Application of Max Ent and Network analysis

Nicola Dainese, Stefano Mancone, Francesco Vidaich

I. INTRODUCTION

In this project we are going build three different models based on Maximum Entropy (Max Ent) principle in order to study the dataset of Barro Colorado Forest Census Plot Data. In Max Ent models one have to decide what are the relevant observable variables that a model should be capable of explain and use their experimental averages as constraints that enable us to determine analytically or computationally the coupled Lagrange parameters. In order to compute averages on a single dataset we divided the original area of $1000 \times 500\ m^2$ in $N = 200$ subplots of $50 \times 50\ m^2$ and considered them as independent configurations (or replicas) of the same system. Thus all the constraints we are going to use will be evaluated on this ensemble of subplots.

II. HAMILTONIAN OF A MAX ENT MODEL FOR GENERIC CONSTRAINTS

In this section we derive the general form of an Hamiltonian given some constraint functions f_r (this is just another way of thinking about the observable variables). We use the index $a = 1, \dots, n$ to characterize each possible configuration x of the system (and not just the N observed) and refer to the probability of a given configuration as $p_a = p(x_a)$. We start considering a generic entropy

$$S[\{p_a\}_{a=1,\dots,n}] = -K \sum_{a=1}^{n} p_a ln(p_a) \tag{1}$$

with K > 0 and following constraints

$$C_0 = (\sum_{a=1}^{n} p_a) - 1 = 0 \tag{2}$$

$$C_r = (\sum_{a=1}^{n} p_a f_r(x_a)) - \langle f_r(x) \rangle_{obs} = 0 \tag{3}$$

with $r = 1, \dots, m$. We maximize the sum of these three equations imposing

$$\frac{\partial (S + \sum_{r=0}^{m} \lambda_r C_r)}{\partial p_a} = 0 \ \forall a \tag{4}$$

We get

$$p_a(\vec{\lambda}) = e^{-1+\frac{\lambda_0}{K}} \exp\{\sum_{r=1}^{m} \frac{\lambda_r}{K} f_r(x_a)\} \tag{5}$$

Imposing normalization (Eq. 2) we get

$$Z(\{\lambda_r\}_{r=1,\dots,m}) = e^{1-\frac{\lambda_0}{K}} = \sum_{a=1}^{n} e^{\sum_{r=1}^{m} \frac{\lambda_r}{K} f_r(x_a)} \tag{6}$$

Finally we can interpret the probabilities p_a of each configuration x_a as Boltzmann weights

$$\begin{aligned} p_a(\vec{\lambda}) &= \frac{1}{Z(\vec{\lambda})} \exp\{\sum_{r=1}^{m} \frac{\lambda_r}{K} f_r(x_a)\} \\ &= \frac{e^{-H(x_a;\vec{\lambda})}}{Z(\vec{\lambda})} \end{aligned} \tag{7}$$

where the Hamiltonian of the system is

$$H(x, \vec{\lambda}) = -\sum_{r=1}^{m} \frac{\lambda_r}{K} f_r(x_a) \tag{8}$$

III. MAX ENT 1

The first model that we are going to use aims to model the presence or absence of the species in a generic subplot of a given area of the Barro Colorado Forest. To do that, we map the presence and absence of a species i to a spin σ_i, that assumes respectively $+1$ and -1 in the two cases. This enable us to work in a familiar framework, that is the one of the Ising model.

A. Analytical derivation of the Lagrange multipliers

Adopting the notation to our specific model, we set $K = 1$, and change the symbols for the configurations, the number of constraints and the constraints functions:
$x_a \to \vec{\sigma}^{(a)},\ m \to S,\ f_r(x) \to \pi_i(\vec{\sigma}) = \sigma_i$
Hence our Hamiltonian is

$$H(\vec{\sigma}) = -\sum_{i=1}^{S} \lambda_i \sigma_i \tag{9}$$

Manipulating the partition function

$$\begin{aligned} Z(\vec{\lambda}) &= \sum_{\{\vec{\sigma}\}} \exp\{\sum_{i=1}^{S} \lambda_i \sigma_i\} = \sum_{\{\vec{\sigma}\}} \prod_{i=1}^{S} \exp\{\lambda_i \sigma_i\} \\ &= \prod_{i=1}^{S} \sum_{\sigma_i = \pm 1} \exp\{\lambda_i \sigma_i\} = 2^S \prod_{i=1}^{S} \cosh(\lambda_i) \end{aligned}$$

