Negatively dependent Boolean variables: concepts and problems

Robin Pemantle

UW–MSR Summer Institute, Alderbrook, 24 August 2015

Story

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

$\mathcal{B}_n$ is a Boolean lattice of rank n

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

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

In case anyone here is unclear on why we are studying collections of zero-one valued variables, here is a word from our sponsor.

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 products containing zeroes and ones—the mathematical building blocks of all computer languages and programs—unless a royalty fee of 10 cents per digit used is paid to the software giant.

"Microsoft has been using the binary system of ones and zeroes ever since its inception in 1975," Gates told reporters. "For years, in the interest of the overall health of the computer industry, we permitted the free and unfettered use of our proprietary numeric systems. However, changing marketplace conditions and the increasingly predatory practices of certain

Why negative dependence?

Two motivations:

1. Sampling
2. Concentration

Sampling

Consider a population of n individuals, of which you wish to sample a random subset. Interpret the random variable X_k as telling you whether the k^{th} individual is in the sample.

For statistical purposes, if i and j have similar properties (e.g, they are neighbors, they are related, they are members of the same PAC), it's best if X_i and X_j are not positively correlated.

We don't necessarily know anything about who is similar to whom, but in the spirit of R. A. Fisher, we can solve this for all possible hidden similarities by making sure that X_i and X_j are negatively correlated for every pair $i \neq j$.

With versus without replacement

The flavor of negative dependence is captured by sampling without replacement. Let X be the mean of a sample of k subjects chosen uniformly from a population of n with replacement and let Y be the mean when you sample without replacement.

Both X and Y have the same expectation (the population mean). It seems obvious that X should have greater dispersion than Y, but the proof, due to Hoeffding (1963) is not trivial.

A recent argument due to Luh and Pippenger (2014) shows that in fact $Y = \mathbb{E}(X|\mathcal{F})$ for an appropriate σ-field f. This is one way to prove convex domination, namely the inequality $\mathbb{E}g(X) \geq \mathbb{E}g(Y)$ for all convex functions g.

πps-sampling problem

Often we would like to ensure that the marginal probabilities $\mathbb{E}X_k$ are equal to, or proportional to, a set of prescribed probabilities $\{\pi_k : 1 \leq k \leq n\}$. This is the so-called πps-sampling problem.

One can consider the further problem of πps-sampling under the requirement of negative dependence.

By the end of the lectures, I will tell you about a number of such schemes and the negative dependence properties they possess.

Some of this is surveyed by Brändén and Jonasson (2012); see also references therein, by Brewer and Hanif (1983) and Tillé (2006).

Concentration inequalities

Often one is interested in the total number of ones, that is, the sum $S := \sum_{k=1}^n X_k$.

Similarly, one might care about a subset sum $S_A := \sum_{k \in A} X_k$.

There is a long literature on tail bounds and limit theorems for sums of random variables under various dependence conditions.

A simple example:

The relatively weak condition of **pairwise negative correlation** implies $\mathrm{Var}\,(S) \leq \sum_{k=1}^n \mathbb{E}X_k(1 - \mathbb{E}X_k)$, and consequent Chebyshev bound $\mathbb{P}(|S - \mathbb{E}S| \geq a) \leq \mathrm{Var}\,(S)/a^2$.

Better negative dependence properties

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

Why do we care about these fancier 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.

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 $\mathcal{B}_n$ is **positively associated** if

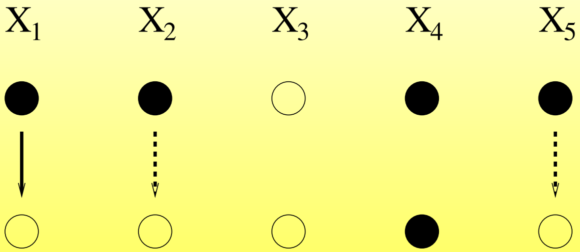
$$\mathbb{E}fg \geq (\mathbb{E}f)\ (\mathbb{E}g)$$

whenever f and g are both monotone increasing on $\mathcal{B}_n$. Taking $f = X_i, g = X_j$ this implies pairwise negative 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. We say that 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 $(\mathbb{P}|X_1 = 1)$ 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).

Negative association

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

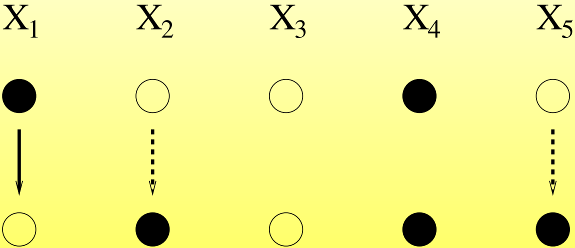
The measure $\mathbb{P}$ on $\mathcal{B}_n$ is **negatively associated** if

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

whenever f and g are both monotone increasing and they depend on disjoint 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.

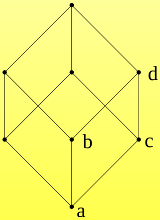
Coupling for negative association



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 $(\mathbb{P}|X_1 = 0)$.

