ONE TERM EDGEWORTH EXPANSION FOR FINITE POPULATION U–STATISTICS OF DEGREE TWO

M. Bloznelis[1], F. Götze

Vilnius University and Bielefeld University

Abstract. By means of Hoeffding's decomposition we represent a finite population $U-$ statistic of degree two by the sum of a linear and a quadratic part. Assuming that the linear part is nondegenerate we prove the validity of one term Edgeworth expansion for the distribution function of the statistic under the optimal (minimal) conditions on the linear part and $2+\delta$ moment condition on the quadratic part. No condition is imposed on the ratio N/n, where N respectively n denote the sample size respectively the population size.

1. Introduction and results

Let $X_1,\dots,X_N$ be a simple random sample drawn without replacement from the set $\mathcal{X}=\{x_1,\dots,x_n\}$, where $n>N$. Given a function $h:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ define the random variable

$$U=\sum_{1\le i<j\le N} h(X_i,X_j).$$

We assume that $\mathbf{E}U=0$ and that the function h is symmetric, i.e., $h(x,y)=h(y,x)$. The statistic U can be decomposed as

$$U=L+Q, \tag{1.1}$$

where $L=\sum_{1\le i\le N}\psi(X_i)$ and $Q=\sum_{1\le i<j\le N}\phi(X_i,X_j)$ denote the linear and the quadratic part of U respectively. Here

$$\psi(x)=(N-1)\,t(x),\qquad t(x)=\frac{n-1}{n-2}\,\mathbf{E}\big(h(X_1,\,X_2)\,|\,X_1=x\big)$$

1991 *Mathematics Subject Classification.* Primary 62E20; secondary 60F05.

Key words and phrases. Edgeworth expansion, finite population, $U-$ statistic.

[1]Research supported by the Alexander von Humboldt Foundation

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

and $\phi(x,\ y) = h(x,\ y) - t(x) - t(y)$. Note that $\mathbf{E}U = 0$ implies $\mathbf{E}h(X_1, X_2) = 0$ and $\mathbf{E}\psi(X_1) = 0$. Moreover, a simple calculation shows that

$$\text{(1.2)} \qquad \mathbf{E}\big(\phi(X_i, X_j)\,|X_k = x\big) = 0, \qquad \forall\, x \in \mathcal{X}, \qquad \forall\, k = 1, \dots, N.$$

In particular, the random variables L and Q are uncorrelated.

Write

$$\tau^2 = npq, \qquad p = N/n, \qquad q = 1 - p.$$

Assume that $\sigma^2 := \operatorname{Var}\psi(X_1) > 0$. Then, for large τ, the distribution of $L/(\sigma\tau)$ can be approximated by the standard normal distribution, using the central limit theorem, see Erdős and Rényi (1959). We approximate the distribution function

$$F(x) = \mathbf{P}\{U \le \tau\sigma x\}$$

by an one term Edgeworth expansion

$$\text{(1.3)} \qquad G(x) = G_U(x) = \Phi(x) - \frac{(q-p)\,\alpha + 3\,\kappa}{6\,\tau}\,\Phi^{(3)}(x),$$

$$\text{(1.4)} \qquad \alpha = \sigma^{-3}\mathbf{E}\psi^3(X_1), \qquad \kappa = \sigma^{-3}\tau^2\mathbf{E}\phi(X_1, X_2)\psi(X_1)\psi(X_2),$$

where $\Phi(x)$ denotes the standard normal distribution function, and $\Phi^{(3)}$ denotes the third derivative of Φ.

We show in Theorem 1 below that under very mild conditions on L and Q the remainder of this approximation is of order $o(\tau^{-1})$ as $\tau \to \infty$.

In order to provide a precise formulation we introduce a sequence of sets $\mathcal{X}_n = \{x_{n,1}, \dots, x_{n,n}\}$ and a sequence of symmetric functions $h_n : \mathcal{X}_n \times \mathcal{X}_n \to \mathbb{R}$, $n = 5, 6, \dots$. Let $X_{n,1}, \dots, X_{n,N_n}$ denote a random sample of size N_n $(N_n < n)$ drawn without replacement from the set $\mathcal{X}_n$ and define

$$U_n = \sum_{1 \le i < j \le N_n} h_n(X_{n,i}, X_{n,j}).$$

Write $p_n = N_n/n$, $q_n = 1 - p_n$ and $\tau_n = (np_nq_n)^{1/2}$. Given n we assume that $\mathbf{E}U_n = 0$ and decompose $U_n = L_n + Q_n$, cf. (1.1). The functions ψ_n, ϕ_n and moments σ_n, α_n, κ_n are defined as ψ, ϕ and σ, α, κ above, but with respect to the statistic U_n. Assume that $\sigma_n > 0$ and write $F_n(x) = \mathbf{P}\{U_n \le \sigma_n\tau_n x\}$. Let $G_n(x) = G_{U_n}(x)$ denote the one term Edgeworth expansion of $F_n(x)$, cf (1.3). Introduce the moments

$$\beta_n = \sigma_n^{-3}\mathbf{E}|\psi_n(X_{n,1})|^3, \qquad \gamma_{\delta,n} = \mathbf{E}\big|\tau_n^2\sigma_n^{-1}\phi_n(X_{n,1}, X_{n,2})\big|^{2+\delta},$$

$$\beta_n^\star(\varepsilon) = \sigma_n^{-3}\mathbf{E}|\psi_n(X_{n,1})|^3\mathbb{I}\{\psi_n^2(X_{n,1}) \ge \varepsilon^2\sigma_n^2\tau_n^2\}, \qquad \varepsilon > 0.$$

For a random variable Z, with finite and positive variance $\sigma_Z^2 = \operatorname{Var} Z$, write

$$\rho_Z(A, B) = 1 - \sup\big\{\big|\mathbf{E}\exp\big\{it\,\frac{Z}{\sigma_Z}\,\big\}\big| : A \le |t| \le B\big\}, \qquad 0 < A < B.$$

Theorem 1. *Assume that $\tau_n \to \infty$ as $n \to \infty$. Suppose that for some $\delta > 0$,*

$$\limsup_n \gamma_{\delta,n} < \infty, \tag{1.5}$$

$$\limsup_n \beta_n < \infty, \tag{1.6}$$

$$\forall\, \varepsilon > 0, \qquad \limsup_n \beta_n^\star(\varepsilon) = 0, \tag{1.7}$$

$$\forall\, \varepsilon > 0, \quad \forall\, B > 1, \qquad \liminf_n \rho_{\psi_n(X_{n,1})}(\varepsilon, B) > 0. \tag{1.8}$$

Then $\quad \Delta_n := \sup_x |F_n(x) - G_n(x)| = o(\tau_n^{-1}) \quad$ *as* $\quad n \to \infty.$

The corresponding result in the case of independent and identically distributed observations was proved by Bickel, Götze and van Zwet (1986).

Note that for asymptotically normal finite population statistics, τ_n plays the same role as $\sqrt{N}$ does in the classical case of i.i.d. observations. In particular, if we assume for simplicity that $\operatorname{Var} L_n = 1$, then the quadratic part Q_n is typically of order $O_P(\tau_n^{-1})$ as $\tau_n \to \infty$.

The main goal of Theorem 1 is to establish the validity of an expansion under minimal conditions on the linear part of the statistic and *without any restriction on the ratio* $p_n = N_n/n$. Indeed the condition (1.8) corresponds to the well known "non-latticeness" condition, the weakest smoothness condition which together with the third moment condition is sufficient for the validity of one term Edgeworth expansion in the classical case of sums of i.i.d. random variables, see, e.g., Petrov (1995). Clearly, (1.6) and (1.7) correspond to the third moment condition in the i.i.d. case. Even more important, is that Theorem 1 applies to an arbitrary sequence of positive integers N_n such that $N_n < n$ and $\tau_n^2 = N_n(n-N_n)n^{-1} \to \infty$.

Clearly, the $2+\delta$ moment condition (1.5) is not optimal.

The asymptotic normality of finite population $U-$ statistics was studied by Nandi and Sen (1963) and later by Zhao and Chen (1990). Berry-Esseen type bounds were established by Zhao and Chen (1990), Kokic and Weber (1990) and as a particular case of more general multivariate combinatorial statistics by Bolthausen and Götze (1992). One term Edgeworth expansions were studied by Kokic and Weber (1990). The validity of the one term expansion with remainder $O(N^{-1})$ was proved by Bloznelis and Götze (1997).

