Solutions Project 7
20 points

1. (2 pts) If p ~ Beta(α,β), then the density is $cp^{\alpha-1}(1-p)^{\beta-1}$ where c is a constant that does not depend on p. The log transformed density is $l(p) = \log(c) + (\alpha-1)\log(p) + (\beta-1)\log(1-p)$. The first derivative is $\frac{d}{dp}l(p) = (\alpha-1)/p - (\beta-1)/(1-p)$ which shows that $p = \frac{\alpha-1}{\alpha+\beta-2}$ is a critical point. For α>1 and β>1, $\frac{d^2}{dp^2}l(p) = -(\alpha-1)/p^2 - (\beta-1)/(1-p)^2$ is always negative, in which case $p = \frac{\alpha-1}{\alpha+\beta-2}$ is the maximum. If both α<1 and β<1, then in fact $p = \frac{\alpha-1}{\alpha+\beta-2}$ is a minimum! If only one of α<1 or β<1, then $p = \frac{\alpha-1}{\alpha+\beta-2}$ could be a maximum or a minimum possibly occurring outside of the sample space $0 \le p \le 1$ for p.

2. (3 pts)

a. The pdf for a single geometric observation y is $f(y \mid p) = p(1-p)^{y-1}$. Thus, the likelihood is $L(y_1,\ldots,y_n \mid p) = p^n(1-p)^{\sum y_i - n}$.

b. The prior is a Uniform distribution on [0,1].

c. The posterior is

$$f(p \mid y_1,\ldots,y_n) \propto \left(p^n(1-p)^{\sum y_i - n}\right) \times 1$$

which is proportional to Beta($\alpha^* = n + 1$, $\beta^* = \sum y_i - n + 1$).

d. The mean of the posterior is $\hat{p}_B = \frac{\alpha^*}{\alpha^* + \beta^*} = \frac{n+1}{\sum y_i + 2}$.

e. The MAP estimator is the maximizer of Beta($\alpha^* = n + 1$, $\beta^* = \sum y_i - n + 1$). By Problem #1 above,

$$\hat{p}_{MAP} = \frac{\alpha^* - 1}{\alpha^* + \beta^* - 2} = \frac{n}{\sum y_i} = \frac{1}{\bar{y}}.$$

which is the same as the MLE for p (Exercise 9.97b).

3. (2 pts)

a. Since n=10 and $\sum y_i = 3055$, then #2c shows that the posterior is Beta($\alpha^* = 11$, $\beta^* = 3046$).

b. By #2d, $\hat{p}_B = \frac{\alpha^*}{\alpha^* + \beta^*} = \frac{11}{3057} = 0.0036$.

c. By #2e, $\hat{p}_{MAP} = \dfrac{\alpha^* - 1}{\alpha^* + \beta^* - 2} = \dfrac{10}{3055} = 0.0033$.

d. By Exercise 9.97b, the MLE is $\hat{p} = 1/\bar{Y}$ = 10/3055 = 0.0033 is the same as the MAP.

e. A 95% credible interval is [0.0018, 0.0060]. This was calculated using the R code

```
> y=c(362,51, 200, 511, 211, 420, 299, 280, 398, 323)
> sum(y)
[1] 3055
> n
[1] 10
> a = 0.05
> qbeta(c(a/2,1-a/2),11,3046)
[1] 0.001798173 0.006009527
```

f. The evidence suggests that with probability .95, the true probability of machine failure on any given day is between 0.0017 and 0.0060.

4. **(4 pts)**

a. For a Beta(α^*, β^*) distribution, the mean is $\alpha^*/(\alpha^* + \beta^*)$. When $\alpha^* = \sum y_i + \alpha$ and $\beta^* = n - \sum y_i + \beta$, then the mean is $\hat{p}_B = \dfrac{\alpha^*}{\alpha^* + \beta^*} = \dfrac{\sum y_i + \alpha}{n + \alpha + \beta} = \dfrac{n}{n + \alpha + \beta}\hat{p}_{MLE} + \dfrac{\alpha}{n + \alpha + \beta}$ where $\hat{p}_{MLE} = \dfrac{\sum y_i}{n}$.

b. Assuming that α and β are constants,

$$E(\hat{p}_B) = \frac{n}{n + \alpha + \beta}E(\hat{p}_{MLE}) + \frac{\alpha}{n + \alpha + \beta} = \frac{n}{n + \alpha + \beta}p + \frac{\alpha}{n + \alpha + \beta}.$$

c. The bias is $Bias(\hat{p}_B) = E(\hat{p}_B) - p = \dfrac{-\alpha - \beta}{n + \alpha + \beta}p + \dfrac{\alpha}{n + \alpha + \beta}$ which is non-zero as long as $\alpha \neq \beta p/(1-p)$.

d. $Var(\hat{p}_B) = \dfrac{n^2}{(n + \alpha + \beta)^2}Var(\hat{p}_{MLE}) = \dfrac{n^2}{(n + \alpha + \beta)^2} \times \dfrac{p(1-p)}{n} = \dfrac{np(1-p)}{(n + \alpha + \beta)^2}$.

e. Since $Var(\hat{p}_B) = \dfrac{n^2}{(n + \alpha + \beta)^2}Var(\hat{p}_{MLE})$ and $\dfrac{n^2}{(n + \alpha + \beta)^2} < 1$, then $Var(\hat{p}_B) < Var(\hat{p}_{MLE})$.

f. Note that $\lim_{n\to\infty} Bias(\hat{p}_B) = \lim_{n\to\infty} \dfrac{-\alpha - \beta}{n + \alpha + \beta}p + \lim_{n\to\infty} \dfrac{\alpha}{n + \alpha + \beta} = 0 + 0 = 0$. Furthermore,

$$\lim_{n\to\infty} Var(\hat{p}_B) = \lim_{n\to\infty} \frac{np(1-p)}{(n + \alpha + \beta)^2} = \lim_{n\to\infty} \frac{p(1-p)}{n + 2(\alpha + \beta) + \dfrac{(\alpha + \beta)^2}{n}} = 0.$$

Application of the General Weak Law of Large Numbers (i.e., the more general version of Theorem 9.1) cinches the proof.

