1997

Some Problems in Probabilistic Tomography

D Applegate

J Reeds

S Scheinberg

Larry A. Shepp
University of Pennsylvania

P Shor

Recommended Citation

Applegate, D., Reeds, J., Scheinberg, S., Shepp, L. A., & Shor, P. (1997). Some Problems in Probabilistic Tomography. *Theory of Probability & Its Applications, 41* (2), 199-209. http://dx.doi.org/10.1137/S0040585X97975435

Some Problems in Probabilistic Tomography

Abstract

Given probability distributions F_1 , F_2 , . . ., F_k on $\mathbf{R}$ and distinct directions θ_1, . . ., θ_k, one may ask whether there is a probability measure μ on $\mathbf{R}^2$ such that the marginal of μ in direction θ_j is F_j, j = 1, . . ., k. For example for k = 3 we ask what the marginal of μ at 45° can be if the *x* and *y* marginals are each say standard normal? In probabilistic language, if *X* and *Y* are each standard normal with an *arbitrary* joint distribution, what can the distribution of X + Y or X - Y be? This type of question is familiar to probabilists and is also familiar (except perhaps in that μ is positive) to tomographers, but is difficult to answer in special cases. The set of distributions for Z = X - Y is a convex and compact set, *C*, which contains the single point mass Z ≡ 0 since X ≡ Y, standard normal, is possible. We show that *Z* can be 3-valued, Z=0, ±a for any *a*, each with positive probability, but *Z cannot* have any (genuine) two-point distribution. Using numerical linear programming we present convincing evidence that *Z* can be uniform on the interval [-ε, ε] for ε small and give estimates for the largest such ε. The set of all extreme points of *C* seems impossible to determine explicitly.

We also consider the more basic question of finding the extreme measures on the unit square with uniform marginals on both coordinates, and show that not every such measure has a support which has only one point on each horizontal or vertical line, which seems surprising.

Keywords

marginal distributions, extreme point, Radon

Disciplines

Probability | Statistics and Probability

Theory Probab. Appl. Translated from Russian Journal
Vol. 41, No. 2

SOME PROBLEMS IN PROBABILISTIC TOMOGRAPHY*

D. APPLEGATE†, J. REEDS†, S. SCHEINBERG‡, L. SHEPP†, AND P. SHOR†

Abstract. Given probability distributions $F_1, F_2, \ldots, F_k$ on $\mathbf{R}$ and distinct directions $\theta_1, \ldots, \theta_k$, one may ask whether there is a probability measure μ on $\mathbf{R}^2$ such that the marginal of μ in direction θ_j is F_j, $j = 1, \ldots, k$. For example for $k = 3$ we ask what the marginal of μ at 45° can be if the x and y marginals are each say standard normal? In probabilistic language, if X and Y are each standard normal with an *arbitrary* joint distribution, what can the distribution of $X + Y$ or $X - Y$ be? This type of question is familiar to probabilists and is also familiar (except perhaps in that μ is positive) to tomographers, but is difficult to answer in special cases. The set of distributions for $Z = X - Y$ is a convex and compact set, C, which contains the single point mass $Z \equiv 0$ since $X \equiv Y$, standard normal, is possible. We show that Z can be 3-valued, $Z = 0$, $\pm a$ for any a, each with positive probability, but Z *cannot* have any (genuine) two-point distribution. Using numerical linear programming we present convincing evidence that Z can be uniform on the interval $[-\varepsilon, \varepsilon]$ for ε small and give estimates for the largest such ε. The set of all extreme points of C seems impossible to determine explicitly.

We also consider the more basic question of finding the extreme measures on the unit square with uniform marginals on both coordinates, and show that not every such measure has a support which has only one point on each horizontal or vertical line, which seems surprising.

Key words. marginal distributions, extreme point, Radon

PII. S0040585X97975435

1. Introduction. Tomography [8] deals with the question of invertibility of the Radon transform of an integrable function f on $\mathbf{R}^2$, i.e., given the line integrals of f, find f. A basic notion of probability theory is that of marginal distribution, which is nothing but the line integrals of the density or measure in one direction. The probabilistic approach has already shed some light on the question of uniqueness of the Radon transform by the theorem of [4] that for any function f on $\mathbf{R}^n$, $0 \leqslant f(x) \leqslant 1$, there is a function g, which takes only 2 values, 0 and 1, for which f and g have exactly the same Radon transform for any line having any one of a finite number of directions. This shows that in the ordinary tomographic framework, the set of all densities g with any n projections (or marginals) contains many elements. By a recent theorem [6] such functions f and g are nearly identical after appropriate smoothing by some kernel.

Logan [10], building on a technique of De Acosta using jointly stable distributions [1], [5], showed there is a pair of standard Cauchy variables X and Y, $\mathbf{P}\{X \in dx\} = \mathbf{P}\{Y \in dx\} = dx/\pi(1 + x^2)$, $-\infty < x < \infty$, for which $X + Y$ is also Cauchy but is centered at $a > 0$, and gave upper and lower bounds on the maximum value of a. The De Acosta–Logan phenomenon is possible only because Cauchy variables have no mean.

