A note on $L^\infty$ convergence of Neumann series approximation in missing data problems

Hua Yun Chen
Division of Epidemiology & Biostatistics
School of Public Health
University of Illinois at Chicago
1603 West Taylor Street, Chicago, IL 60612

Abstract

The inverse of the nonparametric information operator plays an important role in finding doubly robust estimators and the semiparametric efficient estimator in missing data problems. It is known that no closed-form expression for the inverse of the nonparametric information operator exists when missing data form nonmonotone patterns. Neumann series is usually applied to approximate the inverse. Similarly, semiparametric efficient scores in missing data problems are often expressed as solutions to integral equations, which often do not have closed-form solutions even for missing data with the simplest missing patterns. Neumann series also used in this situation to obtain approximations to the efficient scores. However, Neumann series approximation is only known to converge in $L^2$ norm, which is not sufficient for establishing statistical properties of the estimators yielded from the approximation. In this article, we show that $L^\infty$ convergence of the Neumann series approximations to the inverse of the nonparametric information operator and to the efficient scores in missing data problems can be obtained under very simple conditions. This paves the way to the study of the asymptotic properties of the doubly robust estimators and the locally semiparametric efficient estimator in those difficult situations.

Some Key Words: Auxiliary information, Induction, Rate of convergence, Weighted estimating equation.
1 Introduction

Let $Y$ be the full data and $R$ be the missing data indicator for $Y$. A component of $R$ takes value 1 if the corresponding component of $Y$ is observed, 0 otherwise. Denote the observed data by $(R, R(Y))$, where a component of $R(Y)$ equals the corresponding component of $Y$ if the corresponding component of $R$ takes value 1, and is missing if the corresponding component of $R$ takes value 0. Let the density of the distribution for $(R, Y)$ with respect to $\mu$, a product of count measures and Lebesgue measures, be $\pi(Y)f(Y)$. The nonparametric information operator for the missing data problem is defined as $m$, a map from $L^2_0(P_Y)$ to $L^2_0(P_Y)$ such that

$$m(b) = E[E\{b(Y)|R, R(Y)\}|Y]$$

for all $b(Y) \in L^2_0(P_Y)$, where $P_Y$ is the distribution of the full data $Y$ and $L^2_0(P_Y)$ is the collection of all mean-zero square-integrable functions under $P_Y$. See for example, Bickel et al. (1993, page 210), Robins, Rotnitzky, and Zhao (1994), van der Laan and Robins (2003, page 121), and Tsiatis (2006, page 258).

Robins et al. (1994) showed that, when data are missing at random (Rubin, 1976; Little and Robin, 2002) and the probability of observing all the configurations of the full data is greater than a positive constant, i.e.,

$$P(R = r|Y) = P\{R = r|r(Y)\} \text{ and } P(R = 1|Y) \geq \sigma > 0$$

for all missing data patterns $r$, the inverse of $m$ exists in $L^2_0(P_Y)$. Further, those authors showed that doubly robust estimating scores (Lipsitz, Ibrahim, and Zhao, 1999; Robins, Rotnitzky, and van der Laan, 2000) have the form $E[m^{-1}\{b(Y)\}|R, R(Y)]$, which is the optimal estimating scores among a set of augmented inverse probability weighted estimating scores. The semiparametric efficient score is also of this form with a specific $b$. The functional form of $m^{-1}$ is therefore very important in establishing the large sample properties of the doubly robust estimators and the semiparametric efficient estimator and in computing the estimators in practice.

The statistical properties of those estimators are often studied on a one-by-one basis...
and the computation of the estimators are straightforward if closed-form expressions for 
$m^{-1}$ and the efficient scores are found. However, it is known that $m^{-1}$ does not have a 
closed-form expression when missing data form nonmonotone missing patterns. For the 
semiparametric efficient score, closed-form expressions may not exist even for missing data 
with the simplest missing patterns. In those situations, we may still be able to establish 
the theoretical properties of $m^{-1}$ and the efficient score by other means, such the identity 
operator plus the compact operator argument. However, computing $m^{-1}$ can be a challeng-
ing problem. Neumann series has been proposed to approximate $m^{-1}$ for the computation 
of the estimators. The Neumann series expansion of $m^{-1}$ appears as

$$m^{-1} = \sum_{k=1}^{\infty} (I - m)^k,$$

where $I$ is the identity map. However, Neumann series is only known to converge under 
$L^2(P_Y)$ norm, which is insufficient for studying the asymptotic properties of the estimators 
based on the approximation. As far as we know, the $L^\infty$ convergence of the Neumann 
series approximation, which is very important for studying the asymptotic properties of the 
doubly robust estimators and the semiparametric efficient estimators obtained through the 
approximation, has not been established in general.

