Kernel Density Estimation for Undirected Dyadic Data

July 31, 2019

Bryan S. Graham[1], Fengshi Niu[#] and James L. Powell[†]

<u>Initial Draft</u>: January 2019, <u>This Draft</u>: July 2019

[1]Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: `http://bryangraham.github.io/econometrics/`.

[#]Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880, e-mail: fniu@berkeley.edu.

[†]Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880, e-mail: powell@econ.berkeley.edu.

We thank Michael Jansson and Konrad Menzel for helpful conversation as well as audiences at Berkeley, Brown, Toulouse, Warwick, Bristol, University College London and the Conference Celebrating Whitney Newey's Contributions to Econometrics for useful questions and suggestions. All the usual disclaimers apply. Financial support from NSF grant SES #1851647 is gratefully acknowledged.

Abstract

We study nonparametric estimation of density functions for undirected dyadic random variables (i.e., random variables defined for all $n \overset{def}{\equiv} \binom{N}{2}$ unordered pairs of agents/nodes in a weighted network of order N). These random variables satisfy a local dependence property: any random variables in the network that share one or two indices may be dependent, while those sharing no indices in common are independent. In this setting, we show that density functions may be estimated by an application of the kernel estimation method of Rosenblatt (1956) and Parzen (1962). We suggest an estimate of their asymptotic variances inspired by a combination of (i) Newey's (1994) method of variance estimation for kernel estimators in the "monadic" setting and (ii) a variance estimator for the (estimated) density of a simple network first suggested by Holland & Leinhardt (1976). More unusual are the rates of convergence and asymptotic (normal) distributions of our dyadic density estimates. Specifically, we show that they converge at the same rate as the (unconditional) dyadic sample mean: the square root of the number, N, of nodes. This differs from the results for nonparametric estimation of densities and regression functions for monadic data, which generally have a slower rate of convergence than their corresponding sample mean.

JEL Classification: C24, C14, C13.

Keywords: Networks, Dyads, Kernel Density Estimation

1 Introduction

Many important social and economic variables are naturally defined for pairs of agents (or dyads). Examples include trade between pairs of countries (e.g., Tinbergen, 1962), input purchases and sales between pairs of firms (e.g., Atalay et al., 2011), research and development (R&D) partnerships across firms (e.g., König et al., 2019) and friendships between individuals (e.g., Christakis et al., 2010). Dyadic data arises frequently in the analysis of social and economic networks. In economics such analyses are predominant in, for example, the analysis of international trade flows. See Graham (TBD) for many other examples and references.

While the statistical analysis of network data began almost a century ago, rigorously justified methods of inference for network statistics are only now emerging (cf., Goldenberg et al., 2009). In this paper we study nonparametric estimation of the density function of a (continuously-valued) dyadic random variable. Examples included the density of migration across states, trade across nations, liabilities across banks, or minutes of telephone conversation among individuals. While nonparametric density estimation using independent and identically distributed random samples, henceforth "monadic" data, is well-understood, its dyadic counterpart has, to our knowledge, not yet been studied.

Holland & Leinhardt (1976) derived the sampling variance of the link frequency in a simple network (and of other low order subgraph counts). A general asymptotic distribution theory for subgraph counts, exploiting recent ideas from the probability literature on dense graph limits (e.g., Diaconis & Janson, 2008; Lovász, 2012), was presented in Bickel et al. (2011).[2] Menzel (2017) presents bootstrap procedures for inference on the mean of a dyadic random variable. Our focus on nonparametric density estimation appears to be novel. Density estimation is, of course, a topic of intrinsic interest to econometricians and statisticians, but it also provides a relatively simple and canonical starting point for understanding nonparametric estimation more generally. In the conclusion of this paper we discuss ongoing work on other non- and semi-parametric estimation problems using dyadic data.

We show that an (obvious) adaptation of the Rosenblatt (1956) and Parzen (1962) kernel density estimator is applicable to dyadic data. While our dyadic density estimator is straightforward to define, its rate-of-convergence and asymptotic sampling properties, depart significantly from its monadic counterpart. Let N be the number of sampled agents and $n = \binom{N}{2}$ the corresponding number of dyads. Estimation is based upon the n dyadic outcomes. Due to dependence across dyads sharing an agent in common, the rate of convergence of our density estimate is (generally) much *slower* than it would be with n i.i.d. outcomes. This rate-of-convergence is also invariant across a wide range of bandwidth sequences. This property is familiar from the econometric literature on

[2]See Nowicki (1991) for a summary of earlier research in this area.

semiparametric estimation (e.g., Powell, 1994). Indeed, from a certain perspective, our nonparametric dyadic density estimate can be viewed as a semiparametric estimator (in the sense that it can be thought of as an average of nonparametrically estimated densities). We also explore the impact of "degeneracy" – which arises when dependence across dyads vanishes – on our sampling theory; such degeneracy features prominently in Menzel's (2017) innovative analysis of inference on dyadic means. We expect that many of our findings generalize to other non- and semi-parametric network estimation problems.

In the next section we present our maintained data/network generating process and proposed kernel density estimator. Section 3 explores the mean square error properties of this estimator, while Section 4 outlines asymptotic distribution theory. Section 5 presents a consistent variance estimator, which can be used to construct Wald statistics and Wald-based confidence intervals. We summarize the results of a small simulation study in Section 6. In Section 7 we discuss various extensions and ongoing work. Calculations not presented in the main text are collected in Appendix A.

It what follows we interchangeably use unit, node, vertex, agent and individual all to refer to the $i = 1, \ldots, N$ vertices of the sampled network or graph. We denote random variables by capital Roman letters, specific realizations by lower case Roman letters and their support by blackboard bold Roman letters. That is Y, y and $\mathbb{Y}$ respectively denote a generic random draw of, a specific value of, and the support of, Y. For W_{ij} a dyadic outcome, or weighted edge, associated with agents i and j, we use the notation $\mathbf{W} = [W_{ij}]$ to denote the $N \times N$ adjacency matrix of all such outcomes/edges. Additional notation is defined in the sections which follow.

2 Model and estimator

Model

Let $i = 1, \ldots, N$ index a simple random sample of N agents from some large (infinite) network of interest. A pair of agents constitutes a *dyad*. For each of the $n = \binom{N}{2}$ sampled dyads, that is for $i = 1, ..., N-1$ and $j = i+1, \ldots, N$, we observe the (scalar) random variable W_{ij}, generated according to

$$W_{ij} = W(A_i, A_j, V_{ij}) = W(A_j, A_i, V_{ij}), \tag{1}$$

where A_i is a node-specific random vector of attributes (of arbitrary dimension, not necessarily observable), and $V_{ij} = V_{ji}$ is an unobservable scalar random variable which is continuously distributed on $\mathbb{R}$ with density function $f_V(v)$.[3] Observe that the function

[3] In words we observe the weighted subgraph induced by the randomly sampled agents.

$W(a_1, a_2, v_{12})$ is symmetric in its first two arguments, ensuring that $W_{ij} = W_{ji}$ is undirected.

In what follows we directly maintain (1), however, it also a consequence of assuming that the infinite graph sampled from is jointly exchangeable (Aldous, 1981; Hoover, 1979). Joint exchangeability of the sampled graph $\mathbf{W} = [W_{ij}]$ implies that

$$[W_{ij}] \stackrel{D}{=} \left[W_{\pi(i)\pi(j)}\right] \tag{2}$$

for every $\pi \in \Pi$ where $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$ is a permutation of the node indices. Put differently, when node labels have no meaning we have that the "likelihood" of any simultaneous row and column permutation of $\mathbf{W}$ is the same as that of $\mathbf{W}$ itself.[4] See Menzel (2017) for a related discussion.

Our target object of estimation is the marginal density function $f_W(w)$ of W_{ij}, defined as the derivative of the cumulative distribution function (c.d.f.) of W_{ij},

$$\Pr\{W_{ij} \leq w\} \stackrel{def}{\equiv} F_W(w) = \int_{-\infty}^{w} f_W(u)\mathrm{d}u.$$

To ensure this density function is well-defined on the support of W_{ij}, we assume that the unknown function $W(a_1, a_2, v)$ is strictly increasing and continuously differentiable in its third argument v, and we also assume that A_i and A_j are statistically independent of the "error term" V_{ij} for all i and j. Under these assumptions, by the usual change-of-variables formula, the conditional density of W_{ij} given $A_i = a_1$ and $A_j = a_2$ takes the form

$$f_{Y|AA}(w|a_1, a_2) = f_V(W^{-1}(a_1, a_2, w)) \cdot \left| \frac{\partial W(a_1, a_2, W^{-1}(a_1, a_2, w))}{\partial v} \right|^{-1}.$$

In the derivations below we will assume this density function is bounded and twice continuously differentiable at w with bounded second derivative for all a_1 and a_2; this will follow from the similar smoothness conditions imposed on the primitives $W^{-1}(\cdot, \cdot, w)$ and $f_V(v)$.

