Multivariate Bernoulli distributions

Ising Model and negative dependence with actuarial applications

E Marceau with H Cossette, B Côté, A Mutti, P Semeraro

EM, HC: Université Laval (Québec, Canada)
BC: University of Waterloo (Waterloo, Canada)
AM, PS: Politecnico di Torino (Torino, Italia)

60th Actuarial Research Conference
July 29 - August 1, 2025 Toronto, Canada
Thursday 31 July 2025, 9:00-10:00am

Introduction and motivations

Jacob Bernoulli (1654-1705) introduced the univariate Bernoulli distribution.

Key features of the Bernoulli distribution:

- Binary nature: success/failure, occurrence or not, etc.
- Advantages: simplicity, interpretability, flexibility, adaptability.

Multivariate Bernoulli distributions are important in various contexts:

- Actuarial Science, Machine Learning, QRM, Bio-informatics, Natural Language Processing, Etc.

Examples of applications:

- Actuarial science: fire occurrences for a portfolio of property insurance contracts.
- Epidemiology: disease occurrences for a specific population.
- Meteorology: daily rainfall occurrences on several sites.
- Machine learning: binary classification for fraud detection.
- Credit risk modelling: defaults for a credit portfolio.

Introduction and motivations

Individual Risk Model in Actuarial Science :

- Consider an insurance portfolio of d risks $\boldsymbol{X} = (X_1, \dots, X_d)$, where

$$X_v = \begin{cases} B_v, & I_v = 1 \\ 0, & I_v = 0 \end{cases}, \qquad v \in \mathcal{V} = \{1, \dots, d\}. \tag{1}$$

- $\boldsymbol{I} = (I_1, \dots, I_d)$: vector of d occurrence rvs, where $I_v \sim Bern(q_v)$, for $v \in \mathcal{V}$.
- $\boldsymbol{I}$ follows a multivariate Bernoulli distribution, with joint cdf $F_{\boldsymbol{I}}$.
- B_v = loss amount rv of risk v, if $I_v = 1$, for $v \in \mathcal{V}$.
- $\boldsymbol{B} = (B_1, \dots, B_d)$: vector of d independent positive loss amount rvs.
- $\boldsymbol{I}$ and $\boldsymbol{B}$ are independent.
- Fire Insurance for property $v \in \mathcal{V}$:

$$X_v = \begin{cases} B_v, & I_v = 1 & \text{fire occurs at or spreads to property } v \\ 0, & I_v = 0 & \text{otherwise} \end{cases}.$$

- Number of occurrences rv for a subset $A \subseteq \mathcal{V}$ of risks: $N_A = \sum_{v \in A} I_v$, where $N_\emptyset = 0$ and $N_\mathcal{V} = N$.
- Aggregate loss amount rv for a subset $A \subseteq \mathcal{V}$ of risks: $S_A = \sum_{v \in A} X_v$, where $S_\emptyset = 0$ and $S_\mathcal{V} = S$.
- Today, we will focus on the distribution of N_A.

Agenda

The choice of $F_{\boldsymbol{I}}$ has many implications in tasks performed using IRM:

- Distributions of N_A and S_A for any subset $A \in \mathcal{V}$.
- Risk measures, Premium rating and claim reserving.
- Contributions under risk sharing and risk allocation.
- Decision making.

Two main objectives of the talk:

- Study multivariate Bernoulli distributions associated with the tree-structured Ising model whose dependence structure is encrypted on a tree.
- Investigate extreme negative dependence and the strongly Rayleigh property within the family of multivariate Bernoulli distributions.

Agenda:

- Part 1: Fréchet class of multivariate Bernoulli distributions with fixed marginals.
- Part 2: Tree-structured Ising models with fixed marginals [Côté et al., 2025].
- Part 3: Extreme negative dependence and the strongly Rayleigh property within the family of multivariate Bernoulli distributions [Cossette et al., 2025].

Part 1: Fréchet class

Let $\mathcal{B}_d$ be the class of all d-dimensional multivariate Bernoulli distributions.

- We define $\mathcal{B}_d(\boldsymbol{q}) \subset \mathcal{B}_d$ as the Fréchet class of all d-dimensional multivariate Bernoulli distributions with fixed parameters $\boldsymbol{q} = (q_1, \dots, q_d)$, $q_v \in (0,1)$, $v \in \mathcal{V}$.
- Notation: $q_\bullet = \sum_{v \in \mathcal{V}} q_v$.
- Fréchet bounds:
  - If $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$, then
  $$W(i_1, \dots, i_d) \le F_{\boldsymbol{I}}(i_1, \dots, i_d) \le M(i_1, \dots, i_d), \quad \text{for all } (i_1, \dots, i_d) \in \{0,1\}^d.$$
  - The function W is the lower Fréchet bound:
  $$W(i_1, \dots, i_d) = \max(F_1(i_1) + \cdots + F_d(i_d) - d + 1, 0), \quad (i_1, \dots, i_d) \in \{0,1\}^d.$$
  - The function M is the upper Fréchet bound:
  $$M(i_1, \dots, i_d) = \min(F_1(i_1), \dots, F_d(i_d)), \quad (i_1, \dots, i_d) \in \{0,1\}^d.$$
  - $F_v(0) = 1 - q_v = \overline{q}_v$ and $F_v(1) = 1$, $v \in \mathcal{V}$.
- For $d \ge 2$, M is the cdf of a vector of comonotonic rvs.
- For $d = 2$, W is the cdf of a vector of countermonotonic rvs.
- For $d > 2$ and $0 < q_\bullet \le 1$, W is the cdf of a vector of mutually exclusive rvs.
- For $d > 2$ and $1 < q_\bullet < d - 1$, W is not a cdf.

Part 1: Fréchet class

Let $\boldsymbol{I} = (I_v, v \in \mathcal{V})$ be a d-dimensional vector of Bernoulli rvs with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$.

- Joint pmf: $p_{\boldsymbol{I}}(\boldsymbol{i}) = \Pr(I_1 = i_1, \dots, I_d = i_d), \quad \boldsymbol{i} = (i_1, \dots, i_d) \in \{0,1\}^d$.

- Joint pgf $\mathcal{P}_{\boldsymbol{I}}$ is a multi-affine polynomial with positive real coefficients:

$$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = \mathrm{E}\left[t_1^{I_1} \cdots t_d^{I_d}\right] = \sum_{\boldsymbol{i} \in \{0,1\}^d} p_{\boldsymbol{I}}(\boldsymbol{i}) t_1^{i_1} \dots t_d^{i_d}, \quad \boldsymbol{t} \in [-1,1]^d.$$

- A multi-affine polynomial $\mathcal{P}$ is a polynomial in which each variable has a degree at most one.
- Helpful to investigate the dependence properties of $\boldsymbol{I}$.
- Powerful tool for aggregation: using Theorem 1 of [Wang, 1998], we find the pgf of $\mathcal{P}_N$ from the multivariate pgf of $\boldsymbol{I}$:

$$\mathcal{P}_N(t) = \mathcal{P}_{\boldsymbol{I}}(t_1, \dots, t_d)\big|_{t_1 = \dots = t_d = t}, \quad t \in [-1,1]. \tag{2}$$

- Efficient evaluation of risk allocations based on generating function: see [Blier-Wong et al., 2025].

Applications of (2):

1. Identify the distribution of N from $\mathcal{P}_N$ obtained with (2), when it is possible.
2. Compute the values of pmf f_N from the pgf $\mathcal{P}_N$ with FFT algorithm.
3. FFT and pgf: see Ch. 30 of [Cormen et al., 2022]

Part 1: Fréchet class

Independence and joint pgf:

- Let $\boldsymbol{I}$ be a vector of independent rvs with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$.
- Joint pgf of $\boldsymbol{I}$ = multiaffine polynomial with 2^d non-zero terms:

$$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = (1 - q_1 + q_1 t_1) \times \cdots \times (1 - q_d + q_d t_d), \quad \boldsymbol{t} \in [-1,1]^d.$$

Comonotonicity and joint pgf:

- Let $\boldsymbol{I}$ be a vector of comonotonic rvs with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ such that $0 < q_1 < \cdots < q_d < 1$.
- Joint pgf of $\boldsymbol{I}$ = multiaffine polynomial with $d+1$ non-zero terms:

$$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = 1 - q_d + (q_d - q_{d-1}) t_d + (q_{d-1} - q_{d-2}) t_{d-1} t_d + \cdots + (q_2 - q_1) t_2 \ldots t_d + q_1 t_1 \ldots t_d, \quad \boldsymbol{t} \in [-1,1]^d.$$

Mutual exclusivity and joint pgf:

- Let $\boldsymbol{I}$ be a vector of mutual exclusive rvs [1] with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$, $0 < q_1 < \cdots < q_d < 1$ and $\sum_{v \in \mathcal{V}} q_v \leq 1$.
- Joint pgf of $\boldsymbol{I}$ = multiaffine polynomial with $d+1$ (or d, if $\sum_{v \in \mathcal{V}} q_v = 1$) non-zero terms:

$$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = 1 - q_\bullet + q_1 t_1 + q_2 t_2 + \cdots + q_d t_d, \quad \boldsymbol{t} \in [-1,1]^d.$$

[1] See [Dhaene and Denuit, 1999] and [Puccetti et al., 2015] about mutual exclusivity.

Part 1: Fréchet class

Example inspired by *Wall Drawing 413* of the artist Sol LeWitt:

- Let $F_{\boldsymbol{I}^{(m)}} \in \mathcal{B}_9(\boldsymbol{q})$ where $q_\bullet = 3$, for $m = 1, 2, 3$.

v	1	2	3	4	5	6	7	8	9
q_v	0.22	0.40	0.16	0.45	0.50	0.33	0.20	0.38	0.36

- We present 50 samples of $\boldsymbol{I}^{(m)}$, for $m = 1, 2, 3$.
- We used the *acton* sequential gradients scientific color maps[2] from [Crameri, 2018] to visually represent the magnitude of the mean parameter q for each component of the vector.
- Each grid corresponds to $\boldsymbol{I} = (I_1, \ldots, I_9)$, where a 1 is represented by a colored square and a 0 is represented by a white square.

$$\Rightarrow \begin{bmatrix} I_1 & I_2 & I_3 \\ I_4 & I_5 & I_6 \\ I_7 & I_8 & I_9 \end{bmatrix}$$

[2] Concerning scientific color usage, see [Crameri et al., 2020] and [Bujack et al., 2017].

Part 1: Fréchet class

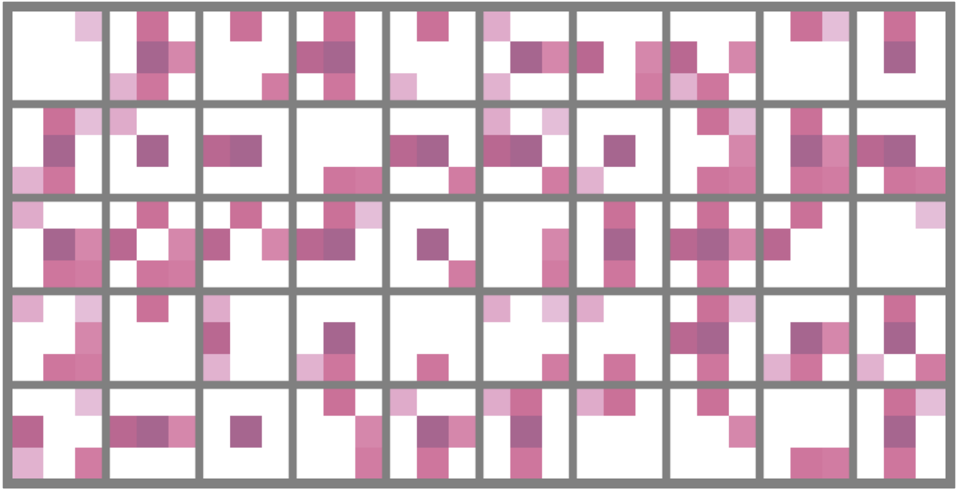


Figure: 50 samples of $\boldsymbol{I}^{(1)} \sim$ MB Distribution 1. Q&A: $F_{\boldsymbol{I}^{(1)}} = ?$ (a) Comon; (b) Indep; (c) Neither.