Lattice conditions

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



Say that $\mathbb{P}$ satisfies the **positive lattice condition** if $\mathbb{P}(b)\mathbb{P}(c) \leq \mathbb{P}(a)\mathbb{P}(d)$ for every diamond (a, b, c, 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 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 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 by RP.

The next slide reviews four reasonably useful properties, each of which is strictly stronger than the last.

In each case, a consequent concentration inequality will be given.

Negative dependence hierarchy

1. Pairwise negative correlation: $\mathbb{E}X_iX_j \leq (\mathbb{E}X_i)(\mathbb{E}X_j)$.
2. Negative cylinder property: $\mathbb{E}\prod_{k\in A} X_k \leq \prod_{k\in A}\mathbb{E}X_k$.
3. 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.
4. Strong Rayleigh: definition TBA.

Negative dependence hierarchy

1. Pairwise negative correlation: $\mathbb{E}X_iX_j \leq (\mathbb{E}X_i)(\mathbb{E}X_j)$.
2. Negative cylinder property: $\mathbb{E}\prod_{k\in A} X_k \leq \prod_{k\in A} \mathbb{E}X_k$.
3. 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.
4. Strong Rayleigh: definition TBA.

1. NC implies $\mathbb{P}(|S - ES| \geq a) \leq n/(4a^2)$

Negative dependence hierarchy

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

1. NC implies $\mathbb{P}(|S - ES| \ge a) \le n/(4a^2)$
2. NCP implies Gaussian bounds: $\mathbb{P}(S - \mathbb{E}S \ge a) \le \exp(-2a^2/n)$

Negative dependence hierarchy

1. Pairwise negative correlation: $\mathbb{E}X_iX_j \le (\mathbb{E}X_i)(\mathbb{E}X_j)$.
2. Negative cylinder property: $\mathbb{E}\prod_{k\in A} X_k \le \prod_{k\in A} \mathbb{E}X_k$.
3. Negative association: $\mathbb{E}fg \le (\mathbb{E}f)(\mathbb{E}g)$ whenever f and g are increasing functions on $\mathcal{B}_n$ measurable with respect to disjoint sets of coordinates.
4. Strong Rayleigh: definition TBA.

1. NC implies $\mathbb{P}(|S - ES| \ge a) \le n/(4a^2)$
2. NCP implies Gaussian bounds: $\mathbb{P}(S - \mathbb{E}S \ge a) \le \exp(-2a^2/n)$
3. NA implies a self-normalized CTL: $(S - \mathbb{E}S)/\mathrm{Var}\,(S)^{1/2} \to \chi$

Negative dependence hierarchy

1. Pairwise negative correlation: $\mathbb{E}X_iX_j \leq (\mathbb{E}X_i)(\mathbb{E}X_j)$.
2. Negative cylinder property: $\mathbb{E}\prod_{k\in A} X_k \leq \prod_{k\in A} \mathbb{E}X_k$.
3. 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.
4. Strong Rayleigh: definition TBA.

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

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.

NC and NC^+ in terms of generating functions

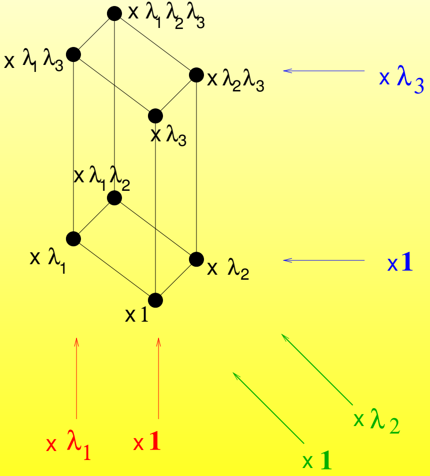
In terms of the generating function, NC is expressed by

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

If we require this not just at $(1, \ldots, 1)$ but at all points in the positive orthant, for all **x** we get the so-called **Rayleigh** property.

Probabilistically, this is the property that all measures produce from $\mathbb{P}$ by **external fields** are NC; here an external field is a reweighting of each $\omega \in \mathcal{B}_n$ by $\lambda_1^{\omega_1} \cdots \lambda_n^{\omega_n}$ for some fixed positive real parameters $\lambda_1, \ldots, \lambda_n$.

External field



Definition of strong Rayleigh

Definition 2

A measure $\mathbb{P}$ on $\mathcal{B}_n$ is said to be strong Rayleigh if its generating function F satisfies

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

for all $\mathbf{x} \in \mathbb{R}^n$ (negative coordinates now allowed!)

This strengthening, while not intuitive, makes algebraic methods more powerful because the variables are no longer constrained.

Definition of strong Rayleigh

Definition 2

A measure $\mathbb{P}$ on $\mathcal{B}_n$ is said to be strong Rayleigh if its generating function F satisfies

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

for all $\mathbf{x} \in \mathbb{R}^n$ (negative coordinates now allowed!)

This strengthening, while not intuitive, makes algebraic methods more powerful because the variables are no longer constrained.

VERY USEFUL FACT:

Definition of strong Rayleigh

Definition 2

A measure $\mathbb{P}$ on $\mathcal{B}_n$ is said to be strong Rayleigh if its generating function F satisfies

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

for all $\mathbf{x} \in \mathbb{R}^n$ (negative coordinates now allowed!)

