

An introduction to the maximum entropy approach and its application to inference problems in biology

Original

An introduction to the maximum entropy approach and its application to inference problems in biology / De Martino, A., De Martino, D.. - In: HELIYON. - ISSN 2405-8440. - 4:4(2018), p. e00596. [10.1016/j.heliyon.2018.e00596]

Availability:

This version is available at: 11583/2976719 since: 2023-03-10T08:38:19Z

Publisher:

Elsevier

Published

DOI:10.1016/j.heliyon.2018.e00596

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Received:
31 January 2018
Revised:
31 March 2018
Accepted:
3 April 2018

Cite as: Andrea De Martino,
Daniele De Martino. An
introduction to the maximum
entropy approach and its
application to inference
problems in biology.
Heliyon 4 (2018) e00596.
doi: [10.1016/j.heliyon.2018.e00596](https://doi.org/10.1016/j.heliyon.2018.e00596)



Review Article

An introduction to the maximum entropy approach and its application to inference problems in biology

Andrea De Martino^{a,b,*}, Daniele De Martino^c

^a Soft & Living Matter Lab, Institute of Nanotechnology (NANOTEC), Consiglio Nazionale delle Ricerche, Rome, Italy

^b Italian Institute for Genomic Medicine (IIGM), Turin, Italy

^c Institute of Science and Technology Austria, Klosterneuburg, Austria

* Corresponding author. Current address: Statistical Inference and Computational Biology Unit, Italian Institute for Genomic Medicine (IIGM), via Nizza 52, 10126 Turin, Italy.
E-mail address: andrea.demartino@roma1.infn.it (A. De Martino).

Abstract

A cornerstone of statistical inference, the maximum entropy framework is being increasingly applied to construct descriptive and predictive models of biological systems, especially complex biological networks, from large experimental data sets. Both its broad applicability and the success it obtained in different contexts hinge upon its conceptual simplicity and mathematical soundness. Here we try to concisely review the basic elements of the maximum entropy principle, starting from the notion of ‘entropy’, and describe its usefulness for the analysis of biological systems. As examples, we focus specifically on the problem of reconstructing gene interaction networks from expression data and on recent work attempting to expand our system-level understanding of bacterial metabolism. Finally, we highlight some extensions and potential limitations of the maximum entropy approach, and point to more recent developments that are likely to play a key role in the upcoming challenges

of extracting structures and information from increasingly rich, high-throughput biological data.

Keywords: Systems biology, Molecular biology, Mathematical bioscience, Computational biology, Bioinformatics

1. Introduction

It is not unfair to say that the major drivers of biological discovery are currently found in increasingly accurate experimental techniques, now allowing to effectively probe systems over scales ranging from the intracellular environment to single cells to multi-cellular populations, and in increasingly efficient bioinformatic tools, by which intracellular components and their putative interactions can be mapped at genome and metabolome resolution. Yet, at least to some degree, these approaches still appear hard to integrate into quantitative predictive models of cellular behaviour. In a sense this is not surprising. Even if we possessed detailed information about all sub-cellular parts and processes (including intracellular machines, their interaction partners, regulatory pathways, mechanisms controlling the exchange with the medium, etc.), it would be hard to build a comprehensive mechanistic model of a cell, and possibly even harder to infer deep organization principles from it. In large part, this is due to the fact that cells have an enormous number of degrees of freedom (e.g. protein levels, RNA levels, metabolite levels, reaction fluxes, etc.) which, collectively, can take on an intimidatingly large number of physico-chemically viable states. On the other hand, experiments necessarily probe only a tiny portion of these states. Therefore, understanding how all internal variables might coordinate so that certain “macroscopic” quantities, like the cell’s growth rate, behave as observed in experiments is quite possibly a hopeless task. In addition, these models would most likely require some tuning of the multitude of parameters that characterize intracellular affairs, rendering overfitting a very concrete prospect. At the same time, though, the deluge of data coming from both sides (experiments and bioinformatics) begs for the development of bridges connecting them, not just as descriptive frameworks and predictive tools but also as guides for novel targeted experiments and bioengineering applications.

The problem appears to be that of finding a reasonable ‘middle-ground’ between full fledged mechanistic approaches and qualitative phenomenological descriptions based on coarse-grained quantities only. Perhaps following the lesson of thermodynamics (which, one might argue, has faced a similar question of bridging microscopic and macroscopic descriptions of physical systems), an increasing number of studies is undertaking a route different from – and in many ways inverse of – the mechanistic one. The key issues to be faced along such route are the following: To what degree do experimental results constrain the space of allowed

states of a living system? Can we learn something about internal variables from experiments that probe a relatively small number of states? Is there a way to perform reliable statistical inference on the values of un-observed internal variables from empirical data? Ultimately, all of these questions boil down to the problem of inferring probability distributions (or, in other terms, statistical models) from limited data. This age-old challenge dating back to the origin of probability theory (see e.g. Laplace's 'principle of indifference', Ref. [1], Ch. 2) has found a self-consistent answer, the only such answer under certain conditions, in the so-called *principle of maximum entropy*.

Over the past decade, entropy maximization or closely related ideas have been repeatedly employed for the analysis of large-scale biological data sets in contexts ranging from the determination of macromolecular structures and interactions [2,3,4,5,6,7,8,9,10,11,12,13,14] to the inference of regulatory [15,16,17,18,19] and signaling networks [20,21,22,23] and of the organization of coding in neural populations [24,25,26,27,28,29,30,31,32,33,34,35,36,37]; from the analysis of DNA sequences (e.g. for the identification of specific binding sites) [38,39,40] to the study of the HIV fitness landscape [41,42,43]; from the onset of collective behaviour in large animal groups [44,45,46] to the emergence of ecological relationships [47,48,49,50,51,52,53,54,55,56,57]. The type of insight derived from these models is remarkably diverse, from fundamental organization principles in structured populations to specific gene-gene interaction networks. Revealingly, the challenge of dealing with high dimensional and limited data posed by biology has in turn stimulated the search for novel efficient implementations of the maximum entropy principle at the interface between computational biology, statistical physics and information theory, leading to an impressive improvement of inference schemes and algorithms. Future ramifications of these studies are likely to explore new application areas, as more/better data become available, theoretical predictions get sharper, and computational methods improve. In many ways, the maximum entropy approach now appears to be the most promising provider of 'middle grounds' where empirical findings and bioinformatic knowledge can be effectively bridged.

Several excellent reviews, even very recent ones, cover the more technical aspects of maximum entropy inference from the viewpoint of statistical physics, computational biology or information theory (see e.g. [58,59,60,61,62]). Our goal here is to provide a compact, elementary and self-consistent introduction to entropy maximization and its usefulness for inferring models from large-scale biological data sets, starting from the very basics (i.e. from the notion of 'entropy') and ending with a recent application (a maximum entropy view of cellular metabolism). We mainly hope to convey its broad applicability and potential to deliver new biological insight, and to stimulate further cross-talk while keeping mathematics to a minimum. A few basic mathematical details are nevertheless given in the Supplementary Material for sakes

of completeness. Given the vastness of the subject, we shall take a shortcut through most of the subtleties that have accompanied the growth of the field since the 1940s, focusing instead on the aspects that (we believe) are of greater immediate relevance to our purposes. A partial and biased list of additional ingredients and new directions will be presented in the Discussion. The interested reader will however find more details and food for thought in the suggested literature.

For our purposes, the path leading to entropy maximization can start from the intuitively obvious idea that, when extracting a statistical model from data, one should avoid introducing biases other than those that are already present in the data, as they would be unwarranted and discretionary. For instance, if we had to model a process with E possible outcomes (like the throw of a dice) and had no prior knowledge of it, our best guess for a probability law underlying this process would have to be the uniform distribution, where each outcome occurs with probability $1/E$. In essence, the framework of entropy maximization generalizes this intuition to more complex situations and provides a recipe to construct the ‘optimal’ (i.e. least biased) probability distribution compatible with a given set of data-derived constraints. Central to it is, of course, the concept of ‘entropy’.

2. Main text

2.1. Entropy and entropy maximization: a bird’s eye view

The notion of ‘entropy’ as originated in thermodynamics is usually associated to that of ‘disorder’ by saying that the former can be regarded as a measure of the latter. The word ‘disorder’ here essentially means ‘randomness’, ‘absence of patterns’, or something similar. While not incorrect, these words clearly require a more precise specification to be useful at a quantitative level. As we shall see, once the stage is characterized more clearly, the entropy of a system (e.g. of a population of cells) with prescribed values for certain observables (e.g. the rate of growth of the population) quantifies the number of distinct arrangements of its basic degrees of freedom (e.g. the protein levels, RNA levels, metabolic reaction rates, etc. of each cell) that lead to the same values for the constrained observables. The larger this number, the larger the entropy. In this sense, the entropy of a system is really a measure of the microscopic multiplicity (the ‘degeneracy’) underlying its macroscopically observable state. What makes entropy a powerful inference tool is closely connected to this characterization.

