Where Statistics Went Wrong Modeling Random Variation

Haim Shore

Ben-Gurion University of the Negev. Dep. of Indust. Eng. and Manag..

August, 2022

Where Statistics Went Wrong Modeling Random Variation

Abstract. The large variety of models of random variation, developed within Statistics to describe nature, stands in stark contrast to the general trend towards unification of the "objects of enquiry", observed in other branches of science. This suggests that perhaps we are wrong in how we model observed random variation. In this paper, we claim that random variation, observed in nature, results from two interacting sources of variation: "Identity" (formed by process/identity factors and ultimately reflected in the mode) and "Error" (formed by non-process/non-identity factors). This dichotomy defines two extreme states: Stable-identity state (there is only error variation), and lack-of-identity state (identity instability is indistinguishable from error variation; error cannot be defined). These give rise to two types of distributions, respectively: [1] Identity-full distributions (mean and standard deviation, STD, are mutually unrelated; mode merges with the mean); [2] Identity-less distributions (mean and STD differ by location and/or scale only). In this paper, we define a new lack-of-identity property (a generalized lack-of-memory property), and related terms. Six propositions are articulated and proved. The new "Random-identity paradigm" is empirically evaluated via conjectures, shown to be supported by examples from current literature. A general model of random variation is developed and its properties probed.

Keywords: Generalized memoryless property; general model of random variation; identity-full/identity-less distributions; identity instability; random-identity paradigm.

AMS 2010 subject classifications: Primary 62; Secondary E10.

1. Introduction

The development of thousands of statistical distributions to-date is puzzling, if not bizarre. An innocent observer may wonder, how in most other branches of science the historical development shows a clear trend towards unifying the "objects of enquiry" (forces in physics; properties of materials in chemistry; human characteristics in biology), this has not taken place within the mathematical *modelling of random variation*? Why in Statistics, as the branch of science engaged in modeling random variation observed in nature, the number of "objects of enquiry" (statistical distributions) keeps growing?

In other words: Where has Statistics gone wrong modeling observed random variation?

Based on new insights, gained from a recent personal experience with data-based modeling of surgery time (resulting in a trilogy of published papers, Shore 2020ab, 2021), we present in this paper a new paradigm to modeling observed random variation. A fundamental insight is a new perception of how the latter is generated, and how it affects the form of the observed distribution. Observed random variation is perceived to be generated not by a single source of variation (as the common concept of "random variable", r.v., implies), but by two *interacting* sources. One source is "Identity", formed by "identity factors". This source is represented in the distribution by the mode (if one exists), and it may generate identity-variation. A detailed example for this source, regarding modeling of surgery times, is presented in Shore (2020a). Another source is an interacting error, formed by “non-identity/error factors”. This source generates error variation (separate from identity variation). Combined, the two interacting sources generate the *observed* random variation. The random phenomenon, generating the latter, may be in two extreme states: An identity-full state (there is only error variation), and an identity-less state (identity factors become so unstable as to be indistinguishable from error factors; identity vanishes; no error can be defined). Scenarios, residing in between these two extreme states, reflect a source of variation with *partial* lack of identity (LoI).

The new "Random Identity Paradigm", attributing two contributing sources to observed random variation (rather than a single one, as to date assumed), has far reaching implications to the true relationships between location, scale and shape moments. These are probed and demonstrated extensively in this paper, with numerous examples from current Statistics literature (relate, in particular, to Section 3).

In this paper, we first introduce, in Section 2, basic terms and definitions that form the skeleton for the new random-identity paradigm. Section 3 addresses implications of the new paradigm in the form of six propositions (subsection 3.1) and five predictions (presented as conjectures, subsection 3.2). The latter are empirically supported, in Section 4, with examples from the published Statistics literature. A general model for observed random variation (Shore, 2020a), bridging the gap between current models for the two extreme states (normal, for identity-full state; exponential, for the other), is reviewed in Section 5, and its properties and implications probed. Section 6 delivers some concluding comments.

2. "Lack of Identity", "The identity-less Property", "Identity-less/Identity-full Distributions" — definitions with examples

Due to instability in at least one of its identity-forming factors (as ultimately reflected in the mode), a process may become random-identity. Outwardly, this is observed as "Lack of Identity" (LoI), to various degrees. Earlier, we loosely used the concept of LoI to describe two extreme process states (extreme LoI), one where all variation is due to error, another when the distinction disappears between internal and external variabilities (generated by identity-factors and by error, respectively). This results in a total loss of identity and, consequently, in an identity-less distribution (like the memoryless exponential).

In this section, we deliver formal definitions of LoI and the identity-less property, define Identity-less/ identity-full distributions and address some properties that they own (other properties are related to in the form of propositions and conjectures in Section 3).

Definition of Lack of Identity (LoI – None, partial or total)

A random phenomenon (like a random process) loses identity altogether, to become identity-less (total LoI), when its internal factors (identity factors) become so unstable as to be indistinguishable from non-identity/error factors. Conversely, a process reaches maximal identity to become identity-full (no LoI) when its identity factors are stable (producing signal only) so that the only source of variation is error (Henceforth, it is assumed that error is multiplicative; a multiplicative error becomes additive for an identity-full process; relate to the model in Section 5).

Equivalently, a process is identity-less when its mean and STD measure same "entity", namely, they differ by location and/or scale only (STD is expressible as a parameter-free linear transformation of the mean). Conversely, a process is identity-full when its mean and STD are mutually unrelated, namely, each may be changed without causing change in the other.

A random phenomenon is random-identity/partial-LoI (namely, neither identity-less, nor identity-full) when its identity factors contribute to the observed random variation, yet some distinction is preserved between signal-affecting and noise-affecting factors (identity and error factors, respectively). This shows in the existence of a mode distinct from the mean (as later elaborated on, Sections 3 and 4).

Affiliated with an identity-less source (forming identity-less observed random variation) is an Identity-less Distribution.

Defining an Identity-less Distribution

An identity-less distribution has the property that, if truncated to the left at any point, the truncated distribution is identical to that prior to truncation, except possibly for a change in scale. In other words, when one observes the right tail of the distribution at any truncation point — the truncated distribution looks the same (apart from change in scale).

Since the set of identity-less distributions includes, as a subset, memoryless distributions (like the exponential and the geometric), we denote this property the "generalized memory-less property", or, alternatively, the "identity-less property". We now define it technically.

A distribution owns the identity-less property if for any two points of truncation, x_i and x_j, the right tail of the truncated distribution looks the same (except possibly for change in scale):

$$P(X \geq x_i + K_i\delta \mid X > x_i) = P(X \geq x_j + K_j\delta \mid X > x_j), \qquad (1)$$

where $\{x_i, x_j\}$ are points of truncation, and $\{K_i, K_j\}$ are respective scale parameters, possibly dependent on the point of truncation.

Examples for identity-less distributions

The uniform-distribution, defined on the interval $\{a,b\}$ and left-truncated at x_i:

Prior to truncation:

$$P(X \geq a + \delta) = 1 - \frac{\delta}{b-a};$$

After left-truncating at x_i ($a<x_i<b$):

$$P(X \geq x_i + \delta \mid X > x_i) = \frac{P(X \geq x_i + \delta)}{P(X \geq x_i)} = \frac{[b-(x_i+\delta)]/(b-a)}{(b-x_i)/(b-a)} = 1 - \frac{\delta}{b-x_i}. \qquad (2)$$

As per the above definition, the right tail looks the same, except for a change in scale (from $1/(b-a)$ to $1/(b-x_i)$). Therefore, the uniform-distribution is identity-less. This explains why the distribution function (the cumulative density function, CDF), considered as an r.v., pursues a uniform-distribution (namely, it is identity-less).

The exponential-distribution, with parameter λ:

$$P(X \geq 0 + \delta) = \exp(-\lambda\delta);$$

$$P(X \geq x_i + \delta \mid X > x_i) = \frac{P(X \geq x_i + \delta)}{P(X \geq x_i)} = \frac{\exp[-\lambda(x_i+\delta)]}{\exp(-\lambda x_i)} = \exp(-\lambda\delta). \qquad (3)$$

Note, that apart from possible change in scale (like with the uniform-distribution), the expressions for the right tails look the same prior to truncation and after.

We are aware of three identity-less distributions (per the above definition): the uniform, the exponential and the geometric. We cannot ascertain that these are the only ones.

Note, that the uniform-distribution is identity-less but not memory-less (as per the traditional definition of the memory-less property).

There are some immediate implications to the above definition, specifying important properties of identity-less/identity-full distributions. These are presented as propositions (with proves) and as conjectures, respectively, in subsections 3.1 and 3.2. In Section 4 the conjectures are supported via *theory-based* examples (from the statistics literature), and also *empirically* supported via an arbitrarily-selected set of twenty-seven distributions.

3. **Implications of the random-identity paradigm (propositions and conjectures)**

3.1 Propositions

Proposition 1 (mode of identity variation): An identity distribution, defined as a model for identity variation, either does not own a mode, or, if one exists, it resides at either bound of the distribution support (at the lower bound, if observed random variation is non-bounded from above; at the upper bound, otherwise).

