Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 121–126

REMARKS AND OPEN PROBLEMS IN THE AREA OF THE FKG INEQUALITY

BY KUMAR JOAG-DEV,[1] L. A. SHEPP and RICHARD A. VITALE
University of Illinois, Bell Laboratories and Claremont Graduate School

The FKG inequality is an effective device when the requisite assumptions can be verified. Sometimes these have to be approached circuitously. This is discussed with reference to past uses and suggestions for work on the range of applicability. New areas of potential application are also presented.

1. Sufficiency and Necessity of the Conditions for the FKG Inequality. The FKG inequality in its original form (Fortuin, Ginibre and Kasteleyn (1971)) states that if (a) Γ is a distributive lattice i.e. order isomorphic to an algebra of subsets of a set, (b) f and g are increasing on Γ, (c) μ is a positive function on Γ with

$$\mu(x)\mu(y) \leqslant \mu(x \wedge y)\mu(x \vee y) \quad \text{for all } x,y, \tag{1.1}$$

then

$$\Sigma f(x)\mu(x)\Sigma g(y)\mu(y) \leqslant \Sigma f(x)g(x)\mu(x)\Sigma\mu(y). \tag{1.2}$$

A simple example of how the FKG inequality can be used in a combinatorial setting is the following. Suppose A, A_i are fixed subsets of $N = \{1, \ldots, n\}$ and k, k_i are given integers, $i = 1, \ldots, r$. Choose a subset of S of N at random by choosing each element to be in S independently with probability p, fixed. Let $\bar{A}_i = |A_i \cap S|$. Then

$$P[\bar{A} \geqslant k|\bar{A}_i \geqslant k_i, i \leqslant r] = a_r \geqslant P[\bar{A} \geqslant k] = a_0.$$

To prove this let Γ be the set of all subsets S of N ordered by inclusion, and let $f(S) = \chi(\bar{A}_i \geq k_i, i \leq r)$, $g(S) = \chi(\bar{A} \geq k)$, and $\mu(S) = 1$. It is easy to verify that (a)–(c), (1.1) hold and this gives the result. The result may not seem surprising until it is realized that a_r is not always increasing in r. Indeed with $n = 2$, $A = \{1\}$, $A_1 = \{1,2\}$, $A_2 = \{2\}$ with $p = ½$ gives a counterexample since $a_0 = ½ < a_1 = ⅔ > a_2 = ½$. This class of problems was posed by Frank Hwang and will be further developed elsewhere.

We will see that FKG is often hard to apply even when one feels it should apply. This may also be illustrated by Hwang's example: It can be shown by a direct argument that

$$P[\bar{A}_i \geqslant k_i, i \leqslant r|\bar{A} \geqslant k] \geqslant P[\bar{A}_i \geqslant k_i, i \leqslant r|\bar{A} = k],$$

But Shepp does not see just now how to give an FKG proof. The obvious choice $g(S) = \chi(\bar{A} \leqslant k)$, $\mu(S) = \chi(\bar{A} \geqslant k)$, and f as before yields the desired conclusion but (1.1) fails. Is there a reordering of Γ to make an FKG proof?

FKG themselves point out that (1.1) is not necessary and one could assume the alternate condition

$$2\mu(I)\mu(O) \geqslant \Sigma'\mu(x)\mu(y) \tag{1.1'}$$

[1] Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command under Grant No. 82–K–0007. The United States Government is authorized to reproduce and distribute reprints for government purposes not withstanding any copyright notation thereon.

AMS 1980 subject classifications. Primary 60E15, Secondary 62H20, 06D99

Key words and phrases: Association, distributive lattice, FKG inequality, positive quadrant dependence, total ordering.

which may hold without (1.1), where I and O are the extreme elements of Γ and the sum is over all other elements.

Shepp (1982) shows by an example the slack in (a)–(c), (1.1), where by merely redefining $<$ in the lattice one gets to satisfy (a)–(c), (1.1) and hence obtain (1.2) where the "natural" ordering fails to satisfy (a)–(c), (1.1). Note (1.2) does not depend on "$<$". Ahlswede and Daykin (1978) and others (see survey on FKG by Graham (1983)) give more general versions, also not necessary.

The question then arises as to whether it is possible to find conditions which are necessary or at least closer to being necessary. In this regard, it may be interesting to note that for a distributive lattice of length 2, (1.1) and (1.1′) are equivalent and necessary. In this case, there are two possible structures:

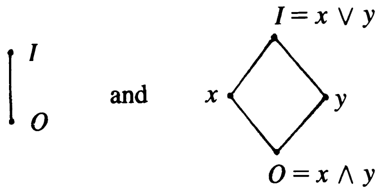

FIGURE 1. Structures for a 2-length lattice.