Part 1: Fréchet class

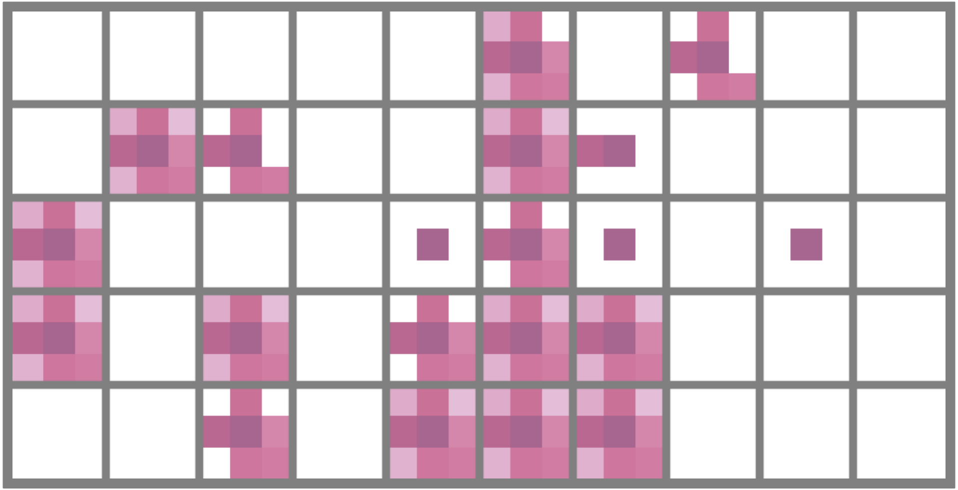


Figure: 50 samples of $\boldsymbol{I}^{(2)} \sim$ MB Distribution 2. Q&A: $F_{\boldsymbol{I}^{(2)}} = ?$ (a) Comon; (b) Indep; (c) Neither.

Part 1: Fréchet class

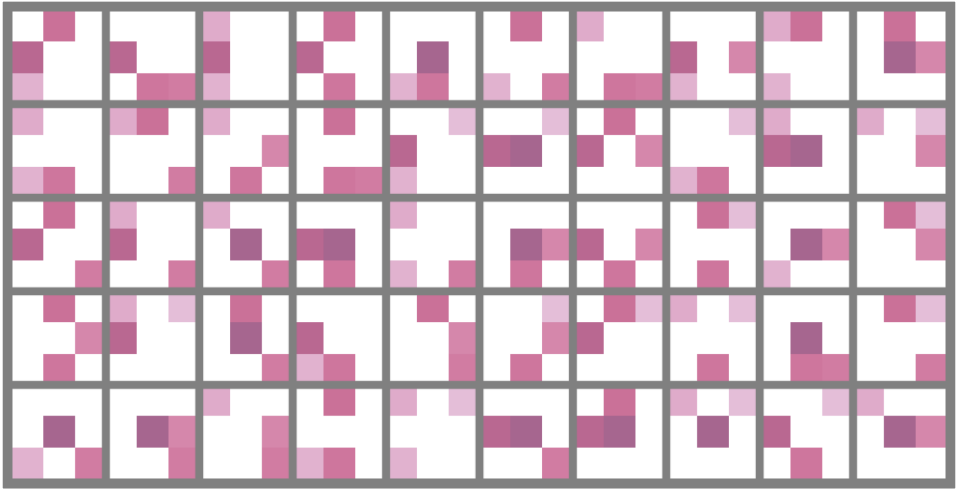


Figure: 50 samples of $\boldsymbol{I}^{(3)} \sim$ MB Distribution 3. Q&A: $F_{\boldsymbol{I}^{(3)}} = ?$ (a) Comon; (b) Indep; (c) Neither.

Part 1: Fréchet class

The inspiration . . .

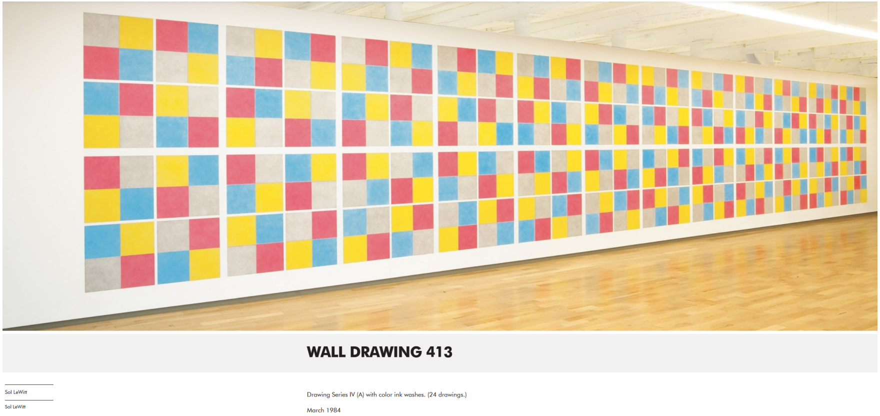


Figure: Sol LeWitt, Wall Drawing 413, MASS MoCA

Part 1: Fréchet class

Let $\boldsymbol{I} = (I_1, I_2)$ with $F_{\boldsymbol{I}} \in \mathcal{B}_2(\boldsymbol{q})$ with $\boldsymbol{q} = (q_1, q_2)$:

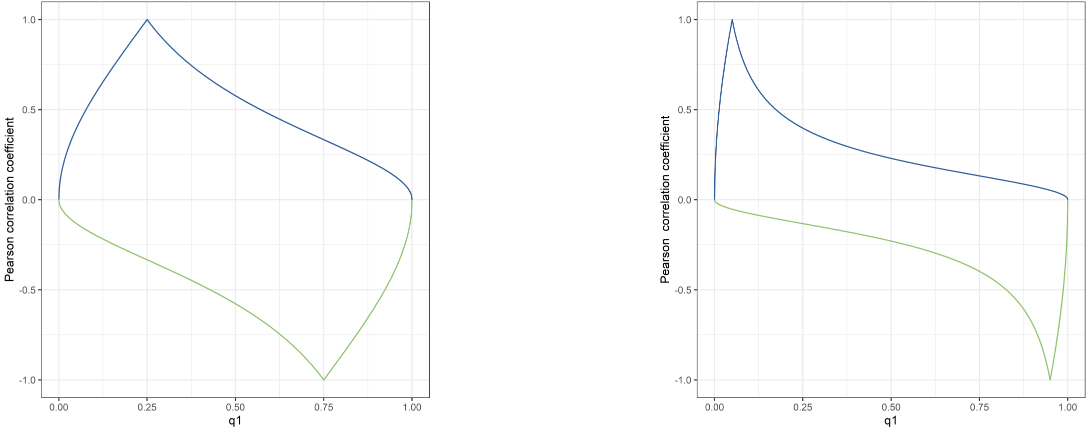
- Pearson correlation coefficient: $\rho_P(I_1, I_2) = \alpha_{(1,2)}$ with $\alpha_{(1,2)} \in A(q_1, q_2)$, where
$$A(q_1,q_2) = \left[-\min\left(\sqrt{\frac{q_1 q_2}{\overline{q}_1\overline{q}_2}}, \sqrt{\frac{\overline{q}_1\overline{q}_2}{q_1 q_2}}\right), \min\left(\sqrt{\frac{\overline{q}_1 q_2}{q_1\overline{q}_2}}, \sqrt{\frac{q_1\overline{q}_2}{\overline{q}_1 q_2}}\right)\right] \subseteq [-1,1].$$
- If $\boldsymbol{I}$ is countermonotonic, then $\alpha_{(1,2)} = -\min\left(\sqrt{\frac{q_1 q_2}{\overline{q}_1\overline{q}_2}}, \sqrt{\frac{\overline{q}_1\overline{q}_2}{q_1 q_2}}\right)$.
- If $\boldsymbol{I}$ is comonotonic, then $\alpha_{(1,2)} = \min\left(\sqrt{\frac{\overline{q}_1 q_2}{q_1\overline{q}_2}}, \sqrt{\frac{q_1\overline{q}_2}{\overline{q}_1 q_2}}\right)$.
- Values of joint pmf $p_{I_1,I_2}(i_1,i_2)$:

i_1, i_2	0	1
0	$\overline{q}_1\overline{q}_2 + \alpha_{(1,2)}\sqrt{q_1\overline{q}_1 q_2\overline{q}_2}$	$\overline{q}_1 q_2 - \alpha_{(1,2)}\sqrt{q_1\overline{q}_1 q_2\overline{q}_2}$
1	$q_1\overline{q}_2 - \alpha_{(1,2)}\sqrt{q_1\overline{q}_1 q_2\overline{q}_2}$	$q_1 q_2 + \alpha_{(1,2)}\sqrt{q_1\overline{q}_1 q_2\overline{q}_2}$

(3)

- Assuming $d = 2$ and mixed marginals $\Rightarrow$ the Pearson correlation coefficient $\alpha_{(1,2)}$ defines $F_{\boldsymbol{I}}$.
- Joint pgf :
$$\mathcal{P}_{I_1,I_2}(t_1,t_2) = p_{I_1,I_2}(0,0) + p_{I_1,I_2}(1,0)t_1 + p_{I_1,I_2}(0,1)t_2 + p_{I_1,I_2}(1,1)t_1 t_2, \qquad (t_1,t_2) \in [-1,1]^2.$$

Part 1: Fréchet class



(a) $q_2 = 0.25$ (b) $q_2 = 0.05$

Figure: Bounds on the Pearson correlation coefficient

Part 1: Fréchet class

Results in [Fontana and Semeraro, 2018] about $\mathcal{B}_d(\boldsymbol{q})$:

- $\mathcal{B}_d(\boldsymbol{q}) =$ convex polytope with a finite number $n_d(\boldsymbol{q})$ of extremal points $r^{[1]}, \ldots, r^{[n_d(\boldsymbol{q})]}$.
- For any $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ with pmf $p_{\boldsymbol{I}}$, there exists

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{n_d(\boldsymbol{q})}) \in [0,1]^n$$

with $\sum_{j=1}^{n_d(\boldsymbol{q})} \eta_j = 1$ such that

$$p_{\boldsymbol{I}}(\boldsymbol{i}) = \sum_{j=1}^{n_d(\boldsymbol{q})} \eta_j r^{[j]}(\boldsymbol{i}), \quad \boldsymbol{i} \in \{0,1\}^d.$$

Many approaches have been proposed to study MB distributions, namely

1. Multivariate Bernoulli distributions defined with common mixtures.
2. Multivariate Bernoulli distributions defined with multivariate copulas.

Today:

- We investigate multivariate Bernoulli distributions within the framework of graphical models.
- We examine negative dependence within the family of multivariate Bernoulli distributions.

Part 1: Fréchet class

Example: Assume $\boldsymbol{q} = (\frac{1}{4}, \frac{1}{7}, \frac{1}{3})$. $\mathcal{B}_3(\boldsymbol{q})$ = convex polytope with $n_d(\boldsymbol{q}) = 11$ extremal points $r^{[1]}, \ldots, r^{[11]}$.

- $r^{[1]}$: pmf of a random vector $\boldsymbol{I}$ where $\boldsymbol{I}$ = vector of mutually exclusive rvs with joint cdf

$$F_{\boldsymbol{I}}(\boldsymbol{i}) = \max(F_1(i_1) + F_2(i_2) + F_3(i_3) - 2; 0), \quad \boldsymbol{i} \in \{0,1\}^3$$

- $r^{[10]}$: pmf of a random vector $\boldsymbol{I}$ where $\boldsymbol{I}$ = vector of comonotonic rvs with joint cdf

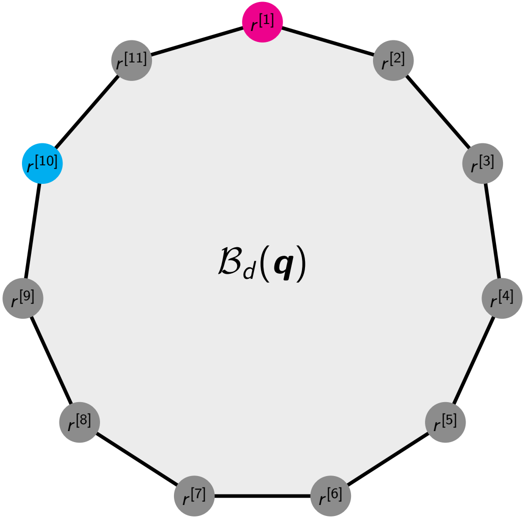
$$F_{\boldsymbol{I}}(\boldsymbol{i}) = \min(F_1(i_1); F_2(i_2); F_3(i_3)), \quad \boldsymbol{i} \in \{0,1\}^3$$

- The pmf of the vector of independent rvs is **not** an extremal point of $\mathcal{B}_3(\boldsymbol{q})$.

$\boldsymbol{i}$	$r^{[1]}(\boldsymbol{i})$	$r^{[2]}(\boldsymbol{i})$	$r^{[3]}(\boldsymbol{i})$	$r^{[4]}(\boldsymbol{i})$	$r^{[5]}(\boldsymbol{i})$	$r^{[6]}(\boldsymbol{i})$	$r^{[7]}(\boldsymbol{i})$	$r^{[8]}(\boldsymbol{i})$	$r^{[9]}(\boldsymbol{i})$	$r^{[10]}(\boldsymbol{i})$	$r^{[11]}(\boldsymbol{i})$
(0,0,0)	0.274	0.417	0.417	0.524	0.524	0.56	0.607	0.607	0.667	0.667	0.637
(1,0,0)	0.25	0.107	0.25	0	0	0.107	0	0.06	0	0	0
(0,1,0)	0.143	0	0	0	0.143	0	0.06	0	0	0	0
(1,1,0)	0	0.143	0	0.143	0	0	0	0	0	0	0.03
(0,0,1)	0.333	0.333	0.19	0.226	0.083	0.19	0	0	0	0.083	0
(1,0,1)	0	0	0	0.107	0.25	0	0.25	0.19	0.19	0.107	0.22
(0,1,1)	0	0	0.143	0	0	0	0.083	0.143	0.083	0	0.113
(1,1,1)	0	0	0	0	0	0.143	0	0	0.06	0.143	0

Source of the values of the pmf $r^{[1]}, \ldots, r^{[11]}$: § 5.3.1 in [Fontana and Semeraro, 2018]

Part 1: Fréchet class



Part 2: Tree-structured Ising model with fixed marginals

The tree-structured Ising model is a specific undirected graphical model defined on a tree that is used to describe the stochastic behavior of a Bernoulli random vector $\boldsymbol{I}$.

