Gibbs Sampling: Assigned Problems

Prof. Dr. Narayan Prasad Adhikari
Central Department of Physics
Tribhuvan University Kirtipur, Kathmandu, Nepal

December 27, 2022

Assigned Problems

- A problem can be solved in group of at most 3 students
- Each of you must prepare and submit a written report ~ 7 pages
- It must contain background of Gibbs sampling, theory behind your problem, algorithm, codes you write to solve the problem, results and discussion.

Assigned Problems: 1

Suppose the joint distribution of $x = 0, 1, .., n$ and $0 \leq y \leq 1$ is given by

$$p(x, y) = \frac{n!}{(n-x)!\,x!} y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \tag{1}$$

note that x is discrete while y is continuous. While the joint density is complex, the conditional densities are simple distributions. To see this, first recall that a binomial random variable z has density proportional to

$$p(z \mid q, n) \propto \frac{q^z(1-q)^{n-z}}{z!\,(n-z)!} \text{ for } \ 0 \leq z \leq n \tag{2}$$

where $0 < q < 1$ is the success parameter and n the number of traits, and we denote $z \sim B(n, p)$. Likewise recall the density for $z \sim Beta(a, b)$ a beta distribution with shape parameters a and b is given by

$$p(z \mid a, b) \propto z^{a-1}(1-z)^{b-1} \text{ for } \ 0 \leq z \leq 1 \tag{3}$$

■ Assigned Problems: 1

With these probability distributions in hand, note that the conditional distribution of x (treating y as a fixed constant) is

$$x \mid y \sim B(n, y) \tag{4}$$

i.e. binomial distribution while

$$y \mid x \sim Beta(x + \alpha, n - x + \beta) \tag{5}$$

The power of the Gibbs sampler is that by computing a sequence of these univariate conditional random variables (a binomial and then a beta) we can compute any feature of either marginal distribution. Use Gibbs sampling algorithm to estimate above conditional distributions. You may take initial values as $n = 20$, $\alpha = \beta = 0.5$. Also take $y_0 = 0.5$
I will add a few more to this problem....

Assigned Problems: 2

- Integration Problem: estimate

$$V = \int_0^1 \int_0^1 \int_0^1 xyz \ln(x+2y+3z)\sin(x+y+z)dxdydz.$$

Gibbs Sampling Algorithm: first choose $p(x,y,z) = \sin(x+y+z)/C$ with

$$C = \int_0^1 \int_0^1 \int_0^1 \sin(x+y+z)dxdydz = \cos(3) - 3\cos(2) + 3\cos(1) - 1,$$

then

$$V = C\int_0^1 \int_0^1 \int_0^1 xyz \ln(x+2y+3z)p(x,y,z)dxdydz.$$

Initialize first state $\mathbf{X} = (1/2, 1/2, 1/2)$;
For $n = 1 : N$
  a) set $i = \lceil 3U \rceil$, $U \sim Uniform(0,1)$;
  b) generate $X_i \sim from$ "pdf" $sin(X + S_{-i})$;
  c) $V(n) = CX_1X_2X_3 \ln(X_1 + 2X_2 + 3X_3)$;
End.
Step b) details: using $S_{-i} = \sum_{j\neq i} X_{jn}$, set

$$U = \frac{1}{\cos(S_{-i}) - \cos(1+S_{-i})}\int_0^{X_i} \sin(S_{-i}+t)dt \;= \frac{\cos(S_{-i}) - \cos(X+S_{-i})}{\cos(S_{-i}) - \cos(1+S_{-i})},$$

so $X_i = \cos^{-1}\left(\cos(S_{-i}) - U(\cos(S_{-i}) - \cos(1+S_{-i}))\right) - S_{-i}$.

Assigned Problems: 2

Further hints:

1. You generate arrays for X
2. Set N and find C
3. Find i=3*random()
4. S=sum(X)-X(i)
5. t=cos(S)
6. X(i)=arccos(t-random()*(t-cos(1+S)))
7. I(n) =C*product(X)*log(x1x2x3)*X

You analyze your results and discuss pros/cons of Gibbs Sampling

■ Assigned Problems: 3

There is a similar analysis for $\int_0^1 \int_0^1 \int_0^1 xyze^{x+y+z} \sin(x+y+z)dxdyd$

Complete it.

