Homework 8 Solutions
Statistics 200B
Due April 4

1. Consider a Bayesian model in which, conditional on unknown parameter λ, $X_1, \ldots, X_n$ are iid with exponential PDF

$$f(x|\lambda) = \frac{1}{\lambda}e^{-x/\lambda}$$

for $x > 0$.

   (a) Find the Jeffreys prior for λ.
   (b) Is the Jeffreys prior proper? Why or why not?
   (c) Find the posterior distribution for λ using the Jeffreys prior.
   (d) Is this posterior proper? Why or why not?

   Solution:

   (a) The Jeffereys prior for λ is

$$f(\lambda) \propto \sqrt{I(\lambda)} = \sqrt{-E\left(\frac{d^2}{d\lambda^2}\log f(x|\lambda)\right)} = 1/\lambda.$$

   (b) From (a), $f(\lambda) = c/\lambda$, where c is a constant unrelated to λ.

$$\int_0^\infty f(\lambda)d\lambda = \int_0^\infty \frac{c}{\lambda}d\lambda = c\log\lambda\,|_0^\infty = \infty.$$

   So $\int_0^\infty f(\lambda)d\lambda \neq 1$, and the Jeffereys prior $f(\lambda)$ is not a proper prior.

   (c) By Bayes rule,

$$f(\lambda|X^n) \propto f(X^n|\lambda)f(\lambda) \propto \frac{1}{\lambda^n}e^{-\sum_{i=1}^n X_i/\lambda}\frac{1}{\lambda} = \frac{1}{\lambda^{n+1}}e^{-\sum_{i=1}^n X_i/\lambda}.$$

   By recognizing it as the kernel of $InverseGamma(n, \sum_{i=1}^n X_i)$, we know that $\lambda|X^n \sim InverseGamma(n, \sum_{i=1}^n X_i)$, which is the posterior distribution of λ.

(d) Since $n > 0$, $\sum_{i=1}^{n} X_i > 0$, $InverseGamma(n, \sum_{i=1}^{n} X_i)$ is a valid distribution. Therefore, the posterior is proper.

2. Consider again the earthquake example from Homework 7, Problem 2. Calculate the posterior distribution for λ, this time using the Jeffreys prior. Make a plot of the two posterior densities, under your subjective prior and under the Jeffreys prior. Include a sentence comparing the two.

Solution:

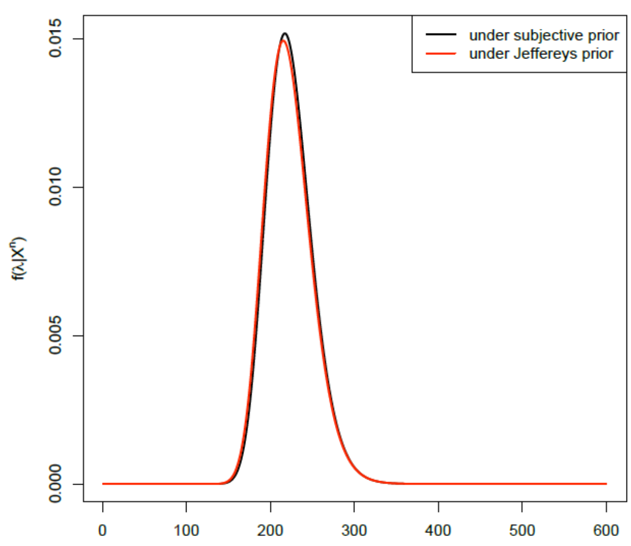
In Problem 2 of Homework 7, we used a subjective prior $\lambda \sim InverseGamma(a, b)$, with $a = 3.33$ and $b = 851.27$. (Your subjective prior may be different.) The corresponding posterior is

$$\lambda | X^n \sim InverseGamma(a+n, b+\sum_{i=1}^{n} X_i) = InverseGamma(69.33, 15276.06).$$

By using the Jeffereys prior, from Problem 1, the posterior is

$$\lambda | X^n \sim InverseGamma(n, \sum_{i=1}^{n} X_i) = InverseGamma(66, 14424.79).$$

The two posterior distributions are plotted in the plot below.



The `R` code is

```
library(MCMCpack)
lambda <- seq(1,600,by=0.1)
post_sub <- dinvgamma(lambda, shape=69.33, scale=15276.06)
post_jef <- dinvgamma(lambda, shape=66, scale=14424.79)
plot(x=lambda, y=post_sub, type="l", lwd=2, xlab=expression(lambda),
    ylab=expression(paste("f(",lambda,"|",X^n,")")))
lines(x=lambda, y=post_jef, col=2, lwd=2)
legend("topright", legend=c("under subjective prior",
    "under Jeffereys prior"), lwd=2, col=1:2)
```

The posterior mean under the Jefferey's prior is slightly smaller than the posterior mean under the subjective prior. The reason is that the Jeffereys prior does not use any knowledge about the occurrence of the earthquake, while the subjective prior supposes that the earthquake occurs more often than observed, and thus increases the posterior mean.