In this article, we show that the Neumann series approximations to $m^{-1}$ and to the 
efficient scores in many missing data problems converge in $L^\infty(P_Y)$ under a little stronger 
conditions than those of Robins et al. (1994). The conditions are very easy to check. The 
rate of convergence is also obtained in the proof. This paves the way to rigorously studying 
the statistical properties of the doubly robust estimators and the semiparametric efficient 
estimator in general.

2 $L^\infty$ convergence of Neumann series approximation to $m^{-1}$

We assume that conditions in (1), which are the same as those assumed in Robins et al. 
(1994), are true. In addition, we assume that $f(y)$ is bounded away from zero and infinite, 
which is often assumed in proving the doubly robust property of the optimal estimating
score $E[m^{-1}\{b(Y)\}]$. See for example, Robins, Rotnitzky, and van der Laan (2000) and van der Laan and Robins (2003).

In the following, we first prove the $L^\infty(P_Y)$-convergence property for the mapping $(I - m)^k$ when $k \to \infty$. Let $Y_1$ be a subset of variables in $Y$ and $P(\cdot|Y_1)$ be the condition probability of $Y$ given $Y_1$. Let $Q_1$ denote the collection of all the missing data patterns that include $Y_1$ in the observed variable list. Let $Q_2$ denote the remaining missing data patterns. Let

$$m_i(s) = \sum_{r \in Q_i} P(R = r|Y)E\{s|R = r, r(Y)\}$$

for $i = 1, 2$. Note that $m = m_1 + m_2$. From now on, we use $Y_1 \cap Y_2$ to denote variables in both $Y_1$ and $Y_2$.

Lemma 1. $||I - m_1||_{L^2(P(\cdot|Y_1))} \leq 1 - \sigma$, where

$$||s||_{L^2(P(\cdot|Y_1))} = \left\{ \int s^2 dP(Y|Y_1) \right\}^{1/2}.$$

Further

$$||m_2(s)||_{L^2(P(\cdot|Y_1))} \leq K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|Y_1 \cap r(Y))))},$$

where $K$ is a constant independent of $s$. Hence,

$$||(I - m)(s)||_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma)||s||_{L^2(P(\cdot|Y_1))} + K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|Y_1 \cap r(Y))))}.$$

Proof: Note that $m_1$ is a self-adjoint operator under the inner product defined in $L^2(P(\cdot|Y_1))$ for any fixed $Y_1$. By virtually the same argument as in proving that $||I - m||_{L^2(P)} \leq 1 - \sigma$ (Robins et al., 1994), it can be shown that $||I - m_1||_{L^2(P(\cdot|Y_1))} \leq 1 - \sigma$.

Since $P(R = r|Y) \leq 1 - P(R = 1|Y) \leq 1 - \sigma$ for any $r \in Q_2$ when $Y \neq Y_1$, it follows that

$$||m_2(s)||_{L^2(P(\cdot|Y_1))} = || \sum_{r \in Q_2} P\{R = r|r(Y)\}E\{s|R = r, r(Y)\}||_{L^2(P(\cdot|Y_1))} \leq \sum_{r \in Q_2} ||P\{R = r|r(Y)\}E\{s|R = r, r(Y)\}||_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma) \sum_{r \in Q_2} ||E\{s|R = r, r(Y)\}||_{L^2(P(\cdot|Y_1))}$$
\[
\leq (1 - \sigma) \sum_{r \in Q_2} K||s||_{L^2(P(\cdot|Y_1 \cap r(Y)))}^N.
\]

The last inequality is due to the uniform boundedness of the conditional densities derived from \(Y\).

**Lemma 2.** \(||(I - m)^N(s)||_{L^\infty(P)} \leq CN(c(Y))(1 - \sigma)^N||s||_{L^\infty(P)},\) where \(C\) is a constant independent of \(N\), and \(c(Y)\) is the cardinality of \(Y\), i.e., the number of variables in \(Y\).

