Maximum entropy approach to characterization of random media

E. V. Vakarin, J. P. Badiali

Laboratoire d'Electrochimie et de Chimie Analytique UMR 7575 CNRS-ENSCP-Universite Pierre et Marie Curie, 4 Place Jussieu, 75252 Cedex 05, Paris, France

Abstract. A maximum entropy approach to estimation of statistical properties of random media through indirect probes is analyzed. It is shown that there is an explicit link between the probe response to an external driving and the information rate. This allows one to relate the response specificities (singularities or inflection points) to the amount of information on the surrounding medium. Advantages of the maximum entropy approach in recovering the power-law distributions under physically justified constraints are demonstrated.

Key Words: Maximum Entropy Inference, Random Media, Power-Law Distribution.

INTRODUCTION

Characterization of disordered media through indirect probes is one of the methods being extensively developed in the context of adsorption into porous matrices [1] or on heterogeneous surfaces [2]. Essentially similar problems appear in various fields, from sociology to astrophysics. In a quite broad sense a random medium can be represented by a probability distribution $f(\sigma)$ of some relevant quantity, σ. The distribution is *a priori* unknown. In order to determine it, the medium is coupled to a system with a known response function $\theta(\mu|\sigma)$, conditional to the medium state σ. In other words, we deal with a driven deterministic (in terms of statistical mechanics or thermodynamics) system in contact with a fluctuating background. The system relaxation to an equilibrium (or stationary) state is usually much faster than the medium fluctuations. Then the experimental data $\theta(\mu)$ can be represented as an average over the medium (background) fluctuations. For concreteness we consider a continuous spectrum of σ

$$\theta(\mu) = \overline{\theta(\mu|\sigma)} = \int d\sigma f(\sigma)\theta(\mu|\sigma) \tag{1}$$

where μ is the driving parameter and the overline denotes the averaging. Adsorption in porous matrices could serve as a prototypic example, where $\theta(\mu)$ is an adsorption isotherm, $f(\sigma)$ is a pore size distribution and the driving parameter μ is the adsorbate chemical potential or pressure.

Therefore, one deals with an inverse problem of extracting $f(\sigma)$ from $\theta(\mu)$. Several theoretical schemes have been developed in this context. One of the simplest approaches is to assume a given functional form of $f(\sigma)$ with a set of adjustable parameters [2, 3, 4], which can be found from a reasonable fit to the data through eq. (1). Despite of its

conceptual simplicity a predictive power of this technique is quite low, because there is no systematic link between the input functions $\theta(\mu)$, $\theta(\mu|\sigma)$ and a large number of fitting parameters. Nevertheless, in some cases [4] a fitting to inflection points in $\theta(\mu)$ gives quite stable and reasonable predictions. Assuming that the kernel $\theta(\mu|\sigma)$ is Langmuirian and that the data $\theta(\mu)$ are known as an analytic function, Sips has shown [5] that, in the framework of eq. (1), $\theta(\mu)$ is a Stieltjes transform of $f(\sigma)$. This gives a straightforward evaluation of the distribution, provided that the transformation exists, but the method is bound to a quite limited set of the kernels $\theta(\mu|\sigma)$. Another set of methods treats eq. (1) numerically, as an integral equation for $f(\sigma)$. In general, this is an ill-posed problem, that may have no or many solutions depending on a combination of $\theta(\mu)$ and $\theta(\mu|\sigma)$. On the other hand, the solution is inavoidably affected by a noise (e.g., due to experimental errors) in $\theta(\mu)$. Such that small variations in the data can produce large errors in the solution. In order to overcome this problem one usually invokes additional conditions derived from some prior knowledge on the solution. One of the widespread procedures is the linear regularization [6] which involves damping parameters and, thus, subjectively deforms the solution. In addition, this method requires the data $\theta(\mu)$ for a very large range of the drivings μ, that is not always accessible experimentally. A maximization of information entropy (MaxEnt) [7] is an alternative approach to retrieving a distribution from the constraint (1). This is an inductive inference method, based on a Bayesian interpretation of probabilities [8], that allows to make predictions under conditions of incomplete information [9]. Despite of numerous applications of these methods in a variety of domains [9, 10, 11, 12, 13, 14] some key features still remain poorly understood. This generates a series of questions specific to the MaxEnt approach as well as of general importance for the characterization problem.

