May 2007

Information, measure shifts and distribution diagnostics

Roger J. Bowden*

Summary

Reflexive shifting of a given distribution, using its own distribution function, can reveal information. The shifts are changes in measure accomplished via simple Radon-Nikodym derivatives. The separation of the resulting left and right unit shifted distributions reveals the binary entropy of location. Recombining the shifts via convex combinations yields information-based asymmetry and spread metrics as well as insight into latent structure. The resulting diagnostics are applicable even for long tail densities where distributional moments may not exist, and independently also of benchmarks such as the Gaussian.

Key Words: Asymmetry, entropy bandwidth, locational entropy, distribution spread, reflexive shifting, unit distribution shifts.

MSC Classifications: 60E05; 94A17.

*School of Economics and Finance, Victoria University of Wellington, and Kiwicap Research Ltd. Emails roger.bowden@vuw.ac.nz; roger.bowden@kiwicap.co.nz . Phone + 64 4 472 4984, fax 4983, postal address School of Economics and Finance, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand.

1. Introduction

The central idea developed in this paper is that shifting a given distribution in a reflexive manner, by using its own distribution function, can reveal information. The resulting operation can be applied in such contexts as latent mixtures to reveal more of the underlying structure. The shift framework can also be useful in originating metrics relating to spread and asymmetry of a distribution. The latter generally assume a benchmark such as the normal distribution, or that symmetry is just a matter of odd-order moments. But in very long tail densities such as the Levy distribution, excess kurtosis and asymmetry metrics based on moments do not even exist. The present paper utilises entropy as the basis for a more systematic approach. The entropy concept employed for the purpose is binary entropy applied at any specified point to the upper and lower half sets defined with the 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. The locational entropy so defined 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.

Locational entropy is not just an arbitrary measurement number. It can be identified as the difference between two 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 Radon-Nikodym derivatives, which in this context are scaling functions derived from the distribution function itself, the reflexivity aspect. The wider the parent distribution's spread, the more the separation of the right and left hand unit shifts. In other words, unit distribution shifts reveal information.

Symmetry and spread aspects can be explored using convex combinations of the unit shifts, most usefully an equal weighting of the left and right hand unit shifts, which can be called the central shift. The points at which the centrally shifted density intersects the parent density define an entropy bandwidth. In turn, 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 in a way that is independent of any assumed benchmark such as the Gaussian, or of the requirement that relevant moments exist. Construct validity is helped by the existence of some useful invariance properties. If the subject distribution admits a standardisation, then the asymmetry metric is an invariant for that type of distribution.

The scheme of the paper is as follows. Section 2 reviews some entropy concepts, and their relationship with the log odds function. The core results of the paper are contained in section 3, which covers unit left and right hand shifts and convex combinations, followed by the relationship with locational entropy together with some invariance properties. The case of the centred shift, which is an equally weighted combination of unit right and left hand shifts, is explored in section 4. This is applied to develop asymmetry and spread metrics, with a further discussion of how unit shifts can reveal information structure. Section 5 offers some concluding remarks, canvassing multivariate extensions, as well as cross shifting by means of an ancillary distribution.

2. Locational entropy

The object of analysis in this section 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. It will also be convenient to assume that *F(x)* is absolutely continuous and a density *f(x)* exists, over the entire support. Sometimes it will be necessary to make distinction between a specific value *X* of the random variable (i.e. $x = X$) and the generic outcome *x*; the specific *X* will then be referred to as a 'marker' or 'marker value'.

Let *X* be such a value and define two subsets $L(X) = \{x \le X\}, R(X) = \{x > X\}$ corresponding to the left and right hand division of the support space at boundary *X*. The log-odds function will be defined as

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

This is the logarithm of the odds that a point *x* taken at random will lie in *L(X)* relative to *R(X)*. For a generic point *x*, the log-odds is related to the density by

$$f(x) = \lambda'(x)F(x)(1-F(x). \qquad (2)$$

A familiar information concept for continuous distributions is differential entropy, defined as

$$\kappa = E[\ln(f(x)]. \qquad (3)$$

This is a useful measure of the uncertainty in the distribution as a whole, although unlike Shannon entropy for discrete distributions, it can have negative values. It often appears in the form of relative entropy or Kullback-Leibler entropy, which takes the expectation, with respect to a base distribution, of the log ratio of an alternative density to the base density. Additional entropy metrics, such as mutual information and conditional entropy, will play a tangential role in what follows. Standard references are Kullback and Leibler (1951), Kullback (1968), Pinsker (1964).

Given a marker value X, the binary entropy quantity

$$\kappa_l(X) = -[F(X)\ln(F(X) + (1-F(X))\ln(1-F(X))] \qquad (4)$$

will be defined as the *locational entropy* at *X*. For generic marker values, the function $\kappa_l(x)$ will be referred to as the locational entropy function. Where a specific marker value is relevant, the upper case *X* will be used.

Basic properties of the locational entropy function can be summarised as follows.

