A review of negative dependence

Robin Pemantle
University of Pennsylvania

MWPC43, 21 October 2022

A review of negative dependence

This talk surveys developments over the last 50-60 years.

Most of the work is not mine, but I have intersected it on several occasions, as a contributor and as a user.

Undoubtedly, the Strong Rayleigh property of Borcea, Brändén and Liggett (JAMS, 2009), is one of the crowning achievements of the theory. They connect negative dependence to the geometry of zero sets of complex polynomials. Their theorem "SR $\Rightarrow$ NA" proves a strong conclusion (negative association) under a surprisingly checkable hypothesis. The result covers many cases of interest.

Developments since then include further consequences of SR, distributional consequences of the geometry of zero sets, and the more general notion of Lorentzian polynomials and measures.

Notation

Some notation throughout the lectures for collections of random variables taking the values 0 and 1:

$\mathcal{B}_n := \{0,1\}^n$ is a Boolean lattice of rank n with the product partial order.

$\mathbb{P}$ is a probability measure on $\mathcal{B}_n$

The random variable X_k is the k^{th} coordinate, i.e., $X_k(\omega_1, \ldots, \omega_n) = \omega_k$.

A probability measure μ is said to stochastically dominate another law ν (written $\mu \succeq \nu$) if $\mu(A) \geq \nu(A)$ for all sets A that are upwardly closed in the partial order. This is equivalent to the existence of a coupling measure Q on $\mathcal{B}_n^2$ supported on $\{(x,y) : x \leq y\}$ such that the projection of Q onto the first coordinate is ν and onto the second is μ.

Some definitions and properties

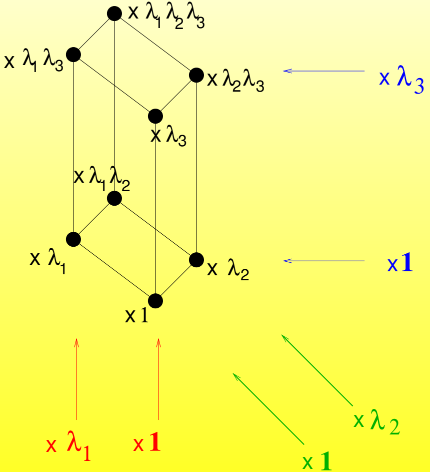
Many of these properties make sense for joint laws of real random variables as well as binary random variables. For reasons of time, I will stick to binary variables. See Karlin and Rinott (JMAA, 1980) for a discussion of properties of bivariate densities that lead to various kinds of positive or negative dependence.

If $\lambda_1, \ldots, \lambda_n$ are positive real numbers, the measure defined by

$$\mu_\lambda(\omega) := \frac{1}{Z} \mu(\omega) \prod_{j=1}^{n} \lambda_j^{\omega_j} \qquad Z := \sum_{\omega}' \mu(\omega') \prod_{j=1}^{n} \lambda_j^{\omega_j'}$$

is said to be a result of applying an external field to μ. If P is a property of measures on $\mathcal{B}_n$, then P^+ holds for μ if P holds for all measures obtained from μ via external fields.

External field



Negative dependence hierarchy

Each of the following properties is implied by the next.

- **NC** Pairwise negative correlation: $\mathbb{E} X_i X_j \leq (\mathbb{E} X_i)(\mathbb{E} X_j)$.
- **NCP** Negative cylinder property: $\mathbb{E} \prod_{k \in A} X_k \leq \prod_{k \in A} \mathbb{E} X_k$.
- **NA** Negative association: $\mathbb{E} fg \leq (\mathbb{E} f)(\mathbb{E} g)$ whenever f and g are increasing functions on $\mathcal{B}_n$ measurable with respect to disjoint sets of coordinates.
- **SR** Strong Rayleigh (definition TBA)

Motivation for studying negative dependence properties

Motivation: tail bounds on sums

Now that we have some defined terms, we can go back and discuss why anyone wants to study negative (or positive) dependence.

One motivating factor for the study of negative dependence was to get tail bounds on the sum $S := \sum_{j=1}^{k} X_j$.

1. NC implies $\mathbb{P}(|S - ES| \geq a) \leq n/(4a^2)$
2. NCP implies Gaussian bounds, via bounds on $\mathbb{E}e^{\lambda S}$:
   $\mathbb{P}(S - \mathbb{E}S \geq a) \leq \exp(-2a^2/n)$
3. NA implies a self-normalized CLT: $(\mathrm{S} - \mathbb{E}\mathrm{S})/\mathrm{Var}\,(\mathrm{S})^{1/2} \to \chi$ (Newman, 1982, bounding difference of char. fn.)
4. SR implies Gaussian tail bounds for all Lipschitz functionals on $\mathcal{B}_\mathrm{n}$ (details will be given later in the lecture).

Motivation: Controlling the probability of events

Another use of positive or negative dependence is to control the effects of conditioning.

Algorithms on graphs such as searches, message passing, and other local parallel algorithms can suffer the "curse of knowing too much," making the performance of the algorithm difficult to analyze because the effects of earlier steps on later conditional distributions is hard to describe. If this information all goes in the same direction, one can at least get one-sided bounds on the effects of conditioning.

Case study: spanning trees

Fix the dimension d and define a random graph G_N to be a uniformly chosen spanning tree of the centered box of volume $(2N + 1)^d$ in the usual nearest neighbor graph on $\mathbb{Z}^d$. How do we know that G_N approaches a weak limit, which is now called the free uniform spanning forest on $\mathbb{Z}^d$? We look at three proofs of this.

