

Lab 1 Notes

January 17, 2023

*Instructor: Abel Rodriguez**TA: Aparav Venkat*

*These notes are meant to accompany the implementations discussed during the lab. They contain derivations for mathematical quantities estimated in the lab and other miscellaneous information. They have not been proofread and may contain typos. Please email aparav@uw.edu if you catch errors.*¹

Converting a 2D lattice to an adjacency matrix

In the Ising model, we work with a regular 2D lattice. An intuitive data structure to store the spins of each node is a 2D matrix. For example we may store observations on a 2D lattice, \mathbf{X} , in a matrix M with the element $M_{i,j}$ denoting the value taken by the node at lattice position (i, j) .

While this matrix structure is convenient for certain scenarios, some calculations require us to use an explicit adjacency matrix. This conversion is not difficult but it can be tedious, often leading us to miss edge cases. Here, we will describe one way to perform this transformation.

The first task is to assign a unique ID to each node. Suppose we have a $m \times n$ lattice. Then, we can assign a node at position (i, j) the unique ID $n(i - 1) + j$. This is equivalent to “unrolling” the matrix M into a vector. The following R code demonstrates how to convert between the lattice position and the unique vector position. Notice that when converting from the vector position back to the lattice position, we need to handle the edge case corresponding to the right boundary carefully.

```
lat_to_vec <- function(i, j, m, n) {
  k <- n*(i-1) + j
  return(k)
}

vec_to_lat <- function(k, m, n) {
  i <- floor((k-1) / n) + 1
  j <- (k %% n)
  if (j == 0) {
    j <- n
  }
  return(c(i, j))
}
```

¹Thanks to Jess Kunke for catching errors in the notes.

Using this, it is easy to write code that converts a 2D lattice to an adjacency matrix. These helper functions are available on Canvas.

Solutions to Lab 1

1. The dataset is available at <https://sites.stat.washington.edu/peter/stoch.mod.data.html>. Please download `Data.Set.5`.

- (a) The implementation should be straightforward at first blush. However, there are two issues. First, the textbook has some typos. This makes it difficult to find out whether your implementation is correct. The error seems to come from how the author parameterizes the interaction effects. Throughout the literature, you will see two different parameterizations,

$$p(x) \propto \exp \left\{ \beta \sum_{i,j} x_i x_j \mathbb{I}(i, j \text{ are neighbors}) \right\},$$

$$p(x) \propto \exp \left\{ \frac{1}{2} \beta \sum_{i,j} x_i x_j \mathbb{I}(i, j \text{ are neighbors}) \right\}.$$

The difference between the two (which, as in this case, is often not made obvious) is that the top parameterization *double counts* interaction terms and the bottom one doesn't. The textbook starts out by double counting interactions. But Figure 4.6 does not double count interactions. This is very confusing. Hopefully the implementation provided in this lab makes the distinctions clear.

Second, the textbook suggests a Metropolis-Hastings sampler for sampling from the auxiliary distribution for exponential tilting. This can be somewhat slow in terms of computation. An alternate is the Swendson-Wang (SW) sampler which you will implement in HW 1. For this lab, we will use the SW sampler. The implementation will become available on Canvas after HW 1 is due.

Now, on to calculating the asymptotic covariance matrix.

$$\frac{\partial^2 \lambda}{\partial \theta_i^2} = - \frac{\frac{\partial^2 Z(\theta)}{\partial \theta_i^2} \cdot \frac{1}{Z(\phi)} \cdot \frac{Z(\theta)}{Z(\phi)} - \left(\frac{\partial Z(\theta)}{\partial \theta_i} \cdot \frac{1}{Z(\phi)} \right)^2}{\left(\frac{Z(\theta)}{Z(\phi)} \right)^2}$$

$$\frac{\partial^2 \lambda}{\partial \theta_1 \partial \theta_2} = - \frac{\frac{\partial^2 Z(\theta)}{\partial \theta_1 \partial \theta_2} \cdot \frac{1}{Z(\phi)} \cdot \frac{Z(\theta)}{Z(\phi)} - \left(\frac{\partial Z(\theta)}{\partial \theta_1} \cdot \frac{1}{Z(\phi)} \right) \left(\frac{\partial Z(\theta)}{\partial \theta_2} \cdot \frac{1}{Z(\phi)} \right)}{\left(\frac{Z(\theta)}{Z(\phi)} \right)^2}$$

