Solutions to Homework 5
BIOS 7731

1. **BD 3.2.3. In Problem 3.2.2 preceeding, give the MLE of the Bernoulli variance $q(\theta) = \theta(1-\theta)$ and gives the Bayes estimate of $q(\theta)$. Check whether $q(\widehat{\theta}_B) = E(q(\boldsymbol{\theta})|\mathbf{x})$, where $\widehat{\theta}_B$ is the Bayes estimate of θ.**

   The MLE of the Bernoulli random variable is $\overline{X}$. Then, by the invariance property of MLEs, the MLE estimate of the variance is $q(\widehat{\theta}_{MLE}) = \overline{X}\Big(1-\overline{X}\Big)$.

   From lecture, for squared loss the Bayes estimate is the posterior mean $\widehat{\theta}_B = \frac{\sum_{i=1}^n X_i + r}{r+s+n}$. Plug-in estimate for $q(\theta)$ is $(\widehat{\theta}_B)(1-\widehat{\theta}_B)$:

$$
\begin{aligned}
q(\widehat{\theta}) &= \Big(\frac{\sum_{i=1}^n X_i + r}{r+s+n}\Big)\Big(1 - \frac{\sum_{i=1}^n X_i + r}{r+s+n}\Big) \\
&= \frac{\Big(\sum_{i=1}^n X_i + r\Big)\Big(r+s+n-\sum_{i=1}^n X_i - r\Big)}{(r+s+n)^2} \\
&= \frac{\Big(\sum_{i=1}^n X_i + r\Big)\Big(n+s-\sum_{i=1}^n X_i\Big)}{(r+s+n)^2}
\end{aligned}
$$

   We also showed $\theta\Big|\vec{X} \sim Beta\Big(r + \sum_{i=1}^n X_i, n - \sum_{i=1}^n X_i + s\Big)$. To simplify our calculations, let $a = r + \sum_{i=1}^n X_i$ and $b = n - \sum_{i=1}^n X_i + s$. Recall $E[\theta|\vec{X}] = \frac{a}{a+b}$ and $Var(\theta|\vec{X}) = \frac{ab}{(a+b)^2(a+b+1)}$.

$$
\begin{aligned}
E\Big[q(\theta_B)\Big|\vec{X}\Big] &= E\Big[\theta\Big(1-\theta\Big)\Big|\vec{X}\Big] \\
&= E[\theta - \theta^2\Big|\vec{X}] \\
&= E[\theta|\vec{X}] - E[\theta^2|\vec{X}] \\
&= E[\theta|\vec{X}] - \Big(Var(\theta|\vec{X}) + \Big(E[\theta|\vec{X}]\Big)^2\Big) \\
&= \frac{a}{a+b} - \Big(\frac{ab}{(a+b)^2(a+b+1)} + \frac{a^2}{(a+b)^2}\Big) \\
&= \frac{a}{a+b}\Big(1 - \frac{b}{(a+b)(a+b+1)} - \frac{a}{a+b}\Big) \\
&= \frac{a}{a+b}\Big(\frac{a+b-a}{a+b} - \frac{b}{(a+b)(a+b+1)}\Big) \\
&= \frac{a}{(a+b)^2}\Big(b - \frac{b}{a+b+1}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{ab}{(a+b)^2}\Big(1 - \frac{1}{a+b+1}\Big) \\
&= \frac{ab}{(a+b)^2}\Big(\frac{a+b+1-1}{a+b+1}\Big) \\
&= \frac{ab}{(a+b)(a+b+1)} \\
&= \frac{\Big(r + \sum_{i=1}^{n} X_i\Big)\Big(n + s - \sum_{i=1}^{n} X_i\Big)}{(r+s+n)(r+s+n+1)}
\end{aligned}
$$

(Note, you can also use what we learned about $E[\theta(1-\theta)]$ for the Beta distribution - see hint from class problems)

Although the numerator is the same, the denominator is different, therefore $q(\widehat{\theta}_B) < E\Big[q(\theta)\Big|\vec{X}\Big]$. Unlike the MLE, you cannot just plug in the Bayes estimate into a function q to get the Bayes estimate of $q(\theta)$.

An alternative solution includes using Jensen's Inequality.

