STA360 HW2

Ken Ye

1

a) $\mathbb{E}[Y_i] = \mathbb{E}[\mathbb{E}[Y_i \mid \theta]] = \mathbb{E}[\theta] = \frac{\eta}{\eta+\eta} = \frac{1}{2}$

(Tower Rule)

$V[Y_i] = \mathbb{E}[\mathbb{E}[Y_i \mid \theta]] + V[\mathbb{E}[Y_i \mid \theta]]$ (Law of Total Variance)

$$= \mathbb{E}[\theta(1-\theta)] + V[\theta]$$

$$= \int_0^1 \theta(1-\theta)\, p(\theta)\, d\theta + V[\theta]$$

$$= \int_0^1 \theta(1-\theta) \frac{\Gamma(\eta+\eta)}{\Gamma(\eta)\Gamma(\eta)} \theta^{\eta-1}(1-\theta)^{\eta-1}\, d\theta + V[\theta]$$

$$= \frac{\Gamma(2\eta)}{\Gamma(\eta)\Gamma(\eta)} \int_0^1 \theta^{\eta}(1-\theta)^{\eta}\, d\theta + V[\theta]$$

$$= \frac{\Gamma(2\eta)}{\Gamma(\eta)\Gamma(\eta)} \frac{\Gamma(\eta+1)\Gamma(\eta+1)}{\Gamma(2\eta+2)} \underbrace{\int_0^1 \frac{\Gamma(2\eta+2)}{\Gamma(\eta+1)\Gamma(\eta+1)} \theta^{\eta}(1-\theta)^{\eta}\, d\theta}_{=1} + V[\theta]$$

$$= \frac{\cancel{\Gamma(2\eta)}\, \cancel{\eta}\, \cancel{\Gamma(\eta)}\, \eta\, \cancel{\Gamma(\eta)}}{\cancel{\Gamma(\eta)}\cancel{\Gamma(\eta)}(2\eta+1)\, 2\cancel{\eta}\, \cancel{\Gamma(2\eta)}} + V[\theta]$$

$$= \frac{\eta}{2(2\eta+1)} + \frac{\eta\eta}{(\eta+\eta)^2(\eta+\eta+1)}$$

$$= \frac{\eta}{2(2\eta+1)} + \frac{\cancel{\eta^2}}{4\cancel{\eta^2}(2\eta+1)}$$

$$= \frac{2\eta+1}{4(2\eta+1)}$$

$$= \frac{1}{4}$$

b) $\mathbb{E}[Y_1 Y_2] = Pr(Y_1 = 1 \wedge Y_2 = 1)$

$$= p(y_1, y_2)$$

$$= \int_0^1 p(y_1, y_2 \mid \theta)\, p(\theta)\, d\theta$$

$$= \int_0^1 \theta^{\overset{\to 2}{y_1+y_2}} (1-\theta)^{\overset{\to 0}{2-y_1-y_2}} \frac{\Gamma(2\eta)}{\Gamma(\eta)\Gamma(\eta)} \theta^{\eta-1} (1-\theta)^{\eta-1}\, d\theta$$

$$= \frac{\Gamma(2\eta)}{\Gamma(\eta)\Gamma(\eta)} \int_0^1 \theta^{\eta+1} (1-\theta)^{\eta-1}\, d\theta$$

$$= \frac{\Gamma(2\eta)}{\Gamma(\eta)\Gamma(\eta)} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \text{ where } a = \eta+2,\ b = \eta.$$

$$= \frac{\Gamma(2\eta)}{\Gamma(\eta)\cancel{\Gamma(\eta)}} \frac{\Gamma(\eta+2)\cancel{\Gamma(\eta)}}{\Gamma(2\eta+2)}$$

$$= \frac{\cancel{\Gamma(2\eta)}(\eta+1)\cancel{\eta}\cancel{\Gamma(\eta)}}{\cancel{\Gamma(\eta)}(2\eta+1)(2\cancel{\eta})\cancel{\Gamma(2\eta)}}$$

$$= \frac{\eta+1}{4\eta+2}$$

c) $\mathrm{Corr}[Y_1, Y_2] = \dfrac{\mathrm{Cov}[Y_1, Y_2]}{\sqrt{V[Y_1]}\sqrt{V[Y_2]}}$

