
Gibbs states

Brook's theorem states that a positive probability measure on a finite product may be decomposed into factors indexed by the cliques of its dependency graph. Closely related to this is the well known fact that a positive measure is a spatial Markov field on a graph G if and only if it is a Gibbs state. The Ising and Potts models are introduced, and the n -vector model is mentioned.

7.1 Dependency graphs

Let $X = (X_1, X_2, \dots, X_n)$ be a family of random variables on a given probability space. For $i, j \in V = \{1, 2, \dots, n\}$ with $i \neq j$, we write $i \perp j$ if: X_i and X_j are independent *conditional* on $(X_k : k \neq i, j)$. The relation \perp is thus symmetric, and it gives rise to a graph G with vertex set V and edge-set $E = \{(i, j) : i \not\perp j\}$, called the *dependency graph* of X (or of its law). We shall see that the law of X may be expressed as a product over terms corresponding to complete subgraphs of G . A complete subgraph of G is called a *clique*, and we write \mathcal{K} for the set of all cliques of G . For notational simplicity later, we designate the empty subset of V to be a clique, and thus $\emptyset \in \mathcal{K}$. A clique is *maximal* if no strict superset is a clique, and we write \mathcal{M} for the set of maximal cliques of G .

We assume for simplicity that the X_i take values in some countable subset S of the reals \mathbb{R} . The law of X gives rise to a probability mass function π on S^n given by

$$\pi(\mathbf{x}) = \mathbb{P}(X_i = x_i \text{ for } i \in V), \quad \mathbf{x} = (x_1, x_2, \dots, x_n) \in S^n.$$

It is easily seen by the definition of independence that $i \perp j$ if and only if π may be factorized in the form

$$\pi(\mathbf{x}) = g(x_i, U)h(x_j, U), \quad \mathbf{x} \in S^n,$$

for some functions g and h , where $U = (x_k : k \neq i, j)$. For $K \in \mathcal{K}$ and $\mathbf{x} \in S^n$, we write $\mathbf{x}_K = (x_i : i \in K)$. We call π *positive* if $\pi(\mathbf{x}) > 0$ for all $\mathbf{x} \in S^n$.

In the following, each function f_K acts on the domain S^K .

7.1 Theorem [54]. Let π be a positive probability mass function on S^n . There exist functions $f_K : S^K \rightarrow [0, \infty)$, $K \in \mathcal{M}$, such that

$$(7.2) \quad \pi(\mathbf{x}) = \prod_{K \in \mathcal{M}} f_K(\mathbf{x}_K), \quad \mathbf{x} \in S^n.$$

In the simplest non-trivial example, let us assume that $i \perp j$ whenever $|i - j| \geq 2$. The maximal cliques are the pairs $\{i, i + 1\}$, and the mass function π may be expressed in the form

$$\pi(\mathbf{x}) = \prod_{i=1}^{n-1} f_i(x_i, x_{i+1}), \quad \mathbf{x} \in S^n,$$

so that X is a Markov chain, whatever the direction of time.

Proof. We shall show that π may be expressed in the form

$$(7.3) \quad \pi(\mathbf{x}) = \prod_{K \in \mathcal{K}} f_K(\mathbf{x}_K), \quad \mathbf{x} \in S^n,$$

for suitable f_K . Representation (7.2) follows from (7.3) by associating each f_K with some maximal clique K' containing K as a subset.

A representation of π in the form

$$\pi(\mathbf{x}) = \prod_r f_r(\mathbf{x})$$

is said to *separate* i and j if every f_r is a constant function of either x_i or x_j , that is, no f_r depends non-trivially on both x_i and x_j . Let

$$(7.4) \quad \pi(\mathbf{x}) = \prod_{A \in \mathcal{A}} f_A(\mathbf{x}_A)$$

be a factorization of π for some family \mathcal{A} of subsets of V , and suppose that i, j satisfies: $i \perp j$, but i and j are not separated in (7.4). We shall construct from (7.4) a factorization that separates every pair r, s that is separated in (7.4), and in addition separates i, j . Continuing by iteration, we obtain a factorization that separates every pair i, j satisfying $i \perp j$, and this has the required form (7.3).

