Supplement to "On the Theoretical Properties of the Network Jackknife"

Qiaohui Lin
qiaohui.lin@utexas.edu

Robert Lunde
rlunde@utexas.edu

Purnamrita Sarkar
purna.sarkar@austin.utexas.edu

Department of Statistics and Data Sciences
University of Texas at Austin

Abstract

We use alphanumeric numbering to avoid confusion with numbering in the main paper. Section A contains proof of Theorem 1. We present the proof of Theorem 2 in Section B and the proof of Theorem 3 in Section C. In Section D we present the proof of Proposition 1. Finally we present proof of Proposition 2 in Section 3.4 and additional theoretical results in Section F. We conclude with additional experimental results in Section G.

A Proof of Theorem 1

In what follows, let $\boldsymbol{\xi}_n = (\xi_i)_{1\leq i\leq n}$ and $\boldsymbol{\eta}_n = (\eta_{ij})_{1\leq i<j\leq n}$. Furthermore, we will let $\boldsymbol{\xi}_{n,i}$ denote the vector formed by removing node i and $\boldsymbol{\eta}_{n,i}$ denote the (concatenated) vector formed by removing all elements containing row or column index i.

Proof. Let $Z_{n,i} = g(\boldsymbol{\xi}_{n,i}, \boldsymbol{\eta}_{n,i})$ denote the functional calculated on an induced subgraph of $n-1$ nodes excluding node i. As before, let $Z_{n-1} = Z_{n,n}$. Construct the following martingale difference sequence:

$$d_i = E(Z_{n-1}|\Sigma_i) - E(Z_{n-1}|\Sigma_{i-1}) \tag{A.1}$$

Here, we consider a filtration introduced by Borgs et al. (2008), which was originally used to establish exponential concentration for certain subgraph frequencies in the dense regime.

Let $\Sigma_0 = \{\emptyset, \Omega\}$, $\Sigma_1 = \sigma(\xi_1)$, $\Sigma_2 = \sigma(\xi_1, \xi_2, \eta_{12})$, $\Sigma_3 = \sigma(\xi_1, \xi_2, \xi_3, \eta_{12}, \eta_{13}, \eta_{23})$ and so forth up to n. The filtration we consider has the following interpretation: for each time $1 \leq t \leq n$, suppose that we observe a $t \times t$ adjacency matrix induced by the nodes $\{1, 2, \ldots, t\}$. Then, Σ_t captures all of the randomness in the corresponding induced subgraph. We may visualize Σ_i as a σ-field generated by a triangular array so that:

$$\Sigma_i = \sigma \left\{ \begin{array}{ccccc} \xi_1 & \eta_{12} & \dots & \eta_{1,i-1} & \eta_{1i} \\ & \xi_2 & \dots & \eta_{2,i-1} & \eta_{2i} \\ & & \dots & .. & \\ & & \xi_{i-2} & \eta_{i-2,i-1}, & \eta_{i-2,i} \\ & & & \xi_{i-1} & \eta_{i-1,i} \\ & & & & \xi_i \end{array} \right\};\ \Sigma_{i-1} = \sigma \left\{ \begin{array}{cccc} \xi_1 & \eta_{12} & \dots & \eta_{1,i-1} \\ & \xi_2 & \dots & \eta_{2,i-1} \\ & & \dots & .. \\ & & \xi_{i-2} & \eta_{i-2,i-1} \\ & & & \xi_{i-1} \end{array} \right\}$$

Observe that $Z_{n-1} - E(Z_{n-1}) = \sum_{i=1}^{n} d_i$, d_i is Σ_i measurable, and $E(d_i|\Sigma_{i-1}) = 0$. Therefore, the variance of Z_n can be written as:

$$\text{Var } Z_{n-1} = E\left(\sum_{i=1}^{n} d_i\right)^2 = \sum_{i=1}^{n} E(d_i^2) + 2\sum_{i<j} E(d_i d_j)$$

Now, for $i \neq j$, observe that:

$$\begin{aligned} E(d_i d_j) &= E(E(d_i d_j|\Sigma_i)) = E(d_i)E(d_j|\Sigma_i) \\ &= E(d_i)(E[E(S_n|\Sigma_j)|\Sigma_i] - E[E(S_n|\Sigma_{j-1})|\Sigma_i]) = 0 \end{aligned}$$

For the jackknife estimate, we have that:

$$E\left(\sum_{i=1}^{n}(Z_{n,i} - \bar{Z}_n)^2\right) = \sum_{i<j} \frac{E(Z_{n,i} - Z_{n,j})^2}{n} = \frac{(n-1)\cdot E(Z_{n,1} - Z_{n,2})^2}{2}$$

We also denote by $\Sigma_{i:j}$, the sigma field containing all information of random variables $\xi_i, \dots, \xi_j$, and $\eta_{k\ell}, i \leq k < \ell \leq j$. Now define $\mathcal{A}$ as $\Sigma_{3:i+1}$. Since Z_{n-1} is invariant to node-permutation, $\mathcal{A}$ is independent of $\sigma(\xi_2, \eta_{23}, \dots, \eta_{2n})$ and $\sigma(\xi_1, \eta_{13}, \dots, \eta_{1n})$,

$$E(Z_{n,1}|\mathcal{A}) = E(Z_{n,2}|\mathcal{A})$$

Define:

$$U = E(Z_{n,1}|\Sigma_{i+1}) - E(Z_{n,1}|\mathcal{A}), \quad V = E(Z_{n,2}|\Sigma_{i+1}) - E(Z_{n,2}|\mathcal{A}) \qquad \text{(A.2)}$$

Then, using the fact that $E[X^2|\Sigma_{i+1}] \geq E[X|\Sigma_{i+1}]^2$ for some Σ_{i+1} measurable r.v. X, we have:

$$E(Z_{n,1} - Z_{n,2})^2 \geq E[E(Z_{n,1}|\Sigma_{i+1}) - E(Z_{n,2}|\Sigma_{i+1})]^2 = E(U - V)^2 \qquad \text{(A.3)}$$

