Empirical Likelihood and Uniform Convergence Rates for Dyadic Kernel Density Estimation

Harold D. Chiang and Bing Yang Tan

ABSTRACT
This article studies the asymptotic properties of and alternative inference methods for kernel density estimation (KDE) for dyadic data. We first establish uniform convergence rates for dyadic KDE. Second, we propose a modified jackknife empirical likelihood procedure for inference. The proposed test statistic is asymptotically pivotal regardless of presence of dyadic clustering. The results are further extended to cover the practically relevant case of incomplete dyadic data. Simulations show that this modified jackknife empirical likelihood-based inference procedure delivers precise coverage probabilities even with modest sample sizes and with incomplete dyadic data. Finally, we illustrate the method by studying airport congestion in the United States.

1. Introduction

Consider a weighted undirected random graph consisting of vertices $i \in \{1, \ldots, n\}$ and edges consisting of real-valued random variables $(X_{ij})_{i,j \in \{1, \ldots, n\}}$. Suppose $X_{ij}$ are identically distributed with density $f(x) = f_{X_{12}}(x)$. The dyadic kernel density estimator (KDE) of Graham et al. (2019), evaluated at a design point $x \in \text{supp}(X_{12})$, is defined as

$$\hat{f}_n(x) = \left(\frac{n}{2}\right)^{-1} \sum_{1 \leq i < j \leq n} 1_{hn}(x - X_{ij}) \cdot K \left(\frac{x - X_{ij}}{hn}\right).$$

Applications of dyadic data are numerous in different fields of the sciences and social sciences, such as bilateral trade, animal migrations, refugee diasporas, transportation networks between different cities, friendship networks between individuals, intermediate product sales between firms, research and development between universities, social networks between legislators, to list a few. See, for example, Graham (2019) for a recent and comprehensive review. This paper investigates the asymptotic properties of and proposes new, improved inference methods for dyadic KDE both complete and incomplete data.

The main contributions of this article are two-fold. First, we establish uniform convergence rates for dyadic KDE under general conditions. These rates are new; no uniform rate is previously known for KDE under dyadic sampling in the literature. Second, we develop inference methods by adapting a modified version of the jackknife empirical likelihood (JEL) proposed by Matsushita and Otsu (2021). The proposed modified JEL (mJEL) procedure is shown to be asymptotically valid regardless of the presence of dyadic clustering. In practice, dyadic clustered data are often incomplete or contain missing values. We extend our modified JEL procedure to cover the practically relevant case of incomplete data under the missing at random assumption. Extensive simulation studies show that this modified JEL inference procedure does not suffer from the unreliable finite sample performance of the analytic variance estimator, delivers precise coverage probabilities even under modest sample sizes, and is robust against both dyadic clustering and iid sampling for both complete and incomplete dyadic data. Finally, we illustrate the method by studying the distribution of congestion across different flight routes using data from the United States Department of Transportation (U.S. DOT).

In an important recent work, Graham et al. (2019) first propose and examine a nonparametric density estimator for dyadic random variables. Focusing on pointwise asymptotic behaviors, they demonstrate that asymptotic normality at a fixed design point can be established at a parametric rate with respect to number of vertices, an unique feature of dyadic KDE. This implies that its asymptotics are robust to a range of bandwidth rates. They also point out that degeneracy that leads to a non-Gaussian limit, such as the situation pointed out in Menzel (2021), does not occur for dyadic KDE under some mild bandwidth conditions. For inference, they propose an analytic variance estimator that corresponds to the one proposed by Fafchamps and Gubert (2007) (henceforth FG) under parametric models and establish its asymptotic validity. In their simulation, they discover that the FG variance estimator often leads to imprecise coverage with moderate sample sizes. Our proposed modified JEL procedure for dyadic KDE complements the findings of Graham et al. (2019) by providing an alternative that has improved finite sample performance.
In practice, the presence of missing edges is an important feature of network datasets, and a researcher working with network data often observes incomplete sets of edges. To our knowledge, none of the existing theoretical work in the literature that focuses on the asymptotics of dyadic KDE has tackled this issue. Under the missing at random assumption, we extend our theory for modified JEL to cover this practically relevant situation.

Ever since the seminal papers of Owen (1988, 1990), empirical likelihood has been extensively studied in the literature. For a textbook treatment of classical theory, see Owen (2001). Some recent developments under conventional asymptotics include Hjort et al. (2009) and Bravo et al. (2020), and many more. See Chen and Van Keilegom (2009) for a recent review. Jackknife empirical likelihood was first proposed in the seminal work of Jing et al. (2009) for parametric $U$-statistics and has since been studied by Gong et al. (2010), Peng et al. (2012), Peng (2012), Wang et al. (2013), Zhang and Zhao (2013), Matsushita and Otsu (2018) and Chen and Tabri (2020), to list a few, for different applications. Most of the aforementioned works focus on actual $U$-statistics under iid sampling. Matsushita and Otsu (2021) demonstrate that JEL cannot only be modified for certain unconventional asymptotics of iid observations, but can also be applied to study edge probabilities for both sparse and dense dyadic networks studied in Bickel et al. (2011). This is achieved by incorporating the bias-correction idea of Hinkley (1978) and Efron and Stein (1981). In a context where controlling bias and spikes—from the presence of the bandwidth in kernel estimation—presents extra challenges, our modified JEL for dyadic KDE provides a nonparametric counterpart to the network asymptotic results in Matsushita and Otsu (2021).