Since $i \perp j$, π may be expressed in the form

$$(7.5) \quad \pi(\mathbf{x}) = g(x_i, U)h(x_j, U)$$

for some g, h , where $U = (x_k : k \neq i, j)$. Fix $s, t \in S$, and write $h|_t$ (respectively, $h|_{s,t}$) for the function $h(\mathbf{x})$ evaluated with $x_j = t$ (respectively, $x_i = s, x_j = t$). By (7.4),

$$(7.6) \quad \pi(\mathbf{x}) = \pi(\mathbf{x})|_t \frac{\pi(\mathbf{x})}{\pi(\mathbf{x})|_t} = \left(\prod_{A \in \mathcal{A}} f_A(\mathbf{x}_A)|_t \right) \frac{\pi(\mathbf{x})}{\pi(\mathbf{x})|_t}.$$

By (7.5), the ratio

$$\frac{\pi(\mathbf{x})}{\pi(\mathbf{x})|_t} = \frac{h(x_j, U)}{h(t, U)}$$

is independent of x_i , so that

$$\frac{\pi(\mathbf{x})}{\pi(\mathbf{x})|_t} = \prod_{A \in \mathcal{A}} \frac{f_A(\mathbf{x}_A)|_s}{f_A(\mathbf{x}_A)|_{s,t}}.$$

By (7.6),

$$\pi(\mathbf{x}) = \left(\prod_{A \in \mathcal{A}} f_A(\mathbf{x}_A)|_t \right) \left(\prod_{A \in \mathcal{A}} \frac{f_A(\mathbf{x}_A)|_s}{f_A(\mathbf{x}_A)|_{s,t}} \right)$$

is the required representation, and the claim is proved. \square

7.2 Markov and Gibbs random fields

Let $G = (V, E)$ be a finite graph, taken for simplicity without loops or multiple edges. Within statistics and statistical mechanics, there has been a great deal of interest in probability measures having a type of ‘spatial Markov property’ given in terms of the neighbour relation of G . We shall restrict ourselves here to measures on the sample space $\Sigma = \{0, 1\}^V$, while noting that the following results may be extended without material difficulty to a larger product S^V , where S is finite or countably infinite.

The vector $\sigma \in \Sigma$ may be placed in one–one correspondence with the subset $\eta(\sigma) = \{v \in V : \sigma_v = 1\}$ of V , and we shall use this correspondence freely. For any $W \subseteq V$, we define the *external boundary*

$$\Delta W = \{v \in V : v \notin W, v \sim w \text{ for some } w \in W\}.$$

For $s = (s_v : v \in V) \in \Sigma$, we write s_W for the sub-vector $(s_w : w \in W)$. We refer to the configuration of vertices in W as the ‘state’ of W .

7.7 Definition. A probability measure π on Σ is said to be *positive* if $\pi(\sigma) > 0$ for all $\sigma \in \Sigma$. It is called a *Markov (random) field* if it is positive and: for all $W \subseteq V$, conditional on the state of $V \setminus W$, the law of the state of W depends only on the state of ΔW . That is, π satisfies the *global Markov property*

$$(7.8) \quad \pi(\sigma_W = s_W \mid \sigma_{V \setminus W} = s_{V \setminus W}) = \pi(\sigma_W = s_W \mid \sigma_{\Delta W} = s_{\Delta W}),$$

for all $s \in \Sigma$, and $W \subseteq V$.

The key result about such measures is their representation in terms of a ‘potential function’ ϕ , in a form known as a Gibbs random field (or sometimes ‘Gibbs state’). Recall the set \mathcal{K} of cliques of the graph G , and write 2^V for the set of all subsets (or ‘power set’) of V .

7.9 Definition. A probability measure π on Σ is called a *Gibbs (random) field* if there exists a ‘potential’ function $\phi : 2^V \rightarrow \mathbb{R}$, satisfying $\phi_C = 0$ if $C \notin \mathcal{K}$, such that

$$(7.10) \quad \pi(B) = \exp\left(\sum_{K \subseteq B} \phi_K\right), \quad B \subseteq V.$$

We allow the empty set in the above summation, so that $\log \pi(\emptyset) = \phi_\emptyset$.

Condition (7.10) has been chosen for combinatorial simplicity. It is not the physicists’ preferred definition of a Gibbs state. Let us define a *Gibbs state* as a probability measure π on Σ such that there exist functions $f_K : \{0, 1\}^K \rightarrow \mathbb{R}$, $K \in \mathcal{K}$, with

$$(7.11) \quad \pi(\sigma) = \exp\left(\sum_{K \in \mathcal{K}} f_K(\sigma_K)\right), \quad \sigma \in \Sigma.$$

It is immediate that π satisfies (7.10) for some ϕ whenever it satisfies (7.11). The converse holds also, and is left for Exercise 7.1.

