The Annals of Probability
1974, Vol. 2, No. 1, 112–130

MIXTURES OF DISTRIBUTIONS, MOMENT INEQUALITIES AND MEASURES OF EXPONENTIALITY AND NORMALITY[1]

By Julian Keilson and F. W. Steutel

University of Rochester and Technische Hogeschool Twente

The central limit theorem and limit theorems for rarity require measures of normality and exponentiality for their implementation. Simple useful measures are exhibited for these in a metric space setting, obtained from inequalities for scale mixtures and power mixtures. It is shown that the Pearson coefficient of kurtosis is such a measure for normality in a broad class $\mathscr{D}$ containing most of the classical distributions as well as the passage time densities $s_{mn}(\tau)$ for arbitrary birth-death processes.

0. Introduction. Many limit theorems of interest to probability theory describe the convergence in distribution of a sequence of random variables. Application of such theorems in some special context is generally hampered by one's inability to say how close one is to the limit distribution. Measures of such proximity have been available, but have either been empirical and inadequately understood (e.g., kurtosis as a measure of normality), largely useful as a theoretical tool (e.g., the Lévy metric [17] page 215), or of limited value by virtue of the breadth of the class of distributions to which it addresses itself (e.g., the Berry–Esséen bound [5]).

By restricting the space of distributions considered, and focusing particular attention on the distance to the limit distribution of interest, one can exhibit simple tractable numerically accessible measures of this distance. One finds, for example, that for a certain broad class $\mathscr{D}$ of distributions (cf. Section 8), the classical Pearson coefficient of kurtosis appears as the distance of interest in a metric space setting (cf. Theorems 8.2 and 8.3). As a second example of interest, one finds that, in the class $\mathscr{E}_1$ of completely monotonic densities f_X, the entity $(\sigma^2/\mu^2)_X - 1$ is the distance to exponentiality in a metric space setting. The squared coefficient of variation (σ^2/μ^2) arises several times in such a context. Even the Pearson coefficient has this character in a hidden form.

The basic spaces dealt with are mixture spaces of two types, scale parameter mixtures and exponent parameter mixtures (power mixtures), and several important such spaces are discussed. A space of interest dealt with at length is the space $\mathscr{L}_2$ of mixtures of symmetric normal distributions. The extent of this class is surprisingly large. Many of the classical distributions of statistics and

Received September 13, 1972; revised May 14, 1973.

[1] Supported in part by the Office of Naval Research under Contract Nos. N00014-68-A-0091 and N0014-67-A-0398-0014.

AMS 1970 *subject classifications.* Primary 60E05, 60F05; Secondary 60J80.

Key words and phrases. Mixtures of distributions, moment inequalities, measures of exponentiality and normality, birth-death processes, weak convergence, log-concavity, log-convexity, completely monotone densities, total positivity, sojourn times.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability.* www.jstor.org

probability theory are represented in $\mathscr{L}_2$ or in the class $\mathscr{D}$ simply related to $\mathscr{L}_2$. All passage time densities for birth-death processes, for example, are found in $\mathscr{D}$.

Systems of moment inequalities for our mixtures of distributions are presented. These are directly related to the log-convexity of the moments of the mixing distributions, and related moment properties associated with log-concavity and log-convexity.

We denote rv's by capitals, their df's, pdf's and cf's by F, f and ϕ respectively, indexed by the rv.

If Z is an rv, then an rv that is distributed as $-Z$ is denoted by Z^D. In expressions of the form $Z + Z^D$, we shall always implicitly assume independence.

1. Mixtures of distributions and some of their properties.

1.1. *General mixtures.* We shall be concerned with df's of the form

$$F_X(x) = \int_{-\infty}^{\infty} K(x, w)\, dF_W(w) , \tag{1.1}$$

where F_W is any df and the kernel K has the properties: $K(x, w)$ is a df in x for every w in the support of F_W, $K(x, w)$ is Borel measurable in w for every x. Distributions with df's of the form (1.1) will be called mixtures, F_W will be called the mixing df. For a given kernel K the corresponding class of mixtures is denoted by $\mathscr{K}$. The support of the mixing df's may be restricted to part of the real line; here it will usually be $[0, \infty)$.

For an arbitrary kernel K, the correspondence between X and W is not one-to-one, i.e., there may be different F_{W_1} and F_{W_2} yielding the same F_X. Every class of mixtures has the properties stated in the following easily verified propositions.

PROPOSITION 1.1. *If K is a given kernel, then $\mathscr{K}$ is closed under mixing.*

PROPOSITION 1.2. *If K is absolutely continuous in x for all w in the support of F_W, with density $k(x, w)$, then F_X is absolutely continuous with*

$$f_X(x) = \int_{-\infty}^{\infty} k(x, w)\, dF_W(w) . \tag{1.2}$$

1.2. *Scale mixtures.* Many known classes of distributions are mixtures with kernels of the form

$$K_Y(x, w) = F_Y(x/w) , \tag{1.3}$$

where F_Y is a df, and mixing distributions with support in $[0, \infty)$. The corresponding pdf's and cf's take the form

$$f_X(x) = \int_0^{\infty} f_Y(x/w) w^{-1}\, dF_W(w) , \tag{1.4}$$

$$\phi_X(t) = \int_0^{\infty} \phi_Y(tw)\, dF_W(w) . \tag{1.5}$$

In terms of rv's we have ($=_d$ denoting equality in distribution)

$$X =_d YW , \tag{1.6}$$

where Y and W are independent. For a given Y, a mixture of this type will be called a scale mixture of Y-distributions, and the class of these mixtures referred to as a W-mixture of Y-distributions.

Much of our attention will be given to such scale mixtures and the presence of a one-to-one correspondence between X and W is of importance. We outline a proof of

PROPOSITION 1.3. *If* $|Y| > 0$ a.s., $E|Y|^\varepsilon < \infty$ *for some* $\varepsilon > 0$, *and if* $X \in \mathscr{K}_Y$, *with* $E|X|^\varepsilon < \infty$, *then* $EW^\varepsilon < \infty$, *and there is a one-to-one correspondence between* X *and* W.

PROOF. Clearly F_X is uniquely determined by F_W. For the converse, we first take $|X|$, and hence W, positive (a.s.). If now $X =_d YW_1 =_d YW_2$, then $E|X|^t = E|Y|^t EW_1^t = E|Y|^t EW_2^t$ for $0 \leqq t \leqq \varepsilon$. As $E|Y|^t > 0$, it follows that $EW_1^t = EW_2^t$ for $0 \leqq t \leqq \varepsilon$. Now, for $0 \leqq \operatorname{Re} z \leqq \varepsilon$ EW_1^z and EW_2^z exist and are analytic. As they coincide for $0 \leqq t \leqq \varepsilon$ it follows that $EW_1^z = EW_2^z$ for $0 \leqq \operatorname{Re} z \leqq \varepsilon$. Furthermore, as $EW^{it} = E\exp(it \log W)$ is a cf, it follows from the uniqueness theorem for cf's that $\log W_1 =_d \log W_2$, and hence that $W_1 =_d W_2$. If X is zero with probability p, then W is zero with probability p, and $(F_W - p)/(1-p)$ is uniquely determined by $(F_X - p)/(1-p)$ as above.

It is sometimes useful to keep the following simple observations in mind (cf. (1.6)):

(i) a scale mixture of a scale mixture of Y distributions is a scale mixture of Y-distributions;

(ii) a W-mixture of Y distributions is a Y-mixture of W-distributions if Y is nonnegative.

