The pseudomarginal approach

[Based on Andrieu,C. and Roberts, G.O. (2009)]
Tea Talk

Maria D. Lomeli-Garcia

Gatsby Computational Neuroscience Unit

August 16, 2013

Table of Contents

Motivation: Intractable densities

$$\pi(\theta) = \int \pi(\theta, \mathbf{z}) \Lambda(\mathrm{d}\mathbf{z})$$

where θ might be a parameter of interest and $\mathbf{z}$ are latent variables. As usual, the target distribution $\pi(\theta)$ cannot be evaluated analytically but the joint $\pi(\theta, \mathbf{z})$ can be.

Some examples:

1. Hidden Markov Models
2. Mixture Models
3. Diffusion processes observed at discrete times [1]
4. Model selection

---

[1] Stramer and Bognar (2011)

What can we do?

1.

$$\begin{aligned}\pi(\theta) &= \int \pi(\theta, \mathbf{z})\mathrm{d}\mathbf{z} \\ &= \int \frac{\pi(\theta, \mathbf{z})}{q_\theta(\mathbf{z})} q_\theta(\mathbf{z})\mathrm{d}\mathbf{z} \\ &= E_{q_\theta(z)}\left[\frac{\pi(\theta, \mathbf{z})}{q_\theta(\mathbf{z})}\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \frac{\pi(\theta, \mathbf{z}_i)}{q_\theta(\mathbf{z}_i)}\end{aligned}$$

i.e. we can rewrite it as an expectation w.r.t. some density $q_\theta(\mathbf{z})$, obtain a Montecarlo estimate of it and use it whenever we might need to evaluate the density of interest.

What can we do? (More fancy stuff)

2. Numerical integration

   ☹The approximation can be quite poor when the dimensionality of $\mathbf{z}$ is high.

3. Data augmentation scheme and use MCMC
   i.e. Compute full conditionals for sampling $\theta \mid \mathbf{z}$ and $\mathbf{z} \mid \theta$
   (either directly or using Metropolis within Gibbs, slice sampling within Gibbs, etc.)

   ☹This can result in strongly correlated samples (θ_i, z_i)

4. **Pseudomarginal scheme** ☺
   Combines the computational efficiency of sampling directly from $\pi(\theta)$ and implementation ease of introducing auxiliary variables.

Which one of this approaches leads to a valid sampler? i.e. has $\pi(\theta)$ as invariant distribution

Some pseudocode

TABLE 1
Comparison of the marginal, MCWM and GIMH algorithms

Step	Marginal	MCWM	GIMH
0. *Given:*	θ and $\pi(\theta)$	θ and $\pi(\theta)$	θ, Z and $\tilde{\pi}^N(\theta)$
1. *Sample:*	$\theta^* \sim q(\theta, \cdot)$	$\theta^* \sim q(\theta, \cdot)$	$\theta^* \sim q(\theta, \cdot)$
		$\begin{cases} Z \sim q_{\theta}^{N}(\cdot), \\ Z^* \sim q_{\theta^*}^{N}(\cdot) \end{cases}$	$Z^* \sim q_{\theta^*}^{N}(\cdot)$
2. *Compute:*	$\pi(\theta^*)$	$\begin{cases} \tilde{\pi}^N(\theta), \\ \tilde{\pi}^N(\theta^*) \end{cases}$	$\tilde{\pi}^N(\theta^*)$
3. *Compute:* $r =$	$\frac{\pi(\theta^*)q(\theta^*,\theta)}{\pi(\theta)q(\theta,\theta^*)}$	$\frac{\tilde{\pi}^N(\theta^*)q(\theta^*,\theta)}{\tilde{\pi}^N(\theta)q(\theta,\theta^*)}$	$\frac{\tilde{\pi}^N(\theta^*)q(\theta^*,\theta)}{\tilde{\pi}^N(\theta)q(\theta,\theta^*)}$
4. *With prob.* $1 \wedge r$*:*	$\vartheta = \theta^*$	$\vartheta = \theta^*$	$\begin{cases} \vartheta = \theta^*, \\ \mathfrak{z} = Z^* \end{cases}$
otherwise:	$\vartheta = \theta$	$\vartheta = \theta$	$\begin{cases} \vartheta = \theta, \\ \mathfrak{z} = Z \end{cases}$

Figure: What is the difference between column 2 and 3?

