Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 275–286

U-Statistics CLT Using Cumulants and a Free Version

Arup Bose[1] and Apratim Dey[2]
[1]*Stat-Math Unit*
Indian Statistical Institute, Kolkata, India
[2]*Department of Statistics*
Stanford University, California

Received: 25 April 2020; Revised: 18 June 2020; Accepted: 11 August 2020

Abstract

It is well-known that a standardised U-statistic based on i.i.d. observations is asymptotically normal. We first give a proof of this using cumulants. Then we consider U-statistics which are based on non-commutative variables. We show that a standardised U-statistics of freely independent identically distributed non-commutative random variables converges to a semi-circle variable. The proof is based on free cumulants. We also discuss briefly the degenerate case.

Key words: U-statistics; Central limit theorem; Degenerate U-statistics; Law of large numbers; Non-commutative probability; Freely independent random variables; Cumulants and free cumulants.

AMS Subject Classifications: 60F05, 46L53

1. Introduction

Let $X_1, X_2, ..., X_n$ be independent and identically distributed (i.i.d.) random variables, defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and with a common distribution F. Suppose $h : \mathbb{R}^m \to \mathbb{R}$ is a Borel measurable function which is symmetric in its arguments. Let

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < ... < i_m \leq n} h(X_{i_1}, ..., X_{i_m}). \tag{1}$$

Then U_n is the well-known U*-statistic* with kernel h and has found extensive uses in statistics. Bose and Chatterjee (2018) contains a wealth of material on the properties of U-statistics. A fundamental distributional limit result for U-statistic is the U-statistics Central Limit Theorem (UCLT).

Theorem 1: (UCLT) Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with a common distribution F and

$$\int_{\mathbb{R}^m} |h(x_1, ..., x_m)|^2 dF(x_1)...dF(x_m) < \infty.$$

Correponding Author: Arup Bose
Email: bosearu@gmail.com

Let $\sigma_1^2 := Var(\mathrm{E}(h(X_1, X_2, ..., X_m)|X_1)) > 0$. Then

$$\sqrt{n}(U_n - \theta) \stackrel{w}{\to} N(0, m^2\sigma_1^2)$$

where $\stackrel{w}{\to}$ denotes weak convergence, $N(0, m^2\sigma_1^2)$ denotes the normal distribution with mean 0 and variance $m^2\sigma_1^2$ and $\theta = \int_{\mathbb{R}^m} h(x_1, ..., x_m)dF(x_1)...dF(x_m)$.

The standard proof proceeds by considering the sum of projections of U_n on the subspace $\mathcal{L}_n := \{\psi_1(X_1) + \cdots + \psi_n(X_n) : \psi_1, \ldots, \psi_n \in L^2(F)\}$ and showing that in the limiting case, as $n \to \infty$, the contribution of this sum is most important. The remaining terms are negligible. Then the classical CLT for the sample mean is applied to complete the proof.

With Theorem 1 as the backdrop, our goal in this article is three-fold.

(i) Show how Theorem 1 can be derived using cumulants and thereby avoid projections.

(ii) Establish the CLT for U-statistics in the *non-commutative* set up, by using free cumulants.

(iii) Establish a limit theorem for degenerate U-statistics in the non-commutative set up.

We address these three goals in the next three sections.

2. Proof of UCLT Based on Cumulants

Suppose $Y_1, \ldots Y_n$ are random variables with joint moment generating function

$$M_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \mathrm{E}\left[\exp\left\{\sum_{j=1}^{n} t_j Y_j\right\}\right],\ t_1, \ldots, t_n \in \mathbb{R}.$$

which is finite in a neighbourhood N of 0 in $\mathbb{R}^n$. In that case, the joint cumulant generating function is defined as

$$C_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \log M_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n),\ (t_1, \ldots, t_n) \in N$$

which also has a power series expansion of the form

$$C_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \sum_{k_1,\ldots,k_n=0}^{\infty} \frac{t_1^{k_1} \ldots t_n^{k_n}}{k_1! \ldots k_n!} c_{k_1,\ldots,k_n}(Y_1, \ldots, Y_n), (t_1, \ldots, t_n) \in N.$$