To make things more precise, one can consider a classical, highly stylized example. Imagine having N identical balls distributed in K urns so that n_i balls are placed in the i -th urn, with $i = 1, \dots, K$ and $n_1 + \dots + n_K \equiv \sum_{i=1}^K n_i = N$. (To fix ideas, one can think of the N balls as the N different cells and of the K urns as K distinct

configurations for the basic degrees of freedom of each cell. An arrangement $\{n_i\}$ defined by specific values of n_1, \dots, n_K therefore represents how cells are distributed over the allowed internal states, with n_1 cells in state 1, n_2 in state 2, and so on.) By simple combinatorics (see [63], Ch. 3), the number of ways in which the N balls can be placed in the K urns leading to the same values of n_1, \dots, n_K is given by the multinomial coefficient

$$\Omega = \frac{N!}{n_1!n_2! \cdots n_K!} . \tag{1}$$

This number¹ describes the ‘microscopic’ degeneracy underlying the specific arrangement of balls described by the numbers $\{n_1, \dots, n_K\}$ and grows fast with N . Actually, for sufficiently large N , (1) turns out to be well approximated (see Supplementary Material, Sec. S1) by

$$\Omega \simeq e^{NH} , \tag{2}$$

where

$$H = - \sum_{i=1}^K \frac{n_i}{N} \ln \frac{n_i}{N} . \tag{3}$$

The quantity H defined in (3) is the *entropy* of the arrangement described by the urn occupation numbers n_1, \dots, n_K . Note (see Supplementary Material, Sec. S2) that $H \geq 0$. In a nutshell, Equations (2) and (3) say that, for large N , some arrangements $\{n_1, \dots, n_K\}$ can be realized in a huge number of ways (as a matter of fact, in a number of microscopic ways that is exponentially large in N), and that the entropy H ultimately quantifies this number. On the other hand, some specific arrangements can have a very small degeneracy. For instance, the arrangement with N balls in urn 1 and no balls elsewhere (i.e. with $n_1 = N$ and $n_i = 0$ for $i \neq 1$) can be realized in a unique way, having $H = 0$ and $\Omega = 1$.²

Note that the quantity $n_i/N \equiv p_i$ represents the fraction of balls appearing in urn i in arrangement $\{n_1, \dots, n_K\}$ or, equivalently for our purposes, the probability that a ball selected at random and uniformly comes from the i -th urn. Therefore H is a function of the probabilities $\{p_i\}$ ($i = 1, \dots, K$), i.e.

$$H \equiv H[\{p_i\}] = - \sum_{i=1}^K p_i \ln p_i , \tag{4}$$

¹ To give an idea, for $N = 10$ and $K = 3$, the arrangement with $n_1 = 5, n_2 = 3$ and $n_3 = 2$ has $\Omega = 2520$. If $K = 4$, the arrangement with $n_1 = 5, n_2 = 3$ and $n_3 = n_4 = 1$ has instead $\Omega = 5040$ (twice as large). If $N = 100$ and $K = 3$, the arrangement with $n_1 = 50, n_2 = 30$ and $n_3 = 20$ has instead $\Omega \simeq 10^{42}$.

² Notice however that many arrangements have $H = 0$. In particular, any arrangement with N balls in one urn and none in the others has $H = 0$. Hence there are K ways to realize an $H = 0$ arrangement, corresponding to the fact that there are K possible choices for the urn in which to place the N balls.

and the condition $\sum_{i=1}^K n_i = N$ simply corresponds to the fact that probabilities should sum to one, i.e. $\sum_{i=1}^K p_i = 1$.

It is rather intuitive that, if balls were ‘thrown’ into urns randomly (i.e. so that each of the N balls has equal probability of ending up in any of the K urns), the resulting arrangement $\{n_1, \dots, n_K\}$ would much more likely be one with large Ω (and entropy) than one with small Ω (and entropy). In particular, the most likely arrangement $\{n_1^*, \dots, n_K^*\}$ (or, equivalently, the most likely distribution of probabilities $\{p_1^*, \dots, p_K^*\}$) should coincide with that carrying the largest degeneracy, or maximum entropy, satisfying the constraint $\sum_{i=1}^K n_i^* = N$ (or, equivalently, $\sum_{i=1}^K p_i^* = 1$). In this sense, the safest bet on the outcome of an experiment in which N balls are randomly assigned to K urns would be to place money on the maximum entropy (MaxEnt) distribution.

This is the gist of the maximum entropy principle: if one is to infer a probability distribution given certain constraints, out of all distributions $\{p_i\}$ compatible with them, one should pick the distribution $\{p_i^*\}$ having the largest value of (4). The only constraint considered in the above example of balls and urns is the normalization of probabilities, i.e. the fractions p_i should sum to one: $\sum_{i=1}^K p_i = 1$. In this case, the MaxEnt distribution is uniform, namely $p_i^* = 1/K$ for each $i = 1, \dots, K$ (see Supplementary Material, Sec. S3). However, constraints can involve other quantities, leading to different MaxEnt distributions (see Supplementary Material, Sec. S4 for a few simple examples). This just reflects the diverse information that constraints inject into the inference problem in each case.

It is important to understand that, because they correspond to maximal underlying degeneracy, MaxEnt distributions are the least biased given the constraints: any other distribution compatible with the same constraints would have smaller degeneracy and therefore would artificially exclude some viable (i.e. constraint-satisfying) configurations of the underlying variables. In other terms, a MaxEnt distribution is completely undetermined by features that do not appear explicitly in the constraints subject to which it has been computed.

These ideas, which ultimately make the maximum entropy principle the central conceptual tool for inferring probability distributions subject to constraints, have been placed on firmer and firmer mathematical ground starting from the 1940s. In our view, three classical results are especially noteworthy in the present context.

Firstly, landmark work by Shannon [64] and Khinchin [65] formally characterized H , Equation (4), as the *only* function complying with a set of *a priori* requirements (known as Shannon–Khinchin axioms) to be satisfied by a measure of the ‘uncertainty’ or ‘lack of information’ associated to a probability distribution $\{p_i\}$. Here, ‘uncertainty’ relates in essence to how (im)precisely one can identify the configuration of basic degrees of freedom from knowledge of the distribution $\{p_i\}$.

In this sense, a larger uncertainty corresponds to a larger underlying degeneracy Ω and hence to a larger entropy H . Therefore, MaxEnt distributions compatible with given constraints formally correspond to those that maximize the uncertainty on every feature except for those that are directly encoded in the constraints.

A strict characterization of MaxEnt distributions is instead encoded in a result known as the ‘entropy concentration theorem’ [61]. In short, and referring to the urn-and-balls example discussed above, it rigorously quantifies the observation that, when N is sufficiently large, the number of microscopic states underlying the MaxEnt distribution is exponentially larger (in N) than the number of microscopic states underlying any other distribution. This is also seen from (2), albeit at a heuristic level. Denoting respectively by Ω^* and H^* the degeneracy and the entropy of the MaxEnt distribution, one sees that, for sufficiently large N ,

$$\frac{\Omega^*}{\Omega} \simeq e^{N(H^* - H)} \quad (5)$$

for any Ω and H corresponding to a distribution different from the MaxEnt one. Because H^* is the maximum value attained by the entropy, $H^* - H > 0$. Hence, (5) states that microscopic arrangements underlying the MaxEnt distribution are more numerous than those underlying any other distribution by an exponentially large (in N) factor. In turn, for large enough N , observing an arrangement of balls that corresponds to a distribution different from the MaxEnt one is exponentially (in N) less likely.

Finally (and perhaps most importantly for our purposes), H has been shown to be the *only* quantity whose constrained maximization allows for least-biased inference satisfying certain generic logical requirements (known as Shore–Johnson axioms) [66]. This result ultimately provides a rigorous basis for using the maximum entropy principle as a general inference technique, independently of the meaning assigned to Eq. (4). In other words, by maximizing H one is not looking for a state of maximum indeterminacy (apart from constraints), but rather following the only recipe for self-consistent inference having certain desirable properties. In this sense, the maximum entropy principle ‘simply’ allows to infer least-biased, constraint-satisfying probability distributions in a mathematically rigorous and logically sound manner.

It is clear at this point that the nature of the microscopic variables and the constraints one wants to impose are crucial in the business of using maximum entropy inference in general, and specifically to obtain information about biological systems from complex, high-dimensional data. In addition, the concrete usefulness of MaxEnt distributions besides their theoretical appeal is not *a priori* obvious. Mathematical arguments guarantee that they can provide a compact statistical description of a dataset that is least-biased and compatible with empirical observations. But can that description be employed e.g. for predictive purposes?

2.2. Maximum entropy inference in biology: the case of gene interaction networks

The answers to the questions posed above clearly depend on the specifics of the system under consideration and of the available data, and require a case-by-case discussion. Applications of the maximum entropy framework to biology have however become more numerous as the size and quality of data sets has increased, and currently range from protein science and neuroscience to collective animal behaviour and ecology. Such an impressive span suggests that at least some aspects must be recurrent across many if not all of these instances (see the discussion presented in [67] for a broader perspective). It is on these and on the lessons that can be drawn from them that we shall try to focus now. For sakes of clarity and simplicity, we shall phrase things in the context of the study of gene expression. Similar considerations can however be formulated in almost all of the cases listed above.

We begin by re-considering the urn-and-balls model described above in the gene expression scenario. The N balls would now represent N cellular samples whose complete expression profiles (e.g. RNA levels) have been experimentally characterized, while the K urns would represent all possible expression profiles. An expression profile of R genes is described by a vector $\mathbf{x} = \{x_i\}$, where x_i stands for the expression level of gene i , with $i = 1, \dots, R$. Usually, the expression of a large set of genes is monitored ($R \gg 1$). A specification of an expression profile for each of the N cells corresponds to an assignment of the N balls to the K urns, and fully describes our experimental sample. Note that the number of possible vectors \mathbf{x} , corresponding to the number K of urns, is in principle huge. By contrast, the number of samples (i.e. N) is typically much smaller than K , so that experiments will vastly under-sample the space of possible expression profiles.