This strengthening, while not intuitive, makes algebraic methods more powerful because the variables are no longer constrained.

VERY USEFUL FACT: For multi-affine functions, (1) is equivalent to F being nonzero on $\mathbb{H}^n$ where H is the open upper half plane. This property, called **stability**, has been well studied.

Stability theory in 3 slides (and two pictures)

It is not the place to take a long detour into complex function theory.

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 content of the lectures you 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

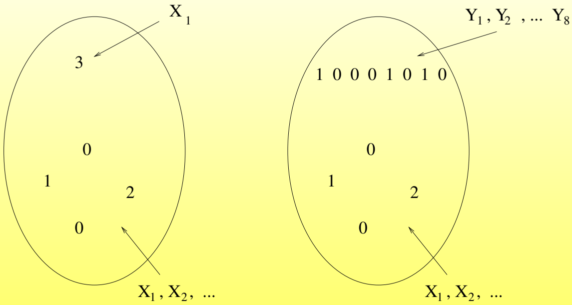
Let $X_1, \ldots, X_n$ be nonnegative integer random variables, all bounded by M. Polarization means replacing X_1 by Boolean variables $\{Y_1, \ldots, Y_M\}$ such that, conditional on $X_1, \ldots, X_n$, the Y variables are exchangeable and sum to X_1.

Lemma 3

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 4 (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.

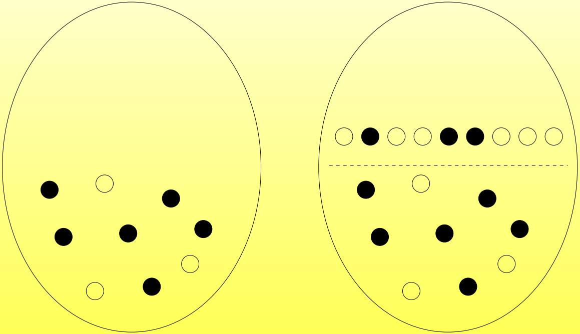
Definition 5

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 6

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.

Time to start reaping benefits

Fruitfulness of the strong Rayleigh property rests on these.

1. Strong conclusions.

   It implies negative association, and all that follows.

2. Relatively checkable hypotheses.

   There are many classes of examples.

Time to start reaping benefits

Fruitfulness of the strong Rayleigh property rests on these.

1. Strong conclusions.

    It implies negative association, and all that follows.

2. Relatively checkable hypotheses.

    There are many classes of examples.

I hope I have convinced you of the strong conclusions, pending a proof of SR implies NA. But...

How is this crazy hypothesis checkable?

Time to start reaping benefits

Fruitfulness of the strong Rayleigh property rests on these.

1. Strong conclusions.

   It implies negative association, and all that follows.

2. Relatively checkable hypotheses.

   There are many classes of examples.

I hope I have convinced you of the strong conclusions, pending a proof of SR implies NA. But...

How is this crazy hypothesis checkable?

We need to discuss closure properties...

Elementary closure properties

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

Elementary 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.

Elementary 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.

Elementary 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.

Elementary 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 $X_1 + X_2$: $F(x_1, x_1, x_3, \ldots, x_n)$.

Elementary 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: $\dfrac{\partial F}{\partial x_j}$ and $F - x_j \dfrac{\partial F}{\partial x_j}$ are stable if F is.

Not so obvious closure properties

Three more closure properties hold that are probabilistically meaningful but less automatic.

Not so obvious closure properties

Three more closure properties hold that are probabilistically meaningful but less automatic.

1. External field: $F(\lambda_1 x_1, \ldots, \lambda_n x_n)$ is stable if F is.

Not so obvious closure properties

Three more closure properties hold that are probabilistically meaningful but less automatic.

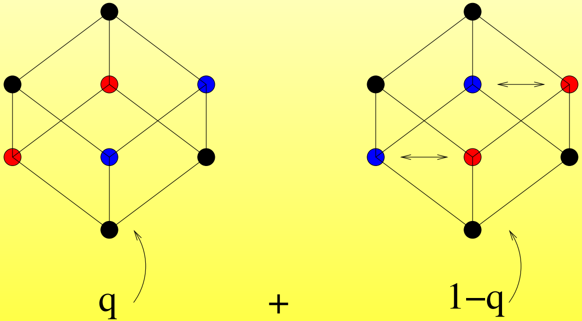
1. External field: $F(\lambda_1 x_1, \ldots, \lambda_n x_n)$ is stable if F is.
2. Stirring, that is, replacing F by a convex combination of F and $F^{ij} := F(x_1, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n)$.

Not so obvious closure properties

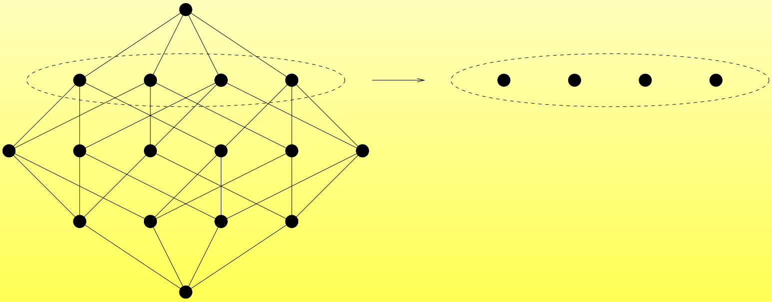
Three more closure properties hold that are probabilistically meaningful but less automatic.

