Microcanonical origin of the maximum entropy principle for open systems

Julian Lee*
Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Korea

Steve Pressé†
Physics Department, Indiana University–Purdue University Indianapolis (IUPUI), Indianapolis, Indiana 46202, USA
(Received 17 August 2012; published 15 October 2012)

There are two distinct approaches for deriving the canonical ensemble. The canonical ensemble either follows as a special limit of the microcanonical ensemble or alternatively follows from the maximum entropy principle. We show the equivalence of these two approaches by applying the maximum entropy formulation to a closed universe consisting of an open system plus bath. We show that the target function for deriving the canonical distribution emerges as a natural consequence of partial maximization of the entropy over the bath degrees of freedom alone. By extending this mathematical formalism to dynamical paths rather than equilibrium ensembles, the result provides an alternative justification for the principle of path entropy maximization as well.

DOI: 10.1103/PhysRevE.86.041126 PACS number(s): 05.20.Gg, 05.70.−a

I. INTRODUCTION

The microcanonical ensemble describes a closed system with fixed parameters specifying the macroscopic state. In contrast, the canonical ensemble describes a system in thermal contact with a heat bath with a fixed temperature. It is often described as a special limit of the microcanonical ensemble, with the microcanonical system consisting of two weakly coupled subsystems where one, designated the bath, is much larger than the other subsystem. In this regard, the microcanonical formulation of statistical physics can be argued to be more fundamental.

For a closed system, the basic postulate of the statistical mechanics is that all the microscopic states consistent with the specified macroscopic parameters have equal probability of being occupied by the system [1–3]. The canonical distribution is then obtained by dividing up the closed system into the system of interest and the heat bath, the latter being much larger than the former, and then summing over all the bath microstates [1,3,4].

An alternative derivation of the canonical distribution follows from the maximum entropy principle (MEP) [1,2,5,6]. Here the Gibbs-Shannon entropy, $H(\{p_i\}) = -\sum_i p_i \log p_i$, is maximized under the constraint that the expectation value of energy, $\sum_i p_i E_i$, is fixed to a certain value, where p_i and E_i are the occupation probability and the energy of the microstate labeled as i.

No transparent connection between both approaches just described has been established, which would require explaining why fixing the temperature of the heat bath is equivalent to fixing the mean energy of the open system. In fact, the temperature of the heat bath determines the most probable value of open system energy, which is determined by the condition that the derivative of the system entropy with respect to energy should be equal to the inverse of the heat bath temperature. In the limit of infinite system size, the relative energy fluctuations from the expectation value vanish and the energy expectation value becomes essentially the same as the most probable value of energy. However, this equivalence breaks down once we consider a small open system. Now that nanoscale systems are routinely probed [7], a clear demonstration of this equivalence, that does not take the infinite system size limit, is necessary.

In this work, we show that the MEP for the open system simply follows from the MEP applied to the closed universe consisting of an open system and a heat bath. The MEP for the open system follows from partial maximization of the Gibbs-Shannon entropy of the closed universe over the bath degrees of freedom. Since it is well known that MEP yields the microcanonical ensemble for a closed system, our result shows the equivalence between the microcanonical and MEP derivations of the canonical ensemble.

The mathematics of the present work can be reinterpreted in the context of dynamical systems. In this light, we will discuss how the microcanonical origin for the MEP provides an alternative justification to MEP applied to dynamical systems with mean flux constraints [8–14].

II. MAXIMUM ENTROPY PRINCIPLE FOR CLOSED AND OPEN SYSTEMS

In this section, we briefly highlight the formalism necessary for the remainder of this work. We introduce the target functions used to derive the canonical (i.e., open) and microcanonical (i.e., closed) distributions from the MEP.