Notice that conditional on $\mathcal{A}$, U is a function of $\{\xi_2, \eta_{23}, \dots, \eta_{2,i+1}\}$, while V is a function of $\{\xi_1, \eta_{13}, \dots, \eta_{1,i+1}\}$. Thus, U and V are conditionally independent. Then, since $\mathcal{A} \subset \Sigma_{i+1}$, by the tower property of conditional expectations, we have that:

$$E(U-V)^2 = E(U^2) - 2E(UV) + E(V^2) = E(U^2) + E(V^2) - 2E(E(U|\mathcal{A})E(V|\mathcal{A}))$$

$$= E(U^2) + E(V^2),$$

Now, we expand $E(U^2)$ as follows:

$$\begin{aligned}
E(U^2) &= E((E(Z_{n,1}|\Sigma_{(i+1)}) - E(Z_{n,1}|\mathcal{A}))^2] \\
&\overset{(i)}{=} E((E(Z_{n,1}|\Sigma_{2:i+1}) - E(Z_{n,1}|\Sigma_{3:i+1}))^2] \\
&\overset{(ii)}{=} E[(E(Z_{n,n}|\Sigma_{1:i}) - E(Z_{n,n}|\Sigma_{1:i-1}))^2] \\
&= E[(E(Z_{n-1}|\Sigma_i) - E(Z_{n-1}|\Sigma_{i-1}))^2] = E(d_i^2)
\end{aligned}$$

Step (i) holds because the random variables associated with node 1 are not present in $Z_{n,1}$. Step (ii) holds because $\xi_1, \dots \xi_n$ and $\eta_{ij}, 1 \leq i < j \leq n$ are i.i.d random variables, and $E[Z_{n,1}|\Sigma_{2:i+1}]$ ($E[Z_{n,1}|\Sigma_{3:i+1}]$) and $E[Z_{n,n}|\Sigma_{1:i}]$ ($E[Z_{n,n}|\Sigma_{1:i-1}]$) are equal in distribution.

Similarly, $EV^2 = Ed_i^2$, $E(U-V)^2 = 2Ed_i^2$. Thus,

$$E(Z_{n,1} - Z_{n,2})^2 \geq E(U-V)^2 = 2Ed_i^2 \tag{A.4}$$

$$E\left(\sum_{i=1}^{n}(Z_{n,i} - \bar{Z}_n)^2\right) = \frac{n-1}{2}E(Z_{n,1}-Z_{n,2})^2 \geq (n-1)Ed_i^2 = \text{Var } Z_{n-1} \tag{A.5}$$

□

B Proof of Theorem 2

For notational convenience, let $Z_n = \hat{P}(R)$ and let $Z_{n,i}$ denote the subgraph frequency defined in Eq 15 with node i removed:

$$Z_{n,i} = \rho_n^{-e} \frac{1}{\binom{n-1}{p}\ |\text{Iso}(R)|} \sum_{S \sim R,\ i \notin V(S)} \mathbb{1}(S = G_n) \tag{A.6}$$

We first present a lemma that will be used in the proof. An identity relating the mean of leave-one-out jackknife estimates to a U-statistic plays an important role in the proof of jackknife consistency for U-statistics. Using a novel combinatorial argument, we show that a similar identity holds for normalized subgraph counts:

Lemma B.1. *Letting $Z_{n,i}$ and Z_n be defined as above, we have that:*

$$\bar{Z}_n := \frac{1}{n}\sum_{i=1}^{n} Z_{n,i} = Z_n$$

Proof. For a subgraph with p nodes and e edges, denote the number of this subgraph in G_n as Q. Denote the number of subgraphs node i is involved in as Q_i. We now analyze $\sum_{i=1}^{n} Q_i$. For each vertex set with cardinality p, a given subgraph is counted once from each vertex. Therefore, $\sum_{i=1}^{n} Q_i = pQ$.

Observe that $Z_{n,i} + Q_i = Q$ since the set of subgraphs that do not contain node i and the set of subgraphs that contain node i are disjoint and their union gives the set of subgraphs counted in Q. It follows that:

$$\frac{1}{n}\sum_i Z_{n,i} = \frac{\frac{1}{n}\sum_i (Q - Q_i)}{\binom{n-1}{p}\rho_n^e} = \frac{(n-p)Q}{n\binom{n-1}{p}\rho_n^e} = \frac{Q}{\binom{n}{p}\rho_n^e} = Z_n.$$

□

Now, we introduce the limiting value of the scaled variance, which represents the value we are aiming for with the jackknife. Bickel et al. (2011) show that the asymptotic behavior of $\hat{P}(R)$ is driven by a U-statistic corresponding to the edge structure of the subgraph. For a subgraph R with $V(R) = \{1, \ldots, p\}$, define the kernel:

$$h(x_1, \ldots, x_p) = \frac{1}{|Iso(R)|} \sum_{S \sim R,\ V(S)=\{1,\ldots,p\}} \prod_{(i,j)\in E(R)} w(x_i, x_j) \tag{A.7}$$

Theorem 1 of Bickel et al. (2011) establishes that:

$$n \cdot \text{Var } \hat{P}(R) \to \sigma^2$$

where $\sigma^2 = p^2\zeta$ is the variance of the U-statistic with kernel h (see for example, Serfling (1980), page 192) and $\zeta = \text{Var}(E(h(\xi_1, \ldots, \xi_p)|\xi_1))$. We will now scale the jackknife variance by n to study its asymptotics. Let:

$$\alpha_i = Z_{n,i} - E(Z_{n,i}|\boldsymbol{\xi}_n), \quad \beta_i = E(Z_{n,i}|\boldsymbol{\xi}_n) \tag{A.8}$$

For simplicity we will use $\bar{\alpha}_n$ (or $\bar{\beta}_n$) to denote the average of α_i (or β_i). Now, consider the following signal-noise decomposition:

$$\begin{aligned} n \cdot \sum_{i=1}^n (Z_{n,i} - \bar{Z}_n)^2 &= n \cdot \sum_{i=1}^n (\alpha_i - \bar{\alpha}_n + \beta_i - \bar{\beta}_n)^2 \\ &= n \cdot \sum_{i=1}^n (\alpha_i - \bar{\alpha}_n)^2 + 2n \cdot \sum_{i=1}^n (\alpha_i - \bar{\alpha}_n)(\beta_i - \bar{\beta}_n) \\ &+ n \cdot \sum_{i=1}^n (\beta_i - \bar{\beta}_n)^2. \end{aligned} \tag{A.9}$$

