// 2012 Theory I #2

Notice that (Y_i, R_i, X_i) are iid so that by integrating out X_i (which are iid) we obtain that (Y_i, R_i) are iid. Thus R_iY_i are iid

2a) By SLLN,

$$\frac{1}{n}\sum_{i=1}^{n} R_iY_i \xrightarrow{a.s.} E[R_iY_i] = E\left(E\left[R_iY_i \mid X_i\right]\right) = E\left(E[R_i \mid X_i]E[Y_i \mid X_i]\right)$$

or

$$\frac{1}{n}\sum_{i=1}^{n} R_iY_i \xrightarrow{a.s.} E[R_iY_i] = E\left(R_i E[Y_i \mid R_i]\right)$$

$$= \sum_{r=0}^{1} r\, E[Y_i \mid R_i = r] P(R_i = r) = E[Y_i \mid R_i = 1] P(R_i = 1)$$

Also by SLLN

$$\frac{1}{n}\sum_{i=1}^{n} R_i \xrightarrow{a.s.} ER_i = P(R_i = 1)$$

$$\infty > EY_i^2 = E\left(E[Y_i^2 \mid R_i]\right)$$
$$= E[Y_i^2 \mid R_i = 1] P(R_i = 1)$$
$$> E[Y_i^2 \mid R_i = 1]$$

Thus, by Slutsky's / CMT

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} R_iY_i}{\sum_{i=1}^{n} R_i} = \frac{\frac{1}{n}\sum_{i=1}^{n} R_iY_i}{\frac{1}{n}\sum_{i=1}^{n} R_i} \xrightarrow{a.s.} \frac{E[Y_i \mid R_i = 1] P(R_i = 1)}{P(R_i = 1)} = E[Y_i \mid R_i = 1]$$

Next we calculate $\mathrm{Var}\left(\begin{bmatrix} R_iY_i \\ R_i \end{bmatrix}\right)$ for use with the CLT.

$$\mathrm{Var}[R_iY_i] = E\left(\mathrm{Var}[R_iY_i \mid R_i]\right) + \mathrm{Var}\left(E[R_iY_i \mid R_i]\right)$$

$$= E\left(R_i^2 \mathrm{Var}[Y_i \mid R_i]\right) + \mathrm{Var}\left(R_i E[Y_i \mid R_i]\right)$$

$$= E\left(R_i \mathrm{Var}[Y_i \mid R_i]\right) + \mathrm{Var}\left(R_i E[Y_i \mid R_i]\right)$$

$$= \sum_{r=0}^{1} r\, \mathrm{Var}[Y_i \mid R_i = r] + E\left\{\left[R_i E[Y_i \mid R_i] - E\left(R_i E[Y_i \mid R_i]\right)\right]^2\right\}$$

$$= \mathrm{Var}[Y_i \mid R_i = 1] + \sum_{r=0}^{1} \left\{r\, E[Y_i \mid R_i = r] - E[Y_i \mid R_i = 1] P(R_i = 1)\right\}^2 P(R_i = r)$$

$$= Var[Y_i \mid R_i = 1] + \left(E[Y_i \mid R_i = 1] P(R_i = 1)\right)^2 P(R_i = 0)$$
$$+ \left(E[Y_i \mid R_i = 1] - E[Y_i \mid R_i = 1] P(R_i = 1)\right)^2 P(R_i = 1)$$

$$= Var[Y_i \mid R_i = 1] + \left\{E[Y_i \mid R_i = 1] P(R_i = 1)\right\}^2 P(R_i = 0)$$
$$+ \left\{E[Y_i \mid R_i = 1] P(R_i = 0)\right\}^2 P(R_i = 1)$$

$$= Var[Y_i \mid R_i = 1] + \left(E[Y_i \mid R_i = 1]\right)^2 P(R_i = 0) P(R_i = 1)$$

$$= Var[Y_i \mid R_i = 1] + \left(E[Y_i \mid R_i = 1]\right)^2 Var[R_i]$$

$$Cov[R_i Y_i, R_i] =$$

$$= E[R_i^2 Y_i] - E[R_i Y_i] E R_i$$

$$= E[R_i Y_i] - E[R_i Y_i] E R_i = E[R_i Y_i]\left\{1 - P(R = 1)\right\}$$

$$= E[Y_i \mid R_i = 1] Var[R_i]$$

Let $g(a,b) = \frac{a}{b}$, then $\frac{\partial}{\partial a} g(a,b) = \frac{1}{b}$, $\frac{\partial}{\partial b} g(a,b) = -\frac{a}{b^2}$