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



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

Proofs

Proofs

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

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.

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 $\{\pi_i : 1 \le i \le n\}$ 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*.

Theorem 7

For any choice of parameter values, the measure $\mathbb{P}'$ is strong Rayleigh. 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 Poisson sampling law has marginals $p_1, \ldots, p_n$. All of these produce the same law, which maximizes entropy among laws with marginals $p_1, \ldots, p_n$.

Proof: $\mathbb{P}$ is trivially SR (e.g., because it is a product). $\mathbb{P}'$ is SR by closure under conditioning. The remaining facts are well known, e.g., Brändén and Jonasson (2012) or Singh and Vishnoi (2013).

Exclusion processes

Closure of SR under stirring, applied in a continuous manner, produces the following result.

Exclusion processes

Closure of SR under stirring, applied in a continuous manner, produces the following result.

Theorem 8

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.

Closure of SR under stirring, applied in a continuous manner, produces the following result.

Theorem 8

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.

One application to sampling occurs when the ground set is not $\{1, \ldots, n\}$ but some set of configurations understood only locally. A Monte Carlo scheme starts with a set of configurations, then repeatedly picks one at random and swaps it out for a neighbor. Then the configuration after T swaps is SR.

Pivot sampling

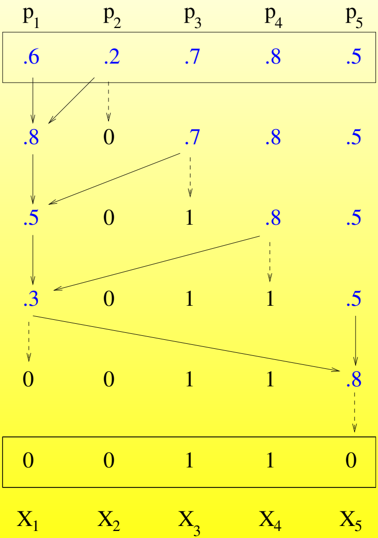
Once more let $\{p_i\}$ be probabilities summing to an integer $k < n$. Recursively, define a sampling scheme as follows.

If $p_1 + p_2 \leq 1$ then set X_1 or X_2 to zero with respective probabilities $p_2/(p_1 + p_2)$ and $p_1/(p_1 + p_2)$, then to choose the variables other than what was set to zero, run pivot sampling on $(p_1 + p_2, p_3, \ldots, p_n)$.

If $p_1 + p_2 > 1$, do the same thing except set one of X_1 or X_2 equal to 1 instead of 0 and the other to $p_1 + p_2 - 1$.

This method is very quick and does not involve having to compute auxilliary numbers such as the numbers p_i in conditional Bernoulli sampling. Brändén and Jonasson (2012) show that several πps-sampling procedures, including pivot sampling, are SR.

Example of pivot sampling



Probability of this was 3/4

Probability of this was 7/15

Probability of this was 8/13

Probability of this was 5/8

Probability of this was 2/10

Proof that pivot sampling is SR

Proof that pivot sampling is SR

Induct on n. Assume WLOG that $p_1 + p_2 \leq 1$. Let $\mathbb{P}$ denote the law of pivot sampling with probabilities $(p_1 + p_2, p_3, \ldots, p_n)$. By induction $\mathbb{P}$ is strong Rayleigh.

Proof that pivot sampling is SR

Induct on n. Assume WLOG that $p_1 + p_2 \leq 1$. Let $\mathbb{P}$ denote the law of pivot sampling with probabilities $(p_1 + p_2, p_3, \ldots, p_n)$. By induction $\mathbb{P}$ is strong Rayleigh.

Let $\mathbb{P}'$ be the product of the degenerate law δ_0 with $\mathbb{P}$, that is, sample from $\mathbb{P}$ then prepend a zero. This is trivially SR as well.

Proof that pivot sampling is SR

Induct on n. Assume WLOG that $p_1 + p_2 \leq 1$. Let $\mathbb{P}$ denote the law of pivot sampling with probabilities $(p_1 + p_2, p_3, \ldots, p_n)$. By induction $\mathbb{P}$ is strong Rayleigh.

Let $\mathbb{P}'$ be the product of the degenerate law δ_0 with $\mathbb{P}$, that is, sample from $\mathbb{P}$ then prepend a zero. This is trivially SR as well.

Let $\mathbb{P}''$ be the law $q\mathbb{P}' + (1-q)(\mathbb{P}')^{12}$, obtained from $\mathbb{P}'$ by transposing 1 and 2 with probability $q := p_1/(p_1 + p_2)$.

Proof that pivot sampling is SR

Induct on n. Assume WLOG that $p_1 + p_2 \leq 1$. Let $\mathbb{P}$ denote the law of pivot sampling with probabilities $(p_1 + p_2, p_3, \ldots, p_n)$. By induction $\mathbb{P}$ is strong Rayleigh.