2. Proof

Before proving Theorem 1 we introduce some notation and formulate some auxiliary lemmas.

For $k = 1, 2 \ldots,$ denote $\Omega_k = \{1, \ldots, k\}$. Write $\Omega = \Omega_n$ and $\mathcal{A}^c = \Omega \setminus \mathcal{A}$, for $\mathcal{A} \subset \Omega$. For a complex valued function g, write $\|g(x)\| = \sup_{x\in\mathbb{R}} |g(x)|$. By c we denote a positive absolute constant.

Let $(X_1, \dots, X_n)$ denote a random permutation of the ordered set $(x_1, \dots, x_n)$ of elements of $\mathcal{X}$ which is uniformly distributed over the set of all such permutations. For any $f : \mathcal{X} \to [0, +\infty)$ and any $\{i_1, \dots, i_k\} \subset \Omega$, the following inequality holds

$$\mathbf{E}\big(f(X_i)|X_{i_1}, \dots, X_{i_k}\big) \le \frac{n}{n-k}\,\mathbf{E} f(X_i), \qquad \forall i \in \Omega \setminus \{i_1, \dots, i_k\}. \tag{2.1}$$

Write $\psi'(x) = -\psi(x)$ and define the statistic

$$U' = L' + Q', \qquad L' = \sum_{1 \le i \le N'} \psi'(X_i'), \qquad Q' = \sum_{1 \le i < j \le N'} \phi(X_i', X_j'),$$

based on the sample $X_1', \dots, X_{N'}'$, where $X_i' = X_{N+i}$ and $N' = n - N$. Write $p' = N'/n$, $q' = 1 - p'$, $\tau' = (np'q')^{1/2}$ and let $G_{U'}(x)$ denote the one term Edgeworth expansion of the probability $\mathbf{P}\{U' \le \sigma'\tau' x\}$, cf. (1.3), where $(\sigma')^2 = \mathbf{E}(\psi'(X_1'))^2$ and where the moments α', κ' are defined by means of (1.4) with respect to the random variables $\psi'(X_i')$ and $\phi(X_i', X_j')$.

Lemma 2.1. *We have $\tau' = \tau$, $\sigma' = \sigma$, $G_{U'}(x) = G(x)$, for every $x \in \mathbb{R}$, and $U' = U$ almost surely.*

Proof of Lemma 2.1. The identity $U' = U$ follows from $L = L'$ and $Q = Q'$ almost surely. The last two identities are consequences of $\mathbf{E}\Psi(X_1) = 0$ and (1.2) respectively.

Let us prove Theorem 1. We shall assume without loss of generality that $\sigma_n^2 = \tau_n^{-2}$, for every n.

Note that condition (1.5) is used only to prove (3.11) below. In the remaining steps of the proof we use the weaker condition

$$\limsup_n \gamma_n < \infty, \qquad \gamma_n = \gamma_{0,n}. \tag{2.2}$$

Given n, write $\psi_n(X_{n,i}) = \hat{X}_i + \tilde{X}_i$, where

$$\hat{X}_i = \psi_n(X_{n,i})\mathbb{I}\{\psi_n^2(X_{n,i}) \le 1\}, \quad \tilde{X}_i = \psi_n(X_{n,i})\mathbb{I}\{\psi_n^2(X_{n,i}) > 1\}, \quad 1 \le i \le N_n.$$

Note that for large n, $\sigma_x^2 := \operatorname{Var} \hat{X}_1 > 0$. To prove this inequality write

$$\sigma_x^2 = \mathbf{E}\hat{X}_1^2 - (\mathbf{E}\hat{X}_1)^2 = \sigma_n^2 - \mathbf{E}\tilde{X}_1^2 - (\mathbf{E}\hat{X}_1)^2 = \tau_n^{-2} - o(\tau_n^{-3}).$$

Here we used $|\mathbf{E}\hat{X}_1| = |\mathbf{E}\tilde{X}_1| \le \mathbf{E}|\tilde{X}_1|^3 = o(\tau_n^{-3})$, and $\mathbf{E}\tilde{X}_1^2 \le \mathbf{E}|\tilde{X}_1|^3 = o(\tau_n^{-3})$, by (1.7). We shall assume that n is large enough so that $\sigma_x > 0$.

Introduce the functions $g_1 : \mathcal{X}_n \to \mathbb{R}$ and $g_2 : \mathcal{X}_n \times \mathcal{X}_n \to \mathbb{R}$,

$$g_1(x) = \sigma_x^{-1}\tau_n^{-1}\big(\psi_n(x)\mathbb{I}\{\psi_n^2(x) \le 1\} - \mathbf{E}\hat{X}_1\big), \qquad g_2(x,y) = \sigma_x^{-1}\tau_n^{-1}\phi_n(x,y).$$

Observe that $\mathbf{E}g_1(X_{n,1}) = 0$ and $\mathbf{E}g_1^2(X_{n,1}) = \tau_n^{-2}$. Write

$$U_n^* = \sum_{1\le i\le N_n} g_1(X_{n,i}) + \sum_{1\le j<k\le N_n} g_2(X_{n,j}, X_{n,k}), \qquad F_n^* = \mathbf{P}\{U_n^* \le x\},$$

$$G_n^*(x) = \Phi(x) - \frac{(q_n - p_n)\alpha_n^* + 3\kappa_n^*}{6\tau_n}\Phi^{(3)}(x), \qquad \alpha_n^* = \tau_n^3\mathbf{E}g_1(X_{n,1})^3,$$

$$\kappa_n^* = \tau_n^5\mathbf{E}g_1(X_{n,1})g_1(X_{n,2})g_2(X_{n,1}, X_{n,2}), \quad \beta_{k,n}^* = \tau_n^k\mathbf{E}|g_1(X_{n,1})|^k, \quad k = 3, 4.$$

Lemma 2.2. *Assume that $\sigma_n^2 = \tau_n^{-2}$, for every n, and $\tau_n \to \infty$ as $n \to \infty$. Assume that (1.6)-(1.8) and (2.2) hold. Then for all $\varepsilon > 0$ as $n \to \infty$*

$$\sigma_x\tau_n = 1 + o(\tau_n^{-1}), \qquad \tau_n\sigma_x^{-1}\mathbf{E}\hat{X}_1 = o(\tau_n^{-1}), \tag{2.3}$$

$$\alpha_n^* = \alpha_n + o(1), \qquad \kappa_n^* = \kappa_n + o(1), \tag{2.4}$$

$$\beta_{3,n}^* = \beta_n + o(1), \qquad \beta_{4,n}^* \le (\tau_n + o(1))\big(\varepsilon\beta_n + \beta_n^\star(\varepsilon)\big) + o(\tau_n^{-2}), \tag{2.5}$$

$$\text{for all} \quad B > 1, \qquad \liminf_n \rho_{g_1(X_{n,1})}(\varepsilon, B) > 0 \tag{2.6}$$

The proof of Lemma 2.2 is elementary.

