The Annals of Applied Probability
1995, Vol. 5, No. 1, 322–326

A TRIANGLE INEQUALITY FOR COVARIANCES OF BINARY FKG RANDOM VARIABLES[1]

BY J. VAN DEN BERG AND A. GANDOLFI

CWI and Universita di Roma Tor Vergata

For binary random variables $\sigma_1, \sigma_2, \dots, \sigma_n$ that satisfy the well-known FKG condition, we show that the variances and covariances satisfy

$$\mathrm{Var}(\sigma_j)\,\mathrm{Cov}(\sigma_i, \sigma_k) \geq \mathrm{Cov}(\sigma_i, \sigma_j)\,\mathrm{Cov}(\sigma_j, \sigma_k), \qquad 1 \leq i, j, k \leq n.$$

This generalizes and improves a result by Graham for ferromagnetic Ising models with nonnegative external fields.

1. Introduction. In several fields, especially statistical mechanics, problems of the following type frequently occur: a stationary sequence $\sigma_0, \sigma_1, \dots$ of binary-valued random variables is given, and one is interested in the existence of the limit, as $n \to \infty$, of $(-1/n) \log \mathrm{Cov}(\sigma_0, \sigma_n)$. (The inverse of such a limit is called *correlation length*.) For instance, each σ_i could be the spin at the vertex $(i,0,\dots,0)$ of a d-dimensional Ising ferromagnet.

In this type of situation inequalities of the form $\mathrm{Cov}(\sigma_i, \sigma_k) \geq \mathrm{Cov}(\sigma_i, \sigma_j)$ $\mathrm{Cov}(\sigma_j, \sigma_k)$ are useful, since they imply subadditivity and hence [if $\mathrm{Cov}(\sigma_0, \sigma_1) \neq 0$] existence of the limit mentioned above. More sophisticated examples of the usefulness of such inequalities, in combination with subadditive ergodic theorems, are given by van Enter and van Hemmen (1983).

For ferromagnetic Ising models with all external fields having the same sign, the inequality mentioned above has been proved by Graham(1982). We show that Graham's result holds for a much larger class of probability distributions. In particular, in the case of Ising ferromagnets, it holds for *arbitrary* external fields.

Before we state our results, we need some preliminaries.

DEFINITION 1. Let $\sigma_1, \sigma_2, \dots, \sigma_n$ be *binary* random variables. The word "binary" here means that each σ_i can take only two possible values, denoted by a_i and b_i. Let Ω denote the product space $\{a_1, b_1\} \times \{a_2, b_2\} \times \cdots \times \{a_n, b_n\}$, and let μ be the distribution of $(\sigma_1, \dots, \sigma_n)$. We say that the collection $\sigma_1, \dots, \sigma_n$ satisfies the FKG condition [see Fortuin, Kasteleyn and Ginibre (1971)] if, for

Received January 1994; revised June 1994.

[1]The research that led to this paper was done in the fall of 1993, when the authors participated in the Programme on Random Spatial Processes at the Isaac Newton Institute for Mathematical Sciences, Cambridge, U.K. Both authors were financially supported by the Isaac Newton Institute.

AMS 1991 *subject classifications*. 60E15, 60K35.

Key words and phrases. FKG inequality, correlation inequalities, Ising model, correlation length.

all $\alpha, \beta \in \Omega$,

$$\mu(\alpha \wedge \beta)\mu(\alpha \vee \beta) \geq \mu(\alpha)\mu(\beta). \tag{1}$$

Here $\alpha \wedge \beta$ is given by $(\alpha \wedge \beta)_i = \min(\alpha_i, \beta_i)$ and $\alpha \vee \beta$ is given by $(\alpha \vee \beta)_i = \max(\alpha_i, \beta_i)$, $i = 1, \dots, n$.

REMARKS. (a) It is well known that if the collection $\sigma_1, \dots, \sigma_n$ satisfies the FKG condition, then every subcollection also satisfies that condition.

