36-705 Intermediate Statistics Homework #10 Solutions

December 8, 2016

Problem 1 [20 pts.]

By setting the first derivative of the loss function equal to 0 we obtain:

$$
\begin{aligned}
&\frac{\partial R(\beta)}{\partial \beta} = 0 \\
\Longrightarrow\ & \mathbb{E}\Big[-2X(Y-\beta^T X)\Big] = 0 \\
\Longrightarrow\ & 2\Lambda\beta - 2\alpha = 0 \\
\Longrightarrow\ & \beta_* = \Lambda^{-1}\alpha.
\end{aligned}
$$

The loss function $R(\beta)$ is strictly convex so β_* is its unique minimum.

Problem 2 [25 pts.]

Recall that

$$\mathbb{E}[|Y - g(X)|] = \mathbb{E}\{\mathbb{E}[|Y - g(X)| \,\big|X]\}.$$

The idea is to choose c such that $\mathbb{E}[\big|Y - c\big| \mid X = x]$ is minimized. Now define:

$$r(c) = \mathbb{E}[|Y - c| \,\big|\, X = x] = \int |y - c| p_{Y|X=x}(y)dy.$$

The function $h_y(c) = |y - c|$ is differentiable everywhere except when $y = c$. Thus for $c \neq y$

$$h_y'(c) = \begin{cases} 1 & c > y \\ -1 & c < y \end{cases} = \mathbb{1}(c > y) - \mathbb{1}(c < y).$$

Since Y is continuous and has a density function, $P(Y = c) = 0$. So to minimize $r(c)$ we can differentiate under the integral sign and set the derivative equal to 0 to obtain:

$$\begin{aligned} r'(c) &= \int h_y'(c) p_{Y|X=x}(y)dy = \int_{-\infty}^{c} p_{Y|X=x}(y)dy - \int_{c}^{\infty} p_{Y|X=x}(y)dy \\ &= 2\int_{-\infty}^{c} p_{Y|X=x}(y)dy - 1 = 0 \\ &\iff \int_{-\infty}^{c} p_{Y|X=x}(y)dy = \frac{1}{2}, \end{aligned}$$

so that $c = m(x)$, which is the median of $p_{Y|X=x}(y)$. It is a minimum since $r'(c) < 0$ for $c < m(x)$ and $r'(c) > 0$ for $c > m(x)$. Since m minimizes $\mathbb{E}[|Y - c| \,\big|\, X = x]$ at every x for any g we get

$$\mathbb{E}[|Y - g(X)| - |Y - m(X)|\big|X = x] \geq 0$$

which implies

$$R(g) - R(m) = \mathbb{E}[|Y - g(X)| - |Y - m(X)|] = \mathbb{E}\{\mathbb{E}[|Y - g(X)| - Y - m(X)|\big|X]\} \geq 0.$$

Problem 3 [25 pts.]

We can write

$$\hat{\beta} = \frac{\sum_{i=1}^n Y_i W_i}{\sum_{i=1}^n W_i^2} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i W_i}{\frac{1}{n}\sum_{i=1}^n W_i^2}.$$

By the Weak Law of Large Numbers, $\frac{1}{n}\sum_{i=1}^n Y_i W_i \xrightarrow{p} \mathbb{E}[YW]$ and $\frac{1}{n}\sum_{i=1}^n W_i^2 \xrightarrow{p} \mathbb{E}[W^2]$. Then by the Continuous Mapping Theorem,

$$\hat{\beta} \xrightarrow{p} \frac{\mathbb{E}[YW]}{\mathbb{E}[W^2]}.$$

We see that

$$\begin{aligned}
\mathbb{E}[YW] &= \mathbb{E}[(\beta X + \epsilon)(X + \delta)] \\
&= \mathbb{E}[\beta X^2 + \beta X\delta + X\epsilon + \epsilon\delta] \\
&= \beta\mathbb{E}[X^2] + \beta\mathbb{E}[X]\mathbb{E}[\delta] + \mathbb{E}[X]\mathbb{E}[\epsilon] + \mathbb{E}[\epsilon]\mathbb{E}[\delta] \\
&= \beta\mathbb{E}[X^2] + \beta\mathbb{E}[X]\cdot 0 + \mathbb{E}[X]\cdot 0 + 0\cdot 0 \\
&= \beta\mathbb{E}[X^2]
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[W^2] &= \mathbb{E}[(X + \delta)^2] \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X\delta] + \mathbb{E}[\delta^2] \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[\delta] + (\mathbb{E}[\delta])^2 + \mathbb{V}(\delta) \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X]\cdot 0 + 0^2 + \tau^2 \\
&= \mathbb{E}[X^2] + \tau^2.
\end{aligned}$$

We conclude that

$$\hat{\beta} \xrightarrow{p} \frac{\beta\mathbb{E}[X^2]}{\mathbb{E}[X^2] + \tau^2}.$$

So $\hat{\beta} \xrightarrow{p} a\beta$ where

$$a = \frac{\mathbb{E}[X^2]}{\mathbb{E}[X^2] + \tau^2}.$$

Problem 4 [30 pts.]