Another alternative solution is noticing that

$$E[q(\theta)|\vec{X}] = E[\theta(1-\theta)|\vec{X}] = E[\theta|\vec{X}] - E[\theta^2|\vec{X}] = q(\hat{\theta}_B) - Var(\theta|\vec{X}).$$

2. **BD 3.2.5. Suppose $\theta \sim \pi(\theta)$, $(X|\theta = \theta) \sim p(x|\theta)$.**

   (a) **Show that the density of X and θ is**

$$f(x, \theta) = p(x|\theta)\pi(\theta) = c(x)\pi(\theta|x)$$

   where $c(x) = \int \pi(\theta)p(x|\theta)d\theta$.

$$f(x, \theta) = p(x|\theta)\pi(\theta) = \pi(\theta|x)p(x) = \pi(\theta|x)\int f(x, \theta)d\theta$$
$$= \pi(\theta|x)\int p(x|\theta)\pi(\theta)d\theta = \pi(\theta|x)c(x)$$

   (b) **Let $l(\theta, a) = (\theta - a)^2/w(\theta)$ for some weight function $w(\theta) > 0, \theta \in \Theta$. Show that the Bayes rule is**

$$\delta^* = E_{f_0(x)}(\theta|x)$$

   where

$$f_0(x, \theta) = p(x|\theta)[\pi(\theta)/w(\theta)]/c$$

   and

$$c = \int\int p(x|\theta)[\pi(\theta)/w(\theta)]d\theta dx$$

   is assumed to be finite. That is if π and l are changed to $a(\theta)\pi(\theta)$ and $l(\theta, a)/a(\theta)$, $a(\theta) > 0$, respectively, the Bayes rule does not change.

   Minimizing this loss function (not shown) gives us the following Bayes rule:

$$\delta^* = \frac{E[\theta/w(\theta)|X]}{E[1/w(\theta)|X]}$$
$$= \frac{\int \theta\pi(\theta|x)/w(\theta)d\theta}{\int \pi(\theta|x)/w(\theta)d\theta}$$
$$= \frac{\int \theta p(x|\theta)\pi(\theta)/(c(x)w(\theta))d\theta}{\int p(x|\theta)\pi(\theta)/(c(x)w(\theta))d\theta}$$
$$= \frac{\int \theta p(x|\theta)\pi(\theta)/w(\theta)d\theta}{\int p(x|\theta)\pi(\theta)/w(\theta)d\theta}$$
$$= \frac{\int \theta p(x|\theta)\pi(\theta)/w(\theta)d\theta}{c}$$
$$= \int \theta f_0(x, \theta)d\theta$$
$$= E_{f_0}(\theta|x)$$

   Alternative solution is using Problem 1.4.24.

3. **BD 3.2.8**

(a) **Suppose that $N_1, \ldots, N_r$, given θ are multinomial $M(n, \theta)$, $\theta = (\theta_1, \ldots, \theta_r)^T$, and that θ has the Dirichlet distribution $\mathcal{D}(\alpha)$, $\alpha = (\alpha_1, \ldots, \alpha_r)^T$, defined in Problem 1.2.15. Let $q(\theta) = \sum_{j=1}^r c_j \theta_j$, where $c_1, \ldots, c_r$ are given constants. If $l(\theta, a) = [q(\theta) - a]^2$, find the Bayes decision rule δ^* and the minimum conditional Bayes risk $r(\delta^{*x}|x)$.**

Given,

$$p(n|\theta) = \frac{n!}{n_1! \ldots n_r!} \theta_1^{n_1} \ldots \theta_r^{n_r}$$

and

$$\pi(\theta) = \frac{\Gamma(\sum_{j=1}^r \alpha_j)}{\prod_{j=1}^r \Gamma(\alpha_j)} \theta_1^{\alpha_1 - 1} \ldots \theta_r^{\alpha_r - 1},$$

the posterior distribution of θ is

$$\pi(\theta|N = n) \propto p(n|\theta)\pi(\theta)$$
$$\propto \theta_1^{n_1 + \alpha_1 - 1} \ldots \theta_r^{n_r + \alpha_r - 1},$$

which is a Dirichlet distribution $\mathcal{D}(\beta)$ where $\beta = (n_1 + \alpha_1, \ldots, n_r + \alpha_r)^T$.

From class, for squared loss the Bayes decision rule that minimizes the posterior risk is

$$\delta^* = E[q(\theta)|X] = E[\sum_{j=1}^r c_j \theta_j |X] = \sum_{j=1}^r c_j E[\theta_j|X] = \sum_{j=1}^r c_j \frac{\beta_j}{\sum_{j=1}^r \beta_j} = \sum_{j=1}^r c_j \frac{\alpha_j + n_j}{a_0 + n},$$

where $\alpha_0 = \sum_{j=1}^r \alpha_j$ and $n = \sum_{j=1}^r n_j$.

