“When an error ceases to be error” — On the process of merging the mean with the standard deviation and the vanishing of the mode

Haim Shore

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, POB 653, Beer-Sheva 84105, Israel. ***shor@bgu.ac.il*****.**

January 2022

Short title: Merging of mean and vanishing of mode

Abstract: The mean and standard deviation are represented by separate parameters and a common parameter in the normal and exponential distributions, respectively. What is the process by which a transition from separate to common parameters takes place? How is the mode affected, and what is the distribution in intermediate states? In this paper, motivated by prior published research on modelling surgery duration, we denote this transition an *identity-losing process*, and introduce a new *random-identity paradigm* to address *random-identity processes*. The latter fluctuate between two extreme states — “Identity-full state” (only error generates variation; process pursues *identity-full distribution*, like normal); “Identity-less state” (signal and noise maintain a parameter-free linear relationship; process pursues *identity-less distribution*, like exponential or uniform). A general framework to address random-identity processes is developed — new terminology is introduced, a model is developed and predicted properties of distributions are articulated as conjectures, empirically corroborated with theory-based examples.

Keywords: Generalized memoryless property; General model of random variation; Process-identity instability; Random-identity paradigm; Random-identity processes. *AMS 2010 subject classifications*: Primary 60; Secondary E05.

1. Introduction

Random-identity processes are all around us. Yet, we have failed to recognize them. Consider, for example, surgeries classified into medically-specified subcategories/specialties (by any classification system like CPT, Current Procedural Terminology). While some subcategories have stable work-content, resulting in surgery duration (SD) that is normally distributed (or nearly so), others display lack of typical work-content, resulting in SD that is memoryless (exponentially distributed; Shore, 2020). Both types of surgery subcategories, and those in between, demonstrate random-identity processes, with identity (work-content within subcategory) stable to various degrees. When identity is constant (no work content variation between surgeries within subcategory), complete separation is maintained between the mean and standard deviation (STD). This results in SD normally distributed. Compare this scenario with that encountered when identity is lost altogether, namely, surgeries (within subcategory) have no characteristic work-content. An example is surgeries performed in an emergency-surgery room, serving any emergency. In that scenario, and others like it, the mean and STD merge to produce an exponentially distributed SD. Most subcategories of surgeries probably belong in a third group of surgeries, where identity (surgery work-content) is random to a certain degree. In this group, identity is not constant, neither is it lost altogether; Rather, process identity becomes random, with mean and STD that share some parameters (unlike with the normal distribution), yet STD cannot be expressed as a parameter-free linear combination of the mean (as with the exponential distribution). An example is the gamma distribution, where STD is proportional to the mean, yet with a parametric coefficient. Indeed, this three-way partition of surgery subcategories can be generalized to all work-processes, which may be classified as repetitive (stable work content), semi-repetitive (work-content somewhat varies between cycles) or non-repetitive (no characteristic work-content). In a recent paper, Shore (2020) introduced a new explanatory bi-variate model for SD that has been empirically validated, using a database of ten thousand surgery times, with surgeries classified into about 130 medically-specified subcategories. A basic assumption therein was that due to different levels of work-content instability (within subcategory), SDs for different subcategories pursue differently-shaped distributions (like the exponential, the normal or the lognormal). Furthermore, the new model was demonstrated to span a family of distributions that delivers good representation to diversely-shaped current statistical distributions, accurately preserving the first five moments of the fitted distribution. Surgery work-content (technically defined therein) was termed — surgery identity. The new model had been applied to develop a statistical process control (SPC) scheme for SD in Shore (2021a) and to estimate utilization rate of operating rooms (Shore, 2021b).

Generalizing the approach in Shore (2020), we introduce in this paper a general framework to address stochastically-stable random-identity processes. It is assumed that a certain variation-generating process, denoted herewith *the fundamental process*, underlies all random-identity processes. It is intuitively described in the next Section 2. In Section 3 we formally define "Process-identity" and related concepts — "Random-identity", partial and complete “Lack of identity” (LoI) and “Identity-full / Identity-less” distributions, linking all to a new generalized memoryless property, termed "The identity-less property". General properties of distributions, as deduced from the new paradigm — separately for identity-full distributions, identity-less distributions, and those in between — are articulated in the form of a proposition and five conjectures, described and explained in Section 4. Empirical support for the validity of the conjectures, based on theory-based results from the statistics literature and on a sample of twenty-seven distributions, is given in Section 5. A model for the fundamental process, as articulated for surgery duration in Shore (2020), is articulated, in terms of the random-identity paradigm, in Section 6. Section 7 delivers final conclusion.

2. How does the fundamental process generate variation?

For an observed random process-response, originating in a possibly random-identity process, the fundamental process fluctuates between two *extreme states* (scenarios):

- **A normal ("healthy") state**, which maintains complete separation between signal and noise factors; The response distribution then has two separate sets of parameters, one for *signal* (like the mean), another for *noise* (like the STD); The respective process is denoted an *identity-full process,* and the associated distribution an *identity-full distribution*;
- **A pathological ("ill-behaved") state**, where no distinction is possible between signal and noise process factors (a single set of factors determines both signal and noise); The response distribution then has an STD that may be expressed as a parameter-free linear combination of the mean; The respective process is denoted an *identity-less process*, and the associated distribution an *identity-less distribution*.

We are reminded that a parameter-free linear relationship between two sets of observations implies that the two sets represent measurements of same "entity", only on different scales. Table 1 summarizes the distinction between the two extreme scenarios.

Insert Table 1 about here

Examples for intermediate scenarios, positioned between the "healthy" (normal) and the "ill-behaved" (pathological), as applied to surgery times, have been described in Shore (2020). Most stochastically-stable processes, observed in nature, are probably random-identity (neither identity-full, nor identity-less). Therefore, we assume in this paper that most current uni-modal response distributions describe variation observed in responses associated with random-identity processes, namely, the distributions represent realizations of the fundamental (variation-generating) process. This leads to a bi-partition of distributions:

- **Category A**: Distributions of r.v.s (random variables) that represent direct observation of natural processes. These distributions may fluctuate between the identity-full and identity-less states, with all implications for distribution properties that this may entail (as implied by the random-identity paradigm). In particular, skewness and kurtosis values are confined to the region spanned between the extreme points of identity-full and identity-less states;
- **Category B**: Distributions of r.v.s that represent *functions* of r.v.s. These distributions may submit to the random-identity paradigm (as described by Conjectures IV and V in Section 4), but not necessarily so. Consequently, these distributions may have arbitrary values for skewness and kurtosis.

