Common Probability Patterns Arise from Simple Invariances

Steven A. Frank

Department of Ecology & Evolutionary Biology, University of California, Irvine, CA 92697, USA;
safrank@uci.edu; Tel.: +1-949-824-2244; Fax: +1-949-824-2181

Abstract: Shift and stretch invariance lead to the exponential-Boltzmann probability distribution. Rotational invariance generates the Gaussian distribution. Particular scaling relations transform the canonical exponential and Gaussian patterns into the variety of commonly observed patterns. The scaling relations themselves arise from the fundamental invariances of shift, stretch and rotation, plus a few additional invariances. Prior work described the three fundamental invariances as a consequence of the equilibrium canonical ensemble of statistical mechanics or the Jaynesian maximization of information entropy. By contrast, I emphasize the primacy and sufficiency of invariance alone to explain the commonly observed patterns. Primary invariance naturally creates the array of commonly observed scaling relations and associated probability patterns, whereas the classical approaches derived from statistical mechanics or information theory require special assumptions to derive commonly observed scales.

Keywords: measurement; maximum entropy; information theory; statistical mechanics; extreme value distributions

---

"It is increasingly clear that the symmetry [invariance] group of nature is the deepest thing that we understand about nature today. I would like to suggest something here that I am not really certain about but which is at least a possibility: that specifying the symmetry group of nature may be all we need to say about the physical world, beyond the principles of quantum mechanics.

The paradigm for symmetries of nature is of course the group symmetries of space and time. These are symmetries that tell you that the laws of nature don't care about how you orient your laboratory, or where you locate your laboratory, or how you set your clocks or how fast your laboratory is moving" (Weinberg [1] (p. 73)).

"For the description of processes taking place in nature, one must have a *system of reference*" (Landau and Lifshitz [2] (p. 1)).

1. Introduction

I argue that three simple invariances dominate much of observed pattern. First, probability patterns arise from invariance to a shift in scaled measurements. Second, the scaling of measurements satisfies invariance to uniform stretch. Third, commonly observed scales are often invariant to rotation.

Feynman [3] described the shift invariant form of probability patterns as:

$$\frac{q(\mathcal{E})}{q(\mathcal{E}')} = \frac{q(\mathcal{E}+a)}{q(\mathcal{E}'+a)} \tag{1}$$

in which $q(\mathcal{E})$ is the probability associated with a measurement, $\mathcal{E}$. Here, the ratio of probabilities for two different measurements, $\mathcal{E}$ and $\mathcal{E}'$, is invariant to a shift by a. Feynman derived this invariant

ratio as a consequence of Boltzmann's equilibrium distribution of energy levels, $\mathcal{E}$, that follows from statistical mechanics:

$$q\left(\mathcal{E}\right) = \lambda e^{-\lambda \mathcal{E}} \tag{2}$$

Here, $\lambda = 1/\langle \mathcal{E} \rangle$ is the inverse of the average measurement.

Feynman presented the second equation as primary, arising as the equilibrium from the underlying dynamics of particles and the consequent distribution of energy, $\mathcal{E}$. He then mentioned in a footnote that the first equation of shift invariance follows as a property of equilibrium. However, one could take the first equation of shift invariance as primary. The second equation for the form of the probability distribution then follows as a consequence of shift invariance.

What is primary in the relation between these two equations: equilibrium statistical mechanics or shift invariance? The perspective of statistical mechanics, with Equation (2) as the primary equilibrium outcome, dominates treatises of physics.

Jaynes [4,5] questioned whether statistical mechanics is sufficient to explain why patterns of nature often follow the form of Equation (2). Jaynes emphasized that the same probability pattern often arises in situations for which physical theories of particle dynamics make little sense. In Jaynes' view, if most patterns in economics, biology and other disciplines follow the same distributional form, then that form must arise from principles that transcend the original physical interpretations of particles, energy and statistical mechanics [6].

Jaynes argued that probability patterns derive from the inevitable tendency for systems to lose information. By that view, the equilibrium form expresses minimum information, or maximum entropy, subject to whatever constraints may act in particular situations. In maximum entropy, the shift invariance of the equilibrium distribution is a consequence of the maximum loss of information under the constraint that total probability is conserved.

Here, I take the view that shift invariance is primary. My argument is that shift invariance and the conservation of total probability lead to the exponential-Boltzmann form of probability distributions, without the need to invoke Boltzmann's equilibrium statistical mechanics or Jaynes' maximization of entropy. Those secondary special cases of Boltzmann and Jaynes follow from primary shift invariance and the conservation of probability. The first part of this article develops the primacy of shift invariance.

Once one adopts the primacy of shift invariance, one is faced with the interpretation of the measurement scale, $\mathcal{E}$. We must abandon energy, because we have discarded the primacy of statistical mechanics, and we must abandon Jaynes' information, because we have assumed that we have only general invariances as our basis.

We can of course end up with notions of energy and information that derive from underlying invariance. However, that leaves open the problem of how to define the canonical scale, $\mathcal{E}$, that sets the frame of reference for measurement.

We must replace the scaling relation $\mathcal{E}$ in the above equations by something that derives from deeper generality: the invariances that define the commonly observed scaling relations.

In essence, we start with an underlying scale for observation, z. We then ask what transformed scale, $z \mapsto T_z \equiv \mathcal{E}$, achieves the requisite shift invariance of probability pattern, arising from the invariance of total probability. It must be that shift transformations, $T_z \mapsto a + T_z$, leave the probability pattern invariant, apart from a constant of proportionality.

Next, we note that a stretch of the scale, $T_z \mapsto bT_z$, also leaves the probability pattern unchanged, because the inverse of the average value in Equation (2) becomes $\lambda = 1/b \langle T_z \rangle$, which cancels the stretch in the term $\lambda \mathcal{E} = \lambda T_z$. Thus, the scale T_z has the property that the associated probability pattern is invariant to the affine transformation of shift and stretch, $T_z \mapsto a + bT_z$. That affine invariance generates the symmetry group of scaling relations that determine the commonly observed probability patterns [7–9].

The final part of this article develops rotational invariance of conserved partitions. For example, the Pythagorean partition $T_z = x^2(s) + y^2(s)$ splits the scaled measurement into components that add invariantly to T_z for any value of s. The invariant quantity defines a circle in the xy plane with a

conserved radius $R_z = \sqrt{T_z}$ that is invariant to rotation around the circle, circumscribing a conserved area $\pi R_z^2 = \pi T_z$. Rotational invariance allows one to partition a conserved quantity into additive components, which often provides insight into underlying process.

If we can understand these simple shift, stretch and rotational invariances, we will understand much about the intrinsic structure of pattern. An explanation of natural pattern often means an explanation of how particular processes lead to particular forms of invariance.

2. Background

This section introduces basic concepts and notation. I emphasize qualitative aspects rather than detailed mathematics. The final section of this article provides historical background and alternative perspectives.

2.1. Probability Increments

Define $q(z) \equiv q_z$, such that the probability associated with z is $q_z \Delta \psi_z$. This probability is the area of a rectangle with height q_z and incremental width $\Delta \psi_z$.

The total probability is constrained to be one, as the sum of the rectangular areas over all values of z, which is $\sum q_z \Delta \psi_z = 1$. When the z values are discrete quantities or qualitative labels for events, then the incremental measure is sometimes set to one everywhere, $\Delta \psi_z \equiv 1$, with changes in the measure $\Delta \psi_z$ made implicitly by adjusting q_z. The conservation of probability becomes $\sum q_z = 1$.

If a quantitative scale z has values that are close together, then the incremental widths are small, $\Delta \psi_z \to \mathrm{d}\psi_z$, and the distribution becomes essentially continuous in the limit. The probability around each z value is $q_z \, \mathrm{d}\psi_z$. Writing the limiting sum as a integral over z, the conservation of total probability is $\int q_z \, \mathrm{d}\psi_z = 1$.

The increments may be constant-sized steps $\mathrm{d}\psi_z = \mathrm{d}z$ on the z scale, with probabilities $q_z \mathrm{d}\psi_z = q_z \mathrm{d}z$ in each increment. One may transform z in ways that alter the probability expression, q_z, or the incremental widths, $\mathrm{d}\psi_z$, and study how those changes alter or leave invariant properties associated with the total probability, $q_z \mathrm{d}\psi_z$.