We start by bounding the third sum, which is the signal in our decomposition. Observe that β_i is a U-statistic with the kernel h defined in (A.7); therefore, by Theorem 1 and its following discussions of Chapter 5 in Lee (1990), we have that:

$$n \cdot \sum_{i=1}^n (\beta_i - \bar{\beta}_n)^2 \xrightarrow{P} \sigma^2 \tag{A.10}$$

The result will follow if we show that the remaining two sums in the decomposition are negligible. If the first sum is negligible, the Cauchy-Schwarz inequality would imply that:

$$n \cdot \sum_{i=1}^{n}(\alpha_i - \bar{\alpha}_n)(\beta_i - \bar{\beta}_n) \leq n \cdot \sqrt{\sum_{i=1}^{n}(\alpha_i - \bar{\alpha}_n)^2 \cdot \sum_{i=1}^{n}(\beta_i - \bar{\beta}_n)^2} \xrightarrow{P} 0$$

It remains to show that: $n \cdot \sum_{i=1}^{n}(\alpha_i - \bar{\alpha}_n)^2 \xrightarrow{P} 0$. Now, observe that:

$$\sum_{i=1}^{n}(\alpha_i - \bar{\alpha}_n)^2 = \sum_{i=1}^{n}\alpha_i^2 - n\bar{\alpha}_n^2$$

Expanding the square for $\sum_{i=1}^{n}\alpha_i^2$ we have that:

$$\begin{aligned}\sum_{i=1}^{n}\alpha_i^2 &= \sum_{i=1}^{n}(Z_{n,i} - E(Z_{n,i}|\boldsymbol{\xi}_n))^2 \\ &= \sum_{i=1}^{n}\binom{n-1}{k}^{-2} \sum_{S\sim R,\ i\notin V(S)} (\rho_n^{-e}\psi(S) - W(S)) \sum_{T\sim R,\ i\notin V(T)} (\rho_n^{-e}\psi(T) - W(T))\end{aligned}$$

where $\psi(S)$ and $W(S)$ are given by:

$$\begin{aligned}\psi(S) &= \frac{1}{|Iso(R)|} \prod_{(i,j)\in E(S),\ S\sim R} A_{ij} \ \times \prod_{(i,j)\in \overline{E(S)},\ S\sim R} 1 - A_{ij}, \\ W(S) &= \frac{1}{|Iso(R)|} \prod_{(i,j)\in E(S),\ S\sim R} w(\xi_i,\xi_j) \ \times \prod_{(i,j)\in \overline{E(S)},\ S\sim R} 1 - \rho_n w(\xi_i,\xi_j)\end{aligned}$$

and $\overline{E(S)}$ are $(i,j) \in V(S)\times V(S)$ that are not contained in $E(S)$. Now, similar to Lee (1990), we group elements in the sum based on the number of elements in $V(S)\cap V(T)$. For each $|V(S)\cap V(T)| = c$, there are $n-2p+c$ terms in total. It follows that:

$$\begin{aligned}\sum_{i=1}^{n}\alpha_i^2 &= \binom{n-1}{p}^{-2} \sum_{c=0}^{p}(n-2p+c) \sum_{|V(S)\cap V(T)|=c} (\rho_n^{-e}\psi(S) - W(S))(\rho_n^{-e}\psi(T) - W(T)) \\ &= \binom{n-1}{p}^{-2} \sum_{c=0}^{p}(n-2p+c) \sum_{|V(S)\cap V(T)|=c} \gamma(S,T), \quad \text{say.}\end{aligned}$$

Now we turn to $n\bar{\alpha}_n^2$;

$$\bar{\alpha}_n = \frac{1}{n}\sum_i Z_{n,i} - \frac{1}{n}\sum_i E(Z_{n,i}|\boldsymbol{\xi}_n) \overset{(i)}{=} Z_n - E(Z_n|\boldsymbol{\xi}_n)$$

Equality (i) follows from Lemma B.1. Now expanding $\bar{\alpha}_n^2$ in a similar manner, we have that

$$\bar{\alpha}_n^2 = \frac{(n-p)^2}{n}\binom{n-1}{p}^{-2}\sum_{c=0}^{p}\sum_{|V(S)\cap V(T)|=c}\gamma(S,T),$$

Then,

$$\begin{aligned}\sum_{i=1}^{n}(\alpha_i-\bar{\alpha}_n)^2 &= \binom{n-1}{p}^{-2}\sum_{c=0}^{p}\left(n-2p+c-\frac{(n-p)^2}{n}\right)\sum_{|V(S)\cap V(T)|=c}\gamma(S,T)\\ &= \sum_{c=0}^{p}\sum_{|V(S)\cap V(T)|=c}\left(c-\frac{p^2}{n}\right)\cdot\binom{n-1}{p}^{-2}\gamma(S,T)\end{aligned}$$

Now, taking expectations, we have that:

$$\begin{aligned}&E\left(\binom{n-1}{p}^{-2}\sum_{c=0}^{p}\sum_{|V(S)\cap V(T)|=c}\gamma(S,T)\right)\\ &= E\left(\binom{n-1}{p}^{-2}\sum_{c=0}^{p}\sum_{|V(S)\cap V(T)|=c}\Big(\rho_n^{-e}\psi(S)-W(S)\Big)\Big(\rho_n^{-e}\psi(T)-W(T)\Big)\right)\\ &= \big(1-o(1)\big)\cdot E\left[\operatorname{Var}(\hat{P}(R)\mid\boldsymbol{\xi}_n)\right] = o\left(\frac{1}{n}\right)\end{aligned}$$

where the last line follows from the proof of Theorem 1 of Bickel et al. (2011).

Now, by Markov inequality, we have that

$$n\cdot\sum_{i=1}^{n}(\alpha_i-\bar{\alpha}_n)^2 \xrightarrow{P} 0 \tag{A.11}$$

and the result follows.

