Negative dependence in sampling

PETTER BRÄNDÉN

Royal Institute of Technology

JOHAN JONASSON

Chalmers University of Technology and Göteborg University

Abstract

The strong Rayleigh property is a new and robust negative dependence property that implies negative association; in fact it implies conditional negative association closed under external fields (CNA+). Suppose that $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$ are two families of 0-1 random variables that satisfy the strong Rayleigh property and let $Z_i := X_i + Y_i$. We show that $\{Z_i\}$ conditioned on $\sum_i X_i Y_i = 0$ is also strongly Rayleigh; this turns out to be an easy consequence of the results on preservation of stability of polynomials of Borcea & Brändén (2009). This entails that a number of important πps sampling algorithms, including Sampford sampling and Pareto sampling, are CNA+. As a consequence, statistics based on such samples automatically satisfy a version of the Central Limit Theorem for triangular arrays.

AMS Subject classification : 60J10

Key words and phrases: Pareto sampling, Rayleigh property, Sampford sampling,

uniform spanning tree

Short title: Negative dependence in sampling

1 Introduction

Suppose that in a population of N units, each unit i is equipped with some quantity of interest, y_i. In order to estimate some interesting function of the y_i's, e.g. the population total $T = \sum_i y_i$, we want to draw a sample of n units. With no other information, we would typically pick this sample uniformly among all n-subsets of $[N]$. However in many situations one has access to auxiliary information in the form of some quantity a_i, which is believed to be roughly proportional to y_i. This extra information could e.g. be older data. In this situation it is easy to argue that in order to get higher precision of estimators, one should pick different items with different probabilities; indeed one should take $\pi_i = \mathbb{P}(X_i = 1) = Ca_i$. (Here X_i is the indicator that item i is in the sample and the normalizing constant is chosen so that $\sum_i \pi_i = n$.) This is easily argued from the form of the standard Horvitz-Thompson estimator of T:

$$\hat{T} = \sum_i \frac{y_i}{\pi_i} X_i$$

which gets zero variance if y_i is exactly proportional to a_i.

Hence it is of interest to consider how to best go about picking a sample of fixed sample size n and prescribed inclusion probabilities π_i, so called πps-sampling. It turns out to be surprisingly difficult, indeed in fact impossible, to find a πps-sampling method which is "best" from all points of view. This has led to a large number of suggestions; indeed Brewer & Hanif (1983) contains more than fifty different πps sampling methods. A more recent reference that treats sampling algorithms extensively is the book by Tillé (2006).

The most important properties on the wish list for a sample are simplicity, efficiency (it should not take a computer too long to determine the sample), high entropy and statistical amenability. With statistical amenability, we mean that a Central Limit Theorem, like Proposition 1 below, or something similar holds. Traditionally, one has had to establish this for each method separately. However by Proposition 1, this follows automatically for most reasonable estimators if it can be shown that the sample is *negatively associated*. Here it will be demonstrated that simple applications of a new method, based on generating polynomials, establish that all of the most common πps-sampling methods are in fact negatively dependent in a very strong sense and in particular negatively associated.

2 Negative association and generating polynomials

A family $X = (X_1, \ldots, X_N)$ of random variables is said to be *positively associated* (PA) if

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)]$$

for all increasing functions f and g. We say that X is *negatively associated* (NA) if the reverse inequality holds for all increasing f and g such that there exists $I \subseteq [N] := \{1, \ldots, N\}$ such that f depends only on $X_I := \{X_i : i \in I\}$ and g only on $X_{[N]\setminus I}$. Positive association and negative association are very useful properties for drawing conclusions on correlation inequalities and limit results. For example, either of the two properties implies versions of the central limit theorem under mild additional assumptions, see e.g. Yuan *et al.* (2003) and Patterson *et al.*

(2001) and the references therein. For example, Theorem 2.4 of Patterson *et al.* (2001) states the following.

