4. Fractional Imputation (Part 3)

1. Semiparametric fractional imputation for missing covariates problem

- Suppose that we are interested in estimating $\theta$ in $f(y \mid x; \theta)$.

- Now, $y$ is always observed and $x$ is subject to missingness.

- Assume MAR in the sense that $Pr(\delta = 1 \mid x, y)$ does not depend on $x$.

- In this case, the mean score equation for $\theta$ is

  $$
  \sum_{i=1}^{n} \left[ \delta_i S(\theta; x_i, y_i) + (1 - \delta_i) E\{S(\theta; X, y_i) \mid y_i\} \right]
  $$

  where

  $$
  E\{S(\theta; X, y_i) \mid y_i\} = \frac{\int S(\theta; x, y_i) f(y_i \mid x; \theta) g(x) dx}{\int f(y_i \mid x; \theta) g(x) dx}
  $$

  where $g(\cdot)$ is the density for the marginal distribution of $x$.

- If $g(x) = g(x; \alpha)$ for some $\alpha$, the parametric FI for this case can be used to compute

  $$
  E\{S(\theta; X, y_i) \mid y_i\} \approx \sum_{j=1}^{m} w_{ij}^*(\theta, \alpha) S(\theta; x_{ij}^{*(j)}, y_i)
  $$

  where

  $$
  w_{ij}^*(\theta, \alpha) \propto \frac{f(y_i \mid x_{ij}^{*(j)}; \theta) g(x_{ij}^{*(j)}; \alpha)}{h(x_{ij}^{*(j)})}
  $$

  and $h(\cdot)$ is the proposal density for $x$.

- Parameter $\alpha$ is a nuisance parameter in the sense that we are not interested in estimating $\alpha$ but its estimation is needed to estimate the parameter of interest.
• Semiparametric approach: use a nonparametric model for $g(\cdot)$ but still uses a parametric model for $f(\cdot)$.

• Without loss of generality, assume that the first $r$ units are responding in both $x$ and $y$ and the remaining $n - r$ units are missing in $x$.

• A nonparametric model for $g(x)$ is that it belongs to the class of $G = \{g(x) = \sum_{i=1}^{r} \alpha_i I(x = x_i); \sum_{i=1}^{r} \alpha_i = 1, \alpha_i \geq 0\}$. Note that the dimension of parameter $\alpha$ is $r$, which can go to infinity asymptotically.

• EM algorithm using semiparametric fractional imputation

Step 0. For each unit with $\delta_i = 0$, $r$ imputed values of $x$ are assigned with $x_i^{(j)} = x_j$. Let $\alpha^{(0)}_k = 1/r$.

Step 1. At the $t$-th EM iteration, compute the fractional weight

$$w_{ij(t)}^* = \frac{f(y_i \mid x_i^{(j)}; \theta^{(t)})\alpha_j^{(t)}}{\sum_{k=1}^{r} f(y_i \mid x_i^{(k)}; \theta^{(t)})\alpha_k^{(t)}}$$

where $\theta^{(0)}$ is the MLE of $\theta$ using only full respondents.

Step 2. Using $w_{ij(t)}^*$ and $(x_i^{(j)}, y_i)$, update the parameters by solving the imputed score equation:

$$\sum_{i=1}^{r} S(\theta; x_i, y_i) + \sum_{i=r+1}^{n} \sum_{j=1}^{r} w_{ij(t)}^* S(\theta; x_i^{(j)}, y_i) = 0 \quad (1)$$

and

$$\hat{\alpha}_{(t+1)}^j = \frac{1}{n} \left\{ 1 + \sum_{i=r+1}^{n} w_{ij(t)}^* \right\} . \quad (2)$$

Step 2 uses (1) and (2) to update the parameters, which are consistent equations obtained by differentiating the log likelihood of $\theta$ and nuisance parameters $\alpha = (\alpha_1, \ldots, \alpha_r)^T$.