C Proof of Theorem 3

Proof. Let $Z_{n,i} = (Z_{n,i}(1),\dots Z_{n,i}(d))$, where d is a constant w.r.t n and each entry corresponds to a count functional with node i removed. Each count functional may involve subgraphs of different sizes. We will use a Taylor expansion around $\bar{Z}_n$.

$$\begin{aligned}f(Z_{n,i}) &= f\big(\bar{Z}_n\big)+\nabla f(\zeta_i)^T(Z_{n,i}-\bar{Z}_n)\\ &= f\big(\bar{Z}_n\big)+\nabla f(\mu)^T(Z_{n,i}-\bar{Z}_n)+\underbrace{(\nabla f(\zeta_i)-\nabla f(\mu))^T(Z_{n,i}-\bar{Z}_n)}_{E_i},\end{aligned}$$

where $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{id}) = c_i Z_{n,i} + (1 - c_i)\bar{Z}_n$ for some $c \in [0, 1]$. Thus, we also have:

$$f(Z_{n,i}) - \overline{f(Z_{n,i})} = \underbrace{\nabla f(\mu)^T (Z_{n,i} - \bar{Z}_n)}_{I_i} + \underbrace{E_i - \frac{1}{n}\sum_i E_i}_{II_i} \tag{A.12}$$

For the first part we see that,

$$n\sum_i (I_i)^2 = n\nabla f(\mu)^T \left(\sum_i (Z_{n,i} - \bar{Z}_n)(Z_{n,i} - \bar{Z}_n)^T\right) \nabla f(\mu) \tag{A.13}$$

We will first show that the inner average of the above expression converges to the covariance matrix of $Z_{n,i}$ (recall that here we are considering a finite dimensional vector). Extending the same argument in Eq A.9 to finite dimensional $Z_{n,i}$'s (and α_i and β_i's defined in Eq A.8),

$$\begin{aligned} &n\sum_i (Z_{n,i} - \bar{Z}_n)(Z_{n,i} - \bar{Z}_n)^T \\ &= n\sum_i \big((\alpha_i - \bar{\alpha}_n)(\alpha_i - \bar{\alpha}_n)^T + (\alpha_i - \bar{\alpha}_n)(\beta_i - \bar{\beta}_n)^T + (\beta_i - \bar{\beta}_n)(\alpha_i - \bar{\alpha}_n)^T \\ &\qquad + (\beta_i - \bar{\beta}_n)(\beta_i - \bar{\beta}_n)^T\big) \end{aligned}$$

By Theorem 9 of Arvesen (1969) we have that:

$$n\sum_i (\beta_i - \bar{\beta}_n)(\beta_i - \bar{\beta}_n)^T \xrightarrow{P} \Sigma \tag{A.14}$$

Above, Σ is the covariance matrix of a multivariate U-statistic with kernels $(h_1, \ldots, h_d)$, where each h_j is the kernel corresponding to the count functional in the j^{th} coordinate of the vector Z_n (see Eq A.7). Now combining Eq A.14 with Eq A.13 we see that,

$$\begin{aligned} \left| n\sum_i (I_i)^2 - f(\mu)^T \Sigma f(\mu) \right| &\leq \|\nabla f(\mu)\|^2 n \sum_i \|\alpha_i - \bar{\alpha}_n\|^2 \\ &\quad + 2n\|\nabla f(\mu)\|^2 \sum_i |(\alpha_i - \bar{\alpha}_n)^T (\beta_i - \bar{\beta}_n)| \end{aligned} \tag{A.15}$$

The first part is $o_p(1)$ by an analogous argument leading to Eq A.11. For the second part, we see that an application of Cauchy Schwarz inequality gives:

$$n\sum_i |(\alpha_i - \bar{\alpha}_n)^T (\beta_i - \bar{\beta}_n)| \leq \sum_{j=1}^{d} \sqrt{\left(\sum_i n(\alpha_i(j) - \bar{\alpha}_n(j))^2\right)\left(n\sum_i (\beta_i(j) - \bar{\beta}_n(j))^2\right)}$$

The first part inside the square root is $o_p(1)$ due to Eq A.11, and the second part is $O_p(1)$ by Eq A.10. Using this in conjunction with Eq A.15 and since

$\|\nabla f(\mu)\|$ is bounded, we see that:

$$\left| n\sum_i (I_i)^2 - \nabla f(\mu)^T \Sigma \nabla f(\mu) \right| = o_p(1)$$

All that remains now is to show that part II_i in Eq A.12 is negligible even when summed and multiplied by n. First note that $(II_i)^2 \leq E_i^2$.

$$\begin{aligned} n\sum_i (II_i)^2 &\leq n\sum_i |(\nabla f(\zeta_i) - \nabla f(\mu))^T (Z_{n,i} - \bar{Z}_n)|^2 \\ &\leq \max_i \|\nabla f(\zeta_i) - \nabla f(\mu)\|^2 \left(n\sum_i (Z_{n,i} - \bar{Z}_n)^T (Z_{n,i} - \bar{Z}_n) \right) \end{aligned} \tag{A.16}$$

Theorem 2 shows that the second part in the RHS of Eq A.16 is $O_p(1)$. We will now show that the first part is asymptotically negligible.

Observe that:

$$\begin{aligned} \max_i \|\zeta_i - \mu\| &\leq \max_i \; c_i \|Z_{n,i} - \mu\| + \max_i \; (1 - c_i)\|\bar{Z}_n - \mu\| \\ &\leq \sqrt{d} \cdot \max_{i,j} |Z_{n,i}(j) - \bar{Z}_n(j)| + 2\|\bar{Z}_n - \mu\| \\ &\leq \sqrt{d} \cdot \max_j \sqrt{\sum_{i=1}^n \left(Z_{n,i}(j) - \bar{Z}_n(j)\right)^2} + 2\|Z_n - \mu\| \end{aligned}$$

Above, $\bar{Z}_n = Z_n$ by Lemma B.1. The first term on the RHS converges in probability to 0 from our Theorem 2. By Theorem 1 of Bickel et al. (2011), $\|Z_n - \mu\|$ is also negligible. Since $\max_i \|\zeta_i - \mu\| = o_p(1)$ and ∇f is continuous at μ, by continuity, we have that $\max_i \|\nabla f(\zeta_i) - \nabla f(\mu)\|^2 = o_p(1)$. Since the second term on the RHS of Eq A.16 is $O_p(1)$ from our previous argument and the first term is $o_p(1)$, it follows that the LHS of Eq A.16 is $o_p(1)$.

