arXiv:2009.00038v2 [stat.ML] 20 Oct 2020

UNCERTAINTY QUANTIFICATION FOR MARKOV RANDOM FIELDS

PANAGIOTA BIRMPA* AND MARKOS A. KATSOULAKIS†

Abstract. We present an information-based uncertainty quantification method for general Markov Random Fields. Markov Random Fields (MRFs) are structured, probabilistic graphical models over undirected graphs, and provide a fundamental unifying modeling tool for statistical mechanics, probabilistic machine learning, and artificial intelligence. Typically MRFs are complex and high-dimensional with nodes and edges (connections) built in a modular fashion from simpler, low-dimensional probabilistic models and their local connections; in turn, this modularity allows to incorporate available data to MRFs and efficiently simulate them by leveraging their graph-theoretic structure. Learning graphical models from data and/or constructing them from physical modeling and constraints necessarily involves uncertainties inherited from data, modeling choices, or numerical approximations. These uncertainties in the MRF can be manifested either in the graph structure or the probability distribution functions, and necessarily will propagate in predictions for quantities of interest. Here we quantify such uncertainties using tight, information-based bounds on the predictions of quantities of interest; these bounds take advantage of the graphical structure of MRFs and are capable of handling the inherent high-dimensionality of such graphical models. We demonstrate our methods in MRFs for medical diagnostics and statistical mechanics models. In the latter, we develop uncertainty quantification bounds for finite-size effects and phase diagrams, which constitute two of the typical predictions goals of statistical mechanics modeling.

Key words. Markov Random Fields, Uncertainty Quantification, Information Theory, Probabilistic Inequalities, Ising model, Long range interactions

AMS subject classifications. 62H22, 82B20, 94A17

1. Introduction.

Probabilistic graphical models (PGM) constitute one of the fundamental tools for Probabilistic Machine Learning (ML) and Artificial Intelligence (AI), allowing for systematic and scalable modeling of uncertainty, causality, domain knowledge, and data assimilation, [36, 47, 34]. The main idea behind PGMs is to represent complex models and associated learning processes using random variables and their interdependence through a graph. We achieve it by constructing structured, high-dimensional probabilistic models, involving many parameters, nodes, and edges, from simpler ones with few parameters, nodes, and edges, thus allowing for distributed probability computations, and by incorporating available data, exploiting graph-theoretic model representations. PGMs are generally classified into Markov Random Fields (MRF) defined over undirected graphs, and Bayesian Networks, defined over Directed Acyclical Graphs [47] that represent causal relationships between random variables, as well as mixtures of those two classes, [34]. Furthermore, the modeling flexibility of PGMs also allows to combine dynamics, data, and deep learning in Hidden Markov Models [36, 48, 45], as well as in recent work brings together multi-scale modeling, physical constraints, and neural networks, [67, 39].

In this paper, we focus on MRFs which are defined over undirected graphs. For instance, they arise in statistical mechanics where interactions between particles are usually bi-directional, or when there may be no inherent evidence for causality (directionality) and thus undirected graphs are the appropriate structure for probabilistic models, [36, 47, 69]. Some of the applications of MRFs include image segmentation,

*Department of Mathematics and Statistics, University of Massachusetts, Amherst, U.S.A (birmpa@math.umass.edu).

†Department of Mathematics and Statistics, University of Massachusetts, Amherst, U.S.A (markos@math.umass.edu.

image denoising [47, Sec. 4.2], text processing [62, 54], bioinformatics [59], computer vision [43], Markov logic networks, [19], Gaussian Markov networks [47, Sec. 7.3], artificial intelligence [34], and statistical mechanics [53, Sec. 19.4]. Overall, MRFs provide a fundamental unifying modeling tool for statistical mechanics, probabilistic machine learning, and artificial intelligence.

Learning MRFs can be based on available data, e.g. for learning the graph we refer to [47, 30, 41] for score-based methods, [53, 42] for independence tests on the graph, while maximum likelihood or Bayesian methods can be used for parameter identification, [47]. On the other hand, MRFs in statistical mechanics can be constructed from physical modeling and related constraints, [66, 53]. Therefore, the learning stage of MRFs necessarily involves uncertainties inherited from data, modeling choices, compromises on model complexity, or numerical approximations. These uncertainties in the MRF can be manifested either in the graph structure or the probability distribution functions, and necessarily will propagate through the graph structure and the corresponding structured probabilistic model in the predictions for quantities of interest (QoIs). To understand and quantify the impact of such uncertainties on model predictions, in this paper we present an information-based uncertainty quantification (UQ) method for general MRFs.

Model Uncertainty in Probabilistic Models: In general probabilistic models, uncertainties arising just from the fluctuations of the QoIs, associated with a given probabilistic model p, are referred to as aleatoric and occur when sampling p, [13]. They are handled by well-known tools, e.g. central limit theorems, concentration inequalities, Bayesian posteriors, MCMC, generalized Polynomial Chaos, etc. In contrast to this more standard type of uncertainty quantification, in MRFs, due to the learning process described earlier, we have *model* uncertainties (also known as epistemic), both in the structure (graph) and the probabilistic model itself–including parametric ones.

Next, we briefly describe the information-theoretic formulation of model uncertainty for general probabilistic models, without assuming any graphical model structures, see [37] for more details. To practically address model uncertainty, we typically compromise by constructing a surrogate or approximation or baseline model p. We construct families $\mathcal{Q}$ of (non-parametric) *alternative models* $\tilde{p}$ to compare to p, while the "true" model p^*, which may be intractable or partly unknown, should belong to $\mathcal{Q}$; for this reason we can refer to $\mathcal{Q}$ as the ambiguity set, typically defined as a neighborhood of alternative models around the baseline p:

$$\mathcal{Q} = \mathcal{Q}^\eta = \left\{\tilde{p} : d(\tilde{p}, p) \leq \eta\right\}, \tag{1.1}$$

where $\eta > 0$ corresponds to the size of the ambiguity set and $d = d(\tilde{p}, p)$ denotes a probability metric or divergence. The next natural mathematical goal is to assess the baseline model "compromise" and understand the resulting biases for QoIs f when we use p for predictions instead of the real model $p^* \in \mathcal{Q}$. We define the predictive uncertainty (or bias) for the QoI f when we use the baseline model p instead of any alternative model $\tilde{p} \in \mathcal{Q}$ (including the real one p^*) as the two worst case scenarios:

$$I^{\pm}(f, p; \mathcal{Q}^\eta) = \sup_{\tilde{p}\in\mathcal{Q}^\eta}/\inf \ \{E_{\tilde{p}} f - E_p f\} : \tag{1.2}$$

where $E_{\tilde{p}} f$ denotes the expected value of the QoI f. Therefore, (1.2) provides a robust performance guarantee for the predictions of the baseline model p for f within the ambiguity set $\mathcal{Q}^\eta$. This robust perspective for general probabilistic models p is known in Operations Research as *Distributionally Robust Optimization* (DRO), e.g. [33, 35]. While the definition (1.2) is rather natural and intuitive, it is not obvious that it is practically computable since the neighborhood $\mathcal{Q}^\eta$ is infinite-dimensional. However it becomes tractable if we use for metric d in (1.1) the Kullback-Leibler (KL) divergence $R(\tilde{p}||p)$. Accordingly, η is a measure of the confidence in KL we put in the baseline model p. In recent work [13, 21, 37], it has been demonstrated that $I^{\pm}(f,p;\eta)$ (an infinite dimensional optimization problem) is directly computable using the variational formula (follows directly from the Donsker Varadhan variational principle, [21]):

$$I^{\pm}(f,p;\mathcal{Q}^\eta) = \pm \inf_{c>0} \left[\frac{1}{c} \log \int e^{\pm c(f-E_p f)} p(dx) + \frac{\eta}{c} \right]. \tag{1.3}$$

In this formula we recognize two main ingredients: η is model uncertainty from (1.1) while the Moment Generating Function (MGF) $\int e^{\pm cf} P(dx)$ encodes the QoI f at the baseline model p. In [21, 37] the authors have developed techniques to compute (exactly or approximately via asymptotics [21]) as well as to provide explicitly upper and lower bounds on $I^{\pm}(f,p;\mathcal{Q}^\eta)$ in terms of concentration inequalities [37]. Tightness, i.e when the sup and inf in (1.2) are attained by an appropriate measure $\tilde{p}$ have also been studied in [37]. Finally, related UQ bounds have been derived for Markov processes using variational principles and functional inequalities [5], and in rare events [2, 22].

Main results: The main thrust of our results here is to build on the aforementioned perspective for information-based UQ, in order to develop UQ methods for MRFs, and to address their specific UQ challenges. In particular, here we address both structure (graph) and probabilistic uncertainties–including parametric ones–using tight, information-based bounds on the predictions of QoIs; although these new UQ bounds rely on (1.2), they specifically, (a) take advantage of the graphical structure of MRFs, and (b) are capable of handling the inherent high-dimensionality of such graphical models, i.e. there is a necessity for scalable UQ in the size of the system, namely the number of nodes in MRFs.

Regarding the scalability issue, in [44] the authors tested various model uncertainty metrics in defining $d(\tilde{p},p)$ in (1.1) such as the Hellinger distance and χ^2 divergence and inequalities, such as Csiszar-Kullback–Pinsker and the Hammersley-Chapman-Robbins inequalities, [65], in order to bound the model bias with respect of a QoI in the spirit of (1.3). It was shown that among these bounds the only one that scales with the dimension of the model p is (1.3) and $d(\tilde{p},p)$ should be the KL divergence.

Once we have settled to the use of the KL divergence for the aforementioned scalability reasons, we turn our attention to the baseline MRF p, the ambiguity set (1.1) and the corresponding alternative MRFs $\tilde{p}$. Based on the earlier discussion on model uncertainty for MRFs arising from statistical learning of graph models or physical modeling, we introduce a unifying perspective of three general types of alternative models $\tilde{p}$, based on their relative structure to the baseline p: Type I MRFs where the graph structures (nodes and edges) are identical to the baseline p and the parameters of probability distributions are different, Type II where the nodes are the same, but the edges and parameters are different. Finally, Type III where the nodes, structure, and parameters are all different.

In general, MRFs satisfy the specific conditional independence properties discussed in subsection 2.2. Contrary to Bayesian Networks, their distributions cannot always be factorized by a product of local conditional distributions or local functions over the graph. The celebrated *Hammersley-Clifford Theorem*, also known as *Fundamental Theorem of Random Fields* [47, 40, 53], guarantees such a factorization along maximal cliques of the graph under the assumption that $p > 0$. Here, we make such an assumption for both baseline and Type I-III MRFs. Consequently, the KL divergence is finite without requiring absolute continuity with respect to p.

We take advantage of all the above and we study UQ problems by developing a unified strategy for Type I and II MRFs. We focus on the two primary ingredients of (1.2), namely the KL divergence and the MGF, and how they manifest themselves on MRFs. The latter also explains the reasons Type III models are not covered here. In KL divergence, the factorization discussed earlier is a crucial tool for its simplification and numerical calculation. It allows us to compare local discrepancies in parameters and structure between the baseline p and alternative models. We call these discrepancies *excess factors of Type I-II* given p. We develop a unifying method for computing the excess factors by interrelating the maximal cliques of alternative MRFs and the baseline MRF p. In particular, when only parameters have changed (Type I), there is one to one correspondence between maximal cliques. Changes on the set of edges (Type II) lead to different sets of maximal cliques and one needs to examine the nature of the new edges and their impact on the maximal cliques of p. On top of that, different sets of nodes (Type III) leads to a drastically new structure that makes such interrelation of maximal cliques hard to achieve. Therefore, this case is not examined here, although the Wasserstein metric or the Γ-Divergence, [23] in place of the KL divergence in defining (1.1), could potentially stimulate new ways to compute the "distance" between Type III MRFs and the baseline. As for the MGF, the choice of QoIs is determinant. We focus on two different QoIs; those that are directly or indirectly involved in the models (e.g. sufficient statistics) as well as characteristic functions defined on events of interest.

Regarding tightness discussed earlier, we find specific distributions that the derived UQ bounds for MRFs are attainable. In addition, we go beyond that, and pose the question: *Given a QoI and a baseline MRF p, what are the possible associated undirected graphs such that the conditional independence properties implied by the graphs are satisfied by the distributions?* Such a question introduces the concept of *tightness at a graph level*. As we see, there are cases where we can explicitly determine the associated graphs and others (when the structure is different than the baseline) that depend on the QoI. In the latter case, we give an example that points out a unifying method to construct the right graph or at least, a set of possible graphs.

Demonstration of UQ for MRFs: We first demonstrate all the above concepts and UQ methods in a fairly simple and low dimensional MRF example from medical diagnostics. Subsequently, we implement our approach on several high-dimensional statistical mechanics models. In particular, we develop UQ bounds for finite size effects and phase diagrams, which constitute two of the typical predictions goals of statistical mechanics modeling and both require scalable UQ methods.

Specifically, we consider as a baseline model p an Ising-spin system with Kac-type interactions, see [55]. We consider such a model because it combines sufficient complexity–since it is not a mean field model–but it is still analytically fairly tractable to serve as a good benchmark problem for high-dimensional MRF. Examples of alternative models $\tilde{p}$ considered here are Ising models with perturbed interaction potentials with respect to the baseline. More specific examples include 1) models with truncated interactions to facilitate computational implementations, [66], and 2) perturbations by a long-range interaction (even longer than a Kac interaction). As we discuss in Section 6, these systems are typically defined in bounded domains with suitable boundary conditions. This can mean that the configuration outside of the domain is given, hence we need to formulate a conditional Gibbs measure in graph language. To have a graph description of these systems, MRFs need to be modified to account for boundary conditions by using reduced Markov Random Fields (rMRFs) (see [47]). Typical questions we address in these examples include the following: (i) How to capture the phase diagram of a perturbed model through its comparison with the baseline phase diagram by bounding the model bias. (ii) How to truncate an interaction so as the phase diagram of the baseline model and the truncated one are close within a prescribed tolerance.

Related methods: We note that existing general-purpose UQ & sensitivity analysis methods, e.g., gradient and ANOVA-based methods,[61, 58, 27] cannot handle UQ with model uncertainties, due to their inherently parametric nature, while it is not clear how they can take advantage of the graphical, causal structure in MRFs. Furthermore, there is earlier work on model uncertainty that represents missing physics with a stochastic noise but without the detailed structure of a graphical model, [49, 63]. In our work, there is a natural structure embedded in the model uncertainty, arising through the graph structure of the MRFs.

Sensitivity analysis has also a long history in statistical mechanics, known as linear and nonlinear response theory, [57, 3], addressing the impact of small and larger parametric perturbations respectively. These types of methods are covered by our approach, as models with perturbed weights are clearly of Type I.

Furthermore, in contrast to these results, a key point in our work here, also immediately clear from (1.3), is that the model perturbations we can consider are not necessarily small. For instance, the parameter η in (1.3) does not need to be small, allowing for global and non-parametric sensitivity analysis; the latter since the KL divergence allows us to consider models outside a specific parametric family, e.g. comparing statistical mechanics models with different potentials. Similarly, we explicitly compute the UQ bounds for large perturbations in a medical diagnostics example.

Sensitivity analysis in MRFs has been also studied in [12]. The authors tackle fundamental questions such as bounding belief change between Markov networks with the same structure but different parameter values. They propose a distance measure and bound the relative change in probability queries by the relative change in parameters (Type I). Global sensitivity in parameters has been studied in [15]. In particular, the authors developed an algorithm that checks the robustness of a MAP configuration i.e. the most likely configuration, in discrete probabilistic graphical models under global perturbations. The present work goes beyond local or global parametric sensitivity analysis that allows us to consider perturbations in both parameters and edges of the graph of the MRF and examines their impact on the prediction of specific QoIs. Special cases of our results for mean field and nearest neighbor Ising models were considered earlier in [44]. Finally, we note that parametric sensitivity analysis for the other class of (directed) probabilistic graphical models, namely Bayesian Networks, was developed in [14] using similar tools to [12]. Parametric sensitivity analysis based on mutual information for multi-scale partial differential equations and neural networks informed by Bayesian Network priors was developed in [68] and [64]. Model uncertainty quantification based on information theory inequalities in the spirit of (1.3) were recently introduced for Bayesian Networks arising in chemical sciences,

[28].

This article is organized as follows: We start with some concepts from graph theory to fix notation and then we give a brief background of MRFs (Section 2). The background behind rMRFs is supplementary to MRFs and is provided in the Appendix A. The idea of graph interconnections, the impact on distributions, alternative models, and addressed questions are illustrated through a motivating example in Section 3.1. We formally introduce all these concepts and the main results in Section 4. The main result provides UQ bounds for rMRFs, preparing the ground for applications to statistical mechanics models. In Section 5, we revisit the motivating example and we answer the addressed questions. Section 6 is devoted to UQ for finite size effects, scalability, and finally UQ for phase diagrams for generic interactions and the Ising-Kac model. In the remaining sections of the Appendix, we discuss the Ising-Kac model, we provide the technical background required for the UQ analysis of Section 6 (e.g Lebowitz-Penrose (LP) limit and thermodynamic pressure), we include the proofs of the main results, and explicit calculations of the UQ for medical diagnostics example and statistical mechanics.

2. Preliminaries.

2.1. Definitions from Graph Theory.

We start with some notation and terminology from graph theory. A **Graph** is a data structure $\mathcal{G}$ consisting of a set of **nodes**, $\mathcal{V} = \{1, 2, \dots, N\}$ and a set of **edges** $\mathcal{E}$, i.e. all pairs of nodes $i, j \in \mathcal{V}$ which are connected by an edge, denoted by (i, j). An edge can be directed, denoted by $i \to j$ or undirected, denoted by $i - j$. A graph is **directed** [resp. **undirected**] if all the edges are directed [resp. undirected]. The nodes $i, j \in \mathcal{V}$ are **adjacent** if and only if $(i, j) \in \mathcal{E}$. The **neighborhood** of node i, denoted by $\mathcal{N}_i$ is the set of nodes which i is adjacent. For sets of nodes A, B, C, C **separates** A **from** B, denoted by $\{i \in A\} \perp_{\mathcal{G}} \{j \in B\} \mid \{k : k \in C\}$, if and only if when we remove all the nodes in C there is no path connecting any node in A to any node in B. Lastly, if $\mathcal{M} \subset \mathcal{V}$, the **induced subgraph** of $\mathcal{G}$ is defined as $\mathcal{G}[\mathcal{M}] = (\mathcal{M}, \mathcal{E}')$ where $\mathcal{E}'$ includes all the edges $(i, j) \in \mathcal{E}$ such that $i, j \in \mathcal{M}$.

2.2. Conditional Independence Properties and MRFs.

In this subsection, we define three conditional independence properties that are necessary for MRFs.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathbf{Y} = \{Y_i\}_{i=1}^{|\mathcal{V}|}$ be a set of random variables that each one is attached to a node and $|\mathcal{V}|$ denotes the cardinality of $\mathcal{V}$.

- **Pairwise Markov property (P)**: Any two non adjacent variables are conditionally independent (**CI**) given the rest, i.e. a conditional joint can be written as a product of conditional marginals; CI is denoted by $Y_i \perp Y_j \mid \{Y_k : k \neq i, j\}$,
- **Local Markov property (L)**: Any variable Y_i is conditionally independent of all the others given its neighbors, that is $Y_i \perp \{Y_k : k \notin \mathcal{N}_i\} \mid \{Y_k : k \in \mathcal{N}_i\}$,
- **Global Markov property (G)**: If A, B, C are sets of nodes then any two sets of variables, $\mathbf{Y}_A = \{Y_i : i \in A\}$ and $\mathbf{Y}_B = \{Y_i : i \in B\}$ are conditionally independent given a separating set of variables $\mathbf{Y}_C = \{Y_i : i \in C\}$, that is $\mathbf{Y}_A \perp \mathbf{Y}_B \mid \mathbf{Y}_C$.

It is obvious that **(G)** implies **(L)** which implies **(P)**.

DEFINITION 2.1. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph where $\mathcal{V} = \{1, 2, \dots, N\}$ is the set of nodes and $\mathcal{E}$ is the set of edges. Let also consider a set of random variables $\mathbf{Y} = (Y_i)_{i \in \mathcal{V}}$ indexed by $\mathcal{V}$ where each Y_i takes values on a finite set $\mathcal{S}$. Their joint probability distribution is denoted by p. We say that $(\mathbf{Y}, p)$ is a* **Markov Random Field (MRF)** *iff* **(G)**, **(L)** *and* **(P)** *are satisfied.*

As MRFs are defined on an undirected graph, it does not allow to use chain rule and further describe the probability distribution $p(\mathbf{y})$. A factorization rule for MRFs (i.e. for undirected graphs and the conditional independencies) is important and is provided by Hammersley and Clifford in their unpublished work [40, 38]. To state their result, we need a few more definitions. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph and let $c \subset \mathcal{V}$.

