Lecture 10: Correspondence Analysis and Multiple Correspondence Analysis

In this lecture we explore another two descriptive projection methods - Correspondence Analysis (CA) and Multiple Correspondence Analysis (MA). Correspondence analysis is similar to PCA but rows and columns are treated equivalently and the method aims to describe the dependencies between two variables.

Slide 46

Our working example is a set of 592 women for which we know the color of the hair (a variable that takes 4 values) and the color of the eyes (another variable that takes 4 values). The table on slide 46 is called the contingency table. We will also need notation for column sums, row sums and the sum of all elements.

Slides 47-48

We introduce:
row profiles: r_{ij}
column profiles: o_{ij}
row mass: m_i (= average column profile = weighted average of column profiles)
column mass: c_j (= average row profile = weighted average of row profiles)

The mass is indicative of the relative importance of each row e.g. there are more people with brown eyes than those with blue eyes.

The question is can we use PCA on the row profiles. We first need to center and rescale them.

Slides 49-51

Centering: We subtract the average column profile, not the column average i.e. we take masses into account.
Rescaling: Using the standard deviation is a bad idea so we divide by sqrt{cj}.
Distance: We define the X^2 Euclidian distance between two normalized rows.

All this might look a bit messy now but it will make more sense later.

Slides 52-59

We now perform PCA but we compute the covariance matrix scaled by the masses. Diagonalizing and projecting onto the first two axes doesn't seem very useful. To make it more

interesting, we add the histograms of eye color for each type of hair and the barycenters of eyes-points weighted by the frequency of that type of hair i.e.

(coordinate of brown eyes)*(percent of brown-eyed dark-haired people) +
(coordinate of Hazel eyes)*(percent of Hazel-eyed dark-haired people) + ...

They all lie within the convex hull of the eyes-points. In slide 57 we do the same thing but for the columns. We get the 'opposite' graph i.e. quadrangles of the same shape but different scale.

Slides 60-61

In the previous slides we saw the duality of the row and column analysis. Such a duality is also present with PCA. With PCA however, we rescale using standard deviation and we diagonalize both column and row covariance matrices using the same normalized table. The duality then arises from the properties of diagonalization. In CA, however, the rescaling is different for column and row analysis. The duality arises from the weighted covariance. The computations on slide 61 show that the divergence matrices for the row and column analysis are the same which explains the duality.

Slide 62

Consider the table that we would have if hair color and eye color were independent. Introduce the inertia (given by the formula in the slide). The sum of squares represents the difference between the real table and the theoretical one. Thus, the inertia measured how dependent the rows and columns are and CA finds the axes that best display this dependence.

Slides 64-71

We now want to do a similar thing but for more variables and we use Multiple Correspondence Analysis. The example consists of n subjects taking a questionnaire of 3 questions, having 4, 3 and 4 possible answers (modalities) respectively. First, we transform the normal table to a binary one by encoding an answer with 4,3 and 4 bits respectively. Then, multiply the n x p matrix of 0s and 1s by its transpose to obtain its compact p x p form – the so called Burt table. Note that 'on the diagonal' we have three diagonal matrices corresponding to each question. The ith number on the diagonal for each of them is the number of people that gave answer i to that question. We now run CA on the Burt table. The transition relations and essential properties are given in slides 69-71.

Slide 72

Note that knowing the number of questions Q and the number of people n we can do a more sophisticated computation of the inertia (both the one for modalities and the one for a question).

Slide 73

The computations and conclusions in slide 72 suggest two tricks that can improve MCA results (i.e. decrease inertia or dependencies). One is to group rare modalities e.g. to group countries by

continent, separate continuous modalities in bins or just make them supplementary. The other is to have not too many possible answers for each question.

Slide 74

Consider again the case of two variables. There are three different approaches – using the binary disjunctive table, the Burt table or the contingency table. All of them return the same result. This shows that MCA is just an extension of CA.

Slide 75

Like with PCA, we can increase the quality of the graphs by including supplementary elements e.g. continuous variables. The computations involved are more extensive than the ones for PCA but the results are very powerful.

Slides 76-86

These slides discuss a real world example of using PCA – one of the big successes of the approach. People are asked to rate on a scale of 7 levels around 200 words that best represent human emotion and are universal. Then PCA is run on the resulting table. As expected, the first axis just gives the 'good-bad' property of the words which we already know. The next 5-8 axes turn out to be highly meaningful. They were labeled manually e.g. "duty-pleasure", "heart-reason", etc. (look at slides 81-84). Because of these axes, semiometry turns out to be very useful in areas such as politics and marketing.