Another result in the overlap between probability and tomography, shows that in delicate cases when the unknown function f is restricted to satisfy inequalities, f may be uniquely determined by only two marginals. Indeed, [3] showed that if g is any function on $\mathbf{R}^2$ with $0 \leqslant g \leqslant 1$ and if g has the same line integrals as does f,

*Received by the editors July 4, 1994.
http://www.siam.org/journals/tvp/41-2/97543.html

†AT&T Bell Laboratories, Murray Hill, NJ 07974.
‡Department of Mathematics, University of California at Irvine, Irvine, CA 92697.

the indicator of the unit disk, for lines which are either horizontal or vertical, then $f = g$ a.e.

Here we study another specific question belonging to the intersection of probability and tomography. We would like to describe the class, C, of all distributions F for $Z = X + Y$ when X and Y each separately have a standard normal distribution function (d.f.). Not much more can apparently be said about C besides the general statements that it is a compact and convex set of zero mean distributions. C is alternatively defined by the (hard to apply in the positive direction) duality condition that

$$\int h dF \geqslant 0 \quad \text{whenever } h(x+y) \geqslant \xi(x) + \xi(y) \quad \text{for all } x \text{ and } y, \tag{1.1}$$

where ξ is any bounded measurable function for which $\mathbf{E}\xi(\eta) = 0$ when η is a standard normal random variable (r.v.). To see that (1.1) is necessary for F to be the distribution of an $F \in C$ note that $h(X+Y) \geqslant \xi(X) + \xi(Y)$ so that (1.1) follows since $\xi(X)$ and $\xi(Y)$ have zero mean. The converse is also true but we omit the proof, because: (a) it's apparently not easy, (b) we do not use the converse assertion, (c) it's well known, and (d) we do not really have a proof.

Remark 1. That the set C of all distributions of $Z = X + Y$, where X and Y are normal, is convex follows using mixing: if $Z_1 = X_1 + Y_1$ and $Z_2 = X_2 + Y_2$, then $Z = X + Y$, where $(X, Y) = (X_j, Y_j)$ with probability $\frac{1}{2}$, $j = 1, 2$, has the average of the distributions of Z_1 and Z_2. That C is also compact in the weak topology is easy to see. It follows from the Krein–Milman theorem that C has extreme points. Since $-Y$ is also standard normal, the set C is also defined by $Z = X - Y$.

Remark 2. Any normal $\mathcal{N}(0, \sigma^2)$ with $\sigma \leqslant 2$ belongs to C because we can take X and Y jointly normal. The distribution $\mathcal{N}(0, 2)$ is an extreme point of C because if it is a mixture of others then each must have variance 4, but if $\mathbf{E}(X+Y)^2 = 4$ then $\mathbf{E}XY = 1$ and so $X \equiv Y$ since equality holds in Schwartz's inequality so that $\mathcal{N}(0, 2)$ is extreme. This gives an extreme point with full support, $\mathbf{R}$.

We show in section 2 that if $Z \in C$ (we allow ourselves to talk of a r.v. being in C if its distribution is in C) then

$$\mathbf{P}\big\{|Z| < a + \varepsilon\big\} > 0 \quad \text{for } a = \Phi^{-1}\left(\frac{3}{4}\right) - \frac{1}{2}, \text{ for } \varepsilon > 0$$

and show that this a is best possible. That is, the support of a distribution in C must intersect $(-a, a)$ for $a > \Phi^{-1}(\frac{3}{4}) - \frac{1}{2}$, where Φ is the standard normal cumulative distribution function (d.f.).

We show in section 2 that Z cannot have a genuine two-point distribution although it may have a three-point distribution; for any $a > 0$, Z may be concentrated on $\{-a, 0, +a\}$ with $\mathbf{P}\{Z = 0\} < 1$ and $\mathbf{P}\{Z = a\} = \mathbf{P}\{Z = -a\}$. We find the minimum value of $\mathbf{P}\{Z = 0\}$ among all $Z \in C$ supported on $\{-a, 0, +a\}$ and note that this is an extreme point of C for each $a > 0$.

We give a nonrigorous but very convincing demonstration based on numerical linear programming that $Z = U(-\varepsilon, \varepsilon)$, i.e., the uniform distribution on the interval $(-\varepsilon, \varepsilon)$ belongs to C for $\varepsilon > 0$ sufficiently small. The output of the approximating linear programming problem indicates that no explicit realization of $U(-\varepsilon, \varepsilon)$ as $X+Y$ with X and Y normal is likely to be found.

Finally in section 3, we consider the perhaps more basic problem to find the set of extreme points of the set S of measures μ on the unit square whose x and y marginals are each uniform on $(0, 1)$. It is easy to see that μ may be supported on the graph of any one-one measurable function and that such a μ is extreme in S. However, we show by example that there are extreme measures in S whose support is a set which intersects a.e. horizontal and vertical line in at least two points.

2. Measures in C with small support. We first show that except for $Z \equiv 0$, no Z concentrated on only two points can belong to C, i.e., one cannot have a measure on $\mathbf{R}^2$ supported on two diagonal lines whose x and y marginals are standard normal. For clarity, we first prove this for the points ± 1, i.e., that $Z = \pm 1$ each with probability $\frac{1}{2}$ does *not* belong to C. The proof is easy: if $X + Y = \pm 1$, then on the sample space where X and Y are defined, the r.v.

$$W = \cos \pi X + \cos \pi Y \equiv 0 \tag{2.1}$$

since $\cos(u \pm \pi) = -\cos u$. But the mean value of W is $2\mathbf{E} \cos \pi X = 2 \exp(-\pi^2/2)$ which is not zero. This proof is really an application of condition (1.1). In the general case of $Z = a, b,\ b \neq a$,

$$W = \cos \frac{2\pi X}{b-a} + \cos \frac{2\pi}{b-a} \left(Y - \frac{b+a}{2} \right) \equiv 0 \tag{2.2}$$

while

$$\mathbf{E}W = \exp\left(-\frac{2\pi^2}{(b-a)^2} \right) \left(1 + \cos \pi \left(\frac{b+a}{b-a} \right) \right) \neq 0 \tag{2.3}$$

since we may assume $a < 0 < b$ because Z has mean zero and the cosine does not $= -1$ in $(-\pi, \pi)$.

