BIOS 7731 HW 5 Tim Vigers

3.2.3) Let $X_1, \dots, X_n$ be iid Bern(θ) with the prior $\pi(\theta)$ is Beta(r, s) density. Also, suppose that $\ell(\theta, a)$ is the quadratic loss function $(\theta - a)^2$.

Due to the invariance property of MLEs, the MLE for $q(\theta) = \theta(1-\theta)$ is $q(\hat{\theta})$ where $\hat{\theta}$ is the MLE of $\hat{\theta}$.

First find the MLE for θ by setting the derivative of the log likelihood $(\ell(\theta))$ equal to 0 and solving for $\hat{\theta}$:

$$L_x(\theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i} \quad \text{(Likelihood)}$$

$$\ell_x(\theta) = \Sigma x_i \log(\theta) + (n - \Sigma x_i)\log(1-\theta) \quad \text{(log likelihood)}$$

$$d/d\theta\ \ell_x(\theta) = \frac{\Sigma x_i}{\theta} - \frac{n}{1-\theta} + \frac{\Sigma x_i}{1-\theta}$$

Setting this equal to 0 and solving for θ gives us $\hat{\theta} = \Sigma x_i / n = \bar{X}$. Thus, the MLE for $q(\theta)$ is $\boxed{q(\hat{\theta}) = \bar{X}(1-\bar{X})}$

Next, find the Bayes estimate of $q(\theta)$. Because we are using squared loss, we know that the Bayes estimate $\delta^*(x) = E[q(\theta)|X]$ by BD 3.2.5. Therefore we have:

$$\delta^*(x) = E[\theta(1-\theta)|X] = E[\theta|X] - E[\theta^2|X]$$ ✓

From class, we know that the Bayes estimate of θ, $\hat{\theta}_B$ is $E[\theta|X]$. So,

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

Therefore, the Bayes estimate of $q(\theta) \neq q(\hat{\theta}_B)$ unless $\boxed{E[\theta^2|X] = E[\theta|x]^2.}$

3.2.5 a) Suppose $\theta \sim \pi(\theta)$ and $(X|\theta=\theta) \sim p(x|\theta)$. Using Bayes' Theorem we know that

$$p(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int_t p(x|t)\pi(t)\,dt}$$

If we let $c(x) = \int_t p(x|t)\pi(t)\,dt$, this can be rearranged:

$$p(x|\theta)\pi(\theta) = p(\theta|x)\, c(x). \qquad \square$$

b) Let $\ell(\theta, a) = \frac{(\theta - a)^2}{w(\theta)}$ for $w(\theta) > 0$

which is analogous to the weighted squared prediction error. To find the Bayes estimate, we want to minimize $E[\ell(\theta, a) | X]$:

$$\frac{\int \ell(\theta, a)\, p(x|\theta)\pi(\theta)\, d\theta}{\int p(x|\theta)\pi(\theta)\, d\theta}$$

which can be re-written:

$$\frac{\int (\theta - a)^2 / w(\theta)\; p(x|\theta)\pi(\theta)\, d\theta}{\int p(x|\theta)\pi(\theta)\, d\theta}$$

If we introduce a normalizing constant $c = \iint p(x|\theta)\, \pi(\theta)/w(\theta)\, d\theta$, Formula above becomes:

$$\frac{\int (\theta - a)^2\, p(x|\theta)\, \pi(\theta)/w(\theta) / c\; d\theta}{\int p(x|\theta)\pi(\theta)/c\; d\theta}$$

This is now a case of squared loss for the density $f_0(x, \theta) = p(x|\theta)\, \pi(\theta)/w(\theta) / c$ (see 1.4.24 for the analogous WSPE approach).

Thus, by BD 3.2.5. we know that the Bayes estimate is $E_{F_0}[\theta | X]$. $\square$

3.2.8. a) Suppose that $X_1, \dots, X_r$ given θ are multinomial $M(n, \theta)$ with $\theta = (\theta_1, \dots, \theta_r)^T$ and that θ has the prior distribution $D(\alpha)$, with $\alpha = (\alpha_1, \dots, \alpha_r)^T$. Let $q(\theta) = \sum_{j=1}^{r} c_j \theta_j$ where $c_1, \dots, c_j$ are known constants. Assuming quadratic loss, find the Bayes rule δ^*.

