Lecture 3: Boolean variables and the strong Rayleigh property

Robin Pemantle

University of Pennsylvania

pemantle@math.upenn.edu

Minerva Lectures at Columbia University

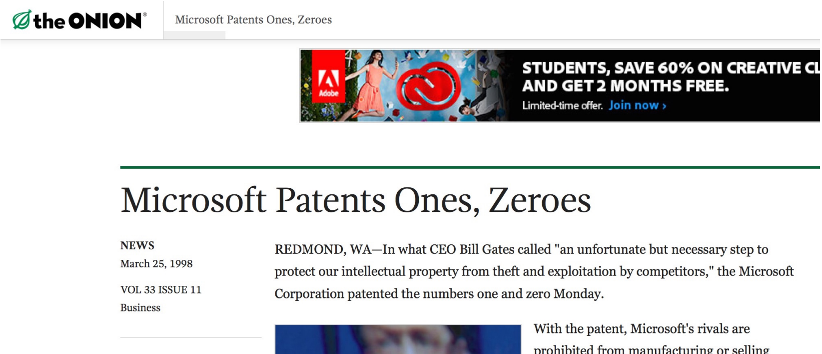
09 November, 2016

Negative dependence for Boolean variables

Binary variables

In this lecture we are interested in **binary variables**, that is, variables taking only the values 0 and 1.

It comes as no surprise that these are of great interest.



the ONION® | Microsoft Patents Ones, Zeroes

Microsoft Patents Ones, Zeroes

NEWS
March 25, 1998

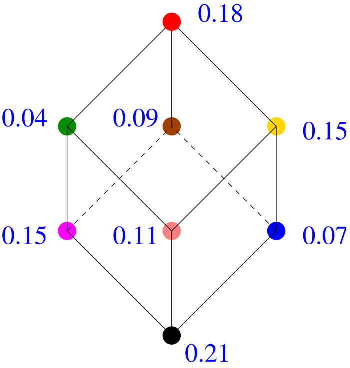
VOL 33 ISSUE 11
Business

REDMOND, WA—In what CEO Bill Gates called "an unfortunate but necessary step to protect our intellectual property from theft and exploitation by competitors," the Microsoft Corporation patented the numbers one and zero Monday.

With the patent, Microsoft's rivals are prohibited from manufacturing or selling

Boolean lattice

A collection of n binary variables can be thought of as a random point in the Boolean lattice $\mathcal{B}_n$ of rank n. The joint law of n binary variables is therefore a probability distribution on this lattice which we can depict by attaching probabilities to the nodes of the lattice.



A probability distribution on the Boolean lattice $\mathcal{B}_3$

Some notation

Let $\mathcal{B}_n := \{0,1\}^n$ denote the Boolean lattice of rank n. The joint law of n binary random variables is a measure μ on $\mathcal{B}_n$. The probability generating function $f = f_\mu$ is given by

$$\mathbf{f}_\mu(\mathbf{x}_1, \ldots, \mathbf{x}_\mathbf{n}) := \sum_{\omega \in \mathcal{B}_\mathbf{n}} \mu(\omega) \prod_{\mathbf{j}=1}^{\mathbf{n}} \mathbf{x}_\mathbf{j}^{\omega_\mathbf{j}} = \mathbb{E}\mathbf{x}^\omega .$$

Substituting $\mathbf{x}_\mathbf{j} = 1$ projects onto a smaller Boolean lattice that "forgets" $\mathbf{x}_\mathbf{j}$.

Substituting $\mathbf{x}_\mathbf{j} = 0$ or taking $(\partial/\partial \mathbf{x}_\mathbf{j})|_{\mathbf{x}_\mathbf{j}=0}$ conditions on $\mathbf{x}_\mathbf{j} = 0$ or 1 respectively.

Negative correlation

Negative correlation between variables $\mathbf{x_i}$ and $\mathbf{x_j}$ can be tested by the following generating function inequality:

$$\mathbf{F}(\mathbf{1})\frac{\partial^2 \mathbf{F}}{\partial \mathbf{x_i}\mathbf{x_j}}(\mathbf{1}) \leq \frac{\partial \mathbf{F}}{\partial \mathbf{x_i}}(\mathbf{1})\frac{\partial \mathbf{F}}{\partial \mathbf{x_j}}(\mathbf{1})\,.$$

I will discuss a hierarchy of negative dependence conditions on $\mathbb{P}$, the strongest of which is called the **strong Rayleigh** property.

Why do we care about these fancy properties?

1. Stronger properties give stronger conclusions.

2. Sometimes to prove a weaker property the only way is to establish a stronger property which is somehow better behaved.

Story of positive and negative dependence

Before defining strong Rayleigh, let's review some of the most natural and most studied positive and negative dependence conditions.