Theoretically speaking, the modified JEL is closely related to the growing literature that uses the idea of accounting for the contributions from both linear and quadratic terms in the Hoeffding-type decomposition of the statistics. This idea emerges in the literature that studies "small-bandwidth asymptotics", in the context of density-weighted average derivative by Cattaneo (2014a, 2014b) and Cattaneo et al. (2013), for the sake of robustness over a wider range of bandwidths. This idea has since been explored in various other contexts such as an inference problem in partially linear models with many regressors in Cattaneo et al. (2018a, 2018b), as well as the various other examples explored in Matsushita and Otsu (2021).

Our results also complement the uniform convergence rates results for kernel type estimators under different dependence settings studied in Einmahl and Mason (2000), Einmahl and Mason (2005), Dony and Mason (2008), Hansen (2008), Kristensen (2009) and so forth. Theoretically, the $\sqrt{n}$-uniform rate of dyadic KDE share common underlying features with the kernel density estimator of Escanciano and Jacho-Chávez (2012), as well as the residual distribution estimator proposed in Akritas and Van Keilegom (2001), both under iid sampling.

In a seminal work, Fafchamps and Gubert (2007) proposed a dyadic robust estimator for regression models, which has since become the benchmark for inference under dyadic clustering. Related works include Frank and Snijders (1994) and Snijders and Borgatti (1999), Aronow et al. (2015), and Cameron and Miller (2014) in different contexts. Asymptotics for statistics of dyadic random arrays are studied by Davezies et al. (2020) under a general empirical processes setting and by Chiang et al. (2020) in high-dimensions. The former show the validity of a modified pigeonhole bootstrap (see, McCullagh 2000; Owen 2007) while the latter use a multiplier bootstrap, both under a non-degeneracy condition. On the other hand, for two-way separately exchangeable arrays and linear test-statistics, Menzel (2021) proposed a conservative bootstrap that is valid uniformly even under potentially non-Gaussian degeneracy scenarios. In this article, we only need to focus on the non-degenerate and Gaussian degenerate cases since the non-Gaussian degeneracy scenario is not a concern for dyadic KDE; Graham et al. (2019) pointed out that non-Gaussian degeneracy can be ruled out under mild bandwidth conditions.

More recently, Cattaneo et al. (2022) study uniform convergence rate and uniform inference for dyadic KDE. Their uniform convergence rate results further improved upon the result by allowing uniformity over the whole support for compactly supported data and enabling boundary-adaptive estimation. They further derive the minimax rate of uniform convergence for density estimation with dyadic data and show that the dyadic KDE is, under appropriate conditions, minimax-optimal. For inference, they obtain uniform distribution theory and provide bias-corrected $t$-statistics-based uniform confidence bands. Their methods differ from the ones we propose, which are based on jackknife empirical likelihood.

1.1. Notation and Organization

Let $|A|$ be the cardinality of a finite set $A$. Denote the uniform distribution on $[0, 1]$ as $U[0, 1]$. For $g : S \to \mathbb{R}$ and $\mathcal{X} \subseteq S$, denote $\|g\|_{\mathcal{X}} = \sup_{x \in \mathcal{X}} |g(x)|$. Write $a_n \lesssim b_n$ if there exists a constant $C > 0$ independent of $n$ such that $a_n \leq C b_n$. Throughout the article, the asymptotics should be understood as taking $n \to \infty$.

The rest of this article is organized as follows. In Section 2, we introduce our model and the dyadic KDE estimator. Our main theoretical results of uniform convergence rates and the validity of different JEL statistics for both complete and incomplete dyadic random arrays are then introduced in Section 3. Section 4 contains the simulation studies while the empirical application can be found in Section 5. A practical guideline on bandwidth choice and proofs of all the theoretical results are contained in the supplementary materials.

2. Model and Estimator

Consider a weighted undirected random graph. Let $I_n = \{(i, j) : 1 \leq i < j \leq n\}$ and $I_\infty = \bigcup_{n=2}^{\infty} I_n$. The random graph consists of an array $(X_{ij})_{(i,j) \in I_\infty}$ of real-valued random variables that is generated by

$$X_{ij} = f(U_i, U_j, U_{(i,j)}), \quad (U_i)_{i \in \mathbb{N}}, (U_{(i,j)})_{i,j \in \mathbb{N}} \overset{iid}{\sim} U[0, 1] \tag{2.1}$$

for some Borel measurable map $f : [0, 1]^3 \to \mathbb{R}$ that is symmetric in the first two arguments. Note that $U_i$ and $U_{(i,j)}$ are independent for all $i \in \mathbb{N}$, $(i, j) \in I_\infty$. While (2.1) may seem to be a specific structural assumption, it is implied by the following low-level condition of joint exchangeability via
the Aldous-Hoover-Kallenberg representation, see for example, (Kallenberg 2006, Theorem 7.22).

