

Simple unity among the fundamental equations of science

Steven A. Frank*

The Price equation describes the change in populations. Change concerns some value, such as biological fitness, information or physical work. The Price equation reveals universal aspects for the nature of change, independently of the meaning ascribed to values. By understanding those universal aspects, we can see more clearly why fundamental mathematical results in different disciplines often share a common form. We can also interpret more clearly the meaning of key results within each discipline. For example, the mathematics of natural selection in biology has a form closely related to information theory and physical entropy. Does that mean that natural selection is about information or entropy? Or do natural selection, information and entropy arise as interpretations of a common underlying abstraction? The Price equation suggests the latter. The Price equation achieves its abstract generality by partitioning change into two terms. The first term naturally associates with the direct forces that cause change. The second term naturally associates with the changing frame of reference. In the Price equation's canonical form, total change remains zero because the conservation of total probability requires that all probabilities invariantly sum to one. Much of the shared common form for the mathematics of different disciplines may arise from that seemingly trivial invariance of total probability, which leads to the partitioning of total change into equal and opposite components of the direct forces and the changing frame of reference.

Keywords: Price equation, invariance, conservation laws, information theory, physical mechanics

Introduction

Problems often concern change. How does natural selection alter a population? How does force change position and velocity? How does climate affect biodiversity?

Many forces act on a system. We cannot know all of them. The Price equation helps by partitioning change into components. One component isolates particular forces. The second component includes everything else.

For a system that changes in a specific way, the equation does not tell us which forces to isolate. We could focus on climate. Or we could focus on a meteor explosion. No matter what forces we isolate, the overall system changes in the same way.

The Price equation's isolation of forces helps because it "focuses attention on the forces, not on the moving body"¹. Force associates with cause.

We may be more interested in understanding cause rather than describing motion.

Different fields focus on different forces. Yet, no matter the focus, disciplines often share a common underlying expression of force and a common partitioning of total change into components.

That unity arises from a simple conservation law for total change. The Price equation partitions the conserved total into a component for the direct forces and a component for the changing frame of reference.

From the Price equation's abstract partitioning, one obtains many of the fundamental mathematical equations of change that recur across disciplines. For example, common expressions of information arise as a simple invariant measurement scale of geometric divergence caused by the forces of change.

In prior articles, I used the Price equation to show the unity among the fundamental equations of many different fields of science^{2,3}. This article emphasizes the basic concepts and geometric intu-

* Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA

web: <https://stevefrank.org>

ition[†].

The Price equation

I begin with an abstract mathematical derivation of the Price equation. I then connect the abstract mathematical equations of change to particular examples.

The examples in later sections include the work done by physical forces, the information gained by the force of natural selection, and the updated Bayesian inference achieved by the “force” of new data. See Frank² for further examples.

The Price equation describes the change in the average value of some property between two populations^{4,5}. A population is a set of things. Each thing has a property indexed by i . Those things with a common property index comprise a fraction, q_i , of the population and have average value, z_i , for whatever we choose to measure by z . See Table 1 for a summary of notation.

Write \mathbf{q} and \mathbf{z} as the vectors over all i . The population average value is $\bar{z} = \mathbf{q} \cdot \mathbf{z} = \sum q_i z_i$, summed over i .

A second population has matching vectors \mathbf{q}' and \mathbf{z}' . Here, q'_i is the fraction of the second population derived from entities with index i in the first population. Similarly, z'_i is the average value in the second population of members derived from entities with index i in the first population. Let Δ be the difference between the derived population and the original population, $\Delta\mathbf{q} = \mathbf{q}' - \mathbf{q}$ and $\Delta\mathbf{z} = \mathbf{z}' - \mathbf{z}$.

The difference in the averages is $\Delta\bar{z} = \mathbf{q}' \cdot \mathbf{z}' - \mathbf{q} \cdot \mathbf{z}$. By using the definitions for $\Delta\mathbf{q}$ and $\Delta\mathbf{z}$, we can write the change in the average as the abstract form of the Price equation

$$\Delta\bar{z} = \Delta\mathbf{q} \cdot \mathbf{z} + \mathbf{q}' \cdot \Delta\mathbf{z}. \quad (1)$$

The first term, $\Delta\mathbf{q} \cdot \mathbf{z}$, is the partial difference of \mathbf{q} holding \mathbf{z} constant. The second term, $\mathbf{q}' \cdot \Delta\mathbf{z}$, is the partial difference of \mathbf{z} holding \mathbf{q} constant. In the second term, we use \mathbf{q}' as the constant value because, with discrete differences, one of the partial change terms must be evaluated in the context of

the second set.

Note that \mathbf{q} has a clearly defined meaning as frequency, whereas \mathbf{z} may be chosen arbitrarily as any values assigned to members. The values, \mathbf{z} , define the frame of reference. Because frequency is clearly defined, whereas values are arbitrary, the frequency changes, $\Delta\mathbf{q}$, take on the primary role in analyzing the structural aspects of change that unify different subjects.