The real numbers $c_{k_1,\ldots,k_n}(Y_1, \ldots, Y_n)$ are called the cumulants of $\{Y_i : 1 \leq i \leq n\}$. If $k_j \neq 0$ for at least two indices j, then $c_{k_1,\ldots,k_n}(Y_1, \ldots, Y_n)$ is called a mixed cumulant of $\{Y_i : 1 \leq i \leq n\}$. We shall use the notation

$$c_j(Y_1, \ldots Y_j) \ \text{ for } \ c_{1,1,\ldots 1}(Y_1, \ldots Y_j).$$

It is easily seen that the if Y, Y_1, Y_2 are random variables, then

$$c_1(Y) = \mathrm{E}(Y),\ c_2(Y) = \mathrm{Var}(Y) \ \text{ and } \ c_{1,1}(Y_1, Y_2) = c_{1,1}(Y_2, Y_1) = \mathrm{Cov}(Y_1, Y_2). \tag{2}$$

In general moments and cumulants are in $1-1$ correspondence and this can be expressed *via* the well-known Möbius function on the set of all partitions of $\{1, \ldots, n\}$, $n \geq 1$. For details of this and other facts, see Nica and Speicher (2006). See also Brillinger (2001) where cumulants have been used extensively to prove limit theorems in time series. We shall need the following facts about cumulants.

Fact 1. Suppose Y_i are independent bounded random variables. Then all their mixed cumulants vanish. This follows easily since the moment generating function factorises.

Fact 2. Cumulants are multi-linear functions of the random variables. This follows from definition.

Fact 3. Y is normally distributed with mean μ and variance σ^2 if and only if its first two cumulants are $c_1(Y) = \mu$ and $c_2(Y) = \sigma^2$ and all other cumulants of Y vanish. This follows immediately from the moment generating function of Y.

Fact 4. Suppose $\{Y_n\}$ is a sequence of random variables such that $c_1(Y_n) \to \mu$, $c_2(Y_n) \to \sigma^2$ and $c_k(Y_n) \to 0$ for all $k \geq 3$. Then $Y_n \xrightarrow{w} N(0,\ \sigma^2)$ where $\xrightarrow{w}$ denotes weak convergence. This is obtained as follows: from the specific nature of the $1-1$ correspondence, it follows that all moments of Y_n converge, and converge to the normal moments. Since the normal distribution is the unique distribution with the normal moments, weak convergence follows.

Proof: [Proof of Theorem 1] We first prove Theorem 1 under the additional assumption that

$$h \text{ is bounded.} \tag{3}$$

By linearity of cumulants, $c_1(U_n) = \theta$ and therefore $c_1(\sqrt{n}(U_n - \theta)) = 0$.

Define

$$\zeta_k(h) = \mathrm{Cov}(h(X_1, ..., X_k, X_{k+1}, ..., X_m), h(X_1, ..., X_k, X_{m+1}, ..., X_{2m-k})).$$

Note that $\zeta_1(h) = \sigma_1^2$. The following formula is standard. It can also be proved easily by using the symmetry of h, equation (2) and linearity of cumulants.

$$c_2(\sqrt{n}(U_n - \theta)) = n \binom{n}{m}^{-2} \sum_{k=1}^{m} \binom{n}{k} \binom{n-k}{m-k} \binom{n-m}{m-k} \zeta_k(h).$$

For any fixed natural number a, we have $\binom{n}{a} \sim \dfrac{n^a}{a!}$ as $n \to \infty$. Therefore, for $1 \leq k \leq m$, we have

$$\binom{n}{k} \binom{n-k}{m-k} \binom{n-m}{m-k} \sim \frac{n^k (n-k)^{m-k} (n-m)^{m-k}}{k!(m-k)!(m-k)!} \sim \frac{n^{2m-k}}{k!(m-k)!(m-k)!}.$$

Thus

$$\lim_{n\to\infty} c_2(\sqrt{n}(U_n - \theta)) = \lim_{n\to\infty} \frac{m!m!}{n^{2m-1}} \sum_{k=1}^{m} \frac{n^{2m-k}}{k!(m-k)!(m-k)!} \zeta_k(h).$$