We next show that for any $a > 0$ there is a pair of standard normal r.v.'s X and Y for which $Y = X - a,\ X,\ X + a$ everywhere, i.e., $Z = Y - X$ is concentrated only on $\{-a, 0, a\}$; a similar construction is possible of course for $Z = X + Y$. Our construction will be explicit and will make $\mathbf{P}\{X = Y\}$ as small as possible. We have seen in section 1 that $X \equiv Y$ is possible and for each a our construction gives an extreme point of C.

Suppose we can find a *nonnegative* function $p(x)$, $-\infty < x < \infty$, for which

$$p(x) + p(x-a) \leqslant \varphi(x) \equiv \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right), \quad -\infty < x < \infty. \tag{2.4}$$

Then we define the measure μ supported on the three lines $y = x - a$, $y = x$, $y = x + a$ as follows:

$$\begin{aligned} \mu\big\{(x,\, y)\colon\ y = x - a, x \in dx\big\} &= p(x-a)\, dx, \\ \mu\big\{(x,\, y)\colon\ y = x, x \in dx\big\} &= \big(\varphi(x) - p(x) - p(x-a)\big)\, dx, \\ \mu\big\{(x,\, y)\colon\ y = x + a, x \in dx\big\} &= p(x)\, dx. \end{aligned} \tag{2.5}$$

Adding the three lines in (2.5) we get $\mu\{(x, y)\colon\ x \in dx\} = \varphi(x)dx$ so that the x-marginal of μ is standard normal. Similarly, from (2.5),

$$\begin{aligned} \mu\big\{(x,\, y)\colon\ y \in dy\big\} &= p(y)\, dy + \big(\varphi(y) - p(y) - p(y-a)\big)\, dy + p(y-a)\, dy \\ &= \varphi(y)\, dy \end{aligned} \tag{2.6}$$

so that the y-marginal of μ is also $\mathcal{N}(0,1)$. Clearly, μ is positive by (2.4), and since μ is concentrated on the lines $y-x=-a,0,a$, it is clear that $Z=Y-X$ takes only the values $-a,0,a$. We want to find the p in (2.4) with the largest integral in order to make $\mathbf{P}\{Z=0\}=1-2\int_{-\infty}^{\infty}p(x)dx$ a minimum.

We need the following lemma, perhaps of interest in itself, which solves a linear programming problem.

LEMMA 1. *Suppose $\varphi_n>0$ is a unimodal sequence, $-\infty<n<\infty$, $\sum\varphi_n<\infty$, and we seek p_n, $-\infty<n<\infty$, with $p_n>0$ and $p_n+p_{n-1}\leqslant\varphi_n$ which will maximize $\sum p_n$. Then this maximum is either the even or the odd sum of φ:*

$$\sup\sum p_n=\min\left(\sum\varphi_{2n},\sum\varphi_{2n+1}\right). \tag{2.7}$$

Proof. Since $p_{2n}+p_{2n-1}\leqslant\varphi_{2n}$ for all n, $\sum p_n\leqslant\sum\varphi_{2n}$ and similarly $p_{2n+1}+p_{2n}\leqslant\varphi_{2n+1}$ so that $\sum p_n$ can be no larger than the right side of (2.7). To show this can be attained, we may suppose that the mode of φ is at zero.

Case 1: $\sum\varphi_{2n}\leqslant\sum\varphi_{2n+1}$ (later we will consider the opposite case). Let

$$p_n=\begin{cases}\varphi_n-\varphi_{n-1}+\varphi_{n-2}-\varphi_{n-3}+\cdots\geqslant 0 & \text{for } n\leqslant 0,\\ \varphi_{n+1}-\varphi_{n+2}+\varphi_{n+3}-\varphi_{n+4}+\cdots\geqslant 0 & \text{for } n>0.\end{cases} \tag{2.8}$$

Then $p_n+p_{n-1}=\varphi_n$ for $n\leqslant 0$ and $p_n+p_{n+1}=\varphi_{n+1}$ for $n>0$ by (2.8). Also $p_1+p_0\leqslant\varphi_1$ because $\sum\varphi_{2n}\leqslant\sum\varphi_{2n+1}$ so p works.

Case 2: $\sum\varphi_{2n}>\sum\varphi_{2n+1}$. Then

$$p_n=\begin{cases}\varphi_n-\varphi_{n-1}+\varphi_{n-2}-\varphi_{n-3}+\cdots>0 & \text{for } n<0,\\ \varphi_{n+1}-\varphi_{n+2}+\varphi_{n+3}-\varphi_{n+4}+\cdots\geqslant 0 & \text{for } n\geqslant 0,\end{cases} \tag{2.9}$$

works in a similar way. This proves the lemma.