The minimum conditional (posterior) Bayes risk is

$$r(\delta^*|X) = E[(q(\theta) - \delta^*)^2|X] = E[(q(\theta) - E[q(\theta)|X])^2|X] = Var(q(\theta)|X).$$

$$Var(q(\theta)|X) = Var(\sum_{j=1}^r c_j \theta_j |X)$$

$$= \sum_{j=1}^r c_j^2 Var(\theta_j|X) + 2 \sum_{j<k} c_j c_k Cov(\theta_j, \theta_k|X)$$

$$= \sum_{j=1}^r c_j^2 \frac{\beta_j(\beta_0 - \beta_j)}{\beta_0^2(\beta_0 + 1)} - 2 \sum_{j<k} c_j c_k \frac{\beta_j \beta_k}{\beta_0^2(\beta_0 + 1)}$$

$$= \frac{1}{\beta_0^2(\beta_0 + 1)} [\sum_{j=1}^r c_j^2 \beta_j^2 - (\sum_{j=1}^r c_j \beta_j)^2],$$

where $\beta_0 = \sum_{j=1}^r \beta_j = \alpha_0 + n$.

(b) **We want to estimate the vector** $(\theta_1, \ldots, \theta_r)$ **with loss function** $l(\theta, a) = \sum_{j=1}^{r} (\theta_j - a_j)^2$**. Find the Bayes decision rule.**

We minimize risk for a sum of squared losses by minimizing the sum of risks for each squared loss. Therefore the Bayes rule for each $j = 1, \ldots, r$ is the expected value of the posterior distribution for θ_j, that is,

$$\delta_j^* = \frac{n_j + \alpha_j}{n + \alpha_0}$$

and $\delta^* = (\delta_1^*, \ldots, \delta_r^*)^T$. Can show that this minimizes the posterior risk by taking the Hessian, or by arguing that each posterior mean minimizes the posterior risk.

4. **Consider a Bayesian model in which the random parameter Θ has a Bernoulli prior distribution with success probability $\frac{1}{2}$. That is,**

$$\pi(\theta) = \begin{cases} \frac{1}{2}, & \theta = 0; \\ \frac{1}{2}, & \theta = 1. \end{cases}$$

Given $\theta = 0$, the random variable X has density f_0 and given $\theta = 1$, X has density f_1.

(a) **Find the Bayes estimate (aka Bayes rule) of θ under squared loss.**

Under squared loss, the Bayes estimate is $E[\theta|X]$. First, find the posterior distribution of θ.

$$p(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\sum_\theta p(x|\theta)\pi(\theta)}$$

$$p(\theta = 0|x) = \frac{f_0(x)\frac{1}{2}}{f_0(x)\frac{1}{2} + f_1(x)\frac{1}{2}} = \frac{f_0(x)}{f_0(x) + f_1(x)}$$

$$p(\theta = 1|x) = \frac{f_1(x)\frac{1}{2}}{f_0(x)\frac{1}{2} + f_1(x)\frac{1}{2}} = \frac{f_1(x)}{f_0(x) + f_1(x)}$$

Then,

$$E[\theta|X] = \sum_\theta \theta p(\theta|x) = 0 * \frac{f_0(x)}{f_0(x) + f_1(x)} + 1 * \frac{f_1(x)}{f_0(x) + f_1(x)} = \frac{f_1(x)}{f_0(x) + f_1(x)}.$$

(b) **Find the Bayes estimate (aka Bayes rule) of θ if $l(\theta, d) = I\{\theta \neq d\}$ (zero-one loss).**

To find the Bayes estimate, we need to minimize the posterior risk

$$E[l(\theta, d)|X] = \sum_\theta I\{\theta \neq d\} p(\theta|x)$$

$$= I\{0 \neq d\} p(\theta = 0|x) + I\{1 \neq d\} p(\theta = 1|x)$$

When $d = 1$,

$$E[L(\theta, 1)|X] = 1 * p(\theta = 0|x) + 0 * p(\theta = 1|x) = \frac{f_0(x)}{f_0(x) + f_1(x)}.$$

When $d = 0$,

$$E[L(\theta, 1)|X] = 0 * p(\theta = 0|x) + 1 * p(\theta = 1|x) = \frac{f_1(x)}{f_0(x) + f_1(x)}.$$

This is minimized by selecting $d = 1$ when $f_0(x) < f_1(x)$ and $d = 0$ when $f_1(x) < f_0(x)$. Therefore, the Bayes rule d is

$$\delta(x) = 1 \ if \ f_1(x) > f_0(x)$$

$$\delta(x) = 0 \ if \ f_1(x) < f_0(x).$$

If $f_1(x) = f_0(x)$ can randomly select $\delta(x) = 1$ or 0 with equal probability.