(10)

Hence we can compute analytically the expected value for each variable σ_i for a given value of the Lagrange parameters

$$\begin{aligned}\langle\sigma_i\rangle_{model(\vec{\lambda})} &= \sum_{\{\vec{\sigma}\}} \sigma_i P(\vec{\sigma}/\vec{\lambda}) \\ &= \frac{\sum_{\sigma_i \pm 1} \sigma_i e^{\lambda_i \sigma_i}}{2cosh(\lambda_i)} \\ &= tanh(\lambda_i)\end{aligned} \tag{11}$$

Now we impose $\forall i$

$$\langle\sigma_i\rangle_{model(\vec{\lambda})} = m_i = tanh(\lambda_i) \tag{12}$$

and inverting the system we find

$$\lambda_i = tanh^{-1}(m_i) \tag{13}$$

Thus we can compute the Lagrange parameters for our Max Ent 1 model, starting from the average presence p_i of each species, mapping it in an average "magnetization" $m_i = 2p_i - 1$ and finally using Eq. 13.

IV. MAX ENT 2

In this section we add to the Max Ent 1 model another constraint, that is the average squared difference between the number of present (S_+) and absent (S_-) species in a single subplot. This leads to a new Hamiltonian

$$H(\vec{\sigma}, \vec{\lambda}, K) = -\frac{K}{S}\sum_{i,j}\sigma_i\sigma_j - \sum_{i=1}^{S}\lambda_i\sigma_i \tag{14}$$

The constraints that we are going to use are:

1. $m_i = \langle\sigma_i\rangle_{model} = C_i(\vec{\sigma})$ with coupled parameters $\lambda_i, i = 1, \ldots, S$
2. $\frac{\langle(S_+ - S_-)^2\rangle_{obs}}{S} = \frac{\langle(\sum_{j=1}^{S}\sigma_j)^2\rangle_{model}}{S} = C_0(\vec{\sigma})$ with coupled parameter $\lambda_0 = K$

A. Estimate of Lagrange parameters with Metropolis and gradient descent

To estimate computationally the parameters for this model the idea is the following:

1. initialize randomly from a normal distribution $\vec{\lambda}^{(0)} = (K^{(0)}, \lambda_1^{(0)}, \ldots, \lambda_S^{(0)})$;
2. use Metropolis algorithm to generate a sample of N configurations $\{\vec{\sigma}^{(a)}\}_{a=1,\ldots,N}$ according to the Hamiltonian $H^{(0)}(\vec{\sigma})$ of the system with the initial parameters;
3. compute the constraint functions $C_i(\vec{\sigma})$, for $i = 0, \ldots, S$, for each of the N configurations and then average the result;
4. compare $\langle C_i^{(0)}\rangle_{model}$ with $\langle C_i\rangle_{obs}$;
5. use a steepest descent algorithm[1] to update the Lagrange parameters according to: $\lambda_a^{(1)} \leftarrow \lambda_a^{(0)} - \eta(\langle C_i^{(0)}\rangle_{model} - \langle C_i\rangle_{obs})$;
6. iterate steps 2-5 until convergence.

Since our goal is to have a model which generates configurations that on average are able to replicate all the observed constraints, we decided to use as a loss function (only for the purpose of visualization) the sum of the squared differences between each constraint evaluated with the model and with the data, i.e. $L(K, \vec{\lambda}) = \sum_{a=0}^{S}(\langle C_a\rangle_{model} - \langle C_a\rangle_{obs})^2$. In the figure below are reported the time evolution of the loss function and of the Lagrange parameters. We can see that after 100 steps both the loss and the parameters reach stable values.

