Online supplement to
Identification, Data Combination and the Risk of Disclosure
Tatiana Komarova,¹ Denis Nekipelov,² Evgeny Yakovlev,³
This version: December 21, 2016

¹Corresponding author, Department of Economics, London School of Economics and Political Science, e-mail: t.komarova@lse.ac.uk.
²Department of Economics, University of Virginia, e-mail: denis@virginia.edu.
³New Economic School, Skolkovo, Moscow, Russia, e-mail: eyakovlev@nes.ru.
A Construction of individual identifiers

This section overviews existing techniques for the construction of individual identifiers, especially in the case of numerical variables.

The key element of our identification argument is based on the construction of the identifying variables $Z^y$ and $Z^x$ such that we can merge some or all observations in the disjoint databases to be able to estimate the econometric model of interest. While we took the existence of these variables as given, their construction in itself is an important issue and there is a vast literature in applied statistics and computer science that is devoted to the analysis of the broken record linkage. For completeness of the analysis in our paper we present some highlights from that literature.

In general the task of merging disjoint databases is a routine necessity in many practical applications. In many cases there do exist perfect cross-database identifiers of individual entries. There could be multiple reasons why that is the case. For instance, there could be errors in data entry and processing, wrong variable formatting, and duplicate data entry. The idea that has arisen in Newcombe et al. [1959] and was later formalized in Fellegi and Sunter [1969] was to treat the record linkage problem as a problem of classification of record subsets into matches, non-matches and uncertain cases. This classification is based on defining the similarity metric between each two records. Then given the similarity metric one can compute the probability of particular pair of records being a match or non-match. The classification of pairs is then performed by fixing the probability of erroneous identification of a non-matched pair of records as a match and a matched pair of records as a non-match by minimizing the total proportion of pairs that are uncertain. This matching technique is based on the underlying assumption of randomness of records being broken. As a result, using the sample of perfectly matched records one can recover the distribution of the similarity metric for the matched and unmatched pairs of records. Moreover, as in hypothesis testing, one needs to fix the probability of record mis-identification. Finally, the origin of the similarity metric remains arbitrary.

A large fraction of the further literature was devoted to, on one hand, development of classes of similarity metrics that accommodate non-numeric data and, on the other hand, development of fast and scalable record classification algorithms. For obvious reasons, measuring the similarity of string data turns out to be the most challenging. Edit distance (see, Gusfield [1997] for instance) is a metric that can be used to measure the string similarity. The distance between the two strings is determined as the minimum number of insert, delete and replace operations required to transform one string into another. Another measure developed in Jaro [1989] and elaborated in Winkler [1999] is based on the length of matched strings, the number of common characters and their position within the string. In its modification it also allows for the prefixes in the names and is mainly intended to linking relatively short strings such as individual names. Alternative metrics are based on splitting strings into individual “tokens” that are substrings of a particular length and
then analyzing the power of sets of overlapping and non-overlapping tokens. For instance, Jaccard coefficient is based on the relative number of overlapping and overall tokens in two strings. More advanced metrics include the “TF/IDF” metric that is based on the term frequency, or the number of times the term (or token) appears in the document (or string) and the inverse document frequency, or the number of documents containing the given term. The structure of the TF/IDF-based metric construction is outlined in Salton and Harman [2003]. The distance measures may include combination of the edit distance and the TF/IDF distance such as a fuzzy match similarity metric described in Chaudhuri et al. [2003].

Given a specific definition of the distance, the practical aspects of matching observations will entail calibration and application of a particular technique for matching observations. The structure of those techniques is based on, first, the assumption regarding the data structure and the nature of the record errors. Second, it depends on the availability of known matches, and, thus, allows empirical validation of a particular matching technique. When such a validation sample is available, one can estimate the distribution of the similarity measures for matched and non-matched pairs for the validation sample. Then, using the estimated distribution one can assign the matches for the pairs outside the validation sample. When one can use numeric information in addition to the string information, one can use hybrid metrics that combine the known properties of numeric data entries and the properties of string entries.