Many classes of distributions that are well defined otherwise can be exhibited as scale mixtures of a given distribution to enable properties to stand out more clearly. In the following sections a number of these classes will be discussed in detail. The most important of these have been listed in the table below.

For any class $\mathscr{K}$ of distributions we will interchangeably write $X \in \mathscr{K}$, $F_X \in \mathscr{K}$, $f_X \in \mathscr{K}$ or $\phi_X \in \mathscr{K}$, if the distribution of X is in $\mathscr{K}$.

1.3. *Power mixtures.* If ϕ_Y is an inf. div. cf, then $\phi_Y{}^w$ is a cf for every $w \geqq 0$. If we randomize with respect to w, we obtain the class $\mathscr{K}_Y{}^*$, defined by: $\phi_X \in \mathscr{K}_Y{}^*$ if and only if

$$\phi_X(t) = \int_0^\infty \{\phi_Y(t)\}^w \, dF_W(w) . \tag{1.7}$$

Mixtures of this kind were introduced by Feller [5]. With slightly more generality, one may consider cf's of the form (1.7) where $\phi_Y{}^w$ is a cf for all w in the support of F_W. For any distribution on the nonnegative integers with pgf $\pi(u)$, say, and any cf $\phi_Y(t)$, the function $\pi(\phi_Y(t))$ is a power mixture of Y-distributions. Power mixtures also occur in branching processes and queueing theory. Examples are provided by busy-period distributions, which have cf's $\beta(t)$ satisfying

TABLE 1

Class $\mathscr{K}_Y$	Y	$f_Y(x)$	$\phi_Y(t)$	$\zeta_K = E\|Y\|^k/E^k\|Y\|$	$\sigma^2 Y/E^2\|Y\|$	ref.
$\mathscr{E}_\alpha$	Γ^α	$e^{-x}x^{\alpha-1}/\Gamma(\alpha) \quad (x > 0)$	$(1 - it)^{-\alpha}$	$\Gamma(\alpha + k)/\{\Gamma(\alpha)\alpha^k\}$	α^{-1}	[7], [20]
$\mathscr{L}_2$	N	$e^{-x^2/2}/(2\pi)^{\frac{1}{2}}$	$e^{-t^2/2}$	$\Gamma\left(\frac{k+1}{2}\right)\pi^{(k-1)/2}$	$\frac{\pi}{2} + 1$	[7], [14], [15]
$\mathscr{U}_n$	U_n	$\left(1 - \frac{x}{n+1}\right)^n \quad (0 < x < n + 1)$	—	$(n + 2)^k \Big/ \binom{n + k + 1}{k}$	$\frac{n+1}{n+3}$	[16]
$\mathscr{E}_\alpha{}^S$	$\Gamma_\alpha{}^S$	—	$(1 + t^2)^{-\alpha}$	$\frac{\pi^{(k-1)/2}\Gamma\left(\frac{k+1}{2}\right)\Gamma^k\left(\alpha + \frac{k}{2}\right)\Gamma^{k-1}(\alpha)}{\Gamma^k(\alpha + \frac{1}{2})}$	$\frac{\Gamma(\alpha)\Gamma(\alpha + 1)}{\Gamma^2(\alpha + \frac{1}{2})}\frac{\pi}{2} - 1$	[7]

$\beta(t) = \phi(t + \lambda i - \lambda i\beta(t))$, where $\phi(t)$ is the cf of the service time (cf. [5]). The kernel for the class $\mathscr{K}_Y{}^*$ is given by

$$K_Y{}^*(x, w) = F_Y{}^{*w}(x) , \tag{1.8}$$

where $F_Y{}^{*w}$ is the w-fold convolution of F_Y with itself, defined by its cf $\phi_Y{}^w$.

One easily verifies that $\phi_{X_1} \cdot \phi_{X_2} = \int_0^\infty \phi_Y{}^w \, d\{F_{W_1} * F_{W_2}\}$, and that, correspondingly, one has

PROPOSITION 1.4. *If in* (1.7) $\phi_Y{}^w$ *is a well-defined* cf *for all* w *in the joint supports of* $F_W{}^{*1/n}$ *for* $n = 1, 2, \cdots$, *and if* W *is* inf. div., *then* X *is* inf. div.

Special cases of (1.7) yield the compound-Poisson and compound-geometric distributions. Trivially, when Y is degenerate, then (1.7) yields all distributions on $[0, \infty)$.

We easily have

PROPOSITION 1.5. *If* Y *is* inf. div., *then* $K_Y{}^*$ *is closed under mixing and convolution.*

Some of the scale mixtures we discuss subsequently can also be considered as power mixtures. This is possible whenever $\phi_Y(t) = \exp(at^b)$ or $\phi_Y(t) = \exp(a|t|^b)$, and this circumstance facilitates the discussion of infinite divisibility.

2. Mixtures of normal distributions. The class of scale mixtures of normal distributions encompasses directly or is simply related to a great many of the distributions of interest to probability theory and statistics. The extent of this class will soon be demonstrated. We start by relating the class to the symmetric stable distributions (cf. [18]) and providing a characterization in terms of complete monotonicity [5].

For $0 < \alpha \leqq 2$ we define the classes $\mathscr{L}_\alpha$ as follows: $\phi_X \in \mathscr{L}_\alpha$ iff ϕ_X is of the form

$$\phi_X(t) = \int_0^\infty e^{-\frac{1}{2}|tw|^\alpha} \, dF_W(w) . \tag{2.1}$$

Some properties of these classes were discussed in an earlier paper [14].

PROPOSITION 2.1. *The class* $\mathscr{L}_\alpha$ *consists of all real* cf's $\phi(t)$ *that are* c.m. (*completely monotone*) *in* $|t|^\alpha$, *i.e., such that* $\phi(|t|^{1/\alpha})$ *is* c.m. *on* $(0, \infty)$.

PROPOSITION 2.2. *If* $\alpha < \beta$, *then* $\mathscr{L}_\alpha \subset \mathscr{L}_\beta$, *and* $\mathscr{L}_\alpha \neq \mathscr{L}_\beta$.

Proposition 2.2 follows from the fact that any function $f(x)$ that is c.m. in x is also c.m. in x^γ for every $\gamma \geqq 1$, i.e., $f(x^{1/\gamma})$ is c.m. in x.

For $\alpha = 2$, we obtain the class of mixtures of normal distributions with cf's

$$\phi_X(t) = \int_0^\infty e^{-t^2w^2/2} \, dW_W(w) , \tag{2.2}$$

and corresponding pdf's (if $F_W(0+) = 0$)

$$f_X(x) = \int_0^\infty \frac{e^{-x^2/(2w^2)}}{(2\pi w^2)^{\frac{1}{2}}} \, dF_W(w) . \tag{2.3}$$

In terms of rv's we have (cf. Table 1)

$$X =_d NW, \tag{2.4}$$

where N is $N(0, 1)$.

PROPOSITION 2.3. *The classes $\mathscr{L}_\alpha$ are closed under mixing and convolution, the class $\mathscr{L}_2$ is also closed under multiplication of densities with suitable renorming if the product is integrable.*

As a direct consequence of Bernstein's theorem for complete monotonicity [5], we have

PROPOSITION 2.4. *The class $\mathscr{L}_2$ contains all symmetric* pdf'*s that are* c.m. *in* x^2, *and hence all symmetric* pdf'*s that are* c.m. *on* $(0, \infty)$.