The original proof (1991) showed convergence via electrical theory. Let H be a finite set of edges of $\mathbb{Z}^d$ containing no cycle. One may express $\mathbb{P}(H \subseteq G_N)$ as a product of resistances, taking each edge of G_N to have resistance 1. A physical law, due to Rayleigh, says that resistances can only decrease when more edges are added. Hence $\mathbb{P}(H \subseteq G_N)$ decreases and must converge to some p_H. The probabilities p_H define the unique weak limit of the uniform spanning tree on G_N.

Better arguments

Shortly thereafter, Feder and Mihail (1992) showed that the uniform spanning tree measure on a finite graph was negatively associated. This is a much stronger result. It follows that the sequence $G_N|_B$ of restrictions of G_N to a given finite region B is stochastically decreasing, whereas the previous proof shows only that the probabilities of cylinder events $H \subseteq G_N$ decrease.

Nowadays we would shortcut the Feder-Mihail proof by noting that the spanning tree measure is known to be determinantal, hence strong Rayleigh, hence negatively associated.

The problem with negative dependence

Why is negative dependence hard?

Useful results were developed for positive dependence long before they were for negative dependence.

Intuitively, this is because if X and Y are positively related and Y and Z are positively related, this is not an obstacle for X and Z to be positively related.

For negative dependence, on the other hand, there is frustration.

For example, if (X, Y, Z) is Markov, and X and Y are negatively correlated and Y and Z are negatively correlated, then X and Y will be positively correlated.

The enemy of my enemy is my friend.

Narrow range for negative dependence

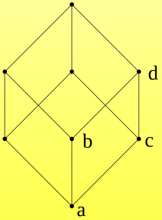
Another manifestation is that the total amount of negative correlation to go around is a lot less than the total amount of positive correlation: $\sum_{i \neq j} \mathbb{E} X_i X_j - (\mathbb{E} X_i)(\mathbb{E} X_j)$ must be between $-n$ and $n^2 - n$.

There is a lot more room for positive correlation (for example when all X_i are the same) than for negative correlation (for example when the sum is constant).

One more indicator of why positive dependence conditions are more harmonious than negative dependence conditions comes from studying conditions of the weights $\mathbb{P}(\omega)$ as lattice functions.

Lattice conditions

A 4-tuple $(\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d})$ of the Boolean lattice $\mathcal{B}_{\mathrm{n}}$ is a **diamond** if b and c cover a and if d covers b and c, where x covers y if $\mathrm{x} \geq \mathrm{y}$ and $\mathrm{x} \geq \mathrm{u} \geq \mathrm{y}$ implies $\mathrm{u} = \mathrm{x}$ or $\mathrm{u} = \mathrm{y}$.



Say that $\mathbb{P}$ satisfies the **positive lattice condition** if $\mathbb{P}(\mathrm{b})\mathbb{P}(\mathrm{c}) \leq \mathbb{P}(\mathrm{a})\mathbb{P}(\mathrm{d})$ for every diamond $(\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d})$. The reverse inequality is called the **negative lattice condition** (NLC).

Positive association

Say that the measure $\mathbb{P}$ on $\mathcal{B}_n$ is **positively associated** if

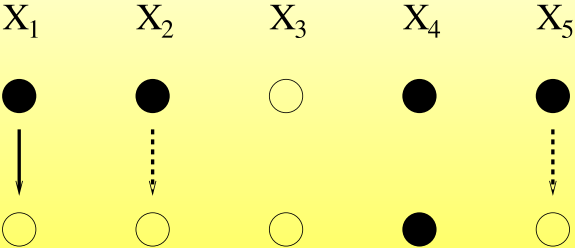
$$\mathbb{E}fg \geq (\mathbb{E}f)\ (\mathbb{E}g)$$

whenever f and g are both monotone increasing on $\mathcal{B}_n$. Taking $f = X_i, g = X_j$ this implies pairwise nonnegative correlation.

Take $f = X_1$ and let $\mathbb{P}_1$ and $\mathbb{P}_0$ denote the conditional distribution of $\mathbb{P}$ given $X_1 = 1$ and $X_1 = 0$ respectively. In this case positive association says $\int g\, d\mathbb{P}_1 \geq \int g\, d\mathbb{P}_0$ for all increasing functions g.

In other words, $\mathbb{P}_1 \succeq \mathbb{P}_0$ (stochastic domination).

Coupling for positive association



One can sample simultaneously from $(\mathbb{P}|X_1 = 1)$ and $(\mathbb{P}|X_1 = 0)$ in such a way that turning off the bit at X_1 also turns off some of the other bits (in this case X_2 and X_5).

FKG

The positive lattice condition is very useful, due to the following result of Fortuin, Kastelyn and Ginibre (1971).

Theorem 1 (FKG)

If $\mathbb{P}$ satisfies the positive lattice condition then $\mathbb{P}$ is positively associated and the projection of $\mathbb{P}$ to any smaller set of variables satisfies both these conditions as well.

The positive lattice condition involves checking the ratios of probabilities of nearby configurations. This is often much easier than computing correlations between bits, which involves summing over all configurations.

Negative association

Negative association is a trickier business because f can't be negatively correlated with itself.