Let $\mu_n = E[Z_n]$. Note that if one counts subgraphs by an exact match as in Bickel et al. (2011) $\mu_n \to \mu$. If one counts subgraphs via edge matching, $\mu_n = \mu$. Thus, both these types of subgraph densities, which asymptotically have the same limit, can be handled by our theoretical results. By Theorem 3.8 in Van der Vaart (2000),

$$\sqrt{n}(f(Z_n) - f(\mu_n)) \rightsquigarrow N(0, \nabla f(\mu)^T \Sigma \nabla f(\mu))$$

This shows that the jackknife estimate of variance converges to the asymptotic variance of $f(Z_n)$.

□

D Proof of Proposition 1

Throughout this section, we will use the notation $x_n \asymp y_n$ to denote $x_n = y_n(1+o(1))$. Before presenting the proof, we present two accompanying lemmas which will be used in the proof of Proposition 1.

Lemma D.1. *Denote $D_i^{(n)}$ the degree of node i in the size n graph.*

$$\sum_{i=1}^{n-1} \mathrm{Var}\left(\frac{D_i^{(n)}}{\binom{n-1}{2}\rho_n}\right) \asymp \frac{4}{n^3} E(\mathrm{Var} \sum_{k,k\neq i} w(\xi_i, \xi_k)|\xi_i) + \frac{4}{n}\mathrm{Var}[E(w(\xi_i,\xi_k)|\xi_i)] + O(n^{-2}\rho_n^{-1}).$$

Lemma D.2. *Denote $D_i^{(n)}$ the degree of node i in the size n graph.*

$$\sum_{i,j,i\neq j} \mathrm{cov}\left(\frac{D_i^{(n)}}{\binom{n-1}{2}\rho_n}, \frac{D_j^{(n)}}{\binom{n-1}{2}\rho_n}\right) \asymp \frac{4}{n} \times 3\mathrm{Var}(E[w(\xi_i,\xi_j)|\xi_i]) + O(n^{-2}\rho_n^{-1})$$

We will use the above to lemmas to prove Proposition 1, which we now present.

Proof. Denote D_n as the total number of edges in graph G_n. By definition,

$$Z_n = \frac{D_n}{\binom{n}{2}\rho_n}$$

Denote $D_i^{(n)}$ the degree of node i in the size n graph. We have that $ED_i^{(n)} = ED_j^{(n)}$ for any node pair. Thus the jackknife estimate of edges for a graph with node i removed is D_n minus the degree of node i. Define

$$\gamma_n = \binom{n-1}{2}\rho_n; \quad \gamma_n' = \binom{n-1}{2}\rho_{n-1} \tag{A.17}$$

Then by definition, we have

$$Z_{n,i} = \frac{D_n - D_i^{(n)}}{\binom{n-1}{2}\rho_n} = \frac{D_n - D_i^{(n)}}{\gamma_n}$$

Then, the jackknife estimate is

$$E\sum_{i=1}^{n}(Z_{n,i} - \bar{Z}_n)^2 = \frac{1}{2n}\sum_{i\neq j} E(Z_{n,i} - Z_{n,j})^2 = \frac{1}{2n}\sum_{i\neq j} E\left(\frac{D_i^{(n)} - D_j^{(n)}}{\gamma_n}\right)^2$$
$$= \sum_{i=1}^{n-1}\mathrm{Var}\left(\frac{D_i^{(n)}}{\gamma_n}\right) - \frac{1}{n}\sum_{i\neq j}\mathrm{cov}\left(\frac{D_i^{(n)}}{\gamma_n}, \frac{D_j^{(n)}}{\gamma_n}\right) \tag{A.18}$$

whereas the total number of degrees in a $(n-1)$ graph is $D_{n-1} = \sum_{i=1}^{n-1} D_i^{(n-1)}/2$ as each edge is counted 2 times from each node. We first obtain an expression for Var Z_{n-1}.

$$\text{Var } Z_{n-1} = \text{Var}\left(\frac{\sum_{i=1}^{n-1} D_i^{(n-1)}/2}{\binom{n-1}{2}\rho_{n-1}}\right) = \frac{1}{4}(n-1)\text{Var}\left(\frac{D_i^{(n-1)}}{\gamma_n'}\right) \tag{A.19}$$

$$+\frac{1}{4}\sum_{i,j,i\neq j}\text{cov}\left(\frac{D_i^{(n-1)}}{\gamma_n'}, \frac{D_j^{(n-1)}}{\gamma_n'}\right) \tag{A.20}$$

For the second term in the R.H.S of Eq A.18, from Lemma D.2, it is easy to check that it is $O(n^{-2})$. Thus scaling Eq A.18 by $n-1$ we have,

$$\begin{aligned}(n-1)E\sum_{i=1}^{n}(Z_{n,i}-\bar{Z}_n)^2 &= (n-1)\sum_{i=1}^{n}\text{Var}\left(\frac{D_i^{(n)}}{\gamma_n}\right) + O\left(\frac{1}{n}\right)\\ = \frac{4}{n^2}E[\text{Var}\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i] &+ 4\text{Var}[E(w(\xi_i,\xi_k)|\xi_i)] + O\left(\frac{1}{n\rho_n}\right) + O\left(\frac{1}{n}\right)\end{aligned} \tag{A.21}$$

Plugging in Lemma D.2 into the second term of R.H.S of Eq A.19 and scaling Eq A.19 by $n-1$, we have

$$\begin{aligned}&(n-1)\text{Var } Z_{n-1}\\ &= \frac{1}{n^2}E[\text{Var}\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i] + \text{Var}[E(w(\xi_i,\xi_k)|\xi_i)]\\ &+ 3\text{Var}[E(w(\xi_i,\xi_k)|\xi_i)] + O\left(\frac{1}{n\rho_n}\right)\\ &= \frac{1}{n^2}E[\text{Var}\sum_{k\neq i} w(\xi_i,\xi_k)|\xi_i] + 4\text{Var}[E(w(\xi_i,\xi_k)|\xi_i)] + O\left(\frac{1}{n\rho_n}\right)\end{aligned} \tag{A.22}$$

The difference between Eqs A.21 and A.22 is:

$$(n-1)E(Z_{n,i}-\bar{Z}_n)^2-(n-1)\text{Var } Z_{n-1} = \frac{3}{n^2}E[\text{Var}\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i]+O\left(\frac{1}{n\rho_n}\right). \tag{A.23}$$

Note that, we also have:

$$\frac{1}{n^2}E[\text{Var}\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i] = \frac{1}{n}E[\text{Var}(w(\xi_i,\xi_k)|\xi_i)] = O\left(1/n\right) \tag{A.24}$$

Eq A.24 establishes Eq **??**. Furthermore, in conjunction with Eqs A.19 and A.18, it also shows that both $(n-1)E\sum_{i=1}^{n}(Z_{n,i}-\overline{Z_n})^2$ and $(n-1)\text{Var } Z_{n-1}$ converge to positive constants. This concludes our proof. □

We now present the proofs of Lemmas D.1 and D.2.