The primacy of frequency change naturally labels the first term, with $\Delta\mathbf{q}$, as the changes caused by the direct forces acting on populations. Because \mathbf{q} and \mathbf{q}' define a sequence of probability distributions, the primary aspect of change concerns the dynamics of probability distributions.

The arbitrary aspect of the values, \mathbf{z} , naturally labels the second term, with $\Delta\mathbf{z}$, as the changes caused by the forces that alter the frame of reference. Those forces that change the frame of reference are sometimes called the inertial forces¹.

It is, of course, possible to interpret the terms in other ways. The equation itself is a simple mathematical identity. That identity has no intrinsic meaning beyond the fundamental partitioning of the change in an average value, $\Delta\bar{z}$, into two components of change.

The Price equation is often written as

$$\Delta z = \text{Cov}(w, z) + E(w\Delta z),$$

in which w is relative fitness, as defined below eqn 2. This expression is equivalent to eqn 1, because $\text{Cov}(w, z) = \Delta\mathbf{q} \cdot \mathbf{z}$ and $E(w\Delta z) = \mathbf{q}' \cdot \Delta\mathbf{z}$. The “E” means the expectation or average. I focus on the form in eqn 1, with vectors and dot products. That form emphasizes the geometry, with natural interpretations in terms of force, distance and constraints imposed by conserved quantities. The commonly used statistical expressions are alternative notations for the same fundamental aspects of distances and geometry⁵.

Conservation of total probability

Probabilities are typically normalized to sum to one, $\sum q_i = \sum q'_i = 1$. That normalization is usually thought of as a trivial aspect of the definition

[†] This article is an adaptation of a prior publication².

Table 1: Definitions of key symbols and concepts.

Symbol	Definition	Equation
\mathbf{q}	Vector of frequencies with $\sum q_i = 1$	(1)
\mathbf{z}	Values with average $\bar{z} = \mathbf{q} \cdot \mathbf{z}$; use $\mathbf{z} \equiv \mathbf{a}, \mathbf{F}$, etc. for specific interpretations	(1)
$\Delta \mathbf{q}$	Discrete changes, $\Delta q_i = q'_i - q_i$, may be large	(1)
$d\mathbf{q}$	Small, differential changes, $\Delta \mathbf{q} \rightarrow d\mathbf{q}$	(7)
\mathbf{a}	Relative change of the i th type, $a_i = \Delta q_i / q_i \rightarrow \dot{q}_i / q_i = \log q'_i / q_i$	(2)
\mathbf{m}	Malthusian parameter, $\mathbf{m} = \log \mathbf{q}' / \mathbf{q}$, log of relative fitness, \mathbf{w}	(15)
\mathbf{w}	Relative fitness, $w_i = q'_i / q_i$, with $\mathbf{m} = \log \mathbf{w}$	(2)
\mathbf{F}	Direct nondimensional forces, may be used for values $\mathbf{z} \equiv \mathbf{F}$	(5)
\mathbf{I}	Inertial nondimensional forces, may be interpreted as acceleration (13)	(12)
$\Delta \mathbf{q} \cdot \mathbf{F}$	Abstract notion of physical work as displacement multiplied by force	(5)
\mathcal{F}	Fisher information, nondimensional expression	(8)
\mathbf{L}	Likelihoods, L_θ , for parameter values, θ ; interpreted as force, $\mathbf{F} \equiv \mathbf{L}$	(17)
$\Delta_{\mathbf{F}}$	Partial change caused by direct forces, e.g., $\Delta \mathbf{q} \cdot \mathbf{F}$ or $\Delta \mathbf{q} \cdot \phi$ or $\Delta \mathbf{q} \cdot \mathbf{L}$	(5)
$\ \cdot\ $	Euclidean vector length, e.g., $\ \mathbf{z}\ $ or $\ \mathbf{F}\ $ or $\ \Delta \mathbf{q}\ $	(4)
\mathbf{r}	Unitary coordinates, $\mathbf{r} = \sqrt{\mathbf{q}}$, with $\ \mathbf{r}\ = 1$ as invariant total probability	(9)

of probability. However, the conservation of total probability profoundly shapes the form of fundamental equations. Seemingly different subjects often share common expressions because they share invariant total probability and its geometric consequences. The Price equation reveals that unity.

To describe the conservation of total probability in the Price equation, define

$$a_i = \frac{\Delta q_i}{q_i} = \frac{q'_i}{q_i} - 1 = w_i - 1. \quad (2)$$

In biology, $w_i = q'_i / q_i$ may be interpreted as relative fitness.

In the Price equation, we can use any value for \mathbf{z} . Let $\mathbf{z} \equiv \mathbf{a}$. The Price equation becomes

$$\Delta \bar{a} = \Delta \mathbf{q} \cdot \mathbf{a} + \mathbf{q}' \cdot \Delta \mathbf{a} = 0. \quad (3)$$

The equality to zero follows because the average of \mathbf{a} is always zero: $\bar{a} = \sum q_i a_i = \sum \Delta q_i = 0$. All of

the changes in probability, Δq_i , must sum to zero in order to keep the total probability constant at one.

