STAT 638 HW 3

Jack Cunningham

3.8)

a)

In order to represent the long-run frequency we want to use a prior that has a strong weight.

So, let's choose $n = 1000$ where the weighted mixture of beta distributions is equivalent to 1000 observations worth of information.

For 20% of coins they behave symmetrically. So we would choose $a + b = 1000 =, a = 500, b = 500$. So:

$$\text{beta(500,500)}$$

For 40% of coins they give a frequency of heads around 1/3. So we would choose $a + b = 1000, a = 1000/3, b = 2000/3$. So:

$$\text{beta}(1000/3, 2000/3)$$

For the remaining 40% of coins they give a frequency of heads around 2/3. So we would choose $a + b = 1000, a = 2000/3, b = 1000/3$. So :

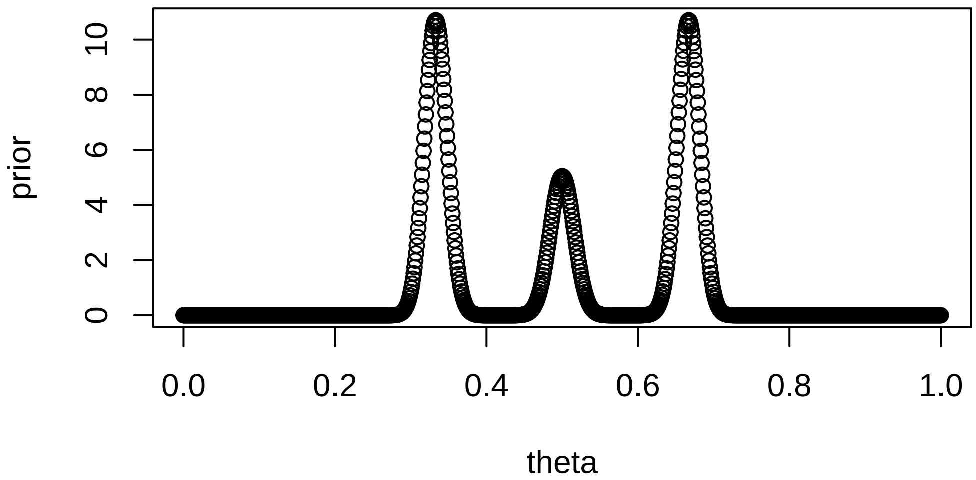
$$\text{beta(2000/3/1000)}$$

In total we have the prior distribution of:

$$\theta \sim \frac{1}{5}\text{beta}(500, 500) + \frac{2}{5}\text{beta}(1000/3, 2000/3) + \frac{2}{5}\text{beta(2000/3,1000/3)}$$

```
theta = seq(0,1,.001)
prior = 1/5*dbeta(theta, 500, 500) + 2/5*dbeta(theta, 1000/3,2000/3) +
```

```
  2/5*dbeta(theta, 2000/3,1000/3)
plot(theta,prior)
```



3.9)

a)

We can use the Galenshore (c, d) distribution as a prior for θ. So:

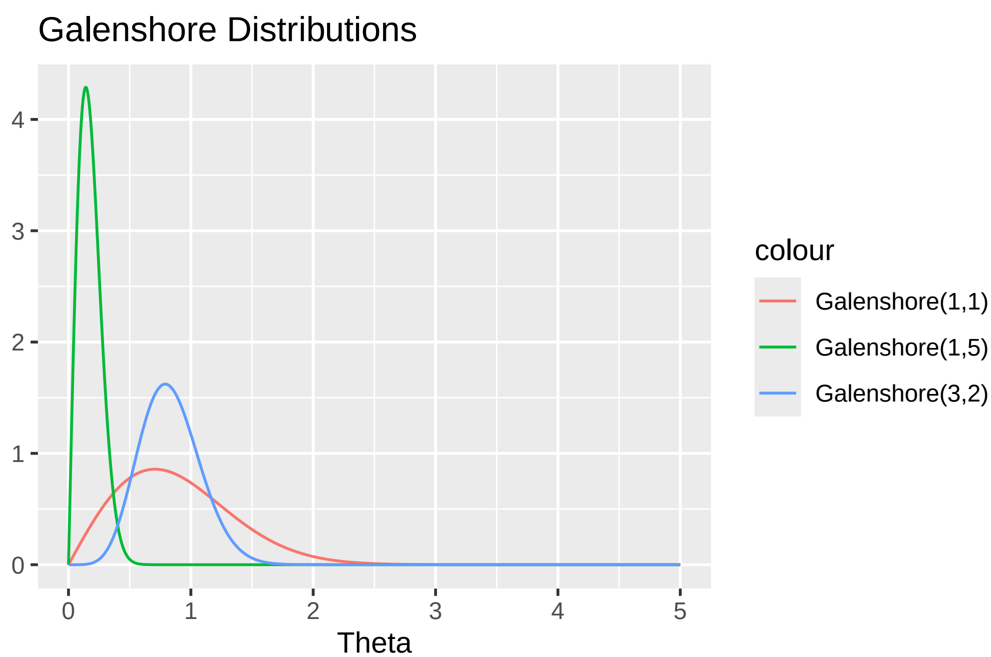
$$p(\theta) = \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-d^2\theta^2}$$

