Entropic criterion for model selection

Chi-h- Yuan Tseng
Computational Biology Laboratory
Department of Physics
National Central University, Chungli, Taiwan 320

Abstract

Model or variable selection is usually achieved through ranking models according to the increasing order of preference. One of the methods is applying Kullback-Leibler distance or relative entropy as a selection criterion. Yet that will raise two questions, why use this criterion and are there any other criteria. Besides, conventional approaches require a reference prior, which is usually difficult to get. Following the logic of inductive inference proposed by Caticha [1], we show relative entropy to be a unique criterion, which requires no prior information and can be applied to different fields. We examine this criterion by considering a physical problem, simple units, and results are promising.

Keyword: Model selection, Inductive inference, Kullback-Leibler distance, Relative entropy, Probability model

PACS: 02.50.Sk, 02.50.Tt, 02.50.Le

1 Introduction

Model or variable selection in process of data analysis is usually achieved by ranking models according to the increasing order of preference. Several methods rooted in this concept such as P-values, Bayesian, and Kullback-Leibler distance method, etc., are some popular examples to provide pertinent selection criteria. P-values method selects models by comparing probability of model given a null model and experimental data sets to a threshold value assessed from same data sets [2]. Yet since this method is restricted to two models and required ad hoc rules to assess threshold value, people has developed Bayesian approaches to overcome these defects [3, 4, 5, 6]. The Bayesian method applies Bayes theorem to update our beliefs and uncertainty about models from prior distributions generated from some prior modeling rules [7]. A preferable model, thereafter, is chosen according to Bayes factor, ratio of posterior distributions of different models. Bayesian Information Criterion (BIC) is one of the most popular model selection criteria [2, 3, 6]. Yet all of these methods require prior information generated from some ad hoc prior modeling rules that suits people’s need.

Aside from Bayesian framework, people also has developed relative entropy, mutual information, or Kullback-Leibler distance based approach [7, 8]. Kullback-Leibler distance measures differences between models and a reference prior for interested systems. The decreasing Kullback-Leibler distance then suggests the increasing order of preference of models. A selection method proposed by Dupuis and Robert on variable selection [8] is based on the evaluation of Kullback-Leibler distance between...
The full model description by complete set of variables for interested system and its approximations, sub-models, described by subset of variables. Given prior information on full model, preferred sub-model is selected when it’s Kullback-Leibler distance reaches a threshold value, estimated by experiences. Since sub-models are projections of full model, there is no need the prior modeling rule to generate prior distribution for sub-model. Yet one still requires prior information on full model. Moreover, it remains questionable to apply Kullback-Leibler distance as a selection criterion even though Dupuis and Robert argued that it is common on choice for information theoretic and intrinsic considerations and computational reasons. Besides, the choice is made because of its property of transitivity and additivity that relate to theory of generalized linear models [8] attempted to apply in breast cancer studies. Our first goal of this work is to answer questions of why is Kullback-Leibler distance not any other criteria. Are there any other entropy based criteria for model selection? A forward, we shall develop an entropy based method to provide ranking scheme for model selection that is free from difficulties encountered in conventional entropic studies. The strategy is closely following the logic of inductive inference proposed by [11] that is to generalize method of maximum entropy (ME) from Jaynes’s version of probability distribution assignments [5] to be a tool of inductive inference initiated by [10] and [11]. This logic differs in one remarkable way from the manner that has in the past been followed in setting up physical theories for example. Normally one starts by establishing a mathematical axiom, and then one tries to append an interpretation to it. This is a very difficult problem; it has affected the development of statistics and statistical physics — what is the meaning of probability distribution and of entropy. The issue of whether the proposed interpretation is unique, or even whether it is allowed, always remains a legitimate objection and a point of controversy. Our procedure is in the opposite order, we first decide what are we talking about and what is our goal, namely, selection criterion for ranking scheme, and only afterward we design the appropriate mathematical axiom, the issue of what is the meaning of probability distributions and of entropy will then never arise.

Based on Cachtia’s logic of inductive inference, we shall derive and present the entropic criterion in next section. It will show relative entropy to be a unique criterion for model selection. We then examine this criterion by considering a complicated physical problem, simple fluids. It has become a well established field over almost three decades. People has developed many approximation models to study and interpret fluid’s properties. Since we have rich theoretical and empirical knowledge of simple fluids from conventional studies to rank those approximation models, it shall provide us a conceivable benchmark for our investigations. Approximation models, mean-field, hard-sphere and improved hard-sphere approximation are considered and briefly presented in section 3-1. We then apply entropic criterion to rank these three models. Details calculations of entropic criterion and comparison against results inferred from conventional analysis are shown in section 3-2. A summary of our discussions is listed in section 4.