It follows from Propositions 2.1, 2.2 and 2.4 that $\mathscr{L}_2$ contains a wide variety of known distributions. We list the following. $\mathscr{L}_2$ contains: the symmetric stable distributions; the Cauchy, Laplace and Student distributions and their mixtures; all rv's of the form $\Gamma_\alpha + \Gamma_\alpha{}^D$ (i.e., $\mathscr{L}_2 \supset \mathscr{E}_\alpha{}^S$, cf. Table 1) with cf's of the form $(1 + t^2)^{-\alpha}$, and all rv's of the form $Z + Z^D$, where the pdf of Z is PF_∞ (cf. [8]). These random variables arise as passage times of interest in birth-death processes [13]. A more general class of rv's of the type $Z + Z^D$ of interest to birth-death processes is given in the following proposition.

PROPOSITION 2.5. *If Z has a* pdf *which is a convolution of* c.m. pdf'*s or a limit of such, then $Z + Z^D \in \mathscr{L}_2$.*

PROOF. If f_Z is c.m., then its L.T. has the representation (cf. [20] page 44)

$$\tilde{f}_Z(s) = \exp\left\{-\int_0^\infty \frac{s}{x(x+s)}\, dm(x)\right\}, \tag{2.5}$$

where m is a measure bounded by Lebesgue measure. The converse is also true. It follows that the L.T. of any finite convolution of c.m. pdf's, i.e., of mixtures of exponential pdf's has a representation of the form (2.5), with m an absolutely continuous measure having a bounded Radon–Nikodym derivative m'. If we now take $\tilde{f}_Z(s) \cdot \tilde{f}_Z(-s)$, and replace s by $-it$, we obtain

$$\begin{aligned} \phi_Z(t)\phi_Z(-t) &= \exp\left\{-\int_0^\infty \frac{2t^2}{x^2(x^2+t^2)}\, x\, dm(x)\right\} \\ &= \exp\left\{-\int_0^\infty \frac{t^2}{y(y+t^2)}\, dn(y)\right\}, \end{aligned} \tag{2.6}$$

where n is absolutely continuous with $n'(x^2) = m'(x)$. Comparing (2.5) and (2.6), we see that the mixture of exponential distributions has been transformed into a mixture of Laplace distributions. As the Laplace distributions are in $\mathscr{L}_2$, $Z + Z^D$ is in $\mathscr{L}_2$ by Proposition 2.3. If $\tilde{f}_Z$ is the limit of a sequence of L.T.'s of the form (2.5), then $\phi_Z(t) \cdot \phi_Z(-t)$ is the limit of a sequence of cf's of the form (2.6). As $\mathscr{L}_2$ is closed under weak convergence, this concludes the proof.

Clearly, all cf's of the form (2.6) are inf. div. In order to discuss convolution more easily, we replace W by $V^{\frac{1}{2}}$ in (2.4). We then have

$$X =_d NV^{\frac{1}{2}} \tag{2.7}$$

with correspondingly

$$\phi_X(t) = \int_0^\infty e^{-t^2v/2}\, dF_V(v)\,. \tag{2.8}$$

From (2.8) it immediately follows that

$$\phi_{X_1}(t) \cdot \phi_{X_2}(t) = \int_0^\infty e^{-t^2v/2}\, d(F_{V_1} * F_{V_2})(v)\,, \tag{2.9}$$

or equivalently, for X_1 and X_2 in $\mathscr{L}_2$ and independent

$$X_1 + X_2 =_d N(V_1 + V_2)^{\frac{1}{2}}\,, \tag{2.10}$$

and hence, rather curiously, for N_1 and N_2 independent and $N(0, 1)$

$$N_1 V_1^{\frac{1}{2}} + N_2 V_2^{\frac{1}{2}} =_d N(V_1 + V_2)^{\frac{1}{2}}\,. \tag{2.11}$$

Clearly, if V is inf. div., then $X =_d NV^{\frac{1}{2}}$ is inf. div.: If $V =_d V_1 + \cdots + V_n$ with identically distributed V_j, then by (2.11) $X =_d N(V_1 + \cdots + V_n)^{\frac{1}{2}} =_d N_1 V_1^{\frac{1}{2}} + \cdots + N_n V_n^{\frac{1}{2}} =_d X_1 + \cdots + X_n$ with identically distributed X_j. It was noted by Kelker [15] that $NV^{\frac{1}{2}}$ may be inf. div. without V being inf. div. It follows that in that case the components of $NV^{\frac{1}{2}}$ cannot all be in $\mathscr{L}_2$.

REMARK 1. In the same way as above it follows that if X is symmetric stable of order α, then V is positive stable of order $\alpha/2$.

More generally one has (cf. [5] page 562).

PROPOSITION 2.6. *If X is symmetric stable of order α, then $X =_d YV^{1/\gamma}$, where Y and V are independent with Y symmetric stable of order $\gamma > \alpha$ and V positive stable of order α/γ. The converse also holds.*

PROOF. By Proposition 2.2 we have for $\gamma \leqq 2$

$$Ee^{itX} = e^{-|t|^\alpha} = \int_0^\infty e^{-|t|^\gamma v}\, dF_V(v) = \int_0^\infty e^{-|tv^{1/\gamma}|^\gamma}\, dF_V(v) = Ee^{itYV^{1/\gamma}}\,, \tag{2.12}$$

where Y and V are independent, Y is symmetric stable of order γ, and, by the uniqueness theorem for L.T.'s, V has L.T. $\exp(-s^{\alpha/\gamma})$.

REMARK 2. As all symmetric stable distributions are in $\mathscr{L}_2$, and as all pdf's in $\mathscr{L}_2$ (cf. (2.3)) are unimodal, all symmetric stable distributions are unimodal (cf. [18] page 158).

3. Other classes of scale mixtures. Other classes of scale mixtures are of interest. The most important of these, perhaps, is the class of completely monotonic densities which are scale mixtures of exponentials. These play a key role in stochastic processes reversible in time [10]. Classes of importance introduced by van Dantzig and discussed recently by Kemperman arise naturally from notions of monotonicity, convexity and their generalization. Of special importance are the one-sided and two-sided unimodal and convex densities.

3.1. *The classes* $\mathscr{E}_\alpha$. These consist of mixtures of gamma-distributions of order α $(0 < \alpha < \infty)$. We have: $\phi_X \in \mathscr{E}_\alpha$ iff

$$\phi_X(t) = \int_0^\infty (1 - itw)^\alpha \, dF_W(w) . \tag{3.1}$$

We list the following propositions without proof.

PROPOSITION 3.1. *If* $\alpha < \beta$, *then*

$$(1 - it)^{-\alpha} = \frac{\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta - \alpha)} \int_0^1 (1 - itu)^{-\beta} u^{\alpha-1}(1 - u)^{\beta-\alpha-1} \, du . \tag{3.2}$$

PROPOSITION 3.2. *If* $\alpha < \beta$, *then* $\mathscr{E}_\alpha \subset \mathscr{E}_\beta$ *(strict inclusion), and* $\lim_{\alpha\to\infty} \mathscr{E}_\alpha = \mathscr{P}$, *the set of all distributions with nonnegative support.*

The class $\mathscr{E}_1$ is of special interest. Two of its properties are noted in

PROPOSITION 3.3. *If* $f_X \in \mathscr{E}_1$, *then* f_X *is completely monotonic, i.e.,* f_X *is of the form*

$$f_X(x) = \int_0^\infty w^{-1} e^{-x/w} \, dF_W(w) . \tag{3.3}$$

Furthermore, f_X *is infinitely divisible* (cf. [19]).