Definition 1 (Joint exchangeability). An \((X_{ij})_{(i,j)\in I_{2n}}\) is said to be jointly exchangeable if for any permutation \(\pi\) of \(\mathbb{N}\), the arrays \((X_{\pi(i),\pi(j)})_{(i,j)\in I_{2n}}\) and \((X_{\pi(i),\pi(j)})_{(i,j)\in I_{2n}}\) are identically distributed.

Remark 1 (Identical distribution). Under this setting, \((X_{ij})_{(i,j)\in I_{2n}}\) are identically distributed.

Suppose that the researcher observes \((X_{ij})_{(i,j)\in I_{2n}}\) for \(n \geq 2\). The object of interest is the density function \(f(x)\) for the conditional density function of \(X\) given \(U\), defined as:

\[ f(x) = \frac{1}{\mathbb{P}(U \cap X = x)} \sum_{i} f(x_{i}) \delta_{x_{i}}(U) \]

where \(\delta_{x_{i}}(U)\) denotes an indicator function of \(x_{i}\) for \(U\).

Theorem 1 (Uniform convergence rates for dyadic KDE). Suppose Assumption 1 holds, then with probability at least \(1 - o(1)\),

\[ \sup_{x \in \mathcal{X}} \left| \hat{f}_{n}(x) - f(x) \right| \lesssim h_{n}^{2} + \frac{1}{n^{1/2}} \]

if \(\inf_{x \in \mathcal{X}} \var{f_{n}(x)} \geq \nu_{0} > 0\) (called the non-degenerate case). Otherwise (called the degenerate case), with probability at least \(1 - o(1)\),

\[ \sup_{x \in \mathcal{X}} \left| \hat{f}_{n}(x) - f(x) \right| \lesssim h_{n}^{2} + \frac{1}{n} + \sqrt{\frac{\log(1/h_{n})}{n^{2}h_{n}}} \]

A proof can be found in Section D.1 of the supplementary materials.

Remark 2 (Convergence rates). Theorem 1 shows the uniform convergence rates under both non-degenerate and degenerate scenarios. In both cases, the bias is \(O(h_{n}^{2})\). In case of degeneracy, the uniform rate coincides with the independent sampling situation with sample size \(2^{(\nu_{0})}\). Besides the bias term, the component \(\sqrt{\frac{\log(1/h_{n})}{n^{2}h_{n}}}\) is the part in which the kernel plays a role (in adding the factor of \(\frac{\log(1/h_{n})}{n^{2}h_{n}}\)). In the nondegenerate case, the stochastic component is \(O_{p}(n^{-1/2})\) because its leading term consists of a sample average of \(\{E[h_{n}^{-1}K((X - X_{ij})/h_{n})|U]\}_{i=1}^{n}\), and these elements are well-behaved under our assumptions. These results agree with the observations made in Graham et al. (2019) for the pointwise behaviors of the dyadic KDE. In a related paper, Graham et al. (2021) consider nonparametric regression with dyadic data and establish uniform convergence rates for a Nadaraya-Watson type estimator under a different setting. More explicitly, their rates are nonparametric rates since the regressors considered are vertex-specific.

3.3. Inference for Complete Dyadic Data with JEL and Modified JEL

We now introduce JEL-based inference procedures for dyadic KDE. Denote the leave-one-out and leave-two-out index sets \(I_{2n}^{(1)} = \{(i, j) \in I_{2n} : i, j \notin \{i, j\}\}\), \(I_{2n}^{(2)} = \{(i, j) \in I_{2n} : i, j \notin \{i, j\}\}\).
For a given $\theta$, define
\[
\hat{\theta} = \left(\frac{n}{2}\right)^{-1} \sum_{(i,j) \in I_n} \frac{1}{h_n} K \left( \frac{x - X_{ij}}{h_n} \right),
\]
\[
\hat{\theta}^{(i)} = \left(\frac{n - 1}{2}\right)^{-1} \sum_{(k,j) \in I_n^{(i)}} \frac{1}{h_n} K \left( \frac{x - X_{kj}}{h_n} \right), \quad i = 1, \ldots, n,
\]
\[
\hat{\theta}^{(ij)} = \left(\frac{n - 2}{2}\right)^{-1} \sum_{(k,l) \in I_n^{(ij)}} \frac{1}{h_n} K \left( \frac{x - X_{kl}}{h_n} \right), \quad 1 \leq i < j \leq n,
\]
and subsequently $S(\theta) = \hat{\theta} - \theta$, $S^{(i)}(\theta) = \hat{\theta}^{(i)} - \theta$, and $S^{(ij)}(\theta) = \hat{\theta}^{(ij)} - \theta$. Furthermore, define the pseudo true value for JEL as
\[
V_i(\theta) = nS(\theta) - (n - 1)S^{(i)}(\theta).
\]
For modified JEL, define
\[
Q_{ij} = \frac{n - 3}{n - 1} \left[ nS(\theta) - (n - 1)(S^{(i)}(\theta) + S^{(j)}(\theta)) + (n - 2)S^{(ij)}(\theta) \right], \quad i < j,
\]
\[
\Gamma^2 = \frac{1}{n} \sum_{i=1}^{n} V_i(\theta)^2, \quad \Gamma_m^2 = \frac{1}{n} \sum_{i=1}^{n} V_i(\theta)^2 - \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q_{ij}^2. \tag{3.1}
\]
For each value of $\theta$, define the pseudo true value for modified JEL by
\[
V^m_i(\theta) = V_i(\hat{\theta}) - \Gamma \Gamma_m^{-1} \{ V_i(\hat{\theta}) - V_i(\theta) \}.
\]
Now, define the modified JEL function for $\theta$ by
\[
\ell^m(\theta) = -2 \sup_{w_1, \ldots, w_n} \frac{1}{n} \sum_{i=1}^{n} \log(n w_i),
\]
s.t. $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$, $\sum_{i=1}^{n} w_i V^m_i(\theta) = 0,$
and $\ell(\theta)$ is defined analogously with $V^m_i$ replaced by $V_i$. The Lagrangian dual problems are
\[
\ell(\theta) = 2 \sup_{\lambda} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \lambda V_i(\theta)) \tag{\ref{eq:dual}},
\]
\[
\ell^m(\theta) = 2 \sup_{\lambda} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \lambda V^m_i(\theta)).
\]
We say $f$ is nondegenerate at $x$ if $\text{var}(f_{X_{12}[U_1]}(x|U_1)) \geq L > 0$, which means that the vertex-specific shock $U_1$ affects the conditional density $f_{X_{12}[U_1]}$. Before stating the next result, which characterizes the asymptotics of JEL and modified JEL for inference, we make the following assumptions.