Proposition 1

(a) For any distribution,

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

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

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

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

$$\kappa_l(X) = -F(X)E[\lambda(x) \mid x \le X] = (1-F(X))E[\lambda(x) \mid x > X], \qquad (5)$$

where the log odds function $\lambda(x)$ *refers to the unconditional distribution function F(x).*

Proof: Definition (4) implies that

$$\kappa_l'(x) = -f(x)\lambda(x). \qquad (6)$$

This has value zero only if $\lambda(x) = 0$, which means that $x=X_m$. Moreover, $\lambda(x)$ changes sign from negative to positive at $x=X_m$, so this is a maximum. The value *ln 2* follows by direct substitution into expression (4). The limiting value in (a) (i) likewise follows from (4), while the average value in (iii) follows by repeated integration by parts, again based on expression (4). Part (b) expression (5) is the integrated version of (6).

□

Locational entropy can be regarded as a measure of the information about the value of *x* derived from knowing whether $x \le X$ *or* $x > X$, measured as the expected reduction in differential entropy that would stem from the conditionality. In terms of entropy theory, it corresponds to the mutual information between variable *x* and a regime indicator function for the partition into either *L(X), R(X)*. Appendix A contains a proof and fuller discussion.

From expression (6), locational entropy is changing more rapidly where the log odds are non zero, i.e. away from the median, at points where the density remains high. If the density *f(x)* is symmetric about its median, then so is the entropy function. Figure 1 depicts the logistic locational entropy function for two values of the logistic dispersion parameter *β*, along with the densities. Locational entropy decays at a much slower rate for distributions with higher spread. For an asymmetric distribution, such as the Gumbel distribution, the locational entropy function is skewed in a fashion similar to the underlying density of the distribution, in this case to the right – there is higher entropy in the right hand tail, reflecting the slower decay rate of the log odds function.

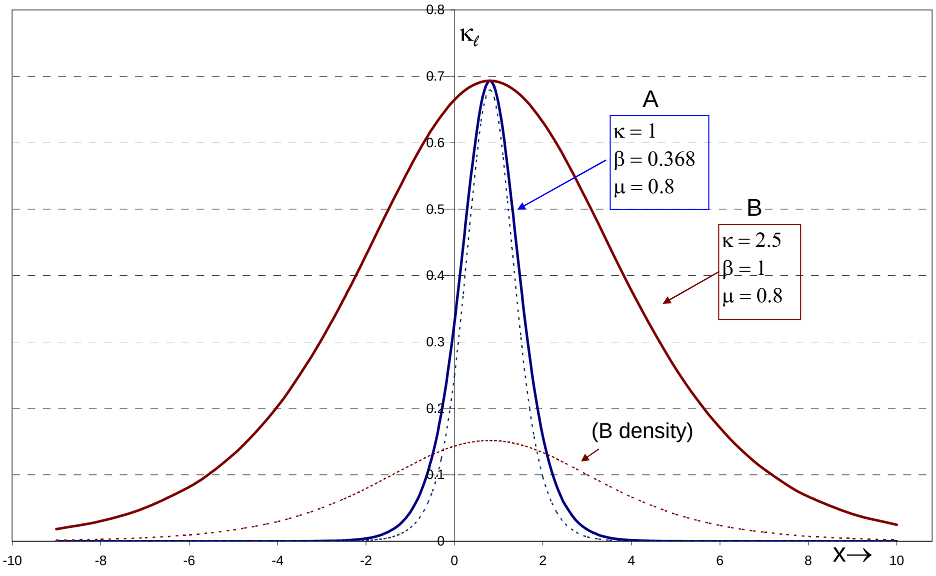

Figure 1: The effect of dispersion on locational entropy

Various decompositions exist involving locational entropy. It can be shown that total distribution entropy κ splits into two contributions, the first being locational entropy as in (4), and the second being the entropies of the conditional distributions for $L(X)$ and $R(X)$, weighted by their respective probabilities of occurring. An analogy is with the between-group and within-group sums of squares in the analysis of variance. The decomposition is in effect that for mutual information (Appendix A).

3. Distribution shifting

Locational 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).

3.1 Unit shifts

Lemma 1

Write

$$\xi_L(x) = -\ln F(x) \qquad (7a)$$

$$\xi_R(x) = -\ln(1 - F(x)). \qquad (7b)$$

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* $Q_R(B) = E[\mathrm{I}_B \xi_R]$.

Proof: It suffices to show that ξ_L and ξ_R are nonnegative random variables such that $E[\xi_L] = E[\xi_R] = 1$. The nonnegativity is immediate from the definitions (7a,b). The unit expectation follows by a straightforward integration by parts.

