FUSIONS OF A PROBABILITY DISTRIBUTION

By J. Elton and T. P. Hill[1]

Georgia Institute of Technology

Starting with a Borel probability measure P on X (where X is a separable Banach space or a compact metrizable convex subset of a locally convex topological vector space), the class $\mathscr{F}(P)$, called the fusions of P, consists of all Borel probability measures on X which can be obtained from P by fusing parts of the mass of P, that is, by collapsing parts of the mass of P to their respective barycenters. The class $\mathscr{F}(P)$ is shown to be convex, and the ordering induced on the space of all Borel probability measures by $Q \preccurlyeq P$ if and only if $Q \in \mathscr{F}(P)$ is shown to be transitive and to imply the convex domination ordering. If P has a finite mean, then $\mathscr{F}(P)$ is uniformly integrable and $Q \preccurlyeq P$ is equivalent to Q convexly dominated by P and hence equivalent to the pair (Q, P) being martingalizable. These ideas are applied to obtain new martingale inequalities and a solution to a cost–reward problem concerning optimal fusions of a finite-dimensional distribution.

1. Introduction. The purpose of this paper is to introduce the notion of a *fusion* of a probability distribution P and to study class properties of fusions and their relationship to classical probabilistic concepts such as convex domination, majorization, martingalizability and dilation.

As a simple concrete example, suppose P is the purely atomic probability distribution with masses $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$ at α, β and γ, respectively. Thinking of P physically, such as the distribution of quantities of various concentrations of a liquid solution (e.g., P represents 1 unit of saline solution of concentration α, 2 units of concentration β and 3 units of concentration γ), it is clear that many other probability distributions may be obtained *irreversibly* from P by fusing parts of P. For example, if all of the components of P are mixed together, the resulting probability distribution is a single atom of mass 1 at the barycenter $(\alpha + 2\beta + 3\gamma)/6$; or, if only half of the α-atom is fused with half the β-atom, the resulting distribution is again purely atomic, but with atoms of masses $\frac{1}{12}$, $\frac{1}{6}$, $\frac{1}{2}$ and $\frac{1}{4}$ at α, β, γ and $(\alpha + 2\beta)/3$, respectively. Each fusion may itself be further fused, resulting in still another distribution. What is the class of all fusions that may be obtained as limits of repeated fusings of a given general distribution, and what properties does this class have? This paper will address these questions in the general settings where P is a Borel probability measure on a separable Banach space, or on a compact metrizable convex

subset of a locally convex topological vector space (l.c.t.v.s.). Special attention is given to the real-valued and finite-dimensional cases, in which setting several new results concerning convex domination and dilations, an answer to a majorization question raised by Marshall and Olkin (1979) and a solution to an applied cost-reward fusion problem are obtained.

For the reader familiar with dilations and balayage, a fusion is almost an antibalayage. The point here is that in many physical experiments a balayage (or unfusing) is simply impossible; the natural process of fusion is an irreversible one and it is in this fusion direction that the action takes place.

2. Preliminaries. Throughout this paper, X will denote either a separable Banach space or a compact metrizable convex subset of a locally convex topological vector space 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 $\overset{\circ}{A}$ the closure and interior of A, respectively, and ∂A is the boundary $\overline{A} \setminus \overset{\circ}{A}$ of A. A sequence x_n in X converges weakly to x (written $x_n \to_w x$) if $f(x_n) \to f(x)$ for 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.

$\mathscr{B}$ will denote the Borel subsets of X, $\mathscr{P}$ the set of Borel probability measures on $(X, \mathscr{B})$, $\delta(x) \in \mathscr{P}$ the Dirac delta measure on $\{x\}$ (single atom of mass 1 at x), $\mathbb{B}^n$ the Borel subsets of Euclidean n-space $\mathbb{R}^n$ and for $P \in \mathscr{P}$, supp P is the support of P. For $A \in \mathscr{B}$, $P|_A$ is defined by $P|_A(B) = P(A \cap B)$. $\{P_n\}$ converges weakly to P ($P_n \to_w P$) means the usual weak convergence of measures in the sense of Billingsley (1968). Throughout this paper, P will always denote an element of $\mathscr{P}$, that is, a Borel probability measure on $(X, \mathscr{B})$ and $\mathscr{L}(Y) \in \mathscr{P}$ is the distribution of the X-valued random vector Y.

Let $A \in \mathscr{B}$ and $P \in \mathscr{P}$. If X is a separable Banach space, say that A has finite first P-moment if

$$(2.1) \qquad\qquad \int_A \|x\| \, dP(x) < \infty;$$

and if X is a compact metrizable convex subset of a l.c.t.v.s., A will always be said to have finite P-moment (this is to avoid having to state separate versions of the same definitions and results).

PROPOSITION 2.1. *If $P(A) > 0$ and A has a finite first P-moment, then there is a unique element $b = b(A, P) \in \overline{\text{co}}(A)$ satisfying*

$$(2.2) \qquad\qquad f(b) = \frac{1}{P(A)} \int_A f \, dP, \qquad \forall \, f \in X^*.$$

PROOF. In the case when X is a separable Banach space, condition (2.1) implies that the identity function on A is Bochner integrable, that is, $\int_A x \, dP(x)$ exists [cf. Diestel and Uhl (1977), page 45] and (2.2) then follows by the linearity of $f \in X^*$. If b were not in $\overline{\text{co}}(A)$, then by the separation version of

the Hahn–Banach theorem [cf. Rudin (1973), page 58] there would exist an $f \in X^*$ such that $f(b) < f(x)$ for all $x \in \overline{\text{co}}(A)$, which contradicts (2.2).

For the case when X is a compact metrizable convex subset of a l.c.t.v.s., see Phelps [(1966), Proposition 1.1, page 4]. □

DEFINITION 2.2. The element $b(A, P)$ in Proposition 2.1 is called the *P-barycenter of A*. Say that *P has a finite first moment* if $b(X, P)$ exists.

REMARKS. If X is finite-dimensional, then it is even true that $b(A, P) \in$ co(A), but in general the closure is needed for infinite-dimensional spaces, as can be seen by taking $X = l_1$, $A = \{e_1, e_2, \ldots\}$, the closed nonconvex subset consisting of all unit coordinate vectors, and P defined by $P(e_n) = 2^{-n}$, $n = 1, 2, \ldots$. Then co(A) is the set of all *finite* convex combinations of the $\{e_j\}$, so $b(A, P) = \sum_{n=1}^{\infty} 2^{-n} e_n \notin$ co(A).

In the infinite-dimensional cases, the assumptions of metrizability and separability are used to facilitate the discussion of weak convergence of measures, but these assumptions are not essential to most of the key ideas in this paper and may be eliminated by the interested reader.

3. Fusions of general probabilities. The main purpose of this section is to define formally the notion of a *fusion* of a probability, analogously to the way measurable functions are defined through indicator functions, simple functions and limits of simple functions; and then to prove several general properties of the class of all fusions of a given distribution.

DEFINITION 3.1. $Q \in \mathscr{P}$ is an *elementary fusion* of P if there is an $A \in \mathscr{B}$ with finite first P-moment and a $t \in [0, 1]$ such that Q is given by

$$Q = \begin{cases} P, & \text{if } tP(A) = 0, \\ P|_{A^c} + tP(A)\delta(b(A, P)) + (1 - t)P|_A, & \text{otherwise.} \end{cases}$$

(In alternative notation, $dQ = I_{A^c} dP + tP(A) d\delta(b(A, P)) + (1 - t)I_A dP$.)

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 $\mathbb{R}^1$ framework), it is also the case that in defining these basic building blocks (elementary fusions) of fusions, it is possible to restrict 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), and further restrictions to subclasses are left to the interested reader. Note that, by definition, P is an elementary fusion of itself (intuitively, fuse nothing, and the result is P).

EXAMPLE 3.2. If $X = \mathbb{R}^1$ and P is the exponential distribution with mean 1, then the whole space X has finite first P-moment, so the Dirac measure at the barycenter 1 is an elementary fusion of P. Another typical elementary fusion of P, formed by taking $A = (0, 5)$ and $t = \frac{1}{3}$, is the mixed (discrete-continuous) distribution with single atom of mass $(1 - e^{-5})/3$ at $(1 - 6e^{-5})/(1 - e^{-5})$, with density $2e^{-x}/3$ on $(0, 5)$ and with density e^{-x} on $(5, \infty)$.

EXAMPLE 3.3. Let $X = \mathbb{R}^1$ and let P be the Cauchy distribution. Then the whole space does not have a P-barycenter, but every bounded measurable subset of positive Lebesgue measure does. By taking A to be a set of the form $(0, \beta)$ with $\beta \gg 1$ and $t = 1$, it is possible to construct an elementary fusion Q of P with the following properties. Given $\varepsilon > 0$ and $N > 0$, Q coincides with P on $(-\infty, 0)$, has a single atom of mass $m \in (\frac{1}{2} - \varepsilon, \frac{1}{2})$ located at $b > N$ and coincides with P on (β, ∞). (This construction will be used later to show that without an assumption of finiteness of first moment, the fusion ordering on $\mathscr{P}$ may fail to be antisymmetric; see Proposition 3.14 and the remarks following it.)

EXAMPLE 3.4. 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 $\{x \in X \colon \|x\| \leq 1\}$ in X. Then A has a finite first P-moment, and since $P(A) > 0$, it follows that A has a P-barycenter $b \in C[0, 1]$. For fixed $s \in (0, 1]$, $x(s)$ is a normally distributed random variable with mean 0 and variance s. Letting $f_s \colon X \to \mathbb{R}$ be the projection $f_s(x) = x(s)$, it is clear by symmetry that $\int_A x(s)\, dP(x) = 0$, so (2.2) implies that $b(s) = 0$, hence $b \equiv 0$. If Q is the elementary fusion of P formed by taking A to be the complement of the unit ball and $t = 1$, 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 $[-1, 1]$.

Next, the elementary fusions will be generalized to the notion of simple fusions. As was the case in defining elementary fusions, there are at least several natural directions in which to proceed. First, the composition-generalization approach is taken, and then another useful approach (matrix simple fusions) is shown to be equivalent.

DEFINITION 3.5. Q is a *simple fusion* of P if there exists a positive integer n and probabilities $\{P_j\}_{j=0}^n \subset \mathscr{P}$ 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.) $\mathscr{S}(P)$ will denote the class of simple fusions of P. Q is a *fusion* of P if there exists $\{P_n\}_{n=1}^\infty \subset \mathscr{S}(P)$ satisfying $P_n \to_w P$; and $\mathscr{F}(P)$ denotes the class of all fusions of P. That is, $\mathscr{F}(P)$ is the weak closure of the set of finite compositions of elementary fusions of P.

For example, if P is purely atomic with exactly two atoms, then $\mathscr{A}(P)$ consists of all purely atomic distributions having the same barycenter as P and having only a finite number of atoms, each of which lies on the closed line segment connecting the two atoms of P, and $\mathscr{F}(P)$ consists of all Borel probability measures which have the same barycenter as P and which have support contained in the closed line segment connecting the two atoms (Proposition 3.13). If P is the Cauchy distribution on $\mathbb{R}^1$, then $\mathscr{F}(P)$ consists of *all* Borel probabilities on $\mathbb{R}^1$, that is, the Cauchy distribution can be fused to obtain *every* other distribution (Proposition 3.14). If P has finite first moment and X is one-dimensional, then the notion of fusion is equivalent to a number of classical partial orderings including convex domination, martingalizability, dilation, smaller-in-mean-residual-life and domination of the Hardy–Littlewood maximal functions and potential functions (Theorem 4.7). Although it is possible to prove these results directly from the definitions, it is much easier to establish an equivalent characterization of $\mathscr{F}(P)$, a characterization which will also facilitate the proof of the convexity of $\mathscr{A}(P)$ and $\mathscr{F}(P)$ (Theorem 3.11), the fact that a fusion of a fusion of P is itself a fusion of P [i.e., $\mathscr{F}(\mathscr{F}(P)) = \mathscr{F}(P)$] (Theorem 3.12) and the fact that if P has a finite barycenter, $Q \in \mathscr{F}(P)$ if and only if (Q, P) is martingalizable (Theorem 4.1).

The next main task is to show that $\mathscr{A}(P)$ is exactly the same as the set of matrix-simple fusions of P (Proposition 3.10). In what follows, Π_n is the set of ordered Borel n-partitions of X, that is, $\Pi_n = \{(A_i)_{i=1}^n : A_i \in \mathscr{B} \ \forall \ i, A_i \cap A_j = 0 \text{ if } i \neq j \text{ and } \bigcup_{i=1}^n A_i = X\}$.

DEFINITION 3.6. Q is a *matrix simple fusion* (m.s.f.) of P if $\exists \ n, k \in \mathbb{N}$, $(A_i)_{i=1}^n \in \Pi_n$ and a (nonnegative) row substochastic $n \times k$ matrix $(t_{ij})_{i=1, j=1}^{n, k}$, with $t_{ij} = 0$ if $b(A_i, P)$ does not exist, so that

$$(3.1) \qquad Q = \sum_{j=1}^k \left(\sum_{i=1}^n t_{ij} P(A_i) \right) \delta(b_j) + \sum_{i=1}^n \left(1 - \sum_{j=1}^k t_{ij} \right) P|_{A_i},$$

where $b_j = (\sum_{i=1}^n t_{ij} b(A_i, P) P(A_i) / \sum_{i=1}^n t_{ij} P(A_i))$ [with the convention that $\vec{0}/0 = \vec{0}$]. Notationally, such a m.s.f. Q will be written as

$$Q = \text{fus}\left((A_i)_{i=1}^n; (t_{ij})_{i=1, j=1}^{n, k}; P \right).$$

If $k = 1$, then Q is called a *column* m.s.f. of P and if $k = 1$, $n = 3$ and $t_{31} = 0$ [i.e., $Q = \text{fus}(A_1, A_2, (A_1 \cup A_2)^c; (t_1, t_2, 0); P)$], then Q is called a *binary* m.s.f. [Technically speaking, for column and binary m.s.f.'s, the fusion proportion matrix is a column vector and should be written as (t_1, t_2, t_3)-transpose, but as no ambiguity arises, it will be written for convenience as a row vector (t_1, t_2, t_3).]

