Large-Sample Inference and Frequency Properties of Bayesian Inference:

- Home Work 2 – discussion

Normal Approximations to the posterior distribution

- Joint posterior
  - Convenient to approximate a unimodal and roughly symmetric posterior density by a normal distribution, centered at the mode.
    - Log posterior approximated as a quadratic function by a Taylor Series expansion.

$$\log p(q \mid y) = \log p(\hat{q} \mid y) + \frac{1}{2}(q - \hat{q})^T \left[\frac{d^2}{dq^2} \log p(q \mid y)\right]_{q=\hat{q}} (q - \hat{q}) + \cdots$$

      - (Mode must be in the interior of the parameter space, and assume that the derivatives exist._
      - First order term is zero since the first derivative at the mode is zero.

- Normal distribution with Unknown mean and variance
  - Using non-informative prior on $(m, \log s)$, expand its posterior around the mode and approximate it by a joint normal

- Interpretation of the posterior density function relative to the density at the mode

- Summarizing posterior distributions by point estimates and standard errors
  - Transformations to an appropriately defined functions of the parameters can improve the normal approximation

- Data reduction and summary statistics
  - Posterior mode and its curvature at the mode
  - If normal approximation not good, these summaries may be misleading.

- Lower dimensional normal approximations
    - For a finite sample size, the normal approximation is typically more accurate for conditional and marginals than for the full joint dist
        - Marginalizing leads to averaging over all other components of the parameter vector.
        - Bioassay Experiment
            - Approximation in the main body may be OK.
                - May lose some of the tail features (skewness, and tail approximation)

Large-Sample Theory

- Amount of data from some fixed sampling distribution, f(y), increases
- Asymptotic normality of the posterior distribution, even if the true distribution of the data is not within the parametric family under consideration
    - Modeled likelihood p(y|q) and prior p(q)
    - If f(y) belongs to this family for some q_0, then in addition to asymptotic normality, ‘consistency’ also holds.
    - If not, then the q_0 is replaced by a density in the class, that is closest to the true f(y) in the Kullback Leibler distance.
    - The Fisher information $J(q)$ is an important component of the asymptotic distribution
- Asymptotic Normality and Consistency
    - Appendix B has the proof under some regularity conditions
    - Continuous function of the parameter and the true q_0 is not on the boundary of the parameter space, twice differentiable
    - Express the second derivative at the mode in terms of the derivatives of the prior and the log-likelihood
    - As n increases the second term converges to its expectation
    - Likelihood dominates the prior in asymptotic sense
        - So with large amount of data, eliciting prior is not that important
        - For small amount of data, prior is critical

- With no data, quality of expert knowledge is highly relevant

Counter Examples to the Theorems

- Under-identified and non-identified parameters
  - Likelihood is same for many different points
    - Recognize the problem exists and get more info or put some constraints on the parameter space
- Number of parameters increasing with sample size
  - Neyman-Scott Problem
- Aliases
  - Mixture models may have two modes
- Unbounded likelihood
  - Mode does not exist
- Improper Prior distributions
  - May lead to improper posterior
- Priors that the point of convergence
- Convergence to the edge of the parameter space
- Approximation not good in the tails

Frequency evaluation of Bayesian Inference

- Large Sample correspondence leads to normal approximation
- Similar interpretation as Frequency theory inference
  - Consistency
  - Efficiency
  - Decision analysis
  - Asymptotic unbiasedness
  - Confidence coverage