- (i) c is called **clique** if any pair of nodes in c is connected by some edge.
- (ii) c is called **maximal clique** if any superset c' of c (i.e $c' \supset c$) is not a clique any more. The set of all maximal cliques of graph $\mathcal{G}$ is denoted by $\mathcal{C}_{\mathcal{G}}$.

Hammersley-Clifford Theorem *A positive distribution $p(\mathbf{y}) > 0$ satisfies one of* **(P), (L)** *and* **(G)** *of an undirected graph G iff p parametrized by some parameters $\mathbf{w} = \{\mathbf{w}_c\}_{c \in \mathcal{C}_{\mathcal{G}}}$ can be represented as a product of clique potentials, i.e*

$$P_{\Psi}^{\mathbf{w}}(\mathbf{y}) \equiv p(\mathbf{y} \mid \mathbf{w}) = \frac{1}{Z(\mathbf{w})} \prod_{c \in \mathcal{C}_{\mathcal{G}}} \Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) \tag{2.1}$$

where $\Psi_c(\mathbf{y}_c \mid \mathbf{w}_c)$ is a non-negative function defined on the random variables in clique c and parametrized by some parameters $\mathbf{w}_c$, and is called **clique potential**. *Also $Z(\mathbf{w})$ is the partition function given by*

$$Z(\mathbf{w}) = \sum_{\mathbf{y}} \prod_{c \in \mathcal{C}_{\mathcal{G}}} \Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) \tag{2.2}$$

Intuitively, the theorem states that given an undirected graph $\mathcal{G}$, the set of all joint distributions that can be factorized in the sense of (2.1) is identical to the set of joint distributions that satisfy the conditional independence properties, under the restriction of strictly positive distributions.

Remark 2.2. Without the assumption of strict positiveness of the joint distribution p, the theorem is not valid. A counterexample has been obtained in [52].

Remark 2.3. The KL divergence or any other f-divergences between a baseline MRF that is assumed nonnegative and alternative MRFs of Type II-III (different structure, see Introduction) could be infinite due to the loss of absolute continuity. In that case, the Wasserstein metric or the Γ-Divergence, [23], could potentially be good alternatives for the KL divergence in defining (1.1). The implementation of the Wasserstein metric or the Γ-Divergence is still unexplored in the context of such MRFs. For this purpose, the development of new methods constitutes an important step towards comparing MRFs with different structures and nonnegative distributions. In this article, we restrict our attention to the Hammersley-Clifford Theorem and we assume strictly positive probability distributions.

3. A motivating Example. In this section, we introduce a motivating example from medical diagnostics. We exploit its simplicity and low dimensionality to demonstrate MRF modeling with parameters and structure learned from data, where we do not assume independence. Most importantly in our context, in this simple model we also demonstrate the types of uncertainties that arise naturally in MRF modeling.

3.1. Medical Diagnostics. Consider the problem of investigating interdependence (structure) and its strength (parameters) between Smoking (S), Asthma (A), Lung cancer (L), and Cough (C), [18]. It is assumed there are prior expert knowledge and data encoded by a probabilistic model (distribution) p^* defined on $\{S, C, L, A\}$. Due to limitations in expert knowledge and data, the true distribution p^* itself may be altogether unknown. This, in turn, forces us to build a surrogate baseline model p, which therefore is uncertain in ways we will specify next, (see uncertainties of Type I-III below).

A baseline MRF. Let us consider a large collection of patient records sampled from p^* denoted by $\mathcal{D} = \{\mathbf{d}[1], \ldots, \mathbf{d}[N]\}$. Using a structure-learning algorithm on the data $\mathcal{D}$ (for instance, greedy score-based structure search algorithm for log-linear models [47, 36]), a model with the structure of $\mathcal{G}$ illustrated in Figure 1 is built, [18]. We assume that the graph is undirected as the directionality associated with the variable dependencies is not known (or is not expected). Subsequently, by parameter learning (for instance, using maximum likelihood estimation [47]) the weights $\mathbf{w}$ become specified from the available data. From now on the resulting model $(\mathcal{G}, \mathbf{w}, p)$ is called the baseline model. Let us consider $S \in \{s_0, s_1\}$, $L \in \{l_0, l_1\}$, $A \in \{a_0, a_1\}$ and

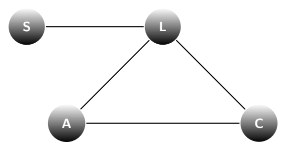

FIG. 1. *MRF structure* $(\mathbf{Y}, p) = (\{S, C, L, A\}, p)$ *over* $\mathcal{G}$ *with joint probability distribution* p.

$C \in \{c_0, c_1\}$. For example, the values s_0 and s_1 can be thought as smoking and non-smoking respectively, a_0 and a_1 can be asthma and no asthma, and so forth. The random variables

$$\mathbf{Y} = \{Y_1, Y_2, Y_3, Y_4\} = \{S, L, A, C\}$$

are accordingly attached to the nodes in $\mathcal{V} = \{1, \ldots, 4\}$ with edges in $\mathcal{E} = \{1-2, 2-3, 2-4, 3-4\}$. The class of maximal cliques is $\mathcal{C}_{\mathcal{G}} = \Big\{\{1,2\}, \{2,3,4\}\Big\}$. As in [18], the joint probability distribution could be a log-linear model ([47], Section 4.4) and thanks to Hammersley-Clifford Theorem, its can be factorized over the maximal cliques with clique potentials

$$\Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) = e^{w_c f_c(\mathbf{y}_c)}, \quad \mathbf{w} = \{\mathbf{w}_c\}_{c \in \mathcal{C}_{\mathcal{G}}}. \tag{3.1}$$

where f_c is often called *feature*.

Alternative models. Both learning steps can induce uncertainties in structure and/or parameters on the baseline. Next, we model and quantify such uncertainties by considering alternative models to the baseline: we focus on graphical models that may have been obtained by learning structure and parameters from either a different data set $\tilde{\mathcal{D}} = \{\tilde{\mathbf{d}}[1], \ldots, \tilde{\mathbf{d}}[\tilde{N}]\}$ or the same data set $\mathcal{D}$ but with different prior (expert) knowledge. We denote the corresponding alternative models $(\tilde{\mathcal{G}}, \tilde{\mathbf{w}}, \tilde{p})$ and assume they can be also represented by a MRF with $\tilde{p} > 0$. We compare $\mathcal{G}$ and $\tilde{\mathcal{G}}$ as well as $\mathbf{w}$ and $\tilde{\mathbf{w}}$. These discrepancies interrelate their particular clique potentials of the Hammersley-Clifford Theorem and are expected to affect predictions of QoIs. Such comparisons and the analysis we develop, reveal how sensitive the baseline model is to changes or uncertainties on parameters (Type I) and structure (Type II & III).

Type I: Let us assume that by learning weights from $\tilde{\mathcal{D}}$ e.g. using maximum likelihood estimation [47], a second model can be obtained with the same structure and new weights denoted by $\tilde{\mathbf{w}}$. Hence, $\tilde{\mathcal{G}} = \mathcal{G}$ (also $\mathcal{C}_{\mathcal{G}} = \mathcal{C}_{\tilde{\mathcal{G}}}$) and $\mathbf{w} \neq \tilde{\mathbf{w}}$. There can be a variety of reasonable assumptions regarding the weights of $\tilde{p}$. The simplest case is when the weight of one maximal clique has found to be different, e.g. the weight on the maximal clique $\{1,2\}$ reduced by 20%. Clique potentials change when the weight in their maximal cliques changes. For example, for $\{1,2\}$, the corresponding clique potential is expressed as

$$\begin{aligned}\tilde{\Psi}_{\{1,2\}}(\mathbf{y}_{\{1,2\}} \mid \tilde{\mathbf{w}}_{\{1,2\}}) &= e^{\tilde{w}_{\{1,2\}}\tilde{f}_{\{1,2\}}(\mathbf{y}_{\{1,2\}})} \\ &= \Psi_{\{1,2\}}(\mathbf{y}_{\{1,2\}} \mid \mathbf{w}_{\{1,2\}})\Phi_{\{1,2\}}(\mathbf{y}_{\{1,2\}} \mid \tilde{\mathbf{w}}_{\{1,2\}})\end{aligned} \tag{3.2}$$

with

$$\Phi_{\{1,2\}}(\mathbf{y}_{\{1,2\}} \mid \tilde{\mathbf{w}}_{\{1,2\}}) = e^{-0.8w_{\{1,2\}}f_{\{1,2\}}(\mathbf{y}_{\{1,2\}})} \tag{3.3}$$

since we consider the simplest case where $\tilde{f}_{\{1,2\}}(\mathbf{y}_{\{1,2\}}) = f_{\{1,2\}}(\mathbf{y}_{\{1,2\}})$. $\tilde{p}$ is then called alternative model of Type I.

Type II: Let $\tilde{\mathcal{G}} \neq \mathcal{G}$ (also $\mathcal{C}_{\mathcal{G}} \neq \mathcal{C}_{\tilde{\mathcal{G}}}$) and $\mathbf{w} \neq \tilde{\mathbf{w}}$. Intuitively, a change on the set of edges can be thought of as structure-learning from either a new data set $\tilde{\mathcal{D}}$ and/or different prior knowledge; see for example Figure 2 where only one new edge has been added.

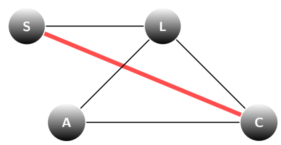

FIG. 2. *A Type II model* $(\tilde{\mathcal{G}}, \tilde{\mathbf{w}}, \tilde{p})$ *over* $\mathbf{Y} = \{S, C, L, A\}$ *with joint probability distribution* $\tilde{p}$. *The associated graph is demonstrated by* $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$ *with* $\tilde{\mathcal{E}} = \mathcal{E} \cup \{1-4\}$. *The new edge is shown in red color.*

The new set of edges and maximal cliques are given by $\tilde{\mathcal{E}} = \mathcal{E} \cup \{1-4\}$ and $\mathcal{C}_{\tilde{\mathcal{G}}} = \Big\{\{1,2,4\},\{2,3,4\}\Big\}$ respectively. Finally, for simplicity we assume that $\tilde{p}$ is another log-linear model with the same clique potential on $\{2,3,4\}$ and different potential on $\{1,2,4\}$ such as:

$$\tilde{\Psi}_{\{1,2,4\}}(\mathbf{y}_{\{1,2,4\}} \mid \tilde{\mathbf{w}}_{\{1,2,4\}}) = e^{\tilde{w}_{\{1,2,4\}}\tilde{f}_{\{1,2,4\}}(\mathbf{y}_{\{1,2,4\}})} \tag{3.4}$$

for some binary function $\tilde{f}_{\{1,2,4\}}$ with its corresponding weights $\tilde{w}_{\{1,2,4\}}$. The key observation is that the new edge $(1-4)$ enlarges the already existing maximal clique $\{1,2\}$ and for that we express $\tilde{\Psi}_{\{1,2,4\}}$ as

$$\tilde{\Psi}_{\{1,2,4\}}(\cdot \mid \tilde{\mathbf{w}}_{\{1,2,4\}}) = \Psi_{\{1,2\}}(\cdot \mid \mathbf{w}_{\{1,2\}})\Phi(\cdot \mid \mathbf{w}_{\{1,2,4\}}, \tilde{\mathbf{w}}_{\{1,2,4\}}) \tag{3.5}$$

with

$$\Phi(\mathbf{y}_{\{1,2,4\}} \mid \mathbf{w}_{\{1,2,4\}}, \tilde{\mathbf{w}}_{\{1,2,4\}}) = e^{\tilde{w}_{\{1,2,4\}}\tilde{f}_{\{1,2,4\}}(\mathbf{y}_{\{1,2,4\}}) - w_{\{1,2\}}f_{\{1,2\}}(\mathbf{y}_{\{1,2\}})} \tag{3.6}$$

As it is evident even with this simple example, Type II uncertainties encoded in (3.6) are not obviously tractable.

Remark 3.1. Our UQ methods rely on the KL divergence as means to measure "distance" between baseline and alternative MRFs. The commonalities in parameters and structure between baseline and alternative models combined with the Hammersley-Clifford Theorem allows the KL divergence to be expressed in a simplified and informative form. In Type III, due to the high degree of discrepancies (parameters, nodes, and edges are all different) such an approach fails to capture the uncertainties in a tractable way. Although, we term such uncertainties as Type III.

Prediction uncertainties. Towards a better understanding of how robust is the distribution p to uncertainties described by alternative models of Type I or II, we investigate how sensitive the baseline model is to alternative models with respect to predictions of (QoIs) e.g. a characteristic function of an event of interest as (5.1). We formulate mathematically these questions for any general QoI f as seeking robust, computable and tight bounds on

$$E_{\tilde{p}}[f] \text{ and the model bias: } \max_{\tilde{p}} |E_{\tilde{p}}[f] - E_p[f]|$$

where the maximum is taken over alternative models $\tilde{p}$ of Type I and II. The performance metric that we deploy to capture model discrepancy for the baseline p is the KL divergence. We revisit the example in Section 5 where we concretely answer the questions above.

4. Main Results. In this section, motivated by the previous example, we present a information-based UQ method on the predictions for QoIs for general MRFs. Namely, given a baseline and an alternative MRF with positive distributions, their graphical representation analyzed over maximal cliques can be classified into three types of interconnections. Thanks to the Hammersley-Clifford Theorem, the factorization of their distributions over maximal cliques allows interrelating the distributions in such a way that KL divergence is computable or well-controlled. As mentioned earlier, we focus on KL divergence as it scales correctly with the dimension of the baseline model [21]. Furthermore, we choose QoIs that are present (e.g sufficient statistics) or indirectly related to the models. Such a class of QoIs covers observables involved in finite size effects and phase diagrams for statistical mechanics models that we examine later. Furthermore, with this choice of QoIs, the MGF generating function with respect to the baseline model can also be computable. To quantify the model uncertainty for MRFs arising from statistical learning of graph models or from physical modeling, our starting point is the Donsker-Varadhan variational principle, [20] which in turn implies the *Gibbs Variational principle* (see [13, 21]):

$$\sup_{\lambda>0}\left\{\frac{-\Lambda_p^f(-\lambda) - R(\tilde{p}\|p)]}{\lambda}\right\} \leq E_{\tilde{p}}[f] \leq \inf_{\lambda>0}\left\{\frac{\Lambda_p^f(\lambda) + R(\tilde{p}\|p)]}{\lambda}\right\} \tag{4.1}$$

In the above bound, $\Lambda_p^f(\lambda)$ is the cumulant generating function computed with respect to the baseline model p and $R(\tilde{p}\|p)$ is the KL divergence between models p and an alternative model $\tilde{p}$:

$$\Lambda_p^f(\lambda) := \log E_p[e^{\lambda f}] \quad \text{and} \quad R(\tilde{p}\|p) := E_{\tilde{p}}\left[\log \frac{d\tilde{p}}{dp}\right] < \infty \tag{4.2}$$

4.1. Mathematical Formulation. Let $(\mathcal{G}, \mathbf{w}, p)$ and $(\tilde{\mathcal{G}}, \tilde{\mathbf{w}}, \tilde{p})$ be two models with $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ being the associated graphs, where $\mathcal{V}$ and $\tilde{\mathcal{V}}$ are the sets of nodes and $\mathcal{E}$ and $\tilde{\mathcal{E}}$ are the sets of edges.

DEFINITION 4.1. *$(\tilde{\mathcal{G}}, \tilde{\mathbf{w}}, \tilde{p})$ and $(\mathcal{G}, \mathbf{w}, p)$ can have one of the following interconnections:*

Type I: $\tilde{\mathcal{V}} = \mathcal{V}$, $\tilde{\mathcal{E}} = \mathcal{E}$ *and* $\tilde{\mathbf{w}} \neq \mathbf{w}$, *or*

Type II: $\tilde{\mathcal{V}} = \mathcal{V}$, $\mathcal{E} \subset \tilde{\mathcal{E}}$ *and* $\tilde{\mathbf{w}} \neq \mathbf{w}$ *or*

Type III: $\tilde{\mathcal{V}} \neq \mathcal{V}$, $\mathcal{E} \neq \tilde{\mathcal{E}}$ *and* $\tilde{\mathbf{w}} \neq \mathbf{w}$.

From now on, we refer to the baseline model when we use the notation $(\mathcal{G}, \mathbf{w}, p)$ and without loss of generality we assume $\mathcal{E} \subset \tilde{\mathcal{E}}$. This assumption simplifies the presentation of our approach but intuitively speaking, the fewer edges an MRF has, the more information it provides, since in a sparser graph, there are more conditional independencies specified.

The results in this section are presented for rMRFs (Appendix A) as we can recall them directly in the UQ analysis of statistical mechanics models in Section 6. The results hold for MRFs and when required, we will be providing more details for their implementation to MRFs.

4.1.1. rMRF models. Let $\mathbf{Y} = \{Y_i\}_{i\in\mathcal{V}}$ be a collection of random variables indexed by a set of nodes $\mathcal{V}$ of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, taking values in some space $\mathcal{Y}^{\mathcal{V}} = \otimes_{i=1}^{\mathcal{V}} \mathcal{Y}_i$. Let $p \equiv p(\cdot|\mathbf{w})$ be a strictly positive joint probability distribution of $\mathbf{Y}$ parametrized by $\mathbf{w}$ such that $(\mathbf{Y}, p(\cdot|\mathbf{w}))$ is a MRF.

Let $\mathbf{u}$ be a context and $\mathcal{M} \subset \mathcal{V}$. If $\mathbf{U} = \{Y_i\}_{i\in\mathcal{M}}$ with $\mathbf{U} = \mathbf{u}$, we construct the corresponding rMRF (see Definition A.1) as follows: let $\mathbf{Z} = \{Y_i\}_{i\in\mathcal{V}\setminus\mathcal{M}}$ and $q(\mathbf{z}|\mathbf{w})$ be the probability distribution factorized according to Proposition A.2 (the analogue of the Hammersley-Clifford Theorem for rMRFs):

$$q(\mathbf{z}) \equiv q(\mathbf{z}|\mathbf{w}) = \frac{1}{Z_{\mathbf{u}}(\mathbf{w})} \prod_{c\in\mathcal{C}_{\mathcal{G}}} \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c)$$

4.2. Alternative models. Under the same collection of random variables, we consider alternative with $\tilde{p}(\cdot|\tilde{\mathbf{w}}) > 0$, parametrized by $\tilde{\mathbf{w}}$ so as $(\mathbf{Y}, \tilde{p}(\cdot|\tilde{\mathbf{w}}))$ is a MRF. Given $\mathbf{u}$ and $\mathcal{M} \subset \mathcal{V}$ as before, we construct the corresponding rMRF $(\mathbf{Z}, \tilde{q}(\cdot|\tilde{\mathbf{w}}))$ with the probability joint distribution, $\tilde{q}(\cdot) \equiv \tilde{q}(\cdot|\tilde{\mathbf{w}})$. The corresponding partition function and clique potentials are denoted by $\tilde{Z}_{\mathbf{u}}(\tilde{\mathbf{w}})$ and $\tilde{\Psi}_{\tilde{c}}[\mathbf{u}](\mathbf{z}_{\tilde{c}} \mid \tilde{\mathbf{w}}_{\tilde{c}})$.

4.2.1. Type I. Let us denote the set of maximal cliques whose weights differ as $\mathcal{B} \subset \mathcal{C}_{\mathcal{G}}$. Then for each $c \in \mathcal{B}$, $\tilde{\mathbf{w}}_c \neq \mathbf{w}_c$. The clique potentials of $\tilde{q}(\cdot|\tilde{\mathbf{w}})$ can be rewritten as

$$\tilde{\Psi}_c[\mathbf{u}](\mathbf{z}_c \mid \tilde{\mathbf{w}}_c) = \begin{cases} \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c)\Phi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c) & \text{, if } c \in \mathcal{B} \\ \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c) & \text{, otherwise.} \end{cases} \tag{4.3}$$

Then $\Phi_c[\mathbf{u}](\cdot \mid \tilde{\mathbf{w}}_c, \mathbf{w}_c) > 0$ is defined on the random variables in clique c and is called *$\tilde{q}$-excess factor of type I relative to q on c.* Cliques where no change on weights has occurred, remain the same.

4.2.2. Type II. In this type, the class of maximal cliques, $\mathcal{C}_{\tilde{\mathcal{G}}}$, is different. The analysis becomes more complicated and clique potentials need to be carefully considered. We look into the nature of one or more new edges by categorizing it as one of the following types: a new edge

(i) can create a totally new maximal clique, see Figure 3, third graph,
(ii) can connect two or more already existing maximal cliques, see Figure 3, second graph,
(iii) can enlarge an already existing maximal clique, see Figure 3, forth graph.