3.2. *The classes* $\mathscr{E}_\alpha{}^S$. These consist of all distributions with cf's of the form

$$\phi_X(t) = \int_0^\infty (1 + t^2w^2)^{-\alpha} \, dF_W(w) \qquad (\alpha > 0) , \tag{3.4}$$

i.e., of all mixtures of $\Gamma_\alpha + \Gamma_\alpha{}^D$, where Γ_α has a gamma distribution of order α. We have

PROPOSITION 3.4. *If* $\alpha < \beta$, *then* $\mathscr{E}_\alpha{}^S \subset \mathscr{E}_\beta{}^S$ *(strict inclusion), and* $\lim_{\alpha\to\infty} \mathscr{E}_\alpha{}^S = \mathscr{L}_2$ (cf. *Section* 2).

PROPOSITION 3.5. *If* $f_X \in \mathscr{E}_1{}^S$, *then* f_X *is completely monotonic on* $(0, \infty)$, *and infinitely divisible.*

3.3. *The monotonicity and convexity classes* $\mathscr{U}_n$. The class $\mathscr{U}_n$ consists of mixtures of densities f_{Y_n}, with

$$f_{Y_n}(x) = \left(1 - \frac{x}{n+1}\right)^n \quad \text{if} \quad 0 < x < n + 1 \tag{3.5}$$
$$= 0 \qquad \text{otherwise} ,$$

and $EY_n = (n + 1)/(n + 2)$. We have $X \in \mathscr{U}_n$ iff

$$f_X(x) = \int_{x/(n+1)}^\infty \left(1 - \frac{x}{(n+1)w}\right)^n \frac{1}{w} \, dF_W(w) . \tag{3.6}$$

These classes have been used by Kemperman [16] with a slightly different norming, and earlier by van Dantzig [4]. For $n = 0$ one obtains all non-increasing densities on $(0, \infty)$, for $n = 1$ all densities on $(0, \infty)$ that are convex. It is possible to give n arbitrary real values > -1, with $\mathscr{U}_{-1}$ representing all distributions on $[0, \infty)$; in the sequel we shall only consider $\mathscr{U}_n$ for integer $n \geqq 0$.

It seems obvious from comparing (3.3) and (3.6) that $\mathscr{U}_n \to \mathscr{E}_1$. We now formally state some properties of the $\mathscr{U}_n$ (cf. [16]).

PROPOSITION 3.6. *$f_X \in \mathscr{U}_n$ iff F_X has an nth derivative $F_X^{(n)}$ in $(0, \infty)$ with $(-1)^{n+1}F_X^{(n)}$ convex in $(0, \infty)$.*

COROLLARY. *$f_X \in \mathscr{U}_n$ for $n \geqq 1$ iff f_X has an $(n-1)$st derivative with $(-1)^{n+1}f_X^{(n-1)}$ convex.*

PROPOSITION 3.7. *$\mathscr{U}_{n+1} \subset \mathscr{U}_n$ for $n = 0, 1, 2, \cdots$ (strict inclusion), and $\lim_{n\to\infty} \mathscr{U}_n = \mathscr{E}_1$.*

For $f_n \in \mathscr{U}_n$, with its mixing df $F_W = G$ having a finite first moment, μ, integration by parts yields

$$(3.7)\qquad \frac{1}{\mu}\int_{((n+1)/(n+2))x}^{\infty} f_n(y)\,dy = \int_{x/(n+2)}^{\infty}\left(1 - \frac{x}{(n+2)w}\right)^{n+1}\frac{1}{w}\,\frac{w}{\mu}\,dG(w)$$
$$= f_{n+1}(x) \in \mathscr{U}_{n+1}\,.$$

Apparently, when G has all moments, and when the procedure leading from f_n to f_{n+1} converges to a pdf f, then f must be completely monotonic. This is in accordance with a result by Harkness and Shantaram [6] to the effect that if l_n and b_n are such that the sequence $\{f_n\}$ defined by a given f, and a recurrence relation

$$f_{n+1}(x) = b_n\{1 - F_n(l_n x)\}\,,$$

converges to f, then F satisfies

$$(3.8)\qquad F(x) = b\int_0^{lx}\{1 - F(y)\}\,dy\,.$$

It also follows from (3.8) that $f(x)$ is completely monotonic.

4. Moment inequalities.

4.1. *Scale mixtures.* Let X be the scale mixture $X = YW$ defined in Sub-section 1.2. For the absolute moments $\beta_{Uk} = E|U|^k$ one has from the independence of X and Y

$$(4.1)\qquad \frac{\beta_{Xk}^{1/k}}{\beta_{X1}} = \frac{\beta_{Yk}^{1/k}}{\beta_{Y1}}\,\frac{\beta_{Wk}^{1/k}}{\beta_{W1}}$$

whenever the moments are finite. The sequence $\beta_{Wk}^{1/k}$ is strictly increasing (cf. [17] page 156) whenever W is non-degenerate and is constant when W is degenerate. The following theorem is immediate.

THEOREM 4.1. *For any scale mixture $X = YW$ one has the infinite system of inequalities*

$$(4.2)\qquad \{\zeta_k^{-1}\beta_{Xk}\}^{1/k} \geqq \{\zeta_j^{-1}\beta_{Xj}\}^{1/j}\,,\qquad k > j,\ j, k = 1, 2, \cdots$$

where $\zeta_k = \beta_{Yk}/\beta_{Y1}^k$ is a fixed sequence of positive numbers for each rv Y. The inequality is strict for all distinct pairs j, k or for none. These inequalities for the completely monotone class $\mathscr{E}_1$ and for related broader classes were given in [1] and [9].

The parameters ζ_k for the different classes of interest are given in Table 1. Of particular importance is the relationship between β_{X1} and β_{X2}. Let $\sigma_X{}^2 = (\beta_{X2} - \beta_{X1}^2)$.

PROPOSITION 4.2. *For any scale mixture $X = YW$, one has*

$$\frac{\sigma_X{}^2}{\beta_{X1}^2} \geqq \frac{\sigma_Y{}^2}{\beta_{Y1}^2} = \zeta_2 - 1 \tag{4.3}$$

with equality if and only if the mixture is pure, i.e., if and only if the mixing distribution is degenerate.

A somewhat stronger version of (4.3) is of interest will be needed subsequently. As for (4.1) one may write

$$\frac{\beta_{X2}}{\beta_{X1}^2} = \frac{\beta_{Y2}}{\beta_{Y1}^2}\frac{\beta_{W2}}{\beta_{W1}^2} = \frac{\beta_{Y2}}{\beta_{Y1}^2}\frac{\sigma_W{}^2}{\beta_{W1}^2} + \frac{\beta_{Y2}}{\beta_{Y1}^2}.$$

Hence

$$\frac{\sigma_X{}^2}{\beta_{X1}^2} - \frac{\sigma_Y{}^2}{\beta_{Y1}^2} = \left(\frac{\beta_{Y2}}{\beta_{Y1}^2}\right)\frac{\sigma_W{}^2}{\mu_W{}^2} \tag{4.4}$$

where $\beta_{W1} = \mu_W = EW$, since W is nonnegative. It follows that the coefficient of variation of the mixing distribution is a measure of the departure of an element X in a class of scale mixtures from purity. (This statement will be given sharper meaning in Section 5.) We note in keeping with this that when $W = W_1 W_2$, where W_1 and W_2 are independent mixing random variables, the coefficient of variation of W exceeds that of W_1 or W_2. Indeed one has from (4.4) for $Y = W_1$, $W = W_2$,

$$\left(\frac{\sigma^2}{\mu^2}\right)_{W_1W_2} = \left(\frac{\sigma^2}{\mu^2}\right)_{W_1} + \left(\frac{\sigma^2}{\mu^2}\right)_{W_2} + \left(\frac{\sigma^2}{\mu^2}\right)_{W_1}\left(\frac{\sigma^2}{\mu^2}\right)_{W_2}. \tag{4.5}$$

The following property of the coefficient of variation is of some interest in itself and will be needed subsequently.