Remark. The proof of the lemma generalizes to show that if the constraints on p are $p_n+\cdots+p_{n-r}\leqslant\varphi_n$ and $p_n\geqslant 0$, $-\infty<n<\infty$, for $r\geqslant 2$ then

$$\sup\sum p_n=\min_{0\leqslant k<r}\sum\varphi_{rn+k}. \tag{2.10}$$

Now suppose that μ is any measure supported on the three lines $y=x-a$, $y=x$, $y=x+a$, which has standard normal marginals. Then μ is easily seen to have a density with respect to linear Lebesgue measure on the lines so that there are nonnegative functions p_+, p_0, p_0 for which

$$\begin{aligned}\mu\big\{(x,y)\colon\ y=x+a,\ x\in dx\big\}&=p_+(x)\,dx,\\ \mu\big\{(x,y)\colon\ y=x,\ x\in dx\big\}&=p_0(x)\,dx,\quad -\infty<x<\infty,\\ \mu\big\{(x,y)\colon\ y=x-a,\ x\in dx\big\}&=p_-(x)\,dx.\end{aligned} \tag{2.11}$$

We want to minimize $\int_{p_0}(x)dx=\mu\{x=y\}$. The marginal condition gives

$$\begin{aligned}p_+(x)+p_0(x)+p_-(x)&=\varphi(x),\\ p_+(x-a)+p_0(x)+p_-(x+a)&=\varphi(x)\end{aligned} \tag{2.12}$$

and subtracting we obtain $p_+(x)-p_+(x-a)=p_-(x+a)-p_-(x)$. If we add this with x replaced by $x-a$, $x-2a$, $x-3a,\ldots$ we get $p_+(x)=p_-(x+a)$, $-\infty<x<\infty$, so that we need $p_+(x) + p_+(x-a) + p_0(x) = \varphi(x)$, $-\infty < x < \infty$. Since $p_0(x) \geqslant 0$ and this takes up the slack, we want to have

$$p_+(x) + p_+(x-a) \leqslant \varphi(x), \quad -\infty < x < \infty,$$

and to maximize $\int p_+(x)dx$. By the lemma, since $\varphi(x+na)$ is unimodal, for each x

$$\sum_{n=-\infty}^{\infty} p_+(x+na) \leqslant \min\left(\sum_n \varphi(x+2na),\ \sum_n \varphi(x+a+2na)\right)$$

and equality holds for the choice of p_+ given in the proof of the lemma. It follows easily that

$$\sup_{p_+} \int p_+(x)\,dx = \int_0^a \min\left(\sum_{n=-\infty}^{\infty} \varphi(x+2na), \sum_{n=-\infty}^{\infty} \varphi(x+a+2na)\right) dx.$$

It is known [2, p. 341] that the standard normal density φ satisfies

$$\text{(2.13)} \qquad \begin{cases} \sum_{-\infty}^{\infty} \varphi(x+na)(-1)^n \geqslant 0 & \text{if } 0 \leqslant x < \frac{a}{2}, \\ \qquad\qquad\qquad\qquad\quad \leqslant 0 & \text{if } \frac{a}{2} < x < a, \end{cases}$$

so that

$$\begin{aligned} \sup_{p_+} \int p_+(x)\,dx &= \int_0^{a/2} \sum_{-\infty}^{\infty} \varphi(x+2na)\,dx + \int_{a/2}^{a} \sum_{-\infty}^{\infty} \varphi(x+a+2na)\,dx \\ &= \sum_{n=-\infty}^{\infty} \Phi\left(\frac{a}{2}+2na\right) - \Phi(2na) + \Phi\big((2n+1)\,a\big) - \Phi\left(2n+\frac{3a}{2}\right) \\ \text{(2.14)} \qquad &= \sum_{n=-\infty}^{\infty} \left(\Phi(na) - \Phi\left(\left(n-\frac{1}{2}\right)a\right)\right)(-1)^{n-1} \end{aligned}$$

which answers the question as to what is the smallest value of $\mathbf{P}\{X = Y\}$ if X and Y are standard normal and $Y = X$ or $Y = X \pm a$. Note that as $a \to \infty$, the right side of (2.14) is very small so $\mathbf{P}\{X = Y\}$ must be close to one, while as $a \to 0$ the right side of (2.14) is nearly $\frac{1}{2}$ so that $\mathbf{P}\{X = Y\}$ can be made very small.

We next show that no $Z \in C$ can omit a symmetric interval about 0 of length greater than $a = 2\Phi^{-1}(\frac{3}{4}) - 1$ and that this a is best possible. Let f: $[0,1] \to [0,1]$ be defined by

$$\text{(2.15)} \qquad f(x) = \begin{cases} x + \frac{1}{2}, & 0 \leqslant x \leqslant \frac{1}{2}, \\ x - \frac{1}{2}, & \frac{1}{2} < x \leqslant 1, \end{cases}$$

and note that if θ is any r.v. uniform on $[0,1]$, so is $f(\theta)$. If X is any standard normal r.v. and we set $Y = \Phi^{-1}(f(\Phi(X)))$, $\Phi(x) = \int_{-\infty}^{x} \varphi$, then $\Phi(X)$ is uniform on $[0,1]$, and so is $f(\Phi(X))$, and hence Y is again standard normal. Since $\big|f(x) - x\big| \geqslant \frac{1}{2}$ everywhere, we see that

$$\text{(2.16)} \qquad \big|\Phi(Y) - \Phi(X)\big| \geqslant \frac{1}{2}$$

on the space Ω on which X is defined. But (2.16) implies that $|Y - X| \geqslant 2a - 1$, where $\Phi(a) = \frac{3}{4}$, since the closest that x and y can be to each other when $|\Phi(y) - \Phi(x)| \geqslant \frac{1}{2}$

is when $y = -x = \pm(\Phi^{-1}(\frac{3}{4}) - \frac{1}{2})$. This proves that there exist X, Y such that $Z = Y - X$ omits the interval $(-a, a)$ with $a = \Phi^{-1}(\frac{3}{4})$.

