Monte Carlo Simulation

Ed Herbst

March 16, 2023

The main event

- Inference: Need to characterize posterior $p(\theta|Y)$.

- Unfortunately, for many interesting models it is not possible to evaluate the moments and quantiles of the posterior $p(\theta|Y)$ analytically.

- Rules of game: we can only numerically evaluate prior $p(\theta)$ and likelihood $p(Y|\theta)$.

- To evaluate posterior moments of function $h(\theta)$, we need numerical techniques.

Estimating Posterior Moments

- We will often abbreviate posterior distributions $p(\theta|Y)$ by $\pi(\theta)$ and posterior expectations of $h(\theta)$ by

$$\mathbb{E}_\pi[h] = \mathbb{E}_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \int h(\theta)p(\theta|Y)d\theta.$$

- We will focus on algorithms that generate draws $\{\theta^i\}_{i=1}^N$ from posterior distributions of parameters in time series models.
- These draws can then be transformed into objects of interest, $h(\theta^i)$, and under suitable conditions a Monte Carlo average of the form

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^N h(\theta^i) \approx \mathbb{E}_\pi[h].$$

- Strong law of large numbers (SLLN), central limit theorem (CLT)...

Direct Sampling

- In the simple linear regression model with Gaussian posterior it is possible to sample directly.

- For $i = 1$ to N, draw θ^i from $\mathcal{N}\left(\tilde{\theta}_T, \tilde{V}_T\right)$.

- Provided that $\mathbb{V}_\pi[h(\theta)] < \infty$ we can deduce from Kolmogorov's SLLN and the Lindeberg-Levy CLT that

$$\begin{aligned} \bar{h}_N &\xrightarrow{a.s.} \mathbb{E}_\pi[h] \\ \sqrt{N}\left(\bar{h}_N - \mathbb{E}_\pi[h]\right) &\Longrightarrow N\big(0, \mathbb{V}_\pi[h(\theta)]\big). \end{aligned} \tag{1}$$

Decision Making

- The posterior expected loss associated with a decision $\delta(\cdot)$ is given by
$$\rho\big(\delta(\cdot)|Y\big) = \int_{\Theta} L\big(\theta, \delta(Y)\big) p(\theta|Y) d\theta.$$
- A Bayes decision is a decision that minimizes the posterior expected loss:
$$\delta^*(Y) = \text{argmin}_d \, \rho\big(\delta(\cdot)|Y\big).$$
- Since in most applications it is not feasible to derive the posterior expected risk analytically, we replace $\rho\big(\delta(\cdot)|Y\big)$ by a Monte Carlo approximation of the form
$$\bar{\rho}_N\big(\delta(\cdot)|Y\big) = \frac{1}{N}\sum_{i=1}^{N} L\big(\theta^i, \delta(\cdot)\big).$$
- A numerical approximation to the Bayes decision $\delta^*(\cdot)$ is then given by
$$\delta^*_N(Y) = \text{argmin}_d \, \bar{\rho}_N\big(\delta(\cdot)|Y\big).$$

Importance Sampling

$$\pi(\theta) = \frac{f(\theta)}{Z} = \frac{p(Y|\theta)p(\theta)}{p(Y)} \tag{2}$$

$f(\cdot)$ is the function we can evaluate numerically.

References: Hammersley and Handscomb (1964), Kloek and van Dijk (1978), and Geweke (1989).

Let g be an arbitrary, easy-to-sample pdf over θ (think normal distribution).

Importance sampling (IS) is based on the following identity:

$$\mathbb{E}_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \frac{1}{Z}\int_\Theta h(\theta)\frac{f(\theta)}{g(\theta)}g(\theta)d\theta. \tag{3}$$

Since $\mathbb{E}_\pi[1] = 1$,

$$Z = \int_\Theta \frac{f(\theta)}{g(\theta)}g(\theta)d\theta.$$

Importance Sampling

(Unnormalized) Importance weight:

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

Normalized Importance Weight:

$$v(\theta) = \frac{w(\theta)}{\int w(\theta)g(\theta)d\theta} = \frac{w(\theta)}{\int Z\pi(\theta)d\theta} = \frac{w(\theta)}{Z}. \tag{4}$$

Can show:

$$\mathbb{E}_\pi[h(\theta)] = \int v(\theta)h(\theta)g(\theta)d\theta. \tag{5}$$

The Details

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

$$w^i = w(\theta^i) = \frac{f(\theta^i)}{g(\theta^i)}. \tag{6}$$

- Compute the normalized importance weights

$$W^i = \frac{w^i}{\frac{1}{N}\sum_{i=1}^{N} w^i}. \tag{7}$$

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

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} W^i h(\theta^i). \tag{8}$$

Note W^i is (slightly) different from v in previous slide.

