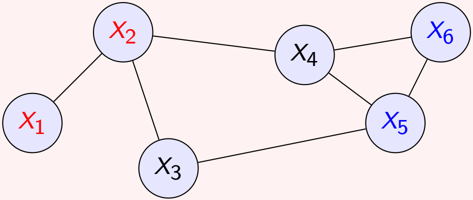
The recent development of dependency graphs

Rui-Ray Zhang
rui.zhang@monash.edu
School of Mathematics, Monash University

Dependency Graphs

Graph G is a dependency graph for random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ if

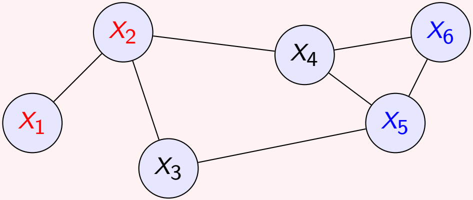
- Vertex set $V(G) = [n] = \{1, \ldots, n\}$.



Dependency Graphs

Graph G is a dependency graph for random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ if

- Vertex set $V(G) = [n] = \{1, \ldots, n\}$.

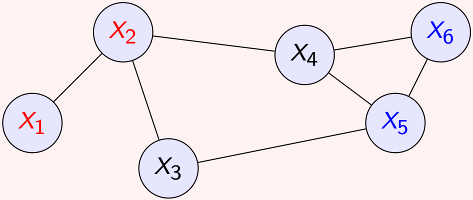


- If disjoint subsets $I, J \subset [n]$ are non-adjacent in G, then $\{X_i\}_{i \in I}$ and $\{X_j\}_{j \in J}$ are independent.
    - In the above example, $\{X_1, X_2\}$ and $\{X_5, X_6\}$ are independent.

Dependency Graphs

Graph G is a dependency graph for random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ if

- Vertex set $V(G) = [n] = \{1, \ldots, n\}$.



- If disjoint subsets $I, J \subset [n]$ are non-adjacent in G, then $\{X_i\}_{i \in I}$ and $\{X_j\}_{j \in J}$ are independent.
  - In the above example, $\{X_1, X_2\}$ and $\{X_5, X_6\}$ are independent.

- There are ones with weaker assumptions, such as the one used in Lovász local lemma.
- The dependency graph for variables may not necessarily be unique, and the sparser ones are the more interesting ones.

Classics

- Concentration under graph-dependence.
  - Janson's concentration bound (2004)

Classics

- Concentration under graph-dependence.
  - Janson's concentration bound (2004)
- The probability of non-existence of certain structures under graph-dependence.
  - Lovász local lemma (1975).
  - Chen-Stein's method (1975).

Classics

- Concentration under graph-dependence.
  - Janson's concentration bound (2004)
- The probability of non-existence of certain structures under graph-dependence.
  - Lovász local lemma (1975).
  - Chen-Stein's method (1975).
  - Janson's inequality (1990).
  - Suen's inequality (1990).

Classics

- Concentration under graph-dependence.
  - Janson's concentration bound (2004)
- The probability of non-existence of certain structures under graph-dependence.
  - Lovász local lemma (1975).
  - Chen-Stein's method (1975).
  - Janson's inequality (1990).
  - Suen's inequality (1990).
- Limiting distribution via dependency graph.
  - Janson's normality criterion via cumulants (1988)

Concentration bounds

A McDiarmid-type bounded difference inequality for graph-dependent variables.

Theorem (Z., Liu, Wang, Wang, 2019)

If we have

1. f is $\mathbf{c}$-Lipschitz,
2. F is a dependency graph for $\boldsymbol{X}$ and $F = \{T_i\}_{i \in [k]}$ is a forest,

Concentration bounds

A McDiarmid-type bounded difference inequality for graph-dependent variables.

Theorem (Z., Liu, Wang, Wang, 2019)

If we have

1. f is $\mathbf{c}$-Lipschitz,
2. F is a dependency graph for $\boldsymbol{X}$ and $F = \{T_i\}_{i \in [k]}$ is a forest,

then for every $t > 0$,

$$\mathbb{P}(f(\boldsymbol{X}) - \mathbb{E}[f(\boldsymbol{X})] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^k c_{\min,i}^2 + \sum_{\{i,j\} \in E(F)} (c_i + c_j)^2}\right),$$

where $c_{\min,i} = \min\{c_j : j \in V(T_i)\}$ is the minimum entry of $\mathbf{c}$ in each tree T_i.

Concentration bounds

Janson, 2004

Let random variables $\{X_i\}_{i \in V(G)}$ be G-dependent such that every X_i takes values in an interval of length $c_i \geqslant 0$. Then, for every $t > 0$,

$$\mathbb{P}\left(\sum_{i \in V(G)} X_i - \mathbb{E}\left[\sum_{i \in V(G)} X_i \right] \geqslant t \right) \leqslant \exp\left(- \frac{2t^2}{\chi_f(G) \sum_{i \in V(G)} c_i^2} \right),$$

where $\chi_f(G)$ is the fractional chromatic number of G.