Defining the function:

```
galenshore = function(theta,c,d){
  return(2/gamma(c)*d^(2*c)*theta^(2*c-1)*exp(-d^2*theta^2) )
}
```

```
theta = seq(0,5,.001)
galenshore_1_1 = galenshore(theta,1,1)
```

```r
galenshore_3_2 = galenshore(theta,3,2)
galenshore_1_5 = galenshore(theta,1,5)
df = tibble(theta, galenshore_1_1,galenshore_3_2,galenshore_1_5)
df |>
  ggplot(aes(x = theta)) +
  geom_line(aes(y = galenshore_1_1, color = "Galenshore(1,1)")) +
  geom_line(aes(y = galenshore_1_5, color = "Galenshore(1,5)")) +
  geom_line(aes(y = galenshore_3_2, color = "Galenshore(3,2)")) +
  labs(x = "Theta", y = "", title = "Galenshore Distributions")
```



b)

We can show that:

$$p(y_1, ...y_n|\theta) = (\frac{2}{\Gamma(\alpha)})^2\theta^{2na}\prod_{i=1}^{n} y^{2a-1}e^{-\theta^2\sum_{i=1}^{n} y_i^2}$$

Then:

$$p(\theta|y_1, ...y_n) \propto p(\theta)p(y_1, ..., y_n|\theta) = \frac{2}{\Gamma(c)}d^{2c}\theta^{2c-1}e^{-d^2\theta^2}(\frac{2}{\Gamma(\alpha)})^2\theta^{2na}\prod_{i=1}^{n} y^{2a-1}e^{-\theta^2\sum_{i=1}^{n} y_i^2}$$

$$= \theta^{2c-1+2na} \prod_{i=1}^{n} y^{2a-1} e^{\theta^2(-d^2 \sum_{i=1}^{n} y_i)}$$

Removed some terms that would not appear to affect the proportion, leaving me with Galenshore $(c + na, d \prod_{i=1}^{n} y_i)$

c)

$$p(\theta|Y_1, ...Y_n) = \frac{2}{\Gamma(c+na)} (d \prod_{i=1}^{n} y_i)^{2(c+na)} \theta^{2(c+na)-1} e^{-(d \prod_{i=1}^{n} y_i)^2 \theta^2}$$

If we were to divide $p(\theta_a|Y_1, ..., Y_n)$ by $p(\theta_b|Y_1, ..., Y_n)$:

$$= (\theta_a^{2(c+na)-1} / \theta_b^{2(c+na)-1}) e^{-(d+\prod_{i=1}^{n} y_i)^2 (\theta_a^2 - \theta_b^2)}$$

It seems the sufficient statistic is $\prod_{i=1}^{n} y_i$.

d)

$$E[\theta|y_1, ..y_n] = \frac{\Gamma(c + na + 1/2)}{d \prod_{i=1}^{n} y_i \Gamma(c + na)}$$

e)

We are looking for:

$$Pr(\tilde{Y} = y|y_1, ..., y_n) = \int Pr(\tilde{Y} = y|\theta, y_1, ..., y_n) p(\theta|y_1, ..., y_n) d\theta$$

$$= \int p(\tilde{y}|\theta) p(\theta|y_1, ..., y_n) d\theta = \int \text{dgalenshore}(a, \theta) \text{dgalenshore}(c + na, d + \prod_{i=1}^{n} y_i)$$

$$= \frac{4}{\Gamma(a)\Gamma(c+na)} y^{2a-1} d \prod_{i=1}^{n} y_i^{2(c+na)} \int \theta^{2a+2c+2na-1} e^{-\theta^2(d^2 \prod_{i=1}^{n} y_i + y^2)} d\theta$$

I'm not sure how to continue from here, I think I made a mistake somewhere along the way here.