To derive the marginal density of W_{ij} note that, by random sampling, the $\{A_i\}$ sequence is independently and identically distributed (i.i.d.), as is the $\{V_{ij}\}$ sequence. Under these conditions, we can define the conditional densities of W_{ij} given $A_i = a$ or $A_j = a$ alone as

$$f_{W|A}(w|a) \equiv \mathbb{E}[f_{W|AA}(w|a, A_j)] = \mathbb{E}[f_{W|AA}(w|A_i, a)],$$

[4]For $\mathbf{W} = [W_{ij}]$ the $N \times N$ weighted adjacency matrix and $\mathbf{P}$ any conformable permutation matrix

$$\Pr(\mathbf{W} \leq \mathbf{w}) = \Pr(\mathbf{PWP} \leq \mathbf{w})$$

for all $\mathbf{w} \in \mathbb{W} = \mathbb{R}^{\binom{N}{2}}$.

and, averaging, the marginal density of interest as

$$f_W(w) \overset{def}{\equiv} \mathbb{E}[f_{W|AA}(w|A_i, A_j)] = \mathbb{E}[f_{W|A}(w|A_i)].$$

Let i, j, k and l index distinct agents. The assumption that $\{A_i\}$ and $\{V_{ij}\}$ are i.i.d. implies that while W_{ij} varies independently of W_{kl} (since the $\{i, j\}$ and $\{k, l\}$ dyads share no agents in common), W_{ij} will not vary independently of W_{ik} as both vary with A_i (since the $\{i, j\}$ and $\{i, k\}$ dyads both include agent i). This type of dependence structure is sometimes referred to as "dyadic clustering" in empirical social science research (cf., Fafchamps & Gubert, 2007; Cameron & Miller, 2014; Aronow et al., 2017). The implications of this dependence structure for density estimation and – especially – inference is a key area of focus in what follows.

Estimator

Given this construction of the marginal density $f_W(w)$ of W_{ij}, it can be estimated using an immediate extension of the kernel density estimator for monadic data first proposed by Rosenblatt (1956) and Parzen (1962):

$$\begin{aligned}\hat{f}_W(w) &= \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=1+1}^{N} \frac{1}{h} K\left(\frac{w - W_{ij}}{h}\right) \\ &\overset{def}{\equiv} \frac{1}{n} \sum_{i<j} K_{ij},\end{aligned}$$

where

$$K_{ij} \overset{def}{\equiv} \frac{1}{h} K\left(\frac{w - W_{ij}}{h}\right).$$

Here $K(\cdot)$ is a kernel function assumed to be (i) bounded ($K(u) \leq \bar{K}$ for all u), (ii) symmetric ($K(u) = K(-u)$), (ii) , and zero outside a bounded interval ($K(u) = 0$ if $|u| > \bar{u}$); we also require that it (iv) integrates to one ($\int K(u)du = 1$). The bandwidth $h = h_N$ is assumed to be a positive, deterministic sequence (indexed by the number of nodes N) that tends to zero as $N \to \infty$, and will satisfy other conditions imposed below. A discussion of the motivation for the kernel estimator $\hat{f}_W(w)$ and its statistical properties under random sampling (of monadic variables) can be found in Silverman (1986, Chapers 2 & 3).

3 Rate of convergence analysis

To formulate conditions for consistency of $\hat{f}_W(w)$, we will evaluate its expectation and variance, which will yield conditions on the bandwidth sequence h_N for its mean squared error to converge to zero.

A standard calculation yields a bias of $\hat{f}_W(w)$ equal to (see Appendix A)

$$E\left[\hat{f}_W(w)\right] - f_W(w) = h^2 B(w) + o(h^2) \qquad (3)$$
$$= O(h_N^2),$$

with

$$B\left(w\right) \overset{def}{\equiv} \frac{1}{2}\frac{\partial^2 f_W(w)}{\partial w^2}\int u^2 K\left(u\right)\mathrm{d}u.$$

Equation (3) coincides with the bias of the kernel density estimate based upon a random ("monadic") sample.

The expression for the variance of $\hat{f}_W(w)$, in contrast to that for bias, does differ from the monadic (i.i.d.) case due to the (possibly) nonzero covariance between K_{ij} and K_{ik} for $j \neq k$:

$$\begin{aligned}
\mathbb{V}\left(\hat{f}_W(w)\right) &= \mathbb{V}\left(\frac{1}{n}\sum_{i<j} K_{ij}\right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i<j}\sum_{k<l} \mathbb{C}(K_{ij}, K_{kl}) \\
&= \left(\frac{1}{n}\right)^2 \left[n \cdot \mathbb{C}(K_{12}, K_{12}) + 2n(N-2)\cdot \mathbb{C}(K_{12}, K_{13})\right] \\
&= \frac{1}{n}\left[\mathbb{V}(K_{12}) + 2(N-2)\cdot \mathbb{C}(K_{12}, K_{13})\right].
\end{aligned}$$

The third line of this expression uses the fact that, in the summation in the second line, there are $n = \frac{1}{2}N\left(N-1\right)$ terms with $(i,j) = (k,l)$ and $N(N-1)(N-2) = 2n(N-2)$ terms with one subscript in common; as noted earlier, when W_{ij} and W_{kl} have no subscripts in common they are independent (and thus uncorrelated).

To calculate the dependence of this variance on the number of nodes N, we analyze $\mathbb{V}(K_{12})$ and $\mathbb{C}(K_{12}, K_{13})$. Beginning with the former,

$$\begin{aligned}
\mathbb{V}(K_{12}) &= \mathbb{E}\left[(K_{12})^2\right] - \left(\mathbb{E}[\hat{f}_W(w)]\right)^2 \\
&= \frac{1}{h^2}\int \left[K\left(\frac{w-s}{h}\right)\right]^2 f_W(s)\mathrm{d}s + O(1) \\
&= \frac{1}{h}\int [K\left(u\right)]^2 f_W(w-hu)\mathrm{d}u + O(1) \\
&= \frac{f_W(w)}{h}\cdot \int [K\left(u\right)]^2 \mathrm{d}u + O(1) \\
&\overset{def}{\equiv} \frac{1}{h_N}\Omega_2(w) + O(1),
\end{aligned}$$

where

$$\Omega_2(w) \overset{def}{\equiv} f_W(w) \cdot \int [K(u)]^2 \mathrm{d}u.$$

Like the expected value, this own variance term is of the same order of magnitude as in the monadic case,

$$\mathbb{V}(K_{12}) = O\left(\frac{1}{h}\right).$$

However, the covariance term $\mathbb{C}(K_{ij}, K_{il})$, which would be absent for i.i.d. monadic data, is generally nonzero. Since

$$\begin{aligned}
\mathbb{E}[K_{ij} \cdot K_{ik}] &= \mathbb{E}\left[\int\int \frac{1}{h^2}\left[K\left(\frac{w-s_1}{h}\right)\right] \cdot \left[K\left(\frac{w-s_2}{h}\right)\right]\right. \\
&\quad \left. \cdot f_{W|AA}(s_1|A_1, A_2) f_{W|AA}(s_2|A_1, A_3)\mathrm{d}s_1\mathrm{d}s_2\right] \\
&= \mathrm{E}\left[\int [K(u_1)]\, f_{W|A}(w - hu_1|A_1)\mathrm{d}u_1\right. \\
&\quad \left. \cdot \int [K(u_2)]\, f_{W|A}(w - hu_2|A_1)\mathrm{d}u_2\right], \\
&= \mathbb{E}\left[f_{W|A}(w|A_1)^2\right] + o(1),
\end{aligned}$$

(where the second line uses the change of variables $s_1 = w - hu_1$ and $s_2 = w - hu_2$ and mutual independence of A_1, A_2, and A_3). It follows that

$$\begin{aligned}
\mathbb{C}(K_{ij}, K_{ik}) &= \mathbb{E}[K_{ij} \cdot K_{ik}] - \left(\mathbb{E}[\hat{f}_W(w)]\right)^2 \\
&= \left[\mathbb{E}\left[f_{W|A}(w|A_1)^2\right] - f_W(w)^2\right] + O(h^2) \\
&= \mathbb{V}(f_{W|A}(w|A_1)) + o(1) \\
&\overset{def}{\equiv} \Omega_1(w) + o(1),
\end{aligned}$$

with

$$\Omega_1(w) \overset{def}{\equiv} \mathbb{V}(f_{W|A}(w|A_1)).$$

Therefore,

$$\begin{aligned}
\mathbb{V}\left(\hat{f}_W(w)\right) &= \frac{1}{n}\left[2(N-2) \cdot \mathbb{C}(K_{12}, K_{13}) + \mathbb{V}(K_{12})\right] \\
&= \frac{4}{N}\Omega_1(w) + \left(\frac{1}{nh}\Omega_2(w) - \frac{2}{n}\Omega_1(w)\right) + o\left(\frac{1}{N}\right) \qquad (4) \\
&= O\left(\frac{4\Omega_1(w)}{N}\right) + O\left(\frac{\Omega_2(w)}{nh}\right).
\end{aligned}$$

and the mean-squared error of $\hat{f}_W(w)$ is, using (3) and (4),

$$\begin{aligned} \text{MSE}\left(\hat{f}_W(w)\right) &= \left(\mathbb{E}[\hat{f}_W(w)] - f_W(w)\right)^2 + \mathbb{V}\left(\hat{f}_W(w)\right) \\ &= h^4 B(w)^2 + \frac{4}{N}\Omega_1(w) + \left(\frac{1}{nh}\Omega_2(w) - \frac{2}{n}\Omega_1(w)\right) \\ &\quad + o(h^4) + o\left(\frac{1}{N}\right) \\ &= O\left(h^4\right) + O\left(\frac{4\Omega_1(w)}{N}\right) + O\left(\frac{\Omega_2(w)}{nh}\right) \end{aligned} \tag{5}$$

Provided that $\Omega_1(w) \neq 0$ and the bandwidth sequence h_N is chosen such that

$$Nh \to \infty, \qquad Nh^4 \to 0 \tag{6}$$

as $N \to \infty$, we get that

$$\begin{aligned} \text{MSE}\left(\hat{f}_W(w)\right) &= o\left(\frac{1}{N}\right) + O\left(\frac{1}{N}\right) + o\left(\frac{1}{N}\right) \\ &= O\left(\frac{1}{N}\right), \end{aligned}$$

and hence that

$$\sqrt{N}(\hat{f}_W(w) - f_W(w)) = O_p(1).$$

In fact, the rate of convergence of $\hat{f}_W(w)$ to $f_W(w)$ will be $\sqrt{N}$ as long as $Nh^4 \leq C \leq Nh$ for some $C > 0$ as $N \to \infty$, although the mean-squared error will include an additional bias or variance term of $O(N^{-1})$ if either Nh or $(Nh^4)^{-1}$ does not diverge to infinity.