The first is easily seen. For the second, take two increasing functions f and g with $f(O) = f(y) = g(O) = g(x) = O$ and $f(x) = f(I) = g(y) = g(I) = I$. Then f and g are positively correlated $\Leftrightarrow$ (1.1) $\Leftrightarrow$ (1.1′).

Neither (1.1) nor (1.1′) is necessary for lattices of length greater than 2, but it is interesting to see heuristically why they are sufficient. The clue lies in the assumption of a *distributive* lattice. Among other things, this means that certain sublattice structures do not occur and, in fact, that locally the lattice looks like the pictured length 2 cases. Thus (1.1) and the distributivity assumption are paired to ensure that things work locally.

As mentioned, this approach is generally too strong. One point of departure for an alternate approach is to hold the distributivity assumption in abeyance. This allows previously forbidden sublattice structures, which we can view as lattices in their own right:

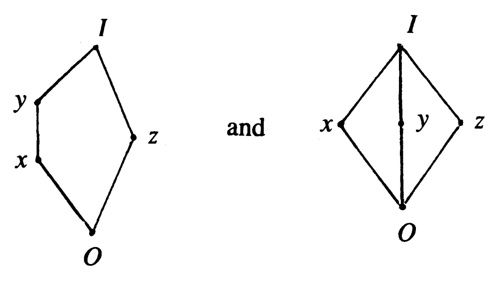

FIGURE 2. Structures for a (nondistributive) lattice.

Consider, for the first structure, increasing f and g with $f(O) = f(z) = g(O) = g(x) = g(y) = O$ and $f(x) = f(y) = f(I) = g(z) = g(I) = 1$. If lattice elements are equiprobable, then $\mathrm{Cov}(f,g) = -{}^1/_5$! A similar example can be given for the second structure. This shows that, for nondistributive lattices, increasing functions are not always positively correlated—at least for equiprobable lattice elements. It would be interesting to see if any general result is possible for the nondistributive case. Subsequent to setting down these remarks, Kemperman's paper (1977, Theorem 7) wich treats the necessity of distributivity was noticed.

The FKG inequality is a powerful device, but it may not be straightforward to use. In fact, as we said, Shepp (1982) treats a problem in which a "natural" ordering fails to satisfy the requirements but another ordering does work. This raises the question of whether there is a systematic way to use the FKG inequality.

One way of formalizing this is as follows:

(Existence) Let a finite set Ω, f, g: $\Omega \to R$, and a probability measure μ on Ω be given. It is desired to find a distributive lattice structure on Ω such that the FKG hypotheses ((a)–(c), (1.1)) hold. When can this be done? Is $\mathrm{Cov}(f,g) \geqslant 0$ close to a condition?

A second question is how long would it take to find a compatible lattice structure.

(Multiplicity) Given the previous set-up (Ω, f, g, μ): what fraction of distributive lattice structures on Ω satisfy the FKG requirements ((a)–(c), (1.1))?

Of course in both of these questions, compatibility of the lattice structure with μ and with the pair (f,g) can be considered separately.

We next discuss two problems to which the FKG inequality seems likely to be able to contribute some insight.

2. A Possible Application to a Partition Problem.

Problem 1. Let a_i, $i = 1, \ldots, m$ be a random sample drawn (without replacement) from $1, 2, \ldots, m+n$ and suppose the remaining numbers are denoted by b_j, $j = 1, \ldots, n$. The sum $S = \Sigma a_i$ has two interpretations.

(i) Let $X_1, \ldots, X_m$; $Y_1, \ldots, Y_n$ be two independent random samples drawn from a single population with a continuous distribution function so that the probability of one or more ties among the observations is zero. Let $R_1, \ldots, R_m$ be the ranks of the X observations among the $m+n$ observations. Then ΣR_i has the same distribution as S, in fact, R_i could be identified with a_i. In this case S is known as the Wilcoxon statistic (with the "null distribution").

(ii) Suppose a_i's are arranged so that $a_1 < a_2 < \ldots < a_m$. Then $a_i - i$ represents the number of b_j's smaller than a_i and it is easy to verify that

$$T = S - m(m+1)/2 = \Sigma_1^m (a_i - i)$$

has a distribution symmetric about $mn/2$ and assumes values, $0, 1, \ldots, mn$.

Further, for an integer k, $0 \leq k \leq mn$,

$$\binom{m+n}{m} P[T=K] = (k;m,n).$$

where $(k;m,n)$ is the number of partitions of k into m parts, with each part $\leqslant n$. In other words, $(k;m,n)$ represents the total number of distinct sequences of non-negative integers $\leqslant n$ and of length m such that each sequence is nondecreasing and the sum of the integers in the sequence is k.

