



Kent Academic Repository

Szydlowski, Arkadiusz (2025) *A note on kernel density estimation for undirected dyadic data*. *Econometric Reviews*, 44 (7). pp. 963-966. ISSN 0747-4938.

Downloaded from

<https://kar.kent.ac.uk/108786/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1080/07474938.2025.2471103>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

A note on kernel density estimation for undirected dyadic data

Arkadiusz Szydłowski

To cite this article: Arkadiusz Szydłowski (2025) A note on kernel density estimation for undirected dyadic data, *Econometric Reviews*, 44:7, 963-966, DOI: [10.1080/07474938.2025.2471103](https://doi.org/10.1080/07474938.2025.2471103)

To link to this article: <https://doi.org/10.1080/07474938.2025.2471103>



© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 20 Mar 2025.



Submit your article to this journal [↗](#)



Article views: 236



View related articles [↗](#)



View Crossmark data [↗](#)

A note on kernel density estimation for undirected dyadic data

Arkadiusz Szydłowski

University of Kent, Canterbury, UK

ABSTRACT

In this note, I show that the \sqrt{N} convergence to the normal distribution holds for the density of outcomes generated from a dyadic network using the seminal result in the U-statistic literature obtained by Frees. In particular, our derivations imply that the main result for the non degenerate case in Graham, Niu, and Powell follows from arguments in Frees.

ARTICLE HISTORY

3 September 2024
18 February 2025

KEYWORDS

Kernel density; Networks; non parametric density estimation; rate of convergence; U-statistic

1. Introduction

Graham, Niu, and Powell (2024) (henceforth, GNP) analyze non parametric estimation of the marginal density of:

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

where $\{A_i\}_{i=1}^N$ and $\{V_{ij}\}_{i,j=1}^N$ are i.i.d. and mutually independent and the function W is symmetric in the first two arguments. Note that this implies that $W_{ij} \perp W_{kl}$ unless at least one of the indices in (i, j) and (k, l) coincide. They show that the kernel density estimator:

$$\hat{f}_W(t) = \frac{2}{N(N-1)} \sum_{i < j} \frac{1}{h_N} K\left(\frac{t - W_{ij}}{h_N}\right)$$

converges to the normal distribution at the parametric rate \sqrt{N} .

Frees (1994) (henceforth, FR) analyzes non parametric estimation of the marginal density of $g(A_1, A_2, \dots, A_m)$, where $\{A_i\}_{i=1}^N$ is an i.i.d. sequence and g is symmetric in all arguments,¹ and shows that the kernel density estimator:

$$\hat{f}_g(t) = \binom{N}{m}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq N} \frac{1}{h_N} K\left(\frac{t - g(A_{i_1}, A_{i_2}, \dots, A_{i_m})}{h_N}\right)$$

converges to the normal distribution at the parametric rate \sqrt{N} .

Intuitively, to see the relationship between the two results, first assume that V_{ij} is drawn from the same distribution as A_i 's. As V_{ij} 's are independent of A_i 's, without loss of generality we can write $W_{ij} \equiv W_{ijk} = W(A_i, A_j, A_k)$. Define the symetrized version of W_{ijk} as:

$$g(A_i, A_j, A_k) = W(A_i, A_j, A_k) + W(A_k, A_i, A_j) + W(A_i, A_k, A_j)$$

(note that W is symmetric in the first two arguments). Now asymptotic \sqrt{N} normality of the kernel density estimate of the density of g follows from the main theorem in FR. Note that, beyond standard

CONTACT Arkadiusz Szydłowski  a.szydowski@kent.ac.uk  University of Kent, Canterbury, UK.

¹Giné and Mason (2007) extend his results to a uniform-in-bandwidth result.

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

conditions on the kernel function, FR requires the density of $g(a, A_j, A_k)$, $w_1(t; a)$, to exist and satisfy $\sup_t E_A |w_1(t; A)|^{2+\delta} < \infty$, which is implied by the smoothness conditions for W and the density of V_{ij} imposed by GNP.