Intuitively, to obtain a matrix fusion of P: Start with a partition of primary sets $A_1, \ldots, A_n$; fuse part of A_1, part of A_2 and so on all together to reduce the measures of $A_1, \ldots, A_n$ accordingly and add a single mass point at the weighted barycenter; then fuse part of the mass left in A_1, part of that in A_2 and so on all together to reduce the measures of the $\{A_j\}$ still further and add

a second single mass point at the new weighted barycenter; and continue this a total of k times. In general, the new measure Q has k new atoms where P had none.

In order to show that two fusions are the same, the following lemma will be used often; its proof is easy and is left to the reader.

LEMMA 3.7. *If Q_1 and $Q_2 \in \mathscr{P}$ have the same barycenter and if $Q_1|_{(a,\,b)^c} = Q_2|_{(a,\,b)^c}$ for some $a, b \in X$, then $Q_1 = Q_2$.*

Recall that, by definition, every simple fusion is the composition of finitely many elementary fusions. The analogous result for matrix simple fusions requires proof.

LEMMA 3.8. *Every matrix simple fusion is the composition of a finite number of binary m.s.f.'s, that is, if Q is an m.s.f. of P, then there exist $Q_1, \ldots, Q_n \in \mathscr{P}$ with $Q_1 = P$, $Q_n = Q$ and such that Q_{i+1} is a binary m.s.f. of Q_i for each $i = 1, \ldots, n-1$.*

PROOF. First it will be shown that every m.s.f. Q of P is the composition of a finite number of column m.s.f.'s. To see this, let $Q = \mathrm{fus}((A_i)_{i=1}^n; (t_{ij})_{i=1,\,j=1}^{n,\,k}; P)$ and assume without loss of generality that $\sum_{i=1}^n t_{ij} > 0$ for all $j = 1, \ldots, k$. If $k = 1$, the conclusion is trivial, since then Q is already a column m.s.f. of P, so assume $k > 1$. By induction, it is enough to show that there exists an $n \times (k-1)$ m.s.f. Q_1 of P and a column m.s.f. Q_2 of Q_1 so that $Q_2 = Q$. Without loss of generality, it may also be assumed that

$$(3.2) \qquad \sum_{i=1}^n P(A_i) t_{ij} > 0 \quad \text{for all } j = 1, \ldots, k.$$

Let $Q_1 = \mathrm{fus}((A_i)_{i=1}^n; (t_{ij})_{i=1,\,j=1}^{n,\,k-1}; P)$ and let $B = \{b_1, \ldots, b_{k-1}\}$, where $b_j = \sum_{i=1}^n t_{ij} b(A_i, P) P(A_i) / \sum_{i=1}^n P(A_i) t_{ij}$, which exists by (3.2) and the defining requirement of m.s.f. that $t_{ij} > 0$ only if $b(A_i, P)$ exists. By combining corresponding columns, it may further be assumed that the $\{b_j\}$ are distinct.

Let $Q_2 = \mathrm{fus}((\hat{A}_i)_{i=1}^{nk}; (\hat{t}_i)_{i=1}^{nk}; Q_1)$, where $(\hat{A}_1, \ldots, \hat{A}_{nk}) = ((A_1 \cap B^c), \ldots, (A_n \cap B^c), (A_1 \cap \{b_1\}), \ldots, (A_n \cap \{b_1\}), \ldots, (A_n \cap \{b_{k-1}\}))$ and

$$\hat{t}_i = \begin{cases} (t_{ik}) \Big/ \left[1 - \sum_{j=1}^{k-1} t_{ij} \right], & \text{for } i = 1, \ldots, n, \\[2ex] (t_{ik}) P(\hat{A}_i) \Big/ \left[P(\hat{A}_i) + \sum_{m=1}^n P(A_m) t_{mk} \right], & \text{for } i = n+1, \ldots, nk. \end{cases}$$

An easy calculation using (3.1) and Lemma 3.7 shows that $Q_2 = Q$, which completes the proof that every m.s.f. is the composition of a finite number of column m.s.f.'s.

To complete the proof of the lemma, it is now enough to show that every column m.s.f. is the composition of a finite number of binary m.s.f.'s. Let Q be

the column m.s.f. given by $Q = \text{fus}((A_i)_{i=1}^n; (t_i)_{i=1}^n; P)$ and then let $Q_1 = \text{fus}((A_i)_{i=1}^n; (t_1, \ldots, t_{n-1}, 0); P)$ and

$$Q_2 = \text{fus}\left(A_n \setminus \{b_1\}, \{b_1\}, (A_n \cup \{b_1\})^c; (t_n, t^*, 0); Q_1\right),$$

where

$$t^* = \frac{t_n P(\{b_1\} \cap A_n)}{P(\{b_1\} \cap A_n) + \sum_{i=1}^{n-1} t_i P(A_i \setminus \{b_1\})}$$

and

$$b_1 = \frac{\sum_{i=1}^{n-1} t_i P(A_i) b(A_i, P)}{\sum_{i=1}^{n-1} t_i P(A_i)}.$$

Using Lemma 3.7, it is easy to check that $Q_2 = Q$ and since Q_1 is a shorter column m.s.f. than Q, and Q_2 is a binary m.s.f. of Q_1, the proof is complete by induction. $\square$

The next lemma is a key step in the proof of most of the main results in this section. It yields an easy proof that the class of simple fusions and hence the class of all fusions (of a given probability) is convex, which in turn is used to show that a fusion of a fusion of P is itself a fusion of P. The fact that a binary m.s.f. is not always just a composition of two, or perhaps three, elementary fusions may seem counterintuitive at first, but examples suggest that in general five are needed. On the other hand, five elementary fusions always suffice, a fact which is not used in the subsequent results in this paper and which may be shown using an argument similar to but much more tedious than the proof of the following weaker result.

LEMMA 3.9. *Every binary m.s.f. of P is a simple fusion of P.*

PROOF. Fix $Q = \text{fus}(A_1, A_2, (A_1 \cup A_2)^c; (t_1, t_2, 0); P)$ and without loss of generality, assume $0 < t_2 < t_1 \leq 1$ (if $t_1 = 0$ or $t_1 = t_2$, then Q is already an elementary fusion of P).

First, suppose that $A_1 = \{a_1\}$ and $A_2 = \{a_2\}$, where $a_1 \neq a_2$ and $P(a_1)P(a_2) > 0$ [if $P(a_1)P(a_2) = 0$ the conclusion is trivial]. It will be shown that such a Q is always the composition of eight elementary fusions.

Let $b_0 = b(\{a_1, a_2\}, P)$. The proof will proceed by using elementary fusions to fuse small masses of P to barycenters $\{b_i\}_{i=1}^6$, each of which has P-measure zero and then to use these temporary atoms to fuse the desired proportions of the masses at a_1, b_0 and a_2 to obtain Q. (In the following, often x is written instead of $\{x\}$ for singleton sets.)

First, let

$$Q_1 = \text{fus}\left(a_1, a_2, \{a_1, a_2\}^c; (\varepsilon_1, \varepsilon_1, 0); P\right),$$

where ε_1 is chosen so that $0 < \varepsilon_1 < \min\{t_1, t_2\}$ and so that $b_1 = b_1(\varepsilon_1) = b(\{b_0, a_2\}, Q_1)$ (which is the point where the next elementary fusion Q_2 will

place its fused mass) satisfies $P(b_1) = 0$. This is possible since $Q_1(b_0)$ *increases* with ε_1 but $Q_1(a_2)$ *decreases* with ε_1, so b_1 moves (continuously and one-to-one) toward b_0 as ε_1 increases, along the line connecting a_2 and b_0; so for all but countably many ε_1, $P(b_1) = 0$.

(At this stage, only a small fraction of the mass desired to be fused from a_1 and a_2 has been fused and placed at b_0, which may already have positive P-measure. Later this must also be corrected.)

Next, let

$$Q_2 = \text{fus}\big(b_0, a_2, \{b_0, a_2\}^c; (\varepsilon_2, \varepsilon_2, 0); Q_1\big),$$

where ε_2 is chosen so that $Q_2(a_2) > (1 - t_2)P(a_2)$, $Q_2(b_0) > P(b_0)$ and, since $b_2 = b_2(\varepsilon_2) = b(\{b_1, a_2\}, Q_2)$ moves toward b_1 as ε_2 increases, ε_2 is also chosen so that $P(b_2) = 0$.

(At this stage, a temporary small mass has been placed at the P-massless point b_1, and a_1, b_0 and a_2 each still have strictly more mass than Q places at these points.)

Let

$$Q_3 = \text{fus}\big(b_1, a_2, \{b_1, a_2\}^c; (\varepsilon_3, \varepsilon_3, 0); Q_2\big),$$

where ε_3 is chosen so that $Q_3(a_2) > (1 - t_2)P(a_2)$ and so that $b_3 = b_3(\varepsilon_3) = b(\{a_1, b_1\}, Q_3)$ and $b_4 = b_4(\varepsilon_3) = b(\{b_2, a_2\}, Q_3)$ satisfy $P(b_3) = P(b_4) = 0$, which is possible since b_3 moves toward a_1 and b_4 toward b_2 as ε_3 increases and since the intersection of two co-countable sets is co-countable.

(There is still too much mass on a_1, b_0 and a_2. The next step will remove the desired amount from a_1 and place it at the P-massless point b_3.)

Let

$$Q_4 = \text{fus}\big(\{a_1, b_1\}, \{a_1, b_1\}^c; ((t_1 - \varepsilon_1)/(1 - \varepsilon_1), 0); Q_3\big),$$

so, in particular, $Q_4(a_1) = (1 - t_1)P(a_1) = Q(a_1)$.

Now the same strategy will be used to remove the desired masses from b_0 and a_2.

Let

$$Q_5 = \text{fus}\big(b_2, a_2, \{b_2, a_2\}^c; (\varepsilon_4, \varepsilon_4, 0); Q_4\big),$$

where ε_4 is chosen so that $Q_5(a_2) > (1 - t_2)P(a_2)$ and so that $b_5 = b_5(\varepsilon_4) = b(\{b_0, b_2\}, Q_5)$ and $b_6 = b_6(\varepsilon_4) = b(\{b_4, a_2\}, Q_5)$ satisfy $P(b_5) = P(b_6) = 0$, which is possible, as before, since b_5 moves toward b_0 and b_6 towards b_4 as ε_4 increases.

Let

$$Q_6 = \text{fus}\big(b_0, b_2, \{b_0, b_2\}^c; (t, t, 0); Q_5\big),$$

where $1 - t = P(b_0)([P(b_0) + \varepsilon_1 P(a_1) + \varepsilon_1 P(a_2)][1 - \varepsilon_2])^{-1}$, so $Q_6(b_0) = P(b_0)$.

(Now the remaining excess mass initially placed at b_0 has been moved to the P-massless point b_5. Finally, the remaining excess mass at a_2 will be moved to P-massless b_6 and then these new atoms $\{b_i\}_{i=1}^6$ will all be combined.)

Let

$$Q_7 = \text{fus}\left(b_4, a_2, \{b_4, a_2\}^c; (s, s, 0); Q_6\right),$$

where $1 - s = (1 - t_2)P(a_2)[(1 - \varepsilon_1)(1 - \varepsilon_2)(1 - \varepsilon_3)(1 - \varepsilon_4)P(a_2)]^{-1}$.

Then

$$Q_7(a_2) = (1 - t_2)P(a_2) = Q(a_2),$$

$$Q_7(a_1) = (1 - t_1)P(a_1) = Q(a_1),$$

$$Q_7(b_0) = P(b_0),$$

and the *mass* $t_1 P(a_1) + t_2 P(a_2) = \sum_{i=1}^{6} Q_7(b_i)$ has been distributed at P-massless points $\{b_i\}_{i=1}^{6}$ in such a way that the *moment*

$$t_1 P(a_1)a_1 + t_2 P(a_2)a_2 = \sum_{i=1}^{6} Q_7(b_i)b_i.$$

Let $B = \{b_1, b_2, b_3, b_4, b_5, b_6\}$ and let

$$Q_8 = \text{fus}(B, B^c; (1, 0); Q_7).$$

From the definition of elementary fusion, it follows that

$$Q_8 = \left(t_1 P(a_1) + t_2 P(a_2)\right)\delta(b_7) + (1 - t_1)P(a_1)\delta(a_1)$$
$$+ (1 - t_2)P(a_2)\delta(a_2) + P|_{\{a_1, a_2, b_7\}^c},$$

where $b_7 = (t_1 P(a_1)a_1 + t_2 P(a_2)a_2)/(t_1 P(a_1) + t_2 P(a_2))$. Thus $Q_8 = Q$, which completes the proof if A_1 and A_2 are singletons.

