Appendix: Consistency and normality of the MLE

For any $\delta > 0$, let $\mathcal{Q}_{k\delta} = \{Q_k \in P_{\mathcal{Y}_k} | \sup_{y_k \in \mathcal{Y}_k} |Q_k(y_k) - Q_{k0}(y_k)| < \delta\}$, $k = 1, 2$. Before showing the consistency of the maximum likelihood estimator, we state a few lemmas that are used in the proof of consistency. Proofs of these lemmas are omitted.

Lemma 1. Suppose that $F_n(y) \to F(y)$ weakly, and that $\varphi(y, h)$ is uniformly bounded and equicontinuous with respect to y over $h \in H$. That is, for any $\epsilon > 0$, there exists $\delta > 0$, and a finite set of $h_k \in H$, $k = 1, \cdots, K_\epsilon$ such that

$$\sup_{|y_1 - y_2| < \delta} \sup_{h \in H} \left\{ \min_k |\varphi(y_1, h) - \varphi(y_2, h_k)| \right\} < \epsilon.$$

Then,

$$\lim_{n \to \infty} \sup_{h \in H} \left| \int \varphi(y, h) dF_n(y) - \int \varphi(y, h) dF(y) \right| = 0.$$

Lemma 2. Suppose that $\sup_{y \in \mathcal{D}} |F_n(y) - F(y)| \to 0$, $\sup_{y \in \mathcal{D}, h \in H} |\varphi_n(y, h) - \varphi(y, h)| \to 0$, and $\sup_{h \in H} ||\varphi(y, h)||_{BV(y)} < +\infty$ where $\mathcal{D}$ is a measurable set and $|| \cdot ||_{BV(y)}$ denotes the variation norm in $y \in \mathcal{Y}$. Then

$$\lim_{n \to \infty} \sup_{h \in H} \left| \int_{\mathcal{D}} \varphi_n(y, h) dF_n(y) - \int_{\mathcal{D}} \varphi(y, h) dF(y) \right| = 0.$$

Lemma 3. Under conditions 1-4, for a single observation, $\{\dot{l}(\gamma, Q_1, Q_2)(h), \gamma \in B, Q_k \in \mathcal{Q}_{k\delta}, k = 1, 2, h \in H_0 \times H_1 \times H_2\}$ is a P_0-Donsker class and $\{\ddot{l}(\gamma, Q_1, Q_2)(h)(h^*), \gamma \in B, Q_k \in \mathcal{Q}_{k\delta}, k = 1, 2, h^*, h \in H_0 \times H_1 \times H_2\}$ is a P_0-Glivenko-Cantelli class.

Lemma 4. Let $b^{-1}(y_3) = \int \int \eta(y_1; y_2; y_3 | \gamma_0) dQ_{10}(y_1) dQ_{20}(y_2)$,

$$\begin{aligned} Q_{n10}(y_1) &= \frac{1}{n} \sum_{i=1}^n \frac{1_{\{Y_{i1} \le y_1\}}}{\frac{1}{n} \sum_{j=1}^n b(Y_{j3}) \int \eta(Y_{i1}; y_2; Y_{j3} | \gamma_0) dQ_{20}(y_2)}, \\ Q_{n20}(y_2) &= \frac{1}{n} \sum_{i=1}^n \frac{1_{\{Y_{i2} \le y_2\}}}{\frac{1}{n} \sum_{j=1}^n b(Y_{j3}) \int \eta(y_1; Y_{i2}; Y_{j3} | \gamma_0) dQ_{10}(y_1)}. \end{aligned}$$

Then $Q_{nk0}(y_k) \to Q_{k0}(y_k)$, uniformally over $y_k \in \mathcal{Y}_k$ P_0-almost sure for $k = 1, 2$. Further,

$$\sup_{y_k \in \mathcal{Y}_k, k=1,2,3} |P(y_1, y_2 \mid y_3, \gamma_0, Q_{n10}, Q_{n20}) - P(y_1, y_2 \mid y_3, \gamma_0, Q_{10}, Q_{20})| \to 0,$$

P_0-almost sure, where $P(y_1, y_2 \mid y_3, \gamma, Q_1, Q_2)$ is the distribution function of $p(y_1, y_2 \mid y_3, \gamma, Q_1, Q_2)$.