**Assumption 2 (JEL and modified JEL for complete data).** Suppose that
1. $f(x) > 0$ at the design point of interest $x$ which lies in $\mathcal{X} \subseteq \text{supp}(X_{12})$.
2. $K$ has bounded support, $nh_n^2 \to \infty$ and $nh_n^{5/2} \to 0$.

**Theorem 2 (Wilks’ theorem for JEL and modified JEL for dyadic KDE with complete data).** Suppose Assumptions 1 and 2 are satisfied, then
\[
\ell^m(\theta) \overset{d}{\to} \chi^2_1.
\]
In addition,
\[
\ell(\theta) \overset{d}{\to} \begin{cases} \chi^2_1, & \text{if } f \text{ is nondegenerate at } x, \\ \frac{1}{2} \chi^2_1, & \text{if } f \text{ is degenerate at } x. \end{cases}
\]
A proof can be found in Section D.2 of the supplementary materials.

**Remark 4 (Asymptotic pivotality).** Theorem 2 shows that the modified JEL is pivotal regardless of whether $f$ is degenerate or not, while JEL is asymptotically pivotal only if $f$ is nondegenerate at $x$. Theorem 2 implies that one can construct an approximate $1 - \alpha$ confidence interval
\[
\mathcal{R}_\alpha = \{ \theta : \ell^* (\theta) \leq c_\alpha \}
\]
for $\ell^* \in \{\ell, \ell^m\}$, where $c_\alpha$ is such that $\lim_{n \to \infty} P(\theta \in \mathcal{R}_\alpha) = \mathbb{P}(\chi^2_1 \leq c_\alpha) = 1 - \alpha$.

**Remark 5 (Conservatism of JEL).** An implication of Theorem 2 is that under degeneracy, JEL is asymptotically conservative. Thus, the tests and confidence intervals based on JEL are, although not always asymptotically precise, still asymptotically valid. On the other hand, although the mJEL is asymptotically precise, simulations in Section 4 shows that it is likely oversized when sample size is small. As such, JEL is still a practical option with its simpler implementation, especially if one prefers to be more conservative with the size control.

**Remark 6 (Uniform inference).** Theorem 2 focuses on pointwise inference. In Cattaneo et al. (2022), the authors pioneer uniform inference using strong approximation and boundary-adaptive kernels. They further develop a robust bias-correction procedure based on the bias-correction idea of Calonico et al. (2018). Their results are based on $t$-statistics. How to generalize an empirical likelihood type procedure to provide uniform inference for dyadic KDE remains an open question.

**Remark 7 (Modified jackknife variance estimator).** A direct implication of the proof of Theorem 2 is the consistency of $\Gamma_m^2$
in (3.1), a bias-corrected jackknife variance estimator of Efron and Stein (1981) adapted to our context. Based on this variance estimator, one can construct an alternative approximate 1 − α confidence interval for θ as \( \hat{\theta} \pm n^{-1/2}z_{\alpha/2}\Gamma_m^{-1} \), where \( z_{\alpha/2} \) is the \((1 - \alpha/2)\)th quantile of the standard normal random variable. This is formalized in the following corollary.

**Corollary 1 (Asymptotic normality with modified jackknife variance estimator).** Suppose Assumptions 1 and 2 are satisfied, then

\[
\sqrt{n}\Gamma_m^{-1}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 1).
\]

### 3.3. Inference for Incomplete Dyadic Data with JEL and Modified JEL

Let us now consider dyadic KDE with randomly missing incomplete data. Suppose the researcher observes

\( X^*_{ij} = Z_{ij}X_{ij} \)

where the random variables \( Z_{ij} \) are Bernoulli(\( p_n \)) that determine whether each edge is observed. Let \( p_n \) be independent from \( (X_{ij})_{(i,j) \in I_n} \), for some unknown probability of observation \( p_n \in (0, 1) \) which can be but is not necessarily fixed in sample size. Define \( N = \hat{p}_n(n) \), \( N_1 = \hat{p}_n(n^{-2}) \), and \( N_2 = \hat{p}_n(n^{-2}) \), with \( \hat{p}_n(n^{-2}) \) being an estimate of \( p_n \).

