STATS 551 Homework 1

Introduction & Single Parameter Models

Due date: 6:00 pm (EST) Jan. 29, 2018

Normal distribution with unknown mean (4×10 points). A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\overline{y} = 70$ kilograms. Assume that weights in the population are normally distributed with unknown mean θ and known standard deviation 10 kilograms. Suppose your prior distribution for θ is normal with mean 60 and standard deviation 20. (1 kilogram $\approx$ 2.20462 pounds)

1. Give your posterior distribution for θ, as a function of n.
2. A new student is sampled at random from the same population and has a weight of $\tilde{y}$ kilograms. Give the posterior predictive distribution for $\tilde{y}$, as a function of n.
3. For $n = 9$, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$.
4. Do the same for $n = 99$.

Guideline for Submission: submit a hard copy (handwritten or printed).

Discrete sample spaces (2×15 points). Suppose there are N cable cars in San Francisco, numbered sequentially from 1 to N. You see a cable car at random; it is numbered 203. You wish to estimate N.

1. Assume your prior distribution on N is geometric with mean 100, i.e.

   $$p(N) = 0.01 \times (0.99)^{N-1}, \text{ for } N = 1, 2, \ldots.$$

   What is your posterior distribution for N?

2. What are the posterior mean and standard deviation of N? (Sum the infinite series analytically or approximate them on the computer.)

Guideline for Submission: submit a hard copy (handwritten or printed).

Nonconjugate single parameter model (3×10 points). Suppose $y_1, \cdots, y_6$ are independent samples from a Cauchy distribution with unknown center θ and known scale 1: $p(y_i|\theta) \propto 1/(1+(y_i-\theta)^2)$. Assume that the prior distribution for θ is uniform on $[0, 100]$. Given the observations $(y_1, \cdots, y_6) = (42, 44.5, 46, 46.8, 47.2, 50)$:

1. Compute the unnormalized posterior density function, $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 100$, for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function $p(\theta|y)$, as a function of θ.

2. Sample 2000 draws of θ from the posterior density and plot a histogram of the draws.

3. Use the samples of θ to obtain 2000 samples from the predictive distribution of a future observation y_7, and plot a histogram of the predictive draws.

Guideline for Submission: submit R markdown (or jupyter notebook) with annotated code followed by results. Discussions about the results should follow the results.

Optional Reading. Read one of the following papers and post your summary and thoughts on Canvas. Bonus points up to 5 will be rewarded.

1. Biostatistics and Bayes, Norman Breslow, Statist. Sci., Volume 5, Number 3 (1990), 269-284.

2. Bayesian Methods in Practice: Experiences in the Pharmaceutical Industry, A. Racine, A. P. Grieve, H. Fluhler and A. F. M. Smith, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 35, No. 2 (1986), pp. 93-150.