2 Entropic criterion for model selection

As mentioned, the selection of one model from a group of models is achieved by ranking those models according to increasing preference. Before we address the issue of what it is that makes one model preferable over another we note that there is one feature we must impose on any ranking scheme. The feature is transitivity: if model 1 described by distribution p_1 is preferred over model 2 described by distribution p_2 and p_1 is preferred over p_3, then p_1 is preferred over p_3. Such transitive rankings are in place by assigning to each p(x) a real number S[p] which we call the "entropy" of p. The numbers S[p] are such that if p_1 is preferred over p_2, then S[p_1] > S[p_2].

Next we determine the functional form of S[p]. The basic strategy [11] is that: (1) if a general rule exists, then it must apply to special cases; (2) if in a certain special case we know which is the best model, then this knowledge can be used to constrain the form of S[p]; and finally, (3) if enough special cases are known, then S[p] will be completely determined. The known special cases are
called the "axiom of ME" and they reject the conviction that one should not change one's mind frivolously, that whatever information was codified in probability distribution \( p(x) \) is important. Three axioms and their consequences are listed below. Detailed proofs are given in [1].

**Axiom 1: Locality.** Local information has local effects. If the constraints that define the probability distribution do not refer to a certain domain \( D \) of the variable \( x \), then the conditional probabilities \( p(x|D) \) need not be revised. The consequence of the axiom is that non-overlapping domains of \( x \) contribute additively to the entropy: \( S[p] = \int dx F(p(x)) \) where \( F \) is some unknown function.

**Axiom 2: Coordinate invariance.** The ranking should not depend on the system of coordinates. The coordinates that label the points \( x \) are arbitrary; they carry no information. The consequence of this axiom is that \( S[p] = \int dx p(x)f(p(x)=m(x)) \) involves coordinate invariants such as \( dx p(x) \) and \( p(x)=m(x) \), where the function \( m(x) \) is a density, and both functions \( m \) and \( f \) are, at this point, unknown. We make a second use of the locality axiom to determine \( m(x) \). When there are no constraints at all and group of different models includes the exact \( P(x) \) for real system, the selected probability model \( p(x) \) should coincide with \( P(x) \); that is, the best probability model \( p(x) \) to real system described by \( P(x) \) is \( P(x) \) itself. On the contrary it suggests that the best probability model \( p(x) \) to \( P(x) \) should be farthest from uniform distribution \( m \). Since exact distribution \( P(x) \) is so complicated it is useful in practical calculations while uniform distribution \( m \) is free from the difficulty. Thus we shall choose uniform distribution \( m \) to be \( m(x) \). At last, we will consider third axiom to determine function \( f \).

**Axiom 3: Consistency for independent subsystems.** When a system is composed of subsystems that are believed to be independent it should not matter whether we treat them separately or jointly. If we originally believe that two systems are independent and the constraints defining the probability distributions are silent on the matter of correlations, then there is no reason to change one's mind. Specifically, if \( x = (x_1,x_2) \), and the exact distributions for the subsystems \( p_1(x_1) \) and \( p_2(x_2) \), then the exact distribution for the whole system should be \( p_1(x_1)p_2(x_2) \). This axiom restricts the function \( f \) to be a logarithm.

The overall consequence of three axioms is that the probability distribution \( p(x) \) should be ranked relative to \( m \) according to their (relative) entropy,

\[
S[p;m] = \int dx p(x) \log \frac{p(x)}{m(x)} = \ln m \int dx p(x) \log p(x) \quad 0.
\]

The derivation has singled out \( S[p;m] \) as the unique entropy to be used for the purpose of ranking probability distributions. Other expressions, \( m \) may be useful for other purposes, but they are not a generalization from the simple cases described in the axioms above. Notice that since \( p(x) \) is ranked relative to a uniform distribution \( m \), which is independent of models. Thus decreasing \( S[p] = \int dx p(x) \log p(x) ) \ln S[p;m] \) indicates there existing more differences between probability model \( p(x) \) and uniform distribution \( m \). Namely, \( p(x) \) that is farther away from uniform distribution carries more relevant information about real system, and \( m \) is more preferable. Before applying entropic criterion to real problem, a sum many of our derivation is given. Based on logic of inductive inference, the answer to questions raised earlier becomes obvious. The use of relative entropy for selection criterion is just what we design to achieve, and needs no further interpretation. Besides, since this criterion is designed based on probability models, it will accommodate to all kinds of probabilistic problems.