Concentration bounds

Janson, 2004

Let random variables $\{X_i\}_{i\in V(G)}$ be G-dependent such that every X_i takes values in an interval of length $c_i \geqslant 0$. Then, for every $t > 0$,

$$\mathbb{P}\left(\sum_{i\in V(G)} X_i - \mathbb{E}\left[\sum_{i\in V(G)} X_i\right] \geqslant t\right) \leqslant \exp\left(-\frac{2t^2}{\chi_f(G)\sum_{i\in V(G)} c_i^2}\right),$$

where $\chi_f(G)$ is the fractional chromatic number of G.

Z., 2022

Under the same setting,

$$\mathbb{P}\left(\sum_{i\in V(G)} X_i - \mathbb{E}\left[\sum_{i\in V(G)} X_i\right] \geqslant t\right) \leqslant \exp\left(-\frac{2t^2}{D}\right),$$

where D is optimised over fractional forest vertex covers of G.

Concentration bounds

Janson, 2004

Let random variables $\{X_i\}_{i\in V(G)}$ be G-dependent such that every X_i takes values in an interval of length $c_i \geqslant 0$. Then, for every $t > 0$,

$$\mathbb{P}\left(\sum_{i\in V(G)} X_i - \mathbb{E}\left[\sum_{i\in V(G)} X_i\right] \geqslant t\right) \leqslant \exp\left(-\frac{2t^2}{\chi_f(G)\sum_{i\in V(G)} c_i^2}\right),$$

where $\chi_f(G)$ is the fractional chromatic number of G.

Z., 2022

Under the same setting,

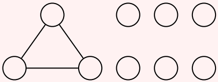
$$\mathbb{P}\left(\sum_{i\in V(G)} X_i - \mathbb{E}\left[\sum_{i\in V(G)} X_i\right] \geqslant t\right) \leqslant \exp\left(-\frac{2t^2}{D}\right),$$

where D is optimised over fractional forest vertex covers of G.

- It is no worse than Janson's, better when G is sparse.
- It generalises to certain decomposable Lipschitz functions, extending McDiarmid's.

Example

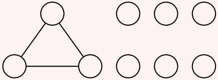
Let $\{X_i\}_{i\in[9]}$ be random indicators with the dependency graph G, and $X = \sum_{i\in[9]} X_i$.



$$\mathbb{P}(X - \mathbb{E}[X] \geqslant t) \leqslant ?$$

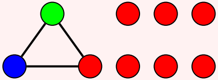
Example

Let $\{X_i\}_{i\in[9]}$ be random indicators with the dependency graph G, and $X = \sum_{i\in[9]} X_i$.



$$\mathbb{P}(X - \mathbb{E}[X] \geqslant t) \leqslant ?$$

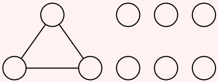
- Janson's idea: to cover vertices with weighted independent sets such that the sum of weights for each vertex equals 1



$$\mathbb{P}(X - \mathbb{E}[X] \geqslant t) \leqslant \exp\left(-\frac{2t^2}{3 \times 9}\right)$$

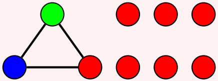
Example

Let $\{X_i\}_{i\in[9]}$ be random indicators with the dependency graph G, and $X = \sum_{i\in[9]} X_i$.



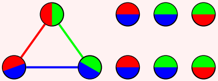
$$\mathbb{P}(X - \mathbb{E}[X] \geqslant t) \leqslant ?$$

- Janson's idea: to cover vertices with weighted independent sets such that the sum of weights for each vertex equals 1



$$\mathbb{P}(X - \mathbb{E}[X] \geqslant t) \leqslant \exp\left(-\frac{2t^2}{3 \times 9}\right)$$

- New idea: to cover vertices with weighted induced forests such that the sum of weights for each vertex equals 1



$$\mathbb{P}(X - \mathbb{E}[X] \geqslant t) \leqslant \exp\left(-\frac{8t^2}{81}\right)$$

The probability of non-occurrences

Let $\{X_i\}_i$ be G-dependent indicators for events and $X = \sum_i X_i$ count the occurrences of events (e.g., existence of certain combinatorial structure, say, $\triangle$ in $\mathscr{G}(n,p)$).

The probability of non-occurrences

Let $\{X_i\}_i$ be G-dependent indicators for events and $X = \sum_i X_i$ count the occurrences of events (e.g., existence of certain combinatorial structure, say, $\triangle$ in $\mathscr{G}(n,p)$).
Let the probability of joint occurrence denote by

$$\mu(S) = \mathbb{E}\left[\prod_{i \in S} X_i\right],$$

for $S \subseteq V(G)$.