□

Corollary:

Any proper convex combination of the form

$$\xi_\theta(x) = \theta\xi_L(x) + (1-\theta)\xi_R(x); \quad 0 < \theta < 1, \qquad (8)$$

also qualifies as a Radon-Nikodym derivative for a change of measure from P to Q_θ *, say.*

□

Lemma 1 implies that for any measurable function *g(x),* $E_L[g(x)] = -E[\ln(F(x))g(x)]$, similarly for regime *R* using (7b). One can then establish the following:

Proposition 1

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

$$f_L(x) = -f(x)\ln(F(x)); F_L(x) = F(x)(1 - \ln(F(x)). \qquad (9a)$$

$$f_R(x) = -f(x)\ln(1-F(x)); F_R(x) = F(x) + (1-F(x))\ln(1-F(x)). \qquad (9b)$$

Proof: Set *g* as the set indicator functions in $\mathbb{R}$ for regimes *L*, *R* in turn. Thus if the regime marker value is *X*, set $g_L(x) = 1$ *for* $x \le X; = 0$, otherwise. Then

$$F_L(X) = E_L[g_L(x)] = \int_{-\infty}^{X} f(x)\ln(F(x))dx.$$

Integrating by parts gives $F_L(X)$, as in expression (9a) with $X = x$, and the density follows by differentiation. Or obtain the latter directly by setting $g(x) = \delta(x - X)$, the Dirac delta function, under smoothness assumptions on f (Lighthill (1959)). Similarly, for the right hand regime R .

□

Corollary:

The density and distribution function corresponding to the convex combination (6) are given by

$$f_\theta(x) = \xi_\theta(x) f(x) = \theta f_L(x) + (1-\theta) f_R(x);$$
$$F_\theta(x) = \theta F_L(x) + (1-\theta) F_R(x) . \qquad (10)$$

□

Multi-step shifting, if desired, can be accomplished via a 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 (9b). Sequential shifting can be a useful way to generate 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.

Figures 2a,b illustrate unit density and distribution function shifts for the logistic distribution. As one should expect, the right shift F_R is first order stochastic dominant over the natural distribution which is in turn dominant over the left shift. Figure 3 depicts shifted densities for the exponential distribution, which is asymmetric, and with support limited to the positive half line. Other details in these graphs, including the centred shift, will be commented on below.

The relative or Kullback-Leibler entropy of the right and left shifted distributions, taken with respect to the original, can be obtained as

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

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 functions $\ln(\xi_L(x))$ and $\ln(\xi_R(x))$ are concave for all the distributions considered in the present paper, so by Jensen's inequality, $-E[\ln(\xi_L(x))] \geq \ln(E(\xi_L(x))] = \ln(1) = 0$. Thus there is an increase of entropy when measured relative to the original distribution, as we should expect – shifting adds information.

In the particular case of the left shift for a 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, $F(x) = \exp(-\exp(-\tilde{x}))$; $\tilde{x} = (x - \mu)/\beta$, 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 $E[\tilde{x}] = \gamma$.