3.14)

a)

Since $Y_1, ..., Y_n \sim$ i.i.d. binary(θ) an appropriate sampling distribution for sum of Y is the binomial distribution:

$$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

The likelihood is:

$$L(\theta|y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

The log likelihood is:

$$l(\theta|y) = \log(\binom{n}{y}) + y\log(\theta) + (n-y)\log(1-\theta)$$

The first derivative of the log likelihood with respect to θ is:

$$\frac{dl(\theta|y)}{d\theta} = \frac{y}{\theta} - \frac{(n-y)}{1-\theta}$$

Setting this equal to zero and solving for θ leads to the MLE estimate $\hat{\theta} = \frac{y}{n}$.

The second derivative of the log likelihood with respect to θ is:

$$\frac{dl(\theta|y)}{d\theta^2} = -y\theta^{-2} - (n-y)(1-\theta)^{-2}$$

Continuing to simplify:

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

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

At this point we substitute in θ for its MLE $\hat{\theta}$ and multiply by -1. So:

$$J(\hat{\theta}) = -\frac{d^2 l(\hat{\theta}|y)}{d\theta^2} = \frac{y - 2\hat{\theta} + n\hat{\theta}^2}{\hat{\theta}^2(1-\hat{\theta})^2}$$

In the numerator we sub in $\hat{\theta} = y/n$:

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

Then we turn one $\hat{\theta}$ into y/n to simplify further:

$$J(\hat{\theta}) = \frac{n}{\hat{\theta}(1-\hat{\theta})}$$

So:

$$J(\hat{\theta})/n = \frac{1}{\hat{\theta}(1-\hat{\theta})}$$

b)

We are looking for a probability density $p_U(\theta)$ such that $\log(p_U(\theta)) = l(\theta|y)/n + c$.

We have:

$$l(\theta|y) = \log(\binom{n}{y}) + y\log(\theta) + (n-y)\log(\theta)$$

So if we divide by n we have:

$$l(\theta|y)/n = \frac{y}{n}\log(\theta) + \frac{n-y}{n}\log(1-\theta) + \frac{1}{n}\log(\binom{n}{y})$$

Knowing that the beta distribution is a conjugate to the binomial distribution we can see if the log transformation of that distribution will fit into the form we are looking for:

$$p_U(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

Taking a log transformation:

$$\log(p_U(\theta)) = (a-1)\log(\theta) + (b-1)\log(1-\theta) + \log(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)})$$

Ignoring the right most term that does not rely on θ we can see that the form of $\log(p_U(\theta))$ matches that of $l(\theta|y)/n$ given a certain choice of a and b:

So:

$$a - 1 = \frac{y}{n}, a = 1 + \frac{y}{n} \quad b - 1 = \frac{n-y}{n}, b = 1 + \frac{n-y}{n}$$

So the probability distribution $p_U(\theta)$ is beta$(1 + \frac{y}{n}, 1 + \frac{n-y}{n})$.

We then need to find the information of this density. Starting with $\log(p_U(\theta))$ we take the first derivative:

$$\frac{d\log(p_U(\theta))}{d\theta} = \frac{a-1}{\theta} + \frac{b-1}{1-\theta}$$

Then the second derivative:

$$\frac{d\log(p_U(\theta))}{d\theta^2} = -\frac{a-1}{\theta^2} - \frac{b-1}{(1-\theta)^2}$$

Then for the information we multiply by -1:

$$-\frac{d(\log(p_U(\theta))}{d\theta^2} = \frac{a-1}{\theta^2} + \frac{b-1}{(1-\theta)^2}$$

We have selected $a = 1 + \frac{y}{n}$ and $b = 1 + \frac{n-y}{n}$:

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

If we were to substitute in our MLE estimate this would reduce to $\frac{1}{\hat{\theta}(1-\hat{\theta})}$:

c)

Using what we know about the relationship between beta priors and binomial sampling distributions the posterior distribution would be beta$(1 + \frac{y}{n} + y, 1 + \frac{n-y}{y} + n - y)$.