THE SCOPE

In this paper we are attempting to address some of these questions. Namely, why in some cases the MaxEnt procedure requires fewer data [10] than other (e.g. regularization) approaches and what are the conditions for that. What is the information content of the data and how does it permit [1, 2] to estimate the distribution characteristics even without entering into fine details of the kernel $\theta(\mu|\sigma)$. What is an optimal choice of the kernel. In particular, why a fitting to inflection points in the data $\theta(\mu)$ gives reasonable predictions [4] and why a derivative kernel

$$\chi(\mu) = \frac{\partial\theta}{\partial\mu} = \int d\sigma f(\sigma) \frac{\partial\theta(\mu|\sigma)}{\partial\mu} \tag{2}$$

is usually preferable [6] over the constraint (1). Finally we demonstrate the efficiency of the MaxEnt approach in combination with the concept of superstatistics [15, 16] in explaining the ubiquity of power-law distributions in natural systems.

CONDITIONAL DISTRIBUTION

First of all let us note that whatever method of retrieving the distribution from the data $\theta(\mu)$ through eq. (1) should give a conditional distribution $f(\sigma) = \varphi(\sigma|\mu)$. That is, the distribution is conditional to the driving μ. This has important consequences. Having found the distribution we can use it to predict averages of other σ-dependent quantities, like $\chi(\mu|\sigma) = \partial\theta(\mu|\sigma)/\partial\mu$

$$\chi_p(\mu) = \int d\sigma \varphi(\sigma|\mu)\chi(\mu|\sigma) \tag{3}$$

On the other hand, $\varphi(\sigma|\mu)$ is found from the condition that eq. (1) becomes an identity. Such that, differentiating the data $\chi(\mu) = \partial\theta(\mu)/\partial\mu$, we obtain the experimental counterpart as

$$\chi(\mu) = \chi_p(\mu) + \int d\sigma \frac{\partial\varphi(\sigma|\mu)}{\partial\mu}\theta(\mu|\sigma) \tag{4}$$

Therefore, the observed response differs from the predicted one, because the distribution changes with the driving μ. Note that this is a general result, independent of the method used to retrieve the distribution from eq. (1). In what follows we will show that the MaxEnt inference procedure can give a deep insight on the origin and the meaning of such a disagreement.

MAXIMUM ENTROPY APPROACH

In this paper we deal with the inversion of eq. (1) in the framework of the maximum-entropy inference scheme, developed by Jaynes [7, 8]. We assume that no prior information on the medium is available. Our uncertainty on the background state can be estimated by an information entropy, which is taken here in the Shannon form

$$H = -\int d\sigma f(\sigma)\ln[f(\sigma)] \tag{5}$$

Maximizing H under the constraint (1) and requiring the normalization for $f(\sigma)$ we get the following conditional distribution $f(\sigma) = f(\sigma|\mu)$ [14]

$$f(\sigma|\mu) = \frac{e^{\kappa\theta(\mu|\sigma)}}{Z}; \qquad Z = \int d\sigma e^{\kappa\theta(\mu|\sigma)} \tag{6}$$

where the Lagrange multiplier κ should be found from the constraint (1). Note that $f(\sigma|\mu)$ should be considered as an *aposteriori* distribution because it is based on the observation $\theta(\mu)$. Therefore, plugging the distribution (6) back to (5) we obtain the amount of uncertainty $H(\mu)$ on the medium state

$$H(\mu) = -\kappa\theta(\mu) + \ln Z$$

If the initial (before the subsystems contact) uncertainty is H_0, then the amount of information [9] $I(\mu)$ on the medium, one can get by driving (e.g. through varying μ)

the deterministic subsystem is given by a reduction in the uncertainty

$$I(\mu) = H_0 - H(\mu)$$

Differentiating, we obtain the information rate

$$\frac{\partial I(\mu)}{\partial \mu} = \kappa \int d\sigma \frac{\partial f(\sigma|\mu)}{\partial \mu} \theta(\mu|\sigma) \tag{7}$$

Combining this with an analog of eq. (4) we arrive at a remarkably simple and extremely useful result

$$\frac{\partial I(\mu)}{\partial \mu} = \kappa \left[\chi(\mu) - \chi_p(\mu)\right] \tag{8}$$

which gives an explicit link among the information rate, the observed response function $\chi(\mu)$ and the model estimation $\chi_p(\mu) = \overline{\partial\theta(\mu|\sigma)/\partial\mu}$. This allows us to analyze the impacts of these ingredients to the information on the medium. In particular, requiring the consistency in (4) : $\chi(\mu) = \chi_p(\mu)$ is equivalent to $\partial I/\partial\mu = 0$, that is, to a necessary condition for a maximum of $I(\mu)$. Therefore, in order to extract a maximum of information at all μ (for given data $\theta(\mu)$ and the model estimation $\theta(\mu|\sigma)$) one has to require $\chi(\mu) = \chi_p(\mu)$, which is equivalent to imposing the differential kernel constraint (2). However, for practical purposes this strategy is limited if the data $\theta(\mu)$ contain a noise (e.g., because of experimental errors). In that case the data differentiation enhances the noise contribution, that, depending on its intensity and frequency, might lead to spurious reflexes in the distribution.