3 A physical problem: simple uids

3.1 Approximation models for simple uids

We shall examine proposed entropic criterion by considering a complicated problem in physics, simple uids (reviews of simple uids can be found in [12], [13], and [14]), in this section. Suppose
a simple uid with density and volume $V$ is composed of $N$ single atom molecules. This uid is described by the Hamiltonian

$$H \left( q_i \right) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + U \text{ with } U = \sum_{i>j} u \left( r_{ij} \right);$$

where $q_i = \left( p_i, r_i \right)$ and the many-body interactions are approximated by a Lennard-Jones pair interaction, $u \left( r_{ij} \right) = 4 \left( \frac{r_{ij}}{\sigma} \right)^{12} - \left( \frac{r_{ij}}{\sigma} \right)^6$ where $r_{ij} = |r_i - r_j|$ and are Lennard-Jones parameters. The probability that the positions and momenta of the molecules lie within the phase space volume $d q_i$ is given by

$$P \left( q_i \right) d q_i = \frac{1}{Z} \exp \left[ - \frac{H \left( q_i \right)}{k_B T} \right];$$

where $\frac{1}{k_B}$ is Boltzmann constant. Since there are $N \left( \frac{1}{2} \right) = 2$ pair interactions, integration of partition function $Z$ in Eq. (6) is impossible to accomplish. $P \left( q_i \right)$ is useless in practical calculations. One strategy to bypass this problem is constructing approximation models that are described by tractable probability distributions. Several approximation models are, therefore, developed according to researcher’s knowledge and experiences in the studies of simple uids in last three decades. Yet we shall consider only three approximation models, mean uid from [15], hard-sphere from [13] or [3] and improved hard-sphere approximation from [7] that are briefly presented in the following to dem onstrate the use of entropic criterion for model selection.

**Mean uid approximation:** Mean uid approximation drastically approximate complicated long range interactions $u \left( r_{ij} \right)$ by an optimized mean uid $v_m \left( r_{ij} \right)$, which is determined by the ME method from [15]. Probability distribution given by mean uid approximation

$$P_m \left( q_i ; \mathbf{r} \right) = \frac{1}{Z_m \mathbf{r}} \exp \left[ - \frac{H_m \left( q_i \right)}{k_B T} \right] + \frac{1}{Z} \exp \left[ - \frac{H \left( q_i \right)}{k_B T} \right];$$

where

$$H_m \left( q_i \right) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \frac{1}{Z_m \mathbf{r}} \exp \left[ - \frac{H \left( q_i \right)}{k_B T} \right];$$

and $\mathbf{r}$ are Lagrange multipliers that enforce the constraint on the expected density $\rho \left( r \right)$ at each point in space and the density is

$$\rho \left( r \right) = \sum_{i=1}^{N} \left( r_i \right);$$

We remark that a constraint on $\rho \left( r \right)$ also constrains the expected number of particles

$$N = \rho N \mathbf{r} = \text{d}^3 \rho \left( r \right);$$

It is convenient to absorb the mean uid $v_m \left( r \right)$ and the multiplier $\mathbf{r}$ into a single potential $V \left( r \right) = v_m \left( r \right) + \mathbf{r}$, the partition function $Z_m \mathbf{r}$ is

$$Z_m \mathbf{r} = \frac{1}{Z} \sum_{n=0}^{\#} \text{d}^3 q_i \exp \left[ - \frac{H \left( q_i \right)}{k_B T} \right];$$

so that

$$P \left( \mathbf{r} ; \mathbf{z} \right) = \frac{1}{Z} \frac{\text{d}^3 r \exp \left[ - V \left( r \right) \right]}{\text{d}^3 r \exp \left[ - v_m \left( r \right) \right]} = \frac{1}{Z} \frac{\text{d}^3 r \left( \mathbf{z} \right)}{\text{d}^3 r \left( \mathbf{z} \right)};$$

where

$$N = \frac{1}{N} \sum_{i=1}^{N} \left| \mathbf{z} \right|.$$
where \( \frac{\hbar^2}{2m} \) and the expected density is

\[
\mathfrak{h} \left( \mathfrak{i} \right) \mathfrak{s}^2 = \frac{M F}{V \left( r \right)} = \mathfrak{s}^2 \mathfrak{d}^2 \mathfrak{n} \left( r \right) = \mathfrak{n} \left( r \right)
\]

or

\[
\mathfrak{n} \left( r \right) = e^{V \left( r \right)} - \frac{1}{3}
\]