• Yang and Kim (2014) developed $\sqrt{n}$-consistency and the asymptotic normality of the SFI estimator of $\theta$ using von Misses calculus and V-statistics theory.
2 Nonparametric fractional imputation

- Bivariate data: \((x_i, y_i)\)
- \(x_i\) are completely observed but \(y_i\) is subject to missingness.
- Joint distribution of \((x, y)\) completely unspecified.
- Assume MAR in the sense that \(P(\delta = 1 \mid x, y)\) does not depend on \(y\).
- Without loss of generality, assume that \(\delta_i = 1\) for \(i = 1, \cdots, r\) and \(\delta_i = 0\) for \(i = r + 1, \cdots, n\).
- We are only interested in estimating \(\theta = E(Y)\).
- Let \(K_h(x_i, x_j) = K((x_i - x_j)/h)\) be the Kernel function with bandwidth \(h\) such that \(K(x) \geq 0\) and
  \[
  \int K(x)dx = 1, \quad \int xK(x)dx = 0, \quad \sigma_K^2 \equiv \int x^2K(x)dx > 0.
  \]
Examples include the following:
- Boxcar kernel: \(K(x) = \frac{1}{2}I(x)\)
- Gaussian kernel: \(K(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)\)
- Epanechnikov kernel: \(K(x) = \frac{3}{4}(1 - x^2)I(x)\)
- Tricube Kernel: \(K(x) = \frac{70}{81}(1 - |x|^3)^3I(x)\)

where
\[
I(x) = \begin{cases} 
1 & \text{if } |x| \leq 1 \\
0 & \text{if } |x| > 1.
\end{cases}
\]
- Nonparametric regression estimator of \(m(x) = E(Y \mid x)\):
  \[
  \hat{m}(x) = \sum_{i=1}^{r} l_i(x)y_i \quad \text{(3)}
  \]
where
\[
  l_i(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_j K\left(\frac{x-x_j}{h}\right)}.
\]
Estimator in (3) is often called Nadaraya-Watson kernel estimator.
• Under some regularity conditions and under the optimal choice of $h$ (with $h^* = O(n^{-1/5})$), it can be shown that

$$E \left[ \left\{ \hat{m}(x) - m(x) \right\}^2 \right] = O(n^{-4/5}).$$

Thus, its convergence rate is slower than that of parametric one.

• However, the imputed estimator of $\theta$ using (3) can achieve the $\sqrt{n}$-consistency. That is,

$$\hat{\theta}_{NP} = \frac{1}{n} \left\{ \sum_{i=1}^{r} y_i + \sum_{i=r+1}^{n} \hat{m}(x_i) \right\}$$

achieves

$$\sqrt{n} \left( \hat{\theta}_{NP} - \theta \right) \longrightarrow N(0, \sigma^2)$$

where $\sigma^2 = E\{v(x)/\pi(x)\} + V\{m(x)\}$, $m(x) = E(y \mid x)$, $v(x) = V(y \mid x)$ and $\pi(x) = E(\delta \mid x)$. A sketched proof for (5) is given in Appendix A.

• We can express $\hat{\theta}_{NP}$ as a nonparametric fractional imputation (NFI) estimator of the form

$$\hat{\theta}_{NFI} = \frac{1}{n} \left\{ \sum_{i=1}^{r} y_i + \sum_{j=r+1}^{n} \sum_{i=1}^{r} w_{ij}^* y_i^{(j)} \right\}$$

where $w_{ij}^* = l_i(x_j)$, which is defined after (3), and $y_i^{(j)} = y_i$. One may consider a sampling of size $m$ from the set of respondents using the fractional weights to reduce the imputation size. Further research is needed.

• Reference


3 Nearest neighbor imputation

- Consider the same setup for nonparametric fractional imputation.

- Note that a nonparametric estimator \( \hat{m}(x) \) in (3) corresponding to boxcar kernel is the local sample average that takes the value of \( y_i \) in the neighbors within range \( h \) in terms of \( x \)'s. In the extreme case, we can make the choice of \( h \) varying with \( x \) so that it can include only the nearest neighbor.

- That is, for the missing \( y_j \), we use the observed covariate \( x_j \) to identify the nearest neighbor of \( y_j \).

- Let \( j(1) \) be the index of the nearest neighbor of \( j \) such that

\[
d \left( x_{j(1)}, x_j \right) \leq d \left( x_k, x_j \right), \quad k = 1, \ldots, r
\]

where \( d \left( x_i, x_j \right) \) is the distance function between \( x_i \) and \( x_j \).

