## [§motivate] 1. Motivating, one-dimensional example

If $f$ and $g$ are increasing functions on $\mathbb{R}$ and $\mu$ is a probabilty measure on $\mathcal{B}(\mathbb{R})$ for for $\mu(f^2) < \infty$ and $\mu g^2 < \infty$ then

$$\mu(f)\mu(g) \le \mu(fg)$$

*Proof.* Expand the left-hand side of

$$\mu^x \mu^y \big(f(x) - f(y)\big)\big(g(x) - g(y)\big) \ge 0$$

□

## [§AD] 2. Generalized FKG

For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ define

$$x \vee y = (x_1 \vee y_1, \dots, x_n \vee y_n) \qquad \text{and} \qquad x \wedge y = (x_1 \wedge y_1, \dots, x_n \wedge y_n)$$

Write $x \le y$ to mean $x \vee y = x$. Say that a function $f$ on $\mathbb{R}^n$ is increasing if it is an increasing function in each of its arguments (for fixed values of the other arguments). Equivalently, $f$ is increasing if $f(x) \le f(y)$ whenever $x \le y$.

AD <1> **Theorem.** *Suppose $f_1, \dots, f_4$ are nonnegative, Borel-measurable functions on $\mathcal{X}^n$, where $\mathcal{X} \subseteq \mathbb{R}$, for which*

AD-ineq <2> $$f_1(x) f_2(y) \le f_3(x \vee y) f_4(x \wedge y) \qquad \text{for all } x, y \in \mathcal{X}^n.$$

*Let $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ be a sigma-finite product measure on $\mathcal{B}(\mathcal{X}^n)$. Then*

$$\mu(f_1)\mu(f_2) \le \mu(f_3)\mu(f_4)$$

*Proof.* Integrate out one coordinate at a time, showing that the key inequality is preserved. Write $x = (X, u)$ and $y = (Y, v)$, where $X = (x_1, \dots, x_{n-1})$ and $Y = (y_1, \dots, y_{n-1})$. Define $\widetilde{f}(X) := \mu_n^u f_i(X, u)$. We need to show that

tf <3> $$\widetilde{f}_1(X)\widetilde{f}_2(Y) \le \widetilde{f}_3(X \wedge Y)\widetilde{f}_4(X \vee Y)$$

The left-hand side of <3> equals

$$\begin{aligned}
&\mu_n^u \mu_n^v f_1(X, u) f_2(Y, v) \\
&\qquad = \mu_n^u \mu_n^v \big(\{u = v\} f_1(X, u) f_2(Y, v)\big) \\
&\qquad\qquad + \mu_n^u \mu_n^v \big(\{u < v\} f_1(X, u) f_2(Y, v) + f_1(X, v) f_2(Y, u)\big)
\end{aligned}$$

The right-hand side of <3> equals

$$\begin{aligned}
&\mu_n^u \mu_n^v f_3(X \wedge Y, u) f_4(X \vee Y, v) \\
&\quad = \mu_n^u \mu_n^v \big(\{u = v\} f_3(X \wedge Y, u) f_4(X \wedge Y, v)\big) \\
&\qquad + \mu_n^u \mu_n^v \big(\{u < v\} f_3(X \wedge Y, u) f_4(X \vee Y, v) + f_3(X \wedge Y, v) f_4(X \vee Y, u)\big)
\end{aligned}$$

On the set $\{u = v\}$, inequality <2> gives

$$f_1(X, u) f_2(Y, v) \le f_3(X \wedge Y, u) f_4(X \vee Y, v)$$

On the set $\{u < v\}$,

$$
\begin{aligned}
A &:= f_1(X,u)f_2(Y,v) \le C := f_3(X\wedge Y,u)f_4(X\vee Y,v)\\
B &:= f_1(X,v)f_2(Y,u) \le C\\
AB &= f_1(X,u)f_2(Y,u)f_1(X,v)f_2(Y,v)\\
&\le f_3(X\wedge Y,u)f_4(X\vee Y,u)f_3(X\wedge Y,v)f_4(X\vee Y,v)\\
&= CD \qquad \text{where } D := f_3(X\wedge Y,v)f_4(X\vee Y,u)
\end{aligned}
$$

If we can show that

ABCD <4>

$$A + B \le C + D$$

then the inequality <3> will follow by pointwise inequalities on the integrands. Inequality <4> is just a rearrangement of the inequality

$$
\begin{aligned}
0 \le (1 - A/C)(1 - B/C) &= 1 - (A+B)/C + (AB)/C^2\\
&\le (C + D - A - B)/C
\end{aligned}
$$

□ And so on.