**Proof:** We prove this lemma by induction based on Lemma 1. Let \(Y_1\) be the set containing a single variable. Let \(Q_1\) be the collection of missing patterns that have \(Y_1\) observed. By Lemma 1,

\[
||(I - m)^N(s)||_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma)^N||(I - m)^N-1s||_{L^2(P(\cdot|Y_1))} + K(1 - \sigma) \sum_{r \in Q_2}||(I - m)^N-1s||_{L^2(P(\cdot|Y_1 \cap r(Y)))}
\]

\[
\leq (1 - \sigma)^N||s||_{L^2(P(\cdot|Y_1))} + K \sum_{k=1}^{N}(1 - \sigma)^k \sum_{r \in Q_2}||(I - m)^N-ks||_{L^2(P(\cdot|Y_1 \cap r(Y)))}
\]

By the definition of \(Q_2\), it follows that \(Y_1 \cap r(Y)\) is empty for any \(r \in Q_2\), which implies that \(||s||_{L^2(P(\cdot|Y_1))} = ||s||_{L^2(P)}\). It follows that

\[
||(I - m)^N(s)||_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma)^N||s||_{L^2(P(\cdot|Y_1))} + K \sum_{k=1}^{N}(1 - \sigma)^N \sum_{r \in Q_2} ||s||_{L^2(P)}
\]

\[
\leq (1 - \sigma)^N||s||_{L^2(P(\cdot|Y_1))} + Kc(Q_2)N(1 - \sigma)^N||s||_{L^2(P)}
\]

\[
\leq K_1N(1 - \sigma)^N \max\{||s||_{L^2(P(\cdot|Y_1))}, ||s||_{L^2(P)}\},
\]

where \(c(Q)\) denote the cardinality of \(Q\) and \(K_1\) is another constant independent of \(N\). Since \(Y_1\) was chosen arbitrarily, the foregoing inequality holds for all \(Y_1\) having a single variable.

Next, suppose that the inequality holds for \(Y_1\) having any \(l\) variables. For \(Y_1\) having \(l + 1\) variables, apply Lemma 1 again to obtain

\[
||(I - m)^N(s)||_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma)^N||s||_{L^2(P(\cdot|Y_1))} + K \sum_{k=1}^{N}(1 - \sigma)^k \sum_{r \in Q_2}||(I - m)^N-ks||_{L^2(P(\cdot|Y_1 \cap r(Y)))}.
\]
By the definition of $Q_2$, $Y_1 \cap r(Y)$ has at most $l$ variables for all $r \in Q_2$. Hence, it follows from the induction that

$$
\|(I - m)^N(s)\|_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma)^N \|s\|_{L^2(P(\cdot|Y_1))} + K \sum_{k=1}^N (1 - \sigma)^N c(Q_2) K_1(N - k)^l \max_{c(Y_0) \leq l} \|s\|_{L^2(P(\cdot|Y_0))}
$$

$$
\leq K(l+1)^N(l+1) (1 - \sigma)^N \max_{c(Y_0) \leq (l+1)} \|s\|_{L^2(P(\cdot|Y_0))}.
$$

It now follows that

$$
\|(I - m)^N(s)\|_{L^\infty(P)} = \text{ess. sup } |(I - m)^N(s)|
$$

$$
\leq CN c(Y) (1 - \sigma)^N \text{ess. sup } \max_{Y_0 \subseteq Y} \|s\|_{L^2(P(\cdot|Y_0))}
$$

$$
\leq CN c(Y) (1 - \sigma)^N \|s\|_{L^\infty(P)},
$$

where ess. sup denotes the essential supremum.

**Theorem 1.** The $m^{-1}$ defined on $L^2(P_Y)$ by the Neumann series maps $L^\infty(P_Y)$ to $L^\infty(P_Y)$ and is continuous invertible in $L^\infty(P_Y)$.

**Proof:** Note that

$$
\|\sum_{k=0}^\infty (I - m)^k\|_{L^\infty(P_Y)} \leq \sum_{k=1}^N \|\sum_{k=0}^N (I - m)^k\|_{L^\infty(P_Y)}
$$

$$
\leq C \sum_{k=1}^\infty N c(Y) (1 - \sigma)^N < +\infty.
$$

This implies that $\|m^{-1}\|_{L^\infty(P_Y)} < +\infty$. It is easy to verify that $m^{-1}$ thus defined satisfies $m^{-1} m = m = I$ in $L^\infty(P_Y)$.

This theorem establishes the $L^\infty(P_Y)$ convergence of the Neumann series. In addition, it gives the rate of convergence of the series, which is a little slower than the geometric convergence rate of the series in $L^2(P_Y)$. The rate of convergence is important in establishing the large sample properties of the doubly robust estimators.
3 $L^\infty$ convergence of Neumann series approximation to efficient scores

The $L^\infty$ convergence of Neumann series approximation to $m^{-1}$ does not directly yield the $L^\infty$ convergence of Neumann series approximation to the efficient scores in missing data problems. However, the method of proof in the previous section can be applied to many missing data problems to obtain $L^\infty$ convergence of Neumann series approximation to the efficient score. This is more useful because the efficient score may not have a closed-form expression even when missing data form the simplest monotone missing pattern. We show in this section that $L^\infty$ convergence of Neumann series approximation to the efficient score holds for missing data in parametric regression, marginal regression, and Cox regression.

