Posterior Samplers: Importance Sampling

Frank Schorfheide
University of Pennsylvania
Econ 722 – Part 1

January 22, 2019
Bayesian Inference

• Ingredients of Bayesian Analysis:
  • Likelihood function $p(Y|\theta)$
  • Prior density $p(\theta)$
  • Marginal data density $p(Y) = \int p(Y|\theta)p(\theta)d\phi$

• Bayes Theorem:
  $$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \propto p(Y|\theta)p(\theta)$$

• Implementation: usually by generating a sequence of draws (not necessarily iid) from posterior
  $$\theta^i \sim p(\theta|Y), \quad i = 1, \ldots, N$$

• Algorithms: direct sampling, accept/reject sampling, importance sampling, Markov chain Monte Carlo sampling, sequential Monte Carlo sampling...
Importance Sampling

- Approximate $\pi(\cdot)$ by using a different, tractable density $g(\theta)$ that is easy to sample from.

- For more general problems, posterior density may be non-normalized. So we write

$$
\pi(\theta) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{f(\theta)}{\int f(\theta) d\theta}.
$$

- Importance sampling is based on the identity

$$
E_{\pi}[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \frac{\int_{\Theta} h(\theta) \frac{f(\theta)}{g(\theta)} g(\theta) d\theta}{\int_{\Theta} \frac{f(\theta)}{g(\theta)} g(\theta) d\theta}.
$$

- The ratio

$$
w(\theta) = \frac{f(\theta)}{g(\theta)}
$$

is called the (unnormalized) importance weight.
For $i = 1$ to $N$, draw $\theta^i \overset{iid}{\sim} g(\theta)$ and compute the unnormalized importance weights

$$w^i = w(\theta^i) = \frac{f(\theta^i)}{g(\theta^i)}.$$

Compute the normalized importance weights

$$W^i = \frac{w^i}{\frac{1}{N} \sum_{i=1}^{N} w^i}.$$  

An approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by

$$\bar{h}_N = \frac{1}{N} \sum_{i=1}^{N} W^i h(\theta^i) = \frac{1}{N} \sum_{i=1}^{N} W^i h(\theta^i).$$
If \( \theta^i \)'s are draws from \( g(\cdot) \) then

\[
E_{\pi}[h] \approx \frac{1}{N} \sum_{i=1}^N h(\theta^i) w(\theta^i), \quad w(\theta) = \frac{f(\theta)}{g(\theta)}.
\]
Since we are generating \( iid \) draws from \( g(\theta) \), it’s fairly straightforward to derive a CLT:

It can be shown (see lecture notes) that

\[
\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \implies N(0, \Omega(h)), \quad \text{where} \quad \Omega(h) = \nabla_g[(\pi/g)(h - \mathbb{E}_\pi[h])].
\]

Using a crude approximation (see, e.g., Liu (2008)), we can factorize \( \Omega(h) \) as follows:

\[
\Omega(h) \approx \nabla_\pi[h] \left(\nabla_g[\pi/g] + 1\right).
\]

The approximation highlights that the larger the variance of the importance weights, the less accurate the Monte Carlo approximation relative to the accuracy that could be achieved with an \( iid \) sample from the posterior.

Users often monitor

\[
ESS = N \frac{\nabla_\pi[h]}{\Omega(h)} \approx \frac{N}{1 + \nabla_g[\pi/g]}.
\]