Target tracking example

Filtering: $\left[X_t \middle| Y_{1:t} \right]$ (main interest)

Smoothing: $\left[X_{1:t} \middle| Y_{1:t} \right]$ (also given with SIS)

However as we have seen, the estimate of this distribution breaks down when t gets large due to the weights becoming degenerate (if we don't resample).

If we resample, most of the values sampled for X_1 will disappear when t gets large (related to the weight breakdown).

So SIS isn't useful for all problems.

Gibbs sampling

Special case of Markov Chain Monte Carlo (MCMC)

Instead of generating independent samples, it generates dependent samples via a Markov chain.

$$X^1 \to X^2 \to X^3 \to \ldots$$

Useful for a wide range of problems.

Popular for Bayesian analyses, but is a general sampling procedure. For example, it can be used to do smoothing in the target tracking example.

Similar to SIS in that the random variable X is decomposed into $X = \{X_1, X_2, \ldots, X_k\}$ and each piece is simulated separately.

However the conditioning structure is different. When sampling X_j, it is drawn conditional on all other components of X.

Gibbs sampler

A) Starting value: $X^0 = \{X_1^0, X_2^0, \ldots, X_k^0\}$

Picked by some mechanism

B) Sample $X^t = \{X_1^t, X_2^t, \ldots, X_k^t\}$ by

1) $X_1^t \sim \left[X_1 \middle| X_2^{t-1}, X_3^{t-1}, \ldots, X_k^{t-1} \right]$

2) $X_2^t \sim \left[X_2 \middle| X_1^t, X_3^{t-1}, \ldots, X_k^{t-1} \right]$

…

$$j) \quad X_j^t \sim \left[X_j \middle| X_1^t, \ldots, X_{j-1}^t, X_{j+1}^{t-1}, \ldots, X_k^{t-1} \right]$$

…

$$k) \quad X_k^t \sim \left[X_j \middle| X_1^t, \ldots, X_{k-1}^t \right]$$

Under certain regularity conditions, the realizations $X^1, X^2, X^3, \ldots$ form a Markov chain with stationary distribution $[X]$.

Thus the realizations can be treated as dependent samples from the desired distribution.

Example: (Nuclear pump failure)

Gaver & O'Muircheartaigh (Technometrics, 1987)

Gelfand & Smith (JASA, 1990)

Observed 10 nuclear reactor pumps

Counted the number of failures for each pump

Pump	Failures (s_i)	Obs Time (t_i)	Obs Rate (l_i)
1	5	94.320	0.053
2	1	15.720	0.064
3	5	62.880	0.080
4	14	125.760	0.111
5	3	5.240	0.573
6	19	31.440	0.604
7	1	1.048	0.954
8	1	1.048	0.954
9	4	2.096	1.910
10	22	10.480	2.099

(Obs Time in 1000's of hours)

(Obs Rate = Failures / Time)

Want to determine the true failure rate for each pump with the following hierarchical model

$$s_i \,|\, \lambda_i \sim \text{Poisson}\left(\lambda_i t_i\right)$$

$$\lambda_i \,|\, \beta \sim \text{Gamma}\left(\alpha, \beta\right)$$

$$\beta \sim \text{IGamma}\left(\gamma, 1/\delta\right)$$

Note: $\beta \sim \text{IGamma}\left(\gamma, 1/\delta\right)$ is equivalent to

$$\frac{1}{\beta} \sim \text{Gamma}\left(\gamma, 1/\delta\right)$$

$$f\left(s_i \,|\, \lambda_i\right) = \frac{\left(\lambda_i t_i\right)^{s_i} e^{-\lambda_i t_i}}{s_i !}$$

$$\pi\left(\lambda_i \,|\, \beta\right) = \frac{\lambda_i^{\alpha-1} e^{-\lambda_i/\beta}}{\beta^{\alpha}\Gamma\left(\alpha\right)}$$

$$\rho\left(\beta\right) = \frac{\delta^{\gamma} e^{-\delta/\beta}}{\beta^{\gamma+1}\Gamma\left(\gamma\right)}$$

Want to determine

1) $\left[\lambda_i \,|\, S\right]$ for each pump i = 1, … , 10

2) $\left[\beta \,|\, S\right]$

where($S = \left\{s_1, \ldots, s_{10}\right\}$)

Note that both sets of these distributions are hard to get analytically.

Can show that

$$p(\lambda|S) \propto \frac{1}{\left(\delta + \sum \lambda_i\right)^{10\alpha+\gamma}} \prod \frac{t_i^{\alpha+s_i} \lambda_i^{\alpha+s_i-1} e^{-\lambda_i t_i}}{\Gamma(\alpha + s_i)}$$

where $\lambda = \{\lambda_1, \ldots, \lambda_{10}\}$.

Note that the λ's are correlated and trying to get the marginal for each looks to be intractable analytically.

Run a Gibbs sampler to determine $[\lambda, \beta|S]$.

From this sampler we can get the desired distributions $[\lambda|S]$ and $[\beta|S]$.

A possible Gibbs scheme

Step 1) Sample $\lambda_1 \sim [\lambda_1 | \lambda_{(-1)}, \beta, S]$

…

Step 10) Sample $\lambda_{10} \sim [\lambda_{10} | \lambda_{(-10)}, \beta, S]$

Step 11) Sample $\beta \sim [\beta | \lambda, S]$

where $\lambda_{(-j)} = \{\lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_{10}\}$

Need the following conditional distributions

$$\lambda_j \sim \left[\lambda_j \middle| \lambda_{(-j)}, \beta, S\right] = \left[\lambda_j \middle| \beta, s_j\right]$$

$$= \text{Gamma}\left(\alpha + s_j, \frac{1}{t_j + 1/\beta}\right)$$

$$\beta \sim \left[\beta \middle| \lambda, S\right] = \left[\beta \middle| \lambda\right]$$

$$= \text{IGamma}\left(\gamma + 10\alpha, \frac{1}{\delta + \sum \lambda_i}\right)$$

