Let (X, Y) denote a bivariate vector with finite fourth moments. Suppose that Y may be missing, so we use R (taking values 1 and 0) to indicate whether Y is observed. We further assume that R and Y are independent given X and that $\pi(x) = P(R = 1|X = x)$ is a known function strictly bounded away from zero. The goal is to estimate $\mu = E[Y]$.

The observed data from n independent subjects consist of $(R_i Y_i, R_i, X_i)$, $i = 1, ..., n$.

(a) (5 points) Show that for any measurable function $g(X)$ such that $E[g(X)^2] < \infty$, the following equality holds:

$$\mu = E\left[g(X) + \frac{R}{\pi(X)}(Y - g(X))\right].$$

(b) (5 points) From (a), we can construct the following estimator for μ (called difference estimator):

$$\widetilde{\mu} = n^{-1} \sum_{i=1}^{n} \left[g(X_i) + \frac{R_i}{\pi(X_i)}(Y_i - g(X_i))\right].$$

Derive the asymptotic distribution for $\widetilde{\mu}$. Show that its asymptotic variance is equal to

$$Var(Y) + E\left[w(X)(Y - g(X))^2\right],$$

where $w(x) = \pi(x)/(1 - \pi(x))$.

(c) (5 points) We choose $g(x) = \gamma x$. Show that the optimal γ to minimize the variance in (b) is $\gamma^* = \frac{E[YXw(X)]}{E[X^2 w(X)]}$. Thus, an estimator for γ^* using the observed data is defined as

$$\widehat{\gamma} = \frac{\sum_{i=1}^{n} R_i Y_i X_i w(X_i)/\pi(X_i)}{\sum_{i=1}^{n} X_i^2 w(X_i)}.$$

Show $\widehat{\gamma} \to_{a.s.} \gamma^*$ and further derive the asymptotic distribution for $\widehat{\gamma}$. You don't need to simplify the final expression.

(d) (5 points) Using $\widehat{\gamma}$, we obtain an estimator for μ as

$$\widehat{\mu} = n^{-1} \sum_{i=1}^{n} \left[\widehat{\gamma} X_i + \frac{R_i}{\pi(X_i)}(Y_i - \widehat{\gamma} X_i)\right].$$

What is the asymptotic distribution for $\widehat{\mu}$? Justify your answer.

(e) (5 points) We now consider $g(x)$ to be any arbitrary function. Show that $g^*(x) = E[Y|X = x]$ is the optimal function for g that minimizes the asymptotic variance in (b). Suggest an estimator using the observed data to estimate this optimal function.

Solution

Note: Students are expected to obtain at least 10 points out of (a)-(c), 5 points out of (d)-(e)

(a) It follows from

$$E\left[\frac{R}{\pi(X)}(Y-g(X))\right] = E\left[E\left\{\frac{R}{\pi(X)}(Y-g(X))|X\right\}\right]$$
$$= E\left[E\left\{\frac{R}{\pi(X)}|X\right\}E\{Y-g(X)|X\}\right] = E[Y-g(X)].$$

(b) By CLT, $\sqrt{n}(\widetilde{\mu}-\mu)\to_d N(0,\widetilde{\sigma}^2)$, where

$$\begin{aligned}
\widetilde{\sigma}^2 &= Var(\frac{R}{\pi(X)}Y - \frac{R-\pi(X)}{\pi(X)}g(X)) \\
&= Var(\frac{R}{\pi(X)}Y) - 2Cov(\frac{R}{\pi(X)}Y, \frac{R-\pi(X)}{\pi(X)}g(X)) + Var(\frac{R-\pi(X)}{\pi(X)}g(X)) \\
&= E[Y^2/\pi(X] - \mu^2 - 2E\left[\frac{R(R-\pi(X))}{\pi(X)^2}Yg(X)\right] + E\left[(\frac{R-\pi(X)}{\pi(X)}g(X))^2\right] \\
&= Var(Y) + E[w(X)Y^2] - 2E\left[w(X)Yg(X)\right] + E\left[w(X)g(X)^2\right] \\
&= Var(Y) + E\left[w(X)(Y-g(X))^2\right].
\end{aligned}$$

(c) The first half is obvious and $\widehat{\gamma}\to_{a.s.}\gamma^*$ follows from SLLN. To obtain the asymptotic distribution, note

$$\sqrt{n}\left\{\left(n^{-1}\sum_{i=1}^n \frac{R_iY_iw(X_i)}{\pi(X_i)}, n^{-1}\sum_{i=1}^n X_i^2w(X_i)\right) - \left(E[\frac{RYw(X)}{\pi(X)}], E[X^2w(X)]\right)\right\} \to_d N(0,\Sigma),$$

where

$$\Sigma = \begin{pmatrix} Var(RYw(X)/\pi(X)) & Cov(RYw(X)/\pi(X), X^2w(X)) \\ Cov(RYw(X)/\pi(X), X^2w(X)) & Var(X^2w(X)) \end{pmatrix}.$$

Then we apply the delta method to obtain

$$\sqrt{n}(\widehat{\gamma}-\gamma^*)\to_d N\left(0, \frac{(1,-\gamma^*)\Sigma(1,-\gamma^*)^T}{E[X^2w(X)]^2}\right).$$

(d) Since

$$\widehat{\mu} = n^{-1}\sum_{i=1}^n\left[\gamma^*X_i + \frac{R_i}{\pi(X_i)}(Y_i-\gamma^*X_i)\right] + (\widehat{\gamma}-\gamma^*)\left[n^{-1}\sum_{i=1}^n X_i\left\{1-\frac{R_i}{\pi(X_i)}\right\}\right],$$

and note

$$\left[n^{-1}\sum_{i=1}^n X_i\left\{1-\frac{R_i}{\pi(X_i)}\right\}\right]\to_{a.s.} 0,$$

we obtain

$$\sqrt{n}(\widehat{\mu} - \mu) = \sqrt{n}(n^{-1} \sum_{i=1}^{n} \left[\gamma^* X_i + \frac{R_i}{\pi(X_i)} (Y_i - \gamma^* X_i) \right] - \mu) + o_p(1),$$

so $\sqrt{n}(\widehat{\mu} - \mu)$ converges in distribution to $N(0, \widetilde{\sigma}^2)$ where $\widetilde{\sigma}^2$ is given in (b) with $g(x) = \gamma^* x$.

(e) Since

$$E[w(X)(Y - g(X))^2] = E[w(X)Var(Y|X)] + E[w(X)(E[Y|X] - g(X))^2],$$

$\widetilde{\sigma}^2$ is minimized when $g(x) = E[Y|X = x]$. We use the empirical data to estimate g using the kernel estimators but each subject is weighted by $R_i/\pi(X_i)$, i.e.,

$$\frac{\sum_{i=1}^{n} R_i/\pi(X_i) K_{a_n}(X_i - x) Y_i}{\sum_{i=1}^{n} R_i/\pi(X_i) K_{a_n}(X_i - x)}.$$