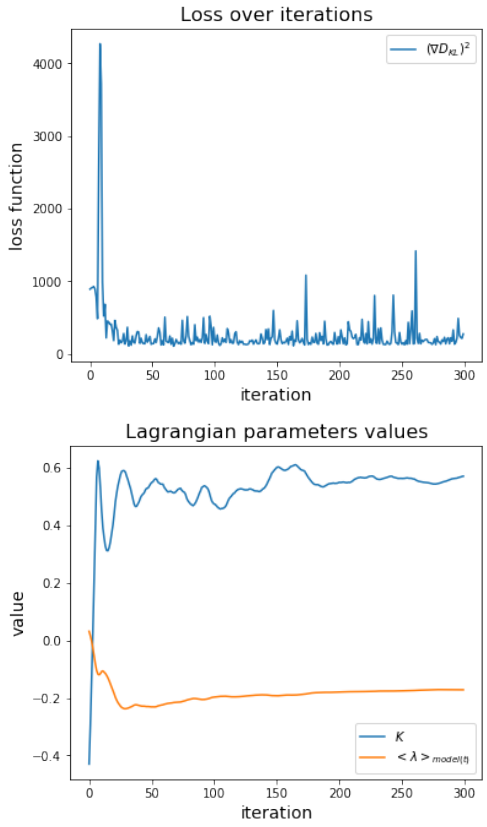

FIG. 1. In the top panel is represented the loss function of our algorithm. In the bottom panel, we plotted the temporal evolution of K and of the mean value of the λ at each iteration. We use the mean in order to compress the information about 299 parameters in a single plot.

B. Theoretical treatment of Random Field Ising Model

In this section we draw a sketch of how one could obtain a theoretical estimate of the Lagrange parameter K of the Hamiltonian of Max Ent 2 (Eq. 14).

This is not a rigorous derivation, because there are a series of assumptions and approximations of which we didn't check the validity. First of all we assume of being in a perturbative regime with respect to the case of Max Ent 1 model, so that we can assume that the λ_i are given by Eq. 13.

Furthermore, we assume that all the λ_i are drawn from the same gaussian distribution $p(h)$ of mean zero and variance $\sigma^2 = Var(\{\lambda_i\})$, i.e.

$$p(h) = \frac{e^{\frac{-h^2}{\sigma^2}}}{\sqrt{2\pi}\sigma} \tag{15}$$

Following the derivation of Schneider and Pytte[2] we use the replica-trick to compute the quenched free-energy $\langle F \rangle_h$ of the system starting from the average over the disorder h of the partition function $\langle Z^n \rangle_h$ of n replicas.

Thus, after taking the limit $n \to 0$, we arrive at the following expression of the quenched free-energy

$$\langle F \rangle_h = S\{Km^2 - \int_{-\infty}^{+\infty} dh \cdot p(h) ln[2cosh(2Km+h]\} \tag{16}$$

and the self-consistency equation for the magnetization, that was obtained from a saddle-point approximation

$$m = m_{SC}(K) = \int_{-\infty}^{+\infty} dh \cdot p(h) tanh(2Km + h) \tag{17}$$

Now we would like to impose the constraint associated to K, that is

$$\frac{\partial ln Z}{\partial K} = \frac{\langle (S_+ - S_-)^2 \rangle}{S} \tag{18}$$

But we can approximate $lnZ \simeq \langle lnZ \rangle_h = -\langle F \rangle_h$ and take the derivative w.r.t. the last term

$$\begin{aligned} -\frac{\partial \langle F \rangle_h}{\partial K} &= \frac{\langle (S_+ - S_-)^2 \rangle}{S} \\ &= -S[m^2 - 2m \cdot m_{SC}(K)] - \\ &- 2KS \frac{\partial m}{\partial K}[m - m_{SC}(K)] \\ &= Sm^2 \end{aligned} \tag{19}$$

where in the last passage we used the self-consistency equation for m (Eq. 17). Thus we can express m as a function of the constraint

$$m_\pm = \pm \sqrt{\left\langle \left(\frac{S_+ - S_-}{S} \right)^2 \right\rangle} \equiv \pm c \tag{20}$$

Substituting this solution in the self-consistency equation for m we get a condition that K must satisfy

$$c = \int_{-\infty}^{+\infty} dh \cdot \frac{e^{\frac{-h^2}{\sigma^2}}}{\sqrt{2\pi}\sigma} tanh(2Kc + h) \tag{21}$$

Although there is not an explicit solution for this equation, we know both c and σ from the data (we can use for σ the variance of the theoretical $\{\lambda_i\}$, regularizing them or removing the divergent ones), thus we can obtain K numerically using Eq. 21. We can see in Fig. 2 that K_{th} is the only value for which the curve $c(K)$ intersects the bisector at the value of c_{obs}, that is our constraint.

