

# STA 602. HW04

Yicheng Shen

9/26/2022

## 1.

(a) It can be shown that the Galenshore distribution can be viewed as an exponential family model.

$$\begin{aligned} p(y | \theta) &= \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2} \\ \text{Rearrange terms} &= \left( \frac{2}{\Gamma(a)} y^{2a-1} \right) \times (\theta^{2a}) \times (e^{-\theta^2 y^2}) \\ &= h(y) c(\phi) e^{\phi t(y)} \end{aligned}$$

which can be specified into

$$\begin{aligned} h(y) &= \frac{2}{\Gamma(a)} y^{2a-1} \\ \phi &= -\theta^2 \\ c(\phi) &= (-\phi)^a \\ t(y) &= y^2 \end{aligned}$$

We can easily get  $d\phi/d\theta = -2\theta$  by the change of variables.

$$\begin{aligned} p(\phi | n_0, t_0) &= \kappa(n_0, t_0) c(\phi)^{n_0} e^{n_0 t_0 \phi} \\ p(\theta | n_0, t_0) &= p(\phi | n_0, t_0) \times \left| \frac{d\phi}{d\theta} \right| \\ &\propto \kappa(n_0, t_0) \theta^{2an_0} e^{n_0 t_0 - \theta^2} \times (-2\theta) \end{aligned}$$

$$\text{Remove constant terms} \propto \theta^{2an_0+1} e^{n_0 t_0 - \theta^2}$$

$$\text{We know that } \text{dgalenshore}(y|a, \theta) \propto y^{2a-1} e^{-\theta^2 y^2}$$

$$\text{Recognize the kernel} \quad 2a' - 1 = 2an_0 + 1$$

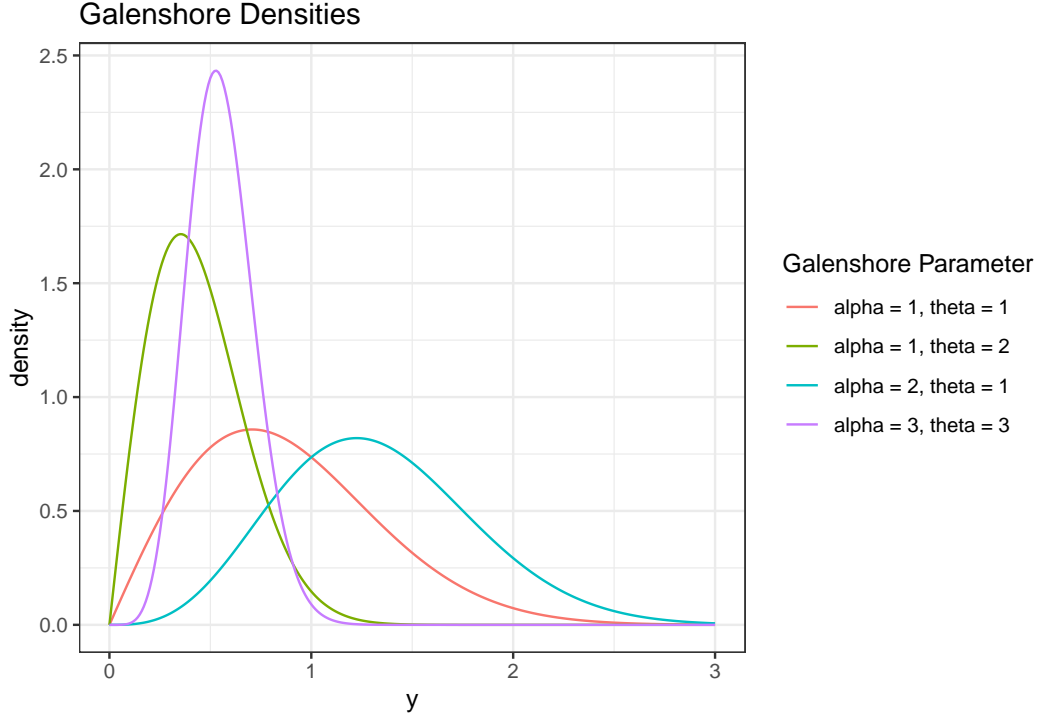
$$-\theta'^2 = -n_0 t_0$$

$$p(\theta | n_0, t_0) \propto \text{dgalenshore}(a', \theta') = \text{dgalenshore}(an_0 + 1, \sqrt{n_0 t_0})$$

