

Article

Probabilistic Inference for Dynamical Systems

Sergio Davis ^{1,2,*}, Diego González ^{1,3}  and Gonzalo Gutiérrez ³¹ Comisión Chilena de Energía Nuclear, Casilla 188-D Santiago, Chile; dgonzalez@gnm.cl² Departamento de Física, Facultad de Ciencias Exactas, Universidad Andres Bello, Sazié 2212, Piso 7, 8370136 Santiago, Chile³ Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653 Santiago, Chile; gonzalo@fisica.ciencias.uchile.cl

* Correspondence: sergio.davis@cchen.cl

Received: 30 April 2018; Accepted: 6 September 2018; Published: 12 September 2018



Abstract: A general framework for inference in dynamical systems is described, based on the language of Bayesian probability theory and making use of the maximum entropy principle. Taking the concept of a path as fundamental, the continuity equation and Cauchy's equation for fluid dynamics arise naturally, while the specific information about the system can be included using the maximum caliber (or maximum path entropy) principle.

Keywords: dynamical systems; bayesian inference; fluid equations

1. Introduction

Dynamical system models are widely used to describe complex physical systems (e.g., the weather), as well as social and economic systems (e.g., the stock market). These systems are usually subject to high levels of uncertainty, either in their initial conditions and/or in their interactions with their environment. From the point of view of constructing predictive models, the optimal description of the time-dependent state of such a system given external constraints is a challenge with promising applications in both fundamental and applied science. This is of course an *inference problem* in which we must choose the most likely solution out of the (possibly infinite) alternatives compatible with the given information we have about the system.

Given all this, it seems that a unified framework for performing inference on dynamical systems may open new possibilities in several areas, including non-equilibrium statistical mechanics and thermodynamics, hydrodynamics (including magnetohydrodynamics), and classical mechanics under stochastic forces, among other possible fields of application. Of course, this vision of inference applied to dynamical systems is not new—the clearest exposition of the ideas that we aim to extend here was given by E. T. Jaynes [1,2], followed by several others [3–5].

In this work, we present some elements for a general framework of inference in dynamical systems, written in the language of Bayesian probability. The first is the master equation, which is shown as a direct consequence of the laws of probability. Next, we develop the treatment of inference over paths from which we obtain the continuity equation and Cauchy's equation for fluid dynamics, and discuss their range of applicability. Finally, we close with some concluding remarks.

2. Why Bayesian Inference?

Unlike the standard (“frequentist”) interpretation of probability theory, in which probabilities are frequencies of occurrence of repeatable events, Bayesian probability can be understood as the natural extension of classical logic in the case of *uncertainty* [6,7]. Bayesian probability deals with unknown quantities rather than identical repetitions/copies of an event or system, and is able to include *prior* information when needed.

The conceptual framework of Bayesian probability provides an elegant language to describe dynamical systems under uncertainty. A straightforward advantage of the Bayesian framework is that one does not need to assume an ensemble of “many identical copies of a system”. A single system with uncertain initial conditions and/or forces is sufficient to construct a theory. The probability of finding this particular system in a given state at a given time would not be a frequency, but rather a degree of plausibility conditioned on the known information. In fact, we can lift even the common assumption of “many degrees of freedom”. The motion of a single particle could be used to construct an internally consistent theory, where the equations of motion for time-dependent probability densities are similar to the ones of hydrodynamics. We will describe in detail both features of the Bayesian formulation in the following sections.

A brief overview of Bayesian notation used in this work follows. We will take $P(Q|\mathcal{I})$ as the probability of a particular *proposition* Q being true given knowledge \mathcal{I} . On the other hand, $\langle G \rangle_{\mathcal{I}}$ will denote the expected value of an arbitrary *quantity* G given knowledge \mathcal{I} , and will be given by

$$\langle G \rangle_{\mathcal{I}} = \sum_{\mathbf{u}} P(\mathbf{u}|\mathcal{I})G(\mathbf{u}), \tag{1}$$

where \mathbf{u} represents one of the possible states of the system.