Proof of Lemma D.1. Applying law of total variance,

$$\sum_{i=1}^{n-1} \text{Var}\left(\frac{D_i^{(n)}}{\gamma_n}\right) = \sum_{i=1}^{n-1} \text{Var}\left[E\left(\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right)\right] + \sum_{i=1}^{n-1} E\left[\text{Var}\left(\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right)\right]. \quad \text{(A.25)}$$

We now show that the second term on the RHS of the above equation is small.

$$\begin{aligned} \sum_{i=1}^{n-1} E\left[\text{Var}\left(\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right)\right] &= \sum_{i=1}^{n-1} E\left[\text{Var}\left(\frac{\sum_{j\neq i} A_{ij}}{\binom{n}{2}\rho_n}\middle|\xi\right)\right] \\ &= \sum_{i=1}^{n-1} E\left(\frac{\sum_{j\neq i}\rho_n w(\xi_i,\xi_j)(1-\rho_n w(\xi_i,\xi_j))}{\binom{n}{2}^2\rho_n^2}\right) \\ &\asymp \sum_{i,j,i\neq j} \frac{\rho_n E[w(\xi_i,\xi_j)]}{n^4\rho_n^2} = O(n^{-2}\rho_n^{-1}) \end{aligned} \quad \text{(A.26)}$$

For the first term on the RHS of Eq A.25, for any fixed i, we have:

$$\begin{aligned} &\text{Var}\left(E\left[\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right]\right) = \text{Var}E\left(\frac{\sum_{k,k\neq i} A_{ik}}{\frac{(n-1)(n-2)}{2}\rho_n}\middle|\xi\right) \asymp \frac{4}{n^4}\text{Var}\left(\sum_{k,k\neq i} w(\xi_i,\xi_k)\right) \\ &\asymp \frac{4}{n^4}E\left(\text{Var}\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i\right) + \frac{4}{n^4}\text{Var}\left(E\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i\right). \end{aligned} \quad \text{(A.27)}$$

Exchanging the sum and expectation in the second term, we can also write,

$$\frac{4}{n^4}\text{Var}\left(E\sum_{k,k\neq i} w(\xi_i,\xi_k)|\xi_i\right) = \frac{4}{n^2}\text{Var}[E(w(\xi_i,\xi_k)|\xi_i)]. \quad \text{(A.28)}$$

Since Eq A.25 involves a sum over $n-1$ identical terms, owing to the fact that $\{\xi_i\}$ are i.i.d, we get the result by multiplying Eq A.27 and A.28 by $n-1$. $\square$

Proof of Lemma D.2. We decompose the covariance into

$$\begin{aligned} \sum_{i,j,i\neq j} \text{cov}\left(\frac{D_i^{(n)}}{\gamma_n}, \frac{D_j^{(n)}}{\gamma_n}\right) &= \sum_{i,j,i\neq j} \text{cov}\left(E\left[\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right], E\left[\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right]\right) \\ &+ \sum_{i,j,i\neq j} E\left[\text{cov}\left(\frac{D_i^{(n)}}{\gamma_n}, \frac{D_j^{(n)}}{\gamma_n}\middle|\xi\right)\right]. \end{aligned} \quad \text{(A.29)}$$

The second term on the RHS of the above equation is small as shown before.

$$
\begin{aligned}
&\sum_{i,j,i\neq j} E\left[\operatorname{cov}\left(\frac{D_i^{(n)}}{\gamma_n}, \frac{D_j^{(n)}}{\gamma_n}\middle|\xi\right)\right] \\
&= \sum_{i,j,i\neq j} E\left[\operatorname{cov}\left(\frac{\sum_{k,k\neq i} A_{ik}}{\gamma_n}, \frac{\sum_{s,s\neq j} A_{js}}{\gamma_n}\middle|\xi\right)\right] \\
&\asymp \frac{1}{n^4\rho_n^2}\sum_{i,j} E[\operatorname{Var}(A_{ij}|\xi)] \\
&\asymp \frac{1}{n^2\rho_n^2}\rho_n E[w(\xi_i,\xi_j)] = O(n^{-2}\rho_n^{-1})
\end{aligned}
$$

For the first term in Eq A.29, for any fixed i and j, we have

$$
\begin{aligned}
&\operatorname{cov}\left(E\left[\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right], E\left[\frac{D_j^{(n)}}{\gamma_n}\middle|\xi\right]\right) \\
&= \operatorname{cov}\left(\frac{\sum_k^{k\neq i} w(\xi_i,\xi_k)\rho_n}{\frac{(n-1)(n-2)}{2}\rho_n}, \frac{\sum_s^{s\neq j} w(\xi_j,\xi_s)\rho_n}{\frac{(n-1)(n-2)}{2}\rho_n}\right) \\
&\asymp \frac{4}{n^4}\operatorname{cov}\left(\sum_{k,k\neq i} w(\xi_i,\xi_k), \sum_{s,s\neq j} w(\xi_j,\xi_s)\right) \\
&= \frac{4}{n^4}\sum_{k,k\neq i}\sum_{s,s\neq j}\operatorname{cov}(w(\xi_i,\xi_k), w(\xi_j,\xi_s)).
\end{aligned}
\tag{A.30}
$$