PROPOSITION 4.3. *For any* rv $V \geqq 0$ *with* $EV^{2r} < \infty$ *the coefficient of variation for* V^α, $(\sigma/\mu)_{V^\alpha}$ *increases with* α *on* $[0, r]$. *The increase is strict except when* V *is degenerate or concentrated on two points, one of which is zero.*

PROOF. It is well known that $\beta_{V\alpha}$ is log-convex on $[0, 2r]$. Let $f(x) = \log \beta_{Vx}$. Then $f'(2x) \geqq f'(x)$ on $I = [0, r]$ and hence

$$\int_b^c f'(x)\,dx \leqq \int_b^c f'(2x)\,dx$$

for $b, c \in I$, i.e., $2f(c) + f(2b) \leqq 2f(b) + f(2c)$ and the result follows. The strictness for V as stated is a consequence of the strict log-convexity.

Special structural characteristics of W may permit other systems of inequalities. If, for example, $f_W(x)$ is log-convex on $(0, \infty)$, then it is known that

$\{\beta_{Wk}/\Gamma(k+1)\}^{1/k}$ increases with k [1]. As for (4.2) we then have

$$\left\{\frac{\beta_{Wk}}{k!}\right\}^{1/k} \nearrow \Rightarrow \left\{\frac{\zeta_k^{-1}\beta_{Xk}}{k!}\right\}^{1/k} \nearrow \tag{4.6a}$$

and, in particular, as for (4.3).

$$\frac{\sigma_X^2}{\beta_{X1}^2} - 1 \geqq 2\left(\frac{\sigma_Y^2}{\beta_{Y1}^2}\right). \tag{4.6b}$$

Similarly, when f_W is log-concave, or when W has log-concave lattice support, then $\{\beta_{Wk}/\Gamma(k+1)\}^{1/k}$ is decreasing in k [1] and we have

$$\left\{\frac{\beta_{Wk}}{k!}\right\}^{1/k} \searrow \Rightarrow \left\{\zeta_k^{-1}\frac{\beta_{Xk}}{k!}\right\}^{1/k} \searrow \tag{4.7a}$$

and, in particular

$$\frac{\sigma_X^2}{\beta_{X1}^2} - 1 \leqq 2\frac{\sigma_Y^2}{\beta_{Y1}^2}. \tag{4.7b}$$

For mixtures of normals ($\mathscr{L}_2$) one has $X = NV^{\frac{1}{2}}$. An almost identical argument to that for (4.6a, b) and (4.7a, b) shows that these inequalities require only minor change when $\{\beta_{Vk}/\Gamma(k+1)\}^{1/k}$ is monotone.

In particular, we find that when $\{\beta_{Vk}/\Gamma(k+1)\}^{1/k}$ decreases with k, then $\{\beta_{Xk}/\Gamma(k/2+1)\}^{1/k}$ decreases with k. For $k = 1, 2$, we obtain

$$\frac{\sigma_X^2}{\beta_{X1}^2} - 1 \leqq \frac{4}{\pi}. \tag{4.8}$$

The inequalities in (4.6a) and (4.6b) are applicable, for example, to F-distributions with random variables of the form

$$X = \Gamma_{m/2}/\Gamma_{n/2},$$

where $\Gamma_{m/2}$ is a gamma variate of half-integral order. The numerator $\Gamma_{m/2}$ may be thought of as the mixing variate and its density is log-convex for $m = 1$, and log-concave for all $m \geqq 2$. Thus (4.6a) and (4.6b) are applicable for $m = 1$ and (4.7a) and (4.7b) for $m \geqq 2$ (for $m = 2$ one has equality).

4.2. *Power mixtures.* Consider the power mixture $X \in \mathscr{K}_Y{}^*$ as introduced in Sub-section 1.3. For the cf one has

$$\phi_X(t) = \int_0^\infty \{\phi_Y(t)\}^w \, dF_W(w), \tag{4.9}$$

and hence by differentiation $EX = EW \cdot EY$, $EX^2 = EW(EY^2 - E^2Y) + EW^2E^2Y = EW\sigma_Y^2 + EW^2\mu_Y^2$.
It follows that for $\mu_Y \neq 0$,

$$\frac{\sigma_X^2}{\mu_X} = \frac{\sigma_Y^2}{\mu_Y} + \mu_Y\frac{\sigma_W^2}{\mu_W}, \tag{4.10}$$

and therefore we have

PROPOSITION 4.4. *For any power mixtures* $X \in \mathscr{K}_Y{}^*$ *with* μ_Y *finite and positive*

$$\frac{\sigma_X{}^2}{\mu_X} \geqq \frac{\sigma_Y{}^2}{\mu_Y}, \tag{4.11}$$

with equality only when the mixing rv *is degenerate.*

For compound-Poisson or compound-geometric distributions (4.11) becomes

$$\left(\frac{\sigma^2}{\mu^2}\right)_X \geqq K\left(\frac{\sigma^2}{\mu^2}\right)_Y \tag{4.12}$$

where K is simply related to the compounding parameter.

5. Metrics for mixture spaces; convergence in distribution. The introduction of a convenient distance function on a space of scale mixtures will enable us in Section 6 to provide measures of exponentiality and normality for elements of mixture classes.

5.1. *Scale mixtures.* Consider the class of mixtures $\mathscr{K}_Y$ consisting of the scale mixtures of Y-distributions, i.e., the set $\mathscr{K}_Y = \{F_X: X = YW;\ Y, W \text{ independent};\ W \geqq 0\}$. Clearly $F_1, F_2 \in \mathscr{K}_Y \Rightarrow p_1F_1(ax) + p_2F_2(ax) \in \mathscr{K}_Y$ for all $p_1 \geqq 0$, $p_2 \geqq 0$, $p_1 + p_2 = 1$, and $a \geqq 0$. We have seen in Proposition 1.3, that under simple, fairly broad conditions, there is one-to-one correspondence between F_W and F_X in $\mathscr{K}_Y$.

A distance function or metric can be associated with $\mathscr{K}_Y$ in a simple way, giving rise to a metric space in which the squared coefficient of variation appears as a key distance. Certain moments will also be needed. We introduce the following space. Let

$$\mathscr{K}_{Yr} = \{F_X: X = YW;\ Y, W \text{ independent};\ W \geqq 0;\ EW = 1;\ EW^r < \infty\}.$$

Let F_1, F_2 in $\mathscr{K}_{Yr}$ be associated with W_1, W_2, and let $\mu_{W_1}(dw)$, $\mu_{W_2}(dw)$ be the corresponding measures. We define the distance function

$$\rho_\theta(F_1, F_2) = \int |w - 1|^\theta |\mu_{W_1}(dw) - \mu_{W_2}(dw)|, \qquad \theta \leqq r. \tag{5.1}$$

The importance of the restriction $EW = 1$ and the utility of the factor $|w - 1|^\theta$ will appear in Section 6. When $\theta = 2$ and W_1 has all support at $w = 1$, we will see in Section 6 that the distance $\rho_2(F_1, F_2)$ coincides with the squared coefficient of variation of W_2, i.e., to $(\sigma^2/\mu^2)_{W_2}$. When $EW_2 = 1$, this is just equal to Var W_2. For measures of distribution shape, e.g., normality or exponentiality, scale is irrelevant, and indicators of shape must be invariant under scale transformation, as is $(\sigma^2/\mu^2)_W$. Transformation of the scale of X and hence W to the subspace for which $EW_2 = 1$ and measurement of distance to the "pure" distribution with $W_1 = 1$ is then appropriate. (See (6.2) and (6.3).)