Proposition 1 *Let* $\{X_{n1}, \ldots, X_{nm_n}\}$, $n = 1, 2, \ldots$ *be a sequence of NA families of random variables with* $\mathbb{E}X_{nk} = 0$ *and write* $s_n^2 := \mathbb{V}\mathrm{ar}(\sum_{k=1}^{m_n} X_{nk})$. *Assume that* $s_n^2 \to \infty$, $s_n^{-2} \sum_j \sum_{i<j} \mathbb{C}\mathrm{ov}(X_{ni}, X_{nj}) \to 0$ *and, for any* $\delta > 0$,

$$\frac{1}{s_n^2} \sum_k \mathbb{E}[X_{nk}^2 \mathbf{1}_{\{|X_{nk}| > \delta s_n\}}] \to 0.$$

Then

$$\frac{1}{s_n} \sum_k X_{nk} \xrightarrow{D} Z$$

as $n \to \infty$, *where* Z *is standard normal.*

In general, it is often very difficult to check if a given family of random variables is PA or NA. The situation improves considerably if one restricts attention to $0/1$-valued random variables. In practice, this does not need to be a severe restriction. For example in a sampling situation, a typical estimator (of e.g. the population total or some other quantity of interest) is a sum of fixed positive quantities attached to the individuals of the sample. Hence Proposition 1 is applicable.

By the FKG inequality (see e.g. Grimmett (1999), Section 2.2) positive association follows from the so called positive lattice condition (PLC)

$$\begin{aligned}
&\mathbb{P}(X_{I\cup J} \equiv 1, X_{(I\cup J)'} \equiv 0)\mathbb{P}(X_{I\cap J} \equiv 1, X_{(I\cap J)'} \equiv 0) \geq \\
&\mathbb{P}(X_I \equiv 1, X_{I'} \equiv 0)\mathbb{P}(X_J \equiv 1, X_{J'} \equiv 0)
\end{aligned}$$

for all finite index sets I and J, where S' denotes the complement of S. Since this condition is local, it is often easier to check than PA directly. Unfortunately the corresponding negative lattice condition (NLC) does not imply negative association. Therefore, NA has turned out to be very difficult to prove and has been established only in a handful out of many situations where it is believed to hold. In a pioneering paper, Pemantle (2000), called out for a search for a useful theory of negative dependence. Since then, the topic has attracted a considerable interest in statistics, computer science, combinatorics and discrete probability, not only for its usefulness, but also for the mathematical challenge it presents. As for statistics in particular, the interest in negative association in fact goes back even further, to the very influential paper of Joag-Dev & Proschan (1983).

Until recently, progress has been made in small steps and has mainly consisted of case studies. However, the recent paper of Borcea *et al.* (2009) provides some significant steps forward for the general theory. One of the key concepts in that paper is the *strong Rayleigh property* of a set of (integer-valued bounded) random variables. The property is defined in terms of the *generating polynomial* of X. For $x \in \mathbb{Z}_+^N$, write $\mu(x) := \mathbb{P}(X_i = x_1, \dots, X_N = x_N)$. The generating polynomial of μ, or of X, is given by

$$F(z) = F_\mu(z) = F_X(z) = \mathbb{E}[z^X] = \sum_x \mu(x) z^x$$

where $z = (z_1, \dots, z_N) \in \mathbb{C}^N$ and $z^x = \prod_{i=1}^N z_i^{x_i}$. Recall that μ is *pairwise negatively correlated* (NC) if

$$\mathbb{P}(X_i = X_j = 1) \leq \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1),$$

for all $1 \leq i < j \leq N$. In terms of the generating polynomials NC translates to that the inequality

$$F \frac{\partial^2 F}{\partial z_i \partial z_j} \leq \frac{\partial F}{\partial z_i} \frac{\partial F}{\partial z_j} \tag{1}$$

holds for $z = (1, \ldots, 1)$ and all $1 \leq i < j \leq N$. A measure μ is NC+ if it is NC when an arbitrary external field, or exponential tilting, is introduced i.e., if for all $a \in \mathbb{R}_+^N$, the measure obtained by replacing $\mu(x)$ by $a^x \mu(x) / \sum_y a^y \mu(y)$ is NC. In terms of generating polynomials NC+ translates to that (1) holds for all $z \in \mathbb{R}_+^N$ and all $1 \leq i < j \leq N$. Due to the interpretation of (1) in electrical network theory Wagner (2008) called NC+ the *Rayleigh property*. It was believed and conjectured, see Pemantle (2000) and Wagner (2008), that the Rayleigh property implies several other important negative dependence properties. However many of these conjectures were disproved in Borcea *et el.* (2009) and Kahn & Neiman (2010). Surprisingly, if the Rayleigh property is innocently altered in the following manner one gets a very robust and useful negative dependence property.

