Name: Qian Liu **ID 443089**

Problem 1

(a) $\bar{X} = \frac{1}{n}\sum_{i=1}^{100} X_i$, X_i are iid for all $i \in (0, \cdots 100)$

$X_i | \theta \sim N(\theta, 50)$

$\bar{X}$ is also normal distributed, $\bar{X} \sim N(\theta, \cancel{\frac{50}{100}}\ 0.5)$

So

$$\frac{\bar{X} - \theta}{\sqrt{0.5}} \sim N(0, 1)$$

$$\theta \in [\bar{X} - N_{(0,1)}^{0.975} \cdot \sqrt{0.5}, \bar{X} + N_{(0,1)}^{0.025} \cdot \sqrt{0.5}]$$

```
> prosample = sort(Michelson$velocity)
> model_var = 50;
> n = length(prosample);
>
> mle = function(y)
+ {
+   a = 0;
+   for(i in prosample){
+     a = a + log(dnorm(x = (i-y)/sqrt(model_var)),0.5)
+   }
+   return (a)
+ }
>
> mle_mean = optim(par=800, fn = mle)$par ; mle_mean
 [1] 852.4023
>
> sample_mean = mean(prosample);
> a = sample_mean - qnorm(0.975)*sqrt(model_var/n);
> b = sample_mean + qnorm(0.975)*sqrt(model_var/n);
> c = c(a,b); c
[1] 851.0141 853.7859
```

Name: Qian Liu ID 443089

(b) ~~P(θ)~~ = Prior $\theta \sim N(\mu_0, \sigma_0^2)$

model $X_i|\theta \sim N(\theta, \sigma^2)$

Posterior distribution $D = \{x_1 \cdots x_n\}$

$$P(\theta|D) = P(D|\theta)P(\theta)/P(D)$$

$$\underset{\theta}{\propto} P(D|\theta)P(\theta)$$

$$= P(\theta)\prod_{i=1}^{n} P(X_i|\theta)$$

$$\propto \exp\left(-\frac{1}{2\sigma_0^2}(\theta-\mu_0)^2\right)\prod_{i=1}^{n}\exp\left(-\frac{1}{2\sigma^2}(X_i-\theta)^2\right)$$

$$= \exp\left(-\frac{1}{2\sigma_0^2}(\theta-\mu_0)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\cancel{\theta}\theta)^2\right)$$

$$= \exp\left(\left(-\frac{1}{2\sigma_0^2}-\frac{n}{2\sigma^2}\right)\theta^2 + \left(\frac{\mu_0}{\sigma_0^2}+\frac{1}{\sigma^2}\sum X_i\right)\theta + c\right)$$

$$= \exp\left(\left(-\frac{1}{2\sigma_0^2}-\frac{n}{2\sigma^2}\right)\theta^2 - 2\frac{\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum X_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}\theta + m^2\right) + c'$$

$$= \exp\left(-\frac{1}{2}\cdot\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)(\theta-m)^2 + c'\right)$$

So: $P(\theta|D) \propto \exp\left(-\frac{1}{2\sigma_n^2}(\theta-m)^2\right)$

$\theta|D \sim N(m, \sigma_n^2)$

where $m = \cancel{\sigma_n}\ \sigma_n^2\cdot\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum X_i}{\sigma^2}\right)$

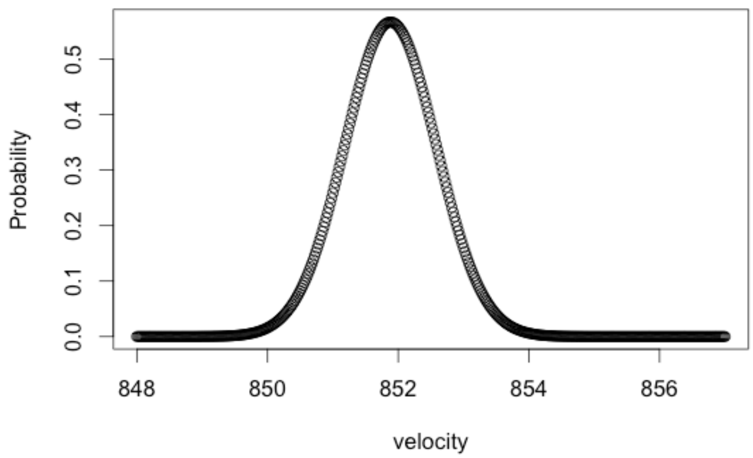
$$\sigma_n^2 = \frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}$$