Proofs of the theorem are divided into three steps. They are proof of existence and consistency, proof of normality, and proof of consistency of the variance estimate.

Proof of existence and consistency of the maximum likelihood estimator: The likelihood has the form

$$\prod_{i=1}^{n} \frac{\eta(Y_{i1}; Y_{i2}; Y_{i3}|\gamma) dQ_1(Y_{i1}) dQ_2(Y_{i2})}{\int \int \eta(y_1; y_2; Y_{i3}|\gamma) dQ_1(y_1) dQ_2(y_2)}.$$

Direct maximization of the likelihood over all distributions for Q_1 and/or Q_2 does not exist. We follow the convention of maximizing the likelihood over all discrete distributions for Q_1 and Q_2. It can be seen that, for any discrete Q_1 and Q_2, there are Q_1 and Q_2 having probability masses only at the observed Y_1 and Y_2 values that make the likelihood no smaller. This means the maximizer, if exists, is achieved with Q_1 and Q_2 having probability masses only at the observed Y_1 and Y_2 values. The likelihood is then a multivariate function of $(\gamma; q_{k1}, k = 1, \cdots, N_1; q_{j2}, j = 1, \cdots, N_2)$ with $\sum_{k=1}^{N_1} q_{k1} = 1$ and $\sum_{j=1}^{N_2} q_{j2} = 1$. From Assumptions 2 and 3, it follows that the likelihood, as a function of $(\gamma, q_{11}, \cdots, q_{(N_1-1)1}, q_{12}, \cdots, q_{(N_2-1)2})$ with $\gamma \in B$ and the others parameters are respectively in

$$\mathcal{Q}_1 = \left\{ (q_{11}, \cdots, q_{(N_1-1)1}) \middle| \sum_{k=1}^{N_1} q_{k1} = 1 \right\} \text{ and } \mathcal{Q}_2 = \left\{ (q_{12}, \cdots, q_{(N_2-1)2}) \middle| \sum_{k=1}^{N_2} q_{k2} = 1 \right\},$$

is continuous. It follows from that $B \times \mathcal{Q}_1 \times \mathcal{Q}_2$ is a compact set that the maximum of the likelihood exists.

From assumption 2, we see that when $dQ_1(Y_{i1}) = 0$ or $dQ_2(Y_{i2}) = 0$ for any $i = 1, \cdots, n$, the likelihood becomes zero. This implies that, for any fixed γ, the maximum of the restricted likelihood over $Q_k, k = 1, 2$, exists and is attained respectively at inner points of $\mathcal{Q}_1 \times \mathcal{Q}_2$. The restricted maximum likelihood estimators of $Q_k, k = 1, 2$, satisfy the score equations derived from

$$P_n l(\gamma, Q_1, Q_2) + \lambda \left(\sum_{k=1}^{N_1} q_{k1} - 1 \right) + \mu \left(\sum_{k=1}^{N_2} q_{k2} - 1 \right),$$

where $P_n l(\gamma, Q_1, Q_2)$ is the log-likelihood. Denote the maximum restricted likelihood estimator by

$(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)$. The score equations for $\hat{Q}_1$ are

$$\frac{\sum_{i=1}^n 1_{\{Y_{i1}=y_{k1}\}}}{nq_{k1}} - \frac{1}{n}\sum_{i=1}^n \frac{\sum_{j=1}^{N_2} \eta(y_{k1}; y_{j2}; Y_{i3}|\hat{\gamma}_n) d\hat{Q}_2(y_{j2})}{\sum_{l=1}^{N_1}\sum_{j=1}^{N_2} \eta(y_{l1}; y_{j2}, Y_{i3}|\hat{\gamma}_n) d\hat{Q}_1(y_{l1}) d\hat{Q}_2(y_{j2})} + \lambda = 0,$$