3. Suppose $X|p_1 \sim Bin(n, p_1)$ and $Y|p_2 \sim Bin(m, p_2)$, with X and Y independent given p_1 and p_2. Let $H_0 : p_1 = p_2$, and suppose under H_0 we assign prior distribution $p_1 \sim Unif(0, 1)$ (and $p_2 = p_1$) and under $H_1 : p_1 \neq p_2$ we assign independent priors $p_1 \sim Unif(0, 1)$ and $p_2 \sim Unif(0, 1)$.

   (a) Calculate $f(x, y|H_1)$. *Hint: Use the independence assumptions.*

   (b) Calculate $f(x, y|H_0)$.

   (c) Use (a) and (b) to compute the Bayes factor BF_{10} for comparing H_1 to H_0.

   (d) Let p_1 denote the "true" batting average for Albert Pujols and p_2 denote the "true" batting average for Ichiro Suzuki. Consider the data

   Pujols: 5146 at bats; 1717 hits

   Suzuki: 6099 at bats; 2030 hits

   treating each at bat as a Bernoulli trial with probability p_1 or p_2 of getting a hit. Calculate BF_{10} for $H_1 : p_1 \neq p_2$ to $H_0 : p_1 = p_2$. Explain the level of evidence this indicates for H_1 relative to H_0.

Solution:

(a)

$$\begin{aligned}
f(x,y|H_1) &= f(x|H_1)f(y|H_1) \text{ (because } X \text{ and } Y \text{ are independent)} \\
&= \int_0^1 f(x|p_1)f(p_1)dp_1 \int_0^1 f(y|p_2)f(p_2)dp_2 \\
&= \int_0^1 \binom{n}{x} p_1^x(1-p_1)^{n-x}dp_1 \int_0^1 \binom{m}{y} p_2^y(1-p_2)^{m-y}dp_2 \\
&\left(\begin{array}{l} \text{Since } p_1 \sim Beta(x+1, n-x+1) \text{ has density} \\ f(p_1) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} p_1^x(1-p_1)^{n-x}, \\ \Rightarrow \int_0^1 p_1^x(1-p_1)^{n-x}dp_1 = \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)} \end{array}\right) \\
&= \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}\binom{m}{y}\frac{\Gamma(y+1)\Gamma(m-y+1)}{\Gamma(m+2)} \\
&= \frac{n!}{x!(n-x)!}\frac{x!(n-x)!}{(n+1)!}\frac{m!}{y!(m-y)!}\frac{y!(m-y)!}{(m+1)!} \\
&= \frac{1}{(n+1)(m+1)}.
\end{aligned}$$

(b)

$$\begin{aligned}
f(x,y|H_0) &= \int_0^1 f(x|p_1)f(y|p_1)f(p_1)dp_1 \\
&= \int_0^1 \binom{n}{x} p_1^x(1-p_1)^{n-x}\binom{m}{y} p_1^y(1-p_1)^{m-y}dp_1 \\
&= \int_0^1 \binom{n}{x}\binom{m}{y} p_1^{x+y}(1-p_1)^{n+m-x-y}dp_1 \\
&= \binom{n}{x}\binom{m}{y}\frac{\Gamma(x+y+1)\Gamma(n+m-x-y+1)}{\Gamma(n+m+2)} \\
&= \frac{n!}{x!(n-x)!}\frac{m!}{y!(m-y)!}\frac{(x+y)!(n+m-x-y)!}{(n+m+1)!} \\
&= \frac{1}{(n+m+1)\binom{n+m}{n}}\binom{x+y}{x}\binom{n+m-x-y}{n-x}.
\end{aligned}$$

(c) The Bayes factor BH_{10} for comparing H_1 to H_0 is

$$BF_{10} = \frac{f(x,y|H_1)}{f(x,y|H_0)} = \frac{\frac{1}{(n+1)(m+1)}}{\frac{1}{(n+m+1)\binom{n+m}{n}}\binom{x+y}{x}\binom{n+m-x-y}{n-x}} = \left(\frac{1}{n+m+2}\right)\frac{\binom{n+m+2}{n+1}}{\binom{x+y}{x}\binom{n+m-x-y}{n-x}}.$$

(d)

$$n = 5146; x = 1717;$$
$$m = 6099; y = 2030.$$

$$BF_{10} = \left(\frac{1}{n+m+2}\right) \frac{\binom{n+m+2}{n+1}}{\binom{x+y}{x}\binom{n+m-x-y}{n-x}} \approx 0.022 < 1.$$

Therefore, the evidence is negative toward H_1. In other words, it supports H_0.