Suppose that X and Y are *any* pair of $\mathcal{N}(0,1)$ r.v.'s on some space. Then for $\varepsilon > 0$,

$$\begin{aligned} &\mathbf{P}\left\{X \in \left(\Phi^{-1}\left(\frac{1}{4} - \varepsilon\right),\ \Phi^{-1}\left(\frac{3}{4} + \varepsilon\right)\right)\right\} = \frac{1}{2} + 2\varepsilon, \\ &\mathbf{P}\left\{Y \in \left(\Phi^{-1}\left(\frac{1}{4} - \varepsilon\right),\ \Phi^{-1}\left(\frac{3}{4} + \varepsilon\right)\right)\right\} = \frac{1}{2} + 2\varepsilon. \end{aligned} \tag{2.17}$$

But the two events in (2.17) must overlap by 4ε in measure so there is a set of measure 4ε on which both X and Y lie in the interval $(\Phi^{-1}(\frac{1}{4} - \varepsilon), \Phi^{-1}(\frac{3}{4} + \varepsilon))$. But if $\mathbf{P}\{|X - Y| \geqslant a\} = 1$ then a must be smaller that $\Phi^{-1}(\frac{3}{4} + \varepsilon) - \Phi^{-1}(\frac{1}{4} - \varepsilon) \to \Phi^{-1}(\frac{3}{4}) - \Phi^{-1}(\frac{1}{4})$ as $\varepsilon \downarrow 0$ so the example omitting $(\Phi^{-1}(\frac{1}{4}), \Phi^{-1}(\frac{3}{4}))$ is best possible as claimed.

The last question on the smallness of the support of $Z = Y - X$ where X and Y are $\mathcal{N}(0,1)$ is whether Z can be uniform on $(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. Since $\mathbf{E}Z^2 \leqslant 4$ we see that $\varepsilon \leqslant \sqrt{12}$ if $U(-\varepsilon, \varepsilon) \in C$. We can get a slightly better upper bound using (1.1). Indeed if we set $f(x) = \cos \alpha x$ then

$$\begin{aligned} f(y) + f(y - \theta) &= \cos \alpha y\,(1 + \cos \alpha\theta) + \sin \alpha y\, \sin \alpha\theta \\ &\leqslant \sqrt{(1 + \cos \alpha\theta)^2 + \sin^2 \alpha\theta} = 2\left|\cos \frac{\alpha\theta}{2}\right| \end{aligned}$$

so setting $y = Y$, $\theta = Z = Y - X$

$$f(X) + f(Y) \leqslant 2\left|\cos \frac{\alpha}{2} Z\right|. \tag{2.18}$$

But if $Z = U(-\varepsilon, \varepsilon)$ then taking expectations in (2.18), we get

$$2 \exp\left(-\frac{\alpha^2}{2}\right) \leqslant 2\frac{1}{\varepsilon}\int_0^\varepsilon \left|\cos \frac{\alpha}{2} u\right| du \tag{2.19}$$

and setting $\alpha\varepsilon = \pi$ gives a better bound than $\sqrt{12}$,

$$\varepsilon \leqslant \frac{\pi}{\sqrt{2 \log \frac{\pi}{2}}} = 3.31 \cdots < \sqrt{12} = 3.46.$$

A better upper bound however is $\varepsilon \leqslant 3.18845$ which can be obtained by numerical linear programming as described below. We are unable to find ε exactly or to even prove rigorously that $U(-\varepsilon, \varepsilon) \in C$ for ε sufficiently small. Indeed the output of the linear program for realizing the (X, Y) distribution at the largest value of ε (see Fig. 1) indicates that this distribution is extremely complicated and is unlikely to be expressible in explicit form. The converse of (1.1) allows us to restate $U(-\varepsilon, \varepsilon) \in C$ in the following equivalent (but apparently useless) form: $U(-\varepsilon, \varepsilon) \in C$ if for all bounded measurable F

$$2\int F(x)\, \varphi(x)\, dx \leqslant \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \sup_{-\infty < x < \infty} \big(F(x) + F(x + \theta)\big)\, d\theta. \tag{2.20}$$

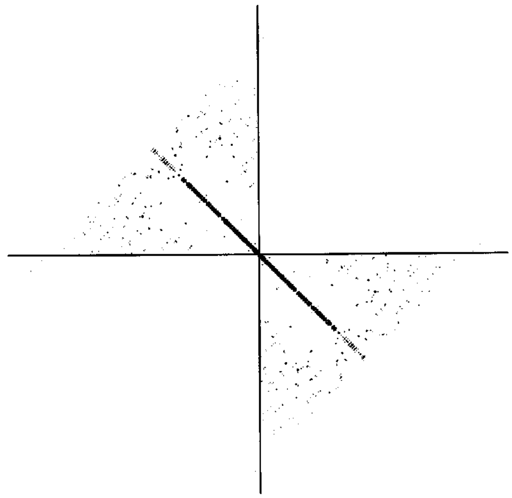

FIG. 1.