Before proceeding, we need to define what we meant by Neumann series approximation to the efficient scores. Neumann series approximation to the efficient score in Robins et al. (1994) and others appears as the successive approximation

$$U_N = S_{F, eff} + P(I - m)U_{N-1} \text{ and } U_0 = S_{F, eff},$$

where $S_{F, eff}$ is the efficient score under the full data model and $P$ is the projection to the closure of the nuisance score space. Note that $P(I - P) = 0$, it is not difficult to show that

$$U_N = (I - Pm)^NS,$$

where $(I - P)S = S_{F, eff}$. This reformulation allows us to express the efficient score directly as

$$\lim_{N \to \infty} E\{(I - Pm)^NS|R, R(Y)\}.$$ 

In this section, we call the sequence $T_N(s) = E\{(I - Pm)^NS|R, R(Y)\}$ the Neumann series approximation to the efficient score. Let $D^N(s) = (I - Pm)^Ns$. It follows from $D^N = (I - P) + P(I - m)D^{N-1}$ that

$$D^{N+1} - D^N = P(I - m)(D^N - D^{N-1}) = (P(I - m))^N (D^1 - D^0).$$

Let $G = P(I - m)$. It follows that, for any $N > 0$ and $p > 0$,

$$D^{N+p} - D^N = \sum_{k=1}^{p} (D^{N+k} - D^{N+k-1}) = \sum_{k=1}^{p} G^{N+k}(-Pm).$$
We will show that
\[ \|G^N\|_{L^\infty(P_Y)} = \left\| \{\mathcal{P}(I - m)\}^N \right\|_{L^\infty(P_Y)} \leq KN^c(1 - \sigma)^N, \tag{2} \]
where \( K \) is a constant independent of \( N \), and \( c \) is a constant depending only on the number of variables subject to missing values. This is sufficient to claim that \( T_N \) converges in \( L^\infty(P_Y) \) because it follows that
\[ \|D^{N+p} - D^N\|_{L^\infty(P_Y)} \leq K^* N^c(1 - \sigma)^N, \tag{3} \]
where \( K^* \) is another constant independent of either \( N \) or \( p \). Further, there exists a constant \( K^+ \), independent of \( N \), such that
\[ \|D^N(s)\|_{L^\infty(P_Y)} \leq K^+ \sigma^{-c} \|s\|_{L^\infty(P_Y)}. \]
Hence, \( \{D^N(s)\}_{N=1}^\infty \) is uniformly bounded and it is a Cauchy sequence in \( L^\infty(P_\eta) \) when \( s \) is uniformly bounded. By the control convergence theorem, \( T_N(s) = E\{D^N s|R, R(Y)\} \) converges in \( L^\infty(P_Y) \).

In the following, we show respectively that (2) is true for parametric regression, marginal regression, and Cox regression.

### 3.1 Missing data in parametric regression

Let \( Y = (V, W, X) \) with density \( p(v|w, x)g(w|x, \beta)q(x) \) where \( g(w|x, \beta) \) is the parametric regression model with a Euclidean parameter \( \beta \in \mathbb{R}^k \), which is of primary interest, and \( V \) is the auxiliary information observed in addition to outcome \( W \) and covariates \( X \). The nuisance parameters for the complete data model are \( (p, q) \). It is not difficult to see in this case that
\[ \mathcal{P}(s) = s(v, w, x) - E(s|w, x) + E(s|x). \]
where \( s \in L^2_0(P_Y) \).

Theorem 2. For any \( N \),
\[ \|G^N s\|_{L^\infty(P)} \leq KN^c(1 - \sigma)^N \|s\|_{L^\infty(P)}, \]
where $c(Y)$ denotes the cardinality of $Y = (V, W, X)$ and $K$ is a constant independent of $N$.

Proof: Consider first that $V$ is not involved, it follows that

$$G(s) = E\{(I - m)(s)|X\}.$$  

For any subset of $X$, denoted by $X_1$, let $G_1(s) = E\{(I - m_1)(s)|X\}$ and $G_2(s) = -E\{m_2(s)|X\}$, where $m_1$ is the sum over patterns that include $X_1$ as observed. It follows by virtually the same argument as that of Lemma 1, that

$$||G_1(s)||_{L^2\{P(\cdot|X_1)\}} \leq (1 - \sigma)||s||_{L^2\{P(\cdot)\}},$$

and that

$$||G_2(s)||_{L^2\{P(\cdot|X_1)\}} \leq K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2\{P(\cdot|X_1 \cap r(Y, X))\}},$$

