Solutions to Homework 6
BIOS 7731

1. **Suppose $X_1, \ldots, X_n$ are IID $Poisson(\theta)$ and consider the prior distribution for θ to be $Gamma(1, \lambda)$.**

   (a) **According to the loss function $l(\theta, a) = \theta^p(\theta - a)^2$, where p is a fixed positive constant, show that the Bayes rule is $\frac{T(X)+p+1}{n+\lambda}$, where $T(X) = \sum_{i=1}^n X_i$.**
   The Bayes rule a^* minimizes $E[\theta^p(\theta - a)^2|X]$. Taking the derivative with respect to a of

$$E[a^2\theta^p - 2a\theta^{p+1} + \theta^{p+2}|X]$$

   and setting it to zero, the Bayes rule is

$$a^* = \frac{E[\theta^{p+1}|X]}{E[\theta^p|X]}.$$

   (Not shown, but the second derivative test can be used to show that this is a minimum.)
   Given the prior distribution for θ as $Gamma(1, \lambda)$, the posterior distribution is

$$\pi(\theta|X) \propto \frac{\theta^{T(X)}e^{-n\theta}}{\prod x_i}\lambda e^{-\lambda\theta},$$

   which is $Gamma(\alpha^{'}, \lambda^{'})$, where $\alpha^{'} = T(X) + 1$ and $\lambda^{'} = n + \lambda$.
   The numerator of a^* is

$$E[\theta^{p+1}|X] = \frac{\lambda^{\alpha^{'}}}{\Gamma(\alpha^{'})}\int_\theta \theta^{p+1}\theta^{\alpha^{'}-1}e^{-\lambda^{'}\theta}d\theta.$$

   The part in the integral is the kernel of a Gamma density, $Gamma(\alpha^{''}, \lambda^{''})$, where $\alpha^{''} = \alpha^{'} + p + 1$ and $\lambda^{''} = \lambda^{'}$ Therefore,

$$E[\theta^{p+1}|X] = \frac{\lambda^{\alpha^{'}}}{\Gamma(\alpha^{'})}\frac{\Gamma(\alpha^{''})}{\lambda^{\alpha^{''}}}.$$

   Similarly, the denominator of a^* is

$$E[\theta^p|X] = \frac{\lambda^{\alpha^{'}}}{\Gamma(\alpha^{'})}\int_\theta \theta^p\theta^{\alpha^{'}-1}e^{-\lambda^{'}\theta}d\theta.$$

   The part in the integral is the kernel of a Gamma density, $Gamma(\alpha^{'''}, \lambda^{'''})$, where $\alpha^{'''} = \alpha^{'} + p$ and $\lambda^{'''} = \lambda^{'}$. Therefore,

$$E[\theta^p|X] = \frac{\lambda^{\alpha^{'}}}{\Gamma(\alpha^{'})}\frac{\Gamma(\alpha^{'''})}{\lambda^{\alpha^{'''}}}.$$

Putting this together,

$$\frac{E[\theta^{p+1}|X]}{E[\theta^p|X]} = \frac{\frac{\lambda^{\alpha'}}{\Gamma(\alpha')}\frac{\Gamma(\alpha'')}{\lambda^{\alpha''}}}{\frac{\lambda^{\alpha'}}{\Gamma(\alpha')}\frac{\Gamma(\alpha''')}{\lambda^{\alpha'''}}}$$

$$= \frac{\frac{\Gamma(\alpha'')}{\lambda^{\alpha''}}}{\frac{\Gamma(\alpha''')}{\lambda^{\alpha'''}}} = \frac{\Gamma(\alpha' + p + 1)}{\Gamma(\alpha' + p)}\frac{\lambda^{\alpha' + p}}{\lambda^{\alpha' + p + 1}} = \frac{\alpha' + p}{\lambda'} = \frac{T(X) + 1 + p}{n + \lambda}$$

Alternative solution is using BD Problem 3.2.5 (pg. 197).

