the slope $\beta(x)$. However, our squared correlation curve $\rho^2(x)$ *does* have an interpretation as a *local* coefficient of determination. To see this, consider Model 2 and suppose that $\mu(x)$ is smooth so that near the point x_0, $\mu(x)$ is nearly linear with slope $\beta(x_0)$. Then for X close to x_0 we can, to a close approximation, write

$$Y \simeq Y_0 = \alpha_0 + \beta(x_0)X + \sigma(x_0)\varepsilon.$$

For Y_0 the coefficient of determination is

$$1 - \frac{Var(Y_0 | X = x)}{Var(Y_0)} = 1 - \frac{\sigma^2(x_0)}{\beta^2(x_0)\sigma_1^2 + \sigma^2(x_0)} = \rho^2(x_0).$$

Thus the correlation curve $\rho^2(x)$ is a local measure of the variance explained by regression in the sense that it is the coefficient of determination for the locally linear approximation Y_0 to Y.

REFERENCES

[1] P. J. Bickel, *Using residuals robustly I: tests for heteroscedasticity, nonlinearity*, Ann. Statist., 6 (1978), pp. 266–291.

[2] R. J. Carroll and D. Ruppert, *Robust estimation in heteroscedastic linear models*, Ann. Statist., 10 (1982), pp. 429–441.

[3] ———, *Transformation and Weighting in Regression*, Chapman and Hall, New York, London, 1988.

[4] F. Galton, *Correlations and their measurement, chiefly from anthropometric data*, Proc. Royal Soc., London, 45 (1988), pp. 135–145.

[5] T. Gasser and H. G. Müller, *Estimating regression functions and their derivatives with kernel methods*, Scand. J. Statist., 11 (1984), pp. 171–185.

[6] P. Hall and R. J. Carroll, *Variance function estimation in regression*, J. R. Statist. Soc., Ser. B, 51 (1989), pp. 3–14.

[7] S. M. Stigler, *The History of Statistics: The Measurement of Uncertainty before* 1900, Belknap Press, Cambridge, Mass., 1986.

[8] ———, *Galton's account of the invention of correlation*, Statist. Science., 4 (1989), pp. 73–86.

FUSIONS OF A PROBABILITY DISTRIBUTION: PRELIMINARY ANNOUNCEMENT*

J. ELTON† AND T. P. HILL†

The purpose of this note is to introduce the notion of a *fusion* of a probability distribution P, and to announce a number of results concerning class properties of fusions and their relationship to classical probabilistic concepts such as convex domination, majorization, martingalizability, and dilation. Proofs of these results will appear elsewhere (see [4] and [5]).

Throughout this paper, X will denote either a separable Banach space or a compact metrizable convex subset of a locally convex topological vector space (l.c.t.v.s.), and X^* will denote the dual space of continuous linear functionals (restricted to X in the latter case). For a subset A of X, I_A is the indicator function of A, A^c the complement of $A, co(A)$ the convex hull of A, $\overline{A}$, and $\mathring{A}$ the respective closure and interior of A and ∂A is the boundary $\overline{A} \setminus \mathring{A}$ of A. A sequence x_n in X converges weakly to x (written $x_n \xrightarrow{w} x$) if $f(x_n) \to f(x)$ for

*This research was partially supported by National Science Foundation grants DMS-87-01691, DMS-89-01267, and DMS-89-14344, and a Fulbright Research Grant.

†Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia 30332.

all $f \in X^*$, and converges strongly to $x(x_n \to x)$ if x_n converges to x in the strong topology. If X is normed, $\|x\|$ will denote the norm of x.

$\mathcal{B}$ will denote the Borel subsets of X, $\mathcal{P}$ the set of Borel probability measures on $(X, \mathcal{B})$, $\delta(x) \in \mathcal{P}$ the Dirac delta measure on $\{x\}$ (single atom of mass 1 at x), B^n the Borel subsets of Euclidean n-space R^n, and for $P \in \mathcal{P}$, supp P is the support of P. For $A \in \mathcal{B}$, $P|_A$ is defined by $P|_A(B) = P(A \cap B)$. $\{P_n\}$ converges weakly to $P\,(P_n \xrightarrow{w} P)$ means the usual weak convergence of measures in the sense of Billingsley [2]. Throughout this paper, P will always denote an element of $\mathcal{P}$, i.e., a Borel probability measure on $(X, \mathcal{B})$ and $\mathcal{L}(Y) \in \mathcal{P}$ is the distribution of the X-valued random vector Y. Let $A \in \mathcal{B}$ and $P \in \mathcal{P}$. If X is a separable Banach space, we say that A *has finite first P-moment* if $\int_A \|x\| dP(x) < \infty$; and if X is a compact metrizable convex subset of a l.c.t.v.s., then A will always be said to have a finite P-moment. For Q and $P \in \mathcal{P}$, P *convexly dominates* Q (written $P \overset{c}{\succeq} Q$) if $\int \varphi dP \leqq \int \varphi dQ$ for all continuous convex functions φ for which both integrals exist.