Let $S_i = \{i,k\}$, and $S_j = \{j,s\}$ be two pairs containing i and j respectively. Some algebraic manipulation yields,

$$
\begin{aligned}
\sum_{k,k\neq i}\sum_{s,s\neq j}\operatorname{cov}(w(\xi_i,\xi_k), w(\xi_j,\xi_s)) = &\sum_{|S_i\cap S_j|=1}\operatorname{cov}(w(\xi_i,\xi_k), w(\xi_j,\xi_s)) \\
&+ \sum_{|S_i\cap S_j|=2}\operatorname{cov}(w(\xi_i,\xi_k), w(\xi_j,\xi_s)).
\end{aligned}
\tag{A.31}
$$

In the R.H.S of the above expression, the second summation has $n(n-1)$ terms, whereas the first has $n(n-1)(n-2)$ terms. Furthermore, for $|S_i \cap S_j| = 2$, it is easy to see that $\operatorname{cov}(w(\xi_i,\xi_k), w(\xi_j,\xi_s))$ is simply the variance of $\operatorname{Var}(w(\xi_i,\xi_k))$ which is positive. For $|S_i \cap S_j| = 1$, W.L.O.G. let $S_i = \{i,u\}$ and $S_j = \{j,u\}$. Conditioned on the shared node ξ_u,

$$
\begin{aligned}
\operatorname{cov}(w(\xi_i,\xi_u), w(\xi_j,\xi_u)) &= \operatorname{cov}[E(w(\xi_i,\xi_u)|\xi_u), E(w(\xi_j,\xi_u)|\xi_u)] \\
&= \operatorname{Var}(Ew(\xi_i,\xi_u)|\xi_u)
\end{aligned}
\tag{A.32}
$$

which is also positive. Hence the contribution of the first sum is of a larger order.

Now we enumerate all the ways in which S_i and S_j can have a node in common, with the constraint of $i \neq j$. For any fixed i and j, s.t. $i \neq j$, $|S_i \cap S_j| = 1$ means that there is 1 common node in $S_i = \{i, k\}$ and $S_j = \{j, s\}$. There are three possible cases, $i = s$, $k = j$, $k = s$. Thus, Eq A.30 can be expanded as (W.L.O.G, suppose $i = s$),

$$\begin{aligned}\text{cov}\left(E\left[\frac{D_i^{(n)}}{\gamma_n}\middle|\xi\right], E\left[\frac{D_j^{(n)}}{\gamma_n}\middle|\xi\right]\right) &\asymp \frac{4}{n^4}[3(n-2)\text{cov}(w(\xi_i, \xi_k), w(\xi_j, \xi_i))] \\ &= \frac{4}{n^3} \times 3\text{cov}(w(\xi_i, \xi_k), w(\xi_j, \xi_i)) \\ &\overset{(i)}{=} \frac{4}{n^3} \times 3\text{Var}(E(w(\xi_i, \xi_k))|\xi_i) \qquad \text{(A.33)}\end{aligned}$$

Step (i) uses an analogous argument from Eq A.32, and conditions on ξ_i.

Eq A.29 involves a sum over all (i, j) pairs, $i \neq j$, , owing to the fact that $\{\xi_i\}$ are i.i.d, we get the result by multiplying Eq A.33 by $n(n-1)$. □

E Proof of Proposition 2

Before we state the proof of our result, recall the following well-known relationship between uniform integrability and convergence of moments. See for example, Theorem 25.12 of Billingsley (1995).

Proposition E.1. *Suppose that $X_n \rightsquigarrow X$ and $\{X_n\}_{n\geq 1}$ is uniformly integrable. Then, $E(X_n) \to E(X)$.*

Now we will prove our proposition below:

Proof. In what follows let $X_n := \tau_n[\hat{\theta}_n - E(\hat{\theta}_n)]$ and $V_n = \tau_n \cdot U_n$. Recall that $U_n = \hat{\theta}_n - \theta$. While our result here is more general, in a jackknife context, $\hat{\theta}_n = Z_n$ following the notation that we use elsewhere. Consider the following decomposition:

$$\tau_n[\hat{\theta}_n - E(\hat{\theta}_n)] = \tau_n[\hat{\theta}_n - \theta] + E(\tau_n[\theta - \hat{\theta}_n])$$

Since $\{V_n^2\}_{n\geq 1}$ is uniformly integrable, it follows that $\{V_n\}_{n\geq 1}$ is also uniformly integrable. Therefore, by Proposition E.1, $E(\tau_n[\theta - \hat{\theta}_n]) \to 0$. By Slutsky's Theorem, it follows that $\tau_n[\hat{\theta}_n - E(\hat{\theta}_n)] \rightsquigarrow U$.

To show that the variances converge to the same value, observe that $E(X_n^2)$ is given by:

$$E(X_n^2) = E(V_n^2) - (E(V_n))^2$$

First, $V_n^2 \rightsquigarrow U^2$ by continuous mapping theorem. Since $\{V_n^2\}_{n\geq 1}$ is uniformly integrable, $E(V_n^2) \to E(U^2)$ by Proposition E.1 again. Finally, $(EV_n)^2 \to 0$ and the result follows. □

F Additional theory

It should be noted that a similar inequality for a closely related procedure has an even simpler proof. This alternative procedure does not require the functional to be invariant to node permutation and allows flexibility with the leave-one-out estimates. However, the resulting estimate is often not sharp. More concretely, let Z_n denote a function of $A^{(n)}$ and let $\widetilde{Z}_{n,i}$ be an arbitrary functional calculated on a graph with node i removed. Consider the following estimator:

$$\widehat{\text{Var}}_{\text{JACK}}\ Z_n = \sum_{i=1}^{n} (Z_n - \widetilde{Z}_{n,i})^2 \tag{A.34}$$

Combining the aforementioned filtration with arguments in Boucheron et al. (2004) leads to the following inequality:

Proposition F.1 (Network Efron-Stein, alternative version)**.**

$$\text{Var}\ Z_n \leq E(\widehat{\text{Var}}_{\text{JACK}}\ Z_n) \tag{A.35}$$

G Additional data analysis results

We first present Tables A.1 and A.2 with details of the networks we used in our real data experiments in Section 4 of the main paper.