Ridder and Moffitt [2007] overviews some techniques for purely numeric data combination. In the absence of validation subsamples that may incorporate distributional assumptions on the “similar” numeric variables. For instance, joint normality assumption with a known sign of correlation can allow one to invoke likelihood-based techniques for record linkage.

B More details on the risk of disclosure and choice of threshold sequences

Propositions 2 and 3 demonstrate that the compliance of the decision rule generated by a particular threshold sequence with a given bound guarantee for the disclosure risk depends on the rate at which the threshold sequence converges towards zero as the sizes of $D^y$ and $D^x$ increase. Informally, consider two threshold sequences $\alpha_N$ and $\alpha_N^*$ where the former converges to zero much faster than the latter so that $\frac{\alpha_N^*}{\alpha_N} \rightarrow \infty$. Clearly, for large enough sizes of the datasets $D^y$ and $D^x$, the sequence $\alpha_N^*$ not only allows more observations to be included in the combined dataset but also gives a greater number of possible combined datasets. In fact, all observations with the values of the constructed identifiers $z_i^x$ between $\frac{1}{\alpha_N}$ and $\frac{1}{\alpha_N^*}$ are rejected by the decision rule implied by the sequence $\alpha_N$ but could be approved by the decision rule implied by the sequence $\alpha_N^*$. In addition, the sequence...
\(\alpha_n^*\) is much more liberal in its definition of the proximity between the identifiers \(z^y_j\) and \(z^x_i\). As a result, the decision rule implied by the sequence \(\alpha_n^*\) generates larger combined datasets. Because the matching information in \((-\frac{1}{\alpha_n}, \frac{1}{\alpha_n})\) is less reliable than that in \((-\infty, -\frac{1}{\alpha_n}) \cup (\frac{1}{\alpha_n}, \infty)\) and linkages for observations with larger distances between the identifiers are decreasingly reliable, the sequence \(\alpha_n^*\) results in a larger proportion of incorrect matches. The effect can be so significant that even for arbitrarily large datasets the probability of making a data combination error does not approach 0. In Proposition 2, where non-disclosure is not guaranteed, and the probability of making a data combination error of the first kind approaches 0 as \(N^y\) and \(N^x\) increase, thresholds used for the decision rule shrink to zero faster than those in Proposition 3, where non-disclosure is guaranteed.

In the remarks below we consider cases when the tails of the distributions of identifiers are geometric or exponential.

**REMARK** (Absence of disclosure guarantees). Here we consider cases when the tails of the distributions of identifiers are geometric or exponential.

(a) Suppose that for small enough \(\alpha > 0\), we have \(\phi(\alpha) = b_1\alpha^{c_1},\ b_1, c_1 > 0\) and \(\psi(\alpha) = b_2\alpha^{c_2},\ b_2, c_2 > 0\). If \(\alpha_n > 0\) is chosen in such a way that

\[
\alpha_n = o\left(\frac{1}{(N^x)^{-\frac{1}{2}}\alpha^{\frac{2}{3}}}\right) \tag{B.1}
\]

then

\[
\inf_{x \in X, y \in Y} \inf_{D^x, D^y} \inf_{i,j} p^{N^y}_{ij}(x, y, D^x, D^y) \to 1 \quad \text{as } N^y \to \infty,
\]

and, thus, non-disclosure is not guaranteed.

(b) Alternatively, suppose that for small enough \(\alpha > 0\), we have \(\phi(\alpha) = b_1e^{-c_1/\alpha},\ b_1, c_1 > 0\) and \(\psi(\alpha) = b_2e^{-c_2/\alpha},\ b_2, c_2 > 0\). If \(\alpha_n \to 0\) is chosen in such a way that

\[
\lim_{N^y \to \infty} N^x e^{-\frac{c_2}{\alpha_n}} = 0, \tag{B.2}
\]

then

\[
\inf_{x \in X, y \in Y} \inf_{D^x, D^y} \inf_{i,j} p^{N^y}_{ij}(x, y, D^x, D^y) \to 1 \quad \text{as } N^y \to \infty,
\]

and, thus, non-disclosure is not guaranteed.