To derive the MSE-optimal bandwidth sequence we minimize (5) with respect to its first and third terms, this yields an optimal bandwidth sequence of

$$\begin{aligned} h_N^*(w) &= \left[\frac{1}{4}\frac{\Omega_2(w)}{B(w)^2}\frac{1}{n}\right]^{\frac{1}{5}} \\ &= O\left(N^{-\frac{2}{5}}\right). \end{aligned} \tag{7}$$

This sequence satisfies condition (6) above.

Interestingly, the rate of convergence of $\hat{f}_W(w)$ to $f_W(w)$ under condition (6) is the same as the rate of convergence of the sample mean

$$\bar{W} \stackrel{def}{\equiv} \frac{1}{n}\sum_{i<j} W_{ij} \tag{8}$$

to its expectation $\mu_W \stackrel{def}{\equiv} \mathbb{E}[W_{ij}]$ when $\mathbb{E}[W_{ij}^2] < \infty$. Similar variance calculations to those for $\hat{f}_w(w)$ yield (see also Holland & Leinhardt (1976) and Menzel (2017))

$$\begin{aligned}\mathbb{V}(\bar{W}) &= O\left(\frac{\mathbb{V}(W_{ij})}{n}\right) + O\left(\frac{4\mathbb{V}(\mathbb{E}[W_{ij}|A_i])}{N}\right) \\ &= O\left(\frac{1}{N}\right),\end{aligned}$$

provided $\mathbb{E}[W_{ij}|A_i]$ is non-degenerate, yielding

$$\sqrt{N}(\bar{W} - \mu) = O_p(1).$$

Thus, in contrast to the case of i.i.d monadic data, there is no convergence-rate "cost" associated with nonparametric estimation of $f_W(w)$. The presence of dyadic dependence, due to its impact on estimation variance, does slow down the feasible rate of convergence substantially. With iid data the relevant rate for density estimation would be $n^{2/5}$ when the MSE-optimal bandwidth sequence is used. Recalling that $n = O(N^2)$, the $\sqrt{N}$ rate we find here corresponds to an $n^{1/4}$ rate. The slowdown from $n^{2/5}$ to $n^{1/4}$ captures the rate of convergence costs of dyadic dependence on the variance of our density estimate.

The lack of dependence of the convergence rate of $\hat{f}_W(w)$ to $f_W(w)$ on the precise bandwidth sequence chosen is analogous to that for semiparametric estimators defined as averages over nonparametrically-estimated components (e.g., Newey, 1994; Powell, 1994). Defining $K_{ji} \overset{def}{\equiv} K_{ij}$, the estimator $\hat{f}_W(w)$ can be expressed as

$$\hat{f}_W(w) = \frac{1}{N}\sum_{i=1}^{N} \hat{f}_{W|A}(w|A_i),$$

where

$$\hat{f}_{W|A}(w|A_i) \overset{def}{\equiv} \frac{1}{N-1}\sum_{j\neq i, j=1}^{N} K_{ij}.$$

Holding i fixed, the estimator $\hat{f}_{W|A}(W|A_i)$ can be shown to converge to $f_{W|A}(w|A_i)$ at the nonparametric rate $\sqrt{Nh}$, but the average of this nonparametric estimator over A_i converges at the faster ("parametric") rate $\sqrt{N}$. In comparison, while

$$\bar{W} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbb{E}}\left[W_{ij}|\, A_i\right],$$

for

$$\hat{\mathbb{E}}\left[W_{ij}|\, A_i\right] \overset{def}{\equiv} \frac{1}{N-1}\sum_{j\neq i, j=1}^{N} W_{ij},$$

the latter converges at the parametric rate $\sqrt{N}$, and the additional averaging to obtain $\bar{W}$ does not improve upon that rate.

4 Asymptotic distribution theory

To derive conditions under which $\hat{f}_W(w)$ is approximately normally distributed it is helpful to decompose the difference between $\hat{f}_W(w)$ and $f_W(w)$ into four terms:

$$\begin{aligned}
\hat{f}_W(w) - f_W(w) &= \frac{1}{n}\sum_{i<j}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j]) && (9)\\
&+ \frac{1}{n}\sum_{i<j}\mathbb{E}[K_{ij}|A_i, A_j] && (10)\\
&- \left(\mathbb{E}[K_{ij}] + \frac{2}{N}\sum_{i=1}^{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)\\
&+ \frac{2}{N}\sum_{i=1}^{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}]) && (11)\\
&+ \mathbb{E}[K_{ij}] - f_W(w) && (12)\\
&\equiv T_1 + T_2 + T_3 + T_{4.}
\end{aligned}$$

To understand this decomposition observe that the projection of $\hat{f}_W(w) = \frac{1}{n}\sum_{i<j}K_{ij}$ onto $\{A_i\}_{i=1}^N$ equals, by the independence assumptions imposed on $\{A_i\}$ and $\{V_{ij}\}$, the U-statistic $\binom{N}{2}^{-1}\sum_{i<j}\mathbb{E}[K_{ij}|A_i, A_j]$. This U-Statistic is defined in terms of the *latent* i.i.d. random variables $\{A_i\}_{i=1}^N$.

The first term in this expression, line (9), is $\hat{f}_W(w)$ minus the projection/U-Statistic described above. Each term in this summation has conditional expectation zero given the remaining terms (i.e., the terms form a martingale difference sequence).

The second term in the decomposition, line (10), is the difference between the second-order U-statistic $\frac{1}{n}\sum_{i<j}\mathbb{E}[K_{ij}|A_i, A_j]$ and its Hájek projection (e.g., van der Vaart, 2000)[5], the third term, line (11), is a centered version of that Hájek projection, and the final term, line (12), is the bias of $\hat{f}_W(w)$. A similar "double projection" argument was used by Graham (2017) to analyze the large sample properties of the Tetrad Logit estimator.

If the bandwidth sequence $h = h_N$ satisfies the conditions $Nh \to \infty$ and $Nh^4 \to 0$, the calculations in the previous section can be used to show that the first, second, and fourth terms of this decomposition (i.e., T_1, $T_{2,}$ and T_4) will all converge to zero when normalized by $\sqrt{N}$. In this case, T_3, which is an average of i.i.d. random variables, will be the leading term asymptotically such that

$$\sqrt{N}(\hat{f}_W(w) - f_W(w)) \xrightarrow{D} \mathcal{N}(0, 4\Omega_1(w)),$$

assuming $\Omega_1(w) = \mathbb{V}(f_{W|A}(w|A_i)) > 0$.

[5]That is the projection of $\frac{1}{n}\sum_{i<j}\mathbb{E}[K_{ij}|A_i, A_j]$ onto the linear subspace consisting of all functions of the form $\sum_{i=1}^N g_i(A_i)$.

If, however, the bandwidth sequence h has $Nh \to C < \infty$ (a "knife-edge" undersmoothing condition similar to one considered by Cattaneo et al. (2014) in a different context), then both T_1 and T_3 will be asymptotically normal when normalized by $\sqrt{N}$. To accommodate both of these cases in a single result, we will show that a standardized version of the sum $T_1 + T_3$ will have a standard normal limit distribution, although the first, T_1, term may be degenerate in the limit.

In Appendix A we show that both T_2 and T_4 will be asymptotically negligible when normalized by the convergence rate of $T_1 + T_3$, such that the asymptotic distribution of $\hat{f}_W(w)$ will only depend on the T_1 and T_3 terms.

We start by rewriting the sum of terms T_1 and T_3 as

$$\begin{aligned} T_1 + T_3 &= \frac{1}{n}\sum_{i<j}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j]) + \frac{2}{N}\sum_{i=1}^{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}]) \\ &\overset{def}{\equiv} \sum_{t=1}^{T(N)} X_{Nt}, \end{aligned}$$

where

$$T(N) \equiv N + n$$

and the triangular array X_{Nt} is defined as

$$\begin{aligned} X_{N1} &= \frac{2}{N}\left(\mathbb{E}[K_{12}|A_1] - \mathbb{E}[K_{12}]\right), \\ X_{N2} &= \frac{2}{N}\left(\mathbb{E}[K_{23}|A_2] - \mathbb{E}[K_{23}]\right), \\ &\vdots \\ X_{NN} &= \frac{2}{N}(E[K_{N,1}|A_N] - \mathbb{E}[K_{N,1}]), \\ X_{N,N+1} &= \frac{1}{n}(K_{12} - \mathbb{E}[K_{12}|A_1, A_2]), \\ X_{N,N+2} &= \frac{1}{n}(K_{13} - \mathbb{E}[K_{13}|A_1, A_3]) \\ &\vdots \\ X_{N,N+N-1} &= \frac{1}{n}(K_{1N} - \mathbb{E}[K_{1N}|A_1, A_N]), \\ &\vdots \\ X_{N,N+n} &= \frac{1}{n}(K_{N-1,N} - \mathbb{E}[K_{N-1,N}|A_{N-1}, A_N]). \end{aligned}$$

That is, $\{X_{Nt}\}$ is the collection of terms of the form

$$\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])$$

for $i = 1, ..., N$ (with $j \neq i$) and

$$\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])$$

for $i = 1, ..., N-1$ and $j = i+1, ..., N$. Using the independence assumptions on $\{A_i\}_{i=1}^N$ and $\{V_{ij}\}_{i<j}$, as well as iterated expectations, it is tedious but straightforward to verify that

$$\mathbb{E}[X_{Nt}|\{X_{Ns}, s \neq t\}] = 0,$$

that is, X_{NT} is a martingale difference sequence (MDS).