By adding more than one new edges, the new maximal cliques of $\tilde{\mathcal{G}}$ can be obtained by a combination of (i), (ii), and (iii). We introduce the following sets:

$$\mathcal{B}_{\cup} = \{\tilde{c} \in \mathcal{C}_{\tilde{\mathcal{G}}} \setminus \mathcal{C}_{\mathcal{G}} : \tilde{c} = \cup_i c_i, \text{ for } c_i \in \mathcal{C}_{\mathcal{G}}\} \tag{4.4}$$

$$\mathcal{B}_{\subseteq} = \{\tilde{c} \in \mathcal{C}_{\tilde{\mathcal{G}}} \setminus \mathcal{C}_{\mathcal{G}} : \text{there exists } c \in \mathcal{C}_{\mathcal{G}} \text{ s.t. } c \subseteq \tilde{c}\} \tag{4.5}$$

$$\mathcal{B}_{\text{new}} = (\mathcal{C}_{\mathcal{G}} \cup \mathcal{B}_{\cup} \cup \mathcal{B}_{\subseteq})^c \tag{4.6}$$

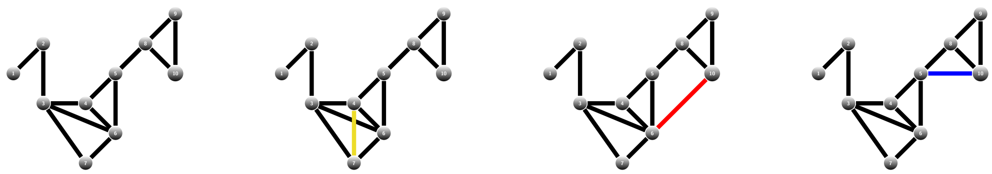

FIG. 3. *(First) Baseline MRF model p demonstrated by graph $\mathcal{G}$. (Second) Alternative model $\tilde{p}$ with the associated graph obtained by adding the yellow edge $(4-7)$ and connecting two maximal cliques of p model, $\{3,4,6\}$ and $\{3,6,7\}$, thus $\tilde{p}$ has a new maximal clique $\{3,4,6,7\}$. (Third) Alternative model $\tilde{p}$ with the associated graph obtained by adding the red edge $(6-10)$ and thus $\tilde{p}$ has a totally new maximal clique $\{6,10\}$. (Forth) Alternative model $\tilde{p}$ with the associated graph obtained by adding the blue edge $(5-10)$ and enlarging the already existing clique, $\{5,8\}$, to $\{5,8,10\}$.*

Then the clique potentials of $\tilde{q}$ can be rewritten as:

$$\tilde{\Psi}_{\tilde{c}}[\mathbf{u}](\mathbf{z}_{\tilde{c}} \mid \tilde{\mathbf{w}}_{\tilde{c}}) = \begin{cases} \prod_{c_i} \Psi_{c_i}[\mathbf{u}](\mathbf{z}_{c_i} \mid \mathbf{w}_{c_i})\Phi^{(ii)}_{\tilde{c}}[\mathbf{u}](\mathbf{z}_{\tilde{c}} \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c) & \text{, if } \tilde{c} \in \mathcal{B}_{\cup}, \\ \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c)\Phi^{(iii)}_{\tilde{c}}[\mathbf{u}]((\mathbf{z}_{\tilde{c}} \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c) & \text{, if } \tilde{c} \in \mathcal{B}_{\subset}, \\ \tilde{\Psi}_{\tilde{c}}[\mathbf{u}](\mathbf{z}_{\tilde{c}} \mid \tilde{\mathbf{w}}_{\tilde{c}}) & \text{, if } \tilde{c} \in \mathcal{B}_{\text{new}} \\ \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c) & \text{, if } \tilde{c} \in \mathcal{C}_{\tilde{\mathcal{G}}} \end{cases} \tag{4.7}$$

where $\Phi^{(ii)}_{\tilde{c}}[\mathbf{u}](\cdot \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c)$ and $\Phi^{(iii)}_{\tilde{c}}[\mathbf{u}](\cdot \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c)$ are some nonnegative functions defined on the variables of $\tilde{c}$. Intuitively, (4.7) tells us that if a new maximal clique $\tilde{c}$ has been created by connecting existing maximal cliques c_i, the clique potential $\tilde{\Psi}_{\tilde{c}}[\mathbf{u}](\mathbf{z}_{\tilde{c}} \mid \tilde{\mathbf{w}}_{\tilde{c}})$ can be written as the product of clique potentials over c_i multiplied by $\Phi^{(ii)}_{\tilde{c}}[\mathbf{u}](\cdot \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c)$. The reasoning is similar when a new maximal clique $\tilde{c}$ has been created by enlarging an existing maximal clique. In both cases, we call $\Phi^{(ii)}_{\tilde{c}}[\mathbf{u}](\cdot \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c)$ and $\Phi^{(iii)}_{\tilde{c}}[\mathbf{u}](\cdot \mid \mathbf{w}_c, \tilde{\mathbf{w}}_c)$ *$\tilde{q}$-excess factors of type II relative to q on $\tilde{c}$*. When $\tilde{c} \in \mathcal{B}_{\text{new}}$, there is no need to express the clique potential through the potentials of $q(\cdot \mid \mathbf{w})$. For simplicity, we assume that clique potentials on common maximal cliques between $\mathcal{G}$ and $\tilde{\mathcal{G}}$ do not change. However, one can consider different potentials and in that case, a term Φ should be introduced similar to (ii) and (iii). For convenience, we establish one last unifying terminology. We call

$$\Phi^{\mathrm{I}}_{\mathbf{u}}(\mathbf{Z}) := \prod_{c \in \mathcal{B}} \Phi_c[\mathbf{u}](\mathbf{Z}_c) \tag{4.8}$$

$$\Phi^{\mathrm{II}}_{\mathbf{u}}(\mathbf{Z}) := \prod_{\tilde{c} \in \mathcal{B}_{\text{new}}} \tilde{\Psi}_{\tilde{c}}[\mathbf{u}](\mathbf{Z}_{\tilde{c}} \mid \tilde{\mathbf{w}}_{\tilde{c}}) \prod_{\tilde{c} \in \mathcal{B}_{\cup}} \Phi^{(ii)}_{\tilde{c}}[\mathbf{u}](\mathbf{Z}_{\tilde{c}}) \prod_{\tilde{c} \in \mathcal{B}_{\subseteq}} \Phi^{(iii)}_{\tilde{c}}[\mathbf{u}]((\mathbf{Z}_{\tilde{c}}) \mid \tilde{\mathbf{w}}_{\tilde{c}}) \tag{4.9}$$

total $\tilde{q}$-excess factor of type I and II relative to q respectively. The total $\tilde{q}$-excess factor of type I relative to q captures all the parameters changes while the total $\tilde{q}$-excess factor of type II relative to q captures all the structural discrepancies. In the case of MRF, we drop the context $\mathbf{u}$ from (4.8) and (4.9) and $\mathbf{Z}$ is replaced by $\mathbf{Y}$.

In Type III, there exists the total $\tilde{q}$-excess factor of type III relative to q. However, due to the high degree of discrepancies, we cannot interrelate maximal cliques of Type III model with q, and by extension each $\tilde{q}$-excess factor cannot be determined.

The next results are straightforward but essential in our calculations. To avoid heavy notation, we remind that $q(\cdot) = q(\cdot \mid \mathbf{w})$ and $\tilde{q}(\cdot) = \tilde{q}(\cdot \mid \tilde{\mathbf{w}})$.

Lemma 4.2. *Let* $(\mathbf{Z}, q)$ *be a rMRF. Then for any alternative rMRF* $(\mathbf{Z}, \tilde{q})$ *of Type* i *with* $\mathrm{i} = \mathrm{I}, \mathrm{II}$ *its partition function is expressed as:*

$$\tilde{Z}_{\mathrm{u}}(\tilde{\mathbf{w}}) = E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}] Z_{\mathrm{u}}(\mathbf{w}) \tag{4.10}$$

where $\Phi^{\mathrm{i}}_{\mathbf{u}}$ *are given by* (4.8) *and* (4.9)*.*

Proof. The proof is based on the method of interrelating the distribution q and $\tilde{q}$, utilizing the total $\tilde{q}$-excess factors relative to q given by (4.8) and (4.9). The explicit computation is provided in Appendix D.1. ∎

The following lemma provides the likelihood ratio between $\tilde{q}$ and q and constitutes the key ingredient for the simplification of (4.2) and the UQ bounds provided in (4.1).

Lemma 4.3. *Let* $(\mathbf{Z}, q)$ *be a rMRF. Then for any alternative rMRF* $(\mathbf{Z}, \tilde{q})$ *of Type* i *with* $\mathrm{i} = \mathrm{I}, \mathrm{II}$*, the corresponding likelihood ratio satisfies:*

$$\frac{d\tilde{q}}{dq} = \frac{\Phi^{\mathrm{i}}_{\mathbf{u}}}{E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]} \tag{4.11}$$

where $\Phi^{\mathrm{i}}_{\mathbf{u}}$ *is given by* (4.8) *and* (4.9)*.*

The proof is omitted as the lemma is a direct consequence of the method of interrelating two distributions discussed above and Lemma 4.2. Note that both results hold for MRFs denoted by $(\mathbf{Y}, p)$ and $(\mathbf{Y}, \tilde{p})$, dropping the context $\mathbf{u}$ from $\Phi^{\mathrm{i}}_{\mathbf{u}}$ in (4.11).

Example 4.4. *In the medical diagnostics example,* $\Phi^{\mathrm{i}}(\mathbf{Y})$ *are given by* (3.3) *and* (3.6)*, representing the total excess factor after reducing the weight on* $\{1, 2\}$ *by 20% and after adding a new link between smoking and cough* $1 - 4$ *respectively. We note that in the latter case,* $\mathcal{B}_{\mathrm{new}} = \mathcal{B}_{\cup} = \emptyset$ *and* $\mathcal{B}_{\subseteq} = \{\{1, 2, 4\}\}$*.*

4.3. Quantities of Interest. We primarily consider QoIs, $f(\mathbf{Z})$, that are directly (e.g. sufficient statistics) or indirectly related to quantities that are involved in the structure of q and $\tilde{q}$. Formally, we consider $f(\mathbf{Z})$ that satisfies

$$\log \Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{Z}) = C_{\mathrm{i}} f(\mathbf{Z}) + \kappa_{\mathrm{i}}(\mathbf{Z}), \quad \mathrm{i} = \mathrm{I}, \mathrm{II}. \tag{4.12}$$

for some constant $C_{\mathrm{i}} \equiv C_{\mathrm{i}}(\mathbf{w}, \tilde{\mathbf{w}}, \mathbf{u}) < 1$ and a function $\kappa_{\mathrm{i}}(\cdot) \equiv \kappa_{\mathrm{i}}(\cdot \mid \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{u})$ that may depend on $\mathbf{w}, \tilde{\mathbf{w}}, \mathbf{u}$. For example, in Section 6, average of spins is a QoI that satisfies (4.12). In particular, Lemma 6.2.2 and the discussion underneath identify $\Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{Z}), C_{\mathrm{i}}$ and $\kappa_{\mathrm{i}}(\mathbf{Z})$ explicitly.

In the medical diagnostics example the QoIs have a different nature e.g. characteristic function of events of interest, see (5.1), hence (4.12) is not applicable. However, the MRF/rMRF formulation, interconnections and alternative models are still the key ingredients to obtain UQ bounds, see Section 5.

4.4. KL divergence. KL divergence is important for the UQ bounds provided in (4.1). By taking into account the above discussion, we show that KL divergence (which is finite due to the positive probabilities q and $\tilde{q}$) depends only on the total $\tilde{q}$-excess factor relative to q given by (4.8) and (4.9). To simplify the notation, we omit the dependence of $\mathbf{Z}$ from κ_i, f and $\Phi^{\mathrm{i}}_{\mathbf{u}}$.

Lemma 4.5. *Let* $(\mathbf{Y}, p^{\mathbf{w}})$*,* $(\mathbf{Y}, \tilde{p}^{\tilde{\mathbf{w}}})$ *be two MRFs defined over graphs* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *and* $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$ *respectively. Let* $\mathbf{u}$ *be a context and* $\mathcal{M} \subset \mathcal{V}$*. We consider the corresponding rMRFs* $(\mathbf{Z}, q), (\mathbf{Z}, \tilde{q})$*.*

a. If $\tilde{q}$ is Type i*, with* $\mathrm{i} = \mathrm{I}$ *or* II*, then the KL divergence is given*

$$
\begin{aligned}
R(\tilde{q}\|q) &= E_{\tilde{q}}\left[\log\frac{\tilde{q}}{q}\right] = E_q\left[\frac{\tilde{q}}{q}\log\frac{\tilde{q}}{q}\right]\\
&= E_{\tilde{q}}[\log\Phi^{\mathrm{i}}_{\mathbf{u}}] - \log E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}] = \frac{1}{E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]}E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\log\Phi^{\mathrm{i}}_{\mathbf{u}}\right] - \log E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}] \qquad (4.13)
\end{aligned}
$$

where $\Phi^{\mathrm{i}}_{\mathbf{u}}$ is defined in (4.8) *and* (4.9) *accordingly.*

b. If $\tilde{q}$ is Type i*, with* $\mathrm{i} = \mathrm{I}$ *or* II*, then for any f satisfying* (4.12)*, the KL divergence is given by*

$$
\mathcal{R}(\tilde{q}\|q) = C_{\mathrm{i}}E_{\tilde{q}}[f] + \frac{E_q\left[\kappa_{\mathrm{i}}\Phi^{\mathrm{i}}_{\mathbf{u}}\right]}{E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\right]} - \log E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\right], \quad \Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{Z}) = e^{C_{\mathrm{i}}f(\mathbf{Z})+\kappa_{\mathrm{i}}(\mathbf{Z})} \qquad (4.14)
$$

Proof. a. We express the KL divergence as follows

$$
R(\tilde{q}\|q) = E_{\tilde{q}}\left[\log\frac{\tilde{q}}{q}\right] = E_q\left[\frac{\tilde{q}}{q}\log\frac{\tilde{q}}{q}\right]
$$

Then, we use Theorem 4.3 and we obtain (4.13). For *b.*, we additionally recall (4.12).

Remark 4.6. As mentioned in Theorem 4.3, the result holds for MRFs denoted by $(\mathbf{Y}, p)$ and $(\mathbf{Y}, \tilde{p})$, dropping the context $\mathbf{u}$ from $\Phi^{\mathrm{i}}_{\mathbf{u}}$.

4.5. UQ bounds. The next theorem is an UQ result on rMRFs that is obtained by consolidating all the above (i.e. total $\tilde{q}$-excess factors relative to q, likelihood ratio, KL divergence and QoIs).

THEOREM 4.7. *Let $(\mathbf{Y}, p)$, $(\mathbf{Y}, \tilde{p})$ be two MRFs defined over graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$ respectively. Let $\mathbf{u}$ be a context and $\mathcal{M} \subset \mathcal{V}$. We consider the corresponding rMRFs $(\mathbf{Z}, q), (\mathbf{Z}, \tilde{q})$. If $\tilde{q}$ is of Type* i*, with* $\mathrm{i} = \mathrm{I}$ *or* II*, then*
(a) for any QoI $f(\mathbf{Z})$, the following bounds hold

$$
\pm E_{\tilde{q}}[f] \le \inf_{\lambda>0}\frac{1}{\lambda}\Big\{\log E_q[e^{\pm\lambda f}] + \frac{1}{E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]}E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\log\Phi^{\mathrm{i}}_{\mathbf{u}}\right] - \log E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]\Big\} \qquad (4.15)
$$

(b) for any QoI $f(\mathbf{Z})$ that satisfies (4.12)*, the following bounds hold:*

$$
\pm E_{\tilde{q}}[f] \le \frac{1}{1-C_{\mathrm{i}}}\inf_{\lambda>0}\frac{1}{\lambda}\Big\{\log E_q[e^{\pm\lambda f}] - \log E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\right] + \frac{E_q\left[\kappa_i\Phi^{\mathrm{i}}_{\mathbf{u}}\right]}{E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\right]}\Big\} \qquad (4.16)
$$

where $\Phi^{\mathrm{i}}_{\mathbf{u}}$ is the total $\tilde{q}$-excess factor relative to q given by (4.8) *and* (4.9)*, κ_{i} and C_{i} are defined in* (4.12)*. Note that when $\tilde{q}$ is of Type* I*, $\tilde{Z}_{\mathbf{u}}(\tilde{\mathbf{w}}) = Z_{\mathbf{u}}(\tilde{\mathbf{w}})$.*

The proof given in Appendix D.2 is based on Lemma 4.5 and the characterization of the exponential integrals. An application to a single parameter exponential family is given in Appendix D.2.

4.6. Tightness for MRFs/rMRFs. Let $\mathcal{V}, \mathcal{E}, \mathbf{u}$ and $(\mathbf{Z}, q)$ be defined in subsection 4.1.1. We introduce the class of rMRFs on $\mathbf{Z}$

$$
\mathcal{P} = \{\text{rMRFs } (\tilde{\mathcal{G}}, \tilde{\mathbf{w}}, \tilde{q}) : \tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}}),\ \mathcal{E} \subseteq \tilde{\mathcal{E}}, \tilde{q} > 0\} \qquad (4.17)
$$

The class consists of all graphs $\tilde{\mathcal{G}}$, with distributions $\tilde{q}$ such that $(\mathbf{Z}, \tilde{q})$ is rMRF of either Type I or II. We define as *rMRF ambiguity set*, the set given by

$$
\mathcal{Q}^{\eta}_{\mathcal{P}} := \mathcal{Q}^{\eta} \cap \mathcal{P} \qquad (4.18)
$$

where $\mathcal{Q}^\eta$ is defined in (1.1) with $d(\tilde{q}, q)$ being the KL divergence. We conclude this section by showing that the inequalities (4.15) and (4.16) are tight i.e. they become an equality for a suitable model $\tilde{q} \in \mathcal{Q}^\eta_{\mathcal{P}}$.

THEOREM 4.8. *Let $(\mathbf{Z}, q)$ be a rMRF defined in subsection 4.1.1 and $f(\mathbf{Z})$ be a QoI with finite MGF $E_q[e^{\lambda f(\mathbf{Z})}]$ in a neighborhood of the origin. Then there exist $0 < \eta_\pm \le \infty$ such that for any $\eta \le \eta_\pm$ there exist probability measures $q^\pm = q^\pm(\eta) \in \mathcal{Q}^\eta_{\mathcal{P}}$, where $\mathcal{Q}^\eta_{\mathcal{P}}$ is given in (4.18), such that (4.15) and (4.16) become an equality. Furthermore, $q^\pm = q^{\lambda_\pm}$ with*

$$dq^{\lambda_\pm} = \frac{e^{\lambda_\pm f}}{E_q[e^{\lambda_\pm f}]} dq \tag{4.19}$$

and $\lambda_\pm$ being the unique solutions of $R(q^{\lambda_\pm} \| q) = \eta$. In particular, the total $q^\pm$-excess factor relative to q denoted by $\Phi^\pm_{\mathbf{u}}$, satisfies

$$\Phi^\pm_{\mathbf{u}} = e^{\lambda_\pm f} \quad \textit{and} \quad C_{\mathrm{i}} = \lambda_\pm, \ \kappa_{\mathrm{i}} = 0 \textit{ respectively.}$$

Proof. See Appendix D.3.

The result holds also for MRFs. The corresponding quantities involved in the theorem are denoted by $p, p^{\lambda_\pm}$ and $\Phi^\pm$.

Remark 4.9. Important points on the theorem are discussed next. For convenience, we use its MRF version. Given a baseline MRF $(\mathbf{Y}, p)$, its associated graph $\mathcal{G}$ and a QoI f, Theorem 4.8 guarantees the existence of probability distributions $p^{\lambda_\pm}$ such that (4.15) and (4.16) become an equality (this is not an unlikely extreme case) and also specifies the distributions explicitly. However, it does not imply how different the associated graphs of $p^\pm$ are, compared to the graph associated to p or grossly speaking, if they are Type I or II. Depending on f, there are cases where this can be determined. In fact, by recalling the Hammersley-Clifford Theorem, we express

$$dp^{\lambda_\pm} = \frac{e^{\lambda_\pm f}}{E_p[e^{\lambda_\pm f}]} \frac{1}{Z(\mathbf{w})} \prod_{c \in \mathcal{C}_\mathcal{G}} \Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) = \frac{1}{Z^\pm(\lambda^\pm, \mathbf{w})} \prod_{c \in \mathcal{C}_\mathcal{G}} e^{\lambda_\pm f} \Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) \tag{4.20}$$

where $Z^\pm(\lambda^\pm, \mathbf{w}) = E_p[e^{\lambda_\pm f}] Z(\mathbf{w})$ is the partition function of $p^{\lambda_\pm}$.