A tree $\mathcal{T}$ is an undirected graph $(\mathcal{V},\mathcal{E})$, connected and acyclic:

- $\mathcal{V} = \{1,2,\ldots,d\}$ = set of vertices
- $\mathcal{E}$ = set of $d-1$ edges
- $\mathrm{path}(u,v)$: path from vertex u to vertex v
  - $\mathrm{path}(u,v)$: sequence of successive edges $e \in \mathcal{E}$
  - the first edge starts at vertex u
  - the last edge ends at vertex v
  - the same edge cannot appear more than once in $\mathrm{path}(u,v)$
- In $\mathcal{T}$: there is no path from a vertex to itself.
- We only consider trees.

$\mathcal{T}_r$: rooted version of a tree $\mathcal{T}$:

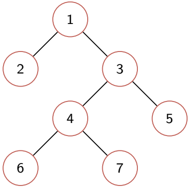
- $r \in \mathcal{V}$ = one specific vertex labelled as the *root*
- For an undirected tree $\mathcal{T}$, one can choose any vertex $r \in \mathcal{V}$ to be the root

Part 2: Tree-structured Ising model with fixed marginals

Working with rooted version $\mathcal{T}_r$ of a tree $\mathcal{T}$ is helpful:

- descendants of vertex $v = \mathrm{dsc}(v)$, $v \in \mathcal{V}$:
  - set of vertices whose path to the root r goes through v
- children of vertex $v = \mathrm{ch}(v)$, $v \in \mathcal{V}$:
  - set of descendants of v that are also connected to v by an edge
- leaf $=$ vertex that has no children
- parent of vertex $v = \mathrm{pa}(v)$, $v \in \mathcal{V} \setminus \{r\}$:
  - the sole vertex connected by an edge to v that is not the children of v
- The root r has no parent

Illustration with $d = 7$ vertices and root $= 1$:



$$\mathrm{dsc}(3) = \{4{,}5{,}6{,}7\}$$
$$\mathrm{ch}(3) = \{4{,}5\} \qquad \mathrm{ch}(2) = \emptyset$$
$$\mathrm{pa}(3) = 1 \qquad \mathrm{pa}(4) = 3$$
$$\mathrm{path}(2{,}5) = \{(1{,}2), (1{,}3), (3{,}5)\}$$
$$\text{leaves} = \{2{,}5{,}6{,}7\}$$

Part 2: Graphical Models

Graphical models are useful at capturing complex dependence relations among rvs.

They are tailored to build high-dimensional multivariate probabilistic models.

Applications: Bioinformatics, image processing, machine learning, etc.

Two large families of graphical models:

- Bayesian Networks, also called directed graphical models.
- Markov Random Fields, also called undirected graphical models.

Markov random fields (MRFs) are used to model high-dimensional vector of rvs:

1. Each node in the graph corresponds to a random variable.
2. Each edge between two adjacent nodes (rvs) capture direct interactions, which make the model interpretable.

Two classical families of MRFs:

- Gaussian graphical model.
- Ising model.

Part 2: Ising model - Natural parametrization

The Ising model is usually defined through the natural parametrization.

Let $\boldsymbol{I} = (I_v, v \in \mathcal{V})$ be an Ising model defined on a graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$.

Under the natural parametrization, the pmf of $\boldsymbol{I}$ is given by

$$p_{\boldsymbol{I}}(\boldsymbol{i}) = \frac{1}{Z}\exp\left(\sum_{v\in\mathcal{V}} \eta_v i_v + \sum_{(v,w)\in\mathcal{E}} \eta_{\{v,w\}} i_v i_w\right), \quad \boldsymbol{i} \in \{0,1\}^d, \tag{4}$$

where

- $\eta_v, v \in \mathcal{V}$ and $\eta_{v,w}, (v,w) \in \mathcal{E}$ are parameters;
- Z is a normalizing constant.

Part 2: Ising model - Natural parametrization

Definition 1 (Markov random field)

A vector of random variables $\boldsymbol{I} = \{I_v, v \in \mathcal{V}\}$ is Markov random field encrypted on the graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$, if it satisfies the **local Markov property**:

$$I_u \perp\!\!\!\perp I_v | (I_j : (u,j) \in \mathcal{E}) \tag{5}$$

for every pair of vertices u,v such that $(u,v) \notin \mathcal{E}$, where $\perp\!\!\!\perp$ marks conditional independence.
We say a MRF is tree-structured if it is encrypted on a tree $\mathcal{T}$.

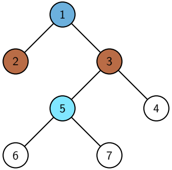


Figure: $I_1 \perp\!\!\!\perp I_5 | (I_2, I_3)$

Part 2: Ising model - Natural parametrization

Hammersley-Clifford theorem: $\boldsymbol{I}$, for which the pmf is non-zero, is a Markov random field encrypted on graph $\mathcal{G}$ if and only if its distribution is Gibbs with respect to $\mathcal{G}$.

Definition 2 (Gibbs distribution)

A distribution is Gibbs with respect to graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ if it is characterized by a pmf of the form

$$p(\boldsymbol{i}) = \frac{1}{Z} \prod_{W \in \mathcal{C}} \varphi_W((i_v : v \in W)), \quad \boldsymbol{i} \in \{0,1\}^d, \tag{6}$$

for some functions $\{\varphi_W, W \in \mathcal{C}\}$, where $\mathcal{C}$ is the set of cliques of the graph and Z is a normalizing constant.

Cliques are sets of vertices such that every vertex is a neighbor to every other.

Part 2: Tree-structured Ising model - Natural parametrization

The following theorem establish the link between tree-structured Ising models and tree structured Markov random fields with Bernoulli marginal distributions.

Theorem 3 (Link)

Under the assumption that joint pmf are non-zero for every configuration on $\{0,1\}^d$, all tree-structured MRFs with Bernoulli marginal distributions are tree-structured Ising models.

Let $\boldsymbol{I}$ be a tree-structured Ising model defined through the natural parametrization:

- $\boldsymbol{I}$ follows a MB distribution: $F_{\boldsymbol{I}} \in \mathcal{B}_d$
- Natural parameterization $\Rightarrow$ major drawbacks for dependence modeling.
- The marginal distributions are not fixed: modification of $\eta_{(u,v)}$ results in a change of q_u and q_v.
- Computation of $p_{\boldsymbol{I}}$ is arduous: we need to sum over $\{0,1\}^d$ find the normalizing constant.
- The computation of the values of the pmf of $\boldsymbol{I}_A = \{I_v, v \in A\}$ for any $A \subset \mathcal{V}$ is also difficult.
- Sampling is not simple.

Part 2: Tree-structured Ising model - Mean parametrization

To circumvent these major drawbacks, the authors of [Wainwright et al., 2008] suggest to define the Ising model under the mean parametrization.

- Assume that $\boldsymbol{I}$ is a tree-structured Ising model defined on the tree $\mathcal{T} = (\mathcal{V},\mathcal{E})$.
- With Bernoulli parameters $\boldsymbol{q}$ and parameters $\boldsymbol{\alpha} = (\alpha_e, e \in \mathcal{E})$. Fix a root $r \in \mathcal{V}$.
- The pmf of $\boldsymbol{I}$ is

$$p_{\boldsymbol{I}}(\boldsymbol{i}) = p_{I_1}(i_1) \cdots p_{I_d}(i_d) \prod_{v \in \mathrm{dsc}(r)} \frac{p_{I_v, I_{\mathrm{pa}(v)}}(i_v, i_{\mathrm{pa}(v)})}{p_{I_v}(i_v) p_{I_{\mathrm{pa}(v)}}(i_{\mathrm{pa}(v)})}, \quad \boldsymbol{i} \in \{0,1\}^d. \tag{7}$$

- $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ = Fréchet class of all d-variate Bernoulli distributions with parameters $\boldsymbol{q} = (q_1, \ldots, q_d)$.
- Notation: $\boldsymbol{I} = (I_v, v \in \mathcal{V})$ follows a Multivariate Bernoulli Distribution, denoted $\boldsymbol{I} \sim MB_d(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$.
- Dependence parameters: $\boldsymbol{\alpha} = (\alpha_{(v,w)}, (v,w) \in \mathcal{E})$, where $\alpha_{(v,w)} \in A(q_v,q_w)$ for each $(v,w) \in \mathcal{E}$.
- $\alpha_{(v,w)} = \rho_P(I_v,I_w)$ for two adjacent rvs I_v and I_w linked by the edge $(v,w) \in \mathcal{E}$.

Part 2: Tree-structured Ising model - Mean parametrization

Assume that $\boldsymbol{I} \sim MB_d(\boldsymbol{q}, \boldsymbol{\alpha}, \mathcal{T})$.

Advantages and original results:

- Only $2d - 1$ parameters $(\boldsymbol{q}, \boldsymbol{\alpha})$ and the configuration of $\mathcal{T}$ specify the multivariate distribution of $\boldsymbol{I} \sim MB$.
- Derive a stochastic representation for $\boldsymbol{I}$.
- Find the Pearson's correlation coefficient for any pair of $\boldsymbol{I}$.
- Design an efficient sampling algorithm
- Derive the multivariate pgf of $\boldsymbol{I}$.

Part 2: Tree-structured Ising model - Mean parametrization

Example: Assume that $\boldsymbol{q} = (1/4, 1/7, 1/3)$ with $\rho_P(I_1, I_2) = 0.1$ and $\rho_P(I_1, I_3) = 0.2$.

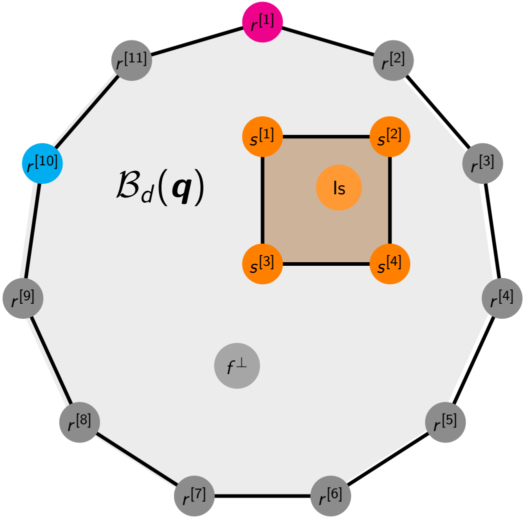
- $\mathcal{B}_3(\boldsymbol{q},\alpha_{(1,2)},\alpha_{(1,3)}) \subset \mathcal{B}_3(\boldsymbol{q})$
- $\mathcal{B}_3(\boldsymbol{q},\alpha_{(1,2)},\alpha_{(1,3)})$ = convex polytope with 4 extremal points $s^{[1]}, s^{[2]}, s^{[3]}, s^{[4]}$
- Let $\boldsymbol{I} \sim MB_d(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$ with $\mathcal{V} = \{1,2,3\}$, $\mathcal{E} = \{(1,2),(1,3)\}$, and $\alpha_{(1,2)} = \rho_P(I_1, I_2)$, $\alpha_{(1,3)} = \rho_P(I_1, I_3)$
- Fixing root $r = 2$, the expression of $p_{\boldsymbol{I}}$ is

$$p_{\boldsymbol{I}}(\boldsymbol{i}) = p_{I_2}(i_2)p_{I_1|I_2=i_2}(i_1)p_{I_3|I_1=i_1}(i_3) = \frac{p_{I_1,I_2}(i_1,i_2)p_{I_1,I_3}(i_1,i_3)}{p_{I_1}(i_1)}, \quad \boldsymbol{i} \in \{0,1\}^3.$$

$\boldsymbol{i}$	$s^{[1]}(\boldsymbol{i})$	$s^{[2]}(\boldsymbol{i})$	$s^{[3]}(\boldsymbol{i})$	$s^{[4]}(\boldsymbol{i})$	$p_{\boldsymbol{I}}(\boldsymbol{i})$
(0,0,0)	0.4488	0.4488	0.5408	0.5408	0.4745
(1,0,0)	0.075	0.1258	0.075	0.1258	0.1002
(0,1,0)	0.092	0.092	0	0	0.0663
(1,1,0)	0.0509	0	0.0509	0	0.0256
(0,0,1)	0.2092	0.2092	0.1172	0.1172	0.1835
(1,0,1)	0.1242	0.0733	0.1242	0.0733	0.0989
(0,1,1)	0	0	0.092	0.092	0.0257
(1,1,1)	0	0.0509	0	0.0509	0.0253