It will now be shown that the general A_1, A_2 case may be reduced to the singleton A_1, A_2 case by two elementary fusions. Let $Q = \text{fus}(A_1, A_2, (A_1 \cup A_2)^c; (t_1, t_2, 0); P)$, again with $0 < t_2 < t_1 \le 1$, and define

$$Q_1 = \text{fus}(A_1, A_1^c; (t_1, 0); P),$$

$$Q_2 = \text{fus}\left(A_2 \setminus \{a_1\}, (A_2 \setminus \{a_1\})^c; (t_2, 0); Q_1\right),$$

where $a_1 = b(A_1, P)$, and

$$Q_3 = \text{fus}\left(\{a_1\}, \{a_2\}, \{a_1, a_2\}^c; (s_1, s_2, 0); Q_2\right),$$

where $a_2 = b(A_2 \setminus \{a_1\}, Q_1)$, $(1 - s_1)Q_2(a_1) = Q(a_1)$ and $(1 - s_2)Q_2(a_2) = Q(a_2)$, with $s_1 > 0$, $s_2 \ge 0$. Since Q and Q_3 agree except possibly where they place their fused masses, Lemma 3.7 implies that $Q_3 = Q$, which, since Q_3 is now in the singleton-set form treated first, completes the proof. $\square$

PROPOSITION 3.10. *Q is a simple fusion of P if and only if Q is a m.s.f. of P.*

PROOF. If Q is a m.s.f. of P, then by Lemmas 3.8 and 3.9, Q is a simple fusion of P. Conversely, since every elementary fusion is clearly a m.s.f. of the form $\text{fus}((A, A^c); (t, 0); P)$ and since the class of m.s.f.'s are closed under composition (which follows easily from Lemma 3.8) and last, since every simple

fusion is by definition a composition of a finite number of elementary fusions, it follows that every simple fusion is a m.s.f. $\square$

Next, Proposition 3.10 will be used to establish the convexity of $\mathscr{F}(P)$, the transitivity of the fusion operation and two results concerning the class of fusions of two-point distributions and of the Cauchy distribution.

THEOREM 3.11. $\mathscr{S}(P)$ and $\mathscr{F}(P)$ are convex.

PROOF. Since the closure of a convex set in a topological vector space is convex [cf. Rudin (1973), page 11], by Proposition 3.10 it suffices to show that the class of matrix simple fusions is convex.

Let $Q_1 = \text{fus}((A_i)_{i=1}^n; (t_{ij})_{i=1, j=1}^{n, k}; P)$, $Q_2 = \text{fus}((\hat{A}_i)_{i=1}^{\hat{n}}; (\hat{t}_{ij})_{i=1, j=1}^{\hat{n}, \hat{k}}; P)$ and fix $\alpha \in (0, 1)$. Define the $n\hat{n} \times (k + \hat{k})$ m.s.f. $\hat{Q}$ of P by

$$\hat{Q} = \left(\left(A_{i_1, i_2} \right)_{i_1=1, i_2=1}^{n, \hat{n}}; \left(t_{i_1, i_2, j} \right)_{i_1=1, i_2=1, j=1}^{n, \hat{n}, k+\hat{k}}; P \right),$$

where $A_{i_1, i_2} = A_{i_1} \cap \hat{A}_{i_2}$, $i_1 = 1, \ldots, n$, $i_2 = 1, \ldots, \hat{n}$ and

$$t_{i_1, i_2, j} = \begin{cases} \alpha t_{i_1, j}, & j = 1, \ldots, k, \\ (1 - \alpha)\hat{t}_{i_2, j-k}, & j = k + 1, \ldots, k + \hat{k}. \end{cases}$$

Then it is easy to check that $\hat{Q} = \alpha Q_1 + (1 - \alpha)Q_2$. $\square$

Next it will be shown that a fusion of a fusion of P is itself a fusion of P; that is, the fusion ordering is transitive.

THEOREM 3.12. If $Q \in \mathscr{F}(P)$ and $R \in \mathscr{F}(Q)$, then $R \in \mathscr{F}(P)$, [that is, $\mathscr{F}(\mathscr{F}(P)) = \mathscr{F}(P))$].

PROOF. Fix $Q \in \mathscr{F}(P)$. It suffices to show that every *elementary* fusion of Q is in $\mathscr{F}(P)$, since by induction it follows that every simple fusion of Q is in $\mathscr{F}(P)$ and therefore that weak limits of simple fusions of Q are also in $\mathscr{F}(P)$, since $\mathscr{F}(P)$ itself is weakly closed.

Let R be an arbitrary elementary fusion of Q with corresponding fusing set A with finite Q-moment and fusion proportion parameter $t \in [0, 1]$ (see Definition 3.1).

If $Q(A) = 0$ or $t = 0$, then $R = Q \in \mathscr{F}(P)$, so further assume without loss of generality that $tQ(A) > 0$. It remains only to show that $R \in \mathscr{F}(P)$, that is,

(3.3) $Q|_{A^c} + tQ(A)\delta(b(A, Q)) + (1 - t)Q|_A \in \mathscr{F}(P).$

The proof of (3.3) will be based on a monotone class argument.

CASE 1. X is a separable Banach space.

For each $T \in \mathscr{P}$ and each A with finite T-moment, define the elementary fusion T_A of T by

$$T_A = T|_{A^c} + tT(A)\delta(b(A, T)) + (1 - t)T|_A.$$

[Note that $T_A = T$ if $T(A) = 0$.]

CLAIM 1. *If A is bounded and $Q(\partial A) = 0$, then $Q_A \in \mathscr{F}(P)$.*

The proof of Claim 1 will proceed in four steps.

(i) There exists $\{P^n\} \subset \mathscr{A}(P)$ with $b(A, P^n) \to b(A, Q)$.

By the definition of $\mathscr{F}(P)$, there exists $\{P^n\} \subset \mathscr{A}(P)$ such that $P^n \to_w Q$. For every $f \in X^*$, $\int_A f \, dP^n \to \int_A f \, dQ$ since $f \cdot I_A$ is bounded and measurable and has discontinuities which constitute a set of Q-measure 0 [Billingsley (1968), page 31] (note the discontinuities of $f \cdot I_A$ are contained in ∂A) and $f \cdot I_A$ is bounded since f is linear and A is bounded. Thus $f(\int_A x \, dP^n(x)) \to f(\int_A x \, dQ(x))$ for all $f \in X^*$, so $\int_A x \, dP^n(x) \to_w \int_A x \, dQ(x)$. A well-known corollary of the Hahn–Banach theorem [Rudin (1973), page 65] implies that for some sequence of convex combinations from $\{P^n\}$ (which shall still be called $\{P^n\}$), $\int_A x \, dP^n(x) \to \int_A x \, dQ(x)$, with convergence in *norm*. By Theorem 3.11, the new sequence $\{P^n\}$ is still contained in $\mathscr{A}(P)$. By the portmanteau theorem [e.g., Billingsley (1968), page 14], $P^n(A) \to Q(A)$ since $Q(\partial A) = 0$, so (i) follows.

(ii) If $B \in \mathscr{B}$ is such that $Q(\partial B) = 0$ and $b(A, Q) \notin \partial B$, then $\delta_{b(A, P^n)}(B) \to \delta_{b(A, Q)}(B)$.

If $b(A, Q) \in B$, it is an interior point of B. Thus $b(A, P^n)$ is in B eventually by (i), so $\delta_{b(A, P^n)}(B) \to 1 = \delta_{b(A, Q)}(B)$.

On the other hand, if $b(A, Q) \notin B$, then it is in the interior of B^c, so, similarly, $\delta_{b(A, P^n)}(B) \to 0 = \delta_{b(A, Q)}(B)$. This proves (ii).

(iii) Let B be as in (ii). Then $P_A^n(B) \to Q_A(B)$.

Since $\partial(A \cap B) \subseteq \partial A \cup \partial B$ and $\partial(A^c \cap B) \subseteq \partial A^c \cup \partial B = \partial A \cup \partial B$, both of these sets $A \cap B$ and $A^c \cap B$ have Q-measure 0. So by the definition of P_A^n and Q_A, the portmanteau theorem again and (ii), statement (iii) follows.

(iv) $\mathscr{D} = \{B \in \mathscr{B} \colon Q(\partial B) = 0, \; b(A, Q) \notin \partial B\}$ is a convergence determining class for $P_A^n \to_w Q_A$ [that is, if $P_A^n(B) \to Q_A(B)$ for all $B \in \mathscr{D}$, then $P_A^n \to_w Q_A$].

Clearly, $\mathscr{D}$ is closed under finite intersections, and it is easy to see by the separability of X that every open set is a countable union of subsets of $\mathscr{D}$ (to see this, just consider open balls centered at some point; as the radii vary, the boundaries are disjoint), so (iv) follows by Billingsley [(1968), Theorem 2.2, page 14], which completes the proof of Claim 1.

Let A_0 be a set in $\mathscr{B}$ with finite Q-moment and define

$$\mathscr{C}_{A_0} = \{B \subseteq A_0 \colon B \in \mathscr{B}, Q_B \in \mathscr{F}(P)\}.$$

CLAIM 2. *$\mathscr{C}_{A_0}$ is a monotone class.*

Let $B_n \uparrow B$, $B_n \in \mathscr{C}_{A_0}$ (the case $B_n \downarrow B$ is similar) and assume $Q(B) \neq 0$. Then $yI_{B_n}(y) \to yI_B(y) \ \forall \ y$, so $\int_{B_n} y \, dQ(y) \to \int_B y \, dQ(y)$ (*norm* convergence) by the dominated convergence theorem [Dunford and Schwartz (1958), page 151, Corollary 16] since $\|y\|$ is integrable on A_0.

Since $Q(B_n) \to Q(B)$, $b(B_n, Q) \to b(B, Q)$ in *norm*. Now let $E \in \mathscr{B}$ satisfy $Q_B(\partial E) = 0$. But $Q_B(\partial E) = 0 \Rightarrow b(B, Q) \notin \partial E \Rightarrow b(B_n, Q) \notin \partial E$ eventually, as before. Thus, as before, $\delta_{b(B_n, Q)}(E) \to \delta_{b(B, Q)}(E)$. Also

$$Q(B_n^c \cap E) \to Q(B^c \cap E),$$

$$Q(B_n \cap E) \to Q(B \cap E),$$

$$Q(B_n) \to Q(B),$$

by the monotonicity of $\{B_n\}$. Hence $Q_{B_n}(E) \to Q_B(E)$, so the portmanteau theorem implies $Q_{B_n} \to_w Q_B$, so $B \in \mathscr{C}_{A_0}$. The case $Q(B) = 0$ is trivial, since then $Q(B_n) \equiv 0$ and $Q_{B_n} = Q_B = Q \in \mathscr{F}(P)$. This proves Claim 2.

CLAIM 3. *Let A_0 be a bounded set in $\mathscr{B}$ with $Q(\partial A_0) = 0$. Then $\mathscr{S} = \{B \subseteq A_0, B \in \mathscr{B}: Q(\partial B) = 0\}$ is a field relative to A_0, contained in $\mathscr{C}_{A_0}$.*

Note that the complement of B relative to A_0 is $B^c \cap A_0$. Now $\partial(B^c \cap A_0) \subseteq \partial(B^c) \cup \partial A_0 = \partial B \cup \partial A_0$, which has Q-measure 0. Similarly, $\partial(B \cup E) \subseteq \partial B \cup \partial E$, so $\mathscr{S}$ is clearly a field. It was shown in Claim 1 that $\mathscr{S} \subseteq \mathscr{C}_{A_0}$.

CLAIM 4. *If A_0 is a bounded open set with $Q(\partial A_0) = 0$, then $\mathscr{C}_{A_0}$ contains all Borel sets in A_0.*

Any open set $B \subseteq A_0$ is the countable union of open subsets B_n of A_0 for which $Q(\partial B_n) = 0$, so the σ-field generated by $\mathscr{S}$ contains all open sets, hence all Borel sets in A_0. But by the monotone class theorem and Claims 2 and 3, so does $\mathscr{C}_{A_0}$, which completes the proof of Claim 4.

To complete the proof of (3.3) (in the case when X is a separable Banach space), observe that if A is a bounded set in $\mathscr{B}$, $A \subseteq A_0$ for some bounded open set A_0 with $Q(\partial A_0) = 0$ and so by Claim 4, $Q_A \in \mathscr{F}(P)$. Finally, if $A \in \mathscr{B}$ has finite Q-moment, there exist $A_n \uparrow A$ with A_n bounded, so by Claim 2, $A \in \mathscr{F}(P)$, since each $A_n \in \mathscr{F}(P)$.

CASE 2. *X is a compact metrizable convex subset of a locally convex t.v.s.*

Since supp P is compact, it may be assumed without loss of generality that A is relatively compact (i.e., $\bar{A}$ is compact), so each $f \in X^*$ is bounded on A; the remainder of the proof then essentially follows that of Case 1. $\square$

PROPOSITION 3.13. *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 $\mathscr{F}(P) = \{Q \in \mathscr{P}: \text{supp } Q \subseteq [\alpha_1, \alpha_2] \text{ and } b(X, Q) = \alpha_1 p + \alpha_2(1 - p)\}$.*

(In other words, starting with a two-point distribution, one can fuse it to obtain any distribution which has the same barycenter and all its mass in the closed line segment $[\alpha_1, \alpha_2]$ connecting those points.)

PROOF. Without loss of generality assume $0 < p < 1$, for otherwise the conclusion is trivial.