We turn our attention to the product in (4.20). Each factor is defined on a maximal clique of $\mathcal{G}$ apart from $e^{\lambda_\pm f}$. We focus on f; Suppose that f is a QoI with domain $\mathrm{Dom}(f)$ and cannot be written as a sum of more than two functions e.g. sample average. If there is a maximal clique c_0 such that $\mathrm{Dom}(f) \subseteq c_0$, then it turns out that all clique potentials of $p^{\lambda_\pm}$ and p are equal except $\tilde{\Psi}_{c_0} = e^{\lambda_\pm f} \Psi_{c_0}$, and hence

$$dp^{\lambda_\pm} = \frac{1}{E_p[e^{\lambda_\pm f}] Z(\mathbf{w})} \underbrace{e^{\lambda_\pm f} \Psi_{c_0}}_{\tilde{\Psi}_{c_0}} \prod_{c \neq c_0} \Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) \tag{4.21}$$

The associated graphs of $p^{\lambda_\pm}$ are apparently of Type I as no change on maximal cliques occurs. If $\mathrm{Dom}(f) \cap c \neq \emptyset$ for more than two maximal cliques c, then the associated graphs to $p^{\lambda_\pm}$ have been changed and thus are Type II. An example is discussed in subsection 5.3. On the other hand, if f can be expressed as a sum of some functions $f = \sum_i f_i$, then we may have more than one candidate graphs associated to $p^{\lambda_\pm}$ either Type I or II. In fact, the exponential can be factorized further (e.g. $e^{\lambda_\pm f} = \prod_i e^{\lambda_\pm f_i}$), giving rise to more than one ways of matching the clique potentials in the sense of (4.21).

5. UQ for Medical Diagnostics. In this section, we revisit the motivating example presented in Section 3, we discuss the model uncertainties and demonstrate the UQ bounds.

5.1. Baseline model. Let us consider the undirected graph in Figure 1, [18] denoted by $\mathcal{G}$. The class of maximal cliques is $\mathcal{C}_{\mathcal{G}} = \Big\{\{1,2\},\{2,3,4\}\Big\}$. The distribution defined over the graph is a log-linear model with clique potentials given by $\Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) = e^{w_c f_c(\mathbf{y}_c)}$, where all the weights $\mathbf{w}_c$, and the binary functions f_c are known. For example, for $c = \{1,2\}$, $\mathbf{w}_{\{1,2\}} = 1.5$ and

$$f_{\{1,2\}}(\mathbf{y}_{\{1,2\}}) = \begin{cases} 1 & , \mathbf{y}_{\{1,2\}} \in \{(s_1,l_1),(s_1,l_0),(s_0,l_0)\} \\ 0 & , \mathbf{y}_{\{1,2\}} \in \{(s_0,l_1)\} \end{cases}$$

Each binary function f_c induces a set $B_c = \{(\omega_1,\omega_2,\omega_3,\omega_4) : f_c(\omega_c) = 1\}$. For example, $B_{\{1,2\}} := \big\{\omega : \omega_{\{1,2\}} \in \{(s_1,l_1),(s_1,l_0),(s_0,l_0)\}\big\}$.

We compare predictions between the baseline and alternatives of Type I and II (see Section 5.2) for the following QoIs:

$$g(\mathbf{Y}) = \mathbf{1}_A, \quad \text{for any event of interest } A \subset \Omega. \tag{5.1}$$

For instance, $A = \{\text{patient is smoker with asthma}\} = \{\omega = (\omega_1,\omega_2,\omega_3,\omega_4) : \omega_1 = s_0, \omega_3 = a_0\}$.

5.2. Alternative models.

5.2.1. Type I. First, we consider the class of log-linear models $\tilde{p}$ over $\mathcal{G}$ with weight change in one maximal clique. Let c be the maximal clique that a weight change occurred. Then the clique potential is given by

$$\tilde{\Psi}_c(\mathbf{y}_c) = e^{\tilde{w}_c f_c(\mathbf{y}_c)} \tag{5.2}$$

The weight after increasing or decreasing by $100a\%$ equals to $\tilde{w}_c = (1+a)w_c$, where $a \in [-1,1]$ stands for the model uncertainty of alternative models of Type I and w_c is the weight on c of the baseline model p. Note that $\mathcal{B} = \{c\}$, where $\mathcal{B}$ defined in subsection 4.2.1.

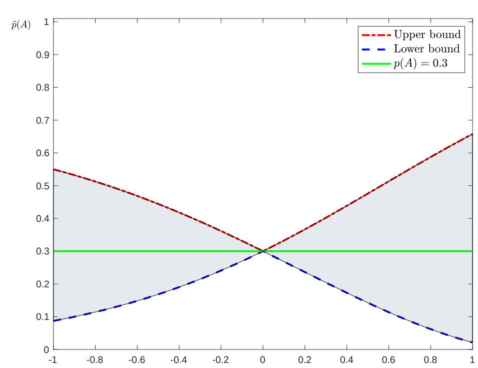

FIG. 4. *For any event of interest, A with $p(A) = 0.3$, the red dashed-dot and the blue dashed curve are the upper bound and lower bound for $\tilde{p}(A)$ provided in (5.4), computed as functions of the weight change a.*

EXAMPLE 5.1. *Let p be as in section 5.1 and $\tilde{p}$ be in the class of log-linear models defined in Section 5.2.1. Also, let B_c be the induced set by the binary function f_c, where c is the maximal clique that a weight change occurred. If $p_{\mathrm{I}} \equiv p(B_c)$ and $a \in [-1, 1]$ (depends on $\tilde{p}$), then*

$$\frac{d\tilde{p}}{dp} = \frac{\Phi^{\mathrm{I}}}{E_p[\Phi^{\mathrm{I}}]} = \frac{e^{aw_c f_c}}{e^{aw_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}}}. \tag{5.3}$$

Furthermore, for any event of interest A, the following holds:

$$\pm\, \tilde{p}(A) \le \inf_{\lambda>0} \frac{1}{\lambda} \left\{ \log\left(\frac{p(A)e^{\pm\lambda} + 1 - p(A)}{e^{aw_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}}} \right) - \frac{aw_c e^{aw_c} p_{\mathrm{I}}}{e^{aw_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}}} \right\} \tag{5.4}$$

Proof. We express $\tilde{\Psi}_c = \Psi_c e^{(\tilde{w}_c - w_c) f_c} = \Psi_c e^{aw_c f_c}$. Then by applying Lemma 4.3, we get (5.3). The bounds provided in (5.4) are direct consequence of Theorem 4.7, (a). In Appendix D.4, we compute all the quantities involved in (4.15) explicitly. ∎

In Figure 4, we demonstrate the UQ bounds provided in (5.4) as functions of the uncertainity parameter a for any event of interest A with $p(A) = 0.3$ and when $p_{\mathrm{I}} = 0.2$.

5.2.2. Type II. We consider the class of log-linear models $\tilde{p}$ over $\tilde{\mathcal{G}}$ with $\tilde{\mathcal{V}} = \mathcal{V}$, $\tilde{\mathcal{E}} = \mathcal{E} \cup e$, where e is a new edge (for example, see Figure 2). We assume that the edge e enlarges an already existing maximal clique in the sense of the analysis in subsection 4.2.2, (4.7). The model uncertainties lie in the binary function $\tilde{f}_{\tilde{c}}$ defined on $\tilde{c}$ and the new weight $\tilde{\mathbf{w}}_{\tilde{c}}$, where $\tilde{c}$ is the enlargement of an existing maximal clique c. The weight $\tilde{w}_{\tilde{c}}$ can also be expressed with respect to w_c: $\tilde{w}_{\tilde{c}} = (1 + a)w_c$. This time $a \in \mathbb{R}$, not necessarily in $[-1, 1]$ as before (e.g $w_c = 1.5$ and $\tilde{w}_{\tilde{c}} = 5$). Then the corresponding clique potential is given by

$$\tilde{\Psi}_{\tilde{c}}(\mathbf{y}_{\tilde{c}}) = e^{\tilde{w}_{\tilde{c}} \tilde{f}_{\tilde{c}}(\mathbf{y}_{\tilde{c}})} = e^{(1+a)w_c \tilde{f}_{\tilde{c}}(\mathbf{y}_{\tilde{c}})}$$

The binary function $f_{\tilde{c}}$ induces a set $B_{\tilde{c}} = \{(\omega_1, \omega_2, \omega_3, \omega_4) : \tilde{f}_{\tilde{c}}(\omega_{\tilde{c}}) = 1\}$. The set $B_{\tilde{c}}$ satisfies one of the following: $B_{\tilde{c}} \cap B_c = \emptyset$ or $B_{\tilde{c}} \cap B_c \neq \emptyset$. Note that $\mathcal{B}_{\subseteq} = \{\tilde{c}\}$ and $\mathcal{B}_{\cup} = \mathcal{B}_{new} = \emptyset$ with $\mathcal{B}_{\subseteq}, \mathcal{B}_{\cup}$ and $\mathcal{B}_{new}$ are defined in subsection 4.2.2. The next proposition gives us the UQ bounds when $B_{\tilde{c}} \cap B_c = \emptyset$.

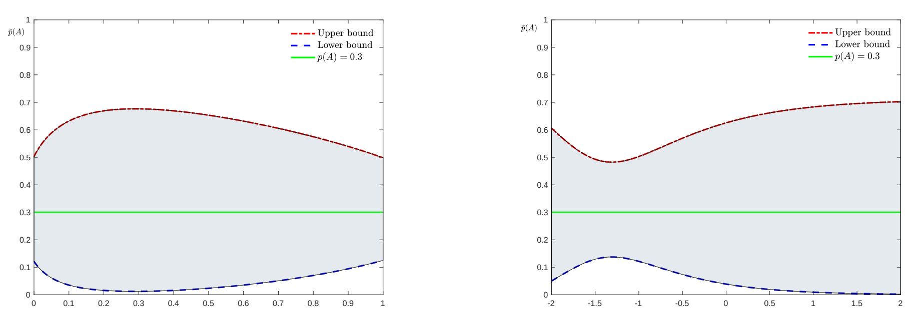

FIG. 5. *A is an event of interest with $p(A) = 0.3$. (Left) For $p_{\mathrm{I}} = 0.2$, $w_c = 1.5$ and $a = -0.2$, the red dash-dot and the blue dashed curves are the upper bound and lower bound for $\tilde{p}(A)$ provided in (5.6), computed as functions of p_{II}. (Right) For $p_{\mathrm{I}} = 0.2$, $w_c = 1.5$, $p_{\mathrm{II}} = 0.7$, the red curve and the blue are the upper bound and lower bound for $\tilde{p}(A)$, computed as functions of the weight change $a \in [-2, 2]$.*

Example 5.2. *Let p be as in section 5.1 and $\tilde{p}$ be in the class of log-linear models defined in Section 5.2.2. Also, let B_c and $B_{\tilde{c}}$ be the induced sets by the binary functions f_c and $\tilde{f}_{\tilde{c}}$ respectively such that $B_{\tilde{c}} \cap B_c = \emptyset$, where $\tilde{c}$ is an enlargement of the maximal clique c. If $p_{\mathrm{I}} \equiv p(B_c)$, $p_{\mathrm{II}} \equiv p(B_{\tilde{c}})$ and $a \in [-1, 1]$, then*

$$\frac{d\tilde{p}}{dp} = \frac{\Phi^{\mathrm{II}}}{E_p[\Phi^{\mathrm{II}}]} = \frac{e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c}}{1 - (1 - e^{(1+a)w_c})p_{\mathrm{II}} - (1 - e^{-w_c})p_{\mathrm{I}}}. \tag{5.5}$$

Furthermore, for any event of interest A, the following holds:

$$\begin{aligned} \pm\tilde{p}(A) \leq \inf_{\lambda > 0} \frac{1}{\lambda} \Bigg\{ &\log\left(\frac{p(A)e^{\pm\lambda} + 1 - p(A)}{1 - (1 - e^{(1+a)w_c})p_{\mathrm{II}} - (1 - e^{-w_c})p_{\mathrm{I}}} \right) \\ &- \frac{w_c e^{-w_c} p_{\mathrm{I}} - (1+a) w_c e^{(1+a)w_c} p_{\mathrm{II}}}{1 - (1 - e^{(1+a)w_c})p_{\mathrm{II}} - (1 - e^{-w_c})p_{\mathrm{I}}} \Bigg\} \end{aligned} \tag{5.6}$$

Proof. The proof is given in Appendix D.5. ∎

Remark 5.3. The case where $B_{\tilde{c}} \cap B_c \neq \emptyset$ is more complicated. However, the KL divergence is still explicitly computable. The computations are in the same spirit of (5.6) and are given in Remark D.2.

In Figure 5, we demonstrate the bounds in 5.6 for any event A with $p(A) = 0.3$ as functions of the uncertainty parameters a (when $p_{\mathrm{I}} = 0.2$, $w_c = 1.5$ and p_{II}=0.7) and p_{II} (when $p_{\mathrm{I}} = 0.2$, $w_c = 1.5$, $a = -0.2$).

5.3. Tightness. Let g be the QoI given by (5.1). By applying Theorem 4.8, there exist $0 < \eta_\pm \leq \infty$ such that for any $\eta \leq \eta_\pm$ there exist probability measures $p^\pm = p^\pm(\eta) \in \mathcal{Q}^\eta_{\mathcal{P}}$, where $\mathcal{Q}^\eta_{\mathcal{P}}$ is given in (4.18), such that (4.15) becomes an equality. Furthermore, $p^\pm = p^{\lambda_\pm}$ with the total $p^{\lambda_\pm}$-excess factor relative to p being $\Phi^\pm = e^{\lambda_\pm \mathbf{1}_A}$. Therefore,

$$dp^{\lambda_\pm} = \frac{e^{\lambda_\pm \mathbf{1}_A}}{p(A)e^{\lambda_\pm} + 1 - p(A)}\, dp \tag{5.7}$$

and $\lambda_\pm$ being the unique solutions of $R(p^{\lambda_\pm} \| p) = \eta$. Depending on the event of interest A, we can determine the graph associated with $p^{\lambda_\pm}$. Specifically, if $A = \cap_i A_i$ with all A_i being defined on the same maximal clique of $\mathcal{G}$ given in Figure 1, then the graph associated with $p^{\lambda_\pm}$ is $\mathcal{G}$ and hence both models are Type I. If at least two A_i, A_j are defined on different maximal cliques, it turns out that the associated graphs are different than $\mathcal{G}$. For example, let $A = \{$patient is smoker with asthma$\} = \{\omega = (\omega_1, \omega_2, \omega_3, \omega_4) : \omega_1 = s_0, \omega_3 = a_0\} = \{\omega : \omega_1 = s_0\} \cap \{\omega : \omega_3 = a_0\}$. Since the total $p^\pm$-excess factor relative to p $\Phi^\pm = e^{\lambda_\pm \mathbf{1}_A}$ cannot be further factorized, it is implied that the new graph has the same set of nodes with an extra edge $1-3$, that is $\tilde{\mathcal{E}} = \mathcal{E} \cup \{1-3\}$. In that case, both models are Type II.

6. UQ for Statistical Mechanics. Large-scale physical systems of interacting particles such as gases, liquids, and solids, are at the core of statistical mechanics and in particular of equilibrium statistical mechanics. The macroscopic properties of a system can be understood through its underlying microscopic description which fundamentally requires the microscopic states and an interaction between microscopic constituents. Statistical mechanics models such as the Ising model provide important foundational models in ML such as Boltzmann machines, [36].

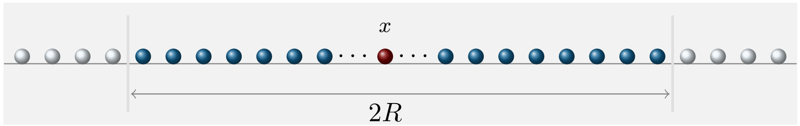

FIG. 6. *One-dimensional Ising spin lattice on Δ (light gray area with blue, red, and white particles). The spin located at $x \in \Delta$ (red particle) interacts only with spins located at y in $B_x(R)$ (blue particles) with strength of interaction $J(x, y)$. The red spin does not interact with the white ones as they are located at distance greater than R from x.*

6.1. Ising Model. An illustrative example is the Ising model, where the space of all microstates is the collection of all spin configurations on a bounded region $\Delta \subseteq \mathbb{Z}^d$:

$$\Omega := \{\pm 1\}^{\Delta} = \Big\{\sigma_\Delta = \{\sigma_\Delta(x)\}_{x\in\Delta} : \sigma_\Delta(x) \in \{+1, -1\}\Big\}$$

as in Figure 6, [55, 47]. An interaction between spins can be short, long range or a combination (such as Lennard-Jones potential, [56]), positive (ferromagnetism), etc, [55, 29, 32]. Here we consider a d-dimensional Ising spin system on Δ with a generic interaction $\mathbf{J} = \{J(x, y) : x, y \in \Delta\}$ satisfying three properties: for all $x, y \in \Delta$ and $z \in \mathbb{R}^d$

$$J(x + z, y + z) = J(x, y) \qquad \text{(translational invariance)} \tag{6.1}$$

$$J(x, y) = J(y, x) \qquad \text{(symmetry)} \tag{6.2}$$

$$\sum_{x\neq 0} |J(0, x)| < \infty \qquad \text{(summability)} \tag{6.3}$$

and an external field, $h \in \mathbb{R}$. Let $R > 0$ be the length of the range of interaction. For $x \in \mathbb{Z}^d$, $B_x(R) = \{y \in \mathbb{Z}^d : \|x - y\|_d \leq R\}$ is the set of all spins that the spin located at x interacts with. For convenience, we denote $B^{\neq}_{x,R} := B_x(R) \setminus x$.

6.1.1. Boundary conditions. Boundary conditions are a fundamental concept in statistical mechanics, [60]. For simplicity, let us assume that Δ is a cube. We consider a system where particles not only interact with particles in Δ, but also with particles "outside" of Δ. Let $\bar{\sigma}_{\Delta^c}$ be a given fixed configuration of spins on the complement of Δ denoted by Δ^c, see Figure 11. The hamiltonian energy of the system is given by:

$$H^{\mathbf{J},h}(\sigma_\Delta|\bar{\sigma}_{\Delta^c}) = H^{\mathbf{J},h}(\sigma_\Delta) - \sum_{x\in\Delta}\sum_{y\in\Delta^c} J(x, y)\sigma_\Delta(x)\sigma_\Delta(y) \tag{6.4}$$

where

$$H^{\mathbf{J},h}(\sigma_\Delta) = -\frac{1}{2}\sum_{x\in\Delta}\sum_{y\in\Delta} J(x, y)\sigma_\Delta(x)\sigma_\Delta(y) - h\sum_{x\in\Delta}\sigma_\Delta(x) \tag{6.5}$$

The Gibbs measure with boundary condition $\bar{\sigma}_{\Delta^c}$ is defined as

$$\mu^{\Delta}_{\mathbf{J},\beta,h}(\sigma_\Delta \mid \bar{\sigma}_{\Delta^c}) = \frac{1}{Z_{\bar{\sigma}_{\Delta^c}}(\mathbf{J}, \beta, h)} e^{-\beta H^{\mathbf{J},h}(\sigma_\Delta|\bar{\sigma}_{\Delta^c})}. \tag{6.6}$$

where $Z_{\bar{\sigma}_{\Delta^c}}(\mathbf{J}, \beta, h) = \sum_{\sigma_\Delta} e^{-\beta H^{\mathbf{J},h}(\sigma_\Delta|\bar{\sigma}_{\Delta^c})}$ is the partition function.

6.1.2. rMRF formulation. A system with configuration as boundary conditions does not admit a MRF description. So, we describe the system using rMRFs. The set of nodes is $\mathbb{Z}^d$, the set of edges can be constructed by looking at all (x,y) such that $\|x-y\|_d \leq R$ and the context is $\mathbf{u} = \bar{\sigma}_{\Delta^c}$, which corresponds to a fixed boundary condition. Then $(\sigma_\Delta, \mu^{\Delta}_{\mathbf{J},\beta,h}(\cdot \mid \bar{\sigma}_{\Delta^c}))$ is a rMRF with maximal cliques $c_x = \{y \in \Delta : y \in B^{\neq}_{x,R}\}$ (spins in c_x interact with all spins in c_x). Let $\mathbf{w} = \{\mathbf{w}_{c_x}\}_{x\in\Delta}$ with $\mathbf{w}_{c_x} = (\mathbf{J}_{c_x}, \beta, h)$ and $\mathbf{J}_{c_x} = \{J(x,y) : y \in c_x\}$. We express each clique potential as

$$\Psi_{c_x} = \exp\left\{\beta\sigma_\Delta(x)\left(h + \frac{1}{2}\sum_{\substack{y\in\Delta\\ y\in B^{\neq}_{x,R}}} J(x,y)\sigma_\Delta(y) + \sum_{\substack{y\in\Delta^c\\ y\in B^{\neq}_{x,R}}} J(x,y)\bar{\sigma}_{\Delta^c}(y)\right)\right\} \tag{6.7}$$

Note that we may resume the full notation when we needed, that is $\Psi_{c_x} \equiv \Psi_{c_x}[\bar{\sigma}_{\Delta^c}](\sigma_{c_x} \mid \mathbf{w}_{c_x})$ where σ_{c_x} is the Ising spin configuration defined on all $y \in c_x$.