for $k = 1, \cdots, N_1$. It follows from $\sum_{k=1}^{N_1} q_{k1} = 1$ that $\lambda = 0$. From this, we see that $\hat{Q}_1(y_1) = \sum_{k=1}^{N_1} q_{k1} 1_{\{y_{k1} \le y_1\}}$, where

$$q_{k1} = \frac{1}{n}\sum_{i=1}^n \frac{1_{\{Y_{i1}=y_{k1}\}}}{D_1(y_{k1}, \hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)},$$

and

$$D_1(y_1, \hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2) = \frac{1}{n}\sum_{i=1}^n \frac{\sum_{j=1}^{N_2} \eta(y_1; y_{j2}; Y_{i3}|\hat{\gamma}_n) d\hat{Q}_2(y_{j2})}{\sum_{l=1}^{N_1}\sum_{j=1}^{N_2} \eta(y_{l1}; y_{j2}, Y_{i3}|\hat{\gamma}_n) d\hat{Q}_1(y_{l1}) d\hat{Q}_2(y_{j2})}.$$

Note that the existence of $\hat{Q}_1$ in the inner part of the set guarentees the maximizer $q_{k1} > 0$ for all $k = 1, \cdots, N_1$ and satisfies $\hat{Q}_1(y_1) = \sum_k q_{k1} 1_{\{y_{k1} \le y_1\}}$. From the expression, we see that

$$\frac{d\hat{Q}_1}{dQ_{n10}}(y_{k1}) = \frac{1}{D_1(y_{k1}, \hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)} \left\{ \frac{1}{\frac{1}{n}\sum_{j=1}^n b(Y_{j3}) \int \eta(y_{k1}; y_2; Y_{j3}|\gamma_0) dQ_{20}(y_2)} \right\}^{-1}.$$

Similar derivation can be carried out for $\hat{Q}_2$ to obtain

$$\frac{d\hat{Q}_2}{dQ_{n20}}(y_{k2}) = \frac{1}{D_2(y_{k2}, \hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)} \left\{ \frac{1}{\frac{1}{n}\sum_{j=1}^n b(Y_{j3}) \int \eta(y_1; y_{k2}; Y_{j3}|\gamma_0) dQ_{10}(y_1)} \right\}^{-1},$$

where D_2 is defined similarly as D_1. Since γ belongs to a compact set, $\hat{\gamma}_n$ has a convergent subsequence. Denote the limit by γ. From Helly selection Theorem (Fuchino and Plewik, 1999), there also exists a subsequence of $\hat{Q}_k$ that converges to a subdistribution Q_k on the continuous points of Q_k, for $k = 1, 2$. Since $\mathcal{Y}_k$ is compact, which ensure Q_k is a distribution function. Hence, $\hat{Q}_k$ converges weakly to Q_k, for $k = 1, 2$.

By repeatedly applying Lemma 1, it follows that $D_1(y_1, \hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)$ converges uniformly to

$$D_1(y_1, \gamma, Q_1, Q_2) = \int \frac{\int \eta(y_1; y_2; y_3|\gamma) dQ_2(y_2)}{\int\int \eta(y_1; y_2, y_3|\gamma) dQ_1(y_1) dQ_2(y_2)} p_{30}(y_3) d\mu_3(y_3).$$

Similarly, $D_2(y_2, \hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)$ converges uniformly to $D_2(y_2, \gamma, Q_1, Q_2)$, defined similarly as D_1. These imply that $Q_1(y_1)$ satisfies

$$Q_1(y_1) = \int_{-\infty}^{y_1} \left\{ \int \frac{\int \eta(y_1; y_2; y_3|\gamma) dQ_2(y_2)}{\int\int \eta(y_1; y_2, y_3|\gamma) dQ_1(y_1) dQ_2(y_2)} p_{30}(y_3) d\mu_3(y_3) \right\}^{-1} p_{10}(y_1) d\mu_1(y_1)$$

and $\hat{Q}_1(y_1)$ converges uniformly over $\mathcal{Y}_1$ to $Q_1(y_1)$. Similar definition and convergence result hold for $Q_2(y_2)$.