Indeed if we let $a = \{f(x) + g(y) + h(y - x)\}$ be the set of bounded measurable $a = a(x, y)$ mapping $\mathbf{R}^2$ to $\mathbf{R}$ of the form

$$a(x, y) = f(x) + g(y) + h(y - x) \tag{2.21}$$

as f, g, h run through the set of bounded measurable functions, then if we define the functional L on a, by $L(1) = 1$, and for a in (2.21),

$$L(a) = \int \big(f(x) + g(x)\big)\, \varphi(x)\, dx + \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} h(u)\, du \tag{2.22}$$

we may verify that $L(a) \geqslant 0$ if $a \geqslant 0$. The Hahn–Banach theorem allows us to extend L to the set of all bounded measurable a which then gives a positive measure μ for which $L(a) = \int a\, d\mu$. This μ gives a measure on $\mathbf{R}^2$ and so $Z(x, y) = y - x$ under μ then realizes $U(-\varepsilon, \varepsilon)$. Unfortunately, (2.20) seems no easier to use as a way to directly construct Z than any other method. We resorted to numerical linear programming described as follows.

From (1.1) and (2.20), $U(-\varepsilon, \varepsilon) \notin C$ if there exist bounded measurable functions F and G such that

$$\begin{gathered} \int_{-\infty}^{\infty} F(x)\varphi(x)dx = 0, \qquad \int_{-\varepsilon}^{\varepsilon} G(\theta)d\theta < 0, \\ F(x) + F(x + \theta) \leqslant G(\theta), \quad -\varepsilon \leqslant \theta \leqslant \varepsilon, \quad -\infty < x < \infty. \end{gathered} \tag{2.23}$$

Without loss of generality, we can assume $F(x) = F(-x)$ and $G(\theta) = G(-\theta)$ and replace (2.23) by

$$\begin{gathered} \int_{0}^{\infty} F(x)\, \varphi(x)\, dx = 0, \qquad \int_{0}^{\varepsilon} G(\theta)\, d\theta < 0, \\ F\big(|x|\big) + F(x + \theta) \leqslant G(\theta), \quad 0 \leqslant \theta \leqslant \varepsilon, \quad -\theta < x < \infty. \end{gathered} \tag{2.24}$$

This is clearly an infinite linear program. We obtained a finite approximation by selecting a set of breakpoints $0 = x_0, x_1, \ldots, x_n$ and requiring F to be piecewise linear with these breakpoints, i.e.,

$$F(x) = \begin{cases} \dfrac{f_i(x_{i+1} - x) + f_{i+1}(x - x_i)}{x_{i+1} - x_i}, & x_i \leqslant x \leqslant x_{i+1}, \\ f_n, & x \geqslant x_n. \end{cases}$$

In this case $G(\theta)$ is also piecewise linear with the same breakpoints. Thus (2.24) becomes a (finite) linear program on the variables $f_0, \ldots, f_n, g_0, \ldots, g_n$. We set $x_i = \frac{i}{n} x_n$ and considered a range of values for n and x_n, finding for each the smallest ε for which the linear program is feasible. For all $x_n = 5$, 10, and 20 and $n = 2^3, 2^4, \ldots, 2^{10}$ the same minimum, $\varepsilon = 3.18845$ was achieved, with the function

$$F(x) = \begin{cases} 1 - \dfrac{4}{3.18845}\,|x|, & -3.188445 \leqslant x \leqslant 3.18845, \\ f(x \pm 6.37688) & \text{otherwise.} \end{cases}$$

This constancy suggests that this is indeed the true minimum for the infinite linear program, and that for $\varepsilon < 3.18845$, $U(-\varepsilon, \varepsilon) \in C$ (from the argument following (2.20)). For $\varepsilon < 3.18845$, the dual to (2.37) provides an approximation to an (X, Y) distribution placing $U(-\varepsilon, \varepsilon) \in C$. However, as was observed earlier, this distribution, shown in Fig. 1, is not readily understood.

To prove the upper bound using the linear program, we can proceed as follows.

For any α, consider the function, $F(x) = F(x \pm 2\alpha)$ defined by

$$\begin{aligned} F(x) &= 1 - \frac{4}{\alpha}|x|, \quad -\alpha \leqslant x \leqslant \alpha, \\ &= -1 + \frac{16}{\pi^2} \sum_{k=0}^{\infty} \frac{\cos((2k+1)\pi x/\alpha)}{(2k+1)^2} \end{aligned}$$

and

$$\begin{aligned} G(\theta) &= \sup_{-\infty < x < \infty} F(x) + F(x + \theta) \\ &= \begin{cases} 2 - \dfrac{4}{\alpha}\,|\theta|, & -\alpha \leqslant \theta \leqslant \alpha, \\ G(\theta \pm 2\alpha) & \text{otherwise,} \end{cases} \\ &= \frac{16}{\pi^2} \sum_{k=0}^{\infty} \frac{\cos((2k+1)\pi x/\alpha)}{(2k+1)^2}. \end{aligned}$$

Clearly,

$$\int_{-\alpha}^{\alpha} G(\theta)\, d\theta = 0.$$

But,

$$\int_{-\infty}^{\infty} F(x)\, \varphi(x)\, dx = -1 + \frac{16}{\pi^2} \sum_{k=0}^{\infty} \frac{\exp(-((2k+1)\pi/\alpha)^2/2)}{(2k+1)^2} \geqslant 0$$

for $\alpha \geqslant 3.18845$. From (2.20), this means $U(-\varepsilon, \varepsilon) \notin C$ for $\varepsilon \geqslant 3.18845$.