2.2. Parametric Scaling Relations

A probability pattern, $q_z \mathrm{d}\psi_z$, may be considered as a parametric description of two scaling relations, q_z and ψ_z, with respect to the parameter z. Geometrically, $q_z \mathrm{d}\psi_z$ is a rectangular area defined by the parametric height, q_z, with respect to the parameter, z, and the parametric width, $\mathrm{d}\psi_z$, with respect to the parameter, z.

We may think of z as a parameter that defines a curve along the path (ψ_z, q_z), relating a scaled input measure, ψ_z, to a scaled output probability, q_z. The followings sections describe how different invariances constrain these scaling relations.

3. Shift Invariance and the Exponential Form

I show that shift invariance and the conservation of total probability lead to the exponential form of probability distributions in Equation (2). Thus, we may consider the main conclusions of statistical mechanics and maximum entropy as secondary consequences that follow from the primacy of shift invariance and conserved total probability.

3.1. Conserved Total Probability

This section relates shift invariance to the conservation of total probability. Begin by expressing probability in terms of a transformed scale, $z \mapsto T_z$, such that $q_z = k_0 f(T_z)$ and:

$$\int q_z \mathrm{d}\psi_z = \int k_0 f(T_z) \mathrm{d}\psi_z = 1$$

The term k_0 is independent of z and adjusts to satisfy the conservation of total probability.

If we assume that the functional form f is invariant to a shift of the transformed scale by a constant, a, then by the conservation of total probability:

$$\int k_0 f(\mathrm{T}_z)\mathrm{d}\psi_z = \int k_a f(\mathrm{T}_z + a)\mathrm{d}\psi_z = 1 \tag{3}$$

The proportionality constant, k_a, is independent of z and changes with the magnitude of the shift, a, in order to satisfy the constraint on total probability.

Probability expressions, $q(z) \equiv q_z$, are generally not shift invariant with respect to the scale, z. However, if our transformed scale, $z \mapsto \mathrm{T}_z$ is such that we can write Equation (3) for any magnitude of shift, a, solely by adjusting the constant, k_a, then the fact that the conservation of total probability sets the adjustment for k_a means that the condition for T_z to be a shift-invariant canonical scale for probability is:

$$q_z = k_0 f(\mathrm{T}_z) = k_a f(\mathrm{T}_z + a) \tag{4}$$

which holds over the entire domain of z.

The key point here is that k_a is an adjustable parameter, independent of z, that is set by the conservation of total probability. Thus, the conservation of total probability means that we are only required to consider shift invariance in relation to the proportionality constant k_a that changes with the magnitude of the shift, a, independently of the value of z. Appendix A provides additional detail about the conservation of total probability and the shift-invariant exponential form.

3.2. Shift-Invariant Canonical Coordinates

This section shows the equivalence between shift invariance and the exponential form for probability distributions.

Let $x \equiv \mathrm{T}_z$, so that we can write the shift invariance of f in Equation (4) as:

$$f(x + a) = \alpha_a f(x)$$

By the conservation of total probability, α_a depends only on a and is independent of x.

If the invariance holds for any shift, a, then it must hold for an infinitesimal shift, $a = \epsilon$. By Taylor series, we can write:

$$f(x + \epsilon) = f(x) + \epsilon f'(x) = \alpha_\epsilon f(x)$$

Because ϵ is small and independent of x, and $\alpha_0 = 1$, we can write $\alpha_\epsilon = 1 - \lambda\epsilon$ for a constant λ. Then, the previous equation becomes:

$$f'(x) = -\lambda f(x)$$

This differential equation has the solution:

$$f(x) = \hat{k}e^{-\lambda x}$$

in which $\hat{k}$ may be determined by an additional constraint. Using this general property for shift-invariant f in Equation (4), we obtain the classical exponential-Boltzmann form for probability distributions in Equation (2) as:

$$q_z = ke^{-\lambda \mathrm{T}_z} \tag{5}$$

with respect to the canonical scale, T_z. Thus, expressing observations on the canonical shift-invariant scale, $z \mapsto \mathrm{T}_z$, leads to the classical exponential form. If one accepts the primacy of invariance, the "energy," $\mathcal{E}$, of the Boltzmann form in Equation (2) arises as a particular interpretation of the generalized shift-invariant canonical coordinates, T_z.

3.3. Entropy as a Consequence of Shift Invariance

The transformation to obtain the shift-invariant coordinate T_z follows from Equation (5) as:

$$-\log q_z = \lambda T_z - \log k$$

This logarithmic expression of probability leads to various classical definitions of entropy and information [3,10]. Here, the linear relation between the logarithmic scale and the canonical scale follows from the shift invariance of probability with respect to the canonical scale, T_z, and the conservation of total probability.

I interpret shift invariance and the conservation of total probability as primary aspects of probability patterns. Entropy and information interpretations follow as secondary consequences.

One can of course derive shift invariance from physical or information theory perspectives. My only point is that such extrinsic concepts are unnecessary. One can begin directly with shift invariance and the conservation of total probability.

3.4. Example: The Gamma Distribution

Many commonly observed patterns follow the gamma probability distribution, which may be written as:

$$q_z = k z^{\alpha\lambda} e^{-\lambda z}$$

This distribution is not shift invariant with respect to z, because $z \mapsto a + z$ alters the pattern:

$$q_z = k z^{\alpha\lambda} e^{-\lambda z} \neq k_a (a + z)^{\alpha\lambda} e^{-\lambda(a+z)}$$

There is no value of k_a for which this expression holds for all z.

If we write the distribution in canonical form:

$$q_z = k e^{-\lambda T_z} = k e^{-\lambda(z - \alpha \log z)} \tag{6}$$

then the distribution becomes shift invariant on the canonical scale, $T_z = z - \alpha \log z$, because $T_z \mapsto a + T_z$ yields:

$$q_z = k e^{-\lambda(a + T_z)} = k_a e^{-\lambda T_z}$$

with $k_a = k e^{-\lambda a}$. Thus, a shift by a leaves the pattern unchanged apart from an adjustment to the constant of proportionality that is set by the conservation of total probability.

The canonical scale, $T_z = z - \alpha \log z$, is log-linear. It is purely logarithmic for small z, purely linear for large z, and transitions between the log and linear domains through a region determined by the parameter α.

The interpretation of process in relation to pattern almost always reduces to understanding the nature of invariance. In this case, shift invariance associates with log-linear scaling. To understand the gamma pattern, one must understand how a process creates a log-linear scaling relation that is shift invariant with respect to probability pattern [7–9].

4. Stretch Invariance and Average Values

4.1. Conserved Average Values

Stretch invariance means that multiplying the canonical scale by a constant, $T_z \mapsto b T_z$, does not change probability pattern. This condition for stretch invariance associates with the invariance of the average value.

To begin, note that for the incremental measure $d\psi_z = dT_z$, the constant in Equation (5) to satisfy the conservation of total probability is $k = \lambda$, because:

$$\int_0^\infty \lambda e^{-\lambda T_z} \, dT_z = 1$$

when integrating over T_z.

Next, define $\langle X \rangle_\psi$ as the average value of X with respect to the incremental measure $d\psi_z$. Then, the average of λT_z with respect to dT_z is:

$$\lambda \langle T \rangle_T = \int \lambda^2 \, T_z \, e^{-\lambda T_z} \, dT_z = 1 \tag{7}$$

The parameter λ must satisfy the equality. This invariance of $\lambda \langle T \rangle_T$ implies that any stretch transformation $T_z \mapsto bT_z$ will be canceled by $\lambda \mapsto \lambda/b$. See Appendix A for further details.

We may consider stretch invariance as a primary attribute that leads to the invariance of the average value, $\lambda \langle T \rangle_T$, or we may consider invariance of the average value as a primary attribute that leads to stretch invariance.

