Revised: April 2010

Information, measure shifts and distribution metrics

Roger J. Bowden*

Summary

Reflexive shifting of a given distribution, using its own distribution function, can reveal information. The shifts are changes in measure such that the separation of the resulting left and right unit shifted distributions reveals the binary entropy of position, called locational or partition entropy. This can be used for spread and asymmetry functions. Alternatively, summary metrics for distributional asymmetry and spread can be based on the relative strengths of left and right hand shifts. Such metrics are applicable even for long tail densities where distributional moments may not exist.

Key Words: Asymmetry and spread measures, distribution shifting, entropy bandwidth, entropy divergence, log odds, partition entropy.

MSC Classifications: 60E05; 94A17.

* Kiwicap Research Ltd., Wellington, New Zealand, and Institute of Finance, Faculty of Mathematics and Economics, Ulm University, Germany. Emails roger.bowden@kiwicap.co.nz; roger.bowden@uni-ulm.de. Phone + 64 4 472 4984, fax 4983; post 11 Waiteata Road, Kelburn, Wellington 6012, New Zealand. Thanks go to a referee for comments that helped to focus the paper, also to seminar participants at Ulm University.

1. Introduction

Descriptive metrics or functions relating to the spread and asymmetry of a distribution are commonly based either upon moments or upon the shape or rate of decay of the density or distribution functions. The present paper develops an alternative way to characterise distributional properties, in terms of locational entropy and associated constructs. The concept of locational or partition entropy refers to binary entropy applied at any specified point to the upper and lower half sets defined with the chosen point as boundary. It is a measure of the uncertainty as to whether a number drawn at random is going to be greater or less than the given marker point. Regarded as a function of the given random variable, partition or locational entropy declines away from the median, and the rate at which it declines provides a measure of the width of the distribution, as well as its symmetry or asymmetry. In this sense the locational entropy schedule itself could be regarded as a spread function.

Partition entropy is not just an arbitrary measurement number. It can be identified as the difference between two ancillary distribution functions that can in turn be regarded as unit shifts to the left and right of the original. The shift operations can be accomplished via simple measure changes in the form of scaling functions derived from the distribution function itself, a reflexivity aspect. The wider the parent distribution's spread, the further the separation of the right and left hand unit shifts. In other words, unit distribution shifts reveal information.

A formalisation with practical applications can be devised in terms of a construct referred to as entropy divergence, which captures the differential spreading induced by the right and left hand shifts on either side of the median. Its maxima and minima on each side of the median can be used to construct metrics for distributional spread and asymmetry, applicable to any distribution, including those for which moments do not exist. The distance between these points is referred to as the entropy bandwidth. The points are also those at which the parent density intersects with the density derived from an equal weighting of the left and right hand unit shifts, called the centred shift. In turn, summary metrics based on entropy bandwidth, and its position relative to the median, can be used to characterise the asymmetry and the spread of the given distribution. The tails used for the purpose are invariant entropy-based probabilities of 16.138% and 83.862% for the lower and upper tails. If the subject distribution admits a standardisation, then the asymmetry metric is an invariant for that type of distribution, independent of its parameters.

The scheme of the paper is as follows. Section 2 establishes the binary entropy concept, called here partition entropy, such that points of equal partition entropy cut off equal tail probabilities. Core methodological aspects appear in section 3, which covers unit left and right hand shifts and convex combinations, followed by the relationship with partition entropy together with some invariance properties. Section 4 utilises the relative left and right hand shifts to develop asymmetry and spread metrics. Section 5 offers some concluding remarks covering information revelation, multivariate extensions, and other perspectives. Proofs throughout are straightforward, and are omitted for brevity.

2. Partition entropy

Entropy concepts such as differential entropy, mutual or conditional entropy are familiar tools of applied probability analysis. Classic references are Kullback & Leibler (1951), Pinsker (1964), and Kullback (1968). The objective of this section is to study the basic properties of an entropy concept that relates to a partition of the support of a random variable, so that this will be a form of binary entropy. The resulting entropy concept can be regarded as the uncertainty of position, rather than of the distribution as a whole. It will be referred to as locational or partition entropy.

A suitable starting point is a probability space $(\Omega, \Im, P)$ on which is defined a real-valued random variable $X(\omega)$, measurable with respect to $\Im$, with probability distribution function *F*. For expositional convenience, the support of *F* will be taken as the entire real line or else $\mathbb{R}_+$, but the extension to a compact subset can readily be made. Section 5 discusses extensions to the multivariate case. A specific value *x* of the random variable X will sometimes be referred to as a 'marker' or 'marker value'. It will also be convenient to assume that *F(x)* is absolutely continuous and a density *f(x)* exists, over the entire support.

The binary entropy quantity

$$h(x) = -[F(x)\ln(F(x)) + (1-F(x))\ln(1-F(x))] \qquad (1)$$

will be defined as the *partition entropy* at *x*. It can be regarded as a measure of the information about the value of the random variable *X* derived from knowing whether $X \leq x \text{ or } X > x$. In terms of standard entropy theory, it corresponds to the mutual information between variable *X* and a regime indicator variable for the partition into either *L(x)* or *R(x)*. For generic marker values, the function *h(x)* will be referred to as the partition entropy function.

Partition entropy can also be measured as the expected reduction in differential entropy that would stem from knowing whether the value of the random variable *X* belongs to *L(x)* or *R(x)*.Total differential entropy $H(f) = -E[\ln(f(x)]$ can be decomposed into the partition or locational entropy *h(x)* plus the conditional differential entropies of the truncated distributions for $X \le x, X > x$, weighted by their respective probabilities of occurring *F(x), 1-F(x)* (Di Crescendo and Longobardi (2002), Asadi et al (2004), (2005)). An analogy is with the between-group and within-group sums of squares in the analysis of variance.

A further point of contact is with the log odds function for the distribution. Given a marker value *x*, define two subsets of the support space $L(x) = \{u \le x\}, R(x) = \{u > x\}$ corresponding to the left and right hand division of the support space at boundary *x*. The logarithm of the odds that a point taken at random will lie in *L(x)* relative to *R(x)* is

$$\lambda(x) = \ln(\frac{F(x)}{1-F(x)}).$$

The complementary function $\bar{\lambda}(x) = -\lambda(x)$ refers to the odds of being in the upper zone *R(x)* relative to *L(x)*. The log-odds function is related to the density by $f(x) = \lambda'(x)F(x)(1-F(x)$, and to distribution entropy by $H(f) = -E[\ln(f(x))] = 2 - E[\ln(\lambda'(x))]$.