$$= \frac{\mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2]}{\sqrt{V[Y_1]}\sqrt{V[Y_2]}}$$

$$= \frac{\frac{\eta+1}{4\eta+2} - \left(\frac{1}{2}\right)^2}{\left(\sqrt{\frac{1}{4}}\right)^2}$$

$$= 4\left(\frac{\eta+1}{4\eta+2} - \frac{1}{4}\right)$$

$$= \frac{4\eta+4}{4\eta+2} - 1$$

$$= \frac{2}{4\eta+2} = \frac{1}{2\eta+1} \quad \text{(graph attached)}$$

d) According to the graph, as η increases from 0 to 1, $corr[Y_1, Y_2]$ decreases, which can be interpreted as that the more confident we are that θ is near $\frac{1}{2}$, the less info Y_1 and Y_2 provide about each other. This makes sense because when we are more certain about θ, we can infer about Y_i from $P(Y_i|\theta)$ more confidently since the observation of Y_i changes our belief about θ less.

2

a) $E[Y_i|\theta] = \theta x_i$, where θ is the average # of birds counted by a volunteer in 1 hour.

b) $p(y_1, \ldots, y_n | \theta) = P(y_1|\theta)\, P(y_2, \ldots, y_n | \theta, y_1)$ → $Y_1, \ldots, Y_n$ independent given θ

$$= P(y_1|\theta)\, P(y_2, \ldots, y_n | \theta)$$

$$\ldots$$

$$= P(y_1|\theta)\, P(y_2|\theta) \ldots P(y_n|\theta)$$

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

$$= \prod_{i=1}^{n} \frac{(\theta x_i)^{y_i} e^{-\theta x_i}}{y_i!}$$

$$\propto \theta^{\Sigma y_i} e^{-\theta \Sigma x_i}$$

$$L(\theta) \triangleq \ln\left(\theta^{\Sigma y_i} e^{-\theta \Sigma x_i}\right)$$

$$= \Sigma y_i \ln(\theta) - \Sigma x_i \theta$$

$$\frac{\partial}{\partial \theta}\left(\Sigma y_i \ln(\theta) - \Sigma x_i \theta\right) = \frac{\Sigma y_i}{\theta} - \Sigma x_i$$

$$= 0 \text{ when } \theta = \frac{\Sigma y_i}{\Sigma x_i} \leftarrow \text{MLE}$$

This MLE makes sense because Σy_i is the total # of birds counted, and Σx_i is the total # of hours. Σy_i divided by Σx_i would give us the # of birds counted / hr on average, so it makes sense that $\theta = \Sigma y_i / \Sigma x_i$, which is the sample mean.

c) $P(\theta \mid \underset{\sim}{y}) = \dfrac{P(\theta)\,P(\underset{\sim}{y} \mid \theta)}{P(\underset{\sim}{y})}$

$$\propto P(\theta)\,p(\underset{\sim}{y} \mid \theta) \qquad \left(\begin{array}{l} \theta \sim \text{gamma}(a,b)\ ;\ P(\theta) = \dfrac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \\ \underset{\sim}{y} \mid \theta \sim \text{pois}(\theta x_i)\ ;\ P(\underset{\sim}{y} \mid \theta) = \prod_{i=1}^{n} \dfrac{(\theta x_i)^{y_i} e^{-x_i}}{y_i!} \end{array} \right)$$

$$\propto \theta^{a-1} \cdot e^{-b\theta}\, \theta^{\Sigma y_i} e^{-\theta \Sigma x_i}$$

$$= \theta^{\Sigma y_i + a - 1} e^{(-\Sigma x_i - b)\theta} \quad \rightarrow \text{ Let } a = \Sigma y_i + a,\ b = \Sigma x_i + b$$

$$= \theta^{a-1} e^{b\theta}$$

$$\propto \text{gamma}(a, b)$$

$$= \text{gamma}(a + \Sigma y_i,\ b + \Sigma x_i) \qquad (\text{substitute back})$$

$$\text{mode}[\theta \mid \underset{\sim}{y}] = \frac{a + \Sigma y_i - 1}{b + \Sigma x_i} = \underset{\text{prior}}{\boxed{\frac{a}{b}}} \frac{b}{b + \Sigma x_i} + \underset{\text{sample mean}}{\boxed{\frac{\Sigma y_i}{\Sigma x_i}}} \frac{\Sigma x_i}{b + \Sigma x_i} - \frac{1}{b - \Sigma x_i}$$