Positive association

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

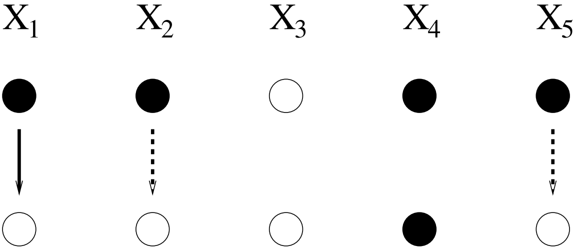
$$\mathbb{E}\mathbf{fg} \geq (\mathbb{E}\mathbf{f})\ (\mathbb{E}\mathbf{g})$$

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

Take $\mathbf{f} = \mathbf{X}_1$ and let $\mathbb{P}_1$ and $\mathbb{P}_0$ denote the conditional distribution of $\mathbb{P}$ given $\mathbf{X}_1 = 1$ and $\mathbf{X}_1 = 0$ respectively. In this case positive association says $\int \mathbf{g}\, \mathbf{d}\mathbb{P}_1 \geq \int \mathbf{g}\, \mathbf{d}\mathbb{P}_0$ for all increasing functions $\mathbf{g}$. We say that $\mathbf{P}_1$ **stochastically dominates** $\mathbb{P}_0$ and write $\mathbb{P}_1 \succeq \mathbb{P}_0$.

$\mathbb{P}_1 \succeq \mathbb{P}_0$ if and only if you can couple them so that the sample from $\mathbb{P}_0$ is obtained from the $\mathbb{P}_1$ sample by changing some ones to zeros (or doing nothing).

Coupling for positive association



One can sample simultaneously from $(\mathbb{P}|\mathbf{X}_1 = 1)$ and $(\mathbb{P}|\mathbf{X}_1 = 0)$ in such a way that turning off the bit at $\mathbf{X}_1$ also turns off some of the other bits (in this case $\mathbf{X}_2$ and $\mathbf{X}_5$).

Negative association

Negative association is a trickier business because $\mathbf{f}$ can't be negatively correlated with itself.

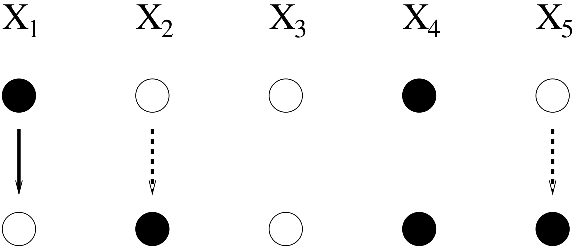
The measure $\mathbb{P}$ on $\mathcal{B}_{\mathbf{n}}$ is **negatively associated** if

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

whenever $\mathbf{f}$ and $\mathbf{g}$ are both monotone increasing and they depend on disjoint sets of coordinates.

Taking $\mathbf{f} = \mathbf{X}_1$, the consequence is that the conditional law of the remaining variables given $\mathbf{X}_1 = 0$ stochastically dominates the law given $\mathbf{X}_1 = 1$. Thus a sample conditioned on $\mathbf{X}_1 = 1$ is obtained from one conditioned on $\mathbf{X}_1 = 0$ by turning some ones into zeros, except the first coordinate, which goes from zero to one.

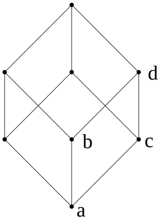
Coupling for negative association



This time, turning off the bit $\mathbf{X}_1$ causes the sample from $(\mathbb{P}|\mathbf{X}_1 = 1)$ to **gain** some ones when it turns into a sample from $(\mathbb{P}|\mathbf{X}_1 = 0)$.

Lattice conditions

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



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

FKG

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

Theorem (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 lattice condition

Unfortunately, the FKG theorem does not hold when the positive lattice condition is replaced by the negative lattice condition.

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

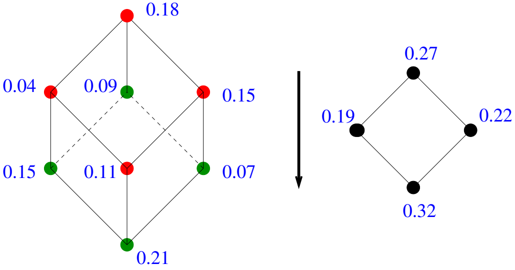
A profusion of properties has been suggested that are somewhat weaker than NA. These are not totally ordered with respect to implication. Many concern the stochastic domination of some conditional distribution of $\mathbb{P}$ by others. The litany is long, including many ultimately failed concepts introduced in [Pem00].

The next slides define some of these properties and describe the implications that hold between them.