Although its odd to use observed data as a prior, we can see in the above form that its pretty much an uninformative prior. The information introduced by the choice of a prior is dominated very quickly as n increases. So I think given a reasonable choice of n there is no problem with using this as a helpful way of setting up an uninformative prior that is mathematically simple.

d)

We are tasked with doing the same for a poisson sampling distribution.

a)

$$p(y|\theta) = \theta^y e^{-\theta} y! \quad, \log(p(y|\theta)) = y\log(\theta) - \theta + \log(y!)$$

$$\sum_{i=1}^{n} \log(p(y|\theta)) = l(\theta|y) = \sum_{i=1}^{n} y_i \log(\theta) - n\theta + \sum_{i=1}^{n} \log(y!)$$

$$\frac{dl(\theta|y)}{d\theta} = \frac{\sum_{i=1}^{n} y}{\theta} - n$$

When we set this to zero we have the MLE $\hat{\theta} = \frac{\sum_{i=1}^{n} y_i}{n}$.

$$\frac{dl(\theta|y)}{d\theta^2} = -\sum_{i=1}^{n} y_i/\theta^2$$

We sub in our MLE $\hat{\theta}$:

$$\frac{dl(\hat{\theta})}{d\theta^2} = -\sum_{i=1}^{n} y_i/(\sum_{i=1}^{n} y_i/n)^2 = -\frac{n^2}{\sum_{i=1}^{n} y_i}$$

Then we have:

$$J(\hat{\theta})/n = \frac{n}{\sum_{i=1}^{n} y_i}$$

b)

We are looking for a probability density $p_U(\theta)$ such that $\log(p_U(\theta)) = l(\theta|y)/n + c$.

We have:

$$l(\theta|y) = \sum_{i=1}^{n} y_i \log(\theta) - n\theta + \sum_{i=1}^{n} \log(y!)$$

So if we divide by n we have:

$$l(\theta|y)/n = \frac{\sum_{i=1}^{n} y_i}{n} \log(\theta) - \theta + \frac{1}{n}\sum_{i=1}^{n} \log(y!)$$

Knowing that the the gamma distribution is a conjugate to the poisson distribution we can see if the log transformation of that distribution will fit into the form we are looking for:

$$p_U(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$$

Taking a log transformation:

$$\log(p_U(\theta)) = a\log(b) - \log(\Gamma(a)) + (a-1)\log(\theta) - b\theta$$

Ignoring the first two terms that do not rely on θ we can see that the form of $\log(p_U(\theta)$ matches that of $l(\theta|y)/n$ given a certain choice of a and b:

$$a - 1 = \frac{\sum_{i=1}^n y_i}{n}, a = 1 + \frac{\sum_{i=1}^n y_i}{n} \quad b = 1$$

So the probability distribution of $p_U(\theta) = \text{beta}(1 + \frac{\sum_{i=1}^n y_i}{n}, 1)$.

We then need to find the information of this density. Starting with $\log(p_U(\theta))$ we take the first derivative:

$$\frac{d\log(p_U(\theta))}{d\theta} = \frac{a-1}{\theta} - b$$

Taking second derivative:

$$\frac{d\log(p_U(\theta))}{d\theta} = -\frac{a-1}{\theta^2}$$

Subbing in a and multiplying by -1:

$$-\frac{d\log(p_U(\theta))}{d\theta} = \frac{a-1}{\theta^2} = \frac{\sum_{i=1}^n y_i/n}{\theta^2}$$

If we were to substitute in our MLE estimate this would reduce to $\frac{n}{\sum_{i=1}^n y_i}$:

c)

Using what we know about the relationship between gamma priors and poisson sampling distributions the posterior distribution would be $\text{gamma}(1+\sum_{i=1}^n y_i/n+\sum_{i=1}^n y_i, 1+n)$. The discussion before is the same here, this serves as a relatively uninformative prior given a reasonable n so this can be considered a posterior for θ.