By CLT

$$\sqrt{n}\left(\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} R_i Y_i \\ \frac{1}{n}\sum_{i=1}^{n} R_i \end{bmatrix} - \begin{bmatrix} E[Y_i \mid R_i = 1]\, P(R_i = 1) \\ P(R_i = 1) \end{bmatrix}\right) \xrightarrow{L} N(0, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} Var[Y_1 \mid R_1 = 1] + (E[Y_1 \mid R_1 = 1])^2 Var[R_1] & E[Y_1 \mid R_1 = 1] Var[R_1] \\ E[Y_1 \mid R_1 = 1] Var[R_1] & Var[R_1] \end{bmatrix}$$

Then by the delta method

$$\sqrt{n}\left(\hat{\mu}_1 - E[Y_1 \mid R_1 = 1]\right) \xrightarrow{L} N(0, \tau)$$

where

$$\tau = \begin{bmatrix} \frac{1}{P(R_1=1)} & -\frac{E[Y_1 \mid R_1 = 1]}{P(R_1 = 1)} \end{bmatrix} \Sigma \begin{bmatrix} \frac{1}{P(R_1=1)} \\ -\frac{E[Y_1 \mid R_1 = 1]}{P(R_1=1)} \end{bmatrix}$$

$$= \frac{1}{[P(R_1=1)]^2}\Big\{ Var[Y_1 \mid R_1 = 1] + (E[Y_1 \mid R_1 = 1])^2 Var[R_1] - 2(E[Y_1 \mid R_1 = 1])^2 Var[R_1]$$
$$+ (E[Y_1 \mid R_1 = 1])^2 Var[R_1] \Big\}$$

$$= \frac{Var[Y_1 \mid R_1 = 1]}{[P(R_1 = 1)]^2}$$

(b) $$E\left[\frac{R_i Y_i}{\pi(X_i)}\right] = E\left(E\left[\frac{R_i Y_i}{\pi(X_i)} \middle| X_i\right]\right) = E\left(\frac{1}{\pi(X_i)} E[R_i | X_i] E[Y_i | X_i]\right)$$

$$= E\left(\frac{1}{\pi(X_i)} P(R_i = 1 | X_i) E[Y_i | X_i]\right) = E\left(\frac{1}{\pi(X_i)} \pi(X_i) E[Y_i | X_i]\right)$$

$$= E\left(E[Y_i | X_i]\right) = EY_i = \mu$$

Use SLLN! Just have to show integrability?

and furthermore

$$E\left[\frac{1}{n}\sum_{i=1}^{n} \frac{R_i Y_i}{\pi(X_i)}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{R_i Y_i}{\pi(X_i)}\right] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Then

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{R_i Y_i}{\pi(X_i)} - \mu\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2} Var\left[\frac{1}{n}\sum_{i=1}^{n}\frac{R_i Y_i}{\pi(X_i)}\right]$$

$$\overset{ind.}{=} \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n} Var\left[\frac{R_i Y_i}{\pi(X_i)}\right] = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\left\{E\left[\frac{R_i^2 Y_i^2}{[\pi(X_i)]^2}\right] - \left(E\left[\frac{R_i Y_i}{\pi(X_i)}\right]\right)^2\right\}$$

$$\le \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\left\{E\left[\frac{1 \cdot Y_i^2}{K}\right] - 0\right\} = \frac{1}{\varepsilon^2 K}\cdot\frac{1}{n^2}\sum_{i=1}^{n} EY_1^2 = \frac{EY_1^2}{\varepsilon^2 K}\cdot\frac{1}{n} \to 0$$

where K ~~as~~ is the ~~value~~ lower bound for $\pi(x)$

Next we want to calculate the asymptotic distribution of $\sqrt{n}(\hat{\mu}_2 - \mu)$.
Since (Y_i, R_i, X_i) are iid, measurable functions of the vector are also iid, in particular ~~the~~ $\frac{R_i Y_i}{\pi(X_i)}$. Thus we may apply the CLT.