To be precise, all the derivations above and below hold for the case when $\hat{f}_W(t)$ forms a non degenerate U-statistics and under standard bandwidth conditions (see Assumption M(c) below). GNP also analyze the degenerate case and the case of "knife-edge" undersmoothing (see their results R4 and R5), which this note does not cover.

2. Main result

The previous discussion imposed some additional assumptions on the model in GNP. Here I show that even without restricting the distribution of V_{ij} (beyond assumptions in GNP) and without symmetrizing the function W in the third argument, the asymptotic \sqrt{N} normality of the kernel density estimator follows from arguments in FR as the shock V gets integrated out in this argument anyway.

Define $f_{W|AA}$ as the marginal distribution of W_{ij} given (A_i, A_j) . We make the same assumptions as the ones used in GNP (pp. 3,5):

- Assumption M.** (a) $f_{W|AA}(w|a_1, a_2)$ is bounded and twice continuously differentiable for all w , a_1 and a_2 .
 (b) K is bounded, symmetric; $K(u) = 0$ if $|u| > \tilde{u}$ for some finite \tilde{u} ; $\int K(u)du = 1$.
 (c) $h_N \rightarrow 0$, $Nh_N \rightarrow \infty$, $Nh_N^4 \rightarrow 0$.

Note that condition (a) implies that $\sup_t E |w_1(t; A_1)|^{2+\delta} < \infty$, where $w_1(t; a)$ is the marginal density of W_{ij} given $A_i = a$. Part (c) assumes undersmoothing and, thus, means that the bias of the kernel estimator goes to zero. Overall, Assumption M implies that the conditions of the main theorem in FR are satisfied with the asymptotic bias $B = 0$.

The following proposition shows that the main result in GNP follows from FR. As in FR, one can prove a slightly more general version of this theorem with the asymptotic bias $B \neq 0$ using the same techniques; however, for simplicity, we concentrate on the case of undersmoothing, as this is the main case in the discussion of GNP. Additionally, for the sake of exposition (as in GNP), we give the result for the second-order U statistic, but the same proof would apply to higher-order U's (as in FR).

Proposition 1. Under Assumption M we have:

$$\sqrt{N}(\hat{f}_W(t) - f_W(t)) \rightarrow N(0, 4\text{Var}(w_1(t; A_1))).$$

Proof. As in FR, we will start with showing that the residual term in the Hoeffding decomposition converges to zero in probability.

Define

$$W_{1N}(a, t) = h_N^{-1} E \left[K \left(\frac{t - W(a, A_2, V_{12})}{h_n} \right) \right] - h_N^{-1} E \left[K \left(\frac{t - W(A_1, A_2, V_{12})}{h_N} \right) \right],$$

and $R_N(t) = \frac{2}{N(N-1)} \sum_{1 \leq i_1 < i_2 \leq N} \tilde{g}(A_{i_1}, A_{i_2}, V_{i_1 i_2}; t)$ where:

$$\begin{aligned} \tilde{g}(a_1, a_2, v_{12}; t) &= \frac{1}{h_N} K \left(\frac{t - W(a_1, a_2, v_{12})}{h_N} \right) - \frac{1}{h_N} E \left[K \left(\frac{t - W(A_1, A_2, V_{12})}{h_N} \right) \right] \\ &\quad - W_{1N}(a_1, t) - W_{1N}(a_2, t) \end{aligned}$$

Lemma A. Let Assumption M hold. Then:

$$R_N(t) = O_p(h_N^{-1/2} N^{-1}).$$

Proof. Note that $E[\tilde{g}(A_{i_1}, A_{i_2}, V_{i_1 i_2}; t) | A_{i_1}] = 0$. We have:

$$\text{Var}(R_N(t)) = \frac{4}{N^2(N-1)^2} \sum_{1 \leq i_1 < i_2 \leq N} \sum_{1 \leq j_1 < j_2 \leq N} E[\tilde{g}(A_{i_1}, A_{i_2}, V_{i_1 i_2}; t) \tilde{g}(A_{j_1}, A_{j_2}, V_{j_1 j_2}; t)]. \quad (1)$$