Our result (8) also gives a useful insight on how to choose $\theta(\mu|\sigma)$. Of course, one may try to guess at random until a satisfactory agreement between $\chi(\mu)$ and $\chi_p(\mu)$ is achieved. A more constructive way is to satisfy the condition $\chi(\mu) = \chi_p(\mu)$ locally, i.e. on a set of characteristic points $\{\mu_k\}$. Then it is clear that the easiest way to do that is to focus on the points μ_k where the condition $\partial I/\partial\mu|_{\mu=\mu_k} = 0$ is satisfied trivially, because both $\chi(\mu_k) = 0$ and $\chi_p(\mu_k) = 0$. The latter is a necessary condition for a plateau or an inflection point in $\theta(\mu)$ and $\theta(\mu|\sigma)$. This explains why a fitting to the inflection points [4] allows to make reasonable predictions. Note that the specificities (steps, kinks, inflections, plateaus, etc.) of the data $\theta(\mu)$ are attributed to some physics. In the context of adsorption probes [1, 2, 4] these peculiarities reflect different adsorption or packing regimes [2, 4]) with some microscopic picture behind. Therefore, it is remarkable that our result (8) translates this physics directly into the required information.

On the other hand, it is not always possible to work under conditions close to a maximum of $I(\mu)$. The plateau in the data could be absent or hardly reachable and the model $\theta(\mu|\sigma)$ is not always exhaustive. In that case it is desirable to find a way of estimating the distribution characteristics in a limited range of drivings μ and in the absence of a detailed microscopic model $\theta(\mu|\sigma)$. From eq. (8) we see that $I(\mu)$ undergoes remarkable changes in a range of μ where $\chi(\mu)$ has a strong maximum ($\theta(\mu)$ makes a step). In that case even a rough estimation of $\theta(\mu|\sigma)$ with a non-singular $\chi_p(\mu)$ is sufficient for a rapid evaluation of at least some background characteristics. Then a limited number of points (limited range of μ) around the data singularity could be enough for retrieving a correct overall picture. Namely this trick is applied for characterizing porous media [1] or heterogeneous surfaces [2] through adsorption probes. The technique consists in a driving

an adsorbate close to the condensation or freezing point (in that case the compressibility $\chi(\mu)$ diverges) which are sensitive to the mean porosity or heterogeneity. It is worth noting again that the MaxEnt result (8) is capable of a straightforward explanation of this experimental fact from a quite general point of view, even without entering into the physics of these phase transitions.

POWER-LAW DISTRIBUTIONS

Despite of the apparently exponential form (6) the actual behavior of $f(\sigma|\mu)$ depends on the nature of the constraint imposed and on a form of the constrained function $\theta(\mu|\sigma)$. Such an ambiguity should not be considered as a shortcoming of the theory. This is a consequence of the fact that we are working under conditions of incomplete information. Then, according to the Bayesian interpretation [7, 8], a probability should be considered as a measure of our ignorance rather than an objective property. Nevertheless, our freedom in choosing the constraints is restricted by the nature of the problem under consideration. In other words, there are constraints which are "naturally" imposed either as design principles [17, 18] or as experimental conditions. In the context of adsorption in porous media [1] the constraint (1) reflects the experimental conditions of measuring the adsorption isotherm. From another point of view $\theta(\mu)$ could be considered as a "target" function [19]. In this case the problem is to find a distribution leading to a given average response $\theta(\mu)$. In particular, in many applications (such as optimal control [20]) it is desirable to constraint the system internal order with the purpose of meeting some survival or functionality objectives. Then it is natural to restrict the phase space by constraining the thermodynamic entropy $S(\rho|\beta)$ (ρ is the number density, β is the inverse temperature). This idea was shortly discussed in a slightly different context [14, 17, 18, 19]. Therefore, in our previous treatment we replace $\theta(\mu|\sigma)$ by $S(\rho|\beta)$, and $\theta(\mu)$ by $\Sigma(\rho)$. Then the distribution (6) closely resembles the Einstein fluctuation formula. Note however that $f(\beta)$ describes the background fluctuations and for $\kappa = 1$ it becomes identical to the distribution of the system fluctuations. In particular, for small fluctuations around an equilibrium state (ρ, β^*) we may expand