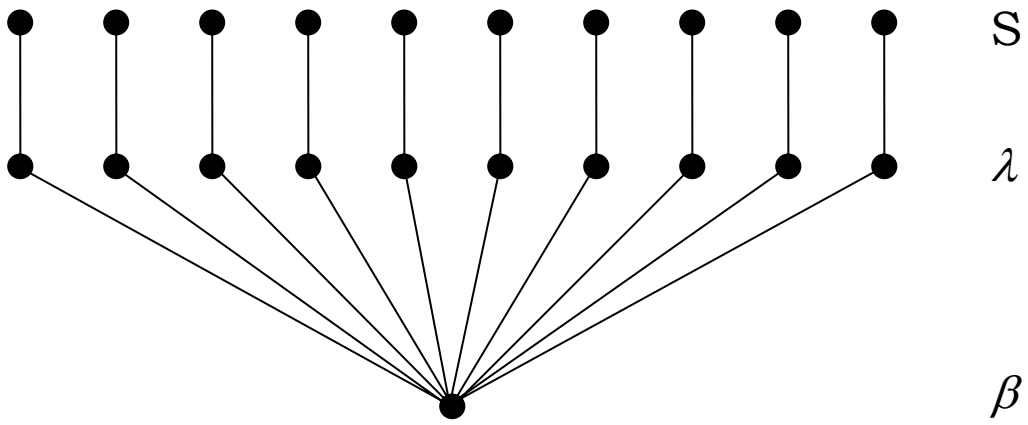
This can be gotten from the joint distribution by including only the terms in the product that contain the random variable of interest

$$[\lambda, \beta, S] = \left(\prod_{i=1}^{10} \frac{(\lambda_i t_i)^{s_i} e^{-\lambda_i t_i}}{s_i!}\right)\left(\prod_{i=1}^{10} \frac{\lambda_i^{\alpha-1} e^{-\lambda_i/\beta}}{\beta^a \Gamma(\alpha)}\right) \frac{\delta^\gamma e^{-\delta/\beta}}{\beta^{\gamma+1}\Gamma(\gamma)}$$

e.g. for λ_j, which terms above have a λ_j in them.

Equivalently, you can do this by looking at the graph structure of the model by only including terms that correspond to edges joining to the node of interest.

e.g. for β, which edges connect with the node for β.



Example Run:

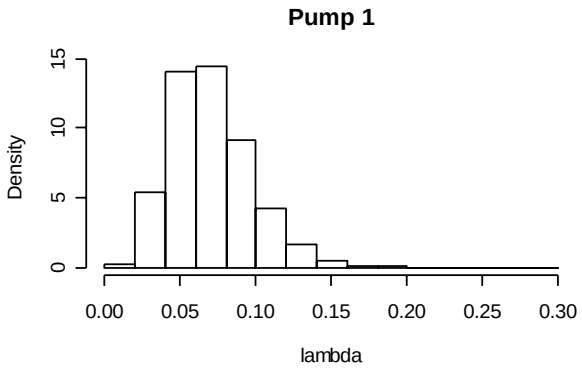
α = 1.8

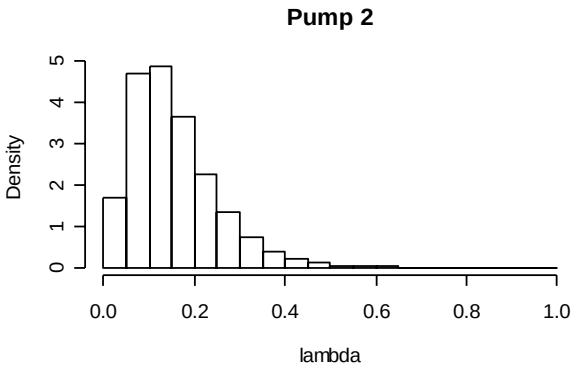
δ = 1

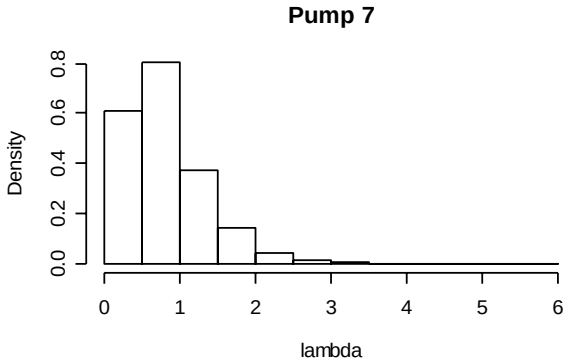
γ = 0.1

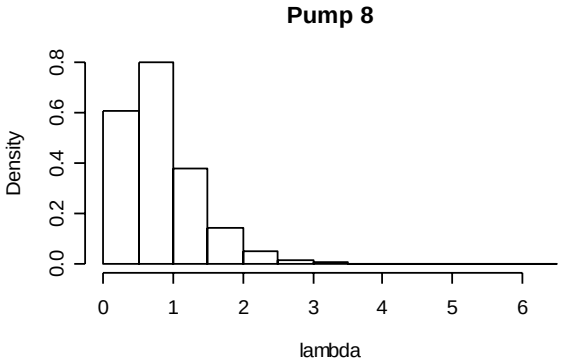
n = 1000

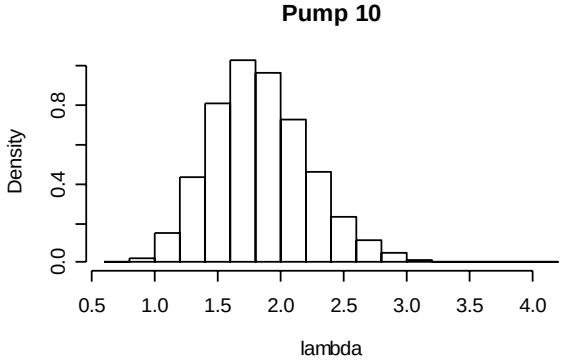
$\beta^0 = \bar{l}$

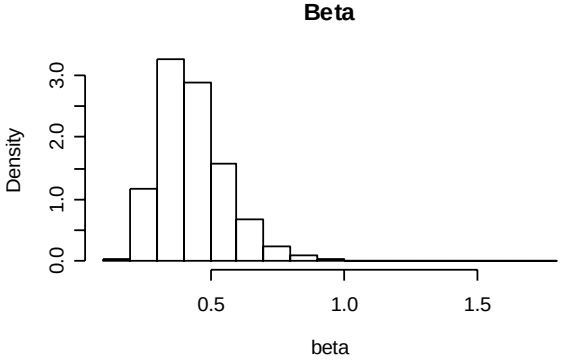

Pump 1
Density
0
5
10
15
0.00
0.05
0.10
0.15
0.20
0.25
0.30
lambda


Pump 2
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
lambda


Pump 7
Density
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
lambda

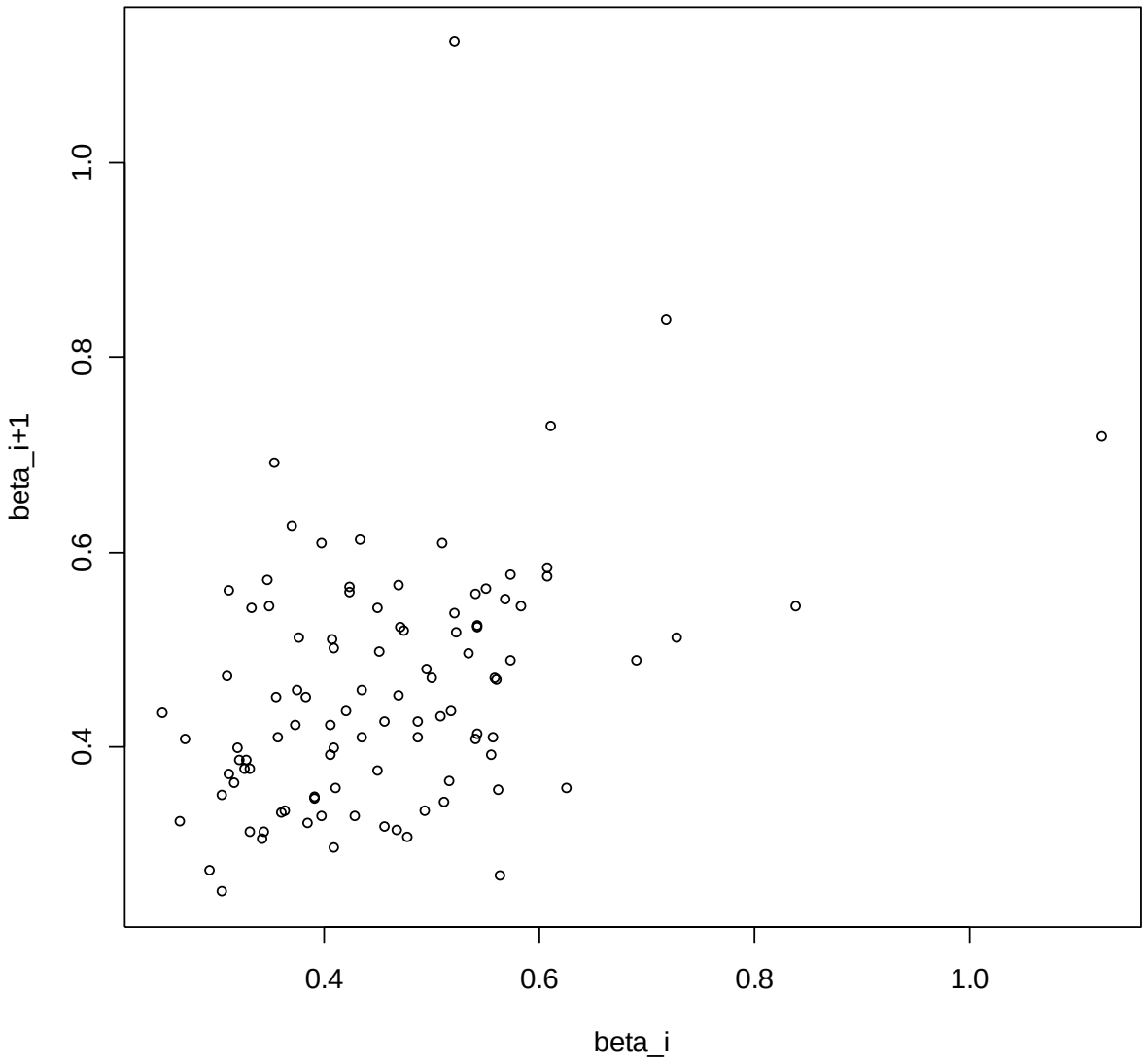

Pump 8
Density
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
lambda


Pump 10
Density
0.0
0.4
0.8
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
lambda

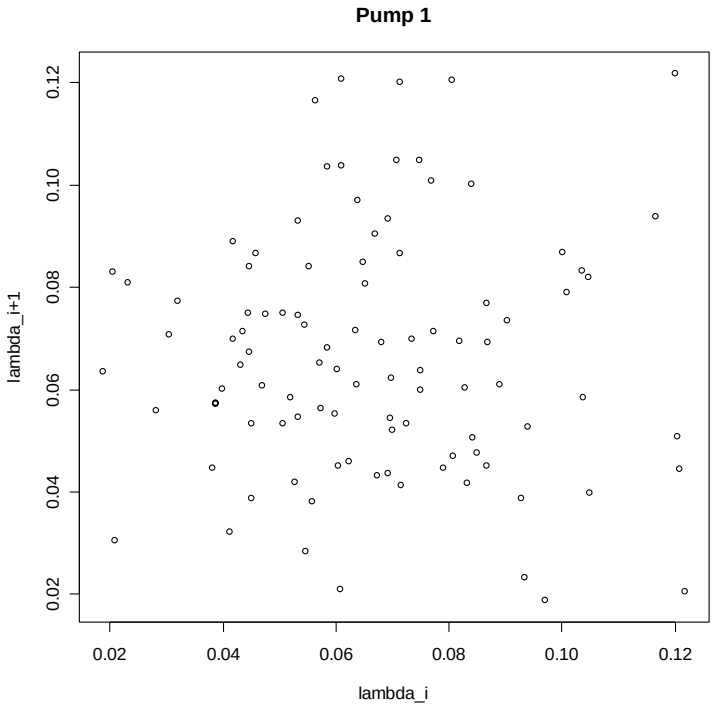

Beta
Density
0.0
1.0
2.0
3.0
0.5
1.0
1.5
beta

Pump	Mean	Median	Std Dev
1	0.0702	0.0668	0.0268
2	0.1542	0.1363	0.0925
3	0.1039	0.0988	0.0399
4	0.1233	0.1206	0.0310
5	0.6263	0.5805	0.2924
6	0.6136	0.6040	0.1351
7	0.8241	0.7102	0.5267
8	0.8268	0.7129	0.5309
9	1.2949	1.2040	0.5776
10	1.8404	1.8121	0.3903