Proof: Identity variation, by definition, cannot have a mode (*within* the allied distribution support). A mode, by definition, represents the most typical value. If observed random variation owns a mode, different from the mean, this implies that variation is generated by both identity-factors and error-factors. Therefore, identity variation, generated by identity-factors, by definition cannot have its own identity (distinct from error), hence cannot own a mode (unless at either bound of the distribution support).
(Note that a model for *identity* variation differs from a model for *observed* random variation; a model for the latter, with its constituent components, is developed in Section 5; In this paper, we assume that observed random variation is unbounded from above.)

Corollary: For a model of identity-variation, a change in mean implies also a change in STD and vice versa (mean and STD are mutually related).

Proposition 2 (relationship between mode and mean): As observed random variation transitions from an identity-full to an identity-less state, the mode departs from the mean to approach either bound of the distribution support.

Proof: Increasing volatility (instability) of identity factors weakens identity (as represented by the observed mode). Therefore, as identity becomes more volatile (obscure), the mode departs from the mean to approach either bound of the distribution support (refer to Proposition 1). For an identity-less state, error becomes undefined, therefore observed random variation becomes "identity" variation only, and the mode reaches either bound (refer to Proposition 1). For an identity-full state, observed random variation becomes error variation (there is no identity variation), therefore the mode merges with the mean.

Proposition 3 (pdf inflexion point): For an identity-less state, the affiliated identity-less probability density function (pdf) has no inflexion point (second derivative is never zero).

Proof: If this property is violated (does not exist), the right tail of a truncated identity-less distribution could not possibly look the same for any truncation point (as this distribution is defined; Refer to Section 2).

Proposition 4 (relationship between STD and the mean for the extreme states): For an identity-full distribution, the mean and STD are mutually unrelated. For an identity-less distribution, STD is a parameter-free linear transformation of the mean.

Proof: The first part of the proposition is self-evident (observable variation is error variation). Regarding the second part, total LoI implies no distinction between identity and non-identity (error) factors. Therefore, the mean and STD measure "same thing", implying that one is a parameter-free linear transformation of the other.

Proposition 5 (relationship between the mode and STD in an identity-full state): For an identity-full distribution, the mode and STD are mutually unrelated.

Proof: By definition, in an identity-full state there is strict separation between signal factors (forming identity) and noise factors (forming error variation). Therefore, the mode merges with the mean, and likewise becomes mutually unrelated to STD (Proposition 4).

Proposition 6 (shape moments for extreme states): For both identity-full and identity-less distributions, shape measures (like skewness and kurtosis) are constant (non-parametric).

Proof: In an identity-full state, identity-factors do not contribute noise, implying that an identity-full distribution is an error distribution, and the mode merges with the mean. Since shape moments change with the mode, however they are invariant to change in mean, this implies that in an identity-full state shape measures cannot be parametric. Conversely, for an identity-less state, a mode, if it exists, resides at either bound of the support of the observed-variation distribution (Proposition 1). Since change in shape is conditioned on change in mode, this implies that in both extreme states shape measures are constant (non-parametric). See also Conjecture 1 in subsection 3.2.

Comment: In this paper, we use the term "observed random variation" to denote the type of random variation, to which the random-identity paradigm applies. The term assumes that there are two types of random variation:

- Random variation *observed* in nature ("Observed random variation");
- Random variation associated with a *function* of r.v.s ("Calculated random variation").

This distinction leads to a bi-partition of distributions:

- **Category A**: Distributions of r.v.s that represent direct observation of nature. These r.v.s may fluctuate between the identity-full and identity-less states, with all implications for distribution properties that this entails (as implied by the random-identity paradigm, and as the latter is expounded in this paper);
- **Category B**: Distributions of r.v.s that represent *mathematical functions* of other r.v.s. (like sample statistics). These distributions may submit to the random-identity paradigm, but not necessarily so. Consequently, these distributions may have arbitrary values of shape moments (like skewness and kurtosis).

Empirical support for this bi-partition has recently been gained in a data-based study, reported in Shore (2020a). In this study, a database of ten thousand surgery times, classified into over 120 medically-specified subcategories, had been statistically analysed (therein, subsection 5.1). Per the random-identity paradigm, and given the large sampling error associated with sample estimates of third moment (*Sk*), we expect subcategory sample-skewness to fluctuate between, say, -0.5 to 2.5 (given that the normal and exponential, the ultimate representatives of identity-full and identity-less states, respectively, have skewness values 0 and 2). A count of the number of subcategories with skewness values in this interval showed that out of 126 medically-specified subcategories, 108 (86%) indeed had skewness value in this interval. This is strong support for the validity of the above bi-partition. In this paper, Category A distributions only are addressed.

3.2 Conjectures

The conjectures herewith are statements about general properties of distributions, as implied by the random-identity paradigm. Empirical evidence (empirical support) for these conjectures is provided in Section 4. Five conjectures are introduced, relating to pure functions (defined prior to Conjecture I, where they are addressed), the mode (Conjecture II), coefficient of variation (CV, Conjecture III), functional "structure" within otherwise purely identity-less functions (Conjecture IV) and non-linear transformations to approach either of the extreme states (Conjecture V).

All conjectures relate to distributions that submit to the random-identity paradigm (Category A distributions, *possibly* B, as addressed in an earlier comment at the end of subsection 3.1).

For Conjecture I we now define a pure function.

Definition: A pure function is a function of independent identity-full or identity-less r.v.s (but not both).

Conjecture I (pure function): A pure function preserves the property (namely, be identity-full/identity-less), provided the relationship between the mean and STD is respectively preserved (see Conjecture IV for scenarios where the expected relationship may not be maintained for a pure function).

Explanation: When all components of the function are identity-less, no source for identity exists. Therefore, the function will also be identity-less, provided the mean and STD are linearly related in a parameter-free relationship (relate to Conjecture IV). Conversely, when all components are identity-full, no identity-factor exists that may generate identity variation. Therefore, the function will also be identity-full, provided the mean and STD are mutually unrelated. This conjecture is supported in subsection 4.1, based on examples.

Conjecture II (mode): The standardized mode (mode divided by STD) and skewness are inversely related (the former is a decreasing function of the latter). Furthermore, as skewness approaches the value associated with an identity-full state (0), the dependence of the mode on skewness weakens until it expires (at Sk=0, the mode assumes the value of the mean, which is unrelated to skewness). Conversely, as skewness approaches the value associated with an identity-less state, the standardized mode either vanishes (become non-distinct), or reside at either bound of the distribution support (relate to Proposition 1).

Explanation: The most prominent manifestation of identity is the mode, and the standardized mode is its best representative. A good analogy is human behaviour. The most frequently observed pattern of conduct is reflection of a person's identity. Conversely, lack thereof is manifested by lack of a typical (most commonly observed) pattern of conduct.

By similar vein, a standardized mode progressing towards either bound of the distribution support is reflection of a process losing identity, with skewness value progressing towards the identity-less value (2, in the exponential case, or any other value of an identity-less distribution, where STD is a parameter-free linear transformation of the mean; relate to Proposition 4). Conversely, when identity-factors are becoming less volatile, identity is stabilized to become more observable. This results in the standardized mode moving towards the mean, with skewness approaching its error value (identity-full value, for example, zero in the normal case).

Note, again, that once an identity-less state has been reached (mode becomes the bound of the distribution support, or is non-distinct), no error can *logically* be defined.
This conjecture is empirically supported in subsection 4.2, based on examples and a sample of existent distributions.

Conjecture III (CV): The same parameter(s) that affect coefficient of variation (CV) also affect skewness. Furthermore, the effect is similar — a decreasing/ increasing relationship between a parameter and CV is preserved in its relationship with skewness. (Note, that we exclude the trivial case when a parameter represents either STD or the mean).

Explanation: According to the new paradigm, CV can be reduced if, and only if, identity instability is reduced, causing observed random variation to become error variation.

Therefore, reducing CV implies convergence to an identity-full distribution and to its skewness value (0). This conjecture is empirically supported in subsection 4.3, based on examples and a sample of existent distributions.

Conjecture IV (functional "structure" within otherwise purely identity-less functions): Any form of "structure", embedded in an otherwise purely identity-less function, may generate identity that prevents the function from becoming identity-less.

Explanation: Functional "structure" may prevent a pure function from displaying the typical relationship between the mean and STD (typical to an identity-less state). Therefore, it might not be identity-less. Examples for "structure" are functional constants (unrelated to distributional parameters of individual r.v.s), interaction effects (as in a product of r.v.s) and active constraints on the function. This conjecture is empirically supported in subsection 4.4, based on examples and a sample of existent distributions.

Conjecture V (transformations to approach extreme states): The shape of the distribution of a random variable may be changed by a non-linear transformation in order to approach the shape associated with either of the extreme states (identity-full or identity-less).

Explanation: In reality, extreme states may be approached by change in the stability of identity-factors. However, it is also well-known that non-linear mathematical transformations of r.v.s may change the shape of the allied distribution (like the Box-Cox transformation). This conjecture states that there are several paths to transform distributional shape (of observed random variation) in order to approach the shape associated with either of the extreme states. Why that is feasible, as a result of the random-identity paradigm, will be explained, and demonstrated with examples, in subsection 4.5.

