Markov Chain Monte Carlo - 2

Let's go back to the Gibbs sampler

Suppose that $\theta = (\theta_1, \ldots, \theta_k)$ is a vector of unknown parameters. Then consider the following Markov chain

1. Start with an initial value $\theta^{(0)}$; set $t = 0$.
2. Sample $\theta_1^{(t+1)}$ from $p(\theta_1 | x, \theta_2^{(t)}, \ldots, \theta_k^{(t)})$.
3. Sample $\theta_2^{(t+1)}$ from $p(\theta_2 | x, \theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_k^{(t)})$.
4. $\ldots$
5. Sample $\theta_k^{(t+1)}$ from $p(\theta_k | x, \theta_1^{(t+1)}, \ldots, \theta_{k-1}^{(t+1)})$.
6. Increase t and return to 2.

Note that at each step a component of the unknown parameter is sampled from its full conditional distribution, given the data, and the current value of all other components of the parameter.

The Gibbs sampler

Indeed, this Markov chain is a special case of MH sampling: using the full conditional distributions as the proposal distribution in an MH sampler gives acceptance probability $A = 1$.

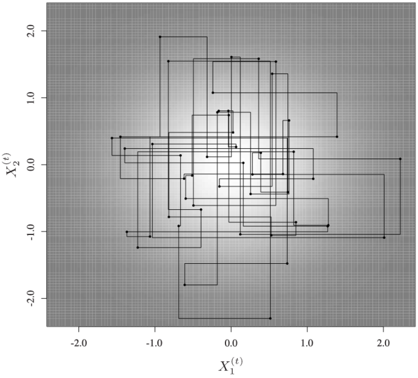
Effect of correlation on the Gibbs sampler

The Gibbs sampler updates one co-ordinate at a time in order to obtain a new sample point.

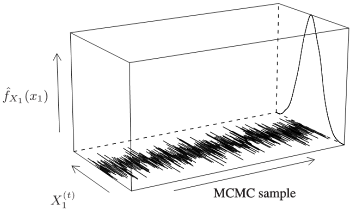
Strong correlation between the elements of $\theta = (\theta_1, \ldots, \theta_k)$ (that is, correlation between the co-ordinates of the target distribution π), will slow down the Gibbs sampler, so that it explores the sample space more slowly.

We illustrate this using the Gibbs sampler for a bivariate normal. See problems from the lab later.

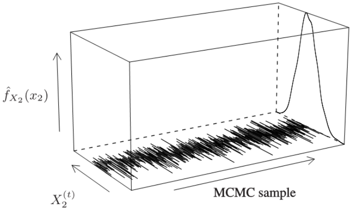
Visual interpretation: Figures due to Ioana A. Cosma.; $\mathbf{X} = \theta$



(a) First 50 iterations of $(X_1^{(t)}, X_2^{(t)})$

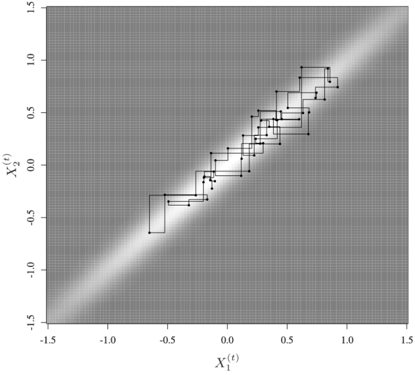


(b) Path of $X_1^{(t)}$ and estimated density of X after 1,000 iterations

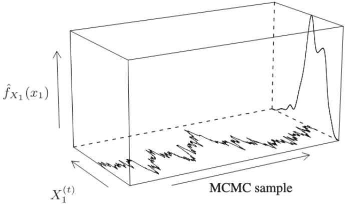


(c) Path of $X_2^{(t)}$ and estimated density of X_2 after 1,000 iterations

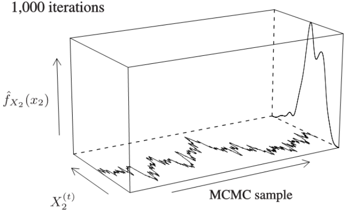
Figure 4.4. Gibbs sampler for a bivariate standard normal distribution with correlation $\rho(X_1, X_2) = 0.3$.