Definition 1 *A measure μ is strongly Rayleigh if* (1) *holds for all $z \in \mathbb{R}^n$ and all $1 \leq i < j \leq N$.*

A reason for the robustness of the strong Rayleigh property is the following alternative definition due to Brändén (2007).

Proposition 2 *Let $X_1, \ldots, X_N$ be 0-1 random variables and μ the corresponding probability measure. Then μ is strongly Rayleigh if and only if the generating polynomial, F, is* stable *in the following sense:*

$$F(z) \neq 0 \textit{ whenever } \Im z_i > 0 \textit{ for all } 1 \leq i \leq N.$$

Using the geometry of zeros of multivariate polynomials, the following important result was proved in Borcea *et el.* (2009).

Proposition 3 *All families $X = (X_1, \ldots, X_N)$ of 0-1 random variables with the strong Rayleigh property are CNA+.*

Here CNA stands for conditional negative association, i.e. the property that X is NA also if we condition on $X_I = x$ for arbitrary $I \subseteq [N]$ and $x \in \{0,1\}^I$. The property CNA+ of course means that CNA holds also when an arbitrary external field is introduced. Using this result, the authors of Borcea *et el.* (2009) show e.g. that the symmetric exclusion process with strong Rayleigh initial distribution is strong Rayleigh for all positive times.

It is not easy to get an intuitive feeling for why the result of Proposition 3 is true. However, in the light of Proposition 2 it is actually not difficult to prove. A short summary goes as follows.

Sketch proof. In terms of the generating polynomial, the operation of *projection*, i.e. regarding only a subset of the X_i's, amounts to setting $z_i = 1$ for i's corresponding to the variables that are disregarded. The operation of *inversion*, i.e. considering $1-X$ instead of X, amounts the replacing the z_i's with z_i^{-1}. *Conditioning* on that one variable X_i is 0 amounts to considering $\partial F/\partial z_i|_{z_i=0}$. Using this, it follows easily from Proposition 2 that the SR property is closed under projections and inversions. Closure under external fields is immediate. Applying the Gauss-Lucas Theorem, the elementary theorem that states that the set of zeros of the derivative of a univariate polynomial (with real or complex coefficients) is always contained in the convex hull of the set of zeros of the polynomial itself, it also readily follows that SR is closed under conditioning. Hence, once one has checked that SR implies NC for $N = 2$, then this also holds for all N. This however follows from elementary calculations.

Now if X is a fixed size sample, the argument is finished via an appeal to the elementary Feder-Mihail result which in particular states that for such a sample, conditional pairwise negative correlations implies CNA. (See Feder & Mihail

(1992).) For the general result (which is not needed in this paper), one also needs to know that any SR measure is the projection of some fixed size SR measure, see Borcea *et el.* (2009). $\square$

Another key result is the following proposition, first proved in Choe *et el.* (2004). It states that the operation of throwing away nonlinear terms of a stable polynomial preserves stability. This result was extended in Borcea & Brändén (2009) (Theorem 2.1), to a precise criterion for when a linear operation on polynomials with complex coefficients preserves stability. We give here a short proof based on the Gauss–Lucas theorem.

Proposition 4 (Choe *et el.* (2004), Prop. 4.17) *If* $P(z) = \sum_{\alpha \in \mathbb{N}^k} a_\alpha z^\alpha$ *is stable, then*

$$Q(z) = \sum_{\alpha \in \{0,1\}^k} a_\alpha z^\alpha$$

is stable or identically zero.

Proof. Write $P(z) = \sum_{j=0}^{K} a_j(\hat{z}) z_1^j$, where $\hat{z} = (z_2, \ldots, z_k)$. Clearly it suffices to prove that $a_0(\hat{z}) + a_1(\hat{z}) z_1$ is stable or identically zero. Since the operation $z_1 \mapsto -z_1^{-1}$ maps the upper half-plane to itself, the inverted polynomial $F(z) := \sum_{j=0}^{K} a_j(\hat{z})(-1)^j z_1^{K-j}$ is stable. It follows from the Gauss–Lucas theorem that $\partial F / \partial z_1$ is stable or identically zero. Hence, by iterating $K-1$ times,

$$R(z) := \partial^{K-1} F / \partial z_1^{K-1} = (K-1)!(K a_0(\hat{z}) z_1 - a_1(\hat{z}))$$

is stable or identically zero. By inverting again and rescaling the variable z_1, we see that $a_0(\hat{z}) + a_1(\hat{z}) z_1$ is stable or identically zero as desired. $\square$

Here is our main result.