6.2. UQ Formulation.

6.2.1. Alternative models. We consider models on a lattice with perturbed interaction in the strength (Type I) and/or range (Type II) such as truncated or long range interaction. Given $\mathbf{J}$ as in subsection 6.1, an interaction $F(x,y)$ satisfying (6.1)-(6.3) with length of range R_F, we say that $\tilde{\mathbf{J}}^{\mathbf{F}} = \{\tilde{J}^F(x,y) : x,y \in \mathbb{Z}^d\}$ is a perturbed interaction if

$$\tilde{J}^F(x,y) = J(x,y)\mathbf{1}_{\|x-y\|_d\leq R} + F(x,y)\mathbf{1}_{\|x-y\|_d\leq R_F} + F(x,y)\mathbf{1}_{\|x-y\|_d> R_F} \tag{6.8}$$

We say that a perturbed interaction is Type I iff

$$R = R_F \text{ and } \operatorname{supp}(F) = \{(x,y) : \|x-y\|_d \leq R_F\}. \tag{6.9}$$

We say that a perturbed interaction is Type II iff

$$R = R_F \text{ and } \operatorname{supp}(F) = \{(x,y) : \|x-y\|_d > R_F\}. \tag{6.10}$$

The rMRF formulation of the system with $\tilde{\mathbf{J}}^{\mathbf{F}}$ goes similarly as in subsection 6.1.2. Note that the graph representation simplifies a possible complexity of J, F and $\tilde{J}^F$ as we connect nodes x,y according to the range of J, F and $\tilde{J}^F$ and assign the corresponding strengths $J(x,y)$, $F(x,y)$ and $\tilde{J}^F(x,y)$.

6.2.2. Total $\tilde{q}_\Delta$-excess factor relative to q_Δ. Let $\tilde{\mathbf{J}}^{\mathbf{F}}$ be defined in subsection 6.2.1 with support given by (6.9) or (6.10), and $q_\Delta(\cdot) := \mu^{\Delta}_{\mathbf{J},\beta,h}(\cdot \mid \bar{\sigma}_{\Delta^c})$ and $\tilde{q}_\Delta(\cdot) := \mu^{\Delta}_{\tilde{\mathbf{J}}^{\mathbf{F}},\beta,\tilde{h}}(\cdot \mid \bar{\sigma}_{\Delta^c})$ be the corresponding Gibbs measures defined in (6.6). It is straightforward (see Appendix D.8) that the total q_Δ-excess factor for $i = \mathrm{I},\mathrm{II}$ is given by

$$\begin{aligned}\Phi^{\mathrm{i}}_{\bar{\sigma}_{\Delta^c}}(\sigma_\Delta) = \exp\Big\{\beta\sum_{x\in\Delta}\sigma_\Delta(x)\Big((\tilde{h}-h) + \frac{1}{2}\sum_{y\in A^{\mathrm{i}}_x\cap\Delta} F(x,y)\sigma_\Delta(y)\\ + \sum_{y\in A^{\mathrm{i}}_x\cap\Delta^c} F(x,y)\bar{\sigma}_{\Delta^c}(y)\Big)\Big\}\end{aligned} \tag{6.11}$$

where for each $x \in \Delta$, $A^{\mathrm{I}}_x = B_x(R)$ and $A^{\mathrm{II}}_x = B_x(R)^c$, with $B_x(R)^c$ being the complement of $B_x(R)$. The total $\tilde{q}_\Delta$-excess factor relative to q_Δ represents the discrepancies between the models which are encoded in all terms of (6.11). In particular, $(\tilde{h}-h)$ and $F(x,y)$ point out how different the external fields and interactions are respectively, as the latter satisfies $F(x,y) = \tilde{J}^F(x,y) - J(x,y)$.

6.2.3. Quantities of Interest. The use of phase diagrams is central in physics and material science. A phase diagram is defined as a graphical representation of equilibrium states under different thermodynamic parameters such as external field h, temperature T and pressure P. It is typically computed in the thermodynamic limit (i.e a limiting process with $\Delta \nearrow \mathbb{Z}^d$ such that the ratio between inter-atomic distances and macroscopic lengths vanishes), [55]. Equilibrium states are characterized by order parameters such as magnetization. For that, we consider the following observable

$$m(\sigma_\Delta) := \frac{1}{|\Delta|}\sum_{x\in\Delta}\sigma_\Delta(x) \tag{6.12}$$

where $|\Delta|$ stands for the volume of a cube $\Delta \subset \mathbb{Z}^d$. As Δ invades the whole $\mathbb{Z}^d$, the expectation of $m(\sigma_\Delta)$ yields the magnetization. Other QoIs could also be considered e.g. correlation functions $v(\sigma_\Delta) = \frac{1}{|\Delta|^2}\sum_{x\in\Delta}\sum_{y\in\Delta}\sigma_\Delta(x)\sigma_\Delta(y)$.

6.3. UQ for finite-size effects. Here we capture the behavior of the average of spins $m(\sigma_\Delta)$ given in (6.12) with respect to the perturbed model $\tilde{q}_\Delta$ by applying the analysis of subsections 4.3, 4.4 and 4.5 under the above statistical mechanics formulation though rMRFs.

Let Δ be a cube in $\mathbb{Z}^d$. Given a configuration $\bar{\sigma}_{\Delta^c}$, the baseline model is an Ising model with interaction $\mathbf{J}$ defined in subsection 6.1. Alternative models are Ising models with a perturbed interaction $\tilde{\mathbf{J}}^{\mathbf{F}}$ defined in subsection 6.2.1.

6.3.1. Cumulant Generating Function. We compute the cumulant generating function defined by (4.2) w.r.t the baseline model q_Δ (the computation is given in (D.8)):

$$\Lambda_{q_\Delta;|\Delta|m(\sigma_\Delta)}(\pm\lambda) = \beta|\Delta|\left(P^\Delta_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}}(\bar{\sigma}_{\Delta^c}) - P^\Delta_{h,\beta,\mathbf{J}}(\bar{\sigma}_{\Delta^c})\right) \tag{6.13}$$

where $P^\Delta_{h,\beta,\mathbf{J}}$ stands for the thermodynamic pressure, [55], defined as

$$P^\Delta_{h,\beta,\mathbf{J}}(\bar{\sigma}_{\Delta^c}) := \frac{Z(\mathbf{J},\beta,h,\bar{\sigma}_{\Delta^c})}{\beta|\Delta|}.$$

6.3.2. KL Divergence. Here we utilize Lemma 4.5 and specify the KL divergence in terms of the quantities involved in (4.14). Then we bound it by using Lemma D.3. Before that, we present a way that bounds the KL divergence with the use of a well-established tool in statistical mechanics referred to as norm-$\|\cdot\|_1$, [60]. After all, we conclude that our approach obtains better uncertainty areas.

Norm-$\|\cdot\|_1$: Let $\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}}(\sigma_X)$ be the following quantity:

$$\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}}(\sigma_X) = \begin{cases} -\frac{1}{2}\beta J(x,y)\sigma_\Delta(x)\sigma_\Delta(y) & ,X=\{x,y\},\ x\neq y, \\ -\beta\sigma_\Delta(x)\big(h+\sum_{y\in B^{\neq}_{x,R}\cap\Delta^c}J(x,y)\bar{\sigma}_{\Delta^c}(y)\big) & ,X=\{x\} \\ 0 & ,\text{otherwise} \end{cases} \tag{6.14}$$

and similarly we define $\Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}(\sigma_X)$. Then,

$$\beta H^{\mathbf{J},h}(\sigma_\Delta|\bar{\sigma}_{\Delta^c}) = \sum_{X:X\cap\Delta\neq\emptyset}\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}}(\sigma_X) \tag{6.15}$$

Also, $\beta H^{\tilde{\mathbf{J}}^{\mathbf{F}},\tilde{h}}(\sigma_\Delta|\bar{\sigma}_{\Delta^c})$ is defined similarly. Then the norm-$\|\cdot\|_1$ of $\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}} - \Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}$ is defined as

$$\|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}} - \Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}\|_1 = \sum_{0\ni X} \|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}} - \Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}\|_\infty \tag{6.16}$$

where $\|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}} - \Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}\|_\infty = \sup_{\sigma_X} |\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}}(\sigma_X) - \Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}(\sigma_X)|$ for $X \subset \mathbb{Z}^d$.

LEMMA 6.1. *Let F be an interaction satisfying (6.1)-(6.3) with support given by (6.9) or (6.10), then*

$$R(\tilde{q}_\Delta\|q_\Delta) \le 2|\Delta|\|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar{\sigma}_{\Delta^c}} - \Phi^{\tilde{h},\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar{\sigma}_{\Delta^c}}\|_1 \le 2\beta|\Delta| \left(|\tilde{h} - h| + \sum_{x\neq 0} |F(0,x)| \right).$$

Proof. See Appendix D.7. ∎

Let us turn to our approach developed in Section 4. We recall the total $\tilde{q}_\Delta$-excess factor relative to q_Δ from subsection 6.2.2 as well as the quantities from Section 4.3, and we express $\log \Phi^{\mathrm{i}}_{\bar{\sigma}_{\Delta^c}}(\sigma_\Delta) = C_{\mathrm{i}}|\Delta|m(\sigma_\Delta) + \kappa_{\mathrm{i}}(\sigma_\Delta)$ with

$$C_{\mathrm{i}} = \beta(\tilde{h} - h) < 1, \quad \kappa_{\mathrm{I}}(\sigma_\Delta) = \beta \sum_{x\in\Delta} \sigma_\Delta(x) \left(\frac{1}{2} \sum_{y\in A^{\mathrm{i}}_x \cap \Delta} F(x,y)\sigma_\Delta(y) \right) + \beta F(\Delta|\bar{\sigma}_{\Delta^c}) \tag{6.17}$$

where $F(\Delta|\bar{\sigma}_{\Delta^c}) = \sum_{x\in\Delta} \sum_{y\in A^{\mathrm{i}}_x\cap\Delta^c} F(x,y)\bar{\sigma}_{\Delta^c}(y)$.

6.3.3. Scalability in system size $|\Delta|$. The high-dimensionality of statistical mechanics models requires scalability. In an UQ standpoint, this is encoded by having meaningful UQ bounds at the thermodynamic limit. The MGF and the KL divergence are scalable quantities as shown in (6.13), Lemma 6.1 and (D.7) (all are multiplied by the size of the system $|\Delta|$) and, by extension the UQ bounds as well. The analysis from now on refers to models of Type I. Although Type II models can be worked on similarly, one example of Type II is discussed in Appendix E.

To get the UQ bounds for $E_{\tilde{q}_\Delta}[m(\sigma_\Delta)]$ we can either apply (4.1) using the crude bound in Lemma 6.1 with $f(\mathbf{Z}) = |\Delta|m(\sigma_\Delta)$:

$$\pm E_{\tilde{q}_\Delta}[m(\sigma_\Delta)] \le \inf_{\lambda>0} \left\{ \frac{P^\Delta_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}} - P^\Delta_{h,\beta,\mathbf{J}}}{\lambda/\beta} + 2\frac{\beta}{\lambda}(|\tilde{h} - h| + \mathcal{F}) \right\} \tag{6.18}$$

or Theorem 4.7 with $\kappa_{\mathrm{i}}(\sigma_\Delta)$ bounded by using (D.7) with $f(\mathbf{Z}) = |\Delta|m(\sigma_\Delta)$:

$$\pm E_{\tilde{q}_\Delta}[m(\sigma_\Delta)] \le \frac{1}{1-\beta(\tilde{h}-h)} \inf_{\lambda>0} \left\{ \frac{P^\Delta_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}} - P^\Delta_{h,\beta,\mathbf{J}}}{\lambda/\beta} + \frac{\beta}{\lambda}\mathcal{F}\left(1 + R_F\frac{|\partial\Delta|}{|\Delta|}\right) \right\} \tag{6.19}$$

with $\mathcal{F} := \sum_{x\neq 0} |F(0,x)|$ which is bounded due to the property (6.3) and $R_F = R$.

6.4. UQ for Phase Diagrams. Here we capture the phase diagram of the perturbed model $\tilde{q}_\Delta$ looking at the magnetization defined in subsection 6.2.3. We study the limit of the bounds obtained in subsection 6.3.3. In fact, let $M(\tilde{\mathbf{J}}^{\mathbf{F}}, \tilde{\beta}, \tilde{h})$ be the limit as $\Delta \nearrow \mathbb{Z}^d$ of $E_{\tilde{q}_\Delta}[m(\sigma_\Delta)]$. Due to scalability we can pass to the limit $\Delta \nearrow \mathbb{Z}^d$ of (6.19):

$$\pm M(\tilde{\mathbf{J}}^F, \beta, \tilde{h}) \le \frac{1}{1-\beta(\tilde{h}-h)} \inf_{\lambda>0} \left\{ \frac{\left(P_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}} - P_{h,\beta,\mathbf{J}}\right)}{\lambda/\beta} + \frac{\beta}{\lambda}\mathcal{F} \right\} \tag{6.20}$$

with $\lim_{\Delta \nearrow \mathbb{Z}^d} P^{\Delta}_{h,\beta,\mathbf{J}} = P_{h,\beta,\mathbf{J}}$ by Theorem 2.3.3.1 in [55] and $\lim_{\Delta \nearrow \mathbb{Z}^d} \frac{|\partial \Delta|}{|\Delta|} = 0$.

Remark 6.2. If we consider the case where $\tilde{\beta} \neq \beta$, the formulae and bounds adjust similarly. When $\tilde{h} = h$, the limit of (6.18) is the same but the last term in the infimum, $\frac{\beta}{\lambda}\mathcal{F}$, is multiplied by 2. Therefore, the uncertainty area of the phase diagram obtained by using the bounds in (6.20) is significantly better.

In the next section, we illustrate the uncertainty area of the phase diagram for both (6.20) and the limit of (6.18) when the baseline model is an Ising-Kac model (i.e. a finite range approximation of mean field interactions). The alternative models are a Kac perturbation, a truncated Kac interaction and perturbation by a long range interaction given in Appendix E.

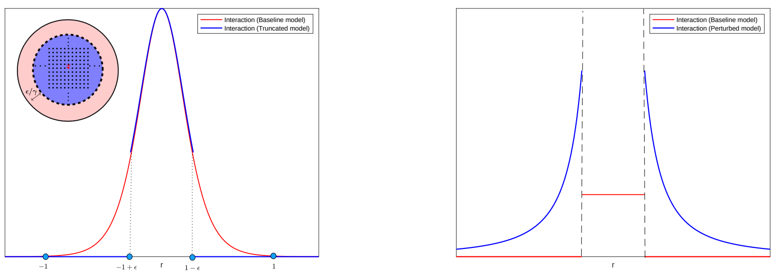

FIG. 7. *(Left) The red curve is a Kac interaction and the blue curve is a truncation of it. The two curves coincide at all r with $|r| \leq 1-\epsilon$. The embedded picture demonstrates the two interactions at the microscopic level. The red particle located at the site $x \in \Delta \subset \mathbb{Z}^2$ interacts with the particles in the blue and the light red through J_γ.The particle interacts only with the particles in the blue area through $\tilde{J}_\gamma^{-J}$ with range $\gamma^{-1}(1-\epsilon)$. (Right) The red curve is an example of Kac interaction (piecewise constant) with $J(r) = \mathbf{1}_{r \leq \frac{1}{2}}(r)$ and the blue curve is a perturbation given by $G(r) = \frac{a}{r^2}\mathbf{1}_{r > \frac{1}{2}}(r)$,for some $a > 0$.*

6.5. Ising-Kac Model. An Ising-spin model with a Kac-type interaction, defined in Appendix B, behaves like a mean field (or Van der Waals model in gas lattice) in the limit with the convexity of free energy emerging naturally in the limit, contrary to mean field or Curie-Weiss models where Maxwell's equal area law is required to refine the non-convex free energy (double well shape), [55]. Such a discrepancy comes from the fact that each spin interacts with all particles in the same way and independently. The idea of Kac was to keep such a picture on large regions but relatively small compared to the range of interaction. Then, the thermodynamical incorrect of the free energy on these large regions looks refined at the scale of interaction. Therefore, the system contains a two-scale behavior that was carried out by introducing a small parameter $\gamma > 0$ known as *Kac scaling.* As we suppose that an Ising spin model is endowed by such an interaction, the model has overall three scales: the lattice spacing is 1, the range of interaction is γ^{-1} while the size of the system is much larger than γ^{-1} and all are well-separated, contrary to the mean field model where the range of interaction is the same as the size of the system.

In the next subsections, Δ is a bounded, $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable region, with $L >> \gamma^{-1}$, see Appendix B.2. The hamiltonian energies and Gibbs measures of the next perturbed models are similar to those in Appendix B (or in subsection 6.1.1). Also, their rMRF formulations are structured analogously to the ones in subsection 6.1.2 and for that we omit it. The thermodynamic pressure for the Ising-Kac model denoted by $P^{\Delta,\gamma}_{\mathbf{J},\beta,h}$ is given in (C.1) and its Lebowitz-Penrose (LP) limit (i.e $\lim_{\gamma\to 0}\lim_{\Delta\nearrow\mathbb{Z}^d}$)

$p_{\mathbf{J},\beta,h}$ is given in (C.2) (see also Appendix C.1, C.2 and C.3 for further discussion).

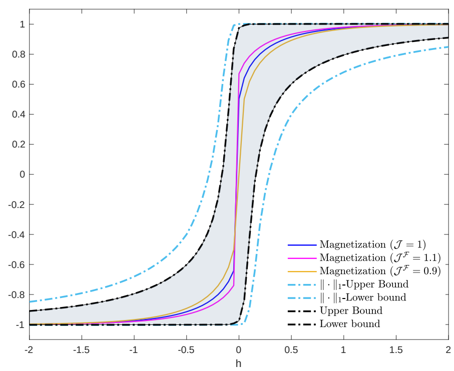

FIG. 8. *The curves in blue, magenta and dark yellow color are the magnetizations of the Ising model with Kac interaction at inverse temperature $\beta = \tilde{\beta} = 1.1$, $\tilde{h} = h$, and total strength $\mathcal{J} = 1$, $\tilde{\mathcal{J}}^F = 1.1$ $(a = 0.1)$ and 0.9 $(a = -0.1)$ (validation) respectively. The black dashed-dot curves are the UQ upper and lower bounds provided by Corollary 6.4 and viewed as functions of $h \in [-2, 2]$. The gray area depicts the size of the uncertainty region. The light blue dashed-dot curves are the UQ upper and lower bounds obtained using norm-$\|\cdot\|_1$.*

6.5.1. Phase Diagram of Perturbed Kac model. Let define a perturbation of a Kac potential.

DEFINITION 6.3. *Let F_γ be an even function satisfying (6.1)-(6.3) and (B.1) with length of range γ^{-1} and $\mathcal{F} := \int_{\mathbb{R}^d} F(r)dr$. We define*

$$\tilde{J}^F_\gamma(x, y) = J_\gamma(x, y) + F_\gamma(x, y), \text{ such that } \mathcal{F} = a\mathcal{J} \tag{6.21}$$

with $a \in [-1, 1]$ and $\mathcal{J}$ is given in Appendix B.1.

The parameter a represents the percentage of increase or decrease of the total strength of interaction $\tilde{\mathcal{J}}^F := \int_{\mathbb{R}^d} \tilde{J}^F(r)dr = (1 + a)\mathcal{J}$.

COROLLARY 6.4. *Let $\tilde{J}^F$ be the interaction given in Definition 6.3. Then, for $\gamma > 0$ small enough, the UQ bounds (6.18) and (6.19) hold for $R_F = R = \gamma^{-1}$ and $\mathcal{F} = |a|\mathcal{J}$. The thermodynamic pressure $P^{\Delta,\gamma}_{\mathbf{J},\beta,h}$ is given in (C.1). Let $M(\tilde{\mathbf{J}}^F, \beta, \tilde{h})$ be the LP-limit of $E_{\tilde{q}_\Delta}[m(\sigma_\Delta)]$. Then, the UQ bounds (6.20) and LP-limit of (6.18) hold with the LP-limit of $P^{\Delta,\gamma}_{\mathbf{J},\beta,h}$ being $p_{\mathbf{J},\beta,h}$ given in (C.2).*

Remark 6.5. We recall from the analysis of the Ising model with a generic interaction J satisfying (6.1)-(6.3), that (6.18) represents crude bounds as norm-$\|\cdot\|_1$ (subsection 6.3.2) has been used, while (6.19) obtained by Theorem 4.7, includes more detail. The difference is illustrated in Figure 8. Furthermore, even if there is a γ^{-1} in the term $2\gamma^{-1}\frac{|\partial\Delta|}{|\Delta|}$ in (6.18), the order of the LP-limit makes it vanish as $L \to \infty$.

Validation. Given β, h, $\mathcal{J}$ and a tolerance $\eta > 0$, we can construct with the use of norm-$\|\cdot\|_1$ and Lemma 6.1 a class of models such that $\mathcal{Q}^{\mathrm{I}}_\eta := \{\tilde{q}_\Delta : 2\beta a\mathcal{J} \leq \eta\}$. This is subclass of $\mathcal{Q}^\eta$ defined in (1.1) with the KL divergence in place of d. In Figure 8, $\beta = 1.1$ and $\mathcal{J} = 1$ while the external field h varies from -2 and 2. The positive parameter $\eta = 0.1$ and the perturbed model with 10% decrease $(a = -0.1)$ of the total strength (magnetization with magenta color) is in $\mathcal{Q}^{\mathrm{I}}_{0.1}$ as demonstrated in dark yellow color.