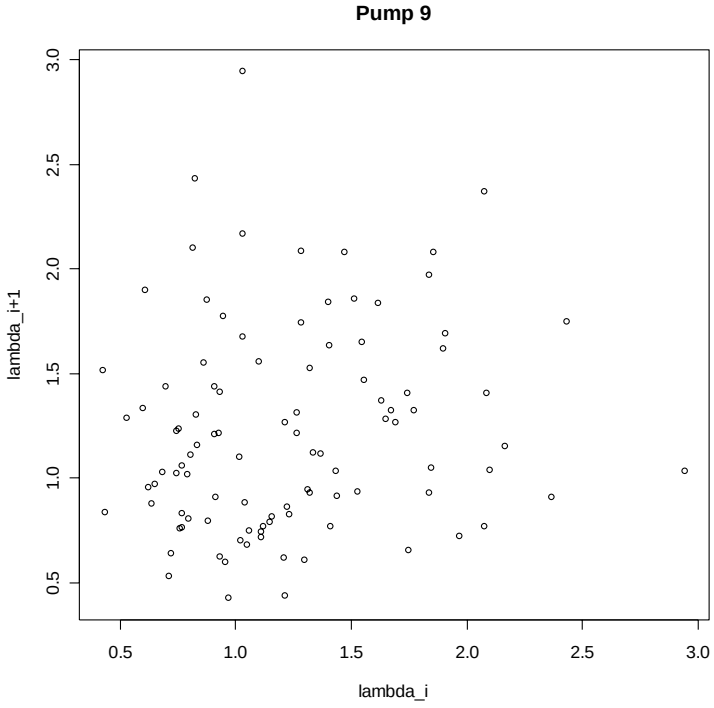
	Mean	Median	Std Dev
Beta	0.4372	0.4161	0.1315



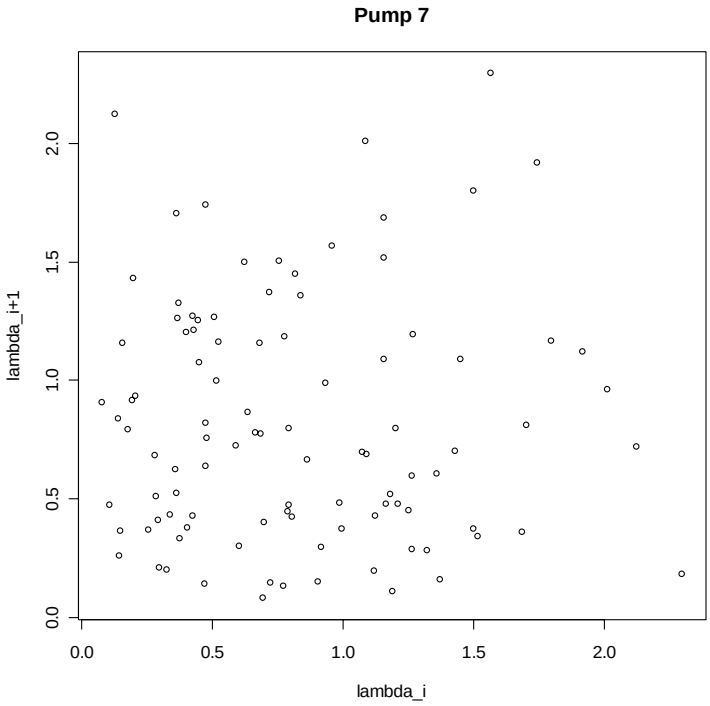
$$\mathrm{Cor}\left(\beta^{i}, \beta^{i+1}\right) = 0.302$$



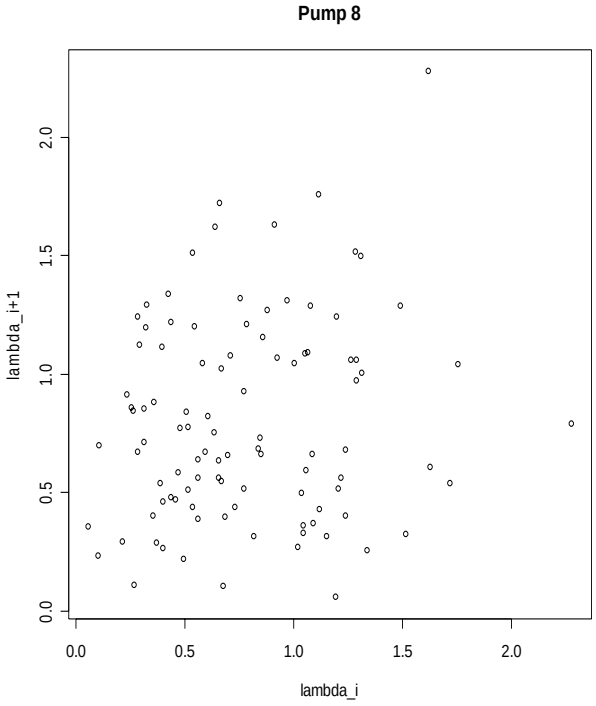
$$\mathrm{Cor}\left(\lambda_1^i, \lambda_1^{i+1}\right) = 0.012$$