Theorem 1 *Let $X = (X_1, \ldots, X_N)$ and $Y = (Y_1, \ldots, Y_N)$ be two independent strongly Rayleigh families of 0-1 random variables and let $Z = X + Y$. Then the law of Z given that $\sum_i X_i Y_i = 0$ is also strongly Rayleigh.*

Phrased in terms of samples, Theorem 1 states that the union of two strongly Rayleigh samples, conditioned on these being disjoint, is strongly Rayleigh.

Proof. Let F and G be the generating polynomials of X and Y respectively. Let H be the generating polynomial of Z given that X and Y are disjoint. We want to show that H is stable. However, the generating polynomial of the unconditional distribution of $X + Y$, i.e. for the convolution of the two corresponding measures, is FG, which is obviously stable. Now H is (up to a constant) derived from FG by removing all terms that are not multi-affine. Hence stability of H follows from Proposition 4. □

Another result that will be of use is the following, which states that if an item in a strongly Rayleigh sample is with a certain probability replaced with a new item previously not in the population, then the new sample is also strongly Rayleigh.

Proposition 5 *Let $X = (X_1, \ldots, X_N)$ be a strongly Rayleigh family of 0-1 random variables such that $\mathbb{P}(X_2 = 1) = 0$. Let I be a 0-1 random variable independent of X and let $Y = (Y_1, \ldots, Y_N)$ be given by $Y_1 = IX_1$, $Y_2 = (1 - I)X_1$ and $Y_j = X_j$, $j = 3, \ldots, N$. Then Y is strongly Rayleigh.*

Proof. This is in fact a special case of Borcea *et el.* (2009), Theorem 4.20. However, since this special case has such a simple proof, let us give it here. If $F = F(z_1, z_3, z_4, \ldots, z_N)$ is the generating polynomial of X, then the generating polynomial of Y is given by $G = pF(z_1, z_3, \ldots, z_N) + (1 - p)F(z_2, z_3, \ldots, z_n)$

where $p = \mathbb{P}(I = 1)$. Fixing $z_3, \ldots, z_N$, we can write $F = az_1 + b$ so that $G = a(pz_1 + (1-p)z_2) + b$. Hence if G has a zero z' with $\Im z'_i > 0$ for all i, then F has the zero $(pz'_1 + (1-p)z'_2, z'_3, \ldots, z'_N)$. In other words, if G is not stable, then neither is F, a contradiction. □

The next section is devoted to πps sampling applications.

3 Applications to sampling

Consider the following rejective πps-sampling method in terms of balls and bins. We have n balls and N bins and we let each unit i in the population correspond to a bin and each choice of an individual for the sample correspond to a ball. Randomly place the balls in the bins independently but with possibly different distributions for different balls; more precisely, let ball number i go to bin number j with probability θ_{ij}. Finally condition on that no two balls go into the same urn. The sample is then defined to consist of the units corresponding to bins that contain one ball. More formally

$$\mathbb{P}(X = x) = C \sum \prod_{i=1}^{n} \prod_{j=1}^{N} \theta_{ij}^{x_{ij}}$$

where the sum is over all $\{x_{ij}\}_{i\in[n],j\in[N]} \in \{0,1\}^{Nn}$ such that $\sum_{i=1}^{n} x_{ij} = x_j$ for each j and $\sum_{j=1}^{N} x_{ij} = 1$ for each i and C is a normalizing constant. Taking inspiration from the pioneering study Dubhashi & Ranjan (1998), we call a sample of this kind a *conditional balls-and-bins sample*, CBAB sample for short. Since any sample of exactly one unit is clearly strongly Rayleigh, it follows from Theorem 1 by induction that:

Theorem 2 *Any conditional balls-and-bins sample is strongly Rayleigh.*

Here are the most important special cases.