where

$$\begin{aligned} \frac{\partial^2 Z(\theta)}{\partial \theta_2} \cdot \frac{1}{Z(\phi)} &= \mathbb{E}_\phi [T_i^2 \exp\{T_1(\theta_1 - \phi_1) + T_2(\theta_2 - \phi_2)\}] \\ &\approx \frac{1}{n} \sum_{j=1}^n (t_i^{(j)})^2 \exp\{t_1^{(j)}(\theta_1 - \phi_1) + t_2^{(j)}(\theta_2 - \phi_2)\} \end{aligned}$$

$$\begin{aligned}\frac{\partial^2 Z(\theta)}{\partial \theta_1 \partial \theta_2} \cdot \frac{1}{Z(\phi)} &= \mathbb{E}_\phi [T_1 T_2 \exp\{T_1(\theta_1 - \phi_1) + T_2(\theta_2 - \phi_2)\}] \\ &\approx \frac{1}{n} \sum_{j=1}^n t_1^{(j)} t_2^{(j)} \exp\{t_1^{(j)}(\theta_1 - \phi_1) + t_2^{(j)}(\theta_2 - \phi_2)\}\end{aligned}$$

This lets us compute the Fisher Information matrix,

$$\mathcal{I} = - \begin{bmatrix} \frac{\partial^2 \lambda}{\partial \theta_1^2} & \frac{\partial^2 \lambda}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \lambda}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \lambda}{\partial \theta_2^2} \end{bmatrix}$$

And the asymptotic covariance matrix is \mathcal{I}^{-1} .

- (b) Following Ghosal and Mukherjee (2020),² the score equations for the maximum pseudo-likelihood estimator (MPLE) are

$$\begin{aligned}Q_n(\beta, B) &= \sum_{i=1}^n m_i(x)(x_i - \tanh(\beta m_i(x) + B)) \\ R_n(\beta, B) &= \sum_{i=1}^n (x_i - \tanh(\beta m_i(x) + B))\end{aligned}$$

where $m_i(x) = \sum_{j=1}^n A_{i,j} x_j$ and A is the adjacency matrix of the 2D lattice. The parameter β corresponds to θ_2 and B corresponds to θ_1 in part (a).

To calculate the asymptotic covariance matrix,

$$\begin{aligned}\frac{\partial Q_n}{\partial \beta} &= - \sum_{i=1}^n \frac{m_i(x)^2}{\cosh^2(\beta m_i(x) + B)} \\ \frac{\partial Q_n}{\partial B} &= - \sum_{i=1}^n \frac{m_i(x)}{\cosh^2(\beta m_i(x) + B)} \\ \frac{\partial R_n}{\partial B} &= - \sum_{i=1}^n \frac{1}{\cosh^2(\beta m_i(x) + B)}\end{aligned}$$

This gives us the Fisher information matrix

$$\mathcal{I} = \mathbb{E} \begin{bmatrix} \frac{\partial Q_n}{\partial \beta} & \frac{\partial Q_n}{\partial B} \\ \frac{\partial R_n}{\partial \beta} & \frac{\partial R_n}{\partial B} \end{bmatrix}$$

Note that in part (a), we did not run into any issues when finding the information matrix as the $[D^2 \log Z(\theta)/Z(\phi)]$ had no dependence on observations X_i . Here, that is not that case. There is no easy way to analytically evaluate this expectation. So we will approximate the Fisher information matrix using random samples (see Seymour (2001)³). Once again, we use the SW sampler to draw observations to estimate $\hat{\mathcal{I}}$. Therefore the asymptotic covariance matrix is $\hat{\mathcal{I}}^{-1}$.

²Ghosal, Promit, and Sumit Mukherjee. “Joint estimation of parameters in Ising model.” *The Annals of Statistics* 48.2 (2020): 785-810. <https://doi.org/10.1214/19-AOS1822>

³Seymour, Lynne. “Estimating the Variance of the Maximum Pseudo-Likelihood Estimator.” *Lecture Notes-Monograph Series* (2001): 281-295. <https://www.jstor.org/stable/4356156>.

To evaluate the finite sample covariance matrix, we need to correct for the convergence rate as seen in Theorem 1.2 of Ghosal and Mukherjee (2020). This amounts to calculating

$$T_n(X) = \frac{1}{n} \sum_{i=1}^n (m_i(x) - \tilde{m}(x))^2$$

$$\tilde{m}(x) = \frac{1}{n} \sum_{i=1}^n m_i(x)$$

Finally, note that the above calculation assumes that $B \neq 0$. If $B = 0$, then see Chatterjee (2007)⁴ for consistency results.