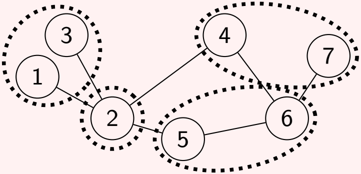
The probability of non-occurrences

Let $\{X_i\}_i$ be G-dependent indicators for events and $X = \sum_i X_i$ count the occurrences of events (e.g., existence of certain combinatorial structure, say, $\triangle$ in $\mathscr{G}(n,p)$).
Let the probability of joint occurrence denote by

$$\mu(S) = \mathbb{E}\left[\prod_{i \in S} X_i\right],$$

for $S \subseteq V(G)$.

Then we have a factorization property



$$\mu(1,3)\mu(2)\mu(4,7)\mu(5,6)$$
$$= \mu(1)\mu(2)\mu(3)\mu(4)\mu(5,6)\mu(7).$$

The probability of non-occurrences

Let $\{X_i\}_i$ be G-dependent indicators for events and $X = \sum_i X_i$ count the occurrences of events (e.g., existence of certain combinatorial structure, say, $\triangle$ in $\mathscr{G}(n,p)$).
Let the probability of joint occurrence denote by

$$\mu(S) = \mathbb{E}\left[\prod_{i \in S} X_i\right],$$

for $S \subseteq V(G)$.

By the principle of inclusion–exclusion and the factorization, we have

$$\mathbb{P}(X = 0) = \mathbb{P}\left(\sum_i X_i = 0\right)$$

The probability of non-occurrences

Let $\{X_i\}_i$ be G-dependent indicators for events and $X = \sum_i X_i$ count the occurrences of events (e.g., existence of certain combinatorial structure, say, $\triangle$ in $\mathscr{G}(n,p)$).
Let the probability of joint occurrence denote by

$$\mu(S) = \mathbb{E}\left[\prod_{i\in S} X_i\right],$$

for $S \subseteq V(G)$.

By the principle of inclusion–exclusion and the factorization, we have

$$\mathbb{P}(X=0) = \mathbb{P}\left(\sum_i X_i = 0\right) = \sum_{S\subseteq V(G)} (-1)^{|S|}\mu(S)$$

The probability of non-occurrences

Let $\{X_i\}_i$ be G-dependent indicators for events and $X = \sum_i X_i$ count the occurrences of events (e.g., existence of certain combinatorial structure, say, $\triangle$ in $\mathscr{G}(n,p)$).
Let the probability of joint occurrence denote by

$$\mu(S) = \mathbb{E}\left[\prod_{i \in S} X_i\right],$$

for $S \subseteq V(G)$.

By the principle of inclusion–exclusion and the factorization, we have

$$\mathbb{P}(X=0) = \mathbb{P}\left(\sum_i X_i = 0\right) = \sum_{S \subseteq V(G)} (-1)^{|S|} \mu(S) = \sum_{S \subseteq V(G)} (-1)^{|S|} \mu(C_1)\mu(C_2)\ldots\mu(C_k),$$

where $C_1 \cup \ldots \cup C_k = S$ and $C_1, \ldots, C_k$ induce maximal pairwise non-adjacent 'polymers' (that is, connected subgraphs; also called 'animals', suggested by Dobrushin, 1996).

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

Formal cluster expansion (Z. 2022)

$$\log(\mathbb{P}(X=0)) \overset{\text{formally}}{=} \sum_{k\geqslant 1} \sum_{(C_1,\ldots,C_k)} \phi(C_1,\ldots,C_k) \prod_{C\in(C_1,\ldots,C_k)} (-1)^{|C|}\mu(C),$$

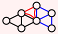
where the sum is over ordered multisets of polymers, whose union is connected.

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

Formal cluster expansion (Z. 2022)

$$\log(\mathbb{P}(X=0)) \overset{\text{formally}}{=} \sum_{k\geqslant 1}\sum_{(C_1,\dots,C_k)} \phi(C_1,\dots,C_k) \prod_{C\in(C_1,\dots,C_k)} (-1)^{|C|}\mu(C),$$

where the sum is over ordered multisets of polymers, whose union is connected.

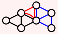
- Polymers (connected subgraphs): $\{\dots,$ , ,  $,\dots\}$.
- Clusters: $\dots,$  $,\dots$.

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

Formal cluster expansion (Z. 2022)

$$\log(\mathbb{P}(X=0)) \overset{\text{formally}}{=} \sum_{k\geqslant 1} \sum_{(C_1,\ldots,C_k)} \phi(C_1,\ldots,C_k) \prod_{C\in(C_1,\ldots,C_k)} (-1)^{|C|}\mu(C),$$

where the sum is over ordered multisets of polymers, whose union is connected.

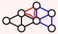
- Polymers (connected subgraphs): $\{\ldots,$ , ,  $,\ldots\}$.
- Clusters: $\ldots,$  $,\ldots$.
- Ursell function ϕ relates to chromatic polynomials.

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

Formal cluster expansion (Z. 2022)

$$\log(\mathbb{P}(X=0)) \overset{\text{formally}}{=} \sum_{k\geqslant 1} \sum_{(C_1,\ldots,C_k)} \phi(C_1,\ldots,C_k) \prod_{C\in(C_1,\ldots,C_k)} (-1)^{|C|}\mu(C),$$

where the sum is over ordered multisets of polymers, whose union is connected.

