

Problem Set #2

MACS 30200, Dr. Evans and Dr. Soltoff

Due Monday, Apr. 30 at 11:30am

1. **2D kernel density estimator (3 points).** The data `BQmat_orig.txt` is a 78×7 matrix of percentages representing the values of a two-dimensional histogram of the percent of the U.S. population that receives all the bequests (inheritances) by a recipient's age (ages 18 to 95, rows) and by a recipient's lifetime income group (7 categories, columns). The seven lifetime income groups are percentiles. Let $prcntl_j$ be the percent of the population in lifetime income group j . The lifetime income groups in the $J = 7$ columns of the `BQmat_orig.txt` data are the following.

$$\mathbf{prcntl} = [0.25, 0.25, 0.20, 0.10, 0.10, 0.09, 0.01], \quad \text{such that} \quad \sum_{j=1}^7 prcntl_j = 1$$

You can read this file into memory using the `numpy.loadtxt` function.

```
bq_data = np.loadtxt('BQmat_orig.txt', delimiter=',')
```

So the $[11, 5]$ -th element of the `bq_data` matrix represents the percent of total bequests (inheritances) received by age-28 and lifetime income group $j = 5$ (80th to 90th percentile of lifetime income).

- (a) Read in the bequests data as a 78×7 NumPy array. Plot the 2D empirical histogram of these data as a 3D surface plot with age and income group on the x -axis and y -axis and the histogram density on the z -axis using a 3D surface plot tool (not a 3D bar histogram tool). Make sure that the axes are labeled correctly. And make sure that your 3D histogram is presented from a perspective that allows a viewer to see that data (don't let the data be hidden by a poor angle of the plot.)
 - (b) Fit a bivariate kernel density estimator to the data. Use a Gaussian kernel. Choose a bandwidth parameter λ that you think is best. Justify your choice of that parameter. Plot the surface of your chosen kernel density estimator. Make sure that the axes are labeled correctly. And make sure that your 3D histogram is presented from a perspective that allows a viewer to see that data. What is the estimated density for bequest recipients who are age 61 in the 6th lifetime income category ($j = 6$, 90th to 99th percentile).
2. **Interaction terms (4 points).** `biden.csv` contains a selection of variables from the 2008 American National Election Studies. Estimate the following linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (1)$$

where Y is the Joe Biden feeling thermometer, X_1 is age, and X_2 is education. Report the parameters and standard errors.

- (a) Evaluate the marginal effect of age on Joe Biden thermometer rating, conditional on education. Consider the magnitude and direction of the marginal effect, as well as its statistical significance.
- (b) Evaluate the marginal effect of education on Joe Biden thermometer rating, conditional on age. Consider the magnitude and direction of the marginal effect, as well as its statistical significance.

For both elements, provide both a written answer and supporting graphical evidence.

3. Parallel computing (3 points). Put parallel computing exercise here.