Target tracking with the Gibbs sampler

As mentioned last time, the smoothing problem, $\left[X_{1:k} \middle| Y_{1:k}\right]$, isn't solved very well with SIS. However it can be done very easily with Gibbs sampling.

Step j, j = 1, … , k – 1

Draw $X_j \sim \left[X_j \middle| X_{j-1}, X_{j+1}, Y_j\right]$

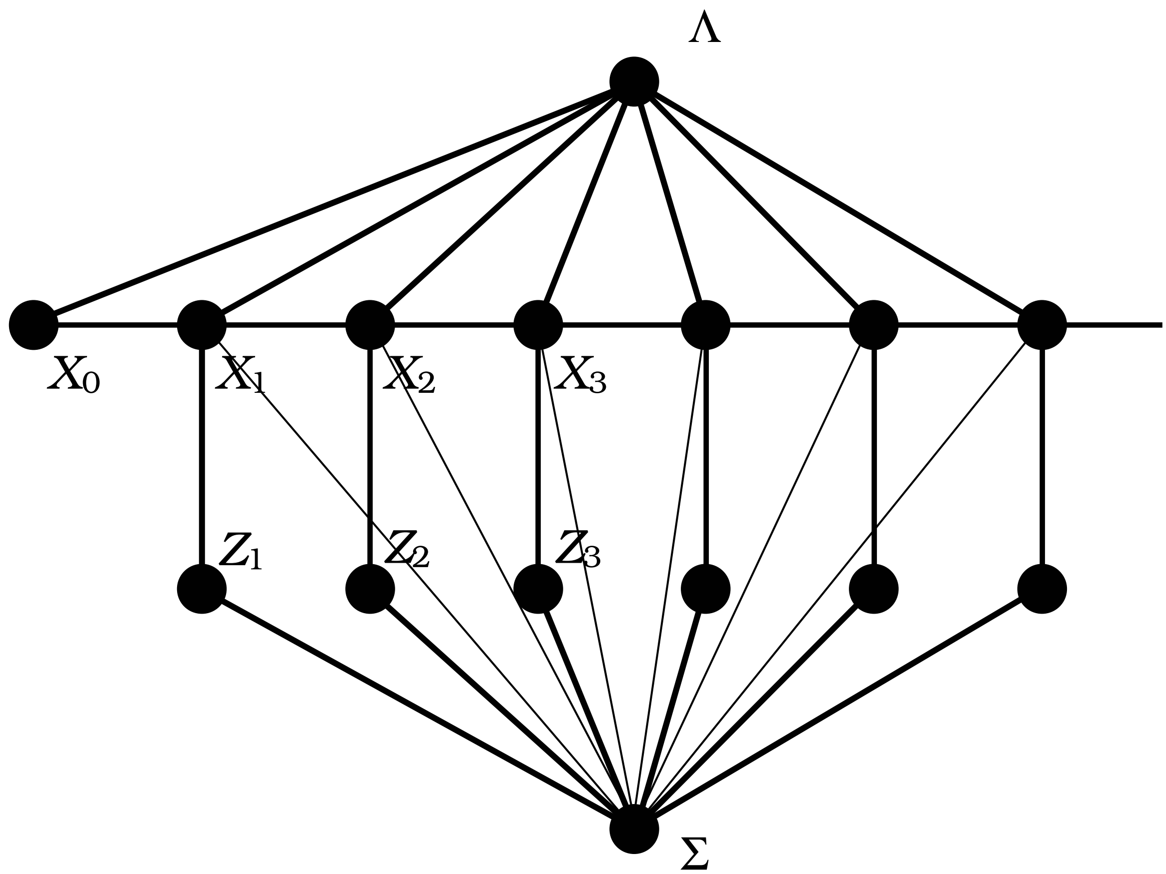
Step k

Draw $X_k \sim \left[X_k \middle| X_{k-1}, Y_k\right]$

As all the components involved in these conditional distributions are normal, each of these conditional distributions are normal, thus are easily sampled.

In the SIS analysis, it was assumed that all of the parameters of the movement and measurement error distributions (all variances) and the starting point were assumed known.

This can easily be relaxed by putting priors on X_0, Λ, and Σ and sampling them as well as part of the Markov chain.



