Btry 694 Class Homework 7

11/20/98

1 Problem 7.1

Haiyue Song

For the Gibbs sampler (7.1.1),

(a) Show that the sequence (X_t, Y_t) is a Markov chain, as is each sequence (X_t) and (Y_t).

(b) Show that $f_X(\cdot)$ and $f_Y(\cdot)$ are, respectively, the invariant densities of the X and Y sequences of (7.1.1).

Proof:

(a) Set $X_0 = x_0$, and for $t = 1, t, \ldots$, generate

$$\begin{aligned} Y_t &\sim f_{Y|X}(\cdot|x_{t-1}) \\ X_{t+1} &\sim f_{X|Y}(\cdot|y_t) \end{aligned} \tag{7.1.1}$$

where $f_{X|Y}$ and $f_{Y|X}$ are the conditional distributions. By the nature of construction, we know (X_t, Y_t) only depends on $(x_{t-1}, y_{t-1}$, so the pair,(X_t, Y_t), is a Markov chain.

We know (X_t) has transition density

$$K(x, x^*) = \int f_{Y|X}(y|x) f_{X|Y}(x^*|y)dy, \tag{1}$$

therefore, (X_t) is independent from X_{t-2}, X_{t-3},. . ., and is simulated conditionally only on x_{t-1}. Thus, (X_t) is a Markov chain. And the same arguments applied to (Y_t).

(b) We know (X_t) has transition kernal (1), we want to show that

$$\pi(B) = \int k(x, B)\pi(dx).$$

$$\begin{aligned}
\int k(x, x^*) f_X(x)dx &= \int\int f_{Y|X}(y|x) f_{X|Y}(x^*|y)dy f_X(x)dx \\
&= \int f_{X|Y}(x^*|y) \int f_{Y|X}(y, x) f_x(x)dxdy \\
&= \int f_{X|Y}(x^*|y) \int f_{XY}(x, y)dxdy \\
&= \int f_{X|Y}(x^*|y) f_Y(y)dy \\
&= \int f_{XY}(x^*, y)dy \\
&= f_X(x^*)
\end{aligned}$$

Hence, f_X is invariant density of (X_t). The same argument applies to (Y_t).

2 Problem 7.4

Mitchell Warachka

See attached graphs for results of gibbs sampler. With n=1000, the mean of the sample was -.0796, variance 3.274, minimum -5.2366, and maximum 5.0482. The mean and variance are close to theoretical 0 and 3.2899.

7.28

part a Show $var(E[X_k|X_0]) = E[E[X_k|X_0]]^2$

$var(E[X_k|X_0])$

$$\begin{aligned}
&= E[E[X_k|X_0]]^2 - (E[E[X_k|X_0]])^2 \\
&\Rightarrow \text{ show second term on right hand side is zero} \\
& \quad (E[E[X_k|X_0]])^2 = (E[E[E...E[X_k|X_{k-1}|X_{k-2}...|X_0]]])^2 \\
& \quad \text{add in additional information sets in form of intermediate observations} \\
&= (E[X_0])^2 = 0
\end{aligned}$$

since process Markov, need only condition back one period and smallest sigma algebra always determines expectation so introducing intermediate observations does not alter original expression. See later martingale discussion.

i.e. when k=2

$$
\begin{aligned}
(E[E[X_2|X_0]])^2 &= (E[E[E[X_2|X_1]|X_0]])^2 \\
&= (E[E[X_1|X_0]])^2 \\
&= (E[X_0])^2 \\
&= 0
\end{aligned}
$$

to show second part $var(E[X_k|X_0]) = var(E[X_{k+1}|X_0])$ use the first result and re express as $E[E[X_k|X_0]]^2 = E[E[E_{k+1}|X_0]]^2$ but using same conditioning argument

$$
E[E[X_{k+1}|X_0]]^2 = E[E[X_k|X_0]]^2 = E[X_0]^2
$$

part b Use property that this process is $X_{k+1} = g(X_k) + v_{k+1}$ where v_{k+1} is independent of x_k and mean zero since $E[X_{k+1}]$ must equal zero, this also constrains $g(X_k) = X_k$ or g to just be identity map and results in X_k being a martingale.