Now let \( I_n = \{(i, j) \in I_n : Z_{ij} = 1\} \) and \( I_n^c_k = \{(i, j) \in I_n : Z_{ij} = 1, \ i, j \neq k\} \). For a fixed \( \theta = f(x) \), define the incomplete dyadic KDE estimator and its leave-out counterparts by

\[
\begin{align*}
\hat{\theta}_\text{inc} &= \frac{1}{N} \sum_{(i, j) \in I_n} \frac{1}{h_n} K\left( \frac{x - X_{ij}}{h_n} \right), \\
\hat{\theta}_\text{inc}^{(i)} &= \frac{1}{N_1} \sum_{(k, j) \in I_n^c_k} \frac{1}{h_n} K\left( \frac{x - X_{kJ}}{h_n} \right), \quad i = 1, \ldots, n, \\
\hat{\theta}_\text{inc}^{(ij)} &= \frac{1}{N_2} \sum_{(k, j) \in I_n^c_k} \frac{1}{h_n} K\left( \frac{x - X_{kJ}}{h_n} \right), \quad 1 \leq i < j \leq n,
\end{align*}
\]

and the JEL pseudo true value for incomplete dyadic data by

\[
\hat{V}_i(\theta) = n \hat{S}(\theta) - (n - 1) \hat{S}^{(i)}(\theta),
\]

where \( \hat{S}(\theta) = \hat{\theta}_\text{inc} - \hat{\theta}, \hat{S}^{(i)}(\theta) = \hat{\theta}_\text{inc}^{(i)} - \hat{\theta} \). In addition, for each \( i < j \), define

\[
\begin{align*}
\hat{S}_i^{(ij)}(\theta) &= \hat{\theta}_\text{inc}^{(ij)} - \hat{\theta}, \\
\hat{Q}_{ij} &= \frac{n - 3}{n - 1} \left[ n \hat{S}_i(\theta) - (n - 1) \hat{S}^{(i)}(\theta) + \hat{S}^{(ij)}(\theta) + (n - 2) \hat{S}_j^{(ij)}(\theta) \right], \\
\hat{\Gamma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{V}_i(\hat{\theta}_\text{inc})^2, \quad \hat{\Gamma}^m = \frac{1}{n} \sum_{i=1}^n \hat{V}_i(\hat{\theta}_\text{inc})^2 - \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{Q}_{ij}^2.
\end{align*}
\]

For each \( i \), define the modified JEL pseudo true value for incomplete dyadic data by

\[
\hat{V}_i^m(\theta) = \hat{V}_i(\hat{\theta}_\text{inc}) - \hat{\Gamma}^m - 1(\hat{V}_i(\hat{\theta}_\text{inc}) - \hat{V}_i(\hat{\theta}_\text{inc})).
\]

Now the modified JEL function for incomplete dyadic data for \( \theta \) can be defined by

\[
\hat{\ell}^m(\theta) = -2 \sup_{w_1, \ldots, w_n} \sum_{i=1}^n \log(w_i), \\
\text{s.t. } w_i \geq 0, \quad \sum_{i=1}^n w_i = 1, \quad \sum_{i=1}^n w_i \hat{V}_i^m(\theta) = 0.
\]

and \( \hat{\ell}(\theta) \) is defined analogously with \( \hat{V}_i^m \) replaced by \( \hat{V}_i \). The Lagrangian dual problems are now

\[
\hat{\ell}(\theta) = 2 \sup_{\lambda} \sum_{i=1}^n \log(1 + \lambda \hat{V}_i(\theta)), \\
\hat{\ell}^m(\theta) = 2 \sup_{\lambda} \sum_{i=1}^n \log(1 + \lambda \hat{V}_i^m(\theta)).
\]

**Assumption 3 (JEL and modified JEL for incomplete data).** Suppose \( Z_{ij} \) are iid independent from \( (X_{ij})_{(i,j) \in I_n} \) for some unknown sequence \( p_n \in (0, 1) \) with \( nhnp_n \to \infty \).

Assumption 3 imposes the random missing structure on the data and establishes a lower bound on the rate of \( p_n \), the probability of observing an edge, to ensure the presence of enough data to establish asymptotic theory. Note that this assumption accommodates sequences of \( p_n \) that converge to zero slowly enough. The condition \( nhnp_n \to \infty \) ensures that we have asymptotically increasing effective sample size.

The following result is an incomplete data counterpart of Theorem 2.

**Theorem 3 (Wilks’ theorem for JEL and modified JEL for dyadic KDE with incomplete data).** Suppose Assumptions 1, 2, and 3 are satisfied, then

\[
\hat{\ell}(\theta) \xrightarrow{d} \chi^2_1.
\]

In addition, if \( f \) is nondegenerate at \( x \) or \( p_n \to 0 \), then

\[
\hat{\ell}(\theta) \xrightarrow{d} \chi^2_1.
\]

A proof can be found in Section D.3 of the supplementary materials.