Basic properties of the partition entropy function, including the relationship with the log odds function, can be summarised as follows.

Proposition 1

(a) For any distribution,

(i) $\lim_{x \to \pm\infty} h(x) = 0$;

(ii) *h (x) has maximum value ln 2 at the median* $x = x_m$ *of the distribution;*

(iii) *The average value* $E[h(x)] = \int_{-\infty}^{\infty} h(x) f(x) dx = \frac{1}{2}$.

(b) For any given marker value x, partition entropy can be written in terms of the conditional expectation of the log odds function as

$$h(x) = -F(x)E[\lambda(u) \mid u \le x] = (1-F(x))E[\lambda(u) \mid u > x], \quad (2)$$

where the log odds function λ *is that of the unconditional distribution function F.*

Expression (2) implies that

$$h'(x) = -f(x)\lambda(x). \quad (3)$$

The partition entropy function will be of significance in what follows because of its connection with distributional spread. Points of any given partition entropy *h* mark off equal left and right hand tail probabilities in the density *f(x)*. More precisely, given a number $0 < p < 1$, set $h_p = -[p\ln(p) + (1-p)\ln(1-p)]$. Given the distribution function *F(x)* and the associated partition entropy function *h(x)*, the equation $h(x) = h_p$ will then have two solutions x_L, x_R such that $F(x_L) = 1 - F(x_R) = p$. Thus the distance $x_R - x_L$ is an indicator of the distribution's spread at the given probability level. The relationship with spread and symmetry is explored further in section 4.

3. Distribution shifting

Partition entropy can be established as the difference between the values of two distribution functions that can be regarded as unit shifts of the parent distribution function to the left and right of the original, respectively. The shifts are accomplished via a process that corresponds to a change of measure, accomplished via appropriately left and right oriented Radon-Nikodym derivatives. For general treatments of measure changes, see Shilov and Gurevich (1977) and Billingsley (1986). In the present context, where the primary focus is on specific unit shifts, the measure theoretic interpretations are not essential, but it is useful to develop things more generally, as other types of shifts can also be contemplated within the same framework.

3.1 Elementary left and right hand shifts

The elementary left and right hand shifts can be generated by a change of measure that can be accomplished with shift factors that are reflexive or endogenous, in the sense that they based on the given distribution function.

Lemma 1

Write

$$\xi_L(x) = -\ln F(x) \tag{4a}$$

$$\xi_R(x) = -\ln(1 - F(x)). \tag{4b}$$

Then the functions ξ_L, ξ_R *qualify as Radon-Nikodym derivatives in a change of measure from P to* Q_L, Q_R *respectively such that for any event* $B \in \Im$, $Q_L(B) = E[\mathrm{I}_B \xi_L]$, *where* $I_\bullet$ *denotes the indicator (membership) function; similarly* $Q_R(B) = E[\mathrm{I}_B \xi_R]$.

Thus ξ_L, ξ_R are nonnegative functions such that $E[\xi_L(x)] = E[\xi_R(x)] = 1$. In a reliability or mortality context, the derivatives of ξ_L, ξ_R correspond to the left and right hand hazard functions:

$$\xi_L'(x) = -\frac{f(x)}{F(x)}; \quad \xi_R'(x) = \frac{f(x)}{1 - F(x)}$$

Thus would refer to the log of the survival function at time x. Figure 1 illustrates $\xi_L(x)$, $\xi_R(x)$ for the Gumbel distribution, which is skewed to the right.

Lemma 1 and expression (4a) imply that for any measurable function $g(x)$, $E_L[g(x)] = -E[\ln(F(x))g(x)]$, similarly for regime R using (4b). One can then establish the following:

Proposition 2

The density and distribution functions for the new measures Q_L and Q_R are given by

$$f_L(x) = -f(x)\ln(F(x));\ F_L(x) = F(x)(1-\ln(F(x)). \tag{5a}$$

$$f_R(x) = -f(x)\ln(1-F(x));\ F_R(x) = F(x) + (1-F(x))\ln(1-F(x)). \tag{5b}$$

Expression (5b) is the right hand tail complement of (5a). In other words, if $\overline{F}(x) = 1 - F(x);\quad \overline{F}_R(x) = 1 - F_R(x)$, then $\overline{F}_R(x) = \overline{F}(x)(1-\ln(\overline{F}(x))$, as in (5a).

The relationship with partition entropy is given by

$$F_L(x) - F_R(x) = h(x). \tag{6}$$

Partition entropy can be obtained as the vertical difference between the left and right shifted distribution functions, marked in as the vertical double headed arrow in figure 2b. One could think of this as an uncertainty test for position. If it makes very little difference to the distribution function to move it either way by one shift, then the position must be known with some certainty. The wider the separation accomplished by the left and right shifts, the greater the inherent locational uncertainty at any given point.

The right and left hand shifted distributions can also be thought of as the result of horizontal translations of the form $y_L = g_L(x); y_R = g_R(x)$ where the function g_L is defined implicitly by $F_R(y_L) = F(y_L)(1-\ln(F(y_L)) = F(x)$; similarly for g_R. In figure 2b, the right horizontal displacement corresponding to the point x is marked in as $y_R(x)$, while the double headed horizontal arrow refers to the displacement spread $y_R(x) - y_L(x)$. In general it is not possible to solve analytically for the horizontal displacements[1].

3.2 Other kinds of distributional shift

Using the elementary right and left hand shifts, a variety of mixture distributions can be defined.

(i) Partial right (similarly left) shifts can be achieved with the Radon-Nikodym derivative $\xi_{\lambda R}(x) = (1-\lambda) + \lambda\xi_R(x); 0 < \lambda < 1$, leading to $F_{\lambda R}(x) = (1-\lambda)F(x) + \lambda F_R(x)$.

