144

Chapter 1

THE AHLSWEDE-DAYKIN THEOREM

Peter C. Fishburn
AT&T Labs–Research, Florham Park, NJ 07932
fish@research.att.com

Lawrence A. Shepp
Rutgers University, Piscataway, NJ 08855
shepp@stat.rutgers.edu

In appreciation to Rudolf Ahlswede

Abstract In 1978, Rudolf Ahlswede and David Daykin published a theorem which says that a certain inequality on nonnegative real valued functions for pairs of points in a finite distributive lattice extends additively to pairs of lattice subsets. It is an elegant theorem with widespread applications to inequalities for systems of subsets, linear extensions of partially ordered sets, and probabilistic correlation. We review the theorem and its applications, and describe a recent generalization to n-tuples of points and subsets in distributive lattices.

1. THE AHLSWEDE-DAYKIN THEOREM

A *lattice* is a partially ordered set $(\Gamma, \prec)$ in which every pair of points $a, b \in \Gamma$ has a unique least upper bound or join

$$a \vee b = \min\{z \in \Gamma : a \preceq z, b \preceq z\}$$

and a unique greatest lower bound or meet

$$a \wedge b = \max\{z \in \Gamma : z \preceq a, z \preceq b\} .$$

The lattice is *distributive* if

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c) \quad \text{for all} \quad a, b, c \in \Gamma$$

or, equivalently, if $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ for all $a, b, c \in \Gamma$. We presume throughout that Γ is finite and recall the useful fact [5, p. 59] that a finite distributive lattice is order-isomorphic for some n to a restriction of $(2^n, \subset)$, the family of subsets of $\{1, 2, \ldots, n\}$ ordered by proper inclusion.

For nonempty $A, B \subseteq \Gamma$, $\vee$ and $\wedge$ are extended to subsets of Γ by

$$\begin{aligned} A \vee B &= \{a \vee b : a \in A, b \in B\} \\ A \wedge B &= \{a \wedge b : a \in A, b \in B\} , \end{aligned}$$

with $A \vee B = \emptyset = A \wedge B$ if A or B is empty. In 1977, Daykin [11] proved that a lattice $(\Gamma, \prec)$ is distributive if and only if

$$|A||B| \leq |A \vee B||A \wedge B| \quad \text{for all} \quad A, B \subseteq \Gamma .$$

This inequality is but one of many implications of a remarkable theorem published the next year by Ahlswede and Daykin [3] that has come to be known as the Ahlswede-Daykin theorem, or the four-functions theorem [6]. For any real-valued function f on Γ, we define the *additive extension* of f, also denoted by f, by

$$f(A) = \sum_{a \in A} f(a) \quad \text{for all} \quad A \subseteq \Gamma .$$

Theorem 1. (Ahlswede-Daykin) *Suppose* $(\Gamma, \prec)$ *is a finite distributive lattice and* $\alpha, \beta, \gamma, \delta : \Gamma \to [0, \infty)$ *satisfy*

$$\alpha(a)\beta(b) \leq \gamma(a \vee b)\delta(a \wedge b) \quad \textit{for all} \quad a, b \in \Gamma \, .$$

Then

$$\alpha(A)\beta(B) \leq \gamma(A \vee B)\delta(A \wedge B) \quad \textit{for all} \quad A, B \subseteq \Gamma \, .$$

When $(\Gamma, \prec) = (2^n, \subset)$ with $\vee = \cup$ and $\wedge = \cap$, the hypothesized inequality, $\alpha(a)\beta(b) \leq \gamma(a \vee b)\delta(a \wedge b)$, has the flavor of *log supermodularity* for a probability distribution μ on the ground set 2^n, defined by

$$\mu(a)\mu(b) \leq \mu(a \cup b)\mu(a \cap b) \quad \text{for all} \quad a, b \in 2^n \, .$$

The hypothesized inequality of Theorem 1 can be viewed as a far-reaching generalization of log supermodularity, which is a key hypothesis of the widely-cited FKG theorem of Fortuin, Kasteleyn and Ginibre [18]. The power of the Ahlswede-Daykin theorem lies in its conclusion that the four-functions inequality hypothesized for individual members of Γ is inherited by subsets of Γ under additive extensions.

Proofs of Theorem 1 are included in [3, 6, 16]. The standard approach is to prove the theorem for $(2^n, \subset)$. The general result for $(\Gamma, \prec)$ order-isomorphic to a restriction of $(2^n, \subset)$ then follows by fixing α, β, γ and δ at 0 on the members of 2^n excluded from the isomorphism. The $(2^n, \subset)$ proof shows that the result holds for $n = 1$ and proceeds by induction on n. The overall proof is pleasantly compact — about one page — in view of the theorem's many implications.

Several of those implications, including the FKG theorem, were proved prior to the publication of [3]. We will not dwell on precedence, but instead will indicate how a variety of results follow from Theorem 1 as the root of a tree-like structure. We classify those results into three types.

Type 1 implications follow more or less directly from Theorem 1 by choosing specific forms for α, β, γ and δ. They include Daykin's inequality for distributivity [11], the FKG theorem [18] and Holley's theorem [22], an inequality of Kleitman [27] and Seymour [33], and the Marica-Schönheim inequality [30].

Type 2 implications use direct applications of Theorem 1 or its type 1 implications, but involve other techniques to arrive at their conclusions. The other techniques often include a reformulation of the problem's structure prior to the direct application, and may have one or more steps that require functional extremization or an examination of limit behavior. Examples include the correlational inequalities for linear extensions of Graham, Yao and Yao [21] and Shepp [34], the so-called xyz inequalities of Shepp [35] and Fishburn [13], and universal correlation theorems of Winkler [39] and Brightwell [8].

Type 3 implications involve structure for which the hypotheses of Theorem 1 or a type 1 or type 2 implication are false, even under reformulations, but which admit perturbations that allow application of preceding results. The perturbed structure is close to the original, and the disparity between the two can be remedied by methods that lead to the desired conclusion. Our primary example of a type 3 implication is a correlation inequality for match sets of random permutations that was conjectured by Joag-Dev [24] and Prem Goel and proved in Fishburn, Doyle and Shepp [17].