By Proposition 2.1, supp Q is contained in the closed line segment $[\alpha_1, \alpha_2]$ for every $Q \in \mathscr{F}(P)$ (see also Theorem 3.20 below), so without loss of generality assume $X = \mathbb{R}^1$.

Suppose Q is a purely atomic distribution with exactly n atoms $a_1, \ldots, a_n$ satisfying

$$(3.4a) \qquad \alpha_1 < a_1 < \cdots < a_n < \alpha_2,$$

$$(3.4b) \qquad \sum_{i=1}^{n} a_i Q(\{a_i\}) = \alpha_1 p + \alpha_2(1-p).$$

Letting $p_i = Q(\{a_i\})$, $i = 1, \ldots, n$, it is easily checked that Q is the m.s.f. of P given by

$$Q = \mathrm{fus}\Big(\big(\{\alpha_1\}, \{\alpha_1\}^c\big); (t_{ij})_{i=1,\,j=1}^{2,\,n}; P\Big),$$

where

$$t_{1j} = \frac{p_j(\alpha_2 - a_j)}{p(\alpha_2 - \alpha_1)}, \qquad j = 1, \ldots, n,$$

and

$$t_{2j} = \left(\frac{p_j}{1-p}\right)\left(\frac{a_j - \alpha_1}{\alpha_2 - \alpha_1}\right), \qquad j = 1, \ldots, n.$$

Then $Q \in \mathscr{F}(P)$, since $\mathscr{F}(P)$ contains all m.s.f.'s of P, by Proposition 3.10. Since the probabilities Q satisfying (3.4a) and (3.4b) are weakly dense in the set of all distributions with support in $[\alpha_1, \alpha_2]$ and barycenter $= \alpha p_1 + \alpha_2(1-p)$ and $\mathscr{F}(P)$ is closed, the proof is complete. $\square$

PROPOSITION 3.14. *Let $X = \mathbb{R}^1$ and P be the Cauchy distribution. Then $\mathscr{F}(P) = \mathscr{P}$.*

PROOF. Applying an elementary fusion of the form in Example 3.3, first to the positive axis and then to the negative axis, shows that for each $\varepsilon > 0$ and $N > 0$, there is a simple fusion Q of the Cauchy distribution (in fact consisting of a composition of only two elementary fusions) which has two atoms of masses $m_1, m_2 \geq \frac{1}{2} - \varepsilon$ located at $-b_1$ and b_2 respectively, where $b_1, b_2 \gg N$. Then, using Proposition 3.13, it is easy to see that Q can be further fused to closely approximate any given distribution with support in $[-N, N]$. Taking weak limits completes the proof. $\square$

REMARKS. It can now be seen that the fusion ordering is *not* in general antisymmetric. Let X and P be as in Proposition 3.14. Thus if Q is the translation of the Cauchy distribution by 1, $Q \in \mathscr{F}(P)$ and $P \in \mathscr{F}(Q)$ from Proposition 3.14, yet $P \neq Q$. However, if P and Q both have a finite first moment, then it will be seen in Corollary 3.24 that $Q \in \mathscr{F}(P)$ and $P \in \mathscr{F}(Q) \Rightarrow P = Q$; see also the remarks following Corollary 3.17.

Next it shall be shown that if Q is a fusion of P, then Q is convexly dominated by P, but the converse is in general not true (it is, however, if P has a finite first moment; see Theorem 4.1).

DEFINITION 3.15. For Q and $P \in \mathscr{P}$, *P convexly dominates Q* (written $P \succcurlyeq_c Q$) if $\int \phi \, dP \geq \int \phi \, dQ$ for all continuous convex functions ϕ for which both integrals exist.

PROPOSITION 3.16. *Let $Q \in \mathscr{F}(P)$. If (i) ϕ is convex and continuous on $\overline{\text{co}}(\text{supp } P)$ and ϕ is P-integrable; or if (ii) $K = \overline{\text{co}}(\text{supp } P)$ is compact and $\exists \ \phi_n \downarrow \phi$ such that ϕ_n is continuous and convex on K (so ϕ is also convex) and if ϕ is P-integrable, then*

$$\int \phi \, dQ \leq \int \phi \, dP.$$

COROLLARY 3.17. *If $Q \in \mathscr{F}(P)$, then $Q \preccurlyeq_c P$.*

REMARKS. If P has a finite first moment, then Corollary 3.17 can also be viewed as a generalization of Jensen's inequality. Proposition 4.11 below shows that no continuity assumption in Proposition 3.16 is necessary if X is finite-dimensional.

A result of Mokobodzki [cf. Alfsen (1971), page 44] implies that if ϕ is convex on a compact convex subset K of a locally convex topological vector space and if ϕ is upper-semicontinuous, then $\exists \ \phi_n \downarrow \phi$, ϕ_n continuous and convex on K. But a convex Borel function on K need not be upper-semicontinuous; see the remarks following Proposition 4.11. On the other hand, if ϕ is defined on *all* of X and is convex and bounded above on some nonempty convex open set, then ϕ is continuous [cf. Schaefer (1971), page 68]. And if X is finite-dimensional, every convex function on all X is continuous. Of course there are also many examples of *discontinuous* convex ϕ which satisfy (ii); for example, $\phi(x) = 0$ for $x \in (0, 1)$ and $\phi(0) = \phi(1) = 1$ for $K = [0, 1] \subset \mathbb{R}^1 = X$.

PROOF OF PROPOSITION 3.16. To establish part (i), first consider elementary fusions. Suppose $Q = \text{fus}((A, A^c); (t, 0); P)$, so $dQ = I_{A^c} \, dP + (1 - t)I_A \, dP + tP(A) \, d\delta(b)$, where $b = b(A, P)$. If $P(A) = 0$, then $Q = P$ and the conclusion is trivial, so suppose $P(A) > 0$. Then

$$\int \phi \, dQ = \int_{A^c} \phi \, dP + (1 - t)\int_A \phi \, dP + tP(A)\phi(b).$$

But since $b = \int_A x \, dP / P(A)$, Jensen's inequality implies

$$\phi(b) \leq \int_A \phi(x) \, dP(x) / P(A),$$

so $\int \phi \, dQ \leq \int \phi \, dP$.

By induction, conclusion (i) then holds for all simple fusions Q of P. Suppose $Q_n \to_w Q$, where each Q_n is a simple fusion of P. Assume first that $\phi \geq 0$ and that ϕ is continuous and convex [on $\overline{co}(\text{supp } P)$]. For $M \geq 0$, let $\phi^M = \min(\phi, M)$. Then $\int \phi^M \, dQ_n \to \int \phi^M \, dQ$ by weak convergence, so

$$\int \phi^M \, dQ \leq \limsup_n \int \phi \, dQ_n \leq \int \phi \, dP \quad \text{for all } M.$$

Letting $M \to \infty$, $\int \phi \, dQ \leq \int \phi \, dP$ follows by the monotone convergence theorem.

Next let ϕ be an arbitrary convex continuous function on $\overline{co}(\text{supp } P)$. For $L \leq 0$, let $\phi_L = \max(\phi, L)$ which is continuous and convex. Then $\int (\phi_L + |L|) \, dQ \leq \int (\phi_L + |L|) \, dP$, so $\int \phi_L \, dQ \leq \int \phi_L \, dP$. Then $\phi_L \to \phi$ as $L \to -\infty$, so since $|\phi_L| \leq |\phi|$ (recall $L \leq 0$) the conclusion (i) follows from the dominated convergence theorem.

Part (ii) follows immediately from (i) and the dominated convergence theorem, since $\phi_n \leq \phi_1$ which is bounded above on K and as in case (i), it is enough to establish the conclusion for nonnegative ϕ. $\square$

The infinite-dimensional conclusion of the next corollary will be strengthened later (Theorem 4.2) to uniform integrability, but will be used in its present form to show that martingalizability implies fusion (Theorem 4.1).

COROLLARY 3.18.   *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 $\mathscr{F}(P)$ is tight. Moreover, if X is finite-dimensional, then P has a finite first moment if and only if $\mathscr{F}(P)$ is tight.*

PROOF.   By Proposition 3.16, $\int \|x\| \, dQ \leq \int \|x\| \, dP$ for all $Q \in \mathscr{F}(P)$, so

$$Q(\|x\| > \lambda) \leq \lambda^{-1} \int \|x\| \, dQ \leq \lambda^{-1} \int \|x\| \, dP \quad \text{for all } Q \in \mathscr{F}(P).$$

Since the right-hand side does not depend on Q, this shows $\mathscr{F}(P)$ is tight.

To prove the second part, suppose, without loss of generality, that $X = \mathbb{R}^n$ equipped with the l_1-norm and assume $\int_X \|x\| \, dP = \infty$. Then at least one orthant A of $\mathbb{R}^n$ satisfies $\int_A \|x\| \, dP = \infty$; suppose further, without loss of generality, that A is the positive orthant $\mathbb{R}^n_+$. Then for all Borel subsets B

of A,

$$\int_B \|x\| \, dP(x) = \int_B \sum_{j=1}^n |x_j| \, dP(x) = \int_B \sum_{j=1}^n x_j \, dP(x) = \sum_{j=1}^n \int_B x_j \, dP(x)$$

$$= \sum_{j=1}^n \pi_j \left(\int_B x \, dP(x) \right) \le \sum_{j=1}^n \left\| \int_B x \, dP(x) \right\| = n \left\| \int_B x \, dP(x) \right\|,$$

where π_j is the projection onto the jth coordinate and $x_j = \pi_j(x)$.

Now given N, there exists an $r \ge 0$ so that $P(B_r) > P(A)/2 > 0$ and $\int_{B_r} \|x\| \, dP(x) > nN$, where $B_r = \{x \in \mathbb{R}^n_+ : \|x\| \le r\}$. Then $\|\int_{B_r} x \, dP(x)\| > N$, so $\|b(B_r, P)\| > N$, which implies that the elementary fusion $Q = \text{fus}(B_r, B_r^c; (1,0); P)$ satisfies $Q(x \in A : \|x\| > N) > P(A)/2$, so $\mathscr{F}(P)$ is not tight, since N is arbitrary. $\square$

In general (i.e., without a moment or similar condition), $\mathscr{F}(P)$ may not be tight, as is seen immediately from Proposition 3.14, since if P is the Cauchy distribution on $\mathbb{R}^1$, then $\mathscr{F}(P) = \mathscr{P}$.

As the next example shows, the converse of Corollary 3.17 does not hold in general, that is, Q may be convexly dominated by P without being in $\mathscr{F}(P)$ if P does not have a first moment, even in the finite-dimensional case.

EXAMPLE 3.19.   Let $X = \mathbb{R}^2$, let P be the (one-dimensional) Cauchy distribution supported on the x-axis and let Q be the Cauchy distribution supported on the y-axis. Since the only convex functions c for which $\int c \, dP$ and $\int c \, dQ$ both exist are those c which are identically zero on both axes (and hence zero everywhere by convexity), P trivially dominates Q convexly (and vice versa), but clearly $Q \notin \mathscr{F}(P)$ since $\text{supp } Q$ is not contained in $\overline{\text{co}}(\text{supp } P)$ (see Theorem 3.20 below).

[Observe that a two-dimensional Cauchy example was needed, since in $\mathbb{R}^1$, it follows from Proposition 3.14 that $Q \in \mathscr{F}(P)$ for *all* $Q \in \mathscr{P}$, in particular for any Q convexly dominated by P.]

The next result generalizes the main idea behind the last example.

THEOREM 3.20.   *If* $Q \in \mathscr{F}(P)$, *then* $\overline{\text{co}}(\text{supp } Q) \subseteq \overline{\text{co}}(\text{supp } P)$.

PROOF.   The conclusion follows immediately from Proposition 2.1 and the definition of fusion if Q is a simple fusion of P. The general case then follows easily using the portmanteau theorem. $\square$

It shall now be shown that if P and Q have finite first moments, then a very special class of convex functions is separating, namely, the positive parts of affine functions, or *wedge functions*. That is, if $\int a^+ \, dP = \int a^+ \, dQ$ for all continuous affine functions a, then $P = Q$. Surprisingly, these functions do *not*, however, determine convex domination: $\int a^+ \, dP \ge \int a^+ \, dQ$ for all affine functions a does *not* imply $\int \phi \, dP \ge \int \phi \, dQ$ for all convex functions. An

example of this will be given in $\mathbb{R}^2$ in the next section, answering in the negative a differently formulated question raised by Marshall and Olkin (1979). No such example is possible in $\mathbb{R}^1$ [see Theorem 4.7(vii) below]. This gives some insight into why the connection among convex domination and dilations and fusions is so much simpler in $\mathbb{R}^1$ than in higher dimensions.

The proof that wedge functions are separating is very easy in finite-dimensions, using only the well-known fact that probability measures are determined by their values on half-spaces. In infinite dimensions, the proof will be reduced to the finite-dimensional case by using the so-called approximation property (AP) of Grothendiek (1955).

The following is a suitable definition of the approximation property for our purposes.

DEFINITION 3.21.   A t.v.s. X has the AP if for every compact subset K of X and every open neighborhood V of 0 in X, there exists $T: X \to X$ a continuous linear operator of finite rank such that $Tx - x \in V$ for all $x \in K$. That is, the identity operator can be uniformly approximated on compact sets by an operator of finite rank.

Enflo showed in a famous counterexample [Enflo (1973)] that not every space has the AP, so one must embed the space in one which does have the AP; this works fine for our problem, since the measure then just lives on a subspace.