We next show that $[\mathscr{K}_{Yr}, \rho_\theta(F_1, F_2)]$ is a metric space.

THEOREM 5.1. *For* $0 \leqq \theta \leqq r$, $\rho_\theta(F_1, F_2)$ *is a metric for distributions in* $\mathscr{K}_{Yr}$, *i.e., for all* F_j *in* $\mathscr{K}_{Yr}$ *we have*

(a) $\rho_\theta(F_1, F_2) \geqq 0$
(b) $\rho_\theta(F_1, F_2) = 0 \Leftrightarrow F_1 = F_2$
(c) $\rho_\theta(F_1, F_2) = \rho_\theta(F_2, F_1)$
(d) $\rho_\theta(F_1, F_3) \leqq \rho_\theta(F_1, F_2) + \rho_\theta(F_2, F_3)$.

PROOF. The nonnegativity, and symmetry of (a) and (c) are trivial. The triangle inequality of (d) is available from the classical argument $|\mu_1 - \mu_3| = |\mu_1 - \mu_2 + \mu_2 - \mu_3| \leqq |\mu_1 - \mu_2| + |\mu_2 - \mu_3|$. For property (b) we note that equality requires that the measures coincide away from the point $w = 1$ and hence at the point $w = 1$ as well. For $\rho_\theta = 0$ when $F_1 = F_2$ see Proposition 1.3.

The metric given in (5.1) provides a basis for discussing convergence in distribution in the homogeneous mixture space $\mathscr{K}_{Yr}$. The following theorem is immediate.

THEOREM 5.2. *Let $(X_j)_1^\infty$ be a sequence of random variables in $\mathscr{K}_{Yr}$. The sequence converges weakly to $X_0 \in \mathscr{K}_{Yr}$ if $\rho(F_j, F_0) \to 0$ as $j \to \infty$.*

5.2. *A related metric for $\mathscr{L}_2$.* For mixtures of normals $\mathscr{L}_2$ we have seen that (cf. (2.7) to (2.11)) the representation $X = NV^{\frac{1}{2}}$ is advantageous in that $X_1 + X_2 \leftrightarrow V_1 + V_2$ for convolution. In this and other similar cases it is useful to employ as the distance between distributions F_1, F_2 in $\mathscr{L}_2$ the metrics (5.1) for the corresponding distributions of V with $EV = \operatorname{Var} X = 1$, i.e.,

$$\rho_\theta(F_1, F_2) = \int |v - 1|^\theta |\mu_{V_1}(dv) - \mu_{V_2}(dv)| \,. \tag{5.2}$$

Clearly $\rho_\theta(F_1, F_2)$ again satisfies the conditions for a metric (cf. Theorem 5.1). The case $\theta = 2$, as we shall see in the next section, is of special interest.

6. Measures of exponentiality and normality. Two important limit theorems in probability theory center about exponentiality and normality. Exponentiality is of importance to rare events (cf. [12] and [3] page 257), and normality to the central limit theorem. Such limit theorems are often weakened in their practical impact by the lack of good criteria for knowing how close one is to the limiting distribution. The metric given in (5.1) provides such a measure of closeness in a setting sufficiently broad to be of interest.

Let us suppose that X_0 in Theorem 5.2 is equal in distribution to that of the basic rv Y, i.e., that $X_0 =_d YW_0$ with $W_0 = 1$, so that the measure for W_0 has all support at $w = 1$. Then for $X \in \mathscr{K}_{Y2}$

$$\begin{aligned}\rho_{W2}(F_X, F_{X_0}) &= \int (w - 1)^2 |\mu_W(dw) - \mu_{W_0}(dw)| \\ &= \int (w - 1)^2 \mu_W(dw) = E(W - 1)^2\end{aligned}$$

i.e.,

$$\rho_{W2}(F_X, F_{X_0}) = \operatorname{Var} W = \sigma_W^2/\mu_W^2 \,. \tag{6.1}$$

It follows that the squared coefficient of variation of the mixing distribution is a measure of distance to the distribution F_Y in the sense of Theorem 5.2, i.e., is the metric of key interest on the scale mixture space. The choice of index

$\theta = 2$ for $\rho_{W\theta}$ is clearly dictated by convenience and tractability. The emergence of the classical coefficient of variation in this setting, however, is somewhat striking.

It is convenient to have a measure of distance from F_X to a pure F_{X_0}-type df (i.e., the df of cX_0 for a positive c) for all $X \in \mathscr{K}_Y$ having a mixing distribution with a finite second moment. We therefore define

$$\rho_{W2}(F_X, F_{\mu_W X_0}) = \rho_{W2}(F_{X/\mu_W}, F_{X_0}) = \sigma_W{}^2/\mu_W{}^2 . \tag{6.2}$$

If $X_0 =_d N$, then ρ_{W2} is a measure of distance to "normality" (for all $X \in \mathscr{L}_2$ having a finite second moment) rather than a measure of distance to F_N. However, for the more general distance (6.2), Theorem 5.2 does not hold.

The metric can be given directly in terms of X and Y from (4.4). One has

$$\rho_{W2}(F_{X/\mu_W}, F_{X_0}) = \frac{\beta_{Y1}^2}{\beta_{Y2}} \left\{ \frac{\sigma_X{}^2}{\beta_{X1}^2} - \frac{\sigma_Y{}^2}{\beta_{Y1}^2} \right\} . \tag{6.3}$$

Exponentiality and normality are of greatest practical importance. In Section 7, we will see how the metric (5.1) is useful for studying convergence to exponentiality in birth-death processes. In Section 8 we will examine normality in the context of the central limit theorem, where variants of (5.1) are often needed.

7. The rate of convergence to exponentiality of sojourn time in birth-death processes. As an illustration of the utility of our results for the study of limit behavior, we consider convergence to exponentiality in birth-death processes. Let $N(t)$ be such a process governed by upward and downward transition rates λ_n, μ_n with $\mu_0 = 0$, $\lambda_n > 0$ for $n \geqq 0$, $\mu_n > 0$, $n \geqq 1$. It is known that when $\lambda_n/\mu_n \to 0$ as $n \to \infty$, visits to high states become rare, and passage times $T_{0,n}$ from state 0 to state n converge in distribution to exponentiality with suitable scaling [11]. A related result for the convergence of the sojourn time $T_n{}^+ = T_{n,n+1}$ on the set of states $0 \leqq m \leqq n$ initiated by a downward transition from state $n+1$ is described by:

THEOREM 7.1.

$$\frac{\lambda_n}{\mu_n} \to 0 \Rightarrow \frac{T_n{}^+}{ET_n{}^+} \to_d \Gamma_1 .$$

PROOF. The rv $T_n{}^+$ has a completely monotonic density (see, e.g., [13]). Hence $T_n{}^+$ is a scale mixture $T_n{}^+ = \Gamma_1 W_n$ and from (4.4)

$$C_n = \left(\frac{\sigma^2}{\mu^2}\right)_{T_n} = 1 + 2\left(\frac{\sigma^2}{\mu^2}\right)_{W_n} . \tag{7.1}$$

We will show that C_n goes to one as $n \to \infty$, so that Theorem 7.1 follows from Theorem 5.2. From (1.5a) and (1.5b) of [10] we have

$$C_n - 1 = \frac{\mu_n}{\lambda_n} \frac{E^2 T_{n-1}^+}{E^2 T_n{}^+} (C_{n-1} + 1) \leqq \frac{\lambda_n}{\mu_n} (C_{n-1} + 1) .$$

One then has

$$\frac{C_n + 1}{2} \leqq 1 + \gamma_n + (\gamma_n \gamma_{n-1}) \cdots + (\gamma_n \gamma_{n-1} \cdots \gamma_1) \tag{7.2}$$

where $\gamma_n = \lambda_n / \mu_n$. It follows from Lemma 3.1 of [11], that $(C_n + 1)/2 \to 1$ as $n \to \infty$ so that $C_n \to 1$ as needed.