The question of which type characterizes a particular implication is subject to personal judgment and can depend on available proofs, so we acknowledge a degree of latitude in our choices. Nevertheless, we have found the classification useful for an appreciation of the role of the Ahlswede-Daykin theorem, and proceed accordingly.

Section 2 of the paper discusses type 1 implications, section 3 describes type 2 implications, and section 4 outlines our perturbation approach to the match set problem with random permutations. We then conclude with a recent generalization of the Ahlswede-Daykin theorem due to Rinott and Saks [31, 32] and Aharoni and Keich [2].

Prior surveys of much of the material we cover are presented by Graham [19, 20], Winkler [40] and Fishburn [16]. We have borrowed freely from these sources and acknowledge our indebtedness to Ron Graham and Peter Winkler.

2. TYPE 1 IMPLICATIONS

We assume throughout this section that $(\Gamma, \prec)$ is a finite distributive lattice. Our first implication of Theorem 1 takes $\alpha = \beta = \gamma = \delta = \mu$ with $\mu : \Gamma \to [0, \infty)$. Then *log supermodularity* for μ, i.e.,

$$\mu(a)\mu(b) \leq \mu(a \vee b)\mu(a \wedge b) \quad \text{for all} \quad a, b \in \Gamma \ ,$$

which becomes the hypothesized inequality of Theorem 1, implies the same form for additive extensions:

$$\mu(A)\mu(B) \leq \mu(A \vee B)\mu(A \wedge B) \quad \text{for all} \quad A, B \subseteq \Gamma \ .$$

When $\mu \equiv 1$ is added to the hypotheses, log supermodularity is automatic and Theorem 1 yields Daykin's inequality $|A||B| \leq |A \vee B||A \wedge B|$ for all $A, B \subseteq \Gamma$.

Log supermodularity also underlies the following lattice version of the FKG theorem. We say that $f : \Gamma \to \mathrm{R}$ is *nondecreasing* if

$$a \prec b \Rightarrow f(a) \leq f(b), \quad \text{for all} \quad a, b \in \Gamma \ .$$

Theorem 2. (FKG) *Suppose* $\mu : \Gamma \to [0,\infty)$ *is log supermodular. Then for all nondecreasing* $f, g : \Gamma \to \mathrm{R}$,

$$\left[\sum_\Gamma \mu(a)f(a)\right]\left[\sum_\Gamma \mu(a)g(a)\right] \le \left[\sum_\Gamma \mu(a)\right]\left[\sum_\Gamma \mu(a)f(a)g(a)\right] .$$

Proof. It is easily seen that the conclusion is invariant to the addition of a constant c to f and g, so we assume that f and g are positive. Then define α, β, γ and δ for Theorem 1 by $f\mu$, $g\mu$, $fg\mu$ and μ, respectively. For example, $\alpha(a) = f(a)\mu(a)$. The hypotheses of Theorem 2 then imply those of Theorem 1, and the conclusion of Theorem 1 implies that of Theorem 2 when $A = B = \Gamma$. ∎

Several implications of the FKG theorem will be noted later. Other implications and related results are available in Kemperman [26], Joag-Dev, Shepp and Vitale [25], van den Berg and Kesten [38], van den Berg and Fiebig [37], Hwang and Shepp [23], Burton and Franzosa [10], and Bollobás and Brightwell [7].

A probabilistic version of the FKG theorem arises by taking $(\Gamma, \prec) = (2^n, \subset)$ with $\vee = \cup$ and $\wedge = \cap$. Let $\mathcal{B}_n$ denote the Boolean algebra of subsets of 2^n, so each object in $\mathcal{B}_n$ is a set of subsets of $\{1, 2, \ldots, n\}$. We say that $A \in \mathcal{B}_n$ is an *up-set* (order filter) if $(a \in A, a \subset b) \Rightarrow b \in A$, and a *down-set* (order ideal, simplicial complex) if $(a \in A, b \subset a) \Rightarrow b \in A$. Clearly, A is an up-set if and only if its complement $2^n \setminus A$ is a down-set. We normalize $\mu \ge 0$ so that $\sum\{\mu(a) : a \in 2^n\} = 1$, and view its additive extension μ as a probability measure on $\mathcal{B}_n$. The expected value of f with respect to μ is $E(f, \mu) = \sum_{a\in 2^n} \mu(a)f(a)$.

Theorem 3. (FKG) *Suppose* μ *is a probability measure on* $\mathcal{B}_n$ *and* $\mu(a)\mu(b) \le \mu(a \cup b)\mu(a \cap b)$ *for all* $a, b \in 2^n$. *Then*

(1) $E(f, \mu)E(g, \mu) \le E(fg, \mu)$ *for all nondecreasing* $f, g : 2^n \to \mathrm{R}$;

(2) $\mu(A)\mu(B) \le \mu(A \vee B)\mu(A \wedge B)$ *for all* $A, B \in \mathcal{B}_n$,

(3) $\mu(A \cap B) \ge \mu(A)\mu(B)$ *for all up-sets* $A, B \in \mathcal{B}_n$.

Comments. (1) is tantamount to the inequality of Theorem 2 under normalization. (3) is immediate from (1) by taking $f = 1$ on A, 0 otherwise, and $g = 1$ on B, 0 otherwise. In (2), $A \vee B = \{a \cup b : a \in A, b \in B\}$, which is not generally equal to $A \cup B$. In fact, if A and B are up-sets then $A \vee B = A \cap B$. ∎

An intermediate result between Theorems 1 and 2 was established by Holley [22]. It says that if $\mu_1, \mu_2 : \Gamma \to [0, \infty)$ satisfy $\sum_\Gamma \mu_1(a) = \sum_\Gamma \mu_2(a)$ and

$$\mu_1(a)\mu_2(b) \le \mu_1(a \vee b)\mu_2(a \wedge b) \quad \text{for all} \quad a, b \in \Gamma ,$$

then $\sum_\Gamma \mu_1(a)f(a) \geq \sum_\Gamma \mu_2(a)f(a)$ for every nondecreasing $f : \Gamma \to \mathrm{R}$. The proof by Theorem 1 is similar to the proof of Theorem 2. We add a constant to f to make it positive, define α, β, γ and δ by μ_1, $f\mu_2$, $f\mu_1$ and μ_2, respectively, then use Theorem 1 with $A = B = \Gamma$ to obtain Holley's conclusion. When μ_1 and μ_2 are probability measures on $\mathcal{B}_n$ that satisfy

$$\mu_1(a)\mu_2(b) \leq \mu_1(a \cup b)\mu_2(a \cap b) \quad \text{for all} \quad a, b \in 2^n ,$$

Holley's theorem says that

$$E(f, \mu_1) \geq E(f, \mu_2) \quad \text{for every nondecreasing} \quad f : 2^n \to \mathrm{R} .$$

We mention several further results for $\mathcal{B}_n$.

