Statistica Sinica **16**(2006): Supplement, S1∼S6

DATA-ADAPTIVE SEQUENTIAL DESIGN FOR CASE-CONTROL STUDIES

Malay Ghosh and Bhramar Mukherjee

University of Florida

Supplementary Material

Lemma 1. *Assume that* $\lim_{n\to\infty}\inf n^{-2}\sum_i\sum_j(x_i-x_j)^2h'(\gamma+\beta x_i)h'(\gamma+\beta x_j)$ >0, *a.s. and* $\sum_{i=1}^n x_i^2=O_p(n)$. *Then,*
(i) $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)=O_p(n^{-1})$;
(ii) $E(\boldsymbol{\eta}|\mathcal{D}_n)=\hat{\boldsymbol{\eta}}_n+O_p(n^{-1})$;
(iii) $E[(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)^\top|\mathcal{D}_n]=\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)+O_p(n^{-3/2})$.

Proof of Lemma 1. (i) With some algebraic manipulations,

$$|\boldsymbol{I}_n(\hat{\eta}_n)|=\sum_i\sum_j(x_i-x_j)^2h'(\gamma+\beta x_i)h'(\gamma+\beta x_j).$$

Hence, by our assumptions, $\boldsymbol{I}_n^{-1}(\boldsymbol{\eta})=O_p(n^{-1})$. By the first order Taylor expansion of $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)$ around $\boldsymbol{\eta}$, we have, $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)=O_p(n^{-1})$. This proves (i).

(ii) To establish (ii), we write,

$$E(\boldsymbol{\eta}|\mathcal{D}_n)=\hat{\boldsymbol{\eta}}_n+\frac{P_n}{Q_n},$$

where

$$P_n=\int(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)\exp\Big[-\frac{1}{2}\{(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)^\top\boldsymbol{I}_n(\hat{\eta}_n)(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)+(\boldsymbol{\eta}-\boldsymbol{m})^\top\boldsymbol{W}^{-1}(\boldsymbol{\eta}-\boldsymbol{m})\}\Big] \times\Big(1+K_n(\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_n)+R_n(\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_n)\Big)d\boldsymbol{\eta}; \quad (1)$$

and,

$$Q_n=\int\exp\Big[-\frac{1}{2}\{(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)^\top\boldsymbol{I}_n(\hat{\eta}_n)(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)+(\boldsymbol{\eta}-\boldsymbol{m})^\top\boldsymbol{W}^{-1}(\boldsymbol{\eta}-\boldsymbol{m})\}\Big] \times\Big(1+K_n(\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_n)+R_n(\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_n)\Big)d\boldsymbol{\eta}; \quad (2)$$

Now by standard square completion technique, we have the term inside the exponential of (1) and (2) as,

$$\begin{aligned}
&(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)^\top \boldsymbol{I}_n(\hat{\eta}_n)(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)+(\boldsymbol{\eta}-\boldsymbol{m})^\top \boldsymbol{W}^{-1}(\boldsymbol{\eta}-\boldsymbol{m})\\
=&\Big[\boldsymbol{\eta}-(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W}^{-1})^{-1}(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})\hat{\boldsymbol{\eta}_n}+\boldsymbol{W}^{-1}\boldsymbol{m})\Big]^\top(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W}^{-1})\\
&\times\Big[\boldsymbol{\eta}-(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W}^{-1})^{-1}(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})\hat{\boldsymbol{\eta}_n}+\boldsymbol{W}^{-1}\boldsymbol{m})\Big]\\
&+(\hat{\boldsymbol{\eta}}_n-\boldsymbol{m})^\top(\mathbf{I}_n^{-1}(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W})^{-1}(\hat{\boldsymbol{\eta}}_n-\boldsymbol{m})^\top. \qquad (3)
\end{aligned}$$

Note that,

$$\begin{aligned}
&(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W}^{-1})^{-1}(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})\hat{\boldsymbol{\eta}_n}+\boldsymbol{W}^{-1}\boldsymbol{m})\\
&=(n^{-1}\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+n^{-1}\boldsymbol{W}^{-1})^{-1}(n^{-1}\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})\hat{\boldsymbol{\eta}_n}+n^{-1}\boldsymbol{W}^{-1}\boldsymbol{m})\\
&=\hat{\boldsymbol{\eta}}_n+O_p(n^{-1}).
\end{aligned}$$

The last equality follows since $n^{-1}\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})=O_p(1)$, by assumption. Also,

$$\frac{\partial^3 l_n(\hat{\boldsymbol{\eta}})}{\partial\eta_k\partial\eta_l\partial\eta_m}\Big|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n}=O_p(n).$$

