

Entropy minimization within the maximum entropy approach

E.V. Vakarin, J.P. Badiali

IRCP UMR 8247 CNRS / ENSCP Chimie ParisTech, 11 rue P. et M. Curie, 75231 Cedex 05, Paris, France

E-mail: eduard.vakarin@upmc.fr

Abstract. A procedure of the minimization of the maximum entropy with respect to an external driving force μ is proposed. In application to coupled dynamic-stochastic systems such an approach allows one to reduce the uncertainty in estimating the dynamic counterpart (a probe or a model) parameters. In its turn this permits to estimate an optimal driving path giving a maximal information on the probability distribution $f(x|\mu)$ of the stochastic counterpart with a given probe/model $\theta(\mu|x)$. It is found that the functional form of the model should be similar to the observed/measured one $\theta(\mu)$, while the minimum uncertainty is reached when the distribution becomes independent of the driving μ .

1. Introduction

The term *Complex Systems* has emerged as a recognized field of research in many domains (from physics to biology). Nevertheless a strict definition of the complexity is still lacking. However there are several common generic features such as multicomponent nature, emergence, non-linear or even feedback relations among the components. In this situation one has to work under conditions of incomplete or partial information on the system. Thus a deterministic description becomes almost impossible and one has to resort to statistical/probabilistic methods. One of them is the maximum entropy approach[1] which is extensively invoked in the solution of inference problems appearing in various fields . The approach allows one to calculate a probability distribution of relevant variables from an incomplete knowledge of the microscopic mechanisms governing the system evolution. The central idea is to find a least biased distribution consistent with the information at hand. Therefore, the scheme operates with two main ingredients: the entropy measure (that estimates the uncertainty associated with the probability distribution) and the constraints encoding the available information. In this context various entropy functionals are introduced (see [2] for a recent review) in order to derive various non-exponential probability distributions observed in nature. These are generalizations of the famous Shannon form (4) that appears from the generalized forms as some limiting case. However, many of these generalized entropies are non-additive for statistically independent subsystems. This fact with its consequences is still a subject of extensive debates [2, 3].

In a situation when many generalized entropy forms are being proposed [4] it might seem that the maximum entropy approach suffers from an ambiguity [5]: "in a sense that any probability distribution seems to be derivable from the maximization of any entropic measure if an appropriate constraint is used". In this respect it has been argued [5] that such an ambiguity



does not appear when one searches for a parametric family of distributions $f(x|\mu)$. Here x is a relevant fluctuating variable and μ is an external parameter independent of x . For instance, such a parameter could enter through an observed mean value $\theta(\mu)$ [6] or the distribution width varying in response to some external perturbation μ , like in the process of insertion into a host matrix [7]. On the other hand, we have to note that we are not free to modify the constraints without logical necessity because they contain the information (or our theoretical model) on the processes running in the system of interest.

Since we deal with a parametrized distribution $f(x|\mu)$ its entropy $H(\mu)$ is also a function of the parameter μ . This opens up a possibility to search for a minimum of $H(\mu)$ as a function of the driving μ . In other words, with a given observable $\theta(\mu)$ and its model estimation $\theta(\mu|x)$ we may try to find an "optimal" set of parameters reducing the uncertainty on the behavior of $f(x|\mu)$. This idea was briefly sketched in [8].

In this paper we explore such a possibility in application to coupled dynamic-stochastic systems [9, 10, 11], where the dynamic counterpart may serve as a probe for determining the behavior $f(x|\mu)$ of the stochastic subsystem. In this way the maximum entropy approach allows one to find a functional form of the distribution $f(x|\mu)$ while the subsequent minimization of the maximized entropy permits to reduce the uncertainty on the relevant parameters.

2. Maximum entropy approach

Let us first summarize some basic principles of the maximum entropy approach. In the light of what is discussed above, we consider a parametrized entropy functional $H(\mu)$ of a quite general form [8]

$$H(\mu) = D(\Phi[f(x|\mu)]) \quad (1)$$

Here H is simultaneously a function of the parameter μ and a functional of the parametrized distribution $f(x|\mu)$, $D(t)$ is some function with appropriate behavior. The functional $\Phi[f(x|\mu)]$ can be represented as

$$\Phi[f(x|\mu)] = \int dx \varphi(f(x|\mu)), \quad (2)$$

where $\varphi(p)$ is another function ensuring suitable overall properties (continuity, concavity,...). Combining eqs. (1) and (2) we arrive at

$$H(\mu) = D\left(\int dx \varphi(f(x|\mu))\right) \quad (3)$$

It is clear that practically all generalized entropy functionals appearing in the current literature can be deduced from the definition above (except the forms involving the distribution derivatives like the Fisher form). For instance, for $D(t) = t$, $\varphi(p) = -p \ln p$ we recover the Shannon entropy

$$H_S(\mu) = - \int dx f(x|\mu) \ln f(x|\mu) \quad (4)$$

Let us consider a problem of maximizing $H(\mu)$ under two natural constraints - the normalization

$$\int dx f(x|\mu) = 1 \quad (5)$$

and some known average $\theta(\mu)$ of a conditional response function $\theta(\mu|x)$

$$\theta(\mu) = \overline{\theta(\mu|x)} = \int dx f(x|\mu) \theta(\mu|x). \quad (6)$$

Here the overbar denotes the corresponding average taken with the distribution $f(x|\mu)$. Usually the average $\theta(\mu)$ corresponds to the observed behavior (e.g. experimental data) and the

conditional function $\theta(\mu|x)$ represents our theoretical estimation (a model) of this behavior for a given state x of the random environment. The latter is characterized by a probability distribution $f(x|\mu)$ depending on the "driving" parameter μ (for instance, this could be an external field, pressure or chemical potential).