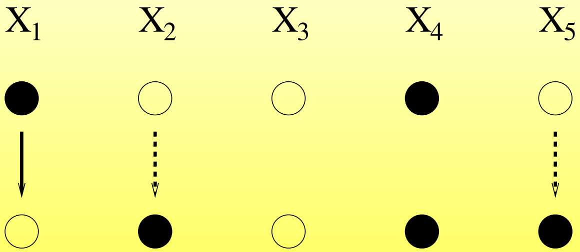
The measure $\mathbb{P}$ on $\mathcal{B}_n$ is **negatively associated** if

$$\mathbb{E}fg \leq (\mathbb{E}f)\ (\mathbb{E}g)$$

whenever f and g are both monotone increasing and they depend on <u>disjoint</u> sets of coordinates.

Taking $f = X_1$, the consequence is that the conditional law of the remaining variables given $X_1 = 0$ stochastically dominates the law given $X_1 = 1$. Thus a sample conditioned on $X_1 = 1$ is obtained from one conditioned on $X_1 = 0$ by turning some ones into zeros, except the first coordinate, which goes from zero to one.

Coupling for negative association



This time, turning off the bit X_1 causes the sample from $(\mathbb{P}|X_1 = 1)$ to **gain** some ones when it turns into a sample from $(\mathbb{P}|X_1 = 0)$.

Negative lattice condition

Unfortunately, there is not a version of the FKG theorem holding when the positive lattice condition is replaced by the negative lattice condition. In particular, NA does not follow from this, and the NLC is not stable under forgetting one of the variables.

As a result, negative association is very difficult to check!

A profusion of properties has been suggested that are somewhat weaker than NA. These are not totally ordered with respect to implication. Many concern the stochastic domination of some conditional distribution of $\mathbb{P}$ by others.

A mess of properties

You can ignore the details and view the next few slides impressionistically

Lattice based conditions

Needless to say, the is no corresponding theorem for the negative lattice condition. Both lattice conditions are closed under external fields, an extreme case of which is conditioning on the value of a variable. But the negative lattice condition is not closed under ignoring variables (projecting to a smaller lattice), so is not a natural condition.

Just as we use $+$ for a property extended to external fields, let h-P denote the property P holding hereditarily, that is, for all subsets of variables.

In this notation, for example, the condition for real variables introduced by Karlin and Rinott (JMAA, 1980) called S-MRR_2, would, for binary variables, be called h-NLC^+.

JNRD

The notion of Joint Negative Regression Dependence (JNRD) is inherited from the theory of real random variables. In the binary context it means that if A is some subset of the set $[n]$ of indices, then the conditional law $(X_j : j \notin A | X_i = x_i : i \in A)$ is stochastically decreasing in the partial order on binary vectors **x** in $\{0, 1\}^A$.

Implications

By 2000, this was the state of affairs, as documented in a paper of mine[1].

$$\begin{tikzcd}
\mathrm{NA}+ \arrow[r] \arrow[d] & \mathrm{JNRD}+ \arrow[r] \arrow[d] & \begin{array}{l}\text{S-MMR}_2\\ \text{h-NLC}+\end{array} \arrow[d] \\
\mathrm{CNA} \arrow[r] \arrow[d] & \mathrm{JNRD} \arrow[r] & \text{h-NLC} \\
\mathrm{NA} & &
\end{tikzcd}$$

Here, CNA means NA after conditioning on the values of any variables. But, what checkable condition implies NA?

[1] This is my paper with the most citations but also the least success in terms of results proved and conjectures that have not been disproved.

Ultra-log-concavity

Consider the case where $S = \sum_{j=1}^n X_j$ is the sum of independent (but not IID) Bernoullis. The generating polynomial $f(z) := \mathbb{E}z^S = \displaystyle\sum_{j=0}^{n} \mathbb{P}(S = j) z^j$ has all real roots.

Theorem 2 (Newton, 1707)

If $\mathbb{E}z^S$ has all real roots then the sequence

$$\left\{ \frac{\mathbb{P}(S = j)}{\binom{n}{j}} \right\}$$

is log-concave.

Ultra-log-concavity is not closed under external fields but I hoped (wrongly) that ULC+ is a natural property implying NA.

One thing turned out to be true

Natural properties of measures on $\mathcal{B}_n$ should hold for $\mu \times \nu$ if they hold for μ and ν. For ULC, this means that the convolution of two ULC measures should be ULC. In my infamous paper, I conjectured this but could not prove it.

But something good came out of this...

One thing turned out to be true

Natural properties of measures on $\mathcal{B}_n$ should hold for $\mu \times \nu$ if they hold for μ and ν. For ULC, this means that the convolution of two ULC measures should be ULC. In my infamous paper, I conjectured this but could not prove it.

But something good came out of this...

Theorem 3 (Liggett, 2001)

The convolution of two ULC measures is ULC.

Those of you who knew Tom might guess how he proved it: he found an inspired combination of algebraic identities and inequalities.

Even bad ideas can be good

This didn't do a whole lot on its own: it was only a corroboration of the fact that ULC might be important.

The best thing that came out of this was that Tom was hooked on this mess of a non-theory, and that he too thought ULC, hence real-rootedness, might play a role.

> **** passage of several years, during which Liggett managed to connect with Petter Brändén, who is an expert in the geometry of roots of multivariate polynomials. ****

Multivariate polynomials? Yes, the generating function

$$f(z_1, \ldots, z_n) = \sum \mathbb{P}(\mathbf{X} = \mathbf{x}) \prod_{j=1}^{n} z_j^{x_j}$$

turns out to be the key player.