Given the data (i.e. the measured expression profiles), the problem is that of inferring a probability distribution $p(\mathbf{x})$ of expression profiles that is (i) least-biased with respect to unavailable information, and (ii) consistent with empirical constraints. According to the maximum entropy principle, we have to find the distribution $p(\mathbf{x})$ that maximizes the entropy

$$H = - \sum_{\mathbf{x}} p(\mathbf{x}) \ln p(\mathbf{x}) \quad , \quad (6)$$

subject to data-derived constraints, the above sum being formally carried out over all possible expression profiles. Whether a quantity should be constrained or not is ultimately determined by whether one can reliably estimate it from data or not. In the instances encountered most often and of greater practical relevance, constraints involve low-order moments of the underlying variables, especially averages (first moments) and correlations (second moments). This is because the statistically

accurate computation of moments requires more and more experimental samples as the order of the moment increases, so that higher-order moments are generically harder to estimate than lower-order ones. We shall therefore consider only the simplest case in which the mean expression levels of each gene and the gene-gene correlations are required to match those derived from data, namely

$$\overline{x_i} = \frac{1}{N} \sum_{a=1}^N x_{i,a} \quad , \quad \overline{x_i x_j} = \frac{1}{N} \sum_{a=1}^N x_{i,a} x_{j,a} \quad , \quad (7)$$

where the index a runs over samples from 1 to N , while i and j range over genes from 1 to R . (It should however be kept in mind that this aspect is ultimately limited by data availability only.) Hence, we must look for the distribution $p(\mathbf{x})$ that maximizes (6) with the usual normalization condition

$$\sum_{\mathbf{x}} p(\mathbf{x}) = 1 \quad , \quad (8)$$

and such that the mean expression levels and the correlations match empirical ones (7), i.e.

$$\sum_{\mathbf{x}} x_i p(\mathbf{x}) = \overline{x_i} \quad \text{for all } i = 1, \dots, R \quad , \quad (9)$$

$$\sum_{\mathbf{x}} x_i x_j p(\mathbf{x}) = \overline{x_i x_j} \quad \text{for all } i \text{ and } j \quad . \quad (10)$$

Such a distribution $p^*(\mathbf{x})$ can be computed as shown in Supplementary Material, Sec. S5 (see also [16]) and reads

$$p^*(\mathbf{x}) = \frac{1}{Z} e^{\sum_{i=1}^R \beta_i x_i + \sum_{i < j} \gamma_{ij} x_i x_j} \quad , \quad (11)$$

where Z , β_i ($i = 1, \dots, M$) and γ_{ij} ($i, j = 1, \dots, M$ with $i \leq j$) are constants known as ‘Lagrange multipliers’ that are introduced to enforce the constraints [68]. Equation (11) is often referred to as the ‘pairwise MaxEnt probability distribution’ [24,28,40,62], as it involves at most couplings between pairs of variables through the last term in the argument of the exponential. Clearly, this is due to the fact that only moments up to the second are constrained.

Eq. (11) provides a formal solution to our problem. To fully evaluate it, though, the values of the Lagrange multipliers Z , β_i and γ_{ij} have to be computed self-consistently from (8), (9) and (10).³ As (11) represents, after all, the least-biased data-informed model for the expression profiles, solving this problem amounts to inferring the model’s parameters from data. Performing this task in a realistic context with $R \gg 1$ genes, which is ultimately the key for the effective implementation of the maximum entropy framework with biological data sets, can be an extremely

³ Note that, because $\overline{x_i x_j} = \overline{x_j x_i}$, the number of constraints is $1 + R + R(R+1)/2$ (where the three addends correspond respectively to (8), (9) and (10)), and it matches the number of constants to be determined. This makes the problem of retrieving the values of the above parameters mathematically well defined.

challenging computational problem. Luckily, a number of mathematically subtle but computationally effective methods have been developed for this goal over the past decade at the interface between statistical physics and computer science. A discussion of these techniques is however beyond our scopes, and excellent and up-to-date overviews can be found e.g. in [62,69]. We shall henceforth assume that the parameters β_i 's and the γ_{ij} 's have been computed, and focus on their physical and biological interpretation.

Based on the MaxEnt distribution (11), one sees that β_i measures the intrinsic propensity of gene i to be expressed, as larger (resp. smaller) values of β_i favor expression profiles \mathbf{x} with larger (resp. smaller) values of x_i . On the other hand, the γ_{ij} 's characterize the strength of pairwise gene-gene interactions as well as their character (via their signs: positive for positive interactions, negative for negative ones). Hence the γ_{ij} 's can in principle yield regulatory information that may be scaled up to the reconstruction of an effective genome-resolution gene-gene interaction network. In [16], for instance, such coefficients were used to infer regulatory interactions in *S. cerevisiae*, after expression profiles were experimentally characterized in cultures at different time points (representing the different samples) via microarrays. Ultimately, knowledge of the γ_{ij} 's allowed to extract a putative, highly-interconnected gene-gene interaction network that emphasized a few hub regulators (including ribosomal and mitochondrial genes as well as genes involved in TOR signaling) implicating global mechanisms devoted to the coordination of growth and nutrient intake pathways.

The ability to bring to light interconnections between genes belonging to different functional categories is a major advantage of the maximum entropy method over alternatives based on the straightforward analysis of correlations, such as clustering techniques. While the latter naturally focus on the identification of genes having a similar expression profile (and therefore tend to group functionally related genes together), the γ_{ij} 's point to a refined notion of correlation. The origin of this fact is especially transparent when the x_i 's are taken to be continuous unbounded variables ranging from $-\infty$ to $+\infty$ (a reasonable approximation whenever expression levels are quantified via centered log-fluorescence values). In this case, Eq. (11) describes a multivariate Gaussian distribution and it can be shown that the matrix of γ_{ij} 's is related to the *inverse* of the matrix of Pearson correlation coefficients, rather than to the correlation matrix itself [16,62]. This makes a substantial difference. Indeed, the covariance of the expression levels of two genes (say, A and B) can be large both when A and B are mutually dependent (e.g. when A codes for a transcription factor of B) and when, while mutually independent, they both separately correlate with a third gene C . In the latter case, though, the behaviour of C would explain the observed correlation between A and B . Specifically, by conditioning on the expression level of C , one would see that A and B are roughly uncorrelated. In

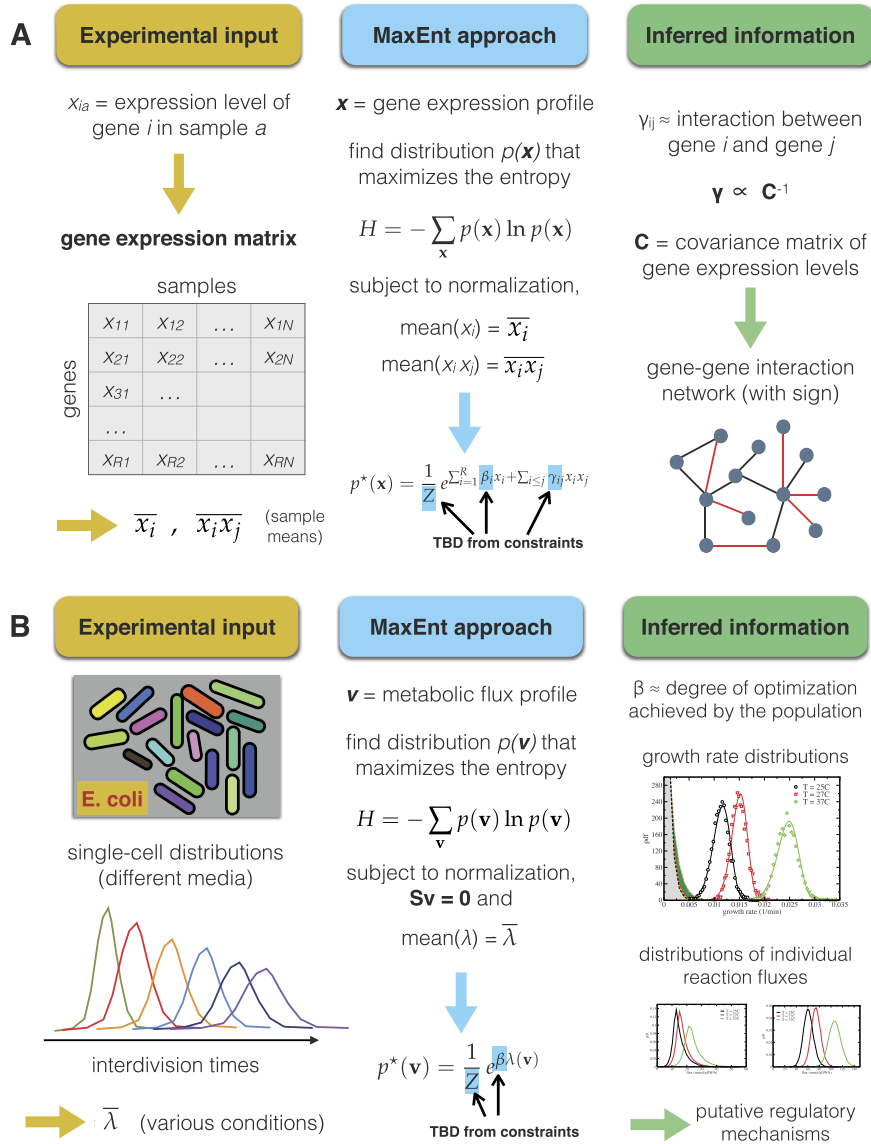


Figure 1. Sketch of the two examples of applications of the maximum entropy principle to biological data analysis discussed in the text. (A) Inference of gene interaction networks from empirical expression data (see Sec. 3 for details). (B) Inference of genome-scale metabolic flux patterns from empirical growth rate distributions in bacteria (see Sec. 4 for details). For each case, we describe schematically the empirical input (left column), the formulation of the maximum entropy inference problem (middle column), and an example of the inferred biological insight (right column).

other terms, the correlation matrix captures the unconditional correlation between variables and therefore contains effects due to both direct and indirect mechanisms. On the other hand, its inverse describes the correlations that remain once the indirect effects are removed [70,71], and thereby provides a more robust and consistent characterization of the interactions between variables. (See Figure 1A for a summary of the scenario just discussed.)

Naturally, some adjustments must come in place when one is to go beyond the Gaussian case. Yet, it can be argued that the above picture is quite generic. While the term ‘interaction’ can acquire different meanings in different cases, the MaxEnt distribution focuses on the most relevant part of the correlations and is therefore capable of extracting a more reliable interaction structure from data than that obtainable via a more standard correlation analysis. This property lies in our view at the heart of the success encountered by the maximum entropy method in several applications, in biology as well as in other fields [72,73,74]. The above example also shows the centrality of empirical data for maximum entropy inference. In some cases, though, constraints that are not derived from experiments can be employed, together with empirical ones, to guide the inference. An example of this is found in the maximum entropy approach to the analysis of metabolic networks.

