# Markov processes follow from the principle of Maximum Caliber

Hao Ge[1,2],* Steve Pressé[3], Kingshuk Ghosh[4], and Ken A. Dill[5]
[1] *School of Mathematical Sciences and Centre for Computational Systems Biology Fudan University, Shanghai 200433, PRC*
[2] *Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.*
[3] *Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158*
[4] *Department of Physics and Astronomy, University of Denver, CO 80208*
[5] *Laufer Center for Physical and Quantitative Biology, Stony Brook University, NY 11794.*
(Dated: June 23, 2011)

Markov models are widely used to describe processes of stochastic dynamics. Here, we show that Markov models are a natural consequence of the dynamical principle of Maximum Caliber. First, we show that when there are different possible dynamical trajectories in a time-homogeneous process, then the only type of process that maximizes the path entropy, for any given singlet statistics, is a sequence of identical, independently distributed (i.i.d.) random variables, which is the simplest Markov process. If the data is in the form of sequentially pairwise statistics, then maximizing the caliber dictates that the process is Markovian with a uniform initial distribution. Furthermore, if an initial non-uniform dynamical distribution is known, or multiple trajectories are conditioned on an initial state, then the Markov process is still the only one that maximizes the caliber. Second, given a model, MaxCal can be used to compute the parameters of that model. We show that this procedure is equivalent to the maximum-likelihood method of inference in the theory of statistics.

PACS numbers:

## I. INTRODUCTION

E.T. Jaynes' principle of maximum entropy (MaxEnt) has wide applications in engineering and science [1–3], and serves as one description of the foundation for equilibrium statistical mechanics [7, 8]. In recent years, this principle has been generalized for treating time-dependent statistical phenomena, and is sometimes called the principle of maximum caliber (MaxCal) [5, 9–12].

Markov processes [14], whose transition densities are described by Chapman-Kolmogorov equations [6], are a common starting point in modeling stochastic dynamics. Here we justify from a maximum entropy standpoint "why start with a Markov process?"

In the present paper, we show that the application of MaxCal to time-homogeneous data precisely yields a Markov process. A Markov process model comes with a set of parameters, i.e., the transition probabilities. These are precisely related to the Lagrange multipliers generated by the principle of MaxCal. In determining these parameters, the MaxCal approach coincides with the method of maximum likelihood in the statistics of parameter estimation.

The theory of Maximum Caliber, in other words the idea of maximizing path entropy subject to constraints, was first tackled in the context of Markov processes by Filyukov and Karpov in 1967 [4], as far as the current authors are aware. In this work, it was assumed that a trajectory could be broken down as a Markov chain. Motivated by arguments from Khinchin, Filyukov and Karpov argued that maximizing the breadth of the trajectory ensemble based on data provided on the Markov transition probabilities was a recipe for inferring a trajectory ensemble consistent with observation. Similar reasoning has been used in recent work [9–13, 15, 16, 18].

Here we work in reverse and ask, given particular constraints is it possible to infer that the Markov model is the one that maximizes the entropy? The answer will turn out to be yes as maximum entropy is well known to set correlations between independent probabilities(or more generally, independent subsets of a map) to zero unless data provides evidence for correlation. This derives from axioms used in obtaining the entropy formula, namely "subset independence" [17, 21]; a property originally enforced by Shannon through his "composition property" as a logical consistency requirement for any inference[22]. Hence, given independent data on transition probabilities, a trajectory becomes the product of transition probabilities. We verify the Markov process as following from consistency conditions for path probabilities as noted by a. Kolmogorov[19] and maximum entropy.

*Electronic address: gehao@fudan.edu.cn

## II. THE MAXIMUM-CALIBER APPROACH

Suppose we have a stochastic process with a discrete state space $S = \{1, 2, \cdots, N\}$.

Consider its trajectories of length T, and denote the probability for the trajectory $\{i_0, i_1, \cdots, i_T\}$ as $p_{i_0 i_1 \cdots i_T}$ where $i_n$ is the state visited at time $n$. The path entropy, $S(T)$, is

$$S(T) = -\sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \log p_{i_0 i_1 \cdots i_T}. \qquad (1)$$

The above, Eq.( 1), is the correct measure which, when maximized subject to constraints, yields the least biased set $\{p_{i_0 i_1 \cdots i_T}\}$ consistent with observation [17, 21]. Furthermore, if $\{p_{i_0 i_1 \cdots i_T}\}$ is a path probability, then it must satisfy certain logical consistency requirements first established by A. Kolmogorov [19].

Specifically, for any $T$ equal 0 or any positive integer, the probability distribution has to satisfy

$$\sum_{i_{T+1}} p_{i_0 \cdots i_T i_{T+1}} = p_{i_0 \cdots i_T}, \qquad (2)$$

i.e. the distribution of the process during the time interval $\{0, 1, 2, \cdots, T\}$ should be the marginal distribution of the process during the time interval $\{0, 1, 2, \cdots, T+1\}$.