We have

$$Var\left[\frac{R_1 Y_1}{\pi(X_1)}\right] = E\left[\frac{R_1^2 Y_1^2}{[\pi(X_1)]^2}\right] - \left(E\left[\frac{R_1 Y_1}{\pi(X_1)}\right]\right)^2$$

$$= E\left[\frac{R_1 Y_1^2}{[\pi(X_1)]^2}\right] - \mu^2 = E\left(E\left[\frac{R_1 Y_1^2}{[\pi(X_1)]^2}\,\middle|\, X_1\right]\right) - \mu^2$$

$$= E\left(\frac{1}{[\pi(X_1)]^2} E[R_1 | X_1] E[Y_1^2 | X_1]\right) - \mu^2$$

$$= E\left(\frac{E[Y_1^2 | X_1]}{\pi(X_1)}\right) - \mu^2$$

Therefore by CLT,

$$\sqrt{n}(\hat{\mu}_2 - \mu) \xrightarrow{L} N\left\{0,\ E\left(\frac{E[Y_1^2 | X_1]}{\pi(X_1)}\right) - \mu^2\right\}$$

2c) $$E\left\{g(X_i)\left[1-\frac{R_i}{\pi(X_i)}\right]\right\} = E\left\{E\left[g(X_i)\left(1-\frac{R_i}{\pi(X_i)}\right)\middle| X_i\right]\right\}$$

$$= E\left\{g(X_i)\left[1-\frac{E[R_i|X_i]}{\pi(X_i)}\right]\right\} = E\left\{g(X_i)\left[1-\frac{\pi(X_i)}{\pi(X_i)}\right]\right\} = 0$$

Thus $E[\hat{\mu}_g] = \mu$. We want to calculate $Var[\bullet]$. We have

$$Var\left\{\frac{R_iY_i}{\pi(X_i)} + g(X_i)\left[1-\frac{R_i}{\pi(X_i)}\right]\right\}$$

$$\boxed{\text{write this as } \frac{R_i}{\pi(X_i)}(Y_i - g(X_i)) + g(X_i)?}$$

$$= Var\left[\frac{R_iY_i}{\pi(X_i)}\right] + 2Cov\left[\frac{R_iY_i}{\pi(X_i)}, g(X_i)\left(1-\frac{R_i}{\pi(X_i)}\right)\right] + Var\left[g(X_i)\left(1-\frac{R_i}{\pi(X_i)}\right)\right]$$

We calculate the pieces individually.

$$Cov\left[\frac{R_iY_i}{\pi(X_i)}, g(X_i)\right] = E\left[\frac{R_iY_ig(X_i)}{\pi(X_i)}\right) - E\left[\frac{R_iY_i}{\pi(X_i)}\right]E\left[g(X_i)\right)$$

$$= E\left(E\left[\frac{R_iY_ig(X_i)}{\pi(X_i)}\middle| X_i\right]\right) - \mu E[g(X_i)]$$

$$= E\left(g(X_i)E[Y_i|X_i]\right) - \mu E[g(X_i)]$$

$$\operatorname{Cov}\left[\frac{R_i Y_i}{\pi(X_i)}, \frac{g(X_i) R_i}{\pi(X_i)}\right] = E\left[\frac{R_i Y_i g(X_i)}{[\pi(X_i)]^2}\right] - E\left[\frac{R_i Y_i}{\pi(X_i)}\right] E\left[\frac{g(X_i) R_i}{\pi(X_i)}\right]$$

$$= E\left(\frac{g(X_i) E[Y_i | X_i]}{\pi(X_i)}\right) - \mu E[g(X_i)]$$

$$\operatorname{Var}\left[g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right] = E\left\{[g(X_i)]^2\left(1 - 2\frac{R_i}{\pi(X_i)} + \frac{R_i}{[\pi(X_i)]^2}\right)\right\}$$

$$- \left\{E\left[g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right]\right\}^2 \leftarrow = 0$$

$$= E\{[g(X_i)]^2\} - 2E\{[g(X_i)]^2\} + E\left\{\frac{[g(X_i)]^2}{\pi(X_i)}\right\}$$

$$= E\left\{\frac{[g(X_i)]^2}{\pi(X_i)}\right\} - E\{[g(X_i)]^2\}$$

$$\operatorname{Var}\left[\frac{R_i Y_i}{\pi(X_i)}\right] = E\left(\frac{R_i Y_i^2}{[\pi(X_i)]^2}\right) - \mu^2 = E\left(\frac{E[Y_i^2 | X_i]}{\pi(X_i)}\right) - \mu^2$$