The target function of the closed system is

$$H(\{p_i\}) + \lambda \sum_i p_i (1 - \delta_{E_i E}) + \nu \left(\sum_i p_i - 1 \right), \quad (1)$$

whereas that for the open system is

$$H(\{p_i\}) - \beta \left(\sum_i p_i E_i - \epsilon \right) + \nu \left(\sum_i p_i - 1 \right). \quad (2)$$

*jul@ssu.ac.kr
†stevenpresse@gmail.com

In both expressions, $H(\{p_i\})$ is the Gibbs-Shannon entropy. In Eqs. (1) and (2), ν is the Lagrange multiplier which enforces normalization of the p_i, that is, $\sum_i p_i = 1$. The Lagrange multipliers λ and β set the two different constraints imposed on closed and open systems. For closed systems, Eq. (1), only energy E_i coinciding with the closed system E is accepted. Thus the Lagrange multiplier λ enforces the constraint $\sum_i p_i(1-\delta_{E_i E})=0$. In the case of the open system, the Lagrange multiplier β enforces a constraint, $\sum_i p_i E_i - \epsilon = 0$, on the average system energy ϵ. These target functions are to be minimized with respect to p_i values and the Lagrange multipliers to fully determine the values of these variables.

For a closed system, the variation of Eq. (1) yields the uniform distribution describing the microcanonical ensemble

$$p_i = \frac{\delta_{E,E_i}}{\Omega(E)}, \tag{3}$$

where $\Omega(E)$ is the total number of microstates with energy E.[1] For an open system, Eq. (2) yields the Boltzmann distribution from the canonical ensemble

$$p_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}, \tag{4}$$

where β is determined by the condition $\sum_k E_k e^{-\beta E_k}/\sum_i e^{-\beta E_i} = \epsilon$. We call this derivation the MEP derivation of the canonical ensemble.

It is well known how the canonical ensemble, Eq. (4), is derived from the microcanonical ensemble [1,3,4] by considering the closed universe consisting of an open system and the heat bath surrounding it. To wit

$$\begin{aligned} p_i \propto \Omega_{\text{bath}}(E_{\text{tot}} - E_i) &= \exp[S_{\text{bath}}(E_{\text{tot}} - E_i)] \\ &\simeq \exp[S_{\text{bath}}(E_{\text{tot}}) - \beta E_i], \end{aligned} \tag{5}$$

where the fixed constant $E_{\text{tot}} = E_i + E_{\text{bath}}$ is the total energy of the closed universe, consisting of the open system and the heat bath. Since the bath energy is assumed much larger than the system energy, we keep only the leading term in the expansion where $\beta^{-1} = [\frac{dS_{\text{bath}}}{dE_{\text{bath}}}]^{-1}$, which is also a fixed constant. We call this derivation the microcanonical derivation of the canonical ensemble.

The connection between microcanonical and MEP derivations of the canonical distribution is not clear. The parameter β in Eq. (2) is introduced as the Lagrange multiplier for constraining the mean energy, whereas β in Eq. (5) is the inverse of heat bath temperature, the derivative of the heat bath entropy with respect to energy.

In the next section, it is shown that the MEP for the open system can indeed be derived using a two-step maximization of Gibbs-Shannon entropy for the entire closed universe. Since the MEP applied to the closed universe yields the microcanonical ensemble, the equivalence between the microcanonical and MEP derivations of Boltzmann distribution is thus established.

III. THE MEP FOR OPEN SYSTEMS EMERGES FROM THE MEP FOR CLOSED SYSTEMS

We begin with the target function for the closed universe composed of the open system plus bath as follows:

$$\begin{aligned} &H_{\text{tot}}(\{p_{ia}\}) + \nu\left(\sum_{i,a} p_{ia} - 1\right) + \lambda \sum_{i,a} p_{ia}(1 - \delta_{E_i+E_a,E_{\text{tot}}}) \\ &= -\sum_{i,a} p_{ia}\log p_{ia} + \nu\left(\sum_{i,a} p_{ia} - 1\right) \\ &\quad + \lambda \sum_{i,a} p_{ia}(1 - \delta_{E_i+E_a,E_{\text{tot}}}), \end{aligned} \tag{6}$$

where indices i and a label the microstates of the open system and the heat bath, respectively. Denoting the total number of the open system and the heat bath microstates as A and B, respectively, there are AB total p_{ia} variables, of which $AB-1$ are independent because of the normalization condition on p_{ia}.