**Theorem 3** states the asymptotic distributions of the proposed JEL and modified JEL statistics for incomplete data. As in **Theorem 2**, the modified JEL is asymptotically pivotal regardless of the asymptotic regime. However, with incomplete data, the asymptotic pivotality of the proposed JEL statistic happens not only under nondegeneracy but also when the proportion of observed data goes to zero slowly asymptotically. In such a scenario, the term that consists of the randomness induced by the missing process (which is conditionally independent) dominates the original leading asymptotic term, which is potentially not pivotal. Note that the construction of confidence intervals in **Remark 4** can be adapted with \( \ell^* \in \{\ell^m, \ell\} \).
4. Simulation Studies

We consider four sets of data-generating processes (DGP) in our simulations. The first two follow

\[ X_{ij} = \beta U_i U_j + U_{(i,j)} \]

with \( U_{(i,j)} \overset{iid}{\sim} N(0,1) \) and \( U_i = -1 \) with probability 1/3 and equals 1 otherwise. We set \( \beta \in \{0, 1\} \), where \( \beta = 1 \) is the sufficiently nondegenerate DGP considered in Graham et al. (2019). This has the density

\[ f(x) = \frac{5}{9} \phi(x - 1) + \frac{4}{9} \phi(x + 1), \]

where \( \phi \) is the density of a standard normal random variable. Meanwhile, \( \beta = 0 \) corresponds to the degenerate case where the true DGP is iid standard normal with density \( f = \phi \). The last two sets of DGPs are the incomplete data counterparts of the first two, with the same DGPs except that not all of the edges are observed following the set-up in Section 3.3. For these cases, we set \( p_n = 0.5 \). We use the rule of thumb bandwidths \( h_n \) and \( h_{inc} \) proposed in (A.1) and (A.2) in Section A of the supplementary materials, respectively. For complete dyadic data, we consider five alternative inference methods that are theoretically robust to dyadic clustering: (i) Wald statistic with FG variance estimator for dyadic KDE proposed in (Graham et al. 2019, eq. (22)) (FG), (ii) Wald statistic with the leading term of FG (IFG), (iii) jackknife empirical likelihood (JEL), (iv) Wald statistic with modified jackknife variance estimator (mJK) from Remark 7, and (v) modified jackknife empirical likelihood (mJEL). Note that (iii)-(v) are studied in this article. For incomplete dyadic data, we consider the complete counterparts of (iii) and (v) from Section 3.3. Throughout the simulation studies, we set the design point \( x = 1.675 \), consistent with the simulation study in Graham et al. (2019). Each simulation is iterated 5000 times.

Tables 1–4 show the simulation results under these different settings. In general, the simulation supports our theoretical findings. For complete data, under both the dyadic and iid DGPs, both mJEL and mJK-based confidence intervals enjoy close to nominal coverage rates even with moderate sample sizes, while the FG estimator is oversized with small sample sizes and converges to the nominal size when \( n \) is large. The original JEL and IFG estimators are conservative but asymptotically valid under the dyadic DGP. Consistent with Remark 5, the JEL estimator is always severely conservative under the iid DGP, as is the IFG estimator. The behaviors of JEL and mJK-based intervals under complete data are similar to those we observed with complete data; JEL is conservative but asymptotically converging to the correct size in the non-degenerate case while mJEL remains quite precise under both degenerate and nondegenerate DGPs.

5. Empirical Application: Airport Congestion

Flight delays caused by airport congestion are a familiar experience to the flying public. Aside from the frustration experienced by affected passengers, these delays are an important cause of fuel wastage and result in the release of criteria pollutants such as nitrogen oxides, carbon monoxide and sulfur oxides. As discussed in Schlenker and Walker (2016), these releases harm the
health of people living near airports. To mitigate these harms, it is important to understand the distribution of congestion across routes, or flights between pairs of airports.

In this empirical application, we apply the dyadic KDE, JEL and modified JEL procedure to understand the distribution of airport congestion, as measured by taxi time, across pairs of airports in the United States. The taxi time of a flight is defined as the sum of taxi-out time (the time between the plane leaving its parking position and taking off) and taxi-in time (the time between the plane landing and parking). In this context, the edges $X_{ij}$ and $X_{jk}$ represent flights between airport $i$ and airport $j$ and flights between airport $j$ and airport $k$, respectively; the fact that flights share a common airport $j$ induces dependence between $X_{ij}$ and $X_{jk}$. Therefore, this is a natural setting to apply dyadic KDE and modified JEL.

Data on flight taxi times are obtained from the U.S. DOT Bureau of Transportation Statistics (BTS) Reporting Carrier On-Time Performance dataset, which compiles data on U.S. domestic nonstop flights from all carriers which earn at least 0.5% of scheduled domestic passenger revenues. From the flight-level data for June 2020, we collapse taxi time into various summary statistics (mean, 95th percentile, maximum) for each origin-destination airport pair, with the summary statistics being taken over flights in both directions so that the resulting network is undirected. Noting that a missing edge means that there were no nonstop flights between the vertices the edge connects, we apply the KDE, JEL, and modified JEL to the resulting network. For simplicity, the sample is restricted to the 100 largest airports by number of departing flights.