Defining the variance of this MDS as

$$\begin{aligned}\sigma_N^2 &\stackrel{def}{\equiv} \mathbb{E}\left(\sum_{t=1}^{T(N)} X_{Nt}\right)^2 \\ &= \sum_{t=1}^{T(N)} \mathbb{V}(X_{Nt}),\end{aligned}$$

we can demonstrate asymptotic normality of its standardized sum – $\frac{1}{\sigma_N}\sum_{t=1}^{T(N)} X_{Nt}$ – by a central limit theorem for martingale difference triangular arrays (see, for example, Hall & Heyde (1980), Theorem 3.2 and Corollary 3.1 and White (2001), Theorem 5.24 and Corollary 5.26). Specifically, if the Lyapunov condition

$$\sum_{t=1}^{T(N)} \mathbb{E}\left(\frac{X_{Nt}}{\sigma_N}\right)^r \to 0 \tag{13}$$

holds for some $r > 2$, and also the stability condition

$$\sum_{t=1}^{T(N)} \left(\frac{X_{Nt}}{\sigma_N}\right)^2 \xrightarrow{p} 1, \tag{14}$$

holds then

$$\begin{aligned}\sum_{t=1}^{T(N)} \frac{X_{Nt}}{\sigma_N} &= \frac{1}{\sigma_N}(T_1 + T_3) \\ &\xrightarrow{D} \mathcal{N}(0,1).\end{aligned} \tag{15}$$

From the calculations used in the MSE analysis of Section 3 we have that

$$\begin{aligned}
\sigma_N^2 &= \mathbb{V}(T_1) + \mathbb{V}(T_3) \\
&= \frac{\mathbb{E}[K_{ij}^2]}{n} + \frac{4\mathbb{V}(\mathbb{E}[K_{ij}|A_i])}{N} + O\left(\frac{1}{n}\right) \\
&= \frac{\Omega_2(w)}{nh} + \frac{4\Omega_1(w)}{N} + O\left(\frac{1}{n}\right) + O\left(\frac{h^2}{N}\right),
\end{aligned}$$

so, taking $r = 3$,

$$\frac{1}{\sigma_N^2} = O(N)$$

assuming $\Omega_1(w) > 0$ and $Nh \geq C > 0$. In the degenerate case, where $\mathbb{V}(\mathbb{E}[K_{ij}|A_i]) = \Omega_1(w) = 0$, we will still have $(\sigma_N)^{-2} = O(nh) = O(N)$ as long as the "knife-edge" $h \propto N^{-1}$ undersmoothing bandwidth sequence is chosen.

To verify the Lyapunov condition (13), note that

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^3 &\leq 8\mathbb{E}\left(\frac{K_{ij}}{n}\right)^3 \\
&= \frac{8}{n^3}\frac{1}{h^3}\int \left[K\left(\frac{w-s}{h}\right)\right]^3 f_W(s)ds \\
&= \frac{8}{n^3h^2}\int [K\left(u\right)]^3 f_W(w-hu)du \\
&= O\left(\frac{1}{n^3h^2}\right)
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
\mathbb{E}\left(\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)^3 &\leq \frac{8^2}{N^3}\mathbb{E}\left(\mathbb{E}[K_{ij}|A_i]\right)^3 \\
&= \frac{8^2}{N^3}\mathbb{E}\left(\int K\left(u\right) f_{W|A}(w-hu|A_i)du\right)^3 \\
&= O\left(\frac{1}{N^3}\right).
\end{aligned} \tag{17}$$

Putting things together we get that

$$\begin{aligned}\sum_{t=1}^{T(N)} \mathbb{E}\left(X_{Nt}\right)^3 =& n\mathbb{E}\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^3 \\ &+ N\mathbb{E}\left(\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)^3 \\ =& O\left(\frac{1}{(nh)^2}\right) + O\left(\frac{1}{N^2}\right) \\ =& O\left(\frac{1}{N^2}\right)\end{aligned}$$

when $Nh \geq C > 0$ for all N. Therefore the Lyapunov condition (13) is satisfied for $r = 3$, since

$$\begin{aligned}\sum_{t=1}^{T(N)} \mathbb{E}\left(\frac{X_{Nt}}{\sigma_N}\right)^3 &= O(N^{3/2}) \cdot O\left(\frac{1}{N^2}\right) \\ &= O\left(\frac{1}{\sqrt{N}}\right) \\ &= o(1).\end{aligned}$$

To verify the stability condition (14), we first rewrite that condition as

$$\begin{aligned}0 &= \lim_{N\to\infty}\left(\frac{1}{\sigma_N^2}\sum_{t=1}^{T(N)}\left(X_{Nt}^2 - \mathbb{E}\left[X_{Nt}^2\right]\right)\right) \\ &= \lim_{N\to\infty}\left(\frac{1}{N\sigma_N^2}\sum_{t=1}^{T(N)}(R_1 + R_2)\right)\end{aligned} \tag{18}$$

where

$$\begin{aligned}R_1 \equiv& N\sum_{i=1}^{N}\left[\left(\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)^2 - \mathbb{E}\left(\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)^2\right] \\ &+ N\sum_{i<j}\left[\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^2 - \mathbb{E}\left[\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^2 \middle| A_i, A_j\right]\right]\end{aligned}$$

and

$$R_2 \equiv N\sum_{i<j}\left[\mathbb{E}\left[\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^2 \middle| A_i, A_j\right] - \mathbb{E}\left[\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^2\right]\right].$$

Since $1/N\sigma_N^2 = O(1)$, the stability condition (18) will hold if R_1 and R_2 both converge to zero in probability.

By the independence restrictions on $\{U_{ij}\}$ and $\{A_i\}$, the (mean zero) summands in R_1 are mutually uncorrelated, so

$$\begin{aligned}
\mathbb{E}\left[R_1^2\right] \equiv N^2 \sum_{i=1}^{N} \mathbb{E}\left[\left(\left(\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)^2 - \mathbb{E}\left(\frac{2}{N}(\mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}])\right)^2\right)^2\right] \\
+ N^2 \sum_{i<j} \mathbb{E}\left[\left(\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^2 - \mathbb{E}\left[\left(\frac{1}{n}(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])\right)^2 \middle| A_i, A_j\right]\right)^2\right] \\
= O\left(\frac{\mathbb{E}\left(\mathbb{E}[K_{ij}|A_i]\right)^4}{N}\right) + O\left(\frac{N^2 \mathbb{E}\left(K_{ij}\right)^4}{n^3}\right).
\end{aligned}$$

But, using analogous arguments to (16) and ((17),

$$\mathbb{E}\left[\mathbb{E}[K_{ij}|A_i]^4\right] = O\left(1\right)$$

and

$$\mathbb{E}\left[K_{ij}^4\right] = O\left(\frac{1}{h^3}\right),$$

so

$$\begin{aligned}
\mathbb{E}\left[R_1^2\right] &= O\left(\frac{1}{N}\right) + O\left(\frac{N^2}{(nh)^3}\right) \\
&= O\left(\frac{1}{N}\right) \\
&= o(1),
\end{aligned}$$

under the bandwidth condition that $1/nh = O(1/N)$. So R_1 converges in probability to zero. Moreover, R_2 is proportional to a (mean zero) second-order U-statistic,

$$\begin{aligned}
R_2 &= \frac{1}{n} \sum_{i<j} \frac{N}{n} \left[\mathbb{E}\left[(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])^2 \middle| A_i, A_j\right] - \mathbb{E}\left[(K_{ij} - \mathbb{E}[K_{ij}|A_i, A_j])^2\right]\right] \\
&\equiv \frac{1}{n} \sum_{i<j} p_N(A_i, A_j),
\end{aligned}$$

with kernel having second moment

$$\begin{aligned}
\mathbb{E}\left[p_N(A_i, A_j)^2\right] &= O\left(\frac{N^2}{n^2} \mathbb{E}\left(\mathbb{E}[K_{ij}^2|A_i, A_j]\right)^2\right) \\
&= O\left(\frac{N^2}{n^2} \cdot \frac{1}{h^2}\right) \\
&= O(1) \\
&= o(N),
\end{aligned}$$

again imposing the bandwidth restriction $1/nh = O(1/N)$. Thus by Lemma 3.1 of Powell et al. (1989), R_2 converges in probability to its (zero) expected value.

Since conditions (13) and (14) both hold, a central limit theorem for martingale difference triangular arrays implies

$$\frac{1}{\sigma_N}(T_1 + T_3) \xrightarrow{D} \mathcal{N}(0,1).$$

A final step is to used this result to obtain the asymptotic distribution of $\hat{f}_W(w)$. Because

$$\frac{1}{\sigma_N} = O\left(\sqrt{N}\right),$$

we have that T_2 and T_4 are asymptotically negligible after standardization with σ_N^{-1} (see Appendix A),

$$\frac{T_2}{\sigma_N} = O_p\left(\sqrt{\frac{N}{n}}\right) = o_p(1)$$

and

$$\frac{T_4}{\sigma_N} = O\left(\sqrt{N}h^2\right) = o(1),$$

so that

$$\begin{aligned}\frac{1}{\sigma_N}\left(\hat{f}_W(w) - f_W(w)\right) &= \frac{1}{\sigma_N}\left(T_1 + T_2 + T_3 + T_4\right) \\ &\xrightarrow{D} \mathcal{N}(0,1).\end{aligned}$$

When $Nh^4 \to 0$ and $Nh \to \infty$,

$$N\sigma_N^2 \to 4\Omega_1(w)$$

and

$$\sqrt{N}\left(\hat{f}_W(w) - f_W(w)\right) \xrightarrow{D} \mathcal{N}(0, 4\Omega_1(w))$$

as long as $\mathbb{V}(\mathbb{E}[K_{ij}|A_i]) > 0$.