Thus the galenshore can be chosen as the conjugate prior.

```
dgalenshore = function(y, a, theta) {  
  (2 / gamma(a)) * theta^(2 * a) * y ^ (2 * a - 1) * exp(-(theta^2) * y^2)  
}  
y = seq(0, 3, by = 0.01)  
sim_dat = rbind(  
  data.frame(y = y, density = dgalenshore(y, 1, 1), dist = 'alpha = 1, theta = 1'),
```

```
data.frame(y = y, density = dgalenshore(y, 1, 2), dist = 'alpha = 1, theta = 2'),
data.frame(y = y, density = dgalenshore(y, 2, 1), dist = 'alpha = 2, theta = 1'),
data.frame(y = y, density = dgalenshore(y, 3, 3), dist = 'alpha = 3, theta = 3')
)
ggplot(sim_dat, aes(x = y, y = density, group = dist, color = dist)) +
  geom_line() + labs(title = "Galenshore Densities", color = "Galenshore Parameter")
```



(b) It is already proved that this belongs to exponential family, the posterior of  $\phi$  is thus given by

$$p(\phi \mid y_1, \dots, y_n) \propto p(\phi \mid n_0 + n, n_0 t_0 + n \bar{t}(\mathbf{y}))$$

We can therefore modify the expression

$$\begin{aligned} p(\theta \mid n_0, t_0) &\propto \text{dgalenshore}(a', \theta') = \text{dgalenshore}(an_0 + 1, \sqrt{n_0 t_0}) \\ p(\theta \mid Y_1, \dots, Y_n) &\propto p(\theta \mid n_0 + n, n_0 t_0 + n \bar{t}(\mathbf{y})) \\ &\propto \text{dgalenshore}\left(a(n_0 + n) + 1, \sqrt{n_0 t_0 + n \bar{t}(\mathbf{y})}\right) \\ &\propto \text{dgalenshore}\left(a' + an, \sqrt{\theta'^2 + n \bar{t}(\mathbf{y})}\right) \\ &= \text{dgalenshore}\left(a' + an, \sqrt{\theta'^2 + \sum_{i=1}^n y_i^2}\right) \end{aligned}$$

(c) Since we now know that the posterior of  $\theta$  would still be Galenshore distribution, we can simplify it

into

$$p(\theta \mid Y_1, \dots, Y_n) = \text{dgalenshore} \left( a' + an, \sqrt{\theta'^2 + \sum_{i=1}^n y_i^2} \right)$$

$$\frac{p(\theta_a \mid Y_1, \dots, Y_n)}{p(\theta_b \mid Y_1, \dots, Y_n)} \propto \frac{\theta_a^{2(a'+an)-1} e^{-\theta_a^2(\theta'^2 + \sum_{i=1}^n y_i^2)}}{\theta_b^{2(a'+an)-1} e^{-\theta_b^2(\theta'^2 + \sum_{i=1}^n y_i^2)}}$$

$$\propto (\theta_a/\theta_b)^{2(a'+an)-1} e^{-(\theta'^2 + \sum_{i=1}^n y_i^2)(\theta_a^2 - \theta_b^2)}$$

Hence,  $\sum_{i=1}^n y_i^2$  is our sufficient statistic.

(d) From the formula for the expectation of a Galenshore distribution

$$p(\theta \mid Y_1, \dots, Y_n) = \text{dgalenshore} \left( a' + an, \sqrt{\theta'^2 + \sum_{i=1}^n y_i^2} \right)$$

$$\mathbb{E}(\theta \mid y_1, \dots, y_n) = \frac{\Gamma(a' + 1/2)}{\theta' \Gamma(a')}$$

$$= \frac{\Gamma(a' + an + 1/2)}{\sqrt{\theta'^2 + \sum_{i=1}^n y_i^2} \times \Gamma(a' + an)}$$

(e) The PPC is given as

$$p(\tilde{y} \mid y_1 \dots y_n) = \int_0^\infty p(\tilde{y} \mid \theta) p(\theta \mid y_i) d\theta$$

$$= \int_0^\infty \frac{2}{\Gamma(a)} \theta^{2a} \tilde{y}^{2a-1} e^{-\theta^2 \tilde{y}^2} \frac{2}{\Gamma(a' + an)} (\theta'^2 + \sum_{i=1}^n y_i^2)^{(a'+an)} \theta^{2(a'+an)-1} e^{-(\theta'^2 + \sum_{i=1}^n y_i^2)\theta^2} d\theta$$

$$\text{Get constant terms out} = \frac{2}{\Gamma(a)\Gamma(a' + an)} \tilde{y}^{2a-1} (\theta'^2 + \sum_{i=1}^n y_i^2)^{(a'+an)} \int_0^\infty 2\theta^{2(a+a'+an)-1} e^{-\theta^2(\tilde{y}^2 + \theta'^2 + \sum_{i=1}^n y_i^2)} d\theta$$

$$= \frac{2\Gamma(a + a' + an)}{\Gamma(a)\Gamma(a' + an)} \frac{\tilde{y}^{2a-1} (\theta'^2 + \sum_{i=1}^n y_i^2)^{(a'+an)}}{(\tilde{y}^2 + \theta'^2 + \sum_{i=1}^n y_i^2)^{a+a'+an}}$$