Table A.1: Details of college networks for first real data experiment (see Figure 3 of main paper)

	Caltech	Williams	Wellesley
Nodes	769	2790	2970
Edges	16656	112986	94899
Average Degree	43.375	63.927	81.023

Table A.2: Details of college networks for second real data experiment (see Figure 4 of main paper)

	Berkeley	Stanford	Yale	Princeton	Harvard	MIT
Nodes	22937	11621	8578	6596	15126	6440
Edges	852444	568330	405450	293320	824617	251252
Average Degree	74.332	97.819	94.544	88.952	109.040	78.040

For our real data experiments, (Section 4 of main paper) we compared subsampling with jackknife on the three colleges (see Figure 3). For simplicity, for the second experiment comparing three pairs of college networks (see Figure 4), we only showed the confidence intervals obtained using jackknife. Here, in Figure A.1, for completeness, we present confidence intervals for test sets constructed from the six college networks using both jackknife and subsampling with different choices of b. This again shows that jackknife CI's mostly are in agreement with those obtained from subsampling.

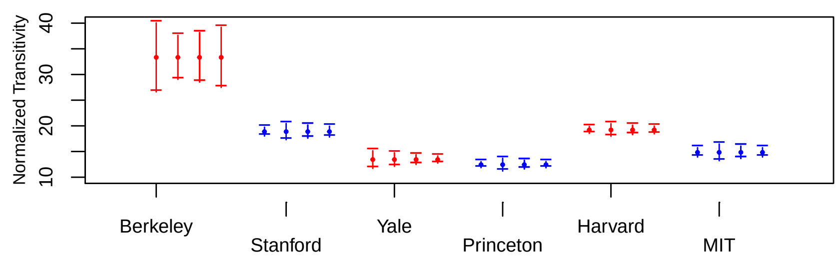

Figure A.1: Confidence intervals of subsampling and jackknife in calculating triangle, two-star densities and normalized transitivity in the example of six college Facebook networks test sets. The four CIs for each college are in the order of jackknife, subsampling with b=0.05n, b=0.1n, and b=0.2n respectively.

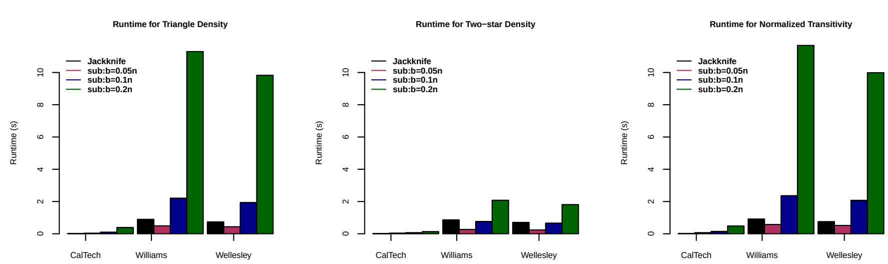

Figure A.2: Computation time of jackknife compared to subsampling in calculating triangle, two-star densities and normalized transitivity in the example of three college Facebook networks.

In addition, we show the timing results our real data experiments. Figure A.2 shows computation time of the three college example of Facebook network data (see Figure 3). We demonstrate the triangle, two-star densities and normalized

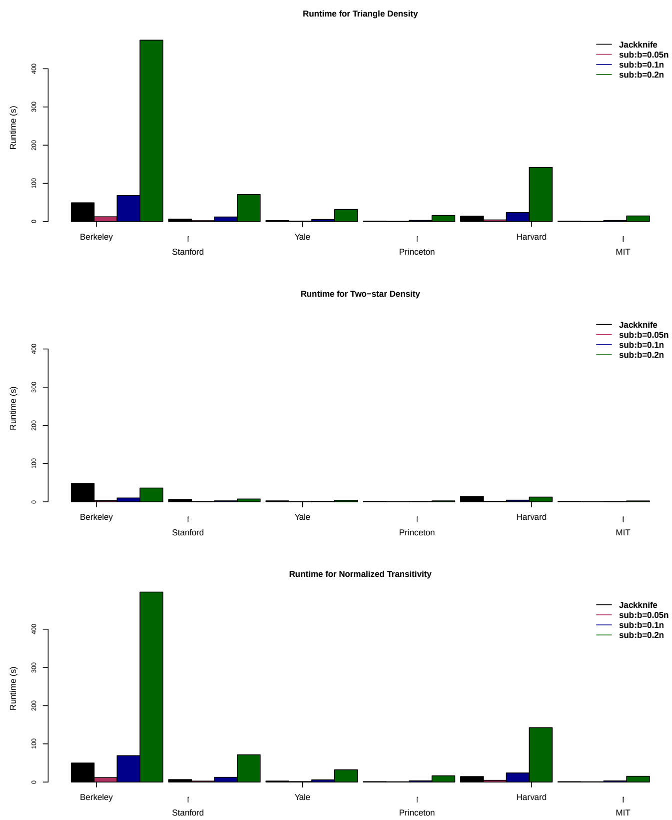

Figure A.3: Computation time of jackknife compared to subsampling in calculating triangle, two-star densities and normalized transitivity in the example of six college Facebook networks test sets.

transitivity variance computation time using jackknife and subsampling with $b = 0.05n$, $b = 0.1n$ and $b = 0.2n$, $B = 1000$ in each college network.

In Figure A.3, we show the computation time of variance estimation for the same statistics on the test sets for the same set of algorithms. Since we split training and test set in half, the training sets have approximately the same time.

These figures show that, it is possible to implement jackknife in a computationally efficient manner when there is nested structure in the subgraph counts. In all these cases, we see that for the larger networks, subsampling with large b is often considerably slower than jackknife.

References

Arvesen, J. N. Jackknifing U-statistics. *The Annals of Mathematical Statistics*, 40(6):2076–2100, 1969. 7

Bickel, P. J., Chen, A., and Levina, E. The method of moments and degree distributions for network models. *Annals of Statistics*, 39:38–59, 2011. 4, 6, 8

Billingsley, P. *Probability and Measure*. John Wiley and Sons, New York, 1995. 13

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801 – 1851, 2008. 1

Boucheron, S., Lugosi, G., and Bousquet, O. Concentration inequalities. In Bousquet, O., von Luxburg, U., and Rätsch, G. (eds.), *Advanced Lectures in Machine Learning*, pp. 208–40. Springer, New York, 2004. 14

Lee, A. J. *U-statistics: Theory and Practice*. CRC Press, Boca Raton, Florida, 1990. 4, 5

Serfling, R. J. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980. 4

Van der Vaart, A. W. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. 8