4.2. Alternative Measures

Stretch invariance holds with respect to alternative measures, $d\psi_z \neq dT_z$. Note that for q_z in Equation (5), the conservation of total probability fixes the value of k, because we must have:

$$\int k e^{-\lambda T_z} d\psi_z = 1$$

The average value of λT_z with respect to $d\psi_z$ is:

$$\int \lambda T_z k e^{-\lambda T_z} d\psi_z = \lambda \langle T \rangle_\psi$$

Here, we do not have any guaranteed value of $\lambda \langle T \rangle_\psi$, because it will vary with the choice of the measure $d\psi_z$. If we assume that $\langle T \rangle_\psi$ is a conserved quantity, then λ must be chosen to satisfy that constraint, and from the fact that λT_z occurs as a pair, $\lambda \langle T \rangle_\psi$ is a conserved quantity. The conservation of $\lambda \langle T \rangle_\psi$ leads to stretch invariance, as in the prior section. Equivalently, stretch invariance leads to the conservation of the average value.

4.3. Example: The Gamma Distribution

The gamma distribution from the prior section provides an example. If we transform the base scale by a stretch factor, $z \mapsto bz$, then:

$$q_z = k_b e^{-\lambda(bz - \alpha \log bz)}$$

There is no altered value of λ for which this expression leaves q_z invariant over all z. By contrast, if we stretch with respect to the canonical scale, $T_z \mapsto bT_z$, in which $T_z = z - \alpha \log z$ for the gamma distribution, we obtain:

$$q_z = k e^{-\lambda_b b T_z} = k e^{-\lambda T_z}$$

for $\lambda_b = \lambda/b$. Thus, if we assume that the distribution is stretch invariant with respect to dz, then the average value $\lambda \langle T \rangle_z = \lambda \langle z - \alpha \log z \rangle$ is a conserved quantity. Alternatively, if we assume that the average value:

$$\lambda \langle T \rangle_z = \lambda \langle z - \alpha \log z \rangle = \lambda \langle z \rangle - \lambda \alpha \langle \log z \rangle$$

is a conserved quantity, then stretch invariance of the canonical scale follows.

In this example of the gamma distribution, conservation of the average value with respect to the canonical scale is associated with conservation of a linear combination of the arithmetic mean, $\langle z \rangle$, and the geometric mean, $\langle \log z \rangle$, with respect to the underlying values, z. In statistical theory, one would say that the arithmetic and geometric means are sufficient statistics for the gamma distribution.

5. Consequences of Shift and Stretch Invariance

5.1. Relation between Alternative Measures

We can relate alternative measures to the canonical scale by $\mathrm{d}\mathrm{T}_z = \mathrm{T}'\mathrm{d}\psi_z$, in which $\mathrm{T}' = |\mathrm{d}\mathrm{T}_z/\mathrm{d}\psi|$ is the absolute value of the rate of change of the canonical scale with respect to the alternative scale. Starting with Equation (7) and substituting $\mathrm{d}\mathrm{T}_z = \mathrm{T}'\mathrm{d}\psi_z$, we have:

$$\lambda \left\langle \mathrm{TT}' \right\rangle_\psi = \int \lambda^2 \mathrm{T}_z \mathrm{T}' e^{-\lambda \mathrm{T}_z} \mathrm{d}\psi_z = 1$$

Thus, we recover a universally conserved quantity with respect to any valid alternative measure, $\mathrm{d}\psi_z$.

5.2. Entropy

Entropy is defined as the average value of $-\log q_z$. From the canonical form of q_z in Equation (2), we have:

$$-\log q_z = \lambda \mathrm{T}_z - \log k \tag{8}$$

Average values depend on the incremental measure, $\mathrm{d}\psi_z$, so we may write entropy [11] as:

$$\left\langle -\log q_z \right\rangle_\psi = \left\langle \lambda \mathrm{T}_z - \log k \right\rangle_\psi = \lambda \left\langle \mathrm{T} \right\rangle_\psi - \log k_\psi$$

The value of $\log k_\psi$ is set by the conservation of total probability, and λ is set by stretch invariance. The value of $\langle \mathrm{T} \rangle_\psi$ varies according to the measure $\mathrm{d}\psi_z$. Thus, the entropy is simply an expression of the average value of the canonical scale, T, with respect to some incremental measurement scale, ψ, adjusted by a term for the conservation of total probability, k.

When $\psi \equiv \mathrm{T}$, then $k_\psi = \lambda$, and we have the classic result for the exponential distribution:

$$\left\langle -\log q_z \right\rangle_\mathrm{T} = \lambda \left\langle \mathrm{T} \right\rangle_\mathrm{T} - \log \lambda = 1 - \log \lambda = \log e/\lambda$$

in which the conserved value $\lambda \langle \mathrm{T} \rangle_\mathrm{T} = 1$ was given in Equation (7) as a consequence of stretch invariance.

5.3. Cumulative Measure

Shift and stretch invariance lead to an interesting relation between $-\log q_z$ and the scale at which probability accumulates. From Equation (8), we have:

$$-\frac{1}{\lambda}\,\mathrm{d}\log q_z = \mathrm{d}\mathrm{T}_z = \mathrm{T}'\mathrm{d}\psi_z$$

Multiplying both sides by q_z, the accumulation of probability with each increment of the associated measure is:

$$-\frac{1}{\lambda}\,q_z\,\mathrm{d}\log q_z = q_z\mathrm{d}\mathrm{T}_z = q_z\mathrm{T}'\mathrm{d}\psi_z$$

The logarithmic form for the cumulative measure of probability simplifies to:

$$-\frac{1}{\lambda}\,q_z\,\mathrm{d}\log q_z = -\frac{1}{\lambda}\,\mathrm{d}q_z = q_z\mathrm{d}\mathrm{T}_z$$

This expression connects the probability weighting, q_z, for each incremental measure, to the rate at which probability accumulates in each increment, $dq_z = -\lambda q_z dT_z$. This special relation follows from the expression for q_z in Equation (2), arising from shift and stretch invariance and the consequent canonical exponential form.

5.4. Affine Invariance and the Common Scales

Probability patterns are invariant to shift and stretch of the canonical scale, T_z. Thus, affine transformations $T_z \mapsto a + bT_z$ define a group of related canonical scales. In previous work, we showed that essentially all commonly observed probability patterns arise from a simple affine group of canonical scales [7–9]. This section briefly summarizes the concept of affine invariant canonical scales. Appendix B provides some examples.

A canonical scale $T(z) \equiv T$ is affine invariant to a transformation $G(z)$ if:

$$T\left[G(z)\right] = a + bT(z)$$

for some constants a and b. We can abbreviate this notion of affine invariance as:

$$T \circ G \sim T \tag{9}$$

in which "$\sim$" means affine invariance in the sense of equivalence for some constants a and b.

We can apply the transformation G to both sides of Equation (9), yielding the new invariance $T \circ G \circ G \sim T \circ G$. In general, we can apply the transformation G repeatedly to each side any number of times, so that:

$$T \circ G^n \sim T \circ G^m$$

for any nonnegative integers n and m. Repeated application of G generates a group of invariances, a symmetry group. Often, in practical application, the base invariance in Equation (9) does not hold, but asymptotic invariance:

$$T \circ G^{(n+1)} \sim T \circ G^n$$

holds for large n. Asymptotic invariance is a key aspect of pattern [12].

6. Rotational Invariance and the Gaussian Radial Measure

The following sections provide a derivation of the Gaussian form and some examples. This section highlights a few results before turning to the derivation.

Rotational invariance transforms the total probability $q_z dT_z$ from the canonical exponential form into the canonical Gaussian form:

$$\lambda e^{-\lambda T_z} dT_z \mapsto v e^{-\pi v^2 R_z^2} \, dR_z \tag{10}$$

This transformation follows from the substitution $\lambda T_z \mapsto \pi v^2 R_z^2$, in which the stretch-invariant canonical scale, λT_z, becomes the stretch-invariant circular area, $\pi v^2 R_z^2$, with squared radius $v^2 R_z^2$. The new incremental scale, $v dR_z$, is the stretch-invariant Gaussian radial measure.

We can, without loss of generality, let $v = 1$ and write $\Lambda = \pi R_z^2$ as the area of a circle. Thus, the canonical Gaussian form:

$$q_z d\psi_z = e^{-\Lambda} dR_z \tag{11}$$

describes the probability, $-\log q_z = \Lambda$, in terms of the area of a circle, Λ, and the incremental measurement scale, $d\psi_z$, in terms of the radial increments, dR_z.

