STAT 489-01: Bayesian Methods of Data Analysis

Problem Set 5

Assigned 2017 March 2
Due 2017 March 9

Show your work on all problems! Be sure to give credit to any collaborators, or outside sources used in solving the problems. Note that if using an outside source to do a calculation, you should use it as a reference for the method, and actually carry out the calculation yourself; it's not sufficient to quote the results of a calculation contained in an outside source.

1 Unfinished Business

Consider the probability distribution from problems 2 and 3 of the exam, which was a Dirichlet(y_1,y_2,y_3) distribution for ϕ_1, ϕ_2 and $1-\phi_1-\phi_2$, as well as the Gaussian approximation in the parameters $\theta_j = \ln\phi_j - \ln(1-\phi_1-\phi_2)$, $j=1,2$. (Refer to the exam solutions for the relevant formulas.)

(a) Write a program or script which, given values of y_1, y_2, and y_3, generates samples from the exact Dirichlet distribution and the Gaussian approximation (this can be done using the `rdirichlet()` function from the `gtools` R library and either the `rmvnorm()` function from the `mvtnorm` library or the `mvrnorm()` function from the `MASS` library), and uses those samples to produce

- **(i)** Ternary plots of each of the two samples in ϕ_1, ϕ_2 and $1-\phi_1-\phi_2$ space
- **(ii)** Scatter plots of each of the two samples in θ_1, θ_2 space
- **(iii)** Estimates, from each of the two samples, of $E(\theta_1|\mathbf{y},I)$, $E(\theta_2|\mathbf{y},I)$ $V(\theta_1|\mathbf{y},I)$, $V(\theta_2|\mathbf{y},I)$ and $\mathrm{Cov}(\theta_1,\theta_2|\mathbf{y},I)$.

(b) Apply your script when $(y_1,y_2,y_3)=(51,31,21)$.

(c) Apply your script when $(y_1,y_2,y_3)=(6,4,3)$.

(d) Apply your script when $(y_1,y_2,y_3)=(4.5,2.5,0.5)$.

2 Jaynesian Evidence

Jaynes defines the "evidence" for a hypothesis H given a state of knowledge X (which might be prior information I, so $X=I$, or prior information along with some data D, in which case $X=D,I$ as

$$e(H|X) = 10\log_{10}\frac{\Pr(H|X)}{\Pr(\overline{H}|X)} = 10\log_{10}\frac{\Pr(H|X)}{1-\Pr(H|X)} \tag{2.1}$$

where $\overline{H}$ is the logical negation of H. Since this is something different from what is usually called "evidence", we will refer to it as the "Jaynesian evidence", and quote it in units of decibels (dB). One can use Bayes's theorem to show that if you add new data D, the Jaynesian evidence is modified by an additive term:

$$e(H|D,X) = e(H|X) + 10\log_{10}\frac{\Pr(D|H,X)}{\Pr(D|\overline{H},X)} \tag{2.2}$$

(a) Suppose we have a bag containing three four-sided dice, two with one red face and three white faces, and one with two red faces and two white faces. I draw a die out but don't show it to you. Let H_1 and H_2 be the hypotheses that the die has one and two red faces, respectively. Show that the approximate Jaynesian evidences for these hypotheses are $e(H_1|I) \approx +3\,\mathrm{dB}$ and $e(H_2|I) \approx -3\,\mathrm{dB}$.

(b) Suppose I toss the die and report to you the color of the bottom face. Let R_1 represent a result of "red" on this first toss and W_1 represent a result of "white". Calculate $\Pr(R_1|H_1,I)$, $\Pr(R_1|H_2,I)$, $\Pr(R_1|\overline{H_1},I)$, $\Pr(R_1|\overline{H_2},I)$, $e(H_1|R_1,I)-e(H_1|I)$, and $e(H_2|R_1,I)-e(H_2|I)$. Repeat the calculation for W_1 and obtain $e(H_1|W_1,I)-e(H_1|I)$, and $e(H_2|W_1,I)-e(H_2|I)$.

(c) Let $D_{n,r}$ represent some series of n tosses containing r reds, and let R_{n+1} or W_{n+1} represent a result of red or white, respectively, on the $n+1$st toss. Show that $\Pr(R_{n+1}|D_{n.r},H_j,I) = \Pr(R_{n+1}|H_j,I)$, $\Pr(R_{n+1}|D_{n.r},\overline{H_j},I) = \Pr(R_{n+1}|\overline{H_j},I)$, $j = 1,2$, and therefore that $e(H_j|R_{n+1},D_{n,r},I) - e(H_j|D_{n,r},I) = e(H_j|R_1,I) - e(H_j|I)$, i.e., the same exchange of evidence between the competing hypotheses occurs with each observed red toss. (And likewise for white.)