We can first study universal aspects of the canonical invariant form based on \mathbf{a} in eqn 3. We can then derive broader results by simply making the coordinate transformation $\mathbf{a} \mapsto \mathbf{z}$, yielding the most general expression of the abstract Price equation in eqn 1.

Distance, force and work

The first term of the Price equation expresses an abstract notion of physical work. The nondimensional work is the product of a force acting on an entity multiplied by the distance that the entity moves.

For example, consider natural selection as a force. The strength of natural selection multiplied by the distance moved by the population is the work ac-

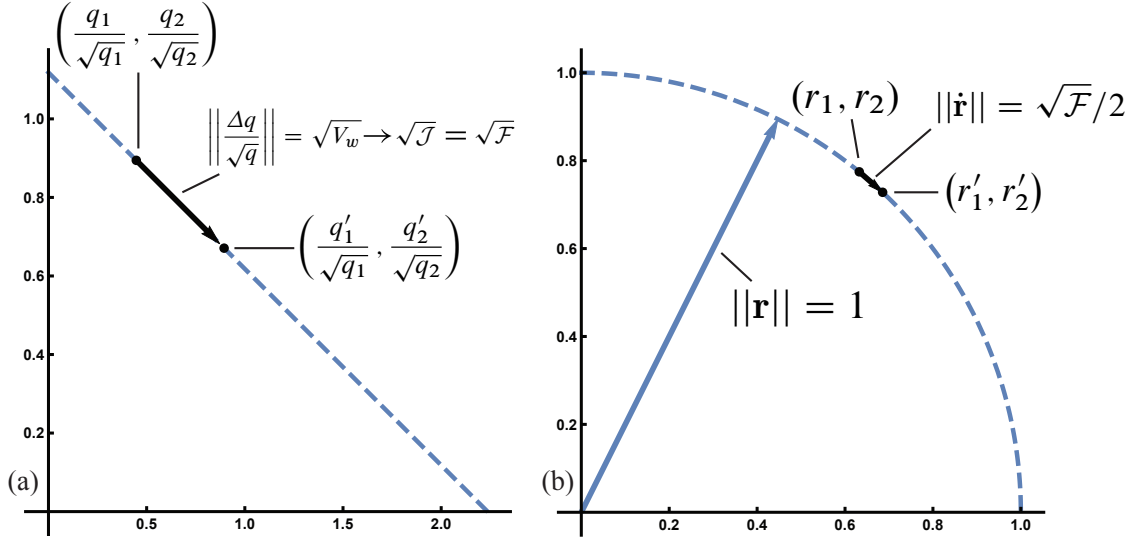


Figure 1: Geometry of change by direct forces. (a) The abstract physical work of the direct forces as the distance moved between the initial population with probabilities, \mathbf{q} , and the altered population with probabilities, \mathbf{q}' . For discrete changes, the probabilities are normalized by the square root of the probabilities in the initial set. The distance can equivalently be described by the various expressions shown, in which V_w is the variance in fitness from population biology, \mathcal{J} is the Jeffreys divergence from information theory, and \mathcal{F} is the Fisher information metric which arises in many disciplines. (b) When changes are small, the same geometry and distances can be described more elegantly in unitary square root coordinates, $\mathbf{r} = \sqrt{\mathbf{q}}$, with $\dot{\mathbf{r}} \equiv d\mathbf{r}$. The symbol “ \rightarrow ” denotes the limit for small changes. From Frank².

completed by natural selection.

When we think about distance, force and work abstractly, we obtain general insight. Natural selection and other problems arise as special cases. To see that generality, begin by writing the standard Euclidean geometry vector length as the square root of the sum of squares

$$\|\mathbf{z}\| = \sqrt{\sum z_i^2}. \quad (4)$$

For any vector \mathbf{z} , the first term of the Price equation is

$$\Delta\mathbf{q} \cdot \mathbf{z} = \|\Delta\mathbf{q}\| \|\mathbf{z}\| \cos \omega,$$

in which ω is the angle between the vectors $\Delta\mathbf{q}$ and \mathbf{z} . If we interpret $\mathbf{z} \equiv \mathbf{F}$ as an abstract, nondimensional force, then

$$\Delta_{\mathbf{F}} \bar{z} = \Delta\mathbf{q} \cdot \mathbf{F} = \|\Delta\mathbf{q}\| \|\mathbf{F}\| \cos \omega \quad (5)$$

expresses an abstract notion of work as the distance moved, $\|\Delta\mathbf{q}\|$, multiplied by the component of force acting along the path of motion, $\|\mathbf{F}\| \cos \omega$.

This expression for work arises in the first term of

the Price equation as the partial change in response to the direct forces, $\Delta_{\mathbf{F}} \bar{z}$.