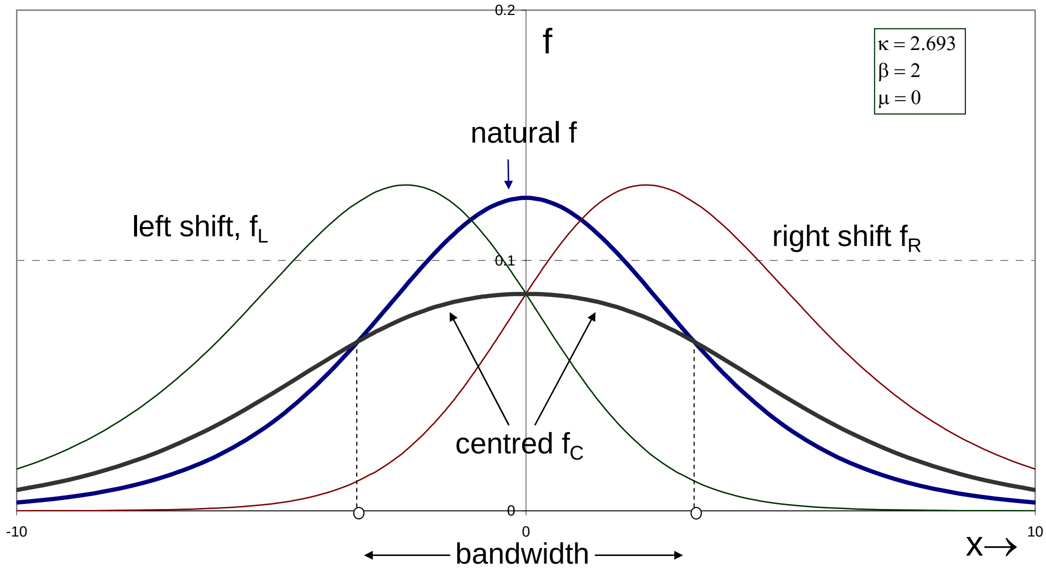

Figure 2a: Density shifting, logistic distribution

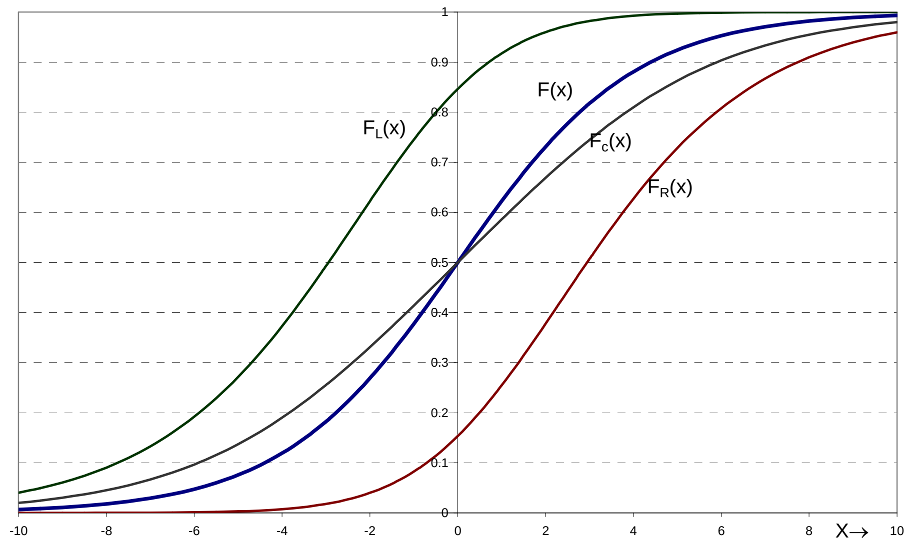

Figure 2b: Logistic cumulatives

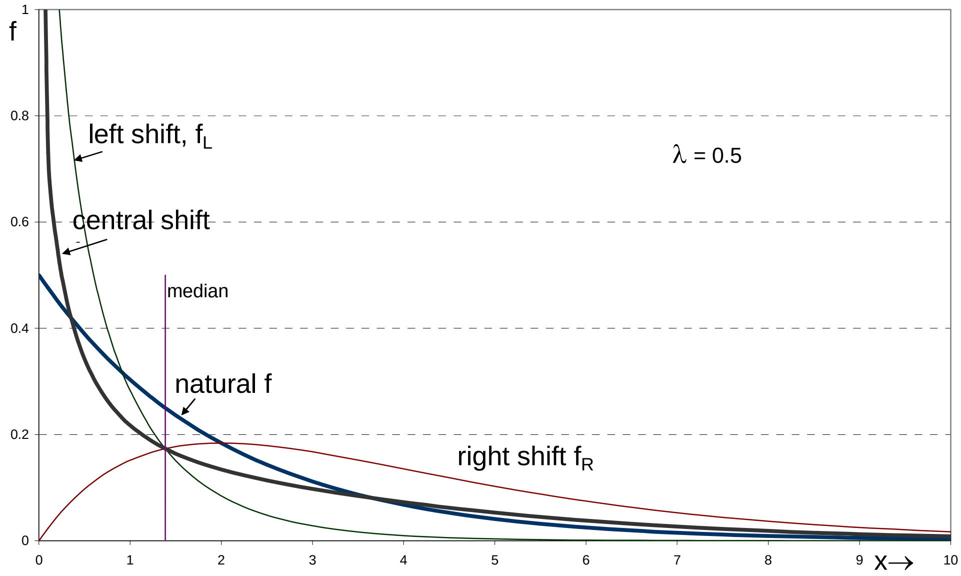

Figure 3: Density shifting, exponential distribution

3.2 Properties

Left and right shifted distributions and densities have a number of relativities, both to each other and to locational entropy.

(a) Locational entropy

Using expression (4) together with (9a,b), it follows that

$$F_L(x) - F_R(x) = \kappa_\ell(x) . \qquad (11)$$

Locational entropy can be obtained as the difference between the left and right shifted distribution functions. 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.

(b) Relativities at the median

A number of relational properties hold at the median X_m of the natural distribution.

(a) $F_\theta(X) = \frac{1}{2} + (\theta - \frac{1}{2}) \ln 2$

(b) $F_L(X_m) - F_R(X_m) = \kappa_\ell(X_m) = \ln 2$

(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 (9a,b), together with the definition (1) 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

$$f_L(X_m) = f_R(X_m) = f(X_m)\ln 2 . \qquad (12)$$

So the left and right hand densities must intersect at the median at 0.6931 of the value of the natural density value, and this property must be shared by every convex combination as in expression (10). Figures 2a,b illustrate; the centred density has θ = ½.

4. Entropy bandwidth and distribution diagnostics

The most important special case of a convex combination is θ = ½, which will be called the 'centred shift'. The associated density and distribution functions are:

$$f_c(x) = \frac{1}{2}(f_L(x) + f_R(x)) ; \qquad F_c(x) = \frac{1}{2}(F_L(x) + F_R(x))$$

The centred densities are marked in figures (2a,b) for the logistic distribution. As noted above they must intersect with the unit right and left hand shifts at a common point, corresponding to the natural median X_m . If the natural density is symmetric, then the centred density is also symmetric (e.g. figure 2a), but this will not be true if the original is asymmetric. Figure 4 illustrates with a Gumbel distribution, with the dispersion parameter $\beta = 1$.