Note that if $1 < k \leq m$ then $\dfrac{n^{2m-k}}{n^{2m-1}} \to 0$. Therefore, only the $k = 1$ term will survive and we get

$$\lim_{n\to\infty} c_2(\sqrt{n}(U_n - \theta)) = m^2\zeta_1(h) > .0$$

Now we will show that $c_j(\sqrt{n}(U_n - \theta)) \to 0$ for all $j \geq 3$. We observe that, by multi-linearity of cumulants,

$$c_j(\sqrt{n}(U_n - \theta)) = \frac{n^{j/2}}{\binom{n}{m}^j} \sum_{I_1, I_2, \ldots, I_j} c_j(h(X_{I_1}), ..., h(X_{I_j})) \tag{4}$$

where $I_1, ..., I_j$ are ordered $m-$tuples $(i_1 < i_2 < \cdots < i_m)$ with each $1 \leq i_r \leq n$ for $1 \leq r \leq m$, and $h(X_I) := h(X_{i_1}, ..., X_{i_m})$ if $I = (i_1, ..., i_m)$.

Now we make the following observations. Fix I_1. Suppose at least one of $I_2, ..., I_j$ does not have any index common with I_1. Then by independence, the corresponding $c_j = 0$.

Let us count the remaining cases. If we fix I_1, then are $O_m(n^{m-1})$ such choices for each of $I_2, ..., I_j$, giving a total of $O_{m,j}(n^{(m-1)(j-1)})$ choices. Finally, I_1 can be chosen in $\binom{n}{m} = O_m(n^m)$ ways. Therefore, the total count of the remaining cases is $O_{m,j}(n^{m+(m-1)(j-1)})$. Note that the we have a common uppper bound for all the cumulants c_j that correspond to these cases.

Hence if $j \geq 3$,

$$c_j(\sqrt{n}(U_n - \theta)) = O_{m,j}(\frac{n^{j/2}}{n^{mj}} \times n^{m+mj-j-m+1}) = O_{m,j}(n^{1-(j/2)}) \to 0 \text{ as } n \to \infty.$$

Hence the proof is complete by an application of Fact 4, under the extra condition (3).

To relax this assumption, we use a standard truncation argument. Define

$$\tilde{h}(x_1, \ldots, x_m) = h(x_1, \ldots, x_m)I(|h(x_1, \ldots, x_m)| \leq B).$$

Let $\tilde{U}_n^{(B)}$ be the corresponding U-statistic. Since $\tilde{h}$ satisfies (3),

$$\sqrt{n}(U_n^{(B)} - \theta_n^{(B)}) \xrightarrow{w} N(0,\ m^2\tilde{\zeta}_1(\tilde{h})) \text{ as } n \to \infty.$$

It is not hard to show that (use DCT)

$$\zeta_1(\tilde{h}) \to \zeta_1(h) \text{ as } B \to \infty.$$

Moreover, it is also easy to show, by using the variance formula developed above for any U-statistics, that

$$\lim_{B\to\infty} \lim_{n} V\Big(\sqrt{n}(U_n^{(B)} - \theta_n^{(B)}) - \sqrt{n}(U_n - \theta)\Big) = 0.$$

This completes the proof of Theorem 1. □

Remark 1: (a) The special case of $m = 1$ yields the standard CLT for the mean: if $\{X_i\}$ are i.i.d. with mean 0 and variance 1, then $\sum_{i=1}^{n} X_i/\sqrt{n}$ converges weakly to the standard normal distribution. This cumulant based proof avoids the use of characteristic function.

(b) By extending the above argument, and an appropriate extension of Fact 4 to multivariate normal, it can be shown that if we have several U-statistics then after the needed centering and scaling, they converge jointly to a multivariate normal distribution. We omit the details.

3. UCLT for Free Variables

Free Probability refers to an extension of classical probability to certain non-commutative spaces. One of its central notions is free independence which is a natural notion of independence available in different types of non-commutative probability spaces. Very strong connections between free independence and random matrices were discovered by Voiculescu (see Voiculescu (1991)). A nice combinatorial introduction to free probability is available in Nica and Speicher (2006). Probabilists and statisticians have been increasingly drawn to aspects of free probability, specially in the context of high dimensional random matrices. For some flavour of its application in high dimensional time series, see Bose and Bhattacharjee (2018).

The non-commutative probability space that we shall work with is the $*$-probability space. We shall briefly describe its basic ingredients. For a detailed introduction see Nica and Speicher (2006).

Recall that, an algebra $\mathcal{A}$ over complex numbers is called a $*$-algebra if it contains a unity $1_{\mathcal{A}}$, and is endowed with an antilinear $*$ operation which maps $a \in \mathcal{A}$ to $a^* \in \mathcal{A}$ and which satisfies $(a^*)^* = a$ and $(ab)^* = b^*a^*$ for all $a, b \in \mathcal{A}$.