$X_{k+1} = X_k + v_{k+1}$

$$
\begin{aligned}
&\Rightarrow \; var([X_{k+1}|X_0]) = var([X_k|X_0]) + var([v_{k+1}|X_0]) \\
&\Rightarrow \; var([X_{k+1}|X_0]) - var([X_k|X_0]) = var([v_{k+1}|X_0]) \geq 0 \\
&\Rightarrow \; E[var([X_{k+1}|X_0]) - var([X_k|X_0])] \geq 0
\end{aligned}
$$

$$\Rightarrow \quad E[var([X_{k+1}|X_0])] \geq E[var([X_k|X_0])]$$

this means that $var(X_k|X_0)$ is an increasing and positive sequence which allows monotone convergence to be used to pass limit as k tends to infinity through expectations in variance.

$$\begin{aligned}
lim_{k->\infty} var(E[X_k|X_0]) &= var(lim_{k->\infty} E[X_k|X_0]) \\
&= var(E[X_\infty|X_0]) \quad \text{by martingale convergence theorem} \\
&= E[var(X_0)] \\
&= \sigma^2
\end{aligned}$$

MATLAB CODE:

```
function [x]=logistic(n);


%Gibbs sampler for logistic distribution with u=0, B=1


x=zeros(n,1);
%start value
x(1)=0;


for (i=1:n)
```

```
        y=exp(-x(i))/((1+exp(-x(i)))^2);

        %y between 0 and .25 (maximum lower than normal, thicker tails)
        %increase range of k to get comparable variance

        k=22*unifrnd(0,y);
        x(i)=unifrnd(-k,k);
end
```

3 Problem 7.29

Matt Tom

Suppose that we have a two-dimensional random variable

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

Then the conditional densities for the Gibbs sampler are:

$$X_2|y_1 \sim \mathcal{N}\left(\rho y_1, 1-\rho^2\right),$$

$$Y_2|x_2 \sim \mathcal{N}\left(\rho x_2, 1-\rho^2\right).$$

First, show that for the X-chain, the transition kernel is

$$K(x_1, x_2) = \frac{1}{2\pi(1-\rho^2)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(x_2-\rho y)^2}{1-\rho^2}} e^{-\frac{1}{2}\frac{(y-\rho x_1)^2}{1-\rho^2}} dy$$

We know that $f(X_2|X_1) = \int\limits_{-\infty}^{\infty} f(X_2|Y_1) f(Y_1|X_1)$. Plugging in the conditional densities

$$f(X_2|Y_1) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2}\frac{(x_2-\rho y)^2}{1-\rho^2}}$$
$$f(Y_1|X_1) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2}\frac{(y-\rho x_1)^2}{1-\rho^2}}$$

gives the desired result.

Show at $X_i \sim \mathcal{N}(0,1)$ is the invariant distribution. I think I'll hold off on this one.

Show that $X_2|x_1 \sim \mathcal{N}(\rho^2 x_1, 1-\rho^4)$. Let's look at the exponent in the integral in part A. Forget the factor $-\frac{1}{2(1-\rho^2)}$. What's left is

$$\begin{aligned}
(x_2-\rho y)^2 + (y-\rho x_1)^2 &= (1+\rho^2)y^2 - 2y(\rho x_2 + \rho x_1) + x_2^2 + \rho^2 x_1^2 \\
&= (1+\rho^2)\left(y^2 - 2y\left(\frac{\rho x_2+\rho x_1}{1+\rho^2}\right) + \left(\frac{\rho x_2+\rho x_1}{1+\rho^2}\right)^2\right) \\
&\qquad - \frac{(\rho x_2+\rho x_1)^2}{1+\rho^2} + \frac{x_2^2+\rho^2 x_2^2+\rho^2 x_1^2+\rho^4 x_1^2}{1+\rho^2}\,. \\
&= (1+\rho^2)\left(y - \frac{\rho x_2+\rho x_1}{1+\rho^2}\right)^2 + \frac{x_2^2 - 2\rho^2 x_1 x_2 + \rho^4 x_1^2}{1+\rho^2} \\
&= (1+\rho^2)\left(y - \frac{\rho x_2+\rho x_1}{1+\rho^2}\right)^2 + \frac{1}{1+\rho^2}\left(x_2 - \rho^2 x_1\right)^2
\end{aligned}$$