Hereditary properties

Definition (Hereditary properties)

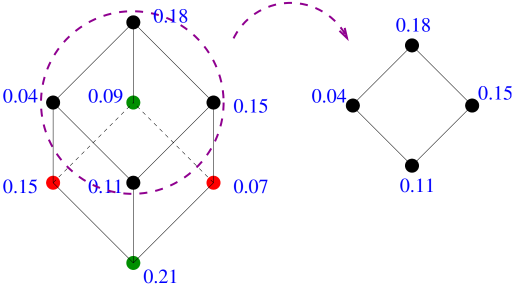
Given a property of the atom sizes in $\mathcal{B}_{\mathbf{n}}$, such as the negative lattice condition, we prepend **h** *to denote that it should hold hereditarily, that is for every projection onto a subset of the variables (forgetting, the others, that is, integrating them out).*



Conditional properties

Definition (Conditional properties)

For properties defined in terms of the random variables, prepend **C** *to say that they hold for sublattices, that is, when conditioning on fixed values of some of the variables.*

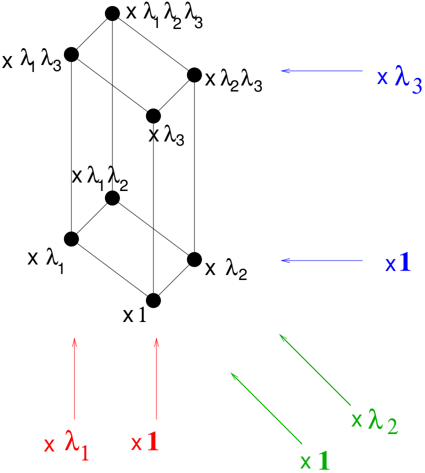


External fields

The superscript $+$ denotes a property continuing under impostion of an **external field**:
the weights $\{\mu(\omega)\}$ are replaced by

$$\mathbf{C}\,\mu(\omega)\prod_{\mathbf{i}=1}^{\mathbf{n}}\lambda_{\mathbf{i}}^{\omega(\mathbf{i})}$$

for $\lambda_1,\ldots,\lambda_{\mathbf{n}}>0$.

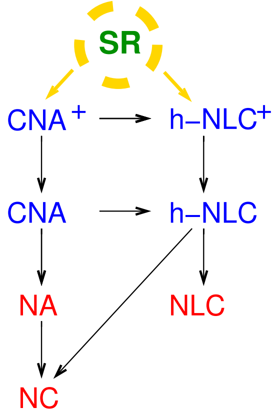


Negative dependence hierarchy

$$
\begin{tikzcd}
\mathrm{CNA}^{+} \arrow[r] \arrow[d] & \text{h–NLC}^{+} \arrow[d] \\
\mathrm{CNA} \arrow[r] \arrow[d] & \text{h–NLC} \arrow[d] \arrow[ddl] \\
\mathrm{NA} \arrow[d] & \mathrm{NLC} \\
\mathrm{NC} &
\end{tikzcd}
$$

CNA^{+}	CNA under external fields
h-NLC^{+}	Hereditary CNA^{+}
CNA	Conditional NA
h-NLC	Hereditary NLC
NA	Negative association
NLC	Negative lattice condition
NC	Negative correlation

Negative dependence hierarchy



SR	Strong Rayleigh
CNA$^+$	CNA under external fields
h-NLC$^+$	Hereditary CNA$^+$
CNA	Conditional NA
h-NLC	Hereditary NLC
NA	Negative association
NLC	Negative lattice condition
NC	Negative correlation

The half-plane property: real stable polynomials of several variables

Multivariate stability

Definition (stable)

A real or complex polynomial $\mathbf{q}$ *in* $\mathbf{d}$ *variables is said to be* **stable** *if* $\mathbf{q}(\mathbf{z}_1, \ldots, \mathbf{z_d}) = 0$ *implies not all coordinates* $\mathbf{z_j}$ *are in the open upper half plane.*

In one variable, a real polynomial $\mathbf{f}$ is stable if and only $\mathbf{f} \in \mathbf{RR}$. In more than one variable it is a lot more complicated.

Example (bilinear functions)

The real polynomial $\mathbf{a} + \mathbf{bx} + \mathbf{cy} + \mathbf{dxy}$ is stable if and only if $\mathbf{ad} \leq \mathbf{bc}$.

Easy properties

Proposition (easy closure properties)

The class of stable polynomials is closed under the following.