(ii) The mixtures can combine the unit right and left handed shifts. Thus

$$\xi_\theta(x) = \theta\xi_L(x) + (1-\theta)\xi_R(x); \quad 0 < \theta < 1$$

gives rise to

$$f_\theta(x) = \xi_\theta(x) f(x) = \theta f_L(x) + (1-\theta) f_R(x);$$
$$F_\theta(x) = \theta F_L(x) + (1-\theta) F_R(x).$$

The most important[2] special case is θ = ½, which will be called the 'centred' shift in what follows. In this case,

$$\xi_c(x) = \frac{1}{2}(\xi_L(x) + \xi_R(x)) = -\frac{1}{2}\ln(F(x)(1-F(x)).$$

The resulting density $f_c(x) = \frac{1}{2}(f_L(x) + f_R(x))$ will intersect the natural density *f(x)* at points where $\xi_c(x) = 1$. Figure 1 illustrates the natural and centred shift functions, figures 2(a),(b) the density and distribution functions, in each case for a Gumbel distribution. Figure 1 also depicts the centred shift function for a logistic distribution with the same median and total entropy. Relative to the symmetric logistic, the positively skewed Gumbel is displaced to the right and this is manifested in the centred shift.

(iii) The probabilities in mixture shifts can alternatively be based on the parent distribution *F(x)*. To accomplish this, the partition entropy function *h(x)* can itself act as a Radon Nikodym derivative via $\xi_h(x) = 2h(x)$. The associated density and distribution functions represent probability weighted combinations of the elementary left and right hand shifts:

$$f_h(x) = \xi_h(x) f(x) = 2[F(x) f_L(x) + (1-F(x)) f_R(x)]; \qquad (7)$$
$$F_h(x) = F(x) F_L(x) + (1-F(x)) F_R(x).$$

The mixture distribution function shifts F_c and F_h have the same median as the parent *F(x)* but $f_c(x_m) < f(x_m) < f_h(x_m)$, so they cross the parent distribution function from different directions, illustrated for the Gumbel distribution in figure 2b.

A number of invariance properties existing at the median x_m of the natural distribution, which may be summarised as follows:

(a) $F_\theta(x_m) = \frac{1}{2} + (\theta - \frac{1}{2})\ln 2$, for convex combinations;

(b) $F_L(x_m) - F_R(x_m) = h(x_m) = \ln 2$;

(c) $F_L(x_m) + F_R(x_m) = 1$.

Putting $\theta = 0.5$, property (a) implies that $F_c(x_m) = F(x_m)$, i.e. that the centred and natural distributions intersect at the natural median. Property (b) says that regardless of the distribution, the vertical difference between the right- and left- shifted distribution functions has the same value at the median, namely *ln2*. Property (c) says that at the median, the distribution functions sum to unity. Properties (b) and (c) are consistent with property (a) by setting $\theta = 1, 0$. Further relativities hold with respect to the densities. From expressions (5a,b), together with the definition of the log odds function, it follows that:

$$f_R(x) - f_L(x) = f(x)\lambda(x) .$$

In particular at the median $\lambda(x_m) = 0$, so

$$f_L(x_m) = f_R(x_m) = f(x_m)\ln 2 .$$

The left and right hand densities must intersect at the median at *ln2* = 0.6931 of the value of the natural density value, and this property must be shared by every convex combination as in expression (7b).