PQ <5> **Corollary.** *Suppose $P$ and $Q$ are probability measures on $\mathcal{B}(\mathcal{X}^n)$ with densities $p = dP/d\mu$ and $q = dQ/d\mu$ with respect to a product measure $\mu$. Suppose*

$$p(x)q(y) \le p(x\wedge y)q(x\vee y) \qquad \text{for all } x, y \in \mathcal{X}^n$$

*Then*

$$Pf \le Qf$$

*for each increasing function $f$ that is both $P$- and $Q$-integrable.*

*Proof.* Without loss of generality $f$ is bounded and nonnegative. [Truncate; recenter; Dominated Convergence.] Define

$$
\begin{aligned}
f_1(x) &= p(x)f(x)\\
f_2(x) &= q(x)\\
f_3(x) &= p(x)\\
f_4(x) &= q(x)f(x)
\end{aligned}
$$

Check that

$$
\begin{aligned}
f_1(x)f_2(y) &= f(x)p(x)q(y)\\
&\le f(x\vee y)p(x\wedge y)q(x\vee y) = f_3(x\wedge y)f_4(x\vee y)
\end{aligned}
$$

□ Invoke Theorem <1>.

P <6> **Corollary.** *Suppose $P$ is a probability measure with a density $p = dP/d\mu$ with respect to a product measure $\mu$, for which*

$$p(x)p(y) \le p(x\wedge y)p(x\vee y) \qquad \text{for all } x, y \in \mathcal{X}^n$$

*If $f$ and $g$ are increasing, $P$-square integrable functions on $\mathcal{X}^n$ then*

$$Pf(x)g(x) \ge (Pf)(Pg)$$

*That is, $f$ and $g$ are positively correlated as random variables under $P$.*

*Proof.* Once again reduce to the case where $f$ is nonnegative. Define

$$
\begin{aligned}
f_1(x) &= p(x)f(x)\\
f_2(x) &= p(x)g(x)\\
f_3(x) &= p(x)\\
f_4(x) &= p(x)f(x)g(x)
\end{aligned}
$$

Check that

$$\begin{aligned} f_1(x)f_2(y) &= f(x)g(y)p(x)p(y) \\ &\le f(x\vee y)g(x\vee y)p(x\wedge y)p(x\vee y) = f_3(x\wedge y)f_4(x\vee y) \end{aligned}$$

□ Invoke Theorem <1>.

[§Ising]

## 3. Application to Ising measures on $\mathbb{Z}^2$

The Ising model gives a joint distribution for an infinite collection of random variables $\{X_i : i \in \mathbb{Z}^2\}$, indexed by the (***sites***) (lattice points) in the lattice $\mathbb{Z}^2$, with each $X_i$ taking values in $\{-1, +1\}$. In fact, the construction of the whole joint distibution is quite subtle. One starts by defining joint (conditional) distributions for $\{X_i : i \in A\}$ for various finite subsets $A$ of $\mathbb{Z}^2$. These distributions satisfy a consistency condition (described in Lemma <14> below) tt enables them to be pasted together to form a joint distribution over all sites in $\mathbb{Z}^2$.

The lattice $\mathbb{Z}^2$ is thought of as the set of vertices in an infinite graph whose edge set $\mathcal{E}$ consists of all all pairs $e = \{i, j\}$ of sites separated by a Eucliden distance 1. For example, the set of neighbors of a site $i = (i_1, i_2)$ is

$$\partial\{i\} := \{(i_1, i_2 + 1), (i_1, i_2 - 1), (i_1 + 1, i_2), (i_1 - 1, i_2)\}.$$

There are four edge with site $i$ as one vertex.