4. The owner of a ski shop must order skis for the upcoming season. Orders must be placed in quantities of 25 pairs of skis. The cost per pair of skis is \$50 if 25 are ordered, \$45 if 50 are ordered, and \$40 if 75 are ordered. The skis will be sold at \$75 per pair. Any skis left over at the end of the year can be sold (for sure) at \$25 per pair. If the owner runs out of skis during the season, she will suffer a loss of "goodwill" among unsatisfied customers. She rates this loss at \$5 per unsatisfied customer. For simplicity, suppose the owner feels that demand for the skis will be 30, 40, 50, or 60 pairs of skis, with probabilities 0.2, 0.4, 0.2, and 0.2, respectively.

   (a) Describe the parameter space Θ and the space of possible actions $\mathcal{A}$.
   (b) What is the prior distribution?
   (c) For each possible $\theta \in \Theta$ and $a \in \mathcal{A}$, compute the loss. (The loss in this case may be negative, representing a good outcome for the shop owner.) Display these possibilities in a matrix.
   (d) What is the Bayes rule? That is, what action minimizes the Bayes risk? Note that in this example, there is no data, so the frequentist risk is the same as the loss.

   Solution:

   (a) Let the parameter θ be the quantity demanded. The parameter space is

   $$\Theta = \{30, 40, 50, 60\}.$$

   Since we want the quantity ordered to be very close to the quantity demanded, and each order must be in quantities of 25 pairs of skis. The space of possible actions is

   $$\mathcal{A} = \{25, 50, 75\}.$$

(b) The prior distribution is as follows.

θ_0	$P(\theta=\theta_0)$
30	0.2
40	0.4
50	0.2
60	0.2

(c) The matrix representing the loss is as follows.

	$L(\theta,\hat{\theta})$	$\hat{\theta}$ = 25	50	75
	30	-600	-500	-375
θ	40	-550	-1000	-875
	50	-500	-1500	-1375
	60	-450	-1450	-1875

(d) The Bayes risk is

$$r(\theta,\hat{\theta}) = \sum_{\theta_0} R(\theta_0,\hat{\theta})P(\theta=\theta_0) = \sum_{\theta_0} L(\theta_0,\hat{\theta})P(\theta=\theta_0)$$
$$= 0.2L(30,\hat{\theta}) + 0.4L(40,\hat{\theta}) + 0.2L(50,\hat{\theta}) + 0.2L(60,\hat{\theta}).$$

For different $\hat{\theta}$, the corresponding $r(\theta,\hat{\theta})$ are summarized below.

	$\hat{\theta}$ = 25	50	75
$r(\theta,\hat{\theta})$	-530	-1090	-1075

Since $\hat{\theta}=50$ gives the smallest $r(\theta,\hat{\theta})$, it is the Bayes rule.

5. Suppose $X|p \sim Binomial(n,p)$ and $p \sim Beta(\alpha,\beta)$.

   (a) Calculate the posterior risk $r(\hat{p}|x)$ for an arbitrary estimator $\hat{p}$.

   (b) For a given x, what value of $\hat{p}(x)$ minimizes the posterior risk? Use this to construct a Bayes estimator.

   (c) What is the posterior risk for the Bayes estimator you found in (b)? How does it relate to the posterior distribution?

Solution:

(a) By Bayes rule,

$$f(p|x) \propto f(x|p)f(p) \propto p^x(1-p)^{n-x}p^{\alpha-1}(1-p)^{\beta-1} = p^{\alpha+x-1}(1-p)^{\beta+n-x-1}.$$

So $p|x \sim Beta(\alpha + x, \beta + n - x)$, and

$$f(p|x) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} p^{\alpha+x-1}(1-p)^{\beta+n-x-1}.$$