2.3. Maximum entropy approach to cellular metabolism

Novel experimental techniques employing e.g. microfluidic devices are capable of probing growth in bacterial populations at single cell resolution, yielding detailed data that monitor growth in thousands of individual cells over many generations [75,76]. These experiments have quantified a number of features linking gene expression and metabolism to overall control mechanisms in proliferating bacteria [77,78,79,80,81]. While the emergent picture is being increasingly refined, tracking its ‘microscopic’ origin, and particularly the causes of growth rate fluctuations, is largely an open problem. Given the time scales involved in these processes, it is reasonable to think that the regulatory layer controlling energy metabolism is crucially involved in establishing this scenario. Indeed, substantial empirical evidence is connecting growth physiology and heterogeneity to metabolic activity in bacteria [77]. Experimental approaches to characterize the fluxes of intracellular metabolic reactions can provide a population-level picture of the mean activity of central carbon pathways [82]. On the other hand, cellular metabolic networks have been mapped, for many organisms, at the scale of the whole genome [83], and a host of self-consistent computational frameworks exist that can connect metabolic phenotypes (i.e. patterns of material fluxes through the enzyme-catalyzed network of intracellular reactions that processes nutrients into macromolecular building blocks, free energy and biomass) to physiological observables such as the growth rate [84]. The question is whether the reverse problem of relating measured growth rates to flux states of genome-scale networks is feasible. In particular, given the distribution of growth rates found in experiments, can one infer the state of the underlying metabolic network, e.g. the rates of individual reactions? Such knowledge might provide important insight into the metabolic bottlenecks of growth in proliferating cells, which could be especially useful in view of the limited experimental accessibility of intracellular reactions.

To progress along this route, it is important to characterize the map between metabolic networks and growth rates in some more detail. The so-called Constraint-Based Models (CBMs) [85] represent highly effective *in silico* schemes to obtain information about cellular metabolic activity from minimal input ingredients known at genome resolution. To be more specific, the central assumption behind CBMs is that, because of the timescale separation between metabolic (fast) and regulatory (slow) processes, over a sufficiently long time (compared to the cell's cycle) metabolic reactions operate close to a non-equilibrium steady state (NESS) where metabolite and enzyme levels are stationary. Under this homeostatic scenario, a viable configuration of fluxes through enzyme-catalyzed reactions can be represented by a vector $\mathbf{v} = \{v_i\}_{i=1}^R$ (with i indexing reactions) that satisfies the linear system of equations taking the matricial form

$$\mathbf{S}\mathbf{v} = \mathbf{0} , \quad (12)$$

where \mathbf{S} denotes the reaction network's matrix of stoichiometric coefficients. From a physical viewpoint, if exchange fluxes with the surrounding medium are included in \mathbf{S} (as usually done), the above conditions simply express the fact that, at stationarity, the total mass of each chemical species should be conserved, and correspond to Kirchhoff-like laws: the overall production (including external supply) and consumption (including excretions) fluxes of every metabolite should balance. \mathbf{S} is obtained from genome-scale reconstructions and has R columns (one per reaction, numbering to about 1,200 in an organism like *E. coli*) and C rows (one per chemical species, amounting to several hundreds for *E. coli*). Therefore (12) compactly represents C linear equations with R unknowns (the individual fluxes v_i). Any flux vector \mathbf{v} solving (12) corresponds in principle to a viable NESS flux pattern of the metabolic network specified by \mathbf{S} , and the structure of stoichiometric matrices usually allows for an infinite number of solutions. In particular, when ranges of variability of the type $v_{i,\min} \leq v_i \leq v_{i,\max}$ are specified for each flux, reflecting empirical kinetic, thermodynamic or regulatory priors (see e.g. [86] for an overview of these data-driven factors), solutions of (12) form a particular kind of set called a 'convex polytope' in mathematical jargon [87]. We shall denote convex polytopes by the letter \mathcal{P} . On the other hand, to each of the solutions, i.e. to each point in \mathcal{P} , CBMs associate a unique value for the biomass output (the growth rate), which we denote as $\lambda(\mathbf{v})$. Therefore, any rule to sample points from \mathcal{P} (i.e. to generate solutions of (12)) will in turn yield a distribution of values for the growth rate. Usually, though, convex polytopes corresponding to genome-scale stoichiometric matrices \mathbf{S} have very high dimensionality (e.g. several hundreds for *E. coli*), and therefore sadly tend to escape both imagination and computational analysis. We will not detail here how points from \mathcal{P} can be generated. For our purposes, it will suffice to say that efficient algorithms exist that allow to extract solutions of (12) with any desired probability distribution for any metabolic network reconstruction, i.e. any \mathbf{S} (see e.g. [84,88,89]).

Given this setup allowing to link metabolic phenotypes to their corresponding growth rates, we are interested in characterizing the inverse map. Specifically, what can we learn about flux configurations from empirical growth rate distributions? Following the maximum entropy idea, one can start by studying the least-biased distribution compatible with data. The simplest data-borne constraint one can impose concerns the mean growth rate. In addition, however, one can inject extra information by requiring that flux vectors are viable NESS of the metabolic network, i.e. that they belong to the polytope of solutions of (12) for the organism under study. This is a substantial change compared e.g. to the case of regulatory networks discussed in the previous section. There, no restriction on the vectors \mathbf{x} of expression profiles applied as no specific assumption on the ranges of variability and mutual dependence of individual expression levels was made. Here, instead, we are effectively adding a (reasonable and motivated) guess for the underlying model for flux profiles: they should satisfy (12) with pre-determined ranges of variability on each flux. The MaxEnt flux distribution for this case can be computed in full analogy with the previous ones (see Supplementary Material, Sec. S6), and it turns out to be given by

$$p^*(\mathbf{v}) = \begin{cases} \frac{1}{Z} e^{\beta\lambda(\mathbf{v})} & \text{if } \mathbf{S}\mathbf{v} = \mathbf{0} , \\ 0 & \text{otherwise} , \end{cases} \quad (13)$$

where Z and β are the Lagrange multipliers respectively ensuring normalization ($\sum_{\mathbf{v}} p^*(\mathbf{v}) = 1$) and the given mean growth rate. The MaxEnt distribution is only defined on the feasible space \mathcal{P} where $\mathbf{S}\mathbf{v} = \mathbf{0}$ (i.e. flux vectors must be viable). In addition, the probability to observe a certain \mathbf{v} depends on its growth rate $\lambda(\mathbf{v})$, while the parameter β quantifies the inferred “degree of optimality”. In a nutshell, a sufficiently large value of β causes p^* to concentrate around flux vectors yielding large values of the growth rate λ . On the other hand, as β decreases towards more and more negative values, the MaxEnt distribution selects metabolic phenotypes growing more and more slowly. For $\beta = 0$, p^* becomes the uniform distribution over the polytope, in which case each viable flux vector \mathbf{v} is assumed to be equally probable.

When applied to modeling data describing steady growth of *E. coli* populations using the genome-scale metabolic network reconstruction given in [90], Eq. (13) turned out to reproduce not just the mean growth rates obtained in different experiments for a number of growth medium/strain combinations, but the entire distributions [91]. The fact that the exponential form (13), which ultimately depends on the single parameter β (see Supplementary Material, Sec. S6), coincides with empirical distributions confirms the observation that growth rate distributions are one-parameter functions (i.e., the variance is a function of the mean, at odds with Gaussian distributions where the mean and the variance are separate parameters) [79]. Strikingly, recent work has shown that the maximum entropy approach, besides

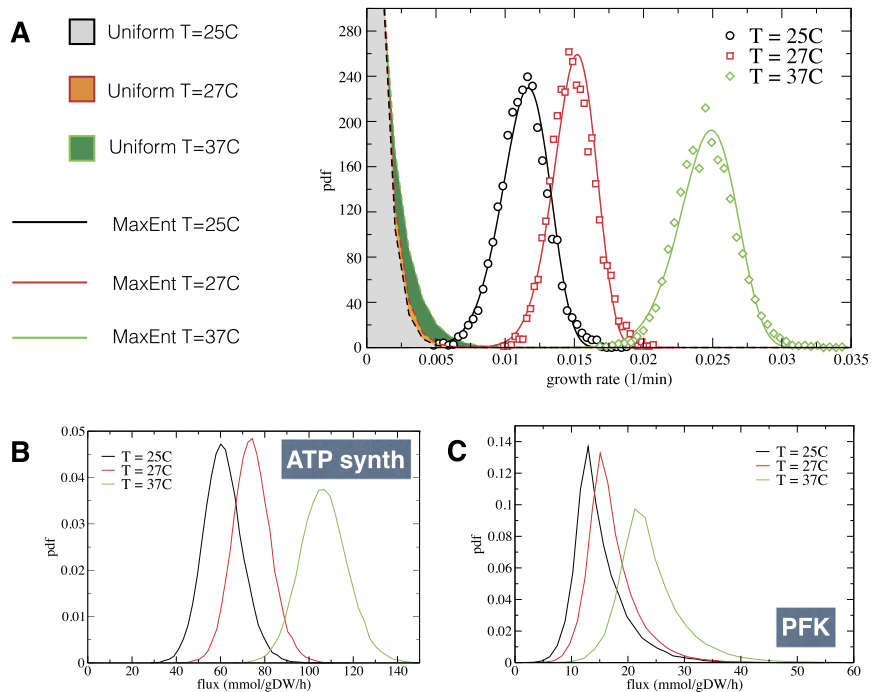


Figure 2. Maximum entropy modeling of growth rate distributions describing *E. coli* growth at different temperatures. Data are taken from [76]. (A) Empirical distributions (markers) are shown together with the MaxEnt distributions obtained by fitting β to match the corresponding means (continuous lines), for three different temperatures. For comparison, the dashed lines and the corresponding shaded areas describe the growth rate distributions corresponding to uniform samplings of the solution spaces of the metabolic model, Eq. (12), in the three cases. In each case, such distributions are described by Eq. (14), with $a = 0$, $b = 22$ and different values of λ_{\max} . (B) and (C): inferred distributions of the ATP synthase flux (B) and of the flux through phosphofructokinase (PFK) (C) at different temperatures.

capturing the statistics of the growth rate, is also capable of describing the behaviour of intracellular fluxes belonging to the central carbon processing pathways, which can be estimated by mass spectrometry, without additional assumptions [92]. Equation (13) is likewise capable of describing bacterial growth distributions obtained in a fixed medium at different temperatures (see Figure 2A), while two examples of predictions for how metabolic flux distributions will be modulated as the growth temperature increases are displayed in Figures 2B and 2C. (See Figures 1B for an overview of this case.)