(b) **Is the Bayes rule minimax? If so, for what values of λ and p is it minimax?**

For this Bayes rule, we need to find the values of λ and p such that the risk is constant (i.e., it does not depend on θ).

$$Risk(a^*) = E[\theta^p(\theta - a^*)^2] = \theta^p E[(\theta - a^*)^2] = \theta^p(Var(a^*) + bias(a^*)^2)$$

Since $T(X)$ is Poisson, then

$$Var(\frac{T(X) + 1 + p}{n + \lambda}) = \frac{n\theta}{(n + \lambda)^2},$$

$$E[\frac{T(X) + 1 + p}{n + \lambda}] = \frac{n\theta + p + 1}{n + \lambda}.$$

and

$$bias(a^*)^2 = (\frac{n\theta + p + 1}{n + \lambda} - \theta)^2 = \frac{(n\theta + p + 1 - \theta(n + \lambda))^2}{(n + \lambda)^2} = \frac{((p + 1) - \theta\lambda)^2}{(n + \lambda)^2}.$$

Therefore,

$$Risk(a^*) = \frac{\theta^p}{(n + \lambda)^2}(n\theta + ((p + 1) - \theta\lambda)^2)$$

$$= \frac{\theta^p}{(n + \lambda)^2}((p + 1)^2 + \theta(n - 2\lambda(p + 1)) + \theta^2(\lambda^2)).$$

To obtain constant risk,

i. If $p = 0$, then find λ such that the θ and θ^2 terms disappear. For the θ^2 term, we need $\lambda^2 = 0$. Therefore, set $\lambda = 0$. Then for the θ term , we need $-2\lambda(p + 1) + n = 0$. Therefore, $\lambda(p + 1)$ must equal $\frac{n}{2}$, and since $p = 0$, then λ must be equal to $\frac{n}{2}$, which contradicts the value above ($\lambda = 0$).

ii. Alternatively, set $p = -1$ for the θ and θ^2 terms to disappear, then $\lambda = 0$ for constant risk.

In either case, the Gamma prior distribution is not possible nor with the limits of Bayes rules (Thm 3.3.3). Therefore, there is no $p > 0$, $\lambda > 0$ such that this Bayes rule has constant risk and this Bayes rule is not minimax.

2. **Empirical Bayes: Consider estimation of regression slopes $\theta_1, \dots, \theta_p$ for p pairs of observations, $(X_1, Y_1), \dots, (X_p, Y_p)$, modeled as independent with $X_i \sim N(0, 1)$ and $Y_i|X_i = x \sim N(\theta_i x, 1)$.**

   (a) **Following a Bayesian approach, let the unknown parameters θ_i be iid random variables from $N(0, \tau^2)$. Find the Bayes estimate of θ_i in this Bayesian model with squared error loss.**

   The Bayes estimate under squared loss is $E[\theta|X, Y]$. The posterior distribution for θ_i is

   $$p(\theta|x_i, y_i) \propto \pi(\theta_i)p(x_i)p(y_i|x_i, \theta_i).$$

   These are all normal distributions and therefore, by completing the square (details skipped), we obtain

   $$\theta_i|x_i, y_i \sim N(\frac{x_i y_i \tau^2}{x_i^2\tau^2 + 1}, \frac{\tau^2}{x_i^2\tau^2 + 1}).$$

   The posterior mean is the Bayes estimate $\hat{\theta}_i = \frac{x_i y_i \tau^2}{x_i^2\tau^2+1}$.