- Polymers (connected subgraphs): $\{\ldots,$ ,  ,  $,\ldots\}$.
- Clusters: …,  ,….
- Ursell function ϕ relates to chromatic polynomials.

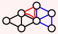
- Cluster expansion (Mayer, 1941) of partition function of a hard–core model with activity of each polymer C being $(-1)^{|C|}\mu(C)$.

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

Formal cluster expansion (Z. 2022)

$$\log\left(\mathbb{P}(X=0)\right) \overset{\text{formally}}{=} \sum_{k\geqslant 1} \sum_{(C_1,\dots,C_k)} \phi(C_1,\dots,C_k) \prod_{C\in(C_1,\dots,C_k)} (-1)^{|C|}\mu(C),$$

where the sum is over ordered multisets of polymers, whose union is connected.

- Polymers (connected subgraphs): $\{\dots,$ , ,  $,\dots\}$.
- Clusters: ...,  ,....
- Ursell function ϕ relates to chromatic polynomials.

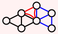
- Cluster expansion (Mayer, 1941) of partition function of a hard–core model with activity of each polymer C being $(-1)^{|C|}\mu(C)$.
- Scott and Sokal (2005) observed that
  - the Lovász local lemma (1975) condition (equivalently, Dobrushin's criterion, 1996) gives an absolute convergence criterion;

The exponential formula reduces the sum to connected subgraphs
(idea: any simple graph is the vertex-disjoint union of connected simple graphs), we have

Formal cluster expansion (Z. 2022)

$$\log\left(\mathbb{P}(X=0)\right) \overset{\text{formally}}{=} \sum_{k\geqslant 1} \sum_{(C_1,\ldots,C_k)} \phi(C_1,\ldots,C_k) \prod_{C\in(C_1,\ldots,C_k)} (-1)^{|C|}\mu(C),$$

where the sum is over ordered multisets of polymers, whose union is connected.

- Polymers (connected subgraphs): $\{\ldots,$ ,  ,  $,\ldots\}$.
- Clusters: $\ldots,$  $,\ldots.$
- Ursell function ϕ relates to chromatic polynomials.

- Cluster expansion (Mayer, 1941) of partition function of a hard–core model with activity of each polymer C being $(-1)^{|C|}\mu(C)$.
- Scott and Sokal (2005) observed that
  - the Lovász local lemma (1975) condition (equivalently, Dobrushin's criterion, 1996) gives an absolute convergence criterion;
  - the tight instance for lower bound is when every polymer is K_1 (also by Shearer, 1985).

$\mathbb{P}(\text{no triangles in } \mathscr{G}(n,p))$

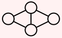
X counts triangles in $\mathscr{G}(n,p)$

- Ruciński, 1988: if $\mathbb{E}\Big[$  $\Big] = np^2 = o(1)$, then X is asymptotically Poisson.

$\mathbb{P}$(no triangles in $\mathscr{G}(n,p)$)

X counts triangles in $\mathscr{G}(n,p)$

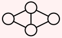
- Ruciński, 1988: if $\mathbb{E}\Big[$  $\Big] = np^2 = o(1)$, then X is asymptotically Poisson.
- Janson-Łuczak-Ruciński, 1990 (July 1987, Poznan):

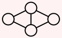
$\exp\left(-\dfrac{\mathbb{E}\Big[\right.$  $\left.\Big]}{1-p^3}\right) \overset{\text{FKG}}{\leqslant} \mathbb{P}(X=0) \overset{\text{Janson's inequality}}{\leqslant} \exp\Big(-\mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big]\Big),$

$\mathbb{P}$(no triangles in $\mathscr{G}(n,p)$)

X counts triangles in $\mathscr{G}(n,p)$

- Ruciński, 1988: if $\mathbb{E}\Big[$  $\Big] = np^2 = o(1)$, then X is asymptotically Poisson.
- Janson-Łuczak-Ruciński, 1990 (July 1987, Poznan):

$\exp\left(-\dfrac{\mathbb{E}\Big[\right.$  $\left.\Big]}{1-p^3}\right) \overset{\text{FKG}}{\leqslant} \mathbb{P}(X=0) \overset{\text{Janson's inequality}}{\leqslant} \exp\Big(-\mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big]\Big),$

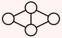
if $\mathbb{E}\Big[$  $\Big] = o(1)$, then

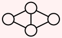
$\mathbb{P}(X=0) = \exp\Big(-\mathbb{E}\Big[$  $\Big] + o(1)\Big).$

$\mathbb{P}(\text{no triangles in } \mathscr{G}(n,p))$

X counts triangles in $\mathscr{G}(n,p)$

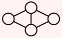
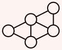
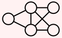
- Ruciński, 1988: if $\mathbb{E}\Big[$  $\Big] = np^2 = o(1)$, then X is asymptotically Poisson.
- Janson-Łuczak-Ruciński, 1990 (July 1987, Poznan):