There has been a wealth of literature on the theory of partitions. For the most recent source, see Andrews (1976, Section 3.2).

The problem we are concerned with here is the unimodality of the distribution of T, that is

$$(k;m,n) - (k-1;m,n) \geqslant 0, \quad 0 \leqslant k \leqslant mn/2,$$

which is directly connected with exceedances of a_i over b_j's. Dynkin (1950) proved the unimodality of T; however, the proof is based on the representation theory of Lie algebras. Recently, Hughes (1977) and Stanley (1980, 1981) discuss several problems regarding unimodality of sequences arising from Lie algebras. So far, no direct combinatorial or probabilistic proof is available. One possible approach is to compare the conditional probabilities of T, governed by the order structure of a_i and b_j's while one or more R_i are fixed. Such probability comparisons are similar to those appearing in Shepp (1982). The lattice structure and possible partial orderings are the same as in that paper. Could the FKG inequality again succeed to give a simple proof?

3. A Restriction to Linear Functions.

Problem 2. Let (X_1,X_2) be a pair of real random variables. The conclusion of the FKG inequality for the measure generated by (X_1,X_2) is that

$$\operatorname{cov}[f(X_1,X_2),g(X_1,X_2)] \geq 0, \tag{3.1}$$

for every pair of (co-ordinatewise) nondecreasing functions f,g. This property of positive dependence for (X_1,X_2) was termed as "association" by Esary, Proschan and Walkup (1967). A weaker notion of positive dependence called "positive quadrant dependence" (PQD) is defined by requiring

$$F_{X_1,X_2}(u,v) \geqslant F_{X_1}(u)F_{X_2}(v), \quad \text{for all } (u,v) \tag{3.2}$$

where F with the appropriate subscripts denotes the distribution function. Lehmann (1966) studied PQD and showed that (3.2) is equivalent to

$$\operatorname{cov}(h_1(X_1),h_2(X_2)) \geq 0, \tag{3.3}$$

for every pair of nondecreasing functions h_1,h_2. Although it is easy to show that (3.2) (or (3.3)) does not imply (3.1), if f,g are restricted to linear nondecreasing functions then the implication does hold. This was proved by Shaked (1982). We give a very simple proof. Note that, we want to prove the following:

$$P[X_1>x_1,X_2>x_2] \geqslant P[X_1>x_1]P[X_2>x_2] \quad \text{for all } (x_1,x_2)$$

implies

$$P[\Sigma_1^2 a_iX_i > c,\ \Sigma_1^2 b_iX_i > d] \geq P[\Sigma_1^2 a_iX_i > c]P[\Sigma_1^2 b_iX_i > d].$$

for a_i, b_i nonnegative and arbitrary constants c, d.

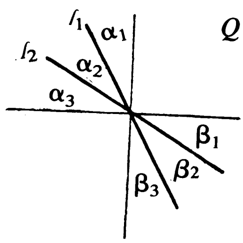

FIGURE 3. Quadrants.

Proof. Let Q, α_i, β_i be the probabilities of the regions as shown, created by interesecting lines l_1, l_2 representing $\Sigma_1^2 a_ix_i = c$ and $\Sigma_1^2 b_ix_i = d$ respectively. We have to show that

$$Q+\alpha_1+\beta_1 \geqslant (Q+\alpha_1+\beta_1+\beta_2)(Q+\alpha_1+\alpha_1+\alpha_2+\beta_1). \tag{3.4}$$

given that

$$Q \geqslant (Q+\Sigma_1^3\alpha_i)(Q+\Sigma_1^3\beta_i). \tag{3.5}$$

However, from (3.5) it is easy to check that

$$Q+\alpha_1+\beta_1 \geqslant (Q+\alpha_1+\Sigma_1^3\beta_i)(Q+\beta_1+\Sigma_1^3\alpha_i)$$

which implies (3.4).

Remark. Notice that in the above proof the quadrant could very easily be replaced by a region defined by an intersection of half planes other than $X_1 > x_1$ and $X_2 > x_2$, the only requirement being that it is *contained* in the intersection of $\Sigma a_i X_i > c$ and $\Sigma b_i X_i > d$.

We will set up an analogy for the measures on lattices. Let Γ be a lattice. Suppose T_1, T_2 are total ordering relations. Consider a partial ordering P induced by T_1, T_2 as follows:

Definition. $x \geq_p y$ *if* $x \geq_{T_1} y$ *and* $x \geq_{T_2} y$.

