Stat 215b (Spring 2004): Lab 6

B. M. Bolstad
bolstad@stat.berkeley.edu

Due Apr 29, 2004

This lab consists of two parts. In both parts the focus will be on methods of fitting smooth models. You should write up your lab such that the analysis is divided into two different sections. However, you need not divide your introduction, methods, conclusions etc. That is you don't have to write two completely separate reports. Please try to have some fun with this one.

1 Part I - Seismic wave attenuation

Following an earthquake, seismic waves radiate out from the source. The amplitude of their oscillations falls off as the hypocentral distance, r, from the source increases. Strong motion seismometers are scattered about and record, for example, the maximum horizontal acceleration, A. An important quantity derived for each earthquake is its magnitude, m.

The above data are very important to seismic engineers and architects, who are concerned with the design of buildings likely to withstand anticipated earthquakes and to seismologists interested in the structure of the earth. For example, it may be desired to predict A for given m and r.

Geophysical theory suggests an attenuation law of the form

$$\log A = \alpha + \beta m - \log r + \gamma r$$

but other functional forms may provide better descriptions.

Joyner and Boore, *Bulletin of the Seismological Society of America*, 1981, collected data for $I = 23$ California earthquakes. This data may be denoted $A_{ij}, m_i, d_{ij}, j = 1, 2, \ldots J_i, i = 1, 2, \ldots, I$, where i refers to a particular earthquake, m_i is its magnitude, and d_{ij} is the epicentral distance of the j^{th} seismometer recording the i^{th} event. Take the hypocentral distance to be given by $r = \sqrt{d^2 + 7^2}$.

These data may be found on the website in the file `earthquakes.dat`. Column 1 indexes the event, column 2 provides the magnitudes, column 4 the epicentral distance in kilometers, and column 5 the maximum horizontal acceleration in units of g (the acceleration due to gravity).

Consider attenuation relations as statistical models, being specific concerning assumptions. Specifically,

1. Fit and assess the model

$$\log A_{ij} = \alpha + \beta m_i - \log r_{ij} + \gamma r_{ij} + \epsilon_{ij}$$

where the ϵ_{ij} are i.i.d. $N(0, \sigma^2)$.

2. Provide a 99% confidence interval for the maximum acceleration at an epicentral distance $d = 50$ km from a magnitude 7.0 event.

3. You will have noticed some outliers. Refit the model using an appropriate M-estimator, and recalculate your confidence interval from part (2). Comment on how this compares with other approaches you might have used to cope with the presence of outliers.

4. Consider the additive model

$$\log A_{ij} = \phi(m_i) + \gamma(r_{ij}) + \epsilon_{ij}$$

with ϕ and γ being unknown smooth functions. Compare your fitted model with that derived from the geophysical theory. Is there statistically significant deviation from the geophysical model? You should answer this formally, while explaining the approximate nature of your results.

2 Part II - January Temperature Distribution in USA

The data that we will analyze gives the normal average January minimum temperature in degrees Fahrenheit with the latitude and longitude of 56 U.S. cities. (For each year from 1931 to 1960, the daily minimum temperatures in January were added together and divided by 31. Then, the averages for each year were averaged over the 30 years.). We wish to explore whether there are any interesting geographical features in the temperature distribution. The datafile is `UStemp.dat`

1. Try to reproduce the image *TempDist1.png* on the website. Hint: You should work with negative longitudes (think of this as degrees west of the prime meridian).

2. Use *loess* to fit a model predicting average January minimum temperature using latitude and longitude.

3. Using your fitted model predict the temperature at a grid of points and try to show visually what you discover. Be sure to interpret this somewhat geographically.

4. Outline a method of choosing in reasonable manner the value of the smoothing parameter. Redraw the image you used in question 3 for several different values of the smoothing parameter. Comment on your results.

5. Now take a different approach. Examine pairwise plots of JanTemp with Latitude and then with longitude. What do you notice?

6. Fit a linear model to predict JanTemp using longitude. Be sure to state your model, assumptions, conclusions. Remember your interpretation should be in a geographical context.

7. Fit a linear model to predict JanTemp using latitude. Be sure to state your model, assumptions, conclusions.

8. Fit a linear model to predict JanTemp using both latitude and longitude. Be sure to state your model, assumptions, conclusions.

9. Compute the residuals to the model using latitude as a predictor. Plot the residuals against longitude. What do you observe?

10. Fit a smooth function with the residuals as the response variable and longitude as your prediction variable. You may use kernel regression, loess, smoothing splines or another method of your own choosing.

11. Outline and use a method of choosing the value of the smoothing parameter in question 10. Show the resulting fits for several different values of the smoothing parameter.