Because this is again using quadratic loss, the Bayes rule δ^* is $E[q(\theta)|X]$. Also, because the Dirichlet distribution is a conjugate prior (BD problem 1.2.15), we know that $p(\theta | x) \sim D(\alpha + x)$ with $\alpha = (\alpha_1, \dots, \alpha_r)^T$ and $x = (x_1, \dots x_r)^T$. Therefore,

$$E[q(\theta)|X] = E\left[\sum_{j=1}^{r} c_j \theta_j | X\right] = \sum_{j=1}^{r} c_j E[\theta_j | X]$$

Based on the problem hint, this means that the Bayes rule is:

$$\delta^*(x) = \sum_{j=1}^{r} c_j \frac{\alpha_j + x_j}{\alpha_0} \quad \text{where} \quad \alpha_0 = \sum_{j=1}^{r} \alpha_j + x_j,$$

Based on BD example 3.2.1, we know that the Bayes risk $r(\pi, \delta^*)$ is

$$E\left[\left(q(\theta) - E[q(\theta)|X]\right)^2\right] = E\left[E\left[\left(q(\theta) - E[q(\theta)|X]\right)\right]^2 | X\right]$$

In other words, this is the expected value of the posterior variance:

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

where $\beta_j = x_j + \alpha_j$ and $\beta_0 = \sum_{j=1}^{r} \beta_j$. Because this isn't a function of θ,

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

c) To estimate vector $(\theta_1, \ldots, \theta_r)$ with $\ell(\theta, a) = \sum_{j=1}^{r} (\theta_j - a_j)^2$, we simply need to find a general formula to minimize $(\theta_j - a_j)^2$. Minimizing each part of the sum will also minimize the sum. Once again we have quadratic loss, so we have:

$$\delta^*(\underset{\sim}{x}) = E[\theta_j | X]$$

We know that $E[\theta_j \mid X] = \dfrac{\alpha_j + x_j}{\sum_{j=1}^{r} \alpha_j + x_j}$

So the Bayes decision rule for $(\theta_1, \dots, \theta_r)^T$ is $\left(E[\theta_1 \mid X], \dots, E[\theta_r \mid X]\right)$.

4. Consider a Bayesian model in which the parameter θ has a prior distribution of Bern$(1/2)$. Given $\theta = 0$ the R.V. X has density $f_0(x)$ and given $\theta = 1$ X has density $f_1(x)$.

First we write the marginal distribution of X:

$$p(x) = \sum_{\theta} p(x \mid \theta)\pi(\theta) = \sum_{\theta} f_0(x) 1(\theta = 0) + f_1(x) 1(\theta = 1)(1/2)^{\theta}(1/2)^{1-\theta}$$

which simplifies to:

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

Next we use this to find the posterior distribution of θ:

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

$$= \frac{f_0(x)\,1(\theta=0) + f_1(x)\,1(\theta=1)\,\cancel{1/2}}{f_0(x) + f_1(x)\quad \cancel{1/2}}$$

Assuming squared loss, we can now find the Bayes estimate $\delta^*(x) = E[\theta \mid X]$:

$$\delta^*(x) = \sum_\theta \frac{\theta\left(f_0(x)1(\theta=0) + f_1(x)1(\theta=1)\right)}{f_0(x) + f_1(x)}$$

$$= \frac{1}{f_0(y)+f_1(y)}\left(0 + f_1(x)\right) = \frac{f_1(x)}{f_0(x)+f_1(x)}$$

b) To find the Bayes estimate using the loss function $\ell(\theta, d) = 1(\theta \neq d)$, we need to minimize $E[\ell(\theta,d) \mid X]$, which is

$$\sum_\theta \ell(\theta,d)\,p(\theta \mid x)$$

So essentially we have a 2×2 table:

	$\theta = 1$	0
$d = 1$	0	1
0	1	0

The expected loss for $d=0$ is:

$$\sum_{\theta} \frac{1(\theta \neq 0)(f_0(x)1(\theta=0)+f_1(x)1(\theta=1))}{f_0(x)+f_1(x)}$$

which equals $\frac{f_1(x)}{f_0(x)+f_1(x)}$ ✓

The expect loss for $d=1$ is:

$$\sum_{\theta} \frac{1(\theta \neq 1)(f_0(x)1(\theta=0)+f_1(x)1(\theta=1))}{f_0(x)+f_1(x)}$$

which equals $\frac{f_0(x)}{f_0(x)+f_1(x)}$ ✓

Therefore $d=0$ incurs less loss when $f_1(x) < f_0(x)$ and $d=1$ is "better" when $f_0(x) < f_1(x)$.

✓

5. Let $X_1, \ldots, X_n$ be iid Uniform $(0, \theta)$ with $\theta > 0$. By Casella & Berger Theorem 5.4.4 we know the density of the j^{th} order statistic X_j is:

$$F_{X_j}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j}$$

So, the distribution of the n^{th} order statistic from a uniform $(0, \theta)$ is:

$$f_{X_n}(x) = n \frac{1}{\theta}\left(\frac{x}{\theta}\right)^{n-1} I^{x}_{(0,\theta)}.$$

Because X_n is complete and sufficient, we know that to find the UMVU estimator we simply need to find a function of X_n, $T(X)$, with expected value equal to θ. First, find the expected value of X_n:

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

$$= \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{\theta^n} \left(\frac{x^{n+1}}{n+1} \Big|_0^\theta \right)$$

So, $E[X_n] = \theta \dfrac{n}{n+1}$ and an unbiased estimator as a function of X_n is:

$$\frac{n+1}{n} X_n = \hat{\theta}$$

Next, we find the Fisher information number:

$$I(\theta) = E_\theta \left[\left(\frac{d}{d\theta} \log f(x|\theta) \right)^2 \right]$$

