APPM 5720

Solutions to Problem Set Four

1. The Bayes rule is the δ that minimizes the Bayes risk

$$
\begin{aligned}
R_\delta &= \mathsf{E}[L(\Theta, \delta(X))] = \int\int L(\theta, \delta(x)) f(\theta|x) f(x)\, d\theta\, dx \\
&= \int\int \tfrac{(\delta(x)-\theta)^2}{\theta(1-\theta)} f(\theta|x) f(x)\, d\theta\, dx \\
&= \int \left[\int \tfrac{(\delta(x)-\theta)^2}{\theta(1-\theta)} f(\theta|x)\, d\theta\right] dx\, dx
\end{aligned}
$$

To mimimize the Bayes risk with respect to δ, it is sufficient to minimize the inner integral with respect to δ. We know, in the case of mean squared error loss, that the minimizing δ is the posterior Bayes estimator for θ. However, on this problem, the loss function is close to, but not quite the same as, mean squared error. One can show that the posterior distribution for θ given x is $Beta(x+1, n-x+1)$. The inner integral is

$$\int \frac{(\delta(x)-\theta)^2}{\theta(1-\theta)} \frac{1}{\mathcal{B}(x+1, n-x+1)}\, \theta^x (1-\theta)^{n-x}\, d\theta.$$

Pulling that $\theta(1-\theta)$ denominator over, we have

$$\int (\delta(x)-\theta)^2 \frac{1}{\mathcal{B}(x+1, n-x+1)}\, \theta^{x-1} (1-\theta)^{n-x-1}\, d\theta.$$

which is starting to look like integration against a different Beta pdf. We may write the integral as

$$\frac{\mathcal{B}(x, n-x)}{\mathcal{B}(x+1, n-x+1)} \int (\delta(x)-\theta)^2 \frac{1}{\mathcal{B}(x, n-x)}\, \theta^{x-1} (1-\theta)^{n-x-1}\, d\theta.$$

This now looks like the expected mean squared error loss against a $Beta(x, n-x)$ "posterior". We know then that this is minimized, with respect to δ by taking δ to be the "posterior" mean. Thus, we get

$$\boxed{\delta = \delta(X) = \mathsf{E}[\text{“}\Theta\text{”}|X] = \frac{X}{X+n-X} = \frac{X}{n}.}$$

---

2. The risk associated with the first decision rule/estimator is

$$R_{\delta_1}(\sigma^2) = \mathsf{E}[(S^2 - \sigma^2)^2]$$

Since S^2 is an unbiased estimator of σ^2, this is just the variance of S^2. This variance is difficult to compute for a general distribution for the X_i. However, for the normal case we have here $(X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2))$ we know that $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$. Thus, we have that

$$
\begin{aligned}
R_{\delta_1}(\sigma^2) &= Var[S^2] = Var\left[\tfrac{\sigma^2}{n-1} \cdot \tfrac{(n-1)}{\sigma^2} S^2\right] \\
&= \tfrac{(\sigma^2)^2}{(n-1)^2} Var\left[\tfrac{(n-1)S^2}{\sigma^2}\right] \\
&= \tfrac{(\sigma^2)^2}{(n-1)^2} \cdot 2(n-1)
\end{aligned}
$$

since $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ and the variance of a χ^2 random variable is 2 times its degrees of freedom parameter.

In summary, we have

$$R_{\delta_1}(\sigma^2) = \frac{2(\sigma^2)^2}{n-1}.$$

Let $S_2^2 = \frac{n-1}{n}S^2$. Note that this is no longer an unbiased estimator of σ^2. The frequentist risk is

$$R_{\delta_2}(\sigma^2) = \mathsf{E}[(S_2^2 - \sigma^2)^2].$$

This is not the variance of S_2^2 since the mean of S_2^2 is not σ^2. It is, however, the mean squared error of the estimator S_2^2. So, we have

$$R_{\delta_2}(\sigma^2) = MSE(S_2^2) = Var[S_2^2] + [B(S_2^2)]^2$$

where $B(S_2^2)$ is the bias:

$$B(S_2^2) = \mathsf{E}[S_2^2] - \sigma^2 = \mathsf{E}\left[\frac{n-1}{n}S^2\right] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2.$$

Now,

$$Var[S_2^2] = Var\left[\frac{n-1}{n}S^2\right] = \left(\frac{n-1}{n}\right)^2 Var[S^2] = \left(\frac{n-1}{n}\right)^2 \cdot \frac{2(\sigma^2)^2}{n-1}.$$

Simplifying and putting this all together, we get

$$R_{\delta_2}(\sigma^2) = \frac{(2n-1)(\sigma^2)^2}{n^2}.$$

One can show formally that

$$\frac{2n-1}{n^2} < \frac{2}{n-1}.$$

for all $n = 1, 2, \ldots$ At least covince yourself by pluginng in some values for n and/or plotting both.