Furthermore, according to Percus-Yevick approximation, one can introduce a useful quantity, radial distribution function \( g \left( r \right) = \mathfrak{n} \left( r \right) = \mathfrak{m} \left( r \right) \) that measures probability of observing particle at distance \( r \) while another particle at origin. It gives information of liquid structures that can be measured directly through x-ray and neutron diffraction experiments. Notice that when two particles are not correlated, radial distribution function \( g \left( r \right) = 1 \), \( g \left( r \right) \) is vanished when two atoms are repelled.

Hard-sphere approximation: One replaces short-range repulsion by hard-sphere potential with an optimal hard-sphere diameter \( r_d \), which is determined by ME method as well \[17\]. Probability distribution given by hard-sphere approximation is

\[
P \left( r \right) = \frac{1}{Z} e^{H \left( r \right) - \mathfrak{d}^2 \mathfrak{n} \left( r \right) - \mathfrak{m} \left( r \right)}
\]

where the Hamiltonian is

\[
H \left( r \right) = e^{U \left( r \right)} + U \left( r \right)
\]

with

\[
U \left( r \right) = \mathfrak{d}^2 \mathfrak{n} \left( r \right)
\]

The partition function and the free energy \( F \left( T, V, N \right) \) obtained by Percus-Yevick approximation \[13\] are

\[
Z = \int dq e^H \left( q \right) = e^{F \left( T, V, N \right)}
\]

and

\[
F \left( T, V, N \right) = N k_B T \left( 1 + \ln \frac{r_D}{r_d} + 4 - \frac{3}{2} \left( \frac{r_D}{r_d} \right)^2 \right)
\]

where the packing fraction, \( \frac{1}{6} \) where \( \frac{r_D}{r_d} \) with \( \frac{r_D}{r_d} = \frac{N}{V} \):

Improved hard-sphere approximation: Although there are several improved hard-sphere models like Barker and Henderson \[12\] and WCA theories \[18\] etc., they were not constructed by probability models directly, and are inappropriate in this investigation. We consider another improved approximation model obtained from ME method that is a probability model and has been proved to be competitive to those theories \[17\]. The crux of this model is that we consider whether the correct choice should have been some other value \( r_d = r_d + \mathfrak{r} \) rather than the optimal \( r_d = r_d \) in original hard-sphere approximation. As discussed in \[17\] this is a question about the probability of \( r_d, P \left( r_d \right) \). Thus, we are uncertain not just about \( q \) but also about \( r_d \) and what
we actually seek is the joint probability of \( q_i \) and \( r_d, P_f (q_i, r_d) \). Once this joint distribution is obtained our best assessment of the distribution of \( q_i \) is given by the marginal over \( r_d, \\
\int d_{r_d} P_d (q_i, r_d) \\
Z \\
= d_{r_d} P_d (r_d) P_{hs} (q_i, r_d), \\
(18) \\
where the distribution of diameters is given by \\
P_d (r_d) = e^{S P_{hs}} P |_{r_d} \int_{1/2}^{1} d_{r_d} = e \frac{F_0}{U} \frac{1}{1/2} \frac{1}{1/2} \frac{d_{r_d}}{d_{r_d}}; \\
(19) \\
(r_d) = N, r_d \frac{4+g}{2} r_d ^{4} \frac{4-2}{2}, \\
defi ned by \( r_d \) and the partition functions and \( U \) are given by \\
Z = e^{F_0} \frac{u}{u} with u = d_{r_d} \int_{1/2}^{1} \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} e \frac{F_0}{U}, \\
(20) \\
and \\
\int_{1/2}^{1} d_{r_d} P_{hs} \int_{1/2}^{1} d_{r_d} = e^{F_{0}} \int_{1/2}^{1} d_{r_d} P_{hs} (r_d, r_d) \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} = e^{F_{0}} \int_{1/2}^{1} d_{r_d} P_{hs} (r_d, r_d) \\
(21) \\
with \( F_0 = P_{hs} (N, V, N, X, r_d) + \frac{1}{2} N \int_{1/2}^{1} \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} (r_d, r_d) \). In addition, one have to consider proper local interaction in model to generate correct liquid structure by requesting \( N \) to be experimental particle number \( N_{eff} \) (please refer to [7] for more discussions). By recognizing that diameters other than \( r_d \) are not ruled out and that a more honest representation is an average over all hard-sphere diameters we are effectively replacing the hard-spheres by a soft-core potential.