For a given birth-death process, the quality of the asymptotic approximation is available directly from (7.2) which provides an upper bound for our distance $(C_n + 1)/2 = (\sigma^2/\mu^2)_{W_n}$ from exponentiality. Explicit evaluation of $(\sigma^2/\mu^2)_{W_n}$ is available from (1.5a) and (1.5b) of [10].

It should be noted that when $\lambda_n/\mu_n \to 0$, one also has T_n^- the first time down from state n converging to exponentiality. Here the argument is much simpler. Indeed, as shown in [10] (cf. (1.1) and (1.9)), T_n^- is a mixture of an exponential random variable with weight $\mu_n/(\lambda_n + \mu_n)$ and a second positive random variable. The resul then follows.

Theorem 7.1 could also have been obtained from a more general theorem for rare events [12], and the formalism of [10]. The quality of the exponential approximation for finite n, however, would not have been available.

8. The rate of convergence to normality in the central limit theorem. We have seen (Section 2) that a mixture of normals X, i.e., an element of $\mathscr{L}_2$, may be represented either in the form $X = NW$ or the form $X = NV^{\frac{1}{2}}$. Two families of metrics were seen in Section 5 to be of interest. The first given in (5.1) is defined in terms of a metric on the space of df's of W. For the parameter value $\theta = 2$, the distance to normality with $W = W_0 = 1$ took the form (6.2)

$$\rho_{W2}(F_X, F_{\mu_W N}) = \sigma_W{}^2/\mu_W{}^2 \,. \tag{8.1}$$

An alternate metric defined in terms of the random variable V was given in (5.2). When the parameter $\theta = 2$, this becomes (cf. (6.1)), for the subspace with $EV = \operatorname{Var} X = 1$ (see below)

$$\rho_{V2}(F_X, F_N) = \sigma_V{}^2/\mu_V{}^2 \,. \tag{8.2}$$

The distance ρ_{V2} may be extended in the same way as ρ_{W2} (cf. (6.2)) to include all $X \in \mathscr{L}_2$ having a finite fourth moment. We define

$$\rho_{V2}(F_X, F_{\sigma_X N}) = \rho_{V2}(F_{X/\sigma_X}, F_N) = \sigma_V{}^2/\mu_V{}^2 \,, \tag{8.3}$$

which can be used as a measure of distance to normality for all $X \in \mathscr{L}_2$ having a finite fourth moment.

The metric (8.1) is more natural than (8.2) in that it deals directly with the mixing rv W. Moreover, the distance $\sigma_W{}^2/\mu_W{}^2$ is available whenever F_X has a second moment since $EX^2 = EN^2 EW^2 = EW^2$, and is more appropriate to the central limit theorem where the second moment is intrinsic. The metric (8.1), however, is not a convenient tool for discussing the rate of convergence to normality since for independent convolution one does not have $X_1 + X_2 \leftrightarrow W_1 + W_2$.

For the representation $X = NV^{\frac{1}{2}}$ we do have $X_1 + X_2 \leftrightarrow V_1 + V_2$ as we have seen in Section 2. For this reason the metric (8.2) is attractive. The requirement that X have a fourth moment to permit V to have a variance is a mild handicap in practice.

We note that if $X = NV^{\frac{1}{2}}$, $EX^2 = \sigma_X^2 = EV$, i.e., distributions in $\mathscr{L}_2$ with finite second moments have finite first moments for their V distributions. As a direct consequence, one finds a simple correspondence between the central limit theorem in $\mathscr{L}_2$ and the weak law of large numbers.

THEOREM 8.1. *Let $(X_j)_1^\infty$ be a sequence of independent random variables in $\mathscr{L}_2$ with $\sigma_j^2 = EX_j^2 < \infty$. Let $(V_j)_1^\infty$ be any corresponding sequence of independent* rv's *with $X_j =_d N_j V_j^{\frac{1}{2}}$. Then* (cf. (2.11))

$$\frac{S_K}{\sigma[S_K]} = \frac{\sum_1^K X_j}{(\sum_1^K \sigma_j^2)^{\frac{1}{2}}} \to_d N \Leftrightarrow \frac{T_K}{ET_K} = \frac{\sum_1^K V_j}{E[\sum_1^K V_j]} \to_d \delta(1)$$

where $\delta(1)$ has all support at unity.

PROOF. From the representation $X = NV^{\frac{1}{2}}$, and the closure of $\mathscr{L}_2$ under convolution we have

$$\frac{S_K}{\sigma[S_K]} \leftrightarrow \frac{\sum_1^K V_j}{\sigma^2[S_K]} = \frac{\sum_1^K V_j}{E[\sum_1^K V_j]} = \frac{T_K}{E[T_K]}. \tag{8.4}$$

Also one has for corresponding sequences $(X_n)_1^\infty$ and $(V_n)_1^\infty$

$$\phi_{X_n}(t) = \int_0^\infty e^{-t^2 v/2}\, dF_{V_n}(v). \tag{8.5}$$

$T_K/ET_K \to_d \delta(1)$ then implies that $S_K/\sigma[S_K] \to_d N$ and conversely. This may be seen, for example, from the continuity theorem for Laplace transforms and the related uniqueness theorem, since $\phi_X((2s)^{\frac{1}{2}})$ is a Laplace transform when X is in $\mathscr{L}_2$ (cf. [5] XIII).

The structure of the metric (8.2) and its simple relation to convolution imply that in the class $\mathscr{L}_2$ the addition of any independent increment X_2 to $X_1 \neq 0$ when X_2 is closer to normality than X_1 brings $X_1 + X_2$ closer to normality. Indeed, one has

$$\left(\frac{\sigma^2}{\mu^2}\right)_{V_1+V_2} = \left(\frac{\sigma^2}{\mu^2}\right)_{V_1} \left(\frac{\mu_{V1}}{\mu_{V1} + \mu_{V2}}\right)^2 + \left(\frac{\sigma^2}{\mu^2}\right)_{V_2} \left(\frac{\mu_{V2}}{\mu_{V1} + \mu_{V2}}\right)^2 \leqq \left(\frac{\sigma^2}{\mu^2}\right)_{V_1}. \tag{8.6}$$

Moreover, for sums of independent identically distributed rv's X_j in $\mathscr{L}_2$, the normality increases (strictly) with the number of summands, the sole exception being when the summands are normal. This is a direct consequence of

$$\left(\frac{\sigma^2}{\mu^2}\right)_{\sum_1^K V_j} = \frac{1}{K}\left(\frac{\sigma^2}{\mu^2}\right)_{V_1}. \tag{8.7}$$

It should also be noted that $(\sigma^2/\mu^2)_{V^{\frac{1}{2}}} \leqq (\sigma^2/\mu^2)_V$, a special case of Proposition 4.3. This implies that $\rho_{W2}(F_X, F_N)$ the distance corresponding to (8.1) tends to decrease when $\rho_{V2}(F_X, F_N)$ decreases. The monotonicity in the norm ρ_{W2} in cases of interest such as those above is not clear.