6.5.2. Phase diagram of Truncated Potential. From a computational point of view, macroscopic properties of high dimensional systems can be studied through simulation models where one can consider an appropriate truncated interaction which can reduce the computational overhead associated with the interaction [66, Chapter 3]. In our context, a truncated interaction can be defined as follows:

DEFINITION 6.6. *Let $0 < \epsilon < 1$. We define the truncated interaction as*

$$\tilde{J}^{-J}(0,r) = \begin{cases} J(0,r) & , |r| \leq 1-\epsilon \\ 0 & , \textit{otherwise} \end{cases} \tag{6.22}$$

The parameter ϵ quantifies the truncation of the interaction as means the cut off area in the domain of J.

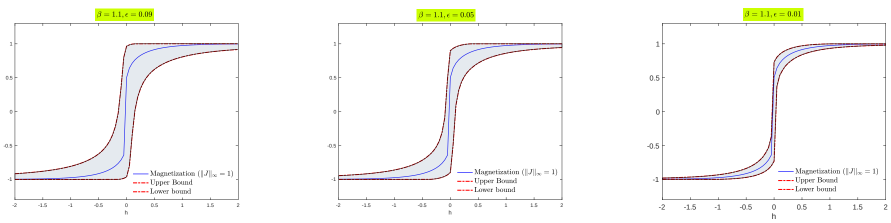

FIG. 9. *The three graphs demonstrate the uncertainty area in gray color for different values of ϵ. In all graphs, the blue solid line is the magnetization of d-sing model with Kac interaction at inverse temperature $\beta = 1.1$, $\|J\|_\infty = 1$ and $\tilde{h} = h$. The black dashed-dot curves are the upper and lower bound of magnetization of the truncated interaction $\tilde{J}^{-J}$, viewed as functions of h. (Left) $\epsilon = 0.09$. (Center) $\epsilon = 0.05$. (Right) $\epsilon = 0.01$.*

The truncated model can be viewed as Type II. However, to be consistent with the assumption $\mathcal{E} \subset \tilde{\mathcal{E}}$ in Definition 4.1, we view it as perturbed interaction of Type I arising from the subtraction of J (also explains the notation $\tilde{J}^{-J}$ in (6.22)) on regions of radius greater than $1-\epsilon$ as illustrated in Figure 7.

COROLLARY 6.7. *Let $\tilde{J}^{-J}$ be the interaction given in Definition 6.6. Then, for $0 < \epsilon < 1$ and $\gamma > 0$ small enough, the UQ bounds (6.18) and (6.19) hold for $R_{-J} = \gamma^{-1}$ and $\mathcal{F} \leq \epsilon\|J\|_\infty$. The thermodynamic pressure $P^{\Delta,\gamma}_{\mathbf{J},\beta,h}$ is given in (C.1). Let $M(\tilde{\mathbf{J}}^F, \beta, \tilde{h})$ be the LP-limit of $E_{\tilde{q}_\Delta}[m(\sigma_\Delta)]$. Then, the UQ bounds (6.20) and LP-limit of (6.18) hold with the limit of $P^{\Delta,\gamma}_{\mathbf{J},\beta,h}$ being $p_{\mathbf{J},\beta,h}$ given by (C.2).*

Remark 6.8. Given $\beta, \|J\|_\infty$, we can choose $\epsilon \equiv \epsilon(\beta, \|J\|_\infty)$ sufficiently small. Consequently, the phase diagram of the two models are close to each other as the uncertainty area is very small (Figure 9). The parameter ϵ quantifies the length of the area that one cuts off the initial interaction.

The same methods are applicable to other perturbations, e.g. the very long range in Appendix E and perturbations in "contexts"/configuration as boundary conditions.

Acknowledgments: The research of M. K. was partially supported by the NSF HDR TRIPODS CISE-1934846. The research of P. B. and M. K., was partially supported by the Air Force Office of Scientific Research (AFOSR) under the grant FA-9550-18-1-0214.

Appendix A. Reduced Markov Random Fields (rMRFs). Let $\mathbf{Y} = \{Y_i\}_{i\in\mathcal{V}}$ be a MRF indexed by a set of nodes $\mathcal{V}$ (finite or infinite) of a graph $\mathcal{G}$. Let us consider $\mathcal{M} \subset \mathcal{V}$. Let also $\mathbf{U} = \{Y_i\}_{i\in\mathcal{M}}$ and $\mathbf{u}$ be an assignment to them, namely $\mathbf{U} = \mathbf{u}$. If $\mathbf{Z} := \{Y_i\}_{i\in\mathcal{V}\setminus\mathcal{M}}$, how does the underlying graph corresponding to $\mathbf{Z} \mid \mathbf{U} = \mathbf{u}$ look like? Can the conditional probability $p(\mathbf{z} \mid \mathbf{U} = \mathbf{u})$ still keep a product structure/factorization as the joint distribution given in (2.1)? To answer the questions, we need a special class of MRF which is called *reduced Markov Random Fields* (rMRFs).

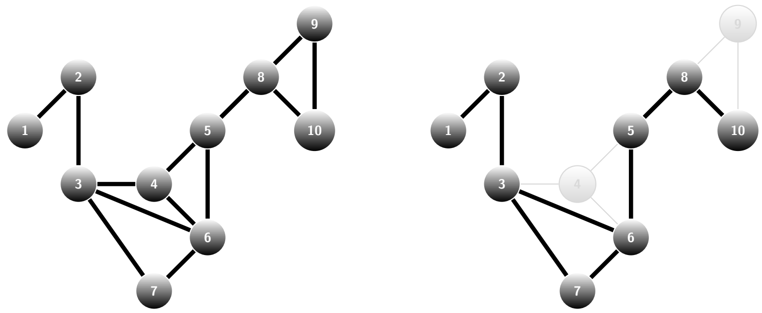

FIG. 10. *The set of nodes is* $\mathcal{V} = \{1, \cdots, 10\}$ *and* $\mathcal{M} = \{4, 9\}$. *Left:* $\mathbf{Y} = \{Y_i\}_{i=1}^{10}$ *with joint distribution* p *is a MRF over* $\mathcal{G}$. *The set of maximal cliques is given by* $\mathcal{C}_\mathcal{G} = \{\{1,2\},\{2,3\},\{3,4,6,\},\{3,6,7\},\{4,5,6\},\{5,8\},\{8,9,10\}\}$. *Right:* $\mathbf{Z} = \{Y_i\}_{i\in\mathcal{V}\setminus\mathcal{M}}$ *with joint distribution* q *is the corresponding rMRF over* $\mathcal{G}'$ *with* $\mathbf{U} = \{Y_4, Y_9\}$ *and* $\mathbf{u} = \{u_4, u_9\}$. *The rMRF is demonstrated by removing the node 4 and 9 (faded nodes) from the graph* $\mathcal{G}$. $\mathcal{C}_\mathbf{U} = \{\{3,4,6,7\},\{4,5,6\},\{8,9,10\}\}$ *while* $\mathcal{C}_\emptyset = \{\{1,2\},\{2,3\},\{5,8\}\}$.

DEFINITION A.1. *Let* $\mathbf{Y} = \{Y_i\}_{i\in\mathcal{V}}$ *be a collection of random variables indexed by a set of nodes* $\mathcal{V}$ *(finite or infinite) of a graph* $\mathcal{G}$. *If* $(\mathbf{Y}, p)$ *is a MRF,* $\mathbf{u}$ *a context,* $\mathcal{M} \subset \mathcal{V}$ *and* $\mathbf{U} = \{Y_i\}_{i\in\mathcal{M}}$, *we define as* **reduced Markov Random Field**, *a MRF* $\mathbf{Z} = \{Y_i\}_{i\in\mathcal{V}\setminus\mathcal{M}}$ *indexed by the set of nodes* $\mathcal{V}\setminus\mathcal{M}$ *of the subgraph* $\mathcal{G}[\mathcal{V}\setminus\mathcal{M}]$ *with joint distribution* $\mathbb{Q}$ *such that*

$$q(\mathbf{z}) \equiv \mathbb{Q}(\mathbf{Z} = \mathbf{z}) := p(\mathbf{z} \mid \mathbf{U} = \mathbf{u}). \tag{A.1}$$

Therefore, $\mathbf{Z} \mid \mathbf{U} = \mathbf{u}$ could be thought as a induced subgraph of $\mathcal{G}$ with set of nodes $\mathcal{V}\setminus\mathcal{M}$, that is eliminating any node corresponding to random variables $\mathbf{U}$ and any edge adjacent to them. Furthermore, according to Definition A.1, $\mathbf{Z}$ is clearly MRF and therefore the conditional probability $p(\mathbf{z} \mid \mathbf{U} = \mathbf{u})$ is expected to have a product structure. All the above are summarized in the following proposition:

PROPOSITION A.2. *Let* $\mathbf{Y}$ *be a MRF with probability distribution* $p > 0$ *parametrized by some parameters* $\mathbf{w} = \{\mathbf{w}_c\}_{c\in\mathcal{C}_\mathcal{G}}$ *given in* (2.1) *and let* $\mathbf{U}, \mathbf{Z}$ *be defined as in the beginning of the subsection. Then,* q *parametrized by* $\mathbf{w}$ *is expressed as*

$$q^{\mathbf{w}}(\mathbf{z}) \equiv p(\mathbf{z} \mid \mathbf{U} = \mathbf{u}, \mathbf{w}) = \frac{1}{Z_\mathbf{u}(\mathbf{w})} \prod_{c\in\mathcal{C}_\mathcal{G}} \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c) \tag{A.2}$$

where for every $c \in \mathcal{C}_\mathcal{G}$

$$\Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c) := \Psi_c(\mathbf{z}_c, \mathbf{u}_c \mid \mathbf{w}_c) \tag{A.3}$$

Moreover, $Z_\mathbf{u}(\mathbf{w})$ *is given by*

$$Z_\mathbf{u}(\mathbf{w}) = \sum_{\mathbf{Y}} \prod_{c\in\mathcal{C}_\mathcal{G}} \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c) \tag{A.4}$$

We refer to [47] and [53] for further discussion about MRFs, rMRFs and the proof of the Hammersley-Clifford Theorem and Propostion A.2.

A.0.1. Partition of the class of maximal cliques. We further investigate the structure of the class of all maximal cliques. Precisely, we collect $c \in \mathcal{C}_{\mathcal{G}}$ such that $\mathbf{U} \cap \mathbf{Y}_c \neq \emptyset$. This leads to partition the set of maximal cliques $\mathcal{C}_{\mathcal{G}} = \mathcal{C}_{\mathbf{U}} \sqcup \mathcal{C}_{\emptyset}$ with

$$\mathcal{C}_{\mathbf{U}} = \{c : \mathbf{U} \cap \mathbf{Y}_c \neq \emptyset\} \quad \text{and} \quad \mathcal{C}_{\emptyset} = \{c : \mathbf{U} \cap \mathbf{Y}_c = \emptyset\}. \tag{A.5}$$

(see example shown in Figure 10). On top of that, the partition of $\mathcal{C}_{\mathcal{G}}$ makes the joint distributions q take the form

$$q(\mathbf{z}) = P_{\Psi}^{\mathbf{w}}[\mathbf{u}](\mathbf{z}) = \frac{1}{Z_{\mathbf{u}}(\mathbf{w})} \prod_{c \in \mathcal{C}_{\emptyset}} \Psi_c(\mathbf{y}_c \mid \mathbf{w}_c) \prod_{c \in \mathcal{C}_{\mathbf{U}}} \Psi_c[\mathbf{u}](\mathbf{z}_c \mid \mathbf{w}_c) \tag{A.6}$$

Appendix B. Coarse-Graining, Kac and Hamiltonian Estimates.

B.1. Ising spin system with Kac-type interaction. We consider the Ising spin system in $\mathbb{Z}^d$, $d \geq 1$, with external magnetic field, $h \in \mathbb{R}$, and Kac potential given by

$$J_\gamma(x, y) = \gamma^d J(\gamma x, \gamma y), \quad x, y \in \mathbb{Z}^d \tag{B.1}$$

where γ is a positive parameter sufficiently small and J is a non-negative (ferromagnetic interaction), even, symmetric function (i.e $J(r, r') = J(r', r)$ for every $r, r' \in \mathbb{R}^d$), translational invariant (i.e $J(r, r') = J(r' + a, r + a)$ for every $r, r' \in \mathbb{R}^d$ and $a \in \mathbb{R}^d$) function such that $J(r) = 0$ for all $|r| > 1$, $\int_{\mathbb{R}^d} J(r)dr = \mathcal{J}$ and $J \in C^2(\mathbb{R}^d)$. The use of $\mathbf{J}$ stands for the collection of $J(x, y)$, that is

$$\mathbf{J} = \{J_\gamma(x, y)\}_{\mathbb{Z}^d \times \mathbb{Z}^d}. \tag{B.2}$$

A special case of Kac-type interaction is the piecewise constant:

$$J_\gamma^{\mathrm{pwc}}(x, y) = \gamma^d \mathbf{1}_{|x-y| \leq \frac{\gamma^{-1}}{2}} \tag{B.3}$$

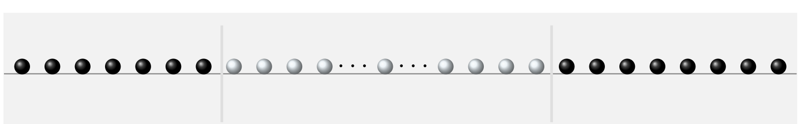

FIG. 11. *One-dimensional Ising spin lattice on Δ (white spins) with configuration boundary conditions on the complement of Δ denoted as $\bar{\sigma}_{\Delta^c}$ (black spins).*

Remark B.1. As γ becomes smaller, more particles are included in a spin neighborhood with γ^{-1} diameter and at the same time, the strength of the interactions becomes weaker.

Hamiltonian energy. For any subdomain $\Delta \subset \mathbb{Z}^d$, we define the Hamiltonian energy of a spin configuration σ_Δ, given a configuration $\bar{\sigma}_{\Delta^c}$ on its complement (see Figure 11) as

$$\begin{aligned} H_\gamma^{\mathbf{J},h}(\sigma_\Delta \mid \bar{\sigma}_{\Delta^c}) = &-\frac{1}{2} \sum_{x \neq y \in \Delta} J_\gamma(x, y) \sigma_\Delta(x) \sigma_\Delta(y) - \sum_{\substack{x \in \Delta, \\ y \in \Delta^c}} J_\gamma(x, y) \sigma_\Delta(x) \bar{\sigma}_{\Delta^c}(y) \\ &-h \sum_{x \in \Delta} \sigma_\Delta(x) \end{aligned} \tag{B.4}$$

Finite volume Gibbs measure. The Gibbs measure given a fixed boundary condition $\bar{\sigma}_{\Delta^c}$ is defined as follows:

$$\mu^{\Delta,\gamma}_{\mathbf{J},\beta,h}(\cdot \mid \bar{\sigma}_{\Delta^c}) = \frac{1}{Z_{\bar{\sigma}_{\Delta^c}}(\mathbf{J},\beta,h)} e^{-\beta H^{\mathbf{J},h}_{\gamma}(\sigma_\Delta;\bar{\sigma}_{\Delta^c})}, \tag{B.5}$$

where β is the inverse temperature and $Z_{\bar{\sigma}_{\Delta^c}}(\mathbf{J},\beta,h)$ is the normalization (partition function). To simplify the notation, we shall often drop γ and the given configuration in the complement of Δ from the Gibbs measure, resuming the full notation when needed, and therefore we write $\mu^{J,h}_{\beta,\Delta} \equiv \mu^{\bar{\sigma}_{\Delta^c},J,h}_{\beta,\Delta,\gamma}$.

B.2. Coarse-graining. We divide $\mathbb{R}^d$ into cubes of side $l = \gamma^{-1/2}$. We denote by $\mathcal{P}^{(l)}_{\mathbb{R}^d}$ the partition of $\mathbb{R}^d$. Namely, for every $i \in l\mathbb{Z}^d$ we set

$$I_{\gamma,i} = \{r \in \mathbb{R}^d : i_k \le r_k \le i_k + l, k = 1,\dots,d\} \tag{B.6}$$

(r_k and i_k being the k-th coordinate of r and i). Then we call

$$\mathcal{P}^{(l)}_{\mathbb{R}^d} = \{I_{\gamma,i} : i \in l\mathbb{Z}^d\}, \tag{B.7}$$

the collection of all the above cubes.

Definition B.2 ([55]). *(1) A function $f(r)$ is $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable, if it is constant in each cube $I_{\gamma,i}$, $i \in l\mathbb{Z}^d$.*

(2) A region $\Delta \subset \mathbb{R}^d$ is $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable, if it can be written as a union of cubes of $\mathcal{P}^{(l)}_{\mathbb{R}^d}$ (or its characteristic is $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable).

(3) Any $\Delta \subset \mathbb{Z}^d$ can be identified as a union of cubes with length 1.

(4) The size of each cube is given by

$$|I_{\gamma,i}| = |I| = l^d = \gamma^{-d/2} \tag{B.8}$$

for every $i \in l\mathbb{Z}^d$. For notational simplicity, we drop γ from $I_{\gamma,i}$.

For any bounded region Δ $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable, we denote $\mathbf{\Delta} := \Delta \cap \mathbb{Z}^d$. Hence, $\mathbf{I}_i = I_i \cap \mathbb{Z}^d$.

B.3. Coarse-grained Interaction. We introduce a new interaction $\bar{J}_\gamma$ which describes the interaction between cubes. More precisely, for every $i, j \in l\mathbb{Z}^d$ with $i \neq j$, we consider

$$\bar{J}_\gamma(i,j) = \frac{1}{|I|^2} \sum_{x\in\mathbf{I}_i} \sum_{y\in\mathbf{I}_j} J_\gamma(x,y), \tag{B.9}$$

and for $i = j$, we define

$$\bar{J}_\gamma(i,i) = \frac{1}{|I|(|I|-1)} \sum_{x\in\mathbf{I}_i} \sum_{\substack{x\in\mathbf{I}_i,\\ y\neq x}} J_\gamma(x,y) \tag{B.10}$$

LEMMA B.3. *For fixed and small $\gamma > 0$, for any $x \in \mathbf{I}_i$ and any $y \in \mathbf{I}_j$, $i, j \in l\mathbb{Z}^d$ with $i \neq j$, we have*

$$|J_\gamma(x,y) - \bar{J}_\gamma(i,j)| \leq \gamma^{d+\frac{1}{2}} \|DJ\|_\infty \mathbf{1}_{|x-y|\leq 2\gamma^{-1}}. \tag{B.11}$$

Also, for any $i \in l\mathbb{Z}^d$ and any $x, y \in \mathbf{I}_i$, we have

$$|J_\gamma(x,y) - \bar{J}_\gamma(i,i)| \leq \gamma^d \|J\|_\infty \tag{B.12}$$

Proof. Let $x \in \mathbf{I}_i$ and any $y \in \mathbf{I}_j$, $i, j \in l\mathbb{Z}^d$ with $i \neq j$, we have

$$\begin{aligned}
|J_\gamma(x,y) - \tilde{J}_\gamma(x,y)| &= |J_\gamma(x,y) - \frac{1}{|I|^2} \sum_{z\in\mathbf{I}_i} \sum_{w\in\mathbf{I}_j} J_\gamma(z,w)| \\
&\leq \frac{1}{|I|^2} \sum_{z\in I_i} \sum_{w\in I_j} |J_\gamma(x,y) - J_\gamma(z,w)| \\
&\leq \frac{1}{|I|^2} \sum_{z\in I_i} \sum_{w\in I_j} \gamma^d \|DJ\|_\infty \gamma |x-y-z+w| \mathbf{1}_{|x-y|\leq\gamma^{-1}} \\
&\leq \frac{1}{|I|^2} |I|^2 \gamma^d \|DJ\|_\infty \gamma\gamma^{-1/2} \mathbf{1}_{|x-y|\leq\gamma^{-1}} \\
&= \gamma^{d+\frac{1}{2}} \|DJ\|_\infty \mathbf{1}_{|x-y|\leq\gamma^{-1}}
\end{aligned}$$

We prove (B.12) similarly. ∎

B.4. Coarse-grained Hamiltonian Energy. In this section we analyze the Hamiltonian energy by using the new interaction defined in (B.9) and the estimates in Lemma B.3. We start by introducing some notation: for any $r \in \mathbb{R}^d$, we define the following quantity as *block spin configuration*:

$$\sigma^{(\gamma^{-1/2})}(r) := \frac{1}{|I|} \sum_{x\in\mathbf{I}_r} \sigma_{I_i}(x) \tag{B.13}$$

so that

$$\sigma^{(\gamma^{-1/2})}(r) = \frac{1}{|I|} \int_{I_r} \sigma^{(1)}(r') dr'$$

Let $\Delta \subset \mathbb{R}^d$ be $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable region. We denote by $\mathcal{M}^{(\gamma^{-1/2})}_\Delta$ all $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable functions on Δ with values in

$$M^{(\gamma^{-1/2})} := \{-1, -1 + \frac{1}{\gamma^{-d/2}}, \ldots, 1 - \frac{1}{\gamma^{-d/2}}, 1\} \tag{B.14}$$

For any bounded $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable region Δ and $m_\Delta \in \mathcal{M}^{(\gamma^{-1/2})}_\Delta$, we define as *coarse-grained Hamiltonian energy*

$$\begin{aligned}
\bar{H}^{\bar{\mathbf{J}}}_{\gamma,h}(m_\Delta; m_{\Delta^c}) := &\int_\Delta \phi_{\beta,h}(m_\Delta(r)) dr + \frac{1}{4} \int_\Delta \int_\Delta J_\gamma(r,r')[m_\Delta(r) - m_\Delta(r')]^2 dr dr' \\
&+ \frac{1}{2} \int_\Delta \int_{\Delta^c} J_\gamma(r,r')[m_\Delta(r) - m_{\Delta^c}(r')]^2 dr dr' \\
&- \frac{1}{2} \int_\Delta \int_{\Delta^c} J_\gamma(r,r') m_{\Delta^c}(r')^2 dr dr' \\
&+ \frac{1}{\beta} \int_\Delta I(m_\Delta(r)) dr
\end{aligned} \tag{B.15}$$

where

$$I(m) := -\frac{1-m}{2}\log\frac{1-m}{2} - \frac{1+m}{2}\log\frac{1+m}{2} \tag{B.16}$$

with

$$\phi_{\mathbf{J},\beta,h}(m) := \left\{-\frac{\mathcal{J}}{2}m^2 - hm\right\} - \frac{1}{\beta}I(m) \tag{B.17}$$

We recall that $\mathcal{J} = \int_{\mathbb{R}^d} J(r)dr$.