Strong Rayleigh distributions

Generating functions

Given a set of random variables $X_1, \ldots, X_n$ taking values in $\mathbb{Z}^+$, the associated generating function is the polynomial in n variables defined by

$$F(x_1, \ldots, x_n) = \sum_{a_1, \ldots, a_n} \mathbb{P}(X_1 = a_1, \ldots, X_n = a_n) \, x_1^{a_1} \cdots x_n^{a_n} \, .$$

When the variables $\{X_n\}$ are Boolean, the corresponding generating function is multi-affine: no powers can by higher than 1.

A useful identity computes the probability of all 1's in a set A:

$$\mathbb{E} \prod_{k \in A} X_k = \frac{\partial}{\partial x_{k_1}} \cdots \frac{\partial}{\partial x_{k_r}} F(1, \ldots, 1)$$

where $k_1, \ldots, k_r$ enumerates A.

Definition of strong Rayleigh

In terms of the generating function, NC is expressed by

$$\mathbb{E} X_i X_j \leq (\mathbb{E} X_i)(\mathbb{E} X_j) \Longleftrightarrow F(\mathbf{1}) \frac{\partial^2 F}{\partial x_i x_j}(\mathbf{1}) \leq \frac{\partial F}{\partial x_i}(\mathbf{1}) \frac{\partial F}{\partial x_j}(\mathbf{1}) \,. \qquad (1)$$

NC^+ requires this not just at $(1, \ldots, 1)$ but at all points $\mathbf{x}$ in the positive orthant; this is known as the **Rayleigh** property.

$$\mathbb{E}_{\mathbf{x}} X_i X_j \leq (\mathbb{E}_{\mathbf{x}} X_i)(\mathbb{E}_{\mathbf{x}} X_j) \Longleftrightarrow F(\mathbf{x}) \frac{\partial^2 F}{\partial x_i x_j}(\mathbf{x}) \leq \frac{\partial F}{\partial x_i}(\mathbf{1}) \frac{\partial F}{\partial x_j}(\mathbf{x}) \,. \quad (2)$$

Definition (strong Rayleigh)

The law $\mathbb{P}$ with generating function F is called **strong Rayleigh** *if* (2) *holds for all $\mathbf{x} \in \mathbb{R}^n$, not just $\mathbf{x}$ in the positive orthant. In other words, NC persists under positive or negative external fields.*

$R^+ \to \mathbb{R} \to \mathbb{C}$

In fact, understanding SR requires not only allowing $\mathbf{x}$ to be negative but allowing arguments of generating functions to be complex as well!

A function $F : \mathbb{C}^n \to \mathbb{C}$ is called stable if F is nonzero on $\mathbb{H}^n$, where H is the open upper half plane.

Generating functions of probability measures on $\mathcal{B}_n$ are multi-affine, meaning each monomial has degree zero or one in each variable. A key result in the development of the strong Rayleigh property is:

Theorem (Borcea, Branden and Liggett, 2009, Lemma 4.1)

$\mathbb{P}$ *is SR if and only if its generating function F is stable.*

Some things that preserve the SR property

It is not the place to take a long detour into the theory of stable functions.

Instead, I will state two results whose proofs require this detour.

These results are very intuitive when stated probabilistically. Once we accept them, the remaining results can be argued in a more or less self-contained manner.

Further details may be found in the original source Borcea, Brändén and Liggett (2009) or in my (2012) survey.

Polarization

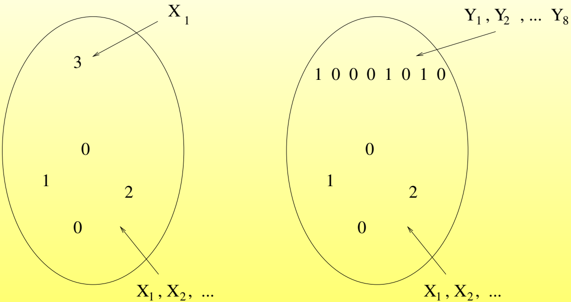
Let $X_1, \ldots, X_n$ be nonnegative integer random variables, all bounded by M. Polarization means replacing X_1 by Boolean variables $\{Y_1, \ldots, Y_M\}$ such that, conditional on $X_1, \ldots, X_n$, the Y variables are exchangeable and sum to X_1.

Lemma 4

If the generating function for $X_1, \ldots, X_n$ is stable then the generating function for $Y_1, \ldots, Y_M, X_2, \ldots, X_n$ is stable.

The polarization construction can be described in algebraic terms, without reference to probability, and is proved via the Grace-Welsh-Szegö Theorem.

Splitting X_1 into exchangeable variables $Y_1, \ldots, Y_m$



On the left is a sample from a distribution on positive integers where all variables are bounded by $M := 8$.

On the right, given that $X_1 = 3$, this variable was replaced by 8 binary variables, three of which were chosen to be 1, uniformly among the $\binom{8}{3}$ possibilities.

Homogenization

Often algebra works better with homogeneous polynomials. A generating function F is homogeneous if and only if the random variable $S := \sum_{k=1}^n S_k$ is constant.

Lemma 5 (Homogenization Lemma)

Let F be a stable polynomial in n variables with nonnegative real coefficients. Then the (usual) homogenization of F is a stable polynomial in $n+1$ variables.

The proof uses hyperbolicity theory, showing that nonnegative directions are in the cone of hyperbolicity.