(b) The same remark as above holds for the conditional distribution of a subcollection, given the values of the others. More precisely, if the collection $\sigma_1, \dots, \sigma_n$ satisfies the FKG condition, then for all $J \subset \{1, \dots, n\}$ and all $\alpha \in \prod_{i \in J} \{a_i, b_i\}$, the conditional distribution of the collection σ_i, $i \notin J$, given the event $(\sigma_j = \alpha_j,\ j \in J)$, satisfies the FKG condition.

By the FKG theorem we mean the result by Fortuin, Kasteleyn and Ginibre (1971), that if μ satisfies the FKG condition, then for all increasing functions f and g on Ω,

$$E(fg) \geq E(f)E(g). \tag{2}$$

Here E denotes expectation with respect to μ, and a function f is called increasing if $f(\alpha) \geq f(\beta)$ whenever $\alpha_i \geq \beta_i$ for all i.

REMARKS. (i) In particular, the inequality (2) implies that for all $1 \leq i, j \leq n$, $\mathrm{Cov}(\sigma_i, \sigma_j) \geq 0$. This special case of the FKG theorem can also be proved more directly from the definition quite easily.

(ii) The FKG theorem is widely used in statistical mechanics. For a different kind of application, to certain replacement algorithms in computer storage problems, see van den Berg and Gandolfi (1992).

2. Results. Our main result, Theorem 1 below, follows quite easily from the following identity, which holds for *any* triple of binary random variables.

PROPOSITION 1. *Let X, Y, Z be binary random variables. Then*

$$\begin{aligned}
&[\mathrm{Var}(Y)\,\mathrm{Cov}(X,Z) - \mathrm{Cov}(X,Y)\,\mathrm{Cov}(Y,Z)] \\
&\qquad \times [p\ \mathrm{Var}^+(X) + (1-p)\,\mathrm{Var}^-(X)] \\
&\quad = [p\ \mathrm{Cov}^+(X,Z) + (1-p)\,\mathrm{Cov}^-(X,Z)] \\
&\qquad \times [\mathrm{Var}(X)\,\mathrm{Var}(Y) - \mathrm{Cov}^2(X,Y)].
\end{aligned} \tag{3}$$

Here Cov^+ *and* Var^+ (Cov^- *and* Var^-) *denote covariance and variance with respect to the conditional distribution given Y takes the maximal (minimal) of its two possible values, and p is the probability that Y takes its maximal value.*

PROOF. If we increase the two possible values of Y by the same amount, nothing changes in (3). If we multiply them by the same factor, both sides of (3) are multiplied by the square of that factor. Therefore, it is clear that, without loss of generality, we may assume that the two possible values of Y are 0 and 1. Now let μ^+ and μ^- be the conditional distribution given Y takes the value 1 and 0, respectively. Express everything in (3) in the obvious way in terms of p and expectations with respect to μ^+ and μ^-. For instance,

$$\begin{aligned}\mathrm{Var}(Y) &= p(1-p), \mathrm{Cov}(X,Z) = E(XZ) - E(X)E(Z)\\ &= pE^+(XZ) + (1-p)E^-(XZ) - (pE^+(X) + (1-p)E^-(X))\\ &\quad \times (pE^+(Z) + (1-p)E^-(Z)), \mathrm{Cov}(Y,Z) = E(YZ) - E(Y)E(Z)\\ &= pE^+(Z) - p(pE^+(Z) + (1-p)E^-(Z)).\end{aligned}$$

Then work out the multiplications so that both sides of (3) become sums of products of factors p, $1-p$, $E^\bullet(X)$, $E^\bullet(Z)$ and $E^\bullet(XZ)$, where $\bullet$ stands for $+$ or $-$. Then it appears that the l.h.s. and r.h.s. of (3) are indeed equal. □

REMARK. The above proof, although quite laborious, is straightforward. Of course, proving a given identity and obtaining that identity are different things. The above proof gives no insight at all as to how the identity was found. The following sketch does show how it was obtained. It is not meant as an alternative proof of the identity (we skip several details, like existence of inverse functions and differentiability), but only to show what leads one to guess it.

