

Simple EP

Swupnil Sahai

September 22, 2015

1. Unknown 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 , with the values stabilizing immediately after the first iteration.

Table 1: Posterior of ϕ (K = 5)

Parameter	True	Prior	HMC	EP Iter 1	EP Iter 2	EP Iter 3
$E(\phi)$	200	0	200.0165	200.0181	200.015	200.0199
$Var(\phi)$	—	20	0.01214	0.01433	0.01243	0.01314

2. Unknown Means (+ Hierarchical Mean)

Formulation

We then try a slightly more interesting model with local unknown means drawn from a global unknown 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 same posterior of ϕ , regardless of the number of sites.

Table 2: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	200	0	200.07	200.23	200.24	200.11
$Var(\phi_1)$	–	20	2.03	2.08	1.99	1.99
Time	–	–	201 s	81 s	73 s	69 s

Additionally, as in the non-hierarchical case, we find that the posterior stabilizes after the first iteration, regardless of the number of sites.

Table 3: Posterior of ϕ (K = 10)

Parameter	True	Prior	HMC	EP Iter 1	EP Iter 2	EP Iter 3
$E(\phi)$	200	0	200.07	200.05	200.21	200.02
$Var(\phi)$	–	20	2.03	1.89	1.98	1.58

Table 4: Posterior of ϕ (K = 25)

Parameter	True	Prior	HMC	EP Iter 1	EP Iter 2	EP Iter 3
$E(\phi)$	200	0	200.07	200.27	200.15	199.98
$Var(\phi)$	–	20	2.03	2.01	2.55	2.08

3. Unknown 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 $\phi_1 = 200$ and $e^{\phi_2} = 5$, and then sampling 1000 data points from each group j , we give EP and HMC the following priors:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1000 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

Results

EP and HMC converge to same posterior of ϕ for a small number of sites. When using 25 sites, however, the center for the hierarchical mean is pulled too much towards the prior. This is surprising given that the Stan calls during this EP run are doing a great job of discovering the local means, just not the hierarchical mean. Overall, this suggests that stronger priors for the hierarchical parameters are needed when the number of sites is large.

Table 5: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	200	200	200.52	199.89	199.33	18.32
$SD(\phi_1)$	—	100	0.64	0.56	0.52	NA
$E(\phi_2)$	1.61	0	1.51	1.50	1.45	5.13
$SD(\phi_2)$	—	1	0.10	0.10	0.11	NA
$Cov(\phi_1, \phi_2)$	—	0	-0.0041	-0.0088	-0.0131	0.92
Time	—	—	1137 s	452 s	467 s	296 s

Fortunately, as in the previous model, we find that the posterior stabilizes after the first iteration.

Table 6: Posterior of ϕ (K = 10)

Parameter	True	Prior	HMC	EP Iter 1	EP Iter 2	EP Iter 3
$E(\phi)$	200	0	200.52	199.33	200.87	200.62
$SD(\phi)$	—	100	0.64	0.52	0.77	0.64
$E(\phi_2)$	1.61	0	1.51	1.45	1.48	1.50
$SD(\phi_2)$	—	1	0.10	0.11	0.10	0.09
$Cov(\phi_1, \phi_2)$	—	0	-0.0041	-0.0131	-0.0335	0.0145

4. Slope and Intercept (+ Hierarchical Mean and Variance for Slope and Intercept)

Formulation

Finally we try a hierarchical linear model with local unknown slopes and intercepts drawn from global unknown slope/intercept mean and variance, as well as a global model variance:

$$\begin{aligned}\alpha_j &\sim N(\phi_1, e^{\phi_3}) \\ \beta_j &\sim N(\phi_2, e^{\phi_4}) \\ x_{ij} &\sim U(0, 50) \\ y_{ij} &\sim N(\alpha_j + \beta_j x_{ij}, e^{\phi_5})\end{aligned}$$

Generating 50 α_j and β_j 's from $\phi_1 = -100, \phi_2 = 10, e^{\phi_3} = 5, e^{\phi_4} = 4$, and $e^{\phi_5} = 20$, and then sampling 100 data points from each group j , we give EP and HMC the following priors:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1000 & 0 & 0 & 0 & 0 \\ 0 & 1000 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}\right)$$

Results

EP and HMC converge to same posterior of ϕ for a small number of sites. When using 25 sites, however, the posterior variance is quite larger.

Table 7: Posterior of ϕ

Parameter	True	Prior	HMC	EP K=5	EP K=10	EP K=25
$E(\phi_1)$	-100	0	-98.7	-99.2	-99.3	-99.2
$V(\phi_1)$	—	1000	0.95	0.64	0.62	1.60
$E(\phi_2)$	10	0	9.64	9.49	9.54	10.18
$V(\phi_2)$	—	1000	0.37	0.37	0.24	0.08
$E(\phi_3)$	1.61	0	1.72	1.83	1.33	1.40
$V(\phi_3)$	—	2	0.026	0.045	0.069	0.056
$E(\phi_4)$	1.39	0	1.46	1.48	1.41	1.44
$V(\phi_4)$	—	2	0.011	0.010	0.012	0.018
$E(\phi_5)$	2.99	0	3.00	3.00	3.00	3.01
$V(\phi_5)$	—	2	0.0001	0.0001	0.0001	0.0001
Time	—	—	81 s	61 s	48 s	60 s