Under "knife-edge" bandwidth sequences, such that $Nh \to C > 0$, we have instead that

$$N\sigma_N^2 \to 4\Omega_1(w) + C^{-1}\Omega_2(w)$$

and

$$\sqrt{N}(\hat{f}_W(w) - f_W(w)) \xrightarrow{D} \mathcal{N}(0, 4\Omega_1(w) + C^{-1}\Omega_2(w)).$$

Degeneracy

Degeneracy arises when $\mathbb{V}(\mathbb{E}[K_{ij}|A_i]) = \Omega_1(w) = 0$. In terms of the underlying network generating process (NGP), degeneracy arises when the conditional density of W_{ij} at w

given $A_i = a$ is constant in a (i.e., when $\mathbb{V}\left(f_{W|A}\left(w|\, A_i\right)\right) = 0$).

As a simple example of such an NGP, let A_i equal -1 with probability π and 1 otherwise; next set

$$W_{ij} = A_i A_j + V_{ij}$$

with V_{ij} standard normal. In this case the conditional density $f_{W|A}\left(w|\, A_i\right)$ is the mixture

$$f_{W|A}\left(w|\, A_i\right) = \pi\phi\left(w + A_i\right) + (1-\pi)\,\phi\left(w - A_i\right)$$

with $\phi\left(\cdot\right)$ the standard normal density function. Unconditionally the density is

$$f_W\left(w\right) = \left[\pi^2 + (1-\pi)^2\right]\phi\left(w-1\right) + 2\pi\left(1-\pi\right)\phi\left(w+1\right).$$

Observe that, if $\pi = 1/2$, then $f_{W|A}\left(w|\, A_i = 1\right) = f_{W|A}\left(w|\, A_i = -1\right) = f_W\left(w\right)$ and hence that $\mathbb{V}\left(f_{W|A}\left(w|\, A_i\right)\right) = 0$.[6] Degeneracy arises in this case, even though there is non-trivial dependence across dyads sharing an agent in common. If $\pi \neq 1/2$, then $\mathbb{V}\left(f_{W|A}\left(w|\, A_i\right)\right) > 0$, but one still might worry about "near degeneracy" when π is close to 1/2.

Menzel (2017) shows that under degeneracy, the limit distribution of the sample mean, $\bar{W}$, equation (8) on on page 7 above, may be non-Gaussian. This occurs because (i) the T_1 and T_2 terms in a double projection decomposition of $\bar{W}$ analogous to the one used here for $\hat{f}_W\left(w\right)$ will be of equal order *and* T_2, the Hájek Projection error, may be non-Gaussian (as is familiar from the theory of U-Statistics, e.g., Chapter 12 of van der Vaart (2000)).

The situation is both more complicated and simpler here. In the case of the estimated density $\hat{f}_W\left(w\right)$, if the bandwidth sequence $h = h_N$ satisfies the conditions $Nh \to \infty$ and $Nh^4 \to 0$, then T_2 will be of smaller order than T_1 and hence not contribute to the limit distribution irrespective of whether the NGP is degenerate or not. In particular, under degeneracy the Liaponuv condition (13) continues to hold for $r = 3$ since

$$\sum_{t=1}^{T(N)} E\left(\frac{X_{Nt}}{\sigma_N}\right)^3 = O\left(\frac{1}{\sqrt{nh}}\right)$$

and it follows straightforwardly that $\frac{1}{\sigma_N}\left(\hat{f}_W\left(w\right) - f_W\left(w\right)\right)$ continues to be normal in the limit.

The "knife-edge" undersmoothing bandwidth sequence is primarily of interest because it results in a sequence where both T_1 and T_3 contribute to the limit distribution. In practice this does not mean that the researcher should set $h = h_N \propto N^{-1}$. Based on the theoretical analysis sketched above, we recommend choosing a sequence that tends

[6] Degeneracy also arises when $w = 1$.

to zero slightly faster than mean squared error optimal sequence where $h = h_N \propto n^{-1/5}$.[7]

Under such a sequence we will have

$$\sqrt{N}(\hat{f}_W(w) - f_W(w)) \xrightarrow{D} \mathcal{N}(0, 4\Omega_1(w))$$

under non-degeneracy and

$$\sqrt{nh}(\hat{f}_W(w) - f_W(w)) \xrightarrow{D} \mathcal{N}(0, \Omega_2(w))$$

under degeneracy. Although the rate of convergence of $\hat{f}_W(w)$ to $f_W(w)$ is faster in the case of degeneracy this will not affect inference in practice as long as an appropriate estimate of σ_N is used; that is working directly with $(\hat{f}_W(w) - f_W(w))/\sigma_N$ ensures rate-adaptivity. Note also that, in the absence of degeneracy, the MSE optimal bandwidth sequence could be used. By slightly undersmoothing relative to this sequence, we ensure that the limit distribution remains unbiased in case of degeneracy.

5 Asymptotic variance estimation

To construct Wald-based confidence intervals for $\hat{f}_W(w)$, a consistent estimator of its asymptotic variance is needed. When $Nh \to C < \infty$, the asymptotic variance depends on both

$$\Omega_2(w) \overset{def}{\equiv} f_W(w) \cdot \int [K(u)]^2 \mathrm{d}u$$

and

$$\Omega_1(w) \overset{def}{\equiv} \mathbb{V}\left(f_{W|A}(w|A_i)\right).$$

In this section we present consistent estimators for both of these terms.

A simple estimator of $\Omega_2(w)$ is

$$\tilde{\Omega}_2(w) = \frac{h}{n} \sum_{i<j} K_{ij}^2, \tag{19}$$

the consistency of which we demonstrate in Appendix A:

$$\tilde{\Omega}_2(w) \xrightarrow{p} \Omega_2(w). \tag{20}$$

The estimator $\tilde{\Omega}_2(w)$ uses the second moment of K_{ij} instead of its sample variance to estimate $\Omega_2(w)$; in practice we recommend, similar to Newey (1994) in the context of

[7] In practice "plug-in" bandwidths that would be appropriate in the absence of any dyadic dependence across the $\{W_{ij}\}_{i<j}$ might work well; although this remains an unexplored conjecture.

monadic kernel-based estimation, the less conservative alternative:

$$\begin{aligned}\hat{\Omega}_2(w) &\equiv h\left(\left(\frac{1}{n}\sum_{i<j}K_{ij}^2\right)-\left(\hat{f}_W(w)\right)^2\right)\\ &= h\left(\frac{1}{n}\sum_{i<j}\left(K_{ij}-\hat{f}_W(w)\right)^2\right)\\ &= \tilde{\Omega}_1(w)+o_p(1)\\ &= \Omega_1(w)+o_p(1).\end{aligned}$$

We next turn to estimation of

$$\Omega_1(w)=\mathbb{V}\left(f_{W|A}(w|A_1)\right)=\lim_{N\to\infty}\mathbb{C}(K_{ij},K_{ij})$$

where $i\neq k$. A natural sample analog estimator, following a suggestion by Graham (TBD) in the context of parametric dyadic regression, involves an average over the three indices i, j, and k:

$$\begin{aligned}\hat{\Omega}_1(w) &\equiv \frac{1}{N(N-1)(N-2)}\sum_{i\neq j\neq k}(K_{ij}-\hat{f}_W(w))(K_{ik}-\hat{f}_W(w))\\ &\equiv \binom{N}{3}^{-1}\sum_{i<j<k}S_{ijk}-\hat{f}_W(w)^2,\end{aligned}$$

for $S_{ijk}=\frac{1}{3}\left(K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}\right)$.[8] In Appendix A we show that

$$\hat{\Omega}_1(w)\xrightarrow{p}\Omega_1(w). \tag{21}$$

Inserting these estimators, $\hat{\Omega}_1(w)$ and $\hat{\Omega}_2(w)$, into the formula for the variance of $\hat{f}_W(w)$ yields a variance estimate of

$$\hat{\sigma}_N^2=\frac{1}{nh}\hat{\Omega}_2(w)+\frac{2(N-2)}{n}\hat{\Omega}_1(w). \tag{22}$$

[8]See also the variance estimator for density presented in Holland & Leinhardt (1976).

We end this section by observing that the following equality holds

$$\begin{aligned}\hat{\sigma}_N^2 =& \frac{1}{n^2}\sum_{i<j}\left(K_{ij}-\hat{f}_W(w)\right)^2 \\ &+ \frac{2(N-2)}{n}\left(\frac{1}{N(N-1)(N-2)}\sum_{i\neq j\neq k}(K_{ij}-\hat{f}_W(w))(K_{ik}-\hat{f}_W(w))\right) \\ =& \frac{1}{n^2}\left(\sum_{i<j}\sum_{k<l}d_{ijkl}(K_{ij}-\hat{f}_W(w))(K_{kl}-\hat{f}_W(w))\right),\end{aligned}$$

where

$$d_{ijkl} = 1\{i=j, k=l, i=l, \text{or } j=k\}.$$

As Graham (TBD) notes, this coincides with the estimator for

$$\mathbb{V}(\bar{W}) = \mathbb{V}\left(\frac{1}{n}\sum_{i<j}W_{ij}\right)$$

proposed by Fafchamps & Gubert (2007), replacing "$W_{ij}-\bar{W}$" with "$K_{ij}-\bar{K}$", with $\bar{K} \stackrel{def}{\equiv} \hat{f}_W(w)$ (see also Holland & Leinhardt (1976), Cameron & Miller (2014) and Aronow et al. (2017)). Our variance estimator can also be viewed as a dyadic generalization of the variance estimate proposed by Newey (1994) for "monadic" kernel estimates.