5. **Let $X_1, \dots X_n$ be iid Uniform$(0, \theta)$, with $\theta > 0$.**

(a) **Show that the density of the largest order statistic $X_{(n)}$ is $p(x, \theta) = n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$, where $I_{(0,\theta)}(x)$ is an indicator for $0 < x < \theta$.**

Using Casella & Berger (CB Thm 5.4.4) the pdf of the nth order statistic is

$$p(x, \theta) = nf_X(x)F_X(x)^{n-1}.$$

Here $f_X(x) = \frac{1}{\theta}I_{(0,\theta)}(x)$ and $F_X(x) = \frac{x}{\theta}I_{(0,\theta)}(x)$, therefore

$$p(x, \theta) = n\theta^{-n}x^{n-1}I_{(0,\theta)}(x).$$

(b) **Find an unbiased estimator for θ based on $X_{(n)}$ and determine its variance.**

$$E[X_{(n)}] = \int_0^\theta nx\theta^{-n}x^{n-1}dx = \frac{n\theta}{n+1}.$$

Therefore, an unbiased estimator for θ based on $X_{(n)}$ is $\hat{\theta} = \frac{n+1}{n}X_{(n)}$. The variance of $\hat{\theta} = \frac{n+1}{n}X_{(n)}$ is

$$\frac{(n+1)^2}{n^2}Var(X_{(n)}) = \frac{(n+1)^2}{n^2}(E[X_{(n)}^2] - E[X_{(n)}]^2).$$

Since,

$$E[X_{(n)}^2] = \int_0^\theta x^2n\theta^{-n}x^{n-1}dx = \frac{n\theta^2}{n+2},$$

then

$$\frac{(n+1)^2}{n^2}Var(X_{(n)}) = \frac{(n+1)^2}{n^2}(\frac{n\theta^2}{n+2} - (\frac{n\theta}{n+1})^2),$$

which simplifies to $\frac{\theta^2}{n(n+2)}$.

(c) **Find the Fisher information matrix $I(\theta)$. Show that the Fisher information inequality does NOT hold for the UMVU estimator in b).**

Since $X_{(n)}$ is complete and sufficient for θ (see CB Example 6.2.23) any function of $X_{(n)}$ that is an unbiased estimator for θ is UMVU (see CB Thm 7.3.23). Therefore, $\hat{\theta} = \frac{n+1}{n}X_{(n)}$, is UMVU.

For an exponential family or if assumptions I and II hold, the Fisher information inequality holds and the variance for an unbiased estimator would be bounded by $I^{-1}(\theta)$, where

$$I(\theta) = E[(\frac{d}{d\theta}\log p(X, \theta))^2].$$

For the uniform case, assumptions I and II do not hold and this is not an exponential family. Although it won't have the same properties as the Fisher Information Number, we still calculate the expectation $E[(\frac{d}{d\theta}\log p(X,\theta))^2]$.
First,

$$\log p(X,\theta) = -\log\theta$$

and

$$I(\theta) = E[(\frac{d}{d\theta}\log p(X,\theta))^2] = \frac{1}{\theta^2}$$

Using $I_n(\theta) = nI(\theta)$, the Fisher Information Inequality does not hold

$$\frac{\theta^2}{n} > \frac{\theta^2}{n(n+2)}.$$

Although we cannot use the inequality to show UMVU for $\hat{\theta} = \frac{n+1}{n}X_{(n)}$, it is still UMVU. Since $X_{(n)}$ is complete and sufficient for θ any function of $X_{(n)}$ that is an unbiased estimator for θ is UMVU.

6. **Let** $X_1, \ldots, X_n$ **be iid** $N(\theta, 1)$

(a) **Show that the unbiased estimator of** θ^2 **is** $\overline{X}^2 - \frac{1}{n}$**.**

$$\begin{aligned}
E\Big[\overline{X}^2 - \frac{1}{n}\Big] &= E\big[\overline{X}^2\big] - \frac{1}{n} \\
&= Var(\overline{X}) + \Big(E[\overline{X}]\Big)^2 - \frac{1}{n} \\
&= Var\Big(\sum_{i=1}^{n} \frac{X_i}{n}\Big) + \Big(E\Big[\sum_{i=1}^{n} \frac{X_i}{n}\Big]\Big)^2 - \frac{1}{n} \\
&= \frac{nVar(X_i)}{n^2} + \Big(\frac{nE[X_i]}{n}\Big)^2 - \frac{1}{n} \\
&= \frac{n}{n^2} + \Big(\frac{n\theta}{n}\Big)^2 - \frac{1}{n} \\
&= \frac{1}{n} + \theta^2 - \frac{1}{n} \\
&= \theta^2
\end{aligned}$$

Therefore $\overline{X}^2 - \frac{1}{n}$ is the unbiased estimator of θ^2.