3. Dynamical Evolution of Probabilities

Consider a discrete-time system that can transit between n possible states $\{x_1, \dots, x_n\}$ at different times. If we denote by $X(t)$ the state of the system at time t , we have that the joint probability of being in state a at time t and in state b at time $t + \Delta t$ is given by

$$P(X(t) = a, X(t + \Delta t) = b|\mathcal{I}) = P(X(t + \Delta t) = b|X(t) = a, \mathcal{I}) \cdot P(X(t) = a|\mathcal{I}). \tag{2}$$

By summing over a in Equation (2) and taking $b = x$, we have

$$P(X(t + \Delta t) = x|\mathcal{I}) = \sum_a P(X(t + \Delta t) = x|X(t) = a, \mathcal{I}) \cdot P(X(t) = a|\mathcal{I}), \tag{3}$$

while by summing over b and taking $a = x$, we have

$$P(X(t) = x|\mathcal{I}) = P(X(t) = x|\mathcal{I}) \sum_b P(X(t + \Delta t) = b|X(t) = x, \mathcal{I}). \tag{4}$$

From these two identities (Equations (3) and (4)), we can construct the discrete-time difference,

$$\begin{aligned} \frac{P(X(t + \Delta t) = x|\mathcal{I}) - P(X(t) = x|\mathcal{I})}{\Delta t} &= \sum_{x'} \left[\frac{P(X(t + \Delta t) = x|X(t) = x', \mathcal{I})}{\Delta t} \right] P(X(t) = x'|\mathcal{I}) \\ &\quad - P(X(t) = x|\mathcal{I}) \sum_{x'} \left[\frac{P(X(t + \Delta t) = x'|X(t) = x, \mathcal{I})}{\Delta t} \right]. \end{aligned} \tag{5}$$

This equation is a discrete-time form of the celebrated “master equation” [8–10], involving time-dependent transition probabilities. The case where $P(X(t + \Delta t) = x|X(t) = x', \mathcal{I})$ is independent of t (i.e., a function only of the initial state x' and final state x) is more commonly known as the master equation in the literature. In the continuous-time limit when $\Delta t \rightarrow 0$, we can write this equation as:

$$\frac{\partial \rho}{\partial t} = \sum_{x'} W_t(x' \rightarrow x)\rho(x'; t) - \rho(x; t) \sum_{x'} W_t(x \rightarrow x'), \tag{6}$$

where the instantaneous density ρ is given by $\rho(a; t) := P(X(t) = a|\mathcal{I})$ and we have defined the (continuous-time) *transition rate* W_t as

$$W_t(a \rightarrow b) := \lim_{\Delta t \rightarrow 0} \frac{P(X(t + \Delta t) = b | X(t) = a, \mathcal{I})}{\Delta t}. \tag{7}$$

In this sense, the master equation as written in Equation (5) is a direct consequence of the laws of probability, and its validity is universal whenever we have transitions between states. Please note that there is no requirement for the system to be Markovian: regardless of the form of the joint probability

$$P(X(t + \Delta t) = x, X(t) = x', X(t - \Delta t) = x'', \dots | \mathcal{I}),$$

there is always a *marginal* model

$$P(X(t + \Delta t) = x, X(t) = x' | \mathcal{I}) = \int dx'' \dots P(X(t + \Delta t) = x, X(t) = x', X(t - \Delta t) = x'', \dots | \mathcal{I}) \tag{8}$$

which yields the transition probability in Equation (7) as:

$$P(X(t + \Delta t) = x | X(t) = x', \mathcal{I}) = \frac{P(X(t + \Delta t) = x, X(t) = x' | \mathcal{I})}{\int d\eta P(X(t + \Delta t) = \eta, X(t) = x' | \mathcal{I})}. \tag{9}$$

It is for this probability that Equation (5) holds. In general, the transition rate W_t will most probably be time-dependent, due to the fact that it captures the dependence of the previous history of the system up to t . It follows from this that all probabilities of time-dependent quantities must evolve in time according to Equation (5) (or (6) in the case of continuous time) for some (possibly time-dependent) transition probability (rate). This continuous-time master equation (Equation (6)) is more general than the continuity equation, as it includes the case where some quantities can be created or destroyed during a process. However, time evolution under global and local conservation laws is a fundamental case that can also be readily obtained from the Bayesian formalism, as we will see in the following sections. As is well-known, the continuous-time master equation can be approximated in the limit of infinitesimally small transitions to obtain the Fokker–Planck equation [9,11], but in the next section we start from the existence of continuous paths as a postulate.