Theorem 4. ([27, 33]) *Suppose $A, B \in \mathcal{B}_n$. If A is an up-set and B is a down-set, then $2^n|A \cap B| \leq |A||B|$. If both A and B are up-sets or down-sets, then $2^n|A \cap B| \geq |A||B|$.*

Proof. The up-sets conclusion is immediate from Theorem 3(3) on taking $\mu(a) = 2^{-n}$ for each $a \in 2^n$. The other conclusions follow from complementation. ∎

The next theorem involves systems of set differences. Its proof requires a few steps beyond what is immediate from Theorem 1 and could be considered a boundary case between types 1 and 2. For $A, B \in \mathcal{B}_n$, let

$$A - B = \{a \setminus b : a \in A, b \in B\} .$$

Theorem 5. ([30]) *For all $A, B \in \mathcal{B}_n$, $|A - B||B - A| \geq |A||B|$.*

Proof. Let $\mathbf{n} = \{1, 2, \ldots, n\}$. Using Daykin's inequality, we have

$$\begin{aligned} |A||B| &= |A||\{\mathbf{n} \setminus b : b \in B\}| \\ &\leq |A \vee \{\mathbf{n} \setminus b : b \in B\}||A \wedge \{\mathbf{n} \setminus b : b \in B\}| \\ &= |\{a \cup (\mathbf{n} \setminus b) : a \in A, b \in B\}||\{a \cap (\mathbf{n} \setminus b) : a \in A, b \in B\}| \\ &= |\{\mathbf{n} \setminus (a \cup \mathbf{n} \setminus b) : a \in A, b \in B\}||\{a \setminus b : a \in A, b \in B\}| \\ &= |\{b \setminus a : a \in A, b \in B\}||A - B| \\ &= |B - A||A - B| . \; \blacksquare \end{aligned}$$

The implication

$$|A - A| \geq |A|$$

of Theorem 5 is known as the *Marica-Schönheim inequality*. Additional facts about the Marica-Schönheim inequality and close relatives are included in Daykin and Lovász [12], Ahlswede and Daykin [4], Aharoni and Holzman

[1] and Lengvárszky [29]. Although their proofs go well beyond our type 1 designation, we mention some of their results here before we discuss other type 2 implications in the next section. For the following composite theorem, parts (1) and (4) are proved in [1], (2) is proved in [12], and (3) is proved in [4]. In part (1), we say that A is *weakly separating* [1] if for all distinct i and j in $\{1, 2, \ldots, n\}$, $\{a \in A : i \in a\} = \{a \in A : j \in a\}$ implies that both sets equal A or both are empty. In addition, $\mathcal{B}_s$ denotes the family of sets of subsets of s for $s \in 2^n$.

Theorem 6. *Suppose $A, B \in \mathcal{B}_n$.*

(1) If A is weakly separating, then $|A - A| = A$ if only if there is a partition of $\{1, 2, \ldots, n\}$ into s and t, an up-set S in $\mathcal{B}_s$, and a down-set T in $\mathcal{B}_t$ such that $A = \{a \cup b : a \in S, b \in T\}$.

(2) If $|A| \geq 2$ then there is a bijection $\phi : A \to A$ such that $\phi(a) \neq a$ for all $a \in A$, and $a \setminus \phi(a) \neq b \setminus \phi(b)$ for all $a \neq b$ in A.

(3) If for every $a \in A$, $b \subseteq a$ for some $b \in B$, then $|A - B| \geq |A|$.

(4) If for all a, $a' \in A$, $(a \setminus a') \cap b = \emptyset$ for some $b \in B$, then $|A - B| \geq |A|$.

Part (1) essentially covers all cases of equality for the Marica-Schönheim inequality, and (2) is a strengthened version of the inequality for $|A| > 1$. Part (3) provides a first-order generalization of the Marica-Schönheim inequality, and (4) strengthens (3) by weakening its hypothesis.

Lengvárszky [29] proves that an analogue of the Marica-Schönheim inequality holds for $(\Gamma, \prec)$ when $a - b$ for $a, b \in \Gamma$ is defined in a particular way with $A - B = \{a - b : a \in A, b \in B\}$ for $A, B \subseteq \Gamma$. The paper also considers $|A - A| \geq |A|$ when the lattice is not necessarily distributive.

3. TYPE 2 IMPLICATIONS FOR LINEAR EXTENSIONS

We assume throughout this section that $(X, \prec)$ is a finite partially ordered set. We do not assume that $(X, \prec)$ is a lattice, let alone a distributive lattice, so implications of the Ahlswede-Daykin and FKG theorems will involve construction of distributive lattices for application of those theorems.

The section focuses on linear extensions of $(X, \prec)$, where $(X, \prec_0)$ is a *linear extension* of $(X, \prec)$ if $<_0$ linearly orders X and $x \prec y \Rightarrow x <_0 y$ for all $x, y \in X$. We say that $x, y \in X$ are *incomparable* in $(X, \prec)$ if $x \neq y$ and neither $x \prec y$ nor $y \prec x$. We let $\mathcal{L}$ denote the set of all linear extensions of $(X, \prec)$ and set $N = |\mathcal{L}|$. We recall [36] that if x and y are incomparable in $(X, \prec)$ then $x <_0 y$ for some linear extension in $\mathcal{L}$, so $\prec = \cap\{<_0 : (X, <_0) \in \mathcal{L}\}$.