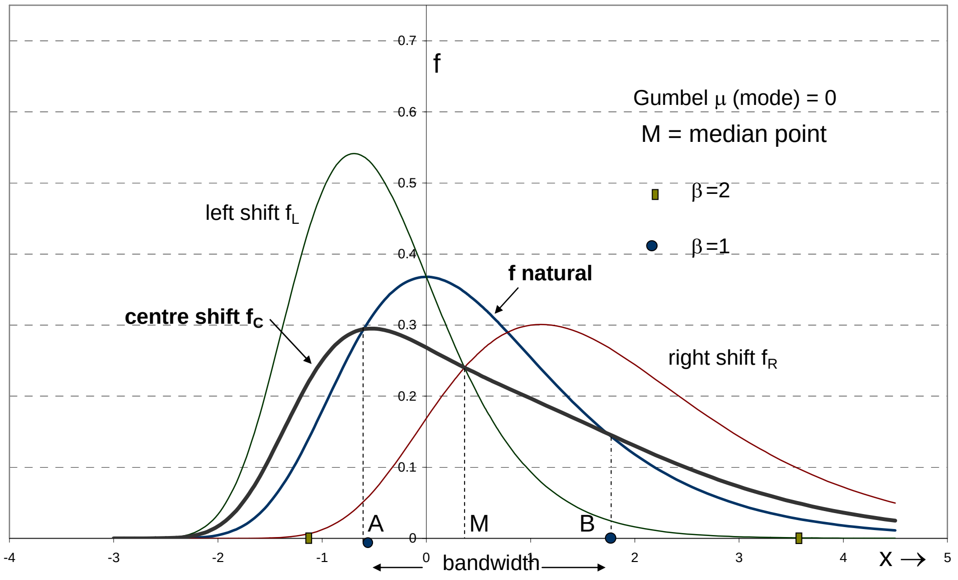

Figure 4: Gumbel distribution: centred shift

The centred shift can be explicitly written in the form

$$f_c(x) = \xi_c(x) f(x), \tag{13a}$$

where

$$\xi_c(x) = -0.5\ln[F(x)(1-F(x))]. \tag{13b}$$

Integrating both sides of expression (13a) over the entire support, and using expression (2) and (3), gives an expression for differential or total entropy as $\kappa = 2 - E[\ln(\lambda'(x)]$.

4.1 Entropy bandwidth and its uses

The notion of entropy bandwidth and associated symmetry and spread metrics flow more or less immediately from the following invariance properties.

Proposition 2

(a) The natural density and its centred shift intersect just twice at points x_L^*, x_R^* *such that*

$$F(x^*) = 0.5(1 \pm \sqrt{1-4e^{-2}}), \tag{14}$$

so that $F(x_L^*) \approx 0.161378; \quad F(x_R^*) \approx 0.838622$.

(b) At the points of intersection, $\kappa_\ell(x_L^*) = \kappa_\ell(x_R^*) = F(x_L^*) - F(x_R^*) \approx 0.441948$*, so that* x_L^*, x_R^* *are points of binary entropy symmetry on each side of the median.*

Proof: Any point *x** where the natural *f(x)* and centred *$f_c(x)$* densities intersect must satisfy $\xi_c(x^*) = -0.5\ln[F(x^*)(1-F(x^*))] = 1$. This yields two solutions $F(x^*) = 0.5(1 \pm \sqrt{1-4e^{-2}})$, as in part (a). Part (b) follows because $\ln(F(x_L^*) = \ln(1-F(x_R^*))$ and locational entropy is symmetric as between *F(x), (1-F(x),* with the actual value obtained from expression (11). Thus the vertical distance between the right and left hand shifted distribution functions must be the same.

□

The distance $\Delta = x_R^* - x_L^*$ will be referred to as the 'entropy bandwidth' of the given distribution. It is marked as such in figures 2a and 4.

A useful bandwidth property arises where the distribution admits a standardisation induced by a linear affine transformation, usually identified with location and scale parameters. Thus suppose that a two parameter distribution admits a standardisation $\tilde{x} = \frac{x-\mu}{\beta}$ such that $F(x;\mu,\beta) = F(\tilde{x};0,1) = \tilde{F}(\tilde{x})$, say. The normal, logistic and Gumbel are all instances. From expression (14), 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 4, the intervals AM, MB will expand in proportion.

The bandwidth can be used for metrics of asymmetry or spread that do not depend upon the normal distribution as benchmark, or of moment existence.

(a) Asymmetry metric

Distributional symmetry can be cast as referring to whether locational entropy is equally concentrated on either side of the median. The centring of the bandwidth interval relative to the median can be taken as an asymmetry metric:

$$\nu = \frac{x_R^* - X_m}{x_R^* - x_L^*} - \frac{1}{2} \qquad (15)$$

Positive (negative) values would indicate positive (negative) skewness. The limits are ½, - ½. Thus for the Gumbel distribution of figure 4, 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.

A useful invariance property arises where the distribution admits a standardisation. In this case the asymmetry metric ν, as defined by expression (15), 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, for the exponential distribution of figure 3 the symmetry metric comes out as ν = 0.1863; the bandwidth asymmetry is very evident from the figure. The Levy distribution (see below) is even more asymmetric.