3.2 Discussions

3.2.1 Entropic criterion analysis

Now, we present proposed entropic criterion to rank three approximation models for simple fluids. In last section before presenting the actual ranking scheme inferred from detail analysis of liquid structures and thermodynamic properties obtained by these approximations against computer simulations and experimental data. According to proposed entropic criterion, ranking scheme is obtained by calculating entropy \( S \) of probability distributions \( P_i \) of \( P_{hs} \) approximations. Substituting Eq. (12) into \( S \) of \( P_{hs} \), entropy per particle number of \( m \) \( P_{hs} \) approximation \( P_{hs} \) is \\
\int_{1/2}^{1} d_{r_d} P_{hs} (q_i, r_d) \log \frac{P_{hs} (q_i, r_d)}{P (q_i)} = e^{F_{0}} \int_{1/2}^{1} d_{r_d} P_{hs} (q_i, r_d) \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} \\
(22) \\
is obtained with the help of Eq. (2), (3), (10), and \( q_{m \in r} (r) = n_{m \in r} (r) = . Next, entropy of hard-sphere approximation \( P_{hs} \) is calculated by differentiating free energy \( F_{hs}, \) Eq. (17), with respect to temperature, \\
\int_{1/2}^{1} d_{r_d} P_{hs} (q_i, r_d) \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} = e^{F_{0}} \int_{1/2}^{1} d_{r_d} P_{hs} (q_i, r_d) \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} \\
(23) \\
S \int_{1/2}^{1} d_{r_d} P_{hs} = F_{hs} (N, V, N, X, r_d) + \frac{3}{2} \frac{1}{2} \frac{1}{2} \frac{d_{r_d}}{d_{r_d}} \\
= \frac{5}{2} \log \frac{3}{2} \frac{4}{3} \frac{2}{(l \frac{1}{2})}. \
(23)
Table 1: This table lists values of third term in Eq. (25) denoted by HS3 and sum of last two terms in Eq. (25) denoted by IHS3 for three different fluid densities and temperatures T.

| T    | 0.55       | 0.65       | 0.7       |
|------|------------|------------|-----------|
|      | HS3        | HS3        | HS3       | IHS3      | IHS3      | IHS3      |
| 107.82K | -1.758     | -2.284     | -2.297    | -3.182    | -2.603    | -3.682    |
| 161.73  | -1.614     | -2.153     | -2.081    | -2.881    | -2.342    | -2.282    |
| 328.25  | -1.376     | -1.832     | -1.744    | -2.338    | -1.944    | -2.614    |

At last, entropy of probability distribution $P_{hs}$ given by improved hard-sphere approximation is obtained by substituting Eq. (19) and (12) into Eq. (18) rest. Because $U_{hs}$ is Heaviside step function, one can write spatial part of $P_{hs}$ as

$$P_{hs}^0(r) = \frac{Z_r}{d\Delta} \exp F(t_d) \text{ with } F(t_d) = N \frac{4}{(1/P)^3}$$

Since $P_d(t_d)$ is vanished when $r > n$ and $r < n$, integrating $P_d(t_d) \exp F(t_d)$ from zero to $r > n$ in Eq. (24) will give a constant value, which defines a new quantity $F$. Therefore, $S P_{hs}$ is given by

$$S P_{hs} = N_{eff} k_B = \frac{5}{2} \log 3 F$$

where $V^0 = \int_0^{R_1} r^3 P_{hs}^0(r) \, dr$. Next, we compare values of Eq. (20), (24), and (25). The only difference between entropies of mean-eld Eq. (20) and of hard-sphere approximation Eq. (24) is the third term contributed by potential part. Since radial distribution function $g_{m \ell}(r)$ in Eq. (22) is vanished within range of strong repulsive forces, $r = 0$ and $r_1$ and becomes one after $r > r_1$, this result leads to a constant integration of third term that is far smaller than total ild volume $V$. Therefore, entropy of $P_{m \ell}$ is approximated to