An especially important feature that the maximum entropy approach brings to light is the fact that the value of β that provides the best fit to experiments corresponds to mean growth rates that are significantly smaller (usually between 50% and 80%) than the maximum growth rate achievable in the same growth medium according to the CBM prediction, which we denote as λ_{\max} . In other terms, based on the experimental data sets considered in [76,91], bacteria appear unable to strictly maximize their

growth rates. On the other hand, the distribution of growth rates implied by (13) at $\beta = 0$ (i.e., when all flux patterns satisfying (12) are equally likely) is of the form

$$q(\lambda) = A\lambda^a(\lambda_{\max} - \lambda)^b, \quad (14)$$

where a and b are constants that depend on the metabolic network reconstruction one is employing⁴ and A is a normalization constant. The growth-rate landscape depicted by (14) is extremely heterogeneous, with a small set of states with growth rates close to λ_{\max} living in a huge sea of slow-growing flux patterns (see the dashed curves in Figures 2A). This suggests that any random noise added to a fast-growing flux pattern will overwhelmingly likely causes a drastic growth rate reduction.

Taking these observations together, one sees that the space of allowed flux patterns can affect, in a quantitatively measurable way, the growth rate distributions that a population of cells in a given medium will achieve. In particular, one could argue that empirical growth rate distributions found in experiments at single-cell resolution might be explained in terms of a trade-off between the higher fitness of fast-growing phenotypes and the higher entropy (numerosity) of slow-growing ones that is established due to the action of noise. This idea has been tested in a mathematical model of an *E. coli* population evolving in time in the CBM-based fitness landscape (14). Indeed, when bacterial growth rates, along with driving replication, were assumed to fluctuate in time according a small diffusive noise in the feasible space (which contrasts, with high probability, the tendency of the population to concentrate around the fastest growth rates achievable), a scenario that is essentially identical to that described by the MaxEnt distribution was recovered despite starting from very different premises [91]. A more careful mathematical study of the same dynamical model has predicted, among other things, that, within such a scenario, response times to perturbations should be inversely correlated to the difference between the maximum achievable growth rate and the population average [93]. In other terms, populations growing sub-optimally may be more efficient in responding to stresses, a prediction that in principle can be tested experimentally.

Given that a single number, i.e. β , appears to provide a full description of empirical growth rate distributions via (13), it would be important to have a more thorough understanding of its physical and biological meaning. According to (13), a larger β implies a faster mean growth rate, but can one point to the physico-chemical and biological determinants of growth that contribute to establishing its value in a bacterial population? More specifically, can one identify the factors that limit β ? In principle, one would expect the growth medium to play a central role in the answer to these questions. However, the picture emerging from a mathematical model in which

⁴ For instance, one has $a = 22$ and $b = 0$ for *E. coli*'s on the carbon catabolic core, but $a = 171$ and $b = 3$ for the full genome-scale network given in [90].

β can be computed from first principles is more involved. In particular, Ref. [94] considered a population growing in a certain environment and characterized by a growth rate distribution $r(\lambda)$. Imagine sampling N_0 individuals from that population, each carrying its intrinsic growth rate sampled from $r(\lambda)$, and subsequently planting them as the initial inoculum in a new growth medium with carrying capacity K . By assuming that growth follows the basic logistic model, it was found that the N_0 seeds evolve in time into a population with growth rate distribution proportional to $\exp(\beta\lambda)$, where β is now a quantity that depends in a mathematically precise way on the capacity-to-inoculum ratio K/N_0 and on the growth rate distribution from which the seeds were sampled, i.e. on $r(\lambda)$. The former dependence encodes, as expected, for the growth medium via K . Interestingly, though, the presence of N_0 says that the population maintains a memory of initial conditions. On the other hand, the fact that β is also a function of $r(\lambda)$ points, perhaps unexpectedly, towards history-dependence. This theoretical picture, which characterizes the maximum entropy scenario at a deeper level and provides quantitative support to some possibly intuitive facts (such as history- and inoculum-dependence of growth properties), has been in part confirmed by empirical data on cancer growth rates [94]. Clearly, this oversimplified model does not account explicitly for factors like direct cell-cell interactions or feedbacks between growth and regulation or nutrient availability. Still, it is interesting that the parameter β , which in principle is introduced here only to enforce a constraint in the maximum entropy scenario, can be given a well defined physical interpretation. The integration of further empirical data and possibly constraints (e.g. concerning individual fluxes) will hopefully provide new insight into this picture. Preliminary results obtained in this direction are encouraging [95].

While the maximum entropy idea has been employed within CBMs for specific purposes like objective function reconstruction, metabolic pathway analysis or to compute distributions of individual fluxes or chemical potentials over the polytope [96,97,98,99,100,101,102,103,104,105,106,107,108,109], the approach just discussed presents an overall view of cellular metabolism that differs significantly from that of mechanistic CBMs such as Flux Balance Analysis [110] or related ideas [111,112,113], despite the fact that both rely on essentially the same physical NESS assumption via (12). In many ways, the two frameworks appear to be complementary. Flux Balance Analysis is capable of describing the optimal metabolic states of fastest growth achievable by a cell in a multitude of environmental and intracellular conditions. The maximum entropy approach, instead, can clarify in quantitative terms how far from the optimum an actual population is, and what population-level feature might be shaping the observed growth rate heterogeneity. These views might become more tightly linked upon further investigating the underlying regulatory mechanisms, molecular interactions or trade-offs limiting growth.

2.4. Beyond the basic framework

The major advantage of the maximum entropy principle lies perhaps in its ability to cope effectively with limited data. The space of states accessible to a living system is huge and always undersampled in experiments. Still, as long as the available data permit the estimation of basic statistical observables with sufficient accuracy, a variety of efficient computational methods exist that allow to compute the parameters of the MaxEnt distribution reliably, leading to a compact, least-biased and mathematically sound representation of a complex, high-dimensional interacting system. We have attempted to describe the fundamentals of entropy maximization and its use for biological applications, opting to focus on two of perhaps the simplest instances involving the study of biological genome-scale networks. Different applications do not require conceptual changes to the approach we presented, but may rely on a more careful and/or involved definition of the state variables (expression profiles or metabolic flux profiles in the examples we considered). Yet some of the issues that we have chosen to leave aside so far now deserve a deeper (albeit brief) discussion.

In first place, we have seen that computing the MaxEnt distribution is akin to constructing a statistical model of the system one is interested in. The inferred model will inevitably depend on the empirical information encoded in the constraints under which entropy is maximized. In turn, while MaxEnt models necessarily reproduce the information used to build them, their predictive power will depend on the encoded constraints as well. In many cases, maximum entropy models can correctly reproduce correlations of order higher than those included as constraints (see e.g. [16,24] for examples, and [114] for a broader theoretical analysis). Still, predictions concerning other quantities may turn out to be incorrect. Assuming the imposed constraints are factual, this can happen essentially for a unique reason: the constrained quantities do not, by themselves, localize the distribution over states where the new observations are matched. In other terms, more (or different) constraints are required. Clearly, as constraints map to physical or biological ingredients, missing constraints can point to useful insight about of the system under study. An example of this is shown in [115], where some phenomenological aspects of the coexistence between fast-growing and persister phenotypes in bacterial populations are explained in terms of a maximum entropy approach with a constraint on growth rate fluctuations.

Secondly, in our presentation we have taken for granted that cells are indistinguishable (i.e., the arrangement $\{n_1, \dots, n_K\}$ is indistinguishable from the arrangement obtained by exchanging the positions of two balls) and that the set of allowed states is discrete (e.g., K urns). As it turns out, these are the simplest and ideal conditions for applying the maximum entropy principle. In many of

the biological applications discussed in the literature, it can be argued that both conditions are satisfied. However, this is not generally so. For instance, reaction fluxes are continuous variables and the space of allowed flux configurations is therefore continuous rather than discrete. While nothing invalidates the maximum entropy principle in such cases and the conclusions are unchanged, some more care is needed in setting the stage for it in presence of continuous variables. The interested reader will find more details e.g. in [61,116].

Thirdly, MaxEnt models like that described by (11) implicitly postulate, via the constraint (10), that the coefficients γ_{ij} are symmetric, i.e. $\gamma_{ij} = \gamma_{ji}$. This amounts to assuming that the gene-gene interaction network is *a priori* undirected, which obviously is at odds with a host of biological evidence. To overcome this limitation, which evidently occurs in the maximum entropy approach whenever correlations are constrained, one has to abandon the framework discussed here and resort to methods that infer dynamical models (as opposed to static ones that can be fully described by the probability to observe a certain state). The reasons are rooted in the fundamental distinction between equilibrium and off-equilibrium processes, a nice discussion of which can be found in [60]. Efficient computational methods have been developed to deal with such situations as well, and we again refer the reader to [69] for a recent overview. A specific generalization of the maximum entropy idea that focuses on inferring distributions of *dynamical trajectories* in the configuration space is known as ‘Maximum Caliber’ [117] (see also [118]). Applications of these ideas are presented e.g. in [119,120,121,122,123] (see also [60] for a review).

Fourth, in our presentation we have tacitly assumed that (i) a MaxEnt distribution exists (i.e., that, once the mathematical problem is constructed, there is a distribution that actually maximizes the entropy subject to the imposed constraints), and that (ii) it is unique (as intuitively desirable to avoid ambiguities, and in compliance with one of the Shore–Johnson axioms). Whether this is the case, ultimately depends on the imposed constraints. For our purposes, it should suffice to say that whenever the constraints are linear functions of the probability distribution $\{p_i\}$, as are e.g. (8), (9) and (10), as well as in all instances discussed here, both (i) and (ii) are true. Yet, in certain situations this may not be the case [124]. The maximum entropy approach then loses some of its appeal and the problem of performing robust inference from limited data has to be treated on a case-by-case basis with extra care.

A final important point we have so far not addressed concerns the treatment of measurement errors affecting the features used to constrain entropy maximization. In practice, inference is always performed under uncertain constraints, which in turn leads to uncertain estimates of the inferred parameters (i.e., for instance, of the inferred interaction structure). The straightforward application of the Maximum Entropy idea discussed so far effectively ignores this aspect. In order to account for it, one can extend the MaxEnt framework presented here in the direction of Bayesian

inference, which explicitly deals with probability distributions of parameters given the data [125]. The reconciliation of the Maximum Entropy and Bayesian approaches poses several challenges and has a long and intriguing history [126,127], up to very recent applications [6,11,128,129,130]. A related aspect concerns MaxEnt inference with missing or incomplete data, a well-known example of which is discussed in [48,131]. Such cases have been long considered in the inference literature [132] and specific methods have been developed to deal with them [133,134,135]. As more high-resolution data probing large-scale biological networks become available, these techniques are highly likely to be extended well beyond their current application domains.