4. Empirical support of conjectures (expounded in subsection 3.2)

In this section, we deliver empirical evidence for the conjectures, based either on known theory-based results, or on general properties of distributions, as displayed by a set of twenty-seven distributions (same as in Shore, 2015). The distributions are detailed in Appendix. Their first three moments are given in Supplementary Material. Some of the distributions have a mode that is either constant, or is expressible in terms of the distribution's parameters. These serve to demonstrate properties of the standardized mode (as implied by the random-identity paradigm) in the examples of Conjecture II (subsection 4.2). Distribution notation pursues Mathematica™.

4.1 Conjecture I (pure functions that preserve the property)

Example 4.1a **(A randomly-stopping sum with only identity-less components, exponential and geometric):** Consider a randomly-stopping sum, S_N, comprising N i.i.d exponential variates with parameter λ, where N is a r.v.:

$$S_N = X_1 + X_2 + .. + X_N \ , \ N \geq 1. \tag{4}$$

For N pursuing any distribution, other than the geometric, S_N will have some identity (relative to the definition above), even though its individual elements, $\{X_i\}$, do not (since they are exponential).

However, suppose that N also pursues an identity-less distribution (the geometric, a discrete analogue of the exponential). This implies that both individual additive elements comprising S_N and their number (N) are ruled by identity-less distributions (signal and noise represented by same parameter(s)). By Conjecture I, we expect S_N also to be identity-less (exponentially distributed). Using known formulae for the mean and variance of random sums:

$$\mu_{S_N} = E(X)E(N)\ ;\ \sigma^2_{S_N} = Var(X)E(N) + [E(X)]^2 Var(N)\ , \qquad (5)$$

we introduce $1/\lambda$ for the mean and STD of the exponential (with parameter λ), with $1/p$ and $(1\text{-}p)/p^2$ for the mean and variance, respectively, of the geometric (with parameter p), to obtain:

$$\mu_{S_N} = E(X)E(N) = (\frac{1}{\lambda})(\frac{1}{p})\ ;$$

$$\sigma^2_{S_N} = Var(X)E(N) + [E(X)]^2 Var(N) = (\frac{1}{\lambda})^2[\frac{1}{p} + \frac{(1-p)}{p^2}] = \mu^2_{S_N}\ . \qquad (6)$$

As predicted, the mean and STD are identically equal, indicating that the random sum is also identity-less (exponential). Example 4.3d articulates conditions for CV, under which, consistent with the random-identity paradigm, a randomly-stopping sum pursues an identity-full distribution (normality).

Example 4.1b (A randomly-stopping sum with only identity-less components, uniform): Consider another scenario of a randomly-stopping sum, S_N (as in Example 4.1a). For N constant, the sum of N *i.i.d* standard uniform r.v.s follows the Irwin–Hall distribution. However, suppose that S_N comprises N i.i.d uniformly distributed r.v.s with support {0, *b*}, and N is distributed as discrete uniform with support {1, n}. Both random components are identity-less (since the uniform is identity-less, as shown in Section 2). Deriving the unconditional mean and variance of S_N, similarly to Example 4.1a, we obtain mean and variance of a uniform distribution with support {0, b(1+n)/2} (calculations not given here as they are easily reproducible). This is strong corroboration that S_N is uniform, namely, a purely identity-less function.

Example 4.1c (A compound Gaussian distribution): For a Gaussian r.v., Y, with a normal independently distributed mean (namely, all distributional components are identity-full), the unconditional distribution of Y is Gaussian (Wikipedia, entry "Compound probability distribution").

Example 4.1d (A weighted-sum of identity-less r.v.s and identity-less weights): Consider the following weighted sum:

$$Y = pX_1 + (1-p)X_2 \ , \qquad (7)$$

where X_1 and X_2 are *i.i.d* exponential variates with a common parameter λ. Suppose that p is constant. It is easy to show that the conditional distribution of Y, given p, is not exponential (mean and STD are not equal unless p equals 0 or 1). The parameter p is a "built-in" structure that blocks the response becoming identity-less (refer to Conjecture IV). However, suppose that the composition of the two components of Y is not stable, and p is uniformly distributed on {0,1}. The uniform distribution is identity-less (Section 2). Therefore, when it appears as a weighting distribution in a mixture distribution, it renders the weighted sum, per Conjecture I, also identity-less (all sources of variation are identity-less). It is easily shown that the (unconditional) mean and STD of Y are equal ($1/\lambda$), implying that Y is identity-less (exponential).

Example 4.1e (A compound distribution with an identity-less parameter, two cases):

Case 1 (Poisson-exponential): Consider a r.v., Y, having conditional Poisson distribution, given parameter α:

$$Y/\alpha \sim Poisson[\alpha]. \qquad (8)$$

The Poisson maintains a common parameter for signal and noise (like the exponential), Furthermore, for α integer, it may be shown that the mode has multiple values, α and α-1, resembling, in that sense, the identity-less discrete uniform distribution (that has more than a single mode). It also resembles an identity-full distribution, since the mode and the mean merge.

However, the criterion for these two types of distributions, as addressed earlier, is the relationship between STD and the mean. The Poisson does not maintain any, therefore the Poisson is neither identity-full, nor identity-less. Assume now that parameter α is random and exponentially distributed with parameter λ. It may be easily shown that:

$Y \sim$ Geometric[1/(1+λ)], (9)

namely, the unconditional distribution of Y is identity-less (geometrically distributed).

We learn from this example that a function of r.v.s may become identity-less, even when not all of its constituents are strictly identity-less (like the Poisson, in this case), conditioned on these constituents having each a single parameter for both the mean and STD. Such cases need further study.

Case 2 (Uniform-exponential): Consider a r.v., Y, having conditional uniform distribution with support {0,α}:

$Y / \alpha \sim Uniform[\{0,\alpha\}].$ (10)

Suppose that α is exponentially distributed with parameter λ. It may be easily shown that its probability density function (pdf), mean and CV are, respectively:

$f(x)= \lambda$ gamma(0,λx), $x>0$;

$\mu=1/(2\lambda)$; (11)

$CV= \sigma/\mu= (5/3)^{(1/2)}$.

The second derivative of the cumulative density function (CDF) is that of the exponential ($-\lambda^2 e^{-\lambda x}$) divided by ($\lambda x$). Both derivatives have roots (becoming zero) only at infinity, implying that the pdf does not have an inflexion point (as typical to identity-less distributions). Furthermore, the mean and STD differ by scale only (CV is constant). It is deduced that the distribution is identity-less, in conformance with the random-identity paradigm.

Example 4.1f (A sum and a ratio of two independent normal r.v.s):

Case 1: Consider two independent r.v.s, X_1 and X_2, with means and standard deviations $\{\mu_1,\sigma_1\}$ and $\{\mu_2,\sigma_2\}$, respectively. The sum: $Z=X_1+X_2$, has mean which is the sum of the means, and variance which is the sum of the variances. Therefore, the mean and STD of Z are mutually unrelated, and according to Conjecture I, Z is identity-full (Z is normal).

Case 2: Consider two independent r.v.s, X_1 and X_2, with means and standard deviations $\{\mu_1,\sigma_1\}$ and $\{\mu_2,\sigma_2\}$, respectively. The ratio: $Z=X_1/X_2$, has mean which depends on three parameters (σ_1 excluded), while the variance depends on all four parameters (both the mean and STD of Z can be easily obtained via Mathematica™). We realize that the mean and STD are not mutually unrelated: STD may be changed by σ_1 with no effect on the mean, yet a change in mean (via change in any of its three parameters) is also a change in STD. Therefore, Z cannot be identity-full and is not normal (as is well known).

4.2 Conjecture II (mode)

Example 4.2a (relationship between mode and skewness): To learn empirically how the mode is related to skewness, we select thirteen distributions (of the set of twenty-seven, Appendix). These are (numbers as therein):

{1, 2, 3, 4, 7, 11, 12, 13, 14,17, 18, 20, 21}.

Note that all distributions have mode that is either constant or expressible in terms of the parameters. We first attach arbitrary values to the parameters of the distributions (Set 1), and then arbitrarily double all parameters (Set 2). This is done to avoid a possible bias in selecting parameters' values.

The relationships between the standardized mode and skewness for the two sets are displayed, respectively, in Figures 1 and 2.

We learn that Conjecture II is indeed empirically supported. As skewness approaches the identity-full value (say, $0\leq Sk\leq 0.5$), and factors sets, which determine signal and noise, start to separate, values of the mode become arbitrary (large dispersion). Conversely, as skewness grows (say, Sk>0.5), there is an inverse relationship with the standardized mode, a clear indication of a process losing identity.