2. (a) (θ_i, σ_i^2) corresponds to the mean and variance of the luminosity of segment i .
- (b) ϕ has the usual interpretation in Ising model. If $\phi > 0$, neighbors are more likely to belong to the same segment. If $\phi = 0$, neighbors are independent. And if $\phi < 0$, then neighbors are more likely to belong to different segments. For the image segmentation problem, it is typical to assume that $\phi > 0$.
- (c) Here, we will use a Gibbs sampler. The overall scheme looks like
 - (1) Initialize $\theta^{(0)}, \sigma^{2,(0)}, Z^{(0)}, \phi^{(0)}$
 - (2) For $t = 1$ to T
 - (i) For $i = 1$ to N
 - Sample $Z_i^{(t)} \sim P(Z_i | Z_1^{(t)}, \dots, Z_{i-1}^{(t)}, Z_{i+1}^{(t-1)}, \dots, Z_N^{(t-1)}, Y, \theta^{(t-1)}, \sigma^{2,(t-1)}, \phi^{(t-1)})$
 - (ii) For $k = 1$ to K
 - Sample $\theta_k^{(t)} \sim P(\theta_k | \theta_1^{(t)}, \dots, \theta_{k-1}^{(t)}, \theta_{k+1}^{(t-1)}, \dots, \theta_K^{(t-1)}, Y, Z^{(t)}, \sigma^{2,(t-1)}, \phi^{(t-1)})$
 - Sample $\sigma_k^{2,(t)} \sim P(\sigma_k^2 | \sigma_1^{2,(t)}, \dots, \sigma_{k-1}^{2,(t)}, \sigma_{k+1}^{2,(t-1)}, \dots, \sigma_K^{2,(t-1)}, Y, Z^{(t)}, \theta^{(t)}, \phi^{(t-1)})$
 - (iii) Sample $\phi^{(t)} \sim P(\phi | Z^{(t)}, Y, \theta^{(t)}, \sigma^{2,(t)})$.

The conditional distributions for Z, θ, σ^2 are derived below. You will derive the conditional distribution for ϕ and implement its sampler in HW 1.

$$\begin{aligned}
 P(Z_i | Z_{-i}, \theta, \sigma^2, Y, \phi) &\propto P(Z_i, Z_{-i}, Y, \theta, \sigma^2, \phi) \\
 &\propto P(Y_i | Z, \theta, \sigma^2, \phi) P(Z_i | Z_{-i}, \theta, \sigma^2, \phi) \\
 &= P(Y_i | \theta_{Z_i}, \sigma_{Z_i}^2) P(Z_i | Z_{-i}, \phi) \\
 \implies P(Z_i = k | Z_{-i}, \theta, \sigma^2, Y, \phi) &= \frac{P(Y_i | \theta_k, \sigma_k^2) P(Z_i = k | Z_{-i}, \phi)}{\sum_{j=1}^K P(Y_i | \theta_j, \sigma_j^2) P(Z_i = j | Z_{-i}, \phi)} \\
 P(Z_i = k | Z_{-i}, \phi) &= \frac{\exp\{\phi N_{\delta(i)}(k)\}}{\sum_{j=1}^K \exp\{\phi N_{\delta(i)}(j)\}}
 \end{aligned}$$

where $N_{\delta(i)}(j)$ is the number of neighbors of node i that have the color j .

$$\begin{aligned}
 P(\theta_j | Z, \theta_{-j}, \sigma^2, Y, \phi) &\propto P(Y, Z, \theta_j, \theta_{-j}, \sigma^2, \phi) \\
 &\propto P(Y_{Z(j)} | \theta_j, \sigma_j^2) P(\theta_j)
 \end{aligned}$$

⁴Chatterjee, Sourav. "Estimation in spin glasses: A first step." The Annals of Statistics 35.5 (2007): 1931-1946.
<https://doi.org/10.1214/009053607000000109>

$$\begin{aligned}
P(\sigma_j^2 \mid Z, \sigma_{-j}^2, \theta, Y, \phi) &\propto P(Y, Z, \sigma_j^2, \sigma_{-j}^2, \theta, \phi) \\
&\propto P(Y_{Z(j)} \mid \theta_j, \sigma_j^2) P(\sigma_j^2)
\end{aligned}$$

where $Y_{Z(j)} = \{Y_i : Z_i = j\}$.

We will use conjugate priors for (θ, σ^2) i.e., $\theta_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \lambda)$, $\lambda > 0$ and $\sigma_j^2 \stackrel{\text{i.i.d.}}{\sim} \text{Inv-Gamma}(\alpha, \beta)$, $\alpha, \beta > 0$.