Survmeth 895

Midterm Examination

Due date: March 15, 2017

Note: Please solve the problems on your own and do not consult with anybody. If you have any questions then please e-mail them to the instructors. Submit your own solutions and include computer code as a part of your solutions.

1. A study was conducted to evaluate the effectiveness of anesthetics used in dental surgery. Three different anesthetics were compared in two different kinds surgery, periodontic (Stratum 1) or endodontic (Stratum 2). The primary binary outcome of interest is whether or not the patient reported pain during the surgery. The following table provides the data from the experiment on 76 patients in Stratum 1 and 97 patients in Stratum 2. In each cell, the numerator is the number patients reporting pain during the surgery and the denominator is the sample size.

Stratum	Anesthetic A	Anesthetic B	Anesthetic C
1	11/23	8/24	6/29
2	9/29	4/31	5/37

   (a) In a dental practice, 70% of the surgeries are endodontic and 30% are periodontic. The objective is to infer about the population proportions of patients who would report pain during the surgery under each Anesthetic. Assuming a large population, obtain the posterior distributions of the population proportion of patients who would report pain when Anesthetic A is administered for a suitable choice of a prior distribution.

   (b) A common quantity used to compare anesthetics is the relative risk of pain, the ratio of the population proportion of patients who would have reported pain under Anesthetic A to those who would have reported pain under Anesthetic B. Simulate the posterior distribution of the relative risk of pain under A relative to B, and the relative risk of pain under A relative to C. How strong is the evidence in favor of Anesthetic B or C when compared to the current standard, Anesthetic A?

2. A large survey of 5000 respondents is being planned. Two different protocols, A and B, are under consideration. These involve differing incentives, the number and type of advance letters, the number of call attempts and other factors that may potentially improve participation rates. A pilot study (not to be part of the actual survey) was conducted to evaluate these two protocols. In this study, fifty subjects were assigned to Protocol A and 50 to Protocol B. Two indicators were measured on these 100 subjects. The ultimate response status (1=Respondent and 0=Nonrespondent)and the cost associated with each subject up until the final disposition of the response status(but not

the interview cost). This cost was determined using a unique computerized logging system that kept track of the number of minutes spent on each subject by each staff member in the survey operation unit.

```
Protocol A:

Final Disposition:

1 0 0 1 0 1 0 1 1 0 0 1 1 1 0 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 1 1
0 1 1 1 1 1 1 0 1 1 1 1

Cost (in $s):

32 30 36 38 48 44 32 33 27 41 30 46 29 27 35 33 47 36 29 31 34 47 39 29 32
42 31 43 35 30 33 36 29 31 35 42 31 26 36 32 32 32 31 36 31 37 37 47 31 29

Protocol B:

Final disposition:

0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 0
1 1 1 1 1 1 1 1 1 1 0 0

Cost (in $s):

38 38 43 37 51 46 47 53 47 47 35 47 45 37 55 45 48 40 45 47 40 47 36 37 47
46 43 33 44 39 46 40 42 40 46 35 41 41 49 43 34 47 48 42 44 38 40 37 42 41
```

(a) What is the posterior distribution of the number of respondents if Protocol A were used in the actual survey for a suitable choice of a prior distribution?

(b) Another objective is to obtain the posterior distribution of the total cost for 5000 respondents if Protocol A were used. Inspect the cost data, develop suitable model assumptions including prior distributions, and simulate the posterior distribution of the total cost.

(c) Repeat (a) and (b) for Protocol B.

(d) One parameter of interest useful in comparing the two protocols is the average cost per respondent. That is, the total cost for 5000 respondents divided by the number of respondents. Simulate the posterior distribution of the difference in the cost per respondent between Protocols A and B. Based on the draws, quantify the evidence in favor of one Protocol over the other. What are your recommendations?

3. For $i = 1, 2, \ldots, n$, let y_i conditional on λ_i be independent Poisson random variables with mean λ_i. Let λ_i be identical and independent Gamma random variables with

parameters α and β. Let α and β be independent Gamma random variables with parameters (a, b) and (c, d) respectively. Assume that a, b, c and d are known. Derive all the relevant conditional distributions for implementing a Gibbs sampling algorithm. That is, derive the following conditional distributions,

- $\pi(\lambda|y, \alpha, \beta)$
- $\pi(\alpha|\lambda, y, \beta)$
- $\pi(\beta|\lambda, y, \beta)$

4. A large departmental store handles about 20,000 accounts receivable per month. A 2% sample ($k = 400$) of accounts was verified each month over a 2-year period ($n = 24$ months). The number of accounts found to be in error per month (out of 400) were (in order of magnitude) 0, 0, 1, 1, 2, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, 9, 10, 10, 13, 14, 17. Assume that the time pattern is erratic. The objective is to obtain the posterior distribution of the total number of accounts with error over the two-year period. Develop appropriate model assumptions, including prior distributions, and carry out your analysis. Under your model assumptions including your choice of prior distri, simulate the posterior distribution of the total number of accounts with errors. Obtain 95% equal tail Bayesian credible interval for the total number of accounts with errors. How will you simulate the posterior distribution of number of accounts with errors during the next two year period?