(b) Spread

The entropy bandwidth Δ of the interval could also be taken as an index of the spread of the distribution. For the normal distribution, the entropy bandwidth would be 1.9776σ. For the logistic 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.

The Levy distribution provides an instructive case study for asymmetry and spread. 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 the ratio x/c. For the Levy distribution, the usual skewness and excess kurtosis metrics are undefined, as the relevant moments do not exist. Figure 5 sketches the shifted densities for $c = 1$; note the position of the median. The entropy bandwith is 23.6004 and the asymmetry metric $\nu = 0.4285$, 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.

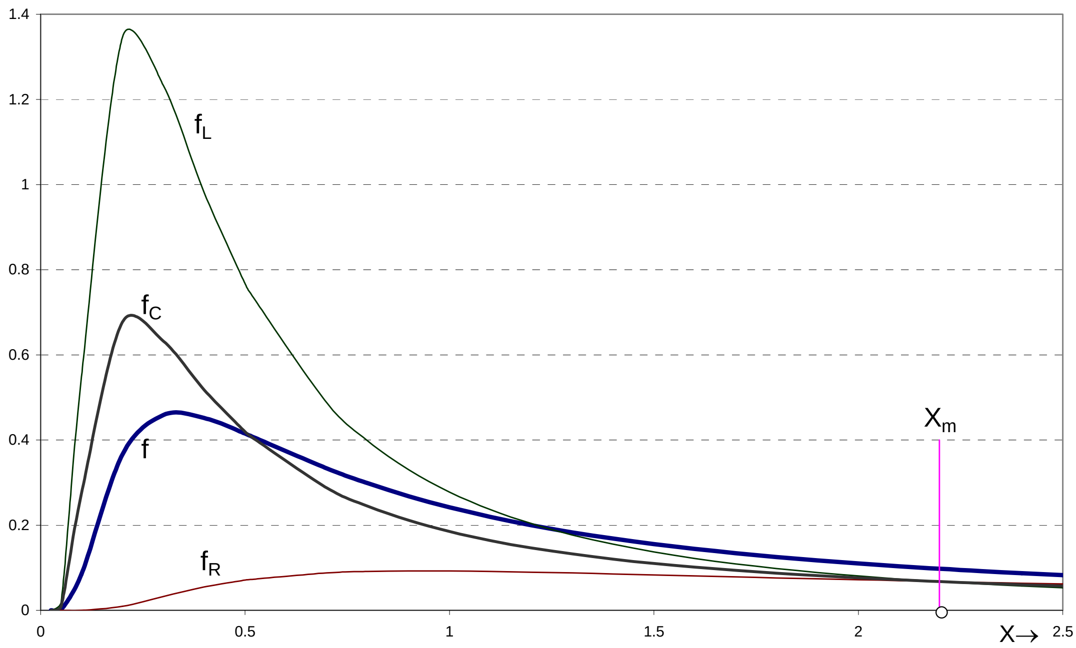

Figure 5: Unit shifts for the Levy density

4.2 Shift operations and information revelation

It is implicit in the foregoing that unit left and right shifts can be used to reveal the informational structure; separation width is an indicator of locational entropy. A more informal way to reveal information is to compare the centrally shifted distribution with the original. The following discussion is intended to illustrate this.

Figure 6 depicts a scenario under which an investor is contemplating a new venture that can result in one of two alternative future scenarios by the end of a development period. In the good case, a density f_g will apply to the capital value of the project. In the bad case, a density f_b will result. The distributions in both cases are imagined to be logistic, with the parameters μ (median, mean), β (dispersion) as indicated. The good state distribution has higher differential or total entropy (κ) – our investor hopes it has plenty of blue sky potential. The investor does not know which of the two distributions, good or bad, will apply, but assesses the probability of the good state as $\theta = 0.275$. The resulting density f is a mixture, as illustrated. Indeed, mixture densities have been suggested as a systematic way to model asymmetry; see McLachlan and Peel (2000).

The mixture aspect is not immediately apparent (e.g. to another investor without the same prior knowledge). However, figure 7 shows what happens when unit left and right shifts are applied, followed by the central shift. The central shifted distribution now has a quite obvious second mode, there is a wide entropy bandwidth (double headed arrow) and a fair degree of asymmetry is revealed. The second investor might reasonably conclude that there is an underlying mixture distribution, one element of which seems to have considerable upside potential.