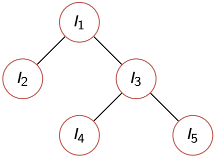
Table: Values of the pmfs $s^{[1]}, \ldots, s^{[4]}$ calculated using [4ti2 team,], and $p_{\boldsymbol{I}}$

Part 2: Tree-structured Ising model - Mean parametrization



Part 2: Tree-structured Ising model - Mean parametrization

Example: $\boldsymbol{I} \sim MB_5(\boldsymbol{q},\alpha,\mathcal{T})$ with $\mathcal{T} = \{\mathcal{V},\mathcal{E}\}$ where $\mathcal{V} = \{1,2,3,4,5\}$ and $\mathcal{E} = \{(1,2),(1,3),(3,4),(3,5)\}$.



$$\boldsymbol{A}_{\mathcal{T}_1} = \begin{bmatrix} 1 & \alpha_{(1,2)} & \alpha_{(1,3)} & 0 & 0 \\ \alpha_{(1,2)} & 1 & 0 & 0 & 0 \\ \alpha_{(1,3)} & 0 & 1 & \alpha_{(3,4)} & \alpha_{(3,5)} \\ 0 & 0 & \alpha_{(3,4)} & 1 & 0 \\ 0 & 0 & \alpha_{(3,5)} & 0 & 1 \end{bmatrix};$$

Fixing root $r = 1$, the expression of $p_{\boldsymbol{I}}$ is

$$p_{\boldsymbol{I}}(\boldsymbol{i}) = p_{I_1}(i_1) \times \frac{p_{I_1,I_2}(i_1,i_2)}{p_{I_1}(i_1)} \times \frac{p_{I_1,I_3}(i_1,i_3)}{p_{I_1}(i_1)} \times \frac{p_{I_3,I_4}(i_3,i_4)}{p_{I_3}(i_3)} \times \frac{p_{I_3,I_5}(i_3,i_5)}{p_{I_3}(i_3)}, \quad \boldsymbol{i} \in \{0,1\}^5. \tag{8}$$

Remarks:

- Only $2d - 1 = 9$ parameters specifiy the distribution of $\boldsymbol{I}$: $q_1, \ldots, q_5$ and $\alpha_{(1,2)}, \alpha_{(1,3)}, \alpha_{(3,4)}, \alpha_{(3,5)}$.
- Edges $=$ direct interactions $\Rightarrow$ interpretable model.

Part 2: Stochastic representation

Stochastic representations of vector of rvs are important:

- Understand how dependence arises among the components of a vector of rvs.
- Design efficient algorithms for simulating samples of a vector of rvs.
- Derive expectation of integrable functions of the components of the vector of rvs.

We need the Binomial thinning operator (introduced in [Steutel et al., 1983]):

- Let K be a nonnegative integer-valued rv.
- The Binomial thinning operator $\circ$ is

$$\alpha \circ K := \begin{cases} \sum_{l=1}^{K} J_l^{(\alpha)}, & K > 0 \\ 0, & K = 0 \end{cases},$$

  where $\{J_l^{(\alpha)},\ l \in \mathbb{N}\}$ = sequence of iid Bernoulli rvs of parameter $\alpha \in (0,1)$.
- Interpretation: $\alpha \circ K = \alpha$-*fraction* of K.
- Applications :
  - Dynamics of models for time series of count data:[Weiß, 2008], [Scotto et al., 2015]
  - Study self-decomposability of discrete rvs: [Steutel and van Harn, 1979]

Part 2: Stochastic representation

Theorem 4 (Stochastic representation)

Assume that $\boldsymbol{I} \sim MB_d(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$ *and let*

$$\sigma_{(v,w)} = \alpha_{(v,w)}\sqrt{q_v q_w (1-q_v)(1-q_w)}, \quad (v,w) \in \mathcal{E},$$

and fix a root $r \in \mathcal{V}$.

Then $\boldsymbol{I}$ *admits the following stochastic representation:*

$$I_r \sim Bern(q_r)$$

and

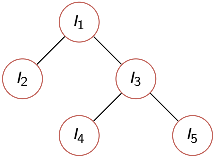
$$I_v = \left(q_v + \frac{\sigma_{(\mathrm{pa}(v),v)}}{q_{\mathrm{pa}(v)}}\right) \circ I_{\mathrm{pa}(v)} + \left(q_v - \frac{\sigma_{(\mathrm{pa}(v),v)}}{(1-q_{\mathrm{pa}(v)})}\right) \circ (1 - I_{\mathrm{pa}(v)}),$$

for $v \in \mathrm{r}$.

Convention: $\alpha \circ 0 = 0$.

Part 2: Stochastic representation

Example: Assume that $\boldsymbol{I} \sim MB_5(\boldsymbol{q},\alpha,\mathcal{T})$ with $\mathcal{T} = \{\mathcal{V},\mathcal{E}\}$ illustrated as follows:



$$\boldsymbol{A}_{\mathcal{T}_1} = \begin{bmatrix} 1 & \alpha_{(1,2)} & \alpha_{(1,3)} & 0 & 0 \\ \alpha_{(1,2)} & 1 & 0 & 0 & 0 \\ \alpha_{(1,3)} & 0 & 1 & \alpha_{(3,4)} & \alpha_{(3,5)} \\ 0 & 0 & \alpha_{(3,4)} & 1 & 0 \\ 0 & 0 & \alpha_{(3,5)} & 0 & 1 \end{bmatrix};$$

Stochastic representation of $\boldsymbol{I}$ assuming that $r = 1 \Rightarrow$ rooted tree $= \mathcal{T}_1$

$$I_1 \Rightarrow \begin{cases} I_2 = \left(q_2 + \frac{\sigma_{(1,2)}}{q_1}\right) \circ I_1 + \left(q_2 - \frac{\sigma_{(1,2)}}{(1-q_1)}\right) \circ (1 - I_1) \\ I_3 = \left(q_3 + \frac{\sigma_{(1,3)}}{q_1}\right) \circ I_1 + \left(q_3 - \frac{\sigma_{(1,3)}}{(1-q_1)}\right) \circ (1 - I_1) \Rightarrow \begin{cases} I_4 = \left(q_4 + \frac{\sigma_{(3,4)}}{q_3}\right) \circ I_3 + \left(q_4 - \frac{\sigma_{(3,4)}}{(1-q_3)}\right) \circ (1 - I_3) \\ I_5 = \left(q_5 + \frac{\sigma_{(3,5)}}{q_3}\right) \circ I_3 + \left(q_5 - \frac{\sigma_{(3,5)}}{(1-q_3)}\right) \circ (1 - I_3) \end{cases} \end{cases}$$

where $\alpha \circ 0 = 0$.

Part 2: Pearson's correlation coefficient

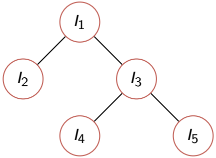
Theorem 5 (Pearson's correlation coefficient)

If $\boldsymbol{I} \sim MB_d(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$, the Pearson correlation coefficient of (I_v,I_w) is given by

$$\rho_P(I_v, I_w) = \prod_{e \in \mathrm{path}(v,w)} \alpha_e, \quad v \neq w \in \mathcal{V}.$$

$\Rightarrow$ *Exponential decreasing: Value of $\rho_P(I_v, I_w) \downarrow$ as path's length $|\mathrm{path}(v,w)| \uparrow$*

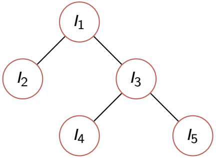
Example: $\boldsymbol{I} \sim MB_5(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$ with $\mathcal{T} = \{\mathcal{V},\mathcal{E}\}$ illustrated as follows:



	I_1	I_2	I_3	I_4	I_5
I_1	1	$\alpha_{(1,2)}$	$\alpha_{(1,3)}$	$\alpha_{(1,3)}\alpha_{(3,4)}$	$\alpha_{(1,3)}\alpha_{(3,5)}$
I_2		1	$\alpha_{(1,2)}\alpha_{(1,3)}$	$\alpha_{(1,2)}\alpha_{(1,3)}\alpha_{(3,4)}$	$\alpha_{(1,2)}\alpha_{(1,3)}\alpha_{(3,5)}$
I_3			1	$\alpha_{(3,4)}$	$\alpha_{(3,5)}$
I_4				1	$\alpha_{(3,4)}\alpha_{(3,5)}$
I_5					1

Part 2: Sampling Algorithm

Illustration of Sampling Algorithm: $\boldsymbol{I} \sim MB_5(\boldsymbol{q},\alpha,\mathcal{T})$ with $\mathcal{T} = \{\mathcal{V},\mathcal{E}\}$ illustrated as follows:



$$\boldsymbol{A}_{\mathcal{T}_1} = \begin{bmatrix} 1 & \alpha_{(1,2)} & \alpha_{(1,3)} & 0 & 0 \\ \alpha_{(1,2)} & 1 & 0 & 0 & 0 \\ \alpha_{(1,3)} & 0 & 1 & \alpha_{(3,4)} & \alpha_{(3,5)} \\ 0 & 0 & \alpha_{(3,4)} & 1 & 0 \\ 0 & 0 & \alpha_{(3,5)} & 0 & 1 \end{bmatrix};$$

Detailed steps of the Sampling Algorithm assuming root $r = 1$:

1. Simulate $I_1 \sim Bern(q_1)$.
2. If $I_1 = 0$, simulate $I_2 \sim Bern(q_2 - \frac{\sigma_{(1,2)}}{1-q_1})$, else $I_2 \sim Bern(q_2 + \frac{\sigma_{(1,2)}}{q_1})$.
3. If $I_1 = 0$, simulate $I_3 \sim Bern(q_3 - \frac{\sigma_{(1,3)}}{1-q_1})$, else $I_3 \sim Bern(q_3 + \frac{\sigma_{(1,3)}}{q_1})$.
4. If $I_3 = 0$, simulate $I_4 \sim Bern(q_4 - \frac{\sigma_{(3,4)}}{1-q_3})$, else $I_4 \sim Bern(q_4 + \frac{\sigma_{(3,4)}}{q_3})$.
5. If $I_3 = 0$, simulate $I_5 \sim Bern(q_5 - \frac{\sigma_{(3,5)}}{1-q_3})$, else $I_4 \sim Bern(q_5 + \frac{\sigma_{(3,5)}}{q_3})$.

Advantages of Sampling Algorithm based on the stochastic representation:

1. It is easy to implement.
2. It performs well for large values of d.

Part 2: Multivariate pgf

Theorem 6 (Multivariate pgf)

Assume that $\boldsymbol{I} \sim MB_d(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$, *let* $\sigma_{(u,v)} = \alpha_{(u,v)}\sqrt{q_u q_v(1-q_u)(1-q_v)}$, *for* $(u,v) \in \mathcal{E}$, *and fix a root* $r \in \mathcal{V}$.

Then the multivariate pgf of $\boldsymbol{I}$ *is given by*

$$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = (1-q_r)\prod_{i\in\mathrm{ch}(r)} \zeta_i^{\mathcal{T}_r}(\boldsymbol{t}_{i\mathrm{dsc}(i)}) + q_r t_r \prod_{i\in\mathrm{ch}(r)} \xi_i^{\mathcal{T}_r}(\boldsymbol{t}_{i\mathrm{dsc}(i)}), \quad \boldsymbol{t}\in[-1,1]^d, \tag{9}$$

with $\zeta_v^{\mathcal{T}_r}(\boldsymbol{t}_{v\mathrm{dsc}(v)})$ *and* $\xi_v^{\mathcal{T}_r}(\boldsymbol{t}_{v\mathrm{dsc}(v)})$ *recursively defined for every* $v \in \mathcal{V}\backslash\{r\}$*:*

$$\zeta_v^{\mathcal{T}_r}\left(\boldsymbol{t}_{v\mathrm{dsc}(v)}\right) = \left(1-q_v+\frac{\sigma_{(\mathrm{pa}(v),v)}}{1-q_{\mathrm{pa}(v)}}\right)\prod_{i\in\mathrm{ch}(v)}\zeta_i^{\mathcal{T}_r}\left(\boldsymbol{t}_{i\mathrm{dsc}(i)}\right) + \left(q_v-\frac{\sigma_{(\mathrm{pa}(v),v)}}{1-q_{\mathrm{pa}(v)}}\right)t_v\prod_{i\in\mathrm{ch}(v)}\xi_i^{\mathcal{T}_r}\left(\boldsymbol{t}_{i\mathrm{dsc}(i)}\right); \tag{10}$$

$$\xi_v^{\mathcal{T}_r}\left(\boldsymbol{t}_{v\mathrm{dsc}(v)}\right) = \left(1-q_v-\frac{\sigma_{(\mathrm{pa}(v),v)}}{q_{\mathrm{pa}(v)}}\right)\prod_{i\in\mathrm{ch}(v)}\zeta_i^{\mathcal{T}_r}\left(\boldsymbol{t}_{i\mathrm{dsc}(i)}\right) + \left(q_v+\frac{\sigma_{(\mathrm{pa}(v),v)}}{q_{\mathrm{pa}(v)}}\right)t_v\prod_{i\in\mathrm{ch}(v)}\xi_i^{\mathcal{T}_r}\left(\boldsymbol{t}_{i\mathrm{dsc}(i)}\right), \tag{11}$$

where products taken over an empty set are equal to 1 *by convention.*

Part 2: Multivariate pgf

Applications of multivariate pgfs (generally) and Theorem 6 (specifically):

- Identify distribution of functions of $\boldsymbol{I}$ (when possible).
- Theorem 6 allows one to handle the multivariate pgf of $\boldsymbol{I}$ recursively, using (9), (10), and (11).
- It is convenient to implement (9), (10), and (11) (with FFT algorithm) for computational purposes.
- One may choose any root r.