- **Conditional Poisson sampling.** Assign to each unit a probability parameter p_i and let each unit be contained in the sample with probability p_i independently of other units, but condition on that the number of chosen units is exactly n. (So in the end, of course, the units are not independent.)

  An equivalent way of describing this is to pick n units independently according to the probabilities $cp_i/(1-p_i)$ and condition on that all the chosen units are distinct.

  It is well known that CP sampling gives maximum entropy under the resulting inclusion probabilities. CP sampling is simple, but not efficient. The main problem is that the p_i's do not coincide with the resulting conditional inclusion probabilities. Finding the correct p_i's for the desired inclusion probabilities means to solve a huge system on non-linear equations, which is time consuming and can usually only be done numerically. (Numerical algorithms for finding approximate solutions are efficient, but if one for some reason insists on the exact right inclusion probabilities, this is indeed a problem.) Also, given the correct p_i's, there is the issue of implementation. However, using the straightforward rejective method, the number of attempts one has to make to get exactly n units in the sample, is of order $\sqrt{n}$, which is not a big problem (unless the size N of the population is of a much higher order).

  That CP sampling is CNA+ was observed in Dubhashi *et el.* (2007). The strong Rayleigh property now follows immediately from Theorem 2, since

CP sampling is the special case of CBAB sampling where the balls have equal probability distributions over bins.

- **Sampford sampling.** This method goes back to Sampford (1967). (See also Bondesson & Grafström (2011) for a recent extension.) This ingenious method consists of two choices and one conditioning step. First pick an item according to the probabilities π_i/n. Then pick a CP sample of size $n-1$ according to the parameters $c\pi_i/(1-\pi_i)$ (c normalizing) and finally condition on that the CP sample does not contain the first unit. The resulting sample is then the union of the first unit and the CP sample. Formally,

$$\mathbb{P}(X = x) = C \sum_{i=1}^{N} (1-\pi_i) x_i \prod_{j=1}^{N} \left(\frac{\pi_j}{1-\pi_j} \right)^{x_j}.$$

  The ingenuity lies in the far from trivial fact that this method actually gives the correct inclusion probabilities π_i. Sampford sampling has very close to maximum entropy and is simple. It is clearly more efficient than CP sampling since no parameters need to be calculated. It now follows that Sampford sampling is also strongly Rayleigh. This follows directly from the definition, the strong Rayleigh property of CP sampling and Theorem 1. Alternatively, one can observe that Sampford sampling is a CBAB sample where one of the balls has bin distribution $\{\pi_i/n\}$ and the others have bin distribution $\{c\pi_i/(1-\pi_i)\}$.

  Remark. Sampford sampling can be very inefficient if it is implemented in the way of placing balls in urns until no urn gets more than one ball. Indeed the expected running time will be exponential in n. However, a Sampford sample should be implemented as described above, i.e. as a CP sample and one additional item conditioned on being disjoint. This algorithm will need order $\sqrt{n}$ attempts to succeed. (To optimize efficiency, one should start with the "additional" item and then start the Poisson sample with that item, so that one can abort the attempt at once if that item gets chosen again.)

- **Pareto sampling.** This method, introduced by Rosén (1997), is commonly used in practice. Let $U_1, \ldots, U_N$ be iid uniform $(0, 1)$ random variables, let $V_i = U_i/(1 - U_i)$ and $W_i = V_i/\tau_i$, where the τ_i's are parameters, which are to be adjusted so as to give the desired inclusion probabilities. The sample consists of the items with the n lowest W_i's. Since the V_i's have distribution function $x/(1 - x)$ and density $1/(1 - x)^2$, conditioning on the index and value of the n'th smallest W_i leads to the following expression:

$$\mathbb{P}(X = x) = C \sum_{i=1}^{N} \frac{c_i}{\tau_i} x_i \prod_{j=1}^{N} \tau_j^{x_j}$$

where

$$c_j = \tau_j \int_0^\infty \frac{x^{n-1} G(x)}{1 + \tau_j x} dx$$

and $G(x) = \prod_{k=1}^{N} (1 + \tau_k x)^{-1}$, see e.g. Traat *et el.* (2004). Compared with Sampford sampling, Pareto sampling has the drawback that the τ_i's usually have to be calculated numerically. On the other hand, given the parameters, the sample comes out very quickly (unless N is very large) and without rejections.