(d) Now suppose we allow for a thousand-to-one chance that we were actually misinformed about the contents of the bag, and I somehow drew a die with three red faces. Calling this hypothesis H_3, we have $e(H_3|I) = -30\,\mathrm{dB}$. (In principle the prior Jaynesian evidences for H_1 and H_2 should go down slightly, but it's a negligible difference which we can ignore to this level of approximation.) Show that while $\Pr(R_{n+1}|D_{n.r},H_j,I) = \Pr(R_{n+1}|H_j,I)$ as before, now with $j = 1,2,3$, $\Pr(R_{n+1}|D_{n.r},\overline{H_j},I) \neq \Pr(R_{n+1}|\overline{H_j},I)$. Write $\Pr(R_{n+1}|D_{n.r},\overline{H_1},I)$ in a form depending on $\Pr(H_2|D_{n.r},I)$ and $\Pr(H_3|D_{n.r},I)$, and similarly for $\Pr(R_{n+1}|D_{n.r},\overline{H_2},I)$ and $\Pr(R_{n+1}|D_{n.r},\overline{H_3},I)$.

(e) Consider the case where $\Pr(H_2|D_{n.r},I) \gg \Pr(H_3|D_{n.r},I)$; what is the approximate change in evidence $e(H_1|R_{n+1},D_{n.r},I) - e(H_1|D_{n.r},I)$ associated with observing a red toss? What about when $\Pr(H_2|D_{n.r},I) \ll \Pr(H_3|D_{n.r},I)$? What about for a white toss in each case?

(f) Consider the corresponding limiting cases to find approximate evidence changes for H_2 and H_3 resulding from a red toss, and from a white toss.

(g) Download the data `http://ccrg.rit.edu/~whelan/courses/2017_1sp_STAT_489/data/ps05_prob2.dat`
which represent the result of a series of tosses, 1 representing red and 0 representing white. Plot the evidences $e(H_j|D_{n,r},I)$, $j = 1,2,3$ (where r is the number of reds in the first n tosses) for the first n results versus n from 0 to 60. (It is probably easier just to calculate $e(H_j|D_{n,r},I)$ all at once than to try to calculate each incremental adjustment along the way.)

3 Bayes Factor and Occam Factors

Suppose we have collected data $y_1, \ldots, y_n$ and are trying to evaluate two models: M_0 says the data are a sample from a $N(0, \sigma^2)$ distribution, and M_1 says they are a sample from a $N(\theta, \sigma^2)$ distribution, where σ is a known quantity and the same for both models. Suppose that M_1 assigns a Gaussian prior $p(\theta|M_1, I) = (2\pi\sigma_\theta^2)^{-1/2} e^{-\theta^2/(2\sigma_\theta^2)}$ to the value of the parameter θ.

(a) Construct the "evidence" $p(\mathbf{y}|M_0, I)$ for model M_0 as a function of the data $\mathbf{y}$, using previous results and simplifying as much as possible.

(b) Construct the sampling distribution $p(\mathbf{y}|\theta, M_1, I)$ for model M_1 (which is also the likelihood function).

(c) Construct the "evidence" $p(\mathbf{y}|M_1, I)$ for model M_1.

(d) Find the maximum-likelihood value $\widehat{\theta}(\mathbf{y})$ which maximizes $p(\mathbf{y}|\theta, M_1, I)$.

(e) Construct the Bayes factor $\mathcal{B}_{10} = p(\mathbf{y}|M_1, I)/p(\mathbf{y}|M_0, I)$ comparing the two models.

(f) Construct the maximized likelihood ratio $p(\mathbf{y}|\widehat{\theta}(\mathbf{y}), M_1, I)/p(\mathbf{y}|M_0, I)$, and show that it is always greater than or equal to one.

(g) Write the Bayes factor $\mathcal{B}_{10}$ as a product of three quantities: i) The maximized likelihood ratio from the previous part, ii) a constant, equal to the value when $\overline{y} = 0$ (this is the Occam factor), iij) whatever's left.

(h) Choosing a non-informative prior for model M_1 corresponds to taking the limit as $\sigma_\theta \to \infty$. What happens to each of the three parts of the Bayes factor in that limit?

(i) Suppose $\sigma_\theta = 5\sigma/\sqrt{n}$. Plot, as a function of $\overline{y}/\sqrt{\sigma^2/n}$, the log of the Bayes factor, along with the log of the three quantities above i) over the range $-3 < \overline{y}/\sqrt{\sigma^2/n} < 3$, and ii) over the range $-15 < \overline{y}/\sqrt{\sigma^2/n} < 15$.

4 Gelman Chapter 6, Exercise 7