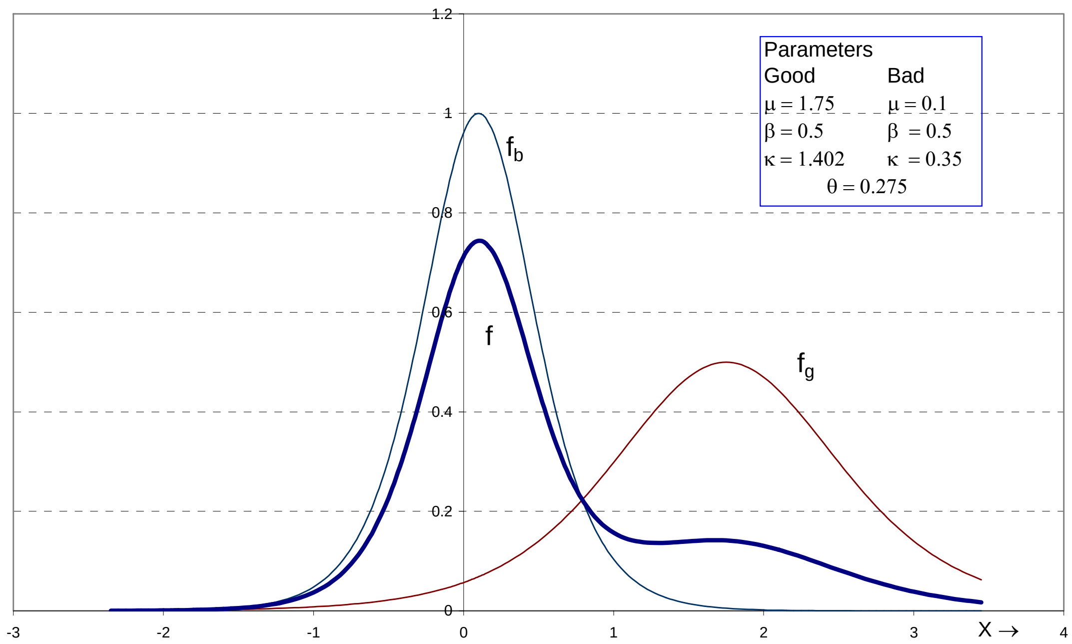

Figure 6: Original density mixture

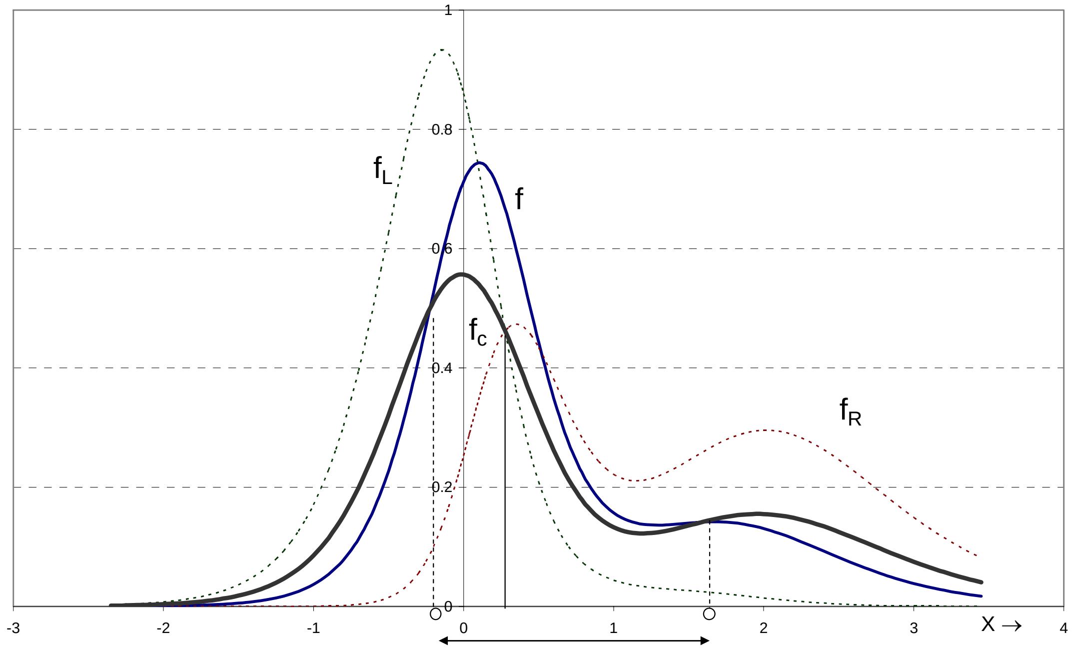

Figure 7: Shifted densities

5. Concluding remarks

Distribution with long tails are of interest in a number of contexts: reliability theory, investments including 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. The purpose of the present paper has been to put forward an approach to distribution diagnostics that calls upon information theory rather than on moments. The approach could perhaps be described as information mining; there is no better place to look for information than in the distribution itself, hence the reflexivity terminology.

It is possible to extend the approach in various ways. Bowden (2007) utilises a more directional approach to entropy that is useful in particular contexts such as options pricing or financial risk management. The approach does call on distributional benchmarking, in that case employing an alternative measure change that establishes a local logistic-equivalent distribution. In that sense it is more limiting than the present paper, which does not require any particular benchmark. Two further possible extensions are as follows.

(a) 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 ***X***, in which case $\boldsymbol{h} \propto \frac{\partial F}{\partial \boldsymbol{x}}]_{\boldsymbol{x}=\boldsymbol{X}}$, i.e. the steepest ascent direction. Once the direction is decided, the univariate framework applies as in the foregoing discussion, with the understanding that locational entropy and distribution shifting are specific to the designated direction.