This dual partition is maintained throughout the paper, with emphasis naturally put on distributions of Category A.

Empirical support for this distinction, between two categories of distributions, has been gained in Shore (2020). In this study, a database of ten thousand surgery times, classified into medically-specified subcategories, had been statistically analyzed (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 sample-skewness to fluctuate between, say, -0.5 to 2.5 (given that the normal and exponential have values 0 and 2, respectively). A count of the number of subcategories with skewness in this interval showed that out of 126 medically-specified subcategories, 108 (86%) indeed had skewness in this interval. This is strong corroboration for the validity of the two-category partition.

In the rest of the paper, and unless otherwise specified, we assume that the extreme left point of a distribution support is zero (for a finite left support) or minus infinity (otherwise). Also, for reasons to be expounded later (see Conjecture I in Section 4), we assume that the best measure of signal (which delivers best representation to process identity) is the standardized mode.

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

Due to randomness in at least one of its identity-defining dimensions, a process becomes random-identity. Outwardly, this is observed as "Lack of Identity" (LoI) to various degrees. This process property is tightly linked to the identity-less property, as we shall soon define and demonstrate. Earlier, we loosely used the concept of LoI to describe an extreme process state (extreme LoI), when distinction between internal and external variabilities disappears, resulting in a memoryless distribution (like the exponential, henceforth defined as an identity-less distribution). In this subsection, 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 a proposition and conjectures in Section 4).

Definition of Lack of Identity (LoI, partial or complete)

A process lacks identity altogether (complete LoI) when its internal factors (identity-related factors) behave like external factors (error-related factors) so that no distinction is feasible between the two sets of factors. Formulated differently, no process-identity exists on which a *multiplicative* error can be defined (it is assumed throughout the paper that error is multiplicative; this assumption is empirically supported by statistical analysis of the database in Shore, 2020; a multiplicative error becomes additive for an identity-full process; relate to the model in Section 6). Due to the inherent lack of separation, in an identity-less process, between the two sets of factors (internal and external), a single set of common parameters determines both the signal (mean) and the noise (STD). Therefore, STD may be expressed as a parameter-free linear transformation of the mean. Conversely, a process reaches maximal identity when there is total separation (no overlap) between signal-affecting and noise-affecting factors. Furthermore, the former set of factors comprises only internal/process factors, the latter only external/non-process factors. Due to this separation, two distinct non-overlapping sets of distributional parameters determine, respectively, the response mean and STD. A process is random-identity when its LoI is partial, and some process factors participate in producing both response signal and noise. Response distributions, associated with processes having partial LoI, do have shape parameters (this will be explained shortly).

A direct result of process "identity-less-ness" is an Identity-less Distribution.

Definition of 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 we observe 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 "extended 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):

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

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

[1] 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 ($a<x_i<b$):

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

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).

[2] The exponential-distribution, with parameter λ:

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

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

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

We are aware of three identity-less distributions (per the above definition): the uniform distribution, the exponential-distribution and the geometric distribution. We cannot ascertain that these are the only identity-less distributions. Later, we define properties that identity-less/identity-full distributions own. These might serve in the future to identify response distributions belonging to the two categories. Note, that the uniform-distribution is identity-less but not memoryless (as per the traditional definition of memoryless-ness).

There are some immediate implications to the above definition, specifying important properties of identity-less/identity-full distributions. These are presented as a proposition and conjectures in the next Section 4. The conjectures are empirically validated, via theory-based examples (from the statistics literature) and a set of twenty-seven distributions, in Section 5.

4. Implications of the random-identity paradigm (a proposition and conjectures)

4.1. A proposition

For an identity-less distribution — the mode either does not exist (as per the uniform distribution), or it is equal to the extreme left-point of the distribution support (like the exponential distribution).

Proof: If this property is violated, the right tail of a truncated distribution could not possibly look the same for any truncation point (as identity-less distribution is defined). Therefore, either there is no mode, or the mode is at the extreme left point of the distribution support (for a positively-skewed distribution).

A second implication is that an identity-less distribution has a probability density function (pdf) with no inflexion point (second derivative is never zero).

We put these two properties as requirements for a model of the fundamental process in Section 6, and then show, for the developed model (originally introduced in Shore, 2020), that the component representing internal variation (identity variation), once becoming identity-less, indeed satisfies these two requirements.

4.2. Conjectures

The conjectures in this subsection are statements about general properties of distributions, as implied by the random-identity paradigm. Empirical evidence (empirical validation) for these conjectures is provided in Section 5. Five conjectures are introduced, relating to the standardized mode (Conjecture I), coefficient of variation (CV, Conjecture II), shape moments (Conjecture III), purely identity-full/identity-less functions of r.v.s (Conjecture IV) and to the effect of built-in structure in functions of r.v.s (Conjecture V). All conjectures relate to distributions that submit to the random-identity paradigm (Category A distributions, possibly B; find details in Section 1).

Conjecture I (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 the identity-full state (0), the dependence of the mode on skewness weakens until it expires (when Sk=0, the mode may assume any value). Conversely, as skewness approaches the value associated with an identity-less state (2 in the case of the exponential, 0 in the case of the uniform), the standardized mode vanishes (becoming zero, as in the case of the exponential, or non-distinct, as in the case of the uniform).