Plugging this into the formula in part A gives

$$\begin{aligned}
K(x_1,x_2) &= \frac{1}{2\pi(1-\rho^2)}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{1+\rho^2}{1-\rho^2}\left(y-\frac{\rho x_2+\rho x_1}{1+\rho^2}\right)^2} e^{-\frac{1}{2}\frac{(x_2-\rho^2 x_1)^2}{(1-\rho^2)(1+\rho^2)}}dy \\
&= \frac{1}{\sqrt{2\pi(1-\rho^4)}}e^{-\frac{1}{2}\frac{(x_2-\rho^2x_1)^2}{1-\rho^4}}\int_{-\infty}^{\infty}\sqrt{\frac{1+\rho^2}{2\pi(1-\rho^2)}}e^{-\frac{1}{2}\frac{1+\rho^2}{1-\rho^2}\left(y-\frac{\rho x_2+\rho x_1}{1+\rho^2}\right)^2}dy\;, \\
&= \frac{1}{\sqrt{2\pi(1-\rho^4)}}e^{-\frac{1}{2}\frac{(x_2-\rho^2x_1)^2}{1-\rho^4}}
\end{aligned}$$

which is what we wanted.

Back to Part B. Suppose that X_1 has a standard normal distribution. Then

$$\begin{aligned}
f(x_2) &= \int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)f_{X_1}(x_1)dx_1 \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-\rho^4)}}e^{-\frac{1}{2}\frac{(x_2-\rho^2x_1)^2}{1-\rho^4}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}dx_1 \; .
\end{aligned}$$

Adding the exponents and forgetting about the $-\frac{1}{2}$ gives

$$\begin{aligned}
\frac{(x_2-\rho^2x_1)^2}{1-\rho^4}+x_1^2 &= \frac{x_2^2-2\rho^2x_1x_2+\rho^4x_1^2}{1-\rho^4}+\frac{x_1^2-\rho^4x_1^2}{1-\rho^4} \\
&= \frac{x_1^2-2\rho^2x_1x_2+\rho^4x_2^2+x_2^2-\rho^4x_2^2}{1-\rho^4} \\
&= \frac{(x_1-\rho^2x_2)^2}{1-\rho^4}+\frac{x_2^2-\rho^4x_2^2}{1-\rho^4} \\
&= \frac{(x_1-\rho^2x_2)^2}{1-\rho^4}+x_2^2 \; .
\end{aligned}$$

Playing on a theme gives

$$\begin{aligned}
f(x_2) &= \int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1) f_{X_1}(x_1) dx_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^4)}} e^{-\frac{1}{2}\frac{(x_1-\rho^2 x_2)^2}{1-\rho^4}} dx_1 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}. \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}
\end{aligned}$$

We can conclude that $X_1 \sim \mathcal{N}(0,1)$ is the invariant distribution.

Show that we can write $X_k = \rho^2 X_{k-1} + U_k$ **where the** U_k **are i.i.d.** $\mathcal{N}(0, 1-\rho^4)$**, and that** $\mathrm{cov}(X_0, X_k) = \rho^{2k}$**. Deduce that the covariances go to 0.** That $X_k = \rho^2 X_{k-1} + U_k$ follows straight from part B. From here we can say that $X_k = (\rho^2)^k X_0$ plus some random variable independent of X_0. Thus,

$$\mathrm{cov}(X_k, X_0) = \mathrm{cov}\left((\rho^2)^k X_0, X_0\right) = \rho^{2k} \mathrm{Var}(X_0).$$

If we assume that the variance of X_0 is 1, then we conclude that $\mathrm{cov}(X_k, X_0) = \rho^{2k}$. Since ρ must be less than 1, the sequence $\{\mathrm{cov}(X_k, X_0) : k = 1, 2, \ldots\}$ must converge to 0.