Details on multivariate pgfs:

- § 34.2.1 of [Johnson et al., 1997];
- Chapter 3 of [Flajolet and Sedgewick, 2009];
- Appendix A of [Axelrod and Kimmel, 2015].

Part 2: Multivariate pgf

Herbert Wilf, Generatingfunctionology (1994):

A generating function is a clothesline on which we hang up a sequence of numbers for display.

Figure: Morning Hockey League (MHL) Hockey Jerseys (Source: LMH)

Part 2: Chow-Liu Algorithm

The Chow-Liu Algorithm allows us to find the maximum likelihood tree underlying the tree-structured Ising model, [Chow and Liu, 1968]

It has two steps:

1. Step 1: Creating a matrix with the information between pairs of rvs.
2. Step 2: Applying a maximum spanning tree algorithm such as Kruskal's algorithm or Prim's algorithm [3].

Step 1 has three substeps:

1. Step 1.1: Determine the Pearson correlation coefficient matrix.
2. Step 1.2: Compute the bivariate pmfs using (3).
3. Step 1.3: Use (12) to compute the information matrix.

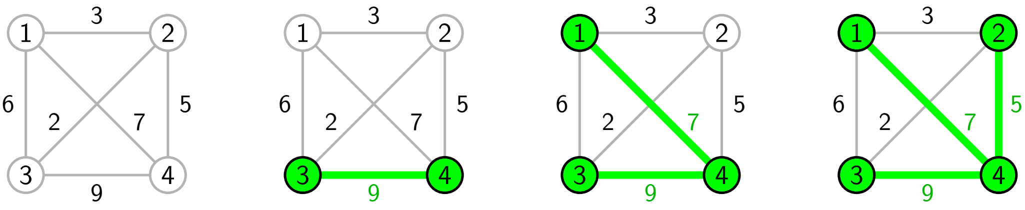
$$\mathcal{I}(I_1, I_2) = \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} \Pr(I_1 = i_1, I_2 = i_2) \ln \left(\frac{\Pr(I_1 = i_1, I_2 = i_2)}{\Pr(I_1 = i_1) \Pr(I_2 = i_2)} \right). \tag{12}$$

[3] See, e.g., [Sedgewick and Wayne, 2011] for details

Part 2: Chow-Liu Algorithm

Illustration of Prim's algorithm:

- Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}_1, \boldsymbol{w})$ be a weighted graph and $\mathcal{T} = (\mathcal{V}, \mathcal{E}_2)$ be the maximum spanning tree.
- Then, we iteratively add the maximum weight edge in $\mathcal{E}_1$ which connects a new vertex in $\mathcal{T}$ to $\mathcal{E}_2$ until all vertices in $\mathcal{V}$ are connected in $\mathcal{T}$.
- Each weight is a multiple of $\frac{1}{20}$.



(a) Initial state (b) Iteration 1 (c) Iteration 2 (d) Iteration 3

We illustrate the procedure in the next example.

Part 2: Numerical Example using Precipitation Data

Precipitation (rainfall and snowfall) data from 19 weather stations in Canada 1970-2025.

- Source = [Goverment of Canada, 2025]: 12662 observations after data cleaning.
- We apply the Chow-Liu algorithm to build a tree-based Ising model.
- Training data = 8500 randomly samble observations; test data = 4162 remaining observations.

City Number v	City	q_v	City Number v	City	q_v
1	EDM	0.3700	11	WHI	0.3892
2	VIC	0.4526	12	SEP	0.4855
3	WIN	0.3582	13	ROU	0.5446
4	FRE	0.4327	14	THU	0.4216
5	STJ	0.5849	15	OTT	0.5282
6	HAL	0.4578	16	CAL	0.3199
7	TOR	0.4026	17	KEL	0.3879
8	CHA	0.4906	18	PRI	0.5316
9	QUE	0.5607	19	CHU	0.5346
10	REG	0.3458			

Table: Estimated values of q_v, $v \in \mathcal{V} = \{1, \ldots, 19\}$ [4]

[4] See the Appendix of the matrix of estimated Pearson correlation coefficient.

Part 2: Numerical Example using Precipitation Data

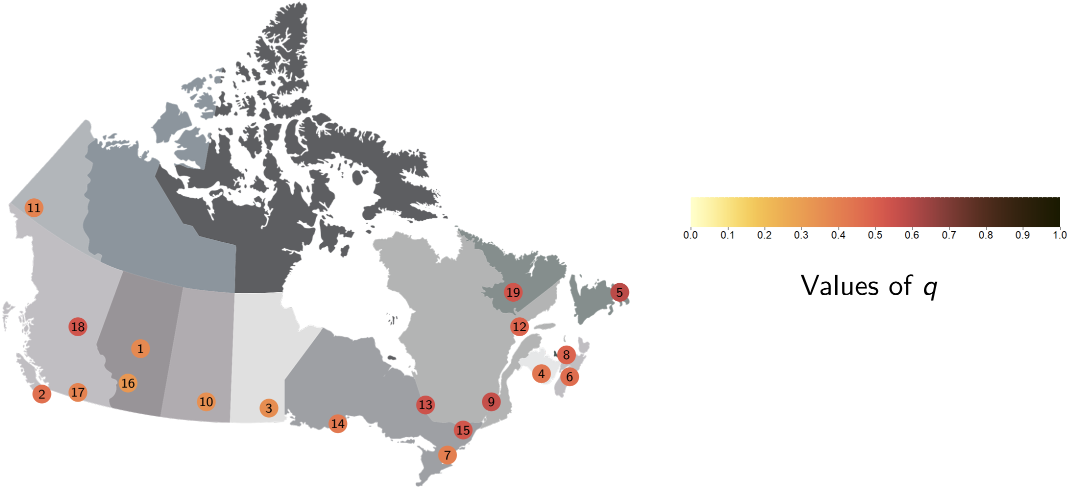


Figure: Estimated value of q_v at each station v, $v \in \mathcal{V}$, on a map of Canada [Pixabay, 2025].

Part 2: Numerical Example using Precipitation Data

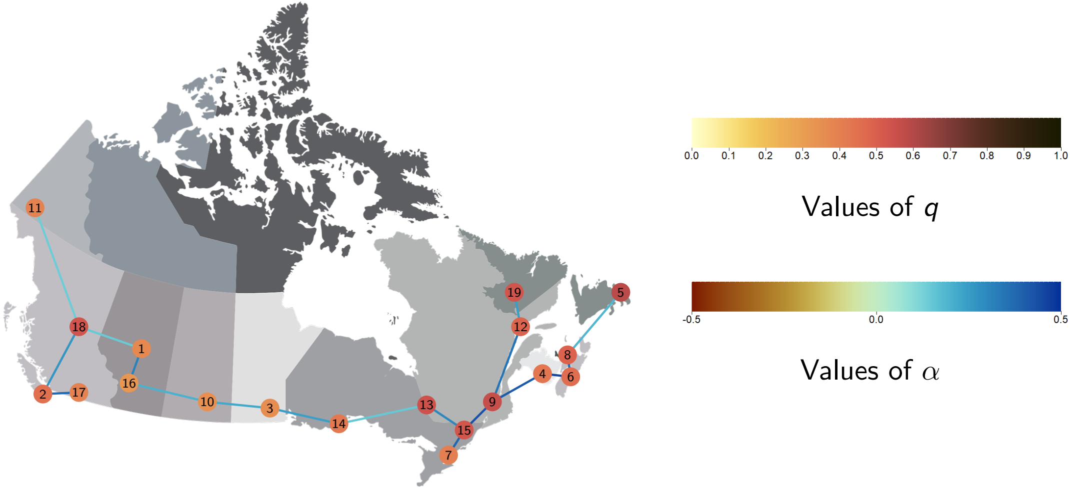


Figure: Tree-structured Ising model: the tree found using the Chow-Liu algorithm

Part 2: Numerical Example using Precipitation Data

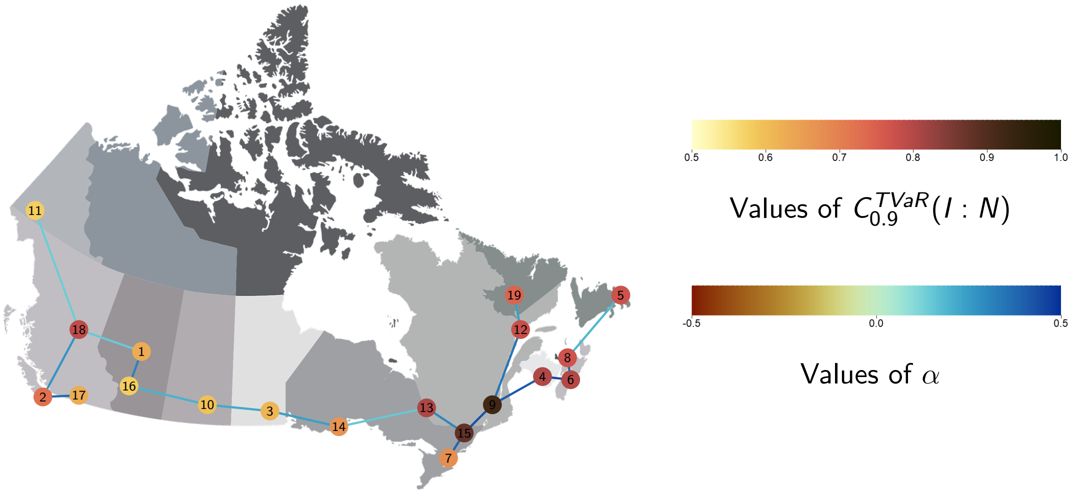


Figure: Tree-structured Ising model: contributions of each stations (centrality)

Part 2: Numerical Example with 4 trees

Impact of the shape of the tree:

- We consider four tree structures with $d = 31$ vertices $\mathcal{V} = (1, \ldots, 31)$
- Structure A: $\mathcal{T}^A = (\mathcal{V}, \mathcal{E}^A) =$ 31-vertex-sized star
- Structure B: $\mathcal{T}^B =$ 5-nary tree of radius 2
- Structure C: $\mathcal{T}^C =$ binary (2-nary) tree of radius 4
- Structure D: $\mathcal{T}^D =$ 31-vertex-sized series tree

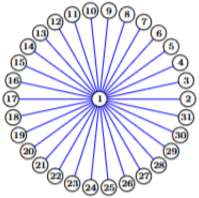
Assumptions:

- $q_v^{\mathcal{S}} = q = \frac{10}{31}$, $v \in \mathcal{V}$, structure $\mathcal{S} \in \{A, B, C, D\}$.
- $\alpha_e^{\mathcal{S}} = \alpha = \frac{1}{2}$, $e \in \mathcal{E}$, structure $\mathcal{S} \in \{A, B, C, D\}$.

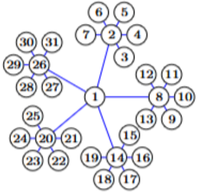
We compute values of two common risk measures in actuarial science and QRM:

- Value-at-Risk: $\mathrm{VaR}_\kappa(N) = F_N^{-1}(\kappa) = \inf\{x \in \mathbb{R}, F_N(x) \geq \kappa\}$, for $\kappa \in (0,1)$.
- Tail-VaR: $\mathrm{TVaR}_\kappa(N) = \frac{1}{1-\kappa} \int_\kappa^1 F_N^{-1}(u) \mathrm{d}u$, for $\kappa \in (0,1)$.