Now canceling out the common terms in P_n/Q_n, we may observe that, whenever $\boldsymbol{\eta}\sim N_2((\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W}^{-1})^{-1}(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})\hat{\boldsymbol{\eta}_n}+\boldsymbol{W}^{-1}\boldsymbol{m}),(\mathbf{I}_n(\hat{\boldsymbol{\eta}_n})+\boldsymbol{W}^{-1})^{-1})$,

$$E[(\eta_k-\hat{\eta}_{nk})(\eta_l-\hat{\eta}_{nl})(\eta_m-\hat{\eta}_{nm})(\eta_p-\hat{\eta}_{np})]=O_p(n^{-2}),$$

for all (k,l,m,p). Hence, from (1)–(3), we have, $P_n=O_p(n^{-2}.n)=O_p(n^{-1})$. Similarly, $Q_n=1+O_p(n^{-1/2})$. Thus $P_n/Q_n=O_p(n^{-1})$. This proves (ii).

(iii) For proving (iii), writing $\boldsymbol{S}_n^{-1}=\boldsymbol{I}_n(\hat{\boldsymbol{\eta}}_n)+\boldsymbol{W}^{-1}$, arguments similar to those used in (ii) give,

$$E[(\eta_i-\hat{\eta}_{ni})(\eta_j-\hat{\eta}_{nj})|\mathcal{D}_n]=s_{nij}+O_p(n^{-\frac{3}{2}}), \qquad (4)$$

for all i,j, where s_{nij} is the (i,j)-th element of $\boldsymbol{S}_n$. But, by applying a standard matrix inversion formula, we have,

$$\begin{aligned}
\boldsymbol{S}_n&=(\boldsymbol{I}_n(\hat{\boldsymbol{\eta}}_n)+\boldsymbol{W}^{-1})^{-1}\\
&=\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)-\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)(\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)+\boldsymbol{W})^{-1}\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)\\
&=\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)+O_p(n^{-\frac{3}{2}}). \qquad (5)
\end{aligned}$$

Hence, by (4) and (5), we get,

$$E[(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_n)^\top|\mathcal{D}_n]=\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\eta}}_n)+O_p(n^{-\frac{3}{2}}). \qquad (6)$$

This proves (iii) and completes the proof of Lemma 1.

Theorem 1. *For the stopping time N as defined in equation (24) of the main text, namely, for*

$$N = \inf\{n(\geq m) : n \geq (\frac{G_n}{c})^{\frac{1}{2}}\}, \tag{7}$$

where, $G_n = n\text{Var}\,(\beta|\mathcal{D}_n)$, we have,

(i) $P(N < \infty) = 1$;

(ii) $cN^2 \xrightarrow{P} [\Sigma(r^*)]^{-1}$ *as* $c \to 0$;

(iii) $L_N(c)/\rho(c) \xrightarrow{P} 1$ *as* $c \to 0$, *where* $\rho(c) = \inf_{S\in\mathcal{T}} E(L_S(c)) = 2c^{1/2}[\Sigma(r^*)]^{-1/2}$;

(iv) $E[L_N(c)]/\rho(c) \to 1$ *as* $c \to 0$. *The A.P.O. rule is first order efficient or asymptotically optimal (A.O.).*

Proof of Theorem 1. Proof of part (i) in Theorem 1 follows immediately from the definition of N.

$$\begin{aligned} P(N = \infty) &= \lim_{n\to\infty} P(N > n) \\ &\leq \lim_{n\to\infty} P(n < (\frac{G_n}{c})^{\frac{1}{2}}). \end{aligned}$$

The result follows since $G_n \xrightarrow{P} [\Sigma(r)]^{-1}$ as $n \to \infty$.

(ii) Use the inequality

$(G_N/c)^{1/2} \leq N \leq m + (G_{N-1}/c)^{1/2}$ or $G_N \leq cN^2 \leq c[m^2 + G_{N-1}/c + 2m(G_{N-1}/c)^{1/2}]$. The result follows since $G_N \xrightarrow{P} [\Sigma(r^*)]^{-1}$ as $c \to 0$.

(iii) Use the identity

$L_N(c) = N^{-1}G_N + cN = 2(cG_N)^{1/2} + N^{-1}(G_N^{1/2} - c^{1/2}N)^2$. Since the second term in the right hand side is $o_p(c^{1/2})$, the result follows by dividing all sides by $\rho(c)$. (iv) In view of (iii) it suffices to show that $L_N(c)/\rho(c)$ is uniformly integrable in $c \leq c_0$. First by the same inequality as used in (ii), for $c \leq c_0$,

$$\begin{aligned} \frac{L_N(c)}{\rho(c)} &\leq \frac{c^{\frac{1}{2}}N}{|\Sigma(r^*)|^{-\frac{1}{2}}} \leq \frac{c^{\frac{1}{2}}\left(m + \frac{G_{N-1}}{c}\right)^{\frac{1}{2}}}{|\Sigma(r^*)|^{-\frac{1}{2}}} \\ &\leq \frac{c^{\frac{1}{2}}\left(m^{\frac{1}{2}} + \frac{G_{N-1}^{\frac{1}{2}}}{c^{\frac{1}{2}}}\right)}{|\Sigma(r^*)|^{-\frac{1}{2}}} \leq \frac{c_0^{\frac{1}{2}}m^{\frac{1}{2}} + G_{N-1}^{\frac{1}{2}}}{|\Sigma(r^*)|^{-\frac{1}{2}}} \end{aligned} \tag{8}$$