Further, it follows that

$$-\frac{1}{n}\sum_{i=1}^{n}\log D_1(Y_{i1},\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)\to-\int\{\log D_1(y_1,\gamma,Q_1,Q_2)\}\,p_{10}(y_1)d\mu_1(y_1).$$

and

$$-\frac{1}{n}\sum_{i=1}^{n}\log D_2(Y_{i2},\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)\to-\int\{\log D_2(y_2,\gamma,Q_1,Q_2)\}\,p_{20}(y_2)d\mu_2(y_2).$$

It now follows that

$$\begin{aligned}
P_n\left\{l(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)-l(\gamma_0,Q_{1n0},Q_{2n0})\right\} &= P_n\left\{\log\frac{d\hat{Q}_1}{dQ_{n10}}(Y_1)\right\}+P_n\left\{\log\frac{d\hat{Q}_2}{dQ_{n20}}(Y_2)\right\}\\
&+P_n\left\{\log\frac{\eta(Y_1;Y_2;Y_3\mid\hat{\gamma}_n)}{\eta(Y_1;Y_2;Y_3\mid\gamma_0)}\right\}-P_n\left\{\log\frac{\int\int\eta(y_1;y_2,Y_3\mid\hat{\gamma}_n)d\hat{Q}_1(y_1)d\hat{Q}_2(y_2)}{\int\int\eta(y_1;y_2,Y_3\mid\gamma_0)dQ_{n10}(y_1)dQ_{n20}(y_2)}\right\}\\
&\to P_0\left\{l(\gamma,Q_1,Q_2)-l(\gamma_0,Q_{10},Q_{20})\right\}\geq 0.
\end{aligned}$$

It follows from Assumption 1, i.e., (γ_0,Q_{10},Q_{20}) is the unique maximizer of the Kullback-Leibler information, that $(\gamma,Q_1,Q_2)=(\gamma_0,Q_{10},Q_{20})$. Since every sequence has a convergent subsequence and every convergent subsequence has the same limit, the sequence $(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)$ converges to the common limit (γ_0,Q_{10},Q_{20}).

Proof of asymptotic normality: Note that the maximum likelihood estimator satisfies the score equation,

$$P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)(h_0,h_{1n},h_{2n})=0,$$

where

$$\begin{aligned}
\dot{l}_{(\gamma,Q_1,Q_2)}(Y_1,Y_2,Y_3)(h_0,h_1,h_2) &= h_0^T\left[\frac{\partial}{\partial\gamma}\log\eta(Y_1;Y_2;Y_3\mid\gamma)-E\left\{\frac{\partial}{\partial\gamma}\log\eta(Y_1;Y_2;Y_3\mid\gamma)\Big|Y_3\right\}\right]\\
&+h_1(Y_1)-E\left\{h_1(Y_1)|Y_3\right\}+h_2(Y_2)-E\left\{h_2(Y_2)|Y_3\right\},
\end{aligned}$$

with E being the expectation taken under the odds ratio model with parameter (γ,Q_1,Q_2), h_0 is bounded, $h_{1n}\in H_{1n}$ is a collection of bounded variation functions with jumps only at the observed

y_1 values and $\int h_{1n}(y_1)d\hat{Q}_1(y_1) = 0$, and $h_{2n} \in H_{2n}$ is a collection of bounded variation functions with jumps only on observed y_2 values and $\int h_{2n}(y_2)d\hat{Q}_2(y_2) = 0$. Define an inner product on $H_0 \times H_1 \times H_2$ as

$$< (h_0, h_1, h_2), (h_0^*, h_1^*, h_2^*) >= h_0^T h_0^* + \int h_1(y_1)h_1^*(y_1)dQ_{10}(y_1) + \int h_2(y_2)h_2^*(y_2)dQ_{20}(y_2).$$

Note that

$$\begin{aligned}\sqrt{n}P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)(h_0,h_{n1},h_{n2}) &= P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)\{0,\sqrt{n}(h_{1n}-h_1),\sqrt{n}(h_{2n}-h_2)\} \\ &\quad +\sqrt{n}P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)(h_0,h_1,h_2),\end{aligned}$$