$$\mathrm{Cor}\left(\lambda_9^i, \lambda_9^{i+1}\right) = 0.091$$



$$\mathrm{Cor}\left(\lambda_7^i, \lambda_7^{i+1}\right) = 0.063$$



$$\mathrm{Cor}\left(\lambda_8^i, \lambda_8^{i+1}\right) = 0.142$$

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, $j = 1, \ldots, k-1$

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

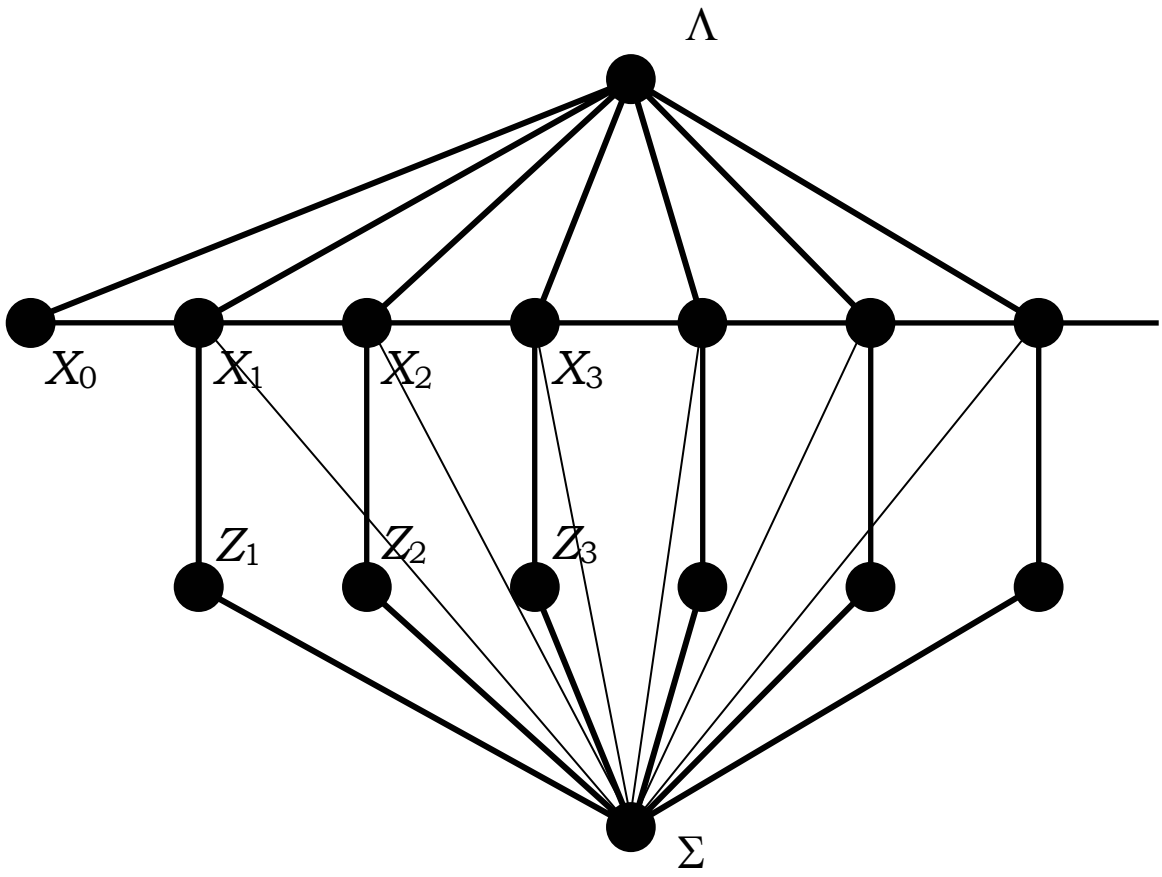
Step k

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, $j = 1, \ldots, k-1$

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:

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

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).

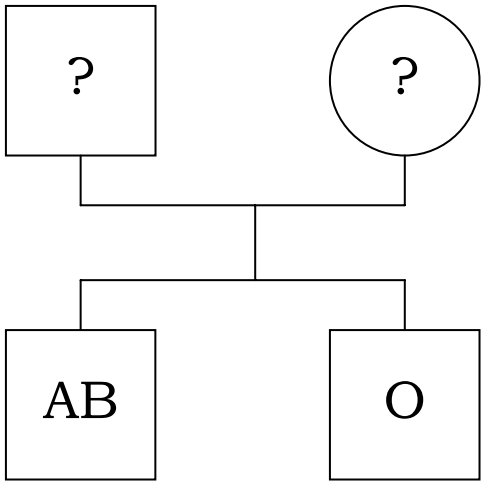
Note that detailed balance $\Rightarrow$ stationarity but stationarity doesn't imply detailed balance.

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

$$\begin{aligned} & P\left[Dad = AO \;\&\; Mom = BO\right] \\ &= P\left[Dad = BO \;\&\; Mom = AO\right] \\ &= 0.5 \end{aligned}$$

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 $\beta^0 = \frac{\bar{l}}{\alpha}$.

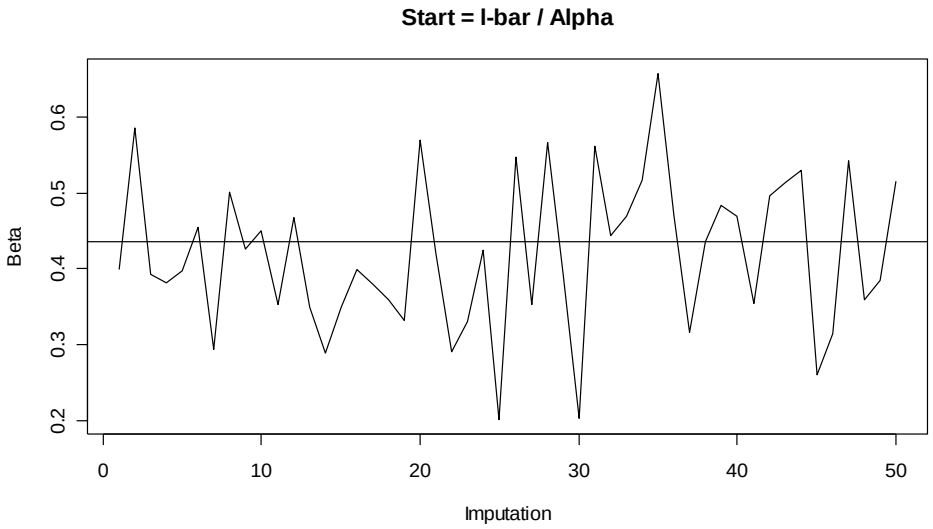
   In target tracking example, set starting positions at each time to average observed positions, the differences of these to get the velocities.