■ Assigned Problems: 4- Binomial, beta and Poisson

- Treat n as unknown with $\pi(n) = Poisson(\lambda)$, ($\lambda$ known).
- $X \sim Binomial(n, \theta)$; $\pi(\theta) = Beta(a, b)$
- The joint distribution for (X, θ, n) is:

$$\binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\left(e^{-\lambda}\frac{\lambda^n}{n!}\right)$$

- $x = 0, 1, \ldots, n$; $0 < \theta < 1$; $n = 0, 1, 2, \ldots$.
- Again, interested on $f(x)$ but impossible to find in analytic-closed form.
- Alternatively (get rid of constants),

$$f(\theta, x, n) \propto \binom{n}{x}\theta^{a+x-1}(1-\theta)^{b+n-x-1}\frac{\lambda^n}{n!}$$

Assigned Problems: 4- Binomial, beta and Poisson

- For Gibbs Sampling, find full conditionals:
- $f(x|\theta, n) \propto \binom{n}{x}\theta^x(1-\theta)^{n-x} \propto Binomial(n,\theta)$,
- $\pi(\theta|x, n) \propto \theta^{a+x-1}(1-\theta)^{b+n-x-1} \propto Beta(a+x, b+n-x)$,
- $\pi(n|\theta, x) \propto \binom{n}{x}\frac{\lambda^n}{n!}(1-\theta)^{n-x} \propto \frac{[\lambda(1-\theta)]^{n-x}}{(n-x)!}; n = x, x+1, x+2, \ldots,$
- If we set, $z = n - x$, then $z \sim Poisson(\lambda(1-\theta))$
- Set $(x^{(0)}, \theta^{(0)}, n^{(0)})$. For $i = 1, 2, 3, \ldots,$
    - Sample $x^{(i)} \sim Binomial(n^{(i-1)}, \theta^{(i-1)})$.
    - Sample $\theta^{(i)} \sim Beta(a + x^{(i)}, b + n^{(i-1)} - x^{(i)})$.
    - Sample $n^{(i)} = x^{(i)} + z, z \sim Poisson(\lambda(1-\theta^{(i)}))$
    - Repeat until *convergence* is reached.

■ Assigned Problems: 4- Binomial, beta and Poisson

Plot n vs iterations, histograms of n and that of x for
$x_0 = 1, \theta^0 = 0.5$ and $n^0 = 3$

■ Assigned Problems: 5

Study the example discussed in lectures about the bivariate normal distribution

■ Assigned Problems: 6

Study the example discussed in lectures about Hierarchical model

pump	1	2	3	4	5	6	7	8	9	10
Number failures	5	1	5	14	3	19	1	1	4	22
observation time	94.32	15.72	62.88	125.76	5.24	31.44	1.05	1.05	2.10	10.48

Assigned Problems: 6

Study the example discussed in lectures about Hierarchical model

We want to compute the posterior distribution of the parameters. We are particularly interested in the mean of the distribution of the λ_i, i.e., the mean of $Gamma(\alpha, \beta)$. The mean of this gamma distribution with fixed α, β is α/β. So we need to compute the mean of α/β over the posterior distribution. We can write this as a ratio of high dimensional (ten or eleven) integrals, but that is hard to compute. So we use the Gibbs sampler to sample λ, β from the posterior. Note that this is an 11 dimensional sampler. So we need the conditional distributions of each λ_i and of β.

Assigned Problems: 7

One Dimensional Ising Model

1.1 The Algorithm

1. **Inizialization**. We start choosing an initial state, that is an initial configuration for the spins $\vec{\sigma}_0$. For example, we can make all spins up-pointing.

2. **Trial move**. We propose a move $\vec{\sigma}_0 \rightsquigarrow \vec{\sigma}_{trial}$ as follows: we randomly choose a spin and flip it.

3. **Acceptation**. If we have flipped, say, the j^{th} spin, we have $\vec{\sigma}_0 = (\sigma_1, \dots, \sigma_j, \dots, \sigma_N)$ and $\vec{\sigma}_{trial} = (\sigma_1, \dots, -\sigma_j, \dots, \sigma_N)$.

   We evaluate the number:

   $$w = \frac{\pi(\vec{\sigma}_{trial})}{\pi(\vec{\sigma}_0)} = \exp\left(-\beta\left[\mathcal{H}(\vec{\sigma}_{trial}) - \mathcal{H}(\vec{\sigma}_0))\right]\right) \tag{5}$$

   that is:

   $$w = \exp\left(-2\beta B\sigma_j - 2\beta J\sigma_j(\sigma_{j-1} + \sigma_{j+1})\right) \tag{6}$$

   Then we generate a uniform random number $r \in (0, 1)$ and:

   (a) if $r \leq w$, accept the move, defining $\vec{\sigma}_1 = \vec{\sigma}_{trial}$;

   (b) if $r > w$, reject the move, defining $\vec{\sigma}_1 = \vec{\sigma}_0$.

4. **Iteration**. Then, we use $\vec{\sigma}_1$ as the new starting point, and go back to point 2.

■ Assigned Problems: 7

One Dimensional Ising Model

You take B=0. Then carry on simulation for 100 spins. Carry on 1000 sweeps. One sweep is equal to number of Monte Carlo Steps equal to number of spins.

