STAT 625 - Advanced Bayesian Inference

Meng Li

Department of Statistics

Importance sampling (IS) I

- Importance sampling is used for computing expectations using a random sample drawn from an approximation to the target distribution.
- Let $p(\theta|y)$ be the posterior distribution density and $q(\theta|y)$ the unnormalized density.
- Suppose we are interested in $\mathrm{E}(h(\theta)|y)$, but we cannot generate random draws of θ from $p(\theta|y)$.
  - If we can generate random draws of θ from $p(\theta|y)$, then the integral can be evaluated by a simple average of simulated values.
- If $g(\theta)$ is a probability density from which we can generate random draws, then we have

$$\begin{aligned}\mathrm{E}(h(\theta|y)) = \int h(\theta|y)p(\theta|y)d\theta &= \frac{\int h(\theta)q(\theta|y)d\theta}{\int q(\theta|y)d\theta} \\ &= \frac{\int [h(\theta)q(\theta|y)/g(\theta)]g(\theta)d\theta}{\int [q(\theta|y)/g(\theta)]g(\theta)d\theta}\end{aligned}$$

Importance sampling (IS) II

- The above expression can be estimated based on S draws $\theta^1, \ldots, \theta^S$ from $g(\theta)$ by

$$\frac{\frac{1}{S}\sum_{s=1}^{S} h(\theta^s) w(\theta^s)}{\frac{1}{S}\sum_{s=1}^{S} w(\theta^s)},$$

  where

$$w(\theta^s) = \frac{q(\theta^s|y)}{g(\theta^s)}$$

  are called *importance weights*.
- We prefer some $g(\theta)$ such that q/g is roughly constant.
- We would like to avoid the case where importance weights vary substantially.
- The worse possible scenario is when the important weights are almost degenerate (are small with high probability but huge with a low probability). This happens, for example, if q has wide tails compared to g.

Efficiency of IS

- It is often helpful to examine a histogram of the logarithms of the largest importance weights.
- Estimates from importance sampling is often poor if the largest ratios are too large (relative to the average). Small importance weights are not a concern as they have little influence on the IS estimates.
- If the weights have finite variance, the effective sample size can be approximated by

$$S_{\text{eff}} = \frac{1}{\sum_{s=1}^{S} (\tilde{w}(\theta^s))^2},$$

where $\tilde{w}(\theta^s) = w(\theta^s) / \sum_{s'=1}^{S} w(\theta^{s'})$ are normalized weights

Importance resampling

- *Importance resampling* (also called sampling-importance resampling or SIR) can be used to obtain independent samples with equal weights.
- Once S samples are drawn from g, say $\theta^1, \ldots, \theta^S$, a sample of $k < S$ samples can be simulated as follows.
  1. Sample a value θ from $\{\theta^1, \ldots, \theta^S\}$ with respect to the probability vector $(\tilde{w}(\theta^1), \ldots, \tilde{w}(\theta^S))$.
  2. Sample a second value using the same procedure, but excluding the already sampled value from the set.
  3. Repeatedly sample without replacement $k-2$ more times.