Let $\mathbb{P}'$ be the product of the degenerate law δ_0 with $\mathbb{P}$, that is, sample from $\mathbb{P}$ then prepend a zero. This is trivially SR as well.

Let $\mathbb{P}''$ be the law $q\mathbb{P}' + (1-q)(\mathbb{P}')^{12}$, obtained from $\mathbb{P}'$ by transposing 1 and 2 with probability $q := p_1/(p_1 + p_2)$.

$\mathbb{P}''$ is strong Rayleigh by closure under stirring.

Proof that pivot sampling is SR

Induct on n. Assume WLOG that $p_1 + p_2 \leq 1$. Let $\mathbb{P}$ denote the law of pivot sampling with probabilities $(p_1 + p_2, p_3, \ldots, p_n)$. By induction $\mathbb{P}$ is strong Rayleigh.

Let $\mathbb{P}'$ be the product of the degenerate law δ_0 with $\mathbb{P}$, that is, sample from $\mathbb{P}$ then prepend a zero. This is trivially SR as well.

Let $\mathbb{P}''$ be the law $q\mathbb{P}' + (1-q)(\mathbb{P}')^{12}$, obtained from $\mathbb{P}'$ by transposing 1 and 2 with probability $q := p_1/(p_1 + p_2)$.

$\mathbb{P}''$ is strong Rayleigh by closure under stirring.

But $\mathbb{P}''$ is the law of pivot sampling on $(p_1, \ldots, p_n)$; this completes the induction.

Determinantal measures

The law $\mathbb{P}$ on $\mathcal{B}_n$ is *determinantal* if there is a Hermitian matrix M such that for all subsets A of $[n]$, the minor $\det(M|_A)$ computes $\mathbb{E} \prod_{k \in A} X_k$.

It is easy to see that determinantal measures have negative correlations. The diagonal elements give the marginals. By the Hermitian property, the determinant of $M|_{ij}$ must be less than $M_{ii} M_{jj}$. Thus,

$$(\mathbb{E} X_i)(\mathbb{E} X_j) = M_{ii} M_{jj} \leq M_{ii} M_{jj} - M_{ij} \overline{M_{ij}} = \mathbb{E} X_i X_j \, .$$

SR property for determinantal measures

Theorem 9

Determinantal measures are strong Rayleigh.

SKETCH OF PROOF: By the theory of determinantal measures, the eigenvalues of M must lie in $[0,1]$. Taking limits later if necessary, assume they lie in the open interval.

Then $F = C \det(H - Z)$ where Z is the diagonal matrix with entries $(x_1, \ldots, x_n)$ and $H = M^{-1} - I$ is positive definite. This is a sufficient criterion for stability (Gårding, circa 1951).

The following result, although we will not use it, is interesting.

Theorem 10 (Lyons)

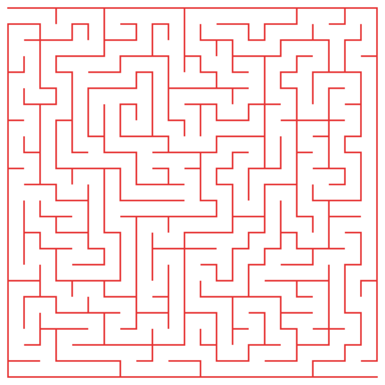
Given probabilities $p_1, \ldots, p_n$ summing to an integer $k < n$, we can always accomplish πps-sampling via a determinantal measure.

PROOF: The sequence $\{p_i : 1 \leq i \leq n\}$ is majorized by the sequence which is k ones followed by $n - k$ zeros. This majorization is precisely the criterion in the Schur-Horn Theorem, for existence of a Hermitian matrix M with $p_1, \ldots, p_n$ on the diagonal and eigenvalues consisting of 1 with mulitplicity k and 0 with multiplicity $n - k$. The matrix M defines the desired determinantal processes.

Spanning tree measures

Let $G = (V, E)$ be a graph with positive edge weights $\{w(e)\}$.

The weighted spanning tree measure is the measure WST on spanning trees proportional to $\prod_{e \in T} w(e)$.



Spanning trees are strong Rayleigh

The WST is determinantal; see, e.g., Burton and Pemantle (1993).

It follows that the random variables $\{X_e := \mathbf{1}_{e \in T}\}$ have the strong Rayleigh property. In particular, they are NA.

Oveis Gharan et al. (2013) use the strong Rayleigh property for spanning trees in a result concerning TSP approximation. From the strong Rayleigh property, they deduce a lower bound on the probability of a given two vertices simultaneously having degree exactly 2.

Further properties of SR measures useful for TSP

They make use of SR $\Rightarrow$ NA, as well as the following properties.

1. Let $\mathbb{P}$ be SR and let $S := \sum_{k=1}^n X_k$. Then S has the same law as the sum of independent Bernoullis. In particular, 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.

Further properties of SR measures useful for TSP

They make use of SR $\Rightarrow$ NA, as well as the following properties.

1. Let $\mathbb{P}$ be SR and let $S := \sum_{k=1}^{n} X_k$. Then S has the same law as the sum of independent Bernoullis. In particular, 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.