We consider the three different types of constraints.

**(1) Constraint on the mean of the singlet distribution.** First, we suppose we only know the mean number of time intervals during which the system dwells in each state $m$, $A_m$, in a trajectory of length T, where

$$A_m = \sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \sum_{k=0}^{T} \delta_{i_k, m}, \qquad (3)$$

where $\delta_{i,j} = 1$ when integers $i = j$, and 0 otherwise, $\sum_{k=0}^{T} \delta_{i_k, m}$ is the number of time intervals during which state $m$ is occupied in the trajectory $\{i_0, i_1, \cdots, i_T\}$, and $\sum_m A_m = T + 1$. We call this data singlet statistics.

Finally, we maximize Eq. (1) subject to constraints on our singlet statistic imposed as Lagrange multipliers. That is, we maximize the Caliber, $S(T) - \sum_m \lambda_m A_m(T)$ with respect to $p_{i_0 i_1 \cdots i_T}$. We conclude that

$$p_{i_0 i_1 \cdots i_T} \propto e^{-\sum_m \lambda_m \sum_{k=0}^{T} \delta_{i_k, m}} = \prod_{k=0}^{T} e^{-\lambda_{i_k}}. \qquad (4)$$

Thus including normalization the probability now becomes

$$\begin{aligned} p_{i_0 i_1 \cdots i_T} &= \frac{\prod_{k=0}^{T} e^{-\lambda_{i_k}}}{\sum_{i_0 i_1 \cdots i_T} \prod_{k=0}^{T} e^{-\lambda_{i_k}}} \\ &= \frac{\prod_{k=0}^{T} e^{-\lambda_{i_k}}}{(\sum_{i=1}^{N} e^{-\lambda_i})^{T+1}} = \prod_{k=0}^{T} p_{i_k}(T), \qquad (5) \end{aligned}$$

By definition

$$A_m = \sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \sum_{k=0}^{T} \delta_{i_k, m} = (T+1) p_m(T). \qquad (6)$$

As a final note on this example, we consider the result of imposing the consistency condition, Eq. 2, to the distribution obtained above. From this we obtain

$$\sum_{i_{T+1}} p_{i_0 \cdots i_T i_{T+1}} = \prod_{k=0}^{T} p_{i_k}(T+1), \qquad (7)$$

but since $\sum_{i_{T+1}} p_{i_0 \cdots i_T i_{T+1}} = p_{i_0 i_1 \cdots i_T} = \prod_{k=0}^{T} p_{i_k}(T)$, it then follows that

$$\prod_{k=0}^{T} p_{i_k}(T+1) = \prod_{k=0}^{T} p_{i_k}(T), \qquad (8)$$

which gives

$$p_{i_k}(T) = p_{i_k}(T+1) \qquad (9)$$

for any $i_k$. The proof is trivial if we were now to change $i_T$ to another state $i_T^{'}$. This would give us

$$\prod_{k=0}^{T-1} p_{i_k}(T+1) \cdot p_{i_T^{'}}(T+1) = \prod_{k=0}^{T-1} p_{i_k}(T) \cdot p_{i_T^{'}}(T) \qquad (10)$$

which, when compared with Eq. (8), yields

$$\frac{p_{i_T^{'}}(T+1)}{p_{i_T}(T+1)} = \frac{p_{i_T^{'}}(T)}{p_{i_T}(T)}.$$

Then since both the summation of all $p_i(T+1)$ and $p_i(T)$ are equal to one, hence Eq. (9) follows.

Under such constraints, MaxCal thus yields an identical, independent distributed (i.i.d.) process. Eq. (4) is the statement of independence and Eq. (9) is the statement of the identical distributions.

**(2) Constraint on pairwise statistics.** Now we consider instead a situation in which the constraint, instead, is on the pairwise statistics for each step $m \to n$ in the time period $[0, T]$, i.e.

$$A_{mn} = \sum_{i_0, \cdots, i_T} p_{i_0 \cdots i_T} \sum_{k=0}^{T-1} \delta_{i_k, m} \delta_{i_{k+1}, n}. \qquad (11)$$

Here $\sum_{k=0}^{T-1} \delta_{i_k, m} \delta_{i_{k+1}, n}$ is just the number of occurrence of the transition $m \to n$, and $\sum_{m,n} A_{mn} = T$.