4.3 Conjecture III (CV)

Example 4.3a (The Central Limit Theorem, CLT): Let $\{X_1, X_2, \ldots, X_n\}$ be a set of n independent r.v.s with a common mean, μ, and STDs $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, respectively. Assume that this set represents a random sample of n observations. The sample mean has CV equal to:

$$CV = (\frac{1}{n\mu})\sqrt{\sum_{i=1}^{n}\sigma_i^2} \tag{12}$$

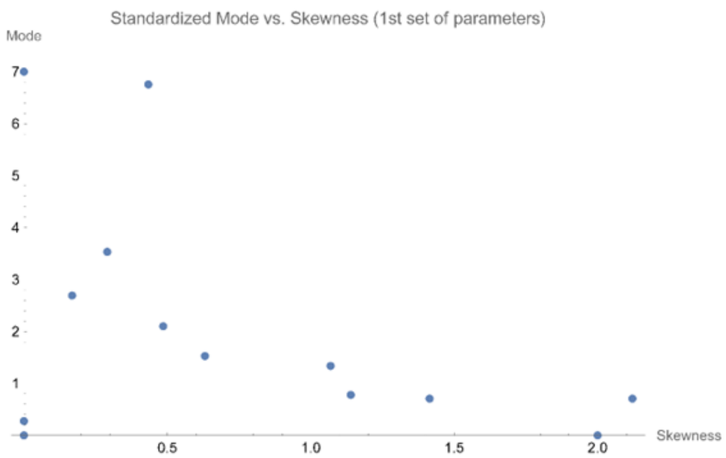

Figure 1. Standardized mode *vs.* skewness for a sample of thirteen distributions (with arbitrary parameter values). Note the increased dispersion for small Sk values (say, Sk≤0.5).

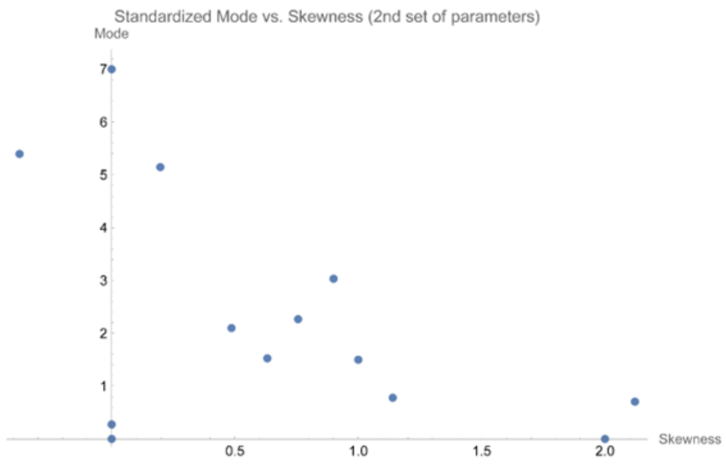

Figure 2. Standardized mode *vs.* skewness for a sample of thirteen distributions (with all parameter values of Figure 1 doubled). Note the increased dispersion for small Sk values (say, Sk≤0.5).

Since increasing *n* decreases CV, the distribution of the average asymptotically converges to an identity-full distribution, namely, the normal distribution (as CLT asserts, consistent with Conjecture III).

Example 4.3b (Asymptotic normality with asymptotic zero CV, irrespective of CLT): The relationship between asymptotic normality and CV tending to zero is a straightforward outcome of the new paradigm. It shows up in all cases where a distribution tends to normality — same parameter(s) that cause a distribution to tend to normality also cause CV to tend to zero (as mandated by the new paradigm).

Examples:

- The binomial distribution with parameters $\{n,p\}$, and the Poisson distribution with parameter λ:

$$CV_{Bin}=\frac{\sqrt{(1-p)/p}}{\sqrt{n}}\,;\ CV_{Pois}=\frac{1}{\sqrt{\lambda}}; \tag{13}$$

- The negative binomial distribution with parameters {n,p}:

$$CV_{NB} = \frac{1}{\sqrt{(1-p)n}}; \tag{14}$$

- The gamma distribution with parameters {α,β}, where β is a shape parameter:

$$CV_{Gam} = \frac{1}{\sqrt{\beta}}; \tag{15}$$

- The beta distribution with parameters {α,β}. Evans *et al*. (1993) states (p. 36) that as the ratio α/β remains constant and both α and β tend to infinity, the beta variate tends to the standard normal. For the beta distribution, CV is:

$$CV_{Beta} = [\frac{\beta}{\alpha(\alpha+\beta+1)}]^{1/2} \tag{16}$$

As implied by the new paradigm, same condition for a beta variate to tend to normality also causes *CV* to tend to zero.

- The lognormal distribution with parameters {α,β}:

$$Sk = CV(3+CV^2). \tag{17}$$

Example 4.3c (Skewness and CV for sum of N i.i.d exponential variates): Consider the sum of N i.i.d exponential variates, with parameter λ:

$$S_N = X_1 + X_2 + .. + X_N \text{ , } N \geq 1. \tag{18}$$

The distribution of S_N is known as Erlang. It is in fact a gamma distribution, with an integer parameter, *N*. The skewness of S_N (*Sk*) and its *CV* are:

$$Sk = \frac{2}{\sqrt{N}} \text{ ; } CV = \frac{1}{\sqrt{N}} \text{ .} \tag{19}$$

We learn that as N increases, the relative *STD* (*CV*) decreases (due to diminishing internal noise relative to the signal, identity of S_N is stabilized). Concurrently, S_N approaches normality. Conversely, as $N\rightarrow 1$, *CV* increases and *Sk* tends to the exponential value (2).

Example 4.3d (Under what constraints does a randomly-stopping sum become normal): The CV of a randomly-stopping sum, S_N, is given by:

$$CV_{S_N}^2 = \frac{CV_X^2}{\mu_N} + CV_N^2 \, , \tag{20}$$

where CV_X and CV_N are coefficients of variation of X and N, respectively, and μ_N is the mean of N. By the new paradigm, normality will be restored asymptotically if, and only if, CV_{SN} tends to zero, namely, both CV_X and CV_N tend to zero (assuming μ_N is non-zero and finite). For example, suppose that X is the average of a random sample of n_1 observations and N is binomially distributed with parameters $\{n_2,p\}$. Eq (20) implies, based on the new paradigm, that as both n_1 and n_2 grow — S_N would tend to normality.

Note that for a random sum S_N:

$$K_{S_N}(t) = K_N[K_X(t)] \, , \tag{21}$$

where K_{SN}, K_N *and* K_X are the respective cumulant generating functions of S_N, N and X. This implies that for any choice of distributions for X and N, the cumulant generating function of S_N can be derived to ascertain whether S_N is indeed asymptotically normal (as predicted by the random-identity paradigm).

Example 4.3e (Relationship between CV and skewness in statistical distributions): For the set of twenty-seven distributions, related to earlier, we assign arbitrary values to the parameters of the distributions. We then delete negatively-skewed distributions and extremely positively-skewed distributions (with skewness measure higher than 6.5) to obtain a subset of twelve distributions. The following significant linear-regression relationship between *CV* and *Sk* is obtained (corr.=0.8821; *p*=E-6; sample size=12):

$$CV = 0.3014 + 0.2049(Sk) \tag{22}$$

Figure 3 displays a scatter-plot of *CV* vs. skewness for this sample of theoretical distributions. This empirical relationship can be anticipated (and explained) only by the random-identity paradigm. As addressed in subsection 5.1 of Shore (2020a), this relationship is reproduced by statistical analysis of surgery times of 126 subcategories of surgeries, derived from a database of nearly ten thousand surgeries (refer to Figure 3 therein). The generalized exponential distribution (addressed in Section 5 and in Supplementary Material) also reproduces this relationship (Shore, 2020a).

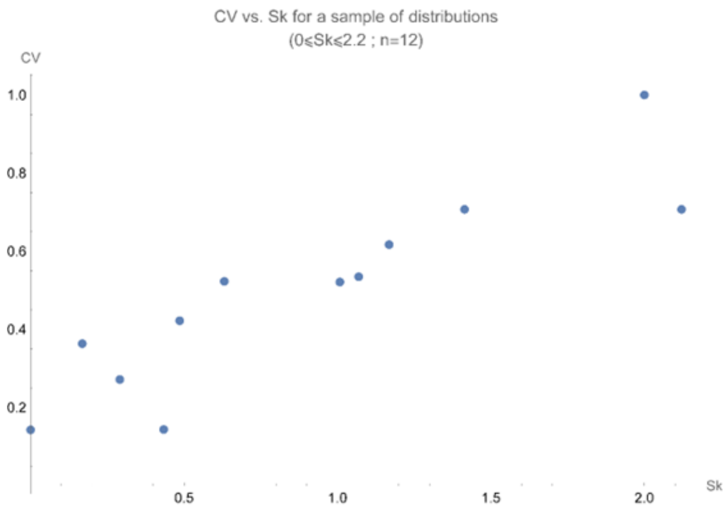

Figure 3. Coefficient of variation (*CV*) *vs.* skewness (*Sk*) for twelve different distributions with arbitrary parameter values.