LEMMA B.4. *Let Δ be any bounded $\mathcal{P}^{(l)}_{\mathbb{R}^d}$-measurable region Δ, then there exists a constant $C > 0$ such that the following estimate holds:*

$$\left| H^{\mathbf{J}}_{\gamma,h}(\sigma_\Delta;\bar\sigma_{\Delta^c}) - \bar H^{\bar{\mathbf{J}}}_{\gamma,h}(\sigma_\Delta^{(\gamma^{-1/2})};\bar\sigma_{\Delta^c}^{(\gamma^{-1/2})})\right| \le C|\Delta|\gamma^{1/2}, \tag{B.18}$$

where $\sigma_\Delta^{(\gamma^{-1/2})}$ and $\bar\sigma_{\Delta^c}^{(\gamma^{-1/2})}$ are defined in (B.13).

Appendix C. Estimates for the thermodynamic pressure of an Ising-Kac model.

For any external field $h \in \mathbb{R}$, inverse temperature $\beta > 0$ and any bounded region $\Delta \subset \mathbb{Z}^d$ the following quantity is what is called the thermodynamic pressure:

$$P^{\Delta,\gamma}_{\mathbf{J},\beta,h}(\bar\sigma_{\Delta^c}) := \frac{\log Z_{\bar\sigma_{\mathbf{I}^c}}(\mathbf{J},\beta,h)}{\beta|\Delta|} \tag{C.1}$$

C.1. Upper and Lower Bound for the Thermodynamic Pressure,([55]). For every $\beta > 0$, $h \in \mathbb{R}$ and bounded region, $\Delta \subset \mathbb{R}^d$, we define

$$p_{\mathbf{J},\beta,h} := -\inf_{m\in[-1,1]}\{-hm + \phi_{\mathbf{J},\beta,0}(m)\} \tag{C.2}$$

If $\epsilon(\gamma) = \gamma^{1/2} + \gamma^{d/2}\log\gamma^{-1}$, then the following bounds hold: there exist constants $c, c' > 0$ such that

$$P^{\Delta,\gamma}_{\mathbf{J},\beta,h}(\bar\sigma_{\Delta^c}) \le p_{\mathbf{J},\beta,h} + \left(c\frac{\gamma^{-1}}{L} + c\epsilon(\gamma)\right), \qquad \textbf{Upper Bound} \tag{C.3}$$

Let m^* be the minimizer of $\phi_{\mathbf{J},\beta,h}$, then $p_{\mathbf{J},\beta,h} = -\phi_{\mathbf{J},\beta,h}(m^*)$, then

$$P^{\Delta,\gamma}_{\mathbf{J},\beta,h}(\bar\sigma_{\Delta^c}) \ge p_{\mathbf{J},\beta,h} - |\phi_{\mathbf{J},\beta,h}([m^*]_\gamma) - \phi_{\mathbf{J},\beta,h}(m^*)| - c\epsilon(\gamma) - c'\frac{\gamma^{-1}}{L}, \qquad \textbf{Lower Bound} \tag{C.4}$$

where $[m^*]_\gamma$ is the value in (B.14) closest to m^*.

C.2. Limit as $\Delta \nearrow \mathbb{Z}^d$ and then $\gamma \to 0$. By using the estimates for the hamiltonian energy given in (B.18), (C.3) and (C.4) we can prove that

$$\limsup_{\gamma\to 0}\lim_{\Delta\nearrow\mathbb{Z}^d} P^{\Delta,\gamma}_{\mathbf{J},\beta,h}(\bar\sigma_{\Delta^c}) \le p_{\mathbf{J},\beta,h} \tag{C.5}$$

$$\liminf_{\gamma\to 0}\lim_{\Delta\nearrow\mathbb{Z}^d} P^{\Delta,\gamma}_{\mathbf{J},\beta,h}(\bar\sigma_{\Delta^c}) \ge p_{\mathbf{J},\beta,h} \tag{C.6}$$

and therefore if $P^{\gamma}_{\mathbf{J},\beta,h} := \lim_{\Delta\to\mathbb{Z}^d} P^{\Delta,\gamma}_{\mathbf{J},\beta,h}$, then

$$\lim_{\gamma\to 0} P^{\gamma}_{\mathbf{J},\beta,h} = p_{\mathbf{J},\beta,h} = -\inf_{m\in[-1,1]}\{-hm + \phi_{\mathbf{J},\beta,0}(m)\}. \tag{C.7}$$

Hence, the thermodynamic pressure converges to the mean field pressure at the LP-limit, namely

$$\lim_{\gamma\to 0}\lim_{\Delta\nearrow\mathbb{Z}^d} P^{\Delta,\gamma}_{\mathbf{J},\beta,h} = p_{\mathbf{J},\beta,h}$$

where $p_{\mathbf{J},\beta,h}$ is defined in (C.2). The convexity properties are provided by the limit as $\Delta \nearrow \mathbb{Z}^d$ and then preserved by $\gamma \to 0$.

C.3. Thermodynamics of an Ising-spin model with a Kac potential. It is shown that when $\gamma > 0$ is sufficiently small, the phase diagram of an Ising-spin model with a Kac potential is close to the phase diagram of a mean field model. Precisely, in [11, 9] (see also [55]) it is proved that for $d \geq 2$, if $h \neq 0$ then there exists a unique DLR measure, [56]. If $h = 0$ there is a critical value of inverse temperature $\beta_c(\gamma) > 0$ such that for any $\beta < \beta_c(\gamma)$, there exists one DLR measure while for $\beta > \beta_c(\gamma)$ there are at least two distinct DLR measures $\mu^{\pm}_{\beta,\gamma}$. Finally, there is an absence of phase transition when γ is kept small (for more details see [55, 56] and references therein).

Appendix D. Proofs.

D.1. Proof of Lemma 4.2. In the following computation we use either (4.3) for type I or (4.7) for type II:

$$\begin{aligned}
\tilde{Z}_{\mathrm{u}}(\tilde{\mathbf{w}}) &= \sum_{\mathbf{z}}\prod_{\tilde{c}} \tilde{\Psi}_{\tilde{c}}[\mathbf{u}](\mathbf{z}_{\tilde{c}} \mid \tilde{\mathbf{w}}_{\tilde{c}}) \\
&= \sum_{\mathbf{z}}\prod_{c} \Psi_{c}[\mathbf{u}](\mathbf{z}_{c} \mid \mathbf{w}_{c})\Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{z}) \\
&= \sum_{\mathbf{z}}\Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{z})\prod_{c} \Psi_{c}[\mathbf{u}](\mathbf{z}_{c} \mid \mathbf{w}_{c}) \\
&= Z_{\mathrm{u}}(\mathbf{w})\sum_{\mathbf{z}}\Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{z})\prod_{c} \Psi_{c}[\mathbf{u}](\mathbf{z}_{c} \mid \mathbf{w}_{c})\frac{1}{Z_{\mathrm{u}}(\mathbf{w})} \\
&= Z_{\mathrm{u}}(\mathbf{w})E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}(\mathbf{z})]
\end{aligned}$$

D.2. Proof of Theorem 4.7. We are mostly based on the proof of the characterization of the exponential integrals (see, e.g. [20]). Let the probability measure R be defined by

$$dR/dq = e^{f(\mathbf{Z})}/E_q[f(\mathbf{Z})].$$

Note that $\mathcal{R}(\tilde{q}\|q) < \infty$, since $q, \tilde{q} > 0$. Thus,

$$-\mathcal{R}(\tilde{q}\|q) + E_{\tilde{q}}[f(\mathbf{Z})] = -\mathcal{R}(\tilde{q}\|R) + \log E_q[e^{f(\mathbf{Z})}] \leq \log E_q[e^{f(\mathbf{Z})}]. \tag{D.1}$$

where for the last inequality we use that $\mathcal{R}(\tilde{q}\|R) \geq 0$ and $\mathcal{R}(\tilde{q}\|R) = 0$ iff $\tilde{q} = R$ [20, Lemma 1.4.1]. For part a., we combine (4.13) of Lemma 4.5 and (D.1) and we get

$$E_{\tilde{q}}[f(\mathbf{Z})] \leq \log E_q[e^{f(\mathbf{Z})}] + \frac{1}{E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]}E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\log\Phi^{\mathrm{i}}_{\mathbf{u}}\right] - \log E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]$$

By replacing $f(\mathbf{Z})$ to $\pm\lambda f(\mathbf{Z})$, we obtain

$$\pm E_{\tilde{q}}[f(\mathbf{Z})] \leq \frac{1}{\lambda}\Big\{\log E_q[e^{\pm\lambda f(\mathbf{Z})}] + \frac{1}{E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]}E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}}\log\Phi^{\mathrm{i}}_{\mathbf{u}}\right] - \log E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]\Big\}$$

By optimizing over $\lambda > 0$ (see [13] and [51]), the following tight estimates are obtained:

$$\pm E_{\tilde{q}}[f(\mathbf{Z})] \leq \inf_{\lambda>0} \frac{1}{\lambda}\Big\{ \log E_q[e^{\pm\lambda f(\mathbf{Z})}] + \frac{1}{E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]} E_q\left[\Phi^{\mathrm{i}}_{\mathbf{u}} \log \Phi^{\mathrm{i}}_{\mathbf{u}}\right] - \log E_q[\Phi^{\mathrm{i}}_{\mathbf{u}}]\Big\}$$

Part b. is proved similarly, utilizing (4.14) instead of (4.13).

Example D.1. *(Single-parameter exponential families) This is a straightforward example and a simple illustration of the ideas in the proof of part b., Theorem 4.7, giving us insights on how well the ideas work together with a rearranging argument. The simplicity of this example arises from the fact that the exponential family is single parametric and therefore the structural part is not present. The probability density function of a random variable X with range $R(X)$, is given by*

$$p^{\theta}(x) = P^{\theta}(X = x) = e^{\theta\phi(x) - F(\theta)}$$

taken with respect to some measure $d\nu$ where $F(\theta) = \log \int_x e^{\theta\phi(x)}\nu(dx)$ and $\phi(x)$ is a real-valued function also known as sufficient statistic. Suppose a second probability density function of the same single-parameter exponential family associated with ϕ

$$p^{\theta+\zeta}(x) = P^{\theta+\zeta}(X = x) = e^{(\theta+\zeta)\phi(x) - F(\theta+\zeta)}$$

for some $\zeta < 1$. One may want to investigate how sensitive the model is in such a change in θ by ζ with respect to $\phi(X)$ as means to bound $E_{P^{\theta+\zeta}}[\phi(X)]$ or to find the error in replacing the first distribution by the "perturbed" one and phrased as bound $E_{P^{\theta+\zeta}}[\phi(X)] - E_{P^{\theta}}[\phi(X)]$. The second exponential family is apparently a perturbation on parameters by ζ, so we can think of the model as Type I. In addition, after employing UQ bounds, the cumulant generating function and KL divergence are the two main ingredients to compute: for any $\lambda > 0$,

$$\Lambda^{\phi}_{P^{\theta}}(\lambda) = \log E_{P^{\theta}}[e^{\lambda\phi(X)}] = F(\theta+\lambda) - F(\theta)$$
$$R(P^{\theta+\zeta}\|P^{\theta}) = \zeta E_{P^{\theta+\zeta}}[\phi(X)] - \log E_{P^{\theta}}[e^{\zeta\phi(X)}]$$

The above expression for KL divergence comes from the calculation of expressing $F(\theta+\lambda)$ in terms of $F(\theta)$ and for that every term is computed with respect to P^{θ}. By substituting the quantities to the UQ bounds and by doing a delicate rearrangement of terms that is feasible because the QoI is a sufficient statistic for the model, we get

$$\pm E_{P^{\theta+\zeta}}[\phi(X)] \leq \frac{1}{1-\zeta} \inf_{\lambda>0} \left\{ \frac{F(\theta+\lambda) - F(\theta)}{\lambda} + \frac{1}{\lambda} \log E_{P^{\theta}}[e^{\zeta\phi(X)}] \right\}$$

D.3. Proof of Theorem 4.8. The existence and the explicit form of the distribution $q^{\pm}$ relies on [37], Theorem 2. Consequently, given a QoI f, we identify the total $\tilde{q}$-excess factor relative to q explicitly, that is $\Phi^{\pm}_{\mathbf{u}} = e^{\lambda_{\pm} f}$. However, the new element is that by utilizing the Hammersley-Clifford Theorem, $q^{\pm}$ defined on $\mathbf{Z}$ are rMRFs, lie in the class $\mathcal{Q}^{\eta}_{\mathcal{P}}$ and the total $\tilde{q}$-excess factor relative to q is explicitly determined.

D.4. Proof of Example 5.1. We prove (5.4): Let A be an event of interest. We define the QoI $g(\mathbf{Y}) = \mathbf{1}_A$ and we compute all the quantities involved in (4.15) explicitly. Let us start with the cumulant generating function:

$$\Lambda^{f}_{p}(\lambda) = \log E_p[e^{\lambda g}] = \log\left(\sum_{\mathbf{y}\in A} e^{\lambda g} p(\mathbf{y}) + \sum_{\mathbf{y}\notin A} e^{\lambda g} p(\mathbf{y})\right)$$
$$= \log\left(e^{\lambda} p(A) + 1 - p(A)\right)$$

We now go through the computation of $E_p[\Phi^{\mathrm{I}}]$:

$$\begin{aligned} E_p[\Phi^{\mathrm{I}}] &= \sum_{\mathbf{y}} \Phi^{\mathrm{I}}(\mathbf{y})p(\mathbf{y}) = \sum_{\mathbf{y}} e^{aw_c f_c(\mathbf{y}_c)} p(\mathbf{y}_c) \\ &= \sum_{\mathbf{y}\in B_c} e^{aw_c f_c(\mathbf{y}_c)} p(\mathbf{y}) + \sum_{\mathbf{y}\notin B_c} e^{aw_c f_c(\mathbf{y}_c)} p(\mathbf{y}) \\ &= e^{aw_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}}. \end{aligned}$$

Similarly, we prove that

$$E_p[\Phi^{\mathrm{i}} \log \Phi^{\mathrm{i}}] = aw_c e^{aw_c} p_{\mathrm{I}} \tag{D.2}$$

Overall, by recalling (4.13) the KL divergence equals to

$$R(\tilde{p}\|p) = \frac{aw_c e^{aw_c} p_{\mathrm{I}}}{e^{aw_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}}} - \log\left(e^{aw_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}}\right)$$

D.5. Proof of Example 5.2. The proof goes similarly as the proof of Example 5.1. The cumulant generating function is the same as in Example 5.1. Let us compute the expected value of the total $\tilde{p}$-excess factor of type II relative to p with respect to p:

$$\begin{aligned} E_p[\Phi^{\mathrm{II}}] &= \sum_{\mathbf{y}} \Phi^{\mathrm{II}}(\mathbf{y})p(\mathbf{y}) = \sum_{\mathbf{y}} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) \\ &= \sum_{\mathbf{y}\in B_c} e^{aw_c f_c(\mathbf{y}_c)} p(\mathbf{y}) + \sum_{\mathbf{y}\in B_{\tilde{c}}} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) + \sum_{\mathbf{y}\notin B_c \cup B_{\tilde{c}}} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) \\ &= e^{(1+a)w_c} p_{\mathrm{II}} + e^{-w_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}} - p_{\mathrm{II}}. \end{aligned} \tag{D.3}$$

We split the sum into the three sums since $B_c \cap B_{\tilde{c}} = \emptyset$. Similarly, we prove that

$$E_p[\Phi^{\mathrm{II}} \log \Phi^{\mathrm{II}}] = -w_c e^{-w_c} p_{\mathrm{I}} + (1+a)w_c e^{(1+a)w_c} p_{\mathrm{II}} \tag{D.4}$$

Overall, by recalling (4.13) the KL divergence equals to

$$R(\tilde{p}\|p) = \frac{-w_c e^{-w_c} p_{\mathrm{I}} + (1+a)w_c e^{(1+a)w_c} p_{\mathrm{II}}}{e^{(1+a)w_c} p_{\mathrm{II}} + e^{-w_c} p_{\mathrm{I}} + 1 - p_{\mathrm{I}} - p_{\mathrm{II}}} - \log\left(-w_c e^{-w_c} p_{\mathrm{I}} + (1+a)w_c e^{(1+a)w_c} p_{\mathrm{II}}\right)$$

Remark D.2. If $B_c \cap B_{\tilde{c}} \neq \emptyset$, then we need to split the sum of (D.3) as follows: Let $U \equiv B_c \cap B_{\tilde{c}}$, then

$$\begin{aligned} E_p[\Phi^{\mathrm{II}}] &= \sum_{\mathbf{y}} \Phi^{\mathrm{II}}(\mathbf{y})p(\mathbf{y}) = \sum_{\mathbf{y}} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) \\ &= \sum_{\mathbf{y}\in B_c\setminus U} e^{aw_c f_c(\mathbf{y}_c)} p(\mathbf{y}) + \sum_{\mathbf{y}\in B_{\tilde{c}}\setminus U} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) + \sum_{\mathbf{y}\in U} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) \\ &\quad + \sum_{\mathbf{y}\notin B_c \cup B_{\tilde{c}}} e^{(1+a)w_c \tilde{f}_{\tilde{c}} - w_c f_c} p(\mathbf{y}) \\ &= e^{(1+a)w_c}(p_{\mathrm{II}} - p(U)) + e^{-w_c}(p_{\mathrm{I}} - p(U)) + e^{aw_c} p(U) + 1 - p_{\mathrm{I}} - p_{\mathrm{II}} + p(U). \end{aligned}$$

Note that $p_{\mathrm{I}}, p_{\mathrm{II}}$ and $p(U)$ are computable as p is known.

D.6. Important Lemma for the proof of the UQ bounds (6.19).

LEMMA D.3. *Let L and $\partial\Delta$ be the side and the boundary of the cube Δ respectively with $L >> R_F$. Then, for any interaction $\mathbf{F} = \{F(x,y) : x, y \in \mathbb{Z}^d\}$ satisfying (6.1)-(6.3) and range R_F, the following holds:*

(i) If the support of F is given by (6.9), then

$$\sum_{x\in\Delta} \sum_{\substack{y\in\Delta^c\\ y\in B^{\neq}_{x,R_F}}} F(x,y) \leq R_F|\partial\Delta| \sum_{x\neq 0} |F(0,x)|.$$

(ii) If the support of F is given by (6.10), then

$$\sum_{x\in\Delta} \sum_{y\in\Delta^c} F(x,y) \leq R_F|\Delta| \sum_{x\neq 0} |F(0,x)|$$

Proof. The bounds are straightforward once we split the sum as follows:

$$\sum_{x\in\Delta} \sum_{\substack{y\in\Delta^c\\ y\in B^{\neq}_{x,R_F}}} F(x,y) = \sum_{\substack{x\in\Delta\\ dist(x,\Delta^c)\leq R_F}} \sum_{\substack{y\in\Delta^c\\ y\in B^{\neq}_{x,R_F}}} F(x,y) + \sum_{\substack{x\in\Delta\\ dist(x,\Delta^c)> R_F}} \sum_{y\in\Delta^c} F(x,y) \leq R_F|\Delta|.$$

where

$$dist(x,\Delta^c) = \inf\{\|x-y\| : y \in \Delta^c\}$$

∎

Remark D.4. When $L << R_F$, both (i) and (ii) are bounded by $R_F|\Delta| \sum_{x\neq 0} |F(0,x)|$.

