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

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

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

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

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

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

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

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

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

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>

$\hat{Q}_K : M_{NP} \to \mathbb{R}$

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

$\hat{Q}_\alpha =$

$\Rightarrow$ we want estimator

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

$$\hat{\alpha}_n = \underset{\alpha}{\text{argmin}} \frac{1}{V} \sum_{i=1}^{V} \sum_{i \in} \Big(Y_i - \sum_{k=1}^{K} \alpha_k \overbrace{\hat{Q}_k(P^0_{n,V})}^{\text{function of empirical dist including covariates}}(W_i)\Big)^2$$

normally make $\Sum \alpha_i = 1$ & $\alpha_i \geq 0$
in R, called nnls function
* works better than LS regression

CV selector

$$\frac{E\left(E_{B_n}\left(P_0 L(\hat{Q}_{k_n}(P^0_{n,B_n})) - L(\hat{Q}_0)\right)\right)}{E\left(P_0 t_{B_n}\left(L(\hat{Q}_{\hat{k}}(P^0_{n,B_n})) - L(Q_0)\right)\right)} \leq 1 + \delta + c \log K(n)/n \quad \forall \delta$$

→ asymptotically goes to 1

Note: # of candidate algorithm plays small role here... only seen in $c \log(K(n))$

! keep $L(Q_k(P_{nB_n}))$ bounded uniformly!

Risk: $R_Q(P_0) = P_0 L(Q)$

10/20/2011 Maya's slides are being used (from DC conference)
↳ & should DL it from bspace.

$$E_L\left[E[Y|L(0), A(0)=1, L(1), A(1)=1]\right] \neq E[Y_{1,1}]$$

b/c $L(1)$ is affected by $A(0)$ normally

Can est. parm $\Psi(P_0)(t) = E\left[E(Y|A(t)=1, Pa(A(t-1))) - E(Y|A(t)=0, Pa(A(t-1)))\right]$

Could estimate summary measure, instead of specific t's, using MSE

$$\underset{\beta}{\text{arg min}} \underbrace{\sum_t \left(\Psi(P_0)(t) - m_\beta(t)\right)^2 h(t)}, \quad \text{where } h(t) = \text{weights}$$

IC would be gradient applied to $\Psi(P_0)(t)$

Marko?

TMLE would fluctuate each of these t-specific models