To initialize OR keep spin in a fixed lattice point assign positive or negative value to each lattice point randomly.

Calculate Energy (E) after each step.

Find $\Delta U = \sqrt{(}E^2 - < E >^2)$

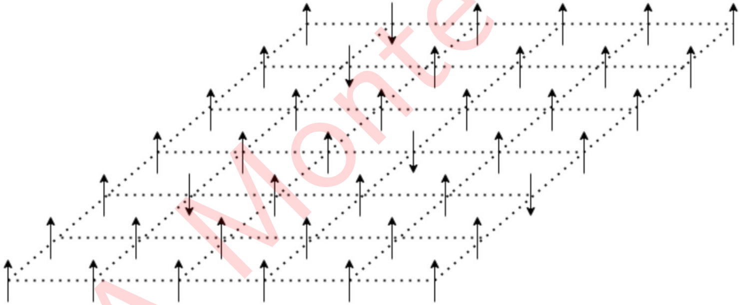
Find $C_v = \frac{\Delta U}{T^2}$

Plot $C_v \; versus \; T$

You do not need to understand all the details but just whatever I mentioned above. I will explain results.

Assigned Problems: 8

Ising model in 2 Dimensions



Assigned Problems: 8

Carry on Monte Carlo Simulation of 2D Ising model according to Metropolis-Hastings Algorithm.

1. Generate a random initial spin configuration $\left(\sigma_{i,j=1}^{N}\right)$ and set temperature T. You can assign values $+$ or $-$ to each point i, j in 2 dimensional lattice
2. Choose a spin $(\sigma_{i,j})$ randomly
3. Calculate energy by taking only four neighbors of the chosen spin $(\sigma_{i,j})$ i.e. $E_{old} = \sigma_{i,j}(\sigma_{(i-1,j)} + \sigma_{(i+1,j)} + \sigma_{(i,j-1)} + \sigma_{(i,j+1)}$
4. Flip the chosen spin i.e. write $\sigma_{i,j} = -\sigma_{i,j}$
5. Now calculate energy of the new configuration of the spins i.e. $E_{new} = -\sigma_{i,j}(\sigma_{(i-1,j)} + \sigma_{(i+1,j)} + \sigma_{(i,j-1)} + \sigma_{(i,j+1)}$
6. if $E_{new} < E_{old}$ accept the flip and go to step 2 for another MC steps
7. Find $en = \frac{E_{new} - E_{old}}{T}$ and geenrate a random number $r \in (0, 1)$
8. Accept the flip if $r > en$ else do not flip the spin
9. Go to step 2

Assigned Problems: 8

Carry on simulation for 10X10 spins in 2 Dimensions. Carry on 1000 sweeps. One sweep is equal to number of Monte Carlo Steps equal to number of spins.
To initialize OR keep spins in fixed lattice point assign positive or negative value to each lattice point.
Calculate Energy (E) after each step.
Find $\Delta U = \sqrt{(E^2 - < E >^2)}$
Find $C_v = \frac{\Delta U}{T^2}$
Plot $C_v \ versus \ T$
You do not need to understand all the details but just whatever I mentioned above. I will explain results.

■ Assigned Problems: 7 & 8 (Optional)

You can also find average magnetization if you want.

$$M = \sum_{i,j=1}^{N} \sigma_{i,j} \tag{6}$$

Also find fluctuations in M $i.e.$ $\chi = \langle M^2 - \langle M \rangle^2 \rangle$
You can also plot χ as a function of T.

Assigned Problems: 9& 10

These two problems are from

Understanding the Metropolis–Hastings Algorithm

Siddhartha CHIB and Edward GREENBERG

Read the accompanying paper before you try to solve these problems.

■ Assigned Problems: 9

7.1 Simulating a Bivariate Normal

To illustrate the M–H algorithm we consider the simulation of the bivariate normal distribution $\mathcal{N}_2(\mu, \Sigma)$, where $\mu = (1, 2)'$ is the mean vector and $\Sigma = (\sigma_{ij})$: 2×2 is the covariance matrix given by

$$\Sigma = \begin{pmatrix} 1 & .9 \\ .9 & 1 \end{pmatrix}.$$

Because of the high correlation the contours of this distribution are "cigar-shaped," that is, thin and positively inclined. Although this distribution can be simulated directly in the Choleski approach by letting $y = \mu + P'u$, where $u \sim \mathcal{N}_2(0, I_2)$ and P satisfies $P'P = \Sigma$, this well-known problem is useful for illustrating the M–H algorithm.