6 Simulation study

Our simulations design is based upon the example used to discuss degeneracy in Section 4. As there we let A_i equal -1 with probability π and 1 otherwise. We generate W_{ij}

$$W_{ij} = A_iA_j + V_{ij}$$

with $V_{ij} \sim \mathcal{N}(0,1)$. We set $\pi = 1/3$ and estimate the density $f_W(w)$ at $w = 1.645$.

We present results for three sample sizes: $N = 100, 400$ and $1,600$. These sample sizes are such that, for a "sufficiently non-degenerate" NGP, the standard error of $\hat{f}_W(w)$ would be expected to decline by a factor of 1/2 for each increase in sample size (if the bandwidth is large enough to ensure that the $\frac{\Omega_2(w)}{nh}$ variance term is negligible relative to the $\frac{2\Omega_1(w)(N-2)}{n} \approx \frac{4\Omega_1(w)}{N}$ one). We set the bandwidth equal to the MSE-optimal one presented in equation (7) above. This is an 'oracle' bandwidth choice. Developing feasible data-based methods of bandwidth selection would be an interesting topic for future research.

Table 1 presents the main elements of each simulation design. Panel B of the table lists "pencil and paper" bias and asymptotic standard error calculations based upon the

Table 1: Monte Carlo Designs

N	100	400	1,600
Panel A: Design & Bandwidth			
π	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
w	1.645	1.645	1.645
$h_N^*(w)$	0.2496	0.1431	0.0822
Panel B: Theoretical Sampling Properties			
$h^2 B(w)$	-0.0033	-0.0011	-0.0004
$\text{ase}\left(\hat{f}_W(w)\right) = \sqrt{\frac{2\Omega_1(w)(N-2)}{n} + \frac{\Omega_2(w)}{nh}}$	0.0117	0.0053	0.0025
$\text{ase}(T_3) = \sqrt{\frac{2\Omega_1(w)(N-2)}{n}}$	0.0098	0.0049	0.0025
$\text{ase}(T_1) = \sqrt{\frac{\Omega_2(w)}{nh}}$	0.0065	0.0021	0.0007

NOTES: Rows 1 through 3 list the basic Monte Carlo design and bandwidth parameter choices. The bandwidths coincide with the MSE optimal one given in equation (7). Panel B gives pencil and paper calculations for the bias of $\hat{f}_W(w)$, as well as its asymptotic standard error (ase), based upon, respectively, equations (3) and (4) in Section 3. The asymptotic standard errors of T_1 and T_3, as defined in Section 4, are also separately given.

expressions presented in Section 3 above. Panel B also presents analytic estimates of the standard deviations of the T_1 and T_3 terms in the decomposition of $\hat{f}_W(w)$ used to derive its limit distribution. In the given designs both terms of are similar magnitude despite the fact that the contribution of the T_1 term is asymptotically negligible in theory.

Table 2 summarizes the results of 1,000 Monte Carlo simulations. The median bias and standard deviation of our density estimates across the Monte Carlo replications closely track our theoretical predictions (compare rows 1 and 2 of Table 2 with Rows 1 and 2 of Panel B of Table 1. Row 3 of the table reports the median "Fafchamps and Gubert" asymptotic standard error estimate. This standard error estimate is generally larger than its asymptotic counterpart. Consequently the coverage of confidence intervals based upon it is conservative (Row 5). The degree of conservatism is declining in sample size, suggesting that – as expected – the "Fafchamps and Gubert" asymptotic standard error estimate is closer to its theoretical counterpart as N grows. Row 4 of the table reports the coverage of confidence intervals based upon standard errors which ignore the presence of dyadic dependence; these intervals – as expected – fail to cover the true density frequently enough.

The simulations suggest, for the designs considered, that the asymptotic theory presented in Sections 3 and 4 provides an accurate approximation of finite sample behavior. Our variance estimate is a bit conservative for the designs considered; whether this is peculiar to the specific design considered or a generic feature of the estimate is unknown.[9] As with bandwidth selection, further exploration of methods of variance estimation in

[9]We observe that our variance estimate implicitly includes an estimate of the variance of T_2, which is negligible in the limit.

Table 2: Monte Carlo Results

N	100	400	$1,600$
median bias	-0.0028	-0.0010	-0.0006
standard deviation	0.0112	0.0051	0.0025
median $\hat{\text{ase}}\left(\hat{f}_W(w)\right)$	0.0173	0.0068	0.0028
coverage (iid)	0.678	0.551	0.390
coverage (FG)	0.995	0.987	0.967

NOTES: A robust measure of the standard deviation of $\hat{f}_W(w)$ is reported in row 2. It equals the difference between the 0.95 and 0.05 quantiles of the Monte Carlo distribution of $\hat{f}_W(w)$ divided by 2×1.645. Row 4 reports the coverage of a nominal 95 percent Wald-based confidence interval that ignores the presence of dyadic dependence. Row 5 reports the coverage properties of a nominal 95 percent Wald-based confidence interval that uses the Fafchamps & Gubert (2007) variance estimate discussed in Section 4.

the presence of dyadic dependence is warranted.

7 Extensions

There are a number of avenues for extension or modification of the simple results for scalar density estimation presented above. One variant of these results would apply when the dyadic variable W_{ij} lacks the idiosyncratic component V_{ij}, i.e., when

$$W_{ij} = W(A_i, A_j),$$

for $\{A_i\}$ an i.i.d. sequence. This case arises when W_{ij} is a measure of "distance" between the attributes of nodes i and j, for example,

$$W_{ij} = \sqrt{(A_i - A_j)^2},$$

for A_i a scalar measure of "location" for agent i. The asymptotic distribution of $\hat{f}_W(w)$ derived above should be applicable to this case as long as the conditional density function $f_{W|A}(w|a)$ of W_{ij} given A_i is well-defined, which would be implied if A_i has a continuously-distributed component given its remaining component (if any) and the function $W(\cdot)$ is continuously differentiable in that component. In the decomposition of $\hat{f}_W(w) - f_W(w)$ for this case, the term corresponding to T_1 would be identically zero (as would $\Omega_2(w)$), but the T_2 term could still be shown to be asymptotically negligible using Lemma 3.1 of Powell et al. (1989) as long as $Nh \to \infty$.

Another straightforward extension of this analysis would be to directed dyadic data, where W_{ij} is observed for all pairs of indices with $i \neq j$ and $W_{ij} \neq W_{ji}$ with positive probability. The natural generalization of the data generation process would be

$$W_{ij} = W(A_i, B_j, V_{ij}),$$

with $\{A_i\}$, $\{B_j\}$, and $\{V_{ij}\}$ mutually independent and i.i.d. with $V_{ij} \neq V_{ji}$ in general. Here the conditional densities

$$f_{W|A}(w|a) = \mathbb{E}[f_{W|AB}(w|A_i = a, B_j)]$$

and

$$f_{W|B}(w|b) = \mathbb{E}[f_{W|AB}(w|A_i, B_j = b)]$$

will differ, and the asymptotic variance of $\hat{f}_W(w)$ will depend upon

$$\Omega_1(w) = \mathbb{V}\left(\frac{1}{2}\left(f_{W|A}(w|A_i) + f_{W|B}(w|B_i)\right)\right)$$

in a way analogous to how $\Omega_1(w)$, defined earlier, does in the undirected case analyzed in this paper.

Yet another generalization of the results would allow W_{ij} to be a p-dimensional jointly-continuous W_{ij} random vector. The estimator

$$\hat{f}_W(w) = \frac{1}{n}\sum_{i=1}^{N-1}\sum_{j=1+1}^{N}\frac{1}{h^p}K\left(\frac{w - W_{ij}}{h}\right)$$

of the p-dimensional density function $f_W(w)$ will continue to have the same form as derived in the scalar case, provided $Nh^p \to \infty$ (or $Nh^p \to C > 0$) as long as the relevant bias term T_4 is negligible. If the density is sufficiently smooth and $K(\cdot)$ is a "higher-order kernel" with

$$\int K(u)\mathrm{d}u = 1,$$
$$\int u_1^{j_1} u_2^{j_2} ... u_p^{j_p} K(u)\mathrm{d}u = 0 \qquad \text{for } j_i \in \{0, ..., q\} \text{ with } \sum_{i=1}^{p} j_i < q,$$

then the bias term T_4 will satisfy

$$\begin{aligned} T_4 &\equiv \mathbb{E}\left[\hat{f}_W(w)\right] - f_W(w) \\ &= O(h^q). \end{aligned}$$

As long as q can be chosen large enough so that $Nh^{2q} \to 0$ while $Nh^p \geq C > 0$, the bias term T_4 will be asymptotically negligible and the density estimator $\hat{f}_W(w)$ should still be asymptotically normal with asymptotic distribution of the same form derived above.

