ASTP 611-01: Statistical Methods for Astrophysics

Problem Set 11

Assigned 2017 November 30
Due 2017 December 11

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 Monte Carlo

Consider an observation where transient events are arriving from all points of the visible sky (above the horizon). An experimenter wishes to test the hypothesis that the locations of these events are isotropic by dividing the sky into approximately (but not quite) equal-area patches and constructing a chi-square statistic out of the expected and observed numbers of events occurring in each patch. Suppose that she chooses six patches, using as coördinates the angle θ down from her zenith and an azimuth ϕ measured around her zenith direction:

- $S_1 \equiv \{0 \leq \theta < \frac{\pi}{4}, 0 \leq \phi < \pi\}$
- $S_2 \equiv \{0 \leq \theta < \frac{\pi}{4}, \pi \leq \phi < 2\pi\}$
- $S_3 \equiv \{\frac{\pi}{4} \leq \theta < \frac{\pi}{2}, 0 \leq \phi < \frac{\pi}{2}\}$
- $S_4 \equiv \{\frac{\pi}{4} \leq \theta < \frac{\pi}{2}, \frac{\pi}{2} \leq \phi < \pi\}$
- $S_5 \equiv \{\frac{\pi}{4} \leq \theta < \frac{\pi}{2}, \pi \leq \phi < \frac{3\pi}{2}\}$
- $S_6 \equiv \{\frac{\pi}{4} \leq \theta < \frac{\pi}{2}, \frac{3\pi}{2} \leq \phi < 2\pi\}$

and observes a total of $n = 100$ events. Then the expected number of events in patch S_i is $E_i = n\Omega_i/(2\pi)$ where $\Omega_i = \int_{\phi_i^{\text{min}}}^{\phi_i^{\text{max}}} \int_{\theta_i^{\text{min}}}^{\theta_i^{\text{max}}} \sin\theta \, d\theta \, d\phi$ is the solid angle of patch S_i, and the chi-squared statistic is

$$W = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} \tag{1.1}$$

where O_i is the actual number of events observed in patch S_i.

a) Work out the expected numbers of events $E_1 = E_2$ and $E_3 = E_4 = E_5 = E_6$.
b) Perform a Monte Carlo simulation with 100,000 iterations as follows. For each iteration:
   i) Generate $n = 100$ (θ, ϕ) pairs each drawn from the distribution

$$f(\theta, \phi) = \frac{\sin\theta}{2\pi} \qquad 0 \leq \theta \leq \pi/2;\ 0 \leq \phi < 2\pi \tag{1.2}$$

   (Hint: factor this into independent pdfs for θ and ϕ separately, and use the pdf for each along with a uniform random number generator.)

ii) Count how many of the sky points fall into each patch.
iii) Construct the chi-squared statistic and note its value.
iv) Generate a p-value from this statistic using the chi-squared distribution with the appropriate number of degrees of freedom.
v) As you go along, keep track of how often $p < 0.05$, $p < 0.01$ and $p < 0.001$. Compare these results to your theoretical expectations.

2 Sampling from a Posterior

Consider an on-off sampling experiment in which we observe $k_{\text{ON}} = 6$ events in an on-source time of $T_{\text{ON}} = 1\,\text{hr}$ and $k_{\text{OFF}} = 9$ events in an off-source time of $T_{\text{OFF}} = 2.5\,\text{hr}$. In terms of the background rate b and the on-source rate $r = b + s$ the posterior, assuming uniform priors on both rates, will be

$$f(b, r|D, I) \propto (bT_{\text{OFF}})^{k_{\text{OFF}}} e^{-bT_{\text{OFF}}} (rT_{\text{ON}})^{k_{\text{ON}}} e^{-rT_{\text{ON}}} \qquad 0 < b < r < \infty \tag{2.1}$$

a) Draw a sample of $N = 10^4$ points from the joint posterior, by drawing pairs of Gamma random variables (e.g., using `scipy.stats.gamma.rvs` with the appropriate shape and scale parameters) and rejecting any pair for which $r < b$. Make a scatter plot of the points in (b, r) space.
b) Change variables from (b, r) to (b, s) where $s = r - b$ and make a scatter plot of the sample in (b, s) space. (Note that you don't need to do any pdf transformations; just calculate the s corresponding to the b and r values from each point in the sample.)
c) Make a histogram ($\sqrt{N}$ bins is likely to look nice) of the s values in the sample. This is an estimate of the marginal pdf for s.
d) Calculate the following quantities from the sample of s values: i) mean, ii) standard deviation, iij) median, iiij) 5th and 95th percentiles. These are estimates of the corresponding quantities from the marginal posterior for s.
e) Change variables from (b, s) to $(\log_{10} b, \log_{10} s)$ and scatter-plot the sample in the new variables. (Again, no Jacobian is necessary; the density takes care of itself.)

3 Markov Chain Monte Carlo

Use the ipython notebook `http://ccrg.rit.edu/~whelan/courses/2017_3fa_ASTP_611/data/ps11.ipynb` to perform and evaluate MCMC simulations for the Bradley-Terry model. Wherever there is an **EXERCISE**, add code at that point in the notebook to perform the requested calculations.