As the posterior distribution is normal distribution. The equal-tailed is also narrowest.

```
> pri_var = 50;
> pri_mean = 800;
> post_var = 1/((1/pri_var)+(n/ model_var));
> post_mean = ((pri_mean/pri_var)+sum(prosample/model_var))*post_var;
> a = qnorm(0.975,mean = post_mean, sd = sqrt(post_var))
> b = qnorm(0.025,mean = post_mean, sd = sqrt(post_var))
> c = c(b,a); c
[1] 850.5022 853.2602
```

Name: Qian Liu   ID 443089

```
> sequ = seq(848,857,0.02); θ
> d = dnorm(sequ,mean = post_mean, sd = sqrt(post_var));
> plot(x= sequ , y = d, xlab = "velocity", ylab = "Probability")
```




```
> Min_inter = function(x)
+ {
+    a = qnorm(x, mean = post_mean, sd = sqrt(post_var))
+    b = qnorm(0.95+x, mean = post_mean, sd = sqrt(post_var))
+    m = abs(b-a);
+    return (m);
+ }
>
> a = optim(par= 0.01, fn = Min_inter)$par; a
[1]0.025
> b = 0.95+a;
> a = qnorm(a,mean = post_mean, sd = sqrt(post_var))
> b = qnorm(b,mean = post_mean, sd = sqrt(post_var))
> c = c(a,b); c
[1] 850.5022 853.2602
```

The posterior distribution is normal distribution. The equal-tailed is also narrowest.

Name: Qian Liu   ID 443089

Problem 2

(a) For prior $\lambda \sim gamma(\alpha, \beta)$

$$p(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \cdot e^{-\beta\lambda}, \quad \lambda > 0$$

where $\Gamma(\alpha) = \int_0^{\infty} \lambda^{\alpha-1} e^{-\lambda} d\lambda$.

Posterior distribution

$$p(\lambda \mid x^{(n)}) = \frac{\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \cdot \lambda^{n} e^{-\lambda\sum_{i=1}^{n} x_i}}{\int \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \lambda^{n} e^{-\lambda\sum_{i=1}^{n} x_i} d\lambda}$$

$$= \frac{(\beta + \sum_{i=1}^{n} x_i)^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{\alpha+n-1} \exp\{-\lambda(\beta + \sum_{i=1}^{n} x_i)\}$$

Set $\alpha = 3$, $\beta = 1$

then $\lambda \mid x^{(n)} \sim gamma(3+n, 1+\sum_{i=1}^{n} x_i)$

(b) For prior $p(\lambda) = 1$

$$p(\lambda \mid x^{(n)}) = p(x^{n} \mid \lambda) \cdot p(\lambda) / p(x^{n})$$

$$\propto (\sum x_i)^{n+1} \lambda^{n} \cdot e^{-\lambda \sum x_i}$$

$\Rightarrow \lambda \mid x^{(n)} \sim gamma(1+n, \sum_{i=1}^{n} x_i)$

(c) for parts (a), $\hat{\lambda}_{mean} = \dfrac{\alpha + n}{\beta + \sum_{i=1}^{n} x_i}$

parts (b) $\hat{\lambda}_{mean} = \dfrac{1+n}{\sum_{i=1}^{n} x_i}$

In part(a), alpha = 3 and beta = 1, mle_mean =3.287695 post_mean_a = 3.26995 ;
In part(b), post_mean_b = 3.353454;