Monte Carlo within Metropolis MCWM

- Z and Z^* are refreshed at each iteration independently of previous samples.
- $\{\theta_i\}$ is still a Markov chain with some transition kernel P^{MCWM}.
- $\pi(\theta)$ is not the invariant distribution ☹

Grouped Independent Metropolis Hastings

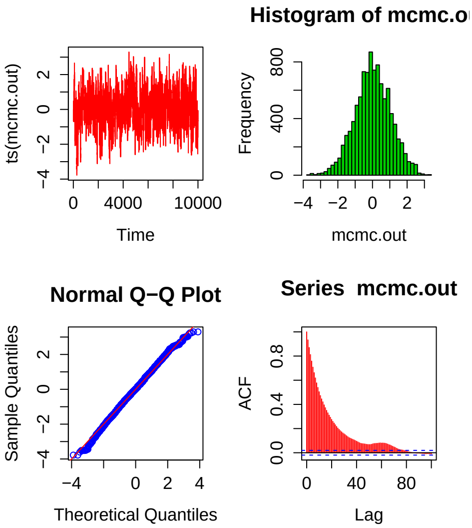
- No fresh Z is sampled at each iteration but Z is recycled from the previous iteration.
- $\{\theta_i\}$ is no longer a Markov chain, but (θ_i, Z_i) is. ☺

So GIMH can be seen as an approximation of a MH algorithm with target $\pi(\theta)$ or as an MH algorithm with target $\tilde{\pi}^N(\theta, Z)$ since the acceptance ratio is

$$\frac{\tilde{\pi}^N(\theta^*)q(\theta^*,\theta)}{\tilde{\pi}^N(\theta)q(\theta,\theta^*)} = \frac{[1/N\sum_{k=1}^N \pi(\theta^*, z^*(k))\prod_{l=1;l\neq k}^N q_{\theta^*}(z^*(l))]q(\theta^*,\theta)q_\theta^N(Z)}{[1/N\sum_{k=1}^N \pi(\theta, z(k))\prod_{l=1;l\neq k}^N q_{\theta}(z(l))]q(\theta,\theta^*)q_{\theta^*}^N(Z^*)},$$

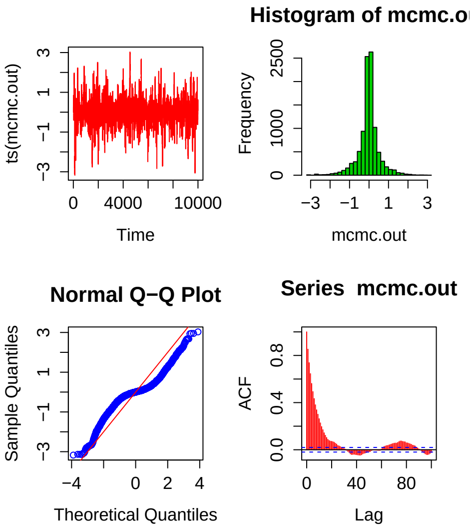
R Code: GIMH

```
pmmcmc<-function(N=100,alpha=0.5)
{
  vec=vector("numeric", n)
  x=0
  oldlik=noisydnorm(x)
  vec[1]=x
  for (i in 2:n) {
    innov=runif(1,-alpha,alpha)
    can=x+innov
    lik=noisydnorm(can)
    aprob=lik/oldlik
    u=runif(1)
    if (u < aprob) {
      x=can
      oldlik=lik
    }
    vec[i]=x
  }
  vec
}
```


Histogram of mcmc.o
ts(mcmc.out)
Time
Frequency
mcmc.out
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
Series mcmc.out
ACF
Lag

R Code: MCWM

```
approxmcmc<-function(n=100,alpha=0.5)
{
  vec = vector("numeric", n)
  x = 0
  vec[1] = x
  for (i in 2:n) {
    innov = runif(1,-alpha,alpha)
    can = x+innov
    lik = noisydnorm(can)
    oldlik = noisydnorm(x)
    aprob = lik/oldlik
    u = runif(1)
    if (u < aprob) {
      x = can
    }
    vec[i] = x
  }
  vec
}
```


Histogram of mcmc.o
ts(mcmc.out)
Time
Frequency
mcmc.out
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
Series mcmc.out
ACF
Lag

R code: Noisy likelihood estimates examples

```
noisydnorm<-function(z)
{
  dnorm(z)*rexp(1,2)
}
noisydnorm<-function(z)
{
  dnorm(z)*rexp(1,0.1+10*z*z)
}
noisydnorm<-function(z)
{
  dnorm(z)*rgamma(1,0.1+10*z*z,0.1+10*z*z)
}
noisydnorm<-function(z)
{
  dnorm(z)*rnorm(1,1)
}
noisydnorm<-function(z)
{
  dnorm(z)*rnorm(1,0,0.1+10*z*z)
}
```

- Nice alternative for dealing with intractable densities.
- Could we modify it for non-linear functions of an intractable density?
- How about derivatives?