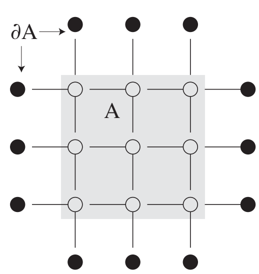
More generally, the boundary $\partial A$ of a set $A \subset \mathbb{Z}^2$ is defined as

$$\partial A := \{j \in A^c : \{i, j\} \in \mathcal{E} \text{ for some } i \text{ in } A\ \}$$

We could also define $\partial A$ as $\cup\{e : e \in \mathcal{E}_A\}\backslash A$, where $\mathcal{E}_A$ denotes the set of all edges $e = \{i, j\}$ for which at least one vertex ($i$, or $j$, or maybe both) is in $A$.

> Remark. The terminology seems a little strange to me, because the boundary of a set in the topological sense is not required to be disjoint from $A$.

neighbors <7> **Example.** Suppose $A$ consists of 9 sites in the form of a $3 \times 3$ grid, the vertices represented by the circles inside the shaded region in the following picture.




The boundary $\partial A$ consists of the 12 sites indicated by the circles filled with black. There are 24 edges in $\mathcal{E}_A$: 12 are between pairs of sites in $A$ and 12 are between a site in $A$ and a site in $\partial A$. □

For each $\beta > 0$ and each $b \in \{-1, +1\}^{\partial A}$ and $x_A \in \{-1, +1\}^A$, define the (conditional) probability that $X_A$ equals $x_A$ by

$$\mathbb{P}_A\{X_A = x_A \mid X_{\partial A} = b\} := p_A(x_A \mid b) := \frac{1}{Z_A(b)} \prod_{\{i,j\}\in\mathcal{E}_A} \exp(\beta x_i x_j)$$

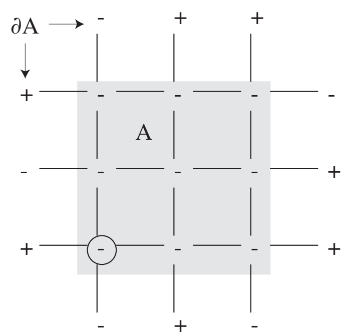
where the standardizing constant,

$$Z_A(b) := \sum_{x_A \in \{-1,+1\}^A} \prod_{\{i,j\} \in \mathcal{E}_A} \exp(\beta x_i x_j),$$

ensures that $\sum_{x_A} p_A(x_A \mid b) = 1$ for each choice of $b$.

Notice that $p_A(\cdot \mid b)$ could be thought of as a density with respect to counting measure $\mu$ on $\{-1,+1\}^A$ and that $\mu$ is a product of the counting measures $\mu_i$ on each of the coordinate subspaces $\{-1,+1\}$.

9prob <8> **Example.** Consider the following configuration, where a plus sign (+) at site $i$ indicates $X_i = +1$ and a minus sign (−) indicates $X_i = -1$. The $x_A$ pattern consists of all $-1$'s.




To calculate $Z_A(b)$, even for the particular $b$ shown, we would have to sum over the $2^9$ values that $X_A$ could take. Too much work. Let me instead calculate $p_A(x_A \mid b)$ up to a constant of proportionality, for the configurations shown.

Each of the 12 edges between sites in $A$ contributes $\beta(-1)(-1) = \beta$ to the exponent. The 12 edges between sites in $A$ and sites in $\partial A$ contribute

$$5\beta(-1)(-1) + 7\beta(-1)(+1) = -2\beta.$$

Thus

$$p_A(x_A \mid b) = e^{10\beta}/Z_A(b)$$

Now suppose that the value $-1$ at the south-west corner of $A$ (the site that is circled) were changed to a $+1$, giving a new $\overline{x}_A$ consisting of eight $-1$'s and one $+1$. What is the value of $p_A(\overline{x}_A \mid b)$? Only contributions for edges with the south-west site as one vertex can change. For $x_A$ the contribution to $p_A(x_A, b)$ was

$$\beta(-1)\big((-1) + (-1) + (-1) + (+1)\big) = 2\beta$$

For $\overline{x}_A$ the contribution changes to

$$\beta(+1)\big((-1) + (-1) + (-1) + (+1)\big) = -2\beta$$

Thus

$$p_A(\overline{x}_A \mid b) = e^{6\beta}/Z_A(b)$$

Notice that the ratio $p_A(\overline{x}_A \mid b)/p_A(x_A \mid b)$ depends only on the values $\{x_j : j \in \partial\{i\}\}$ at the neighboring sites. □