Feynman [3] noted the relation between entropy, radial measure and circular area. In my notation, that relation may be summarized as:

$$\langle -\log q_z \rangle_R = \langle \Lambda \rangle_R$$

However, Feynman considered the circular expression of entropy as a consequence of the underling notion of statistical mechanics. Thus, his derivation followed from an underlying canonical ensemble of particles

By contrast, my framework derives from primary underlying invariances. An underlying invariance of rotation leads to the natural Gaussian expression of circular scaling. To understand how rotational invariance leads to the Gaussian form, it is useful to consider a second parametric input dimension, θ, that describes the angle of rotation [13]. Invariance with respect to rotation means that the probability pattern that relates $q(z, \theta)$ to $\psi(z, \theta)$ is invariant to the angle of rotation.

6.1. Gaussian Distribution

I now show that rotational invariance transforms the canonical shift and stretch invariant exponential form into the Gaussian form, as in Equation (10). To begin, express the incremental measure in terms of the Gaussian radial measure as:

$$\lambda \mathrm{d}\mathrm{T}_z = \pi v^2 \mathrm{d}R_z^2 = 2\pi v^2 R_z \mathrm{d}R_z$$

from which the canonical exponential form $q_z \mathrm{d}\mathrm{T}_z = \lambda e^{-\lambda \mathrm{T}_z} \mathrm{d}\mathrm{T}_z$ may be expressed in terms of the radial measure as:

$$\lambda e^{-\lambda \mathrm{T}_z} \mathrm{d}\mathrm{T}_z = 2\pi v^2 R_z e^{-\pi v^2 R_z^2} \mathrm{d}R_z \tag{12}$$

Rotational invariance means that for each radial increment, $v\mathrm{d}R_z$, the total probability in that increment given in Equation (12) is spread uniformly over the circumference $2\pi v R_z$ of the circle at radius $v R_z$ from a central location.

Uniformity over the circumference implies that we can define a unit of incremental length along the circumferential path with a fraction $1/2\pi v R_z$ of the total probability in the circumferential shell of width $v\mathrm{d}R_z$. Thus, the probability along an increment $v\mathrm{d}R_z$ of a radial vector follows the Gaussian distribution:

$$(1/2\pi v R_z)\, q_z \mathrm{d}\mathrm{T}_z = v e^{-\pi v^2 R_z^2} \mathrm{d}R_z$$

invariantly of the angle of orientation of the radial vector.

Here, the total probability of the original exponential form, $q_z \mathrm{d}\mathrm{T}_z$, is spread evenly over the two-dimensional parameter space (z, θ) that includes all rotational orientations. The Gaussian expression describes the distribution of probability along each radial vector, in which a vector intersects a constant-sized area of each circumferential shell independently of distance from the origin.

The Gaussian distribution varies over all positive and negative values, $R_z \in (-\infty, \infty)$, corresponding to an initial exponential distribution in squared radii, $R_z^2 = \mathrm{T}_z \in (0, \infty)$. We can think of radial vectors as taking positive or negative values according to their orientation in the upper or lower half planes.

6.2. Radial Shift and Stretch Invariance

The radial value, R_z, describes the distance from the central location. Thus, the average radial value is zero, $\langle R \rangle_R = 0$, when evaluated over all positive and negative radial values. Shift invariance associates with no change in radial distance as the frame of reference shifts the location of the center of the circle to maintain constant radii.

Stretch invariance associates with the conserved value of the average circular area:

$$\lambda \langle \mathrm{T} \rangle_R = \pi v^2 \left\langle R^2 \right\rangle_R = \pi v^2 \sigma^2 = \frac{1}{2}$$

in which the variance, σ^2, is traditionally defined as the average of the squared deviations from the central location. Here, we have squared radial deviations from the center of the circle averaged over the incremental radial measure, $\mathrm{d}R_z$.

When $\lambda = v^2 = 1$, we have $\sigma^2 = 1/2\pi$, and we obtain the elegant expression of the Gaussian as the relation between circular area and radial increments in Equation (11). This result corresponds to an average circular area of one, because $\langle 2\pi R^2 \rangle = 2\pi\sigma^2 = 1$.

It is common to express the Gaussian in the standard normal form, with $\sigma^2 = 1$, which yields $v^2 = 1/2\pi$ and the associated probability expression obtained by substituting this value into Equation (10).

6.3. Transforming Distributions to Canonical Gaussian Form

Rotational invariance transforms the canonical exponential form into the Gaussian form, as in Equation (10). If we equate $R_z = \sqrt{T_z}$ and $\lambda = \pi v^2$, we can write the Gaussian form as:

$$q_z \mathrm{d}R_z = \sqrt{\frac{\lambda}{\pi}}\, e^{-\lambda T_z} \mathrm{d}\sqrt{T_z} \qquad (13)$$

in which:

$$\tilde{\sigma}^2 = \langle T \rangle_{\sqrt{T}}$$

is a generalized notion of the variance.

The expression in Equation (13) may require a shift of T_z so that $T_z \in (0, \infty)$, with associated radial values $R_z = \pm\sqrt{T_z}$. The nature of the required shift is most easily shown by example.

6.4. Example: The Gamma Distribution

The gamma distribution may be expressed as $q_z \mathrm{d}\psi_z$ with respect to the parameter z when we set $T_z = z - \alpha \log z$ and $\mathrm{d}\psi_z = \mathrm{d}z$, yielding:

$$q_z \mathrm{d}z = k e^{-\lambda(z - \alpha \log z)} \mathrm{d}z$$

for $z \geq 0$. To transform this expression to the Gaussian radial scale, we must shift T_z so that the corresponding value of R_z describes a monotonically increasing radial distance from a central location.

For the gamma distribution, if we use the shift $T_z \mapsto T_z - \alpha = (z - \alpha \log z) - \alpha$ for $\alpha \geq 0$, then the minimum of T_z and the associated maximum of q_z correspond to $R_z = 0$, which is what we need to transform into the Gaussian form. In particular, the parametric plot of the points $(\pm R_z, q_z)$ with respect to the parameter $z \in (0, \infty)$ follows the Gaussian pattern.

In addition, the parametric plot of the points (T_z, q_z) follows the exponential-Boltzmann pattern. Thus, we have a parametric description of the probability pattern q_z in terms of three alternative scaling relations for the underlying parameter z: the measure $\mathrm{d}z$ corresponds to the value of z itself and the gamma pattern; the measure $\mathrm{d}R_z$ corresponds to the Gaussian radial measure; and the measure $\mathrm{d}T_z$ corresponds to the logarithmic scaling of q_z and the exponential-Boltzmann pattern. Each measure expresses particular invariances of scale.

6.5. Example: The Beta Distribution

A common form of the beta distribution is:

$$q_z \mathrm{d}z = k z^{\alpha - 1} (1 - z)^{\beta - 1} \mathrm{d}z$$

for $z \in (0, 1)$. We can express this distribution in canonical exponential form $k e^{-\lambda T_z}$ by the scaling relation:

$$-\lambda T_z = (\alpha - 1) \log z + (\beta - 1) \log(1 - z)$$

with $\lambda > 0$. For α and β both greater than one, this scaling defines a log-linear-log pattern [8], in the sense that $-\lambda T_z$ scales logarithmically near the endpoints of zero and one and transitions to a linear scaling interiorly near the minimum of T_z at:

$$z^* = \frac{\alpha - 1}{\alpha + \beta - 2} \tag{14}$$

When $0 < \alpha < 1$, the minimum (extremum) of T_z is at $z^* = 0$. For our purposes, it is useful to let $\alpha = \lambda$ for $\lambda > 0$ and to assume $\beta > 1$.

Define T^* as the value of T_z evaluated at z^*. Thus, T^* is the minimum value of T_z, and T_z increases monotonically from its minimum. If we shift T_z by its minimum, $T_z \mapsto T_z - T^*$, and use the shifted value of T_z, we obtain the three standard forms of a distribution in terms of the parameter $z \in (0, 1)$, as follows.