$\exp\Bigg(-\dfrac{\mathbb{E}\big[$  $\big]}{1-p^3}\Bigg) \overset{\text{FKG}}{\leqslant} \mathbb{P}(X=0) \overset{\text{Janson's inequality}}{\leqslant} \exp\Big(-\mathbb{E}\big[$  $\big] + \mathbb{E}\big[$  $\big]\Big),$

if $\mathbb{E}\Big[$  $\Big] = o(1)$, then

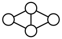
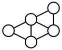
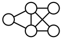
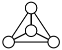
$\mathbb{P}(X=0) = \exp\Big(-\mathbb{E}\big[$  $\big] + o(1)\Big).$

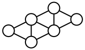
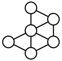
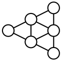
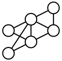
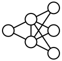
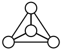
- Wormald, 1996: if $p = o\left(n^{-2/3}\right)$, then

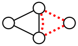
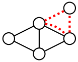
$\mathbb{P}(X=0) = \exp\Big(-\mathbb{E}\big[$  $\big] + \mathbb{E}\big[$  $\big] - \mathbb{E}\Big[$  $\Big] - \mathbb{E}\big[$  $\big] + o(1)\Big).$

$\mathbb{P}$(no triangles in $\mathscr{G}(n,p)$)

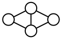
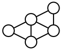
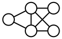
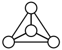
$\mathbb{P}(\text{no triangles in } \mathscr{G}(n,p))$

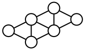
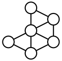
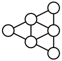
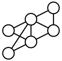
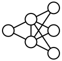
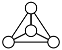
$\log \mathbb{P}(X=0) \overset{\text{formally}}{=} -\mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big] - \mathbb{E}\Big[$  $\Big] - \mathbb{E}\Big[$  $\Big] - \mathbb{E}\Big[$  $\Big]$

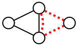
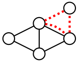
$+ \mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big] + \mathbb{E}\Big[$  $\Big]$

$- \widetilde{\mathbb{E}}\Big[$  $\Big] - \widetilde{\mathbb{E}}\Big[$  $\Big] + \widetilde{\mathbb{E}}\Big[$  $\Big] + \dots$

$\mathbb{P}$(no triangles in $\mathscr{G}(n,p)$)

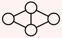
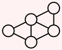
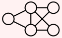
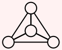
$\log \mathbb{P}(X=0) \overset{\text{formally}}{=} -\mathbb{E}[$  $] + \mathbb{E}[$  $] - \mathbb{E}[$  $] - \mathbb{E}[$  $] - \mathbb{E}[$  $]$

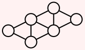
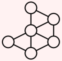
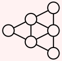
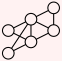
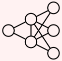
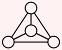
$+\mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $]$

$-\widetilde{\mathbb{E}}[$  $] - \widetilde{\mathbb{E}}[$  $] + \widetilde{\mathbb{E}}[$  $] + \dots$

Stark and Wormald, 2018; Mousset, Noever, Panagiotou, Samotij, 2020

If $p = o\left(n^{-7/11}\right)$, then

$\mathbb{P}(X=0) = \exp\Big(-\mathbb{E}[$  $] + \mathbb{E}[$  $] - \mathbb{E}[$  $] - \mathbb{E}[$  $] - \mathbb{E}[$  $]$

$+\mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + \mathbb{E}[$  $] + o(1) \Big).$

- Mousset, Noever, Panagiotou, Samotij used an alternating cumulant series, whose terms after simplifications are responding to partitions in cluster expansion.

- Mousset, Noever, Panagiotou, Samotij used an alternating cumulant series, whose terms after simplifications are responding to partitions in cluster expansion.

Joint cumulant (1929)

For any multiset of random variables $\{X_i\}_{i\in S}$,

$$\kappa(S) = \sum_{\pi\in\Pi(S)} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{P\in\pi} \mu(P).$$

Also called *semi-invariant* (1889) in probability, *Ursell function* (1927) in physics.

- Mousset, Noever, Panagiotou, Samotij used an alternating cumulant series, whose terms after simplifications are responding to partitions in cluster expansion.

Joint cumulant (1929)

For any multiset of random variables $\{X_i\}_{i\in S}$,

$$\kappa(S) = \sum_{\pi\in\Pi(S)} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{P\in\pi} \mu(P).$$

Also called *semi-invariant* (1889) in probability, *Ursell function* (1927) in physics.
For G-dependent variables and $S \subseteq V(G)$,
if S does not induce a connected subgraph in G, then $\kappa(S) = 0$.

- Isaev, McKay, Z. (2022+):
  accurate asymptotic enumeration of Eulerian orientations, digraphs, oriented graphs.

- Mousset, Noever, Panagiotou, Samotij used an alternating cumulant series, whose terms after simplifications are responding to partitions in cluster expansion.

