Reflection - Bayesian Analysis of Mixtures - Richardson & Green
Nicholas Hoernle   **March 19, 2018**

1. After seeing one data point (x_i) the posterior on μ is (as the normal distribution is symmetric and integrates to 1): $p(\mu \mid x_i) = \frac{\mathcal{N}(x_i|\mu,1)\times 1}{\int \mathcal{N}(x_i|\mu,1)d\mu} = \frac{\mathcal{N}(\mu|x_i,1)}{\int \mathcal{N}(\mu|x_i,1)d\mu} = \mathcal{N}(\mu \mid x_i, 1)$. Therefore, although the prior is 'improper' it still results in a posterior that is a valid distribution (this is certainly not always the case).

2. Using the improper prior in this case is invalid. Above we saw that with one datapoint, the posterior is valid as the datapoint is known to belong to the one Gaussian distribution (with unknown mean μ). Now, with a randomly sampled datapoint it could belong to cluster $z = 0$ and it could belong to cluster $z = 1$. Marginalizing over this ambiguity results in a posterior that does not integrate to 1. This relates to the discussion about the possibility that no observations are allocated to one or more components in a mixture (thus leaving a posterior that does not integrate to 1 if an improper prior is used).

3. Hierarchical bayesian models require the specification of hyperparameters. Sensitivity analysis is the study of the sensitivity of the posterior to perturbations in the values of hyperparameters. This is important as it can highlight where biases are introduced (sometimes unknowingly) through certain choices of the hyperpriors. It can also help to highlight model misspecifications if these sensitivities were not expected. This distinction is seen in Section 5 of the paper where the authors motivate the prior on σ_j based on the least sensitive model structure by placing a hyperprior on $\beta \sim \Gamma(g, h)$. Another example is for MCMC methods where one may wish to investigate the sensitivity to the initial starting parameters. We would like to ensure the convergence of the chain is invariant to the starting points.