Gibbs fields are thus named after Josiah Willard Gibbs, whose volume [95] made available the foundations of statistical mechanics. A simplistic motivation for the form of (7.10) is as follows. Suppose that each state σ has an energy E_σ , and a probability $\pi(\sigma)$. We constrain the average energy $E = \sum_\sigma E_\sigma \pi(\sigma)$ to be fixed, and we maximize the entropy

$$\eta(\pi) = - \sum_{\sigma \in \Sigma} \pi(\sigma) \log_2 \pi(\sigma).$$

With the aid of a Lagrange multiplier β , we find that

$$\pi(\sigma) \propto e^{-\beta E_\sigma}, \quad \sigma \in \Sigma.$$

The theory of thermodynamics leads to the expression $\beta = 1/(kT)$ where k is Boltzmann’s constant and T is (absolute) temperature. Formula (7.10) arises when the energy E_σ may be expressed as the sum of the energies of the sub-systems indexed by cliques.

7.12 Theorem. A positive probability measure π on Σ is a Markov random field if and only if it is a Gibbs random field. The potential function ϕ corresponding to the Markov field π is given by

$$\phi_K = \sum_{L \subseteq K} (-1)^{|K \setminus L|} \log \pi(L), \quad K \in \mathcal{K}.$$

A positive probability measure π is said to have the *local Markov property* if it satisfies the global property (7.8) for all *singleton* sets W and all $s \in \Sigma$. The global property evidently implies the local property, and it turns out that the two properties are equivalent. For notational convenience, we denote a singleton set $\{w\}$ as w .

7.13 Proposition. *Let π be a positive probability measure on Σ . The following three statements are equivalent:*

- (a) π satisfies the global Markov property,
- (b) π satisfies the local Markov property,
- (c) for all $A \subseteq V$ and any pair $u, v \in V$ with $u \notin A, v \in A$ and $u \approx v$,

$$(7.14) \quad \frac{\pi(A \cup u)}{\pi(A)} = \frac{\pi(A \cup u \setminus v)}{\pi(A \setminus v)}.$$

Proof. First, assume (a), so that (b) holds trivially. Let $u \notin A, v \in A$, and $u \approx v$. Applying (7.8) with $W = \{u\}$ and, for $w \neq u, s_w = 1$ if and only if $w \in A$, we find that

$$(7.15) \quad \begin{aligned} \frac{\pi(A \cup u)}{\pi(A) + \pi(A \cup u)} &= \pi(\sigma_u = 1 \mid \sigma_{V \setminus u} = A) \\ &= \pi(\sigma_u = 1 \mid \sigma_{\Delta u} = A \cap \Delta u) \\ &= \pi(\sigma_u = 1 \mid \sigma_{V \setminus u} = A \setminus v) \quad \text{since } v \notin \Delta u \\ &= \frac{\pi(A \cup u \setminus v)}{\pi(A \setminus v) + \pi(A \cup u \setminus v)}. \end{aligned}$$

Equation (7.15) is equivalent to (7.14), whence (b) and (c) are equivalent under (a).

It remains to show that the local property implies the global property. The proof requires a short calculation, and may be done either by Theorem 7.1 or within the proof of Theorem 7.12. We follow the first route here. Assume that π is positive and satisfies the local Markov property. Then $u \perp v$ for all $u, v \in V$ with $u \approx v$. By Theorem 7.1, there exist functions $f_K, K \in \mathcal{M}$, such that

$$(7.16) \quad \pi(A) = \prod_{K \in \mathcal{M}} f_K(A \cap K), \quad A \subseteq V.$$

Let $W \subseteq V$. By (7.16), for $A \subseteq W$ and $C \subseteq V \setminus W$,

$$\pi(\sigma_W = A \mid \sigma_{V \setminus W} = C) = \frac{\prod_{K \in \mathcal{M}} f_K((A \cup C) \cap K)}{\sum_{B \subseteq W} \prod_{K \in \mathcal{M}} f_K((B \cup C) \cap K)}.$$

