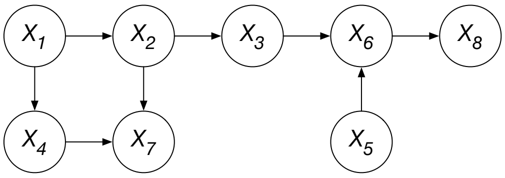
CS5340: Uncertainty Modeling in AI

Tutorial 8

Released: Mar. 20, 2024

1 Gibbs Sampling

You want to run Gibbs sampling on the following graphical model. For each of the random variables below, what is the correct conditional to sample from? **Note:** If there are multiple correct answers, select the one that conditions upon the fewest number of random variables.



Problem 1. Sample x_1.

A. $p(X_1)$ (sample from the prior)

B. $p(X_1|X_2, X_4, X_7)$

C. $p(X_1|X_2, X_4)$

D. $p(X_1|X_2, X_3, X_4, X_7)$

E. $p(X_1|X_2, X_3, X_4, X_7, X_5)$

Problem 2. Sample x_2.

A. $p(X_2|X_1, X_3)$

B. $p(X_2|X_4, X_7)$

C. $p(X_2|X_1, X_3, X_4, X_7)$

D. $p(X_2|X_3, X_7)$

E. $p(X_2|X_1)$

Problem 3. Sample x_3.

A. $p(X_3|X_2, X_6, X_7)$

B. $p(X_3|X_2, X_6)$

C. $p(X_3|X_2)$

D. $p(X_3|X_2, X_5, X_6)$

E. $p(X_3|X_2, X_5, X_6, X_8)$

Problem 4. Sample x_4.

A. $p(X_4|X_1, X_2, X_7)$

B. $p(X_4|X_2, X_7)$

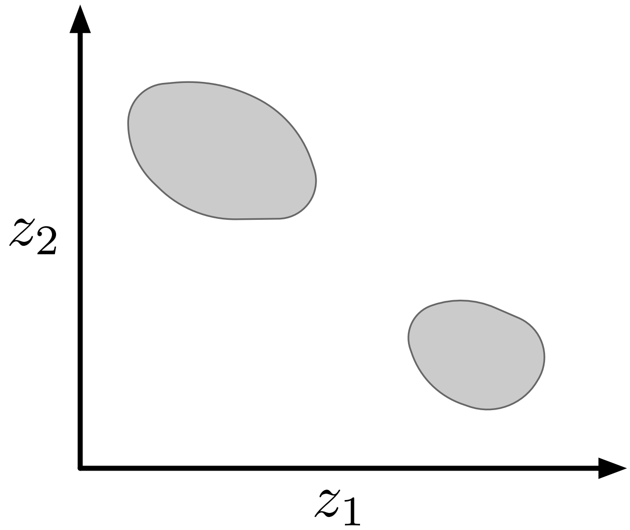
C. $p(X_4|X_1)$

D. $p(X_4|X_1, X_2, X_3, X_7)$

E. $p(X_4|X_1, X_2, X_3, X_7, X_6, X_8)$

2 Discussion Questions

Problem 5. Consider the following distribution of two variables z_1 and z_2 that is uniform over the shaded regions and that is zero everywhere else. **Does the standard Gibbs sampling procedure sample correctly from this distribution**?



Problem 6. Consider the Metropolis-Hastings algorithm. At each step i, we sample a new point from the proposal distribution $q(\mathbf{x}|\mathbf{x}^{(i)})$. In the lectures, we used as an example the simple *isotropic* (spherical) Gaussian centered upon $\mathbf{x}^{(i)}$, i.e.,

$$q(\mathbf{x}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{x}^{(i)}, \sigma^2 \mathbf{I})$$

which is a common choice for continuous variables. The variance σ^2 is a parameter of the proposal distribution. **How sensitive is MH sampling to the parameter σ^2.** What are the respective trade-offs when considering how to set σ^2? *Hint:* Consider a elongated bi-variate Gaussian having strong correlations between its components.

3 The Right Transitions

For each of the matrices below, select True if the matrix is valid *transition matrix* over the states for use in a MCMC algorithm. Select False otherwise. Justify your answer. *Hint:* Look up what properties are required for a transition matrix in an MCMC method.

Problem 7.

$$T = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.8 \\ 0.3 & 0.2 & 0.1 \end{bmatrix}$$

Problem 8.

$$T = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.0 & 0.7 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Problem 9.

$$T = \begin{bmatrix} 0.9 & 0.1 & 0.0 \\ 0.1 & 0.9 & 0.0 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$$

Problem 10.

$$T = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.0 & 0.6 & 0.4 \end{bmatrix}$$

Problem 11.

$$T = \begin{bmatrix} 0.0 & 0.8 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

Problem 12.

$$T = \begin{bmatrix} 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & 0.0 \end{bmatrix}$$