When $\{i_1, i_2\}$ and $\{j_1, j_2\}$ have 0 or 1 element in common the expectation under the sum is zero. Otherwise, the cross-product is bounded by:

$$E[\tilde{g}^2(A_{i_1}, A_{i_2}, V_{i_1 i_2}; t)] \leq h_N^{-1} E \left[K \left(\frac{t - W(A_{i_1}, A_{i_2}, V_{i_1 i_2})}{h_N} \right)^2 \right] + h_N^{-1} E[W_{1N}(A_{i_1}, t)]^2 = O(h_N^{-1})$$

where the first term after the inequality is $O(h_N^{-1})$ by a standard argument, using smoothness of the distribution f_W , and the second term is $O(1)$ by the derivation below. Finally, by the same combinatorial argument as in FR the number of non zero elements in the sum in (1) is of order $O(N^2)$ and we have:

$$\text{Var}(R_N(t)) = O_p(h_N^{-1} N^{-2})$$

which is sufficient for the result. □

Now using the Hoeffding decomposition and [Lemma A](#), we have:

$$\sqrt{N}(\hat{f}_W(t) - E[\hat{f}_W(t)]) = \frac{2}{\sqrt{N}} \sum_{i=1}^N W_{1N}(A_i, t) + o_p(1).$$

First note that due to $Nh_N^4 \rightarrow 0$ we have $E[\hat{f}_W(t)] = f_W(t) + o(N^{-1/2})$. Next, recalling that $w_1(t; A) \equiv f_{W|A}(t|A)$, and that, by the change of variables, we have:

$$h_N^{-1} E \left[K \left(\frac{t - W(a, A_2, V_{12})}{h_n} \right) \right] = \int K(s) w_1(t - sh_N; a) ds,$$

we can write:

$$E[W_{1N}^2(A_i, t)] = \text{Var} \left(\int K(s) w_1(t - sh_N; A_i) ds \right) \leq \int K^2(s) E[w_1^2(t - sh_N; A_i)] ds < \infty$$

where we have used $E[X^2] \geq (E[X])^2$, and the final inequality follows from Assumption [M](#) which implies boundedness of $f_{W|A}$ and K . Finally, using this and a triangular array central limit theorem, we can show that:

$$\frac{2}{\sqrt{N}} \sum_{i=1}^N W_{1N}(A_i, t) \rightarrow^d N(0, 4\text{Var}(w_1(t; A)))$$

□

3. Discussion

In section "Extensions", Graham, Niu, and Powell (2024) conjecture that their derivation of the asymptotic distribution should also apply to an outcome defined as $W_{ij} = W(A_i, A_j)$. Actually, this directly follows from the result in Frees (1994), which shows again the generality and usefulness of his approach. In principle, one can apply the result in FR to any known function of the characteristics of two nodes i and j , for example $g(A_i, A_j) = |A_i - A_j|$, as long as the outcomes $\{A_i\}_{i=1}^N$ are i.i.d. (e.g., due to random assignment).

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

- Frees, E. W. (1994). Estimating densities of functions of observations. *Journal of the American Statistical Association* 89(426):517–525. doi: [10.1080/01621459.1994.10476776](https://doi.org/10.1080/01621459.1994.10476776)
- Giné, E., Mason, D. M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *The Annals of Statistics* 35(3):1105–1145.
- Graham, B. S., Niu, F., Powell, J. L. (2024). Kernel density estimation for undirected dyadic data. *Journal of Econometrics* 240(2):105336 doi: [10.1016/j.jeconom.2022.06.011](https://doi.org/10.1016/j.jeconom.2022.06.011)