Note that, by (2.2) and (1.6), the sequences β_n, κ_n and α_n are bounded. In view of (2.4), (2.5) the sequences α_n^*, $\beta_{3,n}^*$ and κ_n^* are bounded too. In particular, there exists a number, say D, independent of n such that

$$\max\{|\alpha_n|, |\beta_n|, |\kappa_n|, |\alpha_n^*|, |\beta_{3,n}^*|, |\kappa_n^*|\} \le D, \qquad n = 5, 6, \ldots. \tag{2.7}$$

We may choose $D > 1$. Now, it follows from (2.7) that

$$\Big\|\frac{d}{dx}G_n^*(x)\Big\| \le 6D, \quad n = 5, 6, \ldots. \tag{2.8}$$

Lemma 2.3 (Truncation). *Assume that $\sigma_n^2 = \tau_n^{-2}$, $p_n \le 1/2$, for every n, and $\tau_n \to \infty$ as $n \to \infty$. Assume that (1.6)-(1.7) and (2.2) hold. Then as $n \to \infty$*

$$\Delta_n \le \Delta_n^* + o(\tau_n^{-1}), \qquad \text{where} \qquad \Delta_n^* := \sup_x |F_n^*(x) - G_n^*(x)|.$$

Proof of Lemma 2.3. Observe that $p_n \le 1/2$ implies $N_n \le 2\tau_n^2$. We have

$$\mathbf{P}\{L \ne \hat{X}_1 + \cdots + \hat{X}_{N_n}\} \le \sum_{1\le i\le N_n}\mathbf{P}\{\psi_n^2(X_{n,i}) > 1\} = N_n\mathbf{P}\{\psi_n^2(X_{n,1}) > 1\}$$

$$\le N_n\tau_n^{-3}\beta_n^\star(1) = o(\tau_n^{-1}), \tag{2.9}$$

by Chebyshev's inequality and (1.7). The identity

$$\mathbf{P}\{\hat X_1+\cdots+\hat X_{N_n}+Q\le x\}=F_n^*(x+\eta),\qquad \eta=(\sigma_x^{-1}\tau_n^{-1}-1)x-\sigma_x^{-1}\tau_n^{-1}N_n\mathbf{E}\hat X_1,$$

in combination with (2.9) implies $\|F_n(x)-F_n^*(x+\eta)\|=o(\tau_n^{-1})$. Furthermore, by the triangle inequality,

$$\Delta_n\le\Delta_n^*+\|F_n(x)-F_n^*(x+\eta)\|+\delta_n,\qquad \delta_n:=\|G_n^*(x+\eta)-G_n(x)\|.$$

In order to show $\delta_n=o(\tau_n^{-1})$ write $\delta_n\le\delta_{n,1}+\delta_{n,2}$, where

$$\delta_{n,1}=\|G_n(x+\eta)-G_n(x)\|,\qquad\qquad \delta_{n,2}=\|G_n^*(x+\eta)-G_n(x+\eta)\|.$$

By (2.4), $\delta_{n,2}=o(\tau_n^{-1})$. In order to show $\delta_{n,1}=o(\tau_n^{-1})$ we use the bound $|\eta|=|x|o(\tau_n^{-1})+o(\tau_n^{-1})$, see (2.3), and apply the mean value theorem to the function $G_n(x)$, whose derivative satisfies $|\frac{d}{dx}G_n(x)|\le 2D(1+|x|^3)\exp\{-x^2/2\}$, by (2.7).

Proof of Theorem 1. In view of Lemma 2.1, we may assume without loss of generality that $p_n\le 1/2$, for every n. By Lemma 2.3 it suffices to prove $\Delta_n^*=o(\tau_n^{-1})$. To this aim we shall show that for every $\varepsilon\in(0,1)$

$$\limsup_n \Delta_n^*\tau_n\le\varepsilon. \tag{2.10}$$

Fix $\varepsilon\in(0,1)$. Let m_n denote the integer closest to the number $4\rho_n^{-1}\ln\tau_n$. Here $\rho_n=\rho_{g_1(X_{n,1})}\big(0.001/\beta_{3,n}^*,\,\varepsilon^{-1}\big)$ satisfies

$$\liminf_n \rho_n>0, \tag{2.11}$$

by (2.6) and (2.7). Since $m_n=o(\tau_n)$ as $n\to\infty$, we have $m_n<\tau_n^2/2\le N_n/2$, for sufficiently large n. We shall assume that n is large enough so that these inequalities hold.

Split $U_n^*=U_n^\star+\Lambda$, where $\Lambda=\sum_{1\le i<j\le m_n}g_2(X_{n,i},X_{n,j})$, and $U_n^\star=U_n^*-\Lambda$. Write $F_n^\star=\mathbf{P}\{U_n^\star\le x\}$. Using (1.2), it is easy to show, see, e.g., Lemma 5.1 in Bloznelis and Götze (1997), that,

$$\mathbf{E}\Lambda^2\le m_n^2\mathbf{E}g_2^2(X_{n,1},X_{n,2})=O(\tau_n^{-6}\ln^2\tau_n), \tag{2.12}$$

where in the last step we used (2.2), (2.3) and (2.11). Now we are going to apply Slutzky's argument. We have

$$\Delta_n^*\le\|F_n^\star(x)-G_n^*(x)\|+\tau_n^{-2}\|\frac{d}{dx}G_n^*(x)\|+\mathbf{P}\{|\Lambda|\ge\tau_n^{-2}\}, \tag{2.13}$$

where $\mathbf{P}\{|\Lambda| \geq \tau_n^{-2}\} \leq \tau_n^4 \mathbf{E}\Lambda^2 = o(\tau_n^{-1})$, by Chebyshev's inequality and (2.12). It follows from (2.13) and (2.8) that

$$\Delta_n^* \leq \Delta_n^\star + o(\tau_n^{-1}), \qquad \Delta_n^\star = \|F_n^\star(x) - G_n^*(x)\|.$$

Therefore, it suffices to prove (2.10) with Δ_n^* replaced by $\Delta_n^\star$.

By the Berry-Esseen smoothing inequality, see, e.g., Feller (1971), 538 p.,

$$\Delta_n^\star \leq cJ_n + cK_n, \qquad J_n = \int_{|t|\leq T_n} \frac{|\hat{F}_n^\star(t) - \hat{G}_n^*(t)|}{|t|}\, dt, \quad K_n = \frac{1}{T_n} \| \frac{d}{dx} G_n^*(x)\|,$$

$T_n = \frac{\tau_n}{\varepsilon}$. Here $\hat{F}_n^\star(t)$ and $\hat{G}_n^*(t)$ denote the Fourier-Stieltjes transforms of $F_n^\star(t)$ and $G_n^*(t)$. We shall show that, for some absolute constant $c_1 > 0$,

$$J_n \leq c_1 \beta_n \tau_n^{-1} \varepsilon + o(\tau_n^{-1}). \tag{2.14}$$

This bound in combination with (2.7) and the inequality $\tau_n K_n \leq 6D\varepsilon$, see (2.8), implies $\limsup_n \tau_n \Delta_n^\star \leq (6 + c_1) D\varepsilon$ thus completing the proof of Theorem 1.

Given a complex valued function f and numbers $0 \leq a \leq b$, write

$$I_{a,b}(f) = \int_{a<|t|\leq b} \frac{|f(t)|}{|t|}\, dt.$$

We have $J_n = I_{0,T_n}(\hat{F}_N^\star - \hat{G}_N^*) \leq \mathbb{J}_1 + \cdots + \mathbb{J}_4$, where

$$\begin{aligned} &\mathbb{J}_1 = I_{0,H_n}(\hat{F}_n^\star - \hat{F}_n^*), \qquad &&\mathbb{J}_2 = I_{H_n,T_n}(\hat{G}_n^*), \qquad &&\mathbb{J}_3 = I_{H_n,T_n}(\hat{F}_n^\star), \\ &\mathbb{J}_4 = I_{0,H_n}(\hat{F}_n^* - \hat{G}_n^*), \qquad &&H_n = 10^{-3}\tau_n/\beta_{3,n}^*, \qquad &&\hat{F}_n^*(t) = \mathbf{E}\exp\{itU_n^*\}. \end{aligned}$$

In order to prove (2.14) we shall show that

$$\mathbb{J}_i = o(\tau_n^{-1}), \qquad \text{for} \quad 1 \leq i \leq 3, \qquad \text{and} \qquad \mathbb{J}_4 \leq c_1 \beta_n \tau_n^{-1} \varepsilon + o(\tau_n^{-1}). \tag{2.15}$$

Proof of $\mathbb{J}_1 = o(\tau_n^{-1})$. We have $|\hat{F}_n^*(t) - \hat{F}_n^\star(t)| \leq |t|\mathbf{E}|\Lambda|$. Therefore, by (2.12),

$$\mathbb{J}_1 \leq 2H_n \mathbf{E}|\Lambda| \leq 2H_n (\mathbf{E}\Lambda^2)^{1/2} = 2H_n O(\tau_n^{-3} \ln \tau_n) = o(\tau_n^{-1}).$$

Proof of $\mathbb{J}_2 = o(\tau_n^{-1})$. We have

$$\hat{G}_n^*(t) = \exp\{-t^2/2\} + \frac{(q_n - p_n)\alpha_n^* + 3\kappa_n^*}{6\tau_n} (it)^3 \exp\{-t^2/2\}.$$

By (2.7), $|\hat{G}_n^*(t)| \leq (1 + D|t|^3)\exp\{-t^2/2\}$. Using the exponential decay of $\hat{G}_n^*(t)$ as $t \to \infty$ and the fact that $\liminf_n H_n \tau_n^{-1} > 0$, see (2.7), we obtain $\mathbb{J}_2 = o(\tau_n^{-1})$.