   (b) **Determine $E[Y_i^2]$ in the Bayesian model. Using this, suggest a simple method of moments estimator for τ^2.**

   $$E[Y_i^2] = E[E[Y_i^2|X_i, \theta_i]] = E[Var(Y|X_i, \theta_i) + E(Y|X_i, \theta_i)^2]$$
   $$= E[1 + (\theta_i X_i)^2] = 1 + E[\theta_i^2]E[X_i^2],$$

   since X_i is independent of θ_i.

   $$E[Y_i^2] = 1 + (\tau^2 + 0)(1 + 0)) = 1 + \tau^2.$$

   Using the sample second moment $\hat{\mu}_2 = \frac{\sum_{i=1}^p y_i^2}{p}$, a method of moments estimator for τ is

   $$\hat{\tau}^2 = \hat{\mu}_2 - 1.$$

   (c) **Given an empirical Bayes estimator for θ_i combining the simple "empirical" estimate for τ in part b) with the Bayes estimate for θ_i when τ is known in part a).**

   Plugging in the estimator for τ^2 from part b), the empirical Bayes estimator for θ_i is $\hat{\theta}_i = \frac{x_i y_i \hat{\tau}^2}{x_i^2\hat{\tau}^2+1}$.

3. **BD 3.4.2**

   Suppose that there is an unbiased estimate δ of $q(\theta)$ and that $T(X)$ is sufficient. Show that if $l(\theta, a)$ is convex and $\delta^*(X) = E[\delta(X)|T(X)]$, then $R(\theta, \delta^*) \leq R(\theta, \delta)$.

   By the double expectation theorem, the Risk is

   $$R(\theta, \delta) = E[l(\theta, \delta)] = E[E[l(\theta, \delta)|T(X)]].$$

   Since the loss function $l()$ is convex, then by Jensen's Inequality,

   $$E[E[l(\theta, \delta)|T(X)]] \geq E[l(\theta, E[\delta|T(X)])] = E[l(\theta, \delta^*)] = R(\theta, \delta^*),$$

   where $T(X)$ is sufficient so that $\delta^* = E[\delta(X)|T(X)]$ does not depend on θ. Therefore δ^* minimizes the risk for a convex loss function.

4. **BD 3.4.3**

Let $X \sim p(x,\theta)$ with $\theta \in \Theta \subset R$, suppose that assumptions I and II hold and that h is a monotone increasing differentiable function from θ onto $h(\theta)$. Reparameterize the model by setting $\eta = h(\theta)$ and let $q(x,\eta) = p(x, h^{-1}(\eta))$ denote the model in the new parameterization.

(a) Let $I_p(\theta)$ and $I_q(\theta)$ denote the Fisher information, in the two parameterizations.

First, note that since h is a monotone increasing differentiable function, we can take h^{-1}. Also, note that since $\eta = h(\theta)$, then $\partial\eta = h'(\theta)\partial\theta$ and $\frac{\partial\theta}{\partial\eta} = \frac{1}{h'(\theta)}$. To find $I_q(\eta)$, take the following derviative,

$$\frac{\partial}{\partial\eta}\log q(x|\eta) = \frac{\partial}{\partial\theta}\log q(x|\eta)\frac{\partial\theta}{\partial\eta} = \frac{\partial}{\partial\theta}\log p(x|h^{-1}(\eta))/h'(\theta) = \frac{\partial}{\partial\theta}\log p(x|\theta)/h'(\theta).$$

Since

$$I_q(\eta) = E[(\frac{\partial}{\partial\eta}\log q(x|\eta))^2],$$

then

$$I_q(\eta) = E[(\frac{\partial}{\partial\theta}\log p(x|\theta)/h'(\theta))^2] = I_p(\theta)/[h'(\theta)^2] = I_p(h^{-1}(\eta))/[h'(h^{-1}(\eta))^2].$$

Therefore, the Fisher information I is not equivariant under monotone increasing transformations of the parameter.