Then maximizing Eq. (1) subject to constraints on the pairwise statistic imposed using Lagrange multipliers, we conclude that

$$p_{i_0 \cdots i_T} \propto e^{-\sum_{m,n} \lambda_{mn} \sum_{k=0}^{T-1} \delta_{i_k, m} \delta_{i_{k+1}, n}} = \prod_{k=0}^{T-1} p_{i_k i_{k+1}}, \qquad (12)$$

where $p_{i_k i_{k+1}} = e^{-\lambda_{i_k i_{k+1}}}$.

Next, applying the consistency condition, Eq. (2), we find by a method that is similar to the previous case (i.e. only replacing $i_T$ with $i_T^{'}$ and considering its difference from the original form) that

$$p_{i_k i_{k+1}}(T+1) = p_{i_k i_{k+1}}(T). \tag{13}$$

That is, under constraints on transition, the only stochastic process consistent with maximum caliber is the Markovian process with uniform initial distribution. The Markovian property is directly seen from the last proportionality in Eq. (12). We use the equation above to justify that transition probabilities are independent in the way that we used a similar expression to argue independence of the i.i.d. process when we had singlet statistics.

**(3) What if the initial distribution is not uniformly distributed?** If the initial distribution is not uniformly distributed, then the MaxCal method can still be used, but with an entropy conditioned on an initial state as follows

$$S(T|i_0) = -\sum_{i_1,\cdots,i_T} p_{i_1\cdots i_T|i_0} \log p_{i_1\cdots i_T|i_0}, \tag{14}$$

with constraints

$$\langle \text{number of transitions } m \to n|i_0\rangle = \sum_{i_1,\cdots,i_T} p_{i_1\cdots i_T|i_0} \sum_{k=0}^{T-1} \delta_{i_k,m}\delta_{i_{k+1},n} = A_{mn}. \tag{15}$$

Then applying the Lagrange multiplier method, we would also conclude that

$$p_{i_1\cdots i_T|i_0} = \prod_{k=0}^{T-1} p_{i_k i_{k+1}}.$$

Finally, we can easily derive a relation between $A_{mn}$, a property of the data, and $\lambda_{mn}$, the Lagrange multiplier. To do so, we define the *dynamical partition function*

$$Q_d(T) = \sum_{i_1,\cdots,i_T} e^{-\sum_{m,n}\lambda_{mn}\sum_{k=0}^{T-1}\delta_{i_k,m}\delta_{i_{k+1},n}}. \tag{16}$$

Taking the derivative of this dynamical partition function gives the average number of time intervals of dwell, $A_{lp} = -\partial \log Q_d(T)/\partial \log \lambda_{lp}$ (resembling the way that taking derivatives of equilibrium partition functions yield equilibrium averages and higher cumulants). When $T$ is large enough, we have $\sum_l A_{lp}(T)/T \to \pi_p$ where $\pi_p$ is the stationary distribution of state $p$.

## III. DATA ANALYSIS PROCEDURE

We have proved that the MaxCal approach gives a Markov chain. But, in practice, the constraints $A_{ij}(T)$ should be replaced by the sample estimators $\hat{A}_{ij}(T)$.

The unconditioned trajectory probabilities are simply a reweighted sum over the conditioned ones. By selecting the trajectories with the same initial states from an unconditioned ensemble of trajectories, one reproduces the distribution of the conditioned trajectory ensemble. Conversely, by recombining the trajectory with different initial states, one can obtain the distribution for the unconditioned trajectory ensemble.

To be precise, if we had no information regarding the distribution of initial conditions but if we had precise information regarding a particular initial state for a specific measurement, then we could only apply the conditional likelihood function

$$L(\{p_{ij}\}) = \prod_{k=0}^{T-1} p_{i_k i_{k+1}} = \prod_{i,j} p_{ij}^{\hat{A}_{ij}(T)},$$

with constraints $\sum_j p_{ij} = 1$ for each $j$.

Then applying the similar Langrange multiplier method like maximizing the Caliber, we set $H = L - \sum_i \lambda_i(\sum_j p_{ij} - 1)$, where $H$ is the likelihood subject to a constraint on normalization, and

$$\frac{\partial H}{\partial p_{ij}} = \hat{A}_{ij}(T)\frac{L(\{p_{ij}\})}{p_{ij}} - \lambda_i = 0,$$

followed

$$p_{ij} \propto \frac{\hat{A}_{ij}(T)}{\lambda_i},$$

i.e.

$$p_{ij} = \frac{\hat{A}_{ij}(T)}{\sum_j \hat{A}_{ij}(T)}.$$

Knowing an initial distribution $p_i(0)$, simply forces us to add some multiplicative constants in the likelihood function. The same holds for multiple trajectories, except that we must group together the trajectories with different initial states.

Thus, maximizing the conditional caliber in dynamical modeling is equivalent to maximizing the likelihood.