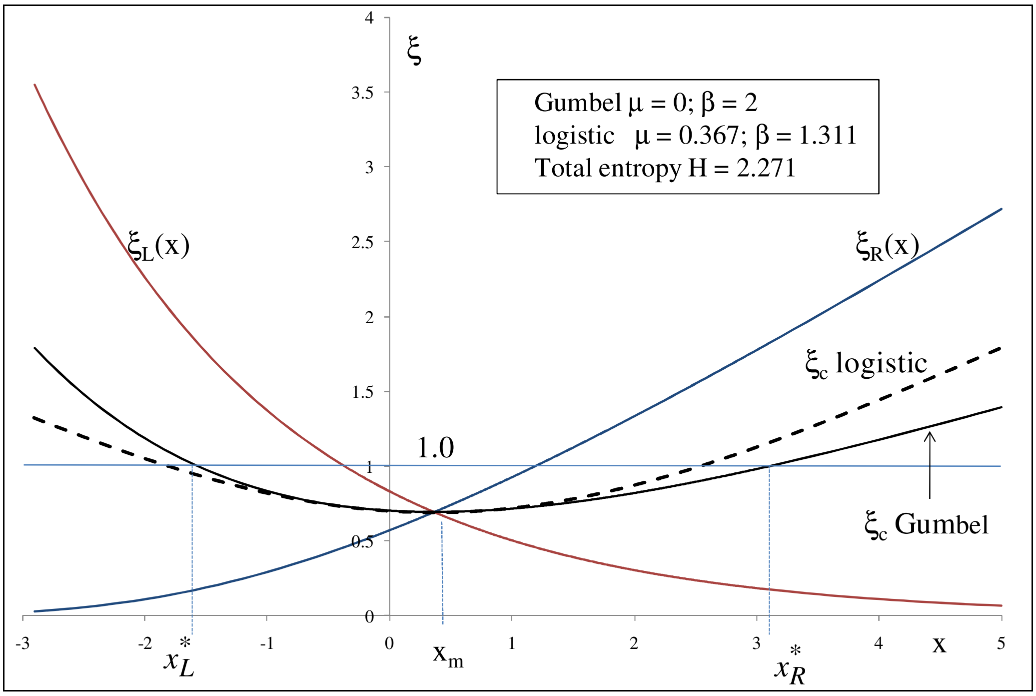

Figure 1: Shift functions: Gumbel distribution

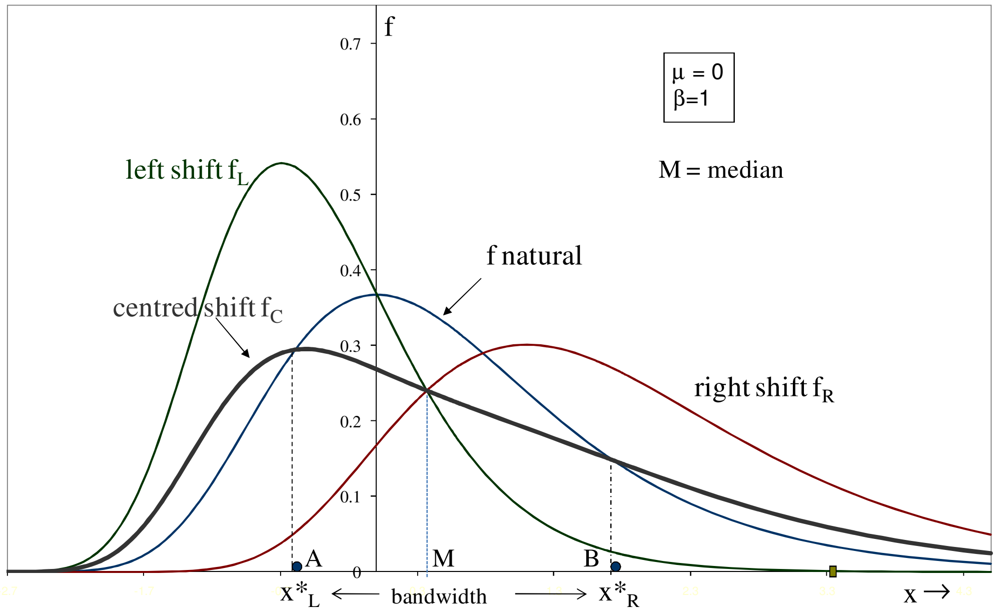

Figure 2a: Gumbel density shifting

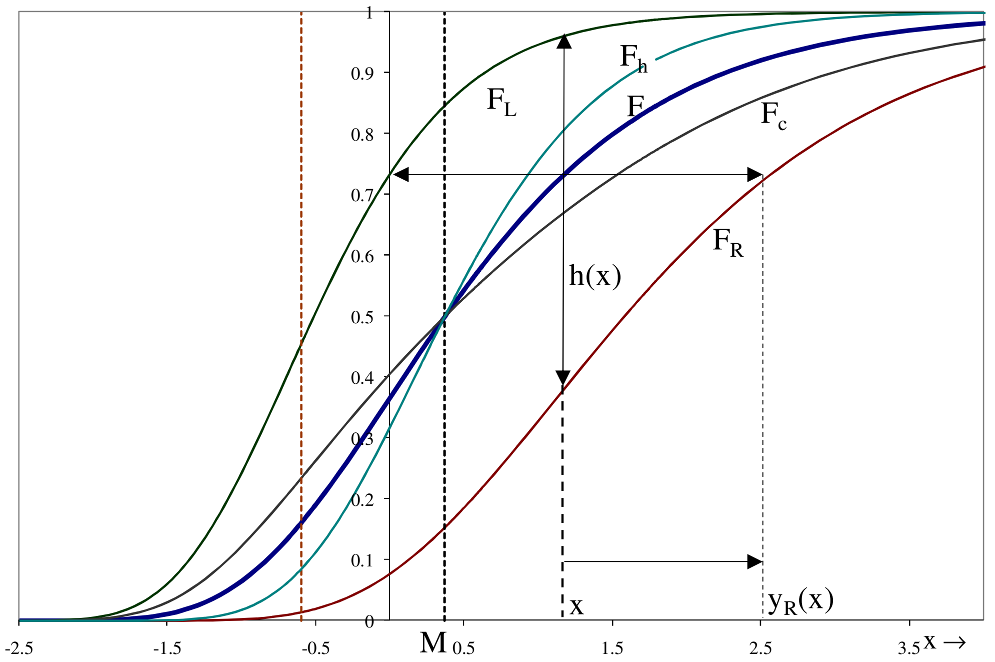

Figure 2b: Gumbel distribution functions

The above are instances of endogenous shifts, where the shift factors are based reflexively on the subject distribution function.A further extension might be to consider cross shifting, where one distribution is used to shift another, so that the shift factor is now exogenous with respect to the distribution to be shifted. Thus starting with *F(x),* one might think of shifting with $\tilde{\xi}_L(x) = -\ln(G(x;\boldsymbol{\theta}))$, where *G* is another distribution function that depends on a set of parameters $\boldsymbol{\theta}$. The new left shifted density would be $\tilde{\xi}_L(x) f(x)$. For this to be valid, one would need $E_f[-\ln(G(x;\boldsymbol{\theta})] = 1$, where the subscript means the expectation with respect to the subject distribution function *F*. This amounts to a restriction on the parameters $\boldsymbol{\theta}$. Thus suppose a normal distribution $N(x;\mu,\sigma^2)$ is to be left shifted by a Gumbel distribution function *G* that depends on a dispersion parameter β and a location parameter *m*. In order to satisfy the Radon Nikodym unit scaling requirement for the shift factor, it is necessary to have $\beta(\mu - m) = \frac{1}{2}\sigma^2$. If this is the case, then the shifted distribution turns out to be itself normal with the same variance σ^2, but with a new mean equal to $\mu - \frac{\sigma^2}{\beta}$; in other words, a simple translation to the left depending on the relative dispersion parameters. On the other hand, the precise meaning of cross shifting is less clear, in terms of partition entropy and related concepts of distributional spread.

3.3 Further properties

1. A left shift could be defined in general as a transformation or rescaling that is first order stochastic dominant (FSD) over the original distribution function, while the original is FSD over the right shift. In the present context, the distribution F_L is certainly first order stochastic dominant (FSD) over the natural distribution, while the latter is FSD over the right unit shift F_R. For, using expression (5a), $F_L(x) - F(x) = F(x)\xi_L(x) \geq 0$. Similarly from (5b), $F(x) - F_R(x) = (1 - F(x))\xi_R(x) \geq 0$, so that *F* is FSD over F_R.

Note also that that for $0 < \theta < 1$ the mixture shifted distribution function $F_\theta(x)$ has a single cross over point from above with *F(x)*, which is a standard test for second order stochastic dominance (SSD). Thus for any constant probability mixture the mixture is SSD over the parent. On the other hand, for the entropic mixture distribution defined by expression (7), the parent $F(x)$ is SSD over $F_h(x)$. For such a mixture,

$F_h(x) - F(x) = \int_{-\infty}^{x} (2h(u) - 1) f(u) du$ and $F_h(x_m) - F(x_m) = 0$. As partition entropy *h(x)* is an increasing function for $x < x_m$, and decreasing for $x < x_m$, it must be that $F_h(x) < F(x); x < x_m$ and $F_h(x) > F(x); x > x_m$. Therefore the entropic mixture distribution function $F_h(x)$ crosses *F(x)* from below at $x = x_m$, which as earlier observed, is a test for SSD.