3. Extremal measures in the uniform (doubly stochastic) case. Let S denote the set of measures on the square, $[0, 1]^2$, whose marginals on x and y are each

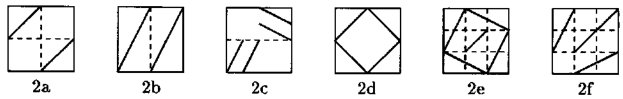

FIG. 2.

uniform on $[0,1]$. S is a convex and compact set of measures. Examples of $\mu \in S$ are given in Figures 2a–2e.

In each case μ is uniform on the line segments pictured (excluding the boundary of the square). We show 2a–2c and 2e are each extreme points of S. Indeed if $\mu = (\mu_1 + \mu_2)/2$ in both of 2a–2b there is only one point along vertical lines so that μ_1 and μ_2 have mass only in the support of μ so that $\mu_1 = \mu_2 = \mu$ and μ is extreme. In case 2c, this is true for $0 \leqslant x \leqslant \frac{1}{2}$ and for $\frac{1}{2} \leqslant y \leqslant 1$ separately, so a similar argument shows that 2c is also an extreme point of S. Since 2d can be written as a convex combination of 2a and its rotation by 90°, 2d is a point of S, but is *not* an extreme point of S.

In the remainder of this section we show that example 2e is an example of an extreme point of S even though its support has more than one point on each horizontal and vertical line. This seems to be the first such example. Lindenstrauss [7] gave a characterization of an extreme doubly stochastic measure μ on $[0,1]^2$ as one such that the set of all functions $f(x)+g(y)$, where f and g are in $L'([0,1],dx)$, are norm-dense in $L'(\mu)$. Lindenstrauss [7] gives an example of such a μ supported on two (piecewise affine) graphs, one over $[0,1)$ and the other over $[c,1)$ (for any c between 0 and 1) for which not every $L'(\mu)$ function can be written as $f(x)+g(y)$. Vitale [9] has given another characterization of the extreme measures in terms of approximability by measures obtained from Borel rectangles. Vitale [9] discusses the example, 2f (due to Shiflett), of an extreme doubly stochastic measure. Note that 2f does not have the property of 2e that each horizontal and vertical line has multiple intersections with the support and so it is easy to prove by the above argument that 2f is extreme.

Let P_h: $(x,y) \mapsto y$ be the horizontal projection from $\mathbf{R}^2$ to $\mathbf{R}^1$. Let $\frac{1}{2} < a < 1$. Divide the unit square into five regions R_i as illustrated in Fig. 3.

Regions R_1 through R_4 are congruent rectangles with side lengths a and $1-a$; region R_5 is a square with side $2a-1$.

Let

D_1	be the line segment from	$(0,\, 1-a)$	to	$(a,\, 0)$,
D_2	be the line segment from	$(a,\, 0)$	to	$(1,\, a)$,
D_3	be the line segment from	$(1,\, a)$	to	$(1-a,\, 1)$,
D_4	be the line segment from	$(1-a,\, 1)$	to	$(0,\, 1-a)$,
D_5	be the line segment from	$(1-a,\, 1-a)$	to	$(a,\, a)$.

The D_i are shown as solid lines in Fig. 3.

Let $\lambda\big|_{D_i}$ be one-dimensional Lebesgue measure supported by D_i, let $A(R_i)$ be the area of R_i and let $L(D_i)$ be the length of D_i and let

$$\mu = \sum_{i=1}^{5} \frac{A(R_i)}{L(D_i)} \lambda\Big|_{D_i}$$

be a weighted combination of the $\lambda\big|_{D_i}$.

THEOREM 1. *μ is an extreme point in S.*

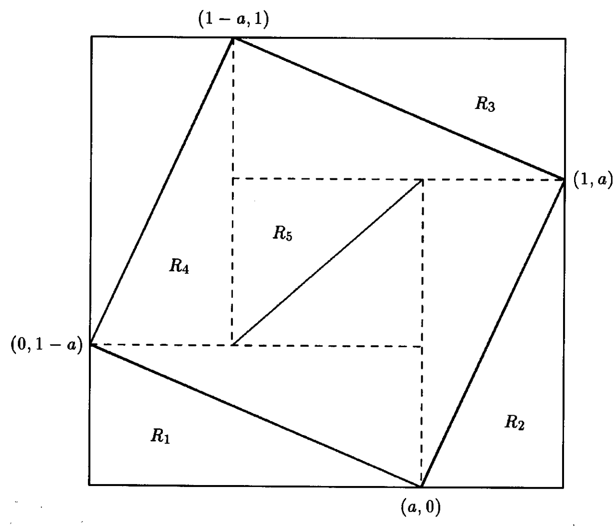

FIG. 3

Proof. That $\mu \in S$ follows from the fact that the marginals of $\frac{A(R_i)}{L(D_i)}\lambda\big|_{D_i}$ are the same as those of two-dimensional Lebesgue measure restricted to R_i, and since the unit square is the disjoint union of the R_i.