In general, if a configuration $x_A$ has $x_i = -1$ then the edges in $\mathcal{E}_{\{i\}}$ contribute

Npm <9> $$\beta x_i N_i = \beta x_i (N_i^+ - N_i^-) \qquad \text{where } \begin{cases} N_i^+ = \sum_{j \in \partial\{i\}} \{x_j = +1\} \\ N_i^- = \sum_{j \in \partial\{i\}} \{x_j = -1\} \end{cases}.$$

That is, we have only to count the number $N_i^+$ of neighbors of site $i$ where $x_j = +1$ and the number $N_i^-$ of neighbors of site $i$ where $x_j = -1$. Of course,

$N_i^+ + N_i^- = 4$, the total number of neighbors of $\{i\}$ in $\mathbb{Z}^2$. If we define a new configuration $\overline{x}_A$ by changing the value only at the site $i$, then that site will contribute $\beta(\overline{x}_i)N_i$ to the exponent defining $p_A(\overline{x}_A \mid b)$, a change of $2\beta N_i$. That is,

flip <10> $$\frac{p_A(\overline{x}_A \mid b)}{p_A(x_A \mid b)} = \exp(2\beta N_i) \qquad \text{if } \begin{cases} \overline{x}_i = +1 = -x_i \\ \overline{x}_j = x_j & \text{for } j \in A\backslash\{i\} \end{cases}$$

Repeated appeals to this simple formula will allow us to verify the conditions of Corollary <5> for two distributions defined by different boundary conditions.

fixedA <11> **Lemma.** *For a fixed finite subset $A$ of $\mathbb{Z}^2$, let $b$ and $b^*$ be two possible boundary conditions for which $b \le b^*$, that is, $b_j \le b_j^*$ for all $j \in \partial A$. Let $f$ be an increasing function on $\{-1,+1\}^A$. Then*

$$\mathbb{P}_A\left(f(X_A) \mid X_{\partial A} = b\right) \le \mathbb{P}_A\left(f(X_A) \mid X_{\partial A} = b^*\right)$$

*Proof.* For simplicity of notation, write $p(x)$ for $p_A(x_A \mid b)$ and $p^*(x)$ for $p_A(x_A \mid b^*)$. From Corollary <5> it enough if we show that

$$p(x)p^*(y) \le p(x \wedge y)p^*(x \vee y) \qquad \text{for all } x, y \in \{-1,+1\}^A.$$

Equivalently, we need to show

ppstar <12> $$\frac{p(x)}{p(x \wedge y)} \le \frac{p^*(x \vee y)}{p^*(y)}$$

Notice that $x \wedge y \le x$ and $y \le x \vee y$. Moreover,

$$\{i : (x \wedge y)_i < x_i\} = D := \{i \in A : x_i = +1, y_i = -1\} = \{i : y_i < (x \vee y)_i\}$$

For convenience, label the sites in $D$ as $1, 2, \ldots, k$. Let $y^{(i)}$ be the configuration obtained from $y$ by changing the values $y_1, y_2, \ldots, y_i$ to $+1$ and let $y^{(0)} = y$. Note that

$$x \wedge y^{(k)} = x \qquad \text{and} \qquad y^{(k)} = x \vee y$$

| | $x_i = -1$ | $x_i = +1$ |
|---|---|---|
| $y_i = -1$ | $x_i \wedge y_i = -1$<br>$x_i \vee y_i = -1$ | $x_i \wedge y_i = -1$<br>$x_i \vee y_i = +1$<br>$y_i = -1$, $y_i^{(k)} = +1$ |
| $y_i = +1$ | $x_i \wedge y_i = -1$<br>$x_i \vee y_i = +1$ | $x_i \wedge y_i = +1$<br>$x_i \vee y_i = +1$ |

It will suffice if we can show that

i.flip <13> $$\frac{p(x \wedge y^{(i)})}{p(x \wedge y^{(i-1)})} \le \frac{p^*(x \vee y^{(i)})}{p^*(y^{(i-1)})} \qquad \text{for } i = 1, 2, \ldots, k,$$

for then inequality <12> will follow by taking products.