Explanation: The most prominent manifestation of identity is the mode. Therefore, the standardized mode is best representative of identity. 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. Similarly, a standardized mode progressing towards zero is reflection of a process losing identity, with skewness value progressing towards the identity-less value (2, or any other value associated with a state where the standard deviation (STD) becomes a parameter-free linear function of the mean). Conversely, when process factors become less variable, process identity is stabilized, causing identity to become more visible. This results in the standardized mode increasing with skewness approaching its error value (the identity-full value, namely, zero). Finally, note that once a process becomes identity-less, no mode can *logically* be defined (standardized mode becomes zero, as with the exponential, or non-distinct, as with the uniform).

This conjecture will be empirically validated in subsection 5.1, based on particular examples and a sample of existent distributions.

Conjecture II (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 parameters are either STD or the mean (as with the normal).

Explanation: According to the new paradigm, CV can be reduced if, and only if, internal instability is reduced, causing the distribution to approach the identity-full scenario (error variation, sourced by factors external to the process, cannot be similarly reduced). Therefore, reducing CV implies convergence to an identity-full distribution. Conjecture II also outlines expected behaviour of CV under various scenarios:

[A] "The Healthy Scenario": Identity-full distributions are symmetric with arbitrary CV values (STD is detached from the mean);

[B] "The Pathological Scenario": Signal/identity factors have rendered noise factors. Therefore, a single set of parameters determines both "signal" (mean) and noise, and STD may be expressed as a parameter-free linear combination of the mean (the two differ by scale only; CV is parameter-free);

[C] "Partial LoI": Partial-identity distributions (partial-LoI distributions) have CV that depends solely on distributional *shape* parameters (same parameters that affect standardized third-degree moment and higher).

This conjecture will be empirically validated in subsection 5.2, based on particular examples and a sample of existent distributions.

Conjecture III (shape moments): Random-identity processes (namely, processes with partial LoI, neither identity-full nor identity-less) have parametric shape moments (standardized third-degree moment and higher). Conversely, shape moments of identity-full and identity-less distributions are parameter-free. This conjecture articulates the idea that the random-identity paradigm points to certain factors that affect shape moments. These factors are described and demonstrated via examples in subsection 5.3.

Explanation: Distribution shape is unaffected by change in location or scale (standardized third-degree moment and higher are invariant to such changes). For identity-full distributions, all parameters affect either the mean or the STD (but not both). Therefore, no parameter can be envisioned that affects both shape moments and one (only) of the first two moments. This will contradict the invariance of shape moments to change in location or scale. Conversely, for identity-less distributions, STD is a parameter-free linear relationship of the mean, namely, the mean and STD differ by scale only (they represent same "entity", measured on different scales). Therefore, no parameter can be envisioned that affects concurrently shape moments and location and scale moments (this would contradict the invariance to location and scale of shape moments).

This conjecture will be empirically validated in subsection 5.3, based on particular examples and a sample of existent distributions.

Conjecture IV (purely identity-full/identity-less functions of random variables): This conjecture relates to functions of r.v.s that are all either identity-full or identity-less. The conjecture states that the distribution of the function preserves the property (be either identity-less or identity-full), provided no structure ("identity") is embedded in the function (in the form of constants, interaction effects and likewise; refer to Conjecture V). Therefore, we denote such — purely identity-full/identity-less functions.

Explanation: Process identity needs process factors that provide identity (namely, have constant value). Once a function comprises only identity-less r.v.s, no source of identity exists. Therefore, the function becomes identity-less, and it is denoted purely identity-less function. Conversely, with only identity-full r.v.s, a function is logically expected to carry this property (there is no source for "losing" identity; signal and noise parameters are separate for all sources of variation). Therefore, the function becomes identity-full, and denoted purely identity-full function.

This conjecture will be empirically validated in subsection 5.4, based on particular examples and a sample of existent distributions.

Conjecture V (structure in a function of r.v.s): Any form of embedded structure, in a function of r.v.s, generate identity that prevents the function from reaching an identity-less state. 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 that the function needs to maintain.

This conjecture will be empirically validated in subsection 5.5, based on particular examples and a sample of existent distributions.

5. Empirical validation of conjectures (expounded in subsection 4.2)

In this section we deliver empirical validation/support of 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 A. Their first three moments are given in Supplementary Material (Section S1). 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 I (subsection 5.1). Distribution notation pursues that in Mathematica™.

5.1. Conjecture I (Mode)

Example 5.1a.

To learn empirically how the mode is related to skewness (subject of Conjecture I), we select thirteen distributions (of the set of twenty-seven). These are (numbers as in Appendix A):

{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 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.

Insert Figs. 1 and 2 about here

We realize that Conjecture I is indeed empirically corroborated. As skewness approaches the identity-full value (say, 0≤Sk≤0.5), and process sets that 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 mode, a clear indication, as foretold by the random-identity paradigm, of a process losing identity.

5.2 Conjecture II (CV)

Example 5.2a. The Central Limit Theorem (CLT)

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

$$CV = \left(\frac{1}{n\mu}\right)\sqrt{\sum_{i=1}^{n}\sigma_i^2} \tag{4}$$

Since increasing n reduces CV, the distribution of the average asymptotically converges to an identity-full distribution, namely, the normal distribution (as CLT asserts; as Conjecture II asserts).

Example 5.2b: Asymptotic normality with asymptotic zero CV (as generalized by the Central Limit Theorem, 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:

$$CV_{Bin} = \frac{\sqrt{(1-p)/p}}{\sqrt{n}}\ ;\ CV_{Pois} = \frac{1}{\sqrt{\lambda}}; \tag{5}$$

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

$$CV_{NB} = \frac{1}{\sqrt{(1-p)n}}; \tag{6}$$

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

$$CV_{Gam} = \frac{1}{\sqrt{\beta}}; \tag{7}$$

- 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:

$$CV_{Beta} = [\frac{\beta}{\alpha(\alpha+\beta+1)}]^{1/2} \tag{8}$$

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}:

$$Sk = CV(3 + CV^2). \tag{9}$$

Example 5.2c: 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:

$$S_N = X_1 + X_2 + .. + X_N \ , \ N \geq 1. \tag{10}$$

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:

$$Sk = \frac{2}{\sqrt{N}} \ ; \ CV = \frac{1}{\sqrt{N}} \ . \tag{11}$$

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 the “healthy” state of normality ($Sk \rightarrow 0$ as $N \rightarrow \infty$). Conversely, as $N \rightarrow 1$, CV increases and Sk tends to the exponential value (2).