Then

$$\operatorname{Var}[\cdot] \overset{\gamma}{=} E\left(\frac{E[Y_i^2|X_i]}{\pi(X_i)}\right) - \mu^2 + 2\left\{E\left(g(X_i)E[Y_i|X_i]\right) - E\left(\frac{g(X_i)E[Y_i|X_i]}{\pi(X_i)}\right)\right\}$$

$$+ E\left(\frac{[g(X_i)]^2}{\pi(X_i)}\right) - E\left([g(X_i)]^2\right)$$

Want to verify that

$$\infty > \operatorname{Var}\left\{ \frac{R_i Y_i}{\pi(X_i)} + g(X_i)\left[1 - \frac{R_i}{\pi(X_i)}\right]\right\}$$

$$= \underbrace{\operatorname{Var}\left[\frac{R_i Y_i}{\pi(X_i)}\right]}_{\leq E Y_i^2 \text{ from (2b)}} + 2\operatorname{Cov}\left[\frac{R_i Y_i}{\pi(X_i)}, g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right] + \operatorname{Var}\left[g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right]$$

$$\operatorname{Cov}\left[\frac{R_i Y_i}{\pi(X_i)}, g(X_i)\right] \overset{\text{Cauchy-Schwarz}}{\leq} \operatorname{Var}\left[\frac{R_i Y_i}{\pi(X_i)}\right] \operatorname{Var}\left[g(X_i)\right] \leq E Y_i^2 \operatorname{Var}[g(X_i)] \overset{\text{By assmpt}}{<} \infty$$

$$\operatorname{Cov}\left[\frac{R_i Y_i}{\pi(X_i)}, \frac{g(X_i) R_i}{\pi(X_i)}\right] \leq \operatorname{Var}\left[\frac{R_i Y_i}{\pi(X_i)}\right] \operatorname{Var}\left[\frac{g(X_i) R_i}{\pi(X_i)}\right] \leq E Y_i^2 \operatorname{Var}\left[\frac{g(X_i)}{K}\right] < \infty$$

$$\operatorname{Var}\left[g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right] \overset{\text{from prev. calculation}}{=} E\left(\frac{[g(X_i)]^2}{\pi(X_i)}\right) - E\left([g(X_i)]^2\right)$$

$$\leq \frac{E\left([g(X_i)]^2\right)}{K} < \infty$$

Thus,

$$P\left(\left|\hat{\mu}_g - \mu\right| > \varepsilon\right) \leq \frac{\operatorname{Var}\left[\hat{\mu}_g\right]}{\varepsilon^2} = \frac{\gamma}{\varepsilon^2 n} \longrightarrow 0$$

and by CLT

$$\sqrt{n}\left(\hat{\mu}_g - \mu\right) \xrightarrow{L} N(0, \gamma)$$

for some choice of $g(\cdot)$

2d) We wish to minimize $Var[\hat{\mu}_g]$. Since we saw in (2c) that $\hat{\mu}_g$ is unbiased and the (Y_i, X_i, R_i) vectors are mutually independent we have

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

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var\left[\frac{R_i Y_i}{\pi(X_i)} + g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E\left\{\left[\frac{R_i Y_i}{\pi(X_i)} + g(X_i)\left(1 - \frac{R_i}{\pi(X_i)}\right)\right]^2\right\} - \frac{\mu^2}{n}$$

Furthermore, since each of the random variables in the above expectation are identically distributed, it suffices to minimize

$$E\left\{\left[\frac{R_1 Y_1}{\pi(X_1)} + g(X_1)\left(1 - \frac{R_1}{\pi(X_1)}\right)\right]^2\right\}$$

wrt some choice of $g(\cdot)$. Denoting the aforementioned r.v. $B(Y_1, R_1, X_1)$ we have

$$E[B(Y_1, R_1, X_1)] = E_{X_1} E_{Y_1, R_1 | X_1}[B(Y_1, R_1, x) \mid X_1 = x]$$

so it is clear that minimizing

$$E_{Y_1, R_1 | (X_1 = x)}[B(Y_1, R_1, x) \mid X_1 = x]$$

wrt g for each x in the support of X_1 minimizes the entire objective function.