A *$*$-probability space* is a pair $(\mathcal{A}, \varphi)$ where $\mathcal{A}$ is a $*$-algebra and φ is a linear functional on $\mathcal{A}$ which satisfies $\varphi(1_{\mathcal{A}}) = 1$, $\varphi(a^*) = \overline{\varphi(a)}$, and $\varphi(a^*a) \geq 0$ for all $a \in \mathcal{A}$.

The elements of $\mathcal{A}$ are called *random elements*. An element $a \in \mathcal{A}$ is called *self-adjoint* if $a^* = a$. Recall that the expectation operator is also linear and satisfies $\mathrm{E}(1) = 1$. Thus, it helps to think of φ as an analogue of the expectation operator.

Example 1: Suppose $(\Omega, \mathcal{F}, P)$ is a classical probability space and E is the expectation operator. Let $\mathcal{A}$ be the set of (complex valued) random variables with all moments finite, where random variables that are almost surely equal, are identified as same. Then $(\mathcal{A}, \mathrm{E})$ is trivially a $*$-probability space. In this case elements of $\mathcal{A}$ commute.

Example 2: A typical example of a $*$-probability space is the algebra $\mathcal{A}$ of all $n \times n$ matrices with random variable entries all whose moments are finite, and for any $A \in \mathcal{A}$, $\varphi(A) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(A)$. The unity is the $n \times n$ identity matrix I, for which $\varphi(I) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(I) = 1$. With $*$ denoting the usual matrix adjoint, $\varphi(A^*) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(A^*) = n^{-1}\overline{\mathrm{E}\,\mathrm{Trace}(A)} = \overline{\varphi(A)}$ since the diagonal entries of A^* are complex conjugates of those of A, and $\varphi(A^*A) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(A^*A) \geq 0$ since all diagonal entries of A^*A are non-negative.

Given random elements $\{a_1, \ldots, a_n\}$, its moments are the quantities $\{\varphi(b_1 \cdots b_k), k \geq 1, b_j \in \{a_1, \ldots, a_n\}$ for all $1 \leq j \leq k\}$. Analogous to cumulants of random variables, there is a concept of *free cumulants* of random elements. These are in $1-1$ correspondence with the moments via the Möbius function on non-crossing partitions of $\{1, \ldots, n\}$, $n \geq 1$. We shall avoid a formal definition. The free cumulants will be denoted by a generic κ. We may note here that for any two random elements a and b, $\kappa_1(a) = \varphi(a)$ and $\kappa_2(a, b) = \varphi(ab) - \varphi(a)\varphi(b)$. Note that in general $\kappa_2(a, b) \neq \kappa_2(b, a)$.

A random element s on a $*$-probability space $(\mathcal{A}, \varphi)$ is said to be a *semi-circle variable* with variance σ^2 if it is self-adjoint and

$$\varphi(s^h) = \begin{cases} \sigma^{2n} C_n = \frac{1}{n+1}\binom{2n}{n}\sigma^{2n}, & \text{if } h = 2n \\ 0, & \text{if } h \text{ is odd.} \end{cases} \tag{5}$$

The numbers $\{C_k, k \geq 1\}$ are known as Catalan numbers and $\{\sigma^{2k} C_k, k \geq 1\}$ define a unique probability measure, known as the *semi-circle distribution* with variance σ^2. It is well-known that $\kappa_1(s) = 0$, $\kappa_2(s, s) = 1$ and all higher order free cumulants of s are 0. Thus, this is the analogue of the standard normal variable X which has the same property for its cumulants.

In the classical set-up, bounded random variables are independent if and only if all their mixed cumulants are 0. Analogously, random elements are "freely independent" or simply *free*, if and only if all their mixed free cumulants vanish.

