Danny Liu
Math143 - Homework 2

Problem 1: Read and write up the Matrix Normal Distribution in your own words.
Just like how the multivariate normal distribution generalizes the normal distribution from one random variable to a vector of random variables, the matrix normal distribution generalizes the multivariate normal distribution from a random vector to a random matrix. Generally, if $\mathbf{X} \sim \mathcal{MN}_{n,p}(\mathbf{M}, \mathbf{U}, \mathbf{V})$, then

$$p(\mathbf{X}|\mathbf{M}, \mathbf{U}, \mathbf{V}) = \frac{\exp(-\frac{1}{2}\text{tr}(\mathbf{V}^{-1}(\mathbf{X}-\mathbf{m})^T\mathbf{U}^{-1}(\mathbf{X}-\mathbf{M})))}{(2\pi)^{np/2}|\mathbf{V}|^{n/2}|\mathbf{U}|^{p/2}}$$

where $\mathbf{X}$ is an $n \times p$ random matrix, $\mathbf{M}$ is the $n \times p$ mean matrix, $\mathbf{U}$ is $n \times n$, $\mathbf{V}$ is $p \times p$. In this case, $\mathbf{U}$ and $\mathbf{V}$ are positive semi-definite matrices that determine what we intuitively think of as the variance in the scalar case and the covariance matrix in the vector case.
We also have $E[\mathbf{X}] = \mathbf{V}$, $E[(\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})^T] = \mathbf{U}\text{tr}(\mathbf{V})$, $E[(\mathbf{X}-\mathbf{M})^T(\mathbf{X}-\mathbf{M})] = \mathbf{V}\text{tr}(\mathbf{U})$. In addition, the maximum likelihood estimate for $\hat{\mathbf{M}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{X}_i$. The covariance parameters do not have a closed form solution but can be solved by iteratively maximizing their gradients according to the formula $\hat{\mathbf{U}} = \frac{1}{kp}\sum_{i=1}^{k}(\mathbf{X}_i - \mathbf{M})\hat{\mathbf{V}}^{-1}(\mathbf{X}_i - \mathbf{M})^T$ and

$\hat{\mathbf{V}} = \frac{1}{kn}\sum_{i=1}^{k}(\mathbf{X}_i - \mathbf{M})^T\hat{\mathbf{U}}^{-1}(\mathbf{X}_i - \mathbf{M})$

Problem 2: Read and write up the Dirichlet distribution in your own words and find an example using Dirichlet (e.g. LDA).
The Dirichlet distribution is the multivariate generalization of the Beta distribution. Hence it is the conjugate prior to the categorial/multinomial distribution.
In general if $\mathbf{X} \sim \text{Dir}(\boldsymbol{\alpha})$, then

$$\frac{1}{B(\boldsymbol{\alpha})}\prod_{i=1}^{K}x_i^{\alpha_i - 1}$$

where $B(\boldsymbol{\alpha})$ is the normalizing constant, $x_i \in [0, 1]$, $\sum_i x_i = 1$, $\alpha_i > 0$ and $E[X_i] = \frac{\alpha_i}{\sum_k \alpha_k}$. We note that $X_k = 1 - \sum_{i=1}^{k-1} X_i$ and that if interpreted as a conjugate prior, α_i is the number of times event X_i has occurred. This clearly extends the Beta distribution where α is interpreted as the number of times a given event occurred and β is when it didn't.

One common use of the Dirichlet distribution is in the Latent Dirichlet Allocation (LDA). LDA is a generative statistical model that allow sets of observations to be explained by unobserved groups. The classic use case is in topic modeling where the observations are documents containing words and the unobserved groups are a set of topics associated with each document. The number of topics k must be specified beforehand and the generative process in which M documents each of length N_i are represented as random mixtures over latent topics assumes two Dirichlet distributions θ_i, the topic distribution for document i, and ϕ_k,

the word distribution for topic k. Then for each word position i, j where $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N_i\}$, a topic $z_{i,j} \sim \text{Multinomial}(\theta_i)$ and word $w_{i,j} \sim \text{Multinomial}(\phi_{z_{i,j}})$ are chosen. Inference is then often done using either Monte Carlo simulations, Gibbs Sampling, Variational Bayes, or MLE.

Problem 3: find an example of a manifold which Prof. Gu hasn't mentioned.
The set of all possible stock portfolios in the S&P 500 form a manifold.