   PROOF: The univariate GF for S is a diagonal of F, hence is SR, hence has all real roots. The remaining properties follow, as will be discussed later.

Further properties of SR measures useful for TSP

They make use of SR $\Rightarrow$ NA, as well as the following properties.

1. Let $\mathbb{P}$ be SR and let $S := \sum_{k=1}^{n} X_k$. Then S has the same law as the sum of independent Bernoullis. In particular, the sequence $\{\mathbb{P}(S = k) : 0 \le k \le n\}$ is ultra-log-concave and it's mode and mean differ by at most one.

   PROOF: The univariate GF for S is a diagonal of F, hence is SR, hence has all real roots. The remaining properties follow, as will be discussed later.

2. **Stochastically increasing levels**: the law $(\mathbb{P} | S = k + 1)$ stochastically dominates the law $(\mathbb{P} | S = k)$.

Further properties of SR measures useful for TSP

They make use of SR $\Rightarrow$ NA, as well as the following properties.

1. Let $\mathbb{P}$ be SR and let $S := \sum_{k=1}^n X_k$. Then S has the same law as the sum of independent Bernoullis. In particular, 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.

   PROOF: The univariate GF for S is a diagonal of F, hence is SR, hence has all real roots. The remaining properties follow, as will be discussed later.

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.

A key lemma

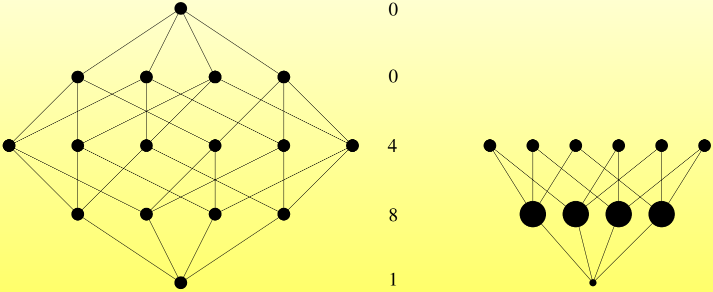
Lemma 11 (rank re-scaling)

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.

Exercise

Exercise: show that SR implies the **stochastic covering property**.

Exercise

Exercise: show that SR implies the **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.

Exercise

Exercise: show that SR implies the **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.

Exercise

Exercise: show that SR implies the **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.

Hint: 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.

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.

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 12 (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.

$\square$

CONCENTRATION INEQUALITIES

Concentration of the sum

Perhaps the most classical theorems in probability theory concern the distribution and tail bounds for sums of independent variables.

The hypothesis of joint independence is very strong. Weakening these has been a major theme. I will repeat two results that came up a few dozen slides ago.

1. Negative cylinder dependence is enough to derive exponential moments, hence Gaussian tail bounds on S.
2. If $\mathbb{P}$ is SR then S has a univariate generating function with all real roots. For example, it is immediate that S has the law of a sum of Bernoullis, from which a CLT follows directly.

Further properties of real-rooted sequences

Proposition 13 (generating polynomials with real roots)

1. *(Newton, 1707) A nonnegative coefficient sequence of a polynomial with real roots is log concave. In fact the sequence is* **ultra-logconcave**, *meaning that $\{a_k / \binom{n}{k}\}$ is log-concave.*
2. *(Edrai, 1953) A polynomial with nonnegative real coefficients has real roots if and only if its sequence of coefficients $(a_0, \ldots, a_n)$ is a* **Pólya frequency sequence**, *meaning that all the minors of the matrix (a_{i-j}) have nonnegative determinant.*
3. *Such a sequence is unimodal and its mean is within 1 of its mode.*

A good survey of such results may be found in Francisco Brenti's (1989) AMS Memoir.

Generalizing the sum

Instead of generalizing the measure, one could ask about concentration inequalities for something other than the sum.

One vein of research concerns maximal inequalities: here the functional to be bounded is typically something like the maximum partial sum $f := \max_{k \leq n} \sum_{j=1}^{k} X_j$.

In the context of Boolean measures, it makes sense to look for bounds that hold for the distribution any reasonably behaved function $f : \mathcal{B}_n \to \mathbb{R}$.

One notion of "well behaved" is to be Lipschitz with respect to the Hamming distance.

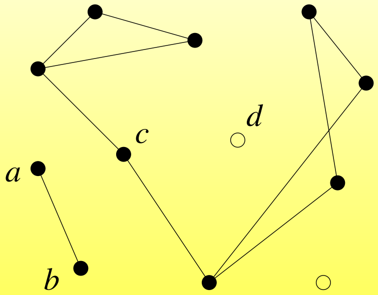
Lipschitz functionals

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 1: $S := \sum_{k=1}^n X_k$ is Lipschitz-1.

Example 2: 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.

Example Lipschitz function counting isolated vertices



The function counting isolated vertices is Lipschitz-2. For example, removing the edge *ab* alters the number of isolated vertices by $+2$, adding an edge *cd* alters the count by -1, and so forth.

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. We don't know, but we can do it if we assume the strong Rayleigh property.

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. We don't know, but we can do it if we assume the strong Rayleigh property.

Theorem 14 (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:

Proof

Sketch of proof:

- Strong Rayleigh measures have the stochastic covering property.