Different from $MLE[\theta] = \dfrac{\Sigma y_i}{\Sigma x_i}$, which is the equivalent of the sample mean, the posterior $\text{mode}[\theta \mid \underset{\sim}{y}]$ not only looks at the sample data (i.e. Σy_i and Σx_i in the formula) but also a and b, which represents the prior belief. When sample size increases, Σx_i increases, and the "weight" $\left(\dfrac{\Sigma x_i}{b + \Sigma x_i}\right)$ on sample mean $\dfrac{\Sigma y_i}{\Sigma x_i}$ increases, so sample mean plays a larger role in determining the posterior $\text{mode}[\theta \mid \underset{\sim}{y}]$ than the prior belief $\dfrac{a}{b}$.

#3

a) $\theta_1 | Y_1 \sim Beta(2+1, 30+18) \Rightarrow \theta_1 | Y_1 \sim Beta(3, 48)$

$\theta_2 | Y_2 \sim Beta(2+16, 30+160) \Rightarrow \theta_2 | Y_2 \sim Beta(18, 190)$

(graph attached)

According to the graph, we can tell that the prior θ distribution (color blue) has the lowest peak / smallest mode and is the most spread-out, which indicates its high variance. This makes sense because we are not very certain about θ prior to the study.

Posterior θ_1 distribution (color red) has the second highest peak/mode and is less spread-out than prior θ, which is due to the fact that the study gives more information helping to consolidate the belief. The mode of posterior θ_1 is not far from that of prior θ, though it's slightly higher.

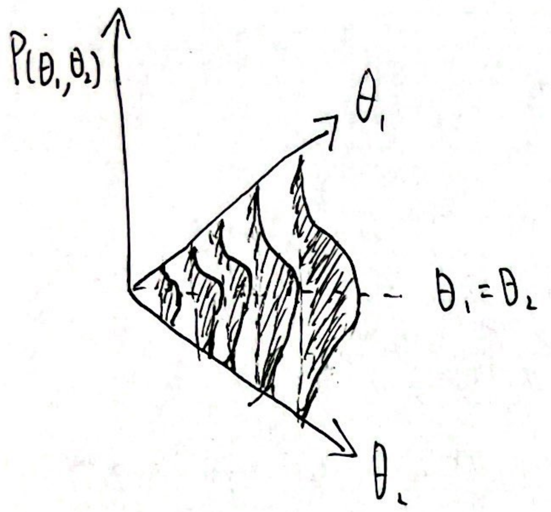
Posterior θ_2 distribution's mode is higher than both prior θ and posterior θ_1, and it has the highest peak and least spread-out distribution, indication of low variance. This makes sense because the robust sample size ($n=176$) helps update the belief, and now we are more confident in it due to the additional info.

b) $Mean[\theta_1 | Y_1] = 0.05883$, 95% C.I. = [0.01255, 0.13714]

$Mean[\theta_2 | Y_2] = 0.08654$, 95% C.I. = [0.0523, 0.12824]

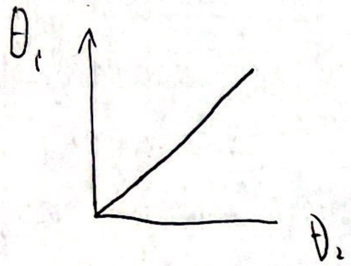
(code attached)

c)



supposed to be 3D, limited artistic talent :)

e.g. bivariate normal



In the joint prior distribution $P(\theta_1, \theta_2)$ drawn above, the mode is always on the line $\theta_1 = \theta_2$, representing the belief that θ_1 and θ_2 are close to each other. The large variance (revealed by the spread-out shape of the joint distribution) tells us that we are highly uncertain about the belief.