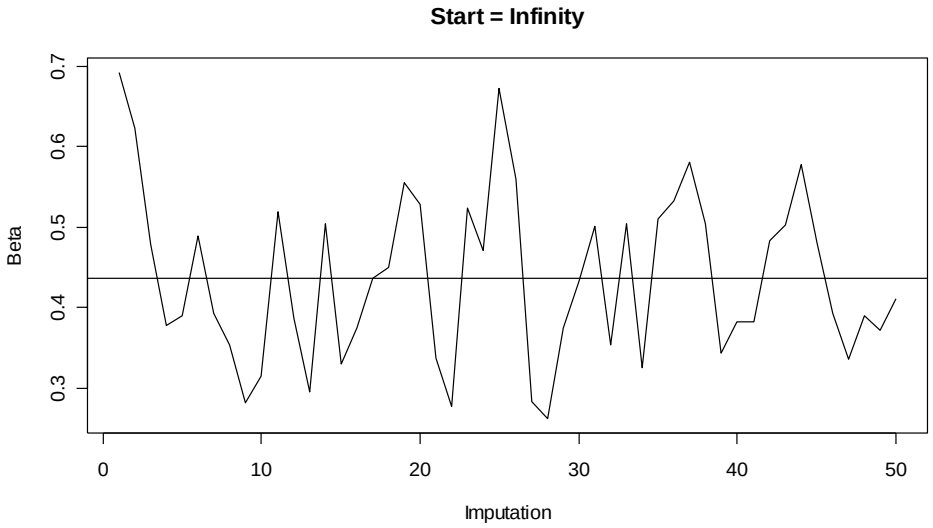
3) Sample from prior

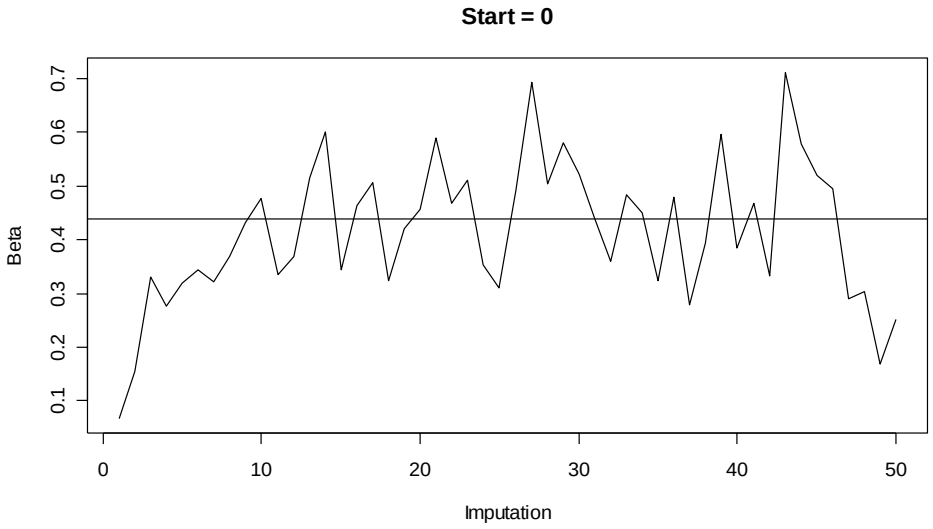
4) Ad hoc choices

   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 = l-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

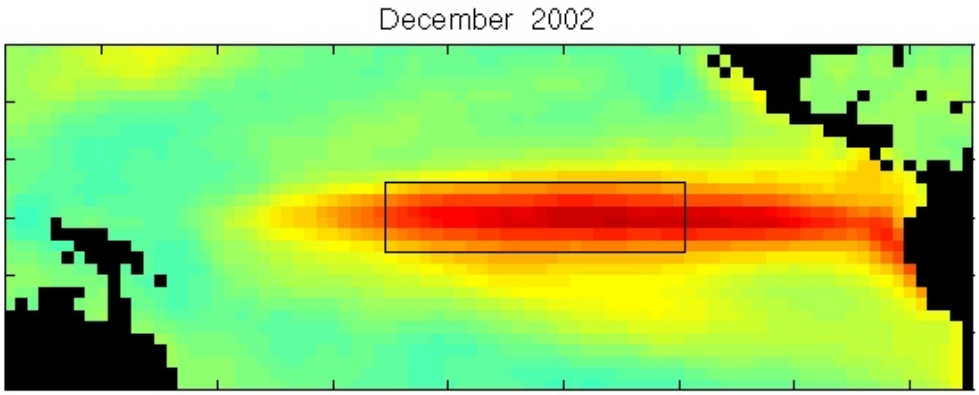

Start = 0
Beta
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
50
Imputation

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

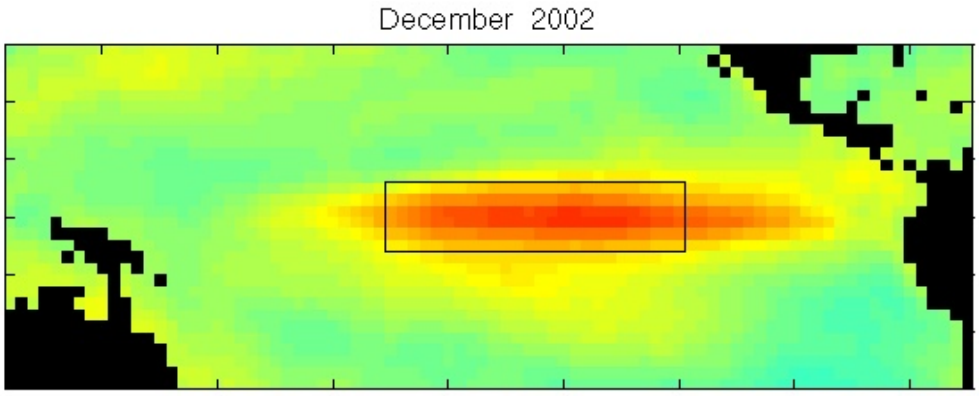
While for this example, the problem clears up quickly, for other problems it can take a while.

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

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

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 (M-H)

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 M-H as for each step,

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

which implies the relationship also 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 $p\left(x \middle| y\right)$ 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).

5) Block at a time

Deal with variables in blocks like the Gibbs sampler. Sometimes referred to Metropolis within Gibbs.

Allows for complex problems to be broken down into simpler ones.

Any M-H style update can be used within each block (e.g. random walk for one block, independence sampler for the next, Gibbs for the one after that).

Allows for a Gibbs style sampler, but without the worry about conjugate distributions in the model to make sampling easier.

Pump Example:

$$s_i \mid \lambda_i \sim \text{Poisson}\left(\lambda_i t_i\right)$$

$$\lambda_i \mid \mu, \sigma^2 \sim \text{LogN}\left(\mu, \sigma^2\right)$$

$$\mu \sim N\left(\nu, \tau^2\right)$$

$$\sigma^2 \sim \text{IGamma}\left(\gamma, \delta\right)$$

Can perform Gibbs on μ and σ^2 but not on λ, due the non-conjugacy of the Poisson and log Normal distributions.

Step i, $i = 1, \ldots, 10$ (M-H):

Sample λ_i from $\lambda_i \left| s, \mu, \sigma^2 \right.$ with proposal

$\lambda_i^* \sim \log N\left(\lambda_i, \theta^2\right)$ (Multiplicative random walk)

$$HR = \frac{\left(\lambda_i^* t_i\right)^{s_i} e^{-\lambda_i^* t_i} \frac{1}{\lambda_i^* \sigma} \phi\left(\frac{\log \lambda_i^* - \mu}{\sigma}\right)}{\left(\lambda_i t_i\right)^{s_i} e^{-\lambda_i t_i} \frac{1}{\lambda_i \sigma} \phi\left(\frac{\log \lambda_i - \mu}{\sigma}\right)} \times \frac{\frac{1}{\lambda_i \theta} \phi\left(\frac{\log \lambda_i - \log \lambda_i^*}{\theta}\right)}{\frac{1}{\lambda_i^* \theta} \phi\left(\frac{\log \lambda_i^* - \log \lambda_i}{\theta}\right)}$$

$$a_{ij} = \min\left(HR, 1\right)$$

Step 11 (Gibbs):