For any $h_1 \in H_1$ and $h_2 \in H_2$, there are $h_{1n} \in H_1$ and $h_{2n} \in H_2$ such that

$$\sqrt{n}(h_{1n} - h_1) = O_{P_0}(1) \text{ and } \sqrt{n}(h_{2n} - h_2) = O_{P_0}(1).$$

Since $\{\dot{l}_{(\gamma,Q_1,Q_2)}(Y_1,Y_2,Y_3)(h_0,h_1,h_2)\}$ is a Glivenko-Cantelli class for $|\gamma - \gamma_0| < \delta$, $\sup_{y_1} |Q_1(y_1) - Q_{10}(y_1)| < \delta$, $\sup_{y_2} |Q_2(y_2) - Q_{20}(y_2)| < \delta$, and bounded (h_0, h_1, h_2), it follows that

$$\sqrt{n}P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)(h_0,h_{n1},h_{n2}) = \sqrt{n}P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)(h_0,h_1,h_2) + o_{P_0}(1),$$

because $P_0\dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)(h_0^*,h_1^*,h_2^*) = 0$ for any $(h_0^*,h_1^*,h_2^*) \in H_0 \times H_1 \times H_2$. Note that

$$\begin{aligned} o_{P_0}(1) &= \sqrt{n}P_n\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)(h_0,h_1,h_2) \\ &= -\sqrt{n}P_0\{\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3) - \dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)\}(h_0,h_1,h_2) \\ &\quad +\sqrt{n}(P_n - P_0)\dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)(h_0,h_1,h_2) \\ &\quad +\sqrt{n}(P_n - P_0)\{\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3) - \dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)\}(h_0,h_1,h_2).\end{aligned}$$

From Lemma 2 and

$$P_0\{\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3) - \dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)\}^2(h_0,h_1,h_2) \to 0$$

uniformly, it follows that

$$\sqrt{n}(P_n - P_0)\{\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3) - \dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)\}(h_0,h_1,h_2) = o_{P_0}(1).$$

Further, since $\dot{l}_{(\gamma,Q_1,Q_2)}(Y_1,Y_2,Y_3)(h_0,h_1,h_2)$ is Frechet differentiable with respect to (γ,Q_1,Q_2) at (γ_0,Q_{10},Q_{20}), it follows that

$$\begin{aligned}
&-\sqrt{n}P_0\{\dot{l}_{(\hat{\gamma}_n,\hat{Q}_1,\hat{Q}_2)}(Y_1,Y_2,Y_3)-\dot{l}_{(\gamma_0,Q_{10},Q_{20})}(Y_1,Y_2,Y_3)\}(h_0,h_1,h_2)\\
&\quad= \sqrt{n}(\hat{\gamma}_n-\gamma_0)^T\sigma_0(h_0,h_1,h_2)+\int\sigma_1(h_0,h_1,h_2)d\sqrt{n}(\hat{Q}_1-Q_{10})\\
&\qquad+\int\sigma_2(h_0,h_1,h_2)d\sqrt{n}(\hat{Q}_2-Q_{20})+o_{P_0}(|\sqrt{n}(\hat{\beta}-\beta_0)|+||\sqrt{n}(\hat{Q}_1-Q_{10})||+||\sqrt{n}(\hat{Q}_2-Q_{20})||),
\end{aligned}$$