- Similarly, the second nearest neighbor of \( y_j \), indexed by \( j(2) \), satisfies

\[
d \left( x_{j(1)}, x_j \right) \leq d \left( x_{j(2)}, x_j \right) \leq d \left( x_k, x_j \right), \quad k \in \{1, \ldots, r\} - \{j(1)\}.
\]

- Let \( y_{j(1)}^* \) and \( y_{j(2)}^* \) be the first nearest neighbor and the second nearest neighbor of \( y_j \), respectively. Then, under some regularity conditions, we can show that

\[
\max_j \left| E \left( y_j \mid x_j \right) - E \left( y_{j(1)}^* \mid x_{j(1)}^* \right) \right| = O_p \left( n^{-1+\alpha} \right) \quad (6)
\]

and

\[
\max_j \left| E \left( y_j \mid x_j \right) - E \left( y_{j(2)}^* \mid x_{j(2)}^* \right) \right| = O_p \left( n^{-1+\alpha} \right) \quad (7)
\]

for any \( \alpha > 0 \). A sketched proof for (6) and (7) is given in Appendix B.

- In fact, we can obtain the same result for \( m \)-nearest neighbors and then use

\[
\hat{\theta}_{NNFI} = \frac{1}{n} \left\{ \sum_{i=1}^n y_i + \sum_{i=r+1}^n \sum_{j=1}^m w_{ij}^* y_i^{(j)} \right\} \quad (8)
\]

where \( y_i^{(j)} = y_{i(j)} \) is the \( j \)-th nearest neighbor of \( i \) and \( w_{ij}^* \) is the fractional weight assigned to \( y_{i(j)}^* \). For \( d(x_i, x_j) = (x_i - x_j)^2 \), we may use a Gaussian kernel to get \( w_{ij}^* \propto \exp \{-d(x_i, x_{i(j)})\} \) with \( \sum_{j=1}^m w_{ij}^* = 1 \).
• Furthermore, we may consider a nearest neighbor based on predictive mean matching:

1. Assume $E(y \mid x) = m(x; \beta)$ for some $\beta$. Fit a working regression model to get $\hat{y}_i = m(x_i; \hat{\beta})$ for all $i = 1, 2 \cdots, n$.

2. Identify the nearest neighbor using $\hat{y}_i$. That is, find $j(1)$ that satisfies

$$d(\hat{y}_{j(1)}, \hat{y}_j) \leq d(\hat{y}_k, \hat{y}_j), \quad k = 1, \cdots, r.$$ 

Similarly, we can identify $m$ nearest neighbors in terms of $\hat{y}_i$.

3. Apply the $m$ nearest neighbors in Step (2) to obtain the nearest neighbor imputation estimator of the form (8).

• Note that we can express $d(\hat{y}_k, \hat{y}_j) = d_{\hat{\beta}}(x_k, x_j)$ for some distance function $d_{\beta}(a, b)$ that also depends on $\beta = \hat{\beta}$. In this case, proving the consistency of the resulting estimator is more complicated. Further research is needed.

• Removes the curse of dimensionality in identifying the nearest neighbor. Very popular in practice but its theory is under-developed.

• Also, we use real value of $y_i$ in the imputation. Such imputation is called hot deck imputation. (Will be covered in Week 5.)

• References


4 Application to measurement error models

- Bivariate data \((x, y)\)
- We are interested in estimating \(\theta\) in \(f(y \mid x; \theta)\) with \(\theta \in \Omega\).
- Instead of observing \((x, y)\), we observe \((w, y)\) where \(w = x + u\) and \(u\) is the measurement error.
- Suppose that we have a separate sample, called calibration sample or validation sample \(V\), such that we observe \(x\) and \(w\) in \(V\).
- In \(V\), we obtain a nonparametric estimator of \(g(x \mid w)\) using a Kernel regression method. Specification of \(g(x \mid w)\) is called Berkson error model in the measurement error literature.
- We also assume that
  \[
  f(y \mid x, w) = f(y \mid x). \tag{9}
  \]
  If condition (9) is satisfies, then the measurement error, \(u = w - x\), is called non-differential and the variable \(w\) is said to be a surrogate for \(x\).
- In the original sample, the imputed values of \(x\) are obtained from
  \[
  f(x \mid w, y) \propto g(x \mid w)f(y \mid x, w)
  \]
  which reduces to, under assumption (9),
  \[
  f(x \mid w, y) \propto g(x \mid w)f(y \mid x). \tag{10}
  \]
  The first component is computed from the validation sampling using a nonparametric regression. The second part is computed using a parametric model assumption, \(f(y \mid x) = f(y \mid x; \theta)\) for some \(\theta\). Thus, the overall estimation problem becomes a semiparametric estimation problem.
- Specifically, the following semiparametric fractional imputation can be used:
1. From the validation sample, use a nonparametric regression technique to obtain \( \hat{E}(x \mid w) = \sum_{i \in V} l_i(w) x_i \), where
\[
l_i(w) = \frac{K_h(w_i, w)}{\sum_{j \in V} K_h(w_j, w)}.
\]

2. For each \( i \) in the original sample, we use \( m = n_V \) imputed values of \( x_i \) by taking all the element of \( x_j \) in \( V \). That is, we use \( x_{i}^{(j)} = x_j \).

