Variations on a Theorem
by Peter de Jong

Giovanni Peccati (Luxembourg University)

Paris, IHP — June 7th, 2022
Introduction

★ **Topic:** multivariate and functional fluctuations of *U-statistics.*

★ *In the last 15 years:* proof of several **fourth moment theorems** for sequences of random variables living in eigenspaces of Markov operators, e.g.: **Gaussian Wiener chaos** (Nualart & Peccati, 2005; Nourdin & Peccati, 2007–10); **Poisson Wiener chaos** (Peccati, Solé, Taqqu & Utzet, 2010; Döbler & Peccati, 2018); **diffusive Markov operators** (Ledoux, 2010; Azmoodeh, Campese & Poly, 2013).

★ **Applications:** mathematical statistics, mathematical physics, stochastic geometry (random geometric graphs & geometry of random fields), ...

★ This line of research was anticipated (in 1990!) by **P. de Jong.**
**Introduction**

★ **Topic**: multivariate and functional fluctuations of \( U \)-statistics.

★ **In the last 15 years**: proof of several *fourth moment theorems* for sequences of random variables living in eigenspaces of Markov operators, e.g.: **Gaussian Wiener chaos** (Nualart & Peccati, 2005; Nourdin & Peccati, 2007–10); **Poisson Wiener chaos** (Peccati, Solé, Taqqu & Utzet, 2010; Döbler & Peccati, 2018); **diffusive Markov operators** (Ledoux, 2010; Azmoodeh, Campese & Poly, 2013).

★ **Applications**: mathematical statistics, mathematical physics, stochastic geometry (random geometric graphs & geometry of random fields), ... 

★ This line of research was anticipated (in 1990!) by **P. de Jong**.
Introduction

★ **Topic**: multivariate and functional fluctuations of *U*-statistics.

★ *In the last 15 years*: proof of several **fourth moment theorems** for sequences of random variables living in eigenspaces of Markov operators, e.g.: **Gaussian Wiener chaos** (Nualart & Peccati, 2005; Nourdin & Peccati, 2007–10); **Poisson Wiener chaos** (Peccati, Solé, Taqqu & Utzet, 2010; Döbler & Peccati, 2018); **diffusive Markov operators** (Ledoux, 2010; Azmoodeh, Campese & Poly, 2013).

★ **Applications**: mathematical statistics, mathematical physics, stochastic geometry (random geometric graphs & geometry of random fields), ...

★ This line of research was anticipated (in 1990!) by P. de Jong.
**Introduction**

- **Topic**: multivariate and functional fluctuations of \( U \)-statistics.

- **In the last 15 years**: proof of several **fourth moment theorems** for sequences of random variables living in eigenspaces of Markov operators, e.g.: **Gaussian Wiener chaos** (Nualart & Peccati, 2005; Nourdin & Peccati, 2007–10); **Poisson Wiener chaos** (Peccati, Solé, Taqqu & Utzet, 2010; Döbler & Peccati, 2018); **diffusive Markov operators** (Ledoux, 2010; Azmoodeh, Campese & Poly, 2013).

- **Applications**: mathematical statistics, mathematical physics, stochastic geometry (random geometric graphs & geometry of random fields), ... 

- This line of research was anticipated (in 1990!) by **P. de Jong**.
**FRAMEWORK, I**

★ Let \((X_1, \ldots, X_n)\) be independent random elements with values in \((E, \mathcal{E})\).

★ For \(k = 1, \ldots, n\), a (square-integrable) \textit{U-statistic} of order \(k\) is a random variable with the form

\[
W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} g(i_1, \ldots, i_k)(X_{i_1}, \ldots, X_{i_k}),
\]

with \(g(i_1, \ldots, i_k) : E^k \to \mathbb{R}\) square-integrable.

★ \(W\) is **symmetric** if the \(X_i\)’s are \textbf{i.i.d.} and \(g(i_1, \ldots, i_k) \equiv g\) is symmetric.