For instance, sequences \(\alpha_n = \frac{a}{(N^x)^{d}}\) when \(a, d > 0\), satisfy this condition.

**Proof.** (a) Let us check that if a sequence \(\alpha_n\) is chosen as in (B.1), then it satisfies (3.10).
In other words, let us check that

\[
\frac{N^x}{\alpha_N^{c_1}} \int_1^\infty \left( \frac{1}{z - \alpha_N} - \frac{1}{z + \alpha_N} \right) \frac{1}{z^{c_1+1}} \, dz \to 0 \quad \text{as } N^y \to \infty.
\]

Indeed,

\[
\frac{N^x}{\alpha_N^{c_1}} \int_1^\infty \left( \frac{1}{z - \alpha_N} - \frac{1}{z + \alpha_N} \right) \frac{1}{z^{c_1+1}} \, dz = \frac{N^x}{\alpha_N^{c_1}} \int_1^\infty \left( 1 - \frac{2\alpha_N}{z + \alpha_N} \right) \frac{(z - \alpha_N)^{-c_2}}{z^{c_1+1}} \, dz.
\]

If \( \alpha_N \) is small enough, then for all \( z \geq \frac{1}{\alpha_N} \), it holds that \( 1 - \left( 1 - \frac{2\alpha_N}{z + \alpha_N} \right)^{c_2} \leq q_1 \frac{\alpha_N}{z + \alpha_N} \) for some constant \( q_1 > 0 \). Therefore, if \( \alpha_N \) is small enough, then for all \( z \geq \frac{1}{\alpha_N} \) we have

\[
\left( 1 - \left( 1 - \frac{2\alpha_N}{z + \alpha_N} \right)^{c_2} \right) \frac{(z - \alpha_N)^{-c_2}}{z^{c_1+1}} \leq q_2 \frac{\alpha_N}{z^{c_1+c_2+2}}
\]

for some constant \( q_2 > 0 \). Finally, note that

\[
\frac{q_2 N^x}{\alpha_N^{c_1-c_2}} \int_1^{\infty} \frac{1}{z^{c_1+c_2+2}} \, dz = \frac{q_2 N^x}{1 + c_1 + c_2} \frac{\alpha_N^{c_2+2}}{c_1 + c_2} \to 0 \quad \text{as } N^y \to \infty
\]

if \( \alpha_N \) is chosen as in (B.1).

(b) Let us check that if a sequence \( \alpha_N \) is chosen as in (B.2), then it satisfies (3.10). In other words, let us check that

\[
N^x e^{\frac{c_1}{\alpha_N}} \int_1^\infty \left( e^{-c_2(z-\alpha_N)} - e^{-c_2(z+\alpha_N)} \right) e^{-c_1 z} \, dz \to 0 \quad \text{as } N^y \to \infty.
\]

Indeed,

\[
N^x e^{\frac{c_1}{\alpha_N}} \int_1^\infty \left( e^{-c_2(z-\alpha_N)} - e^{-c_2(z+\alpha_N)} \right) e^{-c_1 z} \, dz = N^x e^{-\frac{c_2}{\alpha_N}} \frac{e^{c_2 \alpha_N} - e^{-c_2 \alpha_N}}{c_1 + c_2}.
\]

Note that for some constant \( r > 0 \)

\[
e^{c_2 \alpha_N} - e^{-c_2 \alpha_N} \leq r \alpha_N.
\]

Now it is clear that if \( \alpha_N \) is chosen as in (B.2), then (3.10) holds. □

REMARK (Disclosure guarantees). (a) Suppose for \( \alpha \in (0, \bar{\alpha}) \), \( \phi(\alpha) = b_1 \alpha c_1, b_1, c_1 > \)
0, and \( \psi(\alpha) = b_2\alpha^{c_2}, \ b_2, c_2 > 0 \). Let the sequence of \( \alpha_N \to 0 \) (as \( N^y \to \infty \)) be chosen in such a way that

\[
\liminf_{N^y \to \infty} \alpha_N \left( \frac{N^x}{c_2+2} \right) > 0.
\] (B.3)

Then non-disclosure is guaranteed.