The Details

- Refer to the collection of pairs $\{(\theta^i, W^i)\}_{i=1}^N$ as a particle approximation of $\pi(\theta)$.

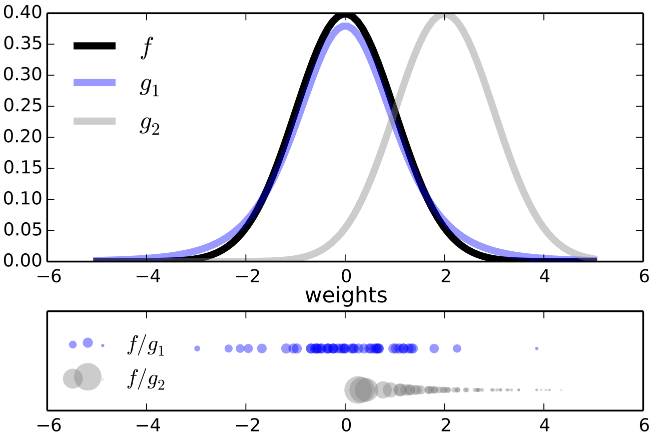
- The accuracy of the approximation is driven by the "closeness" of $g(\cdot)$ to $f(\cdot)$ and is reflected in the distribution of the weights.

- If the distribution of weights is very uneven, the Monte Carlo approximation $\bar{h}$ is inaccurate.

- Uniform weights arise if $g(\cdot) \propto f(\cdot)$, which means that we are sampling directly from $\pi(\theta)$.

Effectiveness of IS depends on similarity of f and g

$$f = \mathcal{N}(0,1), \quad g_1 = t(0,1,5), \quad g_2 = \mathcal{N}(2,1)$$



Only a few draws from $N(2,1)$ have meaningful weight.

$\Longrightarrow$ estimate is based on small sample.

$\Longrightarrow$ estimate will be noisy.

Convergence

- SLLN: If $\mathbb{E}_g[|hf/g|] < \infty$ and $\mathbb{E}_g[|f/g|] < \infty$, see Geweke (1989), the Monte Carlo estimate $\bar{h}_N$ defined in (8) converges almost surely (a.s.) to $E_\pi[h(\theta)]$ as $N \longrightarrow \infty$.

- CLT: A bit more complicated.

Central Limit Theorem

Define the population analogue of the normalized importance weights as $v(\theta) = w(\theta)/Z$ and write

$$\bar{h}_N = \frac{\frac{1}{N}\sum_{i=1}^{N}(w^i/Z)h(\theta^i)}{\frac{1}{N}\sum_{i=1}^{N}(w^i/Z)} = \frac{\frac{1}{N}\sum_{i=1}^{N} v(\theta^i)h(\theta^i)}{\frac{1}{N}\sum_{i=1}^{N} v(\theta^i)}. \tag{9}$$

Now consider a first-order Taylor series expansion in terms of deviations of the numerator from $\mathbb{E}_\pi[h]$ and deviations of the denominator around 1:

$$\begin{aligned}
&\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \\
&= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} v(\theta^i)h(\theta^i) - \mathbb{E}_\pi[h]\right) \\
&\quad -\mathbb{E}_\pi[h]\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} v(\theta^i) - 1\right) + o_p(1) \\
&= (I) - \mathbb{E}_\pi[h]\cdot(II) + o_p(1).
\end{aligned} \tag{10}$$

Central Limit Theorem

Under some regularity conditions, we can apply a multivariate extension of the Lindeberg-Levy CLT to the terms (I) and (II).

The variances and covariance of (I) and (II) are given by

$$\begin{aligned} \mathbb{V}_g[hv] &= \mathbb{E}_\pi[(\pi/g)h^2] - \mathbb{E}_\pi^2[h], \\ \mathbb{V}_g[v] &= \mathbb{E}_\pi[(\pi/g)] - 1, \\ COV_g(hv, v) &= \big(\mathbb{E}_\pi[(\pi/g)h] - \mathbb{E}_\pi[h]\big). \end{aligned}$$

In turn we can deduce that

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \Omega(h)\big), \tag{11}$$

where

$$\Omega(h) = \mathbb{V}_g[(\pi/g)(h - \mathbb{E}_\pi[h])].$$

Accuracy

- Assess the accuracy by computing a Monte Carlo approximation $\bar{h}_N$ multiple times and examine its variability across repeated runs of the posterior sampler.

- If $\bar{h}_N$ satisfies a CLT and the number of draws N is sufficiently large, then the variance across repeated runs of the algorithm (provided this variance is finite for the given N) will approximately coincide with the asymptotic variance implied by the CLT.

