IYSE 6420 Fall 2020 Midterm

Xiao Nan
GT Account: nxiao30
GT ID: 903472104

1. **Bayes Network**. Incidences of diseases A and B (D_A, D_B) depend on the exposure (E). Disease A is additionally influenced by risk factors (R). Both diseases lead to symptoms (S). Results of the test for disease A (T_A) are affected also by disease B. Positive test will be denoted as $T_A = 1$, negative as $T_A = 0$. The Bayes Network is shown in Figure 1. Needed conditional probabilities are shown in Table 1.

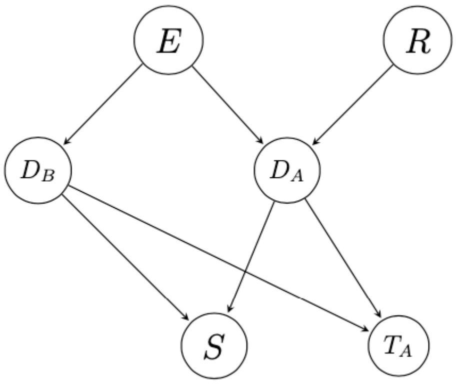


Figure 1: The DAG of the Bayesian networks

Table 1: The known (or elicited) conditional probabilities

E	0	1
	0.8	0.2

R	0	1
	0.7	0.3

D_A	0	1
E^cR^c	0.9	0.1
E^cR	0.4	0.6
ER^c	0.5	0.5
ER	0.3	0.7

D_B	0	1
E^c	0.8	0.2
E	0.3	0.7

S	0	1
$D_A^cD_B^c$	0.95	0.05
$D_A^cD_B$	0.6	0.4
$D_AD_B^c$	0.4	0.6
D_AD_B	0.1	0.9

T_A	0	1
$D_A^cD_B^c$	0.92	0.08
$D_A^cD_B$	0.8	0.2
$D_AD_B^c$	0.15	0.85
D_AD_B	0.03	0.97

(a) What is the probability of disease A $(D_A = 1)$, if disease B is not present $(D_B = 0)$, but symptoms are present $(S = 1)$.

$$P(D_A|D_B^C, S) = \frac{P(D_A, D_B^C, S)}{P(D_B^C, S)} = \frac{P(S|D_A, D_B^C) \times P(D_A, D_B^C)}{P(D_A, D_B^C, S) + P(D_A^C, D_B^C, S)}$$

$$= \frac{P(S|D_A, D_B^C) \times P(D_A, D_B^C)}{P(S|D_A, D_B^C) \times P(D_A, D_B^C) + P(S|D_A^C, D_B^C) \times P(D_A^C, D_B^C)}$$

$$= \frac{0.6 \times P(D_A, D_B^C)}{0.6 \times P(D_A, D_B^C) + 0.05 \times P(D_A^C, D_B^C)}$$

$$P(D_A, D_B^C) = P(D_B^C|ER)P(D_A|ER)P(ER) + P(D_B^C|ER^C)P(D_A|ER^C)P(ER^C) + (D_B^C|E^CR)P(D_A|E^CR)P(E^CR) + (D_B^C|E^CR^C)P(D_A|E^CR^C)P(E^CR^C)$$

$$P(ER) = P(E) \times P(R) = 0.2 \times 0.3 = 0.06$$
$$P(ER^C) = P(E) \times P(R^C) = 0.2 \times 0.7 = 0.14$$
$$P(E^CR) = P(E^C) \times P(R) = 0.8 \times 0.3 = 0.24$$
$$P(E^CR^C) = P(E^C) \times P(R^C) = 0.8 \times 0.7 = 0.56$$

Thus,

$$P(D_A, D_B^C) = 0.3 \times 0.7 \times 0.06 + 0.3 \times 0.5 \times 0.14 + 0.8 \times 0.6 \times 0.24 + 0.8 \times 0.1 \times 0.56 = 0.1936$$

Similarly,

$$P(D_A^C, D_B^C) = P(D_B^C|ER)P(D_A^C|ER)P(ER) + P(D_B^C|ER^C)P(D_A^C|ER^C)P(ER^C) + P(D_B^C|E^CR)P(D_A^C|E^CR)P(E^CR) + (D_B^C|E^CR^C)P(D_A^C|E^CR^C)P(E^CR^C)$$
$$= 0.3 \times 0.3 \times 0.06 + 0.3 \times 0.5 \times 0.14 + 0.8 \times 0.4 \times 0.24 + 0.8 \times 0.9 \times 0.56 = 0.5064$$

Finally,

$$P(D_A|D_B^C, S) = \frac{0.6 \times P(D_A, D_B^C)}{0.6 \times P(D_A, D_B^C) + 0.05 \times P(D_A^C, D_B^C)} = \frac{0.6 \times 0.1936}{0.6 \times 0.1936 + 0.05 \times 0.5064} = 0.821$$

(b) What is the probability of exposure ($E = 1$), if symptoms are present ($S = 1$) and test is positive ($T_A = 1$).