Sample μ from $\mu \left| \lambda, \sigma^2, \nu, \tau^2 \right. \sim N\left(\text{mean}, \text{var}\right)$

where

$$\text{mean} = \text{var}\left(\frac{1}{\sigma^2} \sum \log \lambda_i + \frac{\nu}{\tau^2} \right)$$

$$\text{var} = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1}$$

Step 12 (Gibbs):

Sample σ^2 from

$$\sigma^2 \left| \lambda, \mu, \gamma, \delta \right.$$

$$\sim \text{IGamma}\left(\gamma + 5, \delta + \frac{1}{2} \sum \left(\log \lambda_i - \mu \right)^2 \right)$$

Parameters for run

Burn-in: 1000

Imputations: 100,000

ν = -50

τ^2 = 100

γ = 1

δ = 100

θ^2 = 0.01

Starting values

$$\lambda_i = l_i$$

$$\mu = \frac{1}{10}\sum \log l_i$$

$$\sigma^2 = \frac{1}{9}\sum \left(\log l_i - \mu\right)^2$$

Other options

1) Combine steps 1 – 10 into a single draw.

   With this option all λs change or none do. In the sampler used, whether each λ changes is independent of the other λs.

   The option used is probably preferable, as it should lead to better mixing of the chain.

2) Combine sampling λ, μ, and σ^2 into a single M-H step. Probably suboptimal as the proposal distribution won't be a great match for the joint posterior distribution of λ, μ, and σ^2.

Rejection rates

Having some rejection can be good.

With the multiplicative random walk sampler used, if θ^2 is too small, there will be very few rejections, but the sampler will move too slowly through the space.

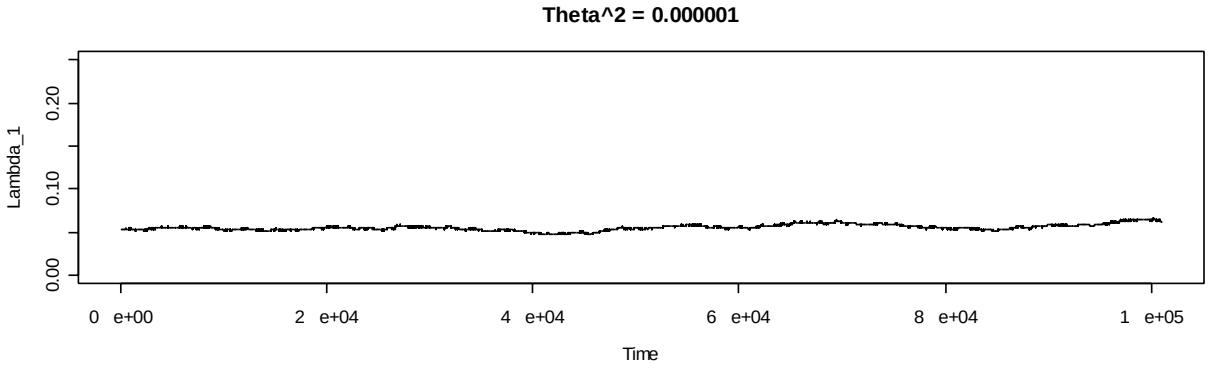
Increasing θ^2 will lead to better mixing, as bigger jumps can be made, though it will lead to higher rejection rates.

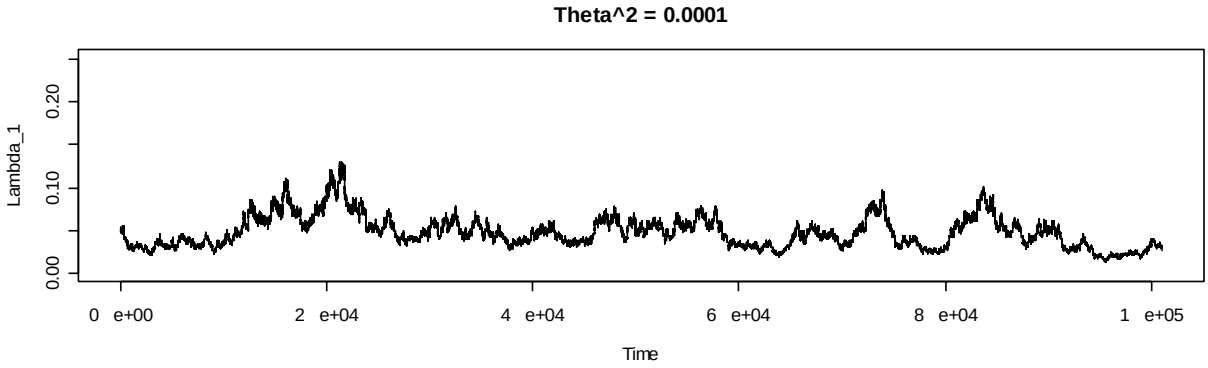
You need to find a balance between rejection rates, mixing of the chain, and coverage of the state space.

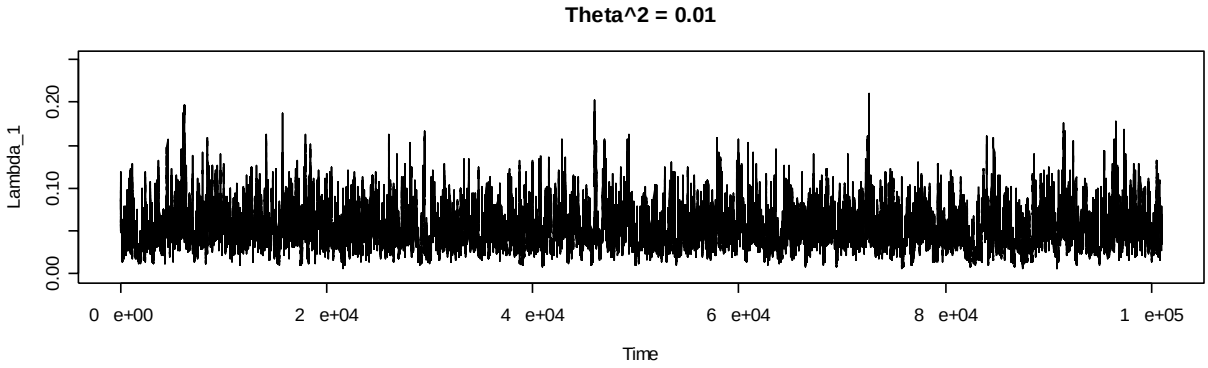
For some problems, a rejection rate of 50% is fine and I've seen reports for large problems using normal random walk proposals the rejection rates of 75% are optimal.