Geometry of change in populations

Divergence between populations

If we let $\mathbf{z} \equiv \mathbf{a}$ describe the relative growth of the various probabilities, $a_i = \Delta q_i / q_i$, then the divergence between populations caused by the directly acting forces can be expressed as

$$\Delta_{\mathbf{F}} \bar{a} = \Delta\mathbf{q} \cdot \mathbf{a} = \sum \left(\frac{\Delta q_i}{\sqrt{q_i}} \right)^2 = \left\| \frac{\Delta\mathbf{q}}{\sqrt{\mathbf{q}}} \right\|^2 = V_w. \quad (6)$$

If we choose to interpret \mathbf{a} as an abstract notion of force, or fitness, acting on frequency changes, then $\Delta\mathbf{q} \cdot \mathbf{a}$ is the work, with magnitude $\|\Delta\mathbf{q} / \sqrt{\mathbf{q}}\|^2$, that separates the probability distribution \mathbf{q}' from \mathbf{q} . In this article, the division of vectors, such as $\Delta\mathbf{q} / \sqrt{\mathbf{q}}$, means elementwise division.

That value of work is equal to the variance in fitness, V_w . From eqn 2, $a_i = w_i - 1$. Thus, $\Delta\mathbf{q} \cdot \mathbf{a} = \text{Cov}(w, a) = V_w$. The variance in fitness

simply describes the geometric divergence between populations caused by the force of natural selection. See Fig. 1a.

Small changes, paths and logarithms

The Price equation provides an exact description for large, discrete changes. Small, continuous changes are included as a special case.

In prior articles, I developed the theory fully for discrete changes. In the remainder of this article, I focus on small, continuous changes. That focus on continuity makes the mathematics simpler and highlights conceptual aspects more clearly.

If we think of the separation between populations as a sequence of small changes along a path, with each small change as $\Delta \mathbf{q} \rightarrow d\mathbf{q}$. This notation means that as the changes, $\Delta \mathbf{q}$, approach zero, we write those changes in differential notation, $d\mathbf{q}$. With that notation

$$\mathbf{a} \rightarrow \frac{d\mathbf{q}}{\mathbf{q}} = d \log \mathbf{q}. \quad (7)$$

With the differential notation, the partial change by the direct forces separates the probability distributions of the two populations by the path length

$$\Delta_F \bar{a} = \Delta \mathbf{q} \cdot \mathbf{a} \rightarrow d\mathbf{q} \cdot \mathbf{a} = \left\| \frac{d\mathbf{q}}{\sqrt{\mathbf{q}}} \right\|^2 = \mathcal{F}, \quad (8)$$

in which \mathcal{F} is an abstract, nondimensional expression of the Fisher information distance metric⁶.

Unitary geometric coordinates

Let $\mathbf{r} = \sqrt{\mathbf{q}}$. Then $\|\mathbf{r}\| = 1$, expressing the conservation of total probability as a vector of unit length, in which all possible probability combinations of \mathbf{r} define the surface of a unit sphere.

The unitary coordinates, \mathbf{r} , also provide a direct description of Fisher information path length as a distance between two probability distributions

$$4\|\mathbf{dr}\|^2 = 4\|d\sqrt{\mathbf{q}}\|^2 = \left\| \frac{d\mathbf{q}}{\sqrt{\mathbf{q}}} \right\|^2 = \mathcal{F}. \quad (9)$$

The constraint on total probability makes square root coordinates the natural system in which to analyze Euclidean distances, which are the sums of squares. See Figure 1b.

Direct forces and frame of reference

Direct and inertial forces

For small changes, we can write the canonical Price equation for the conservation of total probability in eqn 3 as

$$d\bar{a} = d\mathbf{q} \cdot \mathbf{a} + \mathbf{q} \cdot d\mathbf{a} = 0.$$

To emphasize the first term as the direct forces acting on frequency change and the second term as the inertial forces that change the frame of reference, write

$$(\mathbf{F} + \mathbf{I}) \cdot d\mathbf{q} = 0. \quad (10)$$

The first term describes the direct forces

$$\mathbf{F} \equiv \mathbf{a} = \frac{d\mathbf{q}}{\mathbf{q}} = d \log \mathbf{q}, \quad (11)$$

as in eqn 7. The second term describes the inertial forces

$$\mathbf{I} = \frac{\mathbf{q}}{d\mathbf{q}} d\mathbf{a} = \frac{d\mathbf{a}}{\mathbf{a}} = d \log \mathbf{a}, \quad (12)$$

in which multiplication and division of vectors is elementwise, and $d \log$ is an operator acting on nonzero quantities that maps an argument x to dx/x . (Note that $d \log$ maps its argument to its outcome in a single step, rather than as a logarithm and then a differential. Thus, a can be negative here.) This expression for the inertial forces can be expanded as

$$\mathbf{I} = d \log \mathbf{a} = d \log(d \log \mathbf{q}) = d \log^2 \mathbf{q}. \quad (13)$$

The relative differential, $d \log$, describes relative change. The second relative differential, $\mathbf{I} = d \log^2 \mathbf{q}$, describes the relative acceleration in frequency changes. Thus, the inertial forces acting on the frame of reference can be related to an acceleration.