Finally, a particularly useful extension of the kernel estimation approach for dyadic data would be to estimation of the conditional expectation of one dyadic variable Y_{ij} conditional on the value w of another dyadic variable W_{ij}, i.e., estimation of

$$g(w) \equiv \mathbb{E}[Y_{ij}|W_{ij} = w]$$

when the vector W_{ij} has p jointly-continuously distributed components conditional upon any remaining components. Here the Nadaraya-Watson kernel regression estimator (Nadaraya, 1964; Watson, 1964) would be defined as

$$\hat{g}(w) \equiv \frac{\sum_{i \neq j} K\left(\frac{w - W_{ij}}{h}\right) Y_{ij}}{\sum_{i \neq j} K\left(\frac{w - W_{ij}}{h}\right)},$$

and the model for the dependent variable Y_{ij} would be analogous to that for W_{ij}, with

$$\begin{aligned} Y_{ij} &= Y(A_i, B_j, U_{ij}) \\ W_{ij} &= W(A_i, B_j, V_{ij}) \end{aligned}$$

in the directed case (and $B_j \equiv A_j$ for undirected data), with $\{A_i\}$, $\{B_j\}$, and $\{(U_i, V_{ij})\}$ assumed mutually independent and i.i.d. The large-sample theory would treat the numerator of $\hat{g}(w)$ similarly to that for the denominator (which is proportional to the kernel density estimator $\hat{f}_W(w)$); our initial calculations for undirected data with a scalar, continuously-distributed regressor W_{ij} yield

$$\sqrt{N}\left(\hat{g}(w) - g(w)\right) \xrightarrow{D} \mathcal{N}(0, 4\Gamma_1(w)),$$

when $Nh^p \to \infty$ and $Nh^4 \to 0$, where

$$\Gamma_1(w) \equiv \mathbb{V}\left(\frac{\mathbb{E}[Y_{ij}|A_i, W_{ij} = w] \cdot f_{W|A}(w|A_i)}{f_W(w)}\right).$$

If this calculation is correct, then, like the density estimator $\hat{f}_W(w)$ the rate of convergence for the estimator $\hat{g}(w)$ of the conditional mean $g(w)$ would be the same as the rate for the estimator $\hat{\mu}_Y = \bar{Y}$ of the unconditional expectation $\mu_y = \mathbb{E}[Y_{ij}] = \mathbb{E}[g(W_{ij})]$, in contrast to the estimation using i.i.d. (monadic) data. We intend to verify these calculations and derive the other extensions in future work.

A Proofs

Derivation of bias expression, equation (3) of the main text

Under the conditions imposed in the main text, the expected value of $\hat{f}_W(w)$ is

$$\begin{aligned}
\mathbb{E}\left[\hat{f}_W(w)\right] &= \mathbb{E}\left[\frac{1}{h}K\left(\frac{w-W_{12}}{h}\right)\right] \\
&= \mathbb{E}\left[\int \frac{1}{h}K\left(\frac{w-s}{h}\right) f_W(s)\mathrm{d}s\right] \\
&= \int K\left(u\right) f_W(w-hu)\mathrm{d}u \\
&= f_W(w) + \frac{h^2}{2}\frac{\partial^2 f_W(w)}{\partial w^2}\int u^2 K\left(u\right)\mathrm{d}u + o(h^2) \\
&\equiv f_W(w) + h^2 B(w) + o(h^2).
\end{aligned}$$

The first line in this calculation follows from the fact that W_{ij} is identically distributed for all i, j, the third line uses the change-of-variables $s = w - hu$, and the fourth line follows from a second-order Taylor's expansion of $f_W(w - hu)$ around $h = 0$ and the fact that

$$\int u \cdot K(u)\mathrm{d}u = 0$$

because $K(u) = K(-u)$.

Demonstration of asymptotic negligibility of T_2 and T_4

Equation (10), which defines T_2, involves averages of the random variables

$$\begin{aligned}
\mathbb{E}[K_{ij}|A_i, A_j] &= \int \frac{1}{h}K\left(\frac{w-s}{h}\right) f_{W|AA}(s|A_i, A_j)\mathrm{d}s \\
&= \int K\left(u\right) f_{W|AA}(w-hu|A_i, A_j)\mathrm{d}u
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[K_{ij}|A_i] &= \int \frac{1}{h}K\left(\frac{w-s}{h}\right) f_{W|A}(s|A_i)\mathrm{d}s \\
&= \int K\left(u\right) f_{W|A}(w-hu|A_i)\mathrm{d}u
\end{aligned}$$

which are both assumed bounded, so T_2 can be written, after some re-arrangement, as the degenerate second-order U-statistic,

$$T_2 = \frac{1}{n}\sum_{i<j}\left(\mathbb{E}[K_{ij}|A_i, A_j] - \mathbb{E}[K_{ij}|A_i] - \mathbb{E}[K_{ij}|A_j] + \mathbb{E}[K_{ij}]\right)$$

with all summands uncorrelated. This implies, squaring and taking expectations, that

$$\begin{aligned}
\mathbb{E}[T_2^2] &= \frac{1}{n^2}\sum_{i<j}\mathbb{E}[(\mathbb{E}[K_{ij}|A_i,A_j]-\mathbb{E}[K_{ij}|A_i]-\mathbb{E}[K_{ij}|A_j]+\mathbb{E}[K_{ij}])^2] \\
&\leq \frac{5}{n}\mathbb{E}[(\mathbb{E}[K_{ij}|A_i,A_j])^2 \\
&= O\left(\frac{1}{n}\right),
\end{aligned}$$

so

$$T_2 = O_p\left(\frac{1}{\sqrt{n}}\right) = O_p\left(\frac{1}{N}\right).$$

Turning to the fourth term, defined in equation (12), we demonstrated in Section 3 that

$$T_4 = h^2 B(w) + o(h^2) = O(h^2).$$

Demonstration of consistency of $\hat{\Omega}_2(w)$, equation (20) of the main text.

To show result (20) of the main text, we start by showing asymptotic unbiasedness of $\tilde{\Omega}_2(w)$ for $\Omega_2(w)$. The expected value of the summands in (19) equal

$$\begin{aligned}
\mathbb{E}\left[(K_{12})^2\right] &= \frac{1}{h}\int [K(u)]^2 f_W(w-hu)\mathrm{d}u \\
&= \frac{f_W(w)}{h}\cdot\int [K(u)]^2\mathrm{d}u + O(1) \\
&\equiv \frac{1}{h}\Omega_2(w) + O(1) \\
&= O\left(\frac{1}{h}\right),
\end{aligned}$$

from which asymptotic unbiasedness follows, since:

$$\begin{aligned}
\mathbb{E}\left[\tilde{\Omega}_2(w)\right] &= h\left[\frac{1}{h}\Omega_2(w) + O(1)\right] \\
&= \Omega_2(w) + o(1).
\end{aligned}$$

Following the same logic used to calculate the variance of $\hat{f}_W(w)$, we calculate the variance of $\tilde{\Omega}_2(w)$ as

$$\begin{aligned}
\mathbb{V}\left(\tilde{\Omega}_2(w)\right) &= \mathbb{V}\left(\frac{h}{n}\sum_{i<j} K_{ij}^2\right) \\
&= \left(\frac{h}{n}\right)^2 \sum_{i<j}\sum_{k<l} \mathbb{C}(K_{ij}^2, K_{kl}^2) \\
&= \frac{h^2}{n}\left[\mathbb{V}(K_{12}^2) + 2(N-2)\cdot \mathbb{C}(K_{12}^2, K_{13}^2)\right].
\end{aligned}$$

The first term in this expression depends upon

$$\begin{aligned}
\mathbb{V}(K_{12}^2) &= \mathbb{E}\left[K_{12}^4\right] - \mathbb{E}\left[K_{12}^2\right]^2 \\
&= \frac{f_W(w)}{h^3}\cdot \int [K(u)]^4 \mathrm{d}u + O\left(\frac{1}{h^2}\right) - \mathbb{E}\left[K_{12}^2\right]^2 \\
&= O\left(\frac{1}{h^3}\right),
\end{aligned}$$

while the second involves

$$\begin{aligned}
\mathbb{C}(K_{12}^2, K_{13}^2) &= \mathbb{E}[K_{12}^2 K_{13}^2] - \mathbb{E}\left[K_{12}^2\right]^2 \\
&= \frac{1}{h^2}\mathbb{E}\left[\int [K(u_1)]^2 f_{W|A}(w - hu_1|A_1)\mathrm{d}u_1\right. \\
&\quad \left.\cdot \int [K(u_2)]^2 f_{W|A}(w - hu_2|A_1)\mathrm{d}u_2\right] - \mathbb{E}\left[K_{12}^2\right]^2 \\
&= O\left(\frac{1}{h^2}\right).
\end{aligned}$$

Putting things together we have that

$$\begin{aligned}
\mathbb{V}\left(\tilde{\Omega}(w)\right) &= \frac{h^2}{n}\left[\mathbb{V}(K_{12}^2) + 2(N-2)\cdot \mathbb{C}(K_{12}^2, K_{13}^2)\right] \\
&= \frac{h^2}{n}\left[O\left(\frac{1}{h^3}\right) + 2(N-2)\cdot O\left(\frac{1}{h^2}\right)\right] \\
&= O\left(\frac{1}{nh}\right) + O\left(\frac{1}{N}\right) \\
&= o(1),
\end{aligned}$$

which, with convergence of the bias of $\tilde{\Omega}_2(w)$ to zero, establishes (20) of the main text.

Demonstration of consistency of $\hat{\Omega}_1(w)$, equation (21) of the main text.

