Math 339. Exam #1

Eliot Stanton

$\boxed{1}$ $f(y|\theta) = \frac{1}{\theta} y^{\frac{1}{\theta}-1}$, $0<y<1$

$E(Y) = \frac{1}{1+\theta}$

a) $E(Y) = \bar{Y}$ $\quad \frac{1}{1+\theta} = \frac{1}{n}\sum Y_i \quad n = (1+\theta)\sum Y_i \quad \frac{n}{\sum Y_i} = 1+\theta$

$\hat{\theta}_{mom} = \frac{n - \sum Y_i}{\sum Y_i} = \frac{1-\bar{Y}}{\bar{Y}}$

b) $L(\theta) = \frac{1}{\theta} Y_1^{\frac{1}{\theta}-1} \cdot \ldots \cdot \frac{1}{\theta} Y_n^{\frac{1}{\theta}-1} = \frac{1}{\theta^n}\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{\theta}-1}$

$\log L(\theta) = \log(\theta^{-n}) + \log\left[\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{\theta}-1}\right]$

$= -n\log\theta + \frac{1}{\theta}\log\left(\prod_{i=1}^{n} Y_i\right) - \log\left(\prod_{i=1}^{n} Y_i\right)$

$\frac{d}{d\theta}\log L(\theta) = \frac{-n}{\theta} - \frac{1}{\theta^2}\left(\log\prod_{i=1}^{n} Y_i\right) \stackrel{set}{=} 0$

$\frac{-n}{\theta} = \frac{1}{\theta^2}\log\prod_{i=1}^{n} Y_i \quad \hat{\theta}_{MLE} = \frac{-1}{n}\left(\log\prod_{i=1}^{n} Y_i\right)$

$\left[\log\prod_{i=1}^{n} Y_i = \sum_{i=1}^{n}\log Y_i \quad \hat{\theta}_{MLE} = \frac{-\sum_{i=1}^{n}\log Y_i}{n}\right]$

c) If I'm trying to decide between two biased estimators for θ, I would consider both the magnitude of the biases and the variances of the two estimators, looking for a small bias and small variance. To balance these two factors I would calculate the Mean Square Error $MSE(\hat{\theta}) = [\text{Bias}(\hat{\theta})]^2 + \text{Var}(\hat{\theta})$ and choose the estimator with the smallest $MSE(\hat{\theta})$.

$\boxed{2}$ $X_i \sim \text{Normal}(\mu_x, \sigma^2) \quad Y_i \sim \text{Normal}(\mu_y, \sigma^2)$

a) $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2(n-1) \implies E\left(\frac{(n-1)S_x^2}{\sigma^2}\right) = n-1 = \frac{n-1}{\sigma^2}E(S_x^2)$

$E(S_x^2) = \frac{\sigma^2(n-1)}{n-1} = \sigma^2$

$\text{Bias}(S_x^2) = E(S_x^2) - \sigma^2 = \sigma^2 - \sigma^2 = 0$ so S_x^2 is an unbiased estimator for σ^2

b) $Var\left(\frac{n-1}{\theta^2}S_x^2\right) = 2(n-1) = \frac{(n-1)^2}{\theta^4}Var(S_x^2)$

$Var(S_x^2) = \frac{2\theta^4(n-1)}{(n-1)^2} = \frac{2\theta^4}{n-1}$

c) $S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$

$E(S_p^2) = E\left(\frac{(n-1)S_x^2}{n+m-2} + \frac{(m-1)S_y^2}{n+m-2}\right) = \frac{n-1}{n+m-2}E(S_x^2) + \frac{m-1}{n+m-2}E(S_y^2) = \theta^2\left(\frac{n-1}{n+m-2} + \frac{m-1}{n+m-2}\right)$

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

$Bias(S_p^2) = E(S_p^2) - \theta^2 = \theta^2 - \theta^2 = 0$ so S_p^2 is an unbiased estimator for θ^2

d) $Var(S_p^2) = Var\left(\frac{(n-1)S_x^2}{n+m-2} + \frac{(m-1)S_y^2}{n+m-2}\right) = Var\left(\frac{(n-1)S_x^2}{n+m-2}\right) + Var\left(\frac{(m-1)S_y^2}{n+m-2}\right)$

$= \frac{(n-1)^2}{(n+m-2)^2}Var(S_x^2) + \frac{(m-1)^2}{(n+m-2)^2}Var(S_y^2) = \frac{(n-1)^2 \cdot 2\theta^4}{(n+m-2)(n-1)} + \frac{(m-1)^2 \cdot 2\theta^4}{(n+m-2)(m-1)}$

$= \frac{2\theta^4(n-1) + 2\theta^4(m-1)}{n+m-2} = 2\theta^4\left(\frac{n-1+m-1}{n+m-2}\right) = 2\theta^4$