## 2.

(a) The Bayes estimate under means square error loss is the posterior mean, which is  $\frac{10+n\theta}{1+n}$ .

$$\text{Bias} = E[\delta^* \mid \theta] - \theta = \frac{10 + n\theta}{1 + n} - \theta = \frac{10 - \theta}{1 + n}$$

$$\text{Var} = \text{Var}\left(\frac{10 + \sum X_i}{1 + n}\right) = \frac{n\theta}{(1 + n)^2}$$

$$\text{MSE} = \text{Bias}^2 + \text{Var}$$

$$= \left(\frac{10 - \theta}{1 + n}\right)^2 + \frac{n\theta}{(1 + n)^2}$$

$$= \frac{(10 - \theta)^2 + n\theta}{(1 + n)^2}$$

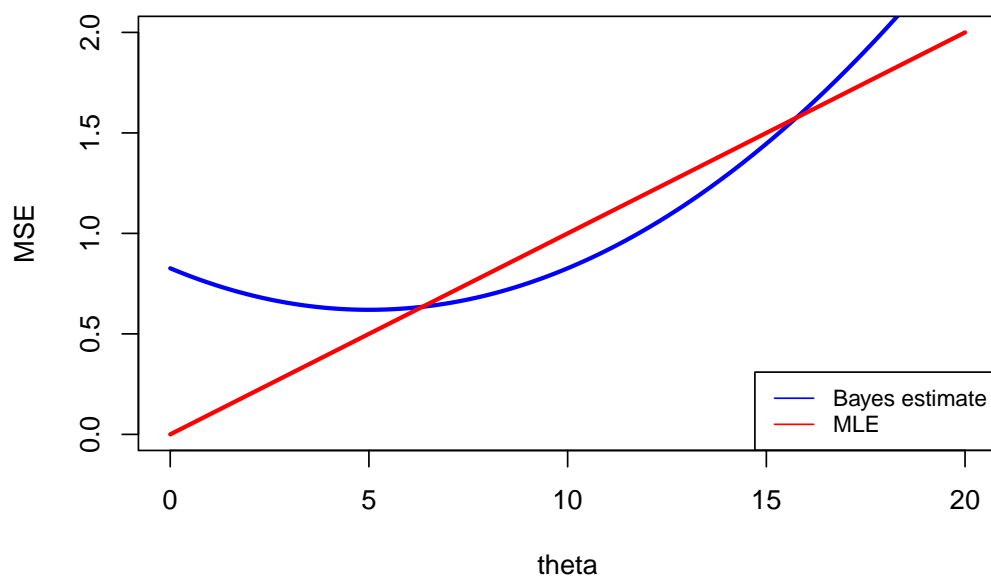
(b) The MLE,  $\bar{X}$ , has zero bias and its MSE is thus

$$\text{Var}\left(\frac{\sum X_i}{n}\right) = \frac{n\theta}{n^2} = \frac{\theta}{n}$$

The plot suggests that the Bayes estimate does better than MLE approximately during the range of  $\theta \in (6.35, 15.75)$ , in other range of  $\theta$ , the MLE has smaller MSE.

In this case, MLE represents the estimator using the sample mean (just the data without any prior guess) whereas Bayes estimate is a compromise between the sampled data and prior. For a  $\theta \sim \text{Gamma}(10, 1)$  prior, it would be a good prior guess if the true  $\theta$  is actually around 10. Therefore, we should expect to see our Bayes estimator to do better around  $\theta = 10$ . If the true  $\theta$  is far away from 10, then our poorly guessed prior negatively affects our Bayes estimates and MLE would do better instead.

```
n <- 10
theta <- seq(0, 20, by = 0.1)
MSE <- ((10 - theta) ^ 2 + n * theta) / ((1 + n) ^ 2)
MLE_Var <- theta / n
plot(theta, MSE, type = "l", col= "blue", ylim=c(0, 2), lwd = 2.5)
lines(theta, MLE_Var, type = "l", col= "red", lwd = 2.5)
legend(x = "bottomright", legend = c("Bayes estimate", "MLE"),
      col = c("blue", "red"), lty = 1, cex = 0.8)
```