(b) A further extension might be to consider cross shifting, where one distribution is used to shift another. Thus starting with *F(x)*, one might think of shifting with $\xi_L(x) = -\ln(G(x;\boldsymbol{\theta}))$, where *G* is another distribution function that depends on a set of parameters $\boldsymbol{\theta}$. The new left shifted density would be $\xi_L(x) f(x)$. 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 shifted left by a Gumbel distribution function *G* that depends on a dispersion parameter β and a location parameter *m*. In order to satisfy the above unit scaling requirement, 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 locational entropy and related concepts of distributional spread.

Appendix A: Locational entropy as mutual information

The following discussion enlarges on an earlier remark that locational entropy has a technical meaning in terms of the mutual information between regime membership and the actual value. It will be helpful to begin by formally extending the dimensionality of the problem. Given two random variables (*x*, α), let their joint and marginal density functions be $\phi(x,\alpha)$, *f(x)* and $p(\alpha)$, respectively. The following differential entropy concepts apply:

Joint entropy: $\kappa_{x,\alpha} = -E_{x,\alpha}[\ln(\phi(x,\alpha)]$; (A1a)

Marginal entropy for *x*: $\kappa_x = -E_x[\ln(f(x)]$; (A1b)

Marginal entropy for α: $\kappa_\alpha = -E_\alpha[\ln(p(\alpha)]$. (A1c)

The subscripts in the above indicate the dimension of variation. Conditional entropy concepts also arise, defined in terms of the conditional density $\phi(x\,|\,\alpha)$:

$$\kappa_{x|\alpha} = -E_{x|\alpha}[\ln(\phi(x\,|\,\alpha))], \quad \text{(A2)}$$

where the subscript denotes that the expectation is taken with respect to the conditional distribution of *x*, given α.

Lemma A1:

Joint entropy can be decomposed as

$$\kappa_{x,\alpha} = \kappa_\alpha + \kappa_m, \quad \text{(A3)}$$

where $\kappa_m = E_\alpha[\kappa_{x|\alpha}]$.

Proof: Using the iterated expectation:

$$-\kappa_m = E_\alpha[E_x[\ln(f(x\,|\,\alpha)] = E_{x,\alpha}[\ln(\frac{f(x,\alpha)}{p(\alpha)}] = E_{x,\alpha}[\ln(f(x,\alpha)] - E_\alpha[\ln(p(\alpha)].$$

Expression (A3) follows from definitions (A1a,c).

□

A final definition sourced from information theory is that of mutual information (Pinsker 1964), which refers to the information that knowing the value of α gives about the value of x . In the present context, this corresponds to

$$I_{a,x} = \kappa_x + \kappa_\alpha - \kappa_{x,\alpha} \quad \text{(A4)}$$

Equivalent definitions apply in terms of the difference between marginal and expected conditional entropy. Although $I_{\alpha,x}$ is a symmetric measure, the subscripts are ordered in the sense that α is imagined to provide information about *x*.

Extending the dimensionality in the above way is a useful way to handle distribution mixtures, where α plays the role of a mixing variable. However for present purposes, it may be noted that locational entropy can be regarded as a degenerate case of bivariate mutual entropy. Given a marker value *X*, we can imagine α to have categorical values labelled as *L*, *R*, with density elements $p(L) = F(X); \quad p(R) = 1 - F(X)$. It is apparent from definition (4) of the text that $\kappa_\alpha = \kappa_\ell(X)$, locational entropy at X. Moreover, set

$$\phi(x, L) = f(x),\ x \le X; \quad = 0\ otherwise,$$

with a complementary definition for $\phi(x,R)$, and set the corresponding log densities identically equal to zero wherever $\phi(x,*)=0$. Then

$$\kappa_{x,\alpha} = -\int_{-\infty}^{\infty}\{\phi(x,L)\ln(\phi(x,L)) + \phi(x,R)\ln(\phi(x,R))\}dx$$
$$= -\{\int_{X}^{\infty} f(x)\ln(f(x))dx + \int_{-\infty}^{X} f(x)\ln(f(x))dx\}$$
$$= \kappa_X .$$

It follows from Lemma A1 that

$$I_{\alpha,x} = \kappa_\alpha = \kappa_\ell(X) .$$

Thus locational entropy has a mutual information dimension, as a measure of information about the value of *x* derived from knowing whether $x \le X \; or \; x > X$.

References

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

Bowden, R.J. (2007) Directional entropy and tail uncertainty, with applications to financial hazard and investments, School of Economics and Finance, Victoria University of Wellington.

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

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

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

Lighthill, M.J. (1959) *An Introduction to Fourier Analysis and Generalised Functions*. New York: Academic Press.

Pinsker, M.S. (1964) *Information and Information Stability of Random Variables and Processes.* San Francisco: Holden Day.

Shilov, G. E., and Gurevich, B. L. (1977) *Integral, Measure, and Derivative: A Unified Approach*, transl. R. A. Silverman. Dover Publications.