(a) First 50 iterations of $(X_1^{(t)}, X_2^{(t)})$



(b) Path of $X_1^{(t)}$ and estimated density of X_1 after 1,000 iterations



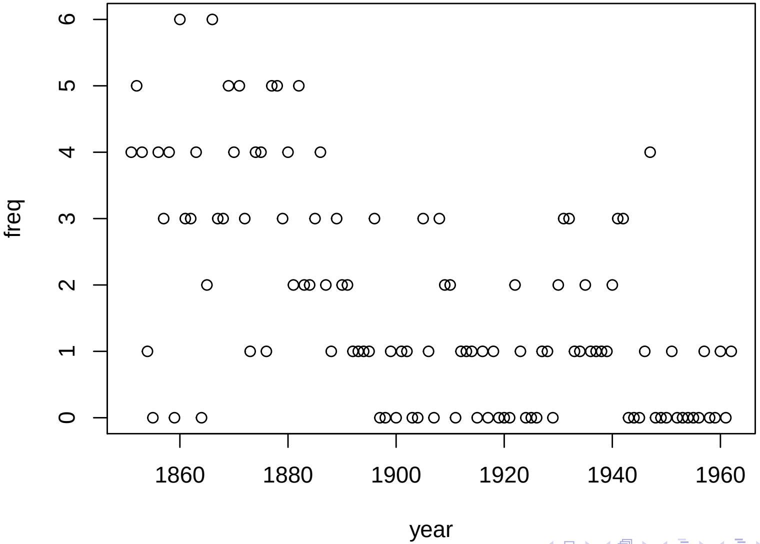
(c) Path of $X_2^{(t)}$ and estimated density of X_2 after 1,000 iterations

Figure 4.5. Gibbs sampler for a bivariate normal distribution with correlation $\rho(X_1, X_2) = 0.99$.

Example: Poisson change point model

Data from Jarret (1979), A note on the intervals between coal mining disasters. *Biometrika* 66, pp. 191–193.
The data gives the dates of explosions killing 10 or more miners, from 1851 to 1962.
Let Y_i be the number of such disasters in year i

```
> library(boot)
> data(coal)
> when <- floor(coal)
> year <- 1851:1962
> freq <- sapply(year, function(x, y) sum(y==x), y=when)
> n <- length(freq)
> plot(year, freq)
```



freq
0
1
2
3
4
5
6
1860
1880
1900
1920
1940
1960
year

Was there a change in the rate of disasters? To (help) answer this question we use the following model.

$$
\begin{aligned}
Y_i &\sim \text{pois}(\lambda_1), \text{ for } i = 1, \ldots, M \\
Y_i &\sim \text{pois}(\lambda_2), \text{ for } i = M+1, \ldots, n \\
\lambda_j &\sim \Gamma(\alpha_j, \beta_j), \text{ for } j = 1, 2 \\
M &\sim U\{1, \ldots, n\}
\end{aligned}
$$

The joint density is given by the product of the conditional densities for each node

$$
\begin{aligned}
&p(\mathbf{y}, \lambda_1, \lambda_2, m) \\
&\quad \propto \prod_{i=1}^{m} \frac{e^{-\lambda_1} \lambda_1^{y_i}}{y_i!} \prod_{j=m+1}^{n} \frac{e^{-\lambda_2} \lambda_2^{y_j}}{y_j!} \lambda_1^{\alpha_1 - 1} e^{-\beta_1 \lambda_1} \lambda_2^{\alpha_2 - 1} e^{-\beta_2 \lambda_2}
\end{aligned}
$$

Collecting like terms we get the posterior

$$\begin{aligned} p(\lambda_1, \lambda_2, m|\mathbf{y}) &\propto e^{-(m+\beta_1)\lambda_1} \lambda_1^{\sum_{i=1}^m y_i + \alpha_1 - 1} \\ &\quad \times e^{-(n-m+\beta_2)\lambda_2} \lambda_2^{\sum_{j=m+1}^n y_j + \alpha_2 - 1} \end{aligned}$$

and thus the conditioned marginals are

$$\begin{aligned} p(\lambda_1|\lambda_2, m, \mathbf{y}) &\sim \Gamma(\alpha_1 + \sum_{i=1}^m y_i, \beta_1 + m) \\ p(\lambda_2|\lambda_1, m, \mathbf{y}) &\sim \Gamma(\alpha_2 + \sum_{j=m+1}^n y_j, \beta_2 + n - m) \\ p(m|\lambda_1, \lambda_2, \mathbf{y}) &\propto \lambda_1^{\sum_{i=1}^m y_i} \lambda_2^{\sum_{j=m+1}^n y_j} e^{(\lambda_2 - \lambda_1)m} \\ &\propto \left(\frac{\lambda_1}{\lambda_2}\right)^{\sum_{i=1}^m y_i} e^{(\lambda_2 - \lambda_1)m} \end{aligned}$$

Note that $p(m|\lambda_1, \lambda_2, \mathbf{y})$ is not a known distribution, but it is finite so we can easily simulate samples from this distribution.

We impliment a Gibbs sampler in R for this model.

Example: `Gibbs_example_coal.pdf`