Any clique K with $K \cap W = \emptyset$ makes the same contribution $f_K(C \cap K)$ to both numerator and denominator, and may be cancelled. The remaining cliques are subsets of $\widehat{W} = W \cup \Delta W$, so that

$$\pi(\sigma_W = A \mid \sigma_{V \setminus W} = C) = \frac{\prod_{K \in \mathcal{M}, K \subseteq \widehat{W}} f_K((A \cup C) \cap K)}{\sum_{B \subseteq W} \prod_{K \in \mathcal{M}, K \subseteq \widehat{W}} f_K((B \cup C) \cap K)}.$$

The right side does not depend on $\sigma_{V \setminus \widehat{W}}$, whence

$$\pi(\sigma_W = A \mid \sigma_{V \setminus W} = C) = \pi(\sigma_W = A \mid \sigma_{\Delta W} = C \cap \Delta W)$$

as required for the global Markov property. \square

Proof of Theorem 7.12. Assume first that π is a positive Markov field, and let

$$(7.17) \quad \phi_C = \sum_{L \subseteq C} (-1)^{|C \setminus L|} \log \pi(L), \quad C \subseteq V.$$

By the inclusion–exclusion principle,

$$\log \pi(B) = \sum_{C \subseteq B} \phi_C, \quad B \subseteq V,$$

and we need only show that $\phi_C = 0$ for $C \notin \mathcal{K}$. Suppose $u, v \in C$ and $u \approx v$. By (7.17),

$$\phi_C = \sum_{L \subseteq C \setminus \{u, v\}} (-1)^{|C \setminus L|} \log \left(\frac{\pi(L \cup u \cup v)}{\pi(L \cup u)} \bigg/ \frac{\pi(L \cup v)}{\pi(L)} \right),$$

which equals zero by the local Markov property and Proposition 7.13. Therefore, π is a Gibbs field with potential function ϕ .

Conversely, suppose that π is a Gibbs field with potential function ϕ . Evidently, π is positive. Let $A \subseteq V$, and $u \notin A, v \in A$ with $u \approx v$. By (7.10),

$$\begin{aligned} \log \left(\frac{\pi(A \cup u)}{\pi(A)} \right) &= \sum_{\substack{K \subseteq A \cup u, u \in K \\ K \in \mathcal{K}}} \phi_K \\ &= \sum_{\substack{K \subseteq A \cup u \setminus v, u \in K \\ K \in \mathcal{K}}} \phi_K \quad \text{since } u \approx v \text{ and } K \in \mathcal{K} \\ &= \log \left(\frac{\pi(A \cup u \setminus v)}{\pi(A \setminus v)} \right). \end{aligned}$$

The claim follows by Proposition 7.13. \square

We close this section with some notes on the history of the equivalence of Markov and Gibbs random fields. This may be derived from Brook's theorem, Theorem 7.1, but it is perhaps more informative to prove it directly as above via the inclusion–exclusion principle. It is normally attributed to Hammersley and Clifford, and an account was circulated (with a more complicated formulation and proof) in an unpublished note of 1971, [129]

(see also [68]). Versions of Theorem 7.12 may be found in the later work of several authors, and the above proof is taken essentially from [103]. The assumption of positivity is important, and complications arise for non-positive measures, see [191] and Exercise 7.2.

For applications of the Gibbs/Markov equivalence in statistics, see, for example, [159].

7.3 Ising and Potts models

In a famous experiment, a piece of iron is exposed to a magnetic field. The field is increased from zero to a maximum, and then diminished to zero. If the temperature is sufficiently low, the iron retains some residual magnetization, otherwise it does not. There is a critical temperature for this phenomenon, often named the *Curie point* after Pierre Curie, who reported this discovery in his 1895 thesis. The famous (Lenz–)Ising model for such ferromagnetism, [142], may be summarized as follows. Let particles be positioned at the points of some lattice in Euclidean space. Each particle may be in either of two states, representing the physical states of ‘spin-up’ and ‘spin-down’. Spin-values are chosen at random according to a Gibbs state governed by interactions between neighbouring particles, and given in the following way.