(a) *Products:* $\mathbf{f}$ *and* $\mathbf{g}$ *are stable implies* $\mathbf{fg}$ *is stable;*

(b) *Index permutations:* $\mathbf{f}$ *is stable implies* $\mathbf{f}(\mathbf{x}_{\pi(1)}, \ldots, \mathbf{x}_{\pi(\mathbf{d})})$ *is stable where* $\pi \in \mathcal{S}_{\mathbf{d}}$*;*

(c) *Diagonalization:* $\mathbf{f}$ *is stable implies* $\mathbf{f}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_3, \ldots, \mathbf{x}_{\mathbf{d}})$ *is stable;*

(d) *Specialization: if* $\mathbf{f}$ *is stable and* $\Im(\mathbf{a}) \geq 0$ *then* $\mathbf{f}(\mathbf{a}, \mathbf{x}_2, \ldots, \mathbf{x}_{\mathbf{d}})$ *is stable;*

(e) *Inversion: if the degree of* $\mathbf{x}_1$ *in* $\mathbf{f}$ *is* $\mathbf{m}$ *and* $\mathbf{f}$ *is stable then* $\mathbf{x}_1^{\mathbf{m}} \mathbf{f}(-1/\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\mathbf{d}})$ *is stable;*

$\square$

Differentiation

Lemma (differentiation)

If $\mathbf{f}$ is stable then $\partial \mathbf{f} / \partial \mathbf{x_j}$ is either stable or identically zero.

PROOF: Fix any values of $\{\mathbf{x_i} : \mathbf{i} \neq \mathbf{j}\}$ in the upper half plane. As a function of $\mathbf{x_j}$, $\mathbf{f}$ has no zeros in the upper half plane. By the Gauss-Lucas theorem, the zeros of $\mathbf{f}'$ are in the convex hull, therefore not in the upper half plane. $\square$

The next property, Wagner calls an "astounding" recent generalization of the Pólya-Schur theorem.

Multivariate Pólya-Schur theorem

This characterizes not just multiplier sequence but all $\mathbb{C}$-linear maps preserving stability. To restrict to multiplier sequences, take $\mathbf{T}(\mathbf{x}^\alpha) = \lambda(\alpha)\mathbf{x}^\alpha$.

Theorem ([BB09b, Theorem 1.3])

The $\mathbb{C}$-linear map $\mathbf{T} : \mathbb{C}[\mathbf{x}] \to \mathbb{C}[\mathbf{x}]$ preserves stable polynomials if and only if either its range is scalar multiples of a single stable polynomial or the series

$$\sum_{\alpha \in (\mathbb{Z}^+)^{\mathbf{d}}} (-1)^{|\alpha|} \mathbf{T}(\mathbf{x}^\alpha) \frac{\mathbf{y}^\alpha}{\alpha!}$$

is a uniform limit on compact sets of stable polynomials in $\mathbb{C}[\mathbf{x}, \mathbf{y}]$.

Sketch of proof:

The crucial step is to show that the closure of the stable polynomials in $\mathbb{C}[\mathbf{x}][[\mathbf{y}]]$ is characterized by requiring stability of a collection of initial segments.

> *A power series* $\sum_\alpha \mathbf{P}_\alpha(\mathbf{x})\mathbf{y}^\alpha$ *whose coefficients are polynomials in* $\mathbf{x}$ *is in the closure of stable polynomials in* $\mathbb{C}[\mathbf{x},\mathbf{y}]$ *if and only if for all* $\beta \in (\mathbb{Z}^+)^{\mathbf{d}}$,
>
> $$\sum_{\alpha \le \beta} (\beta)_\alpha \mathbf{P}_\alpha(\mathbf{x})\mathbf{y}^\alpha$$
>
> *is stable in* $\mathbb{C}[\mathbf{x},\mathbf{y}]$.

Similar techniques as in the univariate case then establish that the $\mathbb{C}$-linear operator defined by such a series preserves stability. □

Multi-affine case

Definition (Strong Rayleigh)

A probability law on $\mathcal{B}_n$ is said to be strong Rayleigh *if its multivariate generating function is stable.*

Generating polynomials for the joint law of binary variables are **multi-affine**, meaning that no individual variable has degree greater than 1.

Theorem (multi-affine equivalence)

If $\mathbf{F}$ *is multi-affine then* $\mathbf{F}$ *is stable if and only if*

$$\mathbf{F}(\mathbf{x})\frac{\partial^2 \mathbf{F}}{\partial \mathbf{x_i x_j}}(\mathbf{x}) \leq \frac{\partial \mathbf{F}}{\partial \mathbf{x_i}}(\mathbf{x})\frac{\partial \mathbf{F}}{\partial \mathbf{x_j}}(\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$.