A few other notations are used in the section. We let μ denote the *uniform* probability measure on $2^{\mathcal{L}}$, so $\mu(L) = 1/N$ for every $L \in \mathcal{L}$. We take $(x <_0 y) = \{L \in \mathcal{L} : x <_0 y \text{ in } L\}$, the set of linear extensions in which $x <_0 y$. The probability of $(x <_0 y)$ under μ is $\mu(x <_0 y)$, with $\mu(x <_0 y)+\mu(y <_0 x) = 1$ when $x \neq y$. Clearly, $\mu(x <_0 y) = |(x <_0 y)|/N$. Finally, we denote by $\cap_I(a_i <_0 b_i)$ the set of linear extensions of $(X, \prec)$ in which $a_i <_0 b_i$ is true for every $i \in \{1, 2, \ldots, I\}$.

Our first two results for the equally-likely linear extensions model consider two-part partitions of X from different perspectives. Their conclusion, $\mu(A \cap B) \geq \mu(A)\mu(B)$, expresses nonnegative correlation between the defined events A and B: the joint occurrence of A and B is at least as probable as the product of their separate probabilities. When $\mu(B) > 0$, $\mu(A \cap B) \geq \mu(A)\mu(B)$ says that $\mu(A|B) \geq \mu(A)$, or that A is at least as likely to occur when B occurs as it is unconditionally.

Theorem 7. ([21]) *Suppose $\{X_1, X_2\}$ is a nontrivial partition of X and $\prec$ linearly orders X_i for $i = 1, 2$. Let $A = \cap_I(a_i <_0 b_i)$ and $B = \cap_J(c_j <_0 d_j)$ for some I and J with all $a_i, c_j \in X_1$ and all b_i, $d_j \in X_2$. Then $\mu(A \cap B) \geq \mu(A)\mu(B)$.*

Theorem 8. ([34]) *Suppose $(X, \prec)$ is the union of disjoint nonempty partially ordered sets $(X_1, \prec_1)$ and $(X_2, \prec_2)$, with $\prec = \prec_1 \cup \prec_2$. With A and B as in Theorem 7, $\mu(A \cap B) \geq \mu(A)\mu(B)$.*

The intuition behind the theorems is that all elementary events for A and B have the form $(x_1 <_0 x_2)$ for $x_1 \in X_1$ and $x_2 \in X_2$, so realization of one of A and B should enhance the likelihood of the other. We note, however, that this intuition is tenuous because $\mu(A \cap B) \geq \mu(A)\mu(B)$ can be false *except* when $(X, \prec)$ has specialized structure as in the theorems' hypotheses. Examples in Shepp [34] and Graham [20, p. 122] show how the conclusion fails for other structures.

Proofs based on the FKG theorem appear in [28, 34] for Theorem 7 and in [34] for Theorem 8. We sketch the proof of Theorem 7 to illustrate constructions that lead to FKG.

Let $(X_1, \prec) = \{x_1 \prec x_2 \prec \cdots \prec x_m\}$ and $(X_2, \prec) = \{y_1 \prec y_2 \prec \cdots \prec y_n\}$ with $m, n \geq 1$. Let Γ be the set of all strictly increasing m-tuples of integers from $\{1, 2, \ldots, m+n\}$, and for $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $\beta = (\beta_1, \ldots, \beta_m)$ in Γ define a reflexive relation $\leq^*$ on Γ by

$$\alpha \leq^* \beta \quad \text{if} \quad \alpha_i \leq \beta_i \quad \text{for} \quad i = 1, \ldots, m\,.$$

Also define $\alpha \wedge \beta$ and $\alpha \vee \beta$ componentwise by

$$(\alpha \wedge \beta)_i = \min\{\alpha_i, \beta_i\}, \quad (\alpha \vee \beta)_i = \max\{\alpha_i, \beta_i\}\,.$$

It follows that $(\Gamma, \leq^*)$ is a distributive lattice (reflexive variety).

We next define a log supermodular function v and nondecreasing functions $-f$ and $-g$ on $(\Gamma, \leq^*)$ as follows. Given $\alpha \in \Gamma$, let α^c be the strictly increasing

n-tuple of integers in $\{1, 2, \ldots, m+n\} \setminus \{\alpha_1, \ldots, \alpha_m\}$, and let σ_α denote the bijection from X onto $\{1, 2, \ldots, m+n\}$ defined by

$$\sigma_\alpha(x_i) = \alpha_i \quad (i = 1, \ldots, m); \qquad \sigma_\alpha(y_j) = \alpha_j^c \quad (j = 1, \ldots, n) .$$

Also let $(X, \prec_A)$ and $(X, \prec_B)$ denote the ordered sets in which $\prec_A = \{(a_1, b_1), \ldots, (a_I, b_I)\}$ and $\prec_B = \{(c_1, d_1), \ldots, (c_J, d_J)\}$. We then define $v, f, g : \Gamma \to \{0, 1\}$ by

$v(\alpha) = 1 \;\Leftrightarrow\;$ the arrangement of X by increasing values of σ_α is a linear extension of $(X, \prec)$;

$f(\alpha) = 1 \;\Leftrightarrow\;$ the arrangement of X by increasing values of σ_α is a linear extension of $(X, \prec_A)$;

$g(\alpha) = 1 \;\Leftrightarrow\;$ the arrangement of X by increasing values of σ_α is a linear extension of $(X, \prec_B)$.

Once log supermodularity and monotonicity have been verified, we use Theorem 2 to conclude that

$$\sum_\Gamma v(\alpha) \sum_\Gamma f(\alpha)g(\alpha)v(\alpha) \geq \sum_\Gamma f(\alpha)v(\alpha) \sum_\Gamma g(\alpha)v(\alpha) ,$$

where the left-to-right sums are the *numbers of linear extensions* of $(X, \prec)$, of $(X, \prec)$ compatible with $\prec_A$ and $\prec_B$, of $(X, \prec)$ compatible with $\prec_A$, and of $(X, \prec)$ compatible with $\prec_B$. Division by N^2 gives $\mu(A \cap B) \geq \mu(A)\mu(B)$. ∎

Our next two theorems show that some instances of nonnegative (Theorem 9) and positive (Theorem 10) correlation do not require strong hypotheses like those in Theorems 7 and 8.