The sampler needs to be modified as

Step 0

Draw $X_0 \sim \left[X_0 \middle| X_1, \Lambda \right]$

Step j, j = 1, … , k – 1

Draw $X_j \sim \left[X_j \middle| X_{j-1}, X_{j+1}, Y_j, \Lambda \right]$

Step k

Draw $X_k \sim \left[X_k \middle| X_{k-1}, Y_k, \Lambda \right]$

Step $k+1$

Draw $\Lambda \sim \left[\Lambda \middle| X_{0:k} \right]$

Step $k+2$

Draw $\Sigma \sim \left[\Sigma \middle| X_{0:k}, Y_{1:K} \right]$

This can be performed by Gibbs sampling if the priors on X_0 is Normal and the priors on Λ and Σ are IGamma.

Conditions for Gibbs Sampling to work

While you can always run the chain, it may not give the answer you want. That is, the realizations may not have the desired stationary distribution.

One-step transitions: $p\left(x \middle| y\right)$

n-step transitions: $p_n\left(x \middle| y\right)$.

Stationary distribution:

$$\pi(x) = \lim_{n\to\infty} p_n\left(x \middle| y\right)$$

If it exists, it satisfies

$$\pi(x) = \int p(x|y)\pi(y)dy$$

A stronger condition which shows that $\pi(x)$ is the density of the stationary distribution is

$$\pi(x)\, p(y|x) = \pi(y)\, p(x|y)$$

holds for all x & y (detailed balance).

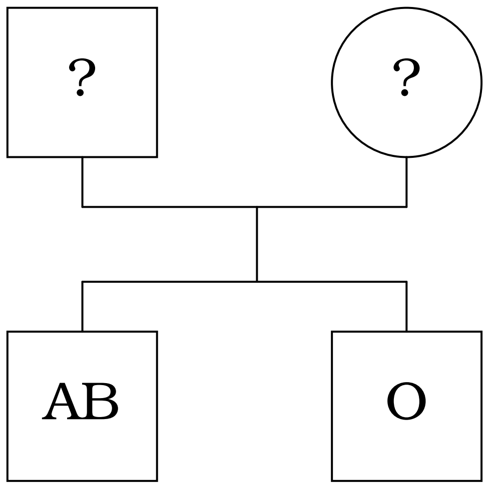
Note that detailed balance $\Rightarrow$ stationarity butstationarity doesn't imply detailed balance.

If the following two conditions hold, the chain will have the desired stationary distribution.

Irreducibility: The chain generated must be irreducible. That is it is possible to get from each state to every other state in a finite number of steps.

Not all problems lead to irreducible chains.

Example: ABO blood types



The children's data implies that the child with blood type AB must have genotype AB and that the child with blood type O must have genotype OO.

The only possible way for the two children to inherit those genotypes if for one parent to have genotype AO and for the other parent to have genotype BO. However it is not possible to say which parent has which genotype with certainty.

By a simple symmetry argument

$$P[Dad = AO \; \& \; Mom = BO]$$
$$= P[Dad = BO \; \& \; Mom = AO]$$
$$= 0.5$$

Lets try running a Gibbs sampler, by first generating mom's genotype given dad's and then dad's given mom's.

Let start the chain with $Dad = AO$.

Step 1: Generate Mom

$$P\left[Mom = AO \middle| Dad = AO\right] = 0$$

$$P\left[Mom = BO \middle| Dad = AO\right] = 1$$

so $Mom = BO$.

Step 2: Generate Dad

$$P\left[Dad = AO \middle| Mom = BO\right] = 1$$

$$P\left[Dad = BO \middle| Mom = BO\right] = 0$$

so $Dad = AO$.

This implies that every realization of the chain has $Mom = BO$ & $Dad = AO$.

If the chain is started with $Dad = BO$, every realization of that chain will have $Mom = AO$ & $Dad = BO$.

The reducible chain in this case does not have the correct stationary distribution. (Well reducible chains don't really have stationary distributions anyway). But running the described Gibbs sampler will not correctly the describe the distribution of the mother and father's genotypes.

Aperiodicity:

Don't want a periodic chain (e.g. certain states can only occur on when t is even)

This violates the idea that each state has a long run frequency marginally.

Starting Points

For every chain you need to specify a starting point. There are a number of approaches for choosing this.

1) Prior means

   In pump example, set $\beta^0 = E[\beta] = \frac{\delta}{\gamma}$.

2) Estimate from data

In pump example, $E[l_i] = \alpha\beta$, so set $\beta^0 = \dfrac{\bar{l}}{\alpha}$.