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

$$P(E|S, T_A) = \frac{P(E, S, T_A)}{P(S, T_A)} = \frac{P(S, T_A|E) \times P(E)}{P(S, T_A)} = \frac{P(S, T_A|E) \times P(E)}{P(S, T_A|E) \times P(E) + P(S, T_A|E^C) \times P(E^C)}$$

We have,

$$\begin{aligned}
P(S, T_A|E) &= P(S|D_A^C, D_B^C)P(T_A|D_A^C, D_B^C)P(D_A^C, D_B^C|E) \\
&\quad + P(S|D_A^C, D_B)P(T_A|D_A^C, D_B)P(D_A^C, D_B|E) \\
&\quad + P(S|D_A, D_B^C)P(T_A|D_A, D_B^C)P(D_A, D_B^C|E) \\
&\quad + P(S|D_A, D_B)P(T_A|D_A, D_B)P(D_A, D_B|E) \\
&= P(S|D_A^C, D_B^C)P(T_A|D_A^C, D_B^C)\{P(D_A^C|ER)P(ER) + P(D_A^C|ER^C)P(ER^C)\}P(D_B^C|E) \\
&\quad + P(S|D_A^C, D_B)P(T_A|D_A^C, D_B)\{P(D_A^C|ER)P(ER) \\
&\quad + P(D_A^C|ER^C)P(ER^C)\}P(D_B|E) \\
&\quad + P(S|D_A, D_B^C)P(T_A|D_A, D_B^C)\{P(D_A|ER)P(ER) \\
&\quad + P(D_A|ER^C)P(ER^C)\}P(D_B^C|E) \\
&\quad + P(S|D_A, D_B)P(T_A|D_A, D_B)\{P(D_A|ER)P(ER) \\
&\quad + P(D_A|ER^C)P(ER^C)\}P(D_B|E) \\
&= 0.05 \times 0.08 \times (0.3 \times 0.2 \times 0.3 + 0.5 \times 0.2 \times 0.7) \times 0.3 \\
&\quad + 0.4 \times 0.2 \times (0.3 \times 0.2 \times 0.3 + 0.5 \times 0.2 \times 0.7) \times 0.7 \\
&\quad + 0.6 \times 0.85 \times (0.7 \times 0.2 \times 0.3 + 0.5 \times 0.2 \times 0.7) \times 0.3 \\
&\quad + 0.9 \times 0.97 \times (0.7 \times 0.2 \times 0.3 + 0.5 \times 0.2 \times 0.7) \times 0.7 \\
&= 0.0906
\end{aligned}$$

Similarly,

$$
\begin{aligned}
P(S,T_A|E^C) = {} & P(S|D_A^C,D_B^C)P(T_A|D_A^C,D_B^C)P(D_A^C,D_B^C|E^C) \\
& + P(S|D_A^C,D_B)P(T_A|D_A^C,D_B)P(D_A^C,D_B|E^C) \\
& + P(S|D_A,D_B^C)P(T_A|D_A,D_B^C)P(D_A,D_B^C|E^C) \\
& + P(S|D_A,D_B)P(T_A|D_A,D_B)P(D_A,D_B|E^C) \\
= {} & P(S|D_A^C,D_B^C)P(T_A|D_A^C,D_B^C)\{P(D_A^C|E^CR)P(E^CR) + P(D_A^C|E^CR^C)P(E^CR^C)\}P(D_B^C|E^C) \\
& + P(S|D_A^C,D_B)P(T_A|D_A^C,D_B)\{P(D_A^C|E^CR)P(E^CR) \\
& + P(D_A^C|E^CR^C)P(E^CR^C)\}P(D_B|E^C) \\
& + P(S|D_A,D_B^C)P(T_A|D_A,D_B^C)\{P(D_A|E^CR)P(E^CR) \\
& + P(D_A|E^CR^C)P(E^CR^C)\}P(D_B^C|E^C) \\
& + P(S|D_A,D_B)P(T_A|D_A,D_B)\{P(D_A|E^CR)P(E^CR) \\
& + P(D_A|E^CR^C)P(E^CR^C)\}P(D_B|E^C) \\
= {} & 0.05 \times 0.08 \times (0.4 \times 0.8 \times 0.3 + 0.9 \times 0.8 \times 0.7) \times 0.8 \\
& + 0.4 \times 0.2 \times (0.4 \times 0.8 \times 0.3 + 0.9 \times 0.8 \times 0.7) \times 0.2 \\
& + 0.6 \times 0.85 \times (0.6 \times 0.8 \times 0.3 + 0.1 \times 0.8 \times 0.7) \times 0.8 \\
& + 0.9 \times 0.97 \times (0.6 \times 0.8 \times 0.3 + 0.1 \times 0.8 \times 0.7) \times 0.2 \\
= {} & 0.12804
\end{aligned}
$$