The histograms, dyadic KDE estimates, pointwise JEL and pointwise modified JEL 95% confidence intervals for the mean, 95th percentile and maximum taxi times between airport pairs in June 2020 are shown in Figure 1(A)–(C). The intervals are obtained by numerically inverting the test statistic for each design point (each whole minute in the range of each statistic). Since both the JEL and modified JEL confidence intervals in Figure 1 are pointwise, a comparison between JEL and mJEL can only be made for each of the finitely many points at which the confidence intervals were calculated. Note that the design points are the same for the JEL and mJEL graphs, making a point-by-point comparison possible.

For applied researchers, the results demonstrate how studying only the mean can be incomplete; unlike the mean, the 95th percentile and (particularly) maximum travel times exhibit positive skewness. Routes with taxi times lying in the positively skewed region cause the most problems, but this distinction is lost when only studying means. For all these statistics, the modified JEL procedure which we propose provides relatively precise estimates even under small sample sizes (100 vertices) and a large proportion of missing edges (72% missing).

6. Conclusion

This article studies the asymptotic properties of dyadic kernel density estimation. We establish uniform convergence rates under general conditions. For inference, we propose a modified jackknife empirical likelihood procedure which is valid regardless of degeneracy of the underlying DGP. We further extend the results to cover incomplete or missing at random dyadic data.
Simulation studies show robust finite sample performance of the modified JEL inference for dyadic KDE in different settings. We illustrate the method via an application to airport congestion.

Despite our focus on density estimation, the inference approach we take can be extended to cover local regression as in Graham et al. (2021), the local polynomial density estimation of Cattaneo et al. (2020b) or local polynomial distributional regression of Cattaneo et al. (2020a) under dyadic data. Another potential direction is to study incomplete data under unconfoundedness or other nonrandom missing models. Finally, in light of the recent paper of Cattaneo et al. (2022) which pioneers uniform inference for dyadic KDE using Wald statistics, it would be interesting to consider ways to adapt modified JEL for uniform inference as well. Investigating the modified JEL under these assumptions provides interesting directions for future research.

Supplementary Materials

The supplementary materials comprise the Supplementary Appendix, application code and data, and simulation code. The Supplementary Appendix includes the proofs of the Theorems in this article and the bandwidth choices for our applications and simulations. The application code includes a Stata do-file for preprocessing the data and a R script that generates Figures 1a, 1b and 1c. The simulation code includes two R scripts, one for the complete case and one for the incomplete case.

Acknowledgments

We thank coeditor Jianqing Fan, the Associate Editor, and three anonymous referees for their constructive comments that have helped us significantly improved the paper. We are indebted to Matias Cattaneo, the anonymous AE, and two anonymous referees for pointing out an error in the earlier version and for suggesting a fix. We benefited from useful comments and discussions with Bruce Hansen, Yuya Sasaki, and the participants at Bristol Econometrics Study Group. We thank Yukitoshi Matushita for sharing their MATLAB simulation code. All remaining errors are ours.