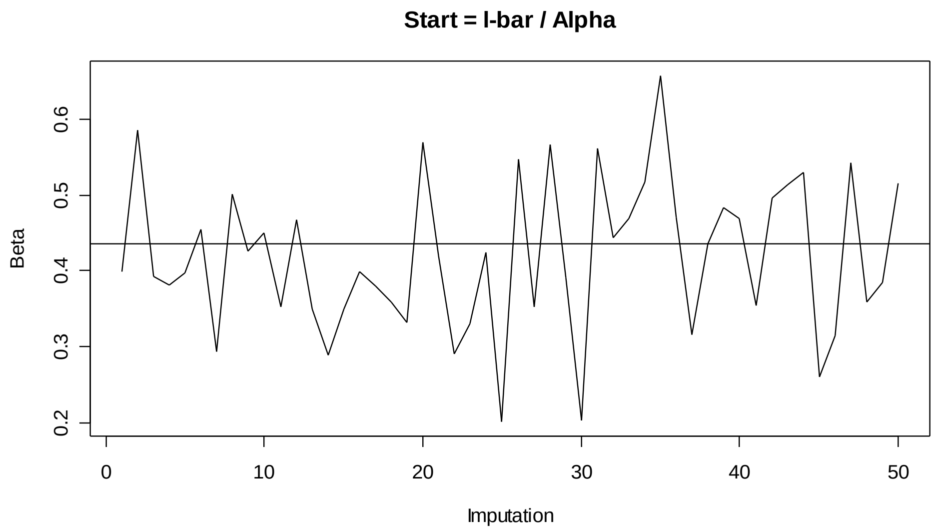
In target tracking example, set starting positions at each time to average observed positions, the differences of these to get the velocities.

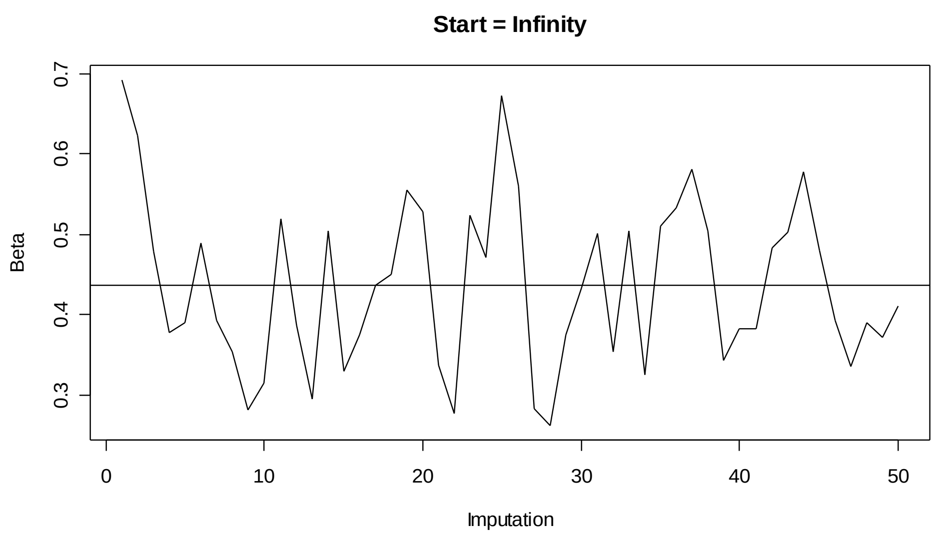
3) Sample from prior

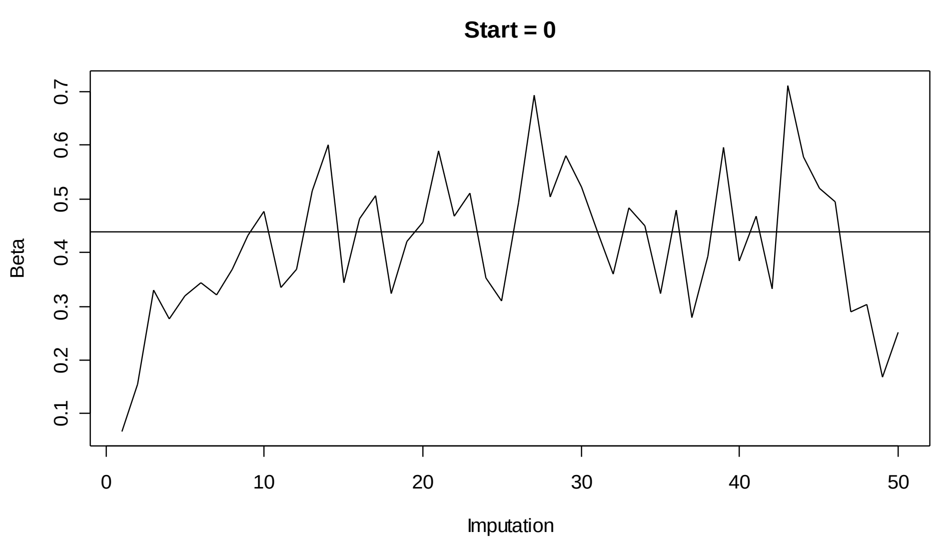
4) Ad hoc choices

In pump example, set $\beta^0 = \infty$

For many problems, this choice can be important. The stationary distribution is an asymptotic property and it may take a long time for the chain to converge.