Theorem 9. (xyz [35]) *For all* $x, y, z \in X$,

$$\mu((x <_0 y) \cap (x <_0 z)) \geq \mu(x <_0 y)\mu(x <_0 z) .$$

Theorem 10. (xyz [13]) *For all mutually incomparable* $x, y, z \in X$,

$$\mu((x <_0 y) \cap (x <_0 z)) > \mu(x <_0 y)\mu(x <_0 z) .$$

Because the nonstrict inequality of Theorem 9 is easily seen to hold when x, y and z are not mutually incomparable, Theorem 10 can be viewed as a strengthening of Theorem 9. We outline a proof of Theorem 9 that uses a limiting argument similar to that used in [34] to prove Theorem 8, and then comment on a substantially different proof for Theorem 10.

Suppose for Theorem 9 that x, y and z are mutually incomparable. Fix an integer $K > |X|$ and let Γ_K be the set of all nondecreasing α from $(X, \prec)$

into $\{1, 2, \ldots, K\}$. Also define $\leq^*$, $\wedge$ and $\vee$ for $\alpha, \beta \in \Gamma_K$ by $\alpha \leq^* \beta$ if $\alpha(x) \geq \beta(x)$, and $\alpha(t) - \alpha(x) \leq \beta(t) - \beta(x)$ for all $t \in X$,

$$\begin{aligned}(\alpha \wedge \beta)(t) &= \min\{\alpha(t) - \alpha(x), \beta(t) - \beta(x)\} + \max\{\alpha(x), \beta(x)\}\\ (\alpha \vee \beta)(t) &= \max\{\alpha(t) - \alpha(x), \beta(t) - \beta(x)\} + \min\{\alpha(x), \beta(x)\}\,.\end{aligned}$$

Then $(\Gamma_K, \leq^*)$ is a (reflexive) distributive lattice.

Now for $a, b \in X$ let $(a < b)_K = \{\alpha \in \Gamma_K : \alpha(a) \leq \alpha(b)\}$. Then both $(x < y)_K$ and $(x < z)_K$ are up-sets in $(\Gamma_K, \leq^*)$. Indeed, for any $t \neq x$, $(\alpha(x) \leq \alpha(t), \alpha \leq^* \beta) \Rightarrow 0 \leq \alpha(t) - \alpha(x) \leq \beta(t) - \beta(x) \Rightarrow \beta(x) \leq \beta(t)$. This shows that the unusual definition of $\leq^*$ is just right for the up-set calculation. It then follows from Theorem 2 with the uniform measure on Γ_K that

$$\frac{|(x < y)_K \cap (x < z)_K|}{|\Gamma_K|} \geq \frac{|(x < y)_K|}{|\Gamma_K|}\,\frac{|(x < z)_K|}{|\Gamma_K|}\,.$$

As $K \to \infty$, the proportion of $\alpha \in \Gamma_K$ that have $\alpha(a) = \alpha(b)$ for $a \neq b$ goes to 0, and it follows by taking limits in the preceding inequality that $\mu((x <_0 y) \cap (x <_0 z)) \geq \mu(x <_0 y)\mu(x <_0 z)$. ∎

Because the limit argument of the preceding proof works only for nonstrict inequality, a different approach is needed for Theorem 10. The following lemma suffices.

Lemma 1. [13] *Suppose x, y and z are mutually incomparable in $(X, \prec)$, and $|X| = n$. Let $N(abc)$ be the number of linear extensions of $\mathcal{L}$ with $a <_0 b <_0 c$ and let*

$$\lambda = \frac{N(yxz)N(zxy)}{[N(xyz) + N(xzy)][N(yzx) + N(zyx)]}\,.$$

Then $\lambda \leq (n-1)^2/(n+1)^2$ if n is odd, $\lambda \leq (n-2)/(n+2)$ if n is even, and for each $n \geq 3$ some $(X, \prec)$ attains the indicated upper bound on λ.

The bulk of [13] is devoted to the proof of Lemma 1, which features two applications of the Ahlswede-Daykin theorem. The first application uses the preceding embedding technique with $K \to \infty$, and the second involves an optimization step that yields the preceding bounds on λ.

To complete the proof of Theorem 10 let

$$T = \frac{N - N(yxz) - N(zyx)}{N(yzx) + N(zyx)}\,.$$

Also let $N(ab) = |\{L \in \mathcal{L} : a <_0 b \text{ in } L\}|$. Because $N(xy) = N(zxy) + N(xzy) + N(xyz)$ and $N(xz) = N(yxz) + N(xyz) + N(xzy)$, rearrangement gives

$$\frac{N(xy)N(xz)}{N[N(xyz) + N(xzy)]} = \frac{T + \lambda}{T + 1}\,.$$

Then $\lambda < 1$ by Lemma 1, so $\mu(x <_0 y)\mu(x <_0 z) < \mu((x <_0 y) \cap (x <_0 z))$. ■

Fishburn [14, 15] comments further on the strict xyz inequality of Theorem 10. Given $|X| = n$, [14] investigates the maximum value of $(T+\lambda)/(T+1)$, i.e., of the xyz ratio $\mu(x <_0 y)\mu(x <_0 z)/\mu((x <_0 y) \cap (x <_0 z))$, but does not completely solve the problem. In [15], an application of Theorem 10 is used in a proof that determines all ordered sets $(X, \prec)$ on n points that maximize $\mu(x <_0 y) = N(xy)/N$ when x and y lie in an m-point antichain for fixed m with $n \geq m \geq 2$.

The conclusion of the xyz inequality, which can be rewritten as $N(xyz)N \leq N(xy)N(yz)$, or

$$\mu(x <_0 y <_0 z) \leq \mu(x <_0 y)\mu(y <_0 z) ,$$

is *universal* in the sense that it holds for all ordered sets. It is therefore natural to ask about other universal correlational inequalities. For example, is it always true that

$$\mu(x <_0 y <_0 z <_0 w) \leq \mu(x <_0 y <_0 z)\mu(z <_0 w)?$$

The answer here is "no", as seen by the partially ordered set $(\{x, y, z, w, t\}, \prec)$ in which $\prec$ consists of the chain $y \prec t \prec w$ plus $y \prec z$, $x \prec w$ and $x \prec z$. Then $\mu(x <_0 y <_0 z <_0 w) = 1/4$, whereas $\mu(x <_0 y <_0 z)\mu(z <_0 w) = 15/64 < 1/4$.