Denote the distribution of (X, Y, Z) by μ. Now define, for $h = (h_1, h_2, h_3) \in R^3$, a new probability distribution μ_h on the same probability space, by

$$\mu_h(x,y,z) = \frac{\exp(h_1x + h_2y + h_3z)\mu(x,y,z)}{N},$$

where N (which depends, of course, on h) is a normalizing constant. Such an operation is very natural in the context of Ising models, where the h_i''s are interpreted as (changes in) the external magnetic fields. We will use properties of the partial derivatives w.r.t. the h_i''s like

$$\frac{\partial}{\partial h_2} E_h(X) = \mathrm{Cov}_h(X,Y), \tag{4}$$

where E_h and Cov_h denote expectation and covariance with respect to μ_h.

Now suppose that we start with $h_1 = h_2 = h_3 = 0$ and then change h_3, but adapt h_1 and h_2 simultaneously so that the expectations of X and Y do not change. Then h_1 and h_2 can be regarded as functions of h_3 and one may ask what the derivatives h_1' and h_2' of these functions (at $h_3 = 0$) are. The calculation of these derivatives can be done in the following two different ways. In the first way, the requirement that the expectations of X and Y remain constant leads [after differentiation and using properties like (4)] to two linear equations in h_1' and h_2'. Solving these gives an expression for h_1'

equal to -1 times the first factor in the l.h.s. of (3) divided by the second factor in the r.h.s. Alternatively, in the second way, we can write

$$E_h(X) = p\, E_h^+(X) + (1-p)E_h^-(X).$$

Since, by construction, $(d/dh_3)p$, $(d/dh_3)E_h(X)$, $(\partial/\partial h_2)E_h^+(X)$ and $(\partial/\partial h_2)E_h^-(X)$ are 0, we now get, after differentiation, one linear equation, in h_1' only. The solution of this equation is -1 times the first factor in the r.h.s. of (3) divided by the second factor in the l.h.s. The equality of the two solutions yields the identity.

Proposition 1 gives the following theorem:

THEOREM 1. *Let $\sigma_1, \dots, \sigma_n$ be a collection of binary random variables satisfying the FKG condition. Then, for any $1 \le i,\ j,\ k \le n$,*

$$\mathrm{Var}(\sigma_j)\,\mathrm{Cov}(\sigma_i, \sigma_k) \ge \mathrm{Cov}(\sigma_i, \sigma_j)\,\mathrm{Cov}(\sigma_j, \sigma_k). \tag{5}$$

PROOF. By the same reasons as in the proof of Proposition 1 above, we may assume, without loss of generality, that each σ_i takes the value 0 or 1. Fix i, j and k and denote σ_i, σ_j and σ_k by X, Y and Z, respectively. Then the triple (X, Y, Z) satisfies the FKG condition [see part (a) of the Remark after Definition 1]. We may assume that each of X, Y and Z is non-degenerate [otherwise both sides of (5) are 0]. By Proposition 1, identity (3) holds. If both $\mathrm{Var}^+(X)$ and $\mathrm{Var}^-(X)$ are 0, then X is apparently "determined" by Y, so $P(X = Y) = 1$ or $P(X = 1 - Y) = 1$. It is easy to see that the latter implies that $\mathrm{Cov}(X, Y)$ is negative and hence is in conflict with the FKG condition, so $P(X = Y) = 1$. However, then both sides of (5) are clearly equal. Concluding, we may assume that the second factor in the l.h.s. of (3) is *strictly* positive. Now we turn to the r.h.s. of (3): the second factor is nonnegative by Cauchy–Schwarz, and the first factor is nonnegative by FKG [see Remarks (a) and (b) after Definition 1 and Remark (i) at the end of Section 1]. Hence the result follows. □