(b) Let $\phi(\eta) = E_q[T(X)]$ and $\psi(\theta) = E_p[T(X)]$, then $\phi(\eta) = \psi(h^{-1}(\eta))$ and the derivatives are

$$\frac{\partial\phi(\eta)}{\partial\eta} = \frac{\partial\psi(h^{-1}(\eta))}{\partial\eta}.$$

To find the derviative of the inverse function h^{-1}, use the chain rule on

$$h(h^{-1}(\eta)) = \eta$$

$$h'(h^{-1}(\eta))\frac{\partial h^{-1}(\eta)}{\partial\eta} = 1$$

and we have $\frac{\partial h^{-1}(\eta)}{\partial\eta} = 1/h'(h^{-1}(\eta))$. Therefore, $\frac{\partial\phi(\eta)}{\partial\eta} = \psi'(h^{-1}(\eta))/h'(h^{-1}(\eta))$. Using the result from part a), the Fisher Information bound

$$B_q(\eta) = (\phi'(\eta))^2/I_q(\eta) = (\psi'(h^{-1}(\eta))/h'(h^{-1}(\eta)))^2 * [h'(h^{-1}(\eta))^2]/I_p(h^{-1}(\eta)),$$

which simplifies to $B_q(\eta) = (\psi'(h^{-1}(\eta))/I_p(h^{-1}(\eta)) = B_p(h^{-1}(\eta))$. Therefore, the Fisher information lower bound B is equivariant under monotone increasing transformations of the parameter.

5. **BD 3.5.1 if $n = 2k$ is even, give and plot the sensitivity curve of the median.**
Let $\theta = \widehat{X}$ be the sample median . Since $n = 2k$ is even, then $n-1$ is odd $\Rightarrow \widehat{\theta}(X_1, \dots, X_n) = X_{(k)}$.
The sensitivity curve is defined as

$$SC(X, \widehat{X}) = n\big(\widehat{\theta}(X_1, \dots, X_n, X) - \widehat{\theta}(X_1, \dots, X_{n-1})\big).$$

Case 1: $X < X_{(k-1)}$ $\widehat{\theta}(X_1, \dots, X_n, X) = \frac{X_{(k-1)}+X_{(k)}}{2}$
Case 2: $X_{(k-1)} \le X \le X_{(k+1)}$ $\widehat{\theta}(X_1, \dots, X_n, X) = \frac{X+X_{(k)}}{2}$
Case 3: $X > X_{(k+1)}$ $\widehat{\theta}(X_1, \dots, X_n, X) = \frac{X_{(k)}+X_{(k+1)}}{2}$
WLOG, assume $X_{(k)} = 0$ Then

$$SC(X, \widehat{X}) = \begin{cases} \frac{nX_{(k-1)}}{2} & \text{if} \quad X < X_{(k-1)} \\ \frac{nX}{2} & \text{if} \quad X_{(k-1)} \le X \le X_{(k+1)} \\ \frac{nX_{(k+1)}}{2} & \text{if} \quad X > X_{(k+1)} \end{cases}$$

The sensitivity curve is bounded as $|X| \to \infty$.

6. **BD 3.5.11 Let μ_0 be a hypothesized mean for a certain population. The (student) t-ratio is defined as $t = \sqrt{n}(\overline{X} - \mu_0)/s$, where $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Let $\mu_0 = 0$ and choose the ideal sample $X_1, \ldots, X_{n-1}$ to have sample mean zero. Find the limit of the sensitivity curve of t as**

Given the $t-$statistic $t_n = \frac{\sqrt{n}(\overline{X}-\mu_0)}{s}$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Let $\mu_0 = 0$ and $\frac{\sum_{i=1}^{n-1} X_i}{n-1} = 0$. $\widehat{\theta} = t$.