Given a measure μ on Γ one may define "marginal distribution functions" F_1, F_2 by $F_i(x) = \mu\{y; y \leqslant_{T_i} x\}$ and the PQD analog would be: for every $x \in \Gamma$,

$$\mu\{y{:}y \leqslant_p x\} \geqslant \mu\{y{:}y \leqslant_{T_1} x\} \cdot \mu\{y{:}y \leqslant_{T_2} x\}. \tag{3.6}$$

In view of the remark above one may ask the following: Suppose (T_1^*, T_2^*) is another pair of linear ordering on Γ such that the induced partial order P^* is weaker than P above, that is

$$y \leqslant_P x \Rightarrow y \leqslant_{P^*} x.$$

Under what conditions would (3.6) be sufficient for the validity of an analogous inequality involving P^* and (T_1^*, T_2^*)?

A related multivariate question is the following. Supose $X_1, \ldots, X_k$ are such that for arbitrary nonnegative constants a_i, b_i and an arbitrary proper subset A of $\{1,2, \ldots k\}$, $\Sigma_{i\in A} a_i X_i$ and $\Sigma_{i\in \bar{A}} b_i X_i$ are PQD, where $\bar{A}$ is the complement of A. This property may be called "disjoint positive linear dependence" (DPLD).

QUESTION. Does DPLD $\Rightarrow$ PLD? Here PLD means $\Sigma_1^k a_i X_i$, $\Sigma_1^k b_i X_i$ are PQD. (This problem is related to some concepts discussed in Joag-Dev (1983)).

It is interesting to see that $k = 3$ is the most crucial while $k = 2$ has already been proved. To see that the case $k = 3$ yields the general result, consider

$$Y_1 = X_1,\ Y_2 = \Sigma_{i\in A} a_i X_i,\ Y_3 = \Sigma_{i\in B} b_i X_i,$$

where A, B are *disjoint* and do not containt 1. The triplet Y_1, Y_2, Y_3 is DPLD. If PLD, it will show that the linear combinations containing *one* common variable would be PQD. Using the same technique successively, the cardinality of $A \cap B$ can be increased to k.

To see the relation between these covariance inequalities and FKG, suppose that L is a product lattice of two components L_1 and L_2 with partial ordering P_1 and P_2 respectively. Suppose μ defined on L satisfies FKG condition with respect to the partial ordering induced by (P_1, P_2). Then it follows that for every pair of nondecreasing functions (f, g) defined on L_1 and L_2 respectively,

$$\mathrm{Cov}[f, g] \geq 0. \tag{3.7}$$

However, it is well known that the validity of (3.7) for every pair of nondecreasing functions does not imply FKG inequality. Suppose now we restrict the nondecreasing functions to those which are *linear*, then the above converse seems to be plausible. In fact, it reduces to having DPLD and PLD conditions equivalent.

REFERENCES

AHLSWEDE, R. and DAYKIN, D. E. (1978). An inequality for the weights of two families of sets, their unions and intersections. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 43* 183–185.

ANDREWS, G. (1976). *Theory of Partitions. Encyclopedia of Mathematics and Its Applications.* Addison-Wesley.

DYNKIN, E. B. (1950). Some systems of weights of linear representations of semisimple lie groups. *Dokl. Akad. Nauk. SSSR (N.S.) 71* 221–224.

ESARY, J., PROSCHAN, F. and WALKUP, W. (1967). Associated random variables with applications. *Ann. Math. Statist. 38* 1466–1474.

FORTUIN, C. M., GINIBRE, J, and KASTELEYN, P. N. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math Phys. 22* 89–103.

GRAHAM, R. L. (1983). Applications of the FKG inequality and its relatives. To appear in *Mathematical Programming Bonn 1982 — The State of the Art,* ed. by Bachem, Grötschel, and Korte. Springer-Verlag.

HUGHES, J. W. B. (1977). Lie algebraic proofs of some theorems on partitions. *Number Theory and Algebra.* H. Zassenhaus, ed., Academic Press, New York.

JOAG-DEV, K. (1983) Characterization of independence via uncorrelatedness. *Ann. Probability 11* 1037–1041.

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

SHAKED, M. (1982). A general theory of some positive dependence notions. *J. Multivariate Anal. 12* 199–218.

SHEPP, L. A. (1982). The XYZ conjecture and the FKG inequality. *Ann Probability 10* 824–827.

STANLEY, R. (1980). Unimodal sequences arising from Lie algebras. *Young Day Proceedings,* 127–136. Marcel Dekkar, Inc., N.Y.

STANLEY, R. (1981). Two combinatorial applications of the Aleksandrov-Fenchel inequalities. *J. Comb. Theory 31* 56–65.