Example 5.2d: Under what constraints does a randomly-stopping sum become normal?

The CV of a randomly-stopping sum, S_N, is given by:

$$CV_{S_N}^2 = \frac{CV_X^2}{\mu_N} + CV_N^2 \ , \tag{12}$$

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 $\{n_2, \mathrm{p}\}$. Eq (12) 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:

$$K_{S_N}(t) = K_N[K_X(t)]\ , \tag{13}$$

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 5.2e: 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 sixteen distributions. The following significant linear-regression relationship between *CV* and *Sk* is obtained (corr.=0.8821; *p*=E-6; sample size=16):

$$CV = \ 0.3014 + 0.2049\left(Sk\right) \tag{14}$$

Figure 3 displays a scatter-plot of *CV* vs. skewness for this sample of theoretical distributions.

Insert Fig. 3 about here

This empirical relationship can be anticipated (and explained) only by the new paradigm. As addressed in subsection 5.1 of Shore (2020), 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 extended exponential distribution (addressed in Section 6 and Appendix B) also reproduces this relationship (Shore, 2020).

Example 5.2f: 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 II). 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 A 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 A 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 A for the identity of the distribution linked to each number). Regarding [15] (Beta distribution) — see details in Example 5.2b.

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, all derived from the random-identity paradigm).

5.3. Conjecture III (Shape moments)

Example 5.3a: An increasing non-linear relationship between STD and the mean 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:

$$\sigma = \alpha\mu^{\beta}, \beta > 1 . \tag{15}$$

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:

$$Y = X^{7} . \tag{16}$$

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

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

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

Example 5.3b: 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 paradigm. Let X be the observed r.v. and Z the standard normal variate. Then, the *inverse* BC transformation, with parameter λ, is:

$$X = [1 + \lambda(\alpha + \beta Z)]^{1/\lambda}, X \geq 0. \quad (17)$$

The author had opportunity to converse both with Box and Cox (find details 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 revokes the original healthy scenario (normal response), with signal and noise represented by separate sets of parameters, and noise represent error variation only (namely, the BC transformation concurrently stabilizes the variance by reducing internal variation). This is achieved by transforming the distribution of the original variate into a two-component mixture-distribution. One component representing distribution of a random variable, defined on a small interval of the mixture-distribution support, and another component representing a constant. In other words, a considerable separation between parameters that represent noise and signal is produced, as required for normality. A detailed numerical example for this phenomenon, the result of power transformation, is given, with respect to the exponential transformation (Manly, 1976), in Example 5.3d below.

Example 5.3c: A too small sample size produces observed response that falsely looks identity-full (normal)

This example, when an observed response *falsely* looks identity-full (normal), is provided in Shore (2020; Section 4). Therein a database of ten thousand surgeries, partitioned into medically-specified subcategories, was analyzed, 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 work-content instability (sample underestimates the true internal/identity variation). Consequently, the distribution of SD, as represented in the subcategory sample, falsely look more symmetric (normal) than it really is. In other words, this bizarre and unexpected effect is compatible with the random-identity paradigm and explained by it.

Example 5.3d: Exponential data transformed into normal data via exponential transformation.

Manly (1976) introduced the exponential data-transformation as an alternative to the Box-Cox transformation:

$$y = \begin{cases} [exp(\gamma x)-1]/\gamma, \gamma \neq 0, \\ x, \gamma = 0 \end{cases} . \qquad (18)$$

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 for the exponential). 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:

$$y = 2*\left(1-\frac{1}{exp\left(0.5x\right)}\right) . \qquad (19)$$

How has this been achieved? We learn that the large range of variation of the exponential *X*, {0,∞}, is transformed into the 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 by a constant (2) a large part of the variation of the original exponential X (say, x>7). In practice, this transformation produces a two-component mixture distribution that is part random (say, for x<7), and part constant (at about y=2, for x>7). A *signal-noise factor* (like that associated with the exponential *X*, where no distinction between signal and noise factors is feasible) is thus transformed (converted) to be more like a signal factor, producing a mixture-distribution support that is in small part random (normally distributed), and mostly constant. This is a typical example of an identity-full distribution, where there is complete separation 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 all 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 γ=0: *Y*=*X*, *X* data are already normal, and the constant component in the mixture distribution disappears altogether (namely, goes to infinity).

As earlier noted (Example 5.3b), a similar analysis may be conducted with respect to the Box-Cox power transformation. This link, between the value of γ and proximity to normality of the *X* data, is a logical outcome of the random-identity paradigm.

Example 5.3e: 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 “healthy” scenario (internal variation vanishing and error becomes sole source of variation). Estimator’s distribution therefore should asymptotically approach symmetry (as mandated by the new paradigm).

5.4. Conjecture IV (Purely identity-full/identity-less functions)

Example 5.4a: A randomly-stopping sum with only identity-less components (exponential and geometric distributions)

Consider a randomly-stopping sum, S_N, comprising N i.i.d exponential variates with parameter λ, where N is a r.v.:

$$S_N = X_1 + X_2 + .. + X_N \;,\; N \geq 1. \tag{20}$$

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 exponentially distributed). 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) in all distributions). By the new paradigm, we expect S_N also to be identity-less (exponentially distributed). Using known formulae for the mean and variance of random sums:

$$\mu_{S_N} = E(X)E(N)\ ;\ \sigma^2_{S_N} = Var(X)E(N) + [E(X)]^2 Var(N)\ , \tag{21}$$

we introduce *1/λ* for the mean and the STD of the exponential (with parameter *λ*), and *1/p* and *(1-p)/p*^{2} for the mean and variance, respectively, of the geometric (with parameter *p*). We obtain:

$$\begin{aligned} &\mu_{S_N} = E(X)E(N) = (\frac{1}{\lambda})(\frac{1}{p})\ ; \\ &\sigma^2_{S_N} = Var(X)E(N) + [E(X)]^2 Var(N) = (\frac{1}{\lambda})^2[\frac{1}{p} + \frac{(1-p)}{p^2}] = \mu^2_{S_N}\ . \end{aligned} \tag{22}$$

As predicted by the new paradigm, the mean and STD are identically equal, indicating that the random sum is exponential (identity-less). In other words, S_N is a purely identity-less function. Example 5.2d articulates conditions, under which, according to the random-identity paradigm, a randomly-stopping sum would pursue normality.