We are grateful to Steve Bellenot for suggesting the proof of the following lemma.

LEMMA 3.22.   *Every l.c.t.v.s. X is a subspace (in both the linear and topological sense) of a l.c.t.v.s. with the AP.*

PROOF.   There exists a separating family F of continuous seminorms on X, such that the sets $\{V_p(\varepsilon): p \in F, \varepsilon > 0\}$ form a local subbase for the topology of X, where $V_p(\varepsilon) = \{x: p(x) < \varepsilon\}$; see Rudin [(1973), page 26–27]. For $p \in F$, X/N_p is a normed space, where N_p is the closed subspace $\{x: p(x) = 0\}$. Now any normed space may be isometrically embedded in some $C(\Omega)$ space, where Ω is compact Hausdorff [Dunford and Schwartz (1958), page 424] and $C(\Omega)$ has the approximation property [Grothendieck (1955), page 185, Proposition 41]. Thus we have $X \to_{\pi_p} X/N_p \to_{i_p} C(\Omega_p)$, where π_p is the quotient map and i_p is an isometric embedding. Define $\psi: X \to \prod_{p \in F} C(\Omega_p)$ by $\psi(x) = (i_p \circ \pi_p(x))_{p \in F}$. By the definition of the quotient and product topologies and the fact that the V_p's form a local subbase for X's topology, $\psi(X)$ is linearly homeomorphic to X, so X may be considered a subspace of $\prod_{p \in F} C(\Omega_p)$. But it is easy to show that a product of spaces with the AP has the AP [Grothendieck (1955), Lemma 19, page 169]. $\square$

By an affine function is meant a function of the form $a(x) = l(x) + b$, where l is linear.

THEOREM 3.23.   *Let P and Q be tight probability measures on a l.c.t.v.s. X [if X is a complete separable metric space, such as a separable Banach space, tightness is automatic; see Billingsley (1986), page 10]. If P and Q both have finite first moments, then*

$$\int a^+ \, dP = \int a^+ \, dQ \quad \text{for all affine } a \Rightarrow P = Q.$$

PROOF.

CASE 1. $X = \mathbb{R}^1$. The right derivative of the function $\phi_P(t) = \int (x - t)^+ \, dP(x)$ is $-P((t, \infty))$, and therefore the distribution of P is uniquely determined by $\phi_P(t)$. But since $x - t$ is affine, $\phi_P = \phi_Q$, so $P = Q$.

CASE 2. $X = \mathbb{R}^n$. It is well known that two probability measures on $\mathbb{R}^n$ are the same if they agree on all half spaces; see Billingsley [(1986), page 396] (the proof uses a Fourier transform argument). A half-space is a set of the form $H = \{x: l(x) \le \alpha\}$ where l is a nonzero real-valued linear function on $\mathbb{R}^n$.

Define a probability measure P_l on $\mathbb{R}^1$ by $P_l(B) = P(l^{-1}(B))$ for B a Borel set in $\mathbb{R}^1$, and similarly for Q_l. Now $P(H) = P_l((-\infty, \alpha])$ and similarly for Q, so to show $P = Q$ it is enough to show $P_l = Q_l$ for all such l.

Let ϕ be any affine function on $\mathbb{R}^1$. Define an affine function a on $\mathbb{R}^n$ by $a(x) = \phi(l(x))$. Now by a change of variables, $\int_{\mathbb{R}^1} \phi^+(t) \, dP_l(t) = \int_{\mathbb{R}^n} \phi^+(l(x)) \, dP(x) = \int a^+ \, dP = \int a^+ \, dQ$ by hypothesis, so $\int \phi^+ \, dP_l = \int \phi^+ \, dQ_l$ for all affine ϕ in $\mathbb{R}^1$. By Case 1, $P_l = Q_l$.

GENERAL CASE.   By Lemma 3.22, assume X has the AP. Let $f \in C(X)$, $\|f\| \le 1$ and fix $\varepsilon > 0$. Choose $K \subset X$, K compact, such that $P(K) > 1 - \varepsilon$ and $Q(K) > 1 - \varepsilon$. Since K is compact and f is continuous, there is an open neighborhood V of zero such that $x \in K$, $y - x \in V \Rightarrow |f(y) - f(x)| < \varepsilon$. Let T be a continuous linear operator on X of finite rank such that $Tx - x \in V$ for all $x \in K$. Thus $|\int f(Tx) \, dP(x) - \int f(x) \, dP(x)| < \varepsilon P(K) + 2P(K^c) \le 3\varepsilon$, and similarly for Q. But $\int f(Tx) \, dP(x) = \int f(y) \, dP_T(y)$, and $P_T = P \circ T^{-1}$ is carried on the range of T which is finite-dimensional. And $\int a^+(y) \, dP_T(y) = \int a^+(Tx) \, dP(x) = \int a^+(Tx) \, dQ(x) = \int a^+(y) \, dQ_T(y)$ for all affine functions a on range T (since $a \circ T$ is affine on X), so $Q_T = P_T$ by Case 2. Thus $|\int f(x) \, dP(x) - \int f(x) \, dQ(x)| \le 6\varepsilon$. Since $\varepsilon > 0$ is arbitrary, $\int f(x) \, dP(x) = \int f(x) \, dQ(x)$ for all $f \in C(X)$, so $P = Q$. □

COROLLARY 3.24.   *If $Q \in \mathscr{F}(P)$ and $P \in \mathscr{F}(Q)$ and either has a finite first moment, then $P = Q$.*

PROOF.   By Corollary 3.17, $\int a^+ \, dP = \int a^+ \, dQ$ for all affine functions a, so the conclusion follows immediately by Theorem 3.23. □

The notion of matrix simple fusion may be generalized to countable partitions as follows; this generalization will be needed in the proof that martingalizability implies fusion (Theorem 4.1 below).

DEFINITION 3.25. Q is a *matrix countable fusion* (m.c.f.) of P if P has a finite first moment and if there exist $(t_{ij})_{i=1, j=1}^{\infty, k}$ nonnegative with $t_i = \sum_{j=1}^{k} t_{ij} \le 1$ for all i and $(A_i)_{i=1}^{\infty}$ a measurable partition of X, so that

$$Q = \sum_{i=1}^{\infty} (1 - t_i) P|_{A_i} + \sum_{j=1}^{k} \delta(b_j) \left[\sum_{i=1}^{\infty} t_{ij} P(A_i) \right],$$

where

$$b_j = \frac{\sum_{i=1}^{\infty} t_{ij} P(A_i) b(A_i, P)}{\sum_{i=1}^{\infty} t_{ij} P(A_i)} \quad \text{(when the denominator } \ne 0\text{)}$$

(it will be shown in a moment that this exists). Notationally, denote such a Q

$$Q = \text{fus}\left((A_i)_{i=1}^{\infty}; (t_{ij})_{i=1, j=1}^{\infty, k}; P \right).$$

PROPOSITION 3.26. *Every matrix countable fusion of P is a fusion of P.*

PROOF. In order to show that every m.c.f. Q of P is in $\mathscr{F}(P)$, it may be shown that Q is the weak limit of m.s.f.'s of P. First, the sum $\sum_{i=1}^{\infty} t_{ij} P(A_i)$ converges because $\sum_{i=1}^{\infty} P(A_i) = P(X)$ converges. To show that $\sum_{i=1}^{\infty} t_{ij} P(A_i) b(A_i, P)$ converges, first consider the case where X is a separable Banach space, in which case

$$\left\| \sum_{i=N}^{M} t_{ij} P(A_i) b(A_i, P) \right\| = \left\| \sum_{i=N}^{M} t_{ij} \int_{A_i} x \, dP \right\|$$

$$\le \sum_{i=N}^{M} t_{ij} \int_{A_i} \|x\| \, dP \le \int_{\cup_{i=N}^{\infty} A_i} \|x\| \, dP.$$

Since $P(\cup_{i=N}^{\infty} A_i) \to 0$ as $N \to \infty$ and $\|x\|$ is integrable, the sequence is Cauchy, hence convergent. On the other hand, if X is a compact metrizable convex subset of a l.c.t.v.s., for each $f \in X^*$,

$$\left| f\left(\sum_{i=N}^{M} t_{ij} \int_{A_i} x \, dP \right) \right| \le \sum_{i=N}^{M} t_{ij} \int_{A_i} |f(x)| \, dP \le \sum_{i=N}^{\infty} P(A_i) \|f\|_{\infty},$$

which converges to 0 as $N \to \infty$. Thus the *scalar* series obtained by applying f converges to a finite limit for each $f \in X^*$, so since the original series lives in the compact set X, the series converges.

Next, let $Q_n = \mathrm{fus}((A_i^n)_{i=1}^{n+1}; (t_{ij}^{(n)})_{i=1, j=1}^{n+1, k}; P)$, where $t_{ij}^{(n)} = t_{ij}$ if $i \le n$ and $= 0$ otherwise, and $A_i^n = A_i$ if $i \le n$, and $= (\bigcup_{k=1}^{n} A_k)^c$ if $i = n+1$. Let

$$b_j^n = \sum_{i=1}^{n+1} t_{ij}^{(n)} P(A_i^n) b(A_i^n, P) \Big/ \sum_{i=1}^{n+1} t_{ij}^{(n)} P(A_i^n)$$

$$= \sum_{i=1}^{n} t_{ij} P(A_i) b(A_i, P) \Big/ \sum_{i=1}^{n} t_{ij} P(A_i).$$

Now $b_j^n \to b_j$, since it was already proved that the numerator and denominators converge. Then

$$Q_n(B) = \sum_{i=1}^{n} (1 - t_i) P(B \cap A_i) + P\left(B \cap \bigcup_{i=n+1}^{\infty} A_i\right)$$

$$+ \sum_{j=1}^{k} \sum_{i=1}^{n} t_{ij} P(A_i) \delta(b_j^n)(B).$$

Observe that for every B, the first term obviously converges to $\sum_{i=1}^{\infty}(1 - t_i) P(B \cap A_i)$ and the second term converges to 0. Now if $b_j \notin \partial B$ for $j = 1, \ldots, k$, then $\delta(b_j^n)(B) \to \delta(b_j)(B)$ and since the set of all such B is a *weak convergence determining class* [Billingsley (1968), page 14], this completes the proof that $Q \in \mathscr{F}(P)$. □

4. Fusions of probabilities with finite first moments. Recall that if Q is a fusion of P, then Q is convexly dominated by P (Corollary 3.17) and that the converse is not true in general, even in finite-dimensional spaces (Example 3.19). However, if P has a finite first moment [i.e., $b(X, P)$ exists] then Q is a fusion of P if and only if Q is convexly dominated by P, as will now be shown.

Throughout this section, X is either a separable Banach space or a compact metrizable convex subset of a l.c.t.v.s., and P is a Borel probability measure on X. Recall that for $Q, P \in \mathscr{P}$, the ordered pair (Q, P) is *martingalizable* if there exists an X-valued martingale (Z_1, Z_2) with $\mathscr{L}(Z_1) = Q$ and $\mathscr{L}(Z_2) = P$ and that a *dilation* on X is a Markov kernel μ from X to X such that for all continuous affine functions ϕ on X, $\phi(x) = \int \phi(r) \mu(dr, x)$ [cf. Phelps (1966) for details]. The main result of this section is the following theorem.

THEOREM 4.1. *If P and Q are Borel 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) *$Q \preccurlyeq_c P$;*
   (iii) *(Q, P) is martingalizable;*
   (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 (1929, 1959) for one-dimensional spaces, by Blackwell (1953), Stein and Sherman for finite-dimensional spaces and Cartier, Fell and Meyer (1964) and Strassen (1965) in various infinite-dimensional settings [see Phelps (1966)]. (Another equivalent condition, which will not be dealt with in this paper, is the Loomis strong ordering [Phelps (1966), page 112], which has applications in the theory of group representations.) The main task here will be to show the equivalence of (i) with (ii)–(iv).

PROOF OF THEOREM 4.1. By the definitions of dilation and martingalizable, it is clear that

(4.1)   (iii) $\Leftrightarrow$ (iv).

Next observe that

(4.2)   (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii)   if X is a separable Banach space,

where the first implication is by Corollary 3.17 and the equivalence follows from Theorem 8 of Strassen (1965), observing that his argument applies to the separable Banach space case as well (as he states) and that his argument shows that only *continuous* convex functions need be considered.

Similarly,

(4.3)   (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iv)

                if X is a compact metrizable convex subset of a l.c.t.v.s.,

where the first implication again follows by Corollary 3.17 and the equivalence is Cartier's result [e.g., Phelps (1966), page 112; note that there X is *not* assumed to be metrizable].

From (4.1), (4.2) and (4.3), it follows that the proof will be complete once it is shown that (iii) $\Rightarrow$ (i). This will be proved first in the Banach space setting and then in the l.c.t.v.s. setting, although there is much overlap in the ideas.

BANACH SPACE CASE. Let (Y_1, Y_2) be a martingale taking values in a separable Banach space X, with underlying probability space $(\Omega, \mathscr{M}, \mu)$ such that $\mu(Y_1 \in B) = Q(B)$ and $\mu(Y_2 \in B) = P(B)$ for all Borel subsets B of X.