The theme of universal inequalities has been pushed to the limit in Winkler [39] and Brightwell [8]. To state their theorems, let $\prec_*$ be an asymmetric binary relation on a set Y. Given an ordered set $(X, \prec)$ with $Y \subseteq X$, let

$$\mu(Y, \prec_*) = \frac{|\{(X, <_0) \in \mathcal{L} : \prec_* \subseteq <_0\}|}{N} .$$

The set of *covering pairs* in $(Y, \prec_*)$ is

$$\Delta(Y, \prec_*) = \{(x, y) \in \prec_* : x \prec_* t \prec_* y \quad \text{for no} \quad t \in Y\} .$$

We say that ordered sets $(Y, \prec_1)$ and $(Y, \prec_2)$ are *compatible* if the transitive closure of $\prec_1 \cup \prec_2$ is irreflexive, i.e., if $\prec_1$ and $\prec_2$ are subsets of a common partial order. In terms of μ as defined here, the xyz inequality of Theorem 9 is

$$\mu(\{x, y, z\}, \{(x, y), (x, z)\}) \geq \mu(\{x, y, z\}, \{(x, y)\})\mu(\{x, y, z\}, \{(x, z)\}) .$$

Theorem 11. ([39]) *Suppose $(Y, \prec_1)$ and $(Y, \prec)$ are compatible finite ordered sets. Then* $\prec_2$

$$\mu(Y, \prec_1 \cup \prec_2) \geq \mu(Y, \prec_1)\mu(Y, \prec_2)$$

for every finite ordered set $(X, \prec)$ *with* $Y \subseteq X$ *if and only if, for all* $x, y, a, b \in Y$,

$$\{(x,y) \in \Delta(Y, \prec_1 \cup \prec_2) \setminus \Delta(Y, \prec_2), (a,b) \in \Delta(Y, \prec_1 \cup \prec_2) \setminus \Delta(Y, \prec_1)\} \Rightarrow (x = a \text{ or } y = b) \ .$$

Theorem 12. ([8]) *Suppose* $(Y, \prec_1)$ *and* $(Y, \prec_2)$ *are compatible finite ordered sets. Then*

$$\mu(Y, \prec_1 \cup \prec_2) \leq \mu(Y, \prec_1)\mu(Y, \prec_2)$$

for every finite ordered set $(X, \prec)$ *with* $Y \subseteq X$ *if and only if* $\prec_1 \cap \prec_2 = \emptyset$ *and, for all* $x, y, a, b \in Y$,

$$\{(x,y) \in \Delta(Y, \prec_1), (a,b) \in \Delta(Y, \prec_2)\} \Rightarrow (x = b \text{ or } y = a) \ .$$

The cases of universal nonnegative correlation in Theorem 11 and universal nonpositive correlation in Theorem 12 are extremely limited. The condition of Theorem 11 says that the covering pairs (x, y) and (a, b) must be related as in the xyz hypothesis, i.e., of the form $\{(x, y), (x, z)\}$ or $\{(x, y), (z, y)\}$. The conditions of Theorem 12 seem even more restrictive.

Additional discussion of the universal correlation theme is provided by Brightwell [9].

4. A TYPE 3 IMPLICATION FOR RANDOM PERMUTATIONS

It is well known that certain instances of the conclusions of Theorems 1 and 2 do not require complete satisfaction of their hypotheses. We illustrate the point with the case of match sets of random permutations from [17].

Let σ be a permutation of $\{1, 2, \ldots, n\}$. The *match set* of σ is its set of fixed points

$$M(\sigma) = \{i \in \{1, 2, \ldots, n\} : \sigma(i) = i\} \ .$$

We assume that all $n!$ permutations of $\{1, 2, \ldots, n\}$ are equally likely and let $\mu(a)$ for $a \in 2^n$ denote the probability that $M(\sigma) = a$, with $\mu(A) = \sum\{\mu(a) : a \in A\}$ for $A \in \mathcal{B}_n$. Thus, when exactly $T(a)$ permutations σ have match set a, $\mu(a) = T(a)/n!$.

Theorem 13. ([17]) *For all up-sets* $A, B \in \mathcal{B}_n$,

$$\mu(A \cap B) \geq \mu(A)\mu(B) \ .$$

An easy corollary, similar to the equivalence of (1) and (3) in Theorem 3, says that if f and g are nondecreasing functions from $(2^n, \subset)$ into R, then $E(fg, \mu) \geq E(f, \mu)E(g, \mu)$. However, Theorem 13 is not a direct implication of Theorem 3 because μ is *not* log supermodular. Although $\mu(a)\mu(b) \leq (a \cup b)\mu(a \cap b)$ for most $a, b \in 2^n$, log supermodularity fails when $|a \cup b| = n - 1 > \max\{|a|, |b|\}$. The reason is that no permutation has exactly $n - 1$ fixed points: if $\sigma(i) = i$ for all but one i then $\sigma(i) = i$ for all i. In other words, $\mu(a \cup b) = 0$ when $|a \cup b| = n - 1$.

Despite the breach of log supermodularity, [17] shows how the Ahlswede-Daykin and FKG theorems can be used to prove Theorem 13. we do this by perturbing μ in ways that assign positive probability to $|a| = n - 1$ such that a perturbed μ satisfies the hypotheses of Theorem 1, or satisfies log supermodularity. Given up-sets A and B, the perturbations leave $\mu(A)$, $\mu(B)$ and $\mu(A \cap B)$ unchanged, so the conclusions of Theorems 1 and 3 can be used for these μ values. Unfortunately, our use of perturbations necessitates examination of many special cases, but this may be an unavoidable cost of the perturbation method.

Although our proof of Theorem 13 is very long, a few comments will indicate one way that the Ahlswede-Daykin theorem is involved. With $T(a) = |\{\sigma : M(\sigma) = a\}|$, it is convenient to work with

$$T_i = T(a) \quad \text{when} \quad |a| = n - i\,,$$

so $T_0 = 1$ (only one permutation has a complete match), $T_1 = 0$ (the breach of log supermodularity, $\sum \binom{n}{i} T_i = n!$, and, by inclusion-exclusion,

$$T_i = i! \sum_{j=0}^{i} (-1)^j / j!\,.$$

The full proof of the theorem assumes that it holds for small n ([24] verifies the result for $n \leq 6$) and considers up-sets A and B that contain every a with $|a| = n - 1$ and do not equal 2^n. The proof divides into two main cases that receive different treatments:

Case 1: $\mu(A \cap B) \geq \mu(A)\mu(B)$ if $A \cup B$ contains a singleton;

Case 2: $\mu(A \cap B) \geq \mu(A)\mu(B)$ if $\min\{|a| : a \in A \cup B\} \geq 2$.