Example 4.3f (General properties of distributions, relating to CV and its components): In this example, we articulate predictions regarding general properties of distributions (as implied by Conjecture III). We use the set of twenty-seven distributions to demonstrate realization of the following predictions:

- **Prediction 1:** If the mean and STD are expressed by different parameters — the distribution is symmetric (*Sk*=0; There is only error variation, which, by definition, is symmetric); **Examples for Prediction 1**: [1], [8], [22], [23] (refer to Appendix for the identity of the distribution linked to each number);
- **Prediction 2:** If *CV* is constant/non-parametric (after possible re-location of distribution support to include zero) — *Sk* is also constant, including, possibly, *Sk*=2 (namely, signal and noise are affected by same set of factors); **Examples for prediction 2:** [2], [12], [14], [17], [18], [20], [24], [26] (refer to Appendix for the identity of the distribution linked to each number); Note that the inverse is not necessarily true — *Sk* can be constant, as in the normal (Prediction 1), while *CV* is not;
- **Prediction 3:** If *CV* varies (it is parametric), same parameters that affect *CV* also affect *Sk* (namely, both *CV* and *Sk* are affected by shape parameters). **Examples for prediction 3:** [3], [4], [5], [6], [7], [9], [10], [11], [13], [15], [16], [19], [21], [25], [27] (refer to Appendix for the identity of the distribution linked to each number). Regarding [15] (Beta distribution) — see details in Example 4.3b.

Note, that all twenty-seven distributions have each been classified into a single group (of the three mutually exclusive groups, defined by the three predictions above). Note, again, that all predictions stem exclusively from the random-identity paradigm.

4.4 Conjecture IV (structure in a function of purely identity-less r.v.s)

Example 4.4a (A randomly-stopping product): Consider the random product, P_N, with distributions as described earlier (Example 4.1a):

$$P_N = X_1 * X_2 * .. * X_N \ , \ N \geq 1. \tag{23}$$

Individual elements, $\{X_i\}$, and their random number, N, are all identity-less (exponential and geometric, respectively). Therefore, as mandated by the new paradigm, P_N seemingly lacks identity (as in Example 4.1a). However, being a product of r.v.s, P_N preserves identity/structure in the form of interaction effects. Therefore, it is not expected to be identity-less (exponential). Indeed, it can be easily shown that the mean and variance of P_N are:

$$\mu_P = \frac{p\mu}{1-\mu(1-p)} \ ; \ \sigma_P^2 = \frac{p\mu^2}{[1-\mu^2(1-p)][1-2\mu^2(1-p)]}, \tag{24}$$

where $\mu=1/\lambda$ is the mean of the exponential variate. We realize, that the mean is equal to the STD only when interaction effects vanish, namely, p=1 (the random product comprises a single exponential variate). Also, as expected, neither is the log of the product exponentially distributed (given the results of Example 4.1a).

Example 4.4b (Sum of N i.i.d exponential variates, N constant): Since all sources of variation (N r.v.s) are identity-less, by the random-identity paradigm the response is expected to be identity-less (exponential). It is not because N is constant, namely, structure is built-in to form identity. Therefore, the sum is gamma distributed, which is not identity-less (see example 4.5a below).

4.5 Conjecture V (transformations to approach extreme states)

Example 4.5a (An increasing non-linear relationship between STD and the mean, with monotone increasing slope, causes asymptotic identity-less-ness; As STD increases at a rate higher than the mean, the response skewness tends to that of an identity-less distribution): In this example, we observe distributions with non-linear relationship between the mean, μ, and STD, σ, of the form:

$$\sigma = \alpha\mu^{\beta}, \beta > 1 \quad (25)$$

Examples:

(1) For the gamma distribution with parameters {α,β}: μ=αβ; σ = μ/√α; sk=2/√α. This implies that for a constant shape parameter α, σ increases *linearly* with μ. Therefore, gamma *does not* tend to the exponential as the mean increases.

(2) Consider a normal variable, *X*, with arbitrary mean of 10 and STD 1 (CV_x=0.1). Define:

$$Y = X^7 \quad (26)$$

This nonlinear transformation produces for *Y* (*without* specifying the distribution):

CV_Y = 0.7183; Sk_Y = 1.968; Ku_Y = 9.963.

Corresponding values for an exponential variate are, {1, 2, 9}. We realize that due to noise increasing more rapidly than the signal (*CV* increases from the original 0.1 to over 0.7), identity is destabilized, and the nonlinear transformation renders an identity-full variable (normal r.v.) into identity-less (as judged by *CV* and third and fourth moments). This result may be anticipated under the random-identity paradigm.

Example 4.5b (An increasing non-linear relationship between STD and the mean, with monotone decreasing slope, causes asymptotic identity-full-ness; As STD starts to increase

at a rate smaller than that of the mean, the response skewness tends to that of an identity-full distribution; Two cases):

Case 1 (Box-Cox Transformation): A Box-Cox (BC) transformation is known to normalize data as well as stabilizing the variance. The link between the two properties has never been satisfactorily explained, unless by the new random-identity paradigm. Let X be the observed r.v. and Z the standard normal variate. The *inverse* BC transformation, with parameter λ, is:

$$X = [1 + \lambda(\alpha + \beta Z)]^{1/\lambda}, X \geq 0. \tag{27}$$

The author had opportunity to converse with both Box and Cox (find report in Shore, 2005, p. 41). I asked them what convinced them that a power transformation normalizes data. Their responses were consistent — "Transformation was conceived based on personal experience".

The new paradigm seems to deliver best explanation. The normalizing BC transformation is supposed to revoke the identity-full scenario (normal response). This is achieved by transforming the distribution of the original variate into a two-component mixture-distribution, with one component representing distribution of a r.v., defined on a small interval of the mixture-distribution support, and another component representing a near-constant. In other words, a considerable separation between parameters that represent noise and signal is achieved. A detailed numerical example for this phenomenon, the result of power transformation, is given, with respect to the exponential transformation (Manly, 1976), in Case 2 below.

Case 2 (*Exponential data transformed into normal via exponential transformation):* Manly (1976) introduced the exponential data-transformation as an alternative to the Box-Cox transformation:

$$y = \begin{cases} [exp(\gamma x) - 1]/\gamma, \gamma \neq 0, \\ x, \gamma = 0 \end{cases} \quad . \tag{28}$$

In Table 1 therein, the author presents data from apparently exponential r.v. (*Sk*=1.962, *Ku*=9.634 *vs.* the theoretic values of *Sk*=2 and *Ku*=9). The parameter γ=-0.5 causes the data to transform to normality (*Sk*=0, *Ku*=3). The negative sign of the parameter delivers the following data transformation from exponential to normal:

$$y = 2 * \left(1 - \frac{1}{exp(0.5x)} \right) \quad . \tag{29}$$

How has this been achieved? We learn that the large support of the exponential *X*, {0,∞}, is reduced into a small interval, {0,2} (approaching an asymptote value of y=2.0; For x=7, y already equals 1.94). This implies that the transformation replaces a large part of the original support by a constant (2) (say, at x>7). In practice, the transformation produces a two-component mixture distribution, one component that is random (say, for x<7), and another that is ner-constant (at about y=2, for x>7). In other words, the transformation *decouples* STD from the mean via converting the original exponential distribution into a two-component mixture distribution, with the first component mostly providing noise, the second component mostly providing signal. This is a typical scenario of an identity-full distribution, where complete separation exists between the signal component (constant) and the noise component.

Similarly, observing again Table 1 in Manly (1976), we find out that *X* data, with skewness of *Sk*=1.027, are transformed to normality with a parameter γ=-0.25, implying an asymptote value of *y*=4 (instead of *y*=2, as earlier shown for *X* data with *Sk*=2). Because the *original X* data are now closer to normality (*Sk*≅1), they look more like the response in a normal scenario (where identity-factors are signal factors, namely, constant). This causes the *constant* part of the mixture distribution, produced by the transformation, to shrink (with the asymptote value increasing from 2 to 4). For $\gamma=0$: $Y=X$, X data are already normal, and the constant component in the mixture distribution disappears altogether (namely, goes to infinity).

As noted earlier (Case 1), a similar analysis may be conducted with respect to the Box-Cox power transformation.

Example 4.5c (Too small sample size produces observed response that falsely looks identity-full, normal): This example, pointing to an observed response *falsely* looking identity-full (normal), is provided in Shore (2020; Section 4). Therein, a database of ten thousand surgeries, partitioned into medically-specified subcategories, was analysed, and the effect addressed of surgery-subcategory size on observed (empirical) distribution of surgery-duration (SD). It is shown that for relatively *small subcategory size* (in the database), the latter is positively correlated with skewness (namely, subcategory skewness increases with subcategory sample size). This strange phenomenon is self-explanatory under the new random-identity paradigm. For poorly represented subcategories, small sample size does not allow adequate reflection (representation) of subcategory actual work-content instability (sample underestimates true identity variation). Consequently, the distribution of SD, as represented in the subcategory sample, falsely looks more identity-full (symmetric, probably normal) than it really is. This bizarre and unexpected effect is compatible with the random-identity paradigm and explained by it.

Example 4.5d (Estimator consistency and asymptotic normality): Let X_n be an unbiased consistent estimator of parameter μ, based on a random sample of n observations. An estimator is consistent if it converges in probability to the true value. Empirical evidence abounds for the asymptotic normality of X_n, for example, ML estimators are known to be consistent and asymptotically normal. However, consistency coupled with asymptotic normality is a direct outcome of the random-identity paradigm. Estimator consistency, by definition, implies stabilization ("convergence in probability") via reduced STD (as *n* becomes larger). This gets the estimator closer to the identity-full scenario (internal variation vanishing, with error becoming sole source of variation). Estimator's distribution therefore should asymptotically approach symmetry, as mandated by the new paradigm.