(b) Suppose for \( \alpha \in (0, \tilde{\alpha}), \ \phi(\alpha) = b_1 e^{-c_1/\alpha}, \ b_1, c_1 > 0 \) and \( \psi(\alpha) = b_2 e^{-c_2/\alpha}, \ b_2, c_2 > 0 \). Let the sequence of \( \alpha_N \to 0 \) (as \( N^y \to \infty \)) be chosen in such a way that

\[
\liminf_{N^y \to \infty} N^x e^{-\frac{c_2}{\alpha_N}} > 0.
\] (B.4)

Then non-disclosure is guaranteed.

For instance, sequences \( \alpha_N = \frac{a}{\log N} \) when \( a > c_2 \), satisfy this condition (in this case, \( \lim_{N^y \to \infty} N^x e^{-\frac{c_2}{\alpha_N}} = \infty \)).

Proof. (a) Let us check that if a sequence \( \alpha_N \) is chosen as in (B.3), then it satisfies (3.11). In other words, let us check that

\[
\liminf_{N^y \to \infty} b_2 c_1 \frac{N^x}{\alpha_N^{c_2}} \int_{1/\alpha_N}^{\infty} \left( \left( \frac{1}{z} - \frac{1}{\alpha_N} \right)^{c_2} - \left( \frac{1}{z + \alpha_N} \right)^{c_2} \right) \frac{1}{z^{c_1+1}} \ dz > 0.
\]

Use \( \left( \frac{1}{z-\alpha_N} \right)^{c_2} - \left( \frac{1}{z+\alpha_N} \right)^{c_2} \) \( 1 - \left( 1 - \frac{2\alpha_N}{z+\alpha_N} \right)^{c_2} \frac{(z-\alpha_N)^{-c_2}}{z^{c_1+1}} \) and note that if \( \alpha_N \) is small enough, then for all \( z \geq \frac{1}{\alpha_N} \),

\[
1 - \left( 1 - \frac{2\alpha_N}{z+\alpha_N} \right)^{c_2} \geq \tilde{q}_1 \frac{\alpha_N}{z+\alpha_N}
\]

for some constant \( \tilde{q}_1 > 0 \). Therefore, if \( \alpha_N \) is small enough, then for all \( z \geq \frac{1}{\alpha_N} \) we have

\[
1 - \left( 1 - \frac{2\alpha_N}{z+\alpha_N} \right)^{c_2} \left( \frac{1}{z-\alpha_N} \right)^{c_2} \frac{1}{z^{c_1+1}} \geq \tilde{q}_2 \frac{\alpha_N}{z^{c_1+c_2+2}}
\]

for some constant \( \tilde{q}_2 > 0 \). Finally, note that if \( \alpha_N \) is chosen as in (B.3), then

\[
\liminf_{N^y \to \infty} \tilde{q}_2 b_2 c_1 \frac{N^x}{\alpha_N^{c_1+1}} \int_{1/\alpha_N}^{\infty} \frac{1}{z^{c_1+c_2+2}} \ dz = \liminf_{N^y \to \infty} \tilde{q}_2 b_2 c_1 \frac{N^x}{1 + c_1 + c_2} \alpha_N^{c_2+2} > 0.
\]

(b) Let us check that if a sequence \( \alpha_N \) is chosen as in (B.4), then it satisfies (3.11). In
other words, we want to check that

$$\liminf_{N^y \to \infty} c_1 N^x e^{\frac{c_1}{N^x}} \int_{\pi_N}^\infty \left( e^{-c_2(z-\alpha_N)} - e^{-c_2(z+\alpha_N)} \right) e^{-c_1 z} \, dz > 0.$$  