4. Fluid Theories in a Bayesian Formulation

We will now consider a dynamical system that follows a path $\mathbf{X}(t) \in \mathbb{X}$ in time, where \mathbb{X} denotes the space of all paths consistent with given boundary conditions. The path $\mathbf{X}(t)$ is not completely known, and we only have access to partial information denoted by \mathcal{I} .

In this setting, Bayesian theory defines a functional $P[\mathbf{X} | \mathcal{I}]$ that is the probability density of the path $\mathbf{X}(t)$ being the “true path” under the known information. For any arbitrary *functional* $\mathcal{F}[\mathbf{X}]$ of the path, we can then write its expected value as a path integral:

$$\langle \mathcal{F} \rangle_{\mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}\mathbf{X} P[\mathbf{X} | \mathcal{I}] \mathcal{F}[\mathbf{X}]. \tag{10}$$

On the other hand, the expected value of any instantaneous *quantity* $A(\mathbf{x}; t)$ is given by

$$\langle A \rangle_{t, \mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}\mathbf{X} P[\mathbf{X} | \mathcal{I}] A(\mathbf{X}(t); t) = \int_V dx \rho(\mathbf{x}; t) A(\mathbf{x}; t), \tag{11}$$

where $\rho(\mathbf{x}; t) := P(\mathbf{X}(t) = \mathbf{x} | \mathcal{I})$ is the instantaneous probability density at time t .

By using a quantity $A(\mathbf{x}'; t) = \delta(\mathbf{x}' - \mathbf{x})$, we see that the probability density itself has a path integral representation:

$$\rho(\mathbf{x}; t) = \int_{\mathbb{X}} \mathcal{D}\mathbf{X} P[\mathbf{X} | \mathcal{I}] \delta(\mathbf{X}(t) - \mathbf{x}). \tag{12}$$

By differentiating Equation (12) with respect to time, we obtain the continuity equation for the instantaneous probability density [12]:

$$\begin{aligned}
 \frac{\partial \rho}{\partial t} &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \frac{\partial}{\partial t} \delta(\mathbf{X}(t) - \mathbf{x}) \\
 &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \left(\nabla \delta(\mathbf{X}(t) - \mathbf{x}) \right) \cdot \dot{\mathbf{X}}(t) \\
 &= -\nabla_{\mathbf{x}} \cdot \left(\int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \delta(\mathbf{X}(t) - \mathbf{x}) \dot{\mathbf{X}}(t) \right) \\
 &= -\nabla_{\mathbf{x}} \cdot \left(\rho \cdot \mathbf{v}(\mathbf{x}; t) \right),
 \end{aligned}
 \tag{13}$$

where $\mathbf{v}(\mathbf{x}; t)$ is the velocity field that describes the flow of probability, given by

$$\mathbf{v}(\mathbf{x}; t) = \left\langle \dot{\mathbf{X}}(t) \right\rangle_{\mathbf{X}(t)=\mathbf{x}, \mathcal{I}} = \frac{1}{\rho(\mathbf{x}; t)} \left\langle \dot{\mathbf{X}}(t) \delta(\mathbf{X}(t) - \mathbf{x}) \right\rangle_{\mathcal{I}}.
 \tag{14}$$

This equation describes the global and local conservation of the probability of finding the system in a given state \mathbf{x} at a time t , and is guaranteed to hold for any system moving continuously in time through paths $\mathbf{X}(t) \in \mathbb{X}$. In the same way, it is possible to derive a dynamical equation for the velocity field $\mathbf{v}(\mathbf{x}; t)$ itself, by differentiating it with respect to time. We have:

$$\begin{aligned}
 \frac{\partial}{\partial t} (\rho \cdot v_{\mu}) &= \frac{\partial}{\partial t} \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \\
 &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \frac{\partial}{\partial t} \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \\
 &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \left(\delta(\mathbf{X}(t) - \mathbf{x}) \ddot{X}_{\mu}(t) + \partial_{\nu} \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \dot{X}_{\nu}(t) \right) \\
 &= \left\langle \delta(\mathbf{X}(t) - \mathbf{x}) \ddot{X}_{\mu}(t) \right\rangle_{\mathcal{I}} - \partial_{\nu} \left\langle \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \dot{X}_{\nu}(t) \right\rangle_{\mathcal{I}} \\
 &= \rho a_{\mu}(\mathbf{x}; t) - \partial_{\nu} (\rho [C_{\mu\nu} + v_{\mu}v_{\nu}]),
 \end{aligned}
 \tag{15}$$