We rewrite p_{ia} in terms of the marginal open system distribution p_i and a conditional bath distribution $p(a|i)$ defined as follows: $p_i \equiv \sum_a p_{ia}$ and $p(a|i) \equiv p_{ia}/p_i$. The number of components for p_i and $p(a|i)$ are A and AB, but the normalization conditions

$$\begin{aligned} &\sum p_i = 1 \\ &\sum_a p(a|i) = 1 \quad (i = 1 \dots A) \end{aligned} \tag{7}$$

reduce the number of independent components to $A-1$ and $AB-A$, respectively, thus making the total number of independent components $AB-1$ as before.

Writing the target function in terms of p_i and $p(a|i)$ we have

$$\begin{aligned} &-\sum_{i,a} p(a|i)p_i \log[p(a|i)p_i] + \sum_i \nu_i\left[\sum_a p(a|i) - 1\right] \\ &\quad + \mu\left[\sum_i p_i - 1\right] + \lambda\sum_{i,a} p_i\, p(a|i)(1-\delta_{E_i+E_a,E_{\text{tot}}}) \\ &= -\sum_i p_i \log p_i - \sum_{i,a} p(a|i)p_i \log p(a|i) \\ &\quad + \sum_i \nu_i\left[\sum_a p(a|i) - 1\right] + \mu\left[\sum_i p_i - 1\right] \\ &\quad + \lambda\sum_{i,a} p_i\, p(a|i)(1-\delta_{E_i+E_a,E_{\text{tot}}}), \end{aligned} \tag{8}$$

where Eq. (7) was used in going from the first to the second line.

Now we can perform the maximization of Eq. (8) in two steps. First, we vary the target function with respect to $p(a|i)$,

[1]For notational simplicity, we assume discrete energy levels here. When the energy levels are continuous, it is more natural to count the number of states with energy values lying between E and $E + dE$ with some small number dE. Then δ_{E,E_i} in Eq. (3) should be replaced by $\theta(E_i - E)\theta(E + dE - E_i)$, with corresponding minor modifications in the following sections.

ν_i, and λ in order to eliminate these for a given p_i. Second, the target function is maximized with respect to the remaining open system variables p_i and μ.

Taking the variation of Eq. (8) with respect to $p(a|i)$, ν_i, and λ and setting them to zero, we get

$$-p_i \log p(a|i) - p_i + \nu_i + \lambda p_i (1 - \delta_{E_i+E_a,E_{\text{tot}}}) = 0, \quad (9)$$

$$\sum_a p(a|i) = 1, \quad (10)$$

$$\sum_{i,a} p_i\, p(a|i)(1 - \delta_{E_i+E_a,E_{\text{tot}}}) = 0. \quad (11)$$

From Eq. (9) we have

$$\begin{aligned} p(a|i) &= \exp\left(\frac{\nu_i}{p_i} - 1\right) \quad (E_a = E_{\text{tot}} - E_i), \\ p(a|i) &= \exp\left(\frac{\nu_i}{p_i} - 1 + \lambda\right) \quad (E_a \neq E_{\text{tot}} - E_i). \end{aligned} \quad (12)$$

Next, the constraints Eq. (10) and Eq. (11) fix the values of ν_i and λ, which yields[2]

$$p(a|i) = \frac{\delta_{E_a,E_{\text{tot}}-E_i}}{\Omega_{\text{bath}}(E_{\text{tot}} - E_i)}. \quad (13)$$

Substituting Eq. (13) into Eq. (8), we now get

$$\begin{aligned} &-\sum_i p_i \log p_i + \sum_{i,a} p_i \frac{\delta_{E_a,E_{\text{tot}}-E_i}}{\Omega_{\text{bath}}(E_{\text{tot}} - E_i)} \log \Omega_{\text{bath}}(E_{\text{tot}} - E_i) + \mu\left(\sum_i p_i - 1\right) \\ &= -\sum_i p_i \log p_i + \sum_i p_i \log \Omega_{\text{bath}}(E_{\text{tot}} - E_i) + \mu\left(\sum_i p_i - 1\right). \end{aligned} \quad (14)$$