The Case 1 proof assumes that $\{1\} \in A$ and uses the FKG theorem and a matching argument in which $b \in B \setminus A$ with $|b| \leq n - 3$ is paired with $b \cup \{1\} \in A \cap B$. The proof for Case 2 uses the Ahlswede-Daykin theorem. Both cases involve perturbations of μ.

In dealing with Case 2, we assume without loss of generality that $A \cap B$ contains all $(n-1)$-sets and work directly with $T(a)$ rather than $\mu(a) = T(a)/n!$. We perturb T to T' on 2^n as follows:

$$T'(a) = \begin{cases} 0 & a = \{1, \ldots, n\} \\ 1/n & |a| = n - 1 \\ T(a) & |a| \leq n - 2\,. \end{cases}$$

This removes weight 1 from $\{1,\ldots,n\}$ and redistributes it evenly over the $(n-1)$-sets. To satisfy the hypothesized inequality of Theorem 1, we first define α and β there by

$$\alpha(a)=\begin{cases}0 & a\notin A\\ T'(a) & a\in A\,,\end{cases}\qquad \beta(b)=\begin{cases}0 & b\notin B\\ T'(b) & b\in B\,.\end{cases}$$

Because all $(n-1)$-sets are in $A\cap B$, we have $\alpha(A)=\mu(A)n!$ and $\beta(B)=\mu(B)n!$. Next, define γ by

$$\gamma(a)=\begin{cases}1/(2n) & a=\{1,\ldots,n\}\\ 0 & a\notin A\cap B\\ T'(a) & \text{otherwise}\,.\end{cases}$$

This gives

$$\gamma(A\vee B)=\gamma(A\cap B)=\frac{1}{2n}+\mu(A\cap B)n!\,,$$

which is slightly greater than $\mu(A\cap B)n!$, so we define $\delta(A\wedge B)$ to be slightly less than $n!$ to make the conclusion of Theorem 1 at A and B agree with $\mu(A\cap B)\geq\mu(A)\mu(B)$. We choose δ constant on sets of fixed cardinality:

$$\delta(a)=\begin{cases}0 & |a|=n\\ 1/n & |a|=n-1\\ 1 & |a|=n-2\\ nT_i & |a|=n-i-1;\ i=2,\ldots,n-2\\ 2nT_{n-2} & |a|=0\,.\end{cases}$$

It follows that, with $\delta_i=\delta(a)$ when $|a|=i$,

$$\delta(A\wedge B)=\sum_{i=0}^{n}\binom{n}{i}\delta_i=2nT_{n-2}+n\sum_{i=1}^{n-3}\binom{n}{i}T_{n-i-1}+\binom{n}{2}+1\,.$$

Given α, β, γ and δ, the Case 2 proof now breaks into a number of subcases for the up-sets A and B that depend on n and $k=n-\min\{|a|:a\in A\cap B\}$. All but a finite number of instances of (k,n) satisfy the hypothesized inequality of Theorem 1, and $\mu(A\cap B)\geq\mu(A)\mu(B)$ is obtained from its conclusion. A few instances here use a further perturbation which increases $\delta_0=2nT_{n-2}$ but leaves all other parts of α through δ unchanged. The instances of (k,n) that do not satisfy the hypotheses of Theorem 1 use other methods to verify $\mu(A\cap B)\geq\mu(A)\mu(B)$.

5. A GENERALIZATION

We conclude by describing a generalization of the Ahlswede-Daykin theorem due to Rinott and Saks [31, 32] and, independently, Aharoni and Keich [2].

The generalization applies to n-tuples $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ in Γ^n for $n \geq 2$, and is identical to the Ahlswede-Daykin theorem when $n = 2$. It is too early to say whether a number of interesting applications will arise for $n \geq 3$, but this seems plausible in view of the usefulness of Theorems 1 and 2.

We assume that $(\Gamma, \prec)$ is a finite distributive lattice and take $n \geq 2$. For each $k \in \{1, \ldots, n\}$, let ϕ_k denote the map from Γ^n into Γ defined by

$$\phi_k(\mathbf{a}) = \vee \{\wedge_{i \in S} a_i : \; S \text{ is a } k\text{-set in } \{1, 2, \ldots, n\}\}$$

for all $\mathbf{a} = (a_1, \ldots, a_n) \in \Gamma^n$. For example, when $n = 3$,

$$\begin{aligned} \phi_1(\mathbf{a}) &= a_1 \vee a_2 \vee a_3 \\ \phi_2(\mathbf{a}) &= (a_1 \wedge a_2) \vee (a_1 \wedge a_3) \vee (a_2 \wedge a_3) \\ \phi_3(\mathbf{a}) &= a_1 \wedge a_2 \wedge a_3 \,. \end{aligned}$$

With 2^Γ the set of subsets of Γ, we extend ϕ_k to $(2^\Gamma)^n$ by letting

$$\phi_k(\mathbf{A}) = \{\phi_k(\mathbf{a}) : \mathbf{a} \in A_1 \times A_2 \times \cdots \times A_n, \mathbf{a} \in \Gamma^n\}$$

for all $\mathbf{A} = (A_1, \ldots, A_n) \in (2^\Gamma)^n$.

Theorem 14. *Suppose* $(\Gamma, \prec)$ *is a finite distributive lattice,* $n \geq 2$, *and* $f_1, \ldots, f_n, g_1, \ldots, g_n : \Gamma \to [0, \infty)$ *satisfy*

$$\prod_{k=1}^{n} f_k(a_k) \leq \prod_{k=1}^{n} g_k(\phi_k(\mathbf{a})) \quad \textit{for all} \quad \mathbf{a} \in \Gamma^n \,.$$

Then

$$\prod_{k=1}^{n} f_k(A_k) \leq \prod_{k=1}^{n} g_k(\phi_k(\mathbf{A})) \quad \textit{for all} \quad \mathbf{A} \in (2^\Gamma)^n \,.$$

The proof in [2] is similar in outline to the proof of Theorem 1 indicated in section 1. It uses $(2^m, \subset)$ in place of $(\Gamma, \prec)$ and proceeds by induction on m after checking the desired result for $m = 0$ and proving it for $m = 1$ with assistance from a result about n-tuples of functions from $\{0, 1\}$ into $[0, \infty)$.