Proof of $\mathbb{J}_3 = o(\tau_n^{-1})$. Denote $\mathcal{I}_0 = \{m_n + 1, \dots, N_n\}$ and $\mathcal{J}_0 = \Omega \setminus \mathcal{I}_0$. Let $\mathbb{A} = (A_1, \dots, A_n)$ be a random permutation of the ordered set $(x_{n,1}, \dots, x_{n,n})$ of elements of $\mathcal{X}_n$, which is uniformly distributed on the set of all such permutations. We split the sample as follows. For $j \in \mathcal{I}_0$, we put $X_{n,j} = A_j$. The observations $X_{n,1}, \dots, X_{n,m_n}$ are drawn without replacement from the (random) subpopulation $\{A_j,\, j \in \mathcal{J}_0\}$. We have

$$U_n^\star = \sum_{1 \le i \le m_n} \big(g_1(X_{n,i}) + v(X_{n,i})\big) + R, \qquad R = \sum_{k \in \mathcal{I}_0} g_1(A_k) + \sum_{j,k \in \mathcal{I}_0, j<k} g_2(A_j, A_k),$$

where given $A_{m_n+1}, \dots, A_{N_n}$, we write $v(x) = \sum_{k \in \mathcal{I}_0} g_2(x, A_k)$. Denote $f_0(t) = \mathbf{E}\big(\exp\{itU_n^\star\} \big| A_{m_n+1}, \dots, A_{N_n}\big)$ and write

$$\mathbb{I}_\star = \mathbb{I}\{\frac{\tau_n}{\varepsilon} v_\star < \frac{\rho_n}{10^2}\}, \qquad \text{where} \qquad v_\star = |\mathcal{J}_0|^{-1} \sum_{k \in \mathcal{J}_0} |v(A_k)|.$$

Observe that $\mathbb{I}_\star$ is a function of $A_{m_n+1}, \dots, A_{N_n}$. We have

$$\hat{F}_n^\star(t) = \mathbf{E} f_0(t) = \mathbf{E} f_0(t) \mathbb{I}_\star + r, \qquad r = r(t) = \mathbf{E} f_0(t)(1 - \mathbb{I}_\star).$$

Clearly, the bound $\mathbb{J}_3 = o(\tau_n^{-1})$ follows from

$$\text{(2.16)} \qquad I_{H_n,T_n}(r) = o(\tau_n^{-1}) \qquad \text{and} \qquad I_{H_n,T_n}(\mathbf{E} f_0 \mathbb{I}_\star) = o(\tau_n^{-1}).$$

To prove the first bound we shall show that $r = O(\tau_n^{-2})$ uniformly in $t \in [-T_n, T_n]$. We have $|r| \le \mathbf{E}|1 - \mathbb{I}_\star|$ and, by Chebyshev's inequality,

$$\mathbf{E}|1 - \mathbb{I}_\star| \le \tau_n^2 \varepsilon^{-2} \rho_n^{-2} 10^4 w_n, \qquad \text{where} \qquad w_n = \mathbf{E} v_\star^2.$$

By Cauchy-Schwartz and symmetry, $w_n \le \mathbf{E} v^2(A_1)$. Using (1.2) it is easy to show that $\mathbf{E} v^2(A_1) \le N_n \mathbf{E} g_2^2(A_1, A_2)$, see Bloznelis and Götze (1997), Lemma 5.1. Invoking (2.2) and (2.3), we obtain $w_n = O(\tau_n^{-4})$ and therefore, $r = O(\tau_n^{-2})$.

To prove the second part of (2.16) we represent the conditional characteristic function f_0 in Erdős–Rényi form. Let $\nu^* = (\nu_1^*, \dots, \nu_{n-N_n+m_n}^*)$ be i.i.d. Bernoulli random variables independent of $\mathbb{A}$ and having probabilities

$$\mathbf{P}\{\nu_1^* = 1\} = p^*, \quad \mathbf{P}\{\nu_1^* = 0\} = q^*, \quad p^* = m_n/(n - N_n + m_n), \quad q^* = 1 - p^*.$$

Write $L_* = \sum_{k \in \mathcal{J}_0} (g_1(A_k) + v(A_k)) \nu_k^*$ and $S_* = \sum_{k \in \mathcal{J}_0} (\nu_k^* - p^*)$. Since the distribution of $U_n^\star$ coincides with the conditional distribution of $L_* + R$ given the event $S_* = 0$, we can write

$$f_0(t) = \lambda_*^{-1} \int_{-\pi}^{\pi} W ds, \qquad W = \mathbf{E}\big(\exp\{it(L_* + R) + isS_*\} \big| A_{m_n+1}, \dots, A_{N_n}\big),$$

where $\lambda_* = 2\pi\mathbf{P}\{S_* = 0\}$ satisfies $\lambda_*^{-1} \le c m_n^{1/2}$, see Höglund (1978). It is shown in Bloznelis and Götze (1997), see formulas (3.14–17), that for $p^* < 1/8$,

$$|W|^2 \le W_1 W_2, \qquad W_1 = \prod_{k\in\mathcal{J}_0} u(A_k), \quad W_2 \le \bigl(1 + 4p^*q^*|t|v_\star\bigr)^{|\mathcal{J}_0|}, \tag{2.17}$$

where $u(x) = 1 - 2p^*q^*(1-\cos(tg_1(x)+s))$. The inequality $p^* < 1/8$ is satisfied for large n, since $m_n = O(\ln\tau_n)$ and $n - N_n + m_n \ge N_n \ge \tau_n^2$, by the inequality $p_n \le 1/2$. We shall assume that n is large enough so that (2.17) holds. Using Cauchy–Schwartz we obtain from (2.17)

$$\mathbf{E}W\mathbb{I}_\star \le (\mathbf{E}W_1)^{1/2}(\mathbf{E}W_2\mathbb{I}_\star)^{1/2}. \tag{2.18}$$

Invoking the inequality $1 + x \le e^x$, $x \in \mathbb{R}$, we get, for $|t| \le T_n$,

$$W_2\mathbb{I}_\star \le (1 + \frac{p^*q^*}{25}\rho_n)^{|\mathcal{J}_0|} \le \exp\{\frac{p^*q^*|\mathcal{J}_0|}{25}\rho_n\} = \exp\{\frac{q^*m_n}{25}\rho_n\}. \tag{2.19}$$

Furthermore, by Hoeffding's (1963) Theorem 4, $\mathbf{E}W_1 \le \bigl(\mathbf{E}u(A_1)\bigr)^{|\mathcal{J}_0|}$. We have $\mathbf{E}u(A_1) \le 1 - 2p^*q^*\rho_n$, by the inequality $1 - \mathbf{E}\cos(tg_1(A_1)+s) \ge \rho_n$ from Lemma 2.1 in Bloznelis and Götze (1997). Therefore,

$$\mathbf{E}W_1 \le (1 - 2p^*q^*\rho_n)^{|\mathcal{J}_0|} \le \exp\{-2p^*q^*|\mathcal{J}_0|\rho_n\} = \exp\{-2q^*m_n\rho_n\}. \tag{2.20}$$

Collecting the bounds (2.19) and (2.20) in (2.18), we get

$$\mathbf{E}W\mathbb{I}_\star \le \exp\{-(1-0.02)q^*m_n\rho_n\} \le \exp\{-0.5m_n\rho_n\}, \tag{2.21}$$

since $q^* = 1 - p^* > 7/8$. It follows from (2.21) that $\mathbf{E}f_0\mathbb{I}_\star \le \lambda_*^{-1}\exp\{-0.5m_n\rho_n\}$. Invoking the bound $\lambda_*^{-1} \le cm_n^{1/2} = O(\ln^{1/2}\tau_n)$ we obtain $\mathbf{E}f_0\mathbb{I}_\star = O(\tau_n^{-2}\ln^{1/2}\tau_n)$ uniformly in t, $H_n \le |t| \le T_n$, and this implies the second part of (2.16).