Probabilistic interpretaion: if $\{X_1, \ldots, X_n\}$ have stable generating function then adding $X_{n+1} := n - \sum_{k=1}^n S_k$ preserves stability.

Symmetric homogenization

Putting these two constructions together yields a natural stability preserving operation within the realm of Boolean measures.

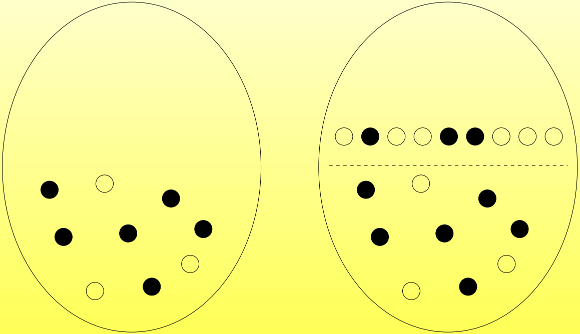
Definition 6

The symmetric homogenization of a measure on $\mathcal{B}_n$ is the measure on $\mathcal{B}_{2n}$ obtained by first adding the variable $X_{n+1} := n - \sum_{k=1}^n X_k$ (homogenizing) and then polarizing: splitting X_{n+1} into n conditionally exchangeable Boolean variables.

Theorem 7

Symmetric homogenization preserves the strong Rayleigh property.

Example of symmetric homogenization



On the left is a configuration in $\mathcal{B}_9$. Symmetric homogenization extends this, on the right, to a configuration on $\mathcal{B}_{18}$ in which the number of new 1's is the number of old 0's and vice versa.

Further closure properties

Here are five properties that preserve SR, for which the proof is more or less immediate from the full definition of stability.

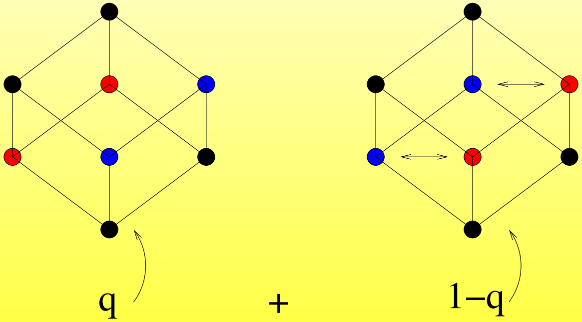
1. Permuting the variables: $F(x_{\pi(1)}, \ldots, x_{\pi(n)}))$ is stable if F is.
2. Merging independent collections: FG is stable if F and G are.
3. Forgetting a variable: setting the **indeterminate** $x_j = 1$. More generally, setting $x_j = a$ for $a \in \mathbb{H}$ preserves stability.
4. Replacing X_1 and X_2 by $X_1 + X_2$: $F(x_1, x_1, x_3, \ldots, x_n)$.
5. Conditioning on X_j: $\frac{\partial F}{\partial x_j}$ and $F - x_j \frac{\partial F}{\partial x_j}$ are stable if F is.

Not so obvious closure properties

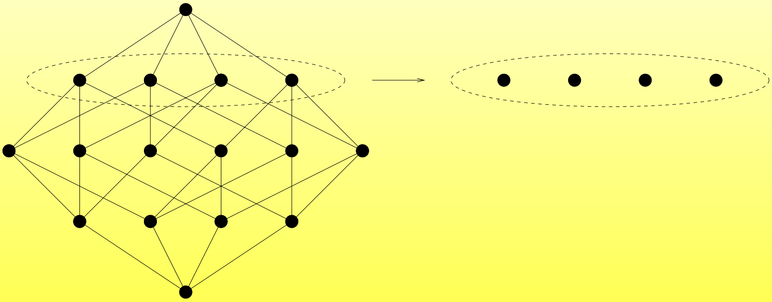
Three more closure properties hold that are probabilistically meaningful, two of which are less automatic.

1. External field: $F(\lambda_1 x_1, \ldots, \lambda_n x_n)$ is stable if F is.
2. Stirring, that is, replacing F by a convex combination of F and $F^{ij} := F(x_1, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n)$.
3. Conditioning on the total, S: $(\mathbb{P} | S = k)$ is SR if $\mathbb{P}$ is.

Stirring



Conditioning on the total



$\mathbf{P}$

$(\mathbf{P} \mid \mathbf{S} = 3)$

Proofs

1. External field. Immediate from the defition of stability: that F be nonvanishing on $\mathbb{H}^n$.

2. Stirring. The nonvanishing of $pF + (1-p)F^{ij}$ may be checked for each fixed set of values of $\{x_k : k \neq i, j\}$ in $\mathbb{H}$. These specializations of F are stable, 2-variable, multi-affine polynomials with complex coefficients. It suffices to check for this class that stability is closed under $F \mapsto pF(x,y) + (1-p)F(y,x)$. This can be done by brute force.

3. Conditioning on the total. Homogenize to obtain the new stable function $G(x_1, \ldots, x_n, y) = \sum_{j=0}^n E_j(x_1, \ldots, x_n) y^j$. Derivatives preserve stability. Differentiating k times with respect to y and $n-k$ times with respect to y^{-1} leaves a constant multiple of E_k.