The measure dz and parametric plot (z, q_z) are the standard beta distribution form; the measure dR_z and parametric plot $(\pm R_z, q_z)$ are the standard Gaussian form; and the measure dT_z and parametric plot (T_z, q_z) are the standard exponential-Boltzmann form.

7. Rotational Invariance and Partitions

The Gaussian radial measure often reveals the further underlying invariances that shape pattern. Those invariances appear from the natural way in which the radial measure can be partitioned into additive components.

7.1. Overview

Conserved quantities may arise from an underlying combination of processes. For example, we might know that a conserved quantity, $R^2 = x + y$, arises as the sum of two underlying processes with values x and y. We do not know x and y, only that their conserved sum is invariantly equal to R^2.

The partition of an invariant quantity into a sum may be interpreted as rotational invariance, because:

$$R^2 = x + y = \sqrt{x}^2 + \sqrt{y}^2$$

defines a circle with conserved radius R along the positive and negative values of the coordinates $\left(\sqrt{x}, \sqrt{y}\right)$. That form of rotational invariance explains much of observed pattern, many of the classical results in probability and dynamics, and the expression of those results in the context of mechanics.

The partition can be extended to a multidimensional sphere of radius R as:

$$R^2 = \sum \sqrt{x_i}^2 \tag{15}$$

One can think of rotational invariance in two different ways. First, one may start with a variety of different dimensions, with no conservation in any particular dimension. However, the aggregate may satisfy a conserved total that imposes rotational invariance among the components.

Second, every conserved quantity can be partitioned into various additive components. That partition starts with a conserved quantity, and then, by adding dimensions that satisfy the total conservation, one induces a higher dimensional rotational invariance. Thus, every conserved quantity associates with higher-dimensional rotational invariance.

7.2. Rotational Invariance of Conserved Probability

In the probability expression $q_z d\psi_z$, suppose the incremental measure $d\psi_z$ is constant, and we have a finite number of values of z with positive probability. We may write the conserved total

probability as $\sum_z q_z = 1$. Then, from Equation (15), we can write the conservation of total probability as a partition of $R^2 = 1$ confined to the surface of a multidimensional sphere:

$$\sum_z \sqrt{q_z}^2 = 1$$

There is a natural square root spherical coordinate system, $\sqrt{q_z}$, in which to express conserved probability. Square roots of probabilities arise in a variety of fundamental expressions of physics, statistics and probability theory [14,15].

7.3. Partition of the Canonical Scale

The canonical scale equals the square of the Gaussian radial scale, $T_z = R_z^2$. Thus, we can write a two-dimensional partition from Equation (15) as:

$$T_z = \sqrt{x_1}^2 + \sqrt{x_2}^2$$

Define the two dimensions as:

$$\sqrt{x_1} = w \equiv w(z,s)$$
$$\sqrt{x_2} = \dot{w} \equiv \dot{w}(z,s)$$

yielding the partition for the canonical scale as:

$$T_z = w^2 + \dot{w}^2 \tag{16}$$

This expression takes the input parameter z and partitions the resulting value of $T_z = R_z^2$ into a circle of radius R_z along the path $(w, \dot{w})$ traced by the parameter s.

The radial distance, R_z, and associated canonical scale value, $T_z = R_z^2$, are invariant with respect to s. In general, for each dimension we add to a partition of T_z, we can create an additional invariance with respect to a new parameter.

7.4. Partition into Location and Rate

A common partition separates the radius into dimensions of location and rate. Define $\dot{w} = \partial w / \partial s$ as the rate of change in the location w with respect to the parameter s. Then, we can use the notational equivalence $H_z \equiv T_z = R_z^2$ to emphasize the relation to a classic expression in physics for a conserved Hamiltonian as:

$$H_z = w^2 + \dot{w}^2 \tag{17}$$

in which this conserved square of the radial distance is partitioned into the sum of a squared location, w^2, and a squared rate of change in location, $\dot{w}^2$. The squared rate, or velocity, arises as a geometric consequence of the Pythagorean partitioning of a squared radial distance into squared component dimensions. Many extensions of this Hamiltonian interpretation can be found in standard textbooks of physics.

With the Hamiltonian notation, $H_z \equiv T_z$, our canonical exponential-Boltzmann distribution is:

$$q_z \mathrm{d}H_z = \lambda e^{-\lambda H_z} \mathrm{d}H_z$$

The value H is often interpreted as energy, with $\mathrm{d}H$ as the Gibbs measure. For the simple circular partition of Equation (17), the total energy is often split into potential, w^2, and kinetic, $\dot{w}^2$, components.

In this article, I emphasize the underlying invariances and their geometric relations as fundamental. From my perspective, the interpretation of energy and its components are simply one way in which to describe the fundamental invariances.

The Hamiltonian interpretation is, however, particularly useful. It leads to a natural expression of dynamics with respect to underlying invariance. For example, we can partition a probability pattern into its currently observable location and its rate of change:

$$e^{-\lambda H_z} = e^{-\lambda w^2} e^{-\lambda \dot{w}^2}$$

The first component, w^2, may be interpreted as the observable state of the probability pattern at a particular time. The second component, $\dot{w}^2$, may be interpreted as the rate of change in the probability pattern. Invariance applies to the combination of location and rate of change, rather than to either component alone. Thus, invariance does not imply equilibrium.

8. Summary of Invariances

Probability patterns, q_z, express invariances of shift and stretch with respect to a canonical scale, T_z. Those invariances lead to an exponential form:

$$q_z \mathrm{d}\psi_z = k e^{-\lambda T_z} \mathrm{d}\psi_z$$

with respect to various incremental measures, $\mathrm{d}\psi_z$. This probability expression may be regarded parametrically with respect to z. The parametric view splits the probability pattern into two scaling relations, q_z and ψ_z, with respect to z, forming the parametric curve defined by the points (ψ_z, q_z).

For the canonical scale, T_z, we may consider the sorts of transformations that leave the scale shift and stretch (affine) invariant, $T \circ G \sim T$, as in Equation (9). Essentially, all of the canonical scales of common probability patterns [7–9] arise from the affine invariance of T and a few simple types of underlying invariance with respect to z.

For the incremental measure scale, $\mathrm{d}\psi_z$, four alternatives highlight different aspects of probability pattern and scale.

The scale $\mathrm{d}z$ leads to the traditional expression of probability pattern, $q_z \mathrm{d}z$, which highlights the invariances that set the canonical scale, T_z.

The scale $\mathrm{d}T_z$ leads to the universal exponential-Boltzmann form, $q_z \mathrm{d}T_z$, which highlights the fundamental shift and stretch invariances in relation to the conservation of total probability.

This conservation of total probability may alternatively be described by a cumulative probability measure, $\mathrm{d}q_z = -\lambda q_z \mathrm{d}T_z$.

Finally, rotational invariance leads to the Gaussian radial measure, $\mathrm{d}R_z$. That radial measure transforms many probability scalings, q_z, into Gaussian distributions, $q_z \mathrm{d}R_z$.

Invariances typically associate with conserved quantities [16]. For example, the rotational invariance of the Gaussian radial measure is equivalent to the conservation of the average area circumscribed by the radial measure. That average circular area is proportional to the traditional definition of the variance. Thus, rotational invariance and conserved variance are equivalent in the Gaussian form.

The Gaussian radial measure often reveals the further underlying invariances that shape pattern. That insight follows from the natural way in which the radial measure can be partitioned into additive components.

9. The Primacy of Invariance and Symmetry

> "It was Einstein who radically changed the *way* people thought about nature, moving away
> from the mechanical viewpoint of the nineteenth century toward the elegant contemplation
> of the underlying symmetry principles of the laws of physics in the twentieth century"
> (Lederman and Hill [17] (p. 153)).

The exponential-Boltzmann distribution in Equation (2) provides the basis for statistical mechanics, Jaynesian maximum entropy and my own invariance framework. These approaches derive the

exponential form from different assumptions. The underlying assumptions determine how far one may extend the exponential-Boltzmann form toward explaining the variety of commonly observed patterns.

I claim that one must begin solely with the fundamental invariances in order to develop a proper understanding of the full range of common patterns. By contrast, statistical mechanics and Jaynesian maximum entropy begin from particular assumptions that only partially reflect the deeper underlying invariances.