3. Conclusions

The existence of reproducible quantitative relationships connecting the growth of a bacterial population to the composition of the underlying cells (e.g. in terms of the RNA/protein ratio) suggests that, when integrated over large enough cell populations, regulatory and metabolic mechanisms can generate stable outcomes in spite of the heterogeneity and noise that affect them all [136]. Significant deviations from the expected outcome are rare. From a statistical viewpoint, one might say that in such cases the law of large numbers (or more precisely the central limit theorem [137]) is at work, and population-level properties will be roughly independent of the way in which cells in the population distribute over allowed states (or, equivalently, that all viable distributions lead to the same population-level properties). On the other hand, this distribution encodes for critical biological information related to robustness, selection and evolvability, and having access to it would be of paramount importance.

With detailed information about individual intracellular processes, one may hope to construct sufficiently comprehensive mechanistic models capable of mapping the behaviour of individual cells, in terms of their regulatory and metabolic activity, to population-level observables. It is however unclear whether such a ‘direct’ approach, whose concrete realization would be severely hampered by the huge number of variables and parameters to be accounted for, would allow to fully uncover how cells are distributed over allowed regulatory states, as the space of states to be explored would be dauntingly large. The ‘inverse’ route consists in trying to infer the distribution from the observed population-level behaviour. As many distributions are likely to be compatible with empirical results, one would ideally want to be able to sample all of them. The space of such distributions can however be prohibitively large for systems as complex as cells. Therefore, the problem of selecting, out of this space, the most informative distribution has to be faced. The maximum entropy principle provides a constructive and mathematically controlled answer: it is the

distribution that maximizes the entropy (4) subject to empirical constraints that yields the optimal choice.

Some of the issues raised above however suggest that, as much as the maximum entropy principle provides powerful means to extract models and/or useful low-dimensional representations from complex, high-dimensional and limited data, there is room to dissect its fundamentals [138,139], re-analyze its use [140,141], or search for alternatives [142]. In our view, besides providing essential theoretical insight, these contributions also highlight some of the main practical challenges that biological datasets pose to computational and theoretical scientists, whose ultimate goals are interpreting them and using them to build e.g. predictive models and *de novo* design protocols. These challenges have ultimately been the key driver behind the massive progress achieved in the study of the so-called ‘inverse problems’ in statistical physics over the past decade. As one can only envision that data will continue to get more and more abundant, of higher quality and increasingly diverse (as novel conceptual schemes emerge), the push for technical improvements and new schemes will likely escalate. In turn, maximum entropy methods may spread further in the coming years, as they are capable of extracting the simplest and least biased conclusions that one can reliably draw from limited data. Some of the new directions that are being probed have already shown promise for applications in biology [143,144].

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Funding statement

This work was supported by the European Union’s Horizon 2020 Research and Innovation Staff Exchange program MSCA-RISE-2016 under grant agreement no. 734439 INFERNET, and by the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under Research Executive Agency (REA) Grant Agreement No. 291734.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2018.e00596>.

Acknowledgements

We gratefully acknowledge Alfredo Braunstein, Anna Paola Muntoni and Andrea Pagnani for useful insight and suggestions.

References

- [1] E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.
- [2] A.S. Lapedes, B.G. Giraud, L. Liu, G.D. Stormo, *Statistics in Molecular Biology and Genetics*, Lect. Notes Monogr. Ser., vol. 33, Institute of Mathematical Statistics, 1999, pp. 236–256.
- [3] F. Seno, A. Trovato, J.R. Banavar, A. Maritan, Maximum entropy approach for deducing amino acid interactions in proteins, *Phys. Rev. Lett.* 100 (7) (2008) 078102.
- [4] M. Weigt, R.A. White, H. Szurmant, J.A. Hoch, T. Hwa, Identification of direct residue contacts in protein–protein interaction by message passing, *Proc. Natl. Acad. Sci.* 106 (1) (2009) 67–72.
- [5] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D.S. Marks, C. Sander, R. Zecchina, J.N. Onuchic, T. Hwa, M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, *Proc. Natl. Acad. Sci.* 108 (49) (2011) E1293–E1301.
- [6] J.W. Pitera, J.D. Chodera, On the use of experimental observations to bias simulated ensembles, *J. Chem. Theory Comput.* 8 (10) (2012) 3445–3451.
- [7] T.A. Hopf, L.J. Colwell, R. Sheridan, B. Rost, C. Sander, D.S. Marks, Three-dimensional structures of membrane proteins from genomic sequencing, *Cell* 149 (7) (2012) 1607–1621.
- [8] B. Roux, J. Weare, On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method, *J. Chem. Phys.* 138 (8) (2013) 02B616.
- [9] A. Cavalli, C. Camilloni, M. Vendruscolo, Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles

- according to the maximum entropy principle, *J. Chem. Phys.* 138 (9) (2013) 03B603.
- [10] M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, E. Aurell, Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models, *Phys. Rev. E* 87 (1) (2013) 012707.
- [11] W. Boomsma, J. Ferkinghoff-Borg, K. Lindorff-Larsen, Combining experiments and simulations using the maximum entropy principle, *PLoS Comput. Biol.* 10 (2) (2014) e1003406.
- [12] B. Zhang, P.G. Wolynes, Topology, structures, and energy landscapes of human chromosomes, *Proc. Natl. Acad. Sci.* 112 (19) (2015) 6062–6067.
- [13] A. Cesari, S. Reißer, G. Bussi, Using the maximum entropy principle to combine simulations and solution experiments, preprint, arXiv:1801.05247, 2018.
- [14] P. Farré, E. Emberly, A maximum-entropy model for predicting chromatin contacts, *PLoS Comput. Biol.* 14 (2) (2018) e1005956.
- [15] P. D’haeseleer, S. Liang, R. Somogyi, Genetic network inference: from co-expression clustering to reverse engineering, *Bioinformatics* 16 (8) (2000) 707–726.
- [16] T.R. Lezon, J.R. Banavar, M. Cieplak, A. Maritan, N.V. Fedoroff, Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns, *Proc. Natl. Acad. Sci.* 103 (50) (2006) 19033–19038.
- [17] P.S. Dhadialla, I.E. Ohiorhenuan, A. Cohen, S. Strickland, Maximum-entropy network analysis reveals a role for tumor necrosis factor in peripheral nerve development and function, *Proc. Natl. Acad. Sci.* 106 (30) (2009) 12494–12499.
- [18] F. Remacle, N. Kravchenko-Balasha, A. Levitzki, R.D. Levine, Information-theoretic analysis of phenotype changes in early stages of carcinogenesis, *Proc. Natl. Acad. Sci.* 107 (22) (2010) 10324–10329.
- [19] V.A. Huynh-Thu, G. Sanguinetti, Gene regulatory network inference: an introductory survey, preprint, arXiv:1801.04087, 2018.
- [20] J.W. Locasale, A. Wolf-Yadlin, Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data, *PLoS ONE* 4 (8) (2009) e6522.

- [21] T.G. Graeber, J.R. Heath, B.J. Skaggs, M.E. Phelps, F. Remacle, R.D. Levine, Maximal entropy inference of oncogenicity from phosphorylation signaling, *Proc. Natl. Acad. Sci.* 107 (13) (2010) 6112–6117.
- [22] Y.S. Shin, F. Remacle, R. Fan, K. Hwang, W. Wei, H. Ahmad, R.D. Levine, J.R. Heath, Protein signaling networks from single cell fluctuations and information theory profiling, *Biophys. J.* 100 (10) (2011) 2378–2386.
- [23] R. Sharan, R.M. Karp, Reconstructing Boolean models of signaling, *J. Comput. Biol.* 20 (3) (2013) 249–257.
- [24] E. Schneidman, M.J. Berry, R. Segev, W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature* 440 (7087) (2006) 1007–1012.
- [25] J. Shlens, G.D. Field, J.L. Gauthier, M.I. Grivich, D. Petrusca, A. Sher, A.M. Litke, E.J. Chichilnisky, The structure of multi-neuron firing patterns in primate retina, *J. Neurosci.* 26 (32) (2006) 8254–8266.
- [26] A. Tang, D. Jackson, J. Hobbs, W. Chen, J.L. Smith, H. Patel, A. Prieto, D. Petrusca, M.I. Grivich, A. Sher, P. Hottowy, A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro, *J. Neurosci.* 28 (2) (2008) 505–518.
- [27] S. Cocco, S. Leibler, R. Monasson, Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods, *Proc. Natl. Acad. Sci.* 106 (33) (2009) 14058–14062.
- [28] Y. Roudi, S. Nirenberg, P.E. Latham, Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't, *PLoS Comput. Biol.* 5 (5) (2009) e1000380.
- [29] G. Tkacik, J.S. Prentice, V. Balasubramanian, E. Schneidman, Optimal population coding by noisy spiking neurons, *Proc. Natl. Acad. Sci.* 107 (32) (2010) 14419–14424.
- [30] I.E. Ohiorhenuan, F. Mechler, K.P. Purpura, A.M. Schmid, Q. Hu, J.D. Victor, Sparse coding and high-order correlations in fine-scale cortical networks, *Nature* 466 (7306) (2010) 617–621.
- [31] F.C. Yeh, A. Tang, J.P. Hobbs, P. Hottowy, W. Dabrowski, A. Sher, A. Litke, J.M. Beggs, Maximum entropy approaches to living neural networks, *Entropy* 12 (1) (2010) 89–106.
- [32] E. Ganmor, R. Segev, E. Schneidman, Sparse low-order interaction network underlies a highly correlated and learnable neural population code, *Proc. Natl. Acad. Sci.* 108 (23) (2011) 9679–9684.