Now suppose $\mu = (\mu_1 + \mu_2)/2$ for $\mu_i \in S$. Clearly the μ_i are supported on $\cup_{i=1}^5 D_i$. Write $\mu_1 = \mu + \delta$, $\mu_2 = \mu - \delta$ for some the signed measure δ. To show μ is extreme it suffices by the previous discussion in this section, to show that δ restricted to $D_1 \cup D_2 \cup D_3 \cup D_4$ is the zero measure.

To this end define a map

$$T\colon \bigcup_{i=1}^4 D_i \to \bigcup_{i=1}^4 D_i$$

as follows. Restricted to D_1, T is horizontal projection into D_2, that is, if $(x,y) \in D_1$, then $T(x,y) = (u,y) \in D_2$. Restricted to D_2, T is vertical projection into D_3, so if $(x,y) \in D_2$ then $T(x,y) = (x,v) \in D_3$. Restricted to D_3, T is horizontal projection into D_4 and restricted to D_4, T is vertical projection into D_1.

It is clear that T restricted to each D_i is affine. Since $\frac{1}{2} < a$, $T(D_1)$ is a proper subset of D_2, $T(D_2)$ is a proper subset of D_3, and so on, so $T^4 = T \circ T \circ T \circ T$ acts as an affine strict contraction on each of D_i, $1 \leqslant i \leqslant 4$.

Since the marginals of μ_1 and μ_2 are the same as those of μ, we see that the marginals of δ must be the (one-dimensional) zero measure.

Suppose $A \subset D_1$ and let $A_0 = [0,1] \times P_h(A)$ be the cylinder set whose horizontal projection is the same as that of A. Then $\delta(A_0) = 0$. But since δ is supported by $\cup_{i=1}^5 D_i$, and since $P_h(D_1) \subset P_h(D_2)$, and since $P_h(D_2)$ is disjoint from $P_h(D_3 \cup D_4 \cup D_5)$, we see that $\delta(A_0) = \delta(A \cup T(A)) = \delta(A) + \delta(T(A))$, so for $A \subseteq D_1$, $\delta(T(A)) = -\delta(A)$. Similarly for $A \subseteq D_i$, $1 \leqslant i \leqslant 4$.

Thus, if $A \subseteq \cup_{i=1}^4 D_i$, $\delta(T^n(A)) = (-1)^n \delta(A)$. So suppose $A \subseteq D_1$ is an interval of length k, with $\delta(A) \neq 0$. Since T^4 is a strict contraction (with contraction coefficient $t < 1$, say) we can see that the length of $T^{4n}(A) = kt^n$ tends to 0 as $n \to \infty$. But this leads to a contradiction:

$$\mu_1\big(T^{4n}(A)\big) = \mu\big(T^{4n}(A)\big) + \delta\big(T^{4n}(A)\big) = \mu\big(T^{4n}(A)\big) + \delta(A)$$

so

$$\lim_{n\to\infty} \mu_1\big(T^{4n}(A)\big) = \lim_{n\to\infty} \frac{A(R_1)}{L(D_1)} \cdot kt^n + \delta(A) = \delta(A),$$

and similarly, $\lim_{n\to\infty} \mu_2(T^{4n}(A)) = -\delta(A)$. So at least one of μ_1 and μ_2 fail to be in S. So, for all intervals $A \subseteq D_1$, $\delta(A) = 0$, and δ is the zero measure on D_1, and hence on D_2, D_3 and D_4 as well.

REFERENCES

[1] R. Chen and L. A. Shepp, *On the sum of symmetric random variables*, Amer. Statist., 37 (1983), p. 237.

[2] W. Feller, *An Introduction to Probability Theory and Applications*, Vol. 2, 2nd ed., John Wiley, New York, 1966.

[3] P. C. Fishburn, J. C. Lagarias, J. A. Reeds, and L. A. Shepp, *Sets uniquely determined by projections on axes* I: *Continuous case*, SIAM J. Appl. Math. 50 (1990), pp. 288–306.

[4] S. Gutmann, J. H. B. Kemperman, J. A. Reeds, and L. A. Shepp, *Existence of probability measures with given marginals*, Ann. Probab., 19 (1991), pp. 1781–1797.

[5] M. G. Hahn, P. Hahn, and M. J. Klass, *Pointwise translation of the Radon transform and the general central limit problem*, Ann. Probab., 11 (1983), pp. 277–301.

[6] L. Khalfin and L. Klebanov, *A solution of the computer tomography paradox and estimating the distances between the densities of measures with the same marginals*, Ann. Probab., 22 (1944), pp. 2235–2241.

[7] J. Lindenstrauss, *A remark on extreme doubly stochastic measures*, Amer. Math. Monthly, 72 (1965), pp. 375–382.

[8] L. A. Shepp and J. B. Kruskal, *Computerized tomography*: *The new medical X-ray technology*, Amer. Math. Monthly, 85 (1978), pp. 420–439.

[9] R. A. Vitale, *Parametrization of doubly stochastic measures*, in Distributions with Fixed Marginals and Related Topics, V. Taylor, ed., IMS Lecture Notes, to appear.

[10] B. F. Logan, *in preparation.*