Since $\hat{f}_W(w)$ is consistent if $Nh^4 \to 0$ and $Nh \geq C > 0$, consistency of $\hat{\Omega}_1(w)$ depends on the consistency of

$$\hat{\mathbb{E}}[K_{12}K_{13}] \equiv \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \sum_{i<j<k} S_{ijk}$$

for $\lim_{N\to\infty} \mathbb{E}[K_{12}K_{13}]$. By the fact that $K_{ij} = K_{ji}$, the expected value of $\hat{\mathbb{E}}[K_{12}K_{13}]$ is

$$\begin{aligned}
\mathbb{E}[S_{ijk}] &= \mathbb{E}\left[\frac{1}{3}\left(K_{ij}K_{ik} + K_{ij}K_{jk} + K_{ik}K_{jk}\right)\right] \\
&= \mathbb{E}\left[K_{12}K_{13}\right] \\
&= \mathbb{E}\left[\int \left[K\left(u_1\right)\right] f_{W|A}(w - hu_1|A_1)\mathrm{d}u_1\right. \\
&\left.\cdot \int \left[K\left(u_2\right)\right] f_{W|A}(w - hu_2|A_1)\mathrm{d}u_2\right] \\
&= \mathbb{E}\left[f_{W|A}(w|A_1)^2\right] + o(1)
\end{aligned}$$

from the calculations in Section 3 above. To bound the variance of $\hat{\mathbb{E}}[K_{12}K_{13}]$, we note that, although $\hat{\mathbb{E}}[K_{12}K_{13}]$ is not a U-statistic, it can be approximated by the third-order U-statistic

$$U_N \equiv \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \sum_{i<j<k} p_N(A_i, A_j, A_k),$$

where the kernel $p_N(\cdot)$ is

$$\begin{aligned}
p_N(A_i, A_j, A_k) &= \mathbb{E}[S_{ijk}|A_i, A_j, A_k] \\
&= \frac{1}{3}\left(\kappa_{ijk} + \kappa_{jik} + \kappa_{kij}\right),
\end{aligned}$$

for

$$\begin{aligned}
\kappa_{ijk} &\equiv \mathbb{E}[K_{ij}K_{ik}|A_i, A_j, A_k] \\
&= \int\int \frac{1}{h^2}\left[K\left(\frac{w - s_1}{h}\right)\right] \cdot \left[K\left(\frac{w - s_2}{h}\right)\right] \\
&\cdot f_{W|AA}(s_1|A_i, A_j) f_{W|AA}(s_2|A_i, A_k)\mathrm{d}s_1\mathrm{d}s_2 \\
&= \int \left[K\left(u_1\right)\right] f_{W|AA}(w - hu_1|A_i, A_j)\mathrm{d}u_1 \\
&\cdot \int \left[K\left(u_2\right)\right] f_{W|AA}(w - hu_2|A_i, A_k)\mathrm{d}u_2.
\end{aligned}$$

The difference between $\hat{\mathbb{E}}[K_{12}K_{13}]$ and U_N is

$$\hat{\mathbb{E}}[K_{12}K_{13}] - U_N \equiv \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \sum_{i<j<k} \left(S_{ijk} - \mathbb{E}\left[S_{ijk}|A_i, A_j, A_k\right]\right),$$

and the independence of $\{V_{ij}\}$ and $\{A_i\}$ across all i and j implies that all terms in this summation have expectation zero and are mutually uncorrelated with common second moment, so that

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\mathbb{E}}[K_{12}K_{13}] - U_N\right)^2\right] &\equiv \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \mathbb{E}[(S_{123} - \mathbb{E}\left[S_{123}|A_1, A_2, A_3\right])^2] \\ &\leq \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \mathbb{E}[S_{123}^2].\end{aligned}$$

But

$$\begin{aligned}\mathbb{E}[(S_{123})^2] &= \mathbb{E}\left[\frac{1}{3}\left(K_{12}K_{13} + K_{12}K_{23} + K_{13}K_{23}\right)\right]^2 \\ &= \frac{1}{9}\left(3\mathbb{E}\left[(K_{12}K_{13})^2\right] + 6\mathbb{E}[K_{12}^2 K_{13}K_{23}]\right),\end{aligned}$$

where

$$\mathbb{E}\left[(K_{12}K_{13})^2\right] = O\left(\frac{1}{h^2}\right),$$

from previous calculations demonstrating consistency of $\Omega_2(w)$, and

$$\begin{aligned}\mathbb{E}[K_{12}^2 K_{13}K_{23}] &= \mathbb{E}\left[\int\int\int \frac{1}{h^4}\left[K\left(\frac{w-s_1}{h}\right)\right]^2 \cdot \left[K\left(\frac{w-s_2}{h}\right)\right] \cdot \left[K\left(\frac{w-s_2}{h}\right)\right]\right. \\ &\quad \left. \cdot f_{W|AA}(s_1|A_1, A_2) f_{W|AA}(s_2|A_1, A_3) f_{W|AA}(s_2|A_1, A_3) \mathrm{d}s_1 \mathrm{d}s_2 \mathrm{d}s_3\right] \\ &= \frac{1}{h}\mathbb{E}\left[\int \left[K\left(u_1\right)\right]^2 f_{W|AA}(w - hu_1|A_1, A_2)\mathrm{d}u_1\right. \\ &\quad \cdot \int K\left(u_2\right) f_{W|AA}(w - hu_2|A_1, A_3)\mathrm{d}u_2 \\ &\quad \left. \cdot \int K\left(u_2\right) f_{W|AA}(w - hu_2|A_1, A_3)\mathrm{d}u_2\right] \\ &= O\left(\frac{1}{h}\right).\end{aligned}$$

These results generate the inequality

$$
\begin{aligned}
\mathbb{E}\left[\left(\hat{\mathbb{E}}[K_{12}K_{13}] - U_N\right)^2\right] &\leq \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \mathbb{E}[(S_{123})^2] \\
&= \begin{pmatrix} N \\ 3 \end{pmatrix}^{-1} \left(O\left(\frac{1}{h^2}\right) + O\left(\frac{1}{h}\right)\right) \\
&= O\left(\frac{1}{N(Nh)^2}\right) \\
&= o(1).
\end{aligned}
$$

Finally, we note that U_N is a third-order "smoothed" U-statistic with kernel

$$
p_N(A_i, A_j, A_k) = \frac{1}{3}\left(\kappa_{ijk} + \kappa_{jik} + \kappa_{kij}\right)
$$

satisfying

$$
\mathbb{E}\left[(p_N(A_i, A_j, A_k))^2\right] = O(1)
$$

by the assumed boundedness of $K(u)$ and the conditional density $f_{W|AA}(w|A_i, A_j)$. Therefore, by Lemma A.3 of Ahn & Powell (1993),

$$
\begin{aligned}
U_n - \mathbb{E}[U_N] &= U_N - \mathbb{E}[S_{ijl}] \\
&= U_N - \mathbb{E}\left[f_{W|A}(w|A_1)\right]^2 + o(1) \\
&= o_p(1).
\end{aligned}
$$

Finally, combining all the previous calculations, we get

$$
\begin{aligned}
\hat{\Omega}_1(w) &= \hat{\mathbb{E}}[K_{12}K_{13}] - \left(\hat{f}_W(w)\right)^2 \\
&= \left(\hat{\mathbb{E}}[K_{12}K_{13}] - U_N\right) + (U_N - \mathbb{E}\left[f_{W|A}(w|A_1)\right]^2) + \mathbb{E}\left[f_{W|A}(w|A_1)\right]^2 \\
&- \left(\left(\hat{f}_W(w)\right)^2 - (f_W(w))^2\right) - (f_W(w))^2 \\
&= E\left[f_{W|A}(w|A_1)\right]^2 - (f_W(w))^2 + o_p(1) \\
&\equiv \Omega_1(w) + o_p(1),
\end{aligned}
$$

as claimed.

References

Ahn, H. & Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2), 3 – 29.

A

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4), 581 – 598. 2

Aronow, P. M., Samii, C., & Assenova, V. A. (2017). Cluster-robust variance estimation for dyadic data. *Political Analysis*, 23(4), 564 – 577. 2, 5

Atalay, E., Hortaçsu, A., Roberts, J., & Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13), 5199 – 5202. 1

Bickel, P. J., Chen, A., & Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics*, 39(5), 2280 – 2301. 1

Cameron, A. C. & Miller, D. L. (2014). *Robust inference for dyadic data*. Technical report, University of California - Davis. 2, 5

Cattaneo, M., Crump, R., & Jansson, M. (2014). Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory*, 30(1), 176 – 200. 4

Christakis, N. A., Fowler, J. H., Imbens, G. W., & Kalyanaraman, K. (2010). *An empirical model for strategic network formation*. NBER Working Paper 16039, National Bureau of Economic Research. 1

Diaconis, P. & Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, 28(1), 33 – 61. 1

Fafchamps, M. & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2), 326 – 350. 2, 5, 2

Goldenberg, A., Zheng, A., Fienberg, S. E., & Airoldi, E. M. (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2), 129–333. 1

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4), 1033 – 1063. 4

Graham, B. S. (TBD). *Handbook of Econometrics*, volume 7, chapter The econometric analysis of networks. North-Holland: Amsterdam. 1, 5, 5

Hall, P. & Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. San Diego: Academic Press. 4

Holland, P. W. & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7, 1 – 45. (document), 1, 3, 8, 5

Hoover, D. N. (1979). *Relations on probability spaces and arrays of random variables.* Technical report, Institute for Advanced Study, Princeton, NJ. 2

König, M. D., Liu, X., & Zenou, Y. (2019). R&d networks: theory, empirics and policy implications. *Review of Economics and Statistics*, 101(3), 476 – 491. 1

Lovász, L. (2012). *Large Networks and Graph Limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society. 1

Menzel, K. (2017). *Bootstrap with clustering in two or more dimensions.* Technical Report 1703.03043v2, arXiv. 1, 2, 3, 4

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9(1), 141 – 142. 7

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(2), 233 – 253. (document), 3, 5, 5

Nowicki, K. (1991). Asymptotic distributions in random graphs with applications to social networks. *Statistica Neerlandica*, 45(3), 295 – 325. 2

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3), 1065 – 1076. (document), 1, 2

Powell, J. L. (1994). *Handbook of Econometrics*, volume 4, chapter Estimation of semiparametric models, (pp. 2443 – 2521). North-Holland: Amsterdam. 1, 3

Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of weighted average derivatives. *Econometrica*, 57(6), 1403 – 1430. 4, 7

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3), 832 – 837. (document), 1, 2

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Boca Raton: Chapman and Hall. 2

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy.* New York: Twentieth Century Fund. 1

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge: Cambridge University Press. 4, 4

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A.*, 26(4), 359 – 372. 7

White, H. (2001). *Asymptotic Theory for Econometricians.* San Diego: Academic Press. 4