Thus we have to find a variation of the following Lagrangian

$$L = H(\mu) + \kappa \left[\int dx f(x|\mu) \theta(\mu|x) - \theta(\mu) \right] + \lambda \left[\int dx f(x|\mu) - 1 \right], \quad (7)$$

where κ and λ are the Lagrange multipliers which should be found from the constraints (5) and (6). The variation procedure $\delta L / \delta f(x|\mu) = 0$ leads to

$$\frac{\partial D(\Phi[f(x|\mu)])}{\partial \Phi[f(x|\mu)]} \frac{\partial \varphi(f(x|\mu))}{\partial f(x|\mu)} = -\kappa \theta(\mu|x) - \lambda \quad (8)$$

In the case of the Shannon entropy $H(\mu) = H_S(\mu)$ (4) this procedure leads to an apparently exponential distribution

$$f(x|\mu) = \frac{\exp(-\kappa \theta(\mu|x))}{\int dx \exp(-\kappa \theta(\mu|x))} \quad (9)$$

Nevertheless, if the relevant variable is x , then the actual form of the distribution (9) depends on the form of the constrained function $\theta(\mu|x)$. For instance, as we have demonstrated [10, 11], if $\theta(\mu|x)$ is of logarithmic form, then (9) transforms into a Γ -distribution. Physically this corresponds to a non-equilibrium stationary state whose thermodynamic entropy is held at a given distance from the maximum that corresponds to the equilibrium. For $\theta(\mu|x) = g(\mu)x^\alpha$ we obtain from (9) a stretched exponential distribution. This means that we can associate exponential distributions to the Shannon entropy only in the case of linear constraint, i.e. $\theta(\mu|x) = g(\mu)x$.

3. Minimal maximized entropy

As is discussed above we are dealing with the parametrized entropy $H(\mu)$. Let us consider now the variation of $H(\mu)$ with respect to the parameter μ . From eqs. (1) and (2) we obtain

$$\frac{\partial H(\mu)}{\partial \mu} = \frac{\partial D(\Phi[f(x|\mu)])}{\partial \Phi[f(x|\mu)]} \frac{\partial \Phi[f(x|\mu)]}{\partial \mu}; \quad \frac{\partial \Phi[f(x|\mu)]}{\partial \mu} = \int dx \frac{\partial \varphi(f(x|\mu))}{\partial f(x|\mu)} \frac{\partial f(x|\mu)}{\partial \mu} \quad (10)$$

Here $\partial D(\Phi[f(x|\mu)]) / \partial \Phi[f(x|\mu)]$ could be a function of μ . Anyway it does not depend on x since $\Phi[f(x|\mu)]$ is a functional of the probability distribution and simultaneously a function of μ . Therefore, combining these two equations we obtain

$$\frac{\partial H(\mu)}{\partial \mu} = \int dx \frac{\partial D(\Phi[f(x|\mu)])}{\partial \Phi[f(x|\mu)]} \frac{\partial \varphi(f(x|\mu))}{\partial f(x|\mu)} \frac{\partial f(x|\mu)}{\partial \mu} \quad (11)$$

Using now the condition (8) for the maximum entropy we arrive at

$$\frac{\partial H(\mu)}{\partial \mu} = -\kappa \int dx \theta(\mu|x) \frac{\partial f(x|\mu)}{\partial \mu} - \lambda \int dx \frac{\partial f(x|\mu)}{\partial \mu} \quad (12)$$

It is clear that the last term disappears because of the normalization condition (5) and we finally get the maximum entropy rate as

$$\frac{\partial H(\mu)}{\partial \mu} = -\kappa \int dx \theta(\mu|x) \frac{\partial f(x|\mu)}{\partial \mu}. \quad (13)$$

Now differentiating the constraint (6) we obtain

$$\frac{\partial \theta(\mu)}{\partial \mu} = \int dx \frac{\partial \theta(\mu|x)}{\partial \mu} f(x|\mu) + \int dx \theta(\mu|x) \frac{\partial f(x|\mu)}{\partial \mu} \quad (14)$$

Combining these two equations we arrive at

$$\frac{\partial H(\mu)}{\partial \mu} = -\kappa \left[\frac{\partial \theta(\mu)}{\partial \mu} - \overline{\frac{\partial \theta(\mu|x)}{\partial \mu}} \right] \quad (15)$$