Since the heat bath is much larger than the open system, we have $\log \Omega_{\text{bath}}(E_{\text{tot}} - E_i) \simeq \log \Omega_{\text{bath}}(E_{\text{tot}}) - \beta E_i$, we get

$$\tilde{H}(\{p_i\}) + \mu\left(\sum_i p_i - 1\right) \equiv -\sum_i p_i \log p_i - \beta \sum_i p_i E_i + \mu\left(\sum_i p_i - 1\right), \quad (15)$$

where an irrelevant constant term was dropped. The MEP target function for an open system, Eq. (2), is now reproduced from Eq. (1). This shows that thermal contact with a heat bath of fixed temperature for an open system is equivalent to constraining the expectation value of energy.

IV. DISCUSSION

It was shown that the MEP for the open system arises in the process of a two-step maximization of the Gibbs-Shannon entropy of the closed universe, first with respect to the bath variables and second with respect to the system variables. This demonstrates the mathematical equivalence of providing contact with the heat bath versus constraining a fixed energy expectation value.

For the sake of concreteness, we used energy as a constraint. However, our constraints could have been other equilibrium macrovariables as well as dynamical constraints such as fluxes. For instance, average fluxes have been used to infer dynamical path probability distribution by maximizing the Gibbs-Shannon entropy over path probabilities subject to constraints on mean fluxes [8–14]. Here we have provided an alternative justification for such methods within a microcanonical framework. We note, however, that the total "universe," consisting of the open system and the "flux bath," cannot be interpreted as a closed universe in the thermodynamic sense, since a system with a nonzero flux is always a system driven by an external agent. For the case of a dynamical system, one just considers a large system with a fixed total flux value. Then, from MEP, all micropaths consistent with this flux value are equally probable. The marginal probability distribution of subsystem paths then follows from the maximization of the path entropy under the constraint of fixed flux expectation value.

ACKNOWLEDGMENTS

We thank Kingshuk Ghosh and Ken Dill for useful discussions. In addition, S.P. acknowledges Ken Dill for his support.

[2]The values of the Lagrange multipliers which satisfy both Eq. (10) and Eq. (11) are $\nu_i = p_i[\log \Omega(E_{\text{tot}} - E_i) - 1]$ and $\lambda = -\infty$. To avoid an infinite Lagrange multiplier for λ, one may use $\tau \equiv e^{\lambda}$ instead of λ.

[1] M. Plischke and B. Bergersen, *Equilibrium Statistical Physics* (Prentice Hall, Englewood Cliffs, NJ, 1989).

[2] L. E. Reichl, *A Modern Course in Statistical Physics* (University of Texas Press, Austin, TX, 1980).

[3] R. K. Pathria, *Statistical Mechanics* (Elsevier, New York, 1972).

[4] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[5] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957).

[6] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, London, 2003).

[7] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002); C. Bustamante, Z. Bryant, and S. B. Smith, Nature **421**, 423 (2003); D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *ibid.* **437**, 231 (2005).

[8] A. A. Filyukov and V. Y. Karpov, J. Eng. Phys. Thermophys. **13**, 326 (1967); **13**, 416 (1967); A. A. Filyukov, *ibid.* **14**, 429 (1968).

[9] E. T. Jaynes, Macroscopic Prediction, in *Complex Systems Operational Approaches in Neurobiology, Physics, and Computers*, edited by H. Haken (Springer-Verlag, Berlin, 1985).

[10] G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102, (2008).

[11] K. Ghosh, K. A. Dill, M. M. Inamdar, E. Seitaridou, and R. Phillips, Am. J. Phys. **74**123, 2006.

[12] S. Pressé, K. Ghosh, and K. Dill, J. Phys. Chem. B **115**, 6202 (2011).

[13] H. Ge, S. Pressé, K. Ghosh, and K. A. Dill, J. Chem. Phys. **136**, 064108 (2012).

[14] C. J. Monthus, J. Stat. Mech.: Theory Exp. (2011) P03008.