Bound for $\mathbb{J}_4$. We shall show that

$$\mathbb{J}_4 \le c_1\beta_{4,n}^*\tau_n^{-2} + o(\tau_n^{-1}). \tag{2.22}$$

This bound in combination with (2.5) and (1.7) implies the second part of (2.15).

In the first step we represent $\hat{F}_n^*(t)$ in Erdős–Rényi form. Let $\mathbb{A} = (A_1,\dots,A_n)$ be a random permutation of (ordered) set $\mathcal{X}_n$. Let $\nu = (\nu_1,\dots,\nu_n)$ be i.i.d. Bernoulli random variables independent of $\mathbb{A}$ and having probabilities $\mathbf{P}\{\nu_1 = 1\} = p_n$ and $\mathbf{P}\{\nu_1 = 0\} = q_n$.

Since the distribution of U_n^* coincides with the conditional distribution of

$$\sum_{1\le i\le n} g_1(A_i)(\nu_i - p_n) + \sum_{1\le i<j\le n} g_2(A_i,A_j)(\nu_i - p_n)(\nu_j - p_n)$$

given the event $\mathbb{B} = \{\sum_{1\le i\le n}(\nu_i - p_n) = 0\}$, we have

$$\hat F_n^*(t) = \lambda^{-1}\int_{|s|\le \pi\tau_n} \mathbf{E}\exp\{iT + iQ\}ds, \quad T = \sum_{1\le i\le n} T_i, \quad Q = \sum_{1\le i<j\le n} Q_{i,j},$$
$$T_i = (tg_1(A_i) + s/\tau_n)(\nu_i - p_n), \qquad Q_{i,j} = tg_2(A_i, A_j)(\nu_i - p_n)(\nu_j - p_n),$$

where $\lambda = 2\pi\tau_n \mathbf{P}\{\mathbb{B}\}$ satisfies $\lambda^{-1} \le (2\pi)^{1/2}$, see Höglund (1978). Therefore, (2.22) is a consequence of the following bounds

$$\int_{|t|\le H_n} \big|\lambda^{-1}\int_{|s|\le\pi\tau_n}(h_1 + h_2)ds - \hat G_n^*(t)\big| dt/|t| \le c_1\,\frac{\beta_{4,n}^*}{\tau_n^2} + c_1\,\frac{\gamma_{0,n}}{\tau_n^2},$$
$$\int_{|t|\le H_n}\int_{|s|\le\pi\tau_n} |\mathbf{E}\exp\{iT + iQ\} - h_1 - h_2|ds\,dt/|t| = o(\tau_n^{-1}),$$
$$h_1 = \mathbf{E}\exp\{iT\}, \qquad h_2 = i^3\binom{n}{2}\mathbf{E}\exp\{i(T_3 + \cdots + T_n)\}V, \qquad V = Q_{1,2}T_1T_2.$$

The first bound is given in Bloznelis and Götze (1997), see formula (4.2). The second bound is proved in Section 3 below.

3. Expansions

Given a measurable function $f : \mathbb{R}^2 \to \mathbb{C}$ and a measurable set $\mathcal{Z} \subset \mathbb{R}^2$ write

$$J_{\mathcal{Z}}(f) = \int_{\mathcal{Z}} |t|^{-1}|f(t)|ds\,dt.$$

The expression $\exp\{itx\}$ will be abbreviated by $\mathrm{e}\{x\}$. Below we drop the subscript n and write τ, p, q, etc. instead of τ_n, p_n, q_n, etc.

In this section we shall prove the bound

$$J_{\mathcal{Z}}(f_1) = o(\tau^{-1}), \qquad f_1 = \mathbf{E}\,\mathrm{e}\{T + Q\} - h_1 - h_2, \tag{3.1}$$

where $\mathcal{Z} = \{(s,t) \in \mathbb{R}^2 : |s| \le \pi\tau,\ |t| \le H_n\}$.

Before proving (3.1) we introduce some more notation. Given a sum $S = s_1 + \cdots + s_k$ we denote $S^{(i)} = S - s_i$ and $S^{(i,j)} = S - s_i - s_j$, for $1 \le i < j \le k$. Furthermore, we write $S_{\mathcal{B}} = \sum_{j\in\mathcal{B}} s_j$, for $\mathcal{B} \subset \Omega_k$. By $\mathbf{E}_{(\mathcal{B})}$, where $\mathcal{B} \subset \Omega$, we denote the conditional expectation given all the random variables, but $\nu_j, j \in \mathcal{B}$. Write $u = u(s,t) = s^2 + t^2$. In the proof we often use the following bounds which are consequences of (1.6), (1.7), (2.2) and Lemma 2.2.

$$\mathbf{E}g_2^2(A_i, A_j) = O(\tau^{-6}), \quad \mathbf{E}T_i^2 \le c\,\frac{u}{n}, \quad \mathbf{E}T_i^4 \le c\,\frac{u^2}{n\tau}\,\frac{\beta_{4,n}^*}{\tau} = \frac{u^2}{n}\,O(\tau^{-1}). \tag{3.2}$$

For $\mathcal{V} = \{(s,t) \in \mathcal{Z} : s^2 + t^2 \le \ln \tau\}$ and $\mathcal{W} = \mathcal{Z} \setminus \mathcal{V}$, we have

$$J_{\mathcal{Z}}(f_1) \le J_{\mathcal{V}}(f_1) + J_{\mathcal{W}}(h_2) + J_{\mathcal{W}}(f_2), \qquad f_2 = \mathbf{E}\,\mathrm{e}\{T+Q\} - h_1.$$

Therefore, (3.1) is a consequence of the bounds

$$J_{\mathcal{V}}(f_1) = o(\tau^{-1}), \qquad J_{\mathcal{W}}(h_2) = o(\tau^{-1}), \qquad J_{\mathcal{W}}(f_2) = o(\tau^{-1}).$$

The bound for $J_{\mathcal{V}}(f_1)$. Write $f_3 = \mathbf{E}\,\mathrm{e}\{T\} iQ$. Expanding in powers of iQ we get

$$\mathbf{E}\,\mathrm{e}\{T+Q\} = h_1 + f_3 + R_1, \qquad |R_1| \le \mathbf{E}Q^2, \tag{3.3}$$

By symmetry $f_3 = \binom{n}{2} f_4$, where $f_4 = \mathbf{E}\,\mathrm{e}\{T\} iQ_{1,2}$. Expanding $\mathrm{e}\{T\}$ in powers of iT_1 and iT_2 and using the fact that $\mathbf{E}_{(\{j\})} Q_{1,2} = 0$, for $j = 1,2$, we obtain

$$f_4 = i^3 \mathbf{E}\,\mathrm{e}\{T^{(1,2)}\} V + R_2, \qquad |R_2| \le \mathbf{E}|Q_{1,2}|(T_1^2|T_2| + T_2^2|T_1|), \tag{3.4}$$

where $V = Q_{1,2} T_1 T_2$. It follows from (3.3) and (3.4) that $|f_1| \le |R_1| + n^2 |R_2|$. We complete the proof of $J_{\mathcal{V}}(f_1) = o(\tau^{-1})$ by showing that

$$\begin{aligned}
|R_1| &\le \mathbf{E}Q^2 = \binom{n}{2} p^2 q^2 t^2 \mathbf{E} g_2^2(A_1, A_2) = t^2\, O(\tau^{-2}),\\
|R_2|^2 &\le 4(\mathbf{E}|Q_{1,2} T_1| T_2^2)^2 \le 4\mathbf{E}Q_{1,2}^2 \mathbf{E} T_1^2 T_2^4\\
&\le 4\mathbf{E}Q_{1,2}^2 \mathbf{E}T_1^2 \,\frac{n}{n-1}\, \mathbf{E}T_2^4 = n^{-4} t^2 u^3 O(\tau^{-3})
\end{aligned}$$

by Cauchy–Schwartz, (2.1) and (3.2).

The bounds for $J_W(f_2)$ and $\mathcal{J}_{\mathcal{W}}(h_2)$. Given a measurable function $f : \mathbb{R}^2 \to \mathbb{C}$ write $f \sim 0$, if $J_{\mathcal{W}}(f) = o(\tau)$. Similarly, write $f \sim g$, if $f - g \sim 0$.