Let $G = (V, E)$ be a finite graph representing part of the lattice. Each vertex $x \in V$ is considered as being occupied by a particle that has a random spin. Spins are assumed to come in two basic types (‘up’ and ‘down’), and thus we take the set $\Sigma = \{-1, +1\}^V$ as the sample space. The appropriate probability mass function $\lambda_{\beta, J, h}$ on Σ has three parameters satisfying $\beta, J \in [0, \infty)$ and $h \in \mathbb{R}$, and is given by

$$(7.18) \quad \lambda_{\beta, J, h}(\sigma) = \frac{1}{Z_1} e^{-\beta H(\sigma)}, \quad \sigma \in \Sigma,$$

where the ‘Hamiltonian’ $H : \Sigma \rightarrow \mathbb{R}$ and the ‘partition function’ Z_1 are given by

$$(7.19) \quad H(\sigma) = -J \sum_{e=(x,y) \in E} \sigma_x \sigma_y - h \sum_{x \in V} \sigma_x, \quad Z_1 = \sum_{\sigma \in \Sigma} e^{-\beta H(\sigma)}.$$

The physical interpretation of β is as the reciprocal $1/T$ of temperature, of J as the strength of interaction between neighbours, and of h as the external magnetic field. We shall consider here only the case of zero external-field, and we assume henceforth that $h = 0$. Since J is assumed non-negative, the measure $\lambda_{\beta, J, 0}$ is larger for smaller $H(\sigma)$. Thus, it places greater weight on configurations having many neighbour-pairs with like spins, and for this

reason it is called ‘ferromagnetic’. When $J < 0$, it is called ‘antiferromagnetic’.

Each edge has equal interaction strength J in the above formulation. Since β and J occur only as a product βJ , the measure $\lambda_{\beta, J, 0}$ has effectively only a single parameter βJ . In a more complicated measure not studied here, different edges e are permitted to have different interaction strengths J_e . In the meantime we shall set $J = 1$, and write $\lambda_\beta = \lambda_{\beta, 1, 0}$.

Whereas the Ising model permits only two possible spin-values at each vertex, the so-called (Domb-)Potts model [202] has a general number $q \geq 2$, and is governed by the following probability measure.

Let q be an integer satisfying $q \geq 2$, and take as sample space the set of vectors $\Sigma = \{1, 2, \dots, q\}^V$. Thus each vertex of G may be in any of q states. For an edge $e = \langle x, y \rangle$ and a configuration $\sigma = (\sigma_x : x \in V) \in \Sigma$, we write $\delta_e(\sigma) = \delta_{\sigma_x, \sigma_y}$, where $\delta_{i, j}$ is the Kronecker delta. The relevant probability measure is given by

$$(7.20) \quad \pi_{\beta, q}(\sigma) = \frac{1}{Z_P} e^{-\beta H'(\sigma)}, \quad \sigma \in \Sigma,$$

where $Z_P = Z_P(\beta, q)$ is the appropriate partition function (or normalizing constant) and the Hamiltonian H' is given by

$$(7.21) \quad H'(\sigma) = - \sum_{e=\langle x, y \rangle \in E} \delta_e(\sigma).$$

In the special case $q = 2$,

$$(7.22) \quad \delta_{\sigma_1, \sigma_2} = \frac{1}{2}(1 + \sigma_1 \sigma_2), \quad \sigma_1, \sigma_2 \in \{-1, +1\},$$

It is easy to see in this case that the ensuing Potts model is simply the Ising model with an adjusted value of β , in that $\pi_{\beta, 2}$ is the measure obtained from $\lambda_{\beta/2}$ by re-labelling the local states.

We mention one further generalization of the Ising model, namely the so-called n -vector or $O(n)$ model. Let $n \in \{1, 2, \dots\}$ and let S^{n-1} be the set of vectors of \mathbb{R}^n with unit length, that is, the $(n-1)$ -sphere. A ‘model’ is said to have $O(n)$ symmetry if its Hamiltonian is invariant under the operation on S^{n-1} of $n \times n$ orthonormal matrices. One such model is the n -vector model on $G = (V, E)$, with Hamiltonian

$$H_n(\mathbf{s}) = - \sum_{e=\langle x, y \rangle \in E} \mathbf{s}_x \cdot \mathbf{s}_y, \quad \mathbf{s} = (\mathbf{s}_v : v \in V) \in (S^{n-1})^V,$$

where $\mathbf{s}_x \cdot \mathbf{s}_y$ denotes the scalar product. When $n = 1$, this is simply the Ising model. It is called the X/Y model when $n = 2$, and the Heisenberg model when $n = 3$.

