To hear and view this Pencast PDF on your computer,
click here to get the latest version of Adobe® Reader®

Ivan's stochastic interventions

9/29/2011 for $O = (W, A, Y)$

$$\left.\begin{array}{l} W = f_W(U_W) \\ A = f_A(W, U_A) \\ Y = f_Y(W, A, U_Y) \end{array}\right\}$$

to intervene, set A rather than allowing equation $\Rightarrow Y_1 = f_Y(W, 1, U_Y)$

$\Downarrow$

Can be rewritten

$W = f_W(U_W)$

$A = \begin{cases} 1 & prob = 1 \\ 0 & prob = 0 \end{cases}$

$Y_1 = f_Y(W, 1, U_Y)$

<u>Question</u>: Why not set $A = \begin{cases} 1 & w/\ prob\ P(W) \\ 0 & w/\ prob\ 1 - P(W) \end{cases}$?

↑ probabilities of getting intervention

so, new notation: $W = f_W(U_W)$

$A_{P\delta} = a$ w/ prob $P\delta(g_0)(A = a, W)$

$Y_{P\delta} = f_Y(A_{P\delta}, W, U_Y)$

$\underbrace{P(O) = P(Y|A,W)\,P(A|W)\,P(W)}$

$P(A|W) = g(A|W)$

$P(W) = Q_W(W)$

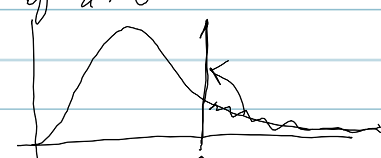
$E(Y|A,W) = \bar{Q}(A, W)$

<u>Examples</u>: 1) $P_\delta(g_0)(A, W) = g_0(\underbrace{A - \delta(W)}_{\text{shifts } +x \text{ as function of covariates}}|W)$

will only focus on this today

2) $P_\delta(g_0)(A = a, W) = \begin{cases} g_0(a|W) & \text{if } a < \delta \\ \sum_{a \geq \delta} g_0(a|W) & \text{if } a = \delta \\ 0 & \text{if } a > \delta \end{cases}$

$\Rightarrow$



↑ all obs on right set to this value (in this ex, $= 0$)

Dealing w/ parameter, can est.

$E(Y_{P\delta})$ or $\underbrace{E(Y_{P\delta} - Y)}$

$= E(Y_{P\delta}) - \underbrace{E(Y)}_{\text{already well know, so can throw out.}}$

This pencast was created with my Livescribe™ smartpen. Click here to see how to create your own.

To hear and view this Pencast PDF on your computer,
click here to get the latest version of Adobe® Reader®

So, $E(Y_{PS}) = \sum_Y y\, P(Y_{PS} = y)$, by def of expectations

<u>Assumptions</u>

1) $A \perp Y_a | W \ \forall\, a \in \mathcal{A}$ (RA assumption)
2) $Y = Y_a \Rightarrow$ if $A = a$, then $Y_a = y$ (def of counterfactuals)

$$P(Y_{PS} = y) = \sum_a \sum_w P(Y_{PS}, A_{PS}, w)$$
$$= \sum_a \sum_w \underbrace{P(Y_{PS} = y \mid A_{PS} = a, W = w)}_{\text{need to identify}} \underbrace{P(A_{PS} = a \mid W = w)}_{P_\delta(g)(A=a|W)} \underbrace{P(W = w)}_{Q_W(w)}$$

Note: 1) $A_{PS} = a \Rightarrow Y_{PS} = Y_a$
↳ comes from consistency def. (assumption #2)

2) $A \perp Y_a | W \Rightarrow A_{PS} \perp Y_a | W$

* think of $A_{PS} = A - \delta(W)$
↳ comes from assumption #1

$$\Rightarrow P(Y_a = y \mid A_{PS} = a, W\ w) \text{ , from note \#1}$$
$$= P(Y_a = y \mid W = w) = P(Y = y \mid A = a, W = w)$$
$\Rightarrow$ Identifiability satisfied

(*) So, $E(Y_{PS}) = \Psi(P) = \sum_Y \sum_a \sum_w y\, P(Y = y \mid A = a, W = w)\, g(a - \delta(w) | w)\, Q_W(w)$

$$= E_P\left[\frac{g_0(A - \delta | W)}{g(A|W)} Y\right] \quad \text{depends on } g(A|W)$$

$$= E_P\left[\bar{Q}(A + \delta, W)\right) \quad \text{depends on } \bar{Q}(A, W)$$

* Recall: EIC of $\Psi(P_n)$ is the gradient
↳ Another way of getting IC: $\frac{\partial}{\partial \varepsilon} \Psi(P(\varepsilon))\big|_{\varepsilon = 0} = E[D(0) S(0)]$