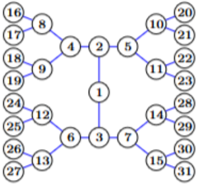
Part 2: Numerical Example with 4 trees



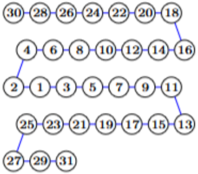
Structure A



Structure B



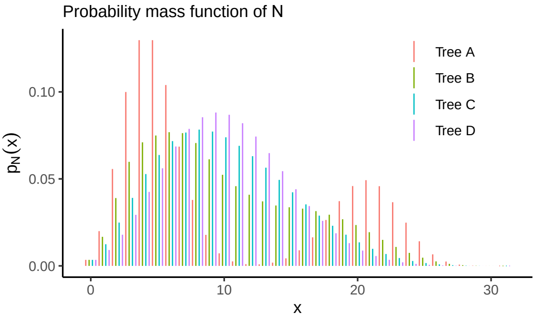
Structure C

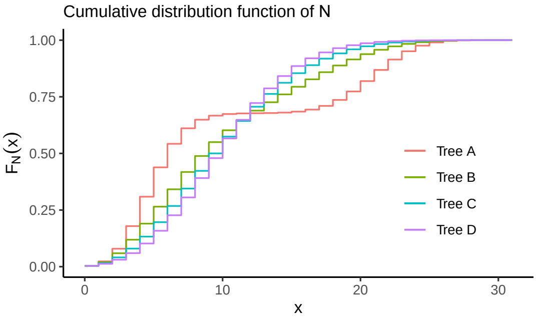


Structure D

	Structure A	Structure B	Structure C	Structure D
$E[N]$	10	10	10	10
$\mathrm{Var}(N)$	60.86	34.91	24.48	19.45

Part 2: Numerical Example with 4 trees





	Structure A	Structure B	Structure C	Structure D
$\mathrm{VaR}_{0.9}(N)$	22	19	17	16
$\mathrm{TVaR}_{0.9}(N)$	23.73	21.48	19.47	18.21
$\mathrm{VaR}_{0.99}(N)$	26	24	23	21
$\mathrm{TVaR}_{0.99}(N)$	26.45	25.56	24.14	22.64

Part 2: Numerical Example with 4 trees

$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_v)$ = contribution of I_v to $\mathrm{TVaR}_{\kappa}(N)$ under Euler's rule, for $v \in \mathbf{V}$

Structure A:

κ	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_1)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_2)$	$TVaR_{\kappa}(N)$
0.9	1	0.758	23.73

Structure B:

κ	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_1)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_2)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_4)$	$TVaR_{\kappa}(N)$
0.9	0.988	0.862	0.647	21.48

Structure C:

κ	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_1)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_2)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_4)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_8)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_{16})$	$TVaR_{\kappa}(N)$
0.9	0.774	0.814	0.758	0.660	0.547	19.47

Structure D:

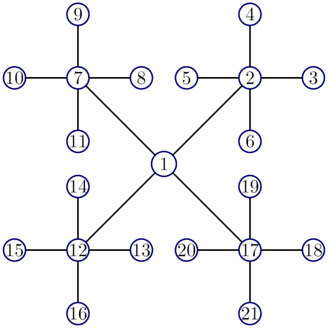
κ	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_1)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_2)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_4)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_8)$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_{16})$	$\mathcal{C}_{\kappa}^{\mathrm{TVaR}}(I_{30})$	$TVaR_{\kappa}(N)$
0.9	0.60237	0.60235	0.60232	0.60218	0.60072	0.49826	18.21

Other applications: conditional mean risk sharing, in the vein of [Denuit and Robert, 2022].

Part 2: Example of marginalization

Let $\boldsymbol{I} \sim MB_d(\boldsymbol{q},\boldsymbol{\alpha},\mathcal{T})$, with $d = 21$ and $\mathcal{T} = (\mathcal{V},\mathcal{E})$ = 4-nary tree of radius 2:

- $q_v^B = q = \frac{5}{21}$, $\forall\ v \in \mathcal{V}$, $\quad \alpha_e^B = \alpha = \frac{1}{3}$, $\forall\ e \in \mathcal{E}$.
- $\boldsymbol{I}$ can take $2^{21} = 2\,097\,152$ values!
- Challenge:
  - Find an efficient approach to compute all values of the pmf of a subset of $\boldsymbol{I}$ ...
  - ...without computing all the $2^{21} = 2\,097\,152$ values of the pmf $p_{\boldsymbol{I}}$ and then marginalizing.



Part 2: Example of marginalization

We compute the 16 values of the pmf of $(I_3, I_7, I_{12}, I_{15})$ using pgf of $\boldsymbol{I}$ and FFT algorithm:

	i_3	i_7	i_{12}	i_{15}	$p_{I_3,I_7,I_{12},I_{15}}(i_3,i_7,i_{12},i_{15})$
1	0	0	0	0	0.392626
2	0	0	0	1	0.074080
3	0	0	1	0	0.061211
4	0	0	1	1	0.059298
5	0	1	0	0	0.101381
6	0	1	0	1	0.019128
7	0	1	1	0	0.027518
8	0	1	1	1	0.026658
9	1	0	0	0	0.112686
10	1	0	0	1	0.021261
11	1	0	1	0	0.020692
12	1	0	1	1	0.020046
13	1	1	0	0	0.034272
14	1	1	0	1	0.006466
15	1	1	1	0	0.011514
16	1	1	1	1	0.011154

Computation time lower than 1 second on Philippe's computer.

Part 3: Extreme negative dependence and the SR property

The theory of negative dependence aims to have a "better understanding of what it means for a collection of random variables to be *repelling* or *mutually negatively dependent*" [Pemantle, 2000].

We consider two streams of research about the theory of negative dependence:

- Stream 1: The authors of [Puccetti et al., 2015] investigate different notions of extreme negative dependence introduced in the literature to generalize countermonotonicity when $d > 2$.
- Stream 2: Answering questions in *Towards a theory of negative dependence* [Pemantle, 2000], the authors of [Borcea et al., 2009] introduce the Strongly Rayleigh (SR) property, which is stronger than the Negatively Associated (NA) property.

Our objective is to compare and combine both streams of research on the *theory of (extreme) negative dependence*.

Part 3: Extreme negative dependence

We consider the following notion of extremal negative dependence.

Definition 7 (Σ-countermonotonic)

A Bernoulli random vector $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ is Σ-countermonotonic if for every subset $\Lambda \subseteq \{1, \ldots, d\}$, the random variables $\sum_{j \in \Lambda} I_j$ and $\sum_{j \notin \Lambda} I_j$ are countermonotonic.

Th 3.7 [Puccetti et al., 2015]: A Σ-countermonotonic element always exists in $\mathcal{B}_d(\boldsymbol{q})$.

Notation: Set of Σ-countermonotonic elements $= \mathcal{B}_d^{\Sigma}(\boldsymbol{q}) \subset \mathcal{B}_d(\boldsymbol{q})$.

We prove that $\mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ is a convex polytope with a finite number of extremal points.

If $q_{\bullet} = m \in \{1, \ldots, m-1\}$, a Bernoulli random vector $\boldsymbol{I}$, that satisfies the property of Σ-countermonotonicity, also satisfies the property of joint mixability:

$$\mathbb{P}\left(I_1 + \cdots + I_d = m\right) = 1.$$

For $m \in \{0, \ldots, d\}$, let $\mathcal{A}_{d,m} \subset \{0,1\}^d$ be the set of d-dimensional binary vectors with sum equal to m:

$$\mathcal{A}_{d,m} = \{(i_1, \ldots, i_d) \in \{0,1\}^d : i_1 + \cdots + i_d = m\}.$$

Part 3: Extreme negative dependence

Assume that $\boldsymbol{I}$ satisfies the property of Σ-countermonotonicity:

- If $q_\bullet \in (m, m+1)$ and $m \in \{1, \ldots, d-2\}$, the support of $\boldsymbol{I}$ is $\mathcal{A}_{d,m} \cup \mathcal{A}_{d,m+1}$.
  - Joint pgf:
  $$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = \sum_{\boldsymbol{i} \in \mathcal{A}_{d,m} \cup \mathcal{A}_{d,m+1}} p_{\boldsymbol{I}}(\boldsymbol{i}) t_1^{i_1} \ldots t_d^{i_d}$$
  - Pmf of N:
  $$f_N(k) = \begin{cases} m+1-q_\bullet, & k = m \\ q_\bullet - m, & k = m+1 \\ 0, & \text{otherwise} \end{cases}.$$
- If $q_\bullet = m$ and $m \in \{1, \ldots, d-1\}$, the support of $\boldsymbol{I}$ is $\in \mathcal{A}_{d,m}$.
  - Joint pgf:
  $$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = \sum_{\boldsymbol{i} \in \mathcal{A}_{d,m}} p_{\boldsymbol{I}}(\boldsymbol{i}) t_1^{i_1} \ldots t_d^{i_d}$$
  - Pmf of N:
  $$f_N(k) = \begin{cases} 1, & k = m \\ 0, & \text{otherwise} \end{cases}.$$
  - Variance of N: $Var(N) = 0$.

Part 3: Extreme negative dependence

Choose a Bernoulli random vector $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$:

- It implies that

$$Var(I_1 + \cdots + I_d) \leq Var(I'_1 + \cdots + I'_d),$$

  for all $\boldsymbol{I}'$ with $F_{\boldsymbol{I}'} \in \mathcal{B}_d(\boldsymbol{q})$.
- Q1: Does it imply that

$$\rho_P(I_v, I_w) \leq 0,$$

  for all pairs in $\boldsymbol{I}$?
- A: No, not for all $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$.
- Q2: Is it possible to identify $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ such that

$$\rho_P(I_v, I_w) \leq 0,$$

  for all pairs in $\boldsymbol{I}$?
- A: Yes. We need to introduce the Strongly Rayleigh (SR) property.

Part 3: Strongly Rayleigh Property

First, we recall the most famous notion of negative dependence.

Definition 8 (Negatively associated (NA))

A d-dimensional random vector $\boldsymbol{I} \in \mathcal{B}_d(\boldsymbol{q})$ is said to be negatively associated (NA) if for any two disjoint set $\Lambda_1, \Lambda_2 \subseteq \mathcal{V}$ and two monotone increasing functions h_1, h_2, the following inequality holds

$$E[h_1(I_j : j \in \Lambda_1)h_2(I_j : j \in \Lambda_2)] \leq E[h_1(I_j : j \in \Lambda_1)]E[h_2(I_j : j \in \Lambda_2)],$$

provided that the expectations are finite [Joag-Dev and Proschan, 1983].

The SR property was introduced in [Borcea et al., 2009]:

- It answers questions raised in the paper *Toward a theory of negative dependence* [Pemantle, 2000].
- [Borcea et al., 2009] is considered as an important achievement in the theory.
- The authors connects negative dependence, geometry, and algebra.
- Key result: SR $\Rightarrow$ NS.
- It is easier to check SR than to check NA.

Part 3: Strongly Rayleigh Property

The SR property is defined for MB distributions in terms of the joint pgf.

Definition 9 (Strongly Rayleigh property)

The rv $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ satisfies the strongly Rayleigh property if its pgf $\mathcal{P}_{\boldsymbol{I}}$ is real stable.

Q: What is a multi-affine polynomial $\mathcal{P}$?

A: Theorem 5.6 of [Brändén, 2007]: A multi-affine polynomial $\mathcal{P}$ with real coefficients is real stable if and only if, for every $1 \leq j_1, j_2 \leq d$,

$$\frac{\partial}{\partial x_{j_1}}\mathcal{P}(\boldsymbol{x})\frac{\partial}{\partial x_{j_2}}\mathcal{P}(\boldsymbol{x}) \geq \frac{\partial^2}{\partial x_{j_1}\partial x_{j_2}}\mathcal{P}(\boldsymbol{x})\mathcal{P}(\boldsymbol{x}), \quad \text{for all } (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{13}$$

Example: Let $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_3(\boldsymbol{q}$ and pgf given by

$$\mathcal{P}_{\boldsymbol{I}}(\boldsymbol{t}) = \frac{1}{2}t_1 + \frac{1}{3}t_2 + \frac{1}{6}t_3.$$

We can verify that $\mathcal{P}_{\boldsymbol{I}}$ is stable using (13) and conclude that $\boldsymbol{I}$ satisfies the SR property.

Part 3: Strongly Rayleigh Property

Assume that $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ satisfies the SR property.

- *SR* $\Rightarrow$ *NA*: If $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$ satisfies the strongly Rayleigh property, then it implies that $\boldsymbol{I}$ satisfies the Negatively associated property.
- Negative pairwise correlation: $\rho_P(I_v, I_w) \leq 0$ for all pairs (I_v, I_w) of $\boldsymbol{I}$.
- Marginalization: The SR property is closed under marginalization: $\boldsymbol{I}_A = (I_v, v \in A)$, where $A \subset \mathcal{V}$ also satisfie the SR property.

Next: Find the distribution of $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ such that $\boldsymbol{I}$ satisfies the SR property.

Part 3: Conditional Bernoulli distributions

Consider $\mathcal{B}_d(\boldsymbol{q})$ with $d \geq 4$ and with $q_\bullet = m \in \{2, \ldots, d-2\}$.

