled population. What is worrying about ADVI (and especially MF-ADVI), is that uncertainty quantities are given (by the approximating distribution) but these can be arbitrarily bad. Kucukelbir gives a case study where full-rank outperforms mean-field in terms of capturing more complicated posterior correlations among latent variables. This is not surprising given the assumptions built into MF and after the Turner discussion. However, even for full-rank ADVI, there is still no guarantee that the converged results even remotely resemble the target distribution and thus uncertainty statistics are unreliable.

In equation (5), we see that the objective $\mathcal{L}$ involves an expectation that would involve an intractable integral. Our goal is to calculate the gradient of this objective. To approximate the gradient via MC sampling, we need the gradient calculation inside the expectation. Elliptical standardization presents a convenient transformation of the variational parameters (with the Jacobian of the transformation being 1) that allows us to use MC sampling to approximate the gradient efficiently.