I find it easiest to work "inside out" with these kinds of derivation:

$$\log f(x|\theta) = \log(n) - n \log(\theta) + (n-1) \log(x)$$

Take the derivative:

$$\frac{d}{d\theta} \log f(x|\theta) = -\frac{n}{\theta}$$

Therefore,

$$I(\theta) = \left(\frac{-n}{\theta}\right)^2 = \frac{n^2}{\theta^2}$$ ✓

In order to examine the Fisher infomation inequality:

$$Var(T(x)) \geq \frac{\left(\frac{d}{d\theta} E[T(x)]\right)^2}{I(\theta)}$$

We next need to find the variance of our unbiased estimator:

$$Var(\hat{\theta}) = Var\left(\tfrac{n+1}{n} X_n\right) = \left(\frac{n+1}{n}\right)^2 Var(X_n)$$

The variance of X_n is:

$$Var(X_n) = E[X_n^2] - E[X_n]^2$$

$E[X_n^2]$ is derived with the same approach I used for $E[X_n]$, and the variance of X_n comes out to:

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

So, the variance of our unbiased estimator $\hat{\theta} = T(X) = {}^{n+1}/_{n} X_n$ is:

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

Plugging this into the information inequality we have:

$$\frac{\theta^2}{n^2+2n} \geq \frac{1}{I(\theta)} = \frac{\theta^2}{n^2}$$

Clearly, this inequality does not hold.

∴ This is because the support of a $U(0,\theta)$ distribution depends on the parameter θ, and therefore the assumptions required for using the information inequality are violated.

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

a) To show that $\bar{X}^2 - 1/n$ is an unbiased estimator of θ^2, first find the expected value of $\bar{X}^2$:

$$E[\bar{X}^2] = Var[\bar{X}] + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \theta^2$$

Because $\sigma^2 = 1$, $E[\bar{X}^2] = \theta^2 + \frac{1}{n}$ which means $E[\bar{X}^2 - 1/n] = \theta^2$.

b) To calculate the variance of $\bar{X}^2 - \frac{1}{n}$, we simply need to find $Var[\bar{X}^2]$ as $1/n$ is a constant:

$$Var[\bar{X}^2] = E[\bar{X}^4] - E[\bar{X}^2]^2$$

We already know the second term from above, and can find $E[\bar{X}^4]$ using Stein's lemma iteratively:

$$E[\bar{X}^4] = E[\bar{X}^3(\bar{X} - \theta + \theta)] = E[\bar{X}^3(\bar{X} - \theta)] + E[\theta \bar{X}^3]$$

$$E[\bar{X}^3(\bar{X} - \theta)] = 1/n \; E[3\bar{X}^2] \text{ by Stein's lemma.}$$
$$= 3/n \left(\theta^2 + \frac{1}{n}\right)$$

Next we use Stein's lemma again on the

second term:

$$\theta E[\bar{x}^3] = \theta E[\bar{x}^2(\bar{y} - \theta + \theta)] =$$
$$\theta\left(E[\bar{x}^2(\bar{y} - \theta)] + E[\theta \bar{x}^2]\right) =$$
$$\theta\left(\tfrac{1}{n} E[2\bar{x}] + \theta(\theta^2 + \tfrac{1}{n})\right)$$

Combining these two parts gives us:

$$E[\bar{x}^4] = \frac{6\theta^2}{n} + \frac{3}{n^2} + \theta^4$$

So, we have:

$$Var(\bar{x}^4) = \frac{6\theta^2}{n} + \frac{3}{n^2} + \theta^4 - \left(\theta^2 + \frac{1}{n}\right)^2$$

which simplifies to:

$$\frac{4\theta^2}{n} + \frac{2}{n^2}$$

Next we need to find the information number $I(\theta)$:

$$I(\theta) = E\left[\left(\frac{d}{d\theta} \log f(x|\theta)\right)^2\right]$$

Again, working "inside out":

$$\log f(x|\theta) = \frac{-(x-\theta)^2}{2} - \frac{1}{2}\log(2\pi)$$

Take the derivative:

$$d/d\theta \log f(x|\theta) = x - \theta$$

Square it:

$$\left(\frac{d}{d\theta} \log f(x|\theta)\right)^2 = (x-\theta)^2$$

Take the expected value to find $I(\theta)$:

$$I(\theta) = \frac{1}{4\sigma^2}$$

Because we did this for a single X_i, we can multiply by n using BD proposition 3.4.2:

$$I(\theta) = \frac{n}{4\sigma^2}$$

✓

Finally, we compare the variance of our estimator to the information inequality bound:

$$\mathrm{Var}(T(X)) \geq \frac{1}{I(\theta)}$$

$$\frac{4\theta^2}{n} + \frac{2}{n^2} \geq \frac{4\theta^2}{n}$$

$\therefore$ The variance of this estimator is greater than the information inequality bound.