Start = I-bar / Alpha
Beta
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
Imputation


Start = Infinity
Beta
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
50
Imputation

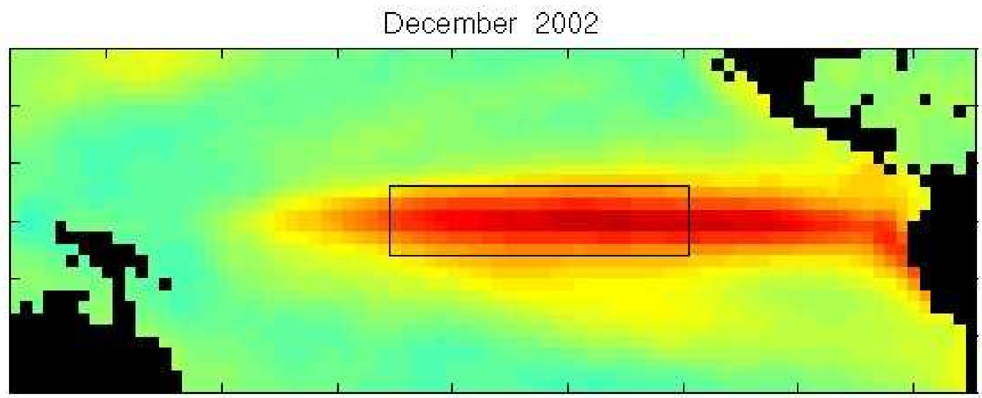

Start = 0
Beta
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
50
Imputation

Starting with $\beta^0 = 0$ (actually 10^{-100}), the initial draws are not consistent with the stationary distribution seen later in the chain.

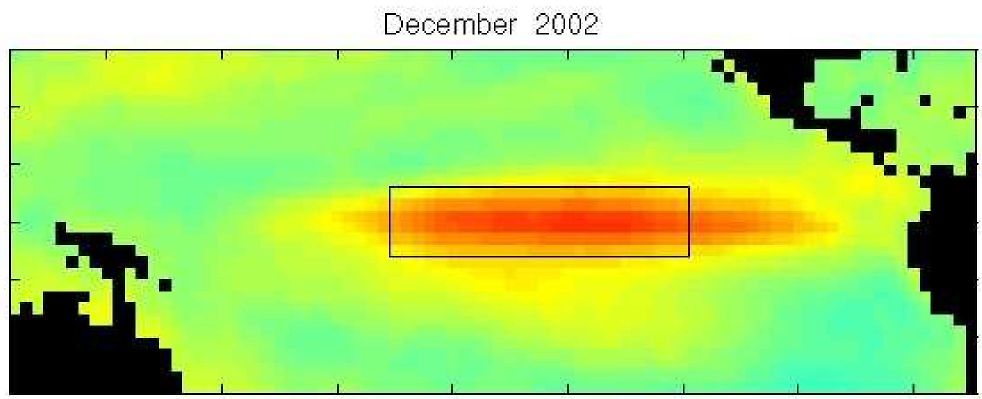
While for this example, the problem clears up quickly, for other problems it can take a while.

This is more common which larger problems, that might have millions, or maybe billions of variables being sampled in a complete single scan through the data. This can occur with large space time problems, such as the Tropical Pacific sea surface temperature predictions discussed at <http://www.stat.ohio-state.edu/~sses/collab_enso.php>.

Forecast map for December 2002 based on data from January 1970 to May 2002



Observed December 2002 map



The usual approach to have a “burn-in” period where the initial samples are thrown away since they may not be representative of samples from the stationary distribution.

The following table contains estimates of the posterior means of the 11 parameters in the pump example with 3 different starting points. The first 200 imputations were discarded and then the next 1000 imputations were sampled.

Pump	$\beta^0 = \bar{l}/\alpha$	$\beta^0 = \infty$	$\beta^0 = 0$
1	0.0688	0.0704	0.0715
2	0.1531	0.1531	0.1575
3	0.1064	0.1024	0.1050
4	0.1234	0.1236	0.1221
5	0.6008	0.6198	0.6319
6	0.6116	0.6145	0.6163
7	0.7744	0.8501	0.8118
8	0.8173	0.8224	0.8190
9	1.2584	1.2748	1.2857
10	1.8393	1.8536	1.8409

β	0.4256	0.4358	0.4334

Often the bigger the problem, the longer the burn-in period desired. However those are the problems where time considerations will limit the total number of imputations that can be done.