Example 5.4b: A randomly-stopping sum with only identity-less components (uniform distribution)

Consider another scenario of a randomly-stopping sum, S_N (as in Example 5.4a). 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 distribution is identity-less, as shown in subsection 3.1). Deriving the unconditional mean and variance of S_N, similarly to Example 5.4a, 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 result is strong corroboration that S_N is a purely identity-less function (uniform function), as indeed anticipated by Conjecture IV.

Example 5.4c: 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 5.4d: A weighted-sum of identity-less r.v.s and identity-less weights

Consider the following weighted sum:

$$Y = pX_1 + (1-p)X_2 \ , \tag{23}$$

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 V). 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, as shown earlier (subsection 3.1). Therefore, when it appears as a weighting distribution in a mixture distribution, it renders the weighted sum, per Conjecture IV, also identity-less (all sources of variation are identity-less). It is easy to show that the (unconditional) mean and STD of *Y* are now equal ($1/\lambda$), implying that *Y* is exponentially distributed.

Example 5.4e: A compound distribution with a random and identity-less parameter (two cases)

Case 1 (Poisson-exponential):

Consider a r.v., *Y*, having conditional Poisson distribution, given parameter α:

$$Y / \alpha \sim Poisson[\alpha]. \tag{24}$$

The Poisson obviously maintains a common parameter for signal and noise (like the exponential), though the mean and standard deviation are not linearly related. Furthermore, if α is an 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). However, the Poisson is not an identity-less distribution, as defined in Section 3. Assume that parameter α is random and exponentially distributed with parameter λ. It may be easily shown that:

$$Y \sim \text{Geometric}[1/(1+\lambda)], \tag{25}$$

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 its constituents are identity-less (in the formal sense), provided a single parameter represents both the mean and the standard deviation. Such cases need further exploration.

Case 2 (Uniform-exponential):

Consider a r.v., Y, having conditional uniform distribution with support $\{0,\alpha\}$:

$$Y / \alpha \sim Uniform[\{0,\alpha\}]. \tag{26}$$

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:

$$f(x)= \lambda \text{ gamma}(0,\lambda x),\ x>0;$$

$$\mu=1/(2\lambda); \tag{27}$$

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

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 (like all identity-less distributions). Furthermore, the mean and *STD* differ by scale only, namely, *CV* is constant. Therefore, it is deduced that this distribution is identity-less, in conformance with the random-identity paradigm.

5.5. Conjecture V (structure in a function of r.v.s prevents identity-less-ness)

Example 5.5a: A randomly-stopping product

Consider the random product, P_N, with distributions as described earlier (Example 5.4a):

$$P_N = X_1 * X_2 * .. * X_N \ , \ N \geq 1. \tag{28}$$

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 5.4a). However, it preserves structure in the form of interaction effects between variables. In other words, a change in any X_i affects the conditional relationship (given $X_i=x_i$) between P_N and the product of the rest of the components. Therefore, the response (P_N) preserves some identity and is not expected to be identity-less (exponential). Indeed, it can be easily shown that the mean and the variance of P_N are:

$$\mu_P = \frac{p\mu}{1-\mu(1-p)} \ ; \ \sigma_P^2 = \frac{p\mu^2}{[1-\mu^2(1-p)][1-2\mu^2(1-p)]}, \tag{29}$$

where $\mu = 1/\lambda$ is the mean of the exponential variate. Obviously, 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 5.4a).

Example 5.5b: Sum of N i.i.d exponential variates (N constant)

Since all sources of variation (the 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.

6. A model for the fundamental Process (Shore, 2020)

In this section, we briefly overview Shore's bi-variate statistical model, adapted here to be expressed in terms of the fundamental process. Twofold *empirical validation* of the model is given in Shore (2020) — 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 is complemented, in this paper (Section 5), by *theory-oriented* evidence. This has been 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 random-identity paradigm.

In compliance with the latter, and with the allied fundamental process (as formerly described), the new model comprises two components, representing two independent sources of variation — internal/identity variation and external/error variation. The former is produced by process-inherent factors (analogously to surgery-inherent factors, technically defined in Shore, 2020). It is represented by Y_i, an r.v. that follows a new generalized exponential-distribution (find details in Appendix B). The latter, external variation, is produced by non-inherent factors (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, 2020):

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

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 standard deviation 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 (2020), this is a plausible feature of the model because it implies that length of surgery, as represented by Y_i, affects the 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 (2020). For the convenience of the reader, we give in Appendix B details about the distribution of Y_i (termed therein the extended exponential-distribution). The final model has three shape parameters, $\{\alpha, \sigma_i, \sigma_e\}$. Parameters $\{\sigma_i, \sigma_e\}$ represent model's internal (identity) and external (error) variation, respectively, and parameter α assumes a value of α=0 (and concurrently σ_e=0) for the exponential scenario (non-repetitive/memoryless process), and a value of α=1 (and σ_i=0), for the normal scenario (a repetitive identity-full process). 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 (2020), the three shape parameters are not independent so that matching two moments is feasible. The location and scale parameters, L and S, may be identified (after the shape parameters have been estimated) via matching the means and STDs (find details in Shore, 2020).

The distribution function (cumulative density function, CDF) of Y is expounded in Shore (2020) 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 "Healthy" normal scenario (α=1). An explanation of this phenomenon is provided in an extended model described in a forthcoming paper. One particularly important property, already addressed in Shore (2020), is the linear relationship between CV and skewness of Y_i distribution (the extended exponential-distribution; refer to Appendix B). This has been earlier addressed in Example 5.2e of Section 5 (Conjecture II).