From the expression for the multivariate normal density, the probability of move (for a symmetric candidate-generating density) is

$$\alpha(x, y) = \min\left\{ \frac{\exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right]}{\exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]}, 1 \right\},$$

$$x, y \in \Re^2. \quad (9)$$

■ Assigned Problems: 9

We use the following candidate-generating densities, for which the parameters are adjusted by experimentation to achieve an acceptance rate of 40% to 50%:

1. Random walk generating density ($y = x + z$), where the increment random variable z is distributed as bivariate uniform, that is, the ith component of z is uniform on the interval $(-\delta_i, \delta_i)$. Note that δ_1 controls the spread along the first coordinate axis and δ_2 the spread along the second. To avoid excessive moves we let $\delta_1 = .75$ and $\delta_2 = 1$.

2. Random walk generating density ($y = x + z$) with z distributed as independent normal $N_2(0, D)$, where $D =$ diagonal$(.6, .4)$.

3. Pseudorejection sampling generating density with "dominating function" $ch(x) = c(2\pi)^{-1}|D|^{-1/2} \exp[-\frac{1}{2}(x-\mu)'D(x-\mu)]$, where $D =$ diagonal$(2, 2)$ and $c = .9$. The trial draws, which are passed through the A–R step, are thus obtained from a bivariate, independent normal distribution.

Assigned Problems: 9

4. The autoregressive generating density $y = \mu - (x - \mu) + z$, where z is independent uniform with $\delta_1 = 1 = \delta_2$. Thus values of y are obtained by reflecting the current point around μ and then adding the increment.

Assigned Problems: 10

7.2 Simulating a Bayesian Posterior

We now illustrate the use of the M–H algorithm to sample an intractable distribution that arises in a stationary second-order autoregressive [AR(2)] time series model. Our presentation is based on Chib and Greenberg (1994), which contains a more detailed discussion and results for the general ARMA(p, q) model.

For our illustration, we simulated 100 observations from the model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \qquad t = 1, 2, \ldots, 100, \quad (10)$$

where $\phi_1 = 1$, $\phi_2 = -.5$, and $\epsilon_t \sim N(0, 1)$. The values of $\phi = (\phi_1, \phi_2)$ lie in the region $S \subset \mathcal{R}^2$ that satisfies the stationarity restrictions

$$\phi_1 + \phi_2 < 1; \qquad -\phi_1 + \phi_2 < 1; \qquad \phi_2 > -1.$$

Following Box and Jenkins (1976), we express the (*exact* or *unconditional*) likelihood function for this model given the $n = 100$ data values $Y_n = (y_1, y_2, \ldots, y_n)'$ as

■ Assigned Problems: 10

For our illustration, we simulated 100 observations from the model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \qquad t = 1, 2, \ldots, 100, \quad (10)$$

where $\phi_1 = 1$, $\phi_2 = -.5$, and $\epsilon_t \sim N(0, 1)$. The values of $\phi = (\phi_1, \phi_2)$ lie in the region $S \subset \mathcal{R}^2$ that satisfies the stationarity restrictions

$$\phi_1 + \phi_2 < 1; \qquad -\phi_1 + \phi_2 < 1; \qquad \phi_2 > -1.$$

Following Box and Jenkins (1976), we express the (*exact* or *unconditional*) likelihood function for this model given the $n = 100$ data values $Y_n = (y_1, y_2, \ldots, y_n)'$ as

$$l(\phi, \sigma^2) = \Psi(\phi, \sigma^2) \times (\sigma^2)^{-(n-2)/2} \times \exp\left[-\frac{1}{2\sigma^2} \sum_{t=3}^{n} (y_t - w_t'\phi)^2\right], \quad (11)$$

where $w_t = (y_{t-1}, y_{t-2})'$,

$$\Psi(\phi, \sigma^2) = (\sigma^2)^{-1} |V^{-1}|^{1/2} \exp\left[-\frac{1}{2\sigma^2} Y_2' V^{-1} Y_2\right] \quad (12)$$

is the density of $Y_2 = (y_1, y_2)'$,

$$V^{-1} = \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{pmatrix},$$

■ Assigned Problems: 10

and the third term in (11) is proportional to the density of the observations $(y_3, \ldots, y_n)$ given Y_2

If the only prior information available is that the process is stationary, then the posterior distribution of the parameters is

$$\pi(\phi, \sigma^2 \mid Y_n) \propto l(\phi, \sigma^2) I[\phi \in S],$$

where $I[\phi \in S]$ is 1 if $\phi \in S$ and 0 otherwise.

How can this posterior density be simulated? The answer lies in recognizing two facts. First, the blocking strategy is useful for this problem by taking ϕ and σ^2 as blocks. Second, from the regression ANOVA decomposition, the exponential term of (11) is proportional to

$$\exp\left[-\frac{1}{2\sigma^2}(\phi - \hat{\phi})' G(\phi - \hat{\phi})\right],$$