ST 790, Homework 4 Solutions
Spring 2017

1. (a) From (1), subtracting μ_0, rearranging, and rescaling, it is immediate that

$$N^{1/2}(\widehat{\mu} - \mu_0) = N^{-1/2} \sum_{i=1}^{N} \left\{ \frac{C_i Y_i}{\pi(V_i)} - \frac{C_i - \pi(V_i)}{\pi(V_i)} h(V_i) - \mu_0 \right\}.$$

Thus, by the definition of an influence function, the influence function of $\widehat{\mu}$, which depends on the observed data, is

$$\varphi(C, CY, V) = \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} h(V) - \mu_0 \tag{1}$$

Trivially, by the same manipulations in the notes, conditioning on V and Y and using MAR,

$$\begin{aligned}
E\{\varphi(C, CY, V)\} &= E\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} h(V) - \mu_0 \right\} \\
&= E\left\{ \frac{E(C|V,Y)Y}{\pi(V)} - \frac{E(C|V,Y) - \pi(V)}{\pi(V)} h(V) - \mu_0 \right\} \\
&= E\left\{ \frac{E(C|V)Y}{\pi(V)} - \frac{E(C|V) - \pi(V)}{\pi(V)} h(V) - \mu_0 \right\} \\
&= E\left\{ \frac{\pi(V)Y}{\pi(V)} - \frac{\pi(V) - \pi(V)}{\pi(V)} h(V) - \mu_0 \right\} \\
&= E(Y - \mu_0) = 0.
\end{aligned}$$

(b) Because the influence function has mean zero, from (1),

$$\text{var}\{\varphi(C, CY, V)\} = E\{\varphi(C, CY, V)^2\} = E\left[\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} h(V) - \mu_0 \right\}^2 \right].$$

The key is to add and subtract

$$\frac{C - \pi(V)}{\pi(V)} E(Y|V)$$

to rewrite this as

$$\begin{aligned}
&E\left[\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} h(V) - \mu_0 \right\}^2 \right] \\
&\quad = E\left[\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} E(Y|V) + \frac{C - \pi(V)}{\pi(V)} \{E(Y|V) - h(V)\} - \mu_0 \right\}^2 \right] \\
&\quad = E\left[\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} E(Y|V, C=1) - \mu_0 + \frac{C - \pi(V)}{\pi(V)} \{E(Y|V, C=1) - h(V)\} \right\}^2 \right],
\end{aligned}$$

where the last equality follows from MAR. This can be written as

$$\begin{aligned}
&E\left[\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} E(Y|V, C=1) - \mu_0 \right\}^2 \right] + E\left[\left\{ \frac{C - \pi(V)}{\pi(V)} \{E(Y|V, C=1) - h(V)\} \right\}^2 \right] \\
&\quad - 2E\left[\left\{ \frac{CY}{\pi(V)} - \frac{C - \pi(V)}{\pi(V)} E(Y|V, C=1) - \mu_0 \right\} \left\{ \frac{C - \pi(V)}{\pi(V)} \{E(Y|V, C=1) - h(V)\} \right\} \right].
\end{aligned} \tag{2}$$

We now show that the cross-product term in (2), which is the covariance between the two squared terms, is equal to zero. In fact, we show a more general result that, for any function $g(V)$ of V,

$$E\left[\left\{\frac{CY}{\pi(V)} - \frac{C-\pi(V)}{\pi(V)}E(Y|V,C=1) - \mu_0\right\}\left\{\frac{C-\pi(V)}{\pi(V)}g(V)\right\}\right] = 0.$$

Write this as

$$E\left\{\frac{CY}{\pi(V)}\frac{C-\pi(V)}{\pi(V)}g(V)\right\} - E\left[\left\{\frac{C-\pi(V)}{\pi(V)}\right\}^2 E(Y|V,C=1)g(V)\right] - \mu_0 E\left\{\frac{C-\pi(V)}{\pi(V)}g(V)\right\}. \tag{3}$$

The third term in (3) is equal to zero immediately by an inner conditioning on V. So condider the other two terms. By conditioning on C and V and using MAR, the rightmost term in (3) becomes

$$E\left\{\frac{CE(Y|V,C)}{\pi(V)}\frac{C-\pi(V)}{\pi(V)}g(V)\right\} = E\left[\frac{C\{C-\pi(V)\}}{\pi(V)^2}E(Y|V,C=1)g(V)\right].$$

Now do an inner conditioning on V, which yields (noting that $C^2 = C$)

$$E\left\{\frac{\pi(V)-\pi(V)^2}{\pi(V)^2}E(Y|V,C=1)g(V)\right\} = E\left\{\frac{1-\pi(V)}{\pi(V)}E(Y|V,C=1)g(V)\right\}.$$

By an inner conditioning on V, the middle term in (3) becomes

$$E\left[\frac{\pi(V)\{1-\pi(V)\}\pi(V)}{\pi(V)^2}E(Y|V,C=1)g(V)\right] = E\left\{\frac{1-\pi(V)}{\pi(V)}E(Y|V,C=1)g(V)\right].$$

Putting this together, we have shown that (3) is equal to zero. Thus, we have

$$E\left[\left\{\frac{CY}{\pi(V)} - \frac{C-\pi(V)}{\pi(V)}h(V) - \mu_0\right\}^2\right] \tag{4}$$
$$= \left[\left\{\frac{CY}{\pi(V)} - \frac{C-\pi(V)}{\pi(V)}E(Y|V,C=1) - \mu_0\right\}^2\right] + E\left[\left\{\frac{C-\pi(V)}{\pi(V)}\{E(Y|V,C=1)-h(V)\}\right\}^2\right].$$

It follows immediately that (4) will be smallest when $h(V) = E(Y|V,C=1) = E(Y|V)$ under MAR. In fact, we have shown that

$$\left[\left\{\frac{CY}{\pi(V)} - \frac{C-\pi(V)}{\pi(V)}E(Y|V) - \mu_0\right\}^2\right] + E\left[\left\{\frac{C-\pi(V)}{\pi(V)}g(V)\right\}^2\right]$$

is smallest when $g(V) \equiv 0$, so that the first term is the smallest the variance of (2) can be.

2. (a)-(c) should have been straightforward. (d)(i) required you to run `proc mi` with the `impute=monotone` option; e.g.,

```
proc mi data=armd out=armdmonoout seed=1518971 nimpute=10;
   mcmc impute=monotone;
   var visual0 visual4 visual12 visual24 visual52;
run;
```

to obtain imputed data sets with monotone missingness. (d)(ii) and (iii) were then also straightforward.

(e) I got

```
               Est     SE          Z    P-val
(a)          -5.6594 (2.5158)   -2.25  0.025
(b)          -5.1928 (2.5683)   -2.02  0.043
(c)   (i)    -5.7649 (3.5642)   -1.62  0.106
     (ii)    -4.6800 (2.6967)   -1.74  0.083
(d)   (i)    -5.6222 (3.5104)   -1.60  0.109
     (ii)    -5.3142 (2.6418)   -2.01  0.044
```

where in (c) and (d) (i) represents subject level weighting and (ii) represents occasion level weighting. You might have noted that deleting the individuals with nonmonotone missingness patterns seems to have a nonnegligible effect on the estimate, and that subject level and occasion level weighting yield qualitatively different inferences. Naive analysis 1 gives a similar esitmate to subject level weighting based on imputation, but the inference is qualitatively different (in terms of “statistical significance”). You might have also compared these results to those from the last homework, where you did full imputation. You probably concluded that the best approach is not clear cut, although the naive approaches are probably not to be recommended on principled grounds.