Thus,

$$
\begin{aligned}
P(E|S,T_A) &= \frac{P(S,T_A|E) \times P(E)}{P(S,T_A|E) \times P(E) + P(S,T_A|E^C) \times P(E^C)} \\
&= \frac{0.0906 \times 0.2}{0.0906 \times 0.2 + 0.12804 \times 0.8} = 0.15
\end{aligned}
$$

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,**

$$\boldsymbol{T_i \sim \mathcal{E}xp(\lambda), f(t \mid \lambda) = \lambda e^{-\lambda t}, t > 0, \lambda > 0}$$

Assume exponential prior on λ,

$$\boldsymbol{\lambda \sim \mathcal{E}xp(2), \pi(\lambda) = 2e^{-2\lambda}, \lambda > 0}$$

(a) Find the posterior distribution of $\boldsymbol{\lambda}$.
(b) Find the Bayes estimator for $\boldsymbol{\lambda}$.
(c) Find the MAP estimator for $\boldsymbol{\lambda}$.
(d) Numerically find 95% equitailed confidence interval for $\boldsymbol{\lambda}$.
(e) Find the posterior probability of hypothesis $\boldsymbol{H_0 : \lambda \le 1/2}$.

(a)
$S = \sum_{i=1}^{n} T_i$ is gamma $\mathcal{G}a(n,\lambda)$
Likelihood,

$$f(s|\lambda) = \frac{s^{n-1}\lambda^n}{\Gamma(n)} e^{-\lambda s}, s > 0, \lambda > 0$$

$$s = \sum_{i=1}^{n} t_i = 3$$

Prior exponential $\mathcal{E}xp(\beta)$,

$$\lambda \sim \beta e^{-\beta\lambda}$$

Then posterior is gamma $\mathcal{G}a(n+1, s+\beta) = \mathcal{G}a(4,5)$ where $\beta = 2, n = 3$

$$\pi(\lambda \mid s) = \frac{(s+\beta)^{n+1}\lambda^n}{\Gamma(n+1)} e^{-(s+\beta)\lambda} = \frac{5^4\lambda^3}{\Gamma(4)} e^{-5\lambda}, \lambda > 0$$

(b)
Bayes estimator is posterior mean,

$$\widehat{\lambda_B} = \frac{n+1}{s+\beta} = \frac{4}{5}$$

(c)
MAP estimator is posterior mode,

$$MAP = \frac{n}{s+\beta} = \frac{3}{5}$$

(d)
95% credible sets

$$[0.218, 1.7535]$$

Matlab code:

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

(e)

$$H_0: \lambda \leq \frac{1}{2}$$
$$P(H_0) = 0.2424$$

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.**

High protein	Low protein
134	70
146	118
104	101
119	85
124	107
161	132
107	94
83	
113	
129	
97	
123	

We want to test the hypothesis on dietary effect. Did a low protein diet result in significantly lower weight gain?
The classical *t* test against one sided alternative will be significant. We will do the test Bayesian way using Gibbs sampler.