(b) **Calculate its variance and show that is greater than the Information Inequality Bound**.

$$\begin{aligned}
Var\Big(\overline{X}^2 - \frac{1}{n}\Big) &= Var\Big(\overline{X}^2\Big) \\
&= E[\overline{X}^4] - \big(E[\overline{X}^2]\big)^2
\end{aligned}$$

Aside:

$$\begin{aligned}
E[\overline{X}^2] &= Var(\overline{X}) + \big(E[\overline{X}]\big)^2 \\
&= \frac{1}{n} + \theta^2
\end{aligned}$$

$$\begin{aligned}
E[\overline{X}^4] &= E[\overline{X}^3(\overline{X} - \theta + \theta)] \\
&= E[\overline{X}^3(\overline{X} - \theta)] + \theta E[\overline{X}^3] \\
&= \frac{1}{n}E[3\overline{X}^2] + \theta\Big(E[\overline{X}^2(\overline{X} - \theta + \theta)]\Big) \\
&= \frac{3}{n}E[\overline{X}^2] + \theta\Big(E[\overline{X}^2(\overline{X} - \theta)] + \theta E[\overline{X}^2]\Big) \\
&= \frac{3}{n}E[\overline{X}^2] + \theta\Big(\frac{1}{n}E[2\overline{X}] + \theta E[\overline{X}^2]\Big) \\
&= \Big(\frac{3}{n} + \theta^2\Big)E[\overline{X}^2] + \frac{2\theta}{n}E[\overline{X}] \\
&= \Big(\frac{3}{n} + \theta^2\Big)E[\overline{X}^2] + \frac{2\theta^2}{n}
\end{aligned}$$

Note: $E[\overline{X}^4]$ is computed by applying Stein's Lemma two times.
So the $Var\Big(\overline{X}^2 - \frac{1}{n}\Big)$ equals

$$\begin{aligned}
Var\Big(\overline{X}^2 - \frac{1}{n}\Big) &= \frac{2\theta^2}{n} + \Big(\frac{3}{n} + \theta^2\Big)E[\overline{X}^2] - \Big(E[\overline{X}^2]\Big)^2 \\
&= \frac{2\theta^2}{n} + \Big(\frac{3}{n} + \theta^2 - E[\overline{X}^2]\Big)E[\overline{X}^2] \\
&= \frac{2\theta^2}{n} + \Big(\frac{3}{n} + \theta^2 - \Big(\frac{1}{n} + \theta^2\Big)\Big)\Big(\frac{1}{n} + \theta^2\Big) \\
&= \frac{2\theta^2}{n} + \frac{2}{n}\Big(\frac{1}{n} + \theta^2\Big) \\
&= \frac{2\theta^2}{n} + \frac{2}{n^2} + \frac{2\theta^2}{n} \\
&= \frac{4\theta^2}{n} + \frac{2}{n^2}
\end{aligned}$$

Information Inequality lower bound is $\frac{\big(\Psi'(\theta)\big)^2}{nI_1(\theta)}$ where $\Psi(\theta) = E[\overline{X}^2] = Var(\overline{X}) + \Big(E[\overline{X}]\Big)^2 = \frac{1}{n} + \theta^2$:

$$\begin{aligned}
I_1(\theta) &= -E\Big[\frac{\partial^2 log(p(X_1,\theta)}{\partial\theta^2}\Big] \\
&= -E\Big[\frac{\partial^2\Big(-log\sqrt{2\pi} - \frac{1}{2}(X_1-\theta)^2\Big)}{\partial\theta^2}\Big] \\
&= -E\Big[\frac{\partial(X_1-\theta)}{\partial\theta}\Big] \\
&= 1
\end{aligned}$$

The Information Inequality lower bound is $\frac{\big(\Psi'(\theta)\big)^2}{nI_1(\theta)} = \frac{(2\theta)^2}{n} = \frac{4\theta^2}{n}$. The variance of the estimator does not meet the inequality bound, $Var\Big(\overline{X}^2 - \frac{1}{n}\Big) = \frac{4\theta^2}{n} + \frac{2}{n^2} > \frac{4\theta^2}{n}$ since $\frac{2}{n^2} > 0$ and it is a strict inequality.