269

Homework #1

Install R, read appendix A of BSM, and install bayesm.

1. A list structure (review pp. 288-293 in BSM) is a very good way to store panel data. Each cross-sectional unit can be a different element in the list. Panel data can be simply a list of lists. This allows for a different number of observations per panel unit as well as even different variables. To develop some familiarity with this structure, this problem will ask you to read in some panel data and perform regressions on each unit, returning an array of regression coefficients.

   Using the help example (`?cheese`) from the cheese dataset in bayesm, read in the cheese data and create a list of lists from the data frame created by reading in `cheese.dat`. Each element in the list should be a list of the form `(y=y, X=X)`. Don't forget to create an intercept. Include both price and display in X. To access the cheese data set, use the R command – `data(cheese)`.

   Write a function to compute regression of y on X taking a list as input.

   Apply this function to the list data created from cheese data and return an m x 3 array of regression coefficients, one for each of the m units.

   X may not be of full column rank for all cross-sectional units. What can you do about this? You can check the rank of a matrix by doing a QR decomposition and inspecting the element `rank` (see `?qr` for details). Note: you can also use the QR decomposition to run the regression, but more about this next week!

2. Show that the determinant of a triangular matrix is the product of the diagonal elements. Hint: show for 2 x 2 and then use induction (e.g. show for n x n assuming (n-1) x (n-1) triangular matrices have determinants that are products of their diagonals).

3. The Cholesky root is the most useful matrix decomposition for Bayesian analysis. It is also one of the most numerically stable decompositions (this means that it can deal with near singular matrices and also matrices with very different scale across the elements without "pivoting"). To obtain a greater intuition for the Cholesky root, consider the following problems:

   a. $z \sim N(0, I)$, let $y = Lz$ where L is lower triangular.
   Compute Var(y).

   b. consider a 3 x 3 case.

   Use the equation $\Sigma = LL'$ to solve for the elements of L in terms of the elements of Σ. Check your computations by using comparison with `chol()` in R for a particular value of Σ.

c. How do we know that the Cholesky root exists for any given pds matrix? One way of establishing this is to construct the Cholesky root from first principles regarding joint and conditional distributions.

Given a correlated vector y, we can orthogonalize y as follows:

$$
\begin{aligned}
&y_1 = \sigma_1 z_1 \\
&-\beta_{21} y_1 + y_2 = \sigma_2 z_2 \\
&-\beta_{31} y_1 - \beta_{32} y_2 + y_3 = \sigma_3 z_3 \\
&\vdots \\
&-\beta_{p,1} y_1 - \ldots - \beta_{p,p-1} y_{p-1} + y_p = \sigma_p z_p
\end{aligned}
$$

where Var(z) = I.

How can you be sure that you can do this? Hint: the parameters are chosen with suggestive symbols!

Given the existence of the system above, how does this establish that there exists a Cholesky root of Σ? Hint: write the system above using matrix notation and construct the Cholesky root directly from these coefficients. You will have to prove that inverses of triangular matrices are also triangular!

Show your R code in the hw not just the runs! BE NEAT. If I can't read it, I won't grade it!