2. The relative entropy of the right and left shifted distributions, taken with respect to the original, can be obtained as

$$E_f[\ln(\frac{f(x)}{f_L(x)})] = -E_f[\ln(\xi_L(x))]; \quad E_f[\ln(\frac{f(x)}{f_R(x)})] = -E_f[\ln(\xi_R(x))] \quad ,$$

where the expectational subscript (*f*) is inserted to remind us that the expectations have to be taken with respect to the natural (unshifted) distribution. The *ln* function is concave, so by Jensen's inequality, $-E_f[\ln(\xi_L(x))] \geq \ln(E_f(\xi_L(x))] = \ln(1) = 0$, and the relative entropies are always positive.

In the particular case of the left shift for any Gumbel (or Fisher-Tippett) distribution, relative entropy is constant, equal to the Euler-Mascheroni $\gamma \approx 0.57721$. To see this, note that for the Gumbel, $F(x) = \exp(-\exp(-\tilde{x})); \quad \tilde{x} = (x - \mu)/\beta$, where μ is the mode and β the dispersion parameter. Now $\xi_L(x) = -\ln(F(x))$ and relative entropy requires $-E[\ln \xi_L(x)]$, which reduces to $E[\tilde{x}] = \gamma$.

3. Multi-step shifting can be accomplished via the recursion

$F_n^L(x) = F_{n-1}^L(x)(1 - \ln(F_{n-1}^L(x)))\ ; n = 1,2,...$

$f_n^L(x) = -f_{n-1}^L(x)\ln(F_{n-1}^L(x))$,

with a similar recursion for the right shifts based on (5b). Sequential shifting can be a useful way to generate new distributional shapes. Thus starting from a symmetric distribution like the logistic, one can generate sequential right shifts that become more and more skewed to the right, with progressively longer right hand tails. However this is not a universal property; the right shifted normal densities continue to be symmetric, with a linear envelope. In general, non trivial stationary solutions do not exist for the above recursions.

4. Entropy bandwidth and distribution metrics

Standard approaches to problems of location, scale or symmetry take the shape of the density or distribution function as the motivation or starting point. In the older literature, the normal distribution was used as a reference base, e.g. for the textbook kurtosis and skewness ratios, alternatively interquartile or decile spreads for distributions where variances or higher order moments do not exist. Point metrics are very useful as summary comparisons of distributions from different families. On the other hand, it is possible to construct special cases by redistributing mass in the tails where an apparent symmetry (according to the point metric) becomes an undetected asymmetry, so that a process of eyeballing the entire distribution is always advisable. More recent treatments, e.g. Doksum (1975), MacGillivray (1986), Benjamini and Krieger (1996), Averous et al (1996), seek to replace spread or asymmetry point or interval style metrics with functions, utilising constructive decompositions[3] of the distribution function with probability values (p) as the underlying argument. For example, a spread function is defined in terms of the difference $F^{-1}(p) - F^{-1}(1-p);\ p > 0.5$.

The present paper takes an alternative point of departure: spread or asymmetry are taken to refer to the profile of uncertainty along the support axis, with uncertainty interpreted as locational entropy (h). In this sense, locational entropy becomes its own spread function. Thus if two distributions A and B have the same median, but the lower and upper solutions $x_L(h_0), x_U(h_0)$ to $h^{-1}(h_0) = x$ are further apart for all numbers $0 < h_0 < \ln 2$, then one could say that distribution A has a uniformly wider spread than does B. This will certainly be true for changes to the scale parameter where the distribution admits a standardisation $\tilde{x} = (x - \mu)/\beta$ in terms of scale and location parameters β, μ, such that $F(x;\mu,\beta) = F(\tilde{x};0,1)$. If the scale parameter β increases, with the location parameter μ adjusted to maintain the same median x_m, then the horizontal translation at any given value h_0 for locational entropy is given by $\nabla x = (\nabla\beta/\beta)(x_0 - x_m)$, i.e. simple proportionality. Figure 3b illustrates for the Gumbel distribution (curves h_A, h_B). Similarly, an asymmetry function could be defined in terms of the relative spread each side of the median, for alternative values of h. Spread or asymmetry functions so constructed are formally equivalent to those based on p values, though there are some visual advantage in using the h function to detect asymmetry (e.g. $h_A(x)$ versus $F_A(x)$ in figure 3b).

4.1 Entropy based point metrics

The search for an entropic point metric, i.e. a summary scalar measure, would rest upon finding a particular entropy value h^* of special significance, at which both spread and asymmetry measures could be defined together in terms of the associated $x_L(h^*), x_U(h^*)$. To serve such a benchmark purpose, the designated entropy number h^* should be invariant over all possible distributions.

In what follows, a suggested benchmark is derived by examining the relative strength of the two shift factors $\xi_L(x), \xi_R(x)$ to the left and right of the median. Locational entropy can be written (expression (6) section 3.1) as the difference between the right and left shifted distribution functions, which in turn are generated by the shift functions $\xi_L(x), \xi_R(x)$. Spread or asymmetry in the underlying distribution function is reflected in the relative behaviour of the two shift functions (e.g. figure 1). The latter are unbounded, but the locational entropy formula weights each by their respective regime probabilities. Thus $h(x) = \zeta_L(x) + \zeta_R(x)$, where the two constituent factors $\zeta_L(x) = F(x)\xi_L(x)$ and $\zeta_R(x) = (1 - F(x))\xi_R(x)$ originate in the left and right shift functions, weighted by their respective regime probabilities. Figure 3a illustrates. The difference $\tilde{h}(x) = \zeta_L(x) - \zeta_R(x)$ between the two components represents the relative power of the left and right hand shift contributions to the locational entropy at any point, and will be referred to as entropy divergence.Written in the form $\tilde{h}(x) = \ln[(1 - F(x))^{1-F(x)} / F(x)^{F(x)}]$, the entropy divergence is analogous to the log odds function $\bar{\lambda}(x) = \ln[(1 - F(x)) / F(x)]$. Unlike the latter, however, it is bounded and $\lim \tilde{h}(x) = 0$ as $x \to \pm\infty$. Its expected value is zero: $E_f[\tilde{h}(x)] = 0$.