- Define inefficiency factor relative to IID sampling,

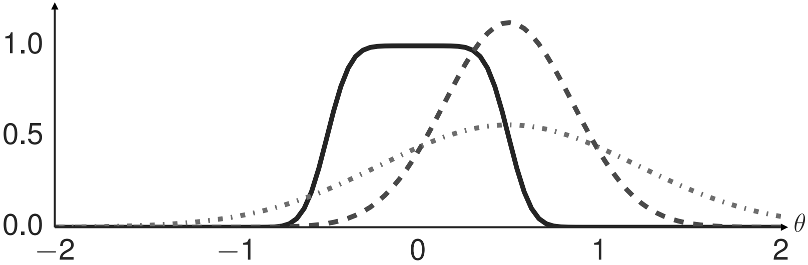
$$\text{InEff}_\infty = \frac{\Omega(h)}{\mathbb{V}_\pi[h]}.$$

If $\text{Ineff}_\infty > 1$ we are worse than iid sampling.

Numerical Illustration

- Let's take a harder $\pi(\theta)$, the set-identified posterior from Moon-Schorfheide (2013).

- Consider diffuse and concentrated importance sample densities g.



Experiment

- Using various N, generate IS approximations for $h(\theta) = \theta$ and $h(\theta) = \theta^2$.

- Calculate estimate of InEff_∞ using $N_{run} = 1000$ Monte Carlo simulations, as well as the exact value [by sampling from $\pi(\theta)$.] Estimates come from:

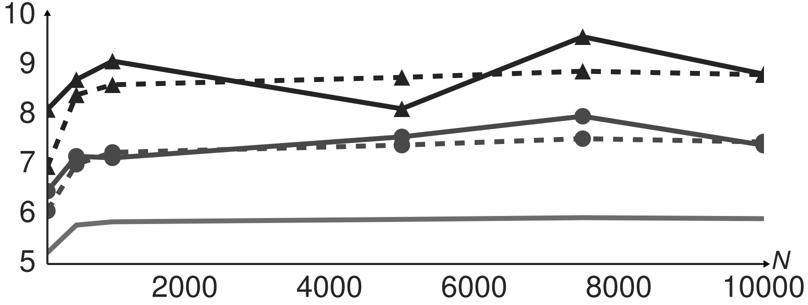
$$\text{InEff}_N = \frac{\mathbb{V}[\bar{h}_N]}{\mathbb{V}_\pi[h]/N}. \tag{12}$$

- Also calculate poor man's version of Inefficiency Factor, because everyone uses it.

$$\text{InEff}_\infty \approx 1 + \mathbb{V}_g[\pi/g]. \tag{13}$$

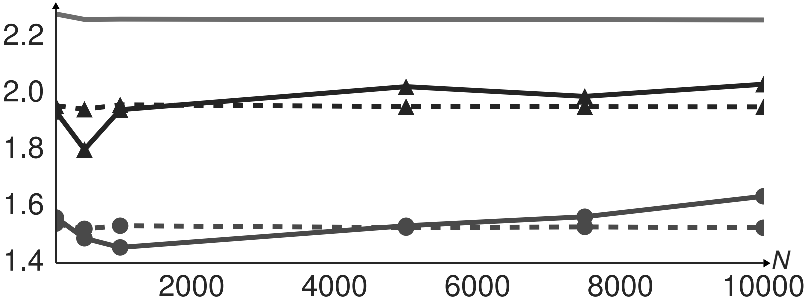
Concentrated IS Density

- solid line = estimates of $\text{InEff}_\infty[h]$, dashed = truth
- triangles = $h(\theta) = \theta$, circles = $h(\theta) = \theta^2$
- grey line = poor man's inefficiency



Diffuse IS Density

- Solid line = estimates of $\mathrm{InEff}_\infty[h]$, dashed = truth
- triangles = $h(\theta) = \theta$, circles = $h(\theta) = \theta^2$
- grey line = poor man's inefficiency



Take aways

- It is important that the importance density g is well-tailored toward the target distribution π!

- Everything is h specific!

- with approximately elliptical posterior, a good importance density can be obtained by centering a fat-tailed t distribution at the mode of π and using a scaled version of the inverse Hessian of $\ln \pi$ at the mode to align the contours of the importance density with the contours of the posterior π.

- Very bad for highly irregular and non-elliptical posteriors...

References

GEWEKE, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317.

HAMMERSLEY, J. M. AND D. C. HANDSCOMB (1964): "Monte Carlo Methods," .

KLOEK, T. AND H. K. VAN DIJK (1978): "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46, 1–19.