★ \(W\) is **degenerate** if

\[
\mathbb{E}[g(i_1, \ldots, i_k)(X_{i_1}, \ldots, X_{i_k}) | X_a : a \in A] = 0,
\]

for all \(A \subsetneq \{i_1, \ldots, i_k\}\).
Let \((X_1, \ldots, X_n)\) be independent random elements with values in \((E, \mathcal{E})\).

For \(k = 1, \ldots, n\), a (square-integrable) \textbf{U-statistic} of order \(k\) is a random variable with the form

\[
W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} g(i_1, \ldots, i_k)(X_{i_1}, \ldots, X_{i_k}),
\]

with \(g(i_1, \ldots, i_k) : E^k \to \mathbb{R}\) square-integrable.

\(W\) is \textbf{symmetric} if the \(X_i\)'s are \textit{i.i.d.} and \(g(i_1, \ldots, i_k) \equiv g\) is symmetric.

\(W\) is \textbf{degenerate} if

\[
\mathbb{E}[g(i_1, \ldots, i_k)(X_{i_1}, \ldots, X_{i_k}) \mid X_a : a \in A] = 0,
\]

for all \(A \subset \neq \{i_1, \ldots, i_k\}\).
FRAMEWORK, I

★ Let \((X_1, ..., X_n)\) be independent random elements with values in \((E, \mathcal{E})\).

★ For \(k = 1, ..., n\), a (square-integrable) **U-statistic** of order \(k\) is a random variable with the form

\[
W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} g(i_1, ..., i_k)(X_{i_1}, ..., X_{i_k}),
\]

with \(g(i_1, ..., i_k) : E^k \to \mathbb{R}\) square-integrable.

★ \(W\) is **symmetric** if the \(X_i\)'s are i.i.d. and \(g(i_1, ..., i_k) \equiv g\) is symmetric.

★ \(W\) is **degenerate** if

\[
\mathbb{E}[g(i_1, ..., i_k)(X_{i_1}, ..., X_{i_k}) | X_a : a \in A] = 0,
\]

for all \(A \subsetneq \{i_1, ..., i_k\}\).
Let \((X_1, ..., X_n)\) be independent random elements with values in \((E, \mathcal{E})\).

For \(k = 1, ..., n\), a (square-integrable) \textbf{U-statistic} of order \(k\) is a random variable with the form

\[
W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} g_{(i_1, ..., i_k)}(X_{i_1}, ..., X_{i_k}),
\]

with \(g_{(i_1, ..., i_k)} : E^k \to \mathbb{R}\) square-integrable.

\(W\) is \textbf{symmetric} if the \(X_i\)'s are \textbf{i.i.d.} and \(g_{(i_1, ..., i_k)} \equiv g\) is symmetric.

\(W\) is \textbf{degenerate} if

\[
\mathbb{E}[g_{(i_1, ..., i_k)}(X_{i_1}, ..., X_{i_k}) \mid X_a : a \in A] = 0,
\]

for all \(A \subsetneq \{i_1, ..., i_k\}\).
**Framework, II**

★ **Fact:** every square-integrable $F = F(X_1, ..., X_n)$ can be uniquely decomposed into a sum of degenerate $U$-statistics of order $k = 0, 1, ..., n$ (**Hoeffding decomposition**).

★ In the case where $X_1, ..., X_n$ are centered and real-valued, classical examples of **non-symmetric and degenerate $U$-statistics** are **homogeneous sums**:

$$ W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{(i_1, ..., i_k)} X_{i_1} \cdots X_{i_k}, \quad a_{(i_1, ..., i_k)} \in \mathbb{R}. $$

★ For **Rademacher variables**: “degenerate $U$-statistics of order $k$” $\Leftrightarrow$ “homogeneous sums of order $k$” $\Leftrightarrow$ “$k$th Walsh chaos”.

★ The **maximal influence** associated with a degenerate $W$ is

$$ \text{Inf}(W) = \max_{i=1, \ldots, n} \sum_{i \in \{i_1, \ldots, i_k\}} \mathbb{E} [g_{(i_1, \ldots, i_k)}(X_{i_1}, ..., X_{i_k})^2]. $$
**Framework, II**

- **Fact**: every square-integrable $F = F(X_1, ..., X_n)$ can be uniquely decomposed into a sum of degenerate $U$-statistics of order $k = 0, 1, ..., n$ (Hoeffding decomposition).