Rayleigh versus strong Rayleigh

An external field replaces $\mathbf{F}(\mathbf{1})$ and its derivatives by $\mathbf{F}(\lambda)$ and its derivatives. Therefore the property h-NLC$^+$ becomes the Rayleigh property

$$\mathbf{F}(\mathbf{x})\frac{\partial^2\mathbf{F}}{\partial \mathbf{x_i}\mathbf{x_j}}(\mathbf{x}) \leq \frac{\partial \mathbf{F}}{\partial \mathbf{x_i}}(\mathbf{x})\frac{\partial \mathbf{F}}{\partial \mathbf{x_j}}(\mathbf{x}) \text{ for all } \mathbf{x} \in (\mathbb{R}^+)^{\mathbf{n}}\,. \qquad (1)$$

The only difference between Rayleigh and **strong Rayleigh** is that strong Rayleigh requires (1) for negative external fields!

Extending positive properties to real properties and real properties to complex properties turns out to be the key to obtaining strong closure properties.

Closure properties of the strong Rayleigh class

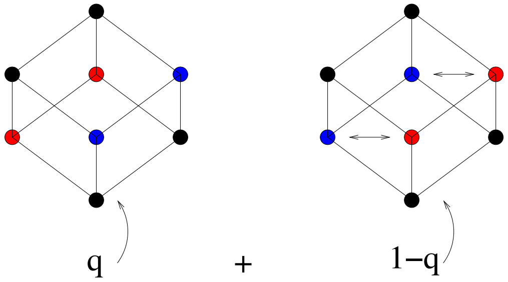
Elementary closure properties

Elementary closure properties of the class of stable polynomials translate into the following elementary closure properties of the class of strong Rayleigh distributions.

1. Permuting the variables: $\mathbf{F}(\mathbf{x}_{\pi(1)}, \ldots, \mathbf{x}_{\pi(\mathbf{n})}))$ is stable if $\mathbf{F}$ is.
2. Merging independent collections: $\mathbf{FG}$ is stable if $\mathbf{F}$ and $\mathbf{G}$ are.
3. Conditioning on $\mathbf{X_j}$: $\dfrac{\partial \mathbf{F}}{\partial \mathbf{x_j}}$ and $\mathbf{F} - \mathbf{x_j}\dfrac{\partial \mathbf{F}}{\partial \mathbf{x_j}}$ are stable if $\mathbf{F}$ is.
4. Forgetting a variable: setting the **indeterminate** $\mathbf{x_j}$ (as opposed to the value of $\mathbf{X_j}$) equal to 1.
5. Replacing $\mathbf{X}_1$ and $\mathbf{X}_2$ by $\mathbf{X}_1 + \mathbf{X}_2$: $\mathbf{F}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_3, \ldots, \mathbf{x_n})$.
6. Reversing a variable: replacing $\mathbf{X}$ by $\mathbf{m} - \mathbf{X}$ where $\mathbf{m}$ is an upper bound for $\mathbf{X}$.
7. Exernal field: $\mathbf{F}(\lambda_1 \mathbf{x}_1, \ldots, \lambda_\mathbf{n} \mathbf{x_n})/\mathbf{F}(\lambda)$ is stable if $\mathbf{F}$ is.

Less elementary: Stirring

8. Stirring: replacing $\mathbf{F}$ by a convex combination of $\mathbf{F}$ and $\mathbf{F^{ij}} := \mathbf{F}(\mathbf{x}_1, \dots, \mathbf{x_{i-1}}, \mathbf{x_j}, \mathbf{x_{i+1}}, \dots, \mathbf{x_{j-1}}, \mathbf{x_i}, \mathbf{x_{j+1}}, \dots, \mathbf{x_n})$.



Proof of stirring property

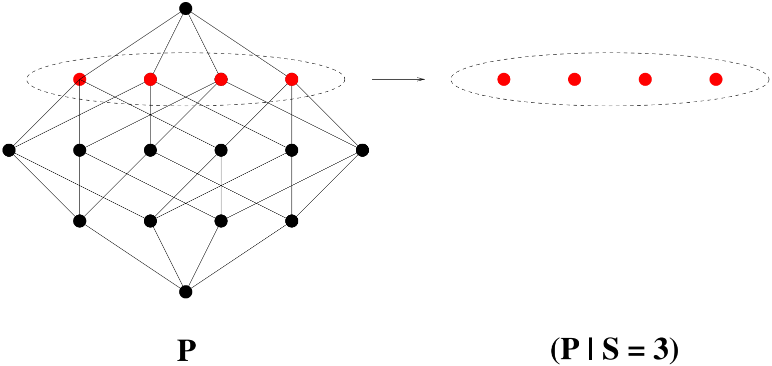
To prove stirring, observe that the nonvanishing of $\mathbf{pF} + (1 - \mathbf{p})\mathbf{F}^{\mathbf{ij}}$ on the upper half-plane $\mathbb{H}^{\mathbf{n}}$ may be checked by checking that for each fixed set of values of $\{\mathbf{x_k} : \mathbf{k} \neq \mathbf{i}, \mathbf{j}\} \in \mathbb{H}^{\mathbf{n}-2}$, the resulting bivariate polynomial is non-vanishing on $\mathbb{H}^2$.