where $Q_2$ is defined accordingly. It also follows by a similar argument as that of Lemma 2 that

$$||G^N(s)||_{L^\infty(P)} \leq C N^c(X) (1 - \sigma)^N ||s||_{L^\infty(P)}.$$

For the general case where $V$ is involved, the map $G$ becomes

$$G(s) = (I - m)(s) - E\{(I - m)(s)|X, W\} + E\{(I - m)(s)|X\}.$$ 

For any $X_1$ that is a subset of $X$, let $Q_1$ denote the collection of missing patterns that includes $X_1$ in the observed variables. Let $Q_2$ denote the remaining missing patterns. Let

$$m_i(s) = \sum_{r \in Q_i} P(R = r|V, W, X)E\{s|R = r, r(V, W, X)\}$$

for $i = 1, 2$. Let $G_1(s)$ be defined as $G(s)$ with $m$ replaced by $m_1$. Let $G_2(s) = G(s) - G_1(s)$.

$$||G_1(s)||^2_{L^2\{P(\cdot|X_1)\}} = \int [(I - m_1)(s) - E\{(I - m_1)(s)|X, W\} + E\{(I - m_1)(s)|X\}]^2 dP(\cdot|X_1)$$

$$= E\left(E[E^2\{(I - m_1)(s)|X, W\}|X] - E^2[E\{(I - m_1)(s)|X, W\}|X]|X_1\right)$$

$$+ E\left[\{(I - m_1)(s)\}^2|X_1\].$$
It follows that

\[ |G_1(s)|_{L^2(P(\cdot|X_1))} \leq (E[\{(I - m_1)(s)\}^2|X_1])^{1/2} \leq (1 - \sigma)||s||_{L^2(P(\cdot|X_1))}. \]

Further,

\[
|G_2(s)|_{L^2(P(\cdot|X_1))}^2 = \int [-m_2(s) + E\{m_2(s)|W, X\} - E\{m_2(s)|X\}]^2 dP(\cdot|X_1) \\
= E \left( E^2\{m_2(s)|W, X\}|X]\right) - E^2\{E\{m_2(s)|W, X\}|X\}|X_1) \\
+ E\{\{m_2(s)\}^2|X_1].
\]

Hence,

\[ |G_2(s)|_{L^2(P(\cdot|X_1))} \leq (E[\{m_2(s)\}^2|X_1])^{1/2} \leq K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|X_1 \cap r(V, W, X)))}. \]

By induction as used in the proof of Lemma 2, it follows that

\[ |G^N(s)|_{L^2(P(\cdot|X))} \leq C N^\alpha(X) (1 - \sigma)^N \max_{X_1 \subseteq X} ||s||_{L^2(P(\cdot|X_1))}. \]

Next, for the case \((W_1, X_1) \subseteq (W, X)\) with \(W_1\) including at least one variable, Define \(m_i\) and \(Q_i\), \(i = 1, 2\), similarly as before. Let

\[ G_1(s) = (I - m_1)(s) - E\{(I - m_1)(s)|W, X\}, \]

and

\[ G_2(s) = G(s) - G_1(s) = -m_2(s) + E\{m_2(s)|W, X\} + E\{(I - m)(s)|X\}. \]

It then follows by a similar argument as in the previous case that

\[
|G_1(s)|_{L^2(P(\cdot|W_1, X_1))} \leq (E[\{(I - m_1)(s)\}^2|W_1, X_1] - E[E^2\{(I - m_1)(s)|W, X\}|W_1, X_1])^{1/2} \\
\leq (1 - \sigma)||s||_{L^2(P(\cdot|W_1, X_1))}.
\]

Further,

\[
|G_2(s)|_{L^2(P(\cdot|W_1, X_1))} \leq \left( \int [-m_2(s) + E\{m_2(s)|W, X\}]^2 dP(\cdot|W_1, X_1) \right)^{1/2} \\
+ (E[E^2\{(I - m)(s)|X\}|W_1, X_1])^{1/2} \\
\leq K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|X_1 \cap r(V, W, X)))} \\
+ (1 - \sigma) \left( ||s||_{L^2(P(\cdot|X_1))} + K \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|X_1 \cap r(V, W, X)))} \right). \]
By induction as used in the proof of Lemma 2, it follows that

\[
|G^N(s)|_{L^2(P(\cdot|V,W,X))} \leq CN^{c(V,W,X)} (1 - \sigma)^N \max_{(W_1,X_1) \subset (W,X)} |s|_{L^2(P(\cdot|V,W_1,X_1))},
\]

where \(C\) is a constant independent of \(N\), and \(c(V,W,X)\) is the cardinality of \((W,X)\).