where in the last line we have defined the acceleration field as

$$\mathbf{a}(\mathbf{x}; t) := \left\langle \ddot{\mathbf{X}}(t) \right\rangle_{\mathbf{X}(t)=\mathbf{x}, \mathcal{I}}
 \tag{16}$$

and the velocity covariance matrix

$$C_{\mu\nu}(\mathbf{x}; t) = \left\langle \dot{X}_{\mu}(t) \dot{X}_{\nu}(t) \right\rangle_{\mathbf{X}(t)=\mathbf{x}, \mathcal{I}} - v_{\mu}(\mathbf{x}; t) v_{\nu}(\mathbf{x}; t).
 \tag{17}$$

By using the continuity equation (Equation (13)) to rewrite the left-hand side as

$$\begin{aligned}
 \frac{\partial}{\partial t} (\rho \cdot v_{\mu}) &= \frac{\partial \rho}{\partial t} v_{\mu} + \rho \frac{\partial v_{\mu}}{\partial t} \\
 &= -\nabla_{\mathbf{x}} \cdot (\rho \cdot \mathbf{v}) v_{\mu} + \rho \frac{\partial v_{\mu}}{\partial t} \\
 &= -\partial_{\nu} (\rho v_{\nu}) v_{\mu} + \rho \frac{\partial v_{\mu}}{\partial t},
 \end{aligned}
 \tag{18}$$

we obtain, dividing both sides by ρ and using Equation (15), that

$$\frac{\partial v_{\mu}}{\partial t} - (\partial_{\nu} v_{\nu}) v_{\mu} - v_{\nu} v_{\mu} (\partial_{\nu} \ln \rho) = a_{\mu} - \partial_{\nu} (C_{\mu\nu} + v_{\mu}v_{\nu}) - (C_{\mu\nu} - v_{\mu}v_{\nu}) \partial_{\nu} \ln \rho.
 \tag{19}$$

The term $v_{\nu} v_{\mu} \partial_{\nu} \ln \rho$ cancels, and we have:

$$\frac{\partial v_{\mu}}{\partial t} - (\partial_{\nu} v_{\nu}) v_{\mu} = a_{\mu} - \partial_{\nu} C_{\mu\nu} - \partial_{\nu} (v_{\mu}v_{\nu}) - C_{\mu\nu} \partial_{\nu} \ln \rho,
 \tag{20}$$

from which we now cancel the term $(\partial_\nu v_\nu)v_\mu$, arriving at

$$\frac{\partial v_\mu}{\partial t} = a_\mu - \partial_\nu C_{\mu\nu} - v_\nu \partial_\nu v_\mu - C_{\mu\nu} \partial_\nu \ln \rho. \quad (21)$$

Rearranging the derivatives of v_μ in the left-hand side, we have

$$\begin{aligned} \left(\frac{\partial}{\partial t} + v_\nu \partial_\nu \right) v_\mu &= a_\mu - \partial_\nu C_{\mu\nu} - C_{\mu\nu} \partial_\nu \ln \rho \\ &= a_\mu - \frac{1}{\rho} \partial_\nu (\rho C_{\mu\nu}) \\ &= a_\mu + \frac{1}{\rho} \partial_\nu \sigma_{\mu\nu}, \end{aligned} \quad (22)$$

which is the Cauchy momentum equation

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla \right) v = \frac{Dv}{Dt} = a(x;t) + \frac{1}{\rho} \nabla \cdot \overleftarrow{\sigma}, \quad (23)$$

with D/Dt the *convective derivative* and $\overleftarrow{\sigma} = -\rho \cdot \overleftarrow{C}$ the stress tensor.

Equations (13) and (23) form a closed coupled system of equations for $\rho(x;t)$ and $v(x;t)$, needing as their only external input the velocity covariance matrix $C_{\mu\nu}$. These equations are then *built-in features* of inference over paths. In a Bayesian approach, they are valid for any system that moves continuously in time. The Cauchy momentum equation includes most notably the Navier–Stokes equation as a particular case [13].