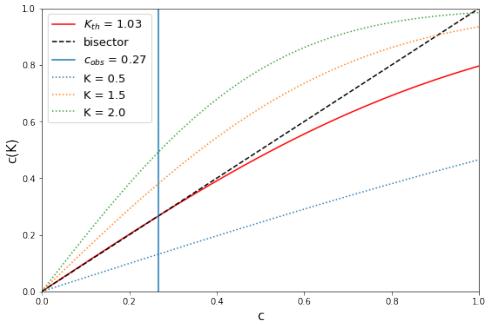

FIG. 2. In this graphic we reported in the x axis c, considered as an independent variable, whereas in the y axis are reported the values $c(K)$ of the self-consistency equation 21, that are parametric curves in K. The vertical line represents the value of the constraint $c_{obs} = 0.27$ and the black dashed line (the bisector) is the site where all consistent solutions of Eq. 21 lie. The curve that has as parameter the theoretical estimate $K_{th} = 1.03$ intersects both the vertical line and the bisector.

Finally, in Fig. 3 we compare this results with the ones obtained via simulation in the last section.

C. Phase diagram of the Hamiltonian

In this section we refer to an approximation of Random Field Ising model in which each λ_i is thought as if it was drawn from a gaussian distribution with variance $\sigma = \sqrt{\frac{\sum_{i=1}^{S} (\lambda_i - \langle \lambda \rangle)^2}{S-1}}$.

We consider the variables K, σ and $\beta = \frac{1}{T}$ and define from them $\hat{K} = \frac{K}{\sigma}$, $\hat{\lambda} = \beta\lambda$ and $\hat{\beta} = \sigma\beta$. It is possible[3] to show that for each $\hat{\beta}$ there is a $\hat{K}$ for which the Hamiltonian of section IV represents a spin system

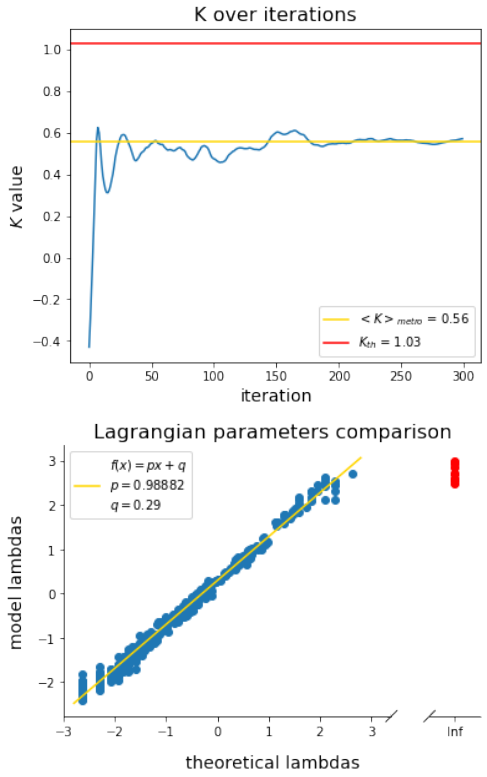

FIG. 3. In the top panel is plotted the temporal evolution of the Lagrange parameter K along with the mean of the last 100 iterations (golden line) and the theoretical expectation. The two predictions are not compatible, but have the same order of magnitude and the same sign, thus in both cases we expect only paramagnetic or ferromagnetic phase (but not antiferromagnetic for example). In the bottom panel we report a scatterplot of the theoretical λ versus the one estimated via simulation. The red points on the top right represent the Lagrange parameters associated with those species which have an observed magnetization $m_i = 1$; as a consequence their theoretical prediction is $\lambda_i = \infty$ and obviously we can not think of approximate them correctly via simulation. The coefficient of the best fit is $p = 0.989 \pm 0.006$, that is slightly less than 1, indicating that the theoretical estimates are in general greater than the computational ones.

in phase transition between the paramagnetic and the ferromagnetic phase.