Note that

$$N^x e^{\frac{c_1}{N^x}} \int_{\pi_N}^\infty \left( e^{-c_2(z-\alpha_N)} - e^{-c_2(z+\alpha_N)} \right) e^{-c_1 z} \, dz = N^x e^{-c_2} \frac{e^{c_2 \alpha_N} - e^{-c_2 \alpha_N}}{c_1 + c_2}$$

and for some constant $\tilde{r} > 0$

$$e^{c_2 \alpha_N} - e^{-c_2 \alpha_N} \geq \tilde{r} \alpha_N.$$  

Thus, if $\alpha_N$ is chosen as in (B.4), then (3.11) holds. □

It can be seen in the two remarks here that the rates of the threshold sequences used for the decision rule can be described in terms of the size of the dataset $D^x$ alone rather than both $D^y$ and $D^x$. This is quite intuitive because in Assumption 2 that database we assumed that $D^y$ contains the subset of individuals from the database $D^x$, and hence $D^x$ is larger. The size of the larger dataset is the only factor determining how many potential matches from this dataset we are able to find for any observation in the smaller dataset without using any additional information from the identifiers.

C Proof of Proposition 5

Fix $\tilde{\theta} \in \Theta_\infty$. Let $\pi \in \Pi^\infty$ be such that $\tilde{\theta}$ minimizes

$$Q(\theta, \pi) \equiv g_\pi(\theta)' W_0 g_\pi(\theta).$$

We can find a sequence $\{\pi^N(\cdot, \cdot)\}$ that converges to $\pi$ uniformly over all $y$ and all $x$. Let $\theta_N$ be any value that minimizes

$$Q_N(\theta, \pi^N) \equiv g_N(\theta)' W_0 g_N(\theta)$$

for the chosen $\pi^N(\cdot, \cdot)$. Clearly, $\theta_N \to \tilde{\theta}$.

First, we establish that $\sup_{\theta \in \Theta} |Q_N(\theta, \pi^N) - Q(\theta, \pi)| \to 0$. Note that

$$Q_N(\theta, \pi^N) - Q(\theta, \pi) = \left( g_N(\theta) - g_\pi(\theta) \right)' W_0 \left( g_N(\theta) - g_\pi(\theta) \right) + 2 g_\pi(\theta)' W_0 \left( g_N(\theta) - g_\pi(\theta) \right).$$

Therefore,

$$\sup_{\theta \in \Theta} |Q_N(\theta, \pi^N) - Q(\theta, \pi)| \leq \sup_{\theta \in \Theta} \| g_N(\theta) - g_\pi(\theta) \|^2 \| W_0 \| + 2 \sup_{\theta \in \Theta} \| g_\pi(\theta) \| \sup_{\theta \in \Theta} \| g_N(\theta) - g_\pi(\theta) \| \| W_0 \|. $$
Conditions (4.15) imply that \( \sup_{\theta \in \Theta} \|g_\pi(\theta)\| < \infty \). Thus, we only need to establish that \( \sup_{\theta \in \Theta} \|g^N(\theta) - g_\pi(\theta)\| \to 0 \). Using condition (4.18), we can show that \( g^N(\theta) \) can be represented as the sum of four terms –

\[
g^N(\theta) = A_{N1} + A_{N2} + B_{N1} + B_{N2},
\]

where

\[
A_{N1} = \frac{\int \int \int \int_{\frac{1}{\alpha_N} > |z^x_i - z^y_i| < \alpha_N} h(x_i) \rho(y_j, x_i; \theta) f_{Y,X \mid Z^x,Z^y}(y_j, x_i | z^y_i, z^x_i) f_{Z^x,Z^y}(z^x_i, z^y_i) \, dz^y_i \, dx_i \, dy_j \, dx_i}{1 - \pi}
\]

and terms \( o_{yx}(1) \) do not depend on \( \theta \) and are such that \( \sup_{y_j \in \mathcal{Y}, x_i \in \mathcal{X}} |o_{yx}(1)| \to 0 \) as \( \alpha_N \to 0 \).