- In the case where $X_1, ..., X_n$ are centered and real-valued, classical examples of non-symmetric and degenerate $U$-statistics are **homogeneous sums**:

$$W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} a(i_1, ..., i_k) X_{i_1} \cdots X_{i_k}, \quad a(i_1, ..., i_k) \in \mathbb{R}.$$  

- For **Rademacher variables**: “degenerate $U$-statistics of order $k$” $\iff$ “homogeneous sums of order $k$” $\iff$ “$k$th Walsh chaos”.

- The **maximal influence** associated with a degenerate $W$ is

$$\text{Inf}(W) = \max_{i=1,\ldots,n} \sum_{i \in \{i_1,\ldots,i_k\}} \mathbb{E}[g(i_1,\ldots,i_k)(X_{i_1},\ldots,X_{i_k})^2]$$
**Fact**: every square-integrable $F = F(X_1, ..., X_n)$ can be uniquely decomposed into a sum of degenerate $U$-statistics of order $k = 0, 1, ..., n$ (Hoeffding decomposition).

In the case where $X_1, ..., X_n$ are centered and real-valued, classical examples of non-symmetric and degenerate $U$-statistics are **homogeneous sums**:

$$W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{i_1, \ldots, i_k} X_{i_1} \cdots X_{i_k}, \quad a_{i_1, \ldots, i_k} \in \mathbb{R}.$$

For **Rademacher variables**: “degenerate $U$-statistics of order $k$” $\Leftrightarrow$ “homogeneous sums of order $k$” $\Leftrightarrow$ “$k$th Walsh chaos”.

The **maximal influence** associated with a degenerate $W$ is

$$\text{Inf}(W) = \max_{i=1,\ldots,n} \sum_{i \in \{i_1,\ldots,i_k\}} \mathbb{E}[g_{i_1,\ldots,i_k}(X_{i_1}, \ldots, X_{i_k})^2].$$
**Framework, II**

- **Fact:** every square-integrable $F = F(X_1, ..., X_n)$ can be uniquely decomposed into a sum of degenerate $U$-statistics of order $k = 0, 1, ..., n$ (Hoeffding decomposition).

- In the case where $X_1, ..., X_n$ are centered and real-valued, classical examples of non-symmetric and degenerate $U$-statistics are **homogeneous sums**:

  $$ W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{(i_1, \ldots, i_k)} X_{i_1} \cdots X_{i_k}, \quad a_{(i_1, \ldots, i_k)} \in \mathbb{R}. $$

- For **Rademacher variables**: “degenerate $U$-statistics of order $k$” $\Leftrightarrow$ “homogeneous sums of order $k$” $\Leftrightarrow$ “$k$th Walsh chaos”.

- The **maximal influence** associated with a degenerate $W$ is

  $$ \text{Inf}(W) = \max_{i=1,\ldots,n} \sum_{i \in \{i_1,\ldots,i_k\}} \mathbb{E}[g_{(i_1,\ldots,i_k)}(X_{i_1},\ldots,X_{i_k})^2] $$
In a beautiful 1990 paper, Peter de Jong proved a surprising result for a sequence \( \{W_n : n \geq 1\} \) of normalised degenerate \( U \)-statistics, that is:

If \( \text{Inf}(W_n) \to 0 \), then \( W_n \) verifies a Central Limit Theorem provided

\[
\mathbb{E}W_n^4 \to 3 = \mathbb{E}N(0,1)^4.
\]
In a beautiful 1990 paper, Peter de Jong proved a surprising result for a sequence \( \{W_n : n \geq 1\} \) of normalised degenerate U-statistics, that is:

If \( \text{Inf}(W_n) \to 0 \), then \( W_n \) verifies a Central Limit Theorem provided

\[ \mathbb{E}W_n^4 \to 3 \quad (= \mathbb{E}N(0,1)^4). \]
**Remarks**

★ *Original proof:* **martingale CLT** + (heavy) combinatorial analysis.