5. Including Particular Knowledge into Our Models

At this point, we have developed a generic framework where no particular details about a system have been included. Clearly all those details have to be contained in $P[\mathbf{X}|\mathcal{I}]$, or rather, in the covariance matrix $C_{\mu\nu}(x;t)$ which can be derived from it. The question remains about how to incorporate these details in the most unbiased manner. In principle, we could start from the *null* assumption of equiprobable paths,

$$P[\mathbf{X}|\mathcal{I}_0] = \text{constant},$$

and add new information \mathcal{R} later on, by updating our probability functional $P[\mathbf{X}|\mathcal{I}_0]$ to a new $P[\mathbf{X}|\mathcal{I}]$, where $\mathcal{I} = (\mathcal{I}_0, \mathcal{R})$. There are essentially two equivalent methods to achieve this, and depending on the actual form of \mathcal{R} , one of them may be more directly applicable than the other.

(1) **Bayes' theorem:** the posterior distribution $P(\mathbf{u}|\mathcal{I}_0, \mathcal{R})$ is given in terms of the prior $P(\mathbf{u}|\mathcal{I}_0)$ by

$$P(\mathbf{u}|\mathcal{I}_0, \mathcal{R}) = \frac{P(\mathbf{u}|\mathcal{I}_0) \cdot P(\mathcal{R}|\mathbf{u}, \mathcal{I}_0)}{P(\mathcal{R}|\mathcal{I}_0)}.$$

This method is most useful when \mathcal{R} is comprised of statements about the states \mathbf{u} (e.g., boundary conditions).

(2) **Principle of maximum entropy:** the posterior distribution $p(\mathbf{u})$ is the one that maximizes

$$\mathcal{S}[p_0 \rightarrow p] = - \int_V d\mathbf{u} p(\mathbf{u}) \ln \frac{p(\mathbf{u})}{p_0(\mathbf{u})},$$

where $p_0(\mathbf{u})$ is the prior distribution. This method is most useful when \mathcal{R} consists of constraints on the final model $p(\mathbf{u})$, usually expressed as fixed expected values.

In Reference [2], Jaynes assumes the continuity equation from the start and derives the flux $\mathbf{J}(x;t) = \rho \cdot v(x;t)$ from symmetry considerations, the central limit theorem, and Bayes theorem.

In our classification, this corresponds to method (1). In Reference [5], Gull recovers Brownian motion by essentially performing discrete-time maximum caliber inference under constraints over location and particle speed, hence corresponding to an application of method (2).

6. The Maximum Caliber Principle

The function $p(\mathbf{u})$ that is closest to our prior probability $p_0(\mathbf{u})$ and is consistent with the constraints \mathcal{R} is the one that maximizes the relative entropy [14,15]

$$\mathcal{S}[p_0 \rightarrow p] = - \int_V d\mathbf{u} p(\mathbf{u}) \ln \frac{p(\mathbf{u})}{p_0(\mathbf{u})}$$

among the set of functions p that are compatible with \mathcal{R} . The negative of this relative entropy, known as the Kullback–Leibler divergence, is commonly used to measure the “informational distance” from p_0 to p . It is important to note that this is a rule of inference and not a physical principle, and therefore it is not bounded by the meaning assigned to the states \mathbf{x} , as long as we can write (Bayesian) probabilities over them.

For the general case of m constraints of the form

$$\langle f_j(\mathbf{u}) \rangle = F_j, \quad (j = 1, \dots, m),$$

the maximum entropy solution starting from $P(\mathbf{u}|\mathcal{I}_0)$ is obtained through the use of m Lagrange multipliers (one for each constraint),

$$P(\mathbf{u}|\mathcal{I}_0, \mathcal{R}) = P(\mathbf{u}|\mathcal{I}_0) \left[\frac{\exp \left(- \sum_{j=1}^m \lambda_j f_j(\mathbf{x}) \right)}{Z(\boldsymbol{\lambda})} \right], \tag{24}$$

where $Z(\boldsymbol{\lambda})$ is the partition function. This is compatible with Bayesian updating, as this posterior distribution is proportional to the prior. The Lagrange multipliers are solutions of the constraint equations in terms of Z :

$$F_j = \frac{\partial}{\partial \lambda_j} \ln Z(\boldsymbol{\lambda}) \quad (j = 1, \dots, m). \tag{25}$$