The Ising and Potts models have very rich theories, and are amongst the most intensively studied of models of statistical mechanics. In ‘classical’ work, they are studied via cluster expansions and correlation inequalities. The so-called ‘random-cluster model’, developed by Fortuin and Kasteleyn around 1960, provides a single framework incorporating the percolation, Ising, and Potts models, as well as electrical networks, uniform spanning trees, and forests. It enables a representation of the two-point correlation function of a Potts model as a connection probability of an appropriate model of stochastic geometry, and this in turn allows the use of geometrical techniques already refined in the case of percolation. The random-cluster model is defined and described in Chapter 8, see also [109].

The $q = 2$ Potts model is essentially the Ising model, and special features of the number 2 allow a special analysis for the Ising model not yet replicated for general Potts models. This method is termed the ‘random-current representation’, and it has been especially fruitful in the study of the phase transition of the Ising model on \mathbb{L}^d . See [3, 7, 10] and [109, Chap. 9].

7.4 Exercises

7.1 Let $G = (V, E)$ be a finite graph, and let π be a probability measure on the power set $\Sigma = \{0, 1\}^V$. A configuration $\sigma \in \Sigma$ is identified with the subset of V on which it takes the value 1, that is, with the set $\eta(\sigma) = \{v \in V : \sigma_v = 1\}$. Show that

$$\pi(B) = \exp\left(\sum_{K \subseteq B} \phi_K\right), \quad B \subseteq V,$$

for some function ϕ acting on the set \mathcal{K} of cliques of G , if and only if

$$\pi(\sigma) = \exp\left(\sum_{K \in \mathcal{K}} f_K(\sigma_K)\right), \quad \sigma \in \Sigma,$$

for some functions $f_K : \{0, 1\}^K \rightarrow \mathbb{R}$, with K ranging over \mathcal{K} . Recall the notation $\sigma_K = (\sigma_v : v \in K)$.

7.2 [191] Investigate the Gibbs/Markov equivalence for probability measures that have zeroes. It may be useful to consider the example illustrated in Figure 7.1. The graph $G = (V, E)$ is a 4-cycle, and the local state space is $\{0, 1\}$. Each of the eight configurations of the figure has probability $\frac{1}{8}$, and the other eight configurations have probability 0. Show that this measure μ satisfies the local Markov property, but cannot be written in the form

$$\mu(B) = \prod_{K \subseteq B} f(K), \quad B \subseteq V,$$

for some f satisfying $f(K) = 1$ if $K \notin \mathcal{K}$, the set of cliques.

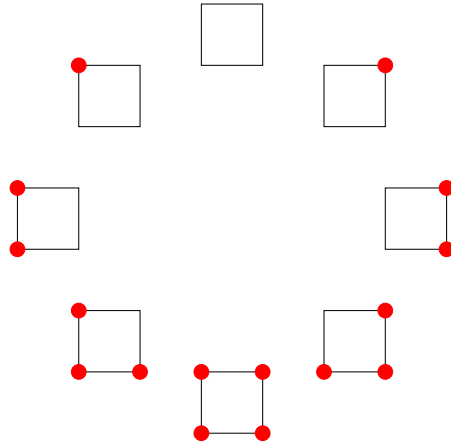


Figure 7.1. Each vertex of the 4-cycle may be in either of the two states 0 and 1. The marked vertices have state 1, and the unmarked vertices have state 0. Each of the above eight configurations has probability $\frac{1}{8}$, and the other eight configurations have probability 0.

7.3 Ising model with external field. Let $G = (V, E)$ be a finite graph, and let λ be the probability measure on $\Sigma = \{-1, +1\}^V$ satisfying

$$\lambda(\sigma) \propto \exp\left(h \sum_{v \in V} \sigma_v + \beta \sum_{e=(u,v)} \sigma_u \sigma_v\right), \quad \sigma \in \Sigma,$$

where $\beta > 0$. Thinking of Σ as a partially ordered set (where $\sigma \leq \sigma'$ if and only if $\sigma_v \leq \sigma'_v$ for all $v \in V$), show that:

- λ satisfies the FKG lattice condition, and hence is positively associated,
- for $v \in V$, $\lambda(\cdot \mid \sigma_v = -1) \leq_{\text{st}} \lambda \leq_{\text{st}} \lambda(\cdot \mid \sigma_v = +1)$.