Proposition 4 implies that \( E \left[ h(X) \rho(Y, X; \theta) \, \mid \, |Z^x| > \frac{1}{\alpha}, |Z^x - Z^y| < \alpha \right] = E[h(X) \rho(Y, X; \theta)] \). Therefore,

\[
A_{N1} = (1 - \pi) E[h(X) \rho(Y, X; \theta)],
\]

and thus,

\[
g^N(\theta) - g_\pi(\theta) = A_{N2} + B_{N1} + B_{N2} - \pi E^x \left[ h(X) \rho(\tilde{Y}, X; \theta) \right].
\]

Note that

\[
\sup_{\theta \in \Theta} \|A_{N2}\| \leq \sup_{y_j \in \mathcal{Y}, x_i \in \mathcal{X}} |o_{yx}(1)| \cdot E \left[ \sup_{\theta \in \Theta} \|h(X) \rho(Y, X; \theta)\| \cdot |Z^x| > \frac{1}{\alpha}, |Z^x - Z^y| < \alpha \right]
\]

\[
= \sup_{y_j \in \mathcal{Y}, x_i \in \mathcal{X}} |o_{yx}(1)| \cdot E \left[ \sup_{\theta \in \Theta} \|h(X) \rho(Y, X; \theta)\| \right] \to 0
\]

as \( \alpha_N \to 0 \). From Assumption 3 (iv), for small \( \alpha_N \) the denominator in \( B_{N1}/\pi \) is the sum

\[
\int \int \int \int_{\frac{1}{\alpha_N} > |z^x_i| > |z^x_i - z^y_i| < \alpha_N} (o_{xz^x} + o_{z^x y}(1) + o_{z^y y}(1) + o_{z^x y}(1)) \times g_2(z^y_j) f_{Y}(y_j) f_X(x_i) \, dz^y_j \, dx_i \, dy_j \, dx_i
\]

\[
+ \int \int \int \int_{\frac{1}{\alpha_N} > |z^x_i| > |z^x_i - z^y_i| < \alpha_N} g_2(z^y_j) g_1(z^x_i) \, dz^y_j \, dx_i,
\]

as \( \alpha_N \to 0 \). From Assumption 3 (iv), for small \( \alpha_N \) the denominator in \( B_{N1}/\pi \) is the sum.
following two terms:

\[ \int \int h(x_i) \rho(y_j, x_i; \theta) f_Y(y_j) f_X(x_i) dy_j dx_i \cdot \int_{|z_i^x| > \frac{1}{\alpha_N}} \int_{|z_i^x - z_j^y| < \alpha_N} g_2(z_j^y) g_1(z_i^x) \, dz_j^y \, dz_i^x + \]

\[ + \int \int h(x_i) \rho(y_j, x_i; \theta) \int_{|z_i^x| > \frac{1}{\alpha_N}} \int_{|z_i^x - z_j^y| < \alpha_N} (o_{z^y z^x}(1) + o_{z^y y}(1) + o_{z^y z^x}(1) + o_{z^y y}(1) o_{z^y z^x}(1)) \times \]

\[ \times g_2(z_j^y) g_1(z_i^x) f_Y(y_j) f_X(x_i) \, dz_j^y \, dz_i^x \, dy_j \, dx_i, \]

where \( o_{y z^y}(1) \) and \( o_{z^y z^x}(1) \) do not depend on \( \theta \) and are such that \( \sup_{|z_i^x| > \frac{1}{\alpha_N}} \sup_{y_j} |o_{y z^y}(1)| \rightarrow 0 \) and \( \sup_{|z_i^x| > \frac{1}{\alpha_N}} \sup_{x_i} |o_{z^y z^x}(1)| \rightarrow 0 \) as \( \alpha_N \rightarrow 0 \). Then \( B_{N_1} - \pi E^* \left[ h(X) \rho(\tilde{Y}, X; \theta) \right] \) is the sum of the following two terms:

\[ \pi E^* \left[ h(X) \rho(\tilde{Y}, X; \theta) \right] \cdot \frac{C_{N_1}}{C_{N_1} + \int \int D_{N_1}(y_j, x_i) f_Y(y_j) f_X(x_i) \, dy_j \, dx_i} - 1 \] (C.5)

and

\[ \pi \cdot \frac{\int \int h(x_i) \rho(y_j, x_i; \theta) D_{N_1}(y_j, x_i) f_Y(y_j) f_X(x_i) \, dy_j \, dx_i}{C_{N_1} + \int \int D_{N_1}(y_j, x_i) f_Y(y_j) f_X(x_i) \, dy_j \, dx_i} \] (C.6)

where

\[ C_{N_1} = \int_{|z_i^x| > \frac{1}{\alpha_N}} \int_{|z_i^x - z_j^y| < \alpha_N} g_2(z_j^y) g_1(z_i^x) \, dz_j^y \, dz_i^x, \]

\[ D_{N_1}(y_j, x_i) = \int_{|z_i^x| > \frac{1}{\alpha_N}} \int_{|z_i^x - z_j^y| < \alpha_N} (o_{z^y z^x}(1) + o_{z^y y}(1) + o_{z^y z^x}(1) + o_{z^y y}(1) o_{z^y z^x}(1)) g_2(z_j^y) g_1(z_i^x) \, dz_j^y \, dz_i^x. \]

The supremum over \( \theta \in \Theta \) of the norm of the term in (C.5) is bounded from above by

\[ \pi E^* \left[ \sup_{\theta \in \Theta} \| h(X) \rho(\tilde{Y}, X; \theta) \| \right] \cdot \frac{C_{N_1}}{C_{N_1} + \int \int D_{N_1}(y_j, x_i) f_Y(y_j) f_X(x_i) \, dy_j \, dx_i} - 1. \]

Because

\[ |D_{N_1}(y_j, x_i)| \leq \sup_{|z_i^x| > \frac{1}{\alpha_N}} \sup_{|z_j^y| > \frac{1}{\alpha_N}} \sup_{y_j, x_i} \sup_{o_{z^y z^x}(1)} \cdot C_{N_1} \]

with \( \sup_{|z_i^x| > \frac{1}{\alpha_N}} \sup_{|z_j^y| > \frac{1}{\alpha_N}} \sup_{y_j, x_i} o_{y z^y}(1) \rightarrow 0 \) as \( \alpha_N \rightarrow 0 \). Hence, (C.5) converges to 0 uniformly over \( \theta \in \Theta \).

The supremum over \( \theta \in \Theta \) of the norm of the term in (C.6) is bounded from above by

\[ \pi \cdot \sup_{\theta \in \Theta} \sup_{|z_i^x| > \frac{1}{\alpha_N}} \sup_{|z_j^y| > \frac{1}{\alpha_N}} \sup_{y_j, x_i} \sup_{o_{z^y z^x}(1)} \cdot \frac{C_{N_1} \cdot E^* \left[ \sup_{\theta \in \Theta} \| h(X) \rho(\tilde{Y}, X; \theta) \| \right]}{C_{N_1} + \int \int D_{N_1}(y_j, x_i) f_Y(y_j) f_X(x_i) \, dy_j \, dx_i}, \]