D.7. Proof of Lemma 6.1. It is not difficult to show (see also Proposition II.1.2 and Lemma II.2.2C in [60]) that

$$\begin{aligned} |\log Z_{\bar\sigma_{\Delta^c}}(\mathbf{J},\beta,h) - \log Z_{\bar\sigma_{\Delta^c}}(\tilde{\mathbf{J}}^{\mathbf{F}},\beta,h)| &\leq \beta\|H^{\mathbf{J},h}(\sigma_\Delta|\bar\sigma_{\Delta^c}) - H^{\tilde{\mathbf{J}}^{\mathbf{F}},h}(\sigma_\Delta|\bar\sigma_{\Delta^c})\|_\infty \\ &\leq |\Delta|\|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar\sigma_{\Delta^c}} - \Phi^{h,\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar\sigma_{\Delta^c}}\|_1 \end{aligned} \tag{D.5}$$

which in turn gives

$$R(\tilde q_\Delta\|q_\Delta) \leq 2|\Delta|\|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar\sigma_{\Delta^c}} - \Phi^{h,\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar\sigma_{\Delta^c}}\|_1 \tag{D.6}$$

since

$$\begin{aligned} R(\tilde q_\Delta\|q_\Delta) = \beta\Big(&E_{\tilde q_\Delta}[H^{\mathbf{J},h}(\sigma_\Delta|\bar\sigma_{\Delta^c})] - E_{q_\Delta}[H^{\tilde{\mathbf{J}}^{\mathbf{F}},\tilde h}(\sigma_\Delta|\bar\sigma_{\Delta^c})]\Big) \\ &+ \log Z_{\bar\sigma_{\Delta^c}}(\mathbf{J},\beta,h) - \log Z_{\bar\sigma_{\Delta^c}}(\tilde{\mathbf{J}}^{\mathbf{F}},\beta,\tilde h) \end{aligned}$$

A straightforward bound yields that

$$\|\Phi^{h,\beta,\mathbf{J}}_{\Delta,\bar\sigma_{\Delta^c}} - \Phi^{h,\beta,\tilde{\mathbf{J}}^{\mathbf{F}}}_{\Delta,\bar\sigma_{\Delta^c}}\|_1 \leq \beta\left(|\tilde h - h| + \sum_{x\neq 0}|F(0,x)|\right).$$

D.8. Proof of Lemma 6.2.2. It is a straightforward computation after subtracting the hamiltonian energies with interaction J and

$$\tilde J^F(x,y) = J(x,y)\mathbf{1}_{\|x-y\|_d\leq R} + F(x,y)\mathbf{1}_{\|x-y\|_d\leq R}, \quad \text{Type I},$$

and

$$\tilde J^F(x,y) = J(x,y)\mathbf{1}_{\|x-y\|_d\leq R} + F(x,y)\mathbf{1}_{\|x-y\|_d> R}, \quad \text{Type II}$$

D.9. Proof of (6.18)**.** For Type I, it suffices to bound $\kappa_{\mathrm{I}}(\sigma_\Delta)$. In fact, we exploit Lemma D.3 and we get

$$0 \le \kappa_{\mathrm{I}}(\sigma_\Delta) \le \beta|\Delta|\left(\frac{1}{2} + 2R\frac{|\partial\Delta|}{|\Delta|}\right)\sum_{x\neq 0}|F(x,y)| \tag{D.7}$$

D.9.1. Cumulant generating function for $f(\mathbf{Z}) = |\Delta| m(\sigma_\Delta)$**.**

$$\begin{aligned}\Lambda_{q_\Delta;|\Delta|m(\sigma_\Delta)}(\pm\lambda) &= \log E_{q_\Delta}[e^{\lambda|\Delta|\frac{1}{|\Delta|}\sum_{x\in\mathbf{\Delta}}\sigma_\Delta(x)}]\\ &= \log\left(\frac{1}{Z_{\bar\sigma_{\Delta^c}}(\mathbf{J},\beta,h)}\sum_{\sigma_\Delta} e^{\lambda\sum_{x\in\mathbf{\Delta}}\sigma_\Delta(x)}e^{-\beta H^{\mathbf{J},h}(\sigma_\Delta|\sigma_{\Delta^c})}\right)\\ &= \log\left(e^{\lambda\sum_{x\in\mathbf{\Delta}}\sigma_\Delta(x)-\beta H^{\mathbf{J},h}(\sigma_\Delta|\sigma_{\Delta^c})}\right) - \log Z_{\bar\sigma_{\Delta^c}}(\mathbf{J},\beta,h)\\ &:= \log Z_{\bar\sigma_{\mathbf{\Delta}^c}}(\mathbf{J},\beta,h\pm\frac{\lambda}{\beta}) - \log Z_{\bar\sigma_{\Delta^c}}(\mathbf{J},\beta,h)\end{aligned} \tag{D.8}$$

Then by using the definition of the thermodynamic pressure in (C.1), we get:

$$\frac{1}{|\Delta|}\Lambda_{q_\Delta;|\Delta|m(\sigma_\Delta)}(\pm\lambda) = \beta\left(P^{\Delta,\gamma}_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}} - P^{\Delta,\gamma}_{h,\beta,\mathbf{J}}\right) \tag{D.9}$$

Appendix E. Phase diagram of a long range perturbation.

E.1. Thermodynamics of a long range perturbation of 1-dimensional Kac model. There is a significant number of works in the literature studying the phase diagram of one-dimensional ferromagnetic Ising model with long range interactions of the form $1/r^k$ with k indicating the decay of interaction and $k \le 2$. For $k < 2$, the occurrence of phase transition has been proved (see [24, 25, 26]). For $k = 2$, the existence of a spontaneous magnetization at low temperature is proved in [31]. The establishment of the existence of phase transition, proving the discontinuity of the magnetization at a critical point, also known as *Thouless effect*, was proved by Aizenman et al in [1]. In [10], the authors study the phase diagram of the system with interaction defined in (E.1) with F given in Definition E.1 as illustrated in the right graph of Figure 7. Precisely, they have shown that there is a critical value of the inverse temperature depending on a and γ sufficiently small such that the system exhibits phase transition.

E.1.1. Phase diagram of a long range perturbation. We consider a one dimensional ferromagnetic Ising spin system with interactions that correspond to a $1/r^2$ long range perturbation of the usual Kac model, see the right picture of Figure 7.

Definition E.1. *Let* J^{pwc}_γ *be as in* (B.3)*, then we define*

$$\tilde J^F_\gamma(x,y) = \begin{cases} J^{\mathrm{pwc}}_\gamma & ,0 \le |x-y| \le (2\gamma)^{-1}\\ F(x,y) & ,\ |x-y| > (2\gamma)^{-1},\end{cases} \tag{E.1}$$

with $F(x,y) = \frac{a}{|x-y|^2}$ *for some number* $a \in (0,\infty)$*, Figure 7 (right).*

The range of the perturbation F is clearly Type II. We derive the UQ bounds as follows:

$$\begin{aligned}\log\Phi^{\mathrm{i}}_{\bar\sigma_{\Delta^c}}(\sigma_\Delta) = \beta\sum_{x\in\Delta}\sigma_\Delta(x)\Big(&\tilde h - h + \frac{1}{2}\sum_{y\in A^{\mathrm{II}}_x\cap\Delta}F(x,y)\sigma_\Delta(y)\\ &+ \sum_{y\in A^{\mathrm{II}}_x\cap\Delta^c}F(x,y)\bar\sigma_{\Delta^c}(y)\Big)\end{aligned} \tag{E.2}$$

then $C^{\mathrm{II}} := \beta(\tilde{h} - h)$ and

$$\kappa_{\mathrm{II}} := \beta \sum_{x\in\Delta} \sigma_\Delta(x)\Big(\frac{1}{2}\sum_{y\in A_x^{\mathrm{II}}\cap\Delta} F(x,y)\sigma_\Delta(y) + \sum_{y\in A_x^{\mathrm{II}}\cap\Delta^c} F(x,y)\bar{\sigma}_{\Delta^c}(y)\Big)$$

We bound κ_{II} based on the following:

$$\begin{aligned}\sum_{x\in\Delta}\sum_{y\in A_x^{\mathrm{II}}\cap\Delta} F(x,y) &\leq |\Delta| \sum_{y\in A_x^{\mathrm{II}}} F(0,y) = |\Delta| \sum_{y\in A_0^{\mathrm{II}}} \frac{a}{y^2} \\ &= \gamma|\Delta| \sum_{y\in A_0^{\mathrm{II}}} \frac{\gamma a}{(\gamma y)^2} \leq C\gamma|\Delta| \end{aligned} \tag{E.3}$$

for some constant C arises from $\sum_{y\in A_0^{\mathrm{II}}} \frac{a}{y^2} < \infty$. Then $\kappa_{\mathrm{II}} \leq 2C\gamma|\Delta|$ and the UQ bounds for long range perturbation with $\beta(\tilde{h} - h) < 1$ are

$$\pm E_{\tilde{q}_\Delta}[m(\sigma_\Delta)] \leq \frac{1}{1-\beta(\tilde{h}-h)} \inf_{\lambda>0}\left\{\frac{P^{\Delta,\gamma}_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}} - P^{\Delta,\gamma}_{h,\beta,\mathbf{J}}}{\lambda/\beta} + \frac{\beta}{\lambda}2C\gamma\right\} \tag{E.4}$$

In the LP-limit we get

$$\pm M(\tilde{\mathbf{J}}^F, \beta, \tilde{h}) \leq \frac{1}{1-\beta(\tilde{h}-h)} \inf_{\lambda>0}\left\{\frac{p_{h\pm\frac{\lambda}{\beta},\beta,\mathbf{J}} - p_{h,\beta,\mathbf{J}}}{\lambda/\beta}\right\} \tag{E.5}$$

REFERENCES

[1] M. Aizenman, J. Chayes, L. Chayes, C. Newman, *Discontinuity of the magnetization in one dimensional $\frac{1}{|x-y|^2}$ percolation, Ising and Potts models*, J. Stat. Phys., 50 (1988), pp. 1–40.

[2] R. Atar, K. Chowdhary, P. Dupuis, *Robust bounds on risk-sensitive functionals via Rényi divergence*, SIAM/ASA J. Uncertain. Quantif., 3 (1) (2015), pp. 18-33.

[3] U. Basu, M. Krüger, A. Lazarescu, C. Maes, *Frenetic aspects of second order response*, Phys. Chem. Chem. Phys. 17 (9) (2015), pp. 6653–6666.

[4] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Courier Corporation, 2007.

[5] J. Birrell and L. Rey-Bellet, *Uncertainty quantification for Markov processes via variational principles and functional inequalities*, arXiv:1812.05174 (2018).

[6] J. Birrell and L. Rey-Bellet *Concentration Inequalities and Performance Guarantees for Hypocoercive MCMC Samplers*, arXiv:1907.11973 (2019).

[7] J. Birrell, P. Dupuis, M. A. Katsoulakis, L. Rey-Bellet, J. Wang, *Distributional Robustness and Uncertainty Quantification for Rare Events*, arXiv:1911.09580 (2019).

[8] A. Bovier, *Statistical Mechanics, Extreme Values, and Disordered Systems*, University Lecture Notes.

[9] A. Bovier, M. Zahradník, *The low-temperature phase of Kac-Ising models*, J. Stat. Phys., 87 (1997), pp. 311–332 .

[10] M. Cassandro, I. Merola, M. E. Vares, *Study of a Long Range Perturbation of a One-Dimensional Kac Model*, J. of Stat Phys., 142 (2011), pp. 487–509.

[11] M. Cassandro, E. Presutti, *Phase transitions in Ising systems with long but finite range interactions*. Markov Process. Related Fields, 2 (1996), pp. 241–262.

[12] H. Chan and A. Darwiche. *Sensitivity analysis in Markov networks*. In International Joint Conference on Artificial Intelligence, (2005).

[13] K. Chowdhary and P. Dupuis, *Distinguishing and integrating aleatoric and epistemic variation inuncertainty quantification*, ESAIM Math. Model. Numer. Anal., 47 (2013), pp. 635–662.

[14] Darwiche A. *Modeling and reasoning with Bayesian networks*. New York: Cambridge University Press ELT, 2009.

[15] J. de Bock, A. Antonucci, and C. P. de Campos. *Global sensitivity analysis for MAP inference in graphical models.* In Neural Information Processing Systems, 27 (2014), pp. 2690-2698.

[16] S. Della Pietra, V. Della Pietra, J. Lafferty. *Inducing features of random fields.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (1997), pp. :380–392,

[17] G. Diezemann, *Nonlinear response theory for Markov processes: simple models for glassy relaxation*, Phys. Rev. E, 85 (2012), pp. 051502.

[18] P. Domingos, Statistical Relational Learning, *Tutorial in ICML*, Oregon State University-Corvallis, OR, USA, https://icml.cc/Conferences/2007/tutorials.html, (2007).

[19] P. Domingos, D. Lowd, *Markov Logic: An Interface Layer for Artificial Intelligence*, Morgan & Claypool (2009).

[20] P. Dupuis, R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley Series in Probability and Statistics, 1997.

[21] P. Dupuis, M. A. Katsoulakis, Y. Pantazis, P. Plecháč, *Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics.* SIAM/ASA J. Uncertain. Quantif., 4(1) (2016), pp. 80–111.

[22] P. Dupuis, M. A. Katsoulakis, Y. Pantazis, L. Rey-Bellet, *Sensitivity Analysis for Rare Events based on Rényi Divergence*, to appear Ann. Appl. Probab.

[23] P. Dupuis, Y. Mao, *Formulation and properties of a divergence used to compare probability measures without absolute continuity*, arXiv:1911.07422 (2019).

[24] F.J. Dyson, *Existence of phase transition in a one-dimensional Ising ferromagnetic*, Comm. Math. Phys., 12 (1969), pp. 91–107.

[25] F.J. Dyson, *Non-existence of spontaneous magnetization in a one-dimensional Ising ferromagnet*, Math. Phys., 12 (1969), pp. 212–215.

[26] F.J. Dyson, *An Ising ferromagnet with discontinuous long-range order*, Comm. Math. Phys., 21 (1971), pp. 269–283.

[27] M. S. Eldred, B. M. Adams, D. M. Gay, L. P. Swiler, K. Haskell, W. J. Bohnhoff, J. P. Eddy, W. E. Hart, J. paul Watson, P. D. Hough, and T. G. Kolda. Dakota, *A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis*: Version 5.0 user's manual, sandia. Technical report, Sandia, (2009).

[28] J. Feng, J. L. Lansford, M. A. Katsoulakis, D. G. Vlachos, *Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences*, Sci. Adv., 6 (2020), eabc3204.

[29] S. Friedli, Y. Velenik, *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*, Cambridge University Press, 2017.

[30] N. Friedman, I. Nachman, D. Peer, *Learning bayesian network structure from massive datasets: The sparse candidate algorithm*, UAI' 99: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, July 1999, pp. 206–215.

[31] J. Fröhlich, T. Spencer, *The phase transition in the one-dimensional Ising model with $\frac{1}{r^2}$ interaction energy.* Commun. Math. Phys., 84 (1982), pp. 87–101.

[32] G. Gallavotti, *Statistical Mechanics, a Short Treatise*, Text and Monographs in Physics, Springer, Berlin, 1999.

[33] L. El Ghaoui, M. Oks, F. Oustry, *Worst-case value-at-risk and robust portfolio optimization: A conic programming approach.* Oper. Res., 51(4) (2003,), pp. 543–556.

[34] Z. Ghahramani, *Probabilistic Machine Learning and Artificial Intelligence*, Nature, 521(7553) 2015, pp. 452-459.

[35] P. Glasserman, X. Xu, 2014. *Robust Risk Measurement and Model Risk*, Quant. Finance 14, 1 (2014), pp. 29–58.

[36] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT press, 2016.

[37] K. Gourgoulias, M. A. Katsoulakis, L. Rey-Bellet, J. Wang, *How Biased Is Your Model? Concentration Inequalities, Information and Model Bias*, IEEE Trans. Inform. Theory, 66 (2020), pp. 3079-3097.

[38] G. R. Grimmett, *Probability on graphs: random processes on graphs and lattices*, Institute of Mathematical Statistics Textbooks 1, Cambridge University Press, 2010.

[39] E. J. Hall, S. Taverniers, M. A. Katsoulakis, D. M. Tartakovsky, *GINNs: Graph-Informed Neural Networks for Multiscale Physics*, arXiv:2006.14807 (2020).

[40] J. M. Hammersley, P. Clifford, *Markov fields on finite graphs and lattices*, 1971.

[41] AJ. Hartemink, DK. Gifford, TS. Jaakkola, RA. Young, *Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks*, Pacific Symposium on Biocomputing, Hawaii, 2001.

[42] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining,*

Inference, and Prediction, Springer Series in Statistics. Springer Verlag, 2001.

[43] X. He, R. S. Zemel, M. A. Carreira-Perpiñìan, *Multiscale conditional random fields for image labeling*. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Washington, DC, USA, 2004, pp. II-II, doi: 10.1109/CVPR.2004.1315232.

[44] M. A. Katsoulakis, L. Rey-Bellet, J. Wang, *Scalable information inequalities for uncertainty quantification*, J. Comp. Phys., 336 (2017), pp. 513–545.

[45] Diederik P. Kingma, M. Welling, *Auto-Encoding Variational Bayes*. In The 2nd International Conference on Learning Representations (ICLR), 2013.

[46] C. Kipnis, C. Landim, *Scaling limits of interacting particle systems*, Springer-Verlag Vol. 320, 1999.

[47] D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.

[48] R. G. Krishnan, U. Shalit, D. Sontag, *Structured Inference Networks for Nonlinear State Space Models*, arXiv:1609.09869, 2016.

[49] L. Landau, E. Lifshitz, *Perspectives in Theoretical Physics* (ed L. P.Pitaevski) pp. 287–297, Pergamon, 1992

[50] S. Lauritzen, *Graphical Models*. Oxford University Press, 1996. ISBN: 0-19-852219-3.

[51] J. Li and D. Xiu, *Computation of failure probability subject to epistemic uncertainty*, SIAM J. Sci. Comput., 34 (2012), pp. A2946–A2964.

[52] J. Moussouris, *Gibbs and Markov Random Systems with Constraints*, J. Stat. Phys., 10 (1974), pp. 11–33. issn: 0022-4715. doi: 10.1007/BF01011714.

[53] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[54] Fuchun Pengand, A. McCallum, *Accurate Information Extraction from Research Papers using Conditional Random Fields*, HLT-NAACL , 2004.

[55] E. Presutti, *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics*, Springer, 2000.

[56] E. Presutti, *From equilibrium to nonequilibrium statistical mechanics. Phase transitions and the Fourier law*, Braz. J. Probab. Stat., 29, Number 2 (2015), pp. 211-281.

[57] D. Ruelle, *A review of linear response theory for general differentiable dynamical systems*. Nonlinearity, 22 (2009), pp. 855–870, doi:10.1088/0951-7715/22/4/009.

[58] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.

[59] K. Sato, Y. Sakakibara, *RNA secondary structural alignment with conditional random fields*, Bioinformatics, 2005 Sep 1;21 Suppl 2:ii237–42. doi: 10.1093/bioinformatics/bti1139. PMID: 16204111.

[60] B. Simon, *The Statistical Mechanics of Lattice Gases*, Vol. 1, Princeton University Press, 2014.

[61] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM Computational Science & Engineering Series: Philadelphia, PA, USA, 2014, pp. 382.

[62] B. Taskar, P. Abbeel, D. Koller, *Discriminative probabilistic models for relational data*, In Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI02), (2002), pp. 485–494, Edmonton, Canada.

[63] S. Taverniers, F. J. Alexander, D. M. Tartakovsky, *Noise propagation in hybrid models of nonlinear systems: The Ginzburg-Landau equation*. J. Comput.Phys., 262 (2014), pp. 313–324, doi: https://doi.org/10.1016/j.jcp.2014.01.015.

[64] S. Taverniers, E. J. Hall, M. A. Katsoulakis, D. M. Tartakovsky, *Mutual Information for Explainable Deep Learning of Multiscale Systems*, arXiv:2009.04570, 2020.

[65] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Science & Business, Media, 2008.

[66] M. Tuckerman, *Statistical mechanics: theory and molecular simulation*, Oxford university press, 2010.

[67] K. Um, E. J. Hall, M. A. Katsoulakis, D. M. Tartakovsky, *Causality and Bayesian Network PDEs for multiscale representations of porous media*, J. Comput. Phys., 394 (2019), pp. 658–678.

[68] K. Um, E. J. Hall, M. A. Katsoulakis, D. M. Tartakovsky, *Causality and Bayesian Network PDEs for multiscale representations of porous media*, J. Comput. Phys., 394 (2019), pp. 658–678.

[69] M. Wainwright, M. Jordan, *Graphical models, exponential families, and variational inference*. Technical Report 649, Department of Statistics, University of California, Berkeley (2003)