Fix $\varepsilon > 0$ and let $(\pi_j)_{j=1}^{\infty}$ be a Borel partition of X with diam $\overline{\mathrm{co}}(\pi_j) < \varepsilon$ for all j. Let $A_i = Y_1^{-1}(\pi_i)$ and $B_i = Y_2^{-1}(\pi_i)$ for $i = 1, 2, \ldots$, and observe that $(A_i)_{i=1}^{\infty}$ and $(B_i)_{i=1}^{\infty}$ each are $\mathscr{M}$-measurable partitions of Ω. Choose N so large that

$$P\left(\bigcup_{i=1}^{N} \pi_i\right) \cdot Q\left(\bigcup_{i=1}^{N} \pi_i\right) > 1 - \varepsilon,$$

and let $b_i = b(\pi_i; P) \; (= \int_{\pi_i} x \, dP / P(\pi_i))$, $i = 1, 2, \ldots$, and note that $b_i \in \overline{\mathrm{co}}(\pi_i)$ by Proposition 2.1.

Let $t_{ij} = \mu(A_j \cap B_i)/\mu(B_i)$ if $\mu(B_i) > 0$ (and $= 0$ otherwise) and note that $\mu(B_i) = P(\pi_i)$ and $t_i := \sum_{j=1}^{N} t_{ij} \leq 1$. Let $\hat{Q}$ be the matrix countable fusion (Definition 3.25) of P given by $\hat{Q} = \mathrm{fus}((\pi_i)_{i=1}^{\infty}; (t_{ij})_{i=1,\,j=1}^{\infty,\,N}; P)$ and note that the associated barycenters of $\hat{Q}$ are

$$a_j = \frac{\sum_{i=1}^{\infty} t_{ij} P(\pi_i) b_i}{\sum_{i=1}^{\infty} t_{ij} P(\pi_i)}, \qquad j = 1, 2, \ldots, N.$$

Then

$$(4.4) \qquad \sum_{i=1}^{\infty} t_{ij} P(\pi_i) = \sum_{i=1}^{\infty} \mu(A_j \cap B_i) = \mu(A_j),$$

so

$$a_j = \sum_{i=1}^{\infty} \frac{\mu(A_j \cap B_i) b_i}{\mu(A_j)}.$$

By (4.4) and the definitions of (π_i), (A_i) and t_i,

$$\hat{Q}(B) = \sum_{i=1}^{\infty} (1 - t_i) P(B \cap \pi_i) + \sum_{j=1}^{N} \delta(a_j)(B)\mu(A_j).$$

Let

$$\tilde{a}_j = \frac{\int_{A_j} Y_1 \, d\mu}{\mu(A_j)} = \frac{\int_{A_j} Y_2 \, d\mu}{\mu(A_j)}$$

(where the second equality follows by the martingale property) and observe that $\tilde{a}_j \in \overline{\mathrm{co}}(\pi_j)$, since $Y_1(\omega) \in \pi_j$ when $\omega \in A_j$. Then

$$\tilde{a}_j = \sum_{i=1}^{\infty} \frac{\int_{A_j \cap B_i} Y_2 \, d\mu}{\mu(A_j)} = \sum_{i=1}^{\infty} \frac{\mu(A_j \cap B_i)}{\mu(A_j)} \frac{\int_{A_j \cap B_i} Y_2 \, d\mu}{\mu(A_j \cap B_i)} = \sum_{i=1}^{\infty} \frac{\mu(A_j \cap B_i)}{\mu(A_j)} \tilde{b}_{ij},$$

where

$$\tilde{b}_{ij} = \frac{\int_{A_j \cap B_i} Y_2 \, d\mu}{\mu(A_j \cap B_i)}.$$

Furthermore, note that $\tilde{b}_{ij}$ is in $\overline{\mathrm{co}}(\pi_i)$, since $Y_2(\omega) \in \pi_i$ when $\omega \in B_i$. Since $\mathrm{diam}\,\overline{\mathrm{co}}(\pi_i) < \varepsilon$, $\|\tilde{a}_j - a_j\| \leq \sum_{i=1}^{\infty}(\mu(A_j \cap B_i)/\mu(A_j))\|\tilde{b}_{ij} - b_i\| \leq 2\varepsilon$ and since $\tilde{a}_j \in \overline{\mathrm{co}}(\pi_j)$,

$$\mathrm{dist}\left(a_j, \overline{\mathrm{co}}(\pi_j)\right) \leq 2\varepsilon.$$

Denote $\hat{Q} = \hat{Q}_\varepsilon$ to indicate the dependence on ε.

Let f be a bounded uniformly continuous real-valued function on X and suppose $|f(x) - f(y)| < d(\varepsilon)$ whenever $\|x - y\| \leq 3\varepsilon$; thus $d(\varepsilon) \to 0$ as $\varepsilon \to 0$. Compute

$$\int f \, dQ = \int f(Y_1) \, d\mu = \sum_{j=1}^{N} \int_{A_j} f(Y_1) \, d\mu + \int_{\cup_{j=N+1}^{\infty} A_j} f(Y_1) \, d\mu$$

and

$$\int f d\hat{Q}_\varepsilon = \sum_{j=1}^{N} \int f(a_j)\mu(A_j) + \sum_{i=1}^{\infty} (1 - t_i)\int_{\pi_i} f(x)P(dx).$$

But

$$\left| \int_{\cup_{j=N+1}^{\infty} A_j} f(Y_1)\, d\mu \right| \leq \|f\|_\infty \mu\left(\bigcup_{j=N+1}^{\infty} A_j \right) \leq \varepsilon \|f\|_\infty$$

and

$$\left| \sum_{i=1}^{\infty} (1 - t_i)\int_{\pi_i} f(x)P(dx) \right| \leq \sum_{i=1}^{\infty} (1 - t_i)P(\pi_i)\|f\|_\infty \leq \varepsilon\|f\|_\infty,$$

where the last inequality follows since

$$1 - t_i = \sum_{j=N+1}^{\infty} t_{ij} = \sum_{j=N+1}^{\infty} \frac{\mu(A_j \cap B_i)}{\mu(B_i)} \leq \frac{\varepsilon}{\mu(B_i)} = \frac{\varepsilon}{P(\pi_i)}.$$

Finally

$$\left| \sum_{j=1}^{N} \int_{A_j} f(Y_1)\, d\mu - \sum_{j=1}^{N} f(a_j)\mu(A_j) \right|$$

$$\leq \sum_{j=1}^{N} \sup_{\omega \in A_j} |f(Y_1(\omega)) - f(a_j)|\mu(A_j) \leq d,$$

since $\|Y_1(\omega) - a_j\| < 3\varepsilon$ for $\omega \in A_j$, which follows because $Y_1(\omega) \in \overline{\mathrm{co}}(\pi_j)$ for $\omega \in A_j$.

Since $d \to 0$ as $\varepsilon \to 0$,

$$\int f d\hat{Q}_\varepsilon \to \int f dQ \quad \text{as } \varepsilon \to 0,$$

which holds for all uniformly continuous bounded f. By Proposition 3.26, $\hat{Q}_\varepsilon \in \mathscr{F}(P)$, which is tight by Corollary 3.18. This implies that $\hat{Q}_\varepsilon \to_\omega Q$, so $Q \in \mathscr{F}(P)$ by the portmanteau theorem, which completes the proof that (iii) $\Rightarrow$ (i) in the Banach space case.

LOCALLY CONVEX TOPOLOGICAL VECTOR SPACE CASE. Let (Y_1, Y_2) be a martingale in a compact metrizable convex subset X of a l.c.t.v.s., with underlying probability space $(\Omega, \mathscr{M}, \mu)$, so that $\mu(Y_1 \in B) = Q(B)$ and $\mu(Y_2 \in B) = P(B)$ for all Borel subsets B of X.

Let U be an arbitrary convex neighborhood of 0, let W be a convex neighborhood of 0 such that $W = -W$ and $W + W + W + W \subset U$ and let V be an open convex neighborhood of 0 with $\overline{V} \subset W$ [cf. Rudin (1973), Chapter 1].

Since X is compact, there is a Borel partition $(\pi_i)_{i=1}^{N}$ of X with $\pi_i \subset V + x_i$ for some $x_1, \ldots, x_j \in X$. Define $(A_i)_{i=1}^{N}, (B_i)_{i=1}^{N}, (b_i)_{i=1}^{N}, (a_j)_{j=1}^{N}$ and $(t_{ij})_{i=1, j=1}^{N, N}$ as in the Banach space case and observe that since V is convex, $b_i \in \overline{\mathrm{co}}(\pi_i) \subset$

$W + x_i$ for each $i = 1, \ldots, N$, and that in this case, $\sum_{j=1}^{N} t_{ij} = 1$. Let $\hat{Q}$ be the matrix simple fusion of P given by $\hat{Q} = \text{fus}((\pi_i)_{i=1}^{N}; (t_{ij})_{i=1, j=1}^{N, N}; P)$ and let

$$\tilde{a}_j = \frac{\int_{A_j} Y_1 \, d\mu}{\mu(A_j)} = \frac{\int_{A_j} Y_2 \, d\mu}{\mu(A_j)},$$

where again the second equality follows by the martingale property. So

$$\tilde{a}_j = \sum_{i=1}^{N} \frac{\mu(A_j \cap B_i)}{\mu(A_j)} \tilde{b}_{ij},$$

where the $\tilde{b}_{ij}$ are defined as in the Banach space case. Observe that again $\tilde{b}_{ij} \in \overline{\text{co}}(\pi_i)$ for each j and i, so $\tilde{b}_{ij} \in W + x_i$ and $\tilde{a}_j \in \overline{\text{co}}(\pi_j) \subset W + x_j$.

Now let f be a continuous real-valued function on X; then $\int f \, dQ = \sum_{j=1}^{N} \int_{A_j} f(Y_1) \, d\mu$ and $\int f \, d\hat{Q} = \sum_{j=1}^{N} f(a_j) \mu(A_j)$. But $b_i - \tilde{b}_{ij} \in W + x_i - (W + x_i) = W - W = W + W$, since $W = -W$. Thus

$$a_j - \tilde{a}_j = \sum_{i=1}^{N} \frac{\mu(A_j \cap B_i)}{\mu(A_j)} \left(b_i - \tilde{b}_{ij} \right)$$

is a convex combination of elements of $W + W$, so $a_j - \tilde{a}_j \in W + W$, since this set is convex. Thus $a_j \in W + W + \tilde{a}_j \subset W + W + W + x_j$, so $Y_1(\omega) - a_j \in (W + x_j) - (W + W + W + x_j) = W + W + W + W \subset U$ for all $\omega \in A_j$. Since f is *uniformly* continuous on X (by the compactness of X), for each $\varepsilon > 0$, there exists an open neighborhood U_ε of 0 such that $x - y \in U_\varepsilon \Rightarrow |f(x) - f(y)| < \varepsilon$. Thus $|\int f \, dQ - \int f \, d\hat{Q}_\varepsilon| \leq \varepsilon \sum_{j=1}^{N} \mu(A_j) \leq \varepsilon$, which proves that $\hat{Q}_\varepsilon \to_w Q$, so $Q \in \mathscr{F}(P)$. $\square$

REMARKS. Actually the implication (i) $\Rightarrow$ (iii) follows directly (without Strassen's and Cartier's results) from the following fact which is not difficult to prove: If $\{Q_n\}$ is uniformly integrable (see Theorem 4.2 below), if $Q_n \to_w Q$ and if (Q_n, P) is martingalizable for all n, then (Q, P) is martingalizable.

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 $\int \phi \, dQ \leq \int \phi \, dP$ for all *nonnegative* continuous convex functions if and only if there is a fusion $\hat{P}$ 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 Pólya (1929).

THEOREM 4.2. *If P is a Borel probability measure on a separable Banach space, then P has a finite first moment if and only if $\mathscr{F}(P)$ is uniformly integrable. More generally, if ϕ: $[0, \infty) \to [0, \infty)$ is convex, nonconstant and nondecreasing, then $\int \phi(\|x\|) \, dP < \infty$ if and only if $\mathscr{F}(P)$ is uniformly*

ϕ-*integrable, that is,*

$$(4.5) \qquad \lim_{\lambda \to \infty} \sup_{Q \in \mathscr{F}(P)} \int_{\|x\| > \lambda} \phi(\|x\|) \, dQ = 0.$$

PROOF. Since ϕ is not constant, there exist $c > 0$ and $\alpha > 0$ such that

$$(4.6) \qquad \phi(\beta) \geq c\beta \quad \text{for all } \beta \geq \alpha.$$

Thus the ϕ-integrability of P implies that P has a finite first moment.

To see (4.5), let $Q \in \mathscr{F}(P)$; by Theorem 4.1 there exists a martingale (Y_1, Y_2) having marginal distributions (Q, P). Thus $(\phi(\|Y_1\|), \phi(\|Y_2\|))$ is a submartingale, so

$$(4.7) \qquad \int_{\phi(\|Y_1\|) > \lambda} \phi(\|Y_1\|) \leq \int_{\phi(\|Y_1\|) > \lambda} \phi(\|Y_2\|),$$

since $\{\phi(\|Y_1\|) > \lambda\}$ is Y_1-measurable. Next observe that $\mathrm{Prob}(\phi(\|Y_1\|) > \lambda) \leq \lambda^{-1} E\phi(\|Y_1\|) \leq \lambda^{-1} E\phi(\|Y_2\|)$. Since $\phi(\|Y_2\|)$ is integrable, for all $\varepsilon > 0$ there exists a $\delta > 0$ satisfying

$$(4.8) \qquad \int_S \phi(\|Y_2\|) < \varepsilon \quad \text{for all measurable } S \text{ with } \mathrm{Prob}(S) < \delta.$$

Take λ so large that $\lambda^{-1} E\phi(\|Y_2\|) < \delta$, so

$$(4.9) \qquad \int_{\phi(\|Y_1\|) > \lambda} \phi(\|Y_1\|) \leq \int_{\phi(\|Y_1\|) > \lambda} \phi(\|Y_2\|) \leq \varepsilon,$$

where the first inequality follows by (4.7) and the second by (4.8). Now (4.5) follows from (4.6) and (4.9). The converse is easy, since $P \in \mathscr{F}(P)$. $\square$

In the Banach space case, a *quantitative* version of the uniform integrability of $\mathscr{F}(P)$ is possible; see Theorem 4.6 below.