1st, define class of parametric submodels
↳ $P(\varepsilon)(Y|A,W) = [1 + \varepsilon S_Y(W,A,Y)]\, P(Y|A,W)$ $\quad E[S_Y | A\ W] = 0$
$g(\varepsilon)(A|W) = [1 + \varepsilon S_A(W,A)]\, P(A|W)$ $\quad E[S_A | W] = 0$
$Q_W(\varepsilon)(W) = [1 + \varepsilon S_W(W)]\, Q_W(W)$ $\quad E[S_W] = 0$

Note: $P_0(\varepsilon) = P_Y(\varepsilon)\, g(\varepsilon)\, Q_W(\varepsilon)$
$$= \underbrace{S_Y + S_A + S_W}_{= S(0)} \text{, the score}$$
$$T(P) = L_0^2(P)$$

This pencast was created with my Livescribe™ smartpen. Click here to see how to create your own.

To hear and view this Pencast PDF on your computer,
click here to get the latest version of Adobe® Reader®

Going back to $\frac{\partial}{\partial \varepsilon} \Psi(P(\varepsilon))\big|_{\varepsilon=0}$

$$\hookrightarrow \Psi(P(\varepsilon)) = \sum_Y \sum_a \sum_w Y\left(1+\varepsilon S_Y(w,a,y)\right) P(Y|a,w)$$
$$\left(1+\varepsilon S_A(a-\delta, w)\right) g(a-\delta|w)$$
$$\left(1+\varepsilon S_W(w)\right) Q_W(w)$$

Note: all we've done is replace functions in $(*)$ w/ our score fcnes.

$$\frac{\partial}{\partial \varepsilon}\Psi(P(\varepsilon))\Big|_{\varepsilon=0} = \sum_Y\sum_a\sum_w Y\left(s_Y(w,a,y) + s_A(a-\delta,w) + s_W(w)\right) P(Y|a,w)\, g(a-\delta,w)\, Q_W(w)$$

$$= \sum\sum\sum Y \frac{g(a-\delta|w)}{g(a|w)} (S_Y + S_W) \underbrace{P(Y|a,w)\, g(a|w)\, Q_W(w)}_{\to P}$$

$$+ \sum\sum\sum Y\, \underbrace{S_A(a-\delta,w) P(Y|a,w)\, g(a-\delta|w)\, Q_W(w)}_{(**)\text{ need to figure this out}} \Rightarrow \text{let } a' = a-\delta$$

$$\Rightarrow (**) = \sum_{a'}\sum_w\sum_Y \underbrace{Y P(Y|a'+\delta, w)}_{\bar{Q}(a+\delta,w)} S_A(a',w)\, g(a'|w)\, Q_W(w)$$

$$= \sum_a\sum_w \bar{Q}(a+\delta, w)\, S_A\, g(a|w)\, Q_W(w)$$

$$= E_P\left[\bar{Q}(A+\delta, W) S_A\right]$$

$$\text{So, } \frac{\partial}{\partial\varepsilon}\Psi(P(\varepsilon))\Big|_{\varepsilon=0} = E_P\left[\frac{g(A-\delta|W)}{g(A|W)} Y (S_Y + S_W)\right] + E_P\left[\bar{Q}(A+\delta, W) S_A\right]$$

$$= E_P\left[\left(\frac{g(A-\delta|W)}{g(A|W)} Y + \bar{Q}(A+\delta, W)\right)(S_Y + S_W + S_A)\right]$$

$$- E_P\left[\frac{g(A-\delta|W)}{g(A|W)} Y S_A\right] - E_P\left[\bar{Q}(A+\delta, W) S_W\right]$$

10/11/2011 Oracle selector = defined by looking at true risk of candidate applied to training sample $\Rightarrow$ and mins true risk.

For loss: $L(\hat{Q}_K, Q_0) = \underbrace{P_0 L(\hat{Q}_K) - \underbrace{P_0 L(Q_0)}_{\text{min at } Q_0}}_{\text{dissimilarities}}$

tells us we will do as well as oracle,

* loss function has to be uniformly bounded.

CV-Risk: $E_{P_n}\left[P^1_{nB_n} L(\hat{Q}\, P^0_{nB_n})\right]$

<u>Super Learner</u>

$\hat{Q}_K : M_{NP} \to \mathbb{R}$ $\Rightarrow$ we want estimator

$\Rightarrow \left\{\hat{Q}_\alpha = \sum_{K=1}^{K} \alpha_K \hat{Q}_K, \vec{\alpha}\right\}$

$\alpha_n = \underset{\alpha}{\operatorname{argmin}}\, E_{B_n}\left[P_{nB_n} L\left(\hat{Q}_\alpha(P^0_{nB_n})\right)\right]$

This pencast was created with my Livescribe™ smartpen. Click here to see how to create your own.