PROPOSITION 1. *If $P(A) > 0$ and A has a finite first P-moment, then there is a unique element $b = b(A; P) \in \overline{co}(A)$, called the P-barycenter of A, satisfying $f(b) = (1/P(A)) \int_A f dP$ $\forall f \in X^*$.*

DEFINITION. Measure $Q \in \mathcal{P}$ is an *elementary fusion* of P if there is an $A \in \mathcal{B}$ with a finite first P-moment, and $t \in [0, 1]$ such that Q is given by $dQ = I_{A^c} dP + tP(A) d\delta(b(A, P)) + (1 - t) I_A dP$. Measure Q is a *simple fusion* of P if there exists a positive integer n and probabilities $\{P_j\}_{j=0}^n \in \mathcal{P}$ satisfying $P_0 = P$, $P_n = Q$ and P_{j+1} is an elementary fusion of P_j for each $j = 0, \ldots, n-1$. (In other words, simple fusions are just finite compositions of elementary fusions.) $\mathcal{S}(P)$ will denote the class of simple fusions of P. Q is a *fusion* of P if there exists $\{P_n\}_{n=1}^\infty \in \mathcal{S}(P)$ satisfying $P_n \xrightarrow{w} P$; and $\mathcal{F}(P)$ denotes the class of all fusions of P. That is, $\mathcal{F}(P)$ is the weak closure of the set of finite compositions of elementary fusions of P.

Intuitively, an elementary fusion simply takes part (a fraction "t") of the mass of a set A and collapses it to the barycenter of A, thereby creating (or enlarging) an atom at that point, and decreasing proportionately the measure of A elsewhere. As is the case in defining the basic building blocks (indicator functions) of measurable functions, where it is usually possible to restrict from general measurable sets to a much smaller class (e.g., to dyadic open intervals in the R^1 framework), it is also the case that in defining these basic building blocks (elementary fusions) of fusions, it is possible to restrict oneself to much smaller classes of sets, for example to relatively compact or bounded sets. However, the elementary fusions here will be taken to be the general ones (via sets with finite first P-moments). Note that, by definition, P is an elementary fusion of itself (intuitively, fuse nothing, and the result is P).

Example 1. Let P be a purely atomic measure with exactly two atoms of mass p and $1 - p$ at points α_1 and α_2, respectively. Then $\mathcal{F}(P) = \{Q \in \mathcal{P}$: *supp* $Q \subseteqq [\alpha_1, \alpha_2]$, and $b(X; Q) = \alpha_1 p + \alpha_2(1-p)\}$.

Example 2. Let $X = R^1$ and P be the Cauchy distribution. Then $\mathcal{F}(P) = \mathcal{P}$.

Example 3. Let $X = C[0, 1]$, the Banach space of continuous real-valued functions on $[0, 1]$ equipped with the sup norm, let P be Wiener measure on X, and let A be the complement of the unit ball $\{x \in X: \|x\| \leqq 1\}$ in X. If Q is the elementary fusion of P formed by fusing all of A (i.e., $t = 1$), then Q is the distribution of a real-valued stochastic process starting at zero, which with probability $P(A)$ never leaves zero, and with probability $1 - P(A)$ looks like Brownian motion conditioned so that all sample paths remain in the interval $[-1, 1]$.

THEOREM 1. *$\mathcal{S}(P)$ and $\mathcal{F}(P)$ are convex.*

THEOREM 2. *If $Q \in \mathcal{F}(P)$ and $R \in \mathcal{F}(Q)$, then $R \in \mathcal{F}(P)$. (That is, the fusion ordering is transitive.)*

The next result which can be viewed as a generalization of Jensen's inequality, says that if Q is a fusion of P, then Q is convexly dominated by P, but the converse is in general not true (for an example, see [4]).