★ In Döbler & Peccati (EJP, 2016) (using *Stein’s method*): Let $W_n = (W_n(1), ..., W_n(d))$, $n \geq 1$, be vectors of degenerate U-statistics such that (i) $\text{Cov}(W_n) \to \Sigma$, (ii) $\max_i \text{Inf}(W_n(i)) \to 0$, (iii) $\mathbb{E}[(W_n(i))^2(W_n(j))^2] \to \Sigma(i, i)\Sigma(j, j) + 2\Sigma(i, j)^2$.

Then, $W_n \Rightarrow N_d(0, \Sigma) + \text{quantitative bounds}.$

★ *Combinatorial component of 2016 proofs:* basically, **unchanged since de Jong (1990).**

★ Careful bookkeeping of constants allows for $d = d_n \to \infty$ as soon as $d_n! \ll \max_\ell \text{Inf}(W_n(\ell))^\alpha$ (some $\alpha > 0$; absolutely not sharp).
REMARKS

★ Original proof: martingale CLT + (heavy) combinatorial analysis.

★ In Döbler & Peccati (EJP, 2016) (using Stein’s method): Let $W_n = (W_n^{(1)},...,W_n^{(d)}), n \geq 1$, be vectors of degenerate U-statistics such that (i) $\text{Cov}(W_n) \to \Sigma$, (ii) $\max_i \text{Inf}(W_n^{(i)}) \to 0$, (iii) $\mathbb{E}[(W_n^{(i)})^2(W_n^{(j)})^2] \to \Sigma(i,i)\Sigma(j,j) + 2\Sigma(i,j)^2$. Then, $W_n \Rightarrow N_d(0,\Sigma) + \text{quantitative bounds}$.


★ Careful bookkeeping of constants allows for $d = d_n \to \infty$ as soon as $d_n! \ll \max_{\ell} \text{Inf}(W_n^{(\ell)})^\alpha$ (some $\alpha > 0$; absolutely not sharp).
**Remarks**

★ **Original proof:** martingale CLT + (heavy) combinatorial analysis.

★ In Döbler & Peccati (EJP, 2016) (using Stein’s method): Let $W_n = (W_n^{(1)}, ..., W_n^{(d)})$, $n \geq 1$, be vectors of degenerate U-statistics such that (i) $\text{Cov}(W_n) \to \Sigma$, (ii) $\max_i \text{Inf}(W_n^{(i)}) \to 0$, (iii) $\mathbb{E}[(W_n^{(i)})^2(W_n^{(j)})^2] \to \Sigma(i,i)\Sigma(j,j) + 2\Sigma(i,j)^2$.

Then, $W_n \Rightarrow N_d(0, \Sigma) + \text{quantitative bounds}$.

★ **Combinatorial component of 2016 proofs:** basically, unchanged since de Jong (1990).

★ Careful bookkeeping of constants allows for $d = d_n \to \infty$ as soon as $d_n! \ll \max_\ell \text{Inf}(W_n^{(\ell)})^\alpha$ (some $\alpha > 0$; absolutely not sharp).
**Remarks**

* Original proof: **martingale CLT** + (heavy) combinatorial analysis.

* In Döbler & Peccati (EJP, 2016) (using **Stein’s method**): Let $W_n = (W_n^{(1)},...,W_n^{(d)}), n \geq 1$, be vectors of degenerate U-statistics such that (i) $\text{Cov}(W_n) \to \Sigma$, (ii) $\max_i \text{Inf}(W_n^{(i)}) \to 0$, (iii) $\mathbb{E}[(W_n^{(i)})^2(W_n^{(j)})^2] \to \Sigma(i,i)\Sigma(j,j) + 2\Sigma(i,j)^2$. Then, $W_n \Rightarrow N_d(0, \Sigma) + \text{quantitative bounds}$.

* Combinatorial component of 2016 proofs: basically, **unchanged since de Jong (1990)**.