A quantitative version of the uniform ϕ-integrability of $\mathscr{F}(P)$ as a consequence of ϕ-integrability of P is also possible, by generalizing Definition 4.3 below to the ϕ-characteristic of P [cf. van der Vecht (1986), page 47]; the proof of the corresponding analog of Theorem 2.6 is essentially the same.

DEFINITION 4.3. For a Borel probability measure P with 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

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

In other words, $r_P(\lambda) = E(\|Y\| \mid \|Y\| \geq \lambda)$, where Y is an X-valued random variable with distribution P. Van der Vecht (1986) has generalizations, an inversion formula and properties and applications of this function in the $\mathbb{R}^1$-framework.

LEMMA 4.4 [van der Vecht (1986), page 49]. *Let S be the supremum of a nonnegative (real-valued) submartingale with last term Y [i.e., $E(Y|\mathscr{F}_t) \geq Y_t$ a.s. $\forall\, t$]. Then $\mathrm{Prob}(S \geq r(\lambda)) \leq \mathrm{Prob}(Y \geq \lambda)\ \forall\, \lambda \geq 0$.*

(This bound is referred to in van der Vecht as the *Blackwell–Dubins bound*, since it stems from a result of Blackwell and Dubins [cf. van der Vecht (1986), page 39] relating such a bound to the Hardy–Littlewood maximal functions of stopped martingales.)

THEOREM 4.5. *Let (Z_1, Z_2) be a nonnegative submartingale on a probability space $(\Omega, \mathscr{A}, \mu)$. Then*

$$\int_{Z_1 \geq r(\lambda)} Z_1\, d\mu \leq \int_{Z_2 \geq \lambda} Z_2\, d\mu \quad \text{for all } \lambda \geq 0.$$

PROOF. First observe that for every nonnegative random variable Z and real number b,

$$(4.10) \quad \int_B Z\, d\mu \leq \int_{Z \geq b} Z\, d\mu \quad \text{for every } B \in \mathscr{B} \text{ with } \mu(B) \leq \mu(Z \geq b).$$

Then calculate

$$\int_{Z_1 \geq r(\lambda)} Z_1\, d\mu \leq \int_{Z_1 \geq r(\lambda)} Z_2\, d\mu \leq \int_{Z_1 \vee Z_2 \geq r(\lambda)} Z_2\, d\mu \leq \int_{Z_2 \geq \lambda} Z_2\, d\mu,$$

where the first inequality follows from the submartingale property, the second since Z_2 is nonnegative and the range of integration is larger, and the third by (4.10) and Lemma 4.4. $\square$

THEOREM 4.6. *Let P be a Borel probability measure with finite first moment on a separable Banach space. Then for all Q in $\mathscr{F}(P)$ and all $\lambda \geq 0$,*

$$\int_{\|x\| \geq r_P(\lambda)} \|x\|\, dQ \leq \int_{\|x\| \geq \lambda} \|x\|\, dP.$$

PROOF. Since $Q \in \mathscr{F}(P)$,

$$\int \phi(\|x\|)\, dQ(x) \leq \int \phi(\|x\|)\, dP(x) \quad \text{for all increasing convex } \phi \colon \mathbb{R} \to \mathbb{R},$$

which follows by Corollary 3.17 since all such ϕ are continuous and $\phi \circ \|(\cdot)\|$ is convex. Using only the one-dimensional version of {Strassen (1965), Theorem 9], it follows that (Q, P) is *submartingalizable*, so the conclusion follows easily from Theorem 4.5. $\square$

If $X = \mathbb{R}^1$ 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 $0 \leq t \leq 1$ (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 van der Vecht (1986) for properties and applications of these functions]. Also, $Q \in \mathscr{P}$ is said to be *smaller in mean residual life* than P if $\int (x - t)^+ Q(dx) \leq \int (x - t)^+ P(dx)$ for all real t. [This ordering has applications in queueing theory; see Stoyan (1983).]

THEOREM 4.7. *If $X = \mathbb{R}^1$ and $P \in \mathscr{P}$ has finite first moment, then the following are equivalent*:

    (i) *Q is a fusion of P;*
   (ii) *$Q \preccurlyeq_c P$;*
  (iii) *(Q, P) is martingalizable;*
  (iv) *there exists a dilation μ of X with $P = \mu Q$;*
   (v) *$H_Q \leq H_P$ and $b(X, P) = b(X, Q)$;*
  (vi) *$U_Q \geq U_P$ 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) *$\int (x \vee t) Q(dx) \leq \int (x \vee t) P(dx)$ for all t, and $b(X, P) = b(X, Q)$;*
  (ix) *$\int_{-\infty}^{t} Q(-\infty, t]\, dt \leq \int_{-\infty}^{t} P(-\infty, t]\, dt$ for all t, and $b(X, P) = b(X, Q)$.*

PROOF. The equivalence of (i)–(iv) follows from the infinite-dimensional result (Theorem 4.1); (iii) ⇔ (vi) follows from Chacon and Walsh [cf. van der Vecht (1986), page 69]; (v) ⇔ (vi) is attributed in van der Vecht [(1986), page 69] to Gilat; and (ii) ⇔ (vii) ⇔ (viii) ⇔ (ix) are in Stoyan [(1983), pages 8–9]. □

REMARKS. In the case where P has support in $[0, \infty)$ and nonzero first moment, the above conditions are also equivalent to "the Lorenz transform of Q is pointwise less than or equal to the Lorenz transform of P and the barycenters are equal," where the Lorenz transform of P, L_P, is $L_P(t) = (b(X, P))^{-1} \int_0^t F^{-1}(s)\, ds$.

The Lorenz ordering has numerous applications in economics as a measure of the distribution of wealth in populations, and many of the other orderings above have extensive application in their finitistic versions; the reader is referred to Marshall and Olkin (1979) and Tong (1980) for the majorization analogs and applications.

If $X = \mathbb{R}^1$, the above results can be used to obtain the following sharp envelope for distribution functions in $\mathscr{F}(P)$.

THEOREM 4.8. *Suppose $X = \mathbb{R}^1$ and that $P \in \mathscr{P}$ has a finite first moment m. Then for all $Q \in \mathscr{F}(P)$,*

$$P\big((-\infty, m_P(x))\big) \leq Q((-\infty, x]) \leq P\big((-\infty, M_P(x)]\big),$$

where for any random variable Y with $\mathscr{L}(Y) = P$,

$$m_P(x) = \inf\{y \colon E(Y|Y \geq y) \geq x\} \quad and \quad M_P(x) = \sup\{y \colon E(Y|Y \leq y) \leq x\}.$$

Moreover these bounds are attained whenever P has no atoms at $m_P(x)$ and $M_P(x)$, respectively.

PROOF. Fix $Q \in \mathcal{F}(P)$. By Theorem 4.7, there exists a martingale (Z_1, Z_2) with $\mathcal{L}(Z_1) = Q$, $\mathcal{L}(Z_2) = P$. The desired inequality then follows easily by Lemma 4.4 (together with an easy calculation to show that the bounds are attained in the nonatomic case). □

EXAMPLE 4.9. Let $X = \mathbb{R}^1$ and P be Lebesgue measure on $[0, 1]$. Then if F is the distribution function of $Q \in \mathcal{F}(P)$, F satisfies:

   (i) $0 \le F(x) \le 2x$ if $0 \le x \le \frac{1}{2}$;
   (ii) $2x - 1 \le F(x) \le 1$ if $\frac{1}{2} \le x \le 1$,

and these bounds are sharp and attained.

REMARKS. A version of Theorem 4.8 which is sharp for *all* P may be obtained by simply taking the P-masses at $m_P(x)$ and $M_P(x)$ into account; this is left as an exercise for the interested reader. It should also be noted that the envelope for $\mathcal{F}(P)$ given in Theorem 4.8 is *pointwise*; as can be seen in Example 4.9, the distribution $F(x) = 2x$ for $x \in [0, \frac{1}{2}]$ does *not* represent a Q in $\mathcal{F}(P)$, since its mean is not $\frac{1}{2}$.

A question raised by Marshall and Olkin (1979) (page 433, converse to B2) is equivalent to the question of whether equivalence of (ii) and (vii) in Theorem 4.7 generalizes to higher dimensions. In other words, are wedge functions (positive parts of affine functions) a determining class of functions for convex domination in $\mathbb{R}^n$ for $n > 1$? The next example shows that they are not, even though they *are* a separating class for $\mathcal{P}$ (Theorem 3.23).

EXAMPLE 4.10. Let $X = \mathbb{R}^2$, $P = (\delta(-2, 0) + \delta(0, -2) + \delta(2, 2) + 3\delta(0, 0))/6$ and $Q = (\delta(-1, -1) + \delta(0, 1) + \delta(1, 0))/3$, let $c_1(x, y) = y \vee 0$, $c_2(x, y) = x \vee 0$ and $c = \max\{c_1, c_2\}$. (Note that c_1 and c_2 are wedge functions, but that c is *not*.) An easy calculation shows that $\int c \, dP = \frac{1}{3} < \frac{2}{3} = \int c \, dQ$, but it will now be shown that $\int w \, dP \ge \int w \, dQ$ for all wedge functions w. Note that if w is a wedge function, $w(x, y) = (ax + by + c)^+$ for some choice of parameters a, b, c. Let $f(a, b, c) = \int (ax + by + c)^+[dP(x, y) - dQ(x, y)]$. It will be shown that $f(a, b, c) \ge 0$ for all a, b, c. It is enough to show this for $|a| + |b| + |c| = 1$, since f is positively homogeneous. Let $\mathcal{A} = \{(x_i, y_i), i = 1, \ldots, 7\}$ be the atoms of P union the atoms of Q. For each subset S of $\{1, 2, \ldots, 7\}$ and $\varepsilon = \{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ with each $\varepsilon_i = 1$ or -1, consider the region of parameter space

$$R_{S, \varepsilon} = \{(a, b, c): ax_i + by_i + c \ge 0 \text{ for } i \in S,$$

$$ax_i + by_i + c \le 0 \text{ for } i \notin S, \varepsilon_1 a \ge 0, \varepsilon_2 b \ge 0, \varepsilon_3 c \ge 0$$

$$\text{and } \varepsilon_1 a + \varepsilon_2 b + \varepsilon_3 c = 1\}.$$

Note that $\{(a, b, c): |a| + |b| + |c| = 1\} = \bigcup_{S, \varepsilon} R_{S, \varepsilon}$, so it is enough to show $f \ge 0$ on each $R_{S, \varepsilon}$. Now $R_{S, \varepsilon}$ is a convex polyhedron and f is an affine

function of a, b, c when restricted to $R_{S, \varepsilon}$, so f takes its minimum on some vertex of $R_{S, \varepsilon}$. The vertices consist of points (a, b, c) such that two of the inequalities in the definition of $R_{S, \varepsilon}$ are replaced by equalities. These correspond to wedge functions whose corner line l: $ax + by + c = 0$ passes through (at least) two points of $\mathscr{A}$, or else passes through (at least) one point of $\mathscr{A}$ and also either $a = 0$ or $b = 0$ [$c = 0$ corresponds to the line passing through $(0, 0)$ which *is* a point of $\mathscr{A}$], or else two of a, b, c are zero. It is obvious for most of these finitely many cases that $f \geq 0$ and very easy (but a little tedious) to compute that $f \geq 0$ in the other cases; the details are omitted.

The analog of this example in the majorization framework of Marshall and Olkin is as follows. Let

$$X = \begin{bmatrix} 0 & 0 & 1 & 1 & -1 & -1 \\ 1 & 1 & 0 & 0 & -1 & -1 \end{bmatrix},$$

$$Y = \begin{bmatrix} 0 & 0 & 0 & -2 & 0 & 2 \\ 0 & 0 & 0 & 0 & -2 & 2 \end{bmatrix}.$$

The question of Marshall and Olkin is whether or not AX majorized by AY for all 1×2 matrices A implies X is majorized by Y. Since $\Sigma c(y_i) = 2 < 4 = \Sigma c(x_i)$, where x_i and y_i are the column vectors of X and Y and c is the convex function in Example 4.10, X is *not* majorized by Y [Marshall and Olkin (1979), page 433, Theorem B1]. However, for every 1×2 matrix A, AX is majorized by AY; this is equivalent to the domination of Q by P above for wedge functions, or may be easily proved directly.

The last result in this section shows that if X is finite-dimensional, the continuity hypotheses in Proposition 3.16 and Corollary 3.17 may be dropped.