We also need the notion of convergence of random elements: Suppose we have a sequence of $*$-probability spaces $(\mathcal{A}_n, \varphi_n)$. Suppose $a_n \in \mathcal{A}_n$ are self-adjoint. Then $\{a_n\}$ are said to *converge in distribution* if $\lim \varphi_n(a_n^k)$ exists for all integers k. We visualize a limit $*$-probability space $\mathcal{A}$, generated by an indeterminate (self-adjoint) element a and with the state $\varphi(a^k) = \lim \varphi_n(a_n^k)$ for all k and extended linearly to the entire algebra. Note that convergence in distribution is *not* the same as the usual weak convergence. However, if $\{\varphi_n(a_n^k), k \geq 1\}$ and $\{\varphi(a^k), k \geq 1\}$ determine unique probability measures,$\{\mu_n\}$ and μ with these as their moments, then the above convergence in distribution implies μ_n converges to μ weakly. Analogous notions hold for joint convergence of several variables.

We shall need the following facts about free cumulants. For proofs see Nica and Speicher (2006).

Fact (a). Suppose Y_i are freely independent random elements in some $*$-probability space. Then all their mixed free cumulants vanish.

Fact (b). Free cumulants are multi-linear functions.

Fact (c). Suppose s is a semi-circle variable with mean 0 and variance σ^2. Then the first two free cumulants of s are $\kappa_1(Y) = 0$ and $\kappa_2(Y) = \sigma^2$. Further all other free cumulants of s vanish.

Fact (d). Suppose $\{y_n\}$ is a sequence of self-adjoint random elements such the $\kappa_1(y_n) \to 0$, $\kappa_2(y_n) \to \sigma^2$ and $\kappa_k(y_n) \to 0$ for all $k \geq 3$. Then $\{y_n\}$ converges to a semi-circle variable with variance σ^2.

We are now in position to give a free version of Theorem 1. Since we are working with random elements which are elements of an algebra, we are restricted to working with only polynomials in the variables of the algebra. Hence the statement of Theorem 1 needs to be modified slightly now. Variables are said to be identically distributed of they have the same moments.

Theorem 2: Suppose $X_1, ..., X_n$ are freely independent self-adjoint identically distributed random elements on a $*$-probability space $(\mathcal{A}, \varphi)$. Suppose $h(x_1, ..., x_m)$ is a self-adjoint polynomial in the m variables $x_1, ..., x_m$, symmetric in its arguments. Let U_n be the U-statistics with kernel h. Let $\theta = \varphi(h(X_1, ..., X_m))$. For $k = 0, 1, ..., m$ define

$$d_k = \kappa_2(h(X_1, ..., X_k, X_{k+1}, ..., X_m), h(X_1, ..., X_k, X_{m+1}, ..., X_{2m-k})).$$

Suppose $d_1 > 0$. Then $\sqrt{n}(U_n - \theta)$ converges in distribution to a semi-circle variable with variance $m^2 d_1$.

Example 3: Suppose $\{X_i\}$ are free and identically distributed. Consider the sample variance

$$s_n^2 = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \frac{(X_i - X_j)^2}{2}.$$

Then s_n^2 is a U-statistic. Suppose without loss of generality $\varphi(X_i) = 0$. By an application of Theorem 2,

$$n^{1/2}\Big(s_n^2 - \varphi(X_1^2)\Big) \to s$$

where s is a semi-circle variable with variance $\varphi(X_1^4) - [\varphi(X_1^2)]^2$. Note that this could be 0 (for example if X_i are free Bernoulli ± 1 with probability 1/2 each) in which case, s is 0.

Example 4: Suppose $\{X_i\}$ are freely independent identically distributed variables where $2[\varphi(X_1)]^2 = \theta$. Let $h(x, y) = xy + yx$ Then

$$\sqrt{n}\Big(U_n - \theta\Big) \to s$$

where s is a semi-circle variable with mean 0 and variance $16\varphi(X_1)^2\Big[\varphi(X_1^2) - [\varphi(X_1)]^2\Big]$.

Proof: [Proof of Theorem 2] The proof is almost a repetition of the proof of Theorem 1. We sketch it. Trivially, $\kappa_1(U_n) = \theta$ and therefore $\kappa_1(\sqrt{n}(U_n - \theta)) = 0$.