The absolute divergence $\left|\tilde{h}(x)\right|$ is maximal at points where the probability weighted left and right hand shifts are maximal relative to one another. There are just two such points x_L^*, x_R^* on either side of the median, with the same value for the absolute entropy divergence.

The points x_L^*, x_R^* can be located as those where the average or centred shift is unity:

$$\xi_c(x^*) = 0.5(\xi_L(x^*) + \xi_R(x^*)) = 1;\ x^* = x_L^*, x_R^*.$$

Figure 1 illustrates. Correspondingly, the absolute entropy divergence can be expressed in terms of the distance between the original and centred distribution function:

$$\left|\tilde{h}(x)\right| = 2\left|F(x) - F_c(x)\right|.$$

It has maxima of equal height to either side of the median at the points x_L^*, x_R^*. The vertical distance between F and F_c is maximised at these points, and the natural and centred densities $f(x), f_c(x)$ intersect (see figure 2a).

Proposition 3 summarises the invariance relationships and numbers that result.

Proposition 3

(a) The points x_L^*, x_R^* of maximum absolute entropy divergence are the two solutions to

$$\xi_c(x^*) = -0.5\ln[F(x^*)(1 - F(x^*))] = 1,$$

given by $F(x^*) = 0.5(1 \pm \sqrt{1 - 4e^{-2}})$ *with* $x_L^* \approx F^{-1}(0.161378)$, $x_R^* \approx F^{-1}(0.838622)$.

(b) The following properties hold:

(i) $F(x_L^*) = 1 - F(x_R^*)$; $F_L(x_R^*) = 1 - F_R(x_L^*)$; $F_L(x_R^*) = 1 - F_R(x_L^*)$

(ii) $h(x_L^*) = h(x_R^*) \approx 0.441948$; $\left|\tilde{h}(x_L^*)\right| = \left|\tilde{h}(x_R^*)\right| \approx 0.146761$.

The two values x_L^*, x_R^* correspond to invariant probabilities of 16.138% for the left and right hand distribution tails. They apply to any distribution. It is of interest to note that they are quite close to the one-sigma tail of the standard normal distribution, which is 15.866%. Figure 3b illustrates the effect of increasing spread (here, higher β), which pushes the right and left hand shifts further apart. The new points x_L^*, x_R^* can be read off from the maxima of the absolute entropy divergence, which shift outwards (dotted arrow).