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)

Proof

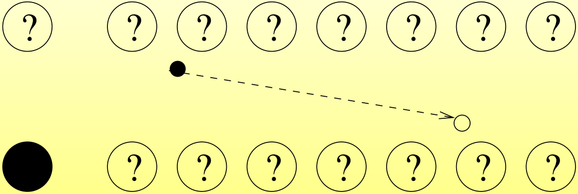
SKETCH OF PROOF:

- 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



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

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

FURTHER EXAMPLES

Sampling without replacement:

Let $r_1, \ldots, r_n$ be positive real weights. Weighted sampling without replacement is the measure on subsets of size k obtained as follows. Let $Y_1 \leq n$ have $\mathbb{P}(Y_1 = \ell)$ proportional to r_ℓ. Next, let $\mathbb{P}(Y_2 = m | Y_1 = \ell)$ be proportional to r_m for $m \neq \ell$. Continue in this way until Y_i is chosen for all $i \leq k$ and let $X_i = \sum_{j=1}^k \mathbf{1}_{Y_j = i}$.

There is a strong intuition that this law should be negatively associated. The law of a sample of size k is stochastically increasing in k, a property shared with strong Rayleigh measures.

Sampling without replacement:

Let $r_1, \ldots, r_n$ be positive real weights. Weighted sampling without replacement is the measure on subsets of size k obtained as follows. Let $Y_1 \leq n$ have $\mathbb{P}(Y_1 = \ell)$ proportional to r_ℓ. Next, let $\mathbb{P}(Y_2 = m | Y_1 = \ell)$ be proportional to r_m for $m \neq \ell$. Continue in this way until Y_i is chosen for all $i \leq k$ and let $X_i = \sum_{j=1}^k \mathbf{1}_{Y_j = i}$.

There is a strong intuition that this law should be negatively associated. The law of a sample of size k is stochastically increasing in k, a property shared with strong Rayleigh measures.

WRONG!

In a brief note in the Annals of Statistics, K. Alexander (1989) showed that weighted sampling without replacement is not, in general, even negatively correlated.

NA but not SR?

Negative association does not imply strong Rayleigh. The only published examples, however, are highly contrived.

NA but not SR?

Negative association does not imply strong Rayleigh. The only published examples, however, are highly contrived.

Problem 1

Find natural measures which are NA but not SR.

NA but not SR?

Negative association does not imply strong Rayleigh. The only published examples, however, are highly contrived.

Problem 1

Find natural measures which are NA but not SR.

There are some candidates: classes of measures that are known to be NA but for which it is not known whether they are SR. I will briefly discuss two of these, namely Gaussian threshhold measures and sampling via Brownian motion in a polytope.

Gaussian threshhold measures

Let M be a positive definite matrix with no positive entries off the diagonal. The multivariate Gaussian $\mathbf{Y}$ with covariances $\mathbb{E} Y_i Y_j = M_{ij}$ has pairwise negative correlations. It is well known, for the multivariate Gaussian, that this implies negative association.

Gaussian threshhold measures

Let M be a positive definite matrix with no positive entries off the diagonal. The multivariate Gaussian $\mathbf{Y}$ with covariances $\mathbb{E} Y_i Y_j = M_{ij}$ has pairwise negative correlations. It is well known, for the multivariate Gaussian, that this implies negative association.

Let $\{a_j\}$ be arbitrary real numbers and let $X_j = \mathbf{1}_{Y_j \geq a_j}$. For obvious reasons, we call the law $\mathbb{P}$ of $\mathbf{X}$ on $\mathcal{B}_n$ a Gaussian threshhold measure.

Gaussian threshhold measures

Let M be a positive definite matrix with no positive entries off the diagonal. The multivariate Gaussian $\mathbf{Y}$ with covariances $\mathbb{E} Y_i Y_j = M_{ij}$ has pairwise negative correlations. It is well known, for the multivariate Gaussian, that this implies negative association.

Let $\{a_j\}$ be arbitrary real numbers and let $X_j = \mathbf{1}_{Y_j \geq a_j}$. For obvious reasons, we call the law $\mathbb{P}$ of $\mathbf{X}$ on $\mathcal{B}_n$ a Gaussian threshhold measure.

Any monotone function f on $\mathcal{B}_n$ lifts to a function $\overline{f}$ on $\mathbb{R}_n$ with $\mathbb{E}\overline{f}(\mathbf{Y}) = \mathbb{E} f(\mathbf{X})$. Thus, negative association of the Gaussian implies negative association of $\mathbb{P}$.

Gaussian threshhold measures

Let M be a positive definite matrix with no positive entries off the diagonal. The multivariate Gaussian $\mathbf{Y}$ with covariances $\mathbb{E}Y_iY_j = M_{ij}$ has pairwise negative correlations. It is well known, for the multivariate Gaussian, that this implies negative association.

Let $\{a_j\}$ be arbitrary real numbers and let $X_j = \mathbf{1}_{Y_j \geq a_j}$. For obvious reasons, we call the law $\mathbb{P}$ of $\mathbf{X}$ on $\mathcal{B}_n$ a Gaussian threshhold measure.