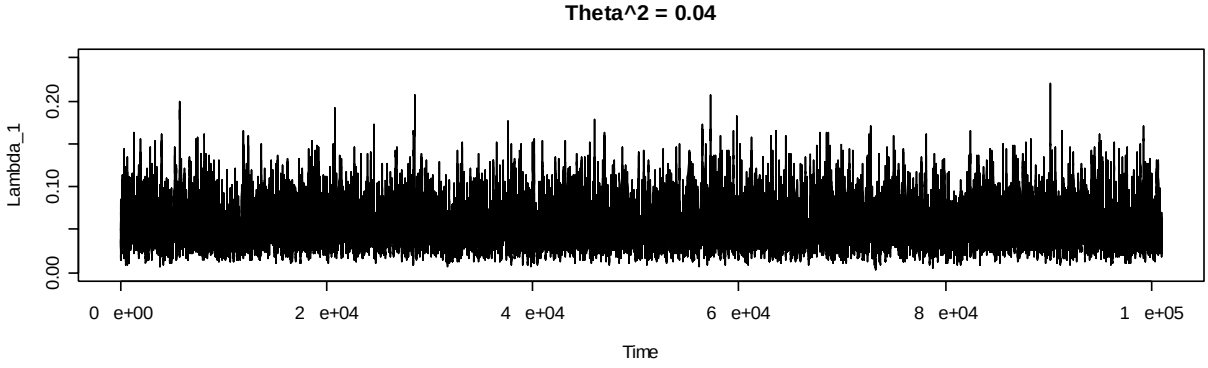
Rejection rates for failure rates proposals under different random walk variances

Pump	0.000001	0.0001	0.01	0.04
1	0.00012	0.00613	0.07045	0.13776
2	0.00009	0.00531	0.03141	0.06130
3	0.00034	0.00784	0.07107	0.13754
4	0.00043	0.01126	0.11705	0.22482
5	0.00028	0.00691	0.05521	0.10705
6	0.00126	0.01442	0.13511	0.26028
7	0.00012	0.00148	0.03027	0.05735
8	0.00007	0.00414	0.02854	0.05824
9	0.00024	0.00559	0.06105	0.12131
10	0.00070	0.01461	0.14790	0.27735


Theta^2 = 0.000001
Lambda_1
0.00
0.10
0.20
0 e+00
2 e+04
4 e+04
6 e+04
8 e+04
1 e+05
Time


Theta^2 = 0.0001
Lambda_1
0.00
0.10
0.20
0 e+00
2 e+04
4 e+04
6 e+04
8 e+04
1 e+05
Time


Theta^2 = 0.01
Lambda_1
0.00
0.10
0.20
0 e+00
2 e+04
4 e+04
6 e+04
8 e+04
1 e+05
Time


Theta^2 = 0.04
Lambda_1
0.00
0.10
0.20
0 e+00
2 e+04
4 e+04
6 e+04
8 e+04
1 e+05
Time

Standard errors in MCMC

As discussed before, the correlation of the chain of the chain must be taken into account when determining standard errors of quantities estimated by the sampler.

Suppose we use $\bar{x}$ to estimate and that the burn-in period was long enough to get into the stationary distribution. Then

$$\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n^2}\left(n + 2\sum_{j=1}^{n-1}(n-j)\rho_j \right)$$

For a reasonable chain, the autocorrelations will die off and so lets assume that they will be negligible for $j > K$. Then the above reduces to

$$\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n^2}\left(n + 2\sum_{j=1}^{K}(n-j)\rho_j \right)$$

If the autocorrelations die off fairly quickly, σ^2 and ρ_j can be estimated consistently (though with some bias) by the usual empirical moments.

Another approach is blocking. Assume that $n = Jm$ for integers J and m. Then let

$$\tilde{x}_j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} x_i; \quad j = 1, \ldots, J$$

Note that $\bar{x} = \bar{\tilde{x}}$. If m is large relative to K, then the correlations between the $\tilde{x}_j$ should negligible and the variance can be estimated as if the $\tilde{x}_j$ were independent.

If the correlation is slightly larger, it might be reasonable to assume that the correlation between $\tilde{x}_j$ and $\tilde{x}_{j+1}$ is some value ρ to be determined, but that correlations at larger lags are negligible. In this case

$$\mathrm{Var}(\bar{x}) \doteq \mathrm{Var}(\tilde{x}_j) \frac{1 + 2\rho}{J}$$

Estimates with $m = 100$

Parameter	$\bar{x}$	SE	ρ
λ_1	0.05290	0.00071	0.36116
λ_2	0.06926	0.00277	0.66197
λ_3	0.07837	0.00106	0.35354
λ_4	0.11053	0.00056	0.10520
λ_5	0.56167	0.01119	0.46975
λ_6	0.60546	0.00237	0.10960
λ_7	0.92318	0.04068	0.67346
λ_8	0.90361	0.03766	0.63510
λ_9	1.82900	0.02884	0.33629
λ_{10}	2.10188	0.00726	0.05263
μ	-2.52492	0.01384	0.41517
σ^2	27.15958	0.09967	0.07579

Estimates with m = 1000

Parameter	$\bar{x}$	SE	ρ
λ_1	0.05290	0.00075	0.13239
λ_2	0.06926	0.00399	0.18756
λ_3	0.07837	0.00088	-0.13079
λ_4	0.11053	0.00045	-0.15794
λ_5	0.56167	0.01205	-0.00838
λ_6	0.60546	0.00226	-0.07845
λ_7	0.92318	0.06081	0.12201
λ_8	0.90361	0.04822	0.04495
λ_9	1.82900	0.03303	0.07779
λ_{10}	2.10188	0.00757	0.06487
μ	-2.52492	0.01981	0.15224
σ^2	27.15958	0.13956	0.29726

Standard error estimates for pump example

	$m = 1000$	$m = 100$	Independent
λ_1	0.000752	0.000710	0.000075
λ_2	0.003992	0.002769	0.000205
λ_3	0.000885	0.001063	0.000111
λ_4	0.000446	0.000555	0.000094
λ_5	0.012051	0.011193	0.001009
λ_6	0.002258	0.002373	0.000439
λ_7	0.060813	0.040679	0.002970
λ_8	0.048219	0.037656	0.002807
λ_9	0.033030	0.028835	0.002945
λ_{10}	0.007568	0.007264	0.001428
μ	0.019808	0.013840	0.005729
σ^2	0.139560	0.099674	0.056767