4 Problem 7.38

Bárbara Patricia González

In Example 7.4.1 we noted that a pair of conditional densities is *functionally* *compatible* if the ratio

$$f_1(x|y)/f_2(y|x) = h_1(x)/h_2(y),$$

for some functions h_1 and h_2. This is a necessary condition for a joint density to exist, but not a sufficient condition. If such a joint density does exist, the pair f_1 and f_2 would be *compatible*.

a) Formulate the definitions of compatible and functionally compatible for a set of densities $f_1, ..., f_m$.

b) Show that if $f_1, ..., f_m$ are the full conditionals from a hierarchical model, they are functionally compatible.

c) Prove the following theorem, due to Hobert and Casella (1998), which shows there cannot be any stationary probability distribution for the chain to converge to unless the densities are compatible:

Theorem 7.5.1 *Let $f_1, ..., f_m$ be a set of functionally compatible conditional densities on which a Gibbs sampler is based. The resulting Markov chain is positive recurrent if and only if $f_1, ..., f_m$ are compatible.*

(*Note*: Compatibility of a set of densities was investigated by Besag 1974, Arnold and Press 1989, and Gelman and Speed 1993.)

5 Solution

a) A set of conditional densities $f_1, ..., f_m$ is compatible if there exists a joint density f which generates them:

$$f_i(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_m) = \frac{f(x_1, ..., x_m)}{\int f(x_1, ..., x_m)dx_i} \text{ for } i = 1, ..., m$$

A set of conditional densities $f_1, ..., f_m$ is functionally compatible if for each fixed $(x'_1, ..., x'_m)$, all of the $\begin{pmatrix} m! \\ 2 \end{pmatrix}$ possible ratios of the form

$$\left\{\frac{\prod_{j=1}^{m} f_{l_j}(x_{l_j}|x_{l_1}, ..., x_{l_{j-1}}, x'_{l_{j+1}}, ..., x'_{l_m})}{\prod_{j=2}^{m} f_{l_j}(x'_{l_j}|x_{l_1}, ..., x_{l_{j-1}}, x'_{l_{j+1}}, ..., x'_{l_m})}\right\}\left\{\frac{\prod_{j=1}^{m} f_{k_j}(x_{k_j}|x_{k_1}, ..., x_{k_{j-1}}, x'_{k_{j+1}}, ..., x'_{k_m})}{\prod_{j=2}^{m} f_{k_j}(x'_{k_j}|x_{k_1}, ..., x_{k_{j-1}}, x'_{k_{j+1}}, ..., x'_{k_m})}\right\}^{-1}$$

is equal to some constant that only depends on $(x'_1, ..., x'_m)$ and is independent of $(x_1, ..., x_m)$, where $(x_{l_1}, ..., x_{l_m})$ and $(x_{k_1}, ..., x_{k_m})$ are different permutations of $(x_1, ..., x_m)$ (Hobert 1994).

b) In the case of a hierarchical model, every f_i is the joint distribution of the data and the parameters considered as a function of one of the parameters. Then the ratios of the form

$$\left\{\frac{\prod_{j=1}^{m} f_{l_j}(x_{l_j}|x_{l_1},...,x_{l_{j-1}},x'_{l_{j+1}},...,x'_{l_m})}{\prod_{j=2}^{m} f_{l_j}(x'_{l_j}|x_{l_1},...,x_{l_{j-1}},x'_{l_{j+1}},...,x'_{l_m})}\right\}\left\{\frac{\prod_{j=1}^{m} f_{k_j}(x_{k_j}|x_{k_1},...,x_{k_{j-1}},x'_{k_{j+1}},...,x'_{k_m})}{\prod_{j=2}^{m} f_{k_j}(x'_{k_j}|x_{k_1},...,x_{k_{j-1}},x'_{k_{j+1}},...,x'_{k_m})}\right\}^{-1}$$

will be constant since every conditional density has the same form.

c) If the densities are compatible, then there is a joint density f that generates them. Now we claim that this is the invariant density, so the chain is positive recurrent.