References

[1] Aharoni, R. and R. Holzman (1993). Two and a half remarks on the Marica-Schönheim inequality. *J. London Math. Soc.* (2) 48:385–395.

[2] Aharoni, R. and U. Keich (1996). A generalization of the Ahlswede-Daykin inequality. *Discrete Math.* 152:1–12.

[3] Ahlswede, R. and D. E. Daykin (1978). An inequality for the weights of two families of sets, their unions and intersections. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 43:183–185.

[4] Ahlswede, R. and D. E. Daykin (1979). Inequalities for a pair of maps $S \times S \to S$ with S a finite set. *Math. Z.* 165:267–289.

[5] Birkhoff, G. (1967). *Lattice Theory*, 3rd ed. Providence, RI, Amer. Mathematical Soc.

[6] Bollobás, B. (1986). *Combinatorics*. Cambridge, Cambridge Univ. Press.

[7] Bollobás, B. and G. Brightwell (1990). Parallel selection with high probability. *SIAM J. Discrete Math.* 3:21–31.

[8] Brightwell, G. R. (1985). Universal correlations in finite posets. *Order* 2:129–144.

[9] Brightwell, G. R. (1986). Some correlation inequalities in finite posets. *Order* 2:387–402.

[10] Burton, R. M., Jr. and M. M. Franzosa (1990). Positive dependence properties of point processes. *Ann. Probab.* 18:359–377.

[11] Daykin, D. E. (1977). A lattice is distributive iff $|A||B| \leq |A \vee B||A \wedge B|$. *Nanta Math.* 10:58–60.

[12] Daykin, D. E. and L. Lovász (1976). The number of values of a Boolean function. *J. London Math. Soc.* (2) 12:225–230.

[13] Fishburn, P. C. (1984). A correlational inequality for linear extensions of a poset. *Order* 1:127–137.

[14] Fishburn, P. C. (1986). Maximizing a correlational ratio for linear extensions of posets. *Order* 3:159–167.

[15] Fishburn, P. C. (1991). A note on linear extensions and incomparable pairs. *J. Combin. Theory Ser. A* 56:290–296.

[16] Fishburn, P. C. (1992). Correlation in partially ordered sets. *Discrete Appl. Math.* 39:173–191.

[17] Fishburn, P. C., P. G. Doyle and L. A. Shepp (1988). The match set of a random permutation has the FKG property. *Ann. Probab.* 16:1194–1214.

[18] Fortuin, C. M., P. N. Kasteleyn and J. Ginibre (1971). Correlation inequalities for some partially ordered sets. *Comm. Math. Phys.* 22:89–103.

[19] Graham, R. L. (1982). Linear extensions of partial orders and the FKG inequality, in: I. Rival, ed., *Ordered Sets*. Dordrecht, Reidel. pp. 213–236.

[20] Graham, R. L. (1983). Applications of the FKG inequality and its relatives, in: *Proceedings 12th International Symposium on Mathematical Programming*. Berlin, Springer. pp. 115–131.

[21] Graham, R. L., A. C. Yao and F. F. Yao (1980). Some monotonicity properties of partial orders. *SIAM J. Algebraic Discrete Methods* 1:251–258.

[22] Holley, R. (1974). Remarks on the FKG inequalities. *Comm. Math. Phys.* 36:227–231.

[23] Hwang, F. K. and L. A. Shepp (1987). Some inequalities concerning random subsets of a set. *IEEE Trans. Information Theory* 33:596–598.

[24] Joag-Dev, K. (1985). Association of matchmakers, mimeo, Department of Statistics, University of Illinois.

[25] Joag-Dev, K., L. A. Shepp and R. A. Vitale (1984). Remarks and open problems in the area of the FKG inequality. *IMS Lecture Notes-Monograph Series* 5:121–126.

[26] Kemperman, J. H. B. (1977). On the FKG inequality for measures on a partially ordered space. *Indag. Math.* 39:313–331.

[27] Kleitman, D. J. (1966). Families of non-disjoint sets. *J. Combin. Theory* 1:153–155.

[28] Kleitman, D. J. and J. B. Shearer (1981). Some monotonicity properties of partial orders. *Stud. Appl. Math.* 65:81–83.

[29] Lengvárszky, Z. (1996). The Marica-Schönheim inequality in lattices. *Bull. London Math. Soc.* 28:449–454.

[30] Marica, J. and J. Schönheim (1969). Differences of sets and a problem of Graham. *Canad. Math. Bull.* 12:635–637.

[31] Rinott, Y. and M. Saks (1991). On FKG-type and permanental inequalities, in: M. Shaked and Y. L. Tong, eds., *Proc. 1991 AMS-IMS-SIAM Joint Conf. on Stochastic Inequalities, IMS Lecture Series*.

[32] Rinott, Y. anada M. Saks (n.d.). Correlation inequalities and a conjecture for permanents. *Combinatorica*.

[33] Seymour, P. D. (1973). On incomparable collections of sets. *Mathematika* 20:208–209.

[34] Shepp, L. A. (1980). The FKG property and some monotonicity properties of partial orders. *SIAM J. Algebraic Discrete Methods* 1:295–299.

[35] Shepp, L. A. (1982). The XYZ conjecture and the FKG inequality. *Ann. Probab.* 10:824–827.

[36] Szpilrajn, E. (1930). Sur l'extension de l'ordre partiel. *Fund. Math.* 16:386–389.

[37] van den Berg, J. and U. Fiebig (1987). On a combinatorial conjecture concerning disjoint occurrences of events. *Ann. Probab.* 15:354–374.

[38] van den Berg, J. and H. Kesten (1985). Inequalities with applications to percolation and reliability. *J. Appl. Probab.* 22:556–569.

[39] Winkler, P. M. (1983). Correlation among partial orders. *SIAM J. Algebraic Discrete Methods* 4:1–7.

[40] Winkler, P. M. (1986). Correlation and order. *Contemp. Math.* 57:151–174.