In exactly the same way, the path (relative) entropy (sometimes known as the *caliber*) is defined as the path integral [1,16–22]:

$$\mathcal{S}[p_0 \rightarrow p] = - \int_{\mathbb{X}} D\mathbf{X} p[\mathbf{X}] \ln \frac{p[\mathbf{X}]}{p_0[\mathbf{X}]},$$

where $p_0[\mathbf{X}] := P[\mathbf{X}|\mathcal{I}_0]$ is the prior path probability. The use of this generalization is justified based on the fact that we can write any path $\mathbf{X}(t)$ in terms of a complete orthonormal basis $\{\mathbf{B}_i\}$,

$$\mathbf{X}(t) = \sum_{n=0}^{N-1} \gamma_n \mathbf{B}_n(t), \quad N \rightarrow \infty, \tag{26}$$

and then there is a one-to-one correspondence between every path $\mathbf{X}(t)$ and its coordinates $(\gamma_0, \gamma_1, \dots, \gamma_{N-1})$. Inference over paths \mathbf{X} then becomes completely equivalent to inference over the coefficients γ , which form a system with N degrees of freedom.

In summary, for the general maximum caliber inference problem we have m constraints, written as

$$\langle \mathcal{F}_j[\mathbf{X}] \rangle = F_j \quad (j = 1, \dots, m), \tag{27}$$

from which the probability functional obtained is

$$P[\mathbf{X}|\mathcal{I}_0, \mathcal{R}] = \frac{1}{Z(\boldsymbol{\lambda})} P[\mathbf{X}|\mathcal{I}_0] \exp\left(-\sum_{j=1}^m \lambda_j \mathcal{F}_j[\mathbf{X}]\right). \quad (28)$$

Any such maximum caliber solution can be cast in the “canonical” form, as

$$P[\mathbf{X}|\mathcal{I}] = \frac{1}{Z(\alpha)} \exp\left(-\frac{1}{\alpha} \mathcal{A}[\mathbf{X}]\right), \quad (29)$$

where \mathcal{A} is a functional, analogous to the Hamilton action of a classical system, and $\alpha > 0$ is a constant with the same physical units as \mathcal{A} . By simple inspection of this canonical form, it is straightforward to see that the most probable path is the one with minimum action:

$$\frac{\delta \mathcal{A}[\mathbf{X}]}{\delta \mathbf{X}(t)} = 0, \quad \forall t. \quad (30)$$

7. An Illustration: Newtonian Mechanics of Charged Particles

As an example of the application of this formalism, consider a “particle” with known square speed v^2 , known instantaneous probability density ρ , and known velocity field $v(\mathbf{x}; t)$ for all times $t \in [0, \tau]$. The corresponding constraints are then

$$\left\langle \left| \dot{\mathbf{X}}(t) \right|^2 \right\rangle = v^2(t) \quad \forall t, \quad (31)$$

$$\left\langle \delta(\mathbf{X}(t) - \mathbf{x}) \right\rangle = \rho(\mathbf{x}; t) \quad \forall \mathbf{x}, t, \quad (32)$$

$$\left\langle \dot{\mathbf{X}}(t) \delta(\mathbf{X}(t) - \mathbf{x}) \right\rangle = \rho(\mathbf{x}; t) v(\mathbf{x}; t) \quad \forall \mathbf{x}, t. \quad (33)$$

The resulting maximum caliber solution is of the form

$$P[\mathbf{X}|\mathcal{I}] = \frac{1}{Z(\alpha)} \exp\left(-\frac{1}{\alpha} \int_0^\tau dt L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t)\right), \quad (34)$$

with the Hamilton action

$$\mathcal{A}[\mathbf{X}] = \int_0^\tau dt L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t), \quad (35)$$

and a Lagrangian defined as

$$L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t) = \lambda_1(t) |\dot{\mathbf{X}}(t)|^2 + \int_V d\mathbf{x} (\lambda_2(\mathbf{x}; t) + \lambda_3(\mathbf{x}; t) \cdot \dot{\mathbf{X}}(t)) \delta(\mathbf{X}(t) - \mathbf{x}), \quad (36)$$

where λ_1 , λ_2 , and λ_3 are Lagrange multipliers. This Lagrangian can be cast into a more familiar form,

$$L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t) = \frac{1}{2} m(t) \dot{\mathbf{X}}(t)^2 - \Phi(\mathbf{X}(t); t) + \dot{\mathbf{X}}(t) \cdot \mathbf{A}(\mathbf{X}(t); t), \quad (37)$$

by simply renaming the Lagrange multipliers and integrating the delta function [18]. Interestingly, this is none other than the Lagrangian for a particle with time-dependent “mass” $m(t)$ in an external “electromagnetic potential” (Φ, \mathbf{A}) . The most probable path under these constraints is determined by the solution of the Euler–Lagrange equation, which reduces to Newton’s second law under a “Lorenz force”,

$$\frac{d}{dt} (m\mathbf{v}) = -\nabla\Phi(\mathbf{x}; t) - \frac{\partial \mathbf{A}(\mathbf{x}; t)}{\partial t} + \mathbf{v} \times (\nabla \times \mathbf{A}(\mathbf{x}; t)), \quad (38)$$

as shown in the Appendix. In particular, it is important to note that it is the constraint on the squared speed $v^2(t)$ that adds the mass $m(t)$ to the model, as $m = 2\lambda_1$, the constraint on the probability density

$\rho(x; t)$ adds the scalar potential $\Phi(x; t)$ to the model, as $\Phi = -\lambda_2$, and finally the constraint on the local velocity field $v(x; t)$ adds the vector potential $A(x; t)$ to the model, because $A = \lambda_3$. Nowhere in the derivation of this Lagrangian have we assumed the existence of charges, electromagnetic fields, or the Lorenz force. The structure that is revealed is the most unbiased under the constraints given in Equations (31) to (33), that is, with approximate knowledge of its location (given by ρ) and velocity “field lines” (given by v). This model could be used for people in a busy street crossing, or vehicles in a city.

8. Concluding Remarks

We have shown that it is possible to construct a fluid theory from Bayesian inference of an abstract system with N degrees of freedom moving along paths $X(t)$, and that this theory automatically includes the continuity equation and the Cauchy momentum equation as built-in features. Moreover, through the use of the Maximum Caliber principle, it is possible to formulate the dynamics of such an abstract system in terms of an action that is minimal for the most probable path, resembling the well-known structures of Lagrangian and Hamiltonian mechanics.

By entering the square speed, instantaneous probability density, and velocity field into our model, a Lagrangian of a “particle” under external fields emerges naturally. This “particle” moves on average according to Newton’s law of motion $F = m \cdot a$ under the Lorenz force, with scalar and vector potentials determined by the known information about the location and velocity lines. In this formulation, the only ingredients that we could call *physical* were the existence of an N -dimensional “particle” moving continuously along (unknown) paths. In this application, we see that position and velocity are the only intrinsic (*real* or *ontological*) properties of the particle at a given time t . On the other hand, the time-dependent mass $m(t)$ and the fields $\Phi(x; t)$ and $A(x; t)$ are emergent parameters (in fact, Lagrange multipliers) needed to impose the constraints on the known information used to construct the model.

Author Contributions: All authors of this paper contributed to the work. S.D. conceived the structure of the paper while D.G. and G.G. provided analysis of previous results and discussions.

Funding: This research was funded by FONDECYT grant 1171127.

Acknowledgments: S.D. and G.G. gratefully acknowledge funding from FONDECYT 1171127.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Derivation of the Lorenz Force from the Lagrangian of a Particle in an Electromagnetic Field

We start from the Euler–Lagrange equation

$$\frac{\partial L}{\partial x_\mu} - \frac{d}{dt} \left(\frac{\partial L}{\partial v_\mu} \right) = 0 \quad (\text{A1})$$

for the Lagrangian in Equation (37),

$$L(x, v; t) = \frac{1}{2} m(t) v^2 - \Phi(x; t) + v_\nu A_\nu(x; t),$$

which gives

$$-\partial_\mu \Phi + v_\nu \partial_\mu A_\nu - \frac{d}{dt} (m v_\mu + A_\mu) = 0. \quad (\text{A2})$$

Replacing the total time derivative of the vector field $A(x; t)$,

$$\frac{d}{dt} A_\mu(x; t) = \partial_t A_\mu + v_\nu \partial_\nu A_\mu,$$

we have

$$\frac{d}{dt}(mv_\mu) + \partial_t A_\mu + v_\nu \partial_\nu A_\mu = -\partial_\mu \Phi + v_\nu \partial_\mu A_\nu, \quad (\text{A3})$$

which can be written as

$$\frac{d}{dt}(mv_\mu) = (-\partial_\mu \Phi - \partial_t A_\mu) + v_\nu (\partial_\mu A_\nu - \partial_\nu A_\mu). \quad (\text{A4})$$

Using the fact that the i -th component of $\mathbf{v} \times (\nabla \times \mathbf{A})$ is given by

$$\begin{aligned} [\mathbf{v} \times (\nabla \times \mathbf{A})]_i &= \epsilon_{ijk} v_j (\epsilon_{klm} \partial_l A_m) \\ &= \epsilon_{ijk} \epsilon_{klm} (v_j \partial_l A_m) \\ &= (\delta_{il} \delta_{jm} - \delta_{im} \delta_{jl}) v_j \partial_l A_m \\ &= v_j (\partial_i A_j - \partial_j A_i), \end{aligned} \quad (\text{A5})$$

we finally obtain

$$\frac{d}{dt}(mv_\mu) = (-\partial_\mu \Phi - \partial_t A_\mu) + [\mathbf{v} \times (\nabla \times \mathbf{A})]_\mu, \quad (\text{A6})$$

which is Equation (38).

References

- Jaynes, E.T. The Minimum Entropy Production Principle. *Ann. Rev. Phys. Chem.* **1980**, *31*, 579–601. [CrossRef]
- Jaynes, E.T. Clearing up Mysteries—The Original Goal. In *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*; Springer: Dordrecht, The Netherlands, 1989; pp. 1–27.
- Dewar, R. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *J. Phys. A Math. Gen.* **2003**, *36*, 631–641. [CrossRef]
- Grandy, W.T. *Entropy and the Time Evolution of Macroscopic Systems*; Oxford Science Publications: New York, NY, USA, 2008.
- Gull, S. Some Misconceptions about Entropy. Available online: <http://www.mrao.cam.ac.uk/steve/maxent2009/images/miscon.pdf> (accessed on 11 September 2018).
- Cox, R.T. Probability, frequency and reasonable expectation. *Am. J. Phys.* **1946**, *14*, 1–13. [CrossRef]
- Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
- Risken, H. *The Fokker-Planck Equation: Methods of Solution and Applications*; Springer: Berlin, Germany, 1996.
- Van Campen, N.G. *Stochastic Processes in Physics and Chemistry*; North Holland: Amsterdam, The Netherlands, 2007.
- Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, NY, USA, 2001.
- Van Campen, N.G. The expansion of the master equation. *Adv. Chem. Phys.* **1976**, *34*, 245–309.
- González, D.; Díaz, D.; Davis, S. Continuity equation for probability as a requirement of inference over paths. *Eur. Phys. J. B* **2016**, *89*, 214. [CrossRef]
- Lamb, H. *Hydrodynamics*; Dover Books on Physics; Dover Publications: New York, NY, USA, 1945.
- Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
- Caticha, A.; Giffin, A. Updating Probabilities. *AIP Conf. Proc.* **2006**, *872*, 31.
- Pressé, S.; Ghosh, K.; Lee, J.; Dill, K.A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115–1141. [CrossRef]
- González, D.; Davis, S.; Gutiérrez, G. Newtonian mechanics from the principle of Maximum Caliber. *Found. Phys.* **2014**, *44*, 923. [CrossRef]
- Davis, S.; González, D. Hamiltonian Formalism and Path Entropy Maximization. *J. Phys. A Math. Theor.* **2015**, *48*, 425003. [CrossRef]
- Hazoglou, M.J.; Walther, V.; Dixit, P.D.; Dill, K.A. Maximum caliber is a general variational principle for nonequilibrium statistical mechanics. *J. Chem. Phys.* **2015**, *143*, 051104. [CrossRef] [PubMed]

20. Wan, H.; Zhou, G.; Voeltz, V.A. A maximum-caliber approach to predicting perturbed folding kinetics due to mutations. *J. Chem. Theory Comput.* **2016**, *12*, 5768–5776. [[CrossRef](#)] [[PubMed](#)]
21. Cafaro, C.; Ali, S.A. Maximum caliber inference and the stochastic Ising model. *Phys. Rev. E* **2016**, *94*, 052145. [[CrossRef](#)] [[PubMed](#)]
22. Dixit, P.; Wagoner, J.; Weistuch, C.; Pressé, S.; Ghosh, K.; Dill, K.A. Perspective: Maximum caliber is a general variational principle for dynamical systems. *J. Chem. Phys.* **2018**, *148*, 010901. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).