3. Compute the fractional weight associated with \( x_{i}^{(j)} \) by
\[
w_{ij(t)}^{*} = \frac{f(y_i \mid x_{i}^{(j)}; \hat{\theta}(t)) K_h(w_i, w_j)}{\sum_{k=1}^{m} f(y_i \mid x_{i}^{(k)}; \hat{\theta}(t)) K_h(w_i, w_k)}.
\]

4. Update the parameter value by maximizing
\[
l(\theta) = \sum_{i} \sum_{j} w_{ij(t)}^{*} \log f(y_i \mid x_{i}^{(j)})
\]
with respect to \( \theta \in \Omega \) to get \( \hat{\theta}(t+1) \).

5. Goto Step 3 until convergence.

- The resulting SMLE (semiparametric maximum likelihood estimator) of \( \theta \) is the solution to
\[
E\{S(\theta; X, y) \mid y, w\} = 0
\]
which is actually the solution to
\[
E\{S(\theta; X, y) \mid y, w; \theta, g\} = 0
\]
where \( g \) is an infinite dimensional nuisance parameter.

- The \( \sqrt{n} \)-consistency of SMLE of \( \theta \) can be established. (Further research is needed.)

- Reference:
Appendix A: Proof for (5)

Regularity conditions:

(C1). Moments’ conditions: \( E(|m(x)|^\alpha) < \infty, E(|y|^\alpha) < \infty \) for some \( \alpha > 2 \).

(C2). Bounded conditions: \( 1 > \pi(x) > d > 0 \) almost surely.

(C3). Smoothness conditions: \( f(x) \) and \( \pi(x) \) have bounded partial derivatives with respect to \( x \) up to an order \( q \) with \( q \geq 2 \), \( 2q > d_x \) almost surely, where \( d_x \) is the dimension of \( x \).

(C4). Kernel function conditions:

1. It is bounded and has compact support.
2. \( \int K(s_1, \ldots, s_{d_x})ds_1 \ldots ds_{d_x} = 1 \),
3. \( \int s_i^lK(s_1, \ldots, s_{d_x})ds_1 \ldots ds_{d_x} = 0 \) for any \( i = 1, \ldots, d_x \) and \( 1 \leq l < q \).
4. \( \int s_i^qK(s_1, \ldots, s_{d_x})ds_1 \ldots ds_{d_x} \neq 0 \).

(C5). Bandwidth conditions: \( nh^{2d_x} \to \infty, \sqrt{n}h^q \to 0 \), as \( n \to \infty \).

Proof

For simplicity, we only consider the case where \( d_x = 1 \). By using standard argument in the kernel smoothing method, it can be shown that

\[
E \left\{ \frac{1}{nh} \sum_{j=1}^{n} \delta_j K\left( \frac{x-x_j}{h} \right) y_j \right\} = \pi(x)f(x)m(x) + O(h^2) 
\] (11)

and

\[
E \left\{ \frac{1}{nh} \sum_{j=1}^{n} \delta_j K\left( \frac{x-x_j}{h} \right) \right\} = \pi(x)f(x) + O(h^2). 
\] (12)

According to (11), (12) and by using Taylor linearization, we have

\[
\hat{m}(x) = \frac{(nh)^{-1} \sum_{j=1}^{n} \delta_j K((x-x_j)/h)y_j}{(nh)^{-1} \sum_{j=1}^{n} \delta_j K((x-x_j)/h)} 
\]

\[
= m(x) + \frac{1}{\pi(x)f(x)} \left\{ \frac{1}{nh} \sum_{j=1}^{n} \delta_j K\left( \frac{x-x_j}{h} \right) y_j - \pi(x)f(x)m(x) \right\} 
\]

\[
- \frac{m(x)}{\pi(x)f(x)} \left\{ \frac{1}{nh} \sum_{j=1}^{n} \delta_j K\left( \frac{x-x_j}{h} \right) - \pi(x)f(x) \right\} + O(h^2) 
\] (13)
Therefore, we have

