IYSE 6420 Fall 2020 Midterm

Xiao Nan

GT Account: nxiao30

GT ID: 903472104

1. **Bayes Network**. Incidences of diseases and depend on the exposure . Disease is additionally influenced by risk factors . Both diseases lead to symptoms . Results of the test for disease are affected also by disease . Positive test will be denoted as , negative as . The Bayes Network is shown in Figure 1. Needed conditional probabilities are shown in Table 1.

A close up of a watch

Description automatically generated

A close up of text on a white background

Description automatically generated

**(a) What is the probability of disease , if disease is not present , but symptoms are present .**

Thus,

Similarly,

Finally,

**(b) What is the probability of exposure , if symptoms are present and test is positive .**

**Hint: You can solve this problem by any of the 3 ways: (i) use of WinBUGS or OpenBUGS, (ii) direct simulation using Octave/MATLAB, R, or Python, and (iii) exact calculation**

We have,

Similarly,

Thus,

2. **Times to Failure**. **Three devices are monitored until failure. The observed lifetimes are 0.9, 1.8, and 0.3 years. If the lifetimes ate modelled as exponential distribution with rate λ,**

**Assume exponential prior on λ,**

**(a) Find the posterior distribution of .**

**(b) Find the Bayes estimator for .**

**(c) Find the MAP estimator for .**

**(d) Numerically find 95% equitailed confidence interval for .**

**(e) Find the posterior probability of hypothesis .**

(a)

is gamma

Likelihood,

Prior exponential ,

Then posterior is gamma where

(b)

Bayes estimator is posterior mean,

(c)

MAP estimator is posterior mode,

(d)

95% credible sets

Matlab code:

|  |
| --- |
| gaminv(0.025,4,1/5)  % ans = 0.218  gaminv(0.975,4,1/5)  % ans = 1.7535 |

(e)

Matlab code:

|  |
| --- |
| gamcdf(0.5, 4, 1/5)  % ans = 0.2424 |

3. **Gibbs and High/Low Protein Diet in Rats**. **Armitage and Berry (1994, p. 111) report data on the weight gain of 19 female rats between 28 and 84 days after birth. The rats were placed on diets with high (12 animals) and low (7 animals) protein content.**

A screenshot of a cell phone

Description automatically generated

**We want to test the hypothesis on dietary effect. Did a low protein diet result in significantly lower weight gain?**

**The classical test against one sided alternative will be significant. We will do the test Bayesian way using Gibbs sampler.**

**Assume that high-protein diet measurements are coming from normal distribution , where is precision parameter,**

**Low-protein diet measurements are coming from normal distribution**

**,**

**Assume that and have normal priors and , respectively. Take prior means as (apriori no preference) and precisions as**

**Assume that and have the gamma and priors with shapes and rates .**

**(a) Construct Gibbs sampler that will sample from their posteriors.**

**(b) Find sample differences . Proportion of positive differences approximates posterior probability of hypothesis . What is this proportion?**

**(c) Using sample quantiles find the 95% equitailed credible set for . Does this set contain 0?**

(a)

Joint distributions,

Thus, conditional distributions

Construct Gibbs sampler, please refer to code q3.m appended. The distribution is shown as below:

Chart, histogram

Description automatically generated

(b)

Sample differences is 12.3150 .

Proportion of positive differences

This is the proportion in the samples that

(c)

95% equtailed credible set

This set contains 0 .

**Matlab Code for Q3 (**q3.m**)**

|  |
| --- |
| %Reference: FALL 2019 -- MIDTERM Online Course ISyE6420 (Penguins)  %------------------------------------------  clear all;  close all;  clc;  %-----------------figure defaults  lw=2;  set(0, 'DefaultAxesFontSize', 17);  fs=14;  msize = 5;  %--------------------------------  randn('state', 10);  data\_1=[134 146 104 119 124 161 107 83 113 129 97 123];  n1=length(data\_1); % n=12  sum\_1=sum(data\_1);  data\_2=[70 118 101 85 107 132 94];  n2=length(data\_2); % n=7  sum\_2=sum(data\_2);  %------------------------------------------  %  nn = 10000+1000;  theta\_1s=[]; theta\_2s=[]; tau\_1s=[]; tau\_2s=[];  % params  theta10=110; theta20=110; tau10=1/100; tau20=1/100;  a1=0.01; a2=0.01; b1=4; b2=4;  % init  theta1=110; theta2=110; tau1=1/100; tau2=1/100;  h=waitbar(0,'Simulation in progress');  for i = 1 : nn  new\_theta1 = normrnd( (tau1\*sum\_1+tau10\*theta10)/(tau10+n1\*tau1), sqrt(1/(tau10+n1\*tau1)) );  par1 = b1+1/2\*sum((data\_1-new\_theta1).^2);  new\_tau1 =gamrnd(a1 + n1/2, 1/par1);  theta\_1s = [theta\_1s new\_theta1];  tau\_1s = [tau\_1s new\_tau1];  theta1=new\_theta1;  tau1=new\_tau1;    new\_theta2 = normrnd( (tau2\*sum\_2+tau20\*theta20)/(tau20+n2\*tau2), sqrt(1/(tau20+n2\*tau2)) );  par2 = b2+1/2\*sum((data\_2-new\_theta2).^2);  new\_tau2 =gamrnd(a2 + n2/2, 1/par2);  theta\_2s = [theta\_2s new\_theta2];  tau\_2s = [tau\_2s new\_tau2];  theta2=new\_theta2;  tau2=new\_tau2;      if i/50==fix(i/50)  waitbar(i/nn)  end  end  close(h)  %  burnin = 1000;  figure(1)  subplot(2,2,1)  hist(theta\_1s(burnin:nn), 70)  xlabel('theta\_1s')  subplot(2,2,2)  hist(theta\_2s(burnin:nn), 70)  xlabel('theta\_2s')  subplot(2,2,3)  hist(tau\_1s(burnin:nn), 70)  xlabel('tau\_1s')  subplot(2,2,4)  hist(tau\_2s(burnin:nn), 70)  xlabel('tau\_2s')  mean(theta\_1s(burnin:nn)) %117.0744  mean(theta\_2s(burnin:nn)) %104.7364  mean(tau\_1s(burnin:nn)) %0.0022  mean(tau\_2s(burnin:nn)) %0.0025  theta\_diff = theta\_1s-theta\_2s;  mean(theta\_diff) %12.3150  sum(theta\_diff>0)/length(theta\_diff) %0.9192  prctile(theta\_diff, 2.5) %-4.9559  prctile(theta\_diff, 97.5) %28.9730 |

REFERENCES

[1] <https://www2.isye.gatech.edu/isye6420/Bank/MidtermFall2019Sol.pdf>

[2] <http://zoe.bme.gatech.edu/~bv20/isye6420/Bank/HW218.pdf>