- [33] E. Granot-Atedgi, G. Tkacik, R. Segev, E. Schneidman, Stimulus-dependent maximum entropy models of neural population codes, *PLoS Comput. Biol.* 9 (3) (2013) e1002922.
- [34] G. Tkacik, O. Marre, T. Mora, D. Amodei, M.J. Berry II, W. Bialek, The simplest maximum entropy model for collective behavior in a neural network, *J. Stat. Mech.* 2013 (03) (2013) P03011.
- [35] U. Ferrari, T. Obuchi, T. Mora, Random versus maximum entropy models of neural population activity, *Phys. Rev. E* 95 (4) (2017) 042321.
- [36] V. Rostami, P.P. Mana, S. Grün, M. Helias, Bistability, non-ergodicity, and inhibition in pairwise maximum-entropy models, *PLoS Comput. Biol.* 13 (10) (2017) e1005762.
- [37] T.A. Nghiem, B. Telenczuk, O. Marre, A. Destexhe, U. Ferrari, Maximum entropy models reveal the correlation structure in cortical neural activity during wakefulness and sleep, preprint, arXiv:1801.01853, 2018.
- [38] G. Yeo, C.B. Burge, Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals, *J. Comput. Biol.* 11 (2–3) (2004) 377–394.
- [39] T. Mora, A.M. Walczak, W. Bialek, C.G. Callan, Maximum entropy models for antibody diversity, *Proc. Natl. Acad. Sci.* 107 (12) (2010) 5405–5410.
- [40] M. Santolini, T. Mora, V. Hakim, A general pairwise interaction model provides an accurate description of in vivo transcription factor binding sites, *PLoS ONE* 9 (6) (2014) e99015.
- [41] A.L. Ferguson, J.K. Mann, S. Omarjee, T. Ndungu, B.D. Walker, A.K. Chakraborty, Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design, *Immunity* 38 (3) (2013) 606–617.
- [42] K. Shekhar, C.F. Ruberman, A.L. Ferguson, J.P. Barton, M. Kardar, A.K. Chakraborty, Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes, *Phys. Rev. E* 88 (6) (2013) 062705.
- [43] J.K. Mann, J.P. Barton, A.L. Ferguson, S. Omarjee, B.D. Walker, A. Chakraborty, T. Ndungu, The fitness landscape of HIV-1 gag: advanced modeling approaches and validation of model predictions by in vitro testing, *PLoS Comput. Biol.* 10 (8) (2014) e1003776.
- [44] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, A.M. Walczak, Statistical mechanics for natural flocks of birds, *Proc. Natl. Acad. Sci.* 109 (13) (2012) 4786–4791.

- [45] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, A.M. Walczak, Social interactions dominate speed control in poising natural flocks near criticality, *Proc. Natl. Acad. Sci.* 111 (20) (2014) 7212–7217.
- [46] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, A.M. Walczak, Dynamical maximum entropy approach to flocking, *Phys. Rev. E* 89 (4) (2014) 042707.
- [47] B. Shipley, D. Vile, É. Garnier, From plant traits to plant communities: a statistical mechanistic approach to biodiversity, *Science* 314 (5800) (2006) 812–814.
- [48] S.J. Phillips, R.P. Anderson, R.E. Schapire, Maximum entropy modeling of species geographic distributions, *Ecol. Model.* 190 (3) (2006) 231–259.
- [49] S. Pueyo, F. He, T. Zillio, The maximum entropy formalism and the idiosyncratic theory of biodiversity, *Ecol. Lett.* 10 (11) (2007) 1017–1028.
- [50] R.C. Dewar, A. Porté, Statistical mechanics unifies different ecological patterns, *J. Theor. Biol.* 251 (3) (2008) 389–403.
- [51] S.J. Phillips, M. Dudik, Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation, *Ecography* 31 (2) (2008) 161–175.
- [52] J. Harte, T. Zillio, E. Conlisk, A.B. Smith, Maximum entropy and the state-variable approach to macroecology, *Ecology* 89 (10) (2008) 2700–2711.
- [53] I. Volkov, J.R. Banavar, S.P. Hubbell, A. Maritan, Inferring species interactions in tropical forests, *Proc. Natl. Acad. Sci.* 106 (33) (2009) 13854–13859.
- [54] F. He, Maximum entropy, logistic regression, and species abundance, *Oikos* 119 (4) (2010) 578–582.
- [55] E.P. White, K.M. Thibault, X. Xiao, Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model, *Ecology* 93 (8) (2012) 1772–1778.
- [56] C.K. Fisher, T. Mora, A.M. Walczak, Variable habitat conditions drive species covariation in the human microbiota, *PLoS Comput. Biol.* 13 (4) (2017) e1005435.
- [57] R. Menon, V. Ramanan, K.S. Korolev, Interactions between species introduce spurious associations in microbiome studies, *PLoS Comput. Biol.* 14 (1) (2018) e1005939.
- [58] E.T. Jaynes, Information theory and statistical mechanics I, *Phys. Rev.* 106 (1957) 620–630.

- [59] E.T. Jaynes, Information theory and statistical mechanics II, *Phys. Rev.* 108 (1957) 171–190.
- [60] S. Pressé, K. Ghosh, J. Lee, K.A. Dill, Principles of maximum entropy and maximum caliber in statistical physics, *Rev. Mod. Phys.* 85 (3) (2013) 1115.
- [61] A. Lesne, Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics, *Math. Struct. Comput. Sci.* 24 (3) (2014) e240311.
- [62] R.R. Stein, D.S. Marks, C. Sander, Inferring pairwise interactions from biological data using maximum-entropy probability models, *PLoS Comput. Biol.* 11 (7) (2015) e1004182.
- [63] R.A. Brualdi, *Introductory Combinatorics*, 5th ed., Prentice-Hall, 2010.
- [64] C. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* 27 (1948) 379–423 and 623–656.
- [65] A.Y. Khinchin, *Mathematical Foundations of Information Theory*, Courier Corporation, 2013.
- [66] J. Shore, R. Johnson, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inf. Theory* 26 (1) (1980) 26–37.
- [67] W. Bialek, *Biophysics: Searching for Principles*, Princeton University Press, 2012, Chapter 6.
- [68] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 2014.
- [69] H.C. Nguyen, R. Zecchina, J. Berg, Inverse statistical problems: from the inverse Ising problem to data science, *Adv. Phys.* 66 (3) (2017) 197–261.
- [70] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 2016.
- [71] G. Varoquaux, R.C. Craddock, Learning and comparing functional connectomes across subjects, *NeuroImage* 80 (2013) 405–415.
- [72] A.L. Berger, V.J.D. Pietra, S.A.D. Pietra, A maximum entropy approach to natural language processing, *Comput. Linguist.* 22 (1) (1996) 39–71.
- [73] G.J. Stephens, W. Bialek, Statistical mechanics of letters in words, *Phys. Rev. E* 81 (6) (2010) 066119.
- [74] J.L. Muñoz-Cobo, R. Mendizábal, A. Miquel, C. Berna, A. Escrivá, Use of the Principles of Maximum Entropy and Maximum Relative Entropy

- for the Determination of Uncertain Parameter Distributions in Engineering Applications, *Entropy* 19 (9) (2017) 486.
- [75] P. Wang, L. Robert, J. Pelletier, W.L. Dang, F. Taddei, A. Wright, S. Jun, Robust growth of *Escherichia coli*, *Curr. Biol.* 20 (12) (2010) 1099–1103.
- [76] Y. Tanouchi, A. Pai, H. Park, S. Huang, N.E. Buchler, L. You, Long-term growth data of *Escherichia coli* at a single-cell level, *Sci. Data* 4 (2017) 170036.
- [77] D.J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova, S.J. Tans, Stochasticity of metabolism and growth at the single-cell level, *Nature* 514 (7522) (2014) 376–379.
- [78] M. Wallden, D. Fange, E.G. Lundius, Ö. Baltekin, J. Elf, The synchronization of replication and division cycles in individual *E. coli* cells, *Cell* 166 (3) (2016) 729–739.
- [79] A.S. Kennard, M. Osella, A. Javer, J. Grilli, P. Nghe, S.J. Tans, P. Cicuta, M.C. Lagomarsino, Individuality and universality in the growth-division laws of single *E. coli* cells, *Phys. Rev. E* 93 (1) (2016) 012408.
- [80] F. Si, D. Li, S.E. Cox, J.T. Sauls, O. Azizi, C. Sou, A.B. Schwartz, M.J. Erickstad, Y. Jun, X. Li, S. Jun, Invariance of initiation mass and predictability of cell size in *Escherichia coli*, *Curr. Biol.* 27 (9) (2017) 1278–1287.
- [81] S. Jun, S. Taheri-Araghi, Cell-size maintenance: universal strategy revealed, *Trends Microbiol.* 23 (1) (2015) 4–6.
- [82] M. Zampieri, K. Sekar, N. Zamboni, U. Sauer, Frontiers of high-throughput metabolomics, *Curr. Opin. Chem. Biol.* 36 (2017) 15–23.
- [83] I. Thiele, B.O. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction, *Nat. Protoc.* 5 (1) (2010) 93–121.
- [84] J. Schellenberger, R. Que, R.M. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, J. Kang, Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0, *Nat. Protoc.* 6 (9) (2011) 1290–1307.
- [85] A. Bordbar, J.M. Monk, Z.A. King, B.O. Palsson, Constraint-based models predict metabolic and associated cellular functions, *Nat. Rev. Genet.* 15 (2) (2014) 107–120.
- [86] J.L. Reed, Shrinking the metabolic solution space using experimental datasets, *PLoS Comput. Biol.* 8 (2012) e1002662.
- [87] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley & Sons, 1998.