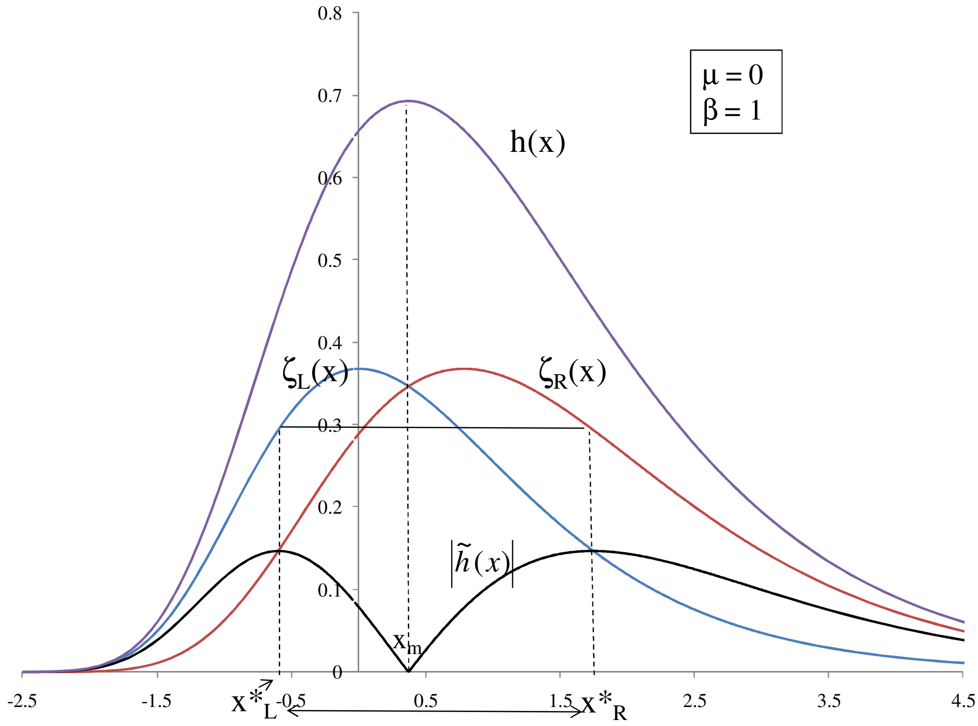

Figure 3a: Absolute entropy divergence function: Gumbel distribution

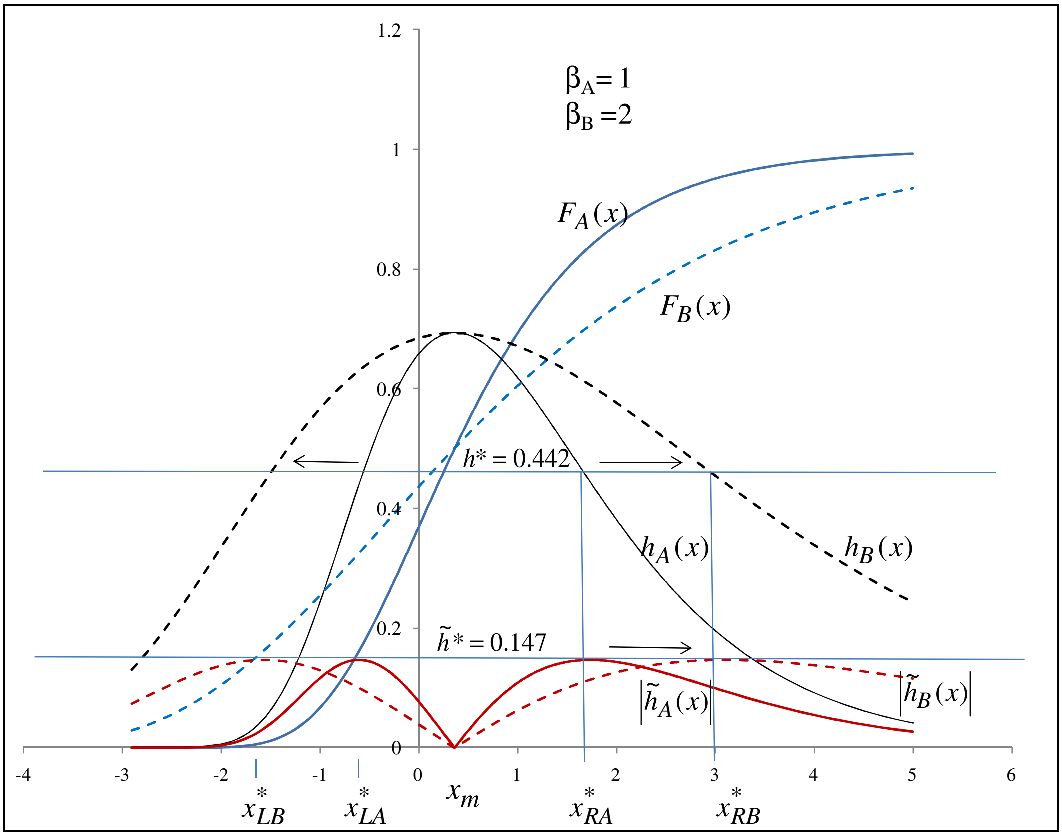

Figure 3b: Entropy divergence and spread: Gumbel distribution.

4.2 Entropy bandwidth and its uses

The notions of entropy bandwidth and associated symmetry and spread metrics flow more or less immediately from the above invariance properties. From Proposition 3, the two points x^*_L, x^*_R mark off equal tail areas of 16.138% underneath the natural density *f(x)*. Partition entropies must then be the same at these points, as must entropy divergence. The distance $\Delta = x^*_R - x^*_L$ will be referred to as the 'entropy bandwidth' of the given distribution. To understand the effect of the entropy bandwidth, imagine in figure 2a some perturbation of the original distribution function F that lengthened the right hand tail at point $B = x^*_R$, while leaving the left hand tail probability of point $A = x^*_L$ unchanged. This amounts to an increase in partition entropy at B relative to A. To preserve equality, point B must move to the right. An increased entropy bandwidth signals the higher spread.

A useful bandwidth property arises where the distribution admits a standardisation in terms of location and scale parameters, as $F(x;\mu,\beta) = F(\tilde{x};0,1)$. The entropy bandwidth is then also standardised on the difference $\tilde{\Delta} = \tilde{F}^{-1}(0.838622) - \tilde{F}^{-1}(0.161378)$. The bandwidth in terms of the unstandardised x will be $\Delta = \beta\tilde{\Delta}$, and hence proportional to the scale parameter β. In terms of a situation such as that depicted in figure (3b), the bandwidth will expand in proportion. For the normal distribution, the entropy bandwidth is 1.9776σ, while for the logistic with dispersion parameter β it is 1.6759β = 0.9240σ. This is consistent with the classic result that the normal is the distribution that maximises differential entropy for a given variance, provided the latter exists. However, the entropy bandwidth can be used for metrics of spread and asymmetry that do not depend upon the normal distribution as benchmark, or of moment existence. The Levy distribution is a case in point (see below).

The entropy bandwidth can also be used to establish a point measure for asymmetry. In this context, distributional symmetry can be cast as referring to whether partition entropy is equally concentrated on either side of the median, for which entropy divergence is a relevant indicator. Asymmetry is measured at the maximum points of probability weighted divergence on each side of the median, which correspond to $\overset{*}{x}_L, \overset{*}{x}_R$. The centring of the bandwidth interval relative to the median can therefore be taken as an asymmetry metric:

$$\nu = \frac{\overset{*}{x}_R - x_m}{\overset{*}{x}_R - \overset{*}{x}_L} - \frac{1}{2}. \tag{8}$$

Positive (negative) values would indicate positive (negative) skewness. The limits are ½, - ½. Thus for the Gumbel distribution of figure (2a), the asymmetry metric would be taken as $\frac{MB}{AB} - \frac{1}{2}$, with a positive value indicating that the distribution is skewed toward the right.

If the distribution admits a standardisation, then the asymmetry metric ν, as defined by expression (8), is an invariant for the distribution type; it is in effect that for the standardised version. For a standard Gumbel with mode $\mu = 0$ and $\beta = 1$, the median is $\mu - \beta \ln(\ln(2) =$ 0.3665, and the entropy bandwidth is the interval (-0.5994, 1.7325), giving an asymmetry metric ν = 0.0858, so the distribution is mildly skewed to the right.

By way of contrast, the Levy distribution provides an instructive case study for asymmetry and spread metrics. The Levy density is defined by $f(x;c) = \sqrt{\frac{c}{2\pi}} \frac{e^{-c/2x}}{x^{1.5}}$ with cumulative $F(x;c) = erfc(\sqrt{0.5c/x})$, where c is a scale parameter. It is standardised via $\tilde{x} = x/c$. The usual skewness and excess kurtosis point metrics are undefined, as the relevant moments do not exist. However, the entropy bandwith is 23.6004 and the asymmetry metric ν = 0.4285, quite close to the theoretical upper limit of 0.5. If the scale constant c is increased to

4, then the bandwidth enlarges to 94.4012, so the distribution has four times the spread. However, the asymmetry metric stays just the same at 0.4285, the invariance arising from the standardisation with respect to the scale parameter.

5. Concluding remarks

Distributions with long tails are of interest in a number of contexts: reliability theory, investments and option pricing, income or wealth distributions, mortality, to name a few. Some of these distributions may not possess the full range of moments; others may have irregular or lumpy bits in their tails, perhaps reflecting mixture elements. In this context, one objective of the present paper has been to put forward an approach to summary distribution diagnostics that calls upon information theory rather than on moments. Some further potential applications may be summarised as follows.