\[
\hat{\theta}_{NP} = \frac{1}{n} \sum_{i=1}^{n} \{ \delta_{i}y_{i} + (1 - \delta_{i})m(x_{i}) \} \\
= \frac{1}{n} \sum_{i=1}^{n} \{ \delta_{i}y_{i} + (1 - \delta_{i})m(x_{i}) \} \\
+ \frac{1}{n^{2}} \sum_{i \neq j}^{n} \frac{(1 - \delta_{i})\delta_{j}h^{-1}K((x_{j} - x_{i})/h) \{ y_{j} - m(x_{i}) \}}{\pi(x_{i})f(x_{i})} + O(h^{2}) \\
\approx \frac{1}{n} \sum_{i=1}^{n} \{ \delta_{i}y_{i} + (1 - \delta_{i})m(x_{i}) \} \\
+ \frac{2}{n(n - 1)} \sum_{i < j}^{n} \frac{1}{2} \frac{(1 - \delta_{i})\delta_{j}h^{-1}K((x_{j} - x_{i})/h) \{ y_{j} - m(x_{i}) \}}{\pi(x_{i})f(x_{i})} \\
+ \frac{(1 - \delta_{j})\delta_{i}h^{-1}K((x_{i} - x_{j})/h) \{ y_{i} - m(x_{j}) \}}{\pi(x_{j})f(x_{j})} + O(h^{2}).
\] (14)

Define

\[
\zeta_{ij} = \frac{(1 - \delta_{i})\delta_{j}h^{-1}K((x_{j} - x_{i})/h) \{ y_{j} - m(x_{i}) \}}{\pi(x_{i})f(x_{i})}, \\
\zeta_{ji} = \frac{(1 - \delta_{j})\delta_{i}h^{-1}K((x_{i} - x_{j})/h) \{ y_{i} - m(x_{j}) \}}{\pi(x_{j})f(x_{j})}
\]

and \(h(z_{i}, z_{j}) = (\zeta_{ij} + \zeta_{ji})/2\) with \(z_{i} = (x_{i}, y_{i}, \delta_{i})\), then \(2 \sum_{i \neq j} h(z_{i}, z_{j})/ \{ n(n - 1) \}\) is the U-statistic. According to U-statistic theory (e.g. Serfling, 1980, Ch. 5), we have

\[
\frac{2}{n(n - 1)} \sum_{i < j} h(z_{i}, z_{j}) = \frac{2}{n} \sum_{i=1}^{n} E \{ h(z_{i}, z_{j}) | z_{i} \} + o_{p}(n^{-1/2}).
\] (15)

Let \(s = (x_{j} - x_{i})/h\), by \(nh^{2} \to \infty, nh^{4} \to 0\) and according to Taylor linearization, we have

\[
E(\zeta_{ij} | z_{i}) = E \left[ \frac{(1 - \delta_{i})\delta_{j}h^{-1}K((x_{j} - x_{i})/h) \{ y_{j} - m(x_{i}) \}}{\pi(x_{i})f(x_{i})} | z_{i} \right] \\
= \frac{1 - \delta_{i}}{\pi(x_{i})f(x_{i})} E \left[ \pi(x_{j}) \frac{1}{h} K \left( \frac{x_{j} - x_{i}}{h} \right) \{ m(x_{j}) - m(x_{i}) \} | z_{i} \right] \\
= \frac{1 - \delta_{i}}{\pi(x_{i})f(x_{i})} \int_{-\infty}^{\infty} \pi(x_{i} + hs) K(s) \{ m(x_{i} + hs) - m(x_{i}) \} f(x_{i} + hs) ds \\
= O(h^{2})
\] (16)
and

\[ E(\zeta_{ji}|z_i) = E\left[ \frac{(1-\delta_j)\delta_i h^{-1}K((x_i - x_j)/h) \{y_i - m(x_j)\}}{\pi(x_j)f(x_j)} \bigg| z_i \right] \]

\[ = \delta_i E\left[ \frac{(1 - \pi_j)h^{-1}K((x_i - x_j)/h) \{y_i - m(x_j)\}}{\pi(x_j)f(x_j)} \bigg| z_i \right] \]

\[ = \delta_i \int \frac{1 - \pi(x_i + hs)}{f(x_i + hs)\pi(x_i + hs)} K(s) \{y_i - m(x_i)\} f(x_i) ds + O(h^2) \]

\[ = \delta_i \frac{1 - \pi(x_i)}{\pi(x_i)} \{y_i - m(x_i)\} + O(h^2). \]  \quad (17)