These specializations of $\mathbf{F}$ are stable, 2-variable, multi-affine polynomials with complex coefficients: $\alpha + \beta\mathbf{x} + \gamma\mathbf{y} + \delta\mathbf{xy}$.

It suffices to check for this class that stability is closed under $\mathbf{F} \mapsto \mathbf{pF}(\mathbf{x}, \mathbf{y}) + (1 - \mathbf{p})\mathbf{F}(\mathbf{y}, \mathbf{x})$, which can be done by brute force.

Less elementary: conditioning on the total

9. Conditioning on the total, $\mathbf{S}$: $(\mathbb{P}|\mathbf{S} = \mathbf{k})$ is SR if $\mathbb{P}$ is.

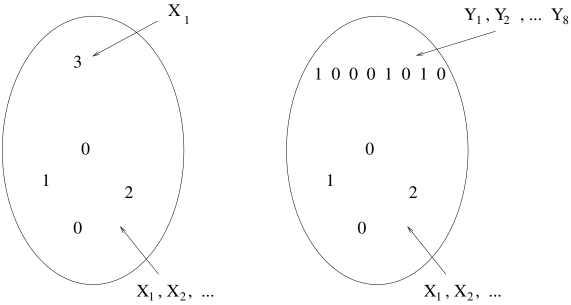


Polarization and homogenization

The proof requires two constructions which allow us to go back and forth between integer and binary variables.

Definition (Polarization)

Let $\mathbf{X}_1, \ldots, \mathbf{X_n}$ *be nonnegative integer random variables, all bounded by* $\mathbf{M}$*. Polarization means replacing* $\mathbf{X}_1$ *by Boolean variables* $\{\mathbf{Y}_1, \ldots, \mathbf{Y_M}\}$ *such that, conditional on* $\mathbf{X}_1, \ldots, \mathbf{X_n}$*, the* $\mathbf{Y}$ *variables are exchangeable and sum to* $\mathbf{X}_1$*.*



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

On the right, given that $\mathbf{X}_1 = 3$, this variable was replaced by 8 binary variables, three of which were chosen to be 1, uniformly among the $\binom{8}{3}$ possibilities.

Polarization

Lemma (Polarization preserves stability)

If the generating function for $\mathbf{X}_1, \ldots, \mathbf{X}_\mathbf{n}$ *is stable then the generating function for the polarization,* $\mathbf{Y}_1, \ldots, \mathbf{Y}_\mathbf{M}, \mathbf{X}_2, \ldots, \mathbf{X}_\mathbf{n}$ *is stable.*

Hint of proof: The polarization construction can be described in terms of functions of a complex variable, without reference to probability. The complex analytic statement is proved via the Grace-Welsh-Szegö Theorem.

Homogenization

Often algebra works better with homogeneous polynomials. Probabilistically, a generating function $\mathbf{F}$ is homogeneous if and only if the random variable $\mathbf{S} := \sum_{\mathbf{k}=1}^{\mathbf{n}} \mathbf{X}_{\mathbf{k}}$ is constant.

Lemma (Homogenization preserves stability)

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

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

Probabilistic interpretaion: if $\{\mathbf{X}_1, \ldots, \mathbf{X}_{\mathbf{n}}\}$ have stable generating function then adding $\mathbf{X}_{\mathbf{n}+1} := \mathbf{n} - \sum_{\mathbf{k}=1}^{\mathbf{n}} \mathbf{S}_{\mathbf{k}}$ preserves stability.

Symmetric homogenization

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

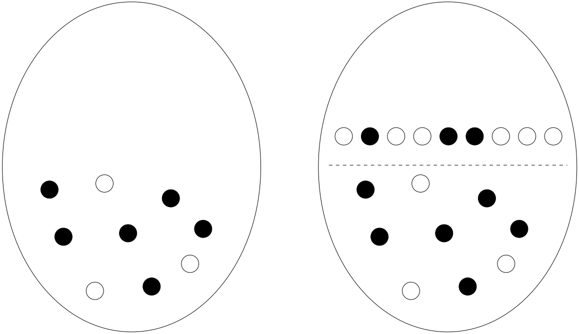
Definition (symmetric homogenization)

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

Theorem

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.

Proof that conditioning on the total preserves SR

Homogenize to obtain the new stable function $\mathbf{G}(\mathbf{x}_1, \ldots, \mathbf{x_n}, \mathbf{y}) = \sum_{\mathbf{j}=0}^{\mathbf{n}} \mathbf{E_j}(\mathbf{x}_1, \ldots, \mathbf{x_n})\mathbf{y^j}$.