Joint cumulant (1929)

For any multiset of random variables $\{X_i\}_{i \in S}$,

$$\kappa(S) = \sum_{\pi \in \Pi(S)} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{P \in \pi} \mu(P).$$

Also called *semi-invariant* (1889) in probability, *Ursell function* (1927) in physics.
For G-dependent variables and $S \subseteq V(G)$,
if S does not induce a connected subgraph in G, then $\kappa(S) = 0$.

- Isaev, McKay, Z. (2022+):
  accurate asymptotic enumeration of Eulerian orientations, digraphs, oriented graphs.
- Stark and Wormald (2018) also have results for $\mathscr{G}_{n,m}$:
  if $d = m\binom{n}{2}^{-1} = o\left(n^{-7/11}\right)$, then

$$\mathbb{P}(X = 0) = \exp\left(-\frac{1}{6}n^3d^3 - \frac{1}{8}n^4d^6 - \frac{1}{2}n^2d^3 + o(1)\right).$$

Asymptotic linearity of hypergraphs

A hypergraph is linear if each pair of hyperedges intersects in at most one vertex.

Asymptotic linearity of hypergraphs

A hypergraph is linear if each pair of hyperedges intersects in at most one vertex. So

$$\begin{aligned} &\mathbb{P}(\text{a random hypergraph is linear}) \\ &= \mathbb{P}(\text{no hyperedge-pairs intersect in more than one vertex}). \end{aligned}$$

Asymptotic linearity of hypergraphs

A hypergraph is linear if each pair of hyperedges intersects in at most one vertex. So

$$\begin{aligned}&\mathbb{P}(\text{a random hypergraph is linear})\\&=\mathbb{P}(\text{no hyperedge-pairs intersect in more than one vertex}).\end{aligned}$$

McKay and Tian, 2020

If $p = o\left(n^{-3/2}\right)$, then

$$\mathbb{P}\left(H_3(n,p) \text{ is linear}\right) = \exp\left(-\frac{1}{4}n^4p^2 + \frac{2}{3}n^5p^3 + o(1)\right).$$

Asymptotic linearity of hypergraphs

A hypergraph is linear if each pair of hyperedges intersects in at most one vertex. So

$$\begin{aligned}&\mathbb{P}(\text{a random hypergraph is linear})\\&=\mathbb{P}(\text{no hyperedge-pairs intersect in more than one vertex}).\end{aligned}$$

McKay and Tian, 2020

If $p = o\left(n^{-3/2}\right)$, then

$$\mathbb{P}(H_3(n,p) \text{ is linear}) = \exp\left(-\frac{1}{4}n^4p^2 + \frac{2}{3}n^5p^3 + o(1)\right).$$

Z. 2022; Wormald and Z., 2022+

If $p = o\left(n^{-7/5}\right)$, then

$$\mathbb{P}(H_3(n,p) \text{ is linear}) = \exp\left(-\frac{1}{4}n^4p^2 + \frac{2}{3}n^5p^3 - \frac{55}{24}n^6p^4 + \frac{3}{2}n^3p^2 + o(1)\right).$$

- In Z. 2022, the truncation error is by Mousset, Noever, Panagiotou, Samotij's results.
- Coefficients match McKay's conjecture based upon numerical simulation.

Asymptotic linearity of hypergraphs

McKay and Tian, 2020

If $d = \binom{n}{3}^{-1} m = o\left(n^{-3/2}\right)$, then via switching method,

$$\mathbb{P}\left(H_3(n,m) \text{ is linear}\right) = \exp\left(-\frac{1}{4}n^4 d^2 - \frac{1}{12}n^5 d^3 + o(1)\right).$$

Asymptotic linearity of hypergraphs

McKay and Tian, 2020

If $d = \binom{n}{3}^{-1} m = o\left(n^{-3/2}\right)$, then via switching method,

$$\mathbb{P}\left(H_3(n,m) \text{ is linear}\right) = \exp\left(-\frac{1}{4}n^4 d^2 - \frac{1}{12}n^5 d^3 + o(1)\right).$$

Wormald and Z. 2022+

If $d = \binom{n}{3}^{-1} m = o\left(n^{-7/5}\right)$, then via perturbation method,

$$\mathbb{P}\left(H_3(n,m) \text{ is linear}\right) = \exp\left(-\frac{1}{4}n^4 d^2 - \frac{1}{12}n^5 d^3 - \frac{1}{24}n^6 d^4 + \frac{3}{2}n^3 d^2 + o(1)\right).$$

Dependency digraphs

Dependency digraphs

A digraph $\mathbf{D} = ([d], E)$ is a dependency digraph for the events $\{A_i\}_{i \in [d]}$ if A_i is mutually independent of all the non-adjacent $\{A_j : (i,j) \notin E\}$ for every $i \in [d]$.