The Kernel of the Gibbs Sampler is given by the conditional density

$$k(x_1^{(t+1)},...,x_m^{(t+1)}|x_1^{(t)},...,x_m^{(t)}) =$$

$$= f_1(x_1^{(t+1)}|x_2^{(t)},...,x_m^{(t)})f_2(x_2^{(t+1)}|x_1^{(t+1)},x_3^{(t)},...,x_m^{(t)})...f_m(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-1}^{(t+1)})$$

So it is enough to prove that

$$f(x_1^{(t+1)},...,x_m^{(t+1)}) = \int k(x_1^{(t+1)},...,x_m^{(t+1)}|x_1^{(t)},...,x_m^{(t)})f(x_1^{(t)},...,x_{m-1}^{(t)})dx_1^{(t)}...dx_m^{(t)}$$

Let $f^{(i)}(x_1,...,x_{i-1},x_{i+1},...,x_m) = \int f(x_1,...,x_m)dx_i$. So we have that

$$\int k(x_1^{(t+1)},...,x_m^{(t+1)}|x_1^{(t)},...,x_m^{(t)})f(x_1^{(t)},...,x_m^{(t)})dx_1^{(t)}...dx_m^{(t)} =$$

$$= \int f_1(x_1^{(t+1)}|x_2^{(t)},...,x_m^{(t)}) f_2(x_2^{(t+1)}|x_1^{(t+1)},x_3^{(t)},...,x_m^{(t)})...f_m(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-1}^{(t+1)})\times...$$

$$... \times f(x_1^{(t)},...,x_m^{(t)})dx_1^{(t)}...dx_m^{(t)} =$$

$$= f_m(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-1}^{(t+1)}) \times ...$$

$$...\times\int f_1(x_1^{(t+1)}|x_2^{(t)},...,x_m^{(t)}) f_2(x_2^{(t+1)}|x_1^{(t+1)},x_3^{(t)},...,x_m^{(t)})...f_{m-1}(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-2}^{(t+1)},x_m^{(t)})\times...$$

$$... \times f^{(1)}(x_2^{(t)},...,x_m^{(t)})dx_2^{(t)}...dx_m^{(t)} =$$

$$= f_m(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-1}^{(t+1)}) \times ...$$

$$... \times \int f_2(x_2^{(t+1)}|x_1^{(t+1)},x_3^{(t)},...,x_m^{(t)})...f_{m-1}(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-2}^{(t+1)},x_m^{(t)}) \times ...$$

$$... \times f(x_1^{(t+1)},x_2^{(t)},...,x_m^{(t)})dx_2^{(t)}...dx_m^{(t)} =$$

$$= f_m(x_m^{(t+1)}|x_1^{(t+1)},...,x_{m-1}^{(t+1)}) \times ...$$

$$\ldots \times \int f_2(x_2^{(t+1)}|x_1^{(t+1)}, x_3^{(t)}, \ldots, x_m^{(t)}) \ldots f_{m-1}(x_m^{(t+1)}|x_1^{(t+1)}, \ldots, x_{m-2}^{(t+1)}, x_m^{(t)}) \times \ldots$$

$$\ldots \times f^{(2)}(x_1^{(t+1)}, x_3^{(t)}, \ldots, x_m^{(t)}) dx_3^{(t)} \ldots dx_m^{(t)} = \ldots$$

$$\ldots = f_m(x_m^{(t+1)}|x_1^{(t+1)}, \ldots, x_{m-1}^{(t+1)}) f^{(m)}(x_1^{(t+1)}, \ldots, x_{m-1}^{(t+1)}) =$$