Given n let $m = m(s,t)$ denote the smallest integer greater than $c_2 n u^{-1} \ln u$, where $c_2 > 0$ is a sufficiently large absolute constant. It is easy to show that for large τ, we have $c_2 \le m \le n/2$, for every $(s,t) \in \mathcal{W}$. We shall assume that n (and therefore, $\tau = \tau_n$) is large enough so that these inequalities hold. Given (s,t) introduce the random variables

$$\xi = \sum_{1 \le i \le m} \xi_i, \qquad \xi_i = t(\nu_i - p)\zeta(A_i), \qquad \zeta(a) = \sum_{j=m+1}^{n} g_2(a, A_j)(\nu_j - p).$$

For $\mathcal{B} \subset \Omega$ and $\mathcal{D} \subset \Omega_m$ denote

$$\tilde{Y}_{\mathcal{B}} = |\mathbf{E}_{(\mathcal{B})}\,\mathrm{e}\{T_{\mathcal{B}}\}|, \qquad Z_{\mathcal{D}} = |\mathbf{E}_{(\mathcal{D})}\,\mathrm{e}\{T_{\mathcal{D}} + \xi_{\mathcal{D}}\}|, \qquad \mathbb{I}_{\mathcal{D}} = \mathbb{I}\{\varkappa_{\mathcal{D}} > c_3\}, \tag{3.5}$$

where $\varkappa_{\mathcal{D}} = \tau^2|\mathcal{D}|^{-1}\sum_{i\in\mathcal{D}}\zeta^2(A_i)$. Proceeding as in proof of Lemma 5.4 in Bloznelis and Götze (1997), see also formulas (4.9–10) ibidem (here we specify constants c_2 and c_3), for every $\mathcal{B}\subset\Omega$, we construct a random variable $Y_{\mathcal{B}}$ depending only on A_i, $i\in\mathcal{B}$, such that, for sufficiently large n, almost surely

$$(3.6)\quad Z_{\mathcal{D}}\le \mathbb{I}_{\mathcal{D}}+Y_{\mathcal{D}}\qquad \tilde{Y}_{\mathcal{B}}\le Y_{\mathcal{B}},\qquad \mathbf{E}(Y_{\mathcal{B}}^2\,|A_i,\,A_j)\le c\,u^{-20},\quad \forall\, i,\,j\in\Omega\setminus\mathcal{B},$$

for every $\mathcal{D}\subset\Omega_m$ and every $\mathcal{B}\subset\Omega$ such that $|\mathcal{B}|\ge m/5$.

Let us prove $J_{\mathcal{W}}(f_2)=o(\tau^{-1})$. Split $T=T_{\mathcal{A}}+T_{\mathcal{B}}$ and $Q=Q_{\mathcal{A}}+Q_{\mathcal{B}}+\xi$, where

$$Q_{\mathcal{A}}=\sum_{1\le i<j\le m}Q_{i,j},\qquad Q_{\mathcal{B}}=\sum_{m<i<j\le n}Q_{i,j},\qquad \mathcal{A}=\Omega_m,\quad \mathcal{B}=\Omega\setminus\mathcal{A},$$

and write $T+Q=W+T_{\mathcal{A}}+\xi+Q_{\mathcal{A}}$, where $W=T_{\mathcal{B}}+Q_{\mathcal{B}}$. In order to prove $f_2\sim 0$ we shall show

$$\begin{aligned}&\mathbf{E}\,\mathrm{e}\{T+Q\}\sim f_5+f_6,\qquad f_6\sim 0,\qquad f_5\sim f_7,\qquad f_7\sim h_1,\\ &f_5=\mathbf{E}v\,\mathrm{e}\{\xi\},\qquad f_6=\mathbf{E}v\,\mathrm{e}\{\xi\}iQ_{\mathcal{A}},\qquad f_7=\mathbf{E}v,\qquad v=\mathrm{e}\{W+T_{\mathcal{A}}\}.\end{aligned}$$

Proof of $\mathbf{E}\,\mathrm{e}\{T+Q\}\sim f_5+f_6$. Expanding $\mathrm{e}\{T+Q\}=v\,\mathrm{e}\{\xi+Q_{\mathcal{A}}\}$ in powers of $iQ_{\mathcal{A}}$, we get $\mathbf{E}\,\mathrm{e}\{T+Q\}=f_5+f_6+R$, where,

$$|R|\le \mathbf{E}Q_{\mathcal{A}}^2=\binom{m}{2}\mathbf{E}Q_{1,2}^2=m^2n^{-2}t^2O(\tau^{-2})=t^2u^{-2}\ln^2 u\,O(\tau^{-2})\sim 0,$$

by symmetry, (3.2) and the choice of m.

Proof of $f_6\sim 0$. Split $\mathcal{A}=\mathcal{A}_1\cup\dots\cup\mathcal{A}_4$, where $\mathcal{A}_i$, $1\le i\le 4$, are disjoint and such that $i\in\mathcal{A}_i$ and $|\mathcal{A}_i|\approx m/4$. Split $Q_{\mathcal{A}}=\sum_{1\le i\le j\le 4}Q_{\mathcal{A}_i,\mathcal{A}_j}$, where $Q_{\mathcal{A}_i,\mathcal{A}_j}$ denotes the sum of $Q_{k,l}$ such that $k\in\mathcal{A}_i$, $l\in\mathcal{A}_j$ and $k<l$. It suffices to prove $f_{i,j}\sim 0$, for every $f_{i,j}=\mathbf{E}v\,\mathrm{e}\{\xi\}Q_{\mathcal{A}_i,\mathcal{A}_j}$. Since the proof for every pair $i\le j$ is almost the same we shall prove only $f_{1,2}\sim 0$. To this aim write $\xi=\xi_*+\xi_{\mathcal{A}_4}$ and $\mathrm{e}\{\xi\}=\mathrm{e}\{\xi_*\}\,\mathrm{e}\{\xi_{\mathcal{A}_4}\}$, where $\xi_*=\xi_{\mathcal{A}_1}+\xi_{\mathcal{A}_2}+\xi_{\mathcal{A}_3}$. Expanding $\mathrm{e}\{\xi_{\mathcal{A}_4}\}=1+\xi_{\mathcal{A}_4}r$, where r is a bounded function of $\xi_{\mathcal{A}_4}$, we obtain

$$f_{1,2}=f'_{1,2}+R\qquad f'_{1,2}=\mathbf{E}v\,\mathrm{e}\{\xi_*\}Q_{\mathcal{A}_1,\mathcal{A}_2},\quad R=\mathbf{E}v\,\mathrm{e}\{\xi_*\}Q_{\mathcal{A}_1,\mathcal{A}_2}\xi_{\mathcal{A}_4}r.$$

Note that $|R|\le \mathbf{E}Z_{\mathcal{A}_3}|Q_{\mathcal{A}_1,\mathcal{A}_2}\xi_{\mathcal{A}_4}|$, see (3.5). By (3.6) and Cauchy–Schwartz,

$$\begin{aligned}|R|&\le \mathbf{E}\mathbb{I}_{\mathcal{A}_3}|Q_{\mathcal{A}_1,\mathcal{A}_2}\xi_{\mathcal{A}_4}|+\mathbf{E}Y_{\mathcal{A}_3}|Q_{\mathcal{A}_1,\mathcal{A}_2}\xi_{\mathcal{A}_4}|\\ &\le(\mathbf{E}\varkappa_{\mathcal{A}_3}Q^2_{\mathcal{A}_1,\mathcal{A}_2})^{1/2}(\mathbf{E}\xi^2_{\mathcal{A}_4})^{1/2}+(\mathbf{E}Y^2_{\mathcal{A}_3})^{1/2}(\mathbf{E}Q^2_{\mathcal{A}_1,\mathcal{A}_2}\xi^2_{\mathcal{A}_4})^{1/2}.\end{aligned}$$

Now the relation $R \sim 0$ follows from (3.6) and the bounds $\mathbf{E}\xi^2_{\mathcal{A}_4} = t^2u^{-1}O(\tau^{-2})$,

$$\mathbf{E}Q^2_{\mathcal{A}_1,\mathcal{A}_2}\xi^2_{\mathcal{A}_4} = |A_1|\,|A_2|\,|A_4|(n-m)\mathbf{E}Q^2_{1,2}Q^2_{4,n} = u^{-3}\ln^3 u\, t^4 O(\tau^{-4}),$$
$$\mathbf{E}\varkappa_{\mathcal{A}_3}Q^2_{\mathcal{A}_1,\mathcal{A}_2} = |A_1|\,|A_2|\tau^2\mathbf{E}\zeta^2(A_3)Q^2_{1,2} = u^{-2}\ln^2 u\, t^2 O(\tau^{-4}).$$