We define the conditional Bernoulli distribution for $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ as follows:

- Consider a vector $\boldsymbol{K} = (K_1, \ldots, K_d)$ of independent Bernoulli random variables with marginal means $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d) \in (0,1)^d$.
- Define the Bernoulli random vector $\boldsymbol{I}^{\boldsymbol{\pi}}$ such that

$$\begin{aligned} p_{\boldsymbol{I}^{\boldsymbol{\pi}}}(\boldsymbol{i}) &= \Pr(I_1^{\boldsymbol{\pi}} = i_1, \ldots, I_m^{\boldsymbol{\pi}} = i_m) \\ &= \Pr(K_1 = i_1, \ldots, K_d = i_d | K_1 + \cdots + K_d = m), \quad \boldsymbol{i} \in \mathcal{A}_{d,m}. \end{aligned}$$

- Define the entropy $H(p_{\boldsymbol{I}})$ by

$$H(p_{\boldsymbol{I}}) = \sum_{\boldsymbol{i} \in \mathcal{A}_{d,m}} -p_{\boldsymbol{I}}(\boldsymbol{i}) \ln(p_{\boldsymbol{I}}(\boldsymbol{i})). \tag{14}$$

- Find π^* that maximizes the entropy $H(p_{\boldsymbol{I}^{\boldsymbol{\pi}}})$ over $\mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ with the constraints $E[I_v^{\boldsymbol{\pi}}] = q_v$, $v \in \mathcal{V}$.
- Then, $\boldsymbol{I} = \boldsymbol{I}^{\boldsymbol{\pi}^*}$ follows a conditional Bernoulli distribution with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ and pmf $p_{\boldsymbol{I}^H} = p_{\boldsymbol{I}^{\boldsymbol{\pi}^*}}$.

Part 3: Conditional Bernoulli distribution

Theorem 10

For any $\boldsymbol{q} \in (0,1)^d$, there exists $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ with pmf $p_{\boldsymbol{I}}$ that maximizes the entropy in the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{q})$ that follows a conditional Bernoulli distribution and that satisfies the SR property.

- Theorem 10 implies that every Fréchet class admits a Bernoulli random vector that satisfies both the SR and the Σ-countermonotonicity properties.
- Conditional Bernoulli distributions have been studied in the framework of sampling with unequal probabilities without replacement, [Chen, 2000].
- In Example 5.1 of [Ghosh et al., 2017], the authors show that $\boldsymbol{I} \sim$ Conditional Bernoulli distribution satisfies the SR property. We show that it also satisfies the Σ-countermonotonicity property.
- Using a property of the SR property ([Borcea et al., 2009]), we introduce the method to define an extension of the Conditional Bernoulli distribution for the case when $q_\bullet \in (m, m+1)$ for $m \in \{1, \ldots, d-2\}$ and $d \geq 3$.

Part 3: Conditional Bernoulli distribution

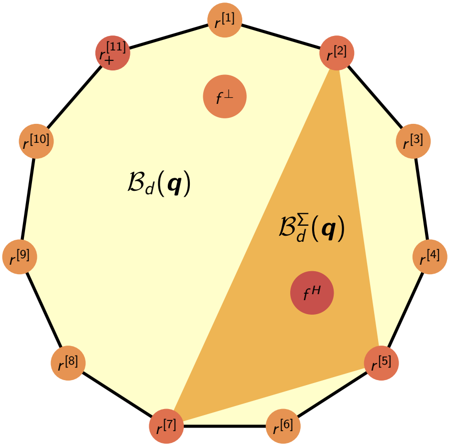


Figure: Representation of $\mathcal{B}_d(\boldsymbol{q})$ and $\mathcal{B}_d^{\Sigma}(\boldsymbol{q})$, where $d \geq 4$, $\boldsymbol{q}_\bullet \in (1,d-1)$.

Part 3: Numerical Illustration 1

Consider $\mathcal{B}_4^{\Sigma}(\boldsymbol{q}) \subset \mathcal{B}_4(\boldsymbol{q})$ where $\boldsymbol{q} = \left(\frac{7}{20}, \frac{9}{20}, \frac{10}{20}, \frac{14}{20}\right)$ and $q_\bullet = 2$:

- $\mathcal{B}_4^{\Sigma}(\boldsymbol{q})$ has 3 extremal points, denoted by $p_{\boldsymbol{I}^{(1)}}$, $p_{\boldsymbol{I}^{(2)}}$, and $p_{\boldsymbol{I}^{(3)}}$.
- $\boldsymbol{I}^{(4)}$ with $F_{\boldsymbol{I}^{(4)}} \in \mathcal{B}_4^{\Sigma}(\boldsymbol{q})$ follows the Conditional Bernoulli distribution.
- $p_{\boldsymbol{I}^{(4)}}$ admits the representation as a convex combination of the 3 extremal points:

$$p_{\boldsymbol{I}^{(4)}}(\boldsymbol{i}) = 0.4235 p_{\boldsymbol{I}^{(1)}}(\boldsymbol{i}) + 0.3090 p_{\boldsymbol{I}^{(2)}}(\boldsymbol{i}) + 0.2675 p_{\boldsymbol{I}^{(3)}}(\boldsymbol{i}), \quad \boldsymbol{i} \in \mathcal{A}_{4,2}.$$

$\boldsymbol{i}$	(1,1,0,0)	(1,0,1,0)	(0,1,1,0)	(1,0,0,1)	(0,1,0,1)	(0,0,1,1)	H
$p_{\boldsymbol{I}^{(1)}}(\boldsymbol{i})$	0	0	$\frac{3}{10}$	$\frac{7}{20}$	$\frac{3}{20}$	$\frac{1}{5}$	1.3351
$p_{\boldsymbol{I}^{(2)}}(\boldsymbol{i})$	0	$\frac{3}{10}$	0	$\frac{1}{20}$	$\frac{9}{20}$	$\frac{1}{5}$	1.1922
$p_{\boldsymbol{I}^{(3)}}(\boldsymbol{i})$	$\frac{3}{10}$	0	0	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{2}$	1.1421
$p_{\boldsymbol{I}^{(4)}}(\boldsymbol{i})$	0.080262	0.092688	0.177050	0.127050	0.242688	0.280262	1.691717

m	$\rho_P(I_1^{(m)}, I_2^{(m)})$	$\rho_P(I_1^{(m)}, I_3^{(m)})$	$\rho_P(I_1^{(m)}, I_4^{(m)})$	$\rho_P(I_2^{(m)}, I_3^{(m)})$	$\rho_P(I_2^{(m)}, I_4^{(m)})$	$\rho_P(I_3^{(m)}, I_4^{(m)})$
1	< 0	< 0	> 0	> 0	< 0	< 0
2	< 0	> 0	< 0	< 0	> 0	< 0
3	> 0	< 0	< 0	< 0	< 0	> 0
4	< 0	< 0	< 0	< 0	< 0	< 0

Quote from Patrizia: *The SR property is equivalent to spreading the negative dependence between all the pairs of* $\boldsymbol{I}$.

Part 3: Numerical Illustration 2

Consider the Fréchet class $\mathcal{B}_9(\boldsymbol{q})$ with the following values $(q_\bullet)$:

0.22	0.40	0.16	0.45	0.50	0.33	0.20	0.38	0.36

$$\Rightarrow \begin{bmatrix} I_1 & I_2 & I_3 \\ I_4 & I_5 & I_6 \\ I_7 & I_8 & I_9 \end{bmatrix}$$

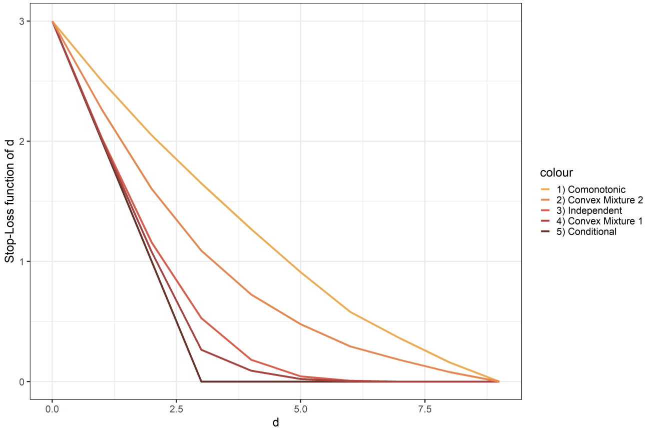
Definition of $\boldsymbol{I}^{(m)}$ with $F_{\boldsymbol{I}^{(m)}} \in \mathcal{B}_9(\boldsymbol{q})$, for $m = 1,2,3,4,5$:

- $\boldsymbol{I}^{(1)}$ = independent components; $\boldsymbol{I}^{(2)}$ = comomonotonic components;
- $\boldsymbol{I}^{(3)} \sim$ conditional Bernoulli distribution (= MB Distribution 3):
  - $p_{\boldsymbol{I}^{(3)}}(i_1, \ldots, i_d) = \begin{cases} > 0, & i_1 + \cdots + i_d = 3 \text{ (84 values)} \\ = 0, & i_1 + \cdots + i_d \neq 3 \text{ (428 values)} \end{cases}$.
  - For example, we have

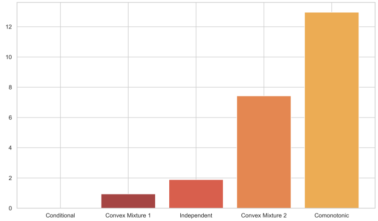
$p_{\boldsymbol{I}^{(3)}}(i_1, \ldots, i_d) = \Pr\Big($  $\Big) = 0.007669$ $\quad$ $p_{\boldsymbol{I}^{(3)}}(i_1, \ldots, i_d) = \Pr\Big($  $\Big) = 0.013192$

- $p_{\boldsymbol{I}^{(4)}}(\boldsymbol{i}) = \frac{1}{2} p_{\boldsymbol{I}^{(1)}}(\boldsymbol{i}) + \frac{1}{2} p_{\boldsymbol{I}^{(3)}}(\boldsymbol{i})$, for $\boldsymbol{i} \in \{0,1\}^9$.
- $p_{\boldsymbol{I}^{(5)}}(\boldsymbol{i}) = \frac{1}{2} p_{\boldsymbol{I}^{(2)}}(\boldsymbol{i}) + \frac{1}{2} p_{\boldsymbol{I}^{(3)}}(\boldsymbol{i})$, for $\boldsymbol{i} \in \{0,1\}^9$.

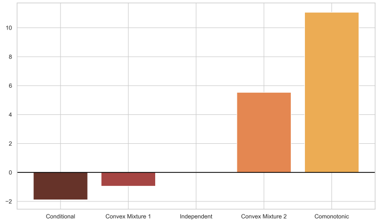
Part 3: Numerical Illustration 2



Stop-loss function $\pi_N(d) = E[\max(N - d; 0)]$



Variance of N

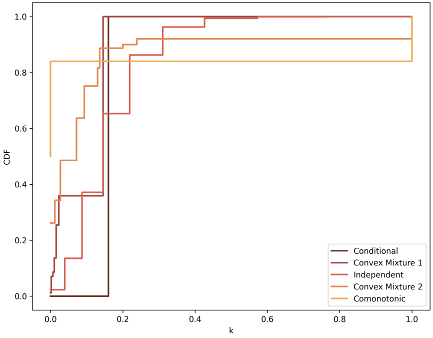


Sum of the covariances of the pairs of $\boldsymbol{I}$

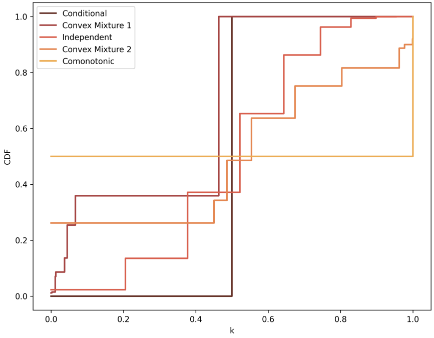
Part 3: Conditional expectation risk sharing rule

Consider a portfolio of d risks represented by the Bernoulli rvs $\boldsymbol{I}$ with $F_{\boldsymbol{I}} \in \mathcal{B}_d(\boldsymbol{q})$.

- The aggregate loss amount rv is $S = bI_1 + \cdots + bI_d = bN$. We assume $b = 1$.
- Contribution of risk v under the conditional expectation sharing rule: $\mathcal{C}_v(\boldsymbol{I}) = E[I_v|N], \quad v \in \mathcal{V}$.
- We compute the values of $\mathcal{C}_v(\boldsymbol{I})$ using joint pgf of $\boldsymbol{I}$ and FFT algorithm [Blier-Wong et al., 2025].
- **Numerical Illustration 2 (cont'd)** Fréchet class $\mathcal{B}_9(\boldsymbol{q})$, where $q_\bullet = m = 3$, $q_3 = 0.16$, $q_5 = 0.5$.



Conditional expectation of risk 3



Conditional expectation of risk 5

Conclusion

Brief summary:

1. In our journey throughout these two research projects about the family of multivariate Bernoulli distributions, we have to work out of our comfort zone.
2. We had to study concepts in geometry, algebra, machine learning, and graph theory, and combine them with our expertise about dependence models.
3. That might explain why we still have fun to do research with our students and colleagues.
4. Maybe, that's what the author of [Su, 2020] means by *Mathematics for Human Flourishing*.

Forthcoming:

1. Consider Markov Random Fields where the underlying dependence structure is encrypted in junction trees.
2. Investigate aggregation methods for multivariate Bernoulli distributions in high dimension.
3. Investigate dependence properties of the tree-structured Ising models.
4. Explore the multiple facets of the strongly Rayleigh property and its applications in actuarial science.

Conclusion

Thanks to . . .

- . . . the organizers for the invitation, notably Jingyi and Dongchen
- . . . Tom for being the moderator of my talk
- . . . Anthony and Philippe for technical support
- . . . my co-authors Alessandro, Benjamin, Hélène, and Patrizia
- . . . the students that I have supervised.