Differentiate $\mathbf{k}$ times with respect to $\mathbf{y}$ and $\mathbf{n}-\mathbf{k}$ times with respect to $\mathbf{y}^{-1}$ to extract the $\mathbf{y^{n-k}}$ coefficient of $\mathbf{G}$.

Stability is preserved under differentiation; the $\mathbf{y^{n-k}}$ coefficient is $\mathbf{E_k}$, the generating function for $(\mathbb{P} \,|\, \mathbf{S} = \mathbf{k})$. $\square$

Consequences of the strong Rayleigh property

Negative association

The most important consequence is proved in the seminal paper of Borcea, Brändén and Liggett [BBL09]:

Theorem (SR implies NA)

Strong Rayleigh measures are negative associated.

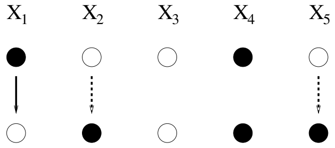
Further useful consequences are as follows.

Definition (Stochastic covering)

A measure μ on $\mathcal{B}_{\mathbf{n}}$ is said to stochastically cover another measure ν on $\mathcal{B}_{\mathbf{n}}$ if $\nu \preceq \mu$ in such a way that the coupling can always be accomplished by changing at most one bit from 0 to 1.

Example (continued)

In a previous example the conditional law $(\mathbb{P} \mid \mathbf{X}_1 = 0)$ stochastically dominated $(\mathbb{P} \mid \mathbf{X}_1 = 1)$. This does not exhibit the stochastic covering relation (though it is possible that SCP holds, witnessed by a different coupling).



Consequences of the strong Rayleigh property

Theorem

Suppose $\mu \in \mathrm{SR}$*, let* $\mathbf{X_i}$ *be the coordinate functions and let* $\mathbf{S} := \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{X_i}$ *be the sum. Let* $\mathbf{p_i} := \mathbb{P}(\mathbf{S} = \mathbf{i})$ *and let* $\mu_\mathbf{i}$ *be the conditional law* $(\mu \,|\, \mathbf{S} = \mathbf{i})$*.*

- **(i)** *ULC: The sequence* $\{\mathbf{p_i}\}$ *is ultra-log-concave*
- **(ii)** *Stochastic increase: The sequence* $\{\mu_\mathbf{k}\}$ *is stochastically increasing.*
- **(iii)** *SCP: The conditional law* $(\mu \,|\, \mathbf{X_i} = 0)$*, restricted to the variables other than* $\mathbf{X_i}$*, stochastically covers the restriction of* $(\mu \,|\, \mathbf{X_i} = \mathbf{k} + 1)$*.*

In the remainder of the lecture (or possibly first thing after the break) I will sketch the proofs of these four results.

Rank-rescaling

First a lemma which follows from the same argument used to prove closure of SR under conditioning on the total.

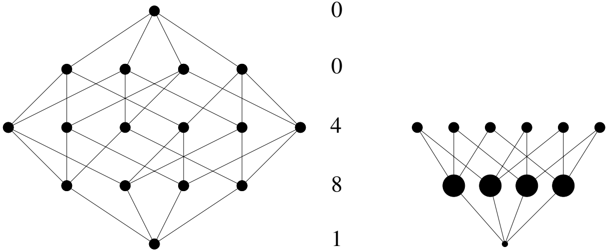
Lemma (rank re-scaling)

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

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

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 $1 + 8\mathbf{x} + 4\mathbf{x}^2$, 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 $\mathbf{b_i} = \delta_{\mathbf{i},\mathbf{k}}$, this is just saying that $(\mathbb{P}|\mathbf{S} = \mathbf{k})$ is SR, which we alredy proved.

2. In general, because the reversed sequence $\{\mathbf{b_{n-k}} : 0 \leq \mathbf{k} \leq \mathbf{n}\}$ is real rooted, we may construct independent Bernoulli random variables $\mathbf{Y}_1, \ldots, \mathbf{Y_n}$ whose law $\mathbf{Q}$ on $\mathcal{B}_\mathbf{n}$ gives $\mathbf{Q}(\sum_{\mathbf{j}=0}^{\mathbf{n}} \mathbf{Y_j} = \mathbf{k}) = \mathbf{b_{n-k}}$ for all $\mathbf{k}$.

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

Example: special case of rank re-scaling

Example (two consecutive levels)

Rank rescaling by the binomial $\mathbf{x}^{\mathbf{k}} + \mathbf{x}^{\mathbf{k}+1}$ shows that a strong Rayleigh distribution restricted to two consecutive levels is still strong Rayleigh.