$$S(\rho|\beta) = S(\rho|\beta^*) - \frac{1}{2\chi(\rho,\beta^*)}(\beta - \beta^*)^2 \tag{9}$$

where $\chi(\rho,\beta^*) = \langle(\beta - \beta^*)^2\rangle$ is the mean-square fluctuation in the system when the background state is fixed at $\beta = \beta^*$. In this approximation the distribution (6) becomes gaussian and κ can be determined combining (9) and (1). Finally we arrive at

$$\overline{(\beta - \beta^*)^2} = \langle(\beta - \beta^*)^2\rangle\,[S(\rho|\beta^*) - \Sigma(\rho)] \tag{10}$$

Note that $S(\rho|\beta^*)$ is the system equilibrium entropy. Consequently $S(\rho|\beta^*) - \Sigma(\rho) \geq 0$ and $\overline{(\beta - \beta^*)^2} \geq 0$ as it should be. Therefore, in order to ensure a given response, $\Sigma(\rho)$, the background should fluctuate coherently with the system fluctuations and with the distance from the equilibrium state. As we will see below, for large fluctuations this tendency also takes place. A quite similar trend has been reported [21] for fluctuations

in the Tsallis statistics [22]. In the limit of $S(\rho|\beta^*) = \Sigma(\rho)$ we return to the standard equilibrium without fluctuations in the background: $f(\beta) = \delta(\beta - \beta^*)$. The system fluctuates according to its response function $\chi(\rho, \beta^*)$.

In order to study large background fluctuations and the system statistics at different time scales we have to introduce an explicit form for $S(\rho|\beta)$. With this purpose we consider an exactly solvable toy model – the ideal gas in contact with a reservoir of fluctuating temperature [15, 16]. This choice is motivated by our goal to extract the most general and essential features, independent of approximations or the system correlations. Therefore, we deal with the entropy per particle $S(\rho|\beta) = 5/2 - \ln(\rho\Lambda^3)$. Here ρ is the number density, $\beta = 1/kT$ is the inverse temperature and Λ is the thermal de Broglie length. Introducing irrelevant scaling constants (making ρ and β dimensionless), $S(\rho|\beta)$ can be reduced to

$$S(\rho|\beta) = const - \ln(\rho) - \frac{3}{2}\ln(\beta) \tag{11}$$

Constraining the average temperature

$$\beta_0 = \int d\beta f(\beta)\beta \tag{12}$$

and the entropy

$$\Sigma(\rho) = \int d\beta f(\beta) S(\rho|\beta) \tag{13}$$

through the inference procedure discussed above, we obtain the following distribution

$$f(\beta) = \frac{\beta^\kappa e^{-\beta/\beta(\kappa)}}{[\beta(\kappa)]^{\kappa+1}\Gamma(\kappa+1)}; \quad \beta(\kappa) = \frac{\beta_0}{\kappa+1} \tag{14}$$

which is precisely the Γ-distribution considered in the superstatistical approach [15, 16], where the exponent κ is related to the noise intensity. In our case the Lagrange multiplier κ should be determined from the entropy constraint (13)

$$\Sigma(\rho) = S(\rho|\beta_0) + \ln(\kappa+1) - \Psi(\kappa) - \frac{1}{\kappa} \tag{15}$$

where $\Psi(\kappa) = d\ln\Gamma(\kappa)/d\kappa$. Thus, the exponent κ is determined by the distance $\Sigma(\rho) - S(\rho|\beta_0)$ from the equilibrium state (ρ, β_0). In particular, for large κ we have found

$$\Sigma(\rho) = S(\rho|\beta_0) + 1/(2\kappa)$$

Therefore, κ is related to the deviation from the standard equilibrium, such that $f(\beta) \to \delta(\beta - \beta_0)$ and $\Sigma(\rho) = S(\rho|\beta_0)$ as $\kappa \to \infty$.

At the short-time scale (of the order of the system relaxation time) the conditional energy distribution ($E = p^2/2m$, at a given temperature β) is Gibbsian

$$f(E|\beta) = \frac{e^{-\beta E}}{\int dE e^{-\beta E}} \tag{16}$$

The long-time behavior of the dynamic system can be represented as a superposition of its short-time statistics and the background fluctuations – the superstatistical approach

[15, 16]. In this spirit the long-time energy distribution can be found by averaging over the temperature fluctuations

$$f(E) = \int d\beta f(\beta) f(E|\beta) = \beta_0 \left[1 - \frac{\beta_0 E}{1-q}\right]^{-q} \tag{17}$$

where $q = \kappa + 2$. For $E = p^2/2m$ we recover the power-law velocity distribution found [15, 16] in the superstatisical approach. Quite similar effects have recently been predicted to occur in driven dissipative inelastic gases [23] and in driven stochastic systems with multiplicative noise [24] or fluctuating mass [25]. In a different context similar power laws where found, applying the maximum-entropy inference to parameterized entropies (Tsallis [22], and Renyi [12]). But the meaning of the entropic parameter is not always clear, while in our case q is directly related to the constraint imposed.