Name: Qian Liu ID 443089

```
> prosample = rexp(50, 3)
> alpha = 3;
> beta = 1;
> mle = function(y)
+ {
+   a=0;
+   for(i in prosample){
+     a = a + log(dexp(x = i, rate = y ),0.5)
+   }
+   return (a)
+ }
>
> mle_mean = optim(par=6, fn = mle)$par; mle_mean
 [1] 3.287695
> post_mean_a = (alpha+length(prosample))/(beta+sum(prosample));post_mean_a
[1] 3.26995
> post_mean_b = (1+length(prosample))/sum(prosample);post_mean_b
[1] 3.353454
```

(D)

In part (a), the interval is [2.449417, 4.207196], range = 1.75778
part (b), the interval is [2.496867, 4.334423], range = 1.83755
The interval in part a is narrower.

```
> a_1 = qgamma(0.975, shape = alpha+length(prosample),scale = 1/(beta+sum(prosample)))
> a_2 = qgamma(0.025, shape = alpha+length(prosample),scale = 1/(beta+sum(prosample)))
> c1 = c(a_2,a_1);c1
[1] 2.449417 4.207196
> b_1 = qgamma(0.975, shape = 1+length(prosample),scale = 1/sum(prosample));
> b_2 = qgamma(0.025, shape = 1+length(prosample),scale = 1/sum(prosample));
> c2 = c(b_2,b_1);c2
[1] 2.496867 4.334423
> > d1 = a_1-a_2;d1
[1] 1.75778
> d2 = b_1-b_2;d2
[1] 1.837556
> d = d1 - d2;d
 [1] -0.07977615
```

Name: Qian Liu ID 443089

Problem 3

The model

$$P(x|\theta) = \binom{n}{x} \cdot \theta^{x} \cdot (1-\theta)^{n-x}$$

for prior $P(\theta) = \dfrac{\theta^{\alpha}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, $\alpha > 1$, $\beta > 1$

The posterior distribution

$$P(\theta|X) \propto P(\theta)P(x|\theta)$$
$$\propto \theta^{\alpha+x-1}(1-\theta)^{n+\beta-x-1}$$

$\theta|x \sim Beta(\alpha+x,\ n+\beta-x)$, $\alpha>1, \beta>1$

A reasonable approximation of the value of the median of ~~a~~ beta distribution, for both $\alpha+x$, and $n+\beta-x$ greater or equal to one, is given by the formula.

$$\delta(x) = \text{median } \theta \simeq \frac{\alpha+x-\frac{1}{3}}{\alpha+x+n+\beta-x-\frac{2}{3}} = \frac{\alpha+x-\frac{1}{3}}{\alpha+\beta+n-\frac{2}{3}}$$

The risk function (same as MSE under squared error loss) is

$$R(\theta,\delta(x)) = MSE[\delta(x)]$$
$$= Var[\delta(x)] + Bias^2[\delta(x)]$$
$$= \frac{n\theta(1-\theta)}{(\alpha+\beta+n-\frac{2}{3})^2} + \left(\frac{\alpha+n\theta-\frac{1}{3}}{\alpha+\beta+n-\frac{2}{3}}-\theta\right)^2$$
$$= \frac{(n\theta-n\theta^2+(\alpha-\frac{1}{3}-\alpha\theta-\beta\theta+\frac{2}{3}\theta)^2)}{(\alpha+\beta+n-\frac{2}{3})^2}$$