As before (we now use the fact that mixed free cumulants vanish for freely independent variables),

$$\kappa_2(\sqrt{n}(U_n - \theta)) = n\binom{n}{m}^{-2} \sum_{k=1}^{m} \binom{n}{k}\binom{n-k}{m-k}\binom{n-m}{m-k} d_k$$

and after similar steps, we get

$$\lim_{n\to\infty} \kappa_2(\sqrt{n}(U_n - \theta)) = m^2 d_1 > 0.$$

Now we will show that $\kappa_j(\sqrt{n}(U_n - \theta)) \to 0$ for all $j \geq 3$. We observe that

$$\kappa_j(\sqrt{n}(U_n - \theta)) = n^{j/2}\binom{n}{m}^{-j} \sum_{I_1, I_2, ..., I_j} \kappa_j(h(X_{I_1}), ..., h(X_{I_j})) \tag{6}$$

where $I_1, ..., I_j$ are ordered $m-$tuples $(i_1 < i_2 < \cdots < i_m)$ with each $1 \leq i_r \leq n$ for $1 \leq r \leq m$, and $h(X_I) := h(X_{i_1}, ..., X_{i_m})$ if $I = (i_1, ..., i_m)$.

Now we count as before and use vanishing of mixed free cumulants when there are at least two freely independent variables, to obtain, for $j \geq 3$,

$$\kappa_j(\sqrt{n}(U_n - \theta)) = O_{m,j}(\frac{n^{j/2}}{n^{mj}} \times n^{m+mj-j-m+1}) = O_{m,j}(n^{1-(j/2)}) \to 0 \text{ as } n \to \infty.$$

The proof is complete once we use Fact (d). □

Remark 2: By extending the above argument, and an appropriate extension of Fact (d), to a semi-circle family (see next section), it can be shown that if we have several U-statistics of non-commutative variables, then they converge jointly to a semi-circle family. We omit the details.

4. Degenerate Case

An obvious question that arises here is what happens under degeneracy *i.e.* when $c_1 = 0$ or $d_1 = 0$? The following result is well-known in the classical case. See Bose and Chatterjee (2018).

Theorem 3: Let $h : \mathbb{R}^m \to \mathbb{R}$ be a symmetric kernel. Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables, such that $\mathrm{E}(h(x_1, X_2, ..., X_m)) = 0$ but

$$\sigma_2^2 := Var(\mathrm{E}(h(X_1, X_2, X_3, ..., X_m)|X_1, X_2))) > 0.$$

Then

$$nU_n \overset{w}{\to} \binom{m}{2} \sum_{k=1}^{\infty} \lambda_k(V_k - 1)$$

where $V_k \overset{i.i.d.}{\sim} \chi_1^2$ and λ_k are eigenvalues of an appropriate integral operator.

The result is first proved for the case $m = 2$. The classical proof (see Bose and Chatterjee (2018)) crucially uses the Fredholm representation: any symmetric kernel ψ : $\mathbb{R}^2 \to \mathbb{R}$ can be written as

$$\psi(x_1, x_2) = \sum_{k=1}^{\infty} \lambda_k f_k(x_1) f_k(x_2)$$

where λ_k are eigenvalues of an appropriate integral operator. The theorem is easy to prove when there are only finitely many non-zero eigenvalues and the general case is tackled by approximation. Then the cases $m \geq 3$ is proved by projections.

The above proof is not suitable for our purposes since our variables are non-commutative. We present a free version of the theorem for degenerate U-statistics based on random elements but with special type of kernels of order 2. It should be possible to extend this result to higher order kernels but we decided to stick to the simplest case.

Theorem 4: Let $h(x_1, x_2) := \sum_{k=1}^{K} a_k(f_k(x_1)f_k(x_2) + f_k(x_2)f_k(x_1))$ where $f_k(x)$ are self adjoint polynomials in the variable x, and a_k are constants. Let $X_1, ..., X_n$ be freely independent self adjoint identically distributed random variables such that the following are true.

(a) For each $1 \leq k \leq K$, $\varphi(f_k(X_1)) = 0$ and $\varphi(f_k^2(X_1)) = 1$.

(b) For each $1 \leq k \neq l \leq K$, $\varphi(f_k(X_1)f_l(X_1)) = 0$.

Define $U_n = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} h(X_{i_1}, X_{i_2})$. Then,

$$nU_n \to 2\sum_{k=1}^{K} a_k(s_k^2 - 1)$$

where $s_1, ..., s_K$ are freely independent standard semi-circular variables.

Example 5: (Examples 3 and 4 continued) Consider the kernel $h(x, y) = xy + yx$ but we now assume that $\varphi(X_i) = 0$ and $\varphi(X_i^2) = 1$. Then it is easy to see that $K = 1$ and conditions (a) and (b) hold.

$$nU_n = n\binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} (X_{i_1}X_{i_2} + X_{i_2}X_{i_1}) \to 2(s^2 - 1)$$

where s is a semi-circle variable with variance 1.