### 3.

(a)

$$\begin{aligned}
 Bias &= E[\delta_5^*|\theta] - \theta = \frac{X + \sqrt{n}/2}{n + \sqrt{n}} - \theta \\
 &= \frac{n\theta + \sqrt{n}/2}{n + \sqrt{n}} - \theta \\
 &= \frac{\sqrt{n}(1/2 - \theta)}{n + \sqrt{n}} \\
 Var &= Var\left(\frac{X + \sqrt{n}/2}{n + \sqrt{n}}\right) \\
 &= \frac{n\theta(1 - \theta)}{(n + \sqrt{n})^2} \\
 MSE &= Bias^2 + Var \\
 &= \frac{n(1/2 - \theta)^2 + n\theta(1 - \theta)}{(n + \sqrt{n})^2} \\
 &= \frac{n[(1/2 - \theta)^2 + \theta(1 - \theta)]}{(n + \sqrt{n})^2}
 \end{aligned}$$

(b) Interestingly, the fifth minimax estimator behaves similar to  $\delta_4$  when  $n = 5$ . It does better than sample mean but worse than stubborn one or Bayes estimator with Beta (12,12) prior when  $\theta$  is close to 0.5. It does better than stubborn one and Bayes with Beta (12,12) prior when  $\theta$  is far from 0.5.

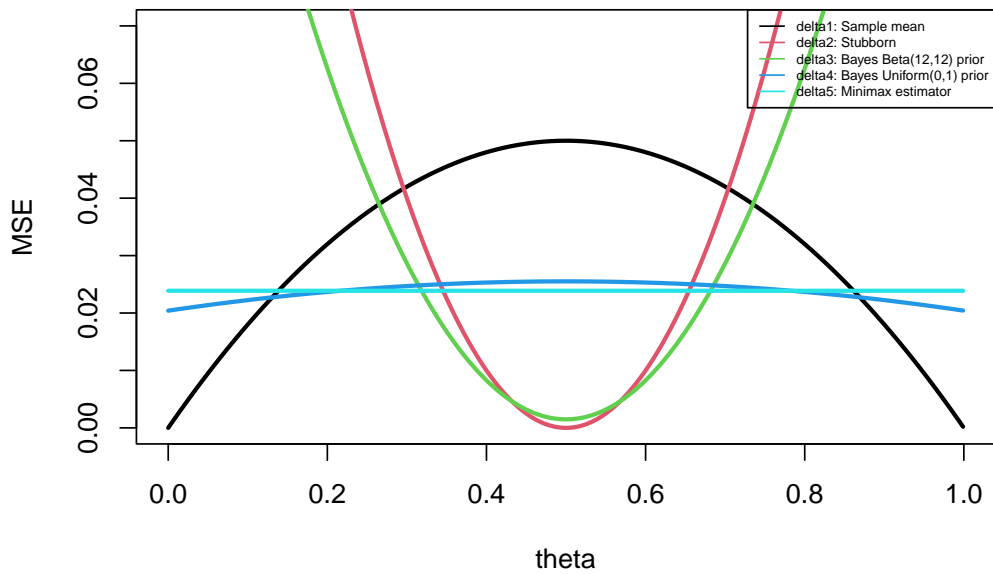
When  $n = 100$ ,  $\delta_5$  is similar to Bayes estimator with Beta (12,12) prior when  $\theta$  is around 0.5. If  $\theta$  is far from 0.5,  $\delta_5$  is much better than the Bayes estimator with Beta (12,12) prior.

This shows that  $\delta_5$  could often mitigate the disadvantages of using a Bayesian approach and making a strong but terrible prior guess.