$$
\begin{aligned}
SC(t, \widehat{t}) &= n\Big(\widehat{\theta}(X_1, \ldots, X_{n-1}, X) - \widehat{\theta}(X_1, \ldots, X_{n-1})\Big) \\
&= n\Big(\widehat{\theta}(X_1, \ldots, X_{n-1}, X)\Big)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\theta}(X_1, \ldots, X_{n-1}, X) &= \frac{\sqrt{n}\overline{X}}{s} \\
&= \frac{\sqrt{n}\big(\frac{\sum_{i=1}^{n-1} X_i}{n} + \frac{X}{n} - 0\big)}{s} \\
&= \frac{\frac{X}{\sqrt{n}}}{\sqrt{\frac{\Sigma(X_i - \overline{X})^2}{n-1}}} \\
&= \frac{\frac{\sqrt{n-1}X}{\sqrt{n}}}{\sqrt{\Sigma(X_i - \overline{X})^2}}
\end{aligned}
$$

Note:

$$
\begin{aligned}
\Sigma(X_i - \overline{X})^2 &= \sum_{i=1}^{n-1}(X_i - \frac{X}{n})^2 + (X - \frac{X}{n})^2 \\
&= \sum_{i=1}^{n-1} X_i^2 - \frac{\sum_{i=1}^{n-1} X_i X}{n} + \frac{(n-1)X^2}{n^2} + X^2 - \frac{2X^2}{n} + \frac{X^2}{n^2} \\
&= \sum_{i=1}^{n-1} X_i^2 + \Big(\frac{n-1}{n}\Big)X^2
\end{aligned}
$$

Therefore

$$
SC(t, \widehat{t}) = n\frac{\frac{\sqrt{n-1}}{\sqrt{n}}X}{\sqrt{\sum_{i=1}^{n-1} X_i^2 + \Big(\frac{n-1}{n}\Big)X^2}}
$$

$$
= n\frac{\frac{\sqrt{n-1}}{\sqrt{n}}X}{|X|\sqrt{\sum_{i=1}^{n-1}\frac{X_i^2}{X^2} + \Big(\frac{n-1}{n}\Big)}} = n\frac{X}{|X|\sqrt{\frac{n}{n-1}\sum_{i=1}^{n-1}\frac{X_i^2}{X^2} + 1}}.
$$

(a) $|x| \to \infty$**,** n **is fixed.**

The limit is

$$\lim_{|X|\to\infty} n \frac{X}{|X|\sqrt{\frac{1}{X^2}\frac{n}{n-1}\sum_{i=1}^{n-1} X_i^2 + 1}} = \pm n,$$

where $\sum_{i=1}^{n-1} X_i^2$ is a constant.

The sensitivity curve is bounded as $|X| \to \infty$.

(b) $n \to \infty$**,** x **is fixed.**

$$\lim_{n\to\infty} n \frac{X}{|X|\sqrt{\frac{n}{n-1}\sum_{i=1}^{n-1} \frac{X_i^2}{X^2} + 1}} = \lim_{n\to\infty} n \frac{X}{|X|\sqrt{\frac{n}{X^2}\sum_{i=1}^{n-1} \frac{X_i^2}{n-1} + 1}}$$

Note that $\sum_{i=1}^{n-1} X_i^2$ can no longer be treated as a constant, since it is growing with n. However, $\frac{\sum_{i=1}^{n-1} X_i^2}{n-1} \to \sigma^2$ by the law of large numbers.

Making the approximation,

$$n \frac{X}{|X|\sqrt{\frac{n}{X^2}\sum_{i=1}^{n-1} \frac{X_i^2}{n-1} + 1}} \approx n \frac{X}{|X|\sqrt{\frac{n}{X^2}\sum_{i=1}^{n-1} \frac{X_i^2}{n-1}}} = \sqrt{n} \frac{X}{\sqrt{\sum_{i=1}^{n-1} \frac{X_i^2}{n-1}}}$$

Therefore,

$$\lim_{n\to\infty} = \sqrt{n} \frac{X}{\sqrt{\sum_{i=1}^{n-1} \frac{X_i^2}{n-1}}} = \infty \frac{\dot{X}}{\sigma} = \infty,$$

for $|\sigma| < \infty$ and $X \neq 0$. The sensitivity curve is not bounded as $n \to \infty$.