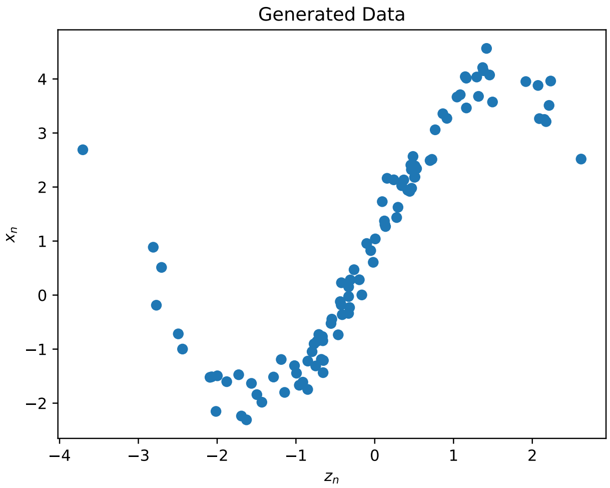
4 Monte-Carlo EM

In this tutorial, we will apply sampling towards learning parameters. We will adapt the EM algorithm so that the E-step is possible even when the posterior does not have a nice analytical form. At a high-level, we will apply Monte-Carlo methods to estimate the expectation of the log joint under the posterior, $\mathbb{E}_{p(z|x,\theta^{old})}[\log p(x, z)]$.

Our model will be similar to the PPCA model covered in a previous tutorial, except that now, we will have a *nonlinear* relationship between z and x. Consider that we have a dataset $\mathcal{D} = \{x_1, \ldots, x_N\}$. Our assumed model is:

$$z_n \sim N(\mu, s^2)$$
$$x_n|z_n \sim N(f(z_n), \sigma^2)$$
$$f(z_n) = \alpha \sin(z_n) + \beta$$

To keep things simple, x_n and z_n are one-dimensional and assume that s^2 and σ^2 are known. The unknown parameters of our model are $\theta = \{\mu, \alpha, \beta\}$. We want to learn θ. There is a code notebook Monte-Carlo_EM.ipynb that accompanies this tutorial (available on Canvas). The figure below shows 100 data points x's generated with parameters ($\mu = 0.0, \alpha = 3, \beta = 1, s^2 = 2, \sigma^2 = 0.1$):



Problem 13. Draw out the DGM for the model described above.

Problem 14. Our dataset only comprises the x's and the z's are latent. We can apply the E-M algorithm to learn the parameters θ. Show that the Q function is given by:

$$Q(\theta, \theta^{old}) = \frac{1}{2}\mathbb{E}_{p(z|x,\theta^{old})}\left[\sum_{n=1}^{N} -\log(2\pi\sigma^2) - \frac{(x_n - \alpha\sin(z_n) - \beta)^2}{\sigma^2} - \log(2\pi s^2) - \frac{(z_n - \mu)^2}{s^2}\right] \quad (1)$$

Problem 15. Given the Q function in the previous problem, argue that to find the new parameters, we only need to minimize two expectations:

A. $\arg\min_{\alpha,\beta} \sum_{n=1}^{N} \mathbb{E}_{p(z_n|x_n,\theta^{old})}[(x_n - \alpha\sin(z_n) - \beta)^2]$

B. $\arg\min_{\mu} \sum_{n=1}^{N} \mathbb{E}_{p(z_n|x_n,\theta^{old})}[(z_n - \mu)^2]$

Problem 16. Let us first approximate the expectations A and B above using self-normalized importance sampling. For example, we will approximate

$$\mathbb{E}_{p(z|x,\theta^{old})}[(x_n - f(z_n))^2] \approx \sum_{i=1}^{M} \tilde{w}_i (x_n - f(z_n^{(i)}))^2$$

where

$$\tilde{w}_i = \frac{\tilde{p}(z_n^{(i)})/\tilde{q}(z_n^{(i)})}{\sum_{j=1}^{M} \tilde{p}(z_n^{(j)})/\tilde{q}(z_n^{(j)})}$$

Assume a Gaussian proposal distribution $q = \mathcal{N}(z_n^{(i)}|\mu^{old}, v^2)$. To compute the importance weight above, work out what should be the:

1. The unnormalized target density $\tilde{p}(z_n^{(i)})$
2. The unnormalized proposal density $\tilde{q}(z_n^{(i)})$

Problem 17. Once the samples and weights are obtained, can the solution to the optimization problem

$$\arg\min_{\alpha,\beta} \sum_{n=1}^{N} \sum_{i=1}^{M} \tilde{w}_i (x_n - \alpha \sin(z_n^{(i)}) - \beta)^2$$

be computed analytically? Likewise, can be solution to

$$\arg\min_{\mu} \sum_{n=1}^{N} \sum_{i=1}^{M} \tilde{w}_i (z_n^{(i)} - \mu)^2$$

be computed in closed form? Also, discuss whether it a good idea to solve these optimization problems to their global maximum.

Problem 18. Next, let us apply MCMC towards computing the above expectations. Recall that the acceptance probability is given by:

$$\min\left\{1, \frac{\tilde{p}(z')q(z|z')}{\tilde{p}(z)q(z'|z)}\right\}$$

Provide an argument that if we use a Gaussian proposal $q(z'|z) = N(z'|z, v^2)$, then $q(z|z') = q(z'|z)$.

Problem 19. Check the accompanying tutorial code Monte-Carlo_EM.ipynb. Explore the code and discuss the effects of the following parameters:

1. The number of iterations.
2. The number of samples.
3. The variance of the proposal distribution.
4. Extra: Using a different proposal such as Metropolis-Adjusted Langevin Algorithm (MALA).