* Careful bookkeeping of constants allows for $d = d_n \to \infty$ as soon as $d_n! \ll \max_\ell \text{Inf}(W_n^{(\ell)})^\alpha$ (some $\alpha > 0$; absolutely not sharp).
Symmetric U-Processes

★ Let \( \{X_i : i \geq 1\} \) be a sequence of i.i.d. random elements.

★ Fix \( k \geq 2 \), and let \( g_n : E^k \to \mathbb{R} \) be a sequence of symmetric, square-integrable and degenerate kernels.

★ The sequential \( U \)-process associated with \( g_n \) is the random function

\[
t \mapsto U(g_n; t) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} g_n(X_{i_1}, \ldots, X_{i_k}), \quad t \in [0, 1].
\]

★ Problem: study the convergence of \( U(g_n; \cdot) \) in the Skorohod space \( D[0, 1] \). We write \( U_n := U(g_n; 1) \), and \( \sigma_n^2 := \text{Var}(U_n) \).
Symmetric $U$-Processes

★ Let $\{X_i : i \geq 1\}$ be a sequence of i.i.d. random elements.

★ Fix $k \geq 2$, and let $g_n : E^k \to \mathbb{R}$ be a sequence of symmetric, square-integrable and degenerate kernels.

★ The sequential $U$-process associated with $g_n$ is the random function

\[
t \mapsto U(g_n; t) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} g_n(X_{i_1}, ..., X_{i_k}), \quad t \in [0, 1].
\]

★ Problem: study the convergence of $U(g_n; \cdot)$ in the Skorohod space $D[0, 1]$. We write $U_n := U(g_n; 1)$, and $\sigma_n^2 := \text{Var}(U_n)$. 
**Symmetric U-Processes**

- Let \( \{X_i : i \geq 1\} \) be a sequence of i.i.d. random elements.

- Fix \( k \geq 2 \), and let \( g_n : E^k \to \mathbb{R} \) be a sequence of symmetric, square-integrable and degenerate kernels.

- The **sequential U-process** associated with \( g_n \) is the random function

\[
t \mapsto U(g_n; t) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} g_n(X_{i_1}, \ldots, X_{i_k}), \quad t \in [0,1].
\]

- **Problem**: study the convergence of \( U(g_n; \cdot) \) in the Skorohod space \( D[0,1] \). We write \( U_n := U(g_n; 1) \), and \( \sigma_n^2 := \text{Var}(U_n) \).
Symmetric U-Processes

★ Let \( \{ X_i : i \geq 1 \} \) be a sequence of \textbf{i.i.d. random elements}.

★ Fix \( k \geq 2 \), and let \( g_n : E^k \to \mathbb{R} \) be a sequence of \textbf{symmetric, square-integrable and degenerate kernels}.

★ The \textbf{sequential U-process} associated with \( g_n \) is the random function

\[
t \mapsto U(g_n; t) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} g_n(X_{i_1}, \ldots, X_{i_k}), \quad t \in [0, 1].
\]

★ \textbf{Problem}: study the convergence of \( U(g_n; \cdot) \) in the \textbf{Skorohod space} \( D[0, 1] \). We write \( U_n := U(g_n; 1) \), and \( \sigma_n^2 := \text{Var}(U_n) \).
Döbler, Kasprzak & Peccati (AAP, 2022)

Let \( \tilde{U}(g_n; \cdot) := U_n(g_n; \cdot)/\sigma_n \) and \( \tilde{U}_n := U_n/\sigma_n \). Assume that

(a) \( \inf(\tilde{U}_n) \to 0 \);

(b) \( |E[\tilde{U}_n^4] - 3| \ll n^{-\eta} \) for some \( \eta > 0 \) + “technicalities”.

Then,

\[ \{U(g_n; t) : t \in [0, 1]\} \Rightarrow \{B(t^k) : t \in [0, 1]\}, \]

in \( D[0, 1] \), where \( B \) is a standard Brownian motion.

Remarks: (i) Condition (b) is checked by using “contraction operators”

(ii) A perfectly working multivariate version is also available.
Döbler, Kasprzak & Peccati (AAP, 2022)

Let $\tilde{U}(g_n; \cdot) := U_n(g_n; \cdot) / \sigma_n$ and $\tilde{U}_n := U_n / \sigma_n$. Assume that

(a) $\inf (\tilde{U}_n) \to 0$;

(b) $|E[\tilde{U}_n^4] - 3| \ll n^{-\eta}$ for some $\eta > 0$ + “technicalities”.