ORCID

Bing Yang Tan © https://orcid.org/0000-0001-8621-5923

References

Akritas, M. G., and Van Keilegom, I. (2001), “Non-parametric Estimation of the Residual Distribution,” Scandinavian Journal of Statistics, 28, 549–567. [907]
Aronow, P. M., Samii, C., and Assenova, V. A. (2015), “Cluster–Robust Variance Estimation for Dyadic Data,” Political Analysis, 23, 564–577. [907]
Bickel, P. J., Chen, A., and Levine, E. (2011), “The Method of Moments and Degree Distributions for Network Models,” Annals of Statistics, 39, 2280–2301. [907]
Bravo, F, Escanciano, J. C., and Van Keilegom, I. (2020), “Two-Step Semiparametric Empirical Likelihood Inference,” Annals of Statistics, 48, 1–26. [907]
Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018), “On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference,” Journal of the American Statistical Association, 113, 767–779. [909]
Cameron, A. C., and D. L. Miller (2014), “Robust Inference for Dyadic Data,” Unpublished manuscript, University of California-Davis. [907]
Cattaneo, M., Feng, Y., and Underwood, W. (2022), “Uniform Inference for Kernel Density Estimators with Dyadic Data,” preprint. [907,908,909,913]
Cattaneo, M. D., Crump, R. K., and Jansson, M. (2013), “Generalized Jackknife Estimators of Weighted Average Derivatives,” Journal of the American Statistical Association, 108, 1243–1256. [907]
——— (2014a), “Bootstrapping Density-Weighted Average Derivatives,” Econometric Theory, 30, 1135–1164. [907]
——— (2014b), “Small Bandwidth Asymptotics for Density-Weighted Average Derivatives,” Econometric Theory, 30, 176–200. [907]
Cattaneo, M. D., Jansson, M., and Ma, X. (2020a), “Local Regression Distribution Estimators,” working paper. [913]
——— (2020b), “Simple Local Polynomial Density Estimators,” Journal of the American Statistical Association, 115, 1449–1455. [913]
Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018a), “Alternative Asymptotics and the Partially Linear Model with Many Regressors,” Econometric Theory, 34, 277–301. [907]
——— (2018b), “Inference in Linear Regression Models with Many Covariates and Heteroscedasticity,” Journal of the American Statistical Association, 113, 1350–1361. [907]
Chen, R., and Tabri, R. V. (2020), “Jackknife Empirical Likelihood for Inequality Constraints on Regular Functionals,” Journal of Econometrics, 221, 68–77. [907]
Chen, S. X., and Van Keilegom, I. (2009), “A Review on Empirical Likelihood Methods for Regression,” Test, 18, 415–447. [907]
Chiang, H. D., Kato, K., and Sasaki, Y. (2020), “Inference for High-Dimensional Exchangeable Arrays,” arXiv preprint arXiv:2009.05150. [907]
Davies, L., D’Haultfoeuille, X., and Guyonvarch, Y. (2021), “Empirical Process Results for Exchangeable Arrays,” Annals of Statistics, 49, 845–862. [907]
Dony, J., and Mason, D. M. (2008), “Uniform in Bandwidth Consistency of Conditional U-Statistics,” Bernoulli, 14, 1108–1133. [907]
Efron, B., and Stein, C. (1981), “The Jackknife Estimate of Variance,” Annals of Statistics, 9, 586–596. [907,910]
Einmahl, U., and Mason, D. M. (2000), “An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators,” Journal of Theoretical Probability, 13, 1–37. [907]
——— (2005), “Uniform in Bandwidth Consistency of Kernel-Type Function Estimators,” Annals of Statistics, 33, 1380–1403. [907]
Escanciano, J. C., and Jacho-Chávez, D. T. (2012), “β-Uniformly Consistent Density Estimation in Nonparametric Regression Models,” Econometrica, 167, 305–316. [907]
Fafchamps, M., and Gubert, F. (2007), “The Formation of Risk Sharing Networks,” Journal of Development Economics, 83, 326–350. [906,907]
Frank, O., and Snijders, T. (1994), “Estimating the Size of Hidden Populations Using Snowball Sampling,” Journal of Official Statistics, 10, 53–53. [907]
Gong, Y., Peng, L., and Qi, Y. (2010), “Smoothed Jackknife Empirical Likelihood Method for ROC Curve,” Journal of Multivariate Analysis, 101, 1520–1531. [907]
Graham, B. S. (2019), “Network Data,” Technical report, National Bureau of Economic Research. [906]
Graham, B. S., Niu, F., and Powell, J. L. (2019), “Kernel Density Estimation for Undirected Dyadic Data,” arXiv preprint arXiv:1907.13630.
Hinkley, D. V. (1978), “Improving the Jackknife with Special Reference to Correlation Estimation,” Biometrika, 65, 13–21. [907]
Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2009), “Extending the Scope of Empirical Likelihood,” Annals of Statistics, 37, 1079–1111. [907]
Jing, B.-Y., Yuan, J., and Zhou, W. (2009), “Jackknife Empirical Likelihood,” Journal of the American Statistical Association, 104, 1224–1232. [907]
Kallenberg, O. (2006), Probabilistic Symmetries and Invariance Principles, New York: Springer. [908]
Kristensen, D. (2009), “Uniform Convergence Rates of Kernel Estimators with Heterogeneous Dependent Data,” *Econometric Theory*, 24, 1433–1445. [907]

Matsushita, Y., and Otsu, T. (2018), "Likelihood Inference on Semiparametric Models: Average Derivative and Treatment Effect," *Japanese Economic Review*, 69, 133–155. [907]

——— (2021), "Jackknife Empirical Likelihood: Small Bandwidth, Sparse Network and High-Dimension Asymptotic," *Biometrika*, forthcoming. [906,907]

McCullagh, P. (2000), "Resampling and Exchangeable Arrays," *Bernoulli*, 6, 285–301. [907]

Menzel, K. (2021), "Bootstrap with Cluster-Dependence in Two or More Dimensions," *Econometrica*, 89, 2143–2188. [906,907]

Owen, A. (1990), "Empirical Likelihood Ratio Confidence Regions," *Annals of Statistics*, 18, 90–120. [907]

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249. [907]

——— (2001), *Empirical Likelihood*, Boca Raton, FL: CRC Press. [907]

Peng, L. (2012), “Approximate Jackknife Empirical Likelihood Method for Estimating Equations,” *Canadian Journal of Statistics*, 40, 110–123. [907]

Peng, L., Qi, Y., and Van Keilegom, I. (2012), "Jackknife Empirical Likelihood Method for Copulas," *Test*, 21, 74–92. [907]

Schlenker, W., and Walker, R. W. (2016), "Airports, Air Pollution, and Contemporaneous Health," *Review of Economic Studies*, 83, 768–809. [911]

Snijders, T. A., and Borgatti, S. P. (1999), "Non-Parametric Standard Errors and Tests for Network Statistics," *Connections*, 22, 161–170. [907]

Wang, R., Peng, L., and Qi, Y. (2013), "Jackknife Empirical Likelihood Test for Equality of Two High Dimensional Means," *Statistica Sinica*, 23, 667–690. [907]

Zhang, Z., and Zhao, Y. (2013), "Empirical Likelihood for Linear Transformation Models with Interval-Censored Failure Time Data," *Journal of Multivariate Analysis*, 116, 398–409. [907]