where σs are

$$\begin{aligned}
\sigma_0(h_0,h_1,h_2) &= -E_0\left\{\frac{\partial^2\log\eta}{\partial\gamma^2}h_0-E_{(\gamma_0,Q_{10},Q_{20})}\left(\frac{\partial^2\log\eta}{\partial\gamma^2}h_0\Big|Y_3\right)\right\}\\
&\quad+E_0\left[\mathrm{Cov}_{(\gamma_0,Q_{10},Q_{20})}\left\{\frac{\partial\log\eta}{\partial\gamma},\frac{\partial\log\eta}{\partial\gamma}h_0+h_1(Y_1)+h_2(Y_2)\Big|Y_3\right\}\right],\\
\sigma_1(h_0,h_1,h_2) &= E_0\left[\left\{E_{(\gamma_0,Q_{10},Q_{20})}\left(\frac{\partial\log\eta}{\partial\gamma}h_0\Big|Y_1=y_1,Y_3\right)-E_{(\gamma_0,Q_{10},Q_{20})}\left(\frac{\partial\log\eta}{\partial\gamma}h_0\Big|Y_3\right)\right\}p_1(y_1|Y_3)\right]\\
&\quad+h_1(y_1)E_0\{p_1(y_1|Y_3)\}-E_0[E_{(\gamma_0,Q_{10},Q_{20})}\{h_1(Y_1)|Y_3\}q_1(y_1|Y_3)]\\
&\quad+E_0[E_{(\gamma_0,Q_{10},Q_{20})}\{h_2(Y_2)|Y_1=y_1,Y_3\}p_1(y_1|Y_3)-E_{(\gamma_0,Q_{10},Q_{20})}\{h_2(Y_2)|Y_3\}p_1(y_1|Y_3)],\\
\sigma_2(h_0,h_1,h_2) &= E_0\left[\left\{E_{(\gamma_0,Q_{10},Q_{20})}\left(\frac{\partial\log\eta}{\partial\gamma}h_0\Big|Y_2=y_2,Y_3\right)-E_{(\gamma_0,Q_{10},Q_{20})}\left(\frac{\partial\log\eta}{\partial\gamma}h_0\Big|Y_3\right)\right\}p_2(y_2|Y_3)\right]\\
&\quad+E_0[E_{(\gamma_0,Q_{10},Q_{20})}\{h_1(Y_1)|Y_2=y_2,Y_3\}p_2(y_2|Y_3)-E_{(\gamma_0,Q_{10},Q_{20})}\{h_1(Y_1)|Y_3\}p_2(y_2|Y_3)]\\
&\quad+h_2(y_2)E_0\{p_2(y_2|Y_3)\}-E_0[E_{(\gamma_0,Q_{10},Q_{20})}\{h_2(Y_2)|Y_3\}p_2(y_2|Y_3)].
\end{aligned}$$

where

$$\begin{aligned}
p_1(y_1|y_3) &= \frac{\int\eta(y_1;y_2;y_3|\gamma_0)dQ_{20}(y_2)}{\int\int\eta(y_1;y_2;y_3|\gamma_0)dQ_{10}(y_1)dQ_{20}(y_2)},\\
p_2(y_2|y_3) &= \frac{\int\eta(y_1;y_2;y_3|\gamma_0)dQ_{10}(y_1)}{\int\int\eta(y_1;y_2;y_3)dQ_{10}(y_1)dQ_{20}(y_2)}.
\end{aligned}$$

We show in the following that $\sigma(h_0,h_1,h_2)=(\sigma_0,\sigma_1,\sigma_2)(h_0,h_1,h_2)$ is a continuous invertible operator on $H_0\times H_1\times H_2$. Note from assumption 5 that

$$<h,\sigma(h)>=-E_0\{\ddot{l}(h,h)\}\geq 0,$$

and the equality holds only when $h=0$. This means that σ is a one-to-one map on $H_0\times H_1\times H_2$. Next, let

$$\sigma(h_0,h_1,h_2)=A(h_0,h_1,h_2)+B(h_0,h_1,h_2),$$

where $A(h_0, h_1, h_2) = (a_0^T h_0, E_0\{p_1(y_1|Y_3)\}h_1(y_1), E_0\{p_2(y_2|Y_3)\}h_2(y_2))$ and

$$a_0 = E_0 \left[Var_{(\gamma_0, Q_{10}, Q_{20})} \left\{ \frac{\partial \log \eta}{\partial \gamma}(Y_1; Y_2; Y_3|\gamma_0) \Big| Y_3 \right\} \right] > 0.$$

Hence, A is continous invertible on $H_0 \times H_1 \times H_2$. Further, it can be routinely verified that B is compact by applying Theorem 2 of Kantorovich and Akilov (1982, pages 326-328) which states that the integral operator on $L^2(P_0)$ is compact if its kernel is summable in $L^2(P_0)$. It then follows from that σ is one-to-one on $H_0 \times H_1 \times H_2$ that σ is continuous invertible on $H_0 \times H_1 \times H_2$.