Finally, for the case \((V_1,W_1,X_1) \subset (V,W,X)\) with \(V_1\) including at least one variable, Define \(m_i\) and \(Q_i\), \(i = 1, 2\), similarly as before. Let

\[
G_1(s) = (I - m_1)(s)
\]

and

\[
G_2(s) = G(s) - G_1(s) = -m_2(s) - E\{(I - m)(s)|W,X\} + E\{(I - m)(s)|X\}.
\]

It follows that

\[
|G_1(s)|_{L^2(P(\cdot|V_1,W_1,X_1))} \leq (1 - \sigma)|s|_{L^2(P(\cdot|V_1,W_1,X_1))}.
\]

Further,

\[
|G_2(s)|_{L^2(P(\cdot|V_1,W_1,X_1))} \leq \left( \int [-m_2(s)]^2 dP(\cdot|V_1,W_1,X_1) \right)^{1/2}
+ (E[E^2\{(I - m)(s)|W,X\}|V_1,W_1,X_1])^{1/2}
+ (E[E^2\{(I - m)(s)|X\}|V_1,W_1,X_1])^{1/2}
\leq K(1 - \sigma) \sum_{r \in Q_2} |s|_{L^2(P(\cdot|(X_1,W_1) \cap r(V,W,X)))}
+ (1 - \sigma)(|s|_{L^2(P(\cdot|W_1,X_1))} + K \sum_{r \in Q_2} |s|_{L^2(P(\cdot|(X_1,W_1) \cap r(V,W,X)))})
+ (1 - \sigma)(|s|_{L^2(P(\cdot|X_1))} + K \sum_{r \in Q_2} |s|_{L^2(P(\cdot|(X_1) \cap r(V,W,X)))}).
\]

Now it follows by induction that

\[
|G^N(s)|_{L^2(P(\cdot|V,W,X))} \leq CN^{c(V,W,X)} (1 - \sigma)^N \max_{(V_1,W_1,X_1) \subset (V,Y,X)} |s|_{L^2(P(\cdot|V_1,W_1,X_1))},
\]

where \(C\) is a constant independent of \(N\), and \(c(V,W,X)\) is the cardinality of \((V,W,X)\).

This along with the boundedness of the density ratios immediately yields

\[
|G^N(s)|_{L^\infty(P)} \leq KN^{c(V,W,X)} (1 - \sigma)^N |s|_{L^\infty(P)},
\]

for some constant \(K\) independent of \(N\).
3.2 Missing data in marginal regression model

Let $W = (W_1, \cdots, W_K)^T$ and $E(W_k|X) = g_k(X_k\beta)$ for $k = 1, \cdots, K$. Let $g(\beta) = \{g_1(X_1\beta), \cdots, g_K(X_K\beta)\}^T$. Let $f(\epsilon)$ be the joint density of $W$ given $X$, where $\epsilon_i = w_i - g_i(x_i\beta)$ and $\epsilon = (\epsilon_1, \cdots, \epsilon_K)$. Let the density for $X$ be $q$ and the density for $V$ given $(W, X)$ be $p$, where $V$ denotes the auxiliary covariate. The nuisance parameter for the complete data model is $(q, f, p)$. For mean-zero square-integrable function $s$,

$$\mathcal{P}s = s(V, X, W) - E(s\epsilon|X)\{Var(\epsilon|X)\}^{-1}\epsilon.$$

The efficient score for $\beta$ under the full data is

$$S_{\beta}^{\text{eff}, F} = \left\{X^T_1 g'_1(X_1\beta), \cdots, X^T_K g'_K(X_K\beta)\right\}\{Var(W|X)\}^{-1}\epsilon.$$

Theorem 3. For any $N$,

$$||G^N s||_{L^\infty(P)} \leq KN^{c(Y)}(1 - \sigma)^N ||s||_{L^\infty(P)},$$

where $c(Y)$ denotes the cardinality of $Y = (V, W, X)$ and $K$ is a constant independent of $N$.