5. (2 pts) (Exercise 8.56 using a Bayesian analysis with a non-informative prior for p).

   (a) The MLE is $\hat{p}_{MLE} = \frac{360}{800} = 0.45$. From the course notes, the posterior is
   $$\text{Beta}(\alpha^\star = \sum y_i + 1 = 361, \beta^\star = n - \sum y_i + 1 = 441),$$
   and the Bayesian estimate is $\hat{p}_B = \frac{\alpha^\star}{\alpha^\star+\beta^\star} = \frac{\sum y_i + 1}{n+2} = \frac{361}{802} = 0.4501$.

   (b) A 98% CI for p is $0.45 \pm 2.326\sqrt{\frac{.45(.55)}{800}} = [0.409, 0.491]$.

   (c) A 98% credible interval for p is $[B_{.01}, B_{.99}] = [0.4095, 0.4911]$ where B_q is the q^{th} percentile from a Beta(361, 441). The R code is

```
> qbeta(c(.01,.99),361,441)
[1] 0.4094955 0.4911195
```

   (d) With probability 0.98, the true percentage of adults who say that movies are getting better is between 41% and 49%.

   (e) Since the credible interval is BELOW 0.5, then the evidence suggests that a minority of adults say that movies are getting better.

6. (2 pts) Consider a SRS $y_1, ..., y_n$ from $N(\mu, \sigma^2)$ when σ^2 is known, and assume an uninformative, flat prior for μ.

   (a) Since $p(\mu) \propto 1$, then
   $$\begin{aligned} p(\mu|y_1, ..., y_n) &\propto \exp\left(-\frac{1}{2\sigma^2}\sum(y_i - \mu)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\sum(-2y_i\mu + \mu^2)\right) \\ &= \exp\left(-\frac{1}{2\sigma^2}(-2n\bar{y}\mu + n\mu^2)\right) \\ &\propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y}^2 - 2\bar{y}\mu + \mu^2)\right) \\ &\propto N(\bar{y}, \sigma^2/n). \end{aligned}$$

   (b) Since the normal posterior is uni-modal and symmetric, then $\hat{\mu}_B = \hat{\mu}_{MAP} = \bar{y}$. Problem #6 in HW5 showed that $\hat{\mu}_{MLE} = \bar{y}$.

7. (5 pts) Let $y_1, ..., y_n$ denote a SRS from a Poisson(λ) distribution (as in Exercise 16.11).

   (a) If the prior is $p(\lambda)$=Gamma(α, β), then the posterior is
   $$\begin{aligned} p(\lambda|y_1, ..., y_n) &\propto \left(\lambda^{\sum y_i} e^{-n\lambda}\right) \times \left(\lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}\right) \\ &= \lambda^{\sum y_i + \alpha - 1} e^{-\lambda\left(\frac{n\beta+1}{\beta}\right)} \\ &\propto \text{Gamma}\left(\alpha^\star = \sum y_i + \alpha, \quad \beta^\star = \frac{\beta}{n\beta + 1}\right) \end{aligned}$$

(b) The posterior parameters are $\alpha^\star = \sum y_i + \alpha, \beta^\star = \frac{1}{n+1/\beta} = \frac{\beta}{n\beta+1}$.

(c) The Bayesian mean of the posterior is $\hat{\lambda}_B = \alpha^\star\beta^\star = \frac{\beta(\sum y_i+\alpha)}{n\beta+1} = \frac{\beta(n\bar{y}+\alpha)}{n\beta+1}$.

(d) In Exam 1, it was shown that $\hat{\lambda}_{MLE} = \bar{y}$.

(e) Assuming that α and β are fixed, then, by 7c, $E(\hat{\lambda}_B) = \frac{\beta(nE(\bar{y})+\alpha)}{n\beta+1} = \frac{\beta(n\lambda+\alpha)}{n\beta+1}$.

(f) As long as $\alpha\beta \neq \lambda$, $\hat{\lambda}_B$ is biased.

(g) By #10c, $Var(\hat{\lambda}_B) = \frac{n^2\beta^2 Var(\bar{y})}{(n\beta+1)^2} = \frac{n\beta^2\lambda}{(n\beta+1)^2}$.

(h) By #10g, $Var(\hat{\lambda}_B) = \frac{n^2\beta^2 Var(\hat{\lambda}_{MLE})}{(n\beta+1)^2}$. Since $\frac{n^2\beta^2}{(n\beta+1)^2} < 1$, then $Var(\hat{\lambda}_B) < Var(\hat{\lambda}_{MLE})$.

(i) From #10e, $\lim_{n\to\infty} E(\hat{\lambda}_B) = \frac{\beta(n\lambda+\alpha)}{n\beta+1} = \lambda$. From #10g, $\lim_{n\to\infty} Var(\hat{\lambda}_B) = \frac{n\beta^2\lambda}{(n\beta+1)^2} = 0$. Application of the more general version of Theorem 9.1 shows that $\hat{\lambda}_B$ is consistent for μ.