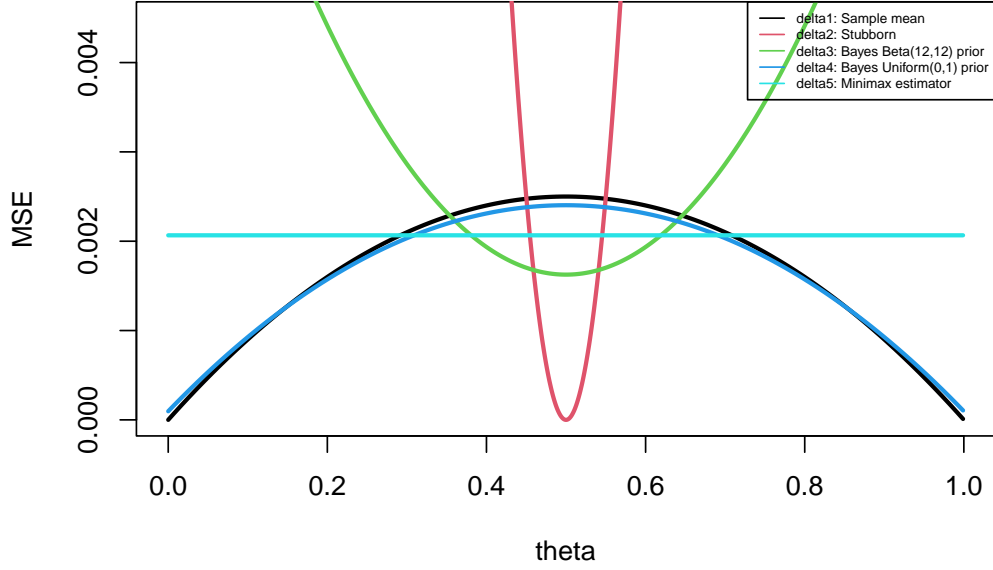
```

n <- 5
theta <- seq(0, 0.999, by = 0.001)
MSE <- theta * (1-theta) / n
MSE_2 <- 0.25 - theta * (1-theta)
MSE_3 <- (n * theta * (1 - theta) + 144*(1-2*theta) ^2) / (n+24)^2
MSE_4 <- (n * theta * (1 - theta) + (1-2*theta) ^2) / (n+2)^2
MSE_5 <- n * ((0.5 - theta) ^ 2 + theta * (1 - theta)) / (n + sqrt(n)) ^ 2
plot(theta, MSE, type = "l", col= 1, ylim = c(0,0.07), lwd = 2.5)
lines(theta, MSE_2, type = "l", col= 2, lwd = 2.5)
lines(theta, MSE_3, type = "l", col= 3, lwd = 2.5)
lines(theta, MSE_4, type = "l", col= 4, lwd = 2.5)
lines(theta, MSE_5, type = "l", col= 5, lwd = 2.5)
legend(x = "topright", legend=c("delta1: Sample mean",
                                "delta2: Stubborn",
                                "delta3: Bayes Beta(12,12) prior",
                                "delta4: Bayes Uniform(0,1) prior",
                                "delta5: Minimax estimator"),
      col = 1:5, lty = 1, cex = 0.5)

```



```
n <- 100
theta <- seq(0, 0.999, by = 0.001)
MSE <- theta * (1-theta) / n
MSE_2 <- 0.25 - theta * (1 - theta)
MSE_3 <- (n * theta * (1 - theta) + 144 * (1 - 2 * theta) ^ 2) / (n + 24) ^ 2
MSE_4 <- (n * theta * (1 - theta) + (1 - 2 * theta) ^ 2) / (n + 2) ^ 2
MSE_5 <- n * ((0.5 - theta) ^ 2 + theta * (1 - theta)) / (n + sqrt(n)) ^ 2
plot(theta, MSE, type = "l", col= 1, ylim = c(0,0.0045), lwd = 2.5)
lines(theta, MSE_2, type = "l", col= 2, lwd = 2.5)
lines(theta, MSE_3, type = "l", col= 3, lwd = 2.5)
lines(theta, MSE_4, type = "l", col= 4, lwd = 2.5)
lines(theta, MSE_5, type = "l", col= 5, lwd = 2.5)
legend(x = "topright", legend=c("delta1: Sample mean",
                                "delta2: Stubborn",
                                "delta3: Bayes Beta(12,12) prior",
                                "delta4: Bayes Uniform(0,1) prior",
                                "delta5: Minimax estimator"),
      col = 1:5, lty = 1, cex = 0.5)
```



(c) The weighted average expression is as the following:

$$\begin{aligned}
 \delta_5 &= \frac{X + \sqrt{n}/2}{n + \sqrt{n}} \\
 &= \frac{n}{n + \sqrt{n}} \frac{X}{n} + \frac{\sqrt{n}}{n + \sqrt{n}} \frac{1}{2} \\
 &= \frac{n}{n + \sqrt{n}} \delta_1 + \frac{\sqrt{n}}{n + \sqrt{n}} \delta_2
 \end{aligned}$$

The weights show that  $\delta_1$  always has larger weight than  $\delta_2$ , and the difference between the weights depends on the sample size,  $n$ . When  $n$  is as small as 5,  $\delta_1$  has a weight of about 0.7. When  $n$  is as large as 100, the weight for  $\delta_1$  grows to 0.9.

When sample size is small, the minimax estimator is a compromise between sample mean estimator and a fixed estimator of 0.5. When sample size is large, the minimax estimator behaves closer to the sample mean estimator (taking more information from the data). Most of the time the minimax estimator is able to obtain relatively low MSE.