THEOREM 3. *If $Q \in \mathcal{F}(P)$, then $Q \overset{c}{\preceq} P$.*

THEOREM 4. *If P is a Borel probability measure on a separable Banach space X such that P has a finite first moment (i.e., $\int \|x\| dP < \infty$), then $\mathcal{F}(P)$ is tight. Moreover, if X is* *finite-dimensional, then P has a finite first moment if and only if $\mathcal{F}(P)$ is tight.*

It is easy to see from Example 2 that if both $Q \in \mathcal{F}(P)$ and $P \in \mathcal{F}(Q)$, it still may happen that $P \neq Q$. If, however, the measures have finite first moments, this cannot happen.

THEOREM 5. *If $Q \in \mathcal{F}(P)$ and $P \in \mathcal{F}(Q)$, and either has a finite first moment, then $P = Q$.*

THEOREM 6. *If P and Q are Borel probability measures on X, where X is a separable Banach space or a compact metrizable convex subset of a locally convex topological vector space, and if P has a finite first moment, then the following are equivalent:*

(i) *Q is a fusion of P;*

(ii) $Q \overset{c}{\preceq} P$;

(iii) *(Q, P) is martingalizable; and*

(iv) *there exists a dilation μ of X with $P = \mu Q$.*

Remarks. The equivalences of (ii), (iii), and (iv), assuming P has a finite barycenter, have been proved in part by Hardy, Littlewood, and Pólya for one-dimensional spaces, by Blackwell [1], Stein, and Sherman for finite-dimensional spaces, and Cartier, Fell, and Meyer [3] and Strassen [9] in various infinite-dimensional settings (see [7]). The main contribution here is the equivalence of (i) with (ii)–(iv).

THEOREM 7. *If P is a Borel probability measure on a separable Banach space, then P has a finite first moment if and only if $\mathcal{F}(P)$ is uniformly integrable. More generally, if φ: $[0, \infty) \to [0,\infty)$ is convex, nonconstant and nondecreasing, then $\int \varphi(\|x\|)dP < \infty$ if and only if $\mathcal{F}(p)$ is uniformly φ-integrable.*

If $X = R^1$ and P has a finite first moment, a number of additional conditions are known to be equivalent to fusions; the next theorem lists some of these. Recall that the *Hardy–Littlewood maximal function* H_P of P is $H_P := (1/(1-t)) \int_t^1 F^{-1}(s)ds$ for $0 \leqq t \leqq 1$ (where F^{-1} is the generalized inverse distribution function of P given by $F^{-1}(s) = inf\{x\colon P(-\infty, x] > s\}$, for $s \in [0, 1]$), and the *potential function* U_P of P is $U_P(t) = -\int |x - t| P(dx)$ (see [11] for properties and applications of these functions). Also, $Q \in \mathcal{P}$ is said to be *smaller in mean residual life* than P if $\int (x-t)^+ Q(dx) \leqq \int (x-t)^+ P(dx)$ for all real t. (This ordering has applications in queueing theory; see [8].)

THEOREM 8. *If $X = R^1$ and $P \in \mathcal{P}$ has a finite first moment, then the following are equivalent:*

(i) *Q is a fusion of P;*

(ii) $Q \overset{c}{\preceq} P$;

(iii) *(Q, P) is martingalizable;*

(iv) *there exists a dilation μ of X with $P = \mu Q$;*

(v) $H_Q \leqq H_P$, *and* $b(X, P) = b(X, Q)$;

(vi) $U_Q \geqq U_p$, *and* $b(X, P) = b(X, Q)$;

(vii) *Q is smaller in mean residual life than P and* $b(X, P) = b(X, Q)$;

(viii) $\int (x \vee t) Q(dx) \leqq \int (x \vee t) P(dx)$ *for all t, and* $b(X, P) = b(X, Q)$;

(ix) $\int_{-\infty}^t Q(-\infty, t]\, dt \leqq \int_{-\infty}^t P(-\infty, t]\, dt$ *for all t, and* $b(X, P) = b(X, Q)$.

DEFINITION. For a Borel probability measure P with a finite first moment on a separable Banach space, the *characteristic* of P, r_p, is the function r_P: $[0, \infty) \to [0,\infty)$ given by

$$r_P(\lambda) = \begin{cases} \int_{\|x\| \geqq \lambda} \|x\|\, dP / P\big(\|x\| \geqq \lambda\big) & if \quad P\big(\|x\| \geqq \lambda\big) > 0 \\ \lambda & if \quad P\big(\|x\| \geqq \lambda\big) = 0. \end{cases}$$

THEOREM 9. *Let $(Z_1, Z_2, \ldots, Z_n)$ be a non-negative submartingale on a probability space $(\Omega, \mathcal{A}, P)$. Then*

$$\int_{Z_1 \geqq r_p(\lambda)} Z_1 dP \leqq \int_{Z_n \geqq \lambda} Z_n dP \quad for\ all \quad \lambda \geqq 0.$$

An application. Many ideal physical laws describe linear mixtures or fusions of various types; one such law is the law of mixtures of concentrations, and another is Raoult's Law of physical chemistry — "the vapor pressure of the component of an ideal solution is proportional to the mole factor of the component."