9.1. Statistical Mechanics

Statistical mechanics typically begins with an assumed, unseen ensemble of microscopic particles. Each particle is often regarded as identical in nature to the others. Statistical averages over the underlying microscopic ensemble lead to a macroscopic distribution of measurable quantities. The exponential-Boltzmann distribution is the basic equilibrium macroscopic probability pattern.

In contrast with the mechanical perspective of statistical physics, my approach begins with fundamental underlying invariances (symmetries). Both approaches arrive at roughly the same intermediate point of the exponential-Boltzmann form. That canonical form expresses essentially the same invariances, no matter whether one begins with an underlying mechanical perspective or an underlying invariance perspective.

From my point of view, the underlying mechanical perspective happens to be one particular way in which to uncover the basic invariances that shape pattern. However, the mechanical perspective has limitations associated with the unnecessarily particular assumptions made about the underlying microscopic ensemble.

For example, to derive the log-linear scaling pattern that characterizes the commonly observed gamma distribution in Equation (6), a mechanical perspective must make special assumptions about the interactions between the underlying microscopic particles.

Some may consider the demand for explicit mechanical assumptions about the underlying particles to be a benefit. However, in practice, those explicit assumptions are almost certainly false and, instead, simply serve as a method by which to point in the direction of the deeper underlying invariance that shapes the scaling relations and associated probability patterns.

I prefer to start with the deeper abstract structure shaped by the key invariances. Then, one may consider the variety of different particular mechanical assumptions that lead to the key invariances. Each set of particular assumptions that are consistent with the key invariances define a special case.

There have been many powerful extensions to statistical mechanics in recent years. Examples include generalized entropies based on assumptions about underlying particle mechanics [18], superstatistics as the average over heterogeneous microscopic sets [19] and invariance principles applied to the mechanical aspects of particle interactions [20].

My own invariance and scaling approach subsumes essentially all of those results in a simple and elegant way, and goes much further with regard to providing a systematic understanding of the commonly observed patterns [7–9]. However, it remains a matter of opinion whether an underlying mechanical framework based on an explicit microscopic ensemble is better or worse than a more abstract approach based purely on invariances.

9.2. Jaynesian Maximum Entropy

Jaynes [4,5] replaced the old microscopic ensemble of particles and the associated mechanical entropy with a new information entropy. He showed that maximum entropy, in the sense of information rather than particle mechanics, leads to the classic exponential-Boltzmann form. A large literature extends the Jaynesian framework [21]. Axiomatic approaches transcend the original justifications based on intuitive notions of information [22].

Jaynes' exponential form has a kind of canonical scale, T_z. In Jaynes' approach, one sets the average value over the canonical scale to a fixed value, in our notation a fixed value of $\langle T \rangle_z$. That conserved average value defines a constraint (an invariance) that determines the associated probability

pattern [23]. The Jaynesian algorithm is the maximization of entropy, subject to a constraint on the average value of some quantity, T_z.

Jaynes struggled to go beyond the standard constraints of the mean or the variance. Those constraints arise from fixing the average values of $T_z = z$ or $T_z = z^2$, which lead to the associated exponential or Gaussian forms. Jaynes did discuss a variety of additional invariances [6] and associated probability patterns. However, he never achieved any systematic understanding of the common invariances and the associated commonly observed patterns and their relations.

I regarded Jaynes' transcendence of the particle-based microscopic ensemble as a strong move in the right direction. I followed that direction for several years [7–9,12]. In my prior work, I developed the intrinsic affine invariance of the canonical scale, T_z, with respect to the exponential-Boltzmann distribution of maximum entropy. The recognition of that general affine invariance plus the variety of common invariances of scale [24,25] led to my systematic classification of the common probability patterns and their relationships [7–9].

In this article, I have taken the next step by doing away with the Jaynesian maximization of entropy. I replaced that maximization with the fundamental invariances of shift and stretch, from which I obtained the canonical exponential-Boltzmann form.

With the exponential-Boltzmann distribution derived from shift and stretch invariance rather than Jaynesian maximum entropy, I added my prior work on the general affine invariance of the canonical scale and the additional particular invariances that define the common scaling relations and probability patterns. We now have a complete system based purely on invariances.

9.3. Conclusions

Shift and stretch invariance set the exponential-Boltzmann form of probability patterns. Rotational invariance transforms the exponential pattern into the Gaussian pattern. These fundamental forms define the abstract structure of pattern with respect to a canonical scale.

In a particular application, observable pattern arises by the scaling relation between the natural measurements of that application and the canonical scale. The particular scaling relation derives from the universal affine invariance of the canonical scale and from the additional invariances that arise in the particular application.

Together, these invariances define the commonly observed scaling relations and associated probability patterns. The study of pattern often reduces to the study of how particular generative processes set the particular invariances that define scale.

Diverse and seemingly unrelated generative processes may reduce to the same simple invariance and, thus, to the same scaling relation and associated pattern. To test hypotheses about generative process and to understand the diversity of natural pattern, one must understand the central role of invariance. Although that message has been repeated many times, it has yet to be fully deciphered.

Author Contributions: Steven A. Frank did all of the research and wrote the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Technical Issues and Extensions

Appendix A.1. Conserved Total Probability

The relations between shift invariance and the conservation of total probability in Section 3.1 form a core part of the article. Here, I clarify the particular goals, assumptions and consequences.

In Section 3.1, I assumed that the conservation of total probability and shift invariance hold. From those assumptions, Equation (4) follows and, thus, also the exponential-Boltzmann form of Equation (5).

I am not claiming that conservation of total probability by itself leads to shift invariance. Instead, my goal is to consider the consequences that follow from a primary assumption of shift invariance.

The justification for a primary assumption of invariance remains an open problem at the foundation of much of modern physics. The opening quote from Weinberg expresses the key role of invariances and also the uncertainty about why invariances are fundamental. My only goal concerns the consequences that follow from the assumption of primary invariances.

Appendix A.2. Conserved Average Values: Equation (7)

Below Equation (7), I stated that the average value $\lambda \langle T \rangle_T = 1$ remains unchanged after stretch transformation, $T_z \mapsto bT_z$. This section provides additional details. The problem begins with Equation (7), repeated here:

$$\lambda \langle T \rangle_T = \int \lambda^2 \, T_z \, e^{-\lambda T_z} \, dT_z = 1$$

Make the substitution $T_z \mapsto bT_z$, which yields:

$$\lambda b \langle T \rangle_T = \int \lambda^2 b^2 \, T_z \, e^{-\lambda b T_z} \, dT_z = 1$$

noting that $T_z \mapsto bT_z$ implies $dT_z \mapsto b dT_z$, which explains the origin of the b^2 term on the right-hand side. Thus, Equation (7) remains one under stretch transformation, implying that $\langle T \rangle_T = 1/\lambda b$.

Appendix A.3. Primacy of Invariance

This article assumes the primacy of shift and stretch invariance. The article then develops the consequences of primary invariance. There are many other ways of understanding the fact that the foundational exponential-Boltzmann distribution expresses shift and stretch invariance, and the Gaussian distribution expresses rotational invariance. One can derive those invariances from other assumptions, rather than assume that they are primary.

Classical statistical mechanics derives shift and stretch invariance as consequences of the aggregate behavior of many particles. Jaynesian maximum entropy derives shift and stretch invariance as consequences of the tendency for entropy to increase plus the assumptions that total probability is conserved and that the average value of some measurement is conserved. In my notation, the conservation of $\langle \lambda T_z \rangle$ is equivalent to the assumption of stretch invariance. Often, this kind of assumption is similar to various conservation assumptions, such as the conservation of energy.

Another way to derive invariance is by the classic limit theorems of probability. Gnedenko and Kolmogorov [26] beautifully summarized a key aspect:

> "In fact, all epistemologic value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity."

The limit theorems typically derive from assumptions such as the summation of many independent random components or, in more complicated studies, the aggregation of partially correlated random components. From those assumptions, certain invariances may arise as consequences.