Hence, it suffices to show that $G_{N-1}^{1/2}$ is uniformly integrable in $c \leq c_0$. This is equivalent to showing $n^{1/2}Var^{1/2}(\beta|\mathcal{D}_n)$ is uniformly integrable in n. This will follow if we can show that $\sup_{n\geq 1} E[n\text{Var}\,(\beta|\mathcal{D}_n)] < \infty$, where the expectation is taken over the distribution of $\mathcal{D}_n$, conditional on $\boldsymbol{\eta}$. Note that,

$$E[\text{Var}\,(\beta|\mathcal{D}_n)] = E[\text{Var}\,(\beta - \hat{\beta}_n|\mathcal{D}_n)] \leq \text{Var}\,(\beta - \hat{\beta}_n). \tag{9}$$

Following Cox and Snell (1968),

$$E(\hat{\beta}_n - \beta|\boldsymbol{\eta}) = \frac{K_1(\boldsymbol{\eta})}{n} + O(n^{-2}),$$

and

$$E[(\hat{\beta}_n - \beta)^2|\boldsymbol{\eta}] = \frac{K_2(\boldsymbol{\eta})}{n} + O(n^{-2}),$$

where $K_1(\boldsymbol{\eta})$ and $K_2(\boldsymbol{\eta})$ are polynomials in the elements of $\boldsymbol{\eta}$. Hence, conditional on $\boldsymbol{\eta}$,

$$n\text{Var}\,(\beta - \hat{\beta}_n) = nE[(\hat{\beta}_n - \beta)^2] - n(E[(\hat{\beta}_n - \beta)])^2 < \infty, \tag{10}$$

uniformly in n. Combining (9) and (10), one obtains, $E[n\text{Var}\,(\beta|\mathcal{D}_n)] < \infty$, hence the proof of (iv).

Suppose T denotes the stopping time for the ACTUAL Bayes rule. Then

$$\begin{aligned} L_T(c) &= T^{-1}G_T + cE(T) \\ &= 2(cG_T)^{\frac{1}{2}} + T^{-1}(G_T^{\frac{1}{2}} - c^{\frac{1}{2}}T)^2 \geq 2(cG_T)^{\frac{1}{2}}. \end{aligned}$$

Bickel and Yahav (1967) have shown that $T/N \to 1$ a.s. as $c \to 0$. Hence, with the same sampling rule as defined in Section 3.1 of the main text, $G_T \xrightarrow{P} [\Sigma(r^*)]^{-1}$ as $c \to 0$. Hence, from the above inequality, and Fatou's Lemma,

$$\liminf_{c \to 0} \frac{E[L_T(c)]}{\rho(c)} \geq 1.$$

But $E[L_T(c)] \leq E[L_N(c)]$ for all c. Hence,

$$\limsup_{c \to 0} \frac{E[L_T(c)]}{\rho(c)} \leq \limsup_{c \to 0} \frac{E[L_N(c)]}{\rho(c)} = 1.$$

Thus $E[L_T(c)]/\rho(c) \to 1$ as $c \to 0$. In other words, the A.P.O. rule N is first order efficient.

Proof of equation (27) in the main text. Equation (27) in the main text states that the expression for $\Sigma(r)$, in the situation with a binary exposure is given by

$$\Sigma(r) = (1-r)\frac{h(\gamma^*(r)+\beta)h(\lambda)h(\gamma^*(r))\overline{h}(\lambda)}{h(\gamma^*(r)+\beta)h(\lambda)+h(\gamma^*(r))\overline{h}(\lambda)}. \tag{11}$$

The expression for $\Sigma(r)$ as given in (8)-(11) of the main text, in the bivariate binary case, may be explicitly computed as follows. Note that the case-control sampling model implies that,

$$\phi_1(x) \propto h(\gamma + \beta x)\phi(x) \qquad \text{and} \qquad \phi_0(x) \propto \overline{h}(\gamma + \beta x)\phi(x),$$

where $\phi(x)$ is the marginal distribution of X. Also,

$$p_1 = \int h(\gamma + \beta x)\phi(x)dx \qquad \text{and} \qquad (1 - p_1) = \int \overline{h}(\gamma + \beta x)\phi(x)dx.$$