Suppose x represents a variable "quality" (such as concentration, or vapor pressure) of a substance which mixes linearly, and further suppose that it costs $c(x)$ to produce one unit of quality x, which then may be sold for $r(x)$. Which distribution should production of this substance follow, and how should it then be mixed in order to maximize the average profit? In other words, if production is according to distribution P, and P is then fused to Q, what are the choices for P and for $Q \in \mathcal{F}(P)$, which will maximize the average profit $\int r dQ - \int c dP$? (For the application below, X is a compact convex subset of R^n.)

DEFINITION. For Borel functions $r, c \colon X \to R$, (Q, P) is (r, c)-optimal if $\int r dQ - \int c dP = sup\{\int r d\widehat{Q} - \int c d\widehat{P} \colon \widehat{P} \in \mathcal{P},\ \widehat{Q} \in \mathcal{F}(\widehat{P})\}$.

DEFINITION. For a function $f \colon X \to R$, let $\check{f}$ denote the *convex closure* of f, that is, $\check{f}(x) = sup\{g(x) | g \colon X \to R,\ g \text{ is convex, and } g \leqq f\}$.

THEOREM 10. *Suppose* $r, c \colon X \to R$ *are upper and lower semicontinuous, respectively. Then* $(\delta(x^*), \sum_{j=1}^{n+1} \alpha_j \delta(x_j^*))$ *is* (r, c)*-optimal,where:*x^* *is any point in* X *satisfying* $r(x^*) - \check{c}(x^*) = max\{r(x) - \check{c}(x) \colon x \in X\}$; *and* $\{(x_j^*, \check{c}(x_j^*))\}_{j=1}^n$ *are any extreme points of the convex set* $\{(x, y) \in R^{n+1} \colon x \in X,\ y \in R,\ y \geqq \check{c}(x)\}$ *which satisfy* $\sum_{j=1}^{n+1} \alpha_j (x_j^*, \check{c}(x_j^*)) = (x^*, \check{c}(x^*))$ *for some* $\{\alpha_j\}_1^{n+1} \geqq 0$, $\sum_{j=1}^{n+1} \alpha_j = 1$.

Remark. It has recently been shown by the authors that if P and Q are finite (i.e., not necessarily probability) measures on X (where again X is a separable Banach space or a compact metrizable convex subset of a l.c.t.v.s) then $\int \varphi dQ \leqq \int \varphi dP$ for all *non-negative* continuous convex functions if and only if there is a fusion $\widehat{P}$ of P which majorizes Q. This result is new even in the finite-dimensional case, and the proofs use a new geometric argument similar in spirit to those of Hardy, Littlewood, and Polya.

Acknowledgment. The authors are grateful to the National Science Foundation and to the Georgia Tech Foundation for travel grants to the second author to present these results at the Second World Congress of the Bernoulli Society in Uppsala in August 1990.

REFERENCES

[1] D. BLACKWELL, *Equivalent comparisons of experiments*, Ann. Math. Stat., 24 (1953), pp. 265–272.

[2] P. BILLINGSLEY, *Convergence of Probability Measures*, John Wiley, New York, 1968.

[3] P. CARTIER, J. FELL, AND P. MEYER, *Comparison des mesures porte et par un ensemble convexe compact*, Bull. Soc. Math. Fr., 92 (1964), pp. 435–445.

[4] J. ELTON AND T. HILL, *Fusions of a probability distribution*, Annals of Probability, to appear.

[5] ———, *Convex domination,dilation,and fusion of general finite measures*, in preparation.

[6] A. MARSHALL AND I. OLKIN, *Inequalities:Theory of Majorization and its Applications*, Academic Press, New York, 1979.

[7] R. PHELPS, *Lectures on Choquet's Theorem*, Van Nostrand, New York, 1966.

[8] D. STOYAN, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley, New York, 1983.

[9] V. STRASSEN, *The existence of probability measures with given marginals*, Ann. Math. Stat., 36, pp. 423–439.

[10] Y. TONG, *Probability Inequalities in Multivariate Distributions*, Academic Press, New York, 1980.

[11] D. VAN DER VECHT, *Inequalities for Stopped Brownian Motion*, CWI Tract 21, Amsterdam, 1986.