It may seem that the derivation of invariances from more concrete assumptions provides a better approach. However, from a mathematical and perhaps ultimate point of view, invariance is often tautologically related to supposedly more concrete assumptions. For example, conservation of energy typically arises as an assumption in many profound physical theories. In those theories, one could choose to say that stretch invariance arises from conservation of energy or, equivalently, that conservation of energy arises from stretch invariance. It is not at all clear how we can know which is primary, because mathematically, they are often effectively the same assumption.

My point of departure is the opening quote from Weinberg, who based his statement on the overwhelming success of 20th century physics. That success has partly (mostly?) been driven by studying the consequences that follow from assuming various primary invariances. The ultimate

basis for those primary invariances remains unclear, but the profoundly successful consequences of proceeding in this way are very clear. These issues are very important. However, a proper discussion would require probing the basis of modern physics, as well as many deep recent developments in mathematics, which is beyond my scope. I simply wanted to analyze what would follow from the assumption of a few simple primary invariances.

Appendix A.4. Measurement Theory

Classical measurement theory develops a rational approach to derive and understand measurement scales [24,25]. Roughly speaking, a measurement scale is defined by the transformations that leave invariant the relevant relations of the measurement process. Different approaches develop that general notion of invariance in different ways or expand into broader aspects of pattern (e.g., Grenander [27]).

This article concerns probability patterns in relation to scale. The key is that probability patterns remain invariant to affine transformation, in other words, to shift and stretch transformations. Thus, different measurement scales lead to the same invariant probability pattern if they are affine similar. I discussed the role of affine similarity in several recent articles [7–9]. Here, I briefly highlight the main points.

Start with some notation. Let $T(z) \equiv T$ be a transformation of underlying observations z that define a scale, T. Each scale T has the property of being invariant to certain alterations of the underlying observations. Let a candidate alteration of the underlying observation be the generator, $G(z) \equiv G$. Invariance of the scale T to the generator G means that:

$$T\left[G(z)\right] = T(z)$$

which we can write in simpler notation as:

$$T \circ G = T$$

Sometimes, we do not require exact invariance, but only a kind of similarity. In the case of probability patterns, shift and stretch invariance mean that any two scales related by affine transformation $T = a + bT$ yield the same probability pattern. In other words, probability patterns are invariant to affine transformations of scale. Thus, with regard to the generator G, we only require that $T \circ G$ fall within a family of affine transformation of T. Thus, we write the conditions for two probability patterns to be invariant to the generator G as:

$$T \circ G = a + bT \sim T$$

and thus, the key invariance relation for probability patterns is affine similarity expressed as:

$$T \circ G \sim T$$

which was presented in the text as Equation (9). My prior publications fully developed this relation of affine similarity and its consequences for the variety of scales that define the commonly observed probability patterns [7–9]. Appendix B briefly presents a few examples, including the linear-log scale.

Appendix B. Invariance and the Common Canonical Scales

The variety of canonical scales may be understood by the variety of invariances that hold under different circumstances. I introduced the affine invariance of the canonical scale in Equation (9). This section briefly summarizes further aspects of invariance and the common canonical scales. Prior publications provide more detail [7–9].

Invariance can be studied by the partition of the transformation, $z \mapsto \mathrm{T}_z$, into two steps, $z \mapsto w \mapsto \mathrm{T}_z$. The first transformation expresses intrinsic invariances by the transformation $z \mapsto w(z)$, in which w defines the new base scale consistent with the intrinsic invariances.

The second transformation evaluates only the canonical shift and stretch invariances in relation to the base scale, $w \mapsto a + bw$. This affine transformation of the base scale can be written as $\mathrm{T}(w) = a + bw$. We can define $\mathrm{T}(w) \equiv \mathrm{T}_z$, noting that w is a function of z.

Appendix B.1. Rotational Invariance of the Base Scale

Rotational invariance is perhaps the most common base scale symmetry. In the simplest case, $w(z) = z^2$. If we write $x = z\cos\theta$ and $y = z\sin\theta$, then $x^2 + y^2 = z^2$, and the points (x, y) trace a circle with a radius z that is rotationally invariant to the angle θ. Many probability distributions arise from rotationally invariant base scales, which is why squared values are so common in probability patterns. For example, if $w = z^2$ and $\mathrm{T}_z \equiv w$, then the canonical exponential form that follows from shift and stretch invariance of the rotationally invariant base scale is:

$$q_z = ke^{-\lambda w} = ke^{-\lambda z^2}$$

which is the Gaussian distribution, as discussed in the text.

Note that the word *rotation* captures an invariance that transcends a purely angular interpretation. Instead, we have component processes or measurements that satisfy an additive invariance constraint. For each final value, z, there exist a variety of underlying processes or outcomes that satisfy the invariance $\sum x_i^2 = z^2$.

The word *rotation* simply refers to the diversity of underlying Pythagorean partitions that sum to an invariant Euclidean distance. The set of invariant partitions falls on the surface of a sphere. That spherical property leads to the expression of invariant additive partitions in terms of rotation.

Appendix B.2. General Form of Base Scale Invariance

The earlier sections established that the canonical scale of probability patterns is invariant to shift and stretch. Thus, we may consider as equivalent any affine transformation of the base scale $w \mapsto a + bw$.

We may describe additional invariances of w, such as rotational invariance, in the general form:

$$w \circ G \sim w \tag{B1}$$

in which $w \circ G \equiv w\left[G(z)\right]$. We read Equation (B1) as: the base scale w is invariant to transformation by G, such that $w \circ G = a + bw$ for some constants a and b. The symbol "$\sim$" abbreviates the affine invariance of w.

For example, we may express the rotational invariance of the prior section as:

$$w(z, \theta) = z^2(\cos^2\theta + \sin^2\theta) = z^2$$

because $\cos^2\theta + \sin^2\theta = 1$ for any value of θ. We can describe rotation by the transformation:

$$G(z, \theta) = (z, \theta + \epsilon)$$

so that the invariance expression is:

$$w \circ G = w\left[G(z, \theta)\right] = w(z, \theta + \epsilon) = z^2$$

Thus, the base scale w is affine invariant to the rotational transformation generator, G, as in Equation (B1). Although this form of rotational invariance seems trivial in this context, it turns out to be the basis for many classical results in probability, dynamics and statistical mechanics.

Appendix B.3. Example: Linear-Log Invariance of the Base Scale

The invariance expression of Equation (B1) sets the conditions for base scale invariances. Although there are many possible base scales, a few dominate the commonly observed patterns [7–9]. In this article, I emphasize the principles of invariance rather than a full discussion of the various common scales.

Earlier, I discussed the log-linear scale associated with the gamma distribution. This section presents the inverse linear-log scale, which is:

$$w(z) = \alpha \log(1 + \beta z)$$

When βz is small, w is approximately $\alpha\beta z$, which is linear in z. When βz is large, w is approximately $\alpha \log(\beta z)$, which is logarithmic in z. This linear-log scale is affine invariant to transformations:

$$G(z) = \frac{(1 + \beta z)^\alpha - 1}{\beta}$$

because $w \circ G = \alpha w \sim w$. The transformation, G, is linear for small magnitudes of z and power law for large magnitudes of z.

The linear-log base scale, w, yields the probability distribution:

$$q_z = ke^{-\lambda w} = k(1 + \beta z)^{-\gamma}$$

for $\gamma = \lambda\alpha$. This expression is the commonly observed Lomax or Pareto type II distribution, which is equivalent to an exponential-Boltzmann distribution for small z and a power law distribution in the upper tail for large z.

We can combine base scales. For example, if we start with w_1, a rotationally invariant scale, $z \mapsto z^2$, and then transform those rotationally invariant values to a linear-log scale, w_2, we obtain $w_2[w_1(z)] = \alpha \log(1 + \beta z^2)$. This scale corresponds to the generalized Student's distribution:

$$q_z = k(1 + \beta z^2)^{-\gamma}$$

For small magnitudes of z, this distribution is linear in scale and Gaussian in shape. For large magnitudes of z, this distribution has power law tails. Thus, a rotationally invariant linear-log scale grades from Gaussian to power law as magnitude increases.

Appendix B.4. The Family of Canonical Scales

The canonical scale, T_z, determines the associated probability pattern, $q_z = ke^{-\lambda T_z}$. What determines the canonical scale? The answer has two parts.