It remains to show that $f'_{1,2} \sim 0$. By symmetry, $f'_{1,2} = \tilde{m}\mathbf{E}v\,\mathrm{e}\{\xi_*\}Q_{1,2}$, where $\tilde{m}$ denotes the number of summands of the sum $Q_{\mathcal{A}_1,\mathcal{A}_2}$. Clearly, $\tilde{m} \le m^2$. Write $\xi_* = \xi_*^{(1,2)} + \xi_1 + \xi_2$ and expand the exponent in powers of $i(\xi_1+\xi_2)$ to get

$$\mathbf{E}v\,\mathrm{e}\{\xi_*\}Q_{1,2} = \mathbf{E}v\,\mathrm{e}\{\xi_*^{(1,2)}\}Q_{1,2} + R, \qquad |R| \le \mathbf{E}Y_{\mathcal{A}_4}|Q_{1,2}(\xi_1+\xi_2)|. \tag{3.7}$$

Furthermore, write $T = T^{(1,2)} + T_1 + T_2$. Expanding the exponent in v in powers of iT_1 and iT_2 and using the identities $\mathbf{E}_{(\{j\})}Q_{1,2} = 0$, for $j = 1, 2$, we get

$$\mathbf{E}v\,\mathrm{e}\{\xi_*^{(1,2)}\}Q_{1,2} = \mathbf{E}\,\mathrm{e}\{W + T^{(1,2)} + \xi_*^{(1,2)}\}Vr, \qquad V = Q_{1,2}T_1T_2, \tag{3.8}$$

where r is a bounded function of T_1 and T_2. Note that the absolute value of the expectation in the right is bounded from above by $\mathbf{E}Y_{\mathcal{A}_4}|V|$. It follows from (3.7) and (3.8) that $|f'_{1,2}| \le 2m^2\mathbf{E}Y_{\mathcal{A}_4}|Q_{1,2}\xi_1| + m^2\mathbf{E}Y_{\mathcal{A}_4}|V|$. This inequality in combination with the bounds

$$\mathbf{E}Y_{\mathcal{A}_4}|Q_{1,2}\xi_1| = t^2u^{-10}n^{-2}O(\tau^{-1}), \qquad \mathbf{E}Y_{\mathcal{A}_4}|V| = |t|u^{-9}n^{-2}O(\tau^{-1}), \tag{3.9}$$

yields $f'_{1,2} \sim 0$. Let us prove (3.9). By Cauchy–Schwartz,

$$\begin{aligned}\mathbf{E}Y_{\mathcal{A}_4}|Q_{1,2}\xi_1| &= p^2q^2t^2\mathbf{E}Y_{\mathcal{A}_4}|g_2(A_1,A_2)\zeta(A_1)| \\ &\le p^2q^2t^2\big(\mathbf{E}Y^2_{\mathcal{A}_4}g_2^2(A_1,A_2)\big)^{1/2}(\mathbf{E}\zeta^2(A_1))^{1/2}.\end{aligned}$$

Invoking the bounds $\mathbf{E}Y^2_{\mathcal{A}_4}g_2^2(A_1,A_2) = u^{-20}O(\tau^{-6})$ and $\mathbf{E}\zeta^2(A_1) = O(\tau^{-4})$, see (3.6) and (3.2), we obtain the first part of (3.9). Similarly, by (3.6), Cauchy–Schwartz and (3.2),

$$\mathbf{E}Y_{\mathcal{A}_4}|V| \le (\mathbf{E}Y^2_{\mathcal{A}_4}Q^2_{1,2})^{1/2}(\mathbf{E}T_1^2T_2^2)^{1/2} = |t|u^{-9}n^{-2}O(\tau^{-1}). \tag{3.10}$$

Proof of $f_5 \sim f_7$. Split $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ so that $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$ and, $\forall i$, $|\mathcal{A}_i| \approx m/2$ and $i \in \mathcal{A}_i$. Write $\mathrm{e}\{\xi\} = \mathrm{e}\{\xi_{\mathcal{A}_1} + \xi_{\mathcal{A}_2}\}$. Expanding in powers of $i\xi_{\mathcal{A}_2}$ and $i\xi_{\mathcal{A}_1}$ we obtain

$$f_5 = \mathbf{E}v\,\mathrm{e}\{\xi_{\mathcal{A}_1}\} + if_8 + R_1, \quad f_8 = \mathbf{E}v\,\mathrm{e}\{\xi_{\mathcal{A}_1}\}\xi_{\mathcal{A}_2}, \quad |R_1| \le \mathbf{E}\mathcal{Z}_{\mathcal{A}_1}\xi^2_{\mathcal{A}_2},$$
$$\mathbf{E}v\,\mathrm{e}\{\xi_{\mathcal{A}_1}\} = f_7 + \mathbf{E}vi\xi_{\mathcal{A}_1} + R_2, \qquad |R_2| \le \mathbf{E}Y_{\mathcal{A}_2}\xi^2_{\mathcal{A}_1}.$$

These expansions in combination with (3.11–14) imply $f_5 \sim f_7$. Let us show

$$R_1 = u^{-1} \ln u \, t^2 o(\tau^{-2-\delta}) + t^2 u^{-20} O(\tau^{-2}) \sim 0, \tag{3.11}$$

$$R_2 = t^2 u^{-20} O(\tau^{-2}) \sim 0, \tag{3.12}$$

$$f_8 = u^{-9} t^2 O(\tau^{-1}) \sim 0, \tag{3.13}$$

$$Ev\xi_{\mathcal{A}_1} = u^{-9} t^2 O(\tau^{-1}) \sim 0. \tag{3.14}$$

Let us prove (3.11). By (3.6), $|R_1| \le \mathbf{E}\mathbb{I}_{\mathcal{A}_1}\xi^2_{\mathcal{A}_2} + \mathbf{E}Y_{\mathcal{A}_1}\xi^2_{\mathcal{A}_2}$. Since $Y_{\mathcal{A}_1}$ does not depend on ν we have, by symmetry, (3.6) and (3.2),

$$\mathbf{E}Y_{\mathcal{A}_1}\xi^2_{\mathcal{A}_2} = |\mathcal{A}_2|(n-m)\mathbf{E}Y_{\mathcal{A}_1}Q^2_{2,n} = t^2 u^{-20} O(\tau^{-2}). \tag{3.15}$$

Furthermore, since $\mathbb{I}_{\mathcal{A}_1}$ does not depend on ν_j, $j \in \mathcal{A}_2$, we have by symmetry and (3.17),

$$\mathbf{E}\mathbb{I}_{\mathcal{A}_1}\xi^2_{\mathcal{A}_2} = |\mathcal{A}_2| t^2 pq \mathbf{E}\mathbb{I}_{\mathcal{A}_1}\zeta^2(A_2) = u^{-1} \ln u \, t^2 o(\tau^{-2-\delta}).$$

This bound together with (3.15) yields (3.11).

To prove (3.12) write, $\mathbf{E}Y_{\mathcal{A}_2}\xi^2_{\mathcal{A}_1} = (n-m)|\mathcal{A}_1|\mathbf{E}Y_{\mathcal{A}_2}Q^2_{1,n}$, by symmetry, and apply (3.6) and (3.2).