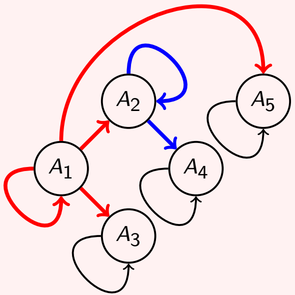
Dependency digraphs

Dependency digraphs

A digraph $\mathbf{D} = ([d], E)$ is a dependency digraph for the events $\{A_i\}_{i\in[d]}$ if A_i is mutually independent of all the non-adjacent $\{A_j : (i,j) \notin E\}$ for every $i \in [d]$.

Example

Let $\{A_i\}_{i\in[5]}$ be events with the following dependency digraph.

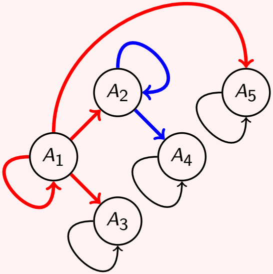


A_1 is independent of A_4.

A_2 is independent of $\{A_1, A_3, A_5\}$.

A natural refinement is

- adjacent vertices → the pairs of 'strongly dependent' events, $D_i \subseteq [d]$ denotes the 'strongly dependent neighbours' of event A_i.
- non-adjacent vertices → the pairs of 'weakly dependent' events.

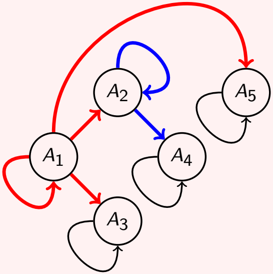


$$\mathbb{P}(A_4 \mid A_1) \approx \mathbb{P}(A_4)$$

$$\mathbb{P}(A_1 \cup A_3 \cup A_5 \mid A_2) \approx \mathbb{P}(A_1 \cup A_3 \cup A_5)$$

A natural refinement is

- adjacent vertices → the pairs of ‘strongly dependent’ events, $D_i \subseteq [d]$ denotes the ‘strongly dependent neighbours’ of event A_i.
- non-adjacent vertices → the pairs of ‘weakly dependent’ events.



$\mathbb{P}(A_4 \mid A_1) \approx \mathbb{P}(A_4)$

$\mathbb{P}(A_1 \cup A_3 \cup A_5 \mid A_2) \approx \mathbb{P}(A_1 \cup A_3 \cup A_5)$

If we change ‘$\approx$’ to ‘$\geqslant$’, then it is the ‘negative dependency graph’ by Erdős and Spencer (1991) to obtain lopsided Lovász local lemma.

New dependency digraphs

For any events $\{A_i\}_{i \in [d]}$ and digraph $\mathbf{D}$ on $[d]$:

New dependency digraphs

For any events $\{A_i\}_{i \in [d]}$ and digraph **D** on $[d]$:

- φ-mixing coefficient is used to measure 'weak dependencies':

$$\varphi = \max_{i \in [d]} \left| \mathbb{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j \mid A_i \right) - \mathbb{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j \right) \right|.$$

New dependency digraphs

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ on $[d]$:

- φ-mixing coefficient is used to measure 'weak dependencies':

$$\varphi = \max_{i\in[d]} \left| \mathbb{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \mid A_i \right) - \mathbb{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \right) \right|.$$

- The influence of 'strongly dependent' events is

$$\Delta_1 = \sum_{i\in[d]} \mathbb{P}\left(A_i \cap \bigcup_{j\in[i-1]\cap D_i} A_j \right) \prod_{k\in[d]\setminus[i]} \mathbb{P}\left(\overline{A_k}\right)$$

$$\Delta_2 = \sum_{i\in[d]} \mathbb{P}(A_i)\,\mathbb{P}\left(\bigcup_{j\in[i-1]\cap D_i} A_j \right) \prod_{k\in[d]\setminus[i]} \mathbb{P}\left(\overline{A_k}\right).$$

New dependency digraphs

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ on $[d]$:

- φ-mixing coefficient is used to measure 'weak dependencies'

$$\varphi = \max_{i\in[d]} \left| \mathbb{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \mid A_i \right) - \mathbb{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \right) \right|.$$

- The influence of 'strongly dependent' events is

$$\Delta_1 \leqslant \Delta_1' = \sum_{i\in[d]} \sum_{j\in[i-1]\cap D_i} \mathbb{P}(A_i \cap A_j),$$

$$\Delta_2 \leqslant \Delta_2' = \sum_{i\in[d]} \sum_{j\in[i-1]\cap D_i} \mathbb{P}(A_i)\mathbb{P}(A_j).$$

New dependency digraphs

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ on $[d]$:

- φ-mixing coefficient is used to measure 'weak dependencies'

$$\varphi = \max_{i\in[d]} \left| \mathbb{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \mid A_i \right) - \mathbb{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \right) \right|.$$

- The influence of 'strongly dependent' events is

$$\Delta_1 \leqslant \Delta_1' = \sum_{i\in[d]} \sum_{j\in[i-1]\cap D_i} \mathbb{P}(A_i \cap A_j),$$
$$\Delta_2 \leqslant \Delta_2' = \sum_{i\in[d]} \sum_{j\in[i-1]\cap D_i} \mathbb{P}(A_i)\mathbb{P}(A_j).$$

- Chen–Stein Poisson approximation involves similar terms as φ, Δ_1', and Δ_2'. Chen (1975) also considered the φ-mixing condition.
- Suen's inequality (1990) also involves Δ_1' and Δ_2'.