Note that $\mathbf{x}^{\mathbf{k}} + \mathbf{x}^{\mathbf{k}+1} + \mathbf{x}^{\mathbf{k}+2}$ is not stable. In general, the restriction of an SR measure to three or more consecutive levels is not SR.

Proof that the sum of SR variables is ULC

Substituting $\mathbf{x_1} = \ldots = \mathbf{x_n} = \mathbf{x}$ preserves the strong Rayleigh property. Thus the total $\mathbf{S} := \sum_{\mathbf{k}=1}^{\mathbf{n}} \mathbf{X_k}$ is univariate stable, i.e., real-rooted.

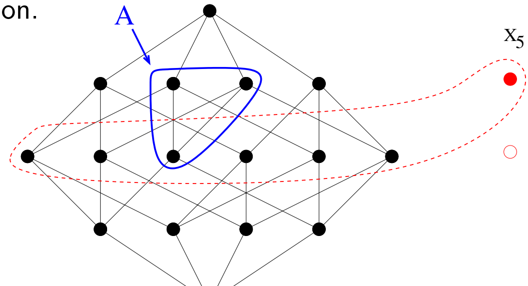
ULC then follows from Newton's inequalities. $\square$

Proof of stochastically increasing levels

Step 1: Restrict to levels $\mathbf{k}$ and $\mathbf{k}+1$ (special case of rank rescaling).

Step 2: Homogenize the measure $(\mathbb{P}|\mathbf{k} \leq \mathbf{S} \leq \mathbf{k}+1)$, yielding a SR measure ν.

Step 3: Negative association (to be proved shortly) implies that the homogenizing variable $\mathbf{X_{n+1}} := \mathbf{1_{S=k}}$ is ν-negatively correlated with any upward event in $\mathcal{B}_\mathbf{n}$. This is the desired conclusion.



Proof of SCP

Step 1: Homogenize $\mathbb{P}$ by adding variables $\mathbf{Y}$ so that the total $\sum_{\mathbf{i}=1}^{\mathbf{n}-\mathbf{k}} \mathbf{Y_i} + \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{X_i}$ is always equal to $\mathbf{n}$. Call the augmented law $\mathbb{P}'$.

Step 2: Let μ and ν are the respective conditional measures on $\mathcal{B}_{\mathbf{n}-1} \times \mathbb{Z}^+$ defined by $\mu = (\mathbb{P}' | \mathbf{X_n} = 1)$ and $\nu = (\mathbb{P}' | \mathbf{X_n} = 0)$.

Step 3: Negative association (to be proved shortly) implies that $\mu \preceq \nu$.

Step 4: Homegenization implies that any coupling of μ and ν witnessing $\mu \preceq \nu$ has precisely one variable increasing from 0 to 1. Restricting to $\mathcal{B}_{\mathbf{n}-1}$ gives a coupling in which at most one variable increases. $\square$

Feder-Mihail lemma

The argument that strong Rayleigh measures are negatively associated requires almost no modification from Feder and Mihail's original proof that spanning tree measures are negatively associated.

Theorem ([FM92, 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.

Proof that SR implies NA

1. Extend $\mathbb{P}$ to $\mathbb{P}'$, the symmetric homogenization.
2. SR implies (ordinary) Rayleigh which implies pairwise negative correlation.
3. The class of homogeneous strong Rayleigh distributions is closed under conditioning. The hypotheses of Feder-Mihail are satisfied, therefore all strong Rayleigh measures are negatively associated.

$\square$

End of Lecture 3

References I

J. Borcea and P. Brändén.
Applications of stable polynomials to mixed determinants.
Duke J. Math., 143:205–223, 2008.

J. Borcea and P. Brändén.
The Lee-Yang and Pólya-Schur programs, I: Linear operators preserving stability.
Invent. Math., 177:541–569, 2009.

J. Borcea and P. Brändén.
The Lee-Yang and Pólya-Schur programs, II: Theory of stable polynomials and applications.
Comm. Pure Appl. Math., 62:1595–1631, 2009.

J. Borcea, P. Brändén, and T. Liggett.
Negative dependence and the geometry of polynomials.
J. AMS, 22:521–567, 2009.

T. Feder and M. Mihail.
Balanced matroids.
Annual ACM Symposium on Theory of Computing, pages 26–38, 1992.

References II

R. Pemantle.
Toward a theory of negative dependence.
J. Math. Phys., 41:1371–1390, 2000.

G. Pólya and J. Schur.
Über zwei Arten von Faktorenfolgen in der Theorie der algebraischen Gleichungen.
J. Reine Angew. Math., 144:89–113, 1914.

D. G. Wagner.
Negatively correlated random variables and Mason's conjecture for independent sets in matroids.
Ann. Comb., 12:211–239, 2008.