Let us prove (3.13). By symmetry again, $f_8 = |\mathcal{A}_2|(n-m)\mathbf{E}v\,\mathrm{e}\{\xi_{\mathcal{A}_1}\}Q_{2,n}$. Write $v\,\mathrm{e}\{\xi_{\mathcal{A}_1}\} = \mathrm{e}\{W' + w + T_2\}$, where $w = \eta + T_n$,

$$\eta = \sum_{i \in \Omega \setminus \mathcal{A}_2} Q_{i,n}, \qquad W' = \sum_{m<i<j<n} Q_{i,j} + \sum_{i \in \mathcal{A}_1} \sum_{m<j<n} Q_{i,j} + T^{(2,n)}.$$

Expanding the exponent in powers of iw and iT_2 and using the fact that W' does not depend on ν_2 and ν_n and $\mathbf{E}_{(\{j\})}Q_{2,n} = 0$, for $i = 2, n$, we obtain $\mathbf{E}v\,\mathrm{e}\{\xi_{\mathcal{A}_1}\}Q_{2,n} = \mathbf{E}\,\mathrm{e}\{W'\}Q_{2,n}T_2wr =: R_3$, where r, with $|r| \le 1$, is a function of T_2 and w. Note that $|R_3| \le \mathbf{E}Y_{\mathcal{A}_3}|Q_{2,n}T_2w|$, where $\mathcal{A}_3 = \mathcal{A}_2 \setminus \{2\}$. Therefore,

$$|f_8| \le n^2\mathbf{E}Y_{\mathcal{A}_3}|Q_{2,n}T_2w| = n^2p^2q^2t^2\mathbf{E}Y_{\mathcal{A}_3}|g_2(A_2, A_n)z_1(z + z_n)|, \tag{3.16}$$

$$z = \sum_{j \in \Omega \setminus \mathcal{A}_2} g_2(A_j, A_n)(\nu_j - p), \qquad z_k = tg_1(A_k) + s/\tau, \quad k \in \Omega.$$

Now (3.13) follows from (3.16) and the bound

$$\mathbf{E}Y_{\mathcal{A}_3}|g_2(A_2, A_n)z_1(z + z_n)| = u^{-9}O(\tau^{-5}).$$

To prove this bound we combine Cauchy–Schwartz and the following bounds, which follow from (3.6), (3.2) and (2.1), $\mathbf{E}z_2^2z_n^2 = u^2O(\tau^{-4})$,

$$\mathbf{E}Y^2_{\mathcal{A}_3}z_1^2 = u^{-20}O(\tau^{-6}), \qquad \mathbf{E}z_2^2z_3^2 \le (n - |\mathcal{A}_2|)pq\mathbf{E}z_2^2g_2^2(A_1, A_n) = uO(\tau^{-6}).$$

The proof of (3.14) is similar to that of (3.13), but simpler.

Proof of $f_7 \sim h_1$. Write $f_9 = \mathbf{E}\,\mathrm{e}\{T\}iQ_{\mathcal{B}}$. Expanding in powers of $iQ_{\mathcal{B}}$, we get $f_7 = h_1 + f_9 + R$, with $|R| \le \mathbf{E}Y_{\mathcal{A}}Q^2_{\mathcal{B}}$. As in (3.12) above, $|R| = t^2u^{-20}O(\tau^{-2}) \sim 0$. It remains to show $f_9 \sim 0$. By symmetry, $f_9 = \binom{n-m}{2}\mathbf{E}\,\mathrm{e}\{T\}Q_{n-1,n}$ Expanding the exponent in powers of iT_{n-1} and iT_n we get $\mathbf{E}\,\mathrm{e}\{T\}Q_{n-1,n} = R_1$, where

$$|R_1| \le \mathbf{E}Y_{\Omega_{n-2}}|Q_{n-1,n}T_{n-1}T_n| = u^{-19}|t|p^2q^2O(\tau^{-5}),$$

see (3.8) and (3.10). We obtain $|f_9| \le n^2|R_1| = u^{-19}t^2O(\tau^{-1}) \sim 0$.

The proof of $\mathcal{J}_{\mathcal{W}}(h_2) = o(\tau^{-1})$ is almost the same as that of $f_9 \sim 0$ in the step $f_7 \sim h_1$ above.

Lemma 3.1. *Assume that (1.5)–(1.7) hold. We have, for all $k \in \Omega_m$,*

$$\forall \mathcal{A}_1 \subset \Omega_m \setminus \{k\}, \qquad \mathbf{E}\mathbb{I}_{\mathcal{A}_1}\zeta^2(A_k) = o(\tau^{-4-\delta}) \qquad \textit{as} \quad \tau \to \infty. \tag{3.17}$$

Proof of Lemma 3.1. Write $u = 2 + \delta$ and $M = \mathbf{E}|g_2(A_1, A_k)|^u$. Observe, that $M = O(\tau^{-3u})$, by (1.5) and (2.3). The bound (3.17) follows from

$$\mathbf{E}|\zeta(A_i)|^u \le c\tau^u M, \qquad \forall i \in \Omega_m, \tag{3.18}$$

$$\mathbf{E}\varkappa^{u/2}_{\mathcal{A}_1} \le c\tau^{2u}M, \tag{3.19}$$

via Hölder's inequality $\mathbf{E}\mathbb{I}_{\mathcal{A}_1}\zeta^2(A_2) \le (\mathbf{E}\varkappa^{u/2}_{\mathcal{A}_1})^{\delta/u}(\mathbf{E}|\zeta(A_2)|^u)^{2/u}$.

Let us prove (3.18). By Rozenthal's inequality, for $\mathcal{B} = \Omega \setminus \Omega_m$,

$$\mathbf{E}_{(\mathcal{B})}|\zeta(A_i)|^u \le c\sum_{j\in\mathcal{B}}|g_2(A_i, A_j)|^u\mathbf{E}_{(\mathcal{B})}|\nu_j - p|^u + c\Big(\sum_{j\in\mathcal{B}}g_2^2(A_i, A_j)\mathbf{E}_{(\mathcal{B})}(\nu_j - p)^2\Big)^{u/2}.$$

Taking the expectation and invoking the inequalities

$$\mathbf{E}_{(\mathcal{B})}|\nu_j - p|^l \le pq, \qquad \sum_{j\in\mathcal{B}}|g_2(A_i, A_j)|^l \le n\mathbf{E}(|g_2(A_i, A_j)|^l\big|A_i), \qquad l \ge 2,$$

we obtain

$$\mathbf{E}|\zeta(A_i)|^u \le c\tau^2M + c\tau^u\mathbf{E}(\mathbf{E}(g_2^2(A_i, A_n)|A_i))^{u/2}$$

Finally, an application of Lyapunov's inequality $(\mathbf{E}|X|)^\eta \le \mathbf{E}|X|^\eta$, $\eta \ge 1$, to the conditional expectation completes the proof of (3.18).

To prove (3.19) we apply the inequality $(k^{-1}\sum_{i=1}^k a_i^2)^{u/2} \le k^{-1}\sum_{i=1}^k |a_i|^u$ and use symmetry to get $\mathbf{E}(|\mathcal{A}_1|^{-1}\sum_{i\in\mathcal{A}_1}\zeta^2(A_i))^{u/2} \le \mathbf{E}|\zeta(A_1)|^u$. Now (3.19) follows from (3.18).

References

Bickel, P.J., Götze, F. and van Zwet, W.R., *The Edgeworth expansion for U-statistics of degree two.*, Ann. Statist. **14** (1986), 1463–1484.

Bloznelis, M. and Götze, F., *An Edgeworth expansion for finite population U-statistics* (1997), Submited to *Bernoulli.*

Bolthausen, E. and Götze, F., *The rate of convergence for multivariate sampling statistics*, Ann. Statist. **21** (1993), 1692–1710.

Erdös, P. and Rényi, A., *On the central limit theorem for samples from a finite population*, Publ. Math. Inst. Hungar. Acad. Sci. **4** (1959), 49–61.

Feller, W., *An Introduction to Probability Theory and Its Applications. Vol II.* (1971), John Wiley & Sons, New York.

Hoeffding, W., *Probability inequalities for sums of bounded random variables*, J. Amer. Stat. Assoc. **58** (1963), 13–30.

Höglund, T., *Sampling from a finite population. A remainder term estimate*, Scand. J. Statist. **5** (1978), 69–71.

Kokic P.N. and Weber N.C., *An Edgeworth expansion for U-statistics based on samples from finite populations*, Ann. Probab. **18** (1990), 390–404.

Nandi, H. K. and Sen, P.K., *Unbiased estimation of the parameters of a finite population.*, Calcutta Statist. Assoc. Bull. **12** (1963), 124–148.

Petrov, V.V., *Limit Theorems in Probability Theory. Sequences of Independent Random Variables* (1995), Oxford University Press, New York.

Zhao Lin Cheng and Chen Xi Ru, *Normal approximation for finite-population U-statistics*, Acta mathematicae applicatae Sinica **6** (1990), 263–272.