5. A model for observed random variation (Shore, 2020a)

In this section, we briefly overview Shore's bi-variate statistical model for surgery duration, adapted here to be expressed in terms of the random-identity paradigm. Twofold *empirical validation* of the model is given in Shore (2020a) — via statistical analysis of a database of ten-thousand surgeries, and by fitting the model's new family of distributions, via a *five-moment* matching procedure, to a large sample of variously-shaped theory-based statistical distributions. Good fit is achieved for all distributions in the sample. The empirical validation of the new random-identity paradigm has been complemented, in this paper, by *theory-based* evidence (refer to Section 4). This was done by addressing known "statistical results" from the statistics literature, derived independently of one another, and showing that these are natural, internally consistent, outcomes of the new paradigm to model observed random variation.

In compliance with this paradigm, the new model comprises two interacting components, representing two independent sources of variation — internal/identity variation and external/error variation. The former is produced by identity-factors (analogously to surgery-inherent factors, technically defined in Shore, 2020a). It is represented by Y_i, an r.v. that follows a new generalized exponential-distribution (find details in Shore 2020a and in Supplementary Material). The latter, external variation, is produced by non-identity/error factors, and it is represented by Y_e, a multiplicative normal error. Consider the following model for the response R (surgery duration in Shore, 2020a):

$$R = S(Y-L) = S[(Y_iY_e)-L] = S[Y_i(1+\varepsilon)-L] = \\ S[Y_i(1+\sigma_e Z)-L], \tag{30}$$

where R is the observed response variable, L and S are location and scale parameters, respectively, Y is the standardized response (L=0, S=1), $\{Y_i,Y_e\}$ are independent r.v.s representing internal and external variation, respectively, ε is zero-mode normal disturbance (error) with STD, σ_e, and Z is standard normal. Assuming $\varepsilon << 1$, Y_e is approximately lognormally distributed.

Note that multiplying Y_i by Y_e implies an interaction effect between the two components (in their relationship to R). As shown in Shore (2020a), this is a plausible feature of the model because it implies that length of surgery, as represented by Y_i, affects STD of the error (Y_e). This is clearly demonstrated with real data therein.

A detailed derivation of the distributions of the model's components (Y_i and Y_e) and their moments are given in Shore (2020a). For the convenience of the reader, we give in Supplementary Material details about the distribution of Y_i (termed therein the extended exponential-distribution). The final model has three shape parameters, $\{\alpha, \sigma_i, \sigma_e\}$. The parameters, $\{\sigma_i, \sigma_e\}$, represent model's identity-variation and error-variation, respectively, and parameter α assumes a value of α=0 (and concurrently σ_e=0) for the exponential scenario (identity-less state), and a value of α=1 (and σ_i=0), for the normal scenario (an identity-full state). The k-th non-central moment of Y, as function of corresponding moments of Y_i and Y_e, is (due to statistical independence, since the multiplicative error is assumed independent of Y_i):

$$E(Y^k) = E(Y_i^k)E(Y_e^k). \tag{31}$$

The form of the distribution of *Y* is determined by the three shape parameters, alluded to earlier, $\{\alpha, \sigma_i, \sigma_e\}$. They may be identified (or estimated), in a moment-matching procedure, by fitting third and fourth standardized central moments (skewness (*Sk*) and kurtosis (*Ku*) measures). Note that, as shown in Shore (2020a), the three shape parameters are not independent so that matching only two moments is feasible. The location and scale parameters, *L* and *S*, may be identified (after estimating shape parameters) via matching the means and STDs (find details in Shore, 2020a).

The distribution function (cumulative density function, *CDF*) of *Y* is expounded in Shore (2020a) both for normal and lognormal errors. A demonstration of the goodness-of-fit of the new model, obtained via five-moment fitting to known diversely-shaped distributions, is also given therein.

Examining values of the parameters obtained (from implementing the fitting procedure), we realize that for most distributions — $0 \le \alpha \le 1.5$, namely, no value of α is below that of the exponential (α=0), and only few exceed somewhat the value associated with the identity-full scenario (α=1). A possible explanation is that not all distributions in the sample are from Category A (the latter describing observed random variation that is confined between the two extreme states, as addressed earlier). One particularly important property, already addressed in Shore (2020a), is the linear relationship between CV and skewness of Y_i distribution (the generalized exponential; relate to Supplementary Material). This has been earlier addressed in Example 4.3e (Conjecture III).

6. Conclusion

A new framework to model observed random variation has been developed. According to the new paradigm, a major source of random variation, revealed in *observed* random variation, is identity instability, coupled, as a secondary source, with a multiplicative error. As identity becomes more unstable, the multiplicative error ceases to be error, until it expires, once identity is lost altogether (an identity-less state). At that point, a merging of the mean with the STD takes place (STD is expressible as a parameter-free linear transformation of the mean), and the mode vanishes (residing at either bound of the distribution support), or it becomes non-distinct (as with the uniform).

A major claim of this article is that there is an identifiable unique process that generates observed random variation. This process generates response variation (*observed* random variation), where, in addition to error, random identity plays a major source of variation, largely ignored or unrecognized to-date. Thus, the new paradigm moves modelling of univariate random variation from unidimensional to bidimensional space. This extension may allow unification, under a single umbrella, of current unimodal statistical distributions (as demonstrated by the highly accurate representation of a large set of current parametric distributions, achieved via five-moment matching; Shore, 2020a).

The new random-identity paradigm introduces an element of uniformity into modelling random variation. It may be first step towards achieving the ideal of "unification of the objects of enquiry" (as related to in the opening paragraph of this paper). Compared to an earlier somewhat-technical attempt in this direction (Shore, 2015), the new one, as expounded in this paper, is based on deep insight into the inner workings of distributions, as summarized in the random-identity paradigm, and its allied propositions and conjectures. Further research is needed to explore the theoretical and practical ramifications of the new paradigm.

Appendix. List of 27 distributions (Section 4; Respective first three moments are displayed in Supplementary Material)

[1] Normal[μ,σ]; **[2]** HalfNormal[θ]; **[3]** LogNormal[μ,σ]; **[4]** InverseGaussian[μ,λ]; **[5]** ChiSquare[v]; **[6]** InverseChiSquare[v]; **[7]** FRatio[n,m]; **[8]** StudentT[v]; **[9]**

NoncentralChiSquare[ν,λ]; **[10]** NoncentralStudentT[ν,δ]; **[11]** NoncentralFRatio[n,m,λ]; **[12]** Triangular[{a,b}]; **[13]** Triangular[{a,b},c]; **[14]** Uniform[{min,max}]; **[15]** Beta[α,β]; **[16]** Chi[ν]; **[17]** Exponential[λ]; **[18]** ExtremeValue[α,β]; **[19]** Gamma[α,β]; **[20]** Gumbel[α,β]; **[21]** InverseGamma[α,β]; **[22]** Laplace[μ,β]; **[23]** Logistic[μ,β]; **[24]** Maxwell[σ]; **[25]** Pareto[k,α]; **[26]** Rayleigh[σ]; **[27]** Weibull[α,β]}.

Mathematica™ - Mathematica is a trademark of Wolfram Research Inc.

References

[1] Evans M, Hastings N, Peacock B. 1993. *Statistical Distributions*. 2nd Edition. Wiley-Interscience. ISBN-10: 0-471-55951-2.

[2] Manly, BFJ. 1976. "Exponential Data Transformations." *Journal of the Royal Statistical Society, Series D (The Statistician)* 25 (1): 37-42. http://www.jstor.com/stable/2988129.

[3] Shore H. 2005. *Response Modelling Methodology: Empirical Modelling for Engineering and Science*. Book. Series on Quality, Reliability and Engineering Statistics: Volume 8. *World Scientific Publishing Inc.* DOI: 10.1142/5708

[4] Shore, H. 2015. "A general model of random variation." *Communications in Statistics – Theory and Methods* 49 (9):1819–1841. DOI: 10.1080/03610926.2013.784990.

[5] Shore, H. 2020a. "An Explanatory Bi-Variate Model for Surgery-Duration and Its Empirical Validation." *Communications in Statistics: Case Studies, Data Analysis and Applications* 6(2): 142-166. DOI: 10.1080/23737484.2020.1740066.

[6] Shore, H. 2020b. "SPC Scheme to Monitor Surgery Duration." *Quality & Reliability Engineering International*. DOI: 10.1002/qre.2813.

[7] Shore, H. 2021. "Estimating Operating-room Utilization-rate for Differently-distributed Surgery Times". *International Journal of Production Research*. On line December, 2021. DOI: 10.1080/00207543.2021.2009141.

