CS&SS/SOC/STAT 563 — Statistical Demography — Spring 2022 - Homework no. 3

Due Monday April 25 at 2:15pm on the course Canvas site.

1. This question is about the female mortality rates in Peru for 1950–2020 extracted in Homework 2.

   (a) Fit the Lee-Carter model to the data using the approximate least squares method described in class.

   (b) Fit the Lee-Carter model to the data using the SVD method described in class, and plot the results, including the estimated parameters and the fitted curves. Compare them to the observed rates graphically.

   (c) Compare the results from the SVD method to those from the approximate least squares method described in class, numerically, graphically and verbally. Find a normalization of the SVD results needed to make the SVD results close to the least squares results.

   (d) Use the Lee-Carter method to obtain probabilistic forecasts of mortality in Peru for 2020-2025. Obtain probabilistic forecasts of the mortality index, k_t. Hence find a confidence interval for the age-specific mortality rate, ${}_nm_x$ for women aged 75-80 in 2020–2025. This should take account only of uncertainty about future values of the morality index.

2. The purpose of this question is to give you some experience with basic Bayesian statistics. You could use any Bayesian textbook, such as the one by Hoff (2009). This question involves Bayesian estimation of a binomial mean. In a sample of 112 people living in Washington who married in 2005, 43 had divorced by 2015. Assume a uniform prior distribution for the proportion, θ, of Washington residents who married in 2005 that had divorced by 2015.

   (a) Find the posterior distribution of θ in analytic form, i.e. as a mathematical expression. [Hint: A uniform distribution is the same as a Beta (1,1) distribution.]

   (b) Simulate a sample of size 1,000 from the posterior distribution. Hence find the posterior mean and a 95% Bayesian confidence interval for θ. [Hint: You can use the rbeta R function.]

   (c) Compute and plot a nonparametric density estimate of the posterior density of θ.

3. This question is about Bayesian estimation of a normal mean. A random sample of 25 students from a high school was taken, and the amount of time each student spent on studying (in hours) during an exam period was measured. The numbers were as follows:

   2.1 9.8 13.9 11.3 8.9 15.7 16.4 4.5 8.9 11.9 12.5 11.1 11.6 14.5 9.6 7.4 3.3 9.1 9.4 6.0 7.4 8.5 1.6 11.4 9.7

   Assume that the true standard deviation of the number of hours is 4. Then find the posterior distribution of the true mean of hours spent studying by the students in the high school, using a normal prior distribution with mean 10 hours and standard deviation 3 hours. Hence find a posterior 95% Bayesian confidence interval for the true mean.