D'Alembert's principle

Substituting the expressions for the direct and inertial forces by the relative change and the relative acceleration of frequencies yields

$$(\mathbf{F} + \mathbf{I}) \cdot d\mathbf{q} = (d \log \mathbf{q} + d \log^2 \mathbf{q}) \cdot d\mathbf{q} = 0. \quad (14)$$

When written in this form, the canonical Price equation of eqn 3 is an abstract, nondimensional generalization of d’Alembert’s principle for probability distributions that conserve total probability^{2,3,7}.

D’Alembert’s principle is a fundamental expression of physical mechanics¹. The principle generalizes Newton’s second law, force equals mass times acceleration. In one dimension, Newton’s law is $F = -mI$, for force, F , and mass, m , times acceleration, $-I$. In my abstract nondimensional expressions, m drops out, so that $F + I = 0$.

D’Alembert generalizes Newton’s law to a statement about motion in multiple dimensions such that, in conservative systems, the total work for a displacement, $d\mathbf{q}$, and total forces, $\mathbf{F} + \mathbf{I}$, is zero. Work is the distance moved multiplied by the force acting in the direction of motion.

In terms of the canonical Price equation with conserved total probability, the change of a probability distribution between two populations can be partitioned into the balancing work components of the direct forces, $d\mathbf{q} \cdot \mathbf{F}$, and the inertial forces, $d\mathbf{q} \cdot \mathbf{I}$. We can often specify the direct forces in a simple and clear way. The balancing inertial forces may then be analyzed by d’Alembert’s principle¹.

Frame of reference

Here is a simple intuitive description of d’Alembert’s principle⁸. You are sitting in a car at rest, and the car suddenly accelerates. You feel thrown back into the seat. But, even as the car gains speed, you effectively do not move in relation to the frame of reference of the car: Your velocity relative to the car remains zero. That net zero velocity can be thought of as the balance between the direct force of the seat pushing on you and the inertial force sending you back as the car accelerates forward.

As long as your frame of reference moves with you, then your net motion in your frame of reference is zero. Put another way, there is a changing frame of reference that zeroes net change by balancing the work of the direct forces against the work of the inertial forces. Although the system is a dynamic expression of changing components, it also

has an overall static, equilibrium quality that aids analysis. As Lanczos¹ emphasizes, d’Alembert’s principle “focuses attention on the forces, not on the moving body ...”. (These two paragraphs are from Frank³.)

Conservative and nonconservative systems

From eqn 8, the work of the direct forces, $d\mathbf{q} \cdot \mathbf{F} = \mathcal{F}$, is the Fisher information path length that separates the probability distributions, \mathbf{q}' and \mathbf{q} . The inertial forces cause a balancing loss, $d\mathbf{q} \cdot \mathbf{I} = -\mathcal{F}$, which describes the loss in Fisher information that arises from the recalculation of the relative forces in the new frame of reference, \mathbf{q}' .

The balancing loss occurs because the average relative force, or fitness, is always zero in the current frame of reference, $\mathbf{q} \cdot \mathbf{a} = \sum q_i(\dot{q}_i/q_i) = 0$. Any direct gain in relative fitness by direct forces, $d\mathbf{q} \cdot \mathbf{F} = \mathcal{F}$, must be balanced by an equivalent loss in relative fitness, $d\mathbf{q} \cdot \mathbf{I} = -\mathcal{F}$, from the changing frame of reference in which relative fitness is calculated.

The movement of probability distributions in the canonical Price equation is always conservative, $d\bar{a} = 0$, so that d’Alembert’s principle holds. When we transform to the general Price equation by $\mathbf{a} \mapsto \mathbf{z}$, then it may be that $d\bar{z} \neq 0$ and the system is not conservative^{2,3,7}. In that case, we may consider constraints on $d\bar{z}$ and how those constraints influence the possible paths of change for $d\mathbf{q}$.

Interpretation of force

I have equated force with change. For example, $\mathbf{F} = \mathbf{a} = d \log \mathbf{q}$. The duality of force and change arises from the following relation. Given the initial condition and the force that acts on a population, we can deduce frequency change. Given the initial condition and the frequency change, we can induce the force⁹.

The deduce-induce relation arises from the notion that force causes change. However, in the abstract mathematics, we only have the relation between force and change. The mathematics does not express primacy of one over the other.

The value of the abstract Price equation arises from its purely mathematical nature. By equating force with relative frequency change, we intentionally blur the distinction between external causes and internal effects. By describing change as the difference between two abstract sets rather than change through time or space, we intentionally blur the scale of change. By separating frequencies, \mathbf{q} , from property values, \mathbf{z} , we intentionally distinguish universal aspects of change between sets from the particular interpretations of property values in each application.