Theorem (Isaev, Rodionov, Z., Zhukovskii; 2020, 2022+)

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ with vertex set $[d]$, we have

$$\left|\mathbb{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right) - \prod_{i\in[d]} \mathbb{P}\left(\overline{A_i}\right)\right| \leqslant \left(1 - \prod_{i\in[d]} \mathbb{P}\left(\overline{A_i}\right)\right)\varphi + \max(\Delta_1, \Delta_2).$$

- The lower bound generalises Dubickas' bound (2008), and the upper bound is via a similar inductive proof.

Theorem (Isaev, Rodionov, Z., Zhukovskii; 2020, 2022+)

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ with vertex set $[d]$, we have

$$\left|\mathbb{P}\left(\bigcap_{i\in[d]}\overline{A_i}\right)-\prod_{i\in[d]}\mathbb{P}\left(\overline{A_i}\right)\right|\leqslant\left(1-\prod_{i\in[d]}\mathbb{P}\left(\overline{A_i}\right)\right)\varphi+\max(\Delta_1,\Delta_2).$$

- The lower bound generalises Dubickas' bound (2008), and the upper bound is via a similar inductive proof.
- If $\varphi=o(1)$, $\Delta_1=o(1)$, and $\Delta_2=o(1)$, then we have the asymptotic independence.

Theorem (Isaev, Rodionov, Z., Zhukovskii; 2020, 2022+)

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ with vertex set $[d]$, we have

$$\left|\mathbb{P}\left(\bigcap_{i\in[d]}\overline{A_i}\right)-\prod_{i\in[d]}\mathbb{P}\left(\overline{A_i}\right)\right|\leqslant\left(1-\prod_{i\in[d]}\mathbb{P}\left(\overline{A_i}\right)\right)\varphi+\max(\Delta_1,\Delta_2).$$

- The lower bound generalises Dubickas' bound (2008), and the upper bound is via a similar inductive proof.
- If $\varphi=o(1)$, $\Delta_1=o(1)$, and $\Delta_2=o(1)$, then we have the asymptotic independence.
  - '*Two Moments Suffice for Poisson Approximations: The Chen–Stein Method*' by Arratia, Goldstein and Gordon (1989); also Chen (1975).

Theorem (Isaev, Rodionov, Z., Zhukovskii; 2020, 2022+)

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ with vertex set $[d]$, we have

$$\left|\mathbb{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right) - \prod_{i\in[d]} \mathbb{P}\left(\overline{A_i}\right)\right| \leqslant \left(1 - \prod_{i\in[d]} \mathbb{P}\left(\overline{A_i}\right)\right)\varphi + \max(\Delta_1, \Delta_2).$$

- The lower bound generalises Dubickas' bound (2008), and the upper bound is via a similar inductive proof.
- If $\varphi = o(1)$, $\Delta_1 = o(1)$, and $\Delta_2 = o(1)$, then we have the asymptotic independence.
  - '*Two Moments Suffice for Poisson Approximations: The Chen–Stein Method*' by Arratia, Goldstein and Gordon (1989); also Chen (1975).
- Max codegrees, max clique-extension counts of random (hyper)graphs, etc. are asymptotically Gumbel.

Theorem (Isaev, Rodionov, Z., Zhukovskii; 2020, 2022+)

For any events $\{A_i\}_{i\in[d]}$ and digraph $\mathbf{D}$ with vertex set $[d]$, we have

$$\left|\mathbb{P}\left(\bigcap_{i\in[d]}\overline{A_i}\right)-\prod_{i\in[d]}\mathbb{P}\left(\overline{A_i}\right)\right|\leqslant\left(1-\prod_{i\in[d]}\mathbb{P}\left(\overline{A_i}\right)\right)\varphi+\max(\Delta_1,\Delta_2).$$

- The lower bound generalises Dubickas' bound (2008), and the upper bound is via a similar inductive proof.
- If $\varphi=o(1)$, $\Delta_1=o(1)$, and $\Delta_2=o(1)$, then we have the asymptotic independence.
  - '*Two Moments Suffice for Poisson Approximations: The Chen–Stein Method*' by Arratia, Goldstein and Gordon (1989); also Chen (1975).
- Max codegrees, max clique-extension counts of random (hyper)graphs, etc. are asymptotically Gumbel.
- Lu and Székely (2009) also introduced a notion of 'positive' dependency graph to obtain upper bounds for Poisson approximation for random matchings.

Open problem

- To establish the absolute convergence of cluster expansion.

Open problem

- To establish the absolute convergence of cluster expansion.
- To design a suitable model for $\mathscr{G}(n, m)$;
  e.g., Féray's weighted dependency graphs (2018) for normality of subgraph counts.

Thanks!