Therefore, we have a "universal" relation among the entropy rate, the observed behavior $\theta(\mu)$ and the average of the model estimation $\theta(\mu|x)$. This relation does not depend on the form of the entropic functional (the form (1) is quite general) and thus could be helpful for drawing conclusions of general validity. A brief discussion of (15) for the case of the Shannon form can be found in [10, 11]. Anyway it is clear that the behavior of $H(\mu)$ depends on what is known/measured $\theta(\mu)$ and on our conditional estimation $\theta(\mu|x)$. Thus it seems logically acceptable to search for a way of reducing the ambiguity. Namely, one can analyze the conditions

$$\frac{\partial H(\mu)}{\partial \mu} = 0, \quad \frac{\partial^2 H(\mu)}{\partial \mu^2} > 0 \quad (16)$$

for a minimum of $H(\mu)$. In such a way we can find some specific values of the parameters at which the uncertainty on the values of the relevant variable x is minimal. Such a procedure might help to determine which "driving path" is more informative. Let us illustrate this on a simple example. Suppose that there are two relevant choices:

$$\theta(\mu) = A\mu^q \quad (17)$$

$$\theta(\mu|x) = x\mu^\alpha \quad (18)$$

, where the value of α is to be estimated. Maximization of the Shannon entropy (4) under these conditions leads to

$$f(x|\mu) = \frac{\mu^{(\alpha-q)}}{A} \exp \left[-x \frac{\mu^{(\alpha-q)}}{A} \right] \quad (19)$$

Estimating the terms in the relation (15) we get

$$\frac{\partial \theta(\mu)}{\partial \mu} = Aq\mu^{q-1} \quad (20)$$

$$\overline{\frac{\partial \theta(\mu|x)}{\partial \mu}} = A\alpha\mu^{q-1} \quad (21)$$

Analyzing the first condition in (16) we arrive at $\alpha \rightarrow q$. From the second condition of (16) we conclude that $\alpha \rightarrow q$ from above or below, depending on the value of q . Based on these arguments and returning back to (19) one can easily see that the minimum uncertainty conditions point up towards choosing the model close to the observed behavior $\theta(\mu)$. In this way the probability distribution becomes independent of the driving μ .

Thus, the minima of $H(\mu)$ occurring at μ_1, \dots, μ_n correspond to the minima of the observed function $\theta(\mu)$ and this is equivalent to a minimization of the distribution width with fixing the positions of the distribution maxima at $x = \mu_1, \dots, \mu_n$. Therefore, equation (15) together with the conditions (16) allow one to translate the specificities of the observed behavior $\theta(\mu)$ into the amount of information on the relevant random variable x . Obviously an interpretation of this correspondence depends on the nature of processes which are encoded in the function appearing in the constraint (6).

4. Conclusion

A minimization procedure of the maximum entropy of a quite general entropic functional with respect to an external driving force μ is proposed. It is demonstrated that for a conditional or parametrized probability distribution $f(x|\mu)$ there is a "universal" relation among the entropy rate and the functions appearing in the constraint. This relation allows one to translate the specificities of the observed behavior $\theta(\mu)$ into the amount of information on the relevant random variable x at different values of the parameter μ . It is clear that we are not free to change the constraint in *ad hoc* manner without some physical (or logical) reasons because it encodes our partial information $\theta(\mu)$ into what we are supposed to know ($\theta(\mu|x)$) and what we are searching for ($f(x|\mu)$). In application to coupled dynamic-stochastic systems such an approach allows one to reduce the uncertainty in estimating the dynamic counterpart (a probe or a model) parameters. In its turn this permits to estimate an optimal driving path giving a maximal information on the probability distribution $f(x|\mu)$ of the stochastic counterpart with a given probe/model $\theta(\mu|x)$. It is found that the functional form of the model should be similar to the observed/measured one $\theta(\mu)$, while the minimum uncertainty is reached when the distribution becomes independent of the driving μ .

References

- [1] Jaynes E T 2006 *Probability Theory, The Logic of Science* (Cambridge: Cambridge University Press)
- [2] Abe S and Okamoto Y 2001 *Nonextensive Statistical Mechanics and its Applications* (Berlin: Springer-Verlag)
- [3] Beck C 2009 *Contemp. Phys.* **50** 495
- [4] Abe S and Thurner S 2008 *Europhys. Lett.* **81** 10004
- [5] Plastino A R, Plastino A and Soffer B H 2007 *Phys. Lett. A* **363** 48
- [6] Vakarin E V, Duda Y and Badiali J P 2006 *J. Chem. Phys.* **124** 144515
- [7] Garcia-Belmonte G, Vakarin E V, Bisquert J and Badiali J P 2010 *Electrochim. Acta* **55** 6123
- [8] Vakarin E V and Badiali J P 2011 *Mod. Phys. Lett. B* **25** 1821
- [9] Beck C 2001 *Phys. Rev. Lett.* **87** 180601
- [10] Vakarin E V and Badiali J P 2006 *Phys. Rev. E* **74** 036120
- [11] Vakarin E V and Badiali J P 2006 *AIP Conf. Proc.* **872** 411