- [88] D. De Martino, M. Mori, V. Parisi, Uniform sampling of steady states in metabolic networks: heterogeneous scales and rounding, *PLoS ONE* 10 (4) (2015) e0122670.
- [89] H.S. Haraldsdottir, B. Cousins, I. Thiele, R.M. Fleming, S. Vempala, CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models, *Bioinformatics* 33 (11) (2017) 1741–1743.
- [90] J.L. Reed, T.D. Vo, C.H. Schilling, B.O. Palsson, An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR), *Genome Biol.* 4 (9) (2003) R54.
- [91] D. De Martino, F. Capuani, A. De Martino, Growth against entropy in bacterial metabolism: the phenotypic trade-off behind empirical growth rate distributions in *E. coli*, *Phys. Biol.* 13 (2016) 036005.
- [92] D. De Martino, A. Andersson, T. Bergmiller, C.C. Guet, G. Tkacik, Statistical mechanics for metabolic networks during steady-state growth, preprint, arXiv: 1703.01818, 2017.
- [93] D. De Martino, D. Masoero, Asymptotic analysis of noisy fitness maximization, applied to metabolism & growth, *J. Stat. Mech.* 2016 (12) (2016) 123502.
- [94] D. De Martino, F. Capuani, A. De Martino, Quantifying the entropic cost of cellular growth control, *Phys. Rev. E* 96 (2017), 010401(R).
- [95] D. De Martino, A. De Martino, Constraint-based inverse modeling of metabolic networks: a proof of concept, preprint, arXiv:1704.08087, 2017.
- [96] A.P. Burgard, C.D. Maranas, Optimization-based framework for inferring and testing hypothesized metabolic objective functions, *Biotechnol. Bioeng.* 82 (6) (2003) 670–677.
- [97] A.L. Knorr, R. Jain, R. Srivastava, Bayesian-based selection of metabolic objective functions, *Bioinformatics* 23 (3) (2006) 351–357.
- [98] J. Heino, K. Tunyan, D. Calvetti, E. Somersalo, Bayesian flux balance analysis applied to a skeletal muscle metabolic model, *J. Theor. Biol.* 248 (1) (2007) 91–110.
- [99] H.C. Chiu, D. Segrè, Comparative determination of biomass composition in differentially active metabolic States, *Genome Inform. Ser.* 20 (2008) 171–182.
- [100] E.P. Gianchandani, M.A. Oberhardt, A.P. Burgard, C.D. Maranas, J.A. Papin, Predicting biological system objectives de novo from internal state measurements, *BMC Bioinform.* 9 (1) (2008) 43.

- [101] C. Martelli, A. De Martino, E. Marinari, M. Marsili, I.P. Castillo, Identifying essential genes in *Escherichia coli* from a metabolic optimization principle, *Proc. Natl. Acad. Sci.* 106 (8) (2009) 2607–2611.
- [102] Q. Zhao, H. Kurata, Maximum entropy decomposition of flux distribution at steady state to elementary modes, *J. Biosci. Bioeng.* 107 (1) (2009) 84–89.
- [103] Q. Zhao, H. Kurata, Use of maximum entropy principle with Lagrange multipliers extends the feasibility of elementary mode analysis, *J. Biosci. Bioeng.* 110 (2) (2010) 254–261.
- [104] D. De Martino, M. Figliuzzi, A. De Martino, E. Marinari, A scalable algorithm to explore the Gibbs energy landscape of genome-scale metabolic networks, *PLoS Comput. Biol.* 8 (6) (2012) e1002562.
- [105] F.A. Massucci, M. DiNuzzo, F. Giove, B. Maraviglia, I.P. Castillo, E. Marinari, A. De Martino, Energy metabolism and glutamate–glutamine cycle in the brain: a stoichiometric modeling perspective, *BMC Syst. Biol.* 7 (1) (2013) 103.
- [106] F.A. Massucci, F. Font-Clos, A. De Martino, I.P. Castillo, A novel methodology to estimate metabolic flux distributions in constraint-based models, *Metabolites* 3 (3) (2013) 838–852.
- [107] D. De Martino, F. Capuani, A. De Martino, Inferring metabolic phenotypes from the exometabolome through a thermodynamic variational principle, *New J. Phys.* 16 (11) (2014) 115018.
- [108] Q. Zhao, A.I. Stettner, E. Reznik, I.C. Paschalidis, D. Segré, Mapping the landscape of metabolic goals of a cell, *Genome Biol.* 17 (1) (2016) 109.
- [109] A. Braunstein, A.P. Muntoni, A. Pagnani, An analytic approximation of the feasible space of metabolic networks, *Nat. Commun.* 8 (2017) 14915.
- [110] J.D. Orth, I. Thiele, B.O. Palsson, What is flux balance analysis? *Nat. Biotechnol.* 28 (3) (2010) 245–248.
- [111] M. Imielinski, C. Belta, H. Rubin, A. Halasz, Systematic analysis of conservation relations in *Escherichia coli* genome-scale metabolic network reveals novel growth media, *Biophys. J.* 90 (8) (2006) 2659–2672.
- [112] J. Heino, K. Tunyan, D. Calvetti, E. Somersalo, Bayesian flux balance analysis applied to a skeletal muscle metabolic model, *J. Theor. Biol.* 248 (1) (2007) 91–110.
- [113] A. De Martino, C. Martelli, R. Monasson, I.P. Castillo, Von Neumann’s expanding model on random graphs, *J. Stat. Mech. Theory Exp.* 2007 (05) (2007) P05012.

- [114] L. Merchan, I. Nemenman, On the sufficiency of pairwise interactions in maximum entropy models of networks, *J. Stat. Phys.* 162 (5) (2016) 1294–1308.
- [115] D. De Martino, Maximum entropy modeling of metabolic networks by constraining growth rate moments predicts co-existence of phenotypes, *Phys. Rev. E* 96 (2017) 060401.
- [116] J.R. Banavar, A. Maritan, I. Volkov, Applications of the principle of maximum entropy: from physics to ecology, *J. Phys. Condens. Matter* 22 (6) (2010) 063101.
- [117] E.T. Jaynes, H. Haken, *Complex Systems: Operational Approaches*, Springer Ser. Synergetics, Springer, Berlin, Heidelberg, New York, 1985.
- [118] K. Ghosh, K.A. Dill, M.M. Inamdar, E. Seitaridou, R. Phillips, Teaching the principles of statistical dynamics, *Am. J. Phys.* 74 (2) (2006) 123–133.
- [119] S. Pressé, K. Ghosh, K.A. Dill, Modeling stochastic dynamics in biochemical systems with feedback using maximum caliber, *J. Phys. Chem. B* 115 (19) (2011) 6202–6212.
- [120] O. Marre, S. El Boustani, Y. Frégnac, A. Destexhe, Prediction of spatiotemporal patterns of neural activity from pairwise correlations, *Phys. Rev. Lett.* 102 (13) (2009) 138101.
- [121] J. Lee, S. Pressé, A derivation of the master equation from path entropy maximization, *J. Chem. Phys.* 137 (7) (2012) 074103.
- [122] H. Ge, S. Pressé, K. Ghosh, K.A. Dill, Markov processes follow from the principle of maximum caliber, *J. Chem. Phys.* 136 (6) (2012) 064108.
- [123] H. Nasser, B. Cessac, Parameter estimation for spatio-temporal maximum entropy distributions: Application to neural spike trains, *Entropy* 16 (4) (2014) 2244–2277.
- [124] L.R. Mead, N. Papanicolaou, Maximum entropy in the problem of moments, *J. Math. Phys.* 25 (8) (1984) 2404–2417.
- [125] C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer Science & Business Media, 2007.
- [126] P. Cheeseman, J. Stutz, On the relationship between bayesian and maximum entropy inference, *AIP Conf. Proc.* 735 (1) (2004) 445–461.
- [127] S.M. Stigler, The epic story of maximum likelihood, *Stat. Sci.* 22 (2007) 598–620.

- [128] S. Olsson, J. Frellsen, W. Boomsma, K.V. Mardia, T. Hamelryck, Inference of structure ensembles of flexible biomolecules from sparse, averaged data, *PLoS ONE* 8 (2013) e79439.
- [129] M. Habeck, Bayesian approach to inverse statistical mechanics, *Phys. Rev. E* 89 (2014) 052113.
- [130] S. Olsson, H. Wu, F. Paul, C. Clementi, F. Noé, Combining experimental and simulation data of molecular processes via augmented Markov models, *Proc. Natl. Acad. Sci.* 114 (2017) 8265–8270.
- [131] S.J. Phillips, M. Dudik, R.E. Schapire, A maximum entropy approach to species distribution modeling, in: *Proceedings of the 21st International Conference on Machine Learning*, ACM, 2004, p. 83.
- [132] R.J. Little, D.B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, 2014.
- [133] S. Wang, D. Schuurmans, Y. Zhao, The latent maximum entropy principle, *ACM Trans. Knowl. Discov. Data* 6 (2012) 8.
- [134] B. Huang, A. Salleb-Aouissi, Maximum entropy density estimation with incomplete presence-only data, in: *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, 2009, pp. 240–247.
- [135] G. Corani, C.P. de Campos, A maximum entropy approach to learn Bayesian networks from incomplete data, in: *Interdisciplinary Bayesian Statistics*, Springer, 2015, pp. 69–82.
- [136] M. Scott, T. Hwa, Bacterial growth laws and their applications, *Curr. Opin. Biotechnol.* 22 (4) (2011) 59–565.
- [137] W. Feller, *An Introduction to Probability Theory and Its Applications: volume II*, John Wiley & Sons, New York, 1968, Chapters VII and VIII.
- [138] T. Obuchi, R. Monasson, Learning probability distributions from smooth observables and the maximum entropy principle: some remarks, *J. Phys. Conf. Ser.* 638 (1) (2015) 012018.
- [139] T. Obuchi, S. Cocco, R. Monasson, Learning probabilities from random observables in high dimensions: the maximum entropy distribution and others, *J. Stat. Phys.* 161 (3) (2015) 598–632.
- [140] E. van Nimwegen, Inferring contacting residues within and between proteins: what do the probabilities mean? *PLoS Comput. Biol.* 12 (5) (2016) e1004726.
- [141] E. Aurell, The maximum entropy fallacy redux? *PLoS Comput. Biol.* 12 (5) (2016) e1004777.

- [142] A. Bejan, S. Lorente, The constructal law and the evolution of design in nature, *Phys. Life Rev.* 8 (2011) 209–240.
- [143] S. Grigolon, S. Franz, M. Marsili, Identifying relevant positions in proteins by critical variable selection, *Mol. BioSyst.* 12 (7) (2016) 2147–2158.
- [144] L. Gresele, M. Marsili, On maximum entropy and inference, *Entropy* 19 (12) (2017) 642.