Consider a fixed $i$. Let $N_i$ be calculated as in <9> for site $i$ using boundary $b$ and values $x \wedge y^{(i-1)}$ in $A$. Let $N_i^*$ be calculated similarly using $b^*$ and $y^{(i-1)}$. The ratio on the left-hand side of <13> equals $\exp(2\beta N_i)$; the ratio on the right-hand side equals $\exp(2\beta N_i^*)$. As $b \le b^*$ and $x \wedge y^{(i-1)} \le y^{(i-1)}$, we must have $N_i \le N_i^*$, from which <13> follows. □

The Lemma shows that $\mathbb{P}\left(f(X_A) \mid X_{\partial A} = b\right)$ is maximized by the $b^+$ consisting of all $+1$'s and maximized by the $b^-$ consisting of all $-1$'s. In particular, this assertion holds when $f_\Lambda(x_A) := \{x_i = +1 : i \in \Lambda\}$ for some subset $\Lambda$ of sites in $A$.

Now consider a similar calculation with the same $f_\Lambda$ but with a $\mathbb{P}_B$ constructed from a larger finite region $B \supset A$ under boundary condition $X_{\partial B} =$

$\rho^+$, with $\rho^+$ consisting of all $+1$'s. The following Lemma will show that the conditional probability of $X_A = x_A$ given $X_{\partial A} = b$ is the same under $\mathbb{P}_B(\cdot \mid X_{\partial B} = \pi)$ is the same as $\mathbb{P}_A(\cdot \mid X_{\partial A} = b)$. It will then follow that

$$\begin{aligned}
&\mathbb{P}_B\left(f_\Lambda(X_\Lambda) \mid X_{\partial B} = \rho^+\right) \\
&\quad = \sum_{b\in\{-1,+1\}^{\partial A}} \mathbb{P}_B\left(X_{\partial A} = b \mid X_{\partial B} = \rho^+\right) \mathbb{P}_A\left(f_\Lambda(X_\Lambda) \mid X_{\partial A} = b\right) \\
&\quad \le \mathbb{P}_A\left(f_\Lambda(X_\Lambda) \mid X_{\partial A} = b^+\right) \sum_{b\in\{-1,+1\}^{\partial A}} \mathbb{P}_B\left(X_{\partial A} = b \mid X_{\partial B} = \rho^+\right) \\
&\quad = \mathbb{P}_A\left(f_\Lambda(X_\Lambda) \mid X_{\partial A} = b^+\right)
\end{aligned}$$

That is, under boundary values set to $+1$'s, the conditional expectation of $f_\Lambda(X_\Lambda)$ decreases as region $A$ expands up to the whole of $\mathbb{Z}^+$. See Kindermann & Snell (1980, Appendix 1) for the arguments leading from here to the conclusion that the conditional distributions for each $X_\Lambda$ converge under such a limit.

DLR <14> **Lemma.** *Suppose $A \subset B$ are two finite subsets of $\mathbb{Z}^2$. Then for every $x_A$, $b$ and $\rho$,*

$$\mathbb{P}_B\left(X_A = x_A \mid X_{\partial A} = b, X_{\partial B} = \rho\right) = \mathbb{P}_A\left(X_A = x_A \mid X_{\partial A} = b\right)$$

*Proof.* The left-hand side of the assserted equality equals the ratio

ratio <15>
$$\frac{Z_B(\rho)\mathbb{P}_B\left(X_A = x_A, X_{\partial A} = b \mid X_{\partial B} = \rho\right)}{Z_B(\rho)\mathbb{P}_B\left(X_{\partial A} = b \mid X_{\partial B} = \rho\right)}$$