7. Conclusion

A new framework to model random variation has been introduced in this paper. According to the new paradigm, a major source for random variation is identity instability, coupled, as a secondary source, with a multiplicative error. As identity becomes more unstable, the multiplication error ceases to be error, until it expires altogether, once identity is lost. At that point, where identity is no more, a merging of the mean with the standard deviation takes place (STD is expressible as a parameter-free linear transformation of the mean), and the mode vanishes, or it becomes indistinct.

A major claim of this article is that there is an identifiable unique process that generates observed random variation, denoted here the fundamental process. The latter 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 a single-dimension space to a bi-dimensional space. The extension to the latter has allowed unification, under a single process, of current uni-modal statistical distributions (as demonstrated by delivering accurate representation to a large set of parametric distributions, preserving well all five moments of the fitted distribution; Shore, 2020). This is a new fashion of looking at random variation, and it introduces an element of uniformity into the art of modelling random variation of observed natural processes, not unlike the ideal advocated in an earlier paper (Shore, 2015), where a general model of random variation was developed.

The new paradigm has also allowed consistent explanation of seemingly unrelated "statistical results", as the five conjectures (Section 4), with their large-scale empirical corroboration (Section 5, with about twenty examples from the statistics literature) have amply demonstrated.

Further research is needed to explore the theoretical and practical ramifications of the new paradigm and of the fundamental process and its modelling (Section 6).

Appendices

Appendix A. List of 27 distributions (Section 5; 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[v,λ]; **[10]** NoncentralStudentT[v,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[v]; **[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]}.

Appendix B. The pdf of Y_i and its moments (Section 6; Shore, 2020)

Suppose that Y_i has probability density function (pdf):

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

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 exponential-distribution" (note that this is different than the definition often used for this term). As α approaches 1, we assume that 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:

$$Z_i = \frac{Y_i - \alpha}{\sigma_i} \; . \tag{B.2}$$

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

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

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

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:

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

This may more compactly be expressed as:

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

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 standard deviation, 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:

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

with:

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

Figures b.1 and b.2 display the mean and standard deviation of Z_i as function of α (Fig. b.1), and the error (Fig. b.2) from modelling STD as function of the mean (via eq. B.7). 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.

Place figs b.1 and b.2 about here

From (B.5) and (B.6), the mean of Y_i and the *m*-th moment around α of Y_i are, respectively:

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

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

or, for moments around zero:

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

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

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. 2020. "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. 2021a. "SPC Scheme to Monitor Surgery Duration." *Quality & Reliability Engineering International*. DOI: 10.1002/qre.2813.

[7] Shore, H. 2021b. “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.

Table 1. The two extreme scenarios of the fundamental process (summary).

Notation for Extreme States		**Characterization**		
Notation 1	**Notation 2**	**Separability (of internal from external factors)**	**Variation**	**Observed response distribution**
"Healthy"	"Identity-full"	Complete separation	Generated solely by external factors	"identity-full" (like Normal; Separate sets of parameters for signal and noise)
"Pathological"	"Identity-less"	No separation (no distinction possible)	No distinction between internal and external variation	"Identity-less" (exponential and others; Single set of parameters for signal and noise; Mean and STD related via parameter-free linear relationship)

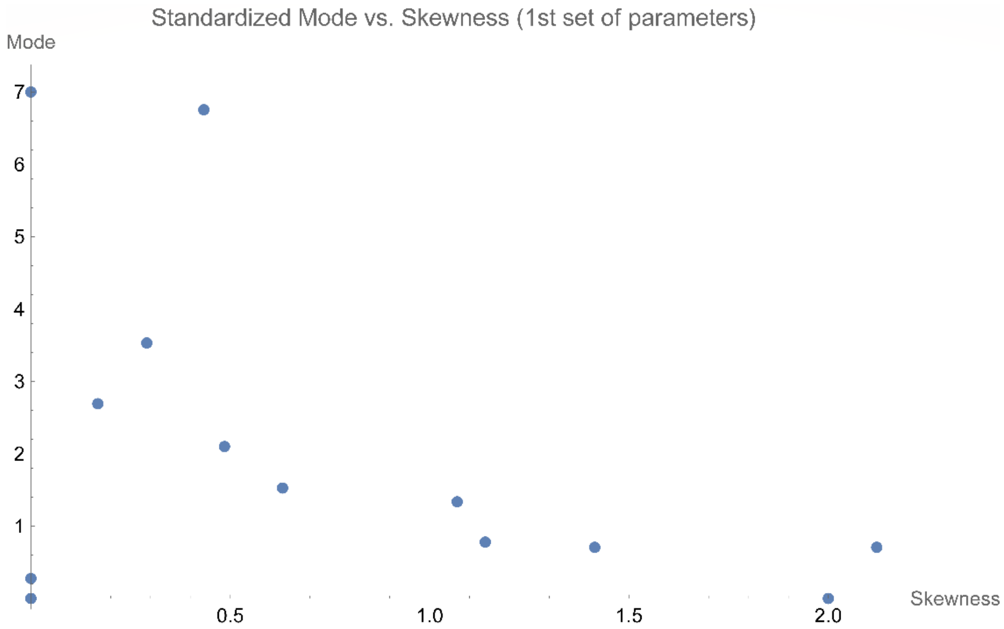

Figure 1. Standardized mode *vs.* skewness for a sample of thirteen distributions (with arbitrary parameter values). Note the increased dispersion for small Sk values (say, Sk≤0.5).