First, each problem begins with a base scale, $w(z) \equiv w$. The base scale arises from the invariances that define the particular problem. Those invariances may come from observation or by assumption. The prior sections gave the examples of rotational invariance, associated with squared-value scaling, and linear to power law invariance, associated with linear to log scaling. When the base scale lacks intrinsic invariance, we may write $w \equiv z$. Earlier publications provided examples of common base scales [7–9].

Second, the canonical scale arises by transformation of the base scale, $T_z = T(w)$. The canonical scale must satisfy both the shift and stretch invariance requirements. If the base scale itself satisfies both invariances, then the base scale is the canonical scale, $T_z = w$. In particular, if the probability

pattern remains invariant to affine transformations of the base scale $w \mapsto \delta + \gamma w$, then the shift- and stretch-invariant distribution has the form:

$$q_z = ke^{-\lambda w} \tag{B2}$$

Alternatively, w may satisfy the shift invariance requirement, but fail the stretch invariance requirement [8,9]. We therefore need to find a canonical transformation $T(w)$ that achieves affine invariance with respect to the underlying shift, $G(w) = \delta + w$. The transformation:

$$T_z = T(w) = e^{\beta w} \tag{B3}$$

changes a shift invariance of w into a stretch invariance of T_z, because:

$$T(\delta + w) = e^{\beta(\delta + w)} = e^{\beta \delta} e^{\beta w} = bT \sim T$$

for $b = e^{\beta \delta}$. We can write $T(\delta + w) = T \circ G$, thus, this expression shows that we have satisfied the affine invariance $T \circ G \sim T$ of Equation (9).

Thus, shift invariance with respect to w generates a family of scaling relations described by the parameter β. The one-parameter family of canonical scales in Equation (B3) expands the canonical exponential form for probability distributions to:

$$q_z = ke^{-\lambda T_z} = ke^{-\lambda e^{\beta w}}. \tag{B4}$$

The simpler form of Equation (B2) arises as a limiting case for $\beta \to 0$. That limiting form corresponds to the case in which the base scale, w, is itself both shift and stretch invariant [8,9]. Thus, we may consider the more familiar exponential form as falling within the expanded one-parameter symmetry group of scaling relations in Equation (B3).

The expanded canonical form for probability patterns in Equation (B4) and a few simple base scales, w, include essentially all of the commonly observed continuous probability patterns [8,9].

Appendix B.5. Example: Extreme Values

In some cases, it is useful to consider the probability pattern in terms of the canonical scale measure, $dT_z = |T'|dz$. Using $T_z = e^{\beta w}$, distributions take on the form often found in the extreme value problems [8,9]:

$$q_z dz = kw' e^{\beta w - \lambda e^{\beta w}} dz$$

in which $w' = |dw/dz|$. For example, $w = z$ yields the Gumbel distribution, and $w = \log z$ yields the Fréchet or Weibull form.

Appendix B.6. Example: Stretched Exponential and Lévy

Suppose the base scale is logarithmic, $w(z) = \log z$. Then, from Equation (B4), a candidate form for probability pattern is:

$$q_z = ke^{-\lambda z^{\beta}} \tag{B5}$$

This important distribution arises in various contexts [9], including the stretched exponential distribution and the Fourier domain spectral distribution that associates with the basic Lévy distributions [12].

In this case, the probability pattern is not shift and stretch invariant to changes in the value of z, because $z \mapsto \delta + \gamma z$ changes the pattern. By contrast, if we start with the base scale $w = \log z$, then the probability pattern is shift and stretch invariant with respect to the canonical scale:

$$T_z = e^{\beta w} = z^{\beta}$$

because the affine transformation of the canonical scale, $z^\beta \mapsto \delta + \gamma z^\beta$, does not alter the probability pattern in Equation (B5), given that we adjust k and λ to satisfy the conservation of probability and the conservation of average value.

The way in which I presented these invariances may seem trivial. If we begin with Equation (B5), then of course, we have shift and stretch invariance with respect to $z^\beta \mapsto \delta + \gamma z^\beta$. However, in practical applications, we may begin with an observed pattern and then try to infer its structure. In that case, analysis of the observations would lead to the conclusion of shift and stretch invariance with respect to the canonical power law scaling, z^β.

Alternatively, we may begin with a theory that includes a complicated interaction of various dynamical processes. We may then ask what invariance property matches the likely outcome of those processes. The conclusion may be that, asymptotically, shift and stretch invariance hold with respect to $z^\beta \mapsto \delta + \gamma z^\beta$, suggesting the power law form of the canonical scale.

In general, the particular invariant canonical scale derives from observations or from assumptions about process. The theory here shows the ways in which basic required invariances strongly constrain the candidate canonical scales. Those generic constraints shape the commonly observed patterns independently of the special attributes of each problem.

References

1. Weinberg, S. Towards the final laws of physics. In *Elementary Particles and the Laws of Physics*; Feynman, R.P., Weinberg, S., Eds.; Cambridge University Press: Cambridge, UK, 1999.
2. Landau, L.D.; Lifshitz, E.M. *Course in Theoretical Physics*, 4th ed.; Butterworth-Heinemann: London, UK, 1980; Volume 2.
3. Feynman, R.P. *Statistical Mechanics: A Set of Lectures*, 2nd ed.; Westview Press: New York, NY, USA, 1998.
4. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
5. Jaynes, E.T. Information theory and statistical mechanics. II. *Phys. Rev.* **1957**, *108*, 171–190.
6. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: New York, NY, USA, 2003.
7. Frank, S.A.; Smith, E. Measurement invariance, entropy, and probability. *Entropy* **2010**, *12*, 289–303.
8. Frank, S.A.; Smith, E. A simple derivation and classification of common probability distributions based on information symmetry and measurement scale. *J. Evolut. Biol.* **2011**, *24*, 469–484.
9. Frank, S.A. How to read probability distributions as statements about process. *Entropy* **2014**, *16*, 6059–6098.
10. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.
11. Wikipedia. Partition Function (Mathematics)—Wikipedia, The Free Encyclopedia. Available online: https://en.wikipedia.org/wiki/Partition_function_(mathematics) (accessed on 19 December 2015).
12. Frank, S.A. The common patterns of nature. *J. Evolut. Biol.* **2009**, *22*, 1563–1585.
13. Bryc, W. *The Normal Distribution: Characterizations and Applications*; Springer-Verlag: New York, NY, USA, 1995.
14. Frieden, B.R. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2004.
15. Frank, S.A. D'Alembert's direct and inertial forces acting on populations: The Price equation and the fundamental theorem of natural selection. *Entropy* **2015**, *17*, 7087–7100.
16. Neuenschwander, D.E. *Emmy Noether's Wonderful Theorem*; Johns Hopkins University Press: Baltimore, MD, USA, 2010.
17. Lederman, L.M.; Hill, C.T. *Symmetry and the Beautiful Universe*; Prometheus Books: Amherst, NY, USA, 2004.
18. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 2009.
19. Beck, C.; Cohen, E. Superstatistics. *Phys. A Stat. Mech. Appl.* **2003**, *322*, 267–275.
20. Hanel, R.; Thurner, S.; Gell-Mann, M. Generalized entropies and the transformation group of superstatistics. *Proc. Natl. Acade. Sci. USA* **2011**, *108*, 6390–6394.
21. Presse, 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.
22. Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37.
23. Frank, S.A. Maximum entropy production from d'Alembert's separation of direct and inertial forces. **2015**, arXiv:1506.07312.

24. Hand, D.J. *Measurement Theory and Practice*; Arnold: London, UK, 2004.

25. Luce, R.D.; Narens, L. Measurement, theory of. In *The New Palgrave Dictionary of Economics*; Durlauf, S.N., Blume, L.E., Eds.; Palgrave Macmillan: Basingstoke, UK, 2008.

26. Gnedenko, B.V.; Kolmogorov, A.N. *Limit Distributions for Sums of Independent Random Variables*; Addison-Wesley: Reading, MA, USA, 1968.

27. Grenander, U. *Elements of Pattern Theory*; Johns Hopkins University Press: Baltimore, MD, USA, 1996.