$$S P_{m \ell} = N_{eff} k_B < S P_{hs} = N_{eff} k_B$$

One thereafter has the inequality equation,

$$S P_{hs} = N_{eff} k_B < S P_{m \ell} = N_{eff} k_B$$

hard-sphere approximation is preferred over mean-eld approximation. Now consider entropy of approximations $P_{hs}$ and $P_{hs}(k_{eff})$. Numerical calculations of third term in Eq. (25) and sum of last two terms in Eq. (25) with three different ild’s densities and temperatures are shown in Table 1 denoted by HS3 and IHS3 respectively as examples. These numerical values show HS3 to be smaller than HS3, namely $S P_{hs} > S P_{hs} = N_{eff} k_B < S P_{hs} = N_{eff} k_B$, and suggest an expected result that improved hard-sphere approximation is preferred over hard-sphere approximation. Therefore, the complete ranking scheme of these three approximations is

$$S P_{hs} > N_{eff} k_B, S P_{hs} > S P_{m \ell} > N_{eff} k_B < 0$$

where the equality in $S P_{hs}$ and $S P_{m \ell}$ will hold when $N_{eff}$ increases to equal to total particle number, which results in improved hard-sphere approximation to reduce to hard-sphere approximation as discussed in (17).
3.2.2 Conventional analysis

Alternatively, one can determine the ranking scheme of these three approximations through exhaustive analysis of comparing liquid structures and their dynamical properties obtained by these approximations against to computer simulations and experimental data (please refer to [15] and [17] for detail). Results showed that mean field approximation only suits for dilute gases and fails to take short-range interaction into account properly. Contrarily, hard-sphere approximation fails to take softness of the repulsive core, which results in less satisfactory prediction of them dynamical properties at high temperature. Yet hard-sphere approximation still provides better description of short-range interactions than mean field approximation does. Furthermore, since no hard sphere approximation attempts to take softness of repulsive core into account pertinently, results showed such an improvement to be come positive with the best perturbative theories so far. One, therefore, can rank these three approximations for simple-uids studies from these analysis as follows, in proved hard-sphere approximation is preferred over hard-sphere approximation and hard-sphere approximation is preferred over mean field approximation. This is exactly the same ranking scheme as indicated by Eq. (28) yet it requires more exhausting efforts.

4 Discussion

There has been abundant theories proposed to construct robust and elegant model or variable selection criteria. We briefly reviewed the rationale and some shortcomings of these methods. P-values method [2] is restricted to two models and requires some ad hoc rules determining threshold value. Although several Bayesian methods ([3], [2], [4], [5], [6]) are proposed to overcome this shortcoming, it still requires prior modeling rule to generate prior distribution. Aside from Bayesian framework, there are relative entropy, mutual information or Kullback-Leibler distance methods for the same goal ([1], [3]). In [2], Dupuis and Robert applied Kullback-Leibler distance to select sub models, projections of a full model given a full model for interested system. Yet this approach still requires prior information on full model and a threshold value. Moreover, it remains questionable to choose Kullback-Leibler distance as the selection criterion even though Dupuis and Robert gave some arguments to defend such a choice. Our first goal is to answer questions of why is Kullback-Leibler distance for selection criterion and are there any other criteria. Therefore, we propose entropic criterion to determine ranking scheme given a group of several models for a system. Following logic of inductive inference proposed by [11] as mentioned in introduction, we answer these two questions by showing relative entropy to be a unique criterion to rank different models for a system. It is just what we design to achieve, and needs no further interpretation. Besides, there is no restriction on types of probability models in this criterion, and it has wide applicability in all kinds of probabilistic problems. Since probability distribution of real system, however, is always intractable that it is useless in practical calculations, we propose to rank probability models relative to a uniform distribution m instead real probability distribution to bypass this defect. Thus decreasing relative entropy S[\eta m] indicates increasing preference of m models. Note that it has no restrictions on number of m models and requires no ad hoc prior modeling rules. At last, we examine this tool by considering a complicated physical problem, simple-uids in this work. Because people has developed many approximations models to study simple-uids, and accumulated rich knowledge in the past, it provides a conceivable benchmark for our investigation. We consider three approximations models, mean field from [15], hard-sphere, and improved hard-sphere approximation from [17] for demonstration. Calculations of entropic criterion of these three approximations straightforwardly gives the same ranking scheme as improved hard-sphere approximation is preferred over hard-sphere approximation and hard-sphere approximation is preferred over mean field approximation, as inferred by thoroughly but exhausting analysis based on our own knowledge and results against to computer simulations and experimental data.
Acknowledgement