Proof: Consider first $Y_1 = X_1$ that is a subset of $X$. Let $G_1(s)$ be defined as $G(s)$ with $m$ replaced by $m_1$. Let $G_2(s) = G(s) - G_1(s)$.

$$||G_1(s)||^2_{L^2(P(\cdot|X_1))} = E\left(E\left[\{(I - m_1)(s)\}^2|X\right] - E\left[\{(I - m_1)s\}\epsilon|X\right]\{Var(\epsilon|X)\}^{-1}
\times E\left[\epsilon(I - m_1)s|X\right]|X_1\right)
\leq E[\{(I - m_1)(s)\}^2|X_1] \leq (1 - \sigma)^2 ||s||^2_{L^2(P(\cdot|X_1))}.$$

Further,

$$||G_2(s)||^2_{L^2(P(\cdot|X_1))} = E\left(E[\{m_2(s)\}^2|X]\right) - E[\{m_2(s)\}\epsilon|X]\{Var(\epsilon|X)\}^{-1}
\times E[m_2(s)|X]|X_1\right)
\leq E[\{m_2(s)\}^2|X_1] \leq K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|X_1 \cap r(Y)))}.$$

12
By induction as used in the proof of Lemma 2, it follows that

$$||G^N(s)||_{L^2(P(\cdot|X))} \leq CN^c(X)(1-\sigma)^N \max_{X_1 \subset X} ||s||_{L^2(P(\cdot|X_1))}.$$ 

Next, for the case $Y_1 \cap (V,W)$ is not empty, let

$$G_1(s) = (I - m_1)(s)$$

and

$$G_2(s) = G(s) - G_1(s) = -m_2(s) - E[(I - m)s \epsilon|X] \text{Var}(\epsilon|X)^{-1}\epsilon.$$

It follows that

$$||G_1(s)||_{L^2(P(\cdot|Y_1))} \leq (1 - \sigma) ||s||_{L^2(P(\cdot|Y_1))}.$$

Further,

$$||G_2(s)||_{L^2(P(\cdot|Y_1))} \leq ||m_2(s)||_{L^2(P(\cdot|Y_1))} + ||E[(I - m)s \epsilon|X] \text{Var}(\epsilon|X)^{-1}\epsilon||_{L^2(P(\cdot|Y_1))}$$

$$\leq ||m_2(s)||_{L^2(P(\cdot|Y_1))} + ||E[(I - m)s]^2|X||_{L^2(P(\cdot|Y_1))}$$

$$\leq K(1 - \sigma) \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|Y_1 \cap r(Y)))}$$

$$+ (1 - \sigma)(||s||_{L^2(P(\cdot|X \cap Y_1))} + K \sum_{r \in Q_2} ||s||_{L^2(P(\cdot|X \cap r(Y)))}).$$

Now it follows by induction that

$$||G^N(s)||_{L^2(P(\cdot|Y))} \leq CN^c(Y)(1-\sigma)^N \max_{Y_1 \subset Y} ||s||_{L^2(P(\cdot|Y_1))},$$

where $C$ is a constant independent of $N$, and $c(Y)$ is the cardinality of $Y$. It follows directly that

$$||G^N(s)||_{L^\infty(P)} \leq CN^c(Y)(1-\sigma)^N ||s||_{L^\infty(P)}.$$

### 3.3 Missing covariates in Cox regression

Suppose that $T$ is the survival time of a subject, which is subject to censoring by censoring time $C$. Given covariate $Z$ (time-independent), $T$ and $C$ are independent. Only $X = T \wedge C = \min(T, C)$ and $\delta = 1_{\{T \leq C\}}$ are observed for survival time and censoring time. $Z$
is subject to missing values. Assume that, given \((T, C, Z)\), the missing data mechanism depends on the observed data \(R(Y) = (X, \delta, R, R(Z))\) only. The density for \((X, \delta, Z)\) is

\[
f(x, \delta | z, \beta, \lambda)p(z) = \lambda^\delta(x)\phi^\delta(\beta z) \exp\{-\Lambda(x)\phi(\beta z)\}g_c^{1-\delta}(x | z)\bar{G}_c(x | z)p(z),
\]

where \(\Lambda(x) = \int_0^x \lambda(t)dt\) and \(g_c\) is the density function of censoring time distribution \(G_c\) and \(\bar{G}_c = 1 - G_c\). Let \(\Lambda_c(x) = \int_0^x \lambda_c(t)dt\), \(\lambda_c = g_c/G_c\), \(dM_T(t) = dN_T(t) - U(t)\lambda_T(t, z)dt\) with \(N_T(t) = 1_{\{X \geq t, \delta = 1\}}\), and \(dM_C(t) = dN_C(t) - U(t)\lambda_C(t, z)dt\) with \(N_C(t) = 1_{\{X \geq t, \delta = 0\}}\), where \(U(t) = 1_{\{X \geq t\}}\). Let \(s\) be mean-zero square-integrable function.

\[
P(s) = \int \left[s(t, 0, Z) - \frac{E\{s(X, \delta, Z)U(t) | Z\}}{E(U(t) | Z)}\right]dM_C(t, Z) + E\{s(X, \delta, Z) | Z\}.
\]