Supplementary Material

S1: List of 27 distributions and their first three moments (mean, variance, skewness; subsections 5.1 and 5.2)

Out[ ]=

$$\{\{1, \text{NormalDistribution}[\mu, \sigma], \mu, \sigma^2, 0\}, \{2, \text{HalfNormalDistribution}[\theta], \frac{1}{\theta}, \frac{-2+\pi}{2\theta^2}, \frac{\sqrt{2}(4-\pi)}{(-2+\pi)^{3/2}}\},$$

$$\{3, \text{LogNormalDistribution}[\mu, \sigma], e^{\mu+\frac{\sigma^2}{2}}, e^{2\mu+\sigma^2}\left(-1+e^{\sigma^2}\right), \sqrt{-1+e^{\sigma^2}}\left(2+e^{\sigma^2}\right)\},$$

$$\{4, \text{InverseGaussianDistribution}[\mu, \lambda], \mu, \frac{\mu^3}{\lambda}, 3\sqrt{\frac{\mu}{\lambda}}\}, \{5, \text{ChiSquareDistribution}[\nu], \nu, 2\nu, 2\sqrt{2}\sqrt{\frac{1}{\nu}}\},$$

$$\{6, \text{InverseChiSquareDistribution}[\nu], \begin{cases}\frac{1}{-2+\nu} & \nu>2\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{2}{(-4+\nu)(-2+\nu)^2} & \nu>4\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{4\sqrt{2}\sqrt{-4+\nu}}{-6+\nu} & \nu>6\\ \text{Indeterminate} & \text{True}\end{cases}\},$$

$$\{7, \text{FRatioDistribution}[n, m], \begin{cases}\frac{m}{-2+m} & m>2\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{2m^2(-2+m+n)}{(-4+m)(-2+m)^2 n} & m>4\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{2\sqrt{2}\sqrt{-4+m}(-2+m+2n)}{(-6+m)\sqrt{n}\sqrt{-2+m+n}} & m>6\\ \text{Indeterminate} & \text{True}\end{cases}\},$$

$$\{8, \text{StudentTDistribution}[\nu], \begin{cases}0 & \nu>1\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{\nu}{-2+\nu} & \nu>2\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}0 & \nu>3\\ \text{Indeterminate} & \text{True}\end{cases}\},$$

$$\{9, \text{NoncentralChiSquareDistribution}[\nu, \lambda], \lambda+\nu, 4\lambda+2\nu, \frac{2\sqrt{2}(3\lambda+\nu)}{(2\lambda+\nu)^{3/2}}\},$$

$$\{10, \text{NoncentralStudentTDistribution}[\nu, \delta], \begin{cases}\frac{\delta\sqrt{\nu}\,\text{Gamma}\left[\frac{1}{2}(-1+\nu)\right]}{\sqrt{2}\,\text{Gamma}\left[\frac{\nu}{2}\right]} & \nu>1\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{(1+\delta^2)\nu}{-2+\nu}-\frac{\delta^2\nu\,\text{Gamma}\left[\frac{1}{2}(-1+\nu)\right]^2}{2\,\text{Gamma}\left[\frac{\nu}{2}\right]^2} & \nu>2\\ \text{Indeterminate} & \text{True}\end{cases},$$

$$\begin{cases}\frac{\delta\sqrt{\nu}\,\text{Gamma}\left[\frac{1}{2}(-1+\nu)\right]\left(\frac{\nu(-3+\delta^2+2\nu)}{(-3+\nu)(-2+\nu)}-2\left(\frac{(1+\delta^2)\nu}{-2+\nu}-\frac{\delta^2\nu\,\text{Gamma}\left[\frac{1}{2}(-1+\nu)\right]^2}{2\,\text{Gamma}\left[\frac{\nu}{2}\right]^2}\right)\right)}{\sqrt{2}\left(\frac{(1+\delta^2)\nu}{-2+\nu}-\frac{\delta^2\nu\,\text{Gamma}\left[\frac{1}{2}(-1+\nu)\right]^2}{2\,\text{Gamma}\left[\frac{\nu}{2}\right]^2}\right)^{3/2}\text{Gamma}\left[\frac{\nu}{2}\right]} & \nu>3\\ \text{Indeterminate} & \text{True}\end{cases}\}, \{11, \text{NoncentralFRatioDistribution}[n, m, \lambda],$$

$$\begin{cases}\frac{m(n+\lambda)}{(-2+m)n} & m>2\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{2m^2\left((n+\lambda)^2+(-2+m)(n+2\lambda)\right)}{(-4+m)(-2+m)^2n^2} & m>4\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{2\sqrt{2}\sqrt{-4+m}\left(n(-2+m+n)(-2+m+2n)+3(-2+m+n)(-2+m+2n)\lambda+6(-2+m+n)\lambda^2+2\lambda^3\right)}{(-6+m)\left(n(-2+m+n)+2(-2+m+n)\lambda+\lambda^2\right)^{3/2}} & m>6\\ \text{Indeterminate} & \text{True}\end{cases}\},$$

$$\{12, \text{TriangularDistribution}[\{a, b\}], \frac{a+b}{2}, \frac{1}{24}(-a+b)^2, 0\}, \{13, \text{TriangularDistribution}[\{a, b\}, c], \frac{1}{3}(a+b+c),$$

$$\frac{1}{18}\left(a^2-ab+b^2-ac-bc+c^2\right), \frac{\sqrt{2}\left(2a^3-3a^2b-3ab^2+2b^3-3a^2c+12abc-3b^2c-3ac^2-3bc^2+2c^3\right)}{5\left(a^2-ab+b^2-ac-bc+c^2\right)^{3/2}}\},$$

$$\{14, \text{UniformDistribution}[\{\min, \max\}], \frac{\max+\min}{2}, \frac{1}{12}(\max-\min)^2, 0\},$$

$$\{15, \text{BetaDistribution}[\alpha, \beta], \frac{\alpha}{\alpha+\beta}, \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}, \frac{2(-\alpha+\beta)\sqrt{1+\alpha+\beta}}{\sqrt{\alpha}\sqrt{\beta}(2+\alpha+\beta)}\},$$

$$\{16, \text{ChiDistribution}[\nu], \frac{\sqrt{2}\,\text{Gamma}\left[\frac{1+\nu}{2}\right]}{\text{Gamma}\left[\frac{\nu}{2}\right]}, \nu-\frac{2\,\text{Gamma}\left[\frac{1+\nu}{2}\right]^2}{\text{Gamma}\left[\frac{\nu}{2}\right]^2}, \frac{\sqrt{2}\,\text{Gamma}\left[\frac{1+\nu}{2}\right]\left(\text{Gamma}\left[\frac{\nu}{2}\right]^2-2\nu\,\text{Gamma}\left[\frac{\nu}{2}\right]^2+4\,\text{Gamma}\left[\frac{1+\nu}{2}\right]^2\right)}{\left(\nu\,\text{Gamma}\left[\frac{\nu}{2}\right]^2-2\,\text{Gamma}\left[\frac{1+\nu}{2}\right]^2\right)^{3/2}}\},$$

$$\{17, \text{ExponentialDistribution}[\lambda], \frac{1}{\lambda}, \frac{1}{\lambda^2}, 2\}, \{18, \text{ExtremeValueDistribution}[\alpha, \beta], \alpha+\text{EulerGamma}\,\beta, \frac{\pi^2\beta^2}{6}, \frac{12\sqrt{6}\,\text{Zeta}[3]}{\pi^3}\},$$

$$\{19, \text{GammaDistribution}[\alpha, \beta], \alpha\beta, \alpha\beta^2, \frac{2}{\sqrt{\alpha}}\}, \{20, \text{GumbelDistribution}[\alpha, \beta], \alpha-\text{EulerGamma}\,\beta, \frac{\pi^2\beta^2}{6}, -\frac{12\sqrt{6}\,\text{Zeta}[3]}{\pi^3}\},$$

$$\{21, \text{InverseGammaDistribution}[\alpha, \beta], \begin{cases}\frac{\beta}{-1+\alpha} & \alpha>1\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{\beta^2}{(-2+\alpha)(-1+\alpha)^2} & \alpha>2\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{4\sqrt{-2+\alpha}}{-3+\alpha} & \alpha>3\\ \text{Indeterminate} & \text{True}\end{cases}\},$$

$$\{22, \text{LaplaceDistribution}[\mu, \beta], \mu, 2\beta^2, 0\}, \{23, \text{LogisticDistribution}[\mu, \beta], \mu, \frac{\pi^2\beta^2}{3}, 0\},$$

$$\{24, \text{MaxwellDistribution}[\sigma], 2\sqrt{\frac{2}{\pi}}\sigma, \frac{(-8+3\pi)\sigma^2}{\pi}, \frac{2\sqrt{2}(16-5\pi)}{(-8+3\pi)^{3/2}}\},$$

$$\{25, \text{ParetoDistribution}[k, \alpha], \begin{cases}\frac{k\alpha}{-1+\alpha} & \alpha>1\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{k^2\alpha}{(-2+\alpha)(-1+\alpha)^2} & \alpha>2\\ \text{Indeterminate} & \text{True}\end{cases}, \begin{cases}\frac{2\sqrt{\frac{-2+\alpha}{\alpha}}(1+\alpha)}{-3+\alpha} & \alpha>3\\ \text{Indeterminate} & \text{True}\end{cases}\},$$

$$\{26, \text{RayleighDistribution}[\sigma], \sqrt{\frac{\pi}{2}}\sigma, \left(2-\frac{\pi}{2}\right)\sigma^2, \frac{(-3+\pi)\sqrt{\frac{\pi}{2}}}{\left(2-\frac{\pi}{2}\right)^{3/2}}\}, \{27, \text{WeibullDistribution}[\alpha, \beta],$$

$$\beta\,\text{Gamma}\left[1+\frac{1}{\alpha}\right], \beta^2\left(-\text{Gamma}\left[1+\frac{1}{\alpha}\right]^2+\text{Gamma}\left[1+\frac{2}{\alpha}\right]\right), \frac{2\,\text{Gamma}\left[1+\frac{1}{\alpha}\right]^3-3\,\text{Gamma}\left[1+\frac{1}{\alpha}\right]\text{Gamma}\left[1+\frac{2}{\alpha}\right]+\text{Gamma}\left[1+\frac{3}{\alpha}\right]}{\left(-\text{Gamma}\left[1+\frac{1}{\alpha}\right]^2+\text{Gamma}\left[1+\frac{2}{\alpha}\right]\right)^{3/2}}\}\}$$