From that

$$\sqrt{n}(P_n - P_0)\dot{l}(Y_1, Y_2; Y_3, \gamma_0)(h_0, h_1, h_2)$$

converges weakly to a tight Gaussian process, it follows that

$$\begin{aligned} &\sqrt{n}(\hat{\gamma}_n - \gamma_0)^T h_0 + \int h_1(y_1) d\sqrt{n}(\hat{Q}_1 - Q_{10}) + \int h_2(y_2) d\sqrt{n}(\hat{Q}_2 - Q_{20}) \\ &= \sqrt{n}(P_n - P_0)\dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)\sigma^{-1}(h_0, h_1, h_2) + o_{P_0}(1), \end{aligned}$$

which converges to a tight Gaussian process with the covariance function

$$P_0 \left\{ \dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)\sigma^{-1}(h_0, h_1, h_2)\dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)\sigma^{-1}(h_0^*, h_1^*, h_2^*) \right\}.$$

It follows from setting $(h_1, h_2) = 0$ that

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0)^T h_0 \to N(0, v(h_0))$$

for any bounded $h_0 \in H_0$, where $v(h_0) = P_0 \left\{ \dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)\sigma^{-1}(h_0, 0, 0) \right\}^2$. This implies that $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ is asymptotically multivariate normal with mean 0 and variance $V = (v_{ij})$ whose element can be obtained by $v_{ij} = \{v(e_i + e_j) - v(e_i) - v(e_j)\}/2$, where e_i is the unit vector with the ith component 1 and all other components 0.

Proof of consistent variance estimate: Let $\hat{\sigma}$ be defined as σ with P_0 and $(\gamma_0, Q_{10}, Q_{20})$ in the expression respectively replaced by P_n and $(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)$. To show the consistency of the variance estimate, note first from Lemma 3 that,

$$\hat{\sigma}(h_0, h_1, h_2)(h_0^*, h_1^*, h_2^*) \to \sigma(h_0, h_1, h_2)(h_0^*, h_1^*, h_2^*)$$

uniformly P_0-almost sure over bounded $h, h^* \in H_0 \times H_1 \times H_2$. it follows from that σ is continuously invertible that $\hat{\sigma}$ is asymptotically continuously invertible almost sure and

$$\hat{\sigma}^{-1}(h_0, h_1, h_2)(h_0^*, h_1^*, h_2^*) \to \sigma^{-1}(h_0, h_1, h_2)(h_0^*, h_1^*, h_2^*)$$

uniformly P_0-almost sure over bounded $h, h^* \in H_0 \times H_1 \times H_2$. Note that

$$\begin{aligned} P_n \left\{ \dot{l}_{(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)}(Y_1; Y_2; Y_3)(h) \right\}^2 &= P_0 \left\{ \dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)(h) \right\}^2 + (P_n - P_0) \left\{ \dot{l}_{(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)}(Y_1; Y_2; Y_3)(h) \right\}^2 \\ &\quad + P_0 \left[\left\{ \dot{l}_{(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)}(Y_1; Y_2; Y_3)(h) \right\}^2 - \left\{ \dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)(h) \right\}^2 \right]. \end{aligned}$$

It follows from the continuity of $P_0 \left\{ \dot{l}_{(\gamma, Q_1, Q_2)}(Y_1; Y_2; Y_3)(h) \right\}^2$ with respect to (γ, Q_1, Q_2) uniformly over h and Lemma 3 on the Gilvenko-Cantelli classes that

$$P_n \left\{ \dot{l}_{(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)}(Y_1; Y_2; Y_3)(h) \right\}^2 \to P_0 \left\{ \dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3)(h) \right\}^2$$

uniformly on $h \in H_0 \times H_1 \times H_2$. It now follows from $\hat{\sigma}^{-1}(h) \to \sigma^{-1}(h)$ that

$$P_n \left\{ \dot{l}_{(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2)}(Y_1; Y_2; Y_3) \hat{\sigma}^{-1}(h_0, 0, 0) \right\}^2 \to P_0 \left\{ \dot{l}_{(\gamma_0, Q_{10}, Q_{20})}(Y_1; Y_2; Y_3) \sigma^{-1}(h_0, 0, 0) \right\}^2.$$