This observation leads to the useful basic identity

$$\frac{\phi_1(x)}{\phi_0(x)} = \frac{1 - p_1}{p_1}\exp(\gamma + \beta x). \tag{12}$$

Using (12) in the expression for $A(r)$ in (12) of the main text, we have,

$$\begin{aligned} A(r) &= \frac{E_0[Xu(\gamma^*(r) + \beta X)\{r\frac{1-p_1}{p_1}\exp(\gamma + \beta X) + (1 - r)\}]}{E_0[u(\gamma^*(r) + \beta X)\{r\frac{1-p_1}{p_1}\exp(\gamma + \beta X) + (1 - r)\}]} \\ &= \frac{E_0[Xu(\gamma^*(r) + \beta X)\{1 + \exp(\gamma^*(r) + \beta X)\}]}{E_0[u(\gamma^*(r) + \beta X)\{1 + \exp(\gamma^*(r) + \beta X)\}]} \\ &= \frac{E_0[Xh(\gamma^*(r) + \beta X)]}{E_0[h(\gamma^*(r) + \beta X)]}. \end{aligned} \tag{13}$$

Table 1. True values of the parameters: $\lambda = -1$, $\beta = 0$, r^*=0.5, $g(r^* = 0.5, \lambda = -1, \beta = 0)$=20.345. Prior parameters: $\mu_\lambda = \mu_\beta = 0$, $\sigma_\lambda = \sigma_\beta = 4, \rho = 0.5$. $\hat{\beta}_{APM}$ denotes the posterior mean obtained by using the Laplace approximation, $\hat{\beta}_{MCMC}$ is the exact posterior mean as obtained by implementing the MCMC numerical integration scheme based on the data at stopping time N. The quantities in the parentheses denote the respective MSE's as estimated from the 500 replications.

c	Mean(N)	Mean(r_N)	Mean(cN^2)	$\hat{\beta}_{MLE}$	$\hat{\beta}_{APM}$	$\hat{\beta}_{MCMC}$
	(Var(N))	(Var(r_N))	(Var(cN^2))	(MSE($\hat{\beta}_{MLE}$))	(MSE($\hat{\beta}_{APM}$))	(MSE($\hat{\beta}_{MCMC}$))
0.05	21.97	0.4976	24.57	0.0131	-0.0125	0.0125
	(8.82)	(0.00488)	(52.28)	(0.7955)	(0.7028)	(0.6879)
0.02	34.58	0.5007	24.28	-0.0251	-0.0420	-0.0398
	(18.43)	(0.00354)	(42.54)	(0.7092)	(0.6275)	(0.6441)
0.005	66.34	0.5046	22.17	-0.0186	-0.0236	-0.0199
	(33.68)	(0.00115)	(18.28)	(0.2722)	(0.2623)	(0.2676)
0.001	143.85	0.5000	20.73	-0.0011	-0.0009	-0.0010
	(53.82)	(0.00051)	(2.84)	(0.1494)	(0.1453)	(0.1421)
0.0001	452.99	0.5024	20.53	-0.0041	0.0032	-0.0042
	(116.78)	(0.00011)	(0.96)	(0.0453)	(0.0451)	(0.0451)

Simulation results for the null case $\beta = 0$.

Next, by (12), (13) and (9) we have,

$$
\begin{aligned}
\Sigma(r) &= E_0[\{x - A(r)\}^2 u(\gamma^*(r) + \beta X)\{r\frac{1-p_1}{p_1}\exp(\gamma + \beta X) + (1-r)\}] \\
&= (1-r)E_0[\{X - A(r)\}^2 h(\gamma^*(r) + \beta X)] \\
&= (1-r)\left[E_0\{X^2 h(\gamma^*(r) + \beta X)\} - \frac{\{E_0(Xh(\gamma^*(r) + \beta X))\}^2}{E_0(h(\gamma^*(r) + \beta X))}\right] \\
&= (1-r)\left[h(\gamma^*(r) + \beta))h(\lambda) - \frac{h^2(\gamma^*(r) + \beta)h^2(\lambda)}{h(\gamma^*(r))\overline{h}(\lambda) + h(\gamma^*(r) + \beta)h(\lambda)}\right] \\
&= \frac{h(\gamma^*(r) + \beta)h(\lambda)h(\gamma^*(r))\overline{h}(\lambda)}{h(\gamma^*(r) + \beta)h(\lambda) + h(\gamma^*(r))\overline{h}(\lambda)}. \qquad (14)
\end{aligned}
$$

Note that in evaluating the expectation E_0, we used the fact that under ϕ_0, $X \sim$ Bernoulli($h(\lambda)$).