We have

$$E_{Y_1,R_1|(X_1=x)}\left[B(Y_1,R_1,x)\,|X_1=x\right] \overset{\text{conditionally ind.}}{=} \sum_{r=0}^{1} E\left[B(Y_1,r,x)\mid X_1=x\right]$$

$$= E\left\{\left[\frac{0\cdot Y_1}{\pi(x)}+g(x)\left(1-\frac{0}{\pi(x)}\right)\right]^2\Big|\,X_1=x\right\}+E\left\{\left[\frac{1\cdot Y_1}{\pi(x)}+g(x)\left(1-\frac{1}{\pi(x)}\right)\right]^2\Big|\,X_1=x\right\}$$

$$= [g(x)]^2+E\left\{\left[\frac{Y_1}{\pi(x)}-g(x)\left(\frac{1-\pi(x)}{\pi(x)}\right)\right]^2\Big|\,X_1=x\right\}$$

Then taking derivatives wrt. $g(x)$ and justifying interchange of limits and integrals by Fubini's theorem we obtain

$$2g(x)-2\,\frac{1-\pi(x)}{\pi(x)}\,E\left[\frac{Y_1}{\pi(x)}-g(x)\left(\frac{1-\pi(x)}{\pi(x)}\right)\Big|\,X_1=x\right]$$

$$=2g(x)-2\,\frac{1-\pi(x)}{\pi(x)}\left\{\frac{E[Y_1|X_1=x]}{\pi(x)}-g(x)\left(\frac{1-\pi(x)}{\pi(x)}\right)\right\}\overset{\text{set}}{=}0$$

$$\Rightarrow\ \frac{g(x)}{\pi(x)}-\frac{1-\pi(x)}{\pi(x)}\cdot\frac{E[Y_1|X_1=x]}{\pi(x)}=0$$

$$\Rightarrow\ g(x)=E[Y_1|X_1=x]\,\frac{1-\pi(x)}{\pi(x)}$$

The second derivative is strictly positive so this is indeed a local min.

$Z_1, Z_2, \ldots$

e) We need an estimator for $E[Y_i | X_i = x_j]$, $j = 1, \ldots, K$. Let $Z_i = (Z_{i1}, \ldots, Z_{iK})$ be a random vector with $Z_{ij} = I(X_i = x_j)$, $j = 1, \ldots, K$.

Thus $Z_1, Z_2, \ldots \overset{iid}{\sim} \text{mult}(1, \gamma)$ where $\gamma_j = P(X_i = x_j)$, $j = 1, \ldots, K$.

Next, define $\hat{\theta}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nK})$ where

$$\hat{\theta}_{nj} = \begin{cases} \dfrac{\sum_{i=1}^n z_{ij} y_i}{\sum_{i=1}^n z_{ij}}, & \sum_{i=1}^n z_{ij} \geq 1 \\ 0, & \text{else} \end{cases}$$

Now by SLLN,

$$\frac{1}{n} \sum_{i=1}^n z_{ij} y_i \xrightarrow{a.s.} E[Z_{1j} Y_1] = E\left(Z_{1j} E[Y_1 | Z_{1j}]\right) = E[Y_1 | X_1 = x_j] P(X_1 = x_j)$$

$$\frac{1}{n} \sum_{i=1}^n z_{ij} \xrightarrow{a.s.} E[Z_{1j}] = P(X_1 = x_j)$$

so that by Slutsky's / CMT

$$\begin{bmatrix} \dfrac{\sum_{i=1}^n z_{i1} y_i}{\sum_{i=1}^n z_{i1}} \\ \vdots \\ \dfrac{\sum_{i=1}^n z_{iK} y_i}{\sum_{i=1}^n z_{iK}} \end{bmatrix} \xrightarrow{a.s.} \begin{bmatrix} E[Y_1 | X_1 = x_1] \\ \vdots \\ E[Y_1 | X_1 = x_K] \end{bmatrix}$$

Since Z_i are iid,

$$P\left(\lim_{n \to \infty} \bigcup_{j=1}^K \left\{ \sum_{i=1}^n z_{ij} = 0 \right\}\right) = 0$$

which is equivalent to

$$\hat{\theta}_n \xrightarrow{a.s} \left(\frac{\sum_{i=1}^{\hat{n}} z_{i1} y_i}{\sum_{i=1}^{\hat{n}} z_{i1}}, \ldots, \frac{\sum_{i=1}^{\hat{n}} z_{ik} y_i}{\sum_{i=1}^{\hat{n}} z_{ik}} \right)$$

should be $(\cdot) - \hat{\theta}_n \xrightarrow{a.s.} 0$

so that by Slutsky's

$$\hat{\theta}_n \xrightarrow{a.s.} \left(E[Y_1 | X_1 = x_1], \ldots, E[Y_1 | X_1 = x_j] \right)$$