To simplify notation, if $e = \{i, j\} \in \mathcal{E}$ write $\Psi_e(x_e)$ for $\exp(\beta x_i x_j)$ and let $x_B = (x_A, b)$. Then the numerator in <15> equals

$$\prod_{e\in\mathcal{E}_B} \Psi_e(x_e) = \prod_{e\in\mathcal{E}_B\backslash\mathcal{E}_A} \Psi_e(x_e) \prod_{e\in\mathcal{E}_A} \Psi_e(x_e)$$

The denominator in <15> equals

$$\sum_{x_A}\prod_{e\in\mathcal{E}_B} \Psi_e(x_e) = \prod_{e\in\mathcal{E}_B\backslash\mathcal{E}_A} \Psi_e(x_e) \sum_{x_A}\prod_{e\in\mathcal{E}_A} \Psi_e(x_e)$$

because no $x_e$ for $e \in \mathcal{E}_B\backslash\mathcal{E}_A$ depends on the coordinates for sites in $A$. The final sum equals $Z_A(b)$. The leading product cancels from the ratio, leaving the ratio that defines $\mathbb{P}_A\left(X_A = x_A \mid X_{\partial A} = b\right)$. □

[§notes]
## 4. Notes

Theorem <1> is due to Ahlswede & Daykin (1978), but the proof comes from Karlin & Rinott (1980). Eaton (1986, Chapter 5) contains a nice exposition.

The original paper of Fortuin, Kasteleyn & Ginibre (1971) stated the result of Corollary <6> for increasing functions defined on a finite distributive lattice. It also contained applications to Physics, including the Ising model.

Preston (1974*a*) noted that finite distributive lattices can always be represented as a collection ofsubsets of some finite set. Equivalently, the points of such a lattice can be represented as $n$-tuples of 0's and 1's, or as $n$-tuples of $\pm 1$'s. Preston (1974*b*, Chapter 3) reproduced a proof Holley (1974), which was expressed as a coupling of two probability measures satisfying the setwise analog of the condition in Corollary <5>. In fact, a general coupling result of Strassen (1965) shows that the Holley result is equivalent to the result asserted by the Corollary.

I need to get the history of FKG straight.

See the survey by Den Hollander & Keane (1986) for more about the history of the FKG inequality and its variants.

See Georgii (1988, Chapter 6) for a detailed, rigorous analysis of the Ising model.

## References


Ahlswede, R. & Daykin, D. (1978), 'An inequality for the weights of two families of sets, their unions and intersections', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **43**, 183–185.

Den Hollander, W. T. F. & Keane, M. (1986), 'Inequalities of FKG type', *Physica* **138A**, 167–182. (Pages 176–177 missing from copy at journal web site.).

Eaton, M. (1986), *Lectures on topics in probability inequalities*, Centre for Mathematics and Computer Science, Amsterdam. CWI tract 35.

Fortuin, C. M., Kasteleyn, P. W. & Ginibre, J. (1971), 'Correlation inequalities on some partially ordered sets', *Communications in Mathematical Physics[?]* **22**, 89–103.

Georgii, H.-O. (1988), *Gibbs Measures and Phase Trasitions*, Walter de Gruyter.

Holley, R. (1974), 'Remarks on the FKG inequalities', *Communications in Mathematical Physics[?]* **36**, 227–231.

Karlin, S. & Rinott, Y. (1980), 'Classes of orderings of measures and related correlation inequalities: 1. Multivariate totally positive distributions', *Journal of Multivariate Analysis* **10**, 467–498.

Kindermann, R. & Snell, J. L. (1980), *Markov Random Fields and their Applications*, Vol. 1 of *Contemporary Mathematics*, American Mathematical Society.

Preston, C. J. (1974*a*), 'A generalization of the FKG inequalities', *Communications in Mathematical Physics[?]* **36**, 233–241.

Preston, C. J. (1974*b*), *Gibbs States on Countable Sets*, Cambridge University Press.

Strassen, V. (1965), 'The existence of probability measures with given marginals', *Annals of Mathematical Statistics* **36**, 423–439.