Assume that high-protein diet measurements $y_{1i}, i = 1, \ldots, 12$ are coming from normal distribution $N(\theta_1, 1/\tau_1)$, where τ_1 is precision parameter,

$$f(y_{1i} \mid \theta_1, \tau_1) \propto \tau_1^{1/2} \exp\left\{-\frac{\tau_1}{2}(y_{1i} - \theta_1)^2\right\}, i = 1, \ldots, 12$$

Low-protein diet measurements $y_{2i}, i = 1, \ldots, 7$ are coming from normal distribution $N(\theta_2, 1/\tau_2)$,

$$f(y_{2i} \mid \theta_2, \tau_2) \propto \tau_2^{1/2} \exp\left\{-\frac{\tau_2}{2}(y_{2i} - \theta_2)^2\right\}, i = 1, \ldots, 7$$

Assume that θ_1 and θ_2 have normal priors $N(\theta_{10}, 1/\tau_{10})$ and $N(\theta_{20}, 1/\tau_{20})$, respectively. Take prior means as $\theta_{10} = \theta_{20} = 110$ (apriori no preference) and precisions as $\tau_{10} = \tau_{20} = 1/100$

Assume that τ_1 and τ_2 have the gamma $\mathcal{G}a(a_1, b_2)$ and $\mathcal{G}a(a_2, b_2)$ priors with shapes $a_1 = a_2 = 0.01$and rates $b_1 = b_2 = 4$.

(a) Construct Gibbs sampler that will sample $\theta 1, \tau 1, \theta 2, and\ \tau 2$ from their posteriors.

(b) Find sample differences $\theta 1 - \theta 2$. Proportion of positive differences approximates posterior probability of hypothesis $H0 : \theta 1 > \theta 2$. What is this proportion?

(c) Using sample quantiles find the 95% equitailed credible set for $\theta 1 - \theta 2$. Does this set contain 0?

(a)

Joint distributions,

$$f(y_1 i, \theta_1, \tau_1) \propto \tau_1^{n/2} \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{n} (y_{1i} - \theta_1)^2\right\} * \tau_{10}^{1/2} \exp\left\{-\frac{\tau_{10}}{2}(\theta_1 - \theta_{10})^2\right\}$$
$$* \tau_1^{a_1 - 1} \exp\{-b_1 \tau_1\}$$

$$f(y_2 i, \theta_2, \tau_2) \propto \tau_2^{n/2} \exp\left\{-\frac{\tau_2}{2}\sum_{i=1}^{n} (y_{2i} - \theta_2)^2\right\} * \tau_{20}^{1/2} \exp\left\{-\frac{\tau_{20}}{2}(\theta_2 - \theta_{20})^2\right\}$$
$$* \tau_2^{a_2 - 1} \exp\{-b_2 \tau_2\}$$