## IV. CONCLUSIONS AND DISCUSSION

Maximum caliber can be used both as a first principle from which to derive stochastic dynamical models and also as a data analysis method. We showed here that Markovian dynamics follows from MaxCal; the Markov property is a natural consequence of maximizing a dynamical entropy over trajectories for time-homogeneous processes. Our treatment can be generalized to handle dynamical systems having finite memory (time delays), corresponding to similar situations in inference for time-series analysis [24].

## V. ACKNOWLEDGEMENT

The authors would like to thank Prof. Hong Qian at the University of Washington for very helpful discussions. We acknowledge NIH grant GM 34993 and 1R01GM090205. In addition, H. Ge acknowledges support by the NSFC 10901040 and the specialized Research Fund for the Doctoral Program of Higher Education (New Teachers) 20090071120003. This article was completed when H. Ge visited the Xie group in the Department of Chemistry and Chemical Biology, Harvard University. H. Ge thanks Prof. Xiaoliang Sunney Xie and his department for their support and hospitality.

---

[1] Jaynes, E.T. 2003. *Probability Theory: The Logic of Science.* Cambridge Univ. Press, London, UK.

[2] Steinbach P.J. et al. 1992. Determination of rate distributions from kinetic experiments. *Biophys. J.*, 61: 235-245.

[3] Gull S. F., Daniell G.J. 1978. Image Reconstruction from incomplete and noisy data. *Nature*, 272: 686-690.

[4] Filyukov A. A., Karpov V. Ya. 1967. Method of the most probable path of evolution in the theory of stationary irreversible processes. *Inzhenerno-Fizicheskii Zhurnal*, 13: 798-804.

[5] Jaynes E.T. 1985. Macroscopic Prediction. *Complex Systems - Operational Approaches in Neurobiology, Physics, and Computers.* H. Haken(eds), Springer-Verlag, Berlin, pp. 254-269.

[6] van Kampen N.G. *Stochastic Processes in Chemistry and Physics.* North Holland Publishing Company(Amsterdam), 1981.

[7] Jaynes, E.T. (1957) Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630.

[8] Jaynes, E.T. (1957) Information theory and statistical mechanics II. *Phys. Rev.* **108**, 171– 190.

[9] Seitaridou, E., Inamdar, M.M., Phillips, R., Ghosh, K. and Dill, K.A. (2007) Measuring flux distributions for diffusion in the small-numbers limit. *J. Phys. Chem. B* **111**, 2288–2292.

[10] Pressé, S., Ghosh, K., Phillips, R. and Dill, K.A. (2010) Dynamical fluctuations in biochemical reactions and cycles S. *Physical Review E* **82**, 031905.

[11] Pressé, S., Ghosh, K. and Dill, K.A. (2011) Modeling Stochastic Dynamics in Biochemical Systems with Feedback Using Maximum Caliber. *J. Phys. Chem. B*, 115, 6202.

[12] Otten, M. and Stock, G. (2010): maximum caliber inference of nonequilibrium processes. *J. Chem. Phys.* 133, 034119

[13] G. Stock, K Ghosh, and K. A. Dill. (2008): Maximum Caliber: A variational approach applied to two-state dynamics. *J. Chem. Phys.* 128, 194102

[14] Chung, K.L. (1982) *Lectures from Markov processes to Brownian motion* Springer-Verlag, New York.

[15] Smith, E. (2011): Large-deviation principles, stochastic effective actions, path entropies, and the structure and meaning of thermodynamic descriptions. *Rep. Prog. Phys.* 74, 046601

[16] Monthus, C. (2011): Non-equilibrium steady states: maximization of the Shannon entropy associated with the distribution of dynamical trajectories in the presence of constraints. *J. Stat. Mech.* P03008

[17] Livesey A.K., Skilling J. *Acta Cryst.*, A41:113-122, 1985.

[18] Ghosh, K. (2011): Stochastic dynamics of complexation reaction in the limit of small numbers. *J. Chem. Phys.* 134, 195101

[19] Andrei N. Kolmogorov. Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin, 1933. English translation (1950): Foundations of the theory of probability. Chelsea, New York.

[20] Jiang, D.-Q., Qian, M. and Qian, M.-P. (2005) *Mathematical Theory of Nonequilibrium Steady-state.* LNM Vol 1833; Springer, New York.

[21] Shore J.E., Johnson R.W. *IEEE Trans. on Information Theory*, IT-26:26, 1980.

[22] Shannon C.E. *Bell Sys. Tech. J.*, 27:379, 1948.

[23] Ge, H. and Qian, H. (2010) The physical origins of entropy production, free energy dissipation and their mathematical representations. *Phys. Rev. E.* **81**, 051133.

[24] Brockwell P.J., Davis R.A.(2009) *Time Series: Theory and Methods. 2nd ed.* Springer-Verlag, New York.