REMARK. It appears from the above proof that the FKG condition is stronger than we really need. Sufficient as that is, if we fix the value of one of the σ_i's, the other random variables have nonnegative covariances (w.r.t. the conditional distribution). However, the well-known property of $\sigma_1, \dots, \sigma_n$ being *associated* is not sufficient in general [μ is associated if it satisfies (2) for all increasing f and g (i.e., if it satisfies the *consequence* of the FKG theorem)]. For instance, let Y_1, Y_2, Y_3, Y_4 be i.i.d. $\{0,1\}$-valued random variables with $P(Y_1 = 0) = P(Y_1 = 1) = 1/2$. Define $X_1 = Y_1Y_2$, $X_2 = Y_2Y_3$ and $X_3 = Y_3Y_4$. Since (Y_1, Y_2, Y_3, Y_4) satifies the FKG condition, it is associated. Hence, since each X_i is an increasing function of (Y_1, Y_2, Y_3, Y_4), (X_1, X_2, X_3) is also associated. However, (5) [with $(\sigma_i, \sigma_j, \sigma_k)$ replaced by (X_1, X_2, X_3)] does not hold, because $\mathrm{Cov}(X_1, X_3) = 0$, while $\mathrm{Cov}(X_1, X_2)$ and $\mathrm{Cov}(X_2, X_3)$ are strictly positive.

COROLLARY 1. *Let* $(\sigma_1, \dots, \sigma_n)$ *be a ferromagnetic Ising system with interactions* $J_{i,j} \geq 0$, $1 \leq i < j \leq n$, *and external fields* $h_1, \dots, h_n$. *For those not familiar with Ising models, the above just means that* $\sigma_1, \dots, \sigma_n$ *are* $\{-1, +1\}$-*valued random variables and their mutual distribution is given by*

$$\mu(\alpha_1, \dots, \alpha_n) = \frac{\exp\left(\sum_{1 \leq i < j \leq n} J_{i,j} \sigma_i \sigma_j + \sum_i h_i \sigma_i\right)}{Z},$$

with Z *a normalizing constant. Our result is that then, for* $1 \leq i, j, k \leq n$,

$$\mathrm{Var}(\sigma_j)\,\mathrm{Cov}(\sigma_i, \sigma_k) \geq \mathrm{Cov}(\sigma_i, \sigma_j)\,\mathrm{Cov}(\sigma_j, \sigma_k). \tag{6}$$

PROOF. It is well known and easy to check that μ satisfies the FKG condition. Hence, the result follows immediately from Theorem 1.

REMARK. The special case of Corollary 1, where all h_is have the same sign, was proved by Graham [see Theorem 1 of Graham (1982)]. In fact his inequality is a little weaker since it does not have the factor $\mathrm{Var}(\sigma_j)$ (which is at most 1) in the l.h.s. His method is Ising-model specific. The main consequence of Corollary 1 is that Graham's condition (that all fields have the same sign) can be dropped.

Acknowledgments. We thank R. Fernández and A. Sokal for referring us to Graham's article and for stimulating remarks, and A. van Enter for informing us of his work with J. L. van Hemmen.

REFERENCES

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

GRAHAM, R. (1982). Correlation inequalities for the truncated two-point function of an Ising ferromagnet. *J. Statist. Phys.* **29** 177–183.

VAN DEN BERG, J. and GANDOLFI, A. (1992). LRU is better than FIFO under the independent reference model. *J. Appl. Probab.* **29** 239–243.

VAN ENTER, A. C. D. and VAN HEMMEN, J. L. (1983). The thermodynamic limit for long-range random systems. *J. Statist. Phys.* **32** 141–152.

CENTRUM VOOR WISKUNDE EN INFORMATICA
KRUISLAAN 413
1098 SJ AMSTERDAM
THE NETHERLANDS

DIPARTIMENTO DI MATEMATICA
UNIVERSITA DI ROMA TOR VERGATA
VIA DELLA RICERCA SCIENTIFICA
00133 ROMA
ITALY