PROPOSITION 4.11. *Let K be a closed convex subset of $\mathbb{R}^n$. Then:*

(i) *if μ is a probability measure on K with barycenter b, then $\phi(b) \leq \int \phi(x) \mu(dx)$ for all convex Borel μ-integrable functions ϕ: $K \to \mathbb{R}$;*

(ii) *if there is a dilation ρ of X so $\mu = \rho\nu$, where μ, ν are probability measures on K having barycenters, then $\int \phi \, d\nu \leq \int \phi \, d\mu$ for all convex Borel μ-integrable functions ϕ: $K \to \mathbb{R}$.*

PROOF. Let S be the minimum closed affine subspace of $\mathbb{R}^n$ such that $\mu(S) = 1$, that is, the affine hull of $\mathrm{supp}(\mu)$. [By affine subspaces we mean $x, y \in S \Rightarrow \alpha x + (1 - \alpha)y \in S$ for all $\alpha \in \mathbb{R}$.]

Note that S is the affine hull of $S \cap K$. Now $\mu(K \cap S) = 1$, so by Proposition 2.1, $b \in S \cap K$, since $b = \int_{K \cap S} x\mu(dx)$ and since $K \cap S$ is closed and convex.

If b is an interior point of $K \cap S$ relative to S, then $\phi(b) = f(b)$ for some affine (automatically continuous) functional f: $S \to \mathbb{R}$ for which $f \leq \phi$ on S.

This is because there is always a nonvertical supporting hyperplane at $(b, \phi(b))$ on the epigraph of ϕ [Stoer and Witzgall (1970), page 142]. Thus

$$\phi(b) = f(b) = \int_S f(x)\mu(dx) \le \int_S \phi(x)\mu(dx)$$

as claimed.

If b is a boundary point of $K \cap S$ relative to S, then there is a supporting closed hyperplane H (relative to S) for $K \cap S$ at b. Here, H is a *proper* affine subspace of S [Stoer and Witzgall (1970), page 103; they call the supporting plane *nonsingular* if it does not contain $K \cap S$ and has nonempty intersection with the interior of $K \cap S$; its existence follows from the fact that b is a boundary point relative to the affine hull of $S \cap K$, which for us is just S as we have noted].

Since one side of H, say H^-, contains no point of $K \cap S$, $\mu(H^-) = 0$. Now if $\mu(H^+) > 0$, then

$$b = \int_{H^+} x\mu(dx) + \int_H x\mu(dx) = \mu(H^+)u + \mu(H)v,$$

where $u \in H^+$, $v \in H$ (the fact that $u \in H^+$ follows from the fact that H^+ is convex and we are in finite dimensions, so it does not matter that H^+ is not closed). But since $b \in H$, this would imply $u \in H$ also, a contradiction.

Thus $\mu(H^+) = 0$ also, so $\mu(H) = 1$. But this contradicts the minimality of S, so b cannot be a boundary point of $K \cap S$ (relative S). This proves (i). To prove (ii), $\mu(B) = \int \rho(x, B)\nu(dx)$, where $x = \int y\rho(x, dy)\ \forall\ x$. By (i), $\phi(x) \le \int \phi(y)\rho(x, dy)\ \forall\ x$, so

$$\int \phi(x)\nu(dx) \le \int\int \phi(y)\rho(x, dy)\nu(dx) = \int \phi(y)\mu(dy). \qquad \square$$

REMARKS. An example of Choquet [Alfsen (1971), page 20] shows that (i) can fail in infinite dimensions, even for a bounded *affine* ϕ of the second Baire class, on a compact convex K.

However, (ii) holds in a general Hausdroff l.c.t.v.s., K compact and convex, for all *upper-semicontinuous* ϕ: $K \to \mathbb{R}$ [Alfsen (1971), page 45]. Even in $\mathbb{R}^2$, a convex Borel function on a compact convex set need not be upper-semicontinuous, so this result does not include our result (ii). (To see this, consider the following modification of an example of Stoer and Witzgall [(1970), page 137]. Let $K = \{(x, y) \in \mathbb{R}^2\colon\ 0 \le x \le 1,\ y^2 \le x\}$ and let f: $K \to \mathbb{R}$ be defined by $f(0, 0) = 0$ and $f(x, y) = y^2/x$ for $x > 0$. Then f is convex, but not upper-semicontinuous at $(0, 0)$.)

Recall from Theorem 4.7 that if P (on $X = \mathbb{R}^1$) has a finite first moment, then Q is a fusion of P if and only if (Q, P) is martingalizable. Since the notion of martingale entails existence of first moments, it might be asked

whether relaxing this requirement, but preserving the fairness of the pair yields an equivalent condition to fusion in the general (nonintegrable P) case. That this is not the case [see Gilat (1977) for definition of the fair process generalization of martingale] can be seen by letting P be Cauchy and $Q = \delta(0)$ [which is in $\mathscr{F}(P) = \mathscr{P}$ by Proposition 3.14] and noting that (Q, P) is not a fair pair in the sense of Gilat, since the conditional moment of Z_2 given $Z_1 = 0$ does not exist.

5. Optimal distributions and fusions for cost–reward problems.
Many ideal physical laws describe linear mixtures of fusions or various types; one such law for mixtures of concentrations was given in the Introduction, and another is Raoult's law of physical chemistry [cf. Barrow (1979), page 279]—"the vapor pressure of the component of an ideal solution is proportional to the mole factor of the component." The main purpose of this section is to apply some of the above fusion results to an applied problem related to such physical laws.

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 $Q \in \mathscr{F}(P)$, which will maximize the average profit $\int r\, dQ - \int c\, dP$?

Throughout this section, it will be assumed that X is a compact convex subset of $\mathbb{R}^n$ (although clearly analogs of these results are possible for the infinite-dimensional case).

DEFINITION 5.1. For Borel functions $r, c \colon X \to \mathbb{R}$, (Q, P) is (r, c)-*optimal* if $\int r\, dQ - \int c\, dP = \sup\{\int r\, d\hat{Q} - \int c\, d\hat{P} \colon \hat{P} \in \mathscr{P}, \hat{Q} \in \mathscr{F}(\hat{P})\}$.

The first result covers the relatively easy case when the cost function c is lower-semicontinuous and convex: It simply says that optimality in this case is attained by producing everything deterministically at some optimal level x^* and not fusing at all.

THEOREM 5.2. *Suppose $r \colon X \to \mathbb{R}$ is upper-semicontinuous and $c \colon X \to \mathbb{R}$ is lower-semicontinuous and convex. Then $(\delta(x^*), \delta(x^*))$ is (r, c)-optimal, where x^* is any vector satisfying $r(x^*) - c(x^*) = \max\{r(x) - c(x) \colon x \in X\}$.*

PROOF. Fix $P \in \mathscr{P}$ and $Q \in \mathscr{F}(P)$. Then

$$\int r\, dQ - \int c\, dP \le \int r\, d\hat{Q} - \int c\, dQ = \int (r - c)\, dQ$$

$$\le r(x^*) - c(x^*) = \int r\, d\delta(x^*) - \int c\, d\delta(x^*),$$

where the first inequality follows by Theorems 4.1 and 4.11(ii) [since $Q \in \mathscr{F}(P)$], the second inequality follows since Q is a probability distribution and the existence of such an x^* follows since $r - c$ is upper semicontinuous. $\square$

DEFINITION 5.3. For a function $f: X \to \mathbb{R}$, let $\check{f}$ denote the *convex closure* of f, that is, $\check{f}(x) = \sup\{g(x)|g: X \to \mathbb{R}, g \text{ is convex and } g \leq f\}$.

THEOREM 5.4. *Suppose* $r, c: X \to \mathbb{R}$ *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* $r(x^*) - \check{c}(x^*) = \max\{r(x) - \check{c}(x): x \in X\};$ *and* $\{(x_j^*, \check{c}(x_j^*))\}_{j=1}^n$ *are any extreme points of the convex set* $\{(x, y) \in \mathbb{R}^{n+1}: x \in X, y \in \mathbb{R}, y \geq \check{c}(x)\}$ *which satisfy* $\sum_{j=1}^{n+1}\alpha_j(x_j^*, \check{c}(x_j^*)) = (x^*, \check{c}(x^*))$ *for some* $\{\alpha_j\}_1^{n+1} \geq 0, \sum_{j=1}^{n+1}\alpha_j = 1$.

PROOF. First observe that since X is compact and $r - \check{c}$ is upper-semicontinuous, that an x^* exists which maximizes $r - \check{c}$. By Caratheodory's theorem, any point $(x^*, \check{c}(x^*)) \in \mathbb{R}^{n+1}$ can be written as a convex combination of at most $n + 1$ extreme points of the set $\{(x, y) \in \mathbb{R}^{n+1}: x \in X, y \in \mathbb{R}, y \geq \check{c}(x)\}$, so there exist $\{\alpha_j\}_1^{n+1} \geq 0, \sum_{j=1}^{n+1}\alpha_j = 1$ and $\{x_j^*\}_1^{n+1}$ satisfying

$$\sum_{j=1}^{n+1} \alpha_j\left(x_j^*, \check{c}(x_j^*)\right) = (x^*, \check{c}(x^*)).$$

Fix $P \in \mathscr{P}$ and $Q \in \mathscr{F}(P)$. Then

$$\int r\,dQ - \int c\,dP \leq \int r\,dQ - \int \check{c}\,dP \leq r(x^*) - \check{c}(x^*)$$

$$= \int r\,d\delta(x^*) - \sum_1^{n+1} \alpha_j\check{c}(x_j^*)$$

$$= \int r\,d\delta(x^*) - \sum_1^{n+1} \alpha_j c(x_j^*)$$

$$= \int r\,d\delta(x^*) - \int c\,d\left(\sum_1^{n+1} \alpha_j\delta(x_j^*)\right),$$
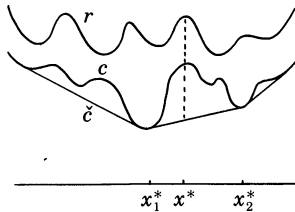

FIG. 1. $(n = 1)$ $\alpha_1 = (x^* - x_1^*)/(x_1^* - x_2^*)$, $P = \alpha_1\delta(x_1^*) + (1 - \alpha_1)\delta(x_2^*)$ *and* $Q = \delta(x^*) \in \mathscr{F}(P)$.

where the first inequality follows since $\check{c} \leq c$, the second by Theorem 5.2 and the second equality since for extreme points x_j^*, $c(x_j^*) = \check{c}(x_j^*)$. Since x^* is the barycenter of the measure $P = \sum_1^{n+1} \alpha_j \delta(x_j^*)$, clearly $\delta(x^*) \in \mathscr{F}(P)$, which completes the proof. $\square$

A typical construction of an (r, c)-optimal pair (Q, P) in the one-dimensional case is shown in Figure 1.

Acknowledgments. The authors are grateful to Professors Alberto Grünbaum and Ronald Fox for several suggestions concerning the physical laws of mixtures, to Professor David Gilat for helpful conversations concerning the relationship of fusion to the notions of balayage and martingalizability and to the referee for several suggestions and corrections.

REFERENCES

ALFSEN, E. (1971). *Compact Convex Sets and Boundary Integrals.* Springer, New York.

BARROW, G. (1979). *Physical Chemistry.* McGraw-Hill, New York.

BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

BLACKWELL, D. and DUBINS, L. (1963). A converse to the dominated convergence theorem. *Illinois J. Math.* **7** 508–514.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

BILLINGSLEY, P. (1986). *Probability and Measure.* Wiley, New York.

CARTIER, P., FELL, J. and MEYER, P. (1964). Comparison des mesures portées par un ensemble convexe compact. *Bull. Soc. Math. France* **92** 435–455.

DIESTEL, J. and UHL, J. (1977). *Vector Measures. Math. Surveys* **15**. Amer. Math. Soc., Providence, R.I.

DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators, Part I.* Interscience, New York.

ELTON, J. and HILL, T. (1991). Convex domination, dilation, and fusion of general finite measures. Unpublished manuscript.

ENFLO, P. (1973). A counterexample to the approximation problem in Banach spaces. *Acta Math.* **103** 309–317.

GILAT, D. (1977). Every nonnegative submartingale is the absolute value of a martingale. *Ann. Probab.* **5** 475–481.

GROTHENDIECK, A. (1955). Produits tensoriels topologiques et espaces nucleaires. *Mem. Amer. Math. Soc.* **16**.

HARDY, G., LITTLEWOOD, J. and PÓLYA, G. (1929). Some simple inequalities satisfied by convex functions. *Messenger Math.* **58** 145–152.

HARDY, G., LITTLEWOOD, J. and PÓLYA, G. (1959). *Inequalities.* Cambridge University Press.

KALTON, N., PECK, N. and ROBERTS, J. (1984). *An F-Space Sampler.* Cambridge Univ. Press.

KAMAE, T. and KRENGEL, U. (1978). Stochastic partial ordering. *Ann. Probab.* **6** 1044–1049.

MARSHALL, A. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

MEYER, P. (1966). *Probability and Potentials.* Blaisdell, Waltham, Mass.

PHELPS, R. (1966). *Lectures on Choquet's Theorem.* Van Nostrand, New York.

RUDIN, W. (1973). *Functional Analysis.* McGraw-Hill, New York.

SCHAEFER, H. (1971). *Topological Vector Spaces*. Springer, New York.
STOER, J. and WITZGALL, C. (1970). *Convexity and Optimization in Finite Dimensions I*. Springer, New York.
STOYAN, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. Wiley, New York.
STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.
TONG, Y. (1980). *Probability Inequalities in Multivariate Distributions*. Academic, New York.
VAN DER VECHT, D. (1986). *Inequalities for Stopped Brownian Motion*. CWI Tract **21**. Math. Centrum, Amsterdam.
WILANSKY, A. (1978). *Modern Methods in Topological Vector Spaces*. McGraw-Hill, New York.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332