Any monotone function f on $\mathcal{B}_n$ lifts to a function $\overline{f}$ on $\mathbb{R}_n$ with $\mathbb{E}\overline{f}(\mathbf{Y}) = \mathbb{E}f(\mathbf{X})$. Thus, negative association of the Gaussian implies negative association of $\mathbb{P}$.

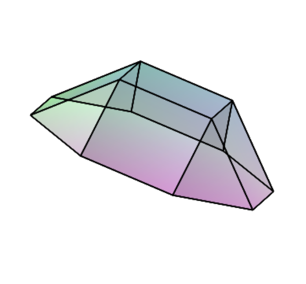
Problem 2

Is the Gaussian threshhold sampling law, $\mathbb{P}$, strong Rayleigh?

Sampling and polytopes

Sampling k elements out of n chooses a random k-set. Suppose we wish to restrict the set to be in a fixed list, $\mathcal{M}$. If we embed $\mathcal{B}_n$ in $\mathbb{R}^n$, then each set in $\mathcal{M}$ becomes a point in the hyperplane $\{\omega : \sum_{i=1}^n \omega_i = k\}$.

The set of probability measures on $\mathcal{M}$ maps to the convex hull of $\mathcal{M}$. The inverse image of $\mathbf{p}$ is precisely the set of measures with marginals given by $(p_1, \ldots, p_n)$. Thus, the πps-sampling problem is just the problem of choosing a point in the inverse image of $\mathbf{p}$.



A point $\mathbf{p}$ in the polytope is a mixture of vertices in many ways, each corresending to a measure supported on $\mathcal{M}$ that has mean $\mathbf{p}$.

Brownian vertex sampling

Based on an idea of Lovett and Meka, M. Singh proposed to sample by running Brownian motion started from $\mathbf{p}$. At any time, the Brownian motion is in the relative interior of a unique face, in which it constrained to remain thereafter. It stops when a vertex is reached. The martingale property of Brownian motion guarantees that this random set has marginals $\mathbf{p}$.

Brownian vertex sampling

Based on an idea of Lovett and Meka, M. Singh proposed to sample by running Brownian motion started from $\mathbf{p}$. At any time, the Brownian motion is in the relative interior of a unique face, in which it constrained to remain thereafter. It stops when a vertex is reached. The martingale property of Brownian motion guarantees that this random set has marginals $\mathbf{p}$.

It turns out that the resulting scheme is negatively associated as long as the set system $\mathcal{M}$ is a **matroid**. This notion generalizes many others, such as spanning trees and vector space bases. For **balanced** matroids it is known that the uniform measure is strong Rayleigh, but even negative correlation can fail for other matroids.

Nevertheless...

Negative association of Brownian vertex sampling

Theorem 15 (Peres and Singh, 2014)

For any matroid $\mathcal{M}$, the random k-set chosen by Brownian vertex sampling is negatively associated.

Sketch of proof: Let f and g be monotone functions depending on different sets of coordinates. Then $f(B_t)g(B_t)$ can be seen to be a supermartingale. At the stopping time, one gets

$$\int fg \, d\mathbb{P} = \mathbb{E} f(B_\tau) g(B_\tau) \leq \mathbb{E} f(B_0) \, \mathbb{E} g(B_0) = \int f \, d\mathbb{P} \int g \, d\mathbb{P} \, .$$

Is Brownian sampling strong Rayleigh?

Problem 3

Is the Brownian sampler strong Rayleigh?

Uniform acyclic subgraph

I will conclude by mentioning some measures where negative dependence is conjectured but it is not known whether any of the properties from negative correlation to srtong Rayleigh holds.

A spanning tree is a connected acyclic graph. A seemingly small perturbation of the uniform or weighted spanning tree is the uniform or weighted acyclic subgraph – we simply drop the condition that the graph be connected.

Problem 4

Is the uniform acyclic graph strong Rayleigh? Is it even NC?

Note: one might ask the same question for the dual problem, namely uniform or weighted connected subgraphs. This problem is open. I don't know offhand whether the two are equivalent.

The random cluster model

The random cluster model is a statistical physics model in which a random subset of the edges of a graph G is chosen. The probability of $H \subseteq G$ is proportional to a product of edge weights $\prod_{e \in H} \lambda_e$, times q^N where N is the number of connected components of H. When $q \geq 1$, it is easy to check the positive lattice condition, hence positive association. When $q \leq 1$, is is conjectured to be negatively dependent (all properties from negative correlation to strong Rayleigh being equivalent for this model).

Problem 5 (random cluster model)

Prove that the random cluster model has negative correlations.

The random cluster model

The random cluster model is a statistical physics model in which a random subset of the edges of a graph G is chosen. The probability of $H \subseteq G$ is proportional to a product of edge weights $\prod_{e \in H} \lambda_e$, times q^N where N is the number of connected components of H. When $q \geq 1$, it is easy to check the positive lattice condition, hence positive association. When $q \leq 1$, is is conjectured to be negatively dependent (all properties from negative correlation to strong Rayleigh being equivalent for this model).

Problem 5 (random cluster model)

Prove that the random cluster model has negative correlations.

Warning: this one has withstood a number of attacks.

THANKS FOR SITTING SO LONG!