Then,

$$\{U(g_n; t) : t \in [0, 1]\} \Rightarrow \{B(t^k) : t \in [0, 1]\},$$

in $D[0, 1]$, where $B$ is a standard Brownian motion.

Remarks: (i) Condition (b) is checked by using “contraction operators”

(ii) A perfectly working multivariate version is also available.
Döbler, Kasprzak & Peccati (AAP, 2022)

Let \( \tilde{U}(g_n; \cdot) := U_n(g_n; \cdot) / \sigma_n \) and \( \tilde{U}_n := U_n / \sigma_n \). Assume that

(a) \( \inf(\tilde{U}_n) \to 0 \);

(b) \( \left| \mathbb{E}[\tilde{U}_n^4] - 3 \right| \ll n^{-\eta} \) for some \( \eta > 0 \) + “technicalities”.

Then,

\[
\{ U(g_n; t) : t \in [0, 1] \} \implies \{ B(t^k) : t \in [0, 1] \},
\]

in \( D[0, 1] \), where \( B \) is a standard Brownian motion.

Remarks: (i) Condition (b) is checked by using “contraction operators”

(ii) A perfectly working multivariate version is also available.
Consider a normalized, non-degenerate (centered and symmetric) $U$-process $U_n$ on $[0, 1]$, along with its Hoeffding decomposition:

$$U_n(t) = \sum_{1 \leq i_1 < i_2 < \cdots < i_p \leq \lfloor nt \rfloor} G_n(X_{i_1}, \ldots, X_{i_p}) = \sum_{k=1}^{p} U_n^{(k)}(t).$$

Döbler, Kasprzak & Peccati (AAP, 2022)

Assume that $\text{Var}(U_n^{(k)}(1)) \to b_k^2$, and that each $U_n^{(k)}$ verifies (an adequate version of) the assumptions of the previous theorem. Then,

$$U_n \Longrightarrow \left\{ \sum_{k=1}^{p} b_k^2 \times t^{p-k} B_k(t^k), \quad t \in [0, 1] \right\},$$

with $B_k$ independent Brownian motions.
GENERALIZING MILLER AND SEN (1972)

Consider a normalized, non-degenerate (centered and symmetric) $U$-process $U_n$ on $[0, 1]$, along with its Hoeffding decomposition:

$$U_n(t) = \sum_{1 \leq i_1 < i_2 < \cdots < i_p \leq \lfloor nt \rfloor} G_n(X_{i_1}, \ldots, X_{i_p}) = \sum_{k=1}^{p} U_n^{(k)}(t).$$

Döbler, Kasprzak & Peccati (AAP, 2022)

Assume that $\text{Var}(U_n^{(k)}(1)) \to b_k^2$, and that each $U_n^{(k)}$ verifies (an adequate version of) the assumptions of the previous theorem. Then,

$$U_n \Longrightarrow \left\{ \sum_{k=1}^{p} b_k^2 \times t^{p-k} B_k(t^k), \quad t \in [0, 1] \right\},$$

with $B_k$ independent Brownian motions.
Let \( r_n \downarrow 0 \) consider \( X_1, \ldots, X_n \) i.i.d. uniform points on the unit cube \( \subset \mathbb{R}^d \), and connect \( X_\ell \) and \( X_j \) if \( 0 < \| X_\ell - X_j \| < r_n \):

We are interested in the changepoint empirical process

\[
S_n(t) := \sum_{1 \leq i \leq \lfloor nt \rfloor < j \leq n} \mathbf{1}_{\{X_i \sim X_j\}}, \quad t \in [0, 1].
\]

Instance of graph-based changepoint detection: see Chen & Zhang (2015)
Example (Change-point Analysis)

* Let $r_n \downarrow 0$ Consider $X_1, ..., X_n$ i.i.d. uniform points on the unit cube $\subset \mathbb{R}^d$, and connect $X_\ell$ and $X_j$ if $0 < \|X_\ell - X_j\| < r_n$:

* We are interested in the changepoint empirical process

$$S_n(t) := \sum_{1 \leq i \leq \lfloor nt \rfloor < j \leq n} 1_{\{X_i \sim X_j\}}, \quad t \in [0, 1].$$

* Instance of graph-based changepoint detection: see Chen & Zhang (2015)
**Example (Changepoint Analysis)**

- Let $r_n \downarrow 0$ Consider $X_1, \ldots, X_n$ i.i.d. uniform points on the unit cube $\subset \mathbb{R}^d$, and connect $X_\ell$ and $X_j$ if $0 < \|X_\ell - X_j\| < r_n$:

- We are interested in the **changepoint empirical process**

$$S_n(t) := \sum_{1 \leq i \leq \lfloor nt \rfloor < j \leq n} 1_{\{X_i \sim X_j\}}, \quad t \in [0, 1].$$

- Instance of **graph-based changepoint detection**: see Chen & Zhang (2015)
**Example (Change Point Analysis)**

⋆ Let $r_n \downarrow 0$ Consider $X_1, ..., X_n$ i.i.d. uniform points on the unit cube $\subset \mathbb{R}^d$, and connect $X_\ell$ and $X_j$ if $0 < \|X_\ell - X_j\| < r_n$:

![Diagram of points and connections](image)

⋆ We are interested in the change point empirical process

$$S_n(t) := \sum_{1 \leq i \leq \lfloor nt \rfloor < j \leq n} 1_{\{X_i \sim X_j\}}, \quad t \in [0, 1].$$

⋆ Instance of graph-based changepoint detection: see Chen & Zhang (2015)
Consider the **sparse regime**: $n r_n^d \to 0$ and $n^2 r_n^d \to \infty$. Then, there exists a constant $c > 0$ (depending on $d$) such that, setting $\sigma_n^2 := c n^2 r_n^d$, and

$$T_n(t) := \frac{S(n,t) - \mathbb{E}[S(n,t)]}{\sigma_n}, \quad t \in [0,1],$$

one has that, if $n^{2-\delta} \ll r_n^d \ll n^{-1}$ for some $\delta > 0$, then

$$T_n \implies \{\sqrt{2} b(t) : t \in [0,1]\},$$

where $b$ is a standard Brownian bridge. In particular:

$$\arg \max_{t \in [0,1]} (-T_n(t)) \implies U_{[0,1]}.$$
**Homogeneous Sums: de Jong & Universality**

Nourdin, Peccati & Reinert (AoP, 2010)

Let $G = \{G_i : i \geq 1\}$ be i.i.d. $N(0, 1)$. Consider a sequence of normalized homogeneous sums of order $k \geq 2$:

$$Q_n(G) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} a^{(n)}_{i_1, \ldots, i_k} G_{i_1} \cdots G_{i_k}, \quad n \geq 1.$$  

If $\mathbb{E} Q_n(G)^4 \to 3$, then

$$Q_n(X) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} a^{(n)}_{i_1, \ldots, i_k} X_{i_1} \cdots X_{i_k} \Rightarrow N(0, 1),$$

for every sequence $X := \{X_i : i \geq 1\}$ of independent centered r.v.'s such that $\sup_i \mathbb{E} |X_i|^{2+\epsilon} < \infty$.

**Recent applications:** Caravenna, Sun & Zygouras (2016++, polymers); Angst & Poly (2021, zeros of random polynomials).
**Homogeneous Sums: de Jong & Universality**

**Nourdin, Peccati & Reinert (AoP, 2010)**

Let $G = \{G_i : i \geq 1\}$ be i.i.d. $N(0,1)$. Consider a sequence of normalized homogeneous sums or order $k \geq 2$:

$$Q_n(G) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} a^{(n)}_{(i_1, \ldots, i_k)} G_{i_1} \cdots G_{i_k}, \quad n \geq 1.$$  

If $\mathbb{E} Q_n(G)^4 \to 3$, then

$$Q_n(X) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} a^{(n)}_{(i_1, \ldots, i_k)} X_{i_1} \cdots X_{i_k} \Rightarrow N(0,1),$$

for every sequence $X := \{X_i : i \geq 1\}$ of independent centered r.v.'s such that $\sup_i \mathbb{E}|X_i|^{2+\epsilon} < \infty$.