which converges to 0 as \( \alpha_N \rightarrow 0 \). Thus, we obtain that

\[ \sup_{\theta \in \Theta} \| B_{N_1} - \pi E^* \left[ h(X) \rho(\tilde{Y}, X; \theta) \right] \| \rightarrow 0. \]
Finally, consider \( \sup_{\theta \in \Theta} \| B_{N^2} \| \). This norm is bounded from above by the sum of 
\[
\sup_{y_j, x_i} |o_{y_jx_i}(1)| \cdot \int_{\Theta} \sup_{\theta \in \Theta} \| h(x_i)\rho(y_j, x_i; \theta)\| f_Y(y_j)f_X(x_i) dy_j dx_i \cdot \frac{C_{N_1}}{C_{N_1} + \int \int D_{N_1}(y_j, x_i)f_Y(y_j)f_X(x_i) dy_j dx_i} \]
and
\[
\sup_{y_j, x_i} |o_{y_jx_i}(1)| \cdot \sup_{|z_i| > \frac{1}{N}} \sup_{|y_j| > \frac{1}{N}} \sup_{|z_j| > \alpha_N} |o_{y_jz_jx_i}(1)| \cdot \frac{C_{N_1} \int \sup_{\theta \in \Theta} \| h(x_i)\rho(y_j, x_i; \theta)\| f_Y(y_j)f_X(x_i) dy_j dx_i}{C_{N_1} + \int \int D_{N_1}(y_j, x_i)f_Y(y_j)f_X(x_i) dy_j dx_i},
\]
and, hence, \( \sup_{\theta \in \Theta} \| B_{N^2} \| \to 0 \) as \( \alpha_N \to 0 \).

To summarize our results so far, we showed that
\[
\sup_{\Theta} \| g^N(\theta) - g_\pi(\theta) \| \leq \sup_{\Theta} \| A_{N^2} \| + \sup_{\Theta} \| B_{N^2} - \pi E^\pi [h(X)\rho(\tilde{Y}, X; \theta)] \| + \sup_{\Theta} \| B_{N^2} \|, 
\]
and thus, \( \sup_{\Theta} \| g^N(\theta) - g_\pi(\theta) \| \to 0 \) as \( \alpha_N \to 0 \). This implies that
\[
\sup_{\Theta} |Q_N(\theta, \pi^N) - Q(\theta, \pi)| \to 0. \tag{C.7}
\]

Now, fix \( \varepsilon > 0 \). Let us show that for large enough \( N^x, N^y, Q(\theta^N, \pi) < Q(\tilde{\theta}, \pi) + \varepsilon \). Indeed, (C.7) implies that when \( N^x, N^y \) are large enough, \( Q(\theta^N, \pi) < Q_N(\theta^N, \pi^N) + \varepsilon/3 \). Also, \( Q_N(\theta^N, \pi^N) < Q_N(\tilde{\theta}, \pi^N) + \varepsilon/3 \) because \( \theta^N \) is an argmin of \( Q_N(\theta^N, \pi^N) \). Finally, (C.7) implies that when \( N^x, N^y \) are large enough, \( Q_N(\tilde{\theta}, \pi^N) < Q(\tilde{\theta}, \pi) + \varepsilon/3 \).

Let \( S \) be any open neighborhood of \( \tilde{\theta} \) and let \( S^c \) be its complement in \( \mathbb{R}^l \). From the compactness of \( \Theta \) and the continuity of \( \rho(\cdot, \cdot) \) in \( \theta \), we conclude that \( \min_{\tilde{\theta}} Q(\tilde{\theta}, \pi) \) is attained. The fact that \( \tilde{\theta} \) is the unique minimizer of \( Q(\tilde{\theta}, \pi) \) gives that \( \min_{\tilde{\theta}} Q(\tilde{\theta}, \pi) > Q(\tilde{\theta}, \pi) \). Denote \( \varepsilon = \min_{\tilde{\theta}} Q(\tilde{\theta}, \pi) - Q(\tilde{\theta}, \pi) \). As we showed above, for this \( \varepsilon \) we have that when \( N^x, N^y \) are large enough,
\[
Q(\theta^N, \pi) < Q(\tilde{\theta}, \pi) + \varepsilon = \min_{S^c} Q(\tilde{\theta}, \pi),
\]
which for large enough \( N^x, N^y \) gives \( \theta^N \in S \). Since \( S \) can be chosen arbitrarily small, this means that \( \theta^N \to \tilde{\theta} \). \( \square \)

References


computational biology, Cambridge University Press.