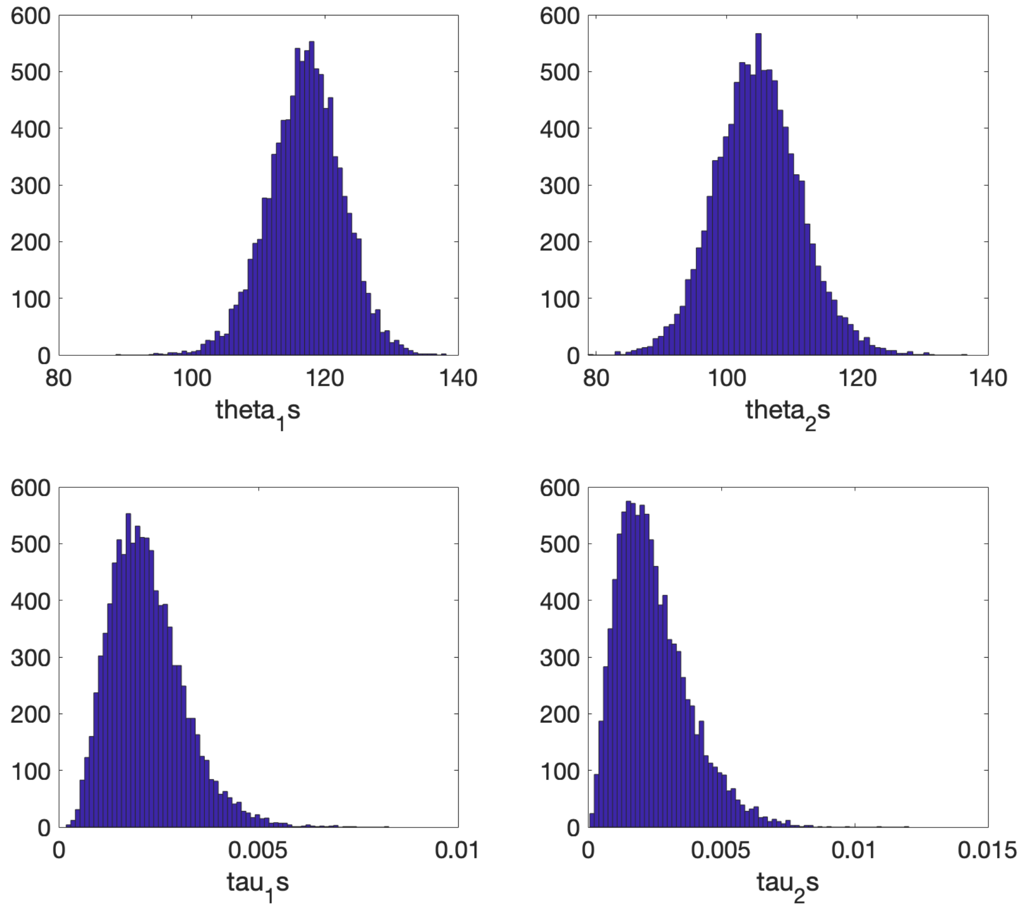
Thus, conditional distributions

$$\pi(\theta_1 \mid \tau_1, y_{1i}) \propto \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{n_1} (y_{1i} - \theta_1)^2\right\} \exp\left\{-\frac{\tau_{10}}{2}(\theta_1 - \theta_{10})^2\right\}$$
$$\propto \exp\left\{-\frac{1}{2}(\tau_{10} + n_1\tau_1)\left(\theta_1 - \frac{\tau_1 \sum_{i=1}^{n_1} y_{1i} + \tau_{10}\theta_{10}}{\tau_{10} + n_1\tau_1}\right)^2\right\}$$
$$\sim \mathcal{N}\left(\frac{\tau_1 \sum_{i=1}^{n_1} y_{1i} + \tau_{10}\theta_{10}}{\tau_{10} + n_1\tau_1}, \frac{1}{\tau_{10} + n_1\tau_1}\right)$$

$$\pi(\theta_2 \mid \tau_2, y_{2i}) \propto \exp\left\{-\frac{\tau_2}{2}\sum_{i=1}^{n_2} (y_{2i} - \theta_2)^2\right\} \exp\left\{-\frac{\tau_{20}}{2}(\theta_2 - \theta_{20})^2\right\}$$
$$\propto \exp\left\{-\frac{1}{2}(\tau_{20} + n_2\tau_2)\left(\theta_2 - \frac{\tau_2 \sum_{i=1}^{n_2} y_{2i} + \tau_{20}\theta_{20}}{\tau_{20} + n_2\tau_2}\right)^2\right\}$$
$$\sim \mathcal{N}\left(\frac{\tau_2 \sum_{i=1}^{n_2} y_{2i} + \tau_{20}\theta_{20}}{\tau_{20} + n_2\tau_2}, \frac{1}{\tau_{20} + n_2\tau_2}\right)$$

$$\pi(\tau_1 \mid \theta_1, y_{1i}) \propto \tau_1^{\frac{n_1}{2}} \tau_1^{a_1-1} \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{n_1}\left(y_{1,i}-\theta_1\right)^2\right\} \exp\{-b_1\tau_1\}$$
$$\sim \mathcal{G}a\left(a_1 + n_1/2, b_1 + \frac{1}{2}\sum_{i=1}^{n_1}(y_{1i}-\theta_1)^2\right)$$
$$\pi(\tau_2 \mid \theta_2, y_{2i}) \propto \tau_2^{\frac{n_2}{2}} \tau_2^{a_2-1} \exp\left\{-\frac{\tau_2}{2}\sum_{i=1}^{n_2}\left(y_{2,i}-\theta_2\right)^2\right\} \exp\{-b_2\tau_2\}$$
$$\sim \mathcal{G}a\left(a_2 + n_2/2, b_2 + \frac{1}{2}\sum_{i=1}^{n_2}(y_{2i}-\theta_2)^2\right)$$

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



(b)
Sample differences is 12.3150 .
Proportion of positive differences $H_0\colon \theta_1 > \theta_2 = 0.9192$
This is the proportion in the samples that $\theta_1 > \theta_2$

(c)
95% equtailed credible set

$$[-4.9559, 28.9730]$$

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