$\boxed{3}$ $\Theta \sim Beta(\overset{\alpha}{7}, \overset{\beta}{4})$

.52 — .76

a) $E(\Theta) = \frac{\alpha}{\alpha+\beta} = \frac{7}{11} = .64$

$sd(\Theta) = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \sqrt{\frac{28}{121 \cdot 12}} = .14$

I will consider values within 1.5 standard deviations of the expected value of my prior to be quite reasonable. I make this choice because while there may be a high likelihood that θ is within one standard deviation of the mean, I don't want to disregard entirely the tails of my prior. Three standard deviations from the mean contains essentially all possible values of θ, but I will look only at the 2nd and 3rd quartiles of that range, which in this case means that .43 to .85 are very reasonable values for θ.

b) $Y|\theta \sim Binomial(20, \theta)$ $f(y|\theta) = \binom{20}{y}\theta^y(1-\theta)^{20-y}$, $y = 1, 2, \ldots, 20$

in this particular survey, $y = 15$ and $n = 20$

This is a Beta-Binomial conjugate family, which means $\theta | Y \sim Beta(\alpha + y, \beta + n - y)$

Here, that means $\theta | Y \sim Beta(22, 9)$, since $\alpha + y = 7 + 15 = 22$ and $\beta + n - y = 4 + 20 - 15 = 9$

$$E(\theta | Y) = \frac{\alpha}{\alpha+\beta} = \frac{22}{31} = .71$$

$$sd(\theta | Y) = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \sqrt{\frac{198}{31^2 \cdot 32}} = .08$$

Our understanding of θ has evolved after observing data. We now expect θ to be around .71, whereas before we thought it would be around .64. We're also slightly more certain of where θ is because the posterior has a standard deviation of .08, as compared to the prior's standard deviation of .14. Whereas I stated earlier that .43 to .85 were reasonable values for θ, my updated range of reasonable values (within 1.5 standard deviations of the mean) is .59 to .83. Essentially, we think a higher proportion of people prefer dogs than we previously thought.

c) In this case $y = 0$ and $n = 1$ in our data.

This means $\theta | Y \sim Beta(7, 5)$ because $\alpha + y = 7 + 0 = 7$ and $\beta + n - y = 4 + 1 = 5$

$$E(\theta | Y) = \frac{\alpha}{\alpha+\beta} = \frac{7}{12} = .58$$

$$sd(\theta | Y) = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \sqrt{\frac{35}{12^2 \cdot 13}} = .14$$

Our understanding of θ has evolved after observing this data. We now expect θ to be near .58, whereas our prior belief was that it would be close to .63. Despite thinking that a smaller proportion of people prefers dogs, we haven't really gained any more confidence in our estimate, since the posterior and prior have the same standard deviation of .14. A range of reasonable values for θ, those within 1.5 standard deviations of the mean, is now .37 to .79, as opposed to .43 to .85. This survey changes our understanding of θ less than the survey in part b, due to the much smaller sample size.

$\boxed{4}$ $Y_i | \lambda \sim Poisson(\lambda)$

a) I'd like to use a Gamma prior w/ Beta as a rate parameter, rather than scale for λ so I can take advantage of the Gamma-Poisson conjugate family. I will tune the prior to find one with an expected value around 12 and a variance around 5 so the range within 1.5 standard deviations is approximately 5 to 20.

$E(\lambda) = \frac{\alpha}{\beta} = 12$ $\alpha = 12\beta$ $Var(\lambda) = \frac{\alpha}{\beta^2} = 5$ $\alpha = 5\beta^2$ $5\beta^2 = 12\beta$ $\beta = \frac{12}{5} = 2.4$ $\alpha = 12(2.4) = 28.8$

I chose $\lambda \sim Gamma(28.8, 2.4)$ as my prior because it has a peak a little above 10 (closer to 12) and it is slightly skewed so that the reasonable range for λ is 5 to 20.

b) With the data $\underset{\sim}{y}|\lambda$ observed, $\sum_{i=1}^{5} y_i = 15+12+5+8+10 = 50$ and $n=5$, so $\bar{y}=10$.

Because the data $y_i|\lambda$ each follow a (iid) Poisson distribution, this is a Gamma-Poisson conjugate family, and $\lambda|\underset{\sim}{y} \sim Gamma(\sum y_i + \alpha, n+\beta)$.

Here, that means $\lambda|\underset{\sim}{y} \sim Gamma(78.8, 7.4)$ since $\sum y_i + \alpha = 50+28.8 = 78.8$ and $n+\beta = 5+2.4 = 7.4$

We now expect λ to be slightly lower, 10.6 instead of 12, which works with our prior fairly well, but we are definitely more certain of what λ is since the variance is much smaller, 1.4 instead of 5.

c) Anthony went wrong specifying a prior by choosing an expected value for λ that was too low, around 3. He should have selected a prior with a higher mean and variance since he apparently shouldn't have been as confident — maybe a flat prior would have been better.