Simple EP

Swupnil Sahai

July 27, 2015

1. Uknown Mean

Formulation

We start by fitting one of the simplest possible models that isn't hierarchical in nature:

$$x_i \sim N(\phi, 1)$$

Generating 1000 data points with $\phi = 200$, we give EP and HMC the following prior:

$$\phi \sim N(0, 20)$$

Results

EP and HMC converge to similar posteriors of ϕ, but the variance for EP decreases after the first iteration.

Table 1: Posterior of ϕ

Parameter	True	Prior	HMC	EP Iter 1	EP Iter 2	EP Iter 3	EP Iter 4
$E(\phi)$	200	0	199.9464	199.9433	199.9411	199.9426	199.9436
$SD(\phi)$	–	20	0.03175	0.03187	0.02531	0.02171	0.01965

2. Uknown Means (+ Hierarchical Mean)

Formulation

We then try a slightly more interesting model with local unknown means drawn from a global uknown mean and known variance:

$$\theta_j \sim N(\phi, 10)$$

$$x_{ij} \sim N(\theta_j, 1)$$

Generating 50 θ_j's from $\phi = 200$, and then sampling 1000 data points from each group j, we give EP and HMC the following prior:

$$\phi \sim N(0, 20)$$

Results

EP and HMC converge to slightly different posteriors of ϕ, with the location biased more and more as the number of sites is increased (i.e. as we spread the computation across more cores).

Table 2: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	200	0	200.93	191.09	188.15	177.06
$SD(\phi_1)$	–	20	1.45	0.60	0.43	0.26
Time	–	–	201 s	116 s	95 s	92 s

HMC Note: HMC "converged"" to strange values of ϕ (e.g. 20) when using only 100 iterations per chain. This was surprising given that the θ_j's were all being recovered perfectly, but ϕ was not. This was additionally surprising given that there were a total of 50,000 data points.

EP Notes: Initially estimation was terrible and I realized that I was trying make the prior sigma within each site equal to 20, which caused the prior sigma overall to be equal to 20/K (because of the way the precision is summed in EP). I then forced the overal prior sigma to be 20, causing each site to have a sigma prior of 20*K. This produced much better results.

Re-run with Stronger Prior

EP performed substantially better when giving a prior of $\phi \sim N(200, 20)$.

Table 3: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	200	200	201.91	201.79	201.75	201.71
$SD(\phi_1)$	–	20	1.42	0.604	0.427	0.26
Time	–	–	183 s	72 s	85 s	127 s

3. Uknown Means (+ Hierarchical Mean and Variance)

Formulation

We then try a model with local unknown means drawn from a global unknown mean and unknown variance:

$$\theta_j \sim N(\phi_1, e^{\phi_2})$$

$$x_{ij} \sim N(\theta_j, 1)$$

Generating 50 θ_j's from $\phi = 200$ and $e^{\phi_2} = 5$, and then sampling 1000 data points from each group j, we give EP and HMC the following priors:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 400 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

Results

EP performed surprisingly poorly when given a relatively non-informative prior.

Table 4: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	200	200	199.66	0.10	0.06	0.02
$SD(\phi_1)$	–	20	0.68	0.58	0.31	0.12
$E(\phi_2)$	1.61	0	1.57	5.20	4.97	4.38
$SD(\phi_2)$	–	1	0.10	0.09	0.07	0.04
$Cov(\phi_1, \phi_2)$	–	0	-0.0064	-0.00067	-0.00035	-0.00007
Time	–	–	1137 s	125 s	88 s	75 s

Re-run with Stronger Prior

EP performed substantially better when given a prior of $\phi \sim N\left(\begin{bmatrix} 200 \\ 0 \end{bmatrix}, \begin{bmatrix} 400 & 0 \\ 0 & 1 \end{bmatrix} \right)$.

Table 5: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	200	200	200.66	200.02	199.99	200.00
$SD(\phi_1)$	–	20	0.67	0.42	0.25	0.11
$E(\phi_2)$	1.61	0	1.57	1.52	1.42	1.09
$SD(\phi_2)$	–	1	0.10	0.10	0.09	0.06
$Cov(\phi_1, \phi_2)$	–	0	0.00034	-0.0036	-0.00015	-0.00026
Time	–	–	937 s	700 s	292 s	237.9 s