This work is partially supported by grant NSC-94-2811-M-008-018 from National Science Council, ROC. Author is grateful for many discussions with colleagues Bao-Zhu Yang and Chien-Chih Chen regarding to practical applications of model selection method in genome and geophysics problems respectively.

References

[1] A. Caticha, Relative Entropy and Inductive Inference, in: G. Erickson, Y. Zhai (eds), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc. 707, 2004 (available from arXiv.org/abs/physics/0311093).
[2] A. E. Raftery, Sociological Methodology 25 (1995) 111.
[3] R. E. Weiss, J. Am. Stat. Ass. 90 (1995) 619.
[4] A. E. Raftery, D. M. adigan, J. A. Hoeting, J. Am. Stat. Ass. 92 (1997) 179.
[5] I. A. Kieseppa, Phil. Sci. 68 (2000) S141.
[6] F. Forbes, N. Peyrard, IEEE Trans. Pattern Analysis and Machine Intelligence, 25 (2003) 1089.
[7] B. V. Bonlander, A. E. Weigend, Selecting Input Variables Using Mutual Information and Nonparametric Density Estimation, in: Proc. of the 1994 Int. Symp. on Artificial Neural Networks, 42-50, 1994.
[8] J. A. Dupuis, C. P. Robert, J. Stat. Planning and Inference 111 (2003) 77.
[9] E. T. Jaynes, Phys. Rev. 106 (1957) 620; E. T. Jaynes Phys. Rev. 108 (1957) 171; E. T. Jaynes: Probability Theory: The Logic of Science, Cam bridge University Press (Cam bridge, 2003).
[10] J. E. Shore, R. W. Johnson, IEEE Trans. Inf. Theory IT-26 (1980) 26; J. E. Shore, R. W. Johnson, IEEE Trans. Inf. Theory IT-27 (1981) 472.
[11] J. Skilling, The Axiom of Maximum Entropy, in: G. J. Erickson, C. R. Smith (Eds), Maximum - Entropy and Bayesian Methods in Science and Engineering, Dordrecht, Kluwer, 1988; J. Skilling, Classic Maximum Entropy, in: J. Skilling (Ed), Maximum Entropy and Bayesian Methods, Dordrecht, Kluwer, 1989; J. Skilling, Quantized Maximum Entropy, in: F. P.ougere (Ed), Maximum Entropy and Bayesian Methods, Dordrecht, Kluwer, 1990.
[12] J. A. Barker, D. Henderson, Rev. Mod. Phys. 48 (1976) 587.
[13] J. P. Hansen, I. R. M c onal: () Theory of Simple Liquids, 2nd edition, A cad. Press (London, 1986).
[14] V. I. K allm anov: Statistical Physics of Fluids, Springer (New York, 2002).
[15] C. Y. Teng, A. Caticha, Maximum Entropy approach to a Mean Field Theory for Fluids, in: C. J. W illiam s (Ed), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc. 659, 2003 (available from arXiv.org/abs/cond-mat/0212198).
[16] G. A. M ancon, F. B. Can eli, J. Chem. Phys. 51 (1969) 4958.
[17] C.-Y. Tseng, A. Caticha, Maximum Entropy Approach to the Theory of Simple Fluids, in: G. Erickson and Y. Zhai (Eds), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc. 707, 2004 (available from arXiv.org/abs/cond-mat/0310746); Tseng, C.-Y. and Caticha, A. Maximum entropy and the Variational Method in Statistical Mechanics: an Application to Simple Fluids, in the process of reviewing for publication in Phys. Rev. E 2004 (available from arXiv.org/abs/cond-mat/0411625).

[18] J. D. Weeks, D. Chandler, H. C. Andersen, J. Chem. Phys. 54 (1971) 5237; J. D. Weeks, D. Chandler, H. C. Andersen, Science 220 (1983) 787.