We see that the form of the probability function of Pareto sampling is the same as that for Sampford sampling, so that Pareto sampling is also a CBAB where all balls but one have the same bin distribution.

A generalization of Pareto sampling is *order sampling*. Here the V_i's are iid random variables of an arbitrary distribution, F, with support on the positive numbers. Again $W_i = V_i/\tau_i$ and the sample consists of the indices with the n smallest W_i's. Order sampling in general is neither a CBAB nor strongly Rayleigh. Indeed, it may not even be NC; taking the V_i's to be exponential leads to the counterexample of Alexander (1989). However if the function $h(\tau, x) = F(\tau x)/(1 - F(\tau x))$ is separable, i.e. can be written on the form $h(\tau, x) = a(\tau)b(x)$, then the probability function of the sample becomes

$$\mathbb{P}(X = x) = C \sum_{i=1}^{N} \frac{c_i}{a(\tau_i)} x_i \prod_{j=1}^{N} a(\tau_j)^{x_j}$$

where

$$c_j = \tau_j \int_0^\infty \frac{b(x)^{n-1} G(x) f(\tau_j x)}{1 - F(\tau_j x)} dx$$

and $G(x) = \prod_{k=1}^{N} (1 - F(\tau_k x))$. One example is e.g. $F(x) = x^\alpha/(1 + x^\alpha)$ for any $\alpha > 0$.

Remark. A recent variant of Pareto sampling, so called conditional Pareto sampling, was recently introduced by Bondesson (2010). Intuitively, this method should also be strongly Rayleigh, but it is not clear to us if the present methods apply.

Another important and well known πps sampling method is *pivotal sampling* (also known as the Srinivasan sampling procedure in the computer science community). Although it may have low entropy, it is extremely simple and efficient and, as we shall see, enjoys all the virtues of negative dependence.

The method is defined inductively on the number of items in the population. In the simplest setting, the items are ordered linearly. Suppose that $\pi_1 + \pi_2 \leq 1$.

Then with probability $\pi_1/(\pi_1 + \pi_2)$ set $X_2 = 0$, $\pi_1' = \pi_1 + \pi_2$ and run pivotal sampling for a sample of size n on the population $1, 3, 4, \ldots, N$ according to $\pi_1', \pi_3, \ldots, \pi_N$. With the complementary probability $\pi_2/(\pi_1+\pi_2)$ use the opposite treatment of items 1 and 2. On the other hand if $\pi_1+\pi_2 > 1$, then with probability $(1 - \pi_1)/(2 - \pi_1 - \pi_2)$, set $X_2 = 1$, $\pi_1' = \pi_1 + \pi_2 - 1$ and run pivotal sampling for a sample of size $n - 1$ on $1, 3, 4, \ldots, N$. With the complementary probability, give items 1 and 2 the opposite treatment.

A more general method is achieved by picking a rooted binary tree on $2N - 1$ vertices with N leaves and placing the units at the leaves in any desired order. Pick, in some predetermined way, two units at leaves with a common neighboring vertex. Then, according to the same formulas as above, determine X_i for one of the two units, remove the two leaves and place the other unit at the common neighbor (which is now a leaf of a smaller tree). Then run pivotal sampling on the new smaller tree.

In Dubhashi *et el.* (2007), it was shown, with quite some effort, that pivotal sampling is NC in the general tree setting and CNA in the linear setting. Here it follows from Proposition 5 and induction that both are in fact strongly Rayleigh. To see this, assume without loss of generality that the two units picked in the inductive step of the method are units 1 and 2 and that $\pi_1 + \pi_2 \leq 1$, and that all pivotal samples on smaller trees are strongly Rayleigh. If we just postulate that $X_2 = 0$ and remove the two leafs and put item 1 at the new leaf, then by the induction hypothesis this gives a strongly Rayleigh sample. However, correcting for the postulation $X_2 = 0$ amounts precisely to the situation covered by Proposition 5.

Remark. One remedy for the low entropy, in particular under linear order, of pivotal sampling is obvious: order the items randomly before drawing the sample. Unfortunately we have not been able to show that this gives a strongly Rayleigh sample, even though we believe this to be the case. Of course, the sample is strongly Rayleigh given the order. However in general a convex combination of strongly Rayleigh measures is not necessarily strongly Rayleigh, indeed not even NC and not even under fixed sample size and fixed inclusion probabilities.