The equation of the curve of phase transition is

$$2\hat{K} = \left(\int_{-\infty}^{+\infty} \frac{d\hat{\lambda}}{\sqrt{2\pi}} \frac{e^{(-\hat{\lambda}^2/2\hat{\beta}^2)}}{cosh(\hat{\lambda})^2} \right)^{-1} \tag{22}$$

We plot in Fig. 4 the curve $\hat{K}(\hat{\beta}^{-1}) = \hat{K}(T/\sigma)$ and confront it with the parameters estimated computationally (labelled as "model") and theoretically. We can see that the Hamiltonian of the model is in the ferromagnetic phase, whereas the theoretical Hamiltonian lies exactly along the curve of phase transition.

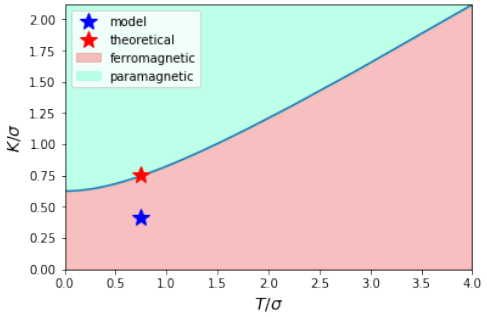

FIG. 4. In the figure is reported the curve of the phase transition between ferromagnetic and paramagnetic phase for a Random Field Ising model. The blue star represents the position of the Hamiltonian parameters estimated in Sec. IV A, that is located in the ferromagnetic region. Instead the red star represents our theoretical estimate K_{th}, that very interestingly it is located exactly on the critical line between the two phases.

V. MAX ENT 3

In this section we take into account both the abundances and the two-points correlations of the observed data to construct a third Max Ent model.

A. Analytical derivation of the Lagrange parameters

The constraints that we use to built Max Ent 3 model are:

1. $\langle x_i \rangle_{obs} = \langle x_i \rangle_{model}$ coupled to λ_i (different from those of the previous sections);
2. $\langle x_i x_j \rangle_{obs} = \langle x_i x_j \rangle_{model}$ coupled to M_{ij}.

Maximizing the entropy of the system together with the constraints we get the following Hamiltonian

$$H(\vec{x}) = \sum_{i=1}^{S} \lambda_i x_i + \frac{1}{2} \sum_{i,j=1}^{S} M_{ij} x_i x_j \tag{23}$$

Because to perform analytic calculations we have to assume that the abundances x are distributed as random variables with domain in all $\mathbb{R}$, to make a better approximation we select only those species that satisfy $x_i - \sigma_{x_i} > 0$; in this way we obtain $S' < S$ species. The procedure now is the standard one that we have seen in section III A. First we compute the partition function performing a gaussian integral. This holds:

$$Z = C \cdot \exp \frac{1}{2} \vec{\lambda}^T M^{-1} \vec{\lambda} \tag{24}$$

Then we compute the expected value of each abundance x_i according to the model and we impose it equal to the

observed average:

$$\begin{aligned}\langle x_i\rangle_{model} &= -\frac{\partial lnZ}{\partial \lambda_i} \\ &= -\sum_{j=1}^{S'}(M^{-1})_{ij}\lambda_j \\ &= \langle x_i\rangle_{obs}\end{aligned} \tag{25}$$

Now, instead of imposing the observed correlation, we use a little trick to impose the observed covariance of the abundances $Cov(x_i, x_j)$:

$$\begin{aligned}\frac{\partial\langle x_i\rangle_{model}}{\partial\lambda_j} &= -(M^{-1})_{ij} \\ &= -\frac{\partial^2 lnZ}{\partial\lambda_j\partial\lambda_i} \\ &= -\langle x_i x_j\rangle_{model} + \langle x_i\rangle_{model}\langle x_j\rangle_{model} \\ &= -Cov(x_i, x_j)_{obs}\end{aligned} \tag{26}$$

Solving the system of Eq. 25 and Eq. 26 we get

$$\begin{aligned}(M^{-1}) &= Cov(x_i, x_j) \\ \lambda_i &= -\sum_{j=1}^{S'} M_{ij}\langle x_j\rangle_{obs}\end{aligned} \tag{27}$$

In Fig. 5 can be seen the distribution of the entries of the interaction matrix for the species that satisfy $x_i - \sigma_{x_i} > 0$.