Thus, we have that $R_{\delta_2}(\sigma^2) < R_{\delta_1}(\sigma^2)$. So, δ_2 dominates δ_1 and therefroe δ_1 is inadmissible.

---

3. Correction: Unless you want to assume that θ is discrete (which is fine), the loss function should be given by a Dirac delta function. A Dirac delta function at a point a, which we will denote as $\Delta_a(x)$ is a function that is zero everywhere except for the point $x = a$ where it has an infinite spike. However, the function, by definition, will integrate to 1 and, when integrated against another function f, will satisfy

$$\int_{-\infty}^{\infty} \Delta_a(x) f(x)\, dx = f(a).$$

The Dirac delta function is usually denoted with a lowercase delta, but we are reserving δ for our decision function.

For this problem in the continuous θ setting, we should take the loss function to be

$$L(\theta, \delta) = 1 - \Delta_\theta(\delta)$$

so that, when $\delta = \theta$ (good estimate/decision), we get 0 loss and we get a loss of 1 for every other decision.

The Bayes risk is then

$$R_\delta = \int \int L(\theta, \delta) f(\theta|\vec{x})\, d\theta\, f(\vec{x})\, d\vec{x}.$$

The inner integral is

$$\int L(\theta,\delta) f(\theta|\vec{x})\,d\theta = \int [1-\Delta_\delta(\theta)] f(\theta|\vec{x})\,d\theta$$

$$\int f(\theta|\vec{x})\,d\theta - \int \Delta_\delta(\theta) f(\theta|\vec{x})\,d\theta = 1 - f(\delta|\vec{x}).$$

So, the Bayes risk is minimized at the δ where $f(\delta|\vec{)}$ (which is $f(\theta|\vec{x})$ with δ plugged in) is maximized. That is, the Bayes rule/estimator is

$$\delta^* = \arg\max_\theta f(\theta|\vec{x})$$

which is the mode of the posterior distribution and hence the MAP estimator.

---

4. Suppose that δ^* is not admissible. Then there exists a decision rule δ such that

$$R_\delta(\theta) \leq R_{\delta^*}(\theta) \quad \forall \theta \in \Omega$$

and

$$R_\delta(\theta_0) < R_{\delta^*}(\theta_0)$$

for at least one $\theta_0 \in \Omega$.

Let

$$c := R_{\delta_*}(\theta_0) - R_\delta(\theta_0) > 0.$$

By continuity of $R_\delta(\theta)$ for all δ, there exists an $\varepsilon > 0$ such that

$$R_{\delta_*}(\theta) - R_\delta(\theta) > c/2$$

for all θ in $A_\varepsilon := \{\theta : |\theta - \theta_0| < \varepsilon\}$.

In this case, we have

$$\begin{aligned} R_{\delta^*} - R_\delta &= \int [R_{\delta^*}(\theta) - R_\delta(\theta)]\, f(\theta)\,d\theta \\ &= \int_A [R_{\delta^*}(\theta) - R_\delta(\theta)]\, f(\theta)\,d\theta + \int_{A^c} [R_{\delta^*}(\theta) - R_\delta(\theta)]\, f(\theta)\,d\theta \\ &\geq \int_A [R_{\delta^*}(\theta) - R_\delta(\theta)]\, f(\theta)\,d\theta \\ &> \tfrac{c}{2} \int_A f(\theta)\,d\theta > 0 \end{aligned}$$

since f is strictly positive on all of the parameter space by assumption.

So, we have that the Bayes risks are ordered as $R_{\delta^*} > R_\delta$ which contradicts the fact that δ^* is the Bayes rule.

Thus, we must have that δ^* is admissible.