So you do want to think about starting values for your chain.

Gibbs sampling and Bayes – Choice of priors

For Gibbs sampling to be efficient, the draws in each step of the procedure need to be feasible.

That suggests that conjugate distributions need to be used as part of the hierarchical model, as was done in pump and target tracking examples.

However conjugacy is not strictly required, as rejection sampling with log-concave distributions might be able to be used in some problems.

This idea is sometimes used in the software package WinBUGS (Bayesian analysis Using Gibbs Sampling).

However for some problems the model you want to analyze is not conjugate and the tricks to get around non-conjugacy won't work.

For example, lets change model for the pump example to

$$
\begin{aligned}
s_i \mid \lambda_i &\sim \text{Poisson}(\lambda_i t_i) \\
\lambda_i \mid \mu, \sigma^2 &\sim \text{LogN}(\mu, \sigma^2) \\
\mu &\sim \text{Logistic}(\nu, \tau) \\
\sigma^2 &\sim \text{Weibull}(\alpha, \beta)
\end{aligned}
$$

Good luck on running a Gibbs sampler on this model (I think).

Other sampling techniques are needed, for this and other more complicated problems.

Metropolis – Hastings Algorithm

A general approach for constructing a Markov chain that has the desired stationary distribution ($\pi_j = \pi(j)$)

1) Proposal distribution:

Assume that $X^t = i$. Need to propose a new state with distribution $q_{ij} = q(j|i)$.

2) Calculate the Hastings' ratio

$$a_{ij} = \min\left\{\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right\}$$

3) Acceptance/Reject step

Generate $U \sim U(0,1)$ and set

$$X^{t+1} = \begin{cases} j & \text{if } U \le a_{ij} \\ i\left(= X^t\right) & \text{otherwise} \end{cases}$$

Notes:

1) Gibbs sampling is a special case of Metropolis - Hastings as for each step,

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = 1$$

which implies the sample holds for a complete scan through all the variables.

2) The Metropolis (Metropolis et al, 1953) algorithm was based on a symmetric proposal distribution ($q_{ij} = q_{ji}$)

$$a_{ij} = \min\left\{\frac{\pi_j}{\pi_i}, 1\right\}$$

So a higher probability state will always be accepted.

3) As with many other sampling procedures, π and q only need to be known up to normalizing constants as they will be cancelled out when calculating the Hastings' ratio.

4) Periodicity isn't a problem usually.

For many proposals, $q_{ii} > 0$ for all i. Also if $a_{ij} < 1$, $P\left[X^{t+1} = i \middle| X^t = i\right] > 0$, thus some states have period 1, which implies the chain is aperiodic.

5) $q_{ij}a_{ij}$ gives the 1-step transition probabilities of the chain (e.g. its $p(x|y)$ in the earlier notation).

6) Detailed balance is easy. Without loss of generality, assume that

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} < 1$$

(which implies $a_{ij} < 1$ and $a_{ji} = 1$)

Then

$$\begin{aligned}\pi_i q_{ij} a_{ij} &= \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \\ &= \pi_j q_{ji} \\ &= \pi_j q_{ji} a_{ji}\end{aligned}$$

7) The big problem is irreducibility. However by setting the proposal to correspond to a irreducible chain solves this.

Proposal distribution ideas:

1) Approximate the distribution. For example use a normal with similar means and variances. Or use a t with a moderate number of degrees of freedom.

2) Random walk

$$q(y|x) = q(y - x)$$

If there is a continuous state process, you could use

$$y = x + \varepsilon; \quad \varepsilon \sim q(\bullet)$$

For a discrete process, you could use

$$q(j|i) = \begin{cases} 0.4 & j = i - 1 \\ 0.2 & j = i \\ 0.4 & j = i + 1 \end{cases}$$

3) Autoregressive chain

$$y = a + B(x - a) + z; \quad z \sim q(\bullet)$$

For the random walk and autoregressive chains, q does not need to correspond to a symmetric distribution (though that is common).

4) Independence sampler

$$q(y|x) = q(y)$$

For an independence sampler you want q to be similar to π.

$$a_{ij} = \min\left\{\frac{\pi_j q_i}{\pi_i q_j}, 1\right\}$$

If they are too different, q_i/π_i could get very small, making it difficult to move from state i. (The chain mixes slowly).