Acknowledgements:

- Department of Mathematical Sciences *Giuseppe Luigi Lagrange*, Politechnico di Torino;
- NSERC Discovery grant program: B Côté, H Cossette, E Marceau;
- Chaire d'actuariat, Université Laval.

Figure: Thank you for your attention! [5]

[5] JB also discovered the constant e by studying compound interest ☺.

References I

4ti2 team.
4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces.

Axelrod, D. and Kimmel, M. (2015).
Branching Processes in Biology.
Springer.

Blier-Wong, C., Cossette, H., and Marceau, E. (2025).
Efficient evaluation of risk allocations.
Insurance: Mathematics and Economics, 122:119–136.

Borcea, J., Brändén, P., and Liggett, T. (2009).
Negative dependence and the geometry of polynomials.
Journal of the American Mathematical Society, 22(2):521–567.

References II

Brändén, P. (2007).
Polynomials with the half-plane property and matroid theory.
Advances in Mathematics, 216(1):302–320.

Bujack, R., Turton, T. L., Samsel, F., Ware, C., Rogers, D. H., and Ahrens, J. (2017).
The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps.
IEEE Transactions on Visualization and Computer Graphics, 24(1):923–933.

Chen, S. X. (2000).
General properties and estimation of conditional Bernoulli models.
Journal of Multivariate Analysis, 74(1):69–87.

Chow, C. and Liu, C. (1968).
Approximating discrete probability distributions with dependence trees.
IEEE transactions on Information Theory, 14(3):462–467.

References III

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2022).
Introduction to algorithms.
MIT Press.

Cossette, H., Marceau, E., Mutti, A., and Semeraro, P. (2025).
Extremal negative dependence and the strongly Rayleigh property.
arXiv preprint arXiv:2504.17679.

Côté, B., Cossette, H., and Marceau, E. (2025).
Tree-structured Markov random fields with Poisson marginal distributions.
Journal of Multivariate Analysis, page 105418.

Crameri, F. (2018).
Scientific colour maps.
Zenodo, 10:5281.

References IV

Crameri, F., Shephard, G. E., and Heron, P. J. (2020).
The misuse of colour in science communication.
Nature Communications, 11(1):5444.

Denuit, M. and Robert, C. Y. (2022).
Conditional mean risk sharing in the individual model with graphical dependencies.
Annals of Actuarial Science, 16(1):183–209.

Dhaene, J. and Denuit, M. (1999).
The safest dependence structure among risks.
Insurance: Mathematics and Economics, 25(1):11–21.

Flajolet, P. and Sedgewick, R. (2009).
Analytic Combinatorics.
Cambridge University Press.

References V

Fontana, R. and Semeraro, P. (2018).
Representation of multivariate Bernoulli distributions with a given set of specified moments.
Journal of Multivariate Analysis, 168:290–303.

Ghosh, S., Liggett, T. M., and Pemantle, R. (2017).
Multivariate CLT follows from strong Rayleigh property.
In *2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 139–147. SIAM.

Goverment of Canada, C. (2025).
Climate Change Canada: Strategies and Initiatives, Environment and Natural Resources, Goverment of Canada.
https://www.canada.ca/en/services/environment/weather/climatechange.html.
Accessed: 2025-07-07.

References VI

Joag-Dev, K. and Proschan, F. (1983).
Negative association of random variables with applications.
The Annals of Statistics, pages 286–295.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997).
Discrete Multivariate Distributions.
Wiley.

Pemantle, R. (2000).
Towards a theory of negative dependence.
Journal of Mathematical Physics, 41(3):1371–1390.

Pixabay, W. (2025).
Canada Map, Vectors, Pixabay.
https://pixabay.com/vectors/map-canada-provinces-territories-2088308/.
Accessed: 2025-07-07.

References VII

Puccetti, G., Wang, R., et al. (2015).
Extremal dependence concepts.
Statistical Science, 30(4):485–517.

Scotto, M. G., Weiß, C. H., and Gouveia, S. (2015).
Thinning-based models in the analysis of integer-valued time series: a review.
Statistical Modelling, 15(6):590–618.

Sedgewick, R. and Wayne, K. (2011).
Algorithms.
Addison-wesley professional.

Steutel, F. W. and van Harn, K. (1979).
Discrete analogues of self-decomposability and stability.
The Annals of Probability, pages 893–899.

References VIII

Steutel, F. W., Vervaat, W., and Wolfe, S. J. (1983).
Integer-valued branching processes with immigration.
Advances in applied probability, 15(4):713–725.

Su, F. (2020).
Mathematics for human flourishing.
Yale University Press.

Tang, W. and Tang, F. (2023).
The Poisson binomial distribution—old & new.
Statistical Science, 38(1):108–119.

Wainwright, M. J., Jordan, M. I., et al. (2008).
Graphical models, exponential families, and variational inference.
Foundations and Trends® in Machine Learning, 1(1–2):1–305.

References IX

Wang, S. S. (1998).
Aggregation of correlated risk portfolios: Models and algorithms.
Proceedings of the Casualty Actuarial Society, pages 848–939.

Weiß, C. H. (2008).
Thinning operations for modeling time series of counts—a survey.
AStA Advances in Statistical Analysis, 92:319–341.

Abstract

Multivariate Bernoulli distributions are essential in the modeling of binary data in a wide variety of contexts, such as actuarial science, quantitative risk management, machine learning, natural language processing, and bioinformatics. In this talk, I will present recent results on two topics related to multivariate Bernoulli distributions.

In the first part of my talk, I will consider tree-structured Ising models, a class of undirected graphical models for Bernoulli random vectors. We introduce a stochastic representation of the components of the Bernoulli random vectors such that the marginal distributions remain fixed. That representation has many advantages: to find the Pearson's correlation coefficient for any pairs of components of Bernoulli random vectors; to design an efficient sampling algorithm; and to investigate properties of the tree-structured multivariate Bernoulli distributions. We also derive an analytic expression for the joint probability generating function of the Bernoulli random vector. The latter is used to build efficient computation methods for the sum of the components of the Bernoulli random vector.

In the second part, I will discuss negative dependence for Bernoulli random vectors. We will provide the essential tools of negative dependence and extremal negative dependence in a common language and the framework of multivariate Bernoulli distributions. We will characterize sigma-countermonotonicity and study the strongly Rayleigh property within this class. We will illustrate those notions using the class of conditional Bernoulli distributions.

To conclude, I will present examples with actuarial applications of the results from the two topics.

Keywords: Ising model; Tree-structured undirected graphical models; Multivariate Bernoulli distribution; Computational Methods; Negative dependence; Strongly Rayleigh Property; Actuarial Applications

Bio

Etienne Marceau is a Full Professor at the School of Actuarial Science at Université Laval. He has been Distinguished Visiting Scholar at UNSW (2019), Visiting Professor at Université Lyon III (2005-2024), and Adjunct Professor at McGill University (2015-2021). His research interests focus on topics which are the intersections of actuarial science, applied probability, and applied statistics, such as dependence modeling and actuarial modeling. He has taught courses on actuarial modeling, risk theory, life contingencies, survival models, and financial mathematics. Strongly believing in HQP training, he has supervised 9 PhD, 31 MSc, and 35 BSc students. He has published more than 60 scientific articles, mostly with HQP, and he is the author of "Modélisation et évaluation quantitative des risques en actuariat" (Springer 2014). He holds the NSERC Discovery Grant and he co-PI on two NSERC CRD grants. He is the co-chair of the ACTRISK laboratory at Université Laval, of the CIMMUL, and a member of the editorial board of Insurance: Mathematics and Economics. He has been chair of CRM-Quantact. He has also been involved with the Society of Actuaries (SOA), as a Faculty for the Course 7 and a Faculty Advisor on the evaluation committee of the CAE of the SOA. He is a member of CIMMUL, CRDM , and CIRCERB. He served as an external research and development advisor, notably for Aon Hewitt, the Quebec Pension Plan and Industrial Alliance.

Part 2: Numerical Example using Precipitation Data

	EDM	VIC	WIN	FRE	STJ	HAL	TOR	CHA	QUE	REG	WHI	SEP	ROU	THU	OTT	CAL	KEL	PRI	CHU
EDM	1	0.039	0.014	0	0	0	0	0	0	0.064	0.02	0	0.001	0.004	0	0.314	0.015	0.146	0
VIC	0.026	1	0.001	0	0	0	0	0	0	0.002	0.038	0	0	0	0	0.012	0.373	0.268	0
WIN	0.069	-0.013	1	0.002	0	0.001	0.004	0	0.005	0.223	0	0.002	0.037	0.247	0.011	0.045	0	0.002	0
FRE	0.014	0.006	-0.031	1	0.033	0.456	0.069	0.171	0.41	0	0	0.144	0.054	0.008	0.178	0	0	0	0.035
STJ	-0.024	0.06	0.02	-0.041	1	0.071	0.002	0.19	0.013	0	0	0.005	0.002	0	0.006	0	0	0	0.001
HAL	0.011	0.036	-0.021	0.456	0.007	1	0.031	0.376	0.187	0	0	0.066	0.025	0.004	0.081	0	0	0	0.016
TOR	-0.01	0.009	-0.003	0.066	-0.041	0.027	1	0.012	0.167	0.001	0	0.059	0.116	0.018	0.385	0	0	0	0.014
CHA	-0.01	0.057	0.003	0.321	0.19	0.376	-0.045	1	0.07	0	0	0.025	0.009	0.001	0.031	0	0	0	0.006
QUE	-0.006	0.016	-0.045	0.41	-0.037	0.286	0.142	0.164	1	0.001	0	0.35	0.131	0.02	0.434	0	0	0	0.085
REG	0.174	0.017	0.223	-0.004	0.023	-0.011	-0.021	0.001	-0.023	1	0.001	0	0.008	0.055	0.003	0.203	0.001	0.009	0
WHI	0.013	0.054	0.009	0.014	0.007	0.008	0.015	-0.006	0.027	-0.036	1	0	0	0	0	0.006	0.014	0.14	0
SEP	0.015	-0.017	-0.023	0.277	-0.018	0.228	0.007	0.211	0.35	0.005	0.024	1	0.046	0.007	0.152	0	0	0	0.244
ROU	-0.031	0.021	0.004	0.105	-0.027	0.04	0.171	-0.021	0.256	-0.027	0.038	0.12	1	0.151	0.302	0.002	0	0	0.011
THU	0.034	-0.004	0.247	-0.039	-0.003	-0.041	0.11	-0.042	-0.023	0.073	0	-0.039	0.151	1	0.046	0.011	0	0.001	0.002
OTT	-0.024	0.014	-0.038	0.247	-0.046	0.174	0.385	0.046	0.434	-0.032	0.018	0.154	0.302	0.014	1	0.001	0	0	0.037
CAL	0.314	-0.03	0.076	0.013	-0.006	0.001	-0.041	0.002	-0.013	0.203	-0.065	0.028	-0.038	0.014	-0.044	1	0.005	0.046	0
KEL	0.138	0.373	-0.01	0.005	0.018	0.015	0.007	0.033	0.014	0.053	0.021	0.001	-0.004	-0.014	0.007	0.14	1	0.1	0
PRI	0.146	0.268	-0.001	0.006	0.014	0.02	0.023	0.03	0.014	0.003	0.14	-0.012	0.018	0.01	0.016	0.002	0.26	1	0
CHU	0.022	-0.031	-0.021	0.006	-0.001	0.012	-0.06	0.004	0.036	0.004	0.028	0.244	0.007	-0.079	-0.011	0.026	-0.037	-0.008	1

Table: Correlation matrix where the inferior triangle corresponds to the empirical values from the training data and the superior triangle corresponds to the model's estimated values

Part 2: Numerical Example using Precipitation Data

Let $N = \sum_{j=1}^{19} I_j$, we can compare the values for the pmf of N for the testing data, the estimated model and an independent case as reference using the Poisson binomial distribution (see [Tang and Tang, 2023]).

k	Empirical $f_N(k)$	Model $f_N(k)$	Independent $f_N(k)$
0	0	0.0005	0.00001
1	0.00264	0.00307	0.00014
2	0.01201	0.01012	0.00112
3	0.02186	0.02372	0.00541
4	0.05022	0.04415	0.01842
5	0.07256	0.06921	0.04676
6	0.08938	0.09472	0.09182
7	0.11893	0.11579	0.1427
8	0.12302	0.1281	0.17811
9	0.12638	0.12897	0.18013
10	0.11917	0.11798	0.14821
11	0.09178	0.09747	0.09921
12	0.07304	0.07206	0.05381
13	0.04998	0.04711	0.02344
14	0.02763	0.0268	0.00808
15	0.01466	0.01299	0.00215
16	0.00481	0.0052	0.00043
17	0.00168	0.00163	0.00006
18	0.00024	0.00036	0.00001
19	0	0.00004	0

Part 2: Numerical Example using Precipitation Data

Using the *devon* scientific color map from [Crameri, 2018], we can visualize the differences in the cdf of N.