Reflection - Problems with Variational EM for Time Series - Turner 2011
Nicholas Hoernle **February 14, 2018**

(1) It is not necessarily *always bad* to use VB to approximate multimodal posteriors, one just has to be very careful to identify the problems that may arise when this is the case. The purpose/use of the model would have to be discussed to make an *always bad* statement but Turner shows that the VI approximation will tend to only approximate one of the modes of the posterior when the modes are significantly separated. This can lead to very misleading results, with the VI approximation having a smaller entropy than the true posterior and thus not only is the distribution badly approximated but the approximating model is also too optimistic about the uncertainty of the posterior. When the separation between the modes is less pronounced (i.e. there is a 'significant bridge' between the modes) then the variational approximation may still be useful.

(2) I believe Turner makes the case that it is not always better to have a larger set of 'Nice' distributions. This certainly can be the case, but he presents a convincing argument for restricting the factorising of the variational family to specific subsets (of the fully factored model) when certain correlation (time or chain dependency) properties are known about the posterior of the time series. He argues that structured approximations can yield a lower bias in the model than a more general approximation that achieves a tighter variational bound.

(3) The interesting relationship between the tightness of free energy and the bias is that the tightest bounds on free energy do not necessarily give the best results in terms of approximating the uncertainty in the model. The findings stress that it is important to evaluate the dependence of the variational bounds on the model parameters (as the free energy is maximised over the approximating distribution q and not over the actual log-likelihood of the parameters ($\log p(T|\theta)$)).