Problem 4

Name: Qian Liu ID 443089

Frequentists believe that scientific inference is not inductive but deductive. Their approach to statistics is also associated with the philosophy and follows Popper's doctrine of falsification. In contrast, Bayesian inference is associated with inductive reasoning and the idea that a model can be replaced by better ones but can never be directly falsified by a significance test. Bayesians may apply 'Bayesian Model Averaging' to minimize the risk, but the 'poor model' is still involved in the model averaging process. For example, frequentists deduct central limit theory and generate some statistic model and test model based on that. For example, when σ is known, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d}$ N(0,1) where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$, E[X] = $\mu, Var[X] = \sigma^2$; And when when σ is unknown, $\frac{\bar{X}-\mu}{s/\sqrt{n}} \xrightarrow{d} T_{(n-1)}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$, E[X] = μ, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} Xi^2 - \bar{X}^2$; They falsify other models and only believe in the model they have deducted. However, Bayesian inference is different, they might create a model Class = {m1,m2,m3…, m_n} , they can define each model and train those models by optimizing those parameters with some training data set. The model class after training can be, for example, m1 : $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d}$ N(0,1); m2: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d}$ Poisson(1) ; m3: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d}$ gamma(1,1);m4: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d}$ beta(1,1); …..; The probability of each model might be like P (m1) = 0.5; P(m2) = 0.1; P(m3)=0.01;…; and $\sum_{i=1}^{n} \mathrm{p(mi)}$ = 1; Then, they average all those models with respect to their weights. Given some data set, 'Bayesian Model Averaging' may work well while the model deducted by frequentists doesn't fit. However, as there is no perfect model that can fit all problems, so it is also possible that BMA does poor performance on other data set while frequentists' model may work well.

Name: Qian Liu **ID 443089**

Appendix

R code for problem 1:

```
prosample = sort(Michelson$velocity)
model_var = 50;
n = length(prosample);

mle = function(y)
{
  a = 0;
  for(i in prosample){
    a = a + log(dnorm(x = (i-y)/sqrt(model_var)),0.5)
  }
  return (a)
}

mle_mean = optim(par=800, fn = mle)$par ; mle_mean

sample_mean = mean(prosample);
a = sample_mean - qnorm(0.025)*sqrt(model_var/n);
b = sample_mean + qnorm(0.025)*sqrt(model_var/n);
c = c(a,b); c

pri_var = 50;
pri_mean = 800;
post_var = 1/((1/pri_var)+(n/ model_var));
post_mean = ((pri_mean/pri_var)+sum(prosample/model_var))*post_var;

a = qnorm(0.975,mean = post_mean, sd = sqrt(post_var))
b = qnorm(0.025,mean = post_mean, sd = sqrt(post_var))
c = c(b,a); c

sequ = seq(848,857,0.02);
d = dnorm(sequ,mean = post_mean, sd = sqrt(post_var));
plot(x= sequ , y = d, xlab = "velocity", ylab = "Probability")

a = qnorm(0.975,mean = post_mean, sd = sqrt(post_var))
b = qnorm(0.025,mean = post_mean, sd = sqrt(post_var))
c = c(b,a); c

Min_inter = function(x)
{
    a = qnorm(x, mean = post_mean, sd = sqrt(post_var))
    b = qnorm(0.95+x, mean = post_mean, sd = sqrt(post_var))
    m = abs(b-a);
    return (m);
}

a = optim(par= 0.01, fn = Min_inter)$par;
b = 0.95+a;
a = qnorm(a,mean = post_mean, sd = sqrt(post_var))
```

Name: Qian Liu   ID 443089

```
b = qnorm(b,mean = post_mean, sd = sqrt(post_var))
c = c(a,b); c
```

R code for problem 2:

```
prosample = rexp(50, 3)
alpha = 3;
beta = 1
mle = function(y)
{
  a=0;
  for(i in prosample){
    a = a + log(dexp(x = i, rate = y ),0.5)
  }
  return (a)
}

mle_mean = optim(par=6, fn = mle)$par; mle_mean
post_mean_a = (alpha+length(prosample))/(beta+sum(prosample));post_mean_a
post_mean_b = (1+length(prosample))/sum(prosample);post_mean_b

a_1 = qgamma(0.975, shape = alpha+length(prosample),scale = 1/(beta+sum(prosample)))
a_2 = qgamma(0.025, shape = alpha+length(prosample),scale = 1/(beta+sum(prosample)))
c1 = c(a_2,a_1);c1


b_1 = qgamma(0.975, shape = 1+length(prosample),scale = 1/sum(prosample));
b_2 = qgamma(0.025, shape = 1+length(prosample),scale = 1/sum(prosample));
c2 = c(b_2,b_1);c2
d1 = a_1-a_2;d1
d2 = b_1-b_2;d2
d = d1 - d2;d
```