Recall the sample variance s_n^2. Now suppose that X_i are free and identically distributed as classical Bernoulli ± 1 with equal probability. Then $\varphi(X_i) = 0$ and $\varphi(X_i^4) = [\varphi(X_i^2)]^2$, so that $n^{1/2}(s_n^2 - 1)$ converges to 0. It can be checked that

$$\begin{aligned} n(s_n^2 - 1) &= n\Big[\frac{1}{n-1}[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2] - 1\Big] \\ &= n[\frac{n}{n-1} - 1] - \frac{n}{n-1}\Big(\sqrt{n}\bar{X}\Big)^2 \to -(s^2 - 1) \end{aligned}$$

where s is a semi-circle variable with variance 1.

As preparation for the proof, we need to extend some of the notions introduced earlier. Suppose $(\mathcal{A}_n, \varphi_n)$, $n \geq 1$ is a sequence of $*$-probability spaces. Let $\{a_{i,n}, 1 \leq i \leq k\}$ be random elements from $\mathcal{A}_n$, $n \geq 1$. They are said to converge jointly if $\varphi_n\left(P(a_{i,n}, a_{i,n}^*, 1 \leq i \leq k)\right)$ converges for every $k \geq 1$ and every polynomial P. Then we can define a limit $*$-probability space $(\mathcal{A}, \varphi)$ where $\mathcal{A}$ is the $*$-algebra generated by polynomials in indeterminates $\{a_i, 1 \leq i \leq k\}$ and the state φ is determined by the limit. That is, for all $k \geq 1$ and all polynomials

$$\varphi\left(P(a_i, a_i^*, 1 \leq i \leq k)\right) = \lim \varphi_n\left(P(a_{i,n}, a_{i,n}^*, 1 \leq i \leq k)\right).$$

We write

$$(a_{i,n}, 1 \le i \le k) \to (a_i, 1 \le i \le k).$$

This is equivalent to saying that for all $j \ge 1$ and for all $1 \le i_1, ..., i_j \le k$,

$$\kappa_j(a_{i_1,n}, \cdots, a_{i_j,n}) \to \kappa_j(a_{i_1}, \cdots, a_{i_j}).$$

A collection $(s_1, ..., s_k)$ of random elements from a $*$-probability space $(\mathcal{A}, \varphi)$ is said to be a semi-circular family if these are self-adjoint semi-circle variables and moreover all mixed free cumulants of order greater than 2 are 0. Note that they are then free if all second order mixed free cumulants are also 0.

We shall also need the following free Central Limit Theorem. This can be proved easily in a few lines by using free cumulants—along the lines mentioned in Remark 1—simply use free cumulants instead of usual cumulants. See Nica and Speicher (2006) for a moment based proof.

Theorem 5: Suppose $\{X_{i,j}, 1 \le i \le k\}, j \ge 1$ are self-adjoint variables which are identically distributed as well as free across $j \ge 1$ in some $*$-algebra such that for all i, j, $\varphi(X_{1,j}) = 0$. Then

$$\frac{1}{\sqrt{n}}(X_{i,1} + \cdots + X_{i,n}, 1 \le i \le k) \to (s_1, \cdots, s_k)$$

which is a semi-circle family in some $*$-probability space $(\mathcal{A}, \varphi_0)$ with $\kappa_2(s_i, s_j) = \kappa_2(X_{i,1}, X_{j,1})$ for all $1 \le i, j \le k$.

Proof: [Proof of Theorem 4]

We observe that

$$\sum_{1 \le i_1 < i_2 \le n} \sum_{k=1}^{K} a_k (f_k(X_{i_1}) f_k(X_{i_2}) + f_k(X_{i_2}) f_k(X_{i_1})) = \sum_{k=1}^{K} a_k (\sum_{i=1}^{n} f_k(X_i))^2 - \sum_{k=1}^{K} \sum_{i=1}^{n} a_k f_k^2(X_i).$$

Therefore

$$nU_n = \frac{2n}{n-1} \sum_{k=1}^{K} a_k \Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_k(X_i) \Big)^2 - \frac{2n}{n-1} \sum_{k=1}^{K} a_k \frac{1}{n} \sum_{i=1}^{n} f_k^2(X_i).$$