Let us summarize the results of the present section.

Theorem 3 *Conditional Poisson sampling, Sampford sampling, Pareto sampling and pivotal sampling specified by any rooted binary tree are all strongly Rayleigh and hence CNA+. This also goes for general order sampling if* $h(\tau, x) := F(\tau x)/(1-F(\tau x))$ *is separable.*

Acknowledgment. We thank Yaming Yu for pointing out the reference Alexander (1989) for us.

Petter Brändén is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation. Johan Jonasson was supported by a grant from the Swedish Research Council.

References

[1] Alexander, K. (1989), A counterexample to a correlation inequality in finite sampling, *Ann. Statist.* **17**, 436-439.

[2] Bondesson, L. (2010), Conditional and restricted Pareto sampling: two new methods for unequal probability sampling, *Scandinav. J. Statist.* **37**, 514-530.

[3] Bondesson, L. & Grafström, A. (2011), An extension of Sampford's method for unequal probability sampling, *Scandinav. J. Statist.* **38**, 377-392.

[4] Borcea, J. & Brändén, P. (2009), The Lee-Yang and Pólya-Schur programs I. Linear operators preserving stability, *Invent. Math.* **177**, 521-569.

[5] Borcea, J., Brändén, P. & Liggett, T. M. (2009), Negative dependence and the geometry of polynomials, *J. Amer. Math. Soc.* **22**, 521-567.

[6] Brändén, P. (2007), Polynomials with the half-plane property and matroid theory, *Adv. Math.* **216**, 302-320.

[7] Brewer, K. R. W. & Hanif, M. (1983), "Sampling with Unequal Probabilities," Springer, New York-Berlin.

[8] Choe, Y. B., Oxley, J. G., Sokal, A. D. & Wagner, D. G. (2004), Homogeneous multivariate polynomials with the half-plane property. Special issue on the Tutte polynomial. *Adv. in Appl. Math.* **32**, 88-187.

[9] Dubhasi, D., Jonasson, J. & Ranjan, D. (2007), Positive influence and negative dependence, *Combin. Probab. Comput.* **16**, 29-41.

[10] Dubhashi, D. & Ranjan, D. (1998), Balls and bins: A study in negative dependence, *Random Structures Algorithms* **13**, 99-124.

[11] Feder, T & Mihail, M. (1992), Balanced matroids, in ”Proc. of the 24th Annual ACM,” ACM Press, New York.

[12] Grimmett, G. (1999), ”Percolation,” Springer, Berlin.

[13] Joag-Dev, K. & Proschan, F. (1983), Negative association of random variables with applications, *Ann. Statist.* **11**, 286-295.

[14] Kahn, J. & Neiman, M. (2010), Negative correlation and log-concavity, *Random Structures Algorithms* **37**, 367-388.

[15] Patterson, R.F., Smith, W. D., Taylor, R. L. & Bozorgnia, A. (2001), Limit theorems for negatively dependent random variables, *Nonlinear Analysis* **47**, 1283-1295.

[16] Pemantle, R. (2000), Towards a theory of negative dependence, J. Math. Phys. **41**, 1371-1390.

[17] Rosén, B. (2007), On sampling with probability proportional to size, J. Statist. Plann. Inference **62**, 159-191.

[18] Sampford, M. R. (1967), On sampling without replacement with unequal probabilities of selection. *Biometrika* **54**, 499-513.

[19] Tillé, Y. (2006), ”Sampling Algorithms,” Springer, New York.

[20] Traat, I., Bondesson, L. & Meister, K. (2004), Sampling design and samle selection through distribution theory, *J. Statist. Plann. Inference* **123**, 395-413.

[21] M. Yuan, C. Su & T. Hu (2003), A central limit theorem for random fields of negatively associated processes, *J. Theoret. Probab.* **16**, 309-323.

[22] D. G. Wagner (2008), Negatively correlated random variables and Mason's conjecture for independent sets in matroids. *Ann. Comb.* **12**, 211-239.

Johan Jonasson

Department of Mathematics

Chalmers University of Technology and Göteborg University

412 96 Gothenburg, Sweden

jonasson@chalmers.se