1. Distributional shifting via the shift factors ξ_L, ξ_R can be of independent interest as a way of revealing information. Thus mixture densities have been suggested as a systematic way to model asymmetry; see McLachlan and Peel (2000). But if the modes of the two distributions being used to form the mixture are obscured, e.g. because of higher variance of one of the constituent distributions, the existence of an underlying mixture element can be easily obscured or hidden. However, applying one or other of the shift operators ξ_L, ξ_R to the natural distribution can make modes more apparent.

2. It is possible to extend the entropy approach itself. Thus Bowden (2010) utilises a more directional approach that assigns conditional entropy values to distribution tails, useful in particular contexts such as options pricing or value at risk in financial risk management. The approach does call on distributional benchmarking, in this case employing an alternative measure change that establishes a local logistic-equivalent distribution. In this sense it is more limiting than the present paper, which does not require any particular benchmark.

3. A corresponding theory for multivariate distributions (as $F(\boldsymbol{x})$ with $\boldsymbol{x}$ a vector valued random variable) would require attention to the definition of direction, i.e. what is to constitute right and left. The directional derivative provides a possible framework. Given a marker point **x**, define a vector $\boldsymbol{x} + z\boldsymbol{h}$, along a suitably normalised direction $\boldsymbol{h}$, where z is a scalar. The entire mass of the distribution is then projected along the direction line, i.e. as the marginal distribution, to obtain a univariate density $f_h(z;\boldsymbol{x})$, which has z as a scalar random variable. A natural direction to take might be that of maximal increase in the parent distribution function F at the given point $\boldsymbol{x}$, in which case $h \propto \frac{\partial F}{\partial \boldsymbol{x}}$, i.e. the steepest ascent

direction. Once the direction is decided, the univariate framework applies as in the foregoing discussion, with the understanding that partition entropy and distribution shifting are specific to the designated direction.

4. The basic task for data analysis, so far as spread and asymmetry metrics are concerned, is to estimate the two inverse probability points corresponding to the 16.13% lower and 83.86% upper entropy bandwidth limits. A variety of approaches would appear to be available. In non parametric terms one could estimate either the parent distribution or else the entropy divergence function $h(x)$, concentrating the window on the neighbourhood of the likely maxima and minima. The revealed local curvature in the latter case can be taken as an indication of likely estimation precision. In general, however, estimation and inference remain as topics for further investigation.

References

M. Asardi, N. Ebrahimi and E.S. Soofi, *Dynamic generalized information measures.* Stat. and Prob. Letts. 71(2005), pp 85-98.

M. Asardi, N. Ebrahimi, G.G. Hamedani and E.S. Soofi, *Information measures for Pareto distributions and order statistics*, in N. Balakrishnan, E. Castillo, and J.M. Sarabia (eds) *Advances in distribution theory, order statistics, and inference* (In Honor of Barry Arnold), Birkhäuser, Boston, (2006).

J. Averous, A.-L. Fougères and M. Meste, *Tailweight with respect to the mode for unimodal distributions*, Stats and Prob. Letters, 28 (1996), 367-373.

K.P. Balanda and H.L. MacGillivray, *Kurtosis and spread*, Canadian Jour. Stats 18 (1990), pp 17-30.

Y. Benjamini and A.M. Krieger, *Concepts and measures for skewness with data analytic interpretations,* Canadian Jour. Stats 24(1996), pp 131-140.

P. Billingsley, *Probability and Measure* (2^{nd} ed.), Wiley, New York: (1986).

R.J. Bowden, *Directional entropy and tail uncertainty, with applications to financial hazard and investments*, Quantitative Finance, forthcoming (2010).

A. Di Crescendo and M. Longobardi, *Entropy based measure of uncertainty in past lifetime distributions*, Jour Applied Prob. 39(2002), pp 434-440.

K.A. Doksum, *Measures of location and asymmetry*, Scandinavian Jour. Stats 2, 11-22.

N. Ebrahimi, S.N.U.A. Kirmani and E.S. Soofi, *Dynamic multivariate information*, Jour. Multivariate Anal. 98 (2007), pp 328-349.

R.A. Groeneveld and G. Meeden, *Measuring skewness and kurtosis*, The Statistician 33 (1984), pp 391-399.

H.L. MacGillivray, *Skewness and asymmetry: Measures and orderings*, Annals of Statistics 44 (1986), 994-1011.

G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, (2000).

S. Kullback, *Information Theory and Statistics*, Dover, New York, (1968).

S. Kullback and R. A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics 22 (1951), pp 79-86.

M.C.Jones, *Families of distributions arising from distributions of order statistics*, Test: Journ. Spanish Stat. Soc., 13 (2004), pp1-43.

M.J. Lighthill, *An introduction to Fourier analysis and generalised functions*. New York, Academic Press, (1959).

M.S. Pinsker, *Information and information stability of random variables and processes,* Holden Day, San Francisco, (1964).

G. E.Shilov and B. L.Gurevich, *Integral, measure, and derivative: A unified approach*, transl. R. A. Silverman, Dover Publications (1977).

---

[1] The first order Taylor series approximations are given by $y_L \approx x - \frac{F(x)}{f(x)}; y_R \approx x + \frac{1-F(x)}{f(x)}$. These can be a guide over some ranges of *x*, but perform poorly for the tails of the distribution.

[2] In addition to its use in measures of spread and asymmetry (section 4), the centred shift is related directly to partition entropy by the censored expectations
$h(x) = 2F(x)(E[\xi_c(X) \mid X \leq x] - 1) = 2(1 - F(x))(1 - E[\xi_c(X) \mid X > x])$.

[3] Thus the decomposition
$F^{-1}(p) = x_m + 0.5[F^{-1}(p) - F^{-1}(1-p)] + 0.5(F^{-1}(p) + F^{-1}(1-p) - 2x_m); x_m = F^{-1}(0.5)$
motivates the spread function as the second right hand term for probability p > 0.5 , and the asymmetry function as [$F^{-1}(1-p) - x_m]/[x_m - F^{-1}(p)]$.