The efficient score for estimating \((\beta, \Lambda)\) can be expressed as

\[
\lim_{N \to \infty} E \left\{\left(I - Pm\right)^Na(X, \delta, Z) | X, \delta, R, R(Z)\right\},
\]

where

\[
a(X, \delta, Z) = \int h_{11}^T \frac{\partial}{\partial \beta} \log \phi(\beta Z)dM_T(t) + \int h_{12}(t) dM_T(t).
\]

Theorem 4. For any \(N\),

\[
||G^N_s||_{L^\infty(P)} \leq K N^{c(Y)} (1 - \sigma)^N ||s||_{L^\infty(P)},
\]

where \(c(Y)\) denotes the cardinality of \(Y = (X, \delta, Z)\) and \(K\) is a constant independent of \(N\).

Proof: Note that

\[
E(s^2 | Z) = \int \left[s(u, 1, Z) - \frac{E\{s(X, \delta, Z)U(u) | Z\}}{E(U(u) | Z)}\right]^2 E\{U(u) | Z\}\lambda_T(u, Z)du
\]

\[
+ \int \left[s(u, 0, Z) - \frac{E\{s(X, \delta, Z)U(u) | Z\}}{E(U(u) | Z)}\right]^2 E\{U(u) | Z\}\lambda_C(u, Z)du
\]

\[
+ E^2\{s(X, \delta, Z) | Z\}.
\]

Hence, \(E[|Ps|^2 | Z] \leq E(s^2 | Z)\), which implies \(||P||_{L^2(P(|Z_1))} \leq 1\) for any \(Z_1 \subset Z\). Since only \(Z\) is assumed to be subject to missing values, it follows that, when \(Y_1 \subset Z\),

\[
||P(I - m_1)||_{L^2(P(|Y_1))} \leq ||I - m_1||_{L^2(P(|Y_1))}
\]

14
and
\[ \|Pm_2\|_{L^2(P(\cdot|Y_1))} \leq \|m_2\|_{L^2(P(\cdot|Y_1))}. \]

Hence, by induction and the convergent rate for \(I - m\) proved in Lemma 2, it follows that
\[ \|G^N(s)\|_{L^2(P(\cdot|Y))} \leq CN^{C(Y)}(1 - \sigma)^N \max_{Y_1 \subset Y} \|s\|_{L^2(P(\cdot|Y_1))}, \]
when \(Y \subset Z\). When \(Y\) includes variables other than \(Z\), let
\[ G_1(s) = \int \{(I - m_1)s\}(u, 0, z)E\{U(u)|Z\}dN_C(u). \]

It follows that
\[ \|G_1(s)\|_{L^2(P(\cdot|Y_1))} = \|(1 - \delta)E\{U(u)|Z\}_{u=X}(I - m_1)(s)\|_{L^2(P(\cdot|Y_1))} \leq \|(I - m_1)(s)\|_{L^2(P(\cdot|Y_1))}. \]

Let \(G_2(s) = G(s) - G_1(s)\). It follows that
\[ G_2(s) = -A_2(A_2A_2)^{-1}A_2^*m_2(s) + \int \{(I - m_1)(s)\}(u, 0, z)E\{U(u)|Z\}U(u)\lambda_C(u, z)du \]
\[ - \int E\{(I - m_1)(s)(X, \delta, Z)U(u)|Z\}dM_C(u, Z) + E\{(I - m_1)(s)(x, \delta, z)|z\}. \]

Hence,
\[ \|G_2(s)\|_{L^2(P(\cdot|Y_1))} \leq K(1 - \sigma)\|s\|_{L^2(P(\cdot|Y_1))} + \|m_2(s)\|_{L^2(P(\cdot|Y_1))}. \]

It now can be shown by induction that
\[ \|G^N(s)\|_{L^2(P(\cdot|Y))} \leq KN^{C(Y)}(1 - \sigma)^N \max_{Y_1 \subset Y} \|s\|_{L^2(P(\cdot|Y_1))}. \]

It follows from this that
\[ \|G^N(s)\|_{L^\infty(P)} \leq KN^{C(Y)}(1 - \sigma)^N \|s\|_{L^\infty(P)}. \]

4 Discussion

We have proved the \(L^\infty\) convergence of the Neumann series approximation to the inverse of the nonparametric information operator, which is critical in constructing doubly robust estimating equations. Neumann series approximation to the semiparametric efficient scores
in missing data problems is also shown to converge in $L^\infty$ for mostly frequently used models in practice. The rate of convergence, which largely determines the asymptotic properties of the estimators, is also obtained in the proofs.

We proved the result under the missing at random assumption. This assumption is not critical as can be seen from the proof. The result can be straightforwardly extended to cover nonignorable missing data problem.
References