According to (14)-(17), we have

\[ \hat{\theta}_{NP} = \frac{1}{n} \sum_{i=1}^{n} \{\delta_i y_i + (1 - \delta_i)m(x_i)\} \]

\[ + \frac{2}{n(n-1)} \sum_{i<j} h(z_i, z_j) + O(n^{-1/2}) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \{\delta_i y_i + (1 - \delta_i)m(x_i)\} \]

\[ + \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{1 - \pi(x_i)}{\pi(x_i)} \{y_i - m(x_i)\} + o(n^{-1/2}) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\delta_i}{\pi(x_i)} y_i - \left\{ \frac{\delta_i}{\pi(x_i)} - 1 \right\} m(x_i) \right] + o_p(n^{-1/2}). \]

Hence, we have (5) with \( \sigma^2 = E \{\sigma^2(x)/\pi(x)\} + \text{var} \{m(x)\} \), where \( \sigma^2(x) = \text{var}(y|x) \).

**Appendix B: Proof for (6) and (7)**

We assume the following conditions.

(C1) Let \( m(x) = E(y|x) \) be a function of \( x \) that satisfies

\[ |m(x_1) - m(x_2)| < C_1 |d(x_1, x_2)|. \]

for some constant \( C_1 \) for all \( x_1 \) and \( x_2 \).

(C2) Let \( m_2(x) = E(y^2|x) \) be a function of \( x \) that satisfies

\[ |m_2(x_1) - m_2(x_2)| < C_2 |d(x_1, x_2)|. \]

for some constant \( C_2 \) for all \( x_1 \) and \( x_2 \).
(C3) For each \( i \in U \), the cumulative distribution of \( d(x) = d(x_i, x) \), denoted by \( F(d) \), satisfies
\[
\lim_{d \to 0} \frac{F(d)}{d} > 0,
\]
where it is understood that \( F(0) > 0 \) satisfies the condition.

**Proof.** Define \( d_{ij} = d(x_i, x_j) \) for \( i \neq j \) and let \( d_{j(1)} \) be the smallest value of \( d_{ij} \) among \( i \in A \cap \{j\}^c \). By definition, \( d_{j(1)} = d(x_j, x_{j(1)}) \). Similarly, we can define \( d_{j(2)} = d(x_j, x_{j(2)}) \). Note that, for \( a_n = n^{1-\alpha} \),
\[
Pr \left( \max_j a_n d_{j(1)} > M \right) = \sum_{j=1}^{n} Pr \left( d_{j(1)} > M/a_n \right)
= n \left[ 1 - F(M/a_n) \right]^n
= n \exp \left[ a_n n^\alpha \log \left( 1 - F(M/a_n) \right) \right]
\leq n \exp \left[ -a_n n^\alpha F(M/a_n) \right],
\]

since \( \log(1 - x) \leq -x \) for \( x \in [0, 1) \). Thus, writing \( t = M/a_n \), we have
\[
Pr \left( \max_j a_n d_{j(1)} > M \right) \leq n \exp \left[ -M n^\alpha F(t) / t \right]
\]
which goes to zero as \( n \to \infty \), since, by (C3), \( F(t) / t \) is bounded below by some constant greater than zero. Thus, we have \( \max_j a_n d_{j(1)} = O_p(1) \) and, by (C1), we have (6).

To prove (7), we use
\[
Pr \left( \max_j a_n d_{j(2)} > M \right) = \sum_{j=1}^{n} Pr \left( d_{j(2)} > M/a_n \right)
= n \left[ 1 - F(M/a_n) \right]^n + n \left[ 1 - F(M/a_n) \right]^{n-1} F(M/a_n)
= n \left( 1 - F(M/a_n) \right)^{n-1} \{ 1 + (n-1)F(M/a_n) \}
= n \exp \left[ (a_n n^\alpha - 1) \log \left( 1 - F(M/a_n) \right) \{ 1 + (n-1)F(M/a_n) \} \right]
\leq n \exp (1) \exp \left[ -a_n n^\alpha F(M/a_n) \right] \{ 1 + nF(M/a_n) \}
= K n^2 \exp \left[ -M n^\alpha F(t) / t \right],
\]
for some \( K \) where \( t = M/a_n \). Thus, for any \( \epsilon > 0 \), we have \( Pr \left( \max_j a_n d_{j(2)} > M \right) \leq \epsilon \) for sufficiently large \( n \). Therefore, we have \( \max_j a_n d_{j(2)} = O_p(1) \) and, by (C1), (A2) is proved. 

\[\Box\]