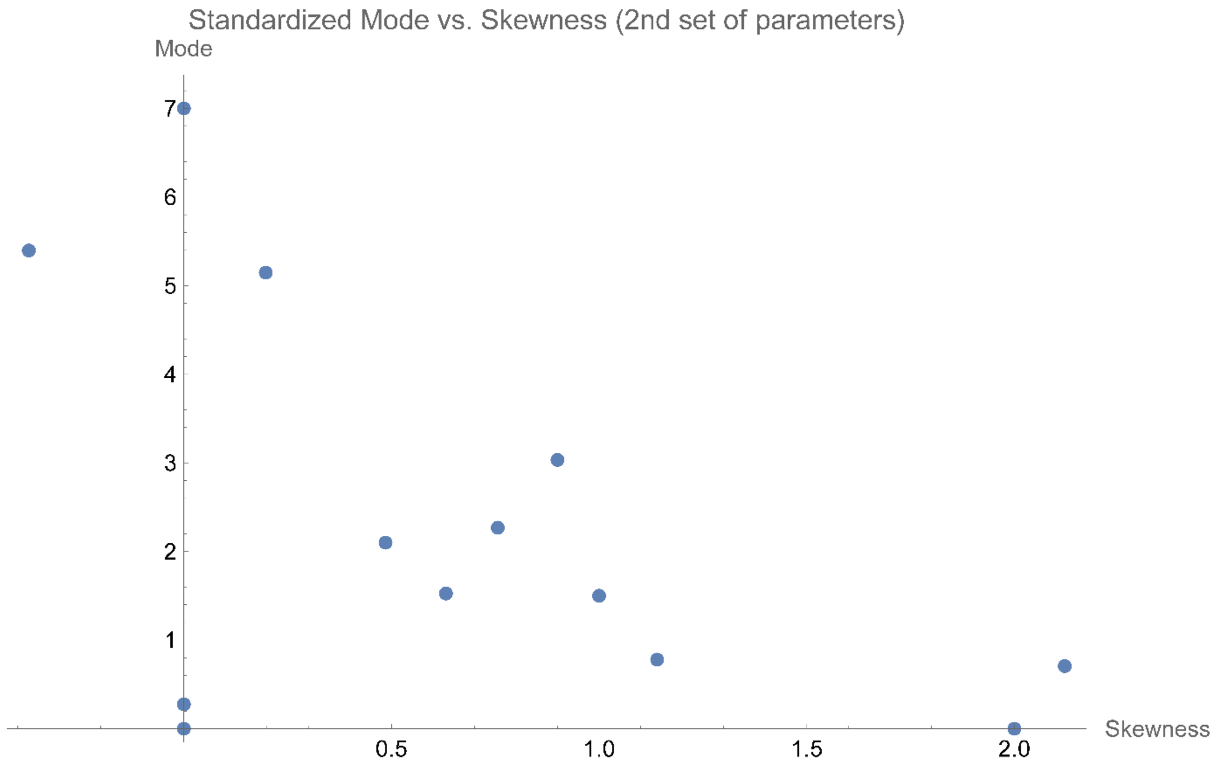

Figure 2. Standardized mode *vs.* skewness for a sample of thirteen distributions (with all parameter values of Figure 1 doubled). Note the increased dispersion for small Sk values (say, Sk≤0.5).

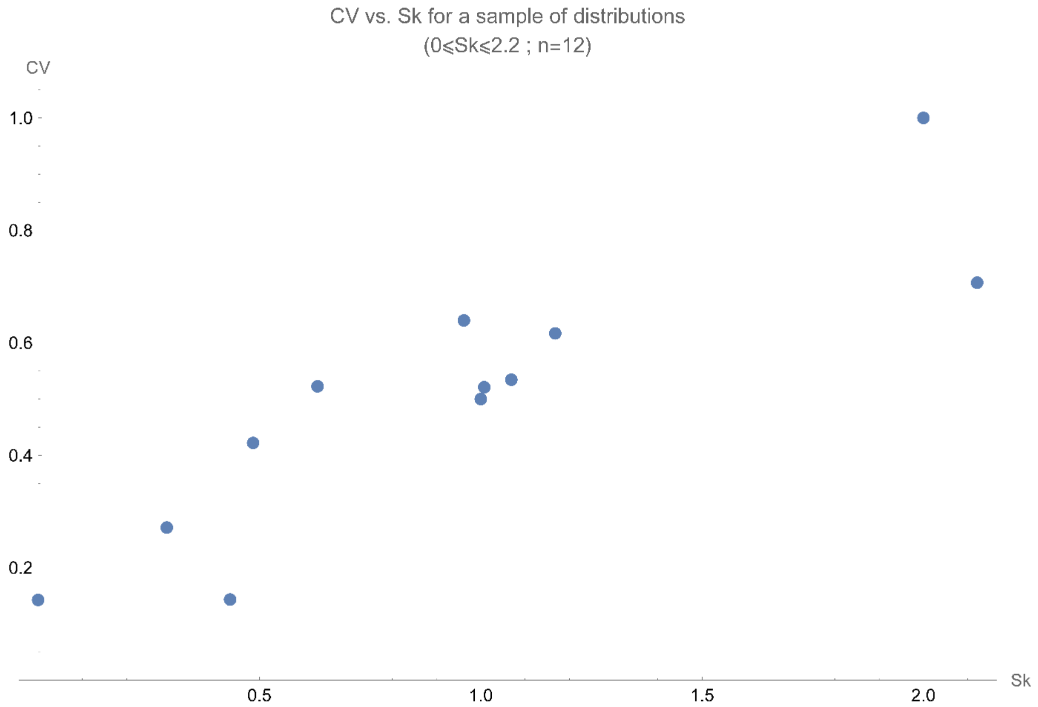

Figure 3. Coefficient of variation (*CV*) *vs.* skewness (*Sk*) for twelve different distributions with arbitrary parameter values (distributions specified in Section 5).