EXAMPLES OF STRONG RAYLEIGH LAWS

Conditioned Bernoulli sampling

Let $\{\pi_i : 1 \leq i \leq n\}$ be numbers in $[0, 1]$. Let $\mathbb{P}$ be the product measure making $\mathbb{E}X_i = \pi_i$ for each i. Let $\mathbb{P}' = (\mathbb{P}|S = k)$. The measure $\mathbb{P}'$ is called *conditioned Bernoulli sampling*. We already know this is SR because it is a product measure, conditioned on the total. What's interesting is you can always arrange to sample this way if your marginals sum to an integer.

Theorem 8

Given any probabilities $p_1, \ldots, p_n$ summing to k, there is a one-parameter family of vectors $(\pi_1, \ldots, \pi_n)$ whose conditional Bernoulli sampling law has marginals $p_1, \ldots, p_n$.

Further examples

- Exclusion process. Let $\mathbb{P}$ be a strong Rayleigh measure on $\mathcal{B}_n$. Suppose for each i, j, the values of X_i and X_j swap at some prescribed, not necessarily constant rates $\beta_{ij}(t)$. Then for fixed T, the law at time t is strong Rayleigh.
- Pivot sampling. (I won't go into this now.)
- Determinantal measures. To prove these are SR, use the fact that $F = C \det(H - \mathrm{d}iag(x_1, \ldots, x_n))$ where H is positive definite, together with Gårding's (1951) criterion for stability.

Further useful properties of SR measures

1. Let $\mathbb{P}$ be SR and let $S := \sum_{k=1}^{n} X_k$. Then S has a real rooted generating function. In particular, it has the same law as a sum of independent Bernoullis, the sequence $\{\mathbb{P}(S = k) : 0 \leq k \leq n\}$ is ultra-log-concave, and it's mode and mean differ by at most one.
2. **Stochastically increasing levels**: the law $(\mathbb{P}|S = k + 1)$ stochastically dominates the law $(\mathbb{P}|S = k)$.
3. The law of $\mathbb{P}$ conditioned on $S \in \{k, k + 1\}$ is strong Rayleigh.

Rank rescaling

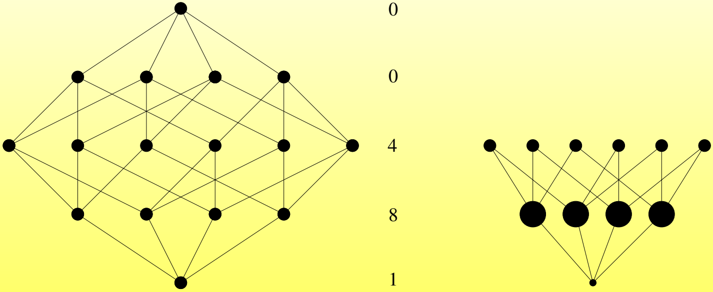
Lemma 9 (rank re-scaling)

Let $\mathbb{P}$ on $\mathcal{B}_n$ be strong Rayleigh and let $\{b_i : 0 \leq i \leq n\}$ be a finite sequence of nonnegative numbers such that $\sum_{i=0}^n b_i x^i$ is stable (equivalently, has only real roots). Then the measure

$$\sum_{i=0}^{n} b_i (\mathbb{P} | S = i)$$

normalized to have total mass 1, is also strong Rayleigh.

Example of rank re-scaling



The sequence $1, 8, 4, 0, 0$ corresponds to the polynomial $1 + 8x + 4x^2$, which has all real roots. A generic measure on $\mathcal{B}_4$ (on the left) becomes a new measure in which ranks 3 and 4 are gone. Points in rank 1 increase in weight by the most, followed by rank 2 and then rank 0. Resulting weights are normalized to sum to 1.

Proof that rank re-scaling preserves SR

PROOF:

1. In the special case $b_i = \delta_{i,k}$, this is just saying that $(\mathbb{P}|S = k)$ is SR, which we alredy proved.

2. In general, because the reversed sequence $\{b_{n-k} : 0 \leq k \leq n\}$ is real rooted, we may construct independent Bernoulli random variables $Y_1, \ldots, Y_n$ whose law Q on $\mathcal{B}_n$ gives $Q(\sum_{j=0}^n Y_j = k) = b_{n-k}$ for all k.

3. The product law $\mathbb{P} \times Q$ is SR (closure under products). By Step (1), the law $(\mathbb{P} \times Q| \sum_{j=0}^{2n} \omega_j = n)$ of the product conditioned on the sum of all the X and Y variables being equal to n is SR as well. Forgetting about the Y variables, this is $\sum_{i=0}^n b_i(\mathbb{P}|S = i)$. $\square$

Cleaning up two arguments from before

Applying the lemma with $b_i := \mathbf{1}_{k \le i \le k+1}$ proves that $(\mathbb{P} | k \le S \le k+1)$ is strong Rayleigh.

To deduce stochastically increasing levels, homogenize the measure $(\mathbb{P} | k \le S \le k+1)$, yielding a SR measure ν. Negative association implies that the homogenizing variable $X_{n+1} := \mathbf{1}_{S=k}$ is ν-negatively correlated with any upward event in $\mathcal{B}_n$. This is the desired conclusion.

Remark: we can't continue and apply the lemma to $b_i := \mathbf{1}_{k \le i \le k+2}$ because $x^k + x^{k+1} + x^{k+2}$ does not have all real roots.

Therefore, $(\mathbb{P} | k \le S \le k+2)$ is NOT in general SR.

A recurring argument

Negative association for the spanning tree measure was first proved by Feder and Mihail (1992). In fact this argument is at the heart of a number of others, so we should be aware, although they state somewhat less, of what their argument showed.