The Class $\mathscr{D}$. In Propositions 2.4 and 2.5 we introduced implicitly and displayed the extent of a class of distributions which we now define formally.

DEFINITION. An rv Z will be said to belong to the class $\mathscr{D}$ if the convolution of the distribution of Z with that of its dual $Z^D =_d -Z$ is a mixture of normals, i.e., if $Z + Z^D \in \mathscr{L}_2$.

We note that $\mathscr{L}_2 \subset \mathscr{D}$, and that the completely monotone class $\mathscr{E}_1 \subset \mathscr{D}$ (cf. Proposition 2.5). $\mathscr{D}$ is clearly closed under convolution but is not closed under mixing. If $Z \in \mathscr{D}$ then $Z + C_1 \in \mathscr{D}$ and $C_2 Z \in \mathscr{D}$ for all real constants C_1, C_2. The metric ρ_{2V} provides a convenient metric for the set of df's in $\mathscr{D}$. Let

$$U = \frac{Z + Z^D}{2^{\frac{1}{2}}} . \tag{8.8}$$

The $2^{\frac{1}{2}}$ factor is introduced so that if $Z \in \mathscr{L}_2$ with $\sigma_Z{}^2 = 1$, then $U \in \mathscr{L}_2$ with $\sigma_U{}^2 = 1$, and $Z = N \Rightarrow U = N$. Let Z_1, Z_2 be elements of $\mathscr{D}$. We define the metric

$$\rho^+(F_{Z_1}, F_{Z_2}) = \rho_{V2}(F_{U_1}, F_{U_2}) \tag{8.9}$$

where ρ_{V2} is the metric (5.2) with $\theta = 2$ for elements of $\mathscr{L}_2$ with unit variance, and U_j is related to Z_j via (8.8).

THEOREM 8.2. *The subspace of $\mathscr{D}$ with $\sigma_Z{}^2 = 1$ and distance function* (8.9) *is a quasimetric space* [2], *i.e., one has*

$$\rho^+(F_Z, F_Z) = 0 , \quad \textit{and the triangle inequality} \tag{8.10a}$$

$$\rho^+(F_{Z_1}, F_{Z_3}) \leqq \rho^+(F_{Z_1}, F_{Z_2}) + \rho^+(F_{Z_2}, F_{Z_3}) . \tag{8.10b}$$

PROOF. (8.10a) is trivial. (8.10b) follows form $\rho^+(F_{Z_1}, F_{Z_3}) = \rho_{V2}(F_{U_1}, F_{U_3}) \leqq \rho_{V2}(F_{U_1}, F_{U_2}) + \rho_{V2}(F_{U_2}, F_{U_3}) = \rho^+(F_{Z_1}, F_{Z_2}) + \rho^+(F_{Z_2}, F_{Z_3})$.

Quasimetric spaces differ from metric spaces only in that $\rho^+(x_1, x_2) = 0$ does not imply that $x_1 = x_2$. This is clearly in keeping with ρ^+ as a measure of distance between distribution forms, with the measure undisturbed when Z_1 or Z_2 is shifted or multiplied by minus one. For departure from normality one has

$$\rho^+(F_Z, F_N) = \frac{\sigma_V{}^2}{\mu_V{}^2} \tag{8.11}$$

where V is the mixing distribution associated with the U obtained from Z.

By the extension of the distance ρ_{V2} (cf. (8.3)) the distance ρ^+ can be extended as follows

$$\rho^+(F_Z, F_{\sigma_Z N}) = \rho^+(F_{Z/\sigma_Z}, F_N) . \tag{8.12}$$

In this way we obtain a distance from normality for all $X \in \mathscr{D}$ with $EX^4 < \infty$.

We will now show that this distance from normality is equal to $\gamma_2/6$ where γ_2 is the classical Pearson coefficient of kurtosis. We have $U = NV^{\frac{1}{2}}$ so that

$$(8.13) \qquad \frac{\sigma_V^2}{\mu_V^2} = \frac{\frac{1}{3}EU^4 - E^2U^2}{E^2U^2} = \frac{1}{6}\left[\frac{E\{Z-\mu_Z\}^4 - 3E^2\{Z-\mu_Z\}^2}{E^2\{Z-\mu_Z\}^2}\right].$$

In summary, we have

THEOREM 8.3. *The distance to normality of a random variable* $Z \in \mathscr{D}$ *in the metric space of Theorem* 8.2 *is* $\gamma_2/6$ *where* γ_2 *is the classical Pearson coefficient of kurtosis.*

Acknowledgment. The authors wish to thank W. J. Hall, R. M. Loynes and G. S. Mudholkar for helpful comments.

REFERENCES

[1] BARLOW, R. E., MARSHALL, A. W. and PROSCHAN, F. (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.* **34** 375–389.

[2] COLLATZ, L. (1966). *Functional Analysis and Numerical Mathematics.* Academic Press, New York.

[3] CRAMÉR, H. and LEADBETTFR, M. R. (1967). *Stationary and Related Stochastic Processes.* Wiley, New York.

[4] VAN DANTZIG, D. (1951). Une nouvelle généralisation de l'inégalité de Bienaymé. *Ann. Inst. H. Poincaré* **12** 31–43.

[5] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

[6] HARKNESS, W. L. and SHANTARAM, R. (1969). Convergence of a sequence of transformations of distribution functions. *Pacific. J. Math.* **31** 403–415.

[7] JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions* **1**, **2**. Houghton Mifflin, Boston.

[8] KARLIN, S. (1968). *Total Positivity.* Stanford Univ. Press.

[9] KARLIN, S., PROSCHAN, F. and BARLOW, R. E. (1971). Moment inequalities of Pólya frequency functions. *Theor. Probabiliry Appl.* **8** 391–398.

[10] KEILSON, J., (1965). A review of transient behavior in regular diffusion and birth-death processes, Part II. *J. Appl. Probability* **2** 405–428.

[11] KEILSON, J., (1966a). A technique for discussing the passage time distributions for stable systems. *J. Roy. Statist. Soc. Ser. B* **28** 477–486.

[12] KEILSON, J. (1966b). A limit theorem for passage times in ergodic regenerative processes. *Ann. Math. Statist.* **37** 866–870.

[13] KEILSON, J. (1971). Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes. *J. Appl. Probability* **8** 391–398.

[14] KEILSON, J. and STEUTEL, F. W. (1972). Families of infinitely divisible distributions closed under mixing and convolution. *Ann. Math. Statist.* **43** 242–250.

[15] KELKER, D., (1971). Infinite divisibility and variance mixtures of the normal distribuion *Ann. Math. Statist.* **42** 802–808.

[16] KEMPERMAN, J. H. B. (1971). Moment problems with convexity conditions I. in *Optimizing Methods in Statistics*, ed. J. S. Rustagi. Academic Press, New York, 115–178.

[17] LOÈVE, M. (1963). *Probability Theory.* Van Nostrand, Princeton.

[18] LUKACS, E. (1970). *Characteristic Functions*, 2nd ed. Griffin, London.

[19] STEUTEL, F. W. (1969). Note on completely monotone densities. *Ann. Math. Statist.* **4** 1130–1131.

[20] STEUTEL, F. W. (1970). Preservation of infinite divisibility under mixing and related topics. *Math. Centre Tracts* **33**, Math. Centre, Amsterdam.

JULIAN KEILSON
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

F. W. STEUTEL
AFDELING WISKUNDE
TECHNISCHE HOGESCHOOL TWENTE
ENSCHEDE, NETHERLANDS