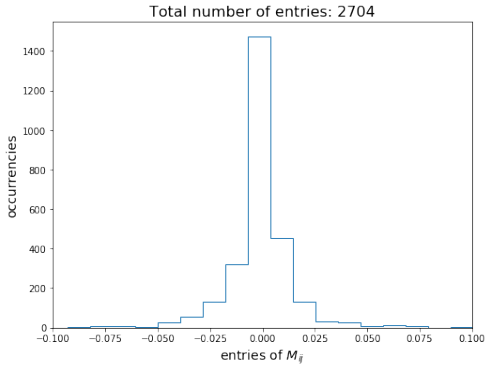

FIG. 5. Histogram of the entries of the interaction matrix M_{ij} for the $S' = 52$ species (52^2 entries). The distribution is centered in zero and more or less symmetric, meaning that both cooperative and competitive interactions are present in the network.

B. Analysis of the interaction network

In this last section we analyze the network given by the adjacency matrix A_{ij} associated to the weighted interaction matrix M_{ij}, where the entries smaller in absolute value of a threshold θ are set to zero. The mapping between the two is straightforward:

$$A_{ij}(\theta) = \Theta(|M_{ij}| - \theta) \tag{28}$$

where Θ is the Heaviside theta. We define θ^* as the greatest threshold for which the adjacency matrix still represents a network of a single connected component. To compute the number of connected components of the system for a given threshold θ we use the method of the laplacian eigenvalues, i.e. the number of connected components is given by the number of zero eigenvalues of the laplacian matrix $L = K - A$ (K is the diagonal matrix containing the degree of each node, i.e. $K_{ij} = k_i\delta_{ij}$). The results are presented in Fig. 6.

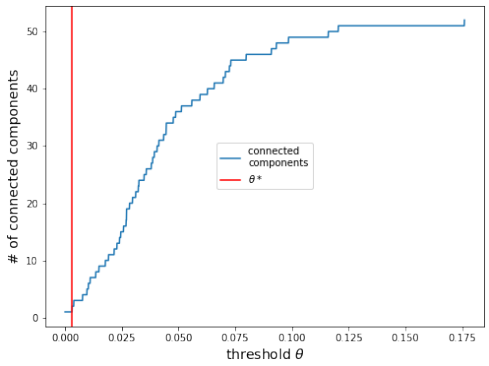

FIG. 6. In this graphic we show the number of connected components of the network as a function of the threshold θ that we put on the entries of the interaction matrix.

We then performed a deeper analysis for the network given by the adjacency matrix $A(\theta^*)$, computing the mean degree $\langle k\rangle$, the standard deviation of the degree σ_k, the diameter of the network D, the global clustering coefficient C, the network assortativity and the maximum betweenness centrality of the network.

To understand the meaning of those quantities we built a benchmark for comparing them, composed by an ensemble of 100 Erdosh-Reny graphs. Each graph is characterized by the same number of nodes of the network and by a probability of connecting two nodes given by $p = \frac{\langle k\rangle}{n-1}$, that is chosen to met the average degree observed in our network $A(\theta^*)$. We consider an ensemble of E-R graphs in order to compute the estimated quantities and not the ones of a single realization. We summarize all the results obtained in the following table:

Quantity	Real graph	ER graph
degree standard deviation	10.28	3.38
diameter	3	2.00
clustering coefficient	0.73 8	0.59
assortativity	-0.24	-0.04
max betweenness centrality	0.05	0.01

[1] Actually we use ADAM as optimizer instead of the plain gradient descent because it is more reliable and stable.

[2] T.Schneider, E.Pytte, Random-field instability of ferromagnetic state, Phys. Rev., 1977.

[3] We refer to https://inordinatum.wordpress.com/2013/01/20/mean-field-solution-of-the-random-field-ising-model/.