**Recent applications**: Caravenna, Sun & Zygouras (2016++, polymers); Angst & Poly (2021, zeros of random polynomials).
Consider **sequential U-processes** associated with normalized non-symmetric U-statistics:

\[ t \mapsto U_n(t) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} g^{(n)}_{(i_1, \ldots, i_k)}(X_{i_1}, \ldots, X_{i_k}), \quad t \in [0, 1]. \]

Assume that

(a) \( \text{Inf}(U_n(1)) \to 0; \)
(b) \( \mathbb{E}[U_n(1)^4] \to 3; \)
(c) \( \mathbb{E}[g^{(n)}_{(i_1, \ldots, i_k)}(X_{i_1}, \ldots, X_{i_k})^4] \leq C\mathbb{E}[g^{(n)}_{(i_1, \ldots, i_k)}(X_{i_1}, \ldots, X_{i_k})^2]^2 \]

Then, \( \{U_n\} \) is relatively compact in \( D[0,1] \) and every adherent point corresponds to the law of a **continuous Gaussian process**.
Döbler, Kasprzak & Peccati (PTRF, 2022+)

Let \( \mathbf{G} = \{ G_i : i \geq 1 \} \) be i.i.d. \( N(0,1) \). Consider a sequence of normalized homogeneous \( U \)-processes:

\[
Q_{\lfloor nt \rfloor}(\mathbf{G}) := \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} a^{(n)}_{(i_1, \ldots, i_k)} G_{i_1} \cdots G_{i_k}, \quad t \in [0,1].
\]

If \( Q_{\lfloor nt \rfloor}(\mathbf{G}) \) converges in \( D[0,1] \) to a continuous Gaussian process, then the same holds for

\[
Q_{\lfloor nt \rfloor}(\mathbf{X}) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor nt \rfloor} a^{(n)}_{(i_1, \ldots, i_k)} X_{i_1} \cdots X_{i_k},
\]

for every sequence \( \mathbf{X} := \{ X_i : i \geq 1 \} \) of independent centered r.v.'s such that \( \sup_i \mathbb{E}|X_i|^4 < \infty \).
**Fractional Products**

- Fix $k \geq 3$, as well as $m = 2, \ldots, k - 1$.
- Write $N = n^m$, and let $\varphi : [n]^m \to [N]$ be one-to-one.
- Consider a **connected** $m$-cover $S_1, \ldots, S_k$ of $[k]$, that is: (i) $\bigcup_i S_i = [k]$, (ii) $|S_i| = m$ and (iii) each index $i \in [k]$ appears in exactly $m$ subsets $S_i$.
- We define

  $$F_N := \{(\varphi(\pi_{S_1} a), \ldots, \varphi(\pi_{S_k} a)) : a \in [n]^k\},$$

  and denote by $\tilde{F}_N$ its symmetrization.
- Then $\tilde{F}_N$ is a symmetric subset of $[N]^k$ s.t. $|\tilde{F}_N| \asymp N^{k/m}$

(Fractional Cartesian Product)
Döbler, Kasprzak & Peccati (PTRF, 2022+)

For every sequence $X := \{X_i : i \geq 1\}$ of independent centered r.v.’s such that $\sup_i \mathbb{E}|X_i|^4 < \infty$,
the empirical process

$$Q_{\lfloor Nt \rfloor}(X) = \frac{1}{|\tilde{F}_N|^{1/2}} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor Nt \rfloor} 1_{\{i_1, \ldots, i_k\} \in \tilde{F}_N} X_{i_1} \cdots X_{i_k},$$

weakly converges to a multiple of $\{B(t^{k/m}) : t \in [0, 1]\}$, where $B$ is a standard Brownian motion (*).

(*) Not achievable by symmetric statistics.
THANK YOU FOR YOUR ATTENTION!