We see that

$$
\begin{aligned}
|\hat{\theta} - \theta| &= \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \mathbb{E}[Y_1] + \mathbb{E}[Y_0] \right| \\
&= \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(1,z)p(z)dz + \int \mu(0,z)p(z)dz \right| \\
&= \left| \left\{ \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right\} - \left\{ \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(0,z)p(z)dz \right\} \right| \\
&\leq \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right| + \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(0,z)p(z)dz \right|.
\end{aligned}
$$

Consider $\left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right|$. Note that $\int p(z)dz = 1$. We see that

$$
\begin{aligned}
\left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right| &= \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}(1, Z_i) \int p(z)dz - \frac{1}{n}\sum_{i=1}^{n} \int \mu(1,z)p(z)dz \right| \\
&= \left| \frac{1}{n}\sum_{i=1}^{n} \int \left(\hat{\mu}(1, Z_i) - \mu(1,z) \right) p(z)dz \right| \\
&= \left| \frac{1}{n}\sum_{i=1}^{n} \int \left(\hat{\mu}(1, Z_i) - \mu(1, Z_i) + \mu(1, Z_i) - \mu(1,z) \right) p(z)dz \right| \\
&= \left| \frac{1}{n}\sum_{i=1}^{n} \int (\hat{\mu}(1, Z_i) - \mu(1, Z_i))p(z)dz + \frac{1}{n}\sum_{i=1}^{n} \mu(1, Z_i) - \int \mu(1,z)p(z)dz \right| \\
&\leq \underbrace{\left| \frac{1}{n}\sum_{i=1}^{n} \int (\hat{\mu}(1, Z_i) - \mu(1, Z_i))p(z)dz \right|}_{A} + \underbrace{\left| \frac{1}{n}\sum_{i=1}^{n} \mu(1, Z_i) - \int \mu(1,z)p(z)dz \right|}_{B}.
\end{aligned}
$$

We will show that A is less than or equal to a value that converges in probability to 0:

$$
\begin{aligned}
\left| \frac{1}{n}\sum_{i=1}^{n} \int (\hat{\mu}(1, Z_i) - \mu(1, Z_i))p(z)dz \right| &\leq \frac{1}{n}\sum_{i=1}^{n} \int |\hat{\mu}(1, Z_i) - \mu(1, Z_i)| \, p(z)dz \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \int \left(\sup_{x^*, z^*} |\hat{\mu}(x^*, z^*) - \mu(x^*, z^*)| \right) p(z)dz \\
&= \int \left(\sup_{x^*, z^*} |\hat{\mu}(x^*, z^*) - \mu(x^*, z^*)| \right) p(z)dz \\
&= \sup_{x^*, z^*} |\hat{\mu}(x^*, z^*) - \mu(x^*, z^*)| \int p(z)dz \\
&= \sup_{x^*, z^*} |\hat{\mu}(x^*, z^*) - \mu(x^*, z^*)| \\
&\xrightarrow{p} 0.
\end{aligned}
$$

Now we will show that B converges in probability to 0. By the Weak Law of Large Numbers,

$$
\frac{1}{n}\sum_{i=1}^{n} \mu(1, Z_i) \xrightarrow{p} \mathbb{E}[\mu(1,z)] = \int \mu(1,z)p(z)dz.
$$

So

$$
\frac{1}{n}\sum_{i=1}^{n} \mu(1, Z_i) - \int \mu(1,z)p(z)dz \xrightarrow{p} 0,
$$

which implies that

$$\left| \frac{1}{n} \sum_{i=1}^{n} \mu(1, Z_i) - \int \mu(1,z)p(z)dz \right| \xrightarrow{p} 0.$$

Now we have shown that

$$\left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right| \leq \sup_{x^*, z^*} |\hat{\mu}(x^*, z^*) - \mu(x^*, z^*)| + \left| \frac{1}{n} \sum_{i=1}^{n} \mu(1, Z_i) - \int \mu(1,z)p(z)dz \right|$$

and the two terms in that sum both converge in probability to 0.
Also, $0 \leq \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right|$. So we conclude that

$$\left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right| \xrightarrow{p} 0.$$

By a similar argument,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(0,z)p(z)dz \right| \xrightarrow{p} 0.$$

Since

$$|\hat{\theta} - \theta| \leq \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1,z)p(z)dz \right| + \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(0,z)p(z)dz \right|,$$

we know that $|\hat{\theta} - \theta| \xrightarrow{p} 0$. Finally, we conclude that $\hat{\theta} \xrightarrow{p} \theta$.