To show the consistency of the numeric differentation in estimating the asymptotic variance when the odds ratio model is correctly specified, note that, for $\gamma = \hat{\gamma}_n + h_0/\sqrt{n}$ and $\gamma_1 = \gamma_0 + h_0/\sqrt{n}$,

$$\begin{aligned} \sqrt{n} P_n \left\{ \dot{l}_\gamma(\gamma, Q_{1\gamma}, Q_{2\gamma}) - \dot{l}_\gamma(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2) \right\} &= \sqrt{n}(P_n - P_0) \left\{ \dot{l}_\gamma(\gamma, Q_{1\gamma}, Q_{2\gamma}) - \dot{l}_\gamma(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2) \right\} \\ &\quad + \sqrt{n} P_0 \left\{ \dot{l}_\gamma(\gamma, Q_{1\gamma}, Q_{2\gamma}) - \dot{l}_\gamma(\hat{\gamma}_n, \hat{Q}_1, \hat{Q}_2) - \dot{l}_\gamma(\gamma_1, Q_{10\gamma_1}, Q_{20\gamma_1}) + \dot{l}_\gamma(\gamma_0, Q_{10}, Q_{20}) \right\} \\ &\quad + \sqrt{n} P_0 \left\{ \dot{l}_\gamma(\gamma_1, Q_{10\gamma_1}, Q_{20\gamma_1}) - \dot{l}_\gamma(\gamma_0, Q_{10}, Q_{20}) \right\}, \end{aligned}$$

where $(Q_{10\gamma}, Q_{20\gamma})$ are the maximum likelihood estimators of (Q_1, Q_2) for fixed γ, and $\dot{l}_\gamma$ is the likelihood score for γ. $(Q_{10\gamma_1}, Q_{20\gamma_1})$ are the least favorite model defined by

$$\begin{aligned} dQ_{10\gamma_1}(y_1) &= u(1 + < \sigma_*^{-1}\{\sigma_1(h_0/\sqrt{n}, 0, 0), \sigma_2(h_0/\sqrt{n}, 0, 0)\}, e_1 >) dQ_{10}(y_1), \\ dQ_{20\gamma_1}(y_2) &= v(1 + < \sigma_*^{-1}\{\sigma_1(h_0/\sqrt{n}, 0, 0), \sigma_2(h_0/\sqrt{n}, 0, 0)\}, e_2 >) dQ_{20}(y_2), \end{aligned}$$

where $e_1 = (1, 0)$ and $e_2 = (0, 1)$, $< a, b >= \sum_k a_k b_k$ for vector $a = (a_k)$ and $b = (b_k)$, $\sigma_*(h_1, h_2) = \sigma(0, h_1, h_2)$, and u and v are taken such that $\int dQ_{10\gamma}(y_1) = 1$ and $\int dQ_{20\gamma}(y_2) = 1$ respectively.

The first two terms on the right-hand side of the equation are $o_P(1)$ from Lemma 3 and the last term converges to $\sigma(h_0, 0, 0)$ following similar arguments to those of Murphy and van der Vaart (2000).

References

Fuchino, S. and Plewik, S. (1999). On a theorem of E. Helly. *Proceedings of the American Mathematical Society*, **127**, 491-497.

Kantorovich, L. V. and Akilov, G. P. (1982). *Functional Analysis,* 2nd ed., Pergamon Press, Oxford.

Murphy, S. A. and Van der Vaart, A. W. (2000). "On profile likelihood (with discussion)," *Journal of American Statistical Association*, **95**, 449-465.

Van der Vaart, A. W. and Wellner, J. A. (1996) *Weak convergence and Empirical processes with application to statistics.* Springer.