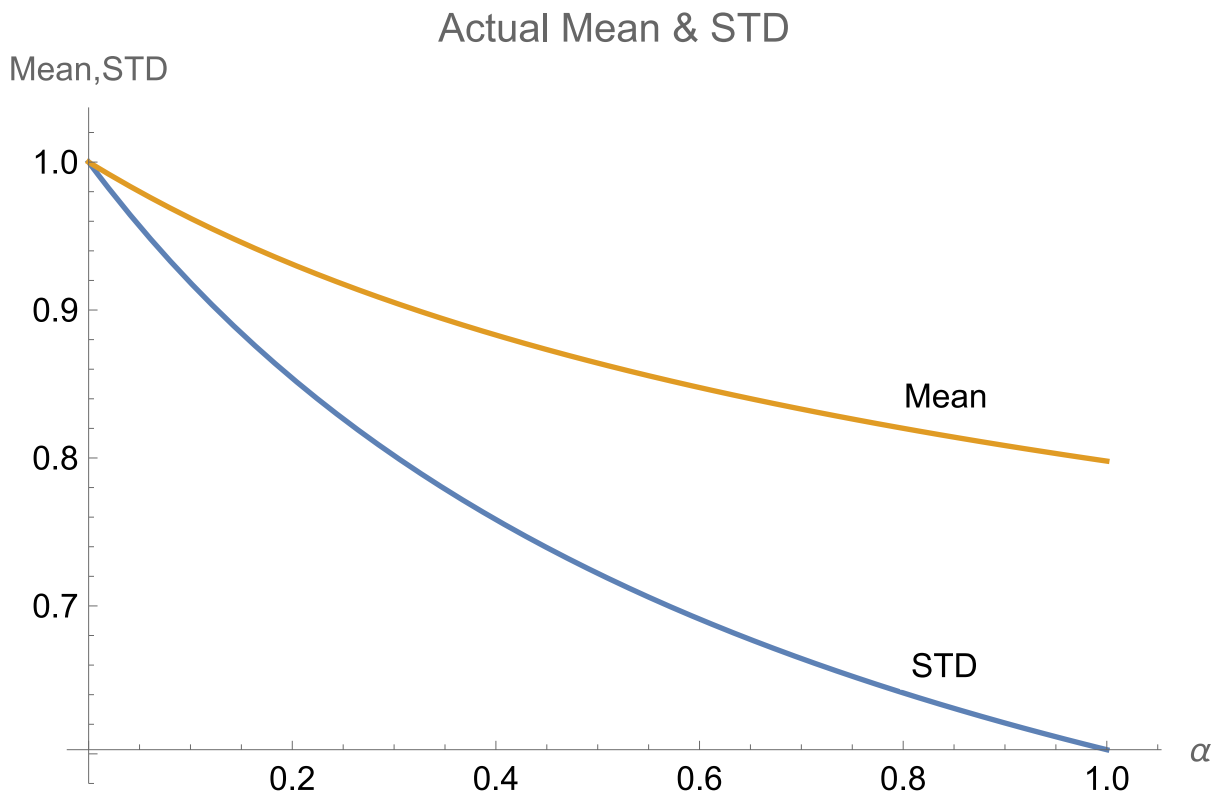

Figure b.1. Standard deviation (STD) and mean of Z_i as function of α (α=0 for the exponential; α=1 for the half-normal)

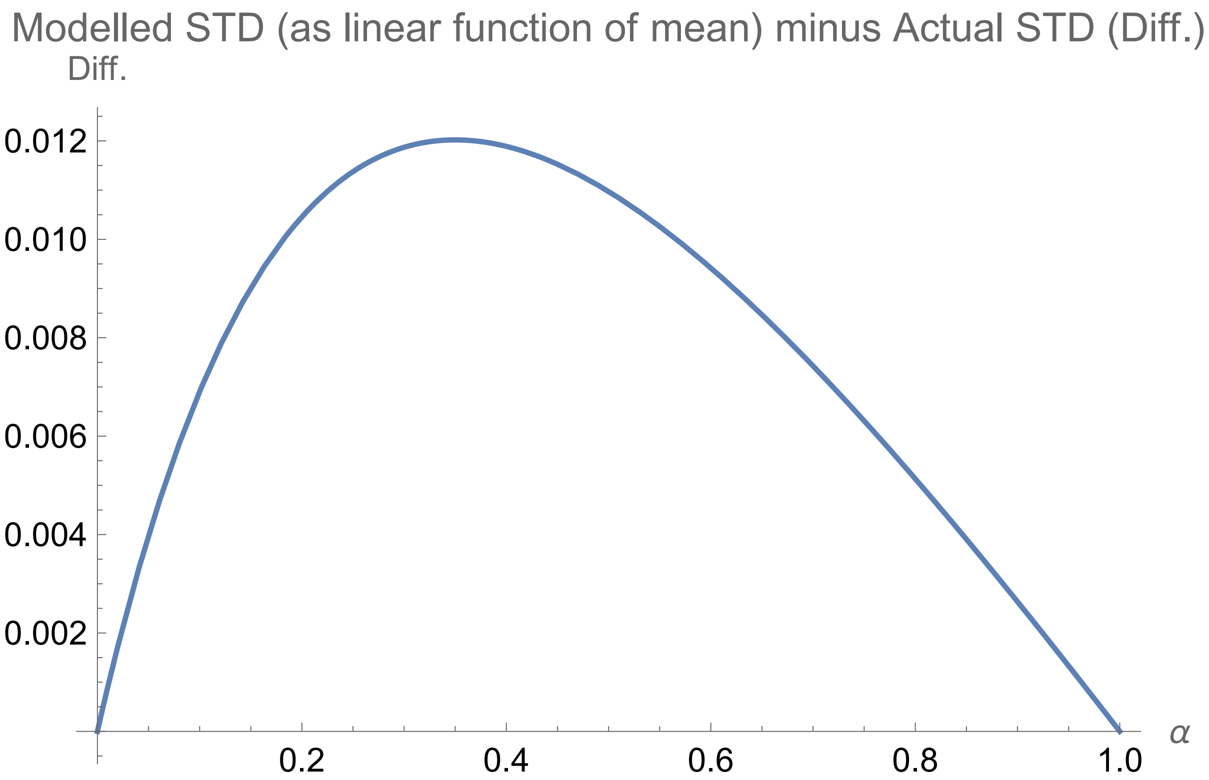

Figure b.2. Error ("Diff.") of standard deviation of Z_i (σ_{zi}), expressed as a linear transformation of the mean: $\sigma_{zi} \cong m\mu_{zi} + (1\text{-}m)$.

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

Supplementary Material

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

Out[•]=

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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