By the free Central Limit Theorem 5,

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} f_1(X_i), \cdots, \sum_{i=1}^{n} f_K(X_i), \sum_{i=1}^{n} (f_1^2(X_i) - 1), \cdots, \sum_{i=1}^{n} (f_K^2(X_i) - 1) \right) \to (s_1, \cdots, s_K, t_1, \cdots, t_K)$$

which is a semi-circular family. Moreover, using conditions (a) and (b), $s_1, \cdots, s_K$ are all freely independent identically distributed semi-circle variables with variance 1. The exact parameters for $(t_1, \cdots, t_K)$ shall not be important to us.

Let

$$A_{n,k,1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_k(X_i) \text{ and } A_{n,k,0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f_k^2(X_i) - 1), 1 \le k \le K.$$

Also let

$$b_{k,0} = s_k,\ b_{k,1} = t_k,\ 1 \le k \le K.$$

Then by definition of joint convergence, for any $j \ge 1$, for any $k_1, ..., k_j \in \{1, ..., K\}$ and for any $\epsilon_1, ..., \epsilon_j \in \{0, 1\}$,

$$\kappa_j(A_{n,k_1,\epsilon_1}, \cdots, A_{n,k_j,\epsilon_j}) \to \kappa_j(b_{k_1,\epsilon_1}, \cdots, b_{k_j,\epsilon_j}).$$

If $j > 1$ and $\epsilon_1, ..., \epsilon_j$ are not all 0, say $\epsilon_1 = 1$ without loss of generality, then using the fact that constants are free of everything,

$$\kappa_j(A_{n,k_1,\epsilon_1}, \cdots, A_{n,k_j,\epsilon_j}) = \kappa_j(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{k_1}^2(X_i), A_{n,k_2,\epsilon_2}, \cdots, A_{n,k_j,\epsilon_j}) \to \kappa_j(t_{k_1}, b_{k_2,\epsilon_2}, \cdots, b_{k_j,\epsilon_j})$$

and therefore, $\kappa_j(\frac{1}{n}\sum_{i=1}^{n} f_{k_1}^2(X_i), A_{n,k_2,\epsilon_2}, \cdots, A_{n,k_j,\epsilon_j}) \to 0$. Further, if $j = 1$, then for any $1 \le k \le K$, $\kappa_1(\frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i) - 1) = 0$ and hence $\kappa_1(\frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i)) \to 1$. This shows the following joint convergence:

$$(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_1(X_i), \cdots, \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_K(X_i), \frac{1}{n}\sum_{i=1}^{n} f_1^2(X_i), \cdots, \frac{1}{n}\sum_{i=1}^{n} f_K^2(X_i)) \to (s_1, \cdots, s_K, 1, \cdots, 1)$$

Therefore, we have

$$nU_n = \frac{2n}{n-1}\sum_{k=1}^{K} a_k(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_k(X_i))^2 - \frac{2n}{n-1}\sum_{k=1}^{K} a_k \frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i) \to 2\sum_{k=1}^{K} a_k(s_k^2 - 1).$$

□

We have crucially used the representation of the kernel. It is not clear how to obtain a limit theorem for a more general kernel. We intend to pursue this direction in future.

Acknowledgements

We are grateful to the Referees for very constructive comments. These have led to a substantial improvement in the presentation. Most of the work in this paper appeared in Chapter 3 of Dey (2019), the Masters dissertation of Apratim Dey under the guidance of Arup Bose. Research of Arup Bose is supported by a J. C. Bose National Fellowship, Department of Science and Technology, Government of India.

References

Bose, A. and Bhattacharjee, M. (2018). *Large Covariance and Autocovariance Matrices.* CRC Press, Boca Raton.

Bose, A. and Chatterjee, S. (2018). *U-Statistics, M_m-Estimators and Resampling.* Springer Texts and Readings in Mathematics.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* SIAM.

Dey, A. (2019). *Random Matrices and Free Probability.* Masters Dissertation, Indian Statistical Institute, Kolkata.

Nica, A. and Speicher, R. (2006). *Lectures on the Combinatorics of Free Probability.* London Mathematical Society Lecture Note Series 335, Cambridge University Press.

Voiculescu, D. (1991). Limit laws for random matrices and free products. *Inventiones Mathematicae*, **104**, **1**, 201-220.