Theorem 10 (Feder and Mihail (1992, Lemma 3.2))

Let $\mathcal{M}$ be a class of probability measures on Boolean lattices that are all homogeneous and pairwise negatively correlated. Suppose $\mathcal{M}$ is closed under conditioning on the value of one of the variables. Then all measures in the class $\mathcal{M}$ are negatively associated.

SR implies NA

With this, we can pay off a debt and prove that SR implies NA.

PROOF THAT STRONG RAYLEIGH MEASURES ARE NEGATIVELY ASSOCIATED:

1. The critical step is that $\mathbb{P}$ can be extended to a homogeneous measure, namely its symmetric homogenization.
2. Observe that SR implies Rayleigh which implies pairwise negative correlation.
3. The class of strong Rayleigh distributions is closed under conditioning. The hypotheses of Feder-Mihail are satisfied, therefore all strong Rayleigh measures are negatively associated.

$\square$

MORE RECENT RESULTS

Stochastic covering property

Say that ν stochastically covers μ if it is stochastically greater and can be coupled so that the sample from ν is either equal to the one from μ or contains precisely one more element.

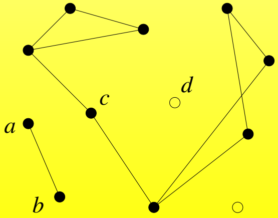
Let μ and ν are the respective conditional measures on $\mathcal{B}_{n-1}$ defined by $\mu = (\mathbb{P} | X_n = 1)$ and $\nu = (\mathbb{P} | X_n = 0)$. A measure is said to have the **SCP** if ν stochastically covers μ, and this holds when $\mathbb{P}$ is replaced by any conditionalization or index permutation.

Proof: stochastic domination follows from negative association. For a homogeneous measure, stochastic covering follows from stochastic domination. In general, $\mathbb{P}$ can be extended to a homogeneous measure (symmetric homogenization), and that's good enough.

Lipschitz functionals

A Lipschitz function $f : \mathcal{B}_n \to \mathbb{R}$ is one that changes by no more than some constant c (without loss of generality $c = 1$) when a single coordinate of $\omega \in \mathcal{B}_n$ changes.

Example: Let $\{1, \ldots, n\}$ index edges of a graph G whose degree is bounded by d. Let Y be a random subgraph of G and let $X_e := \mathbf{1}_{e \in Y}$. Let f count one half the number of isolated vertices of Y. Then f is Lipschitz-1 because adding or removing an edge cannot affect the isolation of an vertex other than an endpoint of e.



Simultaneously generalizing the functional and the measure

Strong tail bounds are available for Lipschitz functions of independent variables. These are based on classical exponential bounds going back to the 50's (Chernoff) and 60's (Hoeffding).

E. Mossel asked about generalizing from sums to Lipschitz functions assuming negative association. This question is still open, but it is true if one assumes the strong Rayleigh property.

Theorem 11 (Pemantle and Peres, 2015)

Let $f : \mathcal{B}_n \to R$ be Lipschitz-1. If $\mathbb{P}$ is k-homogeneous then

$$\mathbb{P}(|f - \mathbb{E}f| \geq a) \leq 2 \exp\left(\frac{-a^2}{8k}\right) .$$

Without the homogeneity assumption, the bound becomes $5 \exp(-a^2/(16(a + 2\mu))$ *where* μ *is the mean.*

Proof

SKETCH OF PROOF:

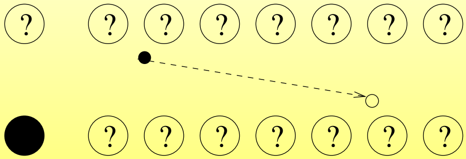
- Strong Rayleigh measures have the stochastic covering property.
- The classical Azuma martingale, $Z_k := \mathbb{E}(f \mid X_1, \ldots, X_k)$ can now be shown to have bounded differences, due to Lipschitz condition on f and coupling of the different conditional laws.

  (See illustration)

Note: this actually proves that any law with the SCP satisfies the same tail bounds for Lipschitz-1 functionals.

Illustration

Look at values of X_i one at a time.



There is a coupling such that the upper row samples from $\mathbb{P}$, the lower row samples from $(\mathbb{P}|X_1 = 1)$, and the only difference is in the X_1 variable and at most one other variable.

A similar picture holds for $(\mathbb{P}|X_1 = 0)$.

Therefore, f varies by at most 2 from the upper to the lower row, hence $|\mathbb{E}f - \mathbb{E}(f|X_1)| \leq 2$.

Application

The proportion of vertices in a uniform spanning tree in $\mathbb{Z}^2$ that are leaves is known to be $8/\pi^2 - 16/\pi^3 \approx 0.2945$. Let us bound from above the probability that a UST in an $N \times N$ box has at least $N^2/3$ leaves.

Letting f count half the number of leaves, we see that f is Lipschitz-1. The law of $\{X_e := \mathbf{1}_{e \in T}\}$ is SR and $N^2 - 1$ homogeneous. Therefore,

$$\mathbb{P}(f - \mathbb{E}f \geq a) \leq 2\exp(-a^2/(8N^2 - 8))\,.$$

The probability of a vertex being a leaf in the UST on a box is bounded above by the probability for the infinite UST. Plugging in $a = N^2(1/3 - 8\pi^{-2} + 16\pi^{-3})$ and replacing the denominator by $8N^2$ therefore gives an upper bound of

$$2\exp\left[\left(\frac{1}{3} - \frac{8}{\pi^2} + \frac{16}{\pi^3}\right)^2 N^2\right] \approx 2e^{-0.0015N^2}\,.$$

A CLT based on geometry of zeros

Distributions on $\{0, \ldots, n\}$ whose generating function F has all real roots obey a CLT, provided their variance goes to infinity. This is because such a distribution is the law of a sum of independent Bernoulli random variables.