Thus, because of the constraint, imposed on the internal order at longer times the system develops avalanches (or energy cascades) at any finite κ. The avalanche size is characterized by $1/\kappa$. Therefore, maintaining the distance from the equilibrium, one can control the size of the rare "catastrophic" events. This can be organized in different ways, such as by powerful energy injections at large velocity scales [23], or through an interplay of additive and multiplicative noises [24].

CONCLUSION

It is shown that the MaxEnt inference approach establishes an explicit link between the amount information on a random medium and a difference of the probe response function and its model estimation. This allows to translate the response specificities into the information rate without entering into the microscopic details determining the probe behavior.

We have found that the major ingredients, relevant to the power-law distributions in composite systems are (i) the widely separated times scales, (ii) non-vanishing background fluctuations. (iii) a constraint, imposed on the overall system, holding the dynamic counterpart in a stationary non-equilibrium state.

The main advantage of the approach developed here is that it avoids parameterized entropy measures and allows one to apply the superstatistical scheme to coupled systems in which the stochastic background does not fluctuate independently of the dynamic counterpart.

In the context of our study the nonextensitivy in systems with power-law distributions is a direct consequence of the "global" nature of the constraint. This makes it impossible to decompose the system into non-correlated parts.

REFERENCES

1. L. D. Gelb, K. E. Gubbins, R. Radhakrishnan, M. Sliwinska-Bartowiak, Rep. Prog. Phys. **62**, 1573 (1999)
2. M. Jaroniec, P. Brauer, Surf. Sci. Rep. **6**, 65 (1986)

3. S. Scaife, P. Kluson, N. Quirke, J. Phys. Chem. B **104**, 313 (2000)
4. D. Dubbeldam, S. Calero, T. J. H. Vlugt, R. Krishna, T. L. M. Maesen, E. Beerdsen, B. Smit, Phys. Rev. Lett. **93**, 088302 (2004)
5. R. Sips, J. Chem. Phys. **16**, 490 (1948); **18**, 1024 (1950)
6. F. Dion, A. Lasia, J. Electroanal. Chem. **475**, 28 (1999)
7. E. T. Jaynes, Phys. Rev. **106**, 620 (1957)
8. E. T. Jaynes, Probability Theory. (Cambridge University Press, 2003)
9. H. Haken, Information and Self-Organization. A Macroscopic Approach to Complex Systems. (Springer-Verlag Berlin, 1988)
10. J. L. Garces, F. Mas, J. Puy, Environ. Sci. Technol. **32**, 539 (1998)
11. T. Horlin, Solid State Ionics **107**, 241 (1998)
12. A. G. Bashkirov, Phys. Rev. Lett. **93**, 130601 (2004)
13. F. Sattin, Phys. Rev. E **68**, 032102 (2003)
14. E. V. Vakarin, J. P. Badiali, Surf. Sci. **565**, 279 (2004);
    E. V. Vakarin, J. P. Badiali, Electrochim. Acta **50**, 1719 (2005)
15. C. Beck, Phys. Rev. Lett. **87**, 180601 (2001)
16. C. Beck, E. G. D. Cohen, Physica A **322**, 267 (2003)
17. Y. Dover, Physica A **334**, 591 (2004)
18. V. Venkatasubramanian, D. N. Politis, P. R. Patkar, AIChE Journal **52**, 1004 (2006)
19. E. V. Vakarin, J. P. Badiali, Phys. Rev. E **74** in press (2006)
20. H. Touchette, S. Lloyd, Phys. Rev. Lett. **84**, 1156 (2000)
21. E. Vives, A. Planes, Phys. Rev. Lett. **88**, 020601 (2002)
22. C. Tsallis, J. Stat. Phys., **52**, 479 (1988)
23. E. Ben-Naim, J. Machta, Phys. Rev. Lett. **94**, 138001 (2005)
24. T. S. Biro, A. Jakovac, Phys. Rev. Lett. **94**, 132302 (2005)
25. M. Ausloos, R. Lambiotte, Phys. Rev. E **73**, 011105 (2006)