S2: The pdf of Y_i and its moments (Sec. 5; Shore, 2020a)

Suppose that Y_i has probability density function (pdf):

$$f_{Y_i}(y) = C_{Y_i} e^{-\frac{1}{1+\alpha}\left(\frac{y-\alpha}{\sigma_i}\right)^{1+\alpha}}, \; y \ge \alpha, \; \alpha > -1, \tag{S.1}$$

with C_{Yi} a normalizing coefficient, and σ_i is an internal-variation parameter. It is easy to realize that at α=1, Y_i becomes left-truncated normal (re-located half normal); It is exponential for α=0.

We denote the distribution of Y_i — the "Extended/generalized exponential distribution" (note that this is different than the definition often used for this term). As α approaches 1, internal variation is vanishing (σ_i=0), Y_i then becomes constant, assumed to be equal to the mode (α=1), and Y becomes normal (with mean of 1) or lognormal. Conversely, for α=0 we expect external disturbance to vanish (σ_e=0, Y_e=1). Therefore, Y_i and Y are both exponential.

Let us introduce:

$$Z_i = \frac{Y_i - \alpha}{\sigma_i} . \quad \text{(S.2)}$$

From (S.1), we obtain the pdf of Z_i:

$$f_{Z_i}(z) = C_{Z_i} e^{-\frac{1}{1+\alpha}(z)^{1+\alpha}}, z \geq 0, -1 < \alpha , \quad \text{(S.3)}$$

$$1 / C_{Z_i} = (1+\alpha)^{\frac{1}{1+\alpha}} \Gamma\left(\frac{\alpha+2}{\alpha+1}\right) . \quad \text{(S.4)}$$

Note that the mode of Z_i is zero (mode of Y_i is α). It is easy to show that the *m*-th non-central moment of Z_i (moment about zero) is:

$$E(Z_i^m) = \frac{\left(\frac{1}{1+\alpha}\right)^{1-\frac{m}{1+\alpha}} \Gamma\left(\frac{1+m}{1+\alpha}\right)}{\Gamma\left(\frac{2+\alpha}{1+\alpha}\right)}, \alpha > -1. \quad \text{(S.5)}$$

This may more compactly be expressed as:

$$E(Z_i^m) = \frac{(A)^{-Am} \Gamma[A(1+m)]}{\Gamma(A)}, A > 0 , \quad \text{(S.6)}$$

where $\alpha \to 1/A - 1$. It is interesting to note a unique property to the distribution of Z_i. This is the near linear relationship between its mean, μ_{zi}, and STD, σ_{zi}. Modelling the linear relationship so that it becomes exact for α=0 (exponential case) and for α=1 (the half normal case), we obtain the linear relationship:

$$\sigma_{Zi} = m\mu_{Zi} + (1-m), \tag{S.7}$$

with:

$$m = \left\{ \frac{1-\sqrt{1-\frac{2}{\pi}}}{1-\sqrt{\frac{2}{\pi}}} \right\}. \tag{S.8}$$

Figures S.1 and S.2 display the mean and STD of Z_i as function of α (Fig. S.1), and the error from modelling STD as function of the mean (via eq. S.7; Fig. S.2). We realize that the approximate merging of the mean with the STD (for any value of α, as evidenced by the approximate linear relationship) is unique for the distribution of Z_i. From (S.5) and (S.6), the mean of Y_i and the *m*-th moment around α of Y_i are, respectively:

$$E\left(Y_i\right) = \alpha + (\sigma_i)\frac{\left(\frac{1}{1+\alpha}\right)^{1-\frac{1}{1+\alpha}}\Gamma\left(\frac{1+1}{1+\alpha}\right)}{\Gamma\left(\frac{2+\alpha}{1+\alpha}\right)},$$

$$E[(Y_i-\alpha)^m] = (\sigma_i{}^m)E(Z_i^m) = (\sigma_i{}^m)\frac{\left(\frac{1}{1+\alpha}\right)^{1-\frac{m}{1+\alpha}}\Gamma\left(\frac{1+m}{1+\alpha}\right)}{\Gamma\left(\frac{2+\alpha}{1+\alpha}\right)}, \tag{S.9}$$

or, for moments around zero:

$$E\left(Y_i^m\right) = \sum_{j=0}^{m}\binom{m}{j}(\alpha)^j(\sigma_i)^{m-j}E(Z_i^{m-j}), \tag{S.10}$$

with $E(Z_i^{m-j})$ taken from (S.5).

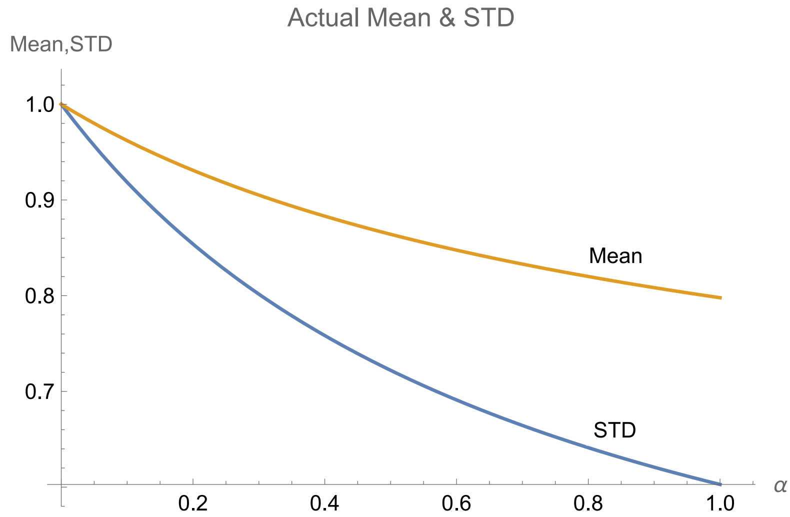

Figure S.1. Standard deviation (STD) and mean of Z_i as function of α (α=0 for the exponential; α=1 for the half-normal)

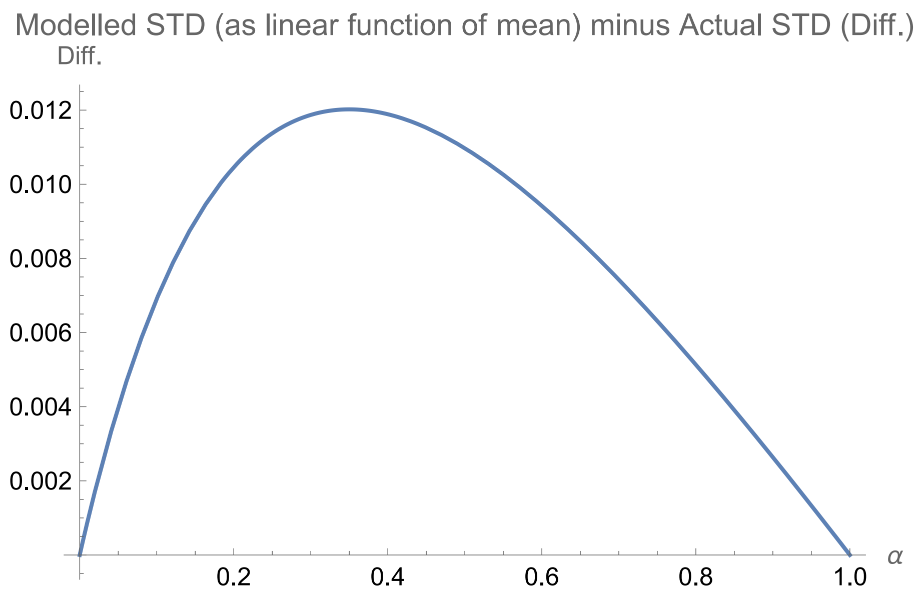

Figure S.2. Error ("Diff.") of standard deviation of Z_i (σ_{zi}), expressed as a linear transformation of the mean: $\sigma_{zi} \cong m\mu_{zi} + (1-m)$.

(Diff. shows "approximate" minus "exact" values; It equals zero for α=0, the exponential case, and for α=1, the half-normal case)