In fact, if the roots are all in the left half-plane, then F is the product of trinomials with nonnegative coefficients, hence the distribution is the sum of independent $\{0, 1, 2\}$-valued random variables, hence the CLT holds again. Similar arguments improve this to requiring F avoid any region avoiding a small sector near the positive real axis.

Theorem (Lebowitz, Pittel, Ruelle and Speer (2016))

If probability distributions $\mathbb{P}_n$ on $\{0, \ldots, n\}$ have generating functions F_n with no zeros in a δ-neighborhood of 1, and the variances grow faster than $n^{2/3}$, then $\{\mathbb{P}_n\}$ satisfies a central limit theorem.

Best 1-avoiding CLT

The hypothesis that $n^{-1/3}\sigma_n \to \infty$ cannot be removed using the technology of [LPRS]. Their proof does not make essential use of the nonnegativity of the coefficients of F.

A natural conjecture is that it can be replaced by $\sigma_n \to \infty$, but this is false as was demonstrated by Michelen and Saharasbudhe (Advances, 2019). Recently, these two proved a quantitative result that appears to be best possible.

Theorem (best geometric CLT)

There is a universal constant C such that if a R.V. X on $\{0, \ldots, n\}$ has generating function F with no zeros within δ of 1, then the self-normalized variable $X_* := (X - \mu)/\sigma$ differs in the CDF sup norm from a standard normal by at most $C \log n/(\delta\sigma)$. This is best possible because for any fixed $C > 0$ there are $\delta > 0$ and $\{X_n\}$ supported on $\{0, \ldots, n\}$ with F_n nonvanishing in a δ-ball around 1, $\sigma_n > C \log n$, and $(X_n - \mu_n)/\sigma_n$ not converging to a normal.

Lorentzian distributions

Another motive for pursuing negative dependence properties has been to settle the conjecture that the random cluster model has negative correlations for $q < 1$.

The hope was to show the RC model was negatively associated by showing it was in a class of models having the NA property.

Unfortunately no one yet knows whether the RC model is strong Rayleigh.

It is, however, in a newly defined class of Lorentzian distributions, which contains the strong Rayleigh distributions.

Lorentzian implies 2-NA, but sadly, not NA

Lorentzian disgtributions were defined in a recent preprint of Brändén and Huh. Their generating functions are homogeneous polynomials of any degree d in the closure of the set of polynomials such that, when differentiated $d-2$ times with respect to any variables, they result in a homogeneous quadratic with Lorentzian signature.

Unfortunately, being Lorentzian only implies NA up to a factor of 2.

This is new information for the RC model: $\mathbb{E}X_iX_j \leq 2\mathbb{E}X_i\mathbb{E}X_j$ for any two edge indicators. It may help settle some open problems on cluster size, even though it is not the big conjecture.

OPEN PROBLEMS

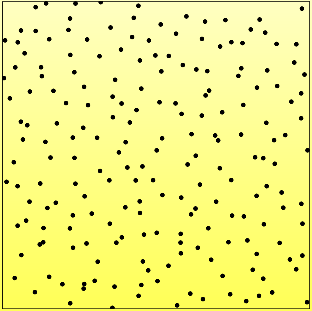
Determinantal point processes

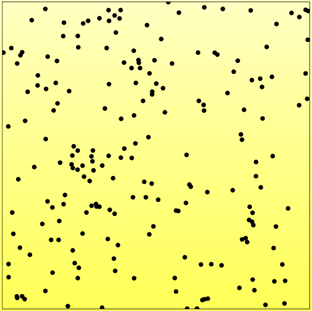
This topic and associated problem goes a bit beyond the realm of measures on finite Boolean lattices.

A point process in $\mathbb{R}^n$ is said to be determinantal if the is a Hermitian kernel $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ such that the joint density of points at $x_1, \ldots, x_k$ is equal to the determinant of the matrix $K(x_i, x_j)_{i,j=1}^n$.

Determinantal point processes satisfy analogous negative dependence properties to discrete negatively dependent collections.

Visual depiction of negative dependence for DPP





Which one is the Poisson?

Open problem # 1: SR in continuous space?

The text by Kulesza and Taskar (2012) on determinantal point processes for machine learning makes no distinction between discrete and continuous space. This makes sense in Computer Science, where continuous space is represented in the end as discrete space (pixels, floating point numbers, etc.) with a very fine mesh.

Mathematically though, we have no theory of the strong Rayleigh property for a point process on a continuous space.

Problem 1

What is the right definition of strong Rayleigh for point processes, and what can be deduced from it? [Nonstandard analysis?]

Open problem #2: random acyclic graphs

Related to the spanning tree measure:

Problem 2

Does the uniform (or weighted) random acyclic subgraph possess any negative dependence properties? Is is conjectured to be NC, and this is still open. If it is, one would imagine it is because it is SR, and this has not been refuted.

Again, we can get this up to a factor of 2 via the new technology.

Theorem 12 (Huh-Brändén)

The uniform measure on random acyclic subgraphs is Lorentzian.

THE END