$$= f(x_1^{(t+1)}, \ldots, x_m^{(t+1)})$$

6 Problem 7.40

Daniel Fink

The Gibbs/EM connection. There is a connection between the EM algorithm and Gibbs sampling in that both have their basis in MC theory. One way to see this is to show that the incomplete data likelihood is a solution to the integral equation of successive substitution sampling and that Gibbs sampling can then be used to calculate the likelihood function. If $L(\theta|y)$ is the incomplete data likelihood and $L(\theta|y, z)$ is the complete data likelihood, define

$$k(z|\theta, y) = \frac{L(\theta|y, z)}{L(\theta|y)}$$

$$L^*(\theta|y) = \frac{L(\theta|y)}{\int L(\theta|y)d\theta}$$

$$L^*(\theta|y,z) = \frac{L(\theta|y,z)}{\int L(\theta|y,z)d\theta}.$$

6.1 Part A

Show that $L^*(\theta|y)$ is the solution to

$$L^*(\theta|y) = \int [\int L^*(\theta|y,z)k(z|\theta',y)dz]L^*(\theta'|y)d\theta' \tag{1}$$

Substituting for $k(z|\theta',y)$ and $L^*(\theta'|y)$ we have

$$\int [\int \frac{L(\theta|y,z)}{\int L(\theta|y,z)d\theta} \frac{L(\theta'|y,z)}{L(\theta'|y)} dz] \frac{L(\theta'|y)}{\int L(\theta'|y)d\theta'} d\theta'$$

Since all integrands are nonnegative finite we invoke Fubini's theorem and switch the order of integration,

$$\int \int \frac{L(\theta|y,z)}{\int L(\theta|y,z)d\theta} \frac{L(\theta'|y,z)}{L(\theta'|y)} \frac{L(\theta'|y)}{\int L(\theta'|y)d\theta'} d\theta' dz$$

Canceling terms we are left with,

$$\frac{1}{\int L(\theta'|y)d\theta'} \int L(\theta|y,z)dz = L^*(\theta|y).$$

6.2 Part B

Show that the sequence $\theta_{(j)}$ from the Gibbs iteration

$$\begin{aligned}\theta_{(j)} &\sim L^*(\theta|y, z_{(j-1)})\\ z_{(j-1)} &\sim k(z|\theta_{(j)}, y)\end{aligned}$$

converges to a random variable with density $L^*(\theta|y)$ as j goes to infinty. How can this be used to compute the likelihood function $L(\theta|y)$?

Let

$$\begin{aligned}g_1(\theta|z) &= L^*(\theta|y, z)\\ g_2(z|\theta) &= k(z|\theta, y)\end{aligned}$$

where y is fixed and known.

Then the two stage Hammersley-Clifford Theorem (7.1.19) can be used to give the joint distribution defined by these two conditional distributions, if such a joint distribution exists,

$$\begin{aligned}f(\theta, z|y) &= \frac{g_1(\theta|z)}{\int g_1(\theta'|z)/g_2(z|\theta')d\theta'}\\ &= \frac{L^*(\theta|y, z)}{\int L^*(\theta'|y, z)/k(z|\theta', y)d\theta'}.\end{aligned}$$

What we want is the distribution of $\theta|y$ so we integrate over z,

$$f(\theta|y) = \int \frac{L^*(\theta|y, z)}{\int L^*(\theta'|y, z)/k(z|\theta', y)d\theta'}dz$$

$$
\begin{aligned}
&= \int \frac{L(\theta|y,z)}{\int L(\theta|y,z)d\theta} \frac{1}{\int \frac{L(\theta'|y,z)}{\int L(\theta|y,z)d\theta} \frac{L(\theta'|y)}{L(\theta'|y,z)} d\theta'} dz \\
&= \frac{1}{\int L(\theta'|y)d\theta'} \int L(\theta|y,z)dz \\
&= L^*(\theta|y).
\end{aligned}
$$

The output from the Gibbs sampler can be used to make inferences about the likelihood function (unnormalized), by simply collecting the sequence $\theta_{(j)}$, into bins or other methods.