For L^2 error loss, the posterior risk is

$$\begin{aligned}
r(\hat{p}|x) &= \int_0^1 L(p,\hat{p}) f(p|x) dp = \int_0^1 (p-\hat{p})^2 f(p|x) dp \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \int_0^1 (p^2 - 2\hat{p}p + \hat{p}^2)^2 p^{\alpha+x-1}(1-p)^{\beta+n-x-1} dp \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \left[\int_0^1 p^{\alpha+x+1}(1-p)^{\beta+n-x-1} dp \right. \\
&\quad \left. -2\hat{p} \int_0^1 p^{\alpha+x}(1-p)^{\beta+n-x-1} dp + \hat{p}^2 \int_0^1 p^{\alpha+x-1}(1-p)^{\beta+n-x-1} dp \right] \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \left[\frac{\Gamma(\alpha+x+2)\Gamma(\beta+n-x)}{\Gamma(\alpha+\beta+n+2)} \right. \\
&\quad \left. -2\hat{p} \frac{\Gamma(\alpha+x+1)\Gamma(\beta+n-x)}{\Gamma(\alpha+\beta+n+1)} + \hat{p}^2 \frac{\Gamma(\alpha+x)\Gamma(\beta+n-x)}{\Gamma(\alpha+\beta+n)} \right] \\
&= \frac{(\alpha+x)(\alpha+x+1)}{(\alpha+\beta+n)(\alpha+\beta+n+1)} - 2\hat{p}\frac{(\alpha+x)}{(\alpha+\beta+n)} + \hat{p}^2.
\end{aligned}$$

(b) By setting

$$\frac{d}{d\hat{p}} r(\hat{p}|x) = -2\frac{(\alpha+x)}{(\alpha+\beta+n)} + 2\hat{p} = 0,$$

$$\Rightarrow \hat{p} = \frac{(\alpha+x)}{(\alpha+\beta+n)}.$$

Also since

$$\frac{d^2}{d\hat{p}^2} r(\hat{p}|x) = 2 > 0,$$

$\hat{p} = \frac{(\alpha+x)}{(\alpha+\beta+n)}$ minimizes $r(\hat{p}|x)$ and is the Bayes estimator.
Note that $\hat{p}$ is the posterior mean. (The posterior is $Beta(\alpha+x, \beta+n-x)$.)
This agrees with the general result for L^2 error loss.

(c) The posterior risk for the Bayes estimator $\hat{p} = \frac{(\alpha+x)}{(\alpha+\beta+n)}$ is

$$\begin{aligned}
r(\hat{p}|x) &= \frac{(\alpha+x)(\alpha+x+1)}{(\alpha+\beta+n)(\alpha+\beta+n+1)} - 2\hat{p}\frac{(\alpha+x)}{(\alpha+\beta+n)} + \hat{p}^2 \\
&= \frac{(\alpha+x)(\alpha+x+1)}{(\alpha+\beta+n)(\alpha+\beta+n+1)} - 2\frac{(\alpha+x)^2}{(\alpha+\beta+n)^2} + \frac{(\alpha+x)^2}{(\alpha+\beta+n)^2} \\
&= \frac{(\alpha+x)(\alpha+x+1)}{(\alpha+\beta+n)(\alpha+\beta+n+1)} - \frac{(\alpha+x)^2}{(\alpha+\beta+n)^2} \\
&= \frac{(\alpha+x)\left[(\alpha+x+1)(\alpha+\beta+n) - (\alpha+x)(\alpha+\beta+n+1)\right]}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \\
&= \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}.
\end{aligned}$$

Since the posterior distribution is $Beta(\alpha+x, \beta+n-x)$, whose variance is

$$\frac{(\alpha+x)(\beta+n-x)}{[(\alpha+x)+(\beta+n-x)]^2[(\alpha+x)+(\beta+n-x)+1]} = \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)},$$

$r(\hat{p}|x)$ is the posterior variance.

6. Let $\Theta = \{\theta_1, \ldots, \theta_k\}$ be a finite parameter space. Prove that the posterior mode is the Bayes estimator under zero-one loss.

Proof:

The Bayes estimator $\hat{\theta}$ minimizes the posterior risk $r(\hat{\theta}|x)$. The zero-one loss function is

$$L(\theta, \hat{\theta}) = \begin{cases} 1 & \theta \neq \hat{\theta} \\ 0 & \theta = \hat{\theta} \end{cases}.$$

Let $f(\theta_i|x)$ be the posterior mass function of θ_i, $i = 1, \ldots, k$. The posterior risk is

$$r(\hat{\theta}|x) = \sum_{i=1}^{k} L(\theta_i, \hat{\theta}) f(\theta_i|x) = \sum_{\theta_i \neq \hat{\theta}} f(\theta_i|x) = 1 - f(\hat{\theta}|x).$$

Since we want $\hat{\theta}$ to minimize $r(\hat{\theta}|x)$, it should maximize $f(\hat{\theta}|x)$. Therefore, the posterior mode (θ_i that maximizes $f(\theta_i|x)$) is the Bayes estimator $\hat{\theta}$ under zero-one loss.