The blurring of cause, effect and scale, and the separation of frequency from value, lead to abstract mathematical expressions that reveal the common underlying structure among seemingly different subjects.

Value of the partition

The conservation of total probability and the constancy of relative success are by themselves trivial. So one might say that the Price equation is simply some notation to describe trivial facts.

However, many fundamental equations from different disciplines follow immediately and easily from the Price equation partition. It seems that each discipline has, in its own way, come to the same essential invariant geometry of an underlying conservative system.

Interpretation in different disciplines reduces to two aspects. First, one must separate the forces of direct interest from those other forces that alter the frame of reference. Second, one must distinguish the underlying conservative foundation from the coordinates of property values for particular problems.

The Price equation does exactly and only those two aspects on which interpretation depends. By focus on those essential aspects, the Price equation brings out the unity of analysis between seemingly different subjects.

My prior publications have shown how key results from different disciplines arise simply and naturally from the Price equation^{2,3,7}. I have already described a generalization of d'Alembert's principle

of physical mechanics. The remainder of this article briefly sketches two additional examples.

Information

The Price equation separates frequencies from property values. That separation shadows Shannon's separation of the information in a message, expressed by frequencies of symbols in sets, from the meaning of a message, expressed by the properties associated with the message symbols. Price¹⁰ was clearly influenced by the information theory separation between frequency and property in his discussion of a generalized notion of natural selection that might unify disparate subjects.

With regard to frequencies, the Price equation simply describes the universal expression of divergence between sets. By contrast, information theory interprets frequencies and changes in frequencies in terms of the information content of messages.

What is the relation between the general, abstract Price equation description of frequencies in relation to the conservation of total probability and the information theory interpretation of frequencies as having some deeper meaning in terms a concept of "information"?

I begin with the Price equation, which has no notion of "information". I show that key quantities and classic expressions of information theory follow immediately from the Price equation. I then consider the following question.

Given that key expressions and results of information theory follow from the abstract Price equation, should we think of those results as deriving from information theory or from the expression of more basic principles of invariant geometry that arise solely from the conservation of total probability? I argue in favor of the second interpretation.

Information expressions from the Price equation

I showed in eqn 8 that the first term of the Price equation is

$$d\mathbf{q} \cdot \mathbf{F} = d\mathbf{q} \cdot \mathbf{a} = \left\| \frac{d\mathbf{q}}{\sqrt{\mathbf{q}}} \right\|^2 = \mathcal{F},$$

in which \mathcal{F} is an abstract, nondimensional expression of the Fisher information distance metric. The second term of the Price equation is

$$d\mathbf{q} \cdot \mathbf{I} = \mathbf{q} \cdot \mathbf{a} = -\mathcal{F}.$$

Thus, d'Alembert's principle for the Price equation

$$(\mathbf{F} + \mathbf{I}) \cdot d\mathbf{q} = \mathcal{F} - \mathcal{F} = 0$$

expresses the conservation of total information. Fisher information¹¹ has occasionally been raised as a candidate for a fundamental principle underlying physics.

By the Price equation, we see Fisher information and the conservation of information arising as simple consequences of the conservation of total probability. There is no essential need for an underlying notion of "information".

The information theory interpretation can be very useful. The point here is to understand the underlying assumptions and mathematics that lead to such expressions.

The Kullback-Leibler divergence^{12,13} is another key expression of information theory

$$\begin{aligned} \mathcal{D}(\mathbf{q}'||\mathbf{q}) &= \sum_i q'_i \log \frac{q'_i}{q_i} = \mathbf{q}' \cdot d \log \mathbf{q} \\ \mathcal{D}(\mathbf{q}||\mathbf{q}') &= \sum_i q_i \log \frac{q_i}{q'_i} = -\mathbf{q} \cdot d \log \mathbf{q}. \end{aligned}$$

This divergence measures the separation between two probability distributions, \mathbf{q} and \mathbf{q}' . The Kullback-Leibler divergence provides an equivalent expression of Shannon information when the divergence is taken from an initial uniform distribution. Thus, the Kullback-Leibler divergence is often described as relative information—the change in information relative to some given initial distribution.

In information theory, it is often useful to consider the sum of the forward and backward divergences, which creates a symmetric measure. That sum is known as the Jeffreys divergence

$$\begin{aligned} \mathcal{J} = \mathcal{D}(\mathbf{q}'||\mathbf{q}) + \mathcal{D}(\mathbf{q}||\mathbf{q}') &= d\mathbf{q} \cdot d \log \mathbf{q} \\ &= d\mathbf{q} \cdot \mathbf{a} = \mathcal{F}. \end{aligned}$$

These results follow when changes are small. For

analysis of discrete changes, see Frank^{2,3}.

Note that, in biology

$$\log \frac{q'_i}{q_i} = \log w_i = m_i, \quad (15)$$

in which log fitness, $\log w_i = m_i$ is often called the Malthusian parameter. The information measures \mathcal{D} , \mathcal{J} and \mathcal{F} can all be expressed in terms of the Malthusian parameter.

The interpretation of information

All of the "information" results in the prior section arose directly from the canonical Price equation's description of conserved total probability. No notion or interpretation of "information" is necessary.

In many disciplines, information expressions arise in the analysis of the specific disciplinary problems. This sometimes leads to the idea that information must be a primary general concept that gives form to and explains the particular results.

Here, the Price equation explains why those information expressions arise so often. Those expressions are simply the fundamental descriptions of force and change within the context of a conserved total quantity. In this case, the conserved total quantity is total probability.

Information does have many useful interpretations⁶. The next section provides an example.

Inference: data as a force

Following Bayesian tradition, denote the force of the data as $\tilde{L}(D|\theta)$, the likelihood of observing the data, D , given a value for the parameter, θ . To interpret a force as equivalent to relative fitness, the average value of the force must be one to satisfy the conservation of total probability. Thus, define

$$w_\theta = L_\theta = \frac{\tilde{L}(D|\theta)}{\sum_\theta q_\theta \tilde{L}(D|\theta)}.$$

We can now write the classic expression for Bayesian updating of a prior, q_θ , driven by the force of new data, $L_\theta = L(D|\theta)$, to yield the posterior, q'_θ , as

$$q'_\theta = q_\theta L_\theta. \quad (16)$$

By recognizing \mathbf{L} as a force vector acting on frequency change, we can use all of the general results derived from the Price equation. For example, the Malthusian parameter of eqn 15 relates to the log-likelihood as

$$\mathbf{m} = \log \frac{\mathbf{q}'}{\mathbf{q}} = \Delta \log \mathbf{q} = \log \mathbf{L}. \quad (17)$$

This equivalence for log-likelihood relates frequency change to the Kullback–Leibler expressions for the change in information

$$\Delta \mathbf{q} \cdot \log \mathbf{L} = \mathcal{D}(\mathbf{q}' || \mathbf{q}) + \mathcal{D}(\mathbf{q} || \mathbf{q}'), \quad (18)$$

which we may think of as the gain of information from the force of the data. Perhaps the most general expression of change describes the relative separation within the unitary square root coordinates as the Euclidean length

$$\Delta \mathbf{q} \cdot \mathbf{L} = \left\| \frac{\Delta \mathbf{q}}{\sqrt{\mathbf{q}}} \right\|^2,$$

which is an abstract, nondimensional expression for the work done by the displacement of the frequencies, $\Delta \mathbf{q}$, in relation to the force of the data, \mathbf{L} .

I defined \mathbf{L} as a normalized form of the likelihood, $\tilde{\mathbf{L}}$, such that the average value is one, $\bar{\mathbf{L}} = \mathbf{q} \cdot \mathbf{L} = 1$. Thus, we have a canonical form of the Price equation for normalized likelihood

$$\Delta \bar{\mathbf{L}} = \Delta \mathbf{q} \cdot \mathbf{L} + \mathbf{q}' \cdot \Delta \mathbf{L} = 0. \quad (19)$$

The second terms show how the inertial forces alter the frame of reference that determines the normalization of the likelihoods, $\tilde{\mathbf{L}} \mapsto \mathbf{L}$. Typically, as information is gained from data, the normalizing force of the frame of reference reduces the force of the same data in subsequent updates.

All of this simply shows that Bayesian updating describes the change in probability distributions between two sets. That change between sets follows the universal principles given by the abstract Price equation.

Prior work noted the analogy between natural selection and Bayesian updating^{14–16}. Here, I emphasized a more general perspective that includes natural selection and Bayesian updating as examples of the common invariances and geometry that unify many topics.

Discussion

How useful is it to understand the common mathematical basis of different disciplines? The generality by itself does not alter the well known results within each discipline. However, three points suggest potential value.

Universality. First, we can evaluate other claims of universality more clearly. For example, information expressions often occur in the fundamental equations of different disciplines. That commonality tempts one to think of information as a primary quantity.

The Price equation shows that information expressions arise from the basic geometry of divergence between populations. Invariance and geometry set the universal foundations.

Force and cause. Second, I support Lanczos'¹ advocacy of d'Alembert's principle because "it focuses attention on the forces, not on the moving body." The principle highlights causal interpretation of the forces that shape complex dynamics.

D'Alembert's principle transforms a changing system into an equilibrium system. The direct forces by themselves cause nonequilibrium dynamics. The addition of the inertial forces brings the total system into equilibrium, $(\mathbf{F} + \mathbf{I}) \cdot d\mathbf{q} = 0$.

Lanczos emphasizes that

By this device *dynamics is reduced to statics*.

This does not mean that we can actually *solve* a dynamical problem by statical methods. The resulting equations are *differential equations* which have to be solved. We have merely *deduced* these differential equations by statical considerations. The addition of the force of inertia \mathbf{I} to the acting force \mathbf{F} changes the problem of motion to a problem of equilibrium.

This quote describes a great benefit of conservation laws in analysis.

In modern physics, d'Alembert's principle is often relegated to a historical footnote. The principle applies only to conservative mechanical systems.

Most mechanical systems are not conservative, because they have forces, such as friction, that prevent changes from being reversed without loss. Because real systems are rarely conservative and reversible, d'Alembert's principle is limited even within its primary domain of mechanics.

These criticisms of d'Alembert are true. But they miss the abstract mathematical power and insight of d'Alembert's expression, emphasized by Lanczos. Much of the great advance of modern physics came from the abstract structure of conservation laws and their invariances, which often have the balancing form of d'Alembert.

The canonical Price equation is a pure abstract expression of d'Alembert's balance between direct and inertial forces. In the canonical Price equation, with focus on frequencies and the change in probability distributions, the abstract system is frictionless, conservative, and reversible. Thus, the Price equation expresses the underlying mathematical structure that unifies so many seemingly different fundamental results of distinct disciplines, which share the same conservation of total probability or a similar conservation.

For example, the equations of change by natural selection can initially be described in terms of changes in relative fitness. Because relative fitness concerns only frequencies, it matches the canonical Price equation, with a total change in relative fitness of zero. Within that equilibrium system, we can partition the change into the direct forces of natural selection and the changing frame of reference caused by altered frequencies.

Typically, one is interested in the change in some trait value of organisms, \mathbf{z} , rather than in the change of relative fitness, \mathbf{a} . Because \mathbf{z} is not conserved in the same way as \mathbf{a} , additional forces may come into play. We can address those additional forces by a variety of powerful supplemental analytical methods, which include Jaynesian maximum entropy as a special case^{2,3}. However, there can never be a complete universal expression that captures all aspects of nonconservative systems.

Particular vs general. Third, the Price equation reveals that many particular explanations in science derive from general underlying principles.

Physical work and information are well defined quantities with very useful applications. Yet the basic equations of change often arise from a general conservation law rather than from the particular qualities of work and information.

To understand the fundamental results for physical work and information, one has to see the underlying invariant structure. Otherwise, one is naturally inclined to favor explanations in terms of particular physical or informational properties.

We certainly do not need the Price equation to understand work or information or the generic forms of probability distributions. But we are much more likely to see the fallacy of the particular if we begin from the Price equation perspective. The Price equation focuses attention on the invariant generic structure, not on the particular details.

In summary, the Price equation expresses the conservation of total probability. That conservation law constrains the dynamics of populations. The invariant geometry of change revealed by the Price equation explains why fundamental results in different disciplines often share the same underlying form.

Acknowledgments

The Donald Bren Foundation supports my research. I completed this work while on sabbatical in the Theoretical Biology group of the Institute for Integrative Biology at ETH Zürich.

References

1. Lanczos, C. *The Variational Principles of Mechanics* (Dover Publications, New York, 1986), 4th edn.
2. Frank, S. A. The Price equation program: simple invariances unify population dynamics, thermodynamics, probability, information and inference. *Entropy* **20**, 978 (2018).
3. Frank, S. A. Universal expressions of population change by the Price equation: natural selection, information, and maximum entropy production. *Ecology and Evolution* **7**, 3381–3396 (2017).

4. Price, G. R. Extension of covariance selection mathematics. *Annals of Human Genetics* **35**, 485–490 (1972).
5. Frank, S. A. Natural selection. IV. The Price equation. *Journal of Evolutionary Biology* **25**, 1002–1019 (2012).
6. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, New York, 1991).
7. Frank, S. A. D’Alembert’s direct and inertial forces acting on populations: the Price equation and the fundamental theorem of natural selection. *Entropy* **17**, 7087–7100 (2015).
8. Wikipedia. Fictitious force — Wikipedia, The Free Encyclopedia (2015). URL http://en.wikipedia.org/w/index.php?title=Fictitious_force&oldid=659661243. [Online; accessed 22-May-2015].
9. Frank, S. A. Natural selection maximizes Fisher information. *Journal of Evolutionary Biology* **22**, 231–244 (2009).
10. Price, G. R. The nature of selection. *Journal of Theoretical Biology* **175**, 389–396 (1995).
11. Frieden, B. R. *Science from Fisher Information: A Unification* (Cambridge University Press, Cambridge, UK, 2004).
12. Kullback, S. *Information Theory and Statistics* (Wiley, New York, 1959).
13. Kullback, S. & Leibler, R. A. On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79–86 (1951).
14. Shalizi, C. R. Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics* **3**, 1039–1074 (2009).
15. Harper, M. The replicator equation as an inference